Lead-Acid Battery

Foreword

Most of the technology we use today in the Bell System has been developed since the invention of the telephone just less than a century ago. Our basic source of stand-by power, however, is still the lead-acid secondary cell, which was invented by Planté in 1859, and which, apart from the introduction by Fauré of pasted electrodes a little later, remains essentially unchanged in design to the present day. The files of the U. S. Patent Office are crammed with disclosures of other secondary cells, and some of these have achieved a limited market; but for most of our business, as for the automotive industry, the lead-acid cell still reigns supreme. The reasons for this are, of course, partly technical and partly economic. The lead-acid cell has an energy density figure that still looks reasonably good when compared with that of its competitors; it can be cycled many times without irreparable damage; the internal resistance is low; and it can be maintained fully charged for long periods by passing through it a small "float" current. At the same time, despite the somewhat limited use of mass production techniques, battery manufacturers have been able to produce cells and sell them to us at a price that compares favorably with the price of other secondary cells.

Nevertheless, the experience of the Bell System with lead-acid secondary cells has not been free from difficulty. The nature of our business places special demands on the performance of a secondary

cell and calls for a design different in important respects from that appropriate to a secondary cell for automotive use. Of course, the cells we buy from the various battery manufacturers are built to designs evolved by them with standby power service in mind. Nevertheless, there have been problems in the field. Maintenance costs have typically been higher and battery life shorter than one might reasonably expect.

Over the years, Bell Laboratories scientists and engineers have worked with the battery manufacturers to assist them in improving their designs. By the early sixties, however, it was clear that the performance of cells in the Bell System was still very far from satisfactory, and there began in fact to be clear indication that efforts to patch up the technology on one front had led to more serious difficulties on others.

In 1964 it was decided, therefore, that a look should be taken at the lead-acid design problem as a whole. This task, which began under F. J. Biondi in the Electronic Components Development Area, and has since expanded to include contributions from a number of other parts of Bell Laboratories and of the Western Electric Company, is now approaching completion. From this work has resulted a new design for the lead-acid cell, which, while retaining unchanged the basic electrochemical principles of the device, incorporates a number of important new design features.

Cells of the new design have been made in experimental numbers at the Nassau Smelting and Refining Company, a Western Electric subsidiary, in Tottenville, Staten Island, New York. This operation has allowed us to work with prototype production machinery under something approximating realistic factory conditions. It is our hope and expectation, however, that we shall find outside vendors interested in manufacturing cells according to the new design for sale to the Bell System.

While realistic life-tests of cells made according to the new design are as yet, of course, in a very early stage, predictions of life expectancy can be made with reasonable assurance on the basis of an understanding of failure modes and a program of accelerated aging. As a result of this work we could, in fact, have designed a cell that would have a predicted life inappropriately long for Bell System purposes, acquired, of course, at an inappropriately high cost. We have therefore picked, on the basis of some necessarily crude economic estimates, a figure of 30 years for the minimum projected life and aimed the design at that figure.

The series of papers that follow describes the technical aspects of

this development. The paper by D. E. Koontz, D. O. Feder, L. D. Babusci and H. J. Luer outlines the use of lead-acid cells in the Bell System and the difficulties experienced with the existing cells, and describes the new design. This paper makes clear that the principal problem with the existing design is the growth in size, with time on float, of the lead-calcium alloy positive grids. The following paper, by A. G. Cannone, Feder and R. V. Biagetti, shows how the grid growth problem has been substantially helped by the use of pure lead, and then still further reduced by designing the shape of the grid to obviate local buckling.

The third paper, by Biagetti and M. C. Weeks, describes the results of experiments with pure tetrabasic lead sulphate as the positive plate paste material instead of the proprietary mixture customarily used. Tetrabasic lead sulphate is shown to be less prone to shrinkage from the grid, because the lead dioxide crystals produced on forming tend to exist as a dense tangle of relatively large crystallites. The next paper, by P. C. Milner, summarizes what is known of the float characteristics of lead-acid cells, setting up a quantitative framework within which the float behavior of any lead-acid cell may be described by quoting a relatively small number of experimentally determinable parameters. T. D. O'Sullivan, Biagetti and Weeks next present the results of a series of electrochemical measurements on the new cell. They describe *inter alia* the float behavior and relate their results to Milner's analysis. They then go on to give the results of field tests on early models of the new cell. The sixth paper, by T. W. Huseby, J. T. Ryan and P. Hubbauer, describes the work leading to the design of the jar and cover for the new cell. During the course of this work it was found that particular attention must be given to the presence of trace amounts of certain impurities, either in the stabilizer for the Polyvinyl Chloride or as solvents that may be introduced in certain ways of sealing the cover to the jar.

The electrochemical activity of these trace impurities is reported in the succeeding paper by A. D. Butherus, W. S. Lindenberger and F. J. Vaccaro, who quote maximum tolerable limits for a number of the impurities likely to be present. The sealing of the cover to the jar is discussed in the paper by D. W. Dahringer and J. R. Shroff, who developed a heat-sealing technique that is mechanically sounder than a solvent seal and also avoids the introduction of excessive amounts of electrochemically active impurities.

The next paper, by L. H. Sharpe, Shroff and Vaccaro, describes the solution of the post seal problem, in which, by means of a flexible

bellows, the post is given some freedom to move relative to the cover without stressing the seal. The tenth paper, by R. H. Cushman, describes a new technique for joining together the positive plates at their peripheries, using a continuous fusion bonding effected by an electrically heated tip. The final paper in this series, by Luer, indicates how cells of the new design can be incorporated into telephone plant to replace the present rectangular cells without the need for extensive redesign of the space set aside for accommodation of cells.

One word on terminology. To an electrochemist, a single jar containing two electrodes is a cell, and a number of these connected together in series constitutes a battery. In the Bell System we tend to refer to a single cell as a *battery*; a number of cells connected together is often called a *string*. Rather than attempt to impose on this set of papers a uniform terminology, I have left usage in this regard to the individual authors to decide. I trust, therefore, that electrochemists will be forgiving if they frequently find a cell called a battery, and power engineers similarly tolerant if they sometimes find a battery called a cell.

C. G. B. GARRETT

Lead-Acid Battery:

Reserve Batteries for Bell System Use: Design of the New Cell

By D. E. KOONTZ, D. O. FEDER, L. D. BABUSCI, H. J. LUER

A new cylindrical lead-acid battery has been developed to provide reserve power to sustain telephone service during power-failure conditions. We discuss in this paper the usage of batteries in the modern telephone plant, review the difficulties encountered with existing cell designs, describe in detail special features of the new design, and outline the scope of the overall development program.

I. INTRODUCTION

For more than 100 years, lead-acid batteries have dominated the field of practical rechargeable batteries. In its modern form, the lead-acid battery is derived from Faure[1] and Sellon[2] who introduced the concept of a conducting lattice-like grid to provide electrical contact and mechanical containment of the particulate pastes of the positive and negative electroactive species. Despite a century of evolution which has seen this battery system optimized for automotive starting, lighting and ignition (SLI) service and become the primary industrial battery system for both stationary (communication) and traction (fork-lift truck, etc.) usage, modern cell designs differ only slightly from Faure's original proposals.

In contrast to the deep-cycle service of industrial truck batteries, and the high-rate discharges typical of batteries for SLI usage, batteries for telephone service are continuously trickle-charged, or "floated" under conditions designed to maintain them in a full state of charge, ready for instantaneous use, should commercial power fail. Discharges are infrequent and at relatively low rates. Bell System requirements emphasize long life and high reliability in the unique float charge use-mode and strive to avoid the economic penalties of excessive maintenance.

Despite these major differences in use-mode, batteries for telephone service differ basically only in their wide size range (50 AH to 7000 AH) and lower specific gravity acid (1.210) from those used for SLI or traction service. Figure 1 shows a representative telephone stationary battery design. The major structural element is the lead alloy grid, a lattice-like frame which serves as the mechanical retainer and electrical contact to the positive and negative active material pastes. Lead oxides mixed with sulfuric acid and water, are pasted into these grid frames and, after drying, converted via electrolytic oxidation or reduction to the positive lead dioxide or negative "spongy" lead electrochemically active species. Microporous separators are interspersed between each pair of positive and negative plates to provide electronic insulation while permitting ionic transport. The separators may be supplemented in some designs by fiberglass mats designed to entrap positive active
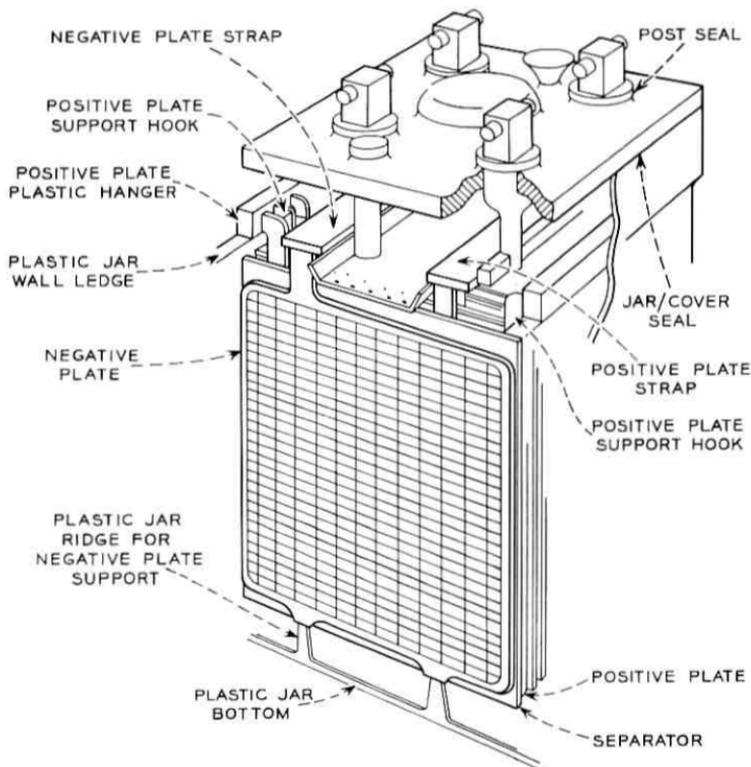

Fig. 1—Typical Pb-acid battery for telephone reserve use.

material which may shed during charge-discharge cycling. In cell elements made up of groups of positive plates with interleaved negative plates and separators, electrodes of like polarity are electrically and mechanically joined by torch-welding to a lead-alloy connecting strap. This in turn is joined to a terminal which projects, via a feed-through seal, through the cell cover, forming the external terminal contact. The entire vertically disposed cell element may either rest on the cell bottom (suitably elevated by ridges to allow collection of shed active material sediment without causing inter-electrode shorts) or be suspended from the jar wall or cell cover.

Because the element weight is supported by the grid frames, the structural integrity of the grid frames has required that soft lead be metallurgically hardened, usually by the incorporation of 3–12 percent Sb, which forms a binary eutectic at 13 percent Sb and produces an increase in tensile strength from 1780 to 7280 psi, and an order of magnitude increase in creep strength. While early workers recognized the effectiveness of Sb in providing the required degree of mechanical hardening (and improving the casting qualities of the grid), the electrochemical activity of the Sb additive was not clearly identified and understood until the mid 1930s, when its influence on the behavior of the negative plate was elucidated by Crennel and Milligan.[3] (The significant reactions which govern the behavior of the lead-acid battery in discharge, recharge and overcharge are shown in Table I.) Haring and others[4] documented the corrosion and dissolution of Sb from the positive grid, its transport through the separator as a complex Sb ion and its redeposition on the negative plate. There it provided sites which facilitated hydrogen evolution (depolarization) on charge and promoted a local action type of self-discharge which ultimately resulted in irreversible sulfation and failure of the negative plate. Thomas and Haring[4] proposed the use of electronegative alkaline earth elements as nondepolarizing alloying additions and in 1948 reported on successful long-term experiments with batteries whose grids were hardened by inclusion of 0.05 to 0.10 percent Ca.[5] These workers demonstrated that cells fabricated with these Pb/Ca alloy grids, when maintained continuously on trickle or "float" charge showed no depolarization or sulfation of the negative plates after a nine-year float test. In this test, a new life limiting process resulted, that is, corrosion and growth of the positive grid in the highly oxidizing environment of float charge usage. Their life tests indicated a critical sensitivity of the corrosion and growth processes to small variations in the alloy composition and to the metallurgical soundness of the casting.[6] The Ca alloy addition in-

TABLE I—REACTIONS IN THE LEAD-ACID BATTERY

Primary Charge-Discharge Reactions

At the Positive Electrode:

$$PbSO_4 + 2H_2O \underset{\text{Discharge}}{\overset{\text{Charge}}{\rightleftarrows}} PbO_2 + 4H^+ + SO_4^- + 2e^-$$

At the Negative Electrode:

$$PbSO_4 + 2e^- \underset{\text{Discharge}}{\overset{\text{Charge}}{\rightleftarrows}} Pb + SO_4^-$$

On Overcharge

At the Positive Electrode:

$$2H_2O \overset{\text{Charge}}{\rightleftarrows} O_2 + 4H^+ + 4e^-$$

$$Pb + 2H_2O \overset{\text{Charge}}{\rightleftarrows} PbO_2 + 4H^+ + 4e^-$$

At the Negative Electrode:

$$2 H^+ + 2e^- \overset{\text{Charge}}{\rightleftarrows} H_2$$

creased the difficulty of grid casting because of its narrow melting range and caused significant analytical and control problems to be introduced into the casting process. The manufacturing problems were successfully overcome and batteries containing lead-calcium grids were introduced into widespread Bell System use in 1950.

During this same period container materials suitable for use in the acidic electrolyte evolved from fragile and expensive glass and ceramics through hard rubber into current polystyrene-based polymers capable of high-speed injection moldings which permitted wide flexibility in jar-and-cover design and resulted in significant economies. Adhesive sealing of containers and covers and sealed feed-through terminals produced cells which eliminated the possibility of internal element maintenance and repair. Elimination of Sb and its depolarization of the hydrogen evolution reaction allowed cells to float at substantially lower currents resulting in a marked decrease in frequency of water additions. These changes stimulated expectations of decreased battery maintenance by telephone company personnel.

In the 20 years since the introduction of the lead-calcium battery, Bell System battery usage has increased by more than an order of magnitude and has been accompanied by significant changes in power plant design and usage.

We discuss in this paper the ways batteries are used in the modern

telephone plant, review the difficulties that have been experienced with existing cell designs, describe a new cell which has been designed to minimize these problems, and outline the scope of the program for its development.

II. BATTERY USAGE

In order to assure uninterrupted telephone service, a continuous, uninterrupted supply of electrical power must be provided. This power is normally purchased from the electric utility serving the area of operation but it is clear that sufficient power must be available to run the plant even during periods when the commercial power is unavailable. Each telephone installation must therefore be provided with a source of reserve power having sufficient capacity to carry the plant through the duration of commercial power interruption.

A typical central office power system is shown schematically in Fig. 2. The telephone plant load is primarily dc and it is this portion of the installation which must be provided with no-break power. A small amount of ac is often included in the plant load. In normal operation the ac load is supplied directly by the commercial source and the dc load by a system of rectifiers driven from the commercial source. Building services such as general lighting and air-conditioning are also supplied directly. The battery is kept connected to the plant at all times and is floated at full charge by the rectifiers.
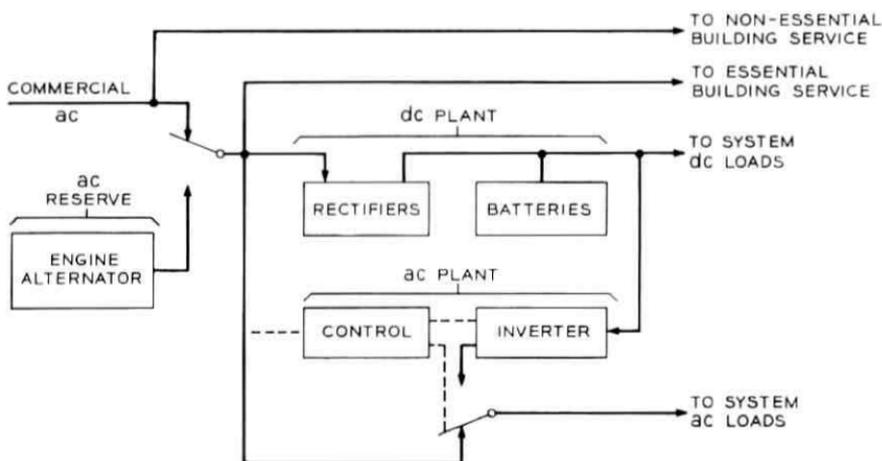

Fig. 2—Telephone office power plant.

In the event of a failure of commercial power, the commercial source is disconnected from the plant and an appropriate alarm is registered. The reserve engine-alternator is then started (either automatically or manually) and, when up to speed, connected to the load. Depending on conditions, the time interval between the onset of the power failure and the load pickup by the reserve engine can vary from less than a minute to more than an hour. During this time interval, the battery provides all the dc power to run the telephone plant. AC power for the telephone equipment is also supplied by the battery through dc–ac inverter plants. This arrangement allows the dc load to be drawn from the batteries without interruption and with minimal switching transients.

The amount of battery reserve which must be provided is highly dependent on the precise conditions of the installation. The significant factors are plant voltage, load current, and required battery reserve time.

While central offices operate principally at 24 and 48 volts, there are a variety of other voltages needed for PBXs, transmission systems and other special services. Table II is a tabulation of the dc voltages for which standard battery plants are presently provided. It also shows the number of cells connected in series to achieve these voltages. Roughly speaking, the number of cells is half the plant voltage (2 volts per cell). However, many plants are provided with additional emergency

TABLE II—DC VOLTAGES OF STANDARD BATTERY PLANTS

Nominal Voltage	Number of Cells
12V	6
24V	9 10 11 12 13 14
48V	23 24 27
130V	63 66 67 69 70
152V	70

TABLE III—POWER OUTAGES IN TELEPHONE OFFICES

No. of Outages Per Year	Duration of Each Outage (% of Total Reported Outages)			
	0–10 Min.	10–60 Min.	1–3 Hr.	>3 Hr.
1–5	37	35	12	5
6–10	3	1	0.5	0
10–20	1	0	0	0
>20	5	0.5	0	0

or "end" cells which are switched onto the end of the battery string during longer discharge periods to maintain the plant voltage within the required limits. This permits the cells in the main battery to discharge to a lower voltage (1.75 V/cell) and thus utilizes the battery more completely.

The current drain on these plants can be anywhere up to 10,000 amperes. The battery string is designed to supply this load for the duration of the expected outage. The amount of reserve time provided depends largely on the local office environment. For plants with good reserve engine back-up located in an area where power failures are historically rare and of short duration, the reserve time provided may be as low as two hours. On the other hand, unattended remote locations without engine back-up (community dial offices, mountaintop relay stations, and so on) may be engineered to run for days on the battery alone. Results of a recent battery survey provided significant statistical information on power outages. These are tabulated in Table III. Large area power failures, such as the Northeast blackout of November 1965, provided practical demonstration of the importance of the reserve battery plant.

The cells we currently use for this service have capacities ranging from 50 to 7000 ampere-hours as tabulated in Table IV. Generally speaking, the smallest available cell having enough capacity to provide the necessary reserve time is used. If the required reserve capacity exceeds that available from a single battery, as many as 8 multiple strings are used in parallel.

Cells of 4000- to 7000-ampere-hour capacity are used in the larger offices. These "tank" cells are of hard rubber case construction and are floor mounted. Cells in the 180- to 1680-ampere-hour range have polystyrene-based plastic cases and are mounted on shelf-type battery racks, generally two tiers high. The smaller 50- and 100-ampere-hour cells currently have polystyrene-based or polypropylene plastic cases

TABLE IV—CELL CAPACITIES USED

| Ampere Hours |
| --- |
| 50 |
| 100 |
| 180 |
| 240 |
| 300 |
| 420 |
| 540 |
| 660 |
| 840 |
| 1080 |
| 1320 |
| 1680 |
| 4000 |
| 5000 |
| 6000 |
| 7000 |
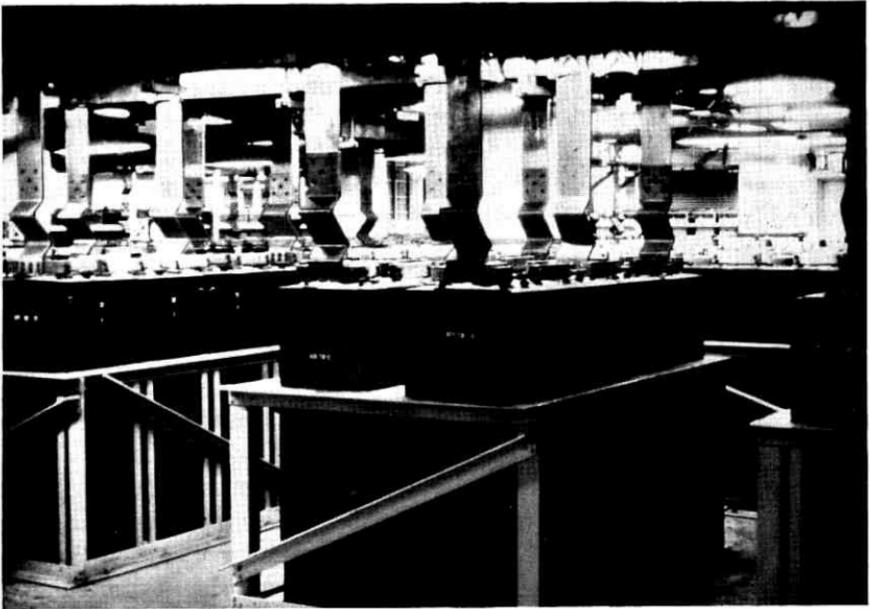
Fig. 3—Central office battery plant with 7000-AH Tank cells.
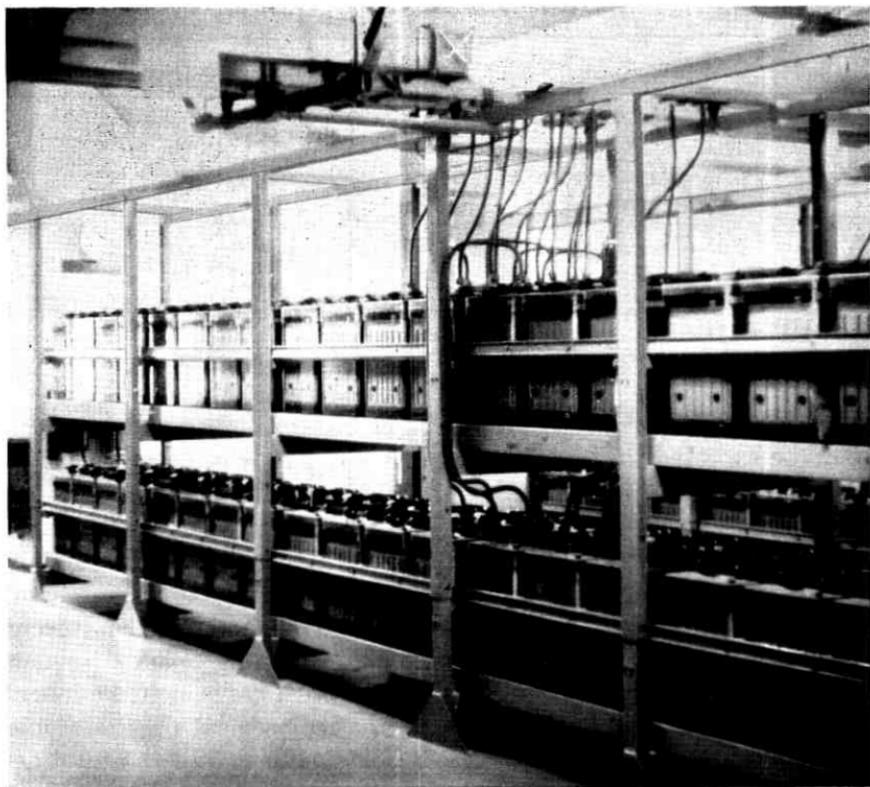
Fig. 4—Central office battery plant with 1680-AH plastic jar cells.

and may be shelf-mounted either on battery racks or in relay racks. Typical central office battery plants are shown in Figs. 3 and 4.

More than half the cells we use are in the 180- to 1680-ampere-hour range. Table V shows the number of lead-calcium cells of this type purchased by the Bell System in 1968. Lead-antimony cells in these sizes accounted for less than 5 percent of the 1968 business. The use of the Pb/Ca cells seems fairly well distributed over the range of sizes with two exceptions. There is a small bulge at 300 to 400 ampere-hours and a much larger one at 1680 ampere-hours. The 1680-AH cells provide more capacity per dollar than any of the other sizes available (including the tank cells) but their use in plants requiring more than five parallel strings is not recommended because of increased maintenance costs. If the maintenance cost of the new cell is lowered as anticipated, this

TABLE V—LEAD-CALCIUM CELLS PURCHASED BY BELL SYSTEM IN 1968

Ampere Hours	No. of Cells
180	7000
240	7150
300	10250
420	14850
540	6600
660	6400
840	5000
1080	5000
1320	4000
1680	43000

Total   109,250

restriction may be relaxed and purchases of the 1680-ampere-hour cell would be expected to rise at the expense of the tank cell.

It is estimated that a total of 1,000,000 to 1,250,000 cells in the 180- to 1680-ampere-hour range are presently in service. It is anticipated that the Bell System demand for these cells will remain at approximately the present level in the immediate future. By the mid 1970s purchases of the new cell are expected to reduce the demand for the current product to virtually zero. The replacement of existing cells by the new cells will continue at least until the mid 1980s after which the longer life of the new cells will take effect and the replacement market should essentially disappear for many years. It is doubtful that the total demand for the new cell will drop however, since the expected demand for new plant should take up the slack.

In this size range, 11 different capacities are furnished in both lead-calcium and lead-antimony grid systems by three manufacturers. The multiplicity of available battery configurations, materials systems and element designs which result from our "end-requirement" specifications create significant problems in the design of mounting and interconnecting arrangements needed to incorporate them into the telephone plant and result in significant confusion in development of suitable practices for their maintenance and use.

In the early 1950s, the Bell System battery plant was relatively small and usage was uncomplicated. Rapid system growth has resulted in continuous expansion to keep pace with new system demands for power and has introduced requirements for offices resistant to earthquake and nuclear attack.[7] Reliability, long life, low maintenance, and thus low annual costs have become increasingly important as size and variety of plant usage has increased. Realization of these objectives

has been complicated by the complexity and multiplicity of installation configurations and compromised by performance deficiencies of the cells currently in use.

III. BATTERY PERFORMANCE IN THE BELL SYSTEM

The introduction of the lead-calcium alloy grid into the manufacture of Bell System batteries in 1950 was accompanied by a significant departure in specification practice. In addition to providing end-requirement electrical and mechanical specifications and "design life" objectives, close control was established over the composition of the Ca grid alloy. This was soon followed by specification and control of composition and properties of the polystyrene-based polymers introduced for battery jars and covers. Significant BTL materials development effort was involved in both these programs. Control of all other details of cell design, materials selection and manufacture was retained by the manufacturers, who were free to initiate changes for cost reduction and product improvement, subject to Western Electric Company and BTL approval. The lack of a proven accelerated aging technique made it difficult to predict the ultimate life of the original designs, or the impact of these changes on the long-term performance of cells which had an anticipated 25-year life expectancy. Field performance problems usually appeared slowly over periods of five to ten years and such early failures as did occur were often considered as isolated manufacturing defects in the absence of widespread field complaints.

In the early 1960s, several serious fires in telephone battery plants became the object of significant telephone company and BTL study. Extensive postmortems of damaged batteries, coupled with simulations in a mock-up battery plant at Murray Hill established conclusively that electrolyte leakage to grounded metal battery racks could produce arcing which might result either in an explosion of hydrogen/oxygen mixtures trapped beneath the cell cover, or in direct ignition of the flammable polystyrene-based plastic jar-and-cover materials. Acid leakage from faulty jar-cover or terminal-cover seals, or from cracked jars was found to be particularly hazardous in higher voltage (> 48 V) battery plants.

More detailed examination revealed that in most cases, leakage could be directly attributed to stresses resulting from corrosion and growth of the positive grid.[8] Corrosion causes the lead-alloy grid to be continually oxidized to PbO_2 . The larger specific volume of the oxidized material creates stresses which cause the grid to grow, crack and
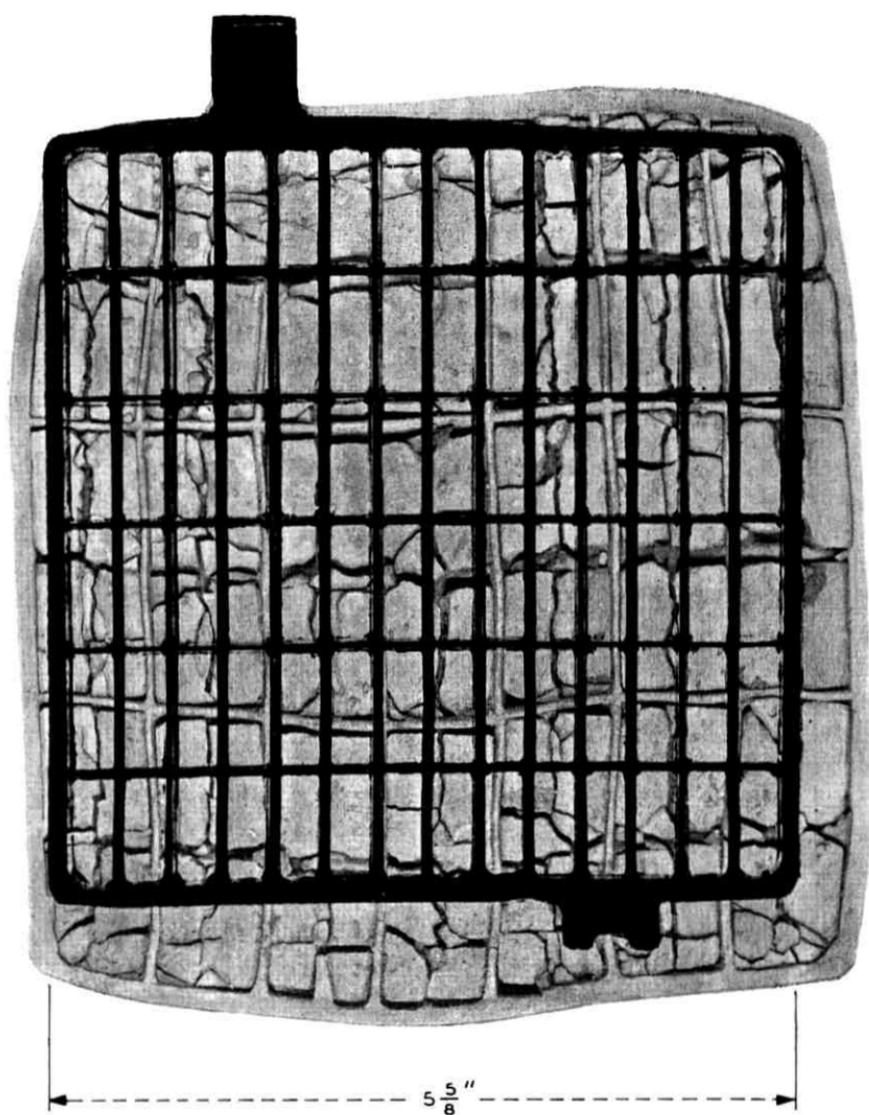
Fig. 5—Corroded positive plate with original grid overlaid to indicate initial dimensions.

eventually lose contact to the active material. The extent of this growth is seen in Figure 5. It shows a corroded and grown positive plate with the original lead-calcium alloy grid overlaid to indicate the initial dimensions. Simulated growth-stress studies revealed the stress-sensitive, brittle polystyrene jars to be incapable of resisting these growth stresses, once grid growth had progressed to the point of contact with the jar walls. Initial stress-induced crazing (Fig. 6) was usually followed by jar fracture (Fig. 7). Likewise, vertical growth stressed and ruptured jar-cover and post-cover seals (Fig. 8). Many examples of each type of failure were found in field examination of telephone battery plants.

Recognition of the mechanical consequences of grid growth stimulated additional concern over the electrochemical integrity of the telephone battery plant. The loss of grid-pellet contact which occurs during grid
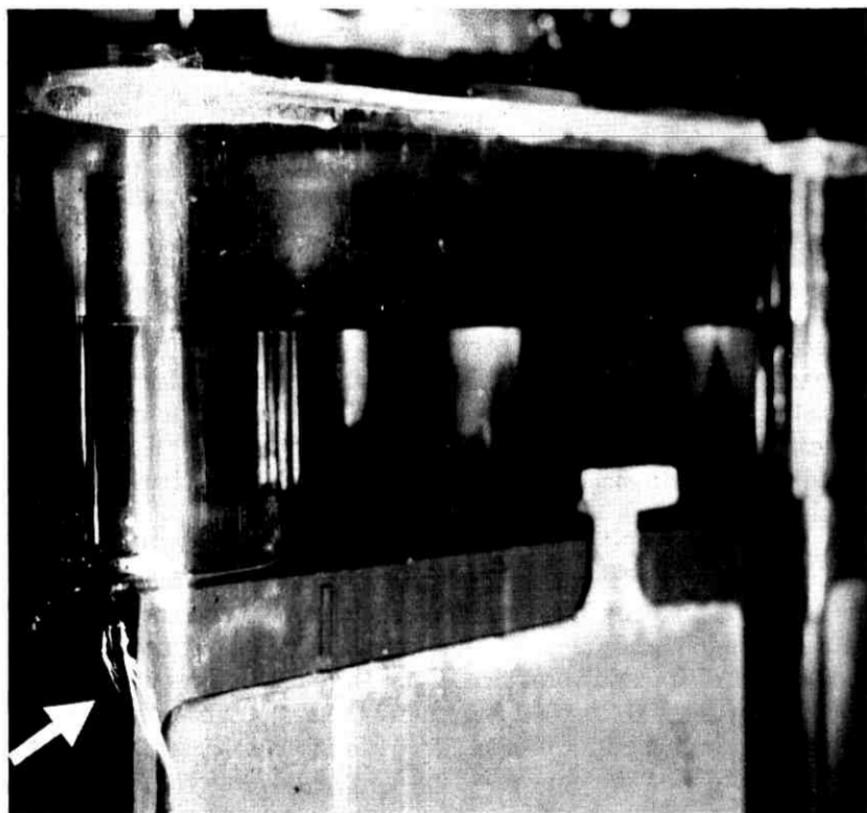

Fig. 6—Crazing in plastic jars.
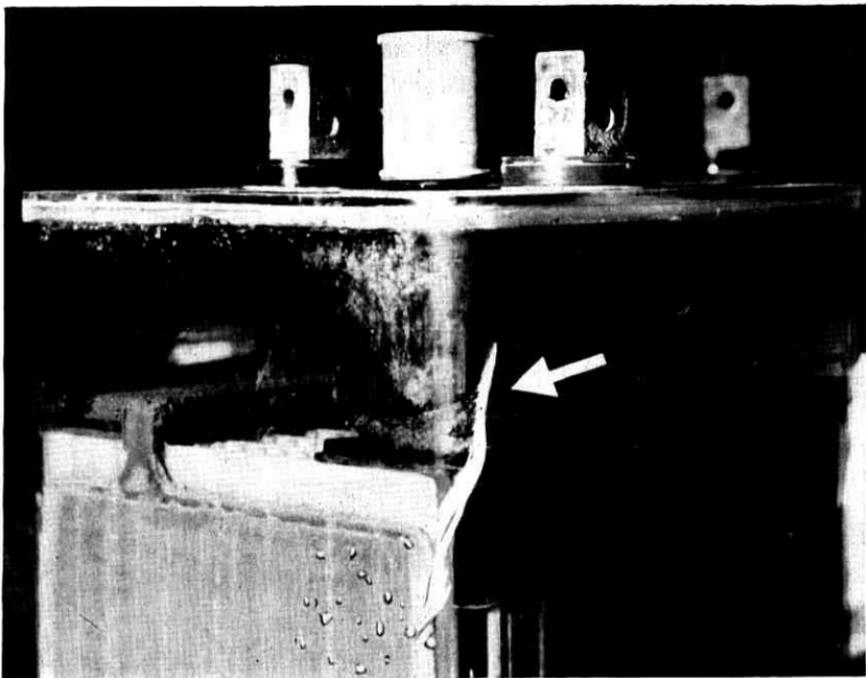
Fig. 7—Crack resulting from craze of Fig. 6.

growth is clearly visible in Fig. 5 and results in severe capacity loss.

Since capacity tests were not routinely performed by the telephone companies to assess the capabilities of their reserve battery plant, a pilot study involving about 50 cells of various ages and manufacturers was carried out with the Pacific Telephone Company. This study revealed capacity to be degrading much more rapidly than had been anticipated.

In order to establish a more accurate understanding of the existing capabilities of the telephone reserve battery plant, an extensive survey was then organized by BTL and implemented with the cooperation of AT&T. Random sampling of cells in all telephone companies (including Bell of Canada) provided discharge information on more than 5000 cells in the 180- to 1680-AH range. Data was taken by telephone company personnel on forms specifically designed to facilitate transfer to punch cards for subsequent computer analysis. Information on cell type, age and manufacturer, battery plant location, type and usage, power-failure statistics and many other parameters were obtained in addition to cell capacity, mechanical conditions and maintenance in-

formation. Computer correlation, analysis and display techniques were used extensively in reviews of this material for both telephone company personnel and individual battery vendors. The results indicated that the life expectancy of lead-calcium cells in system usage would generally fall far short of the anticipated design goal of 25 years to 90 percent of initial capacity with the best designs likely to achieve an average life of no more than 15 years to 75 percent capacity. There was considerable variability in performance and aging characteristics among the various suppliers' products with the earlier, more conservatively engineered designs showing the best long-term performance. In addition it was found that design changes had generally resulted in significant degradation of capacity-aging characteristics with some cells (repre-
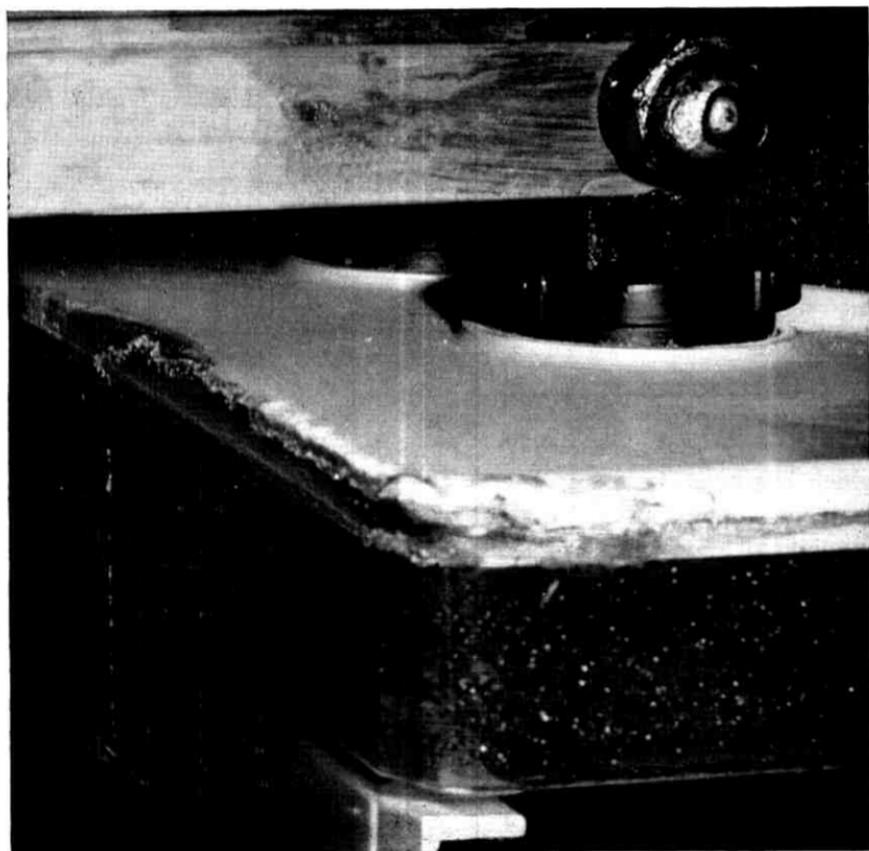

Fig. 8—Cracked and leaking jar-cover seal.

senting a large fraction of the installed plant) fading to less than 50 percent capacity within seven years. Mechanical failures were found to be common to all designs and in many cases had passed unrecognized by telephone company personnel except in catastrophic cases.

While demonstrating the degraded capability of the telephone battery plant, the battery survey documented the need for closer specification and control of cell design, materials and processes in order to insure performance in accord with system objectives. These two factors, coupled with inventory, supply, installation and usage problems resulting from the multiplicity of codes and configurations, stimulated a thorough reexamination and revision of the entire philosophy of Bell System battery design, procurement and use, with the ultimate goal of minimizing the annual cost of battery power to the telephone plant. Specifically, a multifaceted program was initiated in 1964 with the following objectives:

(*i*) To design a lead-acid battery optimized for telephone float service which would incorporate long life (> 30 years) and freedom from mechanical defects.

(*ii*) To provide for close specification of and control over design, materials, processes and components, with the ultimate objective of providing a device in which all facets were controlled and understood.

(*iii*) To introduce standardization and reduction of codes and configurations by adoption of a single unified design.

IV. FEATURES OF THE NEW DESIGN

This paper and those which follow, describe the new lead-acid cell which has been designed; discuss the new processes which have been developed; and provide details of cell performance characteristics, accelerated aging studies and field trial behavior which document its compatibility with the existing telephone power plant and provide a basis from which predictions of life and reliability may be made.

The cell design contains several novel features:

(*i*) The grids are made of pure lead for minimal corrosion and growth and are stacked in a self-supporting structure to minimize distortion.[8]

(*ii*) The grids are conically shaped, for maximum strength, with grid members specially designed to grow in a cooperative, phased manner which provides improved grid/paste contact throughout life and results in a significantly improved capacity aging characteristic.[8]

(*iii*) A new positive paste material, tetrabasic lead sulfate with rod-

like particles which interlock for maximum mechanical stability has been introduced.[9]

(*iv*) Jar and cover are made of a new, transparent, flame-retardant rigid polyvinyl chloride (PVC) material which combines low-cost fabrication, improved impact and craze resistance and increased ease of sealing.[10]

In order to allow reliable fabrication and performance of this design, several new processes have been introduced. These include:

(*i*) A lead-joining technique which provides welded flaw-free positive plate interconnection bonds in multiple, at high speeds compatible with economic manufacturing technology.[11]

(*ii*) A technique for *in situ* casting of a lead-antimony connection rod which provides mechanical and electrical interconnection of the negative plates.

(*iii*) A heat-sealing technique which utilizes directed black body absorption of infrared radiation for localized melting and bonding of the jar-cover sealing surfaces, providing a jar-cover seal capable of supporting over 100 times the weight of the battery.[12]

(*iv*) A redundant post-cover sealing system which provides a rigid epoxy corrosion-restraint sheath on the lead post and is flexibly coupled to the cover to allow stress-free movement of the cell element within the jar.[13]

Central to the development of this new design and the new processes which it incorporates and to our ability to predict its long-term performance was the successful development of techniques for accelerated aging of the electrochemical element, associated materials and supporting systems. Use of elevated temperature aging to predict corrosion, growth and capacity behavior of the electrochemical element, is discussed in detail.[8] Accelerated aging techniques for predicting plastic creep,[10] acid dissolution and interaction effects are also described.[14] Accelerated voltage stressed post-seal corrosion studies are also reported.[13]

Detailed studies of cell charge-discharge and float characteristics are presented in addition to results of initial field trials in working telephone offices.[15]

In this section of the paper we describe the design and construction features of the new cell in some detail. The detailed technical programs which resulted in specific design features are described in the companion articles in the series.
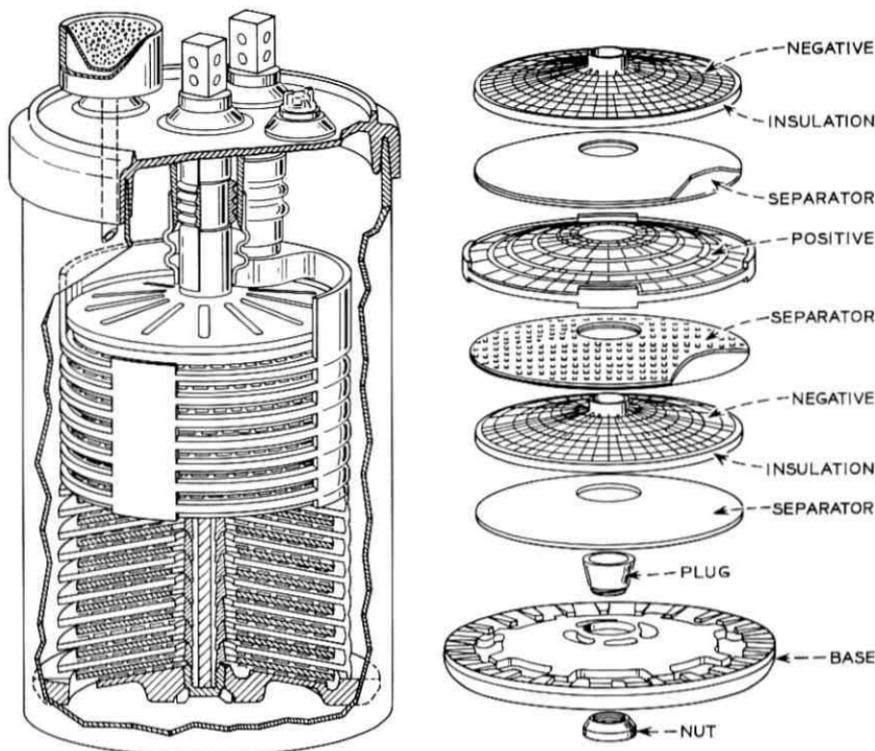
Fig. 9—The Bell System Battery—cutaway and exploded view.

V. CELL DESIGN

Figure 9 shows the new cylindrical cell design in both cutaway and exploded views. Figures 10 and 11 show the pure-lead positive and negative grids in more detail. The plates are conical, cupped to a 10° angle, and stacked, pancake fashion, one upon the other. Microporous rubber separators and fiberglass mats, conically shaped, are interleaved between each pair of positive and negative plates to provide electronic insulation and prevent shorting while allowing ionic and gas transport. Plastic insulating details protect the outer ring of the negative and inner hub of the positive from metallic shorts. The element stack nests on a conically shaped hard rubber base, into which is inserted a blind lead plug secured by means of a hard rubber nut. The upper surface of the lead plug is shaped to receive the conical central hub of the bottom negative plate. The other negatives nest similarly at the center via their conical hubs, forming a hollow central core. A hard

rubber cap provides compression to the top negative plate. The lower end of the negative post is conically shaped to nest in the hub of the top negative plate. After stack assembly and after insertion of a copper rod for conductivity, the entire negative group, including plates, plug and post are interconnected by a so-called "center pouring" process. This process, shown schematically in Fig. 12, introduces molten lead-antimony alloy at the base of the core through a tube which is withdrawn at a controlled rate. Heated nitrogen introduced through the same tube serves both to preheat the core and to prevent oxidation of the lead surfaces. After filling and solidification, this results in a mechanical and electrical connecting rod which is metallurgically bonded to the lead hubs.

The positive plates have vertical connector tabs at their periphery and are joined after assembly by a welding process. All early cells were made with a heliarc welding process which provided only a partial depth bond and was quite time consuming. A new bonding technique has been developed to provide full depth, flaw-free bonds at speeds compatible with manufacturing technology.

In this technique the cell stack is turned with its axis horizontal and
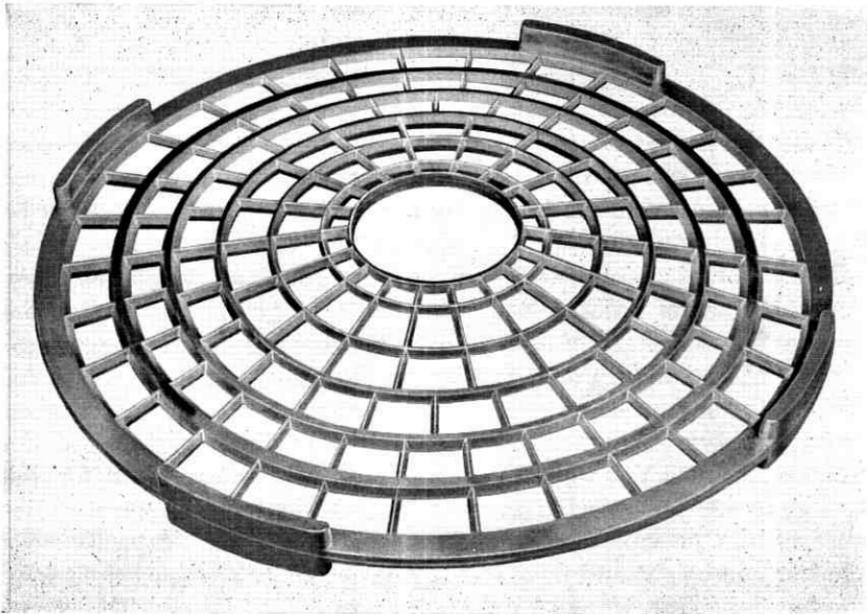

Fig. 10—Positive grid.
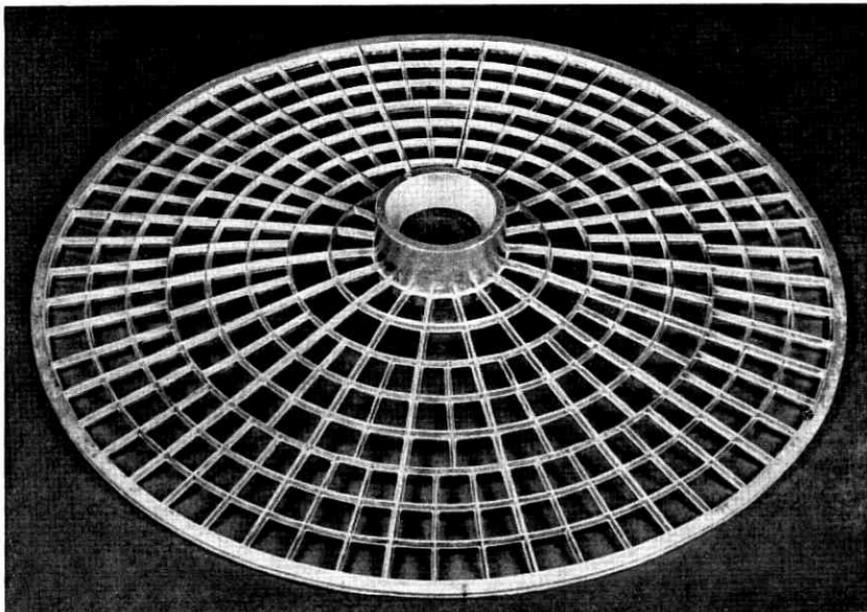
Fig. 11—Negative grid.

a heated tip is passed through from underneath to melt the junction of each pair of positive connector tabs. A curved insulating block which contains the heated tip acts as a traveling mold to contain the molten lead and provide a solidified bond of the same curvature as the original connector tabs. Multiple heads allow at least ten connections to be bonded simultaneously.[11]

Percolation of gas and electrolyte occurs through an annular chimney around the center negative connector and through spaces between the tabs on the periphery of the positive plates. The ten degree conical angle combined with proper venting at the central annulus allows for efficient circulation. The welded unitized element, compressed between the hard rubber cap and base, rests in a simple cylindrical jar made of a flame-retardant PVC which serves primarily as a container for the electrolyte and is subjected therefore only to hydrostatic forces.[10] Because of the pancake stacking arrangement, different capacity cells differ only in height and have identical-sized grids. The jar height may therefore be simply varied to accommodate different sizes. The cover, also made of PVC, provides a heat-sealed shear bond for maximum strength and reliability of the jar-cover seal.[12] For ease of clean-up

and maintenance a dam at the outer edge of the cover contains any electrolyte spillage which might result from specific gravity measurements. In addition, the cover overhangs the jar, thus providing an umbrella action to minimize the possibility of electrolyte spillover accumulating at the seal area. (Acid accumulation at jar-cover seals has proved to be a major cause of fire in batteries of conventional design throughout the Bell System.) Post seals incorporate a rigid epoxy sheath to provide a long leakage path which is coupled to the cover via flexible rubber bellows backed up by a piston-cylinder, accordion-ring type seal to allow for vertical movements which result from any shock incurred during shipment and installation and for vertical element
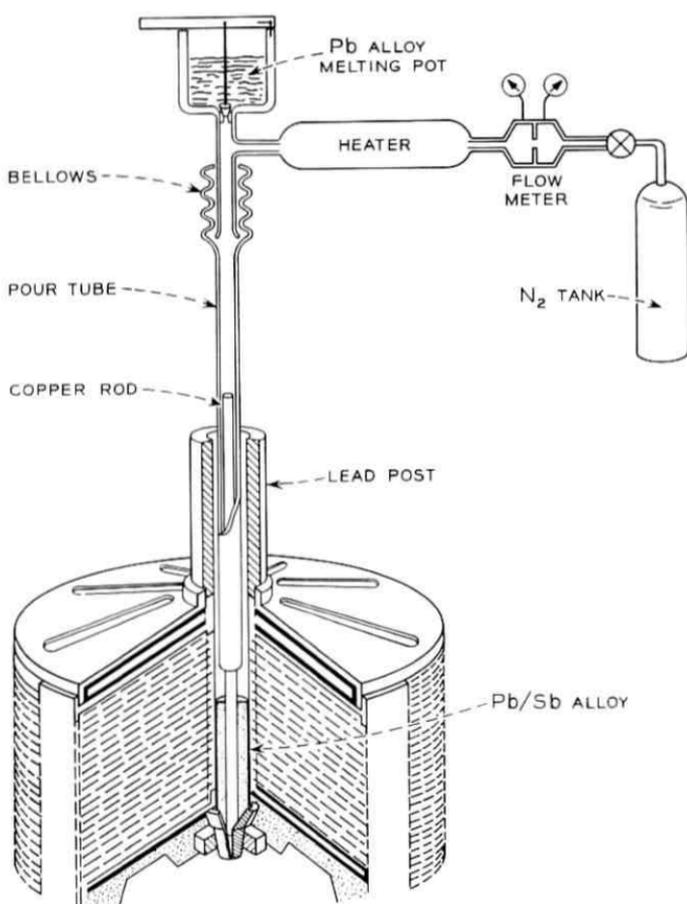

Fig. 12—Centerpouring process—schematic.

growth without overstressing the cover, jar-cover or post seals.[13] The vent incorporates a microporous diffuser which also serves as the funnel for water additions. This has the advantage of retaining all acid condensation and of allowing frequent rinsing and dissolution of salt deposits during normal water addition maintenance. The threaded copper rod inserts have been utilized directly as the terminals for cell interconnections in the early field trial cells. In later designs they will be covered by lead-antimony terminal details cast onto the positive post during fabrication and onto the negative as part of the center pouring operation.

The circular design results from our attempt to minimize the corrosion and control the growth at the positive electrode which is the major life-determining process of lead-acid batteries in telephone float service. The circular conical plates, stacked pancake fashion and joined into a rigid unitized structure, provide a structure of sufficient strength and of minimal loading on the individual elements to allow us to capitalize on the superior corrosion and growth behavior of pure lead which is too soft for use in conventional designs.[8] In this design the stacking arrangement results in loadings of a few pounds per square inch in the most extreme case, well within the long-term creep capability of the pure-lead material. In addition, the circular design allows us to control the contact between the lead grid and the positive paste pellet in such a way as to minimize the capacity loss which occurs in rectangular grids as growth causes the grid to break away from the active material pellet.[8]

A further improvement in grid/paste contact is made possible by the introduction of a new positive active material, tetrabasic lead sulfate.[9] A process has been developed to synthesize this material in the form of interlocking, rod-like particles which minimize shedding and provide optimal grid/paste contact since they do not shrink away from the grid during the pasting, drying or formation processes.

Cells incorporating these features have been built at a battery Design Capability Line (DCL) for laboratory, accelerated and field tests over a size range from 200 to 1000 AH embodying 10.5″ diameter grids which have evolved through three generations of design. Other cells have been built in sizes up to 1680 AH embodying two generations of grids approximately 13″ in diameter. A 1680-AH preproduction prototype design is currently being built at the DCL on prototype manufacturing equipment designed by Western Electric Company. Several hundred of these cells will be furnished to the telephone companies for

TABLE VI—LEAD-ACID CELL FIELD TRIALS

Date	Location	Company	Battery
1/30/69	Murray Hill, N. J.	N. J. Bell Tel. Co.	48V
2/7/69	Gouldsboro, Pa.	AT & T Long Lines	12V
3/21/69	Shirley, N. Y.	AT & T Long Lines	12V
4/11/69	Noyack, N. Y.	AT & T Long Lines	12V
9/16/69	Dover, Missouri	AT & T Long Lines	24V

an extensive field trial during 1970. Specifics of an initial field trial which was instituted in 1969, are shown in Table VI.[15] Figure 13 shows the assembled element of the 420-AH 11″ grid cell furnished for this early trial. Figure 14 shows the completed cell, housed in the same size jar which can contain the elements of the 1680-AH preproduction prototype cell made with 12.5″ diameter positive grids. The 1680-AH cylindrical cell will be electrically interchangeable with and will fit into the same rack space as the conventional 1680-AH designs.

These initial trial cells have furnished important materials and design feedback;[14] demonstrated the durability of the cell in shipment, handling and installation; and, most importantly, served to facilitate operating company acceptance of the new design.

VI. SUMMARY

To provide correction to difficulties observed with commercial cell designs in Bell System use, a new cell has been developed. It utilizes pure-lead grids to minimize grid corrosion and increase cell life. The conical grids are stacked in a self-supporting structure to minimize stress and allow use of the soft lead material. Improved contact to the positive active material results both from grid design and from the use of interlocking particles of a new positive paste, tetrabasic lead sulfate. Rigid, flame-retardant PVC jar-and-cover materials provide improved mechanical strength to the container system and permit introduction of a high reliability jar-cover heat seal. New post seal technology minimizes the possibility of corrosion of the post-connector contact surface during cell life. A Design Capability Line has been established to facilitate introduction of this design into commercial manufacture. Early cells, made at this facility on hand tooling, have been successfully introduced into a limited field trial. A preproduction design is currently being built for a more extensive trial using prototype manufacturing
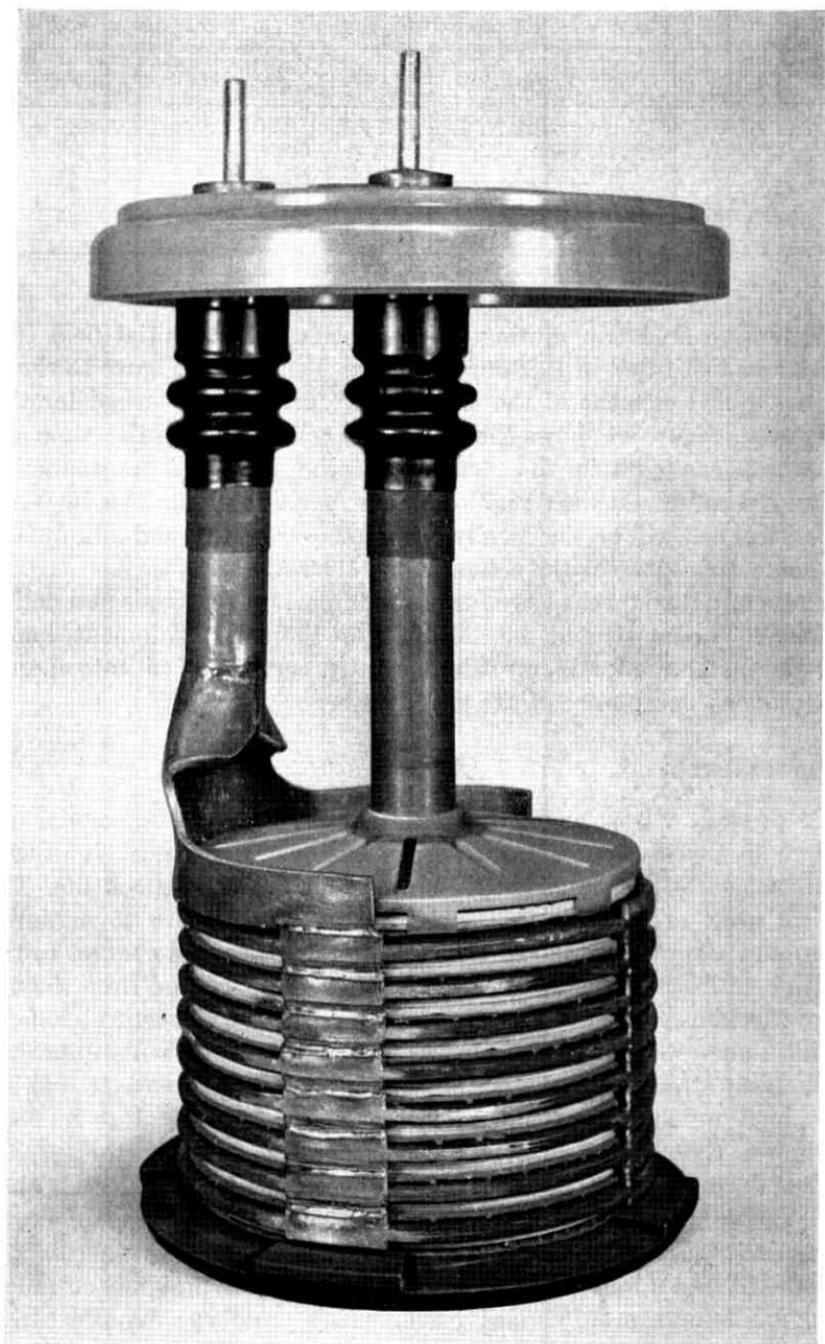
Fig. 13—Element of 420-AH field trial cell.
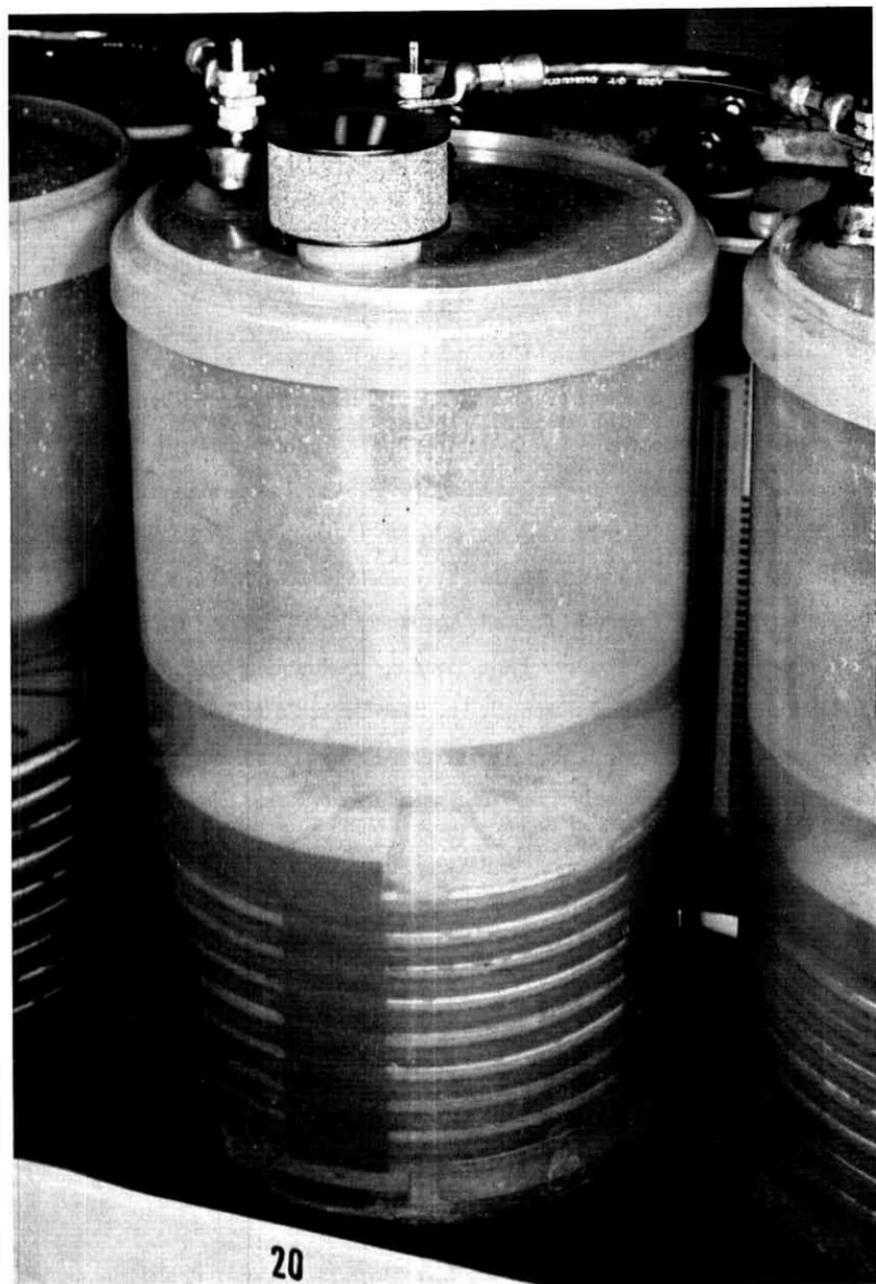
Fig. 14—Complete 420-AH field trial cell.

equipment developed by Western Electric. Accelerated tests have been developed to insure that materials, processes and designs will result in cell life, performance and reliability compatible with Bell System usage.

REFERENCES

1. Faure, M., "Faure Secondary Battery," Electrician, *6* (May 7, 1881), p. 323.
2. Sellon, J. S., "Improvements in Secondary Batteries or Magazines of Electricity," British Patent 3987, issued September 15, 1881, London, England.
3. Crennell, J. T., and Milligan, A. G., "The Use of Antimonial Lead for Accumulator Grids: A Cause of Self-Discharge of the Negative Plates," Trans. Faraday Soc., *27* (January 1931), pp. 103–112.
4. Haring, H. E., and Thomas, U. B., "The Electrochemical Behavior of Lead, Lead-Antimony and Lead-Calcium Alloys in Storage Cells," Trans. Electrochem. Soc., *68* (1935), pp. 293–307.
5. Thomas, U. B., Forster, F. T., and Haring, H. E., "Corrosion and Growth of Lead-Calcium Alloy Storage Battery Grids as a Function of Calcium Content," Trans. Electrochem. Soc., *92* (October 1947), pp. 313–325.
6. Bradley, W. W., unpublished work.
7. Luer, H. J., "Incorporation of the Battery into the Telephone Plant," B.S.T.J., this issue, pp. 1447–1469.
8. Cannone, A. G., Biagetti, R. V., and Feder, D. O, "Positive Grid Design Principles for the Bell System Battery," B.S.T.J., this issue, pp. 1279–1303.
9. Biagetti, R. V., and Weeks, M. C., "Tetrabasic Lead Sulfate as a Paste Material for Lead-Acid Battery Positive Plates," B.S.T.J., this issue, pp. 1305–1319
10. Huseby, T. W., Ryan, J. T., and Hubbauer, P., "Polyvinyl Chloride Battery Jars and Covers," B.S.T.J., this issue, pp. 1359–1376.
11. Cushman, R. H., "Positive Plate Bonding Techniques," B.S.T.J., this issue, pp. 1419–1446.
12. Dahringer, D. W., and Shroff, J. R., "Jar-Cover Seals," B.S.T.J., this issue, pp. 1393–1404.
13. Sharpe, L. H., Shroff, J. R., and Vaccaro, F. J., "Post Seals for the New Bell System Battery," B.S.T.J., this issue, pp. 1405–1417.
14. Butherus, A. D., Lindenberger, W. S., and Vaccaro, F. J., "Electrochemical Compatibility of Plastics," B.S.T.J., this issue, pp. 1377–1392.
15. O'Sullivan, T. D., Biagetti, R. V., and Weeks, M. C., "Electrochemical Characterization of the Bell System Battery: Field Trials of the Battery," B.S.T.J., this issue, pp. 1335–1358.

Lead-Acid Battery:

Positive Grid Design Principles

By A. G. CANNONE, D. O. FEDER, and R. V. BIAGETTI

(Manuscript received April 29, 1970)

In this paper, we present accelerated test data which show the superior anodic corrosion and growth behavior of pure lead as compared to lead calcium and lead-antimony positive grids for lead-acid batteries in float service. We relate differences in growth behavior to differences in metallurgy for these three alloy systems. Pure lead has been incorporated into circular grid designs and tests show these to be a substantial improvement over conventional rectangular grids.

A novel grid design concept has been developed and applied to the design of pure lead circular grids. The rate and direction of growth in the grids are controlled by appropriate geometrical design of the grid members. Batteries incorporating these circular grids increase in capacity with age rather than decrease in capacity as do rectangular grids. The circular grids can be designed to provide lifetimes of several centuries.

I. INTRODUCTION

In order to insure that stand-by lead-acid batteries are always in a full state of charge, the batteries are "floated" at 2.17 V per cell which is 110 mV above the open circuit cell potential (O.C. = 2.060 V for 1.210 sp. gr. H_2SO_4). Under these float conditions, PbO_2 is the thermodynamically stable solid phase at the anodes. Consequently, the lead or lead alloy grid is oxidized to PbO_2. The initial layer of PbO_2 formed on the lead anode is passivating and subsequent corrosion is usually a relatively slow process. The specific volume of PbO_2 is 21 percent greater than that of lead. Therefore, the corrosion product (PbO_2) induces stresses in the lead substrate from which it was formed. The induced stresses cause growth of the grid, which eventually results in failure of the battery. The effect of grid growth is illustrated in Fig. 1 which shows a severely corroded PbCa plate overlayed with a noncorroded grid to show original dimensions. This grid has grown to such an extent that cracking of the PbO_2 pellets and loss of contact with the grid members

1279
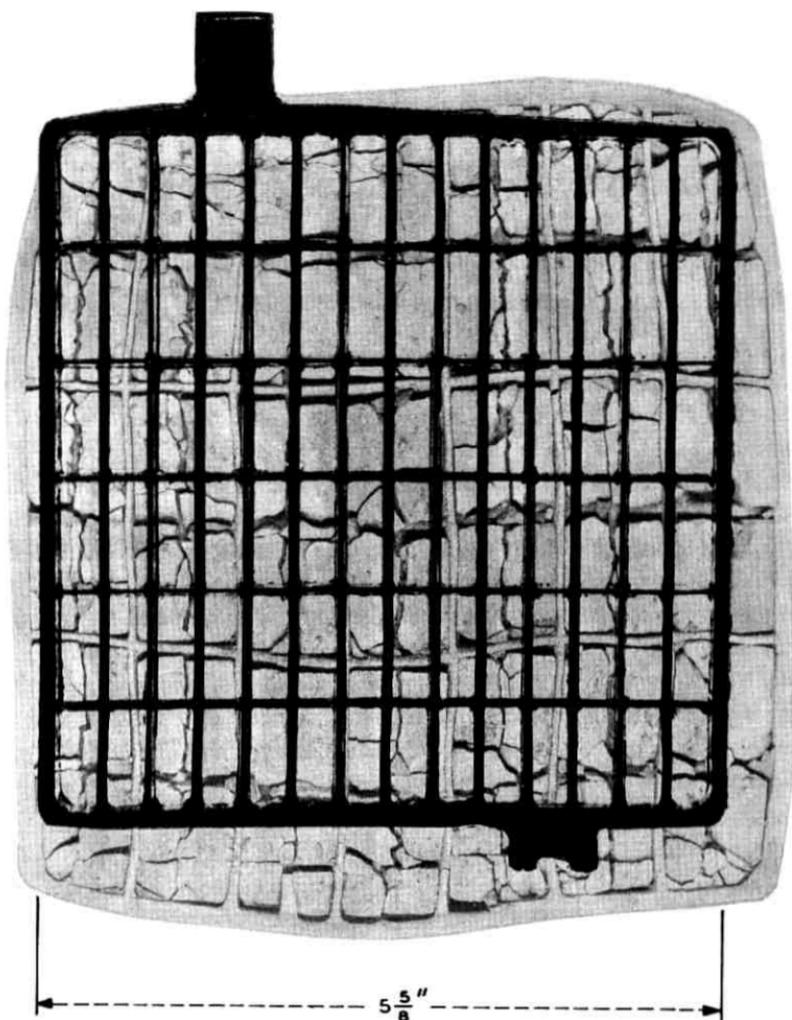
Fig. 1—Severely corroded PbCa grid with overlayed bare grid to show initial dimensions.

have occurred. It is obvious that this must result in a decreased capacity of the plate. Positive plate growth also results in cracking of battery jars, and rupture of jar cover and post cover seals. It was clear that substantial improvements in battery life and performance could be achieved if grid corrosion and growth could be minimized.

It is known that corrosion and growth of PbCa positive grids are

critically related to Ca content.[1,2] Unpublished independent studies by
W. W. Bradley and B. A. Cretella, of these laboratories, have shown
that the best corrosion behavior of PbCa cells in float service were
obtained with cells having low calcium content. This prompted an
investigation into the applicability of pure lead for use in long-term,
float-service, lead-acid batteries.

Several studies on the anodic corrosion of pure lead have been
reported.[3-5] However, these previous investigations are in most cases
not relevant to telephone battery float-service conditions and none
was of help in predicting the float-service behavior of pure lead versus
PbCa and PbSb batteries. Consequently, the anodic corrosion and
growth of grids of pure Pb, PbCa and PbSb were compared by using
accelerated test procedures which were consistent with our float-service
applications.

II. EXPERIMENTAL WORK

The experimental work carried out during this investigation was
divided into two phases. The first phase consisted of corrosion and
growth studies on pure Pb, PbCa, and PbSb grids of conventional
rectangular design (5-5/8" $\times$ 5-7/8" $\times$ 1/4" thick). This study was
designed to establish the validity of the testing procedure and to obtain
comparative Pb, PbCa and PbSb performance data. The second phase
consisted of accelerated corrosion tests designed to evaluate grid designs
and metallurgy developed from results of the first-phase studies.

Corrosion and growth studies were performed on cells assembled in
glass jars in 1.210 sp. gr. H_2SO_4. Both pasted plate and bare grid cells
were used in this study. Other details of cell construction, such as
plate sizes and number of plates, will be given in the appropriate sections
of experimental results. Cells were initially given a series of discharge-
charge cycles to establish zero time performance after which positive
plate dimensions were measured with a micrometer. The cells were
then placed at preselected elevated temperatures and the positive plates
were potentiostated at 80 mV above the reversible $PbO_2/PbSO_4$ poten-
tial at the test temperature. Hg/Hg_2SO_4 electrodes were used for
potentiostatic control. Small cells were heated internally with immersion
heaters, whereas external box ovens were used for the larger cells.
Appropriate thermostats were used for temperature control.

At predetermined intervals the cells were cooled, cycled to establish
capacity, measured to determine dimensional changes and returned to
hot test.

III. RESULTS AND DISCUSSION

The Phase I experiments utilized commercially available PbCa(0.05 percent Ca) and PbSb (6 percent Sb) grids of a design identical to that shown in Fig. 1. These rectangular grids were received already pasted. Identical design pure lead (99.99 percent) grids were cast in this laboratory using a bottom pour, gravity feed mold. In those cases where pure lead pasted plates were used, these were pasted with a commercially available formulation. All cells tested were 5 plate cells—2 positive plates and 3 negative plates—with appropriately located fiberglass mats and microporous rubber separators.

The test results for PbCa, PbSb and pure Pb are shown in Figs. 2, 3 and 4 respectively where the percent horizontal grid growth is plotted versus time at the various temperatures. The horizontal growth data are those for the midpoints of the positive grids. In most cases, each data point is the average growth of the two grids in a given cell. The superior performance of the pure lead grids is illustrated in Fig. 5 where the growth data is plotted for all three lead systems at 82°C.

The growth of the PbCa and pure Pb grids is quadratically time-dependent, whereas the PbSb data are reasonably linear over the range
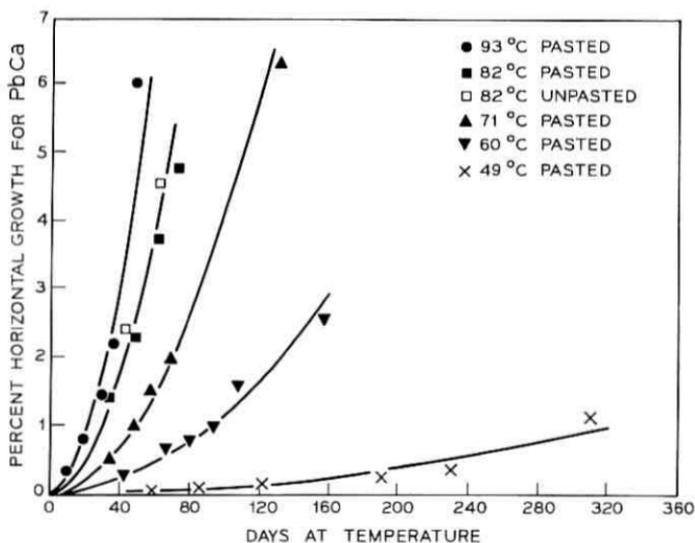

Fig. 2—Percent horizontal growth vs days at temperature for 5-5/8″ × 5-7/8″ × 1/4″ thick PbCa (0.05 percent Ca) grids.
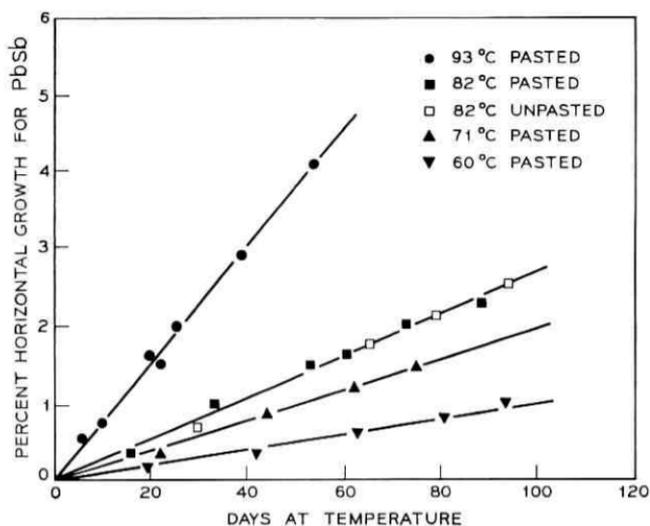
Fig. 3—Percent horizontal growth vs days at temperature for 5-5/8″ × 5-7/8″ × 1/4″ thick PbSb (6 percent Sb) grids.

studied. The curves drawn through the PbCa and pure Pb data points correspond to the equation:

$$\text{Percent Growth} = kt^2 \tag{1}$$

where k is a rate constant and t is time. Curve fitting in accordance with equation (1) was done visually. As can be seen, the curves are a very reasonable fit. The PbSb system shows a linear time dependent growth corresponding to equation (2).

$$\text{Percent Growth} = kt. \tag{2}$$

The temperature dependent rate constants corresponding to equations (1) and (2) are summarized in Table I.

As would be expected, the superior performance of the pure Pb system is also reflected in the ampere-hour capacities of accelerated test cells. Typical behavior is shown in Fig. 6 where the performance of the three Pb systems is shown as a function of time at 93°C. The capacities were normalized in order to eliminate differences in time zero behavior for the three systems. The improved capacity behavior of pure Pb plates may be attributed to the slower growth of pure Pb which enables the grid to maintain contact to the PbO_2 active material for a longer time than for PbCa and PbSb.
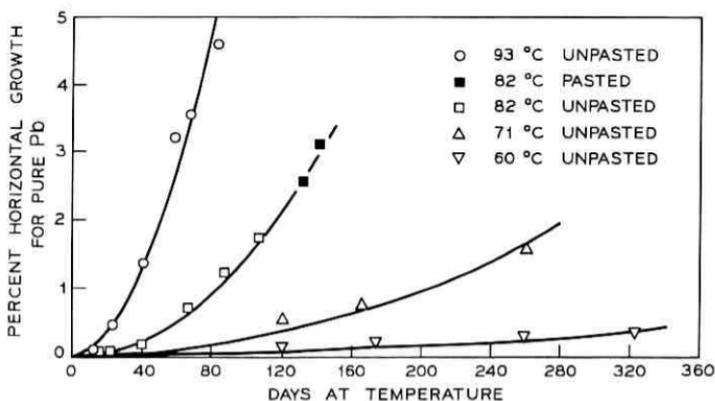
Fig. 4—Percent horizontal growth vs days at temperature for 5-5/8″ × 5-7/8″ × 1/4″ thick pure Pb (99.99 percent) grids.

The reasonable conformance of the PbCa and pure Pb systems to a parabolic growth equation suggests that the growth mechanics for these two systems are similar. Figures 7 and 8 show this to be the case. Corrosion of the PbCa alloy (Fig. 7) is primarily intergranular.[4,6] Figure 8 shows that the corrosion of pure lead is also intergranular. Comparison of Figs. 7 and 8 shows the grain size of the pure lead grid to be considerably larger (approximately 40 times larger) than that of the PbCa grid. This suggests that the superior performance of pure Pb is related to the fewer corrosion sites (grain boundaries) per unit surface area for this system.

3.1 Room Temperature Behavior

In order to predict room temperature behavior, a plot was made of log k [equation (1)] versus $1/T$ for the two materials obeying the parabolic growth equation, Pb and PbCa. In the results, shown in Fig. 9, it is seen that the pure Pb grid data follow the Arrhenius law whereas the PbCa data show divergence at 82°C and 93°C. It is known that PbCa is a temperature-sensitive, age-hardening alloy whose metallurgical properties can be altered significantly by heat treatment.[7] The hardness of any given PbCa alloy is determined by the distribution of the Pb_3Ca phase in the alloy. Maximum hardness is associated with uniform distribution and particle size of the Pb_3Ca phase and prolonged heat treatment agglomerates Pb_3Ca which results in a softer alloy. This suggested that the observed anomalies at 82°C and 93°C might be related to changes in PbCa metallurgy induced by testing at these temperatures.
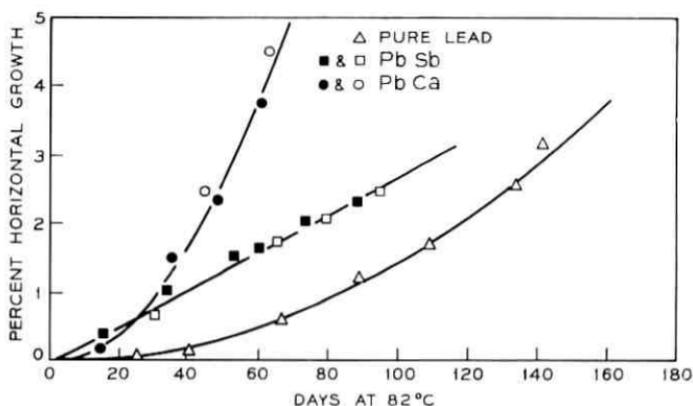
Fig. 5—Percent horizontal growth vs days at 82°C for 5-5/8″ × 5-7/8″ × 1/4″ thick PbCa, PbSb and pure Pb pasted plates.

A second set of accelerated tests on PbCa was carried out in which the grids were heat-treated at 110°C for 11 days prior to testing at elevated temperatures. Accelerated test results on these grids are shown in Fig. 10. Comparison of Figs. 2 and 10 shows that the heat-treated grids grow slower at all temperatures except 93°C. At 93°C the growth curves are essentially the same which indicates that testing at this temperature does indeed induce metallurgical changes in the PbCa system. Figures 11 and 12 are sections of the same PbCa grid before and after heat treatment. There are no obvious differences in grain size or structure which might account for the different growth behaviors. However, tests performed with a Kentron Micro Hardness tester

TABLE I—GROWTH RATE CONSTANTS AT VARIOUS TEMPERATURES FOR $5\frac{5}{8}″ \times 5\frac{7}{8}″ \times \frac{1}{4}″$ THICK PURE Pb, PbSb, PbCa AND HEAT-TREATED PbCa GRIDS

	$k^*(\%/day^2)$			$k^{**}(\%/day)$
$T(°C)$	Pure Pb	PbCa	PbCa (heat treated)	PbSb
93	7.7×10^{-4}	2.0×10^{-3}	1.9×10^{-3}	7.6×10^{-2}
82	1.5×10^{-4}	1.1×10^{-3}	5.5×10^{-4}	2.7×10^{-2}
71	2.5×10^{-5}	4.2×10^{-4}	9.0×10^{-5}	2.0×10^{-2}
60	4.0×10^{-6}	1.2×10^{-4}	2.0×10^{-5}	1.0×10^{-2}
49		1.0×10^{-5}		

* k in equation (1)—% Growth $= kt^2$
** k in equation (2)—% Growth $= kt$
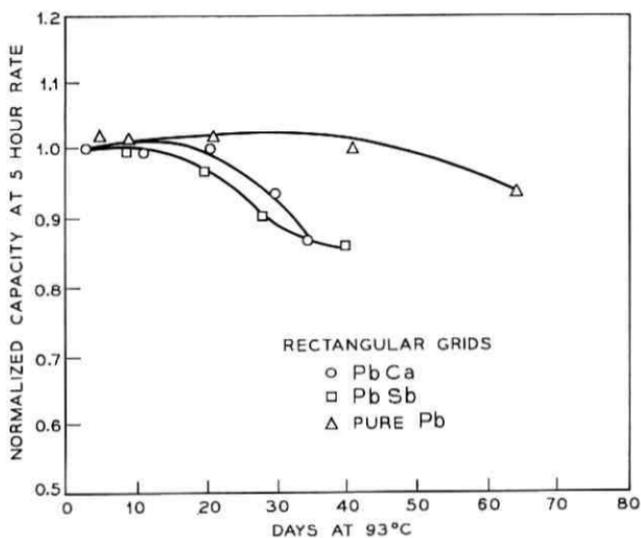
Fig. 6—Normalized 5-hour rate capacities vs days at 93°C for 5-5/8″ × 5-5/7″ × 1/4″ thick PbSb, PbCa and pure Pb grids.
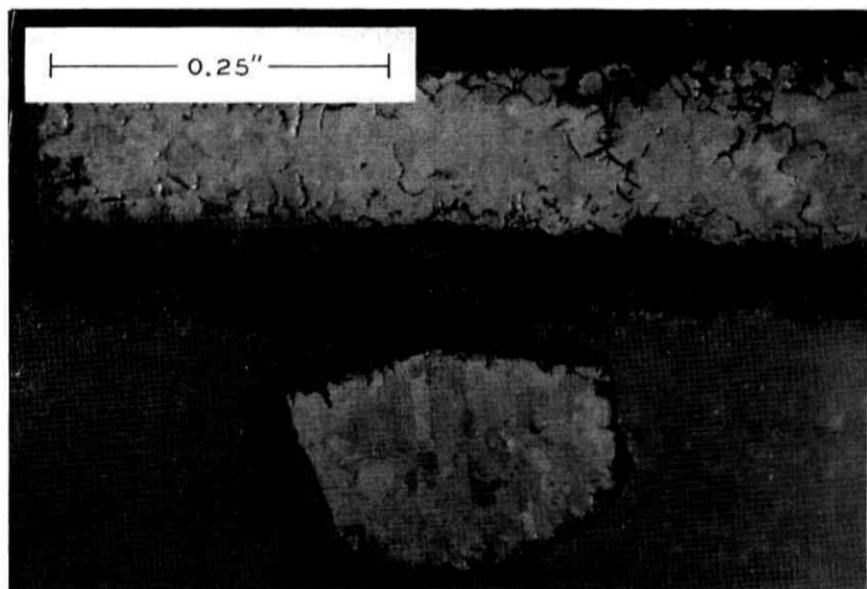

Fig. 7—Section of PbCa (0.05 percent Ca) grid after 36 days at 93°C.
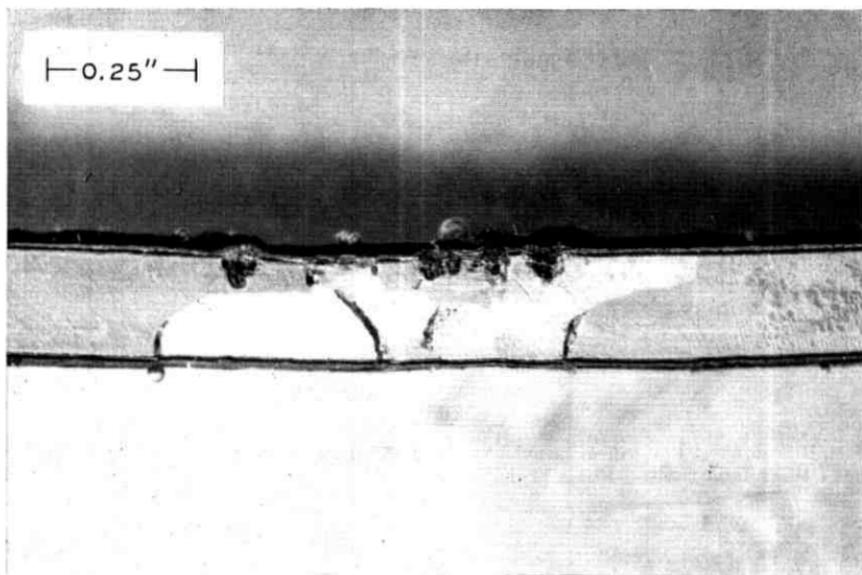
Fig. 8—Section of pure Pb grid after 339 days at 82°C.
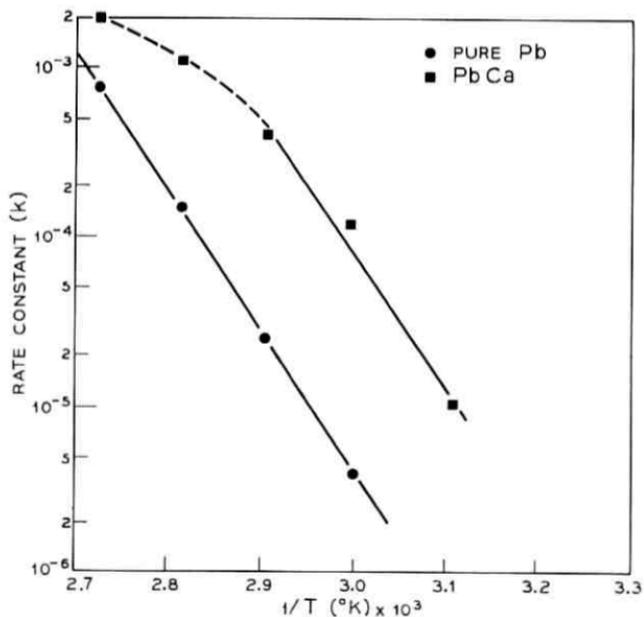
Fig. 9—Rate constant (k) on log scale vs $1/T(°K)$ for 5-5/8'' $\times$ 5-7/8'' $\times$ 1/4'' thick pure Pb and PbCa.
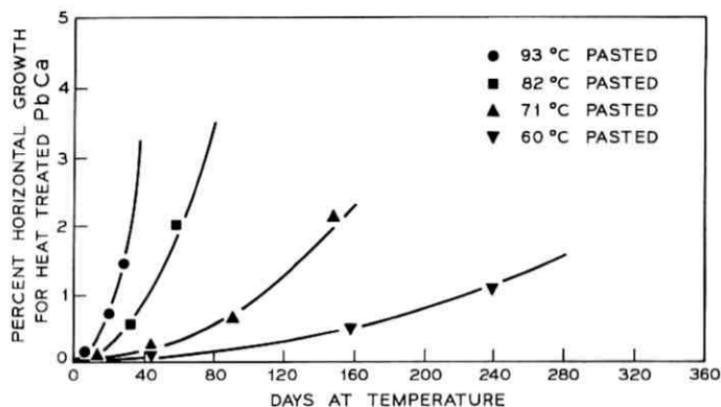
Fig. 10—Percent horizontal growth vs days at temperature for 5-5/8″ × 5-7/8″ × 1/4″ thick heat treated PbCa grids.
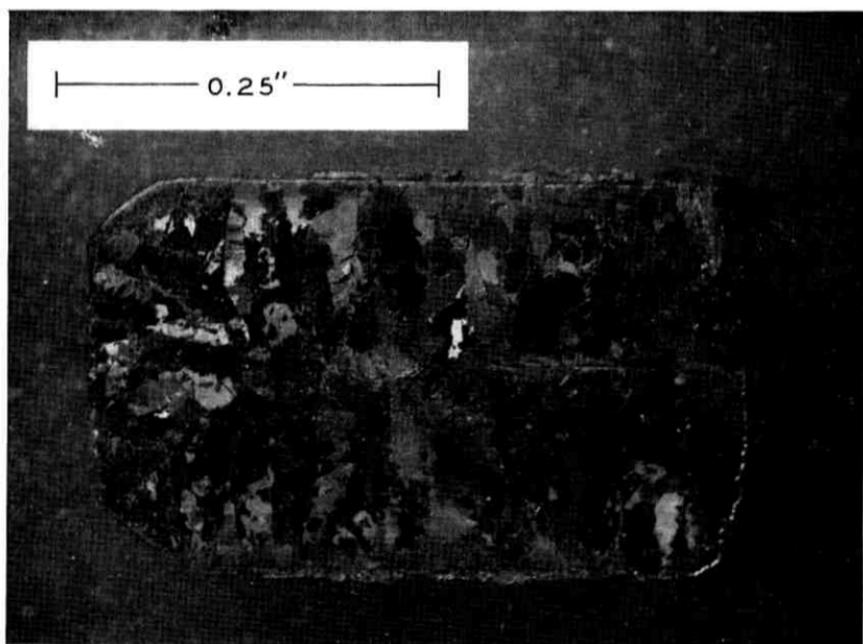
Fig. 11—Section of PbCa (0.05 percent Ca) grid before heat treatment.

|————— 0.25" —————|
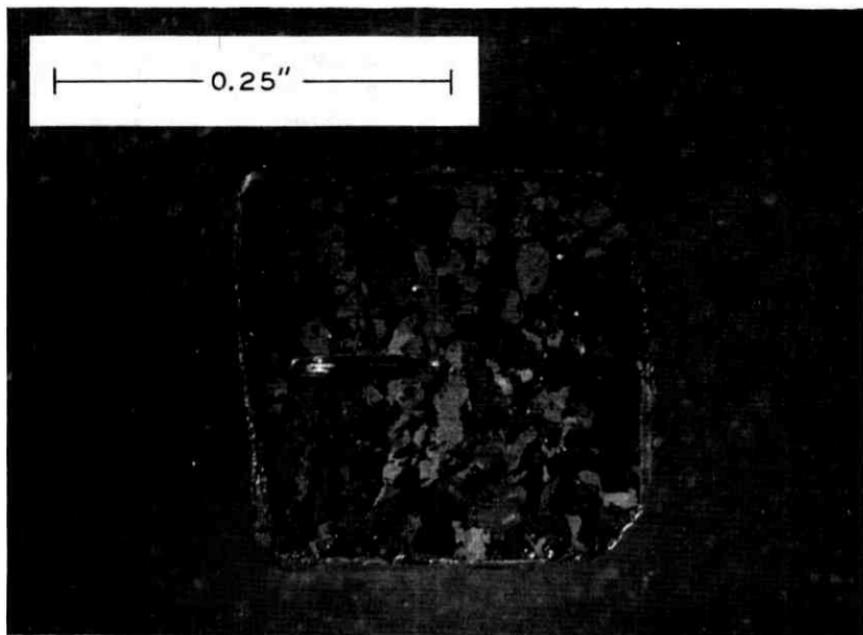
Fig. 12—Section of PbCa (0.05 percent Ca) grid after heat treatment at 110°C for 11 days.

showed the heat treated grids to be somewhat softer than untreated PbCa grids. This is in accord with earlier observations[8] and suggests that the distribution of the Pb_3Ca phase has been altered by heat treatment. It is reasonable to suggest that the corrosion and growth characteristics of PbCa grids are related to the distribution of the Pb_3Ca phase, particularly in the grain boundaries.

Figure 13 shows a log k vs $1/T$ plot for the heat treated PbCa system in addition to the Pb and PbCa data shown in Fig. 9. The data for the heat treated PbCa show Arrhenius behavior indicating that prior heat treatment resulted in stabilization of the PbCa alloy. The rate constant and temperature data for heat treated PbCa are also summarized in Table I. If, as suggested, the growth mechanics are the same for pure Pb and PbCa, then one would expect the slopes (activation energies) of the Arrhenius plots to be the same for these two systems. Figure 9 shows this to be the case with only minor variations.

As mentioned previously, the growth of the PbSb system is linearly time dependent. The fact that the growth behavior for PbSb is different from Pb and PbCa is not unexpected. The PbSb system is more complex
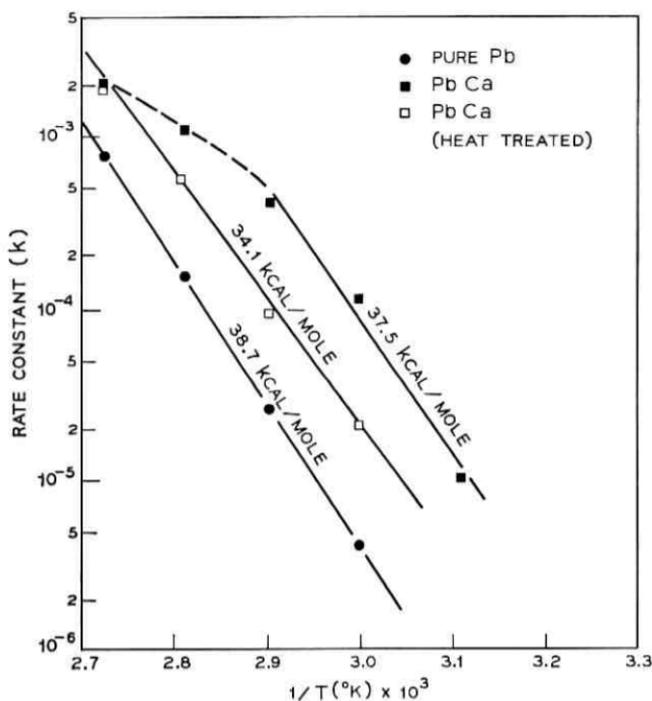
Fig. 13—Rate constant (k) on log scale vs $1/T$ (°K) for 5-5/8″ $\times$ 5-7/8″ $\times$ 1/4″ thick pure Pb, PbCa and heat treated PbCa grids.

because corrosion of the alloy is associated with dissolution of Sb. Since Sb is dissolved, it is not incorporated in the PbO_2 corrosion product and the increase in volume resulting from the conversion of Pb to PbO_2 is significantly less for PbSb than it is for pure Pb or PbCa. For example, a 21 percent increase in specific volume occurs in the corrosion of pure Pb to PbO_2 whereas the corresponding increase for 6 percent PbSb is only 9 percent.[9] The log k vs $1/T$ data for the PbSb grids are plotted in Fig. 14.

Room temperature (25°C) rate constants were calculated using the data in Figs. 13 and 14. These are summarized in Table II along with calculated lifetimes to 1 percent (0.056″) and 4 percent (0.225″) growth. Diagnostic studies have shown that 4 percent growth is an approximate upper limit above which the integrity of the battery may be seriously impaired. Table II shows that pure lead would reach 4 percent growth in 82 years whereas the comparable times for PbCa and PbSb are 16.8 and 13.8 years respectively. This should result in a substantial improve-
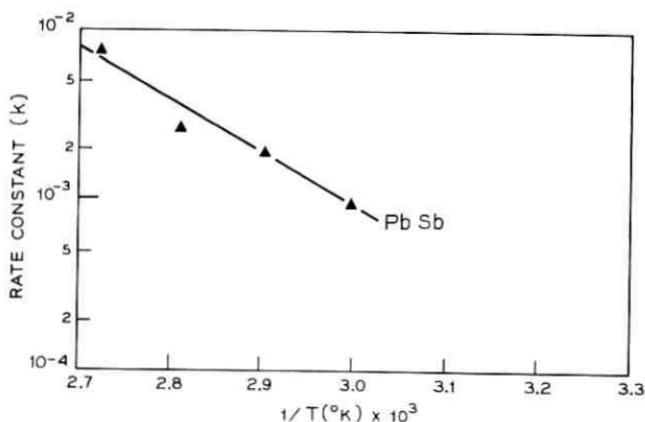
Fig. 14—Rate constant (k) on log scale vs $1/T$ (°K) for 5-5/8″ $\times$ 5-7/8″ $\times$ 1/4″ thick PbSb grids.

ment in battery life for the pure lead as compared to PbCa and PbSb cells. Heat treatment of PbCa grids is effective in reducing the rate of grid growth. However, heat-treated PbCa is still not comparable to pure lead.

Calculation of room temperature rate constants from Arrhenius plots assumes, of course, that changes in mechanism do not occur at lower temperatures. This has not been substantiated under controlled experimental conditions because of the prohibitively long test times required. However, the calculated 16.8 and 13.8 years room temperature lives to 4 percent growth for PbCa and PbSb respectively are consistent with results obtained from diagnostic studies of cells in field use. In view of this, it is reasonable to assume that the calculated 82 years would be realized by the pure Pb cells in field use.

TABLE II—GROWTH RATE CONSTANTS AT 25°C FOR $5\frac{5}{8}″ \times 5\frac{7}{8}″ \times \frac{1}{4}″$ THICK PURE Pb, PbSb, PbCa AND HEAT TREATED PbCa GRIDS

	k at 25°C	Time to 1% Growth (Years)	Time to 4% Growth (Years)
Pure Pb	$4.41 \times 10^{-9}\%/\text{day}^2$	41	82
PbCa	$1.06 \times 10^{-7}\%/\text{day}^2$	8.4	16.8
PbCa (heat treated)	$4.60 \times 10^{-8}\%/\text{day}^2$	12.8	25.6
PbSb	$7.96 \times 10^{-4}\%/\text{day}$	3.4	13.8

3.2 *New Cell Designed*

In order to capitalize on the superior corrosion and growth behavior of pure Pb, a new cell was designed with a self-supporting interlocking structure to minimize the stress on the soft Pb.[10] This section discusses the design and performance of the conical grids which have been developed for the new cell. Figure 15 shows the first (Design "A") of a series of pure lead positive grid designs that were evaluated. The circular grid (10.9″ O.D. × 0.25″) consists of a series of concentric hoops interconnected by both continuous and interrupted radial members. In this circular design, corrosion and growth should cause the diameters of the concentric hoops to increase. The radial members would be buckled as a result of growth.
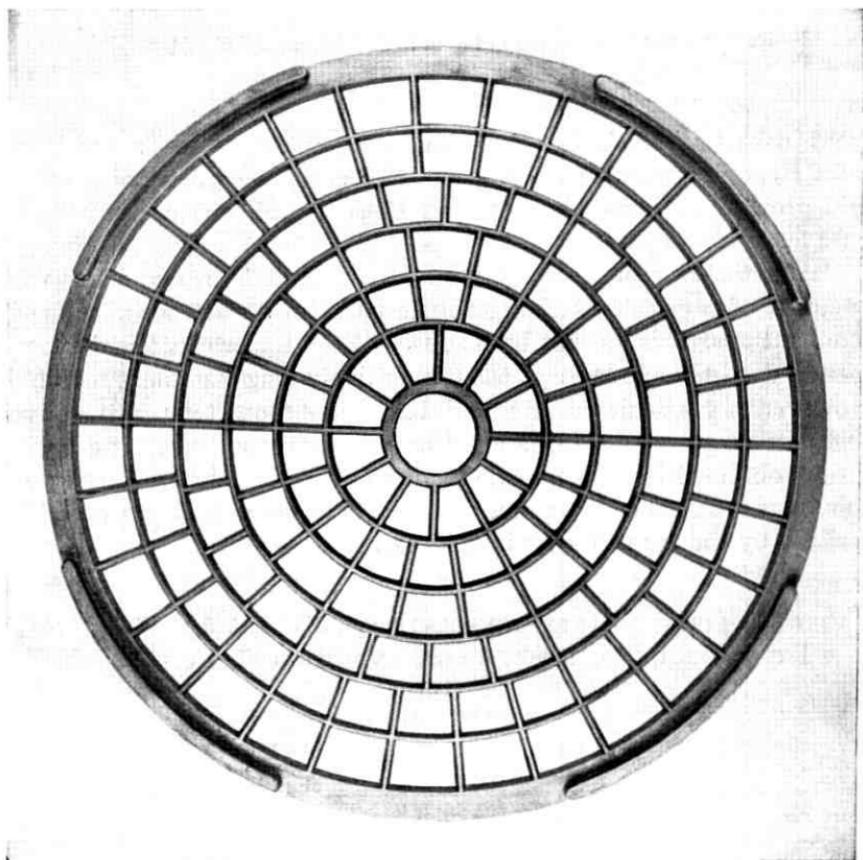

Fig. 15—Design "A" pure Pb grid — 10.9″ O.D. × 0.25″.

Figure 16 shows a different embodiment of this circular grid design concept. This spiral grid (Design "B", 10.9" O.D. × 0.25") has the hoops and radial members prebuckled in an attempt to control the direction of growth of all the members.

Accelerated tests of these positive grid designs were carried out with 160 AH cells (5 hr. rate)—3 positive plates, 4 negative plates. In cells, these grids are shaped to a 10° cone and stacked pancake fashion one atop another. Other details of cell construction are given elsewhere.[10] Results of growth studies for these two grid designs are summarized in Figs. 17 and 18.

The growth data for these two circular designs are roughly linear with time. Since the mechanism of growth is a property of the material and not of grid geometry, growth having quadratic time dependence similar to the rectangular grids was expected. The reason for non-conformance to parabolic growth became apparent after test completion when the cells were disassembled and the grids examined. Figure 19 shows a design "A" grid after 169 days at 93°C. Considerable "mushrooming" has occurred which has substantially distorted the grid from its original 10° conical shape. Other studies showed this same "mushrooming" effect also occurring in the design "B" grids. It was clear that "mushrooming" was the result of unbalanced grid designs in that
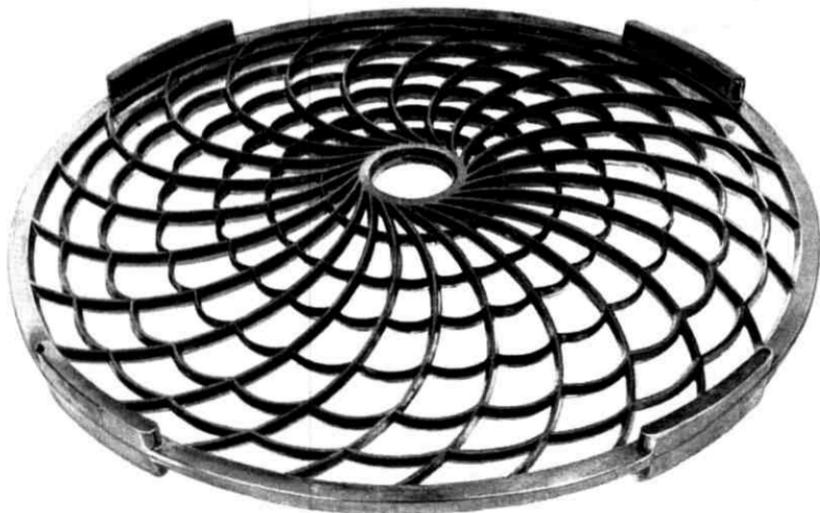
Fig. 16—Design "B" pure Pb grid — 10.9" O.D. × 0.25".
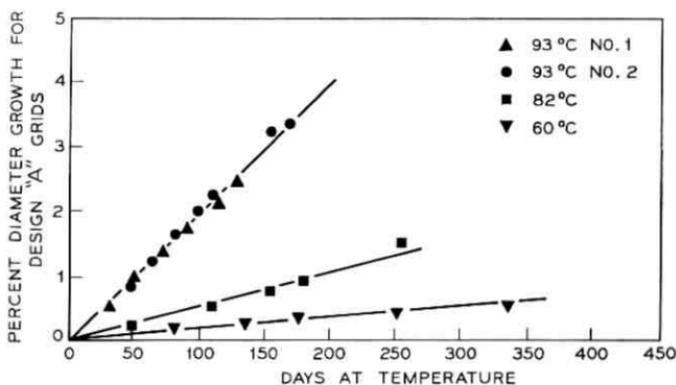
Fig. 17—Percent growth of diameter vs days at temperatire for design "A" pure Pb circular grids.

all the circular hoops were not growing at the same rate. Most of the inner hoops for both grid designs were growing at a faster rate than the outer ring which eventually leads to severe distortion from the original 10° cone.

Evaluation of growth data from the rectangular PbCa grid experiments revealed that the growth of any given member was directly proportional to the surface area (S.A.) and inversely proportional to the cross-sectional area (C.A.) of the member. In effect, the rate constant (k) in equation (1) is a product of two constants, $k_1 k_2$:
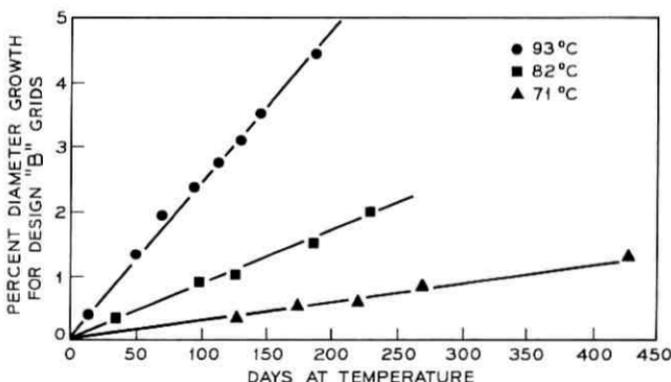

Fig. 18—Percent growth of diameter vs days at temperature for design "B" pure Pb circular grids.
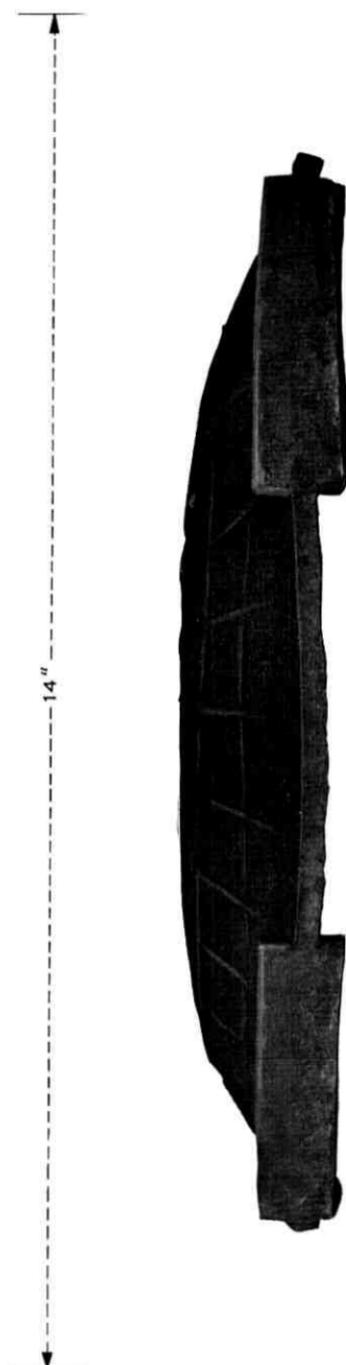
Fig. 19—Design "A" pure Pb circular grid after 169 days at 93°C.

$$\text{Growth} = k_1 k_2 t^{2*} \qquad (3)$$

where k_1 is the rate constant for growth and is a function of only the particular alloy system and k_2 is the S.A./C.A. ratio of a grid member. This concept is further clarified by the data in Table III. In Table III, k_1 [equation (4)] is calculated from detailed growth data (PbCa 82°C) on the three different types of rectangular grid member geometries (Fig. 1).

$$k_1 = \frac{\text{Growth}}{k_2 t^2}. \qquad (4)$$

TABLE III—GROWTH RATE CONSTANTS AT 82°C FOR VARIOUS GRID MEMBERS OF THE $5\frac{5}{8}'' \times 5\frac{7}{8}'' \times \frac{1}{4}''$ THICK PbCa GRID

k_1(in/day^2)	Growth at 68 Days (inches)	k_2(S.A./C.A.)
8.7×10^{-8}	.008	19.9
8.8×10^{-8}	.013	32.0
8.6×10^{-8}	.021	53.0

It is seen in Table III that k_1 is essentially independent of grid member geometry. The conclusion is that growth data obtained on a member of a given geometry enables one to calculate the growth expected for a large variety of geometric shapes provided the alloy system remains constant. Using this approach, the expected growth of the outer ring of the pure lead design "A" grid was calculated and compared to actual growth at 93°C. Figure 20 shows the measured growth to be significantly greater than that calculated. The reason for this lies in the peculiarities of the grid design which leads to "mushrooming." Table IV lists the S.A./C.A. ratios for each of the rings in the "A" design. Since all the ratios are different, the expected rate of growth for each ring would also be different. Table IV also indicates that the three rings closest to the outer ring would grow faster than the outer ring.

It is suggested that during the initial stages of growth, the faster growth of these inner rings is transmitted via the PbO$_2$ pellets to the outer ring. This in effect stretches the outer ring resulting in growth larger than would be expected in the normal course of corrosion. During the intermediate stages of growth, the faster growing inner rings eventually lead to distortion from the original 10° plane resulting in the observed mushrooming. During advanced stages of inner ring growth,

---

* Note that equation (3) expresses growth on an absolute basis rather than the percent basis used previously.
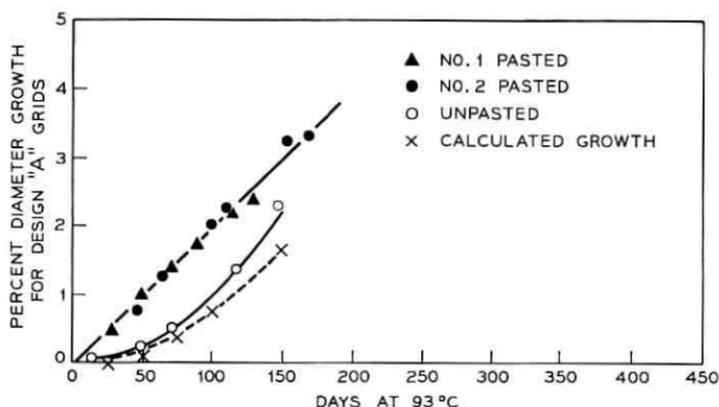
Fig. 20—Experimental and calculated design "A" grid growth vs days at 93°C.

"mushrooming" has become severe and transmission of growth via the PbO_2 paste material is not as effective.

The transmission of growth via the PbO_2 paste material was substantiated in an unpasted grid experiment also shown in Fig. 20. In this case, there is no PbO_2 paste material for transmission of inner ring growth to the outer ring and the growth data correspond much better to that calculated. The growth data given for the "A" and "B" designs represent, therefore, second-order effects controlled by the growth characteristics of the inner ring members of these grids and thus these do not follow the parabolic growth equation. The data, however, still have physical significance and the linear rate constants have been summarized in Table V and plotted in Fig. 21.

The extrapolated 25°C rate constants are 4.62×10^{-5} percent/day and 1.29×10^{-5} percent/day for the "A" and "B" designs respectively. These yield calculated room temperature lives to 4 percent growth

TABLE IV—RATIOS OF SURFACE AREA TO CROSS-SECTIONAL AREA FOR RING SECTIONS OF DESIGN "A" GRIDS

Ring No.	S.A./C.A.
inner	70
2	219
3	331
4	442
5	546
6	657
outer	410

TABLE V—GROWTH RATE CONSTANTS AT VARIOUS TEMPERATURES
FOR "A" AND "B" DESIGN PURE LEAD GRIDS

	$k^*(\%/\text{day})$	
$T°(°C)$	Design "A"	Design "B"
93	2.0×10^{-2}	2.5×10^{-2}
82	5.4×10^{-3}	8.8×10^{-3}
71	—	3.0×10^{-3}
60	1.4×10^{-3}	

k^* corresponds to linear growth equation ($\%$ Growth $= kt$).

(0.420″) of 238 years for the "A" design and 854 years for the "B" design. It is clear that the use of pure lead in a circular design (even though the design is unbalanced) provides positive plates which will not fail as a result of grid growth for extremely long periods of time provided that the ratio of surface area to cross-sectional area of the rings is sufficiently small.

3.3 Third Circular Design

In order to overcome the "mushrooming" problem, a third circular grid was designed in which all the circular hoops were balanced such
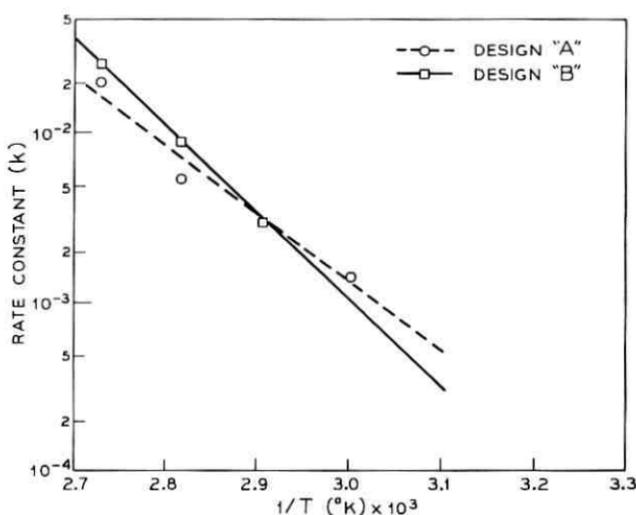

Fig. 21—Rate constant (k) on log scale vs $1/T(°K)$ for "A" and "B" circular pure Pb grid designs.

that they would all grow at the same rate. Since k_1 in Equation (3) is constant for a particular lead alloy system, a balanced circular grid design is obtained by maintaining k_2 (S.A./C.A.) constant for the hoops. Such a balanced grid design is shown in Fig. 22 (10.9″ O.D. × 0.25″) wherein all the circular sections have the same S.A./C.A.—410 in this case. The S.A./C.A. ratio of 410 is the same as that for the outer rings of the "A" and "B" designs. Figure 23 shows the results of 93°C testing on four cells incorporating this "C" design. There is reasonably close correspondence between the data and the calculated behavior for this design. The growth data show some curvature indicating quadratic time dependence as would be expected on the basis of the results obtained on the rectangular grids. However, the growth data do not conform to a simple parabolic expression [equation (1)]. It may be that the accelerated tests have not yet progressed far enough in time for growth behavior conforming to equation (1) to become obvious. Only 93°C data is available for this design and, consequently, Arrhenius treatment to obtain room temperature performance is not possible. However, comparison of the design "C" data with 93°C data of other pure lead grid designs shows the performance of the "C" design to be superior. For
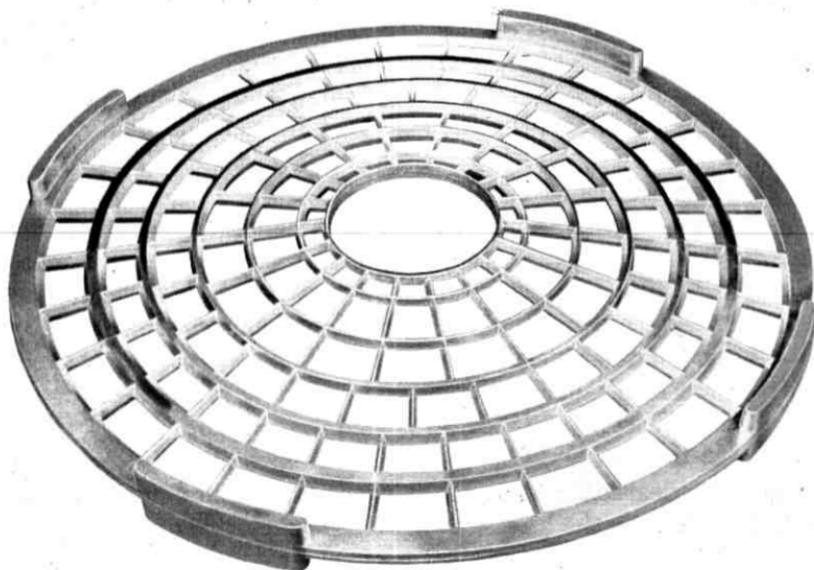

Fig. 22—Design "C" pure lead grid — 10.9″ O.D. × 0.25″.
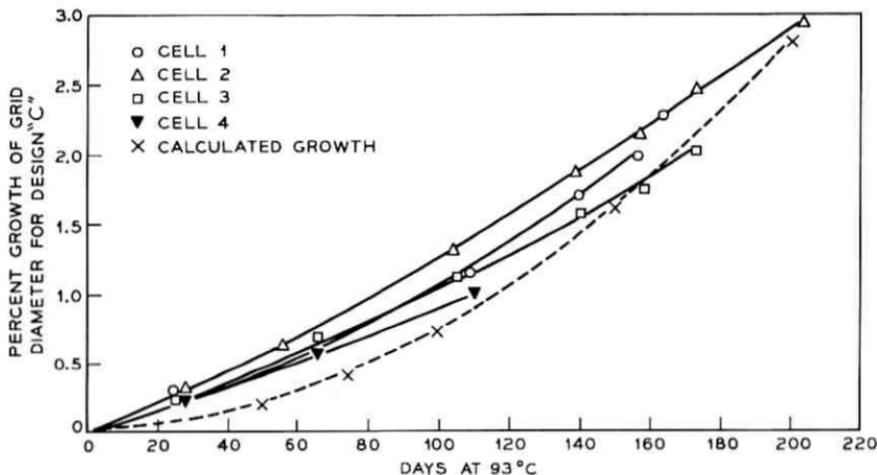
Fig. 23—Percent growth of diameter vs days at 93°C for design "C" pure Pb circular grids.

example, the time to 2 percent growth at 93°C ranges between 148–172 days for design "C" whereas the comparable times are 100, 85, and 50 days for "A", "B" and rectangular designs respectively. Examination of "C" design grids after 93°C testing shows no evidence of "mushrooming" which indicates that this design is indeed balanced. Figure 24 shows a typical "C" plate after 204 days at 93°C. It is seen that the plate is still in excellent condition.

The effects of these various designs and metallurgies are dramatically summarized in Fig. 25. The superior performance of pure lead over PbCa and PbSb in the rectangular design is shown again as it was in Fig. 6. For the "A" and "B" designs the cell capacities increase substantially as corrosion and growth occur. Since the corrosion product is PbO_2, increasing capacity would be expected for grid designs in which contact to the active material was maintained during growth. The capacities of the "A" and "B" designs reach maxima and then degrade. This capacity degradation is a consequence of "mushrooming" for these two designs. Finally the balanced "C" design shows gradually increasing capacity with no maximum reached as of yet. It is reasonable to conclude that capacities of cells made with balanced circular pure lead grids will increase with age rather than decrease as is the case with conventional rectangular cells.
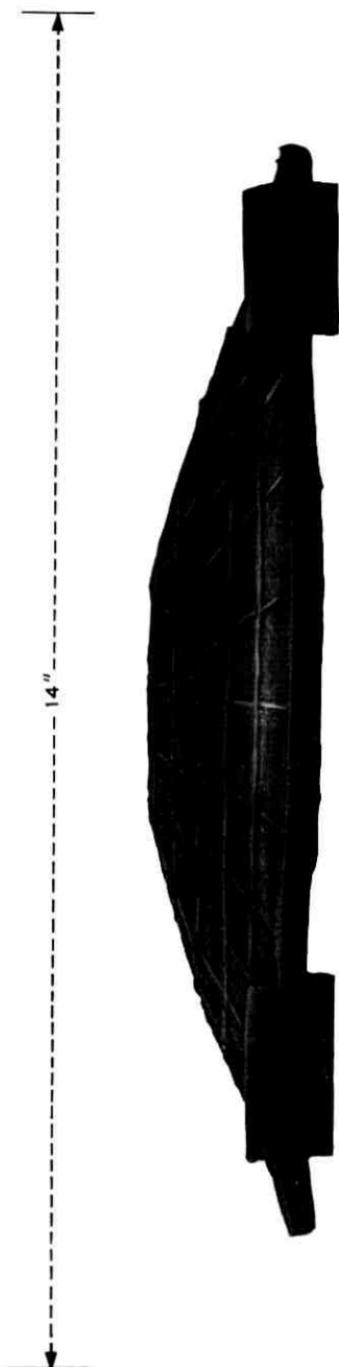
Fig. 24—Design "C" grid after 204 days at 93°C.

14"
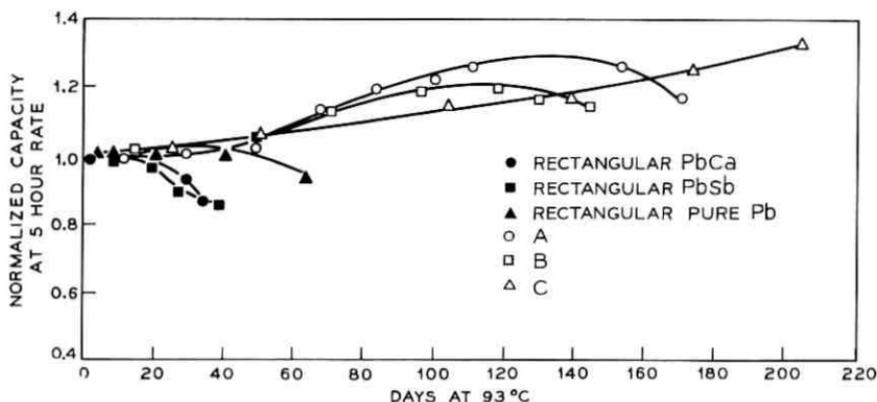
Fig. 25—Normalized 5-hour rate capacity vs days at 93 °C for various grid designs.

IV. SUMMARY AND CONCLUSIONS

Accelerated test data show the superior anodic corrosion and growth behavior of pure Pb as compared to the Ca and Sb alloys when used for positive grids in lead-acid cells for float service. The differences in growth behavior are related to the differences in metallurgy for these three alloy systems. Predicted room temperature lifetimes to 4 percent growth are 82, 16.8 and 13.8 years respectively for pure Pb, PbCa and PbSb rectangular grids.

Similar tests on two pure Pb circular grid designs ("A" and "B") show these to be substantially improved over the conventional rectangular designs. Even though both these designs "mushroom" during corrosion and growth, 238 and 854 years were predicted as the room temperature lifetimes to 4 percent growth for the "A" and "B" designs respectively.

Detailed analyses of the growth behavior of both rectangular and circular grids revealed that rate of growth of a grid member was related to the particular geometry of the member and that "mushrooming" of the circular grid designs resulted from unequal growth of the grid members. This analysis led to the development of a novel circular grid design concept wherein the rate and direction of growth could be controlled by proper sizing of the members. This concept was incorporated into the "C" design grid wherein all the circular members have a fixed surface area to cross-sectional area ratio of 410. Tests on cells incorporating this "C" design have verified the expected improved growth and capacity behavior. It is expected that the "C" grid design

will perform satisfactorily for several centuries. In view of the 30-year design life objective for the new Bell System circular lead-acid battery design,[10] it is clear that the "C" grid design easily meets this objective. Confidence in the expected room temperature behavior of the "C" design has led to the design of another pure lead circular grid. In this newest design, the S.A./C.A. ratio for the circular members has been increased to 600 as compared to 410 for the "C" design. It is expected that this newest design will grow at an absolute rate approximately 1.5 times faster than the "C" design. This "faster" growing grid is more consistent with the design life objective and also provides a reduction in the amount of lead per unit grid volume. Accelerated growth testing of this new design is being initiated and results will be reported at a later date.

V. ACKNOWLEDGMENTS

The authors wish to express their gratitude to those many individuals who have contributed to this manuscript. Special thanks to Messrs. R. A. Landwehrle, P. J. Zsadanyi and J. I. Mount, who were largely responsible for maintenance of the tests, and to Miss M. M. Cook and Mrs. J. R. Hurd for computer handling of the data.

REFERENCES

1. Thomas, U. B., Forster, F. T., and Haring, H. E., "Corrosion and Growth of Lead-Calcium Alloy Storage Battery Grids as a Function of Calcium Content," Trans. Electrochem. Soc., *92* (1947), pp. 313–325.
2. Turner, D. R., and Insinga, L. P., "The Effects of Calcium Content in Lead-Calcium Alloy Grids on the Performance of Lead-Acid Storage Batteries," Abstract No. 33 Electrochem. Soc. Meeting, Oct. 10–14, 1965, Buffalo, N. Y.; published in *Extended Abstracts of Battery Division*, Volume 10, Ann Arbor, Mich.: Electrochem. Soc., 1965, pp. 87–88.
3. Lander, J. J., "Anodic Corrosion of Lead in H_2SO_4 Solution," J. Electrochem. Soc., *98*, No. 6 (June 1951), pp. 213–219.
4. Lander, J. J., and Burbank, N., "Positive-Grid Corrosion in the Lead-Acid Cell: Corrosion Rates of Tin Alloys and the Effect of Acid Concentration on Corrosion," Naval Research Laboratory Report 4076, November 7, 1952.
5. Lander, J. J., "Further Studies on the Anodic Corrosion of Lead in H_2SO_4 Solutions," J. Electrochem. Soc., *103*, No. 1 (January 1956), pp. 1–8.
6. Burbank, J., "Anodization of Lead and Lead Alloys in Sulfuric Acid," J. Electrochem. Soc., *104*, No. 12 (December 1957), pp. 693–701.
7. Schumacher, E. E., "The Nature of Metals in Relation to their Properties," Sci. Monthly, *34* (January 1932), pp. 22–30.
8. Schumacher, E. E., and Bouton, G. M., "Age Hardening Lead Calcium Alloys," Metals and Alloys, *1* (March 1930), pp. 405–409.
9. Simon, A. C., "Stress Corrosion in the Grids of the Lead-Acid Storage Battery," *114*, No. 1 (January 1967), pp. 1–8.
10. Feder, D. O., Koontz, D. E., Babusci, L. D., and Luer, H. J., "Reserve Batteries for Bell System Use: Design of the New Cell," B.S.T.J., this issue, pp. 1253–1278.

Lead-Acid Battery:

Tetrabasic Lead Sulfate as a Paste Material for Positive Plates

By R. V. BIAGETTI and M. C. WEEKS

We present a reproducible method of synthesizing tetrabasic lead sulfate ($4PbO \cdot PbSO_4$) which produces discrete elongated crystals approximately 22 microns long. Tetrabasic lead sulfate undergoes anodic conversion to PbO_2 while maintaining the characteristic morphology of the $4PbO \cdot PbSO_4$ crystals. This results in lead-acid battery positive plates having performance characteristics superior to those fabricated from conventional paste formulations.

We furnish data showing the performance of $4PbO \cdot PbSO_4$ positive plates as a function of plate porosity and present stress cycling data, comparing it with the behavior of typical conventional positive plates.

I. INTRODUCTION

There are two major causes of positive plate capacity degradation in float service lead-acid batteries: (*i*) corrosion and growth of the lead or lead alloy grid,[1] and (*ii*) softening and disintegration of the PbO_2 matrix. The latter has been extensively discussed by J. Burbank.[2]

In order to improve the stability of the PbO_2 matrix, it would be invaluable to have answers to questions concerning the nature of the grid/PbO_2 interface, the nature of the inter-particle bonding in the PbO_2 matrix and the mechanisms of the charge and discharge reactions. It is paradoxical that, although the lead-acid battery has been in use for over a century, answers to these and other fundamental questions are unknown. This is in large part attributable to the complexity of the lead-acid system which in turn is partly the result of extensive variability of materials and processes used in lead-acid battery fabrication.

In conventional positive plate fabrication, the starting material is usually a mixture of some or all of the following species: Pb, PbO (orthorhombic and tetragonal), and Pb_3O_4. The powder is mixed with

water and sulfuric acid during which many complex chemical reactions occur. The reaction product is a paste generally composed of Pb, PbO, $PbSO_4$, Pb_3O_4, $PbO \cdot PbSO_4$, $3PbO \cdot PbSO_4 \cdot nH_2O$ and $4PbO \cdot PbSO_4$. The exact products and their relative concentrations depend upon numerous variables including: starting materials, ratio of water and acid addition, acid concentration, rate of mixing, uniformity of mixing and temperature. The resulting paste is then applied to lead or lead alloy grids and the plates are cured.

During the curing operation, additional chemical reactions take place. The products remain essentially the same as those obtained in the mixing operation but their relative concentrations may change significantly depending upon temperature and relative humidity during the curing operation. The cured plates are immersed in H_2SO_4 and the paste material is anodized to PbO_2. The modification of PbO_2 (α or β) and the microstructure of the PbO_2 matrix will reflect the particular forming conditions—current density, time, acid concentration—as well as all the process and material variables mentioned before. Such variations in plate properties are well-documented in the technical literature.[3-5] It is clear, then, that efforts directed toward improvements in battery performance must also be directed to development of materials and processes that can be defined and controlled. A study of the positive plate active material (PbO_2) was undertaken with the objective of arriving at a simple, direct technique whereby a reproducible and definable lead dioxide mass might be obtained which would yield improved performance characteristics.

It has been established that mechanical stability of the PbO_2 structure in a lead-acid battery positive plate is enhanced by the presence of large prismatic needles of PbO_2.[2,3] Burbank demonstrated that tetrabasic lead sulfate ($4PbO \cdot PbSO_4$), which crystallized as large elongated prisms, underwent anodic conversion to PbO_2 while maintaining a crystal form similar to the original $4PbO \cdot PbSO_4$ crystals.[6] It was suggested that this particular PbO_2 morphology might impart mechanical strength to the PbO_2 mass by a mechanism involving interlocking of these elongated PbO_2 crystals. It was also suggested that $4PbO \cdot PbSO_4$ might be the precursor to prismatic needles of PbO_2 which were found to be present in satisfactory positive plates.

The apparent potential benefits to be derived from the presence of $4PbO \cdot PbSO_4$ in unformed positive plates suggested that positive plates fabricated entirely from $4PbO \cdot PbSO_4$ might have improved performance characteristics. To evaluate the applicability of tetrabasic lead sulfate

to positive plate fabrication, we studied the synthesis of $4PbO \cdot PbSO_4$ and the performance of positive plates fabricated from this material.

II. SYNTHESIS OF TETRABASIC LEAD SULFATE

Tetrabasic lead sulfate was synthesized by reacting a stirred suspension of orthorhombic PbO in water with 8-12N H_2SO_4 in a 5:1 molar ratio.

(i) The stochiometric quantities of PbO and H_2SO_4 were premeasured.

(ii) Water (approximately 0.85 liters per mole PbO) was preheated to 80°C and acidified to pH2 by addition of a small amount of the premeasured quantity of H_2SO_4.

(iii) The PbO was then added to the acidified water and the mixture stirred so as to keep all the PbO in suspension.

(iv) While maintaining the temperature at 80°C to 85°C, the remaining H_2SO_4 was slowly added to the stirred suspension over a one-hour period.

(v) Following acid addition, heating and stirring were continued for 15 minutes after which the heat source was shut off and stirring continued for an additional 1.5 hours.

(vi) The product was allowed to digest for 12 hours, isolated, and dried at 100°C.

(vii) The identity of the material was verified by X-ray diffraction[7] and the composition checked by lead analyses.[8] The product was also examined with an optical microscope.

Synthesis of $4PbO \cdot PbSO_4$ under the above conditions results in a uniform and reproducible pale yellow material which is composed of well-defined elongated prisms as shown in Fig. 1. This crystal morphology is very similar to that reported previously for $4PbO \cdot PbSO_4$ prepared under similar conditions.[6]

The use of the orthorhombic modification of PbO in the synthesis is essential. When tetragonal PbO is used in identical reaction conditions, the reaction is not uniform and the product is a poorly defined mixture of materials. The preacidification of the water prior to addition of the oxide is also critical. When orthorhombic PbO is added to neutral water, the material converts to the tetragonal modification probably by a mechanism involving dissolution and reprecipitation of PbO. The extent of conversion depends upon the dwell time in the reaction vessel
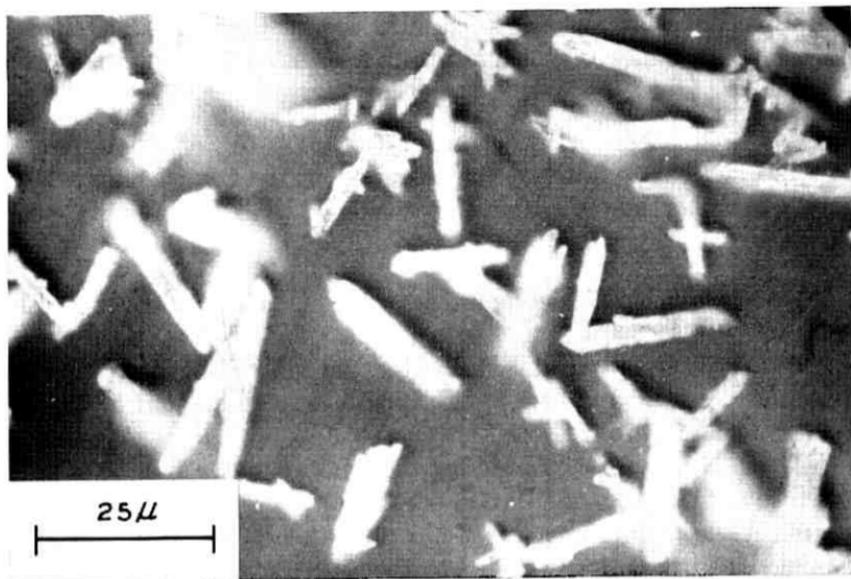
Fig. 1—Photomicrograph of 4PbO·PbSO₄ (500X).

prior to H_2SO_4 addition. When conversion occurs, the H_2SO_4 preferentially reacts with the orthorhombic PbO giving unreacted tetragonal PbO and a product having a sulfate content larger than demanded by the stochiometry of $4PbO \cdot PbSO_4$. The conversion to tetragonal PbO occurs at pH's down to 3.5 although the extent of conversion is less at this lower pH value. At pH2, no conversion is observed and a uniform, reproducible product is obtained.

Control of the temperature to a minimum of 80°C is also essential if one is to obtain a uniform and reproducible product. Below 80°C, a material having a reduced crystal size begins to appear and X-ray diffraction shows the product to be a mixture of compounds. The product prepared at 90°C is identical to that prepared at 80°C and there are no apparent reasons why even higher temperatures could not be used.

Uniformity of the reaction also requires that all the PbO be maintained in suspension. This is a very real problem because of the high density of PbO. Lead oxide which settles in the reaction vessel will not react sufficiently with the added H_2SO_4 and a mixture of products will result.

The 12-hour digestion period does not appear to be essential. Recent experiments indicate that reaction and crystal growth are essentially complete after acid addition is finished and that the product after this period is identical to that isolated after 12-hour digestion.

Finally, the other variables such as PbO/water ratio and acid addition time should be flexible within reasonable limits which will be related to the efficiency of the mixing operation.

To date, roughly 250 batches of $4PbO \cdot PbSO_4$ have been synthesized during this investigation. The batch sizes have ranged between 5 lbs. and 35 lbs., but most have been on the 35 lb. scale. The process has excellent reproducibility and the lead content is readily controllable within the range of 86.6 ± 0.2 percent (Theoretical: 86.60 percent Pb). It was an objective of this study to adequately control the reactants such that the lead content of the product falls within this range. Below 86.4 percent Pb, a product of reduced crystal size becomes apparent. Above 86.8 percent Pb, unreacted PbO appears in the reaction product. The synthesis process also yields a product with reproducible morphology. Detailed particle size analyses[9] were performed on several batches which resulted in the following particle classification:

(i) Average particle length of 22 ± 1 microns with the particle length distribution within any given batch described by a standard deviation of 10 microns.

(ii) Average particle width of 3.5 ± 0.8 microns with the particle width distribution within any given batch described by a standard deviation of 1.7 microns.

III. PREPARATION AND CAPACITY OF TETRABASIC LEAD SULFATE POSITIVE PLATES

A paste of $4PbO \cdot PbSO_4$ suitable for application to lead or lead alloy grids requires only that the powder be mixed with water. Plates pasted with $4PbO \cdot PbSO_4$ are simply air dried, requiring no special curing operations.

The paste does not shrink or crack upon drying, thereby maintaining intimate contact with the grid members. Shrinkage and cracking of paste materials are common problems in the battery industry and a variety of specialized curing operations are often used to overcome these problems. Other studies have shown that the nonshrink property of $4PbO \cdot PbSO_4$ water pastes is attributable to the large crystal size of this material.

We demonstrated this by studying the behavior of plates which were pasted with $4PbO \cdot PbSO_4$ having roughly spherical particles of 3 microns diameter. These were obtained by mechanical grinding of the large $4PbO \cdot PbSO_4$ crystals. Upon drying, these plates showed considerable shrinkage of the material away from the grid frame. Although the

4PbO·PbSO$_4$ in these plates could be formed to PbO$_2$, the effect of shrinkage and loss of contact with the grid frame was reflected in the poor discharge characteristics of these plates. Apparently, the large crystals of 4PbO·PbSO$_4$ (Fig. 1) mechanically interlock to prevent movement of the individual crystals during drying which results in the nonshrink properties of the material.

We obtained detailed performance characteristics of 4PbO·PbSO$_4$ positive plates in the following manner. Five water pastes of 4PbO·PbSO$_4$ of varying densities were prepared by mixing the solid and water in varying proportions. These pastes were applied to preweighed commercially available lead calcium grids (c.a. 0.05 percent Ca). These grids are 2.5″ × 3.0″ × 0.25″ and contain nine paste pellet frames (3 × 3). The pasted plates were air dried at room temperature for two days. The weight of dry paste in each of the five grids ranged from 91 to 110 g.

Five cells were constructed, each consisting of one 4PbO·PbSO$_4$ positive plate located between two comparable pasted negative plates. The 2:1 ratio of negative to positive insured that the cell capacities would be limited by the behavior of the positive plates. The separation between positive and negative plates was 0.25″ and there were no separators placed between the plates. The cells were formed in 1.050 sp. gr. H$_2$SO$_4$ at 0.5 A (33 mA/in^2) for 21 hours followed by 0.3 A (20 mA/in^2) for 240 hours. After this treatment, there was no visual evidence of 4PbO·PbSO$_4$ that had not been converted to PbO$_2$.

The cells were transferred to jars containing 2.5 liters of 1.210 sp. gr. H$_2$SO$_4$. This large volume of H$_2$SO$_4$ was used so that there would be no significant changes in bulk concentration during discharge. The cells were then cycled at various discharge currents ranging from 0.35 A to 2.50 A until all cells gave reproducible results on two successive discharges at the same current. After the cell capacities had stabilized, cell performance data were obtained at 0.13, 0.35, 0.70, 1.00, 1.30 and 2.50 amperes. Plate potentials were measured versus mercury/mercurous sulfate reference electrodes and a positive/reference electrode potential of 0.850 V was taken to be the end of discharge. This positive plate potential (0.850 V) corresponded to a cell voltage of approximately 1.80 V.

The cells were then disassembled. The positive plates were washed in water for 3 hours and dried at 80°C for 18 hours. The PbO$_2$ pellets were removed from each plate and the total amount of PbO$_2$ in each plate was determined by weighing. Porosity data on the PbO$_2$ pellets were obtained with a Numinco Model MIC 901 mercury intrusion porosimeter.

Figure 2 shows the performance of $4PbO \cdot PbSO_4$ plates as a function of porosity. Each data point represents an average of at least two and sometimes three porosity measurements. The dashed line is the 5-hour rate performance (0.90 A) interpolated from the data at the other rates. For the porosity range studied, the data at discharge rates between 2.50 A and 0.35 A show a linear increase in material utilization with increasing porosity.

Although it is generally accepted that positive plate efficiency increases as plate porosity increases, to the authors' knowledge this is the only quantitative data of this type that have been published. Since increasing porosity of the PbO_2 mass facilitates diffusion of the H_2SO_4 electrolyte into the matrix, the performance data in Fig. 2 are strong evidence for a diffusion controlled reaction.

At the low discharge rate of 0.13 A the linear porosity-utilization relationship does not hold (no curve has been drawn through these points). This is not unreasonable since one would expect that for any given plate, there should be a maximum discharge rate below which
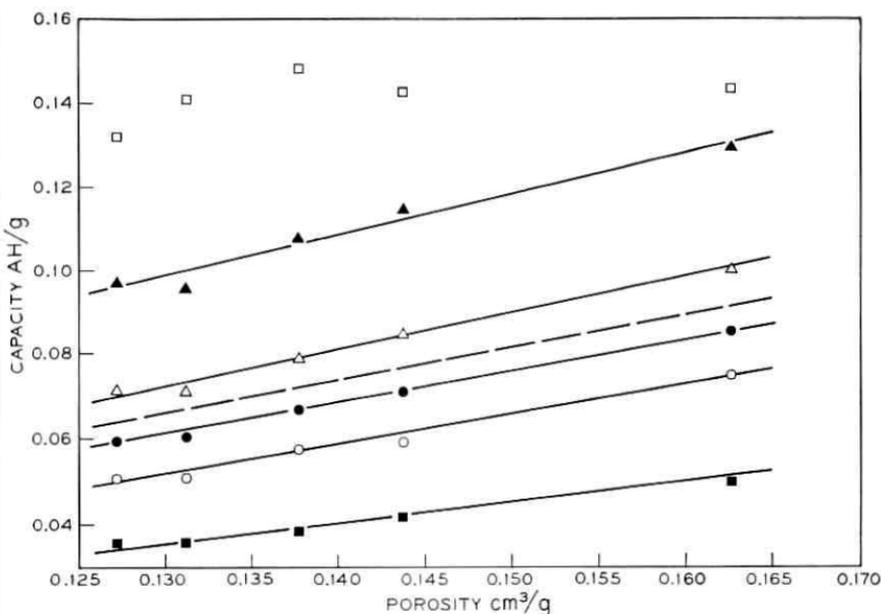

Fig. 2—Capacity (AH/g) vs Porosity (cm³/g) for $4PbO \cdot PbSO_4$ plates.

■ 2.50 A	○ 1.30 A	● 1.00 A
△ 0.70 A	▲ 0.35 A	□ 0.13 A

Dashed line is interpolated 0.90 A (5-hour rate) performance.

the diffusion of electrolyte is sufficient to satisy the demands of the discharge.

For discharge rates sufficiently low where diffusion is not limiting, material utilizations which are independent of porosity would be expected. The 0.13 A data show this to be the case. Excluding the lowest porosity data, the utilizations are reasonably independent of porosity. It may be that 0.13 A is borderline for diffusion control for the porosity range studied here which would account for the lower utilization of the plate having the least porosity.

A porosity of 0.130 cm^3/g corresponds to positive plate loading densities typical of present commercial products. At 0.90 A, which is the rated 5-hour discharge current for this plate, a porosity of 0.130 cm^3/g corresponds to a utilization of approximately 0.065 AH/g. This is equivalent to a utilization of 29 percent (theoretical = 0.224 AH/g) which compares quite favorably with 25–30 percent utilizations realized from commercial plates of comparable design at the 5-hour rate. In addition, data has been accumulated on approximately 2000 plates of varying designs and sizes which correspond closely to positive plate sizes in current commercial product. At tetrabasic lead sulfate loading densities of 68–70 g/in^3 these plates consistently yield capacities of 30 AH/lb. at the 5-hour rate. This is equivalent to 30 percent utilization which again compares favorably with current commercial product.

Figure 3 presents the capacity-porosity data in a different format. The data show that, over the range studied, the less material per unit volume, the greater the capacity per unit volume. These data show that a reduction in the material loading density per unit volume results in an increase in plate efficiency which more than compensates for the reduction in capacity to be expected from the reduced quantity of material in the grid. These unexpected results have very interesting implications. For example, if the 1.0 A data in Fig. 3 is extrapolated to lower material densities, a loading density of 55 g/in^3 will correspond to a utilization of 6.1 AH/in^3. This, in comparison to 4.25 AH/in^3 at a "normal" density of 69 g/in^3, represents a 43.5 percent increase in capacity for a 20.3 percent decrease in material. Unfortunately, there are problems associated with the use of low material densities. As will be seen later, lowering material densities results in reduced plate life. In addition, there is the practical problem of how to achieve low $4PbO \cdot PbSO_4$ plate loading densities. In general, lower loading density plates are obtained by increasing the ratio of water to $4PbO \cdot PbSO_4$ in the preparation of the paste. However, as the ratio of water to $4PbO \cdot PbSO_4$ is increased, the paste becomes more fluid and a point
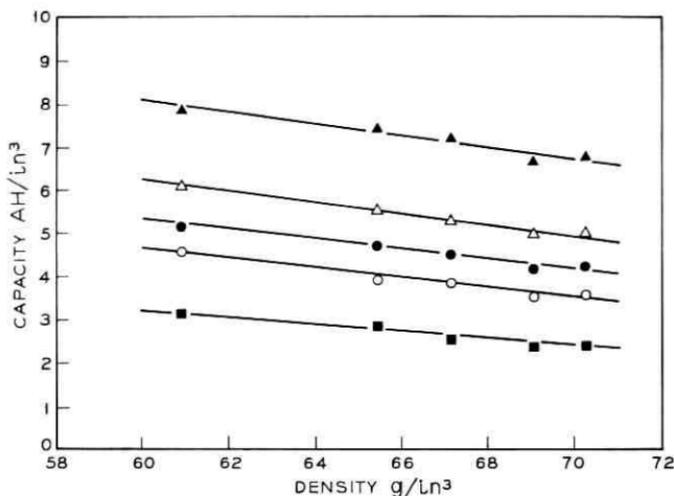
Fig. 3—Capacity (AH/in³) vs density (g/in³) for 4PbO·PbSO₄ plates.

□   2.50 A   O   1.30 A   ●   1.00 A   △   0.70 A   ▲   0.35 A

is reached where the paste is too fluid for application to a grid. A paste density of 60 g/in^3 appears to be the useful low limit for water pastes of $4PbO \cdot PbSO_4$.

In the work reported here, ten days formation was required to convert the $4PbO \cdot SO_4$ plates to PbO_2. In many other experiments, which are not detailed here, 7–10 days formation in 1.050 sp. gr. H_2SO_4 were required to achieve complete conversion of $4PbO \cdot PbSO_4$ to PbO_2. This is significantly longer than the 2–3 days required to form conventional positive paste formulations. The reduced formation efficiency is a result of the comparatively large particle size of $4PbO \cdot PbSO_4$ and is not a property of the chemical composition of the material. This was demonstrated by studying the formation characteristics of

$$4PbO \cdot PbSO_4$$

whose particle size was mechanically reduced. These plates converted completely to PbO_2 after only 2–3 days formation. This strongly suggests that the efficiency of the formation process is related to the material surface area.

IV. CYCLE BEHAVIOR OF TETRABASIC LEAD SULFATE POSITIVE PLATES

Continuous cycling and overcharge of lead-acid cell positive plates accelerate degradation of the PbO_2 matrix. In order to evaluate the

stability of $4PbO \cdot PbSO_4$ positive plates as compared to conventional plates, stress cycling experiments were carried out. These experiments were also designed to evaluate the effect of porosity on the stability of $4PbO \cdot PbSO_4$ positive plates. The details of the experiments are as follows. Five preweighed grids ($2.5'' \times 3.0'' \times 0.25''$), identical to those described previously, were pasted with water pastes of $4PbO \cdot PbSO_4$ of varying densities and air dried at room temperature for 48 hours. We obtained two other plates having identical grids from a commercial battery manufacturer and included them in this experiment for comparative purposes.

Cells were assembled as described previously and were formed in 1.050 sp. gr. H_2SO_4. The cells having the $4PbO \cdot PbSO_4$ positive plates were formed at 400 mA for 10 days whereas the two cells having the commercial positive plates were formed at 400 mA for 2 days followed by 250 mA for 2.5 days. After formation, all positive plates were washed with water, dried at 80°C for 3 hours and weighed to determine the amount of material in the grids.

Each cell was reassembled and transferred to jars containing 2.5 liters of 1.210 sp. gr. H_2SO_4. The large volume of H_2SO_4 was used for the reason stated previously. The seven cells were connected in series and given 7 discharge-charge cycles. The first four discharges were at 1.00 A and the last three at 0.90 A. All discharges were to a 1.75 V end-voltage. After each of these discharges, the series string was recharged over a period of approximately 16 hours at such a rate that 120 percent of the discharge capacity was returned to the highest capacity cell in the string. The positive plates were then removed and the three PbO_2 pellets in the top row of the grid were removed leaving six pellets in each grid. The plates were returned to the cells and the removed PbO_2 pellets were washed with water and dried at 80°C. This pellet material was examined with an optical microscope, studied by X-ray diffraction and analyzed for PbO_2 content.[10]

The reassembled cells were then placed on the following cycle routine. All cells in the string were given a deep discharge at 600 mA (5-hour rate) to 1.75 V once a week and a 1.0 hour discharge at 600 mA every other working day during the week. The deep discharges were followed by approximately 16 hours charge at such a rate that 120 percent of the discharge capacity was returned to the highest capacity cell in the string. The 1.0 hour discharges were followed by 100 percent overcharge over a 23-hour period. This cycle routine was continued until all cells gave less than 4 hours capacity at the 5-hour rate.

A photomicrograph of the unformed paste material from the com-

merical plates used in this experiment is shown in Fig. 4. The material is actually a mixture of white and reddish crystals of varying sizes and differs markedly in both crystal size and shape from $4PbO \cdot PbSO_4$ shown in Fig. 1. Figures 5 and 6 are photomicrographs of PbO_2 from a $4PbO \cdot PbSO_4$ and a commercial plate respectively after the initial seven cycles. X-ray diffraction showed the PbO_2 derived from both the $4PbO \cdot PbSO_4$ and commercial plates to the β-modification. X-ray diffraction analysis of many other $4PbO \cdot PbSO_4$ plates formed in 1.050 sp. gr. H_2SO_4 showed that $\beta \cdot PbO_2$ was exclusively produced. The PbO_2 derived from $4PbO \cdot PbSO_4$ is composed of rather large prismatic crystals resembling the original $4PbO \cdot PbSO_4$ crystals. This observation confirms an earlier similar report.[6] On the other hand, the PbO_2 obtained from the commercial plates is of a much smaller particle size and the crystal shape is not obvious at this magnification.

The results of the cycling experiments are presented in Fig. 7. The material densities were calculated from the measured weights of PbO_2 and the known pasteable volume of the grids. The initial data for the $4PbO \cdot PbSO_4$ grids show, in general, higher initial capacities for those plates having less material. This is in agreement with the quantitative observations reported earlier in Fig. 3. The $4PbO \cdot PbSO_4$ plate data

25μ
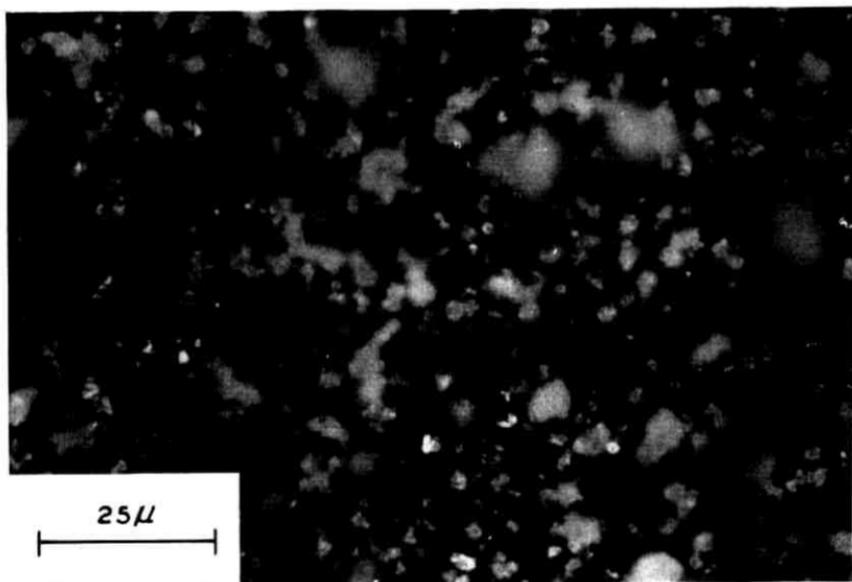
Fig. 4—Photomicrograph of commercial unformed positive paste material (500X).
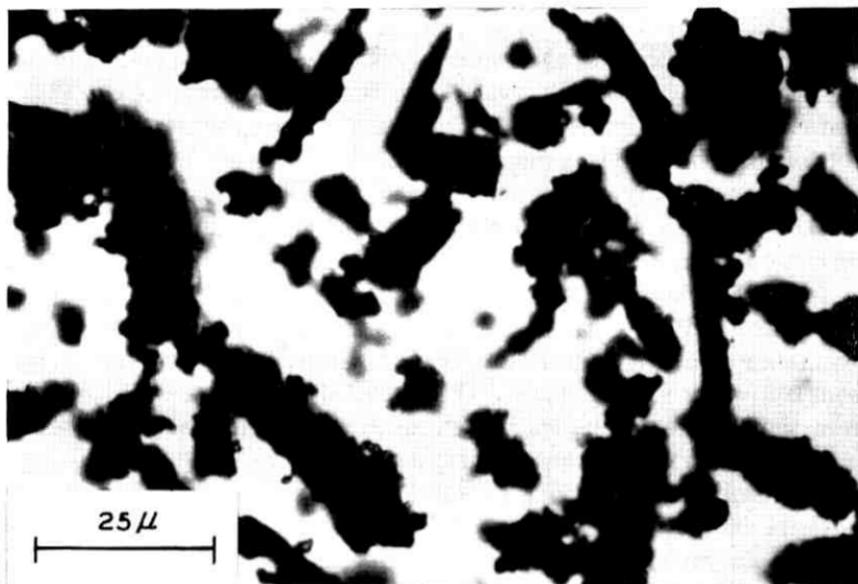
Fig. 5—Photomicrograph of βPbO_2 from $4PbO \cdot PbSO_4$ (500X).
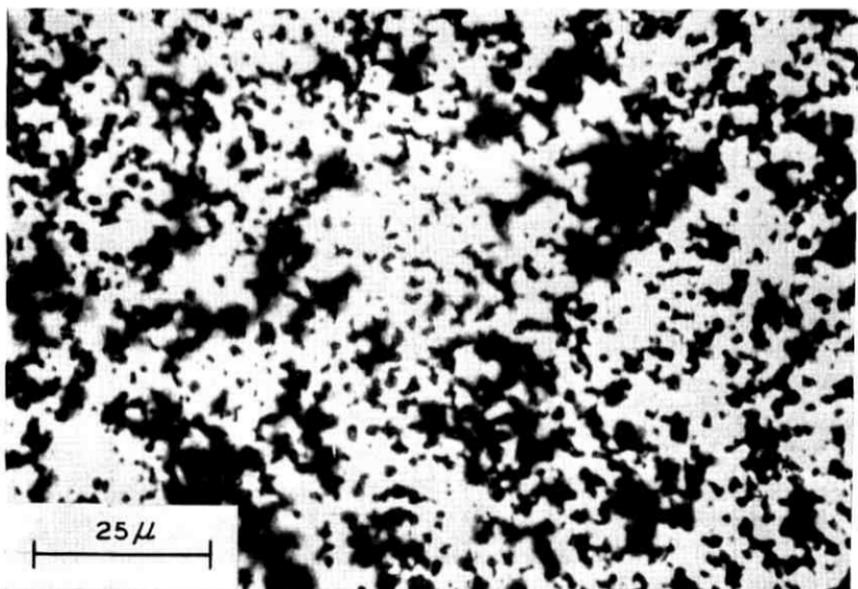
Fig. 6—Photomicrograph of βPbO_2 from commercial plates (500X).
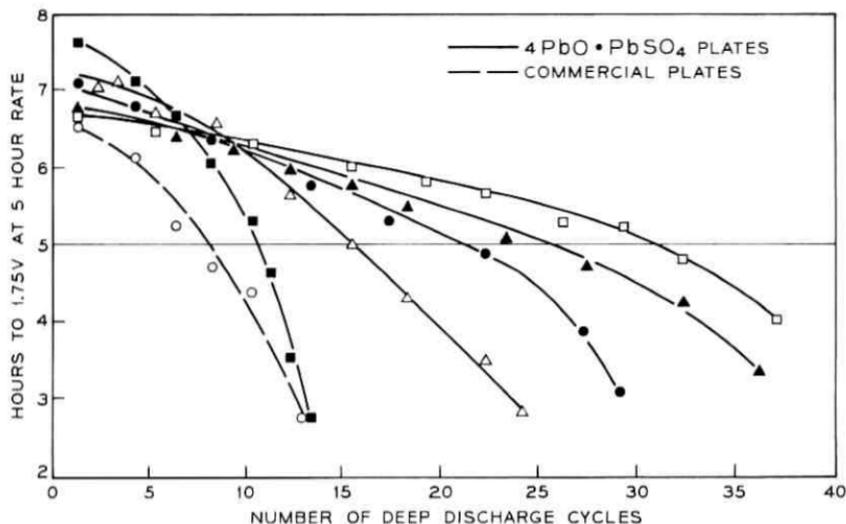
Fig. 7—Time to 1.75 V vs number of deep discharge cycles at the 5 hour rate (600 mA) for 4PbO. PbSO₄ and commercial plates.

○ 66.7 g/in³   ■ 57.7 g/in³   △ 59.5 g/in³
● 66.8 g/in³   ▲ 64.2 g/in³   □ 69.4 g/in³

show a general correlation between material density and cycle life. Cycling of positive plates induces stresses in the PbO_2 matrix since the conversion of βPbO_2 to $PbSO_4$ is associated with a 57 percent increase in volume. The molar volumes are 27.9 cm^3/mole for PbO_2 and 43.8 cm^3/mole for $PbSO_4$. Regardless of whether the stability of the PbO_2 matrix is attributable to mechanical interlocking of the PbO_2 crystals or chemical bonding between PbO_2 particles (it is probably a combination of both), the matrix stability would be related to the extent of inter-particle contact. Consequently, plates with higher material densities would be expected to exhibit longer cycle life than low density plates and, as Fig. 7 shows, this is indeed the case. In all cases the $4PbO \cdot PbSO_4$ plates, regardless of material density, gave better cycle lives than the commercial plates. The data for the two commercial plates are not differentiated since both plates behaved identically.

Capacity failure of all plates was caused by shedding of PbO_2 and the rates of capacity degradation reflect the rates of PbO_2 shedding for these plates. In a comparison of the $4PbO \cdot PbSO_4$ and commercial plates based solely upon material density, the cycle life of the commercial plates would be expected to be in the 20–25 cycle region whereas only 8 cycles were obtained. This demonstrates that particle morphology

plays an important role in determining the PbO_2 matrix stability.

There is an apparent inconsistency in the $4PbO \cdot PbSO_4$ plate data in that the 64.2 g/in^3 plate gave better cycle performance than the 66.8 g/in^3 plate. The densities were calculated from the known grid pasteable volume. Although care was taken to minimize overpasting or underpasting of the grids, slight variations would result in errors in the calculated densities which may account for the above discrepancy.

Pellet samples from all cells were analyzed for PbO_2 to determine if there was any correlation between material density and PbO_2 contents of the $4PbO \cdot PbSO_4$ plates and also to determine if there were any significant analytical differences between the commercial plates and the $4PbO \cdot PbSO_4$ plates. The PbO_2 contents of the seven plates ranged between 90.2 percent and 94.3 percent but they were not significant in relation to the cycle test results.

Although these results involve a comparison with only two plates from the same manufacturer, they are consistent with results reported for similar experiments on lead calcium grids.[11,12] It is reasonable to conclude that the superior performance of the tetrabasic lead sulfate positive plates can be attributed to the characteristic size and shape of the PbO_2 particles obtained from $4PbO \cdot PbSO_4$.

V. CONCLUSIONS

Tetrabasic lead sulfate, having definable and reproducible chemical and physical properties, can be synthesized providing that the reaction parameters are properly controlled. The use of the material in positive plate fabrication is simple and straightforward. Water is the only liquid used in paste preparation and no special plate curing conditions are required. Since water is used, the definability and reproducibility of the $4PbO \cdot PbSO_4$ material enables one to accurately define and control the properties of the resulting paste which allows control of plate performance.

The formed tetrabasic lead sulfate plates yield initial capacities which compare very favorably to current commercial product. A drawback is the additional formation time required. Various methods of reducing the formation time are being investigated. Some of these have been very successful and the effects of these on plate performance are presently being evaluated.

The use of $4PbO \cdot PbSO_4$ results in a significant improvement in the cycle behavior of lead-acid positive plates. Although the cycle data cannot be directly extrapolated to actual float service life, it is clear

that we have realized a substantial improvement in the mechanical stability of the PbO_2 structure. This is consistent with the objective of improving float service life of lead-acid cells used in telephone service.

REFERENCES

1. Cannone, A. G., Feder, D. O., and Biagetti, R. V., "Positive Grid Design Principles," B.S.T.J., this issue, pp. 1279–1303.
2. Burbank, J., "Batteries," *Proc. 3rd International Battery Symposium*, Bournemouth, 1962, New York: Pergamon Press, 1963, pp. 43–60.
3. Burbank, J., "Morphology of PbO_2 in the Positive Plates of Lead Acid Cells," J. Electrochem. Soc., *111*, No. 7 (July 1964), pp. 765–768.
4. Burbank, J., and Ritchie, E. J., "PbO_2 in the Lead Acid Cell," J. Electrochem. Soc., *116*, No. 1 (January 1969), pp. 125–130.
5. Simon, A. C., and Jones, E. L., "Crystallogeneses in the Forming of Plates for the Lead Acid Storage Battery," J. Electrochem. Soc., *109*, No. 9 (September 1962), pp. 760–770.
6. Burbank, J., "Anodic Oxidation of the Basic Sulfates of Lead," J. Electrochem. Soc., *113*, No. 1 (January 1966), pp. 10–14.
7. American Society for Testing and Materials, Power Data File, 6–0308.
8. Vogel, A. I., *Textbook of Quantitative Inorganic Analysis*, New York: John Wiley & Sons, Inc., 1961, pp. 459–587.
9. American Society for Testing and Materials, E-20 62T.
10. Vogel, A. I., *Textbook of Quantitative Inorganic Analysis*, New York: John Wiley & Sons, Inc., 1961, pp. 374–382.
11. Burbank, J., "The Role of Antimony in Positive Plate Behavior in the Lead Acid Cell," J. Electrochem. Soc., *111*, No. 10 (October 1964), pp. 1112–1116.
12. Tudor, S., Weisstuck, A., and Davang, S. H., "The Lead Calcium Grid Battery with Phosphoric Acid and Additive," Electrochem. Tech., *4*, No. 7–8 (July–August 1966), pp. 406–411.

.

Lead-Acid Battery:

Float Behavior of the Lead-Acid Battery System

By PAUL C. MILNER

We discuss in this article the requirements for satisfactory float operation of lead-acid batteries and analyze behavior during float in terms of the kinetic parameters which characterize the electrochemical processes occurring in the overcharge region. We show how knowledge of these parameters may be used to estimate whether adequate battery maintenance can be expected under specified float conditions. The results emphasize the need for cell uniformity, and the treatment provides a basis for determining what variability is permissible.

I. INTRODUCTION

The design of any battery involves many factors, some to be dealt with empirically and some by analysis in light of more fundamental understanding. These factors combine to determine end behavior and, consequently, utility; detailed considerations of them with regard to the Bell System lead-acid battery appear elsewhere in this series of papers. Here, we consider the more general topic of float operation of lead-acid batteries with as much concern for fundamentals as seems appropriate, using the behavior of typical rectangular lead-calcium grid cells to illustrate the treatment.

Although stationary lead-acid batteries, used for load-averaging and reserve power in the Bell System, are commonly maintained by connection to a constant voltage bus, a procedure known as floating, there is surprisingly little published information about the details of this process and the effects of cell and electrode characteristics on its operation. The standard text on storage batteries[1] makes only brief mention of float operation and, in fact, provides very little information even about the overcharge behavior of cells.

The purpose of float is, of course, to maintain a battery in a full state

of charge and to return it to that state after it has undergone discharge. This is accomplished by selecting a bus voltage slightly greater than the open-circuit voltage of the fully-charged battery. To minimize operating voltage variations, it is desirable that the bus voltage be as low as possible. This has the further advantage of avoiding excessive over-charging which increases maintenance and reduces life. On the other hand, if the bus voltage is too low, the rate of recharging may not be adequate, the positive or negative plates may not be kept charged, and the positive grids may start to corrode at a catastrophic rate.

Float operation is, therefore, a matter of a fairly delicate balance as well as a mode of operation dependent on the overall electrical charac-teristics of a battery rather than on some separate indication of its condition. As a result, for satisfactory float operation, the individual cells of a battery must be not only capable of being properly maintained by such a regime but also compatible with one another. The first of these requires a matching or balancing of positive and negative plate behavior in the cells; the second, a matching of cells in the battery. Design, uniformity in manufacture, and conditions of use determine for the most part how well these requirements will be met. Since the basic electrochemical phenomena involved are known (if not completely understood), it is possible to examine in a useful fashion the effects of some of the variables.

The following sections cover, in turn, plate behavior, cell behavior, and battery behavior as they relate to steady-state float operation under conditions relevant to Bell System usage.

II. PLATE BEHAVIOR

The electrochemical reactions of importance in the lead-acid battery are the following, each listed with its calculated, thermodynamically reversible potential in the standard 1.210 sp. gr. H_2SO_4 at 25°C against a mercury/mercurous sulfate reference electrode in the same solution.

$$PbSO_4 + 2H_2O \leftrightarrows PbO_2 + 4H^+ + SO_4^= + 2e^-,$$

$$V_{PbSO_4/PbO_2} = 1.090 \text{ V}; \quad (1)$$

$$2H_2O \leftrightarrows O_2 + 4H^+ + 4e^-, \quad V_{H_2O/O_2} = 0.620 \text{ V}; \quad (2)$$

$$Pb + 2H_2O \leftrightarrows PbO_2 + 4H^+ + 4e^-, \quad V_{Pb/PbO_2} = 0.059 \text{ V}; \quad (3)$$

$$(2Hg + SO_4^= \leftrightarrows Hg_2SO_4 + 2e^-, \quad V_{Hg/Hg_2SO_4} = 0 \text{ V});$$

$$H_2 \leftrightarrows 2H^+ + 2e^-, \quad V_{H_2/H^+} = -0.612 \text{ V}; \quad (4)$$

$$Pb + SO_4^= \leftrightarrows PbSO_4 + 2e^-, \quad V_{Pb/PbSO_4} = -0.971 \text{ V}. \quad (5)$$
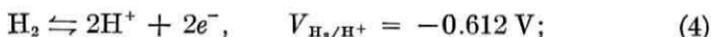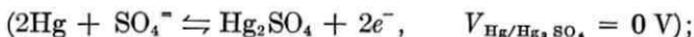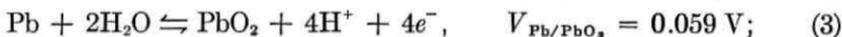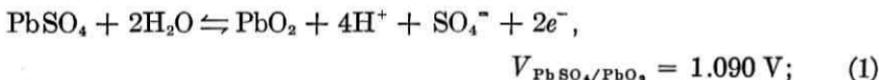
The first and last are the charge-discharge reactions of the positive and negative active materials which provide the capacity of the cell and give it a fully-charged open-circuit potential of $1.090 - (-0.971) = 2.061$ V. The temperature coefficient of the open-circuit potential is about 0.25 mV/°C.[1] At the positive plate, oxygen is evolved at a kinetically controlled rate and the lead of the grid material is oxidized. At the negative plate, hydrogen is evolved at a kinetically controlled rate, and, if oxygen is present, it will be reduced at a diffusion-controlled rate. The corresponding oxidation of hydrogen does not, for kinetic reasons, occur at an appreciable rate at the positive. The reactions of other electroactive species, which may be present in practical systems, are not considered here.

Electrochemical processes under kinetic control generally follow a current density-overpotential relation of the form

$$i = i_0[\exp(\alpha_a F \eta / RT) - \exp(-\alpha_c F \eta / RT)], \tag{6}$$

where i_0, the exchange current density, depends on the concentrations of reacting species and the temperature; α_a and α_c, the anodic and cathodic transfer coefficients, are usually relatively independent of such factors as well as of the potential; and η, the overpotential, represents the departure of the electrode potential from its thermodynamically reversible value, which depends, of course, on concentrations and temperature. F, R, and T are the Faraday, the gas constant and the absolute temperature.

At open circuit and on overcharge in the lead-acid system, the overpotentials for oxygen and hydrogen evolution are large [$\eta_{O_2} \geqq 1.090 - 0.620 = 0.470$ V and $\eta_{H_2} \leqq -0.971 - (-0.612) = -0.359$ V] and only one of the exponential terms is important. This gives rise to Tafel behavior or linear potential-log current relationships for these reactions, and it is convenient for the following discussion to write these in terms of the plate polarizations, that is, the departures of the plate potentials (V_+ and V_-) from those corresponding to the reversible potentials of the charge-discharge reactions, rather than in terms of the overpotentials of the oxygen and hydrogen evolution reactions themselves. Thus, instead of

$$\eta_{O_2} = V_+ - V_{H_2O/O_2} = \frac{RT}{\alpha_{a,O_2}F} \ln \frac{i_{O_2}}{i_{0,O_2}},$$

$$\eta_+ = V_+ - V_{PbSO_4/PbO_2} = \frac{RT}{\alpha_{a,O_2}F} \ln \frac{i_{O_2}}{i_{0+}}, \tag{7}$$

and instead of

$$\eta_{H_2} = V_- - V_{H_2/H^+} = -\frac{RT}{\alpha_{c,H_2}F} \ln \frac{-i_{H_2}}{i_{0,H_2}},$$

$$\eta_- = V_- - V_{Pb/PbSO_4} = -\frac{RT}{\alpha_{c,H_2}F} \ln \frac{i_{H_2}}{i_{0-}}. \tag{8}$$

i_{0+} and i_{0-} are thus the extrapolated oxygen and hydrogen evolution current densities at the positive and negative plate open-circuit potentials, and $(RT/\alpha_{a,o_2}F)$ ln (10) and $-(RT/\alpha_{c,H_2}F)$ ln (10) are the corresponding Tafel slopes, b_+ and b_-.

For a given plate design with a particular thickness and method of manufacture, the ratio of effective surface area to ampere-hour capacity at a specified hourly rate is essentially constant, and it is therefore possible, in the Tafel relationships, to use currents per ampere-hour of capacity in place of current densities. Similarly, for a given cell design, the ratio of positive plate ampere-hour capacity to negative plate ampere-hour capacity is essentially constant, and both Tafel relationships can therefore be written, for a given family of cells, in terms of currents per ampere-hour of positive plate capacity, which normally determines cell capacity. Expressed in this fashion, the extrapolated oxygen and hydrogen evolution currents at the positive and negative plate open-circuit potentials at 25°C are, for typical rectangular lead-calcium grid cells, I_{0+} = 5–10 μA/Ah and I_{0-} = $-(5$–$10)\mu$A/Ah, respectively.[2] The corresponding Tafel slopes at 25°C are b_+ = 70 mV/decade and b_- = -110 mV/decade[2].

This behavior is reasonably well understood on the negative, for, under slightly different conditions of measurement, a battery electrode of 0.55 cm thickness[3] and a planar lead surface[4] have been found to have essentially the same hydrogen evolution characteristics per true (BET) unit area, with $i_{0-} \cong -10^{-4}\mu$A/cm^2 and $b_- \cong -120$ mV/decade for both. The situation with the positive is more complex because of the existence of two modifications of lead dioxide, α and β, and of an as yet unresolved controversy over whether the characteristics of oxygen evolution on the two are the same.[5,6] According to Giner and others,[6] the two forms differ only because of porosity under the conditions where Rüetschi and his coworkers[5] found Tafel slopes of 60 and 110 mV/decade for the α and β forms, respectively, and nearly the same values of i_{0+}, $\sim 10^{-2}\mu$A/cm^2, for both, the true areas being estimated from capacitance measurements. The original work on capacitance,[3] however, included results which indicate $i_{0+} \cong 5 \times 10^{-4}\mu$A/BET cm^2 for a 0.60 cm

thick positive, suggesting that the total surface area of the porous plate is not used in oxygen evolution.

The temperature coefficients of η_+ and η_- for the previously mentioned typical lead-calcium cells are -2.88 mV/°C and $+2.57_4$ mV/°C, respectively.[2] Taking i_{0+} and i_{0-} (and likewise I_{0+} and I_{0-}) to follow the general relation

$$i_{0\pm} = i_{0\pm}^* \exp\left[-E_\pm/RT\right], \tag{9}$$

equations (7) and (8) give

$$\frac{\partial \eta_\pm}{\partial T} = \frac{\eta_\pm}{T} - b_\pm \frac{E_\pm}{\ln{(10)}RT^2}. \tag{10}$$

Assuming, as typical values at 25°C, $\eta_+ = 55$ mV with $b_+ = 70$ mV/decade and $\eta_- = -55$ mV with $b_- = -110$ mV/decade, equation (10) gives $E_+ = 17.8$ kcal/mole and $E_- = 10.2$ kcal/mole. (Neglect of the term $\eta_\pm/T$ gives 16.7 kcal/mole and 9.5 kcal/mole.) Although these lie in the normal range of activation energies for kinetically-controlled processes, they should not be considered as such since η_+ and η_- are related to the open-circuit plate potentials corresponding to reactions (1) and (5) and are therefore not true overpotentials.

At the positive plate, in addition to oxygen evolution, grid corrosion, reaction (3), occurs, with the slow oxidation of lead through a passivating film of PbO_2 . This is, currently, the process which determines the life of properly designed and manufactured lead-calcium cells. As a result, this process has received a good deal of attention but, unfortunately, not much fundamental understanding. With regard to its effect on float operation and the effect of float conditions on it, the following appears to be a reasonable estimate of the pertinent features.

The rate of grid corrosion is, in the short run, relatively independent of potential.[7] It does, however, increase with temperature and, ultimately, with time in processes associated with the end of cell life. As might be expected, increased temperatures reduce the time to the final, irreversible increase, but, equally important, corrosion studies indicate that extended periods (months to years) of positive plate polarizations which are significantly less than about 25 mV have a similar effect. This, then, is an important factor in float operation, and it is generally considered that adequate maintenance under float conditions requires $\eta_+ > 25$ mV. At 25°C, normal rates of lead-calcium grid corrosion correspond to current densities of the order of 0.2 $\mu A/cm^2$ of grid surface or penetration rates of about 3×10^{-4} cm/yr (1.2×10^{-4} in/yr).[7] Typical rectangular positive plates for stationary batteries have about

20 cm^2 of grid surface per ampere-hour of positive capacity. The contribution of corrosion current to float can therefore be expressed in the same units as I_{0+} and I_{0-}, giving $I_c \cong 4$ μA/Ah at 25°C^2. While the potential dependence of I_c can, as a first approximation, be neglected, the temperature dependence is important. Assuming, as for I_{0+} and I_{0-}, that I_c follows a relation of the form

$$I_c = I_c^* \exp [-E_c/RT], \tag{11}$$

the available data[7] indicate a value for E_c for lead-calcium grids of the order of 10.5 kcal/mole. This is a purely empirical quantity and cannot be given any physically meaningful interpretation.

At the negative plate, in addition to hydrogen evolution, the reduction of any oxygen present in solution takes place at a rate controlled by diffusion and thus independent of potential. In 1.210 sp. gr. H$_2$SO$_4$ at 25°C, the solubility of oxygen is about 7×10^{-4} moles/liter,[8] and if its diffusion coefficient is taken to be about 10^{-5} cm^2/sec, the diffusion-limited current due to oxygen reduction from a saturated solution can be calculated from Fick's first law to be about $-3(A/\delta)\mu$A/cm, where A is the effective cross-sectional area through which diffusion to the plate surface can occur and δ is the effective thickness of the diffusion layer. In practice, A may be expected to be largely determined by the area and characteristics (porosity, and so on) of the separator and δ by its thickness, since it is the separator that really defines the paths by which oxygen generated at the positive has access to the negative. For a given cell design (plate thickness), the separator area is proportional to the positive capacity, and thus, like I_{0+}, I_{0-}, and I_c, the oxygen reduction current can be expressed on an ampere-hour basis. For the previously mentioned typical cell at 25°C, this is $I_d = -(20-35)$ μA/Ah,[2] giving, with the Fick's law estimate, $a/\delta \cong 7-12$ cm/Ah, where a is the effective cross-sectional area per ampere-hour. These same factors control the electrolyte resistance, typically 0.14 ohm Ah.[9] With the electrolyte resistivity equal to about 1.2 ohm-cm, this gives $a/\delta \cong 9$ cm/Ah, in good agreement with the previous estimate. The temperature dependence of I_d is given by the relation

$$I_d = I_d^* \exp [-E_d/RT] \tag{12}$$

with $E_d = 13.5$ kcal/mole.[2] In cells, as discussed later, there is one further important restriction on the magnitude of $-I_d$: it cannot, in the long run, exceed the oxygen evolution current at the positives.

For a float current $I_f = I_{0_2} + I_c = -I_{H_2} - I_d$, expressed like the others on an ampere-hour basis, the foregoing results can be combined

to give

$$\eta_+ = b_+ \log \frac{I_f - I_c}{I_{0+}} \tag{13}$$

for the positive and

$$\eta_- = b_- \log \frac{I_f + I_d}{-I_{0-}} \tag{14}$$

for the negative. To include the temperature dependences of the various parameters more explicitly, it is useful to re-write equations (13) and (14) in terms of their values at 25°C, which are here denoted with a superscript zero. Then,

$$\eta_+ = \frac{(T/°K)}{298.15} b_+^0 \left[\log \frac{I_f - r_c I_c^0}{I_{0+}^0} - \log \frac{I_{0+}}{I_{0+}^0} \right] \tag{15}$$

and

$$\eta_- = \frac{(T/°K)}{298.15} b_-^0 \left[\log \frac{I_f + r_d I_d^0}{-I_{0-}^0} - \log \frac{I_{0-}}{I_{0-}^0} \right] \tag{16}$$

where $r_c = I_c/I_c^0$ and $r_d = I_d/I_d^0$. For the typical cell discussed above,

$$b_+^0 = 70 \text{ mV/decade},$$

$$b_-^0 = -110 \text{ mV/decade},$$

$$I_{0+}^0 = 5\text{--}10 \ \mu\text{A/Ah},$$

$$I_{0-}^0 = -(5\text{--}10)\mu\text{A/Ah},$$

$$I_c^0 \cong 4 \ \mu\text{A/Ah},$$

$$I_d^0 = -(20\text{--}35)\mu\text{A/Ah},$$

$$\log \frac{I_{0+}}{I_{0+}^0} = 3.89 \left(\frac{10^3}{298.15} - \frac{10^3}{(T/°K)} \right), \quad \text{(for } E_+ = 17.8 \text{ kcal/mole)};$$

$$\log \frac{I_{0-}}{I_{0-}^0} = 2.23 \left(\frac{10^3}{298.15} - \frac{10^3}{(T/°K)} \right), \quad \text{(for } E_- = 10.2 \text{ kcal/mole)};$$

$$\log r_c = \log \frac{I_c}{I_c^0} \cong 2.29 \left(\frac{10^3}{298.15} - \frac{10^3}{(T/°K)} \right), \text{ (for } E_c \cong 10.5 \text{ kcal/mole)};$$

$$\log r_d = \log \frac{I_d}{I_d^0} = 2.94 \left(\frac{10^3}{298.15} - \frac{10^3}{(T/°K)} \right), \text{ (for } E_d = 13.5 \text{ kcal/mole)}.$$

These data are used in the following sections as a basis for the con-

sideration of float operation. They refer, of course, only to a "typical" lead-calcium cell design, the characteristics of which are representative of one type of cell suitable for this kind of service.

III. CELL BEHAVIOR, PARAMETERS FOR TYPICAL LEAD-CALCIUM CELLS

When a constant float voltage, V_f, is maintained across a cell, it is the cell polarization which is controlled. This is given by $\eta_{cell} = V_f - V_{cell}^{o.c.} = \eta_+ - \eta_-$ and is related to the float current I_f which flows through both positives and negatives by equations (15) and (16). Thus, specification of V_f and the plate parameters determines the condition of the cell during float operation. While the discussion here deals only with the steady-state, it is clear that a cell inadequately maintained in these circumstances will be no better maintained in actual use. From the previous section, adequate maintenance of positive plates is considered to require $\eta_+ > 25$ mV so that grid corrosion is not accelerated. Adequate maintenance of the negatives requires only that $\eta_- < 0$, but in both cases, some margin is obviously desirable.

The quantities of principal interest, then, are I_f, η_+, and η_-. In the light of the previous discussion, these are taken to be determined by the parameters I_{0+}^0, I_{0-}^0, I_d^0, T, and V_f. In all of the calculations, b_+^0, b_-^0, and I_c are assumed invariant and, together with the temperature dependent factors, are assigned the values given earlier. I_{0+}^0, I_{0-}^0, and I_d^0 depend on design and method of manufacture and can vary considerably. Cells with values lying within the limits given for the "typical" cell of the previous section are described in the following as "normal range" cells. As a point of reference, a cell with mid-range values of these parameters is hereafter described as a "median" cell. Temperature and float voltage depend, of course, on conditions of use and can also vary. Standard conditions, as in the Bell System, are taken here to be 25°C and 2.170 V/cell.

For a given set of parameters, I_f and then η_+ and η_- can be calculated from equations (15) and (16), taking into account two restrictions:

(i) η_+ must be positive or zero and η_- must be negative or zero. Solutions giving negative η_+ or positive η_- are physically inadmissible, for in reality the charge-discharge reactions poise the plate potentials at zero polarization and the plates then discharge. At $\eta_+ = 0$, the net discharge rate of the positive is $I_f - I_{0+} - I_c$; at $\eta_- = 0$, that of the negative is $I_f + I_{0-} + I_d$.

(ii) The rate of oxygen reduction at the negative cannot exceed the rate of oxygen evolution at the positive; that is, $-I_d$ can be no greater

than $I_f - I_c$ or, when $\eta_+ = 0$, I_{0+}. Solutions for which this is not so are also physically inadmissible; instead, I_d can be assumed to become approximately equal to $I_c - I_f$ or I_{0+}. In the case of the former, from equation (14), the negative polarization becomes independent of I_f.

The following results are by no means all-inclusive but indicate, instead, some of the effects to be expected.

3.1 *Median Cell, Standard Float Conditions*

$I_{0+}^0 = 7.5$ $\mu A/Ah$, $I_{0-}^0 = -7.5$ $\mu A/Ah$, $I_d^0 = -27.5$ $\mu A/Ah$, $V_f = 2.170$ V/cell at 25°C. In this case,

$$\eta_{cell}/mV = 2170 - 2061 = 109,$$

$$= 70 \log [(I_f - 4)/7.5] + 110 \log [(I_f - 27.5)/7.5],$$

and $I_f = 50.5$ $\mu A/Ah$, $\eta_+ = 55$ mV, $\eta_- = -54$ mV. This represents, more or less, the average behavior of a "typical" or, perhaps more accurately, a desirable cell. Both plate polarizations are good and float operation should be most satisfactory under these conditions.

3.2 *Normal Range Cells, Standard Float Conditions*

$I_{0+}^0 = 5{-}10$ $\mu A/Ah$, $I_{0-}^0 = -(5{-}10)\mu A/Ah$, $I_d^0 = -(20{-}35)\mu A/Ah$, $V_f = 2.170$ V/cell at 25°C. The ranges in the following tables correspond to the range in I_{0+}, 5–10 $\mu A/Ah$.

$I_f/(\mu A/Ah)$	$I_d^0 = -20$ $\mu A/Ah$	$I_d^0 = -35$ $\mu A/Ah$
$I_{0-}^0 = -5$ $\mu A/Ah$	35.3–41.2	47.4–52.8
$I_{0-}^0 = -10$ $\mu A/Ah$	45.5–54.8	56.8–65.7

η_+/mV	$I_d^0 = -20$ $\mu A/Ah$	$I_d^0 = -35$ $\mu A/Ah$
$I_{0-}^0 = -5$ $\mu A/Ah$	56–40	66–48
$I_{0-}^0 = -10$ $\mu A/Ah$	64–49	72–55

η_-/mV	$I_d^0 = -20$ $\mu A/Ah$	$I_d^0 = -35$ $\mu A/Ah$
$I_{0-}^0 = -5$ $\mu A/Ah$	$-(53{-}69)$	$-(43{-}61)$
$I_{0-}^0 = -10$ $\mu A/Ah$	$-(45{-}60)$	$-(37{-}54)$

In all cases the plate polarizations are adequate and float operation should be satisfactory under the standard conditions.

3.3 *Median Cell, Standard Float Voltage, Varied Temperature*

$I_{0+}^0 = 7.5$ $\mu A/Ah$, $I_{0-}^0 = -7.5$ $\mu A/Ah$, $I_d^0 = -27.5$ $\mu A/Ah$, $V_f = 2.170$ V/cell at 5, 15, 25, 35, and 45°C.

	5°C	15°C	25°C	35°C	45°C
$I_f/(\mu A/Ah)$	11.3	24.4	50.5	100.0	190.3
η_+/mV	70	63	55	48	41
η_-/mV	−44	−49	−54	−58	−63

Although there is some reduction in cell polarization with increased temperature because of the rise in the open-circuit cell potential, the median cell still has what would be considered an adequate positive polarization even at the extreme of 45°C (113°F) and should be capable of float at this temperature. However, this would hardly be recommended, for the higher temperature would bring about an increase in the positive grid corrosion current and a consequent decrease in life.

3.4 *Median Cell and Cells at Extremes of Normal Range Behavior, Low (2.150 V/cell) Float Voltage at 25°C*

The lowest positive polarization in the normal range cells occurs for $I_{0+}^0 = 10 \mu A/Ah$, $I_{0-}^0 = -5 \mu A/Ah$, $I_d^0 = -20 \mu A/Ah$, while the highest occurs for $I_{0+}^0 = 5 \mu A/Ah$, $I_{0-}^0 = -10 \mu A/Ah$, $I_d^0 = -35 \mu A/Ah$. The median cell has $I_{0+}^0 = 7.5 \mu A/Ah$, $I_{0-}^0 = -7.5 \mu A/Ah$, and $I_d^0 = -27.5 \mu A/Ah$.

	Lowest η_+	Median Cell	Highest η_+
$I_f/(\mu A/Ah)$	35.5	44.1	50.6
η_+/mV	35	51	68
η_-/mV	−54	−38	−21

In each case, the plate polarizations are adequate and float operation at this slightly lowered voltage should be satisfactory.

3.5 *Varied I_{0+}^0 in Otherwise Median Cell, Standard Float Conditions*

With $I_{0-}^0 = -7.5 \mu A/Ah$, $I_d^0 = -27.5 \mu A/Ah$, and $V_f = 2.170 V/cell$ at 25°C, an increase in I_{0+}^0 from 7.5 $\mu A/Ah$ results in a lowering of η_+ and a corresponding increase in the negative polarization. η_+ drops to 25 mV as I_{0+}^0 increases to 29.4 $\mu A/Ah$, and here the float current is 71.0 $\mu A/Ah$ and $\eta_- = -84$ mV. (η_+ goes to zero when $I_{0+}^0 = 96.9$ $\mu A/Ah$, at which point $I_f = 100.9 \mu A/Ah$ and $\eta_- = -109$ mV.) A decrease in I_{0+}^0 increases η_+ and decreases η_- , and η_- goes to zero when I_{0+}^0 reaches 0.859 $\mu A/Ah$. Here $I_f = 35.0 \mu A/Ah$ and $\eta_+ = 109$ mV. In an otherwise median cell, then, the maximum range of I_{0+}^0 for adequate plate polarizations at 2.170 V/cell and 25°C is only 0.859 to 29.4 $\mu A/Ah$, and this could be considerably reduced at lower float voltages and other temperatures.

3.6 *Varied I_{0-}^0 in Otherwise Median Cell, Standard Float Conditions*

With $I_{0+}^0 = 7.5\ \mu A/Ah$, $I_d^0 = -27.5\ \mu A/Ah$, and $V_f = 2.170\ V/cell$ at 25°C, as I_{0-}^0 becomes more negative than the median $-7.5\ \mu A/Ah$, η_+ increases and η_- approaches zero, reaching that point at $I_{0-}^0 = -247.0\ \mu A/Ah$, where $I_f = 274.5\ \mu A/Ah$ and $\eta_+ = 109\ mV$. As I_{0-}^0 becomes less negative, the reverse takes place, but before η_+ reaches 25 mV, $I_f - I_e$ becomes equal to $-I_d$. This takes place at $I_{0-}^0 = -0.934\ \mu A/Ah$, with $I_f = 31.5\ \mu A/Ah$, $\eta_+ = 39\ mV$ and $\eta_- = -69\ mV$. Beyond this point, assuming $I_d = I_e - I_f$ which makes η_- independent of I_f, the positive polarization decreases, reaching 25 mV at $I_{0-}^0 = 0.689\ \mu A/Ah$, with $I_f = 21.1\ \mu A/Ah$ and $\eta_- = -84\ mV$. (η_+ goes to zero at $I_{0-}^0 = -0.408\ \mu A/Ah$, with $I_f = 11.5\ \mu A/Ah$ and $\eta_- = -109$ mV.) In an otherwise median cell, then, the maximum range of I_{0-}^0 for adequate plate polarizations at 2.170 V/cell and 25°C is -0.689 to $-247.0\ \mu A/Ah$, a range which could be considerably reduced under other float conditions.

IV. BATTERY BEHAVIOR; FLOAT CHARACTERISTICS OF BATTERIES OF TYPICAL LEAD-CALCIUM CELLS

The previous discussion of cells is obviously applicable to batteries provided they consist of identical cells at the same temperature. This is not usually the case because of normal variations in the manufacturing process, as indicated by the range of parameters given for the "typical" cell, and because of non-uniform temperatures under some conditions of use. The requirements for adequate plate maintenance remain the same ($\eta_+ > 25\ mV$, $\eta_- < 0$), but the behavior of the cells in the battery during float now depends on the distribution of the battery polarization among the cells as well as on the distribution of the resulting cell polarizations between the plates. For an n cell battery,

$$nV_f - \sum_n V_{cell}^{o.c.} = \sum_n (\eta_+ - \eta_-),\qquad(17)$$

where V_f is the float voltage per cell and η_+ and η_- are given for the different cells by the appropriate versions of equations (15) and (16). The following results illustrate some situations of interest.

4.1 *12 Cell Battery, Cells from Extremes of Normal Range Behavior, Standard (2.170 V/cell at 25°C) Float Conditions*

The lowest positive polarization in single cell float occurs for $I_{0+}^0 = 10\ \mu A/Ah$, $I_{0-}^0 = -5\ \mu A/Ah$, $I_d^0 = -20\ \mu A/Ah$ and the highest for $I_{0+}^0 = 5\ \mu A/Ah$, $I_{0-}^0 = -10\ \mu A/Ah$, $I_d^0 = -35\ \mu A/Ah$.

(i) With 1 cell from the low η_+ group and 11 from the high η_+ group, $I_f = 55.7$ μA/Ah, the low η_+ cell has $\eta_+ = 50$ mV, $\eta_- = -94$ mV and the high η_+ cells have $\eta_+ = 71$ mV, $\eta_- = -35$ mV. All the plate polarizations are adequate and float operation should be satisfactory, with cell voltages of 2.205 V and 2.167 V.

(ii) With 6 cells from each group, $I_f = 49.9$ μA/Ah, the low η_+ cells have $\eta_+ = 46$ mV, $\eta_- = -85$ mV, and the high η_+ cells have $\eta_+ = 67$ mV, $\eta_- = -19$ mV. As before, the polarizations are adequate, with cell voltages of 2.192 V and 2.147 V.

(iii) With 11 cells from the low η_+ group and 1 from the high η_+ group, the standard float conditions do not provide proper maintenance of the negative in the high η_+ cell, and its polarization goes to zero. The float current, I_f, is 42.7 μA/Ah and in the low η_+ cells, $\eta_+ = 41$ mV and $\eta_- = -72$ mV, which are adequate and give cell voltages of 2.174 V. The high η_+ cell has $\eta_+ = 62$ mV and, with $\eta_- = 0$, the net negative discharge rate is $42.7 - 35 - 10 = -2.3$ μA/Ah at a cell voltage of 2.123 V. For satisfactory operation at 25°C, the float voltage must be increased to more than 2.176 V/cell, at which point the float current is 45.0 μA/Ah and the negative in the high η_+ cell is no longer discharging though its polarization is still zero.

4.2 12-Cell Battery, Median Cells, 2.170 V/cell with One Cell Hot

The median cells have $I_{0+}^0 = 7.5$ μA/Ah, $I_{0-}^0 = -7.5$ μA/Ah, and $I_d^0 = 27.5$ μA/Ah.

With 11 cells at 25°C (77°F) and 1 at 35°C (95°F), the polarization on the hot negative goes to zero. I_f is 53.3 μA/Ah and the 25°C cells have $\eta_+ = 57$ mV and $\eta_- = -59$ mV. In the 35°C cell, $\eta_+ = 26$ mV, which is marginal at best, and really neither set of plates is adequately maintained. The net negative discharge rate in the hot cell is -6.0 μA/Ah, here given by $I_e + I_{0-}$ because the rate of oxygen reduction at the negative is limited by the rate of oxygen generation at the positive under these conditions. This is, of course, a somewhat unlikely situation with regard to temperature but is of interest because individual median cells can be floated adequately even at 45°C.

V. DISCUSSION

While a good deal yet remains to be understood about the processes occurring in the lead-acid system, its behavior during overcharge can at least be formally described in terms of a limited number of experi-

mentally accessible parameters, some of which can be varied by design. These parameters determine the float characteristics and whether or not a cell or battery is capable of being satisfactorily maintained by a particular float condition, as has been discussed. It should be emphasized, perhaps, that in the present state of knowledge, these parameters must be determined experimentally in the assembled cell, modification of them by design is still largely an empirical business, and calculations of float capability must ultimately be verified by float tests.

Nevertheless, it is useful to consider the effects of variations in the parameters, for this gives some estimate of the permissible latitude in design, manufacture and use. For both cells and batteries consisting of identical cells maintained under identical conditions, it has been shown in the foregoing that satisfactory float operation can be expected for quite a range of characteristic parameters and float conditions.

On the other hand, for batteries which do not consist of identical cells which are identically maintained, there is far less latitude. Even certain combinations of what have here been termed "normal range" cells cannot be satisfactorily floated under standard conditions. Variations in temperature among otherwise identical cells can also lead to difficulty. While both these problems have long been recognized, it requires a quantitative treatment to specify the variability that can be tolerated. In general, satisfactory float operation is unlikely to be achieved without good uniformity among cells and their operating conditions.

Finally, there are quantities besides η_+ and η_- which may be of interest in characterizing cell conditions during float. One is the amount of float current being passed in excess of that required to keep the net negative discharge current equal to zero. This is given by $I_f + I_d + I_{0-}$. Another is what might be called the effective dc impedance of the negative, $\partial \eta_- / \partial I_f$, which is given by $b_- \log e / (I_f + I_d)$. These and others are easily calculated.

VI. ACKNOWLEDGMENTS

REFERENCES

1. Vinal, G. W., *Storage Batteries*, 4th ed., New York City: John Wiley and Sons, 1955.
2. Willihnganz, E., unpublished work.

3. Rüetschi, P., Ockerman, J. B., and Amlie, R., "Surface Coverage during Hydrogen and Oxygen Evolution," J. Electrochem. Soc., *107*, No.4 (April 1960), pp. 325–332.
4. Rüetschi, P., and Cahan, B. D., "Anodic Corrosion and Hydrogen and Oxygen Overvoltage on Lead and Lead Antimony Alloys," J. Electrochem. Soc., *104*, No. 7 (July 1957), pp. 406–413.
5. Rüetschi, P., Sklarchuk, J., and Angstadt, R. T., "Stability and Reactivity of Lead Oxides," *Batteries: Proceedings of the 3rd International Symposium*, Collins, D. H., editor, London: Pergamon, 1963, pp. 89–103.
6. Giner, J., Gancy, A. B., and Makrides, A. C., "Preparation and Characterization of Lead Dioxide Electrodes for Reserve Batteries," Final Report to Harry Diamond Laboratories under Contract No. DA-49-186-AMC-265(X), April 1967.
7. Baker, R. A., "The Anodic Oxidation Rate of Pure Lead in Sulfuric Acid," Extended Abstracts, Battery Division, Electrochem. Soc. Houston Meeting, October 10–12, 1960, pp. 53–55, and unpublished results.
8. Gubbins, K. E., and Walker, R. D., Jr., "The Solubility and Diffusivity of Oxygen in Electrolytic Solutions," J. Electrochem. Soc., *112*, No. 5 (May 1965), pp. 469–471.
9. Willihnganz, E., and Rohner, P., "Battery Impedance: Farads, Milliohms, Microhenries," paper presented at A.I.E.E. 1959 Winter General Meeting, New York.

Lead-Acid Battery:

Electrochemical Characterization of the Bell System Battery: Field Trials of the Battery

By T. D. O'SULLIVAN, R. V. BIAGETTI and M. C. WEEKS

(Manuscript received April 29, 1970)

The electrochemical characteristics of pure lead circular grids, pasted plates and seperators were studied individually and then extrapolated to encompass a full battery. We present polarization data for grids and pasted plates. These can be scaled to predict battery behavior when one takes the role of the separator into consideration.

Proper venting of entrapped gases results in capacity behavior of the circular cell which is at least equal to equivalent commercial cells. The overcharge or float behavior was shown in the laboratory to be excellent. Field trials have confirmed the laboratory tests and have shown that the battery has excellent capacity and float characteristics which ensure that this capacity will be retained for long periods of time.

I. INTRODUCTION

In a preceding paper P. C. Milner has outlined a quantitative description of the float characteristics of the lead-acid battery in terms of the various electrode reactions occurring under float conditions and has summarized some of the available data on lead-calcium cells by providing numerical values for the various parameters.[1] In this paper we present the results of measurements on the polarization of the two electrodes of the new Bell System lead-acid battery and analyze them in the framework of Milner's model. On the basis of these measurements, we can make certain statements about the float operation of the new cell.

A second topic covered in this paper is the capacity of the cell and the factors that affect it: discharge rate and cell design, including electrode configuration and venting arrangements. We include, also,

1335

some results of field trials that have been carried out in the New Jersey Bell Murray Hill Central Office and four AT & T Long Lines locations.

II. FLOAT CHARACTERISTICS

To determine the float characteristics of the new cell, it has proved advantageous to study separately the polarizations of the positive and negative plates. During the early stages of the work, it was found that the current drawn by an unpasted grid at some specified overpotential was quite appreciable in comparison with that drawn by a fully formed, fully charged pasted plate at the same overpotential, in spite of the fact that the surface areas, as determined by the B. E. T. technique, are grossly different. In Sections 2.1 and 2.2, therefore, we present polarization data on both pasted and unpasted grids, first for the positive and then the negative.

2.1 *Polarization of Positive Plates and of Unpasted Grids Held Near the Positive Plate Reversible Potential*

Polarization data were obtained by standard techniques. Design "B" pure lead circular positive grids were used in this investigation.[2] The properties of this grid design are as follows: 2.6 lbs. grid weight; 164 in.^{2} grid surface area; 15.2 in.^{3} pasteable volume; 5.0 lbs. pasted plate weight. All pasted positive plates referred to in this paper were pasted with tetrabasic lead sulfate.[3] The plate to be measured was placed in a tank of 1.210 specific gravity sulfuric acid. The tank also contained appropriate circular counter-electrodes and separators made of microporous rubber, placed approximately 5 cm from the plate. Electrode potential measurements were made by connecting through a H_2SO_4 salt bridge to a standard Hg/Hg_2SO_4 reference electrode.

All measurements to be reported in this section were carried out at room temperature. Measurements were carried out either galvanostatically or at constant voltage. The results obtained by these two different methods were found to agree; however, since the galvanostatic method was found in practice to yield reproducible data more rapidly, this method was used for most of the measurements. To perform a set of measurements, the current was set at the desired value and potential measurements using a high impedance digital voltmeter that is accurate to 0.3 mV was begun.

Typically, a time interval of two to ten days would be required for the potential to reach a steady value; at the higher currents a steady state was reached more rapidly. No effort was made to stir the electro-

lyte or to control the dissolved oxygen level. Once a stable electrode potential had been reached, the current would be changed and a new value for the potential obtained in the same manner. A typical experimental run of this kind occupied approximately two months; on this time scale, the polarization data were usually reasonably reproducible and independent of the order in which the points had been obtained.

Measurements were also made on grids that had been held for a prolonged interval (3 months) at a large positive overpotential (> 1.340 V vs Hg/Hg_2SO_4). Here again it was found that a reproducible polarization curve was obtained, displaced, however, from the initial polarization curve.

Figure 1 shows the results of typical sets of measurements and we begin by noting that the following reactions may contribute to the current required to polarize the positive electrode:

$$PbSO_4 + 2H_2O \rightleftarrows PbO_2 + 4H^+ + SO_4^- + 2e^-, \tag{1}$$

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^-, \tag{2}$$

$$Pb + 2H_2O \rightarrow PbO_2 + 4H^+ + 4e^-, \tag{3}$$

$$\text{Oxidation of organic contaminants.} \tag{4}$$
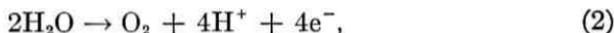
The first thing to note is that Fig. 1 shows that current passed by the unpasted grid at an overpotential similar to that found in Bell System
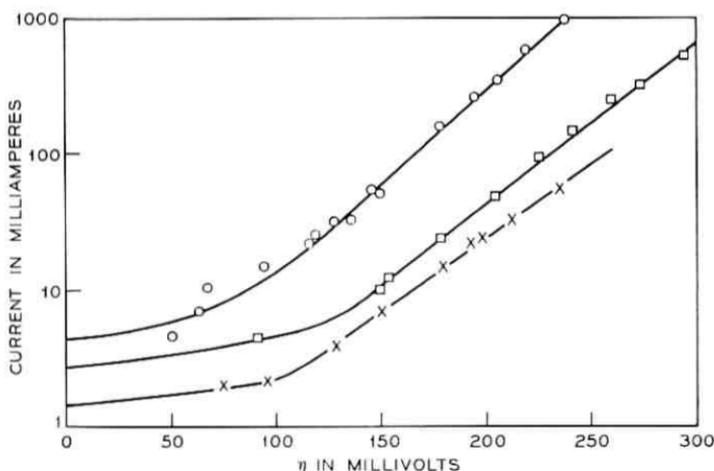

Fig. 1—Polarization data for pasted and unpasted pure-lead circular positive grids. Positives: O, pasted plates; ×, □, bare grids; ×, initial; □, after prolonged high polarization.

usage is only a factor of four lower than that passed by the pasted plate, whereas the surface areas are obviously very different. Measurements of surface area, made by the B. E. T. method, indicated in fact that the ratio of real area to superficial area of the pasted plate is about 20,000 : 1.[4] The current density distribution on porous electrodes is very complex and work is in progress to aid our understanding of this fundamental problem. The data show, however, that the PbO_2 paste utilizes proportionately only a very small fraction of the current supplied to the pasted grid and this implies that the PbO_2 paste material plays a relatively minor role in the evolution of oxygen based on its overall surface area [reaction (2)]. The second thing to note is that the currents passed by the unpasted plates at a given overpotential increase after being held for a prolonged period at a relatively large positive overpotential. We attribute this increase in current to the enhanced corrosion at high overpotentials and the subsequent increase in the true surface area of the electrode.[2]

It would, of course, be desirable to be able to separate the observed currents into the four component parts represented by equations (1) to (4). In particular, it would be of the very greatest interest to determine how much of the current is to be ascribed to reaction (3), since it is this reaction—the corrosion of the grid—that limits the life of the battery.[5-6] Unfortunately, we cannot report on this at this time. The contribution of reaction (1) can be obtained by careful Tafel measurements at very low overpotentials and reaction (4) can be eliminated by using very pure systems. Thus one is left with reactions (2) and (3). Where a soluble species is involved, the associated current may be detected by rotating ring-disc methods. Since the O_2 that is evolved is a soluble species that can be reduced at the ring, one can obtain the fraction of the current that goes into grid corrosion. We have obtained some preliminary ring-disc measurements and will report on this at a later date.

Fitting the data to equation (13) of Milner's paper, we obtain values for the parameters b_+ , and I_0^+ shown in Table I. We also define a parameter "I_e" which is the difference between the intercept of the extension of the Tafel line to $\eta_+ = 0$ and the true extrapolated float current at $\eta_+ = 0$. In using Milner's equation (13), we have in effect obtained a first-order fit to the departure from the Tafel law by subtracting out a constant current which we call "I_e", neglecting, as does Milner, the potential dependence of the corrosion current which Baker reported.[7] Otherwise, the following points deserve comment. The values for b_+ on unpasted grids are consistently higher than the

TABLE I—ELECTROCHEMICAL CHARACTERISTICS
OF BTL POSITIVE ELECTRODES

	b_+ (mV/decade)	"I_e"* ($\mu A/AH$)	I_0+* ($\mu A/AH$)
Pasted Plates	73	60 ± 10	8.5
Unpasted Grids 1 (Initial)	87	26	3.0
Unpasted Grids 2	84	41	3.4

* Data for both pasted and unpasted plates were normalized by dividing by 53 AH which is the 5 hour rated capacity for the positive pasted plate.

"typical value" of 70 mV cited by Milner for lead-calcium pasted plates. Data obtained by us on commercial Pb/Ca pasted grids show values of b_+, I_0 and "I_e," which are comparable to our data obtained for the pure lead pasted plates indicating that the effect is not alloy dependent but results from omission of the paste. This is to be expected since, at time zero at least, there are no fundamental reasons why the Pb/Ca and pure Pb systems should have different Tafel slopes (b_+). These values would be expected to change, however, as the positive plates age and the area for corrosion and oxygen evolution increase.[2] This is indicated by our observation that holding the grids at a relatively high overpotential increased I_0^+ and "I_e." As was stated previously, we attribute this to enhanced corrosion at high overpotentials which has resulted in an increase in the true surface area of the electrode.

2.2 *Polarization of Negative Plates and Negative Grids Held Near the Negative Plate Potentials*

Experimental data were obtained in exactly the same way as those for the positive plates. The geometry, counter-electrode, separator and reference electrode were all the same; the electrolyte was unstirred and no effort was made to control the oxygen content. Again, measurements were made for the most part galvanostatically; steady-state potentials were typically reached in two to ten days. Since corrosion of the negative electrode does not occur, no effect was anticipated of holding this electrode for a long period under high negative polarization conditions. Experimental results have verified this.

The physical properties of the negative plates studied are summarized later (see Section 3.1). Pasted negative plates used in this study were pasted with a commercial negative paste formulation. Figure 2 shows the results of a typical set of measurements on unpasted grids and on pasted plates. We note again that the current passed by the unpasted
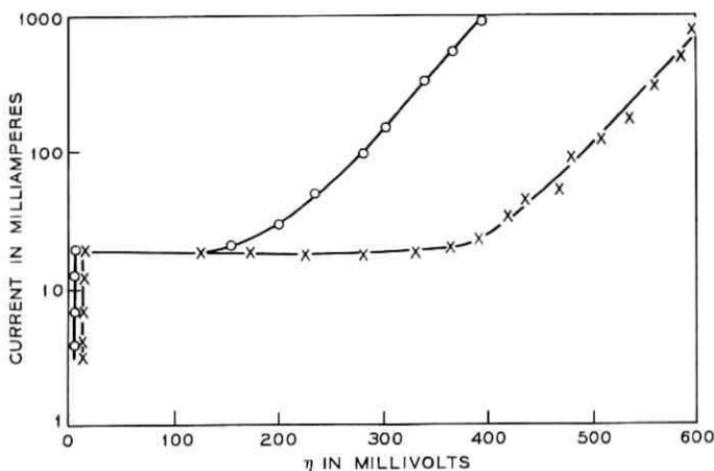
Fig. 2—Polarization data for pasted and unpasted pure-lead circular negative grids. Negatives: O, pasted plate; ×, bare grid.

grids is surprisingly high, although the disparity between the currents and the measured B. E. T. areas is not quite so dramatic as for the positive grids or plates. For a given overpotential, the unpasted grids draw about 3 percent of the current drawn by the pasted plates; the areas, on the other hand, were found to be approximately in the ratio 3000 : 1.[4] Analysis of this discrepancy is further complicated by the influence of adsorbed expanders on the hydrogen evolution reaction.

Figure 2 indicates Tafel behavior at high negative overvoltages, with an additional, potential-independent component showing up in the low overpotential region. Following Milner,[1] we ascribe the Tafel behavior to the reaction:

$$2H^+ + 2e^- \rightarrow H_2 \tag{5}$$

while the potential-independent current is ascribed to the inverse of reaction (ii):

$$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O. \tag{6}$$

Fitting to Milner's equation (14), we derive b— and I_0— from the Tafel asymptote; the limiting current gives I_d directly. Results are shown in Table II.

We note first that the values for b— come close to the theoretical value of 118 mV. The values for I_d are substantially higher than those quoted by Milner for complete cells. As we shall see later, the limiting currents observed in the new Bell System cells are themselves 1 to 2

TABLE II—ELECTROCHEMICAL CHARACTERISTICS
OF BTL NEGATIVE ELECTRODES

	b_- (mV/decade)	I_{0-}* (μA/AH)	I_d* (μA/AH)
Pasted Plates	117	6.5	380
Unpasted Grids	120	.15	380

* Data for both pasted and unpasted plates were normalized by dividing by 53 AH which is the 5 hour rated capacity for the positive pasted plate.

orders of magnitude lower than those in Table I, consistent with Milner's data. We explain the apparent discrepancy with the data shown in Table II as follows. The value of I_d depends on the rate at which oxygen can be transported to the negative electrode. Since, in the experimental geometry used for studying the polarization of isolated plates, the separator is relatively remote from the plates, considerable convection can occur in this system. This would allow more efficient transport of O_2 to the negative plate than would be encountered in a compactly constructed real cell. The potential of the negative electrode would then be poised close to the reversible value at all currents less than I_d as is shown in Fig. 2.

2.3 *Temperature-Dependence of the Polarization at the Positive Plate*

An understanding of the activation energies for all of the various processes occurring under float is of importance, not only because of the guidance it should give as to the preferred float conditions for operation in a given temperature range, but also because it is alternately needed to justify the extrapolations of cell life estimates[2] from measurements at elevated temperatures. The problem is one of great complexity and is not yet fully resolved. The only measurements which we wish to describe at the present time are some that we have made on the temperature-dependence of the current drawn by unpasted grids held in the oxygen evolution domain and all polarized approximately 80 mV at each temperature. Results of these measurements are indicated in Fig. 3.

The activation energy derived from Fig. 3 is approximately 18 K cal/ mole which is similar to that found for the oxygen evolution reaction on Pt.[8] If the evolution of oxygen at the unpasted grid proceeds by a mechanism similar to that at a pasted plate, we may tentatively iden-
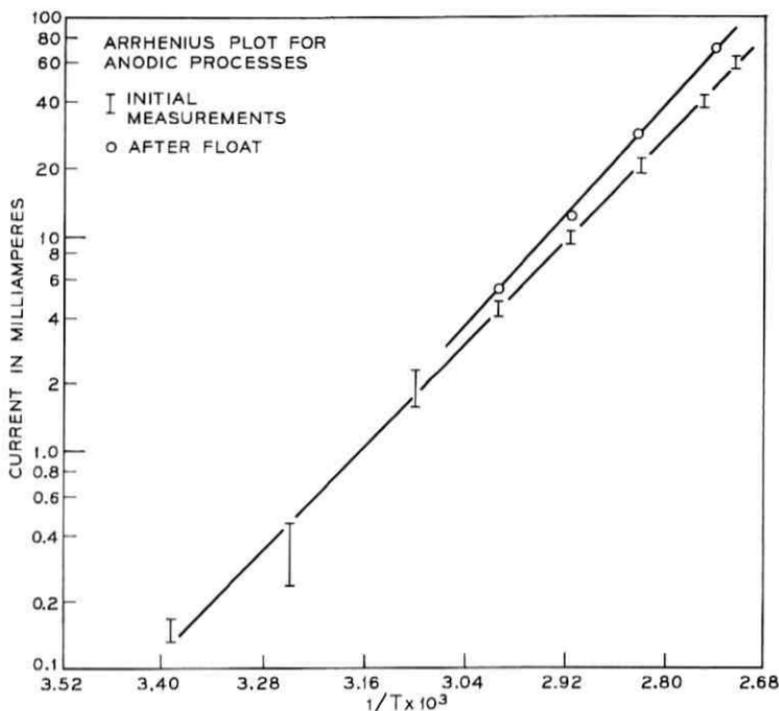
Fig. 3—Log current vs $1/T$ for a pure lead grid held anodic at 80 mV polarization at test temperatures from 25–95°C.

tify this number with the E_+ of Milner's paper, the value of which for "typical cells" is quoted by him as 17.8 K cal/mole, in good concurrence.

2.4 Float Measurements on Complete Cells

We have made a substantial number of polarization measurements on complete cells. A typical result is shown in Fig. 4. If the conditions within the cell were exactly identical to those prevailing in the experimental arrangements described in the preceding sections, we should be able to derive Fig. 4 by adding the polarizations for pasted plates shown in Figs. 1 and 2. It is obvious at a glance that this cannot be done. However, the slopes in Fig. 4 in the high overvoltage region as well as the intercepts do agree reasonably well with the slopes and intercepts of the two Tafel asymptotes in Figs. 1 and 2; it is only in the low overvoltage region that discrepancies occur. Inspection shows the principal difference to be the displacement of the steep part of the negative polarization curve, which may be reasonably associated with

limiting oxygen transfer current I_d, to much lower values of the current. This must be owing, as argued above, to the slow transport of oxygen through the separator.

The practical importance of the results shown in Fig. 4 resides in the fact that typical telephone float conditions (2.17 V) leave the cell precisely in the region where η_- is changing most rapidly. Similar measurements on conventional Pb/Ca cells show this to be the case for this system also. The consequences of this are clear. First, since the precise location of the steep part of the negative polarization curve is sensitive to subtle variations in cell construction, one would expect variations of cell voltages in a battery on float similar to that observed for Pb/Ca batteries. Secondly, small current changes during float will cause the negative plate potential to fluctuate moderately which would be reflected in cell voltage fluctuations. Thus, during periods of telephone plant instability, especially during peak load conditions, moderate fluctuations in cell voltages are to be expected. As we shall see later,
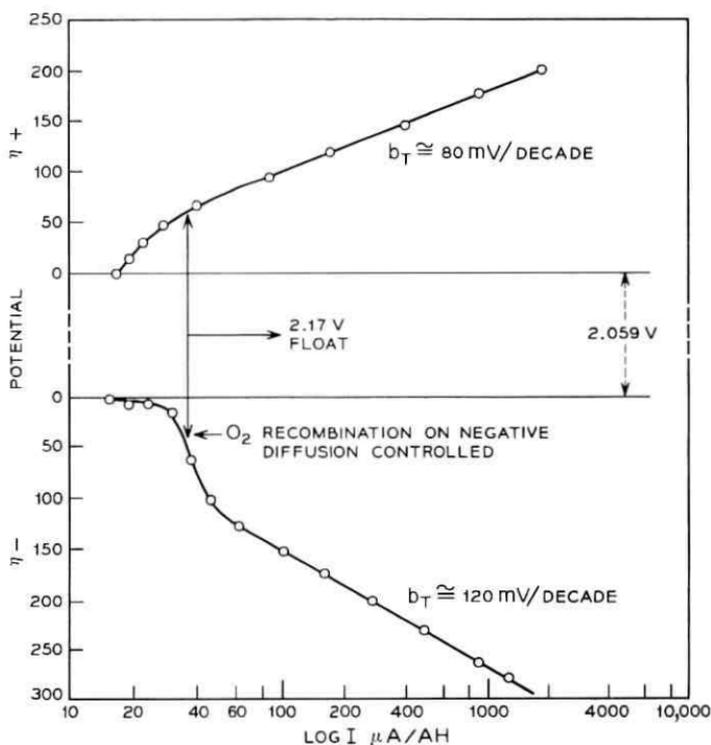

Fig. 4—Polarization data for the Bell System lead-acid cell design.

measurements on field trial cells show this to be the case. In addition, measurements on Pb/Ca cells in the field show the expected similar behavior.

The cell voltage variations and fluctuations observed with pure Pb and Pb/Ca cells are not a problem. They are a natural consequence of the electrochemical characteristics of the negative electrodes of these two battery systems. Figure 4 indicates, and actual measurements confirm, that the positive plate potentials are extremely stable.

The purpose of float operation is to have proper distribution of the total polarization between both the positive and negative electrodes. Close inspection of Fig. 4 shows that if the polarization characteristics of one electrode change seriously enough, one would have the undesirable situation of having the total cell polarization impressed on only one electrode with the other being at open circuit potential. The electrode polarization characteristics are sensitive to the presence of impurities and this is why particular attention has been given in the design of the Bell System battery to the electrochemical effects of trace contaminants.[9]

2.5 Recommendations for Float

It is generally believed that corrosion of the lead grid of the positive plate becomes very severe if the plate is held for any substantial period of time at an overpotential of less than 30 mV.[10] Thus, to ensure long battery life, the polarization on the electrode must be kept above 30 mV. The exact mechanism of this "catastrophic corrosion" is a subject of debate but our work has not progressed far enough to make quantitative statements. As corrosion of the positive grids occurs during the life of the cell on float, the positive plate polarization will show a gradual decrease due to the increasing surface area of the corroding electrode. Therefore, in order to ensure that the positive plate polarization does not fall below 30 mV during the life of the cell, it is essential that the cells have an initial polarization somewhat greater than 30 mV. A time zero positive plate polarization of 60 mV is quite adequate. At the negative electrode, all that is required is that the electrode not discharge. To ensure this, a polarization of only 5-10 mV is adequate in the absence of depolarizers.

For a number of years, Bell System practice has called for float conditions that result in a total overvoltage of 110 mV at 25°C. Examination of Fig. 4 shows that such operation would lead to values for η_+ and η_- shown in Table III. From Table III, it can be appreciated that float operation at 2.17 volts leaves both electrodes in a satisfactory

TABLE III— POSITIVE AND NEGATIVE PLATE POLARIZATIONS FOR
SPECIFIED FLOAT VOLTAGES, FOR THE CELL
CHARACTERIZED IN FIGURE 4

Float Voltage	η_+ mV	η_- mV
2.13	60	10
2.17	70	40
2.21	75	75

range of polarization. Floating at too high a voltage ($\gg$ 2.21) incurs
the penalties of excessive float current and increased corrosion at very
high overpotentials on the positive; floating at too low a voltage, on
the other hand, might incur the risk of discharging the negative.

III. CAPACITY CHARACTERISTICS

It is well known that conventional lead-acid batteries, when dis-
charged at a specified C rate, display a diffusion limited capacity[3] that
is less than the theoretical value based on the weight of the paste and
that also depends on the discharge rate. The dependence on discharge
rate is given empirically by Peukert's equation $I = at^n$; typically,
at a rate of C/5 (5-hour discharge), the capacity utilization of a con-
ventional thick plate (0.25 inch) battery runs around 30 percent. In
this Section, we report capacity figures for the new Bell System lead-
acid battery. It might be a subject for concern that the use of horizontal
rather than vertical grid stacking would lead to reduced capacity be-
cause of poor circulation of the electrolyte. However, this has been
overcome by the adoption of the conical plate design and by leaving
unpasted the grid holes immediately adjacent to the center hub on the
negative as shown in Fig. 5, thereby facilitating percolation of the fluid.
The result has been, as we shall show below, that the new cell has a
capacity equivalent to that of conventional cells.

3.1 *Factors Affecting Cell Capacity*

Our earliest cells had conical plates, with a 10° cone angle, but were
not adequately vented to allow the gases formed on charging the cells
to escape and thus also limiting free circulation of the electrolyte. To
see whether the horizontal stacking might be affecting cell capacity
in an unfavorable manner, a cell was held with its axis at some chosen
angle to the vertical and discharged at a fixed rate. The cell used was a
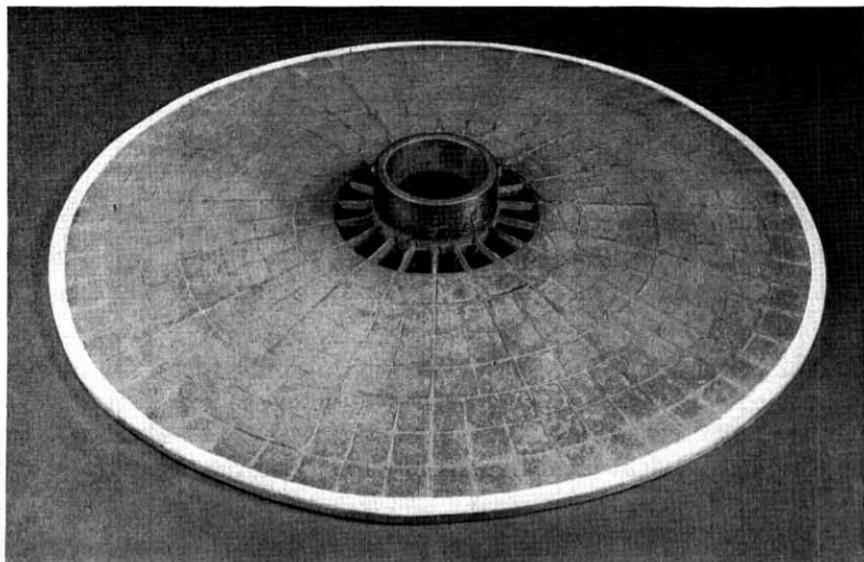
Fig. 5—Pasted negative plate showing vent area.

17-plate cell (8 positives and 9 negatives) using the design "B" positive plates and negative plates described in Table IV. Repeating this experiment at several angles to the vertical, the capacity could then be plotted as a function of orientation (Curve A, Fig. 6). The results clearly show that the capacity is indeed reduced, but rapidly recovers its normal value as the plates are made more vertical. As a result, venting holes were introduced by leaving the innermost ring of the

TABLE IV—PLATE COMPARISONS

| Positives	Design "C"	Commercial
Dimensions	10.9″ Dia. × .250″	9″ × 9.5″ × .250″
Grid Weight	3.2 lbs	3.0 lbs
Paste Vol.	14.0 in³	15.0 in³
Paste Weight	2.2 lbs	2.2 lbs
AH Capacity at 10.5 amps	63 ± 5	64 ± 5
Negatives	Bell System	Commercial
Dimensions	9.8″ Diam. × .187″	9″ × 9.5″ × .170″
Grid Weight	1.5 lbs	1.5 lbs
Paste Vol.	10.8 in³	11.85 in³
Paste Weight	1.9 lbs	2.0 lbs
AH Capacity at 10.5 amps	90 ± 5	90 ± 5
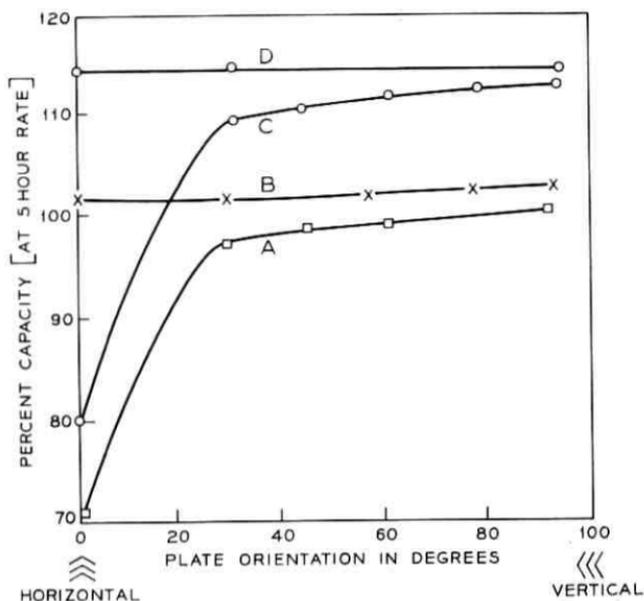
Fig. 6—Capacity vs orientation for the Bell System lead-acid cell design. Curve A, nonvented STD SEP; curve B, vented STD SEP; curve C, nonvented REV SEP; curve D, vented REV SEP.

negative grid holes unpasted (Fig. 5) and removing the innermost ring on the positive electrode. The capacity now no longer showed a significant increase when the cell axis was tilted away from vertical (curve B, Fig. 6). As an additional step, the separator and fiber glass mat were interchanged so that the glass mat was adjacent to the negative plate, thus reversing the configuration used in conventional batteries. This resulted in a uniform increase of around 10 percent in capacity, independent of orientation (Curves C and D, Fig. 6).

The venting problem has thus been solved, but it is worth noting

TABLE V—CELL CAPACITY AS A FUNCTION OF CELL ORIENTATION

Battery Orientation		% Nominal Capacity
Charge	Discharge	
vert.	vert.	114
vert.	hor.	105
hor.	vert.	77
hor.	hor.	76

here that in the process of solving it, clear indication was found that the problem is associated with failure to achieve complete charge. The charging was done at constant current and at least 120 percent of the discharge capacity was put back into the cells. A series of experiments was done with unvented cells held alternately vertical and horizontal on charge and alternately vertical and horizontal on discharge. Capacity figures obtained in this way, expressed as percent of nominal are shown in Table V.

3.2 Capacity Characteristics of the Bell System Cell

Figure 7 shows a discharge curve at the C/5 (84 A) rate for a 420 AH field trial cell along with a similar curve for a conventional cell of comparable size. This cell incorporates the design "C"[2] positive grids. The BTL and commercial cells have approximately equal weights of paste, and it can be seen that the discharge characteristics are very similar. Information on weights and volumes for the grids of the field trial cells is compared in Table V with that for a comparable commercial cell, along with the corresponding plate capacities. It is seen that the physical properties and performance of the design "C" positive plates are es-
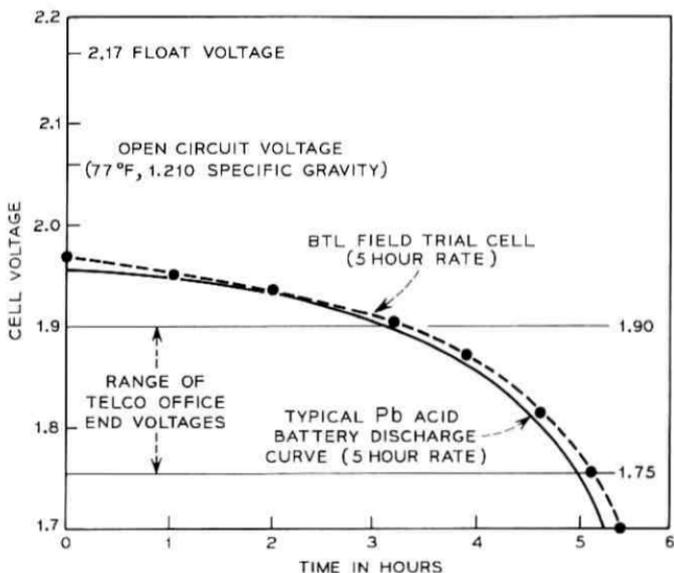

Fig. 7—Five hour rate discharge characteristics of a 420 AH Bell System cell compared to a typical conventional design.
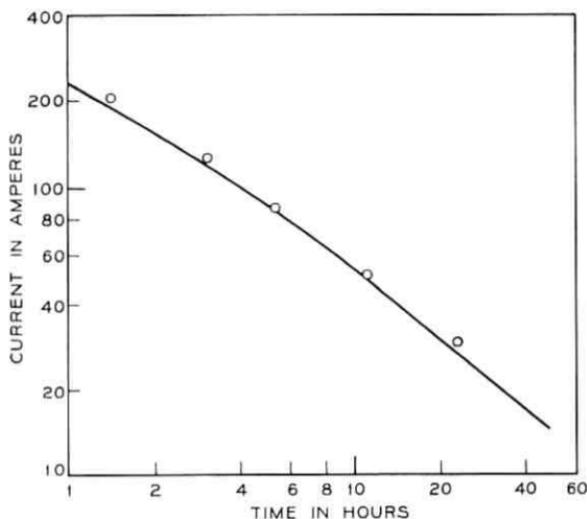
Fig. 8—Capacity vs discharge rate characteristics of field trial cell design compared to typical conventional design. —, performance of typical commercial telephone cells; O, field trial cell.

sentially the same as those for the commercial plate. The same is true for the Bell System and commercial negative plates.

The capacity of a typical field trial cell was also measured as a function of the discharge rate and compared to the performance of an equivalent commercial cell. In all cases recharging was done at the C/10 rate (42 A) until 110 percent of the discharge capacity was returned to the cell. The results are shown in Fig. 8 in which the logarithm of the current is plotted as a function of the logarithm of the time required for the cell voltage to fall to 1.75 volts. According to Peukert's equation, these points should fall on a straight line, the slope of which is related to the design of the cell (plate thickness, acid concentration, and so on). It is seen that, at all currents investigated, the performance of the circular cell is at least equivalent to the commercial cell. We conclude, therefore, that the circular design cell gives satisfactory capacity behavior and that no performance penalties are incurred with this radically new design.

IV. FIELD TRIALS

Field trials of the circular lead-acid cell design[5] were initiated in January 1969. The objectives were to determine any problems associated with shipping, racking, interconnection, and float behavior, and also to
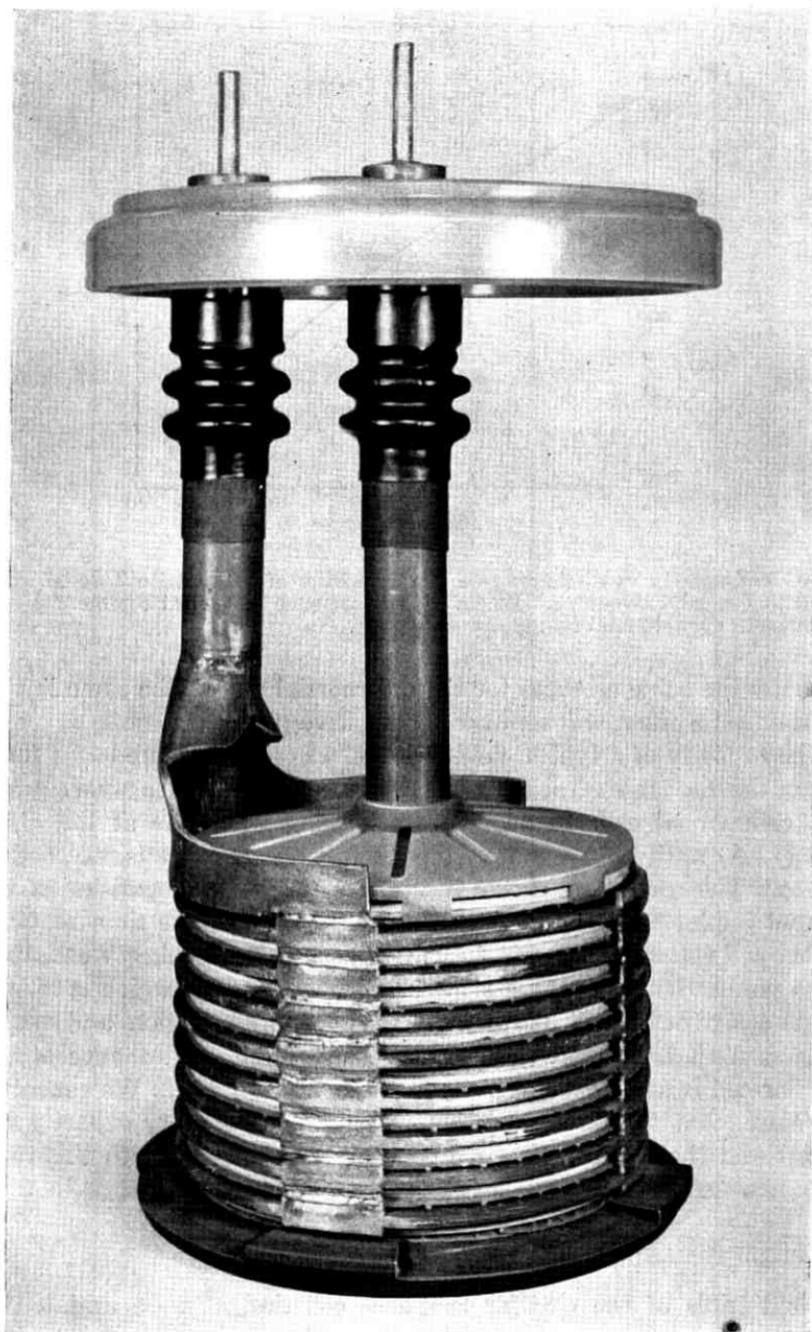
Fig. 9—Field trial cell element.

TABLE VI—FIELD TRIAL LOCATIONS

| Date	Location	Company	No. of Cells
1/30/69	Murray Hill, N. J.	N. J. Bell Tel. Co.	23–(48V)
2/7/69	Gouldsboro, Pa.	AT & T Long Lines	6–(12V)
3/21/69	Shirley Long Island, N. Y.	AT & T Long Lines	6–(12V)
4/11/69	Noyac Long Island, N. Y.	AT & T Long Lines	6–(12V)
9/17/69	Dover, Missouri	AT & T Long Lines	11–(24V) +(2 end cells)
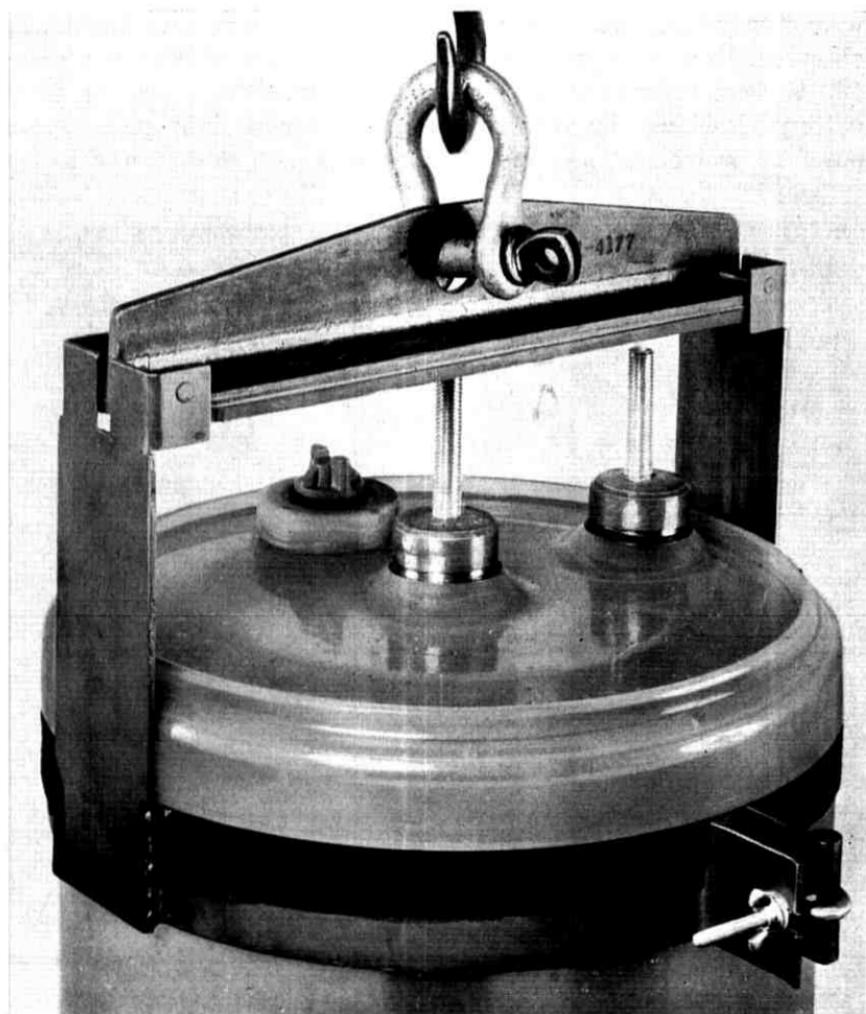
Fig. 10—Clamping fixture for lifting cell by its cover.

provide information concerning any necessary design modifications. The field trial cells (Fig. 9) were manufactured at our Design Capability Line.[5] These are 17-plate cells (8 positive and 9 negative plates) rated at 420 AH at the 5-hour discharge rate and incorporate the Design "C"[2] pure lead circular grid. The cells also incorporate the molded poly-vinylchloride jar and cover.[11]

There are five field trial locations (Table VI) with a total of 54 cells under investigation. All cells were shipped via common carrier in square wood crates which were modified with polyurethane pads to accommodate the circular design. Upon arrival at the field trial sites, all cells were closely examined for physical damage and in no case was damage observed. In a subsequent cross-country shipment, shocks in excess of 500 G were experienced by the crated cells without damage to the battery. Racking, interconnection and turnover into plant service posed no problems. The unpacking and racking operation was, in fact,
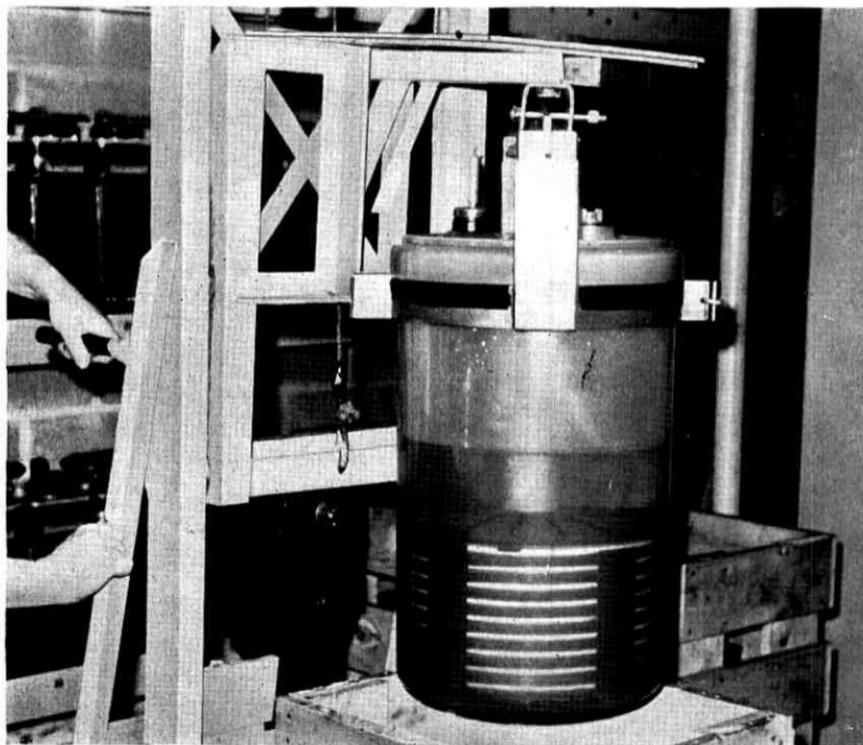

Fig. 11—Field trial cell being installed at Gouldsboro, Pa., TD-2.

Fig. 12—Trial installation at Murray Hill, New Jersey Bell Telephone Co. Central Office.

facilitated by the strength of the jar-to-cover seal[12] which permits the cell to be lifted by its cover. A clamping fixture (Fig. 10) fits under the lip of the cover while the cell is still in the crate. The cell can then be lifted by a hoist (Fig. 11) and moved into position on the rack where the cell is lowered and the clamp is easily removed.

Figure 12 shows the Murray Hill 23-cell (48 V) installation and Fig. 13 is a closeup view of a single cell in that installation. It can be seen that the jar is considerably oversized for the cell element. The reason for this is that the jar was designed to accommodate a 1680-AH cell and our limited production capabilities made it necessary to manufacture smaller cells (420 AH) in order to meet field trial commitments. It should also be noted that the round jar occupies the same rack space as a typical commercially supplied 1680-AH cell.

Lead-acid cells used in the Bell System must meet two major electrochemical requirements: (i) The cells, at the time of installation, must be capable of delivering 90 percent of rated capacity. (ii), At a battery
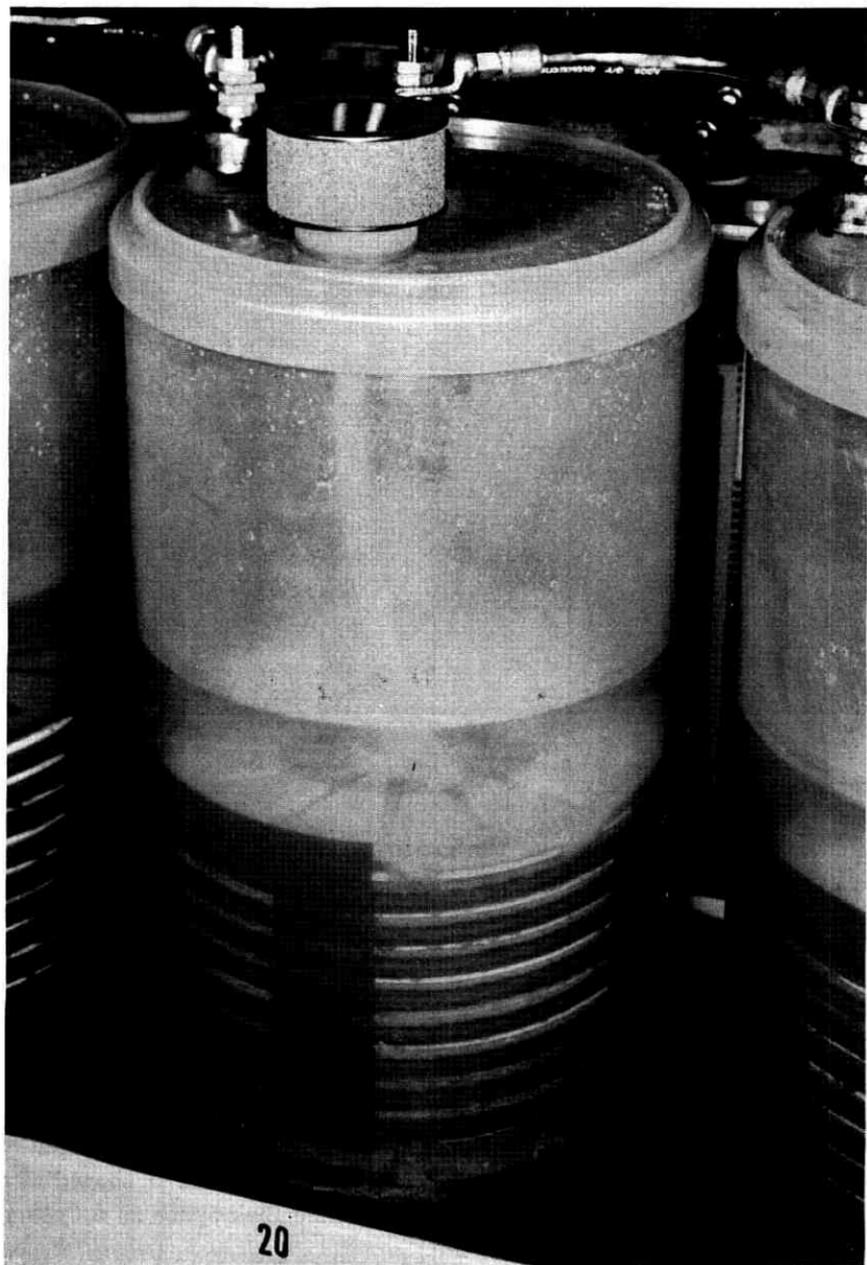
Fig. 13—Closeup of field trial cell.

float voltage of 2.16 to 2.18 volts per cell, all cells must be above 2.13 V minimum. The purpose of the 2.13 V minimum is to ensure adequate polarization to maintain cells in a full state of charge.

Capacity measurements prior to shipment showed that most cells had at least 100 percent capacity and that the 90 percent capacity requirement posed no problem. In addition, capacity measurements have been made on six cells subsequent to installation and these gave 116 to 130 percent of rated capacity.

With regard to the individual cell voltage distribution on float, Table VII shows that all cells meet the minimum 2.13 V requirement. These measurements are made periodically and the data in Table VI are the latest data available from each installation. Plate polarization measurements taken during the early stages of the field trials showed that positive plate polarizations of a few cells were much lower than desired. Specifically, one cell at Murray Hill, one at Gouldsboro and two at Shirley had positive polarizations very close to zero mV. This problem was diagnosed as resulting from cyclohexanone contamination.[9] This solvent was originally used to make jar-to-cover seals for the cells. For the four cells in question, the positive plate polarizations were brought to an acceptable value by boost charging. We feel that this treatment accelerated oxidation of cyclohexanone to electrochemically innocuous materials. Since this treatment, the previously contaminated cells have shown very satisfactory float behavior. Any future problems of this nature have been eliminated by the development of a heat-sealing technique for making jar-to-cover seals.[12] All cells in the Dover, Missouri, installation incorporate jar-to-cover seals made by this method.

Table VI also lists detailed positive and negative plate potential data. Average plate potentials are given for each installation and also the potential ranges. Fluctuations in plant voltage, especially during peak-load conditions, are naturally reflected in fluctuations of individual cell voltages. It was observed, during the course of these measurements, that fluctuations in cell voltage resulted in nearly equivalent fluctuations in negative plate potential, whereas positive plate potentials remained constant. This behavior is expected since Fig. 4 shows that under float conditions the negative plate potential is in a region where small changes in current will cause large fluctuations in that potential; the higher capacitance of the positive, of course, is also a factor.

The data in Table VI show very satisfactory polarizations for both positive and negative plates. Correcting the open circuit positive plate potentials for temperature variations, the voltage extremes of 1.150

TABLE VII—FIELD TRIAL FLOAT DATA

Installation	No. of Cells	Temp. (°F)	Cell		Positive Plate*		Negative Plate*	
			Avg. Voltage	Range	Avg. Voltage	Range	Avg. Voltage	Range
Murray Hill	23	75	2.161	2.131 2.174	1.166	1.156 1.182	-0.995	-0.976 -1.006
Gouldsboro	6	63	2.180	2.159 2.212	1.191	1.182 1.196	-0.999	-0.972 -1.019
Shirley	6	65	2.171	2.159 2.187	1.178	1.166 1.186	-0.992	-0.982 -1.005
Noyac	6	69	2.183	2.168 2.200	1.180	1.179 1.181	-1.004	-0.989 -1.021
Dover	11	72	2.160	2.140 2.170	1.158	1.150 1.164	-1.001	-0.980 -1.011

* Measured vs a Hg/Hg_2SO_4 reference electrode.

and 1.196 correspond to polarizations of 60 and 105 mV respectively. The 60-105 mV polarization range is quite adequate for maintaining positive plates in a healthy state. Similarly the voltage extremes of -0.972 and -1.021 for the negative plate potentials correspond to polarizations of 5 and 53 mV respectively. This polarization range is satisfactory for maintaining negative plates in a fully charged condition.

V. SUMMARY AND CONCLUSIONS

Polarization measurements of individual positive and negative pure-lead circular plates have led to the following significant observations: (i) The differences in current drawn at a given overpotential for pasted plates as compared to unpasted grids are surprisingly small in view of their large differences in surface areas. This difference is much more pronounced for positive plates than for negative plates. (ii) The oxygen diffusion current (I_d) of the negative electrode is very sensitive to the design of the experimental setup. (iii) The Tafel parameters for pure lead are, as expected, very similar to those of Pb/Ca.

These polarization data for grids and pasted plates may be scaled directly to predict battery performance when one takes the oxygen diffusion characteristics of the separator into consideration. Polarization data on completed cells of the circular design show very satisfactory polarizations of both electrodes, adequate for long-term use in telephone float service.

It has been shown that the discharge performance of the circular cell is quite satisfactory. With proper venting, the capacity of the circular cell is at least equivalent to its commercial counterpart.

Field trials, in progress at five locations, show the circular design lead-acid cell to conform to the electrochemical requirements of the Bell System. Capacity data for these cells in the field show better than nominal performance. The voltage distributions of cells in batteries are satisfactory and individual electrode polarization data show these to be adequate.

REFERENCES

1. Milner, P. C., "Float Behavior of the Lead Acid Battery System," B.S.T.J., this issue, pp. 1321–1334.
2. Cannone, A. G., Biagetti, R. V., and Feder, D. O., "Positive Grid Design Principles," B.S.T.J., this issue, pp. 1279–1303.
3. Biagetti, R. V., and Weeks, M. C., "Tetrabasic Lead Sulfate as a Paste Material for Positive Plates," B.S.T.J., this issue, pp. 1305–1319.
4. Amlie, R. F., Ockerman, J. B., and Rüetschi, P., "Absorption of Hydrogen and Oxygen on Electrode Surfaces," J. Electrochem. Soc., 108, No. 4(April 1961), pp. 377–383.

5. Koontz, D. E., Feder, D. O., Babusci, L. D., and Luer, H. J., "Reserve Batteries for Bell System Use: Design of the New Cell," B.S.T.J., this issue, pp. 1253–1278.
6. Ness, P., "Gegenwörtiger Stand Der Forschungsarbeiten an Positiven Elektroden in Bleikakkumulatoren," Electrochim. Acta, 12, No. 2(February 1967), pp. 161–178.
7. Baker, R. A. Extended Abstracts, Battery Division, Electrochemical Society, Houston Meeting, Oct. 1960, p. 53.
8. Bockris, J. O'M., and Srinivasan, S., Fuel Cells: Their Electrochemistry, New York: McGraw Hill, 1969 p. 438.
9. Butherus, A. D., Lindenberger, W. S., and Vaccaro, F. J., "Electrochemical Compatibility of Plastics," B.S.T.J., this issue, pp. 1377–1392.
10. Lander, J. J., "Anodic Corrosion of Lead in H_2SO_4 Solutions," J. Electrochem. Soc., 98, No. 6(June 1961), pp. 213–219.
11. Huseby, T. W., Ryan, J. T., and Hubbauer, P., "PVC Battery Jars and Covers," B.S.T.J., this issue, pp. 1359–1376.
12. Dahringer, D. W., and Shroff, J. R., "Jar-Cover Seals," B.S.T.J., this issue, pp. 1393–1404.

Lead-Acid Battery:

Polyvinyl Chloride Battery Jars and Covers

By T. W. HUSEBY, J. T. RYAN and P. HUBBAUER

(Manuscript received January 14, 1970)

A detailed materials engineering evaluation of a low-cost, high impact, structurally modified rigid polyvinyl chloride (PVC) has demonstrated that battery jars and covers can be molded from PVC in conventional equipment. The battery jars are flame resistant, transparent, and sufficiently tough to withstand considerable abuse without cracking. Although mechanical properties indicate the material becomes slightly more brittle after exposure to 1.220 specific gravity H_2SO_4 at 120°F, the behavior at all times exceeds the currently used styrene-acrylonitrile in toughness. PVC maintains its rigidity beyond 120°F and accelerated tests indicate it will creep less than 1 percent after 20 years at room temperature under all expected loads. Jars weighing 12 pounds and covers weighing 4.5 pounds were each molded in carefully designed molds in a five minute cycle at 400°F without difficulty. The molded parts show very little stress concentration even without expensive annealing treatments and are free of molding defects. The cover diameter is molded 0.005 in. larger than the jar to insure proper fit during assembly. This dimensional accuracy is maintained without requiring special cooling fixtures after molding to restrict uneven shrinkage.

I. INTRODUCTION AND DESIGN CRITERIA

Industrial battery jars historically have been manufactured from either hard rubber or relatively brittle thermoplastics such as styrene and styrene-acrylonitrile copolymer. Each of these materials has a high tensile strength in excess of 10,000 psi. However, they all crack and fracture at very low strains and, although such plastics have a high ultimate strength, they can absorb little energy during impact. The toughness, or work required to bring a material to failure, is related to the area under the stress-strain curve as depicted in Fig. 1. The tough
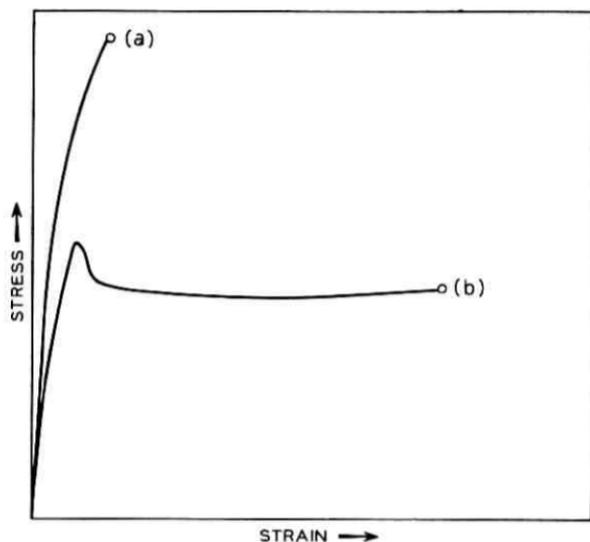
Fig. 1—Typical stress-strain curve for (a) brittle plastic, and (b) tough, ductile plastic.

material illustrated by curve (b) has the ability to yield and dissipate energy prior to breaking. Thus, even after yielding, this material is capable of supporting a load although the deformation may continue to increase. This toughness or impact resistance is gained at the expense of some reduction in rigidity and in failure stress. The rigidity, or stiffness, of a material is related to its modulus of elasticity. The loss in modulus is often due to the introduction of a toughening agent or impact modifier such as a soft rubber well dispersed throughout the hard, brittle matrix. In selecting a plastic for the new Bell System battery jars, emphasis was placed on toughness and impact strength as one of the most desirable and important material improvements.

Clarity for inspection and service convenience is another desirable feature in a battery jar. This is best achieved in heterogeneous, rubber-modified, tough plastics by matching the refractive indices of the two phases. However, since a perfect match is seldom attainable, especially over a sufficiently wide temperature range, it is also beneficial if the size of the dispersed phase remains less than about 0.1 μm in diameter.

In addition, for long-term service, a suitable battery material must have a certain chemical resistance. For example, the material must be free of cracking and mechanical property deterioration in the presence of H_2SO_4 and various other chemicals which might be present in central

offices. It is also important that the material not yield substances that adversely affect the electrochemical performance of the battery. This important property is fully discussed in an accompanying article by A. D. Butherus and W. S. Lindenberger.

Because cells may also be subjected to loads at temperatures as high as 120°F, the long-term creep resistance should also be considered. This is especially important since inclusion of soft impact modifiers into hard plastics increases their tendency to creep.

To reduce the hazard of central office fires, it is desirable for the battery plastic to be flame retardant. Several test methods are available which allow comparisons of this important property to be made. One such method measures the minimum amount of oxygen necessary in a controlled environment to support combustion.[1] Thus, plastics which require an oxygen-rich environment during combustion will be flame retardant in air containing 21 percent oxygen. A comparison is given in Table I of the minimum oxygen concentration necessary to support combustion of most of the common thermoplastics.

Finally, the material must be fabricated by conventional injection molding technology. The molded jars should also not require any annealing treatments to minimize orientation stresses.

II. PROPERTIES OF POLYVINYL CHLORIDE

2.1 General Description

Rigid polyvinyl chloride (PVC) is a commercial, low-cost plastic having a high degree of resistance to solvent crazing and chemical attack. As shown in Table I, no other conventional plastic can match its flame retardancy. Indeed, PVC will support combustion only in environments

TABLE I—OXYGEN INDEX RATING OF COMMON THERMOPLASTICS

Polymer	Oxygen Index (%)
Polyoxymethylene	16
Polyolefins, polystyrene, impact styrene, cellulosics, ABS	17–20
Styrene-PPO copolymers	21–24
Polycarbonate, polyamide	25–28
Polysulfone, PPO, flame retardant polycarbonate copolymers	30–32
Polyvinyl chloride	40
*Polyimide	51
*PTFE	95

* Cannot be processed by injection molding.

containing more than 40 percent oxygen, nearly twice as much oxygen as exists in air. Since it is relatively unstable at high temperatures, its use in injection molding, especially of large or intricate items, has been severely restricted. However, PVC can be stabilized and modified to be processed successfully under many molding conditions. These modified compounds make PVC a material suitable for use in large battery jars.

PVC resin is a high molecular weight polymer containing carbon, hydrogen, and chlorine, whose chemical structure is represented in Fig. 2a. It is transparent and remains very rigid up to temperatures of 160°F. Its glass transition temperature, at 175°F, is closely associated with this loss in rigidity since the modulus decreases rapidly in this temperature range. Its melting temperature is slightly above 400°F. The crystalline fraction is not large, even under the most favorable conditions, owing to polymerization conditions which limit the regularity of the molecular structure. Thus, while the amount of crystallinity may total 10 percent in commerical polyvinyl chloride, especially if samples are carefully annealed, the mechanical behavior of injection molded PVC remains similar to the behavior of amorphous polymers.

2.2 *Thermal Stability*

PVC resin is unstable at high temperatures and degrades prior to melting. To minimize degradation, it must be stabilized and processed below its melting point at temperatures where its viscosity is quite high. This accounts for the difficulty in molding parts from PVC. When the plastic degrades it gives off corrosive hydrogen chloride vapors and forms a discolored, brittle residue. Dehydrochlorination leads to formation of long segments of unsaturated carbon atoms which are responsible for the color change during degradation.[2] This degradation can be minimized by the addition of a small amount of heat stabilizer, often a combination of organic tin compounds.

A significantly more stable polymer can be made by copolymerizing

$$\left(-CH_2 - \underset{\underset{Cl}{|}}{CH} - \right)_n \qquad \left(-CH_2 - \underset{\underset{Cl}{|}}{CH} - CH_2 - \underset{\underset{CH_3}{|}}{CH} - \right)_n$$

(a)   (b)
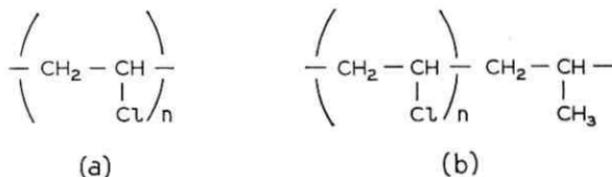
Fig. 2—Structural formulae for polyvinyl chloride and vinyl chloride-propylene copolymer.

a small amount of propylene with the vinyl monomer.[3] The resultant vinyl chloride-propylene copolymer (Fig. 2b) contains about 5 weight percent propylene or about one unit for every 15 vinyl chloride units. This structural modification enhances stability by effectively retarding the evolution of hydrogen chloride. The enhanced stability can be demonstrated in several ways. The Brabender method[4] serves to indicate thermal stability under laboratory conditions approximating the molding process where energy is added both as heat and work. A comparison between the vinyl chloride-propylene copolymer and a PVC of equivalent molecular weight is shown in Table II. The processing time relates

TABLE II—THERMAL STABILITY OF PVC BY BRABENDER METHOD*

Material	Processing Time (Minutes)
Polyvinyl chloride	$5\frac{1}{4}$
Vinyl chloride-propylene copolymer	11

* Manufacturing Standard 17000, Section 1171, Method A.

to the time at 375°F required to cause serious degradation in the sample while it undergoes mechanical working. Since these samples have identical heat stabilizers added in the same amounts, the longer time indicates that the copolymer is more stable to heat and work than the homopolymer.

2.3 Processability

Introduction of propylene does not significantly change the processability of PVC. This can be observed from measurements of viscosity at a temperature typical of injection molding. Figure 3 shows viscosity measurements (η) at 392°F over a range of shear rates ($\dot{\gamma}$).

Because polymer melts are viscoelastic, their flow properties change with shear rate. During injection molding, plastic melts are subject to high shear rates usually exceeding 1000 seconds^{-1}. Most plastics are difficult to mold if the viscosity at high shear rates exceeds about 8000 poise; easily processed materials have high shear-rate viscosities of approximately 1000 poise. Figure 3 shows that the copolymer has nearly the same viscosity as an equivalent molecular weight homopolymer over a wide range of shear rates.

Without an impact modifier the PVC battery jar probably could not reliably withstand the shock loading associated with transportation or an earthquake environment. Therefore, a rather large amount of rubbery
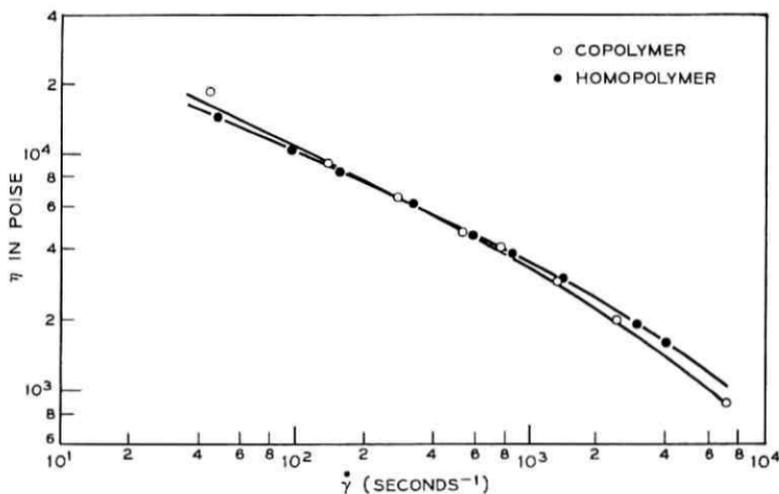
Fig. 3—Melt viscosity versus shear rate for polyvinyl chloride and vinyl chloride-propylene copolymer at 392°F. Weight average molecular weight of both polymers is 80,000.

plastic is used to toughen the PVC compound. The addition of rubber particles to PVC changes its processability. Recent studies of stress relaxation in ABS,[5,6] another impact modified plastic, have shown that above the glass transition temperature of the hard matrix a major effect of the dispersed elastomeric phase is to increase the molecular entanglement. Since entanglement interactions strongly influence the flow behavior, the addition of rubber particles is expected to increase the viscosity. Indeed, an increased viscosity has been observed previously in ABS and in impact modified PVC.[7] Figure 4 illustrates the increase in viscosity due to impact modifier in this compound. Notice, however, that the increase is not so much as to make processing impossible.

Proper dispersion of rubber particles throughout the matrix is a vital factor in developing high impact strength. Figure 5 is a transmission electron micrograph obtained from a sample of the impact modified vinyl chloride-propylene copolymer. Phase contrast is improved by exposing the sample to osmium tetroxide which preferentially stains the rubber particles and thereby increases the electron density.[8,9] The electron staining is also an oxidative process which stiffens the soft rubber particles and aids in the preparation of ultra-thin specimens. The rubber particles in Fig. 5 are largely spherical, well dispersed, and about 0.1 μm in diameter.

To maintain transparency, the rubber particles must have a refractive
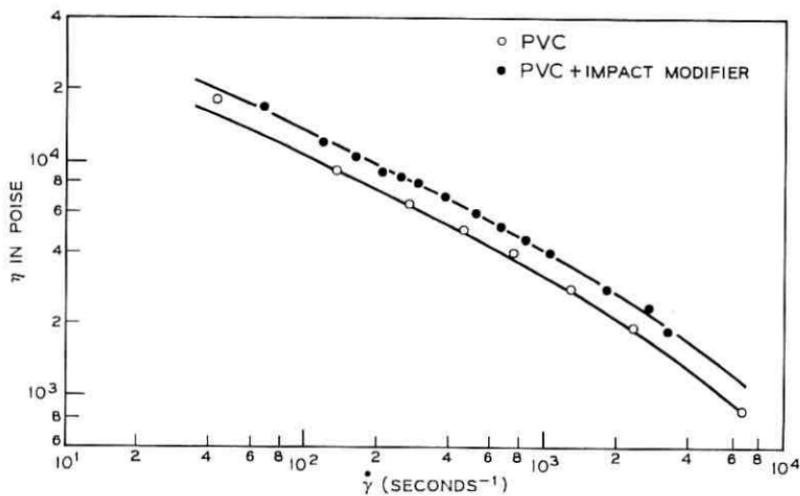
Fig. 4—Melt viscosity versus shear rate for copolymer and impact-modified copolymer at 392°F.
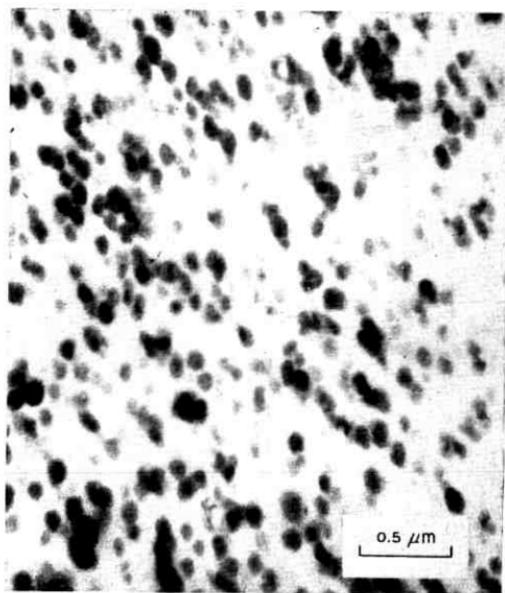
Fig. 5—Transmission electron micrograph of impact-modified copolymer.

index close to that of PVC, 1.537 at 25°C. An easily dispersed material having a refractive index of 1.536 at 25°C is methyl-methacrylate-butadiene-styrene polymer (MBS).[10] In such compounds, methyl-methacrylate and styrene polymers are grafted to styrene-butadiene rubber to improve compatibility with PVC. These impact modifiers usually have a butadiene content approximating 50 percent. Most clear, impact-modified PVC compounds contain MBS as the impact modifier. Since the impact modifier is not flame retardant and it increases viscosity and raises cost, it is necessary to balance these considerations against impact strength.

Processability of PVC is also improved by the addition of a wax. This lubricates the molding equipment during processing and enhances processing characteristics.

2.4 *Mechanical Properties*

The properties of the compound developed for cell jars reflect a combination of properties of the described ingredients. Some commonly measured engineering properties appear in Table III. Especially evident is the improvement in impact strength at the expense of modulus and ultimate strength.

2.5 *Effect of Acid Exposure on Modulus and Strength Properties*

Table IV shows values for some of the mechanical properties evaluated

TABLE III—ENGINEERING PROPERTIES OF IMPACT MODIFIED
POLYVINYL CHLORIDE

		Value	
Property	ASTM Test Method	With Impact Modifier	Without Impact Modifier
Tensile strength (psi)	D-638 Type I	7000	8800
Elongation at break (%)	D-638 Type I	110	36
Impact strength (ft–lb/in notch)	D-256 Method A	28	0.7
Flexural strength (psi)	D-790	9600	13,100
Flexural modulus (psi)	D-790	355,000	416,000
Deformation under load (%) 24 hours at 2000 psi, 50°C	D-621	3.45	1.66
Heat deflection temperature (°K) 264 ps	D-648	144	145
Oxygen index	MS-17000 Section 1226	33	40

TABLE IV—ENGINEERING PROPERTIES OF IMPACT MODIFIED PVC*
AFTER EXPOSURE TO 1.220 SPECIFIC GRAVITY
H_2SO_4 AT 120°F

	Initial	3 Months	6 Months	16 Months
Tensile Strength (psi)	7500	8300	8300	8350
Elongation (%)	50	25	25	18
Modulus (psi)	350,000	350,000	350,000	350,000

* Specimens machined from a battery jar.

at room temperature before and after exposure to battery acid at 120°F for up to sixteen months. Note that the elongation measurements indicate slight embrittlement after exposure to heat and acid. Separate experiments have determined that most of the change is due to heat rather than acid. The tensile strength, for example, increases but levels off after three months. Even after the combination of acid exposure and heat aging, the PVC remains tougher and less brittle than styrene-acrylonitrile is before acid exposure. At no time has the modulus shown any decrease from its initial value. These data show that PVC surpasses in mechanical toughness the plastic currently used for batteries, styrene-acrylonitrile.

2.6 *Effect of Temperature on Mechanical Stiffness*

Because normal battery service temperatures extend up to 120°F, it is of interest to examine modulus through this temperature range. The dynamic mechanical behavior can be used conveniently to measure the temperature-dependent stiffness properties. This nondestructive test measures the steady-state stress response to a sinusoidal deformation. The response can be resolved into two components, one acting in phase with the deformation and one acting in quadrature. That portion of stress in phase with deformation determines the elastic modulus, E', while the quadrature component gauges the damping modulus, E''. The ratio, E''/E', is defined as tan δ. Figures 6 and 7 show the temperature dependence of E' and tan δ, respectively. This information reveals that the modulus of impact modified PVC remains in excess of 250,000 psi up to the maximum operating condition of 120°F. It should be emphasized that this test measures short-time, temperature-dependent behavior, the time scale being inversely proportional to frequency. This behavior is not quantitatively characteristic of longer time response because the mechanical behavior is time-dependent. It is therefore
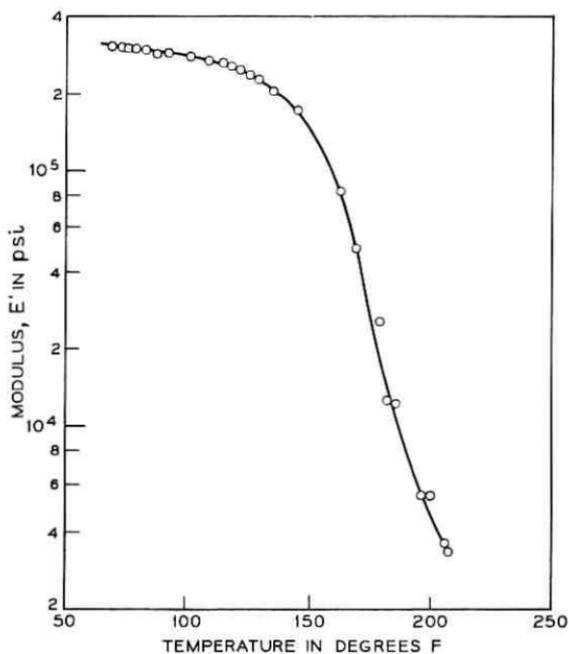
Fig. 6—Dynamic modulus E' versus temperature at 110 Hz for impact-modified polyvinyl chloride.
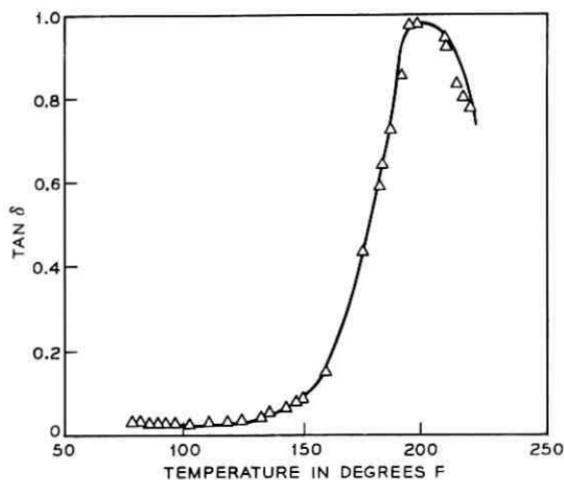

Fig. 7—Loss tangent versus temperature at 110 Hz for impact-modified polyvinyl chloride.

important to inquire into the time-dependent isothermal creep behavior.

2.7 *Long-Term Creep Behavior*

Plastics are known to deform and creep under constant load due to their time-dependent viscoelastic behavior. A characteristic strain-time curve of an amorphous plastic at constant load is shown in Fig. 8. With certain restrictions, the influence of increased temperature is to shift the creep response uniformly along the logarithmic time scale toward shorter times. Thus, the creep response of a plastic at an elevated temperature appears the same in logarithmic time as at a lower temperature, except for a shift in time scale. This relation between temperature and logarithmic time is exploited during accelerated creep testing. Although initially the relationship was exhaustively tested at low loads for amorphous polymers above their glass transition temperatures, it has recently been applied, with good success, to rigid polymers in their glassy state and with higher loads.[11,12]

Figure 9 shows the creep compliance, $J(t) = \epsilon(t)/\sigma$, where $\epsilon(t)$ is the time varying strain and σ is the constant stress, at three temperatures. Note that the compliance appears independent of load at the two lowest temperatures. The master curve in Fig. 10 is obtained by shifting the data at higher temperatures uniformly along the logarithmic time axis to form a continuous curve at 78°F. The displacement in units of logarithmic time is denoted as log a_T, the logarithmic shift factor. The shift in time scale between 78°F and 122°F corresponds to an acceleration by a factor of 300.

Figure 11 shows the creep strain at 122°F under a load of 500 psi,
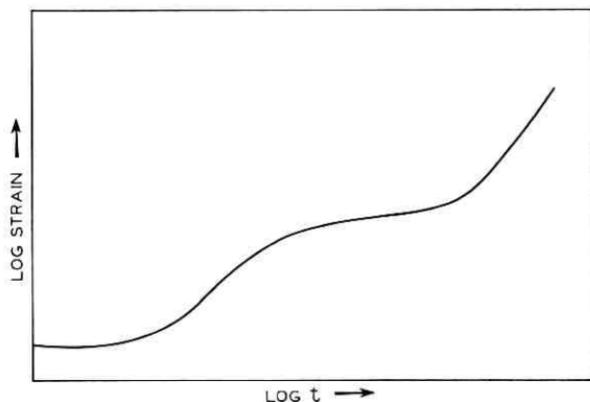
Fig. 8—Typical creep curve for amorphous thermoplastics.
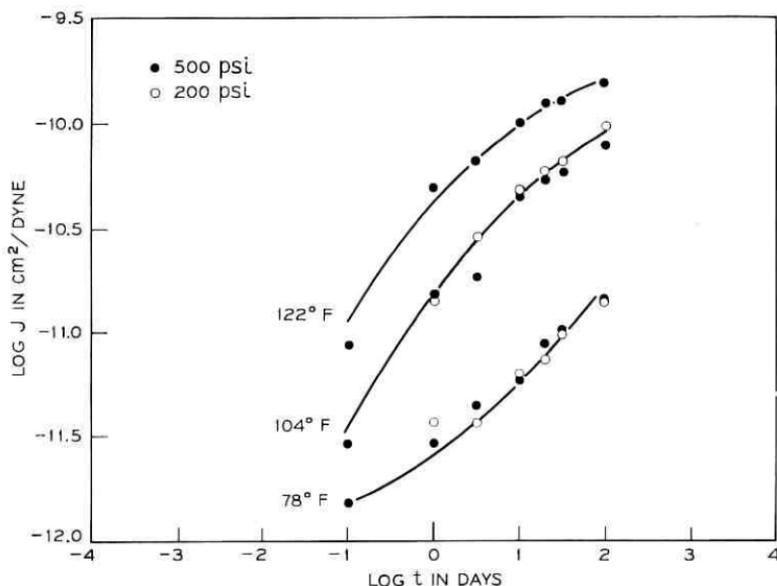
Fig. 9—Creep compliance versus time for impact-modified copolymer.

which exceeds the stress likely to occur in practice. The strain after 24 days at 122°F corresponds to a 20-year creep test at 78°F. Note that this acceleration factor pertains only to time-dependent mechanical properties and not to property changes due to aging, chemical change, deterioration in the presence of acid, or to any electrochemical effects. Figure 11 indicates that the creep response to loads of 500 psi or lower for long periods of time is less than 1 percent at 78°F under the most extreme loading conditions. This is a negligible amount. Indeed, even with continuous exposure to 122°F the estimated 20-year creep remains below 5 percent for loads up to 500 psi.

III. MOLDING TECHNOLOGY

Certain aspects of the molding technology required to produce large molded parts from PVC are of interest. In order to assure ourselves that molding PVC in large battery jars was feasible, it first was decided to mold PVC in existing molds of rectangular shape. With the cooperation of one of the battery manufacturers, the material supplier, and the molder, List 508 battery jars were molded from PVC. The molding weighed 17 pounds and is still the largest item ever molded from PVC.

It is significant that the molder, who had no previous experience with PVC, had no difficulties in molding this large jar.

The cylindrical jar, weighing 12 pounds and having a 1/4 inch wall thickness, is not quite so heavy because it is thinner than the square design. It was molded by a second molder without previous experience with PVC on a 1600-ton reciprocating screw injection molding machine (Fig. 12), one commonly found in many large custom molding shops. The parts are molded on a five minute cycle at 400°F under 17,000-psi injection pressure. The covers weighing 4.5 pounds were molded in a smaller machine, also in a five minute cycle at 400°F under 7500-psi.

The molded parts (Fig. 13) are free of bubbles, burn marks, imperfections, and obvious weld lines. The cylindrical geometry permits the necessary dimensional accuracy to be achieved without using special cooling fixtures after molding to restrict uneven shrinkage. These fixtures are necessary to prevent warping in square geometries. The
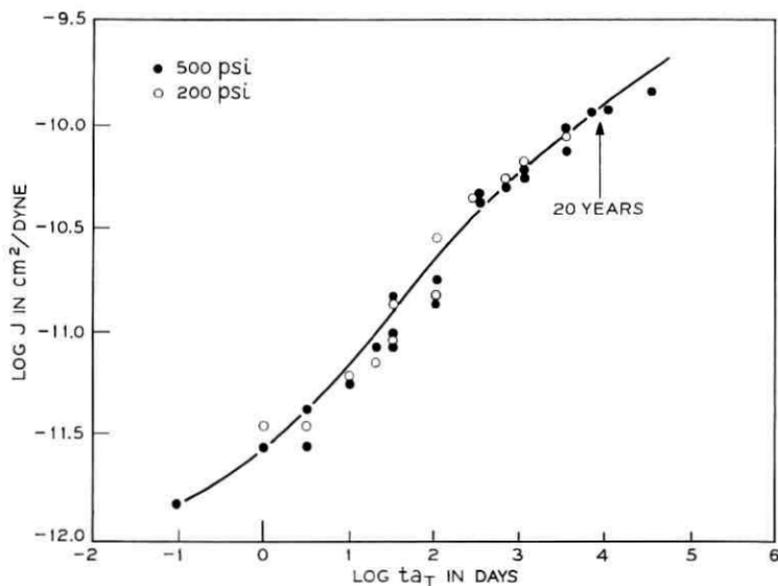

Fig. 10—Master curve of creep compliance versus reduced time for impact-modified copolymer at 78°F.

| T(°F)	Log a_T
78	
104	−1.5
122	−2.5
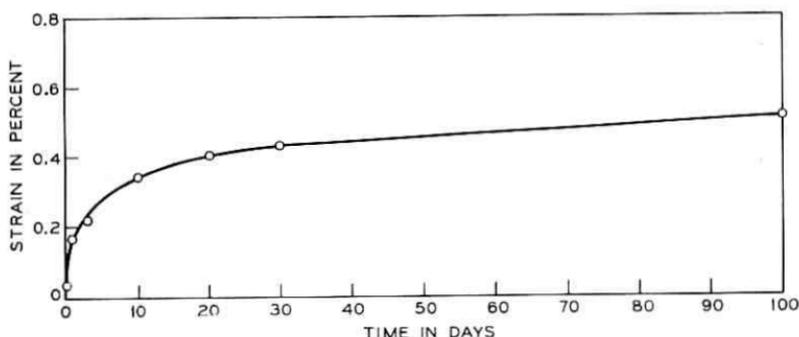
Fig. 11—Creep at 122°F and 500 psi of impact-modified copolymer.

parts are also satisfactorily free of stress concentration, even without expensive annealing treatments, as was demonstrated by examination through polarizing films. Use of two films as a polariscope identifies regions of varying orientation which correspond to areas of high residual stress in the molded part. It is noteworthy that, in this case also, the molder has never encountered difficulties in either starting up, running for extended times, or shutting down his machine.

These extremely high quality moldings were achieved by careful design of both cover and jar molds. Several mold design details which contributed to the successful molding can be pointed out in Fig. 14, a schematic of the jar and cover. Both cover and jar molds are made from prehardened tool steel to eliminate the necessity for hardening and subsequent grinding. The jar mold is equipped with an auxiliary, hydraulically-operated ejector ring on the moving die. This allows the molder to eject on the top surface of the jar and push the jar off the mold. This feature costs very little but allows him to mold jars with only 1/2° draft on each side. A small draft pays off in less wasted space when mounted in service. Hand-operated air valves on the molds for the insides of the jar and cover prevent a vacuum seal from hindering ejection from the mold. Air valves on the molds for the outsides help push the parts onto the moving sides as the molds open.

Cold slug wells are opposite the center gate in both cover and jar mold. This provides a trap for the first amount of cold plastic from the machine. The cover is center gated across the center post hole and machined out later. This gate arrangement insures uniform fill and minimizes stress concentration due to residual orientation. Effort was made to provide maximum cooling in all parts of the mold through special use of additional cold water passages. This is essential for
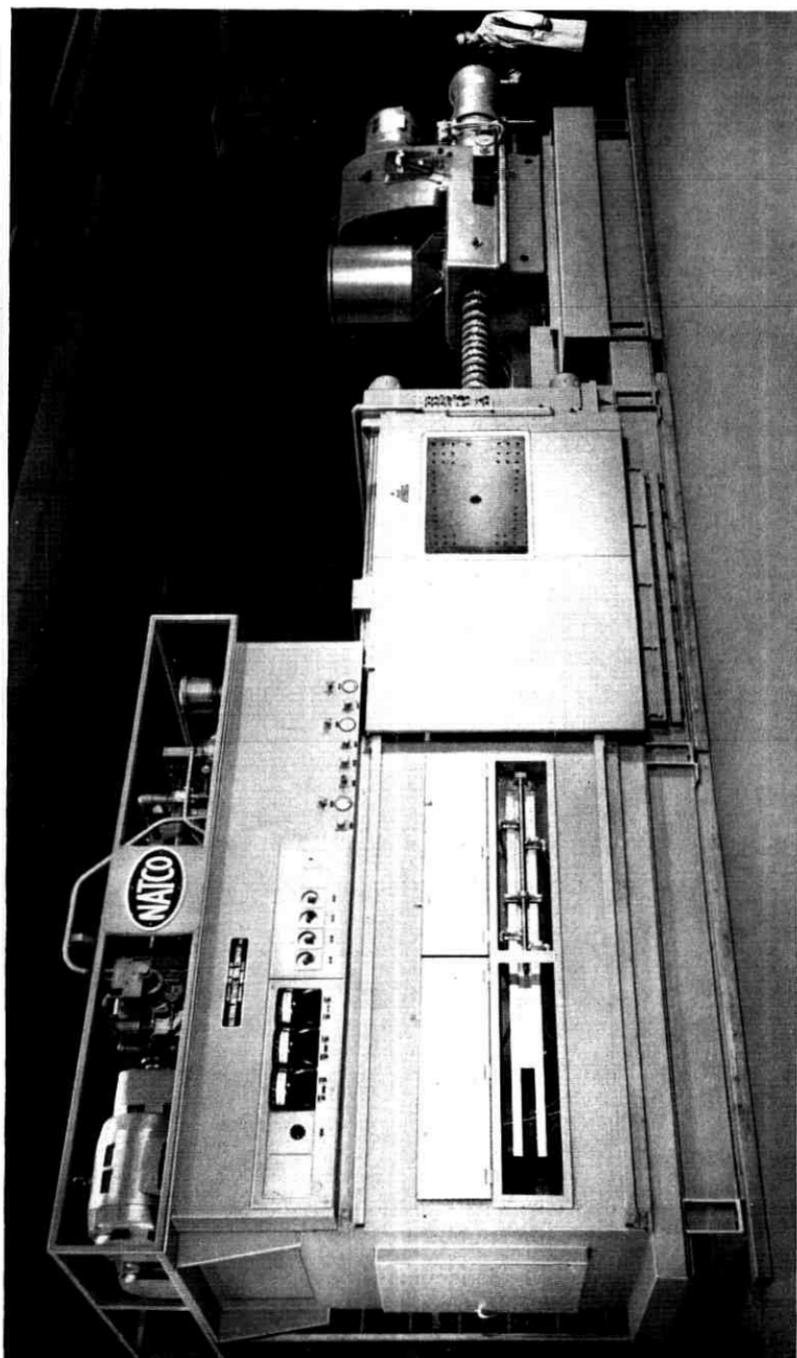
Fig. 12—Injection-molding machine.
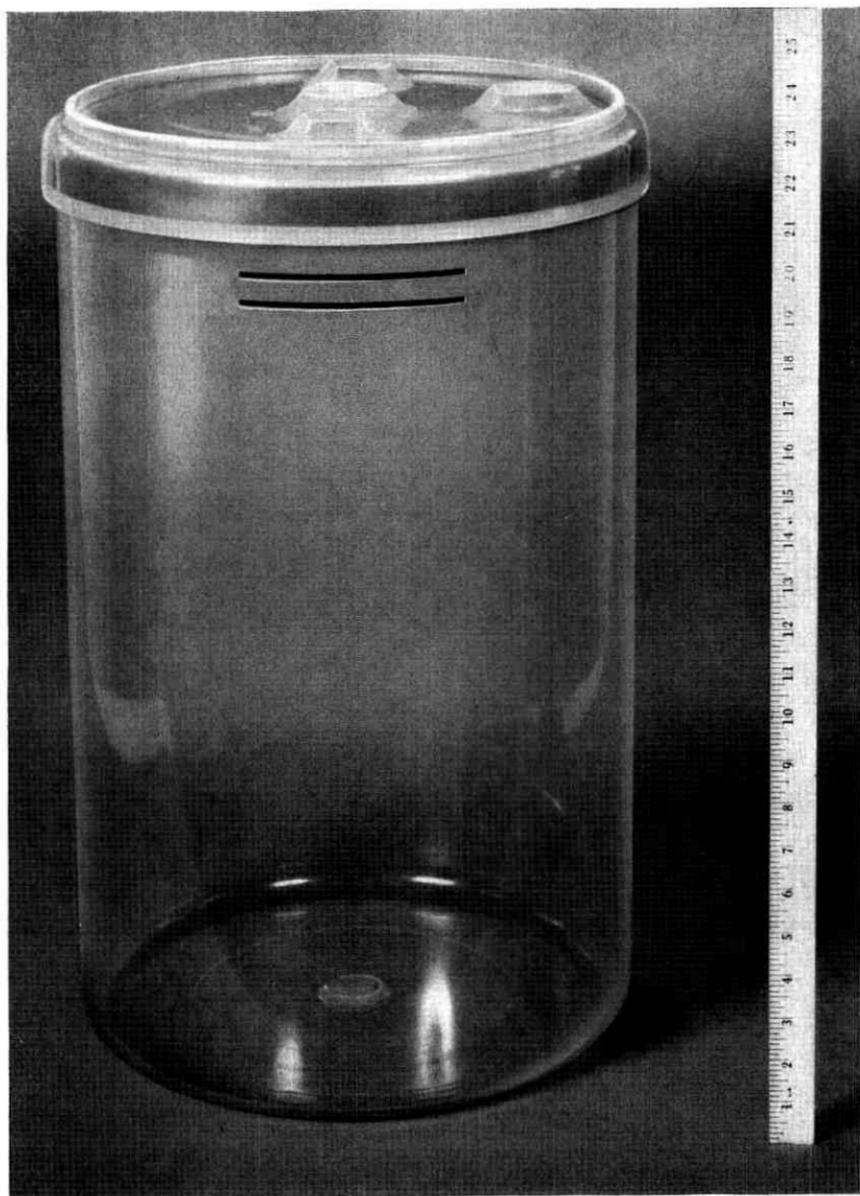
Fig. 13—Injection-molded vinyl chloride-propylene copolymer battery jar and cover.

DIAPHRAGM
CENTER GATE

LOCATION OF
AIR VALVE

LARGE RADII

MINIMUM
DRAFT INSIDE

5/6 INCH WALL
THICKNESS

1/4 INCH WALL
THICKNESS

1/2° DRAFT
PER SIDE

EXTRA FINE
SURFACE FINISH

CENTER GATE WITH
COLD SLUG WELL AND
LOCATING REGISTER

LOCATION OF
AIR VALVES
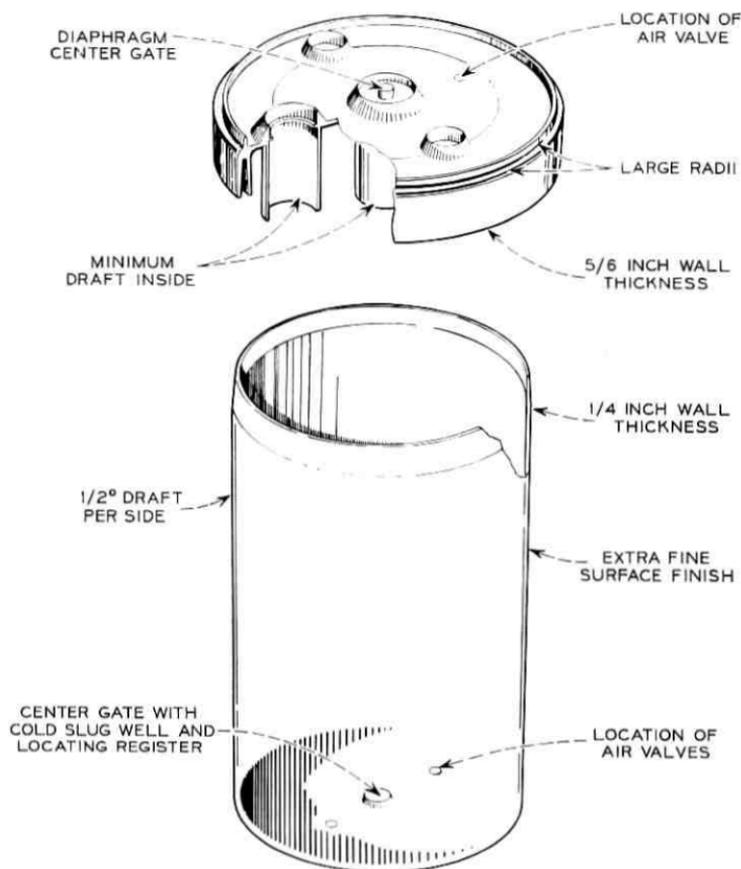
Fig. 14—Illustration of battery jar and cover.

adequate cooling of large parts having uniformly thick sections of 1/4 inch or more. A mold finish of about two microinches was achieved by hand polishing following the final grinding. This extra fine finish was required for satisfactory transparency.

IV. SUMMARY

In conclusion, rigid polyvinyl chloride is a versatile design material exhibiting desirable mechanical properties, flame retardancy, and chemical resistance. Although not as easily processed as many plastics, it can be molded without difficulties using good molding practice. A detailed materials engineering development program has demon-

strated that properly compounded and stabilized PVC can be molded in conventional equipment under controlled manufacturing conditions using a properly designed mold to produce battery jars and covers of high reliability.

V. ACKNOWLEDGMENTS

REFERENCES

1. Fenimore, C. P., and Martin, F. J., "Candle-Type Test for Flammability of Polymers," Modern Plastics, 44, No. 2 (Nov. 1966), pp. 141, 142, 146, 148, 192.
2. Salovey, R., Luongo, J. P., and Yager, W. A., "On the Irradiation of Poly (Vinyl Chloride)," Macromolecules, 2, No. 2 (March-April 1969), pp. 198–200.
3. Cantow, M. J. R., Cline, C. W., Heiberger, C. A., Huibers, D. Th. A., and Phillips, R., "Vinyl Chloride/Propylene Copolymers," Modern Plastics, 46, No. 6 (June 1969), pp. 126, 127, 132, 136 and 138.
4. Bell System Manufacturing Standard 17000, Section 1171, "Torque Rheometer Test for Plastics," Method A.
5. Scalco, E., Huseby, T. W., and Blyler, L. L., "Some Viscoelastic Properties of ABS Polymer," J. Applied Polymer Sci., 12, No. 6 (June 1968), pp. 1343–1353.
6. Bergen, R. L., and Morris, H. L., "The Melt Rheology of ABS Polymers," Proceedings of the Fifth International Congress on Rheology, Vol. 4, Tokyo: University of Tokyo Press, 1969.
7. Malpass, V. E., "Flux and Melt Flow Behavior of Rigid ABS Modified PVC," SPE 27th Annual Technical Conference Preprints, 15 (1969), pp. 55–60.
8. Matsuo, M., Nozaki, C., and Jyo, Y., "Fine Structures and Physical Properties of Plastic/Rubber Two-Phase Polymer Systems," J. Electron Microscopy, 17, No. 1 (1968), pp. 7–15.
9. Kato, K., "Morphological Aspects of ABS (Acrylonitrile-Butadiene-Styrene) Plastics," Japan Plastics, 2, No. 2 (April 1968), pp. 6–17.
10. Ohtsuka, S., Watanabe, H., and Amagi, Y., "New PVC Impact Modifier," SPE 25th Annual Technical Conference Preprints, 12 (1967), pp. 707–716.
11. Malpass, V. E., "Prediction of Long-Term Acrylonitrile-Butadiene-Styrene Relaxation Behavior," J. Applied Polymer Sci., 12, No. 4 (April 1968), pp. 771–788.
12. Rusch, K. C., "The Relaxational Behavior of Heterogeneous Polymers," J. Macromol. Sci., Phys., B2, No. 3 (1968), pp. 419–445.

Lead-Acid Battery:

Electrochemical Compatibility of Plastics

By A. D. BUTHERUS, W. S. LINDENBERGER
and F. J. VACCARO

Because of the float use-mode of reserve batteries in the Bell System, the cell behavior is quite sensitive to small amounts of electroactive organic substances in the cell electrolyte. Premature cell failure has in the past resulted from the introduction of a new plastic without testing its compatibility with the cell electrochemistry. We describe an accelerated test method by which the extraction rate of organic residues from the plastic container into the cell electrolyte is predicted and the electrochemical effects of these residues is determined. These data predict that in normal service no adverse electrochemical effects from interaction of the electrolyte with the jar or cover will occur in the cylindrical lead-acid cell for a minimum of 50 years.

I. INTRODUCTION

1.1 Materials Selection

The general problem of materials selection for battery use is divided into two areas: (*i*) the test material must meet some specified physical property requirements and maintain them over a period of long exposure to battery electrolyte (usually a strong corrosive acid or base solution), and (*ii*) the material must not introduce any solute into the electrolyte which will upset the electrochemical behavior of the cell.

During the early history of lead-acid battery development, the cell was placed in a container made of a material which experience had shown resistant to the sulfuric acid. Since a wide range of polymeric materials was unavailable, glass and lead-lined containers predominated. During the 1930s, hard rubber containers for lead-acid cells were introduced. There was no compatibility testing of this material but, in retrospect, it can be seen that none was needed since:

(*i*) cell lifetimes were short due to failure mechanisms which were faster operating than cell contamination by the container material;

(*ii*) cells were generally cycled, which tended to decompose organic contaminants due to the high charge potentials;

(*iii*) a lead-antimony alloy was used in the grids and the antimony-caused negative overvoltage lowering was so large that it swamped the much smaller increases caused by organic contaminants. These cells required much higher float currents due to this shift and lead-calcium cells which float at much lower currents are preferred for Bell System use.

These characteristics masked the slow-appearing effects caused by organic contamination. With the advent of extremely long-lived, low float current lead-acid cells, organic contamination from container materials can measurably upset cell behavior.

1.2 *Material Testing*

In recent years various plastics have been used in these applications.[1] Styrene, styrene copolymers and rubber-filled styrene plastics have gained the broadest acceptance.[2] With the advent of large new families of plastics, testing for suitable mechanical properties and monitoring their degradation in the particular electrolyte have become commonplace.[3] The effects of the electrolyte on the properties of the plastic can be measured more rapidly. By increasing the test temperature or increasing the severity of the test conditions in other ways (mechanical stress, radiation, and so on).

The second area of material compatibility, which is the effect of the plastic on the battery, has not to our knowledge been rigorously investigated. Weight-change studies are routinely used to determine the amount of electrolyte attack on a polymeric material[4]; that is, if the loss of the plastic is small over some period of time, the material is assumed chemically compatible with the cell. The effect of the small fraction of "lost" material on the cell electrochemistry has not generally been investigated. For typical cyclic battery service in which the cell is routinely charged and discharged, the effects of these organic contaminants are usually small enough to be innocuous, since the cycling of the cell either promotes the oxidation of the contaminants or covers them up in recharge cycles. In the Bell System, however, batteries which serve as a back-up or reserve power supply are floated, that is, maintained in a fully charged state by a potential of 2.17 volts per cell.

In such a use mode, the effects from electrolyte contamination by organic substances can be very harmful to cell life and performance.

1.3 *Example of Electrochemical Incompatibility*

A history of incompatibility problems has existed in the Bell System. Some standard lead-acid cells with porous polyvinyl chloride (PVC) separators were placed in Bell System use between 1954 and 1965. These separators were cheaper and less brittle than the normally used microporous hard rubber units. All of these cells showed needlelike dendritic lead deposits on the negative plates after a few years of service. These dendrites shorted and discharged the cells, behavior which has since been duplicated in this laboratory in similar cells in which the processes were accelerated by elevated temperatures. The cause of this dendritic growth was traced to the PVC separators by the procedure of leaching separator samples with cell electrolyte and introducing this solution into stably-floating cells, which promptly grew needles. Since solutions leached from pure PVC polymer (resin without additives) show no electrochemical effects; the effects observed must have been due to the residual organic solvents and wetting agents used in fabrication of the separators. It was found that the balance of the polarization (discussed subsequently) was markedly disturbed in the doped cells. No further investigations were undertaken and the new separator material was withdrawn from Bell System use.

Two conclusions are drawn from this experience:

(*i*) New materials for fabrication of battery components must be verified as electrochemically compatible with the cell under usemode conditions.

(*ii*) A test procedure must be devised in which the compatibility of new materials can be determined in a relatively short time.

Due to the undesirable properties inherent in the presently used styreneacrylonitrile plastic,[5] a new, rigid PVC copolymer which shows a superior combination of physical properties has been recommended for the jar and cover of the new cylindrical cell. Because of the previous experience with PVC plastic, compatibility testing was mandatory.

There is a need for a test procedure which accelerates the plastic-cell electrochemistry interaction such that a valid choice of materials can be made in a relatively short time (< 12 months). The work reported here is an example of a general test method which we believe to have

broad application in material selection for use in batteries or other electrochemical devices.

II. THE FLOATING CELL

2.1 Description of Float

A battery used as a reserve power source in the Bell System is "floated" when not delivering emergency electrical power; that is, the cells are maintained at full charge by overcharging the cell at a constant voltage of 2.17 V., which is 110 mV above the thermodynamic reversible potential of the cell (Fig. 1). This results in operation in regions where the potentials on the positive and negative plates are largely fixed by the Tafel behavior of the electrodes (the "Tafel lines" on Fig. 1).[6]

2.2 Grid Corrosion and Plate Polarization Balance

It has been shown[7] that grid corrosion in the positive plate is minimized at potentials of 20–30 mV above the reversible potential of the $PbO_2/PbSO_4$ electrode and, to provide a sufficient margin of safety, up to 80 of the 110 mV excess potential should be on the positive plates of the cell, with the remaining 30 mV on the negative plates (Fig. 1).
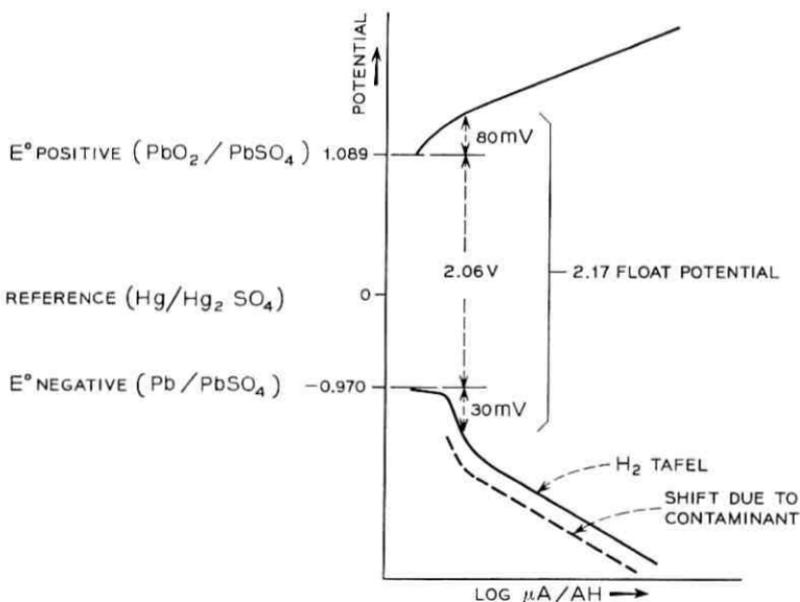

Fig. 1—Float properties of BTL cell, showing effect of contaminants.

The division of this excess potential is determined in large part by the slopes and intercepts of the hydrogen and oxygen Tafel lines[8] and a shift of one or both curves will change the fraction of the excess potential found on each plate. Since these cells must float at the lowest practicable float currents, the cells exist in a potential region where the negative polarization curve has an anomalously steep slope due to oxygen reduction.[8] Because of this, changes in the overpotentials of the cell will have significant effects in the float balance, and an increase in the hydrogen overpotential and/or a decrease in the oxygen overpotential will decrease the fraction of available excess potential on the positive plates of the cell. As the positive polarization drops significantly below 30 mV, the corrosion rate of the grid, which is the failure mechanism of the lead-acid cell on float,[9] increases rapidly, shortening the life expectancy. If the shift were sufficient to place the entire 110 mV on the negative plates, the positive electrodes would remain at their reversible thermodynamic potential and would self-discharge.

In addition to causing overpotential shifts, the organic residues can change the growth habit of the negative electrodes, again affecting cell performance. This can be beneficial in some cases, and organic "expanders" are used to produce a spongy, high surface area, efficient lead anode.[10] The effect can, however, be destructive as in the case of the dentritic structure caused by the PVC separators as described in Section 1.3.

III. ORGANIC CONTAMINANTS AND ELECTRODE BEHAVIOR

A very large number of organic solutes affect electrochemical processes. The role of surface-active organic substances in modifying electrode reactions was noted as far back as 1923[11]; extensive literature has been developed on this subject, though predominantly limited to mercury or platinum electrodes. The usual result of even minute amounts (a few ppm) of electroactive organic substances in an electrolyte is a marked shift in Tafel behavior, especially for hydrogen evolution. A shift of tens of millivolts on a 1 cm² electrode is typically caused by less than 10 ppm of an electro-active contaminant.[12] In general, electrode reactions are quite complex and a rigorous description of the mechanisms and identification of the reaction products and intermediate species is always difficult and often impossible. The inclusion of organic substances at the reaction interface complicates matters further; we therefore will not discuss the mechanisms of the organic molecule-electrode interaction but refer the reader to some of the extant literature.[13-18] It is

sufficient to note that organic contaminants in small amounts can affect Tafel behavior markedly and thus may present a problem for long-lived batteries.

IV. INITIAL TESTING

4.1 *General Description*

Most plastics consist of a mixture of various components. In addition to the polymer resin, stabilizers, antioxidants, lubricants, waxes, impact modifiers (usually another polymer species) and colorants are typically found in plastic formulations.[5] These complicate investigations, for any one or all of these components may introduce sufficient electroactive substances into the electrolyte to upset the float balance. The magnitude of the problem may be seen by noting that a plastic is usually considered resistant to chemical reagents (ASTM D 543) if the percent weight change in the given reagent is 0.1 percent or less over a four-week period under specified conditions and the weight change with time is described by a relatively flat curve.[3] That such a test is inadequate for the present problem is shown by the fact that a weight loss maximum of 0.1 percent from the cylindrical battery jar (normalized to the same test conditions as ASTM D 543) would introduce about 100 ppm organic contaminants into the cell electrolyte. By calculation, this amount of foreign material would cover the actual surface area of the positive and negative electrodes (determined by B.E.T. methods)[19] to a depth more than three monolayers.

To determine if the selected plastic might cause electrochemical problems, equal amounts of thinly-shaved samples of the test plastic,[5] a propylene-PVC copolymer, and the styrene acrylonitrile plastic now in use were subjected to standard sulfuric acid electrolyte (1.210 sp. gr.) for 28 days at 60 C in the apparatus shown in Fig. 2 (this apparatus was used for all subsequent extractions). 60°C was chosen as a temperature high enough to accelerate the chemical processes without getting too close to the glass transition temperature (T_g) of the plastic. (The reasons for this temperature limit are discussed in item *ii* of Section 5.1.)

Since a measurement of the effect of the test plastic relative to the presently used material was desired rather than an absolute value in this preliminary test, the electrochemical effects were grossly exaggerated by the use of a lead electrode of very small area in a rotating disk configuration.[20] This apparatus was used to measure the hydrogen and oxygen overpotential shifts in the test solutions compared to a pure acid solution. The test plastic caused a shift in the hydrogen over-
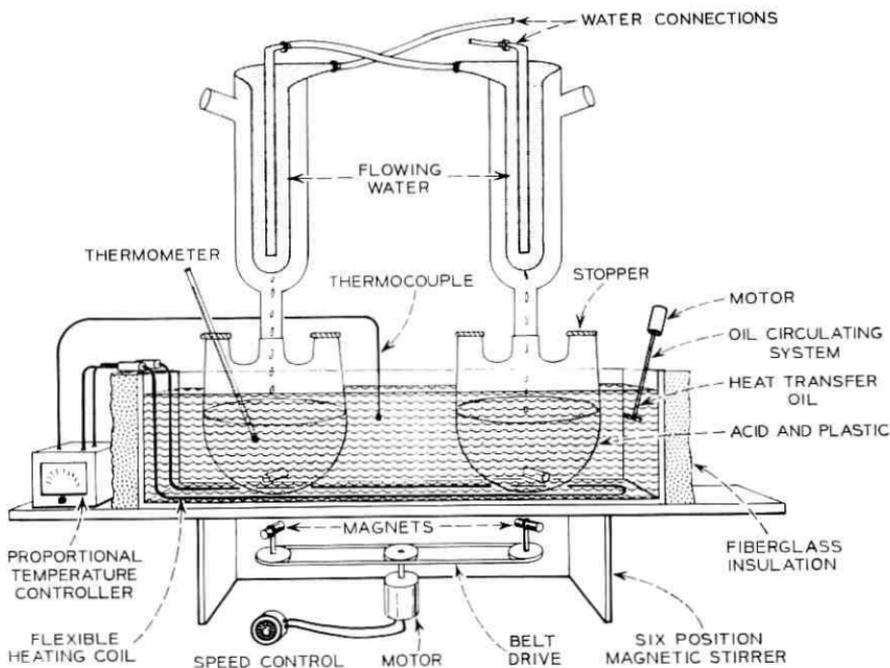
Fig. 2—Extraction apparatus.

potential of about 200 mV while the previously used styrene acrylonitrile caused a 50 mV shift. The oxygen overpotential was not measurably affected, and both solutions caused a slight increase in the lead corrosion current. Thus a source of possible cell deterioration was indicated.

4.2 Relative Effects of Plastic Components

To measure the relative effect of each formulation component independently, samples of the pure components (supplied by the manufacturer) were extracted separately at 60°C with the same volume of 1.210 sp. gr. sulfuric acid as used in the previous test. The amount of each component extracted was equal to the amount of that component found in the thinly shaved test plastic extracted in the previous experiment. As shown in Table I,[21] the individual overpotential shifts are apparently not additive (the mechanism may be competitive—the organic molecule which adheres most tightly to the lead excludes the other species from the main reaction mechanism). It is evident that the organo-tin stabilizer dissolution products produced the major electrochemical effects; in fact, the effects caused by the pure stabilizer were both qualitatively and quantitatively (on a relative scale) identical to

TABLE I[21]—ROTATING DISK ELECTRODE STUDY OF
PLASTICS AND COMPONENTS

Sample	Effects
PVC Co-Polymer	~ -200 mV H_2 overpot. shift slight increase in corrosion current
Styrene Acrylonitrile	~ -50 mV H_2 overpot. shift
PVC Base Resin	No effect
Stabilizer	~ -200 mV H_2 overpot. shift slight increase in corrosion current
Lubricant	~ -100 mV H_2 overpot. shift
Impact Modifier	~ -50 mV H_2 overpot. shift

the effects due to the entire plastic formulation. Therefore, the appearance of electroactive dissolution fragments of the plastic in the electrolyte was monitored by the measurement of the concentration of stabilizer fragments in the sulfuric acid.

This assumption that the electrochemical effect of the plastic leachants was directly related to the stabilizer decomposition was supported by obtaining samples of a plastic compound whose formulation was identical to the test material except for the substitution of a lead-based stabilizer for the normal organo-tin compound. A sample of this experimental plastic and a sample of the lead stabilizer were extracted independently as before, but the extracts both showed identically negligible electrochemical effects. This supports the assumption that the effects produced by the normal test material are effectively caused entirely by the organo-tin stabilizer.*

V. ACCELERATED TEST PROCEDURE

5.1 *Test Assumptions*

The test method evolved from the following facts and assumptions:

(*i*) The stabilizer fragment concentration in the acid will increase with time at a given temperature, probably by diffusion through

---

* This experimental lead-stabilized plastic, though electrochemically innocuous, is unsuitable for fabrication of battery jars since it is opaque and is also almost impossible to injection-mold. Small (0.19 g) plaques were molded only with great difficulty.

the plastic to the electrolyte interface. This process can be accelerated by increasing the test temperature. An apparent activation energy can be measured for the process.

(ii) So long as the test temperature does not closely approach the glass transition temperature of the plastic, the extraction processes occur by mechanisms identical with those operative at room temperature, which permits valid extrapolation of the accelerated reaction rates to room temperature.

(iii) The stabilizer was shown to be completely digested by the 1.210 sp. gr. sulfuric acid and the concentration of tin ions in the electrolyte is thus a precise measure of the amount of stabilizer decomposition.

(iv) It was shown that the appearance rate of tin ions is directly proportional to the ratio of the exposed plastic surface area per unit sulfuric acid volume used in the extraction. The leaching can be accelerated by increasing this ratio (Fig. 3). Other geometric effects were compensated for by using plastic samples of the same physical dimensions for all tests.

5.2 *Test Procedure*

In the test procedure, equivalent quantities of plastic (with equal surface areas) were extracted into equal volumes of standard 1.210 sp. gr. sulfuric acid at 40°, 50°, 60°, and 70°C for up to several weeks. (The time was dependent on the extraction temperature.) Small samples of
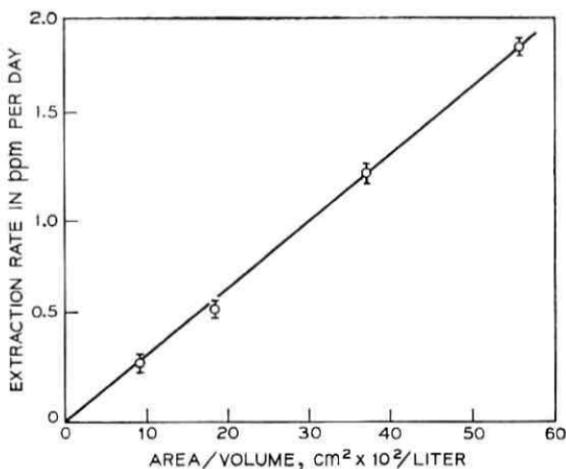

Fig. 3—Extraction rate dependence on plastic surface area/acid volume at 80°C.

acid were removed periodically and analyzed for tin ion.[22] The samples removed were a small enough fraction of the total acid volume so that the change in the plastic surface area/acid volume ratio was small enough that the change in reaction rate which resulted was within the error limits of the measurements.

5.3 Test Results

Resulting concentration data are plotted vs time (Fig. 4) and show linear behavior over the measured temperature range.

A high plastic surface area/acid volume ratio was used to speed up the extraction, and linear extrapolation out to the actual plastic surface area/acid volume ratio present in the BTL cell produced the upper set of time coordinates. A rate equation of the form

$$\frac{d[Sn]}{dt} = k$$
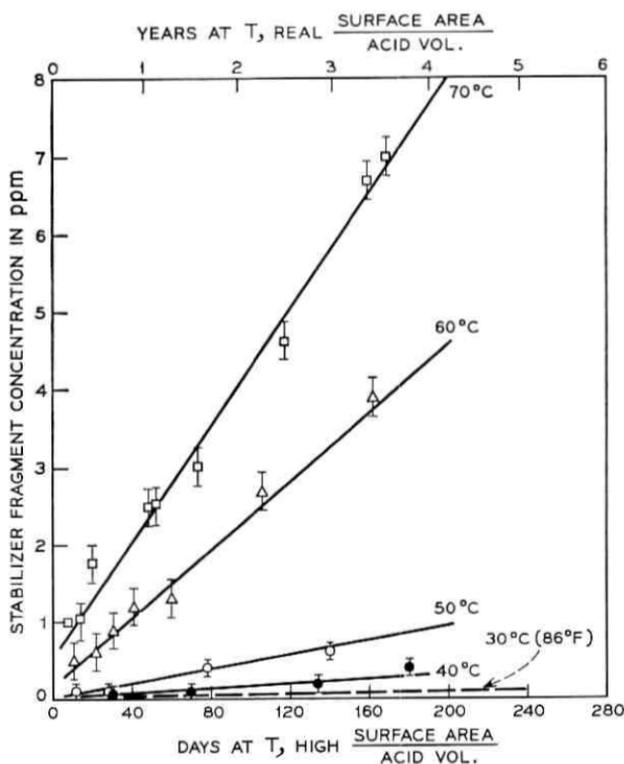
Fig. 4—Extraction rates of stabilizer.
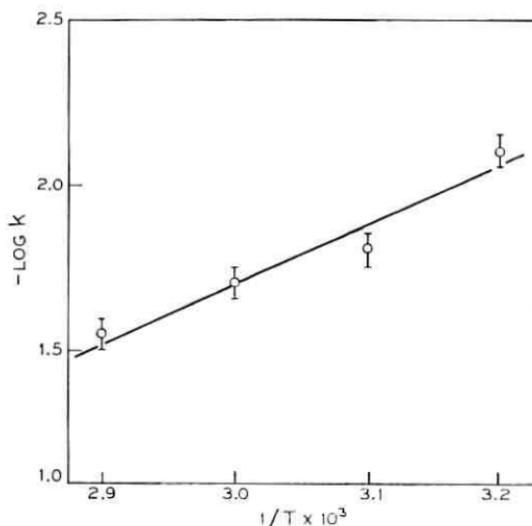
Fig. 5—Arrhenius plot.

was assumed and treatment of these data by the Arrhenius method (Fig. 5) produced an apparent activation energy

$$E_{act} = 10 \pm 1 \text{ kcal/mole.}$$

The use of this value produces the extrapolated room temperature stabilizer decomposition rate described by the dashed line in Fig. 4.

5.4 *Application to Real Cells*

To determine the relationship between contaminant concentration and electrochemical effect in real cells, stably-floating cylindrical cells were doped with extract aliquots of a known stabilizer fragment concentration and the cell characteristics (electrode potentials, float currents and cell potential) were monitored. Four to six weeks were needed after each extract addition for the cell to restabilize. From these data a plot of hydrogen overpotential shift vs stabilizer fragment concentration in the cell was constructed. (Fig. 6)

An elevated temperature accelerated test was run on cells with large amounts of the test plastic immersed in the electrolyte. The conditions duplicated the testing of the experimental microporous PVC separators in which dendritic lead structures caused premature cell-death. No abnormal lead deposits or altered plate structures were noted in this test nor were any changes in float voltages or currents observed.[23]
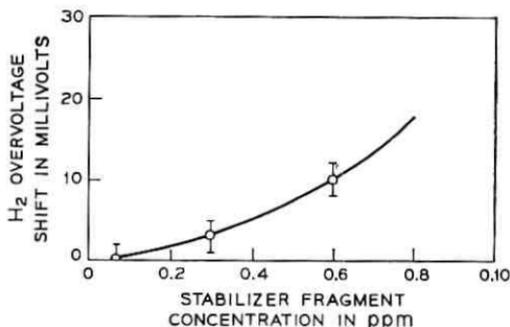
Fig. 6—Overvoltage shift vs stabilizer fragment concentration in BTL cell.

The appearance rate of the stabilizer decomposition products in the battery at room temperature is now known (Fig. 4) and we have established the relationship between the stabilizer fragment concentration and the shift in the hydrogen overvoltage (Fig. 6). We can therefore construct a plot showing the increase in the hydrogen overvoltage (Fig. 7) as a function of time at room temperature. All that remains is the establishment of a maximum allowable overpotential shift and the limit can be read directly from this plot.

Since an increase in the hydrogen overpotential of 50 mV will produce a disastrous increase in the corrosion rate of the positive grid, the allowable shift must be less than this value. It is considered that a maximum of 30 mV is a sufficiently conservative value. Figure 7 shows that even with a worst-case interpretation of the data, the lifetime limit imposed by the organo-tin stabilized test plastic is in excess of 50 years and thus we should not see electrochemical problems in Bell Telephone Laboratories-designed cells due to the plastic case for at least this period of time.

VI. DISCUSSION

Though the experimental work reported here was very specifically related to the immediate problem, we believe it to have wide applicability.

6.1 *Diffusion and Extraction*

Though no assumptions have been made concerning the mechanism of the extraction process and the work described is quite empirical, it is probable that the rate limiting step involves diffusion of the stabilizer
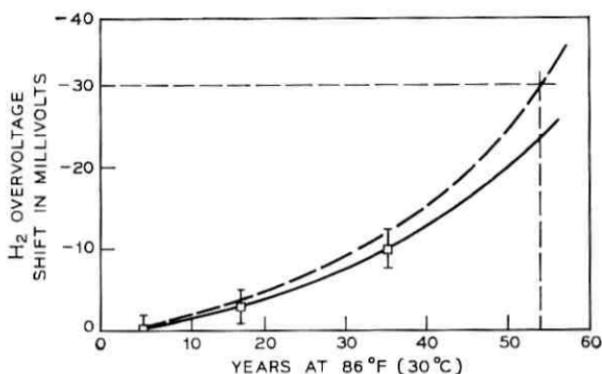
Fig. 7—Life limit of BTL cell due to test plastic.

to the electrolyte/plastic interface. There exists some supporting evidence. When thinly shaved samples of the test plastic were extracted in sulfuric acid, up to 80 percent of the available tin compound was extracted into the acid, but the physical dimensions of the sample shavings were unchanged, and there was little ($<$ 2 percent) weight loss. If the tin extraction were only due to dissolution of the plastic surface, about 80 percent of the sample would have had to dissolve to produce the measured tin concentration. Thus the stabilizer must have diffused through the polymer matrix to the acid. In addition the 10 ± 1 kcal/mole activation energy is comparable with the activation energy of other diffusing species in PVC as shown in Table II.[24,25] The fact that the

TABLE II[24,25]—DIFFUSION IN VARIOUS PLASTICS

Polymer	Diffusant	Activation Energy kcal/Mole
Rigid PVC	N_2	14.79
Rigid PVC	Ar	12.3
Rigid PVC	O_2	13.03
Rigid PVC	CO_2	15.44
Rigid PVC	CH_4	16.8
Rigid PVC	H_2O	9.98
Polyethylene	n-Octadecane	12.4
Polyethylene	DLTP	12.4
Polyethylene	N-ode	12.3
Polypropylene	N-ode	20.5
Polypropylene	DLTP	19.8
Polypropylene	Phenol A	22.3
Poly-4-methylpentene-1	DLTP	14.5
Poly-r-methylpentene-1	Phenol A	14.2

measured 10 ± 1 kcal activation energy is less than any of the values of Table II is comforting since extrapolation based on a low E_{act} will yield conservative time predictions.

6.2 *Future Applications*

If, in the future, methods are available for prediction of diffusion rates in plastics based on the average molecular weight, the physical structure of the polymer, the size and structure of the diffusing molecule and other such parameters, the problem of plastic material selection can be accomplished by *a priori* calculations rather than the time-consuming extraction described in this paper. In addition, plastics may in the future be used as dispensers of a desired additive, slowly releasing small amounts of the dopant which may, for example, be used to maintain a given float balance or retard a particular corrosion process.

Finally, the rotating disk electrode will probably gain importance as a tool for testing compatibility of various solutes in electrochemical devices. Since the electrodes appear to be at least as sensitive as most analytical methods for determination of low solute concentrations, it will probably be calibrated and used directly as an analytical tool, bypassing the time-consuming sampling and separate analytical procedures.

VII. CONCLUSIONS

7.1 *Life Limit of BTL Cell Due to Container Material*

It has been shown that even a "worst-case" treatment of the data results in a projected 50-year minimum life limit for the BTL cell with a container fabricated of the test plastic. This prediction is based on 86°F nominal ambient. If the cells are required to operate at higher temperatures, a corresponding life limit can be calculated by use of the measured 10 ± 1 kcal activation energy.

7.2 *Rotating Disk vs Real Cell Experiments*

The rotating disk provides a quick method for establishing relative levels of electrochemical effects due to a wide range of materials. An example of its utility was demonstrated in early field trials in which a polarization shift was seen in a small number of cells. This resulted in the positive plates floating at their reversible potential and the entire 110 mV excess potential was shifted to the negative plates—a condition that would result in self-discharge of the positive plates. By use of the rotating disk electrode apparatus as a screening tool, it was determined

that cyclohexanone, a solvent used in the jar-cover sealing procedure, was entering the cell electrolyte.

Potential scans through the hydrogen evolution region of the lead disk were made in (1.210 sp. gr.) H_2SO_4 solutions containing various amounts of cyclohexanone. The shift in hydrogen overvoltage (η_{H_2}) relative to uncontaminated acid solution was approximately 50 mV for electrolyte containing 10 ppm cyclohexanone, an amount which was shown to be easily leached from a freshly made cyclohexanone solvent-sealed joint into the battery electrolyte of a BTL cell. The overvoltage shift was toward more negative potentials, the same direction observed on negative plates in the field trial cells. The result of these experiments was the development of an alternate sealing procedure,[26] and the problem has not reappeared in subsequent field trials.

VIII. ACKNOWLEDGMENTS

REFERENCES

1. Vinal, G. W., *Storage Batteries*, New York: Wiley & Sons, 1966, pp. 67–68.
2. Barak, M., "Plastics in Batteries and Fuel Cells," Rubber and Plastics Age, *48*, No. 11 (November 1967), pp. 1202–1205.
3. Seymour, R. B., and Steiner, R. H., *Plastics for Corrosion Resistant Applications*, New York: Reinhold, 1955, pp. 63–80.
4. *1969 Book of ASTM Standards*, Part 27, Philadelphia, Pa.: American Society of Testing Metals, 1969, pp. 152–158.
5. Huseby, T. W., Ryan, J. T., and Hubbauer, P., "PVC Battery Jars and Covers," B.S.T.J., this issue.
6. O'Sullivan, T. D., Biagetti, R. V., and Weeks, M. C., "Electrochemical Characterization of the Bell System Battery," B.S.T.J., this issue, pp. 1335–1358.
7. Lander, J. J., "Anodic Corrosion of Lead in H_2SO_4 Solutions," J. Electrochem. Soc., *98*, No. 6 (June 1951), pp. 213–219.
8. Milner, P. C., "Float Behavior of the Lead-Acid Battery System," B.S.T.J., this issue, pp. 1321–1334.
9. Feder, D. O., "Positive Grid Design Principles for the Bell System Battery," B.S.T.J., this issue, pp. 1279–1303.
10. Lander, J. J., "Further Studies on the Anodic Corrosion of Lead in H_2SO_4 Solutions," J. Electrochem. Soc., *103*, No. 1 (January 1956), pp. 1–8.
11. Ishgarishev, N., and Berkman, S., "Die Überspannung Aut Den Elektroden in Zusammenhang Mit Der Hydvatation Der Ionen," Z. Elektrochem., *28*, No. 1/2 (January 1922), pp. 40–55.
12. Frumkin, A. N., *Advances in Electrochemistry and Electrochemical Engineering*, Delahay, P., editor, New York: Interscience, *1*, 1961, pp. 65–121.
13. Breiter, Manfred, *Advances in Electrochemistry and Electrochemical Engineering*, Delahay, P., editor, New York: Interscience, *1*, 1961, p. 123.

14. Frumkin, A. N., *Advances in Electrochemistry and Electrochemical Engineering*, Delahay, P., editor, New York: Interscience, *3*, 1961, p. 287.
15. Gileadi, Eliezer, *Electrosorption*, New York: Plenum, 1967, pp. 1–18.
16. Frumkin, A. N., *Modern Aspects of Electrochemistry*, Bockris, J. O'M., and Conway, B. E., editors, *3*, Washington, D. C.: Butterworths, 1964, pp. 149–223.
17. Frumkin, A. N., *Transactions of the Symposium on Electrode Processes*, Yeager, F., editor, New York: Wiley, 1961, pp. 1–15.
18. Delahay, P., *Double Layer and Electrode Kinetics*, New York: Interscience, 1965.
19. Braunauer, S., Emmett, P. H., and Teller, E., "Absorption of Gases in Multi-molecular Layers," J. Am. Chem. Soc., *60*, No. 2 (February 1938), p. 309–319.
20. Miller, B., and Visco, R. E., "Ring Disk Amperometry: A Study of Indium Dissolution," J. Electrochem. Soc., *115*, No. 3 (March 1968), pp. 251–258.
21. Miller, B., unpublished work.
22. Eberle, A. R., and Lerner, M. W., "Determination of Tin and Molybdenum in the Nuclear Reactor and Other Materials," Anal. Chem., *34*, No. 6 (May 1962), pp. 627–635.
23. Cannone, A. G., unpublished work.
24. Jackson, R. A., *et al.*, J. Appl. Polym. Sci., *12*, (1968), p. 1297.
25. Tikhomirov, B. P., *et al.*, "Diffusion of Additives in Polyolifins," J. Appl. Polym. Sci., *12*, No. 6 (June 1968), pp. 1297–1309.
26. Dahringer, D. W., and Shroff, J. R., "Jar Cover Seals," B.S.T.J., this issue, pp. 1393–1404.

Lead-Acid Battery:

Jar-Cover Seals

By D. W. DAHRINGER and J. R. SHROFF

A novel jar-cover seal has been designed and fabricated for the new Bell System battery. The seal consists of a tongue-and-groove seal joint design (jar and cover respectively) with an infrared heat absorbing layer at the joint interface. This heat absorbing layer causes the fusion and sealing of the jar and cover surfaces when exposed to high intensity concentrated infrared energy. The seals prepared in this manner are leak free, acid resistant and have extremely high joint strength. A jar-cover seal life equivalent to the useful life of the polyvinyl chloride in the jar is anticipated.

I. INTRODUCTION

The leakage of electrolyte from lead-acid storage batteries has been a chronic problem to consumers. Industrial and individual consumers share such common seal-related deficiencies as electrolyte leakage, corrosion buildup and premature battery failure. Additionally, telephone companies have experienced high maintenance costs, fire and explosion hazards, and interrupted service.[1] These problems can be related to both the inherent characteristics of the lead-acid system and to inadequate joint design considerations. The limited availability of materials resistant to battery acid and the dimensional changes which occur in electrodes and plates are factors, inherent in lead-acid batteries, which result in poor reliability. Poorly designed cap or plug type jar-cover seals on batteries currently used in telephone plant provide very little effective seal area. In addition, they all use adhesives or sealants with low toughness which further contributes to premature battery failure.

These effects are more obvious in batteries for telephone use because of their large dimensions, type of service and long design life.

Early in the development of the new cylindrical battery, studies were undertaken to provide reliable jar-cover seals. The initial performance requirements for the seal were that:

(*i*) The seal joint be permanently leakproof—to both electrolyte and internally generated gas.

(*ii*) The seal material be electrochemically compatible with the lead-acid battery system.

(*iii*) The seal material be stable to all concentrations of battery acid that it might encounter.

(*iv*) The seal be strong enough to support safely the weight of a complete battery (approximately 400 lbs.).

(*v*) The seal be resistant to shock and lateral deformation.

In addition, the seal joint design had to be compatible with an injection-moldable jar and cover and the sealing procedure had to be economical and feasible for production.

II. JAR-COVER SEAL JOINT DESIGN

The circular battery jar geometry and an impact-resistant, unplasticized, polyvinyl chloride (PVC) jar material were selected prior to detailed considerations of the seal design.[2] Both selections ultimately proved to be fortunate in terms of seal joint design and sealing technique.

Initially, consideration was given to the possible use of plug or cap type joints between the jar and cover. These early designs were abandoned in favor of a tongue-and-groove arrangement. Such a design offered advantages of self-alignment with improved shock and static stress resistance.

Figure 1 shows a diagrammatic cutaway view of an assembled jar and cover with a sectional detail of the actual joint area. Incorporated in the design are matching tapers ($4\frac{1}{2}°$) on both the inner and outer surfaces of the jar and corresponding tapers on the cover. This design permits the assembly of the joint with essentially no sliding friction, with a maximum bondline area and with self-adjusting (over a broad range, including zero) bondline thickness.

III. SEALING METHOD

3.1 *Early Sealing Method Investigations*

Before the ultimate sealing technique was selected, several methods were considered. These include:

(*i*) Adhesive bonding with:

    (*a*) solvent or bodied solvent* applied by dipping or brushing,

---

* The solvent used was cyclohexanone which was bodied or thickened with between 15 and 25 weight percent of the jar and cover plastic. (Airco B983 - Propylene-modified PVC.)
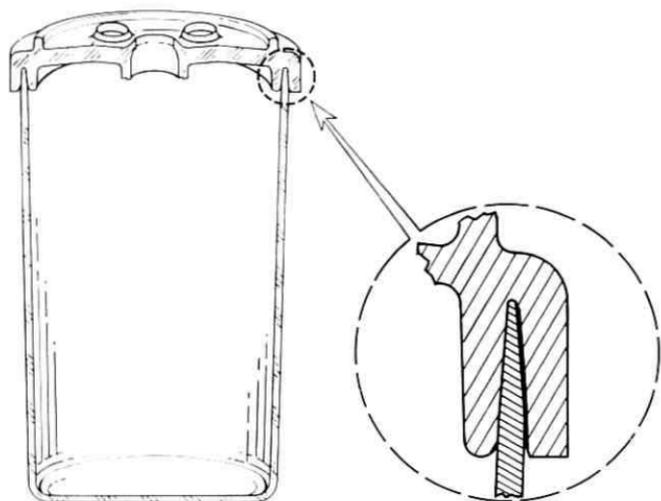
Fig. 1—Jar-cover seal design and assembly.

(b) bodied solvent applied by injection,
(c) frozen solvent gasket, and
(d) solventless adhesive;
(ii) Welding with embedded, heated resistance wire;
(iii) Preheating with hot matching dies;
(iv) Spin welding; and
(v) Ultrasonic welding.

Of these early methods, the solvent-related sealing techniques offered the best combination of production feasibility, low cost and acceptable seal strength. After extensive mechanical testing, several field trial cells were prepared using the solvent-dip technique. Although the mechanical characteristics of these seals were satisfactory, electrode polarization was encountered. Investigations confirmed that contamination of the electrolyte by cyclohexanone from the sealant was responsible for the polarization.[3] Investigation of all of the above sealing methods was superceded by the investigation of the infrared sealing technique described below.

3.2 *Infrared Sealing Technique*

An infrared sealing technique was devised as a practical, production-oriented method for making the jar-cover seal. The method involves concentrating sufficient thermal energy at the jar-cover interface to

permit welding to occur. Specifically, a focused source of high-intensity infrared energy is directed through the outer cover skirt to the interface between the cover and the jar. At this interface, a thin layer of infrared-absorbing material becomes hot, melts, and causes the melting and fusion of the adjacent jar and cover surfaces. The fused layers solidify upon cooling to form the seal.

A quartz line heater* was used in the initial experiments to provide a focal line energy of approximately 125 watts per inch when operated at 110 volts ac. Later, other operating conditions were found to be better (see Section 3.4). The absorbing layer was a 0.0015 to 0.002 inch thick black coating of the same plastic used for the jar and cover molding and containing a small quantity of a carbon black dispersion. With the cover in place, the infrared source was focused on the coated surface at the jar cover interface and the battery rotated with respect to the infrared source until the entire seal was formed. The assembly and sealing steps are shown schematically in Fig. 2.

Times ranging from $\frac{1}{2}$ minute to 2 minutes were necessary, depending on the experimental conditions employed, to achieve complete 40-inch circumferential seals. Shorter sealing times are possible with variations in the infrared source parameters. This will be discussed in a later section.

Of particular importance in the successful use of this technique is the alignment of mating surfaces and the availability of contact pressure at the interface during sealing. An attempt to seal a joint with either a misalignment or a lack of follow-up contact pressure would result in only a one-sided melt and a low strength, potentially leaky, seal. The jar-cover seal design incorporated features to achieve alignment and contact pressure within the range of molding tolerances on the jar and cover diameters, provided the original tapers are accurate.

During the sealing operation, we can observe the point at which a seal forms. As the absorbing layer melts and fuses at the jar-cover interface, the air layer disappears, providing a visible "wetting" of the cover and blackening of the absorbing layer. This characteristic is extremely helpful in quality control evaluation of the seal.

If heating is continued for a period of time beyond that necessary for the seal to form, blistering, burning and charring can result on the outer surface of the cover. This causes an unsightly appearance and degradation of the materials involved. Experiments designed to minimize this problem are discussed in Section 3.4.

---

* Research, Incorporated, Model 5215-5, "Quartz Line Heater" with water-cooled elliptical reflector, and 1200 T3/CL/HT lamp.
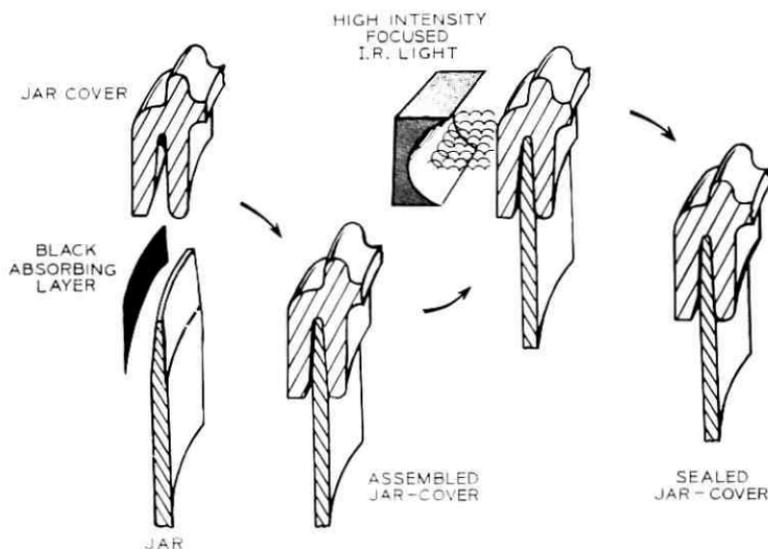
Fig. 2—Jar-cover seal assembly and sealing sequence (schematic).

3.3 *Infrared Absorbing Medium*

The infrared sealing technique was originally evaluated using a black film made from a solution of the jar cover material in cyclohexanone to which a carbon black dispersion was added. Sections 1 and 2 of Table I give the formulation data. Because of handling difficulties with the film, a direct coating technique of the jar lip was devised. Cyclohexanone was originally selected as a solvent because of its solution capacity and relatively low hazard potential. However, this solvent was ultimately replaced by more volatile methyl ethyl ketone (MEK) to reduce possible cell contamination due to residual solvent and to accelerate solvent evaporation. The coating formulation with MEK is shown in Section 3 of Table I. The solids content of this formulation was changed since the MEK is capable of dissolving only up to 15 percent of the PVC.

The present technique does not form a seal at the inner jar-cover interface because little or no energy passes through the absorbing layer at the outer jar-cover interface. Should it prove desirable to seal the inner jar-cover interface, this can be accomplished by applying the absorbing layer to the entire inner jar surface and only the upper half of the outer jar surface. When exposed to the radiation, such an arrangement leads to a two-sided seal.

TABLE I—INFRARED ABSORBING MATERIALS

1. Coating Solution	
25% Airco B983 in Cyclohexanone	— 95.0
Carbon Black dispersion*	— 5.0
Total	100.0

2. Dry Film Components	
Airco B983	— 92.7
Carbon Black	— 2.9
Vinyl resin VYHH	— 4.4
Total	100.0

3. Methyl Ethyl Ketone Based Coating Solution	
15% Airco B-983 in MEK	— 95.0
RBH #5404	— 5.0
Total	100.0

* RBH #5404, Inmont Corp.
Carbon Black   — 15.0
Vinyl resin VYHH   — 22.5
Methyl Ethyl Ketone   — 62.5
Total   100.0

3.4 *Operating Parameters of Infrared Heat Seal Technique*

As explained in Section 3.2, most of our initial experiments were carried out with the lamp operated at 110 volts and the seal in the focal plane. During the course of these experiments it became clear, however, that the margin between seal formation and the onset of blistering and charring was not as wide as one might wish. We have therefore investigated the effects of running the lamp hotter and of moving it further away.

In order to determine which ranges of wavelength are effective for forming the seal and which for causing blistering and charring, the infrared transmission characteristics of the jar or cover material were

obtained, using a Beckman IR-9 spectrophotometer. The infrared transmission spectrum of the PVC (Fig. 3) shows a small absorption peak around 1.2 μ and a region of strong absorption beyond 1.7 μ. The blackened film used for sealing, on the other hand, may be presumed to be strongly absorbing throughout this region. Ideally, therefore, a light source with its emission restricted to the range 0.17 μ to 0.6 μ would be preferred. Since optical filtration is not practical in view of the high radiant flux, some control of the emission spectrum can be achieved by adjusting the temperature of the source. Figure 3 shows computed values for the spectral distribution of the light emitted by the lamp at three different voltages, based on measurements of the color temperature of the filament made with an optical pyrometer. It will be seen that, other things being equal, higher voltage operation should be preferable because a larger fraction of the emitted energy lies in the range of wavelength below 1.7 μ, where it is effective for sealing.

To investigate the effect of varying the voltage applied to the lamp (and its temperature), test specimens were assembled from small squares of jar. These were clamped together with a film of the absorbing material between them. A recording thermocouple was placed immediately behind and in direct contact with the film. The infrared source was then focused on the area of the seal at the thermocouple and temperature
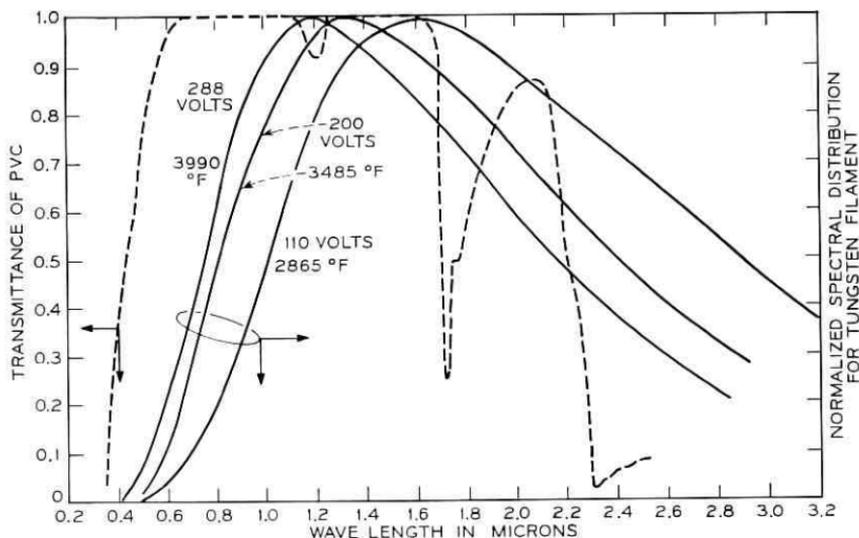

Fig. 3—Infrared transmittance of PVC and emission spectra of tungsten filament lamp.

recorded as a function of time. The instants at which the seal was formed and at which charring became evident were recorded by visual observation. Results are shown in Fig. 4 for three different voltages applied to the lamp. It will be seen that, while there is some decrease with increasing voltage in the time to formation of seal, there is also a substantial decrease in the time to onset of charring. This is to be expected since, even though the *relative* amount of radiation at wavelengths beyond 1.7 μ decreases with increasing filament temperature (Fig. 3), the absolute amount does not. It is possible to compensate for the increase in overall intensity by moving the lamp further away from the seal. Figure 5 shows what happens when the lamp is run at 150 volts and moved progressively further away from the seal. It can be seen that the increase in distance is accompanied by a considerable increase in char time, with only a slight increase in seal time, and that, somewhere between one and two inches, charring no longer occurs. The increased sealing time at the 2-inch distance is compensated by the broader section of seal which can be made with the out-of-focus beam.

IV. MECHANICAL STRENGTH TESTS

Several tests were used to evaluate the breaking strength of jar-cover seals. Shear separation tests were conducted on specimens cut from a
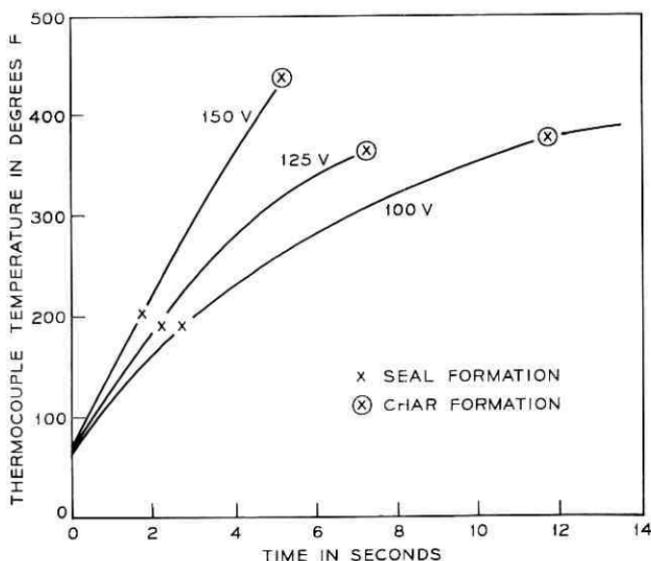

Fig. 4—Thermocouple temperature vs time for various IR source voltages (absorption layer at source focus).
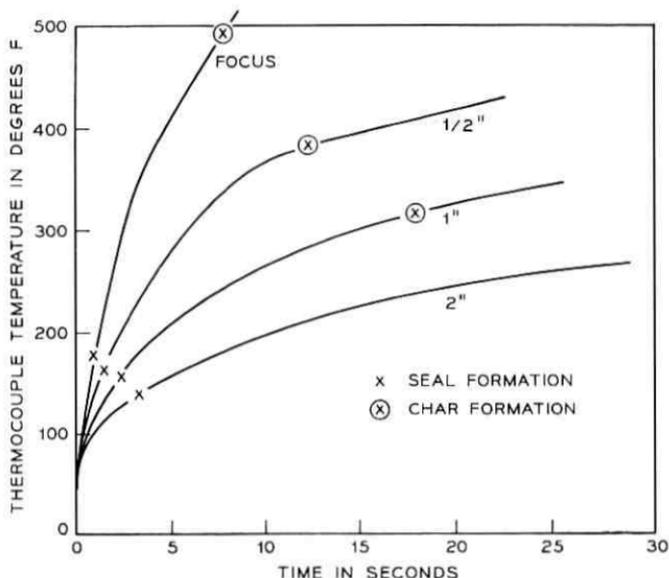
Fig. 5—Thermocouple temperature vs time at various distances from focus at 150 V.

jar-cover seal in the form of a wedge segment of the cover with a one-inch arc. The specimens were placed in a testing fixture and stressed at a rate of 0.1 inch per minute using a Tinius–Olsen "Universal Testing Machine." The force required to separate the seal joint segment was determined and recorded as pounds per circumferential inch of seal joint. The testing fixture with a specimen is illustrated in Fig. 6.

Initial joint strength test results for the infrared sealed jar-cover assembly exceeded the tensile strength of the jar wall using the shear strength fixture and method described above. The actual failing loads averaged about 1500 pounds per inch of seal and generally the failures occurred in the jar wall immediately below the bond area. A few specimens fractured in several places, that is, the top and middle of the outer cover skirt, and a few failed by yielding of the jar wall below the bond area. However, none of these specimens showed any sign of bond area damage or failure. With a total of 40 inches of circumferential seal and each inch capable of supporting three or more times the weight of the battery, there can be little doubt as to the adequacy of the seal strength.

Several jar-cover assemblies, containing 700 pounds of lead weights, were suspended by a ring support under the outer cover skirt. After
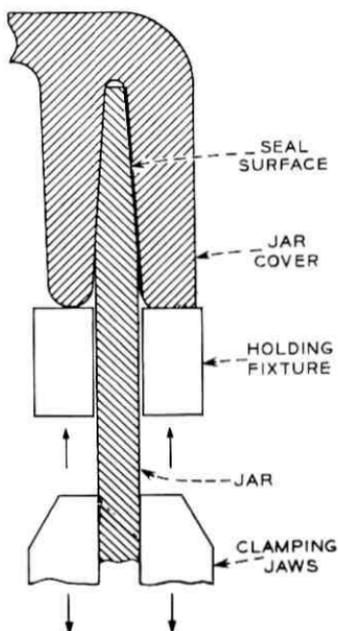
Fig. 6—Jar-cover shear strength test.

three weeks suspension and several deliberate 6″ drops to a concrete surface, no evidence of joint failure could be found.

Additional tests were conducted on sealed empty jars by pressurizing to 5 psi with air and inspecting for leaks. These tests were continued for many months without loss in pressure. Commercial practice for cells currently delivered to the Bell System calls for leak testing at a pressure of only $\frac{1}{2}$ psi for 30 seconds.

V. CONCLUSION

It is demonstrated that reliable, jar-cover seals can be prepared for the new lead-acid battery. Of the variety of sealing techniques investigated, heat sealing by an infrared heat source focused on an absorbing medium placed at the seal interface was found to be economically attractive and production oriented. Seals obtained by this technique are reliable and provide high seal strength. An added advantage of this sealing technique is the ability to determine quality of the seal by simple visual inspection during or after the operation.

VI. ACKNOWLEDGMENTS

REFERENCES

1. Koontz, D. E., Feder, D. O., Babusci, L. D., and Luer, H. L., "Reserve Batteries for Bell System Use: Design of the New Cell," B.S.T.J., this issue, pp. 1253–1278.
2. Huseby, T. W., Ryan, J. T., and Hubbauer, P., "Polyvinyl Chloride Battery Jars and Covers," B.S.T.J., this issue, pp. 1359–1376.
3. Butherus, A. D., and Lindenberger, W. S., "Electrochemical Compatibility of Plastics," B.S.T.J., this issue, pp. 1377–1392.

Lead-Acid Battery:

Post Seals for the New Bell System Battery

By L. H. SHARPE, J. R. SHROFF and F. J. VACCARO

A novel post seal has been designed and fabricated for the new Bell System battery. The seal consists of a cylindrical "clamp" of a rigid cured epoxy resin molded in place on the lead posts and a flexible one-piece butyl rubber structure attached to the epoxy and to the battery jar cover. Accelerated tests show that the seal effectively slows lead corrosion. A useful seal life in excess of 40 years is predicted.

I. INTRODUCTION

The posts on all lead-acid batteries are sealed to prevent the escape of sulfuric acid electrolyte from the battery container. If this highly corrosive aqueous solution of sulfuric acid escaped, it would constitute a hazard and a nuisance in a central office. Electrolyte leaking from the seal of a post can create a high conductivity path to ground. This path, connected to the battery, can lead to heating and arcing which may result in the ignition and burning of adjacent combustible materials. In addition, batteries with leaking post seals require costly periodic maintenance. Leaked electrolyte must be cleaned up, neutralized and replaced.

The post seals of many batteries in float operation leak after being in service for times which are incompatible with Bell System objectives of reliable component performance for twenty to forty years. In this paper, we describe a post sealing system which is developed to be compatible with the expected long-lived behavior of the other components of the new Bell System battery.

Post seals (there are three types) on batteries currently in telephone plant have at least two features in common:

(*i*) They exert circumferential pressure on the post to make the seal.

(*ii*) They have seal lengths of less than one-half inch. One type of seal is flexible so as to accommodate motion of the post, relative to

the cover, due to plate growth in service. The other types are rigid. None of them offers any redundancy in sealing.

Figure 1 pictures each of the three types of post seals currently in use. Each seal has a relatively short length and acid eventually works its way to external electrical connections or to ground. Two of these designs (Types 1 and 2) are rigid sealing systems so that premature
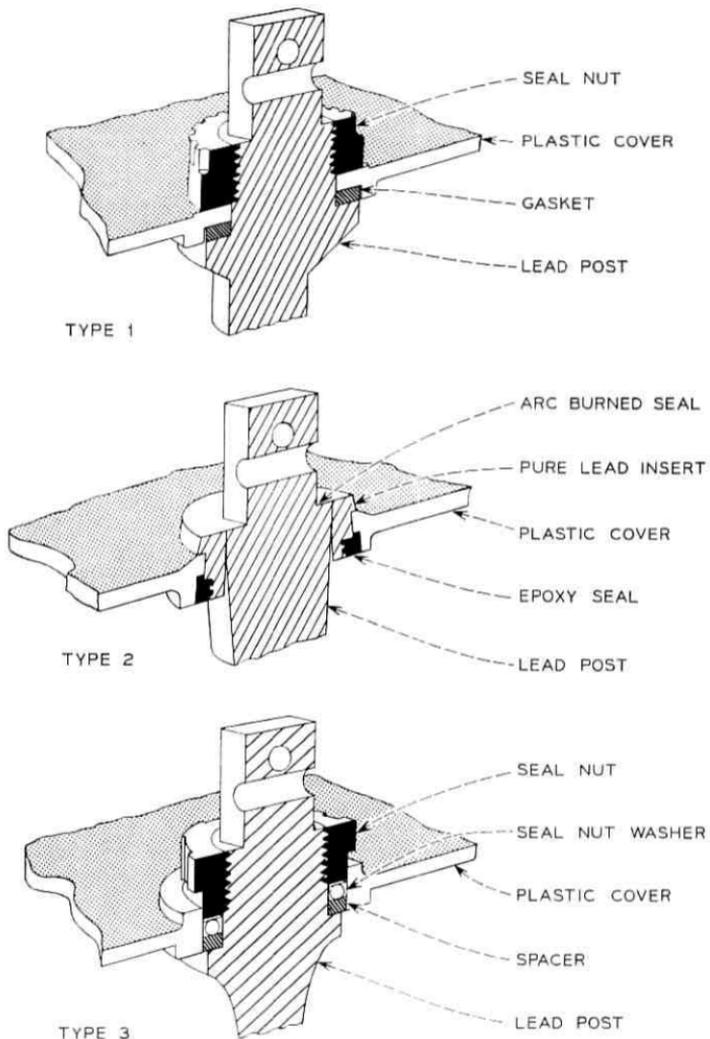

Fig. 1—Current post seals.

failure can occur when element movement (from whatever cause) exerts abnormal stress on the seal.

Because of the hazards involved when post seals leak acid the Bell System defines procedures for maintaining batteries so as to produce a clean and essentially acid-free environment. Proper maintenance requires that the individual cell posts and cover be wiped regularly. On the basis of the current acquisition rate of 100,000 cells/year in KS 15544/L 400 and 500 series and an average life of 13 years, the total cell population should reach 1.3 million over the next few years. At a very conservative estimate of $6/cell/year for labor costs required to clean cells due to post leakage, this generates a total of $7.8 million in costs per year for the Bell System.

The performance criteria for the new battery post-seals are as follows:

(*i*) The seals must be resistant to acid leakage. Acid may be splashed into the space above the elements during transportation or placement; it will be sprayed above its normal level during gassing; it will attempt, by corrosion and capillarity, to move up the lead posts.

(*ii*) The seals must be compatible with the electrolyte and also with the electrochemical conditions at the posts and the electrochemistry of the battery. That is, the seals and the materials comprising the seals must not be chemically degraded either by the electrolyte or by the corrosive conditions at the posts. Additionally, they must not produce material, in time, at either post, which will alter the electrochemistry of the battery,

(*iii*) They must accommodate motion of the plate stack due to vibration in shipment and plate growth in service. The design objective was that a one-inch vertical motion of the plate stacks and posts relative to the cover be permitted, without breaking the seal.

(*iv*) The seals must be readily manufacturable. Further, procedures for their assembly must be compatible with other assembly procedures for the battery.

II. DESIGN AND CONSTRUCTION

Figure 2 shows early post seals, in batteries currently on field trial, which consist basically of three separate parts.

A *primary post seal* of heat-shrinkable, crosslinked polyethylene tubing which is shrunk to fit tightly on the lead post coated with NO-OX-ID-"A", a nonsoap, thickened, paraffinic grease highly resistant to oxidation.
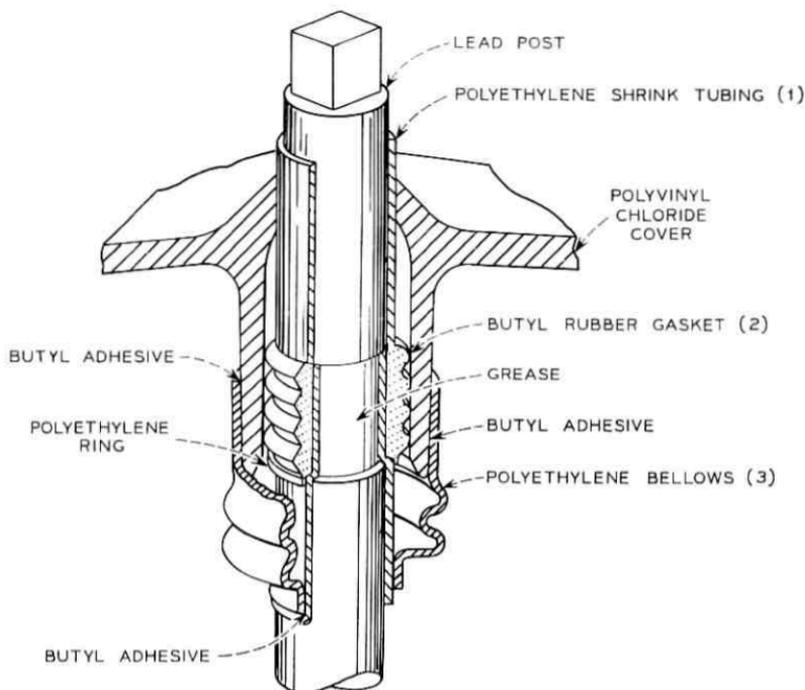
Fig. 2—Three-piece post seal assembly.

A flexible bellows of the same or similar polyethylene (which permits relative motion of the post and cover) joined to the shrink tube and to the skirt of the post hole in the cover with a thermoplastic polyisobutylene adhesive.

A secondary seal which is a corrugated butyl gasket, tightly fitted around the shrink tube, which positions the post centrally in the cover hole and prevents acid and gases from escaping if the bellows should fail. The gasket fits tightly against the wall of the hole but is able to slide against it to accommodate post motion.

These seals were thought to have at least three advantages over seals in batteries presently in service:

(*i*) Corrosion paths to the outside world are longer—an order of magnitude longer than current designs.

(*ii*) They accommodate post motion due to plate growth and element movement during shipment and installation or during an earthquake or nuclear attack.

(*iii*) They provide a redundant electrolyte and gas sealing system.

This three-piece seal, however, has been superseded by other designs mainly because of materials problems and the difficulty of assembling it.

The next logical step in the seal design was the creation of a one-piece seal which combined all of the components of the previous design into a single molded part. This design, which was molded in butyl rubber, is shown in Fig. 3. In it we have the elements of the earlier design—the sleeve, the gasket and flexible bag which performs the function of the bellows of the previous design.

The sleeve is made so as to fit tightly on the post and the seal is assembled as follows. Using the lead post as a plug at the lower end of the sleeve, air pressure is applied at the upper end so that the sleeve inflates slightly. The sleeve is then easily slid over the post, previously coated with either a nonoxidizing grease or a noncuring butyl sealant, into proper position and the air pressure released. The bag is then turned inside out and fastened to the skirt of the cover hole with a butyl adhesive, completing the seal. This one-piece seal takes considerably less time, effort and technique to assemble than the three-piece design.
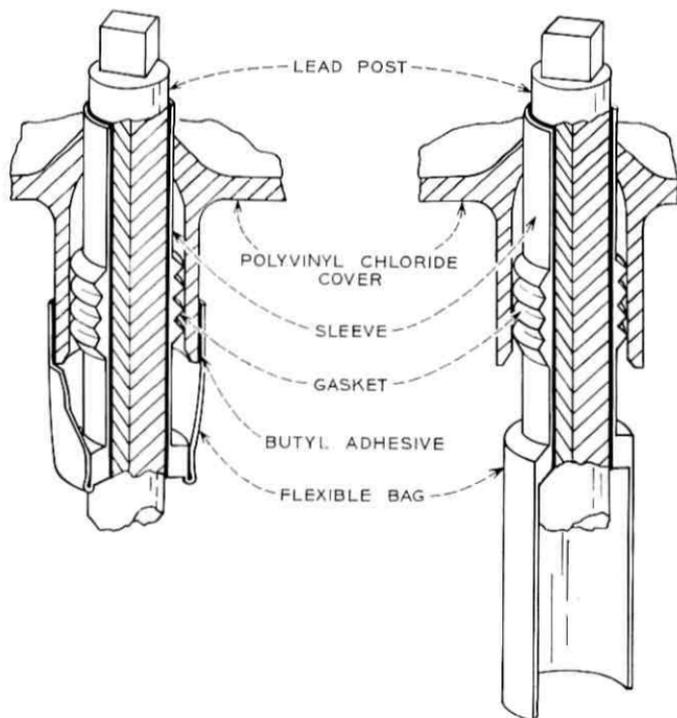
Fig. 3—One-piece post seal assembly.

With respect to the primary post seal, circumferential pressure of the rubber sleeve in combination with the nonoxidizing grease provides a lower limit of corrosion restraint. A multiplicity of clamping rings spaced up the length of the sleeve would provide an additional level of restraint.

A lead-acid cell positive post at a potential above its thermodynamically reversible value will corrode or oxidize. The electrochemical reactions of importance at this post are as follows:

$$2H_2O \rightleftarrows O_2 + 4H^+ + 4e^- \tag{1}$$

$$Pb + 2H_2O \rightleftarrows PbO_2 + 4H^+ + 4e^-. \tag{2}$$
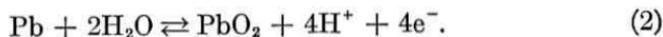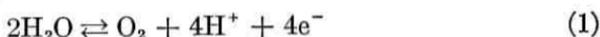
At the positive post, oxygen is evolved at a kinetically controlled rate (equation 1) and lead post material is oxidized (equation 2).

From the above, it is apparent that with corrosion at the positive post there occurs a decrease in density of surface layer material resulting in a volume increase and a larger post diameter. In addition, there is the evolution of oxygen. The volume increase and the evolving oxygen can readily displace a soft or deformable material from the surface of the post and allow electrolyte and resulting corrosion product to progress upward.

When the lead oxide on the negative post is at a polarized potential, it can be reduced and hydrogen liberated. The pertinent electrochemical reactions are as follows:

$$2H^+ + 2e^- \rightleftarrows H_2$$

$$PbO + 2H^+ + 2e^- \rightarrow Pb + H_2O.$$
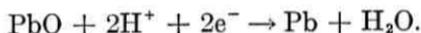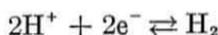
The production of a porous surface with evolving hydrogen could result in rapid capillary flow of acid up the post.

Consideration of the above failure modes dictates the selection of a sealant material which will not readily yield to the stresses developed during lead oxidation-reduction. A restraining cylinder having appropriate mechanical properties and surrounding the post would presumably cause the lead dioxide corrosion product to be produced in the form of a dense barrier offering relatively few capillaries and decreased opportunity for "wicking" of acid up the post. The corrosion rate up the post would then depend upon the compression which the restraining cylinder could provide. The use of a restraining cylinder on the negative post, it is felt, would not result in the generation of high compressive forces in the absence of a volume increase at the surface.

Although there are many mechanical clamping devices available to form a compression seal, material selection and clamp configuration

do present problems. A restraining cylinder of material molded in place on the lead post seemed to offer a relatively simple compression seal in which the complete post length could be utilized. Epoxy resins, cured in place, were selected for the restraining material because they offered favorable chemical and mechanical properties and ease of use.

Figure 4 shows a post seal combining all features of the previous seals with a full length circumferential "clamp" which may be made even more effective by the inclusion of grooves in the lead post. It consists of:

(*i*) The primary post seal—a fairly rigid crosslinked epoxy resin cast around the post-providing a long path length against acid creepage.
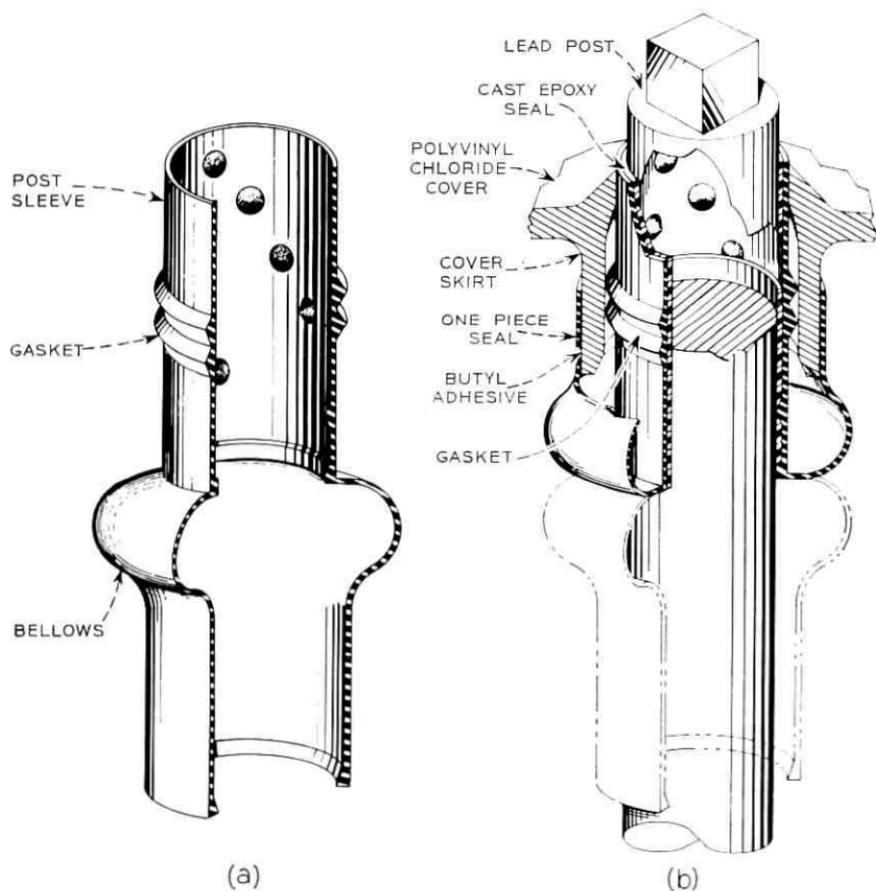

Fig. 4—Final post seal assembly: (a) one-piece post seal, (b) post seal assembly.

(*ii*) The one-piece butyl rubber structure previously described, fitted over the epoxy resin, which provides both the primary and secondary electrolyte and gas seals, that is, the flexible bag and the corrugated gasket.

The sleeve of the one-piece rubber structure has, however, been modified so as to make it serve the function of a mold during casting of the epoxy resin around the post.

Fifteen randomly distributed conical projections are provided on the inner surface of the sleeve. These projections allow proper location of the sleeve on the post providing necessary annular space for molding the epoxy "clamp." The projections also facilitate random flow of the epoxy into the space and flow of air bubbles out. A heavy molded ring at the lower end of the sleeve fits tightly to the post eliminating possibility of epoxy leakage before cure. The curing in place of the epoxy in contact with the sleeve also provides a good joint between the epoxy and the rubber sleeve. The post seal is assembled as follows:

The rubber sleeve is located in place over the post by the conical projections. A supporting fixture is used to enclose the sleeve and prevent bulging during epoxy injection. The epoxy is then injected in the annular cylindrical space between the post and sleeve and cured. The cover is located in the proper position over the corrugated gasket. The bag is then turned inside out and fastened to the skirt of the cover hole with a butyl adhesive, completing the seal.

Materials selection as to compatibility with lead-acid cell performance was of primary concern. After consideration of mechanical properties and design criteria, it had to be insured that the proposed material was unaffected by the acid electrolyte.

The epoxy sealant was evaluated for inertness by performing leaching experiments in 1.210 specific gravity sulfuric acid. The epoxy was first cured in the form of sheets 0.1 inch thick. Pieces cut from these sheets were added to a quantity of acid, maintained at a temperature of $70° \pm 2°C$, such that the ratio of epoxy surface area to solution volume was at least equivalent to that expected in a cell. At predetermined intervals, samples of the solution were removed and cooled to room temperature. After cooling, the solutions were analyzed, by means of the current-potential behavior observed at rotating lead disks under linear potential scan conditions, to determine if electrochemically active materials were present. Scans on the lead disk were both anodic and cathodic in direction to simulate reactions at the negative and positive cell plates. Leaching was continued and scans were made until

the epoxy had become brown in color, indicating severe oxidation. At no period in the acid leach test did the lead disk analysis indicate the presence of materials which would alter the normal operation of a lead-acid cell.

Before acceptance of the butyl rubber for the one-piece seal material, it was also examined by a method similar to the above. Leach solution was tested by rotating lead disk techniques as well as by additions of butyl rubber-acid solutions to charged cells with pasted plates. This material also proved to be clean electrochemically. Materials selection and the details of assembly procedures and techniques are not yet final. However, we expect that the general features of both post seals in the final design of the new Bell System battery will be as we have just described them.

III. ACCELERATED TESTING AND RESULTS

Accelerated degradation of the primary post seal for test purposes was accomplished by potentiostating the lead post with proposed seal at a relative high overpotential in standard battery electrolyte. The increase in corrosion rate at the current densities required to maintain the test overpotentials results in accelerated progress of electrolyte and resulting corrosion product up the lead post. Tests were carried out at normal room temperature ($\approx 75°F$). We did not test at an elevated temperature because we felt that development of thermally induced stresses, heat degradation of materials, and so on, would unnecessarily complicate analysis of seal behavior.

The primary concern of this study was evaluation of seals for the positive post. The corrosion mechanism, stress and environmental conditions at this post, it was felt, require a seal which should more than satisfy the requirements for a seal at the negative post. Preliminary accelerated test data on negative posts appear to justify the above belief.

Primary post seals were prepared for testing by first abrading lead posts 1.72 inch in diameter to brightness with #600 abrasive paper and then wiping carefully with clean rags. The cleaning technique used was chosen to insure minimum lead surface contamination and oxide formation so as to provide a reasonably well-defined surface on which to begin testing.

The epoxy resin plus curing agent was poured into a removable mold surrounding each post and cured to form a collar 0.1 inch thick and several inches long. Coatings less than 0.05 inch thick severely deformed

and disintegrated during test. The posts were then immersed in sulfuric acid (1.210 specific gravity) with the bottom of the epoxy collar submerged in the acid about 1 inch. Several posts were potentiostated at each of three voltages, that is, $+1.16$ V, $+1.30$ V, and $+1.50$ V relative to a $Hg/Hg_2SO_4/H_2SO_4$ (1.210 specific gravity) reference electrode. The voltage $+1.16$ V is the positive post potential of a battery under normal float conditions. The other potentials represent accelerated test conditions. Since the epoxy coating is transparent and the corrosion product dark brown, it was possible to observe and measure visually the progress of corrosion up each post. The precision of this measurement was better than ±0.010 inch. Measurements were made at 60-day intervals to determine a corrosion rate at each voltage.

Initial concern in any accelerated test procedure is that the accelerating parameter influence only the rate at which reactions take place. Alternately stated, accelerated testing should not cause reactions to occur which would not ordinarily occur at unaccelerated conditions. Temperature as an accelerant, for example, was rejected since it was felt that the reaction mechanism might change at elevated temperatures.

Voltage, more precisely current density, was selected as the accelerating parameter. It was then necessary to ascertain that the increased voltage influenced only the reaction rate of corrosion at the positive post. This was accomplished by potentiostating lead posts at each of the three above mentioned positive potentials.

At 60 days, voltage was plotted against corrosion rate on linear coordinate paper (see Fig. 5). The resulting linear relationship for this relative short period indicates the consistency of the reaction mechanism at the various voltages. Calculation of an acceleration factor from these data yields one day at 1.50 V as being equivalent to approximately nine days at 1.16 V (float potential). Plotting data at the 60th day, it was felt, would reflect purely corrosion phenomena with minimum influence due to resultant compressive forces or height of corrosion on the posts. Data at an early time for the 1.16 V test condition, however, would also include substantial measuring uncertainties.

Figure 6 graphically portrays the progress of corrosion up the positive post with time. Two potentials are plotted, $+1.50$ and $+1.16$ V, representing the accelerated and float conditions respectively. It is significant that the slope of the curve for 1.50 V is markedly decreasing with time. This decrease in slope or corrosion rate can perhaps be attributed to the production of compact corrosion product as previously hypothesized. Similar data for 1.16 V would require a number of years to obtain, although the current corrosion rate is approximately 0.2 inch/year.
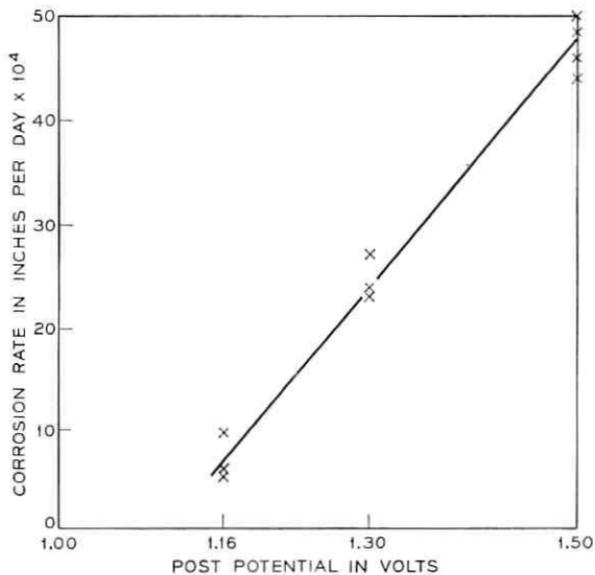
Fig. 5—Corrosion rate as a function of positive post potential.
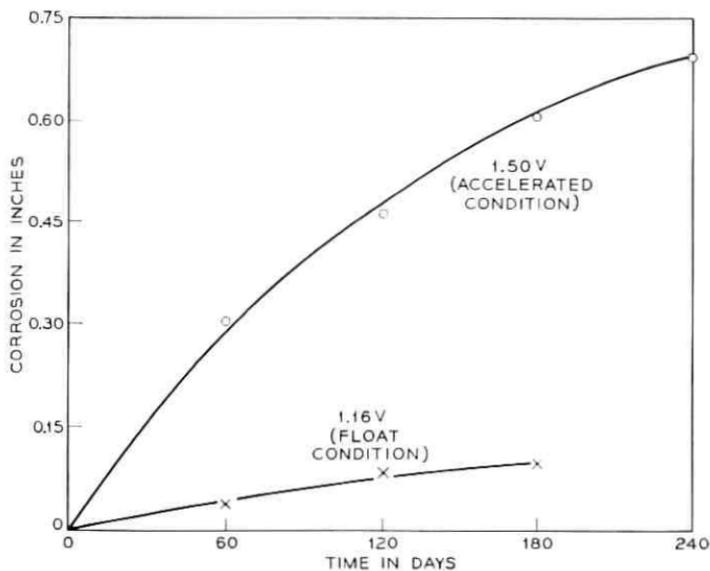
Fig. 6—Corrosion of positive post as a function of time.
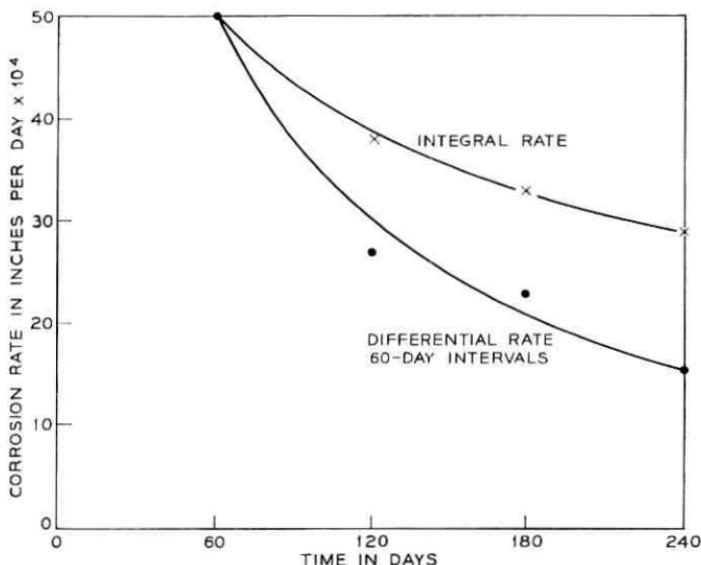
Fig. 7—Corrosion rate of positive post as a function of time (1.50 V).

The decrease in corrosion rate with time is plotted on an integral and differential basis in Fig. 7. It is apparent that the corrosion rate has decreased to 30 percent of its original value from the first (60th day) to the fourth (240th day) 60-day interval. These incremental slope changes contribute to the 40 percent decrease in the integral corrosion rate plot. From the data for the 1.50 V test, it is apparent that the presently observed corrosion rate (0.2 inch/year) at 1.16 V can realistically be expected to decrease to one-half or 0.1 inch/year, which would yield a useful seal life of at least 40 years.

IV. CONCLUSIONS

A novel post-seal has been designed and fabricated for the new Bell System battery. Accelerated tests have been carried out which indicate a useful life for this seal in excess of 40 years. Such a seal life is obtained by:

(i) Utilizing long posts, thus providing long path lengths for corrosion,

(ii) Stifling corrosion effectively by surrounding the posts with a rigid cast epoxy cylinder,

(*iii*) Permitting relative motion of plate stacks and battery jar cover to accommodate stack motion in service, and

(*iv*) Building redundancy into the seal.

V. ACKNOWLEDGMENTS

Lead-Acid Battery:

Techniques for Bonding the Positive Plates

By R. H. CUSHMAN
(Manuscript received April 13, 1970)

Two fusion-bonding processes have been devised for bonding adjacent curved- or flat-lead lugs of the positive plates used in the new Bell Laboratories lead-acid battery. We describe both processes, but emphasize the Mechanical Thermal Pulse (MTP) Continuous Fusion-Bonding Process, which was chosen for use in the prototype battery manufacture. This process is capable of bonding many battery lugs simultaneously with simple equipment and with relatively precise control of bonding parameters.

I. INTRODUCTION

The purpose of this article is to discuss the bonding of the positive plates of the new lead-acid battery to form a complete battery cell. The negative plates are bonded by a center-pour technique described elsewhere in this issue.

Figure 1 shows part of one element of the new battery, using an early prototype positive-plate design. During assembly, negative plates are stacked between positive plates until the element is complete; typically ten or more negatives are used. Each circular positive plate has four or more pure-lead lugs. These must be bonded to the corresponding lugs on the adjacent positive plates.

At the bottom of Fig. 1 are two unpasted positive plates after bonding with a pasted negative plate in position. Note that the battery lugs in this early design were flat to facilitate bonding studies. Figure 2 shows a pair of lugs cut out of the battery plates. This is a later design using curved lugs, approximately 0.188 inch thick and 2.25 inches long and formed out of pure lead. It is desirable to use curved lugs to provide a maximum cross section of the lug, maximum electrical conductance, and simplest overall geometry.

Two new concepts of fusion bonding were developed. Both are to be
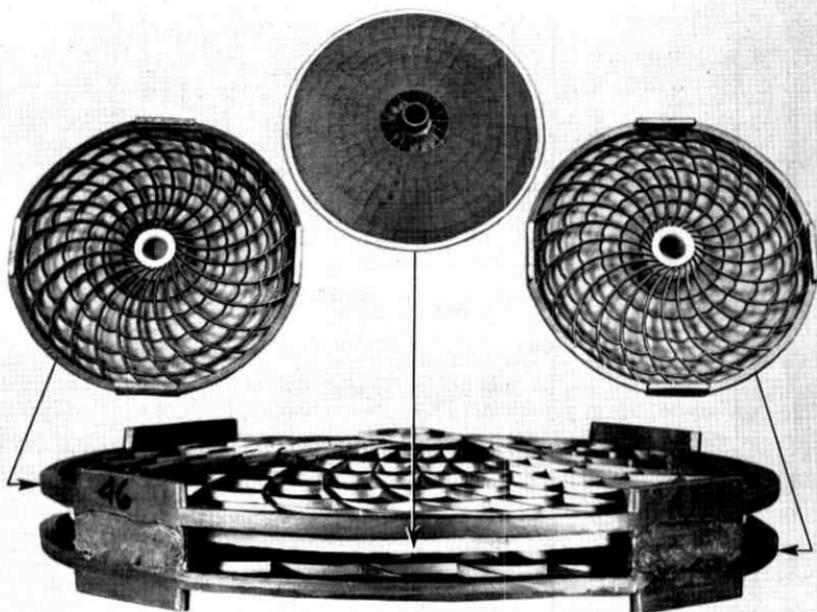
Fig. 1—Exploded and assembly view of three battery plates.

used in manufacture, one for bonding the positive plates and the other for a secondary bonding operation. This article, therefore, considers both processes and discusses bonding of both flat and curved lugs.

II. BONDING REQUIREMENT

The left-hand sketch in Fig. 3 represents a cutaway view of a bond area. Appropriate spacers and insulators are normally interspersed between the negative and positive plates. Note that there is complete access to the outside of the positive-plate lugs in the bond area but extremely limited access to the inside bond area.

A fundamental requirement is that each pair of lugs must be bonded completely without voids or cracks. This requirement is necessitated by the stresses occurring in the lead as a battery is floated. Accelerated life tests have confirmed that the cracks in the bond region tend to widen and deepen, increasing the internal battery resistance. This results in degradation of performance and can eventually progress sufficiently to cause failure.

The first prototype batteries were assembled and bonded with a hand Heliarc welding technique and with the plate axis vertical. This was relatively laborious and time-consuming with typical bonds requiring up to a minute or even more. In addition, considerable difficulty was experienced in achieving crack-free, complete penetration of the lugs.

It was recognized that for faster bonding (*i*) it would be necessary to determine whether a confining, sealing mold could be applied to the present battery geometry to afford reliable containment of the molten lead, and (*ii*) the battery axis should be in a horizontal position so that gravity would help control the molten lead during bonding. An alternate approach would be to attempt to bond the battery plates without the use of any confining, sealing mold or back-up fixture. It will be shown that both approaches were proven feasible but the latter approach proved to be more advantageous.

Several other points in the battery design influenced the choice and design of the bonding method. Due to the unusually long life expectancy for this battery design, contamination of the bond area had to be kept
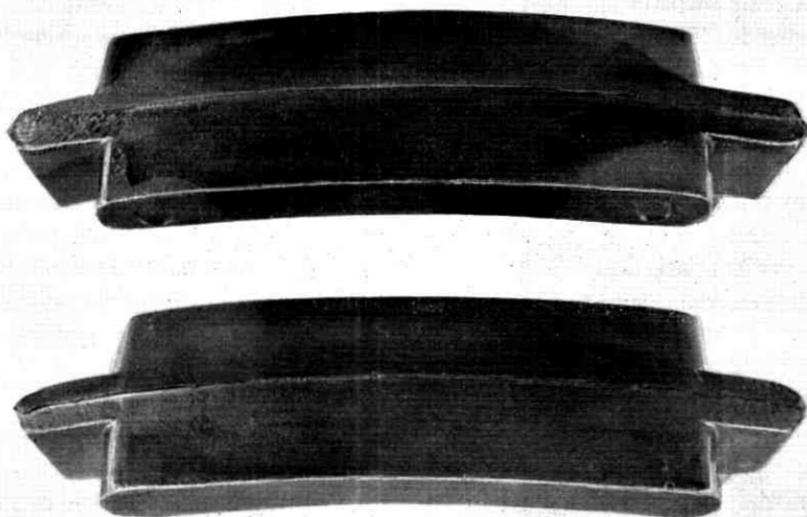
Fig. 2—Two positive plate lugs.
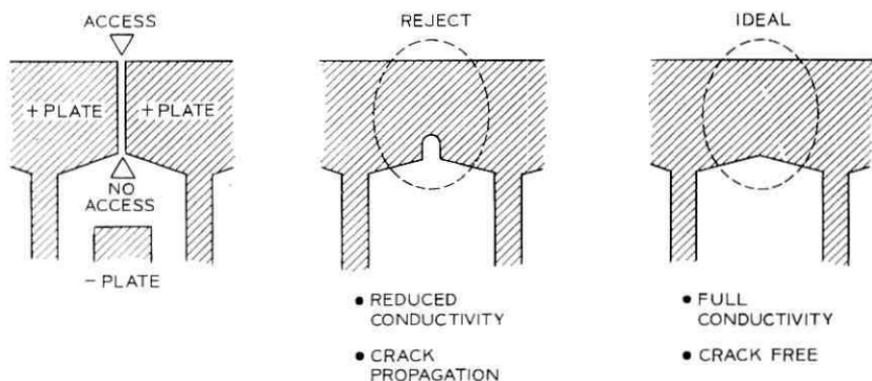
Fig. 3—Positive plate bonding.

to a minimum. In addition, both the magnitude and duration of the peak temperature in the bond area had to be kept to minimum values to prevent thermal damage to the pasted area of the plates, the separators or the insulators.

III. BONDING METHODS

Groups within the Bell System investigated several different bonding methods. The processes included: Heliarc welding, infrared bonding, laser bonding and mechanical thermal pulse or MTP bonding. Table I compares bonding process characteristics. Note that the target temperature, the melting point of pure lead, is approximately 327°C. It can be seen that the MTP processes have the lowest peak temperatures, yet they possess very high heating rates. This has proven useful in minimizing variations in bond temperature. In general, with a given work part and its typical variation in characteristics, the higher the ratio of the effective source temperature to the target temperature of the work part, the more crucial automatic temperature control becomes. In this application, conventional feedback techniques for controlling were not adequate, so a new approach was developed.

Figure 4 illustrates the general MTP principles involved. Thermal energy is stored in an appropriately shaped metal block or ram. This thermal energy is transferred by thermal conduction to the two or more work parts to be bonded. Sufficient pressure is applied between the hot ram and the work parts to assure low thermal resistance values at the interface, making it possible to transfer rapidly relatively large quantities of thermal energy. The transfer is relatively independent of variations

in the characteristics of the work parts. Note that the peak temperature of the ram, and thus of the work parts in the bond area, has a precise, predetermined upper limit—that is, the control temperature of the ram itself. The ram can either be pulse or continuously heated to the desired temperature.

Typical variables in work parts tend to have a strong effect on the thermal energy transferred, but the effect is only minor in the MTP processes. Such variables include emissivity, reflectivity, color, surface roughness, electrical conductivity, volume thermal conductivity, and surface contamination.

IV. MTP HEATED RAM PROCESSES

The MTP fusion-bonding process that makes use of a heated ram is extremely simple in concept. Figure 5 illustrates the basic principles. A carefully shaped, relatively large ram of a material with high thermal conductivity and with high heat capacity, such as copper, is preheated

TABLE I—COMPARISON OF BONDING PROCESSES

Process	Effective Source Temperature	Energy Transfer Method	Energy Transfer Magnitude	Temperature Control Method
Mechanical Thermal Pulse (MTP)	800 °C*	Conduction	Heated ram transfers fixed amount of energy	Determined by ram or tip temperature
Radiant (Incandescent)	3400 °C	Radiation		
Radiant (Arc)	5700 °C	Radiation		
Oxy-Acetylene	3200 °C	Forced convection	Feedback varies energy transfer rate or duration	Determined by total energy transferred
Heliarc	5700 °C	Joule heating plus ion bombardment		
Atomic hydrogen arc	10,000 °C	Joule heating plus ion bombardment		
Laser	~10^8†	Radiation		

* Maximum ram or tip temperature.
† Based on solving the Stefan-Boltzmann equation.
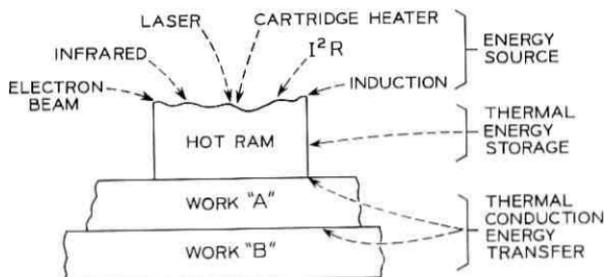
Fig. 4—Principles of processes of bonding with the mechanical thermal pulse (MTP) method.

to a specified temperature. The ram is positioned over the bonding fixture and then plunged with a low force against the two lugs to be bonded contained in a confining back-up fixture.

The thermal conductivity of copper is approximately 11 times greater than that of lead, and copper's heat capacity per unit volume is 2.35 times greater than that of lead. Consequently, the heated ram stores sufficient thermal energy to form a bond between the two lead lugs with a relatively low temperature drop and rapid recovery. Typically, bonds are formed in one to ten seconds. It is possible to combine two or more rams to create two or more bonds simultaneously.

The ram stroke is controlled so that the ram first melts the surface of the lead lugs. Then the ram progressively penetrates to the inside bottom of the lugs as melting continues. The penetration can be readily controlled so that the ram either penetrates completely to the inside bottom of the lug, or to within a predetermined distance from the back of the lug. Consequently, the heated ram offers positive assurance that the thermal energy completely penetrates the entire cross section of the lug.

It is appropriate at this point to compare the MTP heated ram process with alternative bonding techniques. During bond formation between the two lead lugs, the solid, room-temperature lead is heated to temperatures beyond its fusion temperature and changes from a solid to a liquid. Consequently, the electrical, thermal, optical, mechanical and metallurgical characteristics of the lead change appreciably throughout the formation of the bond. Specifically, changes may occur in electrical resistance, emissivity, density, thermal conductivity, ductility, and so on. These changes in characteristics must be taken into account and appropriately compensated for if accurate temperature control is to be maintained. Since the MTP heated ram completely penetrates the lead

lugs, transfer of the thermal energy to the work parts depends primarily upon the relatively unchanging thermal characteristics of the heated ram itself. In addition, since the ram material has a relatively high thermal conductivity, the ram penetration through the center of the bond area assures a minimum temperature differential between the outer and inner surfaces of the bond area.

The shape of the ram is chosen to provide a storage space for the molten lead that is displaced by the volume of the heated ram. Tests have confirmed that this molten lead can be stored in a hollowed-out center section of the ram. After the initial surface melting, the ram starts to penetrate the two lugs, forcing the displaced molten lead into the hollow center. Very little lead, if any, flows outside the confines of the ram itself. As the ram is removed, the molten lead flows back into place to form the bond. The copper ram is plated with approximately one mil of nickel to protect against oxidation and minor abrasion.

Lead oxidizes quite readily and the oxide is an excellent thermal insulator, even in relatively thin layers. Initially, the oxide coating adheres poorly to the surface of the ram. Consequently, with this method it is desirable to wipe the ram with a metal bristle brush after each bond or set of bonds to remove this loosely adhering coating from the ram. The ram can be rapidly and reliably cleaned by a single in-line, one-stroke motion. Appropriate gas jets, protective atmospheres or other techniques might be considered as alternative cleaning methods.

Calculations indicate that from 1000 to 5000 joules are required to bond each pair of lead lugs. The range of energy shown encompasses different bonding speeds. At the most rapid rates investigated—that is, about one second bonding time—minimum thermal energy is required. At easier-to-control, more practical bonding times of from three to seven seconds, the larger quantity of thermal energy is required since a larger
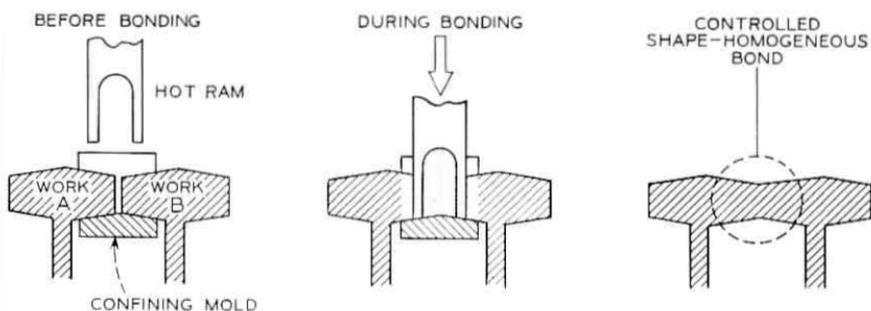
Fig. 5—Fusion bonding with an MTP heated ram.
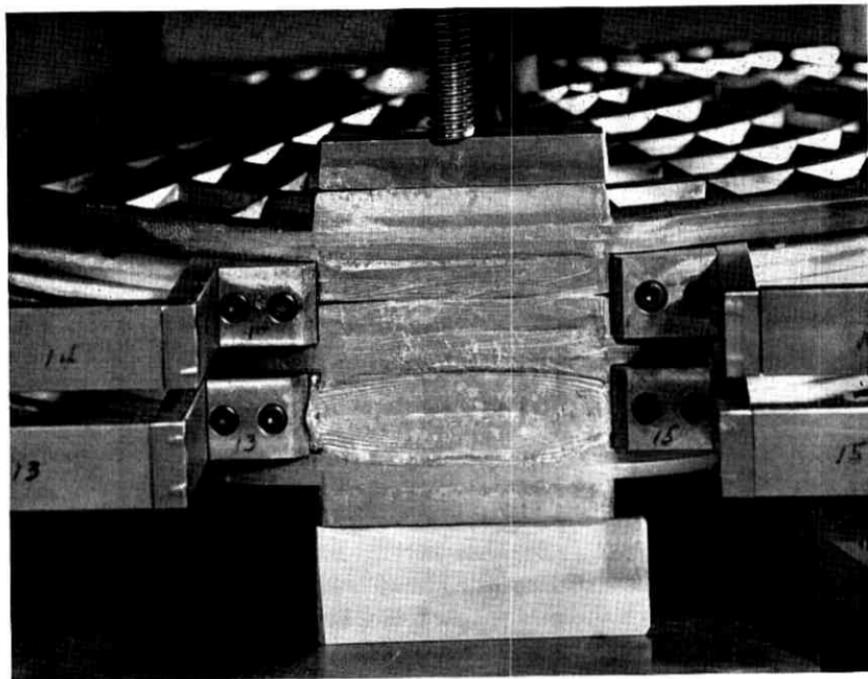
Fig. 6—Positive plate bonding fixture.

volume of the lead work parts absorbs appreciable thermal energy. The optimum time at the present stage of development is approximately five seconds.

A major advantage of the MTP heated ram fusion process is its capability of transferring this relatively sizable quantity of thermal energy rapidly, repeatably and with minimum variation in the peak temperature of the work parts. The MTP ram temperature is typically a maximum of 600 to 700°C; thus the lead temperature is limited to this maximum value. The relatively low temperatures used in MTP fusion processes are beneficial in reducing metallurgical changes in the lead structure, reducing turbulence and vaporization, and eliminating boiling in the bond area.

The forces required to create these bonds are surprisingly low. Typically they range from three to five pounds for a bond length of approximately 2.25 inches.

Figure 6 is a close-up view of a typical back-up fixture as viewed from the top, looking down from the ram's position. Note that two lugs have been bonded and that the adjacent two lugs are in position ready for

bonding. The fixture slides in from both sides and mates to form a containing seal against loss of the molten lead. Figure 7 shows a front and back view of a typical bond along with two metallurgical sections confirming that complete fusion bonds have indeed been achieved.

Tests were made with a chromel-alumel thermocouple bonded to the bottom area of the ram tip. Storage oscilloscope traces of the ram temperature drop during a typical bonding cycle of actual battery lugs are shown in Fig. 8. This chart confirms that the tip temperature initially dropped rapidly to approximately the melting point of lead (327°C) and then more slowly rose to approximately 455°C at the end of six seconds. The quantity of thermal energy required to melt the lead is approximately 40 percent of the energy required to raise the lead from room temperature to 455°C. Consequently, the temperature of the lead rises rapidly to its melting point, then tends to remain there until all the lead in immediate contact with the ram becomes molten. Only then does the temperature of the lead start to rise above its melting point. In addition, the relatively small differential between the temperature of the ram and the melting point of lead helps assure that the molten
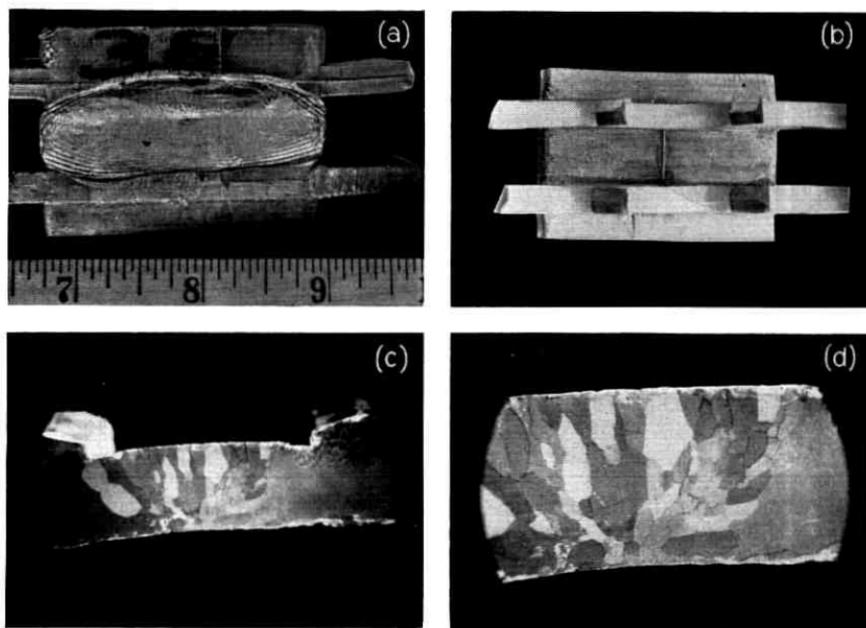
Fig. 7—(a and b) Front and back views of an early test bond; (c and d) 3× and 6× views of the bond area.
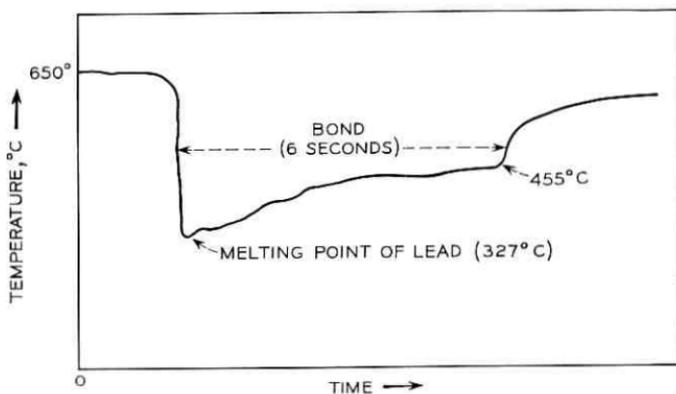
Fig. 8—Oscilloscope trace of ram surface temperature during typical bond.

lead will be heated only to a moderate temperature slightly above its melting point throughout the entire bonding cycle.

Tests verify that 900 to 1000 watts of electrical power are adequate to heat the MTP ram. The ram should be protected by suitable high-temperature thermal insulation to reduce thermal loss and to provide more rapid recovery of temperature between bonds.

The MTP heated ram lead-bonding process is practical for applications where it is possible to incorporate a simple back-up or confining fixture to contain the molten lead during bonding. It has been shown feasible to make excellent bonds with either:

(i) narrow back-up fixtures of metal,
(ii) resilient back-up fixtures of high-temperature plastic or foam such as used for the battery separators or insulators, or
(iii) thin high-temperature glass or plastic, adhesive-coated tape.

Back-up fixture types ii and iii can be left in the battery or can be removed after the bonding is completed. Metal back-up fixtures naturally have to be removed after the bond is completed.

This heated ram process uses simple, low-cost equipment; is capable of forming many bonds simultaneously; and is easy to control and reliable. However, it is clearly best suited for use where free access to the work parts is more readily available. For that reason, it is currently under consideration for use in bonding the lead strap to the periphery of the positive battery plates to create an electrical connection for the subsequent forming process for the lead paste.

V. CONTINUOUS MTP FUSION BONDING PROCESS

5.1 *Principles*

Figure 9 is an artist's rendition of the MTP continuous fusion bonding process. No back-up mold is required. In the heated ram process, the bond is made with the battery axis horizontal and the two lugs to be bonded located at the *top center* of the battery. In the continuous bonding process the battery axis is still horizontal but the lugs are bonded at the *bottom center* of the battery. In this way, gravity helps contain the molten lead on the *inside* (top) of the two lugs being bonded. The molten lead on the outside (bottom) of the two lugs being bonded is now directly contained by the bonding fixture itself. As a consequence, only the heated bonding tip need project within the confines of the circumference of the battery. This markedly simplifies the fixturing and stacking arrangements for the numerous battery plates, separators, and insulators required for each battery cell.

The battery is bonded from the bottom with a carefully shaped heated tip that projects up through a supporting plane. The plane is preferably curved to match the outside contour of the battery lugs. The battery lugs and the tip are moved relative to one another so that the tip progressively is passed along the entire length of the two lugs to be bonded. The lugs melt in the immediate hot-tip area only and quickly solidify as the tip passes by. The molten lead is contained automatically by the supporting plane on the bottom and by the non-molten portions of the lead parts adjacent to the bond area. Typically a time of from 15 to 25 seconds is required to bond two battery lugs using present equipment.

5.2 *Mechanical Design*

To describe the mechanical design of the tip unit, it is convenient to consider first the tip itself, then the tip assembly and the drive mechanism. A view of the tip is given in Fig. 10. The size of the projecting, active portion of the tip is approximately 0.125 inch by 0.2 inch in diameter. The tip is designed to project approximately two-thirds of the way through the parts to be bonded. If the tip projects all the way through the lead parts, there is a tendency to slow down the flow of the molten lead into the cavity left by the tip as it passes by.

The general requirements for an alloy to be used for the tip are that it should be:

(*i*) resistant to high-temperature corrosion,

(*ii*) resistant to molten lead,
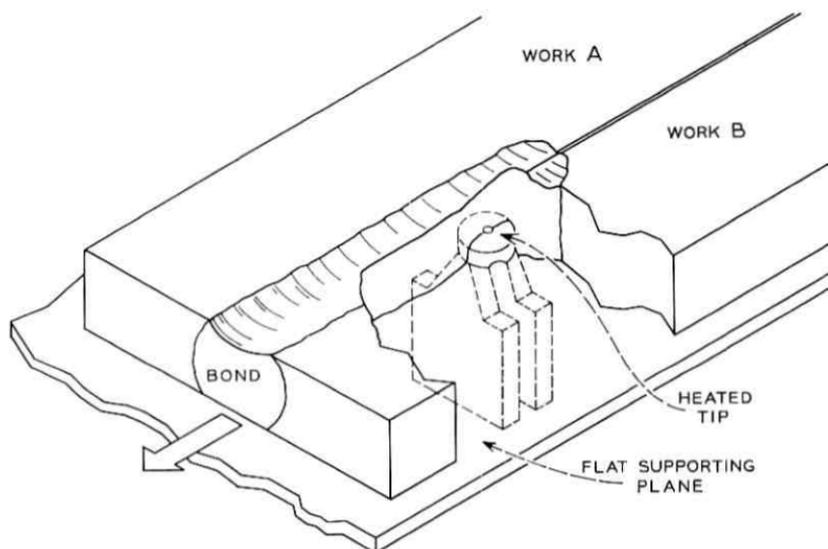
Fig. 9—MTP continuous-fusion bonding process.
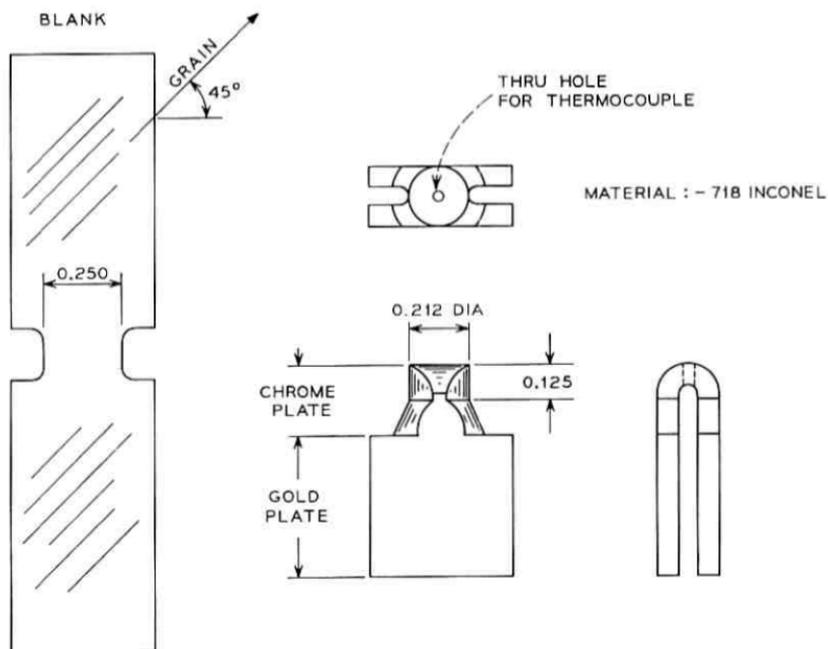
Fig. 10—Details of the heated tip.

(*iii*) formable at room temperature,

(*iv*) machinable with standard tools, and

(*v*) low in cost.

Also it should possess:

(*i*) high compressive strengths at high temperatures, and

(*ii*) high electrical resistivity.

Several high-temperature resistant metal alloys were tested for the heater tip, including Kanthal and Nichrome. The most success in bonding experiments, however, was achieved with Inconel 718 alloy shown in Table II. The tips are stamped out of a flat sheet, thereby reducing their cost to a minimum. The tips can be bent into the final shape at room temperature if the tip is oriented approximately 45° to the rolling direction of the flat sheet. For maximum tip life, the tip temperature during bonding should be kept to as low a value as possible. The reason is that the relationship of tip life versus temperature is exponential. To increase tip life further, the tips are plated with 0.5 to 1 mil of chromium on the active area.

The temperature in the center of the active tip area is continuously sensed with a conventional thermocouple and displayed for the operator. In our laboratory test, we used 20-mil diameter chromel-alumel thermocouple wire, glass-insulated, for these measurements.

Figure 11 shows a section view of the mounting assembly. This represents the heart of the bonding process. First, it provides a reliable connection of low electrical resistance between the power supply welding cables and the heated tip. The electrical resistance of this connection is stable in spite of appreciable temperature excursions and the resultant mechanical stresses. Secondly, the assembly design is modular in concept, allowing for the mounting of a complete assembly within the center-to-center

TABLE II—INCONEL 718 ALLOY, LIMITING CHEMICAL
COMPOSITION IN PERCENT[a]

Nickel (plus Cobalt)	50.00–55.00	Cobalt	1.00 max.
Chromium	17.00–21.00	Carbon	0.08 max.
Iron	Bal.	Manganese	0.35 max.
Columbium		Silicon	0.35 max.
(plus Tantalum)	4.75–5.50	Phosphorus	0.015 max.
Molybdenum	2.80–3.30	Sulfur	0.015 max.
Titanium	0.65–1.15	Boron	0.006 max.
Aluminum	0.20–0.80	Copper	0.30 max.

[a] Conforms to AMS specifications.
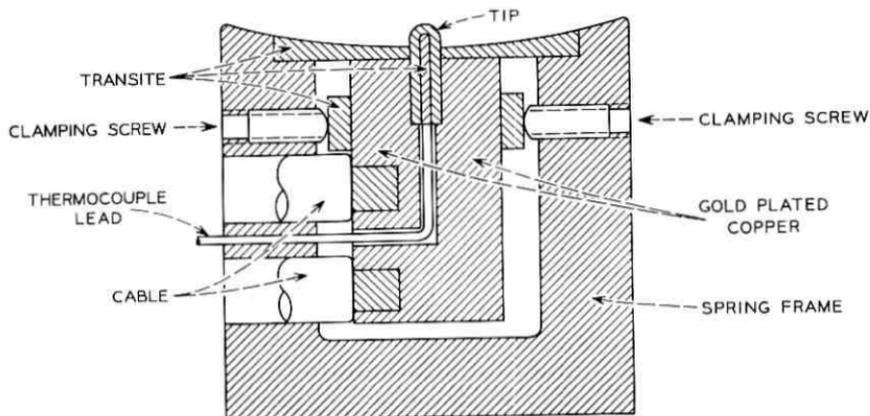
Fig. 11—Section view of the heated tip mounting assembly.

distance of adjacent bonds. Its attachment to the overall bonding machine is simple, allowing rapid and easy replacement.

The stable electrical resistance is achieved by using a compliant outer metal frame to compensate for the relatively large dimensional changes created by the temperature excursions. The frame is designed in the form of a U-shaped spring. The relatively large gold-plated copper blocks are brazed with a silver alloy to the stranded welding cable and are used to make electrical connection to the heated tip. The tip has a close-fitting, thin sheet of high-temperature transite insulation placed in its center around the previously inserted thermocouple wires. The two 0.5-inch square terminal ends of the tip are gold plated to assure minimum electrical interface resistance. The gold-plated copper electrodes are then pressed against the two contact faces of the tip by means of insulated screws inserted in opposite faces of the U-shaped spring frame. The spring frame is designed to flex during operation, thus maintaining a heavy clamping force between the copper electrodes and the heated tip at all times.

The force required to push the tip through the lead has been measured. When the support for the tip does not touch the lead, this force is typically less than ten pounds. When the support for the tip bears relatively firmly against the outside of the lead lugs, however, the force can approach an upper limit of approximately 100 pounds. Depending, then, on the degree of precision of the alignment between the tip support structure and the circumference of the battery lugs, the force required to push the tip through the lugs can vary between 10 and 100 pounds during one bond. Since the bonding speed should be controlled accurately,

this necessitates a motor-drive system that can withstand this variation with minimal change in speed. A constant-speed motor with a large gear reduction ratio worked quite well. The majority of tests were made with motor speeds yielding either 15 to 18 seconds or 26 to 30 seconds per bond.

We have made successful bonds using both (*i*) preset and fixed tip alignment and, (*ii*) spring-loaded, self-compensating tip alignment.

The majority of tests were made with the preset tip alignment, since it simplified the mechanical design. Fixed alignment does, however, require that the shape of the battery plates be accurately controlled. Tests have confirmed that either a compensating or spring-loaded type of tip support structure can be used. The choice between a fixed or compensating design depends largely upon the precision to be expected in the manufacture of the battery plates.

5.3 *Electrical Design*

During the course of the experimental work, it became apparent that it would be advantageous to have two power supplies, one for current to heat the tip and one to supply current flowing through the lead. To understand why, consider the sketches of Fig. 12, showing the heated tip under three conditions. In each of the three sketches, the four points of interest in the following discussion are identified as A, B, C and D. In the idle condition occurring between bonds (left), the tip can readily be heated to a uniform temperature as shown. With just one power supply to heat the tip, no current flows in the lead, and it can be seen in the center sketch that the points of physical contact (B and C) between the heated tip and the work parts have dropped
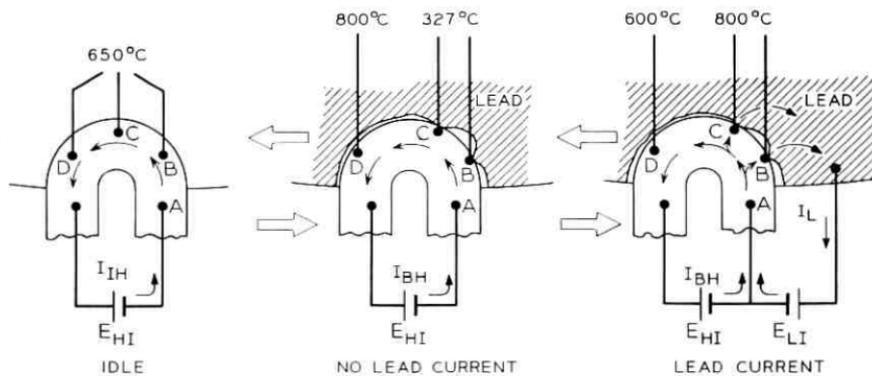
Fig. 12—Heater tip temperature profiles.

to a much lower temperature than the balance of the heated tip. This occurs rapidly and is, of course, exactly the reverse of what is desired. The right-hand sketch of Fig. 12 illustrates what happens when "lead current" is applied by means of a second power supply. This additional current through the tip to the lead lugs serves the useful function of heating the points of contact between the heated tip and the work to a higher temperature than the balance of the tip. This is precisely what is needed. It is interesting to note that as the points of contact (usually two or more) between the heated tip and the lead lugs change in location, the old points of contact that were disrupted automatically cool down and the new points of contact automatically heat up.

Figure 13 is a schematic diagram of the original laboratory power supply for the MTP continuous fusion bonder. The area of most interest
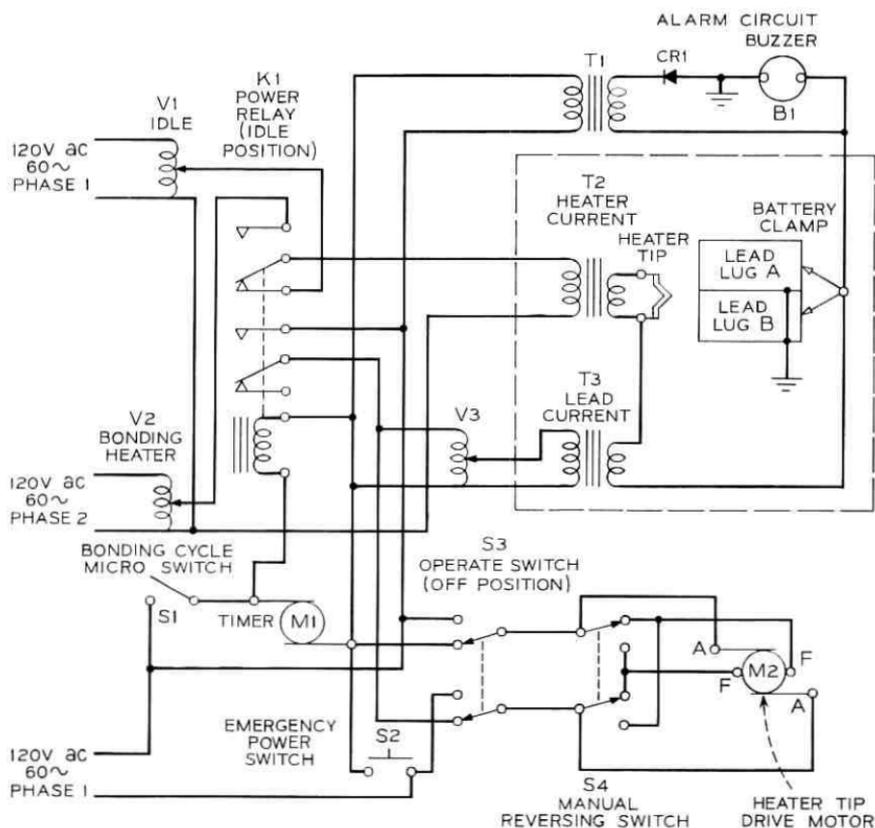
Fig. 13—Schematic diagram of laboratory power supply for MTP continuous-fusion bonder.

is identified by the dashed line. Two step-down transformers, T-2 and T-3, are used to provide the relatively large bonding currents. Both power transformers are, in turn, fed by autotransformers to provide a convenient means of changing bonding currents. Transformer T-2 provides the current for heating the hot tip directly and hence is called the heater-current power supply. Transformer T-3 passes current through the heated tip and thence to the lead battery lugs to be bonded. This second power supply is called the lead-current power supply. As discussed above, the function of this current is to heat the tip hotter in the areas of contact between the tip and the lead battery lugs. The passage of this electrical current through the tip to the lead work parts creates only minimal heating of the lead itself due to the relatively high electrical resistance of the heated tip compared to that of the lead.

It is necessary to provide electrical connection from the lead current power supply to the battery lead lugs. Separate lead current clamps and cables are used for each bond when simultaneous multiple bonds are being made. In this manner the power supply for each heater tip is isolated except for one common point, thus preventing interaction between the different power supplies. Either a simple alarm circuit (step-down transformer T-1 in Figure 13) should be provided to assure that connection to the lead current power supply has been made, or as an alternative, the lead clamp should be locked in position automatically as the heater tip is raised into position ready for bonding.

Since the two power supplies have a common resistance consisting of a portion of the heater tip and a predetermined portion of the total length of the supply cable*, the current flowing in one circuit does affect the current flowing in the other. This coupling can be adjusted to be either aiding or cancelling. Negative coupling is preferred, since this provides maximum heater current for the tip when the lead current is not flowing, and decreases the heater current when the lead current is flowing. In Fig. 12 I_{IH} is the idle heater current, I_{BH} is the bonding heater current, and I_L is the lead current. Note in Fig. 12 that the two ac power supplies are shown as dc batteries. When they are connected to provide negative coupling (that is, with the polarities shown), then the common-current path in the heater tip is limited to the region from points A to C, with most of the current concentrated between points A and B. It can be seen from the center and right-hand views that adding current I_L will reduce the magnitude of current I_{BH} while simultaneously increasing the current flow in regions surrounding points B and C.

---

* Approximately 5 inches in our test using #4 AWG cables approximately 43 inches in length.
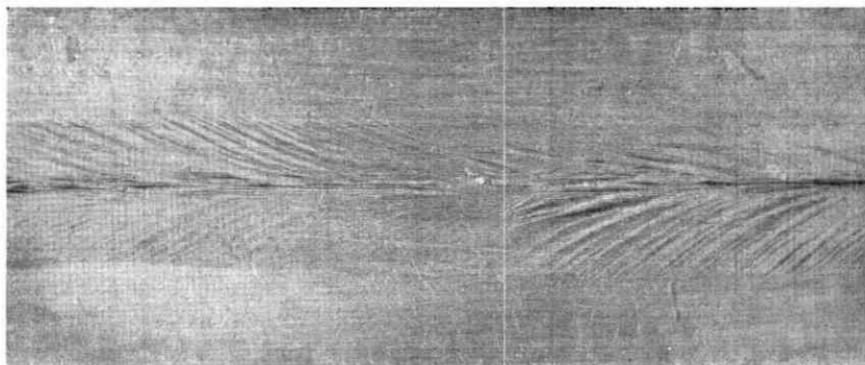
Fig. 14—Top surface of two bonded flat-lead strips.

This increased current generates additional heat energy by joule heating at points B and C. The thermal energy is then transferred to the lead by thermal conduction to create uniform, void-free bonds.

Early experiments using a heated tip of similar design, with the thermocouple mounted in the center of the tip and coupled to a high-speed potentiometric controller, afforded convincing evidence that conventional feedback temperature control did not adequately satisfy this application. For example, if the thermocouple area of the tip was not contacting the lead, the controller would sense a high temperature and would not feed additional power to the tip, even though the actual points of contact between the tip and the lead could be below the melting point of lead. The resultant errors were many hundreds of degrees centigrade and occasionally caused burn-out of the tip.

In summary, two power supplies connected for negative coupling are used to provide the equivalent of a multi-point automatic temperature control system which adequately compensates for random changing points of contact between the tip and the lead work part with minimal electronic circuitry. The idle heater current (see Fig. 13) functions to keep the bonding tip at a moderate temperature between bonds. It can be eliminated if the tip is allowed to cool back to room temperature between bonds and then pulse-heated to bonding temperature just prior to contacting the lugs.

In our initial test, 60-cycle 120 V ac power was used. Two circuits are required having a 120° phase relationship. This is necessary because if both the lead current and the heater current are in phase, the heating rate of the tip is so rapid (typically 500°C/second) that there would be appreciable temperature swings between successive half cycles of

current. The improvement in tip temperature control and life is pronounced when a deliberate out-of-phase condition between the heater current and lead current power supplies is provided.

5.4 *Performance of the Experimental Bonder*

Figure 14 shows the top surface of two flat lead strips, having the same thickness as the battery lugs, that were bonded with this bonding process. Lead oxide was left in place on the lead parts prior to and during bonding. Both lead strips rapidly melted in the center, flowing together while the lead oxide floated to the top as shown. Many test bonds have been made with relatively heavy layers of lead oxide present in the bond area. Reasonable amounts of lead oxide can be floated out in the molten stage to provide good bonds. For manufacture, however, the forming process for the positive plates provides an excessively heavy layer of dark oxide, and it is not deemed practical to attempt to bond through this layer of oxide, since its depth and character are not specified or controlled. Hence, a mechanical or chemical pre-cleaning of the bond area is recommended. Preferably this pre-cleaning should be done as close as possible to the time the bond is to be formed. This is not unduly critical, however, since good bonds can readily be made to plates that have been pre-cleaned several days prior to bonding.

Figure 15 shows front and back views of three typical bonds. These

FRONT   BACK
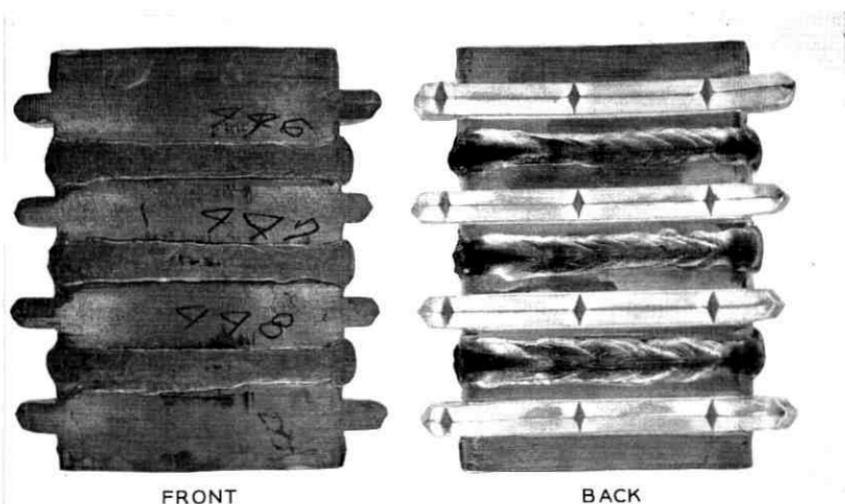
Fig. 15—Photographs of three bonded lugs.

lugs were cut out of complete battery plates for the purpose of the photographs.

The start and end of the bonds were not problems using a constant-speed tip drive with flat lugs. With curved lugs the battery plates are kept stationary and the tip assembly is moved from one end of the lugs to the other. Thus gravity tends to cause the molten lead to fall back into place at both the beginning and end of bonding. This has proved to be quite helpful. The tip is advanced at a constant speed prior to, during, and after bonding. In some cases the additional sophistication of a variable speed, programmed, motor-driven system might be worth considering.

The left- and right-hand sketches in Fig. 16 show a slight projection at the start and end of the bond. The shape of the lugs prior to bonding is shown by the dashed lines. Note the small raised pool of molten metal directly above the leading edge of the tip as it progresses through the bond. As the tip emerges from the bond, a very small quantity of lead, typically one-half the volume of the exposed tip, is pulled out of the bond by the advancing tip. These slight distortions of bond shape are probably minor enough to be ignored. Such distortions can be equalized, however, by tilting the lugs off center so that the leading edge of the lug is slightly higher than the trailing edge.

Figure 17 shows two metallurgical views of sections of bonds. These samples were not etched but were very carefully polished to minimize smearing. Figure 17a shows a view of a typical bond made under the following conditions:

    (*i*) 280 amps RMS lead current,
    (*ii*) 295 amps RMS heater current,
    (*iii*) 830°C bonding temperature,
    (*iv*) 480°C idle, and
    (*v*) 28 seconds bonding time.
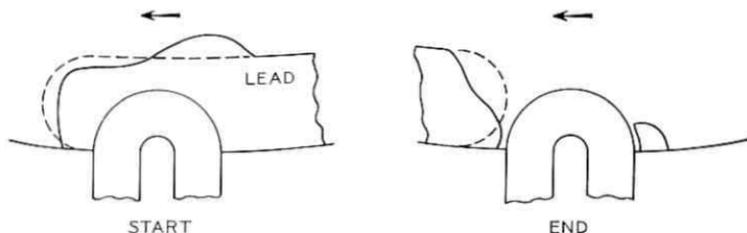
Fig. 16—Bond shape at the beginning and end of a lug.
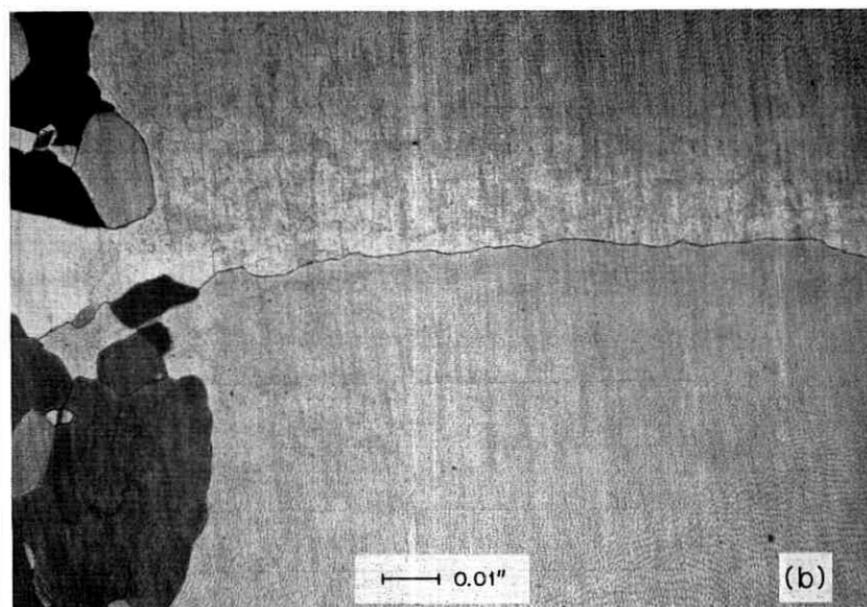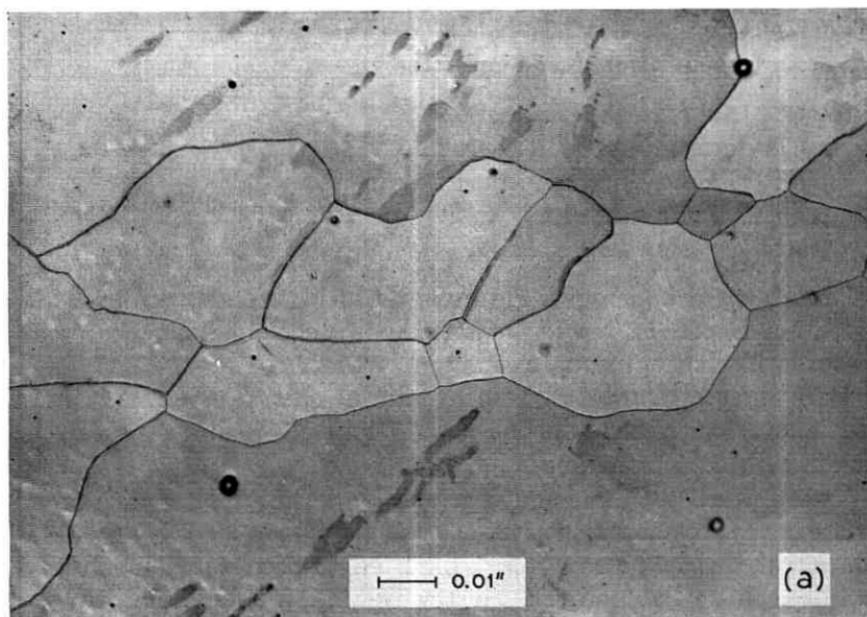
Fig. 17—Metallurgical views of typical bonds. (0.4″ actual = 0.01″).

Figure 17b shows a view of a pair of lugs that were bonded three times. These three bonds were made under the following conditions:

(i) 245 amps RMS lead current,
(ii) 260 amps RMS heater current,
(iii) 1030°C bonding temperature,
(iv) 720°C idle, and
(v) 17 seconds bonding time.

The repeated bonding of the same set of lugs clearly confirms that these bonds can be readily repaired. Also, such metallurgical views indicate that homogeneous, clean bonds are formed.

It requires approximately 6000 joules of energy to form an MTP continuous fusion bond to one pair of lugs 0.188 inch thick by approximately 2.25 inches long.

Two test cells of three and four positive battery plates with appropriate separators and insulators with all the bonds made by the MTP continuous fusion process were submitted to accelerated life tests at Bell Telephone Laboratories, Murray Hill, N. J. In addition, two test cells bonded by the MTP heated ram process underwent accelerated life tests. All four of these first test cells withstood the equivalent of several centuries of normal operation without weld failure.

Analyses of bonded samples were also made to determine if there were traces of nickel or chromium, the major metals involved in the tip. Tests carried out by L. D. Babusci of Bell Telephone Laboratories indicate only trace amounts (below 0.001 percent) of these materials in the bonds.

Figure 18 shows an oscilloscope trace of the temperature sensed at the center of the tip during a typical bonding cycle. There is an ap-
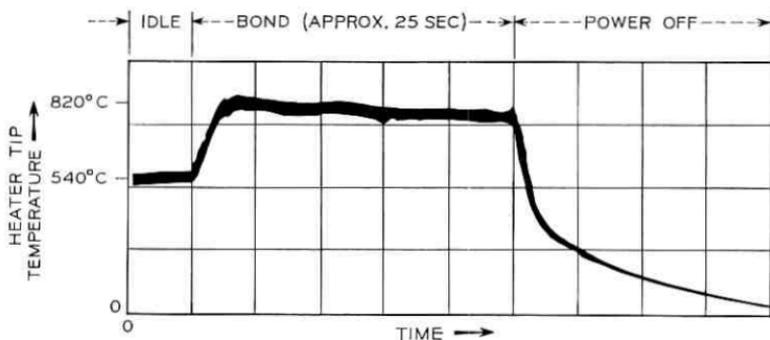
Fig. 18—Oscilloscope trace of tip temperature during typical bond.
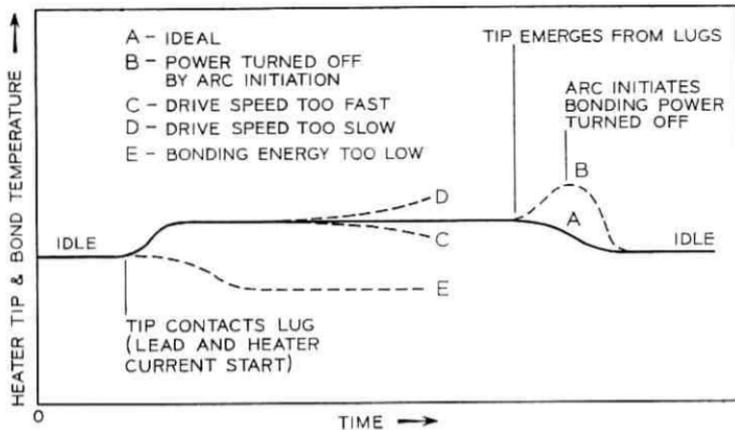
Fig. 19—Variations in tip and bond temperature caused by different power settings and bonding speeds.

preciable temperature differential between the tip and the lead. Consequently, this temperature represents the probable upper limit for the temperature of the lead during the bond. It should, therefore, serve as a convenient dynamic indication of bond quality since it is sensitive to variations in both the bonding equipment and the lead lugs themselves. The temperature during bonding can be displayed on an oscilloscope or a strip chart recorder, or it can be incorporated in a comparison circuit yielding simple "go-no go" type control.

Next, the effects of the different process controls on bond formation is discussed. Tests have indicated that when the idle tip temperature is varied, the variation primarily affects the *beginning* of the bond. Changing the heater current, however, primarily affects the outside or *bottom* of the bond. Changing the lead current, in turn, primarily affects the inside or *top* of the bond. And finally, varying the speed of the bond tends to control the *end* of the bond. Consequently, independent control can be exercised over each of the four surfaces of the bond.

Figure 19 depicts the variations in tip and bond temperature caused by different power settings and bonding speeds. Curve A represents the optimum. It shows a gradual rise from the idle temperature to bonding temperature, with a stable plateau throughout bonding followed by a gradual drop-off at the end of the bond. Curve B confirms the sharp rise in temperature at the end of the bond that results if current is maintained as the tip emerges from the back end of the lug. Curve C shows that if the drive speed is too fast, bonding temperature will start to droop, causing the back end of the bond to become too cold. Con-

versely, Curve D shows the reverse situation when the drive speed is too slow. Finally, Curve E shows the situation when the heater current and lead current settings provide too little power, thus allowing the bonding temperature to drop too low in value.

5.5 *Operation*

The operation sequence is as follows for either individual or multiple bonds:

(*i*)   Each heater tip is adjusted to the desired idle temperature.

(*ii*)   The operate switch (S3 of Fig. 13) is depressed and held depressed throughout the entire bonding process. Once this button is released, the motor is stopped and tip temperatures automatically drop back to the idle position as a "fail safe" mode of operation.

(*iii*) The motor is started, advancing the tip. Once the leading edge of the tip touches the lead lugs, the voltage across each heater tip is switched from the "idle" setting to the higher "bonding" setting by the microswitch S1.

(*iv*)   As the center of the tip lines up with the trailing end of the lug, both the heater and lead currents are turned off by the microswitch S1, switching back to the idle tip heater supply. The motor continues to operate until the tip has advanced at least 0.25 inch past the back end of the lug.

(*v*)   The bonds are complete at this point. To prepare for the next set of bonds, the tip is lowered out of contact with the lugs and is retraced to its starting position. The tip is brushed lightly with a brass bristle brush to remove any lead particles adhering to the tip after each bond. It is then either slid over axially until it is in line for the next adjacent lug, or the battery assembly fixture is rotated an appropriate number of degrees to line up the next row of lugs.

Table III lists typical currents, voltages and temperatures measured at appropriate points in the control circuit for two speed settings. It should be noted that the design of the heater tip influences these values. The listed parameters represent the approximate rms values of voltages and currents as measured or calculated, using the laboratory prototype curved lug bonding fixture with 120 volts ac input to the autotransformers. The two power supplies were connected to provide negative coupling with a 120° phase difference.

In the first prototype bonders, a microswitch was used to turn the heater and lead current voltages on and off for all tips at the beginning and end of each bond. An alternative technique that appears to offer

TABLE III—TYPICAL BONDING PARAMETERS

	Idle		During Bond	
	15–17 Sec. Bonding Time	28–30 Sec. Bonding Time	15–17 Sec. Bonding Time	28–30 Sec. Bonding Time
Heater current power supply				
Step down transformer, T2 — Secondary current	165A	130A	315A	290A
Step down transformer, T2 — Secondary voltage (Open circuit)	1.2V	0.9V	2.2V	2.0V
Lead current power supply				
Step down transformer, T3 — Secondary current	0	0	330A	280A
Step down transformer, T3 — Secondary voltage (Open circuit)	0	0	2.3V	2.0V
Bonding temperature At center of heater tip	760°C	540°C	930°C	820°C

more promise for production applications is individual sensing of the beginning of each lug for each pair of lugs being bonded. This can be done by turning on the lead current before the tip first contacts the lugs. Then, when the tip contacts the lug, the initiation of lead current is sensed with an appropriate transformer, using one lead current cable as a single turn primary. The output of the transformer is used directly to power a relay which turns on the heater current power supply. Standard power relays carry out this function within 25 milliseconds. At a total bonding time of 25 seconds, this represents a tip travel of 0.025 inch and yields higher precision than when microswitches are used.

Turning off the power after the bond is completed is slightly more complex. First, turn-off should be independent of motor speed so as to provide a reasonable tolerance for variations in motor speed under changing load. Second, turn-off should ideally occur at the point when the hot tip has emerged approximately one-half way out from the back end of the lug. The reason is that the tip, as it emerges from the lug, pulls a narrow strand of molten lead with it. Consequently, the path of the lead current from the tip back to the lugs increases in resistance and applies an excessively large quantity of thermal energy to the back end of the bond. This is further aggravated as the tip continues to move, rupturing the now thin strand of molten lead and creating an arc which adds still additional thermal energy.

The optimum method is to turn off both the lead current and heater current upon sensing a predetermined distance of tip travel after initiation of lead current. This can be done with stepper motors or stepper relays, one for each tip, fed by a single electromechanical pulse generating switch operated by the tip drive shaft. All the units would be reset automatically at the end of each bonding cycle. This technique has the advantage that the individual positive plates can be stacked together with less precise alignment without affecting the quality of the bond. In addition, the operator has the freedom of changing the bonding speed without having to readjust the on-and-off mechanisms.

The MTP continuous fusion bonding process can form successful bonds between lugs which are quite severely out of alignment. However, tip temperatures rise above the normal level, reducing tip life. Figure 20 shows three of the most common types of misalignment errors. Such errors usually occur to some degree because of the distortion in shape that results from handling and stacking large numbers of soft, pure-lead plates. However, successful bonds have been made when lugs are radially or tangentially out of alignment by as much as 0.125 inch. This is quite severe when one considers that the lugs are only 0.188 inch thick. This capability for handling misalignment is a significant production advantage.

Figure 21 is a photograph of the second prototype machine built by Western Electric at Kearny, N. J., utilizing the above design principles. This machine, shown installed at the Design Capability Line, was used to confirm that two bonds could be made simultaneously and also to test out the various control techniques. An improved version of this machine with the capability of bonding up to ten pairs of lugs simultaneously is nearing completion.
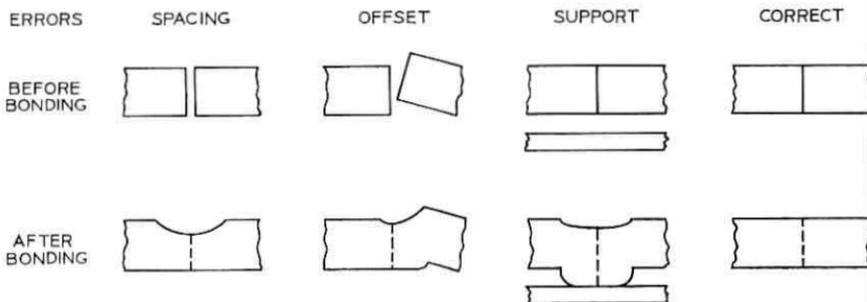
Fig. 20—Misalignment errors and resultant bond shapes.
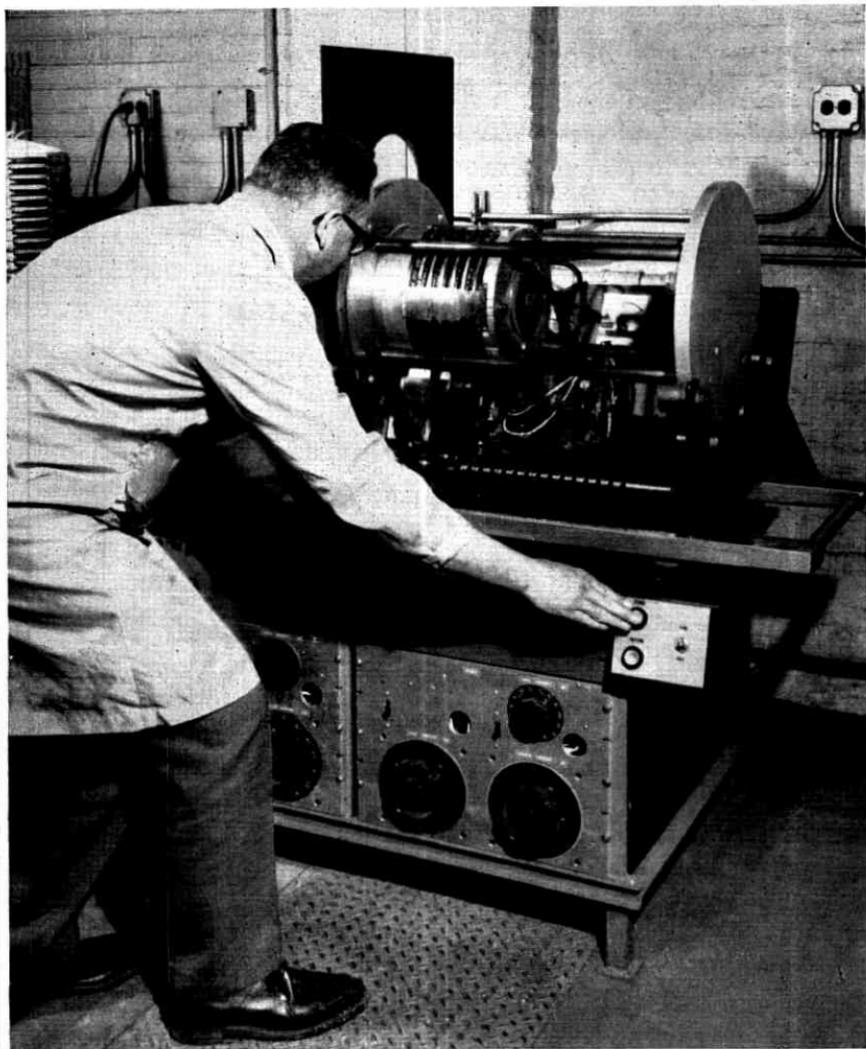
Fig. 21—Prototype MTP continuous-fusion bonder.

As production experience accrues with this new bonding process, an additional simplification will be considered. All bonding tips could be held stationary, while the battery assembly is rotated in a continuous, constant-speed motion for one complete revolution. This approach has already proven feasible.

VI. CONCLUSIONS

This article discussed the basic requirements for bonding the positive battery plate lugs of the new lead-acid battery. Two new bonding concepts were developed specifically to meet the requirements of this program. The MTP continuous-fusion bonding process was chosen for the positive plate bonding due to its many control advantages. These advantages include:

(*i*) simultaneous multiple bonds;

(*ii*) accurate and repeatable temperature control;

(*iii*) simple electronic controls;

(*iv*) individual control capability for the top, bottom, front or back end of the bond;

(*v*) elimination of any back-up or confining fixtures within the circumference of the battery;

(*vi*) automatic initiation and removal of bonding power for each pair of lugs being bonded;

(*vii*) capability of bonding through severe contamination;

(*viii*) capability of bonding severely misaligned lugs; and

(*ix*) rapid bonding speed.

The second bonding process, MTP heated ram fusion bonding, is currently being considered for a secondary application of bonding lead straps to the periphery of the positive battery plates for use in the paste forming process. Additional applications for these processes are currently being studied involving the transfer of up to 90,000 joules of energy per bond. The basic bonding concepts involved in both processes suggest the possible extension of MTP fusion bonding to other low melting temperature metals such as tin, zinc, aluminum, magnesium, and so on.

Lead-Acid Battery:

Incorporating the New Battery into the Telephone Plant

By H. J. LUER

The new Bell System battery has been designed to eliminate the problems experienced in applying the old cells. This article tells how the new cell will be incorporated into the telephone plant and describes a new plastic battery rack which has been specifically designed to accommodate the new cell in an optimum manner.

I. LEAD-ACID BATTERIES—EXISTING PRACTICES

Batteries have been an important part of the Bell System plant since the earliest days. Equipment arrangements have varied in size from a few dry cells in a wall telephone to installations occupying thousands of square feet of floor space in a modern telephone office. The design intent in each of these arrangements is fundamentally identical: to provide a mounting for the battery which protects it from its environment in such a way that neither the safety of operating personnel nor the reliability of the telephone plant is jeopardized. This intent has generally been well satisfied but not always in an economically optimum manner. Specifically, the present 180- to 1680-ampere-hour cells have certain characteristic faults which have severely limited the design options available to those charged with incorporating them into the plant.

1.1 *Equipment Problems with Present Cells*

1.1.1 *Multiplicity of Sizes*

The present product line includes eleven ampere-hour capacities, each of them purchased from three different manufacturers in both lead-calcium and lead-antimony. Since the specifications are of the "end-requirement" type, the manufacturer is free to vary many of the design parameters to suit his own product. As a result, the equipment designer must provide accommodation for 66 sets of physical dimensions.

This has been a major problem and has resulted in a proliferation of mounting arrangements with considerable confusion in the field.

1.1.2 *Cell Fragility*

Another major difficulty arises from the basic construction of the cell. The polystyrene case is fragile; the cover seal is weak; in time, electrolyte seepage can be expected in every installation. Leakage of electrolyte onto a battery rack can raise the potential of the rack sufficiently above ground to constitute a safety hazard. Leakage from two cells on the same rack can establish a current path sufficient to ignite the battery jar material. This has resulted in several disastrous fires and service outages. These characteristics leave the applications engineer very little latitude in mounting design. The cell must be supported over its entire base and can tolerate no continuous loads anywhere else on its outside surface.

1.2 *Dynamic Environments*

The impact of these constraints becomes apparent only when one considers the variety of dynamic environments to which the cells may be exposed in service. Roughly, these fall into these categories: normal service, earthquake areas, and sites hardened against nuclear attack with 2-psi overpressure, 10-psi overpressure, and 50-psi overpressure.

1.2.1 *Normal Service*

The normal service installations are typically as illustrated in Fig. 1. This "soft site" environment is quite benign and involves essentially no shock or vibration. The battery rack is designed as a "minimum cost" item and is only strong enough to support the heaviest cell. The structure consists of several welded steel upright supports and formed steel shelves of various lengths which are laid across the supports and bolted in place. The steel is protected by an acid-resistant paint and by polyethylene sheeting under the cells. The parts are shipped "knocked-down" and assembled on site by Western Electric Company installation personnel. The illustration shows a two-row, two-tier rack but a single-row, two-tier version is available for mounting against a wall and three-tier versions have been used in the past. The cells are supported on their bottom surfaces and are not otherwise braced. Options are provided to extend the upright support structure so as to allow the mounting of cable racks and bus bars over the stands.

These soft site racks are representative of the majority of battery racks used in the Bell System and generally give satisfactory service.

Fig. 1—Battery rack for normal service.

However, in the large areas of the country susceptible to earthquakes, racks require special attention. The entire West Coast has been classified in this category for some time and, because of recent wider recognition of the problem, the requirements for special treatment are gradually being extended eastward—in some cases as far as the Mississippi River.

1.2.2 Earthquake Areas

There are presently two battery racks being used in these earthquake areas. Figure 2 shows a modified soft site stand which has gained wide
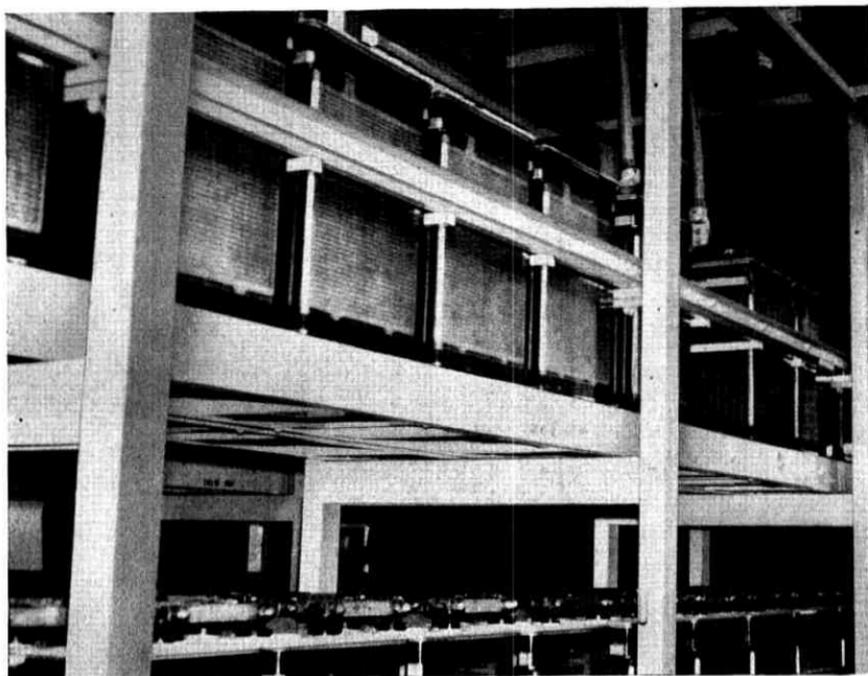
Fig. 2—Soft site rack modified for service in earthquake areas.

popularity. The modifications, which double the installed cost of the stand, consist mainly of "add-on" bracing to support the stand laterally from an overhead ironwork structure and support the cells on the shelves. The shelves themselves are stiffened by the addition of wooden blocks. The battery bracing, an "egg crate" structure bolted to the stand, provides lateral and longitudinal adjustments to accommodate the various sized cells. Because of case fragility, clearance must be provided around the entire cell; it is thus free to rattle around in response to ground motion. Nonetheless, the cell is held on the shelf. This arrangement is only marginally satisfactory but will withstand a mild tremor without damage and will probably ride out a moderate shock without collapse.

The frequently raised question of whether or not these racks would survive a major earthquake has remained unanswered because no Bell System area has experienced such an event since these stands were put into service. However, racks of identical design were in use in the Alaska Communications Building in Anchorage during the 1964 Good Friday earthquake. Figure 3 illustrates the damage sustained by one of these racks. This was a three-tier, two-row rack and employed the

egg-crate bracing. Collapse was apparently caused by a lateral motion of the floor but there was certainly substantial coincident vibration. During collapse, the middle shelf came into contact with the terminals of the lower cells and caused a fire which destroyed some of the lower cells.

Shortly after this incident, a new rack (Fig. 4) was designed which provides additional stiffness and strength through the use of heavier
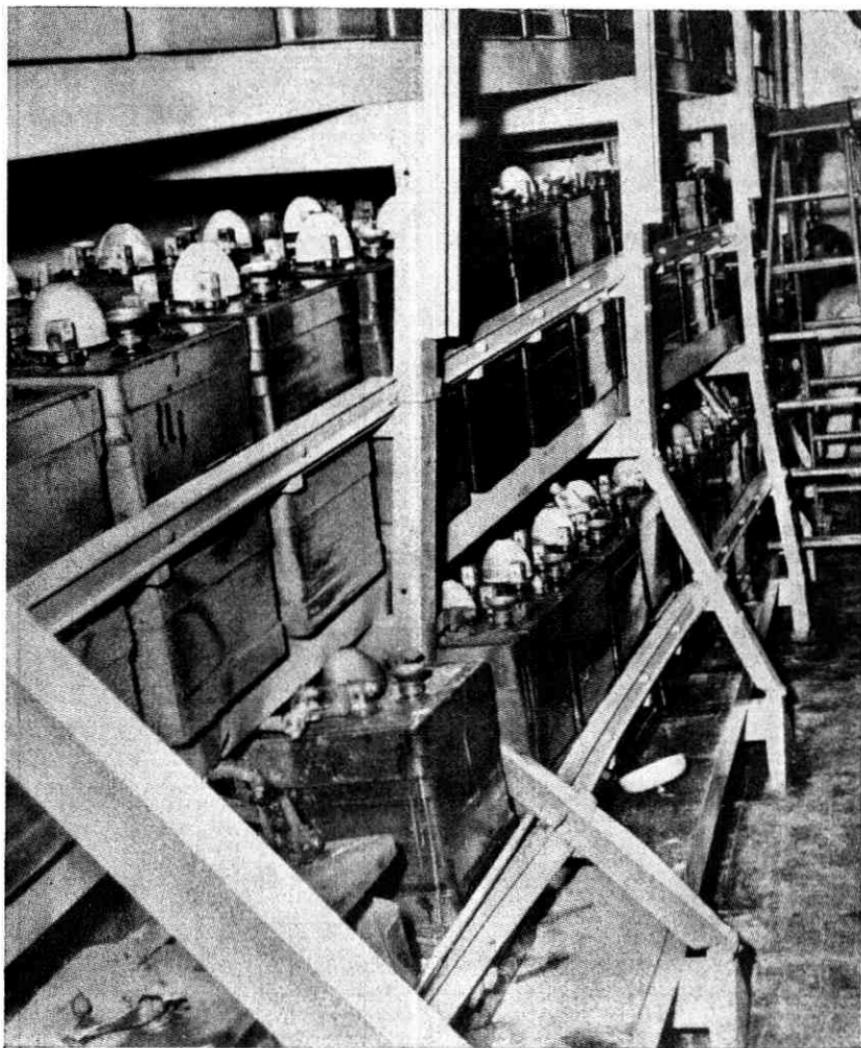
Fig. 3—Battery rack damage resulting from Alaska earthquake, Good Friday, 1964.
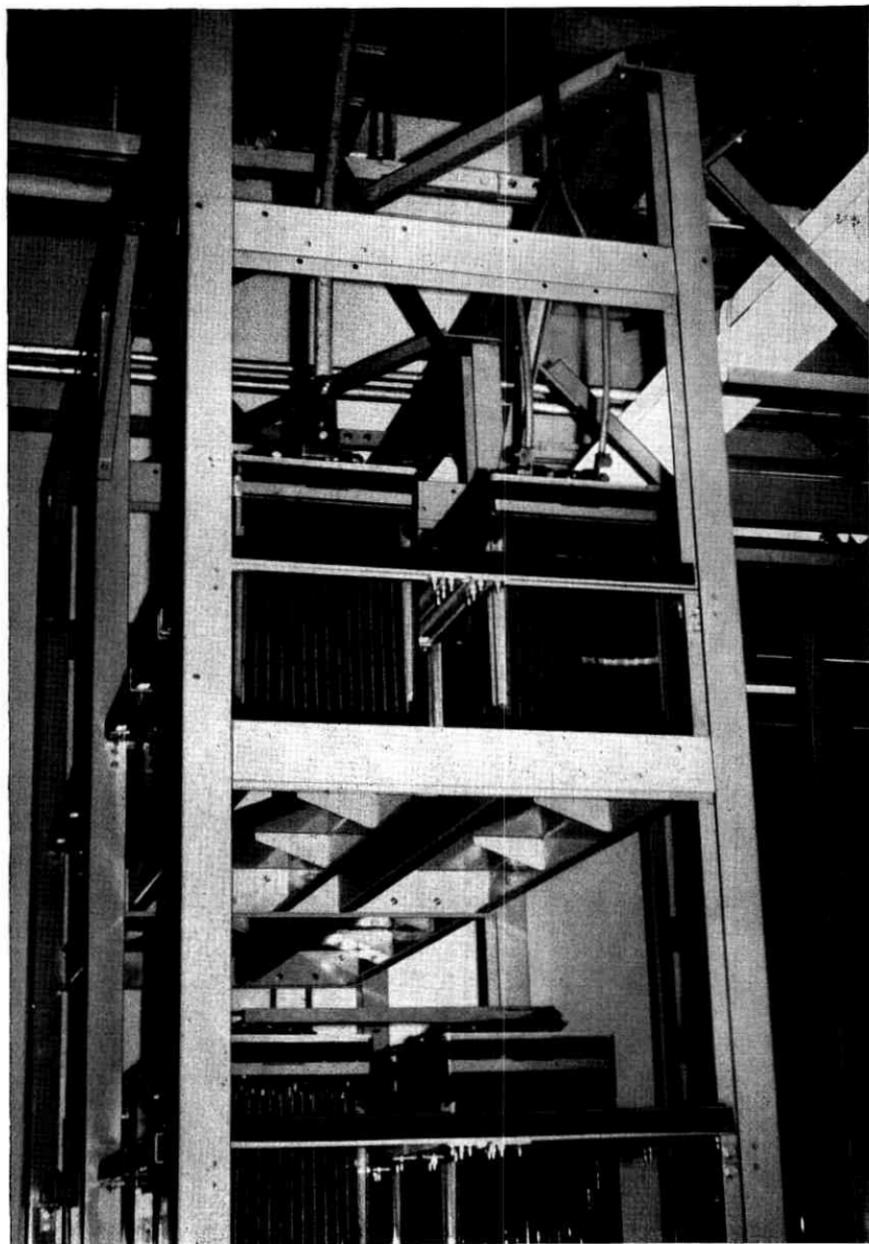
Fig. 4—Battery rack designed for earthquake areas.

steel with diagonal bracing within the rack structure and the welding
of gussets under the shelves. The shelves are bolted between the up-
rights and the egg-crate has been strengthened. The cells must still
be allowed to rattle, however, and thus would be thrown upward if
substantial vertical acceleration were experienced. This rack is now
the recommended standard rack for use in earthquake areas, but there
has been considerable resistance to its general use because it costs three
to four times as much as the earlier model.

1.2.3 *Hardened Sites*

The 2-psi hard sites impose vibration and shock loads which are of
about the same acceleration level as a moderate earthquake and of
shorter duration. It is anticipated that the rack illustrated in Fig. 4
will prove satisfactory for this service.

The 10- and 50-psi hard sites require a much higher order of protection.
Dynamic loads during a nuclear attack are expected to exceed the
capabilities of both the earthquake rack and the standard batteries
themselves.

Figure 5 shows a battery modified for use in these sites. This is essen-
tially the standard polystyrene cell enclosed in a glass-reinforced polyester
frame to provide mounting details and shock protection. The cell is
available in 1680- and 840-ampere-hour versions and is purchased with
hardening details attached at a cost of two to three times that of the
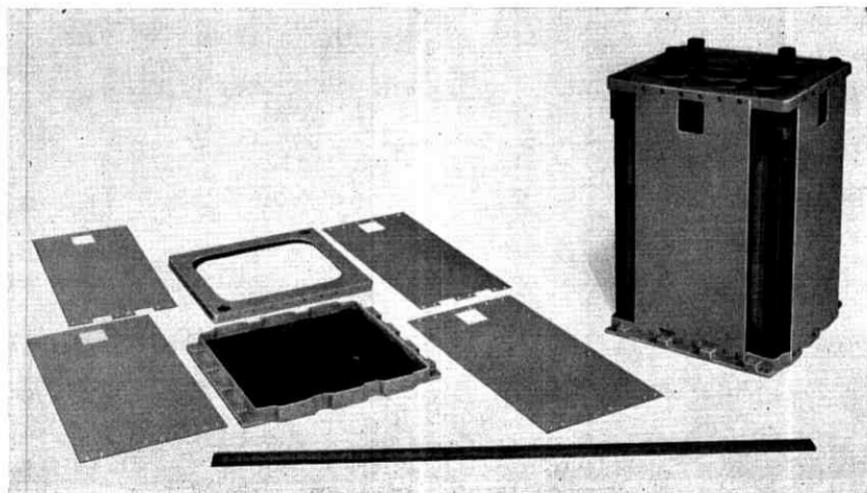
Fig. 5—Specially reinforced cell for service in sites hardened to resist nuclear
attack.

regular cell. Unfortunately, many of these cells leak and, because of the reinforcing frame, they are extremely difficult and costly to clean.

Environmental tests were conducted on this design to assure that it could survive an expected 5-G shock when mounted in the 50-psi site. In order to make certain that the cell is never exposed to more than 5 Gs in service, a new battery rack was designed specifically for hardened site use (Fig. 6). This stand is much stiffer and stronger than the earthquake stand. It is made of heavier gauge steel and has added reinforcement welded to the uprights for additional horizontal strength. Two 2-tier, 2-row racks, separated by an aisle, are joined by a diagonally braced superstructure to provide lateral strength. The hardened cells are bolted to the shelves and are interconnected by means of fine-strand flexible cable instead of the usual rigid bus bars. This allows for some relative motion among the cells with minimal stress on the battery posts.

Since accelerations due to floor motions in a 50-psi hardened site
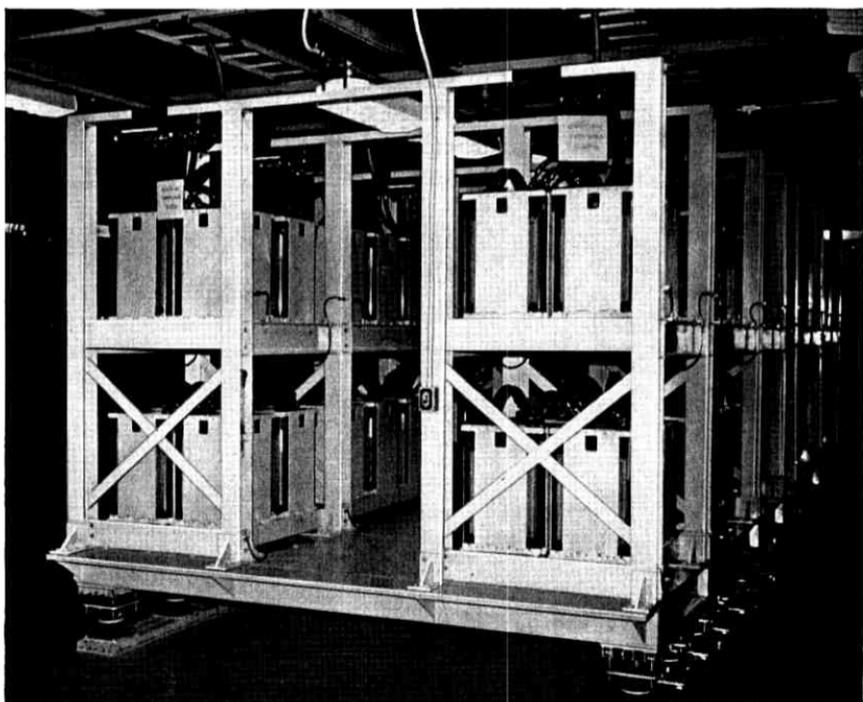
Fig. 6—Spring-mounted battery rack for service in sites designed to withstand 50-psi overpressure.

are expected to exceed the 5-G capability of the cells, the stands in these sites are mounted on a rigid steel platform which in turn is supported on a system of steel or rubber springs. The mass of the stand is matched to the stiffness of the springs to provide an overall natural frequency of about 4 cycles per second. This effectively decouples the stand from the building and assures a sharply defined shock and vibration regime for the cells with expected peak acceleration never exceeding 5 Gs. Because motions in the 10-psi sites are expected to be far less severe and not to exceed 2 Gs at any frequency, the stands in these sites are bolted directly to the floor without a steel platform or suspension springs.

The special cells, stiffened stands, flexible connectors, platforms, springs, and low production rates make these installations 10 to 20 times as expensive as the normal installation. Fortunately, they involve only a small percentage of the total number of batteries being used.

II. THE NEW CELL

Let us now consider the advantages and disadvantages of the new battery from a plant viewpoint.

The most significant feature is the degree of design control which will be established for the new cell. All parts and materials will be controlled by separate Bell System specifications. Fabrication and assembly procedures will be carefully delineated and will be under the surveillance of Western Electric inspection personnel. This assures uniformity in cell size and performance and allows standardization of plant equipment and procedures to accommodate a reasonably stable product line.

2.1 *Reduction in Number of Sizes*

Instead of the present 66 physical sizes, the new cell will be available in only four sizes. Table I lists the present sizes and the replacing new sizes. These four capacities will be packaged in the same diameter jars with four different heights. All corresponding internal parts will be identical for the four sizes, but the number of plates will differ. Terminals and cover details will be identical for all sizes. These sizes have been selected as a compromise between economies deriving from high volume manufacture and those attendant to buying the minimum battery needed for a given installation.

2.2 *New Cell Configurations*

This reduction in available sizes is not without penalty. Figure 7 compares the outline configurations of the new cells with those of the

TABLE I—COMPARISON OF PRESENT-CELL SIZES
AND REPLACING-CELL SIZES

Present Cell	Replacing Cell	Projected Annual Purchases
180 A.H.* 240 A.H.	240 A.H.	14000
300 A.H. 420 A.H.	420 A.H.	25000
540 A.H. 680 A.H. 840 A.H.	840 A.H.	18000
1080 A.H. 1320 A.H. 1680 A.H.	1680 A.H.	52000

* In some installations it may be more economical to replace the 180-ampere-hour cell with parallel strings of the present 100-ampere-hour cell.

old cells. It is apparent that, while total volume is approximately the same, there is a substantial change in geometry. Except for the largest size, the new cells are lower than those they replace but have a larger "footprint."

This change in geometry poses no problem for the 1680-ampere-hour cell in either new installations or in the replacement of existing cells. Figure 8 shows a trial installation in a New Jersey Bell office in which a string of new cells was installed in space previously reserved for 1680-ampere-hour cells of the old design. These field-trial cells have smaller capacity but the outside dimensions are equivalent to the new 1680-ampere-hour cell. There is sufficient space on the shelves for the new cell and the additional height is readily accommodated by the existing stand.

For the smaller sizes, the problem is considerably more difficult. The new cells cannot be exchanged on a one-for-one basis without changing the stand. Also, the present stand design would demand excessive floor space if it were made long enough for the new cells. The solution here is to provide a completely new stand design which takes advantage of the geometry of the new cell to permit three- and four-tier arrangements. Present practice limits battery racks to two tiers because air tends to stratify into layers of different temperature. As a result of this temperature differential, cells near a ceiling connected in a series with others near the floor would self-discharge under normal float-voltage

240-AMPERE HOURS

480-AMPERE HOURS

480-AMPERE HOURS
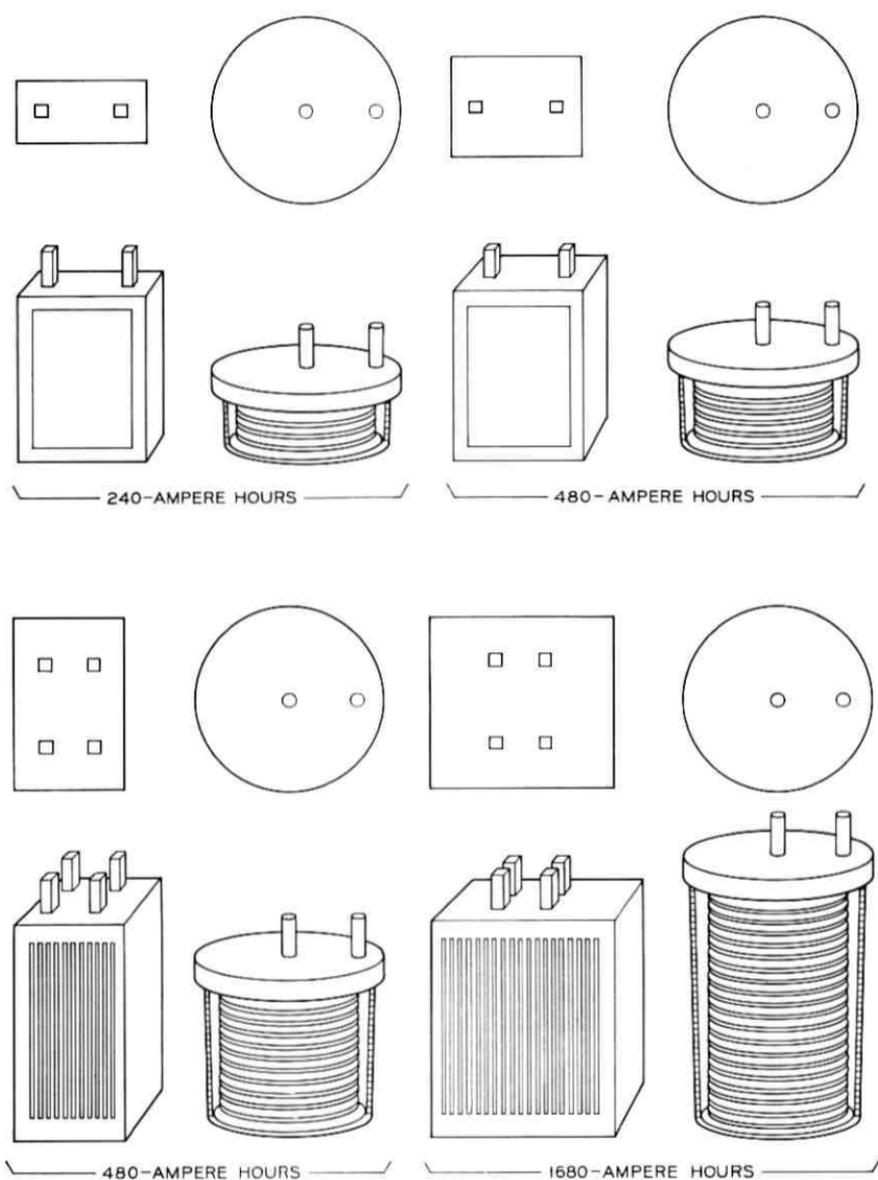
1680-AMPERE HOURS

Fig. 7—Comparison of new round cells with old cells.

Fig. 8—Early trial installation of new cells.

conditions.[1] The new cells will be short enough so that three tiers of the 840- and 420-ampere-hour size and four tiers of the 240-ampere-hour size will be less than 7 feet high (Fig. 9). It is expected that temperature differentials within this height will be acceptable.

Using this approach the stands for the new cells will be only slightly longer than the existing standard arrangements. (Nine inches longer for a 48-volt battery—Fig. 10.) It is expected that this will cause no difficulty in new installations and in most existing sites.

2.3 *Plastic Battery Rack*

The physical characteristics of the new cell permit far greater design latitude than the present battery structure. The cell case has excellent impact strength and can also be subjected to moderately high continuous stresses. Thus it can be safely handled, mounted and clamped by any of its surfaces, including the underside of the cover seal lip. This will permit the elimination of the large variety of present battery stand designs and the development of economical and practical mounting
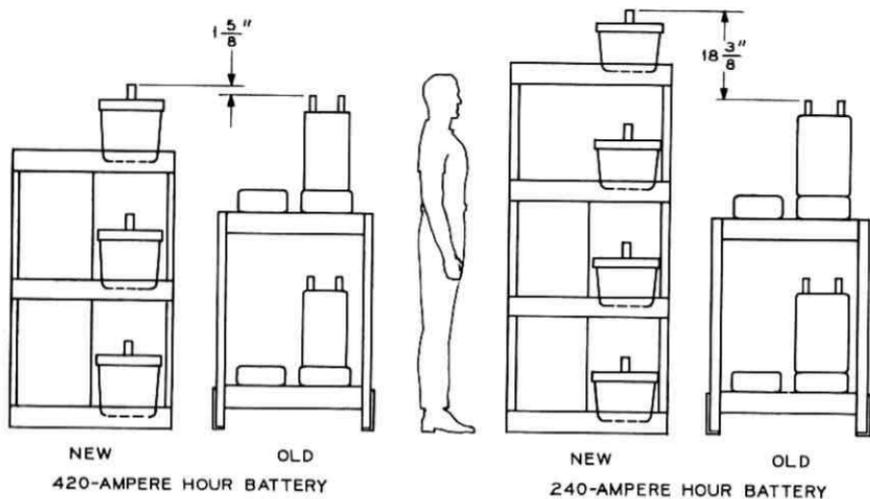
Fig. 9—Three- and four-tier arrangements for smaller sizes of new cells.
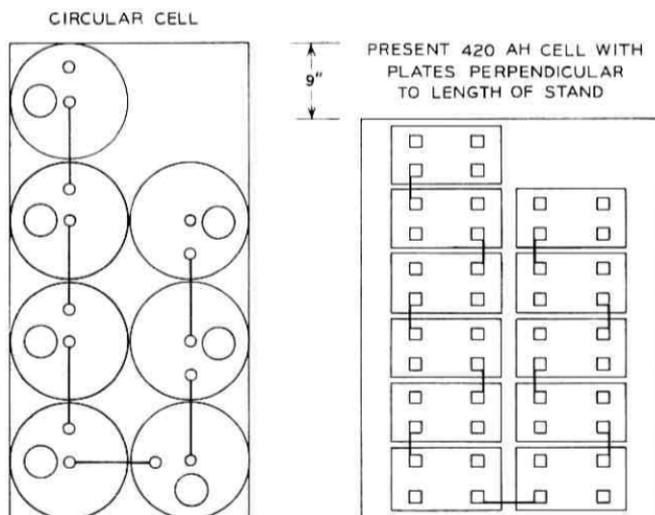
Fig. 10—Plan view of three-tier rack compared with equivalent existing rack.

arrangements which, with small variations, can be used in soft, earth-quake and hardened sites. The design of such a stand is now essentially complete and models are being fabricated for a field trial.

2.3.1 *The Parts*

The two parts from which the stand is assembled are shown in Fig. 11.
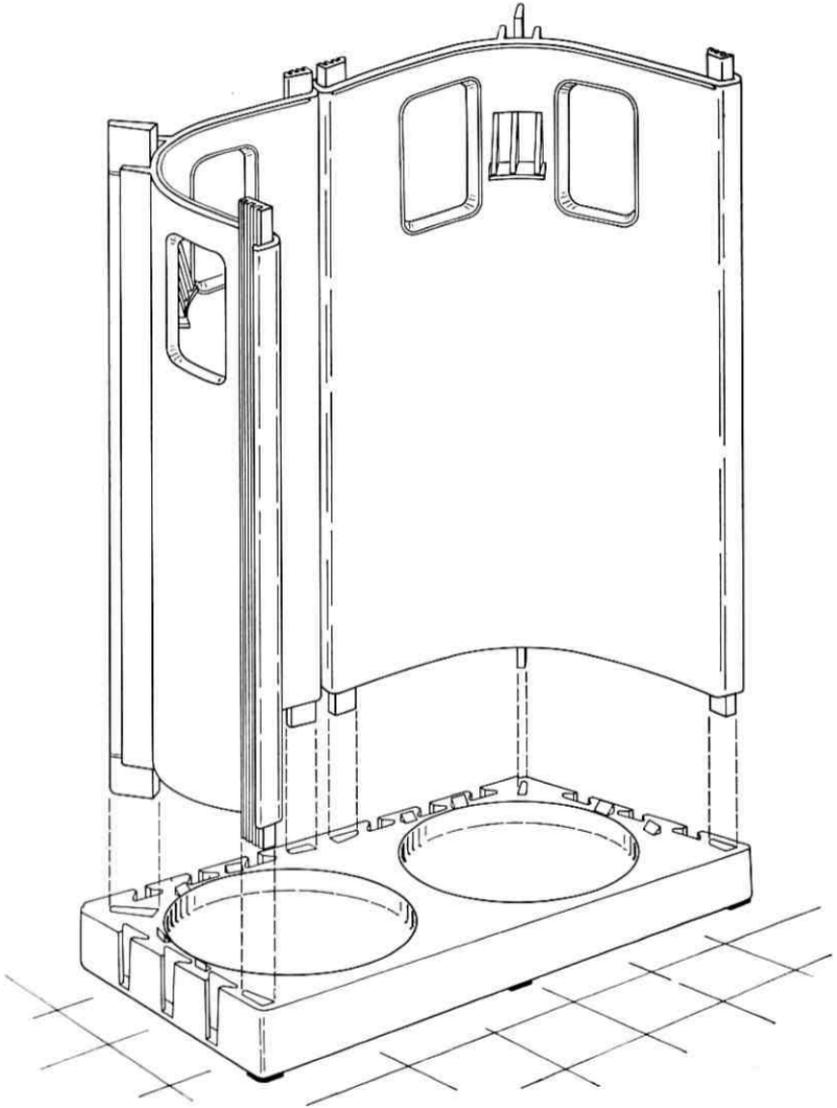
Fig. 11—Parts for new plastic battery rack.
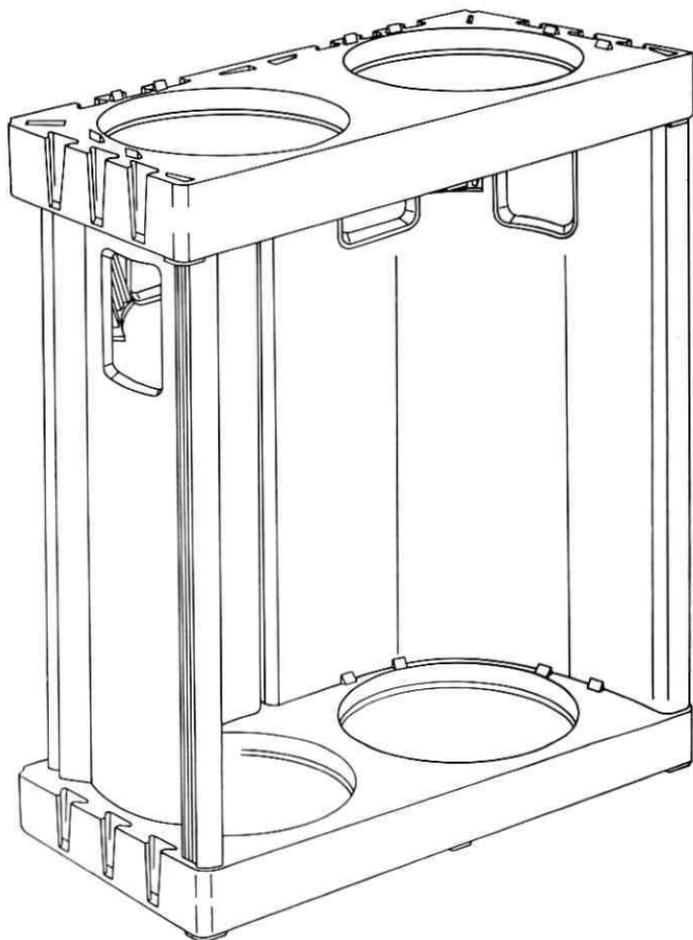
Fig. 12—Assembled basic module of new rack.

These parts are compression molded of fibreglass-reinforced polyester. This material was selected because it is strong, stable, relatively inexpensive and a non-conductor. A non-conductor is desirable for this application to avoid the safety- and fire-hazards alluded to earlier. The plastic is nonflammable in air and acid-resistant and will have color molded in to harmonize with the remaining plant.

To provide positive cell retention, the base part contains two $1\frac{1}{2}$ inch deep wells into which the cells are set. For added rigidity, the back is curved to follow the contour of the cells with access holes provided for cell interconnections. Both the base and the back have deep, molded-
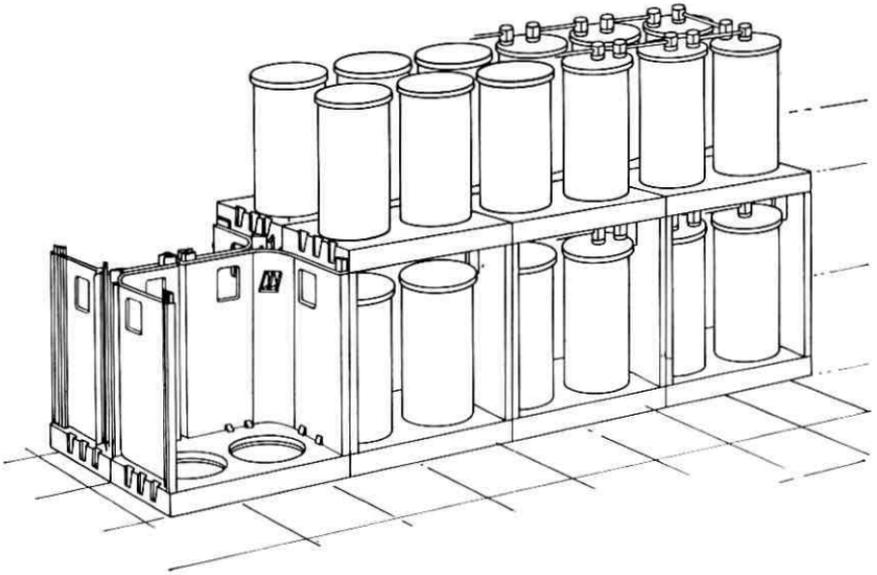
Fig. 13—Free-standing two-row rack.
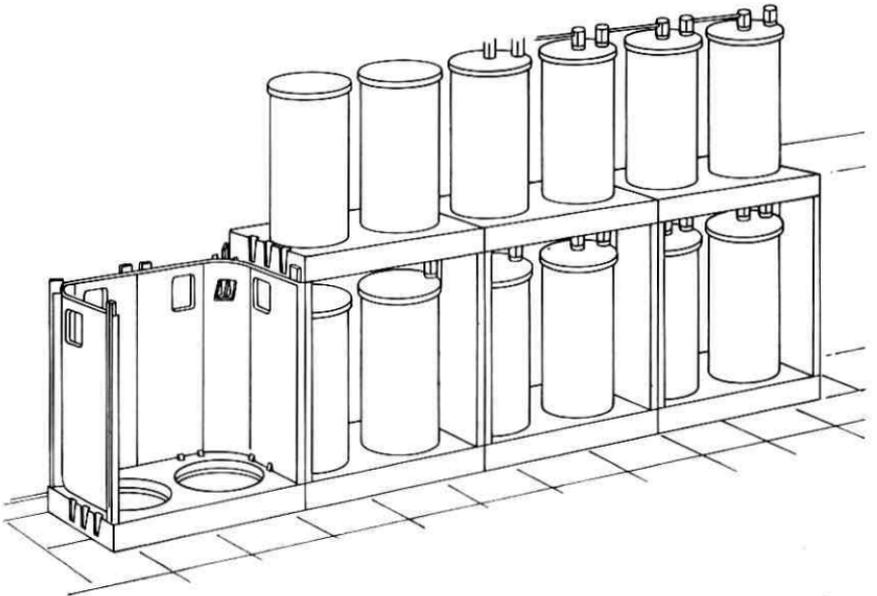
Fig. 14—Single-row rack for wall mounting.

in ribs to provide stiffness and strength. The backs will be available in four different heights to accommodate the various cell sizes. These parts will be shipped loose and assembled on site.

2.3.2 *The Assembled Soft Site Rack*

A basic module comprises two bases and two backs, assembled as shown in Fig. 12, and provides mounting space for four cells. The backs have projections top and bottom which cement into wells provided in the base using an epoxy cement with toothpaste-like consistency.

The modules can be further assembled to provide as many mounting positions as needed. They can be used as either a free-standing two-row stand (Fig. 13) or a single-row stand (Fig. 14) for mounting against a wall.
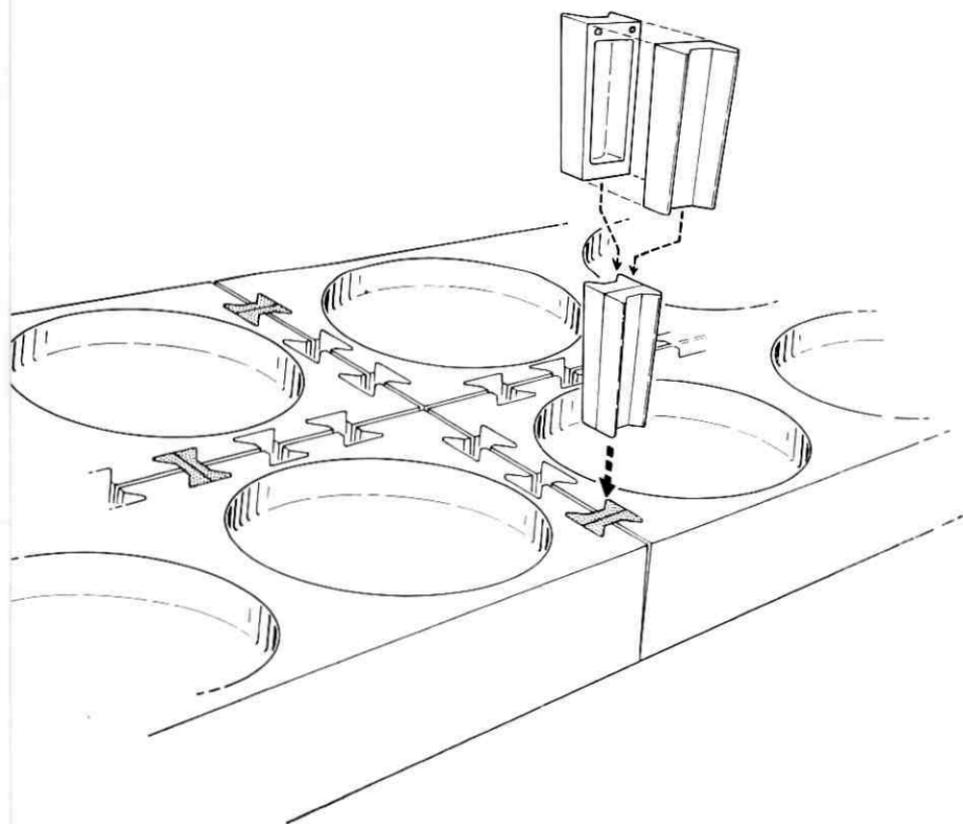
Fig. 15—Interlocking with dovetail keys.

For soft sites, the basic module has sufficient strength to carry its four cells independently. However, for added stability the modules will be interlocked using dovetail keys as shown in Fig. 15. The keys consist of two identical pieces molded of the same material as the rest of the stand and epoxied in place. Although space is provided for up to 12 keys per base, it is expected that 3 will suffice for a soft site.

2.3.3 Cell Installation

These stands will be mounted directly on the floor to reduce height and will allow the installation and removal of individual cells when necessary. Installation procedures are currently tailored to the fragility of the old cell. The cells are wrestled out of their shipping container onto a platform hoist. The platform is then positioned next to the stand at the proper height and the cell pushed onto the stand by hand. To remove a cell, the procedure is reversed, the cell being pulled onto the platform using a strap. This method is effective but on occasion has resulted in cell damage. Since the weight of the new cell can easily be supported by the strength of its cover-to-jar seal, a new hoist is being developed in conjunction with Western Electric installation engineering which will grasp the cell by its cover lip for lifting and positioning during installation and removal.

2.3.4 Earthquake Areas

While designed primarily for soft site use, the stand illustrated in Figs. 13 and 14 does provide some degree of earthquake resistance. Because of the ribs and the shape of the backs, there is sufficient shear strength to avoid stand collapse during a mild shake and the wells in the base will also keep the cells from vibrating off the shelves.

In areas where severe earthquakes are expected, a full complement of dovetail keys will be used and another layer of backs and bases will be added as shown in Fig. 16. Molded into the backs is a small ribbed shelf which is used to lock the cell into the stand by means of a snap-in retainer. This retainer will be injection-molded of the same material as the battery case and will be used only in earthquake and hardened sites. Simple steel brackets and $\frac{5}{8}''$ threaded rods are used to bolt the stand to the floor and, via diagonal bracing, to overhead ironwork. This is very similar to present practice and should find ready acceptance in the field. While dynamic tests are incomplete at the present time, it is expected that this arrangement will be suitable for 2- and 10-psi hard sites as well as the earthquake sites.
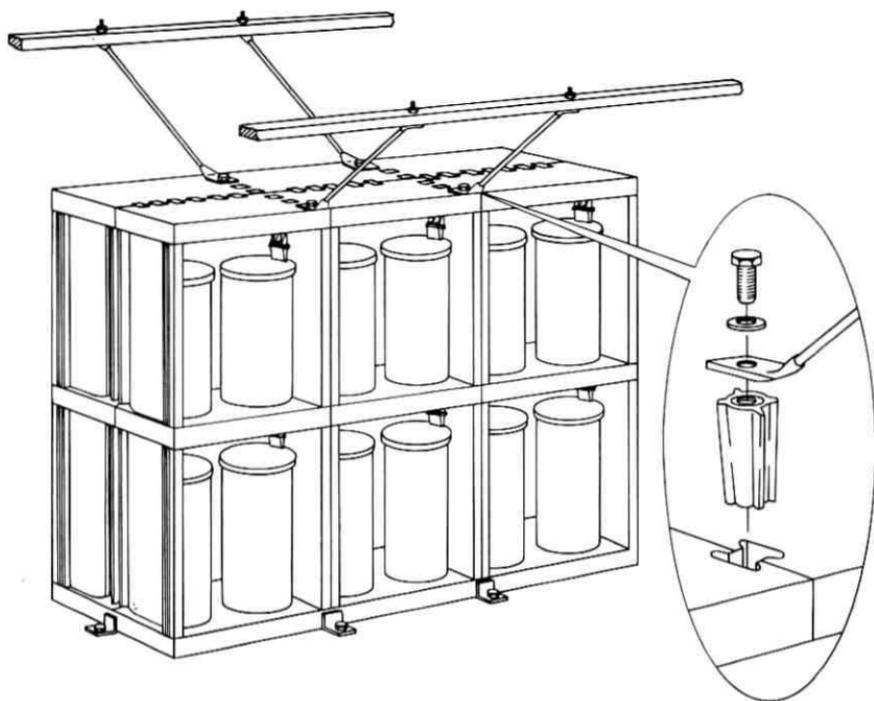
Fig. 16—New rack as installed for service in earthquake areas and in 2-psi and 10-psi hard sites.

2.3.5 50-*psi Hard Sites*

The 50-psi sites require additional modification as illustrated in Fig. 17. Shown is the earthquake stand with channel sections added top and bottom. The stand is raised off the floor and suspended from the ceiling by means of presently available, standard shock isolators. Standard restrainers, which limit the swing during normal service but break during a severe shock, are installed at the bottom. Sufficient clearance is provided around the stand to allow free motion. These arrangements will limit the dynamic loads to about 3.5 Gs in a vertical direction and 1 G in the horizontal direction. It is expected that this will be within the capability of both the new stand and the new cell.

2.3.6 *Costs—New Stand Versus Old*

Present estimates are that the new soft site stand will cost about the same as the old soft site stand but the new earthquake stand will cost only about one-third as much as its present equivalent. Cost of
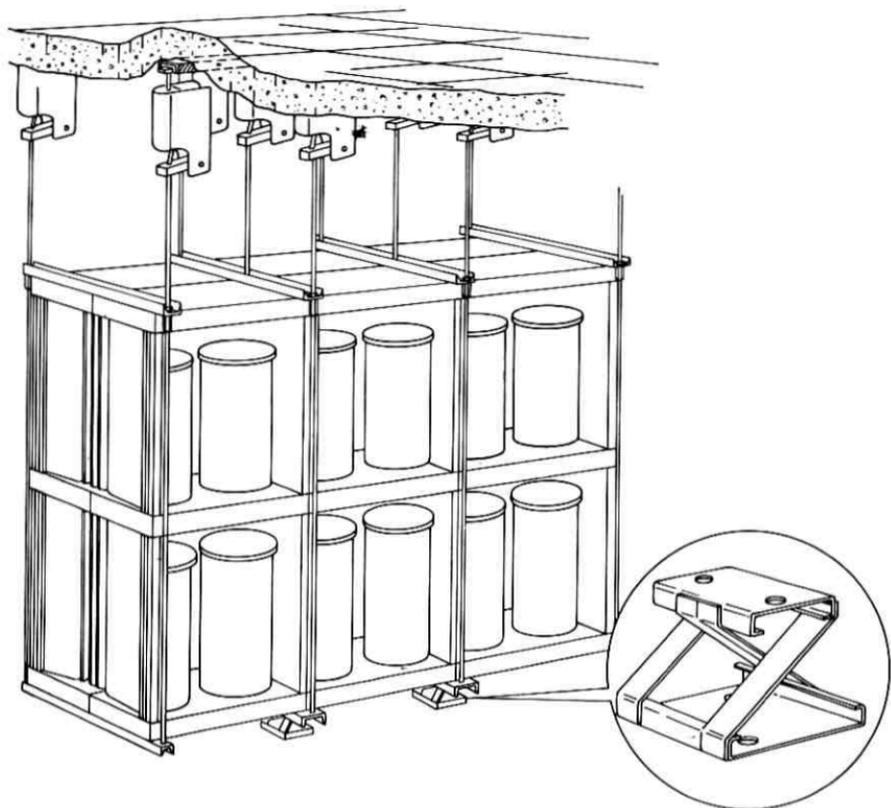
Fig. 17—New rack ceiling-suspended for service in a 50-psi hard site.

the 2- and 10-psi hardened site stands will be similar to that of the earthquake stands and the cost of the 50-psi stand will be about one-fifth that of the present arrangement.

2.4 *Cover Details and Cell Interconnections*

The cells used in early field trials are shown in Fig. 8. They provided only one mounting position for the filling funnel and explosion-proof vent. This caused some difficulty in accessibility for maintenance since the funnel could not be positioned near the aisle in all cases. Additional problems were encountered in interconnecting these cells. Flexible cables with crimped-on lugs were clamped by means of lead-plated nuts to $\frac{1}{2}''$ threaded copper rods which had been cast in place in the positive and negative posts. The stress of tightening the nuts was sufficient in many cases to bend the rods and damage the threads.

As a result of these experiences cells for later field trials will be modified as shown in Fig. 18. The fill tube will be shipped loose and installed in one of two mounting positions after the cell is in place on the stand. This assures that the funnel will always be near the aisle for accessibility. The other mounting position will be plugged (they are both plugged during shipment).

The present standard cell interconnection employs lead-plated copper bus bars bolted to both sides of the post. Depending on the size and configuration of the cell, up to 4 bus bars are used. The later field trial cells will allow a similar arrangement except that only one positive post and one negative post are provided. However, the posts are higher and can still accommodate up to 4 bus bars. This design permits the
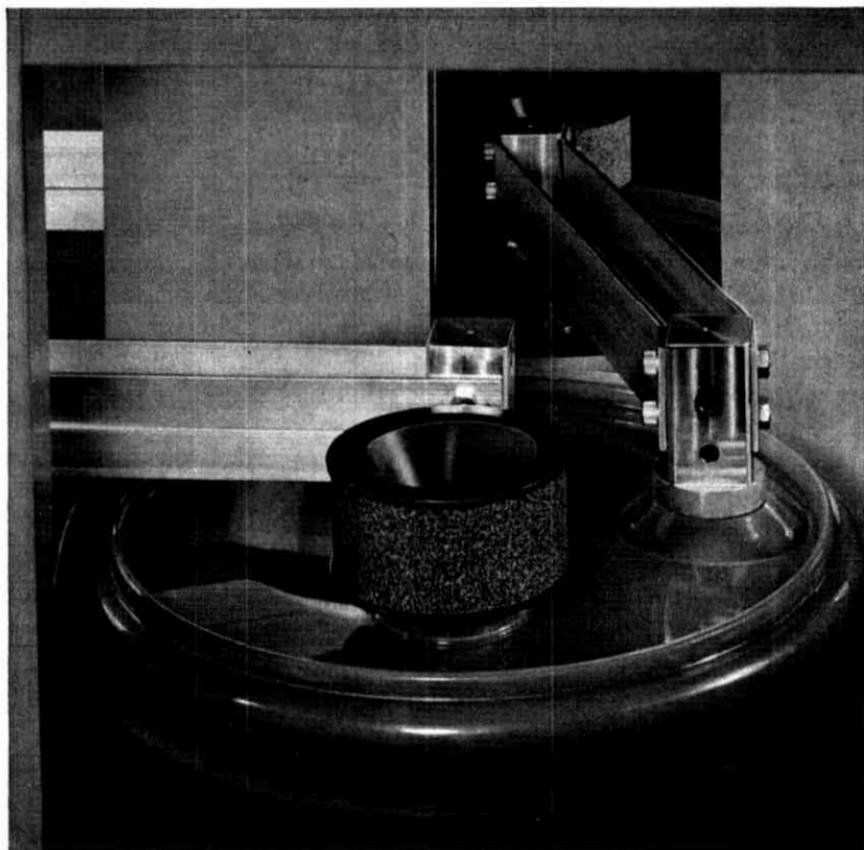
Fig. 18—Top view of new cell showing cover details.

removal of connectors for post cleaning without breaking the electrical connection of the string, and is arranged so that connections can readily be made from either of two perpendicular directions. The post is of lead-antimony alloy and is cast as part of the center-pour operation.

Since the new cell is expected to have a service life which is essentially as long as the equipment it services, consideration is being given to more permanent means of interconnecting cells. Several such methods are presently under development in conjunction with the Western Electric Engineering Research Center, Princeton, New Jersey. The joints will be acid-tight and will provide a reliable, permanent, low-resistance, maintenance-free connection for the lifetime of the battery. It is expected that overall voltage drop within the string will be substantially reduced, thus allowing the cells to be discharged to a lower potential while still maintaining adequate plant voltage. Provision will be made to allow removal and replacement of single cells in case of early failure. It is expected that these techniques will be employed on some of the later field trial cells.

2.5 Battery Plants Within Equipment Lineups

The Operating Companies will also benefit from reduced maintenance due to improved cover seal and post seals. In the past, frequent cleaning of cells has been necessary because of electrolyte leakage and post corrosion. The presence of acid has made it desirable to keep the batteries physically separated from the rest of the telephone equipment. The improved seals will eliminate this objection and allow the installation of power plants within the lineups of the power-using equipment. This would provide shorter and smaller power cables without excessive voltage drop. The new battery stand is designed to be placed in equipment lineups. A single row stand can be installed in a 15-inch-deep lineup with front access only (Fig. 19), and a two row stand in a 30-inch-deep lineup with front-and-rear access. As shown in Fig. 20, doors can readily be provided as an optional add-on accessory where appearance is a factor.

2.6 Testing and Field Trials

We are now engaged in an extensive test program to prove in these new concepts. The cells and stand are being subjected to a series of environmental tests including shock and vibration tests to simulate the shipping, handling, earthquake and nuclear site environments. Preproduction models of the new stands will be installed in several
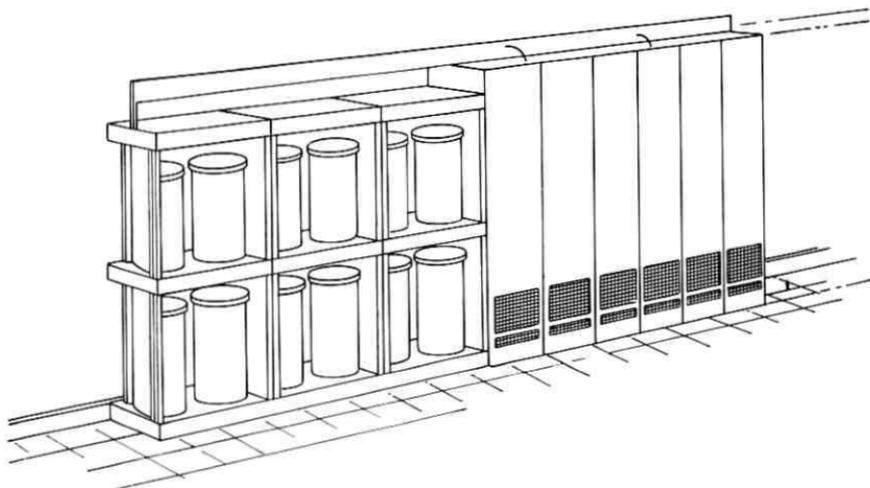
Fig. 19—Single-row rack installed in an equipment line-up.

field trial locations by Western Electric personnel with the new permanent cell interconnections used in some of these sites.

As these tests proceed, we expect to learn more about the advantages and flexibility of the cells and mounting arrangements. These results might considerably temper the approach outlined here, but we expect
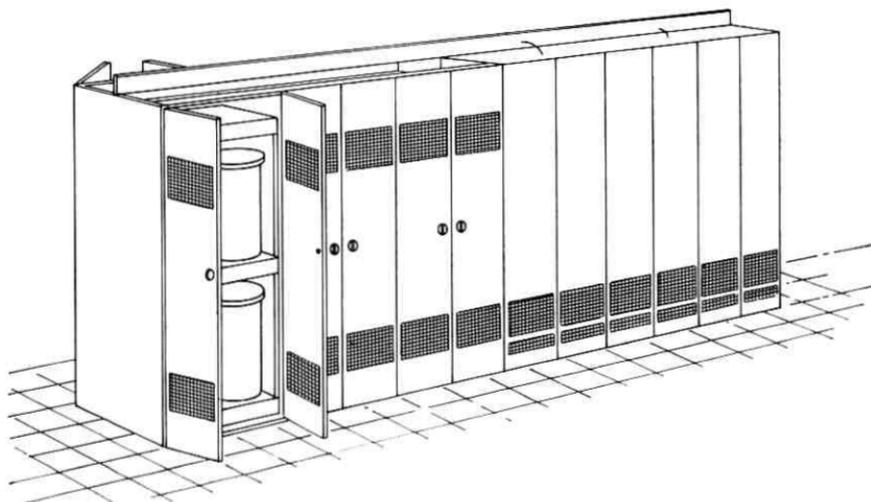
Fig. 20—Two-row stand in an equipment lineup with doors added to improve appearance.

that the integrated systems approach used in the development of the cells and their introduction into the plant will yield immediate and long-term benefit to the Bell System by providing more reliable service at substantially reduced cost.

REFERENCE

1. Milner, P. C., "Float Behavior of the Lead-Acid Battery System," B.S.T.J., this issue, pp. 1321–1334.

Traffic in Connecting Networks When Existing Calls Are Rearranged

By V. E. BENEŠ

Recent mathematical[1,2] *and engineering studies*[3,4] *of the possibility of rearranging existing calls in a connecting network (so as to add more calls) have raised the problem of loss (probability of blocking) for networks operated in this manner. We consider and solve this problem in the context of a Markov traffic model, for a connecting network operated according to the rule that if a new call is blocked, but can be accommodated by rearranging the calls in progress, then it is put up, after some choice of rearrangement.*

I. INTRODUCTION

Each state of the network realizes a given assignment of inlets to outlets, a specification of who is to talk to whom. There is a natural map γ that takes the set S of states into the set A of assignments. The equilibrium probabilities of the process x_t on S we use as a traffic model are complicated and unknown functions of the offered load. But it turns out that, because of the policy of rearranging whenever necessary, $\gamma(x_t)$ is also a Markov process, and one whose state probabilities are easy to calculate.

We give explicit analytical formulas for the equilibrium probability of a given assignment of inlets to outlets, for the probability of n calls in progress. For networks that are one-sided (inlets = outlets) or two-sided (inlets $\cap$ outlets = φ), all the important constants of traffic engineering can be obtained from a *partition function*, a polynomial in the offered load with coefficients depending only on network structure; these constants are the loss, the load carried, the calling rate, and the load variance. These analytical formulas arise from an unexpected connection with the "thermodynamic" model for telephone traffic, described in an earlier paper. As an application, we solve the problem of calculating the loss in a connecting network, made of stages of

rectangular switches, which is rearrangeable except for the fact that the outermost stages are concentrators.

II. STATES AND ASSIGNMENTS

A mathematical model like that of Ref. 5 will be used. The elements of this model separate naturally into combinatorial and probabilistic ones. The former arise from the structure of the connecting network and from the ways in which calls can be put up in it; the latter represent assumptions about the random traffic the network is to carry. We discuss the combinatorial and structural aspects in this section; terminology and notation for these aspects are introduced. The probabilistic aspects are considered in a later section.

A connecting network ν is a quadruple $\nu = (G, I, \Omega, S)$, where G is a graph depicting network structure, I is the set of nodes of G which are inlets, Ω is the set of nodes of G that are outlets, and S is the set of permitted states.[1] Variables x, y, and z at the end of the alphabet denote states, while u and v (respectively) denote a typical inlet and a typical outlet. A state x can be thought of as a set of disjoint chains on G, each chain joining I to Ω. Not every such set of chains represents a state: sets with wastefully circuitous chains may be excluded from S. It is possible that $I = \Omega$, that $I \cap \Omega = \varphi = $ null set, or that some intermediate condition obtains, depending on the "community of interest" aspects of the network ν.

The set S of states is partially-ordered by inclusion $\leq$, where $x \leq y$ means that state x can be obtained from state y by removing zero or more calls. If x and y satisfy the same assignment of inlets to outlets—that is, are such that all and only those inlets $u \in I$ are connected in x to outlets $v \in \Omega$ which are connected to the same v in y (though possibly by different routes), then we say that x and y are equivalent, written $x \sim y$.

We denote by A_x the set of states that are immediately above x in the partial ordering $\leq$, and by B_x the set of those that are immediately below. Thus

$$A_x = \{\text{states accessible from } x \text{ by adding a call}\}$$

$$B_x = \{\text{states accessible from } x \text{ by a hangup}\}.$$

Also, we let A_{cx} be the set of states that could result from x by putting up call c. A call c is new in x if the terminals of c are idle in x. A call c new in x is *blocked* in x if there is no $y \in A_x$ with c in progress in y; it is

completely blocked in x if there is no way of rearranging the calls of x to an equivalent state z, $z \sim x$, such that c is not blocked in z.

The number of calls in progress in state x is denoted by $|x|$. The number of calls which are not completely blocked in x is denoted by $\sigma(x)$, for "successes in x." The functions $|\cdot|$ and $\sigma(\cdot)$ defined on S play important roles in the stochastic process to be used for studying traffic. In addition, we use the notations

β_x = number of idle inlet-outlet pairs completely blocked in state x,

α_x = number of idle inlet-outlet pairs in state x,

and we note that $\alpha = \beta + \sigma$.

It can be seen, further, that the set S of states is not merely partially ordered by $\leq$, but also forms a semi-lattice, or a partially ordered system with intersections, with $x \cap y$ defined to be the state consisting of those calls and their respective routes which are common to both x and y.

An assignment specifies what inlets should be connected to what outlets. The set A of assignments can be represented as the set of all fixed-point-free correspondences from subsets of I to Ω. The set A is partially ordered by inclusion, and there is a natural map $\gamma(\cdot): S \to A$ which takes each state $x \, \varepsilon \, S$ into the assignment it realizes; the map $\gamma(\cdot)$ is a semilattice homomorphism of S into A, since

$$x \geq y \text{ implies } \gamma(x) \geq \gamma(y),$$

$$\gamma(x \cap y) \leq \gamma(x) \cap \gamma(y).$$

Variables a and b are used for members of A. For $a \, \varepsilon \, A$, $|a|$ is the number of inlets (or outlets) which are "busy" if assignment a is specified.

A *unit* assignment is, naturally, one that assigns exactly one inlet to some one outlet, and it corresponds to having just one call in progress. It is convenient to identify new calls c and unit assignments, and to write $\gamma(x) \cup c$ for the larger assignment consisting of $\gamma(x)$ and the call c together, with the understanding of course that none of the terminals of c is busy in $\gamma(x)$.

Remark 1: Not every assignment need be realizable by some state of S. Indeed, it is common for practical networks to realize only a small fraction of the possible assignments. Since we are studying a network operated with rearranging when necessary, blocking will occur in a state x only when a call c idle in x is completely blocked in the

sense that no state realizes $\gamma(x) \cup c$; that is, when $\gamma(x) \cup c \notin \gamma(S)$.

The set $\gamma(S)$ of realizable assignments is also partially ordered by inclusion. We use the notations, for $a \, \varepsilon \, \gamma(S)$,

$$A_a = \{\text{realizable assignments immediately above } a\}$$

$$B_a = \{\text{realizable assignments just below } a\}.$$

$| X |$ is the number of members of a set X.

Remark 2:   For $x \, \varepsilon \, S$, $\sigma(x) = | A_{\gamma(x)} |$.

The set of calls which can be put up in state y is the same for all $y \, \varepsilon \, \gamma^{-1}[\gamma(x)]$.

III. ASSUMPTIONS

A Markov stochastic process x_t taking values on S is used as a mathematical description of an operating connecting network subject to random traffic. A Markov process similar to that of Ref. 1 will be used. This model can be paraphrased in the informal terminology of "rates" by two simple assumptions:

(*i*) The hang-up rate per call in progress is unity.

(*ii*) The calling-rate between an inlet and a distinct outlet, both idle at time u, is $\lambda > 0$.

The transition probabilities of x_t will be described after a discussion of system operation and routing.

It will be assumed that attempted calls to busy terminals are rejected, and have no effect on the state of the network. Successful attempts to place a call are completed instantly with some choice of route, or are rejected, in accordance with some *routing policy*.

It remains to say what happens to blocked or unsuccessful attempts to make a call. We assume that a policy of total rearranging is followed, according to which if a call can be put up at all, by a rearrangement of the existing calls if necessary, then it is completed. We say a call c is *completely blocked* in x if there is no state whatever satisfying the assignment $\gamma(x) \cup c$, that is, if there is no way of rearranging the calls in progress in x so as to create a free path for c. Completely blocked calls are refused, with no change of state.

The map γ is fundamental in our study of the effect of rearranging on blocking. This is because if c is a new call blocked in x which is accommodated by rearranging, then the resulting state of the system is one of the states in $\gamma^{-1}\{\gamma(x) \cup c\}$.

It is convenient to modify the matrix R, used in Refs. 1 and 5, to

cover rearranging as well as routing. This will be done by adding some more non-zero entries to represent ways of putting up calls by network rearrangement, for example, calls that are blocked but not completely blocked. A general description will be given which actually covers the use of rearrangement for unblocked calls.

Consider the set $\gamma^{-1}\{A_{\gamma(x)}\}$, consisting of all states obtainable from x by (possibly) rearranging the calls in x and then adding another call. The equivalence relation $\sim$ of "having the same calls up" (or satisfying the same assignment of inlets to outlets) induces a partition Π_x of $\gamma^{-1}\{A_{\gamma(x)}\}$. It can be seen that Π_x consists of exactly the sets $\gamma^{-1}\{\gamma(x) \cup c\}$ for calls c not completely blocked in x. For $Y \varepsilon \Pi_x$, r_{xy} for $y \varepsilon Y$ is a probability distribution over Y, and $r_{xy} = 0$ in all other cases.

The interpretation of the matrix R is to be this: any $Y \varepsilon \Pi_x$ represents all the ways in which a call c (free and not completely blocked in x) could be added when the network state is x, both with and without rearranging; for $y \varepsilon Y$, r_{xy} is the chance that if call c is attempted in state x, it will be completed (by choice of a route for c and possibly by choice of a rearrangement for x) so as to take the system to state y.

It is to be noted that routing and rearranging are carried out with complete knowledge about the current state of the network.

IV. PRINCIPAL RESULTS

The probabilistic and operational assumptions we have made give rise to a Markov stochastic process x_t taking values on S. This process is determined by its transition rate matrix $Q = (q_{xy})$, given by

$$q_{xy} = \begin{cases} 1 & y \varepsilon B_x \\ \lambda r_{xy} & y \varepsilon \gamma^{-1}(A_{\gamma(x)}) \\ -|x| - \lambda\sigma(x) & y = x \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1: $\gamma(x_t)$ *is a Markov process with transition rate matrix* $\Gamma = (\gamma_{ab})$ *given by*

$$\gamma_{ab} = \begin{cases} 1 & b \varepsilon B_a \\ \lambda & b \varepsilon A_a \\ -|a| - \lambda|A_a| & b = a \\ 0 & \text{otherwise.} \end{cases}$$

Proof: Since x_t is itself a Markov process, it is enough to prove that if $0 < t_1 < \cdots < t_n$ then

$$\text{distr } \{\gamma(x_{t_1}), \cdots, \gamma(x_{t_n}) \mid x_0\}$$

depends only on $\gamma(x_0)$. We have

$$P\{\gamma(x_{t_i}) = a_i, \quad i = 1, \cdots, n \mid x_0\}$$

$$= \sum_{x_i \varepsilon \gamma^{-1}(a_i)} P\{x_{t_i} = x_i, \quad i = 1, \cdots, n \mid x_0\};$$

with $p_{xy}(t) = P\{x_t = y \mid x_0 = x\}$, this is

$$\sum_{x_1 \varepsilon \gamma^{-1}(a_1)} \cdots \sum_{x_n \varepsilon \gamma^{-1}(a_n)} \prod_{i=1}^{n} p_{x_{i-1} x_i}(t_i - t_{i-1})$$

$$= \sum_{x_1 \varepsilon \gamma^{-1}(a_1)} \cdots \sum_{x_{n-1} \varepsilon \gamma^{-1}(a_{n-1})}$$

$$\cdot \prod_{i=1}^{n-1} p_{x_{i-1} x_i}(t_i - t_{i-1}) \sum_{x_n \varepsilon \gamma^{-1}(a_n)} p_{x_{n-1} x_n}(t_n - t_{n-1}).$$

If we can show that for x, $a \varepsilon S \times A$, the function $\varphi_{xa}(\cdot)$ defined by

$$\varphi_{xa}(t) = \sum_{y \varepsilon \gamma^{-1}(a)} p_{xy}(t)$$

depends only on $\gamma(x)$, the rightmost sum above will factor out because it would depend only on a_{n-1}. By iteration it would follow that with z_i an arbitrary element of $\gamma^{-1}(a_i)$

$$P\{\gamma(x_{t_i}) = a_i, \quad i = 1, \cdots, n \mid x_0\} = \prod_{i=1}^{n} \sum_{y \varepsilon \gamma^{-1}(a_i)} p_{z_i y}(t_i - t_{i-1}),$$

whence the theorem.

The matrix function $P(t) = [p_{xy}(t)]$ satisfies the *backward* Kolmogorov equation

$$\frac{d}{dt} p_{xy}(t) = \sum_{z \varepsilon B_x} p_{zy}(t) + \lambda \sum_{z \varepsilon \gamma^{-1}(A_\gamma(x))} r_{xz} p_{zy}(t) - [|x| + \sigma(x)] p_{xy}(t).$$

By integration, this is equivalent to the integral equation

$$p_{xy}(t) = e^{-[|x| + \lambda \sigma(x)]t} \delta_{xy}$$

$$+ \int_0^t e^{-[|x| + \lambda \sigma(x)]u} \left[\sum_{z \varepsilon B_x} p_{zy}(t - u) + \lambda \sum_{z \varepsilon \gamma^{-1}(A_\gamma(x))} r_{xz} p_{zy}(t - u) \right] du.$$

Let $a \varepsilon \gamma(S)$, and sum over $y \varepsilon \gamma^{-1}(a)$ to find, with $q_x = |x| + \lambda \sigma(x)$,

$$\varphi_{xa}(t) = e^{-q_x t} \delta_{\gamma(x)a}$$

$$+ \int_0^t e^{-q_x u} \left[\sum_{z \varepsilon B_x} \varphi_{za}(t - u) + \lambda \sum_{z \varepsilon \gamma^{-1}(A_\gamma(x))} r_{xz} \varphi_{za}(t - u) \right] du.$$

These equations can be solved by successive approximations as

$$\varphi_{xa}(t) = \lim_{n\to\infty} \psi_{xa}^{(n)}(t)$$

with $\psi_{xa}^{(0)}(t) = e^{-q_x(t)} \delta_{\gamma(x)a}$ and

$$\psi_{xa}^{(n+1)}(t) = \psi_{xa}^{(0)}(t)$$
$$+ \int_0^t e^{-q_x u} \left[\sum_{z \epsilon B_x} \psi_{za}^{(n)}(t-u) + \lambda \sum_{z \epsilon \gamma^{-1}(A_{\gamma(x)})} r_{xz}\psi_{za}^{(n)}(t-u) \right] du.$$

We have $q_x = |x| + \lambda\sigma(x) = |\gamma(x)| + \lambda|A_{\gamma(x)}|$, by Remark 1, so $\psi_{xa}^{(0)}$ depends only on $\gamma(x)$. As a hypothesis of induction, suppose that $\psi_{xa}^{(n)}$ depends only on $\gamma(z)$. Then, since $z \epsilon B_x$ implies $\gamma(z) \epsilon B_{\gamma(x)}$ and $|B_x| = |B_{\gamma(x)}| = |x|$, with $\psi_{ba}^{(n)}(t)$ the value of $\psi_{za}^{(n)}(t)$ on $\gamma^{-1}(b)$,

$$\sum_{z \epsilon B_x} \psi_{za}^{(n)}(t-u) = \sum_{b \epsilon B_{\gamma(x)}} \psi_{ba}^{(n)}(t-u)$$

depends only on $\gamma(x)$.

Similarly, since $\sum_{z \epsilon \gamma^{-1}(b)} r_{xz} = 1$ if $b \epsilon A_{\gamma(x)}$,

$$\sum_{z \epsilon \gamma^{-1}(A_{\gamma(x)})} r_{xz}\psi_{za}^{(n)}(t-u) = \sum_{b \epsilon A_{\gamma(x)}} \sum_{z \epsilon \gamma^{-1}(b)} r_{xz}\psi_{za}^{(n)}(t-u)$$
$$= \sum_{b \epsilon A_{\gamma(x)}} \left(\sum_{z \epsilon \gamma^{-1}(b)} r_{xz} \right) \psi_{ba}^{(n)}(t-u)$$
$$= \sum_{b \epsilon A_{\gamma(x)}} \psi_{ba}^{(n)}(t-u),$$

thus $\psi_{xa}^{(n+1)}$ depends only on $\gamma(x)$. It follows that φ_{xa} depends only on $\gamma(x)$. Let $\varphi_{ba}(t) = $ the value of $\varphi_{xa}(t)$ on $\gamma^{-1}(b) = P\{\gamma(x_t)=a \mid \gamma(x_0)=b\}$. By differentiation we obtain

$$\frac{d}{dt}\varphi_{ba}(t) = \sum_{d \epsilon B_b} \varphi_{da}(t) + \lambda \sum_{d \epsilon A_b} \varphi_{da}(t) - [|b| + \lambda|A_b|]\varphi_{ba}(t),$$

whence it follows that the Markov process $\gamma(x_t)$ has the transition rate matrix Γ as stated in the theorem.

Theorem 2: The equilibrium distribution p of $\gamma(x_t)$ is given by

$$p_a = p_0\lambda^{|a|}, \quad p_0 = \left(\sum_{b \epsilon \gamma(S)} \lambda^{|b|} \right)^{-1} = 1/\Phi(\lambda). \tag{1}$$

It satisfies the equation

$$(|a| + |A_a|)p_a = \sum_{b \epsilon A_a} p_b + \lambda \sum_{b \epsilon B_a} p_b, \quad a \epsilon (S). \tag{2}$$

Among all distributions q over $\gamma(S)$ satisfying the condition

$$\sum_{a \epsilon \gamma(S)} q_a|a| = \sum_{a \epsilon \gamma(S)} p_a|a|,$$

p maximizes the entropy functional

$$H = - \sum_{a \in \gamma(S)} q_a \log q_a. \tag{3}$$

Proof: Equation (2) is the equilibrium condition for the matrix Γ of Theorem 1. The solution (1) can be verified by substitution, and the external property (3) of equation (1) is well known.

Remark 3: It can be seen that $\gamma(x_t)$ is the same stochastic process as would be obtained by applying the so-called "thermodynamic" model[6] proposed by the author to the state space $X = \gamma(S)$. This fact is the heuristic reason behind Theorem 2. As a result, all the special features of the "thermodynamic" model are present here, with $\Phi(\cdot)$ playing the role of the partition function.

Remark 4: $\gamma(x_t)$ is a reversible process. Its rate matrix is symmetrizable—that is, a symmetric operator in the space with inner product $\sum_{a \in \gamma(S)} s_a l_a p_a$. From this fact follow useful inequalities for the covariance of $\gamma(x_t)$, as noted in an earlier work.[6] These inequalities have application to sampling error in traffic measurements.

Corollary: The carried load $E \mid \gamma(x_t) \mid (= E \mid x_t \mid)$ is given by

$$E|x_t| = \lambda \frac{d}{d\lambda} \log \Phi(\lambda),$$

and the variance of the load is

$$\left(\lambda^2 \frac{d^2}{d\lambda} + \lambda \frac{d}{d\lambda} \right) \log \Phi - \lambda^2 \left(\frac{d}{d\lambda} \log \Phi \right)^2.$$

Theorem 3: If $\alpha_x = \alpha_{|x|}$, then the probability of blocking is given by

$$1 - \frac{1}{\lambda} \frac{\sum_{a \in \gamma(S)} |a| \lambda^{|a|}}{\sum_{a \in \gamma(S)} \alpha_{|a|} \lambda^{|a|}}.$$

Proof: This proof comes directly from Theorem 2 and Ref. 1, since only completely blocked calls are rejected. Note that in Ref. 1, β counts the blocked calls, whereas here it counts only the completely blocked.

Remark 5: If the network is two-sided with N terminals on a side, then $\alpha_x = (N - |x|)^2$ and the loss is

$$1 - \frac{\dfrac{d}{d\lambda} \Phi}{\left[N^2 - (2N - 1)\lambda \dfrac{d}{d\lambda} + \lambda^2 \dfrac{d^2}{d\lambda^2} \right] \Phi}.$$

If it is one-sided, with T terminals, then $\alpha_x = \binom{T-2|x|}{2}$, and the loss is

$$1 - \frac{2\dfrac{d}{d\lambda}\Phi}{\left[T^2 - T - (4T - 6)\lambda\dfrac{d}{d\lambda} + 4\lambda^2\dfrac{d^2}{d\lambda^2}\right]\Phi}.$$

V. APPLICATION: CONCENTRATING OUTER SWITCHES

Let us consider a connecting network of the familiar type used in recent studies[1,4] and depicted in Fig. 1. For our purposes the $r \times r$ networks in the middle stage might themselves be multi-stage networks; we shall require only that they be rearrangeable, so that if $n \le m$ the whole network itself is rearrangeable. However, we are interested in the case $n > m$ of "concentrating outer switches". We pose and answer the question what is the probability of blocking if $n > m$ and if we follow a policy of complete rearrangement, that is, if the existing calls are assigned new routes whenever this is necessary to accommodate a new, not completely blocked, call. The blocking is of course due entirely to the concentrators. By Theorem 3 and the remark following it, it suffices to calculate the partition function.

It is evident that to calculate the partition function $\Phi(\lambda)$ it is enough to know $|L_k|$, the cardinality of the set L_k of realizable assignments in which k inlets are busy, for $k \ge 0$. Since the network can be obtained from a rearrangeable one by substituting concentrators for square
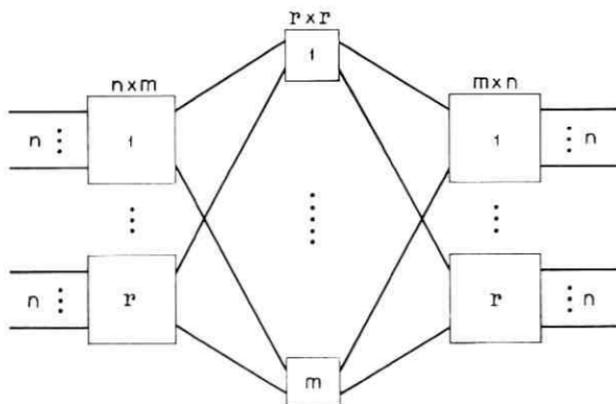
Fig. 1—A connecting network.

TABLE I—VALUES OF $\psi(r, k)$ FOR $n = 3$ AND $m = 2$

r \ k	0	1	2	3	4	5	6	7	8	9
1	1	3	3	0	0	0	0	0	0	0
2	1	6	15	18	9	0	0	0	0	0
3	1	9	36	81	108	81	27	0	0	0
4	1	12	66	216	459	648	594	324	81	0

outer switches, the realizable assignments will be precisely those involving m or fewer busy terminals on each outer switch. With m and n fixed, let $\psi(r, k)$ be the number of ways of choosing k inlet terminals from among the nr available, so that no more than m are from any one inlet switch. Let $i_1 , \cdots , i_k$ be such a choice of inlets, and let $o_1 , \cdots , o_k$ be a similar choice of outlets, also feasible in $\psi(r, k)$ ways. These chosen inlets can be mapped into the chosen outlets in $k!$ ways; each of these assignments will be realizable, and no others involving exactly k terminals will be. Hence in this case,

$$| L_k | = \psi^2(r, k)k!$$

To obtain a recurrence relation for $\psi(r, k)$, let us calculate $\psi(r+1, k)$ in terms of $\psi(r, j)$, $0 \leq j \leq k$. In this case we have one more outer switch available, and it can be seen that there are exactly $m + 1$ ways of using it: With no calls on it, with one call, with two, and so on up to m calls on it. In the first instance there are k on the other r switches, choosable in $\psi(r, k)$ ways; in the second there are $k - 1$ on the other r switches choosable in $\psi(r, k - 1)$ ways, and so on, up to m. If the $(r + 1)$th switch is to have j calls, these can be chosen in $\binom{n}{j}$ ways, but j must not exceed m. Thus

$$\psi(r + 1, k) = \sum_{j=0}^{m} \binom{n}{j} \psi(r, k - j), \qquad 0 \leq k \leq (r + 1)m.$$

It can be seen that $\psi(1, k) = \binom{n}{k}$ for $0 \leq k \leq m$. Introducing the generating function

$$\Psi_r(x) = \sum_{k=0}^{mr} x^k \psi(r, k),$$

we find at once that

$$\Psi_r(x) = \left(\sum_{j=0}^{m} \binom{n}{j} x^j \right)^r = \sum_{k=0}^{nr} x^k \psi(r, k).$$

We recall that if $A(x) = \Sigma a_n x^n$, a_n real, then

$$\Sigma\, a_n^2 x^n = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(x e^{i\theta}) A(e^{-i\theta})\, d\theta.$$

To calculate $\Phi(\lambda)$, finally, we note that

$$\Phi(\lambda) = \sum_{a\,\varepsilon\,A} \lambda^{|a|}$$

$$= \sum_{k=0}^{mr} \lambda^k |L_k|$$

$$= \sum_{k=0}^{mr} \lambda^k \psi^2(r, k) k!$$

$$= \sum_{k=0}^{mr} \lambda^k \psi^2(r, k) \int_0^\infty e^{-u} u^k\, du$$

$$= \int_0^\infty \sum_{k=0}^{mr} (\lambda u)^k \psi^2(r, k)\, du,$$

$$= \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^{\pi} \left(\sum_{j=0}^{m} \binom{n}{j} (u\lambda e^{i\theta})^j \right)^r \left(\sum_{j=0}^{m} \binom{n}{j} e^{-ij\theta} \right)^r d\theta\, du,$$

$$= \frac{1}{2\pi} \int_0^\infty \int_{-\pi}^{\pi} Q^r(u\lambda e^{i\theta}) Q^r(e^{-i\theta})\, d\theta\, du,$$

where $i = (-1)^{\frac{1}{2}}$ and

$$Q(x) = \sum_{j=0}^{m} \binom{n}{j} x^j$$

is the generating function of the binomial coefficients $\binom{m}{i}$ truncated at m. This formula expresses $\Phi(\cdot)$ in terms of known polynomials and constitutes a complete solution of the problem posed, since all interesting quantities can be obtained from the partition function $\Phi(\cdot)$.

For small values of m and n, the recurrence for $\psi(r, k)$ is easily run out to give numerical answers. In Tables I and II we give some values of $\psi(r, k)$ and $|L_k|$ for $n = 3$, $m = 2$, and $r = 1, \cdots, 4$.

TABLE II—VALUES OF L_k FOR $n = 3$ AND $m = 2$

r \ k	0	1	2	3	4	5
1	1	9	18	0	0	0
2	1	36	450	1944	1944	0
3	1	81	2592	10566		
4	1	144	8712			

VI. ACKNOWLEDGMENTS

The author wishes to thank A. Descloux and E. Wolman for comments. Results similar to those in this paper have been obtained by L. A. Bassalygo and B. S. Tsybakov in the Soviet Union.

REFERENCES

1. Beneš, V. E., *Mathematical Theory of Connecting Networks and Telephone Traffic*, New York: Academic Press, 1965.
2. Joel, A. E., Jr., "On Permutation Switching Networks," B.S.T.J., *47*, No. 5 (May–June 1968), pp. 813–822.
3. Waksman, A., "A Permutation Network," J. of the Association for Computing Machinery, *15*, No. 1 (January 1968), pp. 159–163.
4. Tsao-Wu, N., and Opferman, D. C., "On Permutation Algorithms for Rearrangeable Switching Networks," Conf. Rec. of IEEE Int. Conf. on Commun., Boulder, Colorado, June 9–11, 1969, pp. 1029–1034.
5. Beneš, V. E., "Optimal Routing in Connecting Networks Over Finite Time Intervals," B.S.T.J., *46*, No. 10 (December 1967), pp. 2341–2352.
6. Beneš, V. E., "A Thermodynamic Theory of Traffic in Connecting Networks," B.S.T.J., *42*, No. 3 (May 1963), pp. 567–607.

The Flat Plate Problem for a Semiconductor

By J. A. LEWIS

(Manuscript received March 20, 1970)

A closed-form solution for the potential in a current-free semiconductor surrounding a semi-infinite flat plate, carrying a small potential, is derived. We show that, just as in the case of the classical potential problem, there is a weak (square-root) singularity in the electric field at the edge of the plate. The solution also supplies an estimate of the effect of edge curvature, on edge field for a thick plate, found from equipotential contours. For a plate along the positive x-axis the normalized potential u is given by

$$u = e^{\nu}\, \frac{\text{erfc}\,(\xi + \eta)}{2} + e^{-\nu}\left[1 - \frac{\text{erfc}\,(\xi - \eta)}{2}\right],$$

where erfc *is the complementary error function;* ξ, η *parabolic coordinates, such that* $(\xi + i\eta)^2 = x + iy$; *and* x *and* y *are in units of Debye length.*

I. INTRODUCTION

In this paper we calculate the electrostatic potential in a current-free semiconductor surrounding a semi-infinite flat plate carrying a small potential. Our objective is to determine the nature of the singularity in the field at the plate edge. Such information should be useful in the application of finite difference methods to similar, but more complicated, problems. We find that the field near the edge has the same weakly singular behavior as in the classical potential problem, being inversely proportional to the square root of the distance from the edge.

It turns out, more or less fortuitously, that the present boundary value problem has a closed form solution, in terms of exponentials and error functions, so that it has some intrinsic mathematical interest. The solution was originally obtained by a very tortuous path. The problem was first attacked by the Wiener–Hopf technique. Then it was recognized that the inverse of the Fourier transform of the x-

derivative of the solution had a very simple form. After several changes
of integration variable, integration of this derivative led to the closed
form solution.

Here we bypass this involved procedure and derive the solution
directly, first factoring a particular solution and then transforming
to parabolic coordinates. This after-the-fact derivation is admittedly
artificial; in particular, up to now no other nontrivial boundary value
problem has been solved by this "method."

In the sections which follow we first derive the solution, then record
various special forms, and finally consider its behavior in the neighbor-
hood of the plate edge. The method of solution is discussed in Sections
II and III; the results are discussed in Sections IV and V.

II. THE BOUNDARY VALUE PROBLEM

The potential $u(x, y)$ due to a semi-infinite flat plate electrode in a
current-free semiconductor or a quasi-neutral, stationary plasma sat-
isfies the equation

$$\nabla^2 u = u_{xx} + u_{yy} = u, \tag{1}$$

the boundary condition

$$u(x, 0) = 1, \quad \text{for} \quad x \geqq 0, \tag{2}$$

for unit potential on the plate, the symmetry condition

$$u_y(x, 0) = 0, \quad \text{for} \quad x < 0, \tag{3}$$

and the limiting condition

$$u(x, y) \to 0, \tag{4}$$

far from the plate. In the above the Debye length λ_D has been taken
as the unit of length and the plate potential $-\varphi_0(< 0)$ as the unit of
potential. The potential $\varphi(X, Y)$ in dimensional form is then given by

$$\varphi(X, Y) = -\varphi_0 u(X/\lambda_D, Y/\lambda_D).$$

Equation (1) only holds for small electrode potential. Specifically,
for an n-type semiconductor, we require that

$$q\varphi_0/kT \ll 1,$$

for electronic charge q, Boltzmann constant k, and absolute tem-
perature T. In this case the Debye length is given by

$$\lambda_D = (\epsilon kT/q^2 N_d)^{\frac{1}{2}}$$

for semiconductor dielectric constant ϵ and donor number density N_d.

III. THE SOLUTION

Without preamble we now write down the solution to the above boundary value problem, justifying its form after the fact. Its structure is best displayed in a mixed notation. We find

$$u = e^{v}f(\xi + \eta) + e^{-v}[1 - f(\xi - \eta)], \qquad (5)$$

where ξ and η are parabolic coordinates, such that

$$x = \xi^2 - \eta^2, \qquad y = 2\xi\eta,$$

and the function f is given by

$$2f(s) = \text{erfc } s = 1 - \text{erf } s = 2\pi^{-\frac{1}{2}} \int_{s}^{\infty} e^{-t^2} \, dt.$$

By using the identities

$$\xi \pm \eta = r^{\frac{1}{2}}\left(\cos \frac{\theta}{2} \pm \sin \frac{\theta}{2}\right),$$

we may also express u in terms of polar coordinates $r = (x^2 + y^2)^{\frac{1}{2}}$, $\theta = \tan^{-1}(y/x)$.

Now let us attempt to unravel this rather involved expression for u. The occurrence of parabolic coordinates is not surprising. They are natural coordinates for flat plate problems with the family of parabolas $\eta = $ constant degenerating into the positive x-axis for $\eta = 0$ and the orthogonal family $\xi = $ constant giving the negative x-axis for $\xi = 0$. The way in which ξ and η appear in equation (5) and the factors $e^{\pm v}$ are a little more unusual. As we shall see, these two features are tied together.

The exponential factors explicitly display the behavior of the potential far out along the plate, for fixed y and large x, where the potential should approach the one-dimensional potential $e^{-|v|}$. For our purposes they are also essential in simplifying the form of the differential equation in parabolic coordinates. If we transformed equation (1) directly to parabolic coordinates, the resulting equation would be complicated by the presence of the "magnification factor" of the transformation, a function of ξ and η. On the other hand, with an exponential factor removed, the resulting differential equation does not contain the magnification factor. We have a special case of the following general result for equation (1):

Suppose that $u^*(x, y)$ is any particular solution of equation (1). Then $u(x, y) = u^*(x, y)v(x, y)$ also satisfies equation (1) if v satisfies

the equation

$$\nabla^2 v + 2\nabla(\ln u^*) \cdot \nabla v = 0.$$

Now, because the same factor occurs in both the Laplacian and in the product of gradients, this equation has the same form in all orthogonal coordinate systems; that is,

$$v_{\xi\xi} + v_{\eta\eta} + 2[(\ln u^*)_\xi v_\xi + (\ln u^*)_\eta v_\eta] = 0.$$

In general, this identity is of no help, since $\ln u^*$ is a complicated function of ξ and η; but, for $u^* = e^y$, $\ln u^* = y = 2\xi\eta$. Then we have

$$v_{\xi\xi} + v_{\eta\eta} + 4(\eta v_\xi + \xi v_\eta) = 0.$$

Notice the way that ξ and η occur in this equation. If we set

$$v(\xi, \eta) = f(\xi + \eta),$$

the equation reduces to the ordinary differential equation

$$f''(\xi + \eta) + 2(\xi + \eta)f'(\xi + \eta) = 0,$$

an equation satisfied by $f(\xi+\eta) = \text{constant}$ and by $f(\xi+\eta) = \text{erfc}\ (\xi+\eta)$. A similar result obtains for the multiplier of e^{-y}, except that now we find a function of $\xi - \eta$, $g(\xi - \eta)$, say. To satisfy the boundary condition on the plate, we must have $g(\xi) = 1 - f(\xi)$. We are then left with two arbitrary integration constants in f which are determined from the conditions for fixed y and large positive and negative x. The result is equation (5).

IV. THE POTENTIAL AND FIELD IN POLAR COORDINATES

In polar coordinates equation (6) becomes

$$u = \frac{e^{+r\sin\theta}}{2}\left\{1 - \text{erf}\left[r^{\frac{1}{2}}\left(\cos\frac{\theta}{2} + \sin\frac{\theta}{2}\right)\right]\right\}$$
$$+ \frac{e^{-r\sin\theta}}{2}\left\{1 + \text{erf}\left[r^{\frac{1}{2}}\left(\cos\frac{\theta}{2} - \sin\frac{\theta}{2}\right)\right]\right\}.$$

$$(6)$$

Various special values of u are of particular interest. Along the x-axis we have

$$u(x, 0) = \begin{cases} 1, & \text{for } x \geq 0 (\theta = 0, r = x), \\ \text{erfc } |x|^{\frac{1}{2}}, & \text{for } x \leq 0 (\theta = \pi, r = |x|), \end{cases}$$

while on the y-axis ($\theta = \pi/2$, $r = y$)

$$u(0, y) = \frac{e^{y}}{2} \operatorname{erfc} (2y)^{\frac{1}{2}} + \frac{e^{-y}}{2}.$$

Since for large y, $\operatorname{erfc} (2y)^{\frac{1}{2}} \sim (2\pi y)^{-\frac{1}{2}} e^{-2y}$, this vanishes exponentially for large y, as it should.

Near the edge of the plate (that is, for small r)

$$u \approx 1 - 2(r/\pi)^{\frac{1}{2}} \sin \frac{\theta}{2}.$$

The x and y components of electric field are given by

$$u_x = (\pi r)^{-\frac{1}{2}} e^{-r} \sin \frac{\theta}{2}, \tag{7}$$

$$
\begin{aligned}
u_y = {}& -(\pi r)^{-\frac{1}{2}} e^{-r} \cos \frac{\theta}{2} \\
& + \frac{e^{+r \sin \theta}}{2} \left\{ 1 - \operatorname{erf} \left[r^{\frac{1}{2}} \left(\cos \frac{\theta}{2} + \sin \frac{\theta}{2} \right) \right] \right\} \\
& - \frac{e^{-r \sin \theta}}{2} \left\{ 1 + \operatorname{erf} \left[r^{\frac{1}{2}} \left(\cos \frac{\theta}{2} - \sin \frac{\theta}{2} \right) \right] \right\}. \tag{8}
\end{aligned}
$$

Note the simple form for u_x, a separable solution of equation (1) in polar coordinates. The explicit form for u was in fact originally derived by first calculating u_x and then integrating with respect to x, a rather laborious procedure. According to equations (7) and (8), the field near the edge of the plate is weakly singular; that is, $|\nabla u| \sim (\pi r)^{-\frac{1}{2}}$ for small r, just as in the case of the classical potential problem. Along the x-axis

$$
u_x(x, 0) = \begin{cases} 0, & \text{for } x > 0, \\ (\pi |x|)^{-\frac{1}{2}} e^{-|x|}, & \text{for } x < 0, \end{cases}
$$

$$
u_y(x, 0) = \begin{cases} -(\pi x)^{-\frac{1}{2}} e^{-x} - \operatorname{erf} (x)^{\frac{1}{2}}, & \text{for } x > 0, \\ 0, & \text{for } x < 0. \end{cases}
$$

Note that u_y tends very rapidly along the plate to the one-dimensional value $u_y = -1$. The equipotentials near the plate then rapidly become straight and parallel to the plate. Any one of them can be chosen as an electrode in the form of a thick plate with the electric field large, but not infinite, at its edge, which now has a finite radius of curvature. Thus, at least for this family of plates, we can calculate the maximum field strength as a function of edge curvature.

V. THE EFFECT OF EDGE CURVATURE

Suppose that $x = x(y)$ is the equation of the equipotential $u = u_o$; that is,

$$u[x(y), y] = u_o .$$

Then

$$u_x x' + u_y = 0,$$

and

$$u_x x'' + u_{xx} x'^2 + 2u_{xy} x' + u_{yy} = 0.$$

At the edge $y = 0$

$$x = x(0) = x_o , \qquad x'(0) = 0,$$

where $\mathrm{erfc} \mid x_o \mid^{\frac{1}{2}} = u_o$. Thus the radius of curvature of the edge is given by

$$R = 1/\mid x''(0) \mid \; = \; \mid u_x(x_o , 0)/u_{yy}(x_o , 0) \mid$$
$$= \; \mid u_x(x_o , 0) \mid / \mid u(x_o , 0) - u_{xx}(x_o , 0) \mid,$$

using the differential equation (1). Now

$$u(x, 0) = \mathrm{erfc} \mid x \mid^{\frac{1}{2}},$$

for $x \leqq 0$, so that

$$u_x(x, 0) = e^{-\mid x \mid}/(\pi \mid x \mid)^{\frac{1}{2}},$$

and

$$u_{xx}(x, 0) = (1 + \tfrac{1}{2} \mid x \mid)u_x(x, 0).$$

Thus

$$R = 2 \mid x_o \mid / \mid 1 + 2 \mid x_o \mid (1 - 1/w)\mid, \tag{9}$$

where the normalized gradient

$$w = \mid \nabla u(x_o , 0) \mid/u(x_o , 0) = e^{-\mid x_o \mid}/(\pi \mid x_o \mid)^{\frac{1}{2}} \, \mathrm{erfc} \mid x_o \mid^{\frac{1}{2}}. \tag{10}$$

If equations (9) and (10) are evaluated for various x_o , one obtains the edge field strength due to unit potential applied to a plate with the shape $u = u_o$, as a function of the edge radius of curvature, shown in Fig. 1. As R increases, one might expect that the edge field should approach the field on a circular conductor of radius R. This field, shown as a dashed curve in Fig. 1, is given by $K_1(R)/K_0(R)$, where
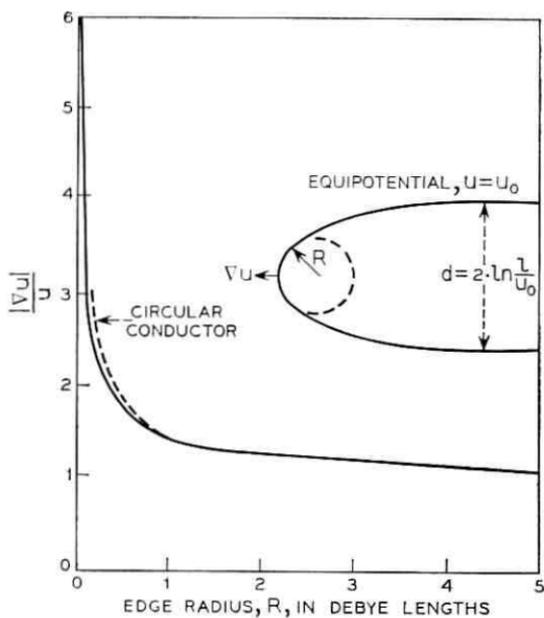
Fig. 1—The edge gradient as a function of edge radius.
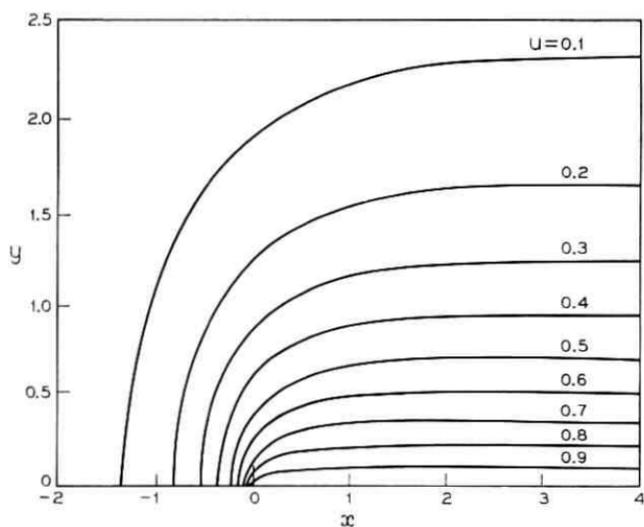
Fig. 2—Equipotentials around a flat plate in a semiconductor.

K_0 and K_1 are the modified Bessel functions of the second kind. Except for very small edge radius, the fields very nearly coincide.

Figure 1 also shows a sketch of a typical equipotential, that is, a thick plate of the family $u = u_o$. The shape shown is typical of moderately small edge radius R. For large R the plate thickness d tends to $2R$, so that thick plates in the family tend to have semicircular noses. On the other hand, thin plates $(u_o \rightarrow 1)$ have "wedge-shaped" noses, R tending to zero like $(1 - u_o)^2$, while d tends to zero like $1 - u_o$.

Figure 2 shows the exact shapes, for $0.1 \leq u_o \leq 1$. Note the sharpness of the equipotential $u = 0.9$, compared with $u = 0.1$, even allowing for the difference in horizontal and vertical scales. For small edge radius then, unless the plate under consideration has a shape similar to the above, the present analysis is not likely to furnish useful estimates of the edge field.

VI. ACKNOWLEDGMENT

Some Calculations on Coupling Between Satellite Communications and Terrestrial Radio-Relay Systems Due to Scattering by Rain

By L. T. GUSLER and D. C. HOGG

Interference coupling due to rain is an important factor in coordinating the shared use of frequencies between satellite-communications and terrestrial microwave radio-relay systems. Calculations of the coupling between a satellite ground station and a radio-relay station due to scattering by rain are given for the frequencies 4, 6, 11, 18.5 and 30 GHz. Several models of the rain environment are considered, one in which the rain falls uniformly over the path, and others involving localized showers in the earth-station beam. The beam of the earth-station antenna is taken to be elevated 30° above the local horizon. Interstation distances from 5 km to 200 km and rain rates up to 400 mm/hr are considered. The analysis includes the near field of the earth-station antenna and shows that attenuation by the rain on the path plays an important role at the higher frequencies. The greatest coupling occurs when the beam of the earth-station antenna is pointed opposite to the azimuth of the relay station. In that case, when the radio-relay antenna response in the direction of the earth station is isotropic, the maximum couplings are −132 dB and −164 dB at 4 GHz with separations of 5 km and 200 km respectively, for a rain rate of 400 mm/hr.

I. INTRODUCTION

Sharing of microwave bands by terrestrial and satellite-communication systems has stimulated investigation of the factors which lead to mutual interference. Of course, antenna characteristics are of prime importance; very low immediate and far side lobes in the antenna

radiation patterns are mandatory in reducing coupling between stations of the two systems. However, even with ideal antennas, there are certain propagation mechanisms that can introduce coupling, one being scatter by precipitation, in particular, rain.

Much knowledge of scattering by rain has been gained through studies with weather radar.[1] Relationships between the backscattering coefficient, the size of the raindrops, the rain rate, and the radio wavelength have been established. These commonly accepted relationships are used in our calculation. Attenuation by the rain is also germane to the calculation, especially at frequencies exceeding 10 GHz, because power from one station that is scattered by a particular volume of rain may well pass through additional precipitation before reaching the other station. The attenuation constants used here are based on measurements with terrestrial systems;[2] for high rain rates they agree with calculations using the Mie theory with a Laws–Parsons drop-size distribution.

In these calculations, the satellite-communications antenna is assumed to be large in terms of wavelengths and to have a well-tapered quasigaussian illumination. As mentioned above, the latter is necessary in practice to produce low immediate and far side-lobe levels. However, this assumption is also advantageous because when the antenna illumination taper is taken to be gaussian, the radiation pattern is readily described analytically at any cross section along the beam from the aperture to infinity.[3,4] This is an important factor because the "near-field zone" of a large space communication antenna may extend for miles; the coupling due to scattering by rain within this near zone can be larger than in the far zone of the pattern and, of course, power scattered in the near zone is not subject to as much attenuation in the path to the antenna.

Several models of the rain environment are discussed to greater and lesser extents. An obvious model is one of uniform rain extending over the entire path between and beyond both stations; this is referred to as Case I in the text. However, this is an unrealistic model where very high rain rates are involved because very heavy rainfall occurs in showers that are quite localized spatially.[5] A more realistic model for heavy rains is that of a shower over only the satellite-communications antenna. This model, treated in Cases II and III, obviates the effect of attenuation over the whole path, which is quite dominant in the case of uniform rain. Case IV, the one that produces the largest coupling between the stations, involves a wall of rain which approaches the satellite communication antenna from along a line which is an

extension of the path beyond that antenna. Since the radiation patterns of both antennas look toward this wall of rain, the effect of attenuation is small.

The computations have been made for rain rates up to 400 mm/hr. The largest coupling occurs when the wall of rain is present (Case IV); this amounts to -132 dB at 4 GHz for a rain rate of 400 mm/hr and a separation of 5 km between stations and applies to cases where the level of the radiation pattern of the terrestrial radio-relay antenna (looking in the direction of the ground station) is isotropic.

It is of interest to compare the co-channel interference resulting from these couplings (of the order -140 dB) with the thermal noise level in the receiver of a typical satellite-communication ground station. Suppose the system noise temperature of a 4-GHz ground station is 70 kelvins within a 40-MHz band; the noise power referred to the input is $P_n = kT_eB$, namely -134 dBW, with the signal typically 25 dB above this value. Radio-relay systems have transmitted power rarely exceeding 10 watts; thus with -140 dB coupling, the interference level would be -130 dBW, 4 dB above the thermal noise. However, if the radio-relay antenna pattern has a sidelobe level say 10 dB above isotropic, the interference would exceed the thermal noise by 14 dB. Since in many systems it is desirable that the interference be several decibels below the thermal noise, couplings of the order -140 dB are indeed significant.

II. GENERAL ANALYSIS

Figure 1 shows a ground-station antenna (T) of a space-communications system with a beam elevated $30°*$ (θ_1) above a path of length d to a station (R) of a terrestrial radio-relay system. The effective radius of the earth, a, and the altitude, h, at which the rain originates are also shown.

For descriptive purposes, let the ground-station be transmitting at wavelength, λ. Because of a well-tapered (quasi-gaussian) illumination, the flux on axis decreases with distance r_1 from the antenna to a volume element of the beam (see Fig. 1) in the following manner:

$$S(r_1) = \frac{E_{or_1}^2}{\eta} = \left[\frac{E_o^2}{\eta r_1^2}\right]\left[\frac{1}{r_1^2} + \left(\frac{\lambda}{\pi \bar{\rho}^2}\right)^2\right]^{-1} \tag{1}$$

η being the free-space impedance and $\bar{\rho}$ the radius at which the aperture

---

* Thirty degrees is typical for the elevation angle of a mid-latitude station observing a synchronous satellite.
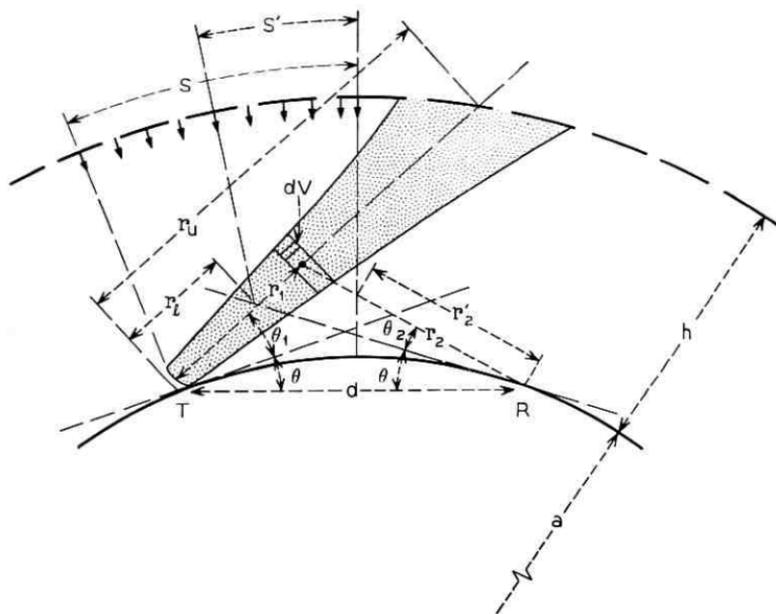
Fig. 1—Geometry of the path; $\theta_1 = 30°$, $\theta = d/2a$, $a = 8500$ km, and the rain originates at height h. s and s' are dimensions of showers (see text).

field is ϵ^{-1} of its axial value, E_o. The volume element is $dV = 2\pi\rho \, d\rho \, dr_1$ where ρ is the radial coordinate of the beam. The flux at a point in the volume element is

$$S(\rho, r_1) = \frac{E_{or_1}^2}{\eta} \exp\left(\frac{-2\rho^2}{\rho_1^2}\right)$$

where ρ_1 is the radius of the annular volume element at which the field is ϵ^{-1} of its axial value (E_{or_1}). Most of the flux is contained within the beam cross section of radius ρ_1, and ρ_1 will be used to define the boundary of the beam in what follows.

The power scattered into an antenna at R (see Fig. 1) with an effective area $A_2(\theta_2)$ in direction θ_2 toward dV is therefore

$$P_R = \int_{r_l}^{r_u} \int_o^{\rho_1} S(\rho, r_1) \frac{\sigma_v}{4\pi r_2^2} A_2(\theta_2) \exp\left[-\alpha(r_1 + r_2)\right] 2\pi\rho \, d\rho \, dr_1 \qquad (2)$$

where α is the attenuation coefficient of the rain and r_2 the distance from the volume element dV to the receiver at R. The limits r_l and r_u on r_1 depend upon the geometry of the rainstorm, and σ_v is the scattering coefficient per unit volume of the rain; these are discussed later.

Substitution of $S(\rho, r_1)$ in equation (2) and integration with respect to ρ results in

$$\int_{r_l}^{r_u} \frac{E_{or_1}^2}{\eta} \frac{\pi \rho_1^2}{2} (1 - \epsilon^{-2}) \frac{\sigma_v}{4\pi r_2^2} A_2(\theta_2) \exp\left[-\alpha(r_1 + r_2)\right] dr_1 .$$

However, the gaussian beam expands such that

$$\rho_1^2 = \bar{\rho}^2 \left[1 + \left(\frac{\lambda r_1}{\pi \bar{\rho}^2}\right)^2\right] \quad \text{and one has, using equation (1)}$$

$$P_R = \frac{E_o^2}{\eta} \frac{\pi \bar{\rho}^2}{2} (1 - \epsilon^{-2}) \int_{r_l}^{r_u} \frac{\sigma_v}{4\pi r_2^2} A_2(\theta_2) \exp\left[-\alpha(r_1 + r_2)\right] dr_1 . \tag{3}$$

But the transmitted power is

$$P_T = \frac{E_o^2}{\eta} \int_o^{\bar{\rho}} \exp\left(\frac{-2\rho^2}{\bar{\rho}^2}\right) 2\pi \rho \, d\rho = \frac{E_o^2}{\eta} \frac{\pi \bar{\rho}^2}{2} (1 - \epsilon^{-2}).$$

Therefore, from equation (3), one obtains

$$\frac{P_R}{P_T} = \int_{r_l}^{r_u} \frac{\sigma_v}{4\pi r_2^2} A_2(\theta_2) \exp\left[-\alpha(r_1 + r_2)\right] dr_1 . \tag{4}$$

Of course equation (4) takes into account the near-field zone of the ground-station antenna; equation (4) is independent of the aperture diameter of that antenna.

For the computations that follow, equation (4) was simplified in two respects. First, the response of the radio relay antenna was taken to be isotropic that is, $A_2 = \lambda^2/4\pi$, independent of θ_2. Other cases computed consider particular configurations where the beams of typical radio-relay antennas with given radiation patterns $A_2(\theta)$ are pointed toward the space station. In some of these cases, the scattered power is simply increased by the gain of that antenna, but instances where this rule does not hold are discussed toward the end of the paper.

Secondly, the scattering by the raindrops is taken to be isotropic, and we use a relationship[6] based upon the Laws and Parsons drop-size distribution

$$\sigma_v = 0.18 \frac{\pi^5}{\lambda^4} 10^{-15} R^{1.6} (m^{-1}),$$

R being the equivalent rain rate in millimeters per hour.* It is known that large raindrops do not scatter isotropically, especially when the

---

* Rigorously, one should evaluate the Stokes parameter and the differential scattering cross sections for all drops in a given rain.

operating wavelength approaches the millimeter band, but the error is not large[6] for the wavelengths considered here. Perhaps more serious is that the drop-size distribution (and indeed the equivalent rain rates themselves) are not well determined for the upper troposphere. Here we will be guided by rain rates measured at ground level,[7] and carry this parameter to 400 mm/hr in certain cases.

The computations are made for the frequencies 4, 6, 11, 18.5 and 30 GHz and the corresponding attenuation coefficients, α, used here (taken in part from Refs. 2 and 6) are: $0.23 \; 10^{-6} R^{1.1}$, $0.69 \; 10^{-6} R^{1.22}$, $2.5 \; 10^{-6} R^{1.28}$, $16 \; 10^{-6} R^{1.09}$ and $55 \; 10^{-6} R^{1.0}$, in units of inverse meters, R being in millimeters per hour.

The several sections that follow deal with various rain geometries. The rain is assumed to originate at a height, h, of four kilometers in all cases.

III. CASE I—RAIN UNIFORM OVER THE PATH BETWEEN THE TWO STATIONS

3.1 Computation

The lower and upper limits in equation (4) are

$$r_l = d \sin \theta / \sin (\theta_1 + 2\theta) \quad \text{and}$$

$$r_u = (a^2 \sin^2 \theta_1 + 2ah + h^2)^{\frac{1}{2}} - a \sin \theta_1$$

for this case. These limits apply specifically when one station is beyond the horizon from the other, a being the effective earth radius and θ_2 the angle between the tangent plane and the chord between the stations as shown in Fig. 1. In the event the two stations are within line-of-sight of one another, the above limits reduce to

$$r_l = 0 \quad \text{and} \quad r_u = h/\sin \theta_1 \; .$$

Computations have been made using both sets of limits (for this case $r_2 = | r_1^2 + d^2 - 2r_1 d \cos (\theta_1 + \theta) |^{1/2}$). At the distances of transition from line-of-sight to beyond-horizon conditions ($d > 50$ km) the data are found to behave properly.

3.2 Results

Plots of coupling between T and R are shown in Figs. 2a through e for the frequencies 4, 6, 11, 18.5 and 30 GHz, interstation separation being a parameter. The abscissa is the equivalent rain rate in millimeters per hour, assumed uniform between stations to a height of four kilometers.
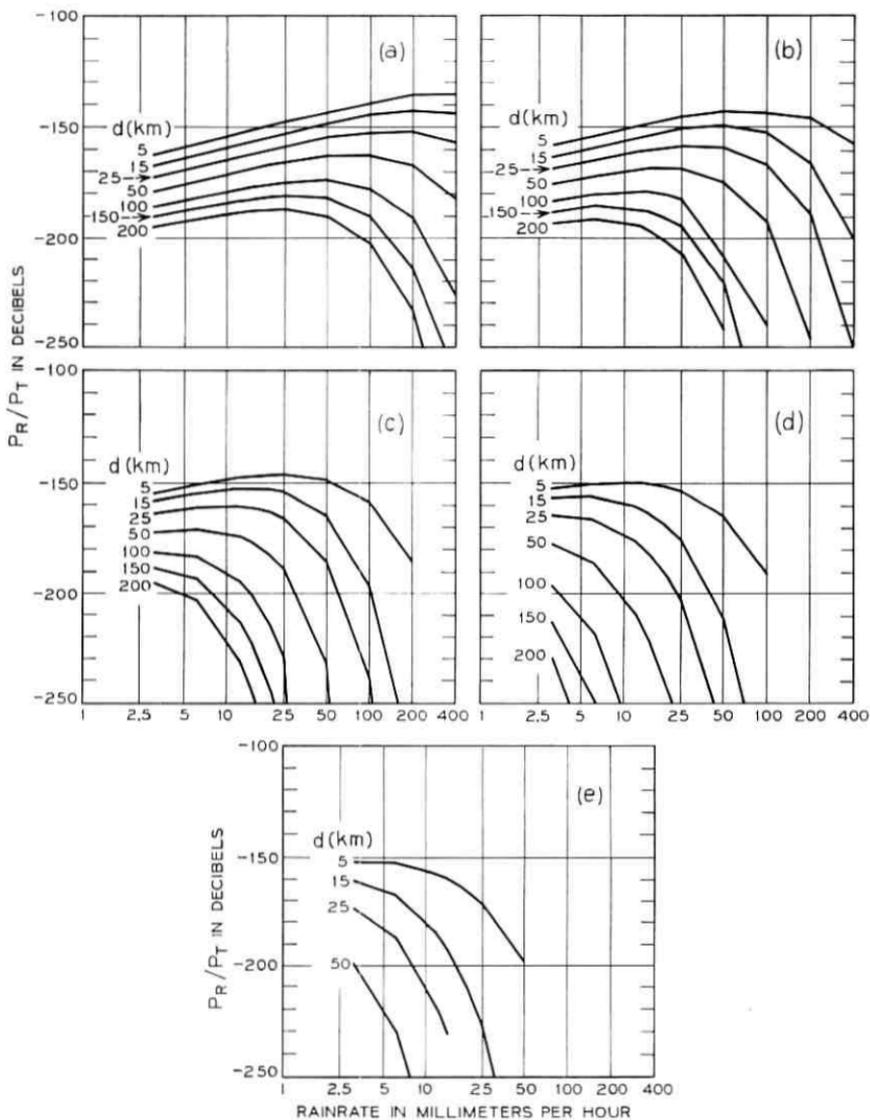
Fig. 2—Coupling for uniform rain originating at a height $h = 4$ km versus rain rate; elevation of ground-station antenna beam is $30°$; isotropic antenna at R. (a) 4 GHz, (b) 6 GHz, (c) 11 GHz, (d) 18.5 GHz, and (e) 30 GHz.

The curves of Fig. 2 are computed for $A_2(\theta) = \lambda^2/4\pi$ in equation (4); that is, for an isotropic receiving antenna at R.* Data for an antenna with a gain of the order 40 dB beamed from R toward T have also been computed; in this case, often one need only add the gain of the antenna (in decibels) to the curves of Fig. 2, a procedure discussed in more detail in Section VII.

An obvious characteristic of the curves in Fig. 2 is that they possess a maximum. As one proceeds from Fig. 2a through e (4 GHz to 30 GHz) the maxima move to ever lower rain rates. This feature evidences the effect of attenuation. Likewise for a given set of curves (Fig. 2a, for example) the maximum occurs at ever lower rain rates as the path length increases, as one would expect in an attenuating medium.

Although these results for uniform rain are instructive, some of them represent situations which do not occur in nature. For example, it is possible that rain rates of the order 200 mm/hr occur over a 5 km path; but it is highly unlikely (or impossible[7]) that such high rates occur over a path of length 200 km. Thus the high rain rate and long path length portion of the data in Fig. 2 should be considered somewhat academic. At the higher frequencies this portion of the data is of little significance anyway (Fig. 2d, for example).

The maximum value of the coupling for Case I appears in Fig. 2a (4 GHz) for the shortest interstation distance (5 km) and the highest rain rate considered (400 mm/hr); it is −136 dB. But maxima of the order −150 dB occur for higher frequencies and the same pathlength at lower rain rates.

IV. CASE II—LOCALIZED SHOWER AT THE SATELLITE STATION

4.1 *Computation*

As shown in Fig. 1, one can introduce another parameter, s, which defines the extent of a shower along the path from the ground-station T toward R. As discussed above, very heavy rainfall is spatially restricted, therefore Case II is more realistic than Case I in the high rain rate regime.

For computation of Case II, both the limits on r_1 and the exponent of the exponential in equation (4) must be modified, the latter because much of the path (r_2) from R to the beam of the ground-station antenna no longer introduces attenuation, the rain being localized near T. Let r_2' be that portion of r_2 with no precipitation. Then the exponent in

---

* Representative of the case where the back or far side lobes of a radio-relay antenna are directed toward the ground station.

equation (4) becomes $-\alpha(r_1 + r_2 - r_2')$ where

$$r_2' = d \, \frac{\left[\sin^2 \theta - \dfrac{1}{\cos \theta} + \dfrac{s}{d}\right]}{\left[\sin^2 \theta + \dfrac{1}{\cos \theta} + \sin \theta \sin (\theta + \theta_2)\right] \cos (\theta + \theta_2)} \, .$$

The limits in equation (3) for this case are

$$r_l = d \sin \theta \sin (\theta_1 + 2\theta) \quad \text{and} \quad r_u = s/\cos \theta_1$$

which for line-of-sight conditions $(\theta = 0)$ reduce to

$$r_l = 0, \quad \text{and} \quad r_u = s/\cos \theta_1 \, .$$

In both cases r_u, and also the effective height of the shower, depend upon its lateral extent, s. Again, the computations show that the transition from line-of-sight to beyond-horizon conditions is well behaved. The expression for r_2 is given in Section III.

4.2 *Results*

Plots of coupling for this case are given in Figs. 3a through e, the antenna at R again being taken to be isotropic. The shower has an extent one kilometer along the path from T and a height of 4 kilometers.

At the higher frequencies in Fig. 3 maxima in the coupling appear, somewhat as in Fig. 2, but the former occur at much higher rain rates than the latter. For example, at 11 GHz, the maxima in Fig. 3c are at about 100 mm/hr, whereas in Fig. 2c no maximum occurs above the rain rates of 25 mm/hr. Since one is dealing with a localized shower it is not surprising that the effect of attenuation sets in at rain rates higher in Fig. 3 than in Fig. 2.

The maximum value attained by the coupling is about -135 dB as shown at 400 mm/hr on Figs. 3a and b for an interstation separation of 5 km. However, maxima of the order -140 dB occur in Figs. 3c through 3e at lesser rain rates.

A feature of the data in Fig. 3 is that the coupling decreases drastically as the distance increases from 50 to 100 km. This decrease is associated with the effective height of the shower, as governed by its width s; beyond $d = 50$ km the scattering volume decreases rapidly. Computations not shown here have been made for showers of lateral extent up to 8 km and, as one would expect, such data lie between those of Figs. 2 and 3. As the extent of the shower increases, the rapid decrease in coupling for distances beyond 50 km becomes less marked.
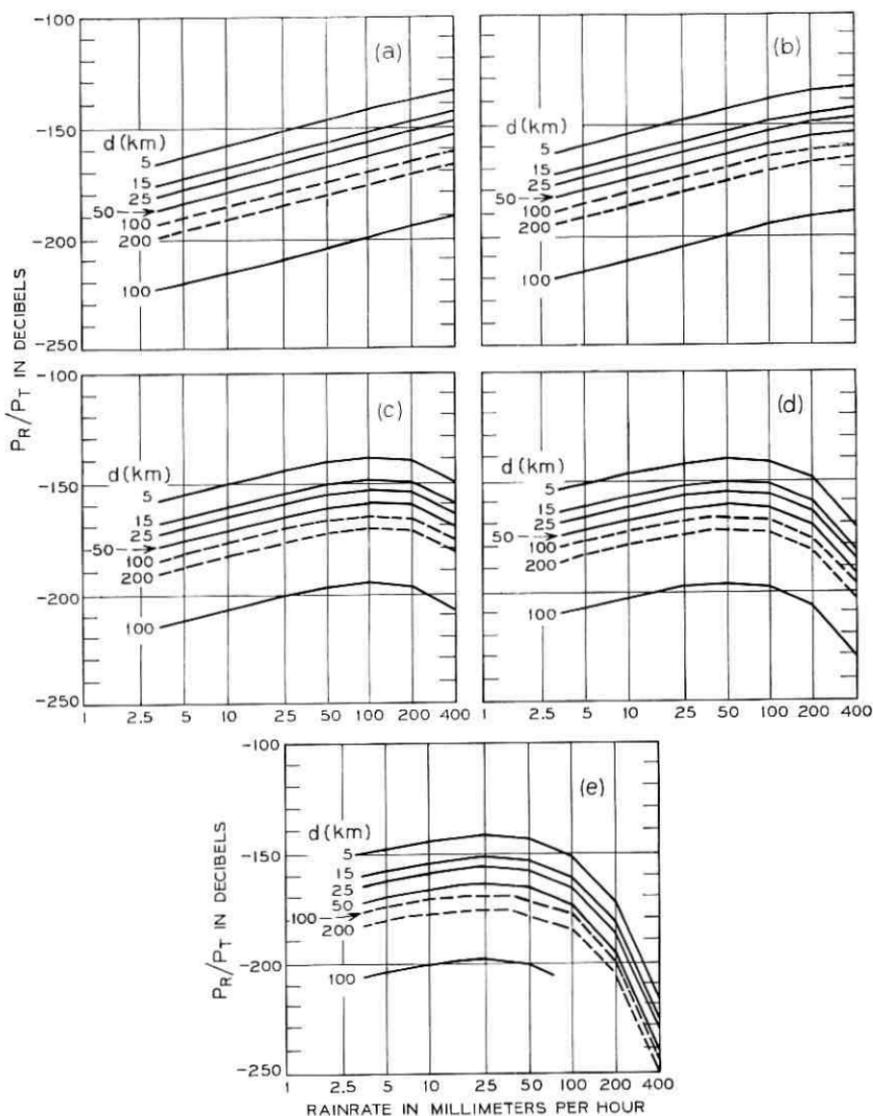
Fig. 3—Coupling for a shower ($S = 1$ km) originating at a height of 4 km above ground station (T) versus rain rate; elevation of ground-station antenna beam is 30°; isotropic antenna at R. (a) 4 GHz, (b) 6 GHz, (c) 11 GHz, (d) 18.5 GHz, (e) 30 GHz. The dashed lines correspond to a shower $S' = 1$ km in extent located at the intersection of the ground-station beam and the tangent plane of the earth (see Fig. 1).

V. CASE III—LOCALIZED SHOWER AT BEAM INTERSECTION

5.1 *Computation*

A modification of the localized shower model (Case II) is achieved by moving the shower (extent s') along the path toward R thereby locating the rain at the intersection of the earth station beam and earth tangent at R as shown in Fig. 1. This geometry results in a scattering volume for the over-the-horizon cases ($d > 50$ km) much larger than that considered in Case II while retaining the limited rain extent.

For the computations of Case III, the lower limit in equation (4) again becomes

$$r_l = d \sin \theta / [\sin (\theta_1 + 2\theta)].$$

The upper limit for this case depends upon whether the earth-station beam emerges from the vertical boundary of the precipitation or from the upper boundary defined by h.

Introducing two new parameters θ' and θ'' where

$$\theta' = 2\theta - \tan^{-1} \{[r_l^2 + d^2 - 2r_l \, d \cos (\theta_1 + \theta)]^{1/2}/a\}$$

and

$$\theta'' = s'/a,$$

the upper limit where the beam emerges from the vertical boundary is

$$r_u = a \sin (\theta' + \theta'')/\cos (\theta_1 + \theta' + \theta''),$$

or where the beam emerges from the upper boundary

$$r_u = (a^2 \sin^2 \theta_1 + 2ah + h^2)^{1/2} - a \sin \theta_1 .$$

For computational purposes, the smaller upper limit is the correct value for a given set of conditions. The path length r_2 from the differential scattering volume is given by

$$r_2 = [r_1^2 + d^2 - 2r_1 \, d \cos (\theta_1 - \theta)]^{1/2}.$$

The exponent in equation (4) must also be further modified because portions of both r_1 and r_2 no longer introduce attenuation. Let r_1' and r_2' be those portions of r_1 and r_2 with no precipitation. The exponent in equation (4) then becomes $-\alpha(r_1 + r_2 - r_1' - r_2')$ where

$$r_1' = r_l ,$$

$$r_2' = [a \sin (2\theta - \theta' - \theta'')]/\cos (\theta_2 + 2\theta - \theta' - \theta''),$$

and

$$\theta_2 = \sin^{-1}\{[(r_1 - r_l)\sin(\theta_1 + 2\theta)]/r_2\}$$

is the elevation angle between r_2 and the tangent plane at R.

5.2 Results

The coupling for distances of 100 and 200 km is shown with the data for Case II as broken lines in Figs. 3a through e, the conditions being identical to Case II except for the lateral displacement of the rain region to the beam intersection. The coupling for distances of 50 km or less have been omitted since the results and configurations are essentially the same as Case II for line-of-sight conditions of 50 km or less.

Comparing the results for Case II and III, the coupling at distances greater than 50 km is greatly increased over the corresponding results for Case II due to the larger scattering volume. At distances greater than 200 km, the scattering volume would be limited by the assumed rain height of 4 km and one would expect a drastic decrease in coupling beyond 200 km.

VI. CASE IV—A WALL OF RAIN NORMAL TO THE PATH EXTENDED BEYOND THE SATELLITE STATION

6.1 Computation

In cases I through III, the region of rain is located on the path between the ground station and the relay station (Fig. 1). However, "off-path" orientations of the ground-station beam relative to the terrestrial radio-relay station are typical of actual system configurations. We examine one such "off-path" orientation in which the ground-station beam points directly away from the radio relay in azimuth. The region of rain is then located on the path extended beyond the earth station, and the energy may be scattered from the boundary of the rain without additional attenuation on the path to the relay station.

The model used for analysis is shown in Fig. 4. A wall of rain normal to the path is located at the intersection of the axis of the ground-station beam and the earth tangent plane from the relay station.

For computation of this case, the limits in equation (4) become

$$r_l = d\sin\theta/\sin(\theta_1 - 2\theta)$$

and

$$r_u = (a^2\sin^2\theta_1 + 2ah + h^2)^{1/2} - a\sin\theta_1.$$
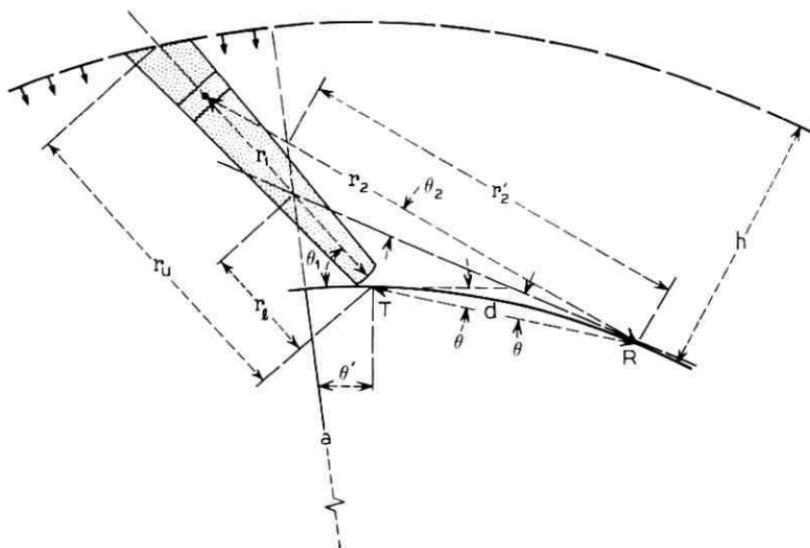
Fig. 4—Geometry for a rain originating at altitude h on the path extended beyond T; $\theta_1 = 30°$, $\theta = d/2a$, $a = 8500$ km.

The path length r_2 from the differential scattering volume is given by

$$r_2 = [r_1^2 + d^2 + 2r_1 d \cos (\theta_1 - \theta)]^{1/2}.$$

The portions of paths r_1 and r_2 with no precipitation, r_1' and r_2' are given by

$$r_1' = r_1 - r_l$$

$$r_2' = a \sin (\theta' + 2\theta) \cos (\theta' + 2\theta + \theta_2)$$

where the angles θ_2 and θ' (Fig. 4) are

$$\theta_2 = \sin^{-1} \{[(r_1 - r_l) \sin (\theta_1 - 2\theta)]r_2\}$$

and

$$\theta' = \tan^{-1} \{[r_l^2 + d^2 + 2r_l d \cos (\theta_1 - \theta)]^{1/2}/a\} - 2\theta.$$

As for Case III, the exponent in Equation (4) then becomes

$$-\alpha(r_1 + r_2 - r_1' - r_2').$$

6.2 *Results*

Plots of the coupling for the extended path scattering are shown in Figs. 5a through e for a rain height of 4 km and earth station elevation
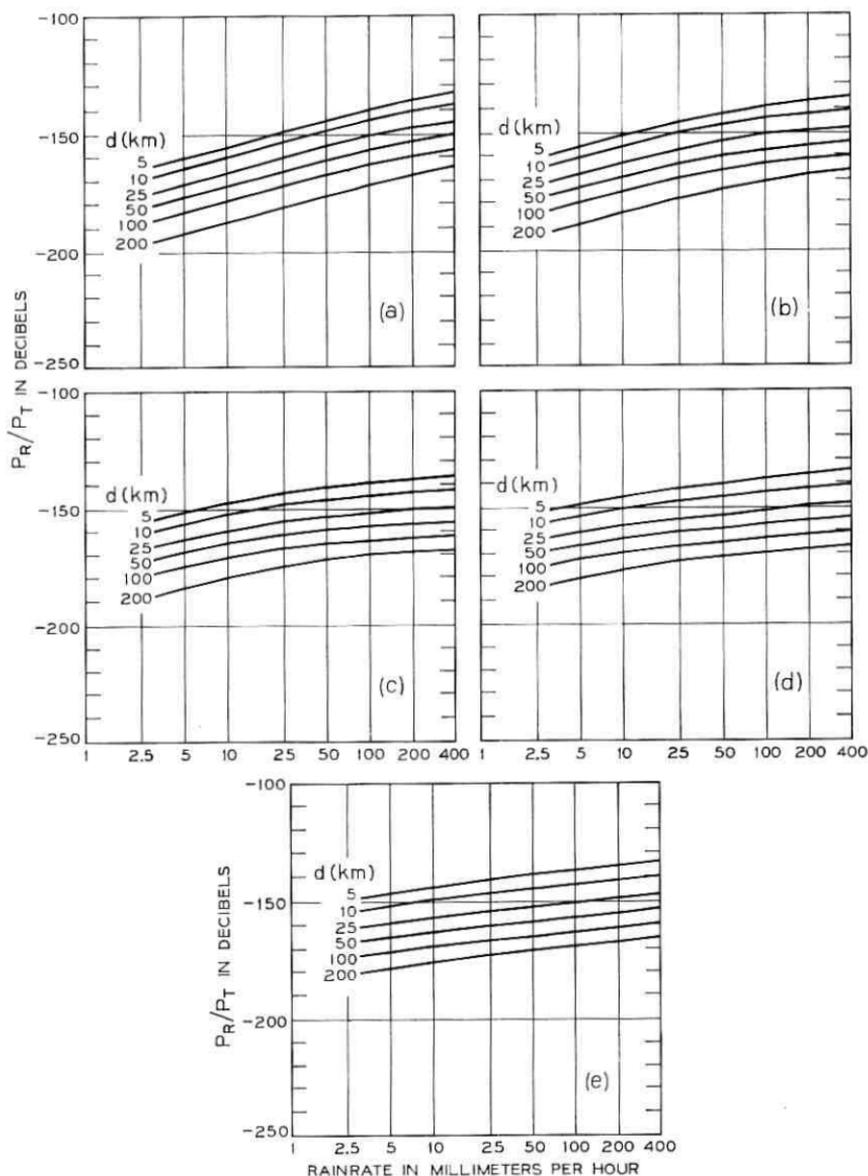
Fig. 5—Coupling for rain originating at a height $h = 4$ km on the extended path versus rain rate; elevation of ground-station antenna beam is $30°$; isotropic antenna at R. (a) 4 GHz, (b) 6 GHz, (c) 11 GHz, (d) 18.5 GHz, (e) 30 GHz.

of 30 degrees. The most notable feature is that the coupling increases with rain rate for all frequencies over the range of rain rates considered. This is in contrast with the previous cases where the curves possessed a maximum at the higher frequencies.

Maximum coupling, approximately -132 dB, occurs at 4 GHz for a separation of 5 km and a rain rate of 400 mm/hr. However, the coupling is reasonably independent (within about 4 dB) of frequency for rain rates greater than about 100 mm/hr. At lower rain rates, the coupling increases with frequency as one would expect; for example, at 5 mm/hr and 100-km separation, the coupling is about -183 dB at 4 GHz and -171 dB at 30 GHz.

VII. EFFECT OF DIRECTIVE ANTENNAS

The effect of the directivity of typical terrestrial radio-relay antennas on the coupling has been computed for all cases treated above. For each case, the beam of the radio relay antenna is assumed to be directed toward the ground station, that is, along the horizon plane (dashed) in Figs. 1 and 4. This is a very pessimistic condition and one hopefully not encountered too often in practice.

The response pattern of the radio-relay antenna $A_2(\theta_2)$ in equation (4) is assumed to be of the form

$$A_2(\theta_2) = A_o[\exp(-K_1\theta_2^2) + \exp - (K_2 + K_3\theta_2)], \qquad (5)$$

θ_2 being the angle measured in degrees from the axis of the relay station beam and A_o the effective area. The first term in equation (5) is the familiar gaussian approximation for the main beam. The second term represents the envelope of the far side lobes of the pattern using constants K_2 and K_3 of 5.5 and 0.115 respectively.

At 4 and 6 GHz, the radio relay antenna is assumed to have a beamwidth of one degree ($K_1 = 2.8$) with 43 dB gain. At 11, 18 and 30 GHz, the beamwidth is taken to be two and a half degrees ($K_1 = 0.41$) with 36 dB gain.

In what follows, the gain realized or coupling relative to the isotropic case is plotted versus rain rate. Thus, the coupling for the directive case is obtained by adding the gain realized to the coupling for the corresponding isotropic case; thus, if the gain realized is 30 dB and the isotropic coupling is -130 dB, the coupling in the directive case is -100 dB.
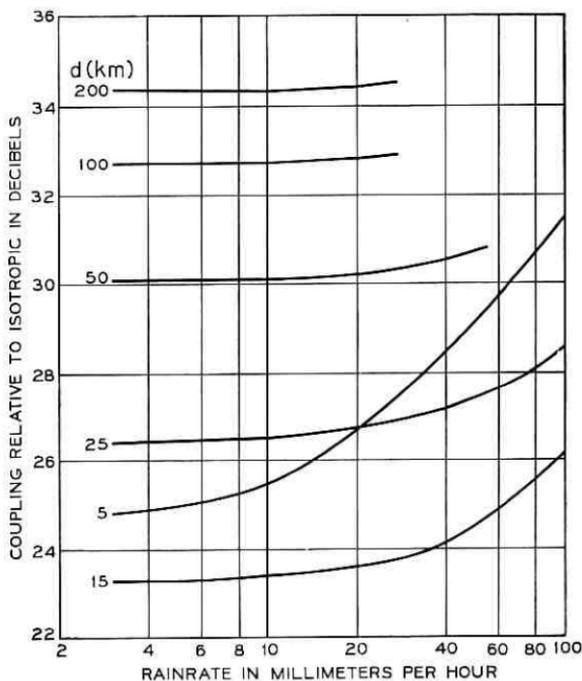
Fig. 6—Coupling for an antenna with 36 dB gain located at R and directed toward T relative to an isotropic antenna at R (gain realized) versus rain rate; frequency 11 GHz.

7.1 Uniform Rain (Case I)

The effect of directivity at 11 GHz is shown in Fig. 6 where the coupling relative to the isotropic case (gain realized*) is plotted versus rain rate for various separations between the ground-station and terrestrial radio-relay antennas. It is clear that at low rain rates most of the nominal gain is realized for the larger separations. At 100 to 200 km separation, the beam of the radio-relay antenna illuminates about the same volume as does an isotropic antenna, the volume in this case being limited by the horizon plane and the upper boundary of the rainstorm (see Fig. 1). But as the separation is decreased to, say, 25 km, Fig. 6 shows that the gain realized is decreased by about 10 dB, the intersections of the ground-station and radio-relay beams define the effective volume whereas an isotropic antenna at R would see essentially all of the ground-station beam.

---

* Of course, antenna gain is a constant but the concept of gain realized in propagating through a given medium is convenient and will be used here.

Figure 6 also shows that the gain realized increases with rain rate, especially for the lesser separations. This is caused by attenuation of the signal scattered from the further (upper) portions of the ground-station beam (see Fig. 1) which, for heavy enough rain rates, produces an effective scattering volume (for an isotropic antenna at R) which is not too much larger than the volume defined by the two beams in the directive case.

A somewhat more subtle effect is the minimum in gain realized which occurs at some separation between 5 and 25 km in Fig. 6. For example, at a rain rate of 50 mm/hr, the gains realized at separations of 50, 25, 15 and 5 km are 31, 27, 24 and 29 dB. Thus, the gain realized begins to increase when the separation is decreased below about 10 km. Now the axis of the beam directed toward space intersects the upper boundary of the rain at a distance of about 7 km along the earth's surface from the ground station and it is at this separation that the total path from R to the upper portions of the volume occupied by the beam (see Fig. 1) is a minimum; that is, r_2 for the upper portions of the beam is a minimum at $d \simeq 7$ km. Thus, for $d < 7$ km the effective volume for an isotropic antenna at R is further decreased and the gain realized with a directive antenna (relative to the isotropic case) increases. For these small separations, the gain realized also increases with rain rate as discussed in the previous paragraph.

At 18 and 30 GHz, the behaviour of the data is similar to that of Fig. 6 but the effect of attenuation is so high that the curvatures are more pronounced. At 4 and 6 GHz, the antenna gain is realized and is relatively independent of separation and rain rate.

7.2 Extended Path (Case IV)

The results for the extended path scattering are shown in Fig. 7 for (a) 4 and 6 GHz and (b) 18.5 GHz. As in the previous case, the gain realized is nearly equal to the nominal gain at the maximum separation distance of 200 km where the common volume is not limited by the relay antenna beamwidth. For lesser separations, the gain realized decreases depending on both the separation and rain rate. The gain realized at the shortest distance (5 km) is reduced below the nominal gain by a maximum of about 15 dB at 4 and 6 GHz and about 11 dB at 18 GHz the 4 dB difference corresponding to the ratio of the assumed beamwidths of one degree at 4 and 6 GHz and two and one half degrees at 18 GHz.

The gain realized at low rain rates with small separations is some-what larger than in Case I. For example, the gain realized at 18 GHz
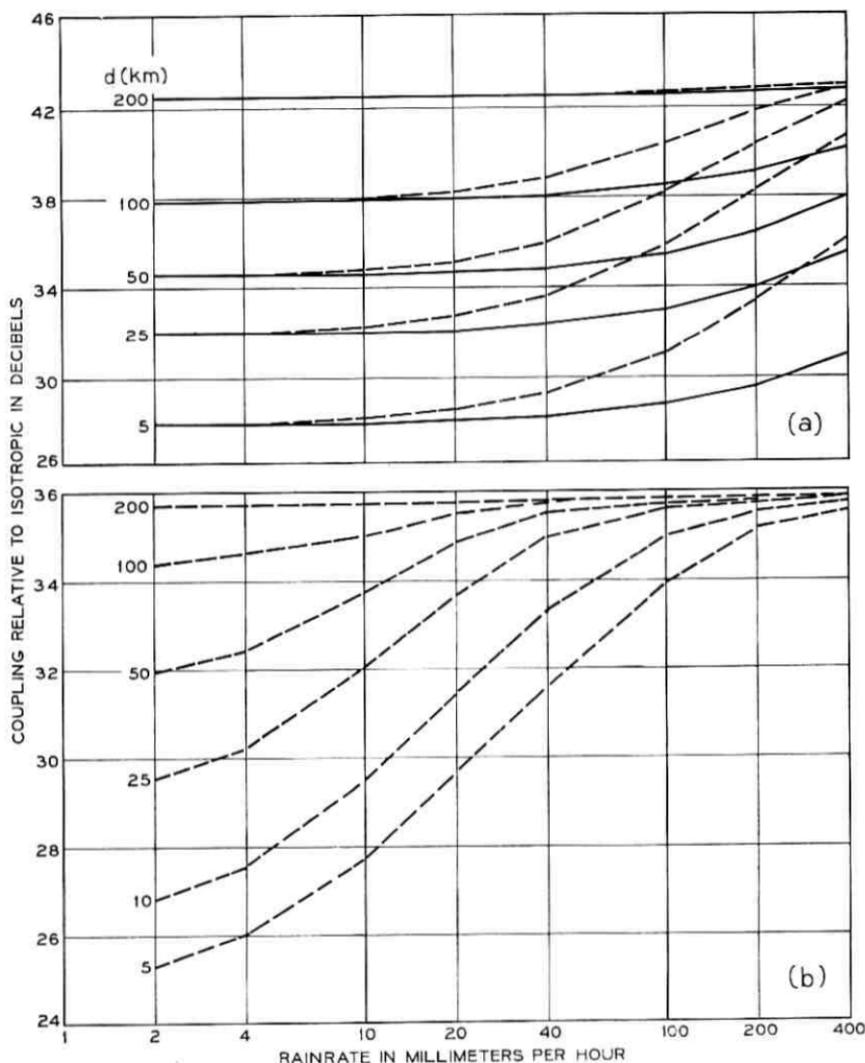
Fig. 7—(a) Coupling for an antenna with 43 dB gain located at R and directed toward T relative to an isotropic antenna at R (gain realized) versus rain rate; frequencies 4 and 6 GHz. (b) Coupling for an antenna with 36 dB gain located at R and directed toward T relative to an isotropic antenna at R (gain realized) versus rain rate; frequency 18 GHz.

and 3 mm/hr (Fig. 7b) with 25 km separation is about 6 dB below the nominal gain whereas the reduction for the Case I (Fig. 6) is 10 dB below nominal. In Case IV, the directive radio relay antenna confines the significant volume to the lower region of the rain; but the significant volume where the isotropic antenna is involved is also near the ground station antenna because r_2 is a minimum there. On the other hand, in Case I, the upper portion of the isotropic scattering volume contributes significantly as discussed in the previous section.

In Case IV, the gain realized also increases with attenuation and hence rain rate as shown in Fig. 7 because the attenuation decreases the effective depth of the isotropic scattering volume. For very high attenuation, most of the energy is scattered from that volume of the rain nearest the ground station and the gain realized is approximately equal to nominal gain; for example, the gain realized is within 2 dB of the nominal gain for rain rates greater than 100 mm/hr at 18 GHz (Fig. 7).

The plots of the gain realized at 11 and 30 GHz (not shown here) are within ±3 dB of the 18 GHz curves, the gain realized being greater at 30 GHz and smaller at 11 GHz. The maximum deviation occurs at 5 km and decreases with increasing separations. At 50 km, the difference is ±2 dB and decreases to only ±1 dB at 100 km.

VIII. DISCUSSION AND SPECIAL EFFECTS

What does all of this mean in terms of what happens in a real-life situation? The various models which have been presented do not represent all the possible orientations and rain conditions that would exist in real situations; however, the coupling caused by realistic rainstorms should range between the results for localized showers (Cases II, III, IV) and those for uniform rain (Case I), depending of course on the spatial distribution of the rain over the path. At 4 and 6 GHz, the effects of rain attenuation are small, at least for realistic sizes of areas with high rain rate, and the results for Cases II and III give a good estimate of the coupling which might be expected for a given rain rate in the ground-station antenna beam. Thus, couplings of the order −140 dB at 5 km to −170 dB at 200 km (Fig. 3a and b) for an isotropic level response at the relay station are typical at 4 and 6 GHz for rain rates of 100 mm/hr or greater.

At frequencies above 10 GHz, attenuation by the rain over the entire path is a dominant factor and it gives rise to large variations in the coupling for the various cases considered. For example, the

coupling at 30 GHz for a uniform rain of 10 mm/hr over a 25 km path (Case I) is −210 dB, whereas, the corresponding coupling with localized showers is −175 dB and −160 dB for Cases II and IV, respectively. However, the maximum couplings with localized showers are about the same as the couplings at 4 and 6 GHz. Thus the spatial distribution of rainfall is an important factor in evaluating the potential interference coupling due to scatter by rain in future systems which may use frequencies exceeding 10 GHz.

Satellite systems above 10 GHz will no doubt use a diversity system of two or more earth stations spaced some few miles apart to avoid outages due to attenuation by rain. Such a path-diversity system will minimize the rain-scatter interference problem since the earth station with the least attenuation and hence with the least rain in its antenna beam will carry the information-bearing signal.

Other factors which will affect the coupling due to scatter should be mentioned. First, the height of the rain for all cases has been taken to be 4 km whereas weather radar data[8,9] show that high effective rain-rates may exist at heights of 10 km. This will extend to horizon for direct rain-scatter coupling to distances of 400 km or more. Also, polarization has been neglected in the calculation; for certain orientations of the ground-station and radio-relay antenna beams, the coupling will be decreased due to this effect.

Another important consideration is that the scattering and attenuation constants, based on empirical data in each case, are not consistent for very high rain rates. Based on these data, we have taken the scattering coefficient to increase as the 1.6 power of rain rate whereas the attenuation constant increases as a lesser power; this gives an inconsistent result at 30 GHz for rain rates greater than about 100 mm/hr in that the attenuation is less than the power lost due to the predicted scatter! Thus, at 30 GHz, and high rain rates, the scattering is less than we have calculated. At frequencies less than 30 GHz, the discrepancy is very small.

In summary, the calculated results show that couplings of the order −140 dB to −170 dB at distances of 5 km to 200 km, respectively, can be expected with an isotropic antenna response at the terrestrial radio relay station, for the small percentages of time associated with heavy rain rates.* These couplings are increased significantly if the response of the relay station exceeds isotropic in the general direction of the earth station. As discussed in the introduction, since couplings

---

* In New Jersey, rain rates of 200 mm/hr have been observed for periods of the order five minutes during a year (0.001 percent of the time).

of the −140 dB are significant, interference due to scatter by rain is a factor which should receive consideration in locating satellite stations. Of course, direct coupling via antenna side lobes and coupling via tropospheric scatter must also be taken into account. In essentially all cases considered here, the coupling due to scatter by rain is less than the line-of-sight coupling but exceeds the median tropospheric scatter coupling between isotropic antennas at the two stations.

There are several previous publications concerned with interference caused by scattering from precipitation [10-14] but experimental data are sparse. We find agreement between the present calculations and the limited data given in Refs. 10 and 12.

IX. ACKNOWLEDGMENT

REFERENCES

1. Proceedings of the Thirteenth Radar Meteorology Conference, American Meteorological Society, McGill University, Montreal, Canada, August, 1968.
2. Semplak, R. A., and Turrin, R. H., "Some Measurements of Attenuation by Rainfall at 18.5 GHz," B.S.T.J., *48*, No. 6, (July-August 1969), pp. 1767.
3. Chu, T. S., "On Coherent Detection of Scattered Light," IEEE Trans. Antennas and Propagation, *AP-15*, No. 5 (Septmember 1967), pp. 703–704.
4. Kogelnik, H., "Imaging of Optical Mode Resonators with Internal Lenses," B.S.T.J., *44*, No. 3 (March 1965), p. 455, equation (50).
5. Freeny, A. E., and Gabbe, J. D., "A Statistical Description of Intense Rainfall," B.S.T.J., *48*, No. 6 (July-August 1969).
6. Gunn, K. L. S., and East, J. W. R., "The Microwave Properties of Precipitation Particles," J. Royal Meteorological Soc., *80*, (1954), pp. 522–545.
7. Hogg, D. C., "Statistics on Attenuation of Microwaves by Intense Rain," B.S.T.J., *48*, No. 9 (November 1969), pp. 2949–2962.
8. Marshall, J. S., Holtz, Clifford D., and Weiss, Marianne, "Parameters for Airborne Weather Radar," Scientific Report MW-48, Stormy Weather Group, McGill University, Montreal, Canada, May 1965.
9. Atlas, D., "Model Atmospheres for Precipitation," *Handbook for Geophysics and Space Environments*, Bedford, Massachusetts: Air Force Cambridge Research Laboratories, 1965, Section 5.2, p. 6.
10. Hogg, D. C., Semplak, R. A., and Gray, D. A., "Measurement of Microwave Interference at 4 Gc Due to Scatter by Rain," Proc. IEEE, *51*, No. 3 (March 1963), p. 500.
11. Dennis, A. S., "Forward Scatter From Precipitation as an Interference Source at Stations Monitoring Satellites," Stanford Res. Memo 2, Project 3773, November 1961.
12. Buige, A., and Levatich, J. L., "Measurement of Precipitation Scatter Effect on Propagation at 6, 12 and 18 GHz," to be published AIAA Conference Proceedings, Los Angeles, California, April, 1970.
13. Altman, F. J., "Design of a Precipitation Scatter Experiment," URSI—Fall Meeting, Austin, Texas, December, 1969.
14. Carey, R. B., and Kalagain, G. S., "Detailed Analysis of FCC/USAF POPSI Project Data," URSI—Fall Meeting, Austin, Texas, December 1969.
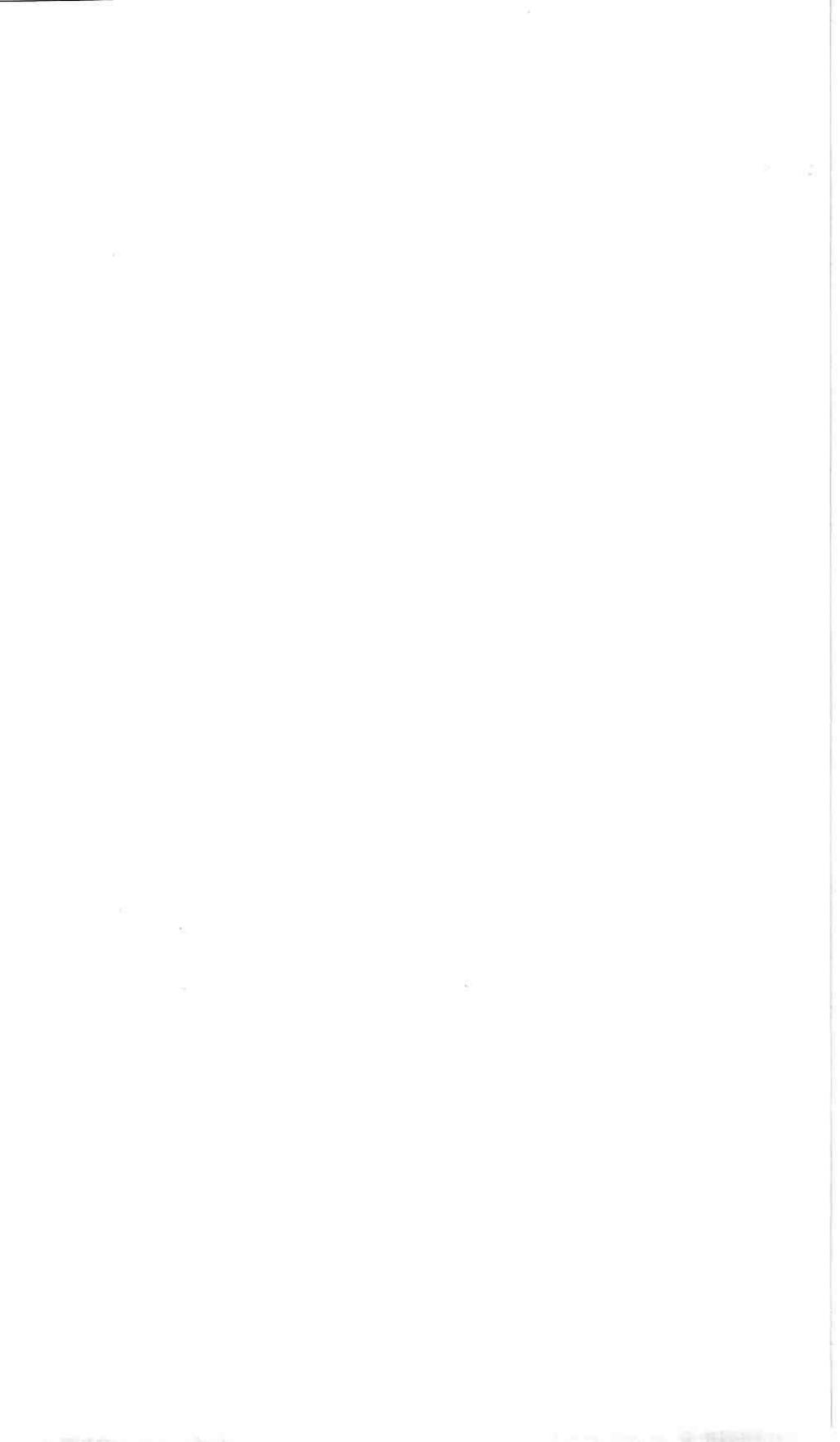
The Number of Fades in Space-Diversity Reception

By ARVIDS VIGANTS

Experimental and theoretical results are presented which show that deep fades due to multipath propagation on line-of-sight microwave links are greatly reduced in number when two vertically separated receiving antennas are used and the stronger of the two received signals is selected as the diversity signal. The experimental results are based on 6 GHz propagation data obtained for a 72-day period on a 28.5 mile path in Ohio, with a 27.5 foot vertical separation of the receiving antennas. The theoretical results are obtained by treating the received signals as correlated Rayleigh distributed random variables. Theoretically predicted variations with fade depth agree with experimental observations. Combination of the theoretical results with experimentally determined parameters provides results which can be used, for example, to calculate the reduction in the number of fades as a function of the vertical separation of the receiving antennas, wavelength, path length, fade depth, and the gain difference (if any) of the receiving antennas.

I. INTRODUCTION

Line-of-sight microwave transmission is affected by a phenomenon called multipath propagation. When this phenomenon is present, the output from a receiving antenna can be practically zero for seconds at a time. However, such deep fades rarely occur simultaneously in the outputs of two vertically separated receiving antennas, and this is the basis for space-diversity reception on line-of-sight microwave links.

The number of fades and the average durations of fades are fundamental properties of the line-of-sight microwave channel. In high performance systems, they may determine the limit of attainable performance. Previous discussions of space diversity do not cover the number of fades and their average durations.[1-3] Experimental

1513

data on these are difficult to obtain because long periods of time are needed to observe a number of fades sufficient for theoretical analysis. We present a set of experimental results obtained from measurements made in 1966 in Ohio. We also present, and this is a major part of the paper, theoretical results on the number of fades and their average durations.

The diversity scheme for which our results are obtained is one where the diversity signal is, at all instants of time, the stronger of two received signals. This scheme is of interest because of its inherent simplicity, and because experimental values for the diversity signal can be obtained easily during computer processing of experimental observations of the received signals. From a broader point of view, this provides information about simultaneous behavior of pairs of received signals that can be used in reliability and interference calculations, and in performance calculations for diversity schemes other than the simple one above.

The number of deep fades is greatly reduced when space-diversity reception is used. We show how to calculate this reduction as a function of the vertical separation of the two receiving antennas, wavelength, path length, depth of fade, and the gain difference (if any) of the receiving antennas. Our results are based on a combination of theory and experiment. The outputs of the receiving antennas during deep fades are treated as random variables that are jointly Rayleigh distributed.[4,5] This provides a description of how various quantities change with fade depth. Parameters describing the occurrence of fading, correlation in space, and the spectral width of the fluctuations are determined from experimental data.

The theoretical expressions describe our experimental observations quite well. However, both theoretical knowledge of fading and the amount of experimental data on fading are limited, and the theoretical model presented here may need modifications as work proceeds and more data become available.

II. EXPERIMENTAL RESULTS

The transmitted power in microwave line-of-sight communication systems is constant (angle modulation is used). Fading information can therefore be obtained from working systems, without interfering with their operation, by installing additional equipment to monitor the power output of the receiving antennas. This was done at West Unity, Ohio, where, in addition, two 8-foot diameter parabolic reflector antennas were installed below existing antennas to provide information
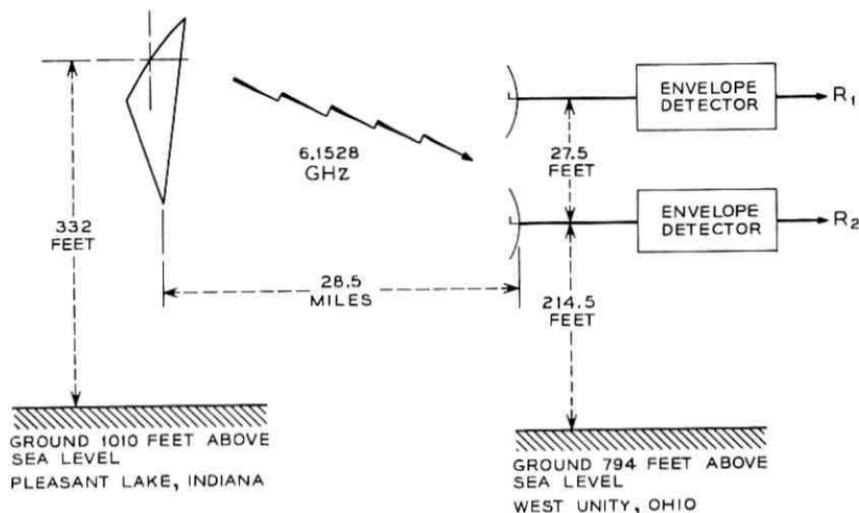
Fig. 1—Essential parameters of the experiment.

about space-diversity reception. We will discuss the fades observed on these two receiving antennas.

The microwave link on which the data were obtained is shown schematically in Fig. 1. The boxes labeled "envelope detector" represent rather sophisticated equipment used in the investigation of fading. The path length was 28.5 miles.* The transmitting antenna at Pleasant Lake, Indiana, was a standard system antenna—a 10-foot horn reflector. The center-to-center vertical separation of the receiving antennas was 27.5 feet. The basic experimental data consisted of observations of the envelopes of received angle-modulated signals centered on 6.1528 GHz.

An important feature of the experiment was long-term coverage. Deep fades are rare events, and the received power must be observed continuously for long periods of time to arrive at a sample of adequate size. Reliability of equipment must be high, and the subsequent processing of the data is a formidable task, even with a computer, because of the high data volume. The data discussed here cover a 72-day period, July 19, 1966, to September 29, 1966.

The power received by the two antennas was sampled essentially simultaneously every 0.2 seconds, converted to a decibel scale, and recorded in digital form for subsequent computer processing (in the

---

* Ground reflections were negligible, and transmission to the bottom antenna had a clearance of about 15 feet (over trees) at 2/3 earth radius conditions.

absence of fading the recording rate was less than the sampling rate). The quantizing steps in the analog-to-digital conversion were somewhat less than 2 dB, which represented a compromise between quantization error and the number of bits needed to encode the power levels (five bits were used). During computer processing, the power level at a sampling instant was assumed to hold until the next sampling instant.

The data were first processed to determine the received power levels on each antenna in the absence of fading. The received power levels during fading were then normalized to the corresponding power levels in the absence of fading. We denote by $20 \log R_1$ and $20 \log R_2$ the normalized dB readings from the top and the bottom receiving antennas respectively. The voltage envelope from the top antenna is R_1, and the voltage envelope from the bottom antenna is R_2. As a consequence of the normalization, both R_1 and R_2 are unity in the absence of fading.

The diversity signal was synthesized during processing by taking the stronger of the two received signals at each sampling instant. The envelope of the diversity signal is

$$R = \max (R_1, R_2). \tag{1}$$

The diversity signal is synthesized from signals that are quantized, and $20 \log R$ can therefore have errors with an upper bound of roughly 2 dB.

A signal is said to be in a fade of depth L and duration τ when its voltage envelope $(R_1, R_2, \text{or } R)$ becomes less than L and remains less than L for τ seconds. During processing, the number of fades of depth L were counted for a set of values of L, and the average fade durations were obtained by dividing the sum of the durations by the number of fades. We restrict our discussion to fades deeper than -20 dB; that is, to values of L that are less than a tenth.

The results for the average fade durations are shown in Fig. 2. (The two theoretical solid lines will be discussed later. The diversity points for fades close to -40 dB are not shown, because an average based on only three observations would not be meaningful.) The durations can be substantial. For example, the average fade duration of a nondiversity signal at -40 dB is about five seconds. Diversity does not reduce these durations appreciably. The reduction is roughly by a factor of two, which can be inferred from a simple heuristic argument as follows. Because the receiving antennas are not very far apart, the fade durations of R_1 and R_2 are roughly the same, but offset in time. As fades become deeper they become shorter, and the offsets start to exceed the fade durations, which results in a large reduction
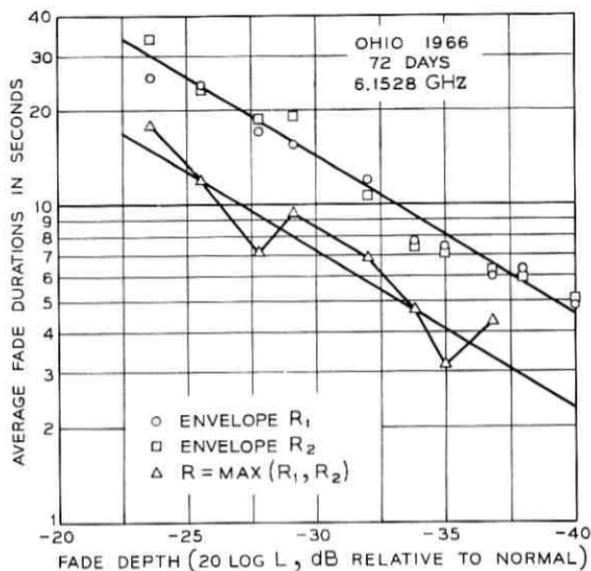
Fig. 2—Average fade durations (Ohio, 1966, 72 days, 6.1528 GHz—the two lines are theoretical).

in the number of deep simultaneous fades (we will see this shortly). Suppose both R_1 and R_2 undergo fades of duration τ. An overlap of the fades, if it does occur, has a value between 0 and τ. There is no preferred overlap position, and $(\tau/2)$ is therefore the average duration of a simultaneous fade.

The results for the number of fades in the 72-day period are shown in Fig. 3. Space diversity has been quite successful in reducing the number of fades. As a matter of fact, the reduction for the deeper fades is so large that the 72-day period was barely adequate for the observation of any −40 dB simultaneous fades at all. From the experimental points, the reduction at −40 dB is by about a factor of twenty. However, this is based on a small sample of only three simultaneous fades at −40 dB. The theoretical curves (discussed later) suggest that, on the average, the reduction at −40 dB will be by about a factor of sixty. There is a small systematic difference in the number of fades of R_1 and R_2. We hope to obtain more information on this from experiments now under way at Palmetto, Georgia.

The conclusion from these observations is that space diversity acts mostly by reducing the number of fades, and that the reduction increases as the fades become deeper.
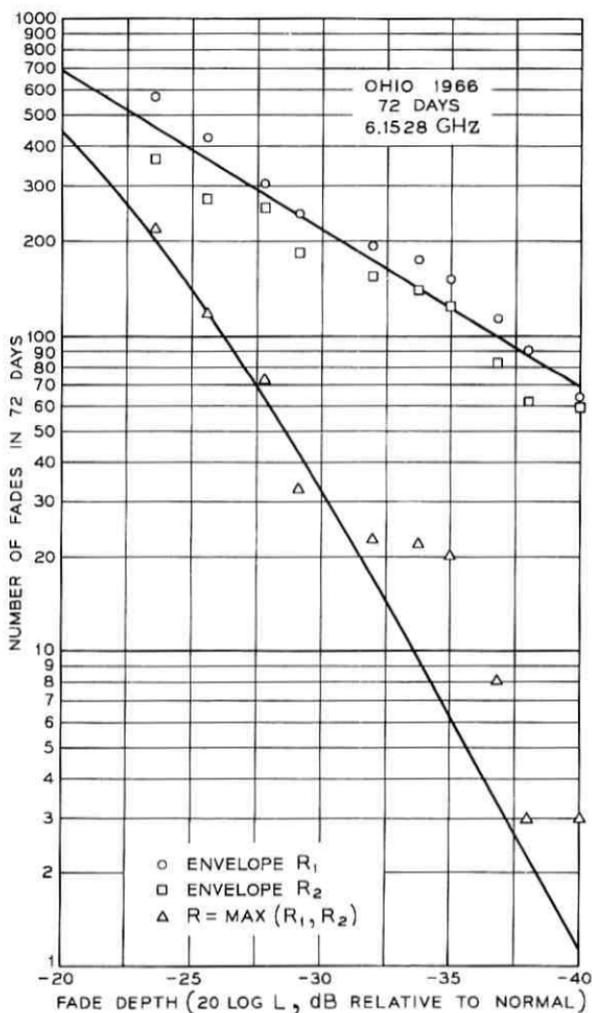
Fig. 3—Number of fades (Ohio, 1966, 72 days, 6.1528 GHz—the curves are theoretical).

III. THEORETICAL MODEL

The number of fades and the average fade durations vary with fade depth. The functional form of the variations can be calculated from probability theory, using methods developed by Rice,[4,5] if various probability density functions associated with R_1, R_2, and R are known.

To arrive at a set of probability density functions for the present

problem, we note that multipath fading normally occurs during relatively short intervals, separated by long periods that contain no fading. Let T_o denote the total time of observation (72 days in the case of the experimental data in the previous section), and let $T(R_1 < L)$ and $T(R_2 < L)$ denote the respective amounts of time during which R_1 and R_2 have values less than L. A mathematical description of experimental observations[3] of these times is

$$T(R_1 < L) = rT_0L^2, \qquad L < 0.1 \tag{2}$$

$$T(R_2 < L) = rT_0L^2, \qquad L < 0.1. \tag{3}$$

The parameter r is proportional to the ratio of time containing fading to total time. It is a function of the length of the line-of-sight path, frequency, time of the year (summer or winter), climate, and terrain.[6]

To introduce some generality, suppose that the gain associated with the bottom antenna is changed, so that the value of R_2 in the absence of fading changes from unity to a value v. If R_2 is still measured on the same scale as R_1,* then equation (3) becomes

$$T(R_2 < L) = rT_0(L/v)^2, \qquad (L/v) < 0.1. \tag{4}$$

The parameter v allows top and bottom antennas of different sizes (the gain difference in dB is 20 log v). If the bottom antenna is smaller than the top antenna, then v is less than unity. This is the case when a second antenna is added below a main antenna on an existing tower, and the size of the second antenna is limited by the allowable wind load on the tower.

The probabilities that R_1 and R_2 respectively are less than L are obtained by dividing $T(R_1 < L)$ and $T(R_2 < L)$ by a normalized time rT_o :

$$\begin{aligned} \text{Pr } (R_1 < L) &= (rT_0)^{-1}T(R_1 < L) \\ &= L^2, \qquad L < 0.1 \end{aligned} \tag{5}$$

$$\begin{aligned} \text{Pr } (R_2 < L) &= (rT_0)^{-1}T(R_2 < L) \\ &= (L/v)^2, \qquad (L/v) < 0.1. \end{aligned} \tag{6}$$

The basis of our theoretical description of fading is the observation that the Rayleigh probability distributions

$$\begin{aligned} \text{Pr } (R_1 < L) &= 1 - \exp{(-L^2)} \\ &\cong L^2, \qquad L < 0.1 \end{aligned} \tag{7}$$

---

* Fade depths of all signals are expressed in decibels relative to the value of the received signal from the top antenna in the absence of fading.

$$\Pr(R_2 < L) = 1 - \exp(-L^2/v^2)$$

$$\cong (L/v)^2, \qquad (L/v) < 0.1 \tag{8}$$

have the same form as equations (5) and (6) for deep fades. This suggests that during deep fades pairs of received envelopes can be treated theoretically as correlated Rayleigh distributed variables. Note that in this theoretical model, as a consequence of equations (7) and (8), the mean square values of the envelopes are

$$\langle R_1^2 \rangle = 1, \qquad \langle R_2^2 \rangle = v^2 \tag{9}$$

where the angle brackets denote averages.

Given that R_1 and R_2 are jointly Rayleigh distributed, one can proceed purely theoretically and obtain expressions for the number of fades and the average fade durations of R. The general expressions for the number of fades are obtained in Appendix A. The probability density functions are listed in Appendix B. The theoretical results can be expected to apply only to deep fades because of the restrictions on L in equations (2) through (8). Reduction to the case of deep fades is carried out in Appendix C, and the results are summarized and discussed in Section IV.

IV. THEORETICAL RESULTS FOR DEEP FADES

The theoretical results for the number of fades and for the average fade durations become relatively simple when fades are deep. Deep in the context here means, first, that

$$L < 0.1, \qquad (L/v) < 0.1, \tag{10}$$

which is just a restatement of the conditions that appeared in the previous section. Second, there are associated conditions that come out of the mathematics in the appendices,

$$q^{-1}L^2 < 0.1, \qquad q^{-1}(L/v)^2 < 0.1. \tag{11}$$

The parameter q describes the correlation of R_1^2 and R_2^2. It is unity when the receiving antennas are very far apart, and it becomes zero when the two receiving antennas merge into one. We give an empirical expression for q when we compare the experimental and theoretical results in Section V. The conditions in equation (11) show how the separation of the receiving antennas affects the range of fade depths for which the simple approximations presented in this section are valid.

The number of fades of the diversity signal is one of the main results. For deep fades, the number of fades of R per unit time is

$$N \cong c(1 + v)(v^2 q)^{-1} L^3,$$ (12)

where the parameter c is proportional to the rms value of the time derivative of R_1,

$$c = [(2/\pi)\langle (dR_1/dt)^2 \rangle]^{\frac{1}{2}}.$$ (13)

The corresponding number of fades of the envelope R_1 per unit time is

$$N_1 \cong cL.$$ (14)

The reduction in the number of fades provided by diversity is therefore

$$F_N = N_1/N$$
$$\cong (1 + v)^{-1} v^2 q L^{-2},$$ (15)

which can be a large number for deep fades (L close to zero).

Numerical estimates of F_N can be obtained using values of q given by equation (19) in Section V. For example, a vertical separation of the receiving antennas that often appears adequate is 40 feet.[1,2] For this, q is about 0.025 at 6 GHz on a 26.5 mile path, which is a typical average path length. For receiving antennas of equal size (v is unity), the reduction in the number of fades is (from equation (15)) about 125 at -40 dB. This means that at the -40 dB level there is, on the average, only one overlap in 125 fades of the nondiversity signals.

The expressions for various quantities, when fades are deep, are summarized in Table I, where the first column refers to R_1 and the second column to R. This shows that the average fade duration of R is

$$\langle t \rangle \cong (1 + v)^{-1} \langle t_1 \rangle$$ (16)

where $\langle t_1 \rangle$ is the average fade duration of R_1. The ratio of $\langle t_1 \rangle$ to $\langle t \rangle$ is independent of fade depth for deep fades, and it becomes two when v

TABLE I—EXPRESSIONS FOR VARIOUS QUANTITIES FOR DEEP FADES

	Nondiversity	Diversity	Improvement
Probability of fade	L^2	$(v^2 q)^{-1} L^4$	$v^2 q L^{-2}$
Number of fades	cL	$c(1 + v)(v^2 q)^{-1} L^3$	$(1 + v)^{-1} v^2 q L^{-2}$
Average duration	$c^{-1} L$	$c^{-1}(1 + v)^{-1} L$	$(1 + v)$

is unity (same size receiving antennas), which supports the heuristic argument given previously in the section on the experimental results.

The reduction in the total amount of time spent in fades is*, from the first row in Table I,

$$F \cong v^2 q L^{-2}. \tag{17}$$

Therefore

$$F_N \cong (1 + v)^{-1} F. \tag{18}$$

When v is unity or less, conditions (10) and (11) translate into a simple statement that the approximations presented here are valid when F is larger than ten.

V. COMPARISON OF EXPERIMENTAL AND THEORETICAL RESULTS

The theory predicts that the number of fades and the average fade durations are certain functions of the fade depth L. This can be verified experimentally. The theory also contains four parameters (r, v, q, c). One of these, the gain difference v, is a design parameter that we control. The other three (r for occurrence of fading, q for correlation in space, c for spectral width of fluctuations) depend on the inhomogeneities of the atmosphere and their interaction with electromagnetic waves. Since we do not know how to determine these parameters theoretically for deep fades on microwave line-of-sight links, we determine them from the experimental data.

In the experiment, the gain difference parameter v is unity. An empirical result[3] for the correlation parameter q is

$$q = (2.75)^{-1}(s^2/\lambda d), \tag{19}$$

where s is the vertical center-to-center separation of the receiving antennas, λ is the wavelength, and d is the path length—all measured in the same units. For the receiving antenna pair discussed here, the value of q is 0.012, which is one of the experimental values[3] on which equation (19) is based. The value of the fade occurrence parameter r is one-half (from Fig. 5 in our previous work[3]).

The theoretical description of the average fade durations of R_1 and R_2 in Fig. 2 is

$$\langle t_1 \rangle = c^{-1} L. \tag{20}$$

---

* We have discussed this previously.[3] The parameter q denotes the quantity $(1 - k^2)$ appearing in the earlier work and in Appendix B here.

Least squares fitting of equation (20) to the experimental points gives*

$$c \cong 2.22 \times 10^{-3} \text{ second}^{-1}. \tag{21}$$

The data are in good agreement with the theoretically predicted dependence on L to the first power. Theory also predicts that the average fade durations of the diversity signal should be shorter by a factor of two, and the corresponding theoretical line in Fig. 2 describes the experimental points quite well. The larger scatter in the diversity points is caused by the small number of observations of deep simultaneous fades.

The number of fades can now be calculated entirely from theory. The number of fades per unit time of the nondiversity signals is given by equation (14). The corresponding number of fades in 72 days is

$$N_1' = rT_0 N_1 \cong 6.9 \times 10^3 L. \tag{22}$$

This equation, shown as the top solid line in Fig. 3, describes the data adequately. Similarly, the number of fades of the diversity signal envelope R in 72 days is, using (12),

$$N' = rT_0 N \cong 1.15 \times 10^6 L^3. \tag{23}$$

The deep fade approximation in this case applies only to fades deeper than about -30 dB, and the curved section of the bottom solid curve in Fig. 3 was therefore computed from expressions in the appendices. The departures from the curve when the number of simultaneous fades becomes small are very likely statistical in nature. Similar curves for frequency diversity (to be published), where we have more than one set of data, show deviations both above and below the theoretical curves in apparently random fashion over fifteen sets of data. Furthermore, as discussed earlier, experimental values for $20 \log L$ can be in error by up to about 2 dB because the diversity signal is synthesized from quantized signals.

---

* The parameter c is related to the spectral width of the fluctuations of R_1 (the spectra of R_1 and R_2 are assumed to differ only by a multiplicative constant, which allows the mean square values to be different). For example, for a Gaussian spectrum

$$w(f) = (2/\pi)^{\frac{1}{2}} f_b^{-1} \exp(-f^2/2f_b^2)$$

the mean square derivative is

$$\langle (dR_1/dt)^2 \rangle = (2\pi f_b)^2$$

which can be substituted into equation (13) to relate c to f_b. We do not have experimentally determined spectra. As a speculation, using the value for c from expression (21), we obtain

$$f_b \cong 4.5 \times 10^{-4} \text{ Hz.}$$

All in all, we think the theory describes the experimental data well. To describe the data we needed only one additional parameter, the spectral width parameter c, beyond those determined previously[3] from amplitude distributions. The theoretical model based on the joint Rayleigh probability distribution therefore appears to give a highly consistent description of deep fades on line-of-sight microwave links.

VI. CONCLUSIONS

The experimental and theoretical results presented here show that, for fading caused by multipath propagation, simultaneous deep fades on vertically separated receiving antennas can be practically non-existent. Furthermore, the theoretical results enable one to describe quantitatively what is meant by practically nonexistent. From a more general point of view, a theoretical description of the output of the receiving antennas may help in modeling the disturbances in the electromagnetic waves incident upon the antennas, on which much remains to be done.

VII. ACKNOWLEDGMENT

We are indebted to W. T. Barnett for many useful discussions. The experimental data come from an experiment to which many people at Bell Telephone Laboratories contributed. In particular, the Multiple Input Data Acquisition System (MIDAS) was designed by G. A. Zimmerman, and the ideas and work of C. H. Menzel were essential to the processing of the data.

APPENDIX A

General Expressions for the Number of Fades

For the diversity signal envelope R, the number of fades of depth $20 \log L$ dB is equal to the number of times R crosses L in an upward direction. Suppose that a time period T is divided into n intervals of length dt such that an interval contains either none or one of the upcrossings. Let m_1 be the number of intervals containing upcrossings made at instants in time when R is R_1, and let m_2 be the number of intervals containing upcrossings made when R is R_2. Then the quantities

$$P_1 = m_1/n, \qquad P_2 = m_2/n \tag{24}$$

are conditional probabilities (the conditions are stated in the next
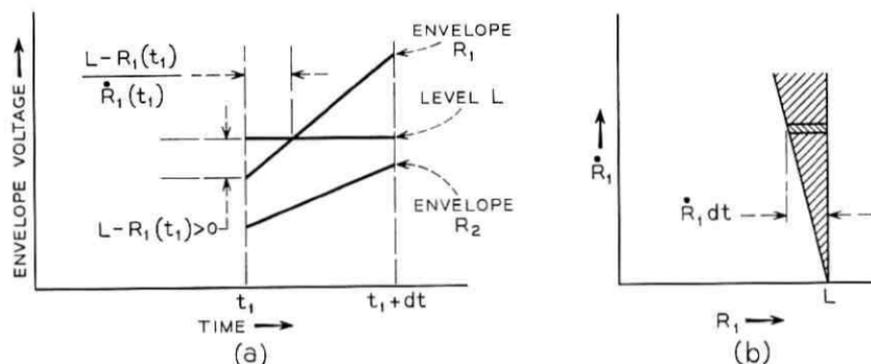
Fig. 4—Level crossing conditions: (a) level crossing in an interval of length dt, and (b) integration in the $(R_1, \dot{R}_1)$-plane.

paragraph) that R_1 and R_2 respectively cross a given level in an interval of length dt. The number of upcrossings of R per unit time is

$$N = (P_1 + P_2)/dt, \tag{25}$$

and the task is to express P_1 and P_2 in terms of the appropriate probability density functions.

To obtain an expression for P_1, we look in detail at an interval dt in which R_1 crosses a level L in an upward direction, as shown in Fig. 4. The interval length dt will go to zero in the limit, and the envelopes in the interval can therefore be represented by straight line segments. The conditions for the indicated level crossing to occur are*

$$R_2(t_1) < L \tag{26}$$

$$\dot{R}_1(t_1) > 0 \tag{27}$$

(the dot denotes a time derivative), and

$$0 < \frac{L - R_1(t_1)}{\dot{R}_1(t_1)} < dt. \tag{28}$$

Integration of the joint probability density function of $(R_1, R_2, \dot{R}_1)$ over the region in $(R_1, R_2, \dot{R}_1)$-space specified by the three inequalities will give the probability P_1. Let $p(R_1, R_2, \dot{R}_1)$ be this probability density. Then,

$$P_1 = \int_0^L dR_2 \int_0^\infty d\dot{R}_1 \int_{L - \dot{R}_1 dt}^L dR_1 p(R_1, R_2, \dot{R}_1) \tag{29}$$

---

* This approach follows Rice,[4] Section 3.3. The additional condition (26) comes in because we are treating a diversity signal.

where part (b) of Fig. 4 helps to visualize the integration with respect to R_1 as prescribed by the inequality (28). As dt approaches zero this becomes

$$P_1 = dt \int_0^L dR_2 \int_0^\infty d\dot{R}_1\ \dot{R}_1 p(L, R_2, \dot{R}_1). \tag{30}$$

The variables R_1 and $\dot{R}_1$ are uncorrelated (a random variable and its derivative are uncorrelated when measured at the same instant of time). If so, then to a good approximation R_2 and $\dot{R}_1$ are also uncorrelated, and*

$$p(R_1, R_2, \dot{R}_1) = p(R_1, R_2)p(\dot{R}_1). \tag{31}$$

Therefore

$$P_1 = dt \int_0^\infty \dot{R}_1 p(\dot{R}_1)\ d\dot{R}_1 \int_0^L p(L, R_2)\ dR_2. \tag{32}$$

The conditional probability P_2 can be obtained from this by an appropriate change of subscripts.

In terms of probability densities, the number of upcrossings of R per unit time is therefore

$$N = \int_0^\infty \dot{R}_1 p(\dot{R}_1)\ d\dot{R}_1 \int_0^L p(L, R_2)\ dR_2$$

$$+ \int_0^\infty \dot{R}_2 p(\dot{R}_2)\ d\dot{R}_2 \int_0^L p(R_1, L)\ dR_1. \tag{33}$$

The number of upcrossings of R_1 per unit time can be obtained from the first term in N by extending the integration with respect to R_2 to infinity, which gives

$$N_1 = \left\{ \int_0^\infty \dot{R}_1 p(\dot{R}_1)\ d\dot{R}_1 \right\} p_1(L). \tag{34}$$

Similarly, the number of upcrossings of R_2 per unit time is

$$N_2 = \left\{ \int_0^\infty \dot{R}_2 p(\dot{R}_2)\ d\dot{R}_2 \right\} p_2(L). \tag{35}$$

Note that the probability density function of R at the value L is

---

* We are using the symbol p to denote probability density functions in general. The arguments of p specify the actual function.

$$p(L) = (\partial/\partial L) \; \text{Pr} \; (R < L)$$
$$= (\partial/\partial L) \; \text{Pr} \; (R_1 < L, R_2 < L)$$
$$= (\partial/\partial L) \int_0^L dR_1 \int_0^L dR_2 \, p(R_1, R_2)$$
$$= \int_0^L p(R_1, L) \, dR_1 + \int_0^L p(L, R_2) \, dR_2 \; . \tag{36}$$

This shows that N can be expressed in the same form as N_1 and N_2 (a constant multiplied by a probability density function) only when the integrals with respect to $\dot{R}_1$ and $\dot{R}_2$ are equal; that is, when these integrals in equation (33) can be factored out.

The crossing rate N can be expressed in terms of the rates N_1 and N_2. Combining equation (33), (34), and (35), we obtain

$$N = N_1 \, \text{Pr} \; (R_2 < L \mid R_1 = L) + N_2 \, \text{Pr} \; (R_1 < L \mid R_2 = L) \tag{37}$$

where $\text{Pr} \; (R_2 < L \mid R_1 = L)$ is the probability that R_2 is less than L, subject to the condition that R_1 is equal to L, and similarly for the second term.

APPENDIX B

Probability Density Functions

To obtain the required probability density functions, the complex envelopes $[R_1 \exp (i\theta_1)]$ and $[R_2 \exp (i\theta_2)]$ are resolved into quadrature components

$$\left.\begin{aligned}
x_1 &= R_1 \cos \theta_1 \\
x_2 &= R_2 \cos \theta_2 \\
x_3 &= \dot{x}_1 = \dot{R}_1 \cos \theta_1 - R_1 \dot{\theta}_1 \sin \theta_1 \\
y_1 &= R_1 \sin \theta_1 \\
y_2 &= R_2 \sin \theta_2 \\
y_3 &= \dot{y}_1 = \dot{R}_1 \sin \theta_1 + R_1 \dot{\theta}_1 \cos \theta_1
\end{aligned}\right\} \tag{38}$$

where the dot denotes a time derivative. The assumption that x_i and y_i $(i = 1, 2, 3)$ have zero means and are normally distributed with second moments

$$\langle x_n x_m \rangle = \langle y_n y_m \rangle = \lambda_{nm} , \qquad n, m = 1, 2, 3 \tag{39}$$

$$\langle x_n y_m \rangle = 0 \tag{40}$$

leads* to the Rayleigh probability density for R_1

$$p_1(R_1) = (R_1/\lambda_{11}) \exp(-R_1^2/2\lambda_{11}) \tag{41}$$

$$\langle R_1^2 \rangle = 2\lambda_{11} \tag{42}$$

with a similar form for the probability density function for R_2.

The probability density function of $\dot{R}_1$ can be obtained in a similar manner. The joint density function in this case is

$$\begin{aligned}
p(\dot{R}_1, \dot{\theta}_1, R_1, \theta_1) &= R_1^2 p(x_1, x_3, y_1, y_3) \\
&= R_1^2 p(x_1) p(x_3) p(y_1) p(y_3) \\
&= \frac{R_1^2}{(2\pi)^2 \lambda_{11}\lambda_{33}} \exp\left\{-\frac{x_1^2 + y_1^2}{2\lambda_{11}} - \frac{x_3^2 + y_3^2}{2\lambda_{33}}\right\} \\
&= \frac{R_1^2}{(2\pi)^2 \lambda_{11}\lambda_{33}} \exp\left\{-\frac{R_1^2}{2\lambda_{11}} - \frac{\dot{R}_1^2 + R_1^2\dot{\theta}_1^2}{2\lambda_{33}}\right\}. \tag{43}
\end{aligned}$$

Integration with respect to θ_1 (from 0 to 2π) and $\dot{\theta}_1$ (from $-\infty$ to ∞) gives

$$p(\dot{R}_1, R_1) = \frac{R_1}{\lambda_{11}(2\pi\lambda_{33})^{\frac{1}{2}}} \exp\left\{-\frac{R_1^2}{2\lambda_{11}} - \frac{\dot{R}_1^2}{2\lambda_{33}}\right\} \tag{44}$$

from which

$$p(\dot{R}_1) = (2\pi\lambda_{33})^{-\frac{1}{2}} \exp(-\dot{R}_1^2/2\lambda_{33}) \tag{45}$$

$$\langle \dot{R}_1^2 \rangle = \lambda_{33}. \tag{46}$$

It is interesting to note that $\dot{R}_1$ is normally distributed.

The joint probability density function of R_1 and R_2 is,[7] in our notation,

$$\begin{aligned}
p(R_1, R_2) = (R_1 R_2/q\lambda_{11}\lambda_{22}) \cdot \exp\left\{-(2q)^{-1}[(R_1^2/\lambda_{11}) + (R_2^2/\lambda_{22})]\right\} \\
I_0\{kR_1R_2/q(\lambda_{11}\lambda_{22})^{\frac{1}{2}}\} \tag{47}
\end{aligned}$$

where I_0 is a modified Bessel function of the first kind, and

$$k^2 = \lambda_{12}^2/\lambda_{11}\lambda_{22} \tag{48}$$

$$q = 1 - k^2. \tag{49}$$

The conditional probability required in equation (37) is, therefore, from the probability density functions given above,

---

* Details can be found in Davenport and Root,[7] Section 8.5, where Rice's[4] work is summarized.

$$\Pr\,(R_2 < L \mid R_1 = L) = \int_0^L dR_2 p(L, R_2)/p_1(L)$$

$$= e^{-\nu} \int_0^x e^{-t} I_0[2(yt)^{\frac{1}{2}}]\, dt \tag{50}$$

where

$$x = q^{-1}(L^2/\langle R_2^2\rangle) \tag{51}$$

$$y = q^{-1}k^2(L^2/\langle R_1^2\rangle). \tag{52}$$

The other conditional probability $\Pr\,(R_1 < L \mid R_2 = L)$ can be obtained from these equations by appropriate changes of subscripts.

APPENDIX C

Deep Fade Approximations

For deep fades (small values of L) simple approximations of the various expressions can be obtained. First of all,

$$p_1(L) \cong 2L/\langle R_1^2\rangle, \qquad (L/R_{1\,\mathrm{rms}}) < 0.1 \tag{53}$$

which immediately leads to

$$N_1 \cong c(L/R_{1\,\mathrm{rms}}), \qquad (L/R_{1\,\mathrm{rms}}) < 0.1, \tag{54}$$

where

$$c = \{(2/\pi)(\langle \dot{R}_1^2\rangle/\langle R_1^2\rangle)\}^{\frac{1}{2}}. \tag{55}$$

The envelope fluctuations can be assumed to fulfill the condition

$$\langle \dot{R}_1^2\rangle/\langle R_1^2\rangle = \langle \dot{R}_2^2\rangle/\langle R_2^2\rangle. \tag{56}$$

If so, then

$$N_2 = c(L/R_{2\,\mathrm{rms}}), \qquad (L/R_{2\,\mathrm{rms}}) < 0.1. \tag{57}$$

Now that the definitions are clear, we can start using the normalized mean square values

$$\langle R_1^2\rangle = 1, \qquad \langle R_2^2\rangle = v^2. \tag{58}$$

In terms of these,

$$N \cong cL\{\Pr\,(R_2 < L \mid R_1 = L) + v^{-1}\Pr\,(R_1 < L \mid R_2 = L)\}. \tag{59}$$

When

$$q^{-1}L^2 < 0.1, \qquad q^{-1}(L^2/v^2) < 0.1, \tag{60}$$

the conditional probabilities become

$$\Pr\,(R_2 < L \mid R_1 = L) \cong q^{-1}(L^2/v^2) \tag{61}$$

$$\Pr\,(R_1 < L \mid R_2 = L) \cong q^{-1}L^2 \tag{62}$$

and

$$N \cong cq^{-1}v^{-2}(1 + v)L^3. \tag{63}$$

The average fade durations of the envelopes R and R_1 are, respectively,

$$\langle t \rangle = N^{-1}\,\Pr\,(R_1 < L, R_2 < L) \tag{64}$$

$$\langle t_1 \rangle = N_1^{-1}\,\Pr\,(R_1 < L). \tag{65}$$

Deep fade approximations of the probabilities are

$$\Pr\,(R_1 < L, R_2 < L) \cong q^{-1}v^{-2}L^4 \tag{66}$$

$$\Pr\,(R_1 < L) \cong L^2, \tag{67}$$

and appropriate substitutions into equations (64) and (65) provide the expressions in Table I.

In cases where the deep fade approximations do not apply, we evaluated the various integrals numerically. It is possible (but we did not do this) to use tables for the conditional probability, equation (50), since it can be expressed in terms of Marcum Q functions[8,9] that occur in communication theory. Similar integrals are denoted by $J(x, y)$ by Luke,[10] who gives references to tables that have to do with hitting a circular target (we could not obtain these tables).

REFERENCES

1. Makino, H., and Morita, K., "The Space Diversity Reception and Transmission Systems for Line-of-Sight Microwave Link," Rev. Elec. Commun. Lab. (Japan), 13, (January–February 1965), pp. 111–129; also "Design of Space Diversity Receiving and Transmitting Systems of Line-of Sight Microwave Links," IEEE Trans. Commun. Tech., COM-15, No. 2 (August 1967), pp. 603–614.
2. White, R. F., "Space Diversity on Line-of-Sight Microwave Systems," IEEE Trans. Commun. Tech., COM-16, No. 1 (February 1968), pp. 119–133.
3. Vigants, A., "Space Diversity Performance as a Function of Antenna Separation" IEEE Trans. Commun. Tech., COM-16, No. 6 (December 1968), pp. 831–836.
4. Rice, S. O., "Mathematical Analysis of Random Noise," B.S.T.J., 23, No. 31 (July 1944), pp. 282–332, and 24, No. 1 (January 1945), pp. 46–156.
5. Rice, S. O., "Distribution of the Duration of Fades in Radio Transmission: Gaussian Noise Model," B.S.T.J., 37, No. 3 (May 1958), pp. 581–635.
6. Barnett, W. T., "Occurrence of Selective Fading as a Function of Path Length, Frequency, and Geography," 1969 USNC/URSI Spring Meeting, April 21–24, 1969, Washington, D. C.
7. Davenport, W. B., and Root, W. L., An Introduction to the Theory of Random Signals and Noise, New York: McGraw-Hill, 1958, Section 8.5.
8. Marcum, J. I., Table of Q Functions, Rand Corporation, January 1, 1950.
9. Schwartz, M., Bennett, W. R., and Stein, S., Communication Systems and Techniques, New York: McGraw-Hill, 1966, Appendix A.
10. Luke, Y. L., Integrals of Bessel Functions, New York: McGraw-Hill, 1962, Chapter 12.

Optimum Phase Equalization of Filters for Digital Signals

By DAVID A. SHNIDMAN

Digital information transmitted by means of a pulse amplitude modulation scheme depends critically on the pulse shape for reliable high speed communications. The pulse shape, in turn, depends in great measure on precise phase equalization. A new technique for the design of phase equalizers, based on digital mean square error, has been developed. This criterion is appropriate for a digital transmission system because it can be related to the system error rate.

A lower bound to the digital mean square error is first obtained by determining the theoretically optimum phase. A physical equalizer consists of a cascade of many (say N) constant resistance all-pass networks. For each of several different values of N, an optimization search over the parameters of the all-pass networks is then done. The smallest value of N which yields an error satisfactorily close to the lower bound is utilized for the optimum physical phase equalizer.

The major benefits derived from using this technique as opposed to the conventional one are:

(i) A significant reduction in the number of all pass sections required.

(ii) More practical element values, facilitating network manufacture.

(iii) A substantial improvement in system performance.

I. INTRODUCTION

Most information transmitted via telephone facilities today is sent over analog channels. It is anticipated, however, that future systems will be largely digital. New services such as *Picturephone*® will be transmitted over toll facilities in a digital format. The Bell System will need to provide a significant digital capacity by the late 1970s. In spite of this emphasis on digital transmission, analog demands will be rapidly increasing for some time to come. The L-4 system, originally conceived to provide increased analog capacity, is now being equipped to transmit

in a digital mode as well. L-4 can therefore temporarily satisfy increasing analog and digital requirements. The eventual transition from analog to digital can be accomplished smoothly by apportioning analog and digital usage according to demand. Similarly, the L-5 system will be able to handle both analog and digital signals. Digital transmission on L-5 is expected to be available earlier than any purely digital system of comparable capacity. These systems with their dual capabilities promise to play an important role in the future of Bell System long-haul transmission.

The scheme (Fig. 1) employed for transmitting digital information over L-4 and L-5 is pulse amplitude modulation. The sequence of input digits, $\{a_k\}$, is used to scale translates (displaced by T seconds) of a given pulse shape. If these pulses do not overlap, then their amplitudes can be determined from samples of the received waveform (assuming noise free transmission) and the scale factors ascertained. Efficient use of channel capacity dictates simultaneous transmission of several signals, each confined to a specific bandwidth. The fulfillment of these bandwidth restrictions causes the transmitted pulses to overlap, and complicates the process of extracting the information digits. For bandwidths of greater than $1/2T$ (for single sideband), there exist a class of pulses, known as Nyquist pulses, which allow, at one point in each T second interval, for the extraction of an interference-free sample. Unfortunately, it is not possible to build filters to realize these pulse shapes exactly or to build sampling equipment with zero pulse widths. Consequently, one is left with an unavoidable amount of "intersymbol interference" such that the output sequence, $\{b_k\}$, will not be an exact duplicate of the input sequence, $\{a_k\}$.

The filters have a second function no less important than pulse shaping, namely, to prevent serious interference between signals in
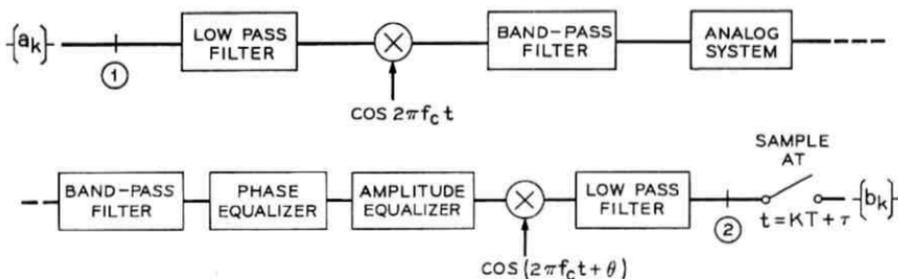

Fig. 1—Model for the digital communication system. [$V(f)$ is the transfer characteristic from 1 to 2.]

different bands. The filters must satisfy out-of-band rejection requirements as well as yield desirable pulse shapes.

With traditional filter design techniques, phase distortion is not considered initially.* An amplitude characteristic, fulfilling out-of-band rejection requirements, is realized as closely to a Nyquist shape as possible. Then the phase is modified by means of a cascade of all-pass sections so as to minimize the "curve fit" mean squared difference (over the pertinent bandwidth) between the resulting delay and a constant delay. This method (of reducing curve fit mean squared error) was initially used in L-4 for the design of filters for digital transmission. The result proved to be unacceptable. The major difficulty with this procedure is the fact that the resultant phase distortion, as determined above, cannot be readily related to an error between an input symbol, a_k, and the corresponding output, b_k. That is, not all errors of equal size as measured by this criterion affect the digital error to the same extent. Any criterion which does relate to differences between the $\{a_k\}$ and the $\{b_k\}$ would necessarily consider both amplitude and phase simultaneously. The digital mean square error, to be defined in the next section, is a criterion which is appropriate for our purposes and is related to the differences between digital input and output.

Although it was originally planned that the amplitude and phase equalization would be designed simultaneously, it was found to be advantageous not to do this initially. The problem becomes much more manageable by separating amplitude and phase designs into two interacting parts. In some cases, it was possible to achieve satisfactory amplitude characteristics by traditional methods. For these cases, the problem becomes one of designing a realizable phase characteristic approximating the optimum phase (which depends on the amplitude and which is not necessarily linear). Therefore, the first new effort is the development of a new technique for phase equalization using a criterion relevant to digital transmission.

II. DIGITAL MEAN SQUARE ERROR

The digital mean square error, D_k, of the kth digit is defined as the mean squared difference between the input, a_k, and its corresponding output, b_k [2,3]; that is,

$$D_k \equiv E\{(a_k - b_k)^2\} \tag{1}$$

---

* In his paper, "Synthesis of Pulse-Shaping Networks in the Time Domain," D. A. Spaulding does consider amplitude and phase simultaneously.[1]

where $E\{\cdot\}$ denotes expectation with respect to the random variables a_k and b_k. If the input sequence is wide-sense stationary and the channel is deterministic, conditions which we shall properly assume, then D_k is independent of k, and the subscript k will be dropped.

It can be shown (see, for example, Ref. 4) that equation (1) can be written as

$$D = m_0 + \int_{-\infty}^{\infty} M(f) V^*(f) \exp(j2\pi f\tau)$$
$$\cdot \left\{ \frac{1}{T} \sum_{\alpha=-\infty}^{\infty} V\left(f - \frac{\alpha}{T}\right) \exp\left[-j2\pi\left(f - \frac{\alpha}{T}\right)\tau\right] - 2 \right\} df \qquad (2)$$

where

$V(f)$ is the transfer characteristic between points 1 and 2 of Fig. 1,
τ is a constant representing a delay in the sampling time,
T is the pulse repetition time interval,

and

$M(f)$ is a factor arising from the correlation between input digits $\{a_k\}$. If we define

$$\rho_n = E\{a_k a_{k+n}\} = \rho_{-n}, \qquad (3)$$

then, assuming convergence, we have for $M(f)$,

$$M(f) = \sum_{n=-\infty}^{\infty} \rho_n \exp(-j2\pi fnT) = \sum_{n=0}^{\infty} m_n \cos 2\pi fnT \qquad (4)$$

where we have defined

$$m_0 = \rho_0$$

and

$$m_n = 2\rho_n, \qquad n \neq 0. \qquad (5)$$

The error, D, can be written in a more convenient form if we define an error term, $\epsilon(f)$, as:

$$\epsilon(f) = \frac{1}{T} \sum_{\alpha=-\infty}^{\infty} V\left(f - \frac{\alpha}{T}\right) \exp\left[-j2\pi\left(f - \frac{\alpha}{T}\right)\tau\right] - 1. \qquad (6)$$

Using equation (6) in equation (2) and noting the fact that the imaginary part of $\epsilon(f)$ is odd, we can write D as:

$$D = m_0 + T \int_{-1/2T}^{1/2T} M(f)[\epsilon(f) - 1][\epsilon^*(f) + 1]\, df$$
$$= 2T \int_{0}^{1/2T} M(f) \mid \epsilon(f) \mid^2 df. \qquad (7)$$

The function $\epsilon(f)$ is a measure of intersymbol interference. If the pulse shape were Nyquist, $\epsilon(f)$ would be zero. The term $M(f)$, due to the correlation of the input digits, acts as a weighting function in the error expression.

$\epsilon(f)$ is determined from the transfer functions of the block diagrams in Fig. 1. We can take advantage of the linearity of the vestigial sideband system and, by translating appropriately, lump several filters together. We define, therefore, the frequency transfer function for the combination of the two low-pass filters as $L(f)$, for the combination of two bandpass filters as $B(f)$, and for the combination of the all-pass networks (that is, the phase equalizer) as $A(f)$. The cable or analog system is represented by $C(f)$. With this notation, we can write the baseband equivalent to the passband portion of the system as:

$$H(f) = \tfrac{1}{2}[A(f + f_c)B(f + f_c)C(f + f_c) \exp(j\theta)$$

$$+ A(f - f_c)B(f - f_c)C(f - f_c) \exp(-j\theta)] \operatorname{rect}\left(\frac{fT}{2}\right) \quad (8)$$

where

$$\operatorname{rect}(x) \equiv \begin{cases} 1, & |x| < \tfrac{1}{2} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

and where f_c is the carrier frequency and θ the phase offset (the difference in phase between the modulation and demodulation cosine functions). The overall transfer function becomes simply

$$V(f) = L(f)H(f). \quad (10)$$

In general $L(f)$, $B(f)$, and $C(f)$ are not measured separately. Instead $R(f)$ is measured where

$$R(f) = \begin{cases} \dfrac{1}{2T} L(f - f_c)B(f)C(f), & f > 0 \\[2mm] \dfrac{1}{2T} L(f + f_c)B(f)C(f), & f < 0 \end{cases} \quad (11)$$

with the result that

$$V(f) = T[R(f + f_c)A(f + f_c) \exp(j\theta)$$

$$+ R(f - f_c)A(f - f_c) \exp(-j\theta)] \operatorname{rect}\left(\frac{fT}{2}\right). \quad (12)$$

If we define $V_1(f)$ as

$$V_1(f) \equiv \frac{g}{T} V(f) \exp(-j2\pi f\tau), \tag{13}$$

where g is a scale factor, then $\epsilon(f)$ becomes, in terms of $V_1(f)$,

$$\epsilon(f) = \sum_{\alpha=-1}^{1} V_1\left(f - \frac{\alpha}{T}\right) - 1 \tag{14}$$

where the limits on the sum reflect the bandlimited nature of $V(f)$.

Since we are initially concerning ourselves with phase equalization only, we want to consider $R(f)$ as fixed and modify $\epsilon(f)$ by our choice of $A(f)$ where $|A(f)| = 1$. The optimum design for $A(f)$ is considered next.

III. OPTIMUM PHASE

In order to obtain a lower performance bound to the optimum design of the phase equilizer, it is necessary to determine, for a given amplitude shape, the phase associated with the minimum error, D. For this purpose, once again consider the error term,

$$\epsilon(f) = \sum_{\alpha=-\infty}^{\infty} \left| V_1\left(f - \frac{\alpha}{T}\right) \right| \exp\left(j\phi_1\left(f - \frac{\alpha}{T}\right)\right) - 1 \tag{15}$$

where $\phi_1(f)$ is the phase of $V_1(f)$.

Because $V_1(f)$ is zero, for $|f| > \frac{1}{T}$, then for $0 \leq f \leq \frac{1}{2T}$

we have

$$|\epsilon(f)|^2$$

$$= \left| |V_1(f)| \exp(j\phi_1(f)) + \left| V_1\left(f - \frac{1}{T}\right) \right| \exp\left(j\phi_1\left(f - \frac{1}{T}\right)\right) - 1 \right|^2$$

$$= \left[|V_1(f)| \cos\phi_1(f) + \left| V_1\left(f - \frac{1}{T}\right) \right| \cos\phi_1\left(f - \frac{1}{T}\right) - 1 \right]^2$$

$$+ \left[|V_1(f)| \sin\phi_1(f) + \left| V_1\left(f - \frac{1}{T}\right) \right| \sin\phi_1\left(f - \frac{1}{T}\right) \right]^2. \tag{16}$$

Clearly, minimizing $|\epsilon(f)|^2$ for each f minimizes D. Let us use the following shorthand notation

$$x = \cos \phi_1(f),$$

$$y = \cos \phi_1\left(f - \frac{1}{\lfloor T}\right),$$

and $\qquad V = | V_1(f) |,$

$$V_\tau = \left| V_1\left(f - \frac{1}{T}\right) \right|.$$

Then for f fixed we wish to minimize

$$J(x, y) = (Vx + V_\tau y - 1)^2 + [V(1 - x^2)^{\frac{1}{2}} + V_\tau(1 - y^2)^{\frac{1}{2}}]^2 \quad (17)$$

subject to the constraints

$$| x | \leq 1 \quad \text{and} \quad | y | \leq 1. \quad (18)$$

Let $\bar{F}$ be the feasible region, that is, the set of points (x, y) which satisfy inequality (18). Let F be the interior of $\bar{F}$. Since $J(x, y)$ is differentiable in F, then the optimum point $(\hat{x}, \hat{y})$ is in F if and only if

$$\left.\frac{\partial J(x, y)}{\partial x}\right|_{(\hat{x}, \hat{y})} = \left.\frac{\partial J(x, y)}{\partial y}\right|_{(\hat{x}, \hat{y})} = 0. \quad (19)$$

If $(\hat{x}, \hat{y})$ lies on the boundary of $\bar{F}$ then equation (18) is not necessarily satisfied.

We can now state the major result concisely as a theorem.

Theorem: If

$$V + V_\tau > 1 \quad \text{and} \quad |V - V_\tau| < 1 \quad (20)$$

then $(\hat{x}, \hat{y})$ ε F and $\hat{x}$ and $\hat{y}$ are determined by

$$\hat{x} = \frac{1 + (V + V_\tau)(V - V_\tau)}{2V}$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (21)$

$$\hat{y} = \frac{1 + (V_\tau + V)(V_\tau - V)}{2V_\tau},$$

yielding an error

$$J(\hat{x}, \hat{y}) = 0:$$

however, if (20) is not satisfied then

$$\hat{x} = 1, \quad \hat{y} = -1, \quad \text{if} \quad V \geq V_\tau \quad (22)$$

or

$$\hat{x} = 1, \qquad \hat{y} = 1, \quad \text{if} \quad V < V_T. \tag{23}$$

Proof: If $(\hat{x}, \hat{y}) \; \varepsilon \; F$ then equation (19) holds and solving for x and y we obtain equation (21). We must show that if inequality (20) holds, then equation (21) is consistent with $(\hat{x}, \hat{y}) \; \varepsilon \; F$. In order to do this, we define σ and Δ as:

$$\sigma = V + V_T - 1 \tag{24}$$

and

$$\Delta = 1 - (V - V_T). \tag{25}$$

Now inequality (20) implies that $\sigma > 0$ and $0 < \Delta < 2$. Therefore for $\sigma > 0$ and $0 < \Delta < 2$ we must obtain the values of $\hat{x}$ and $\hat{y}$ such that $|\hat{x}| < 1$ and $|\hat{y}| < 1$. Using equations (24) and (25) in equation (21), we obtain for $\hat{x}$

$$\hat{x} = \frac{1 + (1 + \sigma)(1 - \Delta)}{2 + \sigma - \Delta} \,,$$

$$= 1 - \frac{\sigma \Delta}{2 + \sigma - \Delta}.$$

Since we have

$$0 < \frac{\sigma \Delta}{2 + \sigma - \Delta} < 2, \tag{26}$$

then $|\hat{x}|$ must be less than one. Similarly, we obtain

$$\hat{y} = \frac{1 - (1 + \sigma)(1 - \Delta)}{\sigma + \Delta} \,,$$

$$= 1 - \frac{2 - \Delta}{1 + (\Delta/\sigma)}. \tag{27}$$

Since the quantity $1 + \Delta/\sigma$ is greater than 1 and $0 < 2 - \Delta < 2$, then $|\hat{y}| < 1$, proving that if inequality (20) is satisfied, then $(\hat{x}, \hat{y}) \; \varepsilon \; F$.

Substituting equation (21) into equation (17), we get

$$J(\hat{x}, \hat{y})$$

$$= [\tfrac{1}{2} + \tfrac{1}{2}(V + V_T)(V - V_T) + \tfrac{1}{2} - \tfrac{1}{2}(V + V_T)(V - V_T) - 1]^2$$

$$+ \left[\left(\frac{4V^2 - 1 - 2V^2 + 2V_T^2 - (V^2 - V_T^2)^2}{2} \right)^{\frac{1}{2}} \right.$$

$$\left. - \left(\frac{4V_T^2 - 1 + 2V^2 - 2V_T^2 - (V^2 - V_T^2)^2}{2} \right)^{\frac{1}{2}} \right] = 0. \tag{28}$$

If inequality (20) is not satisfied, then the point $(\hat{x}, \hat{y})$ is on the boundary. In order to minimize $J(x, y)$ we must eliminate the second term in J; that is, we must have $|\hat{x}| = |\hat{y}| = 1$. If $V \geq V_T$ then $x = 1, y = -1$ minimizes J; otherwise $x = -1, y = 1$ minimizes J.

The optimum phases $\phi_1(f)$ and $\phi_1(f - 1/T)$ are obtained from the equations

$$\phi_1(f) = \cos^{-1} x,$$

and

$$\phi_1\left(f - \frac{1}{T}\right) = \begin{cases} -\cos^{-1} y, & y > 0; \\ -\pi + \cos^{-1} y, & y < 0. \end{cases} \tag{29}$$

In this manner, the optimum phase is determined for all f for the baseband signal $V_1(f)$. Except in the Nyquist roll-off region, $\phi(f)$ is zero. In the demodulation step, there is a similar overlapping in the vestigial roll-off region and an identical procedure can be done provided the two roll-off regions do not overlap. If they do, then the process is nearly the same except that one must first determine, for the overlapping part of the vestigial region, a phase difference between the terms that add together after demodulation and then determine the phase of the baseband waveform. The solutions are, in general, not unique.

IV. REALIZABLE PHASE EQUALIZER

The optimum phase, as specified in the previous section, can not be realized. For physical systems we are restricted to those phase characteristics that can be achieved by cascading sections of constant resistance all-pass networks. A typical constant resistance all-pass network is shown in Fig. 2. Its transfer characteristic is

$$E_2/E_1 = \exp\left(-j\beta(f, f_n, b_n)\right) \tag{30}$$

where (Fig. 3)

$$\beta(f, f_n, b_n) = 2 \tan^{-1} \frac{2}{b_n} \frac{f_n f}{(f_n^2 - f^2)}. \tag{31}$$

Using N such all-pass networks yields an overall response $A(f)$, such that

$$A(f) = \exp\left(-j \sum_{n=1}^{N} \beta(f, f_n, b_n)\right). \tag{32}$$

Substituting equation (32) into equation (12) and using measured characteristics for $R(f)$, we obtain $V_1(f)$, $\epsilon(f)$ and, in turn, D in terms

LATTICE REALIZATION	BRIDGED T REALIZATION

$$Z_a = \frac{\frac{4\pi f_n R}{b_n} s}{s^2 + (2\pi f_n)^2}$$

$$Z_a Z_b = R^2$$

$$Z_a \rightarrow L \lessgtr \dashv\vdash C$$

$$L = \frac{R}{\pi b_n f_n}$$

$$C = \frac{b_n}{4\pi f_n R}$$

$$L_1 = 2\,\frac{R}{\pi b_n f_n}$$

$$0 \le C_1 < \frac{1}{2\pi f_n R}\left[\frac{b_n}{4} - \frac{1}{b_n}\right]$$

$$L_2 = \frac{1}{2}\,\frac{b_n R}{4\pi f_n}$$

$$C_2 = \frac{2\,\frac{1}{\pi b_n f_n R} C_3}{C_3 - \frac{1}{\pi b_n f_n R}}$$

$$C_3 = \frac{b_n}{4\pi f_n R} - 2C_1$$
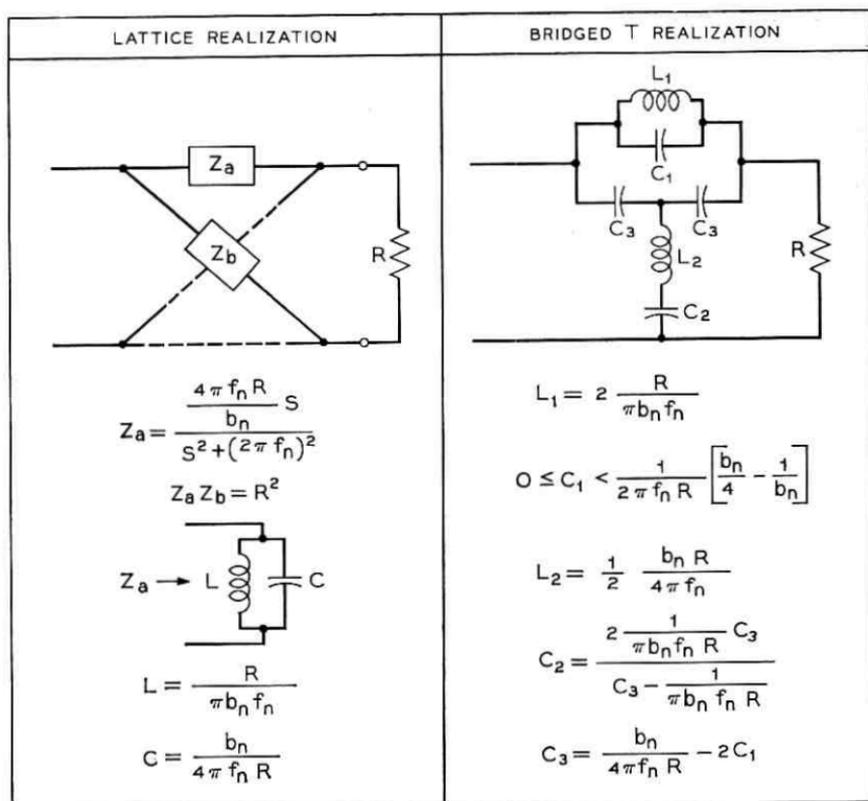
Fig. 2—Two realizations of a constant resistance all-pass network.
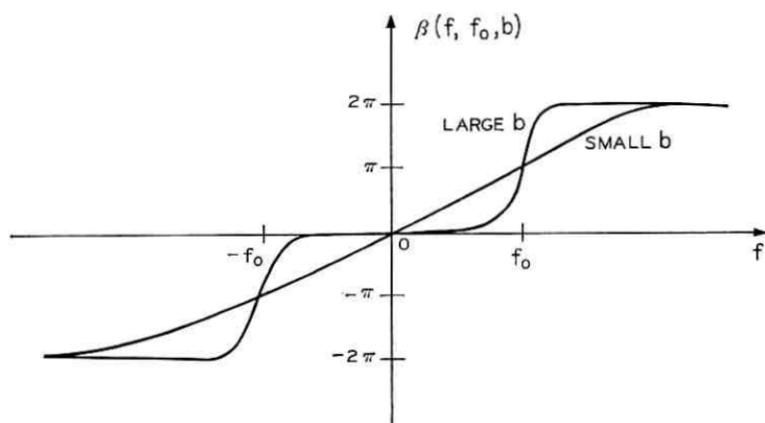
Fig. 3—Typical phase lag characteristics of an all-pass section.

of the b_n and f_n. The next step is the optimization search to choose the set of b_n's and f_n's to minimize D.

V. THE MINIMIZATION OF D

Since D is now seen to be a function of many parameters, it is more appropriately written as

$$D(\mathbf{f}, \mathbf{b}, \tau, g, \theta) = 2T \int_0^{1/2T} M(f) \mid \epsilon(f, \mathbf{f}, \mathbf{b}, \tau, g, \theta) \mid^2 df \qquad (33)$$

where $V_1(f)$ in equation (14) is equal to $gV_2(f)$ and

$$V_2(f) = \left[\mid R(f + f_c) \mid \exp \left[j\left(\phi(f + f_c) - \sum_{n=1}^{N} \beta(f + f_c, f_n, b_n) + \theta \right) \right] \right.$$

$$+ R(-f + f_c) \exp \left[-j\left(\phi(f + f_c) - \sum_{n=1}^{N} \beta(-f + f_c, f_n, b_n) + \theta \right) \right] \right]$$

$$\cdot \exp \left(j2\pi f\tau \right) \operatorname{rect} \left(\frac{fT}{2} \right) : \qquad (34)$$

g is a scale factor for $V_1(f)$; $\mathbf{b}$ and $\mathbf{f}$ are parameter vectors

$$\mathbf{b} \equiv \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{bmatrix}, \qquad \mathbf{f} \equiv \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{bmatrix}, \qquad (35)$$

and $\phi(f)$ is the phase of $R(f)$.

If we adopt the following inner product notation

$$[\alpha(f), \beta(f)]_M = 2T \int_0^{1/2T} M(f)\alpha(f)\beta^*(f) \, df, \qquad (36)$$

then D can be concisely written as

$$D(\mathbf{f}, \mathbf{b}, \tau, g, \theta) = [\epsilon(f, \mathbf{f}, \mathbf{b}, \tau, g, \theta), \epsilon(f, \mathbf{f}, \mathbf{b}, \tau, g, \theta)]_M . \qquad (37)$$

We must first minimize D with respect to the gain, g, the delay, τ, and phase offset, θ, before determining what changes to make in $\mathbf{b}$ and $\mathbf{f}$. After changing $\mathbf{b}$ and $\mathbf{f}$, we repeat the procedure by optimizing g, τ, and θ, and then obtaining new $\mathbf{b}$ and $\mathbf{f}$. After each iteration, we evaluate D and stop when D ceases to improve more than a predetermined amount.

To ascertain the optimum value for g, we solve $\partial D/\partial g = 0$ for g

and obtain

$$g = \frac{\text{Re}\left\{\left[\sum\limits_{\alpha=-1}^{1} V_2\!\left(f - \frac{\alpha}{T}\right), 1\right]_M\right\}}{\left[\sum\limits_{\alpha=-1}^{1} V_2\!\left(f - \frac{\alpha}{T}\right), \sum\limits_{\alpha=-1}^{1} V_2\!\left(f - \frac{\alpha}{T}\right)\right]_M}. \tag{38}$$

g does depend on the phase of $V_2(f)$; however, if the roll-off region is narrow, g is relatively insensitive to it, especially when $\mathbf{b}$ and $\mathbf{f}$ are near their optimum values. We can, therefore, eliminate repeating the calculation of g at each iteration by replacing $V_2(f)$ by $\mid V_2(f) \mid$ in equation (38) to obtain

$$g = \frac{\left[\sum\limits_{\alpha=-1}^{1} \left| V_2\!\left(f - \frac{\alpha}{T}\right)\right|, 1\right]_M}{\left[\sum\limits_{\alpha=-1}^{1} V_2\!\left(f - \frac{\alpha}{T}\right), \sum\limits_{\alpha=-1}^{1} V_2\!\left(f - \frac{\alpha}{T}\right)\right]_M} \tag{39}$$

which, since it is independent of phase, need be calculated only once.

We shall take a similar approach for the evaluation of θ. Optimization with respect to θ would have to be done jointly with optimization with respect to τ. This would enormously complicate the minimization of D, since it would have to be repeated at each iteration. Instead we can choose θ as

$$\theta = -\phi(f_c) + \sum_{n=1}^{N} \beta(f_c, f_n, b_n) + \theta_0 \tag{40}$$

where θ_0 can be chosen so that θ is very close to its optimum value for the final set of $\mathbf{b}$ and $\mathbf{f}$. Thus, as $\mathbf{b}$ and $\mathbf{f}$ improve, the value chosen for θ approaches its optimum. θ_0 depends on $R(f)$ and is determined from the optimum phase.

With g and θ fixed, the optimization with respect to τ has been simplified. τ_{opt} is a solution of the equation

$$\frac{\partial D}{\partial \tau} = 2 \, \text{Re}\left[\epsilon(f, \mathbf{f}, \mathbf{b}, \tau), \frac{\partial \epsilon(f, \mathbf{f}, \mathbf{b}, \tau)}{\partial \tau}\right]_M,$$

$$= 2 \, \text{Re}\left[\epsilon(f, \mathbf{f}, \mathbf{b}, \tau), \sum_{\alpha=0}^{1} j2\pi\!\left(f - \frac{\alpha}{T}\right) V_1\!\left(f - \frac{\alpha}{T}\right)\right]_M = 0. \tag{41}$$

An initial value of τ is chosen as described below and $\partial D/\partial \tau$ is evaluated. If $\partial D/\partial \tau > 0$, then τ is decremented, or if $\partial D/\partial \tau < 0$, then τ is incremented, until $\partial D/\partial \tau$ changes sign. Using the last two values of $\partial D/\partial \tau$, we can employ the method of "false position" (regula falsi)[5] to deter-

mine a value of $\partial D/\partial \tau < 10^{-5}$. The τ yielding this value of $\partial D/\partial \tau$ is accepted as τ_{opt} .

τ_i , the initial value of τ, is chosen to correspond to the best linear mean square error curve fit to the resulting baseband phase. Thus we have

$$\tau_i = -\mu/2\pi$$

where

$$\mu = \frac{\sum\limits_{k=1}^{K} (f_k - f_c)\psi(f_k)}{\sum\limits_{k=1}^{K} (f_k - f_c)^2} , \tag{42}$$

$$\psi(f) = \phi(f) - \sum\limits_{n=1}^{N} \beta(f, f_n , b_n) + \theta,$$

and the f_k are the discrete values of frequency between $f = 0$ and $f = 1/2T$ used in evaluating D by a finite sum.

The minimization is done by a parameter search utilizing the Fletcher–Powell algorithm.[6,7] Since the effectiveness of the method depends on the accuracy with which the gradient is determined, it is particularly advantageous to use an analytic expression for its calculation. This expression is presented in Appendix A.

The computer program to achieve the above minimization was written in its entirety by Mrs. Barbara E. Forman for the CDC 3300 and the IBM 360/75.

Practical problems can arise in performing the parameter search. They are due mostly to the existance of many local minima. The Fletcher–Powell algorithm will approach a local minimum as readily as a global minimum. This difficulty is mitigated by three factors: (i) We have been able to determine, for a given amplitude characteristic, the theoretical phase which will yield the lowest value of D and thereby determine the practical utility of trying to find the global optimum—we have been able to find local minima which yield digital mean square errors insignificantly greater than the lower bound. (ii) We are able to specify the largest acceptable value of D. (iii) The parameter values at the global optimum may be impractical to realize by physical networks.

The system error rate objective for L-4 and L-5 is that the probability of error (P_E) be less than 10^{-8} per regenerative section. Walter E. Norris[8] has been able to relate a P_E of 10^{-8} to a value of D of approximately 10^{-3}. This relationship is not unique, but it is accurate to within a practically

acceptable tolerance. We set our requirement for the theoretical filter at $D = 5 \times 10^{-4}$ since the physical realization will result in a larger error. We will accept a larger D for the implemented filter provided its value for D is reasonably close to this. If a minimization design yields a value of D near the lower bound and below the requirement, it is of no serious consequence that we do not achieve the global minimum. Since our design is based on a mathematical description of the network, it is important, however, that the parameter values be such that the physical realization approximates closely its mathematical representation. It is possible for a local optimum in a favorable parameter region to produce a better result than the global optimum in an unfavorable region due to attendant amplitude distortion and problems in its manufacture.

VI. COMPARISON OF DESIGNS

The data available from the L-4 system affords an opportunity to compare the conventional techniques of phase equalization with our method. The digital sequences on the system will be encoded in a partial response[9] format which yields an $M(f)$ of $M(f) = 4 \sin^2 4\pi fT$. The amplitude characteristic of the cascade of filters for digital transmission is shown in Fig. 4. By conventional methods of phase equalization, 20 all-pass sections were designed to reduce the phase deviation (deviation from linear phase, see Appendix B) to the "best achievable" level across the entire bandwidth. The theoretical performance of the 20-section equalizer (which does not consider the amplitude distortion of the phase equalizer) yields a phase deviation as shown in Fig. 5 and a digital mean square error of 1.45×10^{-4}. The "b" parameters (a measure of the steepness of the slope of the phase characteristics of an all-pass section, see Fig. 3) of several all-pass sections had values well above 25, the highest being 34.39. Values of b above 25 are undesirable for purposes of manufacture. The actual phase deviation as measured by J. S. Ronne[10] is shown in Fig. 6 and yields an error of 1.05×10^{-2}.

An optimum phase deviation for the system, as determined by the method described in an earlier section, is shown in Fig. 7 and yields a digital mean square error of 5.69×10^{-5}. If the phase had been exactly linear, the error would be 7.52×10^{-5}. Our program will attempt to achieve the optimum phase.

We must first determine the number of sections necessary to yield an acceptable error. In Fig. 8, the error, D, for the optimum design is plotted as a function of the number of all-pass sections. The curve shows
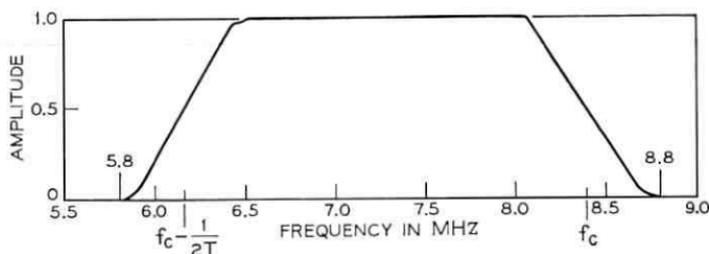
Fig. 4—Amplitude characteristics of cascade of filters.

that one obtains diminishing returns after 10 sections and that these 10 sections have a theoretical error (1.41×10^{-4}) less than that of the original 20-section design. Taking into account the reduced amplitude distortion in realizing 10 sections as opposed to 20 sections, and a minimal theoretical improvement in going from 10 to 20, we would, in all likelihood, achieve better performance from the optimum 10-section design than from the optimum 20-section design. A comparison of calculated amplitude distortions is presented in Fig. 9. It indicates that it is very desirable to use as few sections as is possible.

The optimum 10-section phase equalizer (all of whose b parameters
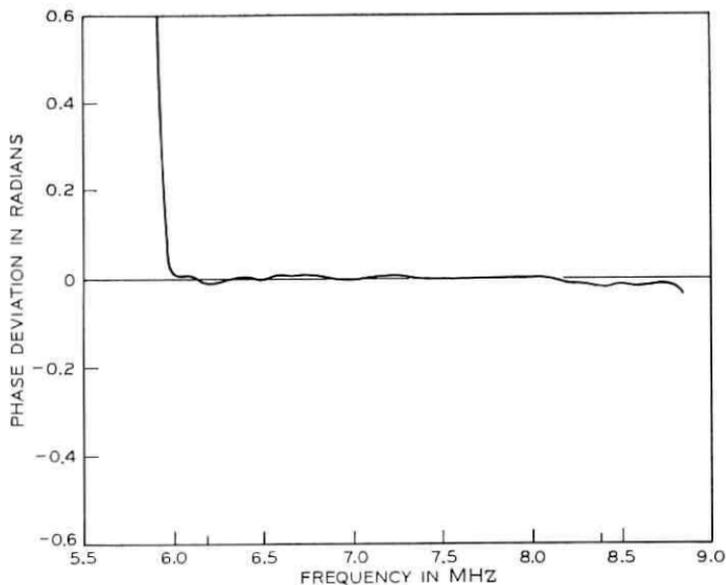
Fig. 5—Theoretical phase deviation of original 20-section design.
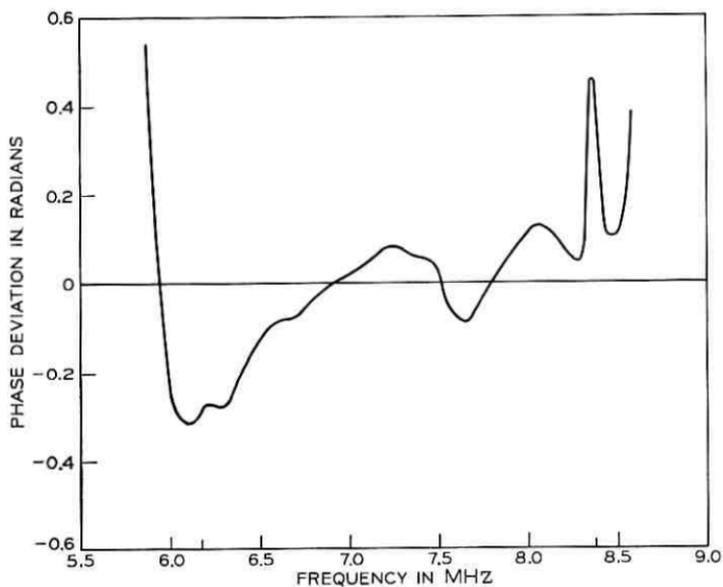
Fig. 6—Phase deviation of original 20-section design as determined from measured characteristics.
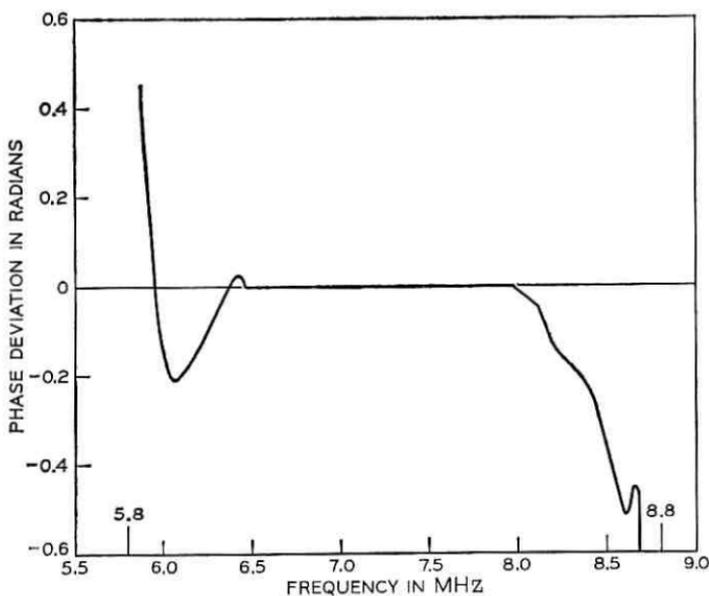

Fig. 7—Phase deviation of optimum phase.
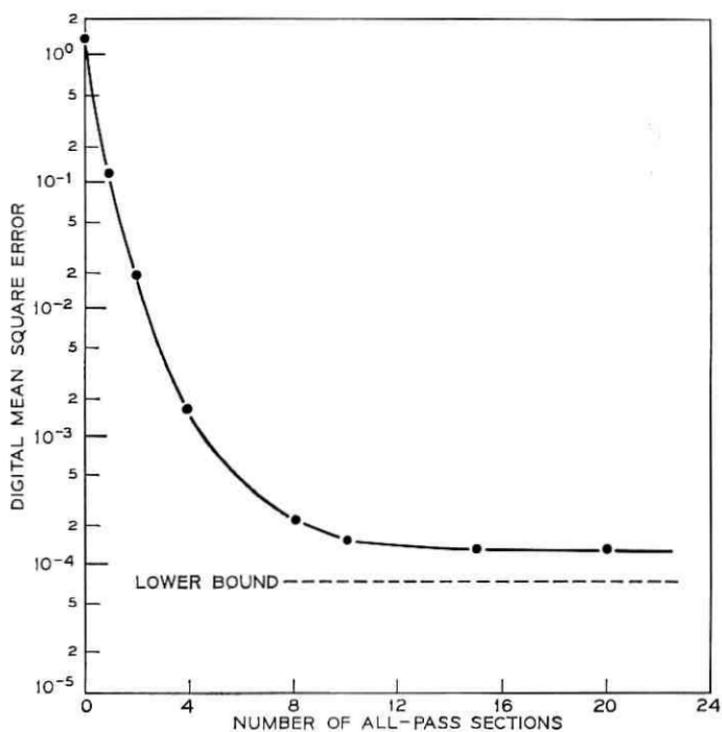
Fig. 8—Minimum digital mean squared error obtained with various numbers of all-pass sections.
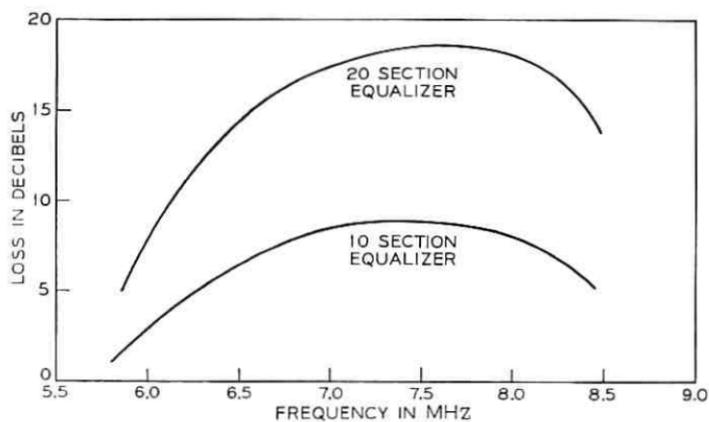
Fig. 9—Calculated loss characteristics of non-ideal phase equalizers.
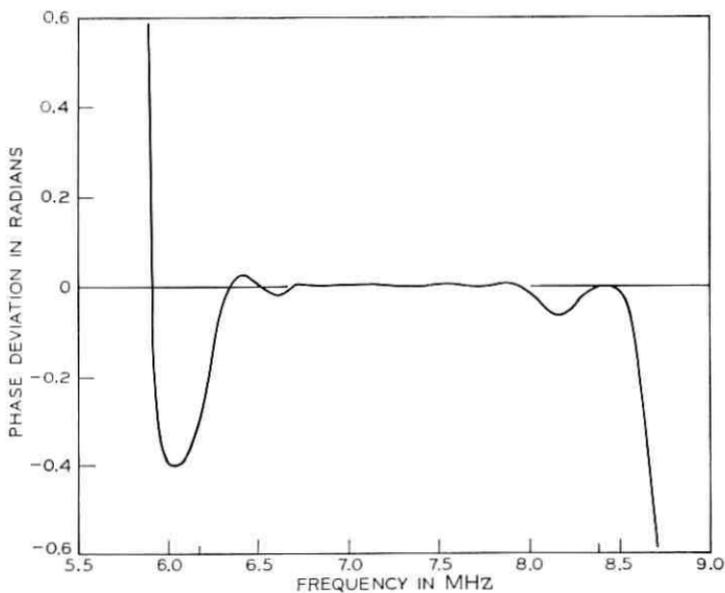
Fig. 10—Theoretical phase deviation of 10-section optimum design.
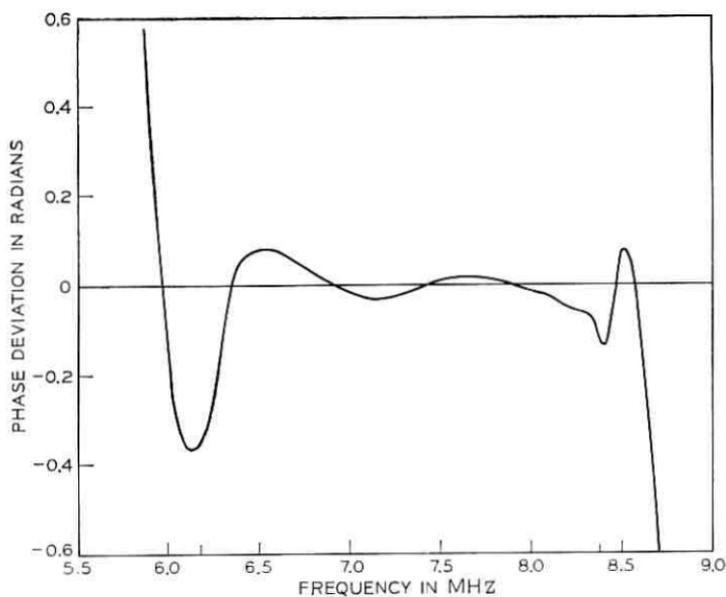

Fig. 11—Phase deviation of optimum 10-section design as determined from measured characteristics.

TABLE I—THE DIGITAL MEAN SQUARE ERROR FOR THE TWO PHASE
EQUALIZER DESIGNS

	Theoretical	Measured
Original Design (20 Sections)	1.45×10^{-4}	1.05×10^{-2}
New Design (10 Sections)	1.41×10^{-4}	1.99×10^{-3}

were equal to about 20) was built and tested. The phase deviation of
the theoretical design is shown in Fig. 10 with a corresponding error
of 1.41×10^{-4}, while the phase deviations determined from the
measurements of J. S. Ronne are shown in Fig. 11. This corresponds to
an error of 1.99×10^{-3}. Table I summarizes the results.

A P_E test was made on both the 10-section and 20-section designs at
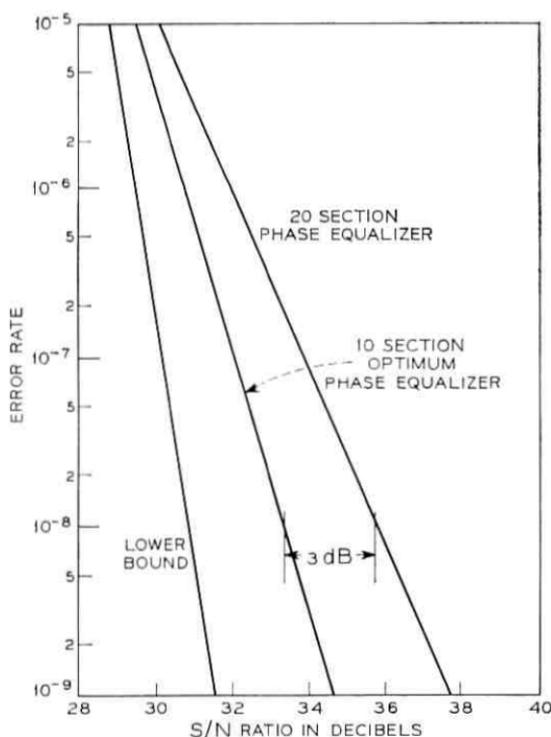
Fig. 12—P_E vs S/N for the 10- and 20-section phase equalizers.

a signal-to-noise (S/N) ratio of 40 dB. The error rate of the 20-section was 2000 times that of the 10-section. The two designs were also tested under the circumstances in which they would be used. Attached to the system is an adaptive equalizer which automatically adjusts to partly compensate for distortions in the system. The P_E of both designs were taken as a function of S/N, with the adaptive equalizer included, and the results are presented in Fig. 12. These measurements indicate that the new design yields 3 dB more S/N margin at $P_E = 10^{-8}$ than the previous design. The lower bound on P_E obtained from additive noise considerations (assuming a Nyquist pulse shape) by F. S. Hill[11] indicates that there is little room for additional improvement through manipulation of the pulse shape.

The advantages of the new design can be summarized as follows: (*i*) it is less expensive (fewer sections), (*ii*) it is easier to manufacture, and, (*iii*) most importantly, it yields better performance.

VII. CONCLUSIONS

The use of the digital mean square error, D, as a design criterion has proved to be of significant value in the design of the phase equalization for the transmission of digital information over the L-4 channel. Further studies indicate that comparable advantages exist in applying this technique to L-5 and other L-4 network designs. The use of the digital mean square error criterion is not limited to the phase equalization described here. Not mentioned above, but necessary, is amplitude equalization to compensate for the amplitude distortion introduced by the physical filters and all-pass sections. The design of this equalization is done optimally using the same techniques.

With conventional techniques, the design of the band limiting filters considers amplitude characteristics alone. The passband of these filters should be designed to minimize D, in which case both the amplitude and phase influence the result. The next step, therefore, is the jointly optimum design of these filters, together with equalization, on a digital mean square error basis.

equalizers, and to T. H. Simmonds, Jr. and F. R. Bies for many helpful discussions.

APPENDIX A

The Gradient

The error D can be written

$$D(\mathbf{f}, \mathbf{b}) = 2T \int_0^{1/2T} M(f) \mid \epsilon(f, \mathbf{f}, \mathbf{b}) \mid^2 df \qquad (43)$$

where, for the band limited $V(f)[\mid V(f) \mid = 0, \mid f \mid > 1/T)]$

$$\epsilon(f) \equiv V_1(f) + V_1\left(f - \frac{1}{T}\right) - 1 \qquad (44)$$

and

$$V_1(f) \equiv g\left\{\mid R(f + f_c) \mid \exp\left[j\left(\phi(f + f_c) - \sum_{n=1}^{N} \beta(f + f_c, f_n, b_n) + \theta\right)\right]\right.$$

$$+ \mid R(-f + f_c) \mid$$

$$\cdot \exp\left[-j\left(\phi(-f + f_c) - \sum_{n=1}^{N} \beta(-f + f_c, f_n, b_n) + \theta\right)\right]\right\}$$

$$\cdot \exp(j2\pi f\tau) \operatorname{rect}\left(\frac{fT}{2}\right).$$

The gradient $\mathbf{G}$ is

$$\mathbf{G} \equiv \begin{pmatrix} \dfrac{\partial D}{\partial b_1} \\[2mm] \dfrac{\partial D}{\partial b_2} \\[2mm] \vdots \\[2mm] \dfrac{\partial D}{\partial b_N} \\[2mm] \dfrac{\partial D}{\partial f_1} \\[2mm] \vdots \\[2mm] \dfrac{\partial D}{\partial f_N} \end{pmatrix} \qquad (45)$$

where

$$\frac{\partial D}{\partial b_n} = 4T \int_0^{1/2T} M(f) \ \mathrm{Re} \ \left\{ \epsilon(f) \left[\frac{\partial V_1^*(f)}{\partial b_n} + \frac{\partial V_1^*\!\left(f - \frac{1}{T}\right)}{\partial b_n} \right] \right\} df,$$

$$\frac{\partial D}{\partial f_n} = 4T \int_0^{1/2T} M(f) \ \mathrm{Re} \ \left\{ \epsilon(f) \left[\frac{\partial V_1^*(f)}{\partial f_n} + \frac{\partial V_1^*\!\left(f - \frac{1}{T}\right)}{\partial f_n} \right] \right\} df,$$

(46)

and

$$\left\{ \begin{matrix} \dfrac{\partial V_1}{\partial b_n} \\[2mm] \dfrac{\partial V_1}{\partial f_n} \end{matrix} \right\} = g \exp{(j2\pi f\tau)} \Bigg[\ | R(f + f_c) | \ \exp\left(j\Big\{ \phi(f + f_c) - \phi(f_n) \right.$$

$$- \sum_{n=1}^{N} [\beta(f + f_c , f_n , b_n) - \beta(f_c , f_n , b_n)] + \theta_0 \Big\}\right)$$

$$\cdot j \left\{ \begin{matrix} \dfrac{-f_n(f + f_c)[(f + f_c)^2 - f_n^2]}{[f_n(f + f_c)]^2 + \dfrac{b_n^2}{4}[f_n^2 - (f + f_c)^2]^2} + \dfrac{f_n f_c (f_c^2 - f_n^2)}{(f_n f_c)^2 + \dfrac{b_n^2}{4}[f_n^2 - f_c^2]^2} \\[6mm] \dfrac{b_n(f + f_c)[f_n^2 + (f + f_c)^2]}{[f_n(f + f_c)]^2 + \dfrac{b_n^2}{4}[f_n^2 - (f + f_c)^2]^2} - \dfrac{b_n f_c (f_n^2 + f_c^2)}{(f_n f_c)^2 + \dfrac{b_n^2}{4}[f_n^2 - f_c^2]^2} \end{matrix} \right\}$$

$$+ \ | R(-f + f_c) | \ \exp\left(-j\Big\{ \phi(-f + f_c) - \phi(f_c) \right.$$

$$- \sum_{n=1}^{N} [\beta(-f + f_c , f_n , b_n) - \beta(f_c , f_n , b_n)] + \theta_0 \Big\}\right)$$

$$\cdot j \left\{ \begin{matrix} \dfrac{f_n(-f + f_c)[(-f + f_c)^2 - f_n^2]}{[f_n(-f + f_c)]^2 + \dfrac{b_n^2}{4}[f_n^2 - (-f + f_c)^2]^2} - \dfrac{f_n f_c (f_c^2 - f_n^2)}{(f_n f_c)^2 + \dfrac{b_n^2}{4}[f_n^2 - f_c^2]} \\[6mm] \dfrac{-b_n(-f + f_c)[f_n^2 + (-f + f_c)^2]}{[f_n(-f + f_c)]^2 + \dfrac{b_n^2}{4}[f_n^2 - (-f + f_c)^2]^2} + \dfrac{b_n f_c (f_n^2 + f_c^2)}{(f_n f_c)^2 + \dfrac{b_n^2}{4}[f_n^2 - f_c^2]} \end{matrix} \right\} \Bigg]$$

$$\cdot \mathrm{rect} \left[\frac{fT}{2} \right].$$

(47)

APPENDIX B

Phase Deviation

The phase deviation is defined as the deviation of the phase from the linear phase which best (in a mean square error sense) fits the phase.

Therefore, if $\psi(f)$ is the phase and $af + b$ is the best linear fit between f_1 and f_2 then the phase deviation ϕ_D is

$$\phi_D(f) = \psi(f) - (af + b)$$

where

$$a = \frac{\int_{f_1}^{f_2} f\psi(f)\, df - \frac{1}{f_2 - f_1}\left(\int_{f_1}^{f_2} \psi(f)\, df\right)\left(\int_{f_1}^{f_2} f\, df\right)}{\int_{f_1}^{f_2} f^2\, df - \frac{1}{f_2 - f_1}\left(\int_{f_1}^{f_2} f\, df\right)^2}$$

and

$$b = \frac{1}{f_2 - f_1}\left[\int_{f_1}^{f_2} \psi(f)\, df - a\int_{f_1}^{f_2} f\, df\right].$$

Discretizing the above equations, we obtain

$$a = \frac{\displaystyle\sum_{k=k_1}^{k_1+n} f_n\psi(f_n) - \frac{\left(\displaystyle\sum_{k=k_1}^{k_1+n} f_k\right)\left(\displaystyle\sum_{k=k_1}^{k_1+n} \psi(f_k)\right)}{n}}{\left[\displaystyle\sum_{k=k_1}^{k1+n} f_k^2 - \frac{\left(\displaystyle\sum_{k=k_1}^{k_1+n} f_k\right)^2}{n}\right]}$$

$$b = \frac{\displaystyle\sum_{k=k1}^{k+n} \psi(f_k) - a\displaystyle\sum_{k=k1}^{k1+n} f_k}{n}.$$

where $f_{k_1} = f_1$, $f_{k_1+n} = f_2$ and

$$\Delta f = f_{k_1+m} - f_{k_1+m-1} = \frac{f_2 - f_1}{n},$$

REFERENCES

1. Spaulding, D. A., "Synthesis of Pulse-Shaping Networks in the Time Domain," B.S.T.J., 48, No. 7 (September 1969), pp. 2425–2444.
2. Tufts, D. W., "Nyquist's Problem—The Joint Optimization of Transmitter and Receiver in Pulse Amplitude Modulation," Proc. IEEE, 53, No. 3 (March 1965), pp. 248–259.

3. Maurer, R. E., "The Optimal Equalization of Random Channels," Ph.D. thesis, Northeastern University, 1968, pp. 231–238.
4. Berger, T., and Tufts, D. W., "Optimum Pulse Amplitude Modulation, Part I," IEEE Trans. on Information Theory, *IT-13*, No. 2 (April 1967), pp. 196–208.
5. Hildebrand, F. B., *Introduction to Numerical Analysis*, New York: McGraw-Hill, 1965, pp. 446.
6. Fletcher, R., and Powell, M. J. D., "A Rapidly Convergent Descent Method of Minimization," Computor Journal, *6*, No. 4 (July 1963), pp. 163–168.
7. Kowalik, J., and Osborne, M. R., *Methods of Unconstrained Optimization Problems*, New York; Elsevier, 1968, Sec. 3.7, pp. 45–48.
8. Norris, W. E., Unpublished work.
9. Franks, L. E., *Signal Theory*, Englewood Cliffs, N.J.: Prentice-Hall, 1969 pp. 219–220.
10. Ronne, J. S., Unpublished work.
11. Hill, F. S., Unpublished work.

A Unified Formulation of Segment Companding Laws and Synthesis of Codecs and Digital Compandors

By H. KANEKO

Segment companding laws are being considered for use in PCM to cater to the possibility of digital processing of uniformly quantized signals. A unified formulation of the segment companding laws is presented, which permits the detailed structure of quantization to be explicitly characterized. Most important, it is shown that this formulation yields algorithms for the systematic synthesis of coders, decoders, and digital compandors. Examples of circuits synthesized from the algorithms are shown for two segment law companding families—the μ-law and the A-law.

I. INTRODUCTION

Instantaneous companding of PCM is used to maintain a reasonably constant signal to quantizing noise ratio for a wide dynamic range of the input voice signal. A logarithmic "μ-law" is a typical "smooth" input/output transfer characteristic that has been approximated in the D1 channel bank realized by a diode compandor.[1,2] In contrast to the smooth law, segment type companding laws, which are piecewise linear approximations to the continuous laws, have also been considered. In early experimental PCM systems a segment law was used primarily to avoid implementation difficulties associated with tracking of diode compandors.[3,4] However, since the mid-sixties segment laws have been reconsidered from the view point of digital linearization to cater to the possibility of digital processing in an integrated digital network. Two of the segment law families have been stressed—the 13-segment A-law,[5,6] and the 15-segment μ-law used in the D2 channel bank.[7]

The major advantage of the segment laws resides in the digital linearization feature, through which digital processing such as digital

filtering, conferencing, net loss adjustment, echo suppression, equalization, companding law conversion, and so on, can be performed efficiently.[8] Also, it has been argued that the segment laws, being well defined laws, should readily permit an international interconnection on a digital basis. However, the defining features of the segment laws have not been clearly drawn.

Many attempts have been made to characterize a particular law by table-look-up, by a set of equations, or by forming algorithms,[5,8] but important parameters defining the particular law have not been delineated. What is most important, the lack of an explicit expression for the family of segment laws has produced difficulty in the understanding and design of codecs and digital compandors.

In this paper a unified formulation is presented, which leads to the systematic synthesis of codecs and digital compandors for the binary code.[*] In the second section, starting from conventional concepts of the μ-law and the A-law, we derive expressions for the decoder output level. Output levels are given in terms of a standard form with segment edge parameter a and centering parameter c. Coder decision levels may also be expressed by the same formulation with the addition of the coding parameter b, thus completely specifying the companding law. This approach avoids table-look-up. Relations between the step size, effective coder input, transfer gain, and other parameters of the law are derived.

In the third section, we show that this formulation is directly related to the codec circuit implementation, and these circuits can be synthesized in a systematic and straightforward manner. Unipolar decoders are first synthesized from the formulation, and the synthesis is extended to bipolar decoders including the sign bit. We also show that the coding parameter b is related to the design of logic in a sequential comparison coder.

Digital compandors can also be synthesized by using the same formulation. In the fourth section we display digital expansion and compression algorithms, and compandor circuits synthesized therefrom. At Bell Telephone Laboratories, pioneering studies on digital compandors have been made by T. P. Stanley, M. R. Aaron and D. G. Messerschmitt in unpublished memoranda. W. L. Montgomery has also done work in this area.[8] The synthesis technique developed here serves to unify these separate works, and makes it possible to design the circuits in a systematic way.

---

[*] Extension to arbitrary bases follows in a straightforward manner and will not be considered.

II. FORMULATION OF SEGMENT LAWS

2.1 *Definition of Variables*

Suppose that a digital compressed code X is composed of m binary digits called "characteristic bits" representing the segment number L, and n binary digits called "mantissa bits" representing the quantizing step V in a segment, as shown in Fig. 1. The total number M of segments in one polarity is equal to 2^m, and the total number N of quantizing steps within a segment is equal to 2^n. Then the digital representation of the compressed signal X is expressed in terms of L and V, as

$$X(L, V) = V + N \cdot L \qquad (1)$$

or

$$V = X \bmod N$$
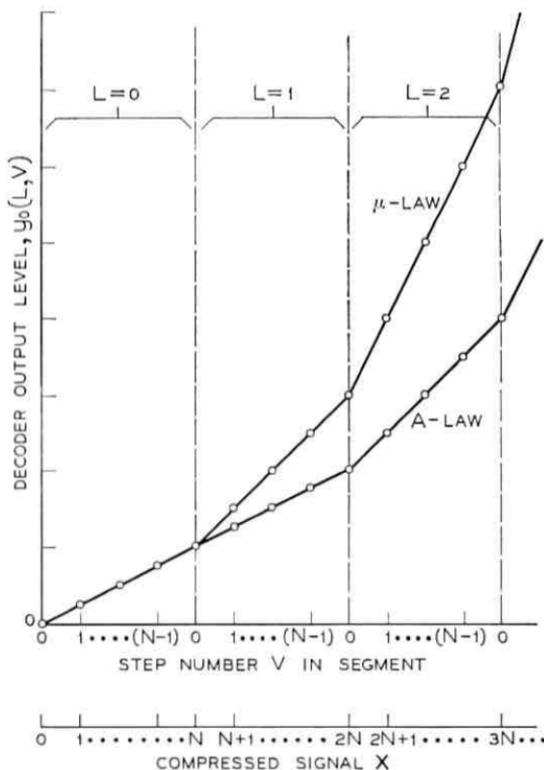
Fig. 1—Segment companding laws—standard form.

where X, L and V are defined in the regions as

$$
\left.\begin{aligned}
X &= \text{digital compressed signal} \in \{0, 1, \cdots, MN - 1\}, \\
L &= \text{segment number} \in \{0, 1, \cdots, M - 1\}, \text{ and} \\
V &= \text{step number within a segment} \in \{0, 1, \cdots, N - 1\}.
\end{aligned}\right\} \quad (2)
$$

The definition of X, L and V beginning from "0" permits a consistent algebraic treatment and yields a straightforward circuit design.

If we denote the decoder output level by y_0, then it is a function of the digital compressed signal and $y_0 = y_0(X) = y_0(L, V)$. The quantizing step size is then given by the difference of two adjacent output levels, and is

$$
\left.\begin{aligned}
\Delta_0(L) &\equiv y_0(L, V + 1) - y_0(L, V), & V &\neq N - 1 \\
\Delta_0'(L) &\equiv y_0(L + 1, 0) - y_0(L, N - 1), & V &= N - 1
\end{aligned}\right\} \quad (3)
$$

where $\Delta_0'(L)$ is the step size at the segment edge. Within a segment the step size $\Delta_0(L)$ is constant and independent of V.

2.2 Definition of the Segment μ-law

We will define the segment μ-law and derive an expression for an output level as a function of L and V. The "so-called" segment μ-law should satisfy the following:

(i) Except for the segment edges, the ratio of the step sizes of the adjacent segments is equal to 2; that is,

$$
\frac{\Delta_0(L + 1)}{\Delta_0(L)} = 2, \qquad L \in \{0, 1, \cdots, M - 2\}, \quad (4)
$$

(ii) $y_0(0, 0) = c$ \qquad (centering), \hfill (5)

(iii) $\Delta_0(0) = 1$ \qquad (normalization). \hfill (6)

Solving the difference equation (4) subject to equations (3), (5) and (6), we have

$$
\begin{aligned}
y_0(L, V) &= 2^L(V + N + a) - N - a + c \\
&= f_\mu(L, V + a) - a + c
\end{aligned} \quad (7)
$$

for an arbitrary real a, where $f_\mu(L, V)$ is the standard form of the segment μ-law, and is

$$
f_\mu(L, V) \equiv 2^L(V + N) - N. \quad (8)
$$

The standard form is interpreted in the following way. If we take a

continuous variable v such that

$$v \in [0, N]$$

then $f(L, v)$ represents a continuous segment μ-law as shown in Fig. 1. The continuous segment μ-law $f_\mu(L, v)$ is a set of segment lines whose segment edges $f_\mu(L, N) = f_\mu(L + 1, 0)$ satisfy the so called μ-law[1] for $x = LN + v$, namely:

$$\frac{x}{E_1} = \frac{\log (1 + \mu y/E_2)}{\log (1 + \mu)} \tag{9}$$

for normalization constants $E_1 = MN$, $E_2 = N(2^M - 1)$, and $\mu = 2^M - 1$. In other words, the continuous standard form gives an exact "chord type" approximation to the smooth μ-law. On this continuous segment law the discrete point V should be defined as the truncation of v.

The output level (7) is then the modified version of the standard form with parameters a and c, and it constitutes a class of "μ-like" laws. The centering parameter c determines the origin, and the edge effect parameter a describes the discontinuity of step size at the segment edges. The continuous segment law varies with parameter a as shown in Fig. 2, and produces discontinuities at the segment edges.

2.3 Extension to the Segment A-law

To obtain the segment A-law,[5] we replace the condition (i) by
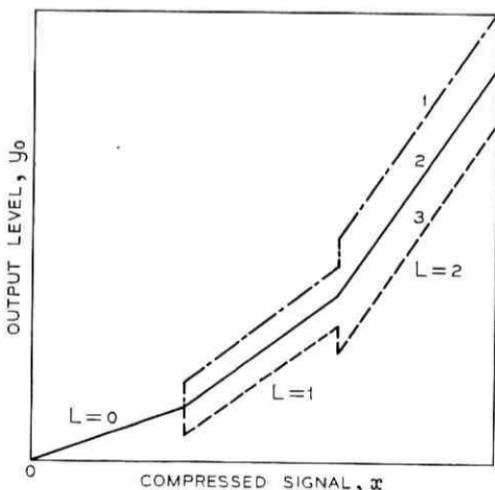
Fig. 2—Effect of segment edge parameter a: Curve 1—$a > 1$; Curve 2—$a = 0$ and Curve 3—$a < 0$.

$(i)'$ Except for the segment edges the ratio of the step size of the adjacent segments is

$$\frac{\Delta_0(L+1)}{\Delta_0(L)} = \begin{cases} 1, & L = 0, \\ 2, & L \in \{1, 2, \cdots, M-2\}. \end{cases} \tag{10}$$

This implies that the 0th and 1st segments constitute a long colinear segment and the step size ratio of the adjacent segments elsewhere is equal to 2. Then solving equation (10) subject to equations (3), (5) and (6) we have

$$y_0(L, V) = f_A(L, V + a) - a + c \tag{11}$$

where $f_A(L, V)$ is the standard form of the segment A-law and is

$$f_A(L, V) = \begin{cases} V, & L = 0, \\ 2^{L-1}(V + N), & L \neq 0. \end{cases} \tag{12}$$

Note that equation (11) is identical with equation (7) for the μ-law, and the output level for the A-law is characterized by the standard form $f_A(L, V)$ in equation (12).

As in the μ-law case, where a discrete V is replaced by a continuous variable $v \in [0, N]$ in the standard form $f(L, V)$, a "chord type" approximation to the so called A-law is obtained. The A-law has been defined as[5,9]

$$\frac{x}{E_3} = \begin{cases} \dfrac{A(y/E_4)}{1 + \log_K A}, & y \in [0, E_4/A] \\[3mm] \dfrac{1 + \log_K (Ay/E_4)}{1 + \log_K A}, & y \in [E_4/A, E_4] \end{cases} \tag{13}$$

for $K = 2$, $A = 128,$* $E_3 = MN$, and $E_4 = N \cdot 2^M$.

Expression (11) applies for any type of segment companding law if the standard form $f_A(L, v)$ is appropriately defined to yield a continuous piece-wise linear approximation to a continuous curve.

2.4 Decision Levels

In the foregoing we have concentrated on the decoder output level because the level of primary importance in companding is the output level. The coder decision level should then be determined to yield an appropriate range of the input signal corresponding to that output level.

---

* The parameter values $A = 87.6$, $K = \epsilon$ (base of natural logarithms) are often used.[5] However, to be consistent with a "chord type" approximation, $K = 2$ and $A = 128$ are used here.

First, the decoder output level should be a good representative of the signal range between two adjacent decision levels. It has been shown elsewhere that for a given set of output levels the locally optimum decision levels, which minimize the mean squared error, should lie halfway between two adjacent output levels irrespective of the input signal distribution.[10] This can be achieved, except at the segment edges, by shifting the output level a half step up or down in the compressed signal domain; that is, the decision level is

$$y_d(L, V) \equiv y_0(L, V + b) \tag{14}$$

where b is the coding parameter and takes on the values $+0.5$ or -0.5. When $b = +0.5$, $y_d(L, V)$ gives the upper bound of the input range corresponding to $X(L, V)$; when $b = -0.5$, it corresponds to the lower bound. The coder decision level y_d implies that an analog input y is encoded into $X(L, V)$ if

$$y_d(X - 1 \mid b = +0.5) \leqq y < y_d(X \mid b = +0.5) \tag{15}$$

for the upper bound approach ($b = +0.5$), and if

$$y_d(X \mid b = -0.5) \leqq y < y_d(X + 1 \mid b = -0.5) \tag{16}$$

for the lower bound approach ($b = -0.5$). In general, these two approaches are identical at all points except at segment edges. Now, for any points including segment edges, it can be shown (Appendix A) that the two approaches coincide if and only if the edge effect parameter a is equal to 0.5. As we will see in Section 3.4, the coding parameter b is closely related to the binary coding logic in a sequential comparison coder.

2.5 Codec Transfer Characteristics

The decoder output level and the coder decision level have been defined. Now, the characteristic from the coder input to the decoder output as shown in Fig. 3 can be expressed explicitly by two equations with the auxiliary variables L, V and the parameters a, b and c; that is,

$$\left. \begin{array}{l} y_0(L, V) = f(L, V + a) - a + c \\ y_d(L, V) = f(L, V + a + b) - a + c \end{array} \right\} \tag{17}$$

where $f(L, V)$ is the standard form defined by equation (8) for the μ-law and equation (12) for the A-law. These equations characterize the precise quantization structures.

The centering parameter c is 0 or 0.5, and when it is 0 the transfer

characteristic at the origin is "mid-tread" as shown in Fig. 3a. When $c = 0.5$, it is "mid-riser" as shown in Fig. 3b.

To show the effect of the parameter a, a few examples of the transfer curves are shown in Fig. 4 for $N = 4$ and mid-tread operation ($c = 0$). As mentioned before, when $a = 0.5$, the transfer curves for $b = +0.5$ and -0.5 are identical except for a shift in the numbering of $X(L, V)$.

2.6 Step Sizes

The step sizes of the output levels and the decision levels are given, except for the segment edges, by

$$\left. \begin{array}{l} \Delta_0(L) = y_0(L, V + 1) - y_0(L, V) \\ \Delta_d(L) = y_d(L, V + 1) - y_d(L, V) \end{array} \right\} V \neq N - 1 \qquad (18)$$

and the corresponding step sizes at the segment edges are given by

$$\left. \begin{array}{l} \Delta_0'(L) = y_0(L + 1, 0) - y_0(L, N - 1), \\ \Delta_d'(L) = y_d(L + 1, 0) - y_d(L, N - 1). \end{array} \right\} \qquad (19)$$

It is easily shown that for non-edge points

$$\Delta_0(L) = \Delta_d(L) = 2^L \qquad (20)$$

for the μ-law. However, for the edge points the step size is given by

$$\left. \begin{array}{l} \Delta_0'(L) = f(L + 1, a) - f(L, N - 1 + a) = \Delta_0(L) \cdot (1 + a), \\ \Delta_d'(L) = f(L + 1, a + b) - f(L, N - 1 + a + b) \\ \qquad = \Delta_d(L) \cdot (1 + a + b). \end{array} \right\} \qquad (21)$$

Step sizes at the segment edges are depicted in Fig. 5 with respect to a.
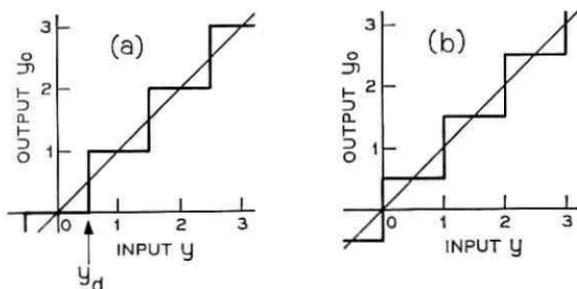

Fig. 3—Effect of centering parameter c, (a) mid-tread ($c = 0$), and (b) mid-riser ($c = 0.5$).
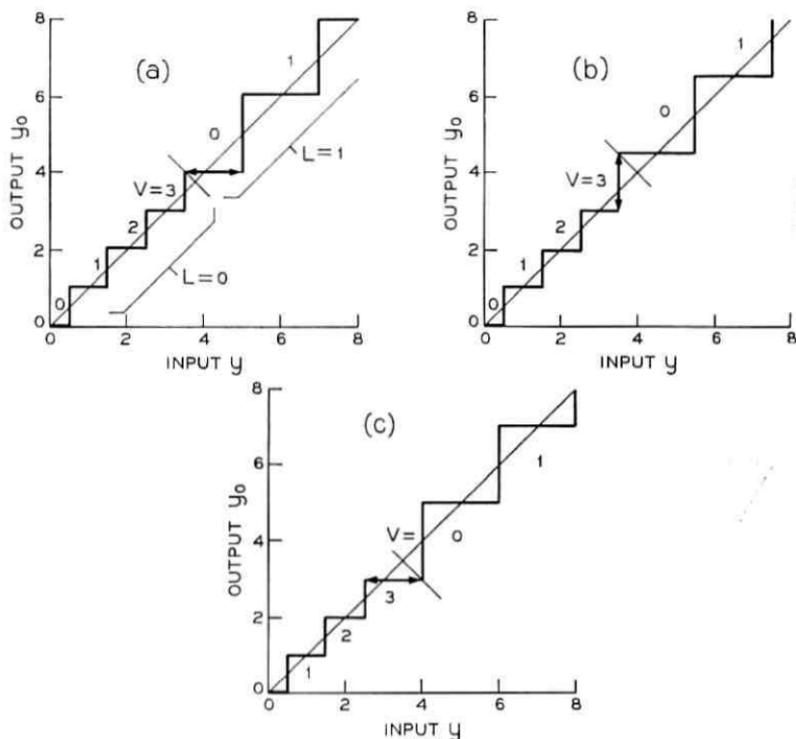
Fig. 4—Transfer staircases with $N = 4$ and mid-tread for (a) RLA ($a = 0$, $b = 0.5$), (b) DLA ($a = 0.5$, $b = 0.5$), and (c) RLA ($a = 1$, $b = -0.5$).

For the A-law, equation (20) is modified to

$$\Delta_0(L) = \Delta_d(L) = \begin{cases} 1, & L = 0, \\ 2^{L-1}, & L \neq 0, \end{cases}$$

and relation (21) applies for the A-law on the understanding that the "segment edges" exclude the colinear segment edge between 0th and 1st.

2.7 Effective Coder Input

Now, we will define an effective coder input $\tilde{y}(L, V)$ such that $\tilde{y}(L, V)$ is the statistical average of the input signal between two adjacent decision levels. If we assume that the input is uniformly distributed between two adjacent decision levels, then $\tilde{y}(L, V)$ is equal to their average as shown below. For $b = +0.5$ (upper bound case)

$$\tilde{y}(L, V) = \begin{cases} \frac{1}{2}[y_d(L, V - 1) + y_d(L, V)], & V \neq 0, \\ \frac{1}{2}[y_d(L - 1, N - 1) + y_d(L, 0)], & V = 0 \quad \text{(edge)}. \end{cases}$$
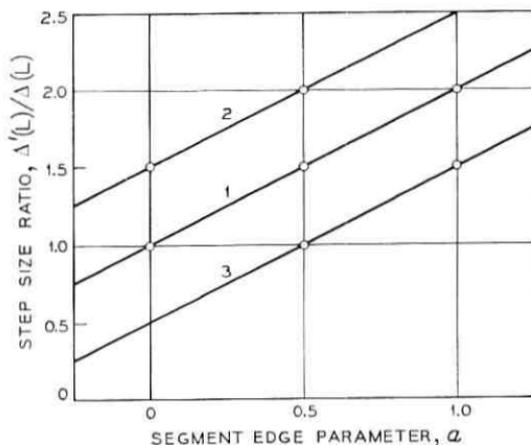
Fig. 5—Step size ratio at the segment edges relative to the lower segment step size: Curve 1—output level; Curve 2—decision level ($b = 0.5$); and Curve 3—decision level ($b = -0.5$).

For $b = -0.5$ (lower bound case)

$$\tilde{y}(L, V) = \begin{cases} \frac{1}{2}[y_d(L, V) + y_d(L, V + 1)], & V \neq N - 1, \\ \frac{1}{2}[y_d(L, N - 1) + y_d(L + 1, 0)], & V = N - 1 \quad \text{(edge)}. \end{cases}$$

Except for the edge points, we have

$$\tilde{y}(L, V) = y_0(L, V) \tag{22}$$

and this is a direct consequence of our definition of y_d in equation (14). At the segment edges (L_e, V_e), it can be shown that

$$\tilde{y}(L_e, V_e) = y_0(L_e, V_e) - \Delta_0(L) \cdot b \cdot (a - 0.5) \tag{23}$$

where edge point means

$$(L_e, V_e) = \begin{cases} (L + 1, 0) & \text{when} \quad b = +0.5, \\ (L, N - 1) & \text{when} \quad b = -0.5. \end{cases} \tag{24}$$

Therefore, except for the segment edges the transfer gain from the effective coder input to the decoder output is unity. At these segment edges, if we define the tracking error ϵ, which is the deviation from unity of the transfer gain from the effective coder input $\tilde{y}$ to the decoder output y_0, then ϵ is

$$\epsilon \equiv 1 - \tilde{y}/y_0$$

$$= \begin{cases} \left| \dfrac{\Delta_0(L)}{y_0} \cdot b \cdot (a - 0.5) \right|, & \text{at segment edges} \\[2mm] 0, & \text{elsewhere.} \end{cases} \tag{25}$$

The ratio $\epsilon/(\Delta_0/y_0)$ is the gain error relative to the step size, and is depicted in Fig. 6, where it is seen that the tracking error is zero over the entire input range if and only if $a = 0.5$.

2.8 Relation Between Parameters

In Section 2.4 it was noted that for a given output level the locally optimal (minimum mean squared error) decision level is half way between adjacent output levels.[10] It can be shown that this optimality condition is satisfied for any X including segment edges, if and only if $a + b = 0.5$, which means $a = 0$ for $b = 0.5$, or $a = 1$ for $b = -0.5$. Since optimal placement is made from the given output level, this is called "Reconstruction (or Output) Level Assignment" (RLA).[8]

On the other hand, for given decision levels, assuming the uniform distribution of the input in each quantization step, the locally optimal (minimum mean squared error) placement of the output levels lies halfway between adjacent decision levels.[10] It can also be proven that this condition is satisfied for any point X including segment edges, if and only if $a = 0.5$, irrespective of $b = \pm 0.5$. Since the optimal approach is based on the given decision levels, this is called "Decision Level Assignment" (DLA).[8] Proof of the above statements is given in Appendix B.
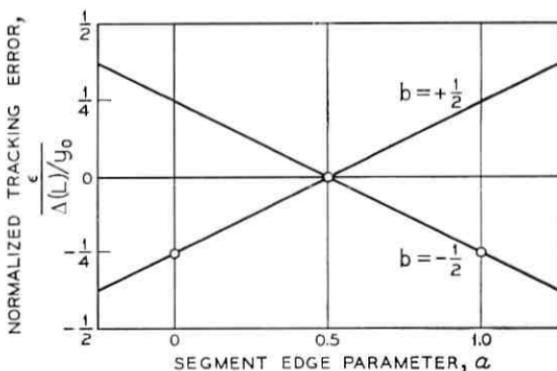

Fig. 6—Tracking error at the segment edges.

As shown previously, when $a = 0.5$ (DLA) the effective tracking error is zero for all inputs, and the upper and the lower bound approaches give identical coding. Furthermore, as will be seen later, the digital compression algorithm and circuits are simpler in DLA than in RLA. The D2 channel bank has been designed according to DLA. Hereafter, our attention is mainly focused on DLA, although the analysis is easily generalized to include other cases.

When $M = 8$, the segment μ-law is generally referred to as the "15-segment μ-law" including negative segments. In contrast, the segment A-law with $M = 8$ is called the "13-segment A-law." In Table I the output levels and the step sizes are listed for the 8-bit (including sign bit) 15-segment μ-law and 13-segment A-law. The maximum and minimum output levels are shown in Table II for ready reference.

Quantizing noise, as calculated for the Laplacian (negative exponential) distributed speech input y_{in}, yields the output signal to quantizing noise ratio shown in Fig. 7. Signal and noise are defined in a band of width equal to half the sampling rate. The average transfer gain $\langle G \rangle$ is chosen so as to minimize the quantizing noise, and is given by $E[y_{in} \cdot y_0]/E[y_{in}^2]$. The idle channel noise, calculated for a zero-mean gaussian input noise, is simply $\mathrm{ICN} = E[y_0^2]$ and is shown in Fig. 7. The centering parameter c is important in the small signal range as is

TABLE I—OUTPUT LEVEL STEP SIZES FOR $m = 3$ AND $n = 4$

segment number L	μ-law non-edge $\Delta_0(L)$	edge points $\Delta_0'(L)$ $a = 0$	$a = 0.5$	$a = 1$	A-law non-edge $\Delta_0(L)$	edge points $\Delta_0'(L)$ $a = 0$	$a = 0.5$	$a = 1$
0	1				1			
		1	1.5	2		1	1	1
1	2				1			
		2	3	4		1	1.5	2
2	4				2			
		4	6	8		2	3	4
3	8				4			
		8	12	16		4	6	8
4	16				8			
		16	24	32		8	12	16
5	32				16			
		32	48	64		16	24	32
6	64				32			
		64	96	128		32	48	64
7	128				64			

TABLE II—MAXIMUM AND MINIMUM OF OUTPUT LEVELS FOR $m = 3$
AND $n = 4$

		mid-tread ($c = 0$)			mid-riser ($c = 0.5$)		
		$a = 0$	$a = 0.5$	$a = 1$	$a = 0$	$a = 0.5$	$a = 1$
μ-law	$y_0(0, 0)$	0	0	0	0.5	0.5	0.5
	$y_0(7, 15)$	3952	4015.5	4079	3952.5	4016	4079.5
A-law	$y_0(0, 0)$	0	0	0	0.5	0.5	0.5
	$y_0(7, 15)$	1984	2015.5	2047	1984.5	2016	2047.5
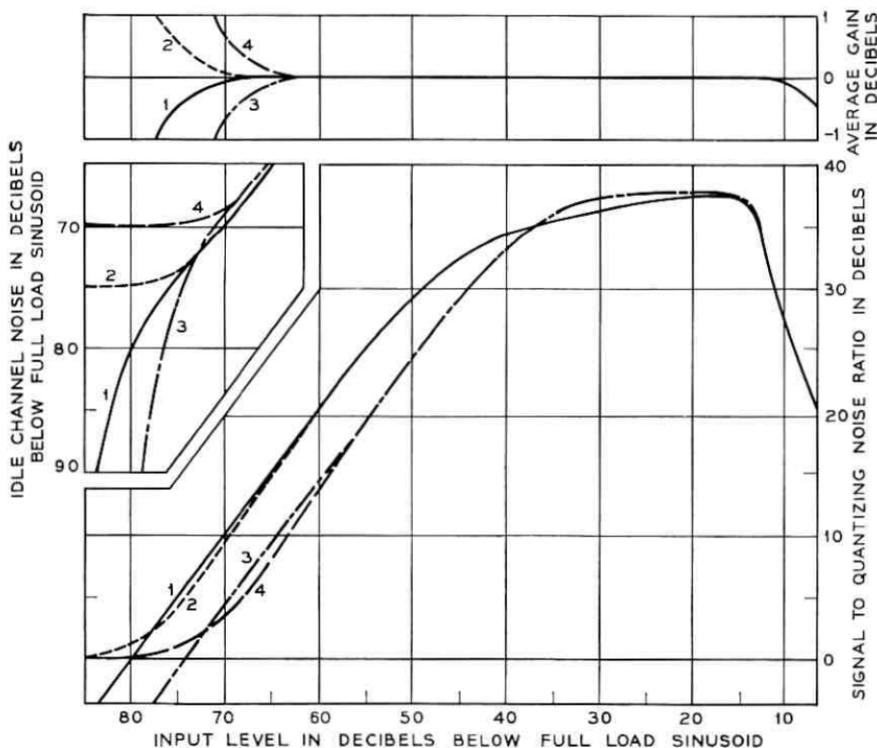
Fig. 7—Average gain, signal-to-noise ratio, and idle channel noise for DLA with $m = 3$ and $n = 4$: Curve 1—μ-law, mid-tread ($c = 0$); Curve 2—μ-law, mid-riser ($c = 0.5$); Curve 3—A-law, mid-tread ($c = 0$); and Curve 4—mid-riser ($c = 0.5$).

well known, and it affects center clipping and idle channel noise. The effect of a on the signal quality is very small. The signal-to-noise ratio relative to the S/N for DLA case (curve 1 in Fig. 7) and the average gain are shown in Figs. 8a and b as curves 1 and 2 for the two RLA cases. The small difference (± 0.14 dB) in S/N in the lower signal range is due to the difference in overload levels. However, if there is a mistracking such that the signal is encoded with $a_1 = 0.5$ and decoded with $a_2 = 1$, the mismatch results in distortion and net loss variation. The penalty in the signal-to-noise ratio and the change in average gain are shown as curve 3 in Figs. 8a and b. Although the penalties are very small, the net loss variation barely meets toll quality requirement, and accumulated impairment in multi-link digital conversions would be intolerable. Therefore, the mixed use of DLA and RLA is to be avoided.

2.9 General Formulation

To summarize this section we give a general expression which is conveniently used hereafter in the synthesis of codecs and digital compandors. From equations (7), (8), (12) and (14) the general expression can be written as

$$y(L, V) = \Delta(L) \cdot (V + P) - Q \tag{26}$$

where for the segment μ-laws
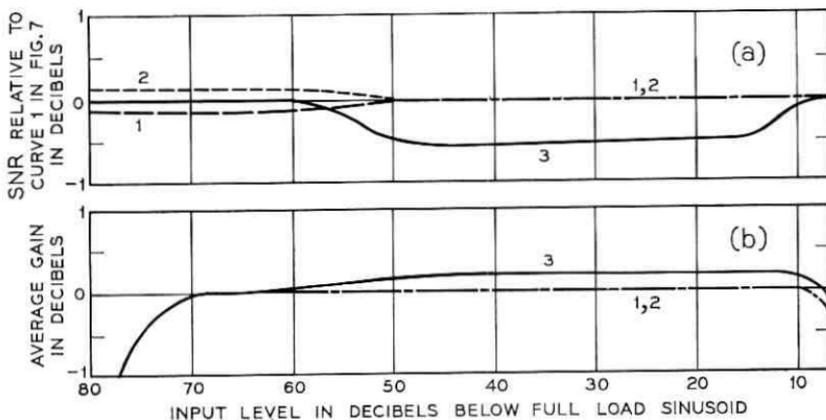
Fig. 8—Effects of parameters on S/N and average gain for μ-law with $m = 3$ and $n = 4$: Curve 1—RLA ($a = 0$, $b = 0.5$); Curve 2—RLA ($a = 1$, $b = -0.5$); and Curve 3—mistracking case coded by DLA and decoded by RLA ($a = 1$).

$$\begin{aligned}
\Delta(L) &= \Delta_0(L) = \Delta_d(L) = 2^L, \\
P &= N + a + b, \\
Q &= N + a - c.
\end{aligned} \right\} \tag{27}$$

Equation (26) can represent either the output level by equating b to 0 or the decision level with $b = \pm 0.5$. For the segment A-laws, expression (27) is replaced by

$$\begin{aligned}
\Delta(L) &= L - \eta, \\
P &= N \cdot \eta + a + b, \\
Q &= a - c,
\end{aligned} \right\} \tag{28}$$

where

$$\eta = \begin{cases} 0, & L = 0, \\ 1, & L \neq 0. \end{cases}$$

III. SYNTHESIS OF CODECS

3.1 Synthesis of Unipolar Decoder

In this section, we show that the representation of the output level and the decision level by equation (26) leads directly to the implementation of the decoder. In the following, attention will be focused mainly on the segment μ-law. Extension to the A-law is readily accomplished by similar procedures.

Equation (26) given by

$$y(L, V) = \Delta(L) \cdot (V + P) - Q$$

is clearly interpreted as adding bias P to V, amplifying it by $\Delta(L)$, and subtracting bias Q from this result, thus obtaining the decoder output $y(L, V)$.

Let us represent the nonlinear code X by an $(m + n)$ bit binary sequence (except for the sign bit e_0), $e_i \in \{0, 1\}$,

$$\begin{aligned}
X &= X\{e_1, e_2, \cdots, e_m, e_{m+1}, \cdots, e_{m+n}\} \\
&= \sum_{i=1}^{m+n} 2^{m+n-i} \cdot e_i.
\end{aligned} \tag{29}$$

Then,

$$
\left.\begin{aligned}
L &= L\{e_1, e_2, \cdots, e_m\} = \sum_{i=1}^{m} 2^{m-i} e_i, \\
V &= V\{e_{m+1}, \cdots, e_{m+n}\} = \sum_{i=1}^{n} 2^{n-i} e_{m+i}.
\end{aligned}\right\}
\tag{30}
$$

From equation (27), $\Delta(L)$ can be written as

$$
\Delta(L) = 2^L = \prod_{i=1}^{m} \{2^{(2^{m-i})}\}^{e_i}
\tag{31}
$$

and is realized by a cascade of switched amplifiers (or attenuators) whose gain is switched between $2^{(2^{m-i})}$ and 1 according to the digit $e_i = 1$ or 0, respectively. Likewise, V in equation (30) is obtained directly from a uniform decoder of any of the known forms[11] operated by $\{e_{m+1}, e_{m+2}, \cdots, e_{m+n}\}$.

An example of the μ-law decoder is shown in Fig. 9a. Biases P and Q are given by $(N + a + b)$ and $(N + a - c)$, respectively. The decoder output is either the decoder output level $y_0(L, V)$ by replacing $b = 0$ in P, or the coder decision level $y_d(L, V)$ by choosing $b = \pm 0.5$.

The decoder circuit for the segment A-law is synthesized in the same way using equation (28), and is shown in Fig. 9b.

3.2 Consideration of the Sign Bit

In the preceding, the sign bit e_0 has been excluded from consideration. In practical cases it is desirable to use a bipolar decoder to cover the whole positive and negative signal ranges. We will give alternative representations of the decoder output in terms of the signed input codes. First, assume that the output level is symmetric with respect to the 0-level line.* Suppose that the sign bit e_0 is equal to 0 for the positive signal region, and 1 for negative region. If we define

$$
\alpha_0 = 1 - 2e_0 \in \{+1, -1\},
\tag{32}
$$

then from equation (26) the bipolar decoder output is given by

$$
y(L, V) = \alpha_0\{\Delta(L)\cdot(V + P) - Q\}
\tag{33}
$$

$$
= \Delta(L)\{\alpha_0 V + \alpha_0 P\} - \alpha_0 Q.
\tag{34}
$$

If we look at the binary representation of $\alpha_0 V = \{e_0, e_{m+1}, e_{m+2}, \cdots, e_{m+n}\}$ as in equation (30), the so-called "sign and magnitude"

---

* This implies that in the mid-tread case the "0" output level is shared between the positive "0" and the negative "0".
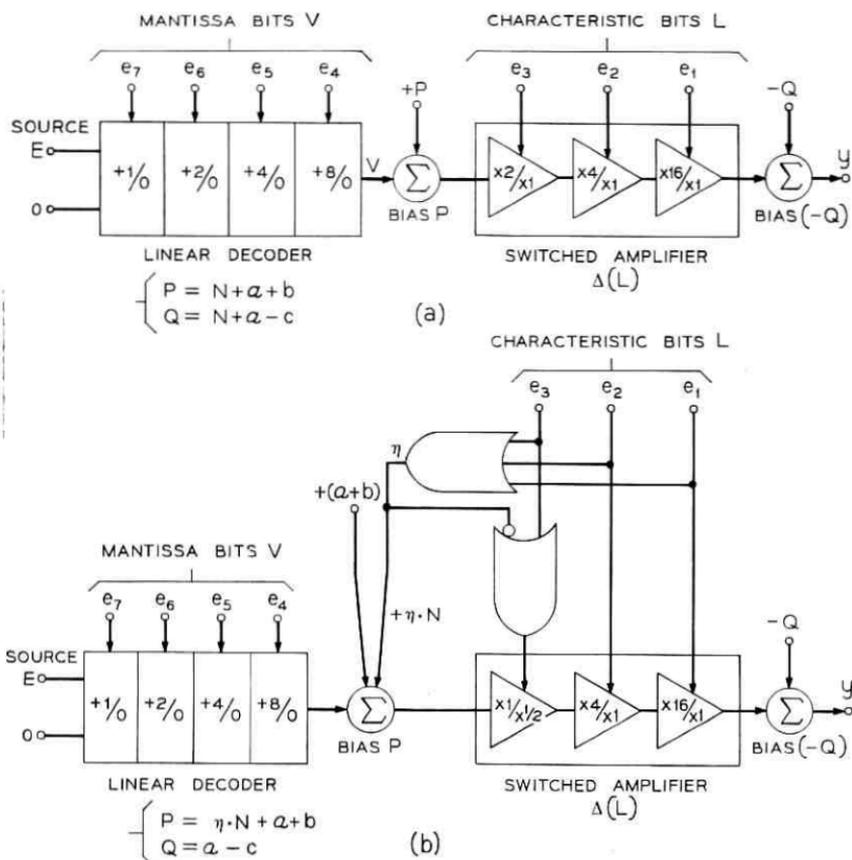
Fig. 9—Synthesis of segment law decoders ($m = 3$, $n = 4$): (a) 15-segment μ-law and (b) 13-segment A-law.

or "folded binary" representation results. Another representation often used in computer arithmetic is the "1's complement" code.[12] If we denote the 1's complement code by

$$V' = V'\{e'_{m+1}, e'_{m+2}, \cdots, e'_{m+n}\} = \sum_{i=1}^{n} 2^{n-i} e'_{m+i},$$

then V' is related to positive V by

$$\alpha_0 V = V' - e_0(N - 1) \tag{35}$$

and also

$$e'_{m+1} = e_{m+i} \oplus e_0 \bmod 2. \tag{36}$$

Therefore, equation (34) can be expressed in terms of V' and

$$y(L, V) = \Delta(L)\{V' - e_0(N - 1) + \alpha_0 P\} - \alpha_0 Q. \tag{37}$$

Note that the modification in the representation applies only to the mantissa bits V, and the characteristics bits L should remain unchanged.

3.3 Synthesis of Bipolar Decoder

We can synthesize the bipolar decoder circuits according to the two alternative representations (33) and (37).*

(*i*) Sign-and-magnitude (folded binary)—The "sign and magnitude" decoder according to equation (35) is obtained by inverting the decoder output of the unipolar decoder, or by inverting all the supply voltages to the unipolar decoder as shown in Fig. 10a. The former requires a precise analog inverter. In the latter circuit, the source switch must be in serial with the decoder switch, and stringent requirements arise due to the wide dynamic range of these analog switches.

(*ii*) 1's complement—Suppose that we use a bipolar source $(+E, -E)$ instead of unipolar source. $(+E, 0)$ By converting the unipolar code to a bipolar code, we have

$$\alpha'_{m+i} = 2e'_{m+i} - 1 \in \{-1, +1\}, \tag{38}$$

and also using equation (32), equation (37) becomes

$$y(L, V) = \Delta(L)\left[\sum_{i=1}^{n} 2^{n-i-1} \cdot \alpha'_{m+i} + \alpha_0\left(P + \frac{N-1}{2}\right)\right] - \alpha_0 Q \tag{39}$$

where $\Delta(L)$, P and Q are given by equation (27). This leads to the decoder circuit shown in Fig. 10b. The weighting currents of the decoder are given by the coefficients of the α's in equation (39). The weighted current $[P + (N - 1)/2]$ that is switched by α_0 is about six times greater than the most significant digit weight $(= 2^{n-2})$ of the linear decoder. Therefore, the problem of the serial switch in Fig. 10a is replaced by a threefold increase in the accuracy requirements of the weighting network here. This appears to be a worthwhile trade in the practical design of codecs.

3.4 Interpretation of Coding Parameter b

As discussed in Section 2.4, the coder decision level y_d is related to the decoder output level y_0 by

$$y_d(L, V) = y_0(L, V + b)$$

---

* We have derived another representation for the "2's complement code" but it is omitted because of its increased complexity over the other two representations.
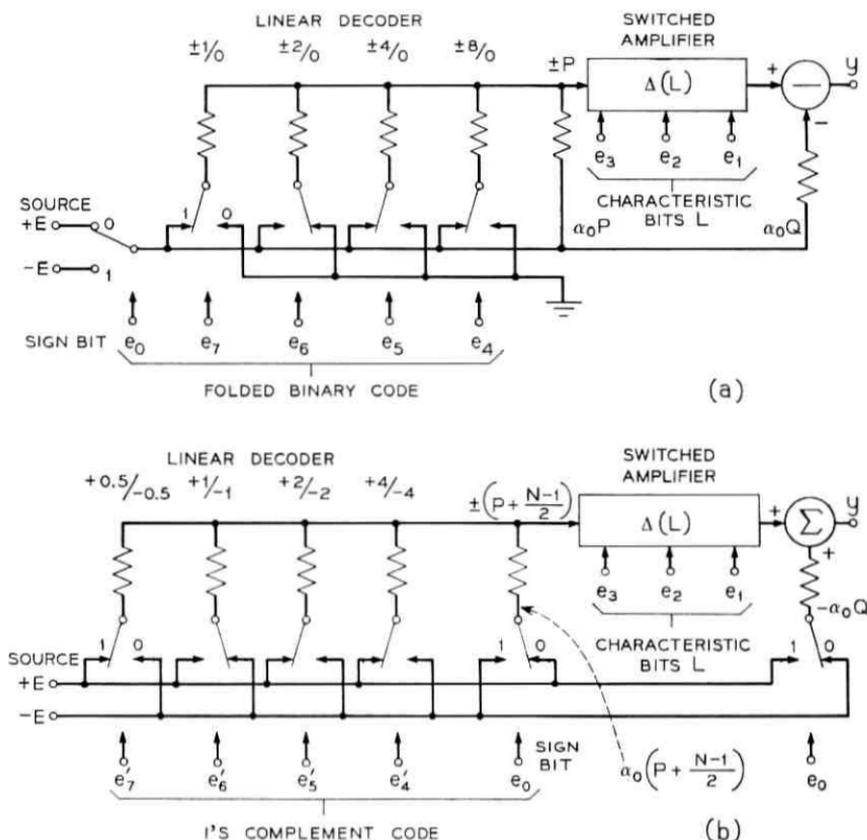
Fig. 10—Synthesis of bipolar decoders operated by (a) folded binary code and (b) 1's complement code.

for $b = \pm0.5$. When $b = +0.5$ the decision level is an upper-bound for the corresponding output level, and when $b = -0.5$ it is a lower bound, in the sense given by inequalities (15) and (16).

In the following we show how the lower or upper bound approach is related through the parameter b to the logic used in implementing a sequential comparison coder.

Referring to Fig. 11, for the lower bound approach ($b = -0.5$) (from the "all reset" state "000"), the first digit trial "1" is made; and the decision level corresponding to "100" is compared with the input. As a result, if the error ϵ defined as

$$\epsilon = \text{input} - \text{local decoded signal}$$
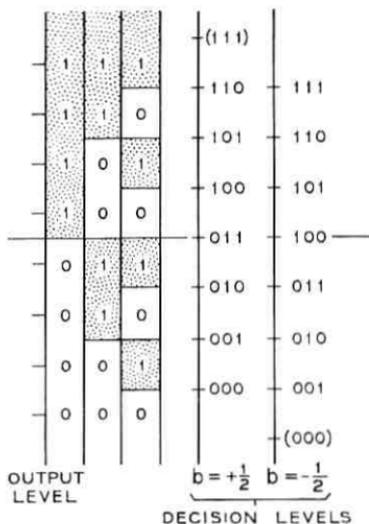
Fig. 11—Upper bound and lower bound decision levels.

is negative the trial "1" is "reset" to "0", and remains unchanged if ϵ is positive. This procedure is repeated digit by digit from the MSD to LSD, and the coding is performed sequentially so as to minimize ϵ.

On the other hand, for the upper bound approach $(b = +0.5)$, the "all reset" is made to "1", the "trial" is made to "0", and the "reset" is made to "0" if $\epsilon > 0$. Therefore, if we take a package of "all reset-trial-conditional reset", it is "0-1-0" for the upper bound approach, and "1-0-1" for the lower bound approach. From this package the circuit design of sequential coding logic and registers follows immediately.

3.5 Synthesis of Sequential Comparison Coder

From the result of Section 3.4, and using the unipolar decoder described in Section 3.1, the synthesis of a unipolar coder is straightforward. Here, we treat only the bipolar coder. It has been shown in Section 3.3 that the bipolar decoder in 1's complement form has an advantage in circuit design. If the upper bound approach is taken $(b = +0.5)$, the "all reset-trial-conditional reset" package is "0-1-0" for a positive signal. For the negative half if $b = +0.5$ the package becomes "1-0-1" because of the 1's complement (inverted) representation. However, the logic package can be uniformly "0-1-0" for the entire range, if we take $b = +0.5$ for positive inputs and $b = -0.5$

for the negative region. Figure 12 shows the schematic of a bipolar sequential comparison coder based on the 1's complement code. The local decoder, which must have the nonsymmetric feature in b, is obtained from the 1's complement decoder of Fig. 10b with $b = 0$. In addition, a constant bias $b = +0.5$ is added, externally. Then, the register and logic are operated on the basis of the 1's complement code, and the coding is performed with the logic package of "0-1-0." The segment characteristic bit should always be a "folded binary code", which is simply the inverse of the 1's complement code. Of course, there are many other alternatives for building a coder, and the synthesis of any of those immediately follows from our formulation.

IV. SYNTHESIS OF DIGITAL COMPANDORS

In Section III we considered the conversion from analogue signal to digital compressed code or the converse. Here, we consider the conversion between the compressed code and the linear code. The conversion from the compressed code to the linear code is called "digital expansion" and the converse is called "digital compression." In Sections 4.1 and 4.2 the digital expandor synthesis is treated, and after that the digital compressor is studied.

4.1 Digital Expansion Algorithm

Suppose that a nonlinear code $X(L, V)$ with bit length $(m + n)$ is digitally linearized to Y. The linearized code Y should correspond
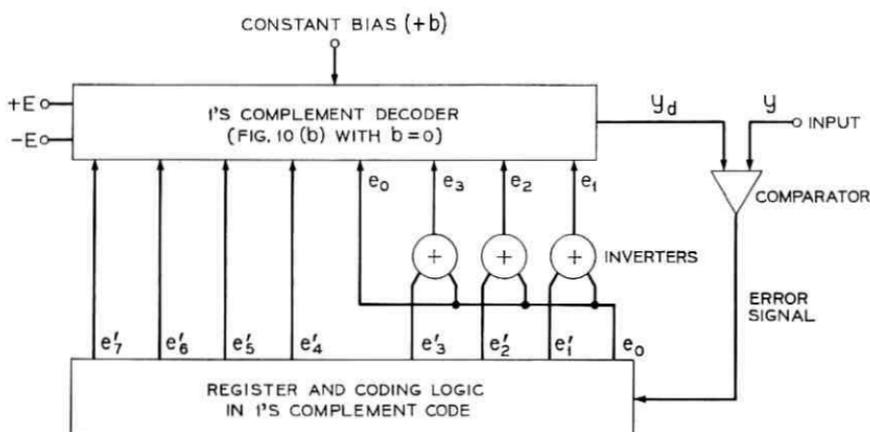

Fig. 12—Bipolar sequential comparison coder with 1's complement code logic.

to the decoder output level

$$Y = y_0(L, V) = f(L, V + a) - a + c. \tag{40}$$

Suppose that the companding law is symmetric for positive and negative signals. We generally exclude the sign bit from consideration. In the case of RLA and mid-tread ($c = 0$), Y is an integer and is expressed by a linear code with length ($2^m + n$) bits. In the cases of DLA ($a = 0.5$) or the mid-riser ($c = 0.5$), Y is expressed by ($2^m + n + 1$) bits including one fractional digit. For example the D2 code ($a = 0.5$, $c = 0$, $m = 3$, and $n = 4$) is basically* expressed in terms of 13-bit binary code without the sign bit.

The expanded signal Y should correspond to the output level. Therefore, from equations (26) and (27) with $b = 0$, Y is given by

$$Y = \triangle(L) \cdot (V + P) - Q \tag{41}$$

where $\triangle(L) = 2^L$ is a shifting operator, by which the binary sequence ($V + P$) is shifted to the left by L bits. Therefore, the algorithm for code conversion from $X(L, V)$ to Y is

Algorithm 1 (expansion): For given $X(L, V)$, Y is obtained such that

   (*i*) Add P to V,
   (*ii*) Shift ($V + P$) by L bits to the left, and
   (*iii*) Subtract Q,

where $P = N + a$ and $Q = N + a - c$. All variables and constants are assumed to be in binary representation.

4.2 *Synthesis of Digital Expandor*

The above algorithm leads directly to a parallel expandor circuit as shown in Fig. 13. The input L is written into a binary counter BC, and the input V is written into a shift register SR through a parallel adder with addend ($N + a$). Then, the binary counter BC counts down to "000", and stops. This interval is equal to L. The SR is shifted by the same clock pulses, and L shifts to the left is performed. Then, the entry of SR is read out through a parallel subtractor with subtrahend ($N + a - c$), thus yielding the parallel expanded output Y.

In practice, a digital expandor is generally followed by a digital processor, in which serial arithmetic logic is most conveniently used when the signal sequence begins with the LSD (Least Significant Digit).[12,13] In the sequel, we concentrate on the serial implementation.

* Except for bit robbing for signalling and framing.[7]
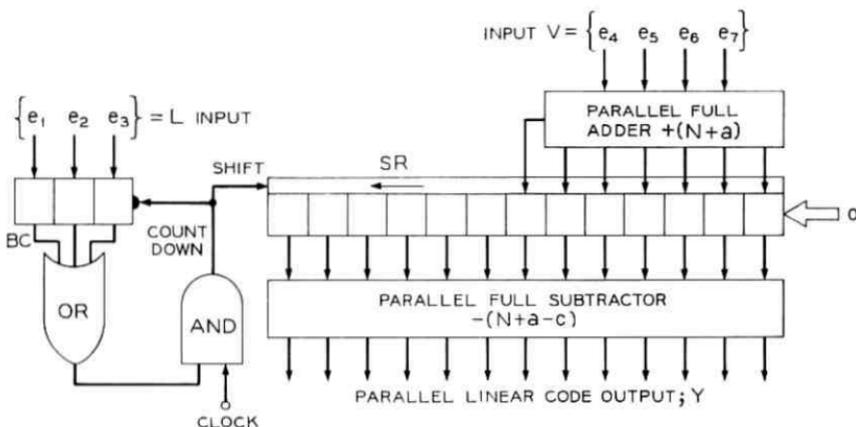
Fig. 13—Parallel digital expandor.

If we replace ($\times 2$) by an operator z, which at the same time represents the delay operator by one clock interval, then Y is given by

$$Y(z) = z^L \cdot \{V(z) + P(z)\} - Q(z) \tag{42}$$

where $V(z)$, $P(z)$ and $Q(z)$ are binary sequences beginning from the LSD. Therefore, a general form of digital expandor can be described schematically as shown in Fig. 14.

Let us consider more specifically the μ-law expandor design. Using the z-transform notation, equation (42) can be written as

$$Y(z) = z^L \{V(z) + z^n + a(z)\} - \{z^n + a(z) - c(z)\}$$
$$= z^L \{V(z) + z^n\} + a(z) \cdot (z^L - 1) + c(z) - z^n \tag{43}$$

where ($\pm$) implies addition/subtraction with carry, if necessary. Assuming DLA [$a(z) = z^{-1}$], mid-tread ($c = 0$), $m = 3$ and $n = 4$, we have

$$zY(z) = z^L[zV(z) + zP(z)] - zQ(z)$$
$$= z^L(z^0 + e_7z + e_6z^2 + e_5z^3 + e_4z^4 + z^5) - z^0 - z^5.$$

This can further be written in the form

$$zY(z) = z^L(e_7z + e_6z^2 + e_5z^3 + e_4z^4 + z^5) + (z^L - 1) - z^5. \tag{44}$$

From this a serial expandor circuit follows as shown in Fig. 15. The inputs L and V are written into a binary counter BC and a shift register SR, respectively. Also "1" is written in the last stage of SR, and the
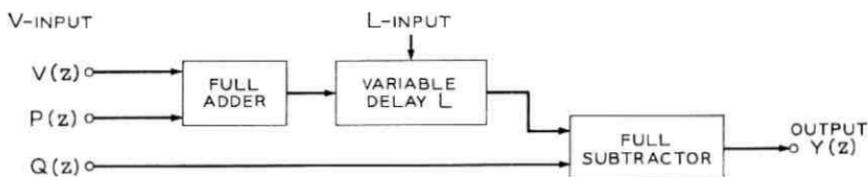
Fig. 14—Basic serial digital expandor.

Fig. 15—Fifteen-segment μ-law expandor with DLA ($a = 0.5$), mid-tread ($c = 0$), $m = 3$, and $n = 4$.

initial entry of SR, which is equal to $(N + V)$, corresponds to the first term in equation (44) without delay z^L. The binary counter BC counts down until it becomes "000" and stops, thus giving a gating signal T_L with length L. Instead of having a separate variable delay as in Fig. 14, SR is designed to wait for L intervals before shifting to the right. This gives a delay corresponding to z^L, and the output of the shift register U, which is equal to the first term in equation (44), is depicted as shown in the timing chart of Fig. 15. Next, notice that the second term $(z^L - 1)$ in equation (44) is a sequence of successive 1's with length L. It is exactly the same as the gating signal T_L. Since T_L does not overlap with the first term U, it can be simply added through an OR circuit to the output of SR. The last term z^5, which is nothing but a timing pulse at the 5th time slot, is subtracted through a half subtractor. Thus, the output LSD-first sequence $z \cdot Y(z)$ is obtained immediately after the write-in of the input.

4.3 Digital Compression Algorithm

In this section, we consider the conversion from a linear code Y to a nonlinear code X. Suppose that a linear code Y is a positive real number representing a signal amplitude normalized by the minimum quantizing step of the compressed code. This code Y is not necessarily an integer, and in general it has q fractional digits in binary representation. This corresponds to the common practice to have more digits in a digital processor than in the input/output digits to reduce the effect of rounding or truncation in arithmetic units (such as digital filters).

As described in Section 2.4 nonlinear coding from an analog signal y to a compressed code $X(L, V)$ is performed by comparing the input with the decision levels. Similarly, in equations (15) and (16), on replacing the analog input y by a discrete linear code Y, we have the following basic digital compression algorithm.

Algorithm 2 (compression: basic):

(*i*) Upper bound approach—Choose $\hat{X}$ such that

$$y_d(\hat{X} - 1 \mid b = +0.5) \leqq Y < y_d(\hat{X} \mid b = +0.5). \tag{45}$$

(*ii*) Lower bound approach—Choose $\hat{X}$ such that

$$y_d(\hat{X} \mid b = -0.5) \leqq Y < y_d(\hat{X} + 1 \mid b = -0.5), \tag{46}$$

where y_d is exactly the decision level defined by equation (14).

It has been shown that both approaches yield identical outputs

if and only if $a = 0.5$ (DLA). Also note that there is an equality only in the lower boundaries. There is the alternative algorithm in which the equality in equations (45) or (46) is given to the upper boundaries. As compared with the coding from analog input y described in Section 2.4, Algorithm 2 is equivalent to analog coding with input $y = Y - 2^{-a-1}$, and the alternative upper boundary algorithm is equivalent to analog coding with $y = Y + 2^{-a-1}$. Both algorithms are asymptotically equivalent to analog coding for large q. We use the former since it is more directly related to circuit implementation.

Though this algorithm is basic, it is not convenient for synthesizing the circuit. We will divide the above algorithm into the $\hat{L}$-search process and the $\hat{V}$-determination processes. Let us introduce a new variable W such that

$$W(\hat{L}, Y) = \{\Delta(L)\}^{-1} \cdot (Y + Q) \tag{47}$$

where $Q = N + a - c$. Then, for the DLA the compression algorithm is given by

Algorithm 3 (compression: DLA): In the case of DLA ($a = 0.5$), the conversion is performed as follows: choose $\hat{L}$ such that*

$$W(\hat{L}, Y) \in [N, 2N) \tag{48}$$

determine $\hat{V}$ such that

$$\hat{V} = \mathfrak{I}[W(\hat{L}, Y)] - N. \tag{49}$$

Here, $\mathfrak{I}[\cdot]$ is the truncation function and defined as

$\mathfrak{I}[x] \equiv$ truncation of x

$= $ integer I such that $I \leqq x < I + 1.$ \tag{50}

We also define the rounding function

$\mathfrak{R}[x] \equiv$ rounding of x

$= \mathfrak{I}[x + 0.5].$ \tag{51}

The derivation of this algorithm is given in Appendix C. If we take the inverse of

$$Y = \Delta(L)(V + P) - Q$$

with respect to V, then

$$V = \{\Delta(L)\}^{-1} \cdot (Y + Q) - P.$$

---

* Semi-closed interval $[N, 2N)$ means $N \leq W < 2N$.

The first term is exactly $W(L, Y)$ and $P = N + a = N + 0.5$ in DLA, and

$$V = W(L, Y) - N - 0.5.$$

By taking the rounding of V and using equation (51), we have

$$\mathfrak{R}[V] = \mathfrak{R}[W(L, Y) - 0.5] - N$$
$$= \mathfrak{I}[W(L, Y)] - N.$$

This is the same as $\hat{V}$ in the algorithm. Therefore, the physical implication of the algorithm is that $\hat{V}$ is obtained by *simply rounding the transformed variable V*.

For the Reconstruction Level Assignment ($a = 0$ for $b = +0.5$, or $a = 1$ for $b = -0.5$) the algorithm is more complicated,[14] and is *Algorithm 4* (compression: RLA): For the RLA, conversion is accomplished in the following way: choose $\hat{L}_0$ such that

$$W(\hat{L}_0, Y) \in [N, 2N), \tag{52}$$

determine $\hat{V}_0$ such that

$$\hat{V}_0 = \mathfrak{I}[W(\hat{L}_0, Y)] - N \tag{53}$$

and correct by

$$X(\hat{L}, \hat{V}) = X_0(\hat{L}_0, Y_0) + \theta \tag{54}$$

where the corrector θ is given by

$$\theta = W_{of} - a \tag{55}$$

for W_{of}, the highest fractional binary digit of $W(\hat{L}_0, Y)$.

The proof is complicated, and so is the circuit, and both are omitted because of less interest in RLA.

4.4 *Synthesis of Digital Compressor*

We will synthesize a digital compressor circuit according to the above algorithms. For simplicity DLA is assumed. Also assume that an input Y is a binary sequence beginning with LSD. In Algorithm 2 the L-search process by equation (48) is accomplished, first by forming a binary sequence $(Y + Q)$, and then shifting bit by bit to the right until equation (48) is satisfied. The criterion (48) implies that the nth digit of W is "1", and the digits of W higher than the nth digit are all "0".

Based on this algorithm an example of the μ-law digital compressor

for DLA, $m = 3$ and $n = 4$ is synthesized as shown in Fig. 16. The input $Y(z)$, consisting of 12 integer digits and q fractional digits, is fed into a serial adder and $Q = (N + a - c)$ is added. Then, the output of the adder is written into the first shift register SR_1 and then to the second shift register SR_2. Let us denote the time instant by t_0 when
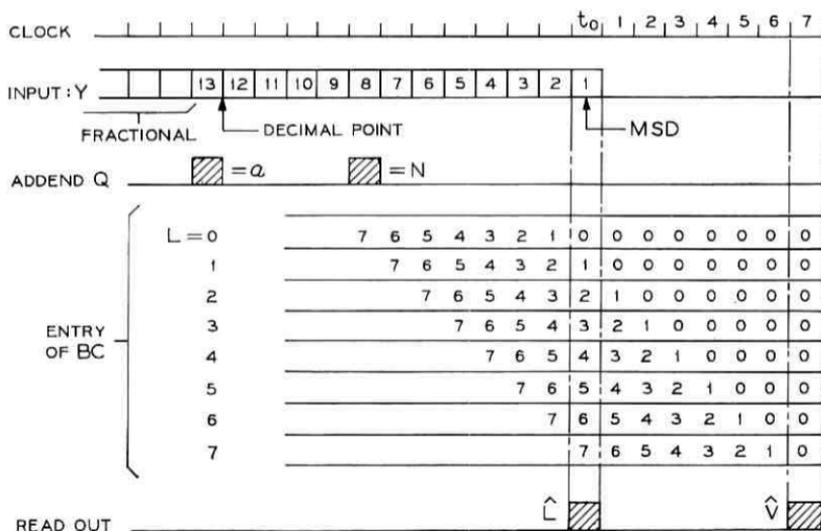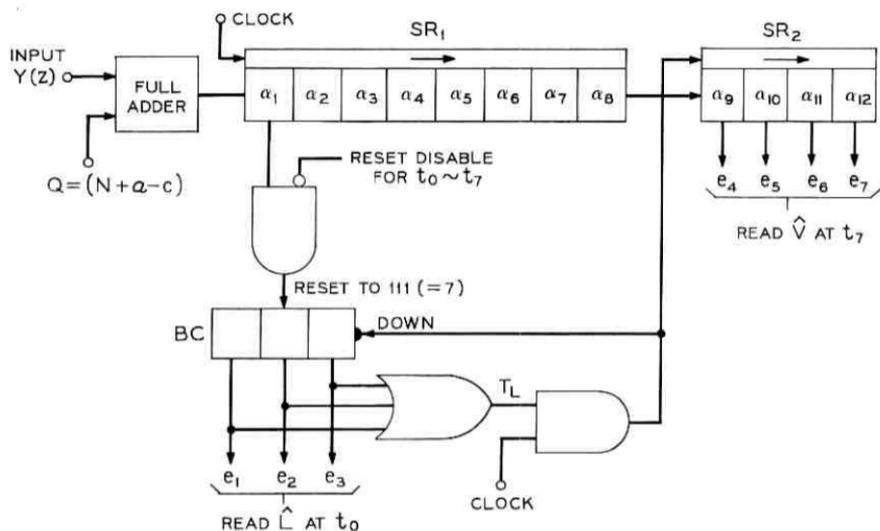

Fig. 16—Fifteen-segment μ-law compressor with DLA ($a = 0.5$), mid-tread ($c = 0$), $m = 3$, and $n = 4$.

the MSD of the input sequence is just stored in the first cell α_1. While the write-in proceeds, the counter BC, which is being counted down by the clock, is reset to the "111" state whenever the entry of the first cell α_1 is 1. Therefore, at the instant t_0 the BC entry is exactly $\hat{L}$, and it is read out at t_0. This also means that the highest order "1" is located at the cell $\alpha_{8-\hat{L}}$, and an additional $\hat{L}$ shifts forces the last entry α_8 of SR_1 to be "1". This is accomplished by shifting SR_1 and SR_2 until the entry of BC becomes "000" at time $t_{\hat{L}}$.

From Algorithm 3 the mantissa output $\hat{V}$ is given by $\Im[W] - N$. Circuit-wise, $\Im[W]$ amounts to discarding the smaller digits below the last stage α_{12} of SR_2, and $(-N)$ means discarding $\alpha_8 = 1$. Thus, $\hat{V}$ is obtained from the cells $(\alpha_9, \alpha_{10}, \alpha_{11}, \alpha_{12})$ of SR_2, which is being stopped after $\hat{L}$ shifts. The $\hat{L}$ output has been read out at time t_0 from BC, thus completing the conversion.

For mid-riser formulation, the circuit can be modified by changing the addend $Q = N + a - c$. A compressor for RLA is also achieved according to Algorithm 4 by adding a correction circuit by θ after the circuit of Fig. 16. The correction circuit may be a parallel adder or a 7-bit binary counter, and obviously the RLA compressor is more complicated than the DLA one.

Among many alternative circuits, a notable feature of the circuit in Fig. 16 is that it can process the input sequence in "real time." While the conversion for a code word is going on, the next code word is being shifted into the SR_1 to prepare for the conversion of the next code word. One might argue that if an input to SR_1 is a sequence beginning from MSD the shift is made from the right to the left simply to yield equation (48). However, doing so one would need additional memory to convert the LSD sequence to the MSD sequence before SR_1.

4.5 *Extension to the Segment A-law*

The synthesis procedure can be extended to the segment A-law using expression (28) instead of (27). The expansion algorithm (corresponding to Algorithm 1) for the A-law is obtained by replacing $P = N \cdot \eta + a$, $Q = a - c$ and the L-shifts by $(L - \eta)$ shifts, where η is equal to 0 for $L = 0$ and 1 otherwise. Following the same procedure, a digital expandor circuit can be synthesized as shown in Fig. 17, for DLA, mid-riser, $m = 3$ and $n = 4$. The A-law expandor differs from the μ-law expandor in Fig. 15 in several respects. The input to the first cell is "η", which is obtained from an OR circuit operating on e_1, e_2, and e_3. The input to the last cell of SR is "1". The gating pulse $T_{L-\eta}$ with length $(L - \eta)$ is obtained by OR operation on e_1

Fig. 17—Thirteen-segment A-law digital expandor with DLA (a = 0.5), mid-riser (c = 0.5), m = 3, and n = 4.

and e_2, and inhibits SR shifting and the SR output. No adder is needed since $Q = 0$.

The compression algorithm (corresponding to Algorithm 3) for the A-law with DLA is obtained by replacing expressions (48) and (49) by

$$W(\hat{L}, Y) \in [\eta N, (1 + \eta)N)$$

and

$$\hat{V} = \Im[W(\hat{L}, Y)] - \eta \cdot N.$$

This algorithm leads to an A-law compressor circuit shown in Fig. 18. The overall performance of this compressor is similar to the μ-law

circuit in Fig. 16. We will indicate only the differences between them. The number of cells in SR_1 is 7. There is no adder in the input since $Q = 0$. The gating pulse $T_{L-\eta}$ is obtained by the same circuit as in the expandor (Fig. 15). To facilitate the 0th segment output $\hat{L} = 0$, a reset pulse t_{-7} is added to BC.

V. SUMMARY

A unified formulation of segment type companding laws has been derived from which the detailed quantization structure of codecs can be explicitly represented. The decoder output levels and the coder decision levels are expressed in terms of standard forms and parameters a, b, and c. The standard form represents a chord-type piecewise linear approximation to smooth μ-law or A-law. The parameter a represents the segment edge effects, b relates to coding procedure, and the parameter c represents the centering. These parameters are typically: DLA ($a = 0.5$ for $b = \pm 0.5$), RLA ($a = 0$ for $b = +0.5$, or $a = 1$ for $b = -0.5$), mid-tread ($c = 0$) and mid-riser ($c = 0.5$). The effect of c is important in small signal range levels. The effect of a on the signal quality is very small, but the mixed use of DLA and RLA approaches introduce small degradations in S/N and net loss tracking.

Synthesis of coders and decoders was described; and it was shown that the design figures such as amount of biasing in the decoder, the
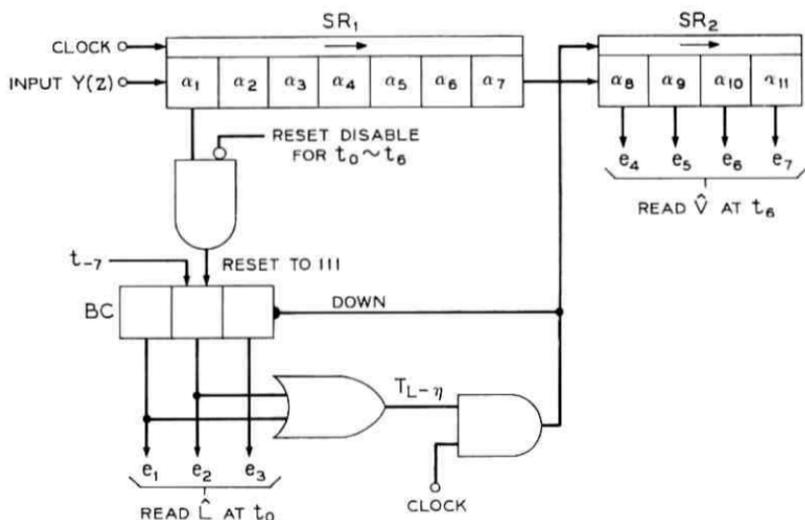

Fig. 18—Thirteen-segment A–law digital compressor with DLA ($a = 0.5$), mid-riser ($c = 0.5$), $m = 3$, and $n = 4$.

type of logic to operate it, and the coding algorithm for the sequential comparison coder can be derived from our formulation. Bipolar codecs are also given in terms of the signed binary code representations.

Another important consequence of this formulation is the systematic synthesis of digital compandors. The expansion algorithm is relatively straightforward. The compression algorithm for DLA was shown to be equivalent to employing the rule of "rounding" for the fractional digits of variable V. Examples of compandor circuits synthesized from these algorithms are shown for the segment μ-law and segment A-law.

The A-law companding circuit appears to be slightly simpler than that for the μ-law, but the difference of the two is not so great as to weigh heavily in the choice between the μ-law and the A-law. It is noted that the formulation and the conversion algorithms developed here can also be conveniently applied for the synthesis of a digital converter between μ-law and A-law or a digital attenuator.

VI. ACKNOWLEDGMENT

APPENDIX A

Proof of Congruence of Equations (15) and (16)

For nonedge points $X(L, V)$, from equations (7) and (14) obviously

$$y_d(X \mid b = +0.5) = y_d(X + 1 \mid b = -0.5) \tag{56}$$

and, hence, (15) and (16) are congruent for nonedge points. For edge points

$$y_d(L, N - 1 \mid b = +0.5) = f(L, N + a - 0.5) - a + c, \tag{57}$$

$$y_d(L + 1, 0 \mid b = -0.5) = f(L + 1, a - 0.5) - a + c$$

$$= f[L, N + 2(a - 0.5)] - a + c. \tag{58}$$

Here, these two equations are linear functions of a. Therefore, the two equations are equal if and only if $a = 0.5$. Thus, equations (15) and (16) are congruent for any X if and only if $a = 0.5$.

APPENDIX B

Proof of the Statements in Section 2.8

For given output levels the optimum decision levels lie halfway between two adjacent output levels.[10] At the nonedge points the proof is obvious from equation (14). At the edge points, the average of $y_0(L, N - 1)$ and $y_0(L + 1, 0)$ is given by

$$\frac{1}{2}\{y_0(L, N - 1) + y_0(L + 1, 0)\} = y_0(L, N - 1) + \frac{1}{2} \Delta_0'(L)$$
$$= y_0(L, N - 1) + \frac{1}{2} \Delta_0(L) \cdot (1 + a)$$
$$= y_d\left(L, N - 1 + \frac{a}{2} \,\middle|\, b = +0.5\right).$$

$$(59)$$

This is equal to $y_d(L, N - 1 \mid b = +0.5)$ if and only if $a = 0$. Also equation (59) is expressed as

$$y_0(L + 1, 0) - \frac{1}{2} \Delta_0'(L) = y_0(L + 1, 0) - \frac{1}{4} \Delta_0(L + 1) \cdot (1 + a)$$
$$= y_d\{L + 1, (1 - a)/4 \mid b = -0.5\}.$$

Therefore, this is equal to $y_d(L + 1, 0 \mid b = -0.5)$ if and only if $a = 1$. This proves that the RLA condition is satisfied if and only if $a = 0$ for $b = +0.5$ or $a = 1$ for $b = -0.5$.

Next, for given decision levels the optimum output levels should be placed halfway between two adjacent decision levels.[10] At the nonedge points this is obviously satisfied. For edge points, since the half way point is equal to $\tilde{y}(L_e, V_e)$, from equation (23) it is shown that $\tilde{y} = y_0$ if and only if $a = 0.5$. Therefore, the DLA condition is satisfied if and only if $a = 0.5$.

APPENDIX C

Derivation of Algorithm 3

Since in DLA $(a = 0.5)$ the upper and lower bound approaches are identical, we will prove Algorithm 3 only for the case of $b = -0.5$. From Algorithm 2, equation (46) can be rewritten as

$$y_d(\hat{L}, \hat{V}) \leq Y < y_d(\hat{L}, \hat{V} + 1).$$

Since $a = 0.5$, this is valid for all $\hat{V} \in \{0, 1, \cdots, N - 1\}$. From equations (56) and (27) the above inequality can be written as

$$\hat{V} + N \leq W(\hat{L}, Y) < \hat{V} + N + 1. \tag{60}$$

Considering the range of $\hat{V} \in \{0, 1, \cdots, N-1\}$, we have as the $\hat{L}$ determining criterion

$$N \leqq W(\hat{L}, Y) < 2N.$$

Then, from equations (50) and (60) we immediately have

$$\hat{V} = \Im[W(\hat{L}, Y)] - N.$$

This completes the derivation of Algorithm 3.

REFERENCES

1. Smith, B., "Instantaneous Companding of Quantized Signals," B.S.T.J., *36*, No. 3 (May 1957), pp. 653–709.
2. Mann, H., Straube, H. M., and Villars, C. P., "A Companded Coder for an Experimental PCM Terminal," B.S.T.J., *41*, No. 1 (January 1962), pp. 1173–226.
3. Schreiner, S. M., and Vallarino, A. R., "48-Channel PCM System," 1957 IRE Nat. Conv. Rec., Part 8, pp. 141–149.
4. Chatelon, A., "Application of Pulse Code Modulation to an Integrated Telephone Network, Part 2—Transmission and Encoding," ITT Elec. Commun., *38*, No. 1 (January 1963), pp. 32–43.
5. CCITT (International Telegraph and Telephone Consultative Committee) Document COM XV-No. 77/E, "Interim Report of Working Party on Question 33/XV," 1966.
6. Bucci, W., "PCM: A Global Scramble for Systems Compatibility," Electronics, *42*, No. 13 (June 23, 1969), pp. 94–102.
7. Henning, H. H., "96 Channel PCM Channel Bank," 1969 Int. Commun. Conf., Denver, Colorado, June 9–11, 1969, pp. 34.17–34.22.
8. Montgomery, W. L., "Digitally Linearizable Compandors with Comments on Project for a Digital Telephone Network," IEEE Trans. on Commun. Technology," *Com-18*, No. 1 (February 1970), pp. 1–4.
9. Purton, R. F., "A Survey of Telephone Speech-Signal Statistics and Their Significance in the Choice of PCM Companding Law," Proc. IEE, *109*, No. 1 (January 1962), pp. 60–66.
10. Max, Joel, "Quantizing for Minimum Distortion," IRE Trans. on Inform. Theory, *IT-6*, No. 1 (March 1960), pp. 6–12.
11. Hoeschele, D. F. Jr., "Analog to Digital/Digital to Analog Conversion Techniques," New York: John Wiley, 1968.
12. Chu, Y., "Digital Computer Design Fundamentals," New York: McGraw-Hill, 1962.
13. Jackson, L. B., Kaiser, J. F., and McDonald, H. S., "An Approach to the Implementation of Digital Filters," IEEE Trans. on Audio and Electroacoustics, *AU-16*, No. 3 (September 1968), pp. 413–421.
14. Kaneko, H., unpublished work.
15. Schaefer, D. H., "Logarithmic Compression of Binary Numbers," Proc. IRE, *49*, No. 7 (July 1961), p. 1219.

Contributors to This Issue

LOUIS D. BABUSCI, A.S., 1941, Newark College of Engineering; Engineering Studies, 1943–44, The Citadel, North Carolina State College, University of Tennessee, Columbia University; Bell Telephone Laboratories, 1963—. Initially, Mr. Babusci was engaged in research and development on sealed nickel-cadmium cells. He has a patent for a process of providing a means of escape of any excessive gas generated within nickel-cadmium cells. Presently he is engaged in research on the lead-acid cell.

VÁCLAV E. BENEŠ, A.B., 1950, Harvard College; M.A. and Ph.D., 1953, Princeton University; Bell Telephone Laboratories, 1953—. Mr. Beneš has been engaged in mathematical research on stochastic processes, traffic theory, servomechanisms, and optimal control. In 1959–60 he was visiting lecturer in mathematics at Dartmouth College. He is the author of General Stochastic Processes in the Theory of Queues (Addison-Wesley, 1963), and of Mathematical Theory of Connecting Networks and Telephone Traffic (Academic Press, 1965). Member, American Mathematical Society, Association for Symbolic Logic, Institute of Mathematical Statistics, SIAM, Mathematical Association of America, Mind Association, Phi Beta Kappa.

R. V. BIAGETTI, B. S., 1960, and M. S., 1962, Providence College; Ph.D., 1966, University of New Hampshire; Bell Telephone Laboratories, 1965—. Since joining Bell Labs, Mr. Biagetti has worked on the lead-acid battery, primarily studying its positive electrode to improve service and reliability. He is Supervisor, Lead-acid Battery Development. Member, American Chemical Society, Sigma Xi.

A. DUANE BUTHERUS, B.A., 1961, Andrews University; Ph.D., 1967, Michigan State University; Bell Telephone Laboratories, 1967—. Mr. Butherus' current interest is the electrochemistry of solid conic conductors and new galvanic cells based on solid electrolytes. Member, AAAS.

ANTHONY G. CANNONE, B.S. (Chemistry), 1958, Seton Hall University; Bell Telephone Laboratories, 1960—. Mr. Cannone has been concerned with vacuum tube materials-processing, cathode development and processing for high power microwave tubes, and the TELSTAR® Projects. He is currently involved with the Bell Telephone Laboratorie; battery, evaluating positive plate designs and cell performance.

R. H. CUSHMAN, studied electrical engineering at Iowa State University, took extension courses at Ohio State University; Western Electric Company, 1961–1970. At Western Electric, Mr. Cushman was Research Leader in charge of the Special Studies Group which engages in the investigation of a wide range of metal bonding and shaping concepts. Prior to that, he had organized and directed his own company involved in the design and manufacture of specialized miniature components for the missile, aircraft and space industries.

DONALD W. DAHRINGER, B.S.Ch.E., 1955, Newark College of Engineering; M.S.Ch.E., 1959, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1961—. Mr. Dahringer has worked in the area of organic materials development. He is currently a Bell System consultant on adhesives and bonding technology. Member, ACS, A.I.Ch.E., NJSPE, Omicron Delta Kappa, Phi Lambda Upsilon, Iota Tau.

D. O. FEDER, B.S., 1946, M.S., 1948, and Ph.D., 1959, Columbia University; Bell Telephone Laboratories, 1954—. Mr. Feder's responsibilities have included studies of electron tube materials and processing; selection and evaluation of materials for use in the TELSTAR® Project; and development and application of nickel-cadmium and lead-acid batteries. He is Head of the Battery Development Department. Member, Electrochemical Society, Sigma Xi, Tau Beta Pi, Phi Lambda Upsilon.

C. G. B. GARRETT, B.A., 1946, M.A., 1950, and Ph.D., 1950, Cambridge University; Bell Telephone Laboratories, 1952—. At Bell Labs., Mr. Garrett has worked on semiconductor surfaces, semiconductor electrochemistry, varactors, organic semiconductors, solid-state and gas lasers, paramagnetic resonance, and nonlinear optics. As Director

of the Electron Device Process and Battery Laboratory, he is responsible for fundamental and exploratory work on chemical processes and on the properties of new structures related to integrated circuits. He is also concerned with the development and applications engineering of batteries. Fellow, American Physical Society.

L. T. GUSLER, B.S., 1962, and M.S., 1964, Mississippi State University; Bell Telephone Laboratories, 1964—. Mr. Gusler has been concerned with the effects of precipitation on microwave transmission. He is currently engaged in studies of the coupling between terrestrial and satellite communications systems due to scattering by rain and aircraft. Member, IEEE, Eta Kappa Nu, Tau Beta Pi.

D. C. HOGG, B.Sc., 1949, University of Western Ontario; M.Sc., 1950, Ph.D., 1953, McGill University; Bell Telephone Laboratories, 1953—. His work has included studies of artificial dielectrics for microwaves, diffraction of microwaves, and over-the-horizon, millimeter wave and optical propagation, sky noise and low-noise antennas. Fellow, IEEE; Chairman U. S. Commission 2 of Union de Radio Scientifique Internationale, Sigma Xi, American Association for the Advancement of Science.

P. HUBBAUER, B.S., 1956, Budapest Technical Institute of Hungary; attended Fairleigh Dickinson University and Newark College of Engineering, Bell Telephone Laboratories, 1967—. Mr. Hubbauer is involved in design and processing problems.

T. W. HUSEBY, B.S. (Ch.E.), 1960, Purdue University; M.S. (plastics engineering), 1961, and Ph.D. (aerospace and mechanical science), 1965, Princeton University; Bell Telephone Laboratories, 1961—. Mr. Huseby initially worked on injection molding and mold design for plastic materials. He then turned to applied research in rheology of polymer melts and thermodynamic studies of polymers. He is now supervisor of a group responsible for the development and testing of plastic materials for Bell System applications. Member, American Physical Society, Society of Rheology, Society of Plastic Engineers, Sigma Xi, Tau Beta Pi.

HISASHI KANEKO, B.S.E.E., 1956, University of Tokyo (Japan); M.S.E.E., 1962, University of California; Dr. Eng., 1967, University of Tokyo; Bell Telephone Laboratories, 1968–70. At Bell Telephone Laboratories, Mr. Kaneko worked on future digital channel banks. Member, IEEE, Institute of Electronics and Communication Engineers of Japan.

D. E. KOONTZ, B.S. in Chemistry, 1945, Youngstown University; M.S. in Chemistry, 1948, and Ph.D. in Chemistry, 1951, University of Pittsburgh; Bell Telephone Laboratories, 1952—. Mr. Koontz was first engaged in studies of electron tube materials and processing. For the TELSTAR® Project, he was responsible for the selection and application of materials for the space vehicle and their evaluation in simulated space environments. Later he was responsible for the materials and chemistry associated with the development of electron tubes, sealed nickel-cadmium batteries, ultrasonic devices, masers, and magnetic and solid-state microwave devices. He is currently Head of the Exploratory Technology Development Department. Member, American Chemical Society, Sigma Xi, Phi Lambda Upsilon, American Society for Testing Materials.

J. A. LEWIS, B.S., 1944, Worcester Polytechnic Institute; M.S., 1948, and Ph.D., 1950, Brown University; Bell Telephone Laboratories, 1951—. Mr. Lewis has worked on problems in piezoelectricity, elasticity, and heat conduction. He is currently concerned with semiconductor problems. Member, American Mathematical Society, Society for Industrial and Applied Mathematics, American Institute of Aeronautics and Astronautics.

W. S. LINDENBERGER, B.S., 1965, Guilford College; 1966, Clarkson College of Technology; Bell Telephone Laboratories, 1967—. Mr Lindenberger's work has been concerned with materials for electronic components. Member, American Chemical Society, American Vacuum Society, The Catalysis Society.

HENRY J. LUER, B.S.M.E., 1956, Newark College of Engineering; Bell Telephone Laboratories, 1956—. Mr. Luer's work has included development of guided missile inertial instruments and guidance systems. He is presently engaged in the physical design of power equipment.

PAUL C. MILNER, B.S., 1952, Haverford College; M.A., 1954, and Ph.D., 1956, Princeton University; Bell Telephone Laboratories, 1957—. Mr. Milner has worked in the areas of electrochemical kinetics and energy storage and is currently head of the Electrochemical Research and Development Department. Member, American Association for the Advancement of Science, American Chemical Society, Electrochemical Society.

T. D. O'SULLIVAN, B.S., 1961, Manhattan College; Ph.D., 1968, Fordham University; Bell Telephone Laboratories, 1966—. Mr. O'Sullivan is engaged in investigating the electrochemical phenomenon associated with the lead-acid battery. Member, American Chemical Society, Sigma Xi, Phi Lambda Upsilon.

J. T. RYAN, B.S. (Ch.E.), 1959, and M.S. (Ch.E.), 1962, Newark College of Engineering; Bell Telephone Laboratories, 1952—. Initially Mr. Ryan worked with casting resins. In 1957 he began his present assignment where he is responsible for polyolefin, polyvinyl chloride, and epoxy transfer molding compounds materials development. Member, American Society for Testing and Materials Committee D-20 on plastics.

LOUIS H. SHARPE, B.S. (Chemistry-Honors), 1950, Virginia Polytechnic Institute; Ph.D. (Physical Chemistry), 1957, Michigan State University; Bell Telephone Laboratories, 1955—. Prior to 1968, Mr. Sharpe was concerned with the phenomenon of adhesion. He presently supervises the Adhesives Engineering and Development Group in the Organic Materials Research and Development Department. Member, American Chemical Society, New York Academy of Sciences, Sigma Xi, Sigma Pi Sigma, Alpha Chi Sigma, Phi Lambda Upsilon; Fellow, American Institute of Chemists. He is Editor-in-Chief of The Journal of Adhesion.

DAVID A. SHNIDMAN, B.S. and M.S. (electrical engineering), 1959, Massachusetts Institute of Technology; Ph.D. (applied mathematics), 1965, Harvard University; Data Sciences Lab at Air Force Cambridge Research Laboratory, 1959–1965; Bell Telephone Laboratories, 1965—.

At AFCRL, Mr. Shnidman was concerned with problems of reliable communications at low signal-to-noise ratios. At Bell Laboratories, as a member of the Coaxial Systems Department, he is involved with problems related to high speed digital signal, transmitted over coaxial lines. Member, IEEE, AAAS, Eta Kappa Nu, Tau Beta Pi, Sigma Xi.

JAYANT R. SHROFF, B.Sc. with honors, University of Bombay; B.S.Ch.E., 1965, University of Florida; M.S.Ch.E., 1967, Drexel Institute of Technology; Bell Telephone Laboratories, 1968—. Mr. Shroff has been engaged in various aspects of Bell System lead-acid battery development program. Areas of activities include materials evaluation, components design, and testing. Member, Phi Lamda Upsilon, American Chemical Society, American Institute of Chemical Engineers.

F. J. VACCARO, B.S. (Chemistry), 1957, and M.A. (Chemistry), 1967, Brooklyn College; Western Electric Company, Inc., 1957–1966; Bell Telephone Laboratories, 1968—. From 1964 to 1966, Mr. Vaccaro was temporarily assigned to the Lead-Acid Battery Development Group at Bell Labs, engaged in life testing, and experimental design and studies. He is presently responsible for the design and evaluation of post seals for the circular lead-acid cell and for the selection and electrochemical evaluation of materials for that cell.

ARVIDS VIGANTS, B.E.E., 1956, City College of New York; M.S. (E.E.), 1957, Eng. Sc.D. (E.E.), 1962, Columbia University; Bell Telephone Laboratories, 1962—. Mr. Vigants has worked on various electromagnetic wave propagation problems, and is currently working on line-of-sight microwave propagation. Member, IEEE, Eta Kappa Nu, Tau Beta Pi, Sigma Xi, Commission 2 of URSI.

MICHAEL C. WEEKS, B.A., (Chemistry), 1967, Taylor University; Bell Telephone Laboratories, 1967—. Since joining Bell Telephone Laboratories, Mr. Weeks has worked in the Battery Development area, primarily in developing the positive paste material used in the new circular lead-acid cell.

B.S.T.J. BRIEFS

On the Computation of the Far Field of Open Cassegrain Antennas

By H. ZUCKER

(Manuscript received February 2, 1970)

The far field radiation patterns of open cassegrain antennas have been previously computed by a double integration[1]. Recently one integration has been eliminated by using a modified stationary phase approximation[2]. The rather remarkable accuracy thus obtained motivated an investigation to estimate the error in this approximation.

In this note we present an alternate formulation for the computation of the far field and provide an additional estimate for the error in the modified stationary phase approximation. It is also shown that for certain illuminations, there exists a closed form solution to the azimuthal integration. For illuminations derived from a linear combination of TE_{1n}^0 and TM_{1n}^0 waveguide modes, the integrals can be expressed in terms of convergent series of Bessel functions.

Based on the aperture field method, the far field electric $\bar{E}_f$ of an open cassegrain antenna is related to the reflected field at the aperture, $\bar{E}_r$, by the following double integral[1]

$$\bar{E}_f = j \frac{\exp(-jkR_a)}{R_a} \int_0^{\theta_m} \int_0^{2\pi} \bar{E}_r(\theta, \phi)$$

$$\cdot \exp[jk \sin \theta_a(x_p \cos \phi_a + y_p \sin \phi_a)]r^2 \sin \theta \, d\theta \, d\phi \qquad (1)$$

where R_a, θ_a, ϕ_a are the far field spherical observation coordinates; λ is the free space wavelength and $k = 2\pi/\lambda$ is the propagation constant; and θ_m is the illumination angle. Referring to Fig. 1, the rectangular x_p and y_p coordinates are related to the spherical θ, ϕ coordinates by

$$x_p = r(\cos \theta_0 \sin \theta \cos \phi + \sin \theta_0 \cos \theta), \qquad (2)$$

$$y_p = r \sin \theta \sin \phi, \qquad (3)$$

$$r = 2f/(a - b \cos \phi), \qquad (4)$$

where

$$a = 1 + \cos \theta \cos \theta_0, \qquad (5)$$
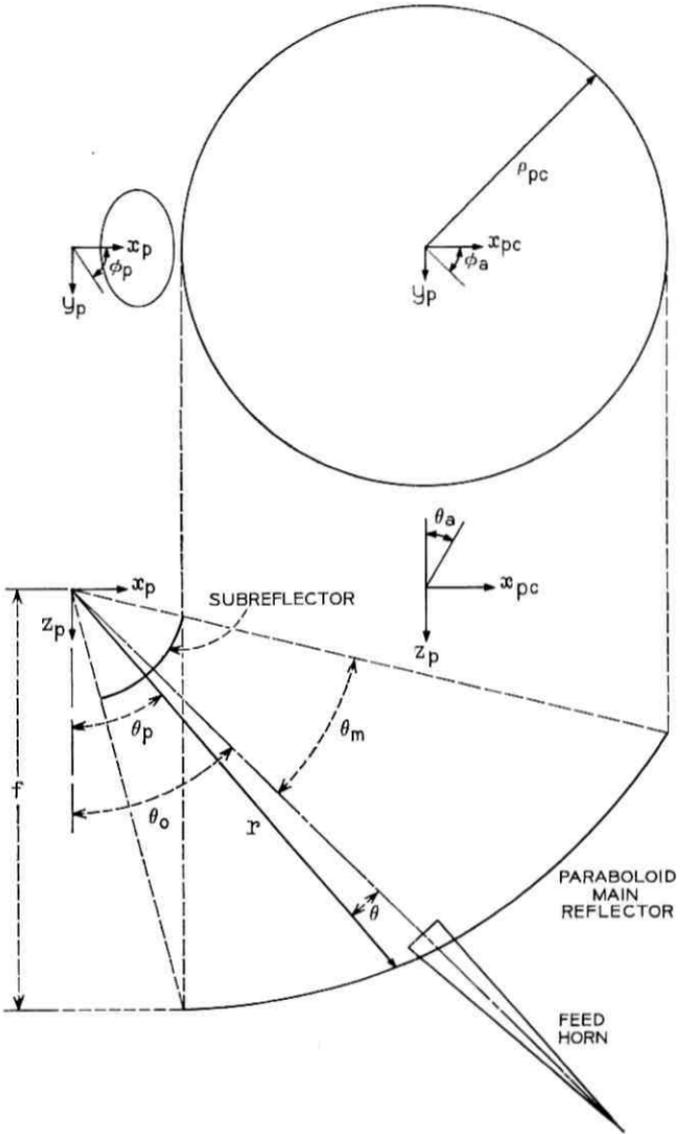
Fig. 1—Open cassegrain antenna.

$$b = \sin \theta \sin \theta_0 , \tag{6}$$

and

$$a^2 - b^2 = c^2, \tag{7}$$

$$c = \cos \theta + \cos \theta_0 . \tag{8}$$

θ_0 is the offset angle, and f is the focal length of the paraboloid.

For combined $TE^0_{1n} - TM^0_{1n}$ mode excitation and for y polarization of the horn feed the reflected electric field at the aperture, $(\bar{E}_{ry})$, can be written

$$(1 + \cos \theta_0)(\bar{E}_{ry}) = 1_{xp} \sin \phi \left[cE_c(0) \cos \phi + (b + a \cos \phi)E_c\left(\frac{\pi}{2}\right) \right]$$

$$- 1_{yp} \left[(a \cos \phi - b)E_c(0) \cos \phi + cE_c\left(\frac{\pi}{2}\right) \sin^2 \phi \right] \tag{9}$$

where $E_c(0)$ and $E_c(\pi/c)$ are the fields radiated by the subreflector in the principal planes $\phi = 0$ and $\phi = \pi/2$ at a distance $2f/(1 + \cos \theta_0)$ from the focal point.

The ϕ integration in equation (1) was recently approximated by a modified stationary phase approximation[2]. An examination of this approximation reveals that the resulting phase factors and the arguments of the Bessel functions have a simple geometric interpretation. Specifically the phase factors are directly related to the location in the x_p, y_p plane of the centers of the $\theta = $ constant circles (see Fig. 2 in Ref. 2) and the arguments of the Bessel functions are related to the radii of these circles.

This suggests, for the ϕ integration, the following coordinate transformation

$$x_p = r_c \cos \phi_1 + x_c , \tag{10}$$

$$y_p = r_c \sin \phi_1 \tag{11}$$

with

$$r_c = \frac{2f \sin \theta}{c} , \tag{12}$$

$$x_c = \frac{2f \sin \theta_0}{c}. \tag{13}$$

This transformation gives in the x_p, y_p plane, for the $\theta = $ constant coordinates the same set of circles as obtained from equations (2) and

(3), and for the ϕ_1 = constant coordinates a set of hyperbolas with the equation

$$x_p^2 - y_p^2 - 2x_p y_p \, \text{ctn} \, \phi_1 + 4f(x_p - y_p \, \text{ctn} \, \phi_1) \, \text{ctn} \, \theta_0 = (2f)^2. \quad (14)$$

The sets of curves which correspond to the transformation (10) and (11) are shown in Fig. 2.

In the θ, ϕ_1 coordinate system the rectangular x, y components of the far field have a much simpler representation. In particular for y polarization these components are

$$(\bar{E}_{fy}) = j \frac{\exp{(-jkR_a)}}{\lambda R_a} \int_0^{\theta_m} (\bar{E}_y) \sin \theta \, d\theta \quad (15)$$

where the subscript y is to designate the y polarization, and the x, y components of $(\bar{E}_y)$ are

$$(E_y)_y = -\frac{4f^2}{(1 + \cos \theta_0)c} \exp{(jx_e \sin \theta_a \cos \phi_a)}$$

$$\cdot \int_0^{2\pi} \left[E_e(0)(a \cos \phi_1 + b) \cos \phi_1 + E_e\left(\frac{\pi}{2}\right)c \sin^2 \phi_1 \right]$$

$$\cdot \frac{\exp{[jkr_e \sin \theta_a \cos (\phi_1 - \phi_a)]}}{a + b \cos \phi_1} \, d\phi_1 \quad (16)$$

and

$$(E_y)_x = \frac{4f^2 \exp{(jkx_e \sin \theta_a \cos \phi_a)}}{(1 + \cos \theta_0)c}$$

$$\cdot \int_0^{2\pi} \left[E_e(0)(a \cos \phi_1 + b) - E_e\left(\frac{\pi}{2}\right)c \cos \phi_1 \right] \sin \phi_1$$

$$\cdot \frac{\exp{[jkr_e \sin \theta_a \cos (\phi_1 - \phi_a)]}}{a + b \cos \phi_1} \, d\phi_1 \,. \quad (17)$$

In deriving equations (15), (16) and (17), the Jacobian, J, of the transformation (10) and (11) was used, which is

$$J = \frac{(2f)^2}{c^3} (a + b \cos \phi_1) \sin \theta. \quad (18)$$

It is readily shown that by using a two term expansion of

$$[1/(a + b \cos \phi_1)] \approx (1/a)(1 - b/a \cos \phi_1),$$

the integrals (16) and (17) give the same values as those obtained by the modified stationary phase approximation but without the term
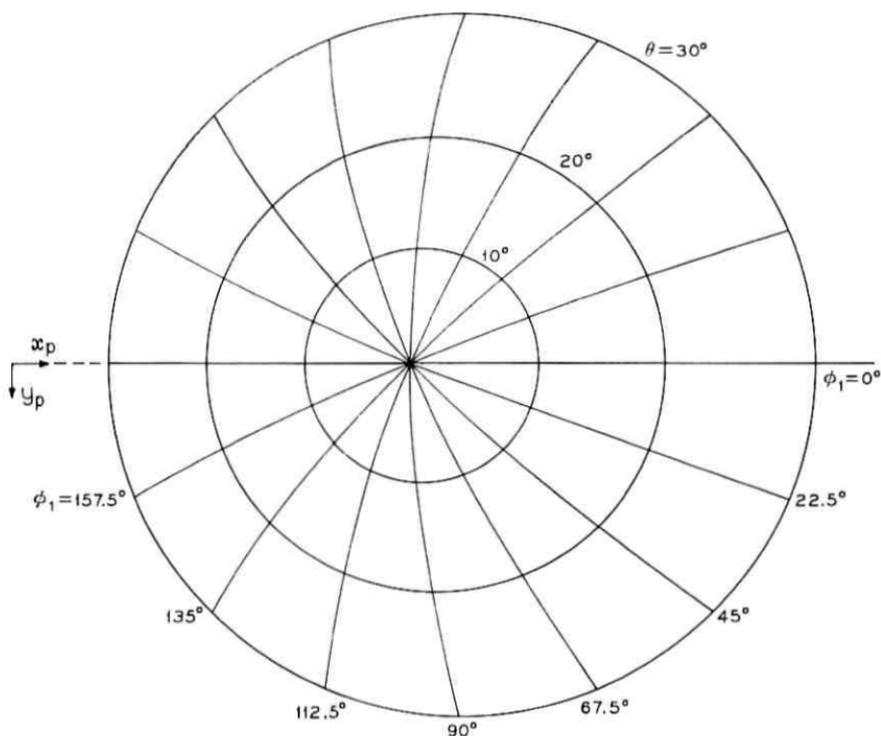
Fig. 2—Transformation of (θ, ϕ_1) coordinates in the (x_p, y_p) plane.

$(\sin \theta \sin \theta_0 \cos \phi_a)^2$ in the denominator. This shows that the modified stationary phase approximation is at least of order $(b/a)^2$, for all observation angles and with the additional property that for large angles off axis it is of order $[(b/a)^2/2kf \sin \theta_a]$.

There are two cases for which the integration of either equations (16) or (17) can be performed in closed form.

(i)
$$aE_c(0) = cE_c\left(\frac{\pi}{2}\right)$$

where a and c are given by equations (5) and (8).

Such illuminations can be approximately realized by combined $TE_{11}^0 - TM_{11}^0$ mode excitations. For this case the integration of equation (16) yields

$$(E_u)_u = -8\pi^2 f^2 \frac{\exp{(jkx_c \sin \theta_a \cos \phi_a)}}{(1 + \cos \theta_0)c} E_c(0) J_0(x) \qquad (19)$$

where J_0 is a Bessel function of order zero with the argument

$$x = kr_c \sin \theta_a . \tag{20}$$

The evaluation of the corresponding expression based on the modified stationary phase approximation yields

$$(E_v)_v = -\frac{8\pi^2 f^2 \exp(jx_c \sin \theta_a \cos \phi_a)}{(1 + \cos \theta_0)c}$$
$$\cdot E_c(0)\left[J_0(x) + \frac{b^2}{a^2 - b^2 \cos^2 \phi_a} \frac{J_1(x)}{x} \cos 2\phi_a \right]. \tag{21}$$

The error in the approximation is the second term in equation (21). That term is zero in the planes $\phi_a = \pi/4$ and $\phi_a = 3/4\pi$. In other planes, the error is of order $(b/a)^2/x$, and is in agreement with the error estimate given above.

The dependence of the second term in equation (21) on θ is shown in Fig. 3 for $\theta_0 = 55°$, $f = 80\lambda$ and $\theta_a = 2.5°$. These parameters are the same as those used in the computations in Ref. 2, Fig. 5. A comparison of the two figures shows good agreement, as would be expected, since the condition for the illumination was approximately satisfied.

(ii) $$cE_c(0) = aE\left(\frac{\pi}{2}\right).$$

These illuminations may also be approximately realized. For this condition the cross-polarized component has a closed form solution given by

$$(E_v)_x = \frac{j8\pi^2 f^2 \exp(jkx_c \sin \theta_a \cos \phi_a)}{(1 + \cos \theta_0)c} E_c(0)\left(\frac{b}{a}\right)J_1(x) \sin \phi_a . \tag{22}$$

An evaluation of the corresponding expression based on the stationary phase approximation gives*

$$(E_v)_x = j\frac{8\pi^2 f^2 \exp(jkx_c \sin \theta_a \cos \phi_a)}{(1 + \cos \theta_0)c}$$
$$\cdot E_c(0)\left(\frac{b}{a}\right)\left[J_1(x) \sin \phi_a + \frac{b^2}{a^2 - b^2 \cos^2 \phi_a} \frac{J_2(x)}{x} \sin 3\phi_a \right]. \tag{23}$$

The error in the modified approximation is the second term and is of the same order as the error term in equation (21).

For illuminations derived from a linear combination of TE_{1n}^0 and TM_{1n}^0 waveguide modes, the integrals (15) and (16) can be expressed in convergent series of Bessel function, by using the following Fourier

---

* Equation (21) in Ref. 2 has a misprint and should be divided by 2.
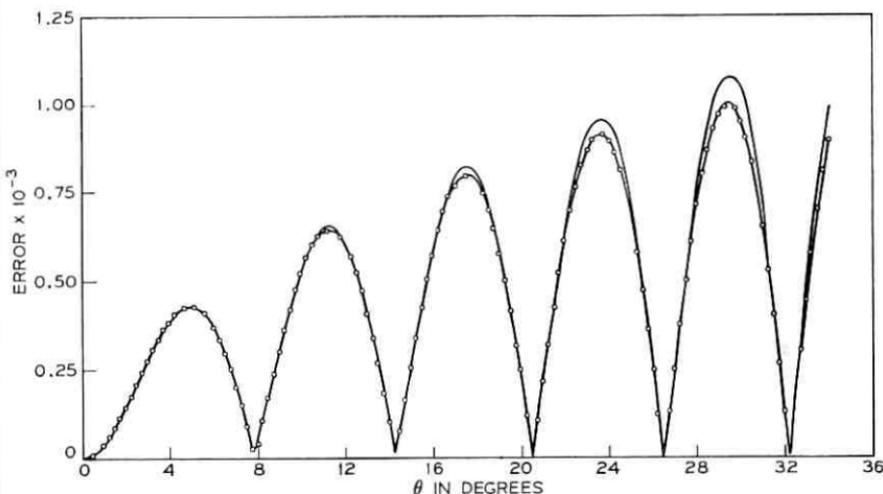
Fig. 3—Error introduced by the modified stationary phase approximation. Angle off axis, $\theta_a = 2.5$ degrees. ———— $\phi_a = 0$ (H plane); $\circ\circ\circ$ $\phi_a = \pi/2$ (E plane).

series expansions

$$\exp[jx \cos(\phi - \phi_a)] = J_0(x) + 2 \sum_{n=1}^{\infty} j^n J_n(x) \cos n(\phi - \phi_a) \tag{24}$$

and

$$\frac{1}{a + b \cos \phi} = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos n\phi \tag{25}$$

with

$$c_n = \frac{2}{\pi} \int_0^{\pi} \frac{\cos n\phi}{a + b \cos \phi} \, d\phi. \tag{26}$$

The integral (26) is tabulated[3] and by using equation (7) gives

$$c_n = \frac{2}{c} (-1)^n \left(\frac{a - c}{a + c}\right)^{n/2}. \tag{27}$$

With these expansions, after some mathematical manipulations the following expressions are obtained for the integrals

$$
\begin{aligned}
(E_v)_v = &-\frac{8\pi f^2 \exp(jx_c \sin \theta_a \cos \phi_a)}{(1 + \cos \theta_0)(a + c)c} \left\{ c \left[E_c(0) + E_c\left(\frac{\pi}{2}\right) \right] J_0(x) \right. \\
&+ \left[aE_c(0) - cE_c\left(\frac{\pi}{2}\right) \right] \left[j\left(\frac{a - c}{a + c}\right)^{\frac{1}{2}} J_1(x) \cos \phi_a \right. \\
&\left. \left. - \frac{2c}{a + c} \sum_{n=0}^{\infty} (-j)^n \left(\frac{a - c}{a + c}\right)^{n/2} J_{n+2}(x) \cos(n + 2)\phi_a \right] \right\}
\end{aligned} \tag{28}
$$

and

$$(E_y)_x = \frac{8\pi f^2 \exp (jx_c \sin \theta_a \cos \phi_a)}{(1 + \cos \theta_0)(a + c)^2 c} \left\{ jb\left[(a + 2c)E_c(0) + cE_c\left(\frac{\pi}{c}\right)\right] \right.$$

$$\cdot J_1(x) \sin \phi_a + 2c\left[cE_c(0) - aE_c\left(\frac{\pi}{2}\right)\right]$$

$$\left. \cdot \sum_{n=0}^{\infty} (-j)^n\left(\frac{a - c}{a + c}\right)^{n/2} J_{n+2}(x) \sin (n + 2)\phi_a \right\}. \tag{29}$$

From equations (5) and (7)

$$\left(\frac{a - c}{a + c}\right)^{\frac{1}{2}} = \tan \theta/2 \tan \theta_0/2. \tag{30}$$

For most antenna designs, equation (30) is much smaller than one, the series in equations (28) and (30) are, therefore, rapidly converging, and can in principle be evaluated to any desired accuracy.

It is noted that in the principal planes, these series can be related to Lommel functions of two variables[3].

The stimulating discussions with N. Amitay and E. R. Nagelberg are gratefully acknowledged by the author.

REFERENCES

1. Cook, J. S., Elam, E. M., and Zucker, H., "The Open Cassegrain Antenna," B.S.T.J., *44*, No. 7 (September 1965), pp. 1255–13001.
2. Ierley, W. H., and Zucker, H., "A Stationary Phase Method for the Computation of the Far Field of Open Cassegrain Antennas," B.S.T.J., *49*, No. 3 (March 1970), pp. 431–454.
3. Gradshteyn, I. S., and Ryzhik, I. M., *Tables of Integrals Series and Products*, New York: Academic Press, 1965, pp. 366, 987.

Holographic Thin Film Couplers

By H. KOGELNIK and T. P. SOSNOWSKI

Recently P. K. Tien and his co-workers have described a prism coupler as a convenient means to feed light into a single mode of a guiding thin optical film.[1] Distributed couplers of this kind are of great interest for integrated optical devices. In this brief we describe thin

and

$$(E_y)_x = \frac{8\pi f^2 \exp (jx_c \sin \theta_a \cos \phi_a)}{(1 + \cos \theta_0)(a + c)^2 c} \left\{ jb \left[(a + 2c)E_c(0) + cE_c\left(\frac{\pi}{c}\right) \right] \right.$$

$$\cdot J_1(x) \sin \phi_a + 2c \left[cE_c(0) - aE_c\left(\frac{\pi}{2}\right) \right]$$

$$\left. \cdot \sum_{n=0}^{\infty} (-j)^n \left(\frac{a - c}{a + c}\right)^{n/2} J_{n+2}(x) \sin (n + 2)\phi_a \right\}. \tag{29}$$

From equations (5) and (7)

$$\left(\frac{a - c}{a + c}\right)^{\frac{1}{2}} = \tan \theta/2 \tan \theta_0/2. \tag{30}$$

For most antenna designs, equation (30) is much smaller than one, the series in equations (28) and (30) are, therefore, rapidly converging, and can in principle be evaluated to any desired accuracy.

It is noted that in the principal planes, these series can be related to Lommel functions of two variables[3].

The stimulating discussions with N. Amitay and E. R. Nagelberg are gratefully acknowledged by the author.

REFERENCES

1. Cook, J. S., Elam, E. M., and Zucker, H., "The Open Cassegrain Antenna," B.S.T.J., 44, No. 7 (September 1965), pp. 1255–13001.
2. Ierley, W. H., and Zucker, H., "A Stationary Phase Method for the Computation of the Far Field of Open Cassegrain Antennas," B.S.T.J., 49, No. 3 (March 1970), pp. 431–454.
3. Gradshteyn, I. S., and Ryzhik, I. M., Tables of Integrals Series and Products, New York: Academic Press, 1965, pp. 366, 987.

Holographic Thin Film Couplers

By H. KOGELNIK and T. P. SOSNOWSKI

Recently P. K. Tien and his co-workers have described a prism coupler as a convenient means to feed light into a single mode of a guiding thin optical film.[1] Distributed couplers of this kind are of great interest for integrated optical devices. In this brief we describe thin

film coupling with a thick dielectric grating. To produce these grating couplers, we have used the materials and techniques of holography where dielectric gratings are known to yield high ($\gtrsim$ 90%) diffraction efficiencies.[2] Independently A. Ashkin and E. Ippen[3] suggested that a grating could be used as a thin film coupling device, and very recently M. L. Dakss, *et al.*,[4] have reported successful light coupling into thin films by means of a (thin) phase grating made of photoresist.

Figure 1 shows a grating coupler. A diffraction grating, placed in the vicinity of the guiding film, diffracts the incident light. If the diffracted wave is phase-matched to a mode of the film, then coupling occurs and light is fed into the film. The grating coupler is a distributed coupler just as the prism coupler, and much of the concepts and the theory developed for the latter[1] can be applied.

In order to make a good and efficient grating coupler one has to (*i*) use lossless and scatterfree materials, (*ii*) suppress unwanted grating orders, and (*iii*) provide for a sufficiently deep spatial modulation of the optical phase shift to achieve strong coupling and the associated short coupling lengths. Point (*i*) restricts us to phase or dielectric gratings. There are two possibilities to satisfy point (*ii*). The first is to use a thick grating and light incident near the Bragg angle. Then, Bragg effects will suppress all but one diffraction order. We shall call this coupler type a "Bragg coupler". The second possibility is to use a grating with a very large number of lines (or fringes) per millimeter. This results in such a large diffraction angle that only one diffraction order can propagate while all the others are beyond cutoff. This large diffraction angle leads to a thin film mode which travels in a direction reverse to that of the incident light, which is why we shall call this coupler a "reverse coupler." In many of the holographic materials which are available today, the achievable refractive index changes are relatively small (10^{-5} to 10^{-2}). It is, therefore, easier to satisfy point (*iii*) by using a Bragg coupler, where the phase shift accumulates throughout the thickness of the grating. The resulting strong coupling leads to short coupling lengths. These are desirable for miniaturization, and they make it easier to maintain the tolerances which are needed for phase-matching.

Figure 1 shows the geometry of a grating coupler, the choice of the coordinate system, and several parameters of interest. In our case the dielectric grating is formed in a layer of gelatin of refractive index n_g which is deposited directly onto the guiding film of index $n_f > n_g$. The grating is characterized by the grating vector **K** which is oriented perpendicular to the fringe planes and has a magnitude
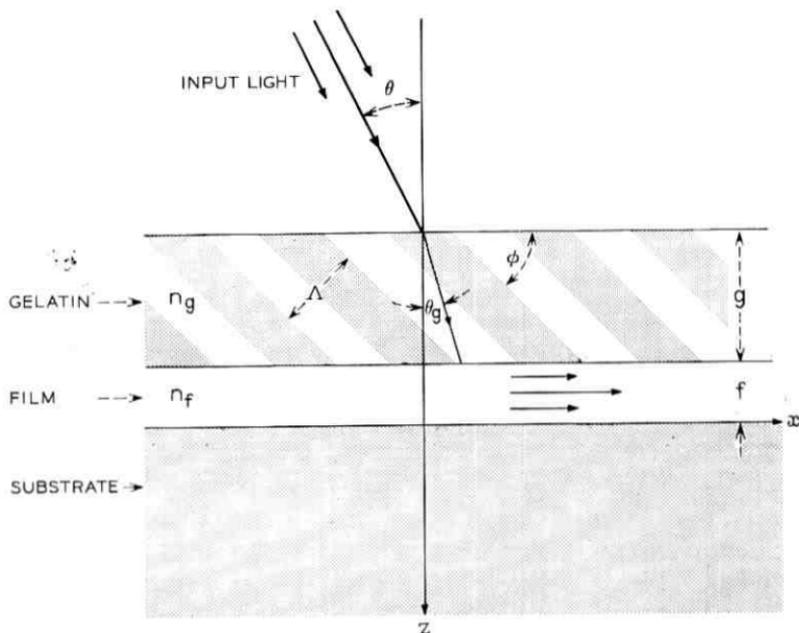
Fig. 1—Cross section of a grating coupler. Λ is the fringe spacing, ϕ the slant angle, and θ the angle of incidence. A transmission grating is shown. For a reflection grating the light is incident from the substrate side.

$$K = 2\pi/\Lambda \qquad (1)$$

where Λ is the fringe spacing. The fringes are slanted with respect to the grating surface by an angle ϕ. The propagation vector of the incident light is $\mathbf{k}_{in}$ and has a magnitude equal to the free space propagation constant $k_0 = 2\pi/\lambda$. In the gelatin the k-vector of this light is $\mathbf{k}_g$ which is of magnitude $n_g k_0$.

The diffracted wave has a k-vector equal to $(\mathbf{k}_g + \mathbf{K})$. It is phase-matched to a film mode when this k-vector has a tangential $(x-)$ component equal to the propagation constant β of the film mode, i.e., when

$$\beta = (\mathbf{k}_g + \mathbf{K})_z = (\mathbf{k}_{in} + \mathbf{K})_z . \qquad (2)$$

Figure 2 shows the k-vector diagrams for the Bragg coupler (a) and the reverse coupler (b). The latter is shown for reasons of comparison. Note that in this latter case the grating vector $\mathbf{K}$ is parallel to the x-axis (which is typical for a thin grating). The cross marks the k-vector of the -1 order which is generated beyond cutoff. The circles of radius

$n_g k_0$ and k_0 respectively indicate the locus of the k-vectors of the input light in the gelatin and in air for a variable angle of incidence. The line spaced a distance $\beta > n_g k_0$ away from the z-axis is the matching line. Phase-matching occurs when the vector sum $(\mathbf{k}_g + \mathbf{K})$ terminates on this line, as shown in the figure. The Bragg condition is obeyed when

$$\cos(\phi - \theta_g) = K/2n_g k_0 \tag{3}$$

where θ_g is the angle of incidence in the gelatin. Geometrically this implies that the vector sum $(\mathbf{k}_g + \mathbf{K})$ would terminate on the $n_g k_0$-circle.

It is clear from Fig. 2a that the Bragg condition and the phase-match condition cannot be met for the same angle of incidence. One can show that there is a minimum possible difference $\Delta\theta_{g\ min}$ between the Bragg angle and the matching angle which is approximately given by

$$\Delta\theta_{g\ min} \approx (n_f - n_g)/n_g \tag{4}$$

where we have assumed that this "detuning angle" is small and that $\beta \approx n_f k_0$. The detuning angle is typically a few degrees of arc. The angular width $2\Delta\theta_{g\ Bragg}$ of the Bragg response between half-power points is[2]

$$2\Delta\theta_{g\ Bragg} = \Lambda/g \tag{5a}$$

for transmission gratings, and

$$2\Delta\theta_{g\ Bragg} = (\Lambda/g)\cot\theta_g \tag{5b}$$

for reflection gratings, where g is the grating thickness. These formulas
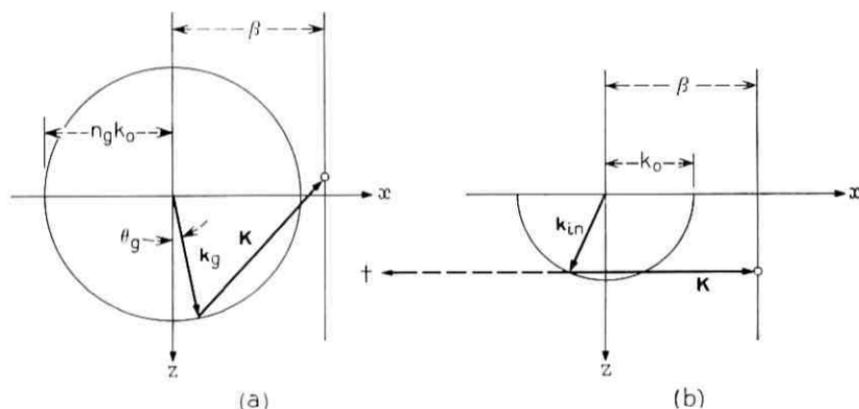
Fig. 2—k-vector diagrams for a Bragg coupler (a) and a reverse coupler (b). $\mathbf{K}$ is the grating vector; β, the propagation constant of a film mode.

assume unslanted gratings. Formulas are available also for slanted gratings[2], and one expects widths of the order of Λ/g, which is typically a few degrees. To take advantage of Bragg effects, we have to bring the matching angle as close as possible to the Bragg angle, and at least onto the wings of the Bragg response. Then we can expect suppression of unwanted orders and sufficiently strong coupling, which is, indeed, what we have observed experimentally.

Our experiments were done with Bragg couplers. We made the dielectric gratings in dichromated gelatin using the preparation and development techniques described in Ref. 5. Gelatin layers of about 4 μm thickness were deposited on the guiding films using a dipcoating technique. The films were low-loss sputtered films of Corning 7059 glass which were kindly supplied by J. E. Goell and D. R. Standley.[6] They were about 0.3 μm thick and had a refractive index of $n_f = 1.62$.

We designed the couplers for light of wavelength $\lambda = 0.6328$ μm, incident perpendicular to the film plane. This prescribes gratings with a slant angle ϕ of about 45° and a fringe spacing of $\Lambda = 0.25$ μm (i.e., 4000 lines/mm). To obtain the right grating parameters, some experimentation is necessary because the gelatin shrinks during development. We produced the desired fringe pattern holographically by exposing the sensitized gelatin at the shorter wavelength of $\lambda' = 0.4416$ μm, which is a line of the cadmium laser. The k-vector diagram of Fig. 3 indicates how the use of this shorter wavelength gives us fringes with the wanted k-vector at convenient angles of incidence. To get the proper interference angles in the gelatin, the two collimated light beams
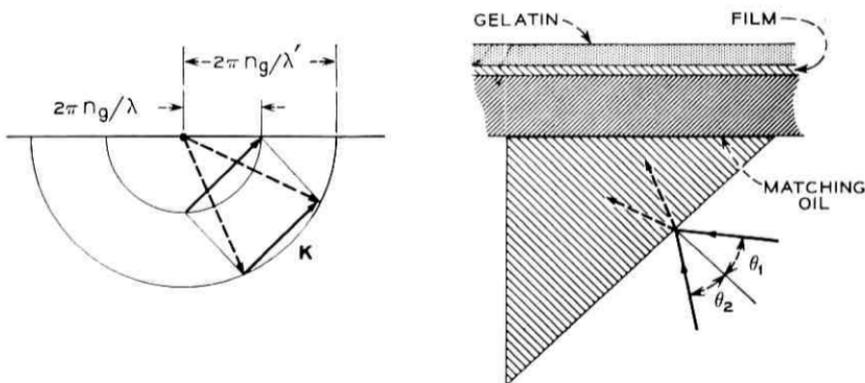

Fig. 3—Arrangement for holographic exposure used to make the grating coupler. The k-vector diagram shows how the grating vector **K** is conveniently produced by the interference of blue light (λ').
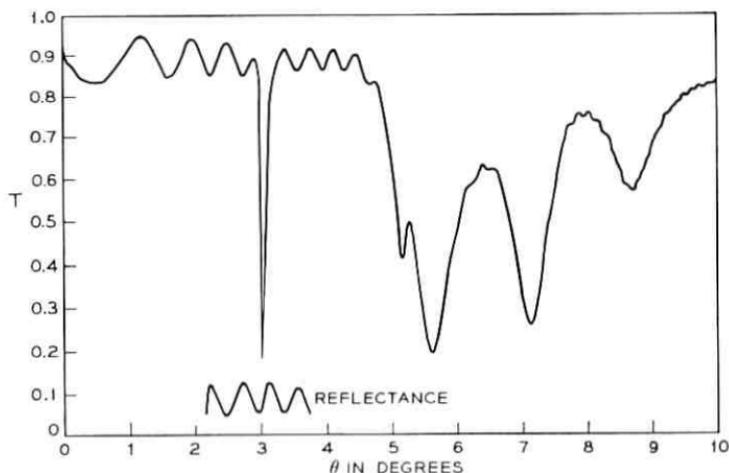
Fig. 4—Angular transmission spectrum for $2w_0 = 0.3$ mm. T is the transmittance of the coupler and θ the angle of incidence (in air). Light is incident from the substrate side.

were entered through a 45° prism which was joined to the film substrate with matching oil as shown in the figure. The angles of incidence on the prism were approximately $\theta_1 = 55°$ and $\theta_2 = 50°$. The edge of the coupler was produced by the shadow of a knife edge in one of the exposing beams about 5 cm away from the gelatin. The size of the resulting gratings was approximately 5 mm × 5 mm.

We made several Bragg couplers and studied their characteristics at 0.6328 μm. We obtained coupling both on transmission and on reflection from the grating, but the better results (which we report below) were obtained with reflecting couplers, i.e., with light incident from the substrate side. We used a gaussian laser beam with its waist positioned near the grating. The beam was positioned at the edge of the coupler. To optimize the coupling further, we varied the waist diameter $2w_0$ and the spacing between the coupler and the waist. For our best coupler we found optimum coupling in a diverging beam with a waist diameter of about $2w_0 = 0.3$ mm and a waist-coupler spacing of 16 cm. For this case the beam diameter at the coupler was $2w = 0.6$ mm. By analyzing the m-lines[1] we found a value for the optimum coupling length which was of the same magnitude. The m-lines also indicated that our film guide supported only one TE and one TM mode.

Figure 4 shows the angular transmission spectrum obtained for TE-coupling with the optimum beam parameters. This plot is obtained

by monitoring the transmittance T through the substrate-film-gelatin sandwich for a variable angle of incidence θ. In the figure we notice a small modulation ($\Delta T \approx 0.1$) superimposed on the spectrum which is due to substrate resonances. The large dip near $\theta = 3°$ is caused by coupling into the film. The angular width of this dip is about $0.1°$. The Bragg condition is obeyed near $\theta = 6°$. The Bragg response is strongly modulated by gelatin and film resonances. It has an overall width of about $3°$. The measured reflectance of the coupler is also shown in the figure. From these data we arrived at the following power balance for optimum coupling: 3 percent of the incident light is absorbed in the substrate, film and gelatin, 8 percent reflected, 18 percent transmitted, and 71 percent coupled into the film.

Coupling to the TM film mode is relatively weaker for the above coupler because it was designed for normal incidence ($\theta \approx 0$) where TM diffraction is at a minimum.[2] The same beam parameters which produced optimum TE-coupling yielded TM-coupling of only 15 percent.

We would like to acknowledge fruitful discussions with J. E. Goell, R. D. Standley and R. Ulrich, and the technical assistance of M. Madden.

REFERENCES

1. Tien, P. K., Ulrich, R., and Martin, R. J., "Modes of Propagating Light Waves in Thin Deposited Semiconductor Films", Applied Physics Letters, *14*, No. 9 (May 1969), pp. 291–294.
2. Kogelnik, H., "Coupled Wave Theory for Thick Hologram Gratings", BSTJ, *48*, No. 9 (November 1969), pp. 2909–2947.
3. Ashkin, A., and Ippen, E., unpublished work.
4. Dakss, M. L., Kuhn, L., Heidrich, P. F., and Scott, B. A., "A Grating Coupler for Efficient Excitation of Optical Guided Waves in Thin Films", Appl. Phys. Letter, *16*, No. 12 (June 1970), pp. 523–525.
5. Brandes, R. G., Francois, E. E., and Shankoff, T. A., "Preparation of Dichromated Gelatin Films for Holography", Applied Optics, *8*, No. 11 (November 1969), pp. 2346–2348.
6. Goell, J. E., and Standley, R. D., "Sputtered Glass Waveguide for Integrated Optical Circuits", BSTJ, *48*, No. 10 (December 1969), pp. 3445–3448.