THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1 Recent Developments:

An Overview

By R. E. STAEHLER and W. S. HAYWARD, Jr.

(Manuscript received December 11, 1978)

This paper presents an overview of the Traffic Service Position System No. 1 in terms of the objectives and design philosophy of the new features that have been continuously added to the system since its initial introduction into service in the Bell System. It also serves as an introduction to the detailed technical papers that follow.

I. INTRODUCTION

Almost all of the telephones in the United States can be used to dial toll calls directly over the Direct Distance Dialing network. Nevertheless, approximately 15 percent of the total number of toll messages are placed with the assistance of toll operators. These toll calls include person-to-person, collect, credit card, hotel-originated calls, calls from coin stations, operator-dialed calls, and certain international calls. On an average business day, more than 13.1 million calls are placed through operators, and, in addition, there are over 5.2 million operator contacts for assistance and originating number identification. At year-end 1978, about 60,000 operators were employed to give this service.

1.1 Capabilities of initial design[1]

The Traffic Service Position System No. 1 (TSPS No. 1) was introduced into the field in January, 1969 (in Morristown, New Jersey) as an operator service system that could be used in conjunction with almost all Bell System local and toll switching systems to allow customers to dial their own operator-assisted calls and to relieve the operator of many tedious tasks required by cord switchboards. The basic objective of the initial system design was to automate routine

functions so that the operator could concentrate on communicating with customers, using human judgment, and establishing calls. The features of the early installations were designed to handle basic functions for coin and non-coin calls, such as automatic call timing and recording, recording of originating directory number, and recording and transmission of customer-dialed numbers. Since TSPS employs stored program control, it was intended that additional features would be added to automate an ever-increasing number of functions as advances in technology made these technically and economically viable. Typical calls not included in the initial design were conference calls, mobile and marine calls, inward to operator calls, and special handling of hotel and motel calls.

TSPS No. 1 provides each operator with a cordless, electronic console which, through indicating lamps and pushbuttons, establishes an operator-machine interface for use in automation of routine functions. TSPS No. 1 increases the number of calls an operator can handle and also quickly makes auxiliary information available which may be needed to serve the telephone customers better.

Customers benefit from these features. Because number recording and timing calculations are performed by the computer-like system, the delay of manual operation is eliminated and calls go through faster. Calls are also billed more accurately. On coin calls, initial and overtime charge calculations are instantly available.

Operating companies benefit because they can provide operator assistance more efficiently. They can, therefore, handle increased traffic levels with the same number of operators. Furthermore, they are able to manage the work force more effectively because of the improved administrative features and to have the added flexibility to locate Operator Office Groups at convenient locations.

Operators benefit by a more attractive working environment, an equitable distribution of calls to their positions, and the satisfaction of serving their customers better.

From 1969 through 1977, five additional issues of generic programs have expanded the feature content in TSPS No. 1. There are now about 135 systems in the continental United States. Over 80 percent of Bell System customers and a large number of customers served by other companies are now served by TSPS. It is estimated that, within the next three years, the Bell System coverage will exceed 90 percent, providing almost universal availability.

1.2 Additions to capability

Since, as shown in Fig. 1, TSPS utilizes stored program control, new features have been continuously and easily introduced into the system.

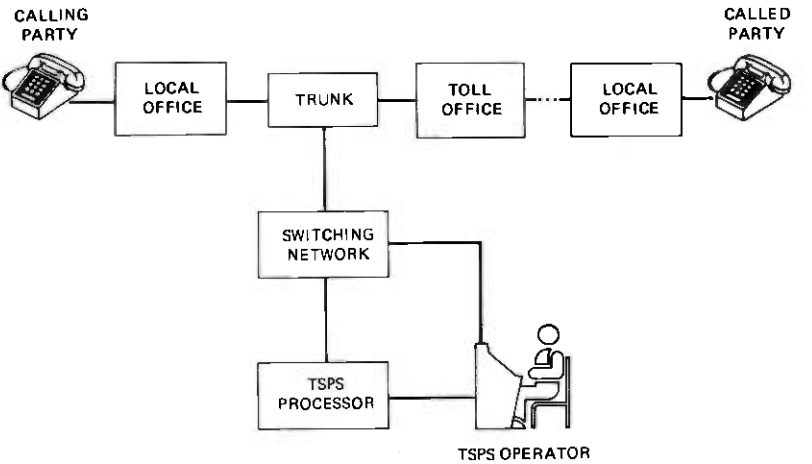


Fig. 1—Basic TSPS No. 1.

1.2.1 Hotel/motel

One of these features is the automation of hotel/motel billing. When a guest at a hotel or motel has made toll calls during his or her stay, the guest must pay the charges for these calls when checking out. These charges are now automatically computed (including taxes) for each call by TSPS. Within a few seconds, the charges are printed on a dedicated teletypewriter on the hotel/motel premises. If the hotel/motel does not have enough toll calling to justify a dedicated teletypewriter, the charges are printed at a telephone company billing center where a clerk telephones the hotel/motel and gives the charges orally—within a few minutes.

Other businesses such as convention centers, marinas, and brokerage houses may also employ this time and charge information feature to allocate telephone charges to their customers.

1.2.2 International call handling

Another of these features is the partial automation of international call handling. When a customer who is not served by an office equipped to handle International Direct Distance Dialing (IDDD) calls wishes to place an international call, he or she simply dials 0 and gives the desired international number to the TSPS operator. The operator then indicates to TSPS via key action on the console that this call is an international call and then enters the number. For a station-to-station call, the TSPS operator releases from the call, and the system automatically forwards the call to a gateway switching office, which then sends the call overseas via satellite or cable. If the call is made on a credit

card, a collect, or a person-to-person basis, the TSPS operator processes the rest of the call in the same manner as a domestic call.

In the case where a customer is served by local offices having IDDD capabilities, the customer who dials a 011 prefix followed by the country code digit and the national number will have his or her call forwarded by TSPS without the assistance of an operator. If the customer dials a 01 prefix followed by the country code and the national number, or if the customer dials 010 only, the TSPS operator will be connected to assist the customer.

1.2.3 Others

In addition, a number of other features such as time and charge quotations for non-coin (non-hotel) calls, recording of charges for directory assistance calls, and maintenance and administration improvements have been introduced into the system.

1.3 Hardware improvements

1.3.1 Semiconductor memory

The original TSPS design utilized piggyback twistor memory based on magnetic storage elements for program and call store. Advances in technology have made it possible to replace the piggyback twistor memory with a semiconductor memory resulting in major reductions in cost, space, and energy consumption.

II. MOST RECENT NEEDS

2.1 Extension to sparsely populated areas

Densely populated areas that generate a large amount of operator-assisted traffic easily justify the capital expenditures for a TSPS, but in a sparsely populated area, the toll switching center is small, and relatively few operators are needed. Savings in operating costs are not sufficient to balance the traffic-insensitive cost of the TSPS central processor. Nevertheless, the improvements provided by TSPS are extremely desirable in these sparsely populated areas. Not only can service be improved, but the difficulties of maintaining a sufficient number of operators, day and night, in small isolated teams could be eliminated.

2.1.1 Remote Trunk Arrangement[2]

To economically provide the benefits of TSPS to the sparsely populated areas, the Remote Trunk Arrangement (RTA), as shown in Fig. 2, has been developed to bring TSPS service from a distant TSPS to one of these locations.

The RTA equipment is located at a toll office in the rural area, with a concentrator network providing the talking path between the incom-

ing trunks and the TSPS operator. Data links (called peripheral control links, PCLs) connect the RTA equipment with the centrally located base TSPS processor which retains the control, logic, records, and centralized access to operators. The operator traffic from one or more RTAs is added to the traffic of a centrally located base TSPS to generate the load necessary to utilize a TSPS economically.

An improved electronic operator console (the 100C) combining training and service features has been designed to be used for both local and remote applications. This configuration is designated Position Subsystem (PSS) No. 2. The distance from the farthest RTA through the TSPS processor to the farthest Operator Office Group can total as much as 1000 miles, so that one TSPS can handle the traffic from a large geographical area.

2.1.2 Design aspects[3-5]

The articles in this issue covering the RTA design aspects include overall description, operational characteristics, transmission considerations, and hardware and software implementation.

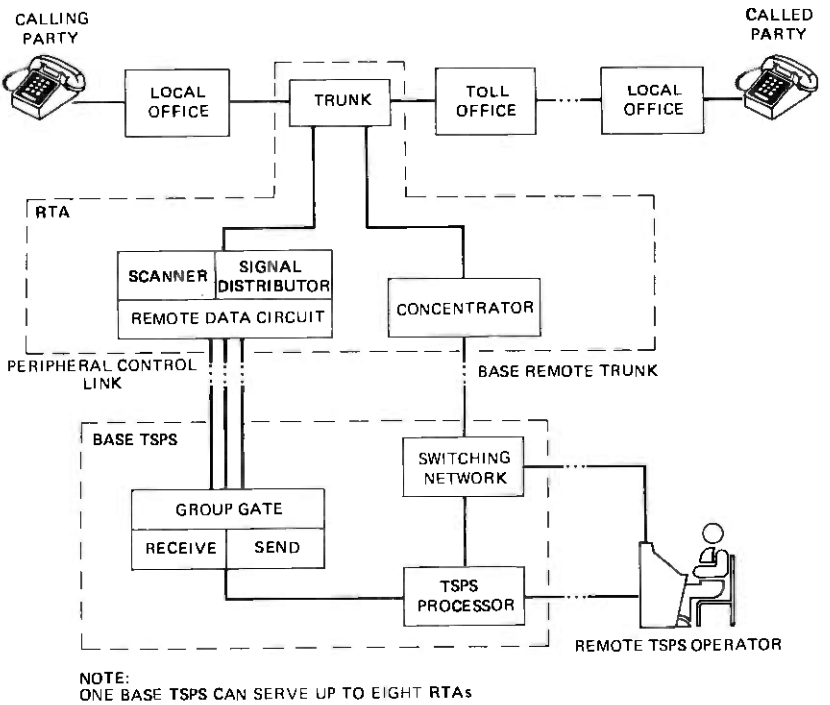


Fig. 2—Remote Trunk Arrangement.

2.2 Relieving operators of routine coin toll call functions

As advances were made in technology, it became increasingly apparent that routine coin toll call functions, such as charge quoting and coin collection, could be fully automated. Confirmation of the technical and economic viability of this approach was obtained via in-depth systems engineering studies supported by test results on laboratory feasibility models.

2.2.1 Automated Coin Toll Service

To further the benefits of TSPS, the Automated Coin Toll Service (ACTS), shown in Fig. 3, has been developed to provide TSPS with the ability to process station paid coin calls automatically, including initial period notification, overtime charge notification, and quotations on time and charges for non-coin originated calls. For communicating with the customer, additional equipment was provided with TSPS which could store digitally encoded speech segments, concatenate these speech segments into sentences, and monitor the coin deposits.

Stored program was added not only to control the new equipment in performing the coin service tasks but also for interfacing the new equipment with the existing TSPS.

2.2.2 Human factor studies[6]

The complex nature of the customer-machine interface in ACTS prompted the implementation of a three-month human factors trial in Chicago, Illinois. The trial was designed to provide ACTS-like service using TSPS operators plus a skeletonized prototype of the ACTS equipment.

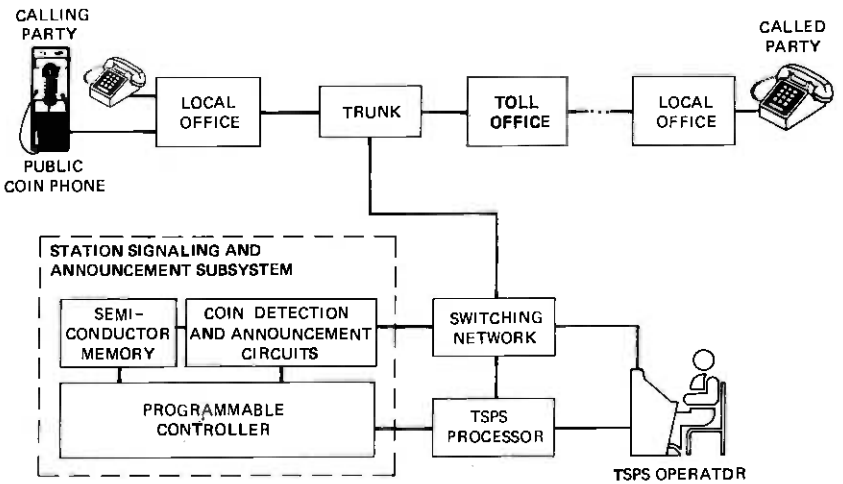


Fig. 3—Automated Coin Toll Service.

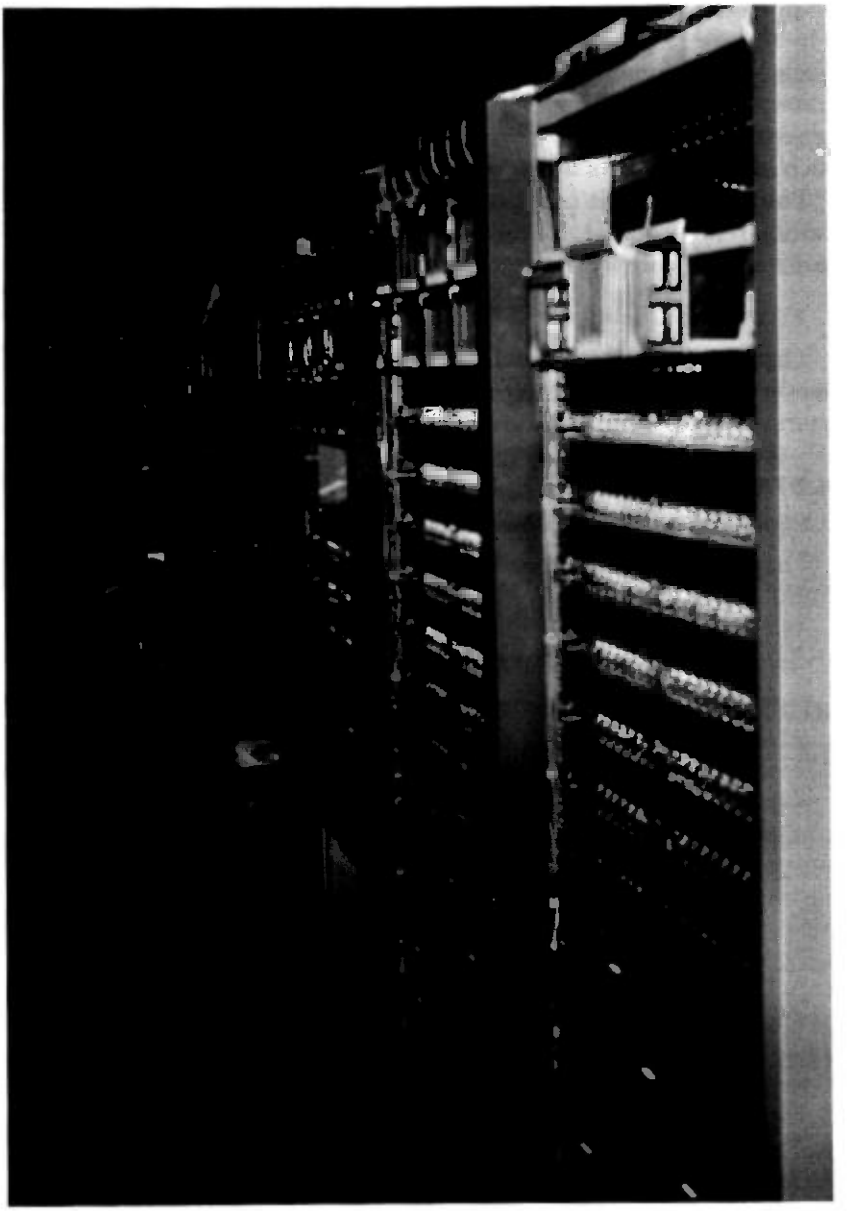


Fig. 4—Remote Trunk Arrangement in Utica, New York.

The TSPS operators were instructed to respond only to dial signals from customers and transmit the initial period length and call charges computed by TSPS to an auxiliary minicomputer via a teletypewriter. The minicomputer in turn sent instructions to the prototype ACTS

equipment to fabricate sentences (initial time interval and call charges) which were transmitted directly to the customer. The coin deposits made by customers in response to the announcements were then transmitted by the TSPS operator to the minicomputer via the teletypewriter. Subsequent prompting announcements and the timing of such announcements were controlled by the minicomputer.

The results of this trial established the basic human factors design parameters essential to the development of a system that would satisfactorily complete coin toll calls without operator assistance.

2.2.3 Design aspects

The articles in this issue covering the ACTS design aspects include details of the human factors trial, system description, operational characteristics, and hardware and software implementation.

III. STATUS

3.1 Remote Trunk Arrangement

The first RTA system was placed in service in Syracuse-Utica, New York, in May, 1976. The RTA equipment in Utica is shown in Fig. 4.

As of year-end 1978, 120 RTAs and 60 PSS No. 2s are in service, with about 20 more RTAs and 10 PSS No. 2s in various stages of installation.

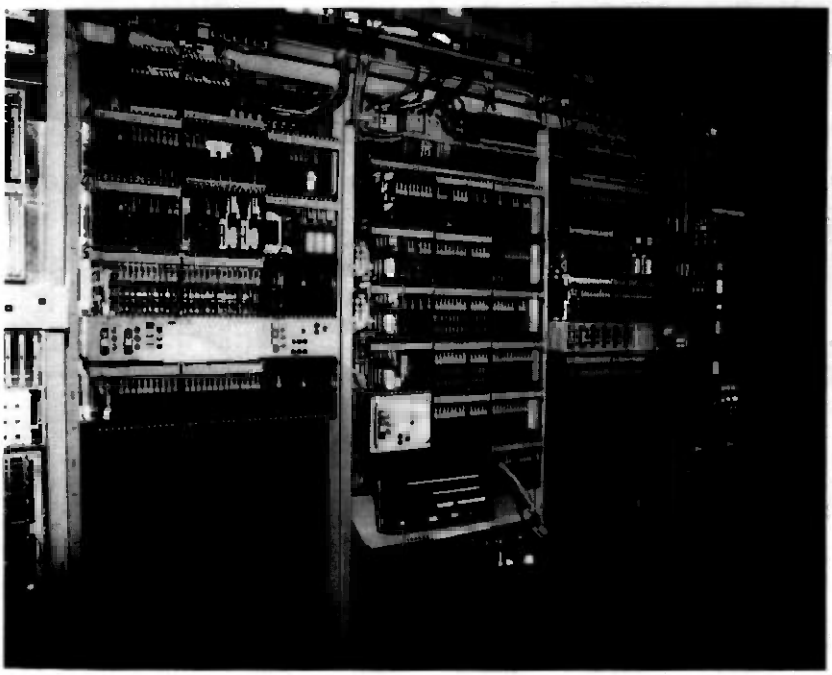


Fig. 5—Automated Coin Toll Service in Phoenix, Arizona.

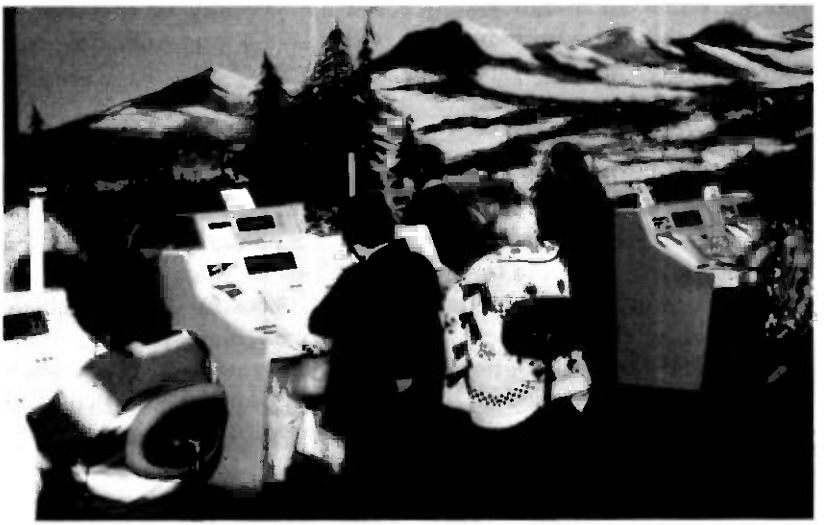


Fig. 6—An Operator Office Group in Phoenix, Arizona.

3.2 Automated Coin Toll Service

The first ACTS system was placed in service in Phoenix, Arizona, in November, 1977. The ACTS equipment in Phoenix is shown in Fig. 5.

As of year-end 1978, 3 ACTS were in service, with about 20 more in various stages of installation. By 1981, it is estimated that ACTS will be installed in all TSPS sites that have significant coin traffic, and that most of the coin paid calls in the Bell System will be handled by the ACTS equipment.

An Operator Office Group located in the suburbs of Phoenix is shown in Fig. 6.

IV. PERFORMANCE

Performance data from the RTA, PSS No. 2, and ACTS sites indicate that all design objectives have been achieved.

V. FUTURE TRENDS[13]

There are plans to exploit the stored program concept further by augmenting the operational software programs to utilize the basic ACTS equipment for handling credit card, collect, and third-number calling without operator intervention.

VI. SUMMARY

This paper has presented a general background for RTA and ACTS additions to TSPS No. 1 as an introduction to the technical papers that

follow. While all design details could not possibly be included, the papers in this issue provide a comprehensive overview of the greatly enhanced utilization of TSPS No. 1.

VII. ACKNOWLEDGMENTS

The development of these projects required the participation of hundreds of people in many organizations in Bell Laboratories, Western Electric, AT&T, and the operating companies. All the authors of this issue are indebted to these organizations for their cooperation and the team effort that culminated in the successful and on-schedule completion of these projects. The authors of this paper also wish to acknowledge the contributions of all the team members whose work is summarized here, as well as those of W. A. Depp, A. E. Spencer, Jr., M. A. Townsend, F. S. Vigilante; and J. C. Dalby, coordinating editor.

REFERENCES

1. R. J. Jaeger and A. E. Joel, Jr., "TSPS No. 1—System Organization and Objectives," B.S.T.J., *49*, No. 3 (December, 1970) No. 3, pp. 2417–2443.
2. T. F. Arnold and R. J. Jaeger, "TSPS/RTA—An Overview of the Remote Trunk Arrangement," Conference Record, *3*, International Conference on Communications ICC 75, p. 46–1.
3. S. M. Bauman, R. S. DiPietro, and R. J. Jaeger, "TSPS No. 1: Remote Trunk Arrangement—Overall Description and Operational Characteristics," B.S.T.J., this issue, pp. 1119–1135.
4. W. Brune, R. J. Piereth, and A. G. Weygand, "TSPS No. 1: Remote Trunk Arrangement Position Subsystem No. 2—Transmission and Signaling Considerations," B.S.T.J., this issue, pp. 1137–1165.
5. A. F. Bulfer, W. E. Gibbons, and J. A. Hackett, "TSPS No. 1: Remote Trunk Arrangement Hardware and Software Implementation," B.S.T.J., this issue pp. 1167–1205.
6. A. Barone-Wing, W. J. Bushnell, and E. A. Youngs, "TSPS No. 1: Automated Coin Toll Service Human Factors Consideration," B.S.T.J., this issue, pp. 1291–1305.
7. M. Berger, J. C. Dalby, E. M. Prell, and V. L. Ransom, "TSPS No. 1: Automated Coin Toll Service Overall Description and Operational Characteristics," B.S.T.J., this issue, pp. 1207–1223.
8. G. T. Clark, D. H. Larson, and K. Streisand, "TSPS No. 1: Automated Coin Toll Service Circuits and Physical Design," B.S.T.J., this issue, pp. 1225–1249.
9. R. Ahmari, J. C. Hsu, R. L. Potter, and S. C. Reed, "TSPS No. 1: Automated Coin Toll Service Software—Operational and Maintenance," B.S.T.J., this issue, pp. 1251–1290.
10. J. J. Stanaway, J. J. Victor, and R. J. Welsch, "TSPS No. 1: Software Development Tools," B.S.T.J., this issue, pp. 1307–1333.
11. J. P. Delatore, D. H. Van Haften, and L. A. Weber, "TSPS No. 1: System Verification and Evaluation Procedures," B.S.T.J., this issue, pp. 1335–1346.
12. G. Riddell, R. T. Steinbrenner, and C. R. Swanson, "TSPS No. 1: Operator Training," B.S.T.J., this issue, pp. 1347–1357.
13. E. M. Prell, V. L. Ransom, and R. E. Staehler, "The Changing Role of the Operator," International Switching Symposium, Paris, France, May, 1979.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Remote Trunk Arrangement: Overall Description and Operational Characteristics

By S. M. BAUMAN, R. S. DiPIETRO, and R. J. JAEGER Jr.

(Manuscript received December 11, 1978)

This paper is an introduction to the Remote Trunk Arrangement feature of the Traffic Service Position System. The design permits the TSPS trunk circuit, which connects the subscriber to the operator, to be located in a distant, rural location. All the logic, records, control, and centralized access to the operators remains at the base unit. The Remote Trunk Arrangement is controlled over a data link and up to eight RTA subsystems can be extended from a single TSPS base unit. The addition of the RTA feature, expansion of the geographical area served by TSPS, and the necessity for handling special operator service traffic, affected several TSPS design parameters and required the addition of several new features. These supporting features and the operational characteristics and design aims of RTA are described.

I. INTRODUCTION

The Traffic Service Position System No. 1 (TSPS) helps the operator more efficiently handle toll calls such as collect, credit card, charge to third number, and coin toll calls. Key elements of the system design were to have all calls served by a single large team of operators, to divide the large team into smaller groups of a maximum of 62 operators for administrative purposes, to permit the location of the groups remote from the base unit near good labor markets, and to achieve high concentrations of traffic to fully utilize the expensive equipment needed to handle the complex, detailed toll calls.

The initial market for TSPS was in large metropolitan areas having high traffic density and high employee turnover. As the needs of the

metropolitan areas were being met, it became apparent that the benefits of TSPS would be desirable in more sparsely populated areas for two major reasons. The first was that, because of their small size—50 position switchboards and smaller—rural operator teams are less cost effective. Moreover, there is a special need to increase efficiency in off-hours when average traffic might not require the services of a single operator. Providing the supporting personnel facilities for small groups added to the cost. The second reason for providing TSPS service in rural areas was to provide uniformity of service to all customers. An increasingly mobile population anticipates the same type of telephone service nationwide.

To provide the benefits of TSPS operation in rural areas, several approaches were investigated, including a scaled-down version of TSPS and the incorporation of TSPS features in a local switching system. The solution chosen is the Remote Trunk Arrangement (RTA). As the name implies, the design permits the TSPS trunk circuit, which provides the brief transmission bridge to the operator, to be located in a distant, rural location. All the logic, records, control, and centralized access to operators remain at the base unit. The RTA is controlled over a data link and provides all the benefits and features inherent in TSPS. Since traffic from a number of RTAs can be handled by a single base unit in addition to the traffic at the base, efficient utilization of the relatively powerful TSPS is achievable. Also, existing TSPS installations not using all the available capacity can be more fully utilized by reaching out into surrounding rural areas.

One RTA can provide up to 496 TSPS trunks which connect to the base over 64 voice circuits. Control of the RTA is over triplicated data lines. To ensure continuity of service, voice transmission and data facilities are split over diverse routings. The transmission objective of having the operator voice levels equivalent to local operator service limits the maximum allowable distance between the most remote RTA to the most distant operator to 1000 miles via the base unit. Various transmission factors establish this limit as explained in Ref. 2.

The original design of TSPS provided for remoting operator groups using a special version of T1 carrier to provide both voice and data tailored to TSPS. This arrangement has been extensively applied, but it poses a limitation to some operating companies because the maximum remoting distance is 80 miles and T carrier routes are not always available to the desired operator location. Therefore, as part of the RTA development, the data link capability for controlling the RTA was so designed that it could also control a remotely located operator group. This data link arrangement has been named the Peripheral Control Link (PCL). The voice and data circuits from the base to the operator positions can now be provided by using circuits on any standard transmission facility. Also, a new version of the TSPS console

was developed to improve the circuit control mechanisms and permit consoles to be used for both training and service. The new console is designated 100C, and the combination of the PCL with the 100C is called Position Subsystem (PSS) No. 2. Position Subsystem No. 2 can be used for both local and remote applications and supersedes the original PSS No. 1. 100C consoles can be located as far as 1000 miles from the base if no RTAs are served by the base.

Several system problems unique to the rural environment had to be solved as part of the RTA development. Closing a rural cord switchboard unit often removed the only 24-hour service unit in the area. Off-hour business service calls, repair service calls, alarms from unattended telephone offices, busy line verification calls, calls from postpay coin telephones and inward assistance calls have been traditionally handled by such switchboards. After due consideration, some of these functions are now handled by new administrative arrangements and some by new designs. A description of the RTA supporting feature designs is contained in Section III of this paper.

The RTA overall description appears in the following sections and the reader is encouraged to consult Refs. 1 and 2 for more details about the hardware and software design.

II. OVERALL DESCRIPTION OF RTA/PSS NO. 2

A simplified view of the RTA is shown in Fig. 1. The TSPS, or base unit, is shown bridging onto toll-connecting trunks between local and toll switching offices. In a similar way, the RTA bridges onto the toll-connecting trunks which home on the toll switching office serving the rural area. The RTA is controlled from the base unit by means of a data link. Voice grade connections between the RTA and the base unit allow the base unit to provide operator assistance as well as other voice frequency functions (e.g., MF digit reception and outpulsing). In this way, all traffic in the complex can be handled by a single operator team and common service circuits. Also, the substantial cost of the TSPS common control equipment is shared by all traffic in the complex.

Although Fig. 1 shows only one RTA, the system is designed to accommodate up to eight RTAs in a single complex. In this way, it has proven possible to pool the traffic of a large geographic area and in some cases to justify a TSPS/RTA complex where a single location is not a candidate for its own TSPS.

Figure 2 shows a more detailed view of RTA. The TSPS base unit is shown in the lower portion of the figure. An RTA, many miles distant, is shown at the top of the figure.

The RTA works under control of the base unit SPC processor and connects to service circuits and positions via the TSPS base unit network. At the RTA location are trunks, a switching network (the

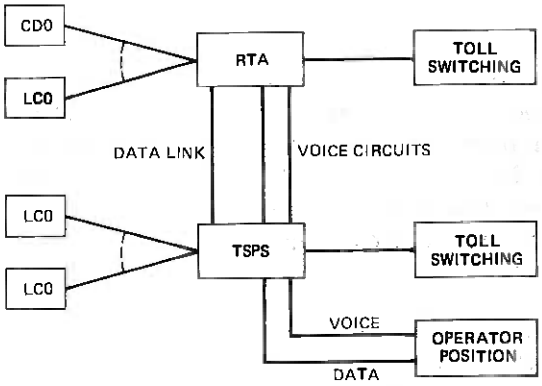


Fig. 1—Simplified diagram TSPS Remote Trunk Arrangement.

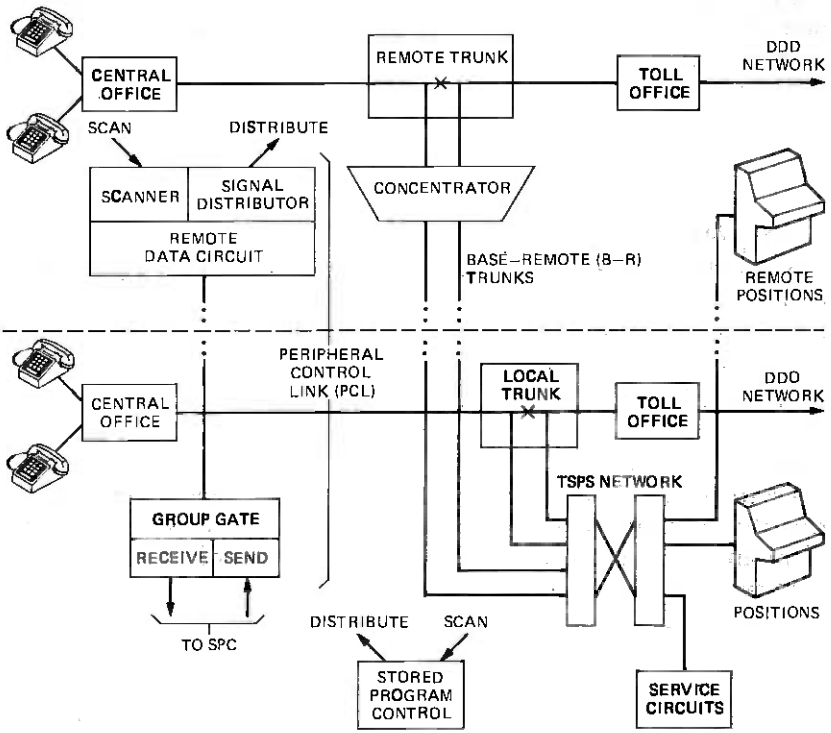


Fig. 2—Detailed diagram TSPS Remote Trunk Arrangement.

concentrator), a test frame, and scanning and signal distribution units. The RTA connects to the base unit over the Peripheral Control Link and over a number of Base-Remote (BR) trunks for voice connection to operators and service circuits.

Solid-state components are used almost exclusively for RTA circuits, and most of these consist of medium-scale TTL integrated circuits. The concentrator matrix, while controlled by integrated circuits, uses small crossbar switches. The compact physical design shown in Fig. 3 is due to this extensive use of integrated circuits. Less than one 20-ft by 20-ft building bay is required to contain an RTA of up to 496 trunks plus up to 64 base-remote trunks. Most of this space is occupied by transmission equipment and trunk frames.

The concentrator to BR trunk connections are engineered to provide a blockage of probability 0.001 to provide minimum delay in connecting operators and service circuits to calls. Since an operator is required on a connection for only a small fraction of the total call duration, a high concentration ratio can be achieved. While this ratio may vary from one site to another, 8:1 is a typical number. This is the reason for supplying a maximum of 64 base-remote trunks. One way of looking at the concentrator is that it is another stage of switching on the base network with very long links (BR trunks) between the first and second stages.

The basic traffic capabilities of TSPS remain essentially unchanged with the addition of RTA. All TSPS features are available to both TSPS

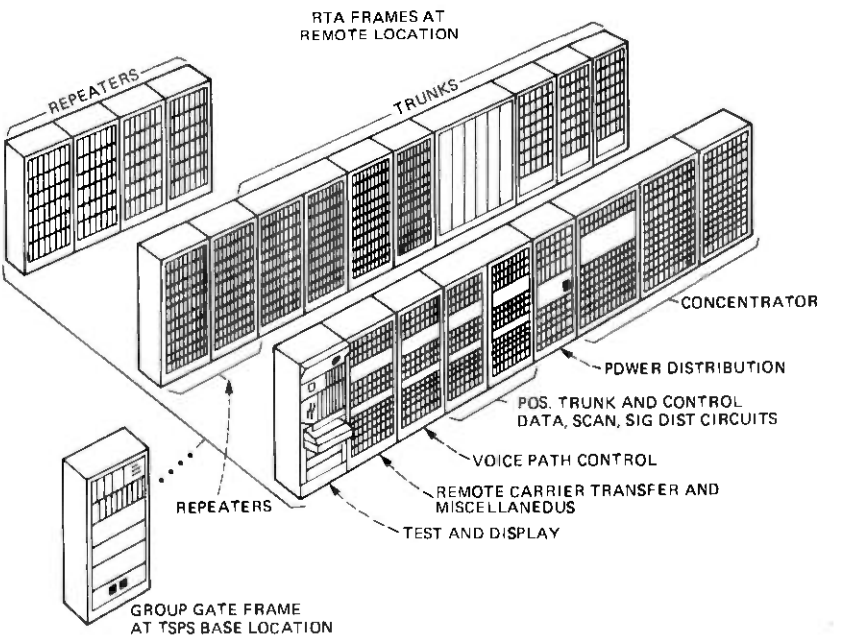


Fig. 3—RTA frames.

base unit customers and RTA customers. The customer will not be able to detect whether service is provided directly by a base unit or through an RTA.

2.1 Peripheral control link

The two-way data capability between the base unit and RTA is provided by a group of elements known collectively as the Peripheral Control Link (PCL) (Fig. 4). The PCL includes the data link itself, the circuits which interface with the data link (group gate and remote data circuit), and the scanning and signal distribution units. The data facilities are standard four-wire voice channels operated full duplex and use 2400-b/s data modems. The Group Gate and Remote Data Circuit perform extensive error detection and provide necessary parallel/serial conversions. Each message over the data link is acknowledged by the receiving end. Messages are retransmitted if mutilated in the original transmission. For reliability, data facilities are provided over two geographically independent routes. Three data lines are provided, with one line serving as a switchable spare for two active lines.

When used with an RTA, the PCL provides the ability to control, via

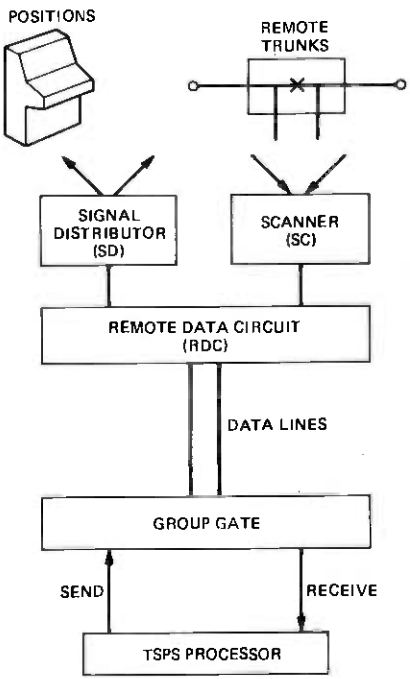


Fig. 4—Peripheral control link.

the signal distributor, the RTA trunk circuits and concentrator. By means of the scanner, the PCL can monitor trunks and report changes of state. This control and monitoring, together with the data communication function of the PCL, provides a set of general capabilities. For the PSS No. 2, the signal distributor and scanner of the PCL are used to control console lamps and detect operator keying actions.

The Position and Trunk Scanner (PTS) is located at the remote end of the PCL. Its function is to autonomously scan key reports from PSS No. 2 positions and DP pulses and supervisory state information from trunks in RTA. This information is reported to the SPC 1A over the data link. The trunks are scanned every 12.5 ms by the PTS which compares this information with last-look data in its memory. For base trunks, timing of supervisory changes of state for the purpose of detecting flashes and disconnects, and filtering out hits (a spurious change of state which must be ignored) is under program control. However, for remote trunks, since delays in transmission of successive changes of state over the data link could distort the relative timing, hit detection and filtering is done for RTA by the PTS. This is accomplished by demanding that changes in supervisory state of trunks persist for at least two successive scans before reporting them to the base unit via the PCL.

The Signal Distributor (SD), located at the remote site, is responsible, for distributing orders to remote trunks, the concentrator, test buffers, a test frame, and to 100C positions. The SD has several features designed to make efficient use of the data link. Commonly used groups of orders from the base to the remote location for control of positions are combined. Hence, one order transmitted from the base has the effect of executing many orders at the remote end. For example, there is a release order which causes all lighted lamps at a 100C position to be extinguished. This saves much data link capacity. A second design feature allows all trunks at an RTA to be initialized with one order. Since there can be as many as 496 trunks at an RTA, this could save as many as 1487 relay orders (three orders per trunk) over the data link. This feaure is used for system initialization. Customer conversations at the time of initialization must be preserved. Therefore, provisions were made in the hardware such that trunks in the talking state are not initialized.

2.2 Circuits controlled by the PCL

The RTA concentrator is controlled from the base unit by call processing software through the PCL signal distributor. The small crossbar switches are arranged in a two-stage network.

The RTA trunk circuits generally provide the same capabilities as the TSPS base unit trunk circuits. However, because they are controlled

over the data link, they differ significantly in design from base unit trunks. They have greater logic capability for sending coded flashing signals and for timing incoming flashing signals. They also have greatly augmented features for testing of the trunk units. With the use of integrated circuits and miniature relays, these complex trunks are similar in size to the functionally simpler TSPS trunks. Potential distortion of the relative timing between consecutive orders during periods of data link congestion necessitated the design of remote trunks with a unique feature. All remote trunks have the ability to generate from one to five winks from a single command. Winks of precise time duration are needed for coin station control using multiwink signaling.

As described earlier, the PCL has also been designed to function with new operator position equipment. The new 100C console contains the necessary electronics to decode information transmitted from the signal distributor, to light or extinguish the appropriate lamps, and to remember the state of all the lamps on the position. The position electronics also detects operator keying actions and provides data to the PCL scanner in appropriate form. Only a few pairs of wires are necessary to connect a position console to the PCL. Thus, the addition of positions to a position subsystem is a relatively simple matter.

The RTA/PSS No. 2 fault recognition and diagnostic programs make use of three maintenance circuits controlled by the PCL. The maintenance buffers are used to set and reset circuit states for reconfiguration and testing purposes. A Diagnostic Controller Circuit is activated over the PCL and used to execute sequences of RTA/PSS No. 2 diagnostic tests stored within its read only memory (ROM). Failing test results are reported back to the base via the PCL and formated for TTY output to the craft. The third maintenance circuit under PCL control is the Test and Display Circuit. This circuit provides the craftsforce at the remote site with a manual interface for control of trunk and position transmission test equipment. It also provides a status display of PCL and controlled circuit equipment.

2.3 Peripheral control link configurations

The RTA/PSS No. 2 system design provides for up to 15 peripheral control links to a base TSPS. There is a maximum of 8 RTA subsystems per TSPS and a maximum of 8 operator groups per TSPS.

The RTA and PSS No. 2 subsystems were each designed to operate when located at a 1000-mile maximum distance from the TSPS base unit. However, for proper transmission of voice and tones, a 1000-mile limit is applied to the sum of the farthest RTA distance to the base plus the farthest PSS No. 2 distance to the base. In other words, if a PSS No. 2 is located in a town 400 miles from the TSPS base and it is the farthest remote group, then the RTA subsystems served by the same TSPS base would have to be within a 600-mile radius of the TSPS base.

2.4 Hardware technology employed for RTA/PSS No. 2

The RTA/PSS No. 2 control circuitry design was one of the early Bell System applications of the Transistor-Transistor-Logic (TTL) integrated circuit technology. The Position and Trunk Control frame, at the remote end of the PCL, also consists of scanning circuitry using opto-isolators and diagnostic circuitry using instructions stored on existing state-of-the-art read-only-memory (ROM) devices.

The 100C positions provided with PSS No. 2 were the first large volume application of the 7-segment light-emitting diode (LED) numeric display. This new display technology was also made available for retrofit into the 100B position provided with PSS No. 1.

The RTA/PSS No. 2 features greatly extended the transmission range of TSPS. To permit such great distances between operators and customers and still maintain good voice quality required the simultaneous development of three key pieces of equipment: the Unified Telephone Circuit (4251 B network), the 1P precision hybrid, and a new three-way/four-wire bridging repeater. Each of these is discussed in Ref. 3.

The Logic Analyzer for Maintenance Planning (LAMP) was used extensively for the generation of circuit-board test-vector information provided to Western Electric. Results were also used to modify board designs to improve test access.

A minicomputer-controlled fault insertion system was developed to automate the large and repetitive task of physical fault insertion for Trouble Locating Manual (TLM) diagnostic data generation. This new system also made available timely feedback to the diagnostic designers about potential problems with diagnostic coverage and resolution.

III. RTA SUPPORTING FEATURES

The addition of the RTA feature, expansion of the geographical area served by a TSPS, and the necessity for handling special operator service traffic, affected several TSPS design parameters and required the addition of several new features. For example, with RTA, the maximum number of states and Numbering Plan Areas (NPAs) served by TSPS were increased from three to eight. In addition, the maximum number of toll switching offices with unique routing patterns serving a TSPS complex was increased from three to eight. When advancing a call to the toll network, TSPS must sometimes change the dialed or keyed digits according to a set of code conversion rules. These rules had to be changed to permit the advancement of incoming, inward, and INWATS calls through TSPS or RTA toll offices located in more than one NPA.

As mentioned in Section I, as TSPS service is extended to a wider area and TSPS capabilities are broadened to handle a wider range of call types, cordboard operations which remain solely to handle Special

Operator Services Traffic (SOST) become less and less efficient. This creates an impetus to handle more of the SOST traffic on TSPS, with the ultimate aim of eliminating the cordboard operation. Among the SOST items that were handled on cordboards, but can now be handled by TSPS, are: inward traffic and postpay coin service. In addition, small central offices connecting to TSPS require compatible trunk equipment. Multiwink coin signaling, an economically attractive method to provide coin signaling to these small offices, was also provided in TSPS/RTA.

3.1 Inward

Inward calls occur when an operator at the originating end of a call requires the assistance of an operator at the terminating end (called the inward operator) so that the latter may perform certain functions that the former cannot.

Examples of inward calls and the functions traditionally performed by the inward operator are:

(*i*) Verification—the inward operator verifies that a line is busy.

(*ii*) Calling hard to reach numbers—the inward operator tries another route for the call where previous attempts have been unsuccessful.

(*iii*) Call-back to coin stations—the call rating, computing, and recording functions are performed by the inward operator.

(*iv*) Call back (to a noncoin station) with time and charges (T&C)—the inward operator performs call rating, computing, and recording functions, and gives the T&C quote to the called party (who *originally* initiated the call).

(*v*) Call back to hotel—the inward operator performs call rating, computing and recording functions, including identification of the hotel guest, and provides the quote of charges to the hotel.

The inward feature allows a TSPS operator to act as the inward operator on an inward call. Figure 5 shows how inward calls are routed to a TSPS operator acting as the inward operator. To reach an inward operator, the originating operator, whether at a cordboard or at a TSPS position, would key forward (NPA)+(TTC)+OC, where NPA+TTC is the routine code to the Terminating Toll Center (obtained from a routing guide as a function of the NPA+NXX portion of the number to be reached by the inward operator). The numbers in parentheses (i.e., (NPA)+(TTC)) are not always present, their presence or absence being determined by the route of the call. The abbreviation OC, which stands for Operator Code, is a 3- to 5-digit number which specifies the nature of the service required from the inward operator. The Operator Codes used, together with their associated call types, are shown in Table I.

The Terminating Toll Center (TTC) recognizes inward traffic from the operator code which it receives, i.e., the OC keyed by the originating operator. Once it recognizes an inward call, the TTC will seize an inward

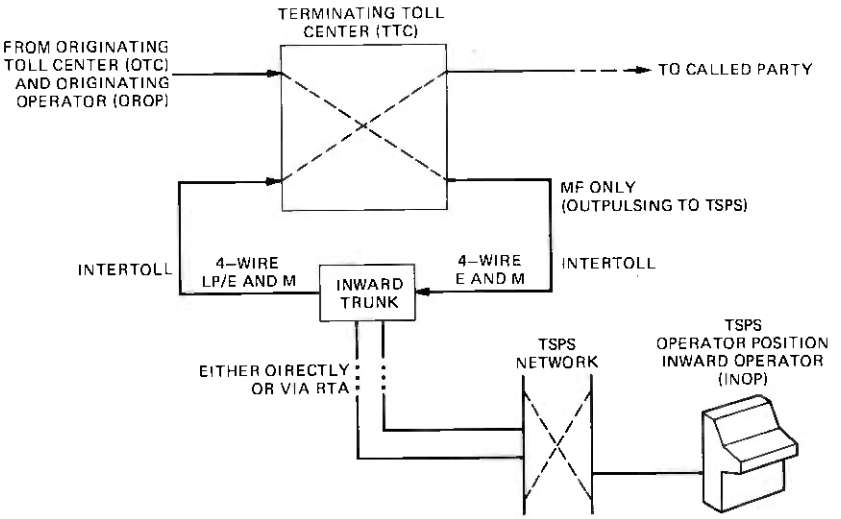


Fig. 5—Routing of inward calls to TSPS INOP and to the called party.

Table I—Operator codes and associated inward call types [originating operator keys (NPA)+(TTC)+Operator Code (OC)]

Operator Code	Inward Call Type
121	General inward, hard-to-reach, verification
1150(1)*	Collect or call-back to coin
1155(1)*	Time & charges (T&C) call back
1156(1)*	Hotel call back

* Additional 1 is local practice option.

trunk to the TSPS serving its toll traffic, and will outpulse the OC to that TSPS. From the OC, the TSPS determines the initial display to be given to the inward operator. The actions taken by the TSPS operator upon arrival of an inward call at the console will depend upon the nature of the inward call.

Inward trunks can be installed at RTAS as well as at a base unit. In either case, as shown in Fig 5, the inward trunk connects on its incoming side to a toll office outgoing trunk, and on its outgoing side to a toll office incoming trunk. The facilities on both sides of the inward trunk are classified as intertoll facilities for transmission purposes. Both base TSPS and RTA inward trunks are four-wire and are used with a three-way, four-wire bridging repeater. E&M signaling is used on the incoming side. The outgoing side can be arranged for either loop or E&M signaling.

Inward trunks may be organized in one trunk group for every toll

office serving the TSPS complex (base unit and RTAs). A single trunk group carrying all inward traffic routed to the TSPS or RTA through a given toll office will optimize the trunk efficiency. Alternatively, having separate trunk groups for each toll office serving the TSPS complex optimizes the grade of transmission on inward calls. Most desirable is the centralization of inward trunks at the highest class toll switch (usually the one serving the base unit) in the TSPS/RTA complex. This provides both good trunk efficiency and the ability to meet or exceed transmission performance objectives.

3.2 Postpay coin service

In postpay coin operation, the coin stations do not have coin collect and return functions. Once a coin is deposited, it is not retrievable. Therefore, collect and return signals are ineffective and are not used in postpay coin operation. Thus the trunks carrying this traffic from local offices to cordboards are simpler than those used in prepay coin operation.

Since the operator cannot return coins on a postpay coin call, a deposit for the initial period is not requested until:

(*i*) The called party answers and

(*ii*) It is determined that the correct called number has been reached.

For collection of overtime charges, there is no difference between postpay and prepay coin operation except that incorrect or questionable deposits cannot be returned.

Any of three methods may be used to bring postpay coin traffic into TSPS from a local office. Under normal conditions (no ANI failure), the first two methods result in automatic identification of postpay coin calls. The third method requires operator intervention to recognize such calls. The three methods are as follows:

(i) Dedicated Postpay Coin Trunks—The only calls routed to TSPS on such trunks are postpay coin calls.

(ii) Combined Postpay Coin-Noncoin Trunks with ANI Screening Identification—This technique is used to identify the kind of call when the local office has ANI. The technique consists of searching the coin band tables in TSPS for the received ANI number to determine whether or not that number corresponds to a coin station.

(iii) Combined Postpay Coin-Noncoin Trunks with Service Tone Identification—For such trunks, the TSPS operator can identify postpay coin calls by the presence of a service tone generated by the local office. The tone is heard shortly after the zip tone when the call is first brought the position. Absence of the service tone signifies a noncoin call.

3.3 Multiwink coin signaling

The Bell System Coin Service Improvement Committee recommended that multiwink signaling coin control be made available to provide additional coin control signals. With TSPS-RTA, BR trunk connections to coin control circuits can be eliminated with multiwink signaling, and multiwink is more economically provided than inband coin control for certain local offices.

The multiwink signaling format employs a series of one to five supervisory on-hook winks from TSPS to the local office outgoing trunks. The signals and their use are shown in Table II.

IV. ADDITIONAL TSPS FEATURES ON THE RTA GENERIC

In parallel with the RTA development, other service and maintenance features were developed. These features were made available with RTA. This section briefly describes those additional features.

4.1 Selective call screening

TSPS is required to provide for selective screening of incoming traffic that was previously handled by cord switchboards. The ability of TSPS to screen calls from various incoming trunks enables the telephone companies to offer institutions and businesses (e.g., hospitals, military installations, press and media, prisons, and university dormitories) the option of restricting the type of charging permitted for toll calls originating over some or all of their lines. Some hospitals, for example, do not want to expend the administrative effort required to collect telephone charges for toll calls from hospital patients. Hospitals could receive this kind of service from cord switchboard operators since the trunk appearance on the cordboard was labeled as hospital and the operator responded accordingly. Given the new ability for TSPS to restrict the kinds of charges allowed, the use of coinless public telephones is also made possible. Coinless public telephones can only be used to place credit card, collect, and bill-to-third-number calls. Coin-

Table II—Multiwink signaling system

Number of On-Hook Winks	Function
1	Operator release
2	Operator attached
3	Coin collect
4	Coin return
5	Ringback

less public telephones are especially attractive to operating companies in places such as major transportation centers to reduce the number of expensive coin stations.

Another feature provided with screening is the charge quotation feature. It allows businesses (such as law firms) to get voice or automatic quotation of telephone charges on calls by account numbers similar to the TSPS hotel/motel feature.

TSPS performs the following basic functions in handling a screened call:

(*i*) TSPS determines if the line is a screened line. This is done with information present in the trunk software register indicating the trunk group as screened or by the presence of that line number in screening line number tables.

(*ii*) If the line is screened, TSPS identifies what type of screening is applicable. A table, called the screening type table, indicates the specific combinations of restrictions allowable. Each line number in the screening line number table has an index to the applicable combination of restrictions in the screening type table.

(*iii*) TSPS also validates the operator's actions. Since the screening restrictions are in memory, the acceptability of the class of billing entered by the operator is checked. If the operator enters an unacceptable class of charge, the system informs the operator by flashing the appropriate key/lamp. The operator then interacts with the customer in an effort to obtain acceptable billing.

(*iv*) Once an acceptable class of charge is entered, a unique mark is placed on the Automatic Message Accounting (AMA) tape indicating that the call was a screened call.

Any of three methods may be used to bring screened traffic into TSPS from a local office:

(*i*) Mixed traffic (combined screen/nonscreen).

(*ii*) Dedicated screened traffic.

(*iii*) ANI 7 information-digit-identified screened traffic.

In the mixed traffic case, traffic with screening conditions and traffic without screening conditions will originate over the same trunks. A search is conducted using the calling digits provided by Automatic Number Identification equipment (ANI) to determine if a screening condition exists on that particular line. If the calling number is not in the search tables, it is assumed that the line does not have a screening condition on it. If an ANI failure occurs, Operator Number Identified (ONI) digits are used for the search.

In the dedicated traffic case, a trunk group is dedicated to screened traffic. If local offices without ANI require screening, dedicated trunks must be used to deter fraud. If the trunk belongs to the dedicated screening trunk group, the calling number obtained by the operator is

checked against the screening line number tables, as in the mixed traffic case, to determine the specific restrictions.

In the case of calls from an electronic local office, an ANI 7 information digit is supplied for screened calls. The ANI information digit is denoted I in the following ANI sequence:

KP-I-7D-ST, where 7D is the 7-digit calling number.

If the ANI information digit or any part of the ANI sequence is destroyed or not received at TSPS, ONI digits are obtained and a search conducted as in the mixed traffic case. Given good ANI, the calling number is then checked against the line number table to determine the specific restrictions.

4.2 More stores/bus (MORSTR)

More stores/bus (MORSTR) is an all-software feature which expands the Stored Program Control (SPC) No. 1A store maintenance structure to cover all expected piggyback twistor/semiconductor memory configurations. It removes the memory capacity limitation imposed by the previous maintenance software (20 stores/bus).

The maximum capacity of the SPC memory was limited by the maintenance software and by processor addressability. It was estimated that the larger TSPS systems with a combination piggyback twistor (PBT) and semiconductor memory stores would exceed the memory capacity limitation imposed by the maintenance software (20 stores/bus) upon advance to subsequent generics. The MORSTR feature was developed to provide maintenance capabilities for a maximum of 24 stores/bus. The SPC store maintenance software with MORSTR can support the various PBT/semiconductor memory combinations up to a limit imposed by the processor with the restriction that no more than 18 of the maximum 24 stores can be of the PBT type.

V. RTA CHARACTERISTICS REQUIRING SPECIAL DEVELOPMENT EMPHASIS

5.1 Necessity for more stringent transmission requirements

The original TSPS base system served by the Position Subsystem (PSS) No. 1 presented a transmission environment confined and simplified by the T1 carrier. The T1 carrier transmission limit meant that remote PSS No. 1 locations were relatively close to the toll office. With the introduction of RTA and PSS No. 2, a complex set of transmission considerations arose due to the large range of subsystem site locations allowed by the peripheral control link design. The TSPS transmission plan was revised to provide engineering and maintenance documentation for TSPS in the new RTA/PSS No. 2 environment.

5.2 Maintenance of unattended remote offices

The design intent for the RTA/PSS No. 2 feature is to provide for the primary maintenance of RTA and PSS No. 2 subsystems by a craftsforce situated at the base TSPS location. Upon identification of a fault in the remote equipment, remote site personnel would then be dispatched to replace or repair the faulty equipment. This requirement to provide base-located maintenance control reflects the general trend to more centralized maintenance in the Bell System.

Maintenance software design for the RTA/PSS No. 2 feature was strongly influenced by this trend. It was complicated by the need to provide also for those operating company situations in which more direct maintenance control was to be exercised by craftspeople situated at the remote locations.

The normal maintenance strategy associated with such items as teletypewriter interactions and visual status displays takes on an added dimension of difficulty when complicated by the introduction of multiple distant equipment sites.

More detail about the maintenance strategy chosen is provided in a subsequent article in Ref. 1.

5.3 Rapid RTA/PSS No. 2 installation buildup

The RTA and PSS No. 2 features are very attractive to operating telephone company planners because of the much increased flexibility for operator team location, the economic and service gains achieved with the replacement of small toll office cordboards, and the ability to prove in new TSPS base offices for cities of moderate size. The early demand was high, and the resulting pressures on Western Electric to supply that market were evident in the final stages of development and deployment.

Due to the efforts of many departments involved with manufacture, testing, documentation, installation, and support at the Western Electric, Columbus works and the regional offices, that heavy initial demand was satisfied. One year after the initial RTA and PSS No. 2 cutover at Utica, New York, on May 16, 1976, a dozen peripheral control links were in service and nearly one hundred had been shipped to the field. By year-end 1978, 120 RTA subsystems and 60 PSS No. 2 groups had been placed into service.

VI. ACKNOWLEDGMENTS

Many of our colleagues have made substantial contributions to the development of the RTA. R. G. Crafton and A. N. Daudelin provided system engineering support and economic analysis. A. Friedes, E. E. Hanna, and S. Lederman provided many of the development requirements and worked to ensure that the requirements were thoroughly

implemented. D. J. Eigen provided design requirements for the selective call screening feature and contributed to the software design effort. G. E. Fowler tirelessly worked to develop call-processing software and later participated with A. W. Robinson and J. J. Serinese in the field test of RTA at Syracuse and Utica, New York.

REFERENCES

1. A. F. Bulfer, W. E. Gibbons, and J. A. Hackett, "TSPS No. 1: Remote Trunk Arrangement Hardware and Software Implementation," B.S.T.J., this issue, pp.1167–1205.
2. W. Brune, R. J. Piereth, and A. G. Weygand, "TSPS No. 1: Remote Trunk Arrangement Position Subsystem No. 2 Transmission and Signaling Considerations," B.S.T.J., this issue, pp.1137–1165.
3. Ibid.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Remote Trunk Arrangement and Position Subsystem No. 2: Transmission and Signaling Considerations

By W. L. BRUNE, R. J. PIERETH, and A. G. WEYGAND

(Manuscript received December 11, 1978)

Transmission and signaling implications are important for the successful introduction of the Remote Trunk Arrangement (RTA) and the Position Subsystem (PSS) No. 2 features into the Traffic Service Position System (TSPS), since both features involve the possibility of substantial distances between the TSPS operators and the customers. Consequently, a new transmission plan including objectives for inserted connection loss (ICL), received noise, balance, and operator sidetone was established and some new transmission equipment was provided. The overall objectives, general and specific considerations, maintenance facilities, signaling implications, and new transmission equipment are described. Emphasis is placed on those aspects that are novel or newly introduced as the result of RTA and PSS No. 2.

I. OVERVIEW

An RTA or PSS No. 2 being located at a remote site several hundred miles distant from the base TSPS presents an entirely new set of transmission and signaling implications. First, there is the peripheral control link or PCL, which extends the control capability of the TSPS processors to the remote locations. Then there are transmission considerations within the RTA and PSS No. 2, modifications required to PSS No. 1, the introduction of new transmission equipment, signaling ramifications, and finally the need for additional impedance balancing.

1.1 The peripheral control link (PCL)

Figure 1 shows the PCL which includes, among other components,

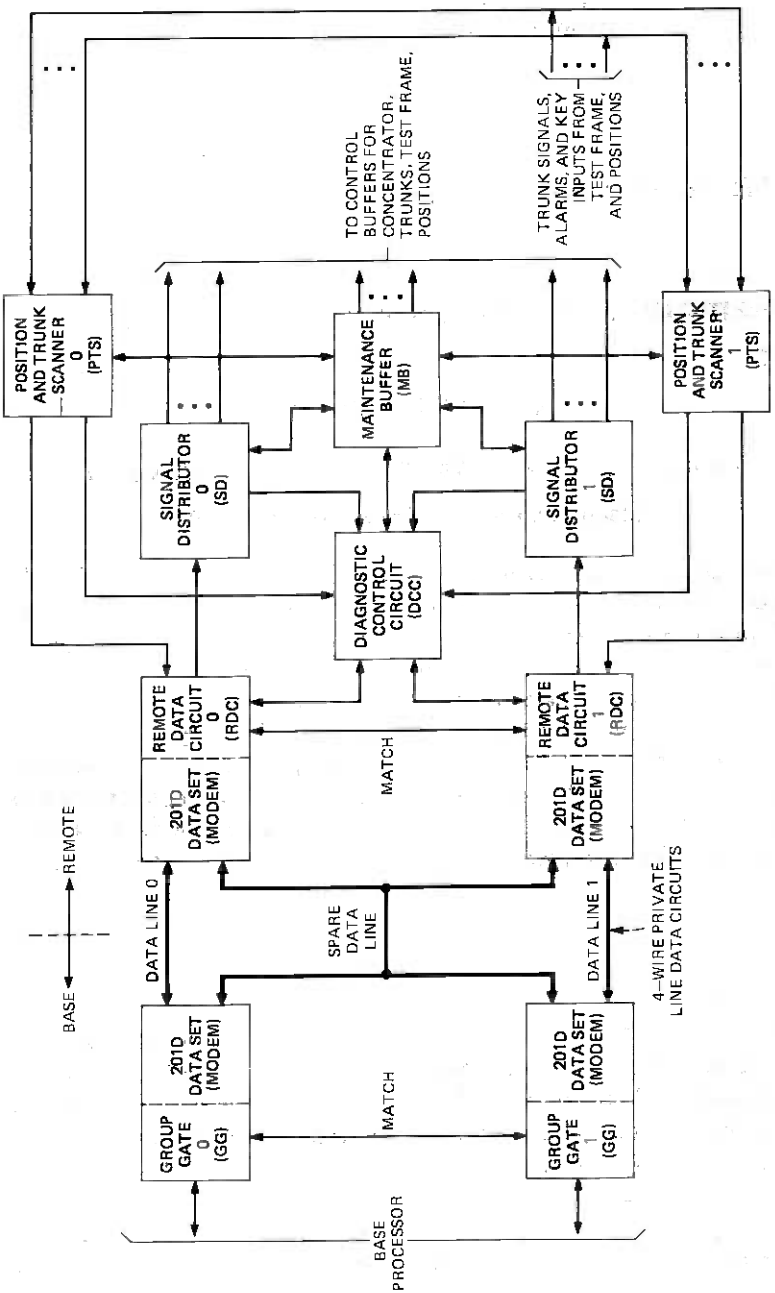


Fig. 1—Peripheral control link.

three data lines used to transmit and receive information to and from the remote sites. System requirements dictate that any standard intertoll grade transmission facilities may be used for all circuits between the base and the RTA or PSS No. 2. Hence, the data transmission circuits must also meet that constraint. Accordingly, the PCL employs 2400 baud modems with the data lines treated as unconditioned, 4-wire private line circuits. This moderate data rate coupled with distances of up to several hundred miles between modems requires that the round-trip delay time required to transmit an instruction to a remote site and receive back an acknowledgment must be limited and must be considered in the design of the data lines.

For reliability, another system requirement specifies that the transmission facilities between base and remote follow at least two diverse routes with one of the two primary data lines in each route. Since, under normal operation, data received on these lines are compared or "matched," a limit must also be placed on the delay differential; that is, on the difference between the delay times of the two routes. Both the delay differential and the round-trip delay times cannot be so restrictive, however, that they substantially restrict the choice of alternate routes between the base and the remote units or severely limit the length of the voice transmission facilities.

1.2 The remote trunk arrangement (RTA)

At the RTA, connection to a TSPS operator is made by bridging onto the toll connecting trunks (TCTS) in a manner similar to that employed at the base. This is shown in Fig. 2. Bridging 900-ohm, 2-wire trunks is accomplished by a simple tap that presents a 450-ohm impedance (two 900-ohm terminations in parallel) to the RTA network (concentrator). Bridging 4-wire trunks requires an associated 3-way, 4-wire bridging repeater which converts the 4-wire bridged tap to 2-wire and which also presents a nominal impedance of 450 ohms to the concentrator. Conversion of 4-wire circuits to 2-wire is always required, since both the RTA and the base TSPS switching networks (RTA concentrator and TSPS trunk and position link circuits) are 2-wire.

The RTA concentrator connects the bridged TCT to a base-remote (BR) trunk, typically a 4-wire intertoll grade facility up to several hundred miles long. Two-wire to 4-wire conversion at both ends of the BR trunk is accomplished by 24V4-type repeaters using a 900-ohm, 4-wire terminating set (4WTS) at the base and a high-impedance 4WTS at the RTA. Finally, the connection is completed through the base network to an operator's position, either a PSS No. 1 or a PSS No. 2.

If the performance of this connection as a function of the length of the BR and operator position trunks is examined, two types of transmission degradation are encountered. The first is echo, as detected by

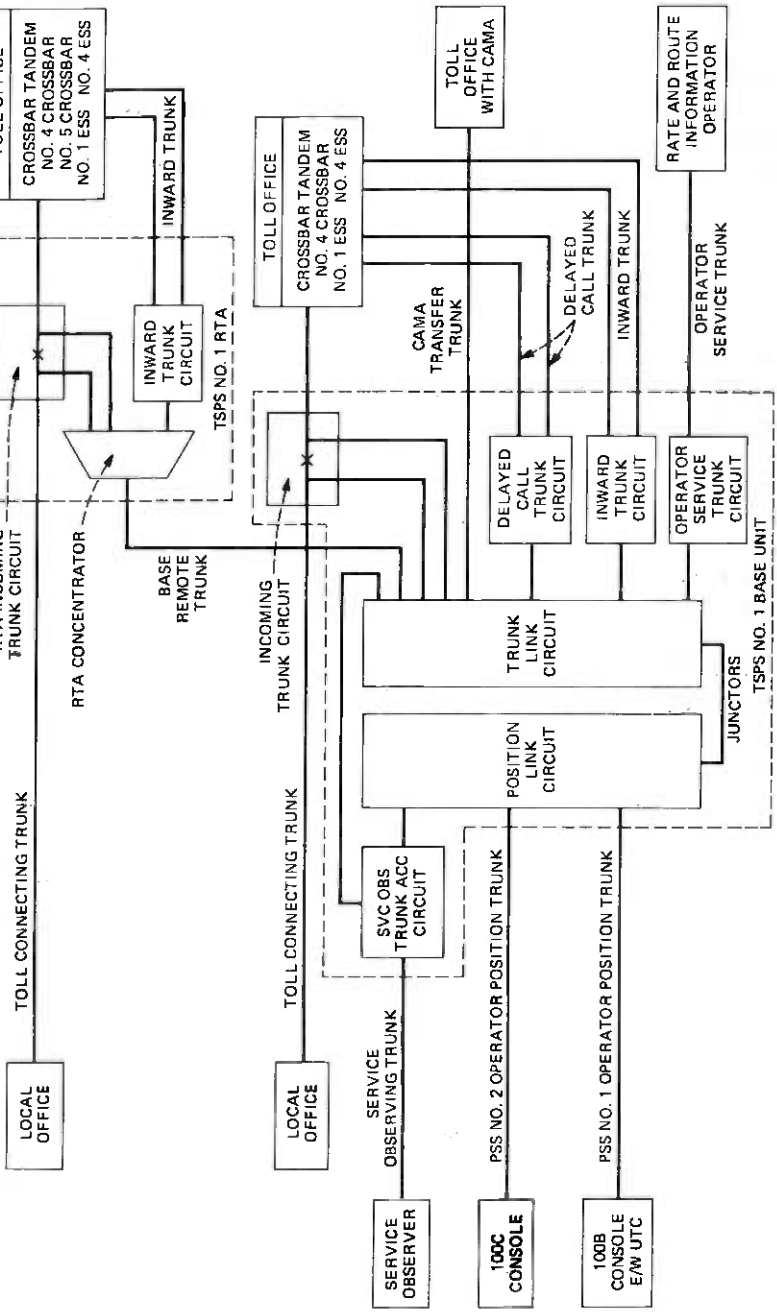


Fig. 2—TSPS No. 1—voice transmission configuration.

the customer(s) and the operator, due to the impedance mismatch at the extra switching point introduced at the RTA concentrator and accentuated by the long BR (and operator position) trunks. As discussed later, echo is controlled through impedance balancing and by a new operator's telephone circuit. The second form of transmission degradation is noise introduced by the long facilities and controlled by the use of compandors on analog facilities or by the use of digital facilities.

1.3 The position subsystem (PSS) No. 2

PSS No. 2 was devised to eliminate the constraints of the original position arrangement (PSS No. 1), which uses unique D1C channel banks and T1 carrier as the transmission facilities for remote installations, and thus is limited in range to approximately 80 miles. Conversely, PSS No. 2 may use any toll grade carrier system and is limited in range only by the noise introduced by the transmission facilities. Hence, received noise is the limiting parameter in the permissible distance between operators and RTAS when standard message grade circuits are used.

The 4-wire operator trunk circuits are connected to the TSPS switching network via 24V4 repeaters and high-impedance 4WTS in the same way as BR trunks at the RTA concentrator. This arrangement permits the PSS No. 2 circuits to be connected to the bridged base TCTS or to BR trunks. At each operator's position, new 4-wire telephone networks provide adjustable sidetone, automatic gain control (AGC), and a voice switched attenuator (VSA).

1.4 The position subsystem (PSS) No. 1

It is necessary and desirable to add RTAS to existing TSPSS which have the original position subsystem arrangements, now designated PSS No. 1. This arrangement employs the equivalent of an unbalanced 4WTS at the base unit end of the position trunks to return a fixed portion of the operator's transmitted signal which serves as sidetone to the operator. No VSA or sidetone adjustment is provided. When an RTA is added, however, the possibility of additional echo from the connection at the RTA makes this an unsuitable arrangement. Hence, the PSS No. 1 unbalanced 4WTS is replaced with a precision balanced 4WTS, and the position is equipped with the same telephone network used in the PSS No. 2. No changes are required in the PSS No. 1 when only PSS No. 2 is added to an existing system.

1.5 Transmission balance

With the introduction of RTA and PSS No. 2, the geographical extent of a single TSPS was transformed from a rather confined area around

a toll office to a system that might cover complete states or even larger areas. Under these circumstances, the distance a call must travel to the operator access point in the network may be significant compared to the length of the average customer-to-customer connection. In those cases, it becomes necessary to treat the operator connection as a toll facility. Toll facilities are characterized by trunks and switching networks which are low loss and transmission balanced and, indeed, these are exactly the requirements that have been introduced into TSPS. Transmission "balance" means matching impedances at all 2-wire to 4-wire junctions to maintain a loss design that allows accurate summation of the losses of the parts while at the same time preventing unwanted signal reflections or "echos" from mismatched junctions.

In the past, balance was not required for TSPS equipped with PSS No. 1 positions, and such systems still do not require balancing. However, whenever a TSPS is equipped with an RTA or is equipped with a PSS No. 2 in which the operator trunks exceed 200 miles in length, the entire system must be balanced.

To facilitate balancing, adjustable network build-out capacitors (NBOCs) are provided by all the 4WTSs or equivalent used in the system. Drop build-out capacitors (DBOCs) are provided or specified at the 2-wire line (or drop) anywhere there is a 4-wire/2-wire transformation on the trunk-link side of the TSPS network.

1.6 Inward trunks

With the introduction of TSPS Generic Program 7, TSPS base units and RTAS have the capability of performing some traditional inward operator functions such as general assistance, time and charges, call-back, and hotel call-back. A special 4-wire inward trunk must be used to handle inward calls. This inward trunk is composed of a toll office outgoing trunk circuit, a TSPS 4-wire bridging arrangement to provide TSPS operator access to the inward trunk, and a toll office incoming trunk circuit. These three circuits are interconnected by 4-wire facilities. An incoming call is switched by the toll office to the outgoing trunk circuit associated with an inward trunk and a TSPS operator is added to the inward trunk through the TSPS or RTA switching network. After the appropriate actions are performed by the operator, the toll office incoming trunk circuit associated with the inward trunk is switched by the toll office to the appropriate toll switching trunk, thereby extending the inward call to the desired end office. The inward trunk remains in the connection between the calling and called customers after the operator has released from the connection.

1.7 New equipment

Several new pieces of transmission-related equipment were introduced as a consequence of RTA and PSS No. 2.

A new unified operator's telephone circuit, or UTC, was designed to provide several improvements such as dual headset operation and supervisor conferencing without transmission degradation. The UTC also provides a voice-switched attenuator (VSA) to control operator talker echo return, an automatic gain control (AGC) circuit to protect the operator from high-level signals, and independent sidetone adjustment to permit meeting sidetone level objectives. Gain and line equalization adjustments are supplied to permit the UTC to properly terminate any standard transmission facility.

To eliminate unwanted echo at the point where BR or operator position trunks are bridged onto a TCT, a new precision-balanced hybrid, the type 1P 4-wire terminating set was introduced. This hybrid is characterized by a high-impedance 2-wire port (approximately 11,600 ohms) and a very high trans-hybrid loss.

Two new 3-way, 4-wire bridging repeaters were also introduced to substantially improve transmission quality when bridging 4-wire circuits. The previous bridging arrangements produced a volume contrast to both the customers and the operator between calls on 2-wire and calls on 4-wire TCTs, with the 4-wire circuits exhibiting approximately 3 to 6.5 dB greater loss, depending on the direction of transmission. The new bridges eliminate this problem and permit gain adjustment in every direction from the bridging point.

1.8 Signaling considerations

Like the base TSPS, the RTA is required to transmit and receive various types of signals. Among these are address signals (i.e., the called and calling numbers), coin control signals, ringback signals, and ringforward signals. The RTA trunk circuits allow each signal to be transmitted in at least two different ways. Unlike the base unit trunks, however, the RTA trunks are not required to generate ±130 V dc signals for coin control and ringback.

The RTA must be able to receive address signals either as dial pulses or multifrequency (MF) tones from originating local offices but is required to forward address information to the toll office only in MF form. Coin control and ringback signals toward the originating office can be either inband (MF tones) or multiple wink (MW), which consists of a specified number of timed on-hook/off-hook dc "winks." Ringforward signals may be either +130 V simplex or a single "wink."

In TSPS/RTA, as in other systems not equipped with common channel interoffice signaling (CCIS), the signals described above must be transmitted over virtually the same paths used for voice transmission. These paths include the BR trunks between the base and the RTA since all the common signaling circuits (e.g., MF receivers, MF outpulsers, coin control and ringback circuits) are located at the base. Hence, care must be exercised when engineering the transmission paths to ensure

that signaling is properly considered. Transmission level points (TLPs) must be carefully controlled to guarantee correct processing of MF signals by the various switching offices involved. MF signaling specifications call for −7dBm per tone signal power at 0- TLP (i.e., −7dBm0). Changes are being incorporated into TSPS to comply with this specification.

II. OBJECTIVES AND GENERAL CONSIDERATIONS

2.1 Voice transmission objectives

A standard TSPS/RTA connection is a 3-party call among the calling customer, the called customer, and the TSPS operator (see Fig. 3). This standard connection was considered to be the same as three 2-way connections, each of which must meet its own transmission objectives.

The transmission objective set for the 2-way customer-to-customer connection is to provide loss-noise-echo performance, which is the same as that of regular direct distance dialed (DDD) toll calls that do not have TSPS access but have the same length. This objective applies whether or not the operator is in the connection and has two corollaries, namely: (*i*) the toll connecting trunk (TCT) satisfies the standard DDD network transmission requirements in both the 2-way and 3-way connection, and (*ii*) the performance provided on calls handled by TSPS is the same as calls handled by the DDD network.

The transmission objective placed on the 2-way operator-to-calling-customer connection was that it have loss-noise-echo performance equivalent to that provided on an average short DDD toll call between 100 and 150 miles in length. This objective is satisfied by the TSPS/RTA transmission design even when there are several hundred miles between the operator position and the TSPS access point on the incoming trunks to the toll office. This performance is obtained because (*i*) the TSPS access point was given the same transmission level as the toll office incoming trunk which reduces the level contrast at the operator position between the received volume from the calling customer and that from the called customer, (*ii*) the BR trunk has zero loss to minimize volume contrast observed by the operator between calls accessed at the base unit and those accessed at an RTA, (*iii*) circuit noise on the trunk facilities was limited, and (*iv*) impedance balance was provided for all trunks to avoid echoes.

Similarly, the transmission objective for the 2-way operator-to-called customer connection was that its loss-noise-echo performance be approximately the same as that of the connection between calling customer and called customer. Two potential transmission degradations that had to be considered in meeting this objective were (*i*) customer and operator talker echoes as a result of the extra 2-wire switching

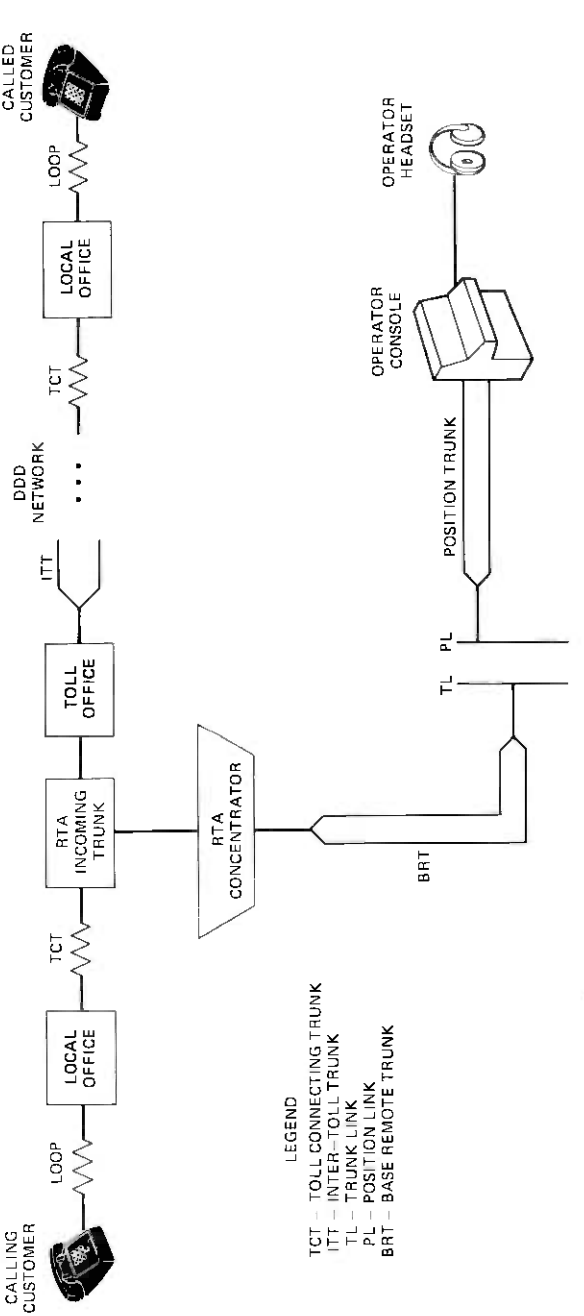


Fig. 3—Connection model of a typical call handled through TSPS via RTA.

point at the RTA and the long BR and operator position trunks, and (*ii*) high noise levels in long BR and operator trunk facilities.

Transmission objectives were also imposed on operator sidetone and trunk facility noise. The operator sidetone objective was made the same as that used for subscribers, namely, an acoustical sidetone path loss of 12 ± 4 dB. Noise objectives on all TSPS associated trunks were made the same as those for the DDD network trunks which use intertoll grade carrier facilities.

2.2 Deviations from requirements

Deviations from transmission requirements were constrained because the sum of many small deviations would cause large overall loss variations between connections. Some loss considerations were:

(*i*) Zero loss was required between the toll office and the base unit or RTA access points on toll connecting and inward trunks. Because any variation in this loss changes the effective TSPS access point on these trunks and, consequently, increases the speech volume from one customer and decreases it from the other, a permissible maximum variation of only 0.5 dB was set. For 2-wire trunks and 2-wire toll switches, the 0 to 0.5 dB loss requirement limits the wiring gauge and length that is used between the base unit (or RTA) and its associated toll office. Four-wire access trunks were aligned to 0 dB loss without difficulty. The permissible loss variation also determined the schedule for trunk maintenance and trunk down limits.

(*ii*) BR trunk loss variations were stipulated to limit the volume contrast for calls between operator and customers. The permitted variation was made the same as that for intertoll trunks in the DDD network. It was recognized that these level changes also affect signaling tones since MF outpulsers are only provided at the base unit so that the MF signaling tones are transmitted to the RTA over the BR trunks.

(*iii*) Loss variations in operator position trunks add to the volume contrasts heard by the customers. Such trunks were given same standard loss variations as intertoll message trunks.

2.3 Transmission level points (TLP)

TSPS conforms to the standard transmission level point (TLP) plan of the Bell System, which makes the local class 5 office a 0 TLP for outgoing signals. Since the trunk loss and its control are primary requirement in TSPS, transmission level points were adjusted as required to meet (*i*) trunk loss requirements, (*ii*) loading objectives for carrier facilities, and (*iii*) standardized toll office transmission level points. To do this, the outgoing TLP of the TSPS was adjusted to conform with the expected received level on the toll-connecting trunk being tested. All trunks internal to the base unit and RTA are considered to be at -3 TLP.

III. VOICE TRANSMISSION CONSIDERATIONS

As discussed in Section I, TSPS Generic 7 introduced the PSS No. 2, the RTA, and inward call-handling features. As a consequence, the RTA concentrator, which is a new switch, and new types of trunks (i.e., PSS No. 2 operator position, BR, and inward trunks) were added to existing TSPS base units as shown in Fig. 2. Modifications were also made in certain existing base unit circuits to meet the TSPS voice transmission objectives discussed in Section II. Each type of trunk associated with a base unit equipped with the RTA, PSS No. 2, and inward call-handling features was designed to satisfy specific transmission requirements to meet transmission objectives. These objectives were stated in terms of loss-noise-echo performance provided to the customer(s) and to the operator on all the possible connections used in TSPS. The loss-noise performance provided on a given type of connection was controlled by setting requirements on (*i*) the inserted connection loss (ICL) and (*ii*) the noise which may be introduced by each of the various trunks in the connections. The talker echo performance provided on a given connection depends upon both the loss and the round-trip signal propagation delay of the talker echo signal path. Because the round trip propagation delay of the echo path depends markedly on the length of a given connection, the loss in the echo path is used to control the echo performance provided on a given connection. For example, in the UTC described in Section 1.7, the loss of the echo path is controlled by introducing additional loss in the receive path when the talker speaks but not in the talker's transmission path. In addition, specific balance requirements are satisfied at all junctions between 4-wire and 2-wire transmission facilities. These specific transmission requirements (i.e., ICL, noise, and balance) are considered in subsequent sections.

Whenever the RTA feature is added to an existing TSPS base unit equipped with the original PSS No. 1 operator positions, the operator position trunks are modified as follows:

(*i*) The existing operator telephone circuits are replaced by the UTC with the voice-switched-attenuator (VSA) and automatic gain control (AGC) features enabled.

(*ii*) For operator positions located remotely from the base units, the DIC channel units at the base unit end of the trunk are replaced by a new DIC channel unit which incorporates the equivalent of a 1P 4-Wire Terminating Set (4WTS).

(*iii*) For operator positions located near the base unit, the 1H 4WTS used in the 24V4 repeater at the base unit end of the trunk are replaced by the 1P 4WTS.

Whenever the RTA feature is added, the service observing circuit used in existing TSPS base units is replaced by an improved circuit.

Prior to the introduction of Generic 7, a 3-way connection among (*i*) the TSPS operator, (*ii*) a service assistance operator, and (*iii*) the customer was permitted. However, with the introduction of Generic 7, such a 3-way connection is no longer provided because of transmission loss and balance problems with the extended range TSPS system. Only a 2-way connection is provided between a TSPS operator and a service assistance operator, during which the calling customer will be put on hold. The circuit modifications outlined above were needed to meet the echo performance objectives discussed in Section 2.1.

3.1 Network arrangements

Figure 2 is a functional block diagram of the voice transmission configuration of the base unit, operator positions, and RTA, and includes the various types of trunks. Typical trunking and switching arrangements at a base unit equipped with the inward call handling, a PSS No. 2, and an RTA are shown in Fig. 4. Figure 5 illustrates the typical trunking and switching arrangements for an RTA installation.

As shown in Fig. 4, the operator position trunks of the PSS No. 2 and PSS No. 1 retrofitted with the UTC as well as the service observing trunks all have 2-wire appearances on the position link circuit of the base unit switching network. These trunks use 4-wire facilities which are converted from 4-wire to 2-wire by a bridging hybrid (i.e., a 1P 4WTS, or equivalent) at the base unit. All the types of trunks shown in Fig. 4 that have appearances on the trunk link circuit of the switching network can be connected to at least one of the three types of trunks having appearances on the position link circuit. All BR trunks use 4-wire facilities while operator service trunks and centralized automatic message accounting (CAMA) transfer trunks may use either 2-wire or 4-wire facilities. When 4-wire facilities are used, a 900-ohm 4WTS (e.g., 1M 4WTS) is used to provide the necessary conversion from 4-wire to 2-wire. All inward trunks, 4-wire TCT, delayed call trunks, and service observing trunks use 4-wire facilities with bridging access to these trunks provided by 3-way, 4-wire bridging repeaters. The equivalent of a 900-ohm 4WTS is incorporated in the bridging repeaters to provide the necessary conversion from 4 wire to 2 wire. Two-wire TCTs are bridged directly at both the RTA and base unit. This is illustrated schematically in Figs. 4 and 5.

As shown in Fig. 5, all connections through the RTA concentrator between a BR trunk and either a TCT or an inward trunk are made on a 2-wire basis. All BR trunks at the RTA use 4-wire facilities and a 1P 4WTS to provide the necessary conversion from 4-wire to 2-wire transmission. The TCT and inward trunk arrangements are essentially the same as that described above for the base unit.

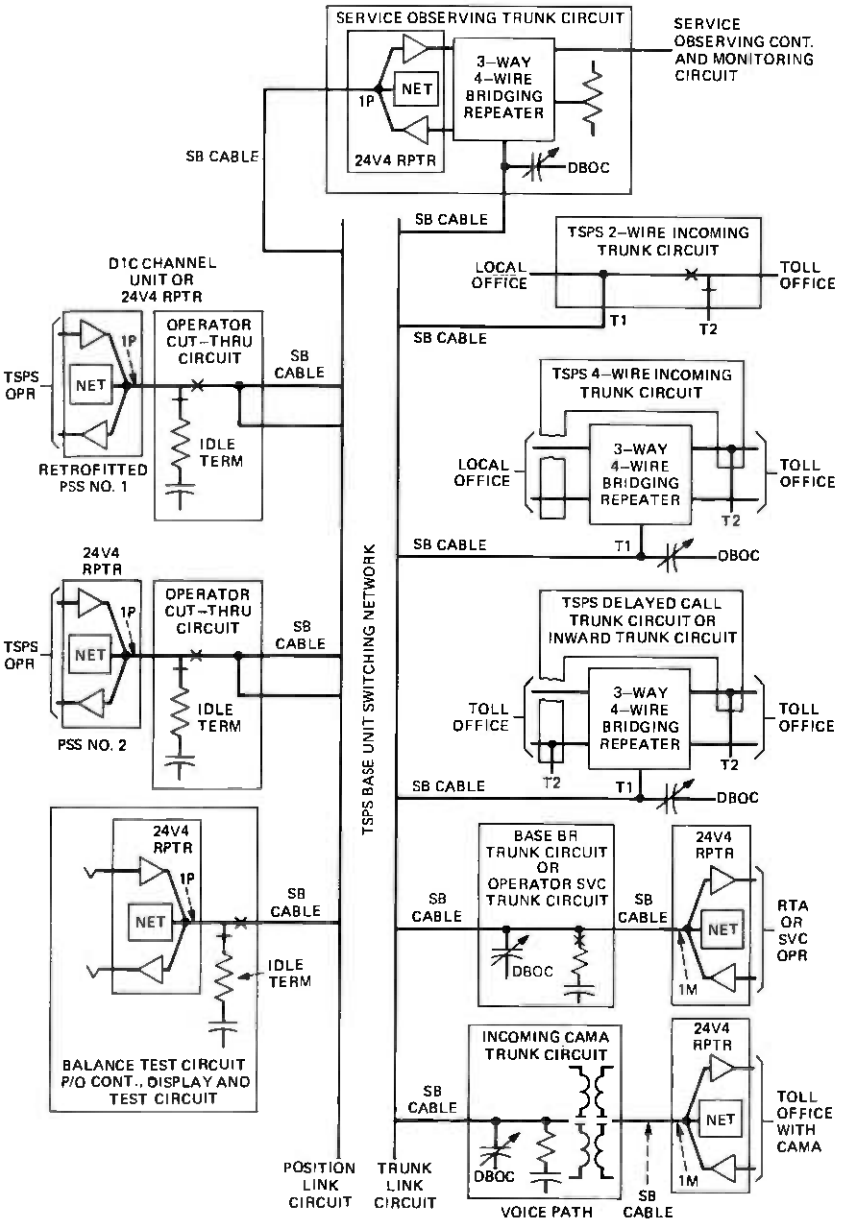


Fig. 4—TSPS base unit trunks and switching arrangements. To simplify the sketch, the T2 port connection to the trunk link circuit of the TSPS base unit switching network is not shown.

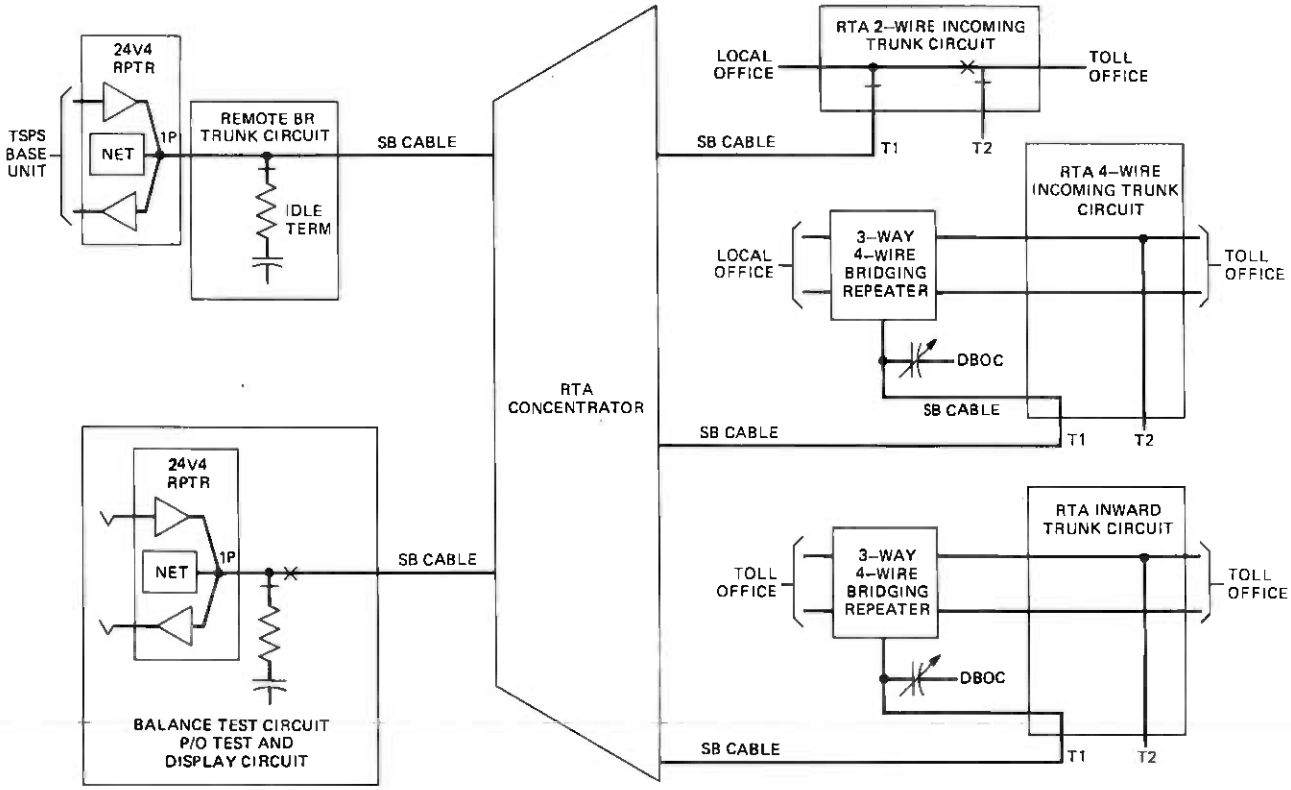


Fig. 5—RTA trunks and switching arrangements. To simplify the sketch, the T2 port connection to the RTA concentrator is not shown.

3.2 Inserted connection loss (ICL)

The ICL requirements on each of the various types of TSPS trunks shown in Fig. 2 are summarized in Table I. It should be noted that the introduction of Generic 7 has not affected the ICL requirements on those trunks which existed before its introduction. The ICL requirements on the new trunks introduced with Generic 7 are discussed next.

The ICL requirements must be the same for TCTs accessed by either a base unit or an RTA as for TCTs not accessed by TSPS to avoid any contrast in transmission performance provided to the customers between directly dialed calls and operator-assisted calls. One additional ICL requirement peculiar to TCTs provided with either base unit or RTA access is that the portion of the TCT between the TSPS bridging point and the outgoing side of the toll office switch should nominally be 0 dB with a maximum of 0.5 dB. This ICL requirement applies for both voice signals and MF signals and is needed (*i*) to equalize the transmission loss between the TSPS operator and the calling customer and between the operator and the called customer to minimize received speech volume contrast and (*ii*) to provide MF address signals to the toll office at the standard signaling power levels.

The ICL for the BR trunks is 0 dB. Since any TSPS operator can be called upon to service calls switched through either the base unit or an RTA, this requirement assures that there will be no significant volume contrast between such calls. It also ensures that the voice frequency MF signals generated at the base unit will reach the TCT bridging point at the standard power level.

The inward trunks introduced with Generic 7 may be considered as two secondary intertoll trunks in tandem, that is: (*i*) a secondary intertoll trunk from the toll office incoming to the base unit (or RTA)

Table I—Inserted connection loss requirements for TSPS trunks

Type of Trunk	ICL (dB)
1. Toll connecting trunk between TSPS No. 1 incoming trunk circuit at the TSPS base unit or RTA and the toll office	0, max. 0.5
2. PSS No. 2 operator position trunk	
a. Operator transmit direction	7
b. Operator receive direction	9
3. Retrofitted PSS No. 1 operator position trunk	
a. Operator transmit direction	7
b. Operator receive direction	9
4. Base-remote trunk	0
5. Inward trunk between TSPS No. 1 inward trunk circuit at the TSPS base unit or RTA and the toll office	0, max. 0.5
6. Delayed call trunk between TSPS No. 1 delayed call trunk circuit and toll office	0, max. 0.5
7. CAMA transfer trunk	
a. Voice path	0, max. 0.5
b. MF signaling path	0, max. 0.5
8. Operator service trunk	0, max. 0.5
9. Service observing trunk	0

for connection to an operator, and (*ii*) a secondary intertoll trunk from the base unit (or RTA) back to the toll office for connection to the desired end office. As such, each of these two sections of an inward trunk is composed of 4-wire facilities having an ICL of 0 dB nominal, with a maximum of 0.5 dB.

The ICL for the PSS No. 2 operator position trunks and for the retrofitted PSS No. 1 operator position trunks is the same. The ICL for both types of operator trunks depends on the direction of transmission. An ICL of 7 dB is used for the operator transmit direction so that the average speech power level from the TSPS operators delivered to the bridging point on the TCTs will be nominally equal to the average speech power level from the calling customers arriving at the same point. An ICL of 9 dB is used for the operator receive direction so that the preferred average speech power will be delivered to the operator when a customer's average speech level is applied to the bridging point on a TCT.

3.3 Noise

The maximum circuit noise permitted on all the various types of TSPS trunks existing before the introduction of Generic 7 was the same as on corresponding type trunks in the message network. This made it possible to use standard toll grade carrier arrangements in the makeup of the facilities for TSPS trunks whenever necessary. The noise limits on those trunks remain unchanged by the introduction of Generic 7 features, with the possible exception of the CAMA transfer trunks, which are discussed later. However, some special noise requirements for the inward, BR, and PSS No. 2 operator position trunks are introduced with Generic 7.

In the event that an inward call is extended to the desired called customer, the inward trunk remains in the overall connection between the calling and called customers after the operator releases from the connection. As discussed in Section 3.2, an inward trunk may be considered as two secondary intertoll trunks in tandem interconnected via a 3-way, 4-wire bridging repeater. To limit any degradation in the quality of the loss-noise performance provided to the calling and called customers, each of the two portions of the inward trunk are composed of low-noise 4-wire facilities such as intertoll grade T-carrier and/or 4-wire metallic facilities.

Any type or combination of types of intertoll grade voice facilities may be used in the makeup of a PSS No. 2 operator position trunks or BR trunks. The amount of noise introduced by the facilities is limited for loss-noise performance reasons. The loss-noise performance provided on customer/operator connections is adversely affected by the additional circuit noise introduced by long PSS No. 2 operator position trunks and long BR trunks.

If a PSS No. 2 is provided at a given TSPS base unit, but not an RTA, the loss-noise performance objective is met on customer/operator connections if the facility route length of analog carrier facilities used in any operator position trunk is no greater than 400 miles. This length limitation applies if the following noise maintenance limits are observed on the analog carrier facilities used in the operator position trunk:

Analog Carrier Length (route miles)	Remove From Service Limit (dBrnC0)	Maintenance Req'd Limit (dBrnC0)
0– 50	40	32
51–100	40	33
101–200	40	35
>200	44	37

If companded L-multiplex analog carrier facilities are used in the operator position trunk, the maximum facility route length may be extended to 1000 miles. It should be noted that the length of a single intertoll grade digital carrier facility (e.g., T-carrier) when used in tandem with an analog carrier facility can be ignored when applying these length restrictions. If the RTA feature is provided but not the PSS No. 2 feature, the above discussion on the makeup of PSS No. 2 operator position trunk facilities applies directly to BR trunks. If both the PSS No. 2 and RTA features are provided at a given TSPS base unit, the analog carrier facility length restrictions indicated above must be apportioned between the operator position trunks and the BR trunks. The length restrictions, of course, apply to all the diverse routes taken by the BR and operator trunks.

The introduction of RTA increases the likelihood that some of the CAMA transfer trunks to a given base unit will be long. The CAMA traffic, formerly handled by operators located near a toll office which is now being served by an RTA, will now most likely be routed over CAMA transfer trunks to the base unit for handling by TSPS operators. This, coupled with the possibility that the base unit may also be equipped with a remote PSS No. 2, places more stringent noise requirements on the CAMA transfer trunks so that their maximum circuit noise is now the same as for the BR trunks discussed above.

3.4 Transmission balance

RTA and PSS No. 2 add a new 2-wire switching point (RTA concentrator) and long lengths of facilities to the overall connection between the customers and the TSPS operator. As discussed in Section 1.5, the extra switching point introduced another potential source of talker echo whenever the trunks to be switched use 4-wire facilities. The long lengths of facilities increase the round-trip propagation delay of any

talker echo signal already existing which makes that echo more disturbing to the talker. To meet echo performance objectives on connections involving TSPS operators, balance requirements must now be placed on the various connections through the base unit and/or through the RTA when at least one of the trunks involved in the connection uses 4-wire facilities. Balance refers to matching circuit impedances to control the magnitude of the signal reflected toward the source at the junction between 4-wire and 2-wire facilities.

In general, the termination of a 4-wire facility and its conversion to a 2-wire facility is accomplished with a 4WTS, or equivalent. A 4WTS contains a balanced hybrid transformer that facilitates the transfer of signal power from the 4-wire receive path into the 2-wire facility and from the 2-wire facility into the 4-wire transmit path. However, some signal power arriving on the 4-wire receive path will be returned on the 4-wire transmit path whenever differences exist between the impedance of the 2-wire facility and the impedance of the balancing network of the 4WTS. The degree of balance between these two impedances will determine what portion of the signal power will be returned to the source.

The balancing networks used in each of the various types of 4WTSs (or equivalent) shown in Fig. 4 and 5 consist of a compromise network plus an adjustable network build-out capacitor (NBOC), which shunts the compromise network. The compromise network is designed to provide impedances over the voice frequency band of 200 to 3400 Hz that match the nominal impedance of the 2-wire circuits which may be connected to that 4WTS. The NBOC is included in the balancing

Table II—Balance requirements on the bridging hybrid at the base unit end of PSS No. 2 and retrofitted PSS No. 1 operator position trunks and the RTA end of base-remote trunks

Type of Trunk Connected to TSPS Bridging Hybrid	Echo Return Loss (dB)		Singing Return Loss (dB)	
	Median	Minimum	Median	Minimum
1. Base unit and RTA				
a. 2-wire toll connecting trunk	15	13	N.S.*	6
b. 4-wire toll connecting trunk	N.S.	19	N.S.	15
c. Inward trunk	24	21	19	16
2. Base unit				
a. Base-remote trunk	24	21	19	16
b. 4-wire delayed call trunk	24	21	19	16
c. 4-wire operator service trunk	24	21	19	16
d. 4-wire CAMA transfer trunk (voice path)	N.S.	21	N.S.	16
e. Service observing trunk	24	21	19	16

* Not specified.

Table III—Balance requirements on 4WTS, or equivalent, associated with various types of 4-wire trunks on connections to a bridging hybrid in TSPS base unit or RTA

Type of 4-Wire Trunk	Echo Return Loss (dB)		Singing Return Loss (dB)	
	Median	Minimum	Median	Minimum
1. Base unit and RTA				
a. Toll connecting trunk	N.S.*	26	N.S.	19
b. Inward trunk	N.S.	26	N.S.	19
2. Base unit				
a. Base-remote trunk	N.S.	26	N.S.	19
b. Delayed call trunk	N.S.	26	N.S.	19
c. Operator service trunk	N.S.	26	N.S.	19
d. CAMA transfer trunk (voice path)	N.S.	26	N.S.	19
3. Service observing trunk	N.S.	26	N.S.	19

* Not specified.

network to simulate the capacitance of the office cabling included between the 2-wire port of the 4WTS and the point of good impedance of the connecting circuit, which is defined as a point of fixed nominal 2-wire impedance. For example, the point of good impedance of a trunk using 4-wire facilities is the 2-wire port of a 4WTS.

The balance requirements on the 1P 4WTS (or equivalent) at the base unit end of PSS No. 2 trunks, retrofitted PSS No. 1 trunks and service observing trunks, or at the RTA end of BR trunks, vary depending upon the type of trunk connected through the base unit network or RTA concentrator to the 1P 4WTS. The balance requirements, listed in Table II as a function of trunk type, are identical regardless of whether the connections are made through the base unit or through the RTA.

The balance requirements on the 4WTS (or equivalent) associated with the various types of 4-wire trunks which can be connected through the base unit or the RTA to a 1P 4WTS are summarized in Table III. Again, balance requirements are identical for a given type of trunk switched through either the base unit or an RTA.

The balance requirements listed in Tables II and III are given in terms of echo return loss (ERL) and singing return loss (SRL) as measured at the 4WTS (or equivalent) using a return loss measuring set (RLMS). An ERL measurement is a weighted average measurement of the return losses for each frequency in the echo frequency range. An SRL measurement is the lesser of two weighted average measurements of the return losses, one measurement covering the frequencies at the lower end of the voice frequency band and the second covering frequencies at the upper end of the voice frequency band.

No resistance build-out capability is provided in the balancing networks of the various 4WTS (or equivalent) used in TSPS applications.

Therefore, to meet specific balance requirements at any given 4WTS maximum allowable lengths of office cable (26-gauge switchboard cable in both base units and RTAs) between the points of good impedance of the trunks and their 2-wire appearance on the base unit switching network or RTA concentrator have been specified. Those segments of switchboard cable in which length must be controlled are labeled "SB CABLE" in Figs. 4 and 5.

In a given base unit or RTA, the value of the NBOC in the balancing networks of all the 1P 4WTSs will be set to some compromise value such that the balance requirements are met on connections through the switching network to all possible 2-wire TCTs. To meet balance requirements when the value of the NBOC is fixed, variation in the shunt capacitances of the switchboard cables shown in Figs. 4 and 5 must be controlled. If the switchboard cabling capacitance variation is sufficiently large, balance requirements will not be met on all possible connections if the value of the NBOC is to remain single-valued.

It should be noted that the shunt capacitance of the switchboard cabling between a 1P 4WTS and the point of good impedance of any of the various trunks using 4-wire facilities shown in Figs. 4 and 5 will, in general, be much smaller than the value of the NBOC in the hybrid balancing network of the 1P 4WTS required to meet balance requirements on connections to TCTs providing 2-wire bridging access. To reduce this cabling capacitance variation, bridged capacitance is added to the shorter 2-wire switchboard cabling paths to increase their effective shunt capacitance and make them equal that of the longer 2-wire cabling paths. This is accomplished by adding a drop build-out capacitor (DBOC) in the shorter 2-wire cabling paths as indicated in Figs. 4 and 5. In some cases, the DBOC is supplied as part of the TSPS or RTA trunk circuit or as part of the 4-wire bridging repeater. In other cases, a separate DBOC is supplied and wired on a bridged basis across the 2-wire switchboard cabling path between the point of good impedance of a given trunk and its appearance on the base unit switching network or on the RTA concentrator.

Terminal balance requirements are imposed on connections in a toll office from any intertoll trunk to any TCT. When the TCT provides TSPS bridging access, the same terminal balance requirements apply. The current toll office terminal balance requirements listed in Table IV are for the most part independent of the type of toll office involved, but are dependent upon the general makeup of the TCT facilities. To meet terminal balance requirements on 2-wire TCTs providing TSPS bridging access, the maximum length and the gauge of the switchboard cabling used to interconnect the base unit or RTA incoming trunk circuit, the toll office, and the point of good impedance of the TCT facilities (e.g., 4-wire terminating set, impedance compensation network) must be controlled as a function of (*i*) the type of toll office, (*ii*) the length and

Table IV—Toll office balance requirements

1. Terminal balance
   a. 2-wire facilities with 2-dB pad (intrabuilding):
      ERL: 22 dB* median, 18 dB minimum
      SRL: 14 dB median, 10 dB minimum
   b. 2-wire facilities or 4-wire facilities with 2-wire extensions (interbuilding):
      ERL: 18 dB median, 13 dB minimum
      SRL: 10 dB median, 6 dB minimum
   c. 4-wire facilities:
      ERL: 22 dB median, 16 dB minimum
      SRL: 15 dB median, 11 dB minimum
2. Through balance
      ERL: 27 dB median, 21 dB minimum
      SRL: 20 dB median, 14 dB minimum

* 20 dB for No. 5 crossbar toll offices only when TCT provides TSPS 2-wire bridging access.

gauge of switchboard cabling internal to a 2-wire toll office, and (*iii*) the value of the NBOC of the balancing network of the intertoll hybrids in a 2-wire toll office.

The terminal balance requirements listed in Table IV also apply to connections through the toll office from any CAMA transfer trunk, TSPS inward trunk, or TSPS delayed call trunk to any TCT. The toll-office through balance requirements also listed in Table IV must be met on connections through the toll office between any TSPS inward or delayed call trunk and any intertoll trunk. These balance requirements have been imposed to limit any degradation in the quality of the echo performance provided on the overall connection resulting from the introduction of additional 2-wire switching points in the overall connection between customers. To limit degradation of the echo performance provided on the overall connection between two customers resulting from the increased round-trip propagation delay of the echo path caused by adding either a TSPS inward or a delayed call trunk in tandem with the overall connection, the maximum route length of the facilities composing each of the two portions of an inward or delayed call trunk must not exceed:

(*i*) Metallic facilities	9 miles
(*ii*) T-carrier facilities	50 miles
(*iii*) T-carrier with metallic extensions	Determined by the requirement that the length of the T-carrier facilities plus 10 times the length of the metallic extension must not exceed 50 miles.

IV. ADDRESS SIGNALING AND SPECIAL SIGNALING

4.1 Signaling methods

RTA trunks utilize classical loop and E&M supervisory signaling methods. Nonsupervisory signal handling, however, is different.

Signaling between the RTA and the local or toll switching offices typically assumes one of two forms: (*i*) ac multifrequency tone bursts and (*ii*) dc pulses or "winks." When MF tones are used, the signaling is usually between the local or toll offices and the base unit with the RTA essentially transparent. This is the case for MF address signals, automatic calling number identification (ANI), and inband coin control and ringback. When dc pulsing is employed, however, signaling is restricted to the relatively short trunks between the RTA and the local and toll offices. This is the case for dc dial pulses and multiple wink (MW) coin control, ringback, and ringforward.

The RTA does not provide ±130 V dc signaling for coin control, but does provide +130 V dc simplex signaling for ringforward toward the toll office.

4.2 Dial pulse receiving

It would be nearly impossible for dc dial pulses (DP) generated at an originating office to pass through the RTA and over the extremely long BR trunks that could be encountered and still be reliably detected by a digit receiver at the base unit. An alternative way of handling DP at the RTA is required. Rather than being transmitted to the base, dial pulses received at an RTA are blocked in the trunk circuit where they are counted and encoded by the Position and Trunk Scanner (PTS) circuit in the PCL. The value of each dialed digit is then transmitted in the interdigital interval to the SPC over the PCL data links and not over the normal voice transmission facilities.

When all the called digits have been obtained, the SPC must determine if the originating office is equipped with ANI (automatic number identification). If it is, an MF receiver must be connected to the T1 port of the RTA incoming trunk via the base networks, a BR trunk, and the RTA concentrator. The trunk circuit is then directed to send an off-hook "wink" to the originating office requesting transmittal of the calling number. The MF tones then follow essentially the same path used to connect an operator to the calling customer. Hence, in the design of the toll-connecting and BR trunks, consideration must also be given to signaling.

4.3 Multifrequency receiving

If the RTA incoming trunk is identified as having MF rather than dial pulse signaling, the SPC establishes a transmission path to a base unit MF receiver in much the same manner as described above to receive the ANI digits after dial pulsing. Once a transmission path is established, the RTA is directed to signal the originating office via a 200-ms off-hook wink to start MF pulsing. After all the called number digits have been registered, the SPC must determine if the originating office

is equipped for ANI and, if so, must signal the office to forward the calling number. This is done by directing the RTA trunk to go off-hook. Thus, in the case of MF originating traffic, the signaling and the voice transmission paths are identical.

4.4 Multifrequency outpulsing

Once the called number has been registered by the SPC, it must be forwarded or outpulsed to the toll office to effect call completion. Since this is normally done at the same time that an operator is connected to the calling customer (acknowledging coin deposits, getting special billing information, etc.), a second transmission path must be established, this time via the T2 port of the incoming trunk circuit. In the case of an RTA call, the MF outpulser circuit is connected to the trunk via a BR trunk and the RTA concentrator.

MF signals transmitted from TSPS to the toll office, therefore, follow a somewhat different route than voice signals, traveling via the T2 port of the trunk circuit while voice signals use only the T1 port. In the case of the relatively simple 2-wire trunk circuits, this makes little difference (see Fig. 6). In the case of 4-wire trunk circuits, however, the MF signals do not go through and therefore do not receive the benefit of the amplifiers in the associated bridging repeaters

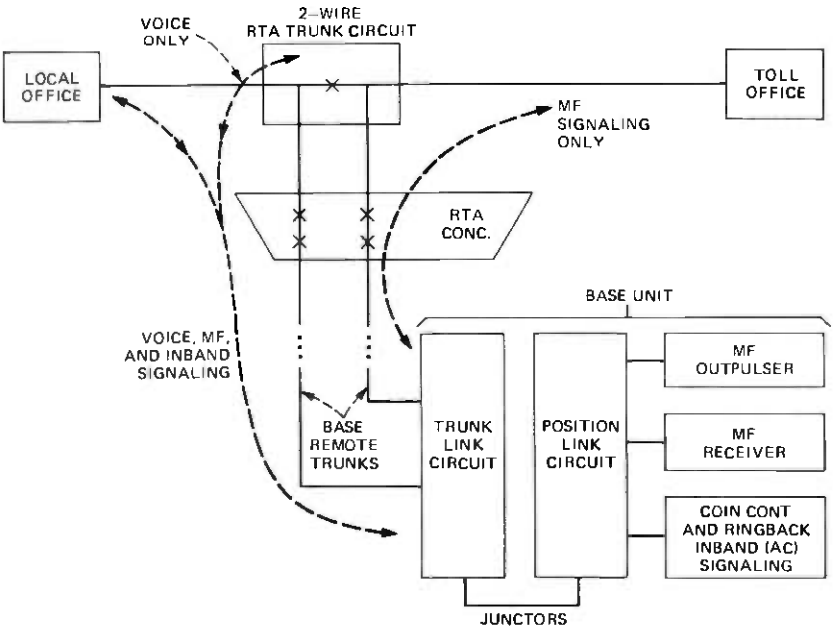


Fig. 6—MF signaling in 2-wire RTA trunk circuits.

(see Fig. 7). Thus, the 4-wire circuits require additional amplifiers to ensure that the proper ratio of MF to voice is transmitted to the toll office.

Care must also be exercised during the design of the trunk circuits and facilities to ensure that they are properly terminated at all times, particularly at the 2-wire/4-wire junctions. Improper termination of the new 1P 4-wire terminating set can lead to unintentional circuit gains or losses with resulting MF signaling difficulties.

For typical DDD calls, the TSPS must transmit the MF signals only as far as the serving toll office which, even for RTA and its toll office, is by design over trunk facilities of no more than 0.5-dB loss. With the introduction of International Direct Distance Dialing (IDDD) in Generic 5, however, TSPS must transmit MF signals to international "gateway" offices which may be more than 2000 miles away. On such calls, TSPS must first transmit a preliminary set of digits to the toll

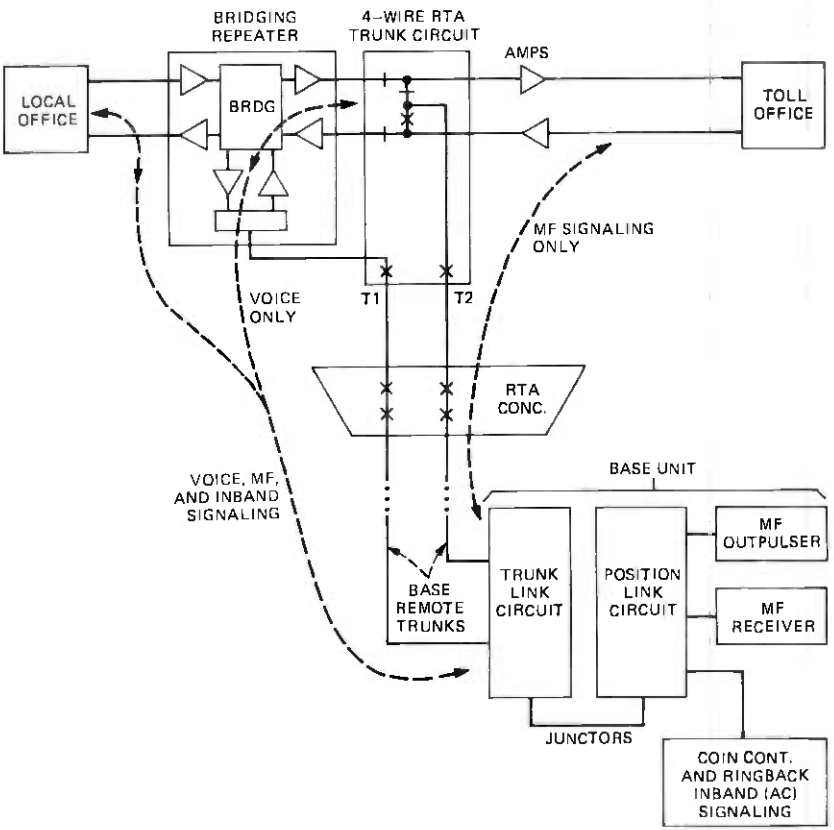


Fig. 7—MF signaling in 4-wire RTA trunk circuits.

office which will cause a connection to the gateway office to be established. Upon receipt of a signal from the gateway office, TSPS must then MF-outpulse the called number.

Transmission facilities of this length may be made up of many segments, be characterized by up to 10 dB of loss, and/or be equipped with echo suppressors. Hence, the MF signals must be transmitted from TSPS at the current standard level of −7 dBm0 per frequency to insure they are not too strong and cause facility overload, clipping, or unacceptably high crosstalk and are not too weak to be successfully received by the gateway offices.

4.5 Coin control and ringback

The RTA trunks are all equipped to provide coin control and ringback by transmitting either inband (ac) or multiple wink (dc) signals. Inband signaling requires the use of the coin control and ringback (CCR) circuit at the base unit to supply the inband frequencies. The SPC establishes a connection between the CCR and the T1 port of the RTA trunk via the base network, a BR trunk, and the RTA concentrator. The RTA trunk is then directed to transmit an alerting on-hook wink toward the local office followed by a quiet interval of 100 ms. Inband frequencies for coin collect (700/1100 Hz), coin return (1100/1700 Hz), or ringback (700/1700 Hz) are then generated by the CCR circuit.

For multiple wink signaling, the RTA trunk circuits are each equipped with an integrated circuit which when properly addressed by the PCL signal distributor circuit, will generate three winks for coin collect, four winks for coin return, or five winks for ringback. One wink and two winks, not currently used in the multiple wink signaling arrangements, are available for future use. The multiwink generator, however, does generate the one wink as an alerting signal for inband signaling and for ringforward (see Section 4.6).

Timing for the multiwink generator is derived from the 12.346-ms Position and Trunk Scanner clock in combination with an integrated circuit decade counter in each trunk circuit. Hence, the winks appear as squared waves with symmetrical 123-ms intervals.

4.6 Ringforward

The RTA trunk circuits can be arranged on an optional basis to provide either wink or simplex versions of the ringforward signal. Timing for this signal is produced by the same MW generator in the trunk circuits used for coin control and ringback. If the +130-V dc supply is connected to the trunk circuit (optional), the ringforward signal will be +130-V simplexed onto the T and R leads toward the toll office. Otherwise, the T and R leads are simply opened and a wink-type signal is sent to the toll office.

V. MAINTENANCE FACILITIES

5.1 CDT modifications

The control, display and test (CDT) frame at the base unit serves as the trunk test panel and also as the primary status display for the TSPS peripherals and subsystems. With the introduction of RTA and PSS No. 2, it was necessary to modify the CDT to accommodate BR trunk and PSS No. 2 position trunk testing and to add a status display for the PCL circuits and data links. A single display is provided which, under control of the SPC, can show the status of any one of the 15 PCLs possible in a given TSPS. This novel display is shown in Fig. 8.

To meet long-range Bell System objectives on test-tone power levels and to prevent crosstalk and other disruptions of carrier systems which are likely to be used for the long BR and operator position trunks, the CDT was modified to permit trunk testing at power levels 10 dB below TLP. Since the base unit, the RTA, and all trunks internal to TSPS (e.g., BR trunks) are considered to be at −3 TLP, the test tones utilized will be at −13 dBm. Coincident with this change, a system for varying the test pad (TP) values was introduced to achieve consistent values of expected measured loss (EML) when testing the difference segments of a TCT. This multilevel testing scheme is discussed later.

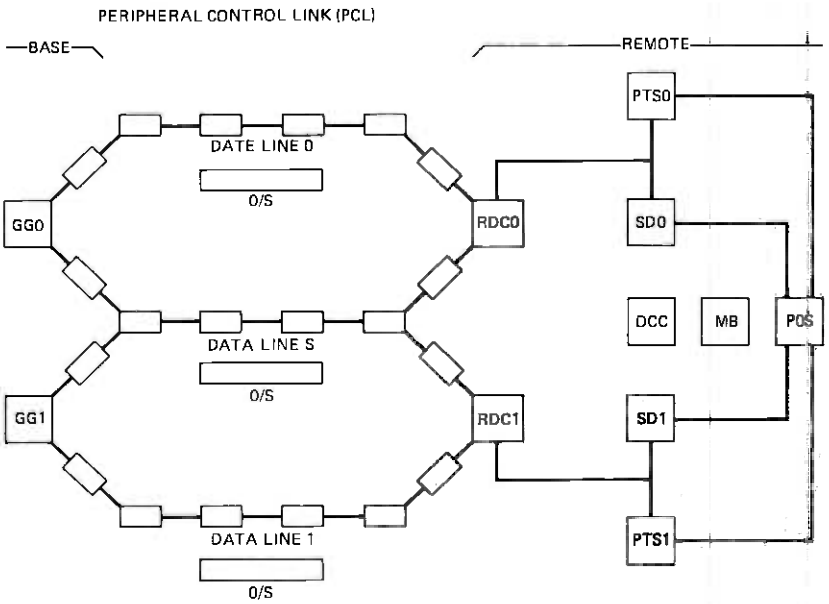


Fig. 8—Control display and test circuit frame—PCL equipment status display.

5.2 Test and display circuit

The test and display circuit (TDC), located at remote RTA and PSS No. 2 sites, provides functions similar to those of the CDT at the base unit. The TDC provides a PCL display, trunk group status indicators, carrier facility and carrier transfer circuit status lamps, and other major circuit status lamps.

A major feature of the TDC is its transmission and noise measuring circuit which is similar to the Automatic Transmission Measuring System 52A Responder. This circuit takes a "snapshot" measurement of the test tone or noise level on a trunk under test and displays the results on an LED digital display. The measured results are also passed in BCD form to the scanner circuit for transmission via the PCL to the base unit. Since the RTA and PSS No. 2 sites are likely to be unattended, this arrangement facilitates one-person testing of BR and position trunks.

On request from the CDT, the TDC is placed in a remote control mode, responding to commands from the SPC transmitted via the PCL. In this mode, the following tests can be made either semiautomatically or by automatic progression from one test to the next and for each trunk in the trunk group:

(i) Near-to-far end loss: Measurement is made by the TDC, transmitted to the base, and printed on the maintenance center TTY.

(*ii*) Far-to-near end loss: Measurement is made on a transmission measuring set built into the CDT.

(*iii*) Near-to-far end noise: Measurement is made as in test (*i*).

(*iv*) Far-to-near end noise: Measurement is made on a 3B noise measuring set built into the CDT.

In its local mode, the TDC is used to test the segments of the TCTs between the RTA and the local offices and between the RTA and the toll office. It can also be used for 2-person testing of BR or operator position trunks. As with the CDT, tests on these circuits are made at 10 dB below TLP. Test jacks are provided at the TDC to facilitate the use of portable test equipment in testing the PCL data links, TCTs, BR trunks, and the teletypewriter data channel.

5.3 Dial access test lines

A new transmission maintenance facility introduced concurrently with Generic 7 is the Dial Access Test Line circuit or DATL. The DATL circuits, located at both the RTA and the base unit, provide the equivalent of a code-100 test line (quiet termination), a code-102 test line (milliwatt 1000-Hz tone) and a code-106 test line (loop-around). These circuits permit maintenance craft at the local and toll offices to reach a TSPS test termination by simply dialing a test code, typically

959-120X for the quiet termination or 959-122X for the milliwatt tone (the "X" signifies that the last digit is immaterial).

Each DATL has two ports, each of which can supply a quiet termination or milliwatt tone. When two trunks in the same trunk group access the DATL at the same time, the SPC will automatically loop the trunks to each other through the DATL. To prevent this arrangement from being used as an unauthorized "meet me" circuit, the DATL is equipped with 60A control units which are designed to pass signals within a very narrow passband centered at 1004 Hz. If energy outside this band and above approximately −30 dBm is detected, the 60A will open the loop. The loop-around path produces an EML twice that of a 1-way connection.

5.4 Maintenance improvements

The introduction of multilevel testing at the CDT and TDC frames constitutes a noteworthy transmission maintenance improvement. In this scheme, the values of the test pad is changed depending upon the direction of the tests, that is, toward the local office, toll office, base unit (for the TDC), or RTA (for the CDT). Testing a toll-connecting trunk associated with TSPS (including RTA) may also involve testing the TCT segments. There are three possible tests: local office to toll office, local office to TSPS, and toll office to TSPS. Note that each switching center can list a given trunk with two different terminating offices, yielding the likelihood of two different EMLs for the same circuit. Test-pad switching at the CDT and TDC eliminates that condition.

Consider a TCT with ICL of 3 dB being tested between a local office and a toll office equipped with a typical 2-dB test pad. The EML at both offices would be 5 dB. However, if the same TCT circuit is tested between TSPS and the toll office when the ICL of the TCT segment is 0 dB, the toll office would now measure 2 dB. To eliminate this inconsistency, 3-dB test pad is inserted at the CDT (or TDC) to make the toll office measurement 5 dB. Similarly, if the TCT segment between the local office and the TSPS is tested, the local office would now measure 3 dB (the ICL value) or possibly 6 dB if the CDT or TDC test pad is 3 dB, as above. Instead, the CDT (TDC) test pad is switched to 2 dB, making the measurement at the local office a consistent 5 dB.

Since the TSPS base unit can be associated with more than one toll office, the CDT is equipped with test pads of 0, 2, and 3 dB to simulate the test pads found in Bell System toll offices. The TDC may have only one toll test pad value, 0, 2, or 3 dB.

VI. CONCLUSION

The introduction of RTA and PSS No. 2 with their long BR and operator position trunks required establishment of a new comprehen-

sive transmission plan. Several existing base unit circuits had to be modified or completely redesigned to meet tightened transmission standards. Equipped with the new 3-way, 4-wire bridging repeaters, unified telephone circuits, and precision 1P 4-wire terminating sets, the RTA and PSS No. 2 provide generally equivalent, if not better, transmission quality to that which previously existed in TSPS.

VII. ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of many colleagues who took part in the establishment of the transmission plan and the development of the equipment described in this article. Although they are too numerous to list completely, several deserve special mention: L. A. Ferrara for his work on the UTC: W. Fischer for trunk signaling; E. Tammaru for data link error analysis; J. J. Hibbert, S. M. Kay, and R. J. Frank for transmission planning; and various members of Department 4161 for their work and cooperation in the development of the new 4-wire bridging repeaters.

REFERENCES

1. T. F. Arnold and R. J. Jaeger, Jr., "TSPS/RTA—An Overview of the Remote Trunk Arrangement," ICC75 Conference Record, *III* (June 1975), pp. 46-1—46-4.
2. J. A. Hackett, K. A. Heller, and L. A. Rigazio, "TSPS/RTA—Reliability and Maintainability in the Hardware Design," ICC75 Conference Record, *III* (June 1975), pp. 46-5—46-10.
3. S. M. Bauman and M. F. Sikorsky, "TSPS/RTA—Call Processing," ICC75 Conference Record, *III* (June 1975), pp. 46-11—46-16.
4. W. L. Brune and R. J. Frank, "TSPS/RTA—Transmission Considerations," ICC75 Conference Record, *III* (June 1975), pp. 46-17—46-21.
5. V. L. Ransom and G. Espinosa, "TSPS/RTA—Application and Planning Information," ICC75 Conference Record, *III* (June 1975), pp. 46-22—46-25.
6. T. F. Arnold, "TSPS Goes to the Country," Bell Laboratories Record *55*, No. 6 (June 1977), pp. 147–153.
7. W. S. Hayward, Jr. and R. E. Staehler, "TSPS No. 1: An Overview," B.S.T.J., this issue pp. 1109–1118.
8. S. M. Bauman, R. S. DiPietro, and R. J. Jaeger, "TSPS No. 1: RTA Overall Description and Operational Characteristics," B.S.T.J., this issue, pp. 1119–1135.
9. A. F. Bulfer, W. E. Gibbons, and J. A. Hackett, "TSPS No. 1: RTA Hardware and Software Implementation," B.S.T.J., this issue, pp. 1167–1205.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Remote Trunking Arrangement: Hardware and Software Implementation

By A. F. BULFER, W. E. GIBBONS, and J. A. HACKETT

(Manuscript received December 28, 1978)

The principal data-handling circuits used between a TSPS base location and remote sites up to 1000 miles distant are described. Included are novel operational programming and fault recognition features required by this large separation between the processor and its peripherals as well as by the uniqueness of certain remote units. The RTA's use of remotely run diagnostics and multi-unit initialization schemes for fast recovery from outages is also covered.

I. INTRODUCTION

Early in the RTA planning process, it was recognized that controlling trunks, networks, and test circuits remotely was a task similar to controlling operator positions, voice path circuits, and test circuits remotely; i.e., the RTA job would be similar, in several aspects, to the work planned for development of a new position subsystem. These configurations are shown in Fig. 1. These job similarities led to development of a set of data-handling circuits known as the Peripheral Control Link (PCL) used to link the remote hardware to the controlling processor at a TSPS base location. A Remote Trunk Arrangement (RTA) consists of a PCL, various 2-wire and 4-wire trunks at the remote site, a concentrator network to temporarily connect these remote trunks to a base-remote trunk, a maintenance buffer, and a trunk test frame. A Position Subsystem (PSS) No. 2 consists of a PCL, several operator positions, some supervisory and chief operator circuits, a voice path control frame, a maintenance buffer, and a test frame.

The PSS No. 2 is only briefly treated in this paper and is mentioned at this time to indicate another consideration in PCL design—a desire

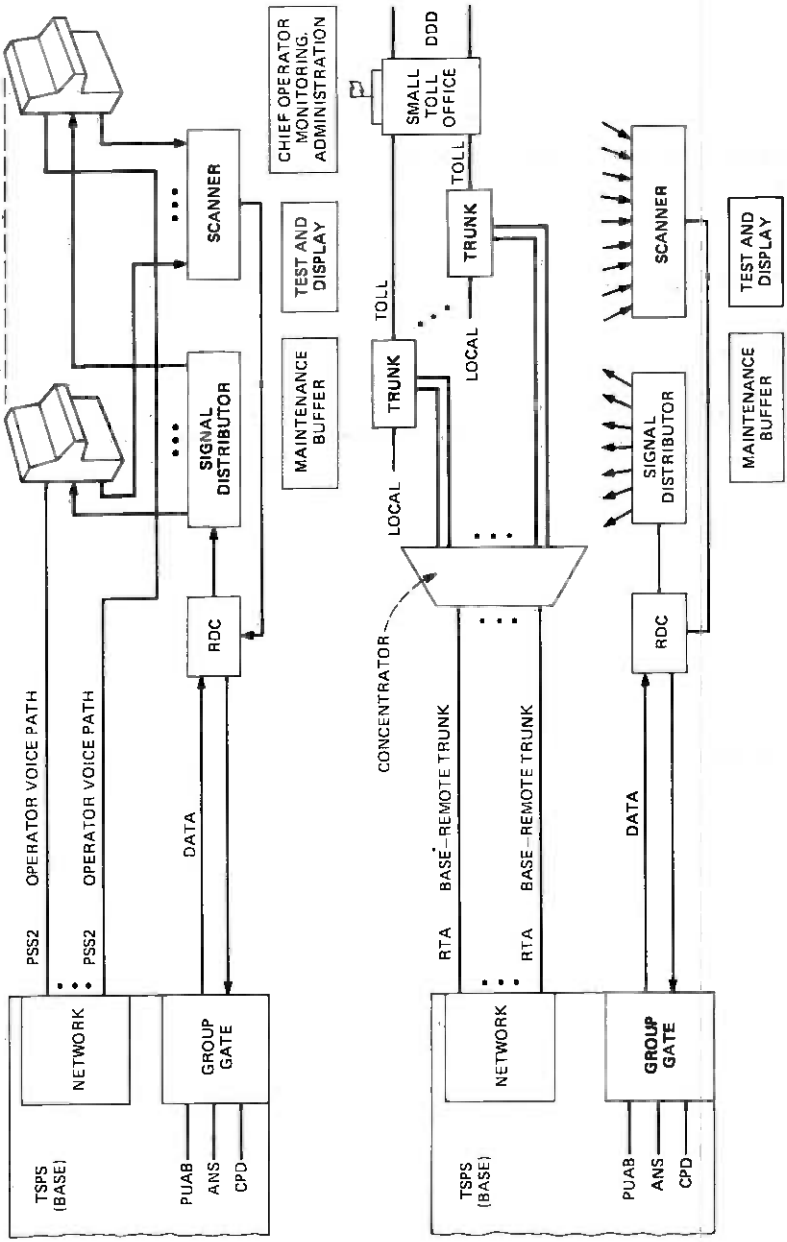


Fig. 1—RTA/PSS No. 2 common equipment.

to keep the PCL general purpose and to have its data rate high enough to handle bursts of operator activity. In both the RTA and the PSS No. 2 configurations, the remote circuits (trunks or positions) are temporarily connected to the base TSPS network in processing a call. In the RTA case, a base-remote trunk is used so that various "service" circuits at the TSPS base location (digit receivers, outpulsers, tones, announcements, etc.) can be connected to the RTA trunk as needed. In the PSS No. 2 case, operators at remote positions are temporarily connected through the base TSPS network to customers placing toll calls requiring operator assistance. The many voice circuits required in each case (up to 64 base-remote trunks or operator voice paths) utilize carrier transfer circuits and carrier failure detection circuits to permit fast, processor-controlled switching to spare carrier groups in the event of trouble. This carrier group switching is similar to that used in other switching designs and is only briefly discussed in Section 2.5.

The PCL includes a base location data conversion circuit known as a group gate, outside plant facilities, remote data circuit, signal distributor, scanner, and diagnostic control circuit. Each of these circuits consists of two identical halves and are arranged in two separate PCL halves with triplicated transmission facilities as shown in Fig. 2. A PCL half is capable of sustaining all communications and control between the Stored Program Control (SPC) and the RTA or PSS No. 2 circuits. When both PCL halves are in service, a state referred to as "duplex,"

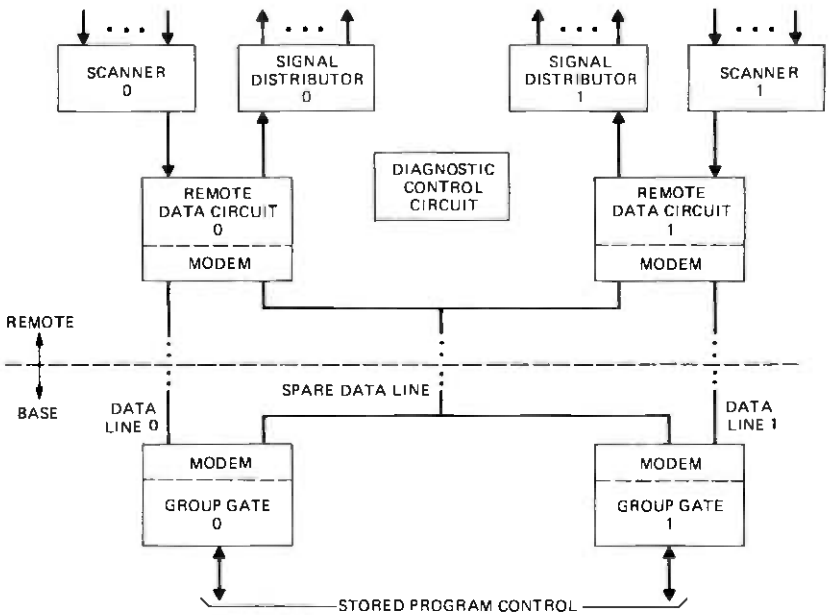


Fig. 2—Peripheral control link.

the PCL throughput remains the same, but the error detecting capabilities of the PCL are increased. Several reliability studies[1-3] have shown the transmission paths to be the weakest link and other studies have shown matching to be one of the best methods of error detection. This led to use of triplicated facilities so that duplex operation *with matching* can be continued even with a facility outage. Operation without matching is also possible so that, while error-detecting capability is lessened, complete RTA or PSS No. 2 operation can continue with two simultaneous facility outages. When a PCL half is out of service, the PCL is said to be in a simplex state.

The RTA equipment not considered part of the PCL is not duplicated. However, control of this part of the RTA is distributed over a variety of individual control devices, such as trunk buffers, concentrator controllers, and maintenance buffers. As a result, failure of any one control device affects only a small part of the RTA complex. For example, a trunk buffer failure would affect no more than 16 trunks.

II. HARDWARE AND PHYSICAL DESIGN

The principal parts of the RTA hardware are the PCL, the concentrator, and the trunks. Each of these parts is described in this section. A test frame, including a 2-way TTY for maintenance messages, is also provided and is described briefly in Section 4.5. The principal parts of the PSS No. 2 hardware are the PCL and the 100C operator positions, and the 100C positions are also described in this section.

2.1 PCL description

The PCL is a general-purpose data link designed to automatically retry failing data words so that fault recognition programs will not be utilized for minor transmission errors. Four principal checks are made on each direction of transmission: parity, cyclic code, sequence check, and matching. In general, the data to be sent are applied to both PCL halves and treated independently by each half. Then, at the far end of the PCL, the data from both halves are matched, and a control bit is examined to see which PCL half will actually execute the order. Certain combinations of matching failures, cyclic code failures, sequence failures, and parity failures still permit valid execution of the transmitted orders, and these possibilities are discussed in Section 4.1.

2.1.1 Group Gate and remote data circuit

The Group Gate (GG) connects the PCL to the address bus, answer bus, and Central Pulse Distributor (CPD) of the TSPS processor. A Group Gate frame provides up to four fully duplicated Group Gates and the common bus input/output circuits to send and receive data to all four. This common bus circuitry initially accepts a 21-bit binary

word from the address bus and passes it to all Group Gates on the frame. An enable signal from the CPD then designates which Group Gate should respond to this particular 21-bit word, i.e., there is a separate enable signal for each Group Gate on the frame. There is also a common enable designating whether this order is "odd" or "even," and circuitry at the remote end continually checks that an alternating pattern of odd and even words is received.

The Group Gate also functions as a parallel-to-serial converter (in the sending direction) and thus converts the 21-bit parallel data into a 21-bit serial bitstream. The odd/even bit mentioned above is added, and a 2-bit "start" code is also added to designate the start of a new transmission and also to designate whether the new word is a long word (*data* from the TSPS processor) or a short word (*scan complete* signal from the processor). The resulting 24-bit serial word (long word) is then passed through a cyclic code generator which appends a 5-bit cyclic code. The resulting 29-bit serial word is fed to a 2400-b/s data set which performs a digital-to-analog conversion and connects to the transmission facility. (The "short word"—an acknowledgment signal from the processor that the last transmission from the distant end was received correctly—is 11 bits long as it is passed to the data set.) In normal (duplex) operation, both halves of a Group Gate receive the 21-bit word from the address bus and generate the 29-bit analog pulse streams to the two transmission paths. However, it is also possible to split the PCL into two (simplex) halves, and in this mode each half can be sent a different order from the processor.

In the receiving direction (Fig. 3), each Group Gate half receives an analog pulse stream into the 2400-b/s data set where it is converted to a digital bitstream. This is passed to a cyclic code check circuit and a parity check circuit. A processor-controlled activity bit designates one Group Gate half as "active" and the other as "inactive." Both halves

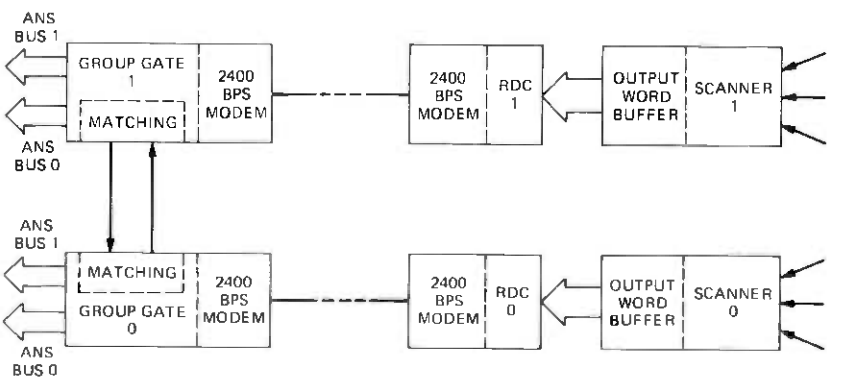


Fig. 3—PCL receive chain.

perform all the actions outlined above, but only the active half controls the matching operation. After the cyclic code and parity checks are performed, the active half moves data from both halves through a serial-mode matching circuit in the active half and determines if the match is valid or invalid. If cyclic code and parity tests have passed and matching is valid, then a signal is sent to the TSPS processor that the Group Gate has information ready to be passed on. If only the active half passed the cyclic code and parity checks and matching with the inactive half is *not* valid, then, again, the processor is signaled that information is present. A CPD pulse from the processor then causes that Group Gate to send its information on the processor answer bus.

For other combinations of invalid data and for transmission "burst" errors[2] affecting more than 1 bit and detected by the cyclic code check, a signal is sent to the remote end to retransmit the data. The processor is also informed that a retransmission or several retransmissions are taking place and, depending upon the severity of the problem, certain fault recognition routines may be executed. For example, the error activity might indicate a bad transmission facility on one PCL half, and the corrective action would be to switch to a spare transmission facility, or the error might be a simple transmission "hit" or "burst" which is cleared up by a retransmission, and no other maintenance activity is required.

The Group Gate can also receive processor data into a maintenance register which sets up any of several specialized maintenance states within a Group Gate half. For example, the cyclic code generator can be configured to purposely output bad codes so that subsequent PCL circuits can be exercised. The sending portion of a Group Gate half can also be connected back to the receiving portion of the same Group Gate half so that data from the TSPS processor can be operated upon by the Group Gate and returned to the processor as part of a diagnostic routine.

The Remote Data Circuit (RDC) is very similar to a Group Gate and, in fact, uses many of the same circuit packs designed for the Group Gate. However, it is located at the remote end of the PCL and connects to the signal distributor and the scanner. Like the Group Gate, it performs parallel-to-serial and serial-to-parallel conversions, parity checks, cyclic code generation and checking, odd/even sequence checking, and matching. The remote data handling circuits are mounted on a double-bay frame such that one side of the frame contains all the 0 half circuits and the other bay contains all the 1 half circuits. The two RDC halves, including their respective 2400-b/s data sets, each comprise one 6-in. mounting plate on the two bays of the frame. This frame, called the position and trunk control frame (for use both in PSS No. 2 and in RTA) is shown in Fig. 4.

Whenever no new data transmissions are available from either the

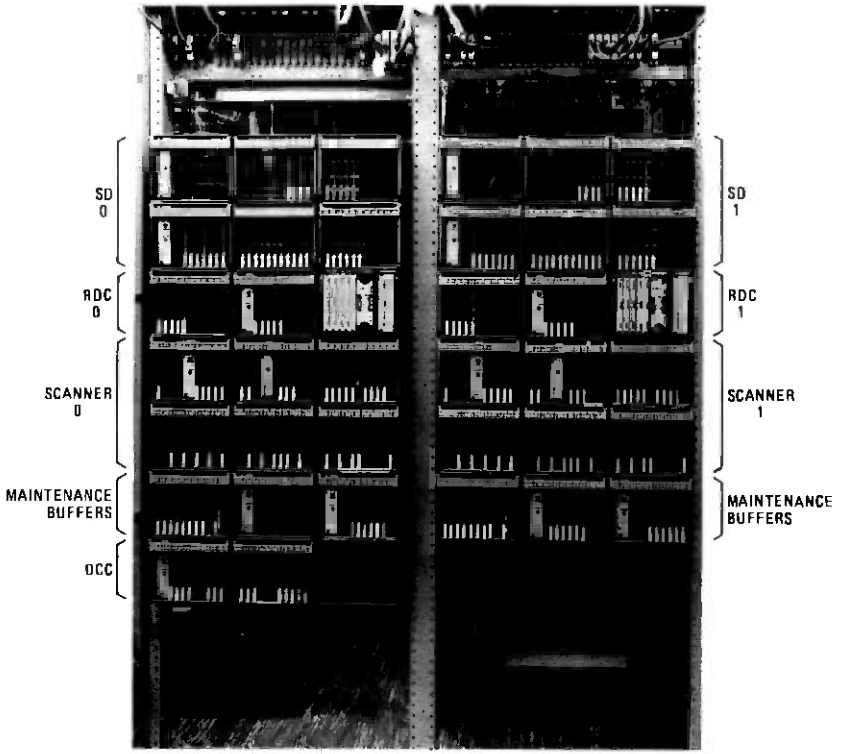


Fig. 4—Position and trunk control frame.

processor (for the Group Gate) or from the remote scanner (for the RDC), a dummy word is sent over the PCL. Thus, the PCL is sending information back and forth every 25 ms even during inactive periods. Both the Group Gate and RDC have the ability to recognize nonreceipt of such interchanges as a mechanism for identifying transmission facility troubles. In addition, the RDC, after several seconds of not receiving any words (dummy or data) from the Group Gate, will split the remote PCL end into two simplex halves. This action is based on the assumption of a trouble which has been recognized at the base (processor) location and has somehow prevented orders (to switch to simplex) from being executed at the remote location. The automatic switch to simplex operation then allows the fault recognition programs to exercise both halves and all three transmission facilities so as to find a good configuration.

2.1.2 Position and Trunk Scanner

The scanner is considered a part of the PCL, but introduces the concept of distributed control, since much of it is physically mounted on the trunk frames and 100C consoles with which it works. This

distributed concept means that only that portion required need be equipped, i.e., as a new trunk frame is added to an office, another piece of the scanner is essentially added to the scanner. This new piece of the scanner was furnished as a part of the trunk frame just added and deals with supervisory signals from those added trunks. It is then connected to the rest of the scanner which is provided as two independent halves, each reporting changes of state to the processor at the TSPS base location. At the base location (in the Group Gates, as described in Section 2.1.1), the two scanner reports are matched before actual transmission to the processor. Mismatch errors are a prime means for detecting scanner malfunctions (see Section 4.1 below).

The scanner is an autonomous circuit continually looking for transitions such as changes in trunk supervision, depression of a key on a position, or test frame and system alarms. It is driven by a clock signal obtained from the 2400-b/s data set used by the RDC and both scanners are run from a single data set clock. However, this clock signal is continually checked and both scanners will switch to the other data set (RDC) clock in case of trouble. For simplex operation, the two scanners each connect to the corresponding individual RDC clocks. Minute differences in the two RDC clocks require a period of synchronization as two simplex scanner halves are reconfigured into a single duplex system.

The scanner examines some 1300 scan points every 12.3 ms and records the state of each in Random Access Memory (RAM). In the *next* 12.3-ms cycle, the previously recorded information is known as the "Last Look" (LL). At any instant of time, the current state of a scan point on which the scanner is currently acting is known as the "Present Look" (PL). The scanner also keeps track (in RAM) of the previously *reported* state of each scan point. This is known as the "Previous State" (PS) and is the state last reported to the TSPS processor. As the scanner examines each point, it compares the PL to the LL and PS. If the PL and LL agree *and* are different from the PS, then a report is made and PS is changed to the new value. The requirement that two scans spaced 12.3 ms apart must agree allows the scanner to disregard all transient conditions producing pulses shorter than 12.3 ms and guards against line and supervisory transients being reported as valid trunk states.

Although both scanners are running on a single clock during duplex operation, there is a low probability chance that one scanner half will see a scan point transition one scan interval ahead of the other. Then, on the next scan, one scanner will make a report while the other scanner has seen the transition for the first time and will not make a report. On the third scan, the second scanner would make its report and from that point on the internal scanner states should be in

agreement. But in the meantime, a mismatch at the Group Gate may have been detected and some fault recognition activity would be triggered needlessly since the mismatch does not indicate a hard fault. To prevent this transient mismatch, a match on the present look between the scanners is performed. If a mismatch occurs at this point, then both scanners ignore the present look in this scan period. On the second scan, if the mismatch was not a hard fault, both scanners will agree and a report will be made in the normal fashion. If the mismatch persists through the second scan, both scanners are again allowed to proceed in a normal fashion but in this case a mismatch will be generated and the hard fault will be detected. A single bit per scan point is required to count the present look mismatches (PLMM) and this bit is kept in RAM along with the LL, PS, and a parity bit for RAM self-checking.

The scanner presents reports to the RDC which adds cyclic code and converts to a serial analog pulse stream. The RDC accepts the scanner reports at a maximum rate of one every 25 ms, depending upon data line activity. However, the scanner *input* rate is based on whatever activity is happening at some interval of time and could conceivably be much greater than a 25-ms rate for brief periods of time, i.e., bursts of operator activity whereby several operators simultaneously push keys or RTA trunk activity involving dial pulse reception on several trunks. Also, the 25-ms output rate may be slowed due to transmission line errors which require retransmissions. To cope with such activity, the scanner contains a 32-word output buffer so that reports can be stored temporarily until they are transmitted. Calculations indicate that, even with bursts of operator and trunk activity, a 32-word buffer will almost always guarantee that no information will be lost.[4] Exceptions can occur during severe (but very infrequent) maintenance activity, when scanner reports accumulate while fault recognition is running maintenance tests under partial outage conditions.

A feature of the scanner is its ability to count dial pulses on the incoming (local office) side of every trunk, recognize the interdigit interval, and report dialed *digits* to the TSPS processor. It also performs timing checks on initial state changes from the local office side of a trunk to recognize "false seizure" conditions so as to not report such false seizures (these can persist for up to 40 ms). These features are based on defining several possible trunk states such as initial seizure, start of open interval of a dial pulse, start of closed interval of a dial pulse, start of interdigital interval, end of pulse, end of digit, disconnect, etc. A trunk progresses from state to state by timing the number of 12.3-ms cycles in which it remains in an on-hook or off-hook condition. Thus, each time the trunk condition changes, a count of 12.3-ms cycles is started (by incrementing a 4-bit count—4 RAM bits—assigned to the

incoming side of each trunk). The actual count reached before the on-hook/off-hook condition changes determines the next state of the trunk. For example, consider a trunk receiving dial pulses which, after a few pulses of a digit, starts counting the 12.3-ms cycles of an off-hook interval. If 2 to 10 cycles are counted and then the trunk condition changes to on-hook, this is considered another pulse of the digit, and the on-hook interval is timed to confirm this. However, if 11 or more (off-hook) 12.3-ms cycles are counted, this is considered an interdigital interval and the previously counted pulses now represent the digit to be reported.

The items scanned by the scanner include trunk supervisory inputs, 100C console keys, and alarm indications from all remote circuits. The reported information consists of scan point type, unit number, and the particular status change of that unit (such as operation of the START TIMING key at a position). A parity bit and a sequence number are added by the scanner. The sequence number is a continuously incremented 2-bit binary number (0-1-2-3-0-1-· · ·) indicating the order in which reports are loaded by the scanner into its 32-word output buffer. The RDC adds a start code, a 5-bit cyclic code as described for the Group Gate in Section 2.1.1, and sends the report back to the TSPS processor. At the processor, this sequence code from the scanner is continually checked and any out-of-sequence condition is a trigger for fault recognition action. This guards against a report being lost with no indication to the processor that the lost report ever existed.

The scanner is constantly performing self-checks as it scans across certain scan points reserved for maintenance self-checks. For example, one scan point is arranged to give on-hook/off-hook sequences that should represent a certain digit count and other circuitry within the scanner checks that this count has been reached. These checks, in addition to parity checks across the RAM, influence an "all-seems-well" bit returned with each scanner report and are constantly examined by the TSPS processor. A failing "all-seems-well" bit triggers immediate fault recognition action.

2.1.3 Signal distributor

The signal distributor takes parallel information from the RDC, strips off the unit designation and unit type information (such as position number, trunk buffer number, concentrator controller number, etc.), and distributes the remaining portion of the order to the designated unit. In performing this distribution, it converts the parallel output of the RDC to a serial bipolar bitstream and adds parity and start bits. The resulting serial word is transmitted over a single pair of wires to an RTA trunk buffer, a 100C console (PSS No. 2), a concentrator controller, a maintenance buffer, or to the test frame. Thus the signal distributor communicates with 70 to 80 units of five different types.

(For a PSS No. 2, this consists of 62 positions, 5 maintenance buffers, 2 test frame buffers, and 1 self-test buffer. For an RTA, this consists of 31 trunk buffers, 5 maintenance buffers, 2 test frame buffers, 40 concentrator controllers, and 1 self-test output). The serial bit-bipolar transmission allows these units to be several hundred cable feet from the signal distributor.

After the serial word is constructed within the signal distributor and loaded into an output shift register, it is then shifted out and, at the same time, looped back into the output register. Thus, when shifting (transmission) is complete, the transmitted word is again in the output register and is rechecked bit by bit for accuracy. Both this check and a subsequent flip/flop activity check made by the receiving unit (100C console, trunk buffer, etc.) influence the check-back signal sent from the signal distributor back to the RDC, and subsequently to the SPC, as an overall check. This shifting action also allows calculation of a parity bit "on-the-fly," i.e., as the information and start bits are shifted out, parity is calculated and, if the shift register recheck passes, the correct parity bit is appended to the end of the serial transmission. Thus the parity bit can be purposely "failed" to keep the receiving unit from using questionable data.

The start code not only identifies the start of a new transmission but also identifies the *type* of unit expected to receive the new information. Each receiving unit contains a start code mask that must match the start code to accept the information bits. Each receiving unit (buffer) also generates either a "check-back" or a "buffer failure" response to the signal distributor each time it receives new information. This response is returned to the distributor over the same pair of wires used to send information to a buffer. This response, together with the signal distributor's internal checks, is used to send a check-back signal to the RDC and from there to the SPC. If the SPC fails to receive this check-back, fault recognition action is triggered, starting with a simple retry and escalating to subsystem reconfigurations if necessary.

The duplex nature of the sending portion of the PCL effectively ends at the RDC and, once RDC matching is complete, only one of the duplicated signal distributors executes the order. The distributor chosen by the RDC is changed on every order to utilize both distributors evenly. If the PCL is split into two independent halves for maintenance purposes, then maintenance orders can use one signal distributor while call processing orders use the other.

The orders to light LED numerical displays in 100C consoles place special requirements on the signal distributor. These orders are always sent as a sequence of 2 to 7 consecutive SPC orders, but only the first contains the address of the desired 100C console. The signal distributor recognizes such orders (usually called "prime" orders since they prime the distributor and the 100C console to expect a sequence) and tem-

porarily stores the position number while passing the rest of the order to the desired position. Subsequent SPC orders will contain pairs of digits to be selected on the console LED display but do not include a position number. Since only one signal distributor executed the "prime" order and remembers the position number, that one distributor handles all of the rest of the sequence and the alternating use of both signal distributors is temporarily suspended.

The "prime" order also informs the distributor as to which of several *types* of LED displays will be forthcoming; e.g., a 12-digit (including blanks) display of a called number, a 1-digit display of elapsed minutes, a 6-digit display of the time-of-day, etc. The distributor includes a 6-unit ring counter that chooses the location, within the operator's numerical display, where the pair of digits just received are to be placed. This ring counter is preloaded by the prime order to start the first pair of digits (next SPC order) in the correct position on the 100C console. The ring counter is then stepped along by each subsequent order so that it is always ready to direct the digits to the proper place. As the signal distributor forms the 23-bit serial word to the 100C console it takes the digit pair *values* from the RDC, the pair *location* from the ring counter, and the 100C *address* from its stored memory of that address, and adds a parity bit computed on-the-fly as the other information goes by.

2.1.4 Diagnostic control circuit

The diagnostic control circuit (DCC) is also considered part of the PCL as shown in Fig. 2. It allows implementation of a remotely run diagnostic concept explained in Section 4.3. The overall description, together with its control by the SPC and its considerable access to remote PCL circuits, are all included in Section 4.3 as part of the RTA maintenance plan.

2.2 Concentrator

2.2.1 Size, ratio, components

The RTA concept is to provide TSPS services for a sparsely populated area from a TSPS in a more built-up area. Therefore, many RTA installations consist of 200 to 400 trunks arranged in many small trunk groups with relatively low occupancy. Initially, only 100 to 150 trunks may be required, with growth to 300 to 400 taking place over several years. A few RTA sites will have larger trunk groups and correspondingly higher occupancy and a few will have trunks exceeding the capacity of a single RTA and use two RTAs in the same building.

Early studies of the potential RTA market and of trunk usage led to the conclusion that the base-remote trunks (from the output side of the concentrator to the controlling TSPS base installation) will be connected to a call for about one-eighth of the total holding time.[7]

Several concentration ratios were evaluated based on trunk numbers mentioned above, and the 8:1 concentrator was selected as the best compromise. The compromise was such that a slightly more expensive concentrator is initially provided but growth, on both the basic concentrator and the subsequent addition of a build-out frame, is easily accomplished. The 8:1 concentrator works well with lightly loaded offices, allows up to 496 incoming trunks, and provides for flexibility in load balancing. With a full complement of base-remote trunks (up to 64), it also works well in moderately heavily loaded offices.

The desire for a low-cost, easily-added-to, and moderately-long-holding-time switch led to use of miniature crossbar switches as the principal concentrator element. In RTA use, the select mechanism of the crossbar switches will accumulate its lifetime operations (estimated at 20 to 25 million operations) in 7 years. Ease of replacing this mechanism plus widespread operating company familiarity with such work also influenced the crossbar switch choice.

2.2.2 The logical concentrator and the physical concentrator

The logical layout of the concentrator is used to understand the connection pattern and to assign switch numbers useful to the TSPS programs controlling concentrator operations. However, the physical construction of the concentrator makes use of 20 × 10 crossbar switches arranged in a rather specialized way to achieve the 16 by 8 and 32 by 8 switches of the "logical" concentrator. This section explains the relationship between these two views of the concentrator.

The concentrator may be obtained in two sizes: a "small" size, able to connect up to 248 remote incoming trunks to as many as 64 base-remote trunks, and a "large" size, which can handle up to 496 remote incoming trunks. In each size, every remote trunk has full access to each of the up to 64 base-remote trunks. The term "remote trunk" is used broadly here to mean any trunk circuit which appears on the inlet side of the concentrator, while the term "remote incoming trunk" means specifically an RTA incoming trunk. A variety of possible remote trunks besides remote incoming trunks includes: hotel/motel auto-quote trunks, inward trunks, and test appearances. With the exception of test appearances, if any of these other types of remote trunks appear on the concentrator, they must replace remote incoming trunks so that the maximum number of the latter is reduced accordingly.

2.2.2.1 The Logical Concentrator. Figure 5 is a simplified depiction of the Logical Concentrator. In the small sizes, the first or trunk stage consists of thirty-two 16 by 8 three-wire crossbar switches, while the second or base-remote stage has eight 32 by 8 three-wire switches. The small Logical Concentrator thus has 512 inlets (2 for each of the 248 remote incoming trunks, 2 for test appearances, and 14 spares), 64 outlets (one for each of the 64 base-remote trunks), and 256 links (one

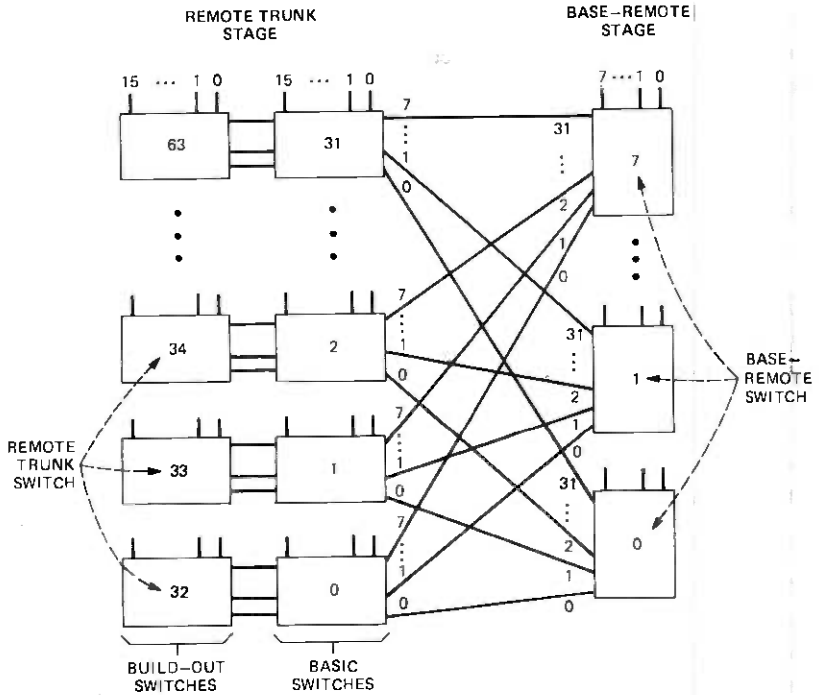


Fig. 5—The logical concentrator and its numbering.

from each of the 32 trunk stage switches to each of the 8 base-remote stage switches).

The large Logical Concentrator is obtained by adding a 16 by 8 three-wire buildout switch to each of the trunk stage switches of the small Logical Concentrator. It thus has the equivalent of thirty-two 32 by 8 trunk stage switches and eight 32 by 8 base-remote stage switches for a total of 1024 inlets (two for each of the 496 remote incoming trunks, two for test appearances, and 30 spares), 64 outlets (one per base-remote trunk), and 256 links.

It is important to note that there is only one possible path or "channel" from a *particular* remote trunk appearance to a *particular* base-remote trunk. By comparison, the base TSPS network has eight possible channels from a given inlet to a given outlet.

To make a connection in the Logical Concentrator from a particular remote trunk to a particular base-remote trunk, a *single* crosspoint is activated on the trunk stage switch containing the remote trunk and another *single* crosspoint is activated on the base-remote stage switch on which the base-remote trunk is terminated. The first crosspoint connects the remote trunk to the link going to the desired base-remote stage switch, and the second connects that link to the proper base-remote trunk.

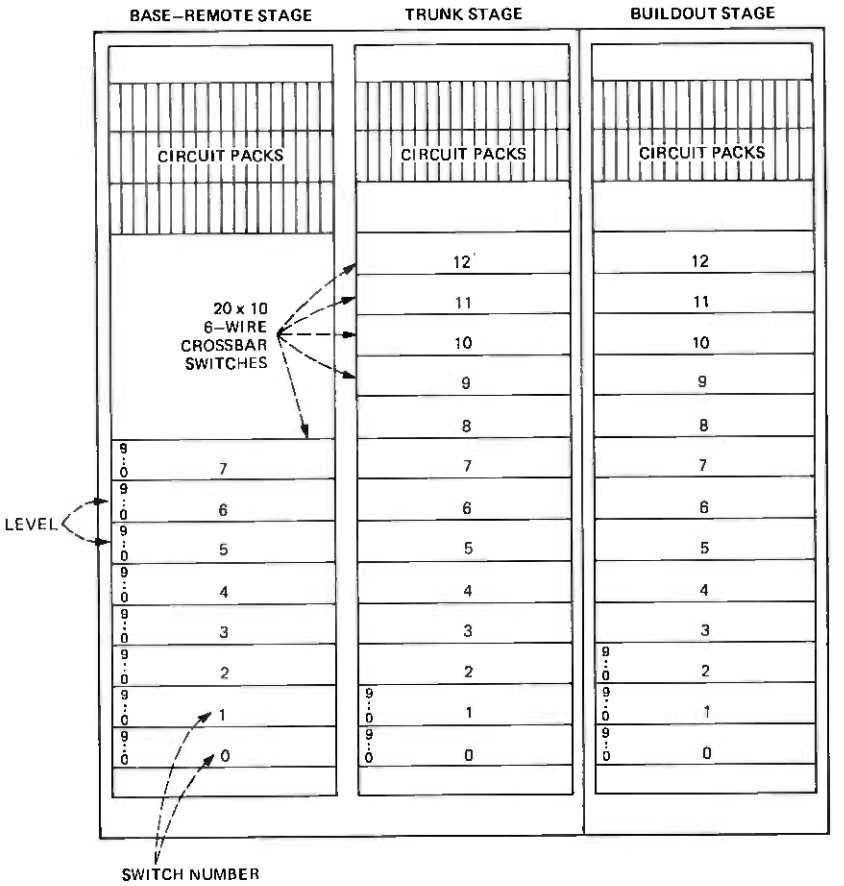


Fig. 6—The physical concentrator frames and their numbering.

2.2.2.2 The Physical Concentrator. Because of limitations on the type of crossbar switches that are available, the actual hardware implementation of the concentrator is quite different from that described above. This hardware implementation is called the Physical Concentrator and is shown in Fig. 6. The small concentrator consists of two bays of crossbar switches. The first, or trunk stage bay, contains thirteen 20 by 10 six-wire crossbar switches, while the second or base-remote stage bay contains eight 20 by 10 six-wire crossbar switches. The large concentrator is obtained by adding a second frame, similar to the trunk stage bay above, called the buildout frame which contains thirteen 20 by 10 six-wire crossbar switches.

To make the 20 by 10 six-wire switches in the physical implementation of the RTA concentrator function logically like 16 by 8 three-wire switches, a rather complex design has evolved. The first step, that of making six-wire switches behave like three-wire switches, is per-

formed using a well-known technique[5] by which levels 0 and 1 are utilized to translate which three of the particular set of six wires are to be used. The result of this translation is that the 20 by 10 six-wire switches operate like 20 by 16 three-wire switches. These are then combined in a fairly complex way to form a unit which operates like the Logical Concentrator.

Thus, in the Physical Concentrator, to make a connection from a particular remote trunk to a particular base-remote trunk, two crosspoints must be activated in the trunk stage switch on which the remote trunk is terminated and two more crosspoints must be activated in one of the base-remote stage switches to which the base-remote trunk is connected. The first of these crosspoints is located in levels 2 to 9 and connects a pair of remote trunk appearances (six wires) to the link going to the desired base-remote stage switch. The second crosspoint is located in levels 0 or 1 and selects which of the pair is desired. The third and fourth crosspoints perform similar functions in the base-remote stage of the concentrator.

2.2.3 Distributed control

All the concentrator switch select and hold magnets are operated by controller circuits arranged on three plug-in circuit packs. The eight groups of packs used for control of base-remote stage switches are physically mounted on the concentrator frame and are always provided. However, the groups of circuit packs controlling each trunk stage logical switch (up to 32) are provided as part of the trunk buffer circuit. An RTA office can have up to 32 trunk buffers, 31 of which can each contain 16 trunks. Thus, as each group of 16 trunks is added to an RTA, a corresponding piece of the concentrator (the three controller circuit packs) is also added. The control of the trunk stage portion of the concentrator is thus distributed over all RTA trunk buffers arranged on up to five separate frames. Some cost saving is obtained by this arrangement, since only the portion of concentrator control actually required by trunks present in the office are purchased. However, of much more importance is the distribution of this control over many equipment units and the resulting overall reliability obtained; i.e., problems affecting a group of trunks affect only that part of the concentrator serving those particular trunks.

The SPC orders distributed to a particular concentrator controller by the signal distributor specify the switch number desired, the particular select and hold magnets to be used, and the function (connect or disconnect). A single trunk stage controller controls two logical switches, since each RTA trunk has two appearances on the concentrator. These two appearances (local office side and toll office side) are provided so that an outpulser connection to the toll office can be set up while an operator connection to the calling customer is established.

[Thus the 8:1 concentration ratio is approximately the ratio of incoming trunks (up to 496) to base-remote trunks (up to 64), while the ratio of trunk appearances is actually 16:1.] In general, the sequence of control orders to establish a concentrator connection is as follows: Trunk stage connect order, followed by a base-remote stage connect order, followed by operation of relays in the selected base-remote trunk. A verification test later performed by the base-remote trunk circuit results in establishment of a control path via the concentrator sleeve lead which locks the trunk stage hold magnet to the base-remote stage hold magnet. Connections are broken by a base-remote stage disconnect order which resets the base-remote stage hold magnet flip-flop and in turn releases both the base-remote and trunk stage hold magnets.

2.3 RTA trunks

The general approach to trunk design for electronic offices has been to keep the trunk simple and temporarily connect it to more sophisticated equipment for such things as digit reception, ANI reception, coin control signal generation, and outpulsing. That this requires many trunk and link orders for each call is of little concern when the trunks and their associated scanners and signal distributors are close to the processor and reached over high-speed buses. However, when the trunks are several hundred miles from the controlling processor, and when excessive numbers of orders can exceed the capacity of relatively slow-speed data paths, the simple trunk concept is not so appropriate. Thus the RTA trunks tend to be more complex than their base-location counterparts and include several features on a per-trunk basis which might otherwise be provided by a separate pool of circuits.

Another feature of each RTA trunk is that it is packaged on a single 5 by 7 in. circuit pack. The single Printed Wiring Board (PWB) construction is made possible by using miniature wire-spring relays, miniature mercury relays, miniature networks, and a specially designed integrated circuit. A typical trunk is shown in Fig. 7. The custom integrated circuit handles several signaling tasks, including coin control signals and ringback signals sent from TSPS to the local office.

The coin control and ringback considerations, including the ability to disable or enable a station set *TOUCH-TONE*® dialing pad at certain points in a call, require that five separate signals be sent from the RTA trunk to the local office. A custom-made IC provided in each trunk provides these "multiple-wink" signals. The ring-forward signal to the toll office is also provided within each RTA trunk. The custom integrated circuit uses a 3-bit binary input set as a command to generate 1 to 5 pulses, each of which turns on (operates) a pair of mercury relays for a timed interval. This pair of mercury relays has contacts arranged in the tip-and-ring leads to provide a reversal to the

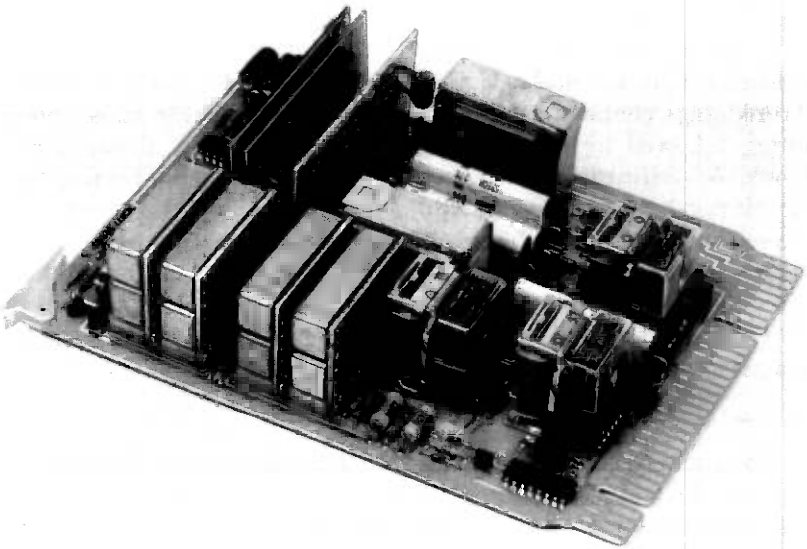


Fig. 7—Typical RTA trunk.

local office. Thus, 1 to 5 timed reversals (winks) can be sent from an RTA trunk. Since this uses only six of the eight possible input states (OFF and one to five winks), the other two are used to operate other relays including a set of "ring-forward" mercury relays which apply +130 V to the toll office side of the trunk. The 3-bit binary input comes from the trunk buffer which generates the 3-bit output from a single SPC order.

With the exception of the use of miniature components and the custom integrated circuit mentioned above, many other talking path and relay states in RTA trunks are similar to previous electronic office trunk designs. The 4-wire trunks utilize the TSPS 3-way 4-wire bridging repeaters mounted on separate repeater bays together with the PWB on the RTA trunk frame. A group of 16 trunks, together with a trunk buffer and appropriate portions of the concentrator controller (Section 2.2) and of the scanner (Section 2.1.2) comprise a *trunk unit.* This trunk unit has each component part arranged to control only the 16 trunks of that unit and is thus completely separate from any other trunk unit. A malfunction in any portion of the trunk unit will thus affect, at most, 16 trunks.

The trunk buffer is similar to all the buffers to which the signal distributor distributes SPC orders and consists essentially of a flip-flop matrix and the necessary control circuitry to set and reset each flip-flop. Output drivers connected to each flip-flop then operate relays in the trunks or send signals to the multiple-wink integrated circuit in

each trunk of this particular trunk unit. The trunk buffer is provided with a "super-initialization" input so that, under severe maintenance conditions, an entire buffer or all buffers in the office can be cleared by a single pair of SPC orders. This initialization path is arranged so that all trunks are cleared *except* those trunks in the cut-through (talking) state. Such trunks are *assumed* to be cleared by the initialization routines but are protected by a circuit design that inhibits the action of the clear signal when in the cut-through state. The corresponding program actions are discussed in Section 3.3. This approach allows rapid system initialization when necessary, while at the same time protecting all customer-to-customer talking paths.

2.4 100C position

As mentioned at the start of this chapter, the Position Subsystem (PSS) No. 2 is only briefly mentioned. By adding 100C consoles to the PCL described in Section 2.1, together with some voice path switching and supervisory circuits, a PSS No. 2 is obtained. The 100C position is treated by the PCL signal distributor as one of several "buffer" circuits to which it distributes orders. The single exception to this is the *sequence* of orders required to light up a numerical display (1 to 12 digits) on the 100C console (see Section 2.1.3). A console consists of two positions, and throughout the TSPS literature "100C console" and "100C positions" are used interchangeably.

The upper portion of the position consists of two large keyshelf printed wiring boards and the LED numeric display PWB. Each keyshelf board contains 30 to 50 key/lamps and some integrated circuits and discrete components associated with those lamps and keys. The single "make" contact of each key connects to a cluster of three diodes, which are in turn connected to three leads of a 9-lead bus. Thus each key generates a 3-out-of-9 code on this 9-lead bus whenever it is depressed. The 9-bit bus is examined by the scanner at regular intervals as discussed in Section 2.1.2. The numerical display PWB contains twelve 7-segment LED numerics and the memory and driving electronics associated with them.

The lower portion of the 100C position contains a dc-to-dc converter and several circuit packs of flip-flops, lamp order buffers, and control circuits. The control circuit decodes the serial bitstream from the signal distributor and determines whether it is part of a numerical display sequence or a single lamp order. The numerical display orders are delivered with little change to the LED display PWB. The lamp orders are formed into a vertical select, horizontal select, and group select output set to address a flip-flop matrix. The result is the setting or resetting of a single flip-flop and generation of a check-back signal to inform the signal distributor that the correct conditions for accessing

a flip-flop were present (a sort of all-seems-well signal). In many cases, the flip-flop connects directly to a console lamp and simply lights or extinguishes that lamp. However, in some cases a console lamp may be placed in a flashing state and in such cases the flip-flop output connects to a lamp order buffer (LOB). These LOBs generally have two flip-flop inputs and a single lamp driver output. One input is used for a steady lamp "on" condition, while the other is used to connect the lamp to a 60 or 120 IPM interrupter circuit. On a typical call, the TSPS processor might initially send several orders to a position causing three or four lamps to light up and then light and extinguish two or three more in the course of the operator's involvement on the call. For coin calls, part of the numerical display will also be lighted, showing the charge and initial period interval for each call.

In summary, the console contains all the lamps, lamp drivers, memory cells (flip-flops), and overall controls to turn on and hold on the various combinations of lamps required for each call. It contains the circuitry to encode up to 84 keys in a 3-out-of-9 format for the scanner. It also contains duplicated input (signal distributor) and output (scanner) capability and the ability to permit input orders to generate output codes directly, with no operator present, for diagnostic purposes.

2.5 Voice circuits

Although the PCL data lines are voice-grade circuits, they are used to transmit only data between the base and remote locations. Other voice-grade circuits provide connections between the RTA trunks and the base office (base-remote trunks), provide for operator talk paths (for PSS No. 2), and also connect the remotely located maintenance teletypewriter to the base location. Standard 4-wire voice-grade circuits are needed for base-remote trunks, and any type toll-grade carrier system with groups of either 12 or 24 circuits may be used.

For reliability, maximum diversification in the carrier groups, including a spare group, is required. This means that the carrier groups should be evenly distributed over two or more carrier route facilities and, within a route, the carrier groups should be distributed as much as possible over independent terminal equipments and power supplies.

Carrier transfer circuitry is provided at the base and remote locations to switch the spare group into service in place of any other group that fails. Switching is controlled by a program at the base location which requires carrier group alarm signals as an input. Since there is only one spare group, loss of more than one carrier group results in partial facility loss and possible delays in handling calls from that RTA.

III. RTA OPERATIONAL SOFTWARE IMPLEMENTATION

When it was first proposed, the RTA operational software was visualized as being relatively straightforward. At the very beginning of an RTA call, a pair of base-remote trunks would be seized and used to "extend" the remote trunk to the base network. The appearances of the base-remote trunk would be arranged to be the software equivalent of those of a base universal trunk. Thus, throughout the succeeding stages of the call, the existing TSPS operational software[6] could be used without significant modification until, when the operator was released, the base-remote trunks would be disconnected and idled.

Early in the planning, however, it was realized that this was not feasible. Not only was it uneconomical to leave both base-remote trunks up throughout the early stages of a call, but economics also dictated that trunk supervisory states be sensed and controlled via the PCL rather than by sensing and controlling the state of base-remote trunks. Moreover, the technology and complexity of the RTA trunks (see Section 2.3) caused the sequence of orders required to control them to be wholly different from that of base trunks so that the software which constructs trunk orders for base trunks could not be used. Also, the structure of the TSPS operational software did not allow the type of changes required for RTA call processing to be made simply or in a small number of low-level modules.

The result was that the RTA operational software design was forced to evolve away from the initial concept of a few relatively large new modules with most of the existing operational software left intact. Instead, many small changes had to be made to most operational programs in the system. These changes were inserted at points where a position or service circuit is connected or released and at points where trunk control orders are formulated. Typically, the changes involved the addition of a software switch based on a test of the location (base or RTA) of the trunk under consideration. This test is facilitated by bits at several points in the trunk parameter register which have a value of 1 if the trunk is an RTA trunk and 0 otherwise. If the trunk is at the base, the switch directs execution of the previously existing program code. If not, a new RTA program leg is used to handle the call.

Thus, most RTA operational software is relatively straightforward and will not be discussed further. However, several program areas are new to RTA and have no analog in the previous TSPS software. These include the control of the RTA concentrator, the base-remote trunk queuing structures, the "virtual" scanner used to maintain the state of remote trunk supervisory scan points, and a new type of audit. Each of these is discussed in the following sections.

3.1 Concentrator control

During a normal TSPS call, various connections must be made to positions and service circuits. If the call is on an RTA trunk, such a connection requires the selection of:

(*i*) An idle base-remote trunk and an idle path from the RTA trunk through the concentrator to that base-remote trunk and
(*ii*) An idle position or service circuit and an idle path through the TSPS base network from the base-remote trunk to that position or service circuit.

Ideally, the path selection routine should hunt both (*i*) and (*ii*) simultaneously in one unified procedure. Such a procedure would be very complex because it would have to consider the RTA concentrator, the set of base-remote trunks, and the base network as one large, complex, six-stage network (two stages at the RTA and four at the base). However, since the probability of success on (*ii*) is virtually independent of the choice made on (*i*), little effectiveness is lost by hunting the path sequentially: first selecting (*i*) and then (*ii*). This allows one potentially complex single procedure to be replaced by two simpler sequential ones. The first of these is handled by the Concentrator Control program—a new program for RTA—while the second is the previous Network Control program with slight modifications.

Also used with slight modification is the path memory annex register,[6] which is linked to the call register during a network connection. It is used to store the identity of the connected circuits, and the paths used, their states, and, in the case of RTA, the concentrator paths and base-remote trunks used and their states.

The Concentrator Control program is a subroutine capable of hunting, connecting, and breaking paths from RTA trunks through the concentrator to the base via base-remote trunks. It has an interface with the other call processing programs which is similar to that of the Network Control program. Like the latter, it has a set of linkage maps, shown in Fig. 8, which are used in the calculations.

Conceptually, hunting and selecting a path for a call through the RTA concentrator, like any connecting network, may be done in two steps:

(*i*) Find all idle paths from the concentrator inlet of a specified RTA trunk appearance to the base network. If no idle paths exist, the call is said to be "blocked." By "idle path" is meant a combination of an idle base-remote trunk and an idle link connecting the trunk stage switch containing the inlet to the base-remote stage switch containing the base-remote trunk.

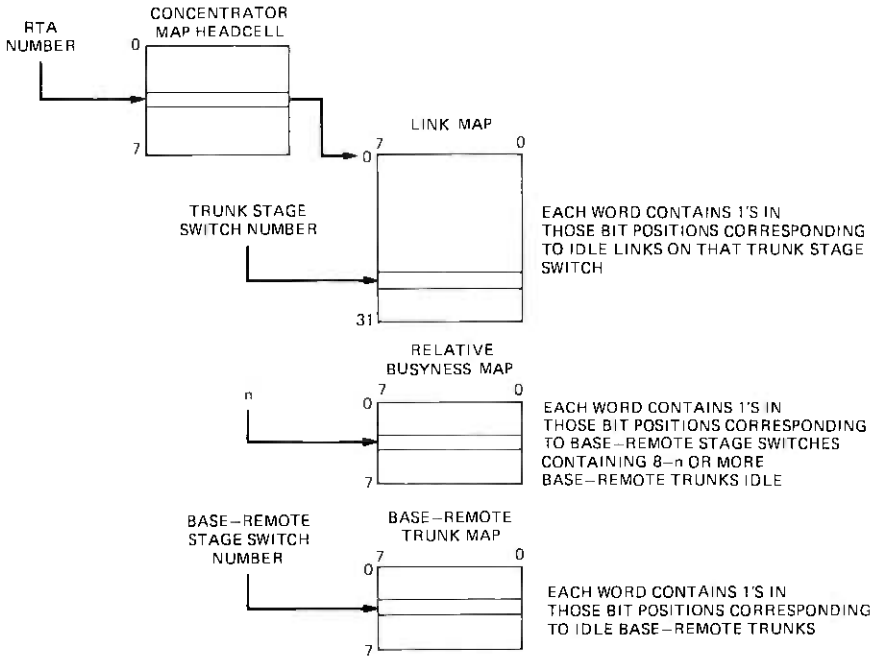


Fig. 8—Concentrator maps.

(*ii*) Select one of these idle paths to be used for the call. The procedure used to do this is called the "routing procedure."

The choice of a routing procedure is important because it affects the blocking probability of the network—the probability that the call is blocked at step (*i*) above— and hence the amount of traffic that the network can carry. One requirement of a routing procedure is that it equalize the long-term mechanical wear of the network crosspoints. Thus, a good routing procedure will minimize network blockage while equalizing crosspoint wear.

An unusual feature of connecting networks such as the RTA concentrator is that none of the routing procedures commonly used for other networks, such as random selection (which is presently used in the TSPS base network) or packing (which is used in the No. 5 and other Crossbar Systems), is a particularly good choice. However, it may be shown[7] that a near-optimal routing procedure is to select the path which goes through the base-remote stage switch which contains the largest number of idle base-remote trunks (with ties broken at random). The base-remote trunk to be used should also be selected at random. Studies show that use of this procedure increases the traffic capacity of the group of base-remote trunks by up to 3/4 Erlang at the

engineered load (about 40 Erlangs for a typical, large RTA) compared to the random routing procedure. This means that one less base-remote trunk is typically required—a substantial saving.

The actual implementation of the path hunting procedure integrates the two steps above into one algorithm. First, a test is made to see if all base-remote trunks are busy. If they are, the call is placed on one of the base-remote trunk queues as described in the next section. If one or more base-remote trunks are idle, the routine reads the proper link busy/idle word from the link map (Fig. 8) and forms the logical product of it with word n of the relative busyness map for n as small as possible such that the result is nonzero. A binary search is used to quickly find the proper value of n. If no such n can be found, the call is blocked. If the call is not blocked, then one of the 1's in the result is selected at random, the corresponding word of the base-remote trunk map is read, and one of the 1's in it is selected at random. The two 1's which are selected determine the path to be used by the call. To connect the path, the program loads the appropriate concentrator control orders in an output buffer, updates the concentrator maps, writes the path information into the path memory annex, and then terminates.

3.2 Base-remote trunk queuing

With RTA, a new group of hardware resources, namely base-remote trunks, is added to the TSPS system. One of these resources is routinely applied to a call at the same time as other resources such as operator positions and service circuits. In addition, on some calls, a connection must be made to two base-remote trunks at the same time (on dial-0, dial pulse, non-ANI calls, for example). And most calls have two base-remote trunks up simultaneously at some point in the call (for position and outpulser). Although the normal engineering of these relatively expensive resources is adequate, the supply may be cut in half by an outage of one of the two diverse routes which the base-remote trunks traverse. Since the base-remote trunks are normally engineered near or over 50 percent occupancy, such an outage can instantly throw the RTA into heavy overload. Although rare, this condition is the most likely RTA failure state because of the excellent reliability of the rest of the equipment.

These facts imply that the group of base-remote trunks must be carefully administered. Also, it was desirable to cleanly integrate the queuing for base-remote trunks into the existing TSPS queuing structures which, although they are basically quite powerful and flexible, have the built in assumption that only one resource may be queued for at a time.

The solution chosen is to have two first-come, first-served base-remote trunk queues served in priority order for each RTA.

If all in-service base-remote trunks are busy, calls wait in one of these queues. When a base-remote trunk becomes idle and available for use, it is connected to the first call on the high priority queue. If there are no calls waiting on the high priority queue, the base-remote trunk is connected to the first call on the low priority queue.

Generally, if no base-remote trunks are available the first time in a call when one is needed, the call waits on the low priority queue. Thereafter on that call, the high priority queue is used whenever base-remote trunk congestion is encountered.

This structure has several advantages. First, it gives priority to calls already in the system on which considerable system resources and real time have already been invested. Also it fits well into the existing "one resource per queue" TSPS queuing structure. For example, if a call needs two base-remote trunks at the same time and none are available, the call waits on the low priority queue until it gets one. Then it waits on the high priority queue and it will get the very next base-remote trunk that becomes available (the fact that it received one base-remote trunk while on the low priority queue means the high priority queue was empty). Thus, the call gets two base-remote trunks in slightly more than one waiting time.

Under conditions of RTA overload (caused by, say, a traffic overload or a route outage affecting up to half the base-remote trunks), simulation studies show that the low priority queue can become quite long while the high priority queue remains very short, usually with either zero or one call in it. In this case, every RTA call waits in the low priority queue exactly once, and thereafter uses the very short high priority queue. The result is that every call receives a roughly equal grade of service no matter how many base-remote trunks it needs during its course. (It is impossible for the high priority queue to become very long in the steady state, for if it were to become long, then the low priority queue would never be served and no new RTA calls would ever get into the system—a contradiction.)

The same simulation studies show that, with this queuing structure, the probability of deadlock is *extremely* low. An example of deadlock is as follows. Consider an idle RTA with N in-service base-remote trunks. Suppose that exactly N dial-0 calls arrive nearly simultaneously. Each would be connected to an operator using a base-remote trunk, thus making all such trunks busy. Each operator would then key in the called number and attempt to initiate outpulsing. Since outpulsing requires a second base-remote trunk, all N calls would be loaded onto the high priority queue. Since no base-remote trunks are available, all N calls will be deadlocked until one of the operators or customers abandon. Because the probability of this occurring was determined to be very low, no defensive measures have been added to the programs.

The programs seek and obtain a base-remote trunk for a call, queuing if necessary, before attempting to obtain a position or service circuit. Since positions are a more expensive resource, the base-remote trunk normally remains connected to the call throughout any position queuing to avoid adding the time required to queue for and connect the base-remote trunk to the operator work time. Thus the position queuing time is added to the base-remote trunk holding time. Because this has the potential to translate an operator overload into a base-remote trunk overload, a complex overload procedure is utilized which disconnects the base-remote trunk from the call if the position queue is very long and uses the high priority queue if necessary to reobtain one quickly.

3.3 The virtual scanner

The TSPS programs often determine the supervisory state of a trunk by directly scanning it. This cannot be done with RTA because the RTA scanner is autonomous. To solve this problem, a software image of the state of each trunk scan point is maintained in temporary memory. This is called the virtual scanner, since the required directed scans can be done by "scanning" (a memory read) this area of memory.

It is implemented by adding a virtual scanner (VS) bit to the two bits already assigned to each scan point[6] (one of these gives the last reported supervisory state while the other indicates that reporting supervisory changes has been temporarily suspended on this trunk).

The VS bit represents the software's best estimate of the current state of the scan point and is always updated whenever a trunk supervisory state change report is received from the RTA.

The only complication occurs when an RTA initialization is required. Such an initialization leaves talking state RTA trunks intact, but idles all others. Often when this happens, the information in the virtual scanner is probably unreliable so that the state of the trunk is no longer considered known.

When this occurs, charging is canceled on all RTA talking calls at the time and the RTA trunks are put into a special software "limbo" state. This state means that the software does not know whether the trunk is idle or talking. When an on-hook report is received on a limbo state trunk, the software assumes the trunk was in the talking state and initiates disconnect actions. Conversely, if an off-hook report is received, the trunk is assumed to have been idle and initial seizure actions commence.

3.4 Audits

The existence of the virtual scanner required a new type of audit program called the "virtual scanner update program" which is quite different from other system audits.

Normally, system audits are run on a routinely scheduled basis. Their purpose is to check the integrity of the software data structures. The virtual scanner update, however, is designed to check the consistency of the virtual scanner with the hardware state of the trunks. It does this by sending out a sequence of special orders which cause the trunk to send back several reports which together indicate its current state. Since doing this on all RTA trunks could overload the PCL, it is done very slowly over several minutes. When it is complete, an output message is printed, stating the number of errors found.

The audit is normally run during off-peak night-time hours and is used to verify the integrity of the virtual scanner and resolve any limbo state trunks which may exist at that time. It is also run after certain rare types of fault recognition activity which cause the scanner word buffer to be initialized with the possibility that one or more RTA trunk supervisory reports may be lost.

IV. RTA MAINTENANCE PLAN

Reliability requirements for an RTA subsystem have been set to meet the overall objectives of a TSPS No. 1 system: less than 2 hours downtime per 40 years service and less than one mishandled call per 10,000 calls. The maintenance plan for RTA is further influenced by the fact that an office is frequently unattended and may be located hundreds of miles from the base TSPS.

Sections 4.1 and 4.2 discuss PCL error detection as implemented in the hardware and software, respectively. Remote diagnostics are covered in Section 4.3. Section 4.4 deals with initialization of RTA circuits, while Section 4.5 describes trunk test capabilities for RTA.

4.1 PCL error detection (hardware)

The PCL data lines may be subjected to errors from a variety of causes. To aid in detecting multibit errors which result in single word errors, each word transmitted is encoded with a 5-bit cyclic error detecting code and a parity bit. In addition to single word errors, multiword errors on a data line may occur as a result of hits and fades. Each word error within a multiword error is treated separately by the group gate and remote data circuits. As a result, single and multiword errors are handled similarly.

As mentioned previously, the group gate and the remote data circuit send data to each other over a data line. An error detected by the receiving circuit must be reported back to the sending circuit so that the word can be transmitted again. This is accomplished by a verification system whereby a "scan-complete" is used to acknowledge a report received by the group gate, and a "check-back" acknowledges

a properly executed order received by the remote data circuit. If an acknowledgment is not received, the word is retried until it succeeds or the retry limit is reached.

A system requirement of RTA is that the voice circuits and data lines may use any transmission facility. This means that the PCL data lines must be able to operate over voice-grade circuits that have a relatively slow data transmission rate. Because of this and the distance (several hundred miles) between base and remote locations, about 45 ms are required to verify that an order or report has been sent without error. Since a PCL data rate of one order every 25 ms is required to deal with momentary call processing peaks and to allow use of as many existing TSPS No. 1 programs as possible, an overlap operation is employed. This means that, before verification of the first word is received, the second word is sent. If an error is detected, the receiving circuit discards both it and the succeeding word. The sending circuit retransmits both words if it doesn't receive verification of the first word. Therefore, the group gate and remote data circuit are designed to retain information about the last two words sent in case repeat transmissions in either direction are necessary. The overall maintenance strategy is strongly influenced by this aspect of the PCL.

The PCL hardware can detect errors other than transmission failures that would be seen by parity or cyclic code check failures. These faults are classed as "send" or "receive" direction faults. The former type will be discussed first.

The normal mode of PCL operation requires that each PCL half independently handle a processor order until it has been received at the remote location by the remote data circuit for each half. The two remote data circuits then compare the orders and if they match, one PCL half (a so-called "active" half) actually distributes the order to a trunk, concentrator, etc. If one half has a "good" order and the other half has received nothing or has a transmission failure, then the half receiving the good order becomes the active half and distributes the order. This mechanism, known as Full ORed Mode, is intended to reduce order retries due to data line transmission errors.

Each time the processor sends an order via the PCL to the remote location, the buffer which is to handle the order generates a checkback report as verification. This checkback is returned to the processor as an indication of correct operation; its absence indicates unsuccessful operation. The use of checkbacks allows for detection of faults beyond the signal distributor but only partly through the RTA buffers; the actual operation of relays and crosspoints are not verified by checkbacks. Those failures which occur beyond the checkback origination point are detected by error analysis programs.

Due to the overlap operation of the PCL described above, the remote data circuit is designed to also fail the order arriving after an order

which results in a checkback failure. This ensures that the second order is not executed before the first, necessary because the second order may depend upon the success of the first order and might lead to a mishandled call if the first order is unsuccessful.

The receive direction faults are those pertaining to the reception of signals originated by RTA circuits. The detection of these faults depends primarily on mismatches between PCL halves. Normally, the two PCL scanners each detect a change of state of an RTA circuit point and generate a report. These two reports are passed back to the group gates where they are compared. Subsequent to receiving the reports, the group gates transmit "scan-complete" words to the remote data circuits so that another pair of reports can be sent to the base.

In addition to matching, however, a cyclic code check is made, an all-seems-well indication from each scanner is checked, a parity check is made, and a sequence code check is made in each report received. The cyclic codes are inserted into the report by the remote data circuit; the all-seems-well, parity, and sequence indications are generated in the scanner. Cyclic codes and parity have been described earlier. A sequence code occupies two bits in each report and is used to eliminate redundant reports received at the group gate. This typically occurs when a "good" report is received but the scan-complete signal fails, causing an unnecessary retransmission of the report by the remote data circuit. If the sequence code check fails, the report is discarded by the group gate. The all-seems-well signal is generated by self-checking circuits within each scanner, as described in Section 2.1.2.

With both PCL halves in service (duplex), one group gate is designated as active. If a mismatch occurs and the active half has a "good" report (one which has correct cyclic code, parity, sequence, and all-seems-well), while the other half has received nothing or has a transmission failure, the good report will be passed on to the processor. However, if the active half has received nothing or has a transmission failure while the other half has a good report, a null scan-complete is generated, causing the report to be retransmitted. This mechanism, known as Half ORed Mode, takes into account the fact that the active group gate controls the matching. (If the program determines that only the active half has a transmission failure or has received nothing, it simply changes the designation of the active half to the one receiving the good report, leaving the PCL in duplex mode.)

In accordance with the circumstances described above, a group gate mismatch results in automatic retransmission of reports. However, if the group gates continuously mismtach several times, the PCL fault recognition program will take action to resolve the mismatch. The tolerance for a few mismatches allows the PCL to automatically recover from a momentary transmission impairment affecting one of the data lines.

Many faults can occur in the scanners which can only be detected by matching. The key characteristic of these faults is the presence of two apparently valid but different words that repeatedly mismatch. The cause of this is either the generation of an extra word or the lack of a generated word in one of the scanners. Under these circumstances, the list of reports in the two memories of the scanners become skewed and a mismatch occurs. The mismatching words represent two scanner inputs, one of which is faulty. This situation is handled by the "receive chain" fault recognition programs, which are discussed in the next section.

4.2 PCL fault recognition (software)

As described in the previous section, the PCL circuits have been designed to detect and correct for most errors which are the result of transmission impairments of short duration. For longer hits, fades, and PCL circuit failures, various other hardware indications, such as check-back failures and group gate alarm ferrods, allow PCL fault recognition programs to maintain working send and receive chains. The send chain refers to those circuits involved in the distribution of a PCL order: the group gate, remote data circuit, signal distributor, and the PCL buffers. The receive chain refers to those circuits involved in the reception of a PCL report: the group gate, remote data circuit, and position and trunk scanner.

If an order has been retried on a duplex PCL without the reception of two successive check-backs for 0.5 second, the send chain fault recognition program will be entered. When this occurs, call processing is suspended and a special test order is sent to an independent buffer. If that order is executed successfully, the original buffer accessed is deemed faulty. If the test order fails, the program splits the PCL so that each PCL half is completely independent with no matching taking place. The resulting simplex operation allows the original failing order to be tried first on one and then on the other PCL half. If the order fails over both halves, the program tries a second test order (to a different buffer) over each half. Based on the results of these tests, the program decides which half is faulty and places that PCL half out of service. Normal call processing is resumed on the good half, and diagnostics are run on the out-of-service half.

For receive chain faults, the fault recognition programs are entered whenever an error rate monitor (administered by software) corresponding to a group gate alarm ferrod exceeds its threshold. The absence of either all-seems-well, good parity, or proper sequence on a PCL half automatically indicates the faulty half. Mismatches between two valid reports involve more detailed processing.

Due to the overlap operation of the PCL, two reports on each half are sent to the base during retransmission. In the case of skewed

reports, chances are that two of the reports received match; that is, one of the reports received on one half may match one of the reports on the other half. The receive chain program checks for this case to eliminate the need for testing all four reports. Only those reports seen by only one half are further tested.

When the receive chain fault recognition program is entered, call processing is suspended. The PCL is again split into two simplex paths and a bypass in the scanners is activated to allow test results to be sent back to the base without disturbing the scanner word buffers. Each scanner is then checked, via test points, for the ability to generate the reports to be tested. These test results are analyzed by the program to determine which PCL half is in trouble. The faulty half is placed out of service, and diagnostics are run on it. Call processing resumes on the in-service half. The flow of reports proceeds from where it was interrupted, with the PCL in the simplex mode. While in the simplex mode, subsequent faults of this type go undetected.

In general, fault recognition program actions are transparent to call processing programs. In the case of a PCL mismatch, the original reports are retained in processor memory so that, once it is determined which PCL half is faulty, the reports already received on the good half can be passed to call processing programs. As already mentioned, the PCL scanners contain word buffers that can retain up to 31 reports which may have occurred while fault recognition tests were being applied. This is sufficient memory to handle a back-up of reports caused by fault recognition activity in all but the very worst cases.

In addition to the send chain and receive chain fault recognition programs, the data line fault recognition program responds to transmission errors reported to the SPC. Error rate monitors, driven by the group gate alarm ferrods, are maintained for each connected data line. When the allowable error rate is exceeded, the line that appears to be faulty is replaced by the spare line. After a delay of a few minutes, the apparently faulty line is reconnected. If its error rate is acceptable, no further action is necessary. If it fails again, however, the spare line is used until the faulty line is repaired or has been determined by subsequent automatic tests to have recovered. The spare data line is also used periodically so that its operational status can be monitored.

The state of the data lines must also be considered by the fault recognition programs when reconfiguration is being performed. If all circuits on a PCL half are working but both data lines that can connect to it are out of service, then that PCL half is not usable.

Although the normal mode of the PCL is the duplex mode, it is possible that a fault could occur while the PCL is simplex. The PCL could be simplex because diagnostics are being run on the out-of-service half or because a circuit on that half is inoperable. If the fault recognition program determines that the active half has a circuit in

trouble while the PCL is in simplex, the out-of-service half will be placed in service and the active half removed from service.

In cases where the out-of-service PCL half is already marked in trouble or has been manually made unavailable to fault recognition and a fault has also been detected on the in-service half, no known working path of PCL units exists between the base and remote locations. Rather than to place the entire PCL out-of-service, a special fault recognition program is invoked which attempts to find a working path, regardless of the software status of the units. This program systematically tests both the send chain and the receive chain of the PCL by using test orders and different combinations of data lines and base peripheral circuits that interface the processor to the group gates. In the event that these tests fail, the PCL will then be placed out of service. However, the program will automatically retest the PCL every few minutes and restore it to service as soon as one working (simplex) path is found. Because the duplex state offers maximum error detection, an out-of-service PCL half is also automatically tested for the absence of units in trouble (or a manual request to leave the half out of service) when the PCL is simplex. If no reason for the simplex configuration is found, the PCL will be restored to the duplex mode.

4.3 Remote diagnostics

The long data paths to RTA and PSS No. 2 installations and the desire to use voiceband facilities resulted in the use of 2400 b/s data rates between the SPC location and the remote locations. If lengthy strings of diagnostic tests were run from the SPC over an in-service PCL half to locate a trouble in the other (bad) PCL half, then call processing would be virtually halted for the duration of the diagnostic. To prevent such adverse call processing effects and still allow thorough diagnostics, the entire set of diagnostic test sequences are placed at the remote end and controlled by a diagnostic control circuit (DCC) which, in turn, is controlled by the SPC. The concept of placing the diagnostics *at* the remote location provides two other very desirable operations as well as minimizing adverse call processing effects. It allows the tests to run very fast since no transmission time is involved, and it allows the DCC to have direct access to various internal points of the circuits being diagnosed which would otherwise be impossible to obtain. With the exception of the scanner with its need for 25 ms to recognize a change of state (see Section 2.1.2), diagnostic orders can run at least an order of magnitude faster than regular PCL orders. This significantly reduces the time between a request for a diagnostic and TTY output of results.

The DCC is essentially a very low-level minicomputer with sequences of tests and expected results of those tests stored in Read-Only-Memory (ROM). Each ROM word is 24 bits long and contains a 3-bit operation code (op code), 20 data bits, and a parity bit. The data bits

usually apply a set of test conditions to a PCL circuit and cause that circuit to take some action. A subsequent ROM word contains an expected result which is matched against the actual result. If the result agrees, then another test word is selected from ROM and the process continues. The access that the DCC has to the various PCL circuits is shown in Fig. 9. Note that *indirectly* the DCC has access to *all* remote circuits; e.g., although it has no direct connection to a 100C console it can send orders to the RDC which can cause the signal distributor to send an order to a console. Similarly, while it has no input from a console it can check the scanner output to see if the scanner received a console output.

When comparisons of actual and expected results fail, a reply (noting the failure) is sent back to the SPC over the in-service PCL half. However, if massive failures are encountered, this reply capability is limited, so as to limit its effect on call processing. Also, massive failures will cause the SPC to terminate the diagnostic early since they quickly pinpoint a source of trouble. Whether or not comparison failures occur, every 64th ROM word causes a "progress report" type reply to the SPC to inform it that the diagnostic is continuing. The "failure report" sent back to the SPC on all mismatches provides a number between 0 and 63. The first 64 words of ROM store these 64 possible failure reports. The action taken by the DCC when a comparison failure (expected result versus actual result) occurs is to just set all but the least

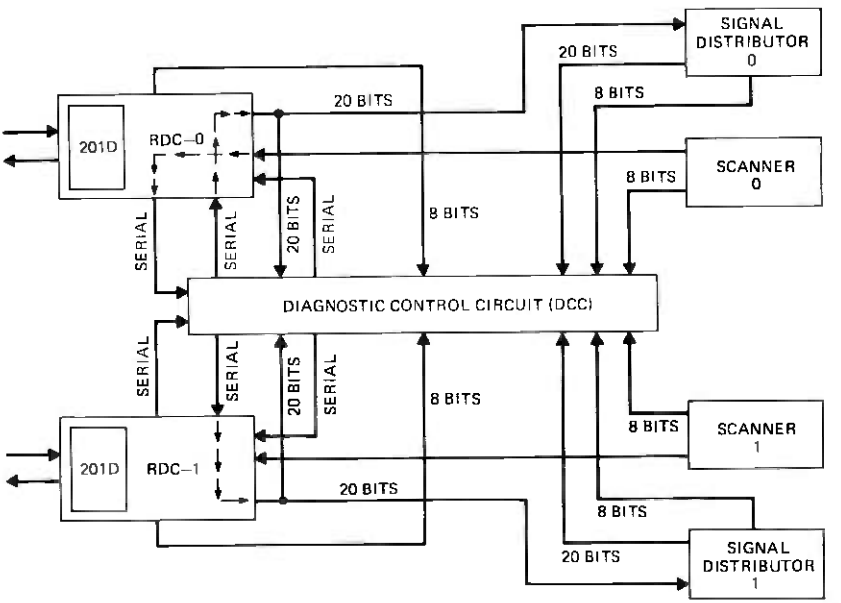


Fig. 9—DCC access to PCL circuits.

significant 6 bits of the address counter to zero. The next ROM word selected will then be one of the first 64 words and as this reply is sent back to the SPC and upper address counter bits are unclamped. With all address counter bits operational, the next word selected is the next word in the diagnostic test sequence. Thus no real program transfer and return function is needed in the DCC since this hardware approach to changing the address counter output does the transfer job and automatically includes the correct return condition.

The SPC controls the DCC by three steps. First, an order is sent over the inservice PCL half to place the out-of-service half RDC in the diagnostic mode, i.e., connect it to the DCC. Second, an order is sent to the DCC (again, over the in-service PCL half) to preload the desired starting address into the DCC address counter. Finally, a "start" command is sent to the DCC to select the first ROM word (at the starting address). This first word of any diagnostic sequence is always an "SPC reply" type word that returns the starting address to the SPC. The SPC checks this against the desired starting address and stops the DCC if the return is incorrect.

Since the SPC knows the starting address and is informed when every 64th subsequent address is reached, it can quickly compute the exact location of a comparison failure (its location in *that* diagnostic sequence) whenever it receives a failure report with its 0-63 tag number. Thus at the end of a diagnostic sequence, the SPC has a pattern of failures and uses these to generate a "trouble number." This number is then compared to a trouble dictionary to locate the failure.

The overall access which the DCC has to each remote RTA circuit can be understood by looking at the 8 possible operation codes together with the expansion of one such operation code. The 8 operation codes (op-codes) are listed below. This listing also gives the reader a glimpse at the DCC programming possibilities and indicates the ease with which each RTA circuit designer could write the appropriate diagnostic.

000 8-bit comparison or 25-ms delay
001 20-bit comparison from RDC
010 RDC test vector (DCC output)
011 20-bit scanner comparison preceded by 25 ms delay
100 Signal distributor test vector (DCC output)
101 20-bit comparison from signal distributor
110 RDC test vector (DCC output)
111 SPC reply word (DCC output)

The 000 op code is expanded within the 20-bit data field since only eight of the data bits are actually needed for an 8-bit comparison. The

order has the form:

23	22							15			11				6				2	0
P	Z_7	Z_6	Z_5	Z_4	Z_3	Z_2	Z_1	Z_0	1	1	1	C/D	X_3	X_2	X_1	1	M	C	B	A	0	0	0

where

Bits 0 to 2 are the op code (000).

CBA = 3 bits used to select the particular 8-lead input desired.

C	*B*	*A*	*Compare Input*
0	0	0	Input from signal distributor
0	0	1	All ones input
0	1	0	Input from RDC
0	1	1	Input from scanner
1	0	0	Spare input set Y
1	0	1	Spare input set X
1	1	0	Input from DCC
1	1	1	All zero input

M = "Mask" bit used to insert 1-bit mask. If M = 1, no mask is used and all 8 bits specified by CBA are compared against $Z_7 - Z_0$. If M = 0, a 1-bit mask is enabled as specified by the $X_3X_2X_1$ bits.

$X_3X_2X_1$ = Mask selection bits used to specify which of the 8 input leads are to be compared.

C/D = Compare/delay bit used to differentiate between use of this order for 8-bit comparisons or for 25-ms delay intervals.

$Z_7 - Z_0$ = 8 comparison bits.

The other seven op codes perform only the function listed and all 20 data bits are either applied by the DCC as a test vector or used by the DCC for a comparison.

The op code listing together with the 8-bit comparison possibilities shown above indicate the large number of internal points of all remote PCL circuits to which the DCC has access. The DCC concept allowed diagnostic points to be placed where they would do the most good. This makes every test/comparison pair of orders a powerful pair and leads to short diagnostic phases. These short phases together with the fast execution times allow most diagnostics to complete in a few seconds. As the diagnostic completes one phase (and before it starts a subsequent phase), a few SPC orders can be sent out to reconfigure the PCL half under test. For example, the RDC can be configured to loop back on itself. The DCC can then send orders to the scanner input side of the RDC—similar to reports normally returned by the RDC to the SPC—and, since the RDC is looped, these orders will come right back

through the receiving half of the RDC. They can then be examined by the DCC or allowed to continue to the Signal Distributor and cause some action there. Again, at the Signal Distributor, the test vector can be examined by the DCC or allowed to propagate even further and cause some other action further downstream. Judicious use of such DCC comparison tests allow very thorough diagnosis of remote circuits.

The DCC also has inputs from itself as well as all 1 and all 0 type inputs. It can use these inputs to run tests on itself and inform the SPC of possible DCC troubles. Massive failures in other circuits usually trigger (via an SPC program decision) this self-test of the DCC to ensure that the massive failures are real and not just the result of an inability of the DCC to make comparisons.

The DCC currently uses about eleven thousand 24-bit ROM words to test out all remote PCL circuits. The memory layout provides three circuit pack locations, each of which can contain $8K$ (K=1024) words so the memory can grow to 24K words if needed. The upper 16K addresses are also available in eight other circuit pack locations, each of which can accommodate a 2K-word PROM circuit pack. With this arrangement, the memory can be placed on ROMs on 1-1/2 circuit packs. Then, as various remote PCL circuit changes and additions take place which would tend to degrade the original diagnostic effectiveness, new versions of selected diagnostic phases can be placed in PROMs on, say, one new memory board plugged into a PROM circuit pack location. A 1-word change in the SPC program can then be used to specify a new starting address for the affected phase (or phases). The old version of the diagnostic is still there, but never gets used while the new (higher numbered) starting address selects *that one phase* from the PROM board. As PROMs are added over a period of years, and several sections of the original (ROM) program become unused, the ROM can then be redone to reflect several years of change.

4.4 RTA initialization

The RTA maintenance buffers contain a variety of distributor points used in conjunction with the operation of several remote circuits. When PCL reconfigurations occur, either by manual request or as a result of fault recognition action, some points in each buffer must be cleared. Some points are used to initialize the remote data circuit, signal distributor, and position and trunk scanner. Each buffer also contains points not connected with a particular PCL half which should not be cleared when reconfiguration takes place. To achieve the desired results, each maintenance buffer has been functionally divided such that selective initializations of these buffers can be performed. The selective initialization feature of the maintenance buffers also provides the capability of regenerating the PCL status display at the remote location upon manual request. In addition to the selective initialization

leads, each maintenance buffer is equipped with a lead which clears all points in the buffer. These leads are pulsed only when the entire PCL is being initialized, such as when a system initialization occurs, or when the PCL has developed faults on both halves.

Under certain circumstances (e.g., RTA or TSPS system initialization), it is desirable to place all the remote and base-remote trunks in an idle state. Since the number of remote trunks (496 max) and base-remote trunks (64 max) in an RTA may be large, individual trunk initialization would be very time-consuming. The trunk super-initialization feature of RTA allows all the remote trunks to be initialized with two orders and all the base-remote trunks to be initialized with one order. This allows for speedy initialization of the RTA trunks when necessary. It should be noted that those trunks already in the talking state will not be affected by the super-initialization orders.

4.5 Test frame capabilities

The test frame at the remote location displays status information and provides control and test facilities including a maintenance teletypewriter. As shown in Fig. 10, the out-of-service status lamps for the PCL equipment are physically arranged to depict the equipment ar-

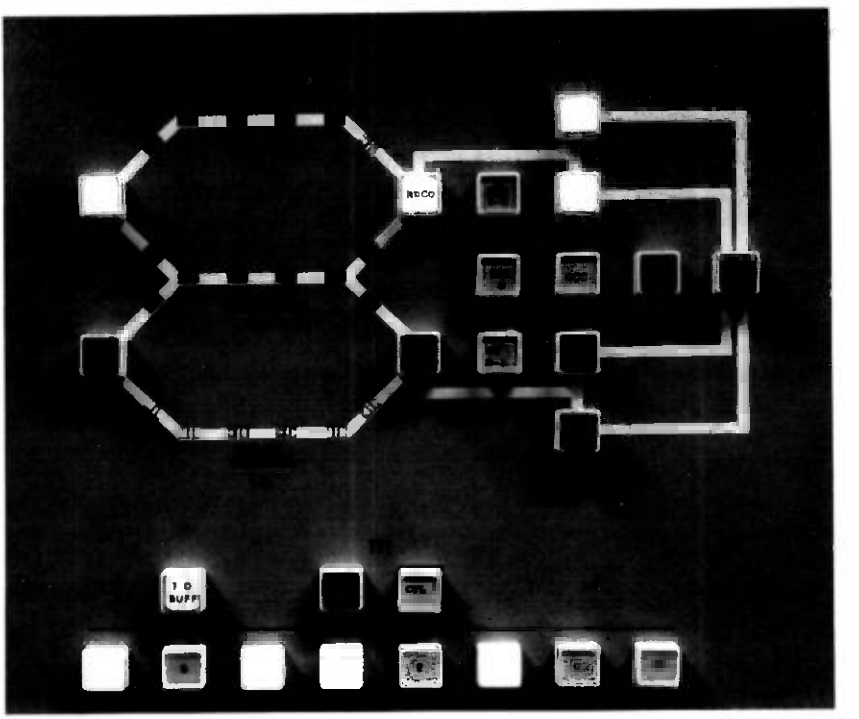


Fig. 10—PCL and voice status displays.

rangement in block diagram form. The three data lines are represented by three rows of light-emitting diodes which, when lit, indicate active working routes. The SW 0, SW REL, and SW 1 keys in the status display enable the remote ends of the data lines to be switched. This capability is provided to allow manual intervention in the rare case that both active data lines fail before the maintenance program at the base location has the opportunity to order the remote equipment to switch.

Several trunk test facilities are provided on the test frame as indicated in Fig. 11. A Dial Access Test Line (DATL) permits local office personnel to check transmission and noise levels on incoming trunks between the local office and the RTA without requiring assistance at the RTA. The Master Test Line (MTL) permits RTA personnel to establish voice communication with any trunk that terminates on the concentrator. The access line permits any trunk to be connected to any one of the test frame facilities which include a voltmeter, transmission level and noise measurement apparatus, a variable frequency milliwatt supply, and a quiet termination.

The test frame voltmeter and transmission measuring devices have a digital readout capability. The measurements can be remotely read

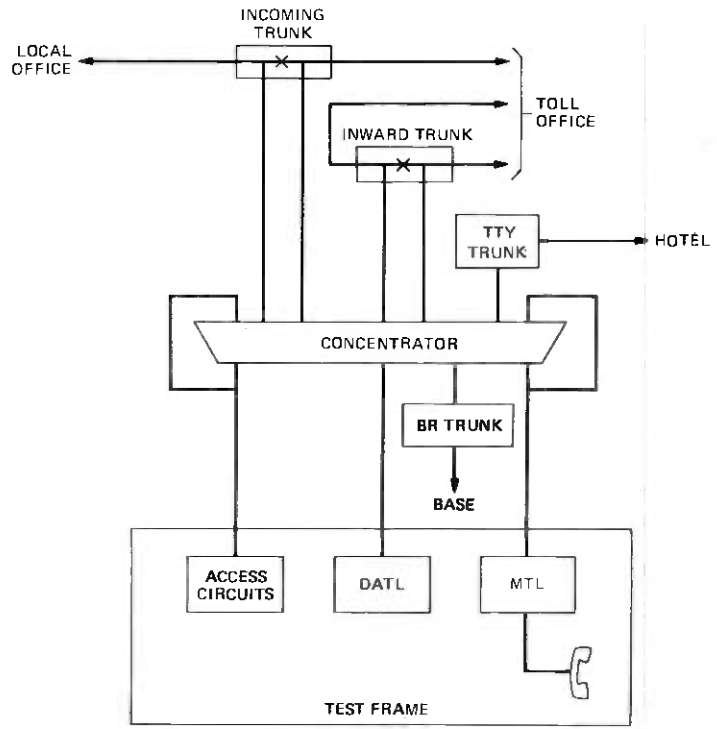


Fig. 11—Trunk test facilities.

by the base program and printed on the base location maintenance teletypewriter. This permits detection, verification, and sectionalization of many trunk problems without anyone present at the RTA.

V. ACKNOWLEDGMENTS

The design and implementation of the RTA hardware and software benefited from the efforts of many persons other than the authors. We would like to particularly acknowledge the contribution of P. S. Bogusz, T. J. Bradley, K. A. Heller, T. Huang, L. C. Kelly, R. V. Kraus, R. N. Markson, J. J. Rielinger, L. A. Rigazio, R. J. Roszkowski, S. M. Silverberg, R. A. Tengelson, R. J. Thornberry, R. A. Weber, and G. Y. Wyatt.

REFERENCES

1. E. Tammaru, "Correlated Error-Rate Results in Data Links with Duplicated Channels," National Telecommunications Conference Record (December 1975), p. 42.26.
2. E. O. Elliot, "Estimates of Error Rates for Codes on Burst-Noise Channels," B.S.T.J., *42*, No. 7 (September 1963), pp. 1977–97.
3. J. A. Hackett, et al, "TSPS/RTA Reliability and Maintainability in the Hardware Design," ICC 75, *3*, pp. 46-5—46-10.
4. A. F. Bulfer and B. D. Wagner, "Calculating Delay and Buffer Overflow Rate of Feedback Data Channels," 7th Hawaii International Conference on System Science, 1974.
5. F. H. Smith and J. M. Catteral, "Uses for the New Crossbar Switch," Bell Laboratories Record, *49*, No. 7 (August 1971).
6. A. W. Kettley, et al, "TSPS No. 1—Operational Programs," B.S.T.J., *49*, No. 10 (December 1970), pp. 2625–2685.
7. A. F. Bulfer, "Blocking and Routing in Two-Stage Concentrators," National Telecommunications Conference Record, December 1975, pp. 19-26 to 19-31.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Automated Coin Toll Service: Overall Description and Operational Characteristics

By M. BERGER, J. C. DALBY, Jr., E. M. PRELL, and V. L. RANSOM

(Manuscript received December 11, 1978)

Automated Coin Toll Service (ACTS) has recently been developed for use on Traffic Service Position System No. 1. ACTS automates the handling of most calls paid for at coin stations, gives time and charge quotations, and provides customer notifications. To accomplish this, a new microprocessor-controlled subsystem is added to TSPS. This subsystem generates announcements and monitors coin deposits. Since ACTS reduces the operator involvement on coin toll calls, it achieves significant savings for the operating companies. ACTS was developed together with several other associated features. This entire package was first put into service in November 1977, in Phoenix, Arizona. This paper gives a functional description of the new subsystem and details the customer interface with ACTS and the other features.

I. INTRODUCTION

To reduce the operating expenses associated with handling coin toll calls at a Traffic Service Position System (TSPS) No. 1,[1] Automated Coin Toll Service (ACTS) has been developed. ACTS permits automated processing of (*i*) the initial contact on most station calls paid for at coin stations (station-paid coin calls), (*ii*) notification at the end of the initial period on all coin-paid calls, (*iii*) overtime charge-due seizure on most coin calls, (*iv*) the customer-requested notification on a call that is not coin-paid, and (*v*) the quotation on time and charge calls. This is accomplished by giving machine-constructed announcements

to the customer and by providing machine recognition of coin deposit signals. Thus, ACTS reduces the operator work time on coin toll, noncoin notification, and time and charge calls, and thereby achieves significant savings for the Bell System.

The technology and concepts of Automated Coin Toll Service evolved over several years. First, the technical feasibility of ACTS was demonstrated in 1973 by building an exploratory model. System engineering studies were conducted in conjunction with American Telephone and Telegraph Company (AT&T) and the operating telephone companies that showed the reduction of operating expenses would offset introductory costs. Three parallel and interrelated development activities emerged. First, a new subsystem was added to TSPS. This subsystem, the Station Signaling and Announcement Subsystem (SSAS), uses a microprocessor called the Programmable Controller (PROCON) and semiconductor memory. The memory is loaded with speech segments that have been digitally encoded. The SSAS retrieves and concatenates these speech segments into sentences. By converting the bit stream to analog signals, SSAS can "speak" to customers. In addition, SSAS monitors coin deposits by the customer to determine when a coin deposit request is satisfied. Second, software and firmware were developed for the TSPS controller which is the Stored Program Control (SPC) unit[1] and the PROCON, respectively. This software allows the TSPS to interface with the subsystem and the subsystem to perform the desired tasks. Third, because of the complexities of the new machine-customer interface, a human factors study was conducted in Chicago, Illinois. In this study, a sampling of coin customers was exposed to the proposed ACTS service, which was simulated from existing positions. This human factors simulation established wording and timing parameters used in the ACTS design.

Part of the technical challenge was to introduce ACTS into live TSPS sites without interrupting the normal call-handling process. Not only was this challenge met, but additional features were introduced at the same time. These features include:

(i) Expanded direct distance dialing to Mexico.
(*ii*) Improved queuing strategy for calls transferred to TSPS for calling number identification for billing purposes (Centralized Automatic Message Accounting, or CAMA, traffic).
(*iii*) Special automated procedures to better control overload conditions.
(*iv*) The ability to more efficiently redistribute (rehome) TSPS trunks to different toll or local offices.
(*v*) Increased coin station test capabilities.
(*vi*) Other miscellaneous features.

These features and ACTS were packaged as a version of TSPS called Generic 8. Following initial service in Phoenix, Arizona on November

26, 1977, this generic was made generally available to the Bell System in mid-1978.

This paper gives an overview of ACTS and some of the other features associated with TSPS Generic 8. It highlights the hardware and software associated with each feature. Additional details of the three development activities that culminated in ACTS are specified in later papers in this special issue.[3-5]

II. COIN TOLL SERVICE PRIOR TO ACTS

TSPS provides a vast improvement in efficiency over cordboards in handling coin toll calls. Not only is the operator's interaction with a customer more efficient and accurate but, in addition, TSPS allows coin customers to dial their own calls, thereby providing faster service.

In a typical TSPS, between 10 and 15 percent of the calls handled by operators are toll calls made from coin stations and paid for with coins deposited at the station by the customer. On these calls in a non-ACTS TSPS, an operator is required (*i*) to quote to the customer the initial period length and charges, and to monitor the amount deposited, (*ii*) to notify the customer at the end of the initial period, and (*iii*) to quote and collect charges due for overtime at the end of the call or after 10 overtime intervals. To assist the operator in processing coin-paid calls, TSPS generally calculates the rates, automatically displays the charges at the operator position, and times the call.

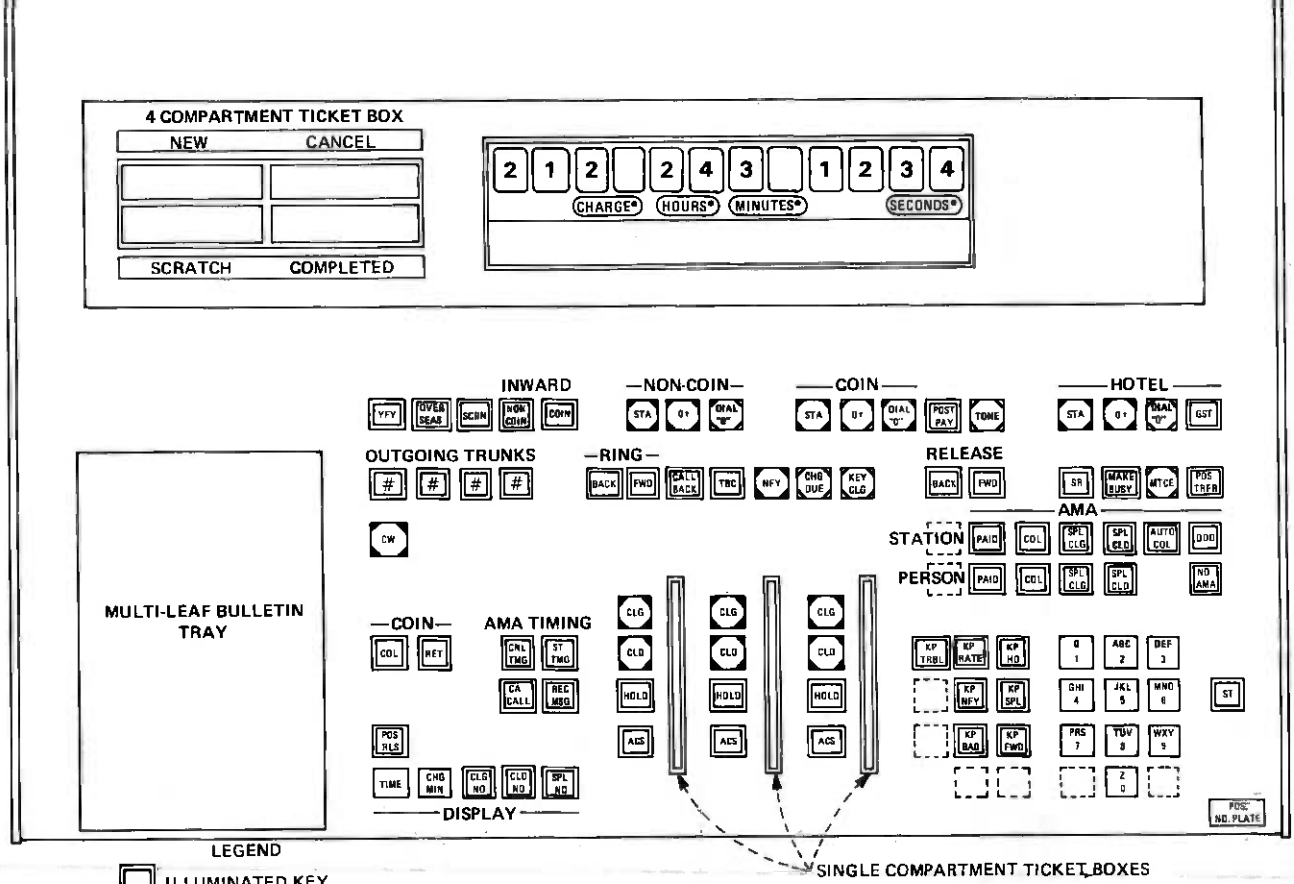
To better understand how the various functions described above are handled by an operator at a TSPS prior to the introduction of ACTS, the following details how TSPS processes a coin-paid call. Figure 1 shows the layout of the keys and lamps on the operator position.

2.1 Initial period contact

To place a station-to-station call at a coin phone served by a TSPS, a customer makes an initial deposit and listens for a dial tone. (In a dial-tone-first area, the initial deposit is not required.) As soon as the dial tone is obtained, the customer dials the digit one* plus the complete 7- or 10-digit called number. The local office determines that this call is a toll call and routes it over a trunk to its associated TSPS. TSPS receives the called and calling number from the local office and determines the charges on the call. The TSPS establishes the necessary connections through its network to bring the call to an operator position. While the operator is responding to the call, the called number is outpulsed to the toll office.

The call arrives at the position with appropriate keys and lamps lit to indicate that this is a coin call. The station initial period charges

* Some areas do not require the digit one to be dialed, depending on local number plan arrangements.


4 COMPARTMENT TICKET BOX
NEW
CANCEL
SCRATCH
COMPLETED
2 1 2 2 4 3 1 2 3 4
CHARGE*
HOURS*
MINUTES*
SECONDS*
INWARD
—NON-COIN—
—COIN—
—HOTEL—
OUTGOING TRUNKS
—RING—
RELEASE
AMA
STATION
PERSON
MULTI-LEAF BULLETIN TRAY
—COIN—
AMA TIMING
—DISPLAY—
SINGLE COMPARTMENT TICKET BOXES
LEGEND
ILLUMINATED KEY
NON-ILLUMINATED KEY
LAMP
TRUNKS TO RATE AND ROUTE, INTERCEPT, INFORMATION CARD SWITCHBOARD, ETC. AS REQUIRED
• THESE DESIGNATIONS ARE NOT VISIBLE UNLESS SIGNIFICANT

and minutes are displayed to the operator, as depicted in Fig. 2. The operator then requests the initial deposit from the customer and depresses the station-paid class-of-charge key to indicate to the system that a station-paid call is being handled. The customer then deposits the coins, and the operator monitors coin signals to determine if the proper deposit has been made. Meanwhile, the call is forwarded through the Direct Distance Dialing (DDD) network.

When the customer deposits the correct amount, the operator acknowledges receipt of the deposit. Then, upon hearing the audible ring of the calling station, the operator prepares the system for automatic timing of the call after called party answer. This is done by depressing

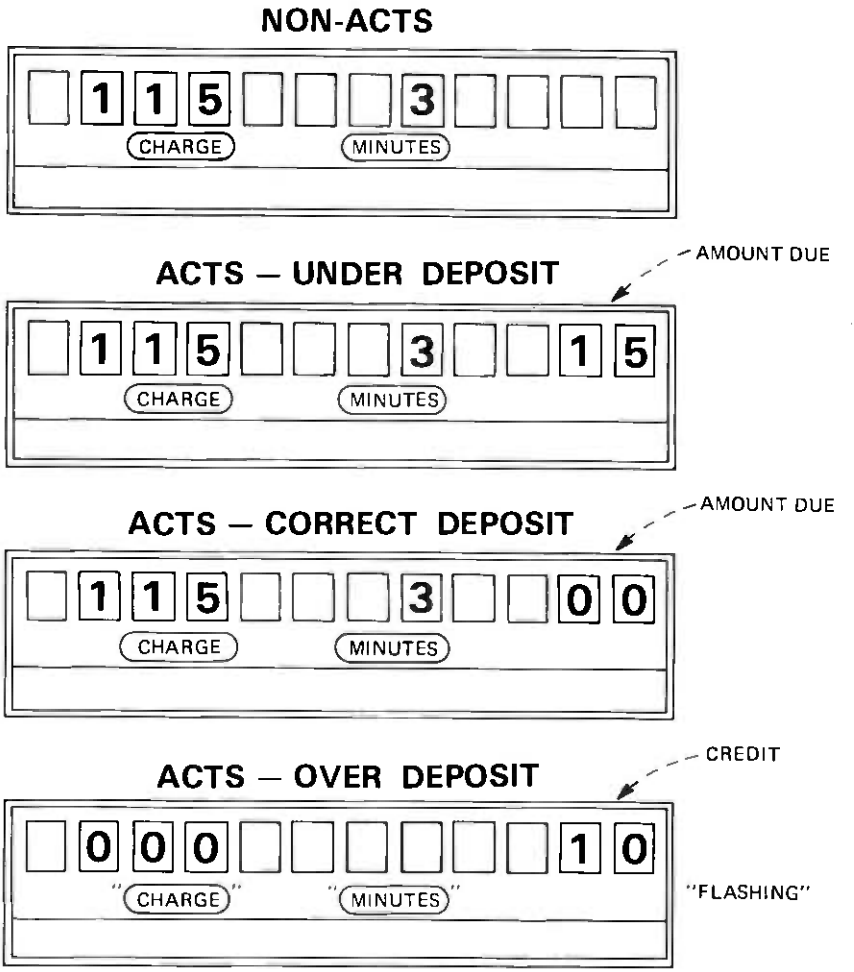


Fig. 2—TSPS No. 1 console numeric display.

the start timing key. The operator then releases the position, concluding the initial contact phase of the calls.

When the customer does not deposit enough money, the operator requests the additional amount. If the call has not yet progressed through the DDD network, the operator stops the forward action of the call by depressing the release forward key. When a full deposit is received, the operator reinitiates the network connection by depressing the start key, then he or she depresses the start timing key, and releases the position.

If a customer lacks proper change, he or she may deposit too much money. The operator acknowledges this overdeposit and instructs the customer to tell the operator that credit is due when overtime is collected.

With the operator released to handle another call, TSPS times the call for the initial period. Before the end of the initial period, TSPS sends a signal to the local office to cause the initial deposit to be collected.

2.2 Notification at the end of initial period

Generally, the call is again connected to an operator at the end of the initial period. This operator is not likely to be the same person who handled the initial contact. The call arrives at the position with the appropriate keys and lamps lit to indicate that this is a notify seizure. The operator notifies the customer that the initial period has ended and requests that the customer signal (by flashing the switchhook) when the call is completed. The operator releases the position and TSPS starts overtime timing.

2.3 Quotation and collection of charges due for overtime

If a call goes into overtime and reaches an elapsed time of 10 overtime intervals (usually 10 minutes), an operator is reconnected to the customer's line to request an intermediate deposit. This deposit is intended to limit coin overtime losses that result when a customer walks away at the completion of the call without paying the charges due. The operator counts the deposits and, when the request is satisfied, releases the position. Additional intermediate deposits are requested at appropriate intervals until the call terminates. Calls in overtime are not interrupted more frequently because of the operator work time penalty.

When the call is concluded, as indicated to TSPS by the calling customer signaling (flashing the switchhook) or either customer hanging up, the call is brought to a position for overtime collection. If the calling customer's phone is on hook, the operator rings the station by operating the ringback key. When the calling party is on the line, the operator requests the charges due and counts the coin deposits. When

the deposit request is satisfied, the operator thanks the customer, depresses a key which signals that coins are to be collected, and releases the position. If the calling customer does not answer the ringback or the operator is unable to obtain full payment for the call, a "walkaway" is recorded on a mark-sense ticket and the operator releases the position.

III. THE ACTS HARDWARE

A new hardware subsystem is added to the TSPS for ACTS, called the Station Signaling and Announcement Subsystem (SSAS). It is composed of a control unit containing a programmable controller called PROCON, an announcement store, and service circuits called Coin Detection and Announcement circuits (CDAs) (see Fig. 3). The control unit and announcement store are duplicated for reliability. One controller is active and the other is standby. The units are not run synchronously, but the temporary memory of each unit is updated on an ongoing basis.

The SSAS control unit contains the PROCON with a read-only program memory and a random access memory for temporary data storage. The SSAS control unit provides an interface with the Stored Program Control (SPC) unit,[1] its mate controller, the announcement memory,

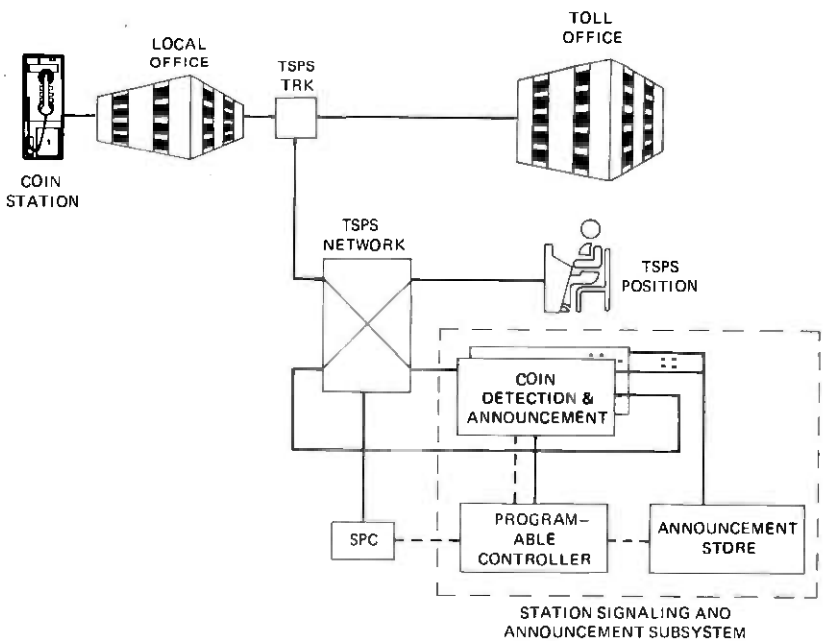


Fig. 3—ACTS block diagram.

and the CDA service circuits. The SPC communicates every 25 ms with the SSAS via software buffers as it does with other peripheral units.

The PROCON handles the internal control, data manipulation, and data transfer required for processing the call through the successive steps from customer contact through deposit acknowledgment. Within each control unit, the PROCON provides a self-checking mechanism for immediate detection of faults.

Prerecorded words and phases are digitally encoded and stored in the announcement store. The announcement store can be expanded in increments of eighty 512-ms speech segments. It is organized to be tolerant to single-bit faults and is diagnosed via the SPC. PROCON forms announcements by directing speech segments from the announcement store to the CDA in the appropriate sequence.

The CDA circuits have appearances on the TSPS switching network and perform the functions of making announcements to customers and of detecting coin deposits. The CDA circuits also are used to monitor for coin deposits when an operator is handling the call. CDAs are traffic-engineered with a typical TSPS requiring from 30 to 40 of these service circuits for ACTS. These service circuits convert the digital bit stream to an analog announcement.

IV. COIN SERVICE WITH ACTS

ACTS uses SSAS to automate the handling of (*i*) initial period coin collections, (*ii*) notifications of the end of the initial period, (*iii*) charge due coin collections, (*iv*) time and charge quotations, and (*v*) notifications on nonpaid calls. This section details the procedures used to automate these call seizures.

4.1 Initial period coin collection

With ACTS, a coin customer initiates a station-to-station call as in the past by dialing one* plus the called number. Since no local office trunk modifications are needed, the local office goes through normal routing of the call to TSPS. After TSPS receives the called and the calling numbers, it determines the rate and computes the initial period charges. However, instead of connecting an operator, TSPS makes network connections that bring the call to a CDA circuit and passes the initial period and charge information to the SSAS. The call is not outpulsed forward until the SSAS completes the initial contact with the customer. In the following sections, the procedures and announcements for the initial collection are described.

* Some areas do not require the one to be dialed.

4.1.1 Initial deposit request

The SSAS constructs the following announcement:

$$\text{“}\left\{\begin{array}{l}\text{X dollar(s)}\\ \text{X dollar(s) and Y cents}\\ \text{Y cents}\end{array}\right\}\text{please.” (2-second pause)}$$

$$\text{“Please deposit }\left\{\begin{array}{l}\text{X dollar(s)}\\ \text{X dollar(s) and Y cents}\\ \text{Y cents}\end{array}\right\}\text{ for the first Z minute(s).”}$$

4.1.2 Initial coin prompt

In studies of customers making coin-paid toll calls, it was observed that many customers ask the operator to repeat the amount before they begin depositing. This even occurred on calls in which the amount was stated twice in the initial request. In view of this, if no deposit is detected within an allotted time after the initial announcement, the SSAS prompts the customer with, for example:

“Please deposit 1 dollar and 50 cents.”

Although a 5.5-second initial coin timing interval is used for the beginning of the call, all timing intervals are designed so that they can be changed in case significant differences in customer behavior are encountered as a result of growing customer familiarity with the service or as a result of demographic differences.

4.1.3 Intercoin prompt

Some customers begin depositing, but stop before the request is satisfied. For example, the customer may lose count of the coins deposited. To accommodate these customers, an intercoin prompt is given. If the time between coin deposits exceeds an allotted time and the deposit is not satisfied, the SSAS prompts the customer by announcing the additional deposit needed in a manner similar to the initial coin prompt.

4.1.4 Acknowledgment of a correct deposit

Data show that over 80 percent of the customers satisfy the deposit request within the allotted time intervals. When the customer deposits the correct amount, the SSAS takes three actions. First, it informs the TSPS to outpulse the call. Second, while the call is being outpulsed, the SSAS acknowledges the deposit to the customer by

“Thank you.”

Third, when the announcement is completed, the SSAS reports the

amount detected to TSPS. Following this report, the TSPS disconnects the CDA circuit and proceeds to time the call.

4.1.5 Acknowledgment of an overdeposit

Some customers lack the proper change and deposit too much money. With ACTS, an overdeposit is acknowledged and credit toward overtime charges is automatically obtained. This eliminates the need for a customer to indicate a credit is due. When the overdeposit is recognized, the SSAS informs TSPS to outpulse the call and to record the amount of the overdeposit. The SSAS acknowledges the overdeposit with the following phase:

"Thank you. You have W cents
credit toward overtime."

If the customer does not use the overdeposit credit and wants a refund, the customer must reach an operator and request a refund.

The SSAS does not time for overdeposits greater than breakage, but as long as the CDA circuit is attached, subsequent deposits are credited towards overtime. Breakage occurs when the denomination of the last coin brings the amount deposited over the amount due (i.e., \$0.05, \$0.10, \$0.15, or \$0.20). As soon as the amount deposited equals or exceeds the amount due, the SSAS informs TSPS to outpulse. To time for additional overdeposits (i.e., prepayment of overtime), the SSAS would have to delay all calls. If the customer overdeposits inadvertently and wishes to redeposit the correct amount, the customer must hang up to have the coins returned. This can be done any time before the called party answers.

4.1.6 Deposits during announcements

The CDA circuits monitor coin deposits during deposit requests. If a coin is detected during a request, the request is truncated immediately. If this coin does not bring the amount deposited up to the amount due, intercoin timing begins. If the amount due is met, the appropriate acknowledgment is transmitted. This allows customers who know the charges for a call to deposit the required amount without listening to the entire deposit request.

4.1.7 Operator assistance

Coin customers may not properly respond to fully-automated service and may need operator assistance. For example, customers who lack the correct change may request that the charges be billed to a credit card, to a third party, or to the called party. To deal with such occurrences, an operator is needed. Thus, if the customer fails to deposit within the allotted time following a prompt, the call times out and receives operator assistance. In addition, if a customer flashes the

switchhook during the coin deposit announcements, the call is also brought to an operator.

When the operator is attached, a new pattern of keys and lamps as well as a new numeric display (see Fig. 2) are lit on the operator console. This informs the operator of an ACTS call and provides the operator with the additional information concerning the amount still to be deposited. The CDA circuit remains connected between the customer and the operator. The circuit continues to monitor coin deposits but makes no announcements. The operator assists the customer in making a full deposit.

When the CDA circuit detects the full deposit, the SSAS reports "deposit satisfied" to the TSPS. The TSPS updates the operator's numeric display and outpulses the call. If the coin deposit tones are not being accepted by the CDA circuit and the operator suspects an equipment malfunction, the operator can override the SSAS and allow the call to progress.

4.2 Notification at the end of the initial period

The SSAS can provide the notification of the end of the initial period on all coin-paid calls whether the initial contact was handled by the SSAS or by an operator. When the initial period timing ends, the call is connected to a CDA circuit instead of an operator. TSPS informs the SSAS of the length of the initial period and the CDA circuit being used. The SSAS then announces to the customer:

"Z minute(s) has ended. Please
signal when through"

where Z ranges between 1 and 6.

When the announcement is complete, the SSAS informs the TSPS that the CDA circuit is idle. The TSPS disconnects the CDA circuit and starts overtime timing.

4.3 Charge due seizures

The ACTS procedures for fully automating overtime charge due seizures on coin-paid calls are presented in this section. The same sequence of deposit requests, coin deposit timing, prompting, and acknowledging described for the initial period deposit request are used for the overtime deposit request. However, the announcement wording is appropriately changed to indicate that money is due for the preceding connection time.

4.3.1 Charge due deposit request

In the same way as before ACTS, if a call lasts a certain number of overtime intervals, an intermediate deposit is requested from the customer. With Generic 8, the number of overtime intervals can be

varied in accordance with the amount of overtime charges (previously it was fixed at 10). Additional intermediate deposits are requested until the call terminates. The SSAS automates the collection at both intermediate and end-of-call overtime charge due seizures on ACTS-handled calls. The collection sequences for end-of-call deposits and intermediate deposits are essentially the same.* Furthermore, these collection sequences are very similar to those used for the initial contact on station (1+) calls.

Without interrupting the conversation path, TSPS connects an idle CDA circuit to the call. TSPS informs the SSAS of the amount due, the number of minutes talked, the CDA circuit being used, and the fact that it is a charge-due seizure. The SSAS generates an announcement similar to that used for initial coin deposit requests.

If the customer overdeposited during the previous collection sequence, the credit is automatically given. The overdeposit is subtracted from the calculated charges due. If no money is due, no actions are taken. If money is due, the TSPS connects an idle CDA circuit to the call. TSPS informs the SSAS of the amount due, the amount of credit, the number of chargeable minutes, the CDA circuit being used, and that it is a charge-due seizure with credit.

To assure the customer that credit is being given, an announcement is used, such as

> "2 dollars and 10 cents please." (2-second pause) "You have 20 cents credit. Please deposit 2 dollars and 10 cents more for the past 10 minutes."

After the charge-due deposit request is made, the deposit timing and, if needed, the prompting announcements described for the initial contact are used. If the deposit request is satisfied, the "Thank You" acknowledgment is given. When the SSAS informs TSPS that the acknowledgment is completed, the CDA circuit is disconnected from the call.

As with initial contact, operator assistance is given if no coin is deposited within a time interval, presently set at 5.5 seconds, or if the customer flashes during the deposit sequence. An operator is also connected if the customer goes on-hook during the announcement sequence.

4.3.2 Walkaways

If the calling customer goes on-hook at the end of a coin-paid call and charges are due, TSPS automatically generates a ringback signal to

* With intermediate deposits, the called party is off-hook. A special coin detection arrangement is used to monitor the calling and called stations.

cause the calling phone to ring. If the calling party answers, an ACTS overtime charge due announcement is made. The start of the announcement is delayed 2 seconds from the time the customer goes off-hook. This allows time for the customer to get the handset to the ear.

However, if the calling party does not answer, TSPS assumes a walkaway. Since the customer did not respond to an automatic ringback, the customer probably will not respond to an operator ringback. Hence, an operator is not connected. Instead, a traffic counter is pegged and a walkaway record is made on the Automatic Message Accounting (AMA) tape. These walkaway records can be processed later to determine patterns so that action can be taken to reduce fraudulent use of coin phones.

4.4 Time and charge quotations

When a customer requests a time and charge (T&C) quotation on a call, the operator instructs the customer to flash and remain off-hook at the end of the call. If the customer follows the operator's instructions and the called party goes on-hook for two or more seconds, the forward connection is released and the call is connected to an idle CDA circuit. TSPS informs the SSAS of the charges, the length of conversation, the CDA being used, and that it is a time and charge quotation. The SSAS generates an announcement, such as

"The charges are 3 dollars and 94 cents plus tax for 12 minutes."

If the calling customer remains off-hook, the quote is repeated 3 seconds later. The SSAS informs TSPS when the second quote is completed, and the call is terminated. If the calling customer goes on-hook during this sequence, TSPS informs the SSAS to suspend the announcement sequence, and the CDA circuit is idled.

The fully automated quotation is given only to calling customers who remain off-hook at the end of the call. If the calling customer goes on-hook at the end of the call, TSPS connects an operator to contact the calling customer and give the T&C quotation. An operator is required because a system ringback may not be answered by the calling customer. For example, a ringback could be answered by a PBX attendant who knows nothing about the call.

When the SSAS gives the T&C quotation, the AMA record specifies that an SSAS quotation was given and the charges the customer was quoted.

4.5 Customer requested notification

On a call that is not coin-paid, a customer can ask to be notified at the end of 1 to 10 minutes. During the initial contact, the operator enters the customer's request into TSPS memory. If such an entry is

made, TSPS connects the idle CDA circuit to the call at the appropriate time. TSPS informs the SSAS of the CDA circuit being used, the number of elapsed minutes, and that it is a notification seizure. The SSAS generates the prescribed tone and announces, for example,

"5 minutes has ended."

When the SSAS informs the TSPS that the announcement is finished, TSPS disconnects the CDA circuit and continues to time the call.

V. ADDITIONAL TSPS FEATURES ON GENERIC 8

In parallel with the ACTS development, other service and maintenance features were developed. Those features are released with ACTS in a software package known as a generic. Since the generic containing ACTS is the eighth for TSPS No. 1, it is called Generic 8. This section briefly describes some of the additional features in Generic 8.

5.1 Expanded dialing to Mexico

This feature expands customer dialing capability on calls to Mexico. The Generic 8 dialing plan for Mexico is eight digits, except that the Northwest border area has a 903 area code followed by seven digits. Presently, calls to the 903 and 905 (Mexico City) area codes are customer dialable. (In the case of Mexico City, the digit "5" is both the third digit of the area code and the first of the eight digits in the dialing plan for Mexico.) Other calls to Mexico are only dialable by an operator, and many areas can only be reached by an inward operator.

The expanded Generic 8 capability allows customers to dial directly, using the international format, calls that currently are only dialable by an operator. Specifically, a customer dials 011 or 01 plus a country code of 52 plus the 8-digit Mexican number. TSPS software recognized the 52 country code and bypasses the 2-stage outpulsing normally done on international calls. TSPS then outpulses 180 plus the 8-digit number for routine handling (6-digit translation) in the domestic toll network. A toll office with trunks to Mexico eventually is reached and outpulses the 8-digit number. In Generic 8, TSPS also continues to handle calls dialed using the 903 and 905 formats.

Besides expanding direct dialing, this feature also simplifies inward calls to Mexico. The TSPS operator keys 52x-121, where x = 5 for Mexico City, 8 for Monterrey, 1 for Chihuahua, or 6 for Hermosillo. TSPS uses the "x" digit to index into a translation table. This table contains the codes to be outpulsed to reach the appropriate Terminating Toll Center.

This feature improves service by allowing the customer to dial calls directly. It also is expected to yield significant economic savings by automating a class of traffic that is experiencing substantial annual growth.

5.2 Transfer CAMA queuing improvement

In some cases, TSPS operators provide Operator Number Identification (ONI) to a CAMA office which will record the billing information. For instance, many TSPSs handle transfer CAMA traffic at night, when CAMA boards are closed and traffic is normally light. If traffic fluctuations produce a relatively heavy load, incoming calls can experience delays before reaching an operator position. Generic 8 has an improved strategy, called delay ratio control queuing, for handling traffic in TSPS offices which perform a transfer CAMA function.

This new queuing strategy tends to maintain a preselected balance between the delays experienced by transfer CAMA and other TSPS customers. A separate queue is established for transfer CAMA calls waiting for a position. Other TSPS customers queue as they do now. The serving rate for transfer CAMA versus other TSPS customers is dynamically modulated by the ratio of the respective queue lengths. This approach has been verified by analytical simulation and field studies to provide better service to CAMA customers while maintaining good service to other TSPS customers.

5.3 Dynamic queuing strategy

The dynamic queuing feature provides measurements and controls of the delays experienced by calls which require an operator position. The delay measurements are used to determine when to light a "calls waiting" lamp indicating to the operators that a moderate number of customers are awaiting assistance. Operators can then expedite call handling. If the delays increase further, the dynamic queuing measurements trigger the application of delay announcements.* These announcements turn away some of the new calls that would otherwise enter the queues.

Previously, the number of TSPS positions staffed was used to index into tables giving critical queue lengths for activating delay announcements and the call waiting lamp. These tables were originally constructed on the assumption of a fixed (60-second) Average Work Time (AWT) per call. Tables based on fixed AWT are insensitive to variations in service rate.

The dynamic queuing feature measures the actual delay encountered by incoming calls. Call abandonments are also directly taken into account, since they affect the measured delay. This accurate delay estimate improves service to TSPS customers. The dynamic queuing feature interacts constructively with the transfer CAMA queuing feature, since separate data can be obtained on the delays seen in the transfer CAMA queue. An integrated approach to lighting the call

* Position disconnect is supplied on transfer CAMA trunks so that reorder can be given at the toll office. The delay announcement is given to other TSPS traffic.

waiting lamp and turning on the delay announcement (or supplying position disconnect) is based on accurate assessments of all the queuing delays in the system.

5.4 Redistributing (rehome) TSPS trunks

The ability to redistribute (rehome) TSPS trunks to a different toll or local offices allows an operating company to react to changing traffic trends. This ability allows TSPS capacity to be efficiently used in adapting to the evolving traffic patterns. Rehoming can be used to relieve congestion at a toll office or when an existing office is being replaced by a more modern electronic office. As an example, perhaps a No. 4A Crossbar office is being replaced by a No. 4 ESS.

Office data parameters for the trunks can be changed en masse. The software limitation of a single office parameter per trunk group is deleted for rehome. The software accommodates a difference (for instance, wink versus delay dial signaling) between the offices.

5.5 Improved coin station tests capabilities

TSPS Generic 8 provides a new coin station test capability. With this capability, a craftsperson can make end-to-end, coin signaling tests between coin stations and TSPS. This new procedure directs the craftsperson at the coin station through a series of tests. The SSAS is utilized to provide appropriate feedback to the craftsperson concerning the denomination of the detected coins. Marginal/stress testing is an integral part of the test. With the test information, the craftsperson is able to determine whether the station is functioning correctly and, if not, what functional area of the station is defective. In addition, certain types of coin-deposit signaling errors detected by the SSAS are recorded with other call billing details. This failure information can be summarized by a telephone company using the new Coin Operational and Information Network (COIN) package.* This information can be used to direct the craftsperson to potentially defective coin stations. Together, the per-call failure information and the quick, simple, and flexible coin station procedures provide the telephone companies with improved detection and resolution of coin system problems.

VI. ACTS DEPLOYMENT AND ECONOMICS

Today, over 75 percent of the coin stations and 80 percent of the average business-day, coin-paid calls in the Bell System are handled on TSPS, and this coverage is increasing each year. The incorporation of ACTS into TSPS eliminates or reduces operator handling of most of

* COIN is an off-line computer package which performs collection, scheduling, and revenue analysis functions for public telephones.

these coin-paid calls, thereby achieving significant operating expense savings for the Bell System.

To achieve these savings, the previously described hardware and software are added to TSPS. In preparation for the introduction of ACTS in the Bell System, new coin stations and coin chassis produced by Western Electric since 1975 have been (and will continue to be) equipped for ACTS coin detection.

On November 26, 1977, the first new TSPS equipped with the ACTS feature was introduced into service in Phoenix, Arizona. In March, 1978, an existing TSPS was first retrofitted with the Generic 8 features in Oakbrook, Illinois.

ACTS was made available on a standard basis to the Bell System operating companies in the middle of 1978.

VII. ACKNOWLEDGMENTS

Several areas at Bell Laboratories made contributions to the overall planning of Automated Coin Toll Service and the other associated features. Significant individual contributions were made by J. J. Coschigano, J. M. Gobat, J. C. Hemmer, F. G. Oram, K. A. Raschke, B. W. Rogers, C. M. Rubald, A. M. Santacroce, S. L. Skarzynski, Jr., and several others.

REFERENCES

1. "TSPS No. 1," B.S.T.J., *49*, No. 10 (December 1970), pp. 2417–2731.
2. Ibid.
3. G. T. Clark, K. E. Streisand, and D. H. Larson, "TSPS No. 1: Station Signaling and Announcement Subsystem: Hardware for Automated Coin Toll Service," B.S.T.J., this issue, pp. 1225–1249.
4. R. Ahmari, J. C. Hsu, R. L. Potter, and S. C. Reed, "TSPS No. 1: Automated Coin Toll Service," B.S.T.J., this issue, pp. 1251–1290.
5. E. A. Youngs, W. J. Bushnell, and A. Baron-Wing, "TSPS No. 1: Automated Coin Toll Service: Human Factors Studies," B.S.T.J., this issue, pp. 1291–1305.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July-August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Station Signaling and Announcement Subsystem: Hardware for Automated Coin Toll Service

By G. T. CLARK, K. STREISAND, and D. H. LARSON

(Manuscript received December 28, 1978)

A new subsystem, Station Signaling and Announcement Subsystem (SSAS), *was added to* TSPS *to provide automated coin toll service. Presented here are descriptions of how this new subsystem generates announcements from digitally stored speech samples, how it responds to coin deposit signals from coin stations, how the announcement circuits and coin tone detection circuits are automatically tested, and how the subsystem is physically packaged.*

I. INTRODUCTION

The new functions needed in TSPS for Automated Coin Toll Service are provided by the newly designed Station Signaling and Announcement Subsystem (SSAS). SSAS delivers voice announcements to customers at coin stations and responds to the coin deposit signals generated at the coin stations.

SSAS can be described from two perspectives, that of a coin toll customer and that of the TSPS processor. From the coin toll customer's viewpoint, it should sound and react the same as or better than the human operators the customer is accustomed to. From the standpoint of the TSPS processor, it operates as an "intelligent peripheral," using the existing instruction and data buses. In response to TSPS instructions, it constructs announcements to request and acknowledge coin deposits. It keeps track of each customer's coin deposits and then reports the amount of the completed deposit to the processor, which then sets up the call. If the customer does not deposit enough coins, SSAS delivers requests for the balance, or finally sends a time-out

report to the processor so an operator can be brought in to assist the customer. To provide reliable service, the SSAS control circuits are duplicated, with one control circuit active and the other standby. They contain self-checking features so faults can be detected promptly and reported to the TSPS processor, enabling a smooth switchover from the active to the standby SSAS controller.

The interface of SSAS with customers is through a number (up to 239 per system) of Coin Detection and Announcement circuits (CDAS). Each CDA provides service to one customer at a time. Each CDA contains a coin tone receiver that recognizes the nickel, dime, and quarter deposit signals from dual-frequency, single-slot, coin stations. Each CDA also contains a digital-to-analog decoder that converts bits at 31,250 b/s into natural-sounding voice announcements. To provide for connecting an operator if the customer has difficulty and to reduce the possibility that operator or automated announcement speech will interfere with correct coin recognition, each CDA also includes a 4-wire, voice-frequency network with an extra port for the operator. Each network contains 4-wire terminating sets and amplifiers to isolate the coin-tone receiver and digital-to-analog decoder from one another. In addition to the customer-serving CDAS, each SSAS contains one unique CDA circuit that functions as a built-in test set.

Figure 1 shows the essential parts of a connection from a customer to an SSAS CDA.

Before the advent of SSAS, a coin call needing TSPS operator service would be connected through a TSPS trunk and the TSPS network to an operator position or to a service circuit for ringing or busy tone. With SSAS, a call identified as eligible for automated coin detection handling is immediately connected to a CDA. If it later develops that an operator is needed, one can be connected to speak to the customer or to listen to the automated announcements and coin signals.

Figure 2 shows the principal parts of SSAS along with the input and output data bus connections to the TSPS processor.

The coin-tone receivers in the CDAS deliver detected coin data to digital registers, also in the CDAS. These registers are scanned periodically, and the data are ultimately relayed back to the TSPS processor. The digital-to-analog decoders that deliver the announcements were adapted from an earlier Bell Laboratories design, Subscriber Loop Carrier 40 (*SLC*™-40). They are preceded by serial buffers that are loaded sequentially, with bursts of 40 bits transmitted at a 1-mHz rate. The bits are decoded at a steady rate of 31,250 b/s.

As mentioned earlier, as many as 239 CDAS can be associated with one SSAS. Since the holding time needed to request and collect coin deposits for a typical call is relatively short, 239 is expected to be enough CDAS to handle the coin toll traffic in large metropolitan TSPS offices. The CDAS are operated by a controller frame and an associated

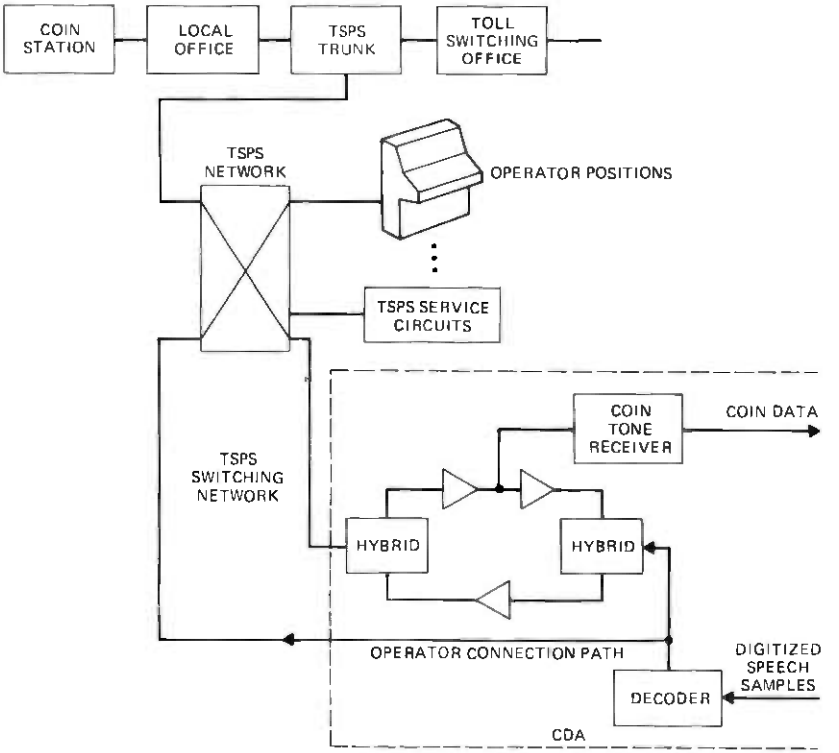


Fig. 1—Connection of SSAS CDA into TSPS.

semiconductor announcement store. For backup in case of failure, there is an identical controller frame/announcement store pair. The announcement store, except for minor modifications to make it accessible either from the SSAS controller or the TSPS processor, is the same design as the TSPS processor main store frames. It may contain up to six memory modules, each containing 32K 47-bit words. Forty bits of each word are data; seven are used for error detection and correction.

The controller frame contains a programmable controller (PROCON) and wired logic which, in response to instructions from the TSPS processor, retrieves samples of digitized speech from the announcement store and distributes them in a multiplexed scheme, with a fixed sequence, 40 bits at a time, to the CDAs. The announcement distribution sequence has 256 time slots, of which 16 are used for test instructions and 240 are used to deliver bits to the 239 CDAs and the one test channel. The distribution sequence is repeated every 1.28 ms, so that each CDA, using bits out of its serial buffer at a rate of 31,250 b/s, for announcement generation, is supplied with precisely enough data to produce uninterrupted announcements consisting of 512-ms segments joined together.

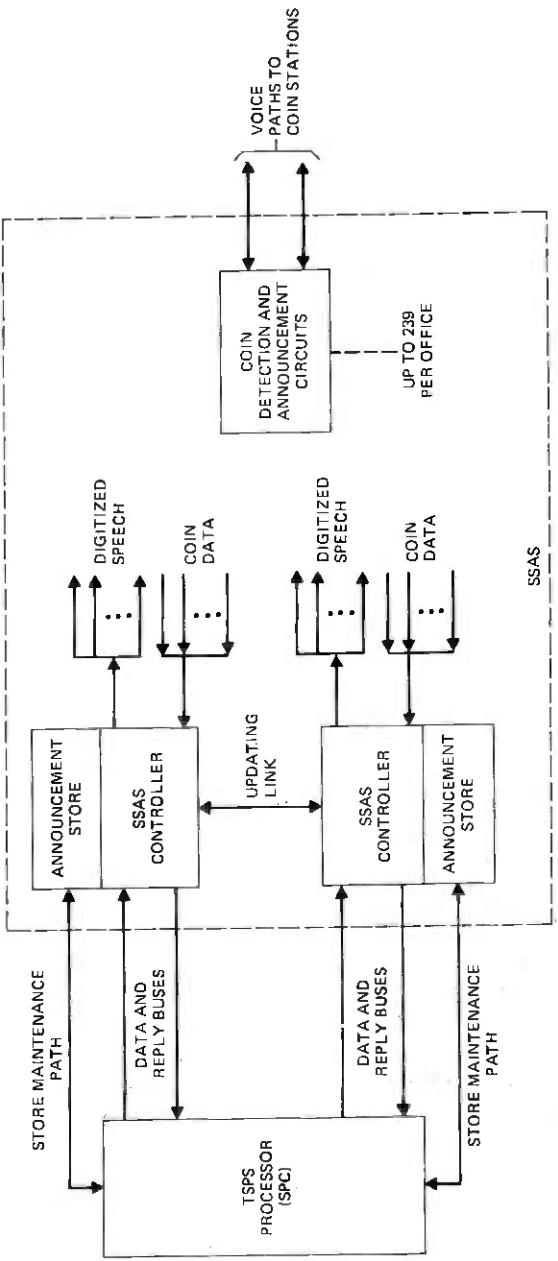


Fig. 2—SSAS block diagram.

If an office were equipped with the full complement of 239 CDAs, it could deliver announcements to 239 customers simultaneously. When there are not 239 CDAs installed, or when not all installed CDAs are in use, a bit pattern representing silence is placed in each inactive time slot.

At the same time that the PROCON in the SSAS controller is distributing 40-bit digitized speech segments to the CDAs, it is sequentially scanning, over separate data paths, registers connected to the outputs of the coin tone receivers in the CDAs to gather data on coin deposits. The coin deposit data, including the time when the deposits occurred, is stored by the PROCON in "scratch pad" memory. The deposit and time data are used to determine future actions, i.e., additional announcements or reports to the TSPS processor when the deposits are sufficient.

Both the active and standby controllers have access (one at a time) to all CDAs. The two controllers do not operate synchronously with matching for error detection. Each, however, contains a close tolerance crystal-controlled clock, so any time difference between them would be only a few microseconds, a difference not noticeable to customers in the event of a switchover. The maintenance and initial loading of the announcement stores takes place over the store maintenance paths shown in Fig. 2. These paths, in fact, are extensions of the existing TSPS processor store buses.

The updating link shown between the two SSAS controllers is a parallel, 17-bit (16 plus parity), 1-mHz link interconnecting the PROCONs in the two controller frames. This link provides two very powerful features. First, if one announcement store is powered down for maintenance, reloading over the store bus would require human action to set up a tape drive and many minutes of loading time. With the updating link, however, the PROCON in the active controller can, interleaved with normal handling of announcements and coin data, transmit in a few seconds the entire contents of the active announcement store to the just-restored inactive announcement store.

Second, while the active controller is dealing with the ever-changing coin collection data stored in its "scratch pad" memory, the same data are being continuously relayed through the updating link to the memory of the standby controller. This makes it possible to perform a planned or unplanned switchover to the standby controller, usually without interrupting announcements or losing track of the coins that as many as 239 customers may be in the process of depositing.

II. DETAILED HARDWARE IMPLEMENTATION

The development of SSAS includes several areas that are felt to be of some general interest and will be discussed further here. These are:

1. Storage, retrieval, and decoding of announcements.

2. Coin deposit signaling and detection.
3. Automated testing of coin detection and announcement circuits.
4. Physical design.

2.1 Announcement storage and processing

The continuing orders-of-magnitude decrease in the cost of digital memories made it clear that the storage of announcements for SSAS should be digital rather than analog. A major decision for SSAS was whether to use writable store (RAM) or read-only memory (ROM) to store the announcement vocabulary. For some recently developed or proposed "talking" systems, ROM is clearly the better choice. For personal calculators for the blind, for example, the vocabulary is well defined to include just the digits zero through nine and the names of the arithmetic operations. Reloading a volatile memory in a portable calculator would be clearly impractical. Also, ROMs can be packaged more densely. For SSAS, however, although the vocabulary for coin traffic might appear to be constant, it is subject to change when call-handling practices change. There may have to be vocabulary differences among operating companies because of differences in their practices concerning overtime collection. More important, when new features are added to SSAS, a significant amount of new vocabulary will have to be added. If the SSAS vocabulary were stored in read-only-memory, the logistics and cost of managing spares, repairs, and additions appear to be objectionable. With a writable memory of the type already in use in TSPS, no special handling of memory units will be needed. New vocabulary can be distributed to operating companies as needed on reels of magnetic tape, using administration procedures already established for distribution of TSPS software updates. Although data in a writable store are subject to loss in the event of a power failure, the back-up power arrangements in Bell System central offices make a shutdown of both announcement memories unlikely; if it does occur, their contents can be restored from a tape stored in the office.

The SSAS announcement vocabulary presently includes 80 512-ms speech segments and the equivalent of 15 more speech segments containing test tones and timing data for automatic self-testing. Each 512-ms segment requires 16,000 bits, stored in the 40-bit data portion of the words at 400 consecutive addresses in the announcement store.

The semiconductor program/data store recently developed for use in TSPS was selected for use with some modification as the SSAS announcement store. Figure 3 is a photograph of this store. This store frame can be equipped with up to six modules, each holding 32,000 47-bit digital words. The 95 presently used vocabulary segments require 38,000 digital words, so the frame presently used in SSAS is equipped

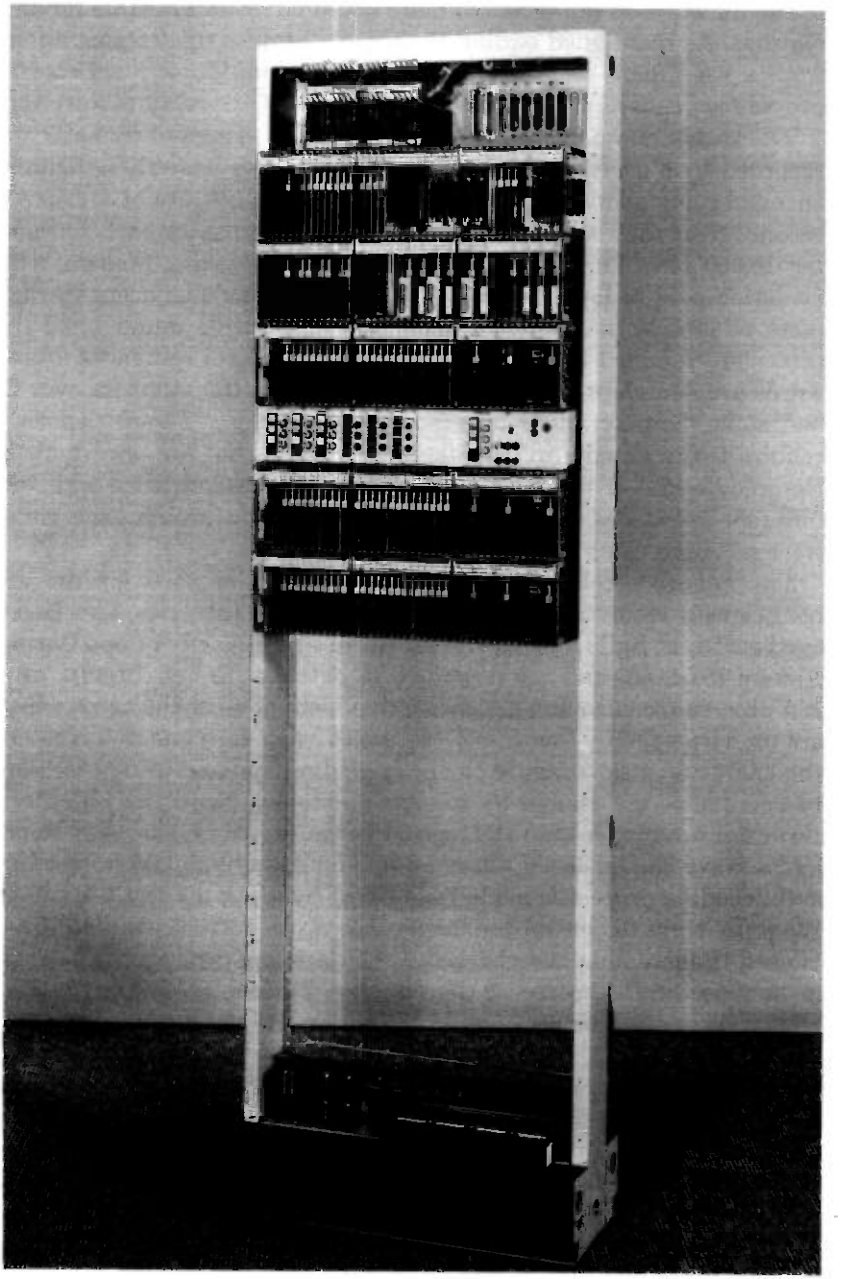


Fig. 3—Announcement store frame.

with only two memory modules. A store frame fully equipped with six modules would accommodate 480 512-ms segments. Since 95 are now in use for ACTS, as many as 385 more could be made available for new features. As mentioned earlier, the announcement store frame differs from a TSPS main store frame only by the addition of a selector to provide access either from the SSAS controller or the TSPS store bus.

The vocabulary of 80 words needed for the SSAS announcements was recorded by a professional announcer. The words were then digitally encoded using adaptive delta modulation with a bit rate of 31,250 per second. Next they were edited into 80 512-ms segments, each containing 16,000 bits. The editing was a subjective listening process, with attention paid to level, pitch, and the silent periods adjacent to each piece of speech, to assure the most natural possible sound when the pieces are rejoined in various combinations to form sentences. Many words are complete in a single 512-ms segment; the numbers over 20 and some common parts of announcements (e.g., "Please deposit") require two segments to complete. The words that are used both in the middle and at the end of sentences are included twice with two different inflections. The 15 equivalent segments containing test tones were recorded using laboratory signal generators.

The adaptive delta modulation (ADM) encoder used for the announcement recording and the decoders used in the CDAS were developed earlier at Bell Laboratories for use in the Subscriber Loop Carrier System (*SLC*-40) and are described in detail in Ref. 1. Briefly, ADM is a process for encoding a signal into a train of ones and zeros which are then decoded by the simple process of using each one to increment the charge on a capacitor in the positive direction and using each zero to decrement the charge by the same amount. Since the bit rate is several times higher than the highest voice frequency, the "stairsteps" in the wave can be easily filtered out. The delta modulation encoding and decoding process is made "adaptive" by using the last few bits in the pulse train to control the magnitude of the current generator that is used to increment the charge on the decoding capacitor. The rates of increase and decrease of the adaptive current generator are controlled by R-C networks tailored to the average parameters of speech syllables.

Adaptive delta modulation produces thoroughly adequate announcement quality using 31,250 bits per second, about half the bit rate that would be needed by 7- or 8-bit pulse code modulation (PCM) using an 8-kHz sampling rate.

The announcement decoder for each of the SSAS CDAS is on a single 6-by-8-inch circuit pack. Each pack includes the decoder circuit from the *SLC*-40 design (adapted for slightly different power supply voltages), a "silence" generator, and a serial input buffer. The silence generator is a flip-flop controlled by a clock to produce alternating

ones and zeros. It is intended to be brought into operation in midword to silence an announcement quickly when a customer starts to deposit coins. This feature allows a substantial speedup in service for customers who already know the charges for calls they are making. The serial input buffer is a self-shifting, first-in/first-out (FIFO) shift register device, available commercially from several manufacturers. It is loaded at a 1.0-mHz rate each time the 40-bit data bursts arrive. It is unloaded continuously into the decoder circuit by the 31,250-Hz clock distributed to all the CDAs.

The sequential distribution of 40-bit announcement segments to all the CDAs is controlled by the PROCON and by wired logic that automatically steps through the 400 store addresses for each speech segment and distributes the data to the 239 CDAs and one test circuit. The retrieval of the words from the 400 consecutive announcement store addresses is controlled from a pair of 256-word recirculating shift registers. These shift registers are used in an alternating fashion. One is loaded by the PROCON with the 256 initial addresses of the announcement segments needed for the next upcoming 512-ms period. One particular address is used to represent silence for inactive or unequipped CDAs. The second shift register, which was previously loaded with the initial addresses for all the announcements (or 512-ms silence segments) in progress, is recirculated 400 times, and each time a one is added to all the addresses. Thus the set of 256 starting addresses is altered at each recirculation to step through the 399 addresses that follow each initial address. These addresses are used to retrieve the announcement data words, which are then distributed in sequence to the CDAs for decoding. After the 400 recirculations, the roles of the two recirculating shift registers are reversed. The recirculation and incrementing begins with the new set of starting addresses, and the just-exhausted shift register is loaded with the still newer set of starting addresses for the next 512 ms of announcements. Each 512-ms interval, when a new set of announcement addresses is loaded, is referred to as a "base period."

2.2 Coin deposit signaling and detection

Coin deposits were reported to an operator for many years through a largely mechanical system. In the widely used 3-slot coin station, the coins rolled and bounced against gongs to produce "bing" for a nickel, "bing-bing" for a dime, and "bong" for a quarter. To provide added flexibility for changes in the initial deposit on local calls and to provide substantially more protection against counterfeit coins and slugs, a new single-slot coin station was introduced in 1966. In this coin station, the coins, after passing through a mechanism that tests them for dimensions, mass, and conductivity, trigger an electromechanical device called a totalizer. After each coin passes, the totalizer resets itself,

and in so doing, momentarily switches on a pulsed-electronic oscillator that produces one "beep" for a nickel, two for a dime, and five for a quarter. In principle, single-frequency "beeps" from such a coin station could be recognized by a tuned electronic detector and counter. However, Bell System experience with multifrequency and *TOUCH-TONE*® signaling indicated that adequate protection against errors caused by speech or noise could be provided only by using dual-frequency "beeps" to represent coin deposits. Accordingly, a low-cost dual-frequency oscillator assembly was designed and introduced into the manufacture and refurbishment of coin stations starting in 1975. This was done deliberately well in advance of the service cutover of SSAS (late 1977) so that a minimum of coin station modification visits would be needed when Automated Coin Toll Service is introduced to an area.

2.2.1 Coin tone receiver operating environment

Coin tone signaling may take place in the presence of ambient speech and noise. Speech interference, for example, may be due to a synthesized announcement in progress, customer speech, an operator talking, or background noise at the coin station at the same time that coins are being deposited. The SSAS coin tone receiver is required to respond correctly to these coin tone signals in the presence of such speech or noise interference. The human ear (and mind) usually has no problem identifying tone signals in the presence of speech and can easily differentiate tone signals from speech. Electronic detection of tone signals, if it involved only receivers tuned to the specific frequencies, would be vulnerable to errors from speech signals since vowel sounds in speech frequently contain frequency components in the recognition band of the coin tone receiver.

The speed of operation of the electromechanical totalizers in the coin stations is affected by temperature, by the dc current available from the loop to the central office, and by wear of the totalizer parts. Consequently, the coin tone receiver must accept a coin deposit signal whose timing varies over a considerable range.

2.2.2 Coin tone receiver overall design philosophy

The basic design requirements can be summarized as follows. The coin tone receiver should:

(*i*) Recognize coin tones from widely ranging coin stations.
(*ii*) Identify a coin station whose performance is outside of requirements.
(*iii*) Operate in the presence of speech.
(*iv*) Reject coin simulations.

The first two requirements are met by accurate timing of the received signal to identify both valid coin deposit sequences and those

from coin stations with defective timing mechanisms. The third and fourth requirements, operating in the presence of speech while rejecting coin simulations, are major considerations and have a significant impact on the receiver's design. As was mentioned earlier, speech may frequently contain sufficiently sustained tone components that simulate coin deposits. By using dual-frequency coin signaling for ACTS, where both tones have to be present simultaneously for a signal to be valid, the coin simulation rate is reduced by a considerable degree. A simple detector that looks only for the presence of the two required frequencies, however, still does not provide adequate simulation immunity. Further receiver protection against coin simulations by speech is obtained by looking at other energy (called guard energy) besides the signal frequencies. If the guard energy is sufficient, the "signal" is assumed to be speech, and tone detection is blocked even though signaling energy may also be present. However, speech from the operator, announcement, or the calling station may be present while coins are being deposited. Thus receivers designed to give good coin simulation protection may be blocked or "talked down" when speech or noise is superimposed on tone signals. This can be seen in Fig. 4, where ambient speech interferes with and partially blocks tone detection and at the same time causes false detection (coin simulation) during the silent interval between bursts. Since the cure for one problem makes the other worse, a compromise but exacting choice must be made in establishing receiver operating parameters.

To reduce the effect of "talkdown" and coin simulation due to the operator and/or announcement, a four-wire terminating set is used to isolate the receiver from speech signals directed toward the originating station. However, because of nonideal return losses associated with trunks, loops, and various terminations, some portion of this speech energy is reflected back to the coin tone receiver. The returned signal

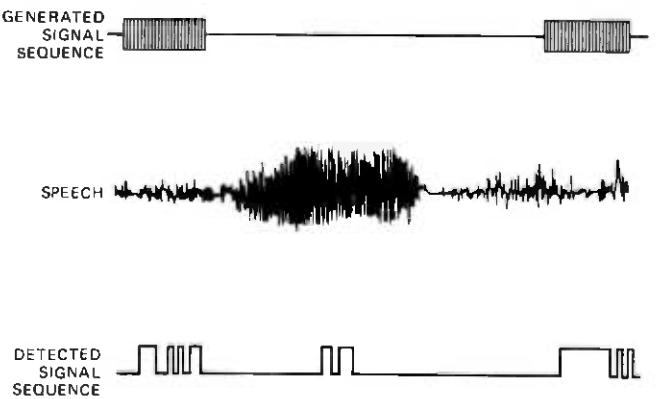


Fig. 4—Coin signal in the presence of speech.

is usually at a considerably lower level so that the probability of interference with signal reception is substantially reduced.

The interfering effect of the announcement is reduced even further by truncating it at the earliest indication of a received coin tone signal. The receiver sends to the ADM circuit pack in the CDA a command to truncate the announcement even before the minimum duration legitimate coin burst is timed. Hence, if a false detection were due to signaling frequency components within the announcement itself, the signal would disappear before a coin simulation could be produced. If the detection were legitimate, truncation would prevent any further talkdown by the announcement.

2.2.3 Coin tone receiver subdivision

Notwithstanding the coin simulation and talkdown protection provided by the four-wire terminating sets and announcement truncation, the receiver itself must have additional safeguards to distinguish between speech and coin tones, yet recognize tones in the presence of speech. This is accomplished in a two-phase processing approach: tone recognition and tone validation. The front-end analog tone recognition portion of the receiver "detects" the signal while providing the initial balance between coin simulation protection and "talkdown" protection; the digital tone validation timing portion then applies different validity standards to various portions of the "detected" signal to determine the coin denomination.

The analog portion (Fig. 5) is similar in principle to existing receiver designs for *TOUCH-TONE* signaling. (*TOUCH-TONE* signaling and receiver design considerations are described in Ref. 2.) It consists of filters, limiters, and detectors which produce a logic 1 output to the timing circuitry when both signal frequencies are simultaneously detected. As was mentioned earlier, tone detection is blocked when sufficient "guard" energy is present. The input bandpass filter (BPF) controls the total frequency range of signals entering the receiver. If this filter is broad, a considerable spectrum of the speech energy entering the receiver will be applied to the limiters. If the speech contains components at the signaling frequencies, it is likely that these components will be dominated by the other speech components, which will "capture" the limiters. Thus the signaling components passing through the bandpass filters, which follow, will not be of sufficient amplitude to operate the detectors. This method of preventing coin simulation is known as "limiter-guard action" and is the technique employed in *TOUCH-TONE* signaling to reduce the incidence of digit simulations. Therefore, a broad input BPF provides good coin simulation protection. If, however, legitimate coin signals are being transmitted while speech is present, they may also be blocked by speech.

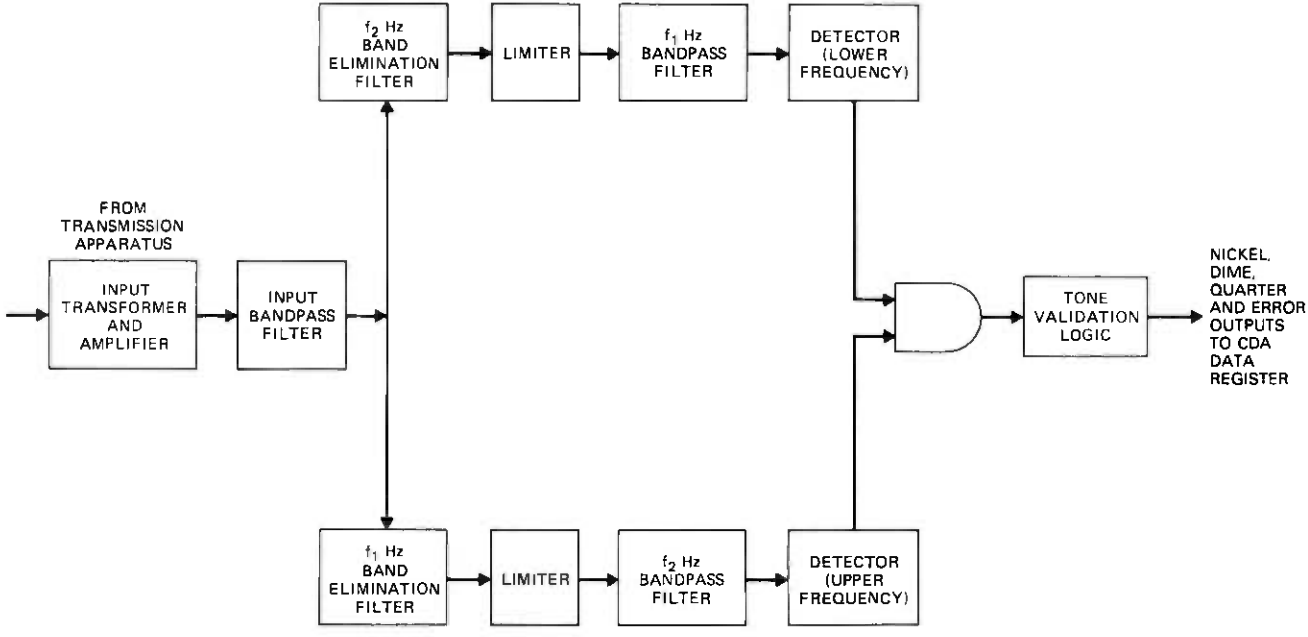


Fig. 5—Coin tone receiver.

Hence, "talkdown" protection will be poor. If the input BPF is narrowed (the input BPF must be at least wide enough to pass the two signaling frequencies), less speech energy can get through and, while "talkdown" is improved, less protection is provided against coin simulation. Based on laboratory and field tests, a filter was selected to optimize the tradeoff between the detector's coin simulation and "talkdown" protection.

The logic signal indicating dual-tone detection is applied to the receiver's tone validation timing circuitry which performs the following functions:

(*i*) Processes the detected signal (refer to Fig. 4 for a sample detected signal) to form "discrete" burst and silent interval logic signals in accordance with the timing algorithm by filling in some gaps while ignoring some burst portions.

(*ii*) Totalizes the number of "discrete" bursts received.

(*iii*) Categorizes each "discrete" burst or silent interval after it is formed and makes a preliminary determination of the coin denomination at that time.

(*iv*) Checks for consistency with previous burst and silent intervals to narrow the denomination possibility for that coin and also check for coin station timing malfunctions.

(*v*) Outputs the coin denomination or timing error signal when certain conditions are satisfied regarding the number of bursts received, the apparent coin category, and the amount of silence since the end of the last burst.

2.2.4 Receiver outputs

As the receiver performs signal timing in accordance with the timing flow diagram in Fig. 6a, certain output conditions may be satisfied and an output representing one of the coin denominations or a timing error (representing an out-of-tolerance coin station) will be delivered to a CDA data register. Before any coin or timing error outputs can be sent, however, the receiver must first deliver a Signal Processing (SP) signal. The SP signal is sent when the timing of the first burst of a suspected coin deposit reaches a certain minimum value. It remains active until a coin identification or timing error output signal is transmitted or the suspected coin deposit is deemed to be a coin simulation. At that time processing for the coin is assumed to be completed. The SP signal is used in the announcement decoder to truncate the announcement (see Section 2.2.2) by switching in the silence generator. The receiver also sends a DST (data strobe) signal along with the outputs, which gates the coin denomination or timing error indication into a data register in the CDA.

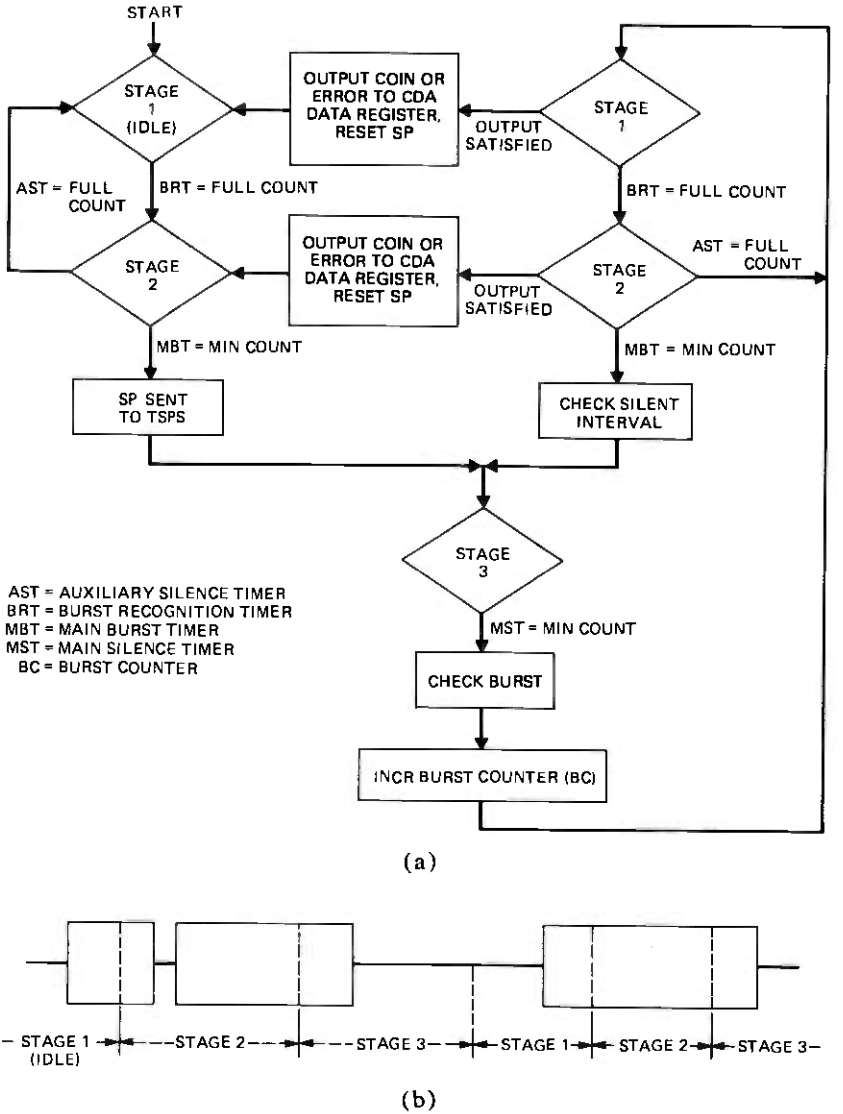


Fig. 6—Timing algorithm. (a) Algorithm flow. (b) Sample detected signal sequence.

2.2.5 Circuit implementation

The tone detection portion of the receiver (see Fig. 5) is based on existing receiver designs for *TOUCH-TONE* signaling. The input bandpass filter, which is critical to the detector's response to speech and also to its performance when tones are being received in the presence of speech, was discussed in detail in Section 2.2.3.

Following the bandpass filter are two band-elimination filters (BEFs) in parallel. The width of these filters is tailored to the tolerances of the coin station oscillators. The first filter rejects the higher of the two signaling frequencies, and the second filter rejects the lower. Thus the output of the high BEF will contain the lower of the two signaling frequencies, and the output of the low BEF will contain the upper signaling tone. The limiters that follow convert the signals to square waves. Hence, at this point, there are two square waves, one out of each limiter.

These signals now pass through bandpass filters centered at the two signaling frequencies. The square wave out of the low frequency limiter passes through a filter centered at the low frequency and a sine wave output is produced. The same occurs for the upper frequency square wave. Following the filters are threshold detectors that detect the sine waves if they are above a certain threshold level. The limiter employs a feedback arrangement to control the limiter operating threshold. The two detector outputs are combined to produce a logic output that is timed by the receiver's tone validation circuitry to determine the denomination of the coin.

Wherever possible, components originally designed for *TOUCH-TONE* signaling are used for design implementation. These include the limiter-threshold generator circuit module, STAR (standard tantalum active resonator) filter circuit modules similar to those for *TOUCH-TONE* signaling, and the detector modules. These circuit modules are hybrid integrated circuits. The analog tone detector circuitry occupies two 4- by 8-inch printed wire boards. The filters are on one board, and the remaining detector circuitry is on the other.

Tone validation timing is done digitally, with the circuitry consisting primarily of small scale integration (SSI) and medium scale integration (MSI) TTL devices. A 1-kHz clock signal synchronizes all tone validation timing operations within the receiver. This clock sets the rate at which the tone detect signal from the analog circuitry is sampled. Based upon whether the dual tone detector is high or low at the sample instant, the various gates, counters, flip flops, etc., are operated in accordance with the tone timing validation algorithm. This tone validation circuitry takes up four circuit packs, making a total of six packs for the entire receiver. Subsequent to deployment of the random logic timing circuitry, a cost reduction was made by having a microprocessor perform the timing algorithm. The entire tone timing validation circuit was placed on one circuit pack, lowering the overall circuit pack count to three.

2.3 Automated testing of announcement and coin detection circuits

The use of digital speech storage and PROCON control in SSAS made it possible, with a very small amount of added hardware, to provide

automated testing of the speech decoding and coin tone receiver circuits.

The adaptive delta modulation decoding circuits in the CDAs are tested by using four single frequency tones digitally stored in the announcement store. Three of the tones are within the frequency band of the voice announcements, and the fourth is just above that band. The three in-band tones verify the flatness of response of the decoder and the transmission network associated with it. The out-of-band tone verifies correct roll-off response of an active low-pass filter included in each CDA to filter out the 31,250-Hz "stairsteps" that result from the decoding process.

Under control of a diagnostic program, each CDA is periodically taken out of service, and the digitized test tones are distributed to it from the standby SSAS controller frame. The TSPS processor simultaneously sets up a network connection from the output of the CDA circuit under test to the CDA test circuit. The CDA test circuit connections for decoder testing are shown in Fig. 7. The CDA test circuit contains a bandpass filter that passes all the test frequencies, followed by a level detector that delivers "go/no-go" responses to indicate when the detected levels are in or out of tolerance. Also included in the CDA test circuit are four active "notch" filters followed by a detector and a smoothing filter. The notch filters remove the fundamental of each test frequency, and the detector responds to the residue, which includes harmonics and noise resulting from the adaptive delta modulation encoding/decoding process. The detector delivers a logic level "no-go" output if the residue is higher than normal. The smoothing filter is needed because, after the fundamental has been removed, the residual combination of harmonics and decoding products is partly random,

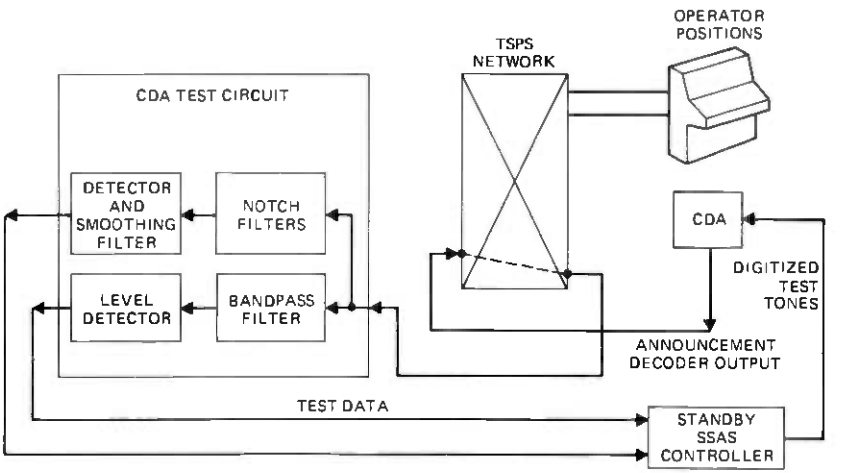


Fig. 7—Testing CDA decoders.

and without smoothing it would cause occasional "false alarm" responses at the detector output.

Testing CDA decoder circuits in the way described above verifies the integrity of the logic circuits and wiring that retrieve and distribute digitized speech segments, and it checks for continuity and proper alignment of the voice frequency transmission apparatus that connects the CDAs to the TSPS trunks.

The coin tone receivers are also tested by using tones stored in the announcement memory. The diagnostic program for the coin tone receivers directs the digitized coin test tones to a CDA decoder circuit pack that is dedicated for this purpose. The output of this decoder is routed through a solid-state switch controlled to produce tone bursts that simulate the coin station output signals that represent nickels, dimes, and quarters. The tone bursts are passed through switchable attenuators and through TSPS network connections to each coin tone receiver in turn. CDA test circuit connections for coin tone receivers are shown in Fig. 8.

A very comprehensive test sequence is applied to the coin tone receivers using the scheme described above. The coin test tones are recorded and stored with nominal frequencies and frequencies just inside and just outside the operating tolerances on both sides of the nominal frequencies. The solid-state switch is operated to produce the nominal tone bursts that check the receiver's detection of the various nominal and edge band frequencies and a variety of non-nominal durations and sequences that exercise the receiver's timing logic. The switchable attenuators control the levels of the tone bursts in the test sequence to provide verification that the coin tone receivers operate over the range of levels that result from coin station and transmission

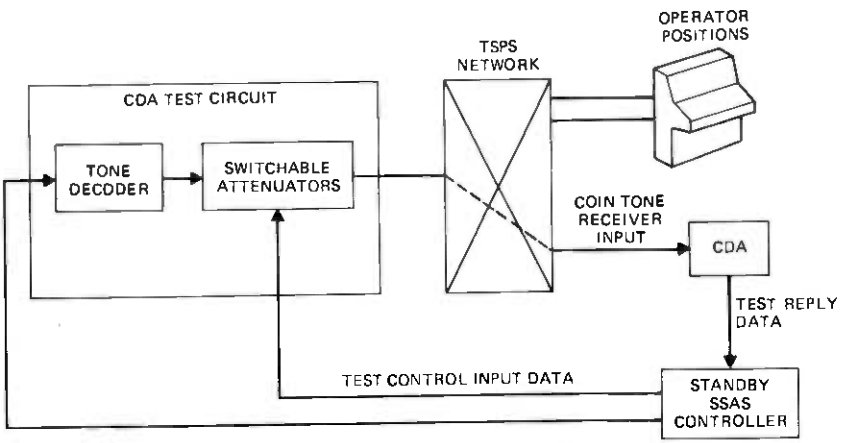


Fig. 8—Testing CDA coin tone receiver.

loss variations. The test sequence for each coin tone receiver is completed automatically in a few seconds.

The CDA test circuit includes four circuit packs identical to a set of four used in each CDA, and five additional packs unique to the test circuit. The CDA test circuit also has a self-test mode. In this test mode, the output of the tone generation and control circuits of the CDA test circuit are temporarily fed back to the input of the filter/detector circuits and a "wrap-around" test sequence is executed.

2.4 Physical design considerations

A complete SSAS installation includes two controller frames, two announcement store frames, from 2 to 15 pairs of frames containing coin tone receivers, announcement decoding circuits, voice frequency terminating sets and amplifiers. The coin tone receivers and announcement decoding circuits are mounted in service circuit frames (up to a maximum of 16 sets per frame) and the voice frequency circuits are similarly mounted in transmission frames.

2.4.1 Controller frame

The controller frame (Fig. 9) is a single-bay frame, 2-ft 2-in. wide and 7-ft high, using a 12-in. deep framework, as do most other TSPS frames. This frame is compatible with all other TSPS frames and requires no special hardware, mounting, or installation arrangements.

The communication bus and interconnection unit at the top of the controller frame contains multi-pin terminal strips for terminating cables from connecting frames. The connections to the announcement store are also made through this unit. Transformers are furnished for connection to the TSPS buses. Bus connectorization for growth is provided, and is compatible with existing office arrangements.

The logic unit in the upper part of the controller frame includes six levels of circuit packs and a control panel unit. Five-volt power and ground return for the circuit packs are provided by a double-sided, printed-wiring back plane for each of the six levels. Each level is split into two halves for power distribution. Five levels can accommodate up to 37 circuit packs on ½-in. centers. The sixth level can accommodate 28 circuit packs on ½-in. centers. A very large percentage of the backplane wiring is automated. The type of circuit pack used is a double-sided printed wiring board, 6 in. by 7 in. in size, with 80 pins. An example of this circuit pack is shown in Fig. 10.

The control panel unit contains the various keys, lamps, and jacks for maintenance and frame control functions.

The memory and control unit provides mounting arangements for the PROCON and its associated memory packs. Connectorized cables are provided within this unit and between parts of the logic unit.

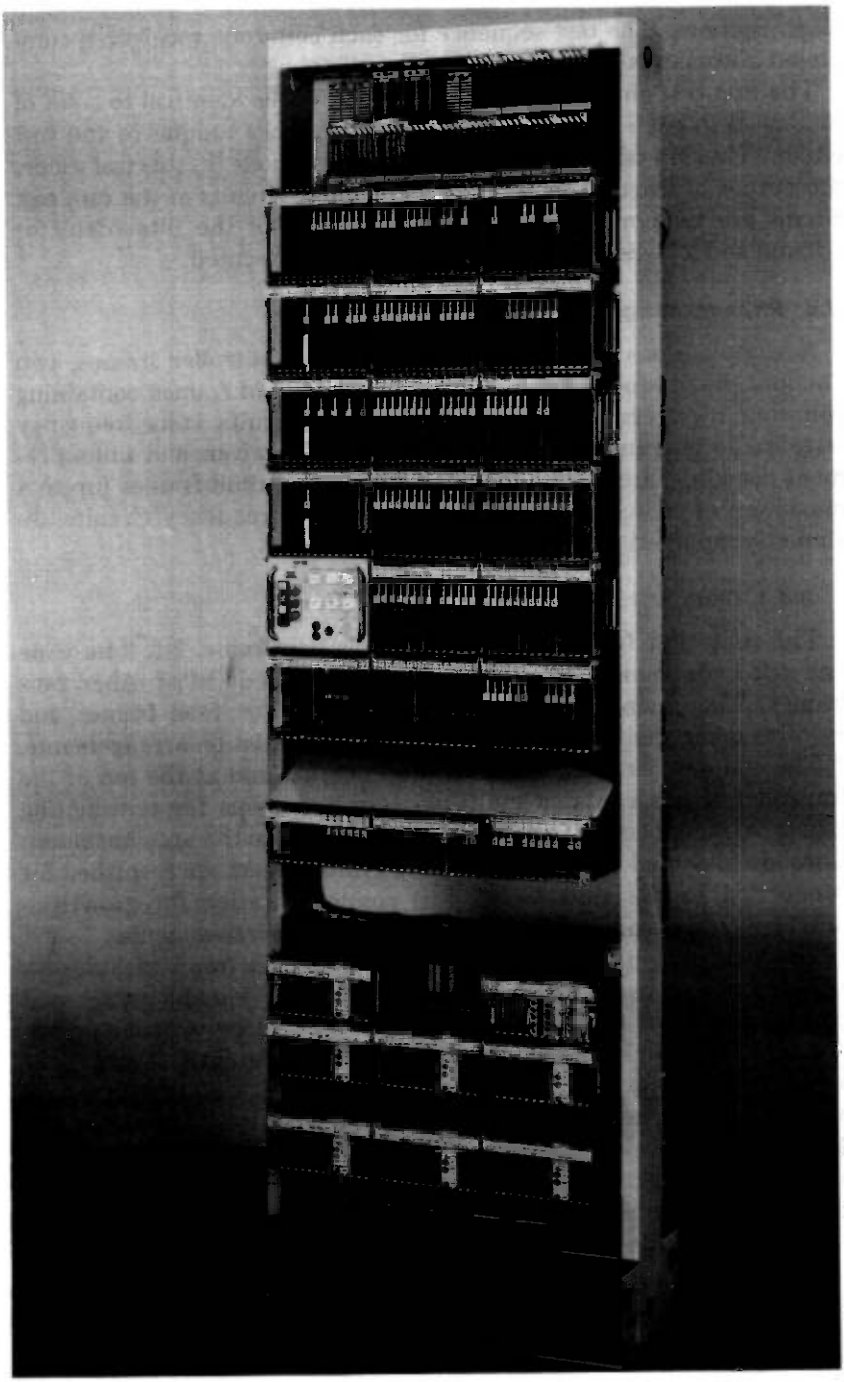


Fig. 9—Controller frame.

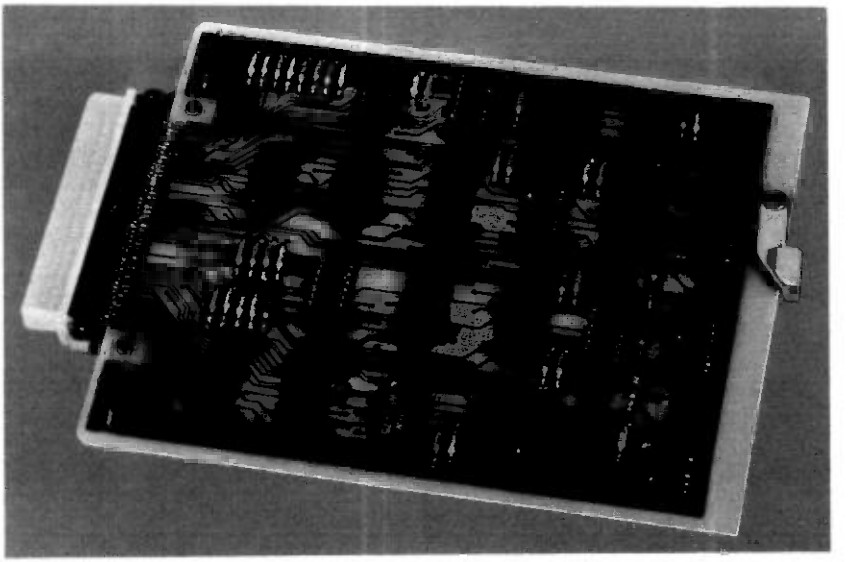


Fig. 10—Controller frame circuit pack.

The power converter unit contains +5 V and −12 V dc-to-dc converters and fuses. These power units are pluggable for easy replacement. A +5 V load fuse and a pilot fuse are provided for each half of the circuit pack levels and for the memory and control units.

The fuse panel unit contains −48 V and +24 V fuses, alarm relays, and power control relays. The −48 V is used to supply the dc/dc converters. The +24 V is used in the circuit packs that contain bus drivers.

2.4.2 Service circuit frame

The service circuit frame (see Fig. 11) is a single-bay frame, 2 ft, 2 in. wide and 7-ft high, using a 12-in. deep framework. Each frame may contain as many as 16 CDAs. A maximum capacity SSAS installation would include 15 service circuit frames to hold 239 CDAs plus one CDA test circuit.

The service circuit frame uses multi-pin terminal strips and bus coupling transformers for terminating cables from other connecting frames. In addition, the unit is arranged to hold three levels of circuited packs. These packs are associated with the 16 CDAs and the circuits that provide switchable access to either controller frame. The +5 V power is supplied individually to each service circuit and to each of the group controllers. Printed-wiring back planes provide +5 V ground return for each pack. A large part of the unit wiring is automated. This unit is always wired for 16 circuits and is equipped by plugging in circuit packs as determined by traffic requirements.

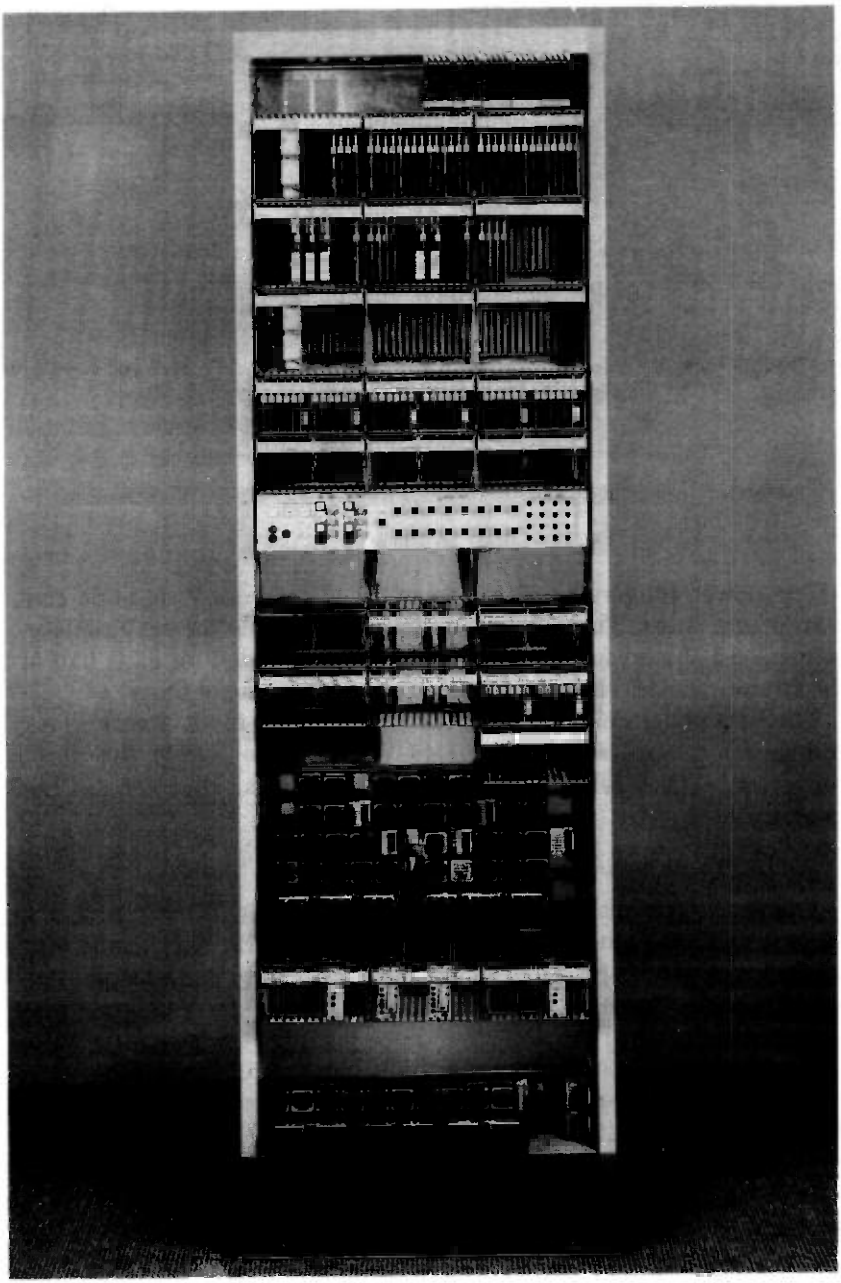


Fig. 11—Service circuit frame.

The control panel unit contains the various keys, jacks, lamps, and switches associated with maintenance and frame control functions. Each CDA is controlled individually.

The power converter unit contains +5, +6, and −6 V dc-to-dc converters and fuses for the +5 V distribution. These power units are pluggable for easy removal. A +5 V load fuse and a +5 V indicator fuse is provided for each service circuit.

The fuse panel unit contains −48 and +24 V fuses, alarm relays, and power control relays. The −48 volt is used to supply the dc/dc power units. The +24 V is used for bus drivers.

Even-numbered frames and odd-numbered frames are fused from separate power distribution frames.

Space for up to 16 coin-tone receivers is provided on all frames except the first frame of the subsystem, where the CDA test circuit unit replaces one of the coin-tone receivers. The coin-tone receivers are connected to their associated CDA circuits by plug-ended cables to facilitate growth and rearrangement due to changes in traffic patterns.

2.4.3 CDA transmission frame

Associated with each service circuit frame is a single-bay frame (see Fig. 12) containing voice frequency transmission equipment.

Growth and rearrangements are accomplished with pluggable apparatus.

2.4.4 Announcement store frame

Associated with each controller frame is an announcement store frame (see Fig. 3) that provides storage for the digitally encoded announcement segments.

The bus unit at the top of the store frame contains terminal strips for interconnecting cables from connecting circuits in the office. Transformers are furnished as part of this unit to connect to the TSPS processor store bus. Store buses are connectorized to facilitate growth.

As mentioned earlier, two memory modules are required for Automated Coin Toll Service. Additional memory modules up to a total of six may be added to provide storage for additional vocabulary words if required by future features.

III. ACKNOWLEDGMENTS

The development of circuits for Automated Coin Toll Service called on the talents of many people at Bell Laboratories. The authors would like to especially thank the following: For design of the dual-frequency coin stations, G. Breitung and J. E. Edington; for encouragement and guidance in the design of coin tone receiving circuits, C. G. Morrison; for skillful laboratory and field testing of coin tone receiving circuits,

Fig. 12—Transmission frame.

M. Schmidt; for guidance and assistance in physical design, R. D. Garside, H. A. Hilsinger, and W. J. Proetta; for skillful electrical design and testing of SSAS circuits, J. O. Croswell, R. L. Grenzow, J. O. Hardy, D. C. Peterson, T. L. Roush, and B. G. Ruel.

REFERENCES

1. R. J. Caniff, "Signal Processing in *SLC*-40, A 40 Channel Rural Carrier Subscriber Carrier," International Conference on Communications, San Francisco, California, June 1975.
2. R. N. Battista, C. G. Morrison, and D. H. Nash, "Signaling System and Receiver for *TOUCH-TONE*® Calling," IEEE, Trans. Commun. Elec., March 1963, pgs. 9–17.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Automated Coin Toll Service: Software

By R. AHMARI, J. C. HSU, R. L. POTTER, and S. C. REED

(Manuscript received December 28, 1978)

The Traffic Service Position System (TSPS) No. 1 operational software for Automated Coin Toll Service provides the logic which controls the handling of coin-originated calls served by the system. The partitioning of responsibilities between the Station Signaling and the Announcement Subsystem (SSAS) are illustrated. The maintenance philosophy, fault detection, diagnostics, and fault recovery aspects of the SSAS are described. The maintenance strategy is centered on a multilevel fault detection scheme in which faults are analyzed and classified according to their degree of seriousness in affecting the SSAS operation.

I. OPERATIONAL SOFTWARE

Automated Coin Toll Service (ACTS) is a feature of the Traffic Service Position System.[1] A Station Signaling and Announcement Subsystem (SSAS) has been added to TSPS to detect and process coin deposits and to construct announcements for coin sent-paid toll customers.

The SSAS contains a programmable controller (a microprocessor), which has a Read Only Memory (ROM) for program and a Random Access Memory (RAM) for transient data. It also has a set of Coin Detection and Announcement circuits (CDAS). These circuits detect coin deposit signals from coin stations and decode digital speech phrases to produce analog announcements. The SSAS also has an Announcement Source and Distributor (ASD) which contains an announcement memory with digitally encoded speech phrases (see Fig. 1). Commands to the SSAS are sent by the TSPS processor over the

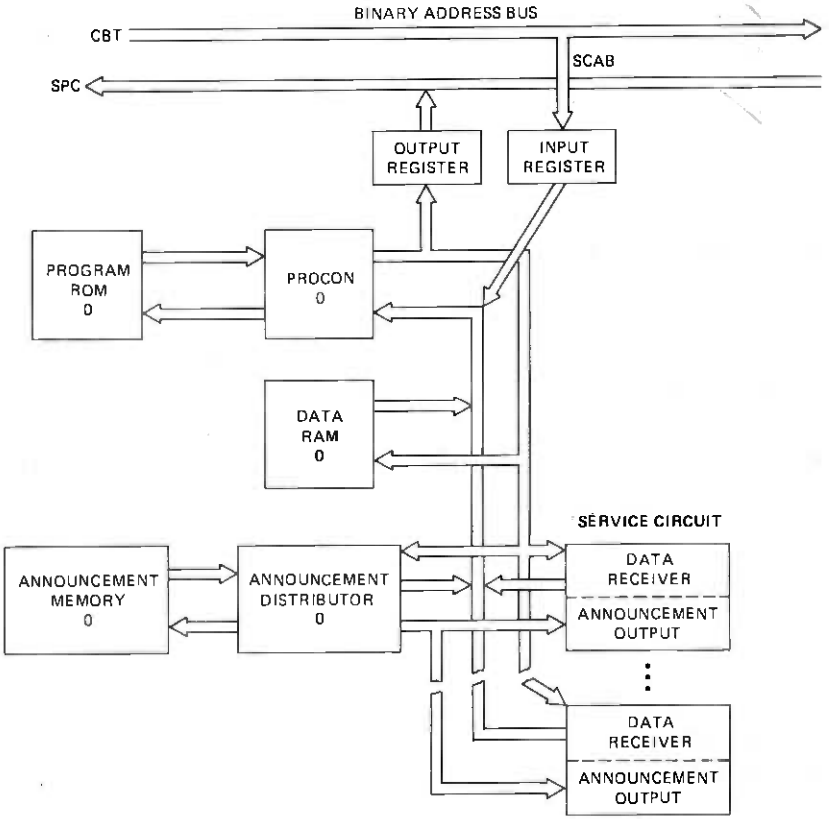


Fig. 1—SSAS configuration.

Peripheral Unit Address Bus (PUAB). Replies to the TSPS processor are sent over the Scan Answer Bus (SCAB).

1.2 Functional description of ACTS call processing

1.2.1 Initial call setup

When the TSPS call connections program* receives a report of an incoming trunk seizure† from the supervisory scan program, it establishes the required network connections for called digit and calling digit reception. (The latter is for Automatic Number Identification [ANI] offices.)

The ANI digit analysis program assumes control until reception of the calling party's number is completed.

---

* This program runs on the TSPS main processor.

† The description that follows deals only with calls that originate on trunks on the base network. Calls that originate on the network of an RTA are also handled by the operational software but not described below.

When the calling party identification has been received, control is returned again to the call connection program. At this time, a general analysis is performed on the information obtained, and the call is marked as 0+, 1+, etc. The 1+ coin-originated calls are candidates for automated treatment. (Coin customers expecting to make deposits to pay for a station-to-station call will dial the call with a "1" prefix or no prefix.)

1.2.1.1 Initial ACTS processing. The next step in processing the call is to determine whether the call can be automated. This is done by a program in the main TSPS processor.

Conditions for Automation. 1+ calls that satisfy the following criteria are candidates for automation during the initial contact on the call.

(*i*) ACTS-Converted Trunk Group. The call must be on an ACTS-converted trunk group. Certain modifications are needed in the coin station to generate dual-frequency coin deposit tones which the Coin Detection and Announcement circuits can recognize. All coin stations served by a trunk group must be modified before any calls on that trunk group can be automated.

(*ii*) Machine Ratable. The call must be machine-ratable; i.e., TSPS must receive or have in office data sufficient rating information to calculate the charges due on a call.

(*iii*) Not a Postpay Coin Originating Station. The call must not be from a postpay station. Coins deposited at postpay coin stations cannot be returned. An operator must verify that the correct party or station has been reached before the customer makes any deposit.

(*iv*) Not a Large Charge Call. The call must have an initial charge less than a certain threshold. Coin station hoppers handle only limited numbers of coins (24 nickels, for example). Operating practices instruct operators to make partial collections for every two to three dollars deposited if the call has large charges. Large charge calls require multiple collections. Hence, an operator is required to verify called party answer before any coins are collected.

(*v*) Automatic Number Identification. The call must have successful Automatic Number Identification. If an ANI failure occurs or the call is Operator Number Identified (ONI), the call cannot be rated. An operator must key the calling number.

Coin deposit monitoring on calls that fail condition (*i*) is not automated but is handled by operators using current procedures. However, notification at the end of the initial period can be automated on all

coin calls, even if dual-frequency oscillators are not installed in the coin stations.

Calls that fail conditions (*ii*), (*iii*), or (*iv*), as well as person-paid, coin-originated calls (which are dialed as 0+ calls), seize a position for the required operator assistance. If the trunk group is ACTS-converted, a Coin Detection and Announcement circuit is attached to assist the operator in counting deposits. Subsequent deposit monitoring for overtime can be fully automated unless the large charge threshold is exceeded.

Calls that fail only condition (*v*) seize a position for the purpose of acquiring the calling number. After the number is keyed in by the operator, further handling of the call is automated.

If the above conditions for automation are met, the call connections program seizes an idle Peripheral Order Buffer (POB) and transfers control to the network control program which loads the POB with orders to establish a connection between the calling customer and a Coin Detection and Announcement circuit. If a CDA is not available, the call connections program seizes an idle POB and transfers control to the network control program which loads the POB with orders to establish a connection to both a position and an outpulsing circuit. The call is subsequently handled as a non-ACTS coin call.

SSAS Processing of Initial Deposit Requests. Processing in the SSAS begins when the Programmable Controller (PROCON) receives an initial deposit request command from the TSPS processor. The command is read by PROCON from the SSAS input registers. The command message includes the number of the CDA handling the call, the initial duration of the call, and the amount to be requested from the customer (see Fig. 2). For purposes of illustration, assume the charge is $1.15 and the initial period is three minutes.

The command field of the message is used as an index into a transfer table. The PROCON transfers to the address retrieved from the table. That program records the call data in the SSAS data RAM, sets the initial state indication for the call, and sends an output command to initialize the CDA circuit.

The PROCON next initiates scanning of the CDA circuit for coin deposits. The PROCON will continue scanning the CDA circuit at least once every 250 ms for the duration of the initial deposit phase of the call. The PROCON also instructs the SSAS announcement control cir-

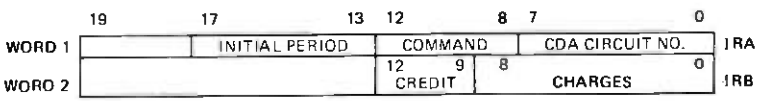


Fig. 2—Typical SSAS input message from TSPS processor. Message is in format as sent from TSPS processor (two 20-bit words).

cuitry to send the first segment of the appropriate announcement phrase to the CDA circuit. It sends subsequent commands for successive portions of the announcement every 512 ms. The announcement used for the call being described is:

"One dollar and fifteen cents,* please." (2-second pause) "Please deposit one dollar and fifteen cents* for the first three minutes."

If the customer deposits during the announcement, the CDA circuit instantly inhibits the announcement. When the PROCON scans the CDA circuit and recognizes the deposit, it ceases sending further announcements, adds the value of the coin deposit to the previous amount deposited, and compares the total with the amount due. Assuming a sufficient deposit has not been made, the PROCON begins timing. If the customer fails to deposit within *five or six seconds*, a prompting announcement is provided indicating the amount still to be deposited. For example, if the customer deposits three quarters in the above example and then stops, the prompt is:

"Please deposit forty cents more."

If the initial deposit announcement completes without a deposit, timing will begin at that point. If the customer has made no deposits and a time-out occurs, the announcement wording for the prompt is:

"Please deposit one dollar and fifteen cents."

Each deposit made causes the PROCON to reset its software intercoin timing register for the call.

The PROCON will report the final results of the initial deposit request to the TSPS processor by loading a message into the SSAS output FIFO buffer (see formats in Fig. 3). There are basically three situations possible.

(*i*) If the customer has failed to deposit within five to six seconds after a prompt, the PROCON sends a reply (Fig. 3, format B) to the TSPS processor indicating this fact so an operator can be connected to provide assistance.

(*ii*) If the customer deposits the exact amount requested, the PROCON sends reply format A which contains the amount deposited.

(*iii*) If the customer overdeposits by using too large a denomination coin, the PROCON sends reply format A. (The TSPS processor

* The announcement format varies slightly, depending on whether the charge involves a dollar amount, a cents amount, or both.

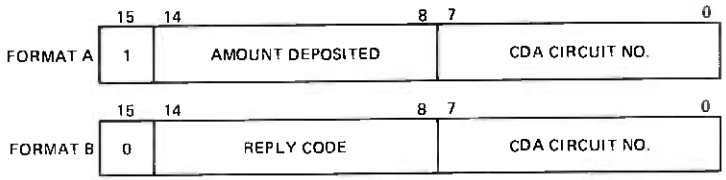


Fig. 3—Formats of SSAS output FIFO buffer.

recognizes the overdeposit and records a credit towards overtime.)

The PROCON acknowledges the latter two cases above by initiating announcements. If the exact amount is deposited, the announcement is:

"Thank you."

If an overdeposit has been made (for example, $1.25 on the $1.15 call described above), the announcement is:

"Thank you, you have ten cents credit towards overtime."

The PROCON continues to scan the CDA for further deposits until the acknowledging announcement is completed and sends a final deposit report to the TSPS processor at the end of the announcement. (This ensures credit for a belated deposit.) Finally, the PROCON places the CDA and the associated call memory in the SSAS data RAM in the idle state. The PROCON performs no further action on the CDA until a new command is received from the TSPS processor.

At any time during the processing described above, the TSPS processor can send a special command which causes the PROCON to idle the CDA circuit and associated call processing memory. (One example of this happening is if the customer hangs up.)

Successful Initial Seizure. Upon receipt of the reply from SSAS indicating an exact deposit or overdeposit, the call is processed to completion. The call connections program seizes an idle POB and transfers control to the network control program which loads the POB with orders to connect an outpulsing circuit. If an outpulsing circuit is not available, the call connections program queues until the circuit is available. The call connections program loads the orders to perform the appropriate relay operations required to complete the connection and activates the POB. Upon successful POB completion, control is returned to the call connections program where the POB is idled. The outpulser loading routine is called next. It loads the digits to be

outpulsed in an outpulsing register and activates sender-attached scanning for the receipt of a sender-attached signal from the toll office. Outpulsing of the called number to that toll office proceeds as described in Ref. 2, page 2658 ff.

Receipt of called party answer and call timing also follow the description in Ref. 2.

Operator Assistance on Automated Initial Seizure. If the customer fails to deposit in response to prompting by the SSAS or flashes the switchhook* to acquire operator assistance, the customer will be connected to an operator. This section discusses the display presented to the operator, the operator actions, and potential race conditions.

If the call must queue for a position, it is given "recall priority" to minimize customer delay. The Coin Detection and Announcement circuit is connected to the calling customer during the queuing interval. If the call is being sent to a position because of a time-out and the customer subsequently satisfies or exceeds the charges while queuing for a position, the SSAS informs the TSPS processor, the call is removed from the position queue, and outpulsing is initiated without operator assistance. When the call reaches the position, the following keys and lamps on the operator console (see Fig. 4) are lit steadily.

The loop access key (ACS)
The appropriate supervision lamps (CLD), (CLG)
The station coin lamp (STA)
The AMA station-paid key (PAID)
The release forward key (FWD)

The lighting of the AMA station-paid key and the release forward key indicates to the operator that an ACTS time-out or a customer switchhook flash has occurred during an ACTS initial contact.

In addition to the above, an ACTS underdeposit display is given to the operator (see Fig. 5). The "charge-minute" designation strip is lit steadily and the numeric field contains, from left to right, up to three digits for charges due, one digit for the initial period and up to three digits for the amount due.

While the call is attached to a position, the operator assists the customer in making a full deposit. The operator must announce the amount due (the right-hand quantity) in the numeric field. The CDA circuit monitors and counts the coin deposits. The numeric display is not automatically updated with each coin deposit. However, the operator can update the display by using the charge and minutes (CHG-

* For a time-out, the SSAS indicates the request for an operator by a message. For a flash, the TSPS processor initiates the request.

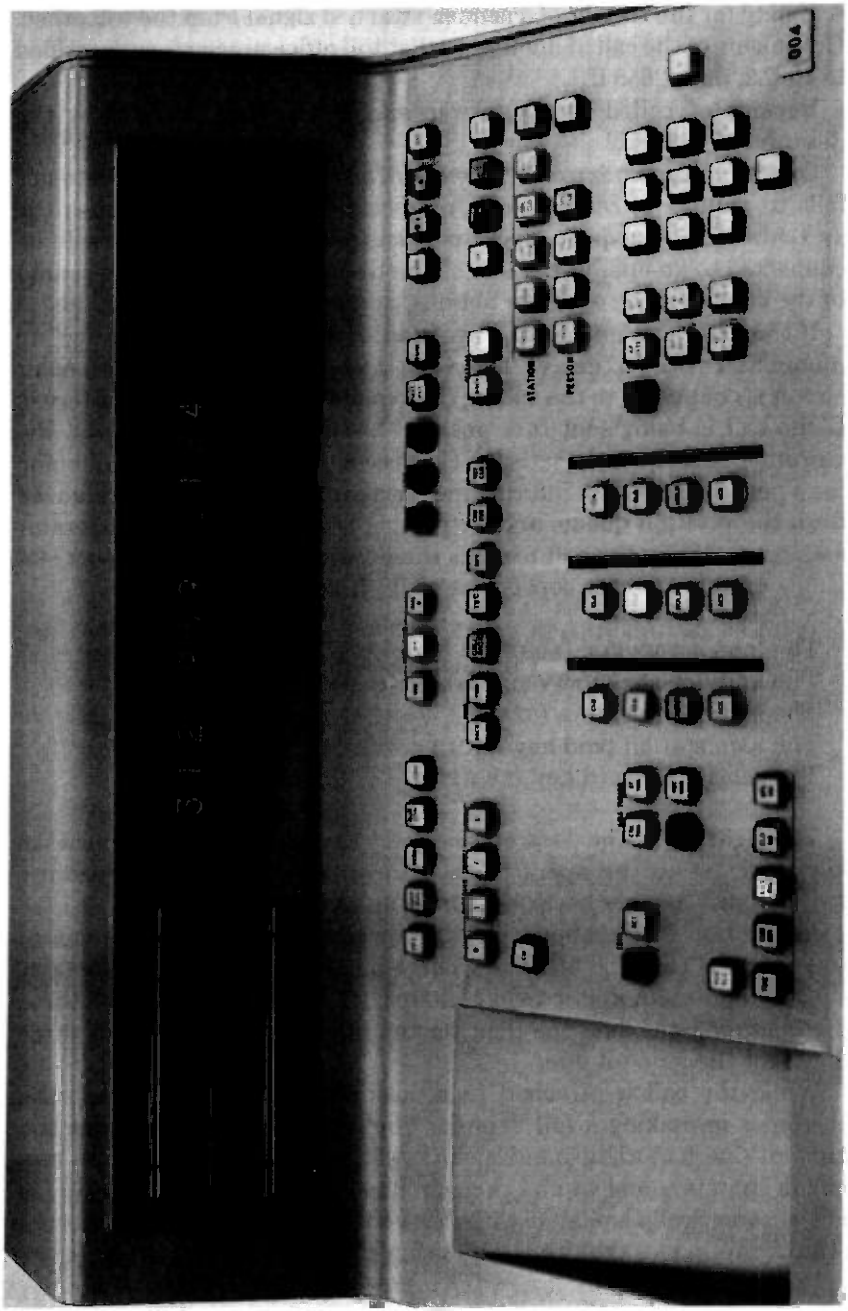


Fig. 4—100B traffic service position key shelf.

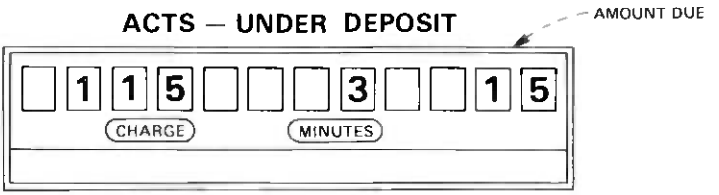


Fig. 5—TSPS No. 1 console numeric display.

MIN) key. (The SSAS will be interrogated for the current amount deposited and the appropriate display presented.)

When the SSAS recognizes that the deposit request is satisfied, the TSPS processor is informed. The forward number is outpulsed and the correct display is presented to the operator (see Fig. 6).

The operator receives several indications that the deposit is satisfied: outpulsing occurs, the amount due display changes to zero, the release forward key darkens, and the start key lights during outpulsing.

After acknowledging the deposit, the operator depresses the start timing* and position release keys.

At position release, the TSPS processor requests the amount detected by the CDA circuit. Upon receipt of the report, the CDA circuit is disconnected from the call. This final report accounts for all money deposited (while the CDA was connected to the call) so no late deposits are missed.

If an overdeposit occurs, the numeric display is changed to an overdeposit display (see Fig. 7). This display has a flashing "charge-minutes" designation strip and a numeric display of zero in the charge field, blanks in the minutes field, and two digits having the overdeposit. The operator informs the customer that credit will be given towards subsequent overtime charges on the call, depress start timing, and position release. On subsequent deposit requests, the overdeposit is automatically subtracted from the charges due.

1.2.1.2 Coin notification. At the end of the initial period, coin sent-paid customers are notified. With ACTS, this no longer requires an operator. The description below applies to coin sent-paid calls from both ACTS and non-ACTS converted stations.

TSPS Processing of Coin Notification. When a call reaches the talking state, it is under the control of the disconnect program. The call is placed on a timing list when the called answer has been established. A time-out occurs 7 seconds prior to the end of the initial period. Subsequent to this time-out, a coin collect sequence† occurs

* Timing on the call is initiated by the TSPS processor only if the called customer answers.

† The coin collect sequence does not occur on postpay coin stations.

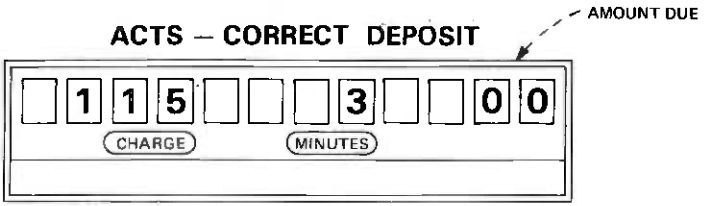


Fig. 6—TSPS No. 1 console numeric display.

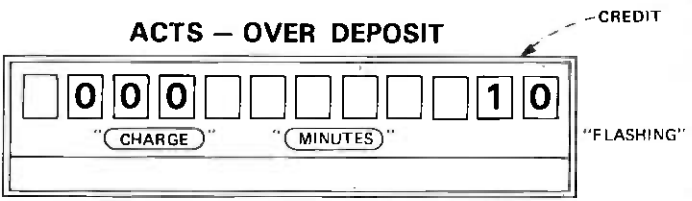


Fig. 7—TSPS No. 1 console numeric display.

and the call is returned to the timing list for the remainder of the initial period.

At the end of the initial period, the disconnect program seizes an idle Peripheral Order Buffer (POB) and transfers control to the network control program which loads the POB with orders to establish a connection between the calling customer and a Coin Detection and Announcement circuit.

SSAS Processing of Coin Notification. When the PROCON receives an input message requesting an end-of-initial-period notification on a coin call, the data accompanying this command are the CDA number and the time of the initial period. In response to this command, the PROCON directs the SSAS announcement control circuitry to give the following announcement:

"Three minutes has ended, please signal when through."

(A three-minute initial period is used for this illustration.) No coin scanning is performed, as no deposits are expected. After the announcement, the CDA hardware and software are idled and an announcement complete reply code (Fig. 3, format B) is sent to the TSPS processor.

1.2.1.3 Charges due seizures. The ACTS procedures for fully automating overtime charges due seizures on coin-paid calls are presented in this section.

Conditions for Automation. The SSAS can fully automate overtime charges due seizures on coin-paid calls on an ACTS-converted trunk group even if the initial contact requires an operator. Since the call is

already rated, no rating or calling number identification (ANI) restrictions apply to overtime seizures. Furthermore, charges due seizures for postpay calls are handled in the same manner as calls from other coin stations. However, there is one additional restriction on charge due seizures. All previous deposit request seizures (either initial contact or charges due) must have been successfully monitored by a CDA circuit. Successful monitoring implies that CDA circuit blocking did not occur and that the CDA circuit did not malfunction. This condition ensures that overdeposit credits are not lost. (If a seizure is not monitored successfully and an overdeposit occurs, the credit is not recorded.) If successful monitoring does not occur, the call is processed by an operator using current practices. In summary, the conditions for automating overtime charge due seizures are:

(*i*) The calling station is on an ACTS-converted trunk group.
(*ii*) The charges do not exceed the large charge threshold* (see Section 1.2.1.1).
(*iii*) All previous seizures were successfully monitored.
(*iv*) CDA must be available.

If a call continues for a certain number of overtime intervals (usually 10) and all the conditions for automation are satisfied, the SSAS automates the changes in seizure. Additional intermediate deposits are requested at successive specified overtime intervals until the call terminates. Then the SSAS automates the final charges due seizure at the end of the call. The collection sequences for end-of-call deposits and intermediate deposits are essentially the same. Furthermore, these collection sequences are very similar to those used for the initial contact on station (1+) calls (Section 1.2.1).

With only a momentary interruption to the conversation path, TSPS connects an idle CDA circuit to the call. TSPS informs the SSAS of the amount due, the number of minutes which have elapsed, the CDA circuit being used, and that it is a charge due seizure.

SSAS Processing of Overtime Deposits. SSAS deals with the two overtime situations (end of call and intermediate overtime deposits) in a similar fashion. After receiving the command from the TSPS processor, the PROCON processes the call in a fashion similar to the initial deposit case and the reports returned to TSPS are also similar. However, the TSPS sends the amount of credit from previous deposits (if any) and the time sent is the elapsed overtime talked (in minutes), not the initial period. The announcement wording is also different. Assume that the call has progressed for five 25-cent overtime periods of two

* The office data provided for the initial-period large-charge threshold is also used to specify the overtime large-charge threshold.

minutes duration each and that the customer had no credit. The announcement would be as follows:

> (Alerting tone) "Please deposit one dollar and twenty-five cents." (0.5-second pause) "One dollar and twenty-five cents for the past ten minutes."

If the customer had a ten-cent credit, for example, the announcement is altered.

> (Alerting tone) "One dollar and fifteen cents please." (2-second pause) "You have ten cents credit. Please deposit one dollar and fifteen cents more for the past ten minutes."

In the intermediate overtime deposit case, both customers are on the call and intend to continue talking. They both can hear the announcement and talk during the deposit interval. In the end-of-call overtime case, only the calling customer is on the call.

As with the initial deposit situation, if a customer stops depositing coins, a prompting announcement is given. Failure to respond to this announcement results in a report to TSPS requesting an operator.

Operator Assistance. If the call times out or the customer flashes the coin station switchhook during the charges due phase of a call, an operator is connected to the call. The following keys and lamps on the operator's console are lit steadily:

(*i*) The loop access key (ACS).
(*ii*) The appropriate supervision lamps (CLG, CLD).
(*iii*) The appropriate coin lamp (STA, 0+, or Dial 0).
(*iv*) The appropriate AMA key (station paid or person paid).
(*v*) The charges due (CHG-DUE) lamp.

These are the same keys and lamps lit by TSPS on charges due seizures prior to ACTS. In addition to these keys and lamps, the ACTS underdeposit display, previously described for initial contact, is lit on the operator's console. The appearance of the ACTS underdeposit display indicates two things to the operator. First, a CDA circuit is attached to the call to monitor the deposits. Second, if the customer overdeposited on a previous seizure, the charges displayed have been corrected by the amount of credit. (Note that, on non-ACTS charge due seizures, the operator must subtract any credit claimed by the customer from the charges displayed.)

The operator assists the customer until a full deposit is received. When the SSAS recognizes that the deposit request is satisfied, the

correct deposit display is given. The operator thanks the customer and depresses position release. Upon releasing the position, TSPS requests the amount detected by the CDA circuit. After TSPS receives this information from the SSAS, the CDA circuit is disconnected.

If the customer overdeposits during an overtime collection with an operator attached, the overdeposit display is lit on the operator's console. If this is a charges due seizure for intermediate collections, the operator should inform the customer that credit will be given on subsequent overtime charges. If the overdeposit occurs at the end of the call, the operator should handle the disposition of the overdeposit in accordance with current operator practices.

If the calling customer hangs up during the ACTS charge due announcement sequence, or the calling customer leaves the phone off-hook, the operator should try to reach the calling customer, if necessary, by ringing back against off-hook. The operator may need to wait for the customer to return. If the operator is unable to obtain full payment for the call, the operator keys a walkaway trouble number. A traffic counter is pegged, and a special bit and the amount of shortage is recorded on the AMA tape.

Even after the decision is made to go to an operator, the CDA circuit still monitors coin deposits. If the deposit is satisfied before the Peripheral Order Buffer (POB) which connects the position is activated, the position seizure is aborted. The SSAS acknowledges the deposit, and the call is released.

If the position is seized and then the deposit is satisfied, the SSAS does not acknowledge the deposit. When the operator's numeric display is lit, it shows (or may quickly change to show) that the deposit is satisfied. The operator should acknowledge the deposit and release the call.

1.2.2 Non-coin features

The announcement without coin detection mode of operation is used to provide announcements for time and charges quotations and notification on calls other than coin sent paid. A different command is used in each case to select a different announcement.

1.2.3 SSAS as a coin detector

The PROCON is also programmed to provide a mode of operation in which it gives no announcements but does monitor for coin deposits. This is used on operator-handled ACTS calls. In this mode, the PROCON informs the TSPS processor when the required deposit or an overdeposit has been made. If the operator collects or returns coins while a call is in this state, TSPS informs the SSAS so the PROCON can appropriately update the amount deposited and the amount due in its data RAM. If

the operator requests that the amount deposited be displayed, the SSAS will transmit that amount to the TSPS processor in response to the appropriate command.

1.2.4 Coin station maintenance and administrative features

One unique aspect of ACTS operation is that inband signals (the coin deposit tones) are transmitted directly from the station through to the TSPS over the voice path. Two special features were developed that are related to coin tone signaling. First, a coin station test capability was developed which allows a craftsperson at the station to test that coin deposits for that station can be detected at the TSPS. Second, a precutover mode of operation was devised to help detect stations that have not been converted to be compatible with ACTS. These features are described below.

Coin station test call. After the craftsperson has performed the usual coin station tests, the TSPS is accessed by dialing a special test code. When TSPS receives this number, it connects the call to an idle CDA and sends a message to the SSAS with a command code indicating that this is a station test call. The remainder of the call is handled by the SSAS except that TSPS supervises the call. If the craftsperson goes on-hook, the TSPS aborts the call by sending a command to SSAS. Also, TSPS will return coins on request from the SSAS.

When the PROCON receives the station test call command, it initializes the CDA, begins coin scanning, and initiates the following announcement:

"Coin Test." (1-second pause) "Please deposit nickel."

If the craftsperson deposits a nickel, the announcement,

"Nickel"

is given, acknowledging the deposit. A dime and quarter deposit are also requested and acknowledged. Then, the craftsperson may make additional deposits which will be acknowledged if received correctly. If no coin is deposited for about six seconds, the deposit request is repeated. If the wrong coin is deposited, a coin return is requested by the PROCON and the test is repeated. If repeated requests for deposits are not satisfied or a total of two minutes elapses on the call, the PROCON issues instructions for the announcement.

"Test has ended."

and reports the time-out to the TSPS, which gives a coin return and disconnects the call.

1.3 PROCON program design

The PROCON has relatively high processing capacity. It has a relatively powerful instruction set which is largely at the assembly language level rather than the microprogramming level, and it manipulates a 16-bit word. It uses a paged program memory and can execute instructions within a page in arbitrary, fixed order. (The next interpage address (displacement) is specified in each 24-bit instruction.) Because of this characteristic, care must be exercised in relocating code to avoid improperly crossing a page boundary. (Such errors are detected during compilation.)

The PROCON's 16K program memory limit requires that the programs be very compact. While an attempt was made to use modern, structured, design and documentation techniques, the limited memory size caused compromises in some areas. (The SSAS application requires a far larger program than most PROCON applications.)

Multiprogramming was achieved by storing all call- (process) related information in the data RAM. Each program retrieves all data from that memory when it begins processing a new call. Each program is a pure process stored in ROM (i.e., the program contains no variable data and cannot alter itself).

The PROCON has very extensive self-checking features. In addition to a variety of parity checks (including next address parity), the arithmetic and logic circuits are duplicated and checked by matching.

1.3.1 PROCON program administration and design standards

Several designers developed separate programs which were integrated into one large program. Loading was accomplished by using a linkage editor and a common file of symbols and macros. All symbols (except labels) and all macros were defined in the common file. No numerical references were permitted, and input/output control designation symbols were keyed to the circuit drawing names for the same leads. Thorough prologue comments were required for each routine. Standard labels and standard symbolic designations and usage for all PROCON internal registers were agreed upon by all programmers.

1.3.2 PROCON active monitor

The basic program loop for the PROCON in the active SSAS (i.e., the SSAS which is processing calls) is called the active monitor. The PROCON executes this program while looking for tasks to perform. In this loop, the PROCON first makes some maintenance checks and then reads a clock to determine if a new 512-ms period has begun. The 512-ms interval is called a base period. Its length is keyed to the length of an announcement speech segment. If a new base period has begun, a

special flag is marked in the system status register.* The PROCON then scans for a message from the TSPS processor. If such a message has arrived, it records the data in the appropriate RAM area.

The PROCON then begins a loop in which each CDA is processed. Each CDA has a dedicated eight-word call register associated with it. (Figure 8 is a typical CDA register layout.) First, the PROCON reads the address of the program which should process the call. This address, which also defines the state of the call, is called a Progress Mark (PM). If the PM is zero, the CDA is idle or unequipped, so no action is taken. A typical PM program is the program which scans a CDA for coins. The PM is stored as a page number and a displacement within the page. The active monitor transfers to the PM routine. If the PM routine determines that the call requires timing, it checks the new base period flag. If that flag is set, the required timing counter is updated.

When the PM program returns to the active monitor, the monitor checks the new base period flag. If that flag is set, all timing counters for the CDA are updated as required and a new announcement segment is selected. (The announcement processing is described in Section 1.3.4.) After the announcement program is completed, the monitor processes the next CDA. When all CDAs have been processed, the monitor clears the new base period bit, checks for various maintenance tasks, and then begins the main loop again.

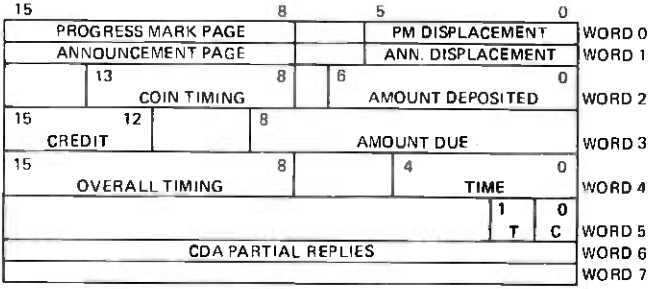
1.3.3 Announcement store layout

Each announcement speech segment consists of 16,000 bits of digital data stored in 400 consecutive announcement store locations. Each location contains 40 bits of data. Decoded at a 31,250 bits/second rate, this corresponds to 512 ms of speech. For programming ease and circuit design convenience, all segments must begin at an address which is a multiple of 16. Most words used fit in one segment (for example, the numbers "one" to "ten"). However, several words or phrases require two segments (for example, the word "eleven" or the phrase "please deposit"). A few phrases require three segments (for example, the phrase "please signal when through").

Because the announcement circuitry bit-synchronizes segments, two or more segments can be blended together with no loss in quality. Therefore, words can be split between 512-ms segments to achieve the best cadence of speech.

Successive segments constituting 1-second and 1.5-second phrases need not occupy consecutive announcement store addresses. However, these segments generally must be used together. In contrast, suffix segments used for compound numbers may be combined with a variety

* One of the PROCON general registers was reserved for critical status indicators.



NOTES:

1. PM = PROGRESS MARK
2. AMOUNT DEPOSITED, AMOUNT DUE AND CREDIT IN NUMBER OF EQUIVALENT NICKELS
3. C–TYPE OF COA
4. T–TRUNK TYPE (2 WIRE = 0,4 WIRE = 1)

Fig. 8—CDA data RAM call register.

of other segments. For example, the suffix "-teen" can be combined with the syllables "thir-," "four-," etc. to form the numbers "thirteen," "fourteen," etc. with minimal use of memory. Similarly, the units place suffixes "-one," "-two," etc. can be combined with the numbers "twenty," "thirty," etc. to form the full required set of numbers (e.g., twenty-one, twenty-two, thirty-one, thirty-two).

Several words are duplicated in the announcement store so that both neutral and falling inflections are available. The falling inflections are used only at the ends of sentences, while the neutral inflections are used elsewhere.

In total, about 80 speech segments 512 ms in length are required for ACTS call handling. About 15 more are required for maintenance purposes.

1.3.4 Announcement table structures

Two major announcement tables are stored in the PROCON ROM. The first is simply a table of addresses. Reading this table using the speech segment number as an index provides the address of the first announcement store location containing the data for that speech segment.

The second table is more complex, as it controls the sequencing of announcement segments to form sentences. The logic for this process is embedded largely in this table rather than the program code. This makes debugging easier and also simplifies the process of adding new announcements. The table consists of program addresses for generalized routines dealing with each step of the announcement. The address

of the current table entry is stored in the second word of the CDA register (see Figure 8). Each announcement routine must decide what speech segment is to be reproduced next and also update the table address in the CDA register. The major situations the various routines must deal with are demonstrated by the following example.

The structure of a portion of a typical announcement is represented schematically in Fig. 9. This is the end of the time-and-charges announcement used for noncoin calls. A typical example of this announcement follows.

> "The charges are one dollar and twenty cents plus tax for seven minutes."

The three major types of announcement routines are each involved in this announcement. One type of routine simply sequences speech segments in fixed order (for example, the words "plus tax for"). The second type of routine chooses a word or sequence of words ("one minute" versus "*M* minutes") based on the value of a parameter (whether *M* is one or greater than one). The third type of routine translates a parameter (*M*) into a speech segment (for example, the number "seven") or a series of segments (for example, "twenty-one"). This last program is a subroutine since the action to assemble the number is the same whether it represents a dollar amount, a cents amount, or a time period.

II. MAINTENANCE

2.1 Overview

2.1.1 TSPS maintenance requirements

In TSPS, as in other Bell System electronic switching systems, the maintenance strategy is based on duplicating vital hardware units[3] and providing hardware checking circuits for other units, using signals to indicate the successful execution of orders and using programs to test the state of the hardware, to detect faults, and to diagnose trouble.

TSPS provides fault recognition and diagnostic programs for all major circuit elements. The purpose of the fault recognition programs is to quickly detect faulty equipment units and, if necessary, remove them from service. Hardware checking circuits are used for trouble detection

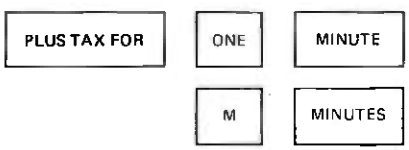


Fig. 9—Typical portion of an announcement phrase.

during operation. When a trouble detection circuit identifies a serious problem in the system, it notifies an interrupt circuit. The interrupt circuit immediately stops operational program processing and transfers control to a fault recognition program associated with the particular trouble indication. The functions of the fault recognition programs are to distinguish nonrepeating troubles (errors) from repeating troubles (faults), and, in the case of faults, to quickly establish an operational system configuration. This is generally done by switching out the faulty unit. When the fault recognition process is completed, operational program processing is resumed at the point of interruption.

Fault recognition is the highest priority function in a real-time system like TSPS. It is of the utmost importance to high system reliability that minimum time be taken away from the operational processing functions because of faulty circuits.

Once the faulty circuit has been taken out of service by the fault recognition program and the system brought back to an operational configuration, diagnostic programs are called in to test the out-of-service unit to isolate the trouble to a small number of circuit packs. A diagnostic program is normally broken into logic testing entities called subphases or phases so that a series of rigorous tests can be performed on the out-of-service circuit unit. Depending on the outcome (pass-fail) of each test, a trouble-locating number is printed on the maintenance teletypewriter. This trouble-locating number is used by the maintenance personnel to reference a trouble-locating manual to isolate the particular fault involved.

2.1.2 SSAS maintenance requirements and anticipated reliability

SSAS reliability and maintainability objectives are based on the fact that a significant part of TSPS traffic is served by the SSAS. The objectives are

(*i*) An average of about one SSAS outage in 10 years.
(*ii*) An average repair time of about 2 hours.

The maintenance plan for SSAS impacts on the reliability objective in three major areas. First, it is necessary to provide fault detection mechanisms (primarily hardware checks) so all failures can be detected rapidly. Second, provisions must be made to minimize the interruption of call processing once a failure is detected, with few or no calls lost. Third, sufficient circuitry must be provided to obtain diagnostic resolution to hold repair time to the 2-hour average as specified above.

The SSAS maintenance strategy relies on a multilevel fault detection scheme in which faults are classified according to their impact on the SSAS operation. A fault that does not affect normal operation of the

SSAS is considered "tolerable." An SSAS side with this type of fault may continue to perform its normal function adequately as long as it is required to do so. As an example, all single bit failures in the SSAS semiconductor announcement memory are tolerated, since they would have very little effect upon the quality of the speech. By using a maintenance strategy that permits an SSAS side to be used with certain "tolerable" faults, the SSAS reliability is further enhanced.

To verify that the SSAS design meets the reliability objectives, a reliability estimate was made by classifying and counting the boards on the SSAS side. The failure rate for each board type was calculated. The failure rate determines the Mean Time Between Failure (MTBF) of an unduplicated SSAS side. The Estimated Downtime (EDT) for the duplicated system was then calculated with the assumption that the mean time to repair for one side down and both sides down are identical. This is a valid assumption because the diagnosis of all critical circuits in SSAS can be done on each side independently. Plugging necessary parameters into the appropriate reliability model showed that the above reliability objectives are well satisfied.

2.2 Announcement Store Maintenance

2.2.1 SPC bus and store switch

The Stored Program Control No. 1A (SPC 1A) is the TSPS main processor.[4] The SPC memory is used for both programs and data. It is partitioned into a maximum of 24 duplicated stores. Each store is permanently assigned to one of the two SPC store buses. An SPC store can be equipped with a piggyback twistor (PBT) or semiconductor memory. Two sequential or complimentary name codes are assigned for an SPC store when equipped with semiconductor memory and one name code when equipped with PBT. The name code serves to identify the stores for addressing purposes. A semiconductor store consists of a semiconductor memory module and its associated semiconductor memory controller, shared by all modules in the same memory frame. The SPC store is the maintainable unit of SPC memory; i.e., it can be removed from or restored to service, diagnosed, and updated with the contents of its duplicate.

The SSAS announcement memory is used to store digitized announcements. An Announcement Store Frame (ASF) is permanently assigned to each side of each SSAS. The SSAS ASF consists of an announcement memory controller and up to six announcement memory modules. To simplify development effort and to take advantage of the memory loading facilities and diagnostic software of the SPC, the SPC semiconductor store frame with a second access port is used as the announcement store frame. This is accomplished by placing a switch on the announcement memory which allows the memory to be connected to

either the SSAS side or the SPC store bus. Furthermore, to reserve the maximum number of memory name codes for the SPC store, all SPC memory access operations use odd parity name codes while all announcement memory access operations use even parity name codes.

Under the normal mode of operation, the switch is set to the SSAS side so that access to the announcement memory can be performed. In the SPC direct access mode, the switch is set to the SPC side so that loading from the Program Tape Unit (PTU) and diagnosing by the SPC can be accomplished. More discussion on loading and diagnostic operations is presented in a later section.

2.2.2 The choice of announcement memory

To handle all TSPS station-paid coin call announcements as well as time and charge quotations for noncoin calls, it was calculated in 1974 that announcement capacity equivalent to about 80 half-second words would be needed. An ultimate memory capacity of 200 to 400 half-second words was forecast at that time if vocabulary for future features as well as space for other TSPS announcements were considered.

A number of storage media were examined for the announcement memory. After some consideration, it was decided that with minor modifications the semiconductor memory used for the SPC 1A store would also be suitable for the SSAS announcement store. This choice of the announcement memory would provide flexibility in areas such as future vocabulary growth and vocabulary reload in the field from a magnetic tape. It was also estimated that most hardware and maintenance software designed for the SPC 1A store could also be applied to the announcement store with only a modest amount of effort.

2.2.3 Diagnostic implementation

The SPC store diagnostic programs were modified to be used for the SSAS announcement store diagnostic. Major objectives for the announcement store diagnostic implementation were specified as follows:

(*i*) Circuits common to both designs to be tested by existing SPC semiconductor memory diagnostic programs.
(*ii*) New announcement store circuits to be tested by the SSAS controller diagnostic.
(*iii*) Same Trouble Locating Manual (TLM) to be used for SPC and SSAS announcement stores.
(*iv*) Similar TTY input/output messages.

Several diagnostic characteristics are common for the SPC and announcement store diagnostic programs: the basic diagnostic unit is a controller and one memory module, seven diagnostic phases are in

the program, the diagnostic terminates on early detection of errors, there is a special bus configuration for increased fault detection, and finally, there are automatic exercise tests for all 32K words. For these common areas, existing SPC store diagnostic programs were applied to the announcement store with only very modest modifications.

On the other hand, some diagnostic characteristics pertaining only to the SSAS announcement store must be added to the existing store diagnostic programs. First, the announcement store has to be switched onto the standby SPC store bus before it can be diagnosed. Second, since during normal operations, the PROCON and ASD circuits in the SSAS controller control the operation of the announcement store, they must be stopped during SPC diagnostic tests. Third, the SSAS controller and memory modules use even parity names while the SPC store uses odd parity names; consequently, diagnostic programs must be capable of distinguishing these two cases and proceeding accordingly.

2.2.4 Loading store from program tape unit

To save development effort, the SPC program used to load from the Program Tape Unit (PTU),[4] compare with PTU tape, and dump the memory contents of the SPC stores to a PTU tape is used to perform the same functions for the SSAS announcement stores. However, since there are some significant differences regarding tape format, loading procedures, and priority levels between the SPC stores and the announcement stores, some changes were made in the PTU program to extend its capabilities.

The requirements for the handling of the tape unit are as follows:

(*i*) The announcement information should be separate from office data and generic programs. A separate tape containing only announcement data should be used for announcement stores.

(*ii*) For the load operation, the craftsperson must force one SPC bus active, simplex, before using the program to access the announcement stores on the other bus.

(*iii*) The announcement tape header should be modified to distinguish it from an SPC tape.

(*iv*) SPC reliability should be given the highest priority. The announcement store should not be accessed if the SPC is not able to run full duplex with all stores operational.

(*v*) Changes must be made to minimize SPC downtime resulting from undetected SSAS store faults.

With these requirements in mind, the craftsperson is instructed not to use the program to access the announcement stores if the SPC cannot run full duplex with all stores operational. With the SPC fully operational, one bus may safely be forced active for the load operation,

thus placing the system in the simplex mode. This choice of which bus to force active must be made such that the standby SPC bus will correspond to the SSAS side requiring PTU actions. This procedure ensures that the announcement stores will never be placed on the active SPC bus. Following this line of conservatism, the announcement stores are normally switched off the standby SPC bus and are connected just prior to a read or write instruction. After this instruction is completed, the stores are removed from the bus.

The SSAS side to be accessed must first be placed in the out-of-service or unavailable state for the compare or dump option. If the operation is to load an SSAS announcement store with data from tape, the SSAS side must be in the unavailable state. This state will not allow SSAS side switching if a fault should be detected on the active side.

2.3 SSAS fault recovery

2.3.1 PROCON detected and reported faults

SSAS fault recovery strategy is based on a fault detection scheme which uses both the SPC and the programmable controller (PROCON) to detect various faults associated with the SSAS hardware. Faults detected by PROCON are reported to SPC. The decision to choose one method or the other was based on the following considerations.

(*i*) SSAS faults, particularly those with serious impacts on call processing, should be detected as quickly as possible.
(*ii*) SPC routine processing time required for detecting SSAS faults should be kept as small as possible.
(*iii*) PROCON should not be responsible for detecting those faults which could affect its integrity or its ability to communicate with the SPC.

Those faults which are detected by PROCON are analyzed before being reported to SPC. PROCON has the capability to analyze available data and reinterrogate appropriate maintenance registers. If failures are indicated, the result of analysis and reinterrogation are sent to the SPC. The reports sent by PROCON are analyzed by the SPC and, if necessary, appropriate commands are sent to PROCON to prevent it from flooding the SPC with unnecessary information.

2.3.1.1 Continuous exercises. Continuous exercise of hardware by PROCON is a means of verifying the correct operation of the hardware and detecting various failures. Exercises are used both on active and standby SSAS sides, but these are particularly important on the standby side since no call processing is taking place. (Call processing exercises the SSAS rather completely.)

The announcement subsystem[5] (i.e., the announcement store and

announcement distributor combined) is among those units which is continuously exercised by PROCON on both the active and standby sides to ensure the integrity of the speech heard by the customers. Any malfunction in these units would result in the appropriate flags being set in a maintenance register called the Error Summary Register (ESR). This register is periodically scanned by PROCON, which analyzes the content and sends the appropriate information to SPC.

2.3.1.2 Announcement subsystem failure reports. PROCON can recognize failures associated with the announcement subsystem by scanning the Error Summary Register (ESR) periodically. The presence of any announcement subsystem failure will result in one or more flags being set in the ESR. If the content of ESR is different from its last look, PROCON sends a message to SPC indicating what trouble has occurred, provided PROCON has received no commands from SPC to stop sending information.

During normal operation, a Hamming and parity check on the address and data received from the announcement stores is performed by hardware[5] before the data are sent on to the service circuits. Once an error is detected, an error flag will be set in the ESR. The 67 bits of address, data, Hamming and parity for the error will be trapped in special maintenance address and data registers.

The content of the ESR is scanned periodically by PROCON. Upon identifying an announcement store error flag in the ESR, PROCON saves the contents of trap registers. Then it performs a read using the saved announcement store failure address. Subsequently, it loads the FIFO output register with the available information. This information includes (see Fig. 10) the content of ESR, the content of failure trap address and data registers (both original and retry values), and an extended reply code used to distinguish the cases where no error is detected upon retry from those where error is also detected upon retry.

Upon receiving a PROCON report, SPC analyzes the reported errors and decides whether a hard fault or a transient error is present. Detected errors and faults are categorized according to their types and seriousness. Types of errors and faults detected by SPC are repeating double bit data fault, repeating single bit data fault, transient double bit data error, transient single bit data error, and repeating single bit address fault. The address and data associated with the errors and single bit data faults are also saved, and are printed hourly or upon manual request on the maintenance teletypewriter for manual troubleshooting. The failing address is also passed to the store diagnostic program to allow extensive tests of the suspected area of the memory.

A feature that will aid in detecting announcement store errors is the announcement store recital program, which sequentially accesses every announcement phrase by loading its address into the maintenance

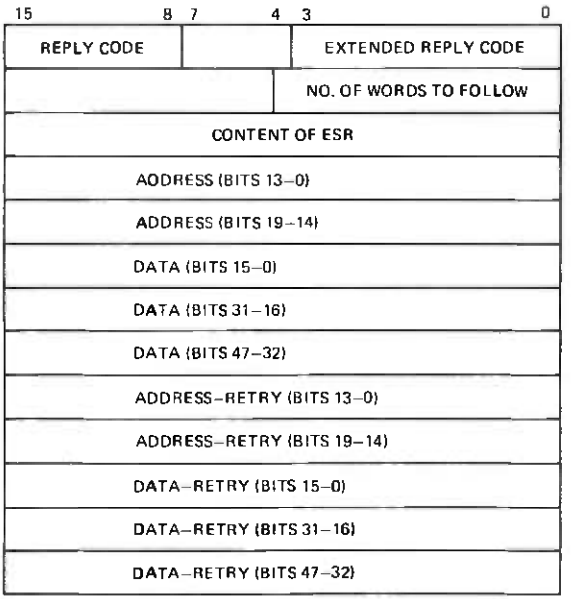


Fig. 10—Format of announcement subsystem failure report.

time slot of a Recirculating Shift Register (RSR). This type of routine exercise will guarantee that the memory used by every speech segment is accessed periodically. If any store error associated with a particular speech segment exists, a flag will be set in ESR notifying PROCON of the error and the address and data associated with the errors will be trapped. This program is implemented on both active and standby SSAS sides.

2.3.1.3 Mate frame buffer failure reports. Communication between the two SSAS sides (see Fig. 11) takes place asynchronously through a first-in, first-out serial memory called the Mate Frame Buffer (MFB). Each transmission through the MFB usually consists of data and an ID field which identifies the data.

During normal operation, all the data associated with the calls in progress are sent from the active side to the standby side through the MFB. This operation enables the standby side to have an up-to-date copy in its Random Access Memory[5] (RAM) of the data associated with each call for use if it is required to change role from standby to active. To test the validity of the data before it is stored, a 1-out-of-16 encoded ID word is associated with the data coming through MFB, to indicate what it contains. The standby PROCON checks for a valid ID format and for proper sequences of IDs. If the standby PROCON detects an error, it informs the SPC by loading appropriate data in its FIFO output register.

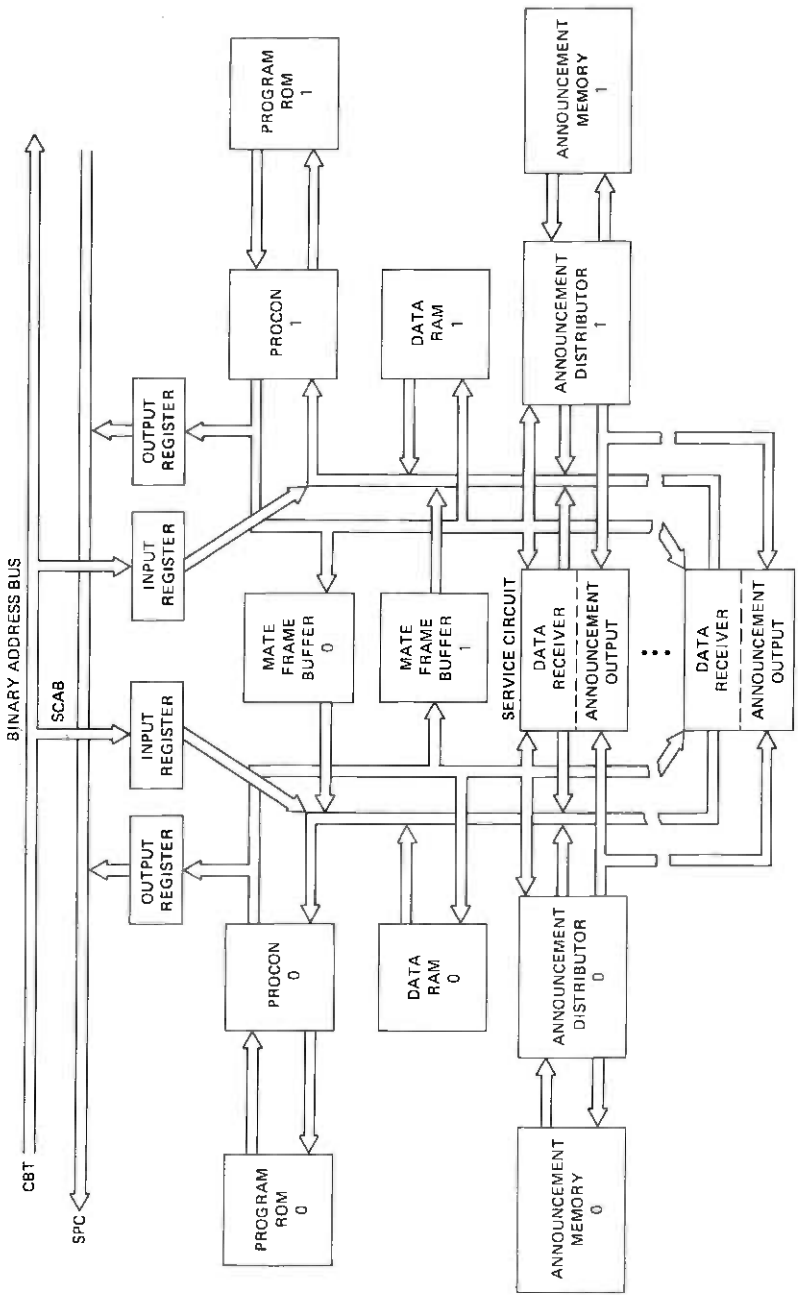


Fig. 11—Duplicated SSAS configuration.

During the unloading of the MFB by the standby side, an overflow condition in the MFB results in an MFB full flag being set in the ESR associated with the active side. This overflow condition could be caused by either an MFB failure or a failure in the standby side which prevents it from unloading the MFB. The MFB full flag in the ESR is monitored by the active PROCON. Upon detecting the MFB full condition, the active PROCON notifies SPC by loading the necessary information in the FIFO output register.

2.3.1.4 CDA error reports. A CDA failure is recognized by PROCON during normal interrogation. All calls except those that are in the announcement mode only require interrogating the CDA reply register for coin deposits. Prior to sending the first interrogation command to a CDA, PROCON reads the reply register to ensure that there are no premature replies from the CDA. In the case of premature reply, all records of the call are erased and a message is sent to SPC indicating a bad CDA. During the course of interrogation, the parity of the data received from the CDA is checked by PROCON. In the case of a parity failure, PROCON again erases the records of the call and informs the SPC of the failure.

During the course of interrogation, PROCON continuously monitors the format and the validity of the data in the CDA reply register. A reply with invalid format is considered a CDA failure. In this case, the records of the call are also erased, and SPC is notified of the CDA failure. Another means of CDA error detection is the CDA maintenance register. If a maintenance bit is set in the reply register, the maintenance register is read to check the type of error. In this case again, PROCON notifies SPC and erases the records of the call. Upon receiving a CDA error report from PROCON, SPC routes the call to an operator position and also takes the bad CDA out of service, so it is not used for any future call. A diagnostic automatically will be requested for the bad CDA to aid the craftsperson in locating the faulty units. If too many CDA failure reports are received within a certain period of time, all associated with the same group controller, SPC then assumes that the group controller associated with the active side is faulty.

PROCON can only detect failures in the digital circuitry of a CDA. Failures in the analog circuitry of a CDA are detected by an error analysis program which resides in SPC. The success rate of attempted calls for a given CDA is analyzed, and if it is below a certain average threshold, the CDA is considered faulty and is taken out of service and automatically diagnosed.

2.3.2 SPC detected and reported faults

2.3.2.1 Maintenance interrupts. Maintenance interrupts are used as a means of identifying faults in SPC-SSAS interface registers. These

types of faults have a serious impact on the operation of SSAS. They can result in isolation of the SSAS, thus crippling the capability of SPC to determine the basic integrity of the SSAS. This basic integrity is a necessary condition for the SPC to have confidence in the capabilities associated with PROCON in detecting SSAS faults.

As with other TSPS peripherals, any malfunction in SPC-SSAS communication is detected by using standard peripheral unit fault recognition techniques that can cause an F-level interrupt (such as central pulse distributor enable verify, and all-seems-well signal failures).[4] Upon occurrence of an F-level interrupt, appropriate fault recognition programs are entered to analyze the faults present in SPC-SSAS interface areas. These programs have the capability of distinguishing serious faults from nonserious faults and reporting them to appropriate reconfiguration programs.

During a scan operation of SSAS by SPC, a malfunction in one or both Scan Answer Buses (SCAB) results in an SPC processor mismatch which in turn causes a C-level interrupt to occur.[4] Upon occurrence of a C-level interrupt, appropriate fault recognition programs are entered which first determine whether SSAS caused the interrupt, and then by performing some tests isolate the faulty SCAB. The nature of failure and its degree of seriousness are also passed to SSAS reconfiguration programs.

2.3.2.2 Periodic scans. SPC detects faults in PROCON itself or in the PROCON peripheral control system by interrogating maintenance registers on the scan answer bus. This interrogation is done periodically (every 25 ms). The information provided by this interrogation (scan) is used by the SPC to determine the sanity of the PROCON for processing calls and to verify the capability of PROCON to detect SSAS faults. Among faults detected by periodic scanning are PROCON all-seems-well failures, Random Access Memory (RAM) all-seems-well failures, PROCON clock-stopped and output register-full indications. Both active and standby SSAS sides are monitored by the periodic scanning. The scan result is analyzed and is used to distinguish serious faults from nonserious faults, and to determine whether or not an immediate action is needed by the SPC.

2.3.2.3 Dead side tests. Despite the fact that extensive error checking circuitry has been built into the SSAS hardware, the possibility still exists that some faults may elude the checks and not be reported to the SPC. Potential faults in this category are nonclassical faults, multiple faults, certain PROCON and RAM failures, etc. A possible consequence of such faults is a dead SSAS side which appears to be free of faults during a period of no normal SPC-SSAS communications. To detect such a failure mode, the SPC performs an exercise periodically. The exercise involves sending a command to an SSAS side and receiving a specific reply. If an SSAS side fails to reply as expected, then the side

is considered to be faulty, and appropriate information is reported to the reconfiguration program.

Since SSAS functional integrity is constantly checked by call processing functions, the dead side test for an active SSAS side will be skipped if traffic is present on the side. More specifically, an SSAS side will not be tested if valid replies are coming through its output register at a regular interval.

2.3.3 Reconfiguration

2.3.3.1 Classification of faults. The response of the SSAS reconfiguration program to various SSAS faults is based on a fault classification scheme which categorizes the faults according to their impact on the normal operation of SSAS. Faults fall into one of four categories which can be summarized as follows:

(*i*) Serious faults. This type of fault corresponds to those which seriously affect the call processing capability of SSAS. For example, a fault which stops normal operation of PROCON is considered to be serious. A serious fault on an active SSAS side will cause an immediate switch to the other side.

(*ii*) Nonserious faults. A nonserious fault is defined as a type which does not seriously affect the call-processing capability of an SSAS side. An active side with this type of fault can continue to perform its normal function adequately as long as it is required to do so. This in turn implies that the switching action to the mate side can be delayed indefinitely, if necessary. For example, a single bit failure on the announcement data can be considered nonserious since its impact on the quality of the announcement is insignificant. As another example, a group controller failure on the active side can also be considered nonserious, since the side can continue to process calls using the remainder of the group controllers and their associated CDAS.

(*iii*) MFB faults. This type of fault is usually an indication of a failure in the MFB itself or the standby SSAS side, with no impact on the active SSAS side. No switching of sides is performed in this case. Instead, the standby side and MFB are both diagnosed for isolation of faulty units.

(*iv*) CDA faults. This type of failure usually represents a problem in an individual CDA. Since there are many CDAS in SSAS, an individual CDA failure does not have a serious impact on the call-processing capability of SSAS. A faulty CDA is marked out of service and will not be used until it is diagnosed and repaired.

2.3.3.2 Smooth side switch. The process of interchanging the active side and standby side is referred to as a side switch. If this is

done so that the flow of announcements is not interrupted, the switch is called a smooth switch.

A smooth switch between two SSAS sides is performed under two circumstances. One case is when the active side has a nonserious fault such as single bit failure in announcement data. In this case, the system can afford to wait as long as needed to switch smoothly. The other situation is when a routine switch to the standby side is desired to test the hardware controlling the reconfiguration and to provide a chance to run a routine set of diagnostics on the previously active side. This process (the smooth switch exercise) is performed three times a day.

A smooth switch is typically characterized by the following actions. First, the SPC sends a command to the active side, telling it to go standby gracefully. Second, the active side brings the standby side in loose synch with the active. Third, the standby side is told to go active gracefully. Since by now the active and the standby are loosely in synch, both sides time themselves until the right moment when the old active goes standby followed almost immediately by the old standby going active. The new active verifies its success by sending a reply to the SPC. Upon receipt of this reply, the SPC changes the SSAS status words to record the new configuration. This type of side switch does not provide any discontinuity in the announcement heard by the customer and also has no effect on the call processing operation.

2.3.3.3 Other types of side switch. The handshaking between the two sides required to perform a smooth switch may not be possible if the active side develops a serious fault such as a PROCON all-seems-well failure. Assuming that the nonactive side is in the standby ready state, the SPC simply takes the faulty side out of service and makes the standby side active. This type of switch is referred to as an *immediate switch*. Since the standby's data RAM is up to date with the active side, only a minor disruption to call processing results. The announcement heard by a customer is interrupted. Approximately a half-second elapses, and then the entire announcement is repeated. In addition, it may result in mishandling a single call. However, this minor disruption of the call processing is considered a small penalty, as hardware faults occur rarely.

Another type of side switch, called a *rough switch*, happens when the active side is doing call processing and the nonactive side is running a routine diagnostic on itself. If the active side develops a serious fault, it is immediately taken out of service. In the meantime, the routine diagnostic on the nonactive side is aborted, and the side is made active. Since the data RAM on the standby side is not up-to-date while a diagnostic is being run, it will be initialized before becoming active. Therefore, all processing on existing calls is aborted and the SSAS starts

afresh. Considering the reliability objectives of the SSAS, the probability of occurrence of a rough switch is estimated to be very small.

2.3.3.4 Tolerance of software errors. Software errors are usually analogous to a serious component failure, causing both SSAS sides to go down and remain down until the problem has been rectified. However, certain classes of software errors could be considered tolerable. An erroneous software instruction which is executed infrequently, if it is considered to be nontolerable, could cause a permanent outage of the SSAS upon its first execution. On the other hand, if at this time the SSAS is reinitialized and restarted, it will continue to perform its normal function until some later time, when it will encounter its second failure, due to the execution of the erroneous software instruction. Following this strategy would substantially decrease the SSAS downtime in cases where infrequently executed software errors are present.

The above considerations provided the basis for developing a strategy which tolerates certain classes of errors in the PROCON software. Upon occurrence of an SSAS outage, the SPC performs an analysis for both sides using past history of SSAS outages. First the side with the fewer past failures is selected. Then after the analysis a decision is made on the nature of the problem for that side, as to whether a hardware fault or a software error exists. In the latter case, the side is reinitialized and used for call processing after a successful initialization, while a diagnostic is requested for its mate side. In the former case, both SSAS sides are diagnosed to pinpoint the hardware malfunctions.

2.3.3.5 Duplex failure. If SSAS becomes totally unavailable, the call processing function performed by the SSAS is aborted and all new calls will be routed to the operators. The new calls are blocked from attempting to cease CDAS for service. Also, several clean-up actions are performed by the SPC to allow for rapid restoration of ACTS. For example the peripheral orders waiting to be sent to the SSAS are searched and disposed of. The SSAS is blocked from providing ACTS until all the clean-up tasks are completed.

2.4 SSAS diagnostics

2.4.1 Controller diagnostic

2.4.1.1 Interface SPC—PROCON. The SSAS diagnostic programs run on one SSAS side at a time, always a nonactive side. The diagnostic request is made either by manual request or by fault recognition programs, either as a result of a fault being detected or as a routine diagnostic. The routine diagnostic is run once a day to do a complete check of the SSAS hardware.

As a direct consequence of the SSAS architecture, the SSAS diagnostic control is divided into two parts; the first part resides in SPC, and the second resides in PROCON's Read Only Memory (ROM) and is under

the control of a standby monitor program which controls the various activities of a nonactive SSAS side, including diagnostics. Other activities associated with the standby monitor include fault recognition, unloading of the Mate Frame Buffer (MFB) and other SPC-requested nondiagnostic functions. SPC retains the overall control over the SSAS diagnostic. SPC initiates SSAS diagnostics by first marking the SSAS side out of service and then sending a diagnostic request to the PROCON. The request indicates that a group of tests referred to as a "phase" is to be run by PROCON. In some cases, this request is also accompanied by a test pattern to be used by PROCON. The standby monitor decodes the request, performs error checking, and invokes the PROCON diagnostic control program. This program further decodes the diagnostic request to select the proper diagnostic phase. Once a diagnostic phase is initiated, the standby monitor exercises control over the execution of the diagnostic. A diagnostic phase performs tests on the SSAS functional block being diagnosed, using predetermined test patterns and sends the test results to SPC through the FIFO output register. The test results are used by the SPC to isolate the particular fault involved.

2.4.1.2 Levels of raw data compression. The diagnostic test results sent by PROCON through the FIFO output register are compared against the expected result by the SPC. The result of that comparison is saved as a binary string of ones and zeros referred to as "raw data" results. The zeros represent the tests passed and ones indicate the tests failed. Although PROCON does considerable data comparison, in many diagnostic phases the test results sent by PROCON are quite voluminous and exceed the capacity of the trouble number generation programs. In such cases, SPC uses a compression algorithm for converting the test results into a compressed raw data bit string. The algorithm uses a simple procedure: If any test result word sent by PROCON does not match the expected value, the SPC diagnostic program will map this mismatch into one raw data bit. The compressed raw data bit string is further reduced by the SPC to a 4-part, 16-digit number referred to as a "fault signature," which uniquely identifies the string. The fault signature is used by the maintenance personnel to reference a trouble-locations manual to isolate the particular fault involved.

In some cases, the fault signature generated by SSAS diagnostic programs fails to provide enough information to clear the trouble associated with the SSAS. In such cases, the maintenance personnel can manually request either compressed raw data or expanded raw data. The raw data printout is used by maintenance personnel as an aid in manual troubleshooting.

2.4.1.3 Power of PROCON test. The SSAS architecture has partitioned the hardware in functional modules (such as PROCON, peripheral

control, announcement memory, service circuit group controllers, etc.).[5] Each of these functional modules are further partitioned to attain the desired level of maintainability and diagnostic resolution.

PROCON access to each functional module results in a simple diagnostic structure consisting of segments that diagnose only one functional module. In general, diagnostic resolution is directly related to the size of the module's diagnostic. Resolution can be improved by reducing the size of the module. A further improvement can be realized by choosing a diagnostic structure that begins with the diagnosis of the most basic functional module and proceeds to test each module by adding diagnostic segments that use the diagnosed modules. These considerations have been taken into account in planning the SSAS diagnostic structure, which provides an average diagnostic resolution of three or fewer circuit packs, with the capability of detecting greater than 90 percent of potential SSAS faults. These diagnostic objectives are required to limit the average repair time to about 2 hours as specified by the SSAS maintainability objectives. PROCON access capability combined with the modular architecture allows the SSAS diagnostic objectives to be achieved.

Another feature that helps to achieve the SSAS diagnostic resolution objective is the bit-sliced architecture used in designing SSAS input and output bus interfaces. The bit-sliced architecture is used to aid the SSAS diagnostics in pinpointing faulty packs and to reduce the number of board codes. This means that the common bits for all registers on a bus are grouped together on one or a few circuit packs (i.e., bit i for all registers on the bus are grouped together). Because of this architecture, it is only necessary for the SSAS diagnostic to know which bit is bad to isolate the faulty circuit boards.

2.4.2 CDA diagnostic

2.4.2.1 SPC-PROCON interaction. The CDA diagnostic is initiated by the SPC either as a result of CDA failures detected by fault recognition programs, error analysis programs, or a manual diagnostic request via the maintenance teletypewriter channel. Also, an automatic progression test of the CDAs is performed to ensure that each CDA is diagnosed once a day.

A service circuit known as the CDA test circuit is used to diagnose CDAs. It has network appearances so it can be connected to the various CDA ports for analog circuit tests.[5] When a CDA diagnostic is requested, the SPC attempts to establish a test configuration if the requested CDA is available for testing. It first verifies that the CDA test circuit is available and the nonactive SSAS side is in the standby ready state. If not, the diagnostic is blocked; otherwise, the CDA test circuit is made maintenance busy, and orders are sent to the active SSAS side to switch

both the CDA test circuit and the CDA under test over to the standby SSAS side. Next, the Signal Distributor (SD) controlled relays in both the CDA and CDA test circuits are released without regard to their previous states, and the two circuits are connected via the TSPS network.

Once the test configuration is established, various phases of the CDA diagnostic are initiated by the SPC. For each phase, the SPC first operates the required SD-controlled relays, and then sends an order to the standby SSAS side to run tests of the specified phase.

The PROCON standby monitor program processes the diagnostic order and transfers PROCON control to the CDA diagnostic program. This program performs the tests of the specified phase and sends the results to the SPC via the FIFO output register. Once the PROCON reply is received, the SPC compares the test results with the expected results. The result of this comparison is the diagnostic raw data. Each bit set to a one in the raw data represents a test failure. The raw data is then converted into a trouble number which is used by the maintenance personnel to isolate the trouble.

When all the phases of the CDA diagnostic have been completed, the SPC releases the relays of the CDA and CDA test circuits, breaks the network connection between them, and sends orders to the standby SSAS side to switch them back to the active side. The CDA test circuit is returned to the maintenance idle state, and the CDA under test is either returned to service or placed out of service, depending upon whether the diagnostic tests have been passed or failed.

2.4.2.2 CDA test circuit use. The CDA test circuit is a special type of SSAS-controlled service circuit. There is one CDA test circuit per SSAS, and it is always equipped as circuit 15 of the service circuit frame 1.

The primary functions of the CDA test circuit can be summarized as follows:

(*i*) It produces simulated coin deposit signals for testing a CDA coin tone receiver.
(*ii*) It detects test tones produced by a CDA announcement decoder.

The CDA test circuit architecture is quite similar to that of the other SSAS service circuits. It consists basically of digital circuitry to interface with SSAS, a tone generator and a tone detector (see Fig. 12). Its digital interface circuitry with SSAS consists of command decoding logic, reply registers, and control logic which provides additional decoding of CDA test circuit commands, and also controls the miniature relays which provide switching of pads and filters.

The tone generator consists of an announcement decoder, an inter-

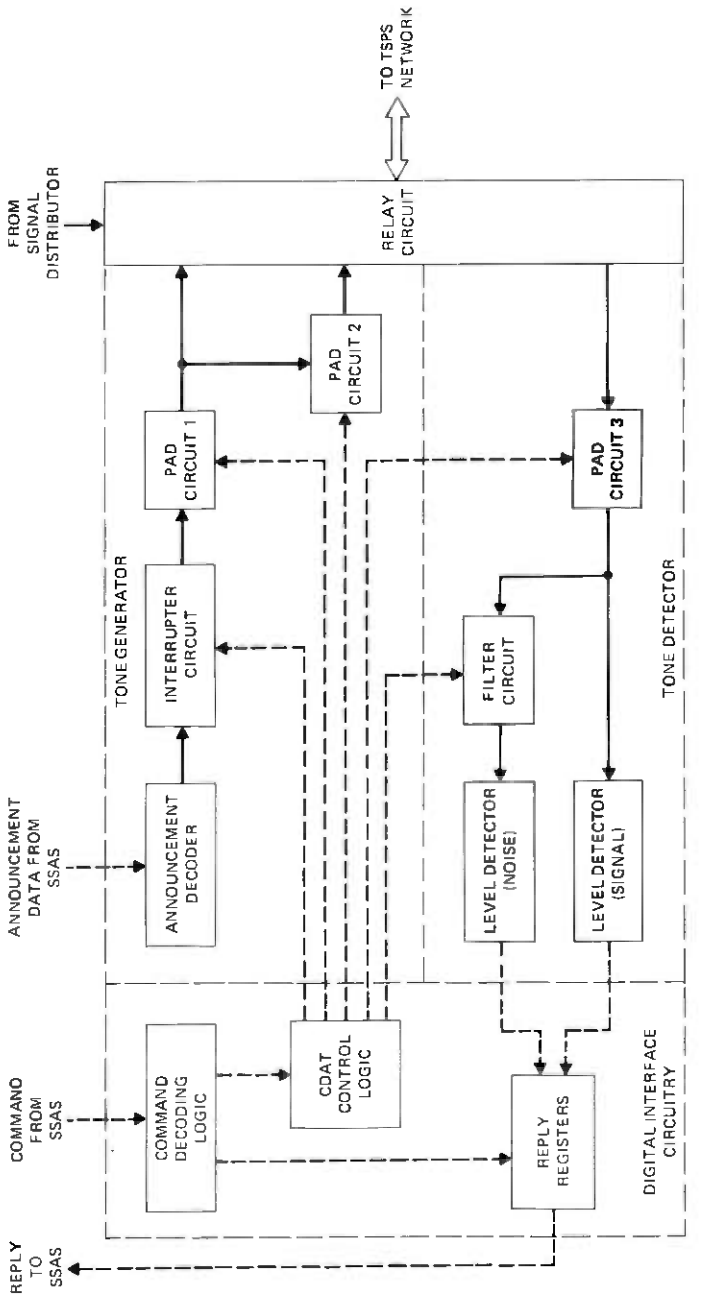


Fig. 12—Block diagram of CDA test circuit.

rupter switch, and a set of switchable attenuation pads. The interrupter switch is used to clamp the analog output of the announcement decoder to ground, thereby silencing the tone generator. The switchable pads provide the level variation needed for a CDA coin tone receiver testing.

The tone detector consists of a switchable pad circuit, level detectors, and a switchable filter circuit. The switchable pad circuit provides level adjustment for signals generated by a CDA announcement decoder. The output of the announcement decoder of a CDA circuit is tested by the level detectors for both signal loss and distortion. During the process of checking for signal distortion, the switchable filter circuit is used to remove the expected fundamental frequency component of the announcement decoder test tone without attenuating any harmonic frequency components.

2.4.3 Use of test tones from announcement store

Many tests performed during the course of CDA diagnostic are aimed at detecting possible malfunctions in either the announcement decoder or the coin tone receiver associated with the CDA under test. Proper operation of the announcement decoder is verified by monitoring its response to a set of digitized test tones. This response is checked for possible abnormalities in both frequency response and output level using a tone detector associated with the CDA test circuit. The coin receiver is tested by using the tone announcement decoder of the CDA test circuit to produce coin deposit signals from data stored in the SSAS announcement memory, and checking the response of the coin tone receiver to deposit signals having various level, frequency, and timing characteristics.

The test tones required for checking the proper operation of the announcement decoder and the coin tone receiver are stored as announcement data. A total of 18 test tones are required for CDA diagnostic tests. The storage requirement for many of these tones is significantly reduced by packing several of them into a single announcement phrase. Combining up to four 125-ms test tones into a single segment reduces the number of the required segments to six. One hundred twenty-five milliseconds allows ample time for the announcement decoder output to stabilize and for the CDA test circuit tone detector to achieve a stable measurement. However, this approach adds additional timing functions for the standby PROCON to perform. PROCON must determine the appropriate time to strobe the tone detector and, for some tests, it must blank out the unwanted tones by operating an interrupter switch within the CDA test circuit which turns off the analog output of the tone announcement decoder.

2.5 CDA alignment and manual tests

2.5.1 Trunk test panel connections

The TSPS Control, Display, and Test (CDT) panel provides a means of manually initiating tests on trunks, operator positions, and service circuits appearing on the TSPS switching network.[6] In conjunction with the ACTS feature, provisions have been made for testing both the CDA and the CDA test circuits from the CDT panel.

Most CDA circuit tests originating from the CDT panel require that test connections be established between the CDA circuit and the CDT circuit. Operation of the Master Test Line–Test key on the CDT panel causes the SPC to connect an MF receiver to the Position Master Test Line (PMTL) of the CDT circuit via the TSPS network. Once this connection is established, the SPC lights the Master Test Line (MTL) lamp on the CDT panel, indicating that codes can be entered via the CDT multifrequency (MF) key set. At this point, a test connection can be requested with any CDA circuit appearing on the TSPS network by entering its trunk group and member number and the camp-on bit via the CDT MF key set. A special display on the CDT panel indicates to the maintenance personnel whether the request was successful, aborted due to system failure, or invalid.

The SPC indicates the status of the requested CDA on a special display lamp. This lamp indicates whether the CDA is idle, traffic busy, maintenance busy, or out of service. If the CDA is busy and camp-on is requested, the SPC reserves it for maintenance by placing it out of service as soon as it becomes idle and notifies the maintenance personnel of its availability via a teletypewriter output message. If the CDA is idle or out of service, the SPC disconnects the MF receiver from the PMTL, extinguishes the MTL lamp, and establishes the test connection between the CDA and CDT circuits.

Several other test features involving CDT panel key actions have also been provided for CDA circuits. By operating appropriate keys, the maintenance personnel can place the CDA connected to the CDT out of service, or return it to service. Also, a display of the ferrod states of the CDA under test can be requested by operating a special key. This key action causes the scan row containing the CDA ferrods to be displayed on the program display of the control and display panel.

2.5.2 Trunk test panel measurements

The CDA circuits contain analog transmission components which provide a part of the talking path between the customer or customers and the CDA announcement decoder. These CDA circuit voice paths must be tested with respect to noise and signal loss whenever transmission problems associated with the circuit are suspected. The trunk

test panel, shown in Fig. 13, provides a convenient means of performing noise and loss tests on CDA circuits that appear on the TSPS network.

Testing of the transmission path of a CDA circuit involves separate tests for performing noise and loss measurements. First, a test connec-

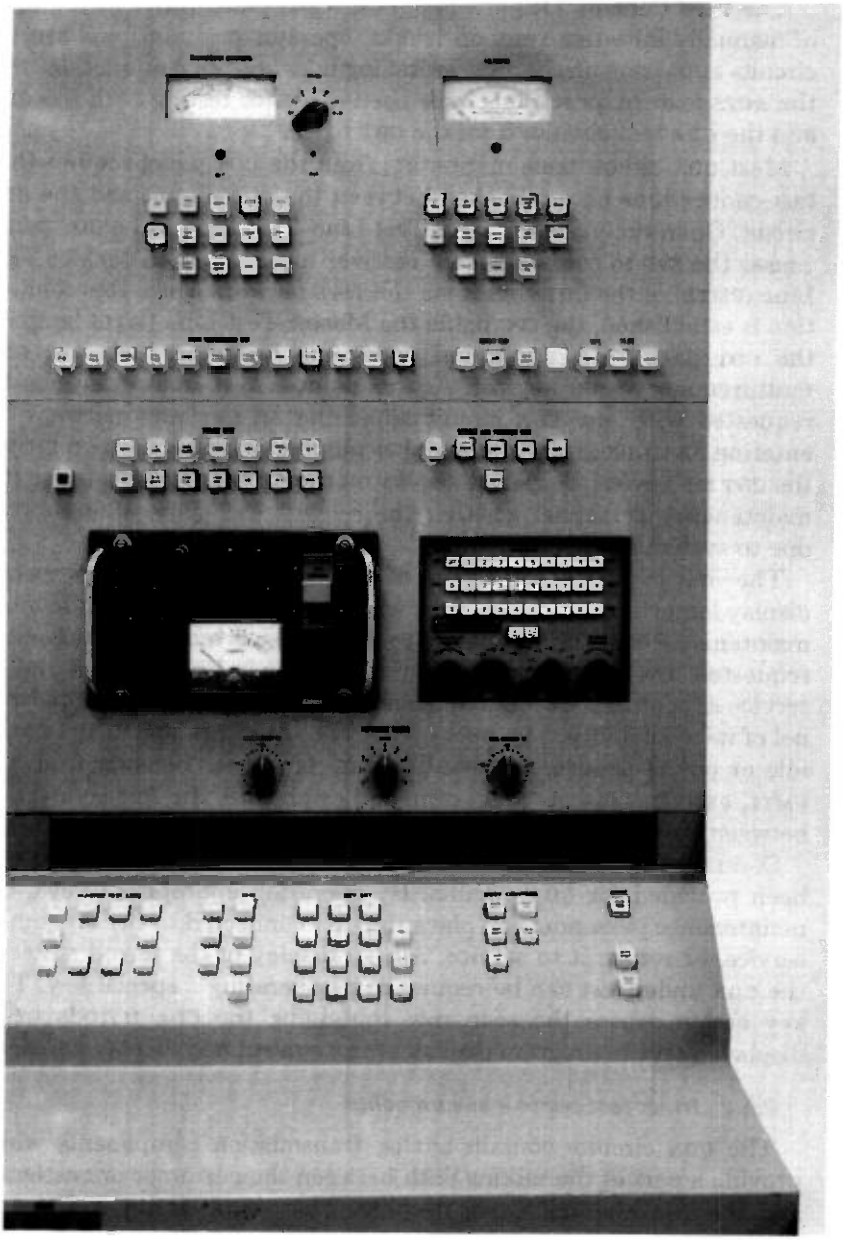


Fig. 13—Trunk test panel.

tion is established and the Master Test Line–Test key is operated to signal the SPC to connect an MF receiver to the PMTL. Next, by entering an appropriate test code and appropriate test control key actions, each of these tests can be initiated. These tests can be repeated or performed in any sequence as long as only one test control key is operated at a time. These tests can also be sequenced under the SPC control by using an appropriate test code. As soon as the code is entered, the SPC establishes the connection for the first test. Upon completion of the first test, the connection for the second test is established. This process is continued until the completion of the last test. At this point, the CDA circuit is disconnected from the CDT panel and is returned to its previous state.

Manual transmission tests of all CDA circuits in the office can be initiated by entering a special test code. The SPC connects each CDA circuit in sequence to the CDT panel. If the circuit is traffic or maintenance busy, it is skipped and a teletypewriter message is printed. After the completion of the tests, the SPC disconnects the CDA circuit from the CDT and connects the next one.

Automatic sequencing of the transmission tests of all CDA circuits can be initiated by entering another special code. The SPC connects each available CDA circuit to the CDT panel, and sequences through each of the tests. Once all of the tests for a CDA circuit are complete, another CDA circuit is connected. Busy circuits which are skipped are identified by teletypewriter messages.

2.5.3 Vocabulary recital

Because the announcement decoding aspect of the SSAS has no existing counterpart in the TSPS, a totally new test feature is provided from the CDT panel. This feature is an announcement vocabulary recital exercise whereby the maintenance personnel can listen to each speech segment in the announcement vocabulary via the CDT telephone head set.

Listening to the announcement output of a CDA circuit provides a manual check not only of the CDA circuit announcement decoder, but also of the SSAS announcement subsystem circuits. When an announcement store is loaded from tape, this feature can be used to verify the contents of the store before it is made active. The announcement recital test can be performed from both active and standby SSAS sides. Two options are provided for the recital test. They can be summarized as follows:

(*i*) The decoding of every announcement segment and phrase in sequence.
(*ii*) The repetitive decoding of a specified announcement segment or phrase.

The announcement recital test on an active SSAS side begins with a CDA circuit connected to the CDT panel. By operating appropriate keys and entering a special test code, the SPC starts the PROCON-controlled announcement exercise on the active side. This exercise accesses the announcement data for a phrase and sends it to the CDA circuit under test. It sequences through every word or phrase in the announcement vocabulary and then stops. Each word or phrase is separated by ½ second of silence.

Entering another test code causes a specified announcement phrase to be repeated at a constant rate. This is controlled by another PROCON exercise which accesses the announcement data for the specified phrase and sends it to the CDA circuit under the test. One-half second of silence is inserted between repetitions. The phrase is repeated until the circuit is released from the CDT panel.

Similar procedures are used for the announcement recital test from a standby SSAS side. Again by entering special codes, appropriate announcement exercises are started. If the SSAS diagnostic is in progress on the standby side, the exercise is not run and a teletypewriter output message is printed.

III. ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of many colleagues who took part in the development of both the operational and the maintenance software described in this article. Among them, several people deserve special mention: J. Atkins, J. R. Connet, D. C. Dowden, I. S. Dowden, R. V. Miller, and R. D. Nafziger for their work in operational programs; M. W. Medin, D. G. Raj-Karne, W. R. Serence, and M. D. Soneriu in maintenance software; and R. H. McGuigan and S. B. Windes in growth and retrofit.

REFERENCES

1. M. Berger, J. C. Dalby, E. M. Prell, and V. L. Ransom, "TSPS No. 1: ACTS Overall Description and Operational Characteristics," B.S.T.J., this issue, pp. 1207–1223.
2. A. W. Kettley, E. J. Pasternak, and M. F. Sikorsky, "TSPS No. 1: Operational Programs," B.S.T.J., *49*, No. 10 (December 1970), pp. 2625–2683.
3. R. J. Jaeger, Jr., and A. E. Joel, Jr., "TSPS No. 1: System Organization and Objectives," B.S.T.J., *49*, No. 10 (December 1970), pp. 2417–2443.
4. G. R. Durney, H. W. Kettler, E. M. Prell, G. Riddell, and W. B. Rohn, "TSPS No. 1: Stored Program Control No. 1A," B.S.T.J., *49*, No. 10 (December 1970), pp. 2445–2507.
5. G. T. Clark, K. E. Streisand, and D. H. Larson, "TSPS No. 1: Station Signaling and Announcement Subsystem; Hardware for Automated Coin Toll Service," B.S.T.J., this issue, pp. 1225–1249.
6. W. K. Comella, C. M. Day, Jr., and J. A. Hackett, "TSPS No. 1: Peripheral Circuits," B.S.T.J., *49*, No. 10 (December 1970), pp. 2561–2623.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July-August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Automated Coin Toll Service: Human Factors Studies

By E. A. YOUNGS, W. J. BUSHNELL, and A. BARONE-WING

(Manuscript received December 28, 1978)

Automated Coin Toll Service replaces operator handling on most toll calls paid for with coins and provides operator handling when customers fail to deposit. Human factors work on ACTS consisted of first documenting the range and frequency of existing operator interactions, then simulating various possible versions of machine-provided service, and last analyzing simulation results to provide service recommendations, performance estimates, and a list of important, but unanswered, questions. The simulation also led to the early development of operator practice and training materials, evaluation tools, and operational requirements, as well as timely discovery of potential problems.

I. INTRODUCTION

Companion papers describe the hardware and software used to provide Automated Coin Toll Service (ACTS).[1-3] This paper describes human considerations which contributed to the design and evaluation of the automated service.

Prior to the development of the hardware and software described in the previous papers, many questions were raised:

(*i*) Would the automated service be an acceptable substitute for Traffic Service Position System (TSPS) operators performing routine functions, e.g., deposit request, coin counting, deposit prompt, deposit acknowledgment, and others?

(*ii*) How would customer depositing (and other behaviors) depend on service design?

(*iii*) How would customer satisfaction with the service depend on the design?

(*iv*) How would operators be affected by the automated service?

These few broad questions implied many specific questions: How should performance be measured? What aspects of customer behavior are most important? and How should these be measured?

The human factors work proceeded in three stages: detailed observations of existing (operator-assisted) coin toll service to comprehend the range and relative frequencies of events; a highly instrumented service trial to obtain performance measurements and customer opinions; and a data analysis and recommendations phase to formulate a final service offering, estimate performance, and isolate potential difficulties for further study. The remainder of this paper is organized around these stages.

Section II briefly describes early observing studies upon which the original service proposal, as well as the provisional announcement and timing schemes, were based. Additional confirming observing studies are also mentioned. Section III describes the usefulness of service simulations—first with operators only, later with computer-driven equipment to provide service and measure customer-service performance. Section IV describes the technical implementation of the computerized simulation, Section V describes simulation results and recommendations, Section VI estimates performance and acceptance, Section VII enumerates unanswered questions, and Section VIII briefly mentions some cost-worth factors to be considered in evaluating the simulation.

II. EARLY HUMAN FACTORS WORK—OBSERVING HOW COIN CALLS ARE PRESENTLY HANDLED BY OPERATORS

At TSPS, toll coin calls are divided into several parts (called position seizures) for efficient operator handling. Each seizure accomplishes a single function: (*i*) initial period deposit request and verification of complete deposit, (*ii*) notification at the end of the initial period that overtime charges would apply unless the call ended immediately, (*iii*) call interruption, deposit request, and verification after each additional 10 minutes of overtime conversation, and (*iv*) ringback (if necessary), deposit request, and verification at the end of conversation. The human factors effort concentrated on replacing operators performing these functions.

The initial service proposal by J. C. Dalby was based on the characteristics of (operator-assisted) station-to-station coin toll calls monitored at Neptune, New Jersey and Chicago, Illinois.[4] Of particular interest were:

(*i*) The wording of the operators' deposit requests.

(*ii*) The time from the deposit request until customers deposited their first coins.
(*iii*) The number of times operators repeated deposit requests before customers started depositing.
(*iv*) The times between coins (intercoin intervals).
(*v*) How frequently customers began depositing, then stopped and asked operators how much was still due.

C. E. Bronell and M. S. Schoeffler[5] conducted a more extensive service measurement study in Seattle, Washington, Miami, Florida, and Rochelle Park, New Jersey. The objective of this study was to reinforce earlier observations and accumulate enough direct observations of operator-handled calls to provide a thorough, general description of the process and of the most common situations likely to be difficult for an automated system. Location differences were also analyzed.

From these observational studies, it became apparent that there are two general types of coin customers: experienced and (those presumed to be) inexperienced. Experienced customers are familiar with coin service. They had correct change and often knew the exact charges before making the call. After a short initial deposit request (e.g., "50 cents please"), experienced customers immediately started depositing and continued until the requested amount was deposited. If the operator used a longer initial deposit request (e.g., "Please deposit 1 dollar and 75 cents for the first 3 minutes"), experienced customers often began depositing while the operator was still talking.

At the other extreme, inexperienced customers are not prepared for the initial deposit request and often ask for a repetition. They began depositing, but occasionally lost count and asked for the remaining amount due. Inexperienced customers often did not have the correct change and had to deposit too much. The operator informed these customers that extra deposits would be credited toward additional charges, if any. Some inexperienced customers dialed the call as if they wanted to pay with coins when they did not wish to do so and were unprepared. (They should have dialed 0 plus the called number.) On these calls, the operator adjusted the method of payment.

As a result of these observations, it was concluded that the sequence of announcements and coin deposit intervals should provide experienced customers with fast efficient service and inexperienced customers with repetitive requests for deposits, prompts for the remaining amount due, and automatic transfer of calls to operators when difficulties were detected. In addition, the automated service should be able to acknowledge extra deposits and provide credit toward additional conversation time.

It was estimated that provisional announcements and deposit inter-

vals would handle approximately 85 percent of initial deposit seizures without operator intervention. On this basis, automated service was an economically sound development. However, it was also recognized that further refinements in the announcements and timing intervals which increased the proportions of mechanized seizures, even by only a few percent, would yield substantial additional savings.

III. FIRST AUTOMATED SERVICE—SIMULATION BY OPERATORS

In the spring of 1975, O. O. Gruenz, Jr., conducted an experiment at Harrison, New York, Fort Washington, Pennsylvania, and New York City, in which the automated service was manually simulated.[6] Several TSPS operators requested deposits and prompted customers using proposed announcements and approximate timing intervals. These operators were instructed to ignore any questions or comments made by customers.

During the experiment, two variants of the provisional deposit request ("Please deposit 35 cents for the first 3 minutes") were tried to reduce customer requests for repetitions of the amounts due:

(*i*) "Thirty-five cents please." (2-second pause) "Please deposit 35 cents for the first 3 minutes,"

(*ii*) "Please deposit 35 cents." (2-second pause) "Thirty-five cents for the first 3 minutes."

Both these announcements were successful in reducing customer needs for additional information. Furthermore, these results indicated that mechanized initial deposit seizures could be increased by several percent with the proper announcement wordings and timing intervals. However, because all record-keeping for this experiment was done manually, it was not feasible to explore many options. Thus, the experiment did not finally answer any design questions; rather, it demonstrated that proper design choices could substantially improve system performance.

3.1 Additional questions about customer/system performance and customer acceptance which justified additional simulation

In addition to announcement wording and deposit timing, early work raised other specific design questions:

(*i*) Would an alerting tone preceding announcements help customers deposit more quickly and/or more reliably?

(*ii*) How would customer/system performance and customer acceptance be affected by the use of a nasal monotone ("machine-like") voice as opposed to more natural-sounding announcements? (There was some speculation that informing customers that they were being machine-served would be beneficial, and it was thought that voice quality was a reliable, non-time-consuming way of doing so.)

(*iii*) Should coin station instruction cards inform customers that service is being machine provided?

Furthermore, a properly designed simulation was expected to:

(*iv*) Identify potential overall system weaknesses, overlooked by even careful examination of individual system elements and stimulate further work where required.

(*v*) Give useful experience with microprocessor-controlled, announcement-generating equipment and software.

(*vi*) Stimulate early development of operator procedures and training materials.

(*vii*) Aid preparation of an appropriate public relations campaign for the introduction of automated service.

(*viii*) Provide some engineering information useful in early site equipment provisioning.

(*ix*) Complement performance measures with customer interviews which would reveal preferences and sensitivities to machine-provided service.

IV. COMPUTER SIMULATION OF ACTS

Various versions of the automated coin toll service were simulated using microprocessor-controlled, announcement-generating equipment. This equipment was originally built by an exploratory development group to demonstrate the feasibility of generating announcements using digitized segments of speech stored in semiconductor memory. However, this equipment lacked much of the hardware and software necessary to provide the fully automated service. First, the coin detection circuits had not been designed and, second, the circuitry and software to interface the microprocessor and the TSPS processors had not been developed. Thus, to simulate the automated service, it was necessary to have TSPS operators relay call and coin deposit information to the microprocessor. A minicomputer and a TTY were used as interface between the operator and the microprocessor. In addition, the minicomputer provided all the necessary timing intervals and a record of the call events on a 9-track tape (see Fig. 1).

The sequence of events on a simulated ACTS call were as follows:

(*i*) The call seized a position and the initial period length and charges were displayed to an operator.

(*ii*) The operator transmitted the information to the minicomputer via the TTY.

(*iii*) The minicomputer sent instructions to generate an announcement (deposit request) which was patched into the customer-operator voice path.

(*iv*) As the customer deposited coins, the operator depressed keys on the TTY which were interpreted by the minicomputer as nickels, dimes, or quarters.

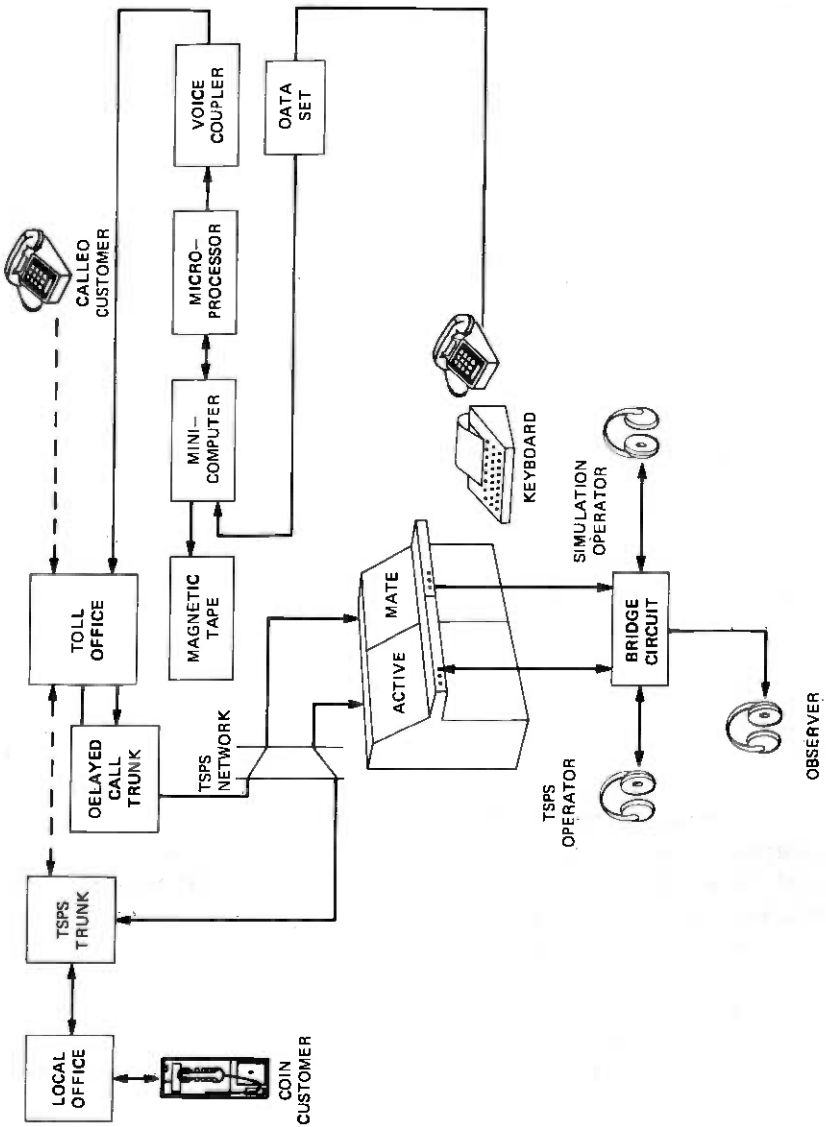


Fig. 1—Simulation block diagram.

The timing and initiation of the announcements were under minicomputer control, as were instructions to the operators. If the customer did not deposit coins, a second operator (sometimes with no previous knowledge of the call) assisted the customer, providing a customer-operator situation much like the eventual service.

Two operator positions (each with two operators) were equipped with TTYs. To ensure that these positions received only station paid coin calls, a special set of program overwrites was installed in the TSPS. In addition, other overwrites were inserted to ensure that calls at the special positions originated from one of a selected set of coin stations. Coin stations were selected to facilitate interviewing, traffic management, station instruction card study, and traffic sampling.

4.1 Administration of the simulation—hourly traffic distribution

Figure 2 illustrates the hourly distribution of simulation traffic and the corresponding distribution for coin-paid toll traffic at Illinois Bell's Great Lakes TSPS. This representative sampling for weekdays was attained by choosing an appropriate mix of 8-hour shifts over the three months of study. The simulation was operated for one shift per weekday, as well as a few weekends.

4.2 Station selection and activation

Stations participating in the trial were those most frequently used for coin-paid toll calls in the host TSPS serving area. A total of 732 stations were activated at various times. However, at busy times during the day, fewer than 100 generally received simulated service. The number of active stations was adjusted dynamically to keep (one or) two trial positions as busy as possible, without overflowing calls to the regular team more than 10 to 15 percent of the time, on the average.

4.3 Varying elements of the service design

To answer the service design questions at hand, a daily schedule of service variants was constructed to sometimes include an alerting tone/no tone, a "machine-like"/natural announcement voice, etc. Overall service was composed of variable elements in a way that allowed the effects of each element to be partialed out by appropriate analyses. Ongoing analyses quickly suggested that some service elements be eliminated from further consideration and that others be tried.

V. CUSTOMER/SYSTEM PERFORMANCE—ANALYSES, RESULTS, AND DESIGN RECOMMENDATIONS

Table I presents some key results derived from simulating over seventeen thousand automated initial deposit seizures. The table

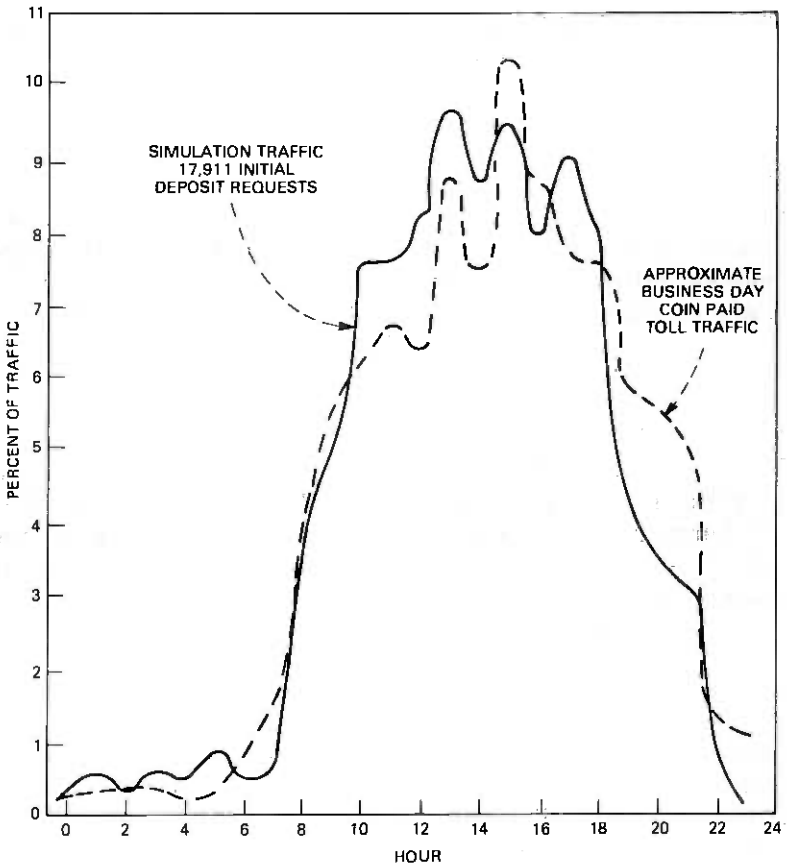


Fig. 2—Approximate business day and simulation study sampling distributions.

shows regression estimates of the effects of design decisions on important service performance measures. Mechanization rate is the percentage of calls handled without operator assistance. Walkaway rate is the percentage of calls going into overtime but not fully paid for by customers "walking away." Abandonment rate is the percentage of calls that customers abandoned prior to receiving busy or ringing signals.

The body of the table contains estimates, as percentages, of the incremental effects of several important service components. For example, the alerting tone is estimated to decrease mechanization rate by 0.2 percent, to increase walkaway rate by 1.4 percent, and to have no measurable effect on abandonments. Service component effect estimates are arbitrarily constrained to a zero sum on each performance measure.

Table I—Estimates of design decision effects on initial deposit request handling performance measures
(√ indicates decisions taken)

Service Components	Estimated Performance Measure Increments		
	Mechanization Rate (%)	Walkaway Rate (%)	Abandonment Rate (%)
Alerting Tone (T)			
Present (Tone)	−0.2	+1.4	Nil
√ Absent (No Tone)	+0.2	−1.4	Nil
Voice Quality (V)			
Machine (Mach)	−0.1	+1.8	+0.2
√ Natural (Nat)	+0.1	−1.8	−0.2
Announcement Wording			
"Please deposit $X.XX for. . ."	−2.2*	+4.7	+0.1
√ "$X.XX please. Please deposit. . ."	+1.3*	−1.7	−0.8
"Please deposit $X.XX, $X.XX for. . ."	+0.9*	−3.0	+0.7
Deposit Interval Allowed			
4.5 secs	−3.7*	−1.6	−0.9
√ 6.0 secs†	+1.3*	+1.5	−0.3
8.0 secs	+2.4*	+0.1	+1.2
Speech Detector (S)			
Present (Det)	−1.0	+0.9	−0.1
√ Absent (No det)	+1.0	−0.9	+0.1
Interactions			
Tone/Voice			
√ Tone-mach or No tone-nat	+0.7	+0.5	+0.3
Tone-nat or No tone-mach	−0.7	−0.5	−0.3
Voice/Speech Detector			
√ Mach-det or Nat—no det	+0.3	−1.5	−0.2
Mach—no det or Nat—det	−0.3	+1.5	+0.2
Constants	86.4	19.8	11.5

* (Column) contrasts statistically different ($p < 0.01$).
† 5.5 secs selected, but 6.0 secs estimates used.

Most individual effects are estimated to be small—too small to be considered statistically different from zero with confidence. Those large enough to reach statistical significance at the 99-percent confidence level are indicated with asterisks. However, estimated effects are additive, in general.

The lower section of the table, labeled "Interactions," presents the estimable two-component interactions. These are, in effect, adjustments to the additive estimates above them. Because of the experimental design, not all such adjustments can be estimated.

To compare the performance of designs, sum estimates corresponding to each service design. For example, the service utilizing tone, natural voice, operators' announcement wording (shown first in the table), a 4.5-s deposit interval, and no speech detector would be expected to have a 10.3-percent lower mechanization rate than the service indicated by the checks (√). By adding the constants given in the bottom row of the table, the actual estimated rates are obtained.

Thus, the checked components result in estimates of 91.3, 14.5, and 10.4 percent for mechanization, walkaway, and abandonment rates, respectively, whereas the example alternative service results in estimates of 81.0, 19.6, and 10.1 percent, respectively—slightly better abandonment performance at the expense of mechanization and walkaway rates.

5.1 Alerting tone effects and recommendations

Prior to the simulation, it was reasoned that a short alerting tone would better prepare customers to understand the announcement requesting deposit and/or better prepare them to deposit promptly in cases when they knew the amount from previous experience. Thus, use of the tone should result in shorter times to first coin and better overall performance.

In addition to the effects presented in Table I, our data indicated that the ½-s alerting tone we used had no significant effect on time to first coin in most situations. However, on intermediate overtime request seizures, the tone boosted the seizure mechanization rate significantly: the tone, apparently, interrupts conversation more effectively than the announcement alone. Thus, the alerting tone is being used only on intermediate deposit seizures.

5.2 Voice quality

As mentioned earlier, a machine-like voice quality was tested in the simulation. It was thought customers might be able to respond more effectively to deposit requests if they realized they were being machine-served and they did not try to converse with the system.

The system performance effects of the machine-like voice were small, mixed, and generally not of practical importance; though walkaway rates are, apparently, increased. Overall, trends favored a natural voice slightly. However, interviewed customers strongly preferred the natural voice. Consequently, natural voice announcements are incorporated in the system.

5.3 Announcement wordings

Fairly early in the simulation study, the operator deposit request phrase (typically), "Please deposit 25 cents for the first 3 minutes," (listed first in Table I), was found to be less effective for the automated service than others. Repetition of the amount decreased average time to first coin and walkaway rate while increasing average proportion of seizures automated.

In retrospect, the improved performance associated with repeating the amount due is not surprising since Bronell and Schoeffler found that operators repeat deposit requests 3 to 15 percent of the time.

Because the automated service would not be able to respond directly to customer repeat requests, improvement is achieved by repeating amounts to everyone who does not deposit immediately.

The request wording chosen for ACTS is "25 cents please." (2-second pause) "Please deposit 25 cents for the first 3 minutes." Most customers begin depositing during the pause, truncating the announcement sequence.

5.4 Deposit and announcement intervals

The primary tradeoff for deposit timing is between allowing short intervals that favor customers who deposit quickly (but occasionally need the amount repeated) or need an operator,* and long intervals that avoid attaching an operator needlessly for customers who deposit slowly.

Announcements that repeat the deposit amount reduce the need for whole announcement repetition. Thus they increase the overall desirability of long interannouncement intervals, i.e., fewer customers must wait for prompting announcements. However, for customers needing an operator, lengthening interannouncement intervals degrades service quality. Therefore, in the final ACTS environments where relatively few customers reach ACTS when they need an operator, relatively longer intervals are desirable; in the environment where more who reach ACTS need an operator, shorter intervals are desirable. The appendix contains a discussion of a potential means of discriminating between customers who need operators and those who do not.

For initial deposit request seizures, 5.5-s interannouncement intervals were incorporated into the system. This choice represents an attempt to balance the mechanization rate benefit of 6.0 s and the walkaway reduction of 4.5 s (see Table I). For intermediate and end-of-conversation seizures, an 11-s interval is followed by operator attachment because repetition of the whole announcement was found to be ineffective in obtaining additional deposits.

5.5 Instruction cards

Bright orange instruction cards were placed on half the simulation coin stations, chosen randomly. Previous studies had indicated that coin station instruction cards are rarely noticed or used by customers. It was reasoned that, if such an attention-getting color was ineffective, no less noticeable card would work.

As anticipated, no practical, consistent changes in customer/system performance resulted from the use of bright orange cards. Customers

* Incorrectly dialing a "1" instead of "0" prefix will result in reaching ACTS instead of an operator. This dialing error is quite common now in areas where "1" is used to specify a station toll call to be paid by coin deposit.

did use stations with orange cards about 6 percent more often on the average, indicating that they noticed the cards from a distance. However, when interviewed about service, immediately after using stations with these cards, customers very rarely remembered any card details, including the color or the statement, "Note: Charges may be requested by a recorded announcement during an equipment trial." Therefore, no special card design or card information was recommended.

VI. CUSTOMER/SYSTEM PERFORMANCE AND CUSTOMER ACCEPTANCE ESTIMATES FROM SIMULATION RESULTS

Simulation results are not only useful in reaching design decisions but also in estimating customer/system performance for the design chosen. However, these estimates are strictly applicable only to the simulation site. Because the Bronell and Schoeffler service measurement study indicated that sites bear many overall similarities as well as significant differences, many findings of the simulation are expected to be accurate more generally. (Field performance of the final ACTS in the simulation site will provide valuable information about the adequacy of the simulation.)

6.1 Times to first coin

Fifty percent of first coins were obtained within 6 seconds of the beginning of the first announcement. Ninety percent of first coins were obtained in 12 seconds, 95 percent in 14 seconds.

6.2 Proportions of seizures mechanized

Seizure mechanization rates are estimated at 91 percent for initial deposit seizures, 64 percent for intermediate deposit seizures, and 76 percent for end-of-conversation deposit seizures. Notification (at the end of the initial period) seizures, because they require no deposit, will be essentially 100 percent automated. Overall, these rates are expected to rise with customer experience, but the rate of increase cannot be predicted from simulation data, since it was conducted over a relatively short (3-month) period with intermittent service at a limited number of stations.

6.3 Abandoned call attempts

Abandoned call attempts, which can occur only on initial deposit seizures, are estimated to be about 10 percent of initial deposit seizures. This rate is expected to decline as customers become accustomed to service. Abandonment from the stations used in the trial, on calls handled normally by operators, was about 7 percent.

6.4 Walkaway rate

The walkaway rate for the design chosen is estimated initially to be about 14.5 percent of calls extending beyond the initial period. How-

ever, the validity of this estimate is difficult to assess. It is considerably higher than the 7 percent for operator-handled calls from the same stations over the trial period. Furthermore, even though no trends were detected during the trial, the walkaway rate is not expected to decrease with customer experience.

Walkaway rate is being carefully monitored in the first ACTS installations to determine the need to adopt alternative overtime collection strategies.

6.5 Customer acceptance of automated service

Interviewed simulation customers generally preferred operator service to automated service. More than a third expressed very strong preference for operator service (8 or 9 on a 9-point scale from 1 = "strongly prefer automated service" to 9 = "strongly prefer operator service"). About 15 percent preferred automated service.

The quality of announcements was the most often noticed clue to customers that they were being machine-served. Seventy-one percent of the interviewed customers receiving the natural voice announcements recognized machine handling; 79 percent recognized the machine-like voice. However, it should be noted that 28 percent of the interviewed customers who received operator service thought their calls had been automated.

Overall, interview data indicate that customers will prefer operators to ACTS initially. As they gain experience with ACTS, acceptance is expected to grow due to increasing familiarity.

VII. QUESTIONS UNRESOLVED BY THE SIMULATION TRIAL

Several important questions were not answered by the simulation study. They will be studied in early ACTS sites:

(*i*) Will long-term customer acceptance require modification of the service?

(*ii*) How will the operator task be affected by ACTS? (Customers reaching operators will often be those who have failed to satisfy ACTS.)

(*iii*) Will walkaway rates remain at tolerable levels?

(*iv*) Should ACTS employ different announcements, timings, etc., in different locations? In the same location over time?

(*v*) Should a voice detector be incorporated into ACTS? (See the appendix for discussion of the voice detector.)

VIII. COST/WORTH ANALYSIS OF SIMULATION AND OTHER HUMAN FACTORS STUDIES

In addition to the savings provided by the initial proposal for ACTS, improvements in the ACTS seizure mechanization rate due to human

factors work will yield substantial additional savings. Operator procedures and associated training materials developed for the simulation study were valuable in preparing for the actual service. Customer/system performance and customer acceptance evaluation tools prepared for the trial are now being used to evaluate early ACTS installations.

In addition, the simulation experiment forced the development, systems engineering, and AT&T operator services organizations to understand the new service in depth early in the development cycle. This understanding permitted early formulation of very detailed operational requirements. The experiment also provided an early indication of some unanticipated development problems. For example, it was discovered that the editing features for analog announcement source tapes were not adequate for producing high-quality announcements. A digital phrase-editing system was consequently proposed and developed in time for the first installation.

IX. ACKNOWLEDGMENTS

We wish to express our deepest appreciation for the efforts of those who supported the human factors studies. Particularly central were the efforts of B. W. Rogers, D. J. Miller, D. E. Confalone, W. K. Comella, and M. S. Schoeffler of Bell Laboratories, as well as the Illinois Bell staff of the Great Lakes TSPS.

APPENDIX

Considerations Leading to Proposal and Evaluation of a Voice Detector to Discriminate Between Customers Who Need an Operator and Those Who Do Not

Manual simulations revealed a problem inherent in the scheme of using machine-controlled announcements to obtain customer deposits: a *long* interannouncement interval is desirable to allow the slowest customers to make deposits without being "rushed," but a *short* interannouncement interval gives better service to customers who need immediate repetition of the announcement or operator assistance. What is needed is a way to discriminate between the two situations.

A potential solution was suggested during the manual simulations when we observed that customers desiring operator assistance often ask questions or make statements during the intervals following announcements, whereas customers preparing to deposit generally do not. For example, customers who say things like, "How much was that, operator?" "Make that a credit card call, operator," or "This is a collect call," need either an announcement repetition or an operator.

A "voice detector" was devised and built to end the ongoing deposit interval if speech occurs. Initially, this causes the announcement to be

repeated. If talking persists, an operator is more quickly attached. Intuitively, this makes possible the provision of long intervals for (silent) customers slow to deposit while responding more promptly to those (speaking) who need announcement repetition or operator assistance.

The simulation trial did not provide a sensitive test of the voice detector for several reasons:

(*i*) Potential benefit is limited *a priori* to those users who need announcement repetition or an operator. The need for announcement repetition was largely eliminated by selecting announcements that state the amount twice. Also, customers needing an operator due to dialing errors ("one plus" instead of "zero plus") were unusually rare in Chicago compared to other cities Bronell and Schoeffler characterized.

(*ii*) No sensitive measures of the service improvements afforded by the voice detector were readily available. (Customers are unaware of the voice detector, and we were able to obtain very limited observer data.)

Our limited data indicate that the voice detector operated as intended. It was triggered in about 56 percent of the cases for which a shortening of the interval was beneficial but acted only 6 percent of the time when it was not potentially beneficial. The majority of cases it missed were due to customers talking during an announcement (when it was disabled because it would have been triggered by the announcement—an easily rectifiable design flaw). However, the voice detector had only a minor impact on primary performance criteria.

Inclusion of the voice detector cannot be recommended on the basis of the data available from the simulation. However, the problem the voice detector was devised to solve still exists and could, at some future time, be shown to justify voice detector inclusion in ACTS or other mechanized services where customer speech is a reliable and useful indicator.

REFERENCES

1. M. Berger, J. C. Dalby, E. M. Prell, and V. L. Ransom, "TSPS No. 1: Automated Coin Toll Service Overall Description and Operational Characteristics," B.S.T.J., this issue, pp. 1207–1223.
2. G. T. Clark, K. Streisand, and D. H. Larson, "TSPS No. 1: Station Signaling and Announcement Subsystem: Hardware for Automated Coin Toll Service," B.S.T.J., this issue, pp. 1225–1249.
3. R. Ahmari, J. C. Hsu, R. L. Potter, and S. C. Reed, "TSPS No. 1: Automated Coin Toll Service," B.S.T.J., this issue, pp. 1251–1290.
4. J. C. Dalby, personal communication.
5. C. E. Bronell and M. S. Schoeffler, personal communication.
6. O. O. Gruenz, Jr., personal communication.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July-August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Software Development Tools

By J. J. STANAWAY, Jr., J. J. VICTOR, and R. J. WELSCH

(Manuscript received December 28, 1978)

This paper is concerned with the development tools, strategies, and methodologies employed by Traffic Service Position System (TSPS) application programmers for the creation and testing of TSPS software. Two environments are described: (i) The support environment provided by an IBM 370/168 and related TSPS support software packages for the production of testable TSPS software. (ii) The support environment provided by the TSPS system laboratory, utility system, and specialized hardware for testing the TSPS software.

I. INTRODUCTION

This paper describes the tools of TSPS software generation and testing in the following environments:

(*i*) The facilities provided via a general-purpose IBM processor for the production of testable TSPS software.

(*ii*) The facilities provided by the TSPS system laboratory for testing new and changed TSPS software in an operational environment.

Section II is concerned with the IBM support environment utilized by TSPS application programmers. The major tools and strategies are described. More important, the methodologies that make use of these tools are explained. Section III is concerned with the testing environment provided by the TSPS system laboratory. This environment combines an actual TSPS machine, a utility system and related software, and special hardware for debugging TSPS software.

II. SOURCE CODE DEVELOPMENT AND GENERATION OF SPC-LOADABLE OBJECT CODE

This section concerns itself with the general environment of TSPS source code creation, modification, and preparation for testing using the TSPS system laboratory and associated utility systems.

2.1 General software support environment of TSPS

The software required to operate a particular TSPS installation is comprised of approximately 300 programs (PIDENTs) totaling over 200 thousand 40-bit Stored Program Control (SPC) 1A machine instructions. The combination of these 300 PIDENTs into an issuable (via Western Electric) software package is referred to as a generic release. There are currently four active TSPS generic releases, each having implemented a major new TSPS feature.

The implementation of new minor enhancements are normally provided by a new release of an existing active generic. All capabilities provided by the software of a lower numbered generic are also provided by the higher numbered generic in addition to the new major feature. The starting point for software development of a new generic will be the current state of the source modules comprising the previous generic. Many of these modules remain unchanged in the new generic. Others are modified to produce a new version of the PIDENT that adds the new capabilities for the new generic. In addition, new PIDENTs are created for the new generic. It is therefore possible to have to maintain as many versions of a PIDENT as there are active generics.

2.1.1 Featuring and a single source environment

Four active generics with 300 PIDENTs per generic could imply that 1200 PIDENT source modules would have to be maintained. This is not the case. TSPS employs the use of "featuring" to maintain a single source module for a PIDENT regardless of the number of distinct versions of that PIDENT. Featuring basically means the following:

> Any addition to a PIDENT source module must be bracketed by feature control directives which, during the assembly of the source module, direct the assembler to either assemble or ignore the bracketed source code. Any existing source lines to be replaced/deleted are likewise bracketed.

This implies that a single source for a PIDENT can be maintained which is capable of generating multiple versions of the PIDENT's object module (i.e., the output of the assembler). By appropriate feature control directives to the assembler, different versions of a PIDENT can be assembled. In speaking of multiple versions of a PIDENT due to four active generics, what actually is meant is that multiple versions of a PIDENT's object module are producible from a single PIDENT source module.

The feature control directives employed by TSPS are:

INFOR feature expression

OUTFOR feature expression
ENDFOR feature expression.

The term "feature expression" in all three is a Boolean expression which is evaluated as "true" or "false" by the assembler. INFOR implies "assemble" the following source lines if the feature expression is true. OUTFOR implies "ignore" the following source lines if the feature expression is true. ENDFOR is the terminating bracket for source lines to be assembled/ignored. INFOR with a true feature expression is the same as OUTFOR with a false feature expression. "Feature expressions" may be combined with Boolean "and," "or," and "not" operations.

Each generic has associated with it sets of feature expressions that always evaluate as true. Every assembly of a PIDENT source module is initiated by the processing of a special control statement which specifies the generic for which the PIDENT is being assembled and therefore the feature expressions which are to be "true" for this particular assembly.

2.1.2 Software development support environment

Software to be executed on the TSPS SPC 1A is generated by means of the computing facilities of an IBM processor located at the Columbus, Ohio branch laboratory of Bell Laboratories. This facility operates the OS/370 operating system with the IBM Time Sharing Option (TSO). The latter point is significant because the TSPS development organization resides at the Indian Hill laboratory in Naperville, Illinois. All accesses to the TSPS software data base are through TSO via dial-up terminal access from Indian Hill. The high-speed data network (VIPERDAE) connecting Columbus and Indian Hill allows hard copy and tape output to be returned to Indian Hill.

TSO provides the needed interactive facilities to both TSPS application programmers and TSPS program administration personnel for data base administration, PIDENT creation and modification, and submission of OS/370 batch jobs for assemblies, loads, etc. It should be mentioned that this interactive environment was new for TSPS with the development of Generic 8. Before Generic 8, the software development environment was punched-card-oriented, with all functions to be performed being initiated via over-the-counter submission of card decks.

2.1.2.1 Creation and modification of PIDENT source modules. The major TSO-provided tool utilized by TSPS application programmers for the creation or modification of PIDENT source is the QED text editor. QED is a powerful, flexible, and general-purpose text editing facility capable of either line number or context editing on a range of various OS/370 file organization types. It should be mentioned at this time that TSPS application programmers do not directly modify existing

PIDENT source modules. Since a single-source module is used to generate (via assembly) the object modules for more than one active generic, it is felt that direct modification of the PIDENT source for any particular generic is too dangerous. Instead, TSPS employs the use of two editors for PIDENT source modification. Via QED, the application programmer is actually creating the editor statements to be processed by the Advanced Processor Editor (APE). APE is a very simple, line-number-oriented editor. It provides only the basic "insert," "replace," and "modify" functions and is specifically designed to operate in conjunction with the TSPS assembler. In almost all cases, APE and the TSPS assembler are executed in sequence in the batch environment of the OS/370. APE applies the edits created via QED to the official PIDENT source and outputs a temporary edited copy of the PIDENT source, which is then assembled. The actual PIDENT source module is not altered during this sequence, although the mechanism does exist in APE to permanently apply the edits to the PIDENT source module, renumber all lines sequentially, and regenerate the PIDENT source module. From this point on, any reference to a PIDENT source module is actually a reference to a PIDENT source module in combination with the official module of APE edits for that PIDENT.

2.1.2.2 The TSPS assembler. The creation/modification of TSPS PIDENT source (source + APE edits) is only the first step in a sequence. This sequence of functions will eventually produce an output from the IBM support machine which is capable of being executed and tested on the TSPS system.

The second step in this sequence is the conversion of the PIDENT source module into an assembled object module suitable for input to the load step. This conversion process combines the execution of the APE editor to produce a temporary modified source module, followed by the assembly of this modified source module by the SPC-SWAP (*Sw*itching *A*ssembly *P*rogram) assembler. Primary outputs of the SPC-SWAP assembler are an object module and an assembly listing corresponding to a particular version of the PIDENT. SPC-SWAP is an excellent, high-powered assembler which possesses an assortment of pseudo-operations for controlling listing format, symbol definitions, etc. SPC-SWAP also includes powerful MACRO definition and usage facilities.

Another facility of the SPC-SWAP assembler which is heavily used by TSPS in its multigeneric environment is the capability to create a special file (referred to as a library) of symbol or macro definitions which can then be accessed by subsequent PIDENT assemblies for the purpose of symbol or macro resolution. It is the library and macro facilities of SPC-SWAP which allow the single-source, multiple-generic environment of TSPS to be viable. The feature control directives described in Section 2.1.1 are actually macros available through the

library facility to each TSPS PIDENT assembly. The control statement which initiates each PIDENT assembly is again a macro (the PACKAGE macro). Execution of the PACKAGE macro, which basically takes a generic name as a parameter, establishes the generic environment of the assembly (i.e., what feature expressions evaluate as "true"; what generic-dependent libraries of symbol, macro, and data definitions will be available during the assembly; what "name" should be given to the produced object module; etc.). Simply by "inserting" via APE a different specification for the PACKAGE macro and reassembling, a single-source module is capable of producing many different generic-dependent versions of its object module.

2.1.2.3 The TSPS loading process. Object modules produced by the SPC-SWAP assembler are not suitable as input to the TSPS. The final step in the sequence of operations which produces SPC-compatible output is the execution of the SPC loader on the IBM support machine. Primary inputs to the SPC loader are the object modules for all PIDENTS comprising a particular generic software release and control directives specifying such things as (*i*) what areas of SPC memory are available for loading PIDENTS, (*ii*) what PIDENTS are to be loaded and, if necessary, at what addresses, and (*iii*) what types of maps and cross references are to be generated. In these respects, the SPC loader is very similar to most relocatable linking loaders. Primary output of the SPC loader is an 800-bits-per-inch, 9-track tape representing the relocated, fully linked generic load module which is capable of being read into SPC memory and executed. Corresponding to the tape output is a printed load map showing memory assignments, available space, etc., and optionally cross-reference listings showing entry point definition and PIDENTS referencing.

An additional capability of the SPC loader is somewhat unique to ESS-type loaders and is heavily used during TSPS software development. When creating a full generic software load of all 300 or so PIDENTS, the SPC loader can be directed to create a special file, called a HISTORY, into which detailed information concerning the generated load is written. The information includes the names of all PIDENTS loaded; the addresses at which they were loaded; how much space each consumed; what entry points each defined; where each entry point was referenced; where and how much free SPC memory remains; what SPC memory was originally available to be loaded; and a complete copy of the relocated, linked, and loaded SPC memory. On a subsequent execution of the SPC loader, the HISTORY file can be reinput to provide the capability for what is known as a partial load. On a partial load, the SPC loader need only be informed of changes to be made to the previous full load represented by the HISTORY. Previously loaded PIDENTS can be unloaded or replaced by new versions, new PIDENTS can be added, additional SPC memory can be made available for loading into, and a

new HISTORY reflecting all changes can be generated. Of particular importance is that, on a partial load, the tape that is generated reflects only the difference between the updated load image and the previous load image that had been saved on the input HISTORY file. This partial load facility of the SPC loader provides SPC application programmers with an incremental load capability that is the basis for one of TSPS's primary software development methodologies to be described.

2.1.2.4 The Interactive Program Administration System. IPAS (Interactive Program Administration System) is a tool developed for use by TSPS and other SPC 1A based systems. IPAS executes in the interactive TSO environment and primarily serves to shield the application programmers and program administration personnel from the complexities of the OS/370 operating system, specifically the nontrivial Job Control Language (JCL) required for batch execution of SPC-related support tools. IPAS is based on the concept of PIDENTs and versions of PIDENTs and utilizes the Bell Laboratories Data Management System (DMS), a hierarchical data base system. IPAS provides the TSPS user with access to QED for line edit file creation and a simple command language for initiating the execution of the SPC-SWAP assembler, the SPC loader, and other minor support tools. IPAS was developed for use with all TSPS generics but to date has been most extensively used on the Generic 8 development.

2.2 TSPS software generation methodologies

The discussion to this point has centered on the support tools and basic implementation strategies available for generating and preparing for execution the TSPS PIDENT source modules. The discussion now turns to the methodologies which make use of these tools and strategies. Two methodologies will be covered, one which applies to software generation in a relatively free administrative atmosphere, and another which applies to software generation in a very tightly controlled change environment. Both have been applicable to the recent Generic 8 development, and in fact both were formulated for the Generic 8 development. Both are equally applicable to the other generics. The two methodologies correspond to the two administrative modes under which TSPS application programmers work. The "development" mode implies the free atmosphere; the "frozen" mode, the tightly controlled atmosphere.

2.2.1 Development mode methodology

The "development" mode primarily applies to the generation of a major new feature, and therefore to a new generic. In this mode, the object is to provide the application programmers with as much freedom as possible in generating the new feature. For this reason, there is little restriction on how the programmers modify existing PIDENTs or create

new PIDENTs. The methodology formulated for this "development" environment is heavily based on the team programming concept and the partial load capabilities of the SPC loader.

The overall software development of the Automated Coin Toll Service (ACTS) feature was divided basically along functional lines. TSPS application programmers were organized in programming teams based on these functions. The implementation of a given function required the modification of some subset of the 300 or so TSPS PIDENTs and the creation of new PIDENTs. The functional subsets of PIDENTs were not necessarily mutually exclusive. Many times multiple functions required modification to the same PIDENT. The startup point for the software development of Generic 8 was the stable state of the TSPS software as it existed for Generic 7 in the second quarter of 1976. TSPS program administration personnel set up the proper PACKAGE macro for Generic 8 and established the Generic 8 dependent macro and symbol libraries which the PACKAGE macro would make available to the SPC-SWAP assembler when performing Generic 8 assemblies. All Generic 7 PIDENTs were then reassembled to produce relocatable Generic 8 versions of the PIDENT object modules. Recall that a PIDENT source is actually the combination of the official (i.e., Program Administration Group [PAG] controlled) PIDENT source module and the official file of APE line edits for that PIDENT. All relocatable Generic 8 object modules were then input to the SPC loader to produce a full Generic 8 load module. This load module was designated the "Base 0" Generic 8 load, capability-wise identical to the Generic 7 state from which it was generated, and loadable and executable in the TSPS system laboratory. The foundation upon which to build Generic 8 was established.

The programming teams were now capable of incrementally modifying this "Base 0" load and testing their function implementation. To illustrate the methodology, assume programming teams 1 and 2. The function of team 1 requires modification of PIDENTs A, B, and C, and the creation of a new PIDENT D. The function of team 2 requires modification of PIDENTs A, E, F, and G. Team 1 would proceed as follows:

(*i*) Exact copies of the official line edit files for PIDENTs A, B, and C would be created using QED. Each would be modified as needed for team 1 function implementation.

(*ii*) The source for new PIDENT D is created using QED.

(*iii*) An SPC-SWAP assembly is initiated via IPAS or other TSO facilities utilizing the copied and modified team 1 line edit files for PIDENTs A, B, and C and the created source file for PIDENT D. The object modules produced by SPC-SWAP are saved as team 1 object modules.

(*iv*) Using the SPC loader, initiated via IPAS, a partial load is gen-

erated based on the HISTORY file corresponding to the "Base 0" Generic 8 load. PIDENTS A, B, and C are replaced by team 1 versions, and PIDENT D is added. The partial load tape produced reflects only the differences between the team 1 load and the "Base 0" load.

(*v*) Team 1 is now capable of overlaying their partial load image on top of a known-to-be-stable "Base 0" load image in the TSPS system laboratory and testing their function unencumbered by new code from other teams.

(*vi*) The cycle can be reiterated as problems are discovered during testing, except that modifications are made to team 1 line edit files.

Team 2 has concurrently been developing their function following the same procedures as outlined for team 1. This approach allows very extensive testing of individual functions, comprising large amounts of new and modified software, to be accomplished prior to a large-scale integration of functions.

The PAG personnel again became involved with the Generic 8 development at periodic intervals (usually 6 to 8 weeks) to produce an updated base load. In preparation for performing the official reassemblies for all PIDENTS modified by the programming teams, a merging of official and team line edit files must take place.

Recalling that both teams 1 and 2 copied the official line edit file for PIDENT A, the PAG personnel would first merge the team line edit files for PIDENT A, and the result would then be merged with the official line edit file for PIDENT A to produce an updated official line edit file. PIDENTS modified by a single team required only a single merge. The final merge with the official line edit file guaranteed that no original official line edits had been inadvertently deleted or modified in such a way as to adversely affect generics. Having completed the merging process, PAG could then make any required modifications to Generic 8 macro or symbol libraries, and then reassemble all modified PIDENTS to produce updated official Generic 8 object modules. These object modules now would contain all the function code tested by the individual teams up to the time the new base was created. PAG then reexecutes the SPC loader to produce a "base $n + 1$" load module and corresponding HISTORY file. The base load image in the TSPS system laboratory is then updated to "base $n + 1$," and the new base can be system-tested to ensure stability.

Although a large amount of new and modified software is introduced with a new base load, the interval between the start of the merging process and the completion of the system testing of a new base averages about 2 weeks. During this interval, the programming teams are able to continue working against "base n," with the stipulation that any additional team line-edit changes must be incorporated with

the updated official line-edit files reflecting "base $n + 1$." Due to heavy testing of individual team functions, a minimum amount of system testing is required to stabilize the new base load, even though the functions had not previously been integrated.

Prior to Generic 8, the "development" mode methodology was based on manually overwriting a load image in the TSPS system laboratory to test new functions or including untested software in new load images. System testing of new load images was an enormous, time-consuming task. Function testing via manual overwrites was more of a hindrance to software development than a help.

2.2.2 *Frozen mode methodology*

The "frozen" mode of software generation applies primarily to active generics which have been issued through the Western Electric Company, and to the latter stages of the development of a new generic. The objective of the "frozen" mode is to maintain the maximum amount of stability in the generic software by providing a highly controlled and documented change environment through which application programmers must make software corrections.

For a generic in the "frozen" mode, software change is initiated in response to a written Trouble Report (TR) documenting a suspected problem or a needed improvement. The "frozen" mode imposes a number of restrictions on the application programmers and PAG as to the manner of software change:

(*i*) No change to a PIDENT can cause the size of the PIDENT's object module to either increase or decrease.

(*ii*) The primary method for testing is not the partial load, but incrementally applied changes to the frozen generic's load image in the TSPS system laboratory.

(*iii*) Line edit changes corresponding to a laboratory change must produce a bit-for-bit match between the PAG-generated (via PIDENT reassembly) next release of the frozen generic and the overwritten old release in the TSPS system laboratory.

(*iv*) All programmer-generated changes must be independently tested and approved by a generic test team.

(*v*) The application programmer must generate a written Correction Report (CR) documenting any changes made in response to a TR.

(*vi*) Not only must the change be independently tested, but the TR, CR, overwrite, and line edits must be approved by a generic software change review committee prior to the line edits being included in the next release of the frozen generic.

To alleviate some of the problems associated with (*i*), (*ii*), and (*iii*), above, a functionally identical set of "patching" directives exists,

available to both the SPC-SWAP assembler and the test laboratory utility system overwrite assembler. In the case of SPC-SWAP, these are TSPS system macros available to any PIDENT assembly. They are merely control directives to the utility system overwrite assembler. These "patching" directives allow the application programmers to add new software, replace existing software, or delete existing software within a PIDENT without altering the assembled size of the PIDENT.

Primary input to utility system overwrite assembler is a deck of punched cards, composed of symbolic SPC source to be assembled, and utility system control directives. A few major restrictions on the change to be assembled are:

(*i*) There is no macro capability. Any macro to be assembled must be manually expanded prior to input.

(*ii*) There is no LIBRARY facility as with the SPC-SWAP assembler. The utility system overwrite assembler must be explicitly informed of the value of any and all symbols referenced by the overwrite but not defined within the overwrite.

(*iii*) No arithmetic beyond addition and subtraction is allowed.

(*iv*) Many data defining pseudo-ops of the SPC-SWAP assembler are not recognized by the overwrite assembler.

Given these restrictions of the overwrite assembler and the previously stated restrictions of the "frozen" mode environment, the change implementation flow prior to Generic 8 (and therefore prior to general TSO usage by TSPS) went as follows:

(*i*) The TR would be received by the application programmer who would be making the software change.

(*ii*) The application programmer would generate an overwrite deck to fix the stated problem and test the fix by temporarily overwriting the generic load image in the system laboratory.

(*iii*) Satisfied with the results, the application programmer would generate the corresponding CR and a deck of line edits for the PIDENT source which produced results identical to the overwrite.

(*iv*) The TR, CR, overwrite deck, and line edit deck would then be submitted to the generic software change review committee for approval. If rejected, back to step (*ii*).

(*v*) If approved, the TR, CR, overwrite deck, and line edit deck are submitted to the generic system test team for independent overwrite testing. If rejected, back to step (*ii*).

(*vi*) If approved, the TR, CR, overwrite deck, and line edit deck are submitted to PAG. The line edit deck is included in the official line edit deck for the PIDENT(s) involved, the TR and CR are filed, and the overwrite deck is saved.

This procedure was followed for each TR requiring a software change. When a new release of the frozen generic was required, PAG would

reassemble all modified PIDENTs and regenerate the generic load, maintaining the starting address and size of each PIDENT. This new load was required to exactly match the overwritten load image in the TSPS system laboratory before it would be released to Western Electric for distribution. The methodology outlined above was successful but possessed inherent and painful shortcomings when it was time to generate and match the next release of the frozen generic. The shortcomings were primarily due to the overwrite assembler and the need to create a separate line edit deck producing identical results. The areas most susceptible to error were the manual symbol resolution and manual expansion of macros required by the overwrite assembler, but not required in the line edits.

For Generic 8, the basic theory of this overwrite methodology, with its checks and approvals, was not substantially altered. There was, however, the development of a new tool, an Overwrite Generation (OVGEN) program, which eliminated the need for separate manual generation of overwrite and line edit decks. OVGEN allowed application programmers to continue to create line edits as files via QED. OVGEN requires the application programmer to include in the line edit file special control directives, identified by the programmer I.D. and a trouble report number, which informed OVGEN of which line edits to extract for overwrite generation. OVGEN, actually a combination of three pre-processors, an SPC-SWAP assembly, and a post-processor, outputs a symbolic overwrite deck for input to the utility system overwrite assembler. The overwrite deck contains all information needed for external symbol resolution. Due to the fact that SPC-SWAP is used to assemble the line edits, the application programmer can utilize macros, libraries for symbol resolution, the SWAP data defining pseudo-ops, etc. In other words, for the application programmer, the environment is quite similar to the partial load environment, with final output being an overwrite deck instead of a partial load tape.

The advantages of the OVGEN procedures in the frozen mode are many.

(*i*) The application programmer need only create the line edits in response to a TR.

(*ii*) The macro facilities of SPC-SWAP are available for use.

(*iii*) Symbols used which are external to the line edit are resolved via pre-processing and SWAP LIBRARY facilities.

(*iv*) Line edits are individually assembled. This leads to far less assembly problems by PAG when the full reassembly of the PIDENT is performed for the next release of the frozen generic.

(*v*) The final match between the new release of the frozen generic and the old release plus overwrites is considerably cleaner due to the overwrites having been generated directly from the line edits.

III. LABORATORY TESTING ENVIRONMENT

TSPS programmers use the TSPS system laboratory complex to test and debug new or changed PIDENTS. One or more programmers working on similar program areas will schedule time in the lab. Thus, this complex has been designed to provide a working environment conducive to high programmer productivity.

Two TSPS laboratories are available to programmers. Both consist of the Stored Program Control (SPC) No. 1A/TSPS complex, the nonresident utility system, and call-oriented simulators. The SPC/TSPS complex allows programmers to test in an environment much like a typical TSPS office. The utility system and simulators provide debugging aids not found in a typical office. This section describes the hardware and software facilities which make up the laboratory complex.

3.1 Hardware configuration

The TSPS laboratory configuration is shown in Fig. 1. This section discusses each component in the configuration except the simulators, which are discussed in Section 3.3.

3.1.1 Stored program control (SPC) 1A

The TSPS is controlled by the SPC 1A.[1] The SPC consists of a processor, memory system, central pulse distributor, signal distributor, master scanner, and a maintenance control center for the man-machine interface. Each of these parts is duplicated for reliability except the maintenance control center.

The SPC processor provides the control for the TSPS by executing the instructions in the memory. The processor cycle is 6.3 μs. The registers available in the processor include a 20-bit Program Address Register (PAR), 20-bit Address Image Register (AIR) which contains the address of the most recent store access, 47-bit Memory Access Register (MAR) for storing the information read from or written into the memory, and seven 20-bit index registers. The index registers are general purpose and may be used for any function.

The word length of the SPC memory is 47 bits (40 bits of information and 7 for error correction). There are 20 bits of addressing. Nineteen bits select the memory word. The remaining bit determines which half of the word is used.

The processor communicates to the peripherals and TSPS by three units: the Central Pulse Distributor (CPD) which allows the SPC to send pulses to points in the system where fast response to an instruction is needed, the signal distributor which allows the SPC to operate or release magnetic latching relays which are connected to output points, and the master scanner which provides status and supervisory inputs to the SPC from the various units in the SPC complex.

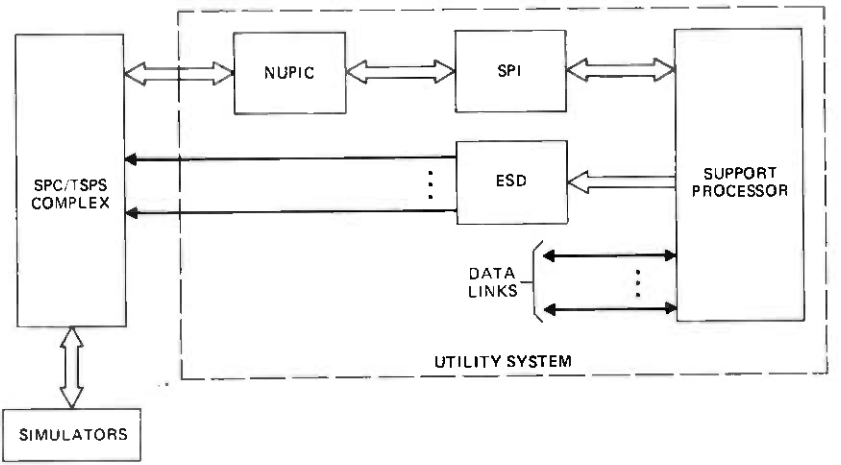


Fig. 1—TSPS system laboratory.

The Maintenance Control Center (MCC) consists of a control and display panel, a teletypewriter, and a program tape unit for loading the SPC 1A memory.

3.1.2 Nonresident utility system

The utility system provides a means for easily loading and reading the SPC memory, for debugging real-time programs in a noninterfering fashion, for controlling devices in the laboratory, and for transmitting or storing files. The advantage of this utility system over the earlier TSPS utilities is that the programs are nonresident to the SPC and that debugging of programs can be done without interfering with the execution of the system's real-time programs.

The utility system is a combination of hardware and software. This section discusses the hardware aspects of the system, and other sections discuss how the utility system is used for software change administration and program testing. The utility hardware consists of a support processor, a support processor interface, a noninteracting utility program interface console, an electronic signal distributor, and serial data links.

3.1.2.1 Support processor. The support processor used in the utility system is a commercial minicomputer. The minicomputer uses the XVMDOS operating system and its peripherals include a 9-track magnetic tape unit, fixed head disk, 3-disk-pack bulk memories, teletypewriter, high-speed printer, card reader, and paper tape reader and punch.

User files and TSPS generic programs are stored on the disk memory. These files can be updated and additional files added by means of the

magnetic tape unit, card reader, or paper tape reader. The user input is via the teletypewriter and card reader. The output is usually over the high-speed printer.

3.1.2.2 Noninterfering utility program interface console/support processor interface (NUPIC/SPI). The interface between the SPC and the support processor is the Noninterfering Utility Program Interface Console (NUPIC) and the Support Processor Interface (SPI). The NUPIC and SPI are two separate circuits but are discussed together because they are so closely related.

The NUPIC interfaces directly with the SPC processors to allow the user to monitor and control the system programs. It provides access to the PAR, AIR, MAR, and the index registers in each SPC processor. It also provides processor clock control and interrupt interfaces. The circuitry in the NUPIC allows the user to load the SPC, read the SPC memory, set program matchers, and have the contents of the SPC registers read and stored. The NUPIC has a man-machine interface consisting of SPC register displays and manual controls. These manual controls are particularly useful if the support processor should fail. The section on program testing will discuss the debugging aids available with the NUPIC.

The NUPIC is controlled by the support processor via the SPI circuit, which provides a 2-way, high-speed data communications channel. Since the support processor and the SPC processors have different word lengths and different cycle times, buffering (core memory) is provided in the SPI. The SPI is used in conjunction with the NUPIC and the support processor for loading the SPC memory, reading the SPC memory, and collecting data during program debugging.

3.1.2.3 Electronic signal distributor (ESD). In testing programs, it is often necessary to have TSPS configured differently due to multiple generics. This means that specified circuits can be connected to or removed from the TSPS buses, equipment can be removed from service, or equipment can be put into service. The ESD provides a means to do this automatically. Up to 2048 distributor points can be individually set or reset by the support processor. These points can be used to control relays, lamps, and logic inputs.

Each user can have a file on the support processor which specifies the generic program and how the ESD points should be set. Section 3.2.1.2 discusses the creation and execution of the user files.

Another use of the ESD is for physical fault insertion. This application is discussed in a later section on trouble location manual generation.

3.1.2.4 Serial data channels. Several serial data channels are available on the support processor. The channels are full duplex, with data rates up to 10K baud. The channels are used for transmitting

files or receiving files for storage. Presently, one of these channels is used for transferring user call load files to and from the microprocessor controlled call simulator (Section 3.3.3.2) and another is used for loading the writable storage unit (used in program development) of the Programmable Controller in the Station Signaling and Announcement Subsystem for the Automated Coin Toll Service[2] (ACTS) feature.

3.2 Laboratory software change administration

The desire to add new features, as well as the need to correct software errors, make it necessary to be able to easily change programs in the TSPS laboratory. This is done differently, depending on the state of the software base being changed. If the code is being developed into a new major generic issue, programmers are free to perform large-scale code modifications and additions. This is termed the development mode. If the base is already an official generic issue, only minor changes and additions can be made in response to specific troubles. This is done to minimize the need for extensive retesting after applying the change. This mode is termed the frozen mode.

This section deals with the tools available for changing the TSPS program in the lab in both of the above modes.

3.2.1 Development mode

As mentioned in Section 2.2.1, the partial load is the primary means of changing code in the development mode. A partial load tape containing all new and changed object code is brought into the TSPS lab by the programmer for debugging. The programmer then uses two support processor programs to load the base load plus the partial load into the SPC. These programs are DZLOAD and ALCFG (Automatic Laboratory Configuration). Once the desired load is in the SPC, the NOVA (Noninteracting Overwrite Assembler) program is used to make minor code revisions until a new partial load tape can be generated.

3.2.1.1 The DZLOAD program. DZLOAD is the interchange and comparator program for SPC code and data residing on magnetic tape, disk, or SPC memory. It allows the user to easily load and verify SPC programs as well as create duplicate copies of SPC memory on disk or tape. The user may choose to store a partial load tape on disk in the support processor, which eliminates the need to carry the magnetic tape into the lab for subsequent debugging sessions.

DZLOAD deals with only one file at a time; however, the support processor can also load multiple files and control the TSPS lab's hardware configuration. The program that does this is called ALCFG.

3.2.1.2 The ALCFG program. The TSPS system laboratories are used to develop and test hardware and software for use at TSPS installations in the field. The laboratory is used to simulate configu-

rations existing in the field and new hardware configurations under development.

The ALCFG program allows the user to easily and quickly change the laboratory's program and hardware configuration. The result is increased lab availability, brought about by a decrease in the manual action required to establish a particular configuration. The user can predefine a configuration and store it on a support processor disk. Thus, calling in the configuration definition under ALCFG will cause the quick reestablishment of the desired lab environment—both software and hardware. The lab hardware is controlled by the Electronic Signal Distributor (ESD). The ESD is explained in Section 3.1.2.3. ALCFG allows the user to easily define a particular ESD state.

During a lab testing and debugging session, the user will probably uncover minor coding errors or oversights. These errors can be temporarily corrected in SPC memory using the NOVA program.

3.2.1.3 The NOVA program. NOVA is a utility program that allows the lab user to overwrite SPC memory. The input to NOVA is symbolic SPC source code residing on punched cards or disk, or typed directly on the user terminal. Since the program being changed resides at a fixed SPC memory location and occupies a fixed amount of storage space, code additions must be incorporated in a "patched" fashion. NOVA is therefore designed to manage a series of patch buffers. These buffers are the actual start and end addresses of spare program memory in the SPC. Definitions of these buffers are covered in Section 3.2.2.2.

NOVA allows the user to specify program changes in either relocatable or absolute fashion. NOVA reads the overwrite statements, assembles them into SPC object code, prints a listing, and stores a binary image of this code on a support processor disk. This binary file is then loaded into SPC memory. Another binary file is also kept by NOVA, namely, the contents of the addresses specified in the overwrite prior to loading the overwrite. This file is identified by a temporary overwrite number. The user can therefore instruct NOVA to flush a specific overwrite out of SPC memory by restoring all addresses to their contents prior to the change.

Many options are available to the NOVA user. Some commonly used ones are (*i*) assemble and produce a listing, (*ii*) assemble, produce a listing, and create a binary file, (*iii*) load the last binary file created, and (*iv*) print the old data along with the overwrite listing. Another option has to do with permanent overwrites. This is covered in the next section.

3.2.2 Frozen mode

The NOVA-assembled overwrite is the primary means of changing SPC code in the frozen mode (see Section 2.2.2). The main emphasis in

this mode is placed on incremental change documentation and testing, and administration of the updated generic base load. The NOVA program is used extensively during this process. A programmer submits, in addition to TR/CR (trouble report/correction report) forms and line edits, a symbolic overwrite to a system test team. This overwrite is designed to fix a particular trouble existing in the base generic. The test team tests the change as a temporary NOVA overwrite and, if accepted, prepares to permanently incorporate it in the base generic.

3.2.2.1 NOVA—Permanent overwrites. The mechanics involved in establishing a permanent NOVA overwrite are very similar to those for temporary overwrites. The main differences are

1. No old data file is kept (permanent overwrites cannot be flushed from SPC memory).
2. The pointers into the patch buffers are permanently changed to indicate that patch used in the overwrite is no longer spare program memory.

In addition, a permanent record of the start and end address of individual patches is kept on disk. This information is extremely useful to PAG when generating a patched load (see Section 2.2.2) corresponding to the updated lab base load. Each patch origin must be defined to allow the SWAP assembler to correctly assemble the patched code into the changed program. The NOVA permanent overwrite listing is used to document the change in the lab until the new PAG load and listings are produced.

3.2.2.2 NOVA—Generic administration. Up to this point in the lab software change section, only one base load is mentioned. However, as stated in Section 2.1, more than one TSPS generic issue is usually active at one time. Consequently, NOVA must be able to properly change and patch programs for each active generic. This generic is identified by the user each time the NOVA program gets called in. NOVA uses this identification to access a set of PIDENT and PATCH files unique to that generic. The administration of these PIDENT and PATCH files is handled by a portion of NOVA, called PIDAM. The system test group usually takes care of this administration. PIDAM allows the user to define and update the PIDENT and patch files for each generic. These files contain a list of all PIDENTs that make up the generic issue along with their start and end addresses, available patch buffers, a store patch map (i.e., a map linking SPC programs to a specific patch buffer), and a list of all permanently loaded patches (see Section 3.2.2.1).

3.3 Laboratory program testing

After the SPC has been loaded with a new generic program (development mode) or changes made to an existing generic program (frozen mode), the programmer is ready to begin testing. The following sec-

tions describe the hardware, software, and simulators available to the programmers for testing programs.

3.3.1 Hardware for program debugging

The NUPIC discussed in Section 3.1.2 gives the user extensive program debugging tools. These tools include matchers, visual displays, and processor controls (Fig. 2).

The matchers available with the NUPIC include:

Address Matchers—There are nine address matchers (one manual) which can be set to indicate when a specified program address is executed or when a specified data address is read or written.

Range Trap Matchers—Two range trap matchers are available. These matchers will indicate when any address within a range of program is reached or when any address within a range of data is accessed.

Bit Matchers—Two 20-bit matchers allow the programmer to match against an address, contents of a memory location, or the contents of an index register. The bit pattern to be matched against can have an associated mask. This mask will indicate which bits in the pattern are "don't cares."

Peripheral Matcher—The peripheral matcher is used to indicate when a particular peripheral order accesses a particular peripheral unit. The peripheral unit address can have an associated mask.

Each of these matchers can be set manually via keys on the NUPIC. All but the manual matcher can also be set automatically. When a matcher "fires," the contents of the SPC registers can be collected by the support processor, analyzed, and printed on the high-speed printer.

The matchers can be set up with different options which will be executed when the matcher fires. These options can be interfering and noninterfering. The noninterfering options include register snaps and transfer traces. The transfer trace gives a record of every transfer that occurs in the program once the matcher fires.

The interfering options cause an interrupt in the SPC program execution and transfer of control to the SPC resident utilities. The options include doing a write into unprotected memory, dumping portions of memory, writing to registers, jumping to a different address, and stopping the SPC.

The visual displays on the NUPIC include binary lamp displays for the major points in each SPC processor such as the index registers and buses. Octal displays are provided for the registers most often used. These include the PAR, AIR, MAR, selected matcher address, and relocatable address (least five significant digits). The PAR, AIR, and MAR displays are available for each processor.

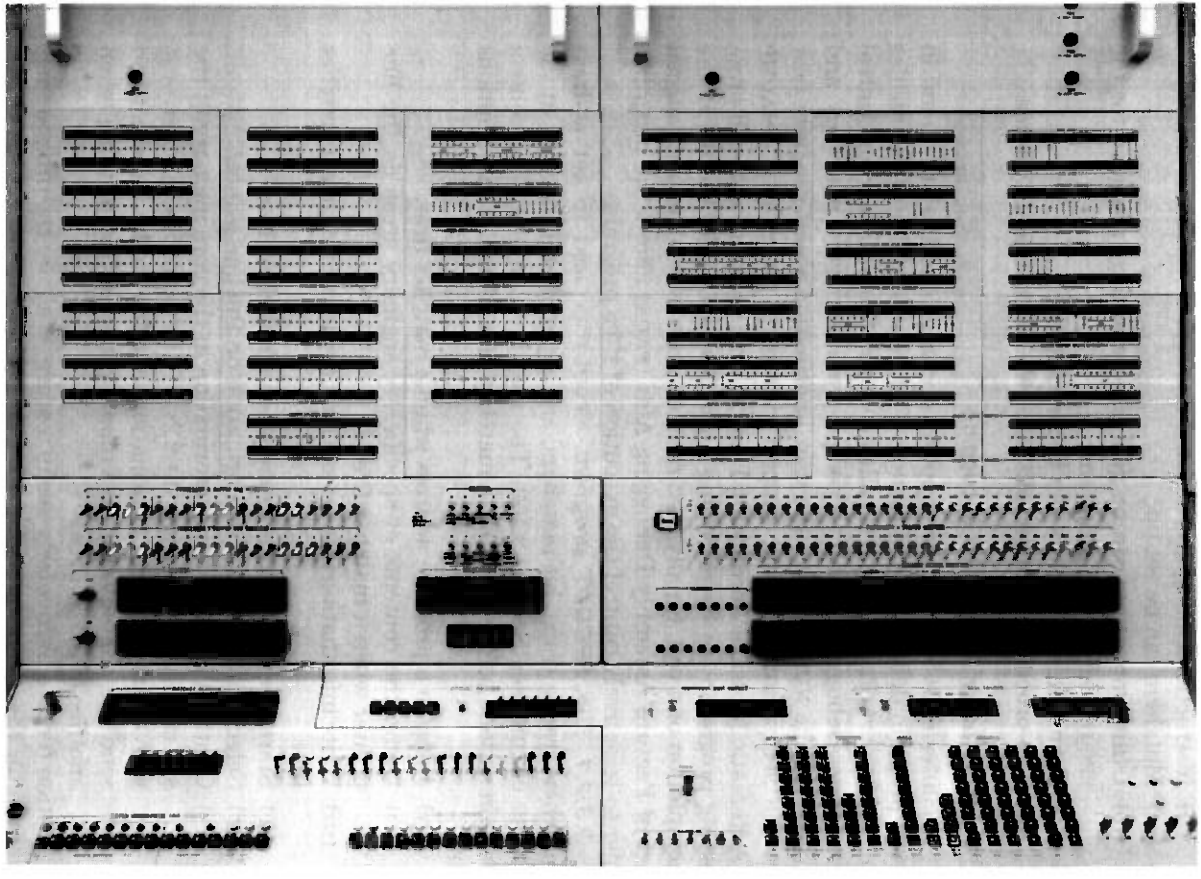


Fig. 2—Noninterfering Utility Program Interface Console.

Several manual control functions are provided to allow the user to selectively control either or both SPC processors.

(*i*) The user can stop or start the processor clock and can step through instructions one cycle or one phase at a time.
(*ii*) Instructions, data, and scanner answers can be inserted.
(*iii*) Matchers can be set up.
(*iv*) Utility interrupts can be generated and flags provided for the SPC resident utility programs.
(*v*) Transient store errors can be simulated.
(*vi*) The processors can be split into two independent systems.

3.3.2 Software for program debugging

Facilities exist in both the support processor and the SPC to aid in TSPS program debugging and testing. The most useful of these resides in the support processor and is called SPCDDT (Stored Program Control Dynamic Debugging Tool). This program activates matchers and traces via the NUPIC to give the user information about program flow in the SPC. Another useful support processor program is AMADMP (Automatic Message Accounting Dump). This program aids call processing programmers by providing formatted AMA information at the completion of a call. Finally, SPC-resident code assembled as a special set of Feature Assembly Debugging Aids (FADA) allows the programmer to selectively control the execution of TSPS programs.

3.3.2.1 The SPCDDT program. SPCDDT is used to set address matchers, range traps, bit matchers, and a peripheral matcher in the NUPIC circuit (see Section 3.1.2). The user has the ability to control certain actions before and after a matcher fires. Some of these options are:

(*i*) Jump to a program address when a matcher fires.
(*ii*) Dump the contents of specific SPC memory areas on the line printer when a matcher fires.
(*iii*) Write information into an SPC memory location when a matcher fires.
(*iv*) Start or end a transfer trace when a matcher fires.
(*v*) Selectively print trace information on the line printer.
(*vi*) Continually store trace information in the core memory of the SPI circuit (see Section 3.1.2.2). If the core fills, overwrite it with the more recent information (overlay mode). Freeze it and send the contents to the support processor for printing when another matcher fires.

Option (*vi*) is a particularly useful feature. It allows the programmer to monitor the entire flow of one (or more) program(s), and print out only the flow prior to an interesting event.

3.3.2.2 The AMADMP program. AMADMP provides formatted AMA

information on the support processor line printer at the completion of a TSPS call. This information is very useful to a programmer in debugging call processing or billing-oriented software.

Normally, the AMA Data Accumulation Program (AMAC) stores the billing information in buffers in TSPS memory. When a buffer is full, AMAC activates another program which records all buffered data on magnetic tape. This tape must then be processed on an off-line computer. The delays and logistics involved in this procedure make it undesirable for debugging purposes. Consequently, AMADMP was written during the development stage of the Remote Trunk Arrangement (RTA) feature of TSPS Generic 7.

AMADMP uses the NUPIC to activate two matchers in the AMAC program. One matcher fires at the start of AMAC's recording of the billing information and the other fires after all information has been buffered for a call. Using the information appearing in the SPC registers at the time the matchers fire, AMADMP requests a dump of the buffer locations used by AMAC. The data from these locations are then formatted in the support processor to make it more readable, and then printed on the line printer.

3.3.2.3 The FADA feature. FADA is a collection of debugging software that can be assembled into a development base load. It consists of special code added to generic TSPS programs as well as a special PIDENT (ECDB). This code does not get released officially for use in live TSPS offices.

FADA was developed during the early stages of Generic 7 as an aid in the recovery of new program loads and a debugging tool. It accomplishes this by allowing the user to selectively inhibit execution of many program functions. This capability makes it possible to simulate many low probability events, cause race conditions, and exercise program failure legs.

3.3.3 Simulators

3.3.3.1 Single-call simulators. Many times in testing programs, the programmer needs the ability to place a single call through the TSPS. Two types of test facilities are available for making single calls. The first of these facilities is the "single-line" simulator. This simulator consists of a local (calling) telephone connected to the local office side of the simulator and the toll (called) phone connected to the toll side of the simulator. The local and toll sides of the simulator are connected to a TSPS incoming trunk. This trunk is dedicated to a particular traffic type. For each traffic type (coin, hotel/motel, RTA), there is a single-line simulator.

Calls are placed by the programmer in the same way a customer would make a particular type of call. The simulator performs all

necessary signaling required by TSPS. The call is recognized by TSPS and handled appropriately by routing it to an operator's position or by connecting the called telephone.

The single-line simulator is used for calls which are to be handled normally. The simulator generates the correct KP, Start (ST) digit, and Automatic Number Identification (ANI) digit for the call. However, there are times when the programmer must have more control over the call. For instance, the programmer may wish to test a call using an improper ST digit or ANI digit. In these cases, the manual trunk test set (also known as the "Burelback Box") is used.

The manual trunk test set allows the programmer to place a single call and to have control over the entire call. The programmer selects the type of TSPS trunk circuit to be used and the call type. By operating switches, the programmer simulates seizure by the local office and responds to supervision from TSPS by keying in the KP digit, the call digits, a ST digit, an ANI digit, and the calling digits. When the toll side of the trunk is seized by TSPS, the programmer generates the toll supervision signals to TSPS. On ACTS calls, the coin tones can also be generated.

3.3.3.2 Multiple-call simulators. There are program bugs which do not show up until there is a substantial traffic load on the system. In TSPS testing, a simulated load can be generated by the Electronic Load Box (ELB) and the Microprocessor Controlled Load Box (MICLOB). Both of these "load boxes" automatically generate calls on multiple TSPS trunk circuits. Each load box simulates the functions of both the local and toll offices.

The characteristics of the ELB are:

(*i*) Generates up to 14 simultaneous calls. All calls must be the same type and the same length.
(*ii*) Can generate 1800 calls per hour.
(*iii*) Works with MF (multifrequency) trunks, 2-wire, or 4-wire, loop or E&M signaling.

To use the ELB, the user selects the number and the type of the calls desired. If more than one call type is required, another ELB must be used. To provide the user with more flexibility in setting up a call load and to provide the capability for coin signaling for ACTS, the MICLOB was developed. MICLOB (Fig. 3) has the following characteristics:

(*i*) It generates up to 32 simultaneous calls of any call mix the user wishes.
(*ii*) The call types include coin, noncoin, hotel/motel, and international.
(*iii*) All call parameters are under user control.
(*iv*) It works with either 2-wire or 4-wire trunks with MF or DP (dial pulsing) signaling and with loop or E&M supervision. Other

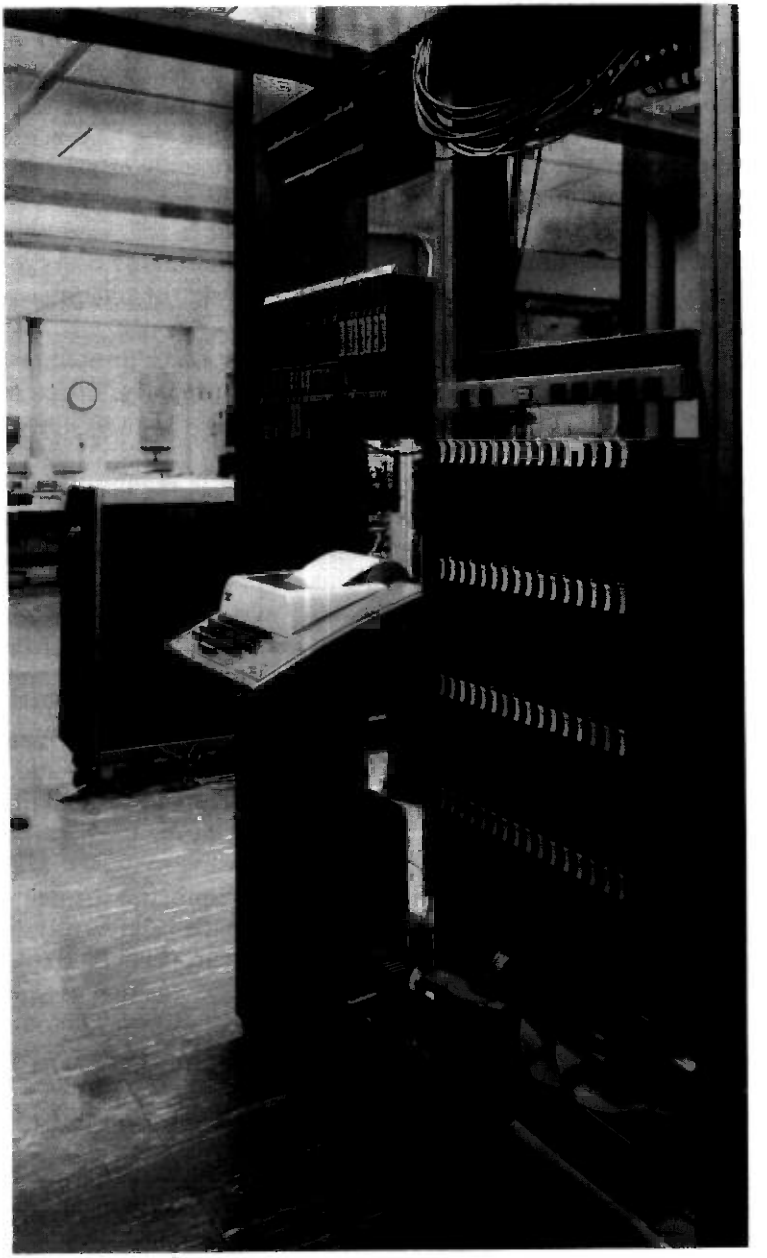


Fig. 3—Microprocessor Controlled Load Box.

trunk types can be easily handled by modifying the microcomputer program and providing the proper trunk interface.

(*v*) The call load can be as high as 10,000 calls per hour for short calls.

The user sets up his call load by entering the call parameters for each trunk via a teletypewriter or a terminal. Once the load is established, the calls can be started or stopped individually or as a group. The user can save his call load parameters on the support processor's disk via a serial data link (Section 3.1.2.4). This call load can be reloaded at a later time. This save/load capability allows users to set up a call load without entering the information via the teletypewriter each time.

Another feature of the MICLOB is the error messages printed on the TTY or terminal whenever calls do not proceed properly. These messages can alert the user to a problem with a trunk circuit or a MICLOB circuit.

In a testing environment where a heavy load is required, the ELB can be used to generate a background load of a particular call type. The MICLOB can be used to generate a load of special calls or of call types not possible with the ELB. In this manner, the TSPS programs can be exercised in the lab much like a live system.

3.3.3.3 Operator simulators. Operator simulators are used to simulate operator actions on calls arriving at positions. These simulators are used when a call load is generated containing calls which require operator assistance. The simulators recognize the call types arriving at the position and generate the required keying sequences. There are two types of operator simulators used in TSPS. The older simulator is hardwired. It can handle all call types except ACTS. Each of these simulators requires an actual position to operate.

The new Microprocessor Operator Position Simulator (MOPS) can handle all call types. Each call type can be handled differently, new call types can easily be added, and existing calls can easily be changed. This simulator can work with or without an actual position available. In this way, less positions are required in the laboratory. The advantage of having calls go to an actual position is that one can observe how the simulator is handling calls or one can manually handle a call if necessary. These functions can be duplicated via a terminal connected to the simulator.

The lamp and display orders sent to a simulator position by TSPS during a call can be converted to lamp names and digital displays and printed on the terminal. By monitoring a particular simulator position, the programmer can observe calls going to the position. The keys on the terminal are programmed to act as position keys so that the programmer can also manually handle a call if desired. The terminal can be remoted if necessary. This feature is useful when the TSPS

operator positions are not at the same location where program testing is being performed.

3.4 Trouble location manual (TLM) generation

The Trouble Location Manual (TLM) contains Trouble Location Numbers (TLN) for a circuit or subsystem. Associated with each TLN is the location of probable failing circuit packs in the circuit. These numbers which are an output from the diagnostic program are used to assist craft in fixing a faulty circuit.

The generation of a TLM for each circuit or subsystem requires that faults be physically inserted in these circuits. The diagnostic program for the circuit being faulted is run. The results from the diagnostic are used to generate the TLN. The collection of TLNs and faulted circuit locations make up the TLM.

Originally, in TSPS, most circuit pack faults could be inserted at the connector. However, with the introduction of RTA and the IGFET (Insulated Gate Field Effect Transistor) memory, circuit packs become much more complex. They contained many integrated circuits in Dual In-Line Packages (DIP). All necessary faults could not be inserted at the connector. To insert faults in the newer circuit packs and to speed up the TLM generation process, a minicomputer-controlled fault insertion technique was developed. This technique, which consists of fault insertion hardware and software, is discussed in the following sections.

3.4.1 Physical fault insertion

The type of faults inserted in the circuits are opens and shorts to desired voltage levels on the circuit pack connectors and the pins of the DIPs on the packs.

The circuit pack which is being faulted is a special version of the standard circuit pack in that the integrated circuits are socket-mounted. The circuit pack is mounted in a special extender board which is plugged into the circuit pack connector. The DIPs on the circuit pack have "daughterboards" inserted between them and their sockets. Both the extender board and "daughterboards" have relays which can be controlled to open or short the circuit pack connector pins and the DIP pins.

These relays are controlled by the support processor by setting selected points in the ESD. The fault insertion hardware decodes this information to operate the selected relays and thus insert the desired fault.

3.4.2 Diagnostic control and raw data accumulation

The support processor contains software to control automatic physical fault insertion and the SPC diagnostic programs. It also contains routines to gather the test results, or raw data, from these diagnostics.

These results are combined into a data base used to produce system TLMs. The support processor software designed to control these processes is called MCFIT (Minicomputer Controlled Fault Insertion Technique). The MCFIT program, together with special code in the SPC to interface with system diagnostics, make up the automatic physical fault insertion control software.

3.4.2.1 The MCFIT program. MCFIT was designed to allow rapid fault insertion and TLM generation. The majority of faults to be inserted are stored in a data base on the support processor's disk. This data base consists of all standard faults that are defined for each type of DIP used in the circuit being faulted. Thus, the user must only input the layout of the circuit pack to be faulted and any modifications to the standard fault list. MCFIT automatically controls the faulting of the entire pack, requiring manual intervention only to move the fault insertion hardware to the next pack.

Once the user has specified the pack layout, MCFIT retrieves the correct list of faults from disk and applies any necessary modifications. This list, together with the subsystem identification and circuit pack location, comprise the MCFIT "work file." MCFIT then executes a fault insertion program which sequentially inserts all faults appearing in the work file. This program controls the fault insertion hardware, activates a special SPC program which requests diagnostics on the specified subsystem unit, and records the failure data on the support processor's magnetic tape unit. It also prints summary data on the line printer. These data point out unexpected diagnostic ATPs (All Tests Passed) and inconsistent failure data for the user to examine.

3.4.2.2 SPC interface software. As mentioned above, the MCFIT program activates a special SPC program to request diagnostics. This SPC program is not part of the official generic, but is loaded into memory at the start of each fault insertion session in the lab. The diagnostic is requested through the NUPIC/SPI interface. An interrupt is generated in the SPC which transfers control to this diagnostic interface program. This program sets the appropriate bits in the diagnostic request words and status words corresponding to the subsystem unit being faulted. The diagnostic sequence proceeds normally in the SPC with one exception. Normally, each diagnostic transfers control to the SPC Diagnostic Output Control Program (DOCP) to print the pass/fail data on the maintenance teletypewriter. This special program, however, intercepts the data passed to DOCP and sends them to the support processor via the NUPIC/SPI.

IV. CONCLUSION

Effective software development depends very heavily on adequate development tools and test facilities. The tools and facilities described

in this paper came about through an evolutionary process. They started out in a much more basic form and were improved and expanded many times before they reached their present state. Thus, effective support software and hardware requires ongoing development. This point must be kept in mind so that programmer productivity can continue to improve.

V. ACKNOWLEDGMENTS

The support tools discussed in this paper are the combined effort of many people. Contributions to these tools were made by B. E. Holmes, J. R. Petty, and E. G. Pflaum in the area of program administration, by G. M. Jensen in the utility software, by M. R. Harder, P. L. Shepherd, and G. L. Taylor in the design of the various simulators, and by R. H. Allen and F. H. Ross in maintaining the utility and system laboratory hardware.

REFERENCES

1. G. R. Durney, H. W. Kettler, E. M. Prell, G. Riddell, and W. B. Rohn, "Stored Program Control No. 1A," B.S.T.J., *49*, No. 10 (December 1970).
2. M. Berger, J. C. Dalby, E. M. Prell and V. L. Ransom, "TSPS No. 1: Automated Coin Toll Service: Overall Description and Operational Characteristics," B.S.T.J., this issue, pp. 1207–1223.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July-August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

System Verification and Evaluation Procedures

By J. P. DELATORE, D. VAN HAFTEN, and L. A. WEBER

(Manuscript received December 11, 1978)

This paper describes the verification and evaluation procedures followed in the development of new features for the Traffic Service Position System (TSPS). Beginning with the definition of new feature requirements, these procedures are adhered to throughout the TSPS development cycle. Hardware and software designs are reviewed to verify that all requirements are met and to ensure that Bell System standards for reliability are maintained. Finally, both system laboratory and site testing are performed to verify the proper implementation of each new feature and to evaluate the overall performance of the TSPS.

I. INTRODUCTION

A significant part of the effort required for any switching system development involves determining whether the system performs properly. The ultimate judge of system performance is the user. In making this judgment, the user considers both system reliability and maintainability. Since Bell System standards for service are very high, measures of acceptable performance are rigorous. Thorough plans are made to ensure that switching systems provide this required high level of performance. These plans begin with the initial concept of a switching system or new feature and continue throughout the development process. Every design decision is based on providing the best possible service at the lowest possible cost. Each decision is evaluated for its impact on the customer who uses the switching system and on the operating company that must administer and maintain it.

This paper discusses the methods used to verify and evaluate system performance of the Traffic Service Position System (TSPS). It also

describes the facilities specifically designed to support the verification and evaluation process.

II. STAGES OF THE TSPS VERIFICATION AND EVALUATION PROCESS

The verification and evaluation process for new TSPS features involves all aspects of system development. This process begins with the determination of feature objectives and continues with the verification that the design requirements for each new feature are consistent with these objectives. Finally, the system is tested to ensure that the implementation reflects the design requirements. This process is divided into several stages. The stages of verification and evaluation discussed in this article are (*i*) requirements reviews, (*ii*) design reviews, (*iii*) circuit analysis or program reviews, (*iv*) unit testing, (*v*) integration testing, and (*vi*) system testing. Figure 1 shows these six

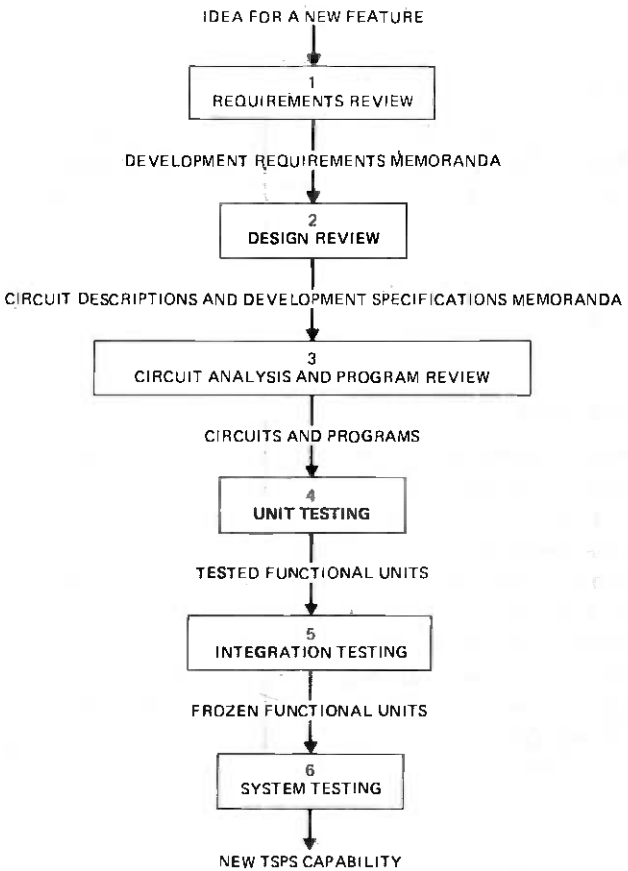


Fig. 1—Six stages in the TSPS verification and evaluation process.

stages and how the results of each stage are used in subsequent stages of the development.

The first stage of the verification and evaluation process begins when a new TSPS feature or system capability is proposed. Each new feature is analyzed to determine the service and maintenance requirements that must be satisfied. This analysis can be very extensive for some new features, such as the Automated Coin Toll Service (ACTS) feature. To perform this analysis for ACTS, an in-service TSPS was modified to simulate ACTS for a comprehensive human factors trial.[1] The results of this trial guided the formulation of the requirements for ACTS. These requirements are documented in development requirements memoranda and form the basis for detailed design and development.

The detailed design and development of a new feature is generally accomplished by partitioning the feature into functional units which are then assigned to development teams for implementation. Continued division of responsibility within a team for the development of a functional unit is dependent on the complexity of that unit. This partitioning applies to both hardware and software portions of the system.

The second stage of the TSPS verification and evaluation process begins when design reviews are held to determine if the conceptual aspects of each functional unit are consistent with the overall system requirements. Once the architectural design of the functional unit has been evaluated in this manner, the development team begins the detailed design of the functional unit. The design of each element in the functional unit is reviewed. Depending on the number and complexity of functional unit interfaces, members of other TSPS development teams may participate at this point in the process. After this review, development specifications memoranda and circuit descriptions are usually written documenting the detailed design.

The hardware development proceeds with the generation of circuit schematic diagrams and the specification of the circuit components used in constructing the circuit. The layout of the circuit components and their interconnections are specified, and prototype circuits are built. The prototype circuits are tested off-line from the system to ensure that they perform as required. Manufacturing tests are also generated for digital circuits using a logic simulation program.[2] After these tests and prototypes of the digital circuits are verified with a circuit pack tester, testing begins in the system laboratory. Once the circuit is verified to operate in accordance with the design intent, manufacturing and factory test information is transmitted to Western Electric. This activity is scheduled so that standard Western Electric-supplied hardware will be available when site testing begins.

The software development is done in parallel with the hardware development. After each program is written, a program review is conducted (*i*) to ensure that all development requirements are met; (*ii*) to verify that the interfaces between programs are correct; and (*iii*) to check that each program instruction is correct. Problems identified and corrected during this program review require much less effort than if they were left to be resolved during later stages of the verification and evaluation process.

The fourth stage in this process is the formulation of functional tests which utilize special test facilities (see Section III) and which ensure that each functional unit meets the design requirements. Upon the completion of unit testing, the fifth stage of the verification and evaluation process occurs when integration testing is performed. Integration tests ensure that each functional unit performs in a total system environment. Specifically, interactions between functional units and hardware-software interfaces are emphasized. Once the integration tests have been successfully run, the functional unit designs are considered frozen, in that changes to a functional unit can only be made on a more formal and controlled basis. The frozen functional units are now ready for system testing, the sixth and last stage in the verification and evaluation process.

During system testing, additional functional tests are written, and the total set of functional tests are performed by an independent group to ensure that any biases held by the design engineers are not reflected in the interpretation of test results. Beginning with this phase of TSPS testing, every change introduced into the system—hardware or software—must be initiated by a trouble report which specifies the seriousness of the problem so that corrections can be generated on a priority basis. Changes must be submitted in a formal manner and approved by a special committee called the change review committee.

Beginning with the system testing phase of the verification and evaluation process, each change is tested incrementally. That is, the correction to each problem, when possible, is considered an independent entity. Each correction can be rejected if not deemed appropriate by either the change review committee or the system testers. This procedure is used (*i*) to provide a high degree of visibility to all changes being introduced into the system, (*ii*) to provide a procedure whereby each change is individually and thoroughly tested, and (*iii*) to provide a procedure whereby design engineers can make changes without repeatedly going through the integration process.

To this end, each change submitted by the design engineer must be accompanied by a very specific test procedure which has been verified before the change is submitted. System testers then take great care to ensure that the test procedure is appropriate and that the change does not invalidate a previously verified functional capability. In some

cases, this can only be done by repeating a long series of previously completed functional tests. This verification and evaluation process is greatly enhanced by a comprehensive set of test facilities discussed in the following section. These test facilities are available to both designers and system testers.

III. SYSTEM LABORATORY TESTING

TSPS verification and evaluation is done at Indian Hill using two system test facilities. Each test facility contains a TSPS and its associated peripheral subsystems (i.e., a Position Subsystem No. 1,[3] Position Subsystem No. 2,[4] Remote Trunk Arrangement,[4] Station Signaling and Announcement Subsystem[5]—see Fig. 2). Minicomputers and var-

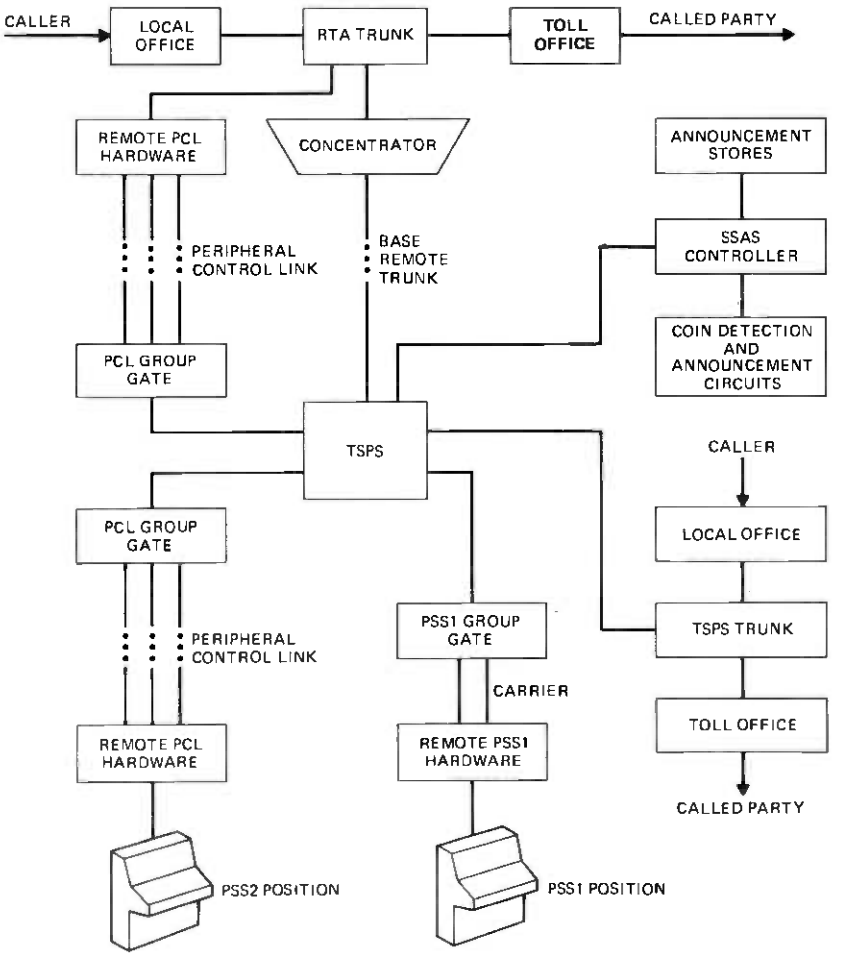


Fig. 2—Traffic Service Position System No. 1.

ious microprocessor-controlled circuits are utilized to produce the desired hardware and software environment.[6] In this environment, the system laboratory user can control and monitor the operation of the TSPS. Figure 3 shows one of the TSPS system laboratories at Indian Hill.

One aspect of controlling the TSPS operation is to establish the desired hardware and software configuration. The minicomputer provides three major capabilities for reconfiguring or modifying the system. First, the minicomputer provides the capability to load the TSPS with a new program, with the system's program stores being automatically reconfigured to match the memory requirements of the new program. Second, portions of the TSPS program can be modified with changes assembled by the minicomputer. This capability is used extensively in the incremental testing of modifications to frozen programs. Third, automatic fault insertion by the minicomputer is used to exercise system diagnostics in the preparation of trouble locating manuals. These manuals are used in conjunction with diagnostic outputs to identify hardware faults. The automatic process by which these faults are inserted helps to increase trouble locating manual resolution by allowing efficient generation of a large number of sample diagnostic results.

Once the desired hardware and software configuration has been established, the minicomputer can then be used to monitor the operation of the TSPS. Data associated with a specific event or combination of events can be recorded and later retrieved by the minicomputer. When correlated with the normal system responses such as teletypewriter messages, those data can be used to resolve problems or to confirm the proper operation of the TSPS.

An important part of testing TSPS features is ensuring that system interactions involving customers and operators are correct. Many inputs processed by a TSPS result from these interactions. To provide similar inputs in the system laboratory, stimuli comparable to those generated by customers and operators are produced using system utilities. A microprocessor-controlled facility is provided to simultaneously generate calls from many trunks. Both the traffic-handling characteristics of each trunk and call types can be changed under program control. Facilities are also provided to simulate the local and toll offices associated with a single TSPS trunk, thus allowing a laboratory user to completely control all stages of an individual call. Microprocessor-controlled operator simulators are provided to automatically handle calls at TSPS operator positions. In addition, calls can be handled manually at an operator position, thereby allowing the system laboratory user to test unexpected operator sequences.

System testing is divided into functional testing and system evalu-

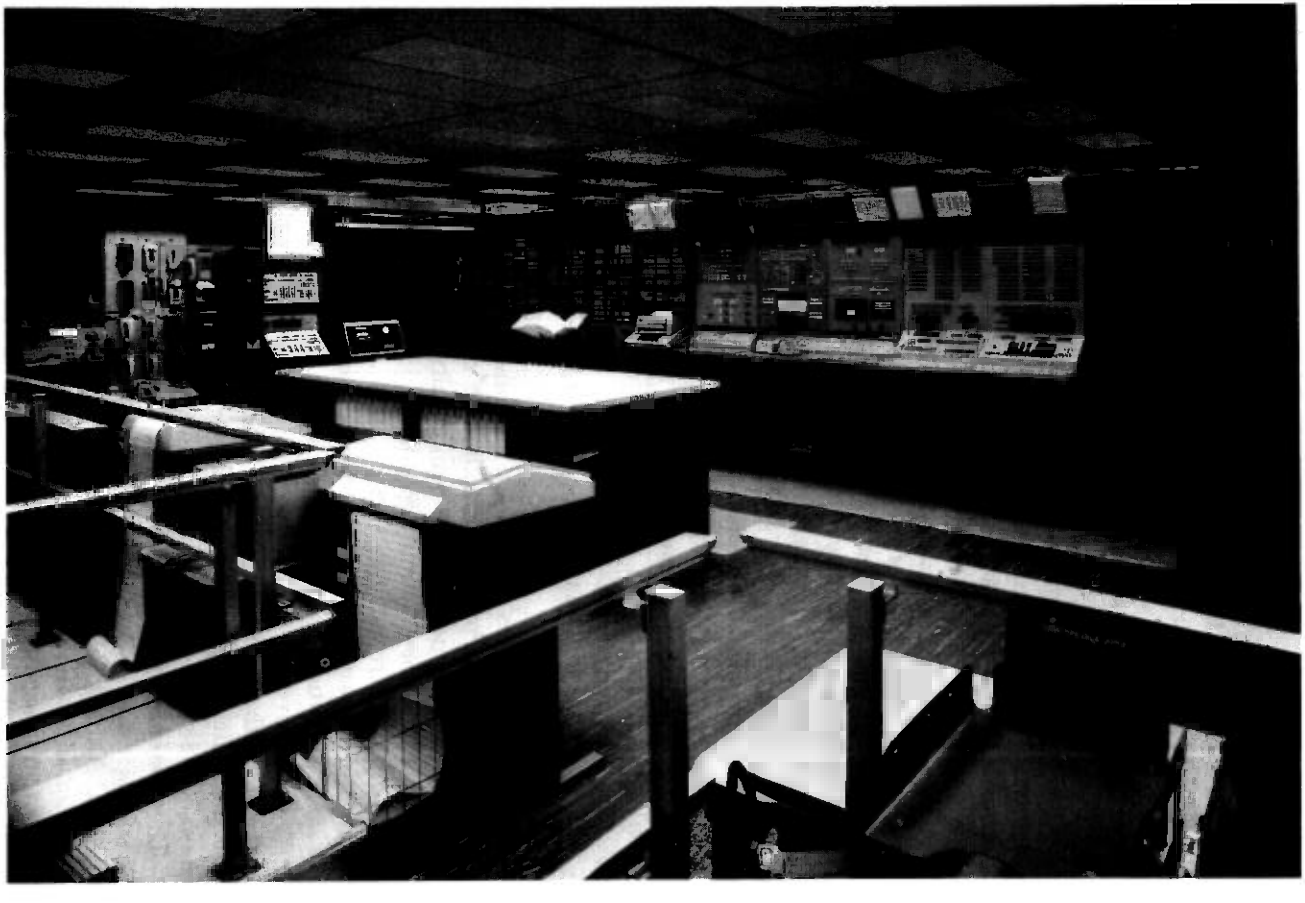


Fig. 3—The TSPS System Laboratory.

ation. Ideally, it would be desirable to write and perform functional tests in the system laboratory for each call processing and maintenance situation the TSPS will encounter. In practice, however, this is impossible. Consequently, during system testing, additional effort is required above and beyond the analysis of functional test results. This effort is referred to as system evaluation. In terms of cause and effect, functional testing involves setting up a prescribed set of initial conditions and then determining whether or not the proper response occurs. However, system evaluation involves observing every improper system response and then determining the cause of that response. This can be a difficult task, which at times is more of an art than a science. Reproducibility is a primary requirement for identifying the cause of a problem. Problems discovered during functional testing are generally reproducible, since a set of initial conditions have been specified for running each functional test. However, problems encountered during system evaluation frequently do not meet this reproducibility criterion. As a result, much more analysis of these types of problems is required.

During system evaluation, several indicators are used to determine that a problem exists. These indicators are:

(*i*) Unexpected teletypewriter output messages.
(*ii*) Loss of service of a hardware unit to the system for no apparently valid reason.
(*iii*) Unexplainable maintenance or call processing activity.

The audit messages[7] printed on the teletypewriter are a primary indication of system problems. Each audit message generally signifies the presence of some program error resulting in an inconsistency in unprotected memory. This memory is continuously updated by different programs to reflect the current state of the system. A detailed analysis of the audit messages will sometimes indicate what caused the error condition. The debugging capabilities of the minicomputer described above are particularly helpful in resolving this type of problem. With these capabilities, system activity which occurred before the audit program detected the error can be analyzed, and the cause of the problem can be identified.

The system laboratory provides a controlled environment where interfaces external to the TSPS have been simulated with system utilities. Although it is possible to test most aspects of the TSPS operation in this environment, increased confidence is built when new features are tested at a newly installed TSPS which interfaces with actual local and toll offices. For these and other reasons, the verification and evaluation process is continued at a test site.

IV. THE SITE TESTING INTERVAL

Each new generic is tested at a TSPS that has not been cut into service. This TSPS has been fully engineered by operating company personnel, and the hardware has been installed by Western Electric during a normal installation interval. Bell Laboratories testing of a generic at this pre-cutover TSPS evaluates the new features in a fully equipped TSPS. It also verifies the accuracy of the information provided to Western Electric and the operating companies on these new features. Testing is done with both local and toll offices to ensure that no interface problems exist. The length of the site test interval varies depending on the number of features being tested, amount and complexity of the new hardware, and size of the program change. Table I summarizes the test site, major new features, and size of the program change for each of the recent TSPS generics.

Testing at the site involves reverifying specific operational and maintenance capabilities for all new features, with the objective of determining whether or not the requirements for each feature are met. In particular, testing focuses on verifying interfaces with local and toll offices. Regression testing is also performed to ensure that previous TSPS capabilities are not adversely affected by the new features. Since site testing is performed after new feature development has been completed, problems identified at the test site are generally more subtle than those previously uncovered during system laboratory testing at Indian Hill. Furthermore, by the beginning of the site test interval, the hardware design has already been proven in the system laboratory. As a result, during this interval the majority of problems identified are in the software. Most are not due to any significant design problems and are easily corrected.

Both functional testing and regression testing at the site are done in a systematic manner; a specific set of tests are performed for each

Table I—Recent TSPS generics

Generic	Test Site	Major New Feature(s)	Size of Program Change (40-bit Words)
Generic 7, Issue 1	Syracuse, New York	Remote Trunk Arrangement, Position Subsystem No. 2	60,000
Generic 7, Issue 2	Saginaw, Michigan	Selective Call Screening, more than twenty stores on a bus	5,000
Generic 8, Issue 1	Phoenix, Arizona	Automated Coin Toll Service	40,000
Generic 8, Issue 2	Montgomery, Alabama	Automated Coin Toll Service with Remote Trunk Arrangement	7,000

feature, the results are observed, and these results are analyzed. The amount of detail with which functional test results are analyzed at the test site is minimal. Instead, any functional tests which fail are documented in trouble reports which are then sent back to the developers at Indian Hill. Extensive analysis at the test site is restricted to problems uncovered during system evaluation. In general, the exact conditions required to bring out these types of problems must be determined at the test site.

The combination of system laboratory testing and site testing during this interval complement each other in the verification and evaluation of a TSPS generic. In the system laboratory at Indian Hill, emphasis is placed on testing individual changes being made to correct specific problems. However, as indicated above, effort at the test site is oriented toward verifying and evaluating functional capabilities rather than testing changes or corrections.

The test facilities provided at the major test sites are comparable to those permanently installed in the system laboratories at Indian Hill. These facilities include the minicomputer, call generation capabilities, and operator simulators described in Section III. Additional capabilities are also provided to remotely control these test facilities so that testing can be done from one location.

The number of problems identified from the time the hardware and software were frozen until the completion of site testing is shown in Table II for each of the recent TSPS generics. These totals are broken down into: (*i*) problems identified at Indian Hill in the system laboratory and (*ii*) problems identified at the test site.

At the completion of the site testing interval, responsibility for monitoring system performance is given to TSPS field support personnel. In addition to continuing the evaluation of TSPS and new feature performance, field support personnel are responsible for updating the system with any necessary changes. To assist in these efforts, the TSPS's maintenance teletypewriter output can be transmitted to the TSPS Diagnostic Center at Indian Hill. Problems detected from this output or reported by the operating company are analyzed, and cor-

Table II—Trouble reports written through the completion of site testing

Generic	TRS Written at the Base Location	TRS Written at the Test Site	Total
Generic 7, Issue 1	600	615	1215
Generic 7, Issue 2	33	20	53
Generic 8, Issue 1	565	800	1365
Generic 8, Issue 2	108	58	166

rections are generated by designers working in conjunction with field support personnel.

V. SYSTEM CAPACITY

The addition of each new feature to TSPS has an impact on the real-time capacity of the TSPS. Overall system real-time capacity varies for each TSPS installation based upon the hardware configuration used and the particular call mix being processed. Each new feature is evaluated during the development cycle to determine its impact on system capacity. This evaluation is verified at the first in-service office or the first office close enough to capacity for a verification to be made. The real-time capacity estimation and verification process is extremely important due to its effect on the long-range planning of the operating companies. To assist the operating companies in their analysis of an individual office's real-time capacity, a program called TSPSCAP is available. This program runs on an off-line computer and is updated with each generic to reflect the addition of new features.

VI. SUMMARY

System verification and evaluation is a process that is interwoven with all aspects of the development of new TSPS features. From the inception of the idea for a new feature, development requirements are evaluated to ensure a proper understanding of the proposed capabilities and to verify that the proposed design will provide these capabilities. After formal reviews, these requirements are specified in development requirements memoranda, and the design is specified in development specifications memoranda and circuit descriptions. Next the design is implemented and each functional unit is verified in the system laboratory. Before the commencement of site testing, both hardware and software are placed in a frozen mode, after which all changes to the system are verified through a formal procedure. Site testing is done in a pre-cutover TSPS engineered by an operating company with hardware supplied and installed by Western Electric. During site testing, functional tests are run to verify the proper implementation of all new features and regression tests are run to ensure the proper operation of previous TSPS capabilities. In addition, system evaluation is done before the site is cut into service, thus resolving many of the more subtle problems. Finally, an analysis of the real-time effect of each new feature is made after cutover to confirm the theoretical analysis made prior to cutover.

This verification and evaluation approach is followed for each new TSPS feature. Adherence to this methodology allows new TSPS generics

to be properly verified and evaluated, thereby insuring Bell System customers of the best possible service.

VII. ACKNOWLEDGMENTS

Verification and evaluation of new TSPS features are accomplished through the efforts of many people in the Operator Systems Laboratory. Due to the laboratory-wide scope of this process, a list of all contributors to its success would be very long. However, the authors would like to acknowledge the special contributions of the following people: J. W. Hopkins, R. C. Jackson, R. Kroning, D. J. Rak, B. D. Reh, M. J. Traube, and D. A. Yackley.

REFERENCES

1. E. A. Youngs, W. J. Bushnell, and A. Barone-Wing, "TSPS No. 1: ACTS: Human Factors Studies," B. S. T. J., this issue, pp. 1291–1305.
2. H. Y. Chang, G. W. Smith, and R. B. Walford, "LAMP: System Description," B. S. T. J., *53*, No. 8 (October 1974), pp. 1431–1449.
3. W. K. Comella, C. M. Day, Jr., and J. A. Hackett, "TSPS No. 1: Peripheral Circuits," B. S. T. J., *49*, No. 10 (December 1970), pp. 2561–2623.
4. R. J. Jaeger, R. S. DiPietro, and S. M. Bauman, "TSPS No. 1: Remote Trunk Arrangement: Overall Description and Operational Characteristics," B. S. T. J., this issue, pp. 1119–1135.
5. M. Berger, J. C. Dalby, E. M. Prell, and V. L. Ransom, "TSPS No. 1: Automated Coin Toll Service: Overall Description and Operational Characteristics," B. S. T. J., this issue, pp. 1207–1223.
6. J. J. Stanaway, J. J. Victor, and R. J. Welsch, "TSPS No. 1: Software Development Tools," B. S. T. J., this issue, pp. 1307–1333.
7. A. W. Kettley, E. J. Pasternak, and M. F. Sikorsky, "TSPS No. 1: Operational Programs," B. S. T. J., *49*, No. 10 (December 1970), pp. 2625–2683.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Traffic Service Position System No. 1:

Operator Training Facilities

By G. RIDDELL, C. R. SWANSON, and R. T. STEINBRENNER

(Manuscript received December 11, 1978)

The evolution of operator training facilities for TSPS No. 1 has involved the development of new program-controlled positions which have much greater flexibility and reliability than preceding training equipments. These PROCON-controlled positions and associated facilities are now the standard means for training TSPS operators. Coincident with the development of these facilities, a minicomputer system for generating master training tapes was designed to facilitate the generating of new training tapes and modifying existing ones to add new operating features.

Since the beginning of telephony, there has been a need to train switchboard operators in the procedures required to handle and complete telephone calls. These procedures started with "on the job" training and gradually improved over the years to specialized training facilities to give the trainees the ability to handle calls before they sat down at a switchboard handling traffic.

With the advent of the crossbar tandem Traffic Service Position in 1961, a new specialized training facility (100A trainer) was designed to train operators in the basics of handling traffic on this new type of switchboard (Fig. 1). This consisted of an operator position with two equipment cabinets. One of the two cabinets contained a paper tape reader, a magnetic tape player, power supplies, and racks for tapes. The second cabinet contained densely packaged electromechanical and solid-state circuitry which performed the necessary trainer functions under control of information on the magnetic and paper tapes.

Customer calls were simulated by verbal passages on the magnetic tapes which also provided synchronizing tones to start the paper tape reader. The paper tape reader and the magnetic tape player were

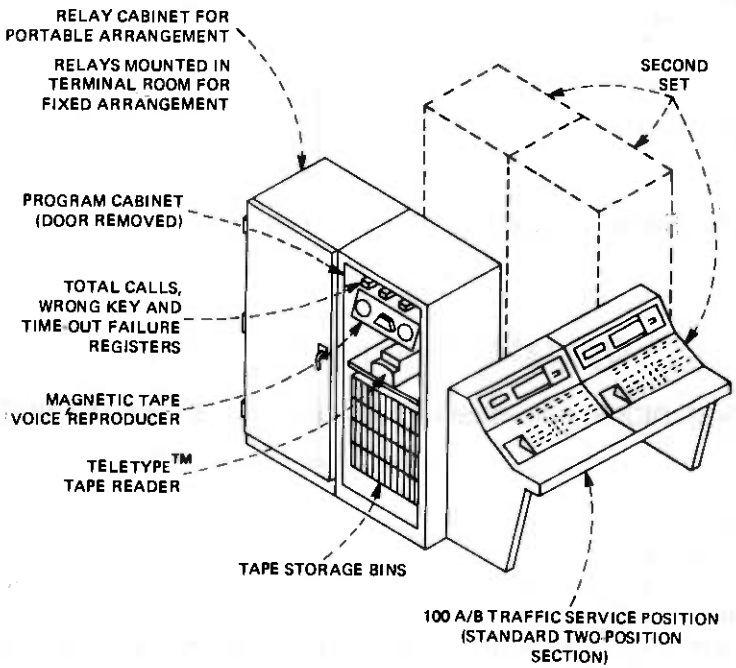


Fig. 1—Operator training equipment.

additionally controlled by the trainee sitting at the training position responding to simulated calls.

In effect, the training system was a simulator which could duplicate nearly all types of calls handled by a TSP. Different types of calls were simulated by different input tapes which were loaded by supervisory personnel.

With the development of TSPS No. 1, this 100A system was modified to provide added capabilities and was classified as the 100B trainer. Essentially, however, the basic operating principles and the physical appearance of the new trainer remained the same. This system has been used to train operators on TSPS call-handling procedures since its first application in 1969. However, as new features have been added to the TSPS, corresponding equipment modifications have been required in the relay circuitry. As there are over 2000 of these trainers in the field, this is a significant problem. In addition to this, the design of these changes becomes more difficult as control logic becomes more complex and as precise operating characteristics of the trainers is required.

To overcome these problems, a new programmable 100C trainer system was designed. Programmability enables new TSPS features to be introduced to the simulator quickly with few or no hardware

modifications. This system also provides a supervisory function not available on the 100B trainer. Furthermore, it uses standard 100C operator positions which can also handle regular traffic when not used as trainers.

This new trainer system shown in Fig. 2 consists of a maximum of eight training positions and one supervisory position. Each position can be switched to handle normal traffic. These positions can also be dedicated for training only and supplied on a stand-alone basis requiring no association with any chief operator group or TSPS No. 1 System. A photograph of a training position is shown in Fig. 3.

A standard 100C position is converted to a trainer by a wiring option added to the existing local cable wiring, and an applique column (Fig. 4) added to the rear of the position. This column contains a microprocessor called the programmable controller (PROCON) (Fig. 5), which provides the necessary control functions for the trainer. In addition to this, translator and temporary memory boards, voice and tone detectors, tone generators, I/O logic control circuitry, diagnostic indicators,

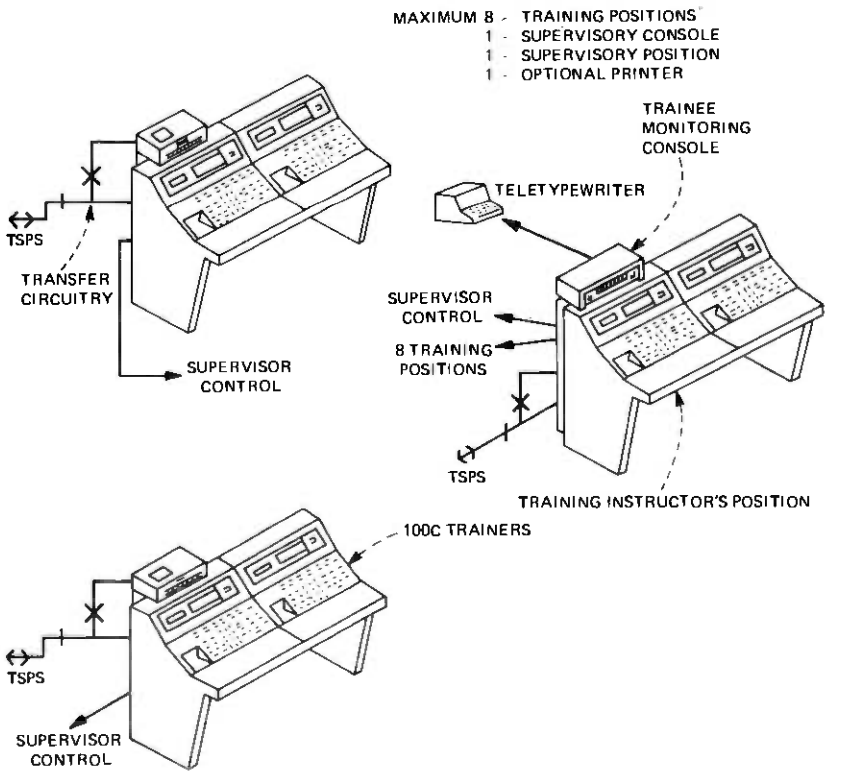


Fig. 2—100C training system.

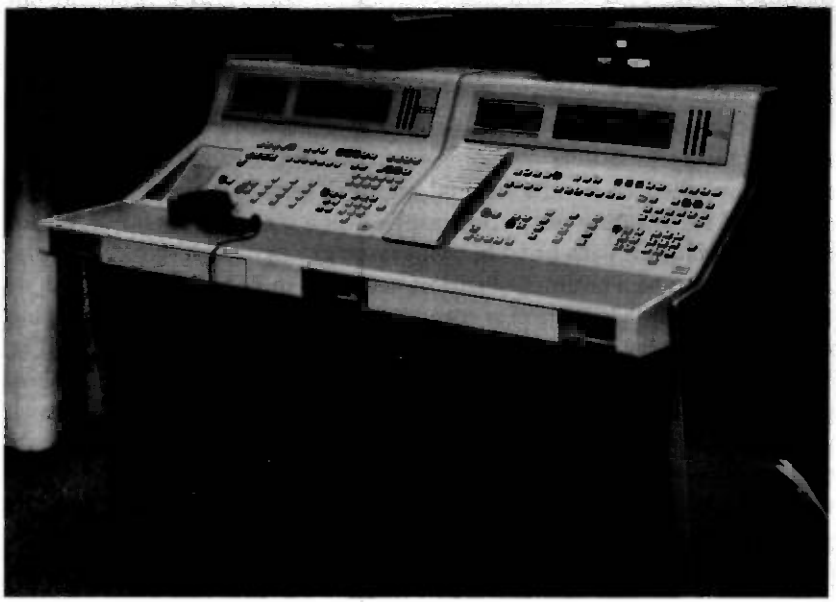


Fig. 3—100C training position.

switching circuitry, supervisory interface control, and isolation circuitry are all contained in the column (Fig. 6).

To provide the necessary combined data and voice inputs to the new trainer, a special cassette tape reproducer was required. This reproducer is largely based on conventional technology in the audio/visual and training fields. The prerecorded tape is in the standard Phillips cassette stereo configuration. However, several special features were required in this application:

(*i*) A precise high speed reverse to enable the trainee to return to the beginning of a call pattern on the tape (called the "repeat" mode).
(*ii*) Remote and local control functions customized for TSPS.
(*iii*) Power supplied to external interface circuitry.
(*iv*) Enclosure designed and stylized for a mounting location on the position accessible to the trainee.

The reproducer is manufactured for the Bell System by an outside supplier. The prerecorded cassette training tape contains the voice and data associated with a maximum of 30 minutes of simulated calls. The tape contains two tracks, one for voice and the other for control tones (Fig. 7). In a position being used for training, the PROCON processes the cassette control tones, operator-keyed actions, and verbal responses to simulate actual calls. The operation of the player is under

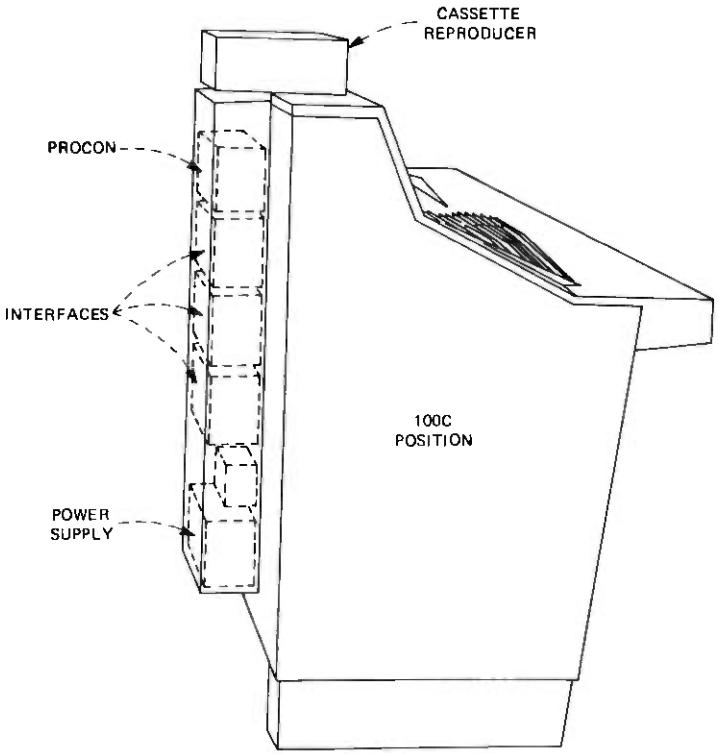


Fig. 4—Conversion of position to trainer.

control of the PROCON, although the trainee has some manual override capability. The trainee now has the ability to load the cassette and initiate and pace the training session, which was not possible in the previous system.

The control of the various components is shown in Fig. 8. The trainee inserts a prerecorded cassette in the reproducer and operates the play button. The cassette advances beyond a point where data, representing a TSPS lamp display, has been presented to the PROCON. A proper response by the trainee (either a key operation and/or verbal response) causes the PROCON to continue reading the tape.

The control data on the cassette are formatted in a 5-bit baudot code representation. Each bit is represented by a different frequency on the tape. A coded 5-bit character is represented by combinations of one to four frequencies simultaneously. These frequencies are present for 80 ms as the tape advances at 1⅞ inches per second (Fig. 7). A 20-ms interval is provided between consecutive characters as the tape advances. An instruction to the trainer generally is provided by two

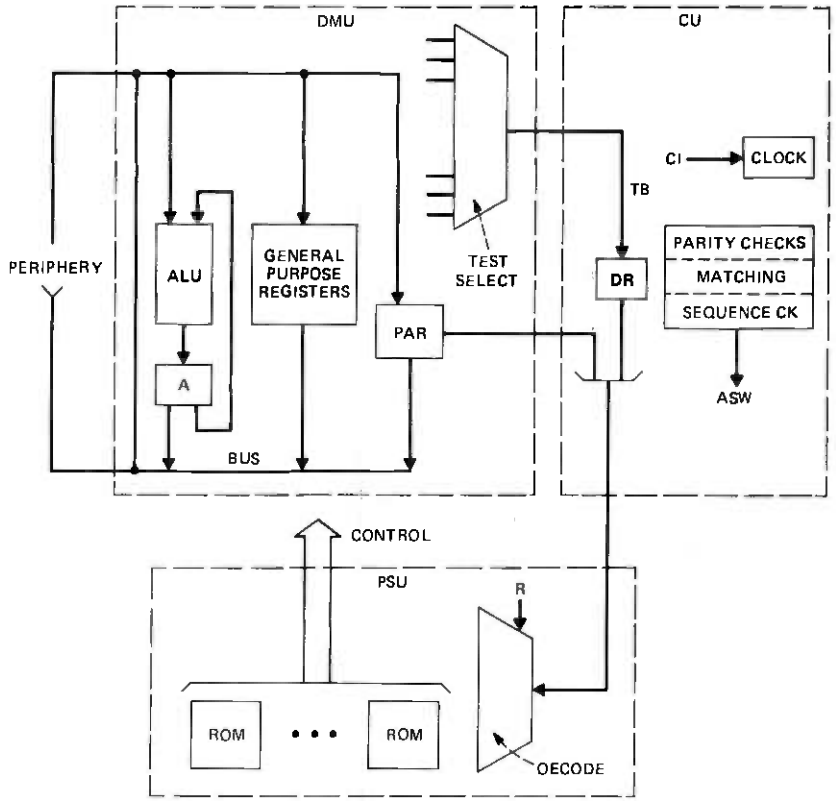


Fig. 5—Architectural overview of PROCON.

successive baudot characters although there are exceptions requiring only one. The data inputs are paired by PROCON and translated into a control function. For digital displays, telephone numbers are stored in the translator memory. There are standardized numbers for all trainees to use regardless of their location in the country. With this technique, a 10- to 12-digit display requires only a 2-character code. In addition, numbers keyed by the operator are stored in temporary memory. Both types of numbers can be displayed, when required, by operation of the corresponding display key at the operator's console. Other digital displays, for example, time and charges, are pre-coded and can be displayed on demand.

Operation of the time display key will produce a time-of-day display. When a position is powered up for the training function, the time is initialized to 8 o'clock. If desired, this can be changed by keying in the desired time on the keyset.

Operation of a key on the console generates one of the 84 possible ⅜

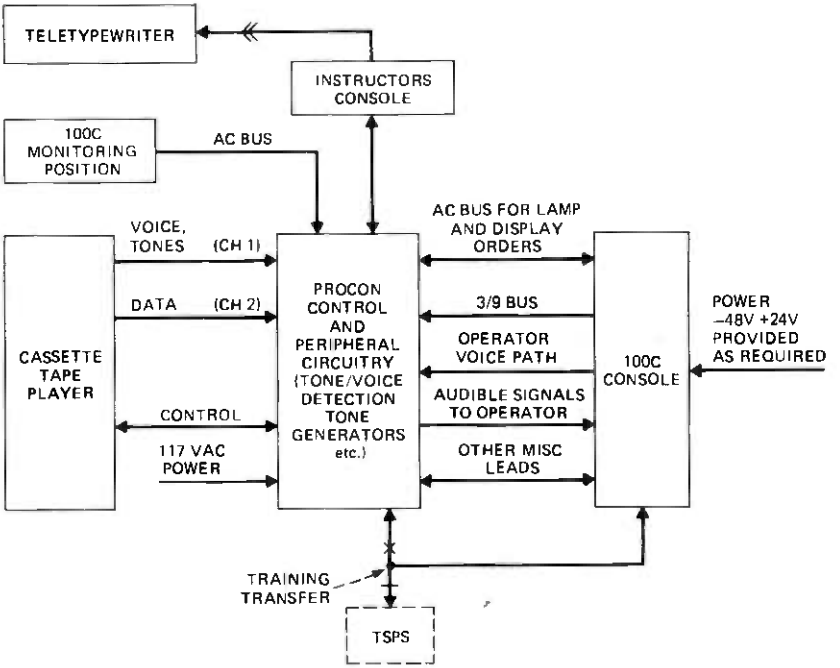


Fig. 6—100C operator training system.

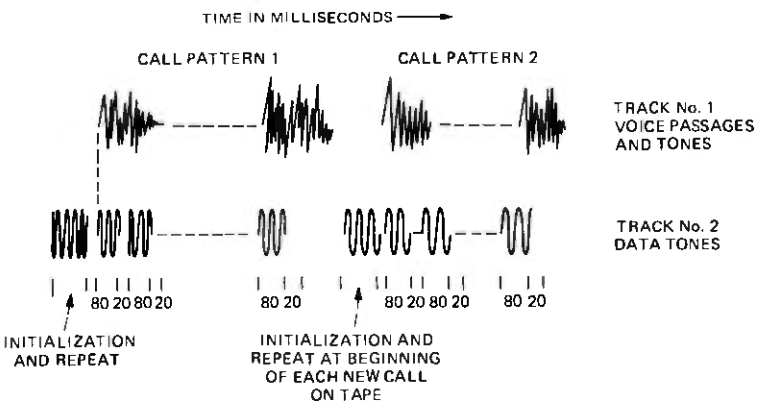


Fig. 7—Training tape information content.

key codes. These ⅜ codes are identified by PROCON through its associated position interface circuit, and appropriate action is taken to control the cassette player, and light, flash, or extinguish lamps on the trainer console.

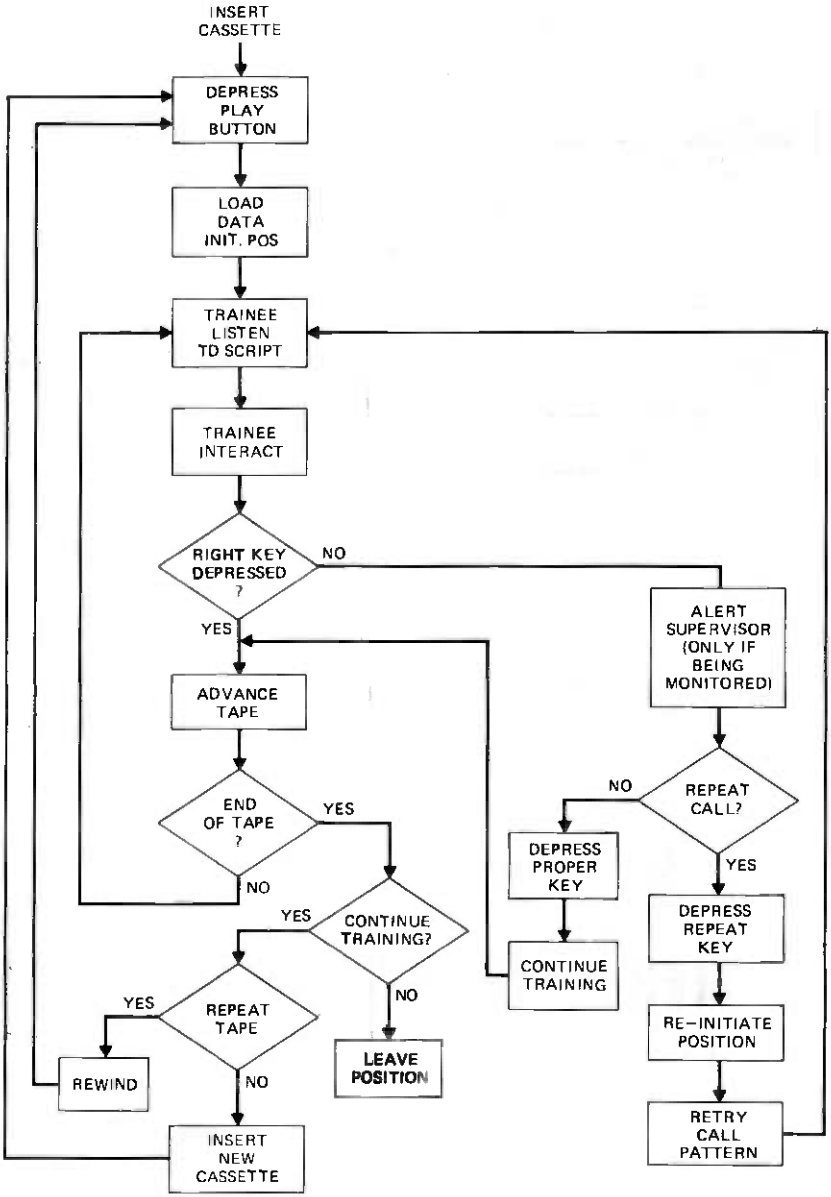


Fig. 8—100C trainer operations.

Verbal responses of the trainee are detected by a voice detector and forwarded to PROCON, which permits the call to progress to the next voice passage or group of data on the cassette tape.

In addition to lamps and display data, the cassette tape data also

control tone signals (ziptone, audible ring, busy) to the trainee. These tones can be present for indeterminate time periods. For these types of signals, tone generators are provided under PROCON control. In general, these tones are initiated or terminated by the trainee key actions.

The standard 100C operator position requires a 24-bit serial data word for each instruction to light a lamp or display pairs of numerical digits. In the 100C trainer, these 24-bit words are identical to those used on the standard 100C position but are generated by PROCON. As the trainee operates keys at the console, the 3% codes generated are translated and matched against expected results. If the correct response is made, the cassette is advanced to present new data. If a match is not made, the trainer does not respond, forcing the trainee to question his/her actions and give the correct responses. A flowchart of the operations of the 100C trainer is shown in Fig. 8.

The 100C operator training system also includes an instructor's console with an associated *DATASPEED* ® 40 printer to provide a record of trainee keying actions. The instructor's console provides a means to supervise trainees while they are learning to handle TSPS traffic. It permits the instructor to monitor the simulated calls and the trainee's responses. As keys are operated by the trainee, the corresponding lamps light on the instructor's console. Incorrect key actions by the trainee cause the corresponding lamp to flash at the instructor's console.

The instructor can select any one of eight positions to monitor. While this one position is being monitored, all other seven positions are also being checked to determine if the trainees are making excessive errors in their responses. Excessive errors generated at a training position cause a corresponding position number lamp to flash at the instructor's console.

A *DATASPEED* 40 printer produces a record of the numbers keyed by the trainees. This can be done on a position-by-position basis by operation of a print key on the console. These records provide information on keying accuracy, as no actual matching of keyed numbers is provided in the 100C trainer programs.

These trainers also require support facilities to provide training tapes. The production of these training tapes require the coordinated efforts of Bell Laboratories, AT&T, Western Electric and general trade suppliers. Initially, the efforts of BTL and the AT&T operator services organization are used to determine the correct operator actions and console displays required on calls. Once this has been determined, data codes required for new features are designed by BTL, and the information is provided to the operator training groups at AT&T. The necessary data patterns are then determined by the operator training

group and entered into the data base of the minicomputer-controlled support system used to produce master training tapes.

To put both the data patterns and voice passages on master tapes, the support system called Automatic Data Entry Console (ADEC) was designed (Fig. 9), and two custom-built systems were provided to the AT&T training organization. This system has the capability of editing tapes and synchronizing the recording of voice and data to accurately simulate actual calls. It is updated and modified as required to provide new features. In order to produce master tapes, a narrative script is provided by the AT&T operator training organization to a professional recording studio which in turn produces a voice master tape. The voice master tape is used by the ADEC to generate the combined master tape consisting of voice passages and data tones on separate tracks. The generating of these tapes can call for several revisions to achieve realistic simulations (Fig. 10).

The combined master tape provided by this system is then transmitted to Western Electric for quality control. Commercial suppliers provide cassette copies of the master tapes for use by the telephone companies. Tight specifications on the generation of the master tape and on the cassette duplication process are required to provide adequate operating margins in the use of audio cassette technology for the basic input to the control system.

The use of the ADEC system has been beneficial in providing capa-

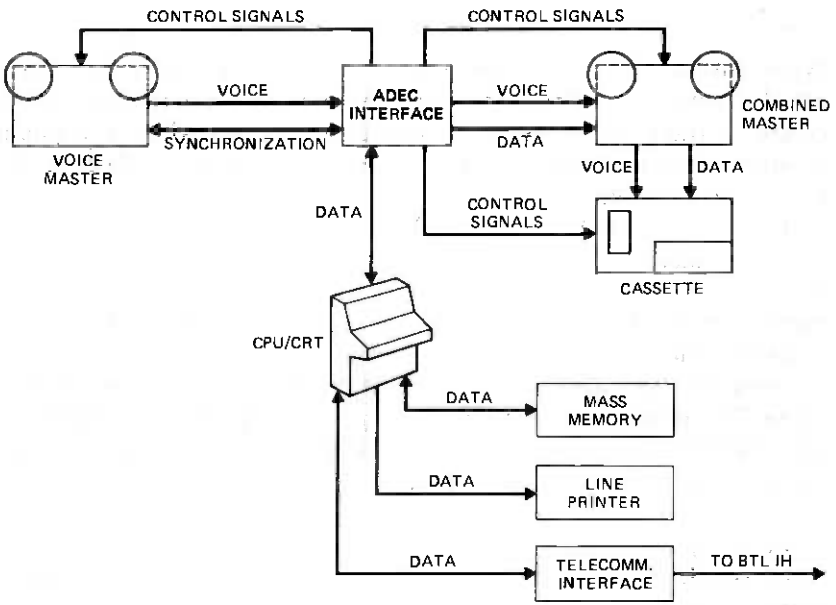


Fig. 9—ADEC system.

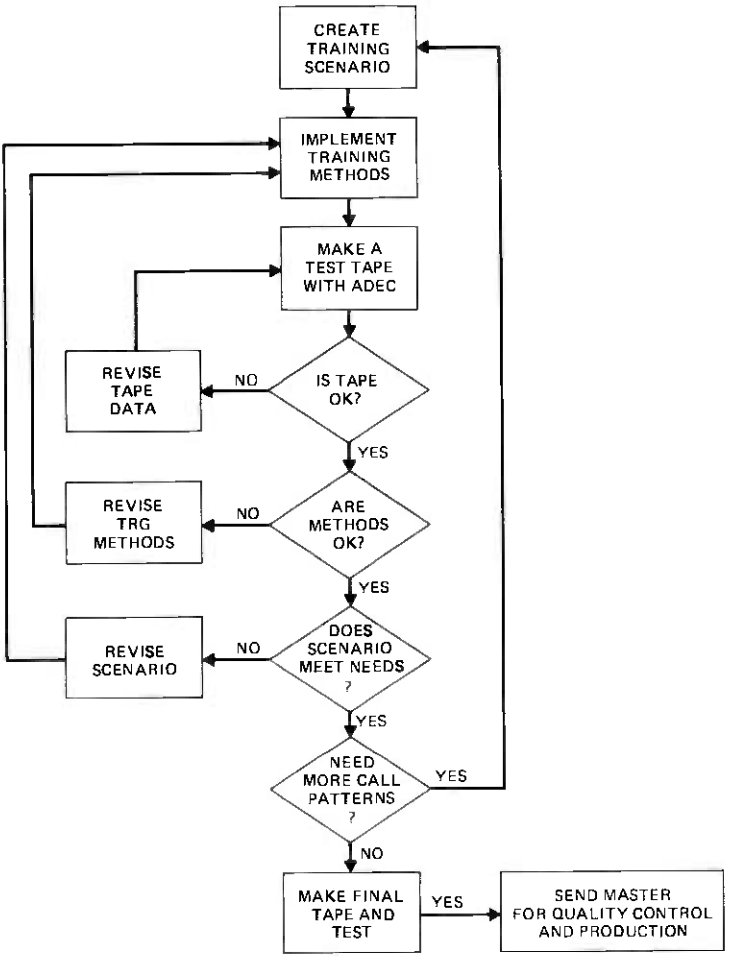


Fig. 10—Procedure for generating combined master tapes.

bilities for modifying and generating new training tapes not previously available with the older trainers. Requirements for training have been expanding continuously, and since the introduction of this system the production of training tapes has greatly improved.

ACKNOWLEDGMENTS

The evolution of Operator Training Systems was dependent upon many people. Of these, the efforts of R. D. Allen, T. W. Benko, R. S. Chiapparolli, Jr., D. F. Loftus, R. J. Radner, D. R. Rueckheim, M. A. Santasieri, R. Stearns, K. E. Vincent, L. Watstein, and others were especially valuable to the success of these systems.

Contributors to This Issue

R. Ahmari, B.S.E.E., 1966, University of Tehran; M.S.(E.E.), 1969, and Ph.D.(E.E.), 1972, Illinois Institute of Technology; Assistant Professor, Manhattan College, New York, 1972–1973; Bell Laboratories, 1973—. Mr. Ahmari has worked on system planning for private networks, design, and development of fault tolerant systems, and system testing for TSPS. He is currently a member of the SPC System Integration Group responsible for testing and integration of various TSPS generics. Mr. Ahmari is a registered professional engineer in the State of Illinois. Member, IEEE.

Ann Barone-Wing, B. A. (psychology), 1974, Mount Holyoke College; Bell Laboratories, 1974—. Ms. Barone-Wing initially worked on the Automated Coin Toll Service (ACTS) simulation study, designed the operator interface, and supervised the customer interviews. More recently, she has contributed to additional studies of operator service mechanization and other human factors projects.

Steven M. Bauman, B.E.E., 1966, Rensselaer Polytechnic Institute, M.E. (Electrical Engineering), 1968, Stevens Institute of Technology; Western Electric, 1966–1968; Bell Laboratories, 1968—. Mr. Bauman worked on Safeguard system design and software development until 1973. He then supervised a group responsible for call processing software for the RTA. In 1977, he was appointed Head, TSPS Feature Programming Department. Member, IEEE.

M. Berger, B.E.E., 1964, City University of New York; M.S.E.E., 1968, Polytechnic Institute of Brooklyn; Airborne Instruments Laboratory, 1964–67; Sperry Systems Management Division, 1967–69; New Jersey Bell 1972–74; Bell Laboratories, 1969–72, 1974—. Mr. Berger was initially involved in developing improved trunk engineering and administrative methods for metropolitan areas. He has been engaged in engineering and planning studies of new features of TSPS No. 1 and

supervised a group conducting planning studies of operator number services systems. At present, he is supervisor of the Operator Systems Engineering Group conducting planning studies and formulating requirements for operator toll and assistance systems. Member, IEEE.

William L. Brune, B.Sc. (Eng. Physics), 1947, Lehigh University; Bell Laboratories 1947–1955; Stromberg Carlson Telecommunications Division, 1955–1959; Bell Laboratories, 1959–. Mr. Brune's early responsibilities at Bell Laboratories were in transmission transformer development. At Stromberg Carlson, he was manager of groups engaged in transmission design and development of carrier multiplex and time division switches. Soon after his return to Bell Laboratories he became supervisor of the Networks, Filters and Equalizers group in the Military Apparatus Laboratory. Since 1970, he has supervised a group engaged in transmission planning for operator and attendant services. Member, Phi Beta Kappa.

Andrew F. Bulfer, B.S. (Electrical Engineering), 1962, Massachusetts Institute of Technology; M.S. and Ph.D. (Electrical Engineering), 1970, Ohio State University; Applied Physics Laboratory, Johns Hopkins University, 1962–1967; Bell Laboratories, 1970—. While at the Applied Physics Laboratory, Mr. Bulfer worked on the analysis and design of naval guided missile systems. Since coming to Bell Laboratories, Mr. Bulfer has worked in the Operator Systems Laboratory on various TSPS No. 1 new feature projects including CATLAS, RTA, PSS No. 2, and ACTS. He is presently supervisor of the Call Programming and Analysis group.

William J. Bushnell, B.S.(E.E.), 1969, Purdue University; M.S.(E.E.), 1971, Polytechnic Institute of Brooklyn; Ph.D. (E.E.), 1975, Polytechnic Institute of New York; Bell Laboratories, 1969—. Mr. Bushnell initially worked on radar system analysis for the Safeguard project. Since 1974, he has been involved in the formulation of requirements and design of ACTS, and in 1977 he began supervising system testing of ACTS and development of other TSPS features. Member, IEEE, Tau Beta Pi, Eta Kappa Nu.

G. T. Clark, B.S.M.E., 1952, Bradley University; Western Electric Company, 1956–1961; Bell Telephone Laboratories, 1961—. Mr. Clark was first engaged in the physical design of step-by-step common

control equipment and later worked on the design of 758C PBX equipment. He has coordinated the physical design of TSPS No. 1 equipment, including detail design of network, position subsystem, TTY trunk and buffer, and the station signaling and announcement subsystem switching equipment. He was also engaged in the physical design of AIS equipment, including the file subsystem No. 2. He is currently working on physical design of new features for TSPS No. 1, including the peripheral system interface frame.

John C. Dalby, Jr., B.S. (Applied Mathematics), 1968, and M.S.E. (Computer, Information, and Control Engineering), 1969, University of Michigan; Masters of Philosophy (Computer Science), 1977, Rutgers University; Bell Laboratories, 1970—. Since joining Bell Laboratories, Mr. Dalby has been involved in TSPS No. 1 development designing call processing and maintenance software and writing system development requirements for new TSPS No. 1 features. Presently, he supervises a group responsible for developing software, hardware, and procedures to add equipment and features to in-service TSPS No. 1 sites without interrupting call processing. Member, Tau Beta Pi.

J. P. Delatore, B. A. (Mathematics), 1963, College of Steubenville; M. A. (Mathematics), 1965, Bowling Green University; Bell Laboratories, 1965—. Mr. Delatore has worked on TSPS program design and TSPS test and evaluation. He worked at AT&T from 1973 to 1975 providing computer-aided service cost methodologies. In 1975 he became supervisor of the TSPS growth and field support group, and is presently supervisor of the TSPS planning group.

Richard S. DiPietro, B.S. (Eng. Sci.), 1970, Northwestern University; M.S.E.E., 1972, New York University; Bell Laboratories, 1970—. Mr. DiPietro worked on Safeguard system performance evaluation and system testing until 1974. He then worked on circuit design and system testing for the remote trunk arrangement. In 1977, he was made supervisor of the TSPS Field Support and Test Group and is currently supervisor of the Call Programming and Operator Actions Group.

Walter E. Gibbons, S.B. (Electrical Engineering), 1974, and S.M. (Electrical Engineering), 1975, Massachusetts Institute of Technology; Bell Laboratories, summer 1972, 1973—. Mr. Gibbons was initially

involved in the exploratory development of a model for the ACTS feature. Subsequently, he has done maintenance programming and system testing for RTA/PSS No. 2 development. He is presently working for the TSPS Processor Department. Member, Sigma Xi.

John A. Hackett, B.S.E.E., 1959, University of Maine; M.S.E.E., 1961, New York University; Bell Laboratories, summer 1958, 1959—. Mr. Hackett initially worked on improvements and call-through testing for step-by-step offices. He then worked on the initial circuit designs and overall architecture of TSPS No. 1. He supervised development of TSPS peripheral circuits and more recent additions to extend TSPS to sparse areas. He is currently supervisor of the operator systems exploratory group. Member, IEEE, Tau Beta Pi, Phi Kappa Phi.

Walter S. Hayward, Jr., A.B., 1943, S.M. (Electrical Engineering), 1947, Harvard University; Bell Laboratories, 1947—. Mr. Hayward has worked in the field of telephone traffic and switching systems engineering. In 1961, he was appointed Head, Electronic Switching Studies Department. In 1964, he was appointed Director of the Traffic Studies Center and is now Director of the Switching Operations Systems Engineering Center engaged in teletraffic studies, billing systems engineering and network data systems engineering. Member, IEEE and ACM.

J. Carl Hsu, B.S.E.E., 1963, National Taiwan University; M.S., 1966; Ph.D. (Computer Science), 1971, UCLA; Senior Development Engineer, NCR, 1966–1970; Bell Laboratories, 1971—. At Bell Laboratories, Mr. Hsu has been involved with call processing and maintenance software design, circuit design, and system planning for TSPS. He is currently supervisor of the TSPS Maintenance Emulation Group.

R. J. Jaeger, Jr., B.A. (Math), 1951, Hofstra University, Bell Laboratories, 1951—. Mr. Jaeger started his Bell System career with the Long Lines Department in New York City in 1941. After serving as a naval aviator in World War II, he returned to Long Lines, and in 1951 he came to Bell Laboratories to work in switching system design. He has worked on No. 4 Toll Crossbar, telegraph switching systems, the first time assignment speech interpolation (TASI) system, panel, step-by-step and the Traffic Service Position System (TSPS). Currently, he is Head of the Feature Planning and Traffic Engineering Department for local ESS systems. Senior Member, IEEE; Member, Kappa Mu Epsilon, Sigma Kappa Alpha.

Daniel H. Larson, B.S.E.E., 1951, Northwestern University; U.S. Army 1954-1955; Communications Development Training Program (CDTP), 1956; Bell Laboratories, 1951—. Mr. Larson has worked on the design and field testing of high-frequency receiver and signal processing circuits for the radars used in the Nike and Safeguard anti-aircraft and anti-missile defense systems. Since 1975, he has worked on circuit design for the Traffic Service Position System. Member, Tau Beta Pi, Phi Eta Sigma, Eta Kappa Nu.

R. J. Piereth, B.S.E.E., 1967, Newark College of Engineering; M.S.E.E., 1969, Rutgers University; Bell Laboratories, 1961–1971; AT&T, 1971–1975; Bell Laboratories, 1975—. Mr. Piereth worked on the No. 101 ESS, No. 1 ESS, No. 2 ESS, Automatic Intercept System, and Traffic Measurements at Bell Laboratories before transferring to AT&T in 1971, where his responsibilities included traffic measurement and force administration systems and equipment. Currently, he supervises a group developing circuits and maintenance programs to provide TSPS with CCIS and busy line verification features. Member, Eta Kappa Nu, Tau Beta Pi, IEEE.

R. L. Potter, B.S.E.E., 1964, State University of New York at Buffalo; S.M.E.E., 1965, Massachusetts Institute of Technology; Bell Laboratories, 1964–1977; AT&T, 1977—. At Bell Laboratories, Mr. Potter has worked on maintenance and operational software for No. 1 ESS and has supervised hardware and software design as well as system performance analysis groups for TSPS No. 1. He is currently Assistant Engineering Manager, Exchange Switching Systems, for the American Telephone and Telegraph Company. Member, IEEE, ACM, AAAS, Tau Beta Pi, Sigma Xi.

Edward M. Prell, B.S.E.E., 1962, University of Kentucky; M.S.E.E., 1964, Columbia University; M.S. (Management Science), 1969, Stevens Institute of Technology; Bell Laboratories, 1962—. Mr. Prell has worked on various aspects of hardware and software associated with the Traffic Service Position System and the Automatic Intercept System. He is currently Head of the Traffic Service Position System Planning Department.

V. L. Ransom, B.S.E.E., 1948, Massachusetts Institute of Technology; M.S.E.E., 1952, Case Institute of Technology; National Advisory Committee for Aeronautics, 1948–53; Bell Laboratories, 1953—. Mr. Ransom was first engaged in the design of a special-purpose digital computer for collecting and processing telephone traffic data. He worked briefly on the operational program for No. 1 ESS arranged for data features. He subsequently supervised a group concerned with planning for traffic measuring and service evaluation systems. In 1970, his efforts were shifted to planning for operator services systems. At present, he is head of a department responsible for systems engineering planning for operator services. Senior Member, IEEE; member, American Association for the Advancement of Science.

Sherman C. Reed, B.S.E.E., 1956, University of Oklahoma; M.S.E.E., 1958, Newark College of Engineering; Bell Laboratories, 1956—. At Bell Laboratories, Mr. Reed had many assignments in ballistic missile defense activities from 1956 to 1974. He began work on TSPS in 1974 and has been involved with call programming and operator actions software design and new feature planning for TSPS. He is currently supervisor of the TSPS No. 1B Development Coordination and System Testing Group. Member, Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon.

George Riddell, M.E.E., College of the City of New York, 1954; Bell Laboratories, 1951—. Mr. Riddell was initially engaged in the design of step-by-step switching systems (SXS) and was involved in the development of SXS common control systems, 4-wire PBXs, answering services switching systems, and other electromechanical switching system design work. In 1963, he was assigned to the TSPS No. 1 System development. He now has the responsibility for TSPS Operator Training Systems and TSPS hardware developments.

R. E. Staehler, B.S.E.E., 1947, The College of the City of New York; M.S.E.E., 1948, Polytechnic Institute of Brooklyn; Bell Laboratories, 1948—. Mr. Staehler's early work was on No. 5 crossbar, toll signaling systems, and trainers for guided missile systems. In 1953, he worked on the development of electronic switching systems, specifically, the processor memory for the experimental central office in Morris, Illinois, and the processor logic and call memory for No. 1 ESS. He was appointed Director of the Electronic Switching Projects Laboratory in 1964 with responsibility for special applications for No. 1

ESS to military and data networks, including No. 1 ESS AUTOVON. In 1968, he was appointed Director of the Electronic Systems Design Laboratory with responsibility for development of the 1A Processor. In 1976, he was appointed Director of the Operator Services Laboratory with responsibility for extending the automation of operator services. Member, IEEE, Eta Kappa Nu, Tau Beta Pi, Sigma Xi.

John J. Stanaway, Jr., B.S.E.E., 1969, University of Virginia; M.S.E.E., 1970, Stanford University; Bell Laboratories, 1969—. Mr. Stanaway has worked on the design of a program tape unit for TSPS, TSPS operator training, and microprocessor test equipment for the TSPS system laboratory. He is currently involved in hardware and software developments for the Automatic Intercept System.

R. T. Steinbrenner, B.M.E., 1958, Union College; M.M.E., 1960, New York University; Bell Laboratories, 1958—. Mr. Steinbrenner was responsible for the development of graphical and alphanumeric display systems and analog recording systems for various Ocean Systems projects. He transferred to Bell System activities in 1971 when he assumed his present position as Head of the Recording Systems Department where he is responsible for the development of audio and digital data storage systems.

Kenneth Streisand, B.S., Queens College, 1959; B.S.E.E., 1959, M.S.E.E., 1960, Columbia University; Bell Laboratories, 1960—. Mr. Streisand has spent most of his laboratory career in voice frequency signaling development. This included work in multifrequency, *TOUCH-TONE®*, and single frequency signaling circuits and specialized projects such as calling line identification (CLI) and automated coin toll system (ACTS). Currently he is working on new features for the remote test system (RTS) used in special service circuit maintenance (SARTS). Member, IEEE, Tau Beta Pi, Eta Kappa Nu.

C. R. Swanson, B.S.E.E., 1950, Iowa State University; Bell Laboratories, 1951—. Mr. Swanson worked on Nike-Ajax, Nike-Hercules, and Nike-Zeus radar systems as well as the ZMAR and Nike-X projects. From 1967 to 1970 he was responsible for the Nike-X missile site radar operations at Kwajalein. He was later Safeguard MSR Software Process Manager and in 1974 was responsible for the integration of the MSR software and hardware at the Safeguard MSR site in North Dakota. In 1974 he assumed responsibilities in hardware and maintenance software of the TSPS project.

D. Van Haften, B. S. and M. S. (Mathematics), 1970, Michigan State University; Ph. D. (Electrical Engineering), 1977, Stevens Institute of Technology; Bell Laboratories, 1970—. Mr. Van Haften initially worked on the Safeguard project. Since joining the Operator Systems Laboratory in 1974, he has been involved in TSPS system testing, field support, and call processing development. Currently he is a member of the Development Coordination and System Test Group. Member, Phi Beta Kappa, Phi Kappa Phi, Pi Mu Epsilon.

John J. Victor, B.S. (Mathematics), 1967, St. Peter's College; M.S. (Computer Science), 1971, Stevens Institute of Technology; Bell Laboratories, 1967–1978; Western Electric, 1978—. Mr. Victor's work at Bell Laboratories included TSPS input/output software design, maintenance software design, control software design, remote maintenance software requirements and design, system testing and integration, and field support. He currently is Department Chief at Western Electric, where his group is responsible for the planning and management of toll operator systems, including TSPS.

Laurance A. Weber, B. E. E., 1945, Cornell University; M. E. E., 1955, Polytechnic Institute of Brooklyn; Bell Laboratories, 1946—. Mr. Weber was initially involved in the design of signaling circuits. Later he participated in the design of circuits for crossbar tandem systems. Following this assignment, he was appointed supervisor in charge of designing data sets for the mechanization of TWX service. He was appointed head of the 101 ESS Design Department in 1960. He has had subsequent assignments in No. 2, No. 2B, and No. 3 ESS. He is presently head of the TSPS Evaluation and Test Department. Member, IEEE, Tau Beta Pi, Sigma Xi, Eta Kappa Nu.

R. J. Welsch, B.S. (Math), 1967, Marquette University; M.S. (Computer Science), 1972, Northwestern University; Bell Laboratories, 1967–1968; U.S. Army, 1968–1970; Bell Laboratories, 1970—. At Bell Laboratories, Mr. Welsch has had experience with No. 1 ESS/ADP, No. 4 ESS, and systems engineering for electromechanical switching systems. From 1974 to 1978, he was a member of the Program Administration and Support Program Group of the Operator Systems Laboratory. He is currently with the Software Development Systems (UNIX) Applications Group of the Microprocessor and Software Technology Laboratory.

Allen G. Weygand, M.E., 1957, Stevens Institute of Technology; M.S.(E.E.), 1959, New York University; Bell Laboratories, 1957–1962; Bellcomm, 1962–1972; Bell Laboratories 1972—. Before transferring to Bellcomm, Mr. Weygand conducted transmission performance studies of various exchange-area voice-frequency systems and short-haul analog carrier systems. While at Bellcomm, he performed communications and tracking systems planning, analysis, and engineering studies in support of the Apollo, Skylab, and space shuttle space flight programs of NASA. Since his return to Bell Laboratories, Mr. Weygand has been engaged in transmission planning and analysis for Bell System operator services and various customer attendant services. Member, Tau Beta Pi.

Edward A. Youngs, B. A. (psychology), 1964, Dartmouth College; M.A. (psychology), 1968, University of North Carolina; Ph.D. (psychology), 1969, University of North Carolina; Bell Laboratories, 1969—. Mr. Youngs initially worked on human factors aspects of electronic key telephones. Later, he contributed in-depth studies of the TSPS operator job. Recently, as supervisor of a human factors group concerned with customer and operator performance and acceptance of new TSPS services, he has participated in the design of Automated Coin Toll Service (ACTS) and other operator service mechanization efforts.

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

Volume 58 July-August 1979 Number 6, Part 2

Evaluation of Adaptive Speech Coders Under Noisy Channel Conditions

By C. SCAGLIOLA

(Manuscript received June 20, 1978)

An experiment has been performed in which the digital transmission of speech coded by adaptive differential PCM was simulated under noisy channel conditions. The experiment was done with two aims: (i) to get information on the subjective effect of channel errors and the influence of various design parameters on the speech quality under various conditions and (ii) to find objective measures for predicting the overall quality of the processed speech over a wide range of circuit conditions. The subjective results show that, for a speech transmission through a channel with bit error probability up to 1/256, best results can be obtained with a slow error recovery, associated with fast quantizer adaptation. The use of slow error recovery and slow quantizer adaptation is preferable for channels with very high bit error rates, like 1/32. Overall subjective quality is well predicted by the sum of two terms: (i) an objective performance measure of the noise present on the output signal, disregarding any effect of level mismatching due to the sensitivity of the adaptation algorithms to channel errors and (ii) a measure of the level mismatching which takes into account both the average gain on the output signal and its fluctuation in time. The best prediction scores are achieved by three newly defined objective performance measures, two-level compensated segmental SNRs, and a spectral signal-to-distortion ratio.

I. INTRODUCTION

The design of digital waveform coders for speech communications must face the inevitable presence of channel errors. Adaptive coders, like ADPCM (adaptive differential PCM), in which the adaptation of the

quantizer step-size is derived from the transmitted binary stream and no error-protected side information is sent to the receiver, may be particularly sensitive to this problem. In fact, a single channel error may cause a multiplicative offset between the signal level at the receiver and that at the transmitter. This offset may persist indefinitely if no error dissipation mechanism is provided.

Recently, some algorithms of quantizer adaptation have been developed that make the effect of a single transmission error die out over time, so the transmitter and receiver can resynchronize their step-size estimates.[1,2] The possibility of obtaining such results is physically due to the fact that these algorithms have an imperfect adaptation: the step size increases more quickly and decreases more slowly for low input levels than for high ones. In this way, the step size is overestimated for low input levels and underestimated for high ones, thereby reducing the dynamic range of the coder. Dynamic range and error dissipation rate vary inversely, and the designer has to balance between them.

In the case of speech transmission, the choice of the appropriate design parameters must be based on a precise evaluation of the subjective quality of the coded speech. Use of the conventional long-term signal-to-noise ratio as an estimator of the subjective quality would be, in this instance, completely misleading, at least because an offset in the signal amplitude between input and output, due to an offset in step sizes caused by an error, will be reflected in a noticeable squared difference between the two waveforms, while it may not be subjectively disturbing.

To study the subjective performance of ADPCM coders operating under both error-free and noisy channel conditions, an experiment has been conducted, as summarized in Section II. The following three sections provide a brief description of the coding method, the definition of several objective performance measures, and the description of the experimental design and testing procedure. Sections VI and VII provide analyses of subjective and objective measurement data. In Section VIII, the results of the previous two sections are discussed and a physical interpretation is given of the principal findings on quality prediction.

II. OVERVIEW OF THE EXPERIMENT

The experiment included 12 different ADPCM coding schemes, which comprised all combinations of two bit rates, two adaptation time constants, and three error dissipation rates. These systems processed a total of 288 speech samples from four talkers (two male and two female), at two different power levels (24 dB apart) and with three different probabilities of independent errors on the channel.

Twenty listeners rated the quality of the processed speech samples

on a scale from 1 to 9. The odd values were associated with the adjectives: unsatisfactory, poor, fair, good, excellent. In addition to the subjective data, a fairly large number of objective performance measures were also taken on the processed speech samples.

The aims of the experiment included the study of:

(*i*) The influence on speech quality of the above design parameters.

(*ii*) The optimum combination of parameters for a given error probability.

(*iii*) The objective measures or combinations of objective measures which are good predictors of speech quality even under noisy channel conditions.

The principal conclusions drawn from the analyses of the subjective and objective data are:

(*i*) Contrary to a common feeling, a very slow error dissipation is sufficient to ensure good robustness of the ADPCM coder to error rates even much higher than those encountered in a normal telephone connection. When no recovery mechanism is provided, a fairly slow adaptation makes the system not very sensitive to channel errors in the range of error rates typical of a telephone connection.

(*ii*) For speech transmission through a channel with bit error rate up to 1/256, best results can be obtained with a very slow error dissipation, associated with fast quantizer adaptation; when the slow error dissipation is associated instead with a slow adaptation, the system becomes fairly robust to very high error rates, like 1/32, at the expense of a slight quality deterioration at low error rates.

(*iii*) Good predictors of subjective quality were found to be two-segmental SNR measures in which a compensation of the level mismatching between input and output was performed on a frame-by-frame basis. The combination of any of these measures with two separate measures of level mismatching further improved the prediction accuracy.

III. ROBUST ADPCM SYSTEM: A BRIEF DESCRIPTION

Figure 1 is a block diagram of the ADPCM coder-decoder used in the experiment. The predictor is a second-order transversal filter, with tap coefficients 1 and −0.5. The step size $\Delta(k)$ is adapted according to the robust algorithm described in Ref. 1, which permits synchronizing the step-size estimates at transmitter and receiver, after a transmission error occurs, during a period of error-free transmission

$$\Delta(k+1) = \Delta^{\beta}(k) \cdot M(I(k)), \tag{1}$$

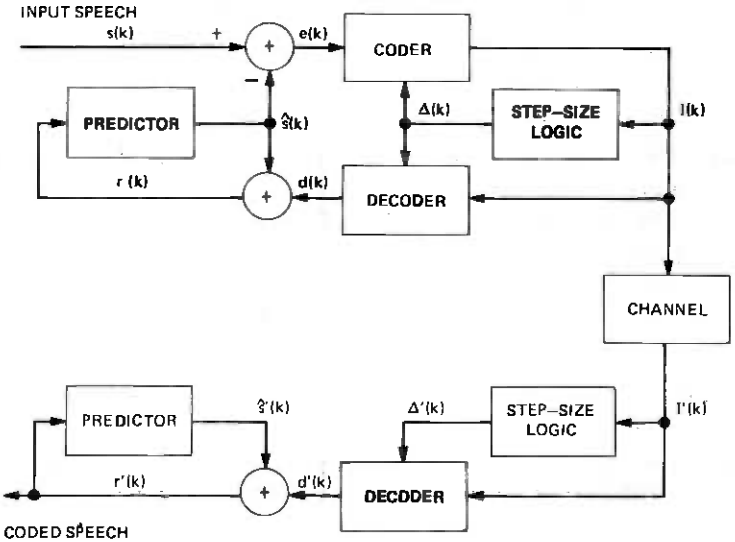


Fig. 1—Block diagram of ADPCM coder-decoder.

where the exponent β, $0 < \beta < 1$, is the decay constant and $M(I(k))$ is the step-size multiplier at time k. The multiplier $M(I(k))$ depends only on the actual code word $I(k)$ and assumes $N = 2^{B-1}$ distinct values $\{M_1, M_2, \cdots M_N\}$, where B is the number of bits used to encode the prediction error $e(k)$.

The decay speed is shown to be independent of the actual code word transmitted and also of the values of the multipliers, being only a function of the decay constant β.[1] When $\beta = 1$, the decay speed is zero and the decay time infinite. With $\beta < 1$, the decay speed increases, but at the same time the dynamic range decreases.

It has been shown that the loading factor (ratio between range of the quantizer and rms quantizer input) is a constant as a function of the input level if $\beta = 1$, but it is a decreasing function of the input level if $\beta < 1$.[1] The form of their relationship is almost linear, with slope approximately inversely proportional to $\log(M_N/M_1)/(1 - \beta)$, as indicated by Fig. 6 of Ref. 1 and as recently proved theoretically by D. Mitra.[3] Therefore, with a small β and a small ratio between the maximum and minimum multipliers, the coder will produce more granular noise for low input level and more overload distortion for high levels.

The values of the multipliers play another important role in the overall performance of the coders under noisy channel conditions. In fact, they determine the magnitude of the initial offset after a single

transmission error. In the worst case, the offset is given by

$$\left| \frac{\Delta'(k)}{\Delta(k)} \right|_{\max} = \frac{M_N}{M_1}. \tag{2}$$

In the present experiment, the multipliers M_i were related by the linear relationship:

$$M_i = [\alpha + C(1 - \alpha)(i - 0.5)]\hat{\Delta}^{(1-\beta)}, \tag{3}$$

where $\hat{\Delta}$ is the step size that gives optimum performance at the desired nominal input level. This was fixed at −21 dBm, i.e., 27 dB under the saturation threshold for the signal that in the internal 16-bit computer representation is 32767. This relationship was chosen because, for $\beta = 1$, the adaptation algorithm coincides with the magnitude estimation algorithm described by Castellino et al.,[4] and for this algorithm more information about the subjective effects of its parameters are available.[5] In eq. (3), C essentially determines the mixture of granular noise and clipping distortion in the decoded prediction error at the nominal level. The parameter α controls mainly the speed of adaptation and hence the ratio between maximum and minimum multipliers.

IV. OBJECTIVE PERFORMANCE MEASURES

Several objective performance indices were measured for each utterance in the experiment. The speech samples used as input to the coders were low-pass filtered at 3.4 kHz before being sampled and converted into digital form by a 16-bit A/D converter operating at 8-kHz sampling rate. Figure 2 is a block diagram of the simulated circuit arrangement for performing coding and measurements. A filter after the ADPCM decoding limits the bandwidth of the output speech as in a real situation. A secondary path provides the reference signal $s_0(k)$ with which the filtered output $r_0(k)$ is compared to compute the objective performance measures. With two identical filters in the main

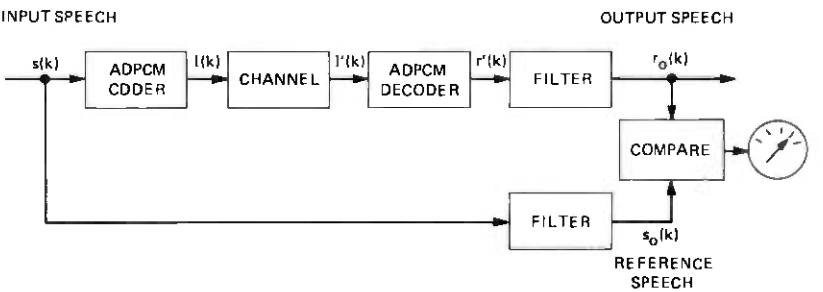


Fig. 2—Block diagram of simulated circuit arrangement for coding and measurements.

and reference paths, only the distortion introduced by the coder is measured. The filters are 5th-order elliptic low-pass with 3.4-kHz cutoff frequency, 0.25-dB in-band ripple, and at least 40 dB stopband attenuation.

The measures are classified in the two categories of time domain measures and frequency domain measures. Frequency-weighted signal-to-noise ratios are included in the first category because they rely strongly on the exact time synchronization of the two waveforms and on the absence of phase distortion.

4.1 Time domain measures

4.1.1 Long-term signal-to-noise ratio (SNR)

$$\text{SNR} = 10 \log \frac{\sum_k s_o^2(k)}{\sum_k [s_o(k) - r_o(k)]^2}, \tag{4}$$

where k ranges over all the samples of the utterance. SNR is the ratio between the long-term signal energy and the long-term noise energy, the noise being defined as the difference between reference signal $s_0(k)$ and output signal $r_0(k)$.

4.1.2 Segmental signal-to-noise ratio (SNR_{seg})

Here the utterance is divided into adjacent segments of J samples each, and the signal-to-noise ratio in each segment is measured in decibels. The noise is still defined as the difference between corresponding samples of reference and output speech. The segmental SNR is the average of these measures over the M segments of the utterance.

$$\text{SNR}_{\text{seg}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\sum_{j=1}^{J} s_o^2(j + mJ)}{\sum_{j=1}^{J} [s_o(j + mJ) - r_o(j + mJ)]^2}. \tag{5}$$

In this experiment, $J = 128$, corresponding to 16 ms segments. This measure, proposed by Noll,[6] was recently found to correlate very nicely with subjective ratings of ADPCM-coded speech.[5] A very important feature added to this basic formula consists in discarding from the computation those segments in which the signal power is below −54 dBm. This threshold, whose value was found to be appropriate for high quality speech,[7] was introduced to avoid a slight idle channel noise having an unduly great negative weight in the overall performance measure. This is done also in all the following time domain measures.

4.1.3 Compensated signal-to-noise ratio (SNR_{com})

This measure was specifically formulated to compensate for the level variations that may occur under channel error conditions when the coder has a slow error dissipation. The difference between reference and output signals due to level offset should not be measured in fact as noise.

To compensate for these level variations, let us formulate the coding process in the mth segment as composed of an amplification of the input signal, the addition of an uncorrelated random noise, and a possible dc component. Therefore, the output process in the mth segment can be written as

$$r_o(k) = g(m)s_o(k) + q(k) + Q. \tag{6}$$

This coincides with the simple linear regression model of $r_0(k)$ on $s_0(k)$. Therefore, the gain factor $g(m)$ is the slope of the regression line of the ouput on the reference signal in the mth segment, and the noise term $q(k)$ is the minimum error made in predicting $r_0(k)$ from $s_0(k)$.[8]

Let us define the signal-to-noise ratio in the mth segment as the ratio between the variances of $g(m) \cdot s_0(k)$ and $q(k)$. This can be shown to be only a function of the correlation coefficient $\rho(m)$ between the reference and output signals in the mth segment:[8]

$$\rho(m) = \frac{JS_{sr}(m) - S_s(m)S_r(m)}{\sqrt{[JS_{ss}(m) - S_s^2(m)][JS_{rr}(m) - S_r^2(m)]}}, \tag{7}$$

where $S_x(m)$ indicates the summation of $x(j)$ and $S_{xy}(m)$ the summation of $x(j) \cdot y(j)$ over the J samples of the mth segment.

Averaging in decibels across the M segments in the utterance, the compensated SNR turns out to be:

$$\text{SNR}_{\text{com}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\rho^2(m)}{1 - \rho^2(m)}. \tag{8}$$

4.1.4 Average gain (G)

This is the average in decibels of the gain factor $g(m)$ defined before, across the M segments. From simple linear regression analysis, the gain $g(m)$ is:[8]

$$g(m) = \frac{JS_{sr}(m) - S_s(m)S_r(m)}{JS_{ss}(m) - S_s^2(m)}. \tag{9}$$

The average gain, which is an indication of how much the output level was increased or decreased on the average, with respect to the input level, is:

$$G = \frac{1}{M} \sum_{m=0}^{M-1} 20 \log g(m). \tag{10}$$

4.1.5 Gain fluctuation (σ_g)

This is simply the standard deviation of the gain $g(m)$, measured in decibels, across the M segments. It is a measure of how much the output level fluctuates owing to transmission errors.

$$\sigma_g = \left[\frac{1}{M}\sum_{m=0}^{M-1}(20 \log g(m))^2 - G^2\right]^{1/2}. \tag{11}$$

4.1.6 Maximum signal-to-noise ratio (SNR_{max})

An alternate way of compensating for the level mismatching between reference and output speech signals was found by defining the noise in the mth segments as

$$\epsilon(k) = s_o(k) - \hat{s}_o(k), \tag{12}$$

where

$$\hat{s}_o(k) = a_1(m) \cdot r_o(k) + a_o \tag{13}$$

is the least-square estimate of $s_0(k)$ based on $r_0(k)$.

The ratio between the variances of the reference signal and the minimum estimation error $\epsilon(k)$ (hence the name "maximum SNR") is again only a function of the correlation coefficient $\rho(m)$, defined by eq. (7).

Averaging again in decibels across the M segments in the utterance, the maximum SNR is

$$\text{SNR}_{\max} = \frac{1}{M}\sum_{m=0}^{M-1} 10 \log \frac{1}{1-\rho^2(m)}. \tag{14}$$

It is readily seen that $\text{SNR}_{\max}$ is always greater than SNR_{com} and that the two measures give essentially different results only for low-quality coding conditions.

4.1.7 Frequency-weighted, segmental, signal-to-noise ratios

This term indicates a fairly large class of measures. In these measures, the frequency axis is partitioned into many bands, usually nonuniform, the reference and output spectra are compared, some performance measure is then computed over each band, and these measures are averaged across the bands. In the measures described below, the spectra are computed over 256 points (32 ms). The segmental measures are obtained by averaging the measures taken every 128 samples (16 ms).

Three measures are reported here. They are described more in Refs. 7 and 9. The partitioning of the frequency axis is effected in those three cases according to the 16 classical articulation bands.[10]

$$\text{SNRF}_1 = \frac{1}{M}\sum_{m=0}^{M-1}\left[\frac{1}{16}\sum_{j=1}^{16} 10\log\frac{S_j(m)}{N_j(m)}\right] \tag{15}$$

$$\text{SNRF}_2 = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{16} L_j(m)\ 10\log\dfrac{S_j(m)}{N_j(m)}}{\sum_{j=1}^{16} L_j(m)} \tag{16}$$

$$\text{SNRF}_5 = \frac{1}{M}\sum_{m=0}^{M-1} 10\log\frac{1}{1+\sum_{j=1}^{16}\dfrac{N_j(m)}{S_j(m)}}, \tag{17}$$

where $S_j(m)$ is the energy of the reference signal $s_0(k)$ in the jth frequency band in the frame m and $N_j(m)$ is the corresponding noise energy. The noise is again defined as the difference between reference and output signals, the latter being preventively divided by the gain $g(m)$ previously defined to compensate for level fluctuations. However $g(m)$ is computed on the 256-point analysis window. In eq. (16), a "loudness weighting" has been introduced. The term $L_j(m)$ corresponds approximately to the subjective loudness in band j, and is computed as

$$L_j(m) = \int_{\text{band } j} |S_r(f)|^{1/2}\, df, \tag{18}$$

where $S_r(f)$ is the spectrum of the gain-compensated output speech $r_0(k)/g(m)$.

4.2 Frequency domain measures

All the spectral measures here presented and used in the experiment are based on the concept of linear prediction or inverse filtering.[11] The speech signal is represented by the pth order autoregressive model:

$$s(k) = \sum_{i=1}^{p} a_i s(k-i) + u(k), \tag{19}$$

where $u(k)$ is the white spectrum excitation function and the a_i's are the coefficients of the inverse filter

$$A_s(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}. \tag{20}$$

The coefficients a_i's are computed to minimize the residual power of the signal at the inverse filter output.

In this paper, the dissimilarity between the spectra of reference and output speech in a given frame is computed essentially by comparing the residual powers of the signals $s_0(k)$ and $r_0(k)$ filtered by the inverse filters $A_s(z)$ and $A_r(z)$, derived from the same two signals. Four residual powers can be computed in the mth signal frame:

(*i*) $P_e(m)$ obtained passing $s_0(k)$ through $A_s(z)$.
(*ii*) $P_d(m)$ obtained passing $r_0(k)$ through $A_s(z)$.
(*iii*) $P_{e'}(m)$ obtained passing $r_0(k)$ through $A_r(z)$.
(*iv*) $P_{d'}(m)$ obtained passing $s_0(k)$ through $A_r(z)$.

Four objective measures based on these concepts are presented in the following paragraphs.

4.2.1 LPC distance measure (D_1)

This measure, proposed by Itakura,[12] is also called log likelihood ratio. The distance between output and reference speech in the mth frame is defined as

$$D_1 = \ln \frac{P_d(m)}{P_{e'}(m)}. \tag{21}$$

It can be shown that D_1 can be expressed in terms of spectral differences between the LPC models of the two frames of speech.[13] Moreover, it results that the spectral difference is most heavily weighted in the peaks of the input speech smoothed spectrum, i.e., in the speech formants.

Interchanging the roles of reference and output speech, a different log likelihood ratio is obtained:

$$D_2 = \ln \frac{P_{d'}(m)}{P_e(m)}, \tag{22}$$

which has the same basic properties as D_1.

The measure used here is actually the arithmetic mean of D_1 and D_2, averaged across the M segments of the utterance:

$$D_I = \frac{1}{2M} \sum_{m=0}^{M-1} \left[\ln \frac{P_d(m)}{P_{e'}(m)} + \ln \frac{P_{d'}(m)}{P_e(m)} \right]. \tag{23}$$

4.2.2 Bharucha index (D_B)

This index is again a distance measure, similar to the log likelihood ratio. It has been formulated by Bharucha[14] and its definition is also

reported in Ref. 7. The basic idea is that of measuring the noise introduced by the coder by "notching out" the speech spectrum by means of a time-varying linear filter, whose transfer function is matched to the inverse of the short-term spectral envelope. The quality index proposed by Bharucha is essentially the average increase in the residual power at the output of the "notch filter," due to coding:

$$D_B = \frac{1}{M}\sum_{m=0}^{M-1} 10 \log \left[\frac{P_s(m)}{P_r(m)} \cdot \frac{P_d(m)}{P_e(m)}\right], \tag{24}$$

where $P_s(m)$ and $P_r(m)$ are the powers of the reference and output speech in the mth segment and provide the appropriate scaling in the measure. It can be shown that an uncorrelated noise component in the output speech is inversely weighted, frame by frame, by the smoothed LPC spectrum of the input speech signal.[7,14] Therefore, the noise has more weight in those frequency bands where the signal energy is low; this is probably in conformity with subjective noise evaluation.

4.2.3 Spectral signal-to-distortion ratios

Following the same basic idea of the Bharucha index, two other measures were derived in the form of signal-to-distortion ratios. In fact, they are measured in decibels, and they increase with increasing quality, like the time domain SNRs. With the first SDR, a distortion power is defined as the difference between the prediction error powers $P_d(m)$ and $P_e(m)$ defined above.

Before taking the difference, however, the term $P_d(m)$ is multiplied by $P_s(m)/P_r(m)$ to compensate for level differences between input and output.

$$\text{SDR}_1 = \frac{1}{M}\sum_{m=0}^{M-1} 10 \log \frac{P_s(m)}{P_d(m) \cdot [P_s(m)/P_r(m)] - P_e(m)}. \tag{25}$$

In the second SDR measure, the difference between the signal-to-prediction error ratios in decibels that is averaged to give D_B is instead computed relative to the input signal-to-prediction error ratio, and this new ratio is again averaged in decibels across the segments of the utterance.

$$\text{SDR}_2 = \frac{1}{M}\sum_{m=0}^{M-1} 10 \log \frac{10 \log[P_s(m)/P_e(m)]}{10 \log[P_s(m)/P_e(m)] \cdot [P_d(m)/P_r(m)]}. \tag{26}$$

In this experiment, D_B, SDR_1, and SDR_2 are computed with an analysis window of 160 samples (20 ms) that was shifted by 128 samples (16 ms) every frame. The inverse filters for computing $P_e(m)$ and $P_d(m)$ were always of the 20th order.

V. EXPERIMENTAL DESIGN AND PROCEDURE

5.1 Circuit conditions

The choice of the experimental design variables was dictated by the criterion of a broad quality range and of each value the variables assumed having caused an effect that could be perceived for at least some combination of the other variables.

Two bit/sample values were chosen, $B = 3$ and $B = 4$, and three values of the decay constant, $\beta = 1$, $\beta = 255/256$, and $\beta = 63/64$.

The condition $\beta = 1$ was included because it should give the broadest dynamic range and because it makes the ADPCM system identical to the earlier schemes.[4,15]

For the step-size multipliers, given by formula (3), C was kept constant for each bit rate and equal to the values 0.65 and 0.41 that were found to be subjectively optimum for $B = 3$ and $B = 4$, respectively.[5] The parameter α was given two values, 0.75 and 0.96875. The first value produces a rather fast adaptation, with a corresponding time constant of 0.5 ms. With $\alpha = 0.75$, the ratio M_N/M_1 is about 4 dB and 5.5 dB for $B = 3$ and $B = 4$, respectively. The second value of α produces a rather slow adaptation, the corresponding time constant being 4 ms. The ratio M_N/M_1 is much smaller, about 0.5 dB and 0.7 dB for 3- and 4-bit/sample, respectively.

Two different input levels, 24 dB apart, were used for every combination of the other conditions. They were $L = -33$ dBm and $L = -9$ dBm, i.e., symmetrical around the nominal input level -21 dBm for which the coder was designed to have optimum performance. The level -21 dBm was not included to keep the dimension of the experiment within reasonable limits of feasibility.

Finally the channel was characterized by three different probabilities of independent errors, $P(e) = 0$, $P(e) = 1/256$, $P(e) = 1/32$.

Summarizing, the experiment included 72 conditions which comprised all the combinations of $B = 3, 4$; $\alpha = 0.75, 0.96875$; $\beta = 1, 255/256, 63/64$; $L = -33, -9$; $P(e) = 0, 1/256, 1/32$.

5.2 Preparation of stimuli

Each experimental condition was simulated four times, using as input signals four sentences spoken by two male and two female talkers. Each talker spoke into a high-quality dynamic microphone, while seated in a sound-proof booth. The digital recordings had been generated by low-pass filtering the amplified microphone signal at 3.4 kHz, and then sampling and converting it into digital form by a 16-bit A/D converter operating at 8-kHz sampling rate.

For simulating the different input levels, the sentences, all previously adjusted to the same mean power level of -21 dBm, were multiplied by a constant factor at the coder input, and then divided by the same

factor at the output. In this way, each processed sentence was listened at the same level, unless channel errors and/or slope overload of the coder had caused output level variations.

From the 288 simulations, two analog test tapes were generated, each containing in a different random order two simulations of each experimental condition, one with a male and one with a female talker. For each talker, 18 different sentences read from a different phonetically balanced list were used, so that in each tape the same sentence appeared only twice.

5.3 Testing procedure

Twenty paid subjects (10 for each tape), all students from junior and senior classes of local high schools, judged the 288 stimuli. They listened to the processed speech binaurally over Pioneer SE 700 earphones at a nominal level of 80 dB_{SPL}, while seated in a double-walled sound booth. As pointed out before, the level of individual sentences varied according to the particular experimental conditions. The total listening time for each group of subjects was about 30 minutes, with a short break after the 80th sentence. After each stimulus, the subjects had 4 seconds to record their judgments. They were asked to rate the quality of the stimuli according to the adjectives: excellent, good, fair, poor, unsatisfactory. Their answer sheet contained 144 rows of short lines divided into nine columns, with the odd ones labeled with the adjectives. In this way, the subjects were allowed to check intermediate ratings, if they chose to do so.

The categorical judgments expressed by the listeners were subsequently converted into numerical scores, assigning value 1 to the category "unsatisfactory," value 9 to the category "excellent," and intermediate integer values to intermediate categories.

Before the actual test sessions took place, the subjects listened to 12 practice sentences different from those used in the experiment, spoken by the same four talkers, and representative of the range of quality they expected in the test.

VI. ANALYSIS OF SUBJECTIVE RESULTS

6.1 Control variables

The purpose of the subjective test was to assess the different behavior of coders in the presence of different operating conditions, namely, channel error rate and input level. Other sources of variability in listener responses are expected to cancel out in the average data for each experimental condition. Before averaging the data for each experimental condition, it was necessary to assess the importance of such extraneous sources of variability, as differences in the way the listeners judged the stimuli and differences due to talker voices.

6.1.1 Listeners

To assess the variability due to listener differences, their responses were analyzed according to MDPREF.[16,17] This is a factor analytic procedure that derives a geometrical multidimensional space representation, in which the stimuli are represented by points and the subjects by vectors. The projections of the points on a vector are the best fit with the scores given to the stimuli by that subject. Basically, MDPREF reveals whether the subjects attended to different psychological attributes in the stimuli or if they attached different weights to each of the various attributes. In the solution for the 72 experimental conditions and the 20 subjects, the first principal component accounted for only 55 percent of the variance, while the remaining 45 percent was distributed over all the other components: 4.2 percent for the 2nd, 3.8 percent for the 3rd, 3.4 percent for the 4th, 3 percent for the 5th, etc.

The fact that 45 percent of the total variance was accounted for by so many axes in an almost uniform fashion indicates that these axes do not represent different perceptual attributes of the stimuli, but that they account only for the "noise" in the subjective data. In other words, in spite of the low variance accounted for by the first axis, it is evident that the listeners attended essentially to the same attributes with the same weights and then only a unidimensional solution exists. Therefore, the mean of the listeners' ratings for each condition were used for the subsequent analyses.

6.1.2 Talkers

An analysis of variance was computed to study the variability of the scores obtained by the talkers of different sex and to assess the validity of averaging the ratings across talkers to perform an analysis of the effect of the design variables. The analysis showed that the difference due to the sex of the talker was highly significant. The average score was 4.40 for female talkers and 5.07 for male talkers. On the other side, however, all the interactions between the sex of the talker and the design variables were not significant. This indicates that the sex of the talker influenced the average value of the ratings, but not the relative ranking of the various experimental conditions. Therefore, the mean ratings across listeners and talkers were used for further analyses, reducing the variability of the data to that due to the physical variables of the coders and the circuits.

6.2 Design variables

In Figs. 3 and 4, the mean ratings across listeners and talkers are shown for each bit rate B and decay constant β as a function of the probability of error $P(e)$, with α and the level L as parameters.

When no error dissipation mechanism is provided (Figs. 3a and 4a), a slower adaptation, i.e., $\alpha = 0.96875$, makes the system less sensitive to the errors. With a slow dissipation, i.e., $\beta = 255/256$ (Figs. 3b and 4b), the slow adaptation is advantageous only at the higher bit error rate, while with no errors the performance appears to be worse than with fast adaptation. With faster error dissipation, i.e., $\beta = 63/64$ (Figs. 3c and 4c) and slow adaptation, the unbalancing of the load factor between low and high input level is very high and the performance at the low level is very degraded even with no errors. With fast adaptation, i.e., $\alpha = 0.75$, the dynamic range is instead very high; besides, even if the performance under error-free conditions is lower than with slower error dissipation, the system is very insensitive to channel errors.

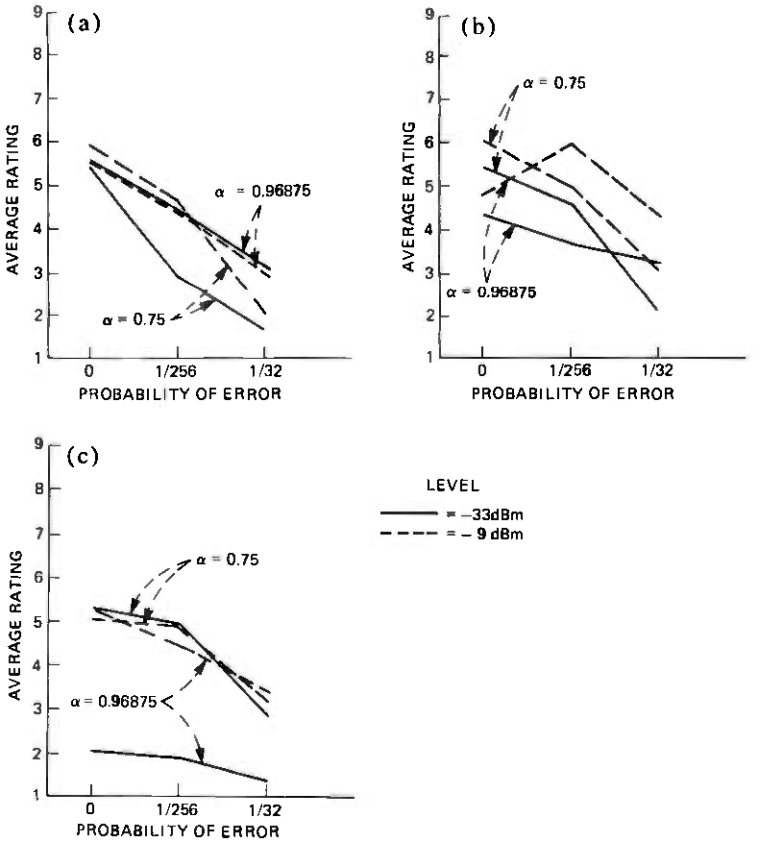


Fig. 3—Mean ratings as a function of probability of error, for $B = 3$: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

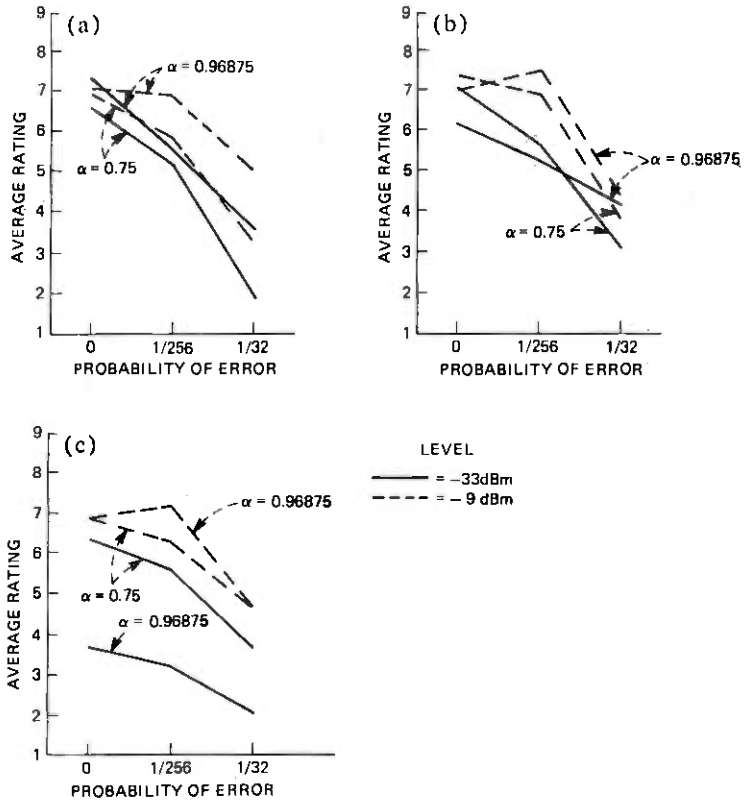


Fig. 4—Mean ratings as a function of probability of error, for $B = 4$: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

6.2.1 *Analysis of variance*

A five-way analysis of variance was performed to evaluate the effect of the experimental variables. The results of the analysis are reported in Table I. The last column reports the P value, that is, the probability that the test statistics assume a value greater than or equal to the computed F ratio, under the null hypothesis, compared with the two significance levels 0.05 and 0.01.

The analysis showed that all the main effects except that due to α are highly significant. The fact that α has no significant effect means that it has a positive effect for certain combinations of parameters and a negative one for others, as shown by the significant interactions.

The interactions between B and α and between B and β were found to be not significant, indicating that α and β have the same effects on the quality of coded speech whether it is a three-bit or a four-bit one. All the other two-way interactions are highly significant. Only the

Table I—Analysis of variance of the mean scores across listeners and talkers

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Ratio	Significance
B	1	31.8928	31.8928	204.7	$P<0.01$
α	1	0.1733	0.1733	1.11	>0.05
β	2	4.9167	2.4583	15.77	<0.01
L	1	19.1570	19.1570	122.96	<0.01
$P(e)$	2	85.5395	43.7697	274.52	<0.01
$B \times \alpha$	1	0.2854	0.2854	1.84	>0.05
$B \times \beta$	2	0.0034	0.0017	0.01	NS
$B \times L$	1	0.7642	0.7642	4.93	<0.01
$B \times P(e)$	2	1.8088	0.9044	5.84	<0.01
$\alpha \times \beta$	2	11.0713	5.5356	35.78	<0.01
$\alpha \times L$	1	3.6933	3.6933	23.87	<0.01
$\alpha \times P(e)$	2	5.0383	2.5191	16.28	<0.01
$\beta \times L$	2	3.8310	1.9155	12.38	<0.01
$\beta \times P(e)$	4	4.7084	1.1772	7.60	<0.01
$L \times P(e)$	2	1.5342	0.7671	4.95	<0.05
$\alpha \times \beta \times L$	2	6.2810	3.1405	20.16	<0.01
Residual	43	6.9980	0.1558		

interaction between L and $P(e)$ is significant at $P < 0.05$ but not at $P < 0.01$. This indicates that the level has an effect almost independent of the probability of error.

The significant interaction between B and L is due to the fact that the difference between the ratings at the two levels is greater on the average for $B = 4$ than for $B = 3$. The significant interaction between B and $P(e)$ is instead due to the fact that the coder with higher bit rate has a greater loss in quality in passing from $P(e) = 0$ to $P(e) = 1/32$.

Of the three-way interactions, only that among α, β, and L is significant. All the other three- and four-way interactions were not significant, and they were pooled in the residual.

VII. QUALITY PREDICTION BY OBJECTIVE MEASURES

To find an objective predictor of the speech quality, linear regression procedures were used. A linear model was chosen not only for its simplicity, but also because in many cases it proved to be adequate to represent the relationship between objective measures and subjective quality. A linear relationship exists, for instance, between the simple SNR and the quality of speech degraded only by the addition of stationary random noise or of speech dependent noise.[18-20] A linear relationship exists also between signal-to-granular noise ratio and probability of overload, and the quality of speech processed by ADPCM coders when no transmission errors are present.[5]

To perform regression analyses, the subjective ratings were averaged across listeners and talkers, and the objective measures were also

averaged, taking the arithmetic mean of the values obtained for each processed sentence. The gain fluctuation σ_g was instead averaged quadratically, taking the square root of the arithmetic mean of the squared values σ_g^2.

Different sets of regression formulas were computed, in which the objective performance measures, like signal-to-noise ratios or spectral distance measures, were used either singly or in combination with the two measures of level mismatching. Figures 5 and 6 show the gain fluctuation and the average gain, both averaged across bit rate, as a function of the probability of error. A few remarks should be made on these figures. Although the two sets of measures have a fairly low correlation of 0.55, the patterns are much alike for low input level and fast adaptation. For the 18 conditions with $L = -33$ dBm and $\alpha = 0.75$, the correlation between G and σ_g is, in fact, 0.96. Therefore, even if in

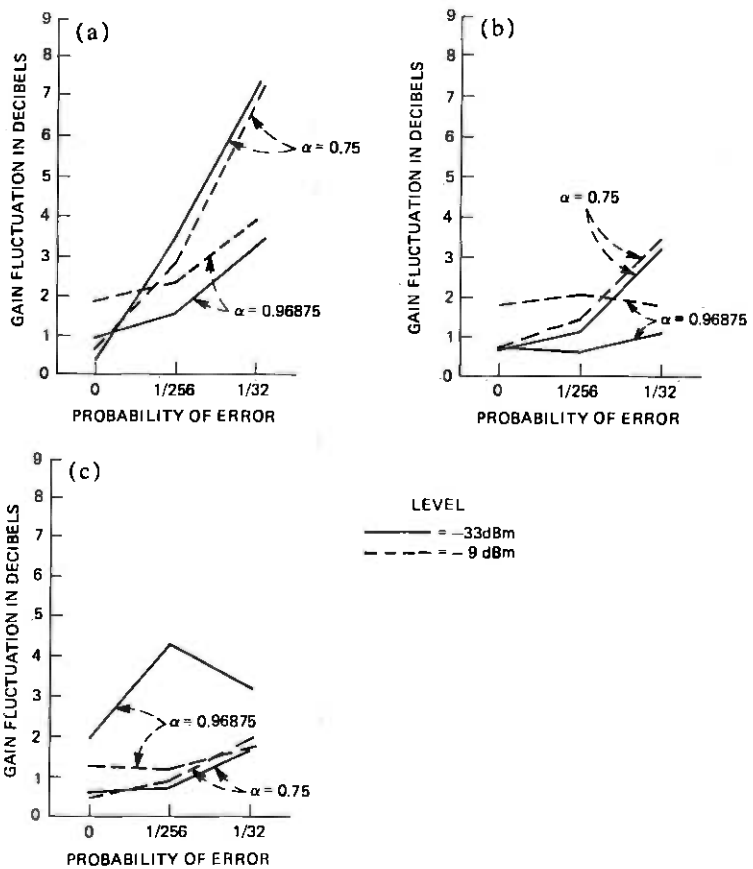


Fig. 5—Gain fluctuation, averaged across bit rate, as a function of probability of error: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

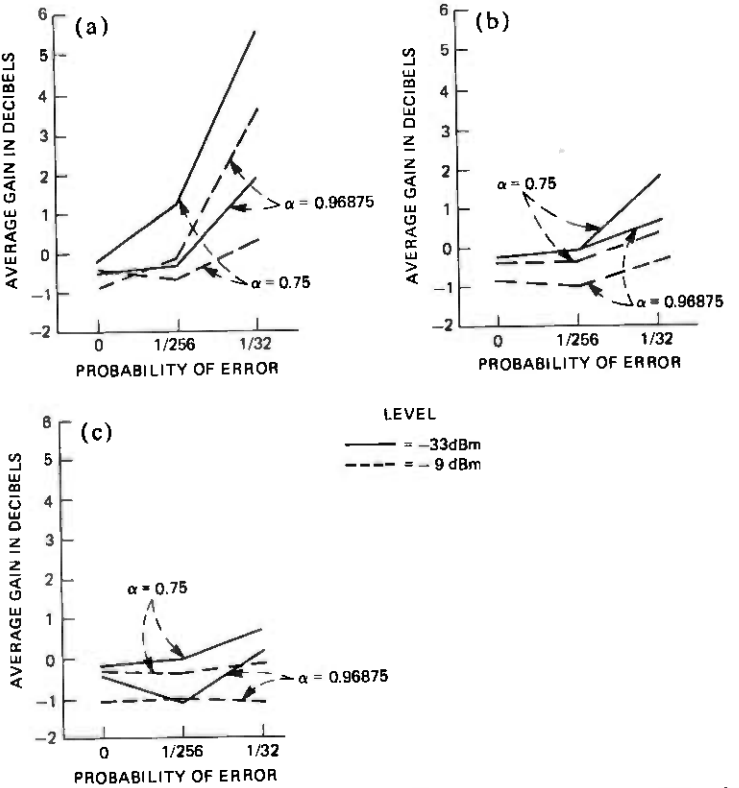


Fig. 6—Average gain, averaged across bit rate, as a function of probability of error: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

a general case the two measures are independent each of the other, in this particular experiment a high value of σ_g is normally associated with a high average gain. This is particularly true if the input level is low, because, owing to the channel errors, the output level tends to be increased. If the input level is high and the quantizer step size is close to its maximum value, the output level is likely to increase only to a smaller extent.

The form of the relationship between the two level mismatching measures and the loss in quality due to the sensitivity of the adaptation algorithms to channel errors was not clear *a priori*, and therefore various nonlinear transformations were tried on those measurement data. No transformation on gain fluctuation proved useful in regression equations, while a compression of the average gain, given by

$$\tilde{G} = G/\sqrt{|G|}, \tag{27}$$

gave better predictions than G, when associated with the other performance measures.

Table II reports the results of the regression analysis. The prediction accuracy is indicated by both the correlation coefficient between the true and predicted subjective scores and the rms prediction error. After each regression analysis, however, a goodness-of-fit test was performed to test normality of prediction errors. The Kolmogorov-Smirnov test[8] was used and in each case the hypothesis of normal distribution was accepted at the 0.20 significance level.

Table II—Formulas for predicting ratings using objective measures

	Formula for Predicting Rating	Correlation	rms error
1	$\hat{R} = 0.156\ \text{SNR} + 2.702$	0.667	1.202
2	$\hat{R} = 0.247\ \text{SNR}_{\text{seg}} + 1.369$	0.873	0.787
3	$\hat{R} = 0.271\ \text{SNR}_{\text{seg}} + 0.260\bar{G} + 1.071$	0.881	0.765
4	$\hat{R} = 0.274\ \text{SNR}_{\text{seg}} + 0.110\sigma_g + 0.767$	0.875	0.780
5	$\hat{R} = 0.302\ \text{SNR}_{\text{seg}} + 0.270\bar{G} + 0.122\sigma_g + 0.393$	0.883	0.756
6	$\hat{R} = 0.336\ \text{SNR}_{\text{com}} - 0.486$	0.911	0.665
7	$\hat{R} = 0.316\ \text{SNR}_{\text{com}} - 0.394\bar{G} - 0.228$	0.935	0.571
8	$\hat{R} = 0.292\ \text{SNR}_{\text{com}} - 0.179\sigma_g + 0.569$	0.923	0.621
9	$\hat{R} = 0.297\ \text{SNR}_{\text{com}} - 0.337\bar{G} - 0.088\sigma_g + 0.254$	0.938	0.561
10	$\hat{R} = 0.389\ \text{SNR}_{\text{max}} - 1.529$	0.913	0.656
11	$\hat{R} = 0.367\ \text{SNR}_{\text{max}} - 0.335\bar{G} - 1.212$	0.931	0.590
12	$\hat{R} = 0.343\ \text{SNR}_{\text{max}} - 0.160\sigma_g - 0.449$	0.923	0.622
13	$\hat{R} = 0.345\ \text{SNR}_{\text{max}} - 0.283\bar{G} - 0.088\sigma_g - 0.668$	0.933	0.581
14	$\hat{R} = 0.267\ \text{SNRF}_1 + 3.906$	0.878	0.773
15	$\hat{R} = 0.251\ \text{SNRF}_1 - 0.295\bar{G} + 3.920$	0.891	0.731
16	$\hat{R} = 0.219\ \text{SNRF}_1 - 0.268\sigma_g + 4.606$	0.909	0.673
17	$\hat{R} = 0.216\ \text{SNRF}_1 - 0.150\bar{G} - 0.235\sigma_g + 4.527$	0.912	0.662
18	$\hat{R} = 0.238\ \text{SNRF}_2 + 3.290$	0.887	0.744
19	$\hat{R} = 0.224\ \text{SNRF}_2 - 0.272\bar{G} + 3.339$	0.899	0.708
20	$\hat{R} = 0.197\ \text{SNRF}_2 - 0.249\sigma_g + 4.050$	0.914	0.656
21	$\hat{R} = 0.195\ \text{SNRF}_2 - 0.140\bar{G} - 0.219\sigma_g + 3.983$	0.916	0.647
22	$\hat{R} = 0.307\ \text{SNRF}_5 + 9.836$	0.855	0.838
23	$\hat{R} = 0.286\ \text{SNRF}_5 - 0.386\bar{G} + 9.499$	0.879	0.770
24	$\hat{R} = 0.244\ \text{SNRF}_5 - 0.323\sigma_g + 9.461$	0.905	0.688
25	$\hat{R} = 0.243\ \text{SNRF}_5 - 0.194\bar{G} - 0.278\sigma_g + 9.318$	0.909	0.671
26	$\hat{R} = -6.514\ D_I + 7.613$	0.797	0.975
27	$\hat{R} = -6.035\ D_I - 0.474\bar{G} + 7.345$	0.837	0.883
28	$\hat{R} = -4.994\ D_I - 0.388\sigma_g + 7.737$	0.878	0.772
29	$\hat{R} = -4.977\ D_I - 0.225\bar{G} - 0.334\sigma_g + 7.592$	0.885	0.751
30	$\hat{R} = -0.519\ D_{II} + 9.154$	0.826	0.910
31	$\hat{R} = -0.482\ D_{II} - 0.423\bar{G} + 8.788$	0.856	0.834
32	$\hat{R} = -0.404\ D_{II} - 0.306\sigma_g + 8.797$	0.868	0.800
33	$\hat{R} = -0.405\ D_{II} - 0.262\bar{G} - 0.241\sigma_g + 8.647$	0.878	0.773
34	$\hat{R} = 0.436\ \text{SDR}_1 - 1.268$	0.850	0.849
35	$\hat{R} = 0.407\ \text{SDR}_1 - 0.442\bar{G} - 0.923$	0.883	0.757
36	$\hat{R} = 0.348\ \text{SDR}_1 - 0.281\sigma_g + 0.518$	0.885	0.751
37	$\hat{R} = 0.353\ \text{SDR}_1 - 0.304\bar{G} - 0.203\sigma_g + 0.254$	0.897	0.712
38	$\hat{R} = 0.947\ \text{SDR}_2 - 0.249$	0.874	0.784
39	$\hat{R} = 0.893\ \text{SDR}_2 - 0.513\bar{G} - 0.028$	0.917	0.642
40	$\hat{R} = 0.766\ \text{SDR}_2 - 0.306\sigma_g + 1.331$	0.918	0.641
41	$\hat{R} = 0.785\ \text{SDR}_2 - 0.353\bar{G} - 0.212\sigma_g + 0.998$	0.934	0.578

Among the objective performance measures taken singly, the best one turns out to be SNR_{max}, with a correlation coefficient of 0.913 and an rms error of 0.656 [formula 10 in Table II]. The compensated signal-to-noise ratio SNR_{com} gives almost the same results, while all the other measures achieve a correlation lower than 0.9. In particular, the conventional, long-term, signal-to-noise ratio has a correlation of only 0.667 and an rms prediction error almost double that of SNR_{max}. The log likelihood ratio D_I is the second-worst predictor when used singly, with a correlation of only 0.797.

When the two measures of level mismatching, i.e., the average gain and the gain fluctuation, are included in the quality prediction formulas, the prediction accuracy is significantly improved, the rms prediction error having a 16-percent decrease on the average. The smallest improvement is displayed by SNR_{seg}. Among all the other measures, SNR_{com} gives the best prediction when combined with $\bar{G}$ and σ_g (formula 9), with a correlation of 0.938 and an rms error of 0.561, about one-quarter of a category. Formulas 13 and 41, which use SNR_{max} and SDR_2, are almost as good as formula 9. The frequency-weighted SNRs also give a fairly good prediction, with correlations over 0.9 and the remaining frequency domain measures, D_I, D_B, and SDR_1 give a slightly poorer prediction.

VIII. DISCUSSION

8.1 *Effects of coder design parameters*

The subjective data have displayed complicated interactions among all the experimental design variables, the strongest interaction being the one between the adaptation constant α and the decay constant β. In fact, each of these two parameters affects different phenomena:

(*i*) The *dynamic range* is reduced when β decreases from unity, but this reduction does not seem to be perceptible for any β if $\alpha = 0.75$. If $\alpha = 0.96875$ and $\beta = 255/256$, a certain reduction in the dynamic range begins to be perceived, producing a loss in quality at the low level of about 1.5 points with respect to the high level. When $\alpha = 0.96875$ and $\beta = 63/64$, the dynamic range is reduced still further, and the loss in quality of the low level with respect to the high one is very large, the average score dropping down from 5.3 to 2.4.

(*ii*) The *effect of the errors* is smaller when α increases or β decreases. For instance, with $\alpha = 0.75$, the loss in quality passing from $P(e) = 0$ to $P(e) = 1/256$ averages 0.49 when $\beta = 63/64$, while it averages 1.59 for $\beta = 1$ and 0.93 for $\beta = 255/256$.

(*iii*) The difference in the effect of the errors between the two input levels is higher for faster adaptation. For instance, passing again from $P(e) = 0$ to $P(e) = 1/256$, the difference between the losses at the two levels averages 1.0 when $\alpha = 0.75$, while it averages 0.45 when $\alpha = 0.96875$.

8.2 Optimum coders

Given a fixed number of bits per sample, a combination of decay constant β and adaptation constant α provides the best output quality for a given probability of error.

In the case of error-free transmission, optimum quality should be attained with no error dissipation mechanism, i.e., $\beta = 1$ which produces theoretically infinite dynamic range. The parameter α is not very critical in that case.[5]

When the coder operates under noisy conditions and the probability of error is in the range of the values encountered in a normal telephone connection or even higher than that (as is the case of $P(e) = 1/256$), a very slow error dissipation associated with fast adaptation provides good robustness to channel errors, without impairing the dynamic range. Actually, in this experiment, the combination $\beta = 255/256$ and $\alpha = 0.75$ provided optimum performance even under error-free conditions.

If the probability of error is as great as 1/32, more typical of mobile radio communications, the best compromise between dynamic range and error sensitivity is obtained by a slow adaptation constant, combined again with a slow error dissipation rate. However the use of faster error dissipation and faster adaptation could be almost as good for this very high error rate.

8.3 Objective measures

One aim of the experiment was to examine a certain number of objective measures of coder performance and to compare them in the light of the actual subjective quality ratings, obtained under very different conditons. With the results of correlation and regression analyses reported in Table II, it is possible to observe strengths and weaknesses of the different measures and to derive general indications on which are the desirable properties of an objective quality measure.

A first indication that emerges from the experimental data is that the conventional long-term SNR is a very poor indicator of the quality of ADPCM coders under noisy channel conditions; this confirms the results obtained with PCM[18] and ADPCM coders[5] in the case of error-free transmission. Therefore, the use of SNR can be completely misleading, when comparing different coders operating under noisy conditions. A noticeable improvement in prediction accuracy is obtained simply by

measuring signal-to-noise ratio segmentally. Being time-segmental is a necessary property of any successful objective quality measure of coded speech.

Table II demonstrates also that, when a coder incorporating an adaptive quantizer is operating under noisy channel conditions, an objective performance measure must not be sensitive to changes or fluctuations of the output speech level. These fluctuations can be measured separately and the value obtained can be combined with the performance measure to improve the accuracy of the subjective quality prediction. It should be noticed, for instance, that $\bar{G}$ and σ_g have a positive coefficient when combined with SNR_{seg} (formula 5 in Table II). This indicates that the level mismatching is weighted too much in SNR_{seg}, which did not incorporate any level compensation.

An important consideration on the subject of objective quality measures is that results from recent experiments[7,9] indicate that frequency-weighted signal-to-noise ratios improve the prediction accuracy, especially when largely different noise spectra are produced by the coders.[7,9] In this experiment, actually the frequency-weighted SNRs did not predict the subjective ratings as accurately as the simpler level-compensated SNRs, namely, SNR_{com} and SNR_{max}. This may depend on the fact that the level compensation is effected by minimizing the rms error on the whole bandwidth; this fact may worsen the measure in same band. If this is the case, it would be a weakness of the frequency-weighted SNR's when measuring coder performance in the presence of channel errors.

A final remark on frequency domain measures: These measures are more general than time domain ones because they are insensitive to short delays or to phase distortion.[9,14] Therefore, they are more easily applicable to the test of coders whose input and output signals are in analog form or which include digital filters. On the other hand, the performance measures used here incorporate a spectral noise weighting that cannot be directly controlled. However, it is encouraging to see that the newly defined spectral signal-to-distortion ratio, SDR_2, provides a very good prediction of subjective ratings when combined with the level mismatching measures.

More work is needed in the field of objective prediction of coder quality. In particular, frequency weighted SNRs should be evaluated more carefully, to derive the appropriate frequency weighting mechanism. In addition, the difference in behavior of the various spectral distortion measures need to be analyzed in more depth.

8.3.1 Estimation of the subjective effect of level mismatching

8.3.1.1 A Physical Interpretation of Prediction Formulas. The results of Table II lend themselves to a nice interpretation. The quality of an

adaptive coder operating in a noisy channel environment (high probability of error) may be considered as composed of two terms: (*i*) "intrinsic" goodness of the speech reproduction, which takes into account the noise due to the coding and to the errors, but not the level mismatching, (*ii*) loss in quality due to the level variations caused by the sensitivity of the adaptation algorithm to channel errors. In formulas, we can write

$$\hat{R} = \hat{R}_I - \hat{R}_L. \tag{28}$$

$\hat{R}_I$ can basically be estimated by any of the performance measures which incorporate gain compensation or, in any case, which are not sensitive to alterations in the output signal level. $\hat{R}_L$ is instead estimated by a linear combination of the gain fluctuation and the average gain, modified according to eq. (27).

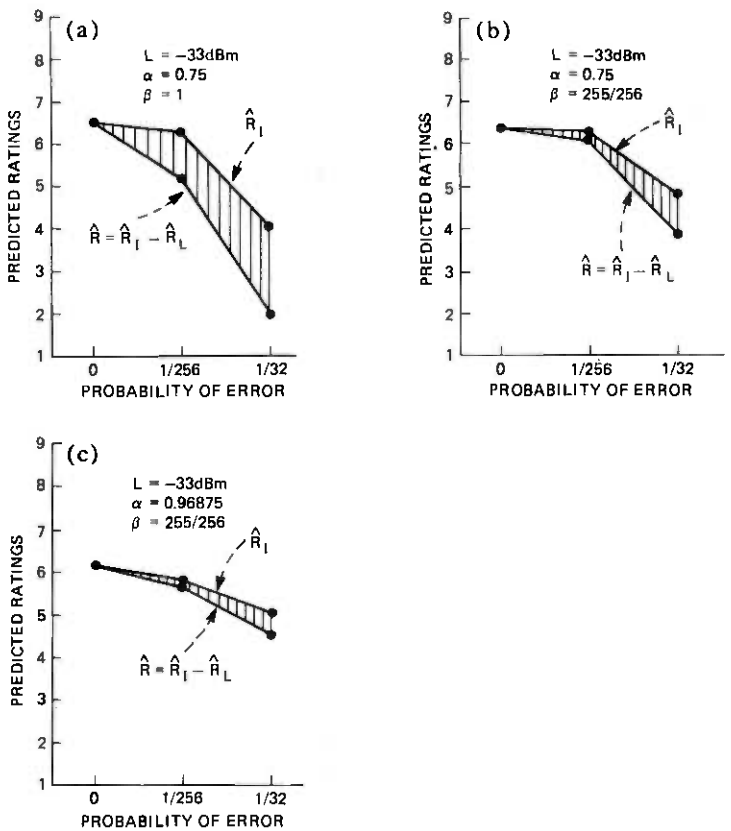


Fig. 7—Predicted overall rating $\hat{R}$ and "intrinsic" goodness $\hat{R}_I$, as a function of probability of error, for 4-bit ADPCM and different combinations of design parameters: (a) $\alpha = 0.75$, $\beta = 1$. (b) $\alpha = 0.75$, $\beta = 255/256$. (c) $\alpha = 0.96875$, $\beta = 255/256$.

8.3.1.2 An Example. In the light of the interpretation given in the previous section, it is possible to give at least a qualitative answer to the question on which is the subjective effect of level mismatching. Figure 7 shows the predicted overall rating $\hat{R}$ and the "intrinsic" goodness $\hat{R}_I$ as a function of the probability of error, for the 4-bit ADPCM coder with low level input and three different combinations of design parameters. $\hat{R}$ was computed according to formula 41 in Table II, while $\hat{R}_I$ was computed discarding the terms involving $\bar{G}$ and σ_g from the same formula:

$$\hat{R}_I = 0.785\ SDR_2 + 0.998.$$

In the case of fast adaptation and absence of error recovery (Fig. 7a), the loss in quality due to level mismatching can be estimated as half category (1 point) for the intermediate error rate and 1 category (2 points) for the high error rate. In the case of fast adaptation and slow error recovery (Fig. 7b), the loss $\hat{R}_L$ is instead reduced to about only half category (1 point) for the high error rate.

Finally, when slow adaptation and slow error recovery are used, simultaneously (Fig. 7c), the loss due to level mismatching can be estimated as only about 0.4 point, i.e., less than a quarter of a category.

IX. ACKNOWLEDGMENTS

The author wishes to thank B. J. McDermott, B. H. Bharucha, R. E. Crochiere, and J. M. Tribolet for several useful discussions and for having provided subroutines for computing some of the objective measures. The author thanks also J. Coker for recruiting the subjects and running the listening tests.

REFERENCES

1. D. J. Goodman and R. M. Wilkinson, "A Robust Adaptive Quantizer," IEEE Trans. Commun., *COM-23*, No. 11 (November 1975), pp. 1362–1365.
2. C. Scagliola, "An Adaptive Speech Coder with Channel Error Recovery," International Conference on Communications, Chicago, Ill., June 1977.
3. D. Mitra, "An Almost Linear Relationship Between the Step-Size Behavior and the Input Signal Intensity in Robust Adaptive Quantization," IEEE Trans. Commun., *COM-27*, No. 3 (March 1979).
4. P. Castellino, G. Modena, L. Nebbia, and C. Scagliola, "Bit Rate Reduction by Automatic Adaptation of Quantizer Step-size in DPCM Systems," International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.
5. B. J. McDermott, C. Scagliola, and D. J. Goodman, "Perceptual and Objective Evaluation of Speech Processed by Adaptive Differential PCM," B.S.T.J., *57*, No. 5 (May–June 1978), pp. 1597–1618.
6. P. Noll, "Adaptive Quantizing in Speech Coding Systems," International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.
7. R. E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A Study of Objective Measures of Speech Waveform Coders," International Zurich Seminar, Zurich, Switzerland, March 1978.
8. A. A. Afifi and S. P. Azen, *Statistical Analysis a Computer Oriented Approach*, New York: Academic Press, 1972.

9. J. M. Tribolet, P. Noll, B. J. McDermott, and R. E. Crochiere, "Complexity vs. Quality for Speech Waveform Coders," IEEE International Conference on Acoustics Speech and Signal Processing, Tulsa, Oklahoma, April 1978.
10. N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Amer. *19* (January 1947), pp. 90–119.
11. J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, New York: Springer Verlag, 1976.
12. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust. Speech and Sig. Proc., *ASSP-23*, (February 1975), pp. 67–72.
13. A. H. Gray and J. D. Markel, "Distance Measures for Speech Processing," IEEE Trans. Acoust. Speech Sig. Proc., *ASSP-24*, (October 1976), pp. 380–391.
14. B. H. Bharucha, "An Objective Measure of Codec Speech Quality," unpublished paper, 1976.
15. N. S. Jayant, "Adaptive Quantization with One-Word Memory," B.S.T.J., *52*, No. 7, (September 1973), pp. 1119–1144.
16. P. Slater, "Analysis of Personal Preferences," Brit. Journal of Statistical Psychology, *13* (November 1960), pp. 119–135.
17. J. D. Carroll, "Individual Differences and Multidimensional Scaling," in *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. I, Shepard, Romney, Nerlove (Eds.), New York: Seminar Press, 1972, pp. 105–155.
18. D. J. Goodman, B. J. McDermott, and L. H. Nakatani, "Subjective Evaluation of PCM Coded Speech," B.S.T.J., *55*, No. 8, (October 1976), pp. 1087–1109.
19. D. L. Richards, *Telecommunications by Speech*, London: Butterworths, 1973, Ch. 4.
20. L. Nebbia and P. Usai, "Influence of Some Types of Noise on Telephone Digital Transmissions," Symposium, "Speech Intelligibility," Liege, November 1973.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July-August, 1979
Printed in U.S.A.

An Adaptive Intraframe DPCM Codec Based Upon Nonstationary Image Model

By N. F. MAXEMCHUK and J. A. STULLER

(Manuscript received November 21, 1978)

This paper introduces a nonstationary model for images and develops an adaptive intrafield DPCM *codec based upon the model. The codec attempts to minimize the mean-square coding error at each sample point in the picture. The quantizer in the resulting adaptive codec is found to be similar to that previously obtained from visual masking considerations. Comparative simulation results using 256 × 256 pixel rasters are given for two- and three-bit/pixel versions of the adaptive codec, the three-bit/pixel Graham codec, and three-bit/pixel previous element* DPCM.

I. INTRODUCTION

This paper introduces a nonstationary model for images and develops an adaptive intrafield codec based upon the model. The codec adaptively estimates both the mean and probable range of values of the next picture sample to be encoded and adapts the predictor and quantizer accordingly. In so doing, the coder attempts to minimize the mean-square coding error (MMSE) at each sample point in the picture. The MMSE distortion measure is generally acknowledged to be a poor indicator of image quality.[1] However, when it is applied on a *point* (rather than area) basis in conjunction with the image model presented here, the coder adaptation and resulting coding quality are found to be comparable to that previously obtained from visual masking considerations. This result follows from a property of human vision, stressed by Graham,[2] concerning the strong connection that exists between image chaos (unpredictability) and the visual system's tolerance to noise-like coding distortion. Because of this property, we obtain good image quality at two bits per pel and excellent quality at three bits per pel in a DPCM codec designed solely using the MMSE criterion—masking phenomena are in large part accounted for automatically when the

source model more adequately represents actual images and when the distortion criterion is applied on a point basis.

II. SOURCE MODEL

This section introduces a nonstationary causal source model for the intrafield video process that will be used to develop the adaptive predictive intrafield codec of Section IV. Motivation for the model is intuitive and follows from an examination of a representative video signal of the type we wish to encode, such as that shown in Fig. 1. This is a frame of two interlaced fields, each having 256 pixels per line and 128 lines, with amplitudes stored as 8-bit quantities. The essential characteristic of this (and any) image is that it is a projective transformation of a collection of physical objects. As a consequence, the image is partitioned into regions of luminance elements whose amplitudes are interrelated by the physical structure of the objects they represent. The result is an array of pixels composed of distinct regions having slowly varying "brightness" and "texture" with abrupt boundaries (the picture outline) separating one region from another. We find it natural to view this array as a field that is partitioned into regions of independent quasi-stationary subfields. Two underlying random phenomena are involved: the random amplitudes of picture elements within a

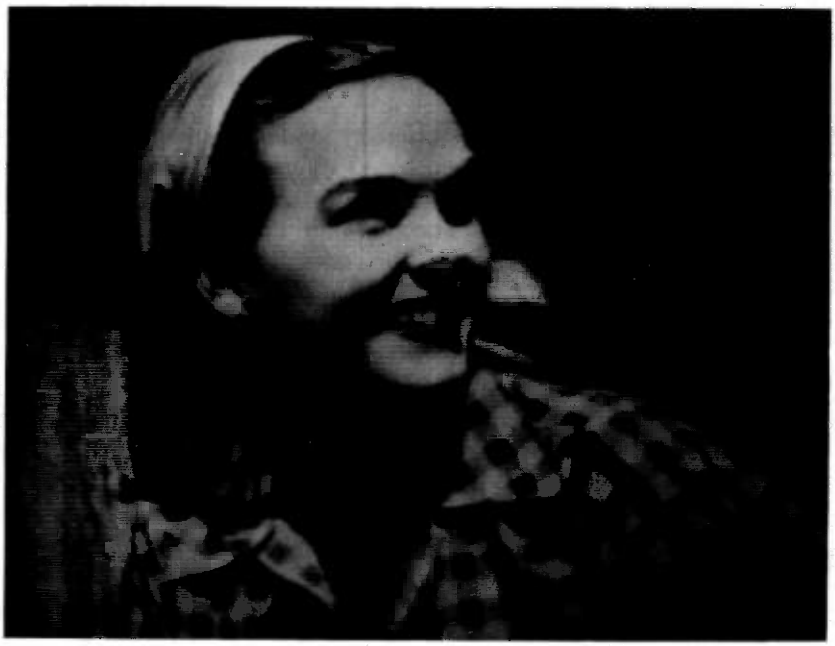


Fig. 1—Checker girl original.

given subfield and the random selection of the subfield with respect to raster coordinates. A source model that incorporates both phenomena is shown in Fig. 2.

Figure 2 models the image generation process as a composite of Q autoregressive sources, $q = 1, 2, \cdots, Q$, and one white source, $q = 0$. Switches S1 and S2 determine which source generates output luminance $s_{mn} \equiv s_t$, where m and n are, respectively, the line number in the field and the column number of the pixel and t is the time the pixel is encountered during conventional line scanning. The autoregressive sources, characterized by predictors 1 through Q and "innovations" process $w_{mn} \equiv w_t$, provide a set of Q possible processes from which the regions of slowly varying brightness and texture of a subfield in an actual image can be approximated. The random variables w_t are assumed zero mean, independent, and characterized by a single known probability density function (pdf) $g(w)$. The predictors F_q in sources $q = 1, 2, \cdots, Q$ are taken to be linear functions of pixels from the local past neighborhood of coordinate $(m, n) \equiv t$. Section IV discusses the specific predictors chosen for the codec of this paper (Table I). Source 0 models those pixels of an actual image that either have no structural relation to previous pixels or whose relation to these pixels is not adequately modeled by sources 1 through Q. Such pixels tend to occur in highly chaotic regions of the image and at certain boundaries at which new subfields are initiated. Since this source represents the extreme of chaos possible in an image, its output is taken to be a

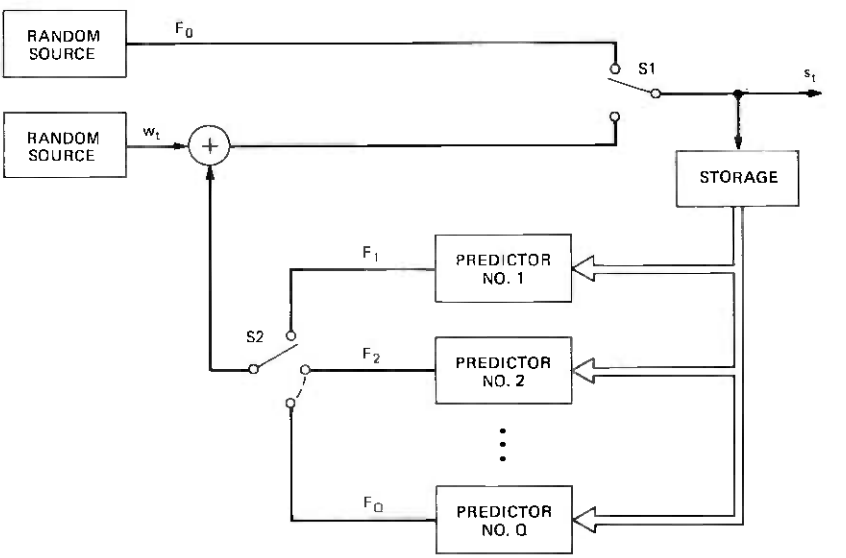


Fig. 2—Source model. Switches S1 and S2 are governed by eq. (2) in the text.

sequence of random variables $F_0 \equiv F_0(m, n)$, each uniformly distributed over [0, 255].

Switches S1 and S2 of the model determine which source is used for final output at each raster coordinate and thus determine image outline as well as more subtle structural changes. "Outline" and "subtle structural changes" are subjectively perceived qualities of an image that are difficult to quantify probabilistically. However, in an actual image "structure" tends to vary slowly: boundaries between regions are an exception, but even here the discontinuity is generally only along one dimension. This suggests that probabilistic information regarding the source in operation at time t can be inferred by appropriate processing of the pixels in the local past vicinity (in the same field) of the pixel in question. To arrive at a source model that characterizes this quasistationary in the simplest way, we model the image source as choosing sources, $q = 0, 1, 2, \cdots, Q$, *independently* according to *unknown* first-order probabilities $P[q; (m, n)]$ that are slowly varying functions of coordinates (m, n). We further assume that the Q textures generated by sources $q = 1, 2, \cdots, Q$ are *a priori* equally likely for a *random* choice of coordinate (m, n):

$$E\{P[q; (m, n)]\} = c \quad q = 1, 2, \cdots, Q \tag{1a}$$

and

$$E\{P[q; (m, n)]\} = \epsilon \ll c \quad q = 0, \tag{1b}$$

where c and ϵ are constants satisfying $\epsilon + Qc = 1$ and the expectation is taken over the raster coordinates.

An alternative approach would be to model the sequence of q_{mn} as stationary Markov. However, this approach was not taken since the assumption of nonstationary *independent* q leads to a relatively simple codec that is robust with respect to both varied picture inputs and channel errors. The assumption of equality of expectations in (1b) leads to mini-max performance with respect to variations in textural content of the picture to be encoded. Biasing this *a priori* distribution toward one predictor would make the codec more susceptible to poor performance on a picture which does not match this distribution. The problem faced by the codec is to estimate probabilities $P[q; (m, n)]$ by suitable processing of past pixel outputs and use the estimates to best advantage for bandwidth compression.

To summarize, the proposed model of the video process has the form (Fig. 2)

$$s_t = \begin{cases} F_0, \text{ with probability } P(0; t) \\ F_q + w_t, \text{ with probability } P(q; t), \quad 1 \le q \le Q, \end{cases} \tag{2}$$

where F_0 is an independent random variable uniform over [0, 255], F_q is a given linear function of pixels in the local past neighborhood of s_t

(Table I), and w_t is an independent zero mean random variable characterized by probability density function (pdf) $g(w)$. Probabilities $p(q;\ t)\ q = 0,\ 1,\ \cdots,\ Q$ vary slowly with respect to at least one coordinate of the raster and satisfy (1); otherwise, these probabilities are unknown. Note that the model embeds the elusive variety of gross image structure in the unknown probabilities $P(q;\ t)$, $0 \leq q \leq Q$. These represent the probabilistic information that the encoder hopes to learn by suitable processing of past image source outputs.

Figure 3 illustrates a representative output generated by the model. In obtaining this output, $g(w)$ was assumed Laplacian, and the $P(q;\ t)$ were estimated from the image of Fig. 1 by a procedure described in Section III. Significant increases in structural similarity to Fig. 1 are possible by modeling the sequence of w_t as nonstationary. In the interest of codec simplicity, however, this additional complexity is not included in the model.

III. ANALYTICAL DEVELOPMENTS

The design of the DPCM codec of Section IV requires specification of both the quantizer and the predictor. Complete statistical information pertinent to this design is contained in the conditional pdf of s_t given

Fig. 3—Representative output of source model. Output of the source model of Fig. 2, where w_t is Laplacian and probabilities $P(q;\ t)$ of eq. (2) are estimated from Fig. 1.

the set of past pixels $\{s_t, t < t\} \equiv S_t^-$. This section describes how this conditional pdf can be etimated at the source output.

It can be easily shown that, for *given* $P(q; t)$, $q = 0, 1, \cdots, Q$, the probability density of s_t conditioned on S_t^- is (for $0 \le s_t \le 255$)

$$p(s_t \mid S_t^-) = \frac{P(0; t)}{255} + \sum_{q=1}^{Q} g(s_t - F_q(S_t^-))P(q; t), \tag{3}$$

where $F_q(S_t^-)$ denotes the qth predictor F_q of s_t as an explicit function of past pixels S_t^-. An estimate of density function (3) is obtained by replacing $P(q; t)$ in the above by its estimate, as described below.

Let the number of times the qth source had been output in a local past region R_t of N points neighboring $(m, n) = t$ (Fig. 4) be denoted by $n(q)$. Due to the nature of the source model, $n(q)$ cannot be measured at the source output. However, a reasonable and computable approximation to it is given by the expectation $E\{n(q)|S_t^-\}$, where the expectation assumes a random selection of (m, n) and is over the density (w_t). By the quasi-stationarity of $P(q; t)$, we then set

$$\hat{P}(q; t) = \frac{E\{n(q) \mid S_t^-\}}{N}, \tag{4}$$

which becomes (appendix)

$$\hat{P}(q; t) = \frac{1}{N} \sum_{j=1}^{N} \hat{P}(q_j = q \mid S_t^-), \tag{5}$$

where $\hat{P}(q_j = q \mid S_t^-)$ is the conditional probability that the jth pixel in R_t (Fig. 4) was output by source q based upon *a priori* probabilities $E\{P[q; (m, n)]\}$ of (1). Further manipulations (appendix) give

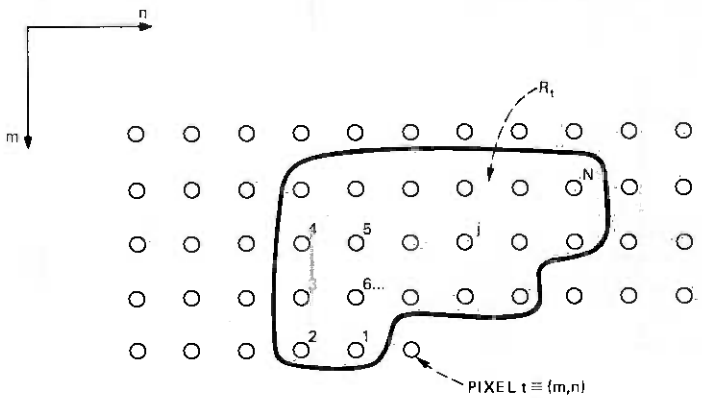


Fig. 4—Illustration of region R_t. This region consists of N pixels in a local past vicinity of pixel (m, n). Note that the numbering of coordinates $j = 1, 2, \cdots, N$ is arbitrary, as are the region boundaries.

$$\hat{P}(q;\, t) = K \frac{\epsilon}{255} \quad q = 0 \tag{6a}$$

$$\hat{P}(q;\, t) = \frac{KC}{N} \sum_{j=1}^{N} g(e_q(t;\, j)) \quad 1 \le q \le Q, \tag{6b}$$

where K satisfies

$$\sum_{q=0}^{Q} \hat{P}(q;\, t) = 1. \tag{7}$$

Equation (5) interprets $\hat{P}(q;\, t)$ as an arithmetic average of *a posteriori* probabilities of q over region P_t, and eq. (6) show how this average can be computed. The term $e_q(t;\, j)$ in (6b) is the difference between the actual value of the jth pixel in region R_t and the predicted value of this pixel given by predictor F_q and is therefore the implied value of the jth innovations variable in R_t under the hypothesis that predictor q was in operation at the source. Explicitly,

$$e_q(t;\, j) = s_t^j - F_q(S_t^{j-}), \tag{8}$$

where s_t^j is the jth pixel in R_t and S_t^{j-} is the set of pixels previous to s_t^j. Equation (6b) estimates $\hat{P}(q;\, t)$, by summing the relative probabilities of the innovations implied under the hypothesis that source q was in operation over region R_t. Note that if $g(\cdot)$ has its peak at zero, then $\hat{P}(q;\, t)$, $1 \le q \le Q$, will be large for those q corresponding to small prediction error $e_q(t,\, j)$ over the N point region. If none of the Q predictors is consistent with past local data, then all terms in the sum of (5b) will be small for $1 \le q \le Q$, and the normalization in (6) will make $\hat{P}(0;\, t)$ large. Further description of (4) to (7) is included in the derivation in the appendix.

The codec described in Section IV predicts s_t by the estimated mean of predictable source outputs:

$$\hat{s}_t = \frac{\sum_{q=1}^{Q} F_q(S_t^-)\hat{P}(q;\, t)}{\sum_{q=1}^{Q} \hat{P}(q;\, t)}. \tag{9}$$

An important characteristic of this prediction rule is its insensitivity to small variations in data S_t^- regardless of the relative values of N and Q. This is in contrast to the covariance method in linear prediction described in a review paper by Makhoul[3] in which small sample size can lead to an ill-conditioned system of equations whose inversion is the adapted predictor. Since (9) is a weighted average of stable (and generally good) estimates F_q, stability persists even for $N < Q$, and some thought indicates that the resulting prediction of $\hat{s}_t$ works in an intuitively reasonable way even if N is only unity.

IV. THE CODEC

In this section, a codec resulting from the source model is described. A block diagram of the encoder is shown in Fig. 5. The codec has been used to code pictures using two and three bits per pel.

The encoder operates by forming Q estimates $F_q(X_t^-)$, $1 \leq q \leq Q$, of source output S_t based upon the previously reconstructed field elements X_t^-. Estimates of source probabilities $P(q; t)$, $0 \leq q \leq 1$, are made with eq. (6) to (7) using previously reconstructed pixels X_t^j in place of S_t^j. Estimates $F_q(X_t^-)$ and probabilities $P(q; t)$ are used to predict the next encoder input pixel s_t, according to (9) and the most likely distribution of values s_t according to (3), with X_t^- replacing S_t^-.

The encoder has been implemented using an $N = 4$ point learning region R_t (Fig. 6) and $Q = 6$ predictors. The predictors used are given in Table I.

Note that with these six predictors the form of predictor (7) can be any one of the most common fixed predictors used in intrafield coders. This varies over the picture so that the best predictor (or best weighted sum) considering the recent past will be used at each sample point.

The pictures which were encoded consisted of 256 lines in two interleaved fields and 256 samples per line. The previous line elements were taken from the previous line in the same field. In this environment, no advantage was obtained by including elements more than one line away in the estimates. Similarly, no visible improvement was obtained using elements that were more than two elements away on the same line. The slope estimator, F_5, and the planer estimator, F_6, were found to be particularly useful in the system which uses two bits

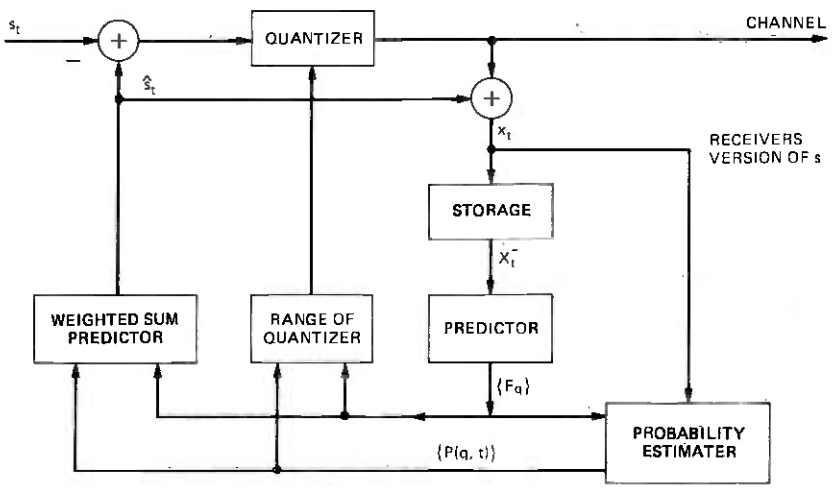


Fig. 5—The encoder.

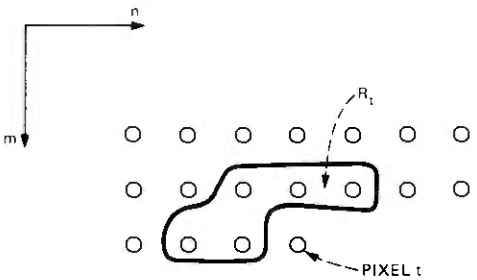


Fig. 6—Four-point region R_t used by codec.

Table I

$F_1(n, m) = \hat{x}(n-1, m)$.
$F_2(n, m) = \hat{x}(n-1, m-1)$.
$F_3(n, m) = \hat{x}(n, m-1)$.
$F_4(n, m) = \hat{x}(n+1, m-1)$.
$F_5(n, m) = 2\hat{x}(n-1, m) - \hat{x}(n-2, m)$.
$F_6(n, m) = \hat{x}(n-1, m) + \hat{x}(n, m-1) - \hat{x}(n-1, m-1)$.

per pel. These estimators allowed the coder to respond more quickly to edges within the picture, and reduced slope overload.

Ideally, the quantizer should be adapted at each point to the estimated probability distribution of s_t. In view of the complex form of (3), this type of redesign is not feasible, and the following ad-hoc curve-fitting technique was used to simplify the adaptation algorithm. The density function $g(w)$ was taken as Laplacian, $g(w) = \alpha/2 \exp(-\alpha \mid w \mid)$. The Max quantizer[4] for this was determined. Each side of the distribution (3) about the mean $\hat{s}_t$ was then approximated by an exponential distribution, and the axis was simply scaled appropriately in codec operation to place the quantization levels. The parameter of each exponential distribution was selected so that it had the same first moment about $\hat{s}_t$ as the corresponding portion of the actual distribution as described in Fig. 7.

When the estimated probability of occurrence of the random estimator is near zero and the estimators $q = 1$ through 6 are identical corresponding to a perfectly flat region in the picture, the parameter of the exponential defining the quantizer assumes its smallest value. In this situation, the parameter of the exponential defining the quantizer is approximately α, the parameter of the Laplacian distribution defining the innovation term in the model. Therefore, α determines the minimum values of the levels of the quantizer, and these, in turn, determine the amount of granularity due to quantization noise in flat regions of the picture and the ability of the coder to respond to unexpected edges. The smaller the value of α, the lower the granular quantization noise; the larger the value of α, the quicker the coder can

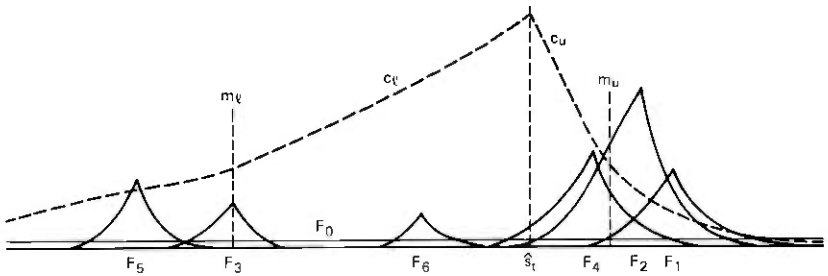


Fig. 7—Illustration of encoder's derivation of predictor and quantizer. F_0 refers to the distribution of the white source output; $\hat{s}_t$ is the weighted sum predictor; m_u and m_ℓ are the upper and lower first moments of the actual distribution about $\hat{s}_t$; and c_u and c_ℓ are the exponential distributions used to determine the quantizer.

respond to edges. Because of this interaction, the values of α were selected experimentally based upon visual examination of a sequence of coded pictures for the two- and three-bit/pel quantizers. For the two-bit/pel quantizer, α was selected so that the minimum value of the inner quantizer level is equal to two picture levels, when the picture is initially quantized into 256 levels. For the three-bit/pel quantizer, the inner quantization level was selected so that the inner quantization level is equal to one picture level.

The random variable F_0 in the source model of Fig. 2 is uniformly distributed over the range of possible values the sample can assume. In implementing the codec, it was found to be desirable to assume that the range of F_0 is somewhat reduced. Limiting the span of F_0 is particularly necessary in the system which transmits two bits per pel. This can be seen as follows. Assume that probability $P(0, t)$ is estimated by the encoder to be close to unity. In this situation, if F_0 has range [0, 255] the four quantization levels will be spread over the entire range of possible sample values. It is then likely that none of the estimators will be close to the reconstructed value x_t even though an estimator can have closely approximated the actual value s_t. Thus, the random estimator may be used for the next sample. This creates an instability in the coder which can propagate into flat regions of the picture. To eliminate this type of instability, the maximum range of the quantizer was limited. To be consistent with limiting the maximum range of the quantizer, the span of F_0 was limited to a symmetrical region about $\hat{s}_t$ of (9). The maximum span of the quantizer was also set experimentally. In the two-bit/pel system, the maximum span of the quantizer was set so that the inner level of the quantizer is eight picture levels. And in the three-bit/pel system, the maximum span of the quantizer was set so that the inner level in the quantizer is four picture levels. In the two-bit/pel system, there are only two quantization levels on each side of the predicted value. In this system, the maximum span of the

quantizer determined the ability of the encoder to track sudden changes in the picture. Therefore, it is necessary to make the maximum quantizer span as large as possible, without making the encoder unstable. In the three-bit/pel system, four quantization levels are on each side of the predictor. In this system, restricting the maximum quantizer span was necessary to prevent the quantizer span from frequently exceeding the range of possible picture levels and wasting quantization levels. This is why a smaller maximum value of the inner quantization level was selected for the three-bit/pel system than for the two-bit/pel system.

In Figs. 8 and 9, the quantization span for various parts of the picture in the two- and three-bit/pel systems is shown. In these pictures, the average of the upper and lower quantization spans is displayed. The white areas correspond to the smallest span of the quantizer and the black levels the largest span. It is interesting to note that the resulting quantizer adaptation is similar to that which would be expected if a masking function were used.[5] However, this quantizer adaptation was arrived at strictly by mathematical techniques, minimizing the expected point mean-squared error with a varying probability distribution of next sample values, rather than by the psychovisual considerations used to derive masking functions. This result is

Fig. 8—Quantizer range adaptation of the two-bit/pel codec.

Fig. 9—Quantizer range adaptation of the three-bit/pel codec.

consistent with Graham's early observations concerning the strong connection between image chaos and the visual system's tolerance to noise-like distortion.[2]

V. RESULTS

This adaptive predictor with a two- and three-bit/pel quantizer has been implemented and compared with an adaptive predictor using Graham's rule[2] and the three-bit/pel fixed quantizer suggested in the Graham paper, and a previous element DPCM encoder with a fixed three-bit/pel quantizer. Our two-bit/pel adaptive predictor has considerably less slope overload than the previous element predictor having three bit/pel quantizer, but is not quite as good as the Graham predictor having a three bit/pel quantizer. Our adaptive predictor with a three bit/pel quantizer has less slope overload than the Graham predictor with a three bit/pel quantizer. In addition, the estimates at edges within the picture are accurate enough to virtually eliminate the edge business in moving sequences which is characteristic of many adaptive predictors. To demonstrate these characteristics, the difference between the original picture of the checker girl, Fig. 1, and the result of processing by these four techniques is shown in Figs. 10 to 13.

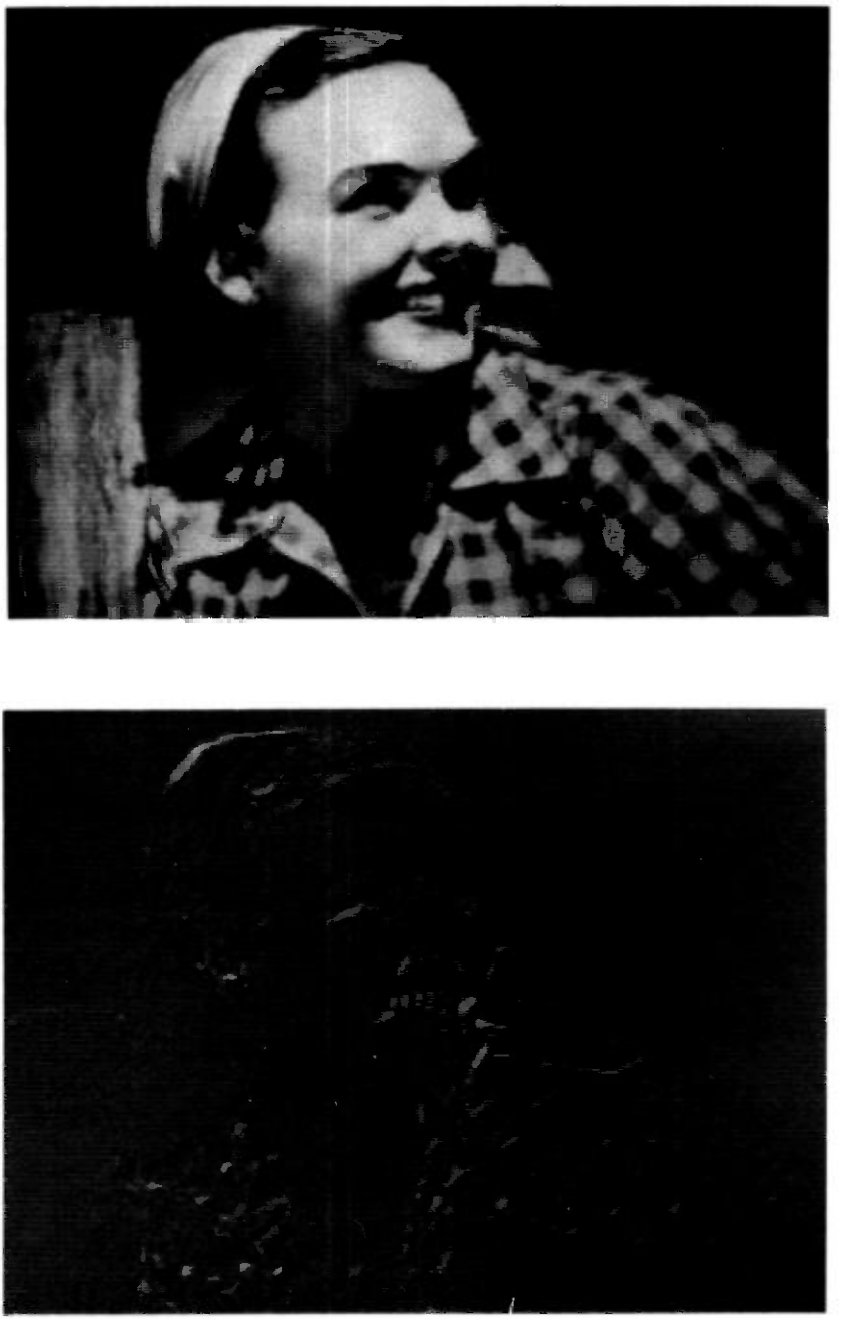


Fig. 10—Two-bit/pel codec performance. Top: Decoder output. Bottom: Difference between decoder output and original.

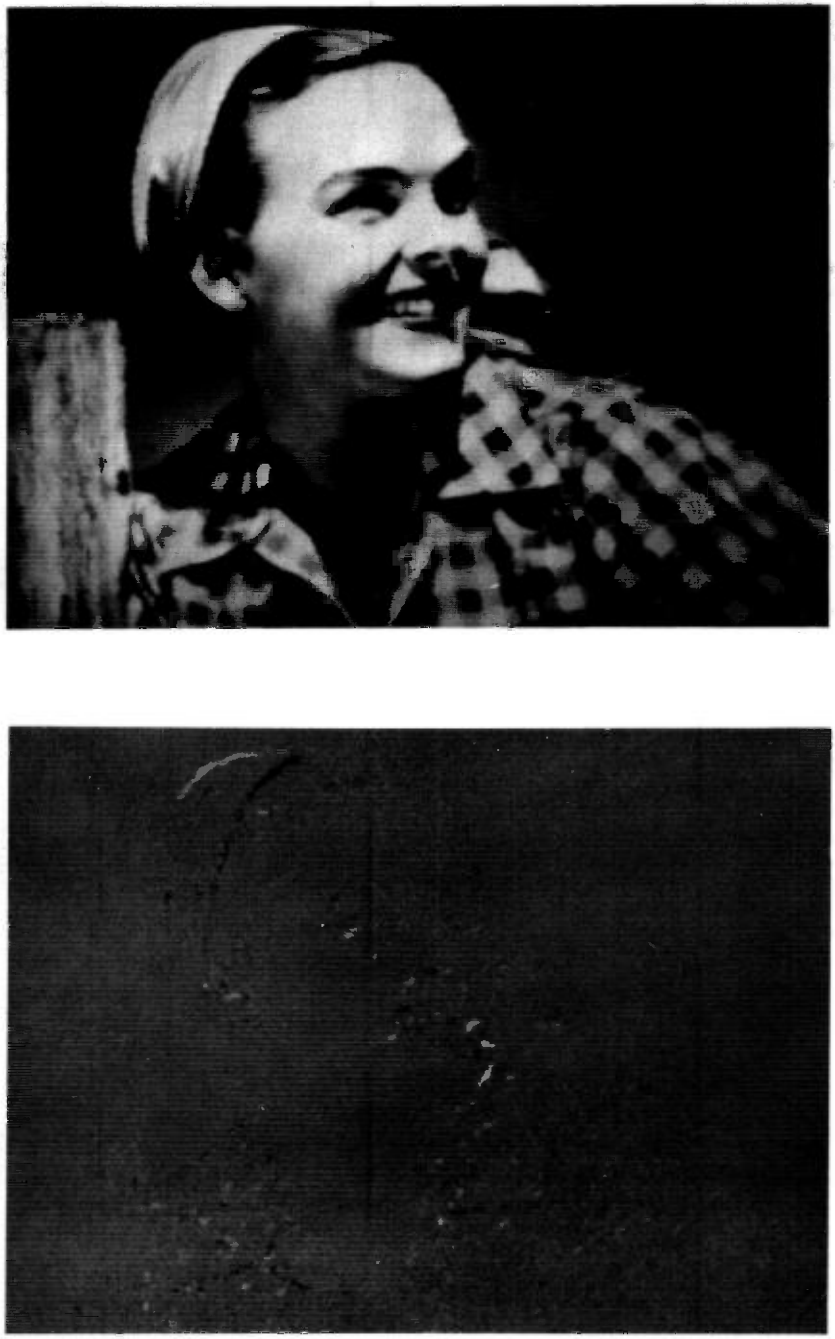


Fig. 11—Three-bit/pel codec performance. Top: Decoder output. Bottom: Difference between decoder output and original.

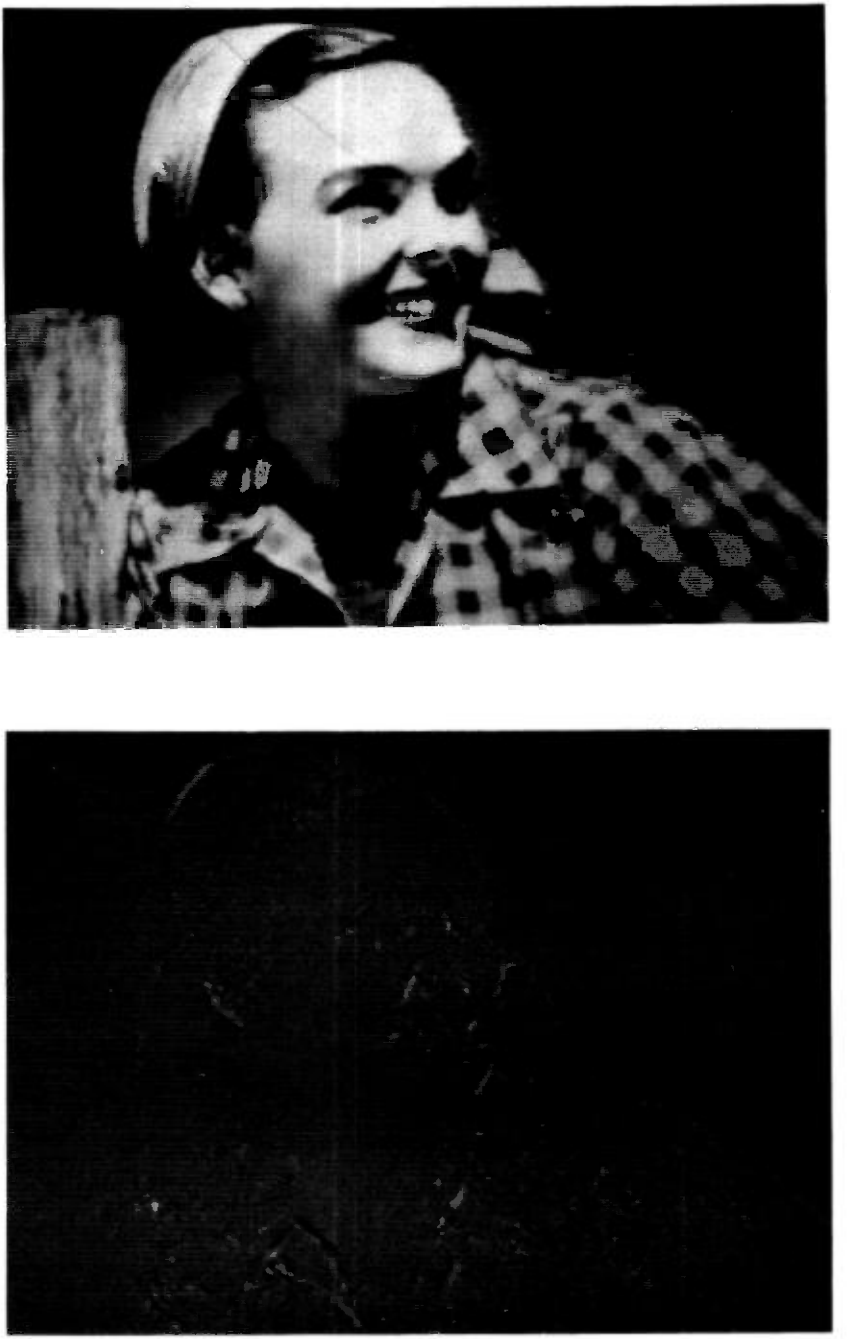


Fig. 12—Performance of three-bit Graham codec. Top: Decoder output. Bottom: Difference between decoder output and original.

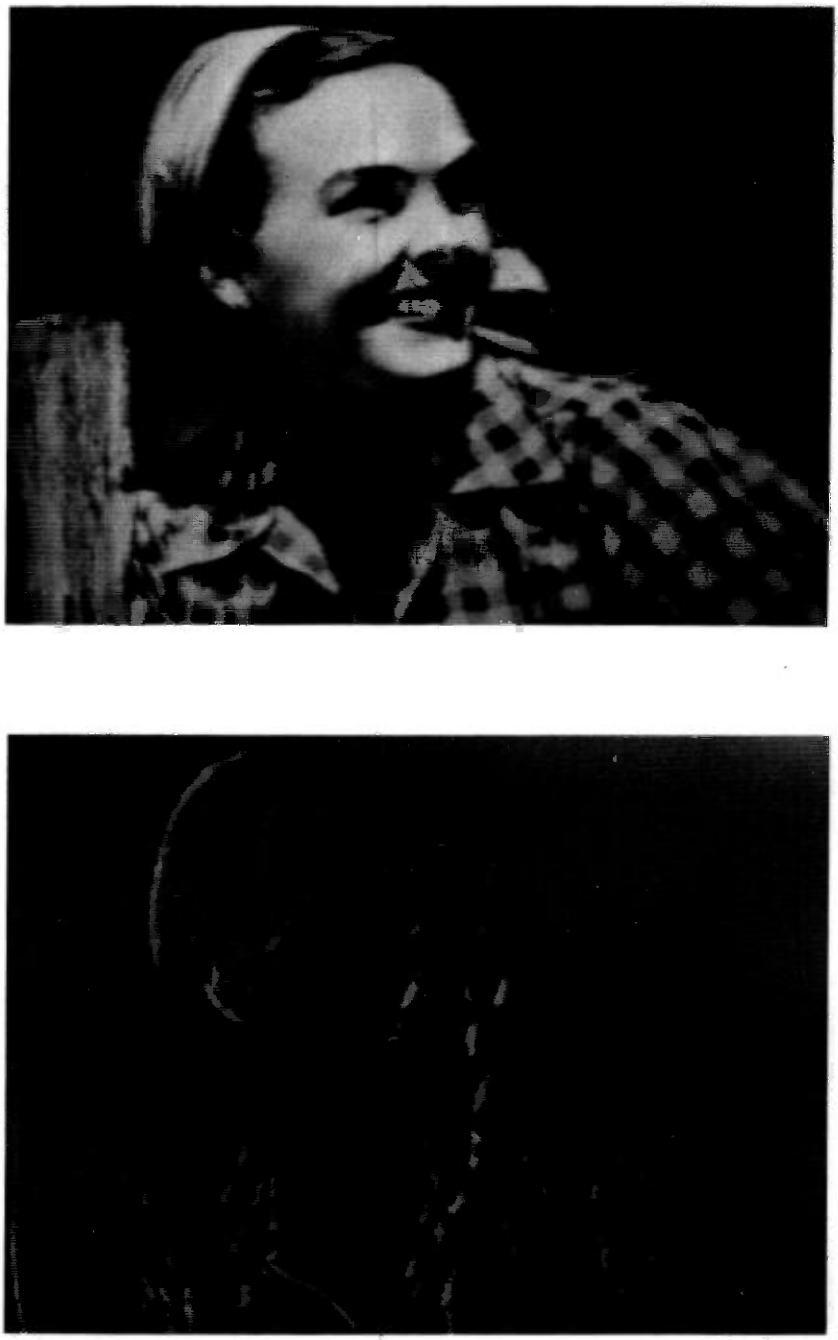


Fig. 13—Performance of three-bit previous element DPCM. Top: Decoder output. Bottom: Difference between decoder output and original.

APPENDIX

This appendix traces the development from eq. (4) to (7) of Section III.

There are $M = (Q + 1)^N$ possible vectors $\mathbf{V} = (q_1, q_2, \cdots q_j, \cdots q_N)$ of source options in the N point region R_t of Fig. 4. Number these vectors $i = 1, 2, \cdots, M$ and let $\mathbf{V}_i$ denote the ith vector. Define $n(q \mid \mathbf{V}_i)$ as the number of components in $\mathbf{V}_i$ that equal the specific value q. Then (4) becomes

$$\hat{P}(q; t) = E\left|\frac{n(q)}{N}\right| \tag{10a}$$

$$= \frac{1}{N}\sum_{i=1}^{M} n(q \mid \mathbf{V}_i)\bar{P}(\mathbf{V}_i \mid S_t^-). \tag{10b}$$

In taking the expectation in (10a) we have treated $t \equiv (m, n)$ as a randomly chosen raster point for which $\bar{P}(q) \equiv E\{P(q; t)\}$ of eq. (1) applies. Because the q are selected independently, $\hat{P}(\mathbf{V}_i \mid S_t^-)$ of (10b) is related to $\bar{P}(q)$ by

$$\hat{P}(\mathbf{V}_i \mid S_t^-) = \frac{P(S_t^- \mid \mathbf{V}_i)}{P(S_t^-)}\bar{P}(\mathbf{V}_i) \tag{11a}$$

$$= \frac{P(S_t^- \mid \mathbf{V}_i)}{P(S_t^-)}\prod_{j=1}^{N}\bar{P}(q_{ij}), \tag{11b}$$

where $p(S_t^- \mid (\cdot))$ denotes the probability density function of the vector of values in S_t^-, and q_{ij} is the jth component of $\mathbf{V}_i$. [The random selection of t cannot affect the independence of the components of $\mathbf{V}_i$ in (11a); hence, (11b)].

Substituting

$$n(q \mid \mathbf{V}_i) = \sum_{j=1}^{N}\delta_{q-q_{ij}} \tag{12}$$

into (10b) and summing over i, (10b) becomes

$$\hat{P}(q; t) = \frac{1}{N}\sum_{j=1}^{N}\bar{P}(q_j = q \mid S_t^-), \tag{13}$$

where $\bar{P}(q_j = q \mid S_t^-)$ is the conditional probability that the jth pixel in R_t was generated by source q. Note that (13) gives $\hat{P}(q; t)$ as an average of *a posteriori* probabilities of q over the region R_t, where

$$\bar{P}(q_j = q \mid S_t^-) = \frac{P(S_t^- \mid q_j = q)}{P(S_t^-)}\bar{P}(q). \tag{14}$$

We now partition the set of pixels S_t^- into (*i*) a subset of pixels future to s_j but past to S_t (call it S_j^+); (*ii*) the pixel s_j; and (*iii*) a subset of

pixels S_t^j past to pixel s_j. The elements in S_j^+ when conditioned on s_j and S_t^j do not depend upon q_j, and it follows after straightforward manipulations that

$$\bar{P}(q_j = q \mid S_t^-) = K\, g(s_t - F_q(S_t^j))\bar{P}(q); \quad 1 \leq q = \leq Q,$$

and

$$\bar{P}(q_j = 0 \mid S_t^-) = \frac{K}{255}\bar{P}(0), \tag{15}$$

where K is a normalizing constant. Substitution of eqs. (15) and (1) into (10) yields (5).

REFERENCES

1. Z. L. Budrikis, "Visual Fidelity Criterion and Modeling," Proc. IEEE, *60* (July 1972), pp. 771 to 779.
2. R. E. Graham, "Predictive Quantizing of Television Signals," IRE Wescon Convention Record, Part 4, 1958, pp. 142 to 157.
3. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, *63* (April 1975), pp. 561 to 580.
4. J. Max, "Quantizing for Minimum Distortion," IEEE, Trans. Inform. Theory, *IT-21* (July, 1975), pp. 373 to 378.
5. A. N. Netravali and B. Prasada, "Adaptive Quantization of Picture Signals Using Spatial Masking," Proc. IEEE, *65* (April 1977), pp. 536 to 548.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Reduction of Transmission Error Propagation in Adaptively Predicted, DPCM Encoded Pictures

By N. F. MAXEMCHUK and J. A. STULLER

(Manuscript received November 21, 1978)

A new technique for reducing transmission error propagation in adaptively predicted, DPCM*-encoded pictures is described. The basis for the technique is a generalization of the notion of predictor output attenuation, described by Graham, to include attenuation of the adaptive prediction function. Simulation results are presented that show that application of the technique to Graham's codec results in significant reduction in error propagation without degradation of picture quality. The technique requires no increase in transmission rate.*

I. INTRODUCTION

This paper presents a new and simple technique to reduce error propagation in DPCM image coders that employ adaptive switching-type prediction. An analytical performance description of this technique has not been obtained. However, simulation results using the Graham[1] adaptation algorithm are presented that demonstrate that—in this case, at least—the technique can provide substantial reduction in channel error propagation without decreasing the transmission rate.

The class of coders considered is those which adaptively choose one of Q fixed predictors F_q, $q = 1, 2, \cdots, Q$, according to a decision rule that operates on the previously reconstructed pixels in the local past vicinity of the element to be predicted. If x_{ij} is the ith pixel on the jth line of the input raster, and y_{ij} is the vector of reconstructed pixels in the local past vicinity of (i, j), then the adaptive predictor has the form

$$\hat{x}_{ij} = F(y_{ij}), \tag{1}$$

where $F(\cdot)$ is one of Q fixed functions $F_q(\cdot)$, $q = 1, \cdots, Q$, with q chosen according to a decision rule operating on y_{ij},

$$q = D(y_{ij}). \tag{2}$$

The encoder and decoder use the same decision rule to determine q, but the decoder must base its decision upon its possibly contaminated version of the reconstructed past scene.

An example of (1) and (2) is given by Graham's predictive encoder. Here $Q = 2$ with

$$\begin{aligned} F_1(y_{ij}) &= y(i-1, j) \\ F_2(y_{ij}) &= y(i, j-1) \end{aligned} \tag{3}$$

and

$$D_{ij} = \begin{cases} 1; & \text{if } |y(i-1, j-1) - y(i, j-1)| \\ & \quad < |y(i-1, j-1) - y(i-1, j)| \\ 2; & \text{otherwise.} \end{cases} \tag{4}$$

It is well known that, for a fixed transmission rate, adaptive prediction generally results in a more accurate coded version of the image, particularly on edges within the picture where large changes in amplitude occur along one dimension. However, a serious problem generally arising from such adaptation is the response of the system to channel errors. Generally, the effect of an error propagates over a larger area of the picture when an adaptive predictor is used than when a fixed predictor is used. This occurs because transmission errors not only (*i*) contaminate the value of the elements used by the receiver in the function $F(\cdot)$ when the receiver's choice of q is correct, but can also (*ii*) cause an error in the receiver's choice of q. Note that effect (*i*) is present in nonadaptive coders and is defined as occurring in adaptive coders when the correct choice of q is made by the decoder. Effect (*ii*) is unique to adaptive prediction and is potentially more grievous since the transmitter and receiver then use different choices for the prediction function $F(,)$—a result that, once started, can propagate. An example of the effect of transmission errors in adaptive DPCM is shown in Fig. 1. In this example, Graham's three-bit codec is used over a binary symmetric channel having bit error probability of 10^{-4}. Figure 2 shows the difference between the output of this system with and without transmission errors.

II. PREDICTOR OUTPUT LEAK

As observed by Graham and others, the effect of transmission errors can be reduced by attenuating the predictor output by a constant α, $0 \leq \alpha \leq 1$. In general, a bias term can be introduced so that $\hat{x}_{i,j}$ assumes the form:

$$\hat{x}_{ij} = \alpha F(y_{ij}) + (1 - \alpha)\eta, \tag{5}$$

Fig. 1—Effect of 10^{-4} channel bit error probability on output picture: Graham 3 bit/pel codec.

where η is a constant in the span of possible picture values. A possible choice of η is the mean of x_{ij},

$$\eta = E\{x_{ij}\}. \tag{6}$$

Other choices, however, can give subjectively better results depending upon context and system nonlinearities.

Equation (5) can be viewed as a weighted combination of *locally inferred* and *globally given* knowledge about x_{ij}. A value of $\alpha < 1$ has the effect of decreasing the memory of the closed reconstruction loop, and the bias term causes the output to tend toward η. The quantity quantized and transmitted has the form

$$x_{ij} - \hat{x}_{ij} = \alpha[x_{ij} - F(y_{ij})] + (1 - \alpha)[x_{ij} - \eta], \tag{7}$$

which is seen to consist of both DPCM and PCM information. As α varies from one to zero, the system changes from DPCM to PCM. Therefore, attenuation of predictor output as in (5) trades error-propagation attenuation with transmission rate. It should be observed that the technique described by (5) will also reduce error propagation in non-adaptive codecs. Also, the technique does not directly address the problem of the *choice* of q and is therefore a remedy more closely connected to type (*i*) errors than to type (*ii*).

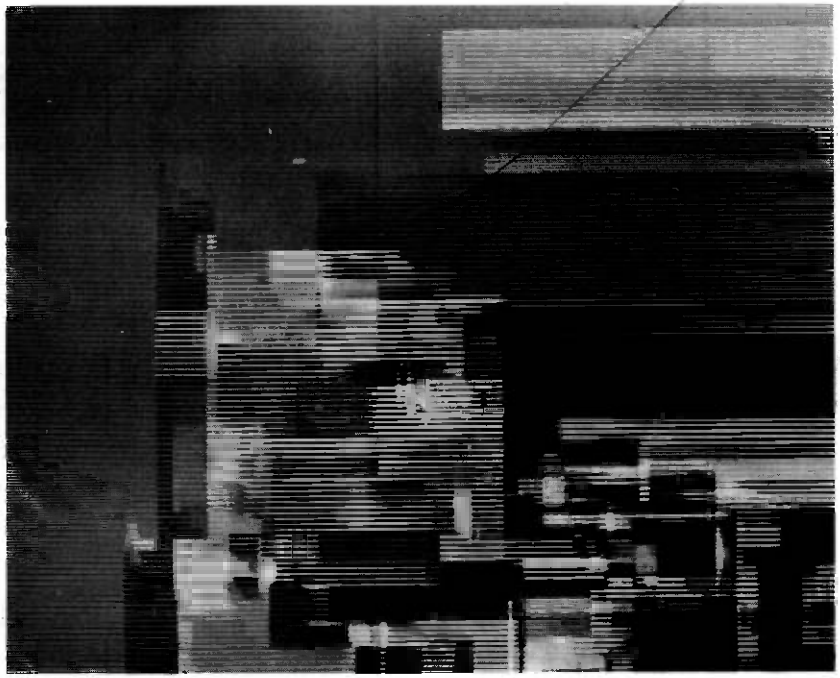


Fig. 2—Difference between output of Graham codec with and without transmission errors.

Using the Graham system, it was found that α could be reduced to $^{15}/_{16}$ before degradation in the output caused by quantization noise becomes visible. The reduction in error propagation resulting from (5) (with η set to 128) is shown in Fig. 3. The difference of the coded picture with and without channel errors is shown in Fig. 4. Although an improvement is obtained with this approach, the next section shows that substantially greater improvement is possible.

III. PREDICTION FUNCTION LEAK

Since the second effect of channel errors in an adaptive codec is to make the value of q uncertain, the receiver loop must in fact estimate the *function* $F(\cdot)$ as well as x_{ij}. In analogy with (5) we introduce a constant β, $0 \leq \beta \leq 1$, and set (at both transmitter and receiver)

$$\hat{F}(\cdot) = \beta F(\cdot) + (1 - \beta)\bar{F}(\cdot), \tag{8}$$

where $\bar{F}(\cdot)$ is a fixed predictor. A reasonable choice for $\bar{F}(\cdot)$ is the mean of F_q,

$$\bar{F}(\cdot) = \sum_{q=1}^{Q} F_q(\cdot)P(q), \tag{9}$$

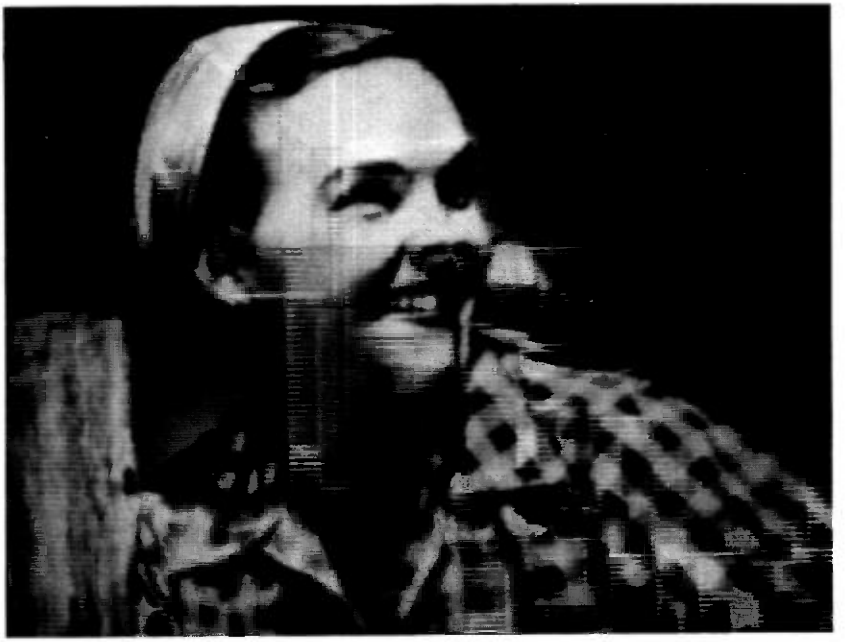


Fig. 3—Reduction of transmission error propagation using predictor output leak ($\alpha = {}^{15}/_{16}$, $\eta = 128$, 10^{-4} bit error probability).

where $P(q)$ is the *a priori* probability of q. Other choices for $\bar{F}(\cdot)$ are possible.

Note that, as β varies from one to zero, a system using $\hat{F}(\cdot)$ will change from fully adaptive DPCM to nonadaptive DPCM. The smaller the value of β, the closer the predictor to being fixed, and the smaller the effect of error propagation due to using the wrong predictor. Note also that (9) is an approach that is applicable only to adaptive codecs. Because of this, we view this technique as a remedy for the second error class (*ii*) described in Section I.

The concept of prediction function leak has been applied to the Graham predictor and has successfully reduced error propagation. The predictor used in this experiment is:

$$\hat{x}_{ij} = \hat{F}(y_{ij}), \tag{10}$$

where $\hat{F}(\cdot)$ is given by (8) using (3) to (4) and

$$\bar{F}(y_{ij}) = \tfrac{1}{2}(y(i-1,j) + y(i,j-1)). \tag{11}$$

It was found experimentally that β could be reduced to be between $\tfrac{3}{4}$ and $\tfrac{1}{2}$ (depending upon the picture) before the ability of the adaptive predictor to respond to edges within the picture was compromised.

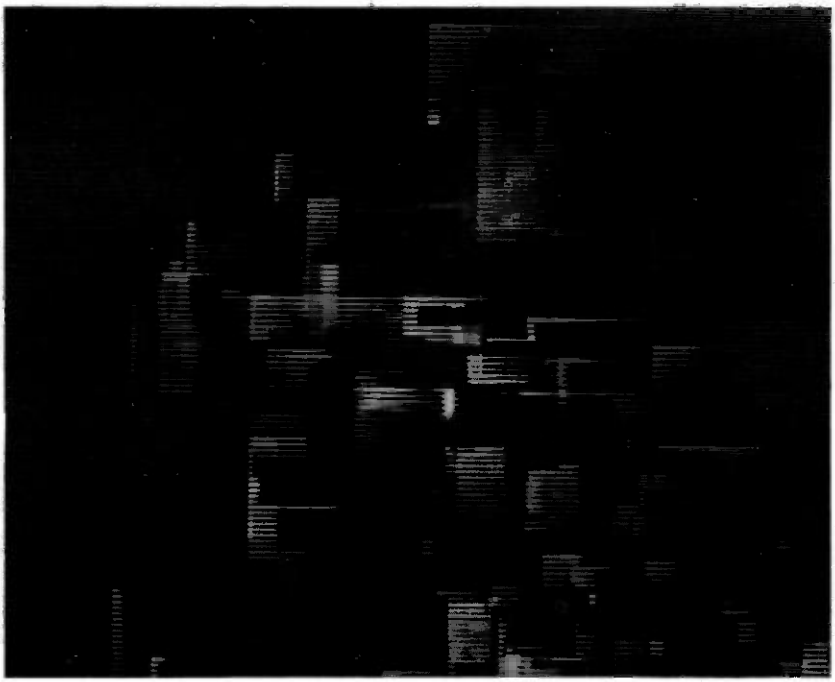


Fig. 4—Difference between output of codec with and without transmission errors ($\alpha = 15/16$, $\eta = 128$).

Fig. 5—Reduction of transmission error propagation using predictor output and prediction function leak ($\alpha = 15/16$, $\beta = 3/4$, $\eta = 128$, 10^{-4} bit error probability).

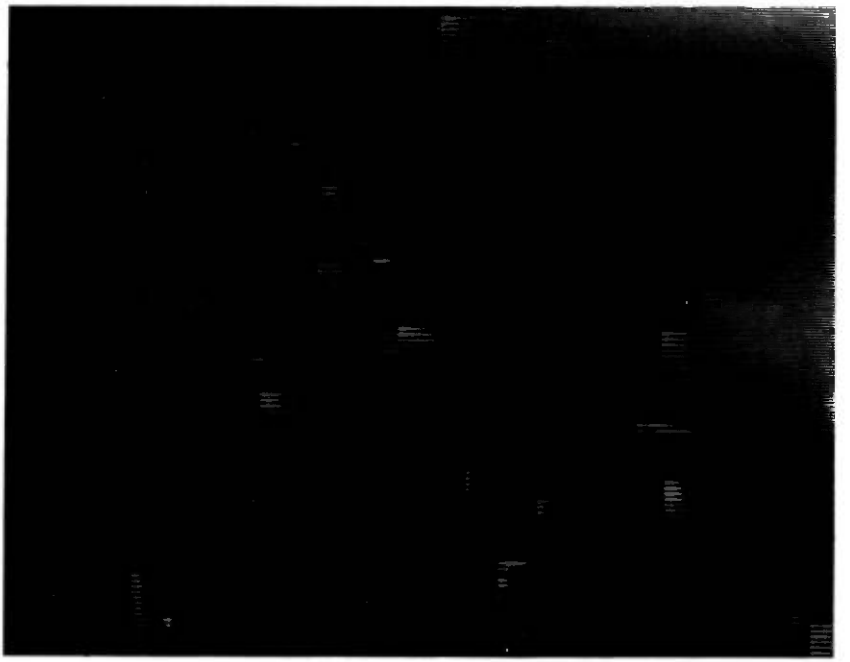


Fig. 6—Difference between output of codec with and without transmission errors ($\alpha = {}^{15}/_{16}$, $\beta = {}^{3}/_{4}$, $\eta = 128$).

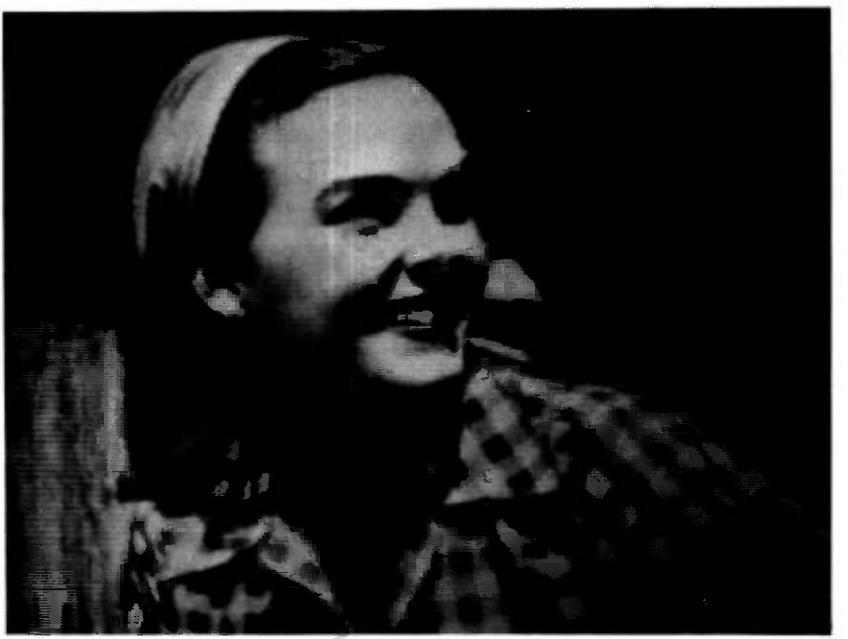


Fig. 7—Reduction of transmission error propagation using predictor output and prediction function leak ($\alpha = {}^{15}/_{16}$, $\beta = {}^{1}/_{2}$, $\eta = 128$, 10^{-4} bit error probability).

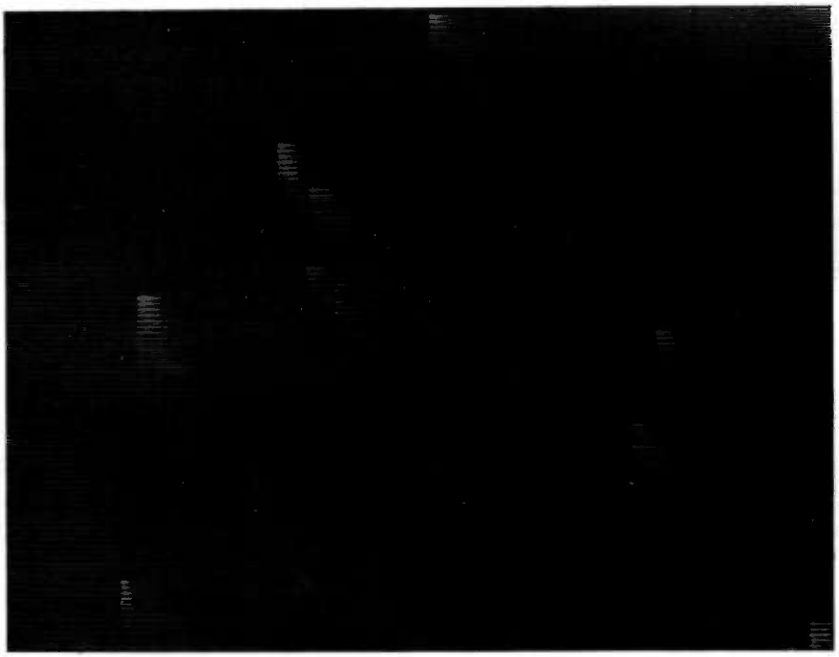


Fig. 8—Difference between output of codec with and without transmission errors ($\alpha = {}^{15}/_{16}$, $\beta = \frac{1}{2}$, $\eta = 128$).

Fig. 9—Quantizing noise of Graham codec.

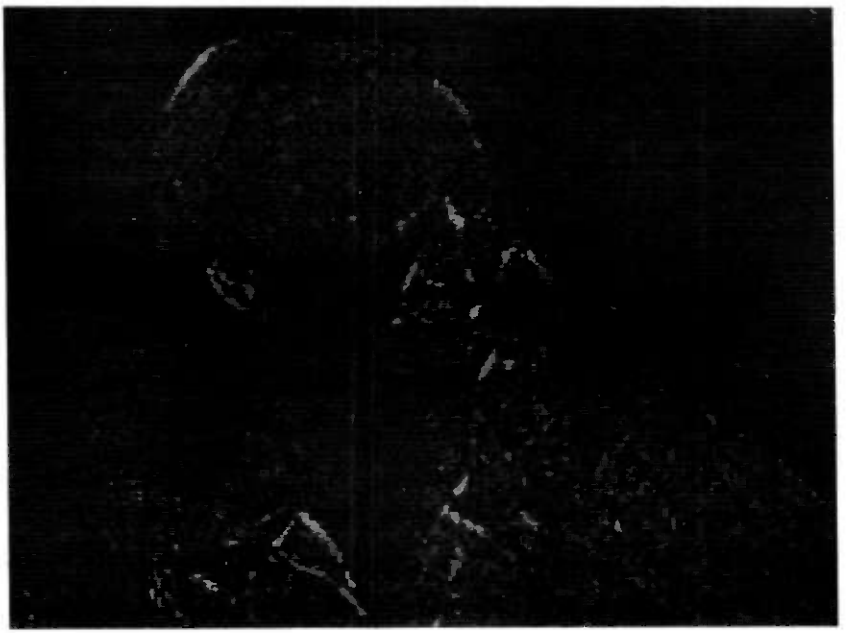


Fig. 10—Quantizing noise for $\alpha = {}^{15}\!/_{16}$, $\beta = 1$.

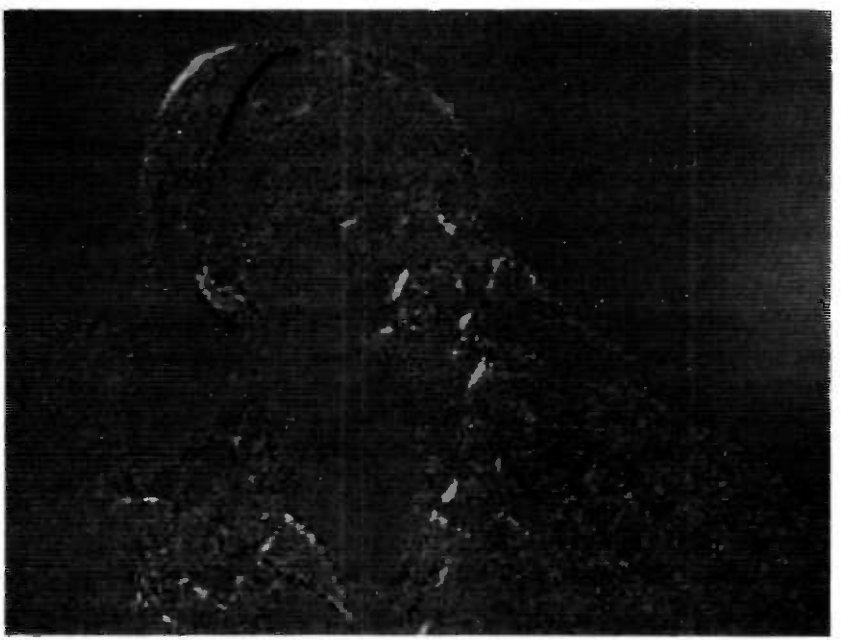


Fig. 11—Quantizing noise for $\alpha = {}^{15}\!/_{16}$, $\beta = {}^{3}\!/_{4}$.

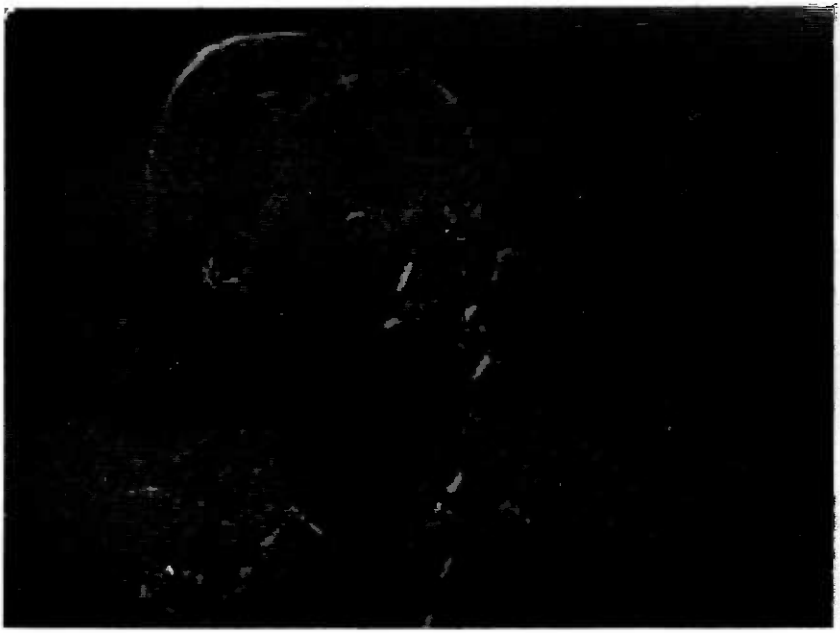


Fig. 12—Quantizing noise for $\alpha = 15/16$, $\beta = 1/2$.

When prediction function leak is used, it is possible to use predictor output leak to further reduce the effects of errors. The estimator is formed as:

$$\hat{x}_{ij} = \alpha\hat{F}(y_{ij}) + (1 - \alpha)\eta$$
$$= \alpha\beta F(y_{ij}) + \alpha(1 - \beta)\bar{F}(y_{ij}) + (1 - \alpha)\eta. \quad (12)$$

Pictures with the same transmission error patterns as those that occurred in the pictures produced using the original Graham predictor were processed using (12). In Fig. 5, $\alpha = 15/16$, $\beta = 3/4$ and $\eta = 128$. Figure 6 is the difference between pictures processed with this predictor with and without transmission errors. In Fig. 7, $\alpha = 15/16$, $\beta = 1/2$, and $\eta = 128$. Figure 8 is the error difference picture of this coder. As seen from these pictures, leaking the prediction function significantly reduces the effect of transmission errors. Analysis of this effect has been hindered by the nonlinear nature of the equations and the fact that the quantities involved exist on a two-dimensional field.

Figures 9 through 12 show the encoding (quantizing) noise for each of the systems previously described in this paper. By comparing Fig. 9 with 11 and 12, it can be seen that, for β equaling 3/4 and 1/2, prediction function leak *reduces* encoding noise along edges within the picture. In fact, pictures transmitted with these values of β over an error-free

channel are preferable to those transmitted with the original Graham codec since edge serration, which sometimes occurs when a switching-type predictor is used, is reduced. Figure 10 shows that the introduction of predictor output leak alone does not produce the same beneficial effect.

REFERENCE

1. R. E. Graham, "Predictive Quantizing of Television Signals," IRE Wescon Convention Record, Part 4, 1958, pp. 142 to 157.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Application of Optimization Theory to the Control of the Optical Fiber Drawing Process

By D. H. SMITHGALL

(Manuscript received January 15, 1979)

The optical fiber drawing process is examined and a feedback control loop identified. The incremental dynamic response of each loop component is determined, and the sensitivity of loop response to system parameters is examined. The control loop is optimized, based upon a mean square error criterion with constraints imposed for periodic disturbances. An expression is derived for the effectiveness of the control loop with respect to sources of system disturbance and found to correlate well with experimental results.

I. INTRODUCTION

With the advent of fiber optics technology has come the potential for use of this technology in high quality telecommunications systems. Such systems require sources, detectors, and fibers superior to those used in the present applications. The fibers in high quality systems must have low loss and dispersion and yet be economically produced. One factor influencing transmission loss in the fiber, particularly at splice locations,[1] is the diameter uniformity. Diameter uniformity is directly related to the manufacturing process and is influenced by the environment in which it is drawn[2] as well as the material from which it is drawn. In additon, large variations in diameter occur during the startup operation, resulting in a material loss of up to 10 percent of the potential fiber.

Much of this wastage can be eliminated and a high degree of fiber uniformity maintained by the judicious design and application of a feedback control on the fiber drawing process. Optimization of such a control requires identification of the distributed, nonlinear drawing process and a quantification of the sensitivities of the process to changes in process parameters.

1.1 The fiber drawing process

Optical fiber is formed by locally and symmetrically heating a cylindrical preform, typically 7 to 25 mm in diameter and 30 and 60 cm in length, to a temperature in the neighborhood of 2000°C. As the preform of diameter D_p is fed into the heat zone at a velocity V_p, the fiber is drawn from the molten material at a velocity V_f, as shown in Fig. 1. Due to the temperatures involved and the tolerances required, the fiber cannot be drawn through a die, and consequently the surface of the molten material is a free boundary whose shape is determined by an equilibrium between the velocity shear gradients and the restraining surface tension. The diameter of the fiber, D_f, is determined, then, by the principle of conservation of mass, which may be written as

$$D_f^2 V_f = D_p^2 V_p + \int_{-\infty}^{t} dw, \tag{1}$$

where w is a random process representing mechanical and thermally-induced disturbances as well as the variations in diameter which occur while the process is establishing its equilibrium condition. The nature of the disturbing influences is illustrated in Fig. 2, where the fiber diameter is plotted for a 500-m length of fiber with constant D_p, V_p, and V_f. An additional source of diameter variation results from changes in preform diameter which are of a slowly varying nature. Once the process "equilibrium" has been established, the noise (diameter vari-

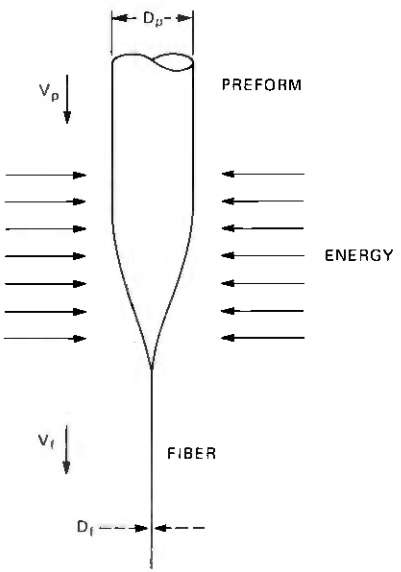


Fig. 1—Geometry of the optical fiber drawing process.

ation) which appears on the fiber can be characterized as band-limited noise riding on a slowly varying bias.

1.2 Identification of process dynamics

Since the fiber diameter is related directly to the two manipulable variables V_f and V_p by the mass conservation principle, and only indirectly to other manipulable variables such as heat source temperature or heat flux, these are the system variables through which a control signal can most effectively be coupled into the process. Experimental results show that the dynamic response of the fiber is two orders of magnitude faster with respect to the drawing velocity than to the feed velocity. Consequently, the control loop shown in Fig. 3 has been determined to be the most effective means of controlling fiber diameter.

The motor-drawing mechanism is typically a pinch-wheel device in which one wheel is driven by a DC motor. The dynamic response of the mechanism used in this investigation is modeled as

$$\frac{V_f(s)}{u(s)} = \frac{0.31}{0.0045s^2 + 0.030s + 1} \frac{\text{meters/second}}{\text{volt}}, \tag{2}$$

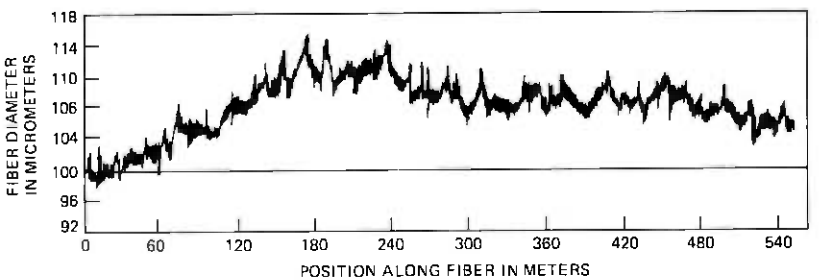


Fig. 2—Diameter profile of uncontrolled fiber with constant feed and drawing speeds.

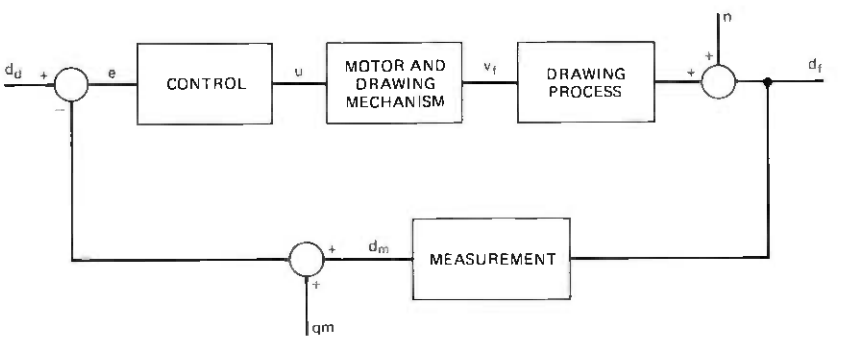


Fig. 3—Block diagram of diameter feedback control loop.

where u is the control signal to the motor. The response of the motor-drawing mechanism is dominated by the inertia of the pinch wheels, while the drawing tension on the fiber has little effect on the response of this component.

The drawing process is a nonlinear, distributed process. Since the diameter is to be regulated about a fixed set point, a linear perturbation model can be used to represent incremental system dynamics. The parameters for this model, as well as the model structure, must be determined experimentally. To obtain response characteristics, the drawing mechanism is excited with a sinusoidal perturbation and the change in draw speed and fiber diameter are measured. Using the technique of Fourier filtering,[3] the relative gain and phase of the drawing mechanism and the process diameter response at each excitation frequency can be determined, and a model can be constructed from the data. It is found that, for nominal velocities up to 1 m/s, the process can be modeled by

$$d_f(s) = \frac{D_f/2V_f}{a_p s^2 + b_p s + 1} v_f(s) + n(s), \tag{3}$$

where d_f and v_f now represent incremental changes in fiber diameter and velocity, D_f and V_f represent the nominal process values, and a_p, b_p are the parameters representing process dynamics. The source of diameter variations, n, is a band-limited, Gaussian process with $E\{n\} = 0$, $E\{n^2\} = \sigma_n^2$. The parameters a_p and b_p are sensitive to certain process parameters and insensitive to others. They have been found to be insensitive to preform diameter, fiber diameter, draw velocity, and the temperature of the heat source over the ranges

$$\begin{aligned} 1950°\text{C} &< T_s < 2150°\text{C} \\ 7\text{ mm} &< D_p < 19\text{ mm} \\ 0\text{ m/s} &< V_f < 1\text{ m/s} \\ 80\ \mu\text{m} &< D_f < 125\ \mu\text{m}. \end{aligned}$$

The response is generally insensitive to changes in preform feed velocity over the range of speeds commensurate with the above values of D_f, V_f, D_p, but is known to change for draw velocities higher than 1 m/s. The dynamic response parameters are also sensitive to the length of the heat zone as shown in Fig. 4. The long heat zone was obtained in a furnace and the short heat zone with laser heat source.[4]

Due to physical limitations, the fiber diameter is measured at some point below the heat zone. Using a forward scattering interference fringe counting technique,[5] the fiber can be measured with an accuracy of 0.25 μm at a rate of 1000 measurements per second. Due to the high measurement rate, the measurement process has no dynamic response to contribute to the loop dynamics. As a result of the digital fringe

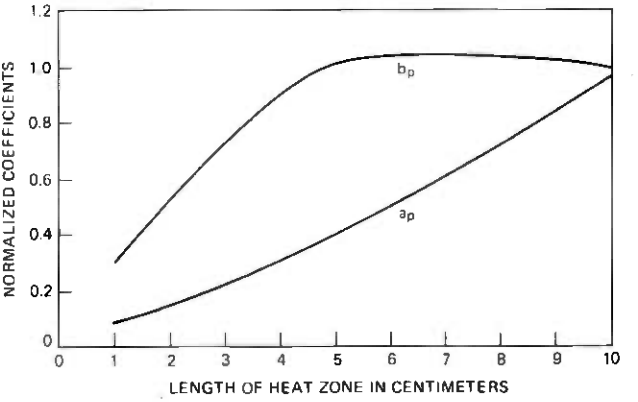


Fig. 4—Variation of process parameters with the length of heat zone.

counting technique, however, a quantization noise, q_m, perturbs the control loop. The effect of q_m can be seen from Fig. 3 to be equivalent to a loop with perfect measurement and a set point which changes $\pm$ q_m at random times. The effect of the noise on the fiber diameter depends upon the rate at which the step changes occur. Experience has shown that the quantization noise becomes significant only when the standard deviation of other system disturbances have been reduced to the level q_m and represents a lower bound of achievable performance for the drawing system.

Since the measurement process must be physically located beneath the heat source, it must also be located some distance away from the point, or region, where the diameter of the molten zone changes in response to variations in the drawing velocity. The distributed nature of the process leads to the expectation that the response to a step change in V_f would result in a distributed change in the boundary of the molten material, as well as a change in fiber diameter. Experience indicates that small perturbations in draw velocity result in diameter variations in the molten zone which are confined to a very small region, near the point where the fiber is formed. Consequently, a "point" at which the molten material changes diameter in response to changes in draw velocity can be defined, and a measurement time delay, T, results which relates the delay distance and nominal drawing velocity.

The resulting model of the measurement process is

$$\frac{d_m(s)}{d_f(s)} = 0.040\, e^{-sT} \text{ volts}/\mu\text{m}, \tag{4}$$

where the quantization noise has been reassociated with the diameter set point. The delay time is typically 0.04 to 0.10 second for a drawing velocity of 1 m/s.

II. DESIGN OF THE CONTROL

The control problem is to reduce the effects of the noise source, n, upon the fiber diameter d_f. The relationship between the two quantities when the feedback loop of Fig. 3 is used is given by

$$\frac{\delta d_f(s)}{n(s)} = \frac{1}{1 + G(s)} = H(s), \tag{5}$$

where $\delta d_f(s) = d_f(s) - d_d$, and $G(s)$ is the forward transfer function

$$G(s) = \frac{d_f(s)}{v_f(s)}\bigg|_{n=0} \times \frac{v_f(s)}{u(s)} \times \frac{u(s)}{d_m(s) - d_d} \times \frac{d_m(s) - d_d}{d_f(s)}. \tag{6}$$

The form of the control circuit is

$$\frac{u(s)}{d_m(s) - d_d} = \frac{K_c}{s} \frac{a_c s + 1}{b_c s + 1}. \tag{7}$$

The integral term is required to remove the slowly varying components of the noise source, and the lead-lag term is used to shape the response curve. Using this control circuit and the component responses (2) to (4), the response curve

$$\left|\frac{\delta d_f(s)}{n(s)}\right|^2_{s=j\omega}$$

as a function of ω is shown in Fig. 5. The response curve shows that low frequency disturbances can be effectively suppressed. There is a loss of control effectiveness as frequency increases to a point where the control loop has no effect upon the disturbances. In the region of the corner frequency ω_c, the noise is amplified. Figures 5 and 6 illustrate the role of the control circuit parameters upon the response curve. The suppression of low frequency variations is affected only by the loop gain. The degree of response peaking around the corner frequency is affected by the gain as well as the lead-lag network parameters. Some peaking resulting from high loop gain must be allowed, since it can only partially be compensated for by adjustment of a_c and b_c. It has been found that +2 dB is an acceptable peak gain for the response curve, from the standpoint of the stochastic disturbances.

The performance of the control loop can be optimized by choice of the parameters K_c, a_c, b_c to minimize the performance index

$$J = \int_0^\infty |H(j\omega)|^2 \, d\omega, \tag{8}$$

subject to the constraint

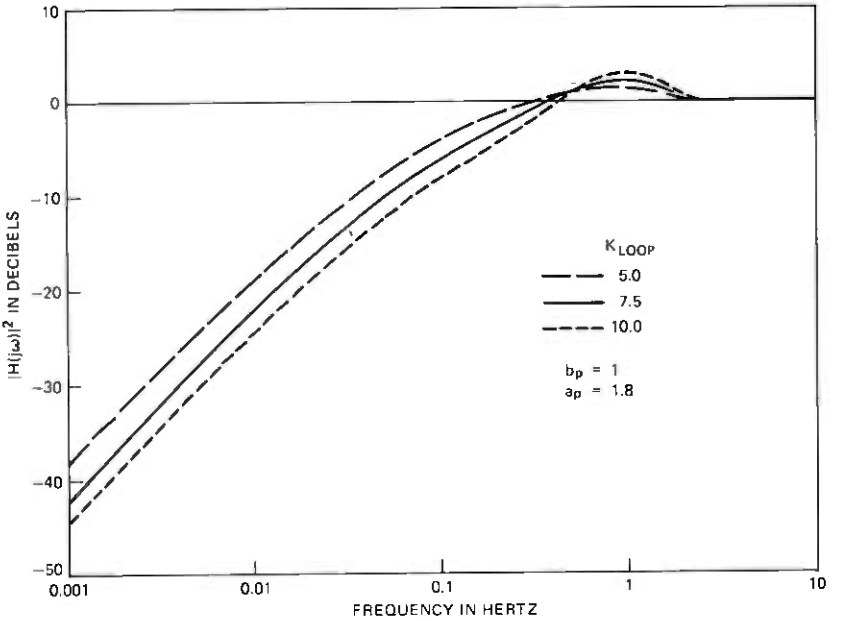


Fig. 5—Sensitivity of control loop performance to loop gain.

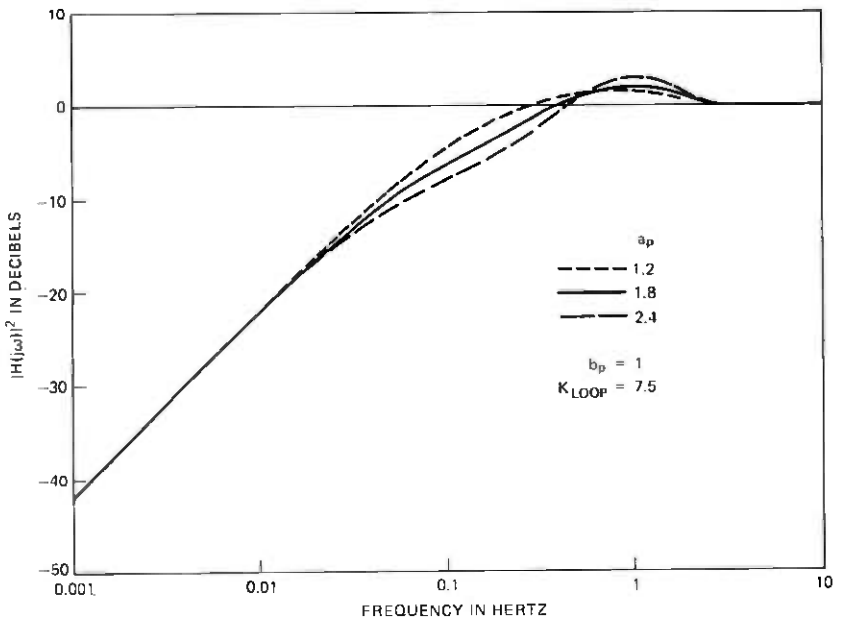


Fig. 6—Sensitivity of control loop performance to lead-lag compensation.

$$\left\{ \max_{\omega} \omega_0 \colon 0 < \omega < \infty \right\} \quad |H(j\omega)|^2 \leq 2 \text{ dB}. \tag{9}$$

The performance index (8) is a measure of the relative noise power transmitted to the fiber diameter through the control loop. Minimization of (8) is equivalent to minimizing

$$J_1 = E\left\{\delta d_f^2(t)\right\}, \quad |n(\cdot)| = 1, \tag{10}$$

for $n(\cdot)$ modeled by white noise. This relationship is established by the relationship between the power spectral densities

$$\Gamma_{d_f} = |H(j\omega)|^2 \Gamma_n, \tag{11}$$

resulting in

$$J_1 = \int_{-\infty}^{\infty} \Gamma_{d_f} \, d\omega = 2\sigma_n^2 J. \tag{12}$$

Due to the constraint (9), which seeks to reduce the effect of periodic disturbances upon the loop design, optimization is most effectively performed in the frequency domain.

The optimum performance for a typical set of drawing system parameters is shown in Fig. 7, for the case of two measurement delay times. Referring to Figs. 5 to 7, of all the system parameters, the effectiveness of the control loop is most sensitive to loop gain and the measurement delay time. Examination of the phase characteristics of the system components reveals that over the range of controllable disturbances the measurement delay contributes the largest phase lag to the loop dynamics. Consequently, it is imperative in drawing system design to minimize the measurement delay. The loop response is sensitive to loop gain because the destablizing tendencies of the measurement delay pull the root loci toward the $j\omega$ axis.[6]

To determine the effectiveness of the control, account must be taken of both the process-related noise, n, and the measurement quantization noise q. The fiber diameter variation is related to these two quantities in the frequency domain by the expression

$$\delta d_f(s) = H(s)n(s) + G(s)H(s)q(s). \tag{13}$$

The power spectral densities are related by

$$\Gamma_{d_f}(\omega) = |H(j\omega)|^2 \Gamma_n(\omega) + |F(j\omega)|^2 \Gamma_q(\omega), \tag{14}$$

where $F(s) = G(s)H(s)$. The noise generating process associated with the fiber drawing process has been experimentally determined and can be modeled as

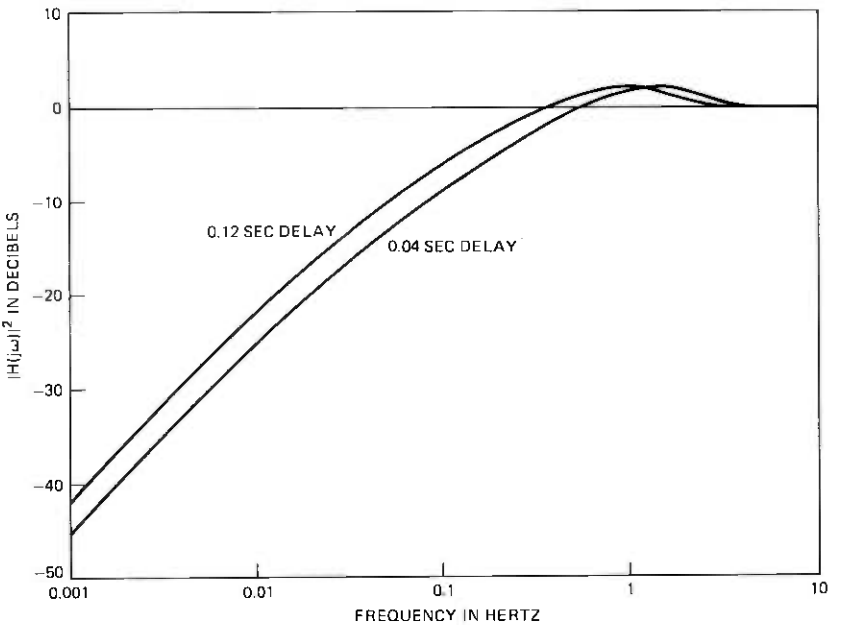


Fig. 7—Sensitivity of control loop performance to measurement time delay. System parameter values are $a_p = 0.0032$, $b_p = 0.080$.

$$\Gamma_n(\omega) = \frac{\frac{\omega_n}{\pi}\sigma_n^2}{\omega^2 + \omega_n^2}, \tag{15}$$

where ω_n is the corner frequency of the noise spectrum.

The quantization noise can be modeled as a random telegraph signal[7] for which the power spectral density function is

$$\Gamma_q(\omega) = \frac{\lambda q_m^2}{\lambda^2 + \pi^2\omega^2}, \tag{16}$$

where λ is the mean number of switchings per unit time. The control function can be approximated as

$$|H(j\omega)|^2 = \frac{\omega^2}{\omega^2 + \omega_c^2}$$

$$|F(j\omega)|^2 = \frac{\omega_c^2}{\omega^2 + \omega_c^2}, \tag{17}$$

where ω_c is the corner frequency of the closed loop transfer function. The variance of the fiber diameter is then determined by computing the auto-covariance function[8] with zero lag:

$$\sigma_d^2 = \frac{1}{1 + \dfrac{\omega_c}{\omega_n}} \sigma_n^2 + \frac{1}{1 + \dfrac{\lambda}{\pi\omega_c}} q_m^2 . \tag{18}$$

Equation (18) illustrates an interesting tradeoff. Whereas it is desirable to reduce the process-induced fiber diameter variation by increasing ω_c insofar as possible, it is done at the expense of allowing additional quantization noise to affect the fiber diameter. The latter source may only be reduced by increasing the resolution of the measurement system.

Experimentally, the effectiveness of the control loop is illustrated in Fig. 8 by the distribution of mean and standard deviations measured on 500-m lengths of controlled and uncontrolled fibers, after the process has reached equilibrium. The mean diameter of the uncontrolled fibers varies over a 4-μm range due to variations in preform diameter, nominal drawing velocities, etc. Using the feedback control, the mean diameter is held to within 0.1 μm of the set point. This deviation from set point is due, in part, to the 0.25 μm measurement resolution. The standard deviations of uncontrolled fibers are distributed over a wide range with the majority of the samples having a

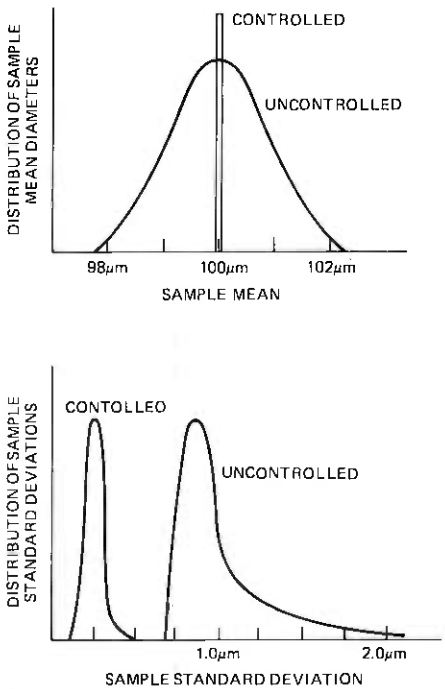


Fig. 8—Measured statistical characteristics of controlled and uncontrolled fibers.

standard deviation of 0.8 μm. Use of the feedback control improves fiber quality with the result that standard deviations below 0.30 μm can be repeatably attained.

Using experimentally determined system parameters, the controlled variance of the fiber diameter given in eq. (18) can be compared to experimental values. If $\omega_n = 0.02$, $\sigma_n = 0.8$ μm, $q_n = 0.25$ μm, $\lambda = 5$, and from Fig. 7 $\omega_c = 0.2$, eq. (18) yields a theoretical standard deviation of 0.27 μm. This value is very close to the median of the experimentally measured sample standard deviations.

III. CONCLUSION

Optimization theory has been applied to the optical fiber drawing process, resulting in a diameter feedback control loop which effectively reduces fiber diameter variations. In addition to examining the sensitivities of control loop performance to system component parameters, the sensitivities with respect to the noise sources were also examined. The expression derived to describe system performance with respect to process-induced diameter variation and measurement quantization noise showed good agreement with experimental results.

REFERENCES

1. C. M. Miller, "Transmission vs. Transverse Offset for Parabolic—Profile Fiber Splices with Unequal Core Diameters," B.S.T.J., *55*, No. 7 (September 1976), pp. 917–927.
2. M. Nakahara et al., "Drawing Techniques for Optical Fibers," Review of the Electrical Communication Laboratories, *26*, No. 3–4 (March–April 1978), pp. 476–483.
3. P. Eykhoff, *System Identification*, New York: John Wiley & Sons, 1974, pp. 378–382.
4. R. E. Jaeger, "Laser Drawing of Glass Fiber Optical Waveguides," Ceramic Bulletin, *55*, No. 3 (1976), pp. 270–273.
5. D. H. Smithgall, L. S. Watkins, and R. E. Frazee, "High Speed Noncontact Fiber Diameter Measurement Using Forward Light Scattering," Appl. Opt., *16*, No. 9 (September 1977), pp. 2395–2402.
6. K. Ogata, *Modern Control Engineering*, Englewood Cliffs: Prentice-Hall, 1970, pp. 346–350.
7. W. B. Davenport and W. L. Root, *Random Signals and Noise*, New York: McGraw-Hill, 1958, pp. 61, 104.
8. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, New York: Holden-Day, 1968, Ch. 6.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

On the Structure of Real-Time Source Coders

By H. S. WITSENHAUSEN

(Manuscript received January 30, 1979)

*The outputs of a discrete time source with memory are to be encoded ("quantized" or "compressed") into a sequence of discrete variables. From this latter sequence, a receiver must attempt to approximate some features of the source sequence. Operation is in real time, and the distortion measure does not tolerate delays. Such a situation has been investigated over infinite time spans by B. McMillan. In the present work, only finite time spans are considered. The main result is the following. If the source is k*th*-order Markov, one may, without loss, assume that the encoder forms each output using only the last k source symbols and the present state of the receiver's memory. An example is constructed, which shows that the Markov property is essential. The case of delay is also considered.*

I. INTRODUCTION

The outputs of a discrete time source with memory are to be encoded ("quantized" or "compressed") into a sequence of discrete variables. From this latter sequence, a receiver must attempt to approximate some features of the source sequence. Operation is in real time, and the distortion measure does not tolerate delays. Such a situation has been investigated over infinite time spans in Ref. 1. In the present work, only finite time spans are considered.

The main result is the following. If the source is kth-order Markov, one may, without loss, assume that the encoder forms each output using only the last k source symbols and the present state of the receiver's memory.

An example is constructed, which shows that the Markov property is essential.

II. THE MODEL

2.1 The causal structure

A source produces a random sequence $X_1, X_2, \cdots, X_T$ where for each $t \in \{1, \cdots, T\}$, X_t is a vector in n_t-dimensional real space. The source

is characterized by the sequence distribution: A given probability measure on the Borel sets of the product space of dimension $\sum_{t=1}^{T} n_t$.

For each t, there is an opportunity for noiseless transmission of a signal Y_t taking q_t possible values. This signal is produced from the X sequence by an encoder. As we consider the problem in real time, causality allows the encoder at t to see only the values $X_1, \cdots, X_t$. The encoders are thus characterized by functions f_t: $R^{n_1+\cdots+n_t} \to \{1, \cdots, q_t\}$, Borel measurable, $t = 1, \cdots, T$.

At the receiving end, the most that could be accessible at stage t is the subsequence $Y_1, \cdots, Y_t$. However, we also want to consider the case of limited memory, as the receiver might not be able to store this whole sequence for large t. The model will be the following.

At $t = 1$, only Y_1 is available, and a discrete variable $Z_1 = r_1(Y_1)$ taking m_1 values is stored in memory. For each $t > 1$, the memory is updated by

$$Z_t = r_t(Z_{t-1}, Y_t), \qquad t = 2, \cdots, T-1,$$

where Z_t takes values in $\{1, \cdots, m_t\}$ and

$$r_t\colon \{1, \cdots, m_{t-1}\} \times \{1, \cdots, q_t\} \to \{1, \cdots, m_t\}$$

is the memory update function.

The purpose of the receiver is to generate a variable V_t in R^{s_t} by

$$V_1 = g_1(Y_1),$$

where $g_1\colon \{1, \cdots, q_1\} \to R^{s_1}$,
and for $t > 1$

$$V_t = g_t(Z_{t-1}, Y_t),$$

where

$$g_t\colon \{1, \cdots, m_{t-1}\} \times \{1, \cdots q_t\} \to R^{s_t}.$$

The interpretation of V_t is that it represents an approximation to something we wish to know at the receiving end about X_t. In particular, one may have $s_t = n_t$ and consider V_t as approximating X_t itself.

The functional relationships described above are symbolized in Fig. 1.

The case of *full receiver memory* is included in this model. One need only identify Z_t with $(Y_1, \cdots, Y_t)$ and r_t with the concatenation function "append."

Furthermore, in this case,

$$m_t = \prod_{k=1}^{t} q_k.$$

2.2 The criteria

The performance of the system is defined by way of a sequence of distortion functions. For each t, a Borel measurable function

$$\psi_t: R^{n_t} \times R^{s_t} \to [0, \infty)$$

is given. Then

$$J_t = E\{\Psi_t(X_t, V_t)\}$$

measures the distortion at stage t. It is possible that J_t be infinite, but it is always well defined, as the composition of Borel functions is Borel and $\Psi_t \geq 0$.

2.3 The optimization problem

The problem to be considered is the following:

Given: The integers: T; $n_1, \cdots, n_T$; $q_1, \cdots, q_T$; $m_1, \cdots, m_T$; $s_1, \cdots, s_T$.

The distribution of the X sequence.

The distortion measures $\Psi_1, \cdots, \Psi_T$.

Choose: The functions $f_1, \cdots, f_T$; $g_1, \cdots, g_T$; $r_1, \cdots, r_T$ (the latter are redundant in the full memory case, i.e., when $m_t \geq q_1 q_2 \cdots q_t$ for all t). A choice of a system of such functions will be called a "design."

In order to:

Minimize (exactly or within ϵ) the sum

$$J = \sum_{t=1}^{T} J_t.$$

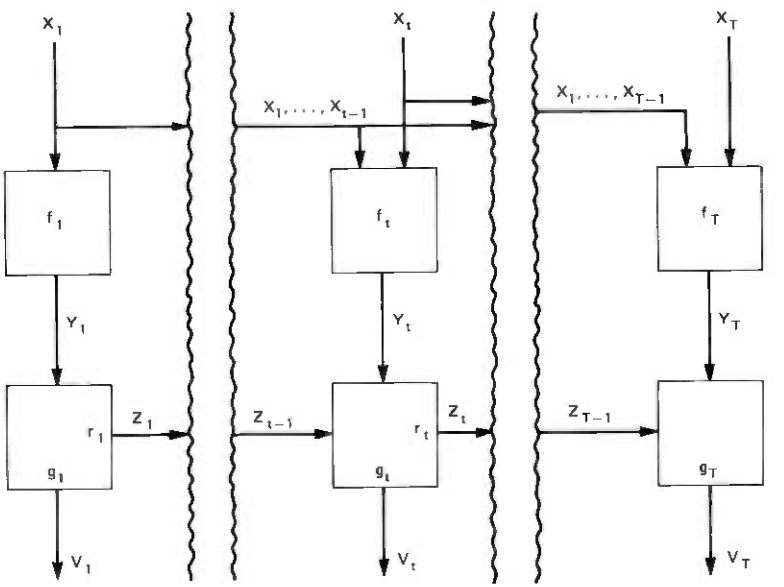


Fig. 1—General system.

Remark that nothing would be gained by having J as a nonnegative linear combination $\sum c_t J_t$ (for instance, with $c_t = e^{-\lambda t}$, a discount factor) because such $c_t \geq 0$ can simply be absorbed into the definition of Ψ_t.

It should be said that the freedom of having n_t, q_t, m_t, s_t, ψ_t depend upon t is not a matter of extra generality, but is essential to the proof techniques used in the sequel.

A design producing the values $(J_1, \cdots, J_T)$ is *at least as good* as a design producing $(J_1', \cdots, J_T')$ when $J_t \leq J_t'$ for all $t \in \{1, \cdots, T\}$. This, of course, implies the much weaker statement that $J = \sum J_t \leq J' = \sum J_t'$.

A design may exist which is at least as good as any other; it is called a *dominant* design. In general, however, no dominant design exists because the set in R^T of achievable vectors $(J_1, \cdots, J_T)$ does not have a corner $(J_1^*, \cdots, J_T^*)$ such that all other points of this set lie in the shifted orthant $J_t \geq J_t^*$, $t = 1, \cdots, T$. Instead, the set may have a *Pareto frontier* of "admissible" vectors, i.e., vectors $(J_1, \cdots, J_T)$ such that no vector $(J_1', \cdots, J_T')$ is achievable that has $J_t' \leq J_t$ for all t with strict inequality for some t.

2.4 Special encoder structures

The encoder f_t at a specific stage $t > k$ is said to have *memory structure of order k*, if there is a Borel function

$$\hat{f}_t: \{1, \cdots, m_{t-1}\} \times R^{n_{t-k+1}+\cdots+n_t} \to \{1, \cdots, q_t\}$$

such that

$$f_t(X_1, \cdots, X_t) = \hat{f}_t(Z_{t-1}, X_{t-k+1}, \cdots, X_t) \quad \text{a.s.}$$

This is equivalent to the assertion that Y_t is measurable on the σ-field generated by $Z_{t-1}, X_{t-k+1}, \cdots, X_t$. In other words, the encoder elaborates Y_t using only the k most recent source outputs $X_{t-k+1}, \cdots, X_t$ and the receiver's current memory Z_{t-1}.

III. THE MAIN THEOREM

The sequence $X_1, X_2, \cdots, X_T$ is said to be kth-order Markov, when, given any block of k consecutive X_t, the parts of the sequence preceding and following this block are conditionally independent. For $k = 1$, this is the ordinary Markov property. Note that the kth-order Markov property holds in a vacuous way if $T < k + 2$.

Most sequences can be approximated by kth-order Markov sequences for sufficiently large k. If this k is small compared to T, then the following main theorem provides a substantial simplification of the encoder optimization problem.

Theorem 1: *Suppose the source is* k*th-order Markov. Then, given any design, there is another design with the following properties:*

(*i*) *The new design differs from the given one only in the choice of encoders.*

(*ii*) *All encoders of the new design have memory structure of order k. (The last encoder f_T can even be made to have memory structure of order* 1.)

(*iii*) *The performance index J of the new design does not exceed the index of the old design.*

Postponing the proof of Theorem 1 to Section V, we comment here on its significance. It says, in particular, that for a Markov source and a receiver with perfect memory, one need only consider encoders which generate each code symbol Y_t using only the current source symbol X_t and the past code sequence $Y_1, Y_2, \cdots, Y_{t-1}$. This result is essentially dependent on the Markov property of the source as can be seen from the following example.

Take $T = 3$ and, for $t = 1, 2, 3$, let $n_t = s_t = 1$, $q_t = 2$, $\psi_t(X_t, V_t) = (X_t - V_t)^2$. Suppose the receiver has perfect memory. Suppose that the source sequence (X_1, X_2, X_3) takes just eight equally probable values, namely (13, 1, 3), (12, 1, 2), (11, 1, 1), (10, 1, 0), (−10, −1, 0), (−11, −1, 1), (−12, −1, 2), (−13, −1, 3).

At the first stage, if one considers only the minimization of J_1, one has a classical quantization problem for X_1. As X_1 takes its values in two separate equiprobable clusters, the minimum of J_1 is attained by letting Y_1 signal the sign of X_1 to identify the cluster. Then $V_1 = \pm 11.5$ and $J_1 = 1.25$. Any other choice of the first encoder yields a strictly larger value for J_1. Furthermore, Y_1 is already sufficient for the attainment of $J_2 = 0$, the second-stage receiver need not even look at Y_2. However, Y_2 can be used to help the third-stage receiver. If one lets Y_2 signal the parity of X_1, then $J_3 = 0$ is attainable by letting Y_3 signal whether $|X_3| \leq 1$ or not.

The design so obtained minimizes J_t for each t (it is "dominant"); *a fortiori*, it minimizes $J = \sum_1^3 J_t$, giving $J = 1.25$. However, the second-stage encoder does not have memory structure of order one.

Is it possible to achieve $J = 1.25$ with memory structure of order one although the source is not Markov? The answer is no for, if one changes the first-stage encoder, this alone will drive J_1 and, *a fortiori*, J above 1.25. But if the first encoder signals the sign of X_1 and the second encoder must have first-order structure, then the second encoder is useless. Indeed, X_2 contains no information not already contained in Y_1, and the receiver remembers Y_1. Now Y_1 is useless to the third-stage receiver,† and a single binary signal Y_3 is insufficient to distinguish among the four possible values of X_3. The best that can be

† $Y_1 = \operatorname{sgn} X_1$ and X_3 are independent.

done is to form Y_3 as in the previous design, giving $J_3 = 0.25$; hence, $J = 1.5$.

The optimum design requires encoder f_2 to "signal ahead" features of X_1 for the later benefit of receiver g_3. This phenomenon is ruled out for sources with the Markov property.

IV. TWO BASIC LEMMATA

All the results in this paper will be derived from two basic lemmata: a "two-stage lemma" and a more complex "three-stage lemma." Once these are obtained, the use of induction and of the technique of "repackaging" random variables will suffice.

4.1 The two-stage lemma

This lemma uses what is, in fact, the basic line of reasoning in Ref. 1. Consider a system with $T = 2$ and any joint distribution of the pair of random vectors (X_1, X_2). Observe that the content Z_1 of the receiver's memory at the beginning of stage 2 is a certain function of X_1; that is,

$$Z_1 = \phi(X_1),$$

where ϕ is a Borel function (in fact, it is the composition of f_1 and r_1). The second (and last) stage is characterized by the functions f_2 and g_2 with (Fig. 2)

$$Y_2 = f_2(X_1, X_2),$$

$$V_2 = g_2(Z_1, Y_2).$$

Lemma 1: *Given a two-stage system with a design in which f_2 does not have memory structure of order* 1, *one can change f_2 (and only f_2) so that it has this structure and the new design is at least as good as the given design.*

Proof: If only f_2 is changed, then J_1, ϕ, and g_2 remain as given. We have to show that, for a suitable change in f_2, J_2 can only decrease. Consider the function

$$F((Z_1, X_2), Y_2) \equiv \psi_2(X_2, g_2(Z_1, Y_2)).$$

As Y_2 is discrete and F is measurable (by its construction), a measurable function $\hat{f}_2$ exists (see the appendix) such that

$$F((Z_1, X_2), \hat{f}_2(Z_1, X_2)) \leq F((Z_1, X_2), Y_2)$$

for all values of Z_1, X_2, Y_2. Hence, by the substitution

$$Z_1 = \phi(X_1)$$

$$Y_2 = f_2(X_1, X_2)$$

$$\psi_2(X_2, g_2(\phi(X_1), \hat{f}_2(\phi(X_1), X_2))) \leq \psi_2(X_2, g_2(\phi(X_1), f_2(X_1, X_2))$$

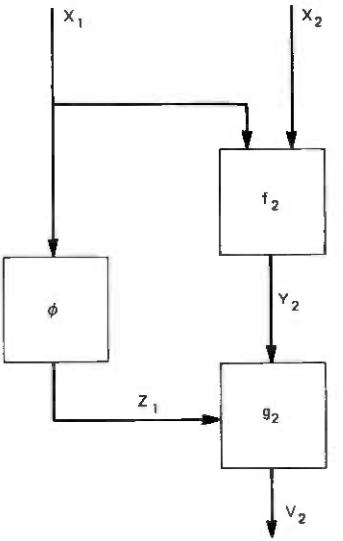


Fig. 2—Two-stage lemma.

holds for all X_1, X_2. As the functions ϕ_1, f_2 and $\hat{f}_2$ are measurable, both sides of the inequality are measurable. Since they are also nonnegative, the inequality persists when taking the expectation of both sides, whether finite or not. This establishes that J_2 can only decrease as claimed.

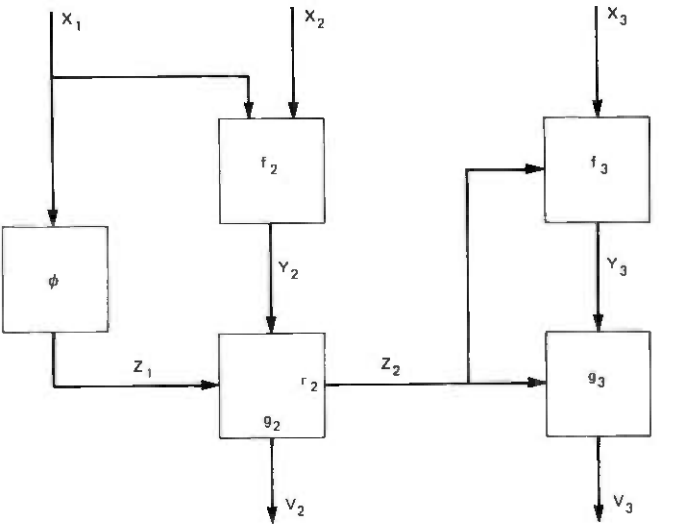


Fig. 3—Three-stage lemma.

4.2 The three-stage lemma

Consider a three-stage system ($T = 3$) with a Markov source. Assume that the last encoder f_3 already has first-order memory structure, while f_2 does not (Fig. 3).

Lemma 2: *Under the above assumptions, one can replace f_2 by an encoder $\hat{f}_2$ having memory structure of order one, without increasing the total cost $J = J_1 + J_2 + J_3$.*

Proof: The first-stage cost J_1 is unaffected by changes in f_2 and the effect of the first-stage design is to generate the receiver memory Z_1 as a certain measurable function $Z_1 = \phi(X_1)$, where ϕ is the composition $r_1 \circ f_1$. By assumption, f_3 can be written in the form

$$Y_3 = f_3(Z_2, X_3)$$

where

$$Z_2 = r_2(Z_1, Y_2).$$

The cost incurred in the last two stages can thus be written

$$\psi_2(X_2, g_2(Z_1, Y_2)) + \psi_3(X_3, g_3(r_2(Z_1, Y_2), f_3(r_2(Z_1, Y_2), X_3))) \equiv F(Z_1, X_2, X_3, Y_2),$$

defining the measurable function F.

Consider now the conditional expectation

$$E\{F(Z_1, X_2, X_3, Y_2) | X_1, X_2\}.$$

Because X_3 is a finite dimensional random vector, a regular conditional distribution of X_3 exists for any condition. In view of the Markov property, conditioning on the pair (X_1, X_2) is equivalent to conditioning on X_2 only. Let $P(dX_3 | X_2)$ be a regular version of this conditional distribution.

Then the conditional expectation under consideration can be written

$$\int P(dX_3 | X_2) F(Z_1, X_2, X_3, Y_2),$$

where Z_1 and Y_2, which depend only on the conditioning variables X_1, X_2, can be treated as fixed. This integral defines a measurable function (nonnegative and possibly extended real-valued)

$$G(Z_1, X_2, Y_2).$$

For any choice of f_2, the sum $J_2 + J_3$ will be given by the expectation of G. Note that X_1 enters G *only* by way of Z_1 and Y_2.

As in Lemma 1, a measurable function $\hat{f}_2$ exists such that, for all Z_1, X_2 and Y_2,

$$G(Z_1, X_2, \hat{f}_2(Z_1, X_2)) \leq G(Z_1, X_2, Y_2).$$

Substituting $Z_1 = \phi(X_1)$, $Y_2 = f_2(X_1, X_2)$ and taking the expectations of both sides of this inequality, implies, by the chain rule, that $J_2 + J_3$ cannot increase when f_2 is replaced by $\hat{f}_2$.

V. PROOF OF THE MAIN THEOREM

To begin with, the situation of the last stage is always a special one, as the following lemma shows.

Lemma 3: *For any source statistics and any design, one can replace the last encoder by one having memory structure of order one, without performance loss.*

Proof: The given T-stage system can be considered as a two-stage system, by setting

$$\tilde{X}_1 = (X_1, X_2, \cdots, X_{T-1})$$

$$\tilde{X}_2 = X_T$$

$$\tilde{Z}_1 = Z_{T-1} = \phi(\tilde{X}_1)$$

$$\tilde{Y}_2 = Y_T$$

$$\tilde{f}_2(\tilde{X}_1, \tilde{X}_2) = f_T(X_1, X_2, \cdots, X_{T-1}, X_T)$$

$$\tilde{g}_2(\tilde{Z}_1, \tilde{Y}_2) = g_T(Z_{T-1}, Y_T)$$

$$\tilde{V}_1 = (V_1, \cdots, V_{T-1})$$

$$\tilde{V}_2 = V_T$$

$$\tilde{\psi}_1(\tilde{X}_1, \tilde{V}_1) = \sum_{t=1}^{T-1} \psi_t(X_t, V_t)$$

$$\tilde{\psi}_2(X_2, V_2) = \psi_T(X_T, V_T),$$

which amounts to a change in notation. Of course,

$$\tilde{n}_1 = \sum_{t=1}^{T-1} n_t,$$

a substantial increase in dimension.

By Lemma 1, there exists an encoder $\hat{\tilde{f}}_2$ which has the structure

$$\tilde{Y}_2 = \hat{\tilde{f}}_2(\tilde{Z}_1, \tilde{X}_2)$$

and whose use does not increase $\tilde{J}_2$. Reverting to the original notation, this corresponds to an encoder $\hat{f}_T$ with the structure

$$Y_T = \hat{f}_T(Z_{T-1}, X_T)$$

whose use does not increase J_T. As the other J_t are unchanged, the lemma is proved.

The above fact is the starting point for the proof of the main theorem with $k = 1$.

Lemma 4: The main theorem holds for $k = 1$. That is, for a Markov source and any design, one can replace the encoders by appropriate encoders having first-order memory structure without increase in the expected cost J.

Proof: Using backward induction, one can first replace f_T by $\hat{f}_T$, as in Lemma 3. Now suppose the encoders for stages $t + 1, t + 2, \cdots, T$ already have memory structure of order one. It must be shown that f_t can be replaced by $\hat{f}_t$ with such structure, without increase in expected total cost. To this effect, the T-stage system can be considered as a three-stage system, in which the third stage has first-order memory structure and the source is Markov, as follows. Set

$$\tilde{X}_1 = (X_1, \cdots, X_{t-1}) \quad \left(\text{thus, } \tilde{n}_1 = \sum_{i=1}^{t-1} n_i\right)$$

$$\tilde{Z}_1 = Z_{t-1} = \phi(\tilde{X}_1)$$

$$\tilde{X}_2 = X_t$$

$$\tilde{Y}_2 = Y_t$$

$$\tilde{Z}_2 = Z_t = \tilde{r}_2(\tilde{Z}_1, \tilde{Y}_2) = r_t(Z_{t-1}, Y_t)$$

$$\tilde{V}_2 = V_t = \tilde{g}_2(\tilde{Z}_1, \tilde{Y}_2) = g_t(Z_{t-1}, Y_t)$$

$$\tilde{\psi}_2(\tilde{X}_2, \tilde{V}_2) = \psi_t(X_t, V_t)$$

$$\tilde{X}_3 = (X_{t+1}, \cdots, X_T), \qquad \left(\tilde{n}_3 = \sum_{i=t+1}^{T} n_i\right)$$

$$\tilde{Y}_3 = (Y_{t+1}, \cdots, Y_T), \qquad \left(\tilde{q}_3 = \prod_{i=t+1}^{T} q_i\right)$$

$$\tilde{V}_3 = (V_{t+1}, \cdots, V_t) = \tilde{g}_3(\tilde{Z}_2, \tilde{Y}_3), \qquad \left(\tilde{s}_3 = \sum_{i=t+1}^{T} s_t\right).$$

The latter relation follows from the fact that each V_θ, $\theta > t$, is given by $g_\theta(Z_{\theta-1}, Y_\theta)$ and the variables $Z_{\theta-1}$, Y_θ are known functions of Z_t, Y_{t+1}, Y_{t+2}, $\cdots$, Y_T using the memory update functions. Then

$$\tilde{\psi}_3(\tilde{X}_3, \tilde{V}_3) = \sum_{\theta=t+1}^{T} \psi_\theta(X_\theta, V_\theta).$$

As the encoders for stages $t + 1, \cdots, T$ already have first-order

memory structure, their effect is to define an encoder

$$\tilde{Y}_3 = \tilde{f}_3(\tilde{Z}_2, \tilde{X}_3)$$

because each of the Y_θ, $\theta > t$, included in $\tilde{Y}_3$ is given by a function $f_\theta(Z_{\theta-1}, X_\theta)$ and the variables $Z_{\theta-1}$, X_θ are known functions of $\tilde{Z}_2$, $\tilde{X}_3$; i.e., of Z_t, X_{t+1}, $\cdots$, X_T using the memory update functions and recursion. The given encoder at stage t has the general form $Y_t = f_t(X_1, X_2, \cdots, X_{t-1}, X_t)$ which translates to

$$\tilde{Y}_2 = \tilde{f}_2(\tilde{X}_1, \tilde{X}_2).$$

The new source $(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$ is Markov since $\tilde{X}_1 = (X_1, \cdots, X_{t-1})$ and $\tilde{X}_3 = (X_{t+1}, \cdots X_T)$ are conditionally independent given $\tilde{X}_2 = X_t$, by the assumed Markov property of the original source.

Thus, the three-stage system satisfies the assumptions of Lemma 2, and $\tilde{f}_2$ can be replaced without loss of total expected cost by $\hat{\tilde{f}}_2$, which has the structure

$$\tilde{Y}_2 = \hat{\tilde{f}}_2(\tilde{Z}_1, \tilde{X}_2).$$

This translates to an encoder

$$Y_t = \hat{f}_t(Z_{t-1}, X_t)$$

for the original problem. Since the notational changes do not influence total cost, the inductive step, and therefore the lemma, is proved.

Note that the above induction is carried out down to $t = 2$ because f_1 cannot help but have the desired structure, albeit trivially so.

Turning to the case of general k, observe that the encoders for the first k stages have memory structure of order k in a trivial way, whatever their design, and for the last stage, Lemma 3 applies. Thus the conclusion of the main theorem holds for $T \leq k + 1$ trivially, as does the assumption on the source. Hence, assume $T \geq k + 2$.

The essence of the proof is a "sliding block" repackaging of the source variables.

Let $$\tilde{X}_t = (X_t, X_{t+1}, \cdots, X_{t+k-1})$$

for $$t = 1, \cdots, \tilde{T} \quad \text{where} \quad \tilde{T} = T - k + 1 \geq 3.$$

Then the sequence $(\tilde{X}, \tilde{X}_2, \cdots, \tilde{X}_{\tilde{T}})$ is Markov. For the variables, let

$$\tilde{Y}_1 = (Y_1, \cdots, Y_k),$$

$$\tilde{Y}_t = Y_{t+k-1}, \quad \text{for } t = 2, \cdots, \tilde{T}$$

$$\tilde{Z}_t = Z_{t+k-1}, \quad \text{for } t = 1, \cdots, \tilde{T}$$

$$\tilde{V}_1 = (V_1, \cdots, V_k),$$

and $$\tilde{V}_t = V_{t+k-1}, \quad \text{for } t = 2, \cdots, \tilde{T}.$$

The functions relating these variables are written as follows:

$$\tilde{Y}_1 = \tilde{f}_1(\tilde{X}_1)$$

summarizes the action of the first k encoders, which will remain unchanged as they already (trivially) have memory structure of order k. For $t > 1$,

$$\tilde{Y}_t = \tilde{f}_t(\tilde{X}_1, \tilde{X}_2 \cdots, \tilde{X}_t) = f_{t+k-1}(X_1, \cdots, X_{t+k-1})$$

where $\tilde{f}_t$ is not uniquely defined by this relation. It can be made unique and measurable by requiring (for example) that for $\theta = 2, \cdots, t$, the function $\tilde{f}_t$ depends on the argument $\tilde{X}_\theta = (X_\theta, \cdots, X_{\theta+k-1})$ only through its last component $X_{\theta+k-1}$. However, any measurable $\tilde{f}_t$ satisfying the identity is acceptable.

The receivers are characterized by their memory updating functions:

$$\tilde{Z}_1 = \tilde{r}_1(\tilde{Y}_1)$$

summarizes the recursive buildup of Z_k from $(Y_1, \cdots, Y_k)$ using $r_1, \cdots, r_k$. For $t > 2$, $\tilde{r}_t$ is defined by

$$\tilde{Z}_t = \tilde{r}_t(\tilde{Z}_{t-1}, \tilde{Y}_t) = r_{t+k-1}(Z_{t+k-2}, Y_{t+k-1}).$$

Likewise,

$$\tilde{V}_1 = \tilde{g}_1(\tilde{Y}_1)$$

summarizes the action of the first k decoders (including their memory updating). For $t > 2$, $\tilde{g}_t$ is defined by

$$\tilde{V}_t = \tilde{g}_t(\tilde{Z}_{t-1}, \tilde{Y}_t) = g_{t+k-1}(Z_{t+k-2}, Y_{t+k-1}).$$

Finally, $\tilde{\psi}_1(\tilde{X}_1, \tilde{V}_1) = \sum_{t=1}^{k} \psi_t(X_t, V_t)$ and for $t > 2$

$$\tilde{\psi}_t(\tilde{X}_t, \tilde{V}_t) = \psi_{t+k-1}(X_{t+k-1}, V_{t+k-1}),$$

where $\tilde{\psi}_t$ depends on argument $\tilde{X}_t$ only through the component X_{t+k-1}. Now Lemma 4 can be applied to this $\tilde{T}$ stage system with Markov source. Without increase in total cost, the encoders $\tilde{f}_2, \cdots, \tilde{f}_T$ can be replaced by encoders $\hat{\tilde{f}}_t$ with first-order memory structure, i.e.,

$$\tilde{Y}_t = \hat{\tilde{f}}_t(\tilde{Z}_{t-1}, \tilde{X}_t), \qquad \text{for } t = 2, \cdots, \tilde{T}.$$

Expressing this in terms of the original variables, the functions f_t for $t = k + 1, \cdots, T$ are replaced by functions $\hat{f}_t$ satisfying

$$Y_{t+k-1} = \hat{f}_{t+k-1}(Z_{t+k-2}, X_t, X_{t+1}, \cdots X_{t+k-1}) \qquad \text{for } t = 2, \cdots, T - k + 1$$

or equivalently

$$Y_t = \hat{f}_t(Z_{t-1}, X_{t-k+1}, \cdots, X_t) \qquad \text{for } t = k + 1, \cdots, T.$$

These $\hat{f}_t$ exhibit memory structure of order k, so that the main theorem is proved.

VI. THE CASE OF DELAYED DISTORTION MEASURES

The basic model of Section II can be modified as follows for the case in which a delay of $\delta > 0$ steps is included in the definition of distortion. The first change is that the variables $V_1, \cdots, V_\delta$ are simply not generated, the receiver spends its first δ periods just accumulating observations of $Y_1, \cdots, Y_\delta$ and updating its memory accordingly.

The second change is that, for $t > \delta$, distortion is measured by a function $\psi_t(X_{t-\delta}, V_t)$ whose expectation defines J_t. The design objective is to minimize

$$J = \sum_{t=\delta+1}^{T} J_t.$$

For this situation, the following structure simplifying result holds.

Delay Theorem: *Suppose that the source is* kth *order Markov and that the distortion is defined with delay* δ. *Then any given design can be replaced, without loss, by one in which the encoders have memory structure of order* $\max(k, \delta + 1)$.

Proof: In case $k \geq \delta + 1$, one can perform the same transformation of the point of view as in the proof of the main theorem. Indeed, this transformation gives cost functions of the form

$$\tilde{\psi}_t(\tilde{X}_t, \tilde{V}_t), \qquad t = 1, \cdots, T - k + 1,$$

where

$$\tilde{X}_t = (X_t, \cdots, X_{t+k-1}), \ \tilde{V}_1 = (V_1, \cdots, V_k), \ \tilde{V}_t = V_{t+k-1}.$$

This is compatible with the delay criterion, as follows:

$$\tilde{\psi}_1(\tilde{X}_1, \tilde{V}_1) = \sum_{t=1}^{k-\delta} \psi_t(X_t, V_{t+\delta})$$

and for $t = 2, \cdots, T - k + 1$

$$\tilde{\psi}_t(\tilde{X}_t, \tilde{V}_t) = \psi_{t+k-1}(X_{t+k-\delta-1}, V_{t+k-1}),$$

where it happens that $\tilde{\psi}_t$ depends upon $\tilde{X}_t$ only through the component $X_{t+k-\delta-1}$.

Therefore the argument of the main theorem applies: One can use encoders with memory structure of order k. In fact, the above shows that this conclusion is valid for any criteria of the form

$$\sum_t \psi_t(X_t, X_{t+1}, \cdots, X_{t+k-1}, V_{t+k-1}).$$

As for the case $k < \delta + 1$, observe that the source is then, *a fortiori*, Markov of order $\delta + 1$. Hence, the first case applies to yield memory structure of order $\delta + 1$, as claimed.

VII. CONCLUDING REMARKS

A few extensions of the results are of interest.

(*i*) The proof of the three-stage lemma goes through under the weaker assumption that f_3 depends upon Z_2, X_2, and X_3.

(*ii*) All the results in this paper remain true for V_t restricted to given subsets of R^{s_t}. This would correspond to quantization levels fixed in advance, as opposed to their selection as part of the design.

(*iii*) Suppose $\delta = 0$, $k = 1$, and the encoder is restricted *a priori* to be a finite state machine of the type

$$W_t = h_t(W_{t-1}, X_t),$$

$$Y_t = f_t(W_{t-1}, X_t),$$

where W_t is a discrete variable representing the contents of the encoder's memory. Then the main theorem implies that it is optimal to take $Z_t = W_t$ and $h_t = r_t$ since this simulation of the receiver's memory produces the argument required for the generation of Y_t. This result was obtained independently by N. T. Gaarder.

VIII. ACKNOWLEDGMENTS

The author is indebted to B. McMillan for bringing this problem to his attention and to D. Slepian for stimulating discussions.

APPENDIX

Let X be a set and $\mathscr{B}$ a σ-algebra of subsets of X. Let Y be a finite set $\{1, \cdots, q\}$. A function $F: X \times Y \to R$ is called measurable if for each y in Y, the function $F(\cdot, y): X \to R$ is $\mathscr{B}$-measurable.

Then it follows that a function $f: X \to Y$ exists such that

$$F(x, f(x)) \leq F(x, y)$$

holds for all $x \in X$ and $y \in Y$ and the function f is $\mathscr{B}$-measurable (which means that $\{x \mid f(x) = y\}$ is in $\mathscr{B}$ for each y).

Since y takes only finitely many values, it is evident that, for each x, one can select an $f(x)$ to satisfy the inequality. However, there may be many x for which the minimizing y is not unique. This creates the need for a choice of values in defining f, and if such a choice were made in a totally arbitrary manner, it is possible that the resulting f not be $\mathscr{B}$-measurable. What is needed is the (elementary) proof that, for a reasonable way to resolve ambiguous choices, the resulting f is automatically $\mathscr{B}$-measurable.

Given $y \in Y$, consider the set A_y of all x for which y is among the minimizing values, this set is measurable because it is defined by a finite number of inequalities among measurable functions, namely, for each $y' \in Y$,

$$F(x, y) \leq F(x, y').$$

The sets A_y cover X but with overlaps. To remove the overlaps, use the numerical indexing of Y to define

$$B_1 = A_1$$

and, for $y > 1$,

$$B_y = A_y - \bigcup_{i=1}^{y-1} A_i.$$

This construction preserves measurability and removes overlap. Thus, if f is defined to take value y on B_y, the desired result is attained. This amounts to stipulating that, when the minimum is attained for more than one element of Y, $f(x)$ is defined as the element with the smallest label.

REFERENCE

1. B. McMillan, "Communicating Systems Which Minimize Coding Noise," B.S.T.J., *48*, No. 9 (November 1969), pp. 3091–3113.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

Current-Access Magnetic Bubble Circuits

By A. H. BOBECK, S. L. BLANK, A. D. BUTHERUS, F. J. CIAK, and W. STRAUSS

(Manuscript received March 9, 1979)

Experimental and theoretical results from our work on current-access technology show promise for high-density, ~10^7 bits/cm^2, and high-frequency, $f > 1$ MHz, bubble devices. We have operated current-access devices where the bubble-driving fields derive from two patterned conducting sheets instead of orthogonal field coils. Margins for generation, propagation, and transfer were studied on 8-μm periods at 1 MHz. These 8-μm period structures typically required 1.5 mA/μm per conducting sheet and dissipated 14 μW/bit. Single conducting-sheet, current-access circuitry also propagates bubbles but offers less design flexibility. We present design criteria, magnetic field equations, and design curves. Implementation of these devices required new magnetic materials with quality factor Q comparable to available garnets, yet higher mobility and lower dynamic coercivity. Of the three systems, $(YLuSmCa)_3(FeGe)_5O_{12}$, $(LaLuSmCa)_3(FeGe)_5O_{12}$, *and* $(LaLuSm)_3(FeGa)_5O_{12}$, *the last appears best suited; some room temperature characteristics of the composition* $La_{0.6}Lu_{2.1}Sm_{0.3}Ga_{0.9}Fe_{4.1}O_{12}$ *are* $4\pi M_s = 470G$, $\mu = 750\ cmS^{-1}\cdot Oe^{-1}$, $d = 1.6\ \mu m$, *and* $\Delta H_c = 2$ *Oe across a bubble for the threshold of motion. Necessary improvements in processing were made with a radio-frequency, chlorine-containing plasma etch which produced metal patterns identical to those of the etch mask. We anticipate that current-access devices, when compared to conventional field-access devices, will achieve higher data rates, lower power consumption per bit, and greater storage densities with existing processing technologies.*

I. INTRODUCTION

We have been investigating the potential of conductor-access bubble circuitry[1] to meet the need for a high-density, high-capacity, low-cost, nonvolatile memory device. There are two basic ways to move bubbles: field-access[2] and current-access (conductor-access). Field-access de-

vices, the only type in use at this time, require a pair of orthogonal drive coils to provide an in-plane magnetic field which, coupled to a bubble data stream via structured permalloy or ion-implanted features,[3] provides the necessary bubble drive forces.

It is difficult, however, to take advantage of the data rates intrinsic in field-access bubble devices as they are scaled down in size. In particular, the volt-ampere product increases with increasing frequency, making economical coil drive circuits difficult to design. Present field-access devices are limited to about a 250-kHz rotating field rate and are thus too slow for applications such as television frame storage. A bit rate per chip of 1 Mb/s or greater would make bubbles attractive to a wider range of potential customers.

In traditional current-access bubble circuitry, meandering conductor strips of relatively high resistance, rather than drive coils, make it possible to provide the necessary currents to move bubbles at stepping rates of 1 MHz and greater with relative ease but give rise to a concern that current-access chips would dissipate excessive power, especially at the higher bit storage densities such as 10^7 b/cm^2.

The disadvantage of the current-access circuits disclosed to date has been the complexity of the conductor patterns themselves. We have developed new configurations, however, that use rather simple apertured sheet geometries and which should dissipate less power as well. Bubble circuits with an 8-μm period, a four-times improvement in storage density over present circuits, are expected to be achieved while maintaining the same design rules of the 16-μm period field-access circuits now in manufacture at Western Electric, Reading.[4]

II. HISTORY OF CURRENT ACCESS DEVELOPMENT

It is not surprising that the very first experiments to control bubble movement used conductor drive. In these first experiments, adjacent posts of a "waffle-iron" ferrite base plate were looped by conductors as shown in Fig. 1.[1] Conductors ϕ_1, ϕ_2, and ϕ_3 driven by unipolar current pulses produced bubble motion with the direction determined by the pulse sequence. Bubble domains had to overlap from one post to another before proper operation could be obtained.

A continued evolution in circuit design next led to the single conductor design of Copeland et al.,[5] shown in Fig. 2. Note the conductor which meanders back and forth across a groove which has been etched into the surface of the epitaxial garnet layer. Undulations in the width of the groove define preferred bubble rest positions which supplement a bipolar current drive to produce bubble motion. It is a basic requirement of this design, however, that a rather precise geometrical fit be maintained between the bubble and the groove pattern. The ideal bubble diameter is, in fact, one-half of the conductor period making

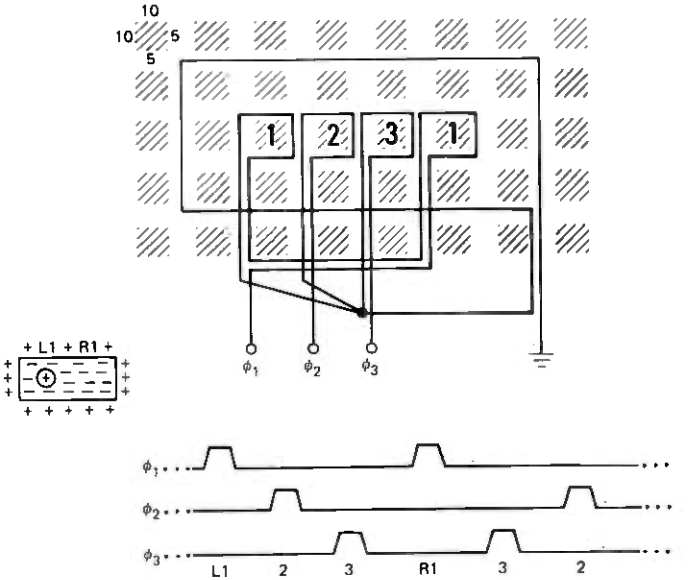


Fig. 1—Schematic of the first bubble shift register circuit. Bubbles in orthoferrites were moved three steps with a waffle-iron base plate.

bubble-to-bubble interactions prohibitive at the maximum storage density and hence the circuit could only be operated with data in every other position. A variation in which grooved tracks were replaced by ion-implanted tracks of similar shape was also operated.[3] About this time, effort in conductor propagation was terminated in favor of "field-access" propagation.

Conductor propagation re-emerged in 1974 when Walsh and Charap[6] proposed a perforated conducting sheet driven by either a rotating or oscillating current as a novel bubble drive. Their computer simulation indicated that operation comparable to that of field-access devices could be accomplished. Their paper went unnoticed.

In 1975, the bubble lattice storage device was announced by Voegeli et al.[7] Conductors were used to move bubbles in the main storage area, the entrance-exit line, and the detector. In the most recent paper on the bubble lattice, Hu et al.[8] state "Ironically, the "simple" shift register propagation of an isolated-bubble has one of the smallest margins in the chip." Hu et al. were somewhat handicapped in their design since bias field compatibility had to be maintained with the main storage area. Nonetheless, their report was not encouraging.

Then in 1977, Dekker et al.,[9] Phillips Research Labs, introduced a design which combined conductor and field access on a major-minor chip. They replaced the usually field-driven major track by a conductor-driven track that operated at a ten-times-higher stepping rate. In

this way, they achieved a fivefold reduction in access time. Typical of the Dekker designs is that illustrated in Fig. 3. The "conductor" is in reality a dual-layer, permalloy-gold sandwich in which the gold layer has been included to lower the sheet resistance and, consequently, the

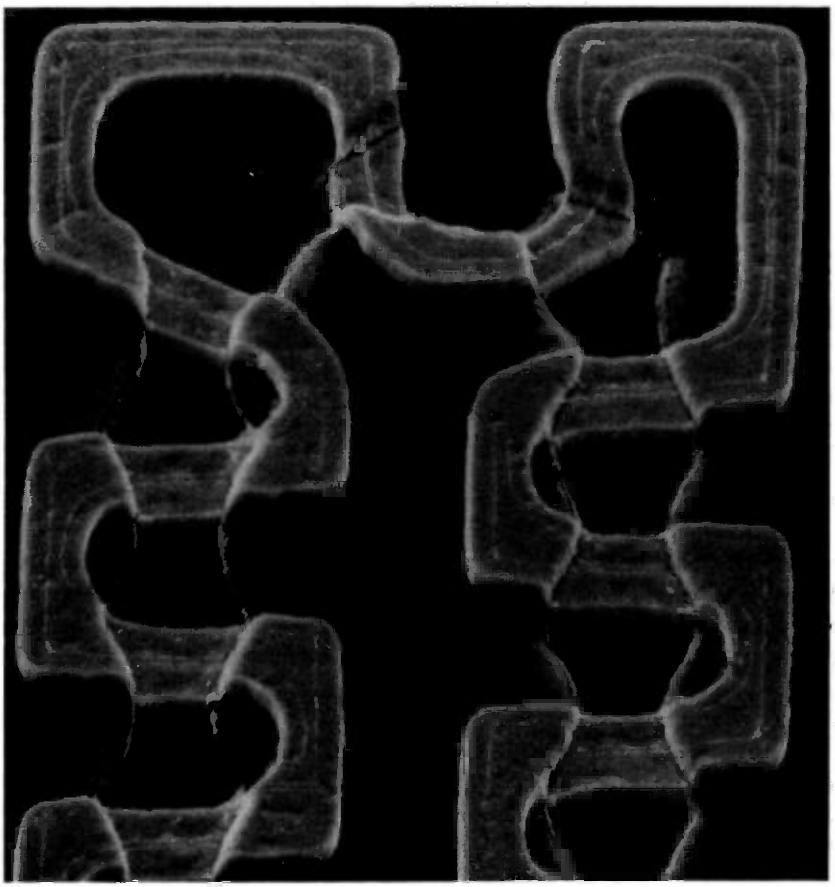


Fig. 2—An SEM of a turn of a conductor-groove circuit. The circuit period is 16 μm.

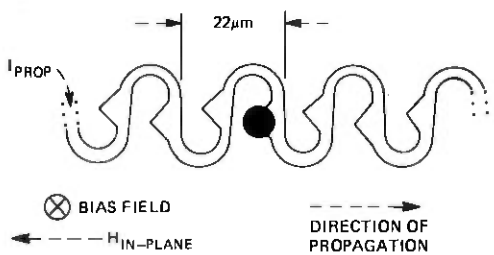


Fig. 3—A current-driven major track. An in-plane field generates poles at points of the "conductor," which is in reality a conductor-permalloy sandwich.

drive dissipation. A nonrotating in-plane field, supplied by the very coils that operate the field-access portion of the chip, magnetizes the triangular permalloy points, converting them to low energy resting positions for bubbles. The primary drive for bubbles comes from areas defined by the conductor loops driven by a bi-polar current to alternately attract and repel bubbles. As a result, bubbles step from loop to loop but always move to the right because of the directionality induced by the triangular points described above. Bubbles attracted to these points offset to the right from their otherwise neutral position within a loop.

Dekker reported a power dissipation of 0.6 mW/b for this structure. It is thus difficult to conceive that their design could be used in an all-conductor chip, since the power dissipation for even a 100-kb chip would be 60 watts. We see in Section VII that chip partitioning can be used to reduce the dissipation considerably, but unlikely, in this case, to a useful level. Copeland's design, on the other hand, did operate at very low power per bit, but it was both wasteful of storage area and difficult to process. It was difficult to process because its minimum feature dimension was a very small portion of the circuit period.

This, then, was the reported status of current access bubble circuits. The expected advantages of higher data rates, ease of driving, and simple package design had not been realized.

III. APERTURED SHEET FUNDAMENTALS

It will be helpful to the understanding of the device structures covered later if we digress at this point to some design considerations peculiar to apertured sheet devices. First, let us consider some analogies that exist between the more familiar permalloy structures and the conductor structures that we introduce in this and the following sections.

An analogy exists between the stray field of an isolated permalloy bar immersed in a uniform magnetic field H and the stray field generated by the distortion of an otherwise uniform current I in a sheet conductor with a slot.[6] In either case, an isolated magnetic dipole is formed. This situation is shown in Fig. 4, where it can be seen that a "plus"-pole-seeking bubble would be attracted to the upper opening of the slot just as it would be attracted to the upper end of the permalloy bar. Note also that the directions of the magnetic field H and the current I are at right angles while producing identically oriented dipoles. We must be careful not to carry this equivalence too far, since the field H is applied over a volume, whereas the current I is confined to a plane. The practical significance is that a slot can completely block a current whereas a permalloy bar cannot completely shunt an applied field.

Next, it is useful to define a sheet current density $J(\text{mA}/\mu\text{m}) = I/\text{W}$,

where I(mA) is the current applied to a sheet of width W(μm). We can relate current density in conductor devices to inplane field intensity in permalloy devices by observing that an infinitely wide conductor carrying a current density J = 1 mA/μm supports an in-plane field of 6.3 Oe. This is illustrated in Fig. 5.

There are, of course, an infinite variety of aperture profiles that can replace the slot of Fig. 4. A number that we have considered are shown in Fig. 6. By fabricating large-scale models, energizing these in a left-to-right direction from a 30-kHz current source and then measuring the signals induced in a uniformly wound single-layer coil scaled in dimension to that of a hypothetical bubble, we obtained the z-field contours also included in the figure. The z-component is that field

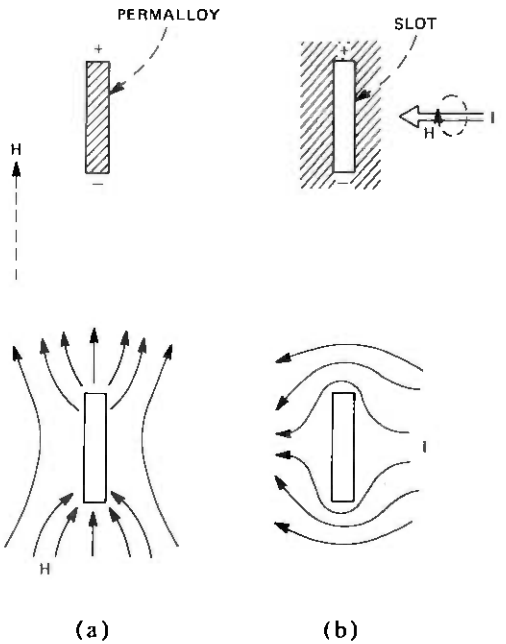


Fig. 4—Magnetic field H applied to a permalloy bar results in a dipole field. Current I applied to a conducting sheet produces a similar field.

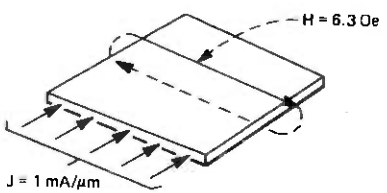


Fig. 5—A conducting sheet, infinite in extent, develops a 6.3-Oe surface field when driven at J = 1 mA/μm.

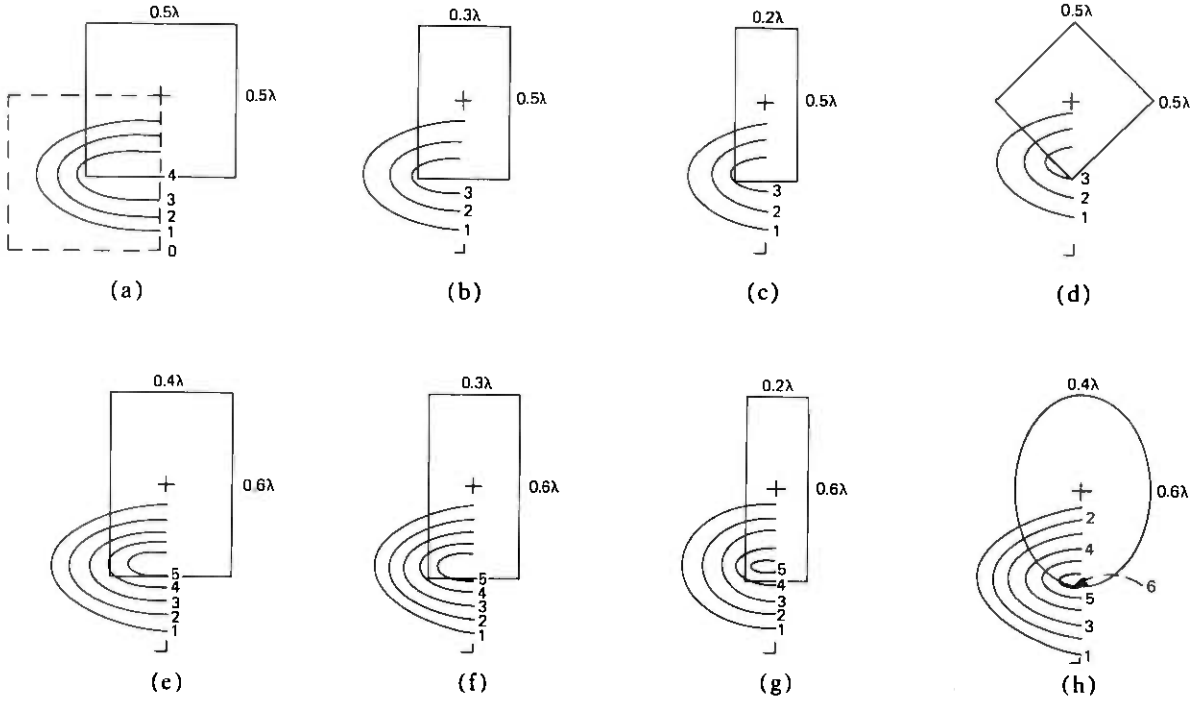


Fig. 6—Contours proportional to the z-component of field measured on scale models are given for apertures (a) through (h). Current flow was from left to right.

component normal to the sheet itself. The results actually apply to a square array of identical apertures positioned at center-to-center separations of λ. Only one quadrant of field is shown, as the others follow by symmetry considerations. Especially note how closely the field peaks and valleys are confined to the vicinity of the apertures. The oversized elements in the lower row give rise to larger fields than the shorter elements of the upper row, but at the expense of a higher power dissipation. We find the shapes "*b*," "*f*," and "*h*" optimal when items such as the ease of fabrication are also considered. Further details on this technique are given in Section X.

Similar results can also be obtained analytically and we have, in fact, devoted most of Section X to this end. One advantage of current-access bubble propagation is that they can be given a more precise mathematical treatment than is possible with field-access devices. Non-linear, permalloy-to-bubble interactions are not involved except, perhaps, in the detector itself. For example, we have calculated the field contours caused by current flow past a circular hole in an otherwise continuous infinite sheet. Two of the resultant design curves are given in Fig. 7, where h is the garnet thickness, a is the hole diameter, and Z_s is the elevation of the conductor sheet above the upper surface of the garnet. These contours give the average field in the garnet for an applied current density of 1 mA/μm. Conceptually, the role of the hole can be viewed as converting an otherwise uniform in-plane field (Fig. 5) into a bipolar field at about a 50-percent efficiency. This compares favorably to the conversion efficiency of a well-designed permalloy feature, which is also approximately 50 percent.

Next we consider the properties of an array of parallel conductor strips such as we encounter when detectors are discussed. The analytical results shown in Fig. 8 have been derived from design curves located in Section X. The parameters apply to a circuit period of 8 μm. For J = 1 mA/μm, the peak-to-peak bias field modulation is 9 Oe, meaning that a bubble "riding" in the field trough will have its bubble-to-strip transition field increased by 4.5 Oe. In an expander-type detector, it will be desirable to increase this transition field to at least 15 Oe so we can conclude that the current density in the detector will be increased to ~4 mA/μm. In conductor circuits, there is ample opportunity to do this.

We now present a problem peculiar to the apertured sheet geometry. Current in the conductor sheet not only generates the local field inhomogenities due to the apertures and an overall inplane field, but it also generates a gross z-field. The situation is shown in Fig. 9, where the field values pertain to J = 1 mA/μm. As might be expected, this field peaks at the edge of the conductor. As a consequence, any bubble located near the edge of a driven conductor sheet experiences a wide

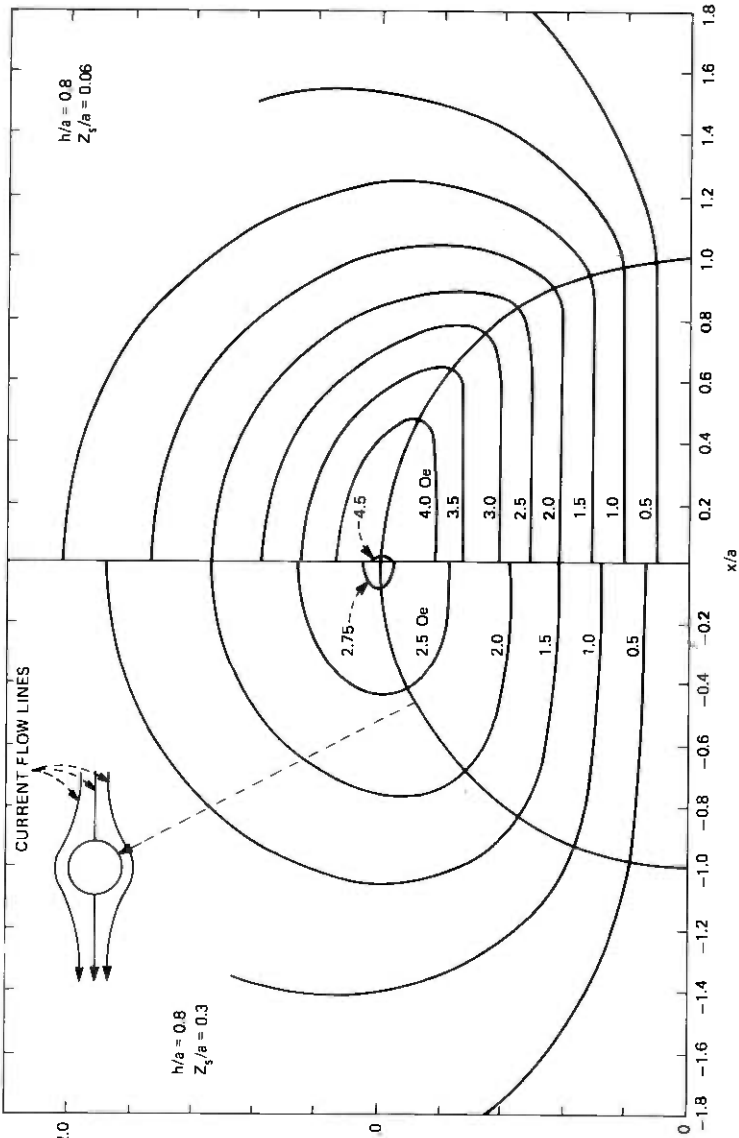


Fig. 7—Calculated z-field contours in the vicinity of a circular hole for a current density $J = 1$ mA/μm.

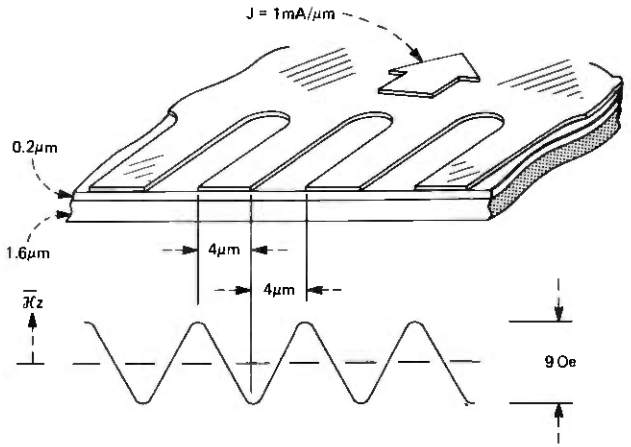


Fig. 8—$\bar{H}z$ variation generated by an array of parallel conducting strips. The dimensions are typical of an 8-μm period expander detector.

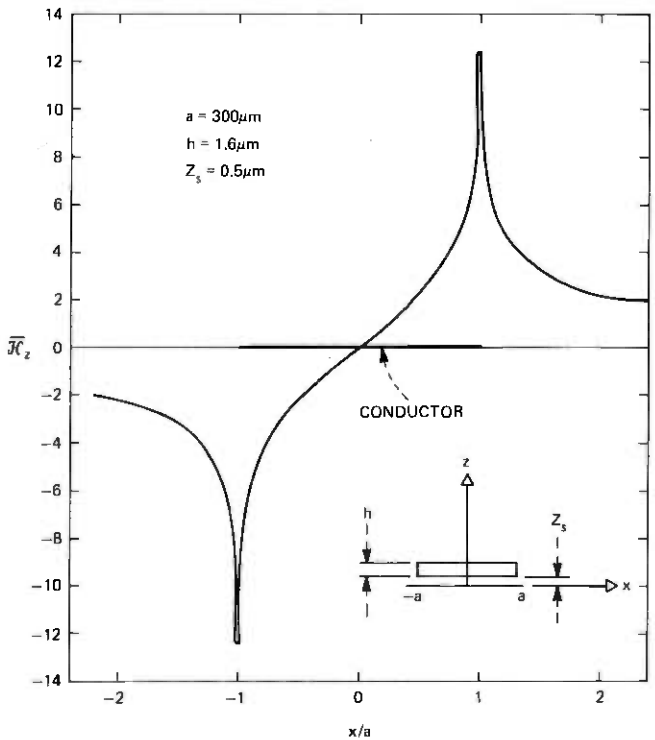


Fig. 9—Curve of the z-field due to a uniform current density J = 1 mA/μm flowing in a conducting sheet of width 2a.

bias change. This problem is substantially alleviated if active bubble circuits are confined to the central 80 percent of a sheet. Other solutions such as the use of a return conductor can also be employed.

Finally, let us contrast the performance of apertures and meandering conductors with respect to field localization and power efficiency. A comparison between the field contour of the undulating conductor given in Fig. 10 and those of Fig. 6 reveal that the fields of the latter more closely resemble those of a dipole. Furthermore, the fields produced by neighboring apertures are seen to be distinct and separate. Such is not the case for the former and, in fact, the edges of the conductor are biased to nearly one-half the peak field.

The structure of Fig. 10 requires one-third the current/cell to produce the same peak field as that of Fig. 6f; however, its resistance/cell is substantially higher. Consequently, there is little difference from a power dissipation standpoint.

IV. ROTATING CURRENT

In this conductor access scheme, the propagate elements are perforations in a continuous sheet of conductor material deposited directly on the epi-garnet. Bubble propagation is accomplished by a pair of

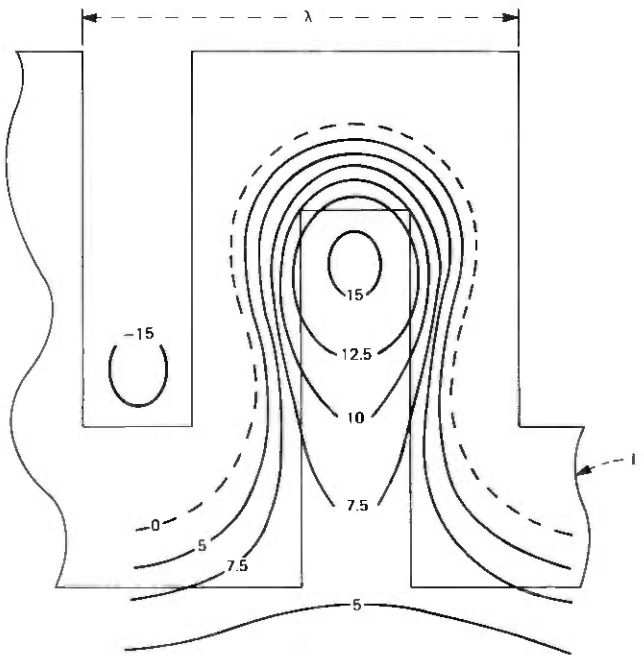


Fig. 10—Measured z-field contours for a meander-line conductor path. Note that the peak field extends toward the boundary of the conductor path.

bipolar currents applied in orthogonal directions to the conductor plane. This structure was first examined by T. J. Walsh and S. H. Charap.[6] Their computer simulations demonstrated the principle of operation. The obvious advantage over a field access device is the absence of field coils. This permits a simpler package and higher operating frequencies, which were previously limited by the field coils.

4.1 Principle of operation

A current is applied to a patterned conductor overlay. This current produces poles around the apertures. As the current is rotated in the conductor plane, the poles rotate about the apertures (see Fig. 11). In this figure, a chevron is the aperture (propagate element). As a current I passes from left to right, poles are established, positive on the bottom and negative on the top of the chevron. The bubble is attracted to the positive poles and will rest at the lower portion of the chevrons as in Fig. 11a. In the following three illustrations, the current is shown rotated in 90-degree steps. As the current rotates, the poles it produces also shift, giving rise to bubble propagation.

4.2 Experimental

Circuits were fabricated with 16-μm and 8-μm periods. Because of the processing advantages of single-gap structures, most test elements

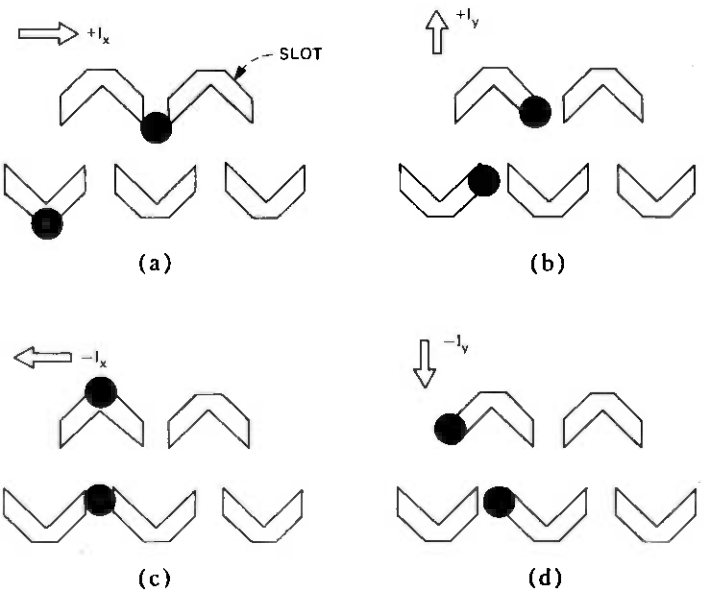


Fig. 11—Principle of operation for a circuit using rotating current for the drive. Motion, (a) through (d), proceeds much as for a single-gap permalloy circuit where chevrons are driven by a rotating field.

were chevrons. Multiple elements, such as the T-bar or X-bar, require tighter line-width control and tax processing capabilities. Portions of the masks used to make the circuits are shown in Fig. 12.

Circuits were patterned in a conductor film of 4500 Å AlCu. Bubbles were viewed in reflection with the conductor sheet serving as the reflecting surface. Four drive pulses are applied in quadrature for propagation, and a typical pulse train is shown in Fig. 13. Some overlap of adjacent current pulses is required. The amount of overlap was varied to ascertain the best operating margins with the lowest power dissipation. It was determined that 33-percent overlap was optimum. Note the slots in the arms of the test structures of Fig. 13. The slots function to keep the current flow lines parallel.

Both chevron and X-bar propagate elements were evaluated. The propagate margins for three chevron-like elements at an 8-μm period are given in Fig. 14. The margins are for various bubble patterns in straight-line propagation. Tests with different bubble patterns gave no indication of bubble-bubble interaction or reduction in propagate margin. Propagation through a 90-degree turn of Fig. 12c was marginal. The X-bar structure of Fig. 12d did not propagate domains' reliability.

4.3 Status

The high current densities required for propagation of 4 to 5 mA/μm as well as the complexities of providing multiple current drives into a common load curtailed efforts on this single-conductor approach. Consequently, problems associated with the design of generators, transfers, turns, and detectors have not been addressed.

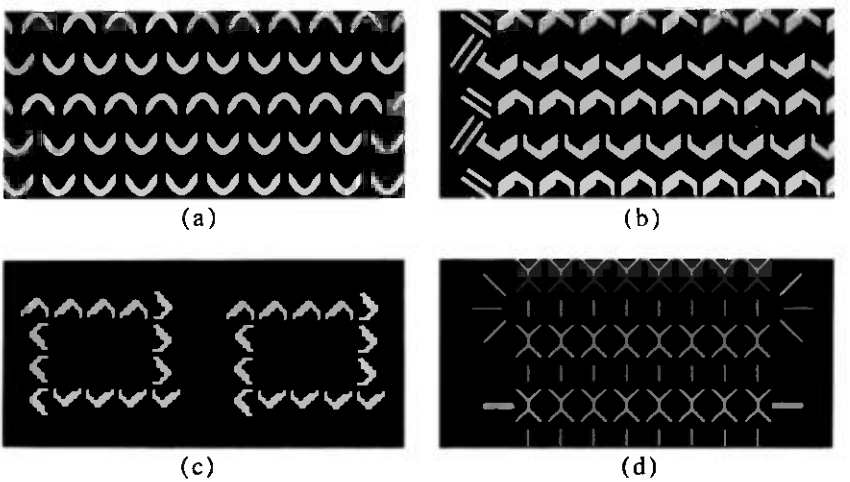


Fig. 12—Test masks used to evaluate rotating current structures.

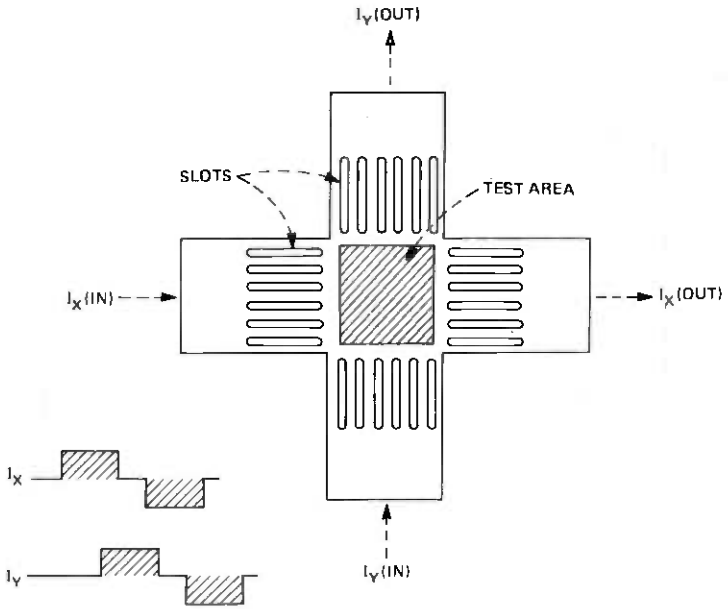


Fig. 13—Details of the overall test circuit. The slots in the lead-in arms prevent the current from "pin-cushioning."

V. CURRENT-ACCESS PROPAGATION BASED ON A STATIC OFFSET FORCE

In this section, we take an in-depth look at propagation circuits that can be driven by a single bipolar current source. The circuits of Copeland and Dekker discussed in Section II and those introduced in the following section, in which the meandering conductor is replaced by an apertured sheet, fall into this category.

5.1 Description

An excellent discussion of the static offset force problem has already been presented by Copeland et al.[10] We expand their results to include the effect of the domain wall coercivity H_C. In the usual single conductor circuit, the current flows in a conductor that crosses back and forth over the bubble track as shown in Fig. 15a. When a bipolar current is applied to the conductor, it produces an approximate standing wave field envelope:

$$H_I = N_I I \sin 2\pi x/\lambda, \tag{1}$$

where the X direction is along the bubble track and H_I is a Z-directed bubble drive field (Fig. 15b). It is obvious that a bubble can respond to

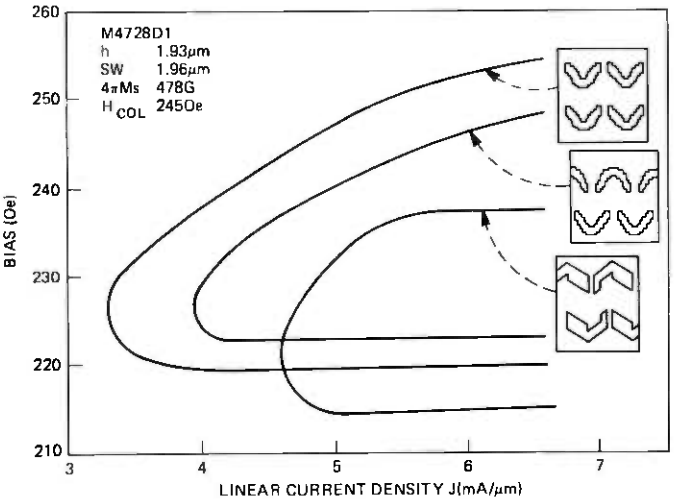


Fig. 14—Operating margins for the three aperture designs indicated.

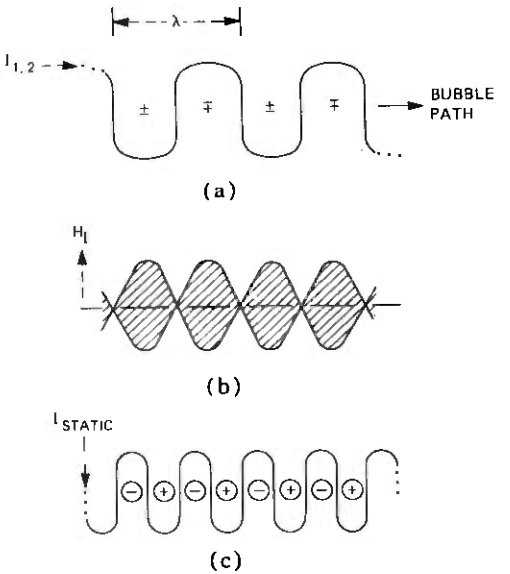


Fig. 15—Schematic layout of a single-conductor circuit. The conductor driven with a bipolar current (a) and the idealized z-field standing-wave pattern produced (b) are shown, as well as a double-frequency meander-conductor driven by a constant current (c).

this standing-wave field pattern by either oscillating back and forth or, more likely, being ejected from the track altogether. Next add the static current I_S (Fig. 15c) which produces yet another field pattern but this time with a double spatial frequency. This field is of the form

$$H_S = N_S I_S \cos(4\pi x/\lambda). \tag{2}$$

In eqs. (1) and (2), N_I and N_S are current-to-field conversion factors.

It is only illustrative, of course, that we use a static current I_S to produce H_S. We have already seen that grooves in a garnet or permalloy points can accomplish the same end and without an added power dissipation. If the dynamic (H_I) and the static (H_S) drive fields are properly proportioned, we achieve propagation as illustrated in Fig. 16. The ideal offset is $\lambda/8$ with an absolute maximum tolerance of $\pm\lambda/8$.

5.2 Conditions for bubble propagation

Before we discuss the constraints that must be imposed on H_I and H_S to obtain bubble propagation, it is helpful to review some bubble fundamentals. A field H applied to a stationary wall causes motion at a velocity v(cm/s) given by the equation

$$\mathrm{v} = \mu(H - H_c), \tag{3}$$

where μ(cm/sec-Oe) is the wall mobility. In the case of a bubble, the change in field intensity across the bubble ΔH is used with the result

$$\mathrm{V} = \frac{\mu}{2}[\Delta H - 8H_c/\pi]. \tag{4}$$

Since the bubble in the circuit (Fig. 15) experiences fields H_I and H_S simultaneously, we combine (1) and (2) to obtain

$$H_T = N_I I \sin 2\pi X/\lambda + N_S I_S \cos(4\pi X/\lambda). \tag{5}$$

There is no general analytical solution for the velocity of a bubble, even assuming H_I sinusoidal and $H_c = 0$. By graphical analysis we can,

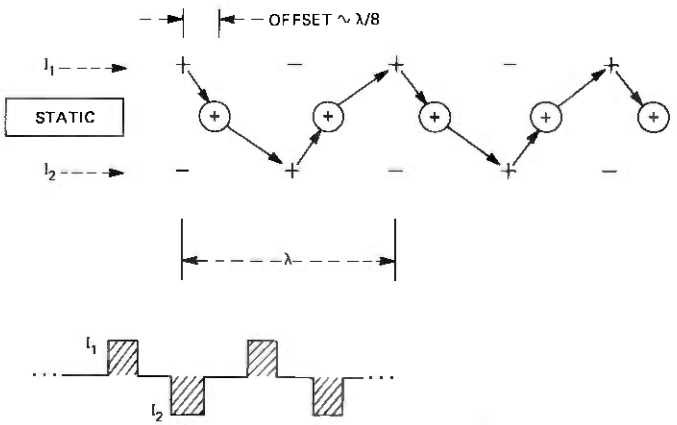


Fig. 16—The unidirectional propagation mechanism for the circuit of Fig. 15.

however, determine the minimum fields needed to overcome the coercivity H_c. From these we can determine the minimum drive currents. The mobility can then be introduced as an iteration of the coercivity under the assumption that the bubble moves at constant velocity.

The drive field H_I and the static field H_S and their sum are plotted in Fig. 17. It is assumed that positive fields increase the bias-causing bubbles to seek locations at which the applied fields are most negative. We treat the case where the bubble diameter is one-fourth the circuit period λ; however, the analysis can be extended to any diameter. Since mobility effects are being neglected, the condition $\Delta H \geqq 8H_c/\pi$ is sufficient to move a bubble. Actually, (4) is valid only for a linear gradient; however, we assume that any field difference at the extremes of the bubble diameter is equivalent to ΔH.

We start with a bubble in position 1 of Fig. 17, where 1 defines the location of the center of the bubble and, initially, with $H_c = 0$. Increasing H_I from zero to the amplitude sketched on the figure shifts the bubble from 1 to 2 and then, as H_I returns to zero, to 3. This completes a half-step of propagation. The other half-step takes place when H_I is reversed in sign. Next, we consider the conditions imposed on H_S when the coercivity H_c is included. There are "dead spots" at H_I (peak), and it is the role of the static field H_S to dislodge bubbles from those positions. The minimum values of H_S for any movement whatsoever is

$$2H_S = 8H_c/\pi. \tag{6}$$

Equation (6) therefore defines the critical minimum value for H_S. To

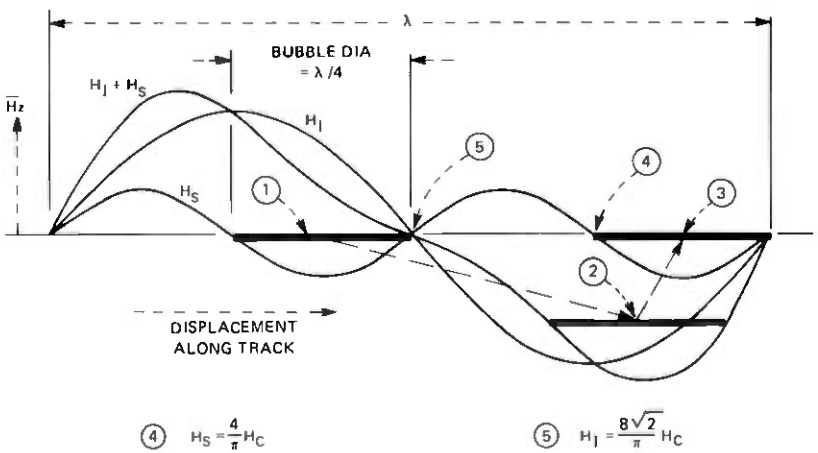


Fig. 17—Idealized z-fields H_I and H_S for a single-conductor propagation circuit with a static offset field. Bubble motion is from 1 to 3.

deduce the critical amplitude of H_I, we set $H_S = 4H_c/\pi$ and assume that there is a bubble at 1. As H_I is increased from zero amplitude, the bubble will move to the right, eventually reaching 5. Now 5 is a minimum gradient location for the bubble and, to overcome the coercivity, the amplitude of H_I must reach

$$H_I = 2\sqrt{2}H_S. \tag{7}$$

Once beyond 5, the bubble moves quickly ahead since it has excess gradient until it stops at 4, where again the field gradient is $8H_c/\pi$. When H_I returns to zero, the static field H_S completes this half-step by nudging the bubble a bit further to the right.

In summary, our two design equations are

$$H_S = 4H_c/\pi \tag{8}$$

and

$$H_I = 8\sqrt{2}H_c/\pi. \tag{9}$$

5.3 Methods to realize a static offset field

It is beyond the scope of this paper to list all the methods proposed to produce a static offset field. Permalloy dots, permalloy points, and embossed garnet features, i.e., grooves, are the three most extensively reported. In the next section, two more approaches are added to that list. They are (*i*) permalloy features contoured in the *z*-direction and polarized by the bias field and (*ii*) ion-implanted low-energy bubble rest positions produced by selectively implanting zones in the surface of the garnet. The remainder of this section, however, is devoted to "grooves," as they are representative of the other approaches.

In *Magnetic Bubbles* by O'Dell,[11] the equivalence of bubbles and current loops is discussed at length. This analogy serves as an excellent introduction to the subject of grooves. It can be shown that the stray field of a bubble and a loop current are identical if the bubble is replaced by a loop current $I_B = 2M_s h$, where the units are mA, gauss and μm, respectively. As an example, if $h = 1.5$ μm and $M_s = 40$ gauss, parameters typical for an 8-μm device, then $I_B = 120$ mA. Shown in Fig. 18a, the "loop current" representation aids in the comprehension of bubble behavior.

For a cylindrical hole in a magnetic layer, a similar equivalent current is $I_H = M_s h$, which, for the parameters above, is 60 mA. For partially etched holes, i.e., craters, the expression is $I_h = M_s \Delta h$ (Figs. 18b and 18c). Since an undulating groove is nothing more than a continuous chain of craters, the analogy can readily be extended to cover that configuration.

We have also included the design curves of Fig. 19, which give the

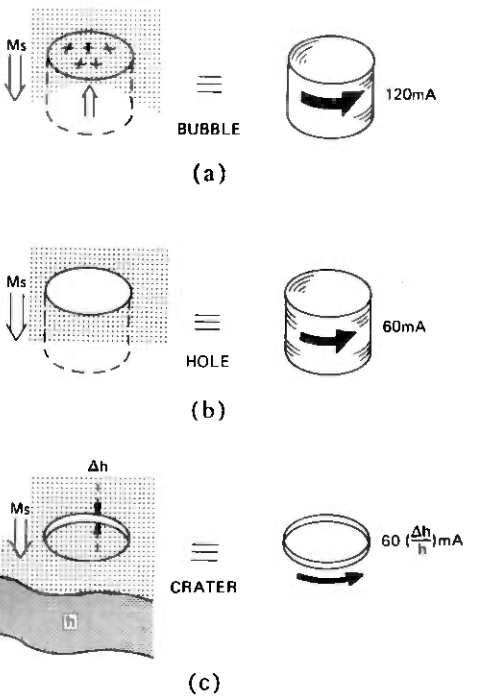


Fig. 18—The loop current magnetostatically equivalent to a typical bubble found in 8-μm period devices is given in (a). Similar relationships hold for a hole (b) and a partial hole (c).

field of a current loop located on the surface of a plate of thickness h. These curves will be useful if a more detailed analysis is attempted. They give the z-field averaged through the plate thickness h and normalized to the field at the center of the loop.

VI. KEYHOLE AND KNOTHOLE CIRCUITS

A class of bubble devices is now introduced, i.e., those driven by currents applied to apertured sheets rather than to arrays of conductors. In this section, only those configurations with a single conductor sheet are described. In Section VII, those circuits with two conductor sheets are covered.

6.1 Keyhole circuits

The idea for keyhole circuits originated from considerations involving the well-known "conductor crossing problem" in permalloy circuits. In particular, we were concerned with the role of the bias field. The bias field necessary to stabilize 6-μm and 1-μm diameter bubbles is about 180 Oe and 450 Oe, respectively. Consequently, the direct

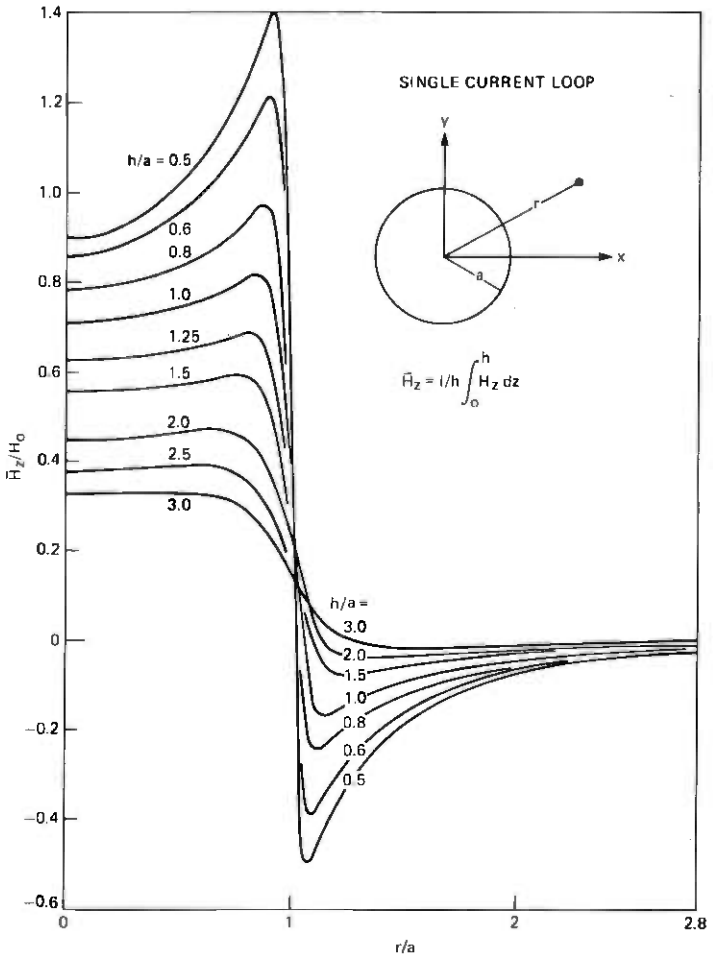


Fig. 19—Z-field contours normalized to the field H_0 at the center of a current loop. These curves are useful to characterize bubble-bubble as well as bubble-etched feature (embossed) interactions.

influence of the bias field on the magnetic state of the permalloy features cannot be neglected in circuits with small bubbles, i.e., high density circuits.

We use the bias field to produce a static offset field by the approach seen in Fig. 20. The field H_{bias} has a component $H_{\|}$ directed along the body of any permalloy feature inclined to the plane of the garnet. Such is the case, then, for a permalloy feature patterned at the edge of a tapered conductor. Magnetic poles that repel bubbles develop at the end of the permalloy feature nearer the garnet, while those that attract

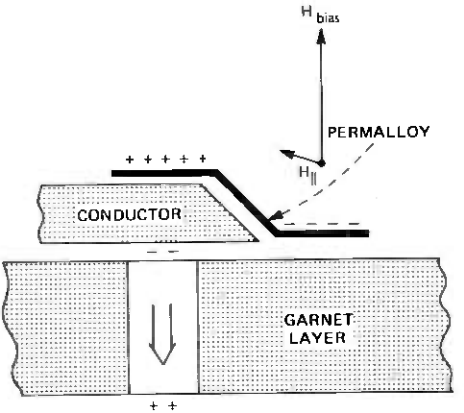


Fig. 20—A permalloy element, located on the edge of a tapered conductor, polarized by a component of the bias field.

bubbles develop at the end farther from the garnet. A reversal of the bias field direction will reverse the sign of the permalloy poles. However, the poles of the bubble also reverse, so the interaction remains unchanged. It is relatively easy to approximate the magnitude of the local field perturbations presented to the bubble. These calculations show that peak fields of a few oersteds to tens of oersteds can be expected.

Bubble propagation circuits capitalizing on this effect were designed; Fig. 21 is typical of those in which the apertures, squares fitted with a tab, are keyhole-like in design; hence, the name. Position and shape of the permalloy bar is seen in more detail in Fig. 22. For the reasons described above, bubbles are attracted to the ends of permalloy bars and repelled from their centers. Bubbles travel from left to right, moving from bar to bar in response to a bipolar current applied vertically to the sheet. They must "tunnel" through barriers as they travel from end-to-end on a given bar, but not as they jump from one bar to another. This asymmetry was apparent in the minimum drive currents of circuits that were tested.

For an 8-μm period circuit operated at 500-kHz, pulse amplitudes were required of 11.2 mA/μm to move along the permalloy and 7.2 mA/μm to bridge successive elements. Construction details were: 0.10-μm SiO_2 prespacer, 0.28-μm AlCu, 0.23-μm SiO_2 spacer, and 0.62-μm permalloy (ploy). With a garnet material that supported 1-μm bubbles, the operating bias range was 495 $\pm$ 17 Oe. A feature of this class of circuits is that they can accommodate a very wide range of bubble diameters. One foremost disadvantage is that they are difficult to process because of the intricacies of the aperture design.

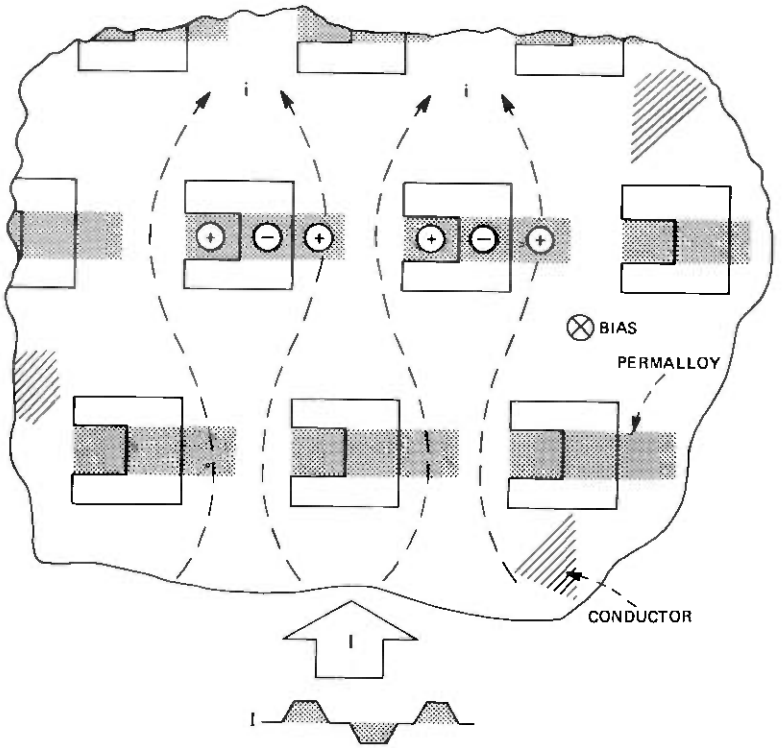


Fig. 21—Bubbles in this keyhole circuit move from left to right in response to a bidirectional current drive. The tab within the hole elevates the end of the permalloy bar but does not interfere with current flow around the aperture.

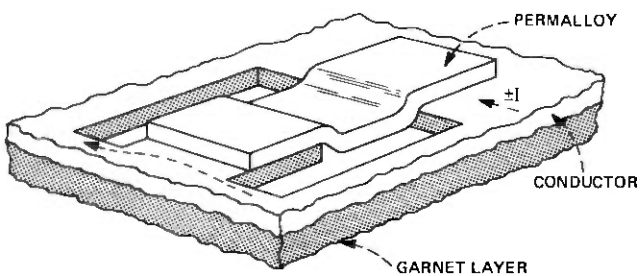


Fig. 22—Perspective view of a keyhole aperture shows the contoured permalloy.

6.2 Knothole circuits

In a knothole circuit, ion-implanted islands are combined with a single-apertured conductor sheet to form a bubble shift register. Uniform ion implantation of a garnet surface is the traditional method of

suppressing hard bubbles. Field-access bubble propagate circuits that use implanted patterns rather than permalloy elements are being considered for next-generation high-density applications. The uniaxial anisotropy of most garnet materials, and especially those treated in this paper, is lowered by any implantation that stresses the lattice. One can consider that bubbles couple to implanted areas because these areas tend to reduce the stray field energy by providing flux closure. The interaction, however, is more complex, as it can involve "charge walls" in the implant layer. Also, bubbles are especially attracted to the implant side of boundaries that partition implanted and nonimplanted areas. Phenomologically, an implanted island behaves like that of an area that has been thinned and as such can be replaced by a loop current. It follows then that the knothole circuit is a physical representation of the hypothetical structure analyzed in Section V.

A straight-line knothole shift register requires two implant islands and a single, simple aperture for each circuit period. Note that a tab in the aperture is no longer necessary and has been removed. Hence, this circuit has been given the descriptive name knothole. The general layout and operation of three parallel shift registers is illustrated in Fig. 23. Bubbles in the center row propagate to the right, those in the

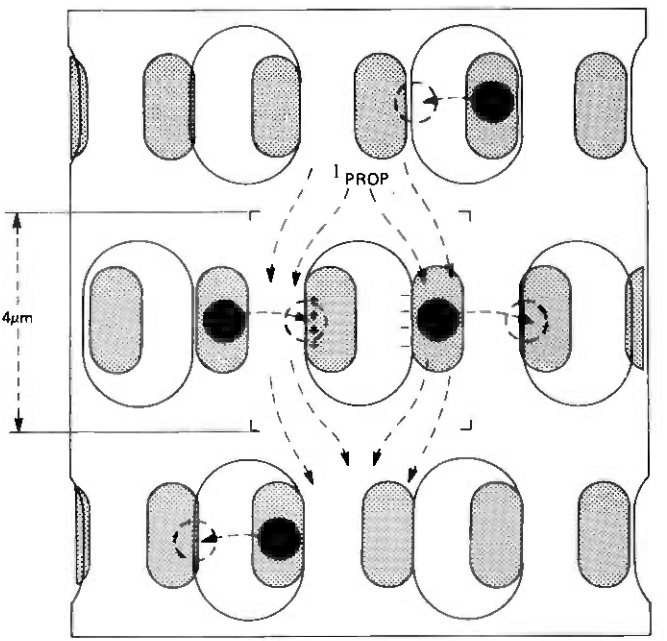


Fig. 23—Bubbles move from rest positions at the ion-implanted islands in response to current flow downward. Each circuit period requires two implanted islands.

outside rows propagate to the left. A bipolar current applied to the sheet drives the bubbles.

The performance of knothole circuits is sensitive to bubble size. To illustrate this point, refer to Fig. 24 where the data are presented for a 4-μm period test circuit using a nominal 1.7-μm bubble material. Islands were given an implant of 2.5×10^{13} Ne at 100 Kev, and no overall implant was included. A static in-plane field of 210 Oe was applied parallel to the current. Bias margin, frequency range, and drive power/bit (based on 0.1 ohms/cell) are outstanding. The direction of propagation is *opposite*, however, to the design direction. The oversize bubbles straddle pairs of implanted areas during operation.

Design of an 8-μm period closed-loop shift register is given in Fig. 25 and data in Fig. 26. Here the nominal bubble diameter is 1.6 μm, which is a good match to the nominal 2 μm by 4 μm implant islands; hence, the direction of propagation is correct in the classical sense, i.e., clockwise around the loop. The islands were implanted at 3×10^{13} with Ne at 50 Kev. Driven at J = 3 mA/μm, the continuous power dissipation is 58 μw/b.

A variety of π-turns were tested during the course of this development, and the most successful of these is incorporated in the circuit of

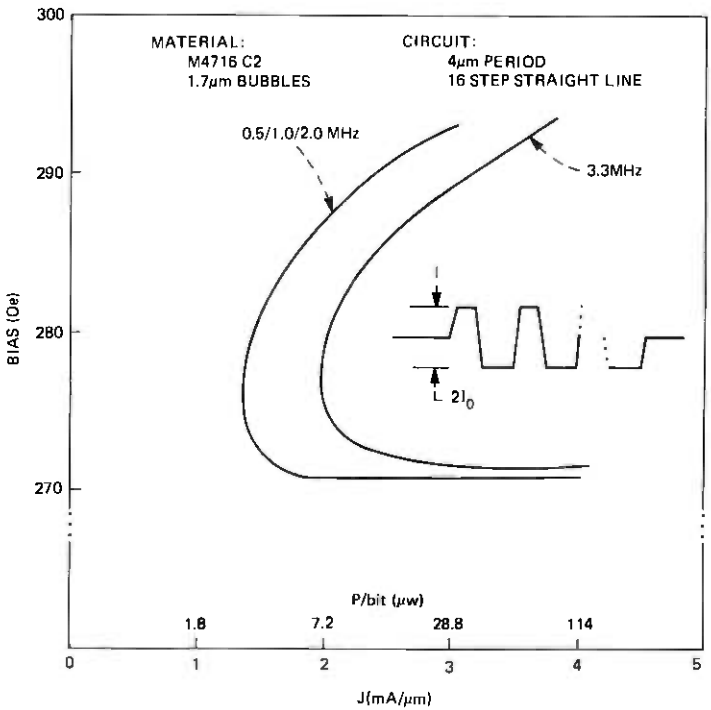


Fig. 24—Operating margins vs frequency for a 4-μm period knothole circuit.

Fig. 25. The shape of the aperture in the turn was optimized from field plots on scale models and iterated until the desired drive gradients were seen. Operating margins of this turn improve when an in-plane field is applied parallel to the direction of the drive current. For the data of Fig. 26, the in-plane field was 100 Oe.

Straight-line propagation parallel to the drive current was achieved with the circuit of Fig. 27. Operating margins are similar to those obtained for propagation normal to the current, and much of the same design criteria apply here. Although the drive currents per cell are lower, the resistance/cell is substantially higher; thus, the power/cell is essentially unchanged.

A detailed description of support functions such as generation and detection is deferred until Section VII, which treats dual conductor circuits.

VII. DUAL CONDUCTOR OPERATION

The knothole configuration with its single perforated sheet and "offset-force" zones does not have the design flexibility that can be

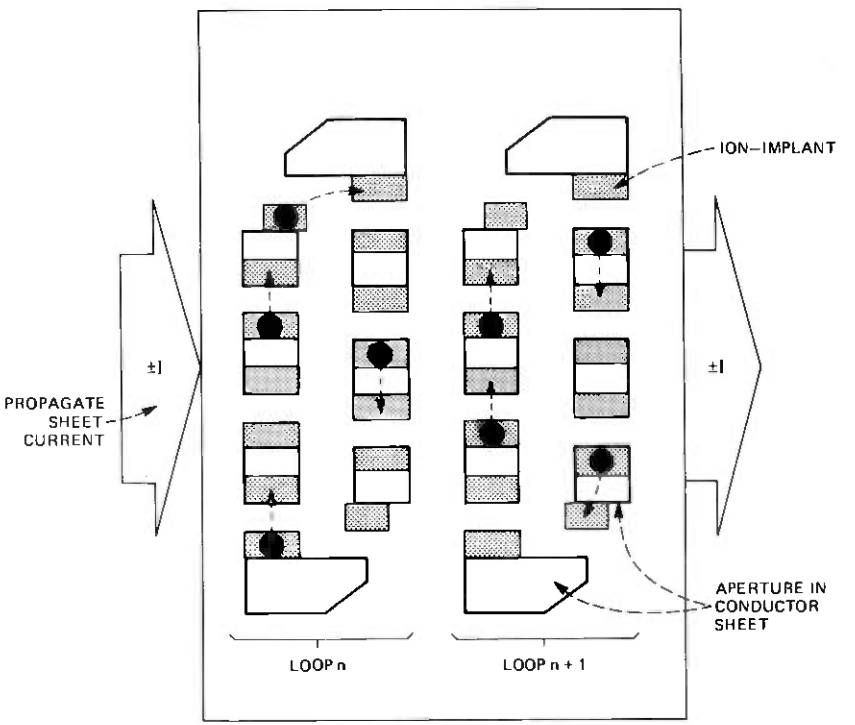


Fig. 25—Details of π-turns used in 7-step shift register loops. Clockwise turns can be mirrored to make counterclockwise turns.

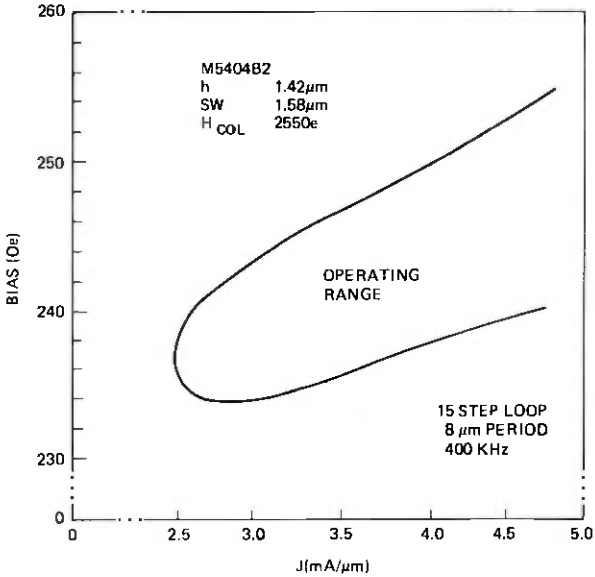


Fig. 26—Data for a 15-step knothole loop designed as in Fig. 25. Bubbles were propagated in 14-step bursts.

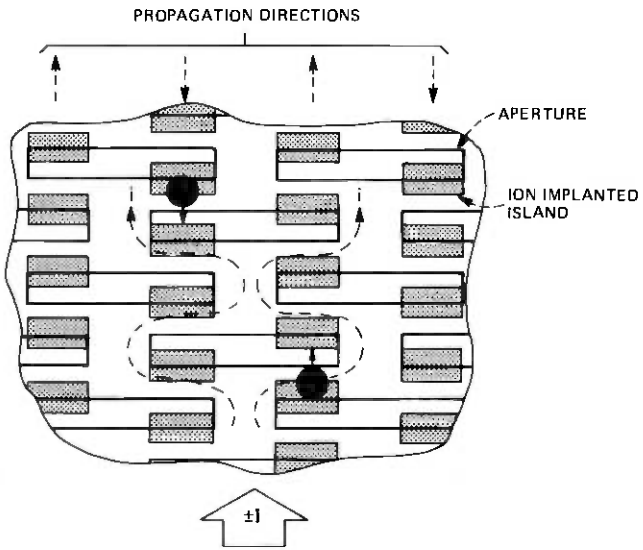


Fig. 27—Propagation parallel to the direction of general current flow.

achieved with a pair of conductor sheets. For example, with the pair of conductor sheets it is possible to transfer a bubble, i.e., switch a bubble from one path to another, or reverse the direction of propagation, simply by altering the propagate pulse sequence. This section deals

with straight-line propagation parallel and perpendicular to the direction of current flow, closed loop design, nucleate generation, detection, and transfer. Also included is a discussion of factors that influence chip design. The discussion is limited to the case where the currents in the two levels flow either parallel or anti-parallel to one another. It will be apparent to the reader, however, that, if the structures are reconfigured so that the currents flow perpendicular to one another, a new family of device structures somewhere between the rotating current structures of Section IV and those of this section could be realized.

Dual conductor circuits are fabricated with a first patterned conductor level either directly on the epitaxial garnet or onto a pre-spacer layer, an insulating layer, and finally a second patterned conductor level. Details of the actual fabrication are covered in Section IX. It will be useful at this time to describe the general layout of the test structure depicted in Fig. 28. All the test data described in the following sections were taken on circuits located within the cross-hatched region which is 320 μm wide and 600 μm long. The paths that lead into this region are shaped to make the current density within the region as uniform as possible.

In line drawings equivalent to processed circuits, for a view looking down onto the second conductor layer, the following conventions will be established and used throughout (see Fig. 29). All apertures in the first conductor level (that nearest the garnet) are outlined by a solid line, whereas apertures in the second level are indicated by a dot pattern. Sheet currents $I_{1,3}$ and $I_{2,4}$ flow in the first and second

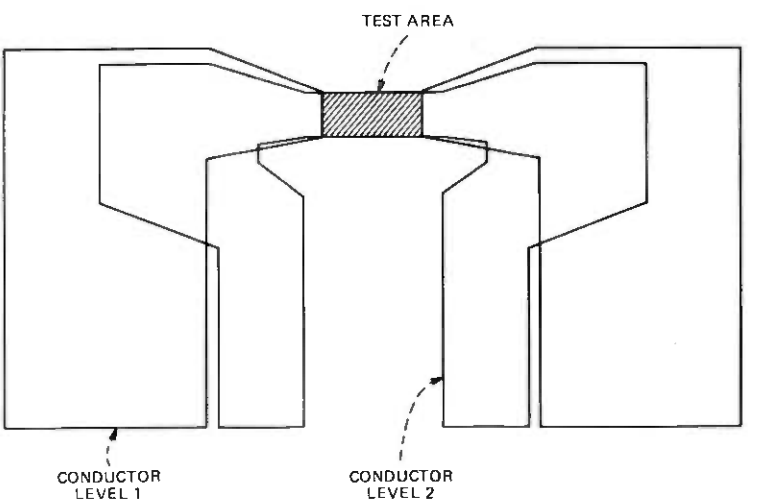


Fig. 28—Layout of the circuit used to test dual-conductor designs. Two or more probe fingers contact each pad area located along the bottom edge. Uniform current density is maintained at the test area.

conductors, respectively, where they generate local bias-field perturbations at positions 1, 3 and 2, 4. For a bubble-stabilizing field H_{bias} directed downward as shown, bubbles are attracted to location 1 for current I_1, location 2 for I_2, etc.

7.1 Propagation normal to the current flow

In Fig. 29, it can be seen that a bubble initially at position 1 propagates upwards (normal to the direction of current flow) when the current pulse sequence [234] is applied. Sequence [321] returns the bubble to its starting position. Straight-line propagation over a further distance is achieved by simply linking like pairs of apertures as in Fig. 30. The length of each aperture is 0.5λ, and the distance between successive like positions is the circuit period, λ. It should also be apparent that the maximum tolerable misalignment of apertures in the vertical direction is 0.25λ.

Typical current-pulse waveshapes to produce the sequence

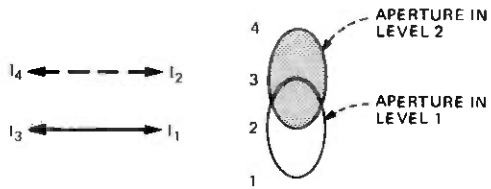


Fig. 29—A four-position, dual-conductor, bubble-stepping circuit. Bubbles are attracted to positions 1 through 4 when currents I_1 through I_4 are applied.

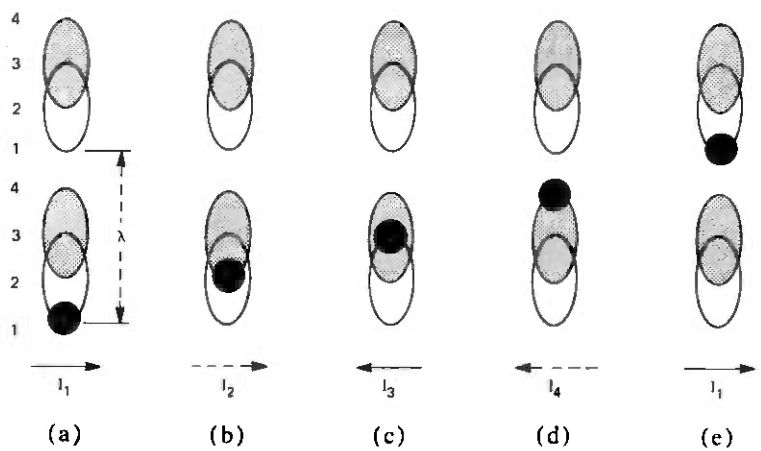


Fig. 30—A time sequence of the operation of a dual-conductor shift register. The direction of propagation will be reversed if the currents are applied in the sequence (e) through (a).

[123412341] for stepping a bubble through two periods are sketched in Fig. 31. Some overlap between successive pulses is helpful to further quantize the traveling-wave bias field minimum that drives the bubble. Also, to get a bubble fully advanced into position for, say, another two-step burst, then a final [41] overlap should be provided.

Success in aperture design is measured in terms of generating the maximum possible field gradient for the lowest power dissipation. Almost any aperture will perturb an otherwise uniform current density and give rise to bubble motion. Some of those tried were shown in Fig. 6, along with field plots obtained on scaled-up models (details in Section X). This field data information coupled with both processing and operational experience points to the rectangular shape 0.3λ by 0.5λ as the most practical choice.

The operating margins for a series of 8-μm period dual-conductor test circuits are presented in Figs. 32 to 38. Circuit data cover both unimplanted and implanted garnets as well as the influence of in-plane fields parallel to the drive currents. A frequency and temperature run are also included. Individual data points are shown wherever they aid in understanding the experiment. Bubbles used in these tests were obtained by subjecting a chip to a large in-plane field. Testing continued until the bubble supply was depleted, and a new supply was initialized.

We report, in Fig. 32, 1-mHz data on an implanted garnet test circuit driven with a two-step to and from sequence [1234123432143214]. Of special interest is the spread in data points, especially at low drive. We suspect that this is a result of the bubble states found in implanted garnets[12-14] and to a greater extent in unimplanted garnets. The domain structure of the implanted layer is also suspected.

The low bias margin is limited by stripout, as is usual for most of the dual-conductor circuits tested. This is especially significant when detection is considered. Free-bubble collapse limits the absolute upper bias, since "hold currents" are not maintained between pulse bursts. If an in-plane field is applied,[15] much of the uncertainty in the data points is eliminated, as is evident in Fig. 33.

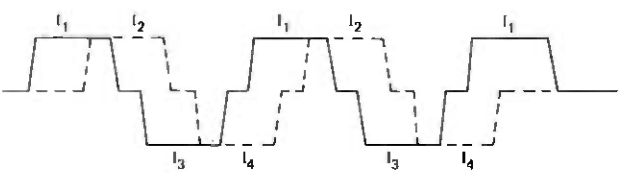


Fig. 31—The current pulse sequence used to step bubbles in a dual-conductor circuit. The overlap of pulses is used to further quantize the propagation field.

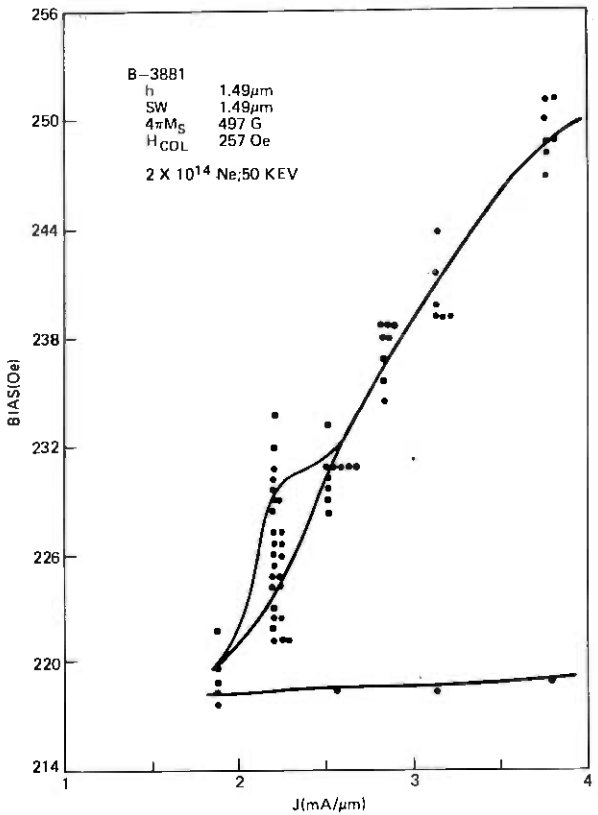


Fig. 32—Bias margin vs drive for an implanted garnet tested at a 1-MHz stepping rate in an 8-μm period dual-conductor circuit. Each data point represents an individual check of the upper bias margin. The spread in the test data indicates that the bubble state and/or the implant layer magnetization does not reproduce from test to test.

When an unimplanted garnet is tested, the upper bias margin often shows two distinct distributions (Fig. 34). Failure modes associated with these curves indicate that two distinct bubble states, probably the $S = 0$ and $S = 1$, are involved. With a sufficiently intense in-plane field, the double distribution disappears. Refer to Fig. 35, where margins for both a 200-Oe and 300-Oe in-plane field are given.

These experiments suggest that an in-plane field is helpful to stabilize bubble propagation in dual-conductor circuits. We do expect the shape of the field gradients to continually redirect bubbles toward the desired path. However, it is known that $S \neq 0$ bubbles do not move along the maximum field gradient, but rather at angles (often near right angles) to the gradient. In Table I, the skew angle calculated for $S = 1$ bubbles are tabulated for the garnet compositions reported in this paper. Further insight into the problem comes from an analysis of

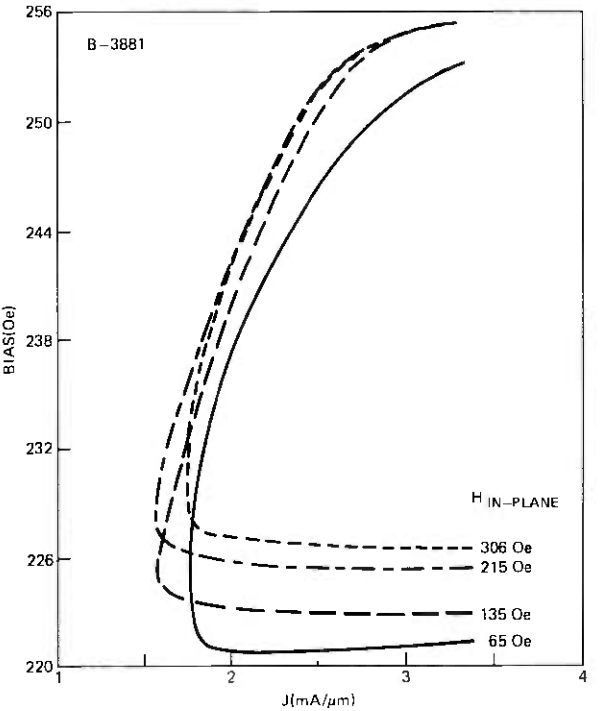


Fig. 33—Margin data for the conditions of Fig. 32, except that an in-plane field from 65 to 306 Oe was applied parallel to the current flow. The spread in data points was virtually eliminated.

Table I—Comparison of material parameters

Material Type	q (ratio)	$4\pi M_s$ (gauss)	ΔH_c (Oe)	μ (cmS/s-Oe)	γ (l/Oe-s)	ψ^* (degrees)	$f_{OPT}^\dagger$ (MHz)
YSmLuCaGe garnet	3.8	524	2.0	300	1.7×10^7	22.5	0.4
LaLuSmGa garnet	4.0	478	1.5	500–1000	1.7×10^7	35–45	0.5–1.0

* Tan $\psi = 4\mu S/\gamma d$ $(S = 1)$.
† $f_{OPT} = 4\mu H_c/\pi\lambda$ $(\lambda = 8\ \mu m)$.

the trajectory of an $S \neq 0$ bubble moving in the dipole field generated at a circular hole. Details are presented in Section X. The choice of whether to cope with bubble states by one of the conventional hard bubble suppression methods or by an in-plane field will be dictated by the method chosen to detect bubbles. It may prove advantageous to shape the apertures, particularly in turns, to take advantage of preferred propagation angles.

In Fig. 36, the results are presented of a frequency run on an unimplanted garnet in a 200-Oe in-plane field. These results are

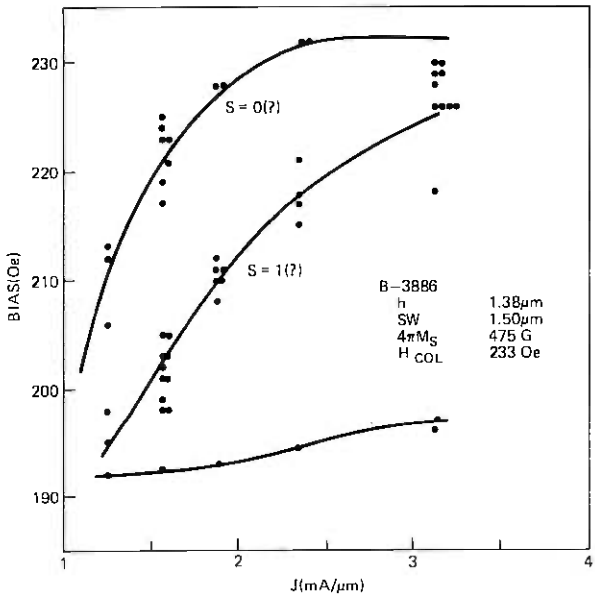


Fig. 34—Bias margins vs drive for an unimplanted garnet tested at 1 MHz in an 8-μm period dual-conductor circuit. Two distinct distributions in the upper bias margin were observed.

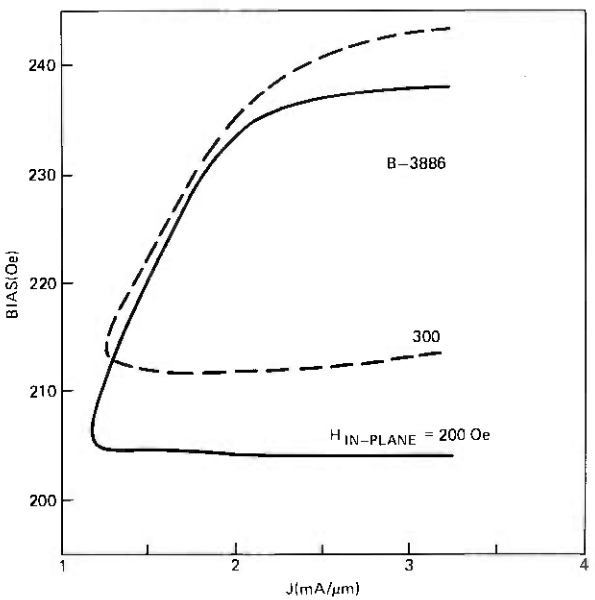


Fig. 35—The data of Fig. 34 repeated except with an in-plane field applied parallel to the current. With unimplanted garnets, substantially higher in-plane fields are needed to tighten the data points.

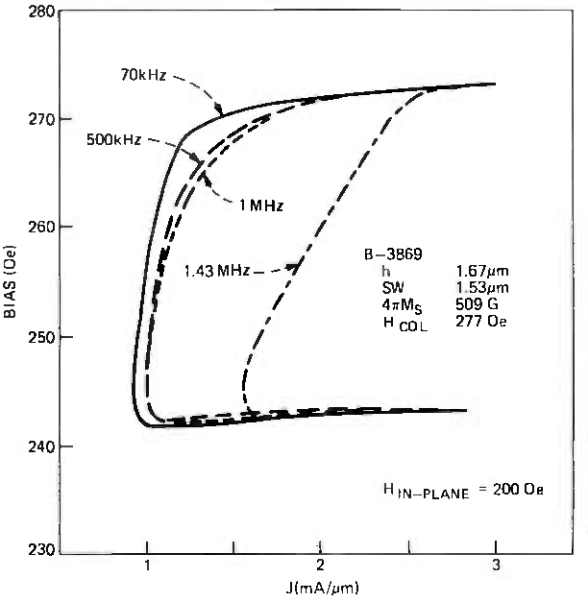


Fig. 36—Operating bias margin vs frequency for an unimplanted garnet film with a 200-Oe in-plane field.

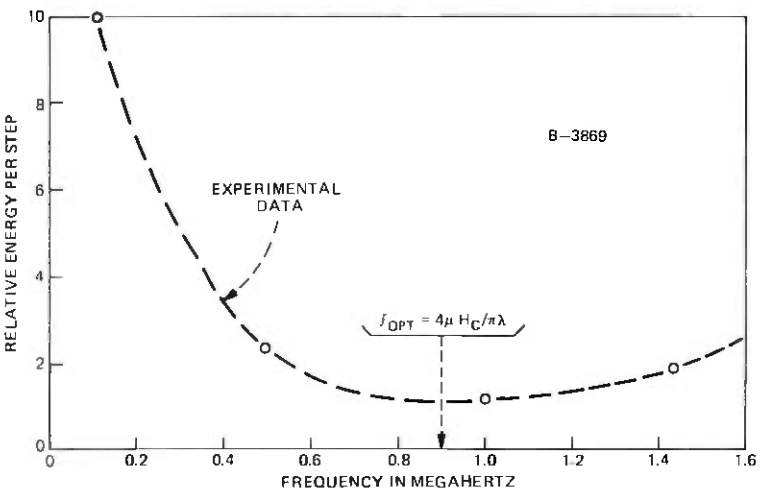


Fig. 37—It is wasteful of power to run current-access circuits except at a frequency f_{OPT} at which the energy dissipated per step is a minimum. Calculated and experimental values of f_{OPT} are compared.

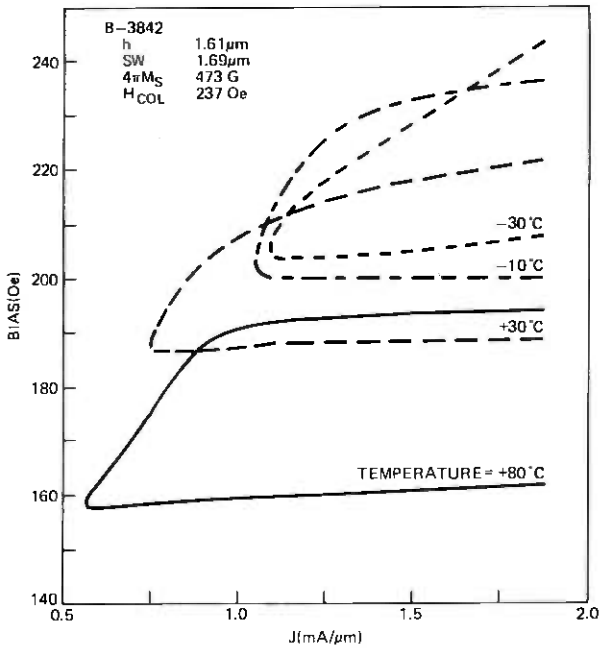


Fig. 38—Operating margins vs temperature taken at 30 kHz with a four-step propagation sequence. The garnet was unimplanted.

consistent with predictions based on calculated drive gradients and measured garnet properties. We can use these results to introduce f_{OPT}, the optimum frequency at which to step a bubble from one point to another to minimize energy dissipation. It can be shown that $f_{OPT} = 4\mu H_c/\pi\lambda$. The relative energy per step (proportional to J^2/f) vs frequency calculated from the data of Fig. 36 is plotted in Fig. 37. The calculated f_{OPT} for wafer B-3869 (H_c = 0.5 Oe, μ = 1,000 cm/sec-Oe) compares favorably to the frequency corresponding to the energy minimum of Fig. 37.

Operating margins at 30 kHz for an unimplanted garnet over the temperature range −30°C to +80°C are given in Fig. 38. These data were furnished by J. L. Smith. The minimum drive increases at low temperature much as expected due to the increase in coercivity. Note especially the very wide margins at 80°C.

7.2 Propagation around a closed loop

Details of a nine-step, bi-directional, closed loop are illustrated in Fig. 39. Note that the slots in the turn are 0.75λ in length and that elements that lead into the turn are repositioned somewhat. Bubbles propagate clockwise for the drive sequence [···12341234···] and

counterclockwise for [···43214321···]. Bias field-drive characteristics for a 6-μm period loop operated at 1 mHz are given in Fig. 40. The margins are quite adequate, as is generally the case for loops at the center of the test area. If z-field cancellation is not used, bubbles in

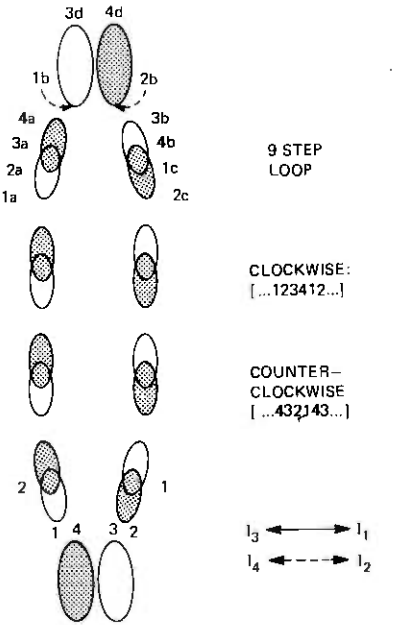


Fig. 39—A 9-step dual-conductor closed loop. The elliptical apertures in the turns have been lengthened for the reasons described in the text.

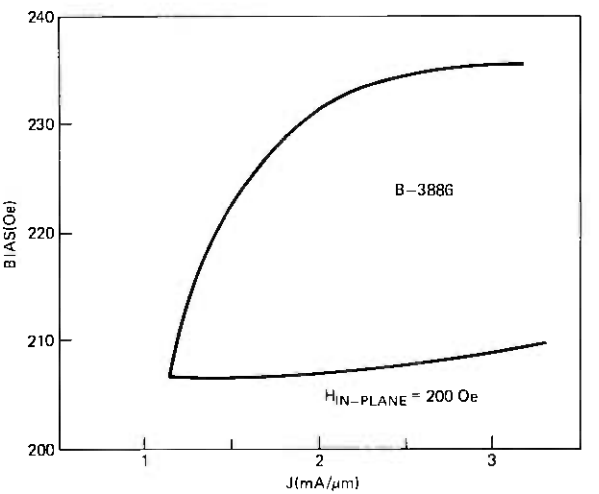


Fig. 40—Bias margin vs drive for a 6-μm period, 7-step loop operated at 1 MHz. The garnet was unimplanted.

loops located near the outer edges of the test area are biased by both an alternating and a constant field. A typical failure mode (Fig. 39) finds bubbles in clockwise motion, jumping to 3d from 2b rather than stepping to 3b. Some modifications were necessary to cope with this problem.

First, the two slots in the turn were lengthened further to 1.25λ and, second, additional apertures were placed beyond the turn itself to help equalize the current distribution. The latter is readily accomplished by interconnecting loops end to end (Fig. 41). Note that the bubbles in adjacent loops move in paths counter to each other just as the gears of a gear train.

Impetus for making the modifications given above came from test circuits in which interconnecting loops spanned the width of the test area. Loops with lengthened turn-elements outperformed all the others in positions 1 and 5, whereas all loops were satisfactory in positions 2, 3, and 4.

These experiments also gave additional evidence of the "gyro" nature of the bubbles in our materials. Consistently, the even-numbered loops outperformed the odd-numbered loops for one direction of

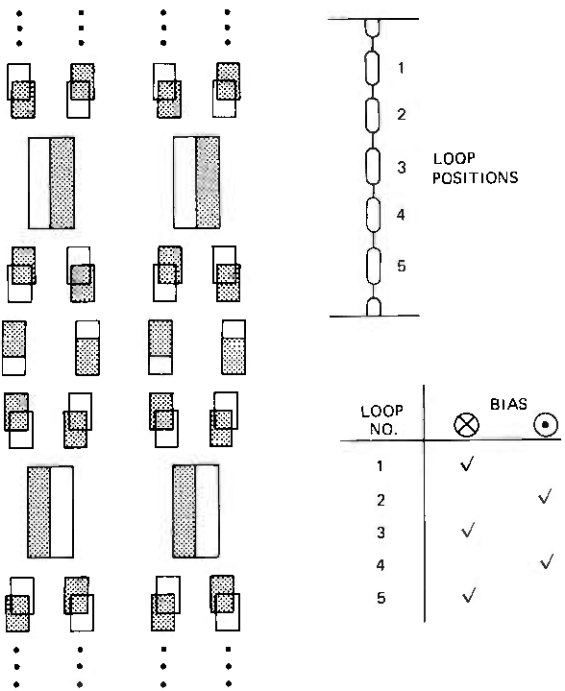


Fig. 41—Loops can be interconnected end-to-end by the method shown. The propagation direction of adjacent loops alternates, i.e., cw, ccw. Performance is related to the propagation sense and bias field direction.

bias and vice versa. One additional observation not as yet understood was made: Operation was generally better if the bias field was directed toward rather than away from the garnet surface.

7.3 Propagation parallel to current flow

In this section, propagation parallel to the direction of general current flow is considered. This is not an absolutely essential operation for a chip layout, since such motion can also be made by combinations of motions normal to current flow and the π-turns discussed in the preceding section. In fact, our earliest shift registers were designed that way. The principle of operation, from Fig. 42a, is straightforward. A bubble starting at the leftmost position 1a proceeds through 1a2a3a4a1234 in response to pulse sequence [123412341].

The actual operation is far from ideal, since propagation down the center track is stable only for a restricted (and oftentimes non-existent)

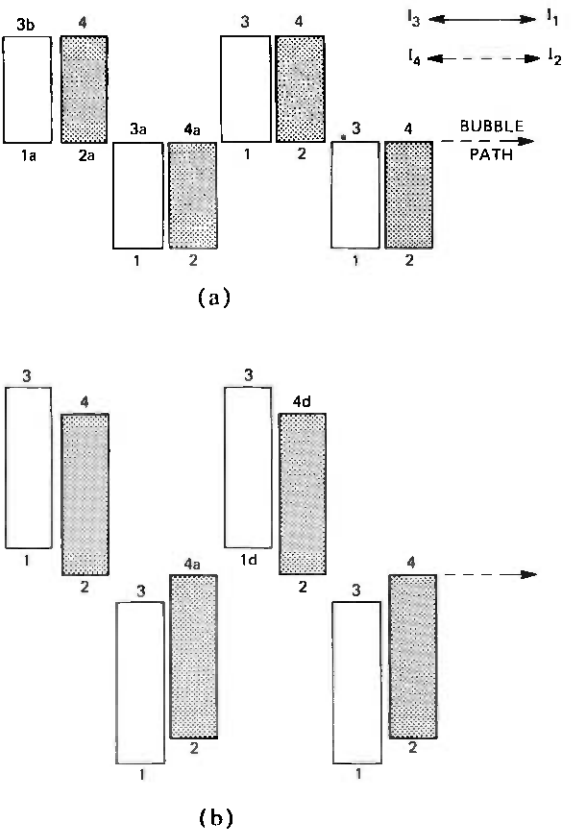


Fig. 42—Circuit for propagation parallel to the current. The original bidirectional design (a) was modified (b) to be unidirectional to improve performance.

set of drive and bias conditions. It is observed that a bubble in transit through 1a2a jumps to 3b rather than proceeding to 3a as desired. This occurs since the current density at 3b, and thus its pole strength, is substantially greater than at 3a, thereby counteracting 3b's increased distance from 2a.

Operation improves substantially if the apertures are increased in length and staggered as shown in Fig. 42b. This modification further removes the offending poles from the intended bubble track and also forces increased current along the bubble track. As a result, the circuit of Fig. 42b gives adequate margins for motion [· · · 12341234 · · ·], however, motion along [· · · 43214321 · · ·] is unreliable because of encounters such as 1d to 4d rather than 4a.

Operating margins for an 8-μm period shift register stepped at a 1-mHz rate are given in Fig. 43. The garnet was not implanted nor was an in-plane field applied. The performance is seen to be quite good.

7.4 Bubble generation

A nucleate bubble generator can either be patterned in one of the propagation conductor levels or it can be provided in a separate conductor level. Only the former design is considered here, since the design of the latter is obvious. Now it is possible to introduce slots in a sheet conductor if the slots are narrow (1/4λ or less) and run essentially parallel to the current flow. Such slots can be used to concentrate current and thus nucleate bubbles, yet permit current flow during normal propagation. Just how this can be done is seen in Fig. 44.

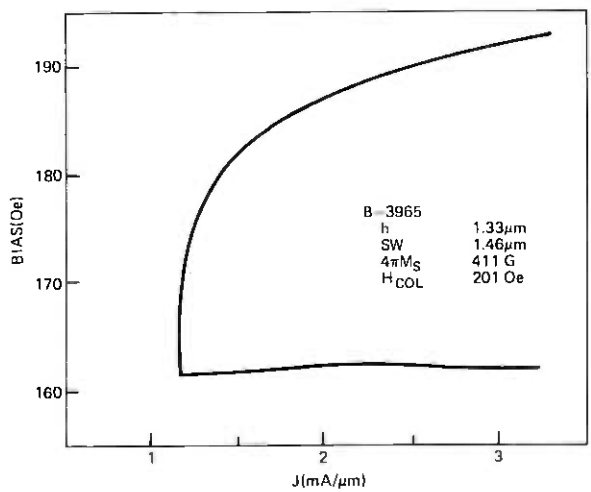


Fig. 43—Bias range vs drive at 8-μm period and 1 MHz for circuit (b) of Fig. 42. The garnet was not implanted.

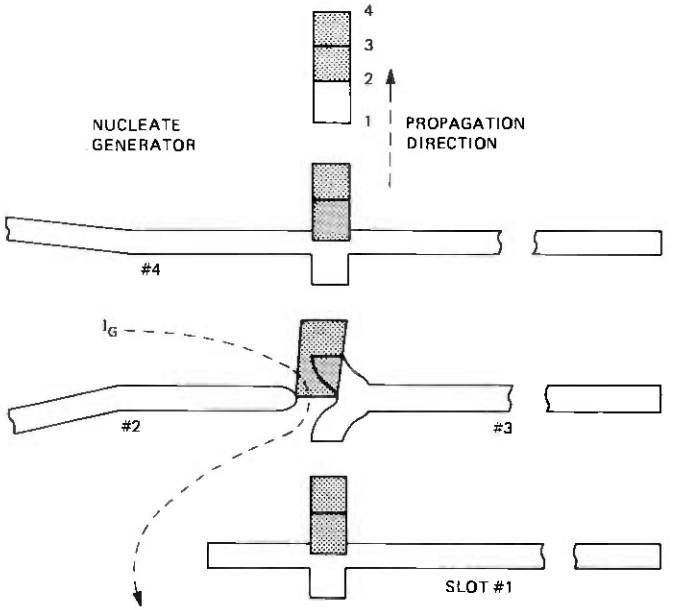


Fig. 44—A nucleate generator can be patterned as part of a conducting level. Slots guide the generator current but do not interfere with normal propagation. Bubbles are nucleated at the tip of slot #2.

The generator is incorporated into level 1 since that ensures the lowest possible generate current. Slots #2 and #4 funnel the generator current I_g from the input pad into the generator proper. A continuation of slot #4 acting in conjunction with slots #1 and #3 forms relatively high resistance paths so that most of the incident current I_g concentrates at the tip of slot #2, nucleating a bubble, before diffusing as it enters the chip area below.

Generator operation has been characterized in a nominally 8-μm period shift register processed on wafer B3954. The drive conditions were: amplitude 1.5 mA/μm, width 1.5 μs with 33 percent overlap. The generate pulse was applied in conjunction with propagate pulse 4. Subsequent propagate pulses move the newly generated bubble along the shift register track. Both adjacent and isolated domains were generated. The 200-ns wide generate pulse had an amplitude range of 230 to 360 mA.

The composite bias margins for generation at and propagation away from the generator were 19 Oe, with the lower end of the margin at stripout. For propagation through the generator, the margins reduce to 15 Oe with most of the margin loss at the low end. To ensure proper motion of bubbles as they pass upwards through the generator structure, a current density $J_{1,3}$ must be maintained between slots #3 and

#4. In our experiments, a single resistor diverted a fraction of the drive current $I_{1,3}$ into the generator pad.

7.5 Detection

Field-access permalloy devices use a chevron expander detector in which a bubble domain is stretched laterally before detection takes place in a magnetoresistance sensor. This approach has the advantage that it permits unlimited expansion without any effect on the data rate of the chip. The exit portion of the detector can be designed either to shrink the strip domain back to a bubble or to discharge the strip domain into a guard rail. The detector we report is similar to that just described and to a detector used with the bubble lattice.[8]

Bubble expansion and contraction is accomplished in the structure of Fig. 45. The analogy to a chevron expander structure is apparent. Factors relating J and $\bar{H}_z$ for an array of parallel conductor strips were introduced in Section III and are detailed further in Section X. In typical geometries, a current density $J = 4$ mA/μm generates a bias field decrement of nearly 20 Oe. In other words, stripout in the detector will be sustained over a 20-Oe range of the bias field. Since all the dual-conductor functions operate down to the strip-to-bubble transition field, we can expect to realize most of that bias field range.

Design curves to estimate the in-plane field H_x due to (*i*) the strip domain and (*ii*) the drive conductor strips are given in Figs. 46 and 47, respectively. At 8-μm period, assuming $4\pi M_s = 500G$ and $J = 4$mA/μm, we expect $H_x = 100$ Oe for a typical strip domain and a maximum field of 20 Oe (depends on detector-to-conductor strip positioning) from the drive conductors. The magnitude of these fields suggests that a thin-permalloy magnetoresistive sensor can be used. A photograph

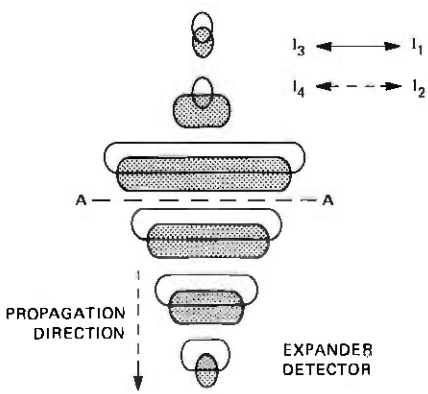


Fig. 45—Details of a dual-level expander detector. Bubbles, entering from the top, are expanded for either magnetoresistive or inductive sensing at A- -A.

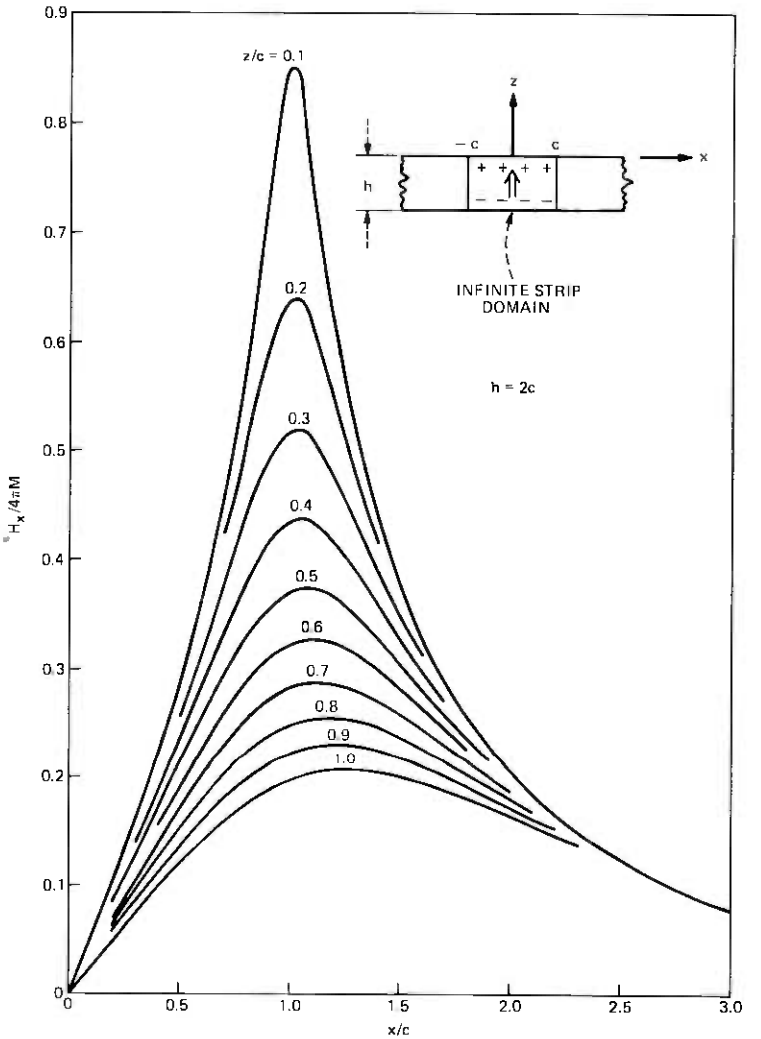


Fig. 46—Calculated in-plane field from a strip domain effective on a magnetoresistive sensor.

of the conductor-expander section complete with a "Chinese character" sensor is seen in Fig. 48. The sensor is processed on top of the pair of slotted conductor sheets.

In current-access devices, there is the opportunity to design a chip without permalloy. For example, if a strip domain is driven by an RF component of the bias field, the variation in the strip's flux can be sensed by a conductor pickup loop. This scheme is being pursued by J. M. Geary, and he has allowed us to present some of his preliminary

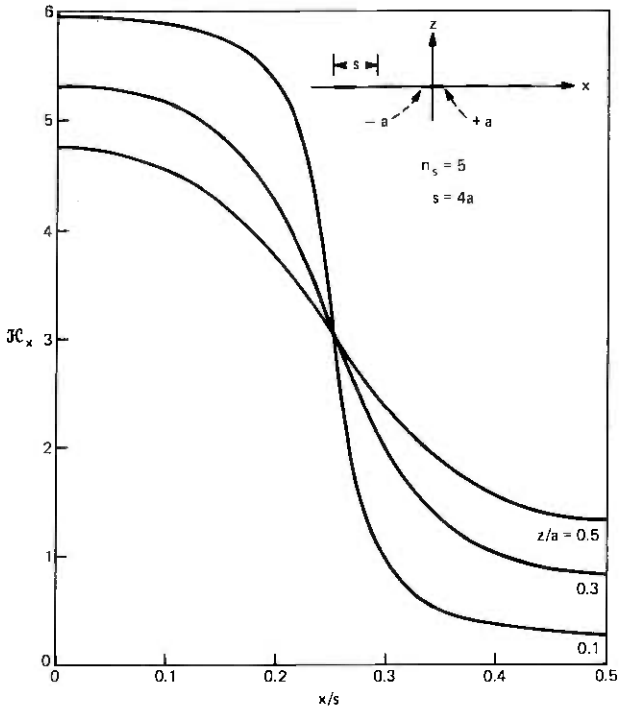


Fig. 47—Calculated in-plane field from the parallel drive strips of an expander detector.

results. A single-turn pickup loop 20 μm in length was patterned in a conducting sublayer at "A···A" of Fig. 45. An output signal of 5 μV was obtained when the width of a strip domain was modulated at 8 MHz (Fig. 49). RF detection in magnetic devices is not new and, in particular, we refer to an excellent article by Benrud et al., which deals with RF detection in thin permalloy-film memories.[16]

7.6 Transfer

Transfer is a particularly useful function since, combined with propagation, generation, and detection, it completes one set of functions that can be used to design a major-minor chip. This is significant, as it is well known that the major-minor organizations give improved performance over single loop shift registers. More important is that a dual-conductor major-minor chip can be partitioned into individually accessed regions with a resultant reduction in chip power dissipation.

One obvious way to implement a transfer is to use a "transfer" conductor patterned as part of a third conductor level to gate bubbles from minor storage loops to a major track and back again in much the

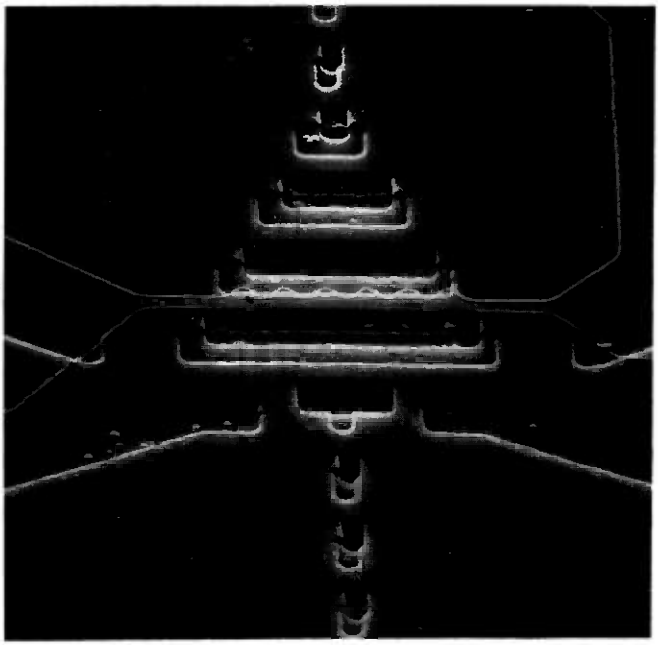


Fig. 48—An SEM photograph of an expander detector with a thin permalloy magnetoresistive sensor. The circuit period is 8 μm.

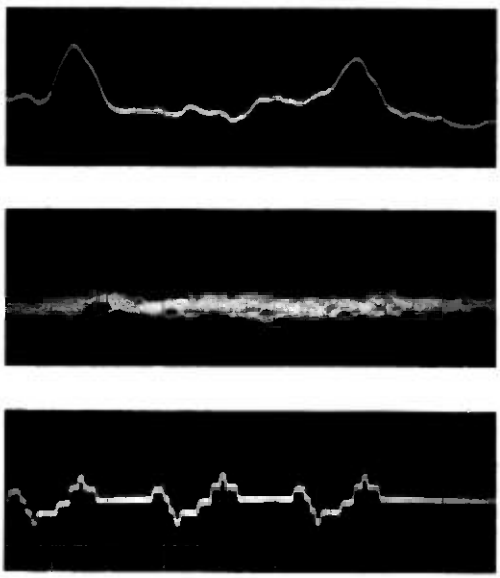


Fig. 49—Strip domains, excited by an 8-MHz RF signal, are detected in a single-loop. The raw signal, approximately 5 μv in amplitude, amplified and demodulated is shown in the upper trace. The bias field was taken above bubble collapse for the middle trace. A composite of the drive current is shown in the lower trace. Horizontal scale is 20 μs/div.

spirit of "field access" permalloy structures. However, perhaps one less processing level will be needed if the transfer can be accomplished within the framework of the propagate conductor levels themselves.

Such a compatible transfer gate can be realized by nesting apertures within one another as per the upper turn of the closed loop in Fig. 50. When this arrangement is used, it is only necessary to modify the propagation pulse sequence to cause bubbles in the transfer position to move into exit paths, whereas bubbles not in the transfer position are only idled backward and forward and always in a manner consistent with the retention of data when normal propagation resumes. Returning to Fig. 50, bubbles in the nine-step loop propagate clockwise with the standard pulse sequence [···12341234···]. However, a bubble located at the transfer position "1a" is transferred to position "4b" by the sequence [123412141234] in which the steps associated with the

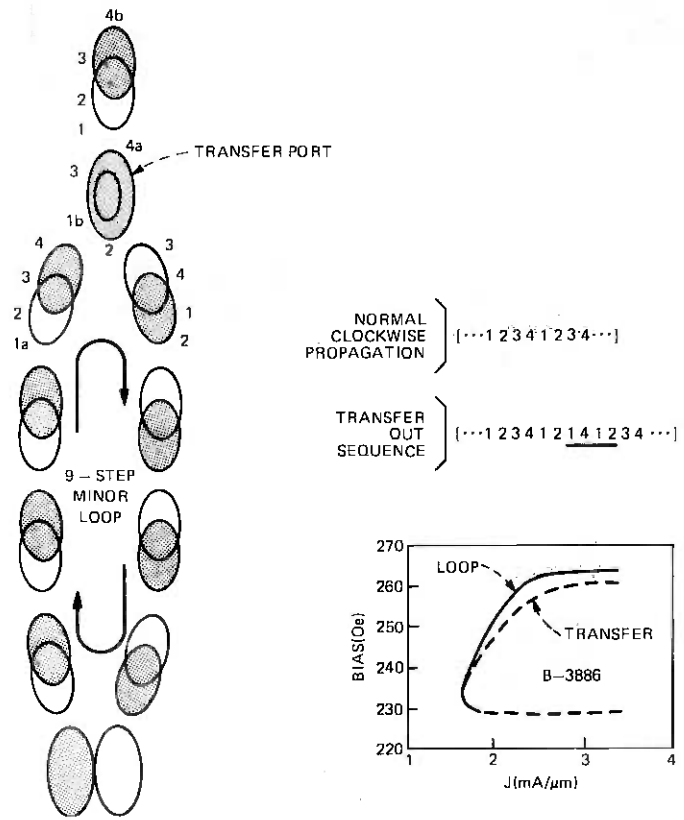


Fig. 50—Design and performance of a transfer port. In the transfer itself, the first level slot is nested within the slot of the second level. The propagation sequence is reprogrammed to transfer.

actual transfer have been underlined. It should be remembered that if data are to be preserved elsewhere, it is important to use only next number transitions in the transfer sequence.

Margins of propagation and transfer taken at 400 kHz are also included in Fig. 50. It can be seen that this transfer has very good margins indeed. More recent data have shown that, when the pulse which moves a bubble from 1b to 4a (4 in the underlined sequence) is lengthened by 0.25 μs, then an otherwise 1-MHz operation can be maintained. That is, only even-odd number transitions are permitted. ted.

7.7 Dual-conductor chip design considerations

There are two distinct aspects to the minimization of dissipation in dual-conductor devices. We discuss the minimization of power consumption on a local basis in Section 7.7.1 and on a chip basis in Section 7.7.2.

7.7.1 Optimization of conductor thickness

Power dissipation will always be a concern in current-access devices. For example, for $J = 1.5$ mA/μm, $\lambda = 8$ μm and $R = 0.1$ $\Omega/\square$, the continuous power dissipation is 14.4 μW/b. Even at $J = 1$ mA/μm and $\lambda = 4$ μm, the power dissipation is 1.6 μW/b. In this section, we turn to the problem of finding the correct thickness for conductors 1 and 2 under the constraint of an intervening insulator of thicknesses. From a practical standpoint, the insulator must withstand any voltage that exists between the two levels. The drive conductors are usually commoned at one end so that the maximum voltage encountered is the sum of the voltage drops of the layers. For the chip design discussed in the next section, the maximum inter-conductor voltage is 7.5 V.

For the purpose of calculating relative power consumption in a 2-conductor device, regard the fields produced by each level as those from an infinitesimally thin conducting sheet at the median plane of the actual conductor. For numerical work, the fields computed from eq. (25) will be used. When the two-conductor levels produce the same field, the relative power per unit area, normalized to the hole diameter $2a$, is

$$P = 2a/t_1 \bar{\mathcal{H}}_1^2 + 2a/t_2 \bar{\mathcal{H}}_2^2, \tag{10}$$

where t_1 and t_2 are the conductor thicknesses. The conductor of thickness t_1 is nearer the garnet. The two-conductor levels are separated by an insulator s units thick, as shown in Fig. 51. Values for P were computed at the maximum of $\bar{\mathcal{H}}$ near $r \approx a$ for $h/a = 0.8$ subject to the height constraint

$$a\tau = t_1 + s + t_2. \tag{11}$$

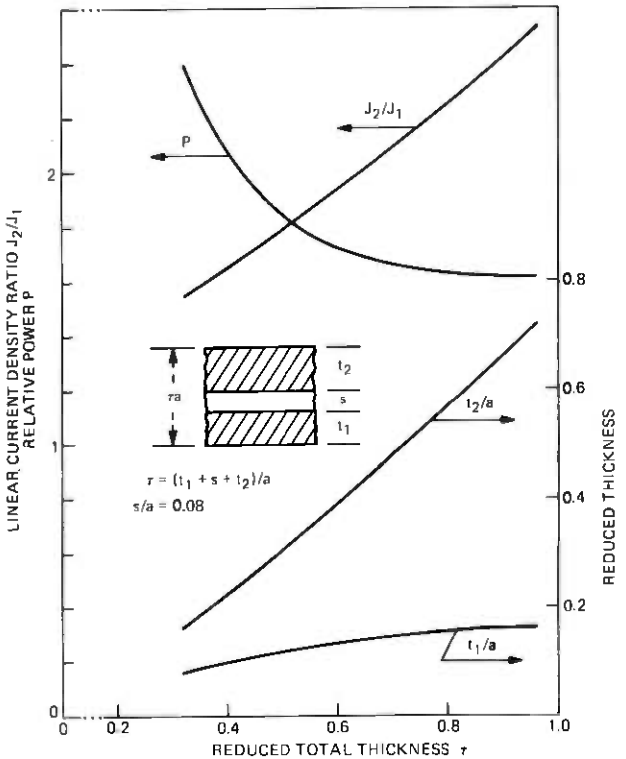


Fig. 51—Nomograph to determine the optimum thicknesses of levels 1 and 2 to minimize power dissipation in a dual-conductor circuit. Analysis based on a circular aperture of diameter $2a$ and garnet thickness $h = a$.

For each of several τ values, the minimum of eq. (10) was found. The results are shown in Fig. 51 together with the ratio of the current densities J_2/J_1 for the two levels. This ratio follows from the requirement of equal fields from the levels: $J_2/J_1 = \bar{\mathcal{H}}_1/\bar{\mathcal{H}}_2$.

The behavior of P as a function of τ represents a trade-off between efficient field production—$\bar{\mathcal{H}}$ large—and efficient power transmission in the conductor levels—t large. The best compromise between these conflicting requirements occurs for $\tau \approx 1$. The broad minimum of P about $\tau \approx 1$ permits a wide choice of τ with only minor changes in power consumption. For example, a 50-percent reduction of τ from 0.96 to 0.48 causes only a 16-percent increase in P. For current experimental devices, typical dimensions are $t_1 = 0.25$ μm, $s = 0.15$ μm, and $t_2 = 0.35$ μm. Using $a = 2$ μm, one finds $\tau = 0.38$. Although a larger value does reduce power consumption, it also increases processing complexity.

7.7.2 *Design of a major-minor chip with partitioning*

We now introduce partitioning of the conducting layers in a dual-conductor shift-register or major-minor chip design. In a partitioned chip, the main storage area and the major track are independently driven with a resultant reduction in chip dissipation. The shift-register and the major-minor chip layouts are very similar except that the shift-register uses parallel tracks rather than minor loops. Both of these designs allow for redundancy in the storage area. Given the functions of propagation, generation, annihilation, and transfer, the design of dual-conductor chips proceeds much as for field-access chips. There are, however, considerations peculiar to these current-accessed chip designs that we shall bring to your attention in this section.

The design and performance characteristics of a hypothetical major-minor chip with an 8-μm period and 262,144-b capacity is outlined in Fig. 52. Similar designs have been proposed by J. L. Smith and P. I. Bonyhard. Bubbles are assumed to move in bursts at a 1-MHz stepping rate. The chip is partitioned into independently accessed storage and detector-major track zones. The resistance everywhere is assumed to be 0.1 Ω/square. Each zone is driven at the current density indicated in the figure. If the entire chip was run continuously at 1 MHz, it would dissipate over 4 W. With partitioning, these zones can be independently accessed, reducing the chip dissipation substantially with no sacrifice in performance. In a partitioned chip it is, of course,

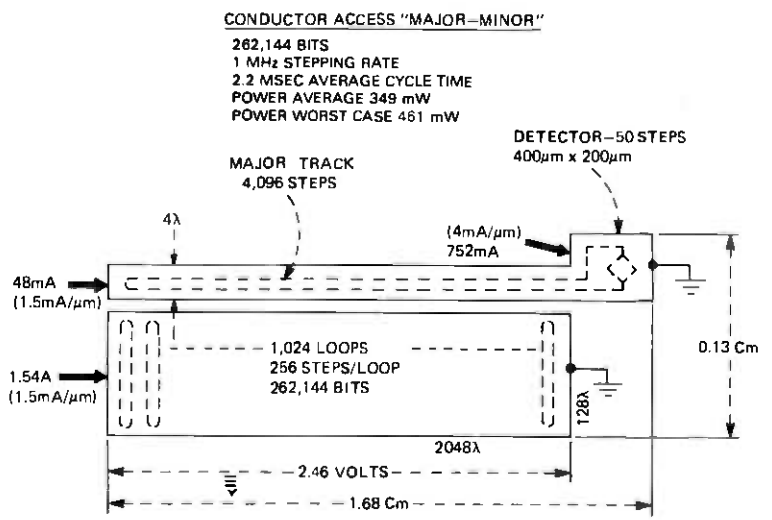


Fig. 52—Proposed design of a partitioned "major-minor" dual-conductor chip. The chip is made rectangular to increase the drive impedance and partitioned to permit selective drive to reduce the on-chip power dissipation.

necessary to move bubbles between zones. Propagation across slots was discussed with generators in Section 7.4.

It is desirable to make the impedance of the storage area compatible with driver circuitry. If the geometry of the storage area is laid out as a square, then it will present about 0.1 Ω resistance to its drivers—completely independent of the storage capacity. However, if the conductor sheet is reconfigured as a rectangle, its resistance can be "transformed" into a better match for semi-conductor drivers.

If the resulting "stick-like" chip is too awkward to handle, it can be folded, since the current (and bubbles) will have no difficulty negotiating the turns. A rectagular chip does have the advantage that it can dissipate heat more readily than a square chip and that it is well configured for a coilless package.

The number of minor loops is determined by noting that 1024 loops are, in binary multiples, the most that can be driven from a 5 V supply. This storage area would dissipate 3.77 W if driven continuously, but its duty factor is very low (0.03) so the average power dissipation is just 110 mW. The detector and major track do run essentially continuously; however, they together represent a very small fraction of the chip area and, in total, dissipate only 239 mW. As a result, the total chip dissipation is but 349 mW. Further information on cycle time, duty factor, and power dissipation is summarized in Table II. If the chip capacity is quadrupled (1024 b/loop), the chip dissipation would only double and the cycle times would remain about the same.

VIII. EPITAXIAL GARNET MATERIAL

In this section, we examine the material requirements for current-access bubble devices. The specific material parameters of particular importance to this technology are discussed and a garnet system proposed which allows relatively high-domain wall mobility along with adequate uniaxial magnetic anisotropy.

8.1 Growth and characterization

All compositions grown for the current-access high-density bubble devices were grown on $Gd_3Ga_5O_{12}$ substrates by the LPE dipping

Table II—Characteristics of partitioned 262-kbit dual-conductor chip

			Average Case			Worst Case		
Zone	Steps	Transit Time (ms)	Cycle Time	Duty Factor	Power (mW)	Cycle Time	Duty Factor	Power (mW)
Minor loops	256	0.256	0.128	0.03	110	0.256	0.06	230
Major track and detector	4146	4.146	4.146	0.97	239	4.146	0.94	231
			4.274	1.00	349	4.402	1.00	461

technique using supercooled melts.[17] Axial rotation was maintained constant at approximately 100 rpm during the growth period. The details of the apparatus and growth techniques have been discussed elsewhere.[18-21] Growth temperatures were restricted to 875 ± 60°C. Lattice parameter measurements were obtained using the HPM method.[22] Magnetic property measurements were obtained using the standard techniques.[23-25]

8.2 Material properties

Table III lists the parameters which must be considered when designing a material to be used in dual conductor-current access devices. Items 1 to 11 in Table III are important parameters to be considered when designing a bubble material for use in most bubble circuit technologies. Items 11 to 13 are of particular importance for dual conductor-current access devices. Because of the high-speed performance inherent in this technology, the mobility of the bubble material is of primary importance. Dynamic coercivity and bubble propagation angle are two other parameters that become increasingly important in small bubble devices. Low dynamic coercivities allow lower drive currents to be used for bubble propagation. A discussion of the importance of the bubble propagation angle is presented in Section 7.3. An approach to increasing the mobility while maintaining adequate q values as well as an approach to lowering the dynamic coercivity is presented later in this section.

Epitaxial garnet films of nominal composition $Y_{1.2}Sm_{0.4}Lu_{0.5}$-$Ca_{0.9}Ge_{0.9}Fe_{4.1}O_{12}$ have been shown to be useful in 16-μm period field-access bubble devices.[26] For 3-μm diameter bubbles, this composition offers the advantages of domain wall mobility slightly in excess of 250 cm s^{-1} Oe^{-1} along the q values of 5 and precise control of the temperature dependence of the magnetic properties by the appropriate adjustment of melt chemistry. Figure 53 shows the material length

Table III—Material parameters for current-access devices

1. Saturation magnetization, $4\pi M_s$ (*Gauss*)
2. Exchange constant, A (ergs cm^{-1})
3. Uniaxial anisotropy, K_u (ergs cm^{-3})
4. Material length parameter, l(μm)
5. Anisotropy field, H_K (Oe)
6. Quality factor, q (dimensionless)
7. Bubble collapse field, H_{col} (Oe)
8. Lattice parameter, L.P. (Å)
9. Magnetostriction coefficient, λ_{111}, λ_{100}
10. Temperature dependence of H_{col}, (% $°C^{-1}$)
11. Bubble mobility, μ (cm/sec-Oe)
12. Dynamic coercivity, ΔH_c (Oe)
13. Bubble propagation angle, ψ (degrees)

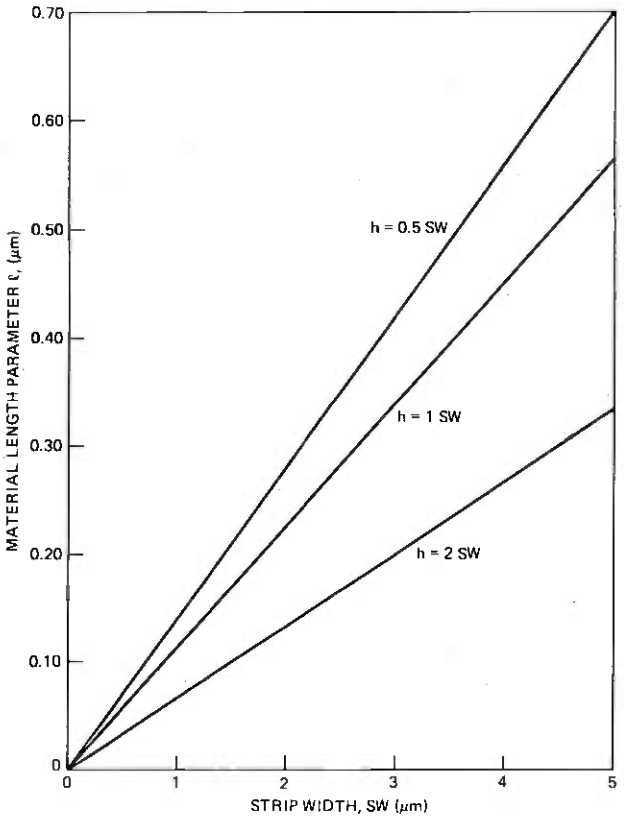


Fig. 53—Material length parameter (l) vs strip width (sw) with the thickness-to-strip-width ratio as a parameter.

parameter (l) as a function of domain strip width (sw) with the thickness-to-strip-width (h/sw) ratio as a parameter. Depending upon the h/sw ratio, the l parameter needed for a strip width of 1.7 μm ranges from 0.115 to 0.235 μm. For the case of h/sw = 1, the material length parameter must be 0.191 μm.

Extending the $(YSmLuCa)_3(FeGe)_5O_{12}$ system to smaller bubbles results in the nominal composition $Y_{1.0}Sm_{0.5}Lu_{0.7}Ca_{0.8}Ge_{0.8}Fe_{4.2}O_{12}$ which exhibits 1.7-μm diameter bubbles and the properties listed in Table IV. Materials grown from the general system $(YSmLuCa)_3(FeGe)_5O_{12}$ have exhibited dynamic coercivities ranging from 1.4 Oe to 3.2 Oe. The higher coercivities being measured are on the larger anisotropy materials, supporting smaller diameter bubbles. Because of the particular importance of bubble mobility in the current-access technology, a technique was needed to maintain a q value of

approximately 3 to 4 for materials exhibiting an l value of 0.16 to 0.19 μm while also increasing the domain wall mobility.

It would be advantageous to be able to decrease the samarium content of the composition listed in Table IV while maintaining the q value, which would result in an increased mobility. One approach to achieving this is to increase the concentration of the most "anisotropy active" constituents. Figure 54 shows the calculated uniaxial anisotropy as a function of the lutetium content of garnet compositions in the three systems $(YLuSmCa)_3(FeGe)_5O_{12}$, $(LaLuSmCa)_3(FeGe)_5O_{12}$, and $(LaLuSm)_3(FeGa)_5O_{12}$. The samarium content and the Curie temperature are used as parameters. Curves A through C are calculated for Ca-Ge substitution of ~0.8 moles per garnet formula unit and a Curie temperature of ~480°K. Curves D and E represent gallium substitution leading to a Curie temperature of ~440°K. We point out that the curves shown in Fig. 54 do not take into account the lattice constant constraint usually imposed on epitaxial garnet films. The compositions required to achieve a lattice constant match with $Gd_3Ga_5O_{12}$ (GGG) substrates (12.383Å) are indicated by the point on each curve marked L.M. Comparison of curves A and B in Fig. 54 reveals that, to decrease the samarium content from 0.5 to 0.3 in the garnet formula and maintain both the anisotropy of ~35000 ergs-cm^{-3} and a lattice parameter match with the GGG substrate, compositions outside the $(YLuSmCa)_3(FeGe)_5O_{12}$ system must be considered. Curve C indicates that the $(LaLuSmCa)_3(FeGe)_5O_{12}$ system allows both the uniaxial anisotropy of ~35000 ergs-cm^{-3} and a lattice match to be achieved. A discussion of the advantages of lanthanum have been presented elsewhere.[27]

It is well known that materials with high $4\pi M_s$ require relatively little diamagnetic ion substitution. The advantages realized in the divalent-tetravalent ion substitution system can be obtained with less complicated crystal compositions when high moment material is required.[28] For materials exhibiting a high $4\pi M_s$, Ca^{2+}-Ge^{4+} substitution could result in too high a Curie temperature making it difficult to

Table IV—Composition: $Y_{1.0}Sm_{0.5}Lu_{0.7}Ca_{0.8}Ge_{0.8}Fe_{4.2}O_{12}$

Lattice parameter	12.383 Å
$4\pi M_s$	510 G
K_u	36,600 ergs cm^{-3}
l	0.19 μm
A	2.7×10^{-7} ergs cm^{-1}
H_k	1800 Oe
q	3.5
μ	~300 cm/S-Oe
ΔH_c (dynamic coercivity)	3.0 Oe

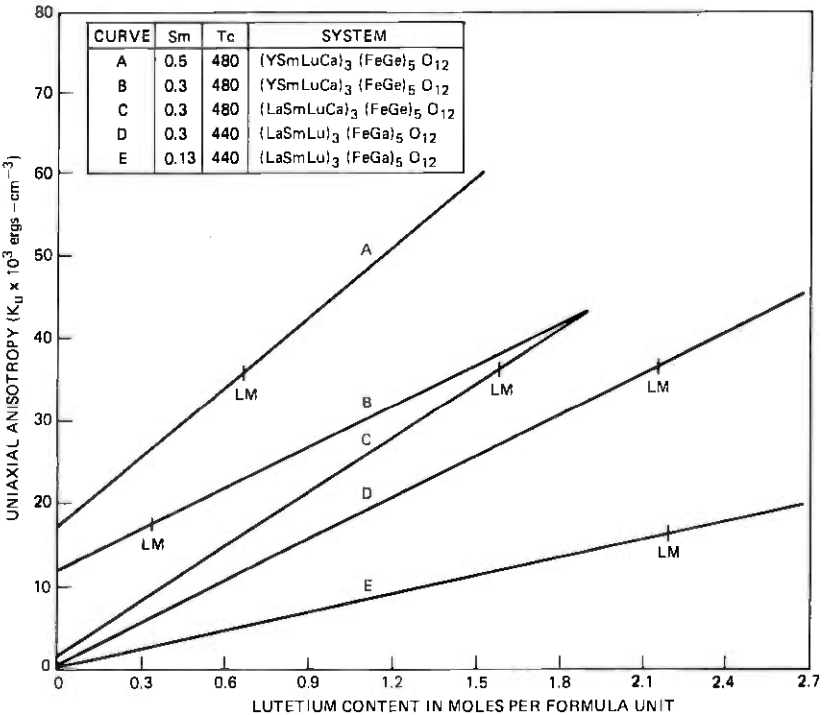


Fig. 54—Calculated uniaxial magnetic anisotropy vs lutetium content for the systems $(YSmLuCa)_3(FeGe)_5O_{12}$, $(LaSmLuCa)_3(FeGe)_5O_{12}$, and $(LaLuSm)_3(FeGa)_5O_{12}$. The samarium content and the Curie temperature are varied. The GGG lattice matching compositions are indicated by the symbol LM on each curve.

control the temperature dependence of the magnetic properties to match the barium ferrite bias magnets.

A slight imbalance of the divalent-tetravalent ion ratio could also result in increased coercivity. The immersion of a gallium garnet substrate into a melt supersaturated with an iron garnet containing Ca^{2+}-Ge^{4+} might be expected to show increased substrate dissolution prior to initial film growth as compared to the same substrate immersed in a melt supersaturated with a gallium containing iron garnet. Considering the complications in melt chemistry and the careful control of growth conditions necessary to minimize compositional fluctuations in Ca^{2+}-Ge^{4+} substituted garnets, a move to simpler film compositions would be advantageous.

Curve D illustrates the anisotropy calculated for the $(LaLuSm)_3(FeGa)_5O_{12}$ system. Note that the lattice match composition occurs at a lutetium concentration of ~2.15 moles per garnet formula unit and that a uniaxial anisotropy of ~35000 ergs-cm^{-3} can be obtained. Table V lists a nominal composition in the

$(LaLuSm)_3(FeGa)_5O_{12}$ system along with some measured magnetic properties. Curve E shows that lowering the samarium concentration to 0.13 moles per garnet formula unit results in a significantly lower anisotropy for the GGG lattice matching composition.

8.3 Dynamic coercivity

The dynamic coercivity of epitaxial bubble garnet films supporting ~1.7-μm diameter bubbles has been measured as a function of film thickness.[29] Figure 55 shows the dynamic coercivity measured by a bubble translation technique as a function of inverse film thickness. All curves are for material exhibiting 1.7-μm diameter bubbles. Curve C is for a sample in the $(YLuSmCa)_3(FeGe)_5O_{12}$ system, while curves A and B represent two compositions in the $(LaLuSm)_3(FeGa)_5O_{12}$ system. Curve A is for a samarium concentration of ~0.13 moles per garnet formula unit, while curve B represents data taken on a sample containing a samarium concentration of 0.3 moles. Note that the gallium material exhibits significantly lower dynamic coercivities than does the Ca^{2+}–Ge^{4+} substituted material.

8.4 Melt chemistry

The phase equilibria observed in the $(YSmLuCa)_3(FeGe)_5O_{12}$ system have been presented previously.[19] Lanthanum does not form an iron garnet, and lutetium iron garnet has been reported being grown only once in bulk form,[30] possibly because of phase equilibria difficulties. As the concentration of lanthanum increases in the pseudo-ternary Flux—Fe_2O_3—$\sum Ln_2O_3$, the garnet phase field would be expected to narrow considerably. Figure 56 shows a section of the pseudo-ternary Flux—Fe_2O_3—$\sum Ln_2O_3$, where $\sum Ln_2O_3$ corresponds to the sum of La_2O_3+Lu_2O_3. The ratio of La_2O_3/Lu_2O_3 was fixed at 0.332 to allow stress-free epitaxial film growth on GGG substrates. Note that all data shown in the pseudo-ternary are for unsubstituted iron garnets. Substitution of diamagnetic ions could be expected to shift the location of the phase boundaries. Addition of gallium results in the garnet-perovskite boundary being displaced towards the Flux—Fe_2O_3 binary. Depending on the overall garnet oxide concentration in the melt, the

Table V—Composition: $La_{0.6}Lu_{2.1}Sm_{0.3}Ga_{0.9}Fe_{4.1}O_{12}$

$4\pi M_s$	466G
K_u	35,500 ergs cm^{-3}
A	2.0×10^{-7} ergs cm^{-1}
l	0.18 μm
H_K	1915 Oe
Q	4.1
μ	500–1000 cm/S-Oe
ΔH_c (dynamic coercivity)	~2 Oe

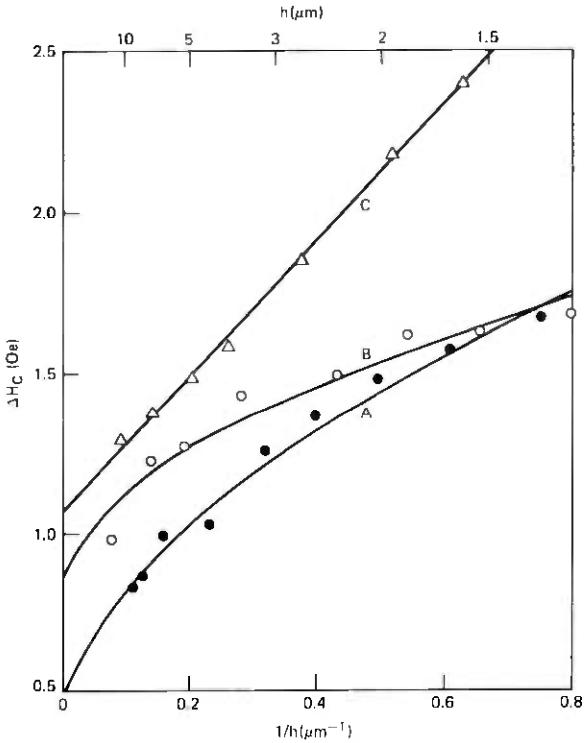


Fig. 55—Bubble dynamic coercivity (ΔH_c) as a function of inverse film thickness for 1.7-μm bubble material. Curve C is for $(YLuSmCa)_3(FeGe)_5O_{12}$. Curves A and B represent two compositions in the system $(LaLuSm)_3(FeGa)_5O_{12}$. Curve A is for a samarium concentration of ~0.13 moles per garnet formula unit. Curve B represents data on a sample containing 0.3 moles of samarium.

range of $Fe_2O_3/\sum Ln_2O_3$ over which garnet is found to be the primary phase field is from 6 to 9. For comparison, the Fe_2O_3/Y_2O_3 ratio range for garnet in the Flux—Fe_2O_3—Y_2O_3 pseudo-ternary is from 12 to 40.

Because of the large size of La^{3+}, one would expect a relatively low value for the La^{3+} distribution coefficient. The definition for the distribution coefficient of lanthanum is:

$$k^{La} = \frac{\left(\frac{La}{La+Lu+Sm}\right)_{crystal}}{\left(\frac{La}{La+Lu+Sm}\right)_{melt}},$$

where the symbols in the fractions represent the number of moles present. Although the value for the effective coefficient will be rate-dependent, the value for k^{La} under average growth conditions is found to be ~0.42.

The use of compositions in the $(LaLuSm)_3(FeGa)_5O_{12}$ system allows less complicated crystal compositions to be used. Lower dynamic coercivities and increased domain wall mobility can be obtained.

IX. PROCESSING

There are two major concerns in processing dual conductor structures. First, the two conductors should both be as close as possible to the epitaxial film supporting the bubbles. Hence, a prespacer (analogous to the prespacer used between conductor and epitaxial layer in standard field access devices) should be as thin as possible or completely omitted. In addition, it is immediately obvious that the dielectric layer between the conductors must be as thin as the limits of dielectric breakdown permit. Second, the very small period devices possible within the material constraints of this technology will require increasingly precise control for etching the fine-aperture metal film pattern. What is required is an etching method that will produce metal patterns dimensionally identical to the etch mask.

9.1 Material requirements

9.1.1 Insulating films

If a prespacer is used, the material and/or process selection is quite unrestrained as the surface is completely planar and the layer thickness

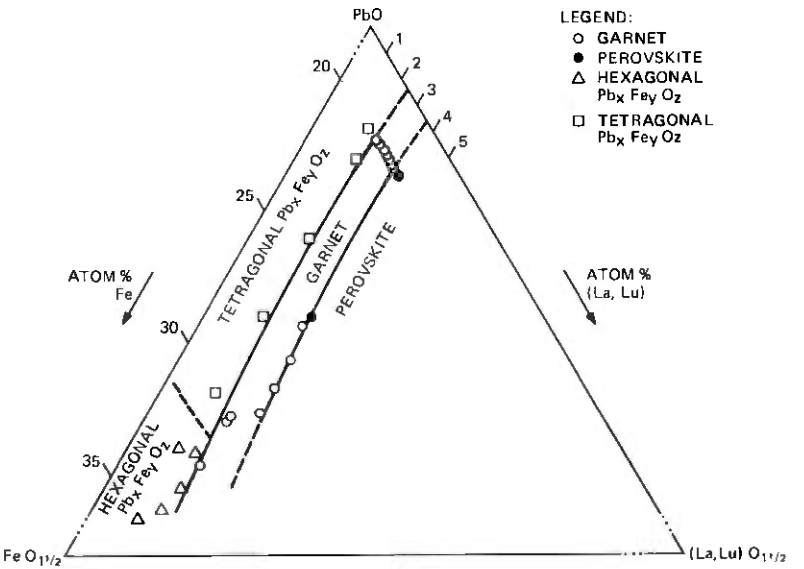


Fig. 56—A section of the pseudoternary Flux—Fe_2O_3—ΣLn_2O_3, where ΣLn_2O_3 represents La_2O_3 + Lu_2O_3.

required is very small. Thus, several avenues can probably be used: RF diode-sputtered SiO_2 or SiN; RF magnetron or S-gun deposited SiO_2 or SiN; plasma-deposited or CVD-deposited SiO_2 or SiN. In all the work reported, prespacers were SiO_2 deposited by RF plasma deposition using an Applied Materials Plasma I radial flow system.

The insulating layer between conductors has more stringent requirements due to the multiplicity of crossovers and the need for minimum thickness. Thus, very good dielectric properties are required and a process that deposits a conformal coating is mandatory, i.e., equivalent deposition rates on vertical and horizontal surfaces are required (Fig. 57). Higher-pressure processes such as plasma-deposition or CVD are clearly preferred over line-of-sight deposition methods such as evaporation or sputtering. A relatively low-temperature (c.f. 250 to 280°C) plasma SiO_2 deposition process has been developed using the Applied Materials Plasma I apparatus.[31] Dual conductor circuits fabricated using the plasma-SiO_2 insulator exhibit breakdown potentials equivalent to those expected for high-quality, thermally grown SiO_2 films of an equivalent thickness. We thus conclude that the step coverage over the dual conductor features is optimal.

9.1.2 Conductor materials

Though any highly conductive metal alloy system may be used, Al or Au alloys appear quite suitable, and most of the work reported has

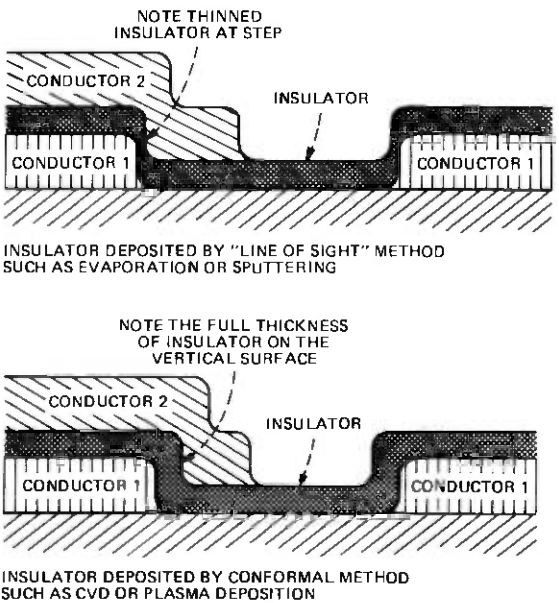


Fig. 57—Effect of conformal vs line-of-sight deposition on insulating film thickness at crossovers.

used Al alloys with low copper content (0.7 to 4 percent). The AlCu alloys are particularly suited because the relatively high current density required for some functions in the dual conductor circuits ($>10^5$ A/cm^2) indicates the use of conductor materials with low electromigration rates. The AlCu alloys have been evaporated using a standard e-gun source and planetary wafer fixturing to maximize step coverage. It is advantageous to carry out this deposition at as low a substrate temperature as possible to prevent stress buildup in the AlCu films.[32]

9.2 Metal patterning

9.2.1 Photolithography

Standard contact-print photolithography was used on devices reported here.[33] We anticipate that other high-resolution methods such as projection printing, X-ray exposure, etc. may be used equally well.

9.2.2 Metal patterning

In standard rotating magnetic field-access bubble devices, the Al conductor layer is relatively simple, and at least for 16-μm period circuits can be etched with standard wet chemical methods. In dual conductor-accessed circuits, however, wet-etch processes are unusable in small period circuits due to the undercutting of the photoresist pattern by the isotropic etch process. This undercutting severely limits the dimensions achievable in conductor patterning. In fact, since these circuits will invariably require that near state-of-the-art resolution be reproduced in the resist mask, no isotropic process or process which biases the conductor line or space dimensions with respect to the photoresist pattern should be used. In the following section, we describe a plasma-etching process that does indeed exactly reproduce the etch mask in the Al film. It is conceivable that other methods may be used, such as the planar anodization method described by D. K. Rose.[34]

For an etch patterning process, very high material etch selectivity will be required. If no prespacer is used, the first-level etch process cannot be allowed to cut into the epitaxial garnet film and thin it selectively. Similar constraints operate when etching the second conductor, since any thinning of the underlying insulator layer will obviously reduce the dielectric breakdown strength between the two conductors. Plasma etching is an obvious choice to satisfy the requirements for etching Al alloy films for high-density dual conductor circuits for the following reasons: (*i*) plasma etching uses a chemical process to remove metal and very high etch selectivity can be achieved for different materials, (*ii*) careful selection and control of plasma process parameters can replicate photoresist patterns in aluminum films without measurable (<0.1 μm) dimensional variation. (A good survey of plasma etch processes can be found in Refs. 35 and 36.)

The particular RF (13.56-MHz) plasma process adapted is based on one developed by D. Wang, et al.[37] The apparatus used was designed by C. J. Mogab and F. B. Alexander,[38] and the etch chamber is shown schematically in Fig. 58. The basic etch process involves the following reaction:

$$Al(s) + 3Cl(plasma) \rightarrow AlCl_3(solid)$$

followed by volatilization of the $AlCl_3$ solid

$$2AlCl_3(solid) \rightarrow (AlCl_3)_2(gas).$$

There is evidence that the anisotropic etch process is promoted by the impact of charged particles on the horizontal surface (Fig. 59).

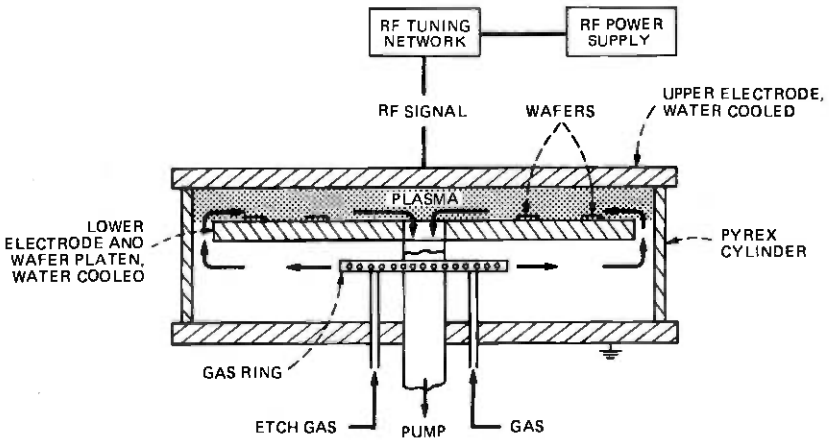


Fig. 58—Radial flow plasma etch apparatus.

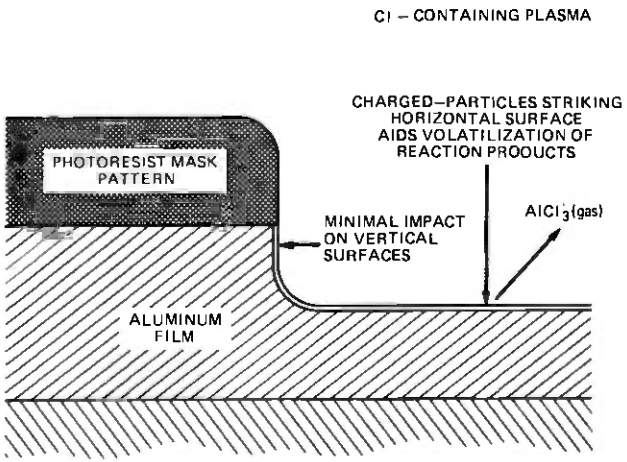


Fig. 59—Schematic of anisotropic etch process.

Obviously, a bias-free etching process will produce near-vertical walls on the metal patterns, which might pose problems in the crossover features in the dual conductor circuits. To date, the combination of conformal plasma SiO_2 and good step coverage in the Al evaporation process have minimized the potential problems, both with respect to dielectric breakdown and also propagate drive margin reduction, which might be caused by high-resistance crossovers. Figures 60 to 63 are SEM photographs of plasma-etched, dual-conductor patterns. Figure 60 shows a horizontal, straight-line propagate path, while Figure 61 shows several vertical loops. Figure 62 shows one of the generator designs tested. Figure 63 shows, at the same magnification, three data storage loops designed and fabricated within identical linewidth and alignment tolerances. The three loops are of 8-, 6-, and 4-μm period.

X. MAGNETIC FIELDS AND BUBBLE MOTION

Early conductor-drive devices[39,40] confined currents to discrete conductor lines narrower than a bubble diameter. The present approach to conductor propagation distributes current over a conducting sheet with openings which provide the necessary spatial variations of magnetic field. To furnish some insights and a design aid for conductor-access propagation, two simple geometries—parallel conducting stripes

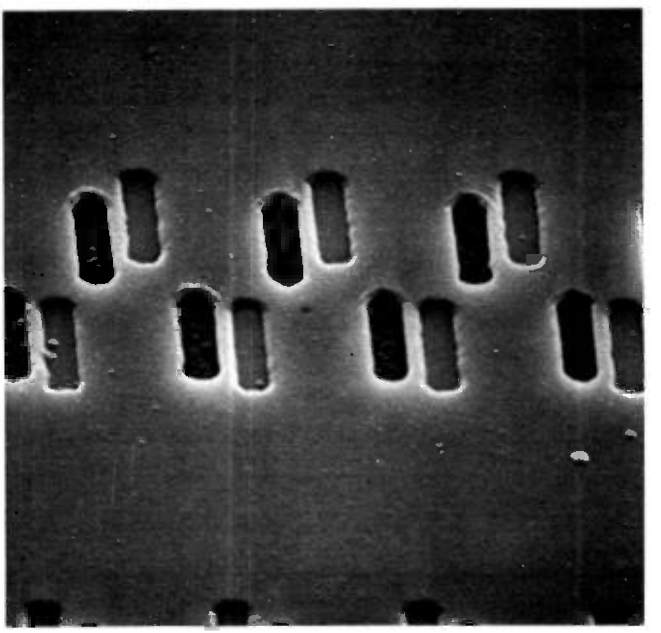


Fig. 60—An SEM of an 8-μm period dual-conductor propagate pattern. The darker apertures are in the first conductor and the lighter in the second conductor pattern. These apertures were etched in the AlCu alloy by a plasma etch technique using a Cl-containing gas (2860X).

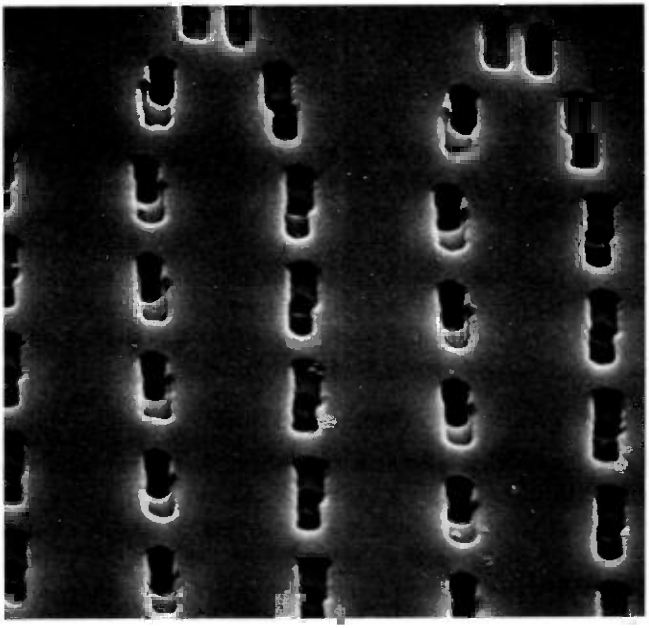


Fig. 61—An SEM of 8-μm memory storage loops (2200X).

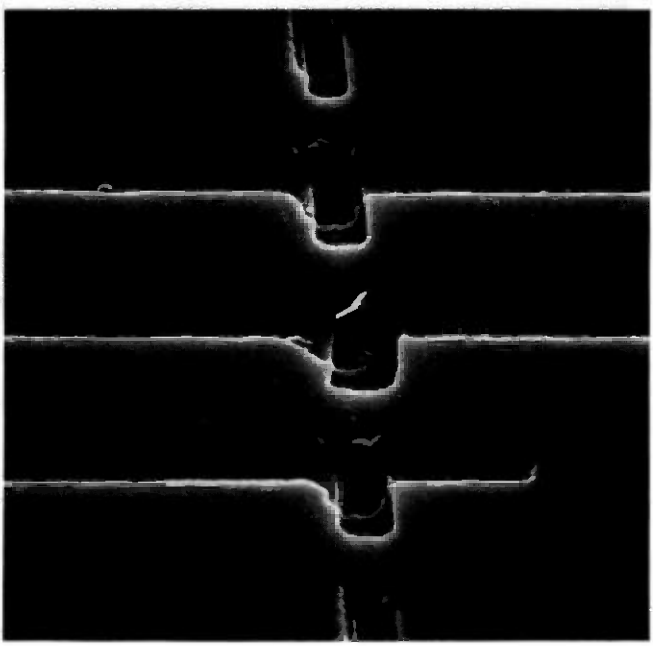


Fig. 62—An SEM of an experimental generator pattern plasma etched in AlCu (3160X).

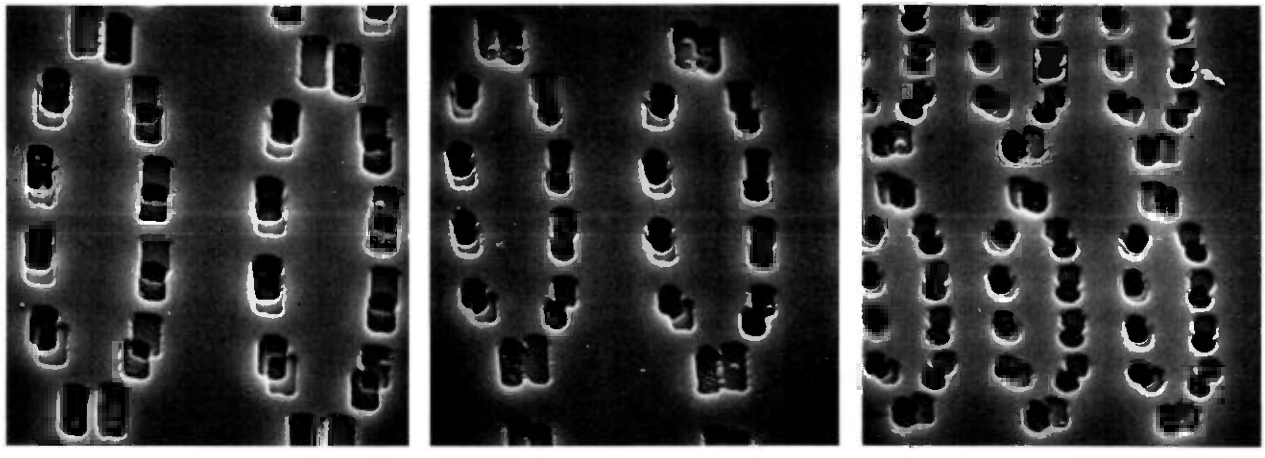


Fig. 63—Data storage loops plasma etched in AlCu films. The three patterns follow identical dimensional design rules (2870X). (a) 8-μm period; (b) 6-μm period; (c) 4-μm period.

and a circular hole in an infinite conducting sheet—were investigated. Bubble motion in a gradient field was also studied.

10.1 Stripe conductors

Consider an infinitesimally thin conducting sheet in the plane $z = 0$. From this sheet, infinite in the y direction, remove material to form conducting stripes of width $2a$ and spacing $s > 2a$; see Fig. 64. Each stripe carries a current $I = -2aJ_o$ in the y direction, where J_o is the constant linear current density of each stripe. In MKS units, the contribution[41]

$$H_z(n) = (J_o/2\pi) \ln[r_2(n)/r_1(n)] \tag{12}$$

of the nth stripe to the z component of the field intensity may be expressed in dimensionless form by the quantity

$$\mathscr{H}_s(n) = \frac{4\pi H_z(n)}{J_o} = \ln\left[\frac{(\xi + \alpha - n\sigma)^2 + \zeta^2}{(\xi - \alpha - n\sigma)^2 + \zeta^2}\right], \tag{13}$$

where the lengths $r_1(n)$ and $r_2(n)$ are defined in Fig. 64, $\xi = x/h$, $\zeta = z/h$, $\alpha = a/h$, and $\sigma = s/h$. For bubble propagation, one is interested in the average intensity

$$\bar{\mathscr{H}}_s(n) = h^{-1} \int_{z_s}^{z_s+h} H_z(n)\, dz, \tag{14}$$

where z_s is the garnet-conductor separation and h the garnet-film thickness. After integrating eq. (14) and summing over the stripes, one finds

$$\bar{\mathscr{H}}_s = \sum_n \sum_{j=1}^{2} (-1)^j [\zeta_j \ln A_{nj} + 2\beta_j(n) \text{Arctan } C_{nj}], \tag{15}$$

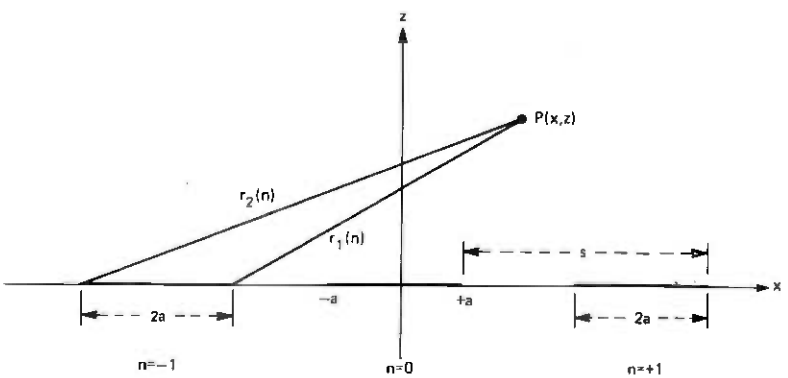


Fig. 64—Coordinate system for an array of infinite conducting stripes. The conductors, $2a$ units wide and spaced s units apart, occupy the plane $z = 0$. Each stripe extends to infinity in the $\pm y$ directions, carries a current I, and bears an index n.

where

$$\mathscr{H}_s = 4\pi\bar{H}_z/J_o = \text{dimensionless average intensity},$$

$$\bar{H}_z = \sum_n \bar{H}_z(n) = \text{average intensity},$$

$$\zeta_1 = z_s/h, \qquad \zeta_2 = \zeta_1 + 1, \qquad \beta_j(n) = \xi + (-1)^j\alpha - n\sigma,$$

$$A_{nj} = [\zeta_j^2 + \beta_2^2(n)]/[\zeta_j^2 + \beta_1^2(n)],$$

$$C_{nj} = \beta_j(n)/[\beta_j^2(n) + \zeta_1\zeta_2].$$

When the number of stripes is small, say, less than 100, the best way to evaluate eq. (15) is by programming the right-hand side as it stands. As the number of stripes n_s is increased, this procedure becomes less attractive because $2n_s$ evaluations of the logarithm and arctangent functions are needed for each value of x. Note that, for $n_s < \infty$, $\mathscr{H}_s$ is not periodic; for a complete numerical description of eq. (15) in the vicinity of the stripes, the number of function evaluations is $2n_xn_\alpha n_s^2$, where n_x is the number of points per period s and n_α the number of α values of interest.

When n_s is large, say, $n_s > 100$, the behavior of $\mathscr{H}_s$ near the array center may be approximated by an infinite number of stripes. This choice permits a significant reduction in computation time but fails to give information about the behavior near the array edges—a matter that will be investigated separately. As $n_s \to \infty$, the dimensionless average intensity, regarded as a function of ξ, has the properties

$$\mathscr{H}_s(\xi) = \mathscr{H}_s(\xi + \sigma) \quad \text{and} \quad \mathscr{H}_s(-\xi) = -\mathscr{H}_s(+\xi). \tag{16}$$

Consequently, one can restrict evaluation of $\mathscr{H}_s$ to the points $0 \leq \xi \leq \sigma/2$. Equation (16) implies $\mathscr{H}_s(0) = \mathscr{H}_s(\sigma/2) = 0$. This result may also be deduced from symmetry considerations. For an origin at the center of an opening between stripes, the same symmetry arguments require $\mathscr{H}_s$ to be antisymmetric about this origin. For numerical evaluation of $\mathscr{H}_s$ in the limit $n_s \to \infty$, eq. (15) was rewritten as

$$\mathscr{H}_s = \mathscr{H}_1 + \mathscr{H}_2, \tag{17}$$

where

$$\mathscr{H}_1 = \sum_{j=1}^{2} (-1)^j\zeta_j\Bigg[\ln A_{oj} + \sum_{n=1}^{N}[\ln A_{nj}A_{-nj}] - (8\alpha\xi/\sigma^2)\Bigg[\zeta(2) - \sum_{n=1}^{N} n^{-2}\Bigg] - [8\alpha\xi(\xi^2 + \alpha^2 - 3\zeta_j)/\sigma^4]\Bigg[\zeta(4) - \sum_{n=1}^{N} n^{-4}\Bigg]\Bigg],$$

$$\mathcal{H}_2 = 2\sum_{j=1}^{2}\sum_{n=-N}^{+N}[(-1)^j\beta_j(n)\text{Arctan } C_{nj}]$$

$$-[16\alpha\xi(\zeta_2^3-\zeta_1^3)/\sigma^4]\left[\zeta(4)-\sum_{n=1}^{N} n^{-4}\right],$$

$\zeta(k)$ = Riemann zeta function.[42]

Equation (17) adds the contributions from the $2N + 1$ stripes nearest the origin as they stand in eq. (15) to the contribution from the remaining stripes approximated by a truncated series in the inverse powers of the stripe index n. Since the leading term in the power series

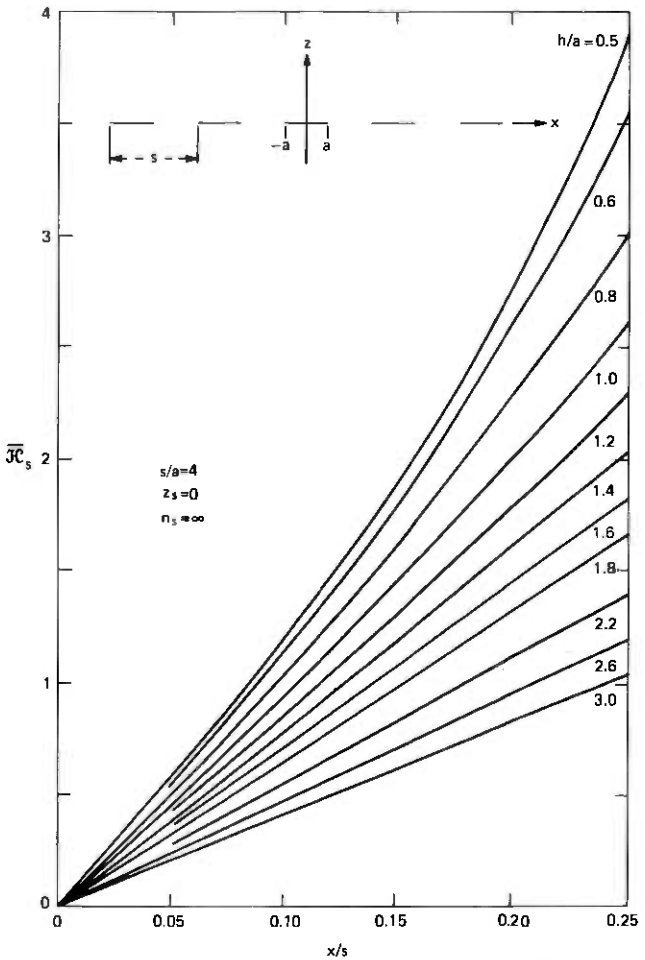


Fig. 65—Dimensionless average value $\overline{\mathcal{H}}_s$ of the z component of magnetic field as a function of x/s. The curves are graphs of eq. (17) for $s = 4a$ and $z_s = 0$. The ordinate $\overline{\mathcal{H}}_s$ is numerically equal to the field in oersteds when the linear current density of each stripe is 1 mA/μm.

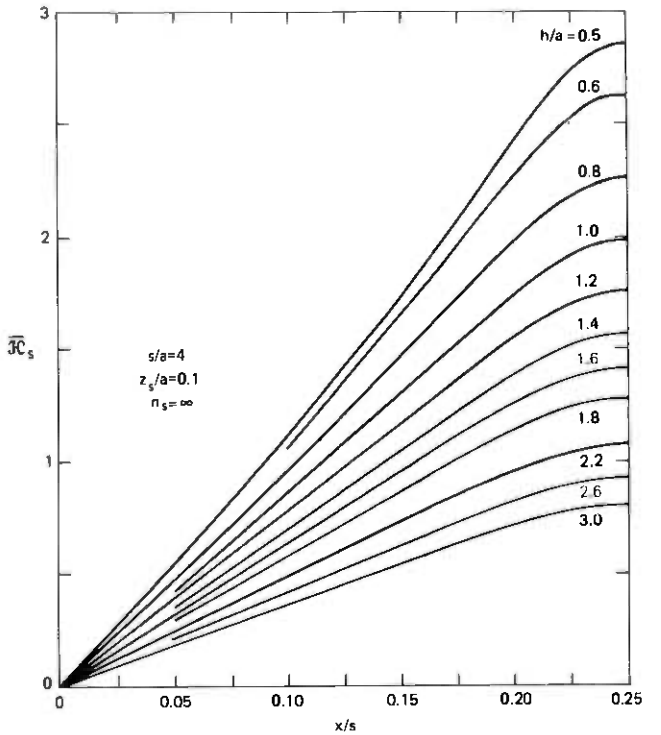


Fig. 66—Same as Fig. 65, except $x_n = 0.1a$.

is n^{-2}, the field at the origin due to a pair of stripes, remote from and symmetric about the origin, decreases as the inverse square of the distance of either stripe from the origin. Accuracy requirements determine N. For the choice $N = 10$, used in the computations presented here, the estimated error in $\mathcal{H}_s$ is less than 10^{-5}. Figures 65 to 69 show graphs of $\mathcal{H}_s$ vs x/s for parameter values useful to the device designer. Curves of $\mathcal{H}_s$ for $z_s = 0$ are unique in that each has a cusp at $x/s = 0.25$, where the slope $\partial\mathcal{H}_s/\partial x$ is not continuous. The choice $s = 4a$ requires calculation only for $0 < \xi \leq \sigma/4$ because $\mathcal{H}_s$ is symmetric about $\xi = \pm\sigma/4$. The origin of this symmetry may be understood by superimposing another current sheet with current density $\mathbf{J} = (J_o/2)\hat{j}$ on the existing stripe pattern of Fig. 64. This addition leaves the z component of field unaltered. It creates, however, a system of adjacent conducting stripes of width $2a$, carrying the current $I = J_o a$ alternately in the positive and negative y direction. Such an array of stripes enjoys symmetry about the points of discontinuity. Contributions to the z component of field from any pair of stripes symmetrically disposed about a selected discontinuity will add; $\mathcal{H}_s$ will, therefore, have extrema at points of current-density discontinuity. To illustrate how the average field intensity decreases with increasing distance from the conduc-

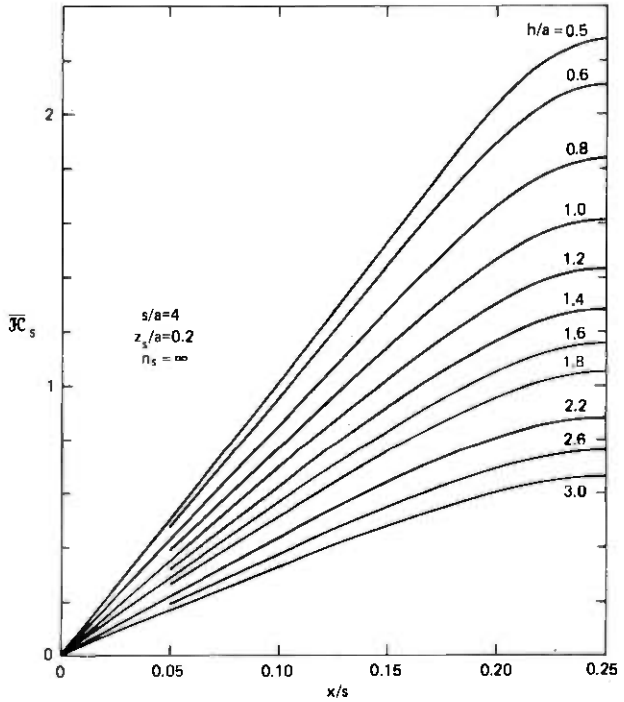


Fig. 67—Same as Fig. 65, except $z_s = 0.2a$.

tor array, values of $\bar{\mathscr{H}}_s$ were computed for several stand-off distances z_s. Since $\bar{\mathscr{H}}_s$ assumes its maximum value at $x/s = 0.25$ for $s/a = 4$, the stripe edge was chosen for the calculations. See Fig. 70.

To convert to cgs units, write $\bar{H}_z = \bar{\mathscr{H}}_s J_o$ Oe and express J_o in A/mm or mA/μm. In other words, the dimensionless quantity $\bar{\mathscr{H}}_s$ gives the average field intensity in oersteds when the linear current density J_o is 1 mA/μm.

The gradient of the average field $\partial \bar{H}_z/\partial x$ can be obtained with only a slight increase in computing effort because the needed function values are available. The dimensionless gradient is

$$\bar{\mathscr{G}} = \left(\frac{4\pi h}{J_o}\right)\frac{\partial \bar{H}_z}{\partial x} = 2\sum_{n}\sum_{j=1}^{2}(-1)^j \text{Arctan } C_{nj}. \tag{18}$$

When the number of stripes n_s is large, one may use the approximation

$$\bar{\mathscr{G}} = 2\sum_{-N}^{+N}\sum_{j=1}^{2}[(-1)^j \text{Arctan } C_{nj}] - (8\alpha/\sigma^2)\left[\zeta(2) - \sum_{1}^{N} n^{-2}\right]$$
$$- (8\alpha/\sigma^4)[3\xi^2 + \alpha^2 - \xi_2^2 - \zeta_2\zeta_1 - \xi_1^2]\left[\zeta(4) - \sum_{1}^{N} n^{-4}\right], \tag{19}$$

an expression correct to the 4th order in the quantity $1/n$. Equation

(19) adds the contribution from the $2N + 1$ stripes nearest the origin to the contribution, approximated by a truncated series in powers of $1/n$, of the infinitely many remaining stripes. Accuracy requirements determine the value of N.

10.2 Infinite conducting sheet with circular hole

Consider again an infinitely thin conducting sheet in the plane $z = 0$. Let the sheet contain a circular hole, extend to infinity, and support

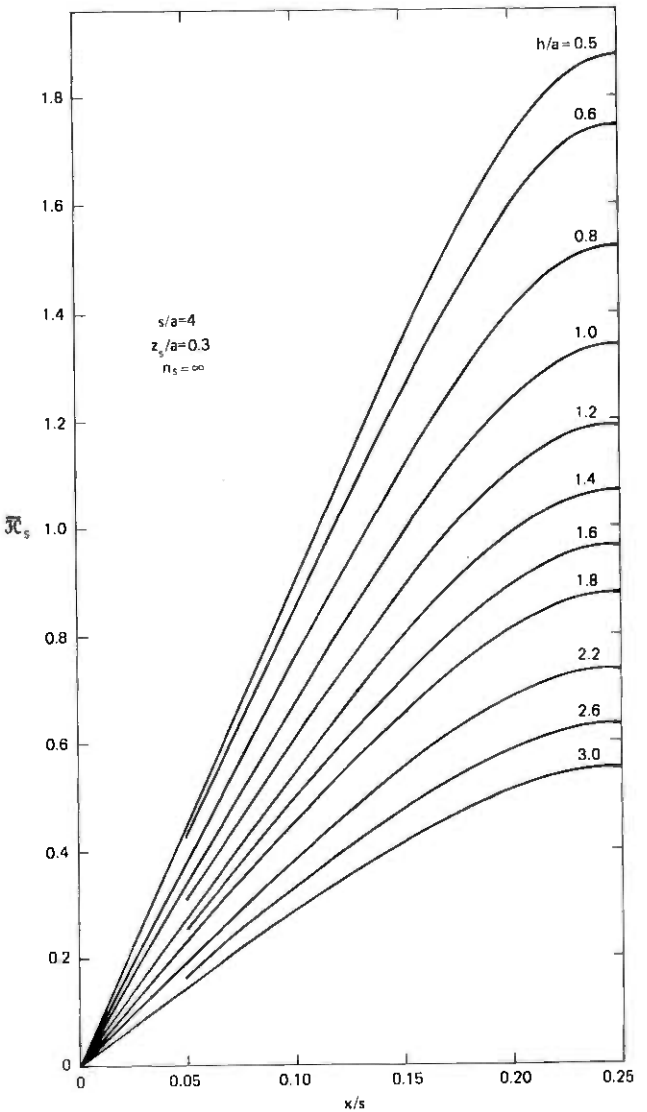


Fig. 68—Same as Fig. 65, except $z_s = 0.3a$.

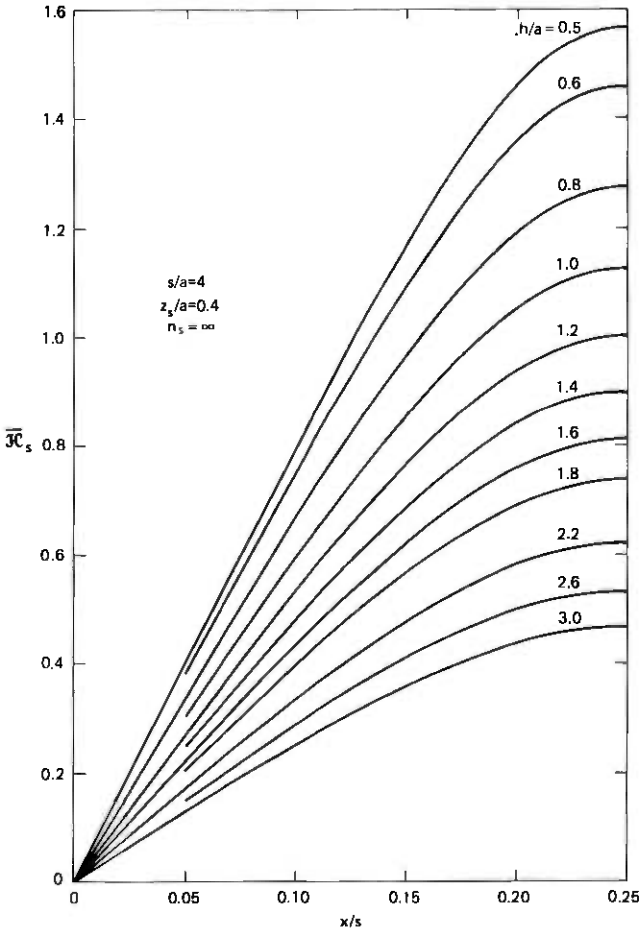


Fig. 69—Same as Fig. 65, except $z_s = 0.4a$.

the constant current density $\mathbf{J} = J_o\hat{\imath}$ at points infinitely removed from the hole. The origin of coordinates is chosen at the symmetry point of the geometry as shown in Fig. 71. The problem to be solved—a means of obtaining the z component of the magnetic field at the point $P(r, \theta)$—falls into two parts: (*i*) solution of a boundary-value problem to yield the linear current density $\mathbf{J}$ in the plane $z = 0$, and (*ii*) integration over the source distribution to obtain the field.

The first of these is discussed by Milne-Thompson[43] in connection with two-dimensional fluid flow around a cylindrical obstacle. The current density is

$$\mathbf{J} = J_o[1 - (a/\rho)^2\cos 2\phi]\hat{\imath} - J_o(a/\rho)^2\sin 2\phi\hat{\jmath}, \qquad (20)$$

where

J_o = constant linear current density at infinity,

a = hole radius,

ρ = radial source coordinate,

ϕ = angular source coordinate.

The reader may quickly verify that eq. (20) satisfies the requisite boundary conditions. The current density is tangential at $\rho = a$ and $\mathbf{J} \to J_o\hat{\imath}$ as $\rho \to \infty$. One can now obtain the magnetic field in MKS units from

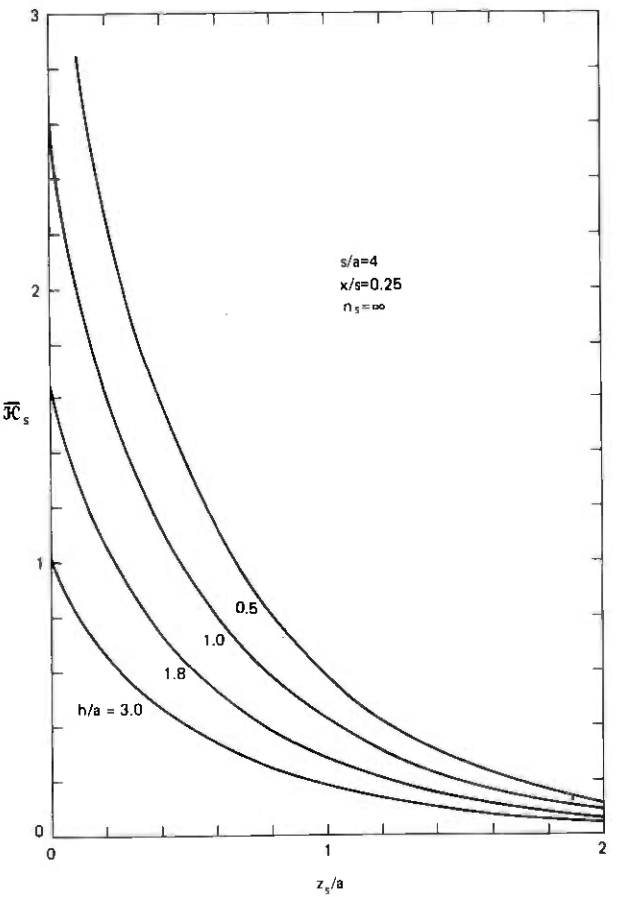


Fig. 70—Dimensionless average value $\bar{\mathcal{H}}_s$ of the z component of magnetic field as a function normalized stand-off distance z_s/a. The curves are graphs of eq. (17) for $s = 4a$ and $x = 0.25s$. The ordinate $\bar{\mathcal{H}}_s$ is numerically equal to the field in oersteds when the linear current density of each stripe is 1 mA/μm.

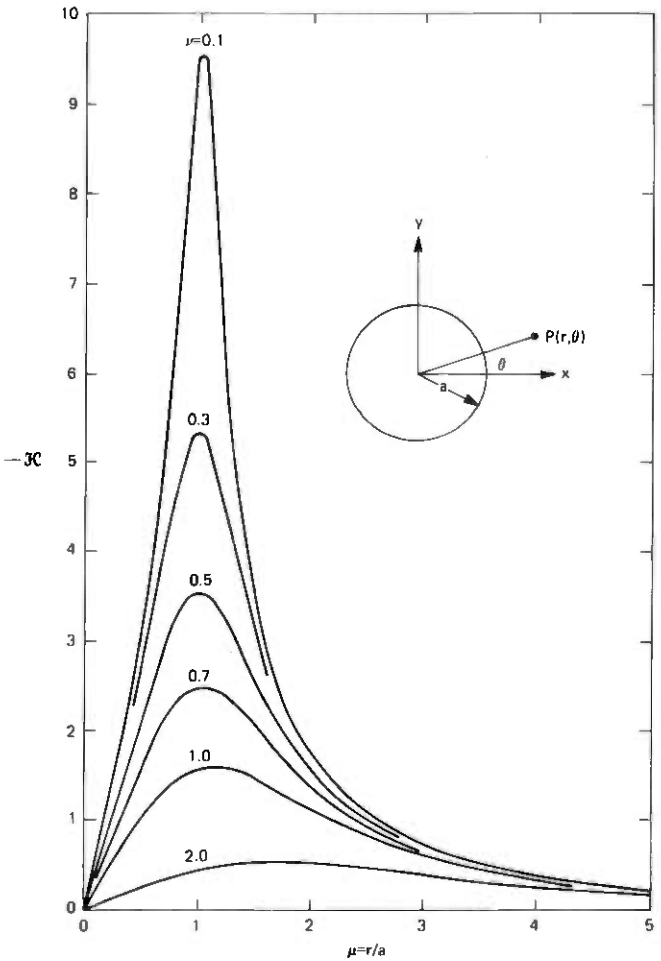


Fig. 71—Current sheet with circular hole. The dimensionless z component of the magnetic field is shown as a function of the reduced distance $\mu = r/a$ for several values of $\nu = z/a$. The curves are graphs of $\mathscr{H}$ defined by the integral below eq. (23). The ordinate $\mathscr{H}$ is numerically equal to the field in oersteds when $\theta = 90°$ and the linear current density approaches 1 mA/μm as $r \to \infty$.

$$4\pi\mathbf{H} = \int_S \frac{\mathbf{J} \times \mathbf{R}}{R^3}\, d\xi\, d\eta, \tag{21}$$

where

$$\mathbf{R} = (x - \xi)\hat{\imath} + (y - \eta)\hat{\jmath} + z\hat{k} \quad \text{and} \quad R = |\mathbf{R}|.$$

The surface integral, with respect to the source coordinates ξ and η, extends over the entire $\xi - \eta$ plane apart from the points $\rho = (\xi^2 + \eta^2)^{1/2} < a$, where the current density vanishes. After projecting

eq. (21) on the z-axis and expressing all components in cylindrical coordinates, one obtains a double integral of the form

$$4\pi H_z = J_o \int_0^{2\pi} d\phi \int_a^{\infty} \Phi(r, \theta, z, \rho, \phi)\, d\rho. \tag{22}$$

All radial integrals are tabulated; see, for example, Gradshteyn and Ryzhik.[44] The remaining angular integration, however, must be done numerically. After appropriate changes of the variables and collecting terms, eq. (22) reduces to

$$4\pi H_z = J_o \mathscr{H} \sin\theta, \tag{23}$$

where

$$\mathscr{H} = 2\int_0^{\pi} (F_1 + F_2 + F_3)\, d\psi = \mathscr{H}(\mu, \nu),$$

$$F_1 = \frac{C_0 + C_1 C + C_2 C^2 + C_3 C^3 + C_4 C^4}{\sigma^2(\sigma^2 - \mu^2 C^2)(\sigma^2 + 1 - 2\mu C)^{1/2}},$$

$$F_2 = C\ln[(\sigma^2 + 1 - 2\mu C)^{1/2} + 1 - \mu C],$$

$$F_3 = \mu(2C^2 - 1)\sigma^{-3}\ln\left[\frac{\sigma(\sigma^2 + 1 - 2\mu C)^{1/2} + \sigma^2 - \mu C}{\sigma - \mu C}\right],$$

$$\mu = r/a, \qquad \nu = z/a,$$

$$\sigma = (\mu^2 + \nu^2)^{1/2}, \qquad C = \cos\psi,$$

$$C_0 = \mu\sigma^2(1 + \sigma^2),$$

$$C_1 = (\sigma^2 + \mu^2)(1 - \sigma^2),$$

$$C_2 = -\mu[\sigma^2(\sigma^2 + 3) + 2\mu^2],$$

$$C_3 = 2\mu^2(\sigma^2 - 1),$$

$$C_4 = 4\mu^3.$$

Numerical integration over the angular variable of eq. (22) employed the automatic numerical quadrature routine QUAD.[45] The dependence of $\mathscr{H}$ on the radius and altitude is shown in Fig. 71. Using this figure and eq. (23), one may now estimate H_z in MKS units. In the hybrid units described earlier, the z component of magnetic field in oersteds is numerically equal to $\mathscr{H}$ when $J_o = 1$ mA/μm and $\theta = \pi/2$.

To obtain information about the average value,

$$\bar{H}_z = \frac{1}{h}\int_{z_o}^{z_o+h} H_z\, dz = \frac{J_o \sin\theta}{4\pi h}\int_{z_o}^{z_o+h} \mathscr{H}(\mu, \nu)\, dz, \tag{24}$$

the dimensionless quantity

$$\bar{\mathscr{H}} = \bar{\mathscr{H}}(\mu, \gamma) = h^{-1} \int_{z_s}^{z_s+h} \mathscr{H}\, dz = \gamma^{-1} \int_{\nu_s}^{\nu_s+\gamma} \mathscr{H}(\mu, \nu)\, d\nu \qquad (25)$$

was computed. Here $\nu_s = z_s/a$ and $\gamma = h/a$. Since several γ values are of interest, one may take advantage of the linear property of integration to reduce the number of function evaluations. Index the γ values of interest with j such that $\gamma_{j+1} > \gamma_j$. To obtain a uniform formula and accommodate Fortran requirements, one must choose $\gamma_1 = 0$ and the smallest value of interest γ_2. Then, with the notation $\bar{\mathscr{H}}_j = \bar{\mathscr{H}}(\mu, \gamma_{j+1})$, one finds

$$\bar{\mathscr{H}}_j = \frac{1}{\gamma_{j+1}} \sum_{i=1}^{j} g_i, \qquad (26)$$

where

$$g_i = \int_{\nu_s+\gamma_i}^{\nu_s+\gamma_{i+1}} \mathscr{H}(\mu, \nu)\, d\nu.$$

Each g_i was evaluated with a 3-point Gaussian rule; for the γ_j used in the computation the estimated error is of order 10^{-4}. The radial slope $\bar{\mathscr{S}} = \partial\bar{\mathscr{H}}/\partial\mu$ was also computed using the approximation

$$\bar{\mathscr{S}}(\mu_k, \gamma_j) \approx [\bar{\mathscr{H}}(\mu_{k+1}, \gamma_j) - \bar{\mathscr{H}}(\mu_{k-1}, \gamma_j)]/2\Delta u, \qquad (27)$$

where Δu is the constant difference between adjacent values of μ. The dependence of $-\bar{\mathscr{H}}$ and $-\bar{\mathscr{S}}$ on r/a is shown in Figs. 72 and 73, respectively. Contour maps of constant $\bar{H}_z$ demonstrate the overall behavior of the field. Figure 7 shows contours of $\bar{H}_z$ = constant for the reduced stand-off distances $z_s/a = 0.06$ and 0.3. For both, $h/a = 0.8$.

10.3 Linear chain of oval holes

Current experimental devices use quasi-elliptic openings rather than circular holes; see, for example, Fig. 29. The work of Richmond[46] with electric fields about a linear grating permits more realistic magnetic field calculations. Consider an infinite chain of identical oval holes on the y-axis, spaced 2b units apart, each with semiaxes $a_2 > a_1$ parallel to the y and x axes, respectively. See Fig. 74 for the oval at the origin. Because of symmetry and the periodicity 2b in the y direction, Laplace's equation need be solved only in the region of Fig. 74. Richmond[46] gives the conformal transformation

$$z = (a_1/c)w + (a_2/\gamma)\cosh^{-1}(\cos\gamma \cosh w) \qquad (28)$$

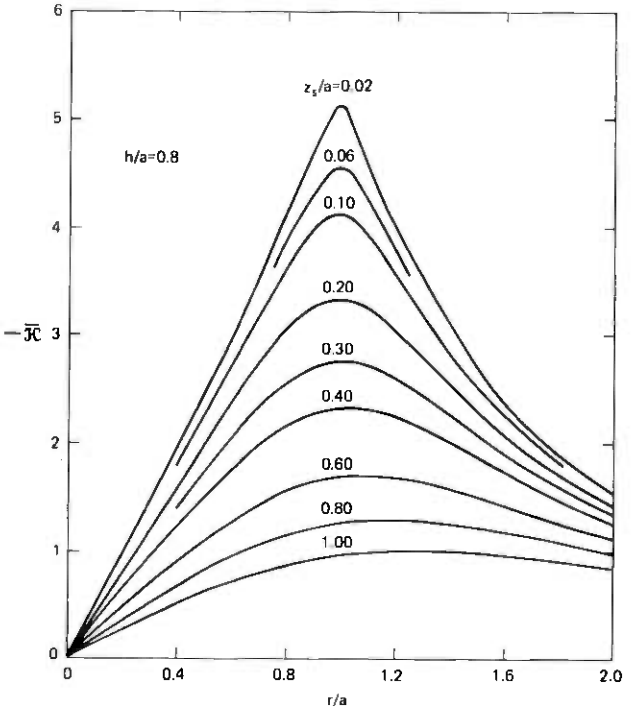


Fig. 72—Dimensionless average value $\bar{\mathcal{H}}$ of z component of magnetic field as a function of r/a for an infinite current sheet with a circular hole. The curves are graphs of eq. (25) for $h = 0.8a$, where a is the hole radius. The ordinate $\bar{\mathcal{H}}$ is numerically equal to the field in oersteds when $\theta = 90°$ and the linear current density approaches 1 mA/μm as $r \to \infty$.

from which one finds the complex linear current density

$$J = \frac{2bJ_o}{\pi}\left[\frac{dw}{dz}\right]^* = \left[\frac{\alpha_2 \gamma c J_o(\cos^2\gamma \cosh^2 w - 1)^{1/2}}{\alpha_1\gamma(\cos^2\gamma \cosh^2 w - 1)^{1/2} + c \cos\gamma \sinh w}\right]^* \qquad (29)$$

outside the oval. Here J_o, a real constant, is the linear current density as $x = \to \infty$, $\alpha_1 = a_1/a_2$, $\alpha_2 = 2b/\pi a_2$, $z = x + iy$, $w = u + iv$, and * means complex conjugate. Inside the oval, $J = 0$. The real parameters c and γ satisfy the nonlinear equations $\cos\gamma \cosh c = 1$ and $c(\gamma\alpha_2 - 1) = \alpha_1\gamma$. Elimination of c from the equations leads to an efficient algorithm, based on Newton's method, for machine computation of γ for practical values of $\alpha_1 \leqslant 1$ and α_2. The iteration scheme

$$\gamma_{i+1} = \gamma_i - \frac{f_i}{f_i'} = \gamma_i + \frac{\gamma_i g_i[\alpha_1\gamma_i + g_i(1 - \alpha_2\gamma_i)]\cos\gamma_i}{\alpha_1\gamma_i^2 + g_i^2\cos\gamma_i}, \qquad (30)$$

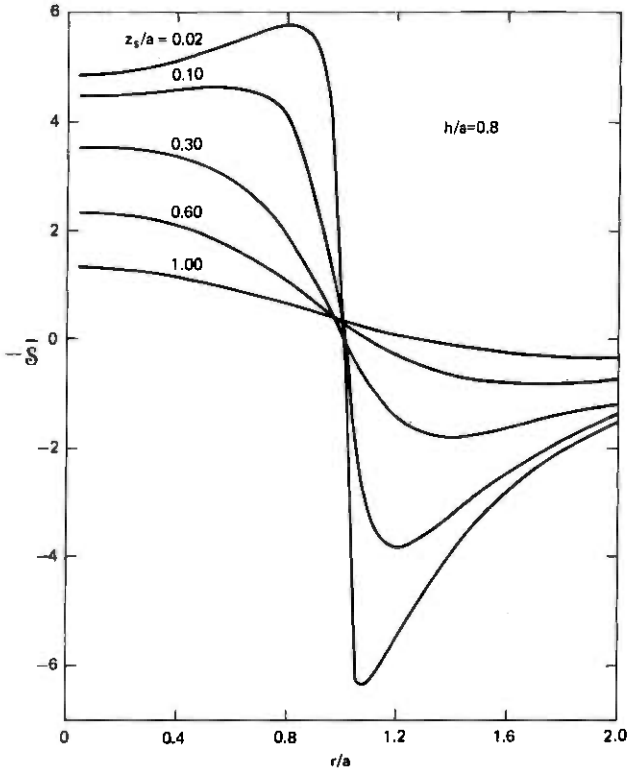


Fig. 73—Dimensionless slope $\mathcal{S} = \partial\mathcal{H}/\partial\mu$ of the z component of magnetic field as a function of r/a for an infinite current sheet with a circular hole. The curves are graphs of eq. (27) for $h = 0.8a$, where a is the hole radius.

where

$$f = \alpha_1 g^{-1} + \gamma^{-1} - \alpha_2,$$

$$g = \ln[(1 + \sin\gamma)/\cos\gamma],$$

$$f' = df/d\gamma,$$

$$f_i = f(\gamma_i),$$

$$f_i' = df/d\gamma \text{ evaluated at } \gamma = \gamma_i,$$

may be started with $\gamma_o = 1/\alpha_2$. If the iteration process ends with γ_j, the parameter c follows from $c = g_j = \ln[(1 + \sin\gamma_j)/\cos\gamma_j]$.

By setting u or v constant in eq. (28), one obtains equipotential or flow lines, respectively. Several flow lines, indexed with the angle variable $\Psi = 180v/\pi$, are shown in Fig. 74. The oval satisfies the parametric equations

$$x/a_2 = (\alpha_1/c)\cosh^{-1}(q/\cos\gamma) \qquad (31a)$$

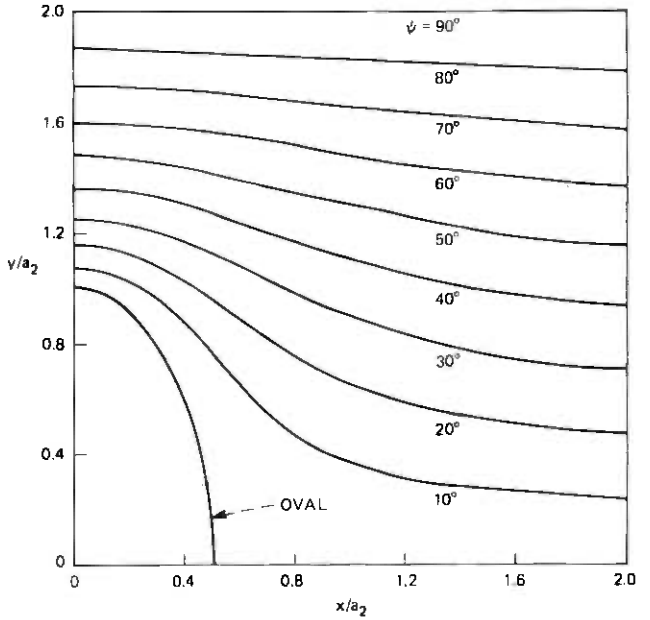


Fig. 74—Current flow near an infinite chain of oval holes. Identical oval holes with semiaxes $a_2 = 2a_1$ and spacing $2b = 4a_2$ occupy the y-axis. The flow lines, indexed with $\psi = 180v/\pi$, were obtained from eq. (28) with v = constant. Note the almost elliptical shape of the oval for the parameter ratios $a_1/a_2 = 0.5$ and $b/a_2 = 2$ used here.

and

$$y/a_2 = (1/\gamma)\cos^{-1}q, \tag{31b}$$

where $0 \leq \cos\gamma \leq q \leq 1$. Richmond[46] discusses the departure from ellipticity. For geometries of practical interest, say, $a_1/a_2 \approx 0.5$ and $b/a_2 \approx 2$, the oval may be regarded as an ellipse. The distribution of current along the line of minimum separation between ovals satisfies eq. (29) with $u = 0$. It was computed from

$$\frac{J}{J_o} = \frac{\alpha_2\gamma c(1 - \cos^2\gamma\cos^2 v)^{1/2}}{\alpha_1\gamma(1 - \cos^2\gamma\cos^2 v)^{1/2} + c\cos\gamma\sin v}, \qquad 0 \leq v \leq \pi/2 \tag{32}$$

and is shown in Fig. 75. The coordinate value associated with v follows from eq. (28).

To complete the field computations, an efficient method for the numerical double integration of the components of eq. (21) must be constructed. This work is in progress.

10.4 Bubble motion

Experimental evidence, Fig. 34, for example, strongly suggests bubbles with different hardness parameters. A study of bubble motion will

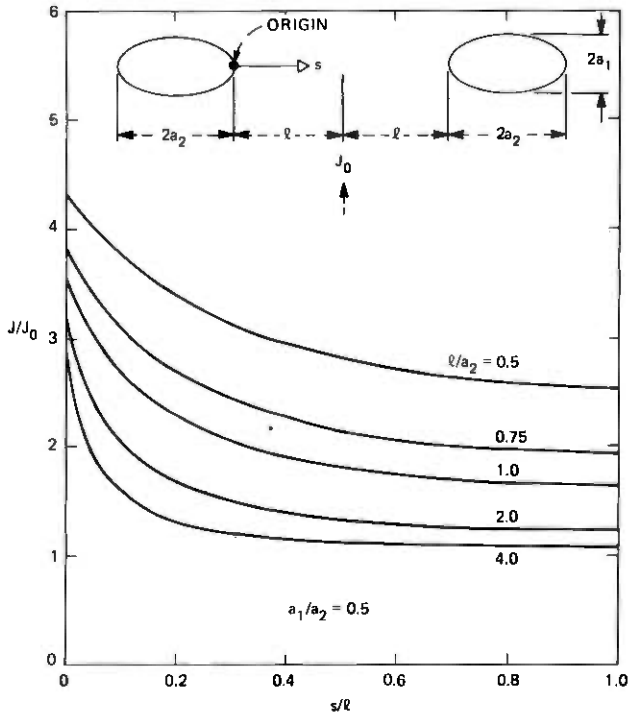


Fig. 75—Distribution of linear current density along the line of minimum separation between ovals of an infinite chain. The current density is normalized to J_0, its value infinitely far from the ovals. The graphs were obtained from eq. (32).

aid the device designer by prediction of trajectories and transit times.

With the assumption that the bubble speed,[47]

$$V = \mu_b[d_b|\nabla\bar{H}_z| - 8H_c/\pi] \geqslant 0, \tag{33}$$

is independent of the hardness or S number, the equations of motion in rectangular coordinates are

$$\dot{x}_i = -VT_{ij}e_j, \tag{34}$$

where V is a scalar point function, μ_b is the bubble mobility equal to one-half the wall mobility, d_b is the bubble diameter, $\bar{H}_z$ is the z component of field averaged over the bubble height, H_c is the coercivity field, $T_{11} = T_{22} = \cos\phi$, $T_{12} = -T_{21} = \sin\phi$, $\hat{e} = \nabla\bar{H}_z/|\nabla\bar{H}_z|$, e_j is the projection of $\hat{e}$ on the jth coordinate axis, and ϕ is the angle between $-\hat{e}$ and the local velocity V. The angle ϕ depends on S according to $\sin\phi = 8SV/\gamma d_b^2|\nabla\bar{H}_z|$ and is regarded as a fixed parameter in eq. (34). Here γ is the gyromagnetic ratio of the material.[48] See also Fig. 76. When applied to bubble motion due to fields from a hole of radius a in a current sheet, eqs. (33) and (34) assume the form

$$\dot{\xi}_1 = -(F_c/\mu^3)[\xi_1\xi_2(\mu\bar{\mathscr{P}} - \bar{\mathscr{H}})\cos\phi + (\mu\xi_2^2\bar{\mathscr{P}} + \xi_1^2\bar{\mathscr{H}})\sin\phi], \quad (35a)$$

$$\dot{\xi}_2 = +(F_c/\mu^3)[\xi_1\xi_2(\mu\bar{\mathscr{P}} - \bar{\mathscr{H}})\sin\phi - (\mu\xi_2^2\bar{\mathscr{P}} + \xi_1^2\bar{\mathscr{H}})\cos\phi], \quad (35b)$$

where

$$F_c = 1 - 8H_c/\pi d_b\,|\nabla\bar{H}_z|,$$

$$|\nabla\bar{H}_z| = (J_o/a)[\bar{\mathscr{P}}^2\sin^2\theta + (\bar{\mathscr{H}}/\mu)^2\cos^2\theta]^{1/2},$$

$$\sin\theta = \xi_2/\mu,$$

$$\cos\theta = \xi_1/\mu,$$

$$\mu = (\xi_1^2 + \xi_2^2)^{1/2},$$

$$\xi_1 = x/a,$$

$$\xi_2 = y/a,$$

$$\bar{\mathscr{P}} = \partial\bar{\mathscr{H}}/\partial\mu.$$

Hybrid units were used for eq. (35); this allows substitution for H_c in oersteds. The current density J_o, however, must be in units of mA/μm. The dot over the variables on the left-hand side of eq. (35) means differentiation with respect to the normalized time $T = t/\tau$, where $\tau = a^2/\mu_b\, d_b\, J_o$. Solutions to these coupled, nonlinear differential equations must, in general, be obtained by numerical techniques. Some special cases, amenable to an analytical approach, are now described.

The equations of motion, eq. (34), assume a simple form when written in polar coordinates:

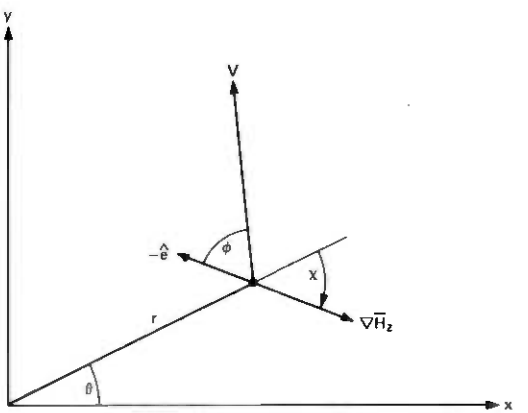


Fig. 76—Coordinate system and symbols used in the equations of bubble motion. Note that the angle χ, as shown here, is a negative quantity because the projection of $\nabla\bar{H}_z$ on the unit vector $\hat{\theta}$ of the polar coordinates (r, θ) is negative.

$$\dot{r} = dr/dt = -V\cos(\phi - \chi), \tag{36a}$$

$$\dot{\theta} = d\theta/dt = +(V/r)\sin(\phi - \chi), \tag{36b}$$

where χ is the angle between $\nabla \bar{H}_z$ and $\hat{r}$ a unit vector in the radial direction. Division of eq. (36b) by (36a) yields the velocity ratio

$$r\dot{\theta}/\dot{r} = v_\theta/v_r = \tan(\chi - \phi). \tag{37}$$

Consider now the special case $\chi = 0$. This implies $\partial \bar{H}_z/\partial\theta$ vanishes. Since ϕ is a constant, eq. (37) may be integrated to furnish the trajectory

$$\theta = \theta_o + \tan\phi \ln(r_o/r), \tag{38}$$

where r_o and θ_o specify the position at some reference time, $t = t_o$. The path is a logarithmic spiral; it is unique in that it is independent of the radial behavior of $\bar{H}_z$. The dependence of $\bar{H}_z$ on r will, however, determine the position and velocity along the spiral as a function of the time by inversion of the quadrature

$$t = t_o + \sec\phi \int_r^{r_o} \frac{ds}{V(s)}. \tag{39}$$

Equation (39) follows from eq. (36a). When the functional dependence of $\bar{H}_z$ on r is sufficiently simple, eq. (39) may be integrated and inverted. For example, when $\bar{H}_z = H_o(r/d_b)^2$, $V = v(r) = (r - \alpha\, d_b)/\tau_o$. Here H_o is a constant, $\alpha = 4H_c/\pi H_o$, $\tau_o = d_b/\mu_w H_o$, and $\mu_w = 2\mu_b$ is the wall mobility. Integration of eq. (39) furnishes

$$r = \alpha d_b + (r_o - \alpha d_b)\exp(-\lambda t), \tag{40}$$

where $\lambda = \tau_o^{-1}\cos\phi$ and $t_o = 0$. Figure 77 shows several trajectories. Bubble motion starts at a point $r = r_o$ marked $t/\tau_o = 0$. From there, bubbles travel along logarithmic spiral paths terminating at the origin. The numerals along the various paths give bubble position in units of t/τ_o for $\alpha d_b/r_o = 0.5$. When $H_c = 0$ and $0 \leqq \phi < 90°$, all bubbles spiral into the origin arriving there as $t \to \infty$.

Two special cases of bubble motion near a circular opening allow integration of eq. (35) by quadrature. For both, $S = \phi = 0$. Consider a bubble located on the y axis at $t = 0$. Using polar coordinates, one finds immediately $\theta =$ constant since $\phi = 0$ and $\chi = 0$ or π; see eq. (36). The radial equation is solved by the quadrature

$$t = \tau \int_0^{r/a} \frac{d\sigma}{|\bar{\mathcal{P}}(\sigma)| - \beta} \tag{41}$$

for motion starting at the origin. Here, $\tau = a^2/\mu_b d_b J_o$ and $\beta = 8H_c a/\pi d_b J_o$ with H_c in oersteds and J_o in mA/μm. Numerical evalua-

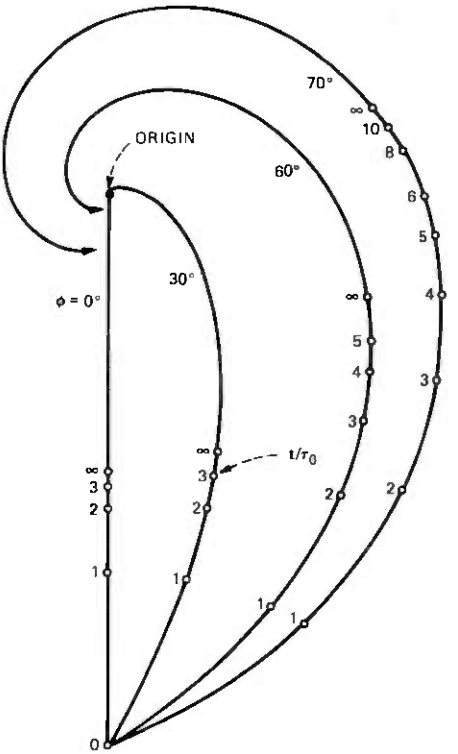


Fig. 77—Logarithmic spiral trajectories described by eq. (38). Bubbles spiral into the origin along these paths when $H_c = 0$. For $H_c > 0$, motion is along the same paths but bubbles do not reach the origin. The timing markers, in units of t/τ_o, show positions for $\alpha d_b/r_o = 0.5$.

tion of eq. (41) for $a = d_b = 2\mu$m, $H_c = 1$ Oe, $\mu_b = 1$ μm/μs-Oe, $J_o = 1$ mA/μm, $z_s/a = 0.06$, and $h/a = 0.8$ shows that 0.70 μs are required for motion from $r = 0$ to $r = 0.8a$. For these parameters, the characteristic time τ equals 2 μs. Arrival at $r/a = 0.8$ completes approximately 90 percent of the possible radial motion.

The second special case occurs when a bubble resides at the radial minimum of $\mathcal{H}$ at $t = 0$. Enter eq. (36) with $\phi = 0$ and $\chi = -\pi/2$. Thus, $r = r_e$ = constant, and the equilibrium value r_e is found from the solution of $\mathcal{P}(\mu) = \partial\mathcal{H}/\partial\mu = 0$ with $\mu_e = r_e/a$. As may be seen from Fig. 72, $r_e/a \approx 1$ for small z_s/a. Equation (36b),

$$\dot{\theta} = V/r_e = \tau_1(\cos\theta - \beta_1), \tag{42}$$

where $\tau_1 = a^2\mu_e^2/\mu_b d_b J_o |\mathcal{H}_e|$, $\beta_1 = 8H_c a\mu_e/\pi d_b J_o |\mathcal{H}_e|$, and $\mathcal{H}_e = \mathcal{H}(\mu_e)$ is the value of $\mathcal{H}$ at the equilibrium radius, has the solution

$$t = \tau_1 \csc\theta_m \ln\left[\frac{\tan(\theta_m/2) + \tan(\theta/2)}{\tan(\theta_m/2) - \tan(\theta/2)}\right] \tag{43}$$

for a bubble at $\theta = 0$ when $t = 0$. The angle θ_m is attained as $t \to \infty$ and satisfies the condition $\cos \theta_m = \beta_1$. For the numerical values used in the previous example together with $\mu_e = 1.00$ and $|\bar{\mathcal{H}}_e| = 4.56$, one finds $\theta_m = 56.0°$ and the time needed for travel from $\theta = 0$ to $\theta = 0.9\theta_m$ is 1.48 μs.

10.5 Discussion

The graphs of Figs. 65 to 69 apply to an infinite number of conducting stripes. As shown in these figures, $\bar{\mathcal{H}}_s = 0$ at the origin, a conductor center, and rises to a maximum at the right conductor edge. The symmetry of $\bar{\mathcal{H}}_s$ about a conductor edge for $s = 4a$ together with the periodicity and antisymmetry expressed by eq. (16) completes the description of $\bar{\mathcal{H}}_s$ for one period. See Fig. 8 for a typical spatial variation of $\bar{\mathcal{H}}_s$ over several periods.

When the number of conducting stripes is finite, $\bar{\mathcal{H}}_s$ still has the wavelike behavior per period as described above. There are, however, essential differences. The function $\bar{\mathcal{H}}_s$ is not periodic. In addition, the wavelike behavior of $\bar{\mathcal{H}}_s$ is superimposed on a gradual change in the mean value of the bias field. This change in bias is due to the current $I = 2an_sJ_o$ flowing in the n_s stripes. The magnitude of the effect can be estimated, using eq. (15), for a single stripe of width $(n_s - 1)s + 2a$. To illustrate the quality of the estimate, consider $n_s = 11$ and $s = 4a$. Choose the origin at the middle of the center stripe and calculate $\bar{\mathcal{H}}_s$ at the center of each stripe and at the corresponding abscissa for a single stripe of width $(n_s - 1)s + 2a = 42a$. Then, for $\alpha = 1$, the numerical results are shown in Table VI. Here, $\bar{\mathcal{H}}_1$ is the normalized intensity for the single stripe. In the design of current-access devices, the shift in local bias must be taken into account. By returning the drive current through a nearby stripe of comparable width, this shift in local bias can be reduced.

To develop current-access propagation structures, large-scale models were made and the magnetic fields due to current flow in them probed with a coil. See Section III. One structure was a stripe pattern which served to calibrate the coil in the following way. The coil of radius r_o and height h was placed over a central stripe with the coil center

Table VI

Stripe Index n	Abscissa x/h	$\lvert\bar{\mathcal{H}}_s\rvert$	$\lvert\bar{\mathcal{H}}_1\rvert$
0	0	0	0
±1	±4.0	0.367	0.385
±2	±8.0	0.760	0.802
±3	±12.0	1.218	1.293
±4	±16.0	1.829	1.995
±5	±20.0	2.928	3.582

directly over the stripe edge. This location permitted maximum flux linkage. One can now calculate the flux threading the coil from

$$\phi = \int B\, dA = 4J_o \int_0^{r_o} \mathscr{H}_s(u)\, du \int_0^{(r_o^2 - u^2)^{1/2}} dv, \tag{44}$$

where the u-v coordinate system has its origin at the coil center with u parallel to x and v parallel to y. The integration proceeds over the first quadrant of the coil cross section; hence, the multiplier of 4. Finally, eq. (44) will give the flux in maxwells when the current density is expressed in A/mm and the coil radius in cm. Numerical integration over u utilized the parameter values $r_o = 0.1$ in., $h = 0.2$ in., and $z_s = 0.025$ in. Using the stripe width $2a = 0.5$ in., one gains entry to Fig. 66 with $h/a = 0.8$ and $z_s/a = 0.1$ to determine the values of $\mathscr{H}_s(u)$. Numerical quadrature yields $\phi = (6.183J_o/\pi)\pi r_o^2$. The result is written in this form because division by πr_o^2 provides the flux density averaged throughout the coil volume. For a sinusoidal drive $J_o = J_m \sin \omega t$, the rms voltage induced in the coil is

$$E_{\text{rms}} = 6.183 N_t \omega r_o^2 J_m \times 10^{-8}/\sqrt{2} = 2.1 \text{ mV} \tag{45}$$

for $N_t = 100$ turns, $\omega/2\pi = 30$ kHz, and $J_m = 1/25.4$ A/mm. This result compares favorably with the measured value of 2.0 mV. With the coil calibration in hand, one can assign absolute field values to voltage measurements from the large-scale models of proposed propagation structures. Apart from a slight increase in labor associated with a two-dimensional numerical integration over the coil cross section, one can also use eq. (25) for coil calibration. This has been done, and the results show a comparably small difference between the measured and calculated coil voltage.

XI. MISCELLANEOUS

This section contains items that do not seem to fit elsewhere. These include a three-level conductor circuit, a single level circuit with a lattice of etched islands, a deep implant layer for hard bubble suppression, and an analysis of bubble-bubble interactions.

11.1 Three-level conductor circuits

Advances in processing techniques will ultimately allow fabrication, at high yield, of structures with a complexity unattainable today. One structure that could become practical is the three-level circuit we discuss at this time.

Bias shift due to current flow in a conductor and its consequences was covered in Section III and, in particular, in Fig. 9. By using three-conductor levels and returning the current of the excited level through

the remaining two, one can ease the bias-shift problem and utilize for storage the full conductor width. Another advantage is simpler electronics; only unipolar drivers are needed.

Two possible propagation structures are shown in Fig. 78. Circuit (a) has slots side by side in each of the three levels. The number of + and − signs indicates the relative pole strengths at the ends of the slots when level 1 is energized with two units of current, half of which returns through each of the remaining levels. The wavelike curve sketched below is an approximation to H_z near the lower ends of the slots. If the levels are excited in turn with pulse currents I_j in the sequence [· · · 123 · · ·], the wave will move to the right and bubble propagation will be to the right. Interchange any two drives in the sequence, such as [· · · 213 · · ·], and bubbles move left. Circuit (b) with in-line oval holes in the three levels generates the indicated relative pole strengths when level 1 is excited. Here also, the sequence

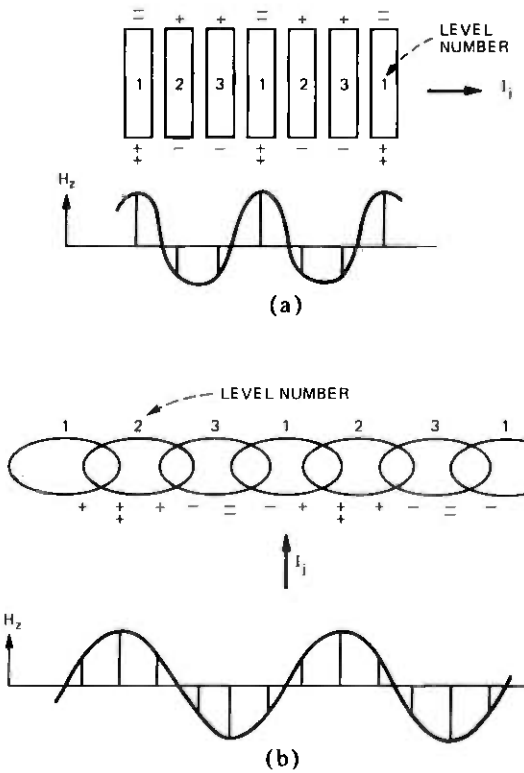


Fig. 78—Two types of three-level circuits are shown: in-line (a) and transverse (b) propagation. In either case, the drive current is returned through the remaining levels to provide z-field cancellation.

[...123...] will cause bubble motion to the right, and the reversal of any two drives in the sequence, to the left.

11.2 Single-level lattice-like circuits

In most single-level conductor circuits, a rather precise fit of the bubble with the feature which produces the static offset field is required. In other words, these circuits cannot tolerate a change in bubble size. A lone exception is that of Dekker in which permalloy points driven by an in-plane field generate the offset field. Other alternatives are possible. From the magnetostatic equivalence of bubbles and cylindrical holes etched in a garnet layer (craters), one can deduce that, at a distance, craters repel bubbles. Therefore, an array of craters can provide interstitial sites that precisely position bubbles. Bubble motion would then be somewhat analogous to electrons jumping from trap to trap in a semiconductor. Devices based on this principle are under study.

11.3 Deep implantation for hard bubble suppression

Data on the performance of implanted and nonimplanted films in dual-conductor devices were presented in Figs. 32 to 35. In particular, Figs. 32 and 34 point out that the minimum drive currents are higher when implanted films are used. This we attribute to a reduction in the effective drive fields because of the shielding effect of implanted layers. If the garnet is implanted at the garnet-substrate interface, this shielding will be eliminated. Experiments with a 2.25-μm thick YSmLuCaGe garnet implanted at 2×10^{16} with H^+ at 300 Kev confirmed this expectation. We anticipate that deep implantation will be especially useful if the garnet layers are less than a micron thick since the buried layer will precisely determine bubble height and, in addition, decouple bubbles from the sometimes troublesome substrate-garnet interface.

It is known that an in-plane field can be used to convert bubbles into the $S = 0$ state; however, the field must be very large if used with small bubble garnets. An analysis by Okabe[49] suggests that the magnitude of this field will be reduced with a very thick implant layer. A thick implant layer is practical only if it is buried.

11.4 Bubble-bubble interactions

We have already stressed the point that some conductor-access circuits can accommodate a wide range of bubble sizes. There is no "gap," in the sense of the permalloy circuits, to be overcome. In an 8-μm period, dual-conductor circuit, for example, bubbles ranging from 4 μm to less than 0.8 μm have been propagated. This tolerance to bubble size is especially useful when bubble-bubble interactions are a concern.

Bubble-bubble interaction imposes the following constraints on the circumstances under which a pair of adjacent bubbles in a track are stable in position when not driven and move in step when driven. In the absence of any drive excitation, the condition

$$H_c \geq 6\pi^2 M_s h r_o^3/\lambda^4, \tag{46}$$

where r_o is the bubble radius assures positional stability. That is, the mutual-interaction field gradients are just balanced by the coercivity H_c. During propagation, the trailing bubble must be driven to overcome both its coercivity and an interaction gradient from the leading bubble. The addition of this interaction is equivalent to a doubling of H_c. The drive current must therefore generate a gradient $\partial H_z/\partial x$ given by

$$2r_o \partial H_z/\partial x \geq 16 H_c/\pi. \tag{47}$$

It follows that, since $J \sim \partial H_z/\partial x$, $P/\text{bit} \sim (H_c r_o)^2$. From (46), it is seen that $r_o \sim H_c^{1/3}$ so

$$P/\text{bit} \sim H_c^{4/3}. \tag{48}$$

Equation (48) tells us that low-coercivity garnets are preferable to high-coercivity garnets since, with their use, the power dissipation will be reduced in spite of the interaction constraint (46).

XII. SUMMARY AND CONCLUSIONS

The purpose of this paper was to introduce new, very high speed, conductor-access devices where apertured conducting sheets replace the serpentine conductors of conventional conductor devices and the rotating-field coils of field-access devices. When apertured sheets are used, the bias field perturbations are highly localized, and therefore the drives can be ranged over wide limits. We have also shown that the impedance of the conducting layers can be made to match that of drivers operated from low-voltage (~5V) power supplies. The design of single, dual, and triple level conductor circuits was discussed.

All the functions for a shift register or a major-minor chip were covered, although some of the designs have not been optimized. For example, we have not as yet detected at a megahertz stepping rate. However, RF detection does look promising, and the redesign of both the detection electronics and the detector is in progress, and the results will be reported later.

The material requirements for the current-access device technology have been discussed. The high-speed performance inherent in this technology places more emphasis on the material parameters of bubble mobility and dynamic coercivity. Using the $(YLuSmCa)_3(FeGe)_6O_{12}$ system as a starting point, a technique has been presented to exercise

the bubble mobility while maintaining adequate uniaxial anisotropy. Compositions in the three systems, $(YLuSmCa)_3(FeGe)_5O_{12}$, $(LaLuSuCa)_3(FeGe)_5O_{12}$, and $(LaLuSm)_3(FeGa)_5O_{12}$, have been compared and the limitations found in the systems discussed. A simplified film composition in the $(LaLuSm)_3(FeGa)_5O_{12}$ systems is observed that exhibits 1.7-μm diameter bubbles along with high mobility and low dynamic coercivity.

Plasma etching has been used to define circuit features in the AlCu films. The anisotropic etch characteristics of the plasma etch process have produced metal patterns dimensionally identical to the photolithographically defined etch masks. Coupled to the high material-etch rate selectivity, the plasma etch process appears to be ideal for producing the very small period devices that current-access technology promises. Circuits have also been successfully fabricated using Au conductors patterned by ion milling technology; for Al conductors, however, the plasma etching has been the more practical etch method. Thin, conformal, insulating layers of SiO_2 have been successfully deposited by plasma deposition at low substrate temperatures, and it appears that this process is particularly suited for fabricating conductor-access devices.

Current-access and field-access devices differ, of course, in many ways. In terms of analytical complexity, the nonlinear permalloy-bubble interactions of the field-access scheme have been removed in current-access devices because permalloy is not used. One may anticipate, therefore, simpler, more frequent, and more accurate analyses in this new technology.

It is too early to conclude that conductor-access devices will become the technology of the next generation bubble devices. In this context, we include all the conductor-access devices now under development here and elsewhere. It is possible, for example, that the apertured sheet drive approach applied to the bubble-lattice file[7] will make that structure more attractive. There are some features that only conductor-access devices possess. If high speed is a necessity, as in video-based systems, conductor circuits are the only option available to the device designer at present. The anticipated small package size will be significant in applications for which space is a premium. The ability to run with logic-level power supplies is another plus for conductor devices. Finally, conductor circuits will require fewer precision analog control circuits since the critical control function to rotating field timing has been eliminated.

XIII. ACKNOWLEDGMENTS

Many people have contributed to this project. The authors thank W. Biolsi, W. L. Swartzwelder, and L. C. Luther for the growth of

garnet films, R. Wolfe for many informative discussions on ion implantation, G. P. Vella-Coleiro, R. D. Pierce, and W. Venard for material characterization, W. E. Hess for test electronics, J. A. Morrison for mathematical advice, and J. L. Blue for fruitful discussions about numerical analysis and programming.

We would also like to acknowledge the efforts of D. L. Fraser, Jr. and D. N. Wang who plasma-etched many of our early circuits, W. A. Johnson for fabricating Au conductor circuits by ion milling, R. O. Miller for constructing our plasma etch station, and R. S. Wagner for foresight and advice in developing the plasma etch process, and J. L. Smith and J. E. Geusic for words of encouragement during moments of frustration. Finally, special thanks should be given to R. T. Anderson for device-coding and to our patient typist, A. L. Becker.

XIV. MEMORIAL

This paper is dedicated to the memory of Edward M. Walters, a close friend and pioneer in magnetic bubbles.

REFERENCES

1. A. H. Bobeck, "Properties and Device Applications of Magnetic Domains in Orthoferrites," B.S.T.J., *46*, No. 8 (October 1967), pp. 1901–1925.
2. A. J. Perneski, "Propagation of Cylindrical Magnetic Domains in Orthoferrites," IEEE Trans. on Magnetics, *MAG-5*, No. 3 (September 1969), pp. 554–557.
3. R. Wolfe, J. C. North, W. A. Johnson, R. R. Spiwak, L. J. Varnerin, and R. F. Fisher, "Ion Implanted Patterns for Magnetic Bubble Propagation," AIP Conf. Proc., No. 10 (1972), pp. 339–343.
4. A. H. Bobeck and I. Danylchuk, "Characterization and Test Results for a 272K Bubble Memory Package," IEEE Trans. on Magnetics, *MAG-13*, No. 5 (September 1977), pp. 1370–1372.
5. J. A. Copeland, J. G. Josenhans, and R. R. Spiwak, "Circuit and Module Design for Conductor-Groove Bubble Memories," IEEE Trans. on Magnetics, *MAG-9*, No. 3 (September 1973), pp. 489–492.
6. T. J. Walsh and S. H. Charap, "Novel Bubble Drive," AIP Conf. Proc. No. 24 (1974), pp. 550–551.
7. O. Voegeli, B. A. Calhoun, L. L. Rosier, and J. C. Slonczewski, "The Use of Bubble Lattices for Information Storage," presented at the 1974 Conf. Magnetism and Magnetic Materials.
8. H. L. Hu, T. J. Beaulieu, D. W. Chapman, D. M. Franich, G. R. Henry, L. L. Rosier, and L. F. Shew, "1K Bit Bubble Lattice Storage Device: Initial Tests," J. Appl. Phys., *49*, No. 3 (March 1978), pp. 1913–1917.
9. E. H. L. J. Dekker, K. L. L. van Mierloo, and R. de Werdt, "Combination of Field and Current Access Magnetic Bubble Circuits," IEEE Trans. on Magnetics, *MAG-13*, No. 5 (September 1977), pp. 1261–1263.
10. J. A. Copeland, J. P. Elward, W. A. Johnson, and J. G. Ruch, "Single-Conductor Magnetic-Bubble Propagation Circuits," J. Appl. Phys., *42*, No. 4 (March 1971), pp. 1266–1267.
11. T. H. O'Dell, *Magnetic Bubbles*, London, England: MacMillan, 1974.
12. W. J. Tabor, A. H. Bobeck, G. P. Vella-Coleiro, and A. Rosencwaig, "A New Type of Cylindrical Magnetic Domain (Hard Bubble)," AIP Conf. Proc., No. 10 (1972), pp. 442–457.
13. J. C. Slonczewski, J. C. Malozemoff, and O. Voegeli, "Statics and Dynamics of Bubbles Containing Bloch Lines," AIP Conf. Proc., No. 10 (1972), pp. 458–477.
14. R. W. Patterson, "Annihilation of Bloch Lines In Hard Bubbles," AIP Conf. Proc., No. 24 (1974), pp. 608–609.
15. H. L. Hu and E. A. Giess, "Hard Bubble Suppression and Controlled State Gener-

ation of One Micron Bubbles in Ion-Implanted Garnet Films," IEEE Transactions on Magnetics, *MAG-11*, No. 5 (September 1975), pp. 1085–1087.
16. V. M. Benrud, G. L. Forslund, M. M. Hanson, R. L. Horst, A. D. Kaske, J. A. Kolling, D. S. Lo, M. J. Nordstrom, H. N. Oredson, W. J. Simon, C. H. Tolman, and E. J. Torok, "Oligatomic Film Memories," J. Appl. Phys., *42*, No. 4 (March 1971), pp. 1364–1373.
17. H. J. Levinstein, S. J. Licht, R. W. Landorf, and S. L. Blank, "Growth of High-Quality Garnet Thin Films from Supercooled Melts," Appl. Phys. Lett., *19*, No. 11 (December 1971), pp. 486–488.
18. S. L. Blank and J. W. Nielsen, "The Growth of Magnetic Garnets by Liquid Phase Epitaxy," J. Crystal Growth, *17* (1972), pp. 302–311.
19. S. L. Blank, J. W. Nielsen, and W. A. Biolsi, "Preparation and Properties of Magnetic Garnet Films Containing Divalent and Tetravalent Ions," J. Electrochem. Soc., *123*, No. 6 (June 1976), pp. 856–863.
20. S. L. Blank, B. S. Hewitt, L. K. Shick, and J. W. Nielsen, "Kinetics of LPE Growth and its Influence on Magnetic Properties," AIP Conf. Proc., No. 10, (1972), pp. 256–270.
21. E. A. Giess, B. E. Argyle, D. C. Cronemeyer, E. Klokholm, T. R. McGuire, D. F. O'Kane, T. S. Plaskett, and V. Sadagopan, "Europium-Yttrium Iron-Gallium Garnet Films Grown by Liquid Phase Epitaxy on Gadolinium Gallium Garnet," AIP Conf. Proc., No. 5, (1971), pp. 110–114.
22. R. L. Barnes, *Advances in X-Ray Analysis*, Vol. 15, K. J. Heinrich, C. F. Barrett, J. D. Newkirk and C. D. Rund, eds., New York: Plenum Press, 1972, pp. 330.
23. R. D. Pierce, "Magnetic Characterization of Bubble Garnet Films in an LPE Growth Facility," J. Crystal Growth, 27 (1974), pp. 299–305.
24. G. P. Vella-Coleiro and W. J. Tabor, "Measurement of Magnetic Bubble Mobility in Epitaxial Garnet Films," Appl. Phys. Lett., *21*, No. 1 (July 1972), pp. 7–8.
25. R. C. LeCraw and R. D. Pierce, "Temperature Dependence of Growth-Induced Magnetic Anisotropy in Epitaxial Garnet Films by Resonance Techniques," AIP Conf. Proc., No. 5, (1971), pp. 200–204.
26. A. H. Bobeck, P. I. Bonyhard, and J. E. Geusic, "Magnetic Bubbles—An Emerging New Memory Technology," Proc. IEEE, *63*, No. 8 (August 1975), pp. 1176–1195.
27. S. L. Blank, R. Wolfe, L. C. Luther, R. C. LeCraw, T. J. Nelson, and W. A. Biolsi, "Design and Development of Single-Layer, Ion-Implantable Small Bubble Materials for Magnetic Bubble Devices," Proc. 24th Annual Conf. on Mag. and Mag. Matl., J. Appl. Phys., March, 1979.
28. M. Kestigian, A. B. Smith, and W. R. Bekebrede, "$(YSmLu)_3(FeGa)_5O_{12}$ for 1 to 3 μm-Diameter Bubble Devices," J. Appl. Phys., *49*, No. 3 (March 1978), pp. 1873–1875.
29. G. P. Vella-Coleiro, F. B. Hagedorn, S. L. Blank, and L. C. Luther, "Coercivity in 1.7-μm Bubble Garnet Films," Proc. 24th Annual Conf. on Mag. and Mag. Matl., J. Appl. Phys., March, 1979.
30. E. R. Czerlinsky and W. G. Field, "Magnetic Properties of Ferromagnetic Garnet Single Crystals," Solid State Phys. Electronics Telecommuns., Proc. Intern. Conf., Brussels, *3* (1958) (Pub. 1960), pp. 488–499.
31. W. A. Johnson, private communication.
32. A. K. Sinha, and T. T. Sheng, "The Temperature Dependence of Stresses in Aluminum Films on Oxidized Silicon Substrates," Thin Solid Films, *48*, No. 2 (February 1978), pp. 117–126.
33. B. J. Roman, private communication.
34. D. K. Rose, "Planar Processing for Magnetic Bubble Devices," IEEE Trans. on Magnetics, *MAG-12*, No. 6 (November 1976), pp. 618–621.
35. C. J. Mogab, and W. R. Harshbarger, "Plasma Processes Set to Etch Finer Lines with Less Undercutting," Electronics (August 31, 1978), pp. 117–121.
36. C. M. Melliar-Smith and C. J. Mogab, in *Thin Film Processes*, J. L. Vossen and W. Kern, eds., New York: Academic Press, 1978, pp. 497–555.
37. H. J. Levinstein, D. N. Wang, Patent Pending.
38. C. J. Mogab and F. B. Alexander, private communication.
39. R. M. Goldstein and J. A. Copeland, "Permalloy Rail-Cylindrical Magnetic Domain Systems," J. Appl. Phys. *42*, No. 6 (May 1971), pp. 2361–2367.
40. R. M. Goldstein, M. Shoji, and J. A. Copeland, "Bubble Forces in Cylindrical Magnetic Domain Systems," J. Appl. Phys., *44*, No. 11 (November 1973), pp. 5090–5095.
41. D. E. Gray, Ed., *American Institute of Physics Handbook*, New York: MacMillan, 1963, p. 5–26.
42. H. B. Dwight, *Tables of Integrals and Other Mathematical Data*, fourth ed., New York: MacMillan, 1961, pp. 12–13.

43. L. M. Milne-Thomson, *Theoretical Hydrodynamics*, fifth ed., New York: MacMillan, 1968, p. 158.
44. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, New York: Academic Press, 1965, pp. 81–85.
45. J. L. Blue, "Automatic Numerical Quadrature," B.S.T.J., *56*, No. 9 (November 1977), pp. 1651–1678.
46. H. W. Richmond, "On the Electrostatic Field of a Plane or Circular Grating Formed of Thick Rounded Bars," Proc. London Math. Soc., *22* (April 1923), pp. 389–403.
47. A. A. Thiele, "Device Implications of the Theory of Cylindrical Magnetic Domains," B.S.T.J., *50*, No. 3 (March 1971), pp. 725–773.
48. T. Hsu, "Control of Domain Wall States for Bubble Lattice Devices," AIP Conf. Proc. No. 24, 1974, pp. 624–626.
49. Yoichi Okabe, "A Model to Predict the Upper Cap-Switch Field of Various Capping Layers," IEEE Trans. on Magnetics, *MAG-14*, No. 5 (September 1978), pp. 602–604.

Contributors to This Issue

Stuart L. Blank, B.S. (ceramic engineering), 1962, Alfred University, New York State College of Ceramics; M.S. (materials science), 1964, Ph.D. (materials science), 1967, University of California; Bell Laboratories, 1969—. Mr. Blank has been involved in investigating the growth of materials by liquid-phase epitaxy, in developing new materials for magnetic bubble device applications, and in transferring the epitaxial growth technology to Western Electric for production of magnetic bubble materials. He is also investigating crystal growth, phase transitions, and defects in oxide materials. At present, he is supervisor of the Epitaxial Materials and Processes group in the Electronic Materials and Processes Department. Member, American Ceramic Society, National Institute of Ceramic Engineers, American Association for Crystal Growth.

Andrew H. Bobeck, B.S. (electrical engineering), 1948, M.S. (electrical engineering), 1949, Purdue University; Bell Laboratories, 1949—. Mr. Bobeck has specialized in the development of magnetic components. His early work involved the first solid-state driven core memory and the twistor memory. His interest in magnetic logic and storage eventually led to the magnetic bubble concept. A significant contribution to magnetic bubble development was the discovery that garnet materials can be prepared with a growth-induced uniaxial anisotropy, an observation that led to the epitaxial garnet films in general use today. At present, he supervises an activity aimed at the development of very high density bubble devices. Distinguished Engineering Alumnus, 1968, Doctor of Engineering (honorary), 1972, Purdue University; Stuart Ballantine medal, 1973, Franklin Institute; co-recipient, 1975, Morris Liebmann award; National Academy of Engineering, 1975; Valdemar Poulsen Gold Medal, 1976, Danish Academy of Technical Sciences. Fellow, IEEE, Tau Beta Pi, Eta Kappa Nu. He has been awarded more than 115 patents.

A. Duane Butherus, B.S. (chemistry), 1961, Andrews University; Ph.D. (chemistry), 1967, Michigan State University; Bell Laboratories,

1967—. Mr. Butherus has done materials investigations in a range of fields including polymers, solid-state ion conductors, and other electrochemically related systems. His current interests are in plasma chemistry and bubble memory processing techniques.

Frank J. Ciak, Bell Laboratories, 1968—. Mr. Ciak has worked in the Device Design Group on the development of magnetic bubble memory devices and associated test equipment. He designed magnet-setting apparatus and in-process bubble device test stations currently in use at Western Electric, Reading, Pa.

Nicholas F. Maxemchuk, B.S.E.E., 1968, City College of New York; M.S.E.E., 1970, Ph.D. 1974, University of Pennsylvania; RCA Laboratories, 1968–1976; Bell Laboratories, 1976—. Since joining Bell Laboratories, Mr. Maxemchuk has been a member of the technical staff of the Electronic and Computer Systems Research Laboratory.

Carlo Scagliola, Dr. Ing. (Electronic Engineering) 1970, University of Pisa, Italy. Mr. Scagliola has been with CSELT (Centro Studi e Laboratori Telecomunicazioni), Turin, Italy since 1970, where he is presently Head, Speech Processing Department. He has been engaged in adaptive speech coding, in assessment of the quality of digitally coded speech, and in studies on automatic synthesis of the Italian language. He served as a consultant at Bell Laboratories from January 1977 through January 1978.

David H. Smithgall, B.S. (E.E.), 1967, M.S. (E.E.), 1968, Ph.D. (E.E.), 1970, Cornell University; Western Electric Engineering Research Center, 1970—. Mr. Smithgall has worked in various areas of process control and has published in the areas of optical waveguides, control, microprocessor applications, and optical fiber measurements. His current interests include the measurement, characterization, and control of optical fiber manufacturing processes. Member, IEEE.

Walter Strauss, B.E.E., 1948, College of the City of New York; Ph.D. (physics), Columbia University, 1961; Department of Electrical Engineering, College of the City of New York, 1948–1953; Bell Laboratories, 1960—. At Bell Laboratories, Mr. Strauss has been concerned with problems related to magnetoelasticity, piezoelectricity, and magnetic devices. Member, American Physical Society, Eta Kappa Nu, New York Academy of Sciences.

John Stuller, B.S.E.E., 1963, Massachusetts Institute of Technology; M.S.E.E.; 1967, University of Southern California; Ph.D., 1971, University of Connecticut; Northrop Nortronics Corp., 1963–1965; TRW Systems Group, 1965–1968; Perkin Elmer Corp. 1970–1971; Department of Electrical Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada, 1971–1974; Bell Laboratories, 1977—. At Bell Laboratories, Mr. Stuller has been engaged in analytical and simulation studies of both intraframe and interframe video coders. His primary area of interest is in the statistical theory of communications and its application.

Hans S. Witsenhausen, I.C.M.E., 1953 and Lic. Sc. Phys., 1956, Université Libre de Bruxelles; Ph.D., 1966, Massachusetts Institute of Technology; Electronic Associates, 1957–63; Electronic Systems Laboratory, 1963–65; Hertz Fellow, 1965–66; Bell Laboratories, 1966—. Mr. Witsenhausen is currently working in the Communications Analysis Research Department. He has published articles on hybrid computation, control theory, optimization, geometric inequalities, and other applied mathematical fields. He was Senior Fellow at Imperial College of Science and Technology, London, in 1972, visiting professor at M.I.T. in 1973, and Vinton Hayes senior fellow at Harvard University in 1975–76. Member, American Mathematical Society, IEEE, Sigma Xi.

Papers by Bell Laboratories Authors

BIOLOGY

Acoustical Properties of the Vocal Tract. M. M. Sondhi, Physics (November 1978), pp. 1–2.

Acoustic Characterization and Computer Simulation of the Air Volume Displaced by the Vibrating Vocal Cords: Lateral and Longitudinal Motion. J. L. Flanagan and K. Ishizaka, Proceedings of the International Symposium on Articulatory Modeling (July 1977), pp. 251–261.

A ^{13}C NMR Study of Gluconeogenesis in Isolated Rat Liver Cells. S. M. Cohen, S. Ogawa, and R. G. Shulman, Frontiers of Biological Energetics, Vol. 2, New York: Academic Press, 1978, pp. 1357–1364.

Early Experience with Sour and Bitter Solutions Increases Subsequent Ingestion. R. M. London, C. T. Snowdon, and J. M. Smithana, Physiology and Behavior (June 1979).

The Solution Conformation of Malformin A. A. E. Tonelli, Biopolymers, *17* (1978), pp. 1175–1179.

CHEMISTRY

Block Copolymer Theory 5. Spherical Domains. E. Helfand and Z. R. Wasserman, Macromolecules, *11* (September-October 1978), pp. 960–966.

^{13}C Chemical Shifts of the Polypropylene "Model" Compounds 3,5-Dimethylheptane and 3,5,7-Trimethylnone. A. E. Tonelli, Macromolecules, *12* (1979), pp. 83–85.

Calculated and Measured ^{13}C NMR Chemical Shifts of the 2,4,6-Trichloroheptanes and Their Implications for the ^{13}NMR Spectra of Poly(vinyl chloride). A. E. Tonelli, F. C. Schilling, W.H. Starnes, Jr., L. Shepherd, and I. M. Plitz, Macromolecules, *12* (1979), pp. 78–83.

CuCl: More Facts Generate More Thoughts on High Temperature Superconductivity. J. A. Wilson, Phil. Mag. B., *38* (1978), pp. 427–444.

The High Temperature Deposition and Evaluation of Phosphorus-or Boron-Doped Silicon Dioxide Films. A. C. Adams, C. D. Capio, S. E. Haszko, G. I. Parisi, E. I. Povilonis, and M. D. Robinson, J. Electrochem. Soc., *126* (February 1979), pp. 313–319.

Measuring the Phosphorus Concentration in Deposited Phosphosilicate Films. A. C. Adams and S. P. Murarka, J. Electrochem. Soc., *126* (February 1979), pp. 334–338.

The Nuclear Modulation Effect in Electron Spin Echoes for Complexes of Cu^{2+} and Imidazole with ^{14}N and ^{15}N. W. B. Mims and J. Peisach, J. Chem. Phys., *69*, No. 11 (December 1, 1978), pp. 4921–4930.

Parametric Counting of Samples Generating Complex X-Ray Spectra. P. F. Schmidt, J. E. Riley, Jr., and D. J. McMillan, Anal. Chem., *51* (February 1979), pp. 189–194.

Piezooptic and Electrooptic Constants of Crystals. D. F. Nelson, Landolt-Bornstein, New Series, Berlin: Springer-Verlag, 1979, pp. 495–505.

Preparation and Characterization of Tetra (2,4-Pentanedionato)-Hexa(Benzotriazolato)-Penta-Cu (II). J. H. Marshall, Inorg. Chem., *17* (December 1978), pp. 3711–3713.

X-Ray Absorption Studies of Metalloproteins. R. G. Shulman, Trends in Biochemical Sciences, *3*, No. 12 (December 1978), pp. N282–N283.

COMPUTING

Application of Multilinear Algebra in Optimal Regulation of Nonlinear Polynomial Systems. M. R. Buric, Proceedings of 16th Annual Allerton Conference on Communications, Control and Computing, *1* (October 4, 1978), pp. 655–663.

Development of the Law of Computer Software Protection. R. O. Nimtz, Journal of Patent Officer Society, *61* (January 1979), pp. 3–43.

IMSMON-A Low Overhead Monitor for IMS Systems. B. Domanski, 1978 Compact Measure Group 1X Conference Proceedings on the Management and Evaluation of Computer Performance (December 5, 1978), pp. 115–123.

Model for the User Services Layer of a Network Architecture. K. Coates, 1969 ACM Computer Science Conference (February 20, 1979), p. 38.

On-Line PWB Design Verification and Associated Data Structuring Problems. J. C. Foster and S. Pardee, Proceedings, International Conference, Interactive Techniques in Computer Aided Design, Bologna, Italy, *78 CH1289-8C* (September 21–23, 1978), pp. 265–271.

What Employers Look for When Hiring Computer Science Graduates. M. M. Irvine, Proceedings of 1978 Annual Conference Association for Computing Machinery, *1*, pp. 404–405.

ELECTRICAL AND ELECTRONIC ENGINEERING

Alpha Particles Tracks in Silicon and Their Effect on Dynamic MOS RAM Reliability. D. S. Yaney, J. T. Nelson, and L. L. Vaniskike, IEEE Trans. Electron Dev., *ED-26* (January 1979), pp. 10–16.

Defect Etch for (100) Silicon Evaluation. D. G. Schimmel, J. Electrochem. Soc., *126* (March 1979), pp. 479–483.

Design of an Integrated Circuit for the TIC Low-Power Line Repeater. P. C. Davis, J. F. Graczyk, and W. A. Griffin, IEEE J. of Solid State Circuits, *SC-14* (February 1979), pp. 109–120.

Digital Communications—The Silent (R) Evolution? M. R. Aaron, IEEE Communications Magazine, *17*, No. 1 (January 1979), pp. 16–26.

The Evolution of the Discrete Crystal Single-Sideband Selection Filter in the Bell System. T. H. Simmonds, Jr., Proc. IEEE, *67*, No. 1 (January 1979), pp. 109–115.

Fatigue Behavior of Flex Cables and Circuits. W. Engelmaier, Electronic Packaging and Production, *19* (February 1979), pp. 110–118.

A 50 MHz Phase and Frequency Locked Loop. R. R. Cordell, C. N. Dunn, J. B. Forney, and W. G. Carrett, 1979 ISSCC Digest of Technical Papers *22* (February 14, 1979), pp. 234–235.

Final State Structure in XPS. G. K. Wertheim, Electron Spectroscopy, Theory, Techniques and Application, Vol. 2, New York: Academic Press, 1978, pp. 259–284.

Foreword-Special Issue on Quantum-Electronic Devices for Optical Fiber Communications. T. Li, IEEE J. of Quantum Electronics, *QE-14* (November 1978), p. 790.

InGaAsP/InP Photodiodes: Microplasma-Limited Avalanche Multiplication at 1-1.3 m Wavelength. T. P. Lee, C. A. Burrus, A. G. Dentai, IEEE J. Quantum Electronics, *QE-15* (January 1979), pp. 30–35.

An Integrated PCM Encoder Using Interpolation PCM Decoder. B. A. Wooley, D. C. Fowlis, J. L. Henry, and C. E. Williams, IEEE J. Solid State Circuits, *SC-14*, No. 1 (February 1979), pp. 20–25.

P-N-P-N Optical Detectors and Light-Emitting Diodes. J. A. Copeland, A. G. Dentai, and T. P. Lee, IEEE J. Quantum Electronics, *QE-14* (November 1978), pp. 810–813.

Self-Contained LED-Pumped Single-Crystal. J. Stone and C. A. Burrus, Fiber and Integrated Optics, *2* (1979), pp. 19–46.

Semiconductor-Laser Self-Pulsing Due to Deep-Level Traps. J. A. Copeland, Electron. Lett. (December 7, 1978), pp. 809–810.

Signal-To-Noise Ratio Statistics for Nondispersive Fading in Radio Channels with Cross Polarization Interference Cancellation. N. Amitay, IEEE Transactions on Communication, *COM-27*, No. 2 (February 1979), pp. 498–502.

Slow Positron Emission from Metal Surfaces. A. P. Mills, P. M. Platzman, and B. L. Brown, Phys. Rev. Lett., *41* (1978), p. 1076.

A Stereo-Fiberscope with a Magnetic Interlens Bridge for Laryngeal Observation. O. Fujimura, T. Baer, S. Niimi, J. Acoust. Soc. Amer., *65*, No. 2 (1979), pp. 478–480.

Supplement to Literature on Digital Signal Processing. J. F. Kaiser and H. D. Helms, Special Publication of IEEE ASSP Society, IEEE, New York, *JH4686-2 ASSP* (1979), p. 237.

MATERIALS SCIENCE

An AES Study of the Surface Composition of Ion-Etched Iron-Chromium Alloys: Effect of Absorbed CO. R. P. Frankenthal and D. E. Thompson, J. Vacuum Sci. and Technol., *16* (January–February 1979), pp. 6–12.

Control of Substrate Tube Diameter During MCVD Preform Preparation. P. D. Lazay and W. G. French, Conference Digest (OSA—Topical Meeting on Optical Fiber Communication) (March 5, 1979), pp. 50–52.

Core Hole Screening in Lanthanide Metals. G. Creceline, G. K. Wertheim, and D. N. E. Buchanan, Phys. Rev. B, *18*, No. 12 (December 15, 1978), pp. 6519–6524.

Effect of pH on the Rate of Corrosion of Gold in Acid Sulfate Solutions. B. S. Duncan and R. P. Frankenthal, J. Electrochem. Soc., *126* (January 1979), pp. 95–97.

The Intermediate Valence State in Rare Earth Compounds; Bulk and Surface Manifestations. G. K. Wertheim, J. Electon. Spectr., *15* (January 1979), pp. 5–14.

The Kinetic Photochemistry of the Marine Atmosphere. T. E. Graedel, J. Geophys. Res., *84* (1979), pp. 273–286.

Mechanical Testing Standards as Developed by the Voluntary Consensus System: Present and Future. A. Fox, Golden Gate Metals and Welding Conference Abstracts, Session 23 (January 29, 1979).

Processing and Magnetic Properties of Low-Loss and High Stability Mn-Zn Ferrites. B. B. Ghate, Mater. Sci. Res., Process. Cryst. Ceram., *11* (1978), pp. 369–379.

Sealed Extrusion-A Coating Method for Photoprinting UV Curable Liquid Resin Films. G. B. Fefferman and T. V. Lake, Fourth Radiat. Int. Conf. (September 1978), pp. 1–18.

Variable Resolution Capability for Multichannel Filter Spectrometers. P. S. Henry, Rev. Sci. Instrum., *50* (February 1979), pp. 185–192.

GENERAL MATHEMATICS AND STATISTICS

Comment on Schruben and Margolin. C. L. Mallows, J. Amer. Statist. Assoc., *73* (September 1978), p. 520.

Complexity Results for Bandwidth Minimization. M. R. Garey, R. L. Graham, D. S. Johnson, and D. E. Knuth, SIAM J. Appl. Math., *34* (May 1978), pp. 477–495.

Factor Analysis and Principal Components: The Bilinear Methods. J. B. Kruskal, The International Encyclopedia of Statistics, The Free Press (1978), pp. 307–330.

Good and Optimal Ridge Estimators. R. L., Obenchain, The Annals of Statistics, *6* (September 1978), pp. 1111–1121.

A Note on Bisecting Minimum Spanning Trees. W. M. Boyce, M. R. Garey, and O. S. Johnson, Networks, *8* (1978), pp. 187–192.

Postscript to "Transformations of Data." J. B. Kruskal, International Encyclopedia of Statistics, New York: The Free Press, Vol. 2, 1978, pp. 1055–1056.

PHYSICS

Amorphous Metal Films by Getter-Sputtering at 25°K. J. J. Hauser, R. J. Schutz, and W. M. Augustyniak, Phys. Rev., *18B* (October 15, 1978), pp. 3890–3896.

Angle-Resolved Photoemission from Surface and Adsorbates. N. V. Smith, Journal de Physique, *39C4* (July 1978), pp. 161–168.

Diamagnetic Structure of Rb in Intense Magnetic Fields. N. P. Economou, R. R. Freeman, and P. F. Liao, Phys. Rev. A, *18* (December 1978), pp. 2506–2509.

The Diffraction of He Atoms at a Si(111) 7 × 7 Surface. M. J. Cardillo and G. E. Becker, Phys. Rev. Lett., *42* (February 1979), pp. 508–511.

Diffraction of Photoelectrons Emitted from Core Levels of Te and Na Atoms Adsorbed on Ni (001). D. P. Woodruff, D.Norman, B. W. Holland, N. V. Smith, H. H. Farrell, and M. M. Traum, Phys. Rev. Lett, *41* (October 16, 1978), pp. 1130–1133.

Dynamic Central Peaks and Phonon Interactions Near Structural Phase Transitions. P. A. Fleury and K. B. Lyons, Proc. of the Int'l. Conf. on Lattice Dynamics, Paris, France (1978), p. 731.

The Effect of Strain-Induced Band-Gap Narrowing on High Concentration Phosphorus Diffusion in Silicon. R. B. Fair, J. Appl. Phys. *50* (February 1979), pp. 860–868.

Effects of Atomic Order in α- and β-phase AgCd Alloys Studied by X-Ray Photoelectron Spectroscopy. G. Crecelius and G. K. Wertheim, Phys. Rev. B, *18*, No. 12 (December 15, 1978), pp. 6525–6530.

Effects of Magnetic Fields on Four Wave Mixing Processes in Atomic Vapors. N. P. Economou, R. R. Freeman, and G. C. Bjorklund, Optics Letters, *3* (December 1978), pp. 209–211.

Light Scattering Determinations of Dynamic Four Point Correlation Functions. P. A. Fleury, Proc. of the NATO Advanced Study Inst. on Correlation Functions and Quasi-particle Transactions in Condensed Matter, New York: Plenum Press, 1978, pp. 325–365.

The Magnetic Behavior of an S = ½ Amorphous Antiferromagnet. R. B. Kummer, R. E. Walstedt, S. Geschwind, V. Narayanamurti, G. E. Devlin, Phys. Rev. Lett, *40*, No. 16 (April 17, 1978), pp. 1098–1100.

Magnetic Field Quantum Beats in Two-Photon Free-Induction Decay. N. P. Economou and P. F. Liao, Optics Letters, *3* (November 1978), pp. 172–174.

Nitric Acid Vapor Line Parameters Measured By Co Laser Transmittance. L. A. Farrow, R. E. Richton, and C. P. Karnas, Appl. Opt., *18* (January 1, 1979) pp. 76–81.

Piezooptic and Elastooptic Coefficients. R. F. S. Hearmon and D. F. Nelson, Landolt-Börnstein, New Series (1979), pp. 505–551.

Real-Time Autocorrelation Interferometer. R. L. Fork and F. A. Beisser, Appl. Opt., *17* (November 15, 1978), pp. 3534–3535.

The Relation of Solitons to Polaritons in Coupled Systems. D. F. Nelson, Solitons and Condensed Matter Physics, Berlin: Springer-Verlag,1978, pp. 187–190.

Thermal Convection and Crystal Growth. P. G. Simpkins, and P. A. Blythe, Physics News in 1977, American Inst. of Physics Booklet, *R-28 1* (November 1977), pp. 52–53.

Time-Resolved Photoluminescence Spectroscopy in Amorphous As_2S_3. M. A. Bosch and J. Shah, Phys. Rev. Lett., *42*, pp. 118–121.

X-ray Photoemission Study of Ce-Pnitides. Y. Beer, R. Hanger, C. Zürdrer, M. Campagna, G. K. Wertheim, Phys. Rev. B, *18* (October 15, 1978), pp. 4433–4439.

PSYCHOLOGY

Subjective Detection of Differences in Variance from Small Samples. R. L. Fike and W. R. Ferrell, Organ. Behav. Hum. Perf., *22* (December 1978), pp. 262–278.

SYSTEMS ENGINEERING AND OPERATIONS RESEARCH

Communications Privacy. A. Gersho, IEEE Commun. Soc. Mag., *16* (November 1978), p. 2.

The Complexity of the Network Design Problem. D. S. Johnson, J. K. Lenstra, and A. H. G. Kan, Networks, *8* (1978), pp. 279–285.

D. C. Welcomes MAT. B. S. Brosius and G. H. Webster, Telephony (January 1, 1979), pp. 19–23.

Digital Techniques for Speech Communications: Some New Dimensions. J. L. Flanagan, Proceedings of the IEEE International Conference on Cybernetics and Society (September 1977), pp. 1–5.

Double Phase Matching Function. D. F. Nelson, J. Opt. Soc. Amer., *68* (December 1978), pp. 1780–1781.

A New Measurement Technique For Telephone Switching Systems. R. J. Jaeger, Jr. and R. C. Stone, Jr., Pacific Telecommunications Conference Proceedings (January 8, 1979), pp. 3C24–3C33.

The Use of MIL-STD-1050 to Control Average Outgoing Quality. B. S. Liebesman, J. Qual. Tech., *11*, No. 1 (January 1979), pp. 36–43.

Errata

F. R. K. Chung and F. K. Hwang, "On Blocking Probabilities for a Class of Linear Graphs," B.S.T.J., *57*, No. 8 (October 1978), pp. 2915–2925.

We acknowledge the help of M. Horgan of England who pointed out an error in our paper. On page 2918, we define

$$f(\beta_{\lfloor (t-1)/2 \rfloor}) = (1 - \alpha_{t/2})^{\beta_{\lfloor (t-1)/2 \rfloor}} \quad \text{if} \quad t \quad \text{is even.}$$

The correct definition should be

$$f(\beta_{\lfloor (t-1)/2 \rfloor}) = (1 - \alpha_{t/2-1}\alpha_{t/2}\alpha_{t/2+1})^{\beta_{\lfloor (t-1)/2 \rfloor}} \quad \text{if} \quad t \quad \text{is even.}$$

F.K.H.

F.R.K.C.

A. N. Netravali, F. W. Mounts, and K. A. Walsh, "Adaptation of Ordering Techniques for Facsimile Pictures with No Single Element Runs," B.S.T.J., *58*, No. 4 (April 1979), pp. 857–865.

Fig. 1 (continued), page 859, is actually Fig. 2 (continued), page 861, and Fig. 2 (continued) is Fig. 1 (continued).

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 6, July–August 1979
Printed in U.S.A.

B.S.T.J. BRIEF

A 30-GHz Scale-Model, Pyramidal, Horn-Reflector Antenna

By R. A. SEMPLAK

(Manuscript received January 24, 1979)

I. INTRODUCTION

In the early 1940s, the pyramidal horn-reflector antenna was invented[1] at Bell Laboratories, Holmdel, New Jersey. It is now in extensive use in the Bell System 4-, 6-, and 11-GHz, transcontinental, microwave, common-carrier, radio-relay network.[2] This antenna is a combination of a square electromagnetic horn and a reflector that is a section of the paraboloid of revolution. The apex of the square horn coincides with the focus of the paraboloid. The antenna is essentially a shielded, offset, parabolic antenna, so that very little of the energy incident on the reflector is reflected back into the feed to produce an impedance mismatch.

Radio interference from adjacent paths limits the number of converging routes of a common carrier microwave radio system, and in recent years, demands have been made to improve the sidelobe performance of the pyramidal horn-reflector antenna. In response, blinders (metallic extensions to the sidewalls of the horn) have been developed[3] which provide a degree of far sidelobe reduction, i.e., sidelobes beyond 35 degrees from the axis of the main beam.

To continue the investigation for ways of reducing the sidelobe levels of this type antenna, a scale model was built. The model has a numerically-machined, precision reflector, and the scaling factor is 7.5, which means that measurements made at a frequency of 30 GHz will represent the performance of a full-size antenna measured at a frequency of 4 GHz. This paper presents and discusses the measurements made on the scaled model at 30 GHz. Comparisons are made with data obtained (at a frequency of 4 GHz) by others on full-sized antennas.[4]

In the discussion that follows, it should be remembered (from Ref. 4) that longitudinal polarization and longitudinal plane indicate that the electric field in the aperture and the plane of antenna rotation, for radiation measurements, are aligned with the pyramidal horn axis, whereas transverse polarization and transverse plane indicate that the electric field in the aperture and plane of antenna rotation are perpendicular to the horn axis.*

II. DISCUSSION

A study of the historical data on the radiation characteristics of the horn-reflector antenna indicates that, since its inception in the 1940s, disagreements between measured data and theoretical values have existed. For example, in the transverse plane with longitudinal polarization, agreement between measured data and theoretical value exists only out to the first sidelobe. With transverse polarization, good agreement extends out to the fourth sidelobe.

Since the aperture of the horn-reflector antenna is illuminated with a dominant waveguide mode, the theoretically obtainable off-axis radiation levels[4] in both the transverse planes for transverse polarization are considerably higher than those obtained for longitudinal polarization in the transverse plane and transverse polarization in the longitudinal plane. For transverse polarization in the transverse plane, one has essentially the equivalent of an aperture with constant illumination across it. Hence, with the antenna mounted on its normal vertical position (i.e., the axis of the horn normal to the earth), the amplitude distributions across the aperture of the antenna are as follows: For longitudinal polarization, the electric field in the longitudinal direction is uniformly distributed across the aperture, and in the transverse direction the field is tapered to a low value at the edges; with transverse polarization, the electric field is uniform across the aperture in the transverse direction and is tapered to a low value at the edges in the longitudinal direction.

Of the many various possible contributors to high sidelobes, the two strongest contenders are: (*i*) higher-order modes being generated in the existing feedhorn, which was designed to accommodate the oversized circular waveguide† used in the system, and (*ii*) surface toler-

* As used in the microwave radio relay system, the horn-reflector antenna is mounted with the axis of the horn normal to the earth's surface. Hence, longitudinal and transverse polarizations could be called vertical and horizontal, respectively. However, the aperture field distribution for each polarization is different, and when the antenna is used as an earth station antenna for satellite communications, or as a radiometer, or simply to obtain radiation patterns in the longitudinal plane (the antenna is mounted on its side), aperture field distributions for so-called vertical and horizontal are now interchanged. To avoid this ambiguity, longitudinal and transverse polarizations which are referred to the axis of the horn are used.

† Higher-order modes can be excited in oversized waveguide; hence to avoid this condition when obtaining radiation patterns, one uses a transition from feedhorn to waveguide transducer.

ances of the parabolic reflector. To examine the first contender, two feedhorns were made, one a scaled version of the existing system feedhorn and the second a new design. This new feedhorn terminates the taper of the horn into the walls of rectangular K-band waveguide, thus ensuring the presence of only the dominant mode. In a sense, this is achieved by opening the parallel walls of the waveguide at a constant angle and extending them until they form the square aperture needed to mate with the rest of the antenna. This, of course, requires that the walls normal to the electric field open first. Carefully machined mandrels were made of both feeds over which copper was electroformed to produce the final models. An examination of the radiation characteristics of the scale-model antenna obtained from using both feeds indicated that the characteristics were essentially identical. Hence, the characteristics contained in the figures to be discussed were obtained using the scale version of the system feedhorn. In the figures, the theoretical values[4] are shown by a dashed line, and the measured data of the full-sized horn-reflector antenna by a broken line. The measured data for the scale-model are shown by the solid line.

For the transverse plane, Figs. 1 and 2 show the measured characteristics for longitudinal and transverse polarizations, respectively. As indicated here, the agreement between scale-model measurements and theory is remarkably good. From these two figures, one can see the improvement in sidelobe level by comparing the results with the broken line curve which indicates the measured response for the full-sized antenna. For example, in Fig. 1 the improvement is about 8 dB and increasing in the region of the fourth sidelobe and beyond, i.e., beyond 8 degrees from the on-axis position. In Fig. 2, a similar comparison shows a 4-dB improvement beginning around 10 degrees.

In the longitudinal plane, the measurements for longitudinal and transverse polarizations were obtained for the scale-mode antenna but are not included here since these patterns, like those of the full-size antenna, are in very good agreement with theory. Similarly, the cross-polarization patterns obtained on the scale model were like those of the full-sized antenna and, since both agree well with theory, are also not included here.

It is evident from these figures that the scale-model measurements are in remarkable agreement with theory. The presumption that the high sidelobe levels of the full-size antenna were due in part to high-order modes being generated by the feedhorn is not valid. This is emphatically demonstrated by the agreement with theory which does not accommodate the presence of higher-order modes. The remaining strong contender for these high sidelobe levels is the accuracy of the full-size parabolic reflector surface. An examination of the full-size structure shows the surface to consist of two half skins fastened to parabolic back ribs by means of 262 (No. 14, Type B) sheet metal

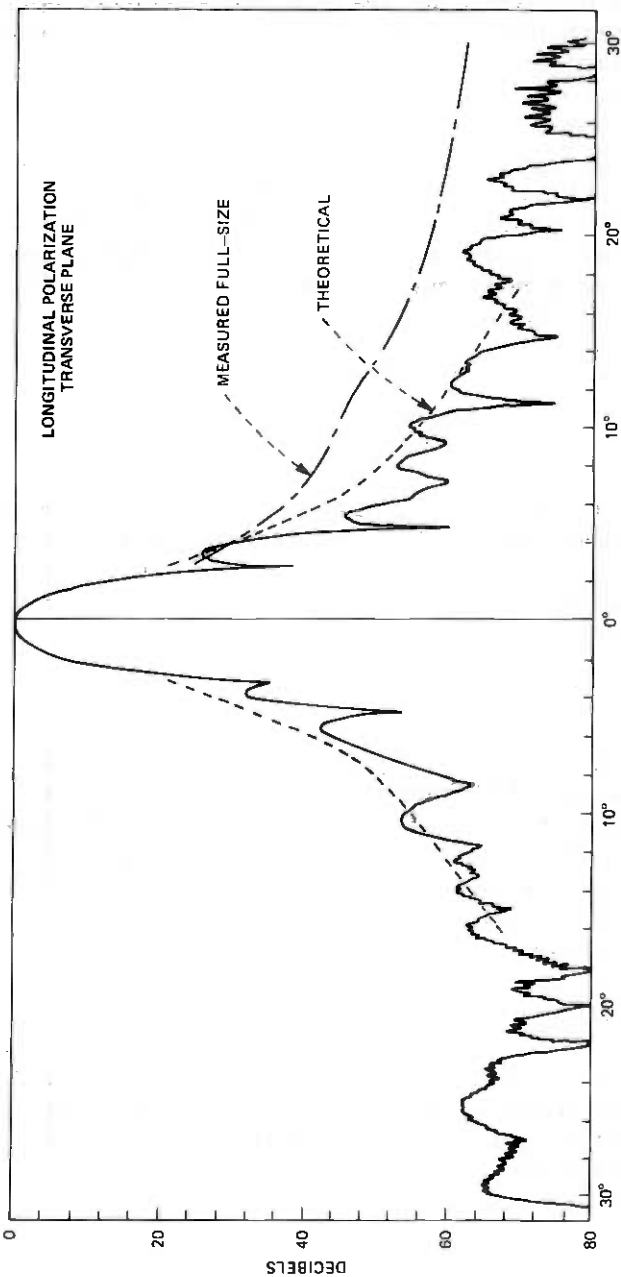


Fig. 1—Radiation pattern in the transverse plane for longitudinal polarization. The dashed line is theoretical, and the broken line represents measurements obtained on the full-size antenna.

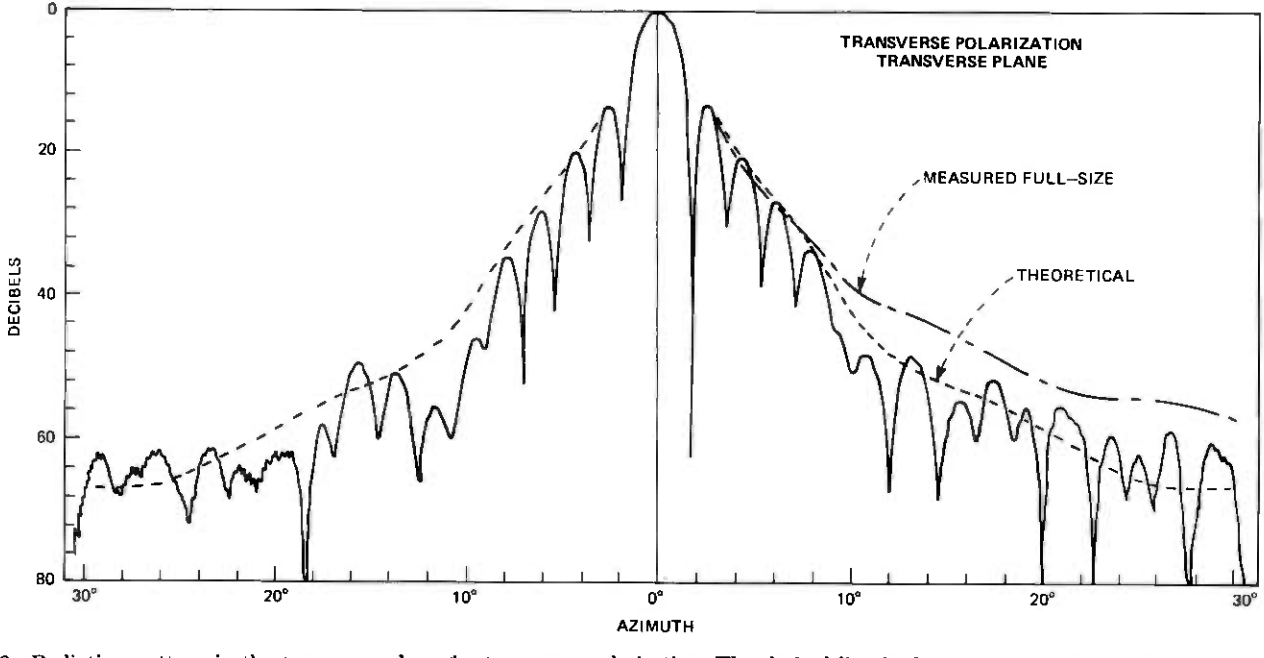


Fig. 2—Radiation pattern in the transverse plane for transverse polarization. The dashed line is theoretical, and the broken line represents measurements obtained on the full-size antenna.

screws and 120 (3/16-inch diameter Universal Head) rivets. Since these fasteners are distributed in a regular fashion, one could envision them as forming an array of scatterers which produce a low level, very broad beam and enhance the far sidelobe region. An investigation of this remaining aspect is not considered in the immediate future since it would entail a reproduction of these discontinuities on the precise surface of the scale-mode. Also, there are many interesting experiments waiting to be done on this scale model which can now serve as a test bed.

III. CONCLUSIONS

Data obtained on a scale-model version of the full-sized horn-reflector (used in the 4-, 6-, and 11-GHz microwave common-carrier, radio relay system) have been presented. From the data, one once again observes how well scale models depict the full-size world. But of most importance was the achievement of agreement between measurement and theory which produced the key finding that past disagreement between theoretical results and experimental data for a full-size antenna in the sidelobe region could not be due primarily to higher order modes being excited in the feedhorn, but could be attributed to surface imperfections of the reflector.

IV. ACKNOWLEDGMENTS

I would like to thank N. F. Schlaack, now retired, of these Laboratories, for sharing his measurements on the full-sized horn-reflector antenna with me. It is with appreciation that I acknowledge A. Quigley for providing the two scaled feedhorns, and the team of P. J. Schoellner and J. A. Swig, Jr., for their shepherding of the fabrication of the scale-model parabolic reflector. The successful results obtained are due to the attention these three gave to their respective parts.

REFERENCES

1. H. T. Friis and A. C. Beck, U. S. Patent 2,236,393.
2. R. W. Friis and A. S. May, "A New Broad-Band Microwave Antenna System," AIEE Trans., Part I, *77*, 1958, p. 97.
3. D. T. Thomas, "Analysis and Design of Elementary Blinders for Large Horn Reflector Antennas," B.S.T.J., *50*, No. 8 (November 1971), p. 2979.
4. A. B. Crawford, D. C. Hogg, and L. E. Hunt, "A Horn-Reflector Antenna for Space Communications," B.S.T.J., *40*, No. 4 (July 1961), p. 1095.