ELECTRICAL COMMUNICATION

ITT

VOLUME 39 • NUMBER 4 • 1964

ELECTRICAL COMMUNICATION

Technical Journal Published Quarterly by
INTERNATIONAL TELEPHONE and TELEGRAPH CORPORATION
320 Park Avenue, New York, New York 10022
President: Harold S. Geneen
Secretary: John J. Navin

CONTENTS

Volume 39 **1964** **Number 4**

EDITOR, Harold P. Westman
ASSISTANT EDITOR, Melvin Karsh

Subscription: $2.00 per year
50¢ per copy

This Issue in Brief

Digitrac Video Correlator for Binary Detection of Radar Targets—The video correlator, as used in the Digitrac air-traffic-control system, is essentially a digital tool to adapt radar video information for large-scale digital data processing. It suppresses noise, interference, and jamming while retaining true targets. Target detection is accomplished by converting the analog video signal into binary *1*'s and *0*'s, delivering a target signal to the display if the number of binary *1*'s (within a preset number of pulse periods) falls between low and high limits.

The binary information is stored in a magnetic core memory. A logic circuit examines the stored information for each range quantum to determine if the number of *1*'s falls within the limits that indicate true targets.

Detection signals from the correlator may be used to supply position coordinates to the computer and/or processed video to the display. The storage capability of the correlator permits data from several radars to be time-shared on one display. The correlator may also be used for narrow-band transmission of radar via telephone lines.

Digitrac Censor Data Processor—Censor, the **CEN**tral proces**SOR** of the Digitrac system, is a real-time parallel digital data processor with a random-access internal magnetic core memory. A common bus line provides 2-way communication with external units, including extra memories and other similar computing systems.

A priority list establishes the order in which calls for facilities are filled, and connection to the memory can be made within 10 microseconds without disturbing computations in the arithmetic unit. If the arithmetic unit must be interrupted, the program is suspended and later resumed when the higher-priority work has been completed.

The arithmetic unit operates at a clock frequency of 5.33 megahertz with a word length of 20 bits. The internal memory stores 2048 or 4096 words of 2×20 bits, with provision for expansion to 32 768 words. The common bus line handles 40-bit words at the rate of 166 000 words per second between any of 240 units that may be connected to it. A memory for numerical constants stores 64 words of 20 bits each.

Digitrac Display System for Air-Traffic Control—In the Digitrac system, radar data are developed in binary-digit form for transmission over the equivalent of a telephone channel to operate plan-position indicators in various control centers.

Part 1 of this paper describes the generation of radar azimuth and range data in digital form for display and for computer input. A series of 2048 equally spaced pulses is produced during each revolution of the radar antenna. A separate alignment pulse establishes the reference point in the pulse train.

A stable generator produces a 280-kilopulse frequency that is divided in a chain of binary counters. The resulting lower frequency is compared with the antenna pulse train, to which it is synchronized by action initiated by an error signal. A phase corrector compensates for nonuniform frequency correction between alignment pulses. The output pulse train is of considerably higher frequency than the antenna pulse train and indicates azimuth with high resolution and precision.

The azimuth train of digits goes to two modulators, each followed by a counter that sets a multiplication factor for the other modulator. This arrangement produces two pulse trains representing the sine and cosine of the azimuth. Two other modulators multiply the sine and cosine values by the radar range sweep length, which has the form of a certain number of pulses at a rate of 2.4 million pulses per second. Thus bursts of pulses are obtained for sine and for cosine in which the numbers of pulses are proportional to X and Y, respectively, and to the distance to the target.

Part 2 is concerned with the generation of symbols for identifying targets on the plan-position indicators as well as on auxiliary tabular displays. As several asynchronous radars may contribute to a single display, the

symbols are inserted during the interval between scans.

A symbol is reproduced from 16 dots on a 32-by-32-dot field, the X and Y coordinates for each dot being stored in binary notation in a memory. A printed card is provided for each character. The positions of the identification symbols are established by the target and move with it. Each character generator can produce 64 characters. Several generators may be used in large installations.

Part 3 describes the plan-position indicators and other auxiliary tabular displays, both using cathode-ray tubes. Either 12- or 16-inch (304.8- or 406.4-millimeter) cathode-ray tubes may be used as plan-position indicators. The digital inputs for bearing, range, symbols, and vector lines are all converted to analog form for use by the tube.

Off-centering controls permit shifts of the picture in both X and Y directions of as much as 4 radii of the screen.

The tabular displays present additional information about targets appearing on the plan-position indicators. Text information is displayed in 6 columns of 12 rows each. Several such displays can be used with one plan-position indicator.

Digitrac Data Links—The speed, accuracy, and flexibility inherent in data transmission are vital in air-traffic control. Data transmissions in Digitrac are made in real time and are continually repeated so that occasional errors are corrected very quickly. Error detection is needed rather than error correction and is obtained by standardized message format and parity check.

A crystal-controlled frequency is divided in a multistage binary counter to produce a clock frequency that controls the transition times of the transmitted pulses. In the receiver the incoming pulses are compared with a similar clock arrangement and pulses are added or subtracted to keep the receiver clock frequency in exact phase with the transmitter clock frequency.

A typical airborne data receiver is shown containing 260 semiconductor integrated circuits and 9 thin-film hybrid circuits.

Congestion in a Loss System When Some Calls Want Several Devices Simultaneously—An extension of the theories developed by Feller and Cohen is presented. This should solve the problem encountered in computing control-circuit memories of a telephone exchange when different kinds of traffic of varying holding times are offered to one group of devices, each call being able to occupy one, two, or several devices simultaneously according to the category of traffic considered.

The theory is developed, and then an example of the above type is examined, followed by a method of numerical computation of blocking probabilities. The results are given for some particular cases in the form of curves, as a comparison with the case in which each category of traffic is handled by its own group of devices.

Realizing the General Two-Terminal-Pair Network with Independently Prescribed Transfer Function and Reflection Coefficient—Proof is given for the realization of a general two-terminal-pair network with independent, prescribed transfer function and with reflection coefficient or its tolerance. Associated problems are discussed. A uniform theory is presented for two-pole and four-pole networks employing inductors, resistors, and capacitors.

The linear and nonlinear relationships between two kinds of loss caused by inductor-capacitor or resistor network elements, have been thoroughly investigated. The power transferred, reflected, and lost within the four-pole network has been accurately analyzed. It has been established how these three types of power change in their relations to each other and how portions of them can be kept fixed. As a result, a prescribed loss characteristic merely must be parallel-shifted when the dissipation is to be compensated for.

In the first two of the five realizing procedures given, resistor-capacitor network elements have been employed in essence. The third procedure has been applied to a ladder network comprising a minimum number of inductor-capacitor elements, compensating for the dissipation of each coil. The fourth procedure is similar to the third and includes the compensation for dissipation of each capacitor. The fifth procedure is again similar to the third, but the number of inductors and capacitors exceeds the minimum; moreover, the perfect coupling of the Brune process is avoided, the loss margin is enlarged, and any equivalent circuit can be added as, for instance, a crystal working directly into the four-pole network.

The four-pole or two-terminal-pair network that is realized and exemplified with the new methods is much better than the previous network, even better than the inductor-capacitor four-pole network with respect to the loss and the reflection coefficient.

Results of Multifrequency Signaling Experiments—International dialing with its increased number of digits requires a fast and highly reliable means for transmitting switching information. Multifrequency-code signaling using 2-out-of-6 frequencies for each code has been tested in the German national telephone network using two systems. One is the pulse system, in which the receiving register, when actuated by a seizure signal, initiates with signals of fixed duration the transmission to it of the codes. The other is the compelled system, in which transmission of the first code seizes the receiving register and continues until its receipt is acknowledged to the transmitting register by a return signal. Following codes are similarly sent and acknowledged.

The choice of signaling frequencies is influenced by line characteristics, chiefly delay, echo, double echo, and harmonic distortion. Line interruptions, switching clicks, and noise must also be considered in signaling. The tests took all these into consideration.

Both systems are technically suitable as they are much faster than conventional dial pulsing. The final choice for an international European system will depend on adaptability to existing networks, reliability, and speed. The pulse system is faster and the compelled system is more reliable.

General Power Relations of Cyclotron Waves—The power exchange between synchronous or cyclotron waves and any arbitrary potential field is discussed without considering particular coupler designs. The interaction mechanism is revealed and Siegman's results are confirmed.

Accelerated Life Testing and Over-Stress Testing of Transistors—The advantages and disadvantages of accelerated and conventional life tests are considered. Methods of relating the results of accelerated tests to normal life are examined. The physical basis of accelerated life testing is discussed. Arrhenius and Eyring equations can be used to relate time to failure and temperature.

On the basis of the Arrhenius equation and using the log-normal failure distribution, the results from high-temperature storage of germanium alloy transistors are used to make predictions that are compared with results of life tests extending to 20 000 hours. The results are then used to compare constant-stress and step-stress accelerated life tests.

Electrical over-stress tests are described in which germanium alloy transistors are subjected to very-high-power pulses of short duration and silicon planar transistors have high currents passed through their emitter junctions in the reverse direction.

Mechanical over-stress tests are exemplified by centrifuge tests on germanium alloy transistors. For one transistor type the failure distribution, although in all cases normal, is markedly dependent on the direction of stress.

Recent Achievements

India Selects Pentaconta—The Indian Post and Telegraph Administration, after a 3-year study of telephone switching systems offered by manufacturers throughout the world, has selected Pentaconta. In a contract signed in New Delhi in May 1964 (Figure 1) provision is made not only for an initial installation of 6 local exchanges and 4 trunk exchanges but also for a telephone manufacturing plant to be built in Bangalore.

The factory, which in 3 years will have an annual production of 100 000 lines of equipment, will be owned and operated by Indian Telephone Industries, a state-owned corporation. International Telephone and Telegraph Corporation will have a minority interest in the factory, will license manufacture under its patents, provide all necessary technical assistance, supply both tooling and test facilities, and train Indians, some of whom will be sent to Belgium and France for experience in our factories there, to serve as key personnel.

The immediate installations of 48 000 lines and 6500 junctions will be in local and trunk exchanges in Bombay, Kanpur, Madras, and New Delhi.

Bell Telephone Manufacturing Company
Belgium

Compagnie Générale de Constructions Téléphoniques
France

Traffic Evaluation for Control of Traffic Signals—Simple time control of traffic lights will not care for short-term or highly localized variations in traffic density. A traffic evaluator has therefore been developed to adjust the signal program to the variations in traffic on a continuing basis.

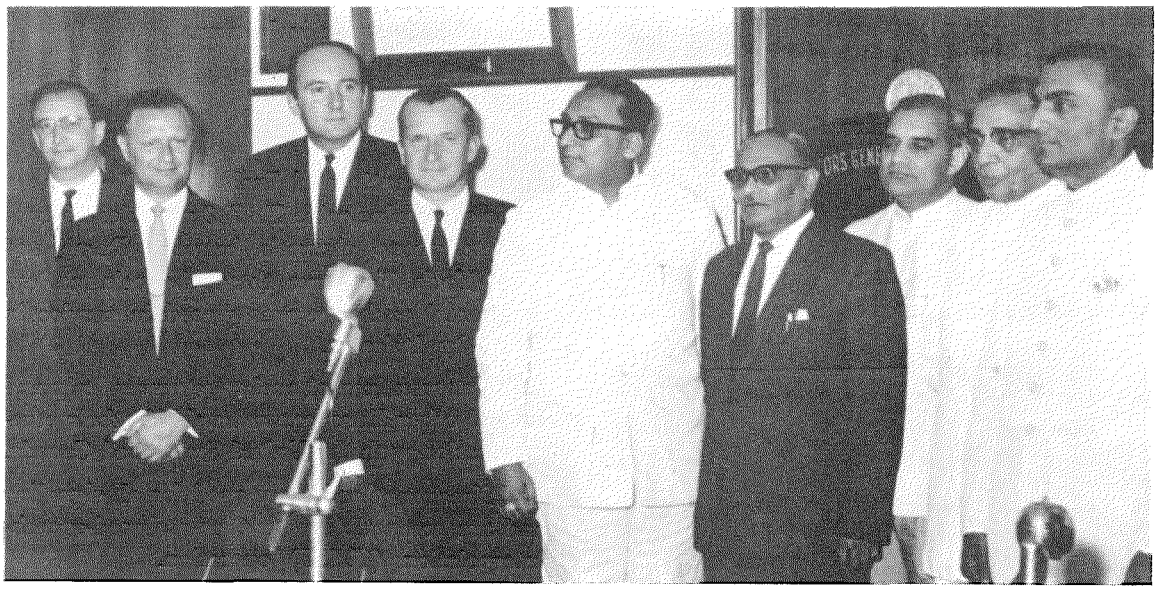


Figure 1—At the signing of the contract for Pentaconta installations and factory in India were, from left to right: Eugene Van Dyck, Manager for Switching, and F. Pepermans, Managing Director of Bell Telephone Manufacturing Company; G. R. Andlinger, President of ITT Far East and Pacific; His Excellency A. Wendelen, Ambassador of Belgium to India; His Excellency Asoke K. Sen, Minister of Law and Communications of India; Uma Shankar, Managing Director of Indian Telephone Industries; His Excellency K. B. Lall, Ambassador of India to Belgium; L. C. Jain, General Manager of the Indian Post and Telegraph Administration, and S. K. Kanjilal, Member of the Indian Post and Telegraph Board.

At critical points, radar monitors supply information to the evaluator for processing at predetermined time intervals. A change in traffic is stored for one time interval and, if it is confirmed by the next report, the signal program is changed accordingly. A change not confirmed is cancelled. Thus frequent changes in signal programs to meet brief fluctuations in traffic density are avoided.

For optimum control three criteria are processed: the number of vehicles, the ratio of traffic in the two paths under control of the signal, and the speed of the vehicles. The number of vehicles merely indicates whether traffic may be handled by the night program. If above this density, the information is evaluated with the other two factors. Five ranges of ratios of traffic in the two paths are used, and the speed information would be of the type that 85 percent of the vehicles are moving slower than some specified speed.

Several evaluators may be interconnected to control traffic over large areas. The radar units are connected over 4-wire lines to counting-pulse equipment, which is similarly connected to the evaluators at control headquarters.

Standard Elektrik Lorenz
Germany

Push-Button Dialing Inaugurated in Austria—From the famous Baroque Room in the Vienna headquarters of the Austrian Post, Telegraphs, and Telephones Administration, on 5 May 1964 its Director General, Dr. Benno Schaginger, made the first call in that country on a push-button subscriber set to Otto Probst, Austrian Minister of Telecommunication and Transportation.

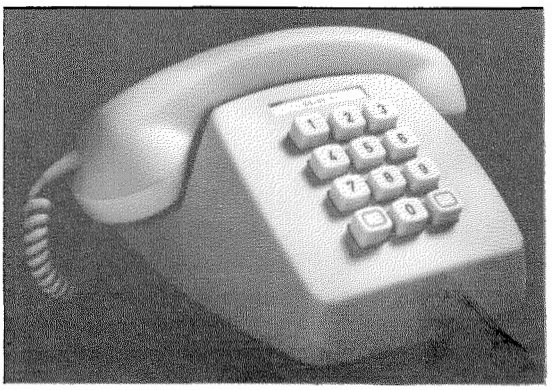


Figure 2—Push-button type of subscriber set being installed in Austria.

The audience of Administration officials and journalists was addressed by Dr. Schaginger and Mr. F. W. Mayer, General Manager of Standard Telephon und Telegraphen, who also demonstrated the advantages of the new system by simultaneously placing calls with the new telephone set, shown in Figure 2, and a conventional dial set.

The initial installation of push-button subscriber sets was limited to the private automatic branch exchange of the Post Administration, the necessary revisions to accommodate the new type of calling having been made in its crossbar switching system. Several push-button sets will be installed for the 15th Congrès de l'Union Postale Universelle in Vienna.

Standard Telephon und Telegraphen
Austria

Rome Intercontinental Exchange—The transatlantic submarine cable *TAT 2* provides 16 two-way telephone channels between Rome and New York. To serve these channels, an experimental exchange was placed in service in Rome by Italcable.

This intercontinental exchange serves Italy and several other European countries in their communication with the United States and Canada.

Multifrequency codes are used for all signaling and are compatible with time-assigned-speech-interpolation (*TASI*) operation.

Fabbrica Apparecchiature per Comunicazioni
Elettriche Standard
Italy

Hawaii-Japan Submarine Cable—On 19 June 1964, the submarine cable from Hawaii via Midway and Guam to Japan was put into serv-

ice. It interconnects at Hawaii with the Compac system, which was opened in December 1963 to link Canada, Hawaii, Fiji, New Zealand, and Australia, as well as with the previously laid cable between Hawaii and the United States.

The cable is 5600 nautical miles (10 400 kilometers) long. It is of the same design as the third transatlantic cable *TAT 3* and accommodates 128 two-way channels. More than 2000 nautical miles (3700 kilometers), including submerged repeaters, were supplied by Standard Telephones and Cables.

Standard Telephones and Cables
United Kingdom

Southeast Asia Submarine Telephone Cable—The first part of Seacom, the Southeast Asia section of the round-the-world telephone cable, will link Hong Kong with Singapore via Jesselton in Sabah as shown in Figure 3. Later links will extend it to Australia where interconnection with the Pacific network will permit telephone service via Canada between Singapore and London.

Standard Telephones and Cables will supply about ⅓ of the 2073 nautical miles (3840 kilometers) of cable of both armored shallow-water and unarmored deep-sea British Post Office types, 70 submerged repeaters of the 95 needed, and 10 undersea equalizer units. It will also provide transductor-controlled units to supply power to the cables.

Standard Telephones and Cables
United Kingdom

Air-Traffic-Control Radar Demonstration—At the inauguration of the new Barkarby plant, various recent developments were demonstrated. The most important was of the new Digitrac system for air-traffic control. By color slides and recorded commentary and sounds, a flight from Stockholm to Copenhagen with procedures from before take-off and enroute was

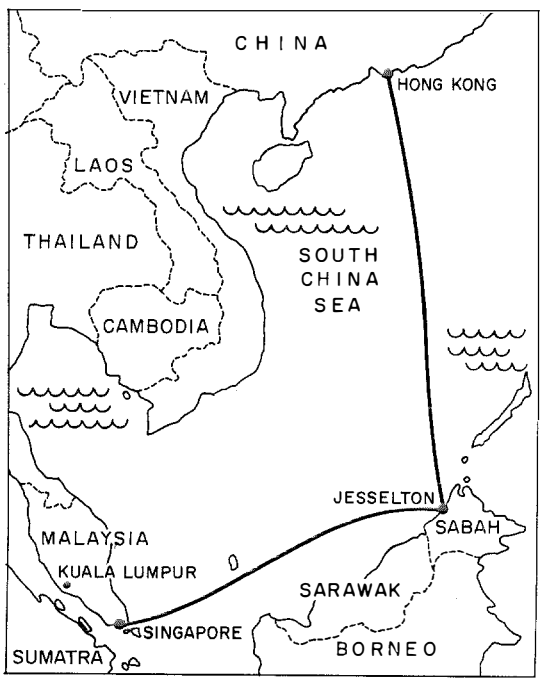


Figure 3—Telephone cable link between Hong Kong and Singapore.

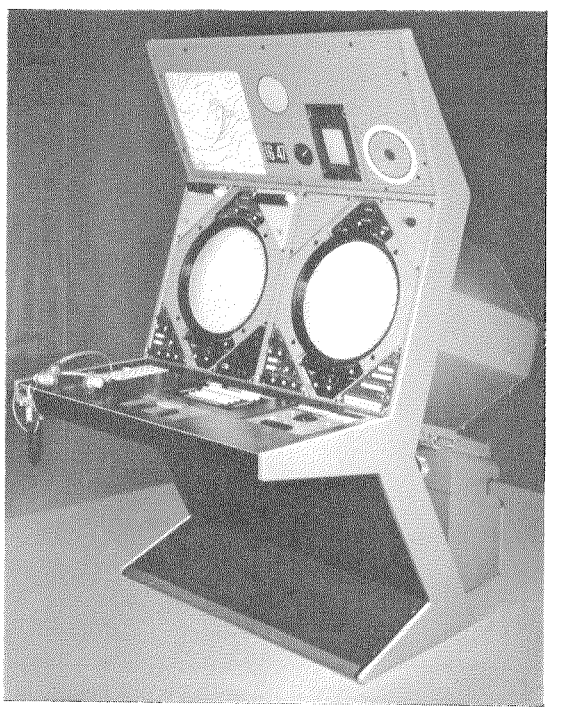


Figure 4—Console for air-traffic-control center.

demonstrated. One of the traffic-control consoles is shown in Figure 4.

Standard Radio & Telefon
Sweden

Microscope Coordinate Reader—An automatic system to determine and record the coordinate positions of subjects being viewed in a microscope is shown in Figure 5. The adjusting screws, including focusing for the depth coordinate are controlled by three reversible motors through flexible shafts. The positions of the shafts are converted into an appropriate output code.

The operator does not touch the microscope adjustments but only controls the motors. Motor speeds are adjustable over a range of 20:1. The accuracy of positioning and reading the shafts controlling the ocular coordinates is within 1/100 turn and of the focusing or depth coordinate is within 1/500 turn.

The electronic apparatus includes a code converter and programer. The latter controls operation of a perforator, which in 2 seconds will print the 3 coordinates of a point. The transistor circuits are mounted on plug-in modular cards.

Bell Telephone Manufacturing Company
Belgium

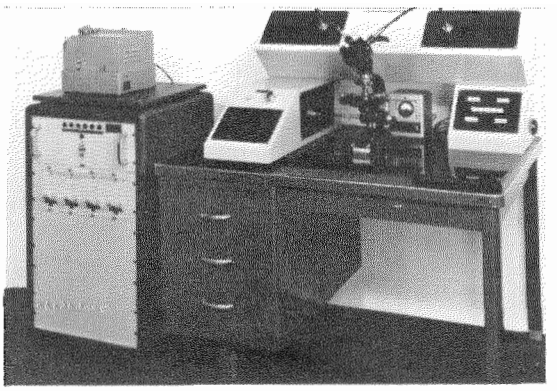


Figure 5—Microscope with motor-operated controls and automatic print-out of adjustment data.

Machine That Learns Tunes—Ordinary recording devices simply copy in, say, magnetic patterns on tape the equivalent waveforms of music or speech that are supplied to a recording head. In contrast, a computer-type device has been developed that analyzes musical tones as rapidly as they are received and stores information on their duration, sequence, and frequency contents. Using its own internally generated notes, it can then play the tune it has learned.

ITT Federal Laboratories
United States of America

Teleprinter for 200 Characters Per Second—Conventional teleprinters operate at 50 bauds, corresponding to 6.6 characters per second. Some designs will double this speed, but to meet demands for 1500 bauds or 200 characters per second new techniques are needed.

For printing at 1500 bauds, the electrolytic tellurium method is used. A 5-by-7 matrix of 35 dots permits any character to be formed. Two electronically controlled printing heads cover the 72-character line. The individual styluses are mounted in a block for insertion in the printing head as a unit. The printing impression is clear, lightproof, does not smear, and can be copied by existing methods.

The transmitter uses photoelectric tape reading. The equipments that make up a station are shown in Figure 6. A transmission channel with a bandwidth of 3 kilohertz is required. The outgoing message may be monitored by the

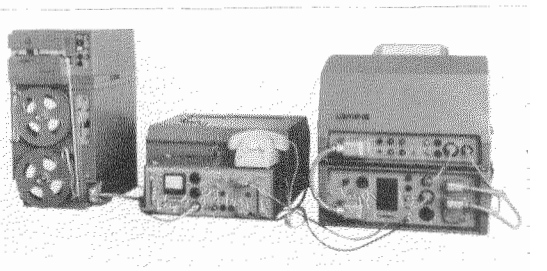


Figure 6—*LO 2000* teleprinter installation capable of working at 200 characters per second.

receiver. The system is intended for private and special networks.

Standard Elektrik Lorenz
Germany

Waffle-Iron Store—Unlike most magnetic thin-film stores, the waffle-iron, which we developed from an original design by Bell Telephone Laboratories, has a closed flux path that combines the advantages of conventional thin-film and ferrite-core stores. These include high packing density, simple construction, low driving current, good output voltage, and absence of creep.

As shown in Figure 7, the flux is returned through a soft ferrite structure. Driving and sense conductors are laid in the grooves in the ferrite plate. The plate is pressed into contact with the thin film, which has a square-loop magnetization characteristic. The thin film is an isotropic iron-nickel alloy electroplated on a polished substrate.

Storage occurs as a diagonal magnetization in the film at the crossroads between the two grooves in which the word and digit currents are flowing simultaneously. The information is read out by applying word current of reversed polarity.

An experimental store of 196 words of 30 bits each is shown in Figure 8. It has a 0.5-microsecond read-and-write cycle. The printed-circuit boards form read and write diode access matrixes. The grooves are 127 microns wide on 380-micron centers. With 4 storage points per bit, there are over 1000 bits per square inch (155 bits per square centimeter). Typical operating values are a word write current of 300 milliamperes, digit current of ±125 milliamperes, read current of 500 milliamperes, and disturbed output pulse of 30 millivolts for 30 nanoseconds.

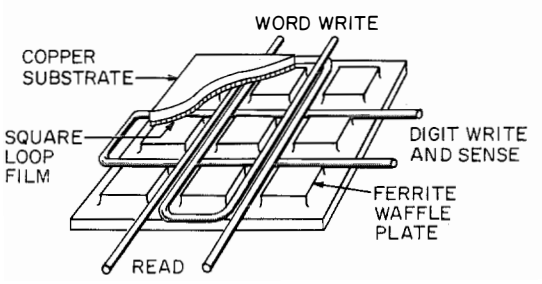


Figure 7—Arrangement of waffle-iron store.

Standard Telecommunication Laboratories
United Kingdom

Pentaconta for Broadcast Program Switching—Pentaconta equipment has been installed for broadcast program switching in the French Maison de la Radio. The equipment was specially designed for low noise and distortion.

The most important of the several assemblies has 40 multiselector frames that give 168 inputs access to 110 outputs. Measurements show that in the most-unfavorable case in the entire installation, the crosstalk attenuation between any circuit and all other circuits including cables and amplifiers exceeds 88 decibels in the band from 40 to 15 000 hertz. Neither phase shift nor measurable noise has been detected.

The results indicate such equipment to be suitable for carrier systems to switch primary groups up to 108 kilohertz and secondary groups up to 600 kilohertz. Correspondingly,

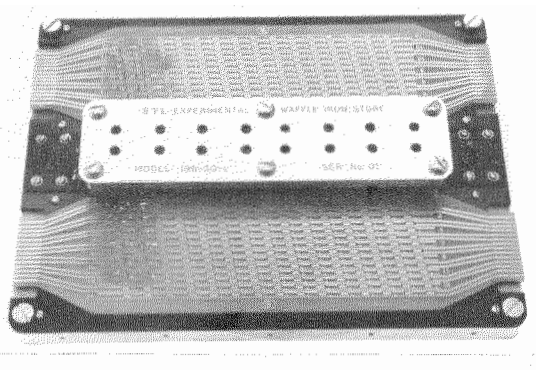


Figure 8—Experimental waffle-iron store for 196 words of 30 bits each.

data transmission at 40 000 bauds could be handled.

Compagnie Générale de Constructions Téléphoniques
France

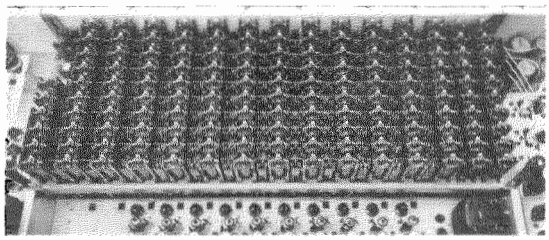


Figure 9—Herkon relays controlled by flat-type relays are used for switching television program signals.

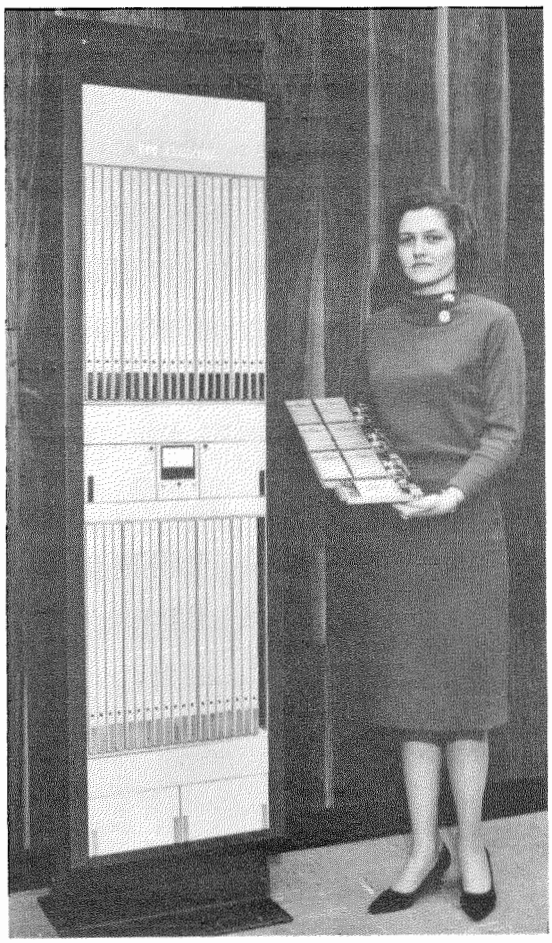


Figure 10—New-generation voice multiplex equipment.

Television Studio Remote-Controlled Program Switching—Herkon relays, having contacts mounted in hermetically sealed gas-filled glass envelopes, are used in a crossbar pattern for switching video signals in a television studio. Flat-type relays, which have too much stray capacitance for switching such high-frequency signals, are used to control the Herkon units as is evident in Figure 9. They are controlled by push buttons on the mixer desk.

The vertical connectors go to the mixer amplifiers and to the preview amplifiers. The horizontal lines go to the cameras, film scanners, and outside sources of programs. The control circuits are arranged to prevent more than one program source from being connected at the same time to an output line.

Standard Téléphone et Radio
Switzerland

Universally Applicable Multiplex Equipment—A new-generation multiplex equipment for voice telecommunication service meets all applicable specifications of the Comité Consultatif International Télégraphique et Téléphonique, Australian Post Office, United States Defense Communications Agency, Electronic Industries Association, and Bell System Practice.

Flexibility is obtained by using 3-channel pregroups that can be combined into 12-channel groups; either pregroups or groups can be through-patched as required by system design

Racks may be mounted back-to-back or back-to-wall. A width of only 21 inches is required for 60 channels. Only 80 modules are needed for 240 channels compared with 800 modules for previous designs.

Failures characterizing equipment built only 4 years ago are avoided by using silicon transistors instead of germanium units, hermetically sealed modules, and only 20 percent as many connectors. About 6 times as many channels

can be accommodated in the same space. A rack of the new equipment is shown in Figure 10.

ITT Telecommunications Group
United States of America

Artificially Grown Quartz—The first commercial plant in Europe for growing quartz crystals has been established at Harlow.

The growing process takes place inside a steel cylinder 10 feet (3 meters) long and 12 inches (30 centimeters) in diameter at a pressure of 24 000 pounds per square inch (1687 kilograms per square centimeter) and a temperature of 400 degrees centigrade. Small seed crystals immersed in a bath of caustic soda in which quartz chips have been dissolved grow as the temperature is slowly lowered.

In Figure 11, large crystals suitable for manufacture of frequency-control units are being taken from the cylinder after growing for less than a month. They show good crystal faces and are virtually free of defects.

Standard Telephones and Cables
United Kingdom

New Factory in Sweden—Provincial Governor Erik Westerlind on 1 June 1964 inaugurated a new office and factory at Barkarby, 10 miles (16 kilometers) northwest of the previous main plant in Stockholm-Bromma. Figure 12 is a view of the structure.

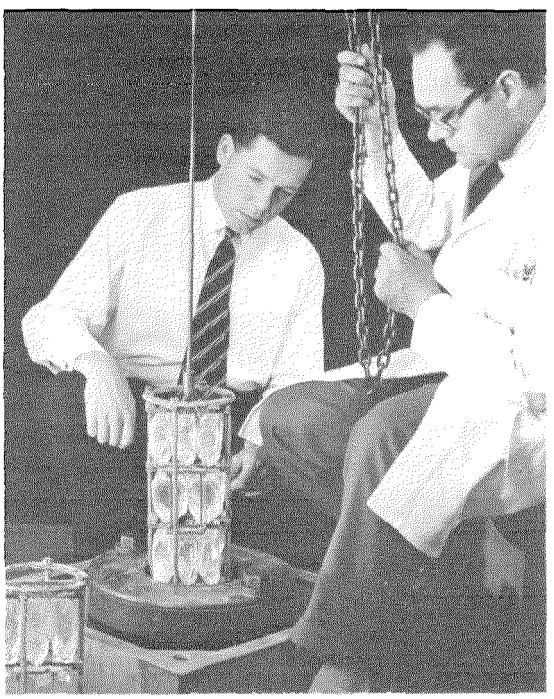


Figure 11—Unloading quartz crystals from the pressure vessel in which they were grown.

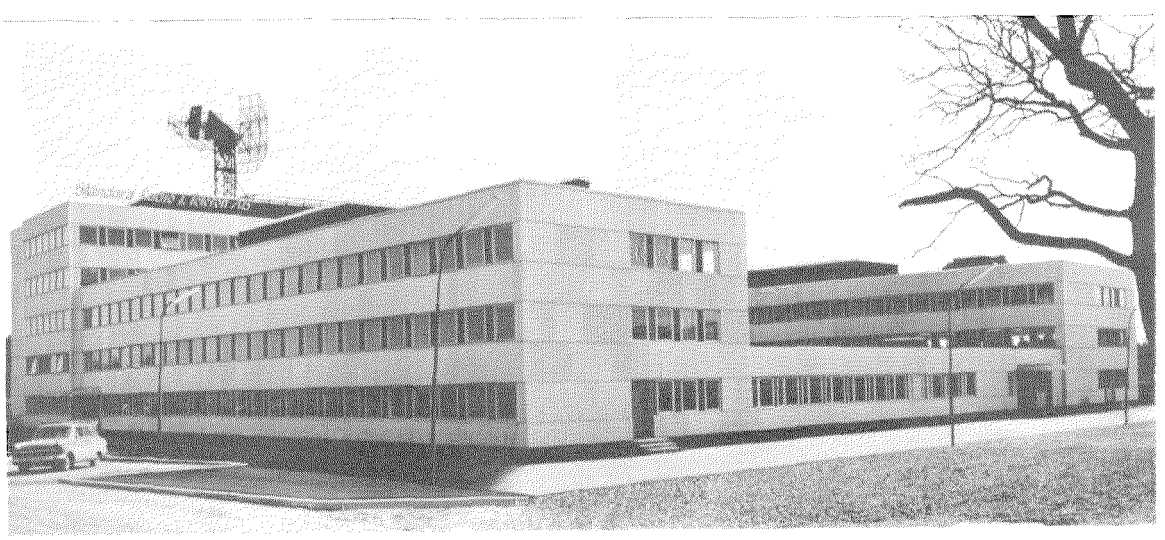


Figure 12—New plant at Barkarby near Stockholm.

The new buildings provide space for headquarters offices and the electronics department. The 5-story office building is flanked by two 3-story wings mainly for laboratory use. A single-story central area of about 5000 square meters (53 800 square feet) is for production of electronic equipment. The total floor area of 18 600 square meters (200 000 square feet) will accommodate about 800 employees. Although it is one of the largest structures of its kind in Sweden, planned expansion will provide for about 2500 employees.

Standard Radio & Telefon
Sweden

Vortac for Amsterdam Airport—The Netherlands Civil Aviation Authority inaugurated on 9 June 1964 its first experimental Vortac installation, which will serve the main civil international airport in that country. For both civil and military use, the facility is at the junction of the airways leading to Schiphol airport near Amsterdam.

To the dual very-high-frequency omnidirectional range (*VOR*) that has been in service for a decade, Le Matériel Téléphonique added a single *AN/GRN-9A* Tacan beacon to complete the Vortac installation.

To improve instrument low-approach operations, a *STAN 7* localizer, supplied by Standard Telephones and Cables, will be added to the present *STAN 8* glide-slope equipment.

Nederlandsche Standard Electric Maatschappij
The Netherlands

Miniature Inductor—A difficulty in making planar miniature inductors on printed-circuit boards is that the conductor thickness must be relatively large to provide a suitable inductance/resistance ratio. Spark machining permits narrow slots to be cut in copper strips of adequate thickness, which may be laminated to insulators. As shown in Figure 13, the tool is two interleaved cutting and spacing tapes wound around a mandrel.

The inductor shown in Figure 14 has an inductance of 6 microhenries and a direct-current resistance of 6 ohms. The tool to make it has 30 turns of tantalum tape 0.0003 inch (7.5 micrometers) thick, spaced by copper tape 10 times as thick wound on a copper tube 0.1 inch (2.5 millimeters) in diameter. The outer diameter of the winding is 0.3 inch (7.5 millimeters). The copper is etched to leave the tantalum standing clear.

A copper sheet of 1 ounce per square foot (30 milligrams per square centimeter) laminated to an insulator can be processed in 4 to 7 minutes.

Standard Telecommunication Laboratories
United Kingdom

Herkon Contacts Perform 500 Million Operations—An accelerated life test of 20 Herkon contacts in a direct-distance-dialing translator showed that each contact has performed over 5×10^8 operations without failure. Electrical characteristics including contact resistance are still within specified tolerances.

In 1959, the Deutsche Bundespost decided to equip all direct-distance-dialing translators with Herkon reed relays. Experience gained within the past 5 years has entirely justified this decision.

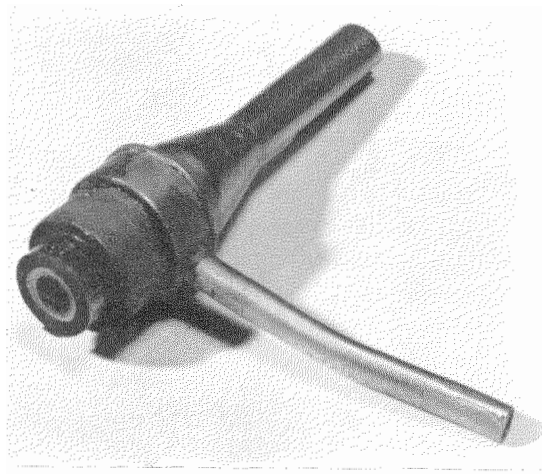


Figure 13—Tool for spark machining spiral inductors on printed-circuit boards.

The assumption that the Herkon contacts would have lives of 10^8 to 10^9 operations, about 20 years, has been well supported by the accelerated tests.

Standard Elektrik Lorenz
Germany

HE-60 Quasi-Electronic Trial Office Completes First Year—The inauguration of the first trial telephone central office using *HE-60* quasi-electronic switching with Herkon sealed reed relays occurred in July 1963 in Stuttgart.* With a capacity of 2000 lines, it has served 500 subscribers. Immediately after cutover, installation of equipment for augmented services was started and all subscribers were connected to it in March 1964. Augmented service provides for either conventional dialing or multifrequency push-button signaling and for subscriber-controlled switching to absentee and changed-number services. A limited number of subscribers also use abbreviated dialing. About 40 of the subscribers have push-button telephone sets.

During system operation of over 2×10^9 component hours, only 1 transistor, 2 germanium diodes, and 2 Herkon contacts had to be replaced. These failures were due to improper handling. None of them led to a service breakdown.

Standard Elektrik Lorenz
Germany

Data Storage, Display, and Transmission—The *KD-5010* system, shown in Figure 15, is for use in data processing and digital communication systems. Over 100 preformed messages of several hundred characters may be selected by a keyboard for display on a cathode-ray tube or for transmission. A full range of alphanumeric characters and special symbols are derived from

* "Quasi-Electronic Telephone Exchange," *Electrical Communication*, volume 38, number 4, page 434; 1963.

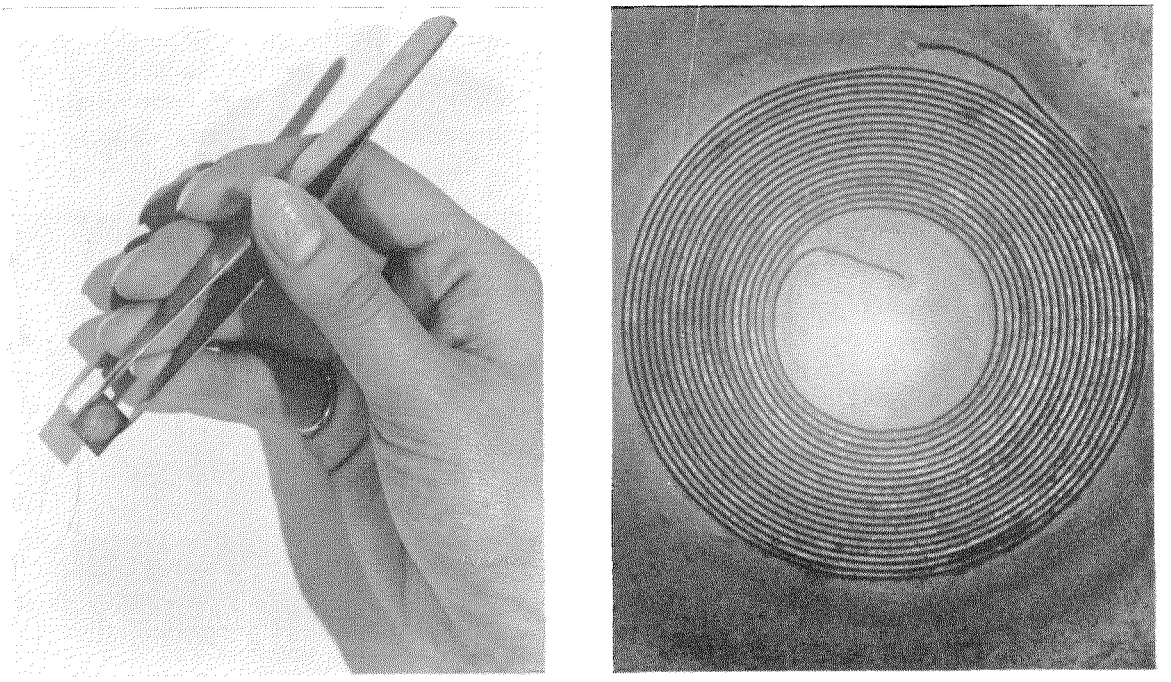


Figure 14—Two views of the miniature spiral inductor machined on printed board.

either monoscopes or a core matrix character generator. They may be initiated by a computer or an operator. Visual and audible alarms can be sent to alert personnel under prescribed conditions.

Specific areas of use include on-line command and control. computer-controlled testing of assemblies and subassemblies in large systems, data analysis, process control, information retrieval. and the evaluation of programs for maximum utilization of available time.

ITT Kellogg Communications Systems
United States of America

Rapidata Telegraph Tests—In the Rapidata *S* system, data are transmitted in blocks with checks for errors and automatic repetition of blocks that are perturbed. Practical tests were made through the courtesy of the transmission department of the French navy.

The test matter was recorded on 8-channel perforated tape and delivered by a photoelectric reader at 1200 bauds to the telephone channel. Received on magnetic tape, it was reproduced at a slower rate to operate a perforator to produce a tape similar to the originating tape.

A preliminary test was successful over a 2000-kilometer (1250-mile) switched telephone channel from Paris to Toulon and return that was unsatisfactory for telephone service. The major test was over manually switched private lines of the navy with transmission from Toulon to Paris. This connection was operated successfully for 10 days under control of the Marine Services.

Compagnie Générale de Constructions Téléphoniques
France

Miniature Selenium Rectifiers for Television Receiver Extra-High Voltage—Type *X80/150* rectifier uses 150 selenium discs stacked in a tube 3 inches (76 millimeters) long. The discs are 0.08 inch (2 millimeters) in diameter and are shown with some stacks in Figure 16.

In Great Britain these rectifiers are replacing the vacuum tubes previously used. Standard receivers require 3 rectifier stacks while battery-operated and small-screen sets need only 2 stacks.

Standard Telephones and Cables
United Kingdom

Trial Electronic Telephone Exchange in Antwerp—After constructing a 10 000-line working model of a time-division-multiplex telephone switching system in 1960, a trial electronic 200-line private automatic branch exchange has

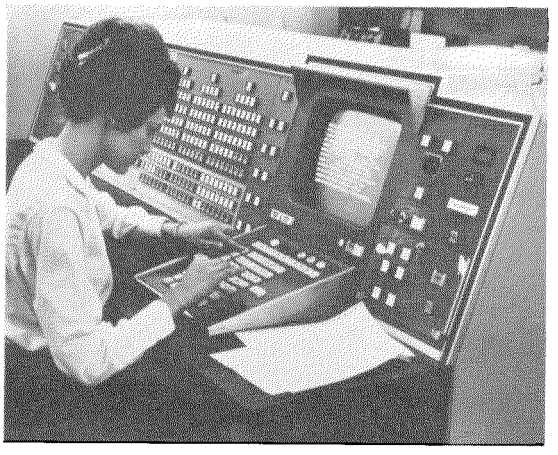


Figure 15—*KD-5010* solid-state data display and transmission system.

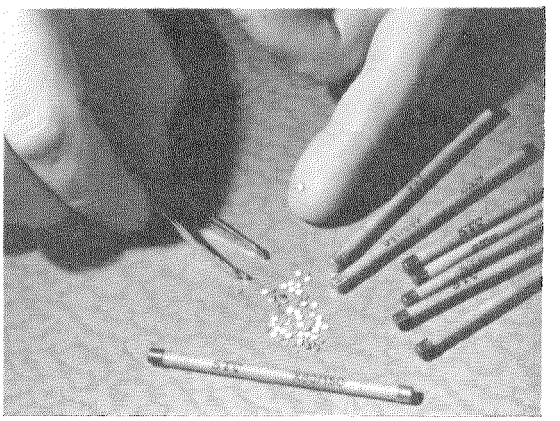


Figure 16—Each stack contains 150 selenium rectifier discs for the extra-high-voltage supply to a television picture tube.

been installed by Bell Telephone Manufacturing Company in Antwerp to handle its regular business traffic. This subexchange is connected to the existing *7E* rotary branch exchange over 13 junctions each for incoming and outgoing traffic.

The time-division switching uses amplitude modulation with resonant transfer. A common 2-wire speech highway is shared sequentially by 25 channels for 4 microseconds each during each 100-microsecond sampling cycle. Semiconductor diodes and transistors are used for switching and logic and ferrite cores for storage. Power consumption is only 800 watts. Either push-button or dial subscriber sets may be connected to all terminals.

Bell Telephone Manufacturing Company
Belgium

Portable Microwave Terminal—The *AN/TRC-112* is a transportable radio terminal for line-of-sight and forward-scatter tactical operation. Individual designs operate at 2, 5, or 8 gigahertz.

Designed for 600-channel tactical pulse-code multiplex and conventional frequency-division multiplex, it uses transistors and integrated circuits. A klystron in the 1-kilowatt output stage of the transmitter is the only vacuum tube in the equipment.

The complete terminal comprising a shelter that houses the transmitter and receiver, two 10-foot (3-meter) paraboloidal antennas, and two power generators can be transported in two small trucks, each towing a trailer mounting a power unit.

ITT Federal Laboratories
United States of America

Complex Computation—An engineer and a computer have solved a problem involving 2002 equations and 13 542 variables, believed to be the largest mathematical problem ever to be solved.

To find the shortest possible noninterfering routes for messages through a network of 62 switches, as indicated in Figure 17, use was made of linear programing. A full year was required to evolve the two original equations that instructed the computer but only 6.5 hours of computer time were needed.

ITT Communication Systems
United States of America

Mexico Telephone Switching Expansion—As part of a 5-year telephone development plan, Mexico has ordered switching equipment for 7200 lines in addition to an earlier order for 12 600 lines. These 19 800 lines will be installed

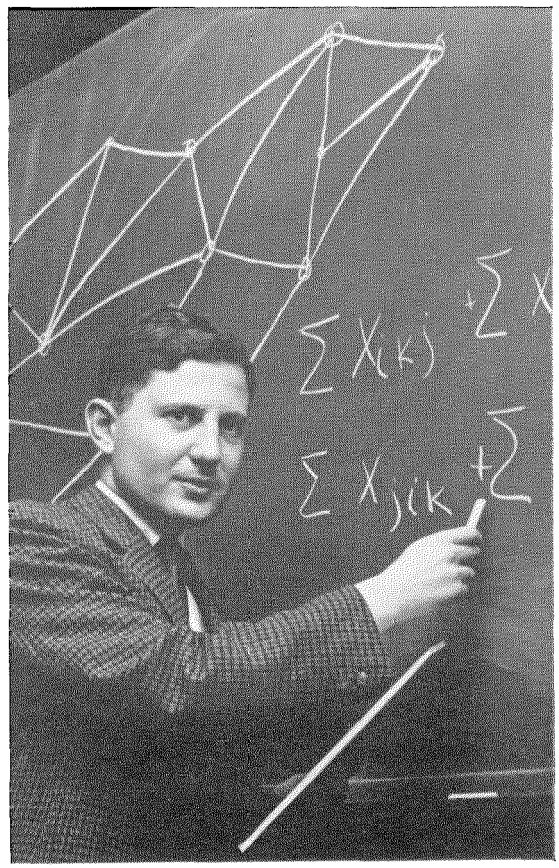


Figure 17—Simeon E. Gordon discusses how to get a call through a network of 62 switches.

in 10 automatic exchanges that have an ultimate capacity of 146 000 lines.

Compagnie Générale de Constructions Téléphoniques
France

Marine Radio Installation—A fleet of 11 tankers, one of which will be the largest ever built in Great Britain, now under construction for British Petroleum Tanker Company will be provided with the newest of marine radio equipment.

The installations will include the new *ST 1200* single- and double-sideband transmitter * recommended for marine use in the 1970's. Shown in Figure 18, this transmitter is suitable for telegraph, telephone, and data transmission from ship to shore.

International Marine Radio Company
United Kingdom

Telephone Operating Position—A new operating desk shown in Figure 19 has been developed for private branch exchanges. The foot, body, and table are separate units made of polyester plastic laminated with glass cloth. The glazed surfaces have a fine satin finish, which may be of any desired color.

The cable ends are wrap connected to a terminal block mounted on the foot, which is done before the body is added. The switching components are mounted in subassemblies and are connected to the terminal blocks by plug-in connectors. As access to the rear of the desk is never needed, desks may be mounted against a wall and back to back. The maximum equipment capacity is 600 buttons or lamps.

The over-all dimensions are 90 by 73 by 95 centimeters (35.7 by 28.6 by 37.5 inches) and a supervisory desk has the same height but is 150 by 80 centimeters (58.7 by 31.5 inches) in width and depth.

Compagnie Générale de Constructions Téléphoniques
France

* F. Hedström, "Marine Transmitter with Single-Sideband Facility," *Electrical Communication,* volume 38, number 4, pages 465–474; 1963.

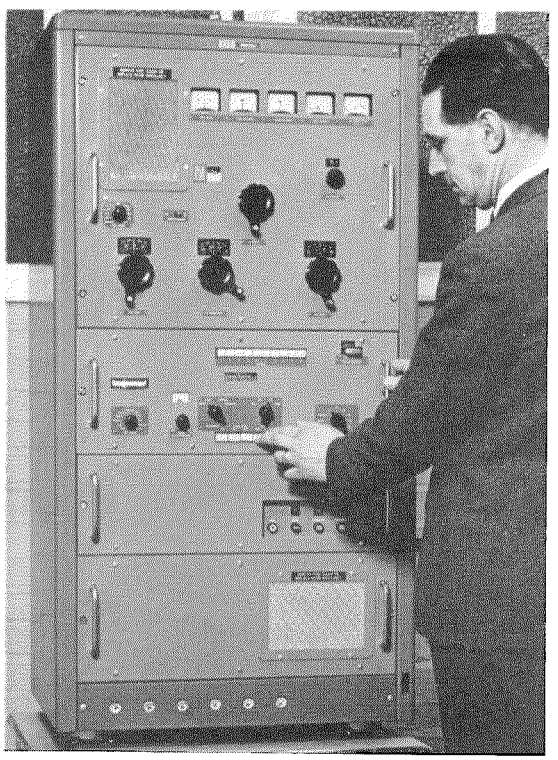


Figure 18—*ST 1200* marine radio transmitter.

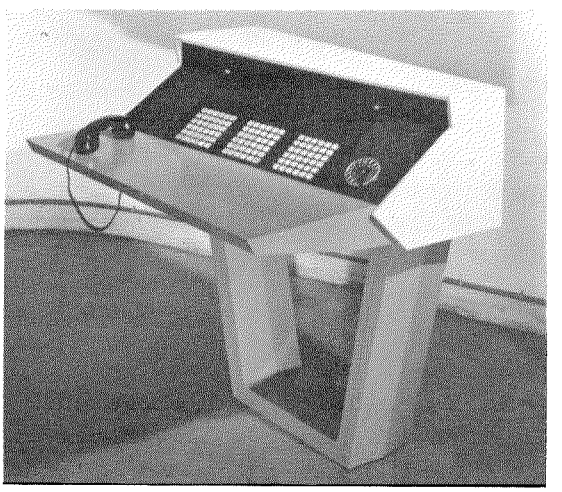


Figure 19—Plastic telephone operating desk for private branch exchanges.

Microwave Link for South Africa—A radiotelephone system will be installed to link Bloemfontein, Port Elizabeth, Capetown, and East London in South Africa.

The system will provide 960 channels, operating in the 6-gigahertz band with 10-watt transmitters. Extensive use will be made of transistors for high reliability and low operating cost. Some links are as long as 70 miles (113 kilometers), about twice that of conventional microwave links.

Standard Telephones and Cables
United Kingdom

Test of Wire-Wrapping Tools—Wire-wrapping tools used in production work are tested daily. To improve both accuracy and speed over manual methods, automatic means have been developed for determining the mechanical resistance of wrapped connections up to the point of stripping off.

At the test position shown in Figure 20, each tool is used to make 5 connections, the results being printed automatically on tape. If any of the connections are outside of the specified limits, the test is interrupted and a lamp lights.

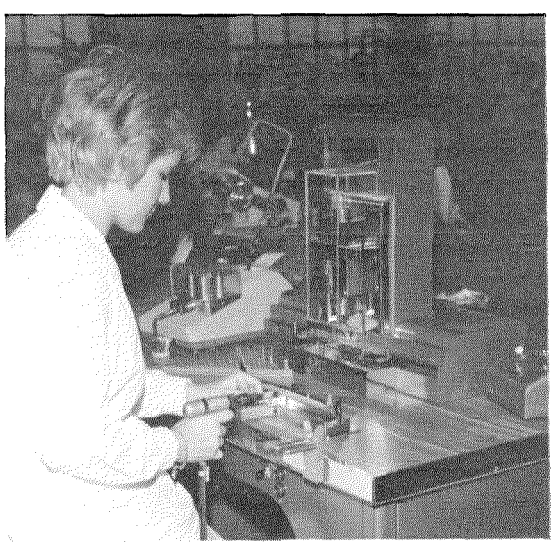


Figure 20—Testing of tools for making wrapped wire connections.

Le Matériel Téléphonique
France

Hull Gets New Telephone Exchange—The only public telephone system in the United Kingdom not operated by the British General Post Office is that of the Hull Corporation Telephone Committee. Alderman R. E. Tennyson, chairman of that committee, recently inaugurated its new "Civic" exchange, which adds 7000 subscribers and also opens the way for subscriber trunk dialing into the British network. Another new exchange and extensions to two existing central offices are in process.

Standard Telephones and Cables
United Kingdom

Infrared Transmission Link—The *PE1AF* infrared-emitting diode has been built into a transmission system (shown in Figure 21) suitable

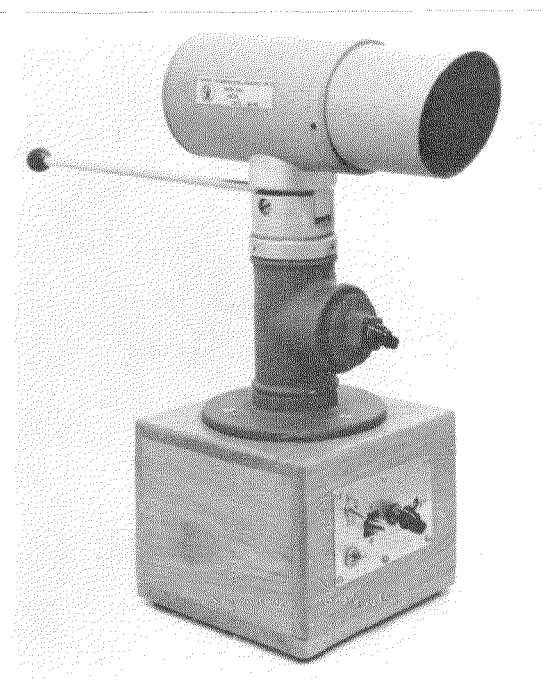


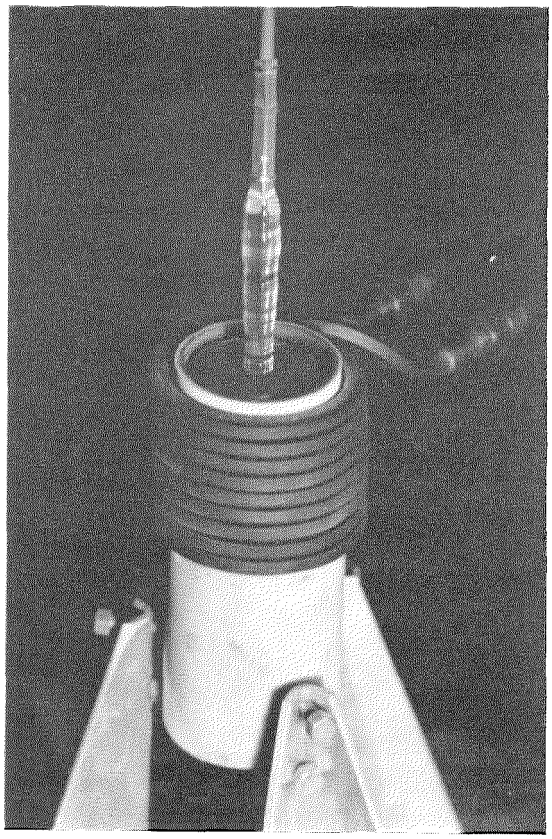
Figure 21—Transmitter using the *PE1AF* infrared source for speech communication.

for speech and music. Such links are useful for communication across canals and rivers and on civil engineering sites or other places where wire lines cannot be used. The narrow angle of the transmitted beam provides secrecy that cannot be obtained with conventional radio links.

Another design modulates the beam at a single frequency and with narrow-band detection will operate over 1000 meters (3280 feet) or more. Applications include intruder detection, monitoring for fog and smoke, and simple alarm signaling. The inherent straightness of the beam enables it to be used as a surveyor's portable baseline, references being obtainable to within 1 or 2 centimeters of the beam axis at any point along the path.

Standard Telecommunication Laboratories
United Kingdom

Figure 22—A calcium-tungstate crystal being pulled from a platinum-plated rhodium crucible.



New Crystals for Lasers—Single crystals of calcium-tungstate, strontium-molybdate, and calcium-molybdate of a purity and size suitable for laser use are difficult to produce. Using the Csochralski method, the required temperatures of 1500 to 1650 degrees centigrade make conventional crucibles unusable.

Rhodium crucibles plated with platinum are satisfactory. The mechanical rigidity is provided by the rhodium and the platinum prevents contamination of the melt by the rhodium-oxide formed at the operating temperature. The pulling of a calcium-tungstate crystal is shown in Figure 22.

Laboratoire Central de Télécommunications
France

Semiconductor Laser—A semiconductor laser now being manufactured in sample quantities is capable, when cooled to the temperature of liquid nitrogen, of converting electric pulses into essentially coherent infrared radiation in the 8500-angstrom-unit region. Typical efficiency is about 5 percent although values as high as 30 percent have been obtained.

Based on accurate orientation of a gallium-arsenide crystal, slices are cut so that the cleavage of the (110) planes can be precisely controlled to occur normal to the slice surfaces. A diffusion of zinc into the heavily selenium-doped starting material produces a p-n junction approximately 20 microns below the surface of the slice. The slices are then cleaved in 0.5-millimeter strips and cut into dice 0.2 millimeter wide. Each die is alloyed to a gold-plated molybdenum disc, which acts as the p contact, and a dot of tin is used as the n connection.

The tiny assembly is mounted on a copper pedestal on a *TO 18* header as illustrated in Figure 23. It is a small efficient light source

capable of extremely fast response (about 5 nanoseconds).

Standard Telecommunication Laboratories
United Kingdom

Methane Filter for Laser—A helium-neon laser has been equipped with a methane-gas filter to block radiation at 33 953 angstrom units. The power of the radiation at 6328 angstrom units derived from the same energy level has thereby been doubled. The presently available power of 400 milliwatts permits increased applications of lasers to telecommunication.

On the theoretical side, the helium-neon laser with methane filter has made it possible to study the excitation mechanism of the energy levels and to measure the efficiency of the cascade laser effects leading to optical-wavelength conversions.

Laboratoire Central de Télécommunications
France

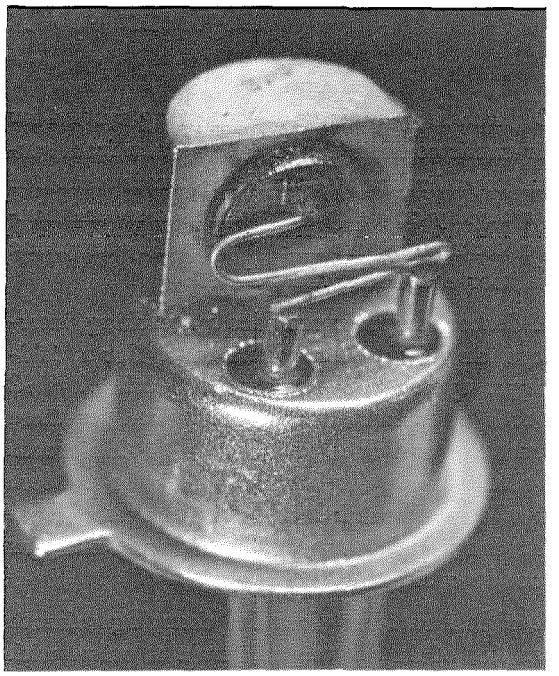


Figure 23—Gallium-arsenide laser assembly.

Schaltungen und Elemente der digitalen Technik (Circuits and Circuit Elements Used in Digital Techniques)

Konrad Bartels and Boris Oklodbizija of Standard Elektrik Lorenz in Pforzheim present in this book an over-all treatment of digital circuit techniques. It is divided into two sections, one on digital circuits and the other on their application.

Recommendations and rules for designing circuits and for rating commonly used components are given, together with examples. A bibliography and comprehensive index facilitate use of the book.

This 156-page book is 15 by 21 centimeters (5.9 by 8.3 inches) in size. It is published by Verlag für Radio-Foto-Kinotechnik, Berlin-Borsigwalde, Germany, at DM 21 per copy.

Digitrac for Air-Traffic Control

Digitrac is a digital data-processing system for air-traffic control. Target information from several asynchronously operated radars is combined for display on cathode-ray tubes in several control centers that are interconnected by telephone lines or equivalent radio links.

A Censor computer provides for tracking, intercept, and flight-plan calculations. For a 200-target system, two memories, each of 8192 words of 40 bits, would be required for these functions, and another memory of 4096 words would provide for identifying symbols. More than one computer may be used together with auxiliary units having fixed programs for routine operations.

A digital encoder gives the changing azimuth position of the antenna as a train of pulses with identification of a reference position. From this train, X and Y digital data are generated. At the plan-position indicators, the digits are converted to X and Y analog signals to deflect the electron beams.

Radar echoes are correlated as a function of the number of hits, pulse length, range, and azimuth to produce "clean" digital signals for transmission to the indicators to brighten the beams at the target positions.

Symbol generators provide X, Y, and beam-brightening data for producing alphanumeric and geometric figures on the indicators to identify targets. The figures are added between radar scans, their positions being originally determined by the operator and moved along with the target by the computer.

The four following papers in this issue of Electrical Communication by members of the staff of Standard Radio & Telefon present details of the major elements of Digitrac and are extensions of the paper "Digitrac for Handling Radar Data," by S. M. Eriksson, which appeared in the preceding number, volume 39, number 3, 1964.

Digitrac Video Correlator for Binary Detection of Radar Targets

KJELL MELLBERG
Standard Radio & Telefon AB; Bromma, Sweden

1. Introduction

The growing complexity of air traffic—both civil and military—has forced air-traffic-control operators to rely more and more on automatic means to prepare surveillance information. Radar is the best means of locating airborne targets, and the plan-position indicator translates this radar information into understandable form. However, human control of air traffic based on observing displays is not sufficient today, because of the high speeds and large numbers of aircraft. Automatic equipment has to be introduced for the routine work, thus freeing the operators for special surveillance and control tasks.

Advanced digital computers can easily assume many types of routine work and also perform certain tasks better than humans (for example, target tracking, track-development estimations, and anti-collision computations). However, the introduction of digital computers in an air-traffic-control system also requires a means of conveying radar information to the computers. Naturally this could be done by the operators with the help of input devices such as keyboards and marker devices (rolling balls, joysticks, et cetera), but this method is normally not good enough. Good target tracking requires fast and accurate target position measurements directly on the radar signal and preferably in real time, which are then supplied to a tracking computer without human intervention.

To fulfill the above requirements, there is a need for equipment between radar and computer to convert analog radar signals to digital form and to process radar signal information. The aim of this processing is to eliminate noise, interference, and signal redundancy, and to leave only extracted target position information, which is delivered to the tracking computer. The target detection involved in this plot extraction normally works on an input video signal of several-hundred kilohertz. Because of this high frequency, it is economical to use a special unit adapted for this type of processing rather than to perform target detection in the tracking computer. Such a target detector may also serve as a noise and interference filter, giving the operators a clear picture on their displays.

Some years ago a digital automatic target detector was designed at Standard Radio & Telefon for digital radar data handling systems. This unit is called a video correlator.

2. Principles of Target Detection

Target plot extraction can be divided into two functions: target detection and the assignment of plot values to a detected target. As will be shown later, plot-value assignment is a simple matter once the time of appearance of a detected target is determined. Target detection is therefore the main problem in processing the radar signal.

Pulse radars give at least two characteristic features to echoes from objects that are small compared with the length in space of the transmitted radar pulse. One is the relatively constant length of the echo, and the second is the correlation in time between successive echoes from the same object. The echoes from successive radar pulses form an echo pattern at a fixed radar range. Such a pattern appears because several radar pulses are normally transmitted during the time that the antenna beam sweeps across the reflecting object. The radar ranges of the echo pattern are fairly constant even for a moving object because of the relatively short time for the antenna beam to pass the object.

These two features can be used to distinguish echoes of small targets from noise and other disturbing signals, which do not exhibit the characteristics mentioned. The echo-length characteristic is normally used by matching the radar receiver bandwidth to the pulse length as closely as possible; this reduces high-frequency

noise but passes long echoes from extended reflecting areas such as ground and clouds. Such long echoes may be eliminated by various techniques of pulse-length discrimination, which is desirable in many cases despite the slight risk of losing true targets that are close to each other.

Correlation may be obtained by integrating the video signal in different ways. The integration must always be synchronized with the radar pulse repetition frequency, so that signals of equal return time in successive radar pulse periods are added together. A well-known video integrator is the plan-position indicator, which performs echo intensity integration by the persistence of its phosphorescent screen. This screen, however, is not an ideal integrator because the picture decays somewhat between successive hits, and furthermore such hits are slightly separated because of the rotation of the intensifying sweep.

Ideal integration can be performed in a delay network (Figure 1), in which the target echoes of N most-recent pulse periods are added. It is well known that in such an ideal integrator, correlated equal signals produce a sum signal that increases linearly with N the number of integrations, whereas noise signals add only in proportion to the square root of N because of the random time and power distribution of such signals. Integration thus enhances the signal-to-noise ratio.

After integration, target detection is facilitated by use of a threshold circuit, through which only signals exceeding a predetermined level may pass. The level is so chosen that an optimal ratio is obtained between target-detection probability and false-signal probability.

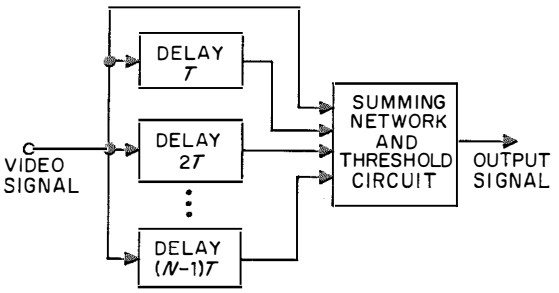


Figure 1—Ideal video integrator with integration cycle of N pulse periods of duration T.

In automatic tracking it is normally necessary that the false-alarm rate be low enough that it does not confuse tracking or overload the tracking computer. Such considerations determine the threshold level, and therefore the detection probability depends on the effectiveness of the integration. The integrator of Figure 1 is difficult to realize in practice for large values of N; furthermore, integration is not optimal unless N equals the number of successive correlated echoes received by the radar from a target. In practice, however, this number is seldom constant. A practical and widely used solution is a single delay line with positive feedback. Although fairly effective, this integrator is not ideal because of unavoidable losses in the feedback path.

Ideal digital integration would be possible if the video signal amplitudes were converted to digital form and stored in a memory for each range distance. Signal values from corresponding range distances could then be added in a digital way. Recent advances in high-speed circuits and core memories have made this idea quite practicable. The digital method has several desirable features.

(A) Great flexibility of control. The integration may be examined at any time without being affected. Among other things the optimal number of integrations may be determined and the center of the echo pattern calculated.

(B) Reliable and stable processing without signal deterioration.

(C) Flexible synchronization. The radar trigger frequency may be allowed to vary at random, which is often desirable.

Unfortunately, the ideal integrator becomes expensive if it is constructed with digital circuits. This requires amplitude numbers of several bits per time quantum (corresponding to the radar pulse length) and a high-speed analog-to-digital

converter. Although it is quite possible to design a memory of the required large capacity, as well as the critical circuits of such a converter, it cannot be done economically today. An economical simplification of the ideal digital integrator is possible without much loss in effectiveness.

3. Binary Correlator

The simplest type of correlator is the so-called binary correlator, which uses only two video amplitude values, signal or no signal. Thus only one bit per time quantum is needed to describe the video signal. At the correlator input the video signal passes through a threshold circuit, where signals above the threshold level are assigned the value of a binary *1*, and signals below this level a binary *0*. The integration in the correlator is then performed by counting the number of binary *1*'s in correlated time quanta. When the number of *1*'s (as counted for a number of pulse periods corresponding to the number of hits per target) exceeds another threshold level, a target signal is delivered. Because the correlator has in effect two signal threshold levels, it is sometimes called a double-threshold detector.

The rough method of quantizing the video signal amplitudes into only two values would seem to decrease the integration effectiveness considerably. However, theoretical studies show that the binary correlator is inferior to the ideal integrator by only about 1 decibel if their integration efficiencies are measured in required signal-to-noise ratios of the video signal for the same integration results.

The correlator is better than the ideal integrator in some cases, for example when strong noise pulses are received. Such pulses may give amplitudes in the integrator that exceed the detection threshold level and thus are taken for targets. The correlator eliminates such false targets, which may be caused by radar interference or pulse jamming.

4. Binary Moving-Window Correlator

To make the digital processing of video information possible, it is first necessary to quantize the information in time quanta. Within each time quantum the information is further quantized in amplitude to standard binary signals, as mentioned in the previous section. For optimal integration the time quanta should be approximately the same size as the radar pulse length. If time is translated into distance, a range quantization is obtained, with a resolution, for example, of 300 meters (1000 feet) for a 2-microsecond radar pulse. In the azimuth direction a natural quantization is obtained in angular increments between successive pulse periods.

Thus any part of the surveillance area is divided into a network of quanta, as illustrated in Figure 2. These quanta are marked with signal amplitudes in the form of binary numbers. Two of the ranges of Figure 2 show a markedly high density of approved signals. Using the assumption that the radar beamwidth would normally permit, say, 6 hits per target, each encircled set of signals in Figure 2 would very probably be reflected from a true target. A special criteria logic in the correlator is used to determine whether a set of signals fulfills the signal density criteria for the specific radar. A typical set of correlation criteria may have the following form for a radar with a maximum of 6 hits per target.

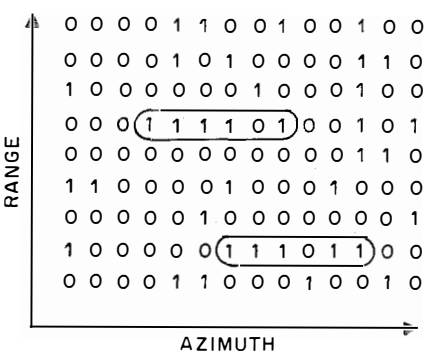


Figure 2—Pattern of target information in core memory matrix. Two probable targets are encircled.

(A) Low limit. At least 3 quanta out of 6 must have a *1*. Any lesser density is considered to be noise.

(B) High limit. Not more than 8 quanta out of 9 may have a *1*. Any higher density is considered to come from ground, clouds, et cetera.

As shown in Figure 2, binary numbers could be stored in 1 shift register per range quantum. The video information for that quantum moves through the shift register as the antenna turns. The register acts as a window moving with the antenna. Therefore the technique is sometimes referred to as the moving-window principle. Obviously it is not economical to use one shift register per range quantum when it is often desirable to cover the radar range with up to 1000 such quanta. In principle, therefore, only one register is used in time multiplex, the information for the other range quanta being stored in a magnetic core memory.

Figure 3 illustrates the principle. The memory contains a number of addresses, one for each range quantum. The addresses are scanned sequentially during the time between adjacent radar transmitted pulses, so that a target return signal from a given distance always coincides in time with the scanning of the same address. The window information within each memory address is successively supplied to a flip-flop shift register, where new information from the radar is added and shifts the previous

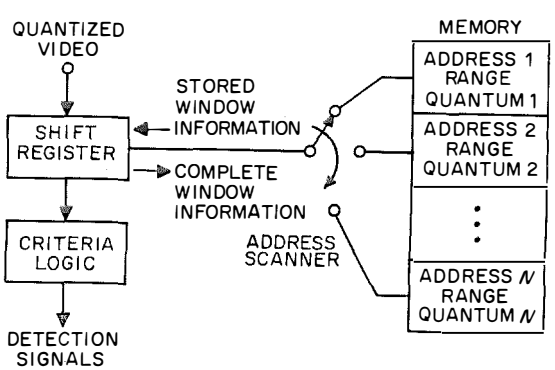


Figure 3—The principle of storing window information in a core memory.

information one step. The contents of the shift register are examined by the criteria logic to determine if a target exists. The logic sends out a detection signal on recognizing a target. The newly completed information in the shift register is then written back into the same location in the memory, the next location is searched, and so on.

To be able to use both the low-limit and high-limit criteria, one obviously must make the number of bits in the window large enough to cover the maximum number of hits per target expected from the radar. Thus for a large number of hits, the shift register and each memory address need a large capacity. A more-economical solution is shown in Figure 4. The basic principle of this solution is to count the number of hits in a binary counter; in this instance the required number of bits for detecting a target per range quantum is proportional only to the binary logarithm of the maximum hit number.

However, to make the correlator work as an ideal integrator, the counter must be supplied from a start/stop logic that determines when an echo pattern starts and stops. This logic may work on a shorter shift register, by means of which it is possible to study the echo density of the incoming signals. The counter is started when the density exceeds a certain start criterion, and is stopped when the density decreases below a stop criterion. The number of counted echoes between start and stop is then compared with the low-limit and high-limit criteria specified for the total echo pattern. A detection signal is delivered from the correlator only if the counted number is between these limits.

5. Plot Extraction

The detection signals from the correlator may be used both as computer input data and as processed video for the plan-position indicator. In the former case, plot values in the form of position coordinates must be determined for the detection signals. This is easily done if the plot is given as range and azimuth values seen by

the radar. The sequential addresses of the correlator memory are given from a binary counter started by the radar trigger pulse. The setting of this counter is proportional at any instant to the radar range of the corresponding memory address, and may thus be sensed by a detection signal to give the range value of the corresponding target. The azimuth value at any instant likewise may be obtained from another counter, which receives pulses from an incremental encoder rotating with the radar antenna.

It is desirable for both computer use and presentation to have the azimuth of the center of a target as a result of the plot extraction. For computer use this is no problem, because the angular distance between start and stop signals may be given by the correlator, and a mean azimuth value can be computed. Another method, which gives the target-center azimuth for presentation also, is to delay the plot extraction by an angle corresponding to the maximum angular distance between a target center and the point corresponding to a stop signal from the correlator. In this case variations of the echo-pattern size do not cause any displacements of the azimuth values. The constant delay is easily obtained as follows: When a stop signal is given for an echo pattern in the correlator between low and high limits, the echo counter is advanced two steps per radar pulse period. When the counter reaches the specified maximum of hits, a detection signal is delivered, the delay of which is constant as measured from the time the radar beam center was pointing at the target. The azimuth center values are then obtained if the alignment of the azimuth unit is displaced by an angle corresponding to the correlator delay.

6. Presentation of Several Radars Simultaneously

If correlated video is used for plan-position-indicator display, the operator sees a clean picture rich in contrast, which simplifies the evaluation. Figure 5 shows half the display disturbed by pulse jamming. The correlator is switched on for the other half of the antenna rotation cycle, giving a clean picture for that period.

For mere presentation it is not essential to have a complete correlator. It would be relatively effective to show all hits received between the start and stop signals of the correlator (that is, the echo counter is not used). Normal echo patterns would be shown on the display, but interference and pulse jamming would be cancelled and noise reduced fairly well.

Use of the complete correlator has several advantages, however. The digital nature of the correlator memory permits it to be used for other purposes besides correlation. One is the storage of a detection signal of any range increment for a certain time. This capability can be used for time-sharing data from several radars on one display. For example, if two radars are to be displayed simultaneously, the display sweeps each radar during alternate pulse periods. If one radar gives a detection signal while the other radar is being displayed, the signal is stored until the first radar is available for display again. Thus no information is lost. Some echoes on the presentation will be shifted in azimuth because of the waiting time, but for

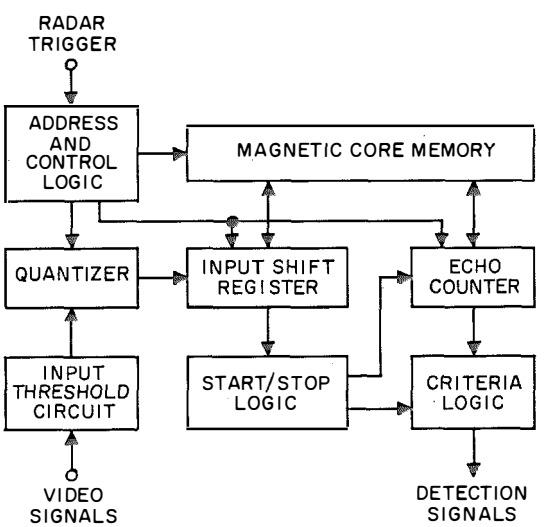


Figure 4—Economical organization of the binary correlator with echo counting.

surveillance this is not very important unless too many radar outputs are being displayed.

Another possible use is to let the detection signals indicate X/Y plots. X/Y values of the radar sweep may be formed continuously by integration in counters of the sweep pulse trains delivered by the digital azimuth and sweep unit [1]. The X/Y values are supplied directly to a digital display as dots. This type of presentation is very effective because it loads the display only with information. The large amount of dead time between presentations can be used for displays from other radars, for marker and identification symbols, for maps, et cetera.

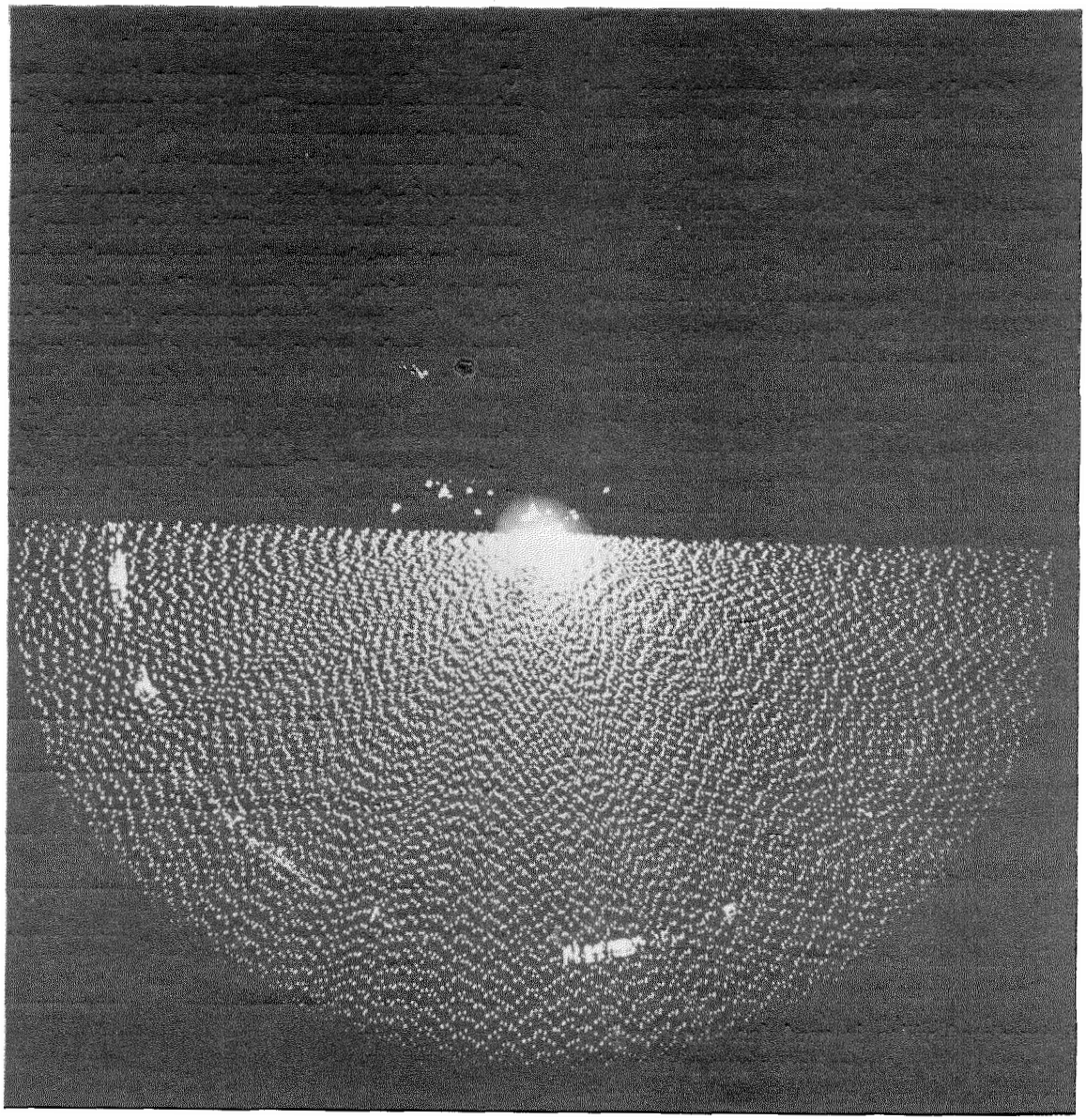


Figure 5—Pulse jamming of radar. The correlator is switched on for half of the antenna rotation cycle, giving a clean picture for that period.

7. Conclusions

The video correlator performs several tasks in a modern radar surveillance system. It acts as a noise, interference, and jamming suppressor for both operator evaluation and automatic data processing. Jamming can be suppressed by random radar triggering. Furthermore, the correlator is an accurate measuring device for digital target plotting, which is essential to automatic tracking. The digital output of the correlator makes possible virtually any type of data processing that may be required.

A very-important processing example not mentioned before is digital narrow-band transmission of radar information. The correlator performs the necessary data reduction and analog-to-digital conversion of radar information to make such transmission possible. Our Natrap system can be used for radar data transmission via telephone lines without loss of information accuracy.

In summary, the correlator is an important instrument to convert radar information to a form suitable for large-scale data processing, which in turn is necessary to control our increasingly complex air traffic.

8. Acknowledgments

The video correlator was designed in detail by Mr. A. Lindholm, who also tested it extensively in cooperation with modern radar stations.

9. References

1. B. Johansson, "Digitrac Display System for Air-Traffic Control; Part 1—Digital Azimuth and Sweep Generation," *Electrical Communication,* volume 39, number 4, pages 495–498; 1964.

2. "Automatic Digital Target Detector for RAE Radar-Data Transmission System," *Civil Aviation Radio News,* number 30, pages 18–24; June 1959.

3. J. I. Marcum, "A Statistical Theory of Target Detection By Pulsed Radar," IRE Transactions on Information Theory, volume IT-6, number 2, pages 59–268; April 1960.

4. L. Gérardin and R. Barbier, "Extraction Automatique d'Informations Radar, Sous Forme Numérique," *L'Onde Électrique,* volume 41, pages 783–794; January 1961.

Kjell Mellberg was born in Angermanland, Sweden, in 1930. In 1956 he received the degree of Master of Science in Electronics at the Royal Institute of Technology in Stockholm.

In 1956 he joined Standard Radio & Telefon in the new digital group, where he worked on the development of digital data transmission equipment. After nearly a year of military service as a staff specialist on air defense, he became leader of the Standard Radio & Telefon group working on the design of automatic air-defense equipment. In 1961 he was appointed head of a new section for computer engineering, and has since been in charge of the design and development of computer systems and computer-type devices for automation.

Mr. Mellberg is a member of the Swedish Association of Engineers and Architects.

Digitrac Censor Data Processor

KJELL MELLBERG
Standard Radio & Telefon AB; Bromma, Sweden

1. Introduction

The digital data processor Censor (**CEN**tral proces**SOR**) is a parallel purely binary machine with a random-access magnetic core memory as an internal store. It can be characterized as a multipurpose computer primarily intended for real-time data processing in control systems. Thus the machine has features that make it very fast in such applications. These features include data bus lines, a semipermanent instruction memory with microprograming facilities, an extensive list of microinstructions, and priority handling of input/output equipment with both program selection and direct access to the memory.

2. Bus Line

A common bus line provides a 2-way information channel connecting all parts of the Censor system (Figure 1). It transfers data and address information between the Censor central unit and external devices, as well as between the external devices themselves. The maximum 2-way transfer rate is 166 000 words of 40 bits each per second. Extra memories similar to the internal one (each of 2048 or 4096 words of 40 bits) may also be connected to the bus line.

A Censor system can also be connected to one or two similar systems via an information gate between the respective bus lines as shown in Figure 2, which also indicates how the arithmetic unit may be connected to the other elements.

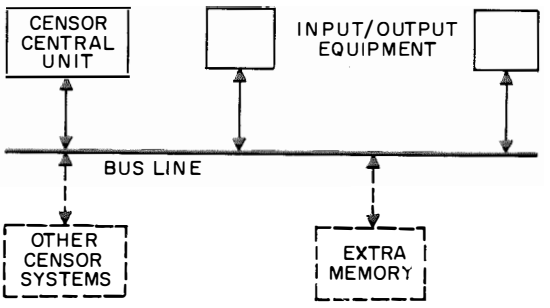


Figure 1—Censor organization.

In this way different systems may share the workload with rapid interchange of data. External equipment in one system may also communicate directly with another system.

3. Direct Access to Memory

Rapid data interchange between external equipment and the central unit is of prime importance in many real-time applications. For nonsynchronous access to the Censor central unit, the program-interruption method has been considered too slow because of the relatively long time needed to preserve data from an interrupted program. Therefore a direct access to the memory has been provided.

To communicate with the memory, an external unit or the arithmetic unit sends a priority call to a memory call priority unit in the Censor, where the relative priorities of all possible calls are stored. In accordance with priority, the calling units are connected to the memory. An external unit having priority can be connected to the memory in as few as 10 microseconds after calling. Such rapid transfers minimize the need for buffer storage in the external units.

The direct-access memory also has the advantage that transfers to and from external equipment can be made without disturbing computations in the arithmetic unit. In the Censor this facility is especially useful, as fixed program subroutines may be run in parallel with the communication between the memory and the external equipment (Section 5).

The number of external units that may be connected to the bus line is 240.

4. Program Selection

In various applications certain external units only write data in or read data from specific memory locations. These data are processed in independently timed program sequences. However, some external units may also require that special programs be run in conjunction with data transfers, or simply for updating purposes.

A common way to organize such multiprogram work is to use the program-interruption technique. This method, however, requires appreciable time to preserve information. Therefore programs normally are not interrupted in the Censor, but are allowed to run their course before a new program is started. In real-time applications, however, there are seldom long programs. The program access time will still be sufficiently short for most cases. The important factors in these applications are normally (A) the amount of work that the machine can perform in a given time and (B) how long external units will have to wait for access to the central unit. The former is increased considerably in the Censor because of the lack of time-consuming interruption operations, and the latter is kept to a minimum thanks to the direct-access memory.

However, if there are programs of such length that the waiting time would be too long, interruption orders can be inserted at convenient intervals. Such an interruption order causes suspension of the program by a program selector. When the higher-priority program is completed, the interrupted program is resumed.

All programs are executed on the basis of priority, which is stored in the program selector. This unit can be built for a maximum of 32 so-called primary-program calls, stored in a priority register. The call signals may be delivered from external units or from a program in progress. Normally, the address in the program next in priority is given by the program selector immediately after the running program is finished. If more than 32 programs are wanted, the 32nd position of the program selector may be used to divert the call to a portion of the memory where start instructions for such secondary programs are stored.

The program selector is also equipped with an internal program modifier, in which changes in program sequences can be initiated by special markers in the processed information. Such markers may be used to announce that special results are obtained in the computations.

5. Special Programing Facilities

A memory word has a length of 2×20 bits. The arithmetic unit has a word length of 20 bits and may make use of either of the memory half words via the data switch shown in Figure 2. Thus it is possible to store two arithmetic words in one memory word. For programing purposes the memory word may be organized in two ways (see Figure 3). Within each memory word, the left 20-bit half word always contains an 11-bit operation code and 5 bits for determining the distribution of the right half word to different registers in the arithmetic and control units. The right 20-bit half word may contain either address to an operand or, in special cases, the operand itself.

The address half word contains one bit for half-word markings, 12 bits for addressing the basic 4096-word memory, 3 bits for memory extensions up to 32 768 words, and 4 bits as spares for possible special features, such as character addressing or indirect addressing.

The operation part of an instruction may also

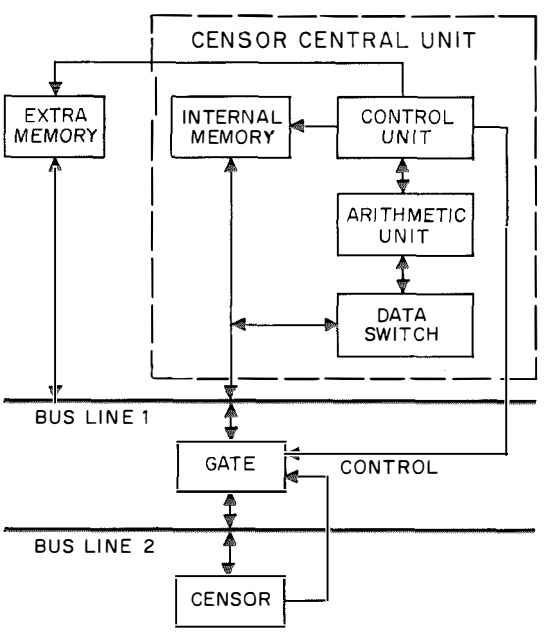


Figure 2—Flexibility of expansion.

be taken from a separate semipermanent 512-word instruction memory in the control unit, where subroutines may be placed. Such subroutines are executed very rapidly because of the high speed of the semipermanent memory (cycle time, 0.75 microsecond). They may also constitute new orders in the instruction list (macro-orders), and for each application may be constructed by the programer (micro-programed) by manually inserting plugs in the instruction memory. One of the basic orders of the instruction list always refers to the instruction memory by a special marker in the code. The rest of this code (8 bits) may be given any even address code of the instruction memory. In this way a subroutine in the instruction memory may be ordered in the same way as a normal order of the instruction list.

To decrease the need for references to the slower normal memory (cycle time, 6 microseconds) during execution of macro-orders, the arithmetic unit is provided with an extra storage register for partial results. The memory in such cases is largely free for communication with external units. Both the memory and arithmetic units are thus used in parallel for different purposes.

The control unit also contains 64 easily changeable numerical constants in a semipermanent constant memory.

The Censor has two address registers, one serving as a program counter and the other as an operand address register. The latter may also be used to address a second computer. To communicate between the memory and external equipment during the running of arithmetic programs, the address part of the external bus line is connected directly to the address circuits of the memory. The Censor in effect has a 3-way address system.

6. Central Processor

Figure 4 is a block diagram of the Censor. The main units are the memory, the arithmetic unit, and the control unit.

The memory is a conventional coincident-current ferrite-core type with 4 modes of operation; read/restore, clear/write, buffer read, and buffer write. The latter 2 are half-cycle operations. The data lines are connected to the external-data bus line and to the instruction bus line in the control unit.

The address lines are connected to an address switch, which chooses a memory address from one of 3 sources: the address bus line for external communication, and 2 address registers for program execution. The address switch is controlled by the control logic.

The arithmetic unit consists of a 20-bit parallel adder and four 20-bit registers: the memory buffer (multiplicand) register, the accumulator register, the multiplier register, and an intermediate storage register for high-speed storage of partial results in computations. In 2-operand operations the operands are placed in the memory buffer and accumulator registers before execution; in single-operand operations

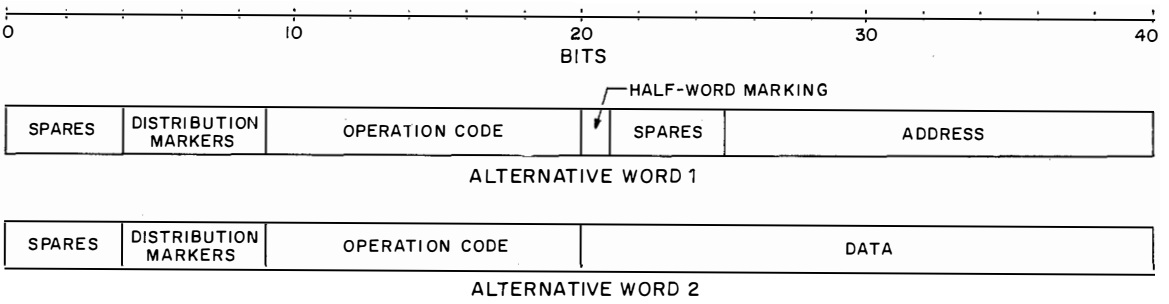


Figure 3—Alternative instruction words.

they are placed in the accumulator. Operation results are always obtained from the accumulator.

The multiplier register can be considered an extension of the accumulator in shift operations (bidirectional shifts) and multiplication, and acts as a quotient register during division. It may be used as a separate register in special applications. This also holds true for the other arithmetic registers because they and certain elements of the control unit are interconnected by an internal data bus line, which permits direct transfers between any two registers connected to it.

The arithmetic unit communicates with the external data bus line via the internal data bus line and the data switch. In this way the arithmetic unit may communicate with any external unit and the memory. The data switch places a 20-bit arithmetic word in the appropriate half of the 40-bit memory word.

The control unit contains two address registers, the memory call priority unit, the program

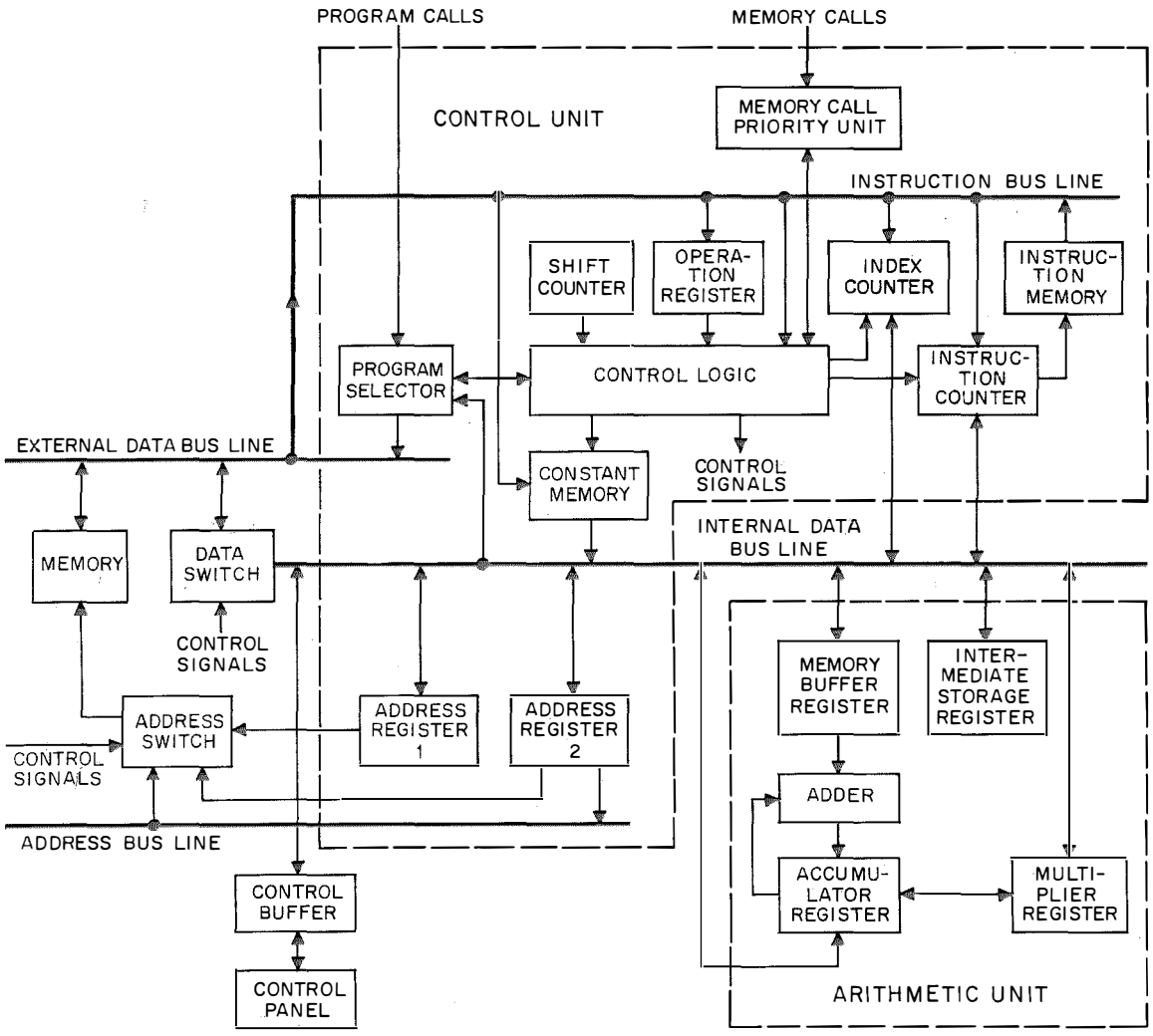


Figure 4—Block diagram of central unit.

selector, an instruction memory with instruction counter for fixed programs, a constant memory, plus control registers and control logic.

Address register 1 is used as an address counter for the normal scanning of memory addresses during the sequential execution of program instructions. Address register 2 is normally used as a storage unit for the operand addresses given by the instructions. It can also be used to address external units and another computer via the address bus line.

The memory call priority unit controls the time of access to the memory from different calling external units and from the program. All connected units have priorities to determine the order of processing their calls. During program execution external units may communicate with the memory without breaking the address sequence of a running program. This is possible thanks to the third means of addressing mentioned above, that is, addressing from external units via the address bus line. During external communication the program then merely has to wait for access to the memory until the memory cycle of the external transfer is completed.

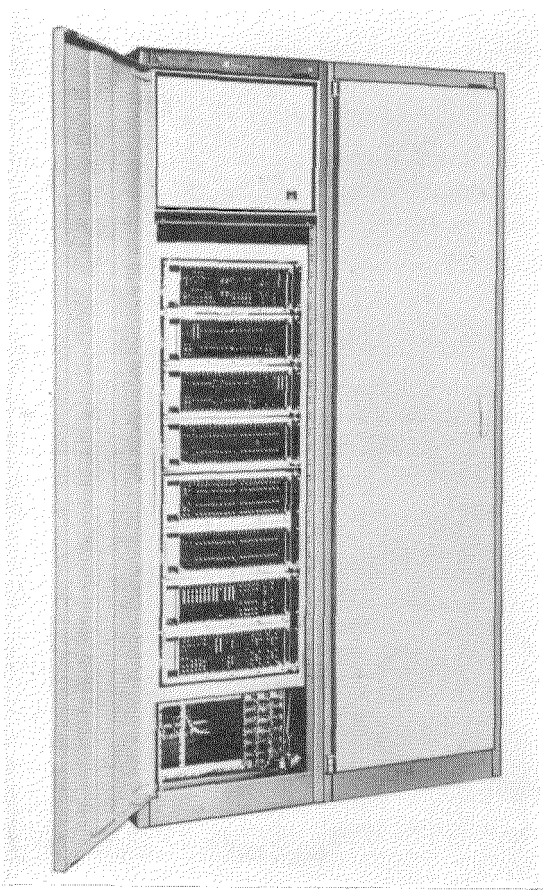


Figure 5—Central unit with one door open.

The control registers consist of an 11-bit operation register, a 5-bit shift counter, and a 9-bit index counter. The operation register is set with the operation part of an instruction. The operation code is decoded in the control logic, from which control signals are delivered to all affected parts of the system.

The shift counter is counted directly for instructions during the shift operation and is continuously compared with the shift order in the operation register. When the two units match, the shifting stops. The index counter is used to count repeated program loops (for example, in iterations and scanning), but may also be used as a storage register for special purposes because it is connected to the internal-data bus line.

The operation orders may also be taken from the instruction memory, where command sequences may be stored semipermanently. The instruction memory consists of 512 diode gates of 11 bits each. Each gate is coded by the insertion of a plug with prewired connections in accordance with the desired code. The diode gates are addressed by an instruction counter having a decoding network.

The instruction counter may be set to a diode-gate address from either the internal data bus line or the instruction bus line. The latter is supplied directly from the core memory. By appropriate operation coding, it is thus possible to start an instruction sequence (macro-order) from the core memory by giving the start address of the sequence to the instruction counter.

The operation register is always set from the instruction bus line. This bus line may be set either from the core memory for stored memory programs or from the instruction memory for semipermanent programs. The choice is made by appropriate operation coding.

The constant memory is constructed in the same way as the instruction memory. It contains up to 64 plug-in units, each corresponding to one numerical constant, which may be looked up for use in any type of program.

The program selector delivers the start address of any selected new program. The program may be placed either in the core memory or in the instruction memory. New program orders may be given to the program selector from either external units or a running program.

For supervision and manual control, a control panel is connected to the internal data bus line and instruction bus line via a control buffer that contains buffer registers, a program for scanning, and lamp amplifiers. Lamps on the panel can show the contents of any register connected to the internal data and/or instruction bus lines. Further facilities are insertion of instructions and data and step-by-step execution of programs.

TABLE 1
GENERAL SPECIFICATION

Characteristic	Rating
Core-memory capacity in words	2048 or 4096, expansible to 32 768
Memory-word length in bits	2 × 20
Memory cycle time in microseconds	6
External-bus-line word length in bits (parallel)	40
External-bus-line speed in words per second (2-way transfers)	166 000
Instruction word in bits (one address):	
Operation code	11
Markers	5
Address including half-word marker and spares	20
Number of basic orders in instruction list	≈100
Word length of arithmetic unit in bits (parallel)	20
Point working	fixed
Clock frequency of arithmetic unit in megahertz	5.33
Elementary operation time of arithmetic unit in microseconds	0.75
Word capacity of semipermanent instruction memory	512
Instruction-memory word length in bits	11
Instruction-memory cycle time in microseconds	0.75
Capacity of constant memory:	
Words	64
Bits per word	20
Maximum number of connected external units	240

7. Input/Output Equipment

The Censor central unit (Figure 5) is provided with a special buffer unit for communication with computer-oriented input/output equipment, such as tape readers, tape punches, printers, and magnetic-tape units. The task of the buffer is to change sets of input character formats to memory words, and conversely to change memory or arithmetic words to sets of output character formats. Thus in many cases input/output equipment may communicate directly with the memory without using the arithmetic unit, a facility that is convenient in on-line operations. The buffer contains a special block-transfer program that acts on whole blocks of information in the memory. For example, if a program results in a large

TABLE 2
COMPUTER SPEED

Operation	Speed in Microseconds		
	Semi-permanent Program	Stored Program*	
		Nonaddress Instruction	Complete Instruction
Addition	0.75	6	12
Multiplication	15.75	23.25	30
Division	16.5	24	30.75
Binary-to-decimal	16.5	24	—
Root extraction	31.5	39	—
$(x^2 + y^2)^{1/2}$†	68.25	75.75	82.50

* Includes time for the memory logic to determine whether the arithmetic unit or external units have demanded access.
† Macro-orders of the type shown are stored in the semipermanent instruction memory and can be ordered from the working memory.

block of data to be printed, the whole block is stored in a specific memory location. Next a print order is given to the input/output buffer together with the start location in the memory and the number of memory words to be printed. The buffer starts printing and takes print information from the memory block by successive memory calls. In that way a program may be finished without being influenced by the printer speed, and when the print order is given the computer may start other programs without being disturbed by the print activity.

8. Capabilities

The characteristics of the Censor computer are given in Table 1 and its overall performance in Table 2. It may operate with the following types of external equipment: tape readers, tape punches, magnetic-tape units, printers, data-link equipment, character generators, plan-position-indicator displays, keyboards, radar-signal digitizers, analog-to-digital and digital-to-analog converters, et cetera.

9. Acknowledgment

The Censor computer has been developed on remarkably short notice and is the result of first-rate teamwork in the data-handling section. Credit for the materialization of the Censor concept goes to Mr. Henry Larson, group leader of systems engineering, and Mr. Sture Jansson, group leader of computer design.

Kjell Mellberg. Biography appears on page 487 of this issue.

Digitrac Display System for Air-Traffic Control
Part 1—Digital Azimuth and Sweep Generation

BENGT JOHANSSON
Standard Radio & Telefon AB; Bromma, Sweden

1. Introduction

The most frequently used way to send azimuth information from the antenna to the plan-position indicator is to use a synchro transmission system. At the receiver end a sine-cosine generator or resolver transforms data to Cartesian coordinates. Analog data processing is used. By careful synchronization of the rotary components of the system, a relatively high degree of accuracy is obtained.

Digitrac uses a digital computer and digital display unit, which makes digital processing of azimuth data the logical choice. The azimuth and range data not only control the display but also provide a continuous flow of information in binary form to the computer. The only two places where the azimuth information appears in analog form are at the antenna shaft and at the input to the deflection coils of the indicator.

Significant features of the azimuth processing (see Figure 1) are the use of a digital azimuth transducer and of an electronic flywheel principle, which makes use of the inertia of the antenna and gives a high degree of noise suppression and azimuth resolution.

2. Azimuth Transducer

Azimuth information is picked off at the antenna by an incremental 2-track encoder that rotates synchronously with the antenna shaft. Theoretically this makes gears unnecessary, thus eliminating gear inaccuracy and backlash, but for practical reasons a 1:1 gear train must be used in most cases. The encoder is of the photoelectric type with built-in photocell amplifiers. The lamp is easy to replace.

One encoder track gives 2048 equally spaced pulses per antenna revolution, with a mark-to-space ratio of approximately 50 percent. This arrangement corresponds to a tachometer, giving information on the momentary rate of rotation. Each edge of the individual counts, however, also gives very-accurate position information, the maximum linearity error being less than 1 minute of arc. This deviation includes all readout errors within the encoder and photocell amplifier.

The other track delivers one count per antenna revolution. This alignment pulse has the same width and angular position as a corresponding pulse in the first track. It provides a reference point for the regenerated azimuth at the data center.

3. Azimuth Pulse Integrator

3.1 Introduction

After transmission to the data center, the encoder tachometer pulses are counted to give an azimuth scale that changes incrementally as the antenna rotates.

This could be done in a simple binary counter that is zeroed once per antenna revolution by the alignment pulse, but the bearing resolution of the electrical angle would then be only 1/2048, or 0.18 degree of arc. This resolution would be inadequate and it could also give rise to a spoky display. A simple count of the

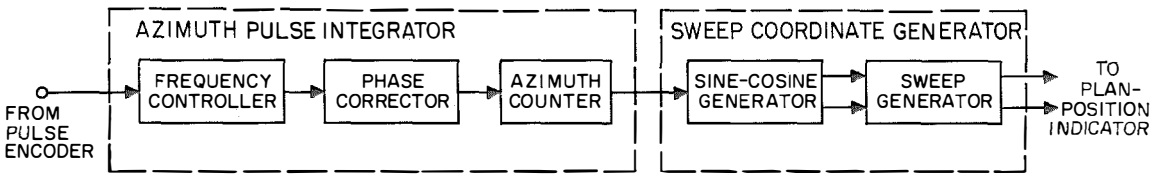


Figure 1—Azimuth and sweep generator.

incoming pulses would also suffer from susceptibility to disturbances on the transmission line or to slight imperfections in the encoder disk. Just one missed tachometer pulse would lead to a steady angular error for the rest of the revolution.

To overcome these troubles the electronic equivalent of a flywheel is incorporated to regenerate the azimuth pulse train and increase the number of pulses per revolution of the antenna over the number received from the antenna.

3.2 Electronic Flywheel

3.2.1 *Frequency Controller*

In the frequency controller of Figure 2, a stable oscillator provides a train of equally spaced pulses at a frequency of 280 000 pulses per second.

In the multiplier this pulse train is acted on by a coefficient that may be adjusted between 0.1 and 1. The output frequency from this multiplier should be the desired frequency of a new azimuth pulse train. To assure the desired frequency, the output pulse train is compared with the lower-frequency tachometer pulse train from the antenna.

The multiplier output passes through two flip-flop circuits to the frequency counter, the output of which is compared with the tachometer pulse train. Any error output changes the multiplying coefficient through the limiter, which permits the coefficient to be changed by only the least-significant bit despite the size of the frequency error. The sampling period and the setting of the limiter are chosen with regard to the inertia of the antenna assembly and of the maximum increment of correction that can be applied later by a phase corrector.

3.2.2 *Phase Corrector*

In the phase corrector the azimuth counter is set to zero by each antenna alignment pulse and counts the intervening pulses from the frequency controller that have passed through a correcting switch. Under ideal conditions for an absolutely constant rate of rotation of the antenna, the azimuth counter should reach the zero-reset point precisely when the alignment pulse arrives. This ideal condition never exists.

The correction applied in the frequency controller is intentionally limited to a certain fraction of the momentary frequency error, with the

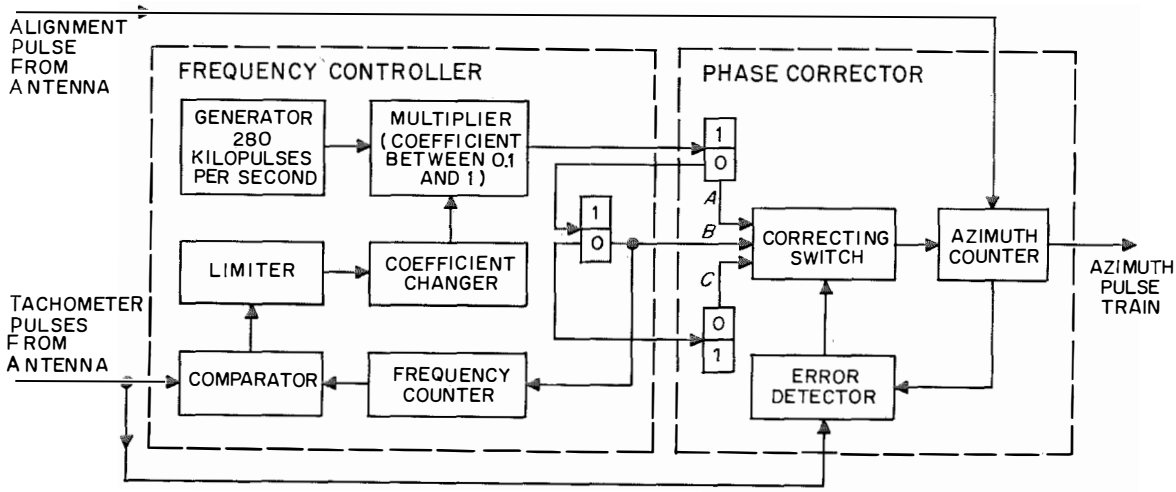


Figure 2—Azimuth pulse integrator.

consequence that a remaining frequency error could be integrated to a phase error. A phase corrector is therefore required.

Those pulses from lead *B* that correspond in timing with the tachometer pulses from the antenna are selected in the azimuth counter and compared in the error detector to produce an error signal.

If the error signal indicates a lag in the pulses from lead *B*, the correcting switch connects lead *A* to the azimuth counter for a limited time. As lead *A* pulses are at twice the frequency of lead *B* pulses, an increased number of pulses go to the azimuth counter to correct for the lag. A lead error connects lead *C*, which halves the frequency, to the azimuth counter to reduce the number of pulses.

The azimuth counter delivers a new azimuth pulse train having many more pulses per second than the tachometer pulse train, with which it is synchronized.

One might object that this again is a sampled, limited frequency control, and that the frequency controller might be omitted. However, it has been proved that the easiest way to preserve a certain degree of flywheel action, without the risk of the azimuth counter locking on the wrong input phase, is to make the correction in two steps.

4. Sweep Coordinate Generator

4.1 Sine-Cosine Generator

The flywheel action in the azimuth integrator results in a pulse train of considerably higher frequency than that of the antenna tachometer pulse train. This new pulse train represents the azimuth information with a high degree of resolution and precision. This information can be read out from the azimuth counter at any time.

However, on its way to the indicator the azimuth data must be converted to digital sine and cosine values of the azimuth, which are also needed as parameters for the computer system. To effect this conversion, the azimuth pulse train goes to two modulators, each of which has a counter at its output (see Figure 3).

The data processing now divides into two parallel channels, one for sine and one for cosine values. The modulators and counters are arranged in such a way that each counter sets the multiplication factor for the modulator of the other channel. Since the integral of a function and the function itself are identical for exponential functions, this arrangement, with proper adjustment of the start values and amplitudes, gives two pulse trains representing sine and cosine for the azimuth.

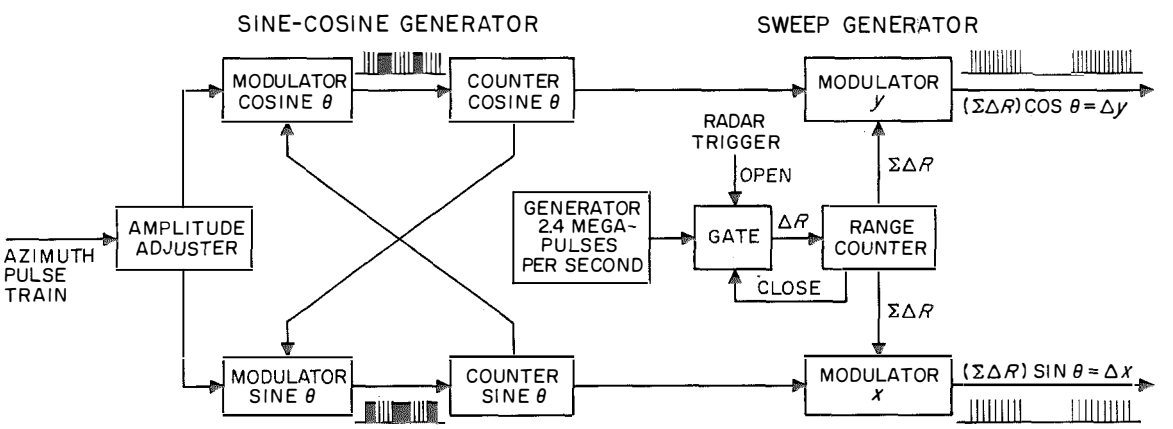


Figure 3—Coordinate and sweep generators.

4.2 Sweep Generator

Two modulators (Figure 3) multiply the sine and cosine values by the radar range sweep length, which has the form of a certain number of equally spaced pulses at a rate of 2.4 million per second. In this way batches of pulses are obtained in which the number of pulses is proportional to X and Y. These two trains of pulse bursts, which are originally triggered by the radar synchronizing pulse, are the two main outputs of the sweep generator.

The high-frequency pulse train representing the radar range sweep length is generated at 2.4 megapulses per second and is gated by the radar trigger pulse. It is integrated in a digital counter, the range counter. The number of pulses generated at any instant corresponds to the distance of a possible target to be presented on the display.

5. Conclusion

The pulse bursts generated by the sweep coordinate generator in synchronism with the radar trigger are supplied to the indicator units. Within each indicator a high-frequency binary counter acting as a sweep integrator processes the coordinate information.

This azimuth processing provides a high degree of precision and noise rejection. It also permits easy readout of azimuth, sine, cosine, and Cartesian coordinates in digital form, with a uniform degree of precision maintained throughout the system.

6. Acknowledgments

The principles for the flywheel function and sine-cosine generation were outlined by Messrs. C. O. Svensson and K. Mellberg. The author also wishes to thank Messrs. B. Alleryd and A. Wäin for fruitful collaboration during development.

7. References

1. Y. Lundh, "Digital Techniques for Small Computations," *Journal of the British Institution of Radio Engineers,* volume 19, number 1, pages 37–44; January 1959.

2. A. A. Voronov and G. N. Sokolov, "A Device for Programming Curves of Second Order, Based on a Digital Integrator," translated from *Avtomatika i Telemekhanika,* volume 20, number 2, pages 176–183; February 1959.

3. H. Z. Yang, "Determination of Maximum Error of a Binary Multiplier," *Automation and Remote Control,* volume 21, number 7, pages 709–713; February 1961. Translated from *Avtomatika i Telemekhanika,* volume 21, number 7, pages 1007–1014; July 1960.

Bengt Johansson was born in Sweden in 1930. He graduated in 1956 from the Royal Institute of Technology with a degree in mechanical engineering.

Mr. Johansson joined Standard Radio & Telefon in 1956 and now is head of the section for servo techniques.

Digitrac Display System for Air-Traffic Control
Part 2—Symbol Generation

BENGT SVENSON
TORSTEN HYLÉN
Standard Radio & Telefon AB; Bromma, Sweden

1. General

Symbols in the Digitrac system are produced by character generators. The total number of symbols required depends on the size of the system. A typical system, able to track about 100 targets simultaneously, needs three generators to produce all required symbols for the plan-position indicators and all data on the tabular displays. The workload is shared by using one generator for the raw-video indicators, one for the synthetic indicators, and the third for the tabular displays.

2. Interscan Presentation

All information concerning the display of symbols and data (for example, coordinates of footpoints,* characters to be selected, and order of distribution) is taken from the memory of the Censor administration computer. For the presentation all target numbers are scanned in the memory one by one. Each character generator has its own scanning program. For flicker-free pictures the repetition rate must be at least 14 per second.

The synthetic indicators and tabular displays can be fully loaded in real time with presentation of both symbols and data. These displays therefore use a symbol scanning frequency of 14 per second for each scanning program (for each target number). To achieve a flicker-free display of symbols on raw-video indicators using interscan technique is more complicated, because the indicator must be connected to different radar stations having different repetition frequencies. The symbol generator cannot be synchronized with a particular radar. Therefore the symbols must be supplied continuously to the indicators and presented during the interscan time of the radars.

Assuming an effective presentation time for the symbols of about 25 percent of the total scan period time, a scanning frequency of 48 per second is necessary for flicker-free presentation of the symbols. Because of the large number of indicators showing different symbols, plus the high scanning frequency required, the speed of the character generator is not high enough to permit flicker-free presentation of all symbols. Targets under close supervision therefore must be differentiated from less-important targets. The former are presented at a flicker-free repetition rate, the latter only twice per second.

3. Principle of Character Generation

Characters are generated by the dot method shown in Figure 1. The number of dots is determined by the bandwidth of the deflection amplifiers in the display units. In this case 16 dots are used to write each character. By integration in the deflection coils, a smoothly written character appears on the screen. Every dot of the character is represented by a binary

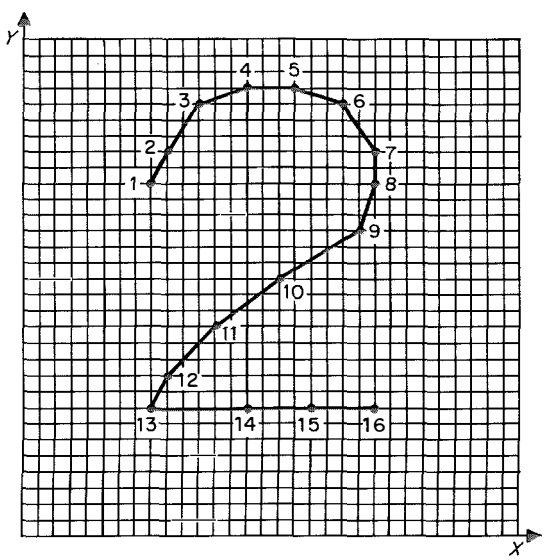


Figure 1—The principle of character generation.

* A footpoint coincides with the center of a target and is the originating point for the symbol.

number in X as well as Y directions. For good flexibility in placing the dots, 5-bit numbers have been chosen so that the 16 dots can be placed anywhere within a square network of 32-by-32 dots. The coordinates for each of the 16 dots of a character are stored in a semi-permanent memory designed as a diode matrix, each character occupying a separate printed card.

Binary values for the X and Y coordinates of the dots are taken from the memory and forwarded through a digital-to-analog converter to the displays. Individual characters can be arranged in groups of 3, as shown in Figure 2. The footpoint of the symbol (position 0) has the same X and Y coordinates as the target. It is also the starting point from which the entire symbol is written, the circle being centered around the target blip. To generate deflection voltages of a group of characters easily, the origin is moved from position 0 to position 1 before the writing starts. After completion of the letter, a new starting point for the circle (position 2) is chosen, then a third point (position 3) for the number, et cetera.

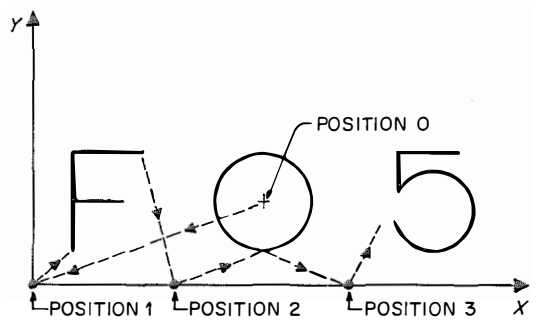


Figure 2—Writing a symbol group.

3.1 Character Generator

The character generator (Figure 3) consists of the following main units: a program unit, a 500-kilopulse generator, a scanning counter, a character shift counter, character matrixes, and two digital-to-analog converters.

The program unit of the character generator receives symbol data and a trigger pulse from the computer. The pulse starts a 60-microsecond delay circuit in the program unit. After

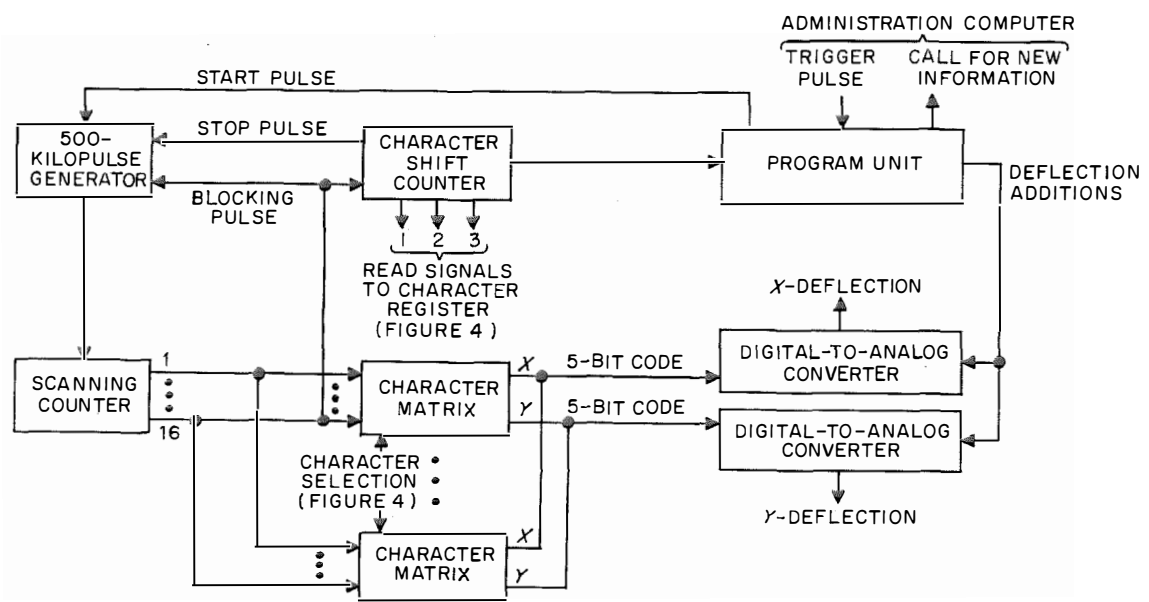


Figure 3—Functional diagram of character generator.

this delay the spot on the indicator has reached a steady state at the footpoint. After 45 microseconds, however, the character writing may be ready to start, and therefore the electron beam is moved to the starting point of the character deflection (position 1 in Figure 2). After the 60-microsecond delay, the program unit delivers a pulse to the generator, which steps the scanning counter forward at a frequency of 500 kilopulses per second. When the counter has reached state 16, a pulse is returned to the generator and blocks it for the following 15 microseconds, after which the generator starts again. Every time the scanning counter reaches state 16, it also steps a character shift counter forward. After this cycle has been completed three times, a stop pulse is returned to the generator.

3.2 Scanning Counter

The scanning counter consists of an ordinary shift register connected as a ring counter. Two microseconds are required to move the electron beam from one dot to the next. Hence 30 microseconds are required to write a 16-dot character. A whole symbol group with 3 characters thus requires 120 microseconds (3×30 for character writing, plus 2×15 for shifting between characters).

All character matrixes are scanned simultaneously in parallel. As only one matrix is used at a time, just one character is selected in each scan. The binary X and Y outputs from the matrixes, together with a voltage for deflection additions, are supplied to digital-to-analog converters that generate the deflection voltages for the character waveforms.

3.3 Character Selection

When the generator is ready to write a new group of characters, it calls the administration computer (Figure 4). Data are supplied via the bus line to a buffer store consisting of three 6-bit registers, one for letters, one for geometric symbols, and one for numbers.

Immediately before the oscillator in the character generator (Figure 3) receives the start pulse, the character shift counter gives a read signal on input 1, opening the associated AND gate (Figure 4). This permits the code to pass from the corresponding register to the character selector via an OR gate. The selector has 64 outputs, each output being connected to a character matrix. The character generator can thus write 64 different characters. When a letter has been written, a new read pulse is received on input 2 and the binary code for a geometric symbol is passed to the character selector. A number is correspondingly triggered by input 3.

3.4 Generation of Brightening Pulses

The plan-position indicators and the tabular displays are supplied with brightening pulses for presentation of characters in the same way as for presentation of targets on the indicators. Each character generator creates these pulses,

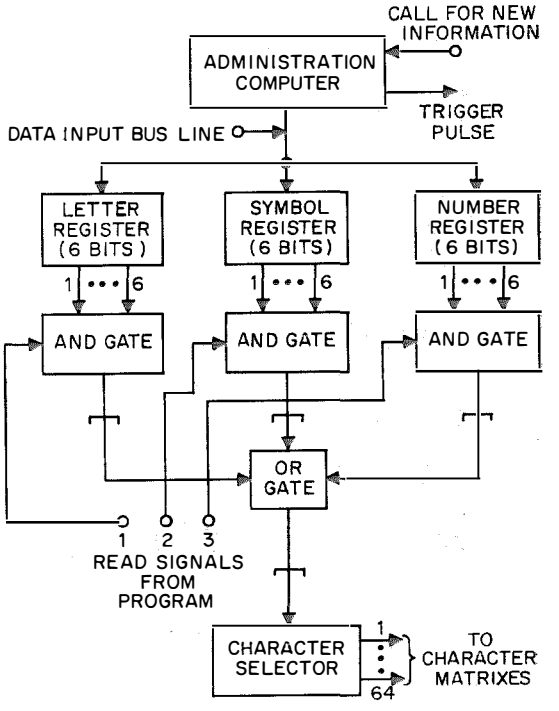


Figure 4—Functional diagram of character register.

which go to the display tubes via a gate system and distribution amplifiers. The gate system determines whether a symbol is to be displayed at a certain indicator.

4. Distribution of Symbols

4.1 Distribution to Raw-Video Indicators

The distribution of symbols among the operators in an air surveillance center, plus the division of symbols into categories of altitudes, speeds, et cetera, are organized by the administration computer as follows.

Each raw-video operator has at his disposal 2 bits per target-information cell in the computer memory (1 bit for a twinkling-display order and 1 bit for a flicker-free-display order). An operator can store symbol information in these memory cells and decide whether a particular symbol will twinkle, be flicker-free, or be displayed at all.

Each operator can also handle 2 flip-flops in the brightening-pulse register, one flip-flop for the twinkling symbol and one for the steady symbol. These flip-flops are set at the same time that the memory gives information to the symbol generator about the symbol to be displayed.

How the flip-flops are set determines how a symbol is presented. Each operator also has the option to watch all tracked targets simultaneously by pushing a button that twinkles all symbols on his screen.

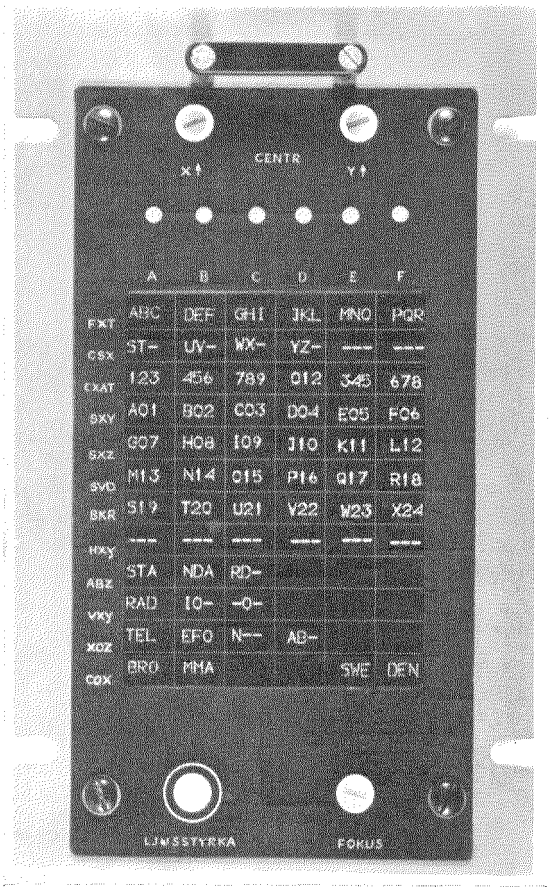


Figure 5—Characters on a tabular display.

4.2 Distribution to Synthetic Indicators

On a synthetic indicator, all presentation is in the form of symbols. These symbols are divided into categories in the Digitrac system (our own fighters, enemy aircraft, et cetera). Targets are placed in categories by adding marks to the symbol information in the administration computer.

If an operator wishes to see a certain category of targets, he presses a corresponding button on his desk. This supplies symbol information with the wanted mark from the memory to a coincidence circuit, which transmits a pulse to open the brightening-pulse gate system. All symbols in the wanted category are then presented one by one on his plan-position indicator.

4.3 Distribution to Tabular Displays

A register in the character generator contains one flip-flop for each tabular display. This register controls the gate system that distributes brightening pulses to the selected displays. When the character generator receives target information for a specific display from

the computer memory, the corresponding flip-flop in the register is set with a binary *1* that opens the brightening-pulse gate.

Target information is displayed in a column with a number of 3-character rows. The screen of the tabular display can accommodate 6 such columns (see Figure 5). The stepping in rows is performed by a tabular program unit, which generates a 4-bit code for each step. A 3-bit code for the stepping in columns is received directly from the computer memory. The tabular program is controlled by the computer.

4.4 Presentation of Pointers

Each operator in a center has his own symbol for marking and reporting to other operators. The pointer data can therefore be given permanent locations in the memory. The brightening pulses for pointers are not distributed as for other symbols, but a relay system permits an operator to transfer these brightening pulses to other operators selected by him.

A special coincidence matrix with permanently connected pointer addresses gives coincidence pulses when their pointer data are read from the computer memory. These pulses are then gated in the relay system, allowing any operator's coincidence pulse to be sent only to those other operators selected to see the pointer. The footpoints for all pointers can thus be sent to all indicators in a center, and the presentation selection is achieved solely through the distribution of brightening pulses.

B. G. O. Svenson was born in Stockholm, Sweden, in 1923. In 1943 he graduated in electrical engineering from the Technical Institute of Hässleholm. In 1949 he received a diploma of radar engineering from Marconi College, Chelmsford, England.

From 1949 to 1953, Mr. Svenson worked for the Royal Swedish Air Force on radar development.

In 1953 he joined Standard Radio & Telefon, where he has been working on radar data presentation. He is now responsible for development of radar data displays at the electronic system division.

K. T. J. Hylén was born in Linderöd, Sweden, in 1932. In 1952 he graduated in radio engineering from the Technical Upper School of Örebro.

From 1952 to 1958 he was employed by Svenska Radio. In 1958 he joined Standard Radio & Telefon, where he is now engaged in development work on digital data presentation equipment.

Digitrac Display System for Air-Traffic Control Part 3—Indicators and Displays

BENGT SVENSON
Standard Radio & Telefon AB; Bromma, Sweden

1. Indicators

1.1 General

The digital plan-position indicator shown in Figure 1 is housed in an octagonal casing that mounts into the operator's desk in Figure 2. The power supply for the indicator occupies a box on the floor behind the desk.

The indicator is assembled from chassis and printed-card subunits. These all swing out on hinges and are therefore easily accessible. Each indicator can accommodate a 12- or 16-inch (305- or 406-millimeter) cathode-ray tube. The inputs to the indicator consist of radar range sweeps, symbols, and vector lines, all converted from digital to analog form, and video signals, supplied in standard analog form. It is also possible to display information received from several radar stations as an integrated picture.

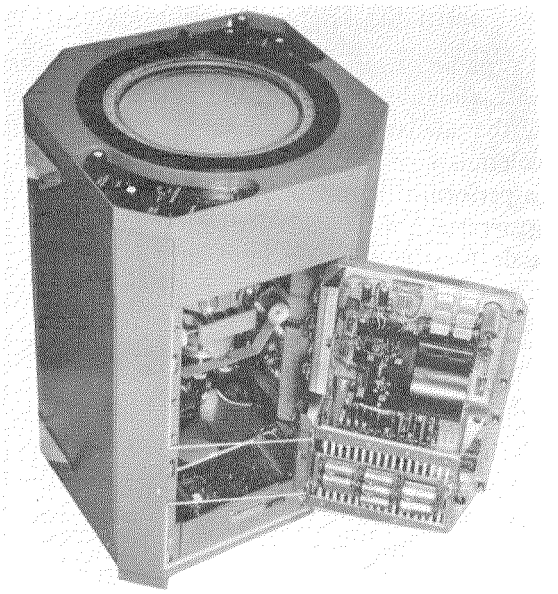
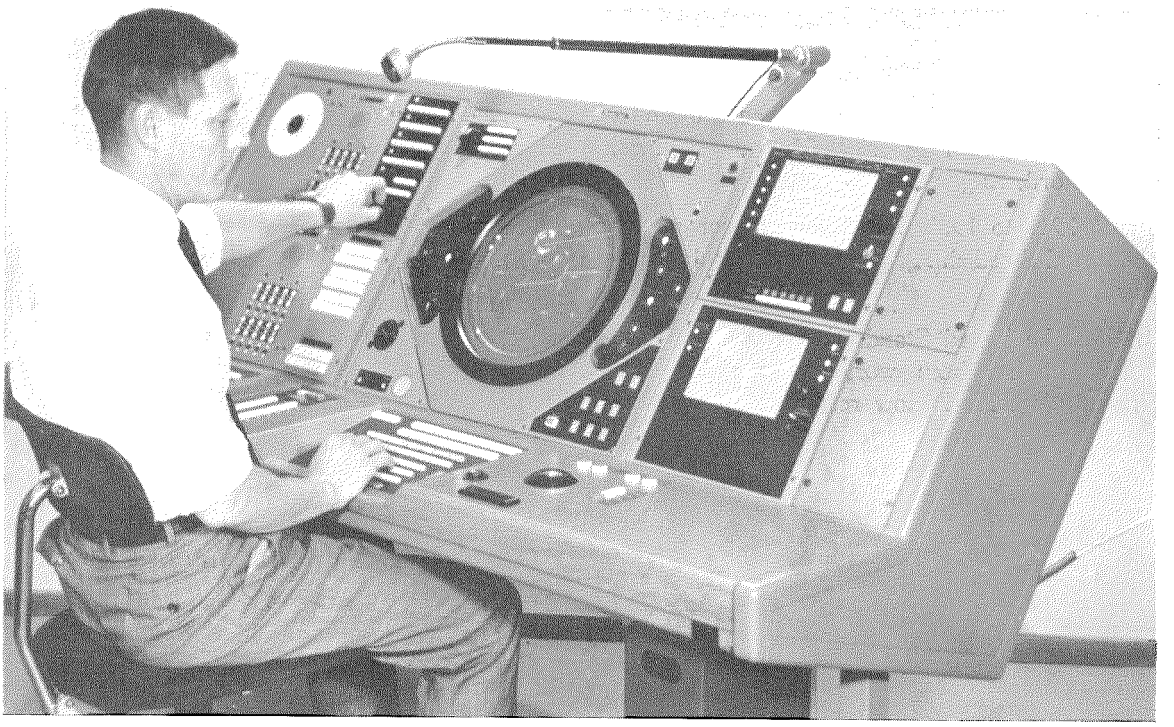


Figure 1—Digital indicator with subunit swung out.

Figure 2—Control desk with indicator in place.



All symbol marks and vector lines are displayed as interscans, and vector lines can be replaced, if desired, by curves from a separate function generator. Instead of presenting both radar sweeps and interscans, the indicator can also be used to display synthetic information.

1.2 Deflection System

Figure 3 is a block diagram of the indicator. Each of the X and Y deflection systems consists of a 12-bit sweep counter, register, digital-to-analog converter, and a sweep amplifier with deflection coil.

For coordinates of X and Y, a sweep generator gives bursts of pulses that vary in number from sweep to sweep. The pulses are counted for every sweep period, and the result is stored in the registers in binary form as 12 bits for each of the X and Y coordinates. The registers send the X and Y binary sweep values in parallel to the digital-to-analog converters. The analog sweep voltages thus formed go to the sweep amplifiers. A zero-setting pulse is delivered to the registers via their counters before the start of each sweep.

All symbols on the indicator are displayed as interscans. Each symbol footpoint is positioned by binary quantities of 12 bits for each X and Y, delivered by the Censor computer in a sequence determined by its programing. During the period between two radar sweeps, the footpoint values go in parallel through the counters to the registers, where they are stored during an interscan period. The symbol values are thus supplied asynchronously with the radar sweep repetition frequency. Display of a symbol can thus be interrupted and appear as only a fraction of a symbol during an interscan period.

The symbol footpoints delivered to the sweep amplifiers must be interrupted about 60 microseconds before the start of a radar sweep, to allow time for zero-setting the deflection system.

The sweep amplifiers that drive the deflection coils for X and Y are designed for direct-current operation. A voltage corresponding to the current through the deflection coils is used for feedback.

Each sweep amplifier has three inputs: the digital-to-analog converter input, the off-centering input, and a symbol character input from an analog symbol generator.

Off-centering can shift the picture in the X and Y directions up to a distance corresponding to four radii of the screen.

Range selection on the indicator is achieved by changing the amplification of the sweep amplifier through variation of the amount of feedback. The amplitude of the symbol character voltage is simultaneously altered to keep the size of the symbols unchanged at different range settings.

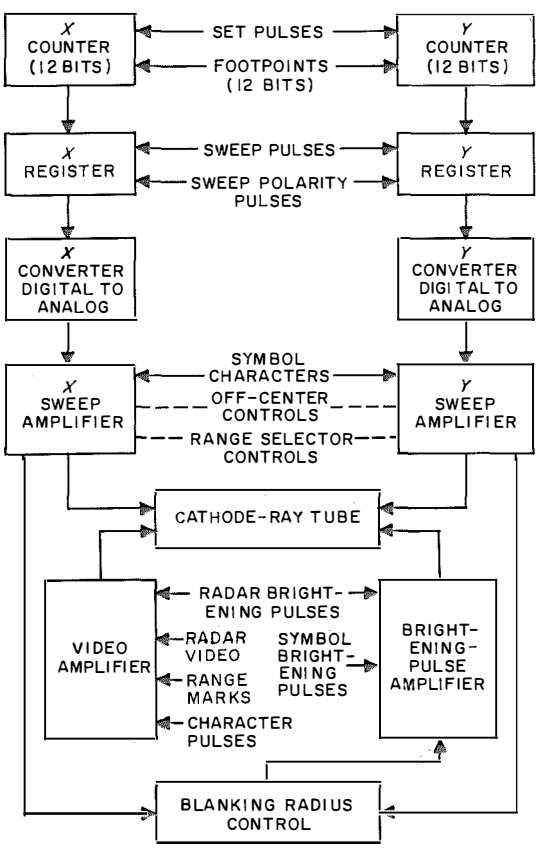


Figure 3—Functional diagram of digital plan-position indicator.

1.3 Video Amplifier

The video amplifier, designed for a bandwidth from 200 hertz to 1 megahertz at the 3-decibel-down points, has inputs for radar video, range marking pulses, and character video. The amplifier also has a gate circuit, controlled by the interscan pulses, which cuts off the radar video input when an interscan appears and thus prevents the symbols from becoming noisy. The character video signal is interrupted when a radar sweep is to be displayed. The range marking pulses are generated in the digital sweep generating system.

The brightening-pulse amplifier handles the brightness pulses of the radar sweep and the interscans. The brightness pulses of the interscans are timed to light the spot 60 microseconds after the footpoint value is set in the register. The timing is performed by the symbol generator.

When the sweep amplitudes reach the edge of the cathode-ray tube, the blanking radius control generates blanking signals from voltages proportional to the deflection currents in the X and Y deflection coils. These voltages are inverted and gated, to obtain a positive voltage proportional to the X and Y deflection, and then added in four resistor networks giving different sums. In a gate circuit the highest of these four sum voltages is selected to generate a blanking pulse at a certain amplitude. By choosing proper values in the resistor networks, it is possible to create a suitable blanking on the cathode-ray tube.

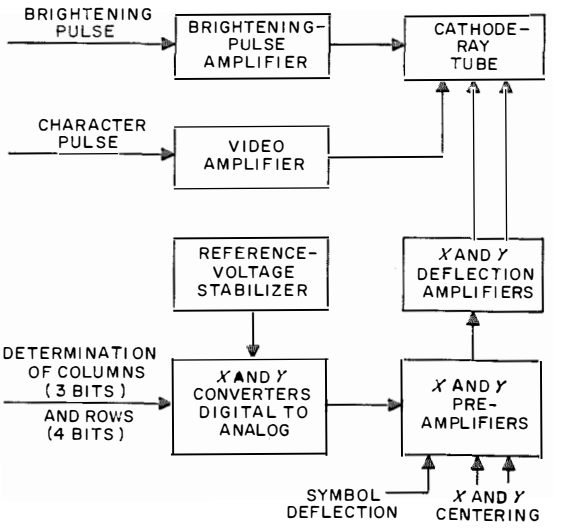


Figure 4—Functional diagram of tabular display.

1.4 Extra-High-Tension Generator and Power Supply

The extra-high-tension generator delivers 15 kilovolts to the first anode and 600 volts to the second anode of the cathode-ray tube, with a stability better than ± 100 volts at 15 kilovolts.

The extra-high-tension generator supplies a controlled-length square-wave pulse to the primary winding of a pulse transformer. The secondary winding provides an output of 7.5 kilovolts. The required 15 kilovolts is obtained by voltage-doubling in the rectifier circuit. Part of this final voltage is compared with the voltage from a zener diode. The difference is used to control the duration of the primary square wave, thus also controlling the 15-kilovolt output. All extra-high-tension parts are immersed in oil.

The stabilizing circuits of the power supply are mounted within the indicator, while the transformer and rectifiers are located in a box outside the indicator. The power supply delivers all power for the indicator and for the two adjacent tabular displays.

1.5 Cooling

The indicator, being fully enclosed, requires forced air cooling.. Connections are provided for an inlet and an outlet air duct. To improve the safety margin and ensure operation for at least half an hour after ventilation failure, the deflection amplifier is provided with a separate fan. A thermal switch is also provided that disconnects the power supply if the temperature exceeds the permissible limit.

2. Tabular Displays

Figure 4 is a block diagram of the tabular display, which presents additional information about targets appearing on the plan-position indicator or processed by the computer. The text information to be written is displayed in 6 columns, each containing 12 rows. The size of the characters is 3 by 4 millimeters (0.12 by 0.16 inch). Several tabular displays controlled from the symbol generator can be used together with a plan-position indicator. The tabular unit is designed to have a rectangular cathode-ray tube 90 by 120 millimeters (3.5 by 4.7 inches) in size. It also uses the same types of deflection amplifiers, deflection coils, and video amplifier as the indicator.

2.1 Writing on Tabular Displays

Usually blocks of three characters are written on each row. For every block of characters, the symbol generator supplies a 3-bit code, determining the column, and a 4-bit code, determining the row. These signals are supplied to the deflection system via digital-to-analog converters for X and Y. The binary codes are maintained as long as the block of characters must be written. The symbol character voltages go to the deflection amplifiers in analog form, as in the plan-position indicator.

The video amplifier receives the character video pulses, and the brightening-pulse amplifier receives a brightening pulse for every block of 3 characters.

B. G. O. Svenson. Biography appears on page 503 of this issue.

Digitrac Data Links

C. O. SVENSSON
C. J. VEDIN
Standard Radio & Telefon AB; Bromma, Sweden

In the control of air traffic, data transmission meets the frequently exacting requirements for speed and accuracy. There may be 50 or more data channels connected to one center for such purposes as narrow-band transmission of radar pictures, bearing information from direction finders, communication with adjacent centers, ground-to-air communication, et cetera.

Data transmission equipment must work compatibly with all kinds of data sources and data consumers. Thus flexibility and adaptability are essential requirements. For this reason it is better to consider the data link as a function rather than as a specific equipment.

1. Bit Synchronization

At the transmitter, a clock controls the transition times of the transmitted pulses. A clock at the receiver must be synchronized with the transmitter clock. Crystal oscillators are used to provide the necessary frequency stability. The crystal frequency is divided by a binary counter to a clock frequency corresponding to the baud figure of the transmission speed. Figure 1 shows the clock pulse generator in a data receiver.

Another requirement is that the clock pulses of the receiver be in phase with the pulses in the data message being received. The binary counter also performs this phasing. The transition times of the data message pulses are compared in the comparator with those of the clock pulses. If the clock pulse is too early, the time difference t will be greater than 0 and the comparator will inhibit 1 pulse to the binary counter. Correspondingly 1 extra pulse will be added if t is less than 0.

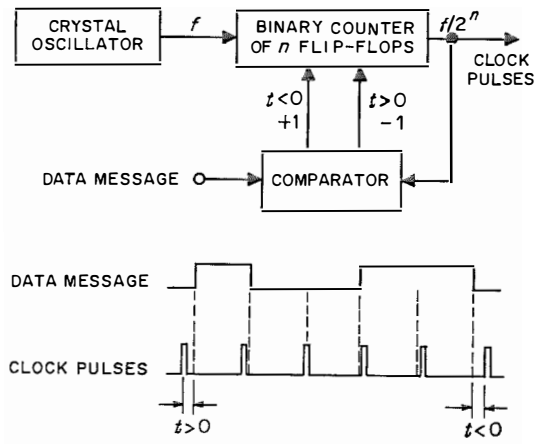


Figure 1—Clock pulse generator in a data receiver.

Thus the phase of the clock pulses is corrected at every transition of the data message. The amount of each correction is $1/f$ seconds, where f is the frequency of the crystal oscillator. The maximum number of such corrections necessary in the most-divergent case is clearly 2^{n-1}, where n is the number of stages in the binary counter. This also determines f as $f = 2^n \cdot B$, where B is the baud figure. It is usually specified that synchronism be achieved within 32 transitions in the data message, which makes the number of stages in the binary counter equal to 6.

The phase control may be regarded as a combined phase and frequency control.

This method of employing an oscillator at a relatively high frequency and a frequency-dividing binary counter also makes it easy to change the baud figure by changing the binary-counter module.

2. Message Synchronization

The proper processing of a message in the receiver also depends on synchronism between the transmitter and receiver on a character-by-character basis. To obtain this synchronism, a start-stop character cycle is normally used, the program unit in the receiver being started by the message itself. For this reason each message is preceded by a start code consisting of a row of binary *1*'s, say n. To prevent the same number of *1*'s from occurring in the message itself, the message is divided into submessages, each of length n-1 bits, separated by binary *0*'s. The length of the start code is optimized rela-

tive to the message length. If the message length is N, then the length of the submessage must be $N^{1/2}$ and the start code $n = N^{1/2} + 1$.

The receiver contains a start-code detector (often combined with the input shift register) that gives a pulse only when the input shift register contains a row of n *1*'s.

3. Addressing

Information from a center is often transmitted to several receivers over the same channel. This is particularly the case in the ground-to-air link, in which several planes receive information from one air traffic center over one common radio channel.

For a receiver to be able to select the information intended for it, an address must be added to each message. For practical reasons the address is given in binary form and is transmitted immediately after the start code.

The terminal equipment then examines all messages in an address detector, which usually consists of a diode gate. The address detector can easily be reprogramed to accommodate address changes.

4. Check Methods

Data transmissions in air-traffic control are made in real time. There is continual repetition of the same or almost the same messages. Occasional errors therefore disturb the transmission only for a short time and there is no need for error correction, only for error detection.

One method often used for error detection is to add one extra bit per submessage, or perhaps per character if this is more practicable, so that the total number of bits is always even or always odd. This is called the parity-check method.

Critical characters in the message, such as command signals, can of course be checked more carefully. One method is to require that two consecutive characters be equal before acceptance. More-sophisticated methods are based, for example, on cyclic codes with greater or lesser redundancy. They may be necessary for one-shot transmissions if errors must be corrected.

It is also desirable to check the operation of the terminal equipment. One successful method used in the Digitrac system is to transmit standard test messages alternately with regular messages and in the same way. Depending on the makeup of the complete system, the test messages can be generated by or stored in the data sources and be evaluated by the data consumers, or the evaluation can be made by the terminal equipment itself.

5. Message Format

It is perhaps possible to standardize message formats, which would permit greater standardization of the terminal units. However, as mentioned earlier, the requirements for number of bits and characters, check methods, and so on, vary from case to case. Therefore the message formats also vary. A typical message format can be seen in Figure 2.

6. Program

The function of a transmitter or receiver in a data link is controlled by an internal program.

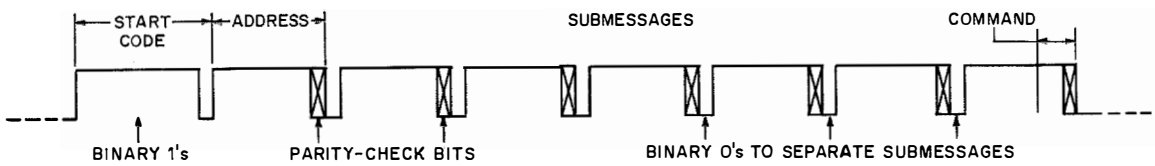


Figure 2—Typical message format.

The program circuits produce the pulses necessary to process the data message.

The two main blocks in these circuits as shown in Figure 3 are a binary program counter and a diode decoding matrix. The program counter is synchronized with the data message and decoded by the matrix. The pulses from the matrix, thus related to the pattern of the data message, control the processing circuits in the equipment.

Depending on the applications of data links, the programs are made fixed or flexible. For special data links of mobile or airborne type, it is economical to use fixed programs. But when several data links and many data messages are involved in a radar data system, it is economical to use flexible programs.

In the Digitrac system each data message is handled exclusively by a transmitting and receiving unit. All transmitting and receiving units are identical in mechanics and cabling. To construct a program required by a data message, the terminal unit (transmitter or receiver) is equipped with printed boards and prewired program contacts according to a special list. By means of the list and a selection of boards and prewired contacts, any data terminal in the system can be programed easily. This arrangement has great advantages, not only for manufacture but also for service and maintenance.

7. Interfaces

There are 4 interfaces in a complete data link: two between the data terminal and the modem equipments, one between the data source and the data terminal transmitter, and the fourth between the data terminal receiver and the data consumer (Figure 3).

7.1 Modem Interfaces

The type and form of the binary signals to be exchanged at the modem interfaces agree with

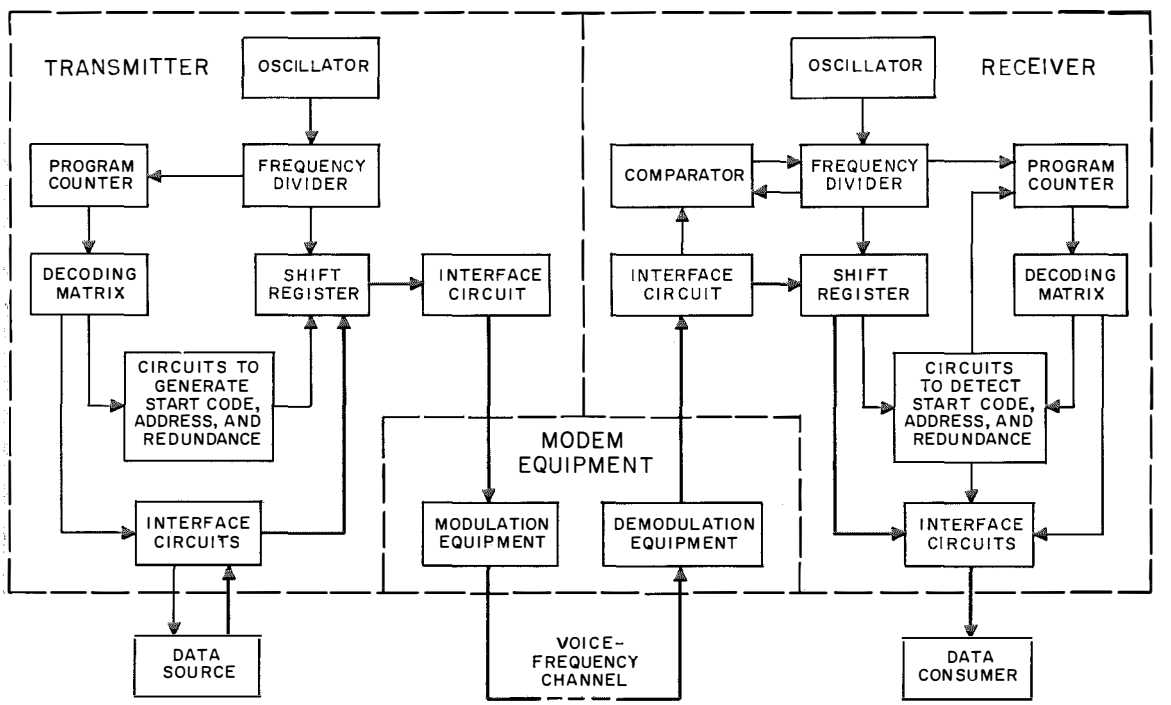


Figure 3—Block diagram of a complete data link. The thick lines show the information path.

the recommendations of the Comité Consultatif International Télégraphique et Téléphonique.

The binary signals are double-current square waves. The interface voltage is nominally 6 volts with respect to signal ground. The signal is considered to be in the mark condition when the voltage is more negative than −3 volts, and in the space condition when the voltage is more positive than +3 volts. The mark and space conditions denote the binary states *1* and *0*, respectively.

7.2 Source and Consumer Interfaces

The interaction between a data transmitter and sources such as switches, encoders, rolling-ball registers, and analog-to-digital registers is simple. They can all be transformed to the same equivalent circuit consisting of a number of 2-position switches that connect each source as required to a common wire. This wire is used by the program circuits to write the binary information into the shift register for serial transmission.

The transfer of signals from a data receiver to consumers such as relays, readouts, digital-to-analog converters, and analog or digital servo equipments is equivalent to writing the information from the shift register into buffer registers. This presents no problems.

The interaction between the data terminal units and the computer memory in a Digitrac system is more complicated. All terminal units involved

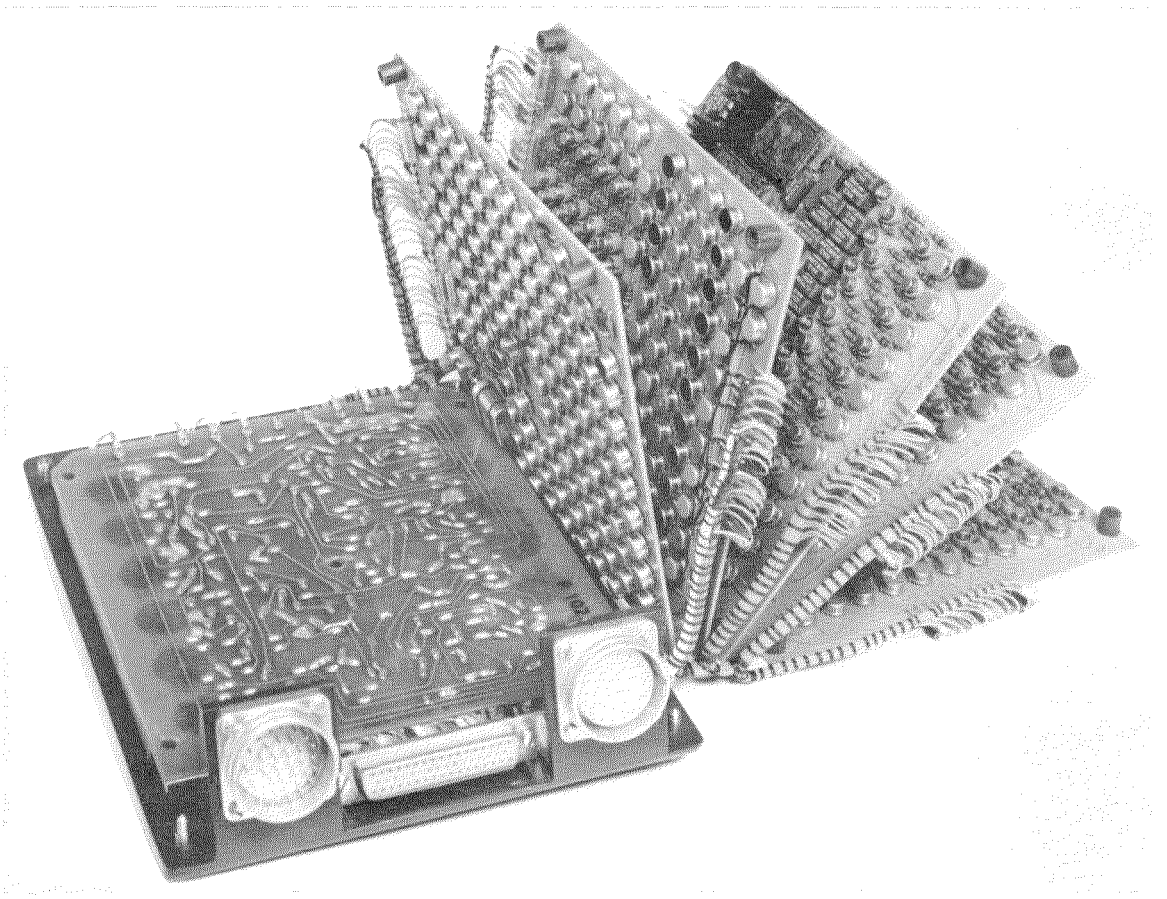


Figure 4---Airborne data receiver including modem equipment and digital-to-analog converters.

work independently. The data information from the central computer memory to the data terminal units is transferred via the common bus line connecting all parts of the system. The time in a transmitter or a receiver program during which the transfer of data must occur is equal to the bit time (the inverted value of the baud figure of the link). The time for the central unit to connect external equipment to the memory is as short as 10 microseconds. Thus there is no risk of losing data, even if all units want data transferred at the same time. All units are given different priorities. A terminal unit with high data speed is given a higher priority than a unit with low data speed.

The transfer of data to a terminal unit is carried out as follows. The data unit calls for transfer; the central unit delays its reply depending on the priority of the calling data unit; the data unit gives the address in the memory in which to read or write the data; the transfer occurs. To meet the requirements, the interface circuits of both units are equipped with very-high-speed amplifiers.

8. Airborne Data Receiver

Figure 4 shows a typical airborne data receiver. It contains 260 semiconductor integrated circuits and 9 thin-film hybrid circuits, in addition to hundreds of conventional parts.

The input signal is frequency shifted, then demodulated, and the message is placed in an input shift register. When a start code is detected, the remainder of the receiving program cycle is started. The address of the message is examined; if the destination is correct, an address flip-flop is set and the rest of the message is accepted. The different characters are processed in turn, error detection is performed, and finally the characters are stored in flip-flop registers. In this particular case some of the values required by the data consumer are 400-hertz signals. Therefore digital-to-analog converters are incorporated in the data receiver and connected to the flip-flop registers. The converters consist of summing resistor networks and amplifiers. The total accuracy of this conversion is of the order of 99.9 percent.

C. O. Svensson was born in Halmstad, Sweden, in 1930. He graduated from the Royal Institute of Technology in physics in 1954.

From 1952 to 1954 he was engaged in research on aerodynamics and thermodynamics for the Swedish Air Force Board.

He joined Standard Radio & Telefon in 1954 and started work on digital techniques the next year. In 1957 he became head of a new section for development, design, and evaluation of data systems.

Mr. Svensson is a member of the Swedish Association of Engineers and Architects.

C. J. Vedin was born in Bromma, Sweden, in 1931. He received a degree in electrical engineering at the Royal Institute of Technology in Stockholm in 1958.

He joined Standard Radio & Telefon in 1957, working as a development engineer. Mr. Vedin is now responsible for the development of data links.

Congestion in a Loss System when Some Calls Want Several Devices Simultaneously*

R. FORTET
CH. GRANDJEAN
Laboratoire Central de Télécommunications; Paris, France

1. Introduction

The central memory in the semielectronic *10 AX* system is divided into "sentences," which give for each call the numbers of the calling and called subscribers, the number of the junctor, the code of the path used in the speech network, et cetera. A particular call requires either one or two sentences depending on whether it is local, outgoing, or incoming. Sentence here means a group of words, each composed of a set of binary digits of varying length.

The required size of such a memory must then be calculated for a given value of the loss probability. In other words, the mathematical problem is to find the loss probability of a group of devices to which are offered several categories of traffic, the calls of a particular category occupying the same number of devices. Each call of the first traffic category occupies one device (for instance, one sentence), each call of the second traffic category occupies two devices simultaneously, and so forth.

The theory developed by Cohen [1] makes it possible to solve the problem that arises when a group of devices is occupied by a number (finite or not) of sources, whatever the distributions of the holding times and of the intervals between calls may be, but under the assumption that one call occupies one device. Incidentally, some basic points in Cohen's theory are left without absolute proof (for example, uniqueness of the solution, and existence and stability of the stationary process).

The theory, elaborated on in this article, is applicable to more-general cases, in which the only assumption is that the rule of seizing the devices takes into account the total number of devices of the group and only the present number of active sources.

Moreover, it is shown that, by reference to Feller's treatment, it is possible to prove absolutely the existence, stability, and uniqueness of the stationary process, at least if the distribution functions possess densities (a restrictive condition that may, perhaps, be superfluous).

This theory is thus particularly relevant to the above-mentioned problem of calculating the size of a central memory to which several categories of traffic are offered. Moreover, each such category may be the sum of several subcategories, the calls of which have the same occupation rule but different average holding times.

The case of two categories of traffic A and B occupying, respectively, one and two devices per call, is thoroughly treated as an example. Recurrence formulas are established making it possible to compute loss probabilities for traffic A and traffic B in a particularly simple way, and the numerical results are given in some curves. The main elements of the mathematical part of the study are in the Appendix, Section 7†.

2. Problem

A number N, finite or not, of sources of calls $S_1, S_2, \cdots, S_j, \cdots$, and a group $\mathcal{G}$ of x telephone devices used for handling the calls originated by this system of sources are considered. It is assumed that the sources behave independently in the following way. The duration T of a conversation of S_j is a random variable with the distribution function $F_j(t) = P_r(T < t)$,

* Presented at International Teletraffic Congress in London, England, 15–21 July 1964.

† Since this paper was written we have become aware of a paper dealing with a similar problem, "Analysis of Mixtures of Wide- and Narrow-Band Traffic," presented by L. A. Gimpelson at the Globecom VI Symposium in Philadelphia, Pennsylvania; June 1964.

which is independent of the history of S_j—before the beginning of this conversation—as well as of the past and present state of $\mathcal{G}$; $F_j(t)$ can depend on j.

Conditionally, if at the moment t, S_j has been engaged in conversation for a duration τ, the probability that this conversation will cease within the time interval t, $t + \mathrm{d}t$ is thus

$$\frac{F_j(\tau + \mathrm{d}t) - F_j(\tau)}{1 - F_j(\tau)} . \qquad (1)$$

If $F_j(t)$ assumes a density $f_j(t)$, the principal part of (1) for $\mathrm{d}t \to 0$ is

$$\frac{f_j(\tau)}{1 - F_j(\tau)} \mathrm{d}t. \qquad (2)$$

It will be assumed that $F_j(+0) = 0$ and that

$$\mu_j = \int_0^{+\infty} t \mathrm{d}F_j(t) = \int_0^{+\infty} [1 - F_j(t)]\mathrm{d}t < +\infty . \qquad (3)$$

In other words, the durations of conversation of S_j have a finite mathematical expectation.

Let it be said that S_j is active if it is engaged in conversation and inactive in the opposite case. Take a moment, for instance the moment 0, when S_j becomes inactive (end of a conversation of S_j); the time T_1 when S_j will make its next call is random and its distribution function is

$$G_j(t) = P_r(T_1 < t).$$

T_1 is independent of the history of S (before the moment 0), and of the past and present state of $\mathcal{G}$. If $\mathcal{G}$ is not in a blocking state at the moment T_1, the call will be accepted and S_j will then become active. Otherwise the call will be refused; then S_j will originate its next call at time $T_1 + T_2$, in which T_2 is random, having a distribution function $G_j(t)$ and being independent of the history of S_j (before time T_1) and of the past and present state of $\mathcal{G}$; acceptance of this call depends on whether or not $\mathcal{G}$ is in a blocking state at the time $T_1 + T_2$. In the former case, S_j will originate its next call at time $T_1 + T_2 + T_3$, in which T_3 is a random variable, independent of the past of S_j (before $T_1 + T_2$) and of the past and present states of $\mathcal{G}$ and its distribution function $G_j(t)$; and so on. At the moment t, when S_j is inactive, it is said that S_j has been inactive for duration τ if $t - \tau$ is the time of the end of the last call originated by S_j (if this call was accepted) or the time of the last call originated by S_j (if this call was refused). Conditionally, if at the moment t, S_j has been inactive for a duration τ, the probability that S_j will originate a call in the time interval t, $t + \mathrm{d}t$ is then

$$\frac{G_j(\tau + \mathrm{d}t) - G_j(\tau)}{1 - G_j(\tau)} . \qquad (4)$$

If $G_j(t)$ has a density $g_j(t)$, the principal part of (4) for $\mathrm{d}t \to +0$ is

$$\frac{g_j(t)}{1 - G_j(\tau)} \mathrm{d}t. \qquad (5)$$

It is assumed that $G_j(+0) = 0$ and that

$$\nu_j = \int_0^{+\infty} t \mathrm{d}G_j(t) = \int_0^{+\infty} [1 - G_j(t)]\mathrm{d}t < +\infty . \qquad (6)$$

Take any moment t; let $j_1, j_2, \cdots, j_k$ be the active sources at that moment; suppose that at that moment t, the source S_j ($j \neq j_\alpha$; $\alpha = 1, 2, \cdots, k$) originates a call; we assume that the rules $\mathcal{R}$ of occupation of the devices of $\mathcal{G}$ by this call bear only on $j_1, j_2, \cdots, j_k$ and j; the particular fact that the call will be accepted or refused, in other words that $\mathcal{G}$ is or is not in a blocking state for this call, depends only on j and $j_1, \cdots, j_k$.

3. Basic Results

With these hypotheses, the evolution of the set ($\Sigma + \mathcal{G}$) of the system Σ of the sources and of $\mathcal{G}$ is a Markov process $\mathfrak{M}$, studied in the Appendix. More precisely, this study is made under the following additional hypothesis H.

Hypothesis H: The F_j and the G_h have densities f_j and g_h, and the ratios

$$\frac{f_j(u)}{1 - F_j(u)}, \quad \frac{g_h(v)}{1 - G_h(v)}, \quad (j, h = 1, 2, 3, \cdots)$$

are bounded.

Then $\mathfrak{M}$ has a unique stationary state, the only state to be studied here; consider the probability, in this state, that at any instant the k $j_1, j_2, \cdots, j_k$ have been active for periods of time comprised, respectively, between u_1 and $u_1 + \mathrm{d}u_1$, u_2 and $u_2 + \mathrm{d}u_2, \cdots, u_k$ and $u_k + \mathrm{d}u_k (u_\alpha > 0, \mathrm{d}u_\alpha > 0; \alpha = 1, 2, \cdots, k)$; while the other sources $h_1, h_2, \cdots$ have been inactive for periods of time comprised, respectively, between v_1 and $v_1 + \mathrm{d}v_1$, v_2 and $v_2 + \mathrm{d}v_2, \cdots (v_\beta > 0, \mathrm{d}v_\beta > 0; \beta = 1, 2, \cdots, N - k)$.

When the $\mathrm{d}u_\alpha$ and $\mathrm{d}v_\beta$ are small, this probability has a principal part that will be represented by

$$\prod_k(j, u; h, v)\mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v_1 \cdots \mathrm{d}v_{N-k}$$

in which j, u, h, and v designate, respectively, the set of the j_α, the u_α, the h_β, and the v_β.

As seen in the appendix, the values of the $\prod_k(j, u; h, v)$ are obtained in the following way. Let $E_k(j_1, \cdots, j_k)$ denote the event that the k sources $S_{j1}, \cdots, S_{jk}$ are active (whatever the present durations of these activities may be), while the other sources are inactive (whatever the durations of such inactivity may be); by virtue of rule $\mathfrak{R}$, certain of the $E_k(j_1, \cdots, j_k)$ are possible and others impossible; let $\sum^*$ designate a summation extended exclusively to the systems $k; j_1, \cdots, j_k$ for which $E_k(j_1, \cdots, j_k)$ is possible; let

$$\omega_k(j_1, \cdots, j_k) = \prod_{r=1}^{k} \left(\frac{\mu_{jr}}{\nu_{jr}} \right) \tag{7}$$

$$\Omega = \sum^* \omega_k(j_1, \cdots, j_k). \tag{8}$$

Then the $\prod_k(j, u; h, v)$ are given by (43), where λ is an absolute constant. For (43) to constitute a law of probability, it is necessary and sufficient that

$$\lambda = \frac{1}{(\prod_{\alpha=1}^{N} \nu_\alpha)\Omega} \tag{9}$$

and therefore that

$$\prod_k(j, u; h, v) = \frac{1}{(\prod_{\alpha=1}^{N} \nu_\alpha)\Omega} \times \prod_{r=1}^{k} [1 - F_{jr}(u_r)] \prod_{s=1}^{N-k} [1 - G_{hs}(v_s)]. \tag{10}$$

In the form (10), rule $\mathfrak{R}$ affects only the value of Ω. Let $\bar{\omega}_k$ be the probability that k particular sources $j_1, j_2, \cdots, j_r, \cdots, j_k$ are active (and have been so for any length of time), while the $N - k$ other sources have been inactive (for any length of time); (10) gives

$$\bar{\omega}_k = \int \Pi_k(u, v)\mathrm{d}u\mathrm{d}v = \frac{1}{(\prod_{\alpha=1}^{N} \nu_\alpha)\Omega} \prod_{r=1}^{k} \mu_{jr} \prod_{s=1}^{N-k} \nu_{hs}$$

that is

$$\bar{\omega}_k = \frac{1}{\Omega} \prod_{r=1}^{k} \frac{\mu_{jr}}{\nu_{jr}}. \tag{11}$$

3.1 Remarks

As given in the Appendix, it is certainly possible to replace hypothesis H with a broader hypothesis; it is probably even possible to eliminate hypothesis H purely and simply, on condition, of course, that hypotheses (3) and (6) are kept.

4. Applications

The preceding scheme contains, as a particular case, most of the 1-stage loss systems without waiting process that have been contemplated. It contains, in particular, the system studied by Cohen [1].

An example of a newer type is the following. The computation of the capacities of the

memories of a telephone exchange leads to the computation of the loss of a group of devices to which a compound traffic is offered; the characteristic property of the different parts of this traffic is to occupy either one or more than one device per conversation. First we shall study the case of a 2-part traffic offered to one group of devices—generalization being immediate—each conversation of the first part of the traffic occupying one device, each conversation of the second part of the traffic occupying two devices. This is clearly a case of Rule $\mathfrak{R}$, set forth in Section 2, since the refusal of a call depends only on the number of active sources.

Let

X = the number of devices of the group considered

$S_i (i = 1, 2, \cdots, M)$ = the set of sources forming the first part A of the traffic occupying one device per conversation

$S'_j (j = 1, 2, \cdots, N)$ = the set of sources forming the second part B of the traffic occupying two devices per conversation

ν_i, ν'_j = the average duration of the intervals between successive calls of sources S_i and S'_j, respectively

μ_i, μ'_j = the average duration of conversations of sources S_i and S'_j, respectively.

4.1 Case 1

According to (11) the probability $\bar{\omega}_{m,n}$ that m particular sources $S_{i1}, \cdots, S_{ir}, \cdots, S_{im}$ of the traffic A, and n particular sources $S'_{j1}, \cdots, S'_{js}, \cdots, S'_{jn}$ of the traffic B, are active, the others being inactive, is given by

$$\bar{\omega}_{m,n} = \frac{1}{\Omega} \prod_{r=1}^{m} \frac{\mu_{ir}}{\nu_{ir}} \prod_{s=1}^{n} \frac{\mu'_{js}}{\nu'_{js}} \tag{12}$$

if this state is a possible state of the system.

Ω is determined by

$$\Sigma \bar{\omega}_{m,n} = 1 \tag{13}$$

in which the summation is extended to all the possible states.

The loss probability $\mathfrak{B}_X(A)$ for the sources of traffic A and the loss probability $\mathfrak{B}_X(B)$ for the sources of traffic B, are given by

$$\begin{aligned} \mathfrak{B}_X(A) &= \Sigma \bar{\omega}_{X-2n,n} \\ \mathfrak{B}_X(B) &= \Sigma (\bar{\omega}_{X-2n,n} + \bar{\omega}_{X-2n-1,n}) \end{aligned} \tag{14}$$

in which the summations are extended to all the values of n corresponding to possible states

$$n = 0, 1, \cdots, \frac{X - X(\text{mod } 2)}{2} .$$

4.2 Case 2

Let us suppose that all the sources S_i are identical.

$$\begin{aligned} \mu_i &= \mu \\ \nu_i &= \nu, \text{ for all } i \end{aligned} \tag{15}$$

and that all the sources S'_j are identical.

$$\begin{aligned} \mu'_j &= \mu' \\ \nu'_j &= \nu', \text{ for all } j. \end{aligned} \tag{16}$$

The probability $P_{m,n}$ that any m sources of traffic A and any n sources of traffic B are active simultaneously, the other sources being inactive (in all $m + 2n$ devices occupied) is given by

$$P_{m,n} = \frac{1}{\Omega} C_M^m \left(\frac{\mu}{\nu}\right)^m C_N^n \left(\frac{\mu'}{\nu'}\right)^n . \tag{17}$$

If the corresponding state is a possible state ($m + 2n \leqslant X$), Ω is determined by the condi-

tion of normalization

$$\sum P_{m,n} = 1 \tag{18}$$

in which the summation is extended to all the possible values of the couple m, n such that

$$m + 2n \leqslant X. \tag{19}$$

The blocking probabilities are given by equations analogous to (14).

4.3 Case 3

Finally, let us consider the case in which the number of sources of each traffic is infinite. It is known that, under fairly widespread conditions, the traffic A and the traffic B tend toward Poissonian traffic.

Calling $P_{m,n}$ the probability that m devices of the group are occupied by traffic A and that $2n$ devices are occupied by traffic B, $X - m - 2n$ devices remaining unoccupied, we obtain (provided that $X - m - 2n \geqslant 0$)

$$P_{m,n} = \frac{1}{\Omega} \frac{A^m}{m!} \frac{B^n}{n!} \tag{20}$$

in which Ω is determined by (18).

Expression (20) can be obtained directly by solving the equation of state, which, assuming an exponential distribution of the holding time, is written

$$\left(\frac{A+m}{h} + \frac{B+n}{h'}\right) P_{m,n} = \frac{A}{h} P_{m-1,n} + \frac{B}{h'} P_{m,n-1} + \frac{m+1}{h} P_{m+1,n} + \frac{n+1}{h'} P_{m,n+1} \tag{21}$$

in which h and h' are, respectively, the average holding times of traffic A and traffic B*.

It is seen that this equation is separable with respect to the indexes m and n, and that its solution is the same as that of the system

$$\begin{aligned}(A+m)P_{m,n} &= AP_{m-1,n} + (m+1)P_{m+1,n} \\ (B+n)P_{m,n} &= BP_{m,n-1} + (n+1)P_{m,n+1}\end{aligned} \tag{22}$$

the solution of which is clearly given by (20). But the result obtained in Section 3 is much more general, since the validity of this solution for any probability density of the holding times, the existence of a stationary regime, and the uniqueness of the solution have been proved.

The complete expression of the coefficient Ω of (20) is given by

$$\Omega = \sum_{j=0}^{r} \sum_{i=0}^{X-2j} \frac{A^i}{i!} \frac{B^j}{j!} \tag{23}$$

where r is the greatest integer $\leqslant X/2$, that is

$$r = \frac{X - X(\mathrm{mod}\ 2)}{2}. \tag{24}$$

A call of traffic A will be lost if all the devices are occupied and the loss probability for traffic A is thus

$$\mathfrak{B}_X(A) = \sum_{m+2n=X} P_{m,n} = \frac{\sum_{j=0}^{r} \frac{A^{X-2j}}{(X-2j)!} \frac{B^j}{j!}}{\sum_{j=0}^{r} \sum_{i=0}^{X-2j} \frac{A^i}{i!} \frac{B^j}{j!}}. \tag{25}$$

A call of traffic B will be lost if all the devices or all but one of the devices are occupied; the loss for traffic B is thus

$$\mathfrak{B}_X(B) = \sum_{\substack{m+2n=X \\ m+2n=X-1}} P_{m,n} = \mathfrak{B}_X(A) + \frac{\sum_{j=0}^{q} \frac{A^{X-1-2j}}{(X-1-2j)!} \frac{B^j}{j!}}{\sum_{j=0}^{r} \sum_{i=0}^{X-2j} \frac{A^i}{i!} \frac{B^j}{j!}}. \tag{26}$$

The index q, at the upper limit of the summation in the numerator, is equal to r if X is an odd number and to $r - 1$ if X is an even number.

* This problem was solved in the case where $h = h'$ by N. Rönnblom (*Tele*, English Edition, number 2, 1959), who obtained the correct solution in a different way with a view to a different application.

We are now going to establish relations between $\mathcal{B}_X(A)$ and $\mathcal{B}_X(B)$ as well as recurrence relations on X, and it will be seen that the computation of (25) and (26) thus becomes very easy.

4.4 Practical Computation of Blockings

Calling ν_X the numerator and δ_X the denominator of the last member of (25) and writing the values of $\mathcal{B}_X(A)$ and $\mathcal{B}_X(B)$, the following relations are found.

(**A**) the numerator of (26) is none other than the numerator of (25) in which X has been replaced by $X - 1$, which means that whether X is odd or even, we may write

$$\mathcal{B}_X(A) = \frac{\nu_X}{\delta_X} \tag{27}$$

$$\mathcal{B}_X(B) = \mathcal{B}_X(A) + \frac{\nu_{X-1}}{\delta_X}. \tag{28}$$

(**B**) Between ν_X and δ_X we obtain the recurrence relation

$$\delta_X = \delta_{X-1} + \nu_X \tag{29}$$

from which it is immediately drawn that

$$\frac{1}{\mathcal{B}_X(A)} = 1 + \frac{\nu_{X-1}}{\nu_X}\frac{1}{\mathcal{B}_{X-1}(A)} \tag{30}$$

$$\mathcal{B}_X(B) = \left(1 + \frac{\nu_{X-1}}{\nu_X}\right)\mathcal{B}_X(A). \tag{31}$$

Computation of the losses by recurrence leads us to the computation of the series ν_X, or more exactly to the computation of the series of ratios ν_{X-1}/ν_X, for which we are also going to establish a recurrence relation.

For this, three consecutive values of X will be considered (see (25) and (27)).

4.4.1 X *Even* $= 2r$

$$\nu_X = \sum_{j=0}^{r} \frac{B^j}{j!}\frac{A^{X-2j}}{(X-2j)!}$$

$$\nu_{X-1} = \sum_{j=0}^{r-1} \frac{B^j}{j!}\frac{A^{X-1-2j}}{(X-1-2j)!}$$

$$\nu_{X-2} = \sum_{j=0}^{r-1} \frac{B^j}{j!}\frac{A^{X-2-2j}}{(X-2-2j)!}$$

$$= \sum_{j=1}^{r} \frac{B^{j-1}}{(j-1)!}\frac{A^{X-2j}}{(X-2j)!}. \tag{32}$$

Multiply ν_{X-1} by A/X and ν_{X-2} by $2B/X$ and add

$$\begin{aligned}\frac{A}{X}\nu_{X-1} + \frac{2B}{X}\nu_{X-2} &= \sum_{j=0}^{r-1}\frac{B_j}{j!}\frac{A^{X-2j}}{X(X-1-2j)!} + \sum_{j=1}^{r}\frac{2}{X}\frac{B^j}{(j-1)!}\frac{A^{X-2j}}{(X-2j)!} \\ &= \frac{A^X}{X!} + \sum_{j=1}^{r-1}\frac{B^j}{j!}\frac{A^{X-2j}}{(X-2j)!}\left(\frac{X-2j}{X} + \frac{2j}{X}\right) + \frac{2}{X}\frac{B^r}{(r-1)!} \\ &= \frac{A^X}{X!} + \sum_{j=1}^{r-1}\frac{B^j}{j!}\frac{A^{X-2j}}{(X-2j)!} + \frac{B^r}{r!}. \end{aligned} \tag{33}$$

It is seen that this sum is identical to ν_X.

4.4.2 X *Odd* $= 2r + 1$

The relations are the same as in (32), with this exception: In the expression ν_{X-1}, the summation is extended from 0 to r (and not from 0 to $r - 1$). That $(A/X)\nu_{X-1} + (2B/X)\nu_{X-2}$ is reduced to ν_X can be verified in the same way as in (33).

Thus, for any X, the following relation can be written

$$\nu_X = \frac{A}{X}\nu_{X-1} + \frac{2B}{X}\nu_{X-2},$$

$$\text{with } \nu_0 = 1,\ \nu_1 = A. \tag{34}$$

Actually only the ratio ν_{X-1}/ν_X is necessary in

the computation of blockings. Taking

$$\frac{1}{\lambda_X} = \frac{\nu_{X-1}}{\nu_X} \tag{35}$$

the relation (34) is written

$$\lambda_X = \frac{A}{X} + \frac{2B}{X \cdot \lambda_{X-1}}, \quad \lambda_1 = A. \tag{36}$$

In sum, if the blockings $\mathfrak{B}_X(A)$ and $\mathfrak{B}_X(B)$ are to be computed, the $\lambda_i (i = 1, 2, \cdots, X)$ will be computed first by means of (36) and the blockings will be given by

$$\frac{1}{\mathfrak{B}_i(A)} = 1 + \frac{1}{\lambda_i \mathfrak{B}_{i-1}(A)}, \qquad i = 1, 2, \cdots, X \tag{37}$$

$$\mathfrak{B}_i(B) = \left(1 + \frac{1}{\lambda_i}\right) \mathfrak{B}_i(A), \qquad i = 1, 2, \cdots, X \tag{38}$$

with

$$\mathfrak{B}_1(A) = \frac{A}{1 + A}.$$

We could just as well take as departure value for the recurrences

$$\lambda_0 = \infty$$
$$\mathfrak{B}_0(A) = \mathfrak{B}_0(B) = 1. \tag{39}$$

Finally, it can be pointed out that (36) and (37) are reduced to the well-known recurrence relations of the Erlang function for $B = 0$ (in other words, as could be expected—when the traffic occupying two devices per call is zero).

4.5 Case of Any Number of Traffic Categories

Let us consider a group of x devices to which the following traffic is offered:

$A_i (i = 1, 2, \cdots, X)$ of average holding time equal, respectively, to $h_i (i = 1, 2, \cdots, X)$ and each conversation of any traffic A_i occupying only one device;

$B_j (j = 1, 2, \cdots, Y)$ of holding time equal to h'_j and each conversation occupying two devices;

$C_k (k = 1, 2, \cdots, Z)$ of holding time equal to h''_k and each conversation occupying three devices, and so on.

From the above, it can be seen that the state equations will be separable and that the parts of the traffic occupying the same number of devices per conversation can be grouped. The values of the different probabilities of occupation will be independent of the average holding time.

Taking

$$A = \sum_1^X A_i$$

$$B = \sum_1^Y B_j$$

$$C = \sum_1^Z C_k$$

and calling $P_{m,n,p,\ldots}$ the probability that at any one instant m conversations of traffic A, n conversations of traffic B, p conversations of traffic C, $\cdots$, are in progress, we can write

$$P_{m,n,p,\ldots} = \frac{A^m}{m!} \frac{B^n}{n!} \frac{C^p}{p!} \cdots P_{0,0,0,\ldots}$$

with

$$\frac{1}{P_{0,0,0,\ldots}} = \sum_{p=0}^{X-\ldots} \sum_{n=0}^{X-3p\ldots} \sum_{m=0}^{X-2n-3p\ldots} \frac{A^m}{m!} \frac{B^n}{n!} \frac{C^p}{p!}.$$

The loss probabilities will be given by equations analogous to (25) and (26). Moreover, it can be seen that (25) and (26) are valid even if the traffic A and the traffic B comprise several quantities of traffic having different average holding times, such that

$$A = \Sigma A_i, \qquad B = \Sigma B_j.$$

4.6 Numerical Results

The values of the blocking probabilities have been computed in a particular case chosen from practical available data for existing exchanges. It has been found that the traffic A, which occupies one device per call, amounts

to about 82 percent of the total offered traffic; the traffic B, which occupies two devices per call, is then 18 percent of the total offered traffic. The loss probabilities $\mathcal{B}_X(A)$ and $\mathcal{B}_X(B)$ have been calculated with a Gamma ET computer for the following eight values of A and B traffic in erlangs.

A	41	82	123	246	369	492	820	1230
B	9	18	27	54	81	108	180	270

These values cover the range of traffic from that of a small exchange to that of a large exchange. Therefore, for any exchange, we can obtain at once the loss probability of the memory under consideration, provided the values of A and B are such that $B/A = 18/82$.

Figure 1 gives the values (from 10^{-1} to 10^{-4}) of $\mathcal{B}_X(A)$ and $\mathcal{B}_X(B)$ versus X for these eight values of A and B.

For comparison, the number of devices necessary to handle traffic A and traffic B when they are offered to two different groups, has been calculated with the Erlang formula.

Two numbers X' and X'' (comparison criteria) can be compared with X.

(**A**) $X' = X'_1 + X'_2$, where X'_1 is the number of devices of the first group handling the traffic A with the loss $\mathcal{B}_X(A)$ and X'_2 is the number of devices of the second group handling the traffic B with the loss $\mathcal{B}_X(B)$.

(**B**) $X'' = X''_1 + X'_2$ where X''_1 is the number of devices of the first group handling the traffic A with the loss $\mathcal{B}_X(B)$.

The value X'' is introduced because $\mathcal{B}_X(B)$ is always higher than $\mathcal{B}_X(A)$ and the loss imposed is generally the same for both A and B. It thus can be said that the loss probability of the memory is $\mathcal{B}_X(B)$.

In Figure 1, two additional scales give the values of X' and X'' so obtained.

To sum up the results, the ratios X'/X and X''/X versus A are shown in Figure 2 for $\mathcal{B}_X(B) = 10^{-1}, 10^{-2}, 10^{-3}$, and 10^{-4}. The savings obtained by commoning the memory are then given in percent by

$$100\left(\frac{X'}{X} - 1\right)$$

if the first comparison criterion is used, and

$$100\left(\frac{X''}{X} - 1\right)$$

if the second comparison criterion is used.

5. Conclusion

This study has two interests.

(**A**) Theoretical interest. It is an extension of the work already done on Markov processes and a generalization of the theories developed in the field.

(**B**) Practical interest. Figure 2 shows the percentage of devices that can be saved by using a single group of devices (memory) to handle two categories of traffic occupying one or two devices per call. It can be seen that these savings decrease when the traffic increases and they increase when the loss decreases. For a loss of 10^{-4}—a typical value—the savings amount to about 4 percent for heavy traffic and 18 percent for light traffic.

6. References

1. J. W. Cohen, "The Generalized Engset Formulae," *Philips Telecommunication Review*, number 18, page 158; 1957.

2. A. Blanc-Lapierre and R. Fortet, "Théorie des Fonctions Aléatoires," Masson Éditeur, Paris; 1955.

7. Appendix

The set $(\Sigma + \mathcal{G})$ of the system Σ of the sources and of $\mathcal{G}$ has its state completely and sufficiently defined by giving:

(**A**) Indexes $j_1, j_2, \cdots, j_r, \cdots, j_k$ of the presently active sources; their number assumed to be $k (k = 0, 1, \cdots, N)$. The other sources are then inactive; their indexes will be denoted by $h_1, h_2, \cdots, h_s, \cdots, h_{N-k}$.

(**B**) Durations $u_1, u_2, \cdots, u_r, \cdots, u_k$ for which the active sources $S_{j1}, \cdots, S_{jk}$ have been active.

(**C**) Durations $v_1, v_2, \cdots, v_s, \cdots, v_{N-k}$ for which the inactive sources $S_{h1}, \cdots, S_{hN-k}$ have been inactive.

The state defined by these parameters will be represented by the symbol $x_k(j, u; h, v)$, the set of these possible states $x_k(j, u; h, v)$ of $(\Sigma + \mathcal{G})$ by $\mathfrak{X}$, and finally the state of $(\Sigma + \mathcal{G})$ at instant t by $x(t)$. Under these conditions, it is clear that the evolution of $(\Sigma + \mathcal{G})$ is a homogeneous Markov process $\mathfrak{M}$, whose transition probability

$$P_r[x(\tau)\epsilon e/x(t) = x]$$

will be designated by $P(\tau - t; x, e)$ [where $\tau > t$, where e is any set of possible states, and where x is any possible state, that is, any one of the $x_k(j, u; h, v)$].

The process $\mathfrak{M}$ is of the permanent discontinuous type studied by W. Feller (refer to [2], chapter VII). The Feller study was made with hypotheses that are not quite satisfied by $\mathfrak{M}$; nevertheless, his methods and results are easily adaptable to the case of $\mathfrak{M}$ if an appropriate hypothesis is made; hypothesis H, defined in Section 3, will be taken. Then it is found that:

(**A**) Under hypothesis H, the transition probability as a function of the set e has a density

$$p_k(j, u; h, v/x; \tau - t)$$

the meaning of which is the following: If e is the set of states $x_k(j, \alpha; h, \beta)$ for which

$$u_1 \leqslant \alpha_1 < u_1 + \mathrm{d}u_1, \cdots, u_k \leqslant \alpha_k < u_k + \mathrm{d}u_k$$

$$v_1 \leqslant \beta_1 < v_1 + \mathrm{d}v_1, \cdots,$$
$$v_{N-k} \leqslant \beta_{N-k} < v_{N-k} + \mathrm{d}v_{N-k}$$

in which the $\mathrm{d}u_r$, $\mathrm{d}v_s$ are >0, $P(\tau - t; x, e)$ is equivalent to

$$p_k(j, u; h, v/x; \tau - t)\mathrm{d}u_1 \cdots \mathrm{d}u_k \mathrm{d}v_1 \cdots \mathrm{d}v_{N-k}$$

when the $\mathrm{d}u_r$, $\mathrm{d}v_s$ are small.

(**B**) As a function of the u, of the v, and of t, p_k satisfies

$$\frac{\partial}{\partial t} p_k + \sum_{r=1}^{k} \frac{\partial}{\partial u_r} p_k + \sum_{s=1}^{N-k} \frac{\partial}{\partial v_s} p_k$$
$$= -\left(\sum_{r=1}^{k} \frac{f_{jr}(u_r)}{1 - F_{jr}(u_r)} + \sum_{s=1}^{N-k} \frac{g_{hs}(v_s)}{1 - G_{hs}(v_s)} \right) p_k \tag{40}$$

which expresses one of the two Kolmogorov equations relating to the process $\mathfrak{M}$.

(**C**) p_k is the only solution of (40) that has the properties of a transition probability for process $\mathfrak{M}$.

It is remarkable that (40) is independent of rule $\mathfrak{R}$, provided that this rule is of the type assumed in Section 2. Now we observe that a function $f(u_1, \cdots, u_k; v_1, \cdots, v_{N-k})$ independent of t and τ and of the form

$$f(u, v) = \lambda \prod_{r=1}^{k} [1 - F_{jr}(u_r)] \times$$
$$\prod_{s=1}^{N-k} [1 - G_{hs}(v_s)] \tag{41}$$

in which λ is a constant, satisfies (40).

From this remark the following consequences can be deduced.

(**A**) Suppose that $\mathfrak{M}$ has a stationary evolution; that is, that there exists an a priori probability

$$\prod(e) = P_r[x(\tau)\epsilon e]$$

independent of τ and compatible with the transition probability $P(\tau - t; x, e)$ of $\mathfrak{M}$; suppose that $\prod(e)$ has a density $\prod_k(j, u; h, v)$ in the same way that $P(\tau - t; x, e)$ has a density $p_k(j, u; h, v/x; \tau - t)$; and finally suppose that, as this is intuitive, $\prod_k(j, u; h, v)$ satisfies (40); that is, as $\prod_k$ is independent of t and τ, that

$$\sum_{r=1}^{k} \frac{\partial}{\partial u_r} \prod_k + \sum_{s=1}^{N-k} \frac{\partial}{\partial v_s} \prod_k$$
$$= -\left(\sum_{r=1}^{k} \frac{f_{jr}(u_r)}{1 - F_{jr}(u_r)} + \sum_{s=1}^{N-k} \frac{g_{hs}(v_s)}{1 - G_{hs}(v_s)} \right) \prod_k \tag{42}$$

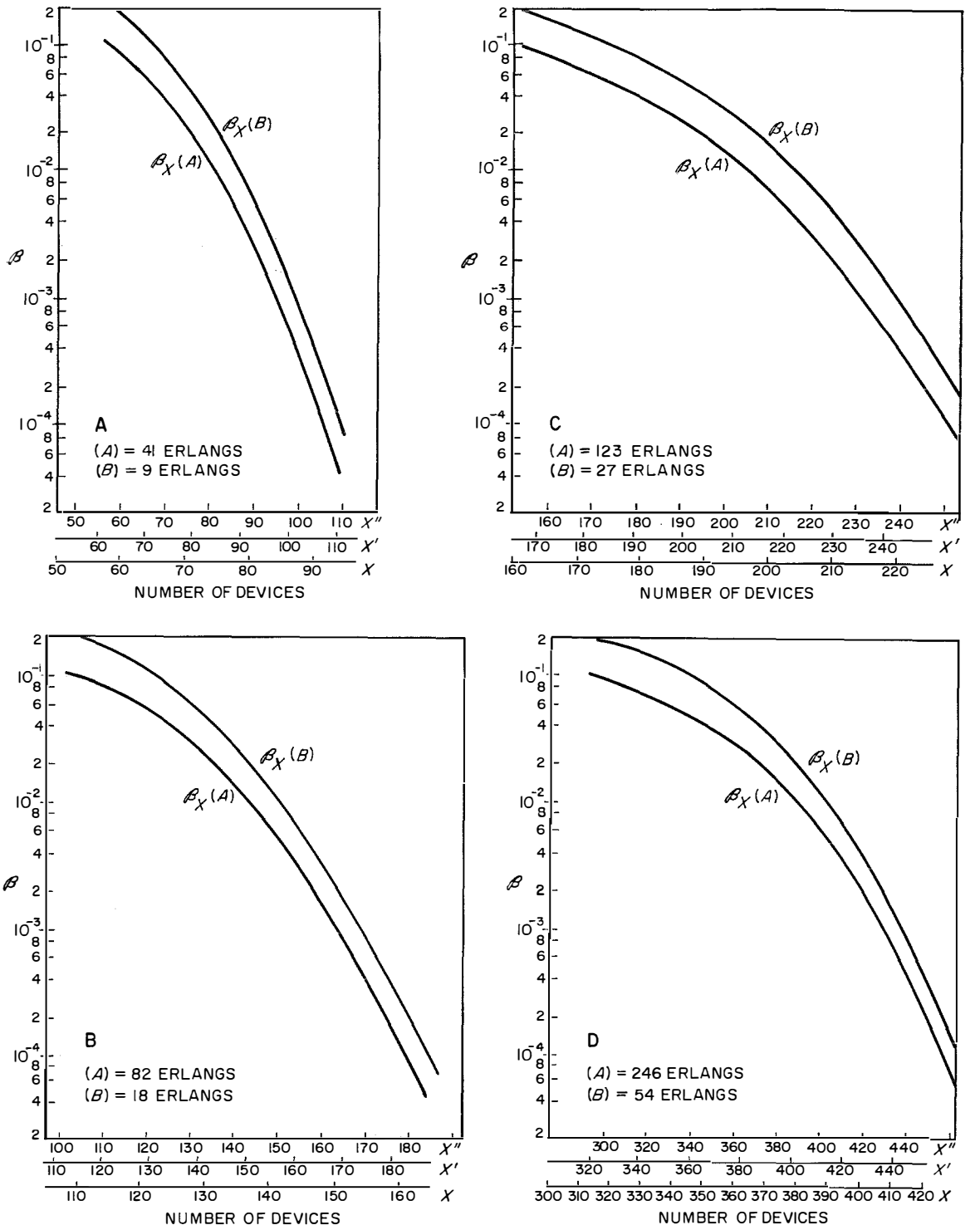
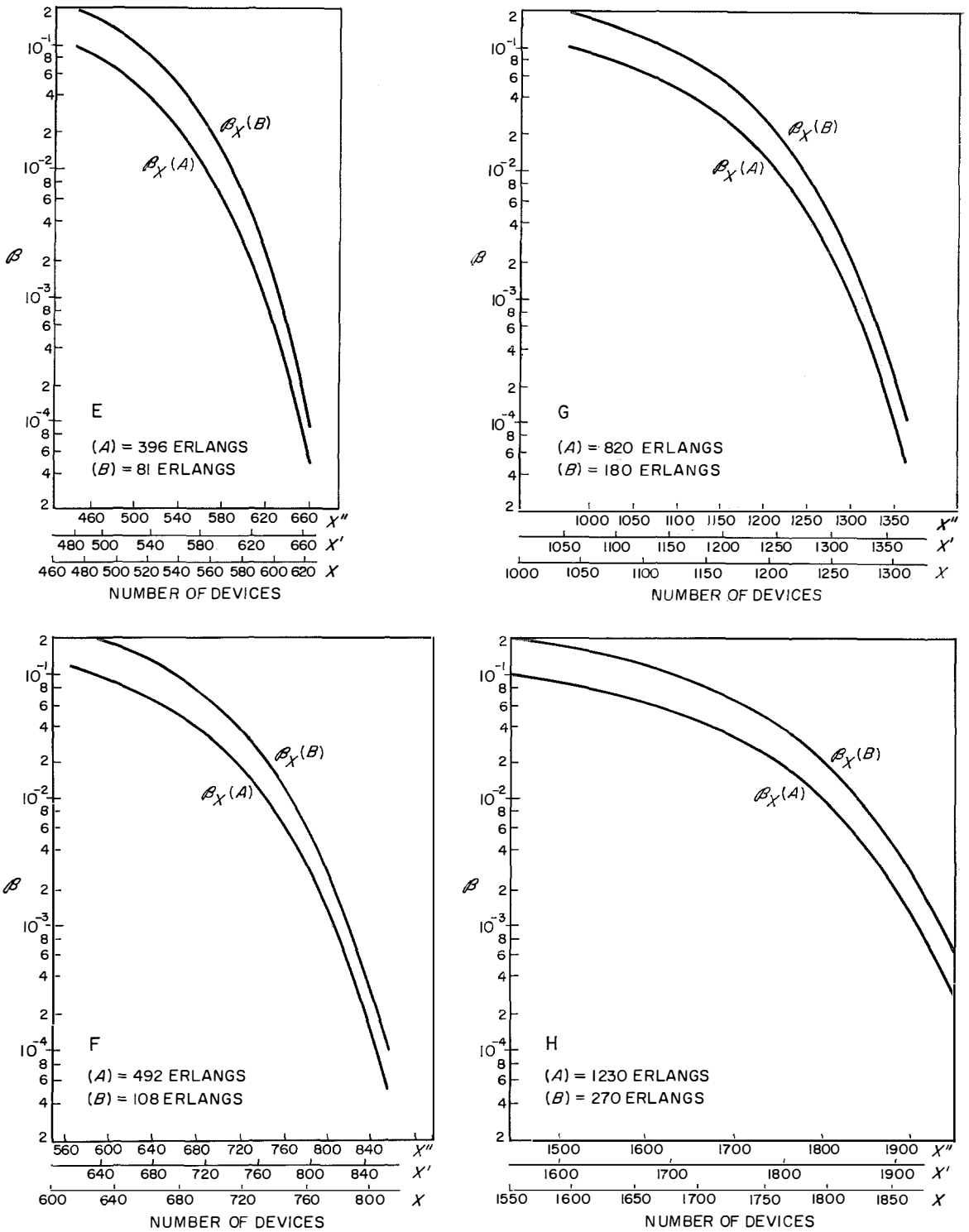


FIGURE 1—These 8 pairs of curves show blocking probabilities as a function of the number of devices. $B_X(A)$ and



$\mathcal{B}_X(B)$ are for traffic types (A) and (B) that require 1 and 2 devices per conversation, respectively.

According to (41), (42) can be satisfied by taking

$$\Pi_k(j, u; h, v) = \lambda \prod_{r=1}^{k} [1 - F_{jr}(u_r)] \times \prod_{s=1}^{N-k} [1 - G_{hs}(v_s)] \quad (43)$$

where the constant

$$\lambda = \lambda_k(j_1, \cdots, j_k)$$

can depend on k and on $j_1, \cdots, j_k$.

Suppose that $\prod_k$ is actually of the form of (43), and consider the corresponding stationary evolution of $\mathfrak{M}$; let t be any instant, and dt any infinitely small positive number; suppose that $x(t) = x_k(j, u; h, v)$ with

$u_1, \cdots, u_k$	positive and finite;
$v_2, \cdots, v_{N-k}$	positive and finite;
v_1	positive, but $< dt$, and therefore infinitely small.

Neglecting events of infinitely small probabilities of an order higher than the first one, the above implies that at the instant $t - dt$, $x(t - dt)$ was a state of the type $x_{k+1}(j_1, u_1 - dt, \cdots, j_k, u_k - dt, h_1, \tau; h_2, v_2 - dt, \cdots, h_{N-k}, v_{N-k} - dt)$, where the value of τ may be arbitrary; and that in the interval $(t - dt, t)$ S_{h1} became inactive, an event whose probability is equivalent to

$$\frac{f_{h1}(\tau)}{1 - F_{h1}(\tau)}\, dt.$$

Taking (43) into account, we get

$$\lambda_{k+1}(j_1, \cdots, j_k, h_1) = \lambda_k(j_1, \cdots, j_k).$$

In other words, the constant λ of (43) is independent of k and of the $j_1, \cdots, j_k$. It is sufficient to use this fact, as in Section 3, to obtain (10) and (11).

(**B**) Let

$$\Delta_k(j, u; h, v/x, \tau - t) = \lambda_k(j_1, \cdots, j_k; x, \tau - t) \times \prod_{\alpha=1}^{k} [1 - F_{j\alpha}(u_\alpha)] \prod_{\beta=1}^{N-k} [1 - G_{h\beta}(v_\beta)]$$

with $\lambda_k(j_1, \cdots, j_k; x, \tau - t)$ independent of the u and the v. As a function of the u and the v, Δ_k is the solution of (40).

We can define values $\bar{\lambda}_k(j_1, \cdots, j_k; x, \tau - t)$ of the $\lambda_k(j_1, \cdots, j_k; x, \tau - t)$ such that the corresponding values $\bar{\Delta}_k$ of the Δ_k have the properties of a transition probability for the process $\mathfrak{M}$. This can be seen by using procedures quite similar to those mentioned on the preceding page. The details, however, are more complicated and will be omitted; but a result of (**C**), above, is that the $\bar{\Delta}_k$ are identified with the p_k: thus we have obtained the p_k constituting the transition probability of $\mathfrak{M}$.

Now it is seen that

$$\lim_{\tau - t \to +\infty} \bar{\lambda}_k(j_1, \cdots, j_k; x, \tau - t) = \frac{1}{\left(\prod_{\alpha=1}^{N} \nu_\alpha\right)\Omega}, \quad (44)$$

Thus, referring to (42), it is seen that $\mathfrak{M}$ not only has a stationary evolution but that it has only one; any evolution of $\mathfrak{M}$ tends toward this unique stationary evolution.

7.1 Possibilities of Generalization

Hypothesis H can certainly be replaced with a broader hypothesis; we have not, however, actually proceeded to such an extension. Moreover, in the equations obtained for the $\prod_k$ and the p_k, the laws of probability of the durations of conversations and of the durations of inactivity are used only through their distribution functions, F_j and G_h, and not by their densities. Thus it is intuitive that hypothesis H might be eliminated purely and simply on condition, of course, that hypotheses (3) and (6) are kept; it would certainly be interesting to prove this beyond question.

Furthermore, the broad generality of (40) and hence of its associated form (41) should be noted; in fact, (40) concerns only the evolution of the set of sources and is worth whatever the rules $\mathfrak{R}$ of seizure of the devices may be; in

the preceding, the rules $\mathfrak{R}$ are used only for determining $\lambda_k(j_1, j_2, \cdots, j_k)$.

7.2 Remark

In the application developed in Section 4, the group of devices is assumed to be fully commoned; this means that as soon as any device is free, it can be occupied by a call from traffic A, and as soon as any two devices are free, they can be occupied by a conversation from traffic B. Stricter conditions could, of course, be introduced (for instance, a call from traffic B requires two adjacent devices).

In such a case, and according to what was said above, (40) and (41) would remain valid; but the determination of the $\lambda_k(j_1, \cdots, j_k)$

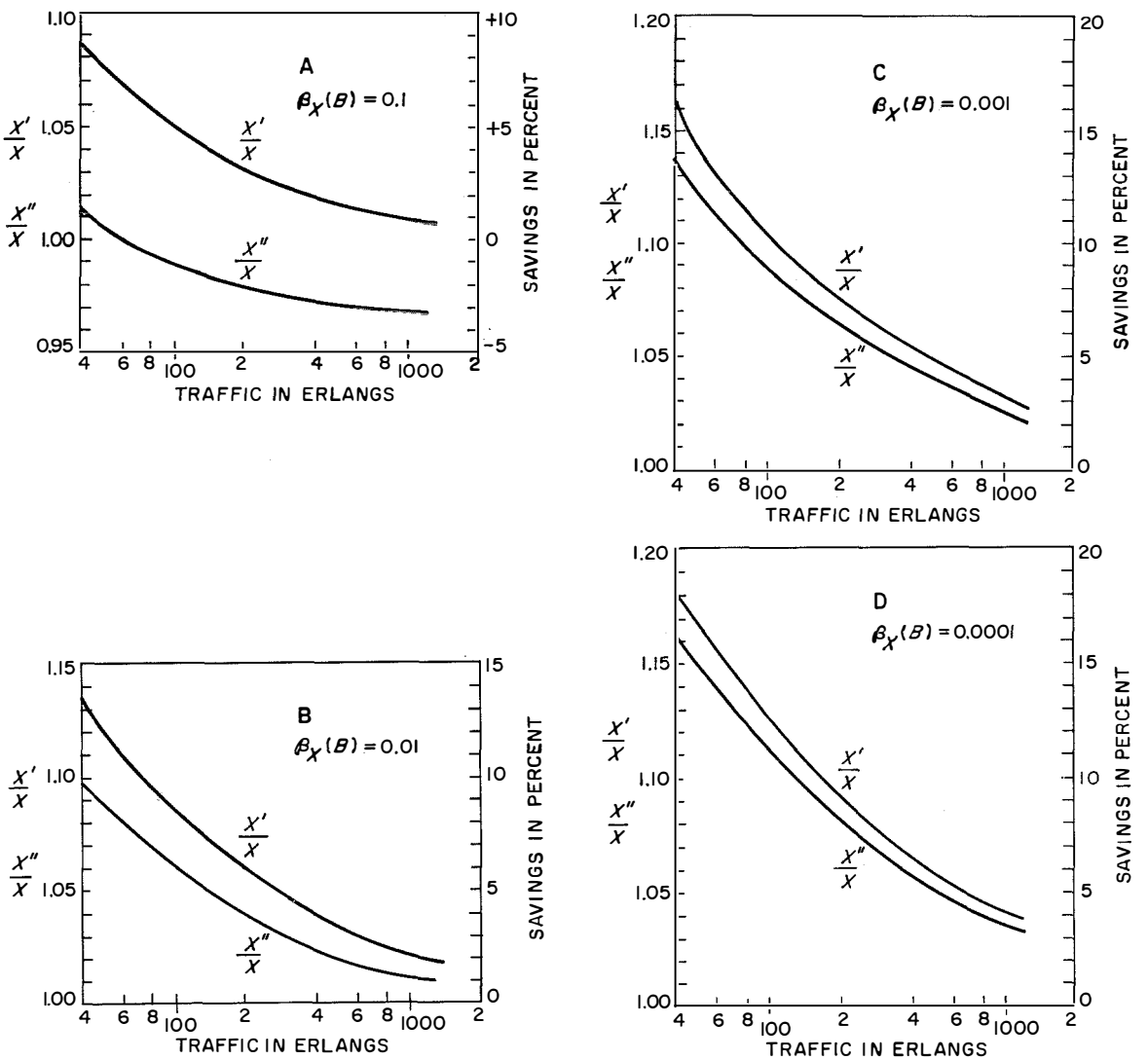


Figure 2—Percentage of devices that can be saved by using a single memory for handling traffic A and traffic B. $B/A = 18/82$.

would raise problems that are quite different and much-more difficult to solve.

Robert Fortet was born in Dordogne, France, in 1912. He was a student of the École Normale Supérieure and obtained the grade of Doctor in Sciences at the University of Paris in 1939.

Since 1947 he has been a consultant on mathematics at Laboratoire Central de Télécommunications; he is now primarily concerned with the statistical theory of telephone traffic.

Professor Fortet occupied the Chair of Theory of Probability and Mathematical Physics at the Faculty of Sciences of Paris since 1959. He is a member of the International Advisory Committee of the International Teletraffic Congress, and a Fellow of the Institute of Mathematical Statistics.

Charles Grandjean was born in Bourges, France, on 27 October 1930. He received the License ès Sciences, and obtained in 1952 the diploma of "Études Supérieures en Calcul des Probabilités et Statistiques."

He worked for a few years in the research laboratories of the Office National d'Études et de Recherches Aéronautiques and of Philips Telecommunications. He joined Laboratoire Central de Télécommunications in 1960 and is in charge of studies related to telephone traffic theory.

Realizing the General Two-Terminal-Pair Network with Independently Prescribed Transfer Function and Reflection Coefficient

NAI-TA MING

Standard Elektrik Lorenz AG; Stuttgart, Germany

1. General

This paper is mainly based on the working-parameter theory, which differs from the insertion-loss theory by only a constant. If the source resistance is equal to the load resistance, both are identical.

Accordingly, terms related to the working-parameter theory should be identified, wherever necessary, by the word "working" (working transfer function, working driving-point impedance, et cetera).

The working driving-point impedance (or admittance) of a 4-pole network is that existing at one end of the network while the other end has the load resistance R. In the normalized case presented here, $R = 1$.

A dissipative 4-pole (or 2-terminal-pair) network is one in which each inductance has a series resistance; each capacitance has a parallel conductance; and the primary or secondary inductance of each transformer has its own resistance in series.

1.1 Problem

Of all the published work on network synthesis employing the insertion-loss theory, none appears to have dealt with the case in which a 4-pole network (hereafter called a 4-pole) is to be realized when the transfer constant (the natural logarithm of the "working" transfer function) and the reflection-coefficient tolerance—or this coefficient itself—are prescribed simultaneously and independently. This case was mentioned in a footnote of [1], but no concrete method for realization was given. For inductor-capacitor 4-poles, only the above-defined transfer constant and the reflection coefficient can be prescribed at the same time in a certain correlation. To this day, however, there has been no useful method of realizing an inductor-capacitor 4-pole when its transfer constant and the reflection-coefficient tolerance (or the reflection coefficient) are prescribed independently and simultaneously.

Theoretically, this problem can be solved using the principle described in this paper. Probably cost has discouraged a systematic treatment of this problem before. So far only the approximation method [2–4] could be used in realizing a 4-pole with compensation for dissipation. Here the reflection coefficient

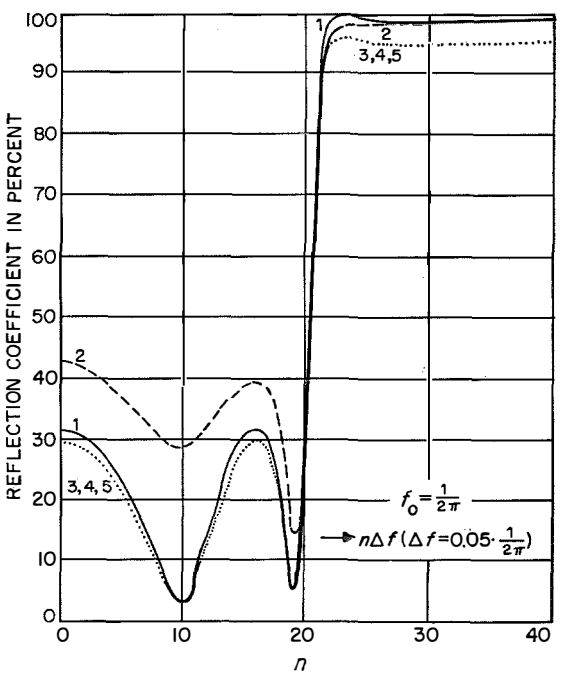


Figure 1—Reflection coefficients of ladder network with prescribed roots of $g(\lambda)$, $h_1(\lambda)$ and $f(\lambda)$. Curve 1 is for a reactance four-pole; curve 2 is for resistor-inductor-capacitor four-pole with consideration of dissipation of each inductor and each capacitor according to the classical approximate method; curve 3 is for resistor-inductor-capacitor four-pole with consideration of dissipation only for inductors according to a new rigorous method without satisfying the prescribed constant forward-loss function; curve 4 is the same as curve 3 but with prescribed constant forward-loss function; and curve 5 is similar to curve 3 but contains an arbitrary prescribed equivalent circuit.

cannot be prescribed in advance, and unfortunately it is much greater than that of a 4-pole without compensation for dissipation. As may be seen from Figure 1, the reflection-coefficient characteristic of a 4-pole with compensation for dissipation by the said approximation method (curve 2) is much worse than that without compensation for dissipation (curve 1) and also worse than that obtained with the method herein described, considering only the coil loss (curve 3). To evolve a general theory applicable to both 2-pole and 4-pole networks with resistor-inductor-capacitor elements, and to improve the above-mentioned unsatisfactory performance, the following investigation has been carried out.

1.2 Result

The following results were obtained.

(**A**) Proof of the realizability of a prescribed (working driving-point) impedance, (that is, the reflection coefficient $T_1^{-1}(\lambda)$ is thus prescribed) by a lossy 4-pole whose transfer constant, apart from a constant factor, is likewise prescribed independently of $T_1^{-1}(\lambda)$.

(**B**) A 2-pole and 4-pole compatibility theory has been obtained that permits design of a 2-pole that will represent a certain desired 4-pole.

(**C**) The loss introduced in a 4-pole by the reactance elements (inductors, capacitors) and the loss introduced in the same 4-pole by resistor elements may have a linear or nonlinear interdependence. This gives rise to the question under what prescribed conditions these two losses can be linear and nonlinear for the general characteristic equation

$$k^2 g(\lambda)g(-\lambda) = f(\lambda)f(-\lambda) + h_1(\lambda)h_1(-\lambda) + j_1(\lambda)j_1(-\lambda).$$

In other words, which loss is that kind of forward dissipation (ratio of the power dissipation in the form of joulean heat to the maximum power transfer) controlling the linear or nonlinear interdependence of the two types of loss. The following two cases are given.

(**A**) Linear and nonlinear dependence for a prescribed loss, apart from a constant factor, and for a prescribed input working driving-point impedance (W_{b1}).

(**B**) Linear and nonlinear dependence for a prescribed loss, apart from a constant factor, and for an input reflection coefficient not exceeding a prescribed tolerance.

In addition, let us consider which internal resistances can be varied and with what results on the loss. Five procedures are given to realize a 4-pole starting from its W_{b1} under a certain condition. In all cases of procedures (**A**) through (**D**) and in most cases of procedure (**E**), the realization is based mainly on the ladder structures.

(**A**) Resistor-capacitor elements with a minimum number of reactance elements, or cascaded 4-pole sections of resistor-capacitor elements (Section 2.2.1).

(**B**) Resistor-inductor-capacitor elements (Section 2.2.2).

(**C**) Resistor-inductor-capacitor elements with a minimum number of reactance elements. Only the inductor dissipation is considered (Section 2.2.3).

(**D**) Resistor-inductor-capacitor elements with a minimum number of reactance elements, considering inductor and capacitor dissipation by approximation (Section 2.2.4).

(**E**) Resistor-inductor-capacitor elements, considering mainly the inductor dissipation (Section 2.2.5).

Procedures (**C**) through (**E**) are applicable in this paper to the attenuation poles on the imaginary axes of the λ plane. The 4-poles obtained by these procedures have ladder structures and many-fewer reactance elements than required in Section 2.1.5, in which 4-poles are presented as proof. Procedure (**A**) and particularly procedure (**C**) are highly suitable eco-

nomically. Procedure (**C**) always results in a reliable network synthesis based on Section 2.1.13 if the dissipation of each coil is sufficiently low. Procedure (**E**) is indispensable if one or more equivalent networks (for example, a crystal circuit) are to be considered as circuit elements in an over-all 4-pole or if the loss of each coil must be very large.

Procedure (**A**) is applied for resistor-capacitor networks and procedure (**C**) for resistor-inductor-capacitor networks. The frequency characteristic for procedure (**A**) is more limited than that of all the others, but it possesses some advantages (better reliability, low cost, suitability for very-low frequencies, et cetera). The characteristic for procedure (**B**) is less limited than that of procedure (**A**) but more limited than that of procedure (**C**). The application of procedure (**B**) depends on certain characteristics of the input working driving-point impedance W_{b1}. In such circumstances one must use this procedure for a ladder network as in the sub-4-terminal network a–a′—2–2′ of Figure 7 for example 5.

Procedure (**D**) for a minimum number of reactance elements cannot always be reliably carried out. This case is not treated here.

Five examples are given in Figures 3 through 7. To be able to compare these examples, their prescribed characteristic functions $g(\lambda)$, $h_1(\lambda)$, and $f(\lambda)$ are identical. Conversely, the five realized 4-poles are used to compute losses and reflection coefficients with the aid of the network analysis. The first example applies to the reactance 4-pole. The second example applies to the 4-pole with compensation for dissipation after the classic approximation method [2–4]. The third example is valid by procedure (**C**) for a 4-pole with compensation for dissipation only of the coil; the highest coefficient of the prescribed $f(\lambda)$ is not retained. The fourth example is identical to the third, but the highest coefficient $f(\lambda)$ is retained as prescribed. The fifth example has been completed with procedures (**E**) and (**B**), but with a greater number of reactance elements. The loss and the reflection coefficients of the 4-poles realized in examples 1 through 5 are numbered accordingly in Figures 1 and 2. The circuit diagrams of these five examples are shown in Figures 3 through 7 with all network-element ratings. It may be readily seen from Figure 1 that the reflection coefficients of examples 3 and 4, realized by the new procedures, are somewhat less than in example 1 and much less than in example 2. The loss of the network examples 3 and 4 in the pass-band range is likewise less than obtained in example 2 with compensation for dissipation by the approximation method, and their attenuation characteristics in the vicinity of the cutoff normalized frequency are steeper than in example 2.

The perfect coupling of the Brune process can

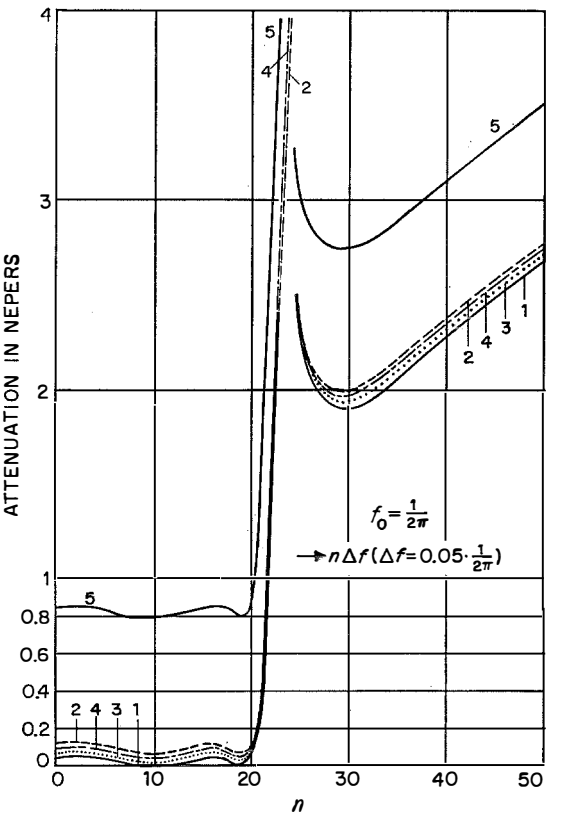


Figure 2—Attenuation curve of ladder network with prescribed roots of $g(\lambda)$, $h(\lambda)$, and $f(\lambda)$. Curves 1 through 5 correspond to the respective cases in Figure 1.

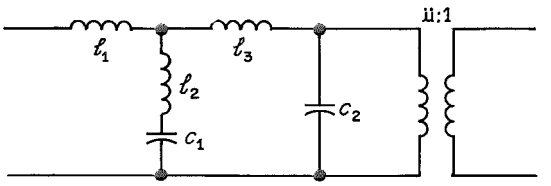


Figure 3—Reactance four-pole for curve 1 in Figures 1 and 2.

$l_1 = 0.863\,90$	$c_1 = 0.487\,20$
$l_2 = 1.459\,8$	$c_2 = 0.872\,70$
$l_3 = 1.710\,6$	$ii = 1.375\,6$ $(ii^2 = 1.892\,3)$

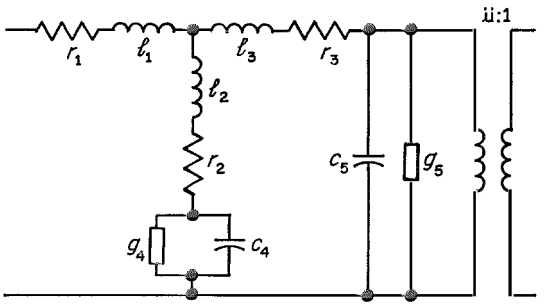


Figure 4—Four-pole for curve 2 in Figures 1 and 2.

$r_1 = 0.008\,721\,35$	$l_1 = 1.406\,67$
$r_2 = 0.013\,677\,8$	$l_2 = 2.206\,09$
$r_3 = 0.016\,324\,9$	$l_3 = 2.633\,05$
$g_4 = 0.001\,998\,86$	$c_4 = 0.322\,397$
$g_5 = 0.003\,169\,66$	$c_5 = 0.511\,236$
$ii = 1.561\,27$	$ii^2 = 2.437\,56$

be avoided by procedure (**E**). Figure 8 shows calculations for part of example 5.

To save computer time, these examples were computed by the new process without any attempt to retain the low Q of each coil.

2. Fundamental Theory and Procedure

2.1 Fundamental Theory

2.1.1

Throughout this paper it is assumed that the internal source resistance R_1 associated with the electromotive force and the load resistance R_2 are normalized ($R_1 = R_2 = 1$), and that the frequency f is also normalized to a certain frequency f_0 so that $\Omega = f/f_0 = \lambda/j$ where $j = (-1)^{1/2}$.

2.1.2

As is well known, each reactance 4-pole must obey the following characteristic equation.

$$k^2 g(\lambda)g(-\lambda) = f(\lambda)f(-\lambda) + h_1(\lambda)h_1(-\lambda) \quad (1)$$

where $g(\lambda)$ is a Routh-Hurwitz polynomial, $f(\lambda)$ is an even or odd function in the form of a polynomial, and $h_1(\lambda)$ is likewise a polynomial. The transfer function $S(\lambda) = kg(\lambda)/f(\lambda)$ and the input echo transfer function (or the inverse value of the reflection coefficient)

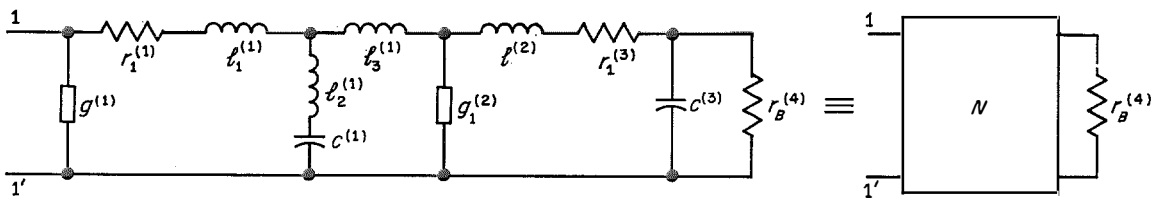


Figure 5—Four-pole for curve 3 in Figures 1 and 2.

$g^{(1)} = 0.021\,375\,355\,404$	$1/g_1^{(2)} = 397\,85.753\,384$	$l_2^{(1)} = 1.457\,893\,079\,4$	$l^{(2)} = 2.253\,987\,448\,1$
$1/g^{(1)} = 46.782\,847\,868$	$r_1^{(3)} = 0.000\,052\,268\,136\,081$	$l_3^{(1)} = -0.542\,997\,869\,27$	$c^{(3)} = 0.872\,654\,550\,96$
$r_1^{(1)} = 0.021\,405\,454\,970$	$r_B^{(4)} = 1.892\,826\,359\,0$	$c^{(1)} = 0.487\,827\,269\,41$	$1/r_B^{(4)} = 0.528\,310\,478\,52$
$g_1^{(2)} = 0.000\,025\,134\,625\,210$	$l_1^{(1)} = 0.865\,271\,561\,33$		

$T_1(\lambda) = kg(\lambda)/h_1(\lambda)$ must comply with the following conditions.

$$|S(\lambda)| \geq 1 \qquad (2A)$$

$$|T_1(\lambda)| \geq 1. \qquad (2B)$$

Then (1) is necessary and also sufficient to realize a reactance 4-pole [2, 4, 5].

2.1.3

A general 4-pole must obey the following characteristic equation.

$$k^2 g(\lambda)g(-\lambda) = f(\lambda)f(-\lambda) + h_1(\lambda)h_1(-\lambda) + j_1(\lambda)j_1(-\lambda). \qquad (3)$$

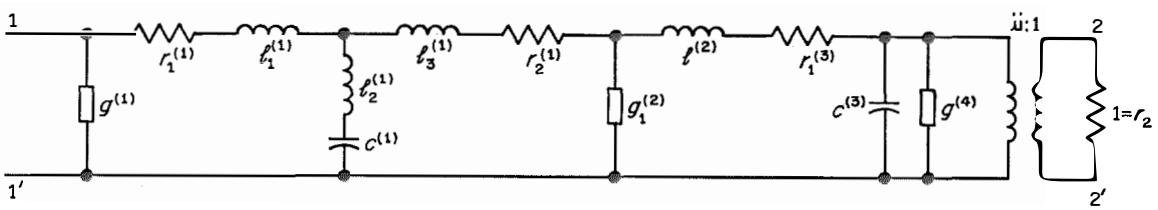


Figure 6—Four-pole for curve 4 in Figures 1 and 2.

$g^{(1)} = 10^{-2} \times 2.137\ 535\ 540\ 4$	$l_3^{(1)} = -10^{-1} \times 5.429\ 978\ 692\ 7$	$g^{(4)} = 10^{-2} \times 2.191\ 577\ 953\ 1$
$1/g^{(1)} = 10 \times 4.678\ 284\ 786\ 8$	$r_2^{(1)} = 10^{-5} \times 2.613\ 406\ 802\ 5$	$1/g^{(4)} = 10 \times 4.562\ 922\ 339\ 1$
$r_1^{(1)} = 10^{-2} \times 2.140\ 545\ 497\ 0$	$g_1^{(2)} = 10^{-5} \times 2.513\ 462\ 522\ 7$	$l^{(2)} = 2.253\ 987\ 445\ 2$
$l_1^{(1)} = 10^{-1} \times 8.652\ 715\ 613\ 3$	$1/g_1^{(2)} = 10^{4} \times 3.978\ 575\ 335\ 7$	$c^{(3)} = 10^{-1} \times 8.726\ 545\ 520\ 9$
$l_2^{(1)} = 1.457\ 893\ 079\ 4$	$r_1^{(3)} = 10^{-5} \times 2.613\ 406\ 800\ 6$	$\ddot{u} = 1.405\ 255\ 922\ 4$
$c^{(1)} = 10^{-1} \times 4.878\ 272\ 694\ 1$		

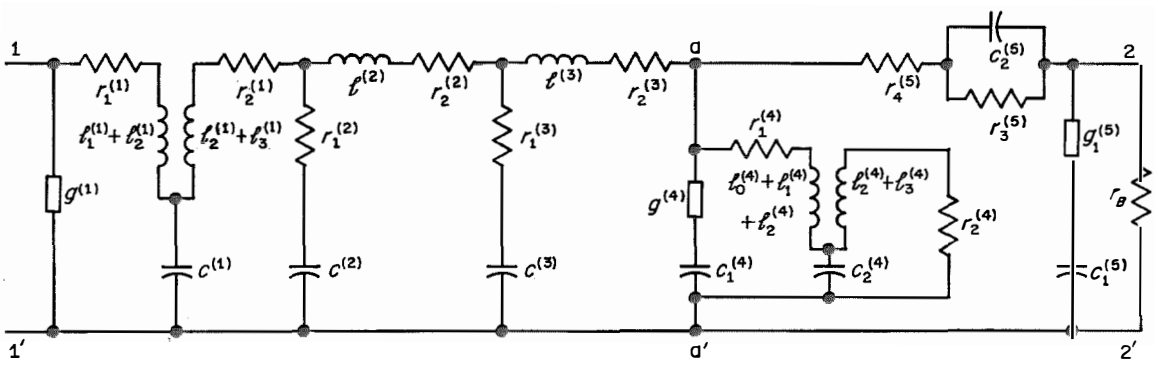


Figure 7—Four-pole for curve 5 in Figures 1 and 2.

$g^{(1)} = 10^{-2} \times 2.137\ 535\ 540\ 4$	$r_2^{(2)} = 10^{-3} \times 2.250\ 000\ 000\ 0$	$l_2^{(4)} = 10^{-2} \times 1.395\ 757\ 794\ 7$
$r_1^{(1)} = 10^{-2} \times 2.140\ 545\ 497\ 0$	$r_1^{(3)} = 10^{-3} \times 2.246\ 709\ 683\ 1$	$c_2^{(4)} = 10^{-1} \times 4.247\ 058\ 877\ 7$
$l_1^{(1)} = 10^{-1} \times 8.652\ 715\ 613\ 3$	$c^{(3)} = 10^{-1} \times 4.450\ 953\ 354\ 3$	$l_3^{(4)} = 10^{-1} \times 2.533\ 200\ 932\ 9$
$l_2^{(1)} = 1.457\ 893\ 079\ 4$	$l^{(3)} = 10^{-6} \times 4.104\ 990\ 254\ 9$	$r_2^{(4)} = 2.591\ 534\ 440\ 2$
$c^{(1)} = 10^{-1} \times 4.878\ 272\ 694\ 1$	$r_2^{(3)} = 10^{-8} \times 4.104\ 990\ 254\ 9$	$r_4^{(5)} = 10^{-1} \times 4.759\ 125\ 504\ 7$
$l_3^{(1)} = -10^{-1} \times 5.429\ 978\ 692\ 7$	$g^{(4)} = 10^{2} \times 1.661\ 096\ 349\ 7$	$c_2^{(5)} = 10^{-2} \times 9.696\ 209\ 487\ 0$
$r_2^{(1)} = 10^{-3} \times 8.320\ 000\ 000\ 0$	$c_1^{(4)} = 10^{-1} \times 1.235\ 174\ 208\ 8$	$r_3^{(5)} = 1.031\ 330\ 852\ 9$
$r_1^{(2)} = 10^{4} \times 3.978\ 574\ 506\ 4$	$r_1^{(4)} = 10^{-6} \times 1.782\ 468\ 101\ 3$	$g_1^{(5)} = 5.426\ 724\ 964\ 3$
$c^{(2)} = 10^{-6} \times 2.513\ 463\ 046\ 6$	$l_0^{(4)} = 10^{-6} \times 1.678\ 074\ 328\ 2$	$c_1^{(5)} = 10^{-3} \times 5.426\ 724\ 964\ 3$
$l^{(2)} = 2.252\ 710\ 274\ 7$	$l_1^{(4)} = -10^{-2} \times 1.322\ 869\ 568\ 3$	$r_B = 5.369\ 567\ 744\ 1$

The proof of this may be found in [1]. Apart from the conditions of (2A) and (2B), the forward-dissipation function $R_1(\lambda) = kg(\lambda)/j_1(\lambda)$ should also comply with the condition

$$|R_1(\lambda)| \geq 1. \qquad (2C)$$

(To distinguish the above dissipation function from that defined or used in [1, 2, 4, 6, and 7], the latter will be called the M function herein).

2.1.4

In their physical significance

$|S(\lambda)|^{-2}$ = ratio of the power transferred by the 4-pole to the maximum power (defined below)

$|T_1(\lambda)|^{-2}$ = ratio of the input reflection power to the maximum power.

$|R_1(\lambda)|^{-2}$ = ratio of the power dissipated as joulean heat by the 4-pole to the maximum power

maximum power = that maximum power consumed by the unit load of the 4-pole with unit source resistance.

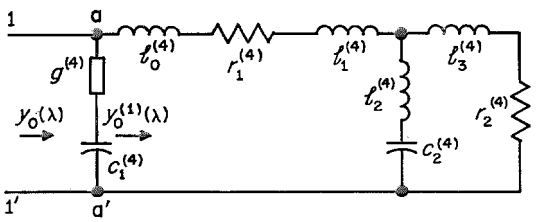


Figure 8—Four-pole to avoid the perfect coupling required by the Brune process.

$g^{(4)} = 166.109\,634\,97$ $\quad l_2^{(4)} = 0.013\,957\,577\,947$

$c_1^{(4)} = 0.123\,517\,420\,88$ $\quad c_2^{(4)} = 0.424\,705\,887\,77$

$l_0^{(4)} = 0.000\,001\,678\,074\,328\,2$ $\quad l_3^{(4)} = 0.253\,320\,093\,29$

$r_1^{(4)} = 0.000\,001\,782\,468\,101\,3$ $\quad r_2^{(4)} = 2.591\,534\,440\,2$

$l_1^{(4)} = -0.013\,228\,695\,683$

2.1.5

To realize a general 4-pole with prescribed $S(\lambda)$, $T_1(\lambda)$, and $R_1(\lambda)$, equation (3) is necessary and also sufficient. The proof that the prescribed functions $S(\lambda)$ and $T_1(\lambda)$ can be realized by a 4-pole may be stated about as follows: If a transform $\lambda = p - \epsilon$ is applied to the variables of $S(\lambda)$ and $T_1(\lambda)$, we obtain $g^*(p)$, $f^*(p)$, and $h_1^*(p)$. From this we have the input driving-point impedance

$$W_{b1}^*(p) = \frac{kg^*(p) + h_1^*(p)}{kg^*(p) - h_1^*(p)} = \frac{P^*(p)}{Q^*(p)}. \qquad (4)$$

Considering

$$N^*(p) = \frac{-P^*(p)}{f^*(p)} = a^*(p) + b^*(p) \qquad (5)$$

as the voltage-transfer function, (5) is realized after [1] by a 4-pole without coupling. Because retransform $p = \lambda + \epsilon$, the 4-pole has only dissipative network elements (each coil has a series resistance and each capacitor has a parallel conductance). After the retransform the input working driving-point impedance of this 4-pole may be written

$$W_{b1}^{(1)}(\lambda) = \frac{P^{(1)}(\lambda)}{Q^{(1)}(\lambda)} = [M_{b1}^{(1)}(\lambda)]^{-1}.$$

Through expansion with a common constant factor ν for $P^{(1)}(\lambda)$ and $Q^{(1)}(\lambda)$, we obtain

$$\nu P^{(1)}(\lambda) = a(\lambda) + b(\lambda).$$

A 2-pole in parallel to the 4-pole input is determined with an admittance function $y_0(\lambda)$ so that the final 4-pole has the prescribed input working driving-point impedance

$$W_{b1}(\lambda) = \frac{kg(\lambda) + h_1(\lambda)}{kg(\lambda) - h_1(\lambda)} = \frac{P(\lambda)}{Q(\lambda)}. \qquad (6)$$

Using the equation

$$y_0(\lambda) = M_{b1}(\lambda) - \ddot{u}_1^{-2} M_{b1}^{(1)}(\lambda) \qquad (7)$$

the voltage transfer ratio $\ddot{u}_1$ of an ideal transformer is determined so that $y_0(\lambda)$ is still an M function. As required, the 2-pole y_0 can be realized by the Brune process [5, 8], the Bott-Duffin-Richards process [9, 10], and the

author's process [4, 6, 7]. In the last-mentioned two processes, reactance elements with compensation for dissipation can be provided. The process [7] yields a ladder network with a minimum number of network elements, but with coupling. The process [4] yields a 4-pole having no coupling, but a number of superfluous elements.

Finally, the ideal transformers can be displaced to the right and merged with the normalized load. The final 4-pole has the transfer function

$$S_w(\lambda) = \ddot{u}_1 v^{-1} S(\lambda). \tag{8}$$

This is the proof for the possibility of realizing a dissipative 4-pole where the transfer function—apart from a constant factor—is prescribed; it has the prescribed input working driving-point impedance but not the minimum number of circuit elements.

2.1.6

For the prescribed function $S(\lambda) = kg(\lambda)/f(\lambda)$ and the tolerance τ (that is, maximum $|T(\lambda)|^{-1} \leq \tau$), it is possible to obtain a 4-pole from a ladder network with a minimum number m (degree of $g(\lambda)$) of reactance elements (Section 2.2.3). For the two independently prescribed functions $S(\lambda) = kg(\lambda)/f(\lambda)$ and $T_1(\lambda) = kg_t(\lambda)/h_1(\lambda)$ (rational fractional functions), numerator and denominator of $S(\lambda)$ or $T_1(\lambda)$ have to be expanded with one or the other factor, a Routh-Hurwitz polynomial. Hence, these prescribed values cannot be realized with a 4-pole having the minimum number m of reactance elements.

If the two prescribed functions $S(\lambda)$ and $T_1(\lambda)$ are two curves, it is possible to determine, by interpolation, first $S(\lambda) = kg(\lambda)/f(\lambda)$. (This can be accomplished, for instance after Cauer's theory [5] or using $\varphi_1(\lambda) = h_1(\lambda)/f(\lambda)$ from an unpublished paper of Cauer for any prescribed attenuation on the band-stop range in the sense of Chebishev). Then the inverse of the square values of $T_1(\lambda)$ is multiplied with $|g(\lambda)|^2$ in the band of interest (for instance $(0, \Omega_1)$); using a suitable interval $\Delta\Omega$, that is, $|T_1(\lambda)|^{-2}|g(\lambda)|^2$, it can always be further approximated by the Weierstrass theory with a polynomial $h_1(\lambda)h_1(-\lambda) = H(x)$ of a degree not higher than that of $g(\lambda)$. Thus we obtain the desired polynomial $h_1(\lambda)$. In the first place, the necessary and sufficient conditions [1] must be taken into account (refer to 2.1.2 and 2.1.3).

2.1.7

The following abbreviations will be used.

$$g(\lambda) = \alpha_m \bar{g}(\lambda) \tag{9A}$$

$$h_1(\lambda) = \gamma_m h(\lambda) \tag{9B}$$

$$f(\lambda) = \beta_n f(\lambda). \tag{9C}$$

α_m, γ_m, and β_n are the highest coefficients of $g(\lambda)$, $h_1(\lambda)$, and $f(\lambda)$, respectively; m is the degree of $g(\lambda)$ or $h_1(\lambda)$; and n that of $f(\lambda)$. Each of the functions $\bar{g}(\lambda)$, $h_1(\lambda)$, and $f(\lambda)$ has a highest coefficient of 1.

$$\hat{g}(\lambda) = kg(\lambda) \tag{10}$$

where k is a constant number.

It is assumed that $g(\lambda)$, $f(\lambda)$, and $h_1(\lambda)$ are relative primes having no common submultiples.

$$\varphi_1(\lambda) = \frac{h_1(\lambda)}{f(\lambda)} \tag{11A}$$

$$\zeta_1(\lambda) = \frac{j_1(\lambda)}{f(\lambda)} \tag{11B}$$

$$\phi_1(x) = \varphi_1(\lambda)\varphi_1(-\lambda) = \frac{h_1(\lambda)h_1(-\lambda)}{f(\lambda)f(-\lambda)} = \frac{H_1(x)}{F(x)} \tag{11C}$$

$$Z_1(x) = \zeta_1(\lambda)\zeta_1(-\lambda) = \frac{j_1(\lambda)j_1(-\lambda)}{f(\lambda)f(-\lambda)} = \frac{J_1(x)}{F(x)} \tag{11D}$$

$$s(x) = S(\lambda)S(-\lambda) = \frac{g(\lambda)g(-\lambda)}{f(\lambda)f(-\lambda)} = \frac{G(x)}{F(x)}. \tag{11E}$$

2.1.8

The general characteristic equation can also be represented in the following forms.

$$1 = |S(\lambda)|^{-2} + |T_1(\lambda)|^{-2} + R_1(\lambda)|^{-2} \quad (12)$$

and

$$|S(\lambda)|^2 = 1 + \phi_1(x) + Z_1(x). \quad (13)$$

It is expedient to use (12) for power and (13) for attenuation. According to (13), the loss is

$$\log_e|S(\lambda)| = \tfrac{1}{2}\log_e[1 + \phi_1(x) + Z_1(x)]. \quad (14)$$

Equation (3) is a mathematical representation applicable both to the 4-pole and to the 2-pole. Since any 2-pole may be considered as a 4-pole and vice versa, this can be taken directly from (3), which gives nothing else but the property of the numerator in the real part of the input working driving-point impedance $W_{b1}(\lambda)$ or of the admittance $M_{b1}(\lambda)$ of a 4-pole with the functions $kg(\lambda)$, $h_1(\lambda)$, and $f(\lambda)$. The numerator of the real portion of $W_{b1}(\lambda)$ or $M_{b1}(\lambda)$ is

$$k^2g(\lambda)g(-\lambda) - h_1(\lambda)h_1(-\lambda) = f(\lambda)f(-\lambda) + j_1(\lambda)j_1(-\lambda). \quad (15)$$

2.1.9

The roots of $f(\lambda)$ are defined as attenuation poles. When $W_{b1}(\lambda)$ can be expanded by a common factor $q(\lambda)$, that is

$$W_{b1}(\lambda) = \frac{q(\lambda)[kg(\lambda) + h_1(\lambda)]}{q(\lambda)[kg(\lambda) - h(\lambda)]} \quad (16)$$

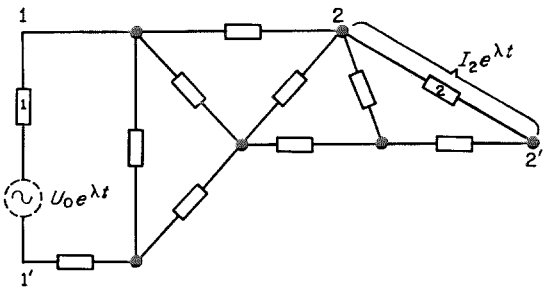


Figure 9—General configuration of a four-pole. The rectangle with the designation 1 or 2 within it is a lossy two-pole; each other rectangle designates any two-pole.

the roots of $q(\lambda)$ are defined as apparent attenuation poles. Equation (3) has then to be multiplied with $q(\lambda)q(-\lambda)$ on both sides.

2.1.10

Generally the loss can be caused by the reactance elements and the resistances and conductances within a 4-pole, but it can also be caused by the reactance elements alone, in which case it will be designated D_l. The loss caused by pure resistances in the 4-pole will be designated D_r. If the over-all loss can be expressed by

$$D = D_l + D_r$$

the D_l and D_r are said to be linear, otherwise they are nonlinear.

2.1.11

In some linear network (Figure 9), an input 1–1′ and an output 2–2′ may be selected. If now a source such as $U_0e^{\lambda t}$ is inserted between the input terminals 1–1′, we have at the output 2–2′ of the 4-pole a response $I_2e^{\lambda t}$.

$$I_2 = \frac{\alpha_{12}(\lambda)}{\alpha(\lambda)}\,U_0 \quad (17)$$

is the solution of the variables across the output 2–2′ of the 4-pole in Figure 9, from a system of integral and differential equations represented, after Laplace transform, as follows.

$$\|A_{st}(\lambda)\| \cdot \begin{Vmatrix} I_1 \\ I_2 \\ \vdots \\ I_n \end{Vmatrix} = \begin{Vmatrix} U_0 \\ \vdots \\ \end{Vmatrix}. \quad (18)$$

$\alpha(\lambda)$ is the determinant to the matrix $A_{st}(\lambda)$ and $\alpha_{st}(\lambda)$ is the algebraic complement to the element $A_{st}(\lambda)$ of the determinant $\alpha(\lambda)$. If now (17) is considered from the purely mathematical viewpoint, then

$$\left[\frac{\alpha_{12}(\lambda)}{\alpha(\lambda)}\right]_{\lambda=\lambda_1} = 0.$$

We then say: λ_1 is an attenuation pole of this network. Therefore λ_1 is the root of $f(\lambda)$.

2.1.12

Conversely, if we start from the input working driving-point impedance $W_{b1}(\lambda)$ (see (3) and (6)) as the initial function to realize a 4-pole of prescribed transfer function $kg(\lambda)/f(\lambda)$ (up to a constant coefficient), then $kg(\lambda)$ and $h_1(\lambda)$ are fixed. During the separation process, I_2 should be zero in all roots of $f(\lambda)$. This will unambiguously fix $f(\lambda)$ by (3).

The 4-pole so realized is shown in Figure 10. What remains to be done is to determine β_n^2 from $f(\lambda)f(-\lambda)$ (refer to (3)). Generally we write

$$\beta_n^2 = \Delta^2 + \delta^2. \tag{19}$$

The general characteristic equation is rewritten

$$k^2g(\lambda)g(-\lambda) = \Delta^2 f(\lambda)f(-\lambda) + h_1(\lambda)h_1(-\lambda) + j_1(\lambda)j_1(-\lambda) + \delta^2 f(\lambda)f(-\lambda)$$

where $j_1(\lambda)j_1(-\lambda) + \delta^2 f(\lambda)f(-\lambda)$ is the new $j_1^*(\lambda)j_1^*(-\lambda)$. After rearrangement in accordance with (13), we obtain

$$|\hat{S}_w|^2 = \frac{k^2g(\lambda)g(-\lambda)}{\Delta^2 f(\lambda)f(-\lambda)} = \frac{\beta_n^2}{\Delta^2} \times \left[1 + \frac{h_1(\lambda)h_1(-\lambda)}{f(\lambda)f(-\lambda)} + \frac{j_1(\lambda)j_1(-\lambda)}{f(\lambda)f(-\lambda)}\right]. \tag{20}$$

The actual loss is therefore

$$\log_e |\hat{S}_w| = \log_e |S(\lambda)| + \log_e \beta_n/\Delta. \tag{21}$$

The losses $\log_e|S(\lambda)|$ and $\log_e(\beta_n/\Delta)$ are linear; however, within $\log_e S(\lambda)$, losses caused both by the reactance elements and by resistors exist. Hence, D_l and D_r are not linearly dependent.

The following three cases differ by the magnitude of the power dissipated within the 4-pole (Figure 10).

$$\left.\begin{matrix}\text{loss in 4-pole} \\ \text{maximum power}\end{matrix}\right\} \begin{matrix}\leq \\ = \\ \geq\end{matrix} \Bigg\} |R_1(\lambda)|^{-2}$$

$$\text{for} \begin{cases} \Delta^2 > \beta_n^2, & \delta^2 < 0 \\ \Delta^2 = \beta_n^2, & \delta^2 = 0 \\ \Delta^2 < \beta_n^2, & \delta^2 > 0. \end{cases} \tag{22}$$

The designation $|R_1(\lambda)|^{-2}$ is the prescribed forward-dissipation function and is equal to

$$\frac{j_1(\lambda)j_1(-\lambda)}{k^2g(\lambda)g(-\lambda)}.$$

If (22) is valid for one point in the imaginary axis, then it is valid for all points in the imaginary axis. The existence of all three cases of (22) depends on the properties of the roots of $f(\lambda)$ and $h_1(\lambda)$ and on the network configuration. A detailed report [21] contains an accurate presentation thereof and the special cases

$$j_1(\lambda) = \kappa f(\lambda)$$
$$j_1(\lambda) = \kappa h_1(\lambda)$$
$$h_1(\lambda) = 0.$$

2.1.13

The following is a treatment of the linear interdependence of D_l and D_r. It is assumed that the reflection-coefficient tolerance $|T_1(\lambda)|^{-1} \leq \tau$, and that the transfer function $S(\lambda) = g(\lambda)/f(\lambda)$ up to a constant factor.

We restrict the roots of $f(\lambda)$ to the imaginary axis of the λ plane and start from the well-known characteristic equation for the reactance 4-pole.

$$g(\lambda)g(-\lambda) = f(\lambda)f(-\lambda) + h_1(\lambda)h_1(-\lambda). \tag{23}$$

As explained in Section 2.1.10, $\log_e S(\lambda) = D_l$. If we select a κ so that $1 + \kappa^2 = k^2$ and

$$\text{maximum } \frac{|h_1(\lambda)|}{k|g(\lambda)|} < \tau \tag{24}$$

for a pass-band range, then we obtain the following characteristic equation for the general (resistor-inductor-capacitor) 4-pole.

$$k^2|g(\lambda)|^2 = |f(\lambda)|^2 + |h_1(\lambda)|^2 + \kappa^2|g(\lambda)|^2.$$

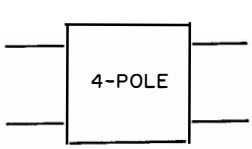


Figure 10—Four-pole.

Considering the actual subdivision of power due to the realizing procedure (Section 2.1.12), the actual transfer function, the echo transfer function, and the forward-dissipation function are designated by $\hat{S}_w(\lambda)$, $\hat{T}_{w1}(\lambda)$, and $\hat{R}_{w1}(\lambda)$, respectively, instead of by $S(\lambda)$, $T_1(\lambda)$, and $R_1(\lambda)$.

Generally

$$|\hat{S}_w(\lambda)|^{-2} = \frac{\Delta^2 \bar{f}(\lambda)\bar{f}(-\lambda)}{k^2 g(\lambda)g(-\lambda)} \qquad (25A)$$

$$|\hat{T}_{w1}(\lambda)|^{-2} = \frac{h_1(\lambda)h_1(-\lambda)}{k^2 g(\lambda)g(-\lambda)} \qquad (25B)$$

$$|\hat{R}_{w1}(\lambda)|^{-2} = \frac{\delta^2 \bar{f}(\lambda)\bar{f}(-\lambda)}{k^2 g(\lambda)g(-\lambda)} + \frac{\kappa^2}{k^2} \qquad (25C)$$

where $R_1(\lambda) = k/\kappa$ is the prescribed transfer loss (forward-dissipation) function for the 4-pole. Through some rearrangement in (23) we obtain the following equation corresponding to (21).

$$\log_e |\hat{S}_w(\lambda)| = \log_e |S(\lambda)| + \log_e k + \log_e (\beta_n/\Delta). \qquad (26)$$

Here $\log_e k + \log_e (\beta_n/\Delta)$ is the loss caused by the resistance elements of the 4-pole.

The relation between the dissipation within the 4-pole and the forward-dissipation function prescribed as well as Δ^2, δ^2, and the highest coefficient β_n^2 are exactly the same as in (22), all three cases being real.

The initial function for realizing the 4-pole is its input working driving-point function presented in (6).

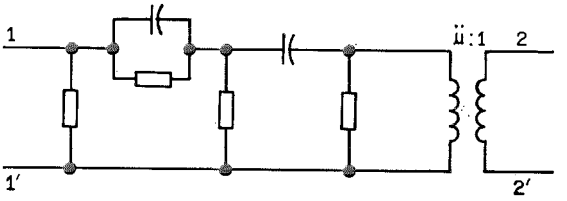


FIGURE 11—Guillemin's procedure for imaginary attenuation poles.

2.1.14

The $|T_1(\lambda)|^{-1} < \tau$ and the $S(\lambda) = g(\lambda)/f(\lambda)$ can also be realized by a network as follows: Writing γ_m of (9B) in the form

$$\gamma_m^2 = \gamma_m^{(1)2} + \gamma_m^{(2)2}$$

we obtain in the pass-band range

$$\text{maximum} \left| \frac{\gamma_m^{(1)} \bar{h}_1(\lambda)}{g(\lambda)} \right| < \tau.$$

This gives the general characteristic equation

$$|g(\lambda)|^2 = |f(\lambda)|^2 + \gamma_m^{(1)2}|\bar{h}_1(\lambda)|^2 + \gamma_m^{(2)2}|\bar{h}_1(\lambda)|^2$$

where $j_1(\lambda) = \gamma_m^{(2)} \bar{h}_1(\lambda)$.

Similar to (25) we obtain

$$|\hat{S}_w(\lambda)|^{-2} = \frac{\Delta^2 \bar{f}(\lambda)\bar{f}(-\lambda)}{g(\lambda)g(-\lambda)} \qquad (27A)$$

$$|\hat{T}_{w1}(\lambda)|^{-2} = \frac{\gamma_m^{(1)2}\bar{h}_1(\lambda)\bar{h}_1(-\lambda)}{g(\lambda)g(-\lambda)} \qquad (27B)$$

$$|\hat{R}_{w1}(\lambda)|^{-2} = \frac{\gamma_m^{(2)2}\bar{h}_1(\lambda)\bar{h}_1(-\lambda)}{g(\lambda)g(-\lambda)} + \frac{\delta^2 \bar{f}(\lambda)\bar{f}(-\lambda)}{g(\lambda)g(-\lambda)}. \qquad (27C)$$

The prescribed forward-dissipation function is represented as

$$R_1(\lambda) = \frac{g(\lambda)}{\gamma_m^{(2)}\bar{h}_1(\lambda)}.$$

The equation corresponding to (26) is

$$\log_e |\hat{S}_w(\lambda)| = \log_e |S(\lambda)| + \log_e (\beta_n/\Delta) \qquad (28)$$

where $\log_e (\beta_n/\Delta)$ is the loss introduced in the 4-pole by resistors only, while $\log_e S(\lambda)$ is the loss caused by reactance elements.

Equation (22) applies here again; but for $f(\lambda)$ and $h_1(\lambda)$ with roots located only on the imaginary axis of the λ plane, the last case $<$ is inapplicable.

The initial function as the input working driving-point impedance to realize a 4-pole is

$$W_{b1}(\lambda) = \frac{g(\lambda) + \gamma_m^{(1)}\bar{h}_1(\lambda)}{g(\lambda) - \gamma_m^{(1)}\bar{h}_1(\lambda)} = \frac{P(\lambda)}{Q(\lambda)}. \qquad (29)$$

2.1.15

Substituting

$$\beta_n^2 = \beta_n^{(1)2} + \beta_n^{(2)2}$$

we obtain

$$|g(\lambda)|^2 = \beta_n^{(1)2}|\bar{f}(\lambda)|^2 + |h_1(\lambda)|^2 + \beta_n^{(2)2}|\bar{f}(\lambda)|^2$$

where $j_1(\lambda) = \beta_n^{(2)}\bar{f}(\lambda)$.

The following equation is valid for the prescribed driving-point impedance in realizing a 4-pole with prescribed attenuation poles.

$$W_{b1}(\lambda) = \frac{g(\lambda) + h_1(\lambda)}{g(\lambda) - h_1(\lambda)}. \qquad (30)$$

Similar to (25) we obtain

$$|\hat{S}_w(\lambda)|^{-2} = \frac{\Delta^2\beta_n^{(1)2}\bar{f}(\lambda)\bar{f}(-\lambda)}{g(\lambda)g(-\lambda)} \qquad (31A)$$

$$|\hat{T}_{w1}(\lambda)|^{-2} = |T_1(\lambda)|^{-2} = \frac{h_1(\lambda)h_1(-\lambda)}{g(\lambda)g(-\lambda)} \qquad (31B)$$

$$|\hat{R}_{w1}(\lambda)|^{-2} = \frac{\beta_n^{(2)2}\bar{f}(\lambda)\bar{f}(-\lambda)}{g(\lambda)g(-\lambda)} + \frac{\delta^2\beta_n^{(1)2}\bar{f}(\lambda)\bar{f}(-\lambda)}{g(\lambda)g(-\lambda)}. \qquad (31C)$$

$$R_1(\lambda) = \frac{g(\lambda)}{\beta_n^{(2)}\bar{f}(\lambda)}$$

is the prescribed forward-dissipation function. The equation corresponding to (26) is

$$\log_e|\hat{S}_w(\lambda)| = \log_e|S(\lambda)| + \log_e\frac{\beta_n}{\beta_n^{(1)}} + \log_e\frac{\beta_n^{(1)}}{\Delta} \qquad (32)$$

where $\log_e(\beta_n/\Delta)$ is the loss caused only by resistors in the 4-pole and $\log_e S(\lambda)$ is the loss caused by reactance components. If the roots of $f(\lambda)$ are located only on the imaginary axis of the λ plane and (30) is realized by a ladder network, the assumptions made will readily be appreciated. Again, on condition that we substitute $\beta_n^{(1)}$ instead of β_n in (22), this equation is applicable to the first two cases.

2.2 Procedure

2.2.1

We now deal with a second-rank C function, defined as a special function of the Brune function, which is a positive real function that is realized only by a 2-pole with circuit elements consisting of resistances R and capacitances C.

If $W_{b1}(\lambda)$ is a second-rank C function, the processes to realize $W_{b1}(\lambda)$ are subdivided into the following four cases in accordance with the distribution of roots of the function $f(\lambda)$.

(**A**) For attenuation poles on the negative real axis of the λ plane, W_{b1} can be realized by a ladder network after the Guillemin process [11]. (See Figure 11.)

(**B**) For attenuation poles in the left λ plane, $W_{b1}(\lambda)$ can be realized also by the Guillemin

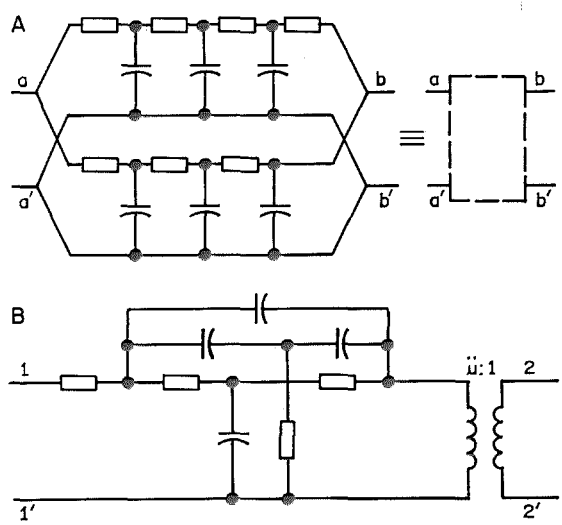


Figure 12—*A* is Guillemin's and Fiakow's modified procedure for sub-two-terminal-pair network *a-a'—b-b'* in *A* of Figure 13. The location of the attenuation pole is arbitrary except on the positive real axis of the λ plane.

B is Dasher's procedure for location of the attenuation pole in the left λ plane including the boundary (no real attenuation pole and only one pair of conjugate complex attenuation poles). This figure is realized, however, for input working driving-point immittance according to the principle in this paper.

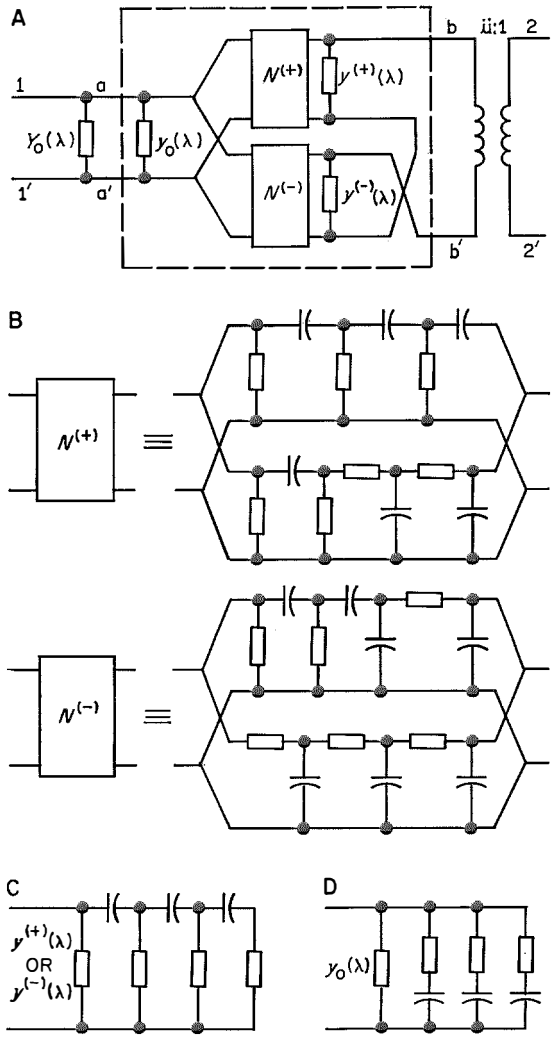


Figure 13—Author's procedure for arbitrary location of attenuation poles. For example, $f(\lambda) = \beta_3\lambda^3 - \beta_2\lambda^2 + \beta_1\lambda - \beta_0$ with $\beta_v > 0$ $(v = 0, 1, 2, 3)$, $g(\lambda) = \gamma_3\lambda^3 + \gamma_2\lambda^2 + \gamma_1\lambda + \gamma_0$ is a Routh-Hurwitz polynomial, but with zero on the negative real axis of the λ plane.

In *A*, either the admittance $y^{(+)}(\lambda)$ or $y^{(-)}(\lambda)$ or both of them are equal to zero. $Y_0(\lambda)$ and $y_0(\lambda)$ can be combined as one.

In *B*, some of the resistances may be equal to zero.

$Y_0(\lambda)$ is similar to either *C* or *D* in the figure. All capacitors are positive and consequently can be realized with shunt dissipation. However, under certain conditions the configuration can be simplified.

process [12], by the application of Section 2.2.2 (Figure 12A), or by the Dasher process [13] by a cascaded 4-pole section parallel-connected from *T* networks (Figure 12B).

(**C**) For attenuation poles except those on the positive real axis of the λ plane, $W_{b1}(\lambda)$ can be realized by Fiakow's modified Guillemin process [12, 14] by a network combining several ladder networks in parallel (Figure 12A).

(**D**) For attenuation poles at any point of the λ plane, $W_{b1}(\lambda)$ can be realized after the principle of Section 2.1.5 and the author's process [1] (Figure 13).

2.2.2

If just $P(\lambda)$ or $Q(\lambda)$ of $W_{b1}(\lambda)$ has only simple roots on the negative real axis of the λ plane or if both $P(\lambda)$ and $Q(\lambda)$ of $W_{b1}(\lambda)$ have only simple roots on the negative real axis of the λ plane, but $W_{b1}(\lambda)$ is not a second-rank *C* function, these cases are treated with only real attenuation poles as follows.

(**A**) In accordance with [5] consider the voltage transfer function

$$N(\lambda) = -\frac{P(\lambda)}{f(\lambda)} = a(\lambda) + b(\lambda). \quad (33)$$

What is wanted is a polynomial $K(\lambda)$ with only such negative real roots; we thus have

$$y_{22}(\lambda) = \frac{B(\lambda)}{K(\lambda)} = \frac{P(\lambda)}{K(\lambda)} - 1 \quad (34A)$$

as a second-rank *C* function and

$$y_{12}(\lambda) = \frac{f(\lambda)}{K(\lambda)}. \quad (34B)$$

In accordance with (**A**) of Section 2.2.1, realizing of $y_{22}(\lambda)$ is accomplished with a ladder network. When an ideal transformer of a suitable voltage-transfer ratio $\ddot{u}_0$:1 is connected to the input of this ladder network, $y_{22}(\lambda)$ and $y_{12}(\lambda)$ are realized.

Designating the input working driving-point admittance to the ideal transformer by $M_{b1}^{(1)}(\lambda)$, we can obtain, with the principle

of (7), a 2-pole in parallel to the 4-pole input, the Brune function or the M function of which is equal to $y_0(\lambda)$. As required, $y_0(\lambda)$ is realized by a 2-pole as described immediately following (7). The ideal transformers can be displaced to the right and finally merged with a normalized load resistance. (See sub-4-terminal network a–a′—2–2′ in Figure 7). For other attenuation poles, a similar treatment is valid in accordance with the proper paragraph of (**B**), (**C**), and (**D**) of Section 2.2.1 [21].

(**B**) If $Q(\lambda)$ has only negative real roots, we substitute the $M(\lambda)$, $Q(\lambda)$, $z_{22}(\lambda)$, $z_{12}(\lambda)$ $W_{b1}{}^{(1)}(\lambda)$, $\ddot{u}_1$, $z_0(\lambda)$, admittance, and impedance for the $N(\lambda)$, $P(\lambda)$, $y_{22}(\lambda)$, $y_{12}(\lambda)$ $M_{b1}{}^{(1)}(\lambda)$, $\ddot{u}_1{}^{-1}$, $y_0(\lambda)$, impedance, and admittance in the proper places of paragraph (**A**) immediately above, the remainder of the text need not be changed. The figure is an inverse network of that in Figure 7. However, it does not always exist.

(**C**) If both $P(\lambda)$ and $Q(\lambda)$ have only simple negative real roots, but $P(\lambda)/Q(\lambda)$ is not a second-rank C function, $W_{b1}(\lambda)$ can be reduced either to case (**A**) or case (**B**).

2.2.3

If $W_{b1}(\lambda)$ has at least one pair of conjugate complex roots in both numerator and denominator, we will confine ourselves to the case that $f(\lambda)$ has only imaginary roots. We want to realize the prescribed $W_{b1}(\lambda)$ and $f(\lambda)$ by a ladder network having a minimum number of reactance elements.

This process is based mainly on the author's work [7] in which an M function is realized by a ladder network with optional attenuation poles; here, however, it is realized by a ladder network with prescribed attenuation poles. Here we designate by x_1 the absolute minimum point of the real portion $U(x)$ of $W_{b1}(\lambda) = U(x) + j\Omega V(x)$, and by $x_{\infty 1}$ the attenuation poles.

If $x_1 = x_{\infty 1}$, the Brune process can be applied directly. If $x_1 \neq x_{\infty 1}$, it is necessary to find a new absolute minimum point $x_1{}^*$ so that $x_1{}^* = x_{\infty 1}$. Applying these principles, either a conductance g in parallel to the 4-pole input must be separated from $M_{b1}(\lambda)$, or a series input resistance r must be separated from $W_{b1}(\lambda)$. The residual function, say, $W_{b1}{}^{(1)}(\lambda) = W_{b1}(\lambda) - r$, the real portion of which has such new minimum $x_1{}^*$, can then be separated by the Brune process. Repeating this process, all attenuation poles are separated. For instance, observing (22) and suitably dividing the remaining constant number into two parts (one of them regarded as the load resistance), the prescribed properties are realized by a 4-pole.

The principles are based mainly on touching a circle K, as an M function of a linear (rational) function, with $W_{b1}(\lambda)$ in a W plane for $\lambda = j(x_{\infty 1})^{1/2}$ in the first order (refer to [6] page 365). After a parallel displacement of the imaginary axis of the W plane to the right by r units, if we make an inverse map of the W plane with new coordinates as well as of the curves therein, the circle K is converted into a circle $K^{(1)}$ (a straight line being a degenerate circle). Also, a transition of $W_{b1}(\lambda)$ to $M_{b1}{}^{(1)}(\lambda) = 1/[W_{b1}(\lambda) - r]$ is obtained.

$K^{(1)}$ and $M_{b1}{}^{(1)}(\lambda)$ contact only in the first order. For our purpose we will make sure that $M_{b1}{}^{(1)}(\lambda)$ is an M function. Applying the principle of Section 2.1.13, we will use the characteristic functions $g(\lambda)$, $h_1(\lambda)$, and $f(\lambda)$ for a reactance 4-pole as initial functions with a sufficiently small figure κ (refer to sentence preceding (24)). Then this process can always be carried out because the displacement of attenuation-pole locations is small.

The configuration of the network corresponds to Figure 14. Generally it is represented as in Figure 15 because of Norton's equivalent network. Here each coil is compensated for dissipation.

2.2.4

We will compensate here for the dissipation of the coil as well as that of the capacitor. If the

roots of $f(\lambda)$, when interpolated, are located on a line of the λ plane shifted by ϵ to the left from and parallel to the imaginary axis, we can treat this realizing problem in the most-rigorous sense. Otherwise we would carry out, in accordance with the well-known predistortion method, a transformation $\lambda = p - \epsilon$ for the variable of the prescribed function $g(\lambda)$. According to the principle and process described in [2–4], we would obtain a new polynomial $\tilde{h}_1^*(p)$ and consequently a reactance 4-pole. By the retransform $p = \lambda + \epsilon$, $\tilde{h}_1^*(p)$ is transformed into $\tilde{h}_1(\lambda)$ and we would obtain a 4-pole with compensation for dissipation, but a bad reflection coefficient; that is

$$M_1 = \eta M_2, \quad \text{for} \quad \eta > 1 \qquad (35A)$$

where in the pass band

$$M_1 = \text{maximum of } \left|\frac{\tilde{h}_1(\lambda)}{Mg(\lambda)}\right| \qquad (35B)$$

and

$$M_2 = \text{maximum of } \left|\frac{h_1(\lambda)}{g(\lambda)}\right|. \qquad (35C)$$

We designate the minimum of $|g^*(p)/f^*(p)|$ by $1/M$, that is

$$\frac{1}{M} = \text{minimum of } \left|\frac{g^*(p)}{f^*(p)}\right|. \qquad (36)$$

Now we treat

$$M^2 g^*(p)g^*(-p) = \tilde{h}_1^*(p)\tilde{h}_1^*(-p) + f^*(p)f^*(-p) \qquad (37)$$

with the process described in Section 2.2.3. κ has to be chosen after the equation

$$\kappa = \left(\frac{\eta^2}{\xi^2} - 1\right)^{1/2} \qquad (38)$$

with $\xi \leq 1$.

ξ has to be determined after

$$\xi \leq \frac{\tau}{M_2}. \qquad (39)$$

If $\xi < \tau/M_2$, the realized 4-pole has a reflection coefficient in the entire pass band less than the given tolerance τ.

If we now carry out a retransform $p = \lambda + \epsilon$ for this 4-pole, we obtain a series resistance for each coil and a parallel conductance for each capacitor (Figure 4). The negative resistances can likewise be eliminated by Norton's equivalent networks. Otherwise we replace the 4-pole section containing negative resistance by an equivalent 4-pole section without coupling, and then retransform this 4-pole section; or we apply a different process,

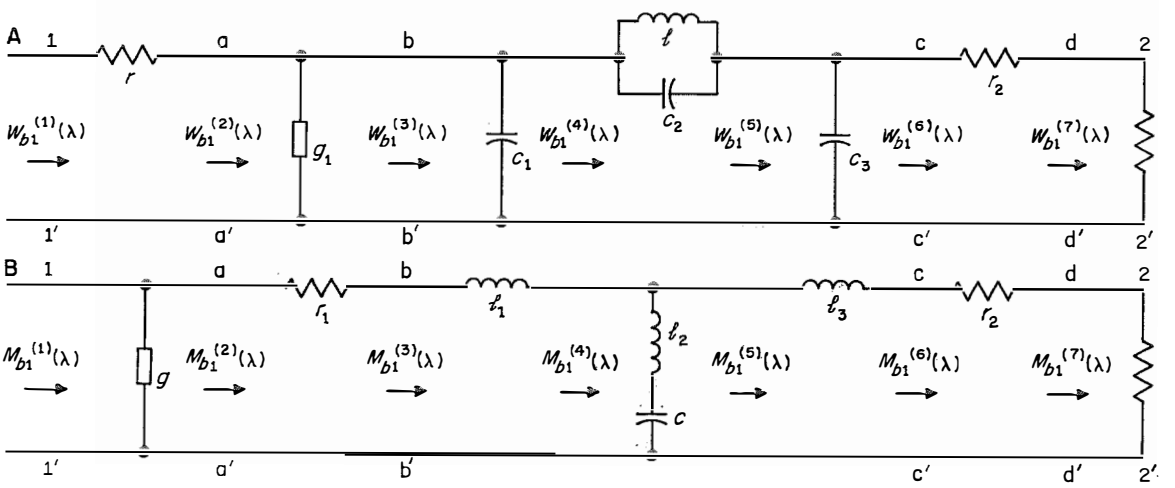


Figure 14—Separation of partial four-pole with prescribed real attenuation pole for input working driving-point impedance or admittance.

not covered in this paper, that is presumably applicable for a sufficiently small quantity of ϵ.

2.2.5

This process comprises, beside the one described in Section 2.2.3, the following process (Figure 7). In accordance with the basic idea, the absolute minimum point x_1 of the real portion $U(x)$ of an input working driving-point impedance is made to coincide with one of the attenuation poles by the partial removal of any general Brune function (in the detailed report both of degrees 0 through 2 and of the general case) of a two-terminal network as impedance function instead of partially removing the simplest Brune function in the form of a polynomial of degree 0 or 1. Apparent attenuation poles produced by this procedure (refer to Section 2.1.9) must be removed by using a 2-terminal partial network-pair. Hence the procedures referred to in Sections 2.2.1 and 2.2.2 are applied, depending on what kind of 2-pole is used for the partial removal.

This procedure calls for investment in more reactance elements, but it offers the following advantages.

(**A**) The perfect coupling in the Brune process for the 2-pole can always be avoided; as an example, Figure 8 shows l_0 by which the perfect coupling is eliminated.

(**B**) An equivalent circuit (for instance, of a crystal) can be a priori incorporated for a prescribed property.

(**C**) There is a wide margin in compensation for dissipation.

(**D**) $W_{b1}(\lambda)$ always can be realized by a ladder structure with compensation for dissipation of each coil, even if there are repeated real attenuation poles (for instance, $j\Omega_{\infty 1}$) or if there is the absolute minimum point x_1 of the real portion $U(x)$ of an input working driving-point impedance $W_{b1}(\lambda)$ of a higher order.

(**E**) Generally, only this procedure leads to the realization of a ladder network containing no coupling whatever, having real attenuation poles, and complying with the independently prescribed transfer function $S(\lambda) = kg(\lambda)/f(\lambda)$ and reflection coefficient $T_1^{-1}(\lambda)$ (or the tolerance thereof).

3. References

1. Nai-Ta Ming, "Theory of Realization of Dissipative Four-Pole Whose Working Transfer Characteristic is a Prescribed Function of Frequency," *Chinese Academic Press* (Peking); 1957. A summary of this article was published in *Hochfrequenztechnik und Elektroakustik*, volume 68, number 6, pages 190–193; January 1960.

2. Nai-Ta Ming, "Verwirklichung von linearen Vierpolschaltungen vorgeschriebener Frequenzabhängigkeit unter Berücksichtigung übereinstimmender Verluste aller Spulen und Kondensatoren", *Archiv für Elektrotechnik*, volume 39, number 7, pages 452–471; 1949.

3. S. Darlington, "Synthesis of Reactance 4-Poles That Produce Prescribed Insertion-Loss Characteristics," *Journal of Mathematics and Physics*, volume 18, pages 452–471; 1939.

4. W. Cauer, "Theorie der linearen Wechselstromschaltungen", volume 2, Akademie-Verlag, Berlin; 1960.

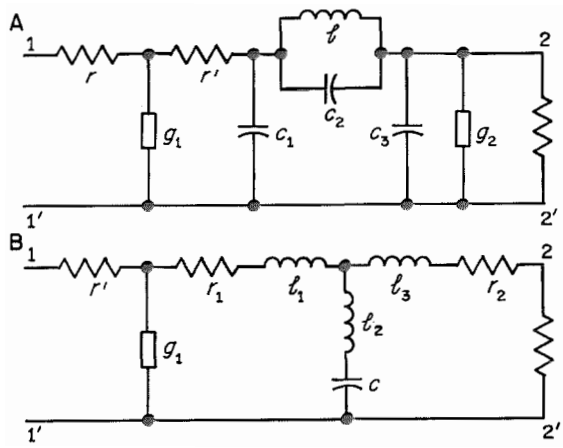


Figure 15—Equivalent four-pole for Figure 14. r' can become zero through Norton's equivalent circuit.

5. W. Cauer, "Theorie der linearen Wechselstromschaltungen", volume 1, first edition 1944: second edition 1954.

6. Nai-Ta Ming, "Verwirklichung von linearen Zweipolschaltungen vorgeschriebener Frequenzabhängigkeit unter Berücksichtigung der Verluste von Spulen und Kondensatoren", *Archiv für Elektrotechnik*, volume 39, number 6, pages 359–387; April 1949.

7. Nai-Ta Ming, "Existenzbeweis zur Realisierung einer Verlustfunktion in linearer Wechselstromschaltung durch Kettenschaltung von Verlustschaltelementen", Science Reports of National Tsing-Hua University, series A: volume 5, number 3, pages 350–377; 1949.

8. O. Brune, "Synthesis of a Finite Two-Terminal Network Whose Driving Point Impedance is a Prescribed Function of Frequency," *Journal of Mathematics and Physics*, volume 10, pages 191–235; 1931.

9. R. Bott and R. J. Duffin, "Impedance Synthesis without Use of Transformer," *Journal of Applied Physics*, volume 20, page 816; 1949.

10. P. I. Richard, "A Special Class of Function with Positive Real Part in a Half-Plane," *Duke Mathematical Journal*, volume 14, pages 777–786; 1947.

11. E. Guillemin, "A Note on the Ladder Development of RC-Networks," *Proceedings of the I.R.E.*, volume 40, number 4, pages 482–485; 1952.

12. E. Guillemin, "Synthesis of RC-Networks," *Journal of Mathematics and Physics*, volume 28, number 1, page 22; April 1949–1950.

13. B. J. Dasher, "Synthesis of RC Transfer Functions as Unbalanced Two-Terminal-Pair Networks," *Transactions of the I.R.E. Professional Group in Circuit Theory*, volume PGCT 1, pages 20–34; 1952.

14. A. D. Fiakow, "Two-Terminal-Pair Networks Containing Two Kinds of Elements Only," *Proceedings of the Symposium on Modern Network Synthesis*, page 63; April 1952.

15. B. V. Bylgakov, "Oscillation," Moscow Technical Literature Press; 1954.

16. J. G. Truxal, "Automatic Feedback Control System Synthesis," McGraw-Hill Book Company, New York; 1955: pages 103–220.

17. W. Bader, "Kopplungsfreie Kettenschaltungen", *Telegraphen-Fernsprech-Funk- und Fernseh-Technik*, volume 31, number 7, page 177; 1942.

18. H. Piloty, "Konjugierte Impedanzen und Weichenfilter", *Telegraphen-Fernsprech-Funk- und Fernseh-Technik*, volume 31, pages 255–265; 1942.

19. H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Company, New York; 1947: page 127.

20. W. Cauer, "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit", *Archive fur Elektrotechnik*, volume 17, pages 355–388; 1926–1927.

21. Nai-Ta Ming, a detailed report to Standard Elektrik Lorenz with the same title as this paper, dated 8 January 1964.

Nai-Ta Ming was born in Jükao, Kiangsu, China, on 8 June 1911. In 1936 he received the Bachelor of Science degree in electrical engineering from the Tsing-Hua University in Peking. In 1940 he graduated as a Diplom-Ingenieur from the Technische Hochschule in Berlin-Charlottenburg, Germany, and in 1944 he received the Dr.-Ing. degree from the same Institute.

From January 1941 to May 1945, he was an assistant to Professor W. Cauer in the Mix & Genest Laboratory, Berlin-Schöneberg, working also at the Mathematical Institute of the University in Göttingen for nearly 9 months.

From 1945 to 1948 he worked on the papers left by Professor Cauer.

From the summer of 1948 until March 1949, he was a guest research worker at the Massachusetts Institute of Technology and at the Harvard Radio Laboratory.

From 1949 to 1952 he was a professor at the Electrotechnical College of Tsing-Hua University and from 1952 to 1958 at the Mathematical Institute of Akademia Sinica Peking, where he founded and headed the Digital Computer Laboratory.

In May 1959 he joined Standard Elektrik Lorenz, working on highly accurate data transmission systems and general network synthesis.

Results of Multifrequency Signaling Experiments

G. A. W. RAHMIG
L. GASSER
Standard Elektrik Lorenz AG; Stuttgart, Germany

1. Introduction

The expansion of subscriber long-distance dialing systems goes hand in hand with attempts to minimize the time required to establish a connection. Better service and more-effective use of costly toll lines are the goals. A means to these ends is the use of multifrequency-code signaling instead of conventional pulse signaling.

In the multifrequency-code system investigated by the authors, the dialed digits as well as the control signals are in the form of a 2-out-of-6 frequency code. These 2-frequency signals are exchanged between registers. End-to-end signaling involves the originating and the intermediate intertoll registers up to the terminating register. In link-by-link signaling, the signals are exchanged between the registers associated with one particular section of the connection to be established.

There are basically two methods of using these signals and both are exemplified by the two signaling systems herein described. In the pulse system the receiving register, actuated by the seizure signal, initiates the transmission of dialed information from the originating register by means of voice-frequency signals having defined length. In the compelled system the 2-frequency signal corresponding to the first digit is transmitted from the originating register simultaneously with the seizure of the receiving register and is continuously applied to the latter. An acknowledging signal from the receiving register causes the forward first digit to be switched off. This cutoff causes the acknowledging signal to discontinue, which then causes the next digit to be transmitted in the forward direction [1, 2].

Both systems have been operated experimentally on toll lines of the German telephone network to determine the effects of transmission-channel properties on multifrequency-code signaling. Since the transmission properties are very similar in the various national networks, the results may be evaluated on a supranational level for Europe.

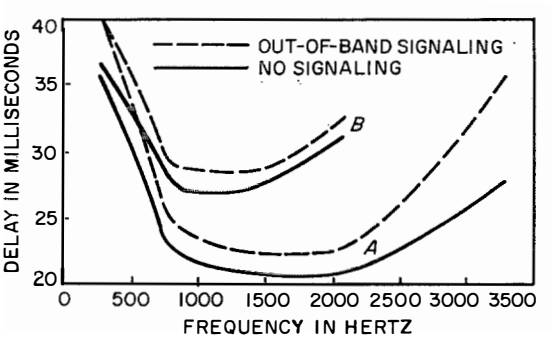


Figure 1—Delay as a function of frequency for two toll circuits. Curves *A* are for eight 300-kilometer (186-mile) carrier sections. Curves *B* are for six of these carrier sections plus two loaded voice-frequency cables, each 83.7 kilometers (52 miles) long.

2. Properties of Transmission Channels

2.1 Delay and Delay Distortion

Figure 1 shows delay as a function of frequency for two toll circuits, using data from the Deutsche Bundespost. For each curve the delay differences Δt between the delay at each frequency indicated and the point of minimum delay are given in Table 1.

TABLE 1
Delay Differences Δt in Milliseconds

Frequency in Hertz	Curves A		Curves B	
	No Signaling	Out-of-Band Signaling	No Signaling	Out-of-Band Signaling
660	6.5	7.8	3.5	4.4
780	3.1	3.4	1.2	1.2
1900	0.2	0.1	3.2	3.3
2100	0.4	0.3	4.2	4.2

Measurements obtained from a 120-channel carrier system corresponded well with data computed for this connection from the values in the table.

While the maximum delay distortion determines the minimum pulse duration in the pulse system, the signal duration in the compelled system adjusts itself to the over-all delay. Hence a compelled system is faster if the delay is small, while on long lines that have substantial delay, the pulse system is faster (neglecting the time required for the backward acknowledging signals).

2.2 Echo and Double Echo

As multifrequency-code end-to-end signaling is being planned in Europe, the delay resulting from long lines must be considered together with echo effects, which increase with the delay. Only the double echo affects signaling in the multifrequency-code systems.

Figure 2 is a sketch relating to double echo. It may be assumed for simplicity that the net-loss variations s_1, s_2 and the balance return losses a_{F1}, a_{F2} are equal in both directions (their totals amount to $2s$ and $2a_F$, respectively). If the multifrequency-code receiver has an operating range a_0 and a safety margin a_s, measured from the lower response limit of the receiver, the following equation applies to elimination of double-echo effects, considering only one circulation.

$$2a_F = a_{F1} + a_{F2} = s_1 + s_2 + a_0 + a_s$$

hence

$$a_F = s + \frac{a_0 + a_s}{2}.$$

For an estimate of the balance return and trans-hybrid losses to be expected, measurements in the Stuttgart area were obtained from hybrids that were connected to side circuits or phantoms of a cable loaded with 80/40 millihenries and having an attenuation of 4 decibels (two toll line transformers were included). In the range of frequencies used for multifrequency codes, Table 2 lists the values recorded for the terminations stated.

TABLE 2
Measured Losses

Trans-Hybrid Loss (Echo) in Decibels	Loop Attenuation (Double Echo) in Decibels	Termination in Ohms
13	29.5	no load
23.5	51.2	400
26.1	58.2	600
27.8	59.0	800

A comparison of the first two columns shows that the loop attenuation is greater than might be expected from an addition of trans-hybrid loss figures (both columns show minimum

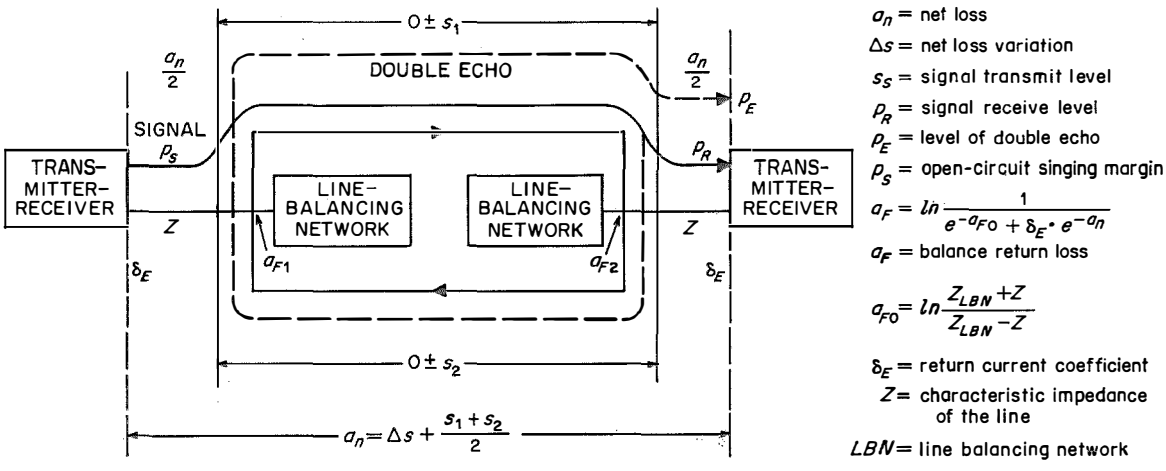


Figure 2—Echo and double echo.

figures). This can be explained by the low probability of the minima of trans-hybrid loss falling into the same frequency. The resulting decrease of double echo by 6 decibels may thus be regarded as a gain.

Considering these results in connection with the characteristic impedances of cables used in the German network, it can be shown that satisfactory protection against effects of double echo is provided when multifrequency-code transmitters and receivers have 600 ohms $\angle$ 0 $\pm$ 15 percent internal and input resistances.

2.3 Distortion Products

Of the two frequencies f_m and f_n making up a multifrequency-code signal, the third-order product of harmonics f_n-$2f_m$ are the most troublesome; this is indicated by a calculation based on the linear multifrequency code. The attenuation of this intermodulation product distortion in carrier systems depends on the magnitude of the signal level; this has to be kept in mind when the transmitting level is to be fixed.

Another fact to be considered is the dependence of harmonic attenuation on the net loss of the carrier sections. Measurements were verified with a calculation covering the net loss as well as the harmonic voltages of the first two sections of a transmission line. Since measurements and calculations were in good agreement, the characteristics were extrapolated to an over-all section net loss of zero decibels. The attenuation a_{d3} then obtained for various signal levels S_s is listed in Table 3.

TABLE 3
Harmonic Attenuation a_{d3} in Decibels

Number of Line Sections	For $S_s = -4.3$	For $S_s = -8.7$
2	46.8	56.0
4	40.7	50.0
6	37.2	46.4
8	34.7	44.0

These values appear periodically only as short-duration minima because of the frequency drift in the single carrier-system sections. The maximum attenuation is about 9 decibels higher, according to measurements. For an adequate signal-to-noise ratio, the transmitting level of an individual frequency should not substantially exceed -8 decibels referred to 1 milliwatt.

2.4 Line Interruptions

The number of line interruptions was rather small. At times, however, interruptions accumulated, primarily in daytime. Readings from fault counters were taken every 15 minutes between 0800 and 1700 hours. These showed that the connection operated without fault for hours while up to 50 interruptions might suddenly appear within 15 minutes.

Other investigators found that these interruptions were caused by dry joints producing resistance variations as a result of rack vibrations. Since dry joints can be largely eliminated by modern test procedures, printed circuits, wrapped connections, et cetera, and since the over-all number of interruptions was relatively low when referred to the total time of observation, there seemed to be no point in blaming interruptions for faults. Excluding the few accumulations, a total of 75 interruptions appeared during the transmission of 2.4 million digits via 6 line sections. These interruptions were observed during 6 consecutive days, mostly in daytime, and amount to not more than 2 interruptions per day per line section.

2.5 Switching Clicks and Noise

The multifrequency-code signaling system is designed for interregister operation. The registers with their transmitting and receiving units are connected to the line only during the dialing process. During this time, the line is not yet switched through to the subscriber end;

hence, the multifrequency-code equipment cannot be disturbed from that end. Therefore no receivers with speech guard circuits are required as is the case, for instance, in the 2-voice-frequency system of the Comité Consultatif International Télégraphique et Téléphonique for the transmission of line signals.

However, wetted circuits exist in relay repeaters; thus when line sections are through-connected or control units are switched on or off, electrical transients may generate clicks having very-high voltage peaks.

There is therefore the danger, for any dialing system using voice frequencies, that clicks might disturb the operation in the toll-line network. This disturbance would become apparent in undesirable operation of one or more signal receivers by frequencies contained in the click spectrum. To prevent this disturbance, measures were tried in switching plants by which clicks were prevented from appearing on the line together with a voice-frequency signal.

Methods used for suppressing clicks also reduce the speed of signaling. Furthermore, full suppression is impossible in end-to-end signaling and other measures must be taken.

To find the best way to suppress clicks, laboratory experiments were performed on all types of wetted contacts of relays employed in the German toll network. The clicks generated were analyzed for frequencies and amplitudes. It was found that the click spectrum resembles an irregular line spectrum, showing a number of maxima and minima. The voltage peaks may be 10 to 100 times the response level of signal receivers. These peaks were caused by overloading of transformers in repeaters and signal receivers. When two repeaters were interconnected, a damped oscillation of about 220 hertz was generated, the frequency being determined by the decoupling transformers, hybrids, and 1-microfarad decoupling capacitors. This oscillation has an initial amplitude so great that transformers and hybrids are overloaded and generate odd harmonics. Of course, this was of no consequence in conventional dialing systems.

If the duration of noise pulses is defined as the time in which the noise voltage exceeds the signal-receiver response voltage, then durations of 15 to 30 milliseconds can be observed.

The above experimental data were checked with other data obtained from carrier channels between Stuttgart and Heilbronn involving 4-wire carrier-system repeaters. As expected, all signal receivers responded to the noise when the signal receiver was connected to the line and when the repeaters were switched through. The measured click voltages confirmed the values measured in the laboratory without transmission lines.

Another source of trouble is selector noise, which can appear at any instant and disturb a transmitted signal. This noise can partly or fully suppress a dialed digit and it occurs particularly often in old-type sliding-contact switching systems. The greater part of this noise is pulse shaped and is caused by vibration of the racks in which the operating selectors are mounted. This vibration produces a microphonic effect across contacts of selector wiper arms already in position. Selector-noise pulses typically form coherent pulse trains that are sometimes separated by short intervals. Each train may last up to 50 milliseconds. The level of selector-noise pulses may rise to −17 decibels referred to 1 milliwatt, while the pulse frequencies show an energy maximum near 1000 hertz. A multifrequency-code signaling equipment that uses transmission lines and is disturbed by such noise must be protected by suitable measures as discussed in Section 4.2.3.

2.6 Compandors

Compandors tend to round the edges of square-wave pulses. To investigate their possible effects on multifrequency-code signals, the latter were

therefore transmitted via lines of the German toll network equipped with compandors. The tube-type compandors have the following ratings: compression ratio ≈ 2; rise time, 2 to 3 milliseconds; fall time, 8 to 10 milliseconds [3, 4].

In the pulse system, no effect was observed on the shaped pulses from lines with compandors. In the compelled system, however, the effects on square pulses were not negligible. The following effects were studied in detail.

(A) The delay in response of the receivers, caused by flattened pulses on lines with compandors.

(B) The delay in response of several compandor-equipped lines in tandem.

(C) The delayed release of signal receivers.

The response delay was 1.5, 2.5, and 3.0 milliseconds, respectively, for 1, 2, and 3 lines using compandors. These results show only a small increase in delay with more than 3 compandor-equipped sections in tandem; the delay in the German network due to compandors may be assumed to be 4 milliseconds maximum. No delay was observed in the release of signal receivers by compandors.

3. Equipment Used for Field Trials

After investigating the properties of available transmission channels, the next step was to operate the laboratory-type multifrequency-code signal transmitters and receivers of both pulse and compelled systems using toll lines. This was done in close cooperation with the Departments of Development and Service of the Deutsche Bundespost, starting in the Stuttgart regional center.

A special test setup was developed for this purpose. The experience gained from the first experiments, which started in 1959, was used to improve the test setup in 1961 for further experimental runs in 1962.

The test setup was accommodated in a rack at the Stuttgart regional center. It permitted the multifrequency-code signaling units to be connected to 2-wire or 4-wire loops, and permitted 4-digit numbers to be transmitted in a 2-out-of-6 frequency code. The received numbers were compared with those transmitted, and the following types of faults were counted: missing frequency, excess frequency, missing digit, excess digit, and wrong digit. A 3000-hertz pilot tone was also used to measure line interruptions exceeding 25 milliseconds.

Either numbers of all possible digit combinations or a fixed number were transmitted automatically. An interruption adjustable in duration could be gated into the line at any time. The equipment was used with a 2-wire system, which required 2 groups of 6 frequencies, 1 each for the forward and backward directions. The groups were spaced by 240 hertz, the frequencies within each group by 120 hertz. The lowest frequency was 540 hertz.

The receiver sensitivity was set with respect to the maximum individual received level. The pulse senders supplied flattened pulses in the first experiments, which involved the pulse system alone, to avoid trouble from clicks caused by switching on the senders. Later, when the compelled system was tested, the pulse senders were adjusted to transmit square-wave signals, which are essential to this system.

The aforementioned faults were metered by counters from which readings were taken daily. During the trial operations, however, counter readings were taken every 15 minutes as this helped to trace faults to the causes producing them.

Both 2-wire voice-frequency and carrier channels and 4-wire carrier channels were employed, as these modes of operation are encountered in direct distance dialing.

One of the 4-wire channels included compandors. During the trial operation, 5 million 4-digit numbers were transmitted through

these channels. The results are discussed separately in Section 4 for each of the two multifrequency-code systems tested.

4. Results

4.1 Pulse System

Since the pulse system was intended to operate without signal conversion in end-to-end signaling, the signal receiver should handle a range of levels of 22.5 decibels, corresponding to the maximum permissible attenuation between two terminal offices in the German network plus a safety margin. If this range is exceeded in international traffic, signal conversion should be provided in border offices. The band from 600 to 2100 hertz is most suitable for the signal frequencies, considering the delay and loss distortion on a long line comprising 6 repeater sections and loaded cables as 2-wire feeders. Linearly spaced signal frequencies were chosen to avoid interference by the second harmonics. After proper consideration of delay distortion, channel transients, and relay response time, a pulse duration of 55 ±10 milliseconds and an interval of 40 ±10 milliseconds were chosen, both being adjustable. Shaped keying was used to concentrate the energy in the signal frequencies.

Direct keying was employed in the trial operation. The response thresholds of all 6 receivers of a group were commonly controlled and adjusted to a level 6 decibels below the maximum level of the received signals. The trial operation verified the ratings and provided information on the effects of reflections, line interruptions, and fluctuations of transmission properties.

4.1.1 *Echo and Double Echo*

Since two signal-frequency bands differing for forward and backward signals were used, the system was immune to echo effects. The bands being 240 hertz apart, the filter attenuation of the receivers can easily be made so high that an echo cannot actuate a receiver.

No effect of double echo on multifrequency-code signaling could be observed in normal operation. As this was ascribed to the higher

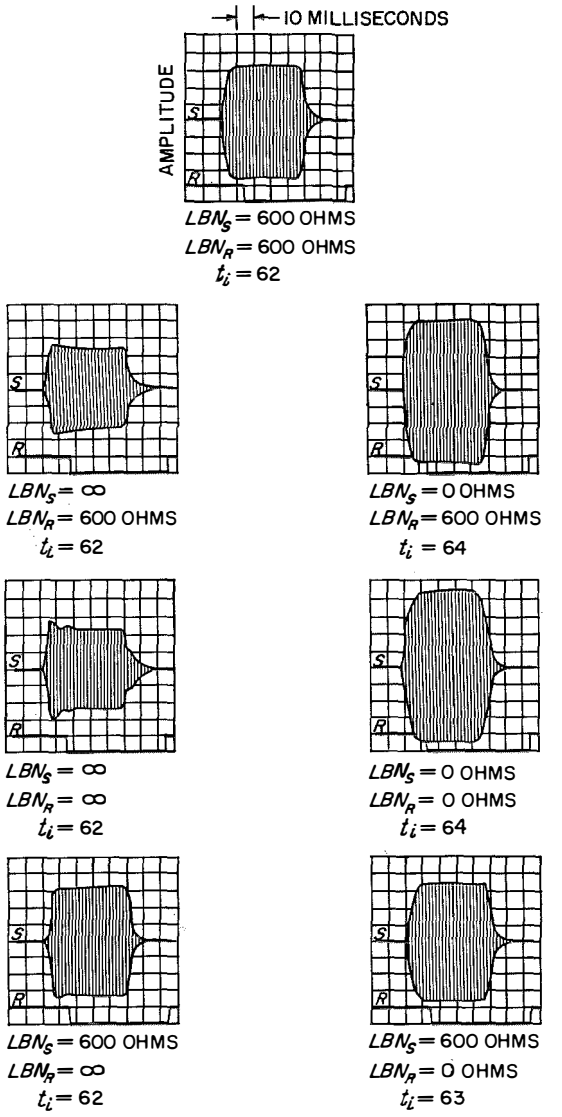


Figure 3—Effect of reflection at 1860 hertz. LBN_S and LBN_R = line balancing networks at the sending and receiving ends. t_i = pulse duration in milliseconds.

net loss of the repeater sections, artificial reflections were generated by changing the line-balancing networks in the signal transmitting and receiving equipments. The resulting distortion of the received multifrequency-code pulse is shown in Figure 3. The line-balancing networks *LBN* in the figure were adjusted to zero, 600 ohms, and infinite resistance for different tests, in which the pulse duration t_i was found by measuring the time that the receiver pulse contacts remained closed.

It was found that because of the response threshold, no trouble was caused by the double echo even when its magnitude exceeded the calculated value. The small increase in t_i associated with short-circuited line-balancing networks may be explained by the increase in threshold level.

These results were the same regardless of the type of channel employed. Even when the loss-fluctuation range specified by the Deutsche Bundespost is fully used, no trouble should be expected from double echo if the input impedance of signal transmitters and receivers is kept within ±15 percent of 600 ohms. The protection against interference by reflections is increased if the receivers have a response threshold automatically controlled by the receiving level of the maximum tone, as in these measurements.

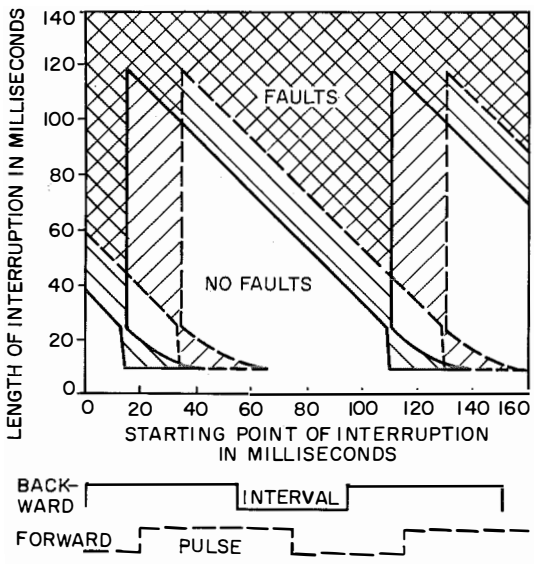


Figure 4—Interruption at outgoing exchange in pulse system.

4.1.2 *Line Interruptions*

To establish the effect of line interruptions on the pulse system, the maximum permissible duration of an interruption was plotted against the starting point of an interruption in Figure 4, referring this starting point to a pulse with a duty cycle of 55 milliseconds on and 40 milliseconds off. It was assumed that the receiver operate and release times were each 16 milliseconds and that the register operating time was 4 milliseconds.

The illustration shows the fault-limiting lines, solid for the backward signal and dashed for the forward signal. Both are identical in shape and shifted only by the time between the forward and backward signals. As may be seen from this diagram, the first part of the signal is lost if the interruption starts before the receiver has responded. Otherwise the interruption tends to shorten the pulse, but it is tolerable if the pulse is still long enough for proper processing. On the other hand, once the receiver has responded, a later interruption must not last so long (9 milliseconds in the available equipment) that one pulse is received as two.

Again, if the interruption after receiver operation is long enough to suppress the rest of the pulse or to leave a portion that cannot be evaluated, it constitutes no interference; on the contrary, it may even last throughout the following interval and a portion of another pulse so long as the remainder of this other pulse can be recognized. The same is true of an interruption starting during the unused remainder of the pulse or during the interval.

This statement applies generally even though it was derived from a case using a special duty cycle and negligible delay distortion. Any other duty cycle merely shifts the limiting lines in Figure 4 without changing their shapes, while

delay distortion between the two signal frequencies has the effect of shortening the pulse since a digit is received only when both signal frequencies are present. Observe that an interruption at another point of the transmission line will cause a proportional shift, but no structural change, in Figure 4.

Theoretical considerations were checked with multifrequency-code equipment connected to a transmission line. The result fully confirmed expectations except that stray faults occurred in a portion of the area indicated to be free of faults in Figure 4; however, this was traced to receiver malfunction caused by the square-wave keying of the pulse after the interruption. Square-wave keying generates a wide frequency spectrum that may cause operation of receivers tuned to other than the signal frequencies if these receivers are not protected by a response delay. This protection was not provided in the test setup.

4.1.3 *Clicks and Noise*

The first experiments were made on permanent connections of the toll network, which involved no selectors and eliminated clicks caused by switching and variations in contact resistances. No effect of crosstalk was observed.

If clicks that persist for 20 to 30 milliseconds are expected, pulses must be lengthened for reliable recognition of a signal. This problem becomes more complex if the channels are disturbed by selector noise, for a train of selector noise pulses might suppress a digit. It has been confirmed by trial that the pulse system is unsatisfactory for operation under such conditions of noise.

4.1.4 *Signaling Speed*

The pulse duration of 55 ±10 milliseconds and interval of 40 ±10 milliseconds that were provided for multifrequency-code signaling and for a satisfactory safety margin were shortened in several steps during the trial. For each step the faults were counted separately. The limiting values were obtained for all types of channel and all attenuation figures to be expected; they were clearly identified by the sudden increase of faults. The limits for a 2-wire line were: pulse duration of 4 milliseconds with an interval of 76 milliseconds, and pulse duration of 56 milliseconds with an interval of 24 milliseconds. This line comprised 6 repeater sections and had, at 800 hertz, a net loss adjusted to 14 decibels and a maximum delay distortion of 4.5 milliseconds. Even when this line was extended through loaded 2-wire feeders, the delay distortion did not increase. Contrary to conditions in carrier channels, the delay is associated with the upper limit of the signal band (see Figure 1). Table 4 lists the failures counted in the transmission of 100 000 four-digit numbers. It may be seen that few faults are caused when the pulse duration is between 6 and 50 milliseconds; almost all can be traced to line interruptions. For the shorter pulse durations listed, however, the faults exceeded interruptions. Only those interruptions exceeding 25 milliseconds were recorded; since short pulses could be disturbed by shorter interruptions, we may assume that the remaining faults were at least partly caused by interruptions shorter than 25 milliseconds.

TABLE 4

TEST RESULTS FOR 100 000 FOUR-DIGIT NUMBERS

Pulse Duration in Milliseconds	Interval in Milliseconds	Faults	Interruptions
4	76	>5000	1
6	74	48	42
10	70	2	0
20	60	6	21
30	50	16	79
40	40	3	4
50	30	0	4
56	24	2085	31

In the transmission line described previously, the pads of outgoing repeater circuits were substituted for by calibration that adjusted the net loss of the line. Thus it was found that the minimum permissible interval remains about

the same for all values of net loss, while the minimum permissible pulse duration is in the range of 5 to 30 milliseconds. Hence the practical minimum durations of pulse and interval are 10 and 26 milliseconds, respectively. These are obtained for the maximum attenuation of 19 decibels between two end offices, the highest value used in the German toll network.

No difference was found in the minimum pulse duration on a 4-wire line with and without compandors. This is explained by the absence of delay in the compandor that follows the rounded edges of the signal pulse.

The tests showed that we could reduce the pulse duration from 55 ±10 to 30 ±10 milliseconds and still retain a safety margin, even under extreme line conditions, provided the signaling is undisturbed. The interval of 40 ±10 milliseconds has proved satisfactory but depends on the receiver design.

To measure the time required by multifrequency-code equipments to transmit one digit, the sending and receiving equipments were directly connected to each other. In both directions, the equipment time t_e was 15 milliseconds. This time increases with distance and line attenuation and becomes 19 milliseconds at an attenuation of 10.5 decibels. The delay introduced by 6 repeater sections covering 1200 kilometers (750 miles) was $t_d = 11$ milliseconds, according to measurements. These data permit us to compute the total transmission time t_t for, say, three digits if t_t is defined as the time between the onset of the first backward pulse and the recognition of the receipt of the third digit.

$t_t = 178$ milliseconds for 1 section for which $a_n = 7$ decibels, $t_e = 17$ milliseconds, and $t_d = 2$ milliseconds.

$t_t = 200$ milliseconds for 6 sections for which $a_n = 10.5$ decibels, $t_e = 19$ milliseconds, and $t_d = 11$ milliseconds.

This corresponds to a signaling speed of 14 to 15 digits per second (assuming undisturbed signaling).

4.2 Compelled System

The range of levels to be handled by a multifrequency-code receiver was originally fixed at 22.5 decibels, and this range was used in tests with the pulse system. In accordance with the recommendations of the Comité Consultatif International Télégraphique et Téléphonique [5], the receivers in the compelled system were adjusted to an expanded range for European international lines and for a sending level, for each of the two frequencies of a signal, of −8 decibels referred to 1 milliwatt at relative level zero with a tolerance of ±2 decibels.

The range of levels for multifrequency-code receivers should include requirements for end-to-end signaling over international circuits. Two such ranges are required.

(A) −6 to −35 decibels referred to 1 milliwatt at relative level zero for a long-distance call covering 9 sections from terminal-to-terminal offices.

(B) −6 to −26 decibels referred to 1 milliwatt at relative level zero for a 4-wire connection.

4.2.1 *Echo and Double Echo*

As the same frequency scheme has been chosen for both signaling systems, the statements regarding echo for the pulse system also apply for the compelled system.

For double echo, the relations given in Section 2.2 show that an operating range $a_0 = 29$ decibels, a safety margin $a_s = 5$ decibels, and a maximum net-loss variation of $s_1 = s_2 = 2$ decibels per carrier-line section (or 18 decibels for 9 sections) would require a balance return loss $a_R = (6 + 17)$ decibels $= 23$ decibels, where the 6 decibels are the product of the net-loss variations and of the square root of the number of sections.

These 23 decibels constitute a relatively high value. In practice, the relations are more favorable for the following reasons.

(A) Because of the level regulation by pilot frequencies that is generally introduced, a lower net-loss variation may be taken into consideration, say, 1 decibel times $n^{1/2}$, where n is the number of sections.

(B) The signal receivers have response thresholds automatically adjusted by the signal level to a response level 6 decibels less than the signal level. This would reduce the permissible signal-to-noise ratio $a_0 + a_s$ of the double echo to 10 decibels.

The result would be a decrease of a_F from 23 to 8 decibels, a value that can still be improved by matching the signal transceivers to the line as indicated in Section 2.2.

The signal receivers of the compelled system had automatic threshold regulation. Receivers and transmitters had terminal impedances of about 600 ohms. Tests with the compelled-system receiver showed no adverse effect from double echo, regardless of the operating channel. (It should be noted that double echo could lengthen the signal transmission time in the compelled system by a maximum of twice the delay time per digit, but would cause no interference.)

4.2.2 *Line Interruptions*

The effect of line interruptions on the compelled system was observed in the same way as described in Section 4.1.2. A delay of 17.5 milliseconds was assumed for the transmission path as this results in the same pulse duration of 55 milliseconds used in the pulse system. The result shown in Figure 5 can thus be directly compared with Figure 4. It is clear that the range of faults in the compelled system is much smaller. This is so because an interruption that does not divide a pulse into two pulses may have any duration and will cause no interference. Only the transmission speed will decrease.

The range of faults will be enlarged and shifted in time if delay and delay distortion differ from those assumed for Figure 5. Thus, a delay of 30 milliseconds will result in fault areas 60 milliseconds high and 85 milliseconds across (dotted sides of triangle in Figure 5). In other words, as verified by tests, even a long delay does not result in the fault probability of the pulse system. Furthermore, the faults in Figure 5 can be overcome by simply transmitting two identical digits in tandem, so that the first digit is followed by a special repetition signal permitting the end of a digit to be recognized. The compelled system can thus be protected against any type of interruption; applied to the pulse system, this remedy would merely eliminate that part of the fault area in Figure 4 caused by dividing a pulse into two.

4.2.3 *Clicks and Noise*

Clicks caused by the switching plant greatly disturb the operation of the compelled system. Disregarding selector noise, we will first assume that the compelled system operates over lines connected by modern low-noise switching plant (crossbar switches, precious-metal-contact rotary selectors, et cetera).

The simplest protection against clicks is to delay the response time of the signal receivers. There is no technical obstacle against placing

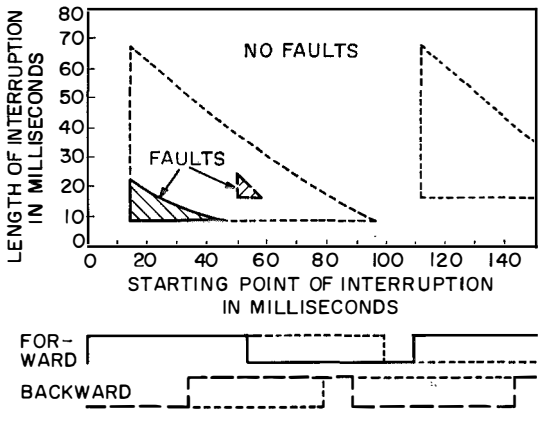


Figure 5—Interruption at outgoing exchange in compelled system. The solid (forward) and dashed (backward) lines each involve delay of 17.5 milliseconds, while the dotted lines are for 30-millisecond delays.

the clicks toward the front end of a digit or between two digits. The required delay is 30 to 35 milliseconds, however, as confirmed by experiments, and this slows down the over-all system speed. This delay comprises the click duration (Section 4.1.3) as well as a safety margin for the receiver response. A better solution would be a receiver that provides additional protection by responding to a frequency outside the signal band. This click receiver will block the signal receivers during a click, but will permit the system to operate at maximum speed in the absence of clicks.

Another improvement consists of suppressing the clicks as much as possible. Since clicks partially result from overloaded receiver input transformers, their amplitude can be reduced by a factor of 10 if the receiver has an input resistance of about 600 ohms and is linear over the whole frequency range. (This feature was absent in the pulse-system receiver.) However, the click amplitude would still be 2 to 5 times the response voltage of the receiver. Nevertheless, the duration of the clicks is reduced because their voltage maxima are attenuated; as a result, the delay introduced can be made shorter and this also increases the system speed.

Repeater circuits could be modified to prevent the charging of capacitors and thus avoid transients.

Each of these solutions is practicable and a trial operation was run to test which was best. In the first stage of the investigation, the method that requires a delay of 30 to 35 milliseconds was considered incompatible with a high-speed signaling system. We also assumed that it would be very hard to introduce modifications to the existing wetting contacts of relays. A click receiver was developed for the trial operation.

The choice of frequency band for the click receiver is important. Analysis showed that the greatest click energy in the German toll network is in the band from 200 to 450 hertz. However, a click receiver operating in this band fails if clicks cover long distances before they reach the receiver. Frequencies from 200 to 450 hertz are delayed longer than frequencies from 1000 to 2000 hertz (Figure 1). Since the signal receivers respond within 12 to 15 milliseconds, they respond to click frequencies in the signal band before the lower click frequencies can actuate the click receiver.

An attempt to place the click-receiver band into the 240-hertz gap between the two signal-frequency groups would encounter two obstacles. Amplitudes of the click spectrum have minima and maxima within the range of various signal frequencies. Hence, a click maximum might reach the signal receiver at the same time that a click minimum reaches the click receiver, which remains inoperative. Moreover, the click receiver would have a relatively narrow operating band and this would delay its response, whereas it should operate faster than the signal receivers to prevent clicks from actuating them.

A favorable solution uses the lower signal-frequency group from 540 to 1140 hertz for the forward direction and the upper group from 1380 to 1980 hertz for the backward direction. A click receiver for frequencies from 300 to 1000 hertz is parallel connected to the upper-group receivers in the outgoing register unit. This wide-band click receiver is very fast and responds to clicks that have passed through 10 or even more carrier-system sections, since the delay distortion is low in this frequency band. A click receiver for 300 to 450 hertz is parallel connected to the lower-group signal receivers. Switching clicks are generated only in tandem offices, hence they cannot cover more than 1 line section in the forward direction and arrive after moderate delay at the receivers of tandem offices. Here again the click receiver promises fast and reliable operation.

An arrangement of this type was used for trial runs on 4-wire lines of the German toll network with satisfactory results. The system operates with sufficient speed. If a click occurs

the signal receivers are blocked before the click reaches them, and unblock after the click decays. This solution is useful only for 4-wire lines because the click receiver associated with the forward direction must operate in the signal band of the acknowledging signals.

The described protection is insufficient if the transmission lines are additionally disturbed by selector noise. This was revealed in the trial operation. If a train of selector-noise pulses reaches a receiving unit simultaneously with a signal pulse, the latter could be suppressed. The operation should be able to wait until the selector noise has decayed, otherwise the digit cannot be identified and acknowledged.

The receivers for lines disturbed by selector noise were arranged on the following principles. A code-checking circuit verifies all receiver outputs for frequencies of genuine digits in the 2-out-of-6 code. Two frequencies being verified, a time circuit is actuated that rejects any digit of shorter duration than, say, 15 milliseconds. Once a digit is accepted, the two associated relays are locked and the memory does not accept new information. Release of the signal receivers and the memory to accept the next digit depends on two conditions: The register should have completed the preceding evaluation, and all 6 signal receivers should be free to accept the next digit (no noise may exist on the line). During this inactive period the arrangement is switched back to normal.

If the receivers are connected to lines disturbed by selector noise, the number of faulty digits depends on the frequency of the selector-noise pulse trains, the noise level, and the receiver-protection time chosen.

If the protective delay time of a receiver is 15 milliseconds and the signaling speed is 10 digits per second, 1 out of 7000 digits is disturbed on the average, as was found in the trial operation. This value changes in proportion to the signaling speed, that is, to the delay. If the protective time is increased to 20 milliseconds, no faults are detected.

For the following reasons, the compelled system can operate on lines heavily troubled by selector noise, whereas the pulse system becomes unsatisfactory. (This assumes the duty cycle of 55 and 40 milliseconds.)

In the pulse system, it is not possible to provide pulse durations capable of outlasting selector-noise pulse trains. The compelled system simply waits automatically until the disturbance is over; a fault can be generated only if the noise causes two receivers to remain in the responsive state longer than the protective time. This is highly improbable if the protective response delay is set correctly, as has been confirmed by experiment.

To provide similar reliability in the pulse system, the pulse duration and interval would have to be more than doubled, while the operating speed would be cut in half. In the compelled system a signal is lengthened automatically only when it is concealed by noise before it can be detected. It should also be noted that longer pulses in the pulse system would increase the probability of a noise pulse train appearing simultaneously with a signal pulse.

4.2.4 *Signaling Speed*

Signals in the compelled system may differ greatly in duration. The system automatically adapts to the line used in a connection. Considering the receiver response time of 15 milliseconds, the channel delay of 0 to 30 milliseconds, and the additional delay that increases with attenuation, the digit duration may vary between 30 and 125 milliseconds. The transmission time for a digit signal plus interval must then be 60 to 250 milliseconds, and the compelled system can thus transmit 4 to 15 digits per second.

These extreme values cover the range from zero loss to the receiver response limit. The latter is reached only when the line properties are very poor; on the average the transmission

speed of the compelled system is 8 to 10 digits per second. Actually these values were confirmed in the trial operation whenever the line was free of selector noise and when the click receiver was used. Selector noise and the delay of the time circuits in the described arrangement naturally reduce this speed. An estimate based on measurements shows that in this

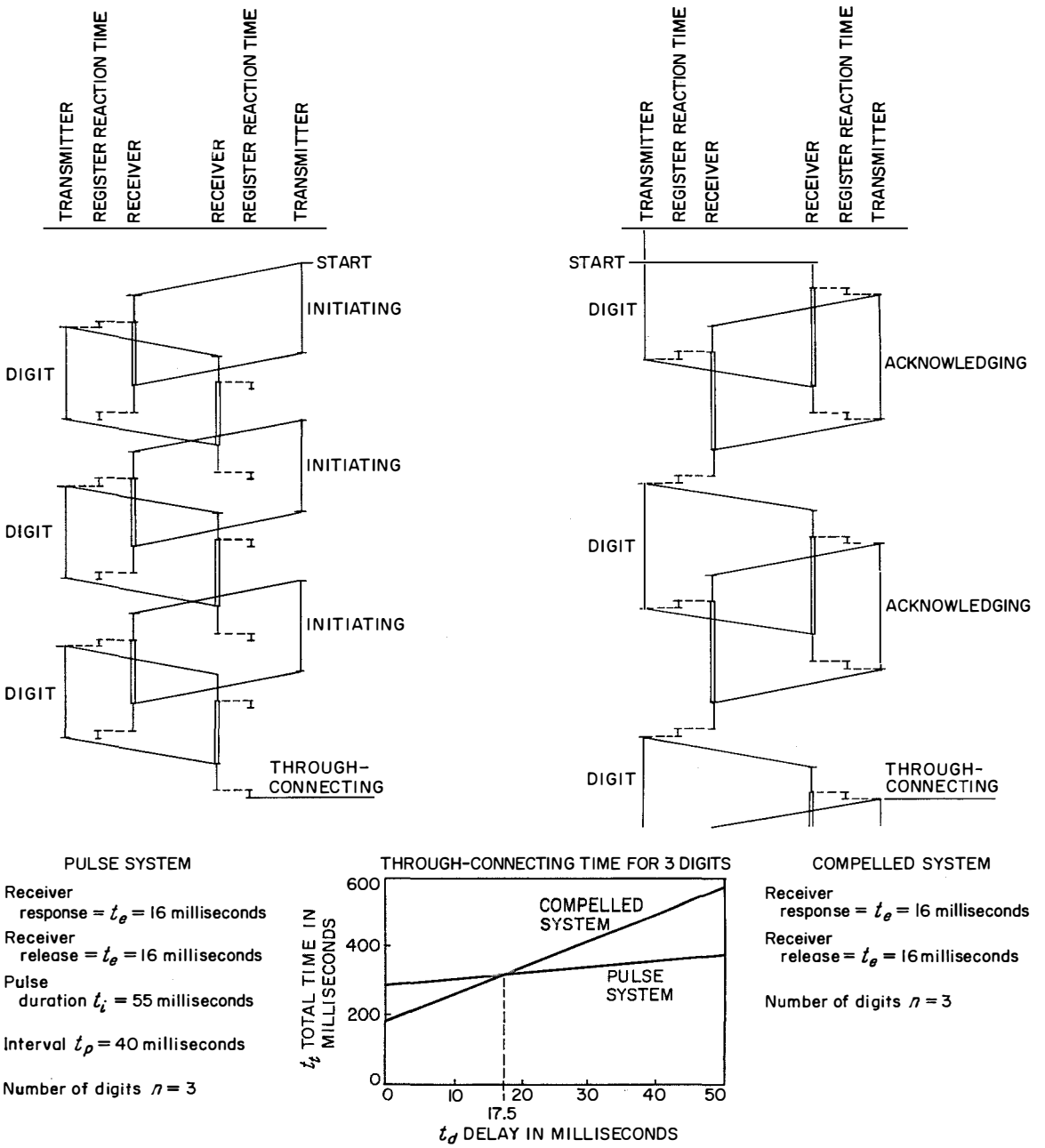


Figure 6—Comparison of pulse and compelled systems.

most-unfavorable case the transmission speed is 3 to 5 digits per second.

5. Conclusions

Trial operations with two multifrequency-code signaling systems have shown that these systems differ both in speed and in reliability. These results stress some characteristic features that bear on the application of each system.

Figure 6 presents a comparison of the two systems. The pulse system is faster, particularly on long-distance lines, because the pulse duration and, hence, the speed depend only on the delay distortion. On short lines this can be a disadvantage: The pulse duration, chosen in consideration of the delay distortion on long lines, cannot be adapted to the conditions of any other transmission line.

If channels are disturbed by selector noise, the pulse system is unsuitable because complete digits may be suppressed.

The speed of the compelled system is automatically adapted to the channel quality. With attachments, the system reliably transmits and receives digits even via channels affected by substantial noise, the only effect produced by such channels being reduced signaling speed. The automatic adaptation of this system to the quality of transmission lines and switching equipment simplifies routine maintenance and is capable of advancing international cooperation without the need for a complicated agreement procedure.

6. References

1. M. Den Hertog, "Interregister Multifrequency-Code Signalling for Telephone Switching in Europe," *Electrical Communication,* volume 38, number 1, pages 130–164; 1963.

2. H. Pausch, A. Pfau, and F. Pfleiderer, "Ein Mehrfrequenzcode-Wählverfahren für weltweiten Fernsprechwählverkehr," *Nachrichtentechnische Zeitschrift,* volume 14, number 11, pages 560–566; 1961.

3. M. Jänke, E. Prenzel, and W. Speer, "Dynamikpresser und -dehner für Fernsprechverbindungen," *Fernmeldetechnische Zeitung,* volume 6, number 10, pages 459–468; October 1953.

4. L. Christiansen, F. Buchholtz, and W. Zaiser, "Das Sechskanal-Kompandersystem Z6NC für den Fernsprechnahverkehr," *Fernmeldetechnische Zeitung,* volume 8, pages 502–511; September 1955. Reprinted as "Compandor System Z6NC for Short-Haul Carrier Telephony," *Electrical Communication,* volume 35, number 1, pages 28–45; 1958.

5. Recommendations of the Comité Consultatif International Télégraphique et Téléphonique, Red Book, volume 6, avis Q16.

Günther A. W. Rahmig was born on 8 February 1928, in Erfurt, Germany. In 1955 he received a Dipl.-Ing. degree in telecommunications from the Technische Hochschule in Stuttgart.

In 1946 he joined Standard Elektrik Lorenz, working as a technician while at college. On graduation, he was assigned to the engineering development of new telephone sets. He is now deputy chief of the laboratory for audio-frequency transmission.

Lorenz Gasser. Biography appears in volume 39, number 2; 1964; page 243.

General Power Relations of Cyclotron Waves

V. DUBRAVEC
Standard Elektrik Lorenz AG; Stuttgart, Germany

1. Introduction

In recent years electron-beam parametric amplifiers have been gaining in importance as low-noise amplifiers operating in the microwave range. Their development was stimulated by the difficulties encountered with low-noise traveling-wave tubes. It has been recognized that in principle it is not possible to remove slow noise waves from the beam; for this reason also the noise figure of a traveling-wave tube cannot become extremely low. Fast noise waves, on the other hand, can be removed from the beam. This suggested the use of fast signal waves for amplification. A single coupler can simultaneously remove the noise and insert the signal. The excited signal wave can then be amplified parametrically.

Hitherto, this amplification method has failed for space-charge waves, primarily because the propagation velocity of such waves is not sufficiently frequency-dependent. Use of cyclotron waves, however, led to success [1]. Although a number of very-good low-noise cyclotron wave amplifiers are available today, the development in this field is not yet complete. One of the main objectives now is to decrease the excessive magnetic field intensity required for frequencies above 1 gigahertz; to achieve this the simple Cuccia coupler must be replaced by some other types.

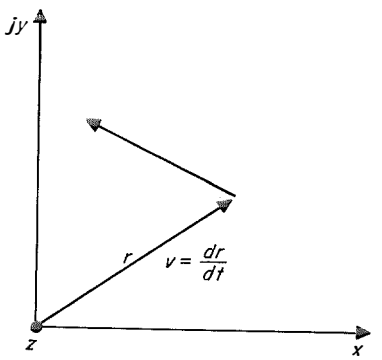


Figure 1—Coordinate system to describe electron motion in homogeneous magnetic field $B = B_z$.

It is of interest to know whether a particular coupler is able to extract the noise from a wave. This depends on whether the energy is to be added to or extracted from a beam when out-coupling the wave. In general it must be assumed that in out-coupling a given wave with various types of coupler, different quantities of power will be extracted from the beam. Consequently, the power transferred does not depend on the state of the beam, but rather on the properties of the coupler. To be able to consider the power transferred by synchronous or cyclotron waves as a state variable, we must restrict ourselves to couplers with potential fields only. Each coupler of this class extracts the same power during demodulation of a given wave. These problems have been investigated by Siegman [2] in his study of the coupling between cyclotron (or synchronous) waves and slow-wave circuits. Since such calculations are based on the simplified equivalent circuit of a slow-wave structure, the results obtained are correspondingly unreliable. For this reason the power exchange between synchronous or cyclotron waves and any arbitrary potential field is investigated here in a general way without considering particular coupler designs. The results thus obtained permit a clear view of the interaction mechanism and confirm Siegman's results.

All symbols used in this paper are explained in the body or in the Appendix. Wherever possible, small letters are used for instantaneous real quantities and capital letters for complex amplitudes. To avoid misunderstandings the complex amplitudes are designated by a caret over the character only for linear vector components and scalar quantities.

2. Fundamentals

The motion of electrons in the homogeneous magnetic field $B = B_z$ is preferably described as in Figure 1 with the aid of the coordinate system 0 (x, jy, z). The transverse (x,y plane)

electron position is given by the radius vector

$$r = x + jy. \tag{1}$$

In general the transverse components of other vector quantities are then of the form

$$a = a_x(x, y, z, t) + ja_y(x, y, z, t) \tag{2}$$

where a_x and a_y are real functions.

In this notation Newton's equation of motion in the transverse plane becomes

$$\frac{d^2r}{dt^2} = -\eta\epsilon + j\omega_c\frac{dr}{dt} \tag{3}$$

where ϵ = electric field, $\eta = |e|/m$ and $\omega_c = \eta B$ = cyclotron frequency. The second term on the right-hand side is proportional to the Lorentz force $f_L = jeBv$, the factor j taking into account the orthogonality between f_L and $v = dr/dt$.

When viewing the motion of a 1-dimensional train of electrons moving with a velocity v_0 in the z direction along the axis, we must consider the z coordinate of an individual electron as a function of time and write

$$\frac{d}{dt}\{r[z(t), t]\} = \frac{\partial r}{\partial t} + \frac{\partial r}{\partial z}\frac{dz}{dt} = \left(\frac{\partial}{\partial t} + v_0\frac{\partial}{\partial z}\right)r. \tag{4}$$

Equation (3) can now be written as

$$\frac{\partial^2 r}{\partial z^2} + \frac{2}{v_0}\frac{\partial^2 r}{\partial z\partial t} + \frac{1}{v_0^2}\frac{\partial^2 r}{\partial t^2} - j\frac{\omega_c}{v_0}\frac{\partial r}{\partial z} - j\frac{\omega_c}{v_0^2}\frac{\partial r}{\partial t} = -\frac{\epsilon}{2U_0} \tag{5}$$

(with $v_0^2 = 2\eta U_0$). For sinusoidal variations with time the field ϵ can be expressed by its circularly polarized components (refer to Appendix)

$$\epsilon = E_+e^{j|\omega|t} + E_-e^{-j|\omega|t}. \tag{6}$$

By introducing (6) into (5) and by also expressing r in terms of circularly polarized components $r_+ = R_+ e^{j|\omega|t}$ and $r_- = R_- e^{-j|\omega|t}$, two separate differential equations can be derived. To simplify the calculation we shall use in (5) only

$$\epsilon = E e^{j\omega t} \tag{7}$$

$$r = R e^{j\omega t} \tag{8}$$

and permit ω to become negative also, to obtain an equation valid for both polarizations. Equation (5) then becomes

$$\frac{d^2R}{dz^2} + j(2\beta_e - \beta_c)\frac{dR}{dz} - \beta_e(\beta_e - \beta_c)R = -\frac{E}{2U_0} \tag{9}$$

where

$$\beta_e = \frac{\omega}{v_0} \tag{10}$$

and

$$\beta_c = \frac{\omega_c}{v_0}. \tag{11}$$

The polarization is now determined solely by the sign of β_c. With $E = 0$, the solution for (9) is

$$R = R_1 e^{-j\beta_1 z} + R_2 e^{-j\beta_2 z} \tag{12}$$

where the phase constants

$$\beta_1 = \beta_e - \beta_c \tag{13}$$

$$\beta_2 = \beta_e \tag{14}$$

are solutions of the characteristic equation of (19) ($R = e^{-j\beta z}$ assumed). The first term of (12) is the cyclotron wave, the second one the synchronous wave. The instantaneous values of these waves are

$$r_1 = R_1 e^{j(\omega t - \beta_1 z)} \tag{15}$$

$$r_2 = R_2 e^{j(\omega t - \beta_2 z)}. \tag{16}$$

For t = constant, these equations yield the instantaneous configuration of the beam pattern along the axis; the beam is wound around the z axis in the form of a helix. For $z = z_0$ = constant, they yield the motion of the point of beam intersection with the plane $z = z_0$. The velocities of these motions are

$$v_{s1} = \frac{\partial r_1}{\partial t} = j\omega r_1 \tag{17}$$

$$v_{s2} = \frac{\partial r_2}{\partial t} = j\omega r_2. \tag{18}$$

These are expressions for the transverse velocities of the beam and not for velocities of individual electrons. To determine the electron velocity we use (4), taking into account (13) and (14).

$$v_1 = \left(j\omega + v_0 \frac{\partial}{\partial z}\right) r_1 = j\omega_c r_1 \qquad (19)$$

$$v_2 = \left(j\omega + v_0 \frac{\partial}{\partial z}\right) r_2 = 0. \qquad (20)$$

We can derive from (20) that the electron helix of a synchronous wave does not rotate, since the transverse velocity of the individual electrons equals zero; the pattern as a whole moves (without deformation) with the velocity v_0 in the z direction.

According to (19), the helix of a cyclotron wave rotates around the z axis with the velocity ω_c, moving simultaneously with the velocity v_0 in the z direction. Thus the rotation of the electron helix is the only difference between cyclotron waves and synchronous waves in a field-free space.

3. Power and Energy Relations

If a beam carrying the charge of Q coulombs per meter is immersed in the transverse electrical field $\epsilon = \epsilon_x + j\epsilon_y$, a beam element of the length dz will absorb from the field the power

$$dP = Q(\epsilon_x v_x + \epsilon_y v_y)\mathrm{d}z = \tfrac{1}{2}Q(\epsilon v^* + \epsilon^* v)\mathrm{d}z \quad (21)$$

where $v = v_x + jv_y$ is the transverse velocity. This power is in general a time-variable instantaneous quantity.

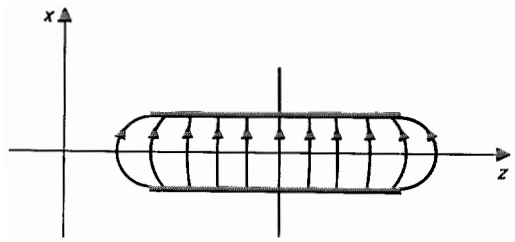


Figure 2—Cuccia coupler.

If the field is not completely transverse, the power exchange with the z component must be taken into account. In practice it is not possible to realize a completely transverse field for the entire beam length, but only for a portion of it. The Cuccia coupler [3] may serve as an example (see Figure 2).

The field between the electrodes is transverse and the stray fields at the electrode edges have an unavoidable z component. This axial component is also unavoidable in a slow-wave circuit—a component that is of considerable practical importance. For a more-detailed investigation of these relations, assume that $\epsilon_z(0, 0, z, t) = 0$ on the axis of the unmodulated beam, as for example in the Cuccia coupler (Figure 2). The modulated beam, deflected to the point $r(t)$, experiences there a field $\epsilon_z(r, z, t) = \epsilon_z(x, y, z, t) \neq 0$.

For small values of x and y, we expand the field $\epsilon_z(x, y, z, t)$ into a Taylor series about $x = y = 0$ and retain only the first terms. With $\epsilon_z(0, 0, z, t) = 0$, then we can write

$$\epsilon_z(x, y, z, t) = x\left(\frac{\partial \epsilon_z}{\partial x}\right)_{x=y=0} + y\left(\frac{\partial \epsilon_z}{\partial y}\right)_{x=y=0}. \qquad (22)$$

If the field is curl free—a condition normally met in practice—and the equations

$$\frac{\partial \epsilon_z}{\partial x} = \frac{\partial \epsilon_x}{\partial z} \qquad (23)$$

$$\frac{\partial \epsilon_z}{\partial y} = \frac{\partial \epsilon_y}{\partial z} \qquad (24)$$

are satisfied at every point and instant of time, we can write for (22)

$$\epsilon_z(x, y, z, t) = x\left(\frac{\partial \epsilon_x}{\partial z}\right)_{x=y=0} + y\left(\frac{\partial \epsilon_y}{\partial z}\right)_{y=x=0} \qquad (25A)$$

or

$$\epsilon_z(r, z, t) = \frac{1}{2}\left[r^*\left(\frac{\partial \epsilon}{\partial z}\right)_{r=0} + r\left(\frac{\partial \epsilon^*}{\partial z}\right)_{r=0}\right]. \qquad (25B)$$

According to (25B) the power delivered by the field ϵ_z to the beam is

$$\mathrm{d}P_z = Q(v_0 + v_z)\epsilon_z(r, z, t)$$
$$= \tfrac{1}{2}Q(v_0 + v_z)\left(r^*\frac{\partial \epsilon}{\partial z} + r\frac{\partial \epsilon^*}{\partial z}\right). \qquad (26)$$

If we neglect the alternating component v_z, which is small in comparison to v_0, then according to (21) and (26), we can write for the total power absorbed by the beam from the field

$$dP = \frac{Q}{2}\left[(\epsilon v^* + \epsilon^* v) + v_0\left(r^*\frac{\partial \epsilon}{\partial z} + r\frac{\partial \epsilon^*}{\partial z}\right)\right]dz. \quad (27)$$

With

$$v = \frac{\partial r}{\partial t} + v_0\frac{\partial r}{\partial z} = v_s + v_0\frac{\partial r}{\partial z} \quad (28)$$

equation (27) becomes

$$dP = \frac{Q}{2}\left[(\epsilon v_s^* + \epsilon^* v_s) + v_0\frac{\partial}{\partial z}(\epsilon r^* + \epsilon^* r)\right]dz. \quad (29)$$

Thus the power calculation has been reduced to only transverse components.

To obtain the mean value of power in the time-periodic case, we introduce the circularly polarized amplitudes according to the equations

$$\begin{aligned} \epsilon &= E_+ e^{j\omega t} + E_- e^{-j\omega t} \\ v_s &= V_{s+} e^{j\omega t} + V_{s-} e^{-j\omega t} \\ r &= R_+ e^{j\omega t} + R_- e^{-j\omega t} \end{aligned} \quad (30)$$

into (29) and take the time average.

As explained in the Appendix, we write

$$d\bar{P} = d\bar{P}_+ + d\bar{P}_- \quad (31)$$

where

$$d\bar{P}_+ = \frac{Q}{2}\left[(E_+V_{s+}^* + E_+^*V_{s+}) + v_0\frac{d}{dz}(E_+R_+^* + E_+^*R_+)\right]dz \quad (32)$$

$$d\bar{P}_- = \frac{Q}{2}\left[(E_-V_{s-}^* + E_-^*V_{s-}) + v_0\frac{d}{dz}(E_-R_-^* + E_-^*R_-)\right]dz. \quad (33)$$

Thus the mean power can be computed separately for different polarizations. Equations (32) and (33) are not time-dependent and can be integrated. The total power delivered by a coupler to the electron waves is

$$\bar{P} = \bar{P}_+ + \bar{P}_- \quad (34)$$

where

$$\bar{P}_+ = \frac{Q}{2}\int_{z_1}^{z_2}(E_+V_{s+}^* + E_+^*V_{s+})dz \quad (35)$$

$$\bar{P}_- = \frac{Q}{2}\int_{z_1}^{z_2}(E_-V_{s-}^* + E_-^*V_{s-})dz. \quad (36)$$

The terms within the pair of parentheses on the right-hand side of (32) and of (33) do not contribute to the integrals, as the integration limits z_1 and z_2 are outside the coupler where $E_+ = 0$ and $E_- = 0$. Consequently the transverse beam velocity $v_s = \partial r/\partial t$, and not the electron velocity $v = dr/dt$, is decisive for the power exchange. This has been proved by Siegman [2] for the particular case of a traveling-wave coupler. With such a coupler we can write for the passive coupling case

$$E_+ = E_{0+}e^{-j\beta z} \quad (37)$$

$$R_+ = R_{0+}e^{-j\beta z} \quad (38)$$

where β is the propagation constant of a certain mode. By introducing (37) and (38) into (32), we obtain

$$d\bar{P}_+ = \frac{Q}{2}(E_+V_{s+}^* + E_+^*V_{s+})dz \quad (39)$$

since

$$\frac{d}{dz}(E_+R_+^* + E_+^*R_+) = 0.$$

In this case the power at every point z is exchanged only with the transverse beam velocity v_s and not with the electron velocity v. With the Cuccia coupler this is no longer the case. By introducing into (32) $E_+(z)$ = constant, for example, as can be done for the homogeneous portion of the field between the capacitor plates, we can write

$$d\bar{P}_+ = \frac{Q}{2}\left[E_+\left(V_{s+}^* + v_0\frac{dR_+^*}{dz}\right) + E_+^*\left(V_{s+} + v_0\frac{dR_+}{dz}\right)\right]dz$$

or

$$d\bar{P}_+ = \frac{Q}{2}(E_+ V_+^* + E_+^* V_+)dz \quad (40)$$

where

$$V_+ = V_{s+} + v_0 \frac{dR_+}{dz} \quad (41)$$

obviously is the circularly polarized amplitude of the electron velocity. The power exchange with the electron velocity v takes place within the homogeneous portion of the Cuccia coupler; within the inhomogeneous portion the exchange mechanism is a combination of both extremes. Despite this complicated behavior, (35) and (36) give the total power exchange with any type of coupler.

Since power exchange with the transverse beam velocity v_s proved to be very important, let us review the left half of (32) and (33) in more detail. For simplicity we write it without polarization signs as one single equation

$$d\bar{P}_s = -\frac{I_0}{2v_0}(EV_s^* + E^* V_s)dz \quad (42)$$

with Q substituted by the direct current

$$I_0 = -Qv_0. \quad (43)$$

While the waves are excited by the field E, we can introduce E from the equation of motion (9) into (42). With minor transformation and consideration of

$$\frac{d^2R^*}{dz^2}R - \frac{d^2R}{dz^2}R^* = \frac{d}{dz}\left(\frac{dR^*}{dz}R - \frac{dR}{dz}R^*\right) \quad (44)$$

$$\frac{dR^*}{dz}R + \frac{dR}{dz}R^* = \frac{d}{dz}(RR^*) \quad (45)$$

the result can be written as

$$\frac{d\bar{P}_s}{dz} = j\beta_e U_0 I_0 \frac{d}{dz}\left[\left(\frac{dR^*}{dz}R - \frac{dR}{dz}R^*\right) - j(2\beta_e - \beta_c)RR^*\right]. \quad (46)$$

The total power delivered by the coupler can be found by the integration of (46) between the limits z_1 and z_2 including the entire coupler region. If the beam enters the coupler unmodulated, that is, $R(z_1) = 0$, we can express the power at the coupler output as

$$P_s(z_2) = j\beta_e U_0 I_0 \left[\left(\frac{dR^*}{dz}R - \frac{dR}{dz}R^*\right) - j(2\beta_e - \beta_c)RR^*\right]. \quad (47)$$

According to (35), $\bar{P}_s(z_2)$ represents the total power $\bar{P}$ delivered by the coupler, because we used the velocity v_s in our calculation.

If the cyclotron wave

$$R = R_1 e^{-j(\beta_e - \beta_c)z} \quad (48)$$

is excited in the coupler, its power can be determined by introducing (48) and

$$\frac{dR}{dz} = -j(\beta_e - \beta_c)R \quad (49)$$

into (47). The power carried by a cyclotron wave is thus

$$\bar{P}_1 = \beta_e \beta_c U_0 I_0 RR^* = \frac{1}{2\eta} I_0 \omega \omega_c RR^*. \quad (50)$$

This is the equation derived by Siegman for a traveling-wave coupler. As has been shown here, it applies to any type of coupler.

If the synchronous wave

$$R = R_2 e^{-j\beta_e z} \quad (51)$$

is excited in the coupler, its power is calculated —similarly to the cyclotron wave—with the aid of (47)

$$\bar{P}_2 = -\beta_e \beta_c U_0 I_0 RR^* = -\frac{1}{2\eta} I_0 \omega \omega_c RR^*. \quad (52)$$

Depending on the sign of ω or β_e, this power is either positive or negative. Since cyclotron wave and synchronous wave are excited simultaneously in almost all cases, their powers can simply be added. Normally, all four waves (two positively and two negatively polarized) are excited simultaneously, their radius vectors and powers being added.

Finally, the total power delivered by the transverse field to the electrons is computed. For this purpose we use the electron velocity

$$V = j\omega R + v_0 \frac{\mathrm{d}R}{\mathrm{d}z} = V_s + V_k \qquad (53)$$

and write

$$\mathrm{d}\bar{P}_T = -\frac{I_0}{2v_0}\left[E(V_s^* + V_k^*) + E^*(V_s + V_k)\right]\mathrm{d}z. \qquad (54)$$

The power $\bar{P} = \bar{P}_s$ exchanged with the beam velocity v_s has been derived earlier, so that now only $\bar{P}_k$ must be calculated.

$$\mathrm{d}\bar{P}_k = -\frac{I_0}{2v_0}(EV_k^* + E^*V_k)\mathrm{d}z = -\frac{I_0}{2}\left(E\frac{\mathrm{d}R^*}{\mathrm{d}z} + E^*\frac{\mathrm{d}R}{\mathrm{d}z}\right)\mathrm{d}z. \qquad (55)$$

Proceeding as with the calculation of P_s, we introduce E from the equation of motion (9) into (55). After minor transformations the result can be written as

$$\frac{\mathrm{d}\bar{P}_k}{\mathrm{d}z} = -U_0 I_0 \frac{\mathrm{d}}{\mathrm{d}z}\left[\frac{\mathrm{d}R}{\mathrm{d}z}\frac{\mathrm{d}R^*}{\mathrm{d}z} - \beta_e(\beta_e - \beta_c)RR^*\right]. \qquad (56)$$

We now obtain the power at the coupler output by integrating (56) from z_1 to z_2 over the entire length of the coupler.

$$\bar{P}_k(z_2) = -U_0 I_0\left[\frac{\mathrm{d}R}{\mathrm{d}z}\frac{\mathrm{d}R^*}{\mathrm{d}z} - \beta_e(\beta_e - \beta_c)RR^*\right]. \qquad (57)$$

The power $\bar{P}_k$ of a cyclotron wave or a synchronous wave of the form (48) or (51) then is expressed alternatively as

$$\bar{P}_{k1} = U_0 I_0(\beta_c^2 - \beta_e\beta_c)RR^* \qquad (58)$$

$$\bar{P}_{k2} = U_0 I_0 \beta_e \beta_c RR^*. \qquad (59)$$

The total power delivered by the transverse field to a synchronous or cyclotron wave is then in accordance with (50) and (58) or (52) and (59)

$$\bar{P}_{T1} = \bar{P}_1 + \bar{P}_{k1} = U_0 I_0 \beta_c^2 RR^* \qquad (60)$$

$$\bar{P}_{T2} = \bar{P}_2 + \bar{P}_{k2} = 0. \qquad (61)$$

For the cyclotron wave, (60) simply expresses the total transverse rotational energy transferred per second through the plane $z = z_0$ = constant. For the synchronous wave, the electrons do not rotate and P_{T2} thus is zero.

Summarizing the results, it can be stated that the synchronous wave power $\bar{P}_2$ is exchanged only with the axial field component, while the cyclotron waves absorb the (positive) rotational energy $\bar{P}_{T1}$ from the transverse field component and the power $(\bar{P}_1 - \bar{P}_{T1})$ from the axial field component.

4. Acknowledgment

This work was sponsored by the German Ministry of Defence.

5. References

1. R. Adler, G. Hrbek, and G. Wade, "The Quadrupole Amplifier, a Low-Noise Parametric Device," *Proceedings of the IRE*, volume 47, number 10, pages 1713–1723; October 1959.

2. A. E. Siegman, "The Waves on a Filamentary Electron Beam in a Transverse-Field Slow-Wave Circuit," *Journal of Applied Physics*, volume 31, number 2A, pages 17–27; January 1960.

3. C. L. Cuccia, "The Electron Coupler—A Developmental Tube for Amplitude Modulation and Power Control at Ultra-High Frequencies," *RCA Review*, volume 10, number 2, pages 270–303; June 1949.

6. Appendix

For sinusoidal quantities, the notation $a = a_x$

$+ ja_y$ can easily be converted into conventional complex notation. The vector components of a vector $\bar{A}$ in complex notation are

$$A_x = |A_x|\, e^{j(\omega t+\varphi_x)} = \hat{A}_x\, e^{j\omega t} \qquad (62)$$

$$A_y = |A_y|\, e^{j(\omega t+\varphi_y)} = \hat{A}_y\, e^{j\omega t}. \qquad (63)$$

To be able to use the form $a = a_x + ja_y$, where a_x and a_y represent real instantaneous vector components, we must write

$$\left.\begin{aligned} a_x &= \tfrac{1}{2}(\hat{A}_x\, e^{j\omega t} + \hat{A}_x^*\, e^{-j\omega t}) \\ a_y &= \tfrac{1}{2}(\hat{A}_y\, e^{j\omega t} + \hat{A}_y^*\, e^{-j\omega t}) \end{aligned}\right\} \qquad (64)$$

and further

$$\begin{aligned} a &= a_x + ja_y \\ &= A_+\, e^{j\omega t} + A_-\, e^{-j\omega t} = a_+ + a_- \end{aligned} \qquad (65)$$

where

$$\left.\begin{aligned} A_+ &= \tfrac{1}{2}(\hat{A}_x + j\hat{A}_y) \\ A_- &= \tfrac{1}{2}(\hat{A}_x^* + j\hat{A}_y^*) \\ a_+ &= \tfrac{1}{2}(A_x + jA_y) \\ a_- &= \tfrac{1}{2}(A_x^* + jA_y^*) \end{aligned}\right\} \qquad (66)$$

or

Solving for linear components, equations (66) become

$$\left.\begin{aligned} \hat{A}_x &= A_-^* + A_+ \\ \hat{A}_y &= j(A_-^* - A_+) \\ A_x &= a_-^* + a_+ \\ A_y &= j(a_-^* - a_+) \end{aligned}\right\} \qquad (67)$$

or

Equations (65), (66), and (67) show the relation between the two forms of notation. According to (65), a general sinusoidal vector quantity a can be represented as the sum of a positively polarized vector a_+ and a negatively polarized vector a_-. In literature on this subject the terms a_+ and a_-^* are frequently used instead of a_+ and a_- as circularly polarized components of the vector $\bar{A} = \bar{\imath}A_x + \bar{\jmath}A_y$.

In this case, only the time factor $e^{j\omega t}$ must be used. The transformation equations (66) can then be written as

$$\left.\begin{aligned} a_+ &= \tfrac{1}{2}(A_x + jA_y) \\ a_-^* &= \tfrac{1}{2}(A_x - jA_y) \end{aligned}\right\}. \qquad (68)$$

When determining the complex power, as for example encountered in the Poynting theorem, we are dealing with a scalar product

$$\begin{aligned} \bar{A}\bar{B}^* &= A_xB_x^* + A_yB_y^* \\ &= \hat{A}_x\hat{B}_x^* + \hat{A}_y\hat{B}_y^* = 2P_c. \end{aligned} \qquad (69)$$

P_c represents the complex power. By introducing into (69) the circularly polarized coordinates from (66), the product becomes

$$\begin{aligned} P_c = \tfrac{1}{2}\bar{A}\bar{B}^* &= \tfrac{1}{2}(\hat{A}_x\hat{B}_x^* + \hat{A}_y\hat{B}_y^*) \\ &= A_+B_+^* + A_-^*B_-. \end{aligned} \qquad (70)$$

The form of the products $A_+B_+^*$ and $A_-^*B_-$ differs because we use the two time factors $e^{j\omega t}$ and $e^{-j\omega t}$. By using only the time factor $e^{j\omega t}$, as is customary in complex notation, the quantities

$$\begin{aligned} A_+' &= A_+ \quad A_-' = A_-^* \\ B_+' &= B_+ \quad B_-' = B_-^* \end{aligned} \qquad (71)$$

must be considered circularly polarized components of the vectors $\bar{A}$ and $\bar{B}$, as stated earlier. With these terms

$$\begin{aligned} P_c &= \tfrac{1}{2}(\hat{A}_x\hat{B}_x^* + \hat{A}_yB_y^*) \\ &= A_+'B_+'^* + A_-'B_-'^*. \end{aligned} \qquad (72)$$

The real part of P_c represents the mean value over the time of the actual power. From equation (70) we derive

$$\begin{aligned} \bar{P} = Re(P_c) = \tfrac{1}{2}[&(A_+B_+^* + A_-^*B_-) \\ &+ (A_+^*B_+ + A_-B_-^*)]. \end{aligned} \qquad (73)$$

On the other hand, the instantaneous power can be written as

$$P(t) = a_xb_x + a_yb_y = \tfrac{1}{2}(ab^* + a^*b). \qquad (74)$$

By introducing the circularly polarized components of a and b into this equation, the resulting terms, which are not contained in the time factor $e^{j2\omega t}$ or $e^{-j2\omega t}$, may be considered as the mean value of $P(t)$ over the time. We can thus write

$$\bar{P} = \overline{P(t)} = \tfrac{1}{2}\overline{(ab^* + a^*b)} = \tfrac{1}{2}[(A_+B_+^* + A_-^*B_-) + (A_+^*B_+ + A_-B_-^*)] \quad (75)$$

with the horizontal line denoting the mean value of the term below the line. This is confirmed by comparison of (75) with (73).

Vladimir Dubravec was born in Yugoslavia on 4 January 1929. He graduated as a radio engineer from the University of Zagreb in 1955.

From 1956 to 1959 he worked with Radio Zagreb as an audio control and studio engineer. In 1959 he joined Standard Elektrik Lorenz, where he is currently in charge of the development of picture tubes.

Accelerated Life Testing and Over-Stress Testing of Transistors *

J. M. GROOCOCK
Standard Telephones and Cables Limited; Footscray, Sidcup, Kent, England

1. Introduction

Complete proof that a transistor has a specified reliability in a particular equipment is established only by successful operation in that equipment for the required period. This information has no direct practical value to the equipment designer, because it cannot be available when decisions are being made about the suitability of the transistor for a particular application. At this time only estimates of its reliability can be made. These estimates are deduced from earlier results using various assumptions, which may or may not be true.

The following three assumptions must often be made in using conventional life test data.

(A) The devices being bought for use now have the same reliability as those on which the life test information was obtained.

(B) The operating condition applying in the equipment is in all ways comparable to that used in the life test.

(C) A quantitative reliability figure based on a given number of life test hours will apply even though the period for which reliability is required may be considerably longer (or shorter) than that in which the life test information was obtained.

For very-high reliability (for example 0·001 per cent for 1000 hours) it is clear that the life test results would have acceptable validity only if the evidence for making the above three assumptions was very strong.

One type of life test that to some extent eliminates the need for these assumptions uses over-stress methods of accelerating failure mechanisms. There are two common definitions of accelerated or over-stress tests. In one, they are defined as tests carried out at stress levels that exceed in some respect the ratings set by the manufacturer. With this definition it is possible to classify all tests as either within rating or accelerated.

For the second definition it is assumed that tests are accelerated or over-stress if carried out at stress levels more extreme than those of the actual practical application. Here the classification of a test as realistic or accelerated depends on the particular application of the device. For the second definition, tests performed at stresses within the ratings of a device will be accelerated for some applications. By this definition also, many if not most of the reliability tests called for in military specifications will be accelerated.

The usefulness of accelerated life tests and over-stress tests always depends on some known or assumed relationship between the results obtained under the accelerated condition and the results obtained under the actual working condition of the device. By definition, the over-stress condition is not of direct interest to the user. In some cases the extrapolation from the over-stress condition to the normal condition is based on very-simple reasoning (for example, a sample of devices of type *A* proved more reliable than a sample of devices of type *B* under the over-stress condition; therefore there is a probability greater than ½ that devices of type *A* will be more reliable than devices of type *B* under normal stress). Another piece of simple reasoning is

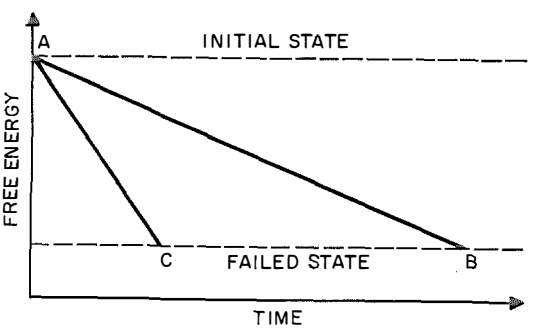


Figure 1—Change in free energy with time.

* Reprinted from *Electronics Reliability and Microminiaturization*, volume 2, number 3, pages 191–204; July–September 1963.

that if a sample of devices showed satisfactory reliability when subjected to a high stress for a short time, it is likely that similar samples will stand up to a lower stress for a longer time. In other cases, knowledge of the physical basis of the failure mechanisms enables more-precise extrapolation to be made.

Accelerated tests carried out on samples can give useful information on the reliability of the batches of transistors from which they are drawn, without any difficult and expensive bonding procedure. Furthermore, extrapolation and interpolation are essential factors in the use of accelerated life tests; the derived relationship between stress level and failure rate enables the results expected from a whole range of values of that particular stress to be inferred. Finally, because small samples can be tested, most of the stresses (thermal, electrical, and mechanical) to which the device may be subjected can be investigated, so that any weaknesses are revealed.

2. Physical Basis of Accelerated Life Tests

Failure of a transistor on life test or in operation results from the transistor changing over a period of time from one thermodynamic state (the initial state) to another state (the failed state) of lower free energy (see Figure 1). This will be generally true for failures progressively affecting the population of transistors; it does not necessarily apply to a very-small proportion of abnormal failures that may not follow the laws of thermodynamics.

In accelerated life tests, an attempt is made to set up an operating condition that accelerates in a controlled manner the transition from the initial state to the failed state, so that in Figure 1 path *AC* is followed rather than path *AB*. However, the change in operating condition will change both the initial state and the failed state of the transistor, as indicated in Figure 2, and under the accelerated condition the transition from the initial state to the failed state will follow path *CD* of Figure 2 rather than path *AC* of Figure 1. Often, however, the initial and failed states for the normal and accelerated conditions will not differ markedly. In these cases the methods of reaction kinetics can be used directly to extrapolate from the results of the accelerated test to the normal condition. In other cases, particularly if any transistor parts change physically between the normal condition and the accelerated condition, the initial and failed states may be changed markedly, and the extrapolation from the accelerated to the normal condition may be so difficult as to be impracticable.

Reaction kinetics indicates that reaction rates are usually affected to a major extent by the concentration of reacting elements and by the temperature. For concentration, reactions are classified as zero, first, or second order, as their reaction rates are independent of or are dependent on the first or second powers of the concentration of some reacting elements. Reaction rates often diminish progressively as some vital reactant is used up. However, concentration is only of secondary importance in accelerated life tests of transistors, because it is not usually adjustable at the will of the experimenter. An example in which concentration might be important is the deterioration of germanium alloy transistors under high-temperature storage, due to the release of water vapour into the gas surrounding the transistor element. If the amount of water available is small, the rate at which it is released might drop as its concentration dropped,

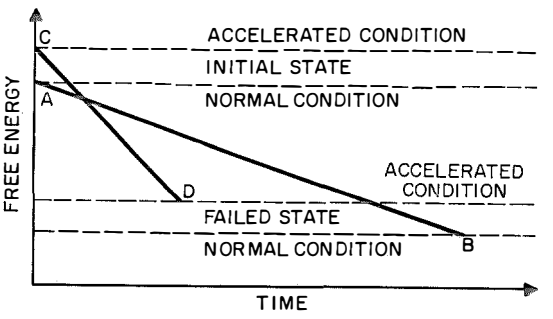


Figure 2—Effect of change in operating conditions.

even under constant temperature conditions. The associated deterioration of electrical characteristics might then approach a constant level, rather than continue toward complete failure.

Reaction kinetics gives a number of expressions relating reaction rate to temperature, and these are discussed below. Reaction kinetics also considers reactions in which rates depend on electrical fields or the transport of material by electrical currents.

Undoubtedly the most-valuable and widely employed method of accelerating failure mechanisms in transistors is to use a storage or operating temperature higher than the one the devices will be subjected to in actual equipments. This usefulness stems from three reasons.

(A) Particularly if storage rather than operation is performed, over-stressing can be carried out easily and cheaply with no special equipment besides the temperature-controlled ovens normally available to any group doing component reliability work.

(B) Many physical and chemical processes that might lead to failure are accelerated by increased temperature.

(C) The relationship governing this acceleration very often follows the Arrhenius equation, and this enables quantitative extrapolations to be made readily.

The Arrhenius equation in its integrated form [1] can be expressed as

$$\frac{dm}{dt} = A \exp\left(-\frac{E}{RT}\right) \qquad (1)$$

where dm/dt is the rate of a chemical reaction, T is the absolute temperature, R is the gas constant (1·987 calories/(degree mole)), and A and E are constants, the latter being the activation energy.

The Arrhenius equation is essentially an expression for relating reaction rate to temperature. It applies directly only if the initial and final states (see Figure 1) are not markedly changed over the temperature range being considered, and only if other factors that affect reaction rate (such as concentration) do not vary significantly.

If the Arrhenius equation is applied to transistor failure, it is assumed that failure is due to a reaction that has progressed to an amount M_0 at the initial state (time $= t_0$) and to an amount M_1 at the failed state (time $= t_1$). Integrating (1) for constant temperature we obtain

$$M_1 - M_0 = \{A \exp(-E/RT)\} \quad (t_1 - t_0).$$

The time for failure $t = t_1 - t_0$ (that is, the time for the reaction to progress by an amount $\Delta M = M_1 - M_0$) is therefore given by

$$t = \frac{\Delta M}{A} \exp\left(\frac{E}{RT}\right)$$

and taking logarithms

$$\log t = \log \frac{\Delta M}{A} + 0{\cdot}434 \left(\frac{E}{RT}\right). \qquad (2)$$

In life testing it is assumed that the amount of reaction ΔM leading to failure can be related directly to some recognizable change in electrical characteristic such as leakage current or gain, or to the occurrence of catastrophic failure. For example, the time for leakage current to reach 100 microamperes is measured at different temperatures and a plot of logarithmic time against the reciprocal of absolute temperature is made; this plot should give a straight line that can be used for making the extrapolation.

The Arrhenius equation, although derived by analogy to a thermodynamic relationship (the van't Hoff equation), is essentially empirical. Several attempts have been made to give it a sound theoretical basis and possibly the most fruitful of these was by Eyring [2]. In its simplest form the Eyring equation can be expressed as

$$\frac{dm}{dt} = BT \exp\left(-\frac{E}{RT}\right) \qquad (3)$$

where B is a constant and the other symbols have the same meaning as in (1). This equation leads to the following relationship between failure time and absolute temperature.

$$\log t = \log \frac{\Delta M}{B} - \log T + 0{\cdot}434 \left(\frac{E}{RT}\right). \quad (4)$$

This equation is more complicated than (2) in that it contains the extra term $\log T$. As a consequence, experimental results cannot be expressed by (4) simply by plotting $\log t$ against $1/T$.

It is apparent that the Arrhenius equation is a simpler approximation of the Eyring equation. For transistor life testing it is doubtful if the greater accuracy obtained by using (4) rather than (2) is justified. This is because the ranges of time and temperature relevant to transistor life testing are very limited (10–100 hours to 10^5–10^6 hours and 75 degrees centigrade to 300–400 degrees centigrade) and the precision achieved with the experimental results and required for extrapolation is not high. An example of the difference obtained by using the two equations is given in Table 1. In this hypothetical example it is assumed that the life of a particular transistor has been shown experimentally to be 100 hours at 352 degrees centigrade and 10 000 hours at 127 degrees centigrade. Using the Arrhenius based equation, the life at 21 degrees centigrade is 10^6 hours (activation energy 10·2 kilocalories/mole), whereas the Eyring based equation gives a life of $0{\cdot}871 \times 10^6$ hours at the same temperature. This example was especially chosen to have a low activation energy and yet simulate a practical situation. It is not possible to reduce the activation energy appreciably below that used in the example and still retain correspondence to practical cases. Because the difference between the results obtained with the Eyring and Arrhenius equations decreases rapidly with increasing activation energy, the results in Table 1 represent about the maximum possible difference.

Transistor failure rates can be influenced by a potential difference applied across a p-n junction and by the current flowing through the junction. In this case the Arrhenius equation (2) becomes modified to

$$\frac{dm}{dt} = f(V)\, f(i) \exp\left(-\frac{E}{RT}\right) \quad (5)$$

where $f(V)$ and $f(i)$ indicate the dependence on voltage and current respectively. Analysis of cases complicated in this way demands factorial life test experiments or modifications of these experiments such as the use of Latin squares [3].

For actual transistors it is possible, of course, to have two or more different failure mechanisms that are accelerated unequally by increase of temperature (that is, having different activation energies). A hypothetical example is illustrated in Figure 3. A device has two completely separate mechanisms, each leading to failure. Accelerated life tests carried out at 162 and 227 degrees centigrade give the two failure times, indicated by X's on Curve A. On the basis of these results, a prediction is made that the failure time at 60 degrees centigrade is 1 million hours. However, for temperature below 120 degrees centigrade, failures due to the second mechanism occur in shorter times than those due to the first; in fact the failure time at 60 degrees centigrade is less than 100 000 hours. In general, the predictions made from accelerated life tests carried out on devices having two or more failure mechanisms, each following the Arrhenius equation, will be either correct or optimistic.

TABLE 1

CALCULATED RESULTS OF TRANSISTOR LIFE TESTING

Equation Used	Temperature in Degrees		Time to Failure in Hours
	Centigrade	Kelvin	
—	352	625	$1{\cdot}000 \times 10^2$
—	127	400	$1{\cdot}000 \times 10^4$
Arrhenius	21	294	$1{\cdot}000 \times 10^6$
Eyring	21	294	$0{\cdot}871 \times 10^6$

For this reason actual results under the condition to which extrapolation is made are unlikely to be better than the extrapolation would indicate, although they may well be worse.

3. Application of Arrhenius Equation to Results Obtained from High-Temperature Storage of Germanium Alloy Transistors

The use of the Arrhenius equation is exemplified by results obtained from storage at 100, 115, and 135 degrees centigrade of 3 samples each of 24 transistors drawn from the *TED540* product line. These *p-n-p* alloy transistors were made before the use of molecular sieve on this product line greatly improved reliability. The storage rating of this type is 75 degrees centigrade.

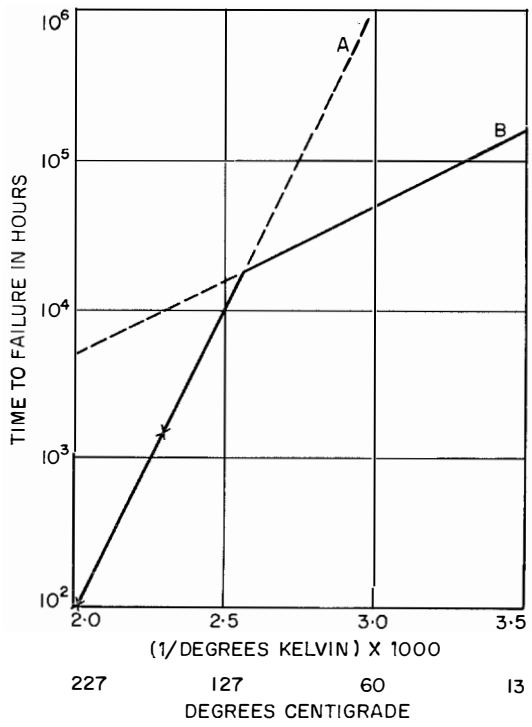


Figure 3—Failure time as a function of temperature for two failure mechanisms. Curves *A* and *B* represent two different failure mechanisms that behave differently as a function of temperature.

A failure criterion of $I_{CBO} = 10$ microamperes with $V_{CB} = -9$ volts was defined. The value was chosen so that a failure required a marked increase in leakage current (the initial distribution was in the range 0·2 to 1·5 microamperes) but was not so high that the number of failures was reduced. The results obtained are plotted as curves *A, B,* and *C* in Figure 4.

Figure 4 also shows the importance that the assumption of a known failure distribution—in this case the log-normal—can have. The logarithmic time scale and the arithmetic probability scale are such that a log-normal failure distribution gives a straight line.

By assuming that the Arrhenius equation is followed and that the failure distribution is log-normal, the distribution of failure times at other temperatures can be calculated. Curve *E* in Figure 4 is such a calculated distribution for storage at 75 degrees centigrade. For comparison with this calculated value, we have the results of storage at 75 degrees centigrade of 135 type-*TED540* transistors. These have now been tested for more than 20 000 hours and the results are given in Curve *D*. The calculated and experimental results agree quite well.

A preliminary report on this experiment has been given [4], but in this case the over-stress tests were carried out on samples of *TED540* transistors made nearly two years later than those used for the long-term test at 75 degrees centigrade. For the results reported here, the accelerated tests were performed on samples of *TED540* transistors made at the same time as those used for the test at 75 degrees, but which had been stored at 25 degrees centigrade until used in the over-stress experiment. Comparison with the results obtained earlier suggested that the reliability of the transistor had been markedly improved during the two-year production period and that the agreement between the extrapolated distribution and the experimental distribution at 75 degrees centigrade was much poorer than that shown in Figure 4.

A more-detailed account of the method of making the extrapolation using the Arrhenius equation and log-normal distribution is given in [4]. The results are also analyzed using the Weibull distribution [5]. In addition, the physical failure mechanism operating in this case is briefly discussed.

Using Curve *D* as the basis, a further extrapolation to 45 degrees centigrade is given as curve *F* in Figure 4. This indicates that derating these transistors severely should reduce this particular failure mechanism to a negligible level. Only 0·01 per cent will have failed in 170 000 hours and the median failure time is more than 500 000 hours. However, it must be pointed out that from the sample stored at 75 degrees centigrade 1 device out of 136 failed in 280 hours. Curve *D* indicates that the proportion of devices expected to fail in 280 hours is negligible, far below 0·01 per cent. This

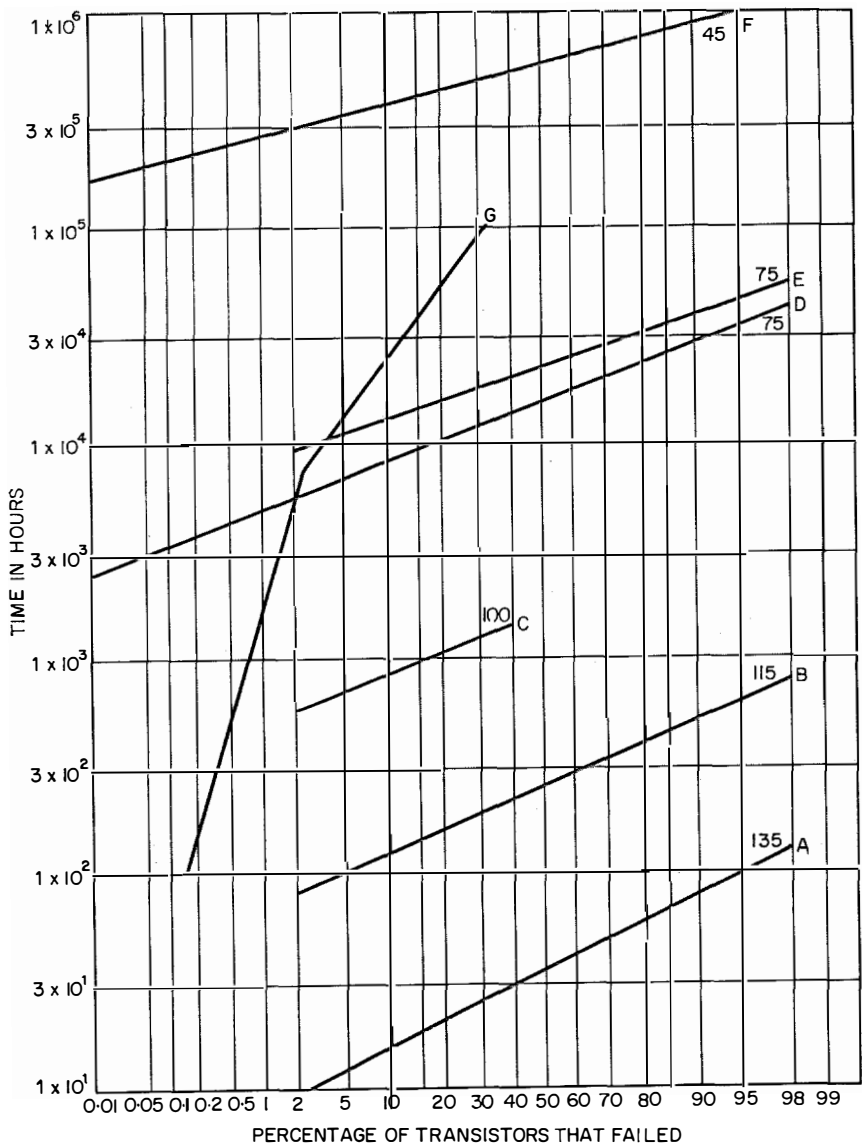


Figure 4—Log-normal plots of time-to-failure for *TED-540* transistors stored at the temperatures indicated in degrees centigrade. Curve *E* was calculated using the measured results in curves *A*, *B*, and *C*, while curve *F* was calculated from the data in Curve *D*. Curve *G* is an operating test of units dissipating between 120 and 240 milliwatts at an ambient temperature of 25 degrees centigrade.

single failure is therefore a rogue and is excluded from the plot. However, the presence of rogues at the level of 0·1 to 1 per cent means that there is little value in extrapolation to lower failure levels, unless some method of eliminating the rogues is available.

In using the log-normal failure distribution to make the extrapolation in Figure 4, straight lines have been fitted by eye to the results plotted on logarithmic probability paper. A more-accurate method of fitting curves could have been used (for example, by calculating the means and standard deviations of the logarithms of the times-to-failure). However, because the samples used were finite in size, the sample means and standard deviations would have been only estimates of the parameters of the population of transistors from which the samples were drawn. Similarly, the times obtained for particular failure proportions are only estimates of the times for the population. Curve *A* indicates that 95 per cent of the devices would fail in 100 hours. However, even if the assumption is correct that the failure distribution is log-normal, it is necessary to say that as many as 13 per cent of the devices could fail in 100 hours to have a 90-per-cent confidence in the result as applied to the population. For the much-greater sample size (135) in Curve *D*, the experimental result is 5-per-cent failures in 7000 hours. In this case the 90-per-cent confidence level, because of the larger sample, is only 8 per cent. These results were taken from curves given by Jones [6].

It is clear that similar margins should be applied when making extrapolations using the Arrhenius equation, to account for the inaccuracies in curve fitting.

Curve *G* in Figure 4 is the result obtained from an operational life test on 601 type-*TED540* transistors for 23 000 hours, 13·8 million transistor-hours in all. Curve *G* shows the results of tests carried out at a number of power levels as summarized in Table 2.

The thermal resistance of the *TED540* transistor is 0·25 degree centigrade per milliwatt; these devices were tested with junction temperatures in the range from 55 to 85 degrees centigrade (25 degrees ambient temperature). Despite the variation in junction temperature, the log-normal distribution is still followed although, as is to be expected, the standard deviation on the log of the failure times is greater than for the comparable storage test. The mean-time-to-failure is 300000 hours, compared with 20 000 hours for storage at 75 degrees centigrade. However, in both cases 2 per cent of failures occur in 6000 hours and the proportion of early failures is higher for the operational test. There is an indication that 12 failures (2 per cent) did not fit the main distribution. The dissipation in these 12 devices averaged 187 milliwatts, indicating that their early failure was not in the main due to higher than average dissipation. This result further emphasizes the need to have a rogue-free population if extrapolations to low percentiles are to have any validity.

TABLE 2

RESULTS OF OPERATIONAL LIFE TEST AT VARIOUS POWER LEVELS

Dissipation Range in Milliwatts	Number of Devices Within Range
120–130	2
130–140	55
140–150	85
150–160	58
160–170	29
170–180	78
180–190	68
190–200	78
200–210	47
210–220	87
220–230	10
230–240	4

4. Step-Stress Testing

There are two methods of carrying out over-stress tests. The first of these was touched on in Section 3. A constant stress (in this case high temperature) was applied to a sample of transistors and the number that failed was recorded at regular intervals, while the survi-

vors were further subjected to the constant stress. The one experiment gives the over-all failure distribution of the sample at the selected level of stress. The precision of the measurements is limited by the frequency at which they can be made. This may be daily or hourly, or alternatively by some continuous method of recording survival. By repeating the experiment on further samples using other levels of stress, the over-all pattern of percentage failed against time and level of stress can be obtained.

The alternative method of carrying out the over-stress tests is to use the step-stress technique developed by Dodson and Howard [7]. A sample of transistors is subjected to low stress for a fixed period. Measurements are then made and the number of failures recorded. The survivors are then subjected for the same period to a higher stress. The procedure is repeated, increasing the stress in regular steps until all the devices have failed. This experiment gives the over-all failure distribution versus level of stress for the selected period.

The experiment can be repeated on further samples using other time periods so that the over-all pattern of percentage failed against time and level of stress can be obtained. The way the method works can be illustrated by using the results given in Curves *A, B,* and *C* of Figure 4. If this experiment, using the step-stress method, had been performed with 3 samples of devices and periods of 100, 300, and 1000 hours, the results given in Table 3 would have been obtained. From this, percentage failed can be plotted against temperature on linear probability paper (using a scale linear in 1/degrees Kelvin) and the information of Figure 4 can be built up.

It is clear that in this particular experiment the step-stress method would have achieved less precision than the constant-stress experiment. This is because measurements were made at short intervals in the latter, whereas measurements could be made only at coarse temperature increments in the former. There are two reasons for this: One is experimental in that there is little point in using temperature steps smaller than 5 to 10 times the precision with which the temperature can be measured and controlled; the second reason is that it is generally assumed in step-stress experiments that stressing at a lower level does not affect the failures occurring at a higher level. This is true only if the steps of stress are fairly coarse. If temperature is the stress used, it is possible to correct for the effect of the lower stress levels [7]. However, the correction is fairly complicated and if this degree of precision is required, it is usually simpler to repeat the experiment, selecting conditions that make the correction unnecessary.

It is the author's opinion that the main value of the step-stress technique is in the early stages of an investigation, when it will quickly indicate the relevant range of stress levels.

5. Electrical Stresses

In addition to the straightforward studies of the effect of current, voltage, and power discussed above, over-stress testing permits a quick appraisal of whether a transistor can stand up to less-usual electrical stresses. Two examples of such appraisals follow.

In the first, 10 *TED504* germanium alloy transistors with molecular sieve were subjected to a series of very-high-power pulses. The pulses had an amplitude of 30 watts, a duration of 12 microseconds, and a mark-to-space ratio of 1:100; the average power dissipation was therefore 300 milliwatts. The *TED504*

TABLE 3
TRANSISTOR FAILURE PERCENTAGES IN STEP–STRESS EXPERIMENT

Period in Hours	Temperature in Degrees Centigrade		
	100	115	135
100	—	5	95
300	—	60	—
1000	17	99	—

transistor has a thermal resistance of 0·25 degree centigrade per milliwatt, so that the average junction temperature was about 100 degrees centigrade. The basic test circuit and the voltage waveforms are given in Figure 5. The test circuit was driven by a commercial pulse generator and subjected the 10 transistors simultaneously to the required operating condition.

After 2400 hours of this test none of the 10 transistors had failed or shown any marked change in gain, leakage current, or breakdown voltage. The detailed results are given in [8]. This life test is one in which, because the device has not shown degradation when subjected to high stress for a relatively short time, it is inferred that the transistor will not degrade if subjected to a lower stress for a long time. This experiment suggests that high-power pulsing does not produce any special mechanism of failure in this type of device; for normal applications, which are usually at powers considerably less than 30 watts, only the average power dissipation need be considered, provided that the duration of the high-power pulses is short compared with the thermal time constant.

In the second experiment, 36 silicon planar transistors, half *BLY10* and half *BFY15,* were divided into equal batches and subjected to reverse emitter currents of 1, 5, or 20 milliamperes. The circuit used is shown in Figure 6. In each case the emitter voltage rating was exceeded and the device driven into the avalanche region. The experiment was designed to reveal the relationship between reverse current and reliability. However, none of the samples has failed after 3000 hours of testing, and the gain, leakage currents, and breakdown voltages have remained stable. Detailed results are given in [8].

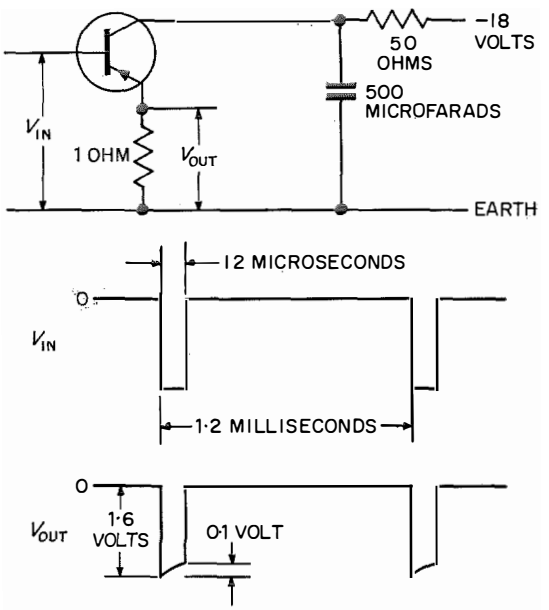


Figure 5—Test circuit and voltage waveforms for high-power-pulse life tests. The amplitude of V_{IN}, approximately 6 to 10 volts, depends on the gain of the transistor.

6. Mechanical Over-Stress Tests

As noted above, accelerated life tests may not always reveal the major failure mode for normal operation. The presence of a small percentage of rogues may also make nonsense of an extrapolation to a low failure rate. This problem is likely to be particularly severe with transistor types such as the silicon planar, the inherent reliability of which is so great that accelerated storage or operating tests usually indicate infinitesimally small failure rates for normal operating conditions. For these devices the actual failure rate may depend entirely on the percentage of rogues. Quality-control procedures then play a major part.

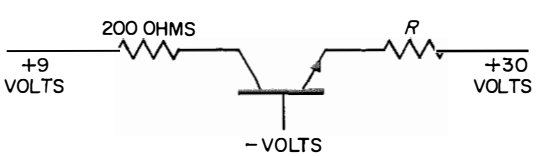


Figure 6—Circuit for applying reverse emitter current to transistors. For reverse emitter currents of 1, 5, and 20 milliamperes, the values of R were 27, 4·7, and 1·2 kilohms.

However, because many of the rogues result from mechanical defects, mechanical over-stress testing can give valuable information. As with all over-stress tests, an attempt is made, based on the failures found at high stress levels, to infer how frequently failures due to a similar mechanism are likely to occur at normal stress levels. It may be that the rogue failures found in, say, an electrical dissipation test, will be drawn entirely from the lower (but non-rogue) percentiles of a mechanical test. Information from a mechanical over-stress test might be used to improve the device, with consequent elimination of some of the rogues.

In principle, all the mechanical stresses to which transistors may be subjected can be examined in over-stress experiments. The relevant stresses include constant acceleration, shock, vibration, thermal cycling, and thermal shock. The method may be illustrated by results obtained from an experiment in which germanium alloy transistors were tested in a centrifuge. A sample of 24 transistors drawn from the *TED540* product line was subjected progressively to increasing accelerations of 5, 15, 25, and 35 thousand gravity units g. The number of failures was recorded at each level of stress and the survivors underwent further stress (step-stress technique). Another similar sample was stressed at 10, 20, 30, and 40 thousand gravity units.

Acceleration tests were made using additional samples of 24 *TED540* transistors for different stress directions relative to the axes of the transistors. Each of the failed transistors was examined to determine the mechanical cause of failure. The failure analysis was not as valuable as it might have been, because it was not practicable to remove the devices from the centrifuge jig as soon as they had failed. All devices were therefore subjected to the maximum stress, even though many of them failed at lower levels of stress.

The results of the main series of experiments are given in Table 4. In some additional tests, 2 failures occurred in direction C at 16 000 gravity units and 24 failures occurred at 32 000

TABLE 4

CUMULATIVE FAILURES OUT OF 24 FOR STRESSES IN DIRECTIONS A THROUGH F

Stress ($S \times 1000$)	Direction of Stress					
	A	B	C	D	E	F
0	0	0	0	0	0	0
5	0	0	1	0	0	0
10	0	3	0	2	—	—
15	1	0	1	5	2	6
20	3	8	8	15	14	12
25	4	9	17	24	21	23
30	6	11	20	—	—	23
35	9	14	24	24	24	24
40	15	19	24	—	—	24

A dash indicates that no test was made.

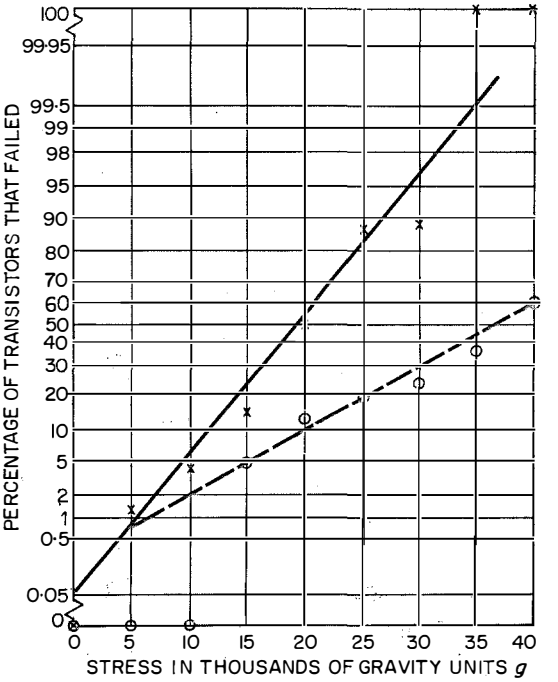


Figure 7—Failure distribution in centrifuge tests of *TED540* germanium alloy transistors. The solid curve represents the combined results for transverse stress directions C, D, E, and F in Table 4, while the broken curve represents the results for longitudinal stress direction A.

gravity units. In direction *D*, 23 failures occurred at 22 000 gravity units.

Results for the 4 transverse stress directions *C*, *D*, *E*, and *F* have been combined and plotted as the solid curve in Figure 7. This is the maximum-probability line obtained by probit analysis [9]. The results indicate that although most failures occur at stresses above 10 000 gravity units, which is far above the requirements of most practical systems, about 1 per cent of the devices may fail at 5000 gravity units and 0·1 per cent at quite-low stresses.

The results for longitudinal stress direction *A* have been plotted as the broken curve in Figure 7. Although the median level of stress causes less than ⅓ the percentage of failures caused by the corresponding stress in the transverse directions, it appears that the percentage of failures may well be higher for the longitudinal direction at stress levels below 5000 gravity units.

7. Conclusions

Accelerated life testing and over-stress testing of transistors can be valuable for predicting reliability relatively quickly and cheaply. A wide variety of stresses including thermal, electrical, and mechanical can be studied so that the possibility of revealing all important failure modes is increased. However, the actual failure rates occurring at normal levels of stress may be markedly higher than predicted, because the quantitative relation between stress and failure rate was not determined correctly, or because failure modes exist that were not revealed by the accelerated tests. For this reason it is clearly desirable to integrate the results from accelerated tests, long-term tests, and large-sample tests.

8. Acknowledgments

Grateful acknowledgment is made to the Services' Valve Test Laboratory and to the Admiralty for permission to include the results of experiments performed in their behalf.

9. References

1. S. Glasstone, "Textbook of Physical Chemistry," second edition, D. Van Nostrand Company, New York; 1946: page 1088.

2. Page 1098 of reference 1.

3. D. N. Chorafas, "Statistical Processes and Reliability Engineering," D. Van Nostrand Company, Toronto; 1960: page 130.

4. J. M. Groocock and R. H. Chilton, *Symposium Digest, Second Symposium of Electronic Equipment Reliability,* Institution of Electrical Engineers, page 43; October 1962.

5. "Reliability Training Text," second edition, Institute of Radio Engineers, New York; 1960.

6. H. C. Jones, "The Relation Between Sample Size and Confidence in Test-to-Failure Reliability Programs," *IRE Transactions on Reliability and Quality Control,* volume PGRQC-16, pages 24–33; June 1959.

7. E. A. Dodson and B. T. Howard, "A Method for the Rapid Evaluation of Semiconductor Devices," IRE Meeting of Professional Group on Electron Devices, Washington, District of Columbia; October 1960.

8. "Life Test Bulletin Number 3," Standard Telephones and Cables Limited, Transistor Division; October 1962.

9. R. A. Fisher and F. Yates, "Statistical Tables for Biological, Agricultural and Medical Research," 5th edition, Oliver and Boyd, London; 1957: pages 9–11 and Table 9.

J. M. Groocock was born in Nottingham, England, in 1929. He was educated at the Imperial College of Science and Technology, London. In 1950 he received the bachelor of science degree in chemistry and in 1955 a

doctorate in philosophy and a diploma for research in physical chemistry.

In 1957 he entered the Armament Research and Development Establishment of the Ministry of Supply. In 1958 he joined Standard Telephones and Cables, where he is now chief quality-assurance engineer of the Transistor Division.

In Memoriam

BRUNO W. SUTTER

Bruno W. Sutter was born in Saint Gall, Switzerland, on 10 July 1899. The Swiss Federal Institute of Technology conferred an electrical engineering diploma on him in 1922.

He joined the Berne Office of Bell Telephone Manufacturing Company as an engineer in the transmission and radio field. Among his engineering activities, he participated in the first transatlantic low-frequency radiotelephone system between Great Britain and the United States of America.

In 1936 he was made chief engineer of Standard Téléphone et Radio in Zurich and two years later he was appointed works manager. Early in 1946 he became sales manager and in 1960, director of business planning. He contributed substantially to the development and expansion of Standard Téléphone et Radio. His wide experience and good judgment were a great asset to the company.

Bruno W. Sutter had many friends throughout Switzerland and abroad, for he traveled widely. He died after a long and painful illness on 9 August 1964. He will be remembered for his conscientious dedication to his work and for his kindness.

VOLUME 39 • NUMBER 4 • 1964