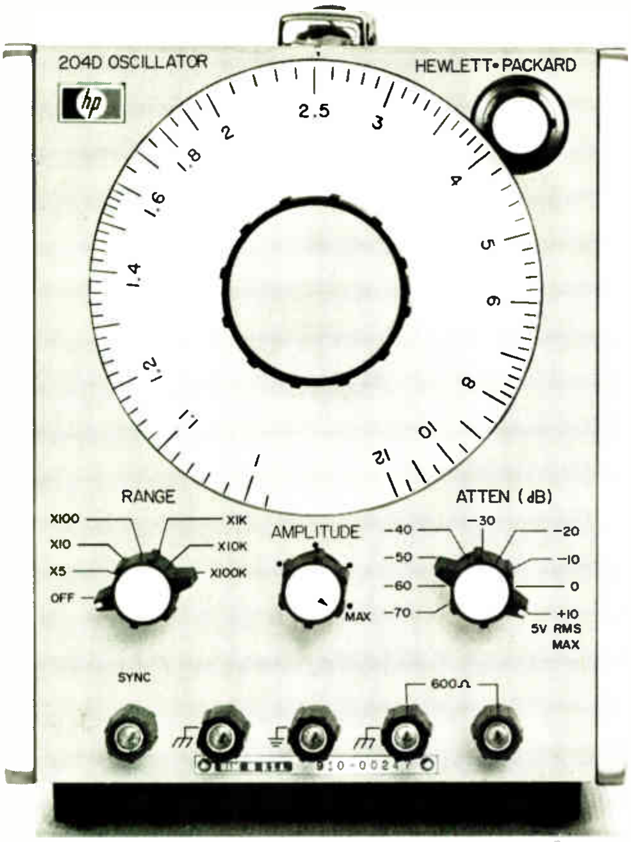
IEEE spectrum

features



Dollar for Dollar— You get more from an HP Oscillator!

Most ac work starts with an oscillator . . . why not start with the best?

And best does not necessarily mean more expensive. It means selecting the right oscillator for your requirements at the right price.

Sounds simple, until you sit down to decide which oscillator is best for you. Then, you have to compare your technical requirements against what is available. Next you make the necessary trade-offs between performance and price to insure that you get the best oscillator for your job at the best price. Finally you have to weigh the intangibles, like company reputation, service and reliability.

But there is an easier way!

HP has been in the oscillator business for a long time. In fact they invented the first stable wide bandwidth Wein - bridge oscillator. And since that time HP oscillators have built a solid reputation for performance and reliability—in fact most of you have used them in the past.

So, all you really have to do to get all the information you need to decide which oscillator is best for you is turn to page 267 of your 1970 HP catalog.

Not only will you get information on all of HP's complete line of oscillators . . . you also get the background information to help make your choice more meaningful.

HP oscillators give you a wide range of choices. Frequencies start as low as 0.00005 Hz and go as high as 32 MHz. You can even get power outputs as high as 5 watts.

And, if you have any questions, just call your local HP field engineer. He will be more than happy to answer your questions and give you any help you may need in selecting the best oscillator for your requirements.

Or, write to Hewlett-Packard, Palo Alto, California 94304. We will send you complete information and application notes on HP's oscillators. Europe: 1217 Meyrin-Geneva, Switzerland. Prices on HP oscillators start as low as $215.

090/3

HEWLETT hp PACKARD

SIGNAL SOURCES

Circle No. 2 on Reader Service Card

the cover

One of the newest minicomputers—Digital Equipment Corporation's PDP-8/E—surrounded by flowchart symbols, used to diagram the logical flow of a program, and mnemonic symbols, used to write a program in assembly language. The article beginning on page 37 discusses the use of minicomputers such as this one in a variety of applications

departments

IEEE SPECTRUM is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: 345 East 47 Street, New York, N.Y. 10017. Cable address: ITRIPLEE. Telephone: 212—752-6800. Published at 20th and Northampton Sts., Easton, Pa. 18042. Change of address can be made effective with any particular issue if received by IEEE Headquarters by the 10th of the preceding month. **Annual subscription:** IEEE members, first subscription $3.00 included in dues. Single copies $1.50. Nonmember subscriptions and additional member subscriptions available in either microfiche or printed form. Prices obtainable on request. **Editorial correspondence** should be addressed to IEEE SPECTRUM at IEEE Headquarters. **Advertising correspondence** should be addressed to IEEE Advertising Department, at IEEE Headquarters. Telephone: 212—752-6800.

OTHER IEEE PUBLICATIONS: IEEE also publishes the PROCEEDINGS OF THE IEEE, the IEEE STUDENT JOURNAL, and more than 30 Transactions for IEEE Groups with specialized interests within the electrical and electronics field. Manuscripts for any IEEE publication should be sent to the editor of that publication whose name and address are shown on page 17 of the January 1970 issue. When in doubt, send the manuscript to E. K. Gannett, Editorial Services, at IEEE Headquarters, for forwarding to the correct party.

IEEE also copublishes, with the Institution of Electrical Engineers of London, England, *Electrical & Electronics Abstracts* and *Computer & Control Abstracts*, as well as their companion titles journals, *Current Papers in Electrical & Electronics Engineering* and *Current Papers on Computers and Control.*

Forum

Readers are invited to comment in this department on material previously published in IEEE SPECTRUM; on the policies and operations of the IEEE; and on technical, economic, or social matters of interest to the electrical and electronics engineering profession.

Cleveland Section

On February 19 this year, the IEEE Cleveland Section held an ambitious panel discussion entitled: ". . . But What Are You Going to Do for Me in the '70s, IEEE?" An audience four times the average attendance listened to leaders of our profession, including 1969 President F. Karl Willenbrock.

As a first step to being responsive to members' demands in programming, the event was a qualified success: It posed the question that was most frequently asked, it brought together employed engineer and his management, a union representative and an IEEE president, members impatient for modernizing the Institute and others who cherished the status quo, and it exposed the severe limitations of the four types of engineers' organizations represented: classical and lobbying (IEEE, NSPE), sounding board and union (CPTP, ASPEP). In other words, the Cleveland Section provided the proper subject, the concerned audience, and four leaders who should have the answers.

Under those circumstances, the title and negative tone of *IEEE Spectrum*'s coverage (Cleveland Section fails meeting objectives, May, p. 105) is, to say the least, unfortunate. The theme question was addressed to IEEE and, although there were no immediate answers, Dr. Willenbrock offered to seek solutions. It should, therefore, be in the interest of the Institute and its publications to encourage efforts such as ours by stressing their positive aspects and thereby assuring your leaders that IEEE does not regard the engineer's socioeconomic problems as unfortunate, but without solution.

Herbert H. Heller
Chairman, IEEE Cleveland Section

Feedback loops in economics

I agree completely with J. M. Green (Forum, June 1970) when he says that engineers should become more involved in social and economic problems; and that technological solutions exist, or can be made to exist, for many of the problems that beset our society. I was disappointed to learn, however, that the suggested preparation for this task is a reading of two novels by Ayn Rand!

Although I am not qualified to comment on the literary merits of Miss Rand's work, I can comment on content. Her novels are polemic in nature, being thinly veiled tracts in favor of an extreme *laissez-faire* capitalism. Like many who espouse this particular idea on an emotional basis, Miss Rand displays a complete lack of appreciation for the problems of stability that one is apt to encounter in a multiloop, nonlinear feedback system as complex as a national economy.

The theorists of the Right have long professed a religious faith in the inherent stability of a totally uncontrolled economy, a faith that I doubt many control engineers will uncritically accept. The theorists of the Left, on the other hand, are morally certain that "capitalism contains the seeds of its own destruction." By this, they mean that they believe it to be unstable; and they are correct if they are referring to a totally unrestrained capitalism.

They are entirely wrong, however, when the capitalist economy is tempered by the application of external feedback loops in an attempt to keep it stable. For the past 40 years or so, governments have been doing this with more or less success. The present gyrations of the U.S. economy (perhaps not observed by Green, who is a Canadian) are, in my opinion, partially the result of controls too lightly applied.

Again returning to Green's initial premise; I most wholeheartedly concur that the engineering profession has much to contribute in the social and economic areas. Let us hope that this contribution can be made on a rational basis, and not upon emotional biases that are, in the last analysis, based on a natural revulsion toward the excesses of the Russian revolution. Had the preceding society been one of free, thinking citizens, that revolution need never have taken place. If we can maintain our freedom, keep our economy on an even keel, and keep our reliance on the power of the human intellect, we may yet avoid our revolution!

Charles E. Hendrix
Pacific Palisades, Calif.

Senses of insects

The article by Hsiao and Süsskind in March *IEEE Spectrum* on "Infrared and

Dynamic range: 10^{-99} to 10^{+99}

Automatic special functions:
a^x, Log_{10}/Log_e, SIN/COS,
SIN^{-1}/COS^{-1}, $x!$, $\frac{1}{x}$,
Radians to Degrees,
Single key Σx, Σx^2, N,
Rectangular to Polar Conversion.

Just plug in
and put it to work.

Weight: Twelve pounds.

Monroe men who provide the most complete technical hardware and software support. Because calculators are their only business. And our only business too. That's how we got our name. Monroe. The Calculator Company

A DIVISION OF LITTON INDUSTRIES

550 Central Avenue, Orange, New Jersey 07050

Circle No. 5 on Reader Service Card.

Microwave Communication by Moths" makes reference to the work that I did on this topic in 1960. I am, however, misquoted in that the work was done on the Common Vapourer moth *(Orgvia antiqua)* and not the corn earworm. Furthermore, I was by no means the first to demonstrate the attraction of males to any empty box that had previously contained a female moth. My experiments were primarily directed toward demonstrating the inadequacy of the then-accepted olfactory theory of mate finding, and I also proposed a radiation theory that associated the female-finding phenomenon with the attraction of moths to lights.

The work of Philip Callahan goes much further than anything that I attempted, but it is interesting to note that Callahan, himself a professional entomologist, pays tribute to the foresight of the nonprofessionals in earlier years. In one of his papers[1] he lists the names of those who suggested electromagnetic radiation as the mechanism of assembling; these include Marais (journalist and attorney, 1939), Grant (electrical engineer, 1948), Duane and Taylor (chemists, 1950), and myself (electrical engineer, 1960). The authors may also be interested to know that my paper in *The Entomologist*, to which they refer, was rewritten with an electrical bias for the *Journal of the Institution of Electrical Engineers* and published in 1961.[2]

One comment in that paper that I should like to reiterate here is that it is an interesting speculation as to whether the assembling of male glowworms would also have been "explained" as an olfactory mechanism, had the wavelength of the radiation that they so obviously use been just a little longer, so as to be just outside our visible band.

E. R. Laithwaite
Imperial College of Science and Technology
London, England

REFERENCES

1. Callahan, P. S., "Are arthropods infrared and microwave detectors?" *Proc. North Central Branch ESA*, vol. 20, 1965, pp. 20–31.
2. Laithwaite, E. R., "The assembling of moths—does radiation play a part?" *J. IEE*, vol. 7, 1961, pp. 500–503.

Correction

In Part I of the article "At the Crossroads in Air-Traffic Control" (*IEEE Spectrum*, June 1970), the column headings in Tables I and II (pp. 28 and 29) for *aircraft handled, departures,* and *overs* should read "Number, thousands," not "Number, hundreds of thousands."

COMPUTER

NON-LINEAR SYSTEMS

Fits Nicely Doesn't It ?

THE NLS SERIES S-5 COMPUTER SYSTEM

- **On line solutions in actual engineering units with visual monitor and display of data.**
- **Up to 1000 3-wire inputs for DC, AC, Frequency Resistance, and Ratio.**
- **Service to a number of input peripherals such as counters, A/D converters, digital voltmeters, accumulators, etc.**
- **Data analysis and output recording with X-Y plotters, magnetic tapes, punched cards and tapes, typewriters, line printers, CRT terminals.**
- **Contact-closure outputs for external process control.**
- **Self-testing and de-bug programs with query-type executive routines.**

An S-5 System installed in a jet engine manufacturing plant is producing valuable data on stress and creep in vital engine parts . . .

Another S-5 is acquiring and processing production test data on the performance of precision transducers. This system controls several independent test stations simultaneously . . .

Other S-5's are operating on board ships to measure earth gravity variations for producing accurate surveys of the ocean floor . . .

NLS has been solving the real problems of instrumentation systems for nearly two decades. We have learned many new languages -- acoustics, thermodynamics, traffic control, chemical process control, turbine engine dynamics, naval architecture, telemetry, surveying, communications, statistics, chromatography, stress analysis . . .

Each S-5 system is programmed in the language of its application. No knowledge of computer programming is required.

With the versatile computer-based S-5 system, NLS continues a tradition of adapting each system to its application. In order to meet our customers' diverse and changing requirements -- and to back our claim that no knowledge of computer programming is required -- we have built an organization of highly skilled professional systems engineers and programming specialists, and we have produced the most comprehensive set of programming tools in the measurement systems industry.

FORTRACE is NLS' new real-time monitor for the S-5 system. Incorporating the latest automatic programming technology, FORTRACE enables the process engineer to develop programs in the language of his process. Embedded in the FORTRACE programming system are automatic mechanisms for Process Monitoring, Data Transformation, Performance Computations, Recording and Display and Control. The unique structure of the FORTRACE programming system provides large-scale computing power at lowest cost.

For further information, and comprehensive literature, contact Non-Linear Systems, Inc., P.O. Box N, Del Mar, California 92014. Call 714-755-1134, TWX 910-322-1132.

NLS

NON-LINEAR SYSTEMS, INC.
DEL MAR, CALIFORNIA

Circle No. 7 on Reader Service Card.

Spectral lines

Engineering—an 'open profession.' Most engineers would argue that engineering is an "open" profession; that is, that anyone with intellectual capability and appropriate academic preparation can enter. To a certain extent, this position is defensible, but it does not stand up under close scrutiny.

One measure of the openness of the profession is to examine the relative percentages of various ethnic, racial, or other types of groups and compare them with the overall population percentages. Although accurate quantitative data are not readily available, it becomes immediately clear that there are in some cases order-of-magnitude differences between the overall population percentages of such groups and the percentages of those groups in engineering.

Let's consider some specific examples and assume that the IEEE membership percentages are typical of the engineering profession. The first conclusion from examining the data is that there are not many women engineers. Although it is not always possible to determine the sex of a member from the name—particularly when initials are used—there appears to be a difference of a factor of more than 50 between the percent of women engineers and the percent of women in the overall population. In some countries, such as the Soviet Union, approximately one third of the engineers are women, but in the United States and a number of other countries it is approximately one percent. Thus the field of engineering does not appear to be a very "open" profession to women in many countries.

Let us now turn to a racial grouping and consider the number of engineers in the U.S. from a minority group such as blacks. Although accurate data are very difficult to obtain, it appears that the percentage of blacks in the IEEE in the U.S. is an order of magnitude lower than in the general population. It is probable that this same difference obtains in the case of other minority groups, both in the U.S. and in other countries.

The reasons for these large differences are certainly complex; they relate, in part, to historical facts, social customs, economic differences, and educational opportunities. However, the question is whether or not an organization such as the IEEE can take any action that would make electrical and electronics engineering more accessible as a career to a wider segment of the general population of the country.

The 1969 Board of Directors took cognizance of this matter by having an *ad hoc* committee—Hubert Heffner and Bernard M. Oliver—carry out a study of the situation. This study was received by the Board, which then voted the following resolution:

> "The engineering profession has ideally been open to all, with opportunities for entrance and advancement limited only by ability and training. However, it is clear that educationally disadvantaged individuals and members of some minority groups have found it difficult to gain access to the profession. Therefore, the IEEE Board of Directors will seek means to increase the opportunities for professional education for such individuals in the field of electrical and electronics engineering. The Board also asks all organizational units of the Institute to make widely known the opportunities for a rewarding career in our profession for any member of a minority group with education and ability."

A number of other societies in other professions in the United States have confronted the minority group question with specific action programs. For example, in the legal profession, a very effective combination of deans of law schools and professional societies have, with support from a foundation, developed a scholarship program to encourage educationally disadvantaged minority group members to go to law school. Similarly, architects and urban planners have obtained, via their professional societies, support from the Ford Foundation to enable them to grant fellowships to minority students to study architecture and planning.

The IEEE as a major engineering society with a widespread organizational structure should be able to play an important role in encouraging individuals from the underrepresented segments of society to enter the electrical and electronics profession. In doing so, it would not only benefit the profession but it would also make a most direct contribution to the betterment of society. Effective means of accomplishing this objective can be developed by the various Groups, Sections, Student Branches, Boards, and committees that carry on the work of the Institute. The membership of this Institute has the collective ability to implement the programs needed to make this profession "open" in fact as well as in theory.

F. Karl Willenbrock

Inductorless filters: a survey

I. Electromechanical filters

Although LC filters would appear to be an integral part of modern electronic equipment, in a world of microminiaturization they may be obsolete. The big question is, what will replace them?

George S. Moschytz *Bell Telephone Laboratories, Inc.*

Since their invention early in the 20th century, filters have played an important role in electronic technology. However, the LC filter, which had been the type most widely used in the past, is now being phased out because of the change in design criterions imposed by the current trend toward microminiaturization. The problem of finding a replacement may be solved by implementing some of the inductorless filter methods described in this two-part article. Part I discusses the two basic groups of electromechanical filters—the monolithic crystal and ceramic types, and the mechanical filter, which is coupled with a transducer. Part II of this article, to appear next month, will consider linear active and digital devices.

In 1831 Michael Faraday formulated the law of electromagnetic induction and self-induction. Some 84 years later, in 1915, G. Campbell and K. W. Wagner utilized Faraday's law in their invention of the first electromagnetic or *LC* wave filter. Significant advances in filter theory and technology then followed rapidly,[1,2] until today, filters have so permeated electronic technology that it is hard to conceive of a modern world without them. Consumer, industrial, and military electronic systems all require some kind of signal filter, and, in the past, *LC* filter networks provided one of the most efficient and economical methods of implementing them.

In spite of the vast resources that have gone into the perfection of *LC* filter theory, technology, and manufacture, and the numerous design tools that have evolved (including easily accessible tables, charts, and, more recently, computer programs), there is a tendency to eliminate *LC* filters from modern electronic equipment wherever possible, because integrated circuits have completely changed the conventional methods and performance criterions previously accepted in electronic designs. Therefore, filter techniques, like all other circuit techniques that do not fit into the new world of microminiaturization, will ultimately be replaced by techniques that do.

The big question is what will replace *LC* filters. The fact that there are so many contenders for this role shows how difficult it is to find an electronically equivalent replacement that is comparable in cost. The answer, which can be only surmised today, very likely will encompass some, if not all, of the approaches to inductorless filter design described in this article.

Electromagnetic filters utilize the energy-storage properties of inductors and capacitors to realize frequency selectivity. Some of the main advantages of these networks, in addition to those already mentioned, are that they (1) dissipate negligible power (passive networks), (2) are unconditionally stable, (3) require tolerances on network components that are of the same order of magnitude as the tolerances of the resulting network response because of low component sensitivities, (4) generate practically no noise, and (5) provide a dc path or total dc isolation, as desired, with no offset voltage. The disadvantages of *LC* filters, as related to the criterions imposed by integrated circuit technology, can be summarized as: (1) bulkiness (weight and volume), (2) low coil Q, resulting in limited circuit Q, and (3) fabrication methods that are incompatible with IC batch-processing techniques. These disadvantages are directly related to the properties of inductors themselves. High-Q wire-wound inductors are inherently bulky since the volume necessary to store a given amount of magnetic energy is much larger than that required by capacitors for the storage of the same amount of electric energy.

As the frequency decreases, the size and cost of inductors increase objectionably, which is one reason for the excessive space occupied by filters in many low-frequency and voice-band communication systems. As more and more of the remaining circuits in these systems, and the digital circuits in particular, are microminiaturized through the use of semiconductor integrated devices, the discrepancy between filter size in comparison with the overall equipment size is becoming increasingly objectionable. Because of this widespread discrepancy in so many electronic systems, much effort has been devoted to reducing the size of the filters. The seemingly obvious approach is to microminiaturize the inductors. However, reducing inductor size to an extent comparable with other miniaturized components results in a drastic reduction in Q,[3] thus decreasing the already limited available network Q.* Furthermore, whereas resistors and capacitors can be realized in the quasi two-dimensional form well suited to integrated circuit implementation, this is not generally true for inductors. For any reasonable performance, at least at frequencies below several megahertz, inductors require a volume that is incompatible with present-day microcircuits.

* When a capacitor is miniaturized by scaling down all of its dimensions while holding its material properties constant, its Q (ratio of susceptance to conductance) remains constant at a given frequency. When an inductor is similarly miniaturized, its Q (ratio of reactance to resistance) decreases as the square of the scaling factor.

FIGURE 1. Generalized electromechanical filter.

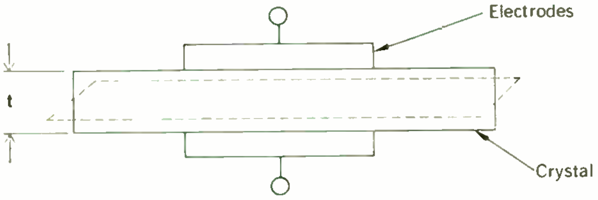


FIGURE 2. Transverse thickness shear wave of AT-cut crystal slice.

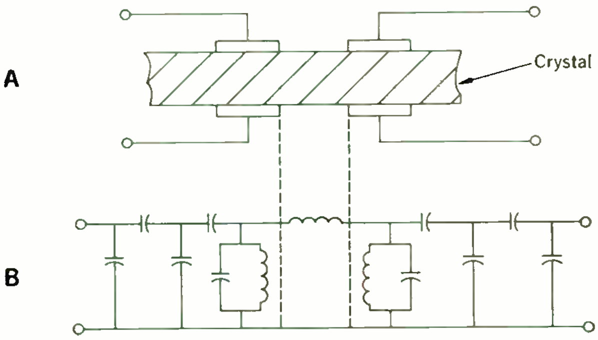


FIGURE 3. Monolithic crystal filter in simplest form (A) and equivalent circuit (B).

It is not surprising that any designer of frequency-selective networks or filters intended for microcircuit implementation or for low-frequency applications will delete, a priori, all electromagnetic items (transformers, inductors, etc.) from his list of possible components. This eliminates all components characterized by two of Maxwell's four equations governing electromagnetic phenomenons and necessitates the inclusion of other components with analogous resonance properties that can be made, with the help of suitable transducers, to interact with electric signals and networks. Certain electromechanical devices that combine mechanical resonators with energy transducers based on the piezoelectric, the magnetostrictive, or the electrostrictive effect come closest to meeting this requirement. Such devices have been used for decades to expand rather than replace the capabilities of *LC* filters. Perhaps the most common replacements for *LC* filters, however, are not those based directly on material phenomenons that exhibit physical characteristics analogous to *LC* filter combinations, but those based on electronic circuits that provide the desired network response *functionally* by applying appropriate analog or digital circuit techniques.

The frequency-sensitive element of the generalized electromechanical filter shown in Fig. 1 is a mechanical transmission device in which mass or moment of inertia and elastic compliance or stiffness interact in resonance at a particular frequency. A direct analogy to electromagnetic resonance can be found in that there is a correspondence of mass to inductance, stiffness to capacitance, and mechanical resonance to electrical resonance. The conversion of the energy within the mechanical resonator into electric energy and the coupling of the converted electric signals to an electric network are obtained by means of an electromechanical transducer, which generally is based on the piezoelectric, magnetostrictive, or electrostrictive effect.

Broadly speaking, there are two basic categories of electromechanical filters: those that provide mechanical resonance properties and the capability for energy conversion in one and the same device, and those that combine two separate materials or devices to perform these functions. The most important groups in the former category are monolithic crystal and ceramic filters; in the latter, it is the group generally called mechanical filters.

Monolithic crystal filters

High-frequency bandpass filters having bandwidths of 0.1 percent or less and high stopband discrimination have traditionally been realized by so-called *crystal filters*, which combine quartz crystals with balanced transformers, inductors, and capacitors.[4-6] Since crystal filters clearly do not fall into the category of inductorless filters, they will not be discussed in detail here. Also, their limited use generally leads to very high costs and their method of implementation to large size and inefficient design. In order to reduce the size and cost without sacrificing performance, a great deal of effort was expended in the last decade to retain the advantages provided by crystals while eliminating the additional electromagnetic components required. This meant, in effect, integrating piezoelectric resonators into monolithic filters.

About five years ago, after the network properties of piezoelectric elements were better understood and as the effects of electrodes on the vibrational properties of piezoelectric devices in the high-frequency range became known, the first complete monolithic bandpass filters were successfully developed.[7] Since that time, monolithic crystal filters have made numerous inroads into the applications and frequency ranges that had previously been considered far beyond their capabilities.[8-10] One reason for their increasingly widespread use is the fact that new manufacturing techniques already have reduced their cost by a half; another is that the resulting filters may be as much as two orders of magnitude smaller than their conventional counterparts.

Briefly, monolithic crystal filters use coupled mechanical vibrations in a piezoelectric material to provide bandpass-type filter functions. A filter consists of a crystalline quartz wafer onto which pairs of metal electrodes are deposited. The operation of the filter is made possible by two factors:

1. The crystal is piezoelectric. It can transfer electric energy into a mechanical form, specifically, into a transverse shear wave (see Fig. 2), and back again. Therefore, in addition to serving as an interresonator coupling medium, it performs the transducing functions.

2. The metal electrodes lower the resonance frequency of the transverse shear wave in the plated region as compared with unplated quartz. As a result, the resonance created in the plated region does not propagate into the areas without electrodes, but remains trapped under the electrodes, which are thin metal films.

In Fig. 3(A), a monolithic filter is shown in its simplest

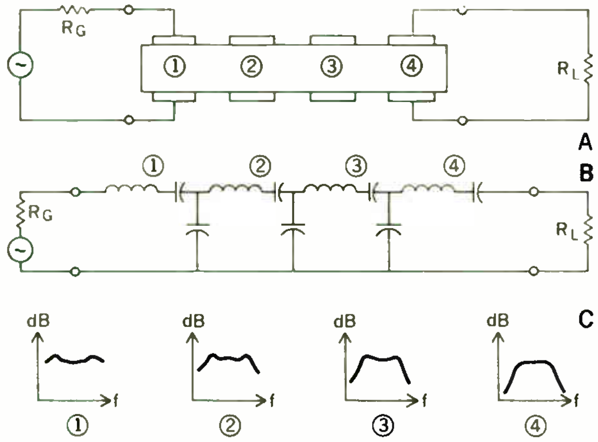


FIGURE 4. A four-resonator filter (A), its equivalent circuit diagram (B), and the frequency-response curves after successive resonators (C).

form. It consists of an input and an output resonator formed by a pair of electrodes deposited onto opposite faces of a quartz crystalline wafer. Vibration, induced by resonance between the plates, decays very rapidly outside the plated region; therefore, it is essentially trapped under the electrodes. Although the vibration is confined to the electrode area, the displacement decays exponentially in the surrounding region and this, mechanically or acoustically, couples adjacent resonators. The dimensions and separation of the resonators determine the coupling, bandwidth, transmission characteristics, and terminal impedance. The equivalent electrical network is shown in Fig. 3(B).

The amount of coupling between adjacent resonators may be controlled within limits since it depends upon the dimensions of the resonators, the thickness of the metal electrodes, and the spacing between resonant regions. In AT-cut crystals, for example, the range of coupling coefficients permits the realization of passbands ranging from about 0.001 percent to about 0.2 percent of the center frequency, if the fundamental mode of the thickness shear vibration is used. (For the large variety of shapes and orientations commonly used for crystals; see, for example, Ref. 11.) The techniques available to shape the wafer permit center frequencies from about 5 MHz to approximately 150 MHz.

In general, the important factors in the monolithic filter are the number of resonant elements, the coefficient of coupling between adjacent resonant elements, and the impedance of the resonance elements. More resonant elements result in a higher rate of cutoff—that is, a more rapid increase of attenuation as a function of frequency in the frequency range adjacent to the flat passband. This is shown qualitatively for a four-resonator filter in Fig. 4. A cross section of the filter with its electrode arrangement is shown in Fig. 4(A), and the equivalent circuit diagram in Fig. 4(B). The frequency-response curves in Fig. 4(C) show the variation of the current with frequency as the input voltage is held at one volt. The improvement in selectivity as the current passes through the resonators is apparent from the increased steepness of slope at the band edges. The "bumps" inside the passband region indicate reflections in the filter. This reflective interaction eventually leads to the desired flat response in the band by the time the current passes the last resonator.

Table I presents a summary of the salient features of monolithic crystal filters as they are available today. The advantages of these filters from the standpoint of *Q*

I. Characteristics of electromechanical filters*

Type	Frequency Range	Pole Q	$\frac{\Delta f}{f}$† ppm/°C	$\frac{\Delta Q}{Q}$† %	Functional Versatility	Functional Accuracy	Tunability	Signal Dynamic Range, dB	Compatibility with HIC Technology
Monolithic crystal	5–150 MHz	1000–250 000	±1	0.1	Fair (bandpass and frequency-rejection networks)	Good (with exception of spurious modes)	Initial tuning good; system adjustment poor	40–80 (depending on proximity to spurious tones)	Good
Ceramic	0.1–10 MHz	30–1500	±100	0.2	Fair (requires additional components for most functions)	Good (with exception of spurious modes)	Initial tuning good; system adjustment poor	40–80 (depending on proximity to spurious tones)	Fair
Mechanical	0.1 Hz–20 kHz	50–5000	±50	0.1	Fair (bandpass and frequency rejection; numerous nonfilter functions possible)	Good	Good	60–80	Poor

* The characteristics listed may be mutually exclusive.
† Frequency range over temperature range from 0° to 60°C and aging.

capability and frequency stability are apparent. In fact, there is no comparable technique to match this performance. However, the frequencies are, and presumably will continue to be, limited to the megahertz range. One possible method of using monolithic crystal filters for low-frequency applications is to heterodyne the low-frequency signal into the crystal frequency range and then to demodulate the filtered signal back to the low frequencies.

Functionally, the versatility of monolithic crystal filters is limited. They are inherently all-pole (more specifically, bandpass) filters with the number of transmission poles corresponding to the number of electrode pairs. As such, they permit the realization of any standard bandpass function (Chebyshev, Butterworth, etc.). Other filter functions (band elimination, transmission zeros, low pass, high pass, etc.) can be realized only with additional components, although a degree of functional variability can be obtained by splitting electrodes. In order to reduce the number of inductors required as additional components, active circuits have been used instead.[12]

Initial filter adjustment and tuning can be carried out very accurately (e.g., 0.1-dB ripple in the passband of, say, an eight-pole bandpass filter can be reproduced consistently). However, because of unwanted vibrations within the quartz plate, invariably there are additional passbands at higher frequencies. The amount of interference with the required frequency response will depend on how close these spurious resonant modes are to the frequency band of interest. The magnitude of the unwanted modes may be reduced to some extent by adjusting the areas of the various resonators appropriately. However, this in turn may degrade the Q of the useful main shear mode and effect the desired out-of-band frequency rejection of the filter.

The mechanical or acoustic coupling, which determines the bandwidth of the filter, depends principally on the electrode separation and, to a lesser extent, on the electrode mass and length. The bandwidth is controlled by electrode separation, mass, and area. It can be adjusted by placing stripes between the electrodes and either adding to these stripes by evaporation or removing them by laser. The center frequency, determined principally by the plate thickness, is fine-tuned by the mass of electrodes. These operations can be tightly controlled in manufacture. However, once the filter is assembled into a system, it is very difficult to make any additional adjustments. Furthermore, since the entire resonator structure is coupled, it is impossible to measure the uncoupled frequency of any one resonator. Adjustments of resonators, therefore, must take this coupling into account, with the help of theoretical expressions derived for this purpose.

Compared with their conventional counterparts, monolithic crystal filters permit considerable size reduction since no transformers, inductors, or other discrete components are needed. Further size reduction and compatibility with hybrid IC techniques also appear possible since thin-film circuitry for passive components and beam-lead semiconductors is now at a stage where these probably can be deposited on the same substrate as a monolithic crystal filter. In addition, once the necessary fixtures, masks, design data, etc., are obtained, filter costs become quite low, because only one quartz plate and one enclosure are needed for each filter.

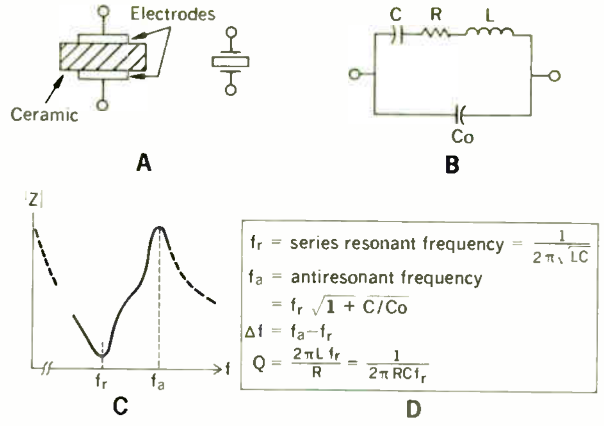


FIGURE 5. Basic two-electrode resonator. A—Cross section and circuit symbol. B—Equivalent circuit. C—Impedance frequency characteristics. D—Design equations.

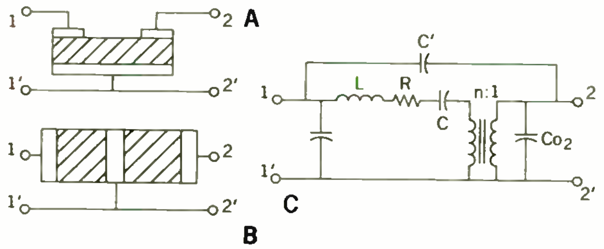


FIGURE 6. Three-electrode resonators. A—Basic form. B—Composite three-electrode resonator obtained by bonding together two single resonators. C—Equivalent circuit.

Typical applications for monolithic crystal filters are in telephone carrier systems (frequency range, 2.6–19 MHz), point-to-point wire and radio transmission systems with single-sideband, double-sideband, or narrow-band frequency modulation, and broadband telephone multiplex systems. All of these systems, in addition to being in the megahertz frequency range, have stringent requirements on Q and frequency stability that are hard to meet economically by other methods.

Ceramic filters

Two major limitations of *monolithic crystal* filters are their restriction to frequencies in the megahertz range and their narrow bandwidths (i.e., high Qs). Both of these limitations can be overcome partially by *ceramic* filters[13-15] because of the difference in material properties (principally the piezoelectric coupling coefficient) between ceramic and crystal filters rather than because of any basic difference in concept. Thus, ceramic filters combine the functions of the mechanical resonator and the electromechanical transducer shown in Fig. 1 due to their combined mechanical resonance and piezoelectric* properties in the same way that monolithic crystal filters do.

In contrast to monolithic crystal filters, ceramic filters are now used more in combinations of individual two- or three-electrode resonators of the kind shown in Figs. 5

* Clearly, we refer here only to those high-Q piezoelectric ceramic materials that are used in ceramic filters.

and 6 than as monolithic structures, although the latter are being considered. Two-electrode resonators may be interconnected with capacitors and/or amplifiers to provide high-order filter configurations. A building-block approach to ceramic filter design[16] in which identical two-electrode resonators are combined with capacitors in a ladder configuration, as shown in Fig. 7, has been suggested. By selecting appropriate capacitor combinations instead of a variety of different resonators, a wide range of frequency responses can be obtained. Composite filters using three-electrode resonators also behave essentially like ladder structures consisting of cascades of resonators coupled either directly or by series or shunt capacitors. Figure 8 shows a two-resonator filter network.

As with crystal filters, efforts have been made, when additional components are necessary to realize a desired network function, to avoid inductors and to use active auxiliary networks instead. In this way it is hoped that ceramic filters can be incorporated into or combined with hybrid integrated circuits. An example of a resonator–operational amplifier (with differential output) combination is shown in Fig. 9(A). The transfer function of this circuit is given by

$$T(s) = \frac{e_o}{e_{in}} = -\frac{R_F}{R_G}\,\frac{Z_b - Z_a}{0.75[R^2 + R(Z_a + Z_b) + Z_aZ_b]^{1/2}} \tag{1}$$

This function corresponds to a hybrid lattice filter of the kind shown in Fig. 9(B), except that it is realized with an operational amplifier instead of a transformer. The frequency response that corresponds to the selectivity of a critically coupled double-tuned LC circuit is shown in Fig. 9(C).

The characteristics of ceramic filters, as determined by the ceramic resonators themselves, are summarized in Table I. The lower frequency range and wider bandwidths (lower Q values) are apparent. In particular, the frequency range covered includes the intermediate frequencies (IF) of AM radios (455 kHz), audio television reception (4.5 MHz), and FM receivers (10.7 MHz). It is for this consumer market that ceramic filters hold the most promise, both as to performance and cost. Ceramic resonators are not newly arrived on the scene, of course; millions of them have been operating in military and commercial equipment for years without failure or deterioration. However, new material resonator and network approaches, as well as automated mass-production techniques, are being applied to make the use of ceramic filters in broadcast receivers and other high-volume applications economically feasible. In fact, the mass production of ceramic IF filters is inherently simpler than that of wire-wound LC circuits and ultimately should result in reduced prices for IF circuitry.

The electrical resonator parameters may be varied over a range depending upon the ceramic material, processing, and resonator geometry. The quantities $\Delta f/f_r$ and C_0 (see Fig. 5) may be adjusted over a wide range by the manufacturing process. This makes the ceramic resonator a more versatile electric circuit element than, for example, the quartz resonator, whose equivalent circuit is the same as that of Fig. 5 but whose $\Delta f/f_r$ is a material constant of small and fixed value.

Compared with the average consumer-type IF transformer, ceramic materials and filters have a better frequency–temperature stability and quality factor Q, the latter by an order of magnitude. A typical radial resonator with f_r = 455 kHz has a diameter of 0.56 cm and a thickness of 0.038 cm. All equivalent circuits of the piezoelectric resonator are valid only in the vicinity of the operating frequency. For radial resonators, the circuit equivalent of Fig. 5 is accurate for frequencies up to approximately $1.5f_r$. At higher frequencies, other resonances (spurious modes) occur due to overtones of the

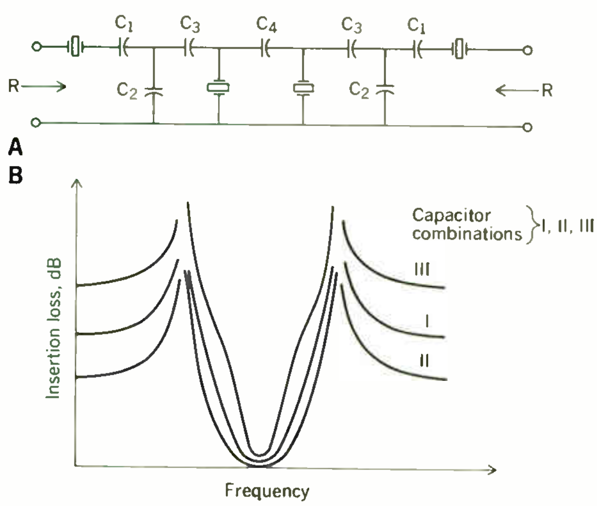


FIGURE 7. Four-resonator network with identical resonators and four different capacitors. A—Circuit diagram. B—Frequency response.

FIGURE 8. Two-resonator version of a composite filter using three-electrode resonators.

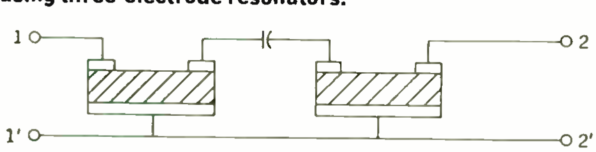
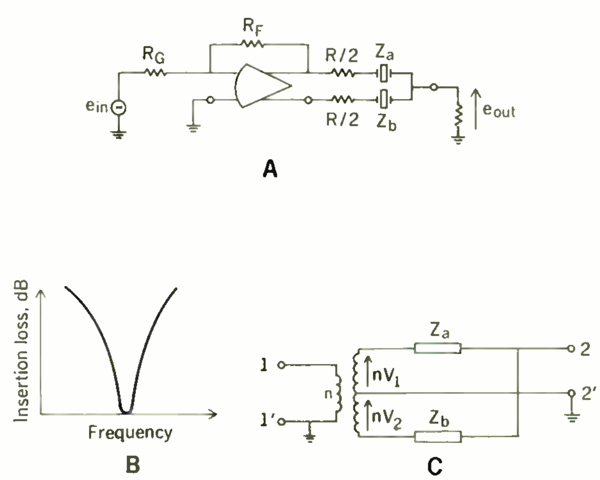


FIGURE 9. Hybrid lattice filter using ceramic resonators. A—Active realization. B—Passive realization. C—Frequency response.



radial mode and to other vibrational modes.

Spurious responses within the individual resonator may be suppressed by optimizing both the processing and the geometry of the resonator. However, some of these approaches are not compatible with economy. For instance, one known method completely eliminates all radial overtones, but results in a more costly resonator and a bulkier, less rugged package. On the other hand, complete elimination of one radial overtone, or partial suppression of two neighboring radial overtones, may be achieved by controlling the resonator geometry, without sacrificing economy, size, or ruggedness.

Although ceramic filters, in their present form, cannot be integrated monolithically, hybrid integrated circuits have been made using thickness-mode ceramic IF filters. Also, radial-mode resonators become small enough at higher frequencies (> 1 MHz) to be included in IC packages. They are not economically competitive with conventional IF transformers at present; however, some experimental circuits are being developed using this technique, and pilot quantities are being evaluated for high-volume commercial systems.

Mechanical filters

In addition to motivating the development of ceramic filters, the need for small and low-cost selective IF filters has also pushed the development of mechanical filters.[17-19] Furthermore, during the last decade work has been carried out in the areas of low-frequency applications and of applications requiring a large number of low-loss coupled resonant circuits to reduce equipment size with mechanical filters.[20-22]

Mechanical filters generally consist of a mechanical resonator, and rely on the electrostrictive, magnetostrictive, or piezoelectric effect of a separate transducer for the interaction between electric and mechanical energy, particularly for a direct relation between electrical and mechanical resonances. Consider, for example, the H-shaped resonator[23] shown in Fig. 10. It consists of two balanced masses (i.e., the bars of the H) connected by a flexible web. As the piezoelectric input transducer contracts and expands, the web flexes and the bars oscillate in opposite directions. The resonance signal can be picked off by a coil near a bar or by another piezoelectric transducer underneath the web. Since the bars rotate counter to each other at their nodal points (i.e., points of least deflection), there is virtually no net transmission of energy through the common plane of the base. Thus, in contrast with most other mechanical resonators, the resonant energy loss that stems from vibrations transmitted by the resonator to its mounting base is avoided. Since the Q of a mechanical resonator gives the ratio of conserved energy to dissipated energy, very high Q values can be obtained.

Another common problem of low-frequency mechanical resonators—their susceptibility to external shock and vibrations—is also greatly reduced. In the same way that transmission of vibrations from the resonator to the mount is prevented by the symmetrical push–pull configuration of the H-shaped resonator, external shocks are prevented from traveling from the mount to the resonator. This process is similar to the rejection of a common-mode signal at the input to a differential amplifier in that any vibrations originating at the base are common to both bars of the resonator and therefore cancel out.

The main characteristics of the H-resonator are summarized in Table I. As shown, this representative mechanical filter is a low-frequency device. The resonant frequencies range from a fraction of a hertz to 20 kHz; Q values can vary anywhere between 50 and 5000. The response of a resonator with a Q of 4000 at 318 Hz is shown in Fig. 11. If appropriate materials are used, dimensional variations due to temperature changes have little or no effect on the balance of the device, and its frequency stability can be very tightly controlled. The resonant frequency can be shifted 20 to 30 percent by balanced pairs of threaded tuning slugs inserted in threaded holes running the entire length of each bar. Low values of Q are obtained by setting the slugs at differing distances from the nodes of the bars. Consequently, this particular mechanical resonator can be accurately tuned initially as well as readjusted after incorporation into a system.

The H-resonator is unconventional as far as mechanical filters go in that it does not use electrostrictive or magnetostrictive transducers; instead, it uses a piezoelectric transducer. This is possible because of its geometrical configuration. Consequently, it can deliver a substantial electric signal into a low-impedance load even though piezoelectric transducers are inherently high-impedance devices. In fact, if the transducer and load are properly adjusted, a voltage gain of as much as two to one can be obtained. The dynamic range of the H-resonator, which is determined by the signal capabilities of the piezoelectric transducer, is large. The output voltage, for example, can be maintained high enough to drive the gate

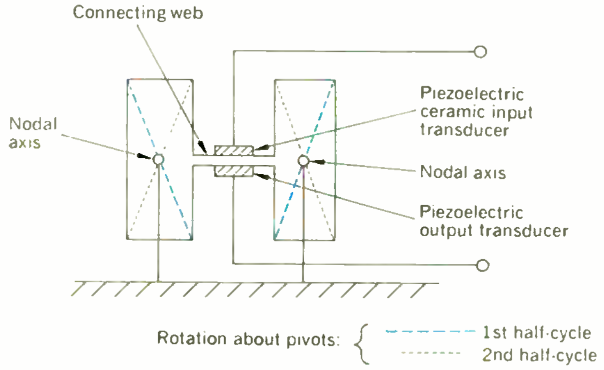


FIGURE 10 (left). H-shaped mechanical filter.

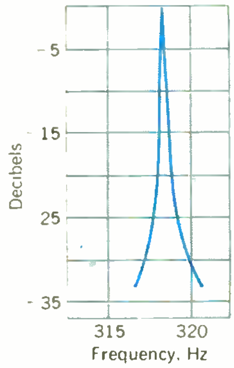


FIGURE 11 (right). Frequency response of mechanical H-resonator: f_c — 318.36 Hz; bandwidth = 0.081 Hz at —3 dB, 1Hz at —20 dB; Q = 4000; insertion loss = —18 dB at 5 volts.

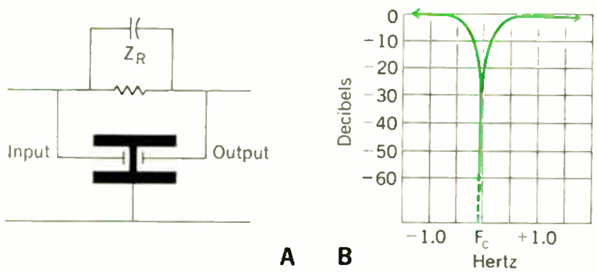


FIGURE 12. Frequency-rejection network using the H-shaped resonator. A—Circuit, showing dual transducer resonator. B—Frequency response: f_c = 60 Hz; bandwidth = 0.56 Hz at −3 dB.

of a silicon-controlled rectifier directly. As with ceramic and monolithic circuits, multiple combinations of inputs and outputs can be achieved by coupling a number of separate transducers to the web of the H-shaped resonator. Since piezoelectric ceramic transducer material is polarized, transducer pairs can be placed so that their responses are in phase or 180 degrees out of phase with each other. By suitably combining in- and out-of-phase transducers, a variety of functions—such as oscillators, FM discriminators, and bandpass and frequency-reject filters—can be obtained. A 60-Hz frequency-rejection filter using the H-resonator with a transducer pair is shown in Fig. 12(A); its frequency response is shown in Fig. 12(B).

The functional versatility of mechanical filters, just as with other electromechanical filters, is quite limited. Inherently these filters are resonators providing bandpass or frequency-rejection characteristics. Thus, to obtain, say, low-pass, high-pass, or elliptic filters with finite transmission zeros, auxiliary circuits, which may be active *RC* in order to avoid inductors, must be used.

Although the H-resonator and other mechanical filters developed recently may be considerably smaller than equivalent *LC* circuits, particularly at low frequencies, they are still far from directly compatible with integrated circuit techniques. However, work has gone on in the last few years to overcome this limitation and reports of mechanical resonators fabricated by batch processes compatible with integrated circuits have been reported.[24-27] One of these, the resonant gate transistor (RGT), is a surface field-effect transistor in which the frequency-determining element is a tiny gold cantilever beam suspended over the region connecting the source to the drain. The mechanical resonance of the cantilever is coupled electrostatically to the source-drain region modulating the electric field on the silicon surface from which it is air-isolated, much as the gate modulates the electric field in a conventional insulated-gate field-effect transistor. The mechanical response of the beam establishes the bandpass properties of the device.

More recently, an improved version of the RGT, the Tunistor, has been reported.[28] This device was designed specifically for integrated circuit implementation and is intended to overcome the limitations of the former. Whereas the RGT operates only up to 50 kHz with a frequency stability of 300 ppm/°C and requires an electrostatic polarization voltage of 20 to 70 volts, the Tunistor has been demonstrated to operate up to 500 kHz with *Q* values of several thousand, a frequency stability of 40 ppm/°C, and no need of an electrostatic polarization voltage. It is expected that frequencies up to several megahertz will be feasible with this device.

The Tunistor differs from the RGT fundamentally in two ways. First, in its monolithic form, the resonator of the Tunistor is formed from silicon instead of gold. This change results in the greatly improved frequency stability mentioned above. It also avoids the problems associated with electroplating the nickel spacer and the gold cantilever required in the RGT and permits the use of a resonator geometry, which extends the frequency range upward. Second, the electrostatic transducers of the resonant gate transistor are replaced by piezoelectric film transducers. This change eliminates the need for the polarization voltage of the RGT and avoids the associated effects on gain and frequency, the need to control cantilever-to-substrate spacing, and the noise resulting from the exposed FET.

Large nonintegrated replicas of the experimental integrated Tunistors, which are fabricated from stainless steel instead of silicon, were also built. Where the resonant frequencies of the silicon devices lie somewhere between 80 and 500 kHz with *Q* ranging between 60 and 1400, the frequency range of the former could be used down to several hundred hertz with *Q* ranging between 150 and 600. Clearly, the Tunistor is no solution to the problem of finding an integrated mechanical filter operating at voice frequencies, since the integrated version is restricted to high frequencies. As with the ceramic and monolithic filters, however, the nonintegrated Tunistor may be compatible with hybrid integrated circuit technology.

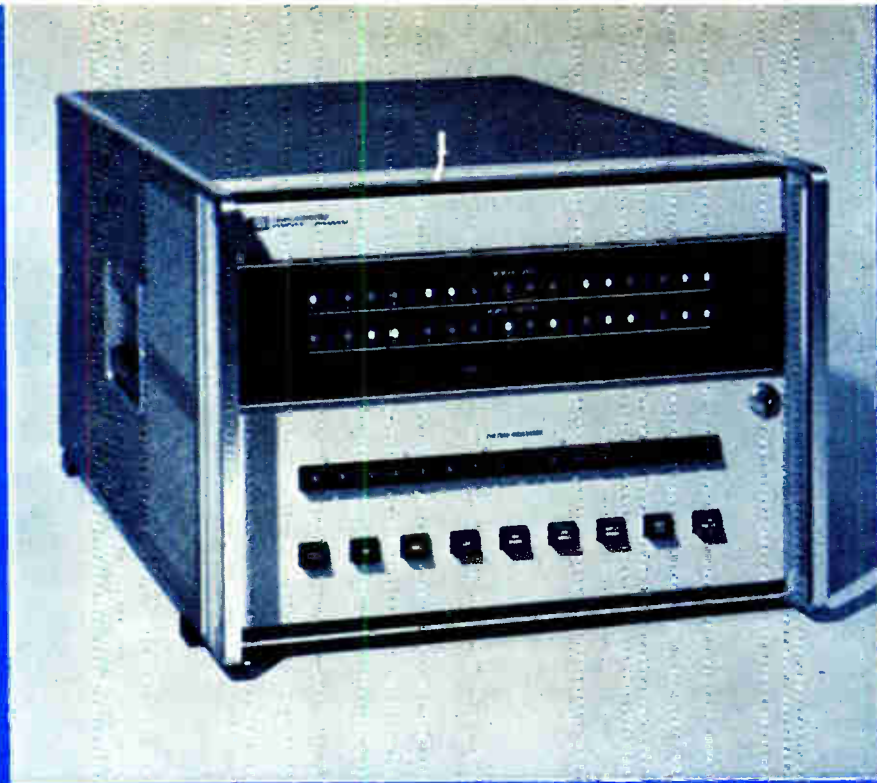
Although the Tunistor represents a very novel approach to the integrated filter problem, much development work remains to be done before these devices will be marketable. In particular, the technology for maintaining adequate tolerances in production must be established. Also, circuit techniques must be found to exploit its advantages (i.e., small size, high *Q*, etc.) and to compensate for its disadvantages (for instance, use of thermal stabilization of the silicon chip to minimize the temperature coefficient artificially). Nevertheless, it would appear that mechanical silicon resonators are feasible, at least at high frequencies. Silicon is attractive as the basic material not only because it is used in IC fabrication in general, but also because as a mechanical resonator it has high-*Q* properties and a relatively low temperature coefficient of frequency. No way has yet been devised to utilize materials with lower temperature coefficients (e.g., crystaline quartz, nickel–iron alloys) in a monolithic IC, although hybrid devices have been used. Since the insertion loss for the miniaturized resonators tends to be high, additional gain stages are generally necessary to provide acceptable signal levels.

The complete article represents essentially the full text of a paper presented at the 1970 Electronic Components Conference, Washington, D.C., May 13–15, and published in the Proceedings of that conference (No. 70 C 12-PMP).

REFERENCES

The complete list of references will appear with Part II of this article.

George S. Moschytz' biography appeared in the January issue of IEEE Spectrum, page 50.



HEWLETT-PACKARD'S 2114B minicomputer is a 16-bit machine.

An IEEE SPECTRUM applications report

Minicomputer applications in the seventies

Small, programmable digital computers are taking over many of the tasks that have been handled in the past by hard-wired logic systems or by large, expensive computers. The strength of the minicomputer is in the fact that it is the cheapest form of digital logic system that you can buy today

Ronald K. Jurgen Managing Editor

The purpose of this report is to give the reader, by means of specific application examples, an overview of the extreme versatility of the minicomputer. The basic system configurations that will be described may be adapted, by analogy, to many other tasks. In this manner we hope to stimulate the reader to think about ways in which a minicomputer might be able to do a job for him at less cost, in less time, in a more thorough manner, or more reliably than that job could be done—if it could be done at all—without the minicomputer.

It is beyond the intended scope of this report to discuss minicomputer architecture, to describe in detail commercially available minicomputers, or to attempt to offer detailed guidelines for the selection of a specific minicomputer model for a system.

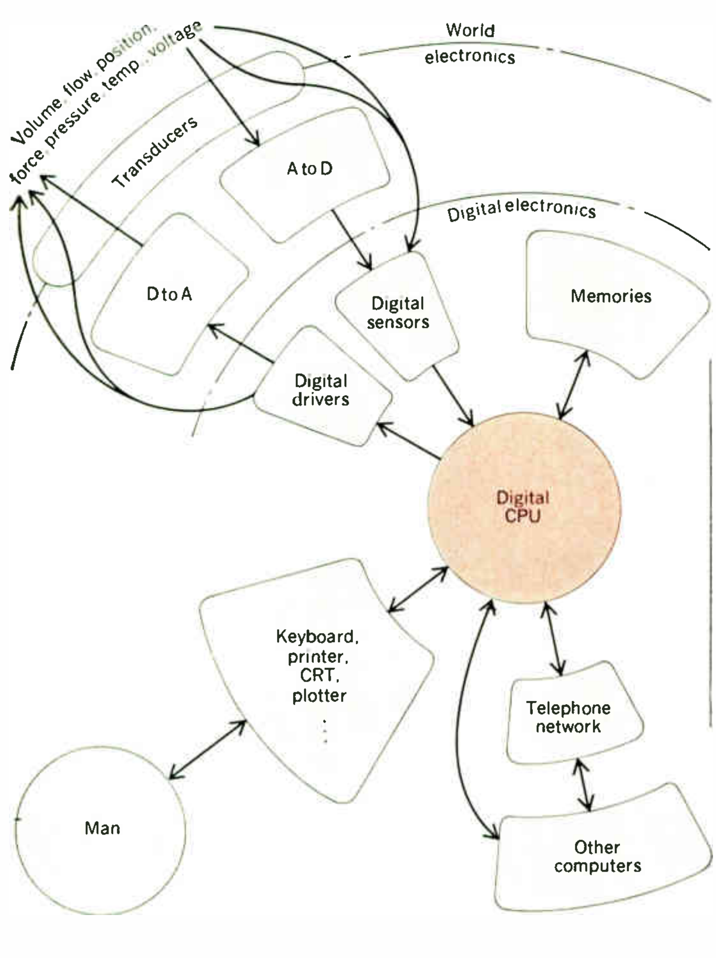
Using as an indicator the publicity given to the minicomputer, one would think that it is a revolutionary new device. It is not. It is the result of engineering know-how in using commercially available integrated circuits to package a physically small computer at low cost.

The term "minicomputer" is catchy but misleading. The "mini" portion of the term is generally appropriate when referring to physical size and cost—and possibly word length and memory size—but not when one is considering computer power. Today's small computers outperform many of yesterday's large computers.

Setting limits on the definition of a minicomputer is an elusive task. A minicomputer is often defined, for example, as a general-purpose, programmable digital computer, small in size, that sells in its basic form for under \$20 000. At the low end of the price range (under \$7500), however, the minicomputer is apt to be a dedicated or single-purpose computer—a process controller, for example—rather than a general-purpose device. At the upper price limit, the \$20 000 selling price is subject to debate. Perhaps the figure should be \$25 000, \$30 000, or even \$40 000. Or should it be less than \$10 000? It depends on your point of view.

For purposes of this report we have not attempted to apply strict limits to the cost of a minicomputer. The significant fact is that there are available commercially small digital computers, relatively low in cost, that can do a remarkable variety of jobs extremely well.

FIGURE 1. A block diagram showing how a digital computer interfaces with man and the real world.



The actual cost of the basic computer is not nearly as important as the cost of the total system in which that computer will be used. Most important of all is whether or not that total minicomputer system cost can be justified for the job that is to be accomplished.

The minicomputer industry is a fast-growing one. The first minicomputer was introduced only about five years ago. Since then—according to Reg A. Kaenel of Bell Telephone Laboratories who chaired a minicomputer session at the recent Spring Joint Computer Conference—more than 10 000 minicomputer installations have been made. The industry has grown, he says, to the point where there are about 100 different minicomputer models available from 62 different manufacturers. For example, at Bell Telephone Laboratories there are 120 minicomputers in use; they consist of 34 different models that have been supplied by 12 different manufacturers.

The proliferation of minicomputer models with varying characteristics can be confusing to a prospective user. An important point to remember, however, is that the specific application of a minicomputer is what determines the value to the user of any one of its characteristics.

Before taking a look at what comprises a typical minicomputer, let us first see how any computer interfaces with man and the real world. Professor Thomas F. Piatkowski of the Thayer School of Engineering at Dartmouth College introduced a recent minicomputer seminar at that college by stating that one can view the computer as a component in the hierarchy of information-transfer devices, as indicated in Fig. 1. He described the illustration as follows: The keypunches, keyboards, printers, plotters, Teletypes, and telephone networks that operate with digital computers are essentially digital in nature. This digital communication is interactive with man typically through printed characters. The central processing unit (CPU), or mainframe, does all of the logical and arithmetical processing associated with the computer. Ordinarily, it comes with a collection of memories—from very fast core memories to slow magnetic and paper tapes. Here, too, all of the devices are essentially digital.

In order to use the computer as a system component to interact with an outside world that is typically nondigital, one must convert the external parameters into equivalent digital signals and also convert digital signals into variables usable in the outside world. In practice, these conversions are accomplished in a number of ways.

Moving from the outside world toward the computer, it might be that the input to the digital sensor can be obtained directly in the outside world. Typically, however, one is dealing with parameters such as volume, flow, position, and force, which are not even electrical in nature. These parameters may be converted to electric analog signals by using analog transducers and then be converted to digital signals. Thus, one can go step-wise through conversion from, say, flow rate to proportional analog voltage to proportional digital signal to digital sensor to computer input. Going in the reverse direction, the computer can influence the outside world by generating digital data that are sent to a driver that has the digital signal converted to analog through a digital-to-analog converter. The analog signal may then go through a transducer and be converted into force or temperature, and so on.

Figure 2 is a block diagram of a simple minicomputer.

It was used by Eric M. Aupperle of the University of Michigan in his role as one of the faculty at the recent excellent seminar on minicomputers sponsored by the National Electronics Conference as part of its continuing series of Professional Growth in Electronics Seminars. Mr. Aupperle explained the system components in this manner: The computer control unit (CCU) in Fig. 2 coordinates all of the other parts of the computer to insure that the logical sequence of operations will be carried out correctly and at exactly the right time. The CCU receives its instructions (the stored program provided by the programmer) from memory via the memory control unit. These instructions are interpreted to produce the specific logical sequence required by each program.

The memory modules usually consist of thousands of ferromagnetic cores. For each core one bit or binary digit can be obtained. It is common to deal with a small collection of bits referred to as the computer's memory word. Usually one or more memory modules are purchased for any given minicomputer application. Slower-speed mass-storage devices such as magnetic tape, disk, or drum may also be used. They are not included in Fig. 2 since they are peripherals rather than part of the minicomputer mainframe. Read-only memories are sometimes included as part of the computer main memory or as an addition to or a substitute for the computer's read/write memory.

The memory control unit serves as the master index for the insertion and retrieval of information from the memory modules. Under direction of the CCU it is able to route memory words to and from all of the other units at very high speeds. The time it takes the memory words to go through the memory control unit is called the memory cycle time. The memory control unit is also able to function independently of the CCU when data are transferred at high speed through the direct memory access unit or when several minicomputers share a common memory.

The arithmetic unit accepts data previously stored in the memory or newly provided and performs various algorithmic operations on these data. Results of these arithmetic operations either may be returned to the computer memory or transferred elsewhere, perhaps to the input/output (I/O) unit. The arithmetic unit contains mainly flip-flop registers and gating circuits to provide both temporary storage and complex logical switching.

Most peripheral devices are connected to the I/O bus, an extension of the I/O unit. Through this bus and unit flow commands or output data from the minicomputer to its attached devices (not shown in Fig. 2). In the other direction flows device status information or input data. This combined information flow may be controlled either by the CCU or by individual devices via the minicomputer's interrupt structure.

When data must be transferred quickly, the direct memory access unit is used. This unit, once set in operation by the CCU, is able to pass data directly between memory and a device without intervention of the program that is being executed.

Now that we have seen how a computer interfaces with man and the real world and what a basic minicomputer looks like, let us next consider a generalized minicomputer system such as might be used in a typical application. Again the writer is indebted to Mr. Aupperle for the illustration in Fig. 3 and its explanation. Figure 3 shows a simplified, basic block diagram for a generalized minicomputer system. Excluded from the illustration for the sake of clarity are direct memory access channels, device controllers, interrupt systems, etc. The host computer shown as one of the blocks in Fig. 3 is usually a large computer to which a minicomputer may be tied in certain applications. Variations of this basic block diagram will be used throughout this report as specific categories of applications are discussed.

In order to impart some sense of order, the applications that comprise the major part of this report have been grouped in four categories: original equipment manufacturer (OEM) applications; stand-alone computing, laboratory, and monitoring applications; process control applications; and communications applications. These categories were selected by Mr. Aupperle, except that we have combined stand-alone computing applications with laboratory and monitoring applications. For each

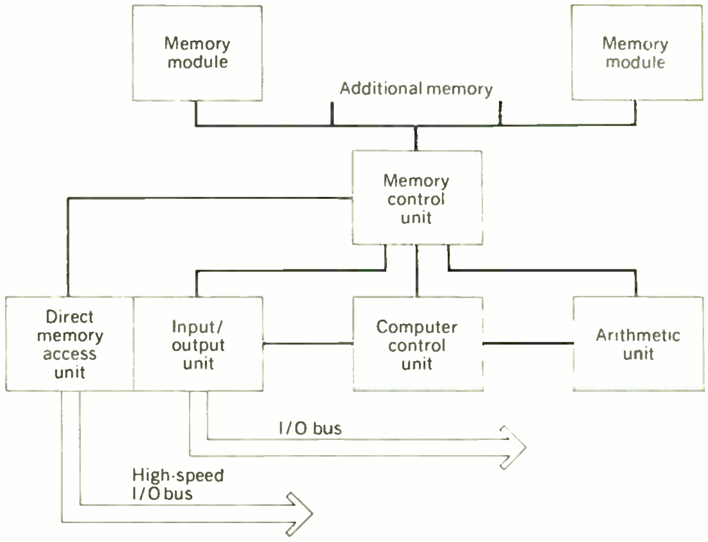
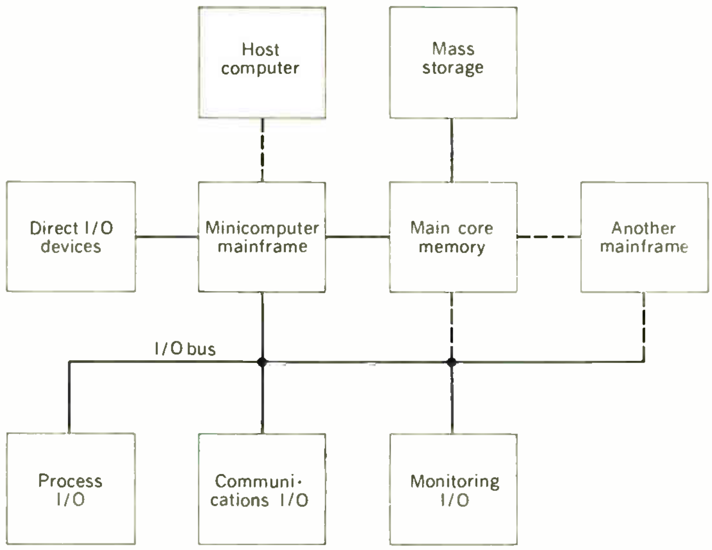


FIGURE 2. Block diagram of a basic minicomputer.

FIGURE 3. A generalized minicomputer system.



category, the basic description quoted and the basic block diagram are Mr. Aupperle's.

It is often difficult to categorize a particular application clearly. In such cases the decision whether to place the application in one category or another has been an arbitrary one. In the OEM category, any one of the application examples could conceivably appear under one or more of the other categories. The distinction here, however, is that the end user is purchasing a system in which the minicomputer is only a component rather than buying the minicomputer as such to incorporate in a system of his own design.

OEM applications

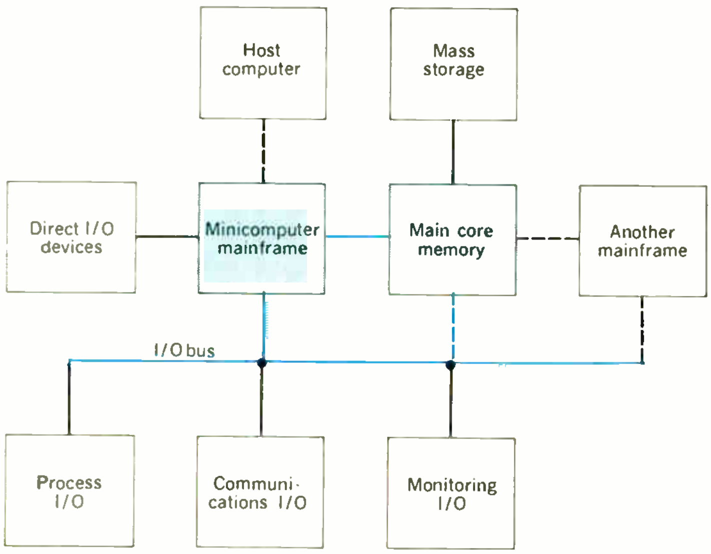


FIGURE 4. Shown in color are those portions of the generalized minicomputer system that pertain to original equipment manufacturer applications.

"Several small computer manufacturers are specializing in the sale of their products as original equipment for other manufacturers to use as components in larger, more complex systems. Clearly, most minicomputers can be used in this way, e.g., as component testers, numerical control machines, automatic weighing systems, and transfer machines. Frequently, this type of application will require only the mainframe and some memory, perhaps even without an input keyboard equipment. In this most basic configuration, minicomputers are extremely inexpensive but the user must be able to develop the necessary interfacing required by his application."

Those portions of the generalized minicomputer system of Fig. 3 that are relevant to OEM applications are indicated in color in Fig. 4. It is assumed, for purposes of Fig. 4, that a program has been put into the computer.

The following specific applications of minicomputers are typical of those that are in the OEM category.

Performance monitoring system

Control Data Corporation's new CDC 5100 minicomputer is designed to be used in environments that preclude the use of standard commercial computers. Typical of such systems is the Ocean Data Equipment Corporation Model CPMS-216 computerized performance monitoring system. It is designed to be carried on ships for test and monitoring purposes.

One specific application for the CPMS-216 is active sonar testing. The application of a computerized system in testing sonar transducers has significant advantages over the manual methods previously used.

Manual testing involved the use of an oscilloscope to test for signal characteristics of amplitude, current, and phase. The time involved in testing a transducer was five to ten minutes. Typical large systems have several hundred transducer elements active simultaneously and manual methods require several days to collect data.

In contrast, the computerized performance-monitoring system can test several hundred elements in a single "ping" of the system. The full acquisition and correlation of real-time data take about 200 milliseconds. Printing the results takes another 5 seconds. In addition to the time saved, the computerized system also provides higher-quality data and a hard-copy record of the results.

Aside from collecting and recording data that had previously been collected manually, the computerized system provides other information such as signal frequency, circuit impedance, and a real-time analysis of collected data.

The CDC 5100 used in this application is a general-purpose 16-bit minicomputer with 4K to 64K memory (K = 1024 words).

Stored-logic numerical control

A programmable all-stored-logic numerical control that is capable of both contouring and multifunction point-to-point control of machine tools incorporates the mainframe of the Westinghouse Prodac 2000 minicomputer as its logic element. In this instance, Westinghouse, the minicomputer manufacturer, is supplying a minicomputer to Westinghouse, the OEM. The new control operates on instructions from any punched paper tape system or from a general-purpose computer as part of a direct numerical control system.

Typical applications for the control are machining centers, multiaxis vertical turret lathes, horizontal boring mills, multiturret chuckers, and engine lathes.

Figure 5 is a block diagram showing activation of a machine tool by conventional contouring control. The motion interpolator activates the servo system that moves the machine members. Auxiliary functions operate on the machine through the external machine interface magnetics. The machine response to the auxiliary functions is fed back through the interface to the control. The components in the first through third columns—input logic, input control, buffers, and motion interpolator—and the auxiliary functions active store in the fourth column are hard-wired functional programs.

The portion of Fig. 5 that is indicated in color is what is replaced by the minicomputer in the new stored-logic numerical control system. All the hard-wired programs of the conventional contouring control are implemented as software in the minicomputer.

The Westinghouse Prodac 2000 minicomputer is a 16-bit model with 4K of core memory expandable to 32K.

Component test system

The Birtcher Corporation's Model 8000 automatic test system can be used to test discrete components, integrated circuits, logic cards, complete logic assemblies, and many other devices and circuits. The system may be used, for example, for production testing, incoming and final inspection, process control, and engineering evaluation.

The test system has two basic parts: the control section and the measurement section. The control section transmits digital control information to the system's measurement section, which then performs its operational functions as instructed by the control section.

Two different control systems are available with the Model 8000—tape or computer control. The tape-controlled version gives economical, medium-speed, automatic testing with a capability of basic go/no-go testing. When a minicomputer is used for the computer-controlled version the test system becomes extremely versatile, operates at high speeds, and is capable of performing a wide range of sophisticated test programs, including data analyses. The computerized version uses a Lockheed Electronics Company MAC 16 minicomputer. A block diagram of the computer-controlled system is shown in Fig. 6.

The MAC 16 is a 16-bit device with 4K to 64K plug-in core memory and a one-microsecond cycle time.

Remote monitoring system

Integrated Systems, Inc., designs and builds "Duo-Scan" remote control, alarm, and telemetering systems. The basic block diagram of the system is shown in Fig. 7. Time division multiplex is used on all inputs. Each input is sequentially examined by the remote encoder and inputs are transmitted repeatedly, one after the other, over a narrow-band communications channel to the master station decoder. The status of all inputs at the remote station is received at the master station and reconverted to parallel information by the decoder. In the Duo-Scan system, this procedure is repeated once, thereby requiring two identical pulse frames before the information received is considered valid.

To provide increased flexibility in the system, Integrated Systems is now using a GRI-909 minicomputer manufactured by GRI Computer Corporation. The computer permits the system to be adapted to changing scan rate requirements and numbers of inputs. It provides logged alarms in special formats and makes calculations and decisions on digitally telemetered quantities to determine if a given reading is out of limits or exceeds a certain rate of change. On-line system analysis can be performed based upon point status and telemetered quantities. The computer can issue control commands when initiated by an operator or automatically in emergency conditions.

The computer interfaces with the system by replacing the Duo-Scan decoder and performs all the manipulations of the code word that the normal decoder would accomplish.

The basic computer master system can handle up to 100 remote substations, each with up to 254 points of control, or, in a data logging system, at each of 100 re-

FIGURE 5. Block diagram showing activation of a machine tool by conventional contouring control. Area in color has been replaced by a minicomputer.

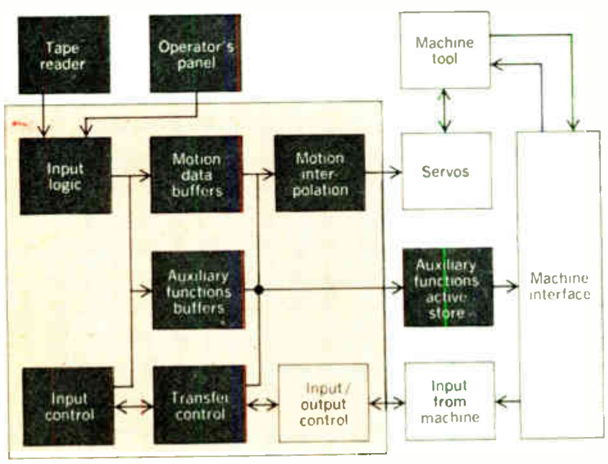
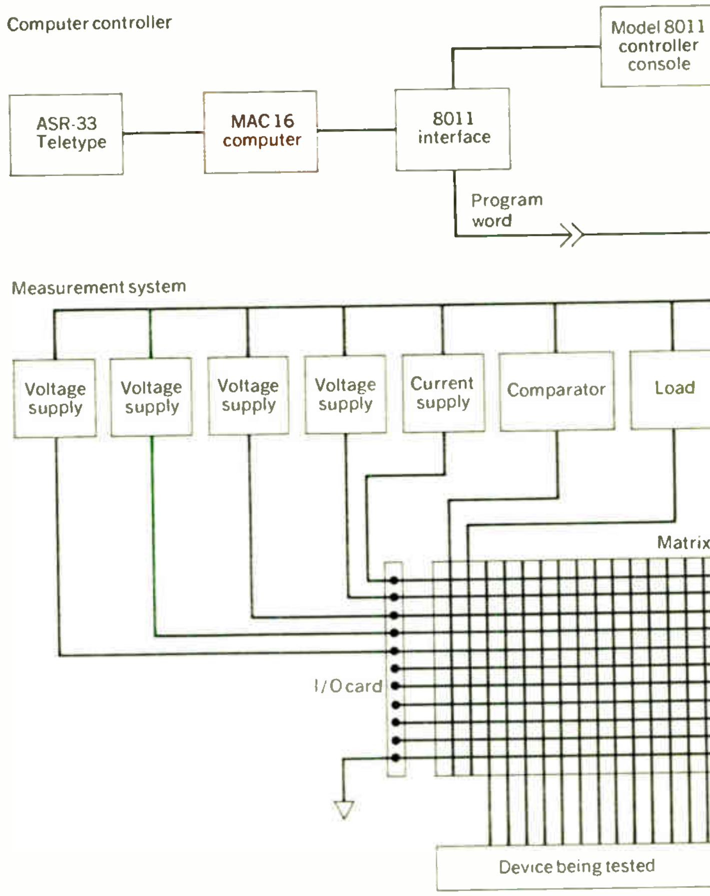


FIGURE 6. Component test system.



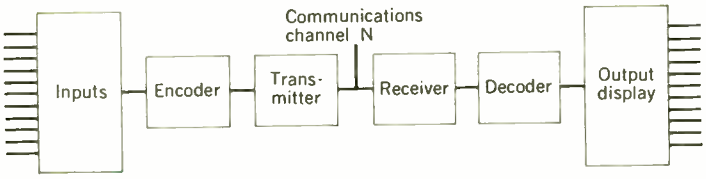


FIGURE 7. Duo-Scan alarm or telemetering system for handling up to 100 remote substations.

mote substations it can continuously monitor and update up to 256 points every 16 seconds.

The GRI-909 computer has a 16-bit parallel-word format, 4K to 32K memory, and 1.76-μs cycle time.

Data-handling system

A good example of a multicomputer hierarchy—a complex system in which smaller computers are subordinated to larger computers in order to save costs and facilitate communication—is the System Seventy designed by Mark Computer Systems and scheduled for installation this fall by the Uni-Card Division of the Chase Manhattan Bank. Based on Data General Corporation's Supernova minicomputer, Mark's System Seventy will be used by Uni-Card for on-line data entry and file inquiry into Uni-Card's IBM 360/50. The availability of the minicomputer-based system will not only facilitate data handling but may also eliminate the need for a second system 360 as a backup. This type of business data processing is common for large computers but has not usually incorporated minicomputers.

In the initial installation, each of the two Supernovas will control 20 DD-70 CRT terminals, also designed by Mark, and each of the minicomputers will be responsible for the entry of individual sales transactions and a variety of system housekeeping tasks such as account status changes and credit authorization. The installation will pave the way for eventual entry of Uni-Card sales from remote terminals located in merchant stores.

In performing all its functions, the System Seventy is designed to look to the 360/50 like a magnetic tape drive. This reduces the user's interface problems substantially and also relieves the larger computer of additional housekeeping tasks.

If the 360 should go out, an exception file of accounts to which credit should not be extended can be transferred from magnetic tape to the Supernova disk. The Supernova can then answer credit authorization inquiries on a yes or no basis while, at the same time, temporarily storing the transaction information until it can be transferred to the 360.

The Supernova is a 16-bit minicomputer with 4K-32K of core memory, read-only memory, and an 800-nanosecond cycle time.

Stand-alone computing, laboratory, and monitoring applications

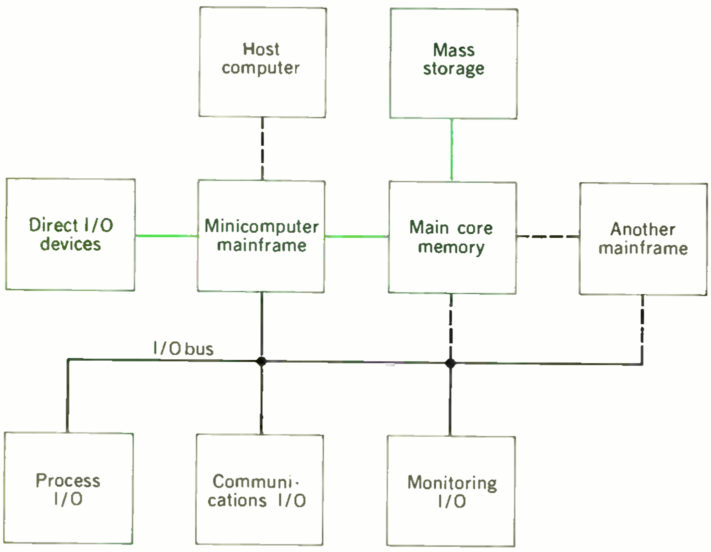


FIGURE 8. Stand-alone computing applications use those portions of the generalized minicomputer system of Fig. 3 that are shown in color.

"Stand-alone computing applications occur typically in university and research laboratories or in industrial engineering departments where real-time (as distinct from batch programming) analysis and computation are required.... Laboratory and monitoring applications represent an extension of stand-alone computing applications where usually a variety of transducers, sensors, multiplexers, and analog-to-digital converters are connected to the basic stand-alone system. Included in this category are systems configured for data acquisition, spectrum analysis, fast Fourier transformations, biological studies, medical research, oceanographic analysis, experimental physics, and many others."

Those portions of the generalized minicomputer system of Fig. 3 that are relevant to stand-alone computing applications are shown in color in Fig. 8. The direct I/O devices usually are Teletypes but paper-tape readers and punches or card readers and line printers might be added as I/O devices. The mass storage facility can be magnetic tape, disk, or drum storage devices—or, possibly, tape cassettes.

Those portions of the generalized minicomputer system of Fig. 3 that are relevant to laboratory and monitoring applications are shown in color in Fig. 9. These applications require a lot of individual tailoring of hardware. In the monitoring I/O block one would find such devices and converters as operational amplifiers, sample and hold circuits, analog-to-digital converters, digital-to-analog converters, digital voltmeters, counters, multiplexers, oscilloscope displays, plotters, programmable signal generators and power supplies, and special-purpose transducers and sensors.

The following selected applications are typical of those in the stand-alone computing, laboratory, and monitoring category.

Test of a new propulsion system

A 250-mi/h (400 km/h) linear induction motor test vehicle, Fig. 10, is being low-speed-tested with the help of a minicomputer by the Garrett Corporation under a contract from the U.S. Department of Transportation.

The heart of the electric-propulsion research is a Varian Data Machines 620/i general-purpose minicomputer located in a nearby instrumentation and telemetry trailer.

The goal of the testing program is to develop practicality studies for the new propulsion method. A vehicle propelled by a linear induction motor is theoretically capable of high speeds because thrust isn't limited by rail-wheel contact. The linear induction motor is a rotary motor that is cut along a radius, unrolled, and laid out flat. This technique gives an air gap between the primary and secondary windings, allowing linear motion between the two. One of the members is lengthened along the path of travel so that motion can be continuous.

In collecting data to evaluate the system, the Varian minicomputer accepts telemetered data into two buffers at the rate of 32 000 readings—i.e., data words—per second. The computer actually serves as a speed regulator for the data coming in from the vehicle. Incoming data are filling up one buffer as the other is feeding data to a magnetic tape. When the first buffer is filled, it starts feeding data to the tape while the second buffer starts accepting incoming data from the sensors on board the vehicle.

This procedure saves test time, according to Garrett engineers, and enables them to record accumulated data on the tapes in the correct format. One reel of magnetic tape covers a 10-minute run and contains more than 20 million data words.

Tapes filled with data from the minicomputer are calibrated and converted to engineering terms by another program fed into the minicomputer that uses an additional 32K of disk storage. The results from later data analysis will enable investigators to determine performance characteristics of the linear induction motors.

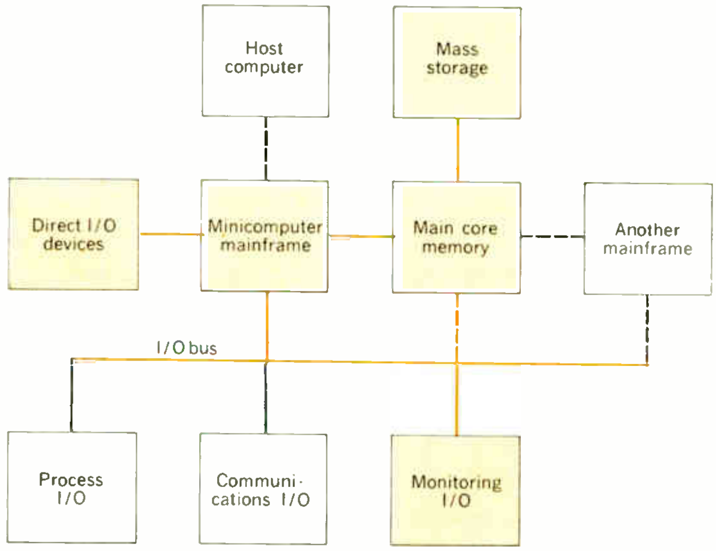


FIGURE 9. Laboratory and monitoring applications use those portions of Fig. 3 shown in color.

The Varian 620/i is a 16-bit machine equipped with an 8K core memory.

Multiprocessing approach to engine testing

Small computers have come into widespread use in all types of development and testing laboratories for combustion engines. The computers are used to increase testing efficiency, which can mean either to enable more tests to be performed per test stand or to get the results of a given test to the engineer sooner.

The functions performed by the computer can be divided into two broad categories—test measurement and control, and generation of test reports. Parameters measured in almost all reciprocating-engine tests include torque, speed, fuel consumption, and a variety of temperatures and pressures. In many test stands, the control consists of throttle and dynamometer load control, which enables programmed performance testing to be done automatically.

A multiprocessing computer system, the MODCOMP III/70 by Modular Computer Systems, is shown in Fig. 11.

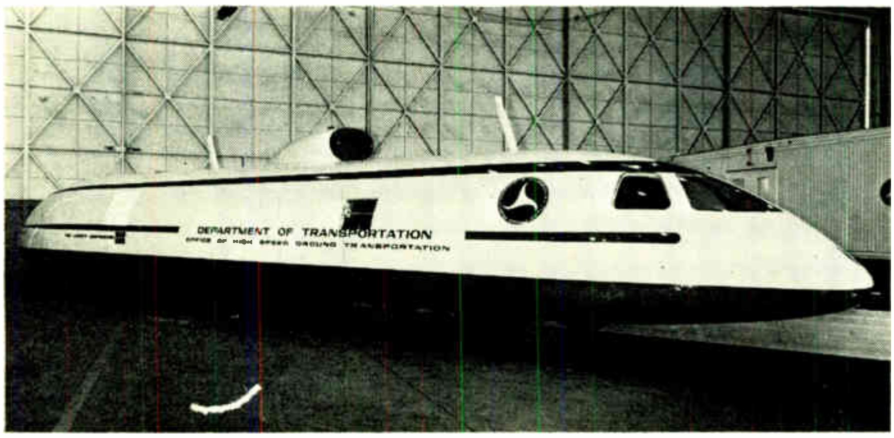


FIGURE 10. Linear induction motor test vehicle that is being low-speed-tested with the aid of a minicomputer.

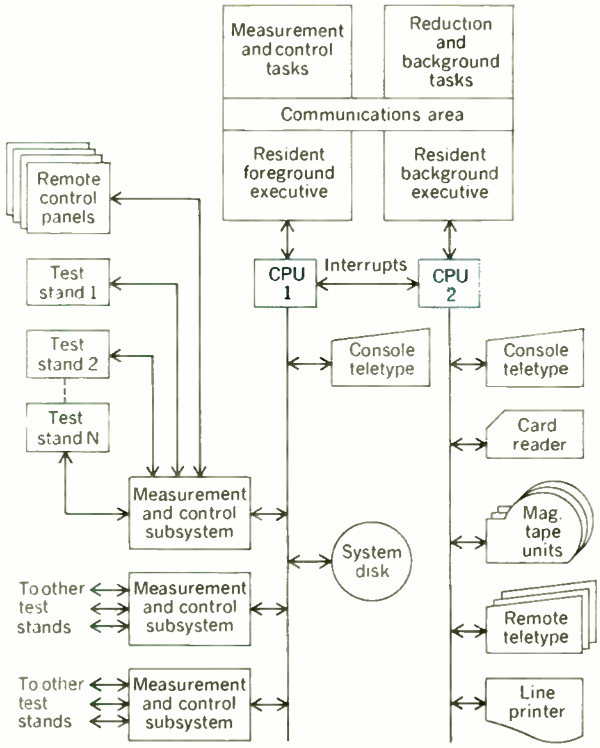


FIGURE 11. Multiprocessing system for engine testing.

It consists of two 16-bit computers that have shared core memory as well as private core memory. Each is capable of interrupting the other and communicating via shared memory. In engine testing, the system is applied as follows:

One CPU is assigned the measurement and control functions for all test stands. Test stand measurements are collected and stored. The second processor is then called to store the test data on magnetic tape for subsequent processing. This second processor produces test reports from the data stored on magnetic tape for archiving. In addition, it is capable of performing independent functions such as program assemblies, compilations, or executions.

A principal advantage of the multiprocessor approach to foreground/background applications is the degree of security provided between the foreground and background functions. For example, CPU 2, Fig. 11, cannot affect the measurement and control subsystem or the memory, interrupts, or any machine states associated with CPU 1. It can only interrupt CPU 1 at a preassigned level and communicate through the common memory area addressable by both computers.

A second advantage of the multiprocessor approach is system backup capability. If CPU 1 or its private memory becomes inoperable, CPU 2 and associated memory can be connected to take over the foreground tasks, including recording data on magnetic tape if required. Thus, testing can continue. Test backlog buildup and idling of test personnel because of computer failure also can be avoided.

The flexibility of the multiprocessor system can be made as great as desired. New test cells and new test programs can be brought on line without requiring system shutdown. Each CPU can have access to the peripherals connected to the other CPU by I/O service requests through the interrupt and common communications facility.

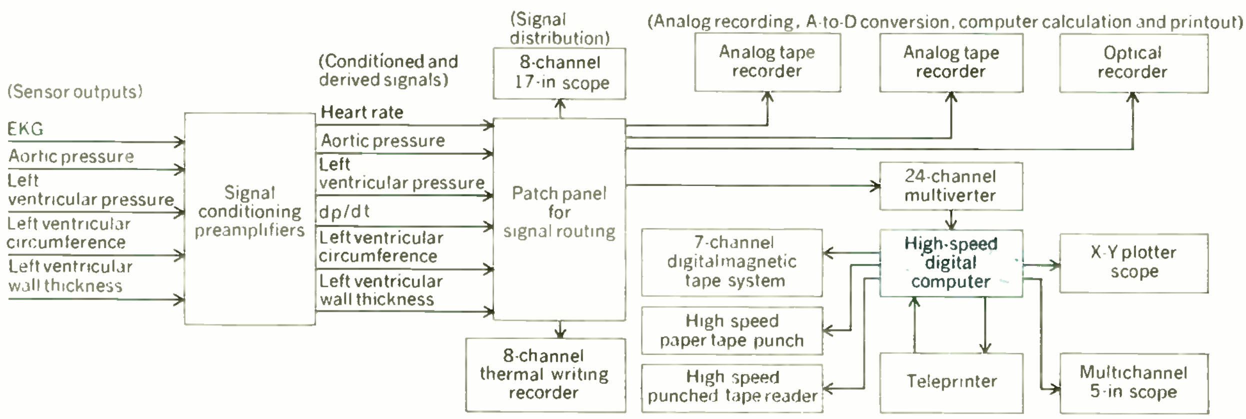
Medical research

The Cardiovascular-Renal Research Laboratory at Howard University monitors cardiac dimensions and pressures in conscious dogs and derives dynamic ventricular performance data with a Hewlett-Packard general-purpose 2116B minicomputer with a 16K memory.

Figure 12 shows the measured variables, conditioned and derived signal paths, and the elements of the data acquisition and processing system used at Howard. With a variety of implanted sensors, the investigators measure, monitor, and record the primary cardiac variables of physical dimensions, pressures, flow rate, ECG, and heart rate of conscious dogs, acquiring the data under computer control. After a selection is made of the segment of data that is of particular interest, the data are used as source information for rapid calculation by the computer of the parameters descriptive of dynamic cardiac performance such as cardiac output, flow, power, wall stresses, and other information.

The minicomputer permits the on-line computation, printout, and display of the measured and derived quantities. It also permits the simultaneous measurement, in

FIGURE 12. Basic elements of a minicomputer system for testing ventricular performance in conscious dogs.



the intact animal, of the variables for a moment-to-moment assessment of ventricular performance.

IC logic card testing

A minicomputer that can accommodate a specially built 160-line buffer to expand its basic 16-line I/O bus capability is making as many as 650 tests of 60-circuit IC logic cards in only milliseconds whereas it previously took a highly skilled operator several hours.

Computer Entry Systems, Inc., a manufacturer of time-shared, multistation key-to-disk systems, is using a Varian 620/i minicomputer to cut costs and personnel training expenses while at the same time guaranteeing that the logic cards used in its systems are 100 percent free of logic circuitry flaws. The minicomputer determines whether various logic circuits are go or no-go and also determines where the failure is when one is detected.

When the minicomputer was introduced into the testing process, Computer Entry Systems built its own 160-pin buffer to expand the minicomputer's output capability and also constructed a low-cost tape transport to interface with it. This tape stores programs developed by the company, retains the assembler for the minicomputer, and also stores a look-up table in memory which holds the addresses for instituting various test procedures and programs.

In a sense, the Varian computer programs itself when it receives the signals that tell it which part of the preformatted tape to read. An untrained, inexperienced operator can then sit down at the test console, enter through Teletype various accounting routines such as item part number, date, etc., and then tell the minicomputer to start the test.

The machine goes to the appropriate portion of the tape, reads the program required, and looks up the appropriate procedure in the look-up table. It then executes all the required tests for the logic card in question and signals go or no-go. If a part has not tested properly, the minicomputer pinpoints the circuit at fault.

Before the minicomputer system was installed, the firm was hiring trained test operators to test the logic cards on a pin-to-pin basis. After training and setting operational procedures and inspection routines, the test operation took several hours.

Process control applications

"This class of applications usually places a minicomputer system in charge of an industrial manufacturing, treating, assemblying, etc., operation. The range of applications also is broad. At one extreme are systems designed largely for monitoring, next comes systems with progressively more control sophistication in addition to their monitoring functions, and, at the other extreme, are systems using adaptive control concepts. While most process control applications occur in continuous operations (e.g., chemical, glass, steel plants), there are considerable interest and activity by discrete part manufacturers in incorporating small computers in their plants. Finally, it should be clarified how this class of applications differs from OEM applications. Truthfully, the distinction is largely one of degree rather than of fundamental differences. In the former case only the mainframe is imbedded in the equipment it controls. Here the mainframe, usually augmented by several peripherals, is viewed more as a separate element in the system. In many cases it is added to already existing equipment and/or shared among several pieces of equipment."*

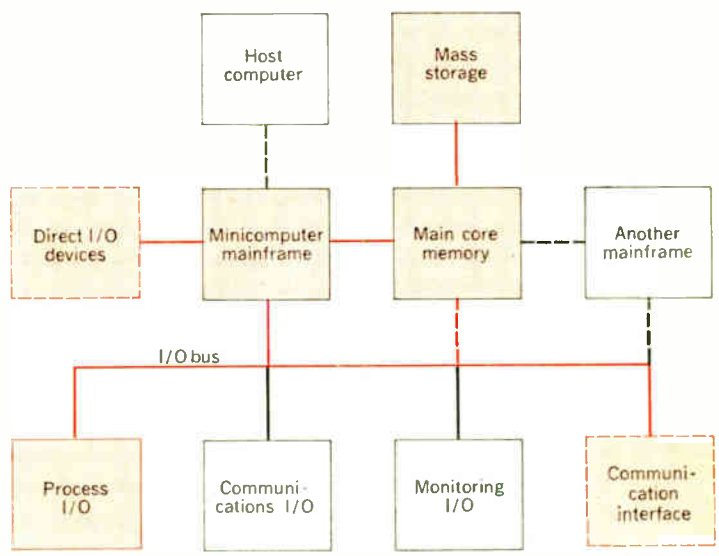


FIGURE 13. Process control applications use the portions of Fig. 3 that are shown in color.

The portions of Fig. 3 that are relevant to a typical process control application are shown in color in Fig. 13. The process control I/O block will contain both analog and digital output devices. Analog output signals cover a broad range of voltage and current levels; digital outputs can be in the form of logic levels, pulsed signals, or contact closures provided by relays. The signals are used to drive and control the actuators of the process equipment.

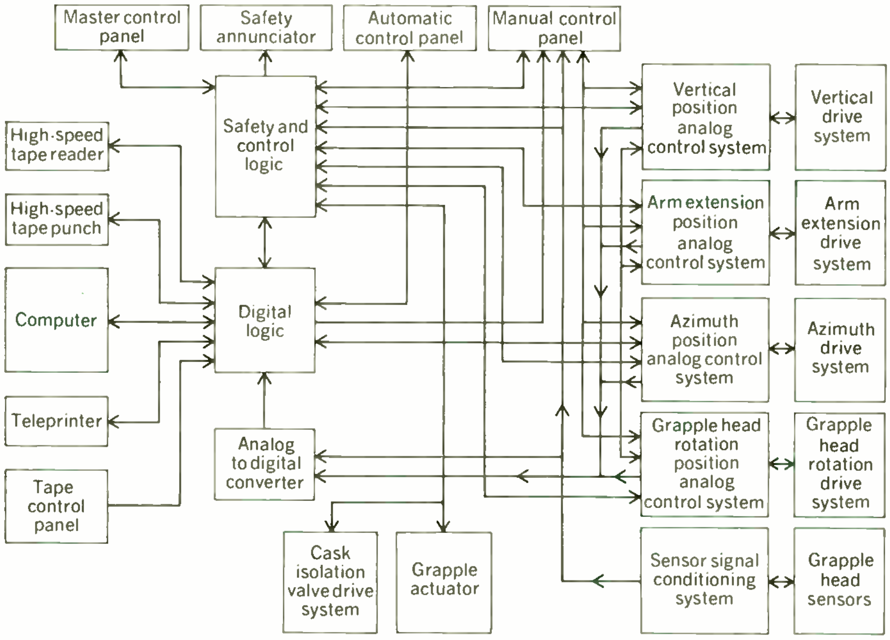
The following applications are typical of those in this category.

* These systems might alternatively be considered with laboratory and monitoring applications and the distinction, if any, is application rather than functional.

Nuclear-fuel-handling system

The precision of a Digital Equipment Corporation PDP-8/S process minicomputer will control refueling operations at the 330-MW (e) Fort St. Vrain Nuclear Generating Station near Denver, Col. Gulf General Atomic developed the automated refueling system for

FIGURE 14. Computerized nuclear-fuel-handling system.



the advanced nuclear plant, which is under construction by GCA for Public Service Company of Colorado.

The reactor core contains nearly 3000 uranium/thorium-loaded graphite blocks located inside a prestressed concrete containment vessel. The hex-shaped blocks, about 80 cm tall, are arranged in vertical stacks some six meters high.

In order to accomplish the transfer of fuel, a large fuel-handling machine is used. (It is 14 meters high and weighs 160 tonnes.) The heart of the machine is the manipulator mechanism and its control system. A block diagram of the system is shown in Fig. 14. There are four controlled freedoms of motion. Automatic control is initiated from the automatic control panel. Included are the digital computer, its I/O interface hardware, and peripheral equipment consisting of paper-tape reader and punch and a teleprinter.

The tape-control panel allows the operator to input information to the computer by means of paper tape without actually using the more complex controls on the computer itself. A multiplexed analog-to-digital converter brings analog signals to the computer from position potentiometers, grappling head translations, and tachometers.

The safety and control logic contains hardware interlocks and failure protection devices. A prime consideration in the design of this system was to maintain a fail-safe condition at all times and to prevent inadvertent operator error. Connecting to the positioning mechanisms are the control interfaces, which are basically analog servo systems.

The computer-controlled system has been completed, mated with the fuel-handling machine, and is now being used to test the machine operation.

Minimization of the time involved in refueling is desirable because no electric power is being produced by the plant during the refueling period. Initial studies indicated that a manually controlled fuel-handling machine would be too time consuming to operate, would be subject to operator error because of the numerous steps involved, and would create a problem in fuel inventory. A number of automated control types were then investigated. Numeric control by paper tape was studied but ruled out due to requirements for decision making during the fuel-handling process and the desirability for a flexible control that could be modified easily as the system was developed and put into operation. A digital-computer-based control system was finally selected. Tests have shown that fuel movement under computer control is three times faster than under manual control, which is provided as a backup mode of operation.

The minicomputer used in this application is a PDP-8/S with 4K memory. It was selected for the application some time ago during the initial design of the system, and can be replaced by either the PDP-8/L or PDP-8/I in future applications.

Glass batch weighing

At the Dartmouth College seminar on minicomputers, T. H. Finger of Owens-Illinois presented a paper on an automated glass batch plant computer system. The description of the system that follows was excerpted from his paper.

Glass batching is that portion of the glass-making process in which various raw materials are stored, weighed, mixed, and delivered to the glass furnace for melting. The glass batch plant that was described in the paper was one that was modernized from a manual weigh system to a process control computer system.

The batch plant is designed to weigh and deliver 1800 tonnes of mixed raw material per day in 4-tonne batches. A typical glass batch is made up of 2540 kg of sand, 815 kg of soda ash, 725 kg of lime, and various minor material. The weigh tolerance on 2540 kg of sand is ± 4.5 kg and on the smaller minor ingredients is ± 0.05 kg.

The batch house is divided into two weigh lines, each of which can operate independently of the other. The raw

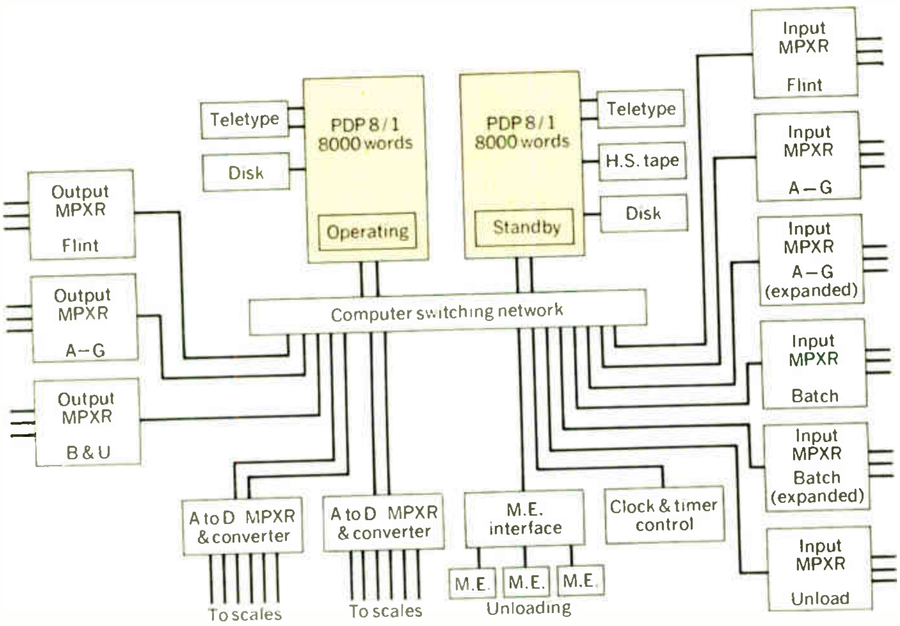


FIGURE 15. Computer system for controlling the operations of a glass batch plant.

materials are weighed into the scales from silo bins and after complete weighing are delivered to a high-speed gathering conveyor that directs them to the check scales. The actual weight from each scale is totaled and compared with the check scale weight to verify that all material was delivered to the check scale. The materials are then delivered to a mixer and, after mixing, are delivered to the furnace requiring the batch. In the plant, there are nine furnaces in the system, two batch bins each, for a total of 18 batch bins to schedule.

The plant modernization will include a total of 22 new scales, new conveyors, new mixers, and a new batch conveying system. The computer system will schedule the furnaces and weigh the raw materials.

The Owens-Illinois approach has been to size the computer system to do the task, as would be done in purchasing any other type of equipment, rather than defining a computer system and then developing enough tasks to justify the expenditure.

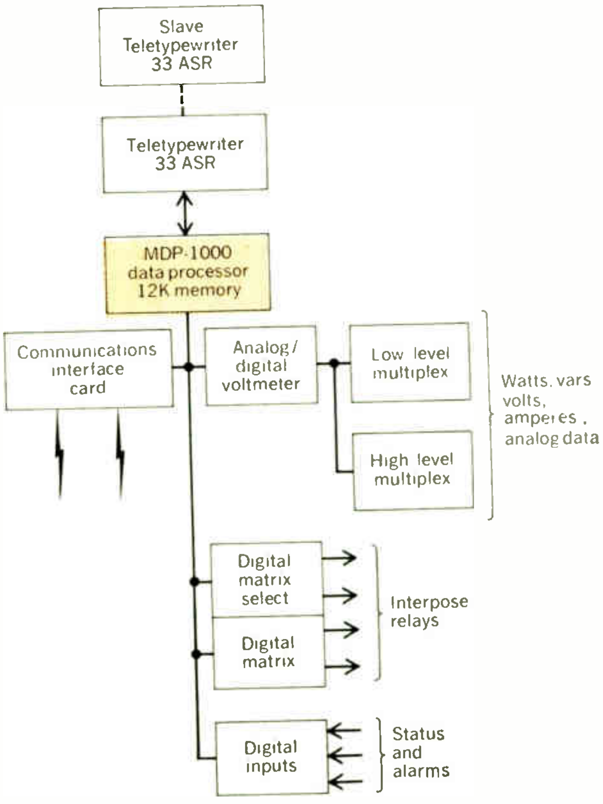
The first function of the computer is for direct control of the turning on and off of equipment. Material flowing into the 22 scales will be controlled, as will the control of the weighed raw material and its flow to the nine furnaces, with various printed outputs as records.

To accomplish these controls various interfaces were designed for process monitoring or reading of process information into the computer—mainly, are switches open or closed? There are six input multiplexers (switch selectors) in the system of 84 points each for a total of 504 contact sense points. Three output multiplexers of 72 points each for a total of 216 relay closures are used. Also included are two analog-to-digital converters and multiplexers of 24 points each for a total of 48 voltage readings.

The process monitoring or checking of various conditions will cover 54 bin level sensors, 120 equipment positions (is a gate open or closed?, is a mixer on or off?), 35 equipment service switches (is the equipment in or out of service, or is a diverter in a correct position?), 35 manual entry decades, and 85 scale unit weights (beam counterbalance weights).

Process control in the system will add and subtract unit weights to the scales as the weighing is performed. In all, there are 42 scale unit weighs, 31 scale gates, 65 raw material gates, 4 conveyors, 18 equipment on–offs, and 18 furnace schedules under process control. Process analog, or the measurement of voltage and the reading of

FIGURE 16. Minicomputer control of a remote electric power substation is accomplished with the system shown.



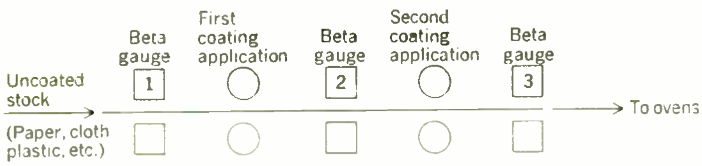
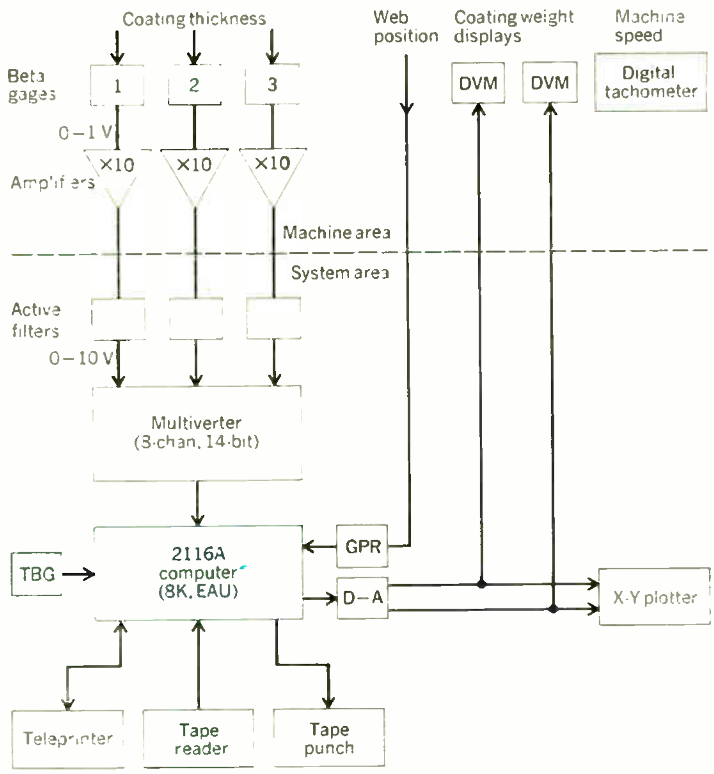


FIGURE 17. Basic components of a system for applying various coatings to uncoated stock.

FIGURE 18. Computer-based system that permits nondestructive testing of coating thickness.



the raw material scales, covers 42 scale weight potentiometers, two power supply voltages, and four test voltages.

A block diagram of the complete computer system is shown in Fig. 15. Two Digital Equipment Corporation PDP-8/I minicomputers, each with 8K core memory and 32K disk storage, are used. One is the operating computer and the other is a standby computer. If the operating computer fails, a computer switching network turns over the process to the standby computer, which is also used for off-line programming for the system.

The interfaces are divided into subgroups so that a failure of one subsystem will not stop the whole process. At the right in Fig. 15 are the input multiplexers for the reading of information from the process into the computer. Six small input multiplexers are used instead of one large one so that failure of one would not affect the other parts of the process. Two analog-to-digital converters and multiplexers are on the left in Fig. 15. Again, two are used for redundancy. At the bottom right are the unloading manual entry stations and the clock and timer control.

A standard electromagnetic weigh system was originally proposed for the modernization. The proposed equipment specifications, the material flow timing layout, point list, and hardware specifications were determined. Next, the hardware and software programming was done. The computer system economics was based upon the replacement of the electromechanical weighing equipment.

A general rule was used that the interface and sensors should cost about the same as the basic computer system and that the software programming should cost about the same as the computer system.

The total process control computer system and software programming is about one half the cost of the standard electromechanical equipment—but that does not include the extra computer. The extra computer system was purchased in lieu of a three-shift service computer maintenance contract.

Control of a remote power substation

Supervisory control of a remote, unattended power station with a Motorola minicomputer has been accomplished by Wisconsin Electric Power for its new Granville Substation. Increasing power demands required system expansion and the need arose for a substation at Granville with a new set of requirements for control, indication, and event recording that exceeded the functional capability of the existing Telememory remote.

The Motorola Control Systems Division MDP-1000 minicomputer is used at Granville to perform the following functions:

1. It receives control commands from a central station and operates the circuit breakers.
2. It stores data, and upon request of the central station, converts the data to binary coded decimal form and transmits it to the central station for display.
3. It scans ampere quantities every two minutes and performs high-limit checks for out-of-limit alarming.
4. It performs logging functions by logging data on the hour and listing all out-of-limit alarms still existing at that time and logging circuit breaker status changes in the order in which they occur and alarming the dispatcher in the event of abnormal operation.

A block diagram of the system is shown in Fig. 16. Changes are reported to the central station as soon as they occur. When the central station issues a control command, the MDP-1000 accepts the command, checks it for validity by performing multiple security checks, and decodes it. The computer then sends a signal via the appropriate digital command module in the I/O module network to operate the correct interpose relay of the two provided for each circuit breaker. One relay closes the circuit breaker, the other trips it. When the command has been carried out, a signal is returned through the correct I/O module to this effect. The computer challenges this signal and requires that its correctness be verified. When correctness is confirmed, the computer initiates a change-of-status signal to the central station, which changes the status display at the central. Status indications also are logged locally. Time of occurrence is added automatically as part of the logging routine.

Status indications are logged in the order of occurrence. The status of 65 points is monitored every 4 ms so that any changes in control positions or measured values will be detected within that time. In order to detect status changes with 4-ms resolution, the 65 points are scanned continuously. The computer checks eight points each of eight digital input cards in the I/O card cage.

In addition to cycling through the 65 points at 4-ms intervals, the MDP-1000 collects quantitative data, does data logging, and reports alarms. Scanning for ampere quantities that exceed the high limit is initiated every two minutes. The computer also stores watts, vars, and voltage levels. At hourly intervals it signals the ASR 33 Teletype through the appropriate I/O module and logs about 35 quantitative data readings indicative of plant operation. These include 15 ampere readings, eight voltage readings, six megawatt readings, and six megavar readings.

The MDP-1000 has 4K to 16K core memory with an access time of 2.16 μs.

Coating thickness detection

Measurement and control of coating weights is of major production concern to the Norton Company. Destructive sampling wastes money in terms of material destroyed and requires production stoppages. Material is also wasted if more than the optimum amount is applied. Conversely, the product will not give satisfaction if coating thicknesses are below specification or vary too much.

A minicomputer-based system permits the nondestructive testing of coating thickness. In this method, coating thicknesses are measured by beta gages that provide an electrical output related to the mass of material in the gage measuring gap. Gages are placed before and after each coating station. The process block diagram is shown in Fig. 17.

The complete system diagram is shown in Fig. 18. The system is under control of a Hewlett-Packard 2116A minicomputer with 8K memory. Web speed (up to 1 meter/second) is measured through a switch closure occurring for every 6 cm of web passing through the machine, which acts as an interrupt to the general-purpose register. This allows 66 ms for measurement, computation, and output, but only 12 ms were required in practice. Since the speed of the web and the web length between reading heads were known, an appropriate delay was inserted between samples of successive beta gage outputs so that the same portion of the web was measured.

Computed digital displays of the two coating weights are provided for the machine operator, using the two outputs of the digital-to-analog converter to drive two three-digit digital voltmeters. A digital tachometer can be switched by the operator to read web speed, coating material supply speed, etc.

In operation, the system measures, computes, and outputs the actual net coating weights to the process operator about five times per second. At the completion of each run (up to 915 meters) the system outputs a run data summary, listing average weights and standard deviations for the backing material and each of the coatings.

Communications applications

"Minicomputers are and will continue to serve many vital communications applications. One example is data concentrating systems where a number of low-speed input devices such as Teletypes and other human input/output terminals are connected to the minicomputer. The minicomputer then concentrates all of this data to transmit it efficiently to some other device, usually a larger computer. Another example is the use of minicomputers to facilitate intercomputer (large CPU systems) data transmission. Still another is the addition of a card reader, line printer, and a communication interface to a minicomputer to build a remote batch terminal for a large central processor. Minicomputer time-sharing systems may be included here too. Such applications have been known for a long time, but have been seriously considered only in the last few years. Considerable activity in data communications will be generated in the future."

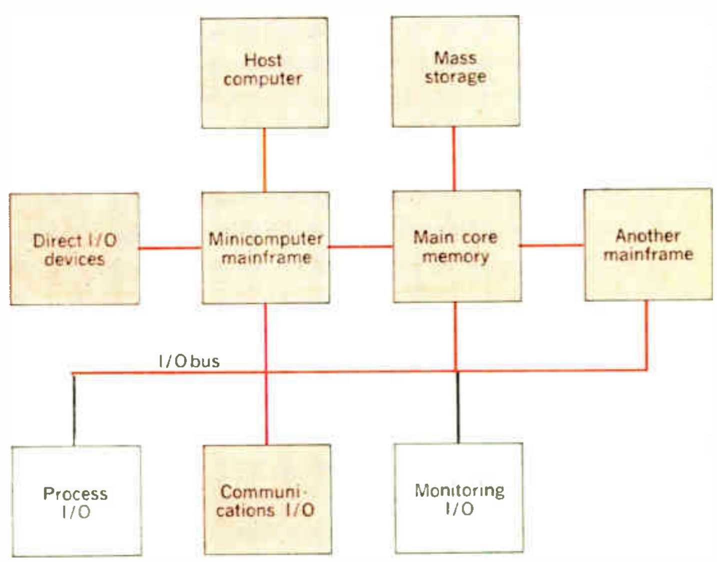


FIGURE 19. The portions of the generalized minicomputer system of Fig. 3 that are used for communications applications are shown in color.

The portions of Fig. 3 that are relevant to communications applications are shown in color in Fig. 19. Mass storage devices are not always included but most systems include one or more direct I/O devices in addition to the communications hardware. Some kind of interface to a host computer is common. The communications I/O block may contain data set controllers to service various common carrier data sets.

The selected applications that follow are typical of communications applications.

Warehouse data handling

A major reason behind the proliferation of minicomputers has been their application in new and frequently never-before-computerized jobs. The chief advantage of the minicomputer in such applications is its ability to provide an improvement in price/performance over the way a job has been done before.

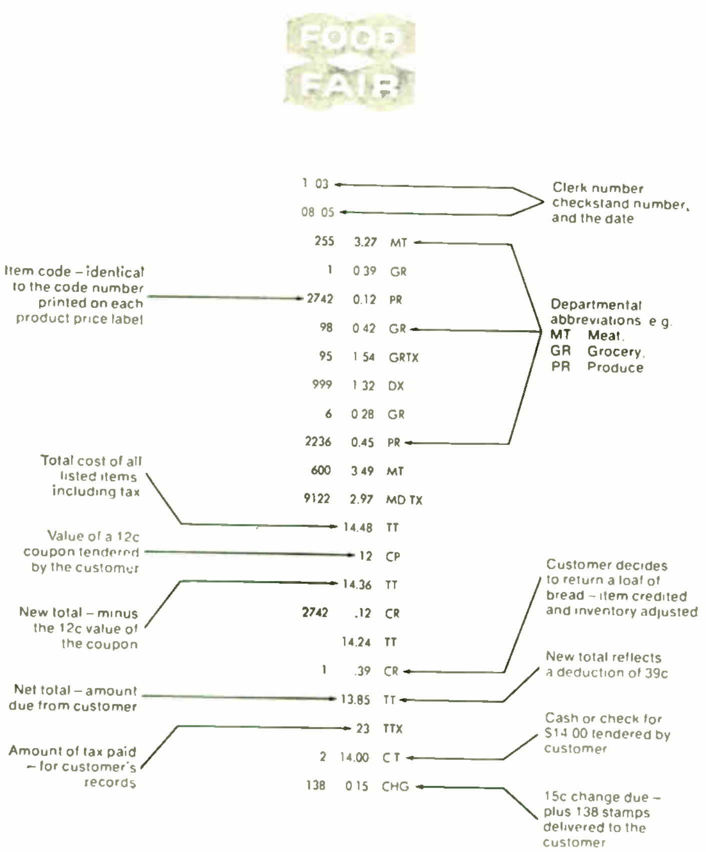
Newly developed uses of minicomputers in noncontrol applications such as warehousing and retailing point up both the versatility of the small computer and the advantages it can provide to the user. One system, developed for warehouse applications around Data General Corporation's Nova minicomputers, consists essentially of a series of free-standing minicomputer systems, each of which may be used as a terminal for a larger computer. Each of the remote minicomputers in the system is tailored to a specific warehousing situation and each is designed for use by unsophisticated warehouse personnel.

The small computer is programmed to operate in a question-and-answer mode. The computer presents the series of questions and warehouse personnel check off the correct answers. Use of the minicomputers on site at the warehouse eliminates the need for more expensive data communication facilities. The small computers may also perform certain data processing on site, including inventory control, payroll, and other duties related to the specific warehouse.

As the on-site computation is completed, the minicomputers, now acting as terminals, condense the data and forward relevant parts over telephone lines to the central computer facility for further data processing and management information on the corporate level.

Small computers acting as terminals also may be used to save communication costs. One such system has been designed for a major retail operation to allow the retailer to take advantage of a party line telephone cost and reduce the number of telephone lines required. In this instance, it proved possible to cut the telephone bill from a potential $10 000 per month to $1200 per month. Equally important was the fact that the retail warehouse network system was designed around existing production facility procedures, thus enabling each computer in the system also to act as a translator between the individual warehouse departments and the machine's central processor.

FIGURE 20. Customer's record of marketing transactions at minicomputer-controlled supermarket.



Supermarket control

The Food Fair Store in Baldwin Hills, Calif., is the scene of a wedding between a cash register and a minicomputer to produce a totally new control system.

The shopper gradually notices that each item in a department has the same color label but its own code number and that odd-weight packages of meats and produce are stamped simply with a price per pound. At the checkout station, the checker taps the code color, code number, and quantity into a counter terminal whose display flashes the correct price for each purchase. Since 60 percent of all products sold are groceries, the system is programmed to assume that a grocery sale of one item has been made unless otherwise indicated.

The checker is completely relieved from concern with such factors as prices, coupons, and sales taxes. Meats and produce are weighed and priced automatically. The computer even notifies the checker of any invalid entry and waits for a correction before the next entry can be punched. The customer is provided an accurate, detailed record as shown in Fig. 20. Officials of Inventory Management Systems, who designed the system for Food Fair, believe it can be improved by having the checker hold a portable electronic scanner before the coded label on an item. The scanner will read the code into the computer system, which will print out, via the terminal, the correct information on the customer's receipt tape.

A block diagram of the system is shown in Fig. 21. It uses a Honeywell DDP-516 in the prototype system or a Honeywell H316 in the standard version. When the checker enters an item code, a multiplexer identifies the particular check-stand number, adds an address, forwards data with the address attached, and converts the data into EIA-compatible signals. The multiplexer shifts levels to convert data from digital and logical numbers to frequency changes that can be handled by the telephone data set. It scans all terminals for information, which is forwarded over a single telephone line to a remotely located data center where another data set converts the frequency back to EIA levels. A communication-line controller then converts the signals into logical levels for computer input.

The computer accepts the data, identifies the source terminal, stores the new data, and performs arithmetic or erasing functions. The computer then sends return messages via the controller and data set, telephone line, data set, and multiplexer back to the specific terminal. Backup provisions include two telephone lines between the store and the data center and an H316 computer on the store premises.

In the standard system, the H316 is the primary computer in a central data center. The Honeywell H112 is used as a controller for a Librascope disk in the store for backup purposes. This backup eliminates the need for more than one lease line per store.

The system's most valuable asset is probably its real-time capability for updating inventory records at the store, warehouse, and within a total retailing network. This approach permits accurate electronic ordering. When the checker records an item purchased (or returned) the inventory of the item at each level is corrected.

System benefits are extensive. Automatic recording is possible, each store's ideal product mix can be determined, capital investment (e.g., cash registers) can be reduced or eliminated, and daily and weekly sales reports by department, as well as weekly reports of low turnover items, are available.

The H316 is a minicomputer with 4K to 16K memory, disk storage option, and 1.6-μs cycle time. The H112 is a 12-bit stored-program controller with 4K to 8K memory and 1.69-μs cycle time.

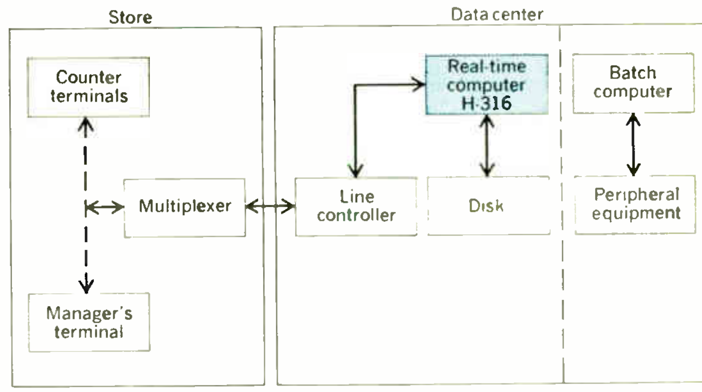


FIGURE 21. Block diagram of a supermarket system controlled by a minicomputer.

Telephone trouble reporting

A Motorola MDP-1000 minicomputer in combination with a MDR-1000 electronic document reader is making more efficient the handling of customer telephone troubles for a western telephone company. The system being used for reporting trouble to the proper service center and initiating corrective action has reduced both costs and reaction time substantially.

When trouble is reported, a clerk checks the appropriate boxes on a printed form. At frequent intervals the forms are collected and placed in the document reader, which reads the cards automatically and transfers the information they contain via dataphone data set to the computer.

In the system, the computer performs the function of separating and classifying the trouble information, then routing it to the proper service center. It also performs an accounting function. In this role, it classifies and summarizes the trouble information by type and source of trouble.

The statewide system consists of nine terminals that originate trouble reports and 16 terminals to which trouble information is routed for action. The computer polls the readers in sequence and transmits the sorted trouble information in a format suitable for Teletype readout.

In-house time-sharing system

The Wang Laboratories Model 3300 Basic is a low-cost, easily operated minicomputer time-sharing system. It accommodates any number of terminals up to 16. With the Wang Model 1103A acoustic coupler, the user may operate from remote locations using a Teletype terminal and standard telephone lines.

The central processor is an 8-bit minicomputer called the 3300. It has 4K memory expandable to 64K and 1.6-μs cycle time.

Since the Wang system uses an extended version of the Basic language developed at Dartmouth College as a problem-solving tool, it is easily used for solving mathematical problems. The steps involved in solving a quadratic equation for its real roots are as follows:

Given the equation

$$2x^2 + 9x + 3 = 0$$

when

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

Turn on the terminal-press ATTENTION key

```
BASIC READY. (The system is ready to go.)
:10 REM THIS SOLVES THE QUADRATIC
:20 LET B = 9, LET A = 2, LET C = 3
:30 LET S = SQR (B ↑ 2 − 4 * A * C)
:40 PRINT "X1 = ", (−B + S)/2*A, "X2 = ", (B−S)/2*A
:Run—
X1 = —0.36254134 X2 = −4.1374586
```

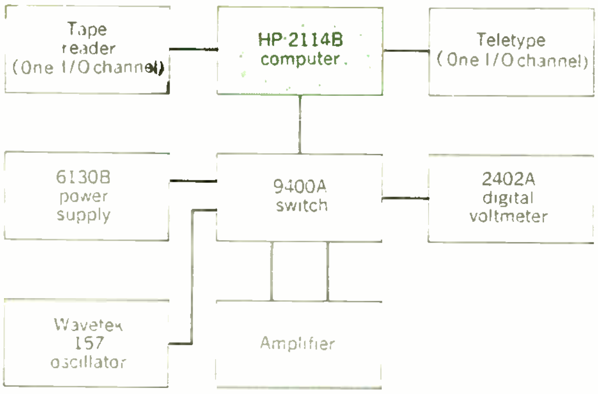
Conclusion

This report has concentrated on presenting specific minicomputer applications to give the reader an insight into the wide range of minicomputer capabilities. As was stated earlier, it is beyond the scope here to discuss in detail how systems are configured or how minicomputers are selected. But one grossly simplified example of a system configuration may serve as an indicator of just how involved the configuration of a complex system may be.

Hewlett-Packard tells the potential user of one of its HP 9500 Series Automatic Test Systems how to configure a system of his own. It involves making a block diagram with the computer in the center, leaving room above for computer input and output devices, with stimulus instruments on the left, measuring instruments on the right, and switching devices on the bottom. After particular instruments and interfaces are selected from information supplied by Hewlett-Packard, the next step is to totalize the I/O channel and memory requirements from the information given about the various instruments.

For memory requirements, you add the expected number of words that would be required for applications test programs to the total required for instrumentation. A minimum of 1500 to 2000 words for test programs is recommended.

The choice of computer and appropriate options must be made last since the computer requirements are not known until all the automatic test system requirements have been fulfilled. Then you make a choice of computer and extended options to satisfy the requirements of total memory and I/O channels. Next you select computer



Model number	Option or accessory number	Description	Words of memory	I/O channels
6130B	J20	Digital voltage source (first one)	(207)	(1)
157	S 134	Waveform synthesizer (Wavetek)	(168)	(1)
9400A	003	Distribution switch (3 relay trees)	(33)	(1)
2402A		Integrating digital voltmeter	(106)	(2)
		Total	514	5
2114B		Computer	2649	7

FIGURE 22. Simplified system for testing an amplifier showing number of memory words and input/output channels required for specific devices.

mainframe options and also power fail interrupt and parity error check, if desired, and cabinet requirements.

The elementary system shown in Fig. 22 might be used for testing an amplifier. This system uses a computer with tape reader and Teletype, a dc power supply, an oscillator, a digital voltmeter, and a switching device. In Fig. 22 are also shown those portions of the tabular information supplied by Hewlett-Packard that apply to the particular instruments and devices used in the block diagram.

The number of words in memory that are required for the instruments totals 514. If the HP2114B minicomputer, for example, is considered for the system, it has available 2649 words of memory and 7 I/O channels. Subtracting the 514 words of memory from the 2649 available leaves 2135 words for the applications test programs. The number of I/O channels available is sufficient without need for I/O channel extenders.

The Hewlett-Packard configuration example works easily because the HP 9500 system was designed to be modular. The amount of core memory used for the computer reflects the fact that it uses a resident interpreter to greatly simplify programming.

Designing a system of your own is not the only alternative, of course. Ronald H. Temple of the NEC minicomputer seminar faculty stressed the fact that the end user of a minicomputer has four avenues open to him: do it himself, get it all from a computer vendor, have a systems house do it, or buy the whole system from an OEM. Doing it yourself is the riskiest, he said, and buying from an OEM is probably the safest.

The prime requisite is to find someone who understands *your* problem. Such a person is the best investment you can make toward your goal of a workable computer system that is not a financial disaster.

To dig deeper . . .

There have been many excellent in-depth reports published on the subject of minicomputers. The following three are typical:

Minicomputers and Communications Preprocessors—Part One, Minicomputer Concepts, Principles, and Applications; Part Two, Descriptions of Individual Minicomputers, Publication's Div., Programming Sciences Corp., 6 East 43 St., New York, N.Y. 10017, 1970, $250.

All About Minicomputers, Datapro Research Corp., 2204 Walnut St., Philadelphia, Pa. 19103, 1970, $10.

Small Control Computers, Equity Research Associates, 55 Broad St., New York, N.Y. 10004, 1969.

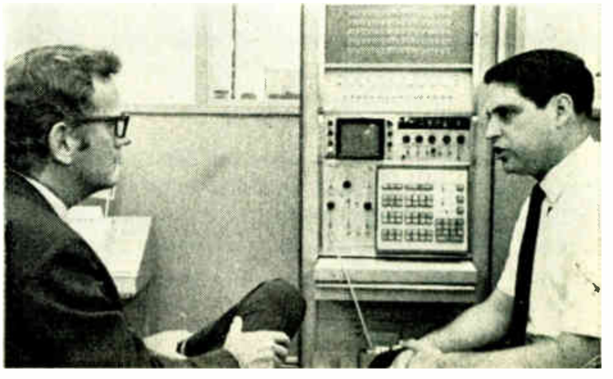
A representative listing of minicomputer manufacturers and their addresses will be sent to any reader who circles number 100 on the Reader Service Card in this issue.

The writer wishes to thank all those people who contributed their time and knowledge to make this report possible and, especially, Eric M. Aupperle and the National Electronics Conference for allowing use of some of their minicomputer seminar material.

Ronald K. Jurgen (SM), managing editor of IEEE Spectrum, has been in the technical editing field for 20 years and has been managing editor of IEEE Spectrum since its birth. He was graduated from Rensselaer Polytechnic Institute in 1950 with the B.S. in E.E. degree (electronics option). He received his basic editorial training as an assistant editor on the staff of Electrical Engineering and then joined Electronics under the editorship of Donald G. Fink, now IEEE's general manager.

In 1953, Mr. Jurgen started, in the capacity of editor, Sutton Publishing Company's monthly journal, Electronic Equipment, now EEE. After five years in that position, he returned to Electronics to be feature editor. His next position was manager of scientific information, on the corporate level, for IBM. He rejoined Sutton to become editor of Industrial Electronics and, in 1961, returned to Electrical Engineering to become editor in 1962.

In the illustration, Mr. Jurgen (left) is shown with Peter R. Roth, project leader for the Hewlett-Packard 5450A Fourier Analyzer (background), a minicomputer-based measurement system.



The exponential crisis

A backlash may be developing in society because the pace of an accelerating technology is outstripping society's ability to acclimate itself to scientific change

Neal Eddy Weston, Mass.

An exponential proliferation of technology imposes serious burdens upon society. Consequently, the engineering community will be forced to deliberate more than it has in the past about the social consequences of its acts. In this way, the best interests of both the engineer and society will be served.

As engineers, we don't often discuss the social consequences of our technology for this smacks of meddling into things outside of our competence. One of the first things we learn on the job is the need to understand basic processes thoroughly before attempting to bend them to our will. We develop a healthy disdain for those things we can't put into unambiguous mathematics. By default then, we have left the messy questions of technological interaction with society largely to the politician, the sociologist, and the economist.

It is essential, I believe, that we overcome this natural aversion. In brief, I argue that technology develops exponentially, whereas man (and hence his social institutions) is limited as to the rate at which he can satisfactorily assimilate the resultant social change. It is essential that we attempt to understand, anticipate, and ultimately control a substantial fraction of these social forces if we are to meet the stressful challenges of this decade. Already, there is cause for concern. Society is less eager to support technology: Many of us have been, or soon will be, forced to seek other employment.

We, as scientists and engineers, know technology to be entirely neutral, merely a tool in the hand of the user. But like Arkwright, Kay, Hargreaves, and other innovators of the Industrial Revolution, we are nonetheless vulnerable to the blind, unreasoned reaction of the mob, which cannot distinguish between technology and the applications to which it is put. Technological proliferation inevitably causes or exacerbates tensions within society, thereby threatening its stability and, as a result, endangering our livelihood. To avoid future troubles and ameliorate present anguish, it behooves us to consider carefully the implications of our acts.

Exponential growth

Let me begin this argument by noting first that technology has indeed developed in an essentially exponential[1] fashion since the days of the Industrial Revolution and the systematic cultivation of technology. There is considerable disparity in the exponential constant from one field to another, for there are fashions in technology as in most human endeavors. For unclear reasons, possibly instinctual, more probably psychological, we seek out those problems believed within our grasp and avoid others that are perhaps more important but have a lesser probability of solution. Progress therefore proceeds in spurts, first in one field, then another, not entirely unlike fitting together a jigsaw puzzle. It is thus not surprising that growth rate varies over the broad areas of technology. But it is remarkable that most indexes show growth rates between three and six orders of magnitude per century. To illustrate this point note that global circumnavigation has decreased from roughly 1.5×10^3 to 1.5 hours, whereas electrical-communications bandwidths have increased from 200 Hz to 200 MHz during the hundred years, 1870–1970. Although there are many more scientists and engineers by virtue of an expanding population (0.1 percent per decade), the principal factor in this growth rate is the increased efficiency of the technological society that frees more and more of us from the requirement for immediately productive labor.

Society is beginning to recognize the dangers inherent in uncontrolled population expansion, but we have devoted little direct attention to the concommitant dangers of undirected technology. Positively increasing exponential processes cannot proceed indefinitely in the real world for these ultimately would violate the laws of mass and energy conservation. Population explosions are eventually limited by either the food supply or the frustrations and apathy caused by overcrowding. Similarly, the technology explosion is limited both by resource availability and, I maintain, by social pressures. Just as man has a choice of how he will limit population—by planning or by starvation—I argue that we may choose between conscious planning and nihilistic, anarchic social rebellion that eventually leads to the regulation of technological growth. Perhaps this view is overstated, but before dismissing it, consider first if man's capacity to accommodate social change has increased.

Man's nature remains unchanged

I claim that man's nature remains substantially unchanged, both biologically and psychologically. To be sure, improved food supply, sanitation, and medicine have enabled men to grow taller and healthier. Improved nutrition may also have slightly increased man's mean intelligence: Better diet in youth more than offsets both geriatric decline (in a longer-lived populace) and a degraded gene pool (by subversion of natural-selection processes).

Social and psychological fashions come and go, but it would be difficult to argue that the basic motivating forces have changed greatly since medieval abnegation of Roman technology. The veneer of education may indeed be substantially thicker, yet it has not significantly altered basic motives or social cohesion. Short of genetic engineering, or its more practicable equivalent, eugenics, man's innate capacity to anticipate, plan for, and assimilate social change is constrained.

In effect, I argue that man is willful and capricious: He does not always serve his own best interests, even when perceived, whether those interests be immediate or deferred. He bumbles along, his will to plan circumscribed by a love of freedom and a sense of adventure. Thus the greatest danger, I believe, is the irrational act: not particularly the unrestrained technology—evil as its consequences may be in some instances—but rather the consequent, unreasoned reaction to it. The danger is either repudiation or abandonment of technology when the social disruption proves excessive. This sort of reaction serves the interests of none, for many, if not most, social ills require continuing technological advance for their solution. In short, I argue that we impose a burden upon society by our exponentially accelerating technology. Since we cannot count upon society, by itself, to improve its capacity for accommodating change, we must attempt to improve society's understanding of what we do, and further—and perhaps most important—must take great care in how we add to the burden of society created by our thoughtless acts.

Infringement upon historic freedoms?

Let us consider some examples of this exponential burden. People find themselves thrust closer together. Habitable land area remains fixed at the same time increased population, communication, and transportation bring us into closer proximity with our neighbor. Of necessity, freedom and independence become curtailed. As in thermodynamics, so too in sociology; stuffing more and more molecules (people) with ever-greater energy into a fixed volume increases the collision rate. Neighbor becomes unpleasantly aware of neighbor. The trend is to expend more energy, dump more waste, and thus fill even more space. The collective environment is thereby degraded. The interaction between environment and life style is too complex to describe in rigorous fashion, yet it is clear that a loss in the quality of life ensues. That man can alter significantly his children's lives through failure to control pollutants has become a part of the "conventional" wisdom.[2] To espouse such arguments is even politically fashionable.

Pollution is a complex, vexing problem fraught with difficulties. I don't raise the issue for its own sake, important as it may be, but rather as but one readily grasped example of the technological burden. Certainly, I am concerned with environmental effects upon my children's health. Yet I am far more concerned that we, the technological practitioners, failed to perceive the problem earlier (or, at least, did nothing to call attention to it or to point out our limited vision). We, therefore, court the greater disaster of public rebellion against, not pollution, but the technology that made that pollution possible—the very same technology that is clearly required to control pollutants in an expanding population. I won't attempt to point out what we might have done differently for that is now academic. (Had we been able to anticipate the problem sooner, there is serious question that the public would have acted on the solution or that less blame would ultimately be placed upon technology. But clearly this is no excuse for not having tried.)

To provide another example of the exponential burden, weapons technology has increased kill capacity some three orders of magnitude during the last century. The minie balls and Gatling guns of a century ago did cause great bloodshed, pain, and death. Today, these weapons seem ludricrous toys. The endless names of Civil War dead were laboriously chiseled into monuments of gleaming elegiac marble. Yet what are those numbered dead when we consider our present potential to annihilate all life by a mistake?

Man rationalizes his weapons, claiming they will be used defensively. But then he works toward first-strike capability, saying that offense is the best defense. Yet the danger is that society ultimately may come to equate this perilous balance of terror with the technology that made it possible.

The argument is not yet complete. In my view men are faced with problems even more serious than simply coping with the possibility of world annihilation. After all, individuals are ultimately more concerned with the immediate quality of life: security of family, freedom from material want, the liberty to pursue the course of our lives in tranquillity, the freedom to speak our minds and follow our own goals. These are more important than life or death. Parlor discussion of "better Red than dead" versus "liberty or death" is idle. Thoughtful men would agree that with sufficient loss in freedom life becomes living death.

I suggest technology has indirectly infringed upon our freedoms. This point can be pressed in several different ways. For example, I have already discussed the pollution of the physical environment and alluded to its effect upon life style. There are other equally important effects. Collectively, we have moved into the city from the farm because agriculture has been made so productive by technology that there are few jobs left on the farm. But in the cities and suburbs to which many of us have moved, we no longer sink the deep roots we once had. Our technologically oriented society places a premium upon job mobility, makes it easier to move when desired, provides social and welfare services, and, perhaps most important, affords engaging alternatives to conventional social intercourse through mass-medium entertainment. Yet it can be argued that all these factors ultimately result in a curtailment of freedom.

Social deterioration may partly result from technology

Once even the most densely populated slums maintained a spirit of community: The juvenile apple thief would be caught and thrashed by the casual onlooker who, as often as not, would know both boy and family and hence would serve on the spot as surrogate parent. Today, we don't know boy or parents; the juvenile thief, lost in the toils of an impersonal legal system, becomes confirmed in his antisocial behavior patterns. The "conventional" wisdom has taught people to avoid involvement; they look the other way and avoid trouble. In my affluent suburb, I am safe from neither my neighbor's son's vandalism nor the professional thief who commutes from the city. The spirit of community is weak and neighbors do not involve themselves in the affairs of others. It is, of course, much worse in the cities: The anonymity born of transitory life styles breeds gangs of young hoods who mug the elderly in broad daylight for psychotic thrills. Kitty Genovese's cries go unanswered. We are still our brother's keeper; it is just that we can no longer recognize him.

The deterioration in social structure is widely noted. Admittedly, the causes of that deterioration are not so clearly understood. The social machine is an exceedingly complex mechanism. However, strong arguments can be adduced to the effect that technology has been a major factor in this decline. Not only has there been a population move to the city as the result of improved agriculture but, perhaps more important, many individuals are nightly glued to their television screens. Which of its inventors and developers foresaw television's present role?

Vladimir Zworykin, for instance, said he had anticipated something quite different, something more cultural. Instead, television's advertisements breed unrealistic goals through the idolatry of youth, sex, and wasteful consumption of superficial products; its news focuses upon the violent, not through any conscious design, but through the elemental pressures of the marketplace; its entertainment blandly bathes our senses in ways calculated not to tax the mentally disadvantaged. Television has become a distorting mirror that we hold up to ourselves—a mirror that reflects a tarnished image, which itself causes change within the social structure. The apparent freedom of *laissez-faire* is illusory—it is only another means for directly altering society. The pressures of the marketplace determine what is seen and so indirectly mold human thought through technology.

Let me elaborate further. I would argue that current injustices, real and fancied, are not greatly different from those of the past. For example, the Spanish-American War, prompted by jingoistic yellow journalism, had less firm moral foundation than even the most acidulous critic would claim for the Vietnam conflict. What are the injustices now inflicted upon our Indians compared with prior butchery and thievery? Even were all Black Panther claims true, how many blacks are murdered, lynched, or terrorized today compared with those victimized during postreconstruction times? Don't misunderstand me; my intent is not to condone injustice, past or present. Regrettably, injustice does occur and civilization must dedicate itself to its exposure and subsequent eradication. The important point here is that injustice is now far more visible than formerly, especially when it involves violent and bizarre acts. Response becomes more immediate, more visceral. Attempt at reasoned argument fails when adrenalin flows. It succumbs to a kind of Gresham's law: the emotional and violent subordinate the thoughtful.

Television programming provides an immediate, readily visible example of the crushing pressures exerted by a proliferating technology upon an inadequately prepared society. I do not argue for television's abolition or censorship. In fact, its capacity for good may well outweigh the harm it does. I wish merely to show that inherent in an exponentially expanding technology is the power to change social values in unanticipated, unintended ways. If this point hasn't yet been made clear, consider the manner in which social values have changed over this past century.

In the days of the frontier, a man ate in proportion to how productively he hunted or farmed. If he shirked, he went hungry: Self-reliance meant survival, dependency meant death. Now our productivity is sufficient to support the weak. Old virtues are no longer essential and, in fact, slowly erode under the pressures of the new. We design and build machines that do the hard, back-breaking work so that physical labor, per se, is no longer to be admired. Craftsmanship is less important than the ability to design the machine that can do a faster job than the obsolete craftsman. Productivity is no longer measured in man-hours of drudgery. The skills of communication, of getting along, and of coordinating activity become more important than the physical skills. Human labor becomes increasingly expensive compared with that of the machine, and so it becomes cheaper to buy a new item than to have the old repaired. Given quick innovation and rapid obsolescence, it becomes cheaper to produce designs with restricted life. Thus values must change. Yet, as previously indicated, rapidly changing values breed uncertainty and society becomes vulnerable to disruption.

The responsibility is ours

It is easy to question the underlying values of a technologically oriented society and we well may quarrel that such questioning is pointless without explicit, workable values to replace the old. But the prime consideration is that many do question our values; that there is an inchoate, growing suspicion of our technology.

We run the danger of having to take responsibility for something that we alone did not create, and that, in any event, was not the result of conscious plan. As technologists, we are thus the potential victims of a society that may conclude, rightly or wrongly, that we have collectively spindled, folded, and mutilated its values.

Of course, no one desires controls of any sort. Perhaps along with other forgotten lessons of history we are fated to relearn the parable of the commons*: In the late 18th century, English villages all had their commons—jointly owned pasturage—where all cattle were free to graze. Each herdsman, being free to graze as many cattle as he wished, would reason, entirely on the basis of self-interest, that he should maximize his share of the free grazing by keeping on the commons the largest herd he could muster. The consequences must have been clear. Overgrazing would destroy the commons. Yet self-interest (as usual) took primacy over the perceived communal interest. The herdsmen suffered doubly: first, by the loss of overgrazed land, and second, by governmental control of what, until then, had been their own affair.

This parable of the commons is relevant in some ways to the present situation of the engineering community. I argue that we must subordinate immediate self-interest to long-term interests. We must recognize our interaction with society and understand that it is no longer possible for us to make decisions either affecting society or our indirect relationship with it without jointly having considered the potential repercussions. Unless we soon take responsibility for controlling the manner in which our skills are applied, we shall find them controlled by others—and quite probably in ways we might not wish.

By any standard, the United States is a wealthy nation. In the material sense this nation can surely afford to support technology at its prior growth rate, and will

* See the article written by Hardin[3]; for a not altogether convincing rebuttal, see Crowe's response.[4] Note that, in any case, the parable runs roughshod over historical complexities. The whole enclosure movement involved deep-seated economic pressures.

probably be able to do so for some time to come. Perhaps society cannot afford not to, for those problems with which we now are just beginning to grapple will ultimately require the application of substantially increased technology if they are to be brought under control. One cannot eat his cake and have it too. There is no going back to the idyllic innocence of prior times. With current population levels, society cannot abandon technology. In fact, more will be required to cope with the problems that technology has produced. I raise only the more obvious examples: Pollution control requires servo loops to optimize combustion, to monitor and perhaps control, in particular, pH, toxicity, and temperature levels; law and order, although primarily a sociopolitical problem, requires electronic equipment for the prevention and detection of crime as well as for aiding in the apprehension and prosecution of criminals; population explosions require improved agronomy and automated distribution systems; inner cities require pollution-free transportation systems using electronic aids for both guidance and control, and perhaps propulsion. As with any attempted prognosis of innovation, the list is not now denumerable. Though most thoughtful men would agree that more technological effort is necessary, engineers continue to change occupation and in doing so discourage others who once might have aspired to an engineering career.

To be blunt, the "conventional" wisdom has concluded that technology doesn't pay or, at least, that it pays insufficiently well to warrant prior levels of support. We as technologists know that conclusion to be dead wrong. Yet, as I have attempted to outline, there is a superficially attractive and quite unconsciously pursued line of reasoning that underlies this loss of faith in technology.

In part, we are blamed for problems we did not create, those that are essential to exponential growth and over which we can exercise little direct control. But we did not even try. Worse yet, we intensified incipient anti-technological attitudes by overselling our skills—in part because of claims made in our behalf by overzealous administrators seeking new contracts and expanded budgets. Yet again, we remained silent. Perhaps we aspired to become entrepreneurs and administrators too and thus were afraid to speak out; but, whatever the reason, we shirked our responsibility.

Specifically, we have defaulted in these ways:

1. The spin-offs inferred by the Congress from our man-in-space proposals just never materialized.

2. Communism was not contained by expensive, ineffectual weapons systems.

3. Air-traffic-control systems did not measure up to expectations.

4. The public has lost its confidence in our cost estimates.

5. Utilities have been operated in ways that have raised questions of subordinated public interest in pollution, safety, reliability, and forecast of requirements.

This list is far from exhaustive. Each of us can add or elaborate upon items; the precise details are not especially important. The point is that we have rather cavalierly added to the stresses imposed by technological proliferation, thereby lowering the tolerable growth rate. It is necessary, I argue, that we seek out a symbiotic relationship with society wherein each helps solve the other's problems. But we have not attempted to do so.

The initiative is still ours

Like herdsmen resenting governmental encroachment, we would much prefer, I believe, to assume the responsibility for regulating our own behavior than to have others do so for us. It is a matter both of pride and efficiency. Those of us, for example, who have actively participated in the DOD procurement cycle are perhaps better qualified to comment upon its present inequities than are outsiders who have had no direct experience. There is often no harsher judgment than that by peers.

At the very heart of the matter is how such regulation is to be effected. It would be presumptuous of me to present detailed plans of action when no consensus now exists recognizing the need for such action. Yet I would argue that it is long past time that the IEEE took upon itself the responsibility for policing the conduct of its membership in a manner not unlike the other professional associations—and do so in a manner cognizant of the complex, multifarious social pressures involved. Further, it would appear appropriate that representative committees be established for the purpose of setting standards of conduct to ensure a minimum of needless aggravation to society at large. Perhaps in extreme cases sanctions might even be applied to corporations, government agencies, and individuals who violated the promulgated standards and, conversely, groups as well as individuals might be rewarded for service in support of the greater needs of technology. Possibly, the IEEE, from time to time, and as the need arose, might prepare position papers for the purpose of calling public attention to potential problem areas. But I don't wish to encumber the principal argument with peripheral, mechanical issues that can be resolved in a reasonable manner by reasonable men. The important point is, I believe, that we accept our responsibility. Of course, there are those who say that the IEEE has become moribund, an ossified institution capable of fulfilling only its historical objectives. I hope they are wrong, for I am convinced that if we cannot control events soon we shall be controlled by them.

REFERENCES

1. Platt, J., "What we must do," *Science*, vol. 166, pp. 1115–1121, Nov. 28, 1969.

2. Richie-Calder (Lord), "Mortgaging the old homestead," *Foreign Affairs*, vol. 48, pp. 207–220, Jan. 1970.

3. Hardin, G., "The tragedy of the commons," *Science*, vol. 162, pp. 1243–1248, Dec. 13, 1968.

4. Crowe, B. L., "The tragedy of the commons revisited," *Science*, vol. 166, pp. 1103–1107, Nov. 28, 1969.

F. N. Eddy (M) attended Harvard College and received the B.S. degree in applied physics (1954). He also did graduate work at Boston University. Following army service and a brief stint at the Raytheon Corporation, he joined Laboratory for Electronics, Incorporated, and there, for eight years, he was involved in radar systems design and signal-processing techniques. Also during this period, he served as consultant on ultrasonic image conversion to the Army Research Office. Since 1965, he has been at the MITRE Corp. working on ionospheric radar signal processing. He is a member of AAAS.

Speech spectrograms using the fast Fourier transform

Increased flexibility and the capability for on-line analysis are the two primary reasons for utilizing a digital computer for the generation and display of speech spectrograms

Alan V. Oppenheim Massachusetts Institute of Technology

An important aid in the analysis and display of speech is the sound spectrogram, which represents a time–frequency–intensity display of the short-time spectrum.[1–3] With many modern speech facilities centering around small or medium-size computers, it is often useful to generate spectrograms digitally, on-line. The fast Fourier transform algorithm provides a mechanism for implementing this efficiently.

A principal reason for the importance of the sound spectrogram in speech analysis is that many speech sounds can be considered to be produced by exciting a resonant cavity, the vocal tract, with either a quasi-periodic or a noiselike excitation. For many applications, then, the speech waveform is characterized by the frequencies of the vocal tract resonances and, for the quasi-periodic excitation, the fundamental frequency of the excitation, both of which are readily apparent on a spectrogram.

Many modern speech research facilities center around small or medium-size computers, which provide a mechanism for carrying out sophisticated studies in speech analysis and synthesis.[4,5] With such a facility, it is sometimes useful to generate speech spectrograms on-line rather than by making an analog recording, which is then analyzed off-line by a spectrograph machine. Furthermore, it is often advantageous to closely relate time displays with spectral displays and to be able to choose bandwidths flexibly—perhaps even time-dependently—during the analysis. Such flexibility is ideally suited to computer implementation. With the use of the fast Fourier transform algorithm for computing spectrums digitally, speech spectrograms can be implemented on small computers with analysis times approximately the same as required with modern analog spectrographic equipment.

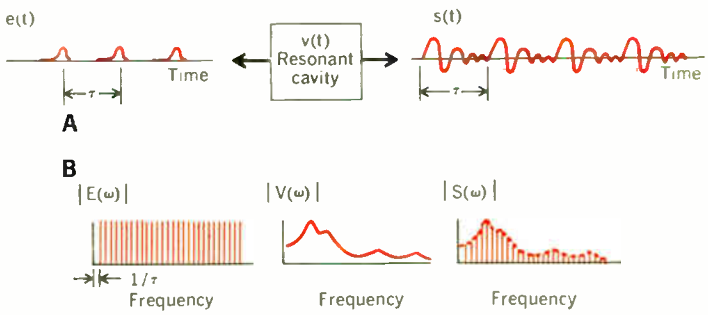
Spectral analysis of speech

A simplified but often useful model for speech production consists of a linear system with a quasi-periodic excitation function for voiced sounds (such as vowels) and a noiselike excitation during unvoiced (fricative) sounds.[6] The linear system represents the vocal cavity. The quasi-periodic excitation function during voiced sounds corresponds to the air flow through the vocal cords. During the production of a sustained vowel, the vocal tract configuration can be assumed to be fixed, and the sound produced corresponds to the response of a resonant cavity. Figure 1(A) represents the production of a steady-state vowel, where the excitation function $e(t)$ corresponds to air flow through the vocal cords, and the output $s(t)$ corresponds to air flow at the lips. If, for example, the vowel produced was the vowel /ah/ as in "father," a typical set of numbers for the first three resonances of the resonant cavity are 750, 1150, and 2500 Hz. Since the resulting vowel is periodic, its spectrum is a line spectrum with harmonics spaced in frequency by the reciprocal of the pitch period (typically about 125 Hz for a man). The envelope of the line spectrum will contain peaks corresponding to the resonant frequencies of the vocal cavity. Figure 1(B) illustrates pictorially the line spectrum corresponding to the excitation,* the spectral envelope corresponding to the frequency response of the vocal cavity, and the composite line spectrum corresponding to the spectrum of the steady-state vowel. During the production of fricative sounds such as /sh/, the excitation is a noiselike waveform produced by turbulence at the lips and teeth. Thus, for fricative sounds, the output is noiselike and has no line spectrum.

During the production of continuous speech, of course, the shape of the vocal cavity is not fixed, and the resonances vary to produce different sounds. Consequently, the linear system corresponding to the resonant cavity in Fig. 1 is time-varying, as is the excitation period. If the variation is not too rapid, it is reasonable to view the system as stationary on a short-time basis, so that a

* The line spectrum corresponding to the excitation will have an envelope representing the spectrum of an individual glottal pulse. For purposes of illustration, it was assumed that the spectrum of the glottal pulses was approximately constant.

FIGURE 1. Simplified picture of speech production.



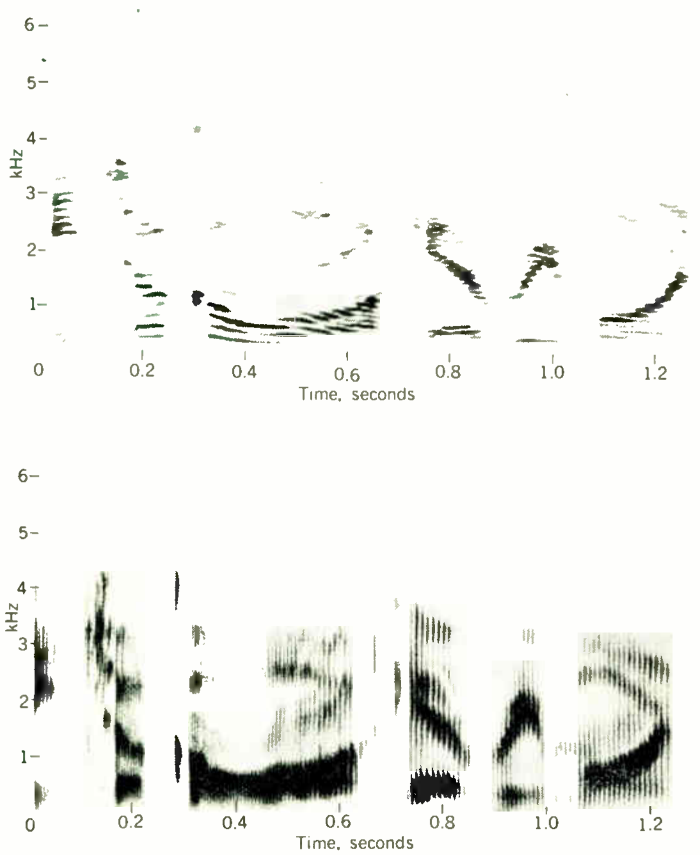


FIGURE 2. Spectrograms of the sentence, "He took a walk every morning," spoken by a male. A—Narrow-band spectrogram. B—Wide-band spectrogram.

FIGURE 3. Equivalent filter characteristic for rectangular time window.

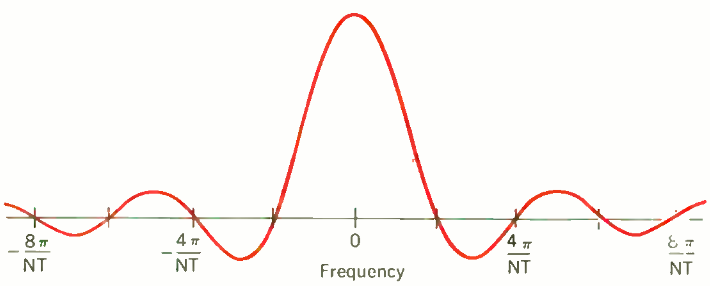
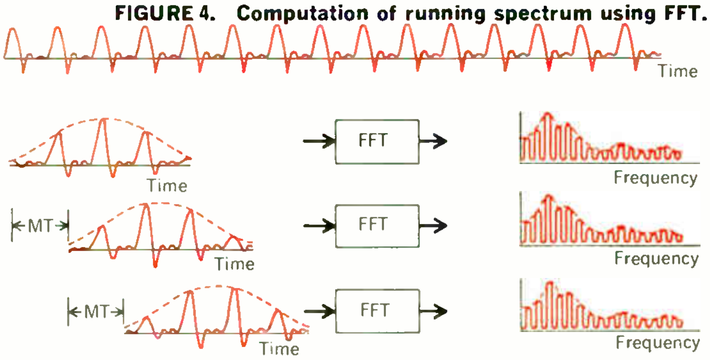


FIGURE 4. Computation of running spectrum using FFT.



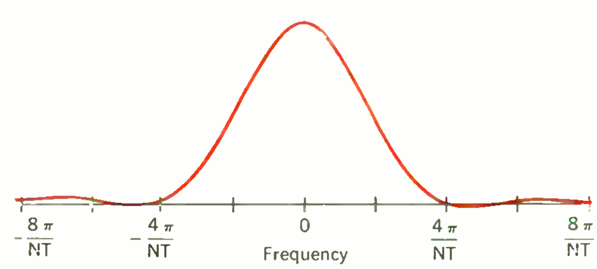
short-time spectral analysis of the speech waveform would exhibit peaks in the envelope corresponding to the resonances of the vocal tract as well as a harmonic structure corresponding to the excitation. As the time window is increased in length, frequency resolution improves and the harmonic structure becomes more evident, but the spectral analysis loses its ability to follow rapid changes.

A spectral analysis with a shorter time window, and consequently a wider frequency window, would provide better time resolution at the expense of spectral resolution; that is, it would tend not to resolve individual pitch harmonics in the spectrum but would be better able to track rapid changes. Because of this tradeoff between time and frequency resolution, it is common in spectral analysis of speech to utilize both narrow-band analysis, corresponding to good frequency resolution and poor time resolution, and wide-band spectral analysis, corresponding to good time resolution and poor frequency resolution. A typical means for obtaining and displaying speech spectrums is the spectrograph machine, for which the analysis corresponds to playing the speech through a bank of equal-bandwidth filters (usually implemented by heterodyning the signal past a single fixed filter). In a narrow-band analysis, the filter bandwidths are typically 45 Hz; for a wide-band analysis, they are 300 Hz. The recording is made on Teledeltos paper. Figure 2 shows narrow-band and wide-band spectrograms for the sentence, "He took a walk every morning," as spoken by a male. In the former, it is clear that the individual pitch harmonics have been resolved in frequency, whereas in the latter they are no longer evident. However, in the wide-band spectrogram, vertical striations can be seen that correspond to individual pitch periods. This is a consequence of the good time resolution of the wide-band spectrogram; they are not evident in the narrow-band case. Also evident on both wide- and narrow-band spectrograms are the frequency regions corresponding to high spectral amplitude designating the vocal tract resonances, referred to as formants.

Computation of spectrograms using the FFT

The fast Fourier transform (FFT) algorithm has assumed tremendous importance as a means for computing spectrums and implementing spectral displays on a digital computer.[7-10] In particular, for implementing on a digital computer spectral analysis of speech similar to that implemented by a spectrograph machine, it is considerably more efficient to carry out the analysis using the FFT algorithm than to implement a filter bank. The FFT is an algorithm for computing the discrete Fourier transform (DFT), defined as

FIGURE 5. Equivalent filter characteristic for Hanning window.



$$F(k) = \sum_{n=0}^{N-1} f(nT)e^{-j\frac{2\pi}{N}nk} \qquad (1)$$

where $f(nT)$ corresponds to equally spaced samples of an analog time function $f(t)$. Assuming that the sampling has been done at a rate equal to or higher than the Nyquist rate ($2f_m$, where f_m is the highest frequency in the analog time function), it can be shown that the magnitude of the kth spectral point $|F(k)|$ in Eq. (1) corresponds to the magnitude that would be obtained at a time $t = (N - 1)T$ when *samples* of the analog function $f(t)$ are played through an analog filter with a frequency response $H(\omega)$ given by

$$H(\omega) = \frac{\sin \frac{NT}{2}\left(\omega - \frac{2\pi k}{NT}\right)}{\left(\omega - \frac{2k}{NT}\right)} \qquad (2)$$

This filter characteristic is sketched in Fig. 3 for $k = 0$. The set of numbers $F(k)$, for k equal to 0 through $N - 1$, then corresponds to the set of outputs from a filter bank, each filter of which has a spectral shape similar to Fig. 3, with a center frequency at $\omega = 2\pi k/NT$.

The computation in Eq. (1) provides only one spectral section, that is, the output of the filter bank at a time $t = (N - 1)T$. To obtain a short-time spectral analysis, we would like to perform this computation at successive instants of time, and, in addition, to be able to modify the filter shape. For example, we may wish to reduce the sidelobes in the filter characteristic in Fig. 3. Furthermore, as we change from a wide-band to a narrow-band analysis, we would be required to change the width of the central lobe in the filter. To determine a running spectrum and provide flexibility in terms of the filter characteristic, the expression in (1) can be modified as

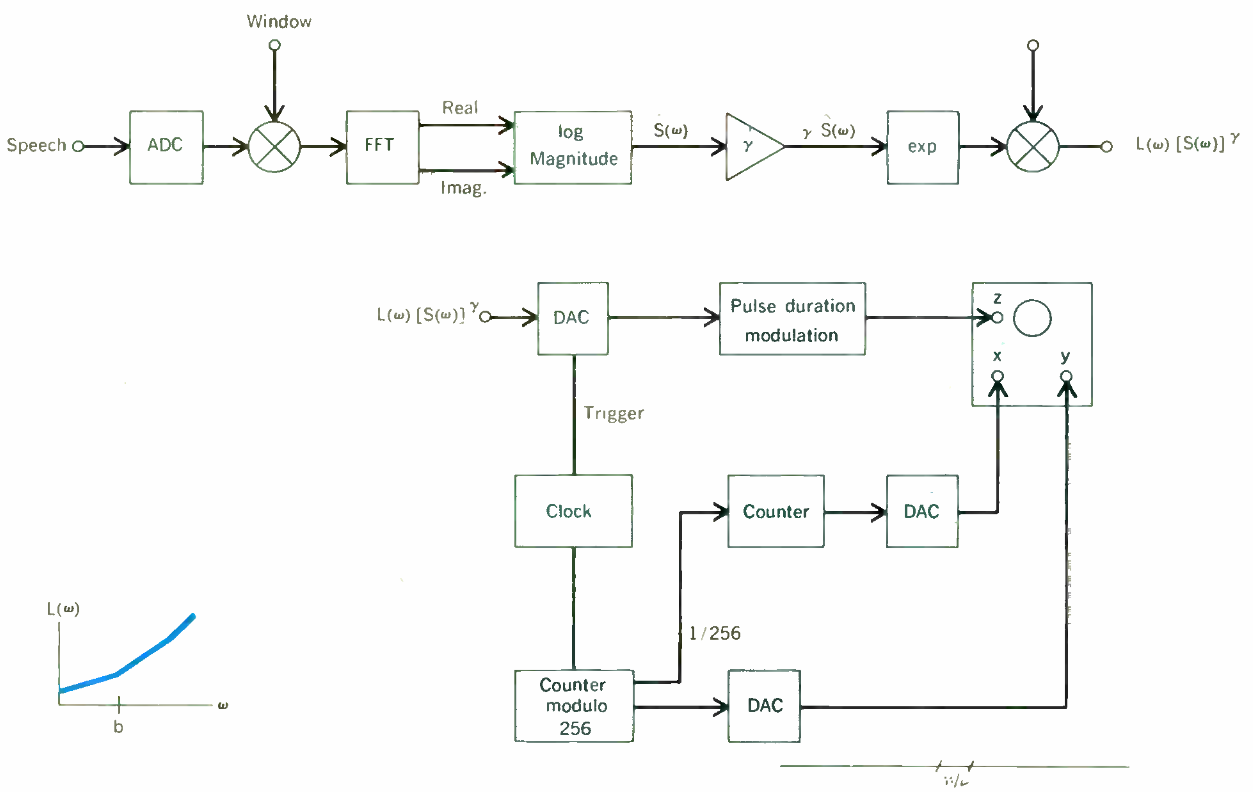
$$F_r(k) = \sum_{n=0}^{N-1} w(nT)f(nT + rMT)e^{-j\frac{2\pi}{N}nk} \qquad (3)$$

Equation (3) introduces two changes. The first is to include a window $w(nT)$ to provide a better spectral characteristic. This is motivated by the fact that since a computation of the discrete Fourier transform as given by Eq. (1) is necessarily restricted to a computation on a finite length of data, there is implicit in (1) a time window imposed on $f(t)$, that is, $f(t)$ is multiplied by a rectangular window with a width equal to NT. It is that rectangular time window that leads to the spectral window shown in Fig. 3. By modifying the rectangular window with some new time window $w(nT)$, it is possible to modify the spectral shape shown in Fig. 3. The second modification incorporated in Eq. (3) corresponds to implementing a spectral analysis of successive sections of the waveform. In other words, the set of numbers $F_r(k)$ represents a computation of the discrete Fourier transform of a section of the analog time function starting at $t = rMT$ and ending at $t = rMT + (N - 1)T$. This corresponds to a filter bank output at time $t = rMT + (N - 1)T$. Successive sections (Fig. 4) are spaced in time by MT.

In a filter-bank implementation of the spectral analysis, the time window $w(nT)$ corresponds to the low-pass prototype of the impulse response of each of the filters. One observation from this is that the spectral analysis described by (3) corresponds to a filter-bank analysis for which the spectral shape of each of the filters in the filter bank is approximately the same. For example, Eq. (3) could not represent a filter bank having constant-Q filters, for which the bandwidth is proportional to the frequency. If a constant-Q analysis were desired, a direct implementation of the filters would be used.[11]

As an example of speech spectrograms obtained by

FIGURE 6. Block diagram for computation and display of spectrograms.



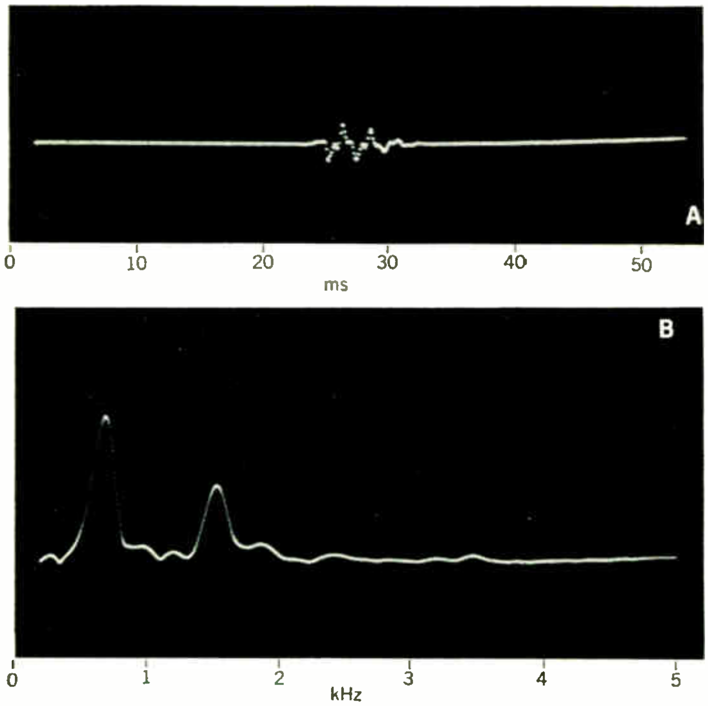


FIGURE 10. A—Speech segment with window applied for wide-band analysis. B—Spectral section from (A).

Figures 7 and 8 show some typical spectrograms obtained with the foregoing system. The sentence again is "He took a walk every morning," as spoken by a male. Figure 7 corresponds to narrow-band spectrograms: In Fig. 7(A), the parameter γ is unity and no frequency shaping has been applied; Fig. 7(B) corresponds to no frequency shaping but a γ of 2/3; Fig. 7(C) represents a γ of unity and linear shaping starting at 1.25 kHz with a slope of 1.6 (1/kHz). Figure 8 represents wide-band spectrograms: Fig. 8(A) corresponds to a γ of unity and no frequency shaping; Fig. 8(B) to no frequency shaping but a γ of 2/3; and Fig. 8(C) to a γ of unity and linear shaping starting at 1.25 kHz with a slope of 1.6 (1/kHz). Figure 8(D) is identical to Fig. 7(C), except for an expanded time scale. In Fig. 9(A) is a typical section of input after the time window has been applied; Fig. 9(B), which shows the magnitude of the DFT of that section, using the narrow band, represents a vertical pass across Fig. 7 at the abscissa marked by the arrows. Similarly, the time function and spectral cross section shown in Fig. 10 are for the wide-band spectrogram, and correspond to the arrows in Fig. 8.

Advantages of FFT-generated spectrograms

This article has discussed and illustrated a procedure for generating and displaying speech spectrograms on a digital computer. In the present system, the analysis time is roughly comparable to that obtainable with modern analog spectrographic equipment. Because of the inherent capabilities of digital computers, there is the potential for considerable flexibility with this method.

The primary advantages of obtaining spectrograms digitally are (1) the increased flexibility and (2) the ability to carry out on-line spectrographic analysis of speech that is being synthesized or processed digitally for other reasons. At present, it does not seem to be efficient or advantageous to carry out *routine* spectrographic analysis of *large* amounts of speech data on a sophisticated computer facility. However, as the cost of small computers and digital hardware decreases, it may eventually be practical and economical to reserve a small facility for preliminary analysis of speech signals, including displays of spectrums and time waveforms.

The author wishes to acknowledge the assistance of Don Johnson and Mrs. Ann Fried.

The work reported in this article was carried out in part at the M.I.T. Lincoln Laboratory (sponsored in part by the U.S. Air Force), and partly at the M.I.T. Research Laboratory of Electronics [supported in part by the U.S. Air Force (Office of Aerospace Research, under contract F19628-69-C-0044)].

REFERENCES

1. Koenig, W., Dunn, H. K., and Lacey, L. Y., "The sound spectrograph," *J. Acoust. Soc. Am.*, vol. 18, pp. 19–49, 1946.

2. Potter, R. K., Kopp, G. A., and Green, H. C., *Visible Speech*. Princeton, N.J.: Van Nostrand, 1947.

3. Kersta, L. G., "Amplitude cross-section representation with the sound spectrograph," *J. Acoust. Soc. Am.*, vol. 20, pp. 796–801, 1948.

4. Denes, P. B., and Mathews, M. V., "Laboratory computers: their capabilities and how to make them work for you," *Proc. IEEE*, vol. 58, pp. 520–531, Apr. 1970.

5. Schroeder, M. R., "Computers in acoustics: symbiosis of an old science and a new tool," *J. Acoust. Soc. Am.*, vol. 45, pp. 1077–1088, May 1969.

6. Flanagan, J., *Speech Analysis, Synthesis and Perception*. New York: Academic Press, 1965.

7. Cooley, J. W., and Tukey, J. W., "An algorithm for the machine computation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, Apr. 1965.

8. Gold, B., and Rader, C., *Digital Processing of Signals*. New York: McGraw-Hill, 1969.

9. Bergland, G., "A guided tour of the fast Fourier transform," *IEEE Spectrum*, vol. 6, pp. 41–52, July 1969.

10. Rothauser, E., and Maiwald, D., "Digitalized sound spectrography using FFT and multi-print techniques (abstract)," *J. Acoust. Soc. Am.*, vol. 45, p. 308, 1969.

11. Arnold, C. E., "Spectral estimation for transient waveforms," presented at IEEE Arden House Workshop, Jan. 1970.

Alan V. Oppenheim (M) received the S.B. and S.M. degrees in 1961 and the Sc.D. degree in 1964, all in electrical engineering, from the Massachusetts Institute of Technology. From 1961 to 1964 he was a member of the M.I.T. Research Laboratory of Electronics and an instructor in the Department of Electrical Engineering. During this period his main activities centered on system theory and communication theory. His doctoral research involved the application of modern algebra to the characterization of nonlinear systems. In 1964 he became an assistant professor in the M.I.T. Department of Electrical Engineering and a staff member of the Research Laboratory of Electronics; in 1969 he was appointed to an associate professorship, after spending 1967–69 as a staff member of Lincoln Laboratory. His present research activities are in the areas of speech communication and digital waveform processing. Dr. Oppenheim is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, and the Acoustical Society of America. With coauthors R. W. Schafer and T. G. Stockham, he received the 1969 IEEE G-AE Senior Award.

Two new photorecording papers from Kodak

Kodak Linagraph Direct Print Papers, Types 1895 and 2022, could be your only choice when...

- you're faced with some laboratory recording, some field recording
- wet processing is impossible
- you need data instantly, and it must last indefinitely
- your recorders are direct-print

Now, if there's an event happening almost anywhere, you can get a high-quality tangible record. With new *Kodak Linagraph* 1895 (standard base) and 2022 (extra-thin base) Papers, you get readable traces the instant you expose to room light . . . crisp, dark traces a minute later. And the image is extremely stable.

No-heat, no-wet process *Linagraph* 1895 and 2022 Papers are designed for use with direct-writing cameras and oscillograph records equipped with high-intensity tungsten, xenon, or mercury-vapor light sources. They produce high-contrast images that are easy to read, reproduce well on most office copiers.

This paper is one of many products designed by Kodak scientists and engineers to solve data-recording needs through photography. The materials and technology to meet your specific requirements are probably available right off the Kodak shelf. Want more information? Just mail the coupon or call our photographic specialists at (716) 325-2000, Ext. 5114.

Department 412L
EASTMAN KODAK COMPANY
Rochester, N. Y. 14650

☐ Send information on *Kodak Linagraph* 1895 and 2022 Papers.

☐ Send new "Kodak Products for Instrumentation" catalog.

☐ Have representative call.

Name

Position

Firm Name

Address

City

State Zip

8-104

Kodak

Circle No. 20 on Reader Service Card

Circle No. 21 on Reader Service Card →

3 IMD SOLUTIONS FOR FREQUENCY CONVERSION PROBLEMS!

1 $ 52 each 1–4 units

Considering the thousands we've sold since 1966, our Model M1 must be one of the best double-balanced mixers available. As a phase detector, pulse or amplitude modulator, up/down frequency converter, or a mixer, it's practically become an industry standard. Typically, at 100 MHz, its isolation is greater than 50 dB and its conversion loss and noise figure is only 6.0 dB. Excellent balance provides very good suppression of the two IMD problems, namely $(2f_{R2}-f_{R1})\pm f_L$ two-tone products, and $mf_L \pm nf_R$ harmonic intermodulation products. And, as with all RELCOM components, every published spec is guaranteed from −54°C to +100°C! The Model M1 has almost everything you need in a broadband mixer. Well, almost.

1D $150 each 1–4 units

9B $100 each 1–4 units

Because, what's really needed to suppress severe intermodulation distortion is a higher level mixer.

That's the reason we invented the dc to 500 MHz Model M1D and its p.c. counterpart, the Model M9B. Designed to solve more severe IMD problems, these high-level units offer a +10 dBm conversion compression level for 18 dB extra two-tone suppression, plus superior f_R harmonic-intermod suppression. Furthermore, thanks to special Schottky-Barrier Diodes, the noise figure is typically 7.0 dB at 100 MHz — unusually low for a high-level mixer. Isolation at 100 MHz is greater than 60 dB. With that kind of performance, the M1D/M9B just about answers everybody's IMD problems.

1E $200 each 1–4 units

9E $150 each 1–4 units

But, to make sure, we also invented the dc to 400 MHz Model M1E, with the Model M9E for printed circuits. These hermetically-sealed mixers virtually eliminate all higher order f_R harmonic IMD, with a +21 dBm conversion compression to provide an extra 40 dB of two-tone suppression. At 100 MHz, the noise figure for the M1E/M9E is typically 7.0 dB (2.0 dB lower than the nearest FET mixer) and the isolation is greater than 55 dB.

We could say more. But we'd rather have you compare our high-level performance for yourself. Circle our Reader Inquiry Number, call our factory sales engineer, or better still, requisition your own evaluation units from our stock. It's easy. The prices are low. The call's on us.

. . . three more performance standards for reliable signal-processing components!

TYPICAL HARMONIC IMD

$mf_L \pm nf_R$ mixing products are indicated by the number of dB below the $f_L = f_R$ output. Data was obtained in a typical 50-ohm system. f_R = 49 MHz at 0 dBm, and f_L = 50 MHz. f_L levels are noted for each unit. Data obtained with f_R at −10 dBm for each unit is available on request.

Harmonics of f_R \ Harmonics of f_L	0	1	2	3	4	5	6	7	8
7	79	69	80	74	83	63	78	60	71
6	90	86	91	91	90	84	93	84	88
5	72	70	71	52	77	46	75	45	73
4	80	79	82	77	82	76	77	72	77
3	51	49	53	51	55	48	54	53	58
2	69	72	79	67	75	66	77	68	75
1	25	0	39	13	45	22	54	37	59
0		36	45	52	63	45	60	71	64

Model M1: f_L at +7 dBm. Performance will improve at lower frequencies, and with f_R at a lower level.

Harmonics of f_R \ Harmonics of f_L	0	1	2	3	4	5	6	7	8
7	>99	79	>99	78	>99	78	>99	81	99
6	>99	>99	>99	>99	>99	>99	>99	>99	>99
5	93	73	87	72	88	66	85	64	82
4	96	80	96	80	95	82	98	78	90
3	63	58	65	60	65	55	64	54	66
2	68	67	76	67	80	66	82	66	83
1	25	0	39	11	50	16	59	19	59
0		39	42	46	58	37	65	49	75

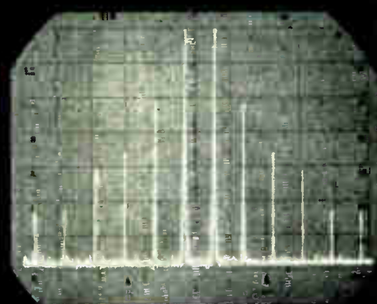
Model M1D: f_L at +17 dBm. With $n \geq 3$, the average harmonic intermodulation suppression of this high-level mixer will be 9 dB greater than a well-balanced, low-level, double-balanced mixer such as the RELCOM Model M1.

Harmonics of f_R \ Harmonics of f_L	0	1	2	3	4	5	6	7	8
7	>99	>99	>99	>99	>99	>99	>99	>99	>99
6	>99	>99	>99	97	>99	>99	>99	>99	98
5	>99	96	>99	95	>99	>99	>99	90	>99
4	88	91	>99	92	90	95	87	94	87
3	81	73	85	69	85	68	85	64	87
2	64	71	62	70	63	70	61	62	64
1	24	0	35	11	42	19	50	39	49
0		29	20	32	24	29	27	30	29

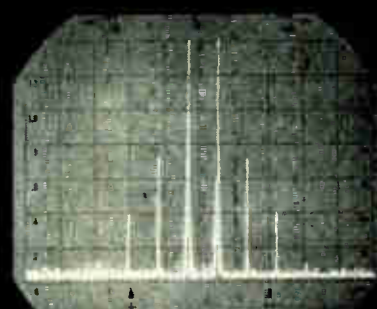
Model M1E: f_L at +27 dBm. With $n \geq 3$, the average harmonic intermodulation suppression of this ultra high-level mixer will be 25 dB greater than a well-balanced, low-level, double-balanced mixer such as the RELCOM Model M1.

TWO-TONE INTERMODULATION

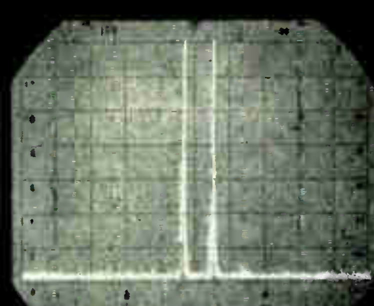
Spectrum displays of two-tone IMD suppression taken under the following conditions: Input Signals: f_L = 352 MHz, f_{R1} = 322 MHz at 0 dBm, f_{R2} = 320 MHz at 0 dBm. Horizontal Scale: 2.5 MHz/cm, centered at ≈31 MHz. Vertical Scale: 10 dB/cm.



Model M1: With a +7 dBm f_L pump signal, the well-balanced Model M1 suppresses many of the two-tone spurs. However, there are still a number of relatively unsuppressed products.



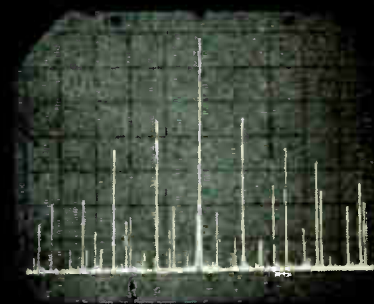
Model M1D: With a +17 dBm f_L pump signal, the Model M1D eliminates the higher order two-tone products from the 60 dB spectrum and attenuates the remaining spurs.



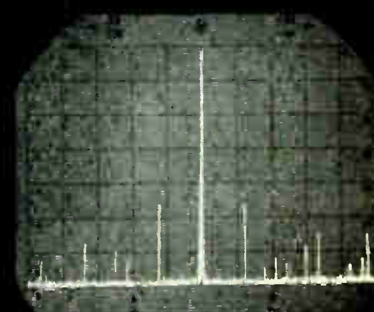
Model M1E: With a +27 dBm f_L pump signal, the Model M1E virtually eliminates all two-tone products from the 60 dB spectrum.

HARMONIC INTERMODULATION

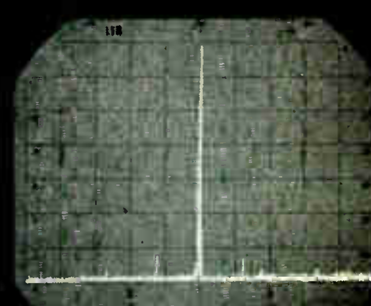
Mixer-Generated intermods are shown for each Model under the following conditions: Input Signals: f_{R1} = 45.6 MHz at −10 dBm, f_{R2} = 27.8 MHz at −2 dBm, f_{R3} = 26.6 MHz at −2 dBm, f_L = 50 MHz with the levels noted below. Horizontal Scale: ≈ 2 MHz/cm, centered at 95.6 MHz. Vertical 10 dB/cm.



Model M1: Here f_L is at +7 dBm. Mixer-generated harmonic intermodulation literally floods the spectrum.



Model M1D: With f_L at +17 dBm the harmonic modulation is reduced an additional 18 dB.

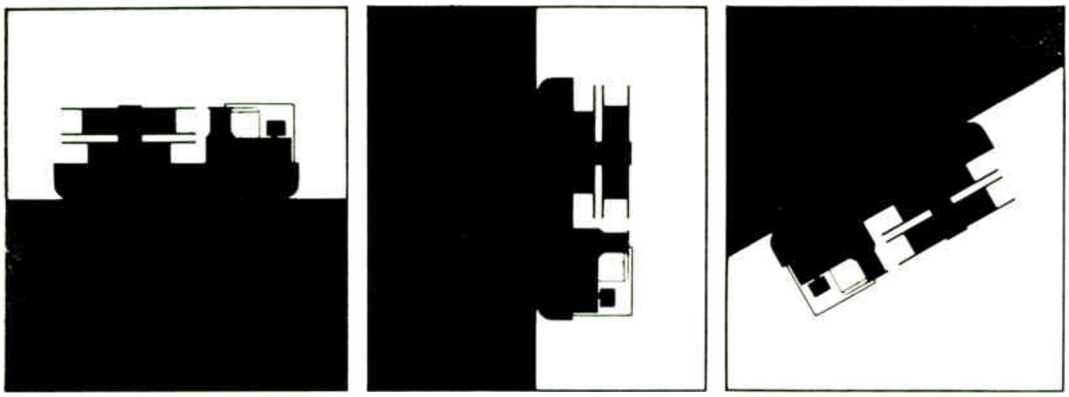


Model M1E: Now f_L is at +27 dBm and th harmonic intermodulation is reduced b nearly 40 dB.

RELCOM

2329 Charleston Road · Mountain View · California 94040

Telephone (415) 961-6265 · TWX (910) 379-6979



Westel's small television recorder operates on its side... on its back...

The WR 201 needs only 6/10 of a cubic foot and weighs just 37 pounds. Yet, it records television pictures of the highest professional broadcast quality: even during high g turns and other difficult maneuvers.

What's more, the WR 201 meets or exceeds the shock and acceleration requirements of Mil Standard 810B, as qualified by an independent test lab. It has also been tested to stringent RFI and EMI parameters.

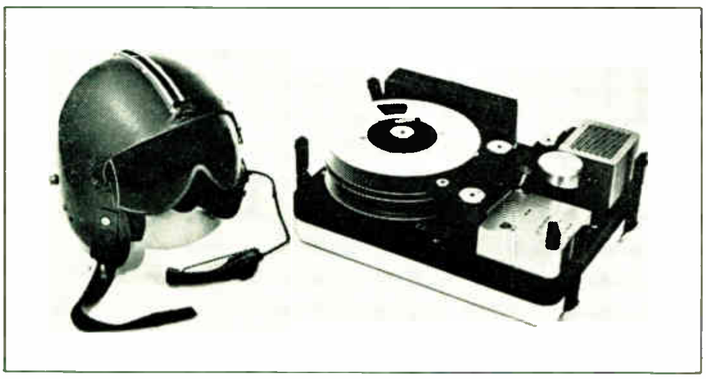
A WRR 301 ground television recorder/reproducer completes the system, giving you capability to handle many differing imaging applications calling for variations in bandwidth, information, and environment: night reconnaissance, using LLTV and FLIR techniques; training; missile guidance; and even studio broadcasting (the system meets all applicable time base stability requirements).

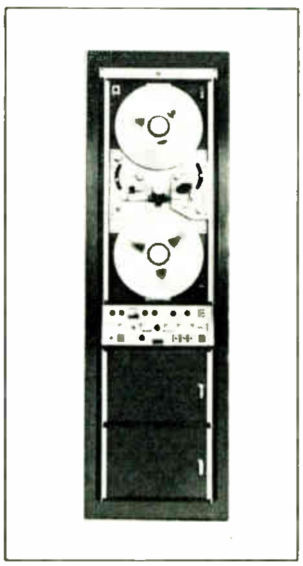
In addition, the WR 201 and the WRR 301 use the same economical one-inch tape and have interchangeable scanning assemblies.

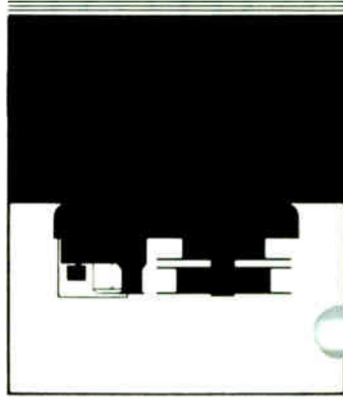
Standardized design yields further cost reductions. In most respects, the WR 201 and the WRR 301 are the same as the units made by Westel for commercial broadcasting applications.

Result? Assured availability of high performance equipment in various television line and field standards, with bandwidths as high as 6.5 MHz. You can even have NTSC color, if required.

The versatile Westel system has been thoroughly tested and has been flying missions in high performance aircraft for the past two years. For the full story, call, or write to: Government Marketing Manager, Westel Company, 2555 Charleston Rd., Mountain View, CA 94040/(415) 961-7145.







and at high g's

Westel Company

2555 Charleston Rd., Mt. View, Calif. 94040

Circle No. 22 on Reader Service Card

Noise in amplifiers

A number of factors contribute to the noise problem, which can drastically limit the effectiveness of amplifiers and other instrumentation, especially at low signal levels

Seymour Letzter, Norman Webster

Princeton Applied Research Corporation

In practically every type of research program in the physical sciences as well as in sophisticated engineering analyses, very small electrical signals must be measured and, in general, the limit of attainable precision and detectability is set by noise. This is true for the physicist and chemist performing nuclear magnetic resonance or spectroscopy experiments, for medical and biological researchers interested in evoked potentials, for geologists measuring small remanent magnetic fields in rock samples, for the metallurgist making Fermi surface measurements, and for the engineer performing vibration analysis and sensitive bridge measurements. These are only a few examples of applications in which noise plays a critical role in limiting measurement precision and signal detectability. This article discusses some of the inherent problems and describes techniques for improving signal-to-noise ratio.

Generally speaking, noise includes all those voltages and currents that accompany a signal of interest and obscure it. There are many different types of noise and they arise in many different sources. Many are directly electrical in nature, such as the noise produced by amplifiers and other instrumentation used to process signals. Others are not inherently electrical, but manifest themselves as electrical fluctuations when some element of the experimental system acts as a transducer (frequently without the knowledge of the experimenter). For example, structure vibration transmitted to coaxial cables can cause signals to be induced in the cables as a result of dimensional, and hence capacitive, changes in the cables. Some of the noises with which the experimenter must contend include

1. Thermal noise arising in the signal source impedance.
2. Noise produced in the instruments used to process the signal. In most instances, front-end preamplifier noise will dominate.
3. Environmental noise, which opens up a host of possibilities, including
   (a) Interference at the power frequency or its harmonics.
   (b) Automotive ignition noise.
   (c) Radio stations.
   (d) Lightning (and this can be remarkably distant).
   (e) Changes in barometric pressure.
   (f) Structure vibration.
   (g) Temperature fluctuations.
4. Statistical fluctuations resulting from the ultimately quantized nature of all measured quantities.

All of these, and many others, can limit the accuracy, precision, and useful sensitivity of measurements. Fortunately, only a few will have a significant effect in any one experiment, and the problem of the experimenter in minimizing noise effects will be less difficult than the foregoing list might lead one to believe.

The environmental interference frequently can be reduced to a negligible level by following "sound experimental practice,"[1] which includes proper grounding, shielding, guarding, and other procedures. With environmental factors under control, the experimenter still faces "quantization fluctuations," along with source-resistance noise and amplifier noise. The former only rarely will prove to be a problem, and where it is a problem there is little the experimenter can do other than to optimize his experiment so that the maximum number of "events" per unit time are measured. This leaves the final two types of noise, source thermal noise and amplifier noise, from which we arrive at the purpose of this article—namely, to provide some insight into the characterization of these noises and their effects, and to indicate how these noise effects can be minimized for a given experimental situation.

SNR at amplifier output terminals (voltage-source driven)

The SNR is a universally accepted quantitative expression for the degree of noise contamination of a signal. Figure 1 shows an amplifier fed from a voltage source E_s with a finite source resistance R_s. Before going on to calculate the SNR for this amplifier, we must first quantitatively characterize the source-resistance noise and the amplifier noise.

Johnson noise. Johnson noise is caused by random motion of thermally agitated electrons in resistive materials. Its instantaneous amplitude is unpredictable, but the probability that it will have an amplitude in an interval dV volts is given by $p(V)\,dV$, where $p(V)$ is the familiar Gaussian probability density function:

$$p(V) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-V^2/2\sigma^2} \tag{1}$$

where the parameter σ is the rms value of the fluctuations and the quantity universally accepted to describe the noise output from a resistor. It can be shown from thermodynamic considerations that the rms value is bandwidth-dependent as follows:

$$\sigma = E_{JN} = (4kTR_sB)^{1/2} \quad (2)$$

where k = Boltzmann's constant = 1.38×10^{-23} J/°K; T = resistor temperature, °K; R_s = resistance, ohms; and B = noise bandwidth, hertz. Johnson noise is "white noise"; that is, its rms value per unit bandwidth (rms density) is constant from dc to frequencies extending into the infrared region. For analytical purposes, the noisy resistor is represented by a noiseless resistor and a noise voltage or current generator as shown in Fig. 2.

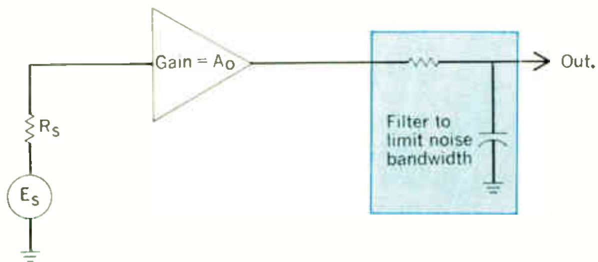


FIGURE 1. Amplifier fed from voltage source through source resistance.

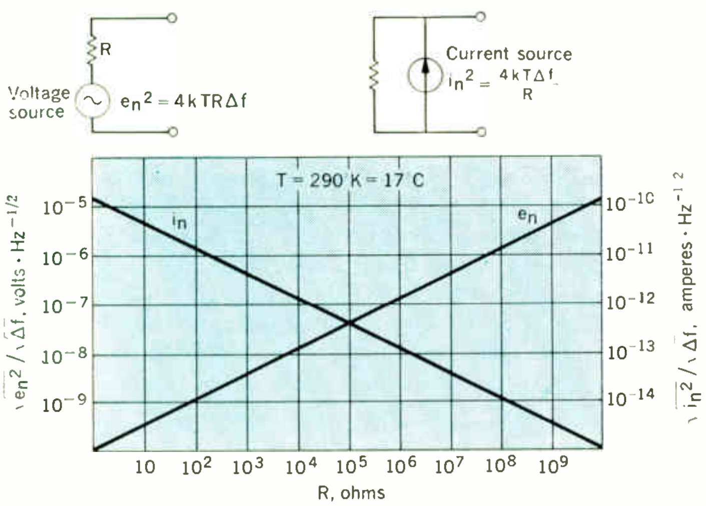
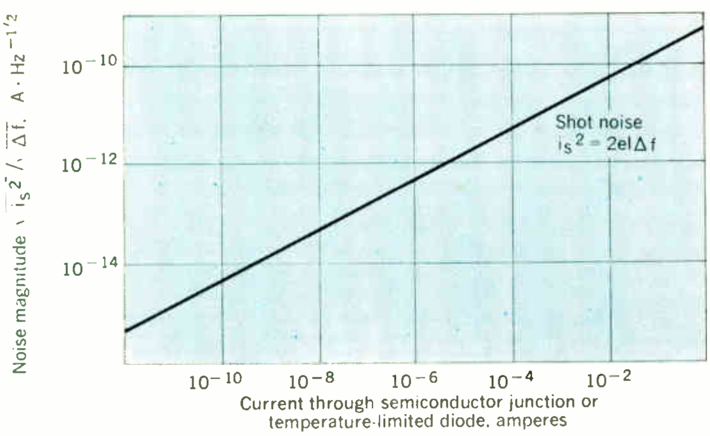


FIGURE 2. Equivalent-circuit representations of thermal noise in resistor R, together with magnitude of rms voltage and current per cycle as a function of R.

FIGURE 3. Shot noise as a function of current.



The source-resistance Johnson noise is the minimum possible noise that can accompany the signal. Other types of noise from other sources may obscure the signal as well, but the Johnson noise will always be present. Besides the Johnson noise that arises in the source resistance, there is additional Johnson noise produced by resistors in the amplifier. This noise also degrades the SNR. However, other sources of noise are also present in the amplifier, and contribute to the total amplifier noise, which is a conglomeration of thermal, shot, and flicker noise of resistors, vacuum tubes, and semiconductor devices.

Shot noise. Shot noise manifests itself as random current fluctuations in vacuum tubes and semiconductor junctions. It is caused by the random arrival of discrete electron charges at anodes, collectors, and drains. Equation (3) depicts the rms value of the shot-noise current in an emission-limited vacuum diode.

$$I_{\text{shot}} = (2eI_{\text{dc}}B)^{1/2} \text{ amperes rms} \quad (3)$$

where e = electronic charge = 1.59×10^{-19} coulomb; I_{dc} = average current through diode, amperes; and B = noise bandwidth, hertz.

Figure 3 shows the shot-noise-density variation of the direct current. Like thermal noise, shot noise exhibits a flat power spectrum, and thus can be referred to as "white" noise.

Flicker noise. Flicker noise is characterized by its spectral composition, and, for most electronic devices, it dominates thermal and shot noise from dc to about 100 Hz. Although flicker noise is detectable in virtually all conducting materials having power applied to them, it is most prominent where electron conduction occurs in granular or semiconductor devices and oxide-coated vacuum-tube cathodes. Composition, deposited carbon, and even metal-film resistors, in that descending order, all show voltage-dependent flicker noise. Flicker noise exhibits a $1/f^n$ power spectrum, with n typically (but not always) having a value in the range of 0.9 to 1.35.

The particularly disturbing thing about flicker noise is that the $1/f$ characteristic seems to hold down to as low a frequency as one cares to measure it. Even the long-term drift in dc transistor or vacuum-tube amplifiers would seem to be a manifestation of very-low-frequency flicker noise. As a result, flicker noise imposes a unique barrier to measurement accuracy. Consider the case where the signal of interest is at dc or at a low frequency in the flicker-noise-dominated frequency region. Because of the $1/f^n$ power-spectrum characteristic, no improvement in signal-to-noise ratio can be derived by increasing the time constant of the measuring device whenever n is equal to or greater than unity, the usual case—in contrast to operation in the white-noise-dominated frequency region, where a significant improvement can be achieved in this manner. It becomes apparent that small-signal measurements should be made at frequencies at which white-noise phenomenons (Johnson and shot noise) dominate, with dc measurements particularly to be avoided.

Total noise and SNR

The aforementioned amplifier noise sources can be lumped together, allowing the amplifier to be characterized as a noiseless amplifier and two fictitious noise generators—voltage and current—connected to the input terminals, as shown in Fig. 4. Noise-density generators e_n and i_n are expressed in units of rms volts/hertz$^{1/2}$ and rms amperes/hertz$^{1/2}$, respectively.

This form of characterization facilitates SNR and NF calculations, as we shall see later. We shall assume that e_n and i_n are white-noise sources and that their cross-correlation function is identically equal to zero. These assumptions, though by no means universally justifiable, do simplify the ensuing mathematics, and the resulting conclusions and statements are valid. Figure 5 illustrates the frequency dependence of these generators in an actual amplifier. Most manufacturers specify the noise content of the amplifier by the noise figure, whose definition is stated later in the article. In any case, whether given e_n and i_n, or NF, the experimenter can predict amplifier performance and optimize his experiment as shown in the following pages.

Because the noise sources are considered to be random and uncorrelated, noise *power* in a system is additive, and the total rms noise is the square root of the sum of the squares of each generator output. Note that the amplifier noise current is treated as a voltage by computing the voltage drop across the source resistance R_s; see Fig. 5. It is not necessary to consider the noise contribution of the low-pass filter resistor when the amplifier gain is large.

The total noise-output voltage is

$$E_{tno} = [4kTR_s + e_n^2 + (i_nR_s)^2]^{1/2}A_0 f_n^{1/2} \text{ volts rms} \quad (4)$$

where A_0 = midband gain and f_n = noise bandwidth = $RC/4$ for Fig. 4, in which R and C are values of output-filter components.

The signal at the output terminal is $E_s A_0$. Consequently, the output signal-to-noise voltage ratio $(\mathrm{SNR})_o$ can be determined by dividing $E_s A_0$ by Eq. (4). The result is as follows:

$$(\mathrm{SNR})_o = \frac{E_s}{[4kTR_s + e_n^2 + (i_nR_s)^2]^{1/2} f_n^{1/2}} \quad (5)$$

Often it is convenient to know the total equivalent noise referred to the amplifier input terminals (E_{tni}). This is easily obtained by dividing Eq. (4) by the midband gain A_0. The result is

$$E_{tni} = [4kTR_s + e_n^2 + (i_nR_s)^2]^{1/2} f_n^{1/2} \text{ volts rms} \quad (6)$$

At this point it should be clear that E_{tni} can be found by measuring the noise at the amplifier output terminal with a true rms voltmeter and dividing the reading by the gain A_0. It should be equally obvious that E_{tni} cannot be found by putting the voltmeter directly across the amplifier's input terminals.

It is evident from the third term of Eq. (4) that the noise contribution of the amplifier is dependent on the magnitude of the source resistance R_s. This leads us to the first of three statements that express the essence of these relationships.

Statement 1. For a given amplifier driven from a voltage source, the SNR is maximum when $R_s = 0$.

Equation (5) also shows the advantage of restricting the bandwidth. By allowing the signal of interest to be transmitted with no more of the high-frequency components getting through than are necessary to carry the "information," the SNR will be improved—hence the output low-pass filter shown in Figs. 1 and 4. Further improvement could be realized by also incorporating a high-pass filter to attenuate noise components below the signal frequency. Bandpass-selective amplifiers having a fast rolloff above and below the signal frequency can frequently be used to great advantage.

Noise figure

A popular figure of merit used to describe an amplifier's quality, insofar as noise is concerned, is the noise figure of the amplifier. Relevant to the circuit in Fig. 4, noise figure, expressed in decibels, can be defined as follows:

$$\mathrm{NF} = 20 \log_{10}\left[\frac{\text{Input voltage SNR (amplifier disconnected)}}{\text{Voltage SNR at amplifier output terminals}}\right] \quad (7a)$$

or, in terms of power,

$$\mathrm{NF} = 10 \log_{10}\left[\frac{\text{Input power SNR (amplifier disconnected)}}{\text{Power SNR at amplifier output terminals}}\right] \quad (7b)$$

With the aid of Fig. 4 and Eq. (5), and by expressing all parameters as power, Eq. (7b) can be written as

$$\mathrm{NF} = 10 \log_{10}\left[\frac{E_s^2/4kTR_sf_n}{E_s^2/[4kTR_s + e_n^2 + (i_nR_s)^2]f_n}\right]$$

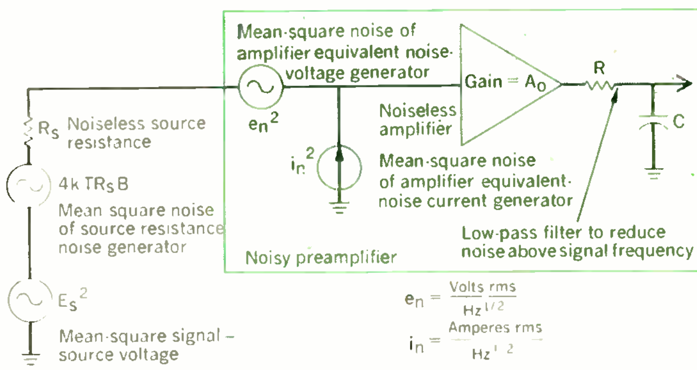
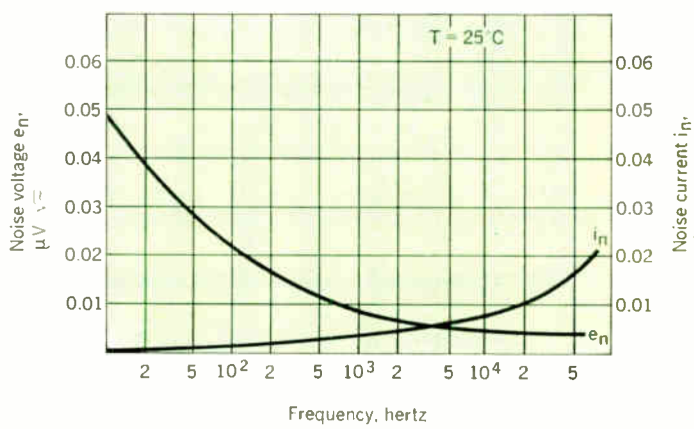


FIGURE 4. Equivalent circuit for noisy amplifier fed from signal E_s with source resistance R_s.

FIGURE 5. Noise current and voltage as a function of frequency for a typical low-noise preamplifier.



Further simplification yields

$$NF = 10 \log_{10}\left[1 + \frac{e_n^2 + (i_n R_s)^2}{4kTR_s}\right] \qquad (8)$$

Akin to noise figure is noise factor, F, defined as

$$F_{\text{power}} = 10^{NF/10} = \left[1 + \frac{e_n^2 + (i_n R_s)^2}{4kTR_s}\right] \qquad (9)$$

or, in terms of voltage,

$$F_{\text{voltage}} = 10^{NF/20}$$

Continuing, the total equivalent noise referred to the amplifier input terminals (E_{tni}) can be shown to be

$$E_{tni} = [4kTR_s f_n]^{1/2} 10^{NF/20} \text{ volts rms} \qquad (10)$$

which brings us to our second statement.

Statement 2. For a *given source resistance*, the least noisy amplifier is the one with the smallest NF. A noiseless amplifier ($e_n = i_n = 0$) has an NF of 0 dB.

Let us now refer back to Eq. (8) as the basis for the next statement.

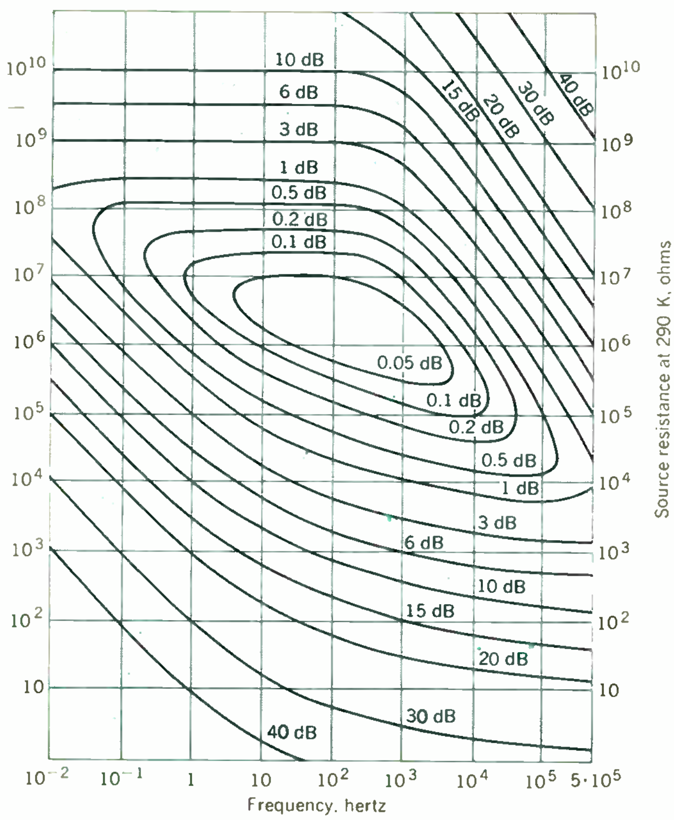
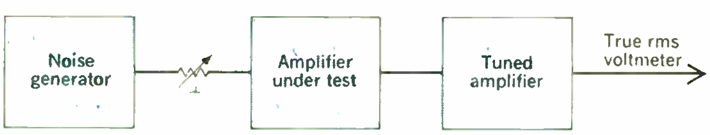


FIGURE 6. Noise-figure contours for Model 113 preamplifier.

FIGURE 7. Equipment hookup for measuring noise figure.



Statement 3. The noise figure increases without limit as the source resistance approaches zero.

Compare statements 1, 2, and 3. Initially, they may seem to appear contradictory. This apparent paradox is easily resolved if one remembers that the noise figure is only a measure of comparison of amplifier noise with the thermal noise developed in an arbitrary source resistance and nothing more. In fact, it is the authors' experience that many researchers overemphasize the amplifier noise figure and neglect consideration of minimum experimental signal. The following example illustrates this point:

Assume that a given experiment provides a 1-μV rms signal and that it is necessary to choose which of two amplifiers will be used to process the signal. The first amplifier, a moderately priced unit, has a noise figure of 20 dB. The second, a very expensive instrument, has a noise figure of 3 dB. Assume the source resistance to be 100 ohms and the noise bandwidth to be 100 Hz. Equation (10) will be used to compute the total input noise for each of these amplifiers. Tabulating the required data, we obtain:

Amplifier 1	Amplifier 2
NF = 20 dB	NF = 3 dB
R_s = 100 ohms	R_s = 100 ohms
f_n = 100 Hz	f_n = 100 Hz

From Eq. (10) the total equivalent input noise (E_{tni}) for amplifier 1 is 130 nV and for amplifier 2 it is 18 nV. Selecting amplifier 2 over amplifier 1 seems logical, but this decision may be uneconomical in light of the assumed 1-μV signal level.

Noise-figure contours

Noise-figure contours are essentially the locuses of points of constant noise figure as a function of source resistance and operating frequency. They allow the user to determine suitable points of operation, such as source resistance and frequency. Equivalent input noise and signal-to-noise ratios can be determined by using the contours in conjunction with Eq. (10). Figure 6 shows the noise figure contours for a PAR Corp. Model 113 preamplifier.

Figure 7 is a simplified sketch of the system used to measure noise figure. A noise generator provides a white-noise source calibrated in microvolts/hertz$^{1/2}$. When the noise generator is shut off, the source resistor R_s and the amplifier produce a voltage reading on the true rms voltmeter of x volts rms. The noise generator is then turned on and its noise voltage is increased until the voltmeter reading is 1.414 times the previously taken x reading. In other words, a calibrated noise voltage equal to E_{tni} is added to E_{tni}, allowing the value of E_{tni} in volts/hertz$^{1/2}$ to be read directly from the calibrated noise source. The following equation can then be used to compute the noise figure at one frequency and for one value of R_s:

$$NF = 10 \log_{10}\left[\frac{(E_{tni})^2}{4kTR_s}\right] \qquad (11)$$

By varying R_s while maintaining a fixed center frequency on the tuned amplifier, one can determine values of noise figure as a function of R_s. By varying the tuned-amplifier center frequency while holding R_s constant, one can determine values of NF as a function of frequency. All values of NF can then be plotted on a single graph to obtain the noise figure contours.

Optimum source resistance and minimum noise factor

In this section we are concerned with finding the value of a source resistance (R_{opt}) that, when "inserted" across the amplifier's input terminals, will yield the minimum noise factor (F_{min}). This investigation is important; the next section will show that transforming an arbitrary source resistance to the optimum source resistance, via a transforming device, will maximize the SNR.

R_{opt} can be found by differentiating Eq. (9) with respect to R_s, setting the derivative to zero, and solving for R_s, as follows:

$$\frac{\partial F}{\partial R_s} = 0 = 4kT \frac{[2(i_nR_s)^2 - e_n^2 - (i_nR_s)^2]}{\text{Denominator } (R_s)} \qquad (12)$$

$$R_{opt} = \frac{e_n}{i_n} \text{ ohms} \qquad (13)$$

Minimum noise factor (power) is found by substituting Eq. (13) into Eq. (9), resulting in

$$F_{min} = 1 + \frac{e_n i_n}{2kT} = 1 + \frac{2e_n^2}{4kTR_{opt}} \qquad (14)$$

Throughout this article we have assumed e_n and i_n to be frequency-invariant; Fig. 8 shows how R_{opt} and NF_{min} vary with frequency when e_n and i_n vary as shown in Fig. 5.

SNR improvement by properly matching R_s to preamplifier

In this section we will show that transforming the voltage source's intrinsic source resistance (R_s) to the optimum resistance (R_{opt}) via a matching transformer, as shown in Fig. 9, achieves a significant improvement in SNR. To simplify the mathematics, we will assume that the transformer is ideal; that it has infinite self-inductance, that it is lossless, and that it has infinite bandwidth. In addition, we assume the amplifier input impedance to be much greater than the reflected source impedance. Let us now proceed to compute the signal-to-noise improvement (SNI) obtainable by using the transformer. To begin, let $a = N_s/N_p$, where N_s and N_p represent the number of secondary and primary turns, respectively. Then the optimum source resistance reflected across the amplifier input terminals is

$$a^2R_s = R_{opt} = \frac{e_n}{i_n}$$

We define the signal-to-noise improvement factor (SNI) as follows:

$$\text{SNI} = \frac{\text{SNR using transformer}}{\text{SNR without transformer}} \qquad (15)$$

If the following analysis shows SNI to be greater than one, the transformer improves the SNR. Therefore,

$$\text{SNI} = \frac{a^2E_s^2}{4kTR_{opt} + e_n^2 + (i_nR_{opt})^2} \div \frac{E_s^2}{4kTR_s + e_n^2 + (i_nR_s)^2}$$

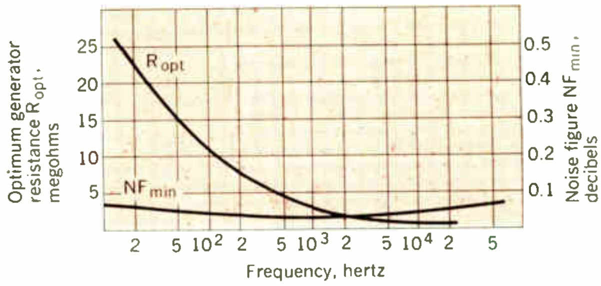


FIGURE 8. Variation of R_{opt} and NF_{min} with frequency for a low-noise preamplifier.

$$\text{SNI} = \underbrace{\left[1 + \frac{e_n^2 + (i_nR_s)^2}{4kTR_s}\right]}_{F_{\text{unmatched}}} \div \underbrace{\left[1 + \frac{2e_n^2}{4kTR_{opt}}\right]}_{F_{\text{min}}}$$

We achieve the result that the SNI equals the unmatched noise factor divided by the minimum noise factor:

$$\text{SNI} = \frac{F_{\text{unmatched}}}{F_{\text{min}}} \qquad (16)$$

It is important to remain consistent with respect to voltage and power when using these equations. In other words, both values of F must be expressed as power to obtain SNI (power), and both values of F must be in voltage to obtain SNI (volts). Usually it is most convenient to determine SNI (voltage) by computing the square root of SNI (power).

Illustrative example. Given a typical low-noise preamplifier to be driven from a low (10-ohm) source resistance. Data supplied by the manufacturer indicate e_n and i_n for the amplifier at the intended operating frequency: $e_n = 10^{-8}$ V/Hz$^{1/2}$ and $i_n = 10^{-13}$ A/Hz$^{1/2}$.

The problem is to pick a transformer that will yield maximum SNR improvement and to calculate that SNR improvement.

From the preceding discussion,

$$R_{opt} = \frac{e_n}{i_n} = \alpha^2R_s = \frac{10^{-8}}{10^{-13}} = 10^5 \text{ ohms}$$

A turns ratio of 1:100 ($a = 100$) is required to transform R_s to R_{opt}. The noise factor for the amplifier without the transformer can be computed by means of Eq. (9), using the given data. Thus,

$$F_{\text{unmatched}} = 1 + \left[\frac{10^{-16} + (10^{-26} \times 10^2)}{1.66 \times 10^{-20} \times 10^1}\right] \approx 600$$

$$(\text{NF} \approx 28 \text{ dB})$$

The noise factor for the amplifier when the transformer is used can be computed from Eq. (14) as follows:

$$F_{\text{min}} = 1 + \left[\frac{10^{-8} \times 10^{-13}}{0.83 \times 10^{-20}}\right] \approx 1.21 \quad (\text{NF} \approx 0.5 \text{ dB})$$

From Eq. (16),

$$\text{SNI}_{\text{power}} = \frac{600}{1.12} = 536 \quad (\text{SNI voltage} = \sqrt{536} \approx 23)$$

It is not always possible to achieve F_{min}, but a compromise can be reached such that, in general, the SNR improvement can be shown to be

$$\text{SNI} = \frac{F_{\text{unmatched}}}{F_{\text{matched}}} \quad (17)$$

where $F_{\text{unmatched}} > F_{\text{matched}} \geq F_{\text{min}}$. Or, in terms of NF,

$$\text{SNI}_{\text{power}} = 10^{(\text{NF}_{\text{unmatched}} - \text{NF}_{\text{matched}})/10} \quad (18a)$$

and

$$\text{SNI}_{\text{voltage}} = 10^{(\text{NF}_{\text{unmatched}} - \text{NF}_{\text{matched}})/20} \quad (18b)$$

It must be pointed out that it is impossible to achieve the SNI indicated by Eq. (16); the imperfections of all real transformers serve to reduce the expected SNI. Contributing sources of noise include winding-resistance thermal noise, Barkhausen noise (which results from the behavior of the magnetic domains in the core material when a signal is applied), mechanical stresses (including vibration), and sensitivity to pickup or interference from magnetic fields. This last source proves particularly bothersome at the power-line frequency and its lower-order harmonics. Heavy shielding is often required to reduce this pickup to an acceptable level. Frequently the physical orientation of the transformer also proves instrumental in achieving minimum pickup. Sensitivity to mechanical stresses and vibration does not often prove to be a serious problem, and where it does, supporting the transformer by suitable shock-absorbent material is usually an adequate solution. Most commonly, the winding resistance thermal noise proves to be the most significant source of transformer noise. Its magnitude is easily computed. The winding resistance of the secondary winding is divided by the square of the turns ratio, and that resistance is added to the primary winding resistance. For example, a 1:100 transformer having a primary winding resistance of 1 ohm and a secondary resistance of 10 000 ohms would have a thermal resistance of 2 ohms referred to the primary (20 000 ohms referred to the secondary). Despite the additional noise introduced by incorporating a transformer into the system, the SNI is increased sufficiently to make use of the transformer well worthwhile.

Figure 10 shows the noise figure contours for a PAR Model 116 preamplifier. The contours on the left were measured under "direct input" conditions. Those on the right were measured with the amplifier's internal transformer incorporated into the input circuit. Front-panel switching allows the transformer to be introduced for improved SNR when working from low source resistances. The following example shows the degree of SNR improvement attainable by using the Model 116 preamplifier with the transformer in a low-source-resistance experiment:

Assuming a small signal at 100 Hz from a source resistance of 100 ohms, the problem is to choose between operating with or without the transformer, and to compute the signal-to-noise improvement.

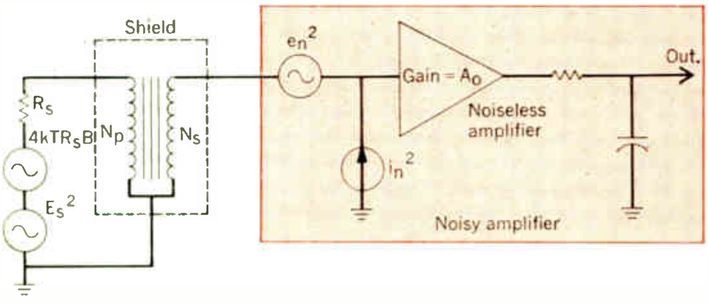
From Figure 10, the preamplifier, directly coupled, has an NF of 15 dB under the given conditions, whereas the same preamplifier, transformer-coupled, has an NF of only 2 dB. Clearly, transformer coupling should be used. By employing Eq. (18b) to compute SNI, we obtain

$$\text{SNI}_{\text{voltage}} = 10^{(\text{NF}_{\text{unmatched}} - \text{NF}_{\text{matched}})/20} = 10^{(15-2)/20} = 4.5$$

The computed SNI in this example will be significantly more accurate than that obtainable in the earlier example, in which it was necessary to work from e_n and i_n and to assume a perfect transformer. The results obtained were idealized and could be approached only under the best of conditions. In the latter example, however, the raw data were measured values of noise figure obtained with the transformer installed, and these values fully reflect the transformer imperfections.

A transformer's passband becomes narrower with increasing source resistance. Therefore, it is essential that the chosen matching transformer's frequency response be relatively flat within the frequency range of interest. Otherwise, the SNI expected may differ significantly from that obtained. We have shown that, for a given voltage source and amplifier, the SNR is maximized when $R_s = 0$. In practice, $R_s \neq 0$ because all signal sources have some value of resistance associated with them. The SNR, however, can be enhanced by transforming a low value of R_s to a value as close to R_{opt} as practical considerations allow. This does not mean a physical resistor is to be inserted in series with R_s to obtain a cumulative resistance equal to R_{opt}. Increasing the source resistance in this manner causes a twofold degradation in the signal-to-noise ratio, as can be readily determined by referring to Eq. (6). First, there is a considerable increase in thermal noise, which results when the thermal noise in the series resistor is added to that developed in the original source resistance. Second, there is an increased influence of the amplifier's noise current because of the voltage drop of this current across the series resistor. The current effect can be particularly damaging, since it varies directly with the total source resistance, in contrast to the thermal noise, which varies as the square root of the source resistance.

FIGURE 9. Use of transformer to match R_s to R_{opt} of preamplifier.



Equivalent noise resistance and equivalent noise temperature

In addition to the widely used equivalent noise generator and noise-figure methods of expressing the noise performance of amplifiers, other methods are also used. One of the most popular is to state an amplifier's series noise resistance R_e and parallel noise resistance R_i, where these quantities are defined by the following expressions[2]:

$$R_e = \frac{e_n^2}{4kT} \quad (19)$$

and

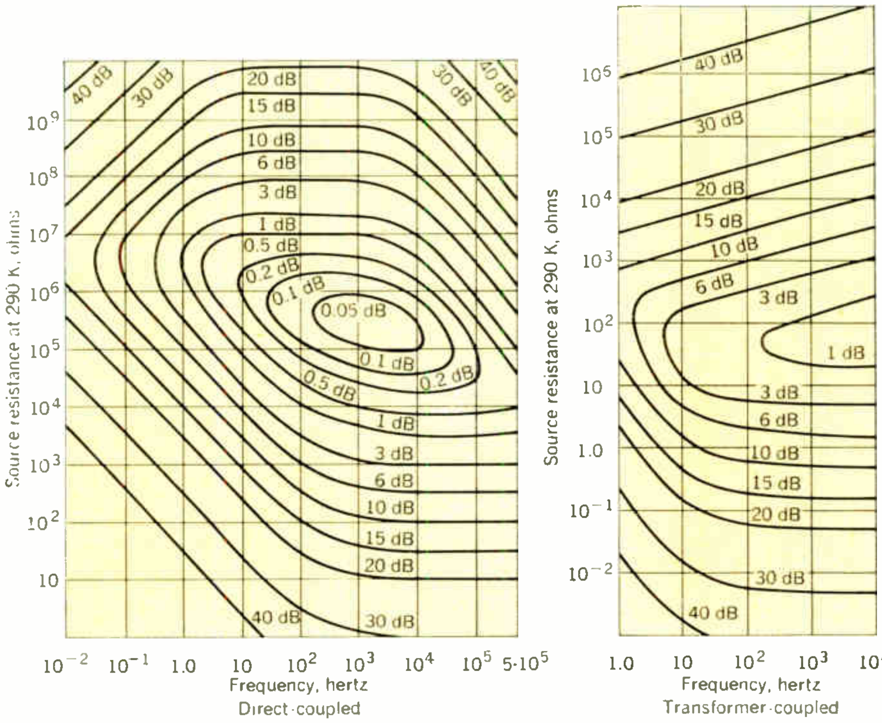


FIGURE 10. Noise-figure contours for Model 116 preamplifier.

$$R_i = \frac{4kT}{i_n^2} \tag{20}$$

where e_n = output of amplifier's equivalent noise-voltage generator; k = Boltzmann's constant; T = absolute temperature, °K; and i_n = output of amplifier's equivalent noise-current generator.

By the use of R_e and R_i, Eq. (8) takes the form:

$$\mathrm{NF} = 10 \log \left[1 + \frac{R_e}{R_s} + \frac{R_s}{R_i} \right] \tag{21}$$

Equation (21) makes it particularly easy to understand how NF varies as a function of R_s. Large noise figures are obtained with R_s either large or small. Minimum NF is obtained when the two terms containing R_s are equal. Differentiating with respect to R_s yields

$$R_{\mathrm{opt}} = \sqrt{R_e R_i} \tag{22}$$

Equation (22) in turn can be substituted into (21) to obtain another expression for minimum noise figure:

$$\mathrm{NF}_{\mathrm{min}} = 10 \log \left[1 + 2\sqrt{R_e/R_i}\right] \tag{23}$$

For the sake of comparison, refer back to Eqs. (13) and (14) to see how R_{opt} and NF are computed from e_n and i_n.

It is interesting to note that a plot of constant-noise-figure contours "contains" R_e and R_i to a good approximation in the 3-dB contour. Over the range in which the upper and lower curves are widely separated, the upper portion of the 3-dB contour represents R_i and the lower portion represents R_e, with the actual resistance values being read directly from the source-resistance scale. At the ends, where the curves converge and close, the approximation is inaccurate because R_e and R_i extend beyond the convergence region and, in fact, cross over, with the result that, at very high and very low frequencies, R_e is larger than R_i. The significance of this relationship is that it gives a convenient way of quickly evaluating the relative noise performance of two amplifiers, where the noise performance of one is expressed as a plot of R_e and R_i, and that of the other is expressed as a plot of contours of constant noise figure.

It is implicit in the preceding discussion of R_e and R_i that attenuating a signal source, either by means of a passive network, or by loading by the input resistance of the amplifier, will degrade the SNR and not just the signal level. This is worth stressing because of the common misconception that such attenuation "reduces the signal and noise equally, leaving the SNR unchanged."

Equivalent noise temperature T_e is another way of expressing the noise characteristics of an amplifier. The effective noise temperature T_e of an amplifier whose input is fed from a particular source resistance, is defined as the increase in source resistance temperature required to produce the observed available noise power at the output of the amplifier, the amplifier being noiseless.[3] Like noise figure, equivalent noise temperature is a function of source resistance and frequency, and thus we can plot contours of equivalent noise temperature that define the noise performance of an amplifier in much the same manner as do contours of constant noise figure. Unlike noise figure, equivalent noise temperature is *not* a function of the source-resistance temperature. Figure 11 is a plot of the contours of constant noise temperature for the amplifier whose noise-figure contours are plotted in Fig. 6. The expression relating the equivalent noise temperature and noise figure is

$$T_e = 290(10^{\mathrm{NF}/10} - 1) \tag{24}$$

Note that the 290 value in this equation is the absolute temperature of the NF measurement. If the NF had been measured at some other temperature, that temperature would be used in place of 290.

The utility of the noise-temperature method of expressing amplifier noise performance becomes clear by considering a case in which the source resistance is at a different temperature than the amplifier—for example, in cryogenic research. Suppose one wanted to estimate E_{tni} for an amplifier, first from the noise figure, and second from the equivalent noise temperature. The appropriate equations are

$$E_{tni} = [4kTR_s f_n]^{1/2}\, 10^{NF/20} \text{ volts rms} \qquad (10)$$

and

$$E_{tni} = [4k(T_s + T_e)R_s f_n]^{1/2} \text{ volts rms} \qquad (25)$$

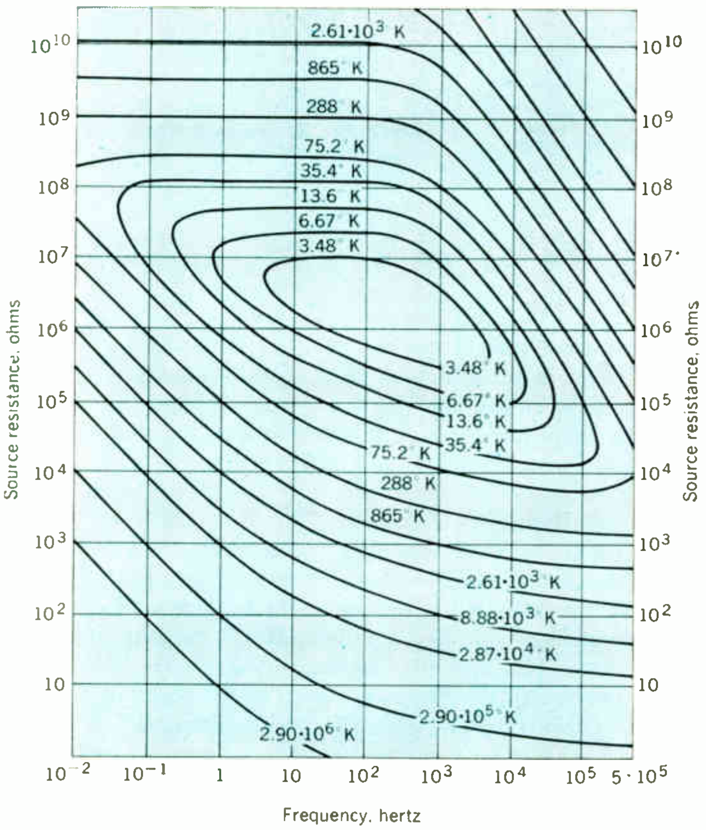
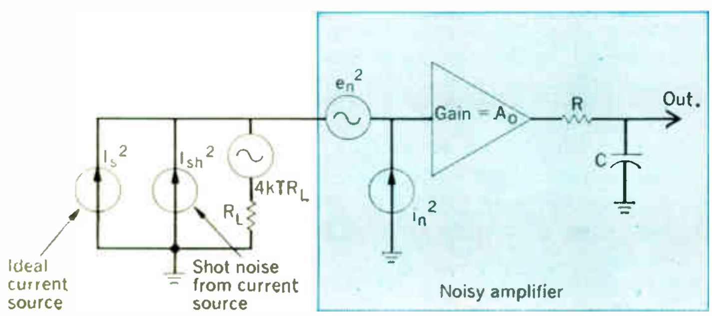


FIGURE 11. Contours of constant equivalent noise temperature for Model 113 preamplifier.

FIGURE 12. Equivalent circuit for noisy amplifier fed from current signal I_s with load resistance R_L.



where T_s is the source-resistance temperature and T_e is the amplifier equivalent noise temperature, both in degrees Kelvin.

If the source resistance and the amplifier were at the temperature of the NF measurements, either equation would quickly yield the desired result. However, if the amplifier were at room temperature and the source resistance at liquid-helium temperature, only Eq. (25) could be used to compute the total noise referred to the input directly. Equation (10) could be used only if NF were reevaluated for the low source-resistance temperature. The required computation, though straightforward, is lengthy, and contains many more opportunities for error than does Eq. (25). In general then, when the noise performance data are in terms of noise figure and it is necessary to operate at temperatures other than that at which the noise-figure measurement was made, it is advisable to use Eqs. (24) and (25) to determine E_{tni}.

The following example further illustrates the advantages of T_e over NF to the cryogenic researcher. Consider the researcher who wishes to optimize an experiment to be conducted at 1 kHz at a source-resistance temperature of 4°K. A cursory glance at Fig. 6 could mistakenly lead him to believe that excellent noise performance is obtainable over a wide range of source-resistance. At 1 kHz, the noise figure changes only 1 dB as R_s varies from 10^4 to 10^6 ohms. Figure 11, however, shows that the noise temperature changes from 75.2°K to 3.48°K, a marked difference, as R_s varies over the same range. It is obvious from Fig. 11, if not from Fig. 6, that at cryogenic temperatures it is particularly important to operate from the optimum source resistance, which, as explained earlier, may be obtained by following a low-resistance sensor by a suitable step-up transformer. Of course, if one were to "correct" Fig. 6 by using Eq. (24), substituting "4" for "290," this advantage would be equally obvious from the NF contours; at 4°K, the NF at 1 kHz changes from 13 to 2.7 dB as R_s varies from 10^4 to 10^6 ohms. The point to be made is that T_e gives an *immediate* indication of the expected noise performance, valid for *all* temperatures, whereas NF must be recomputed to obtain a valid indication of expected noise performance if the source-resistance temperature differs significantly from that at which the NF contours apply.

Output SNR for current-source-driven amplifier

Figure 12 shows an amplifier/filter driven by a current source, such as a photodiode or photomultiplier tube. R_L is the dc load resistor required to complete the dc bias circuit for the source, and I_{sh} is the source's intrinsic shot-noise current.

What can we say about the $(SNR)_o$ and its optimization? The amplifier's input signal voltage is directly proportional to the parallel combination of R_L and the amplifier's ac input impedance; however, for simplicity we assume that the input impedance is much larger than R_L. Therefore, by the use of Eq. (5), the voltage $(SNR)_o$ can be shown to be

$$(SNR)_o = \frac{I_s R_L}{[(I_{sh}^2 + i_n^2)R_L^2 + e_n^2 + 4kTR_L]^{1/2} f_n^{1/2}} \qquad (26)$$

Note from Eq. (26) that the voltage of the signal varies directly with R_L, which leads to the next statement.

Statement 4. $(\mathrm{SNR})_o$ is maximized asymptotically as R_L approaches infinity.

In many respects, the situation is analogous to the one in which the source resistance is very much lower than R_{opt}. When operation is from a source resistance that is either lower or higher than R_{opt}, the situation with regard to noise can, in principle, be improved by using a transformer to match the source resistance to R_{opt} of the amplifier. As shown earlier, this is indeed true for working from R_s values that are much smaller than R_{opt}. However, practical transformer design limitations usually prevent one from improving SNR by using a step-down transformer to match current-source resistances to R_{opt}. As discussed previously, when we are working from a low source resistance, the situation is only worsened by connecting a resistor in series with the source to make R_s equal to R_{opt}; similarly, it can only be worsened when working from a current source by connecting a resistor in parallel with the input to make R_s equal to R_{opt}. The load resistor R_L is such a parallel resistor. However, by making it as large as possible, the shunting effect is minimized and the best possible SNR is obtained.

Hence, for a given preamplifier, the experimenter should use the highest value of R_L possible (regardless of the noise-figure contours for that amplifier). In practice, the input cable's capacitive reactance and the amplifier's finite input impedance constrain the size of R_L.

Where the signal derives from a current source, noise-figure contours should only be consulted to determine the system noise or SNR, or to select the preamplifier to be used.

Maximum SNR when working from a current source will generally be achieved by:

1. Restricting the system bandwidth in accordance with maximum tolerable signal distortion.
2. Keeping R_L as large as possible.
3. Using an extremely high input impedance amplifier having low e_n and i_n. Amplifiers with FET (field-effect transistor) input stages are most appropriate.
4. Keeping the source shot noise to a minimum.
5. Keeping the input cables as short as possible.

Summary

Having borne with us through the discussions of the preceding pages, the reader should now be in a position to reap his well-deserved reward—namely, to take the raw noise data (e_n and i_n, R_e and R_t, NF or F, or T_e) furnished by the preamplifier manufacturer, and to convert the data into some useful numbers that compare amplifiers and allow estimation of the probable noise performance in a given experiment. The paragraphs dealing with transformer impedance matching show that when working with signals from a low source impedance, we can achieve a great improvement in signal-to-noise ratio by matching the source impedance to the optimum noise resistance of the amplifier by means of a step-up signal transformer.

One factor to bear in mind in using most of the equations provided in this article is that they are based on a consideration of source thermal noise and amplifier noise solely. As pointed out at the beginning of the article, there are many other sources of noise as well, and all too often they also degrade the SNR and cannot be neglected. Nevertheless, even though the computed SNR may exceed the actual SNR due to other sources of noise, the equations do define the best operating conditions.

Readers should realize that this article is primarily confined to "front-end" considerations. By using the appropriate method of additional signal processing, enormous additional improvements in signal-to-noise ratio can be obtained. These include the use of lock-in amplifiers, signal averagers, or, in some instances, general signal correlators.

REFERENCES

1. Coor, T., "Signal to noise optimization in precision measurement systems," PAR Tech. Note T-198A, Princeton Applied Research Corp., Princeton, N.J.
2. Faulkner, E. A., "The design of low-noise audio-frequency amplifiers," *Radio Electron. Eng.*, p. 17, July 1968.
3. "IRE Standards on Methods of Measuring Noise in Linear Twoports, 1959," *Proc. IRE*, vol. 48, pp. 60–68, Jan. 1959.

BIBLIOGRAPHY

Baxendall, P. J., "Noise in transistor circuits—pt. 2," *Wireless World*, p. 454, Dec. 1968.

Chaykowsky, O. C., and Moore, R. D., "Signal-to-noise considerations in experimental research," PAR Tech. Note T-196, Princeton Applied Research Corp.

Cherry, E. M., and Hopper, D. E., *Amplifying Devices and Low-Pass Amplifier Design*. New York: Wiley, 1968, Chap. 8.

Hooge, F. N., and Hoppenbrauwers, A. M. H., "$1/f$ noise in continuous thin gold films," *Physica*, vol. 45, p. 386, 1969.

Nittrouer, C. A., "Signal averagers," PAR Technical Note T-162A, Princeton Applied Research Corp.

Radeka, V., "$1/|f|$ noise in physical measurements," *IEEE Trans. Nuclear Science*, vol. NS-16, pp. 17–35, Oct. 1969.

Ridgway, S. L., "The use of a lock-in amplifier for the detection and measurement of light signals," Signal Notes, Princeton Applied Research Corp., Nov. 1967.

Smith, L., and Sheingold, D. H., "Noise and operational amplifier circuits," *Analog Dialogue*, p. 1, Mar. 1969.

Verster, T. C., "Anomalies in transistor low-frequency noise," *Proc. IEEE*, vol. 55, pp. 1204–1205, July 1967.

"How to use noise figure contours," Tech. Note T-226, Princeton Applied Research Corp., Princeton, N.J.

"Representation of noise in linear twoports," *Proc. IRE*, vol. 48, pp. 69–74, Jan. 1960.

Seymour Letzter received the bachelor of electrical engineering degree from the City College of New York and the M.S.E.E. degree from Northeastern University, Boston, Mass. Prior to joining the Princeton Applied Research Corporation, Princeton, N.J., he was employed by the ITT Corporation's Federal Laboratories, The Raytheon Company, and Fairchild Space and Defense Systems. In his present position at PAR, he aids customers in solving complex instrumentation problems, specializing in signal recovery and noise-reduction techniques.

Norman Webster serves as chief technical writer for the Princeton Applied Research Corporation. His formal education, which led to the bachelor of arts degree in German from Rutgers University, was augmented by U.S. Army technical school training in the field of electronics as well as experience as an instrumentation technician at the Princeton-Penn nuclear accelerator. He has been engaged in writing instruction manuals and technical writing for PAR since joining the corporation during 1965.

FIGURE 1. Interior mock-up of a jet air-carrier cockpit at Mohawk Airlines' training center.

At the crossroads in air-traffic control

III. View from the cockpit—reactions of some pilots, automatic flight management, flight-control systems, airborne CAS, 'moving map' displays, 'three-dimensional' area navigation, and others

In addition to ground control, the other vital elements in air-traffic control are the pilot, the on-board electronic instrumentation, and the prompt performance response of the aircraft in executing maneuvers on command

Gordon D. Friedlander **Senior Staff Writer**

Control of air traffic from the ground is only half the story—the airborne electronic equipment, displays, instrumentation, and human response are necessary to complete the loop in any practicable system. A few versions of collision-avoidance systems are entirely under airborne control. Based upon experience, pilots are wary of the hazards at certain major airports; in approaching other terminals, however, they have a sense of confidence that is inspired by the existence of the latest ATC equipment and facilities that afford a high degree of safety to pilots, passengers, and aircraft.

Flying an air carrier consists of hours of sheer boredom—interrupted by seconds of pure terror.

—A veteran airline pilot

Pilots, being human, are subject to the same fears and aversions as earth-bound mortals. Their outward calm, under the stress of inner tension, is part of the discipline of control that is a prerequisite for successful aviators and astronauts.

About two years ago, David Susskind interviewed three airline pilots on his television program, each of whom

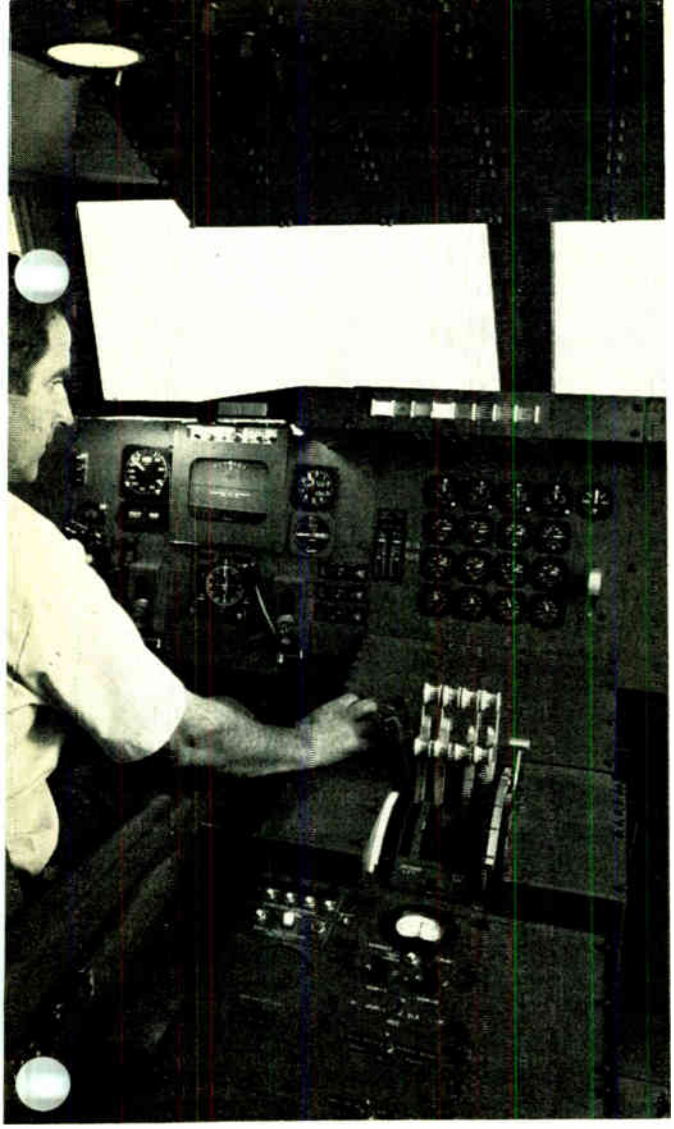


FIGURE 2. Cockpit of Boeing Aircraft's supersonic (SST) transport simulator. The company's flight-management instrumentation is installed as an integral part of the control panel equipment.

had been involved in a tragic air-carrier accident. The crashes, in one instance caused by an obstruction on a runway, obviously made these men acutely aware of the shortcomings in both airborne and ground-traffic control. They discussed, for example, the hazardous conditions for air-carrier operations that exist at some major U.S. airports. Because of nearby vertical obstructions (buildings, smokestacks, radio transmission towers, etc.), surrounding geographic features, or proximity to large cities, these pilots gave poor ratings to LaGuardia (New York), Washington National, Midway (Chicago), Denver, Logan (Boston), and Salt Lake City airports. (In fact, one pilot referred to them as "scareports.")

Some pilots are unhappy about operating in and out of Kennedy International—even under optimum traffic-control conditions. In the words of one veteran: "JFK is a glittering complex of World's Fair pavilion-type terminal buildings that look just great to the air traveler who is not aware that only four of the eight runways are ILS equipped." Malton Airport (Toronto) received high praise as being one of the best planned and equipped air terminals in North America: it is free of dangerous nearby vertical obstructions, it is situated 40 km from downtown Toronto, and all eight of its runways are ILS equipped.

Up front, in the pilot's seat

The writer, who can become readily confused by the instrument assortment on the dashboard of any 1970-model car, is thoroughly awed by the complex array of control knobs, levers, switches, flashing lights, and instrument panels installed within the cockpit of a typical jet air carrier (see Fig. 1). The pilots themselves must become adjusted to this degree of complexity by means of elaborate training programs in which actual flight conditions are computer-simulated with an amazing degree of fidelity. Thus Fig. 1 is actually an exact mock-up of the cockpit of a 227 Vista-Jet (twin-engine) airliner at Mohawk Airlines' Edwin A. Link Training Center in Utica, N.Y.

The present situation in air-traffic control has reached crisis proportions. As of today, only major airports in the U.S. are equipped with a degree of sophisticated electronic equipment that is marginally adequate to handle the ever-increasing volume of air operations in terminal areas. For example, there are only 282 ILS systems presently installed. There is very little ground/air or air/ground automation available on an off-the-shelf basis. The pilot must still execute his maneuvers largely on a manual basis during descent and landing approach. There are, however, a number of interesting concepts for air-traffic-control and in-flight automation that are either in the R&D or installed experimental stages. But the reader should bear in mind that *right now* most pilots depend upon voice communications, analog readouts and displays, and their own backlog of flight experience and expertise in handling their aircraft.

In the ensuing discussion on cockpit instrumentation, we shall be dealing with specific instrumentation and overall systems—either manufactured or proposed—that are capable of locking onto the existing transmission frequencies and channels used in VOR/LOC (omnirange and localizer), Doppler, glide-slope, hyperbolic (loran), and other standard navigational and approach aids. (For a discussion of some of these systems, see Part II, IEEE SPECTRUM, July 1970.) There are, of course, many competitive systems and instruments, but space constraints do not permit descriptions of all of them. Although certain systems will be described in some detail, such descriptions do not imply any commercial endorsement; rather, they are included to afford the reader the highest possible degree of understanding of the functions and use of the airborne equipments and concepts.

Flight management concepts

Automatic Flight Management (AFM) is a concept that is being developed in a ten-year-long R&D program by The Boeing Company. It emphasizes the automation and integration aspects of commercial air-carrier navigation, guidance, flight control, and aircraft equipment management. As such, the component parts of the program are not unlike other developments being undertaken independently within the industry. These components include "area navigation" and the use of moving map displays (both of which will be discussed later in this piece), inertial navigation, digital autopilots, air/ground and ground/air data links, blind-landing techniques, and

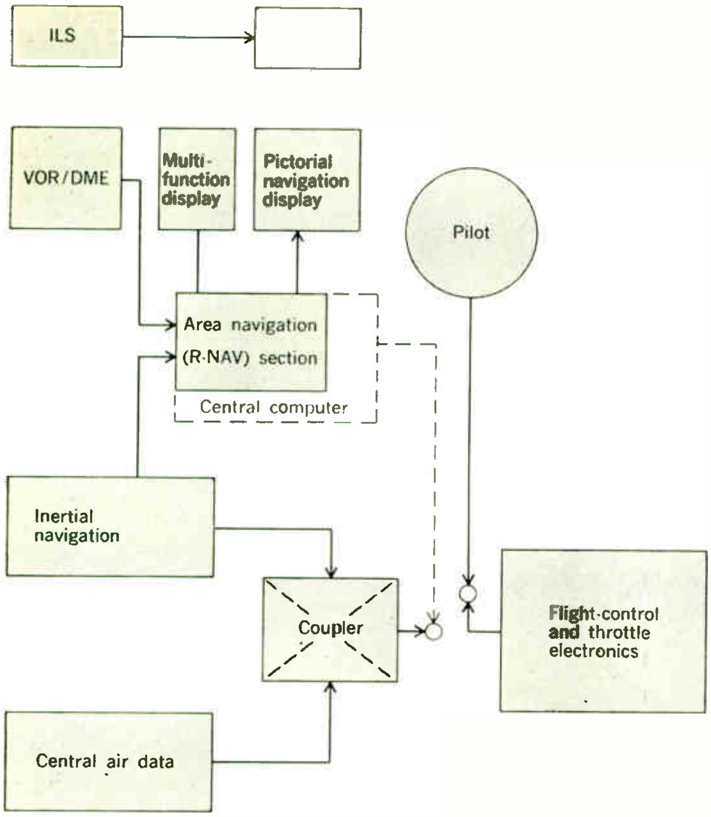
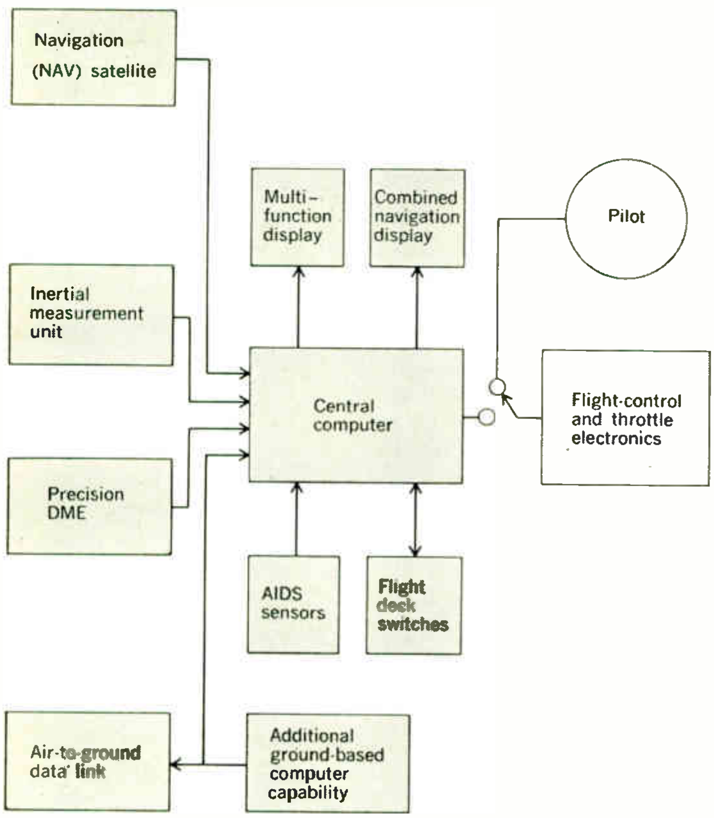


FIGURE 3. Block diagram of the first phase in a two-step transition period toward automatic flight management—the SST prototype stage.

FIGURE 4. Block diagram of the SST production phase of automatic flight management.



fully coordinated flight operations by the air-traffic-control complex. Boeing claims that the identifying feature of AFM is the integration of these and other individual ATC advances to produce a large leap forward in the total state of the art.

The stated objectives of AFM are to assist the growth of commercial aviation by reducing en route traffic delays, improving the utilization of available airspace and airports, and increasing the dependability and efficiency of operations. The AFM concept is also intended to increase the orderliness of air-traffic control in general, and the consistency of each aircraft's flight path in particular, in the interest of improved flight safety.

Statistics reveal that 40 percent of commercial air-carrier accidents from 1959 to 1968 were caused by navigational- and flight-path-control errors (most of which occurred in the descent and landing phases). Boeing believes that the semiautomatic principle of AFM, with an attendant reduction in flight crew workload, should minimize the accidents caused by human error.

The nucleus of the AFM system is an on-board central computer with sufficient logic and memory capabilities to provide precise control of the plane's flight-path performance according to a preprogrammed space–time profile. This flight profile will consist of nine elements: preflight checkout, taxiing from terminal, takeoff, programmed climb-out, en route cruising characteristics, programmed descent and terminal approach, automated landing, taxiing to terminal, and postflight checkout. The semiautomated system (with manual override provisions) will permit the pilot to depend less upon his routine operating skills while simultaneously increasing the effectiveness of his judgment and experience in coping with unusual situations.

One application of the concept will be for use by the Boeing 747s; another will be for the supersonic transports (SSTs) of the future.

Instrumentation and phases of AFM. Figure 2 shows the cockpit of Boeing's SST simulator. The instrument panels directly in front of the pilot contain the mode-control and flight-management displays. The aircraft's operations will be under the guidance of the mode-control panel, which is locked onto the AFM computer. The visual displays on this panel include the flight-path angle display, and four multifunction displays. The latter—which are actually analog readouts plotted along Cartesian coordinate axes—include a fuel gauge, effects of air braking, air-speed profile, and descent profile.

These multifunctional displays, together with audiotone alarm signals, will permit the pilot to monitor every airborne and on-the-ground operation in the nine-element flight profile just mentioned.

A pictorial (moving) map display will be part of the *airborne integrated data system* (AIDS). This general map concept will be discussed separately later in the article.

Essentially then, AFM is an effort to integrate as many as ten discrete computer and display subsystems (see Figs. 3 and 4) into one comprehensive, automated package. The development of AFM will involve a two-step transition period: Phase I, 1970–1975; Phase II, 1975–1980. During the first time slot, the SST prototype (scheduled for 1973) will encompass the functions indicated in Fig. 3. In the latter five-year period, a data link and additional ground-based computer capability, as shown in the Fig. 4 block diagram for the SST production

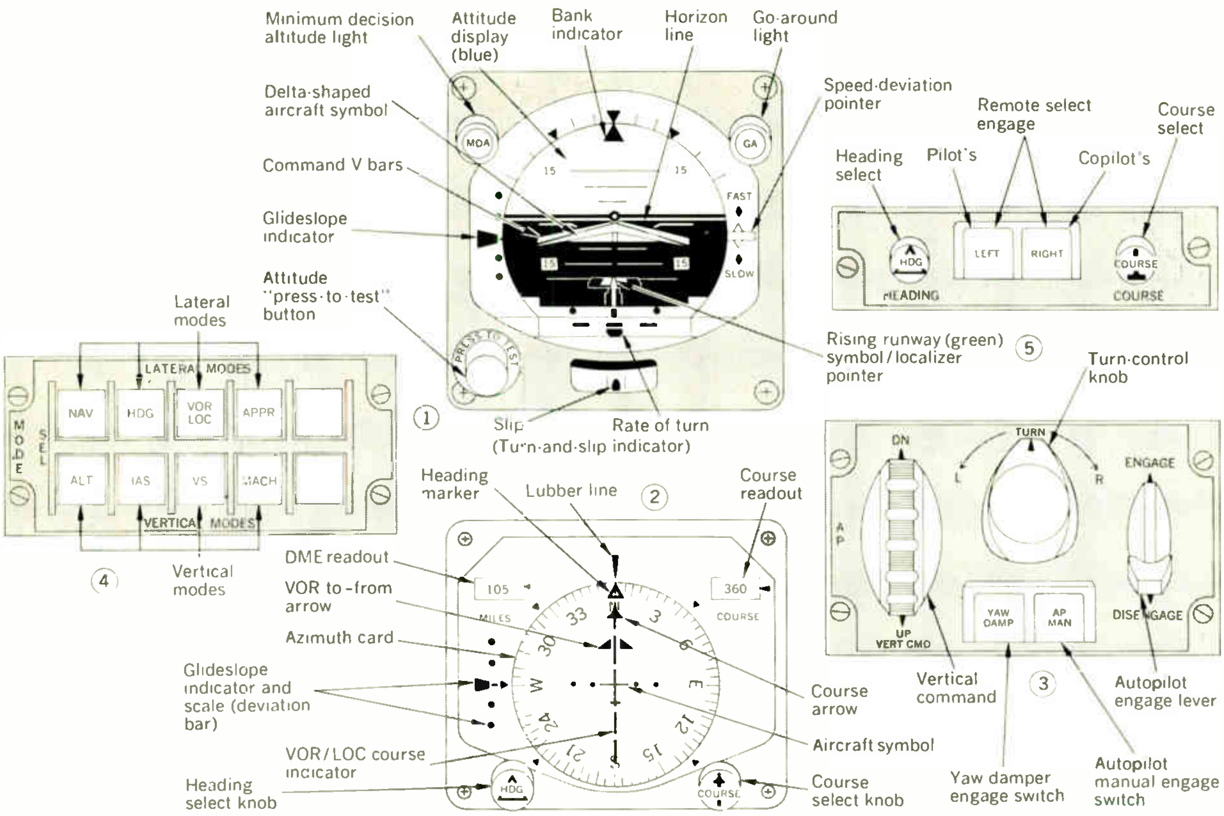


FIGURE 5. Sketch of the complete instrument array for the Collins Radio Company's Flight-Control System.

line, will complete the ten-year AFM evolution. It is expected that 50 percent of the air carriers will have AFM (or a similar system) by 1980.

Collins 'Flight-Control System'

One of the parallel developments to flight management is the Collins Radio Company's on-board Flight-Control System (FCS), which is described by the manufacturer as a combination of the conventional automatic-pilot and flight-director subsystems (the latter monitors aircraft performance). Unlike AFM, however, the Collins hardware is presently available. The system, with its on-board computer, can provide the pilot with three options:

- Complete autopilot control, with simultaneous flight-director commands that the pilot can monitor.
- Manual control of the autopilot while flying flight-director commands.
- Manually flying the computed flight-director commands, with autopilot disengaged.

System description, displays, and controls. Figure 5 shows the complete Flight-Control System's instrument panel array, which consists of the following five components:

Flight-director indicator (1). Gyroscope attitude, ILS-approach displays, radio altitude, speed-command data, and computer commands are combined in this 5-inch (12.7-cm) square instrument. The fixed delta-shaped symbol represents the aircraft. In command modes, the V-bars display the attitude that the delta symbol should assume. The command is fulfilled when the delta is flown into the center of the V-bars (as indicated in Fig. 5).

The "press-to-test" button permits the partial testing of the pitch-and-roll servo systems. The glide-slope (glidepath) pointer and scale indicator are shown at the left side of the instrument.

The runway symbol (at the bottom of the flight-director display) depicts the runway's centerline and shows the lateral displacement from the center of a localizer beam. The radio altimeter drives the runway symbol up vertically during the final 200 feet (61 meters) of descent as it simultaneously indicates localizer deviation.

One portion of the turn-and-slip indicator displays the rate of turn of the aircraft about the yaw axis. A two-needle-width deviation shows a two-minute standard rate turn of 3 degrees per second, and a one-needle-width shows a 1½-degree-per-second, four-minute turn. The slip indicator monitors the slip or skid of the aircraft, and it is used as an aid to coordinated maneuvers.

The speed-deviation pointer, if used, is driven by an external speed-command system and it displays the difference between the actual speed of the aircraft and the computed optimum speed. When not in use, the pointer is deflected out of view.

Annunciator (amber minimum-decision altitude, and green go-around) lights illuminate when the aircraft has descended to a preset radio altimeter decision height, at which time either the "land" or "go-around" decision must be made.

Course indicator (2). The course indicator includes the heading and course selections for the flight director and also the autopilot, and it displays the aircraft's position and heading with respect to the compass and azimuthal heading. A digital readout aids in fast, accurate course

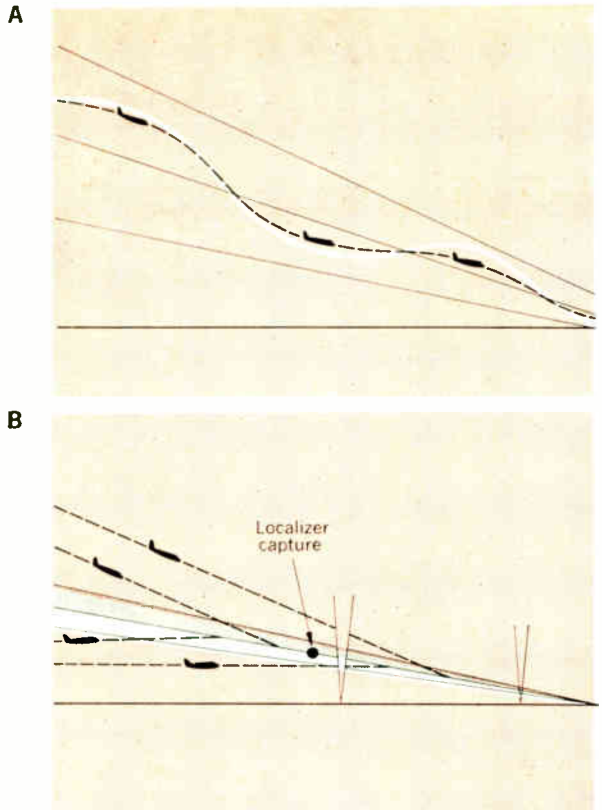


FIGURE 6. A—Profile of glide-slope gain programming in which low-altitude beam irregularities are optimized by gradually reducing the glideslope signal gain and by increasing the application of stored pitch information. The glideslope gain program is initiated by the radio altimeter at 1000 feet (305 meters) and the gain program is controlled to touchdown by absolute radio altitude. B—The automatic glide-slope capture is extended to all ILS approach situations by means of adaptive glide-slope capture, which can occur before or after localizer capture and from above or below the glide-slope centerline.

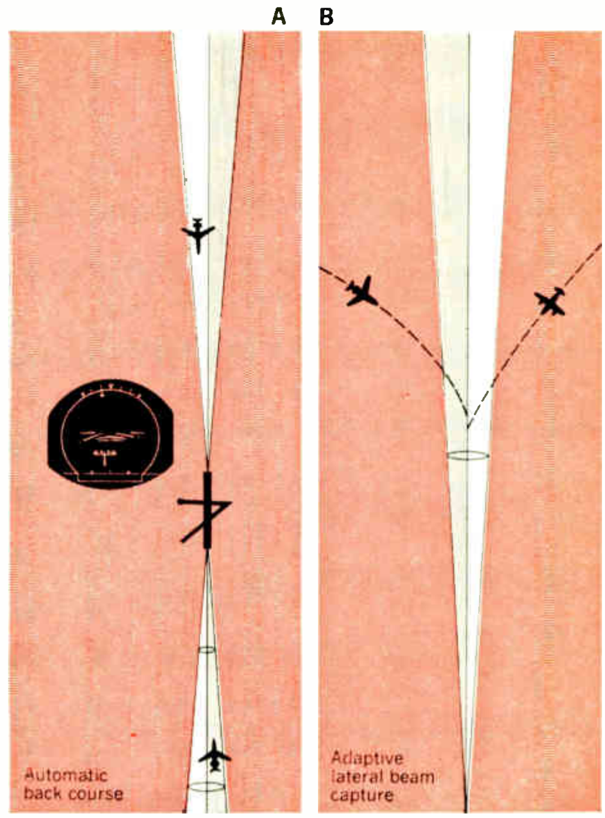


FIGURE 7. A–Sketch of automatic back-course capability of Collins' system, in which the autopilot and flight director back-course localizer intercepts and approaches are made with the same precision as front course approaches. B—Adaptive lateral beam capture increases the operational flexibility by providing automatic VOR and localizer course capture from either the pilot's or ATC's choice of heading. The capture begins at an optimum point computed from beam rate. The effects of speed, distance, intercept angle, and wind drift are automatically compensated for in a smooth turn to the course centerline.

selection. Slant-range (in nautical miles) to a selected DME station, lateral deviation from a selected VOR course, and vertical deviation from the center of the glidescope are also displayed.

As shown in the Fig. 5 diagram (instrument no. 2), the fixed aircraft symbol (when related to the movable parts of the course indicator) shows the aircraft's position and heading in relation to the azimuth card and the course-deviation bar. The heading-select marker is set to the desired heading on the azimuth card by rotating the knob. Once set, the heading marker (an orange delta symbol) will automatically rotate with the azimuth card to give a continuous reading of the selected heading. When the aircraft passes the VOR station, the "to–from" arrow reverses to give the correct indication. The arrow is *not* visible when a localizer frequency is selected.

The (yellow) course arrow is set to the desired VOR radial or localizer course on the azimuth card by rotating the course-select knob. The course counter in the upper-right-hand corner of the instrument improves the accuracy and speed of course selection by giving a digital readout, in degrees of azimuth, of the VOR radial or localizer course indicated by the course arrow. The DME readout (upper left corner) gives the slant range distance in nautical miles.

The to–from arrow always indicates the direction toward the VOR station along the course selected on the course-select knob.

The glide-slope indicator and scale (vertical dots) indicate the position of the aircraft above or below the glide slope.

Autopilot controller (3). This panel is the control center for the autopilot system; it provides the autopilot mode and command functions. The engage lever at the right is used to lock on or lock out the autopilot and it is also employed to engage the yaw damper when YAW DAMP is selected as an independent function. The yaw damper switch is a push-on/push-off solenoid-held switch, and it is used to select yaw-damping action independent of the autopilot operation. The autopilot manual switch is used to uncouple the autopilot from flight-director guidance.

The vertical-command control provides either vertical-speed or pitch command. The nose of the aircraft is lowered when the control is rotated toward DN, and raised when it is rotated toward UP. In the pitch mode of operation, the control commands the aircraft's pitch attitude. Rotation of the control from the UP index to the DN index will command a descent change in altitude of about 500 feet (153 meters) per minute. The turn-control knob is used to command banking when the autopilot is not coupled to the flight director. A center-position detent provides a compass heading hold.

Mode control (4). The operating modes of the pilot's flight director are selected on the mode-control panel. There are eight active and two spare push-on/push-off solenoid-held switches, which, from left to right, top to bottom row, are for

1. Navigation (NAV). Inertial, Doppler, loran, and compound navigation system outputs can be processed by the flight computer to furnish commands to the autopilot and flight director.

2. Heading (HDG). Selection of the heading mode commands a computed bank-angle turn to establish the heading set on the flight-director course indicator.

3. VOR/LOC. This mode provides the automatic intercept and tracking of VOR (omnirange) radials and localizer courses selected on the flight director's course indicator.

4. Approach (APPR). Glide-slope arm, capture, and gain programming (see Fig. 6) are combined with localizer capture in the approach mode. Submode switching from glide-slope arm to glide-slope capture is automatic.

5. Altitude hold (ALT). The aircraft will make a smooth transition to flying the pressure altitude it is at when this button is pressed to engage.

6. IAS hold. The "indicated air speed" at the moment of mode engagement will be maintained by the autopilot and commanded by the flight director's V-bar pitch movements.

7. VS hold. Precisely timed descents can be flown by means of the vertical-speed hold mode. The autopilot will hold and the flight director will command the vertical speed indicated at the moment of mode engagement.

8. Mach hold. Efficient climbs and precisely timed descent speeds can be obtained at the higher altitudes of supersonic flight with the Mach-hold mode. (The flight computer uses information from a central air-data computer to maintain the Mach number existing at Mach hold engagement.)

The two spare buttons are reserved for future add-on mode-control capabilities.

Remote course and heading selector (5). This selector permits both the pilot and copilot to set the heading marker and course arrow on their course indicator from a location that is convenient to reach. In normal operation, both the COURSE and HDG knobs on the course indicators are pushed into the remote position at all times. To set course and heading with the remote selector, the pilot depresses the LEFT button to the push-on position; the pilot's flight-director heading and course may then be set by rotation of the HDG and COURSE knobs on the remote selector. If the LEFT button is depressed again to the push-off position, the pilot's remote select will be turned off and the heading and course selections will remain fixed.

With the pilot's course indicator set for remote operation, depressing the RIGHT button will cause the copilot's heading marker and course arrow to rotate until they match those of the pilot. This feature permits either course indicator to act as a preselect for the other. If *both* buttons are depressed, the two course indicators will track together with heading and course adjustment.

Figure 7 shows two additional operational capabilities of the system: *automatic back course* and *adaptive lateral-beam capture*.

Flight applications. There are eight flight situations in which the pilot can make optimum use of the Flight-Control System's automatic operation. These include takeoff, climb-out and VOR intercept, en route cruising, descent, holding pattern, initiating terminal radar vectors, ILS approach mode, and go-around mode (in case an approach procedure is missed or aborted).

Report on airborne CAS

How CAS works. The airborne collision-avoidance system (recently tested by Martin Marietta's Baltimore Division for the Air Transport Association) uses advanced technologies drawn from four different areas: ATC radar beacon, precision clocking (crystal and atomic), digital data logic and control, and cockpit instrumentation. It is a cooperative air-to-air communication system based on precise time-frequency synchronization in each equipped aircraft. In this instance, synchronization is accurate to about 0.2 μs and one part in 10^8 in frequency for the communication signal in space.

The accurate timing serves two basic purposes. First, it controls the transmission from each aircraft so that such transmission occurs during a time period (message slot) reserved for that aircraft. Each other CAS-equipped aircraft will "listen" during that particular message slot and transmit during its own assigned message slots. This communications scheme, called *time-division multiplexing* (TDM), permits up to 500 aircraft to communicate reliably on a single frequency. Second, the degree of precision maintained among the CAS-equipped aircraft requires that each will transmit within a fraction of a microsecond of the start time for its message slot, and each receiving aircraft can make a direct-path, one-way range measurement on the transmitting aircraft.

The transmission frequency is also coherent in that it is traceable to the extremely accurate basic frequency of the system. Thus, each receiving aircraft can make a Doppler measurement on each transmission to determine the range rate of closing. The ratio of range divided by range rate equals the *time to collision* (or closest approach). This computed value is called "Tau," and it is a major criterion for collision threat evaluation. A fixed *minimum range* is another criterion and it provides threat evaluation in situations (such as one plane overtaking another at the same altitude) in which range rates are too low for reliable Doppler measurement.

All information exchanges occur within the frequency band assigned to the CAS function (1597.5–1622.5 MHz). The frequency is switched sequentially and automatically through that band in 5-MHz increments.

In the system, one aircraft will compare its own altitude with that of all other cooperating aircraft; it will command a CLIMB/DESCEND maneuver if either of the following threat criterions are satisfied: altitude and Tau, or altitude and range threat.

A FLY LEVEL command is employed to check an aircraft's climb or descent when the CAS system detects that the protected altitude band will be penetrated by another aircraft that poses either a Tau- or minimum-range threat. The functions and operational details of the

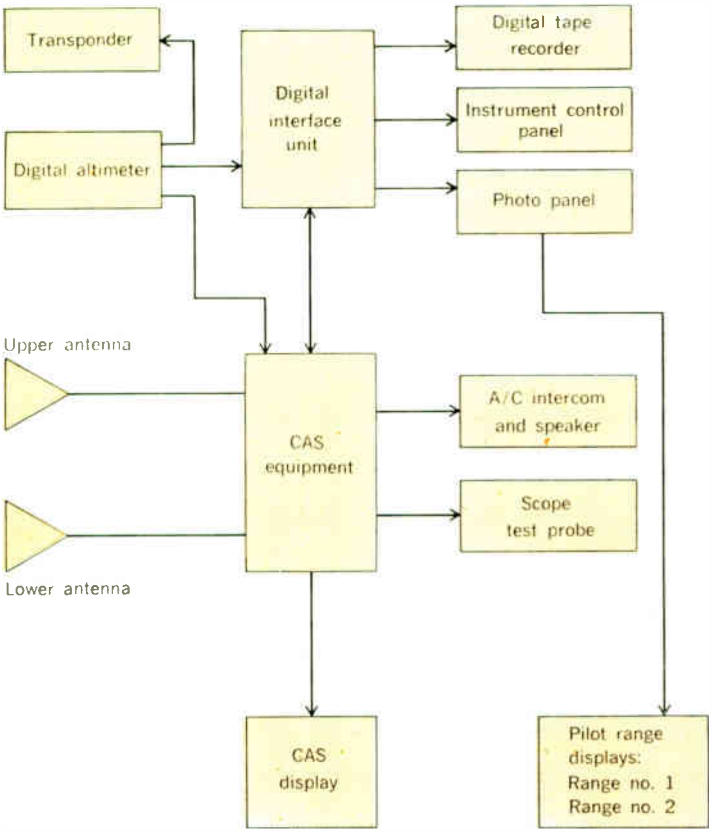


FIGURE 8. Block diagram of instrumentation used in the recent CAS test and evaluation program.

CAS will be further developed under the next eight subheadings.

Aircraft used in test program. Five aircraft—three piston-engine transports, and two small jets—fitted with CAS equipment and special instrumentation were used in the Martin Marietta-conducted testing program that was concluded last March. The piston aircraft were used for all engineering test work and for tests at low closure rates (below 100 knots, or 185 km/h) simulating typical terminal area conditions. The high-performance jets were used for tests of high closure rates (up to 1000 knots, or 1850 km/h) and for high-altitude intercepts. All five aircraft were used simultaneously for system tests during approaches and landings.

The instrumentation. Each of the five aircraft was equipped with a photo panel, which provided a valuable combination of flight data and CAS information that could later be reviewed from the photographic records. Digital tape recording of aircraft altitude and CAS data was used on all planes, except one jet. The altitude-reporting subsystem was the same as that used for the ATC radar beacon system (see Part I of this series). Precise altitude information is of critical importance to CAS since "climb" and "descend" are the primary collision-avoidance maneuvers. Two range displays (with readout in nautical miles) showed the distance separation from other participating aircraft. Figure 8 is a block diagram of the CAS instrumentation employed in the test program.

Cockpit displays. Each manufacturer participating in the test program provided his own advisory and command display, which used selectively illuminated words and symbols. Distinctive audiotone signals were broadcast over the cockpit speaker system for use with the visual preparatory and command signals.

Airspace 'protection envelopes.' The airspace protection envelope surrounding two approaching aircraft—one of which is considered as "your" plane—is bounded by altitude and Tau. A band of airspace extending 700 feet (214 meters) above and below your altitude is considered as "co-altitude." Thus aircraft within the co-altitude zone are considered to be threats. The horizontal extent of the airspace protection envelope is determined by Tau. As indicated in Fig. 9, the larger envelope repre-

FIGURE 9. A simplified diagram of the Tau-zone horizontal airspace protection envelopes.

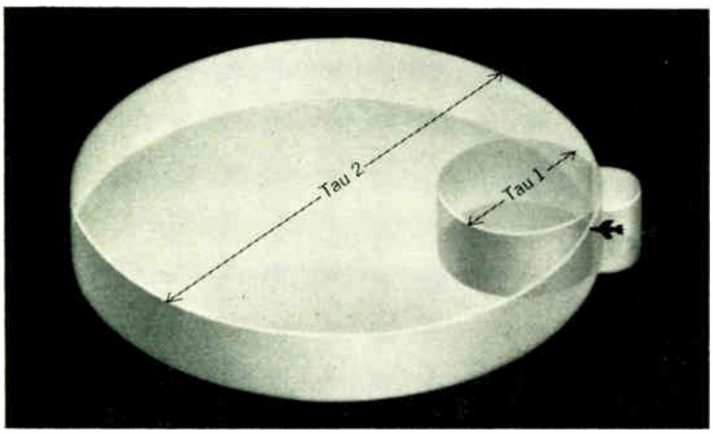
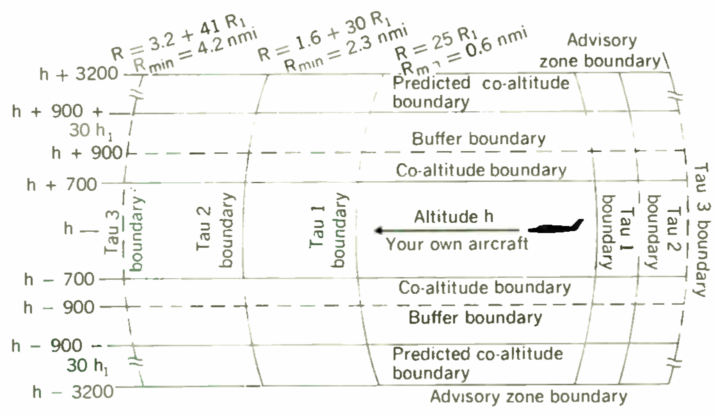


FIGURE 10. Diagram showing how Tau-zone boundaries are a function of the rate of closing of the aircraft involved.



A. CAS protection envelope

Legend: R = Range of boundary, nmi
R_1 = Closure rate, nmi/s
h = Own altitude, ft
h_1 = Own altitude rate, ft/s
R_{min} = Minimum range of boundary, nmi

B. CAS indicator logic

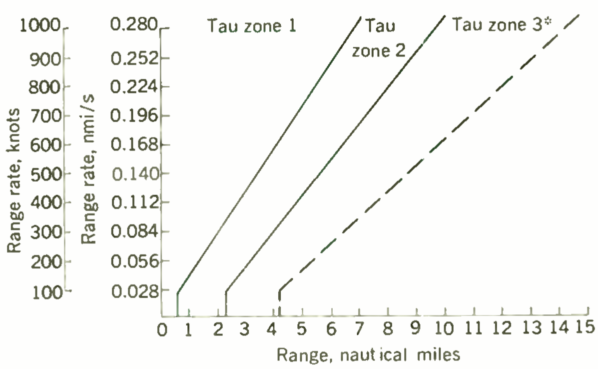
If intruder is in region:	CAS indicator display is:
Co-altitude Tau zone 1	Climb (dive)–Stop turn
Co-altitude Tau zone 2	Prepare to climb (dive)–Stop turn
Buffer Tau zones 1 and 2	Prepare to climb (dive)
Predicted co-altitude Tau zones 1, 2, and 3	Level off–Stop turn
Aircraft above (below) Tau zones 1, 2, and 3	Advisory aircraft above (below)

Notes: The "Tau 3" boundary, as a result of the flight tests, was found to be unnecessary and was dropped. All altitudes are given in feet, since this unit was employed in all tests.

sents "Tau 2"; the smaller, "Tau 1" for one airplane of a two-aircraft (total number involved) intercept.

Aircraft flying above and below the co-altitude protection envelope activate your advisory ("above" or "below") display when within a 700- to 3200-foot (214–976-meter) altitude band. Your own altimeter rate of change is used to expand the envelope when you are climbing or descending. As shown in Fig. 10, the Tau-zone boundaries are normally a function of the rate of closure of the two aircraft involved. When such closure rates are below 80 knots (148 km/h), the boundaries are defined by a minimum range.

Thus, when a co-altitude aircraft, posing a collision threat, penetrates your Tau 2 boundary, it activates your "prepare to..." advisory display (see Fig. 10); penetration of your Tau 1 boundary activates your "maneuver" commands. Threat evaluation is made automatically every three seconds, using the Tau and altitude criterions shown in Fig. 10. Logic values, however, can be modified without any significant change in the hardware displays.



How boundaries are defined

Zone	Minimum range, nmi	Range – axis intercept, nmi	1/slope, seconds
τ_1	0.6	0	25
τ_2	2.3	1.6	30
τ_3	4.2	3.2	41

*Tau 3 zone has been dropped.

FIGURE 11. Graph of Tau-zone boundary and minimum-range criterions as applied to CAS.

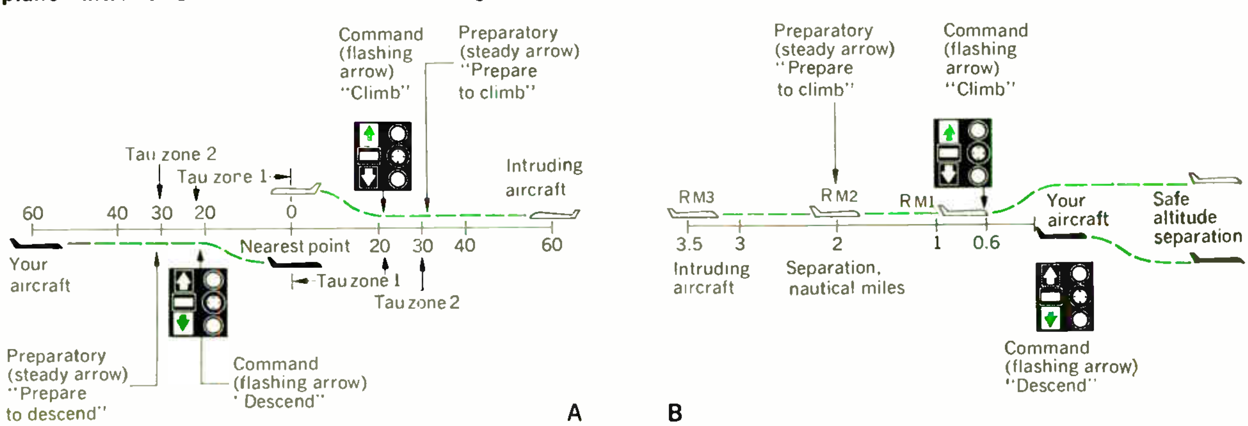
More on Tau-zone boundaries. The "basis of collision avoidance is to provide adequate time for avoidance maneuvers." This may be somewhat of a tautology, but it is the key principle of the entire concept, as represented by Tau. Tau 1, which triggers a maneuver command, was *set* at 25 seconds during the test program. When the rate of closure, or range rate (Fig. 11), between your aircraft and an intruder is 1000 knots (1850 km/h), the Tau 1 maneuver command will occur when both aircraft are 7 nautical miles (13 km) apart. But with a closing rate of only 100 knots (185 km/h), the same maneuver command would occur at a range of about one-half mile (1 km) apart. For closing rates below 80 knots (148 km/h), a minimum range, as shown in Fig. 11, is substituted for Tau. This is indicated by the vertical segment of the Tau-zone boundary line.

The minimum range for each Tau-zone boundary is shown in the Fig. 11 table; these values are set to allow time for the CAS to monitor and display information, for the pilot to react and initiate maneuver commands, and for the aircraft to leave the collision line. The boundary of Tau zone 2, for example (middle sloping line), is defined as a Tau of 30 seconds, or a minimum range of 2.3 nautical miles (4.3 km). Thus, at a closing rate of 1000 knots, the preparatory signal triggered by penetration of zone 2 would activate when the intruder's range was about 9.8 nautical miles (18.1 km).

Final countdown to intercept. Figure 12(A) illustrates how the CAS functions when Tau triggers the CAS commands (above 80 knots closing rate). Although the figure shows the "head-on" closure case, the same computations and display actions apply, regardless of the relative directions of the two aircraft.

Each aircraft in the sketch is within the other's co-altitude protection zone. Thus when an aircraft penetrates Tau zone 2 (Tau = 30 seconds), its display shows a steady red arrow denoting either "prepare to descend" or "prepare to climb." But upon penetration of Tau zone 1, the arrow illumination changes to flashing red, which is the command to "climb" or "descend." In Fig. 12(A)—which is based on data taken from an actual test intercept—the Tau zone 1 indication is displayed at approximately 22 seconds (three seconds earlier than the 25-second criterion set for this zone); however, CAS design

FIGURE 12. A—Diagram illustrating how CAS functions when Tau triggers the CAS commands if the closing rate is more than 80 knots (148 km/h). B—Diagram of a "following plane" intruder situation in which the closing rate is less than 80 knots.



criterions are adequate to provide safe separation with such a time lag, or even when only one of the two aircraft responds to the maneuver command. [By referring to the sketch diagrams shown in Fig. 16 of the second installment of this article (July 1970, p. 79), the reader may be better able to comprehend some of the situations in which the involved aircraft receive these maneuver commands.]

Range alarms at low rates of closure. When closure rates are lower, as in the case of the aircraft shown in Fig. 12(B), the computation of Tau (range divided by range rate) is no longer used as the criterion that triggers the CAS display. Instead, a minimum range is specified that will ensure that a preparatory signal or maneuver command is issued in time to maintain safe aircraft separation. Figure 12(B) covers the case in which the intruding aircraft is within the co-altitude of your aircraft. At about 2 nautical miles (3.7 km) from your plane, the intruder receives a "prepare to climb" signal as indicated in the sketch. At a minimum range (RM1) of 0.6 nautical mile (1.1 km) from your aircraft, the intruder receives a "climb" command. Similarly, your plane would receive a "prepare to descend" advisory when the range is about 2.3 nautical miles (4.3 km), and a command to "descend" when the range has been reduced to 0.6 nautical mile. (For simplicity of illustration, these points have not been shown on the sketch.) The diagram shows the typical case in which two aircraft are headed in the same direction, at slightly different speeds, with one aircraft slowly overtaking the other; but it applies equally well for two aircraft that are displaced laterally and are flying on near-parallel, slowly converging courses.

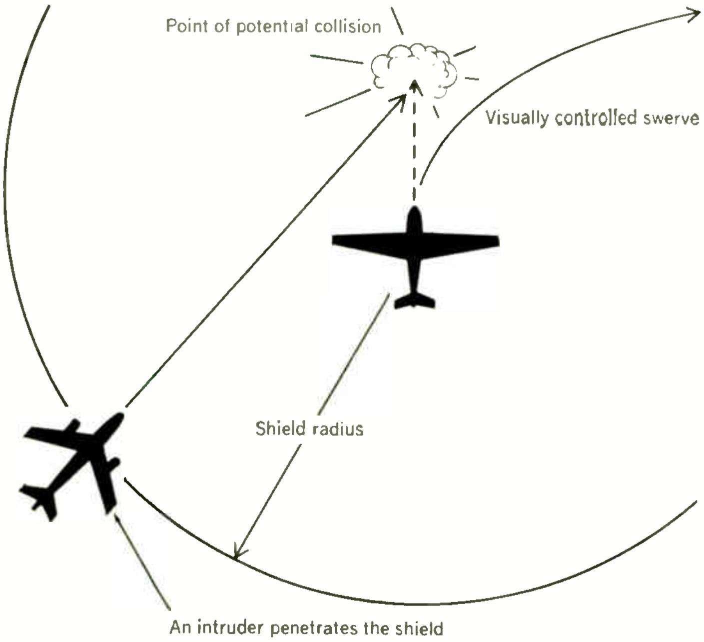


FIGURE 13. Simplified diagram of the "shielded flying" technique, which is a basic principle of RCA's SECANT CAS.

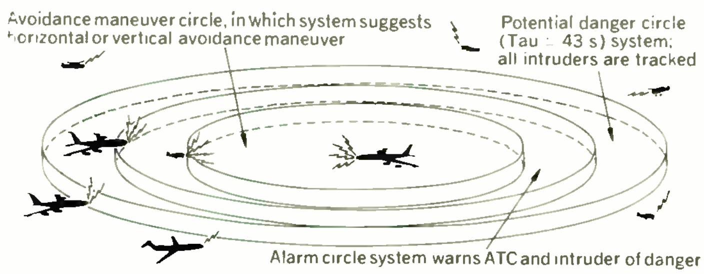
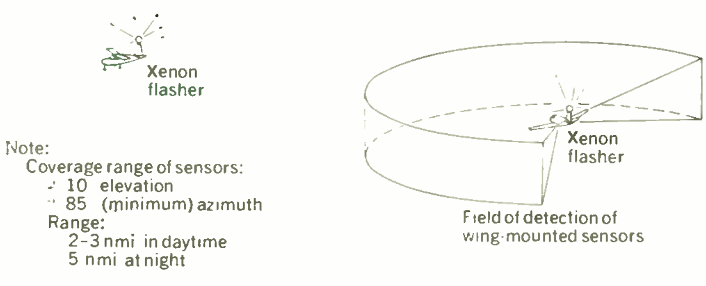


FIGURE 14. Three concentric circles: "potential danger," "alarm," and "avoidance maneuver" from the basic shield envelopes of the SECANT system.

FIGURE 15. Diagram showing the operational fields of view of a low-cost xenon flasher PWI system adapted for use in SECANT CAS.



Operational tests and a conclusion. A total of 187 two-aircraft intercepts were flown under strictly controlled visual-flight conditions, with complete digital- and photo-panel recording of the test operations. Altitude and track profiles were carefully preplanned and precisely flown. Of the 187 intercepts, 51 involved jet aircraft at high altitude and/or high rates of closure. Intercepts and closures were flown and recorded in a wide variety of operational situations to evaluate the system's performance for all azimuthal directions and vertical closures. Jets were flown head-on, at a common altitude, at closing velocities of about 1000 knots. At the low end of the closing speed scale, the Martin 404s (piston aircraft) were flown within 50 feet (15 meters) of each other in a common direction, at and near a common altitude. The jet aircraft were also flown in and out of the patterns of the slow-moving 404 groups.

Finally, the report states as one of its conclusions that "the flight test and evaluation program has demonstrated that CAS equipment built in compliance with the ATA's Air Navigation and Traffic Control (ANTC) Report 117, will prevent collisions in all types of operational situations." (This, of course, applies only to aircraft in which the CAS equipment is installed.)

'Secant': a possible low-cost CAS?

The RCA Corporation is now in the process of developing a CAS, designated SECANT (an acronym for Separation Control of Aircraft by Nonsynchronous Techniques). It is RCA's contention that, in order to be genuinely effective, CAS equipment must be made available to all general aviation aircraft at a reasonable cost, and must not be reserved exclusively for air-carrier and military use.

The feasibility of the system's critical elements has been demonstrated in the laboratory—and the total concept, according to a company spokesman, has been proved by means of computer simulation. The scheme will offer three equipment configurations (listed under the following subheadings), all compatible with each other, and are tailored to varying operational requirements and cost considerations. The simplest version, for private light planes, is expected to be priced between $500 and $1000.

The three airborne versions of SECANT, in order of ascending sophistication, are the Proximity-Warning Indicator (PWI), CAS, and a Traffic-Monitoring System (TMS).

PWI. The PWI is intended for the smaller general aviation aircraft and light military planes. The PWI-equipped aircraft electronically "sees" all other PWI-, CAS-, or TMS-equipped planes, and gives its own pilot an indication of the presence of an intruder within a predetermined range (fixed by the pilot) of his own aircraft. This is called *shielded flying*, in which the pilot decides what "shield radius" (see Figs. 13 and 14) is suitable for his current conditions of flight. He takes into account the speed of his own craft, the visibility conditions, traffic densities, and the types of aircraft that may be in his vicinity, and then sets the radius by turning a panel knob.

When any aircraft penetrates his shield, the pilot is alerted visually (Fig. 15) by his PWI. The alarm, which may consist of a rectilinear "cruciform" array of position lights, will advise the pilot that he "has company" either above, below, at the same level, and to the right or left within his preset shield. It is then up to the pilot to locate the intruder by eyeball observation and to determine whether he is in immediate or potential danger. This decision will govern the pilot's subsequent course of action.

Each SECANT-equipped aircraft would contain a "remitter" (a combination *re*ceiver–trans*mitter*), which would continuously transmit interrogating signals called "probes," and, in turn, would reply to probes from other aircraft. Thus each aircraft also carries a prober/processor as part of its component package. The PWI would separate air traffic into two parts:

- *Pop-up*—An aircraft that is in the airspace within the shield radius of the protected plane (it may or may not be a collision threat).
- *Flak*—All aircraft beyond the shield radius.

CAS. The RCA version of CAS is intended for some military aircraft, executive-jet aviation, air taxis, and smaller air carriers (piston engine and turboprop). It includes a "threat evaluator," which provides a defensive-flying capability by informing the pilot that he is in a threatening encounter situation with a SECANT participant in his traffic pattern. Simultaneously, the system displays the optimum avoidance route. Further, the alarm is given in time to permit the pilot to check the validity of the indicated evasion. If the pilot decides that he should not act unilaterally for fear of upsetting the traffic pattern, the situation is automatically brought to the attention of the air-traffic ground controller over an instant-alert "hot line." This feature should permit the best available coordination of the disengagement. In any event, the CAS would keep a self-generated escape route in reserve should the ATC fail to resolve the impending mid-air encounter by a preset time. The CAS also separates air traffic into two parts:

- Threat—An aircraft that looks as if it will soon be "too close for comfort."
- Flak—All aircraft that are not threats.

TMS. This system is intended for the larger air carriers and military planes; it provides the feature of "preventive flying." This allows the pilot to *keep out of trouble of his own making* by showing him where his potentially dangerous traffic is located and how it is moving from moment to moment. Actually, he is furnished with a traffic-situation display that shows all the SECANT traffic with which he might interact in the next minute or so, thereby enabling him to validate an intended change of course before he makes it. The TMS, which also includes the features of CAS, separates traffic into three categories:

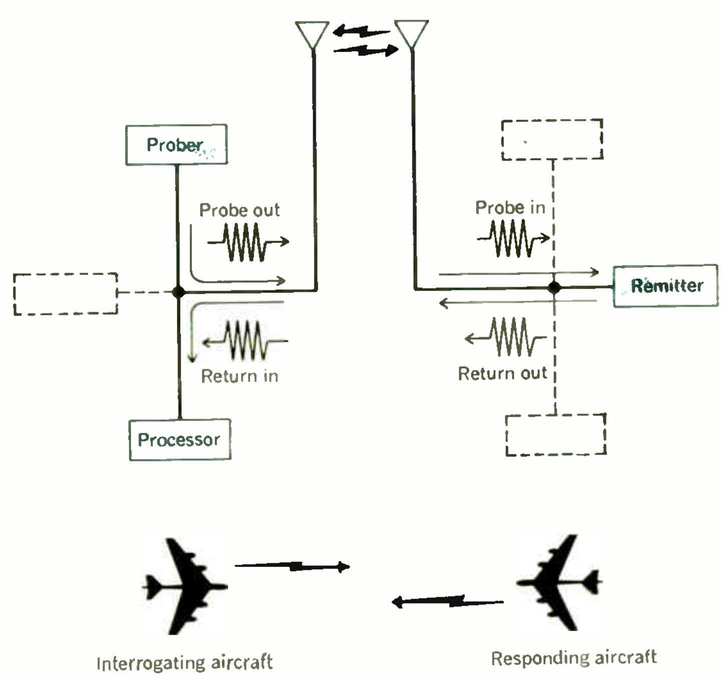
- Threat—An aircraft on a course that indicates it may soon pose a problem.
- Hazard—An aircraft that is not a threat at the moment, but could become threatening in a short time.
- Flak—All aircraft that are not threats or hazards within the traffic pattern.

Signal exchange in SECANT. Figure 16 shows that each aircraft participating in SECANT has interactions between the on-board remitter and the prober/processor. To interrogate its environment, an aircraft transmits probe pulses (short bursts of RF carrier). Any SECANT-equipped aircraft receiving such a probe acknowledges it by transmitting a return pulse. The returns have the same signal structure as the probes, but are shifted in frequency. For each probe pulse received, one return pulse is transmitted (one-to-one correspondence).

All aircraft interrogate and respond simultaneously—and in the same frequency band (1592.5–1622.5 MHz)—without synchronization among themselves. Thus the first job of the processor is to separate the significant returns (hits) from the unwanted returns (fruit) that other interrogators have triggered. When the interrogating aircraft receives a stream of returns, in response to a stream of its probes, it processes the return stream to

1. Reject the fruit.
2. Detect the targets in its vicinity.
3. Make range measurements on the target's trajectory.

FIGURE 16. Block diagram showing the signal interactions between each aircraft in the SECANT system.



4. Decide what to do on the basis of acquired infor-

in the aircraft, which enables the pilot to plot his own,

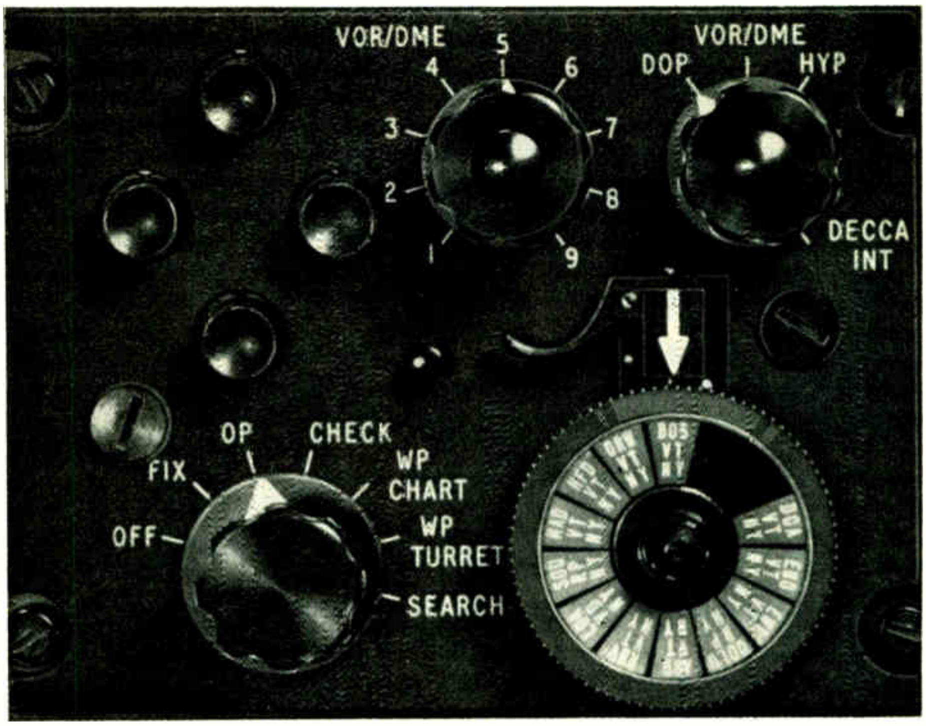


FIGURE 20. The Omnitrac receiver/computer control unit, showing the turret switch that contains 12 digital keys, each of which defines the geographical position of a preselected en route point.

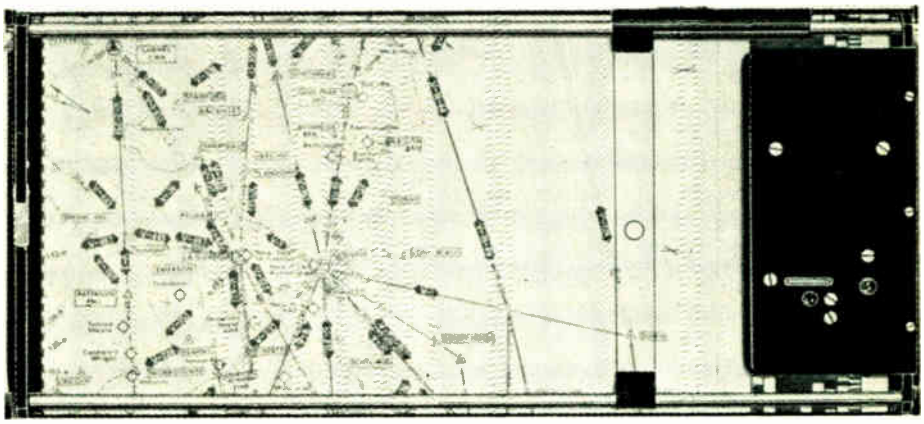


FIGURE 21. Trace of an actual approach on the Omnitrac display chart of an aircraft bound for Kennedy International Airport.

Lines' air-shuttle jet carriers in operation between New York and Boston, and New York and Washington.

The nucleus of Omnitrac is a compact, high-speed on-board digital computer that accepts radio signals from all existing FAA navigational aid transmissions—VOR/DME, hyperbolic (loran) systems, etc.—and converts these data into a rectilinear (x–y-axis) pictorial plot. Figure 19 is a block diagram of the system configuration. As indicated, the computer is locked onto the flight instruments and the autopilot so that the aircraft can be flown along a narrowly defined corridor, or "strip" between terminal airports. Like the Butler system just described, Omnitrac eliminates the requirement that a pilot must fly on a radial, or "spoke," to and from VOR/DME stations and, instead, fly a direct course, without doglegs, from one NAVAID center to the next.

The pictorial display *records* the flight track for an integration of the visual and readout display functions. A single chart roll contains both en route and terminal area maps used in the terminal-to-terminal journey of the aircraft. Route deviations, caused by weather conditions or altered flight plans, can be plotted directly on the chart. The alternative ground-based NAVAIDS are selected by running the appropriate chart section into the pictorial display; this also automatically feeds into the computer all other necessary fixed navigational parameters required by the chart.

The Omnitrac receiver/computer control unit is the principal component of the five-element system. As shown in Fig. 20, it contains a turret switch with 12 digital keys, each of which defines the geographical position of a preselected en route point. When the turret is turned to bring a particular key into use, the computer reads out bearing and distance from the aircraft's current position to the defined point. The pictorial display (flight log) also has an inverse mode in that, by manually setting the pen scriber to any point on the strip map, the computer is fed with—and stores—the selected position and reads out bearing and distance when required.

Figure 21 shows the actual trace of an air carrier's approach into Kennedy International (JFK) Airport by using the Colts Neck and Kennedy VORTAC* stations.

Similar systems—other suppliers

We have described, within the spatial constraints of this three-part series, those systems in which detailed information was made available either by the FAA or the manufacturer. In some instances, requested data were not received by the writer. Therefore, to bridge the gap, we shall at least mention some of the other devices that are either available or under development:

The 'Microvision' unit. The Bendix Corporation's Navigation and Control Division has developed a device that electronically displays, by means of a "heads-up" viewer, a lighted outline of an airport runway in an IFR landing. Depending upon the visibility at the time, the pilot can see either the actual runway or a simulated electronic runway.

Librascope L-193. A report indicates that the Librascope Group of General Precision Systems Inc. has a unit that superimposes all flight commands in coded color on an inner windshield in the cockpit. These commands include those required for landing, take-off speed, altitude, attitude, and heading control.

Electronic attitude director indicator (EADI). A number of commercial airlines have expressed interest in the EADI as a possible replacement for the present electromechanical devices. The Norden Division of United Aircraft, Sperry Rand Corporation, Kaiser Aerospace and Electronic Corporation, and General Electric have these equipments available.

Other firms active in the areas of ILS systems, NAVAIDS, or guidance-and-control systems, include Airborne Instruments Laboratory, FCA's Aviation Equipment Department, Waddell Dynamics (a subsidiary of Cubic Corporation), Bell Aerosystems Company, Laboratory for Electronics Inc., and Lear Siegler Inc.

Comments in conclusion

In this three-installment series, we have presented the views of the FAA, of some of the airline pilots, and of the air-traffic-control personnel. Although each group is obviously working toward the same objective—that of the greatest possible safety and efficiency in every aspect of air operations and traffic control—they are at variance with each other as to the methods and procedures that will ensure these common goals, and there is disagreement on the priority and speed at which certain programs should (or should not) be implemented.

It is only natural that the writer, after conducting many interviews with all parties in researching the background information for this article, would form some subjective opinions and "gut reaction." First of all, the level of criticism of the FAA expressed by the air-traffic-controllers through PATCO, the qualms voiced by some pilots, and the reassurances by the FAA that the overall situation is under control can hardly inspire confidence in the air traveler. Although the FAA points with justifiable pride at the sophisticated Common IFR Room in the New York terminal area, its ARTS III and Metroplex programs, etc., these are systems not yet in general use throughout the U.S. It will take a number of years—and a considerable sum in appropriations—before much obsolescent or obsolete equipment is replaced by the advanced electronic systems that are becoming available.

Our terminal areas and available airspace are becoming dangerously overcrowded because of

1. The sheer volume of the steadily increasing number of air operations, in all aircraft categories, at our major and secondary airports.

2. The lack of alternative high-speed ground transportation facilities over short and intermediate distances (300–1000 km).

3. An insufficient number of primary airports at or near many of our major cities.

Further, it has been alleged by controllers that many incidents in which mid-air collisions were averted by last-second evasive action are no longer reported in compiling "near-miss" statistics; and these occurrences are increasing, not decreasing. The writer is not impressed by the statistics that indicate that the number of fatalities per passenger-mile make air transportation the safest mode of travel. In any air-carrier mishap, one's chances of survivial are very slim; one does not usually walk away from the ground wreckage following a mid-air collision at *any* altitude.

Report of the ATC Advisory Committee. In December 1969, the *Report of Department of Transportation Air Traffic Control Advisory Committee* was released. The following are significant quotations from the document in respect to the serious problems faced by aviation:

"Air traffic is in crisis. The crisis now manifest at a few high density hubs is the direct result of the failure of airports and air traffic control capacity to keep up with the growth of the aviation industry .... Unless strong measures are taken, forces presently in motion will blight the growth of American aviation.

"The demand for all categories of aviation will maintain its high growth rate unless further constrained by an inadequate air traffic system ...

"In light of this projected demand, the Committee sees three critical problems which urgently require solutions if aviation growth is to be accommodated:

1. The shortage of terminal capacity;

2. The need for new means of assuring aircraft separation;

3. The limited capacity and increasing cost of ATC."

And, in the introduction to the report, we find this paragraph:

"While the Committee is recommending specific system characteristics and is proposing a number of high priority system engineering and development programs, it has not attempted detailed designs, nor has it considered specific deployment plans. Nevertheless, it is clear that the approach recommended will require an investment of several billion dollars during the 1970's for airport improvement and new construction if the demand is to be accommodated."

The writer joins the FAA, the air-traffic controllers, the pilots, and the flying public in hoping for the best in the future of aviation.

* A VORTAC is a combination of VOR (omnirange) and TACAN (military tactical navigation) stations.

The writer wishes to thank the FAA officials in Washington, and in the New York terminal area; the air-traffic controllers; the management and pilots of Mohawk Airlines and Eastern Air Lines; and a number of electronics and aerospace firms for their cooperation and assistance in furnishing the background data and information needed for the preparation of this series.

Circuit breakers

Physical and engineering problems

II—Design considerations

The solution to the problems encountered in the design of circuit breakers often consists of numerous compromises, dictated by the contradictory demands generally imposed on these devices by the system

W. Rieder Brown, Boveri & Co., Ltd.

This second installment of a three-part article focuses on circuit-breaker characteristics of general importance, independent of the arc medium chosen. A systematic, though necessarily incomplete, schedule of decisions to be made in the course of the development of one new breaker type will show the complexity of the problems involved.

Although even specialists can hardly distinguish different makes of electric motors and transformers, the great differences in shape and size of circuit breakers are generally fairly obvious. Whereas every motor is made from copper and iron, circuit breakers may use oil, air, SF_6, or vacuum as active mediums, and the pressure range covers 11 orders of magnitude, from 10^{-6} torr to some 100 atmospheres. Thus the designer of a breaker has many degrees of freedom. Moreover, as mentioned in the first installment of this article, contradictory requirements derived from the four fundamental duties (conduction, insulation, current making, and current breaking), and the special duties resulting from specific circumstances—together with requirements concerning reliability, maintenance, size, weight, and price—must, as far as possible, be optimized by a single design.

Therefore each decision, each figure fixed, each material chosen, each distance on a drawing is the result of a great number of compromises with regard to all the physical, functional, technological, and economic points mentioned or omitted in this article.

Fundamental decisions

Before the development of a new type of circuit breaker is started, the following factors must be considered: supply and BIL (basic impulse insulation level) voltage, continuous current, short-time current (maximum dynamic current), make and break current, special duties, and total break time (from tripping to the final interruption of the last pole to clear). Moreover, some general features (whether current-limiting interruption or minimum arc power release)—such as type of service (indoor or outdoor), need for tropicalization, inspection and maintenance requirements, mechanical and electrical life, combination with current transformers, geometric dimensions, weight and price limits, whether single or multiple gap, production figures expected, and development costs and time—must be fixed according to the situation on the market.

Thereafter, the basic concept of the breaker design is outlined, starting with the choice of the arc medium and treatment, the rated voltage of a single interrupting unit, whether equal units or specialized ones should be used for make and break, whether parallel resistors and/or capacitors are to be provided and for which purpose, contact movement and operating mechanism, means of gas generation, gas-pressure values, application of magnetic fields, main insulant (ceramic or organic), and so on.

Finally, there is an immense number of more detailed questions, such as choice of contact material, shape and number of contact points, opening and closing forces, single or parallel contacts, contact distance and movement characteristics, materials (metal, insulant), potential, and shape and diameter of the nozzle (if any).

Some general problems

Current limitation. In any circuit, a fault current can be limited by inserting a high arc resistance fast enough to prevent the increase of the short-circuit current to its prospective value, as is usual in high-power fuses, where the arc is introduced by the melting of a suitable conductor just above the rated continuous current and then produces an arc voltage drop comparable to the supply voltage because of the high power needed to melt the quartz sand in which the arc is embedded. In circuit breakers the power has to be withdrawn from the arc by other means—for example, magnetic-blast breakers, described later. In high-voltage systems, however, the protecting relays do not work fast enough, and the inertia of the moving parts of the usual breakers is too great to limit the current in time, as is possible in low-voltage miniature circuit breakers. Arc voltages of the order of 100 to 1000 kV are connected with large amounts of power and energy. Therefore, current-limiting switching up until now has been confined to low supply voltages, though chiefly for economic rather than physical reasons.

The number of interrupting units. It is not a matter of absolute quality, but of relative economy, whether it is preferable to provide a small number of high-capacity interrupters or a larger number of weaker, but cheaper, units. However, this economy includes not only the price of the actual units, but also the costs of their operating mechanisms, their mountings, their maintenance, and the space they need. Another economic point of view concerns the ability to combine breakers of different voltage and power ratings within a building-block system (Fig. 13). If the units are too powerful, the graduation of types within the system is too coarse, and for intermediate breaking capacities the next powerful breaker becomes too large and expensive. Finally, the testing problem may be decisive, since costs and required power for testing may be substantial.

Split-up of unit duties. It may be pointed out that in high-voltage installations the fundamental duties of current interruption and insulation are generally split between circuit breakers and disconnects (isolators). The disconnects are just able to interrupt a no-load current of a transformer, but they must perform important safety functions: the gap between open contacts is most immediately visible and always has a higher dielectric strength than the insulation to ground, so that an incoming overvoltage will tend to flash over to ground rather than to the disconnected parts of the system. Moreover, there is no creepage path in parallel to the gap but across a grounded metal part (Fig. 14).

The separation of functions may be extended to the units of the circuit breaker itself. It is not a matter of absolute quality, but of relative economy, whether it is preferable to develop and to manufacture one universal interrupter or to delegate the functions of breaking to one and of disconnecting and making to another.

In the latter case, when the current has been interrupted by the breaking units, the disconnecting units in series open and interrupt the current through the resistors in parallel to the breaking units (Fig. 15). Then the breaking units may close again and are ready to interrupt a fault current, made when the disconnecting units close. Obviously this is advantageous, especially for rapid auto-reclosing.

The universal unit has to cover much more divergent demands, but only one type of interrupter must be developed, manufactured, and stocked. The "specialists" are simpler, faster to develop, and more reliable. Therefore the same manufacturer prefers both a universal chamber for a medium-duty breaker for which a compromise is easier and cheaper (Fig. 16), and a division of functions for a heavy-duty type (Fig. 15). The making unit of the latter may be used independently as a high-voltage load switch (Fig. 17) to interrupt the continuous current and to make, but not to break, fault currents.

Break time and contact movement. To avoid destruction of the supply system by the Joulean heat of short-circuit currents, the time lag between tripping and final interruption has to be limited to two or three cycles of the supply frequency. Since the arc does not extinguish reliably unless a certain contact clearance is achieved at current zero, the arcing time is necessarily up to 3/4 of a cycle. Only about 5/4 of a cycle (20 ms) remains for the delay of the tripping mechanism and the acceleration of the moving parts until contacts part.

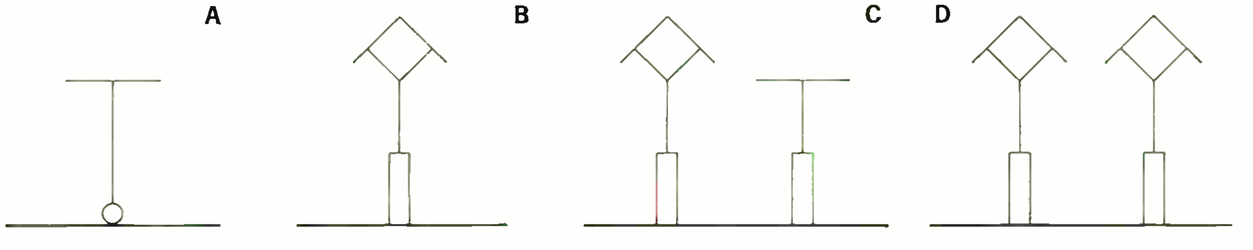
Contact movement has to start earlier: to avoid too slow opening (and too long arcing) on the one hand, and too high accelerating forces (and costs of the operating mechanism) on the other, the contacts part after a certain acceleration time with an opening speed greater than zero. Either the "fixed" contact moves a short way together with the moving one, or the latter slides a short distance along the fixed one.

To avoid overvoltages resulting from the chopping of small currents, the contact clearance must not increase too fast at the beginning, to allow for reignition after chopping at moderate voltage values. By this means, the energy stored in the inductance of the circuit (transformer) is dissipated in successive sparks, as shown in Fig. 18. Then the contacts open faster to reach the minimum interrupting distance within half a cycle of the service frequency. In some breakers they stop again for about the same time, to make sure that one current zero occurs at the optimum gap length independent of the time interval between contact parting and the next current zero, and independent of current asymmetry.

If the optimum interrupting distance is not sufficient for safe continuous disconnection, after arc extinction the contacts move into their final insulating position or special isolating contacts open in series with the interrupters (see Figs. 15 and 19). Unfortunately, the optimum interrupting distance is not universal, but depends instead on the particular duties.

In the interrupters of gas-blast breakers the valves must also be opened within one cycle or less to obtain fully developed gas flow in time. This gives rise to serious mechanical problems due to the transmission of mechanical impulses from the release (which is at ground potential) across the insulators to the valves and contact operat-

FIGURE 13. Series of circuit breakers of different rated voltages and power, composed of equal units (building-block system). A—170 kV, 7 GVA. B—362 kV, 15 GVA. C—550 kV, 25 GVA. D—765 kV, 40 GVA (also shown in Fig. 16).



ing mechanisms in the hot interrupting units. Transmission must be not only fast but also synchronous within small tolerances (1 ms) for each unit of the pole. The problems become the more severe the greater the currents (that is, the heavier the contacts) and the higher the voltages (that is, the longer the insulating distances to ground).

Pneumatic devices are too slow with respect to dead

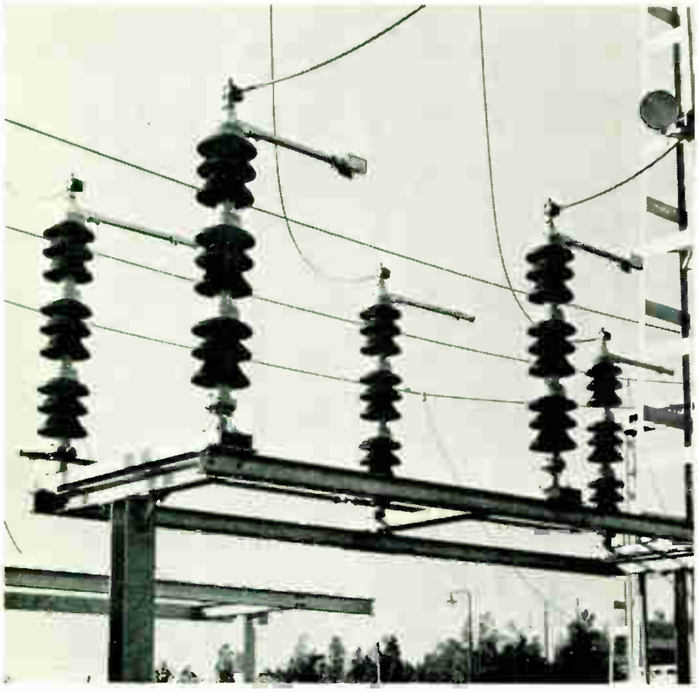
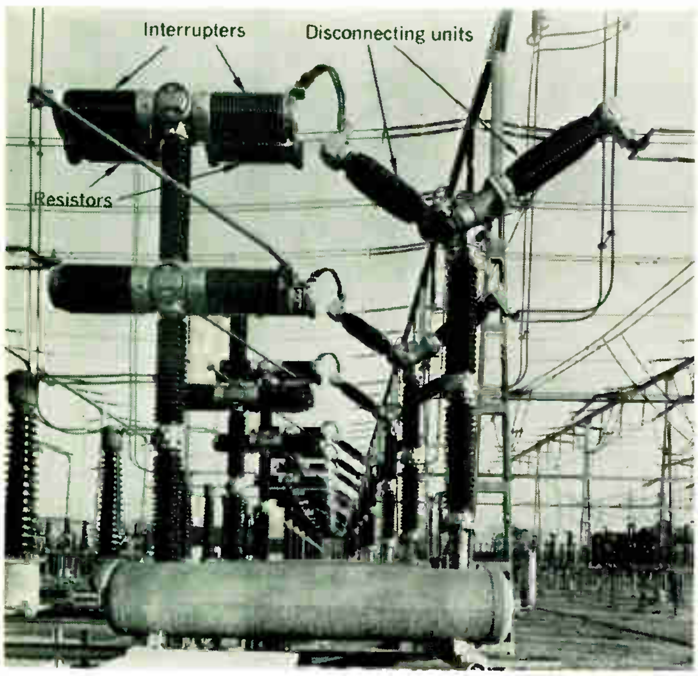


FIGURE 14. Disconnector, rated 145 kV, 800 amperes.

FIGURE 15. Pole of a 250-kV, 20-GVA air-blast circuit breaker consisting of four interrupting and two disconnecting units, with pneumatic mechanical operation.



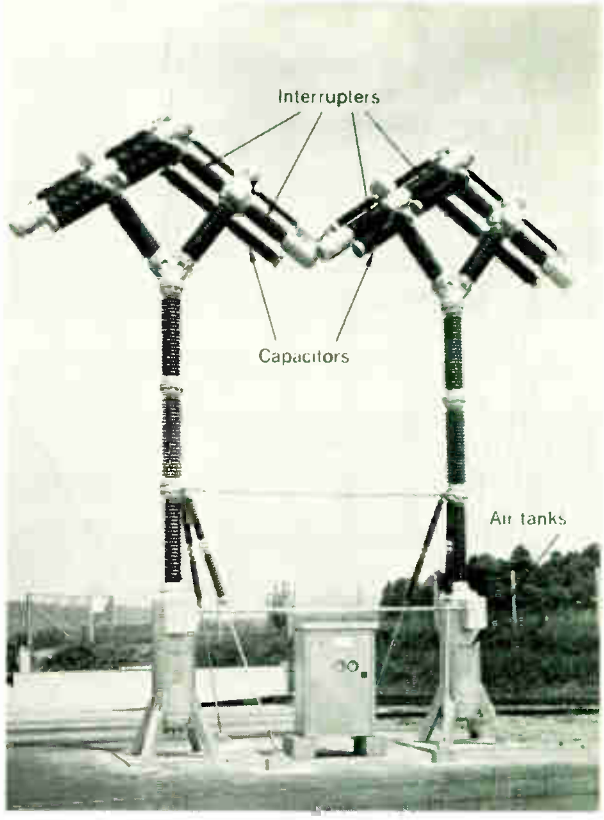
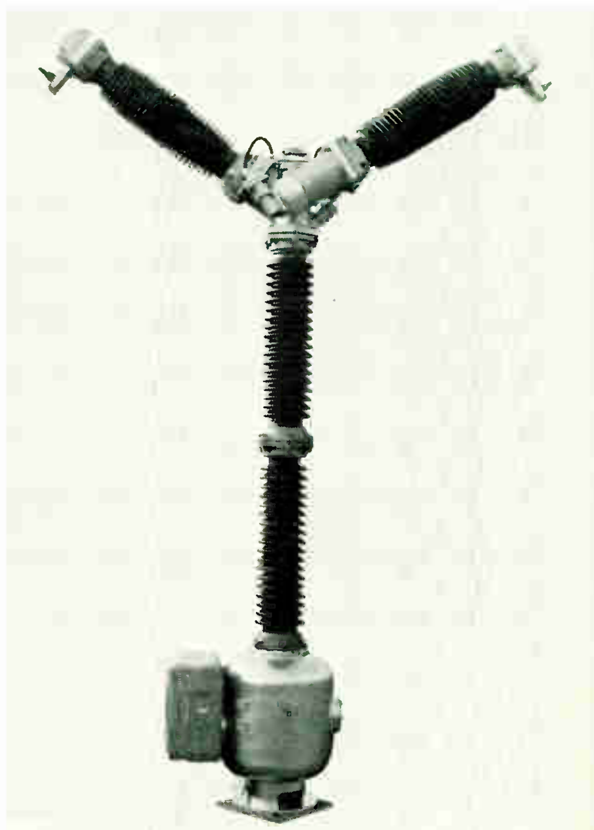


FIGURE 16. Pole of a 765-kV, 40-GVA air-blast circuit breaker consisting of eight universal units with parallel capacitors for voltage grading, with pneumatic mechanical operation.

FIGURE 17. High-voltage load switch (245 kV, 2 kA) using the disconnect units of the breaker shown in Fig. 14.



volumes and the limiting speeds of sound. Hydraulic power transmission is inhibited by the wide range of temperature at which open-air breakers have to work ($-40°$ to $+80°C$) using the same oil. Even solid links, such as rods, waste costly time for elastic deformations, if not prestressed.

Additional difficulties arise in oil breakers, in which the contact travels a comparatively long distance and some arcing time is needed to produce the pressure required for arc extinction.

The actuating power is supplied either by compressed air or by motor-charged springs (Fig. 20), or (in the case of vacuum interrupters which have light contacts and comparatively short travels) electromagnetically.

Resistors and capacitors. As described in last month's installment, it is possible to aid a breaker during the interaction period by a resistor or capacitor in parallel with the arc.

Whether it is preferable to use a greater number of units or to increase the breaking capacity of each of them with the aid of parallel resistors or capacitors depends chiefly on relative economy and special duties rather than absolute quality. In this case the need for stronger insulating columns bearing the units and the auxiliary interrupters necessary to switch off the current across the resistors must be taken into account and compared with other parameters that could cause equivalent advantages, such as increased pressure or nozzle diameter (gas flow).

The value of the resistances needed depends on their purpose.[40] To control the recovery voltage distribution across the units of a multibreak device, one may use capacitors of the order of tens of picofarads or high resistors (100 kilohms) if they are needed. Lower resistances could cause sequential breakdowns; if one of the units fails and is supplied by a high current across the resistors in parallel to the other units, it would not be able to recover, but another unit would also reignite because of the higher voltage stress. The resistances may be either nonlinear or constant.

To aid the breaker in cases of high frequency and high amplitude of recovery-voltage transients, medium resistances (1 kilohm) are useful in certain cases.

The capacity of many breakers is limited by short-line faults. Air-blast breakers can be reinforced efficiently by low resistances of the order of the surge impedance of the line (a few hundred ohms). Of course these carry a considerable current, which has to be interrupted by auxiliary interrupters or disconnecting units as in Figs. 15 and 19. These resistances aid the arc interruption both before and after current zero. For SF_6 breakers the arc voltage is lower and the critical fault distance is shorter than for air-blast breakers; aid seems to be needed mainly just before the arc is cooled down to the critical 2000°K, when the breaker is weak and thermal reignition can easily be forced, whereas afterwards its dielectric strength is high. Therefore resistors are less effective than capacitors at the fast voltage change immediately around zero.

At extra-high voltages the circuit breakers are sometimes misused to reduce the overvoltages that appear naturally (and not due to a weakness of the breaker) when a no-load line is connected to a system. In this case a making resistor, which has to be short-circuited after some milliseconds, can help; however, its optimum value (500 to 1000 ohms) depends on both line length and potential (if the line is precharged after an interruption in the case of autoreclosing).

Contact materials. The choice of contact material is another problem involving many contradictory demands.[41] The contact material has to provide low contact resistance even at moderate contact force; it must not produce disturbing surface layers—even in aggressive atmospheres and after a long time at elevated temperatures. Contact welding has to be prevented in both closed and closing contacts, even at the highest fault currents expected. Contact wear due to mechanical and electrical stresses (at making and breaking) should be minimized,

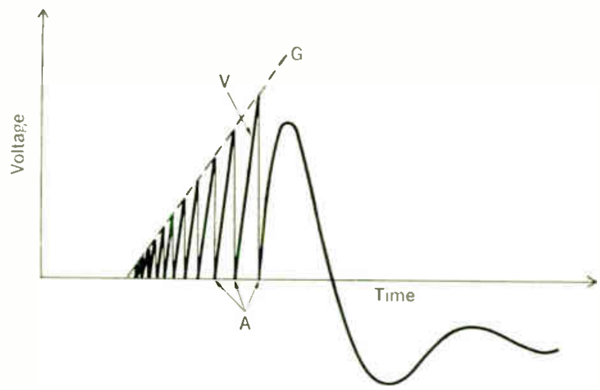
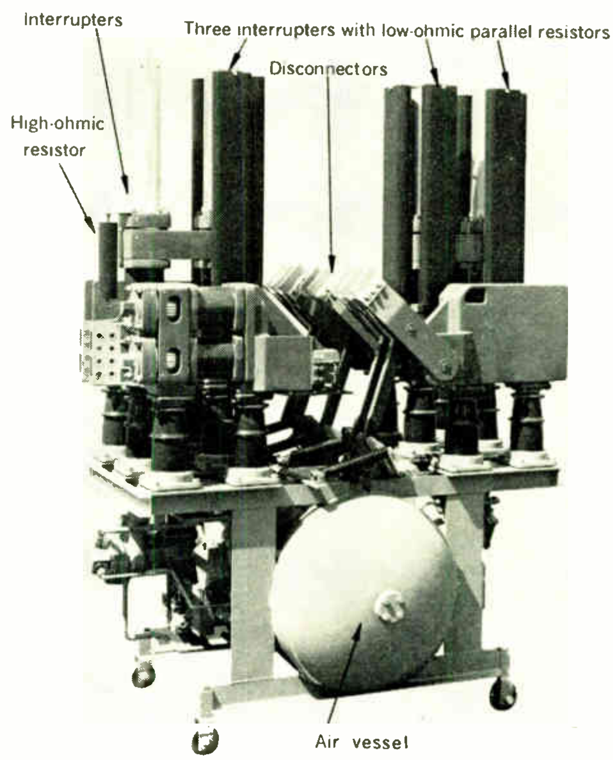


FIGURE 18. Voltage oscillograms during the interruption of a small inductive current by an air-blast breaker. Overvoltages are prevented because of the sufficiently slow contact speed. V: recovery voltage across contact gap. A: spark voltage. G: increasing dielectric strength during contact separation.

FIGURE 19. Indoor air-blast breaker with air disconnector, rated 20 kV, 12kA, 3 GVA. Pneumatic mechanism.



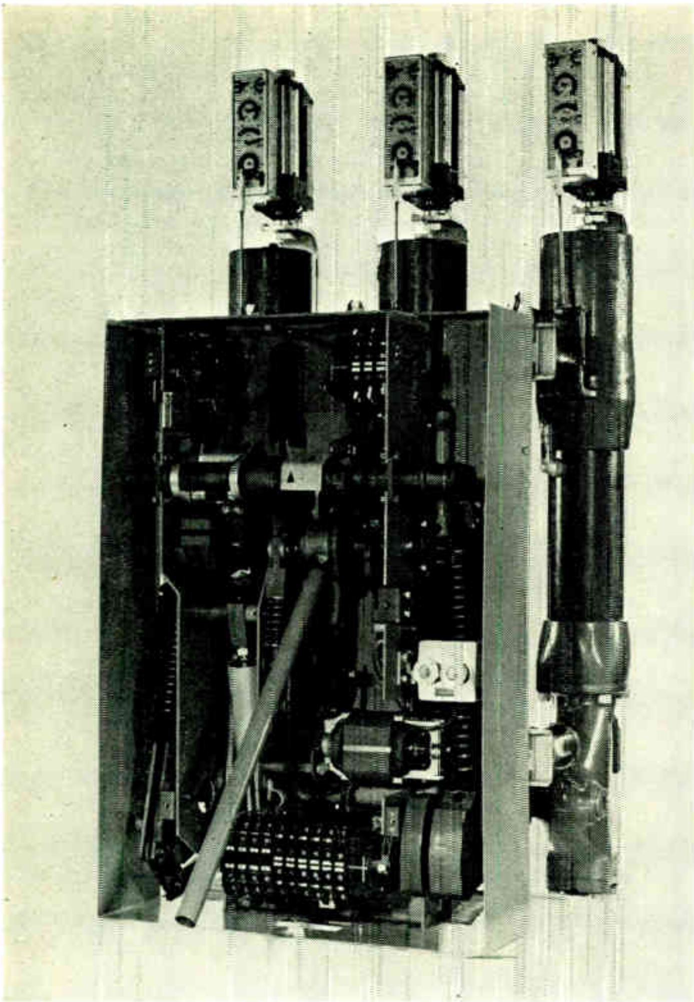


FIGURE 20. Motor-charged spring actuator for low-oil-content indoor circuit breaker.

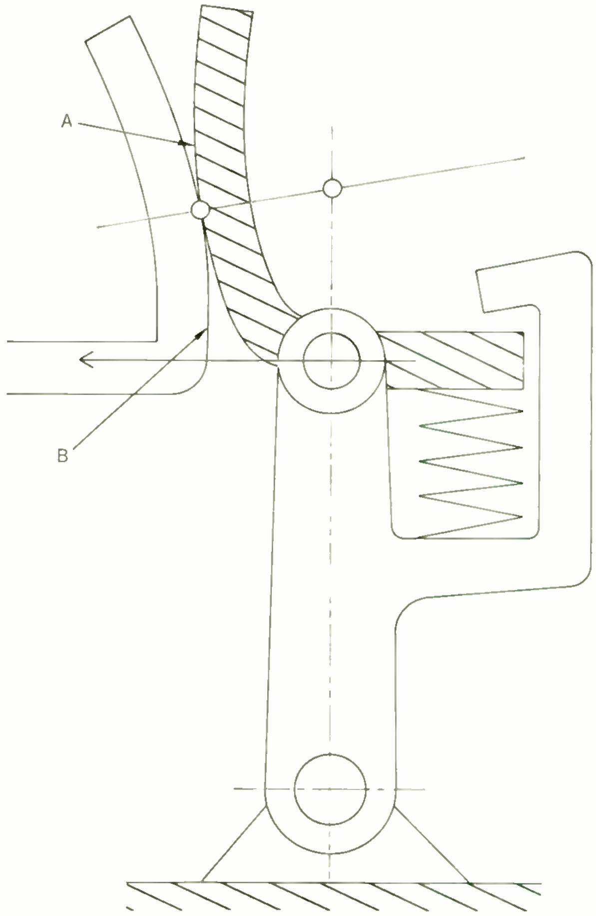


FIGURE 21. Contacts parting in a point (A), which is different from another point (B) that conducts the continuous current.[42]

but the material should be easily machinable and cheap.

A separate compromise has to be made for each type of breaker, depending on the opening and closing forces available; the rated make, continuous, short-time, and break currents; arcing time; arcing medium and gaseous emissions; contact temperature; switching frequency; contact movement (with or without a self-cleaning slide); and conditions of manufacturing and price. Since the claims lead to contradictory physical features (such as hard and soft or high and low melting point), it is not possible to give a list of the features of an ideal contact material. Further difficulties arise because certain different features, such as electrical conductivity, hardness, and melting and boiling point, are caused by the microscopic structure of the crystal lattice and therefore cannot be chosen independently. Some features always appear in combinations (for example, noble and expensive) which cannot be explained easily.

Alloys do not allow us to combine the features of their components in a simple way, because they build a new crystal lattice with new features and always with a higher resistivity and hardness but a lower melting point than the average of their components.

Sintered compounds are more suitable because they have the properties of their components. They may be nonmetallic, such as CdO, WC, and C. Unfortunately, if oxygen is present the most promising combination, silver–tungsten, builds a chemical compound, Ag_2WO_4, that is a very reliable insulator. On the other hand, some compound materials, such as Cu-W, Ag-CdO, and Ag-Ni, are very successfully used and show much better features than might be expected from the components.

Nevertheless, compromises are sometimes not acceptable; often the duties must be distributed between two pairs of contacts connected electrically in parallel. One of them has to carry the continuous current, whereas the other has to make and break it.

Often the designer does not really separate the two pairs of contacts but provides just one material at that part of the contact members that touch first and part last, whereas another material is used at the point of steady contact (Fig. 21). Of course, this combination requires a rather complicated contact movement.

The still more complex problems of vacuum contacts will be discussed in Part III of this article.

REFERENCES

40. Ruoss, E., CIGRE Rept., Group 13, pp. 488–490, 1964.

41. Rieder, W., "Die Beurteilung der Kontaktwekstoffe für elektrische Schaltgeräte," *Bull. SEV*, vol. 53, pp. 830–840, 1962.

42. Pawelka, E., *Einführung in die Schaltgerätetechnik*. Vienna: Springer, 1965.