VOLUME 12 NUMBER 8 IEEE spectrum AUGUST 1975

IEEE spectrum

articles

departments

TEKTRONIX ANNOUNCES

A new concept in portable instrumentation

The TEKTRONIX TM 515 Traveler Mainframe looks like fashionable flight luggage, compact and easy to carry, or slide under an aircraft seat. In reality, it's a five-compartment power module/mainframe that provides power and interface connections for TM 500 plug-in modular instrumentation. Plug in the new (two-wide) SC 502 15-MHz dual-channel oscilloscope, and you have the beginnings of a powerful take-along instrumentation system.

You can optimize a TM 500 system to your needs by selecting from more than 30 plug-in modular instruments. With the TM 515 Traveler Mainframe and SC 502 Oscilloscope as a nucleus, select from DMM's, counters, generators, power supplies, signal processors, and even blank plug-ins for your "home-built" circuits. Intended applications include areas from digital field service to medical, from audio/communications to on-site industrial controls maintenance.

The SC 502 is Tektronix quality, featuring clean triggering characteristics, delay line input, trigger view, trigger holdoff, 1 mV sensitivity, and the capability of working through the rear interface circuit board with other TM 500 instruments. It features a specially brilliant crt designed built by Tektronix for use in areas of high ambient light. Include a DD 501 Digital Delay alongside the SC 502 and gain the capability of delay-by-events—you can then obtain stable digital displays from electromechanical sources like disc drives that would otherwise be too jittery for accurate viewing on any conventional oscilloscope. Include the DC 505A Universal Counter and DM 502 Digital Multimeter to complete your TM 515 package, and discover the benefits of simultaneous counter and DMM capability with trigger level readout at the touch of a push button.

The TM 500 concept lets you take along on field servicing trips the same instruments you use in the lab or for production testing, thereby enabling you to maintain the same standards on the "outside". The SC 502 Oscilloscope, for example, may be used as a bench instrument in any multiple-compartment TM 500 mainframe, and it offers unique systems capabilities, as well, when operated in a rack in the RTM 506.

Contact your local Tektronix Field Engineer or circle the appropriate reader service number for a demonstration of TM 500 instrumentation or additional technical information on the TM 515 Traveler Mainframe and SC 502 Oscilloscope. For an up-to-date TM 500 Catalog write to Tektronix, Inc., P. O. Box 500, Beaverton, Oregon 97077. In Europe write Tektronix Limited, P. O. Box 36, St. Peter Port, Guernsey, Channel Islands.

TEKTRONIX®

committed to technical excellence

On Reader Service Card, Circle No. 3 for Information, No. 4 for Demonstration

the cover

A rippling rainbow dramatizes the high information densities attained by viedo disk technology. White light is diffracted by data tracks less than two micrometers apart on this sample video disk from MCA Laboratories (article, page 34).

IEEE Publications Operations

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—644-7555. Published at 20th and Northampton Sts., Easton, Pa. 18042. **Change of address** must be received by the first of a month to be effective for the following month's issue. Please use the change of address form below. **Annual subscription:** for IEEE members, first subscription, $3.00 included in dues. Single copies, $4.00. 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—644-7760.

SUBSCRIPTIONS, orders, address changes: IEEE Service Center, 445 Hoes Lane, Piscataway, N.J. 08854. Telephone: 201—981-0060. Other Publications: IEEE also publishes more than 30 specialized publications (see SPECTRUM, Jan. 1975, pp. 95–96).

IEEE DEPARTMENTS: IEEE Headquarters: 345 East 47 Street, New York, N.Y. 10017. IEEE Service Center (for orders, subscriptions, address changes): 445 Hoes Lane, Piscataway, N.J. 08854. Telephones: Headquarters, 212—644 + extension: Information 7900, General Manager 7910, Controller 7748, Educational Services 7860, Publishing Services 7560, Region/Section Services 7750, Technical Services 7890. Service Center, 201—981-0060. Washington Office/Professional Services, 202—785-0017.

MOVING? PLEASE NOTIFY US 4 WEEKS IN ADVANCE

Name (Please Print)

Address

City State/Country Zip

MAIL TO: IEEE SERVICE CENTER
CODING DEPARTMENT
445 HOES LANE, PISCATAWAY, N. J. 08854

ATTACH LABEL HERE

- This notice of address change will apply to all publications to which you subscribe.
- List new address above.
- If you have a question about your subscription place label here and clip this form to your letter.

spectral lines

Education, registration, certification, validation

In the July issue we opened the question of the role of government in regulating the practice of engineering. Some states in the U.S., we noted, are going so far as to restrict the use of the title "engineer" to those who are registered in the state. Others are pressing in the direction of doing away with the "industrial exemption," which could require that EEs working for private industry would have to be registered. We concluded on the note that more and more informed observers believe that registration for all engineers, without exception, may be inevitable.

A parallel topic, which space did not permit our addressing, is that of certification. Unlike registration, which is state administered and related largely to establishing minimum basic requirements for the practice of engineering, certification is usually administered by a professional society and implies proficiency and technical competence at an advanced level, often in a relatively narrowly defined specific discipline. (For example, the Society of Manufacturing Engineers provides certification in 12 specialties, ranging from tool engineering to electrical and electronic manufacturing engineering.) Certification can be accomplished in a number of ways, most of them involving some combination of specified education, experience, and formal testing.

In some professions, certification is a requisite for continued licensing (registration) of a practitioner. This hardly seems a likely probability in the case of engineering, at least not until registration becomes more widespread. Nevertheless, while you yourself may be neither registered nor certified, if you've a son (or daughter) studying engineering, the likelihood of his (her) becoming registered and/or certified is surely higher. Yet before registration/certification becomes the norm in our profession, one can anticipate years of debate concerning the relationship between the two, as well as to educational requirements. On the one hand, there exists a fundamental school of thinkers who believe that education should be a prerequisite to registration, and that registration, in turn, should be a prerequisite to certification. On the other hand, many nondegreed engineers endorse direct certification because it gives them an optional route to recognition as a professional.

Beyond the difficulties in precisely defining registration and certification and their relationship to one another, there are a number of further barriers to their extension. There is, of course, the strong implication of limited access to the profession, with its attendant impact on engineering schools (fewer students, a longer cycle per student). There is the expense entailed in administering the programs (generally much higher for certification than for registration). There is the resistance to change on the part of industry managers, who envision restrictions on their latitude in job assignments and salary administration. There is the concern by the professional societies themselves that faulty design work by an engineer certified by a society might open the door to legal action against the society itself. There is the complication, in the case of registration, of incompatibility of registration standards from state to state, and in the case of certification, in incompatibility of standards between specialties as well as between disciplines.

A major factor in both registration and certification concerns currency. Many states do not require periodic reexamination or requalification to prove continuing competence of licensed engineers. But this may change. Certification programs are much more likely to contain a time limit, so that renewal is necessary in, say, three years. Recertification procedures may involve participation in continuing education programs, or reexamination in the applicant's field.

Psychological factors loom as deterrents to registration/certification, too. In electrical engineering, this is, in part, because most senior EE managers are neither licensed nor certified. Some have risen via the "self-made man" route, and attach greater significance to performance than to degrees and credentials. The best certification of competence, they believe, is the ability to keep your job. Some managers feel that registration/certification could lead to over-restriction regarding entry to the profession—effectively screening out the maverick geniuses, the rebel innovators, and those who might enter through the "back door" (mathematicians, chemists, physicists), to the detriment of technical progress and society. (Would an inventor, ultimately, be restricted from gaining a patent, they ask, were he not licensed/certified in the proper field?) Among the skeptics are those who fear that mass registration and certification would tend to produce "standard" (interchangeable, but mediocre) engineers.

IEEE itself intends to embark upon an experimental program on January 1, which has some of the aspects of certain certification programs, but which itself avoids the use of the label "certification." The Institute's preference is to provide the individual engineer with a tool—namely, the formal validation of a structured educational program (comprising courses from ECPD-accredited schools, in-house courses, and IEEE courses), which may be useful to him on the job, in achieving licensing, etc. In its experimental validation program, IEEE will be measuring the individual engineer's reaction to the value of the program.

In the long run, the acceptance of registration/certification will hinge upon its usefulness to and impact upon all of the parties involved—engineer, manager, industry, and the public at large.

Donald Christiansen, Editor

Television on a silver platter

A proliferation of options may spawn some orphans in the contest among developers of video disks

High information density—billions of bits etched in a microfine spiral track and then inexpensively replicated in plastic—that's the common denominator of what promises to be the most exciting home entertainment product since color television: the video disk. Together with a shelf-top player that connects to any standard TV's antenna terminals, single video disks can provide half an hour of full color programming per side—complete with stereo or bilingual soundtrack. However, as test marketing gets underway, enthusiasm for the concept could dissipate into pragmatism because several incompatible, independently developed systems now claim preeminence.

Indeed, lack of commonality among competitive video disks is reminiscent of communications problems that plagued the infamous biblical tower. Disks are flexible, stiff, single-sided, double-sided, shiny, transparent, mechanically grooved, and optically flat. They sport different diameters, different playback speeds, and different base materials. Yet each, with its associated player, does essentially the same job of delivering a compatible U.S. NTSC 525-line (or European PAL 625-line) video format, modulated on a low level RF carrier, to any standard television receiver.

Who's on first?

Major public announcements relating to video disk availability have been made by Philips/MCA, RCA, and Telefunken/Decca. Each has demonstrated working hardware attractively styled for consumer use—not just engineering models. Also actively engaged in development efforts but maintaining a lower profile are Zenith, Thomson-CSF, i/o Metrics, and Sony. So far, only Telefunken has test-marketed a product. Its TED video disk player went on sale in West Germany for $650 this past March. Philips/MCA and RCA both mention late 1976 for introduction of their players at $500 and $400, respectively. In every case, the cost of single preprogrammed disks is projected at less than $10.

Don Mennie Associate Editor

All these schemes can be roughly lumped into two broad categories based on the method used to extract information from a rapidly rotating disk—optical and mechanical. Furthest along with optical playback is the Philips/MCA team that joined forces last summer (see *Spectrum*, Jan. 1975, pp. 67–69), and held public demonstrations of its equipment in New York, March 17–21. Clear color pictures and sound were obtained using single-sided reflective video disks manufactured independently by MCA in California, N. V. Philips in the Netherlands, and Zenith in Chicago. Playing time is 30 to 40 minutes. The Philips/MCA playback transducer is a photodiode powered by a 1-mW helium-neon laser.

Soon afterward, RCA held several invited sessions at Rockefeller Center where its entry, "SelectaVision," built around a mechanically tracked capacitive pickup, was also demonstrated. RCA's video disks are two-sided, and play 30 minutes per side. The only unconventional component in the system is a $10 sapphire stylus that snaps out for easy replacement.

Telefunken's product employs a diamond-tipped mechanical "skid" coupled with a piezoelectric trans-

This video player (lower left) developed by Philips/MCA yields 30 minutes of color programming plus a dual soundtrack (stereo or bilingual mono) from a single-sided "reflective" video disk spinning at 1800 r/min. Playback may be accomplished with any standard TV by connecting the video player to the set's antenna terminals. Production of players for U.S. distribution will be handled by Magnavox. Mastering is done in open air with a laser scribe. Microscopic pits are cut in metal films (or photoresist) deposited on a polished glass disk. The resulting spiral track starts at the *inside* edge of the recorded band and moves outward.

ducer. Tiny bumps on the Telefunken video disk are detected and amplified during playback. The disks themselves run 10 minutes each and are never handled by the operator. Instead, an envelope containing a disk is inserted in the player which automatically positions each disk for play, engages the skid (stylus), and finally returns the disk to the envelope when playback is completed.

Notably, when and if mass acceptance does blossom, it will probably embrace only one of these video players. Each boasts legitimate claims of technical excellence, so the outcome could easily hinge on marketing finesse and the wide distribution of inexpensive, interesting programs.

The optical option

Crucial to Philips/MCA's video player is a supply of reliable, low-cost helium-neon lasers. Lee Mickelson, physicist with MCA Laboratories, Torrance, Calif., says such lasers are presently being delivered at $34 per tube in lots of 5000, with prices to drop below $20 per tube for orders greater than 100 000 units. Guaranteed lifetime is 18 months. The laser provides a monochromatic, low-noise light beam that may be sharply focused for extracting information from a rapidly rotating video disk.

Such disks must rotate at speeds directly related to the power line frequency of a given television system (50 or 60 Hz). A 50-Hz (European) system produces 25 complete pictures (frames) per second on the TV screen (1500 per minute), while a 60-Hz (U.S.) system gives 30 frames per second (1800 per minute). To implement slow and stop motion features, Philips/MCA has selected 1500 r/min and 1800 r/min respectively as video disk rotation speeds. In either system, one complete revolution of the disk yields one completely scanned TV frame.

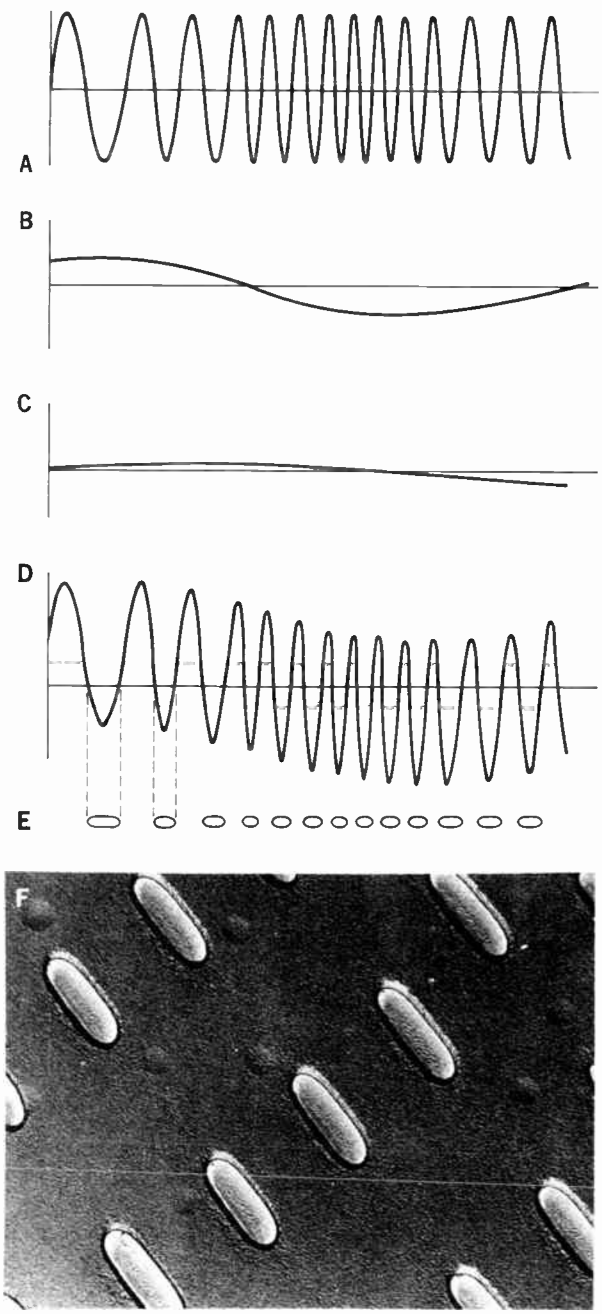
Potentially important and so far unique to the Philips/MCA video disk system is its instant random access capability. Viewers can hunt through a program at high speed while the fast-changing image stays in horizontal and vertical sync. Each frame is individually numbered (00 000 to 54 000 for a typical NTSC half-hour) and the digits can be called up during any viewing mode—stop motion, slow motion, real time, or "hunt"—with a simple push-button control. Any particular frame or sequence can be exactly identified for reference purposes.

Philips/MCA's noncontacting optical readout precludes wear or a reduction in performance of the disk or pickup. Laser light is first focused into a spot. The beam then diverges, passes through several optical elements, including a microscope objective, before impinging on the video disk surface. This surface contains a spiral track of oblong depressions, all having the same width and depth, coated with a highly reflective aluminum layer. The multiplexed video and audio information is described only by the length of the depressions and the spaces between them (Fig. 1).

When the incident light spot (now about 1 μm in diameter) hits a depression, reflected light will be lost because it is diffracted over angles larger than the microscope objective is capable of accepting. This modulation is detected by a photodiode and electronically processed into sound and pictures.

Tracking and timing errors caused by motor speed irregularities, mechanical vibration, and video disk centering tolerances are compensated for by small mirrors in the optical scanning system. They are mounted on transducers (galvanometer or piezoelectric) and can give the laser beam both radial and tangential "nudges," effectively nulling out wow and flutter. Eccentricity of 100 μm can be reduced to less than 0.2 μm. Crosstalk suppression between adjacent tracks is better than 30 dB. Demodulated video signal-to-noise ratio is about 45 dB.

[1] Several different multiplexing schemes for recording audio and video information on a video disk are under evaluation by Philips/MCA. Shown here is the NTSC crossband system where brightness (A), color (B), and sound (C) are modulated on separate carriers. If these signal components are combined and then clipped (D), the signal peaks form a square wave of varying pulse width. This is the signal that modulates the laser scribe. Individual depressions in an actual video disk are 0.1 μm deep, 0.7 μm wide, and between 0.8 and 2.5 μm in length (E,F).



The objective lens is mounted on a movable coil suspended in a magnetic field (like a loudspeaker voice coil). This allows continuous compensation for disk warping or other vertical irregularities. Depth of focus is short, so scratches, dust, and fingerprints on the video disk's exterior are "overlooked" by the readout. The aluminized, information-carrying surface is protected on its active side (facing the lens) by transparent plastic, while the reverse side is shielded by a hard coating. The disk is much less sensitive to general handling than a normal LP phonograph record.

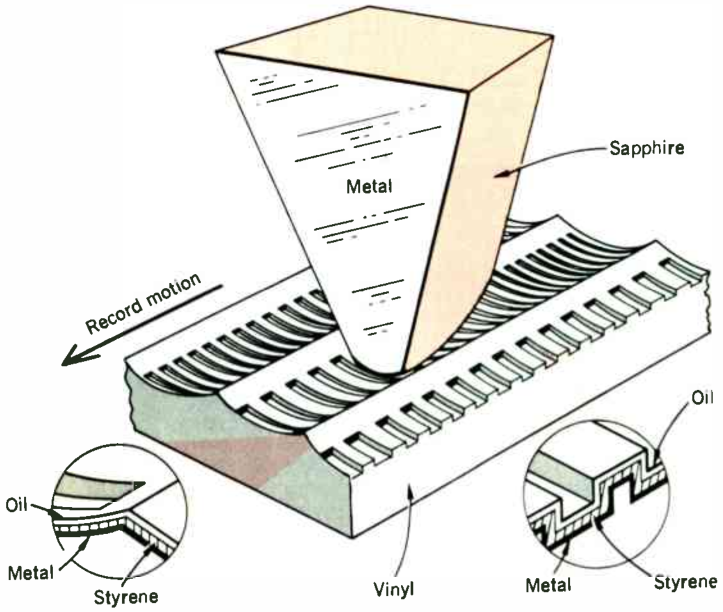
A relatively recent convert to this reflective-type video disk design is the Zenith Radio Corp., Chicago, Ill. Originally working with Thomson-CSF on a transparent video disk that carried information in the form of microscopic pits on the outside surface, Zenith engineers found that fingerprints tended to fill these tiny depressions with fluid, making the recorded information almost invisible (disk plastic and skin oil have a similar refractive index). And the exposed surface was quite sensitive to scratches.

Like Philips/MCA, Zenith engineers have also demonstrated a video player. The Zenith model uses a single mirror, tiltable about two axes, for correcting speed-related playback error. Raw timing errors of about 5 μs are produced by eccentricity alone, and must be reduced a thousand fold to recover a conventional NTSC video signal with standard FM encoding/decoding techniques. Simpler alternatives for video recovery are being actively explored.

Capacitive pickup is groovy

Indeed, simplicity could well determine the successful adaption of video players to the home environment. RCA's approach, retrieving video information with a metalized sapphire stylus tracking a finely grooved disk, boasts a design built from widely available electrical and mechanical components. "Our space-age technology has been kept in the factory," says RCA's Richard Sonnenfeldt, staff vice president, video disk operations, referring to the elaborate and expensive electron beam recorder RCA developed to cut video disk masters.

[2] The disk surface and stylus tip of RCA's video player system (SelectaVision) meet in a spiral track of 2187 grooves per cm. Information is recorded as transverse slots of varying width and separation (length and depth are fixed) in an otherwise smooth groove. Both groove and stylus are metalized to implement capacitive detection of the relief pattern as the disk turns at 450 r/min. Brightness, color, and audio are encoded in the zero crossings of the signal recovered from the disk.



Though the RCA video disk is grooved, the sapphire stylus does *not* follow complex undulations like a phonograph needle. Instead, the stylus tracks a shallow trough, the bottom of which is ribbed with rectangular slots (Fig. 2). These slots are composed of several distinct overlaid sections. Just below thin layers of lubricant and dielectric is a subsurface of metal. The stylus too has a metalized area extending across its face right to the point of contact with the video disk.

Upon rotation, the relief pattern in the disk is detected as a change in capacitance between the stylus electrode and the moving metal subsurface. Actual change is approximately 3.5×10^{-16} farad.

The stylus–disk capacitance is part of a resonant circuit (at about 915 MHz), detuned by capacitance variation (Fig. 3). When driven by a local oscillator of suitable frequency (on the skirt of the resonance curve), this variable-frequency resonant circuit exhibits variable impedance that controls the local oscillator signal as it passes through. The resulting amplitude modulation (impressed on the local oscillator signal) is removed by a diode detector. Output from the detector is an FM signal directly related to the passage of slots under the sapphire stylus (the slots are of continuously varying width). This FM signal is finally demodulated to provide a composite video waveform. Audio components are recovered by appropriate filtering followed by FM detectors.

Small perturbations in playback speed, due to synchronous motor hunting, disk centering, etc., are corrected by an "arm stretcher." This unit consists of a small electromechanical transducer (similar to a moving coil loudspeaker element) which drives the stylus arm back and forth, parallel to the grooves. Error signals to control the arm stretcher are derived from measurements of the color burst frequency as the video disk is played. Dropout compensation is provided by a 63-μs delay line, appropriate sensing circuitry, and video switches that substitute the video from a previous line whenever defects occur on the present line. Video signal-to-noise ratio is pegged at greater than 40 dB.

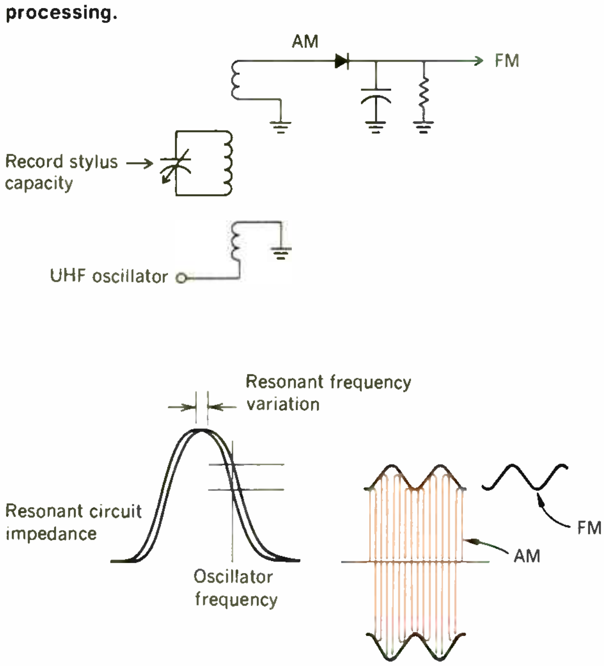
Another major difference between RCA's SelectaVision and the Philips/MCA reflective disk is speed. Though no sluggard at 450 r/min, the RCA disk runs at only one quarter the angular velocity of its optical counterpart. As a result, four complete TV frames are produced during each revolution of the disk, making it less suitable for stop-action and slow motion effects (SelectaVision will offer neither feature when initially marketed in 1976). However, lower rotational speed does simplify the problem of correcting time base errors. RCA engineers explain that the timing error associated with 450-r/min operation is low in frequency, permitting TVs with relatively slow horizontal synchronizing circuits (most U.S. and European made sets) to derive acceptable playback from SelectaVision's output without circuit modifications.

RCA's video disks are expected to last in excess of 500 plays, while stylus life is 300 to 500 hours.

Turning lumps into lumens

Despite recent technical demonstrations indicating video disks are more a soon than someday proposition, "first-on-the-block" fans must stash their disposable income a bit longer—unless they live in West Germany. Starting in March 1975, the TED (Television Disk) system was made available to West German consumers. This is the first video disk and player combination sold anywhere, and represents a cooperative effort between Telefunken of Hannover, Teldec (Telefunken/Decca) of Hamburg, and Decca of London.

[3] As part of a resonant circuit, SelectaVision's variable disk-stylus capacitance causes small changes in tuning. A local oscillator signal is situated along the resonant curve's skirt where variable impedance effects are most pronounced. The resulting AM modulation (impressed on the local oscillator signal) is removed by a diode for further processing.



Like RCA's SelectaVision, Telefunken/Decca's TED system also uses a contacting (mechanical) stylus to pick information off a video disk (Fig. 4). The TED disk has actual "bumps" on the bottom of its finely grooved, elastic surface which are gradually deformed and then sprung back into position by a diamond skid. The quick release of these bumps becomes a mechanical impulse that travels through the skid to a piezoelectric transducer. Amplification and filtering of the output separates the color, brightness, and audio signals. Finally, a low-power local oscillator (tuned to an unused TV channel) is modulated with the recovered program.

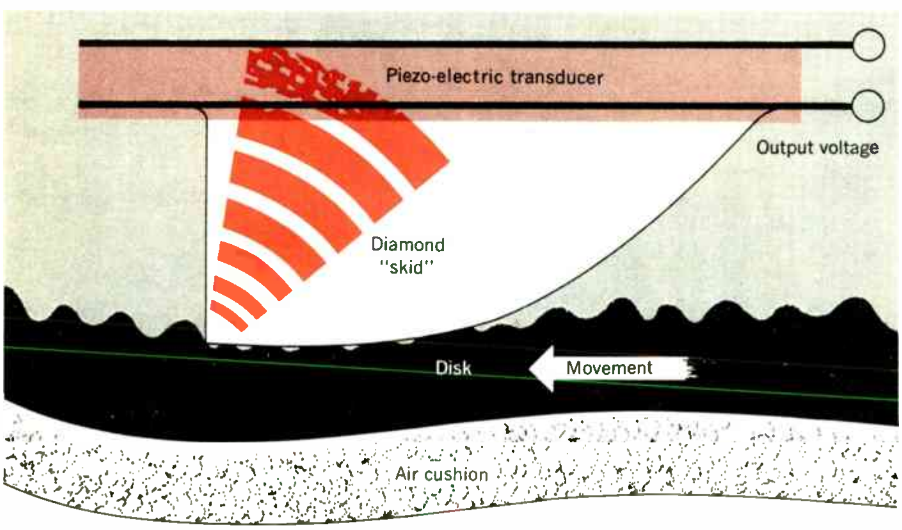
Telefunken/Decca's wafer-thin video disks do not ride on a turntable, but are connected directly to the player's drive motor (Fig. 5). The stationary plate under the disk has several small holes near the motor drive shaft allowing air circulation. Centrifugal force imparted to air molecules under the spinning disk causes a radial flow. Fresh air is drawn up through the holes and a continuous stream is established, "floating" the extremely flexible disk on a near-frictionless cushion. The plate beneath the video disk actually forms a shallow curve, or "saddle," and imparts this bent condition to the moving foil. The advantage is reduced flutter compared to a disk spun perfectly flat.

With a diameter of only 21 cm and rotating at 1500 or 1800 r/min (depending on desired compatibility with local TV scanning rates and power line frequencies), the TED disks play just ten minutes each. Telefunken predicts they could become low-cost video "inserts" for magazines—or even newspapers!

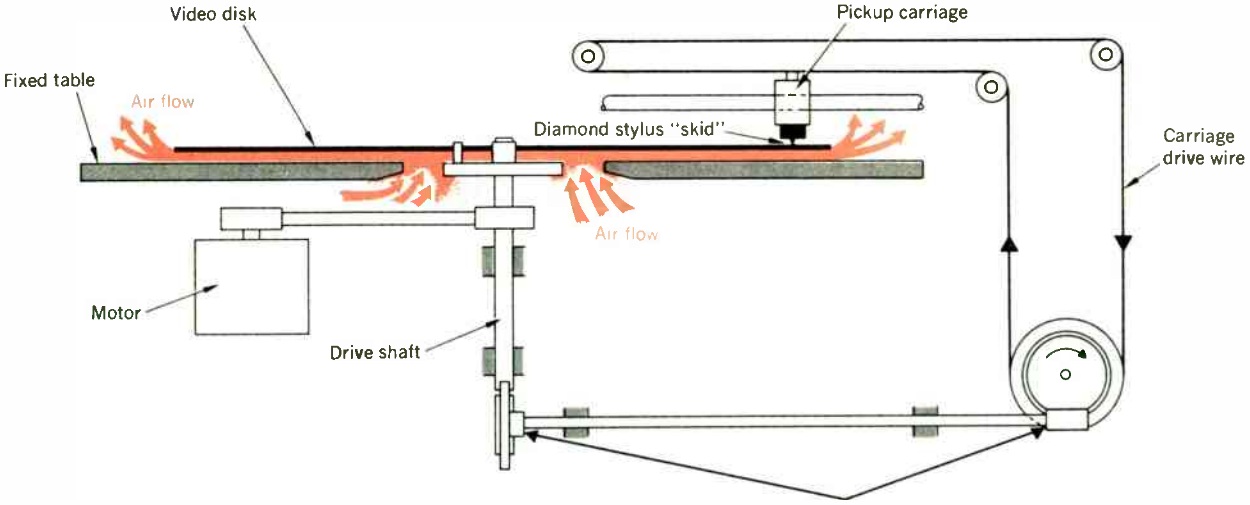
The TED player is unique in that it includes an automatic "self-sharpener" for the diamond skid. After each 10-minute viewing interval, a recessed burnishing wheel is pressed against the stylus for four seconds. The skid is smoothed to its original shape, maintaining image quality during the next play.

Living with the limited run

The eventual dominance of *one* video disk technology could sell players like hotcakes, but what about disks? After all, will anyone watch a feature film over and over the way they listen to an album of music? And what about program selection? Mastering video

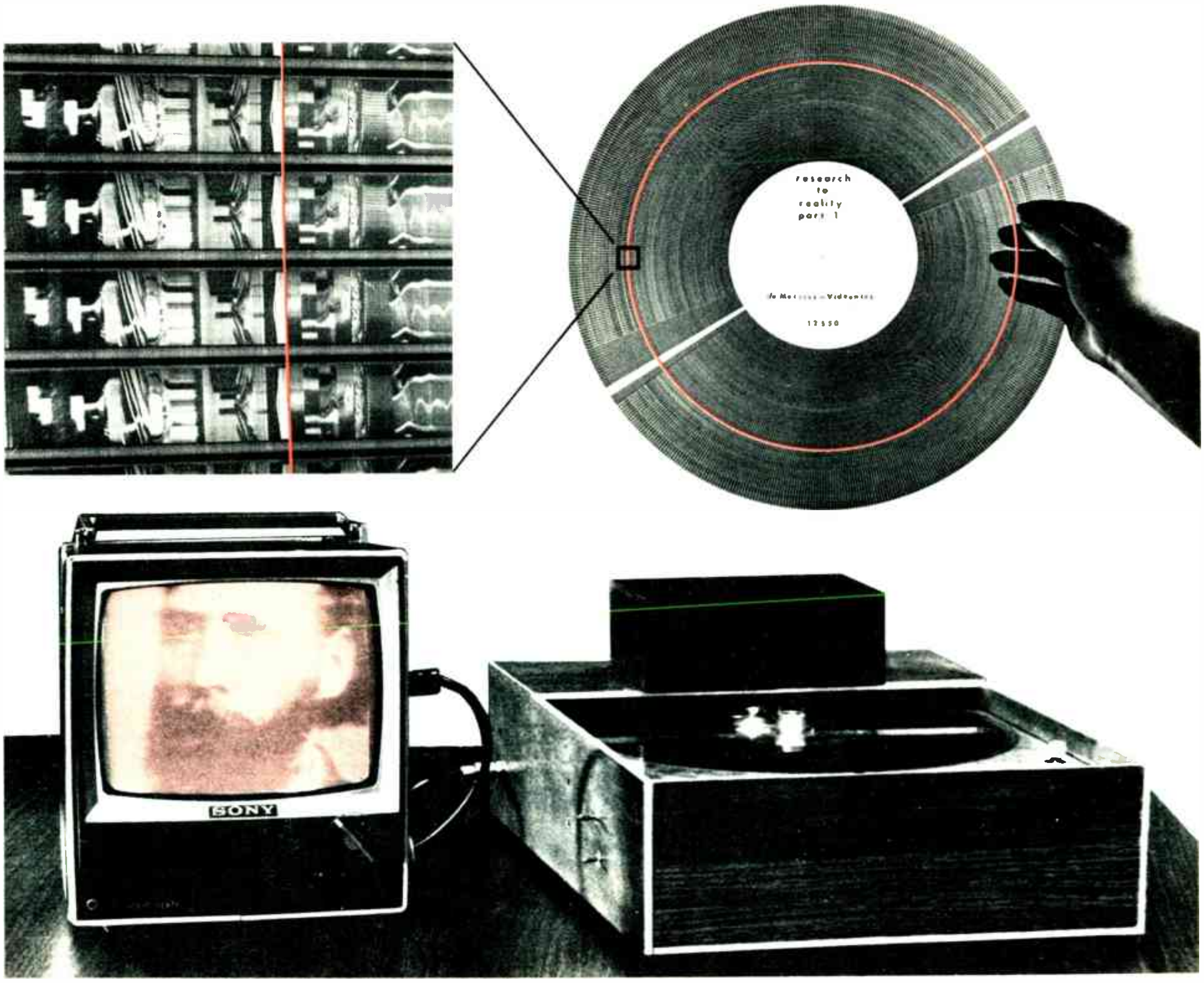


[4] The diamond skid employed by Telefunken/Decca's TED video disk system is many times longer than an individual bit of recorded information. As the flexible, plastic disk moves, the skid gradually compresses the information "bumps" along the bottom of a given groove. Each "bump" is abruptly released upon emerging from under the skid, producing a small mechanical impulse. A piezoelectric transducer cemented to the skid converts these vibrations to a pulse train for amplification and filtering. The TED disk contains about 280 grooves per mm and plays for ten minutes. Video S/N ratio is 40 dB.



[5] Centrifugal force imparted to air molecules beneath a spinning TED video disk sets up an "air cushion" that supports the disk during play. The pickup rides in a special carriage that is geared to follow the disk's nominal groove pitch. Stylus mounting is carefully damped to keep mechanical resonances well outside the operating range, but enough lateral movement is allowed to track small eccentricities and warps.

[6] This experimental video disk from i/o Metrics (upper right) is made from fine-grain holographic film. One full revolution (at 1800 r/min) is equal to one completely scanned TV frame. On playback, the i/o Metrics player connects directly to a standard TV's antenna terminals (bottom). A closer look at the disk (upper left) reveals radial bands that correspond to the horizontal retrace interval (solid black bars) including the front porch and back porch (associated gray bars). The vertical retrace intervals occur only twice per complete frame, but are of considerably longer duration. They can be seen directly opposite each other on a 45 degree diagonal in the upper right photo.



disks with laser cutters and electron beams is expensive. Issuing just guaranteed "hits" will defeat the prospect of providing diverse, unusual programming.

One possible answer is yet another video disk system (employing fine-grain photographic film) under development by the i/o Metrics Corp., Sunnyvale. Calif. Company physicist John A. Jerome believes even tiny runs of 10 to 1000 copies could be handled economically with diazo duplicates of silver emulsion video disk masters.

The i/o Metrics system records at 1800 r/min on a 30-cm (12-inch) diameter disk of unexposed fine-grain film. The film is a special type originally created for holographic photography, but its fine structure allows accurate recording of the high information densities associated with real-time video record/playback. A modulated low-power laser beam is focused through a microscope objective during the recording process, producing a track of varying width and transparency when the latent image is later developed.

While any FM or AM encoding technique allowing adequate bandpass and linearity can be used for mastering, i/o Metrics favors direct analog recording of the video signal, omitting any carrier frequency. This greatly simplifies the record and playback electronics (for video), but requires some juggling to accommodate the soundtrack. Encoding audio in both the horizontal and vertical retrace intervals has been attempted, with vertical retrace recording proving the better alternative. To date, i/o Metrics has publicly demonstrated playback of black-and-white and color images *without* sound (Fig. 6). The best video signal-to-noise ratio obtained thus far is 42 dB.

Though not yet a fully developed consumer entertainment medium, the i/o Metrics player and disk could serve as keen competition to microfiche for high-density, low-cost data storage. Estimates made by i/o Metrics as of October 1974 show that a minimum of 10 000 video frames (10 000 printed pages) can be stored on one video disk at a materials cost of 19 cents per diazo copy (assuming 1000 or more disks are produced). The player itself would cost $300, with recording and duplicating equipment estimated at $15 000 and $40 000, respectively.

The i/o Metrics player features a flexible floating disk—riding on air cushion—much like Telefunken/Decca's TED system. The playback transducer is a photodiode powered by an incandescent lamp.

Tape's trump card?

Underlying the variety of video disk technologies is a commitment to handling extremely high information densities at low cost—a goal almost achieved in several instances. Meanwhile, despite record/playback, editing, and re-use features, video tape has not been able to follow its audio counterpart from professional to consumer applications. Is high-density magnetic recording impractical? Not according to the Sony Corp., Tokyo, Japan.

Nobutoshi Kihara, managing director and general manager for Sony, reports that his engineers have succeeded in reducing video tape consumption to less than one tenth of that associated with typical reel-to-reel or cartridge systems. The Sony system (called Mavica) employs a magnetic audio/video card encased in a protective jacket. This envelope, containing a card, is inserted into a Mavica player that provides automatic handling during record or playback.

Recording density has been increased for the video "Mavicard" by narrowing the width of the recording track and giving up the guard band. A special phase-modulation method is used that aligns the center carrier phase between adjacent tracks. Signals that overlap (a mixture of signals from two tracks) play back without generating a "beat." Also, a skip-field technique is employed that further increases recording density. Slight shifts in tracking cause almost no deterioration of the playback image.

In magnetic video recording, the playback head output depends on the minimum recorded wavelength and the width of the recorded track. To obtain an acceptable signal-to-noise ratio, a wavelength of 2 μm requires a track width of 20 μm. However, the special phase modulation method just mentioned not only eliminates need for a guard band but also allows acceptable playback with an overlap of two or three frames. This means the track pitch can be reduced from 20 μm to almost 8 μm. Video signal-to-noise ratio on playback is reported to be 40 dB.

Based on these techniques, the Mavica system allows five minutes of color recording and playback on a small magnetic card (about 16.5 × 16.6 cm). During playback, card "advance" can be controlled manually to provide slow or stop motion.

Red ink blues

The development of various video players for home use has stretched over a dozen years and consumed millions of investment dollars. For every company developing a player, there are many others whose business could be greatly affected (hopefully, for the better) by consumer acceptance of one dominant technology. TV set manufacturers and movie film companies, for example, could make design modifications and licensing agreements with confidence. The price of a wrong commitment is fearsome indeed.

Just a few years back, video tape cassettes were widely acclaimed as the next logical step in the evolution of home entertainment electronics. One of the most ambitious systems—Cartrivision—was promoted as a complete package *including* a new color TV. Cartrivision's developers sought licensing agreements with set manufacturers, but made no immediate provision to exploit the add-on market.

Despite some very desirable product features (like record/playback), success was not to be. On May 14, 1975, the capital equipment of Cartridge Television Inc.—all purchased new from 1971 to 1973—was sold at public auction in Richmond, Indiana. With a rap of the gavel, Cartrivision passed into history. ◆

Portions of this article are based on video disk papers presented at the Montreux Television Symposium held in Switzerland. May 1975. Symposium coverage was provided by Spectrum Editorial Board member Fritz E. Borgnis.

Semiconductor memories

More users are opting for RAMs, ROMs, and PROMs as chip function densities increase and price per bit drops

In 1970, little if any of the computer mainframe memory market was served by semiconductor memory-bits. Today, semiconductor memories serve anywhere from 30 to 50 percent of this market. Predictions are that, by 1980, well over 60 percent of all computer and minicomputer mainframe applications will make use of semiconductor memories, and this percentage does not account for IBM, which designs its own semiconductor memories for its computers.

In the past, magnetic ferrite cores were the dominant storage medium in computer mainframes. This may still be so today, but the situation has been rapidly changing since the introduction, in 1970, of the p-channel, metal-oxide-semiconductor (PMOS), 1024-bit (1-kb) random-access memory (RAM). More recently, 4-kb, n-channel (NMOS) RAMS have become available, and these are expected to accelerate the inroads of semiconductor technology into mainframe memories. Some designers liken the 4-kb RAM's impact to that of transistor–transistor logic (TTL) on logic ICs. In fact, most new computer and minicomputer designs are now using semiconductor memories. Many of the new designs are exclusively semiconductor.

While 4-kb RAMs have held the semiconductor memory spotlight, as one of the few dynamic segments of an otherwise weakened semiconductor market, much more is happening or about to happen in semiconductor memory technologies. Here are just a few highlights:

- An explosive growth is taking place in the number of bits that MOS read-only memory (ROM) packages can handle, with 32-kb units available now in dual-in-line packages (DIPs), and 64-kb units predicted within a year.
- Reprogrammable MOS ROMs are also getting larger with 8-kb units now available.
- A host of new applications have arisen for both bipolar and MOS ROMs and programmable ROMs (PROMs) in consumer and industrial markets, spurred on by parallel developments in microprocessors. Bipolar PROM prices have dropped markedly.
- Complementary MOS (CMOS), the low-power-dissipation, high-noise-immunity semiconductor process, is being applied in increasing volume to the manufacture of RAMs—1-kb CMOS RAMs that operate from a small battery are already available.
- The newest member of the memory-related semiconductor family is the field-programmable logic array (FPLA), meeting the need in microprocessor and other computer systems for nonstandard logic functions.
- Probably the hottest area, next to 4-kb NMOS RAMs, is integrated-injection-logic (I^2L) memories. I^2L is a bipolar technology that is said to have the density and low-cost advantages of the MOS process.
- Charge-coupled devices (CCDs), low-cost and high-capacity serial-access memories, have now become commercially available in DIPs at up to 16-kb densities.

Roger Allan Associate Editor

NMOS at the top for RAMs

Starting with the Intel 2102 NMOS static 1-kb RAM in 1972, NMOS became the MOS process to offer the best combination of memory access time, density, low power dissipation, and low cost per bit. Compared to the older PMOS process utilized in the highly popular Intel 1103 1-kb dynamic RAM, NMOS offers nearly a factor of two increase in operating speeds due to lower capacitances and threshold voltages. Except for MOSTEK's 4-kb metal-gate NMOS RAM, all new 4-kb dynamic RAMs are silicon-gate NMOS designs (EMM Semi has an NMOS static design).

Unlike static RAMs, dynamic RAMs in most cases require constant memory refreshing from additional support circuitry. However, they are less expensive, are higher in storage density, are much faster, and dissipate less standby and operating power than static RAMs in comparable situations. Access times of 200 ns and less have already been achieved. For small systems such as point-of-sale terminals, the static RAM is more flexible and economical since the higher static RAM costs are easily outweighed by the flexibility afforded by not having many additional supporting components.

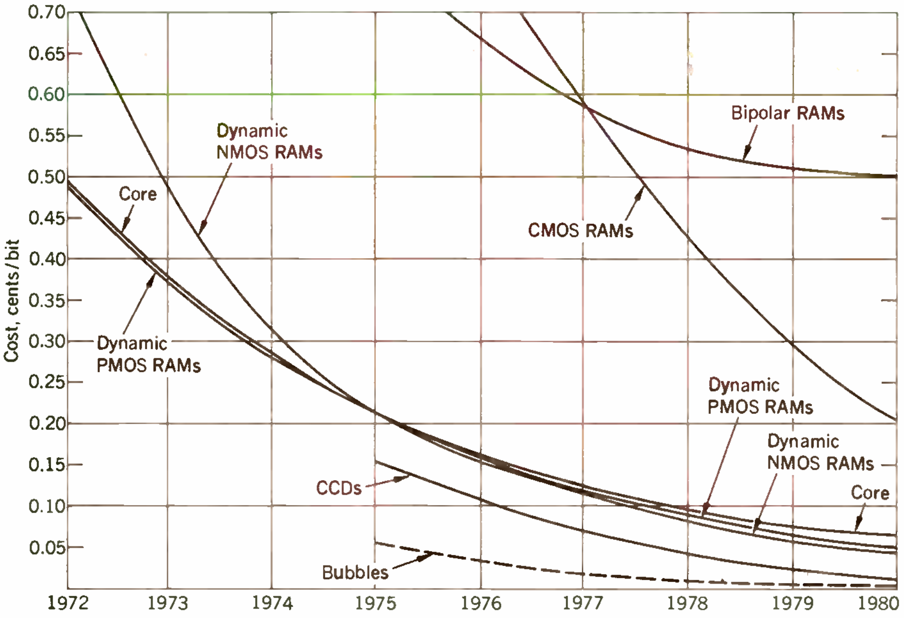
Unlike core memories, where lower costs per bit are usually realized with increasing system storage capacity, semiconductor memories generally rely on increasing package densities to reduce per-bit costs. Current board-level RAM prices of 0.5 to 1¢ per bit (0.25¢ per bit at the component level) are roughly competitive with board-level core prices which are in the 0.25¢ to 0.5¢ per-bit range. By the end of this year, 4-kb RAM prices are expected to decrease to 0.15 to 0.2¢ per bit, at the component level, as larger volumes of them are sold. Figure 1 gives a per-bit cost breakdown, at the component level, for various memory technologies.

Figure 2 shows what the new 4-kb RAM density permits. A 256-kb TTL-compatible system, organized as 32k by 8, with all supporting circuitry, can now be assembled on a 19- by 34-cm (7.5- by 13.5-in) printed-circuit board, using NMOS dynamic 4-kb RAMs, thereby dramatically increasing system bit density.

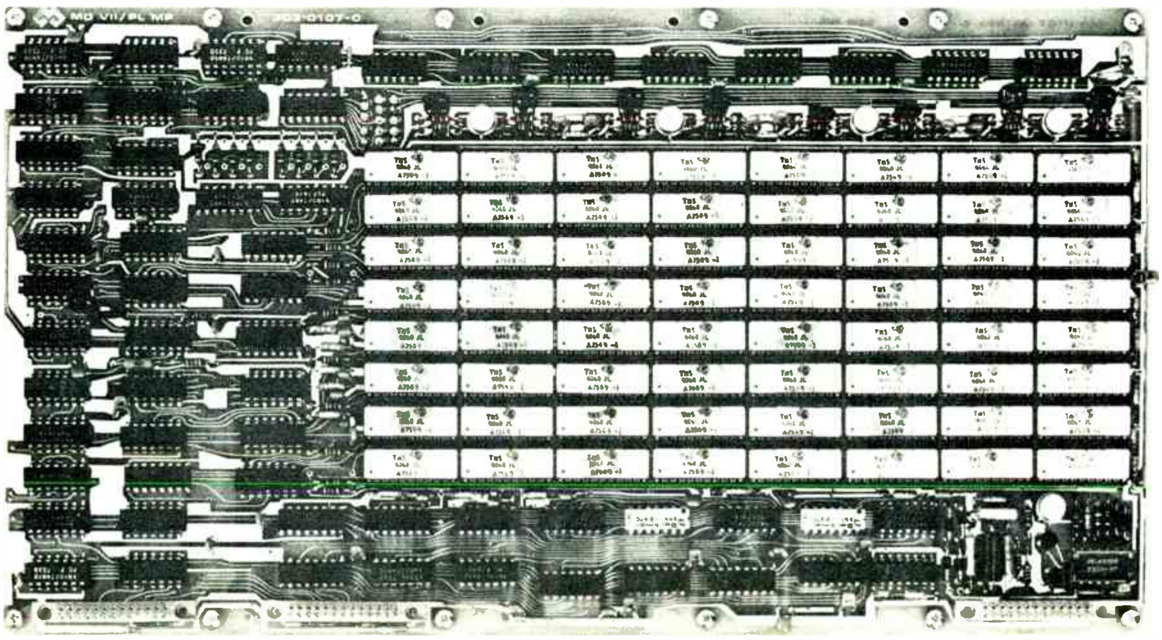
While some semiconductor memory designers are understandably optimistic about the future of RAMs —spearheaded by the 4-kb NMOS dynamic RAMs, as replacements for mass-storage disk and drums—others including users are less optimistic in their expectations and projections. The latter group point out that disk and drum systems will be with us for at least another decade; these systems are nonvolatile, very low in cost, fairly reliable, and relatively compact. A related memory area that may be served by semiconductors is the portable file-storage, now provided by floppy and fixed-head disks. Here, better semiconductor memory reliability and the potential for higher densities and lower costs make semiconductors a very attractive alternative for the 10^6- to 10^8-bit range. However, constant improvements in fixed and moving-head disk and drum systems—by getting larger bit densities on the same surface area—have made possible 10^9-bit systems for about $300 000, or 0.03¢/bit, a price that includes all control electronics. While it is possible that per-bit RAM pricing will be at this level by 1980, disk and drum systems can also be expected to drop in per-bit cost as a result of technological improvements.

The three-way package-pin tug-of-war

The new dynamic NMOS 4-kb RAMs are available in one of three packages: 16-, 18-, or 22-pin DIPs, and there are no less than three different commercially available pinouts, none of which has emerged as the clear standard. This has caused much user confusion. The 16-pin package requires external multiplexing circuitry; with the 18-pin RAM, read-write data must share a common input-output bus. The 22-pin package overcomes both these limitations, but it takes up much more area on a printed-circuit board—some 56 percent more area than a 16-pin RAM and about 45 percent more than an 18-pin RAM. Today, the 22-pin



[1] NMOS dynamic RAMs are expected to drop in cost to 0.04¢/bit by 1980, below the projected 0.06¢/bit price of cores. CCD memory prices are shown starting with this year, when they first became commercially available. Bubble-memory prices are projected throughout since they haven't been commercially available. All prices are at the component level, and do not include peripheral circuitry prices.



[2] Designed around Texas Instruments' 4096-by-1 NMOS dynamic RAMs, this Monostore VII/Planar memory system from Monolithic Systems Corp. packs a 256-kb TTL-compatible system (organized as 32k by 8), with all supporting electronics, on a 19- by 34-cm (7.5- by 13.5-in) printed-circuit board.

package is the one most widely used, but packaging density remains an extremely important factor for many applications, and ever. the proponents of the 22-pin package readily admit that with improvement in operating characteristics, either 16- or 18-pin designs might eventually become the standard.

One large 4-kb RAM user, the Hewlett-Packard Co. of Cupertino, Calif., has started using the 22-pin package in its miniature 21MX minicomputers, and now believes that the 18-pin package from Texas Instruments (which, incidentally is *not* pin-for-pin compatible with National Semiconductor's 18-pin design) is the best package for its system. Texas Instruments is a sole source on its 18-pin RAM while National is backed up by three second sources—Signetics, Monolithic Memories, and Advanced Memory Systems —for its 18-pin RAM. Hewlett-Packard says that the 18-pin RAM gives the density of the 16-pin RAM without sacrificing operating speed for address multiplexing.

As for bipolar RAMs, 1-kb memories are commercially available, with 4-kb RAMs under investigation. Bipolar RAMs with TTL, Schottky-TTL, and ECL technologies will continue to be used for high-performance applications where operating speed is an important consideration. Many 1-kb bipolar RAMs have access times as fast as 45 ns, with ECL providing the fastest speeds.

TTL-compatible static, 1-kb, bipolar RAMs are now opening the way for practical use of bipolar RAMs in large-storage-capacity systems. The ease of interfacing such RAMs allows the design of memory systems in which the control and peripheral circuitry adds little to the system's total access time, as compared to what would be added to the system by slower MOS RAMs.

ROMs, PROMs, and EAROMs proliferate

The expansion of microprocessor and intelligent-terminal applications has meant a tremendous increase in both the types and bit-capacity of bipolar and MOS ROMs available. Plans by some manufacturers to introduce high-capacity 64-kb MOS ROMs by the end of this year could materialize by the time this article is in print. The 32-kb size is already commercially available.

The biggest reasons for increasing ROM densities, usually done with the higher-density and lower-cost MOS process, is to reduce package count on the board and increase the number of functions in a ROM package. Another reason is to decrease the large mask cost. Many MOS ROM manufacturers feel that the possibility of making up an entire computer–compiler system from a few precoded ROM chips is not too far in the future—if MOS ROM bit densities continue to climb, simplifying software program development. Per-bit prices for large MOS ROMs are down to from 0.05 to 0.1c.

Some ROM manufacturers make equivalent-density PROMs, using the same pinout and packaging configuration. Except for Intel's 1602 and National Semiconductor's 5203/5204 MOS electrically alterable ROMs (EAROMs), which have no quartz windows for data erasure by ultra-violet light and thus can be technically considered PROMs (data is written in by electrical pulses), PROMs available today are of the bipolar variety and are available up to 4 kb in density.

Of course, the greatest flexibility in altering one's ROM program at the design stage is provided by EAROMs that allow more than a one-time program change. Their drawbacks include the relatively high voltages needed to reprogram them electrically, and the time it takes to erase the data, usually by ultraviolet light, which can take as much as one-half hour. The metal-nitride-oxide-semiconductor (MNOS) method, is being carefully weighed as an alternative reprogramming method. Nitron, a McDonnell-Douglas subsidiary, National Cash Register, and Westinghouse employ this technique in their reprogrammable ROMs.

The great majority of ROM/PROM manufacturers,

I. Characteristics of semiconductor memory technologies

Technology	Access Time (ns)	Cycle Time (ns)	Operating Power Dissipation per Bit (μW)	Density (bits/chip)
PMOS	300–400	500–600	100	4096
NMOS	150–250	350–450	100	4096
CMOS	50–60	500–600	20	1024
CCD	80–250*	120–500*	10	16 384
Bipolar TTL	50–60	80–100	500	1024
Bipolar ECL	45–50	60–70	500	1024
Bipolar I^2L †	50	100	–	–

* Average.
† Under development.

Testing, testing, testing . . .

Those designers contemplating the use of the new dynamic 4-kb NMOS RAMs better stop for a moment and realize that all is not "peaches and cream." As with any new and complex technology, the 4-kb RAMs demand much testing and system debugging before they can be put to work with confidence.

Many large users are relating stories of large problems in correlating testing of the 4-kb RAMs with the manufacturer. Just determining the correct number and types of tests can take quite a long time. What the manufacturer may consider reasonable tests may not necessarily suit the user for the application at hand. Users report having to build special test fixtures and enumerate special test procedures for some very complex and sophisticated testing. A major instrument manufacturer believes that device makers may optimize performance specifications by choosing test conditions that yield the maximal value of a given parameter.

One user complained of the low yields he was getting after instituting his own burn-in tests, in the plant, to reach a desirable goal of reliability. Hewlett-Packard says it has attained satisfactory reliability figures of 0.11 meantime-between failure per 1000 hours, after in-coming burn-in tests. H-P's objective was a minimum failure rate of 0.15 per 1000 hours, equivalent to core-memory systems.

In fairness to the manufacturers, the 4-kb RAM is a complex IC for which complete and adequate testing procedures have yet to be standardized. As with any new technology, this simply takes time. Even 1-kb RAM tests are still evolving.

including those who make the MOS EAROMs, feel that it would be impractical to try to make a MOS PROM that works in much the same way a bipolar one does. This is due to the high current densities needed to do the programming function, as evidenced by MOS EAROMs. Current state-of-the-art EAROMs require microsecond-wide pulses from 25 to as high as 75 volts, in the 98-percent duty-cycle range. Nevertheless, at least one manufacturer is seriously considering a bipolar fusible-link approach in making a MOS PROM.

Another PROM concept receiving renewed interest due to its interface compatibility with CMOS logic is the programmable diode matrix. Introduced by Harris Semiconductor back in 1967, the diode matrix is an array of diodes in fusible links that can be used, together with small-scale-integration ICs to generate small user-programmable ROMs or PLAs. These circuits are finding new applications in distributed systems where programmable address recognition and instruction decoding is required, as well as in small systems to augment ROM storage when only a small number of bits must be programmable.

There is now better than a 1:1 PROM/ROM ratio in the number of bipolar semiconductor memory components being used by designers. If a similar trend developed in customer requirements for MOS memory components, at least one manufacturer feels that a greatly changed MOS PROM market could result. A MOS PROM, made with the same layout and design rules as a MOS ROM, could be half the price, on a per-bit basis. MOS PROMs are also cheaper and dissipate less power than equivalent-capacity bipolar PROMs.

Bipolar PROMs are programmed in the field by the user (or at the factory at the user's request) by one of two general methods: either by blowing open a fuse link or by creating a short circuit. The most popular fuse material is nichrome, which, after some initial reliability problems (the open fuse tended to regrow), has achieved a measure of reliability that is instilling confidence in many users. Two other fuse materials used are titanium-tungsten by Texas Instruments, which says it is more reliable than nichrome, and Intel's polysilicon. The short-circuit method known as avalanche-induced migration is used by Intersil and involves blowing a diode-junction into a short circuit.

The big news in bipolar PROMs is their plummeting prices. Back in April 1969, when Harris Semiconductor introduced the first PROM (a 512-bit bipolar unit), it sold for around 10c per bit. Bipolar PROM prices are now down to the 0.5c-per-bit level for 4-kb PROMs and 0.4c for 1-kb PROMs, thanks largely to the record numbers being sold for use in low-cost microprocessor circuits, where the freedom to experiment with one's microcomputer program until a level of confidence is reached is a big plus for designers. In fact, some MOS and bipolar memory manufacturers encourage designers to use PROMs before ROMs, particularly for small to moderate-storage applications, until a definite ROM pattern is agreed upon.

CMOS—the power miser

A hot new challenge to the 1-kb NMOS static RAM is the 1-kb CMOS RAM, now available from several manufacturers. Despite the fact that CMOS is somewhat more costly than NMOS, it has much lower power dissipation, and when combined with substrates, made of sapphire or spinel, high-speed silicon-on-sapphire (SOS) and silicon-on-spinal (SOSP) RAMs are possible with active power dissipation low enough to make possible operation from very small batteries, thereby making nonvolatile portable memory packages practical. As a point of comparison, standby power dissipation of a 1-kb CMOS RAM can be as low as 100 μW, against 100 to 150 mW for an NMOS static RAM of the same bit capacity. Even active power-dissipation levels of about 20 mW compared to about 200 mW make CMOS look attractive for low-power-dissipation applications.

Because CMOS is a static design and uses more silicon per cell than other MOS designs, it is generally agreed that the high cost of silicon will make CMOS useful for memories no larger than 1 kb in capacity. One memory manufacturer is known to be planning to introduce a 1-kb CMOS RAM by the end of this year with power drain, in the standby mode, of a mere 10 μA. Intel has commercially available a 5101 1-kb CMOS RAM (organized 256 by 4) with 15 μA maximum standby power drain.

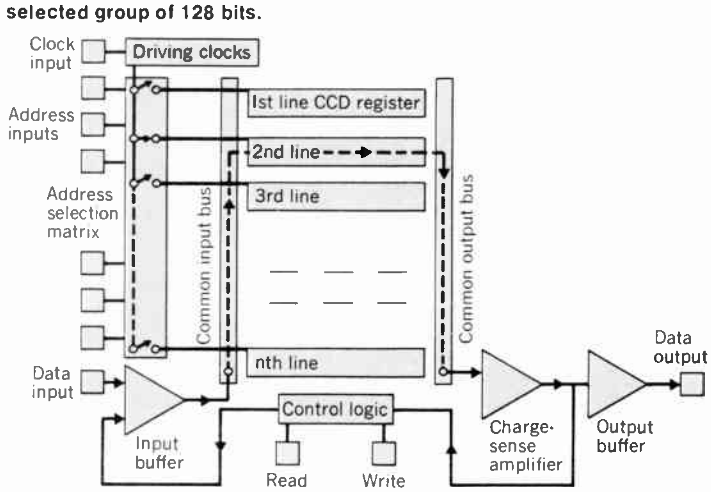
While there are few commercially available CMOS/SOS memories at the moment, work is underway by several companies to produce many such products. RCA has just announced a 1-kb CMOS/SOS static RAM with a 100-ns access time and just 1 mW of standby power dissipation. The company has produced and delivered other non-memory IC products using the CMOS/SOS process.

The emerging FPLA

Whenever the subject of projecting developments for the next few years comes up, designers of semiconductor memories usually include the field-programmable logic array (FPLA) as one of the important developments. In some cases, this is related to microprocessor developments, often where the same company that makes the memory also makes microprocessors.

When implementing random-logic functions in circuit design, a designer can choose from many stan-

[3] Fairchild's LARAM concept for its CCD memories speeds up the usually slow access time of serial-access CCD memories to 80 ns, for the first accessed bit of any selected group of 128 bits.



dard medium-scale-integration (MSI) and large-scale-integration (LSI) building-block ICs to simplify his design. However, he is often left with a good deal of decoding and control logic that doesn't conveniently lend itself to standard building-block ICs. The FPLA is said to be the answer to this problem.

An important development from Intersil is the company's new IM5200 FPLA, the first such commercially available unit. The IC is functionally equivalent to a collection of AND gates which may be ORed at any of its 8 outputs. Using the FPLA's 14 inputs, a designer has a total of 48 product terms and a complexity of over 480 4-input logic gates. Programming the FPLA is by electrical pulses.

Signetics also has an FPLA, its 82S100/101 (three-state/open-collector outputs). This Schottky device with a 50-ns maximum access time and a chip-enable feature has 16 inputs and 8 outputs.

Nonprogrammable PLAs have already been available from several semiconductor manufacturers, and are providing considerably more flexibility than ROMs as substitutes for nonstandard-circuit logic functions, since they can be easily configured to generate a wide variety of logic functions at relatively low cost and high speed.

I^2L for high performance

I^2L is a high-density (comparable to MOS) bipolar technique that, according to some of its advocates, does not invite compromise in performance as far as memories are concerned. Nearly every major semiconductor manufacturer is either actively looking at or about to introduce I^2L memory products (some this year) in RAM and ROM configurations.

There seems to be two schools of thought on making I^2L memories: one group feels that I^2L RAMs will be the next generation of high-density and low-cost components for main memory, while another group sees more use for I^2L RAMs and ROMs as elements for PLAs, arguing that I^2L higher performance far outweighs its lower cost. Neither group however disagrees that I^2L RAMs (static devices) will become lower in per-bit costs than static MOS RAMs, but neither is as confident when it comes to dynamic MOS RAMs. Some designers foresee 4-kb I^2L RAMs for high-speed applications currently not served by 4-kb NMOS dynamic RAMs.

One company, Fairchild Semiconductor, says it is planning to introduce this year RAMs made from its own I^3L® process, a process that combines I^2L with the Isoplanar® technology, providing Isoplanar density, and speed advantages along with I^2L advantages of density and low cost.

Whether I^2L or I^3L processes (or any other process in combination with I^2L) will be competitive in memories is not yet very clear, but as a new technology, I^2L is known be in need of more refinements for better production yields and higher speeds.

Will CCDs move in?

CCDs memories, of which 9- and 16-kb packages have now become commercially available, have been described as the semiconductor memories for high-density low-cost storage, filling the gap between high-performance, higher-cost semiconductor memories, and lower-performance, lower-cost disk and drum memories. An overview of current semiconductor, disk, and drum technologies, however, shows a large area of overlap, between CCDs and the new 4-kb RAMs, in this cost-performance gap. One thing is clear: at the component level, CCDs are currently cheaper than RAMs. However, this isn't yet clear at the system or board level. After all, a CCD is a serial-access form of memory that is, by its nature, slower than RAM for randomly accessing a bit (though it is much faster on a bit-rate basis). CCDs require

The bubble-memory prognosis

One mass-storage technology that has been talked about as a likely future candidate for low-cost high-density memories is that of magnetic-domain bubbles. In fact, there has been more than just talk. Many organizations, including Bell Telephone Laboratories and Rockwell International, to mention two, have demonstrated magnetic-domain bubble viability by building 100-kb bubble chips in the laboratory (Bell produced a 460 554-kb bubble memory in a package roughly 9 by 5 by 2 cm, while Rockwell produced a 100-kb memory on a 300- by 300-mil chip). But beyond building them in the laboratory, no commercial effort to produce them has taken place, this despite the fact that magnetic-domain bubble memory proponents have been saying that they've been ready to do this for the last two or three years. Why haven't bubbles entered the marketplace? Will they ever? Perhaps a closer look at some critics' comments might provide the answer.

First of all, the biggest argument against bubbles is that they are slow devices, and that despite efforts to improve operating speeds, they haven't been proven to be low enough in cost to compete with many other available and planned future mass-memory technologies. Even when their nonvolatility aspects are considered, critics argue, the system per-bit cost must become dramatically cheaper than what is available before bubbles can be considered as serious contenders.

Another argument, this one by semiconductor advocates, is that bubble memory work is performed using 3-μm-metal and 1-μm-gap technologies—advanced technologies supported by Government funds. If such technologies were made available (by Government funding) to semiconductor RAM manufacturers, critics say, 300-kb RAMs could easily be made today. The argument is that if bubble-memory people had to work with standard commercial technologies developed through their own funds, their projections of very low costs per bit wouldn't be as attractive as they now seem.

The fact that there is a limit to what can be done with bubbles—you can't make photosensors, analog circuits, logic elements (opinions differ on this) etc. from bubbles as is possible with semiconductors—makes them less versatile.

In fairness, it should be pointed out that semiconductor technologies themselves were aided by intensive Government funding during the 1950s, and that ultimately, bubble memory technology can mature to become viable commercially. When that might happen, no one is willing to estimate.

more overhead circuitry than RAMs, but they dissipate less power.

There is a large amount of disagreement on the relative cost advantage of CCDs over RAMs, among semiconductor memory designers. Some feel that the cost advantage of CCDs over RAMs, which is at the moment about a factor of 2, has to increase dramatically—and potentially is likely to do so—to be competitive with MOS RAMs for mass-storage disk and drum replacements. Some manufacturers say that until 100-kb (or higher) CCDs are developed, this expected cost advantage will not take place. Others feel that CCDs are clearly competitive with RAMs in block-oriented memory applications.

To get around the slow access times of CCDs, Fairchild developed the line-addressable RAM (LARAM) concept (Fig. 3), which allows access times anywhere from 80 ns (for the first accessed bit in a selected 128-bit string) to 25 ns (for the last bit). Fairchild feels that this will make CCDs competitive in performance and cost with other mass-storage techniques such as fixed-head and head-per-track disks and drums. Fairchild projects CCDs at 0.01¢/bit by 1980, with the LARAM concept.

Right now, two companies, Intel and Fairchild, are supplying 9- and 16-kb CCD memories. At the present rate of progress, a 32-kb CCD memory can be expected within a year.

An important development from Bell Northern Research of Ottawa, Canada, is a 16-kb CCD memory on a chip only 137 by 167 mils. This is considerably smaller than other 16-kb CCD memory chip sizes, and is comparable to the chip size of some newer 4-kb NMOS dynamic RAMs (a criticism of CCD memory by NMOS designers is its comparably larger chip size).

Terabit memories with electron beams

The use of electron beams for accessing memory by bombarding targets of silicon in small CRTs has been under investigation for a number of years by three principals: Stanford Research Institute (SRI), Menlo Park, Calif.; General Electric, Schenectady, N.Y.; and Micro-Bit Corp., Lexington, Mass. Such electron-beam-addressable memories (EBAMs), as referred to by SRI and Micro-Bit, or beam-addressable MOS memories (BEAMOS), as General Electric calls them, have the potential of storing millions of bits of information on silicon targets about one cm², making possible low-cost (projections of 0.02¢/bit after five years) terabit mass-memory systems with a few small CRT memories hooked together. Not only are density and low cost attractive properties—ones available with other nonelectronic techniques—but so are high speeds of a few microseconds in accessing many millions of bits of data. When density, low-cost, and high-speed properties are available in a single memory technology, that technology begins to look very attractive.

General Electric has just introduced a 32-Mb BEAMOS system with a 30-μs access time and a 10-Mb/s transfer rate. The entire system fits into a module 43 cm (12 in) long by 1 cm (4 in) in diameter, and is made up of four silicon targets, each packing densities of 8 Mb/mm², with access times under 10 μs, and 43-Mb/s data transfer rates. Micro-Bit feels that 10^8- to 10^9-bit EBAM systems are possible now.

While BEAMOS was developed primarily for military applications, General Electric says that present costs are about 0.02¢ per bit (at the system level) and projects this to drop to 0.001¢ per bit in the near future.

Those who argue against electron-beam systems point out that such systems are still largely experimental, having been used only in a handful of special applications (mostly military), and have yet to be proven reliable and cost effective. Other disadvantages cited are the use of bulky CRTs, and the necessary bulky power-supply hardware for such systems. Nevertheless, of all the exotic research efforts going on today for mass-storage techniques (holographic, laser, bubbles, thin-film, CCDs, etc.), the EBAM or BEAMOS efforts seem to have the best potential for the future, according to a majority of semiconductor memory designers.

And electron beams are being used for making mass-storage memories in another way. Recently, IBM announced a tenfold increase in LSI memory-chip packing densities, with an 8-kb memory, achieving 5-Mb/in² (32.3-Mb/cm²) densities by using electron-beam lithographic techniques with ion implantation. IBM developed this 8192-bit FET memory chip with an access time of 90 ns (the chip is fully decoded), using electron-beam lithographic techniques with line resolutions down to 1 to 1.5 μm. ◆

For further reading

Several publications dealing with specific areas of semiconductor memories are available from semiconductor manufacturers and other sources, many of which are application notes. Here are a few:

Application note AP-11 from Intel Corp., Santa Clara, Calif., deals with package selection and use of the new 4-kb NMOS RAMS, and is entitled "Which way the 4k .... 16, 18, or 22-pin?".

An excellent reliability study is available on the new 4k RAM in a report by Robert Frankenburg, Hewlett-Packard Co., Cupertino, Calif. The company is a large user of the 4-kb RAM in its 21MX minicomputers.

Another 4-kb reliability report is available from Texas Instrument's Inquiry Answering Service, Dallas, Tex.

A reliability study of the nichrome fusible-link technology as used in PROMs is available from Signetics, Sunnyvale, Calif. It is entitled "Signetics PROM reliability."

National Semiconductor, Santa Clara, Calif., will be making available a "how to" application note on using the new 4-kb RAM.

An excellent review of semiconductor memories is in the August 1975 *Proceedings of the IEEE*: "Techniques for large-capacity digital storage."

The 1975 AFIPS conference proceedings (vol. 44) of the National Computer Conference contains several interesting papers from a session on "Storage technology," dealing with mass-storage memory systems.

Rasmussen on reactor safety

How nuclear power reactor risks are quantified; and nuclear sabotage, theft, shipping, and waste disposal risks put in perspective

In August 1974, the U.S. Atomic Energy Commission (now the Nuclear Regulatory Commission) issued, in draft form, a report called "Reactor Safety Study; an assessment of accident risks in U.S. commercial nuclear plants." Known also as "WASH-1400," the report was compiled by a study group of about 70 experts directed by this author, and it comprises 14 volumes and 3300 pages at a cost of about $3 million.

Aimed mainly at reaching some meaningful conclusions about the risks of nuclear accidents, using current technology, the study provided a quantitative assessment of the risk to the public from reactor accidents. Additional goals of the project were to carry out a realistic safety assessment, as opposed to the conservative approach usually taken in the licensing process for nuclear power plants; to develop methodological approaches needed to perform these assessments, and to acquire an understanding of the limitations of these methods; to identify areas in which future safety research might be fruitfully directed; and to provide an independent check of the effectiveness of the reactor safety practices of industry and the Government.

The results of the Reactor Safety Study indicate that nuclear power plants have achieved a relatively low level of risk compared to many other activities in which our society engages. The study concluded that reactor risks measured in terms of fatal injuries are smaller than many other man-made and natural risks to which we are exposed as a society and as individuals. These other risks include those due to fires, explosions, dam failures, air travel, toxic chemicals, tornadoes, hurricanes, and earthquakes (see Table).

In addition to fatal injuries, the study also estimated various other possible consequences including latent cancers, genetic effects, thyroid illness, and property damage. Although there are consequences with delayed health effects arising from other risks, they are generally poorly known so there was no basis for comparison of the delayed effects with other risks. However, even in the worst accidents it was found that the numbers of latent cancers and genetic defects would be small compared to the normal incidence rate and it is doubtful that they could be detected statistically. This was not true for thyroid illness, however, where the number of cases in a bad accident would be readily detectable. Fortunately, most of these illnesses are treatable by current medical procedures.

The draft of WASH-1400 considered only those accidents that produced measurable effects on the public. Such accidents all involved the release of significant amounts of radioactivity. Since by far the largest amount of radioactivity is contained within the uranium dioxide fuel (a ceramic material), this radioactivity is effectively contained unless the fuel is overheated and melted. For this reason, the study was restricted to examining those accidents that could potentially lead to fuel melting. Although there are accidents other than fuel melting capable of some public consequences, the study concluded that because of their low probability of occurrence and relatively small amount of radioactivity, they do not contribute significantly to the public risk.

The study was restricted to accidents in the power plant itself. Many critics have raised questions about other risks such as:

- Hazards in shipping
- Waste disposal
- Sabotage
- Theft of fuel

Although not covered by the draft of WASH-1400, they have been the subject of other studies (most of which are not risk analyses like WASH-1400). In general, however, because the amounts of radioactivity involved are so much smaller, the consequences of accidents in these parts of the fuel cycle are much smaller than for serious accidents in the power plant itself. (This is, of course, not true for the terrorist use of a nuclear explosive.) These issues are briefly discussed later in this article.

Since the draft report, considerable work has been done to make the consequence calculations more realistic. Improvements in the weather model and the population distribution, in addition to the corrections and improvements suggested by the comments received, are being made. All of these improvements will be included in the final report expected to be issued in the fall. There is no doubt that many of the comments received have been very valuable and will

Probabilities of major man-caused and natural events

Type of Events	Probability of 100 or More Fatalities	Probabilities of 1000 or More Fatalities
Man-caused		
Airplane crash	1 in 2 years	1 in 2000 years
Fire	1 in 7 years	1 in 200 years
Explosion	1 in 16 years	1 in 120 years
Toxic gas	1 in 100 years	1 in 1000 years
Natural		
Tornado	1 in 5 years	very small
Hurricanes	1 in 5 years	1 in 25 years
Earthquake	1 in 20 years	1 in 50 years
Meteorite impact	1 in 100 000 years	1 in 1 000 000 years
Reactors		
100 plants	1 in 10 000 years	1 in 1 000 000 years

Norman C. Rasmussen
Massachusetts Institute of Technology

contribute considerable improvements in both the accuracy and understandability of the final report. In our calculations to date, it appears that none of the consequence curves will change by more than a factor of ±3, the original uncertainty assigned. The latent effects due to long-term radiation exposure will increase by about a factor of 3, however.

Methodology of the study

In order to estimate the probabilities of accidents, the study employed two methodologies called fault trees and event trees. These methodologies had been developed during the last two decades for the purpose of evaluating the reliability and failure rates of complex engineered systems whose failure rates were low enough so that they could not be measured from a reasonable amount of operating experience. They have also proven to be valuable for improving reliability in the design stage before any operating history has been obtained. Basically the techniques are used to break up the possible causes of system failures into various subsystems or components whose failure rate can be estimated from previous experience with that type of subsystem or component.

In the simplified event tree for a pipe break accident (Fig. 1), the basic objective is to identify the possible sequences of events following the break. The different outcomes depend upon the operability of the various systems that affect the course of events. The illustration is a simplified example of the technique. The actual trees are somewhat more complicated.

The fault tree is a logic roughly the reverse of the event tree. It starts with some undesired event and reasons back to the way it may have come about. In Fig. 2, the first few steps in a fault tree for loss of power to the emergency safety features (ESFs) are shown to illustrate the principle. Since either loss of ac or dc power to these features can cause the top event, these two events are coupled to the top event by an "or" gate. In the next level, we see that either off-site or on-site power is sufficient for success so both must be lost to produce the loss-of-ac event. They are coupled to the loss-of-ac event by an "and" gate to indicate this relationship. The trees are developed through many such levels until the failure events are pieces of equipment whose failure rates are known from experience in similar applications. It is then mathematically possible to calculate the probability of the top event. In this case, this probability could then be used for P_B in Fig. 1. Thus, the event tree defines possible accident sequences and the fault tree is used to calculate the probabilities of the events in these sequences.

This type of analysis can be applied to complex

Risks of the coal option

Today, in the opinion of author Rasmussen and many others, the United States has one of two alternatives for generating electric power in increasing amounts—either to use nuclear fuel or to use coal. No other technology can make a substantial near-term contribution to the electrical demand problem. And coal, too, has its severe risks. To the people who worry about the nuclear risk, Prof. Rasmussen argues that, although we have learned a lot about radiation in the last thirty years, we have learned far less about what sulphur dioxide and some of the other pollutants of that nature do to health, or to the environment, and we operate much closer to the limits of known effects of these pollutants than we do to those of radiation.

According to Prof. Rasmussen, more and more studies indicate that there are rather important health and environmental effects from the burning of coal. For example, a recent report by the U.S. Environmental Protection Agency, called "Briefing notes—a status report on sulfur oxides," has stated that the combustion of coal has adverse effects on human health, vegetation, materials, and visibility. Dated April 1974, the report goes on to explain that these effects are caused by acid sulfate aerosols (suspensions of fine solid or liquid particles in gas—in this case, in air) converted in the atmosphere from sulfur dioxide or trioxide. These two chemical compounds are produced during the burning of coal by the oxidation of sulfur contained in the coal. The aerosals are said to have a much smaller particle size than that of sulfate particles of natural origin—like those from soil or sea spray—and they are therefore able to penetrate much deeper into the lungs. In addition, the man-generated aerosols are acidic unlike their natural counterparts.

One of the major conclusions of the EPA status report is that a massive conversion of urban power plants to the burning of high-sulfur fuels (like coal), and the use of tall stacks with supplementary control systems in rural power plants, will most probably greatly increase sulfate concentrations and endanger public health. The report, however, points out that there are many uncertainties in our scientific information base. These are related to problems and information gaps that will probably remain for five to ten years.

Privately, an EPA scientist has expressed a somewhat more optimistic view about the risks from a massive burning of coal. This scientist maintains that as we learn more about sulfur dioxide, the technology to reduce its emission from coal-fired power plants will probably improve, and this may eventually help to reduce the risks associated with the burning of coal for power generation.

Conclusions that are similar to those stated in the EPA report, although somewhat more cautiously phrased, are included in a recent report by the National Academy of Sciences (NAS). Entitled "A Quality and Stationary Source Emission Control," and dated March 1975, this report, which was prepared for the U.S. Senate Committee on Public Works, states that steam electric generating plants that burn coal are major sources of sulfur oxide emissions, especially in the northeastern part of the U.S. (Over 50 percent of total man-created sulfur oxide emissions in the U.S. are produced through the combustion of coal in power plants, and in some regions the percentage is much higher.) The NAS report goes on to say that increased coal-fired electrical generation (about a 100-percent increase is predicted over the next decade) may lead to as much as a 40-percent increase in ambient sulfate levels in urban areas, if power plant emissions are not controlled.

However, emissions of sulfur dioxide from power plants that burn low-sulfur (less than 1 percent of sulfur) coal can be reduced by at least 90 percent, according to the NAS report, by available techniques called lime and limestone scrubbing, that have already been demonstrated on a commercial scale. A similar reduction has been achieved also in power plants that burn medium- or high-sulfur coal and successful operation of a lime scrubber in a power plant burning medium-sulfur, low-chlorine coal has been demonstrated, according to the NAS report, on a commercial scale. To resolve the question of commercial availability of lime scrubbing technology for all coals, indicates the NAS study, experience on commercial scale can and should be obtained for medium- and high-sulfur coals containing systems in which failures of components or combinations of components can lead to system failures. In the case of reactor accidents, it is well suited for estimating the probability of failure of the various emergency safety systems designed to cope with serious events. Generally, the probability of the initiating event (e.g., pipe break) cannot be obtained in this way and must be estimated from operating experience.

Thus, the probability of a core melt accident is determined from the probability of the initiating event and the probability of the failure of systems designed to cope with it. However, the probability of a particular consequence will also depend upon how much of the radioactive material escapes from the containment, what the prevailing weather conditions are, and what population density exists in the contaminated area. In Fig. 3 are included these various inputs to the consequence calculations. As an example, consider the worst accident identified in the study—one which had an estimated probability of occurrence of 10^{-9} per plant per year. This worst case comes about as a result of an initiating event followed by emergency system failures that produce the largest release of radioactivity. This release must occur under the most stable weather conditions with the wind blowing toward the highest population density. Let us quantify this accident for the case of a small pipe break. Experience indicates that a small pipe break has a probability of 10^{-3} per plant per year. The fault tree analysis indicates that the failure of the emergency core cooling system needed to prevent core melt following such a break is about 10^{-2} per demand. The probability that the largest release will result is about 10^{-2} per core melt. Something near the worst weather occurs about 10 percent of the time (i.e., 10^{-1}) and the probability that the wind is blowing toward the highest population density is 10^{-2}. Thus, the overall probability of the small pipe break accident with the largest consequences is:

$$\begin{bmatrix}\text{Probability}\\ \text{of worst}\\ \text{consequence}\end{bmatrix} = \begin{bmatrix}\text{Probability}\\ \text{of initial}\\ \text{event}\end{bmatrix} \times \begin{bmatrix}\text{Probability}\\ \text{of safety}\\ \text{systems}\\ \text{failures}\end{bmatrix} \times \begin{bmatrix}\text{Probability}\\ \text{of containment}\\ \text{failure}\end{bmatrix} \times \begin{bmatrix}\text{Probability}\\ \text{of worst}\\ \text{weather}\end{bmatrix} \times \begin{bmatrix}\text{Probability}\\ \text{of popula-}\\ \text{tion density}\end{bmatrix}$$

$$[10^{-10}] = [10^{-3}] \times [10^{-2}] \times [10^{-2}] \times [10^{-1}] \times [10^{-2}]$$

chlorine beyond 0.04 percent.

The NAS report emphasizes that there has been a rapid advance in the understanding and the application of scrubbing technology, especially in the past year. There is a reasonable expectation, according to the NAS study, that in the near future scrubbers will be available for purchase as routine components of power systems, provided a vigorous development program is pursued. The NAS report, however, goes on to say that there are still many uncertainties in critical variables and relationships affecting ambient concentrations of sulfates, as there are uncertainties concerning the extent of harmful effects of such concentrations on health. Accordingly, the report emphasizes the need for more information, not only on the effects of suspended sulfates on health, but also on the chemistry of the atmospheric conversion of sulfur dioxide into sulfate.

While most studies deal with the impact of the burning of coal as separate from the impact of nuclear technology, a number of scientists have made comparative studies. Lave and Freeburg of Carnegie-Mellon University have produced a very detailed comparison of the public health effects of the two technologies.

When considering routine emissions from coal and nuclear power plants, conclude Lave and Freeburg, the health risk of power plants with pressurized water reactors (PWRs) is about 18 000 times less than that of coal-fired plants. Boiling water reactor (BWR) plants too seem to be less risky than coal, according to the Lave-Freeburg study, by a factor of over 20 (provided they fulfill certain conditions about their stacks). These figures are based, according to Prof. Lave, on data of the late 1960s, and it is understood that changes in the design of recent BWR plants may have already improved the second figure. In addition to lower risks in routine emissions from power plants, the nuclear power technology proves to be less risky in the mining stage than coal, according to Lave and Freeburg. The excess death rate in coal mining is at least ten times that of uranium mining, according to the authors. However, the authors go on to say that strip mining of coal involves much lower accident and chronic disability rates, and insofar as this technology becomes more important in the future, accident and disability rates will shift in favor of coal. (A NAS scientist claims, however, that improvements in deep mining technology have reduced the risks associated with it to about a factor of two only, higher than that of strip mining.) Although complete data are not available on accident and disability rates for other phases of the fuel cycles, the differences between coal and uranium are unlikely to be important, according to Lave and Freeburg, when compared with the estimated differences from mining and milling. The Lave/Freeburg analysis did not include accident risks in either coal or nuclear power plants, but Prof. Lave believes that even inclusion of such data cannot change the results appreciably in favor of coal.

The need for a quantitative environmental comparison of coal and nuclear power generation, and the associated fuel cycles, led to a workshop on that subject sponsored by the National Science Foundation and held on May 27–28 at the MITRE Corporation, McLean, Va. Over 100 attendants from many disciplines participated in the workshop and typical studies presented in that workshop were by people from the Stanford Research Institute, Brookhaven Laboratories, the National Oceanic and Atmospheric Administration, Princeton University, Battelle Northwest Laboratories, Massachusetts Institute of Technology, and the Energy Research and Development Administration. Although no specific conclusions were reached, the multidiscipline discussions following the presentations were found useful by many of the participants of the workshop. A summary of the meeting as well as its entire proceedings will be available from MITRE by the end of this summer.

Considering coal as an alternative to nuclear fuel for electrical power generation, Prof. Rasmussen, believes, in agreement with studies like that by Lave and Freeburg, that the burning of coal is much more damaging than the use of nuclear energy for the generation of electrical power, both environmentally and healthwise; nevertheless, he maintains that, in the U.S., a lot of coal will have to be burned, even with the maximum use of nuclear power, to meet U.S. electricity demands.—*Gadi Kaplan, Associate Editor*

When all other ways of producing this accident are considered the probability becomes a factor of 10 larger. Uncertainties in the data were propagated through the calculations and were estimated to be about ±5 on probability and ±3 on consequence magnitude.

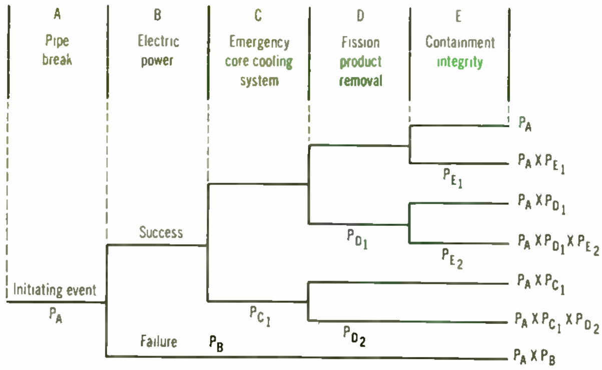
Probability estimation

There are three things about such a calculation of probability that can be questioned (and have been by various commentators). They are: (1) Are the probability values used for the various factors correct? (2) Are the probabilities independent as implied by the multiplication? (3) In obtaining the total probability of such events have all possible accidents been considered? Let us now consider these questions, each of which is relevant to our discussion.

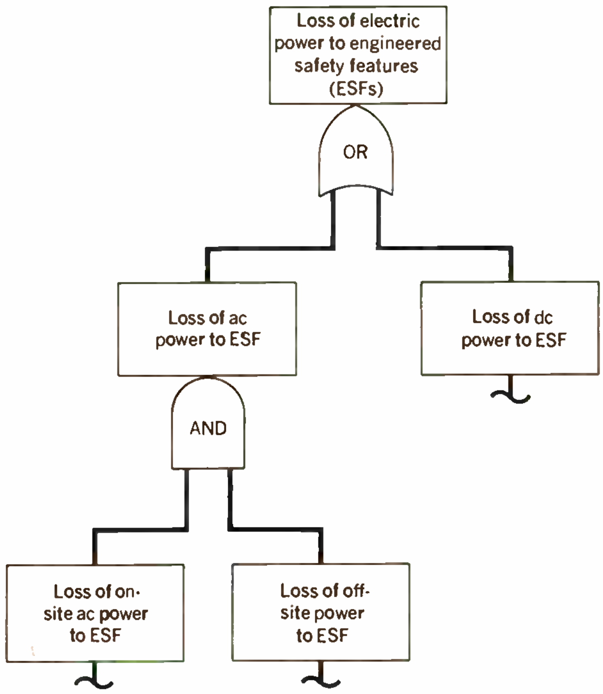
The probability of the initiating event is in almost all cases based on data from operating experience. However, because very few such events have occurred the statistical accuracy of many of the initiating events considered is poor. In the case of pipe breaks, the amount of uncertainty was believed to be plus or minus a factor of 10.

In the case of emergency system failure, the probability of failure was obtained from fault tree analysis mentioned earlier. About 20 such systems were analyzed and they were found to have failure rates on demand of 10^{-2} to 10^{-5}. The emergency shutdown system which is mainly electronic and highly redundant was in the 10^{-4} to 10^{-5} range. The mechanical systems of pumps, pipes, and valves, used to deliver emergency water, were mostly in the 10^{-2} to 10^{-3} range. These analyses consider three contributors to the failure probability—the failure of various individual pieces of hardware, failures due to human interaction including operator error, and maintenance errors and unavailability due to testing procedures that are routinely required. In addition, contribution due to dependencies between failures (i.e., a failure that causes a second failure), called "common mode" failures, had to be considered. It is interesting to note that these latter two factors—that is, human interaction and common mode failures—often made important contributions to total system failure rate. It was, in fact, the omission of such factors in some early papers in this field that led to unrealistically small predictions for the failure rate. Unfortunately, some of the skepticism about these techniques is a result of some of this early work in the field.

A reasonable question to ask is whether or not such techniques have been able to predict system failure rates reasonably accurately. Unfortunately, most of this work done in the United States has been on the analysis of weapons and space systems where there is either very limited experience to check the results or they are classified secret for security reasons. However, this is not true of much of the work that has been done by the Systems Reliability Service of the United Kingdom, where these techniques have been used to analyze commercial systems for well over a decade. The British have examined safety systems in the chemical, nuclear, and conventional power industries. For example, in some 50 systems in the U.K., whose failure rate was predicted prior to service, the observed failure rate was within a factor of



[1] To identify significant accident sequences in nuclear power plants, an "event tree" approach, shown here in a simplified version, was employed in the reactor safety study report known as the WASH-1400 draft. Basically, an event tree is a logic method for identifying the various possible outcomes of a given event, called the "initiating event." Since the "failure" probability *(P)* associated with each event is almost always 0.1 or less, the corresponding "success" probability (1 − *P*), is approximated to be 1.0. Using this approximation, event sequence probabilities are indicated on the right-hand side.



[2] To estimate probabilities of failure of a system, "fault tree" methods have been used in most instances in the reactor safety study. In the portrayed first few steps of a fault tree development, a failure of ac power to the engineered safety features (ESF), or that of dc power to the control systems of the same ESF, will result in total ESF failure.

±4 of the predicted value in 96 percent of the cases. Contrary to some peoples' intuition, the observed failure rate was, on the average, about 25 percent smaller than the predicted rate. This, of course, is a result of the common practice of using conservative values when there are uncertainties in the analysis. As the British experience shows, when such analyses are done carefully by experienced people, these methods can produce quite good results.

Although there's not yet enough nuclear experience to verify that the WASH-1400 predictions have

Criticism of the WASH-1400 draft

Comments on methods of analysis employed in WASH-1400 are included, among other observations, in a report, issued in April by the American Physical Society, on a one-year study on technical aspects of the safety of light water nuclear power reactors. Although resources were not available to the APS study group to carry out an independent evaluation of the quantification aspects of WASH-1400, the group maintains that the event tree and fault tree approach employed in that study (see main article) can have merit in highlighting *relative* strengths and weaknesses of reactor systems, particularly through *comparisons of different sequences* of reactor behavior. However, the APS report goes on to say, based on the members' experience with problems of a similar nature to those analyzed in WASH-1400—ones involving very low probabilities—that the APS study group does not have confidence in the presently calculated *absolute* values of the probabilities of the various event-tree "branches" in models employed in WASH-1400. Also, the group that conducted the APS study estimates substantially larger long-term consequences than those computed in the draft to WASH-1400.

For example, whole body dose cancers and genetic defects were calculated in the APS study to be about 10 000 each following a "reference accident" —that is, an accident with a frequency somewhere between 1 in 20 000 and 1 in 2 000 000 reactor years. In the WASH-1400 draft, the corresponding values were 310 each—that is, lower by a factor of over 30. This is mainly attributed to the neglect in WASH-1400 of the long-term ground dose which is dominated by cesium-137. The APS study took into account a dose over a period of a decade or so from cesium-137, whereas one-day activity by this contaminant only following an accident was included in the computation in WASH-1400. Also, the APS study computed up to about 5600 cancer deaths from lung and thyroid doses. In addition to the 310 cancer deaths that were computed in WASH-1400, where whole body doses only were taken into account. (The lungs get extra beta-ray exposure from inhaled radioactive aerosols, and an extra dose from radioiodines would be concentrated in the thyroid.) The APS report also estimates between 22 500 to 300 000 nonfatal thyroid nodule cases following an average reference accident. No explicit estimate on thyroid nodules for this type of an accident is given in the WASH-1400 draft, according to the APS study. But the APS study group estimated that, based on the assumptions in WASH-1400, that study would have come up with a figure of about 25 000 thyroid nodules for an average reference accident. It is understood, however, that revisions in the long-term effects will be included in the final WASH-1400 document to be published in the fall.

Another report critical of WASH-1400 was issued earlier, in December 1974, as a joint effort by the Sierra Club and the Union of Concerned Scientists (UCS). Prepared by a group of ten scientists and engineers, the report provides a preliminary evaluation of WASH-1400. Although admittedly incomplete and tentative in its conclusions due to the size and complexity of WASH-1400, this report was nevertheless believed by its authors to have potentially useful comments on urgent safety problems that exist in nuclear power reactors.

What is the criticism of WASH-1400 in the Sierra Club/UCS report? First of all, the study group regards it as unlikely, if not impossible, that WASH-1400 was able to identify every important accident sequence—this in view of the complexity of the nuclear technology. The group also claims that WASH-1400's analysis methodology does not judge the design adequacy of equipment and must largely take it on faith. This can introduce serious errors in the event that design mistakes go unnoticed, as they occasionally do, says the Sierra Club/UCS report. The report goes on to criticize WASH-1400's use of fault-tree analysis. While this method of analysis can reflect the dependency of one component failure on another if the dependency is known, in the vast majority of cases studied in WASH-1400, components are assumed to fail independently, according to the Sierra Club/UCS report. Subtle but crucially important dependencies can arise from physical proximity, unexpected component response to the abnormal circumstance of an accident, or design error, says the critical report, and experiences with reactor accidents show this to be an important consideration as such dependencies have badly aggravated otherwise innocuous events.

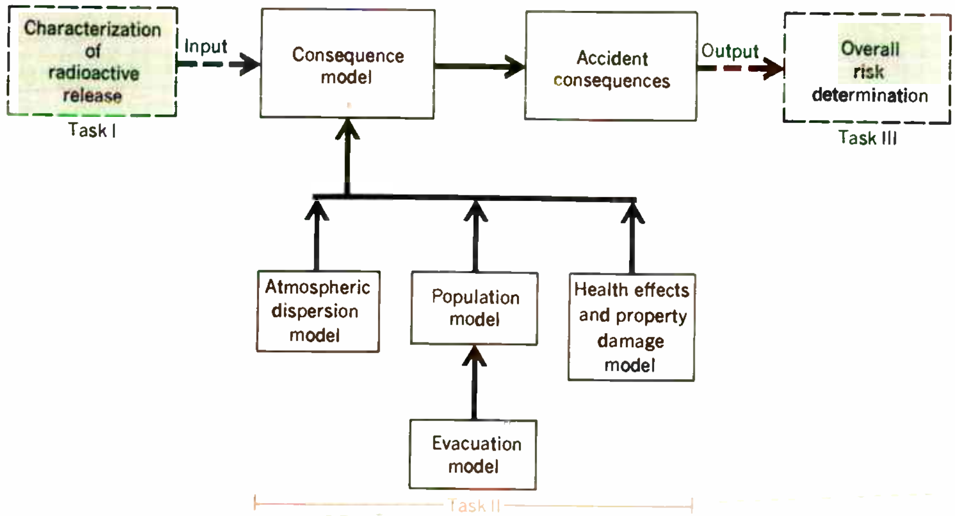
Another drawback of the fault-tree method used in WASH-1400 is, according to the Sierra Club/UCS report, that human error is the least tractable source of error in such analysis. Further, possible errors may have been introduced by a lack of data on structural failures. Again, WASH-1400 neglected secondary failure, except in restricted cases (secondary failure is defined as failure of a component because it is stressed beyond its specified range of performance). Another Sierra Club/UCS criticism suggests that WASH-1400 wrongly assessed risks from common-mode failures (in such events, redundant and presumably independent systems are nevertheless simultaneously disabled through some common cause). In addition, the Sierra Club/UCS group concludes that a pressure vessel rupture represents a very real risk—an assumption rejected in WASH-1400. The Sierra Club/UCS report also claims, in this regard, that the use of failure rate data in WASH-1400 is suspect.

In conclusion, the Sierra Club/UCS study group says it does not believe that the event-tree/fault-tree methods can be employed to determine *absolute* probability values for accidents in order to use these values as proof of the safety of nuclear plants. Such methods, if properly used, can do no more than assist in making comparisons between diverse system designs, in assessing *relative* improvements from system component changes, or in identifying design weak points.

In addition, the Sierra Club/UCS group has concluded that the aggregate consequences to human health of major accidents evaluated in WASH-1400 are *seriously understated* (up to about a factor of 16 in fatalities and acute illnesses, according to the group's "conservative" estimate). And finally, the group concludes that there are serious implications concerning the nuclear program to be drawn from WASH-1400 results that that study fails to acknowledge. For example, the concept of "floating" nuclear power plants becomes highly suspect, claims the Sierra Club/UCS group, in light of the finding in the WASH-1400 draft of a 1 to 17 000 probability of reactor core melt. Also, the consequences and risk from sabotage are seriously understated in WASH-1400.

The UCS even goes one step further, saying that it (the UCS) has concluded that on the basis of all of the criticism of the draft of WASH-1400, and the magnitude of errors in this study, WASH-1400 is discredited and its evaluation of risks from commercial nuclear power plants cannot be relied on.—*G.K.*

[3] The consequences of a given radioactive release depend upon how radioactivity is dispersed in the environment, the number of people and amount of property exposed, and the effects of radiation exposure on people and contamination of property. The principal subtasks involved in analyzing these factors in the reactor safety study are shown.



achieved as good results as those of the British, at least the failure probabilities obtained (i.e., 10^{-2} to 10^{-5} for various systems) are quite consistent with what the British experience has been with typical safety systems in various types of commercial service. The probabilities of containment failure mode, included in the WASH-1400 draft, are also based principally on fault tree analysis.

The probability of the weather condition comes from weather bureau records and the probability of the population density, from U.S. Census Bureau records. The possible uncertainty in these values is therefore quite small.

Independence of probabilities

A simple multiplication of the various probabilities in the above mentioned formula for computing the probability of worst consequences assumes that the probabilities are independent. Suppose, for example, that when the pipe broke, the steam released also broke the containment and caused the emergency system to fail. Clearly, it would not be correct to assume these events are independent. Of course, just such things must be considered in reactor design. Thus, every attempt is made to ensure that the containment is strong enough to contain the pressure of the steam and that the emergency cooling system will operate in the presence of the steam and water that might be released. Nevertheless, since all these systems are part of one plant, in one place, all sources of possible common mode failures cannot be eliminated. Therefore, an important part of the Reactor Safety Study was to search for such dependencies by a careful analysis of the plant and to modify the probabilities wherever such dependencies were identified.

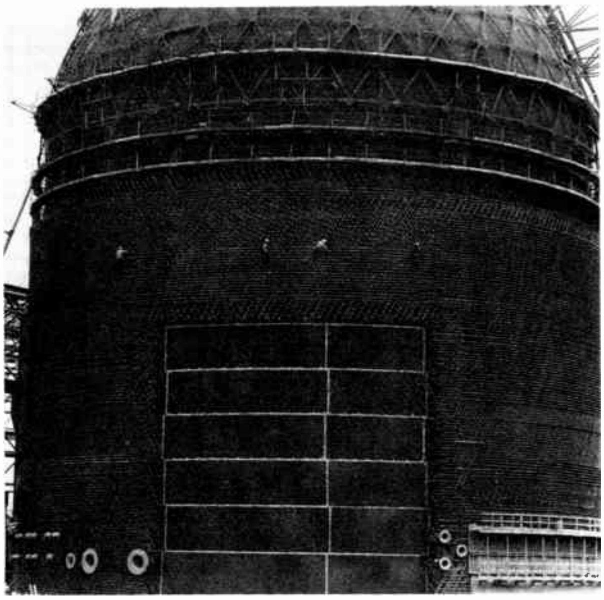
All potential accidents considered?

Such an analysis can only consider those accidents that can be imagined by those doing the analysis. It seems clear that not all the possible ways of failure have yet been observed, so almost surely there are possible accidents that have not been considered. At first glance, this might seem like a limitation that would negate the usefulness of these techniques. However, it must be remembered that the goal of the analysis is to make a reasonably accurate estimate of the failure probability. To do this, all that is required is that all the dominant modes of failure be considered because very rare or unusual failure modes will surely change the overall failure probability very little. Because there has been considerable experience with systems of pipes, pumps, and valves—such as are used in nuclear plants—it is very unlikely that the most likely modes of failure have been overlooked.

What gets irradiated?

In calculating the failure rate of components, we studied the effects of equipment amortization and depreciation due to extreme temperature, pressure, and radiation environment. We think these conditions will not change the probability by more than a factor of two, whereas a factor of ten was included in our probabilities. The usual temperatures and pressures are quite within the range of past experience. The only thing that is different is radiation. Major parts of nuclear power plants are not in a radiation field—in fact, none of the operating equipment is. The only thing that gets irradiated in the nuclear plant is the vessel and the fuel. But the pumps and the pipes and the valves do not receive any significant amount of radiation. After all, people have to go in and work. They are in low-level fields primarily, and these fields are not strong enough to do any material damage. The vessel itself has been carefully studied and the effects of radiation on it are known. About 70 pieces of steel, from the same batch of material the vessel is made of, are put in the vessel before it is put into service, and they are taken out periodically and examined for any deterioration of the vessel.

This brings us to the question of the possibility of risk after the expected lifetime of a nuclear plant (30 to 40 years). There is no such possibility because the fuel is removed from the vessel. However, some of the steel inside the vessel is still radioactive, principally with cobalt-60, with about a five-year half-life. One of the things that could be done at this point is to seal the reactor building up and let the components sit for a while. But each application per plant must include a plan for decommissioning, and every utility that operates nuclear power plants must set aside money in the later years to accumulate a fund for this task. We



Failure modes of a containment system for a nuclear power plant, like this one under construction (in 1971) for the Diablo Canyon project of Pacific Gas and Electric Company in San Luis Obispo County, Calif., were also analyzed among other studies of failure, included in the Reactor Safety Study. (Courtesy Atomic Industrial Forum)

have already decommissioned small reactors, and it is not an impossible task.

Results valid for only first 100 plants

One question frequently asked in discussions of the models and methods employed in the analysis of accident risks in light water reactors (LWRs), of which WASH-1400 is comprised, is: How valid is this analysis for the next generation of reactors, namely liquid metal fast breeder reactors or high-temperature gas-cooled reactors (see *Spectrum*, February 1974, pp. 85–89, and June 1975, pp. 46–55)? The answer is that for the breeder, which is a very different concept and has very different kinds of accidents that could lead to releases than those postulated for LWRs, the part of the study that calculated the probability of release will not be applicable at all, and will have to be redone. Once the release is computed, however, the general model that we have developed could be used. In the report, this was not done because it was felt that there will probably not be a substantial number of breeder reactors in operation in the United States before the year 2000. And it is impossible to predict the risk of plants whose design is as yet unknown.

Further, while the study should be applicable to the first 100 plants in the U.S. (we analyzed the 24th and the 34th plant of the first 100 reactors as samples of that group), the next 100 will probably meet different design criteria and will therefore require modifications in the techniques employed in the study.

Management and safety

A question is sometimes asked whether bad management of a utility during construction and operation of a plant can substantially increase risks in the local area of that particular plant. While this question is not taken up in WASH-1400, it is my feeling that safety in part depends upon those who operate a plant, and that a safety philosophy in the management contributes to the overall safety of the plant. The nuclear industry tries to make all plants meet the minimum standards of Federal codes and regulations, so that no plant can be operated in a careless way. It is certainly not to a utility's advantage to have an accident, because if its own losses are not covered by its insurance, it will cost it severely in repairs. Nevertheless, that does not mean that, occasionally, someone won't be tempted to take a shortcut to save some money. But in my judgment, the NRC's inspection procedures are quite good. A further concern is the size of utilities undertaking nuclear ventures. This is handled by the requirement that the

Radiation doses, in perspective

While opponents to nuclear power express concerns about radiation risks to the general public from power generating nuclear reactors, even under normal operation, Prof. Rasmussen has pointed out that the general population receives radiation doses at a fairly high level as a result of such common occurrences as automobile accidents. According to Prof. Rasmussen, about two million people are seriously injured each year in automobile accidents in the U.S., and, by and large, most of them are taken to emergency rooms and are thoroughly X-rayed. (Current medical practice in accident cases requires extensive X-raying of any patient so that a record will exist that can be used in the event of malpractice suits.) Thus, of these two million people, many receive doses in the range of 0.5–1.0 rem (roentgen-equivalent-man, a unit of dose equivalency). And according to an NRC radiation standard, a person living adjacent to a nuclear plant may not receive more than 5 milirems per year.

To put this in perspective, Prof. Rasmussen points out that, in a serious accident in a nuclear plant, the total man-rem—that is, number of people times their doses—can reach tens of millions. The annual dosages due to automobile accidents in one year total about 10 percent of that. Thus, in a few decades—far less than the probable period of a single nuclear reactor accident—auto-accident-related dosages will far exceed those from a serious and less likely reactor accident.

In addition to this kind of comparison, Prof. Rasmussen notes that we don't yet know what genetic damage chemical elements in the environment are causing, but that it is known that many of them do have genetic effects.

What is a "safe" radiation dose level? Prof. Rasmussen says that there seems to be no conclusive answer to this because the more we learn about radiation, the more we know about the consequences. However, NRC's 5-milirem low limit for a person who lives next to a plant's boundary is well below what that person normally receives from his food and his natural surroundings (about 100 mrem), and the average yearly X-ray dose for people in the U.S. is between 50 and 70 mrem (10 to 15 times the NRC's low limit). A well known report by the National Academy of Sciences on biological effects of ionizing radiation (the BEIR report) has shown that in 1970, on the average, the estimated annual whole body dose rate from nuclear power in the U.S. was about 0.003 mrem/year whereas the natural dose rate amounted to about 102 mrem/year. Medical dose rate for diagnostic purposes was estimated to be about 72 mrem/year in that year. (Being average figures only, these numbers can be much higher in specific segments of the population.)—*G.K.*

Energy balance of nuclear power

How much energy is invested in the construction and the running of a nuclear power plant, and how much energy is available from the plant during its entire service? Questions of this nature have recently been asked, and the answers to these questions vary widely. In an article in *New Scientist*, Dec. 19, 1974, Peter Chapman, director of the Energy Research Group and a senior lecturer in the science faculty at the Open University, claims that a nuclear program in which the number of reactors are steadily increasing could find itself investing more energy each year in the construction of new reactors than is available as output from those already constructed.

A totally different assessment was made by Kenneth Davis, vice president of Bechtel. At an atomic Industrial Forum Conference on the topic of "Accelerating Nuclear Power Plant Construction," held in March in New Orleans, La., Mr. Davis showed that under most pessimistic assumptions, an 1100-MW (electric) nuclear power plant with gross weight of some 580 000 tons pays off its total energy investment in five months of actual initial plant operation, and maybe even less than that. And the energy invested includes that required to make the material and equipment for the plant as well as allowances for all associated fuel cycle facilities. Also included is that energy spent on mining, milling, enrichment, and reprocessing of fuel both for the initial fuel loading and for replacement loadings.

Another recent energy balance analysis for nuclear power plants shows that a single 1000-MW (electric) generating power plant returns over its expected 30-year lifetime more than ten times the capital energy investment, based on the assumption that equal thermal and electric energy units have an identical societal value. However, according to Frank von Hippel, Margaret Fels, and Hartmut Krugmann, authors of an article in the April 1975 issue of the professional bulletin of the Federation of American Scientists, the entire electric utility industry has been built on the premise that, for many purposes, one unit of electrical energy is more valuable than three equal units of thermal energy. According to this view, the net energy from a single power plant is more than three times better than the value stated.

But what about the net energy of an exponentially growing nuclear industry? According to the three authors, a plant currently producing 4 billion kWh (electric) each year can be paired with three plants under construction that are currently consuming together 1.35–2.40 billion kWh (thermal) in construction and initial core loading. The authors indicate that if the higher values of their estimate is correct, then indeed the rate of energy investment in the nuclear sector is significant in comparison to the output (considering equal values to equal electric and thermal units). However, the authors point out that, even under this assumption, the nuclear power industry would still be producing more energy than it is absorbing. And with the current growth rate, indicate the three authors, the nuclear program in the U.S. does not seem to drain significant amounts of energy from the rest of the economy.—*G.K.*

NRC approve a statement of financial responsibility before a license is issued to a utility. No "fly-by-night" outfit can build and operate a nuclear plant.

Nuclear wastes

Although the problem of dealing with nuclear wastes was not treated in WASH-1400, there are some general comments that should be made about this generally misunderstood aspect of nuclear power.

As far as storage of contaminated parts is concerned these parts are comprised of two kinds: low-level types, such as ion exchange resins used to clean the water in nuclear power plants, and high-level types in the form of the spent fuel itself. The radioactivity remains in the fuel after its use. Probably the single fact that makes this a manageable problem is the relatively small amount of the nuclear waste. A 1000-MWe plant will produce only about 2 tonnes of highly radioactive waste a year. (For comparison, a similar size coal station generates yearly about 250 000 tonnes of ash.) Currently, however, the U.S. does not have an operating reprocessing plant for spent fuel from commercial nuclear power plants. (One plant, Nuclear Fuel Services of West Valley, N.Y., did operate till 1971, when its operations were suspended for modifications.) Consequently, although no waste is now being generated by reprocessing plants, spent fuel is accumulating in the storage pools on the plants' sites, and, in a few plants, this is becoming a problem. It was hoped that another reprocessing plant, Allied-General Nuclear Services, in Barnwell, S.C., might be on line about 1976 or 1977, beginning then to reduce this backlog of fuel. But a recent decision regarding plutonium recycle may affect the schedule of operation of that and other reprocessing plants in the U.S.

Current Federal regulations require that the high-level radioactive waste be converted into an insoluble solid within five years after it is produced. Within ten years, this solid waste must be sent to a Federal facility for temporary storage until a permanent disposal method has been established. To date, the Federal government has not announced what permanent disposal method will be used, but there are at least two methods that have been investigated that appear acceptable. These are: to bury the material in deep beds of salt or to bury them in deep granite formations. Most of the radioactivity of the high-level wastes has a 30-year half life so almost all the radioactivity would be gone in a few centuries. However, there is a small component with half lives in the range of 10 000 years or more. With processing techniques now available the amount of such activity can be kept low enough so that the level of these long-lived alpha emitters would not be much greater than that of a rich uranium ore deposit.

Transportation of nuclear material

It is inevitable that as the nuclear industry grows an occasional shipment of radioactive material will be involved in a transportation accident. I believe the risks associated with these accidents will be quite small for several reasons. First, the regulations on shipping containers are very strict. These containers are required to withstand environments such as might be expected in such accidents (i.e., fire, impact, etc.) Second, the material is shipped in solid form and must have decayed to a point where it can no longer

Safety spending—how much is too much?

In the opinion of author Rasmussen, actions have been taken in the U.S. in the name of safety that were not carefully thought out and did not really help safety, but contributed to inflation very seriously. In addition, maintains Prof. Rasmussen, some of these actions lack any statistical analysis to show that they really contribute to safety. Asks Prof. Rasmussen, if a railing is made four feet high instead of three feet high, does it really reduce the risk to people or does it simply make the product cost more?

The Nuclear Regulatory Commission, according to the Professor, has tried to do a good job in improving safety in nuclear power generation, but there have been some areas in which the NRC has required levels of safety that are unwarranted. And in other areas the NRC safety requirements are just about reaching the right level. So it is a question of balance, according to Prof. Rasmussen, and it makes sense to say, on a very cold cash basis: If one can reduce the possibility of this accident, it is worth these many dollars to society and therefore represents a good tradeoff. But Prof. Rasmussen claims that there has been an over-application of safety standards in some areas. The whole safety philosophy is a good one, he says, only its implementation is imperfect. We ought to spend dollars to make nuclear power reactors safe, but this spending is surely out of balance with what we are spending on saving a life on the highway.

A different attitude to the question of spending on safety, as related to light water reactors (LWRs) (see *Spectrum*, June, pp. 46–55) has been recently expressed by the American Physical Society. In a report (released April 28, 1975) on a one-year-long study sponsored by that society, and supported by the NRC and the National Science Foundation, APS expresses its belief that additional investments in safety research on LWRs are warranted by many technological opportunities for the enhancement of such safety. Current spending on reactor safety research, says the APS report, is relatively small, considering the large present and future capital investment in LWRs, and in view of the great social importance of reactor safety.—*G.K.*

generate enough heat to melt itself. Thus, even if the container were to rupture, any contamination should be quite localized and readily cleaned up. Third, the routes for shipment are specified to avoid major centers of large populations. In general, I believe the shipment of these materials under current regulations is somewhat less hazardous than current commercial shipments of highly flammable liquids.

Sabotage and safeguarding

Concerns have recently been voiced about the possible risks to the public by sabotage to nuclear power plants. In WASH-1400, we mentioned that we looked at possibilities of sabotage, not in a detailed way, but enough to convince ourselves that the saboteur could not produce accidents bigger than those we had looked at. In other words, there is no disaster a saboteur could create that would not fit into our class of accidents somewhere. In addition, it would be very difficult for him to create the worst kind of accident, because that requires the worst kind of release, and that involves some conditions a saboteur cannot control. For example, such a disaster would have to occur at the worst time, weather-wise—that is, when the wind was blowing toward the highest population center. And although a very calculating saboteur might wait until such conditions exist, this certainly makes it much more difficult.

What we further said in WASH-1400 was that we know of no way to calculate the probability of a sabotage attempt, and of its success, because it depends on a lot of factors—how many people, how well organized, and how likely it is for such organizations to form with this intent. And history has shown that various kinds of sabotage events have had wildly fluctuating likelihoods. Such things as kidnapping of business executives, in vogue in South America, or hijacking of airplanes, which now seems less fashionable than it was, are cases in point.

But the sabotage topic has many aspects, besides those examined in WASH-1400 and in other studies that we have reviewed to some extent. We have been examining the kinds of things a terrorist might do to cause social disruption. If one lists those according to their difficulty of achievement, one finds that, in most people's opinion, reactors are way down on the list, due to the difficulty and the uncertainty of achieving any specific goal, as compared to some easier, more certain ways of causing disruption. Obviously, it is not in the public interest to come out and say: "Here are ten easier ways to cause trouble." But basically what it boils down to is that if groups are going to try to disrupt society, the nuclear plant is just one of many possibilities, and is therefore one more place that society must defend. I do not believe this provides a reason for stopping nuclear power; it is, however, a very good reason for making nuclear power plants as secure as possible.

Basically, the same argument applies to public concerns about the theft of nuclear fuel by terrorist groups. Once again, I think it is absolutely mandatory that any government that deals in plutonium should supply appropriate security to defend itself against such clandestine operations. On the other hand, if the whole commercial nuclear power program were to be abandoned, would this risk be decreased? Probably not much, since everybody knows that there are tens of thousands of plutonium explosives already built, fabricated, and distributed about the world. In my opinion, any group capable of stealing plutonium and constructing a weapon is also probably capable of organizing an élite force capable of attempting to hijack a nuclear weapon. The little I know of the security of nuclear explosives makes me believe that the likelihood of success in such an attempt is probably very small. The perfect record over the last 30 years indicates that this is true. As in the case of sabotage, the potential risks are significant enough so that very strict security procedures must be maintained and improved as the amounts of plutonium increase. However, I believe this is a manageable problem so that this should not be considered as a reason for stopping nuclear power.

Nuclear insurance

Closely related to the question of nuclear risks is the insurance topic. Last year, the U.S. Congress passed a bill that was to phase out Government participation in insurance gradually and put the responsibility onto the utilities who operate plants. The bill

Nuclear failure rate data upgraded

Substantial improvement over the quality of data available at the time the WASH-1400 report was under development is expected when a new failure rate data manual for nuclear power plants, currently under development, is completed. With its publication, anticipated for spring 1976, the Manual—developed as IEEE Standards Project 500 by a working group formed in the Nuclear Power Engineering Committee of the IEEE Power Engineering Society—is expected to serve as a standard source for nuclear failure rate data. The manual will cover all electrical, electronic, and sensing components in nuclear power plants. Containing data on more than 1400 generic components, it is expected to be the authoritative source of data for the nuclear power industry.—*G.K.*

was vetoed by the President and, according to present legislation (the Price-Anderson Act), the U.S. Government supplements the first 125 million dollars of liability insurance coverage, available to utilities by commercial insurance companies, by bringing the total coverage up to \$560 million. (The premiums paid by the utilities for this Government coverage are about \$100 000 a year per plant.) The Reactor Safety Study estimates that the probability of exceeding the \$560 million damage to parties other than the operating utility itself is less than 10^{-6} per year per plant.

However, the vetoed legislation is worthy of consideration. Its main thrust was that each owner of a nuclear power plant could be liable for somewhere between \$2 and \$5 million per plant, in helping to pay losses that occurred in any accident in any plant in the U.S. What that would have meant was that if a plant had an accident and there were \$200 million dollars worth of claims, the commercial insurance would have paid \$125 million, then each reactor around the country would have been assessed a fraction of the difference and would have had to pay that. The arrangement was to be like that of an insurance pool, and yet different in a number of ways. It would not have been necessary to set aside \$2 million somewhere for this purpose. It would have merely been an obligation, should a case come up, and it would not have depleted working capital the way an insurance pool does. In the latter case, you really put the money in and the pool holds it, and if the experience is good, some years later the pool pays you back. Meanwhile, it ties up capital. According to this arrangement, once there were 100 plants, if each were liable for \$5 million, that would have come to \$500 million—of coverage in excess of the commercial coverage. This liability limit could later increase as more plants were built, and could reach several billions of dollars, depending on how the limits were set.

Unfortunately, as has been said, the insurance bill was vetoed by the President—on a technicality. There was a rider on the bill that said that once WASH-1400 was completed, Congress might reevaluate, and even fundamentally change the bill. The President felt he could not sign into law any bill the Congress could change at a later date. Thus, this insurance issue will have to be considered again by the present Congress.

It is interesting to note that the commercial insurance industry has recently reduced the insurance rates of nuclear power stations. Among the reasons given were the excellent accident record and also the results of the Reactor Safety Study.

The impact of WASH-1400

Since the draft report was issued, it has been the focus of considerable discussion. Although I believe it was a good first step in applying the aforementioned methods to risk assessment, clearly it can be improved considerably. Because of the high level of interest in this topic, I have no doubt that such improvements will be made in the years ahead. In general, both critics and supporters of the study felt that this detailed analysis of complex technology was a worthwhile effort and already requests are being made for similar studies on other aspects of the nuclear industry and also for nonnuclear industries. In general, I believe this is good, although it is not clear that all technological risks can be assessed in this way. For example, the risks associated with the burning of coal are mainly due to normal releases, not accidents. These techniques are therefore not particularly valuable since in the case of coal-burning the problem is basically one of better understanding the health effects of certain pollutants.

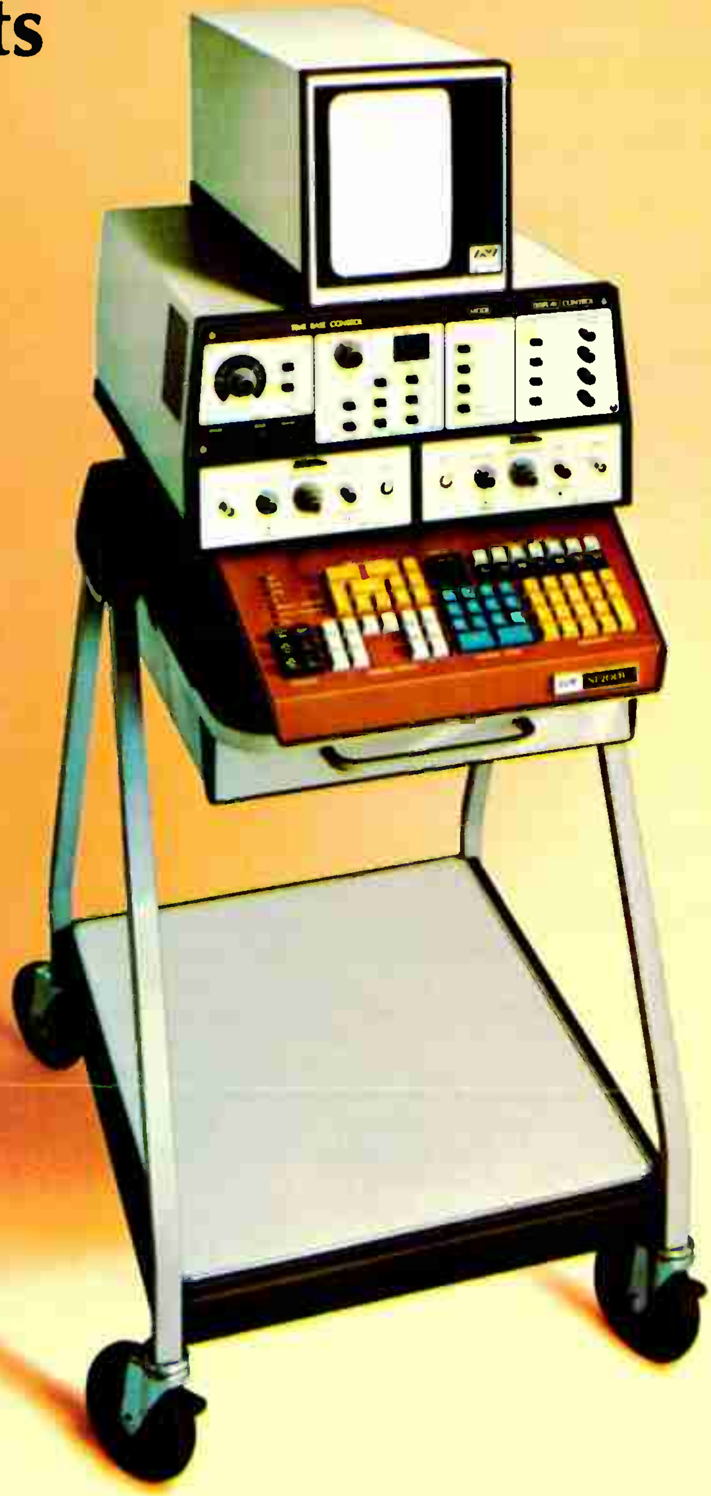
Even though studies such as this one cannot be expected to end all controversy, nevertheless, they can help focus the issues and provide a more logical basis for discussion of those issues. Even skeptics of the quantitative aspects of such analyses see their benefit for comparing various aspects of the risks. Eventually, as more such studies are done and experience shows them capable of reasonable results, I am sure that they will be accepted as a technique for quantifying risks. However, it took more than a decade for these procedures to become accepted in the United Kingdom. Hopefully, the process will be somewhat shorter in the United States.

A final word: Nuclear power, like all power generation known today, brings with it certain risks and problems. However, when these are carefully considered they appear relatively small compared to the benefits that can be obtained. Numerous studies have shown that the effect on public health is less for nuclear plants than today's coal plants. I believe that the energy problem facing the United States today is such that, at least for the rest of this century, nuclear power will have to be an essential part of the solution to the problem. ◆

Norman C. Rasmussen, head of the Department of Nuclear Engineering at the Massachusetts Institute of Technology, was recently appointed by the U.S. Department of Defense to head a panel of high-level civilian scientists to "provide guidance on the allocation of technology base resources and identification of technology targets of opportunity." Since 1972, he has served as a consultant to the Atomic Energy Commission (now the Nuclear Regulatory Commission) and, in that capacity, he directed the Reactor Safety Study described in this article. Dr. Rasmussen received the Ph.D. in Physics from M.I.T. and is the author of over 50 papers.

Meet the all-purpose instrument. It's a digital oscilloscope. It's programmable. It calculates. It utilizes all the power of a microprocessor. It's Norland Instruments

NI 2001



All the function and instrumentation capability you'll probably need is yours in this single mainframe. The first Programmable Calculating Oscilloscope is available now — only from Norland.

Here is the ultimate instrument for the acquisition, processing and manipulation of electrical data. It completely eliminates the need to compromise your requirements with a jumbled array of separate instruments. The NI 2001 combines all the capabilities of a digital oscilloscope and a microprocessor in a single mainframe. It brings you flexibility, convenience, accuracy and reliability you won't find anywhere else.

This — the first Programmable Calculating Oscilloscope — is the product of many years of technological research and instrumentation engineering. With the advent of microprocessor technology, Norland Instruments engineers were quick to recognize the power available to the instrument designer and first applied a microprocessor to the Norland line of medical instruments. That experience ultimately led to the use of microprocessor technology in the development of the NI 2001 — the first truly new generation of test and measurement instrumentation for the industrial and scientific user.

Air safety as seen from the tower

A veteran of more than 33 years in air traffic control says hazards lurk in proposed FAA procedures

In 1975, air travel is still a pretty safe way of quickly getting from one place to another; but there are ominous influences at work that are chipping away at the safety factor. Even in the vaunted air traffic control (ATC) system there are conditions in ferment that presage an erosion of the safety record to which the aviation community has always pointed with pride.

In short, the ATC system in the U.S., like Dr. Frankenstein's creation, may turn out to be a monster that will do in not only its creators, but also the entire flying public—simply because that ATC system now contains the potential for increasing aircraft accidents. Does this sound shocking? I hope so. It is such an alarming situation that many of the people most closely concerned with the regulation, enforcement, and promotion of air safety *have not yet permitted themselves to believe it!* And some of those who are aware are trying to ignore the problem in the hope that it will just conveniently self-destruct.

This is why the implementation of a computerized airborne collision-avoidance system (CAS)—pilot-monitored, aircraft-contained, and independent of ground control—must be achieved as soon as possible ... and much faster than the Federal Aviation Administration's (FAA's) present plans. The continued reliance upon a ground-based controller-monitored system (another generation layered onto the already lopsided, top-heavy first-generation structure) is leading inevitably to reduced safety and efficiency, and less orderly regulation in the movement of air traffic in the United States.

John K. King Consultant

Factors against ground-controlled ATC

Why won't ground-based controller-monitored systems work without being a threat to air safety? Very simple: they were originally intended to separate airline flights during poor-visibility weather, in which instrument flight rules (IFR) governed. Twenty-five to 30 years ago, there weren't too many such flights. Then, the controllers and pilots knew and trusted one another and, together, they worked out any situations not covered by the rules and procedures. Today, however, the system is trying to separate everything in the crowded controlled airspace—almost every type of flight in almost every type of weather. There are airline (air carrier), military, business, pleasure, and sports flights (general aviation). There are also slow-moving helicopters and short takeoff and landing (STOL) types of craft that are airborne.

Furthermore, the number of pilots for each of these categories—as well as the number of controllers—has

O'Hare tower (Chicago) during a peak operations period. In the foreground is the tower supervisor at the master console. In the background are five ATC controllers on duty. Two visual CRT displays of the terminal approach area may be seen at the center and at the right.

increased exponentially. And humans being what they are, one finds highly intelligent and proficient pilots and controllers; but one also finds idiots in the cockpit and control centers—and thousands who fall in between the two categories. The upshot of this situation is that pilots and controllers no longer trust each other implicitly. Today, there are so many rules, regulations, and procedures that nobody—pilot or controller—even pretends to know them all. Thus, it is an impossible milieu in which to achieve a safe, efficient, and orderly flow of air traffic. That the system is working at all is because of the vast amount of airspace between metropolitan air terminals, lots of luck (thus far!), and some dedicated people who struggle within the system to keep it from collapsing.

"The U.S. air traffic control system, like Dr. Frankenstein's creation, may turn out to be a monster that will do in not only its creators, but also the entire flying public . . ."

Some of the factors that are transforming ATC from an accident-preventing to an accident-causing system include:

- Environmentalists and residents in the proximity of airports who force compromises to be made in the rules governing safe flight (such as noise-abatement turns and power reduction during takeoffs).
- Vertical obstructions (tall buildings, smokestacks, TV and radio transmission antennas) that are permitted to penetrate into established safe flight altitudes.
- Inability of the FAA to enforce its safety regulations because of inadequate personnel budgets.
- The temptation in the aviation business to compromise on safety in favor of profits.
- The FAA's "procedures mill" that grinds out reams of rules!
- The air traffic controllers' union, PATCO (which is more concerned with member benefits than with the welfare of the system).
- "Ditto" with regard to the Airline Pilots Association (ALPA).
- The FAA's reliance upon chronic "controller shortages" to obtain money from Congress.

All of these factors combine to militate against safety. In essence, then, the laws of chance are working against the system. Meanwhile, let's "zero in" on a . . .

Detailed analysis of the adverse elements

Well-meaning environmentalists, armed with meager statistics based upon very recent theories and opinions, have selected the airplane as a prime villain. But regulations devised to protect the environment often conflict with regulations devised to keep flying safe. It's not exactly a case of deciding whether air pollution or an aircraft crash is more lethal; nevertheless, the trend toward the need for such a decision *is* established. Real estate developers and unhappy suburbanites living near major airports bring intensive pressure on local governments—and the FAA—to curb aviation activities in the vicinity. It does not matter that many of the developments have been constructed near existing airports; once the home owners move in, they attack the airport as a newly discovered nuisance that dispenses noise pollution!

The usual political result is that restrictions are imposed upon takeoffs and landings, many of which compromise the safe maneuvering of the aircraft and force pilots to act contrary to their best knowledge and instincts as to what constitutes the safest course to fly. Every time a pilot is compelled to negotiate a noise-abatement procedure, aviation safety is degraded to some extent and the ATC system is not functioning in the best interest of safety. As traffic volume grows and environmental lobbying increases, the potential for more aircraft accidents will rise.

The FAA has formulated certain airspace criteria to maintain the safety of aircraft in flight. Almost every time, however, that a TV station, a grain elevator, or

A thumbnail history of ATC, plus personal observations

The ATC system, the best and the safest—if you believe the pitches of those who work and fly in it—was created in the late 1930s out of sheer necessity, because commercial aircraft were getting too "intimate" with each other in terminal areas during instrument flight conditions. The history of ATC, in detail, is available for those who would like to reread the "glory years" of pioneer achievement—and the subsequent duller years of bureaucratic make-do and self-perpetuation.

Suffice it to say that ATC was developed and nurtured to provide the safe, efficient, and orderly flow of air traffic the public relies upon.

Lately, however, ATC has been skewed and jury-rigged—ostensibly in the name of greater safety. Just as there is a lot of romantic nonsense woven into the fabric of early airplanes and pilots, a lot of myths have sprung up about air traffic controllers. These myths have obscured some of the facts and have permitted the ATC system to receive a higher rating than it has deserved for some time.

From the vantage point of more than 30 years with the FAA (18 of them in air traffic control, and the balance in administrative support positions), it is possible to reflect upon the public image of ATC promoted by the FAA as compared with operational practices in its ATC field facilities. It doesn't necessarily add up to improved safety for those who fly.

In the early years, safety was always "top priority," with efficiency and orderliness being sacrificed, if necessary, to ensure maximum safety. Such contingencies occurred every time there were instrument flight conditions, and they became more frequent with the annual increase in air traffic. Thus, by the late 1950s, it was apparent to both pilots and controllers that the amount of inefficiency to provide maximum safety was intolerable. So pilots and controllers began to "cut corners"; rules and procedures governing the movement of aircraft were deliberately violated to achieve a "fingers-crossed" balance between safe flight and efficient flight.

The aviation companies and the CAA*/FAA permitted themselves to be "vaguely aware" of what was happening; but, they preferred not to look too closely, unless pressured

* The Civil Aeronautics Administration, predecessor of the FAA

a commercial high-rise complex wants to penetrate the airspace with an obstruction higher than the established minimum safe flying altitude, the hearing decision seems to come down in favor of the special interest. This forces a raising of minimum safe altitudes or the writing of a special-condition regulation or procedure. In any event, it gives the pilots and the controllers another small dot on their already complex charts and maps, and another amendment to flight procedures to worry about—and to try to remember! So air safety is jeopardized, because each change detracts from already established routine, and the ATC system does not function at its best.

The FAA has a huge problem of enforcing safety criteria because there are so many facets of aviation: the planning, building, and flying of aircraft; the licensing and monitoring of pilots; the necessity for ground and flight training; the licensing and monitoring of mechanics and other ground personnel; the inspection of airports, airways, navigational aids, etc. This is another aspect of the near-impossible job of coping with complex situations that are part and parcel of the total picture of air safety and ATC.

Perhaps the nub of the overall problem is *too much* FAA regulation—but self-regulation hasn't worked too well either, from the Garden of Eden down through history to the ATC system!

Because of budgetary constraints imposed by the Federal government, its agencies, and Congress, the FAA has been obliged to delegate the responsibility for safety inspections to persons and companies in the aviation community. Over the past few years investigations into aircraft accidents have shown conclusively that "self-policing" doesn't lead toward increased air safety.

Safety vs. profits

Another weakening link in the chain of air safety within the ATC system (and closely linked to the subject of self-regulation) is the matter of aviation business profits. Most companies are in the aviation business to make money. The profit motive is a respectable and acceptable reason for being in any business. However, in today's economic climate, the competitive need to cut costs and increase profits is ever-present; so, the decisions in the aviation world are tough ones to make when it boils down to the bottom line of confronting profits with safety. Such agonizing decisions come up regularly in the purchase of safety-oriented equipment, the proficiency of personnel, the maintenance of aircraft—and adhering to the letter of ATC system regulations and procedures.

Calculated risks *are* taken with safety in the fight to keep red ink out of the ledgers. And certainly, past accident investigations have revealed that specific safety advisory messages have been ignored. (At least, the temptation is always present to tilt toward "making another buck" if the noncompliance with the safety rule doesn't seem to be very important.)

Grist from the "procedures mill"

One of the most frustrating features of the ATC setup (and one which poses a further threat to air safety) is the previously mentioned air traffic control procedures mill (a trite, but accurate description!) that grinds away relentlessly in the bowels of the FAA's Air Traffic Service. It daily produces more than its quota of chaff and confusion for both controllers and pilots. Let's dig deeper into this contention: ATC procedures handbooks are revised quarterly; this was originally done to update and keep them abreast of the progressive evolution of aviation technology and flying skills. But, as is often the case with bureaucratic functions, the original intent got lost in the

"That the system is working at all is because of the vast amount of air space between crowded terminal areas, lots of luck, and some dedicated people who are struggling to keep it from collapsing!"

---

by an accident or a near collision.

Over the past 20 years, more corners have been cut and more chances taken in order to keep the ever-increasing traffic moving. Only the advent of radar, which made the jobs of pilots and controllers much easier, has prevented the system from breaking down long before now.

The authorities, of course, will vigorously deny that chances are being taken within the system; however, any eavesdropper at the confidential hangar-flying sessions—those man-to-man let-your-hair-down one-more-martini-please meetings—will hear pilots and controllers recite hairy tales added to the ATC folklore. Some proof of this chance-taking is indicated in the vigorous insistence by pilots and controllers that there be a "no-fault" provision in any system for reporting incidents and midair near-misses. Nobody cares to admit what is really going on "up there" unless he is guaranteed immunity from prosecution for violating air-traffic regulations.

When all concerned with ATC cease deluding themselves and see the system for what it is—a Rube Goldberg device (albeit somewhat more sophisticated) held together by good intentions and a lot of prayers—then a more objective approach can be acquired toward an airborne ATC system.

The need for safe, efficient, and orderly movement of air traffic is increasingly urgent, and the current FAA-managed apparatus simply doesn't provide it, despite the introduction of some new and flashy gadgets. Those who are willling to take a realistic look will have to admit that the needs won't be met in the future either.

The FAA, through its own statistics, admits its inability to provide the ground-based navigational aids, expanded air traffic controller force, and other related services necessary to keep pace with the growth of air operations. The FAA must have controllers to maintain its bureaucratic identity; so, instead of striking out boldly toward a new and different concept for separating aircraft, it insists upon upgrading, modifying, and supplementing its way through generation after generation of the same old basic system, replete with its built-in obsolescence and destined to "self-destruct" in about the year 2000—if not before! The answer *has to be* the implementation of an airborne computerized system.—*J.K.K.*

paper shuffle, and constant revision has become the "way of life" in order to keep a division or branch on the organizational charts and to retain the procedures rewriters on the payroll. Revision for the sake of revision appears to have taken over, and the ground-based ATC system has become ever more cluttered, clogged, and cumbersome. There are simply *too many* ATC procedures on the books, and many pilots and controllers have just quit trying to remember all of them. Thus, they concentrate primarily on those that apply to their current flight routines.

Testimony at the most recent aircraft accident investigations has elicited statements by pilots, controllers, airline officials, and FAA people that indicate a fuzziness, general unfamiliarity, or outright ignorance of the very "rules-of-the-airways" upon which the safety of the ground-based ATC system depends.

With the increase in traffic that has been predicted and the foreseeable lack of any appreciable increase in the capability to handle that projected traffic growth, both pilots and controllers will be increasingly overburdened with just the physical actions of their work. Therefore, the mental strain associated with this fact will result in more and more small—but vital—bits of information being overlooked or forgotten, with consequences that are ominous.

PATCO's power ploys

The controllers' union achieved its foothold on the FAA by means of its notorious "sick-out" of 1970 (which, in less polite terms, was an illegal and unethical strike against the Government of the United States). It was also a clear violation of the oath taken by every controller when he was hired by the Government. As was the case with the postal workers, the Federal government permitted the strikers to get away with their illegal acts.

The FAA rehired most—if not all—of the strikers and thus set the precedent for more sick-outs whenever PATCO decides they are needed. Traffic slowdowns (an admission by the union and the FAA that rules and regulations are circumvented during the usual ATC operations for the sake of "efficiency") were adopted as recently as October 1974, when PATCO negotiations with the FAA were stalled. In these instances, the controllers "adhered to the book"—which simply means they followed to the letter the rules that ensure safe flight (the same rules they often disregard in order to "move the traffic").

An ATC system that is at the mercy of a controller work force, willing to get sick or slow down air traffic on cue from its union, is a dying system. Since the FAA is powerless to stop these actions, this seems to be the most cogent argument for an airborne CAS.—*J.K.K.*

Influence of the "unions"

Another element in the system posing a threat to continued air safety is the Professional Air Traffic Controllers Organization (PATCO), which now represents the controllers on personnel and working environment matters. PATCO negotiates an annual contract with the FAA for the "care and feeding of controllers." Because in the early days the FAA was not as considerate of its controller work force as it might have been, the formation of a union was necessary and inevitable. If the controllers' union leadership becomes stable and responsible, it could do much to relieve some of the safety threats inherent in ATC; but that desired responsibility and stability have yet to be proved.

Nevertheless, PATCO "will call the shots" for the FAA, and there *are* problems in that situation. Controllers who are more concerned with personal benefits than with public service are not disposed to worry very much about safety or efficiency in ATC. The union-inspired attitude of controllers doing only that which is required by strict interpretation of regulations—instead of giving an "extra inch" to volunteer unrequested aid—probably contributed to recent aircraft accidents. Although the controllers were legally and procedurally right, they were, nevertheless, negligent in disregarding available data because they considered it to be the pilot's responsibility to utilize the particular information. With this attitude prevalent, I feel the ground-based ATC system will gradually fall apart unless the controllers can again be inspired to respond "above and beyond the call of duty." Unfortunately, today's unionized controllers give no evidence of being motivated to that extent.

The pilots also have their union to represent them in the same areas of concern to the controllers; this is the Air Line Pilots Association (ALPA), which, from time to time, exhibits some of the same instability and irresponsibility toward ATC as does PATCO. Always vigilant to retain pilot control in the cockpit for the safety and efficiency of airline flights, ALPA, however, rushes to disavow all such responsibility when an air crash or accident occurs. ALPA is very zealous in protecting these "cockpit rights" as it strives to keep the pilot in command (as regulations say he is); but, simultaneously, the union tries to "pass off" possible pilot errors to the controllers. Probably the rashest of ALPA actions was its threat to discontinue the use of cockpit recorders because they were revealing pilot actions that were uncomplimentary to the PR image of airline pilots! Furthermore, in its press releases, ALPA has exhibited an animosity toward controllers, the FAA, and even airline management that is not conducive to a smoothly operating ATC network. Thus, I must reiterate that a computerized airborne system, independent of either pilot or controller action, might remove the causes of such friction. A system that depends upon trust and cooperation between pilots and controllers won't continue to function if these criteria are missing.

The FAA bureaucracy, and a summation

With all the evident need to get away from a ground-based ATC, why does the FAA continue to support it? Why does it give short-shrift to the suggestions to get with an airborne system? The answers are that the FAA can produce much oratory, many statistics, and reams of PR copy to support whatever it really decides to do. As is the case with all government bureaucracies, the appearance of doing something is more important than doing it. Since almost everyone recognizes the need to do something about the outmoded machinery that moves air traffic, the FAA applies a new facade and adds a bumper sticker. Putting a new horn, buzzer, warning light, or radio speaker in the cockpit for controllers to activate or interpret for the pilot may give the appearance of doing something about a new CAS—but it doesn't overcome or upgrade the deficiencies inherent in the basic system.

One must bear in mind that the very existence of the FAA depends upon an ATC apparatus with thousands of controllers. More than 90 percent of its work force is either directly or indirectly related to ATC. The FAA has talked Congress into appropriating large sums of money—sometimes legitimately—with its cries of alarm about the short supply of controllers. A midair crash or an air carrier disaster always triggers the clamor for more controllers *in a hurry;* and, of course, the necessary monies to recruit, hire, train, and retain them.

The FAA's facility staffing tables, which stipulate the number of controllers needed for individual facilities, are highly suspect because the FAA seldom "staffs up" to the number indicated. In addition, the FAA controller-training program is an expensive, wasteful, time-consuming process that is administered in a manner that guarantees a 20–30-percent personnel attrition rate. This, of course, helps to perpetuate the shortage. Further, controllers now have an early-retirement "second-career" training program that is ripping off the taxpayers for millions of dollars—and accentuating the "controller shortage." In my mind, there is no doubt about it: a controller-operated ATC system is essential to the well-being of the FAA—if not to that of the flying public!

The days are numbered for the conventional concepts of fixed airways and approaches to terminal areas, ground-based air traffic controllers, and the ever-multiplying and burdensome procedures and regulations. When it is needed under IFR conditions, the existing ATC system is neither truly efficient nor truly safe. Responsibility for collision avoidance must be taken out of the hands of both controllers and pilots and replaced by a well-programmed real-time computer. It would be disastrous if a credibility gap breached the confidence the public has placed in air travel. With that final warning in mind, let's reject the fallacy that an airborne ATC system may be too expensive, too complicated, too revolutionary, and too visionary. It is needed and its time will come. ◆

In an effort to provide a diversity of views on air safety, *Spectrum* invited **John K. King**, an FAA veteran with more than 30 years' experience, 18 of which were as an air traffic controller, to author this article. Mr. King retired from the agency in 1973. Until March 1, 1975, he was executive director of the Air Traffic Control Association in Washington, D.C., managing the headquarters office. Mr. King's experience in aviation includes that of chief of the FAA's Southwest Region Training Branch, principal advisor on air-traffic training in the FAA's Washington headquarters, and active participant in air traffic control administration. At the present time, he is a lecturer, free-lance writer, and special correspondent for the *Journal of Air Traffic Control*. He operates out of Fort Worth, Texas.

Avionics II

Air safety: the view from the cockpit

An experienced airline pilot appraises the past, present, and future of air safety and air traffic control trends

A veritable revolution in air safety systems is in the offing—a revolution that will drastically alter the face of U.S. air traffic control as we now know it. At present, this revolution is in what might be called its "manifesto stage." By this I mean that a raft of proposals have been offered by the U.S. Department of Transportation (DOT) under the umbrella acronym UG3RD, or Upgraded Third Generation Air Traffic Control System. These proposals, if implemented, would mark the greatest change in air-to-air separation since the original establishment of separation standards in the 1930s. Further, navigation, surveillance, and communications would all be affected in a major way. The question devolves: What, exactly, would be the effect of this proposed automation program known as UG3RD?

The purpose of any air traffic control (ATC) system is to provide for the separation of aircraft within the system's jurisdiction—i.e., all planes operating under instrument flight rules. As requirements and technological capabilities change, ATC systems evolve. During the last 50 years, there have been three generations of such systems. The box on p. 73 provides a "family tree" of ATC systems, but here it should be sufficient to note that the most recent, third generation ATC system—the one in use today—was prompted by the Government's "Beacon Report," which followed a midair collision over the Grand Canyon in the late 1950s.

In many ways, this system was like its predecessor, except for the extensive use of computers. This facet of today's system received a great deal of publicity when it was first introduced—there were promises of vastly improved safety and capacity—but, in retrospect, it appears that this was simply part of a tremendous selling job to instigate the appropriation of large sums of money for what amounted to an internal modernization of the Federal Aviation Administration (FAA).

This modernization created a huge data-processing plant within the FAA. And that plant, with its touted combinations of software on the ground and hardware

William B. Cotton
Air Line Pilots Association

in the sky, was to provide a myriad of improved services. But the facts are that it did little to improve either safety or capacity, and did nothing to improve the productivity of the air traffic specialists.

The conception of a new generation

The FAA's vast new automation base began being installed in the late 1960s at precisely the same time that airport and ATC system capacity were reaching their saturation points. With long delays both in bringing aircraft down and in getting them into the air becoming commonplace in an age of ever-increasing air traffic, the U.S. Department of Transportation decided to commission a study of possible long-term solutions to the problem, assuming unconstrained expansion in aviation. The resulting Air Traffic Control Advisory Committee Report of 1969 has become the bible of ATC planning, and its recommendations constitute what is now known as the UG3RD.

Conceived during a period of strong domestic economy, cheap fuel, and a seemingly endless demand for use of the skies, the UG3RD not surprisingly emphasizes the ground-support facilities of the air traffic service—which were being severely strained. For safety reasons, a concept called intermittent positive control was proposed to limit freedom of flight under visual flight rules (VFR) to the degree necessary to avoid collision, while flights under instrument rules (IFR) were to be much more rigidly controlled from the ground through improved surveillance and data-linked "command clearances" of heading, altitude, and airspeed. This technique was proposed to accomplish terminal area navigation in what was expected to be a sky blackened with airplanes. A whole potpourri of automated functions and services that seemed feasible with the big computers were lumped into the UG3RD, whether or not they were mutually compatible.

As shown in the 1975 edition of the National Airspace System Plan, the UG3RD system currently features several categories of suggested improvements. Some are entire separation concepts, some are hardware only, some are software only, and nearly all contain implied procedural changes. These improvements include: intermittent positive control, a discrete address beacon system (DABS), area navigation (RNAV), a microwave landing system (MLS), upgraded ATC automation, airport surface traffic control, a wake vortex avoidance system, aeronautical satellites, and automation of flight service stations.

Although not apparent from the list, a number of important reductions in separation standards are contemplated. The RNAV Task Force specified a reduction in route spacing from the present 8 to 3 nautical miles. In addition, it is assumed that with the use of the MLS, the spacing between independent parallel approaches will be approximately 2500 feet, and, in addition, lower radar separation minima are being discussed as possibilities with the improved surveillance capabilities of DABS. Although none of these reductions have been subjected to a realistic safety analysis to determine how the risk of collision will be affected, they are considered necessary—and possible. And their implementation, if it occurs, will literally revolutionize air-to-air separation.

The consequences of UG3RD

Navigation, surveillance, and communications, as I have said, are all slated for major changes. Area navigation, airport surface traffic control, and the microwave landing system require new ground and airborne hardware and operating procedures. Their very versatility poses the greatest problem in the organization of the ground flow-management machinery. Surveillance, via the discrete address beacon, would be considerably better than with our present secondary radar. In addition to its more accurate and reliable data-acquisition capability, FAA says a major improvement would be the integral digital data link that permits messages to be sent to or from any air-

In the cockpit of a Boeing 747, preparing for takeoff, are the captain (at left) and his copilot. At right in the foreground is the flight engineer, in front of his control panel.

plane within the radar's range of coverage. Such a link has been proposed to carry the bulk of routine ATC messages, and is an integral part of the plans for advanced metering and spacing. DABS also requires completely new hardware on the ground and in the aircraft.

The aeronautical satellite would make independent surveillance of oceanic flights possible for the first time, which certainly would lead to drastic reductions in the applicable separation standards. The wake vortex avoidance system, if it works, could assure safe landing intervals through responsiveness to existing weather conditions and type of aircraft. In the UG3RD, computers would monitor the live traffic situation, make decisions for the controllers, and issue control clearances to the pilots. And IPC offers the first serious collision-avoidance capability to VFR pilots ever proposed by the FAA's Air Traffic Service.

A pertinent question

The UG3RD, in short, involves major modifications to every aspect of the present ATC, as well as a number of services not previously available. Why is it not labeled "fourth generation"? The answer may be that it was proposed while third-generation equipment was still being installed, and it is simply good salesmanship to call it an extension of the system that airspace users were just beginning to pay for.

More important: Why is a system that seems to promise vastly improved service received so coolly by airspace users? There are a number of answers.

First, the aircraft operator realizes very quickly that none of the third-generation equipment he has just purchased will work in the upgraded version. Transponder and altitude encoding equipment will have to be replaced with DABS equipment, standard VOR-DME by RNAV gear, and ILS replaced by or supplemented by MLS. Even the communications transceiver must be replaced to accommodate the split-channel 25-kHz spacing. In response to the ensuing complaints, the FAA agrees to continue to service the old equipment; on the other hand, its users will forgo any advantages of the "upgraded third."

A second reason for the skepticism arises from a look at the FAA's past performance in developing and implementing its plans. The third-generation equipment and automation programs were child's play compared with what has been proposed for the UG3RD—yet the original schedule for the former has slipped ten years to where the last of the NAS Stage A equipment is just now being installed in traffic control centers around the U.S.

A third reason for user disinterest is a feeling that some of the major components of the program just won't work. For example, the random navigational system (RNAV) offers great potential for increased flexibility and efficiency in the operation of aircraft—but the FAA's plan is to establish a system of RNAV routes similar to what we have now with just VOR, thus destroying both the randomness and efficiency associated with the use of RNAV equipment.

In this same vein, the process of derandomization of arriving aircraft involves two types of navigation. The first type is navigation with respect to a geographical reference frame—finding the optimum path from origin to destination. The second kind of navigation involves merging with some and avoiding other aircraft; in other words, navigating or controlling the flight path with reference to other aircraft known as pertinent traffic. These two types of navigation must be carried out simultaneously, and today's terminal radar procedures have evolved to carry out this task. Area navigation alone, except in a purely strategic four-dimensional form, cannot accomplish the objective. Metering and spacing are designed to implement the second type of navigation, but have insufficient information for efficient geographical navigation. In effect, then, no one part of the proposed terminal navigation receives all the data necessary for the safe, expeditious movement of aircraft—and yet the two types cannot be combined in the UG3RD, as proposed.

The generation concept

Although the generations of air traffic control have no firm definitions themselves, they do have a number of elements in common: separation standards, navigation, communications and surveillance systems, identification of predominant aircraft types, and modes of operation.

The first generation. Separation standards were based on time, route, and altitude separation. Aircraft navigated along routes based on low-frequency-range legs. Communications of position and airway clearance were primarily relayed from the aircraft and traffic facilities through intermediate communication stations, either the flight service stations or the airlines company radio. Independent surveillance of aircraft positions did not exist; thus, except for position reports, it could be called an open-loop control system. Most of the aircraft had speeds of less than 200 mi/h, and were engaged in short-haul transportation. The system was organized because the infant airlines wanted to operate in instrument weather conditions in which the see-and-avoid concept of collision avoidance could not function.

The second generation. After World War II, several technical advances changed the predominant methods of navigation and the aircraft were flying faster, higher, and further. The basic time, route, and altitude separations were still used, but radar provided an independent fix of aircraft position at a much higher information update rate, as well as smaller separation minima. Very-high-frequency omnidirectional aids enabled pilots to fly any course into or out of the stations. However, air traffic services were firmly fixed along the old routes, making the first-generation separation techniques easy to apply, but ruining much of the versatility of the omnidirectional range station for instrument flight. Static-free VHF communications provided direct communications between pilots and air traffic personnel, enabling greater efficiency and flexibility in obtaining modifications to original flight plans as well as information on potentially perilous conditions.

The third generation. This generation brought no new separation standards, although the use of radar separation became more widespread. Navigation and communications remained the same, except for a gradual increase in the amount of navigation by radar vector as an alternative to the holding pattern. This method did not actually lessen the delays, but it did disguise them. As extensive use was made of computers to automate the "housekeeping" functions of flight-plan storage, flight data strip printing, and even the tracking and display of the aircraft targets on the ground controller's scope.—*W.B.C.*

Secondly, IPC, the collision-avoidance service dependent on the widespread and successful implementation of DABS surveillance and data-link equipment, demands one of the most massive and sophisticated data-processing and programming undertakings ever attempted. Even if it should be successful, it will work only where line-of-sight beacon coverage exists—and historically most midair collisions have occurred near airports without radar, and hence not adaptable to the DABS.

Furthermore, the metering and spacing automation program, which attempts to fix the landing interval more precisely than a human controller can, so far has met with complete failure. Much of the difficulty stems from the software's inability to anticipate the various response times of the pilots and the rates of turn, acceleration, deceleration, climb, and descent to expect when the heading, altitude, and airspeed instructions are issued. The human controller's experience with the different aircraft operators and the subtleties of direct voice communications are difficult to better with a fixed computer program. In addition, there is the glaring conflict between RNAV and the metering and spacing program in the terminal area. In the former, the flight path is programmed by the pilot; in the latter, by an air traffic service computer. The two cannot exist simultaneously, but the UG3RD calls for just this impossibility.

The microwave landing system (MLS) affords another example of the poor likelihood of success. It does have several distinct advantages over ILS, primarily in reducing multipath problems and interference from other aircraft, in easier siting of equipment, and in pilot selectivity of azimuth and glide slope. However, most of the FAA's sales effort has been focused on the system's curved-approach capabilities, which are said to afford great terminal area route flexibility and noise relief. In spite of the advanced state of MLS development, this feat has been demonstrated only on airplanes with Flight Directors—and, because of its cost, most IFR aircraft do not contain this necessary equipment.

The last, and most serious, criticism of the UG3RD relates to the implied philosophy of control, the location of responsibility, and the combination of the various features into a system that will fulfill the future navigational requirements of the users. The proposed programs have the look of engineering ideas that arose simply because they became technically possible. They may remotely resemble certain operational or safety requirements of the National Airspace Systems users, but certainly have not been inspired by them.

The air traffic management philosophy of the FAA's Engineering and Development Branch is to locate all functions centrally on the ground. The aircraft are to be fitted into rigidly structured flight paths designed for the convenience of the Air Traffic Service, not for the pilots operating within it. The FAA has not sought to optimize the flight paths or the speed and altitude profiles of aircraft, or even to accommodate such optimization. Instead, the upgraded automation will impair the ability of individual operators to optimize their own flights. The approach to RNAV implementation is an example of this. Random flight planning is the last RNAV function to be considered by the FAA, yet it is the only aspect of the technique capable of producing a substantial benefit.

Responsibility without control

Not only will the UG3RD do nothing to improve a pilot's capability to plan or execute more economical flight paths, but it will erode the pilot's ability to direct those paths. The design of flow control, metering and spacing, and ATC communications by data link decreases the pilot's awareness of the outside environment—weather phenomena, geographical position, near-term future flight path, and relationship to other aircraft in the area. It is this aspect of UG3RD that makes the system totally unacceptable to pilots. Even assuming that the users could accommodate to the poor economics of allowing ground agencies to direct flights, the potential hazards of reduced situational awareness are tremendous. Most accidents attributed to pilot error occur because the pilot was unaware of danger—the malfunctioning of some onboard system, hazardous weather, proximity to terrain or aircraft.

Why, then, is a display that can present effectively the geographical and traffic situation of an aircraft to its pilot not given the highest priority in the FAA's air traffic plans? The answer must be that it does not fit into their philosophy of keeping all control on the ground. However, that philosophy is inconsistent. The pilot's responsibility for the conduct and safety of his flight is strictly maintained by the FAA, especially after an accident, even though the National Transportation Safety Board hearings in such cases usually reveal the pilot's lack of awareness of a dangerous situation. In spite of this, the design of the UG3RD detracts even more from the aircraft operator's sense of awareness and control—yet nowhere does it propose to lessen his legal responsibility.

Obviously, the "upgraded third" has a number of useful features, although they have not yet been combined into a meaningful total system. This will only occur when the FAA accepts its proper part in air traffic management—to provide the pilots with sequence determination, right of way, data on position, identity, and intent of aircraft, and status of the navigation and airport facilities. In other words, FAA's role should be to equip pilots with the data that will enable them to carry out, efficiently and safely, their responsibility to their passengers and to themselves. ◆

William B. Cotton has been active as a flight officer for United Airlines since 1967. He has served as chairman of the National Air Traffic Control Committee of the Air Line Pilots Association for the past five years, and as vice chairman of the Air Traffic Study Group of the International Federation of Air Line Pilots Associations. He has been a spokesman for pilots on the subject of ATC at many meetings and symposia, has published several articles on that topic, and is also an advisor to international bodies studying aircraft separation and oceanic ATC. He received the S.M. degree in aeronautical and astronautical engineering from the Massachusetts Institute of Technology and the B.S. from the University of Illinois, and is a member of Sigma Gamma Tau, the honorary aeroengineering fraternity. *Spectrum* invited Captain Cotton to produce this article because of his experience as an airlines pilot and his special background in ATC problems.

Ampère: father of electrodynamics

The master mathematician, teacher, and practitioner of electrical science is celebrated on the 200th anniversary of his birth

In 1789, at the age of 14, André-Marie Ampère witnessed the murder of his father, the mayor of Lyon, by a mob during the "Reign of Terror," following the French Revolution, when every titled person or official of the overthrown regime was subject to execution or imprisonment without trial. The trauma of this experience, plus the death of his wife in 1803, made him an embittered and reclusive person for most of his life. But it was, perhaps, this introversion that spurred him to the single-minded pursuit of the hallmark discoveries and experimentation in electrical science that have earned him a special niche in the gallery of modern scientific history.

Ampère received the bulk of his primary education at home and, by 12, he was proficient in the highest mathematics of the time. Although he received no formal education, he had mastered Lagrange's *Mécanique Analytique* at 18. From 1796 to 1801, he earned his living, with considerable success, as a private tutor in mathematics, physics, and foreign languages. His brilliance in science and the arts, however, did not pass unnoticed; in 1801, he was appointed to a teaching post at Bourg. In 1802, he became professor of mathematics at the Lycée de Lyon; and, for the next seven years, Ampère also lectured on mathematical analysis at the École Polytechnique in Paris (where he was appointed a professor in 1809).

Anglo-French science—and politics

Warfare, apparently, was conducted on a more "civilized" plateau during the Napoleonic era than it is today. Although Bonaparte was an obsessive "anglophobe," this personal prejudice did not prevent him from permitting the unhampered interchange of ideas and contacts between French and English scientists. The beginning of the 19th century was a turbulent period of renascence and dramatic new concepts in the scientific institutions of both nations. Thus, it was not surprising that, in 1813, at the height of the Napoleonic Wars, Sir Humphry Davy, the famous English chemist, was accorded special permission and safe conduct to travel freely in France because *the object of his visit was scientific.* He was even received cordially by Napoleon, who witnessed some of Davy's noted experiments.

Ampère was apolitical and had no strong feelings about the war being waged against England. In fact, he welcomed the new ideas from Britain. He was always ready and eager to accept novel scientific developments and was most generous in praising his contemporaries. For example, as early as 1810, Ampère sent a letter to Davy which ended ". . . allow me to congratulate myself upon having found this happy occasion to offer you the accolade of my profound respect and also sincere admiration for your great [scientific] discoveries." This was the beginning of a 15-year-long cordial correspondence between the two men, devoted largely to the area of chemistry.

Ampère and the French Academy

One of the foci of scientific activity, especially in electricity, was the French Academy of Sciences. The top scientists of the period reported their discoveries and debated the merits of their theories before that august assemblage. The transactions and minutes of these meetings were duly recorded for future reference. In addition, there was considerable spirited correspondence among French and foreign scientists. For-

Engraving of André-Marie Ampère at the time he was elected to the Royal Institute of France in 1814.

Gordon D. Friedlander Senior Editor

rules and laws. Among these are the following:

- Two electric currents attract when they flow parallel to each other in the same direction, but repel when the flow is parallel in the opposite direction.
- When metallic conductors carrying currents cannot turn into a parallel plane, each conductor tends to move the other conductor into a position parallel to it and in the same direction.
- The repulsions and attractions are completely different from those of electrostatic electricity.
- The relationship of the flow of an electric current and a magnet is covered by the law of attraction for two electric currents, as discovered by Oersted. (This law ensues from the logical deduction that a magnet is a product of electric currents made by the action of iron particles on each other, in a manner similar to the action of the Voltaic pile.)

To ascertain the interaction and mutual action of currents, Ampère suggested two experimental methods: the actual measurement of forces at various distances, and the "null method," which is more accurate. In the latter technique, which determines whether electrical forces are in true balance and equilibrium, Ampère mounted two coils so that they were equally and oppositely affected by the earth's magnetic field, thereby producing an *astatic* pair of coils. By means of this apparatus and similar "home-made" equipment, he observed that

1. A current's effect is reversed when the direction is reversed.
2. The effect of a current is not changed by the configuration of the conductor; thus, a current flowing through a coiled circuit is the same as that flowing through an expanded circuit.
3. The force exerted by a flowing current on an element of another circuit is at right angles to the line uniting them.
4. When all linear dimensions are increased proportionately, the force between two elements of circuits is unaffected, and the current strength remains unchanged.

Ampère's telegraph

Despite his preoccupation with the theoretical aspects of electricity and magnetism, Ampère showed his application side in discussing practical uses for his discoveries:

With as many wires and magnetized needles as there are letters, and by placing each letter on a different needle, one could, with the aid of a pile placed at a distance from the needles . . . create a sort of telegraph with which to write all the details which one might wish to transmit, across whatever obstacles there might be, to a person instructed to observe the letters placed on the needles. By setting up on the pile a keyboard whose keys carry the same letters and [which] *would make the connections when depressed, this means of communication would function with great facility, and the time to operate it would be merely the time necessary to touch a letter on one side and to read it on the other.*

Gauss and Weber, about 20 years later, actually developed an electromagnetic telegraph.

Honors to the end of his career

During his 61-year lifetime, Ampère became a member of the French Academy of Sciences, the Royal Institute of France, a Fellow of the Royal Society of London; he was also elected to the scientific societies and academies of Edinburgh, Cambridge, Geneva, Stockholm, Berlin, Brussels, and Lisbon. He served as an honored professor of the École Polytechnique of Paris and the Royal College of France.

Ampère's personal life was disciplined and joyless (largely the result of the tragic deaths of his father and wife). Nevertheless, his scientific interests covered a broad and diverse range. In addition to electricity, his published works included studies in chemistry, mathematics, physics, psychology, botany, and natural history. From 1800 onward, he evolved a new classification of all the sciences; and, by subdividing them, he arrived at a total of 128 scientific disciplines and subdisciplines—one of which he termed "cybernetics"; another, he called "technesthetics."

Ampère died in Marseilles on June 10, 1836.

The International Electrical Council, meeting in Paris in 1891, designated the practical unit of electrical current as the *ampere*. ◆

Why the world had to wait

Bern Dibner, director of the Burndy Library, correctly includes André-Marie Ampère in his booklet *Ten Founding Fathers of the Electrical Science*. But until 1885, it was almost impossible to find published reports of the works of this great contributor to 19th century science and technology. But there are reasons for this: Ampère's most significant works were performed from around 1820 to 1827. Ampère was born in 1775, and thus his early life spanned the great upheaval—and consequent unpleasantness—of the French Revolution. Publication of the renowned French scientific journal *Comptes Rendus* was suspended by The Directorate—the rather motley group of egalitarian (and anti-science) revolutionaries—that superseded the monarchy. Thus, *Comptes Rendus* was suspended from about 1792 to 1836. And the new series of the work, (from 1836 forward) did not bother in any way to cover prior events or notable discoveries that occurred in the 44-year-long hiatus. So, the important and valuable Ampère papers were not revealed to the general public until 1885—at least 60 years after their author had performed his important experiments.

Recommended Reading

For a good informative monograph on Ampère, see Dibner, B., *Ten Founding Fathers of the Electrical Science*. pp. 25–26. Norwalk: Burndy Library Publications, 1954. Another short piece, revealing the open interchange of ideas among French and British scientists of the early 19th century is by Gardner, K. R., and Gardiner, D. L., "André-Marie Ampère and his English acquaintances," *Brit. Jour. for the Hist. of Sci.*, vol. 2, pp. 235–245, July 1965. Also, the reader may wish to refer to: Tricker, R. A. R., Early Electrodynamics, Elmsford: Pergamon Press, 1965; Herivel, J. W., "Aspects of French theoretical physics in the 19th century," *Brit. Jour. for the Hist. of Sci.*, vol. 3, 1966; and Brown, T., "The electric current in early 19th century physics," *Hist. Stud. in the Phys. Sci.*, vol. 1, 1969.