Speaking of Communication

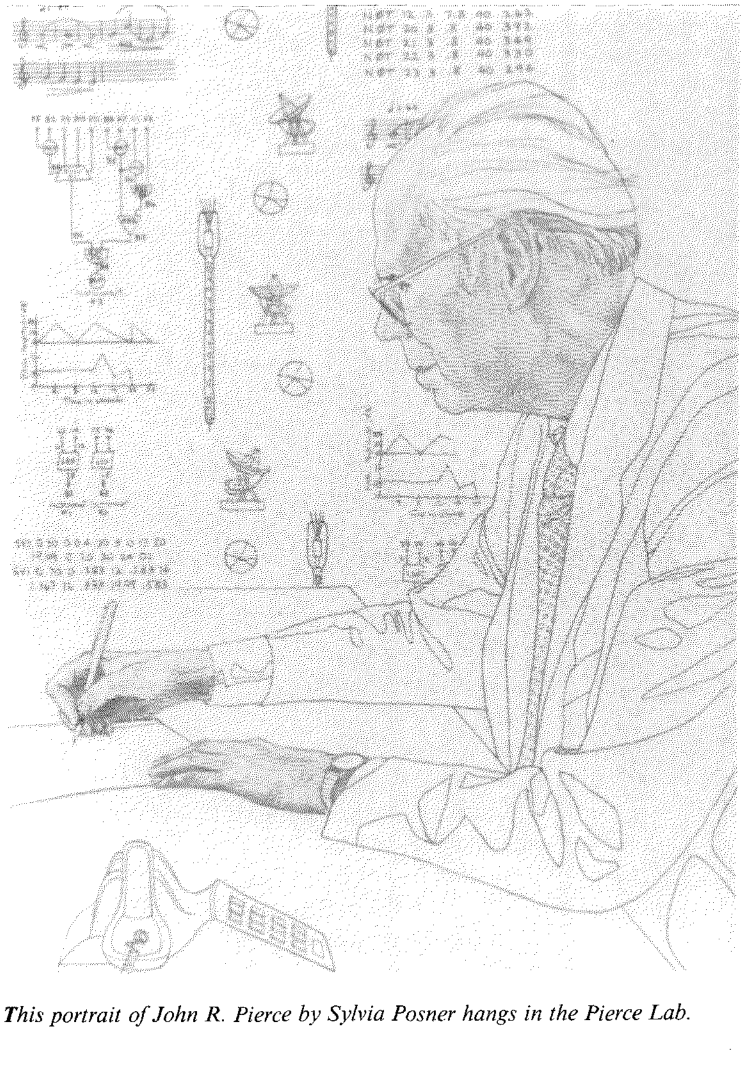
by Edward C. Posner
Rodney M. Goodman
Robert J. McEliece

As communication systems have become more extensive on earth and in space, what are the limits of their energy and complexity? What are the means of protecting transmitted information from error and theft, and what will prevent a logjam of data bits or an infinite queue of waiting telephone calls? While much of the information theory that underlies today's high-performance communication was formulated 40 years ago, it is only recently that very large scale integration (VLSI) has made such electronic communication practical commercially — generating further opportunities for communications researchers.

The communications research group in Electrical Engineering is trying to solve some of these problems and is gaining an international reputation in the process. Caltech is now one of the leading academic institutions world-wide in information and coding theory and in error correction for data storage. We are also, with support from Pacific Bell, one of the very few academic institutions in the country doing teaching and research on circuit switching and, together with other departments on campus and at the Jet Propulsion Laboratory, are at the leading edge of research on associative memory.

The communications group has close interactions on campus with computer science, mathematics, neurobiology, and chemistry and biology, as well as with JPL. Such interactions are beneficial to all, because the techniques of information and communication theory, involving probability and combinatorial reasoning, apply widely in other areas of information processing and networks, and in statistical physics.

John R. Pierce (BS 1933, PhD 1936), professor of engineering emeritus, started the communications group at Caltech in 1971. The Pierce Lab in 214 Steele is named in his honor. We three have been involved with it



This portrait of John R. Pierce by Sylvia Posner hangs in the Pierce Lab.

for varying lengths of time — Posner, a holdover from the Pierce era; McEliece since 1982; and Goodman, the newest arrival, since 1985. The research projects of the communications group are also varied, and we would like in this article to give an overview of this exciting field.

McEliece's research specialties are information theory and error-correcting codes, subjects that have developed from Claude Shannon's important work first published in 1948. This research concerns the problem of communicating reliably over unreliable channels and has wide-ranging applications. One subject his group is investigating is anti-jam communications — devising efficient strategies for communication in the presence of intelligent and adaptive interference. This problem can be viewed as a game with two power-limited players — the communicator and the jammer — and McEliece and his group have been able to combine game theory with information theory to develop several new classes of anti-jam communication strategies using error-correction concepts. Some of the results have been quite surprising and may find application to protecting communications.

McEliece is also interested in the application of information theory to the problem of reliable storage and retrieval of information from computer memories. Here the communication is the transmission of data not in space (from here to there) but in time (from now to then), and the transmission medium is not the electromagnetic "ether" but rather a storage medium, such as semiconductor RAMs, magnetic tapes, and magnetic disks. As device physics technology pushes these storage media toward their physical limits, the reliable retrieval of the stored data becomes more and more difficult. For example, as RAMs become larger and feature sizes shrink, soft (non-permanent) error rates rise due to background alpha particle radiation, circuit noise, and other effects. Error-control coding has become essential if the computer's memory system is to have a reasonable mean time between failure. McEliece and Goodman have developed theories for estimating the mean time between failure of coded systems, which are actually being used by system designers in specifying coded RAM systems. Back in 1948 Shannon's theorems predicted with uncanny accuracy that in most high-performance communication systems it is better in the long run to correct errors (and that is what we are doing here) than to try to prevent them by overwhelming the noise with signal.

Particular research projects in the data-storage area concern the design of error-correcting codes for magnetic tapes and the reliable storage of data on semiconductor RAM chips. The latter involves both the ultimate physical limits of data storage density (combining semiconductor physics with Shannon's theorems) and practical methods for incorporating error-correcting codes right on the high-density RAM chips themselves rather than by adding extra chips.

VLSI can also be used to build powerful decoders for both space and time communication. Recently the communications group developed a single-chip VLSI implementation of a decoder for the important class of Reed-Solomon codes. These codes have applications ranging from deep-space communication (Voyager at Uranus) to cellular radio to high-density data storage on magnetic tapes and computer disks.

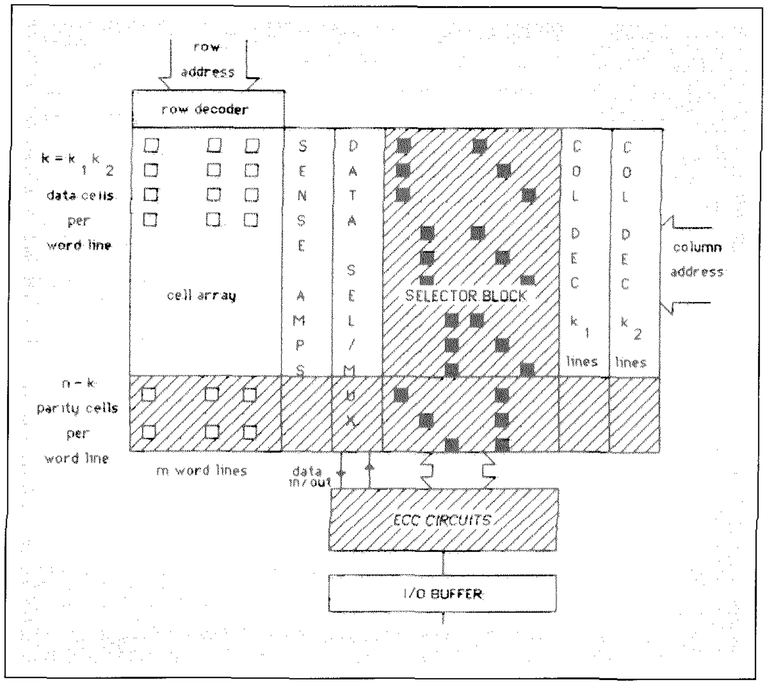
Some very exciting recent work of McEliece, Posner, and Eugene Rodemich (JPL), with both students and faculty in the Electrical Engineering Department and others, involves the asymptotic storage capacity of the associative memory network concept developed by John Hopfield, the Roscoe G. Dickinson Professor of Chemistry and Biology. Information theory and pattern recognition are involved because the concept of an associative neural net has close ties to both fields. John Lambe of the Electronics and Control Division at JPL heads the design and fabrication effort. The group is also working closely on neural nets with James Bower, assistant professor of biology.

Goodman's interests, like those of the communications group in general, lie in digitally getting the information from A to B through space (communications) or time (storage), with minimal distortion (coding), without it being ripped off (cryptography), via the fastest route possible (intelligent computation-intensive algorithms). More specifically, he's interested in digital communication and computer networks, error control and public-key cryptography for data transmission and storage, digital signal processing for speech and vision, VLSI architectures for decoding and signal processing, expert systems, and high-level computer language hardware implementations.

Although diverse, these interests have two strong themes that are fundamental to the digitization of communications networks: information theory (to give us a sound theoretical path to follow), and computation (to give us the most efficient method of implementing algorithms). Computation cannot be divorced from VLSI, because VLSI is changing the way we think about algorithms and their efficiency. For example, it is now possible to build considerable intelligence into even the smallest communications system, and, in fact, intelligence in networks, and our understanding of it, are key requisities for the all-digital public and private networks of the future.

Widespread acceptance of public-switched, all-digital telephone networks will rest partly on the ability to guarantee secure point-to-point communications. The user must be confident that the information cannot be tapped or altered without his knowledge. Authentication is also important in a digital system: How do you know whom you are talking to? On the telephone we recognize a person's voice, but "bits is bits" and a digital bitstream can be impersonated much more easily than an analog signal. This fact puts cryptography, in the form of secret coding and authentication, firmly into the civil communications arena. Public-key cryptosystems (PKCs) offer a means to provide both these features automatically, but there are many practical and theoretical questions that need answering before such systems can be implemented. Indeed PKCs still seem very much the ultimate party trick — it turns out that two people can openly exchange numbers and quickly establish a secure common key, while the eavesdropper has to do an almost impossible amount of computation to get the same result.

Goodman and research fellow Tony McAuley are doing research in three main areas of public-key cryptography. First, they are trying to develop new practical PKC algorithms; they and McEliece have invented some as-yet-unbroken PKCs. They are also developing new "broadcast" PKCs for transmitting a message securely to a number of users. Conventional PKCs are one-to-one systems and cannot handle this. These broadcast systems have great potential application in direct broadcast satellites and in packet-switched data networks, where the number of packet hops can be reduced significantly by using group-addressed packets. Goodman and McAuley are researching several broadcast PKC methods that can trade security for fewer packet hops (useful for data that need to be secure only for a short time). They are also looking at broadcast PKCs that use such network topology as ring local area networks, tree hierarchies, and layered communication systems. Goodman's PKC research also concerns the practical implementation (on VLSI chips) of particular algorithms involving the modular exponentiation of large (e.g. 512-bit) numbers.



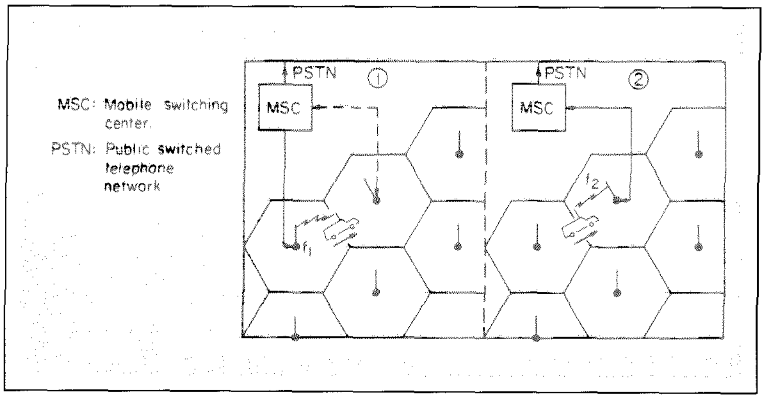
One-megabit RAM with on-chip error correction. The cost is extra chip area needed to store and select the check bits (shaded). The particular coding scheme shown, however, saves area by using the same column and row decode structure as needed in an uncoded chip.

More recently Goodman has been investigating the use of on-chip error correction required by the new ultra-large RAMs. Soft (non-permanent) errors occur so frequently (in terms of the high reliability one demands from computers) that the chip mean time between failure would be very low without error correction. In addition, coding to avoid hard (permanent) errors can be used to further increase lifetime and to increase yields without having to reconfigure spare memory rows by a process known as laser trimming. The main questions to be answered are: How can we implement powerful error-correcting codes on the chip, while using the minimum number of redundant (non-data) cells? And how can we make these systems fast so as to not degrade RAM access time? Goodman and colleagues have developed coding schemes that are intimately linked to the actual RAM structure and will be investigating these algorithms practically using the chip fabrication facilities of the digital communications lab that Goodman set up.

Another application of the digital communications lab is in space communications. A project is currently under way in collaboration with F. Pollara of JPL's Telecommunications Division to examine the applicability to deep space communications of a convolutional code decoding algorithm developed by Goodman. In order to send back high-quality pictures from the outer planets with the low-power spacecraft transmitters, the information is encoded on the spacecraft and decoded back on earth. Extremely low error probability is needed because video compression cannot tolerate errors. Long convolutional error-correcting codes are among the most powerful known but have been impractical to decode optimally. Goodman's algorithm uses the structure of these codes with a pre-computation to significantly improve decoding speeds and hence make decoding long codes a practical possibility.

Somewhat more down to earth is Posner's research on communications traffic and switching. New telephone services make many of the traditional models of communication traffic inapplicable, and advances in queueing theory are needed to model them. Some of the more popular of these new services are cellular radio, alarm reporting and monitoring, electronic messages with images and voice annotation, and interactive services such as home banking and games. In order to cope with the problems of accomodating large numbers of users in these new services, and because of the need to share system facilities efficiently, the development of intelligent systems and networks is essential. One of the most important problems in these systems is the automatic and intelligent allocation of a small number of communication links, channels, or paths among the relatively large number of users. This is indeed a main theme of research in switching and traffic.

Handoff in cellular radio. When a car moves from one cell to another, as determined by signal-strength measurements, the central mobile telephone switching office hands the call off to the new cell site at a new frequency. This is done without any user involvement. If there were no free frequency channel in the new cell (a rare event), the handoff would be blocked and the call disconnected.

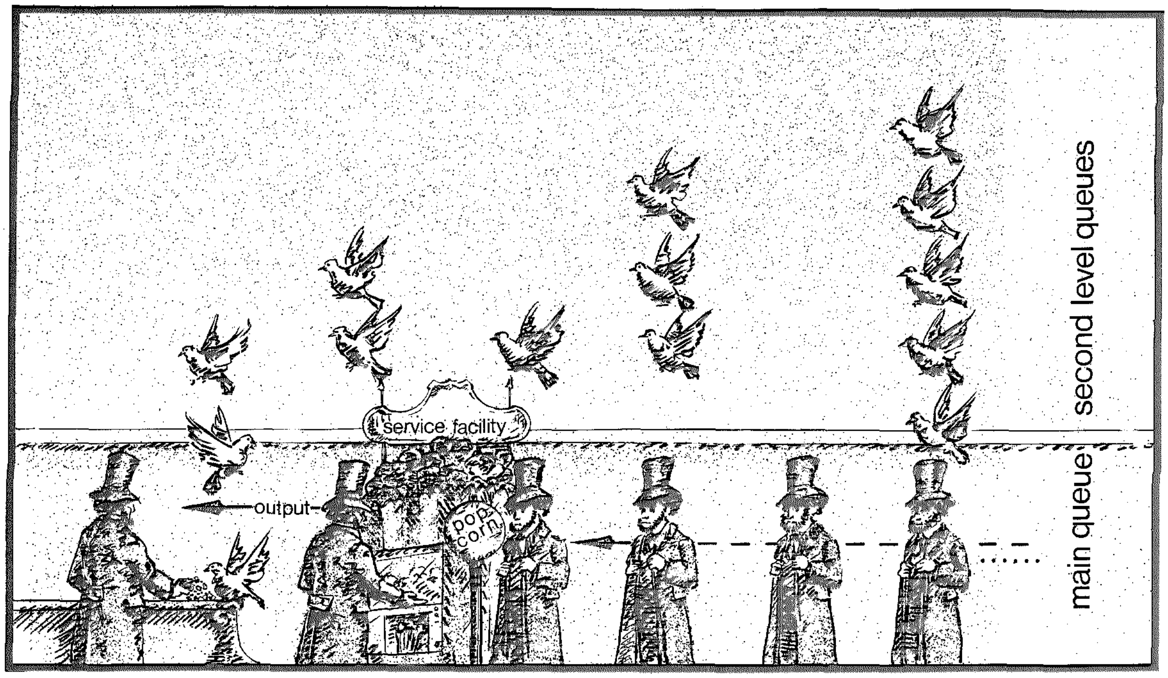

MSC: Mobile switching center.
PSTN: Public switched telephone network
PSTN
①
MSC
f1
PSTN
②
MSC
f2

Cellular radio, which has recently been successfully introduced in the greater Los Angeles area, enables a much larger number of subscribers than before to have mobile telephone service. In the past, communication with mobiles was achieved in a way similar to broadcast radio. A high-power transmitter is positioned at a high point (on Mt. Wilson for the bulk of the LA area) allowing coverage of virtually an entire extended metropolitan area. The transmitter would then serve any mobile wishing to initiate (or receive) a call from the public switched telephone network. The main problem of this approach is the small number of frequency channels available for the service — only 11 for all of Los Angeles. This means that as soon as 11 calls are taking place over the old system, which still exists, it would be saturated and no new call could be accepted. Long delays then ensued.

Cellular radio solves this problem by dividing the service area into a large number of smaller cells, currently with centers about 10 miles apart, each served by its own low-power cell-site transmitter. Due to the faster-than-inverse-square drop in signal power in the urban environment (because the propagation is not line-of-sight), a frequency channel used in a given cell can be reused in another one located fairly close-by while still avoiding co-channel interference (self-jamming). Each frequency channel can be reused dozens of times in the whole service area, so the total number of subscribers that can be simultaneously served increases by that factor. (Some of the increased capacity is due to the use of a higher UHF frequency not previously available to mobile radio.)

The problem of subscribers moving from one cell into another is solved by the handoff concept — disconnecting from the frequency channel of the old cell and switching to a frequency channel of the new cell. The whole process is unnoticed by the subscriber but requires a high level of automation and intelligence at a base switching office. Although the cellular concept was invented by AT&T in 1947, cellular radio first became practical only in the mid-1970s because of advances in integrated electronics. And VLSI makes it extremely practical and affordable now.

With the electronic circuit problems by now largely solved, most of the few remaining difficulties that cellular radio faces deal with traffic. One of the Posner group's first tasks in attacking the problem was to estimate the probability distribution of the occupation



Queue of servers. Each man here has his own loyal queue of pigeons waiting to be fed as soon as the feeders buy popcorn. The pigeons wait in turn to be fed on a first-come, first-served basis. Recent experimental work in front of Steele Laboratory has shown that real pigeons don't behave this way.

time of a frequency channel in a cellular system. This parameter is vital in determining the number of channels needed in each cell. It is also needed to derive policies that maximize the number of customers actually served. An analytical model was derived under simplified assumptions and a computer simulation written for the more general case. The main result of this work was to establish, both theoretically and by simulation, that in most practical situations the simplifying memoryless assumptions (negative exponential service time distribution) used in classical telephony could still be applied without too many modifications. This is an encouraging result because traffic theory is much simpler if we can forget about memory.

Other traffic problems Posner's group is investigating involve the protection of handoff calls from being blocked and disconnected when all new frequency channels are busy, and setting priorities in channel assignments when a mobile telephone system (a modification of the old type) shares the frequency spectrum with mobile dispatchers.

Queueing has always been an intimate part of traffic theory and becomes especially important for these new services. One familiar problem in service management (from supermarkets to banks) has to do with situations in which we are forced to wait for a certain facility shared by a large group of users to become available. Posner's group is considering an interesting case in which these customers themselves constitute service facilities. (They also serve, who only stand and wait.) The initial service facility can be considered as no more than one of the multiple subscribers in the network. That is, any one of them can provide service to other customers, as in the illustration above. Such a multi-level situation can arise in intelligent networks with a hierarchy of computing facilities or when data service providers must wait for data from remote data bases in order to provide service to their customers.

In the context of today's multiple options for residential or business telephone service — a far cry from plain old telephone service (POTS) — we can also think of this situation as a "camping" system. Camping means that a calling subscriber can wait for a busy subscriber to become free and then be automatically connected. When several levels of camping are allowed, what happens when a third customer tries to call any of the camping customers? Dealing with machines rather than people, we might be willing to tolerate some delay and wait for our call to be completed, instead of continuing to call and hoping that eventually we will reach the intended party. A new camping queue can be repeated for each new customer arriving, giving rise to a series of new queues served in an overall first-come-first-served basis. Once a unit or terminal becomes free, it will respond in sequence to all those calls waiting for it. Thus, any call can become a new service facility itself. The study also realistically allows reneging (leaving the queue).

A number of questions arise in this kind of problem: Is it possible for deadlock to occur, with everyone ending up waiting forever? And if deadlock can arise, how can it be avoided? How long are the delays? How do we minimize the average waiting time? Solutions for some of these problems have already been found for particular camping systems under the memoryless constraint for the customer arrival and service processes. In the memoryless case, closed-form solutions have been found for the steady-state probability distribution of the queue sizes, which in turn provide tools to understand this and other related traffic problems. And such solutions will be helping people to communicate better as the integrated services digital networks of the future become available.

In addition to those faculty members mentioned in the article, others whose work is allied to the communications group include Yaser Abu-Mostafa (PhD 84), assistant professor of electrical engineering and computer science; Eric Baum, research fellow in chemistry; Charles Elachi (MS 69, PhD 71) visiting lecturer in electrical engineering from JPL; Dale Harris, visiting lecturer in electrical engineering from Pacific Bell; Carver Mead (BS 56, MS 57, PhD 60), the Gordon and Betty Moore Professor of Computer Science; Vera Pless, visiting professor of mathematics from the University of Illinois-Chicago; Demetri Psaltis, associate professor of electrical engineering, and his student Santosh Venkatesh; Lawrence L. Rauch, recently retired professor of electrical engineering from the University of Michigan and JPL; David Rutledge, associate professor of electrical engineering; Laif Swanson, visiting lecturer in electrical engineering from JPL; P. P. Vaidyanathan, assistant professor of electrical engineering; Henricus van Tilborg, visiting professor of electrical engineering and computer science; and Richard Wilson, professor of mathematics, and his student, Pierre Baldi.

Graduate students involved in the communications group include Khaled Abdel-Ghaffer, Mark Bell, Mario Blaum (Phd 85), Li Fung Chang, Kar-Ming Cheung, Yurdaer Doǧnata, Roch Guérin, Enrique Hernândez, Eric Majani, Phil Merkey, Patrick Smyth, Kumar Swaminathan (PhD 86) and Doug Whiting (PhD 85). Research support has come from the Air Force Office of Scientific Research, AT&T Bell Laboratories, the Defense Advanced Research Projects Agency, Garrett Corp., IBM, NATO, Pacific Bell, the Pacific Telesis Foundation, and Caltech's Program in Advanced Technologies, whose communications company is GTE Laboratories. □