Telecommunications regulation is an important public policy consideration and is presently the topic of much debate. The author presented a technical viewpoint of this topic, particularly with regard to possible applications and how the data processing industry could be involved, during the Keynote Panel Session at IEEE Compcon '78 held on September 6, 1978. His presentation is printed here.

Computing and communications—A perspective of the evolving environment

by L. M. Branscomb

The United States is at a crucial policy juncture with respect to telecommunications regulation. In many areas of national life, debate about public policy is stimulated by problems or deficiencies in the nation's institutions. In this case, the debate is stimulated by a great spectrum of favorable opportunities. The question is, "How can the people best take advantage of these extraordinary opportunities?"

In discussing communications and data processing from a technical point of view, I want to emphasize the great variety of possibilities for serving people's needs and the highly innovative character of the work that remains to be done. I hope this discussion will demonstrate why we cannot expect to regulate into existence the specific forms of electronic services that will best satisfy the public's requirements. This is true because we do not know what all of these types of service may be. We do know their variety will be as great as is the variety of institutions and applications in America.

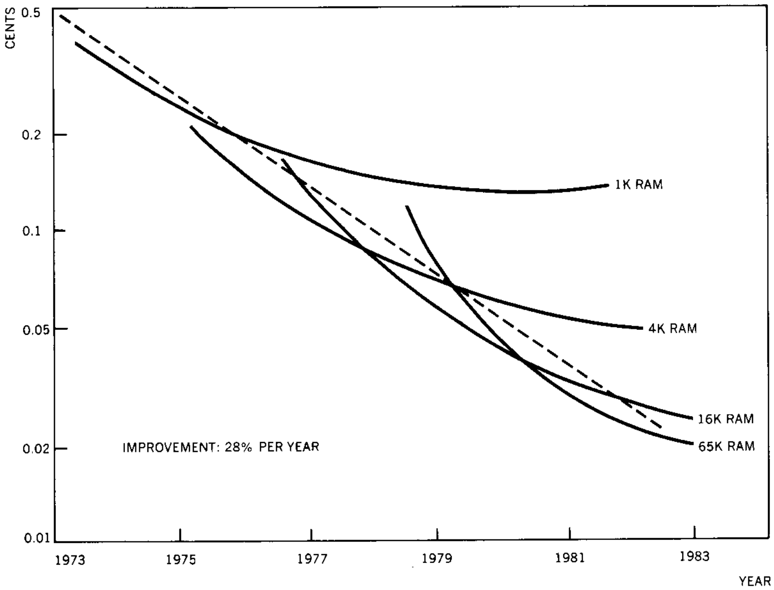
In any case, the development of distributed data processing, still in its early stages, has resulted in many complex technical communications issues being addressed. Thus, while one can specify the general characteristics of the basic telecommunications services that are needed, the details of network structures, provisions for network control, and kinds of application systems must remain open to innovation to ensure the most rapid possible solution of all the problems involved. The data processing industry today is healthy, innovative, and competitive. This innovation is spurred by research and development. *Business Week* magazine indicates that the data processing industry contributes 11 percent of all the privately financed industrial research and development in this country. Both the data processing and telecommunications industries have produced their share of innovations: satellites, optical fiber transmission, microprocessors, magnetic bubbles, and large-scale integration for example. But the richest source of innovation comes from the marriage of new technology and imaginative ideas about public needs.

There is a great variety of new applications, some already widespread, others beginning to be introduced, still others lying in the future. These innovations should be allowed to find their proper place in American life. Travel reservations, electronic banking, point-of-sale transactions, computer conferences, voice message storage and redistribution, facsimile distribution and image processing, interactive graphics, text editing and document distribution, electronic information services via television—are all examples of the rich possibilities.

Some people are impatient about extending the benefits of the technology now being pioneered by the most advanced users—banks, airlines, the offices in large organizations—to the average citizen of lesser means. They want to predict the technology, plan the applications, and then regulate them into existence. When Mao Tse Tung found that excessive centralization didn't work he adopted the policy, "Let 100 flowers bloom." The exciting new technologies, the applications which can be built upon them, and the benefits from these applications are flowers in the field. The policy environment is the fertilizer. None of us wants it to become the defoliant.

Extending the analogy a bit further, the American public, given the proper policy environment, should be allowed to wander through the field of possibilities picking the desirable flowers and leaving the remaining weeds to wilt. Market forces cannot accomplish all the things we need in our society. But letting the risk-taking innovators experiment with their own money, instead of having all innovation centrally planned and financed by regulated rates of return, is not all bad.

Figure 1 Projected improvement in cost per bit of random access memory at chip level (From "Microelectronics" by Robert N. Noyce. Copyright © 1977 by Scientific American, Inc. All rights reserved.)

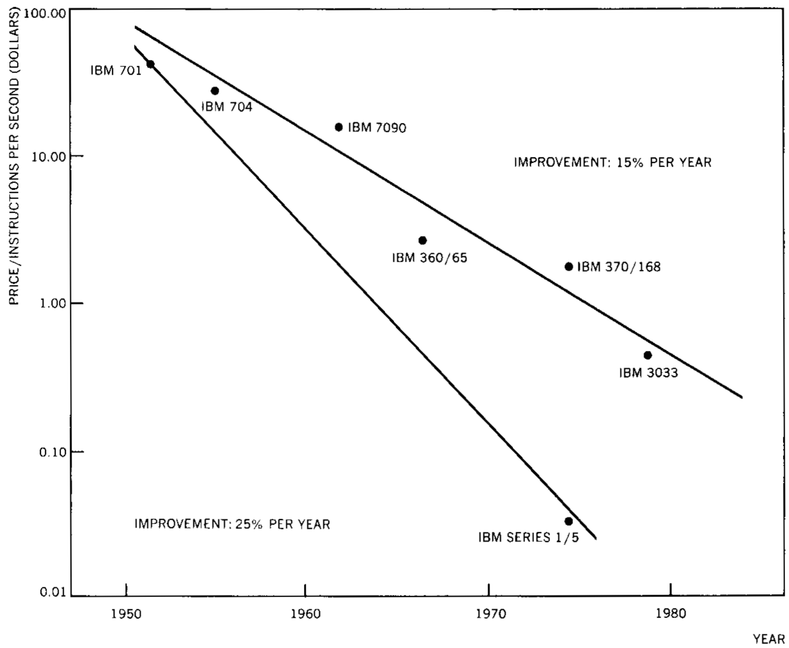


To understand the optimum policy environment for encouragement of innovation, let us take a look at the rate of innovation in the supporting electronic technologies that has been achieved by the data processing industry and the supporting semiconductor industry.

From the point of view of the scientist and the engineer, achieving the results projected in Figure 1 represents a very real challenge. A 28 percent per year improvement in the cost of random access memory bits at a chip level assumes that solutions for a vast set of multidisciplinary problems will be found. Advanced lithography systems must be developed, new surface physics and chemistry processes must be understood, the topological problems of chip layout and testing must be solved before the projected two hundredths of a cent per bit price can be reached by 1983. I have plenty of confidence in the creative ability of scientists and engineers. But one must also have confidence that the demand for memory bits will continue to rise faster than 28 percent per year. This will require a very high rate of application or market innovation to create the demand.

Figure 2 illustrates computer costs and performance history. The 28 percent per year improvement in memory costs, and the associated improvement in logic, cannot be translated directly to the cost of a computer system. That is, the improvement in packaging technologies, wiring capability, power supply costs, circuit de-

Figure 2 History of IBM computer cost/performance for large systems (upper curve) and systems of similar function



sign automation, and, importantly, system program development has all been quite rapid, but not at quite the same rate as the electronic circuits themselves. However, when one considers computers of roughly equivalent function, for example, the IBM 701 in 1950 and the IBM Series I Model 5 in 1975, the improvement is remarkably close to that 28 percent shown for the silicon chips.

The price per instruction per second, shown for large systems, declines at a still dramatic 15 percent per year, the difference being taken up in additional function: for example, much more real memory, dynamic address translation supporting virtual memories and relieving complex, time-consuming storage management tasks, and much richer operating systems.

The communications industry has also been a participant in and beneficiary of rapid technological progress. One example, adapted from the *Bell Laboratory Record*, is the relative cost per circuit mile for new terrestrial transmission systems as a function of the number of circuits carried; that is, the bandwidth of the technology. For paired cables with 10 circuits, the relative cost is over $200 per circuit-mile, whereas at the other end of the scale, for wave guide transmission, the investment cost per circuit can be well under a dollar per mile when 100 000 or more circuits are installed. Fiber optic transmission using lasers promises to overshadow even these improvements eventually. Thus new tech-

Table 1 Examples of emerging technologies

Base:	• Magnetic bubbles • Electron beam lithography • Josephson/GaAs devices
Systems:	• Computer networks • Distributed systems • Data security
Applications:	• Office systems • Facsimile and graphics • Speech filing

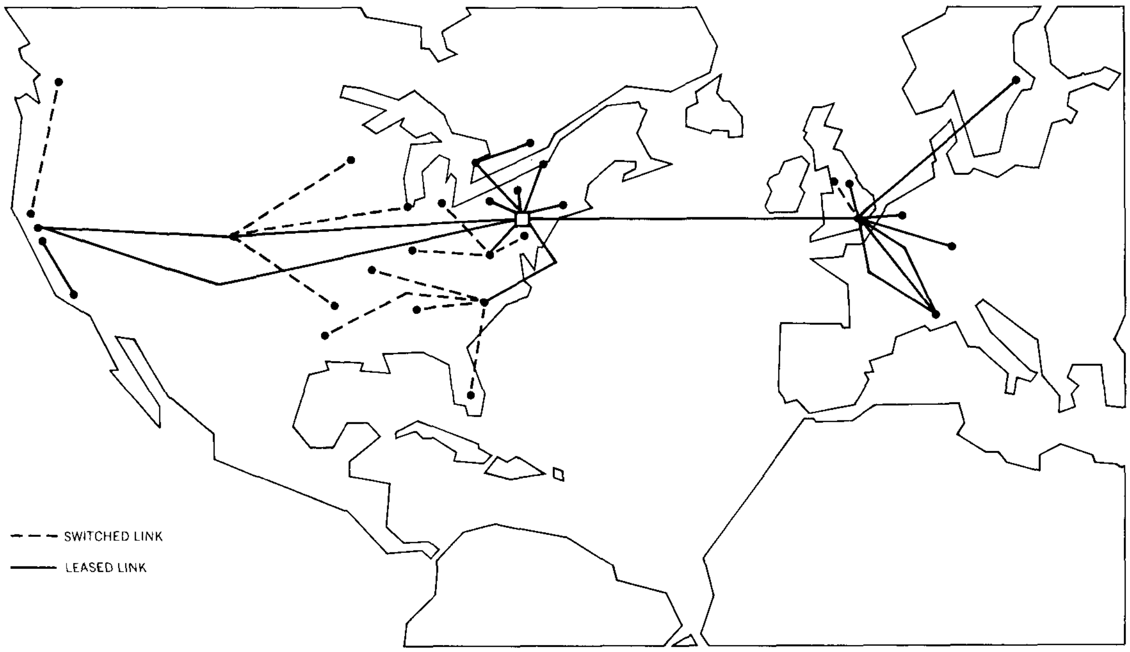
nologies, associated with rising demand for bandwidth, are improving communications costs at a generally accepted rate of about 11 percent per year.

Let me reemphasize then that the improvement in costs for a given function, demonstrated both in data processing and communications, is made possible only by rising demand for the function. Demand must rise faster than the capacity provided by increasingly productive technologies introduced successively. The introduction of such new technologies would be brought swiftly to a halt at any time that the policy environment inhibited the growth of the new applications that show us how to use the new capabilities.

I have discussed emerging base technologies for which three examples are given in the top group of Table 1. Continuation of this kind of progress in base technology certainly depends on advanced work in new devices, new material systems, and new tools, of which examples are given here. But two other kinds of technology are at least as important from the policy point of view. In the middle group of the table are three examples of systems technologies, which today represent the most sophisticated technical effort and involve advanced concepts in system design, architecture, and programming. I wish to go into systems technology questions in more depth. For completeness, let me also note the third group, which are examples of application technologies.

People are always asking, "In the world of the future, will information systems be centralized or distributed, maxi or mini, top down or bottom up?" My answer to all such questions is "yes." Information systems will become increasingly closely adapted to the structure of the organizations that use them and the information flow of the application they address. Even within a single user organization, there may be multiple systems having different structural characteristics or having subsystems which are structured differently. Let me show you two examples of quite dif-

Figure 3 IBM internal network serving scientific and engineering locations



ferent systems. These are both examples of IBM internal systems, since I am most familiar with them, but such examples abound throughout the world of distributed data processing.

Let's start with an example of an IBM internal network which is depicted in Figure 3 and provides message-switching services functionally somewhat similar to ARPANET (Advanced Research Projects Agency Network), although it grew in a completely spontaneous bottom-up kind of style. This network originated when two of our scientific centers had a lot of collaborative work to do and decided to interconnect their computing facilities. It grew slowly at first, then much more rapidly, until today it interconnects almost all major computers in our research and development facilities as well as many manufacturing and headquarters sites. There are now 254 processor nodes in 65 cities in 12 countries connected to this network, generally referred to as SUN (Subsystem Unified Network) internally. It evolved according to the organization of working affinity groups in IBM laboratories. Sister laboratories interconnected, related divisions tied in, and so on until almost all IBM scientific and engineering locations in the world can have access to it.

This network is not optimized for average traffic flow, with arbitrary sources of work at each node and minimum response time criteria, such as might be the case in a commercial, transaction-oriented network. The computing facilities reflect actual work

Table 2 Growth of IBM Consolidated Corporate Data Network (CCDN)

	1975	*1978*
Applications systems	4	14
Control units	475	660
Devices	4130	8200
Displays	2880	6000
Printers	1250	2200
MSG/min (peak periods)	3800	6800
Char/min (peak periods)	1.2×10^6	2.3×10^6
Network response time	3 sec	3 sec

flow involved while the network was growing. As functions were added, new resources were added to accomplish these functions. Work flows throughout the network without any central node being in control.

Message input, editing, and switching are major functions; job and data preparation at one node for remote job entry is also important. Cryptographic facilities are provided under user control for data security.

Although this network grew as a bottom-up, noncentrally managed network, there are now arising a number of centralized functions such as directory services, which one node or another has volunteered to provide. The network uses both switched and leased communication facilities. There are several nodes that share a mass store data facility. There are six different kinds of processors, with three different kinds of job entry methods involved. Network reliability as a whole is not crucial in this situation. Local performance is more important than message exchange performance, and it is very easy to add a new node or to take one off without disrupting anybody else.

Another internal IBM network is what we call the Consolidated Corporate Data Network, CCDN, which is totally different from the SUN network in philosophy. It is sophisticated and complex. It is organized top-down, with central control to support multiple applications, and is also growing very rapidly. Table 2 shows growth rates from 1975 through 1978, indicating how many major applications are running on this network. The number of such application systems grew from four to fourteen during this period. Each of these major application systems can be accessed by 8200 terminals. All are controlled under Systems Network Architecture. This gives the effect of what would be a 114 800 terminal network in a world where each terminal would be tied to one specific application system.

This network is used for order entry, order inquiry, intracompany message service, customer engineering education and other applications.

Today, there are over two million characters per minute flowing over this network in a peak period, or 1.2 billion characters a day. These applications are important to the continuity of IBM's business from minute to minute and hour to hour; the company depends upon the availability of the network. It also depends on the discipline of the applications and the audit trails that flow through them. The fact that the function is the same at every terminal is important in this kind of network, whereas it is not important in the loosely coupled job network I just described.

Many other kinds of networks have been designed. In particular, a number have been designed for specific industry applications, such as supermarket systems, retail point-of-sale systems, electronic banking, airline reservations and insurance systems. Each of these has its own unique characteristics, and they are different from one another in a number of ways.

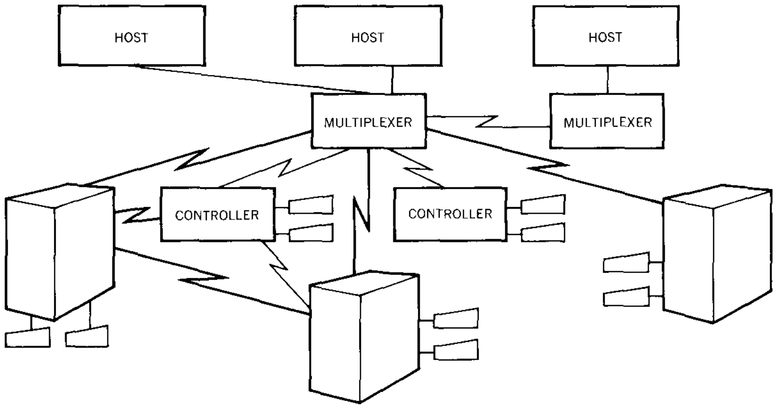
I want to emphasize that networks such as these, which include hundreds of computers and thousands of terminals, each with a specialized set of network characteristics to meet the user environment, are running today using leased and switched facilities—almost entirely analog voice—provided by common carriers. The complexity in these systems is in the computer programming that controls both the application environment, the data flow, and the management of communications traffic on the network.

High-quality digital transmission facilities will reduce errors and reduce costs for such networks. But the user must be free to tailor the network to his specific requirements, which suggests that these communications facilities be as simple and flexible as possible.

System technology for computer networking (Figure 4) is still changing rapidly and presents many new technical problems. For example, many application networks, which in the past were designed as centralized systems, are becoming decentralized, with control over regional subnetworks in a variety of local computers. This seems to be the trend of the future and presents a number of difficult problems which are being addressed in university and industry research laboratories around the world.

No clear definitions distinguish a distributed system from a non-distributed system, in spite of—or perhaps because of—an overwhelming amount of technical literature on the subject. I will not attempt to draw that line; instead, I will assert that a distributed system will usually have at least one of the following entities dis-

Figure 4 Interconnected, top-down, decentralized network illustrates one approach to distributed data processing

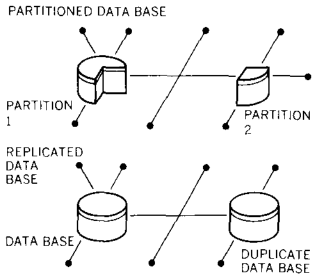


tributed: data, computing, or control. It also makes a lot of difference whether data and computation are shared or not shared. It is not difficult to distribute data among several nodes with each user having a specific node in which the data he is responsible for is located.

Techniques for transmitting pieces of a data set to a second node to create a new, modifiable and forever separate data set are also well known. This is not the case for true shared data bases, which can be updated and accessed simultaneously by programs in different locations, using procedures independent of the location of the data. These data bases are a prolific source of unsolved research topics.

All distributed systems that share data and computing have a common set of problems, which I call control problems. Examples are name and address identification, location of programs and data, access and telecommunications security, routing, flow control, program and data base translation. Many of these can be distributed; others are more naturally centralized.

Figure 5 Distributed data base can be either partitioned among network nodes (top) or replicated



Let's consider, for example, the question of how you control distributed data. There are various ways you can carve up data. Figure 5 shows two of them. One is to partition the data base and the other is to replicate it. Given that data are to be shared among programs that are distributed among nodes, when should you partition? When should you replicate? Or when is partially replicated data appropriate? An example of a partitioned data base would be the regional positive credit record which is updated on each transaction so that the user keeps the data for a particular region in that location and keeps them always up to date, accessing them when needed.

Table 3 Emerging data communications concepts and applications

• Fast circuit switched data networks
• Packet switched networks
• New, better data interfaces (e.g., X.21)
• Prestel and Bildschirmtext*
• Satellite services
• Interactive cable TV

*British and German, respectively, interactive home information services

An example of a replicated data base would be the negative credit file in the local store—a list of bad risks or stolen credit cards—which the user installs overnight, or once a week or whatever, and which stays relatively static and is accessed locally by the point-of-sale terminal when the central system is not available. For inquiry only, replicated data seem best. For update, partitioning seems best. Unfortunately, a lot of applications lie in between. Sometimes the application even needs to shift from one kind of distributed data to another between day and night.

Data base synchronization is also a challenge. How do you keep data consistent in the face of time delays? A satellite communication delay is a conspicuous example of such a time delay. In fact, however, satellite delays tend not to be large compared to delays that already exist in multinode networks which have a significant amount of queuing and processor time consumed in various nodes. So it is important to know when a data base may have been updated partially by one application and is being interrogated by another. It is also important that the inquirer know what the state of the data is at the time it is accessed. Demand assignment broadband satellite channels may help with this problem. A satellite has very high connectivity. Each earth station can see all the data, in a symbolic sense. Thus a pool of bandwidth can be made available to two or more stations on demand. You can imagine replicating a large data base and moving it, perhaps many millions of bytes, in a fraction of a second from one location to another.

The user needs to be able to trade off costs of transmission, costs of computing, response time, and other system characteristics to optimize his total application system, and not have his choices indirectly constrained by the carriers.

I have described a variety of different types of network architectures. I have discussed quite a variety of challenging—and to some extent still unsolved—technical problems involved in the more complex versions of distributed processing. But there is also a substantial variety of communication concepts and applications emerging. Table 3 shows a few of them. Some of these may find wide usage; some may not. For many, their usefulness will depend on the solutions to the kind of distributed systems research problems I have been discussing. For example, a packet network may be good for message switching or inquiry applications but may, nevertheless, be a very inefficient way to transmit large volumes of data or programs from one network node to another. Instead, a reliable, rapidly connecting, digital switched network may be far better. Thus it is appropriate to require each of these communications approaches to stand on its merits, competing among one another and against other similar ideas.

Finally, before leaving the discussion of technologies and applications, let me discuss for a moment the economic incentives for additional innovations. Earlier I indicated that data processing capability has been improving at a rate between 15 and 25 percent per year for some time. Similarly, communication costs have been decreasing at roughly 11 percent per year. The costs of personnel, however, rise with inflation at an average rate of around 6 percent per year.

International Data Corporation estimated that in 1975 total expenses of computing organizations in the U.S. were $26 billion, of which about half was for hardware, a third for professional salaries and about a fifth for communications. The productivity trends I just described motivate the user to seek more computing power and software, and when necessary, more communications capacity, in order to hold down the growth rate of his payroll costs. Most of the complexities in distributed data processing, which I have been discussing, are undertaken with this motivation in mind. Continuous innovation in distributed data processing can solve the difficult networking problems I have discussed and, given adequate communication facilities, can permit the productivity of the people building and using networks to rise along with the productivity trend of the electronic technology itself.

This is the real payoff from the opportunity for innovation, but it is not without risk. What then are the keystones of a public policy that both sustains the rate of innovation and provides the framework for needed national communications?

I am not a lawyer and not an expert on what the boundary of government regulation must be. I accept the notion that some form of basic digital transmission service must be provided by a regulated common carrier or carriers throughout the country. These facilities have been compared to an electronic highway. The electronic highway should restrict the freedom to innovate as little as possible, while providing a low-cost, low-error-rate, universally available service. The key characteristic of such a service is that it provides pure transmission, or transparency as it is often called. An example of a test for transparency, though perhaps not the most important one, is the ability of the communication service to carry encrypted data without the carrier having access to the key.

Given the existence of such pure transmission facilities from the common carriers, it should then be permitted for any entrepreneur to purchase these facilities, utilize them to support any end use application he has in mind, and to offer the entire application to the public without any form of regulatory impediment at all. The crucial policy issue here is the unregulated resale of communications service. I am aware that this has been an active subject of debate before the FCC, and the Commission has vacillated as to whether it has the authority to deregulate the resale of communications. This authority seems to me to be a major point that should be explicitly assured in the proposed revision of the 1934 Communications Act.

Whatever the balance of monopoly versus competitive service, the concept of "regulated competition" is to be viewed with great skepticism. If this self-contradictory concept is then to be avoided, a crisper distinction between regulated common carriage and unregulated, competitive services is required.

The answer, I feel, lies in the deregulation of value-added services. An excellent example with which to test this principle is at issue at this time. A service called Advanced Communication Service (ACS) has been proposed by AT&T to the FCC as a service using the Dataphone® Digital Service (DDS), available now in leased form. This is clearly an entrepreneurial venture which has substantial technical risk, moves far down the path toward the offering of data processing services by a monopoly carrier, and in any case, will cost the Bell System a very substantial amount of money to develop. The fact that it is to be built on DDS indicates that there is a clear boundary between what could be regulated and could be left unregulated. If all of the add-on facilities in ACS, beyond the Dataphone® Digital Service, were left totally unregulated and provided to the customer in a competitive marketplace, the value-added services provided by ACS could succeed or fail against other service alternatives. The cost of the transmission facility for ACS would then be the standard price that DDS charges all its customers. And anyone, carrier or not, should be able to offer an ACS-like service built on the resale of DDS facilities.

Unhappily, the AT&T Consent Decree does not permit AT&T to offer any service that is not regulated by the FCC. This restriction provides additional incentive to expand the regulatory reach of government to cover services which do not need to be regulated, in order to permit AT&T to bring its very substantial technical and financial resources to bear on the new innovative opportunities. In that sense, the antitrust Consent Decree has an anticompetitive effect, for if it were removed, AT&T would be able to organize itself to engage in unregulated competitive services, and thus compete on the same basis as others.

In summary, we have reached a critical policy juncture at a time when technologies, particularly systems and application technologies, are still evolving rapidly and indeed are not always well understood. Predicting technological success is not easy; predicting commercial success of new kinds of services is even less certain. Yet we know what we need in the way of basic national facilities, and we know that a policy of minimum regulation will be most conducive to the innovation that supports progress.

The policymakers need to ensure for the nation transparent data services of the highest quality. Such services should allow new applications of all sorts, including the resale of value-added functions, such as packet networks, on an unregulated basis. These transparent services should provide a homogeneous set of base functions for the user. They would be useful to the unregulated entrepreneur who wants to build a packet, facsimile, or mixed voice-data system. These facilities would also be useful to the user of a distributed data processing system with large volumes of high-speed files for programs to be transmitted from one node to another.

The challenge to the policymaker is to create the rules and institutional arrangements that will permit our best scientists and engineers to vie with one another in their attempts to solve end user problems at the lowest possible cost. It is the public consumer of these services that should determine which ones our society needs and which ones it does not. Marketing and engineering professionals should design them for a competitive marketplace, not for regulatory agencies and courts.

Dataphone is a registered trademark of AT&T.

Reprint Order No. G321-5092.