The Future of Fiber-Optic Computer Networks

Paul E. Green, IBM T.J. Watson Research Center

Fiber-optic networks have had a profound impact on information transmission over their first quarter-century. Orders of magnitude of further improvement are waiting to be accessed.

Fiber-optic communication is only 25 years old, but it would be difficult to exaggerate the impact it has already had on all branches of information transmission, from spans of a few meters to intercontinental distances. This strange technology, involving — of all things — the transmission of messages as pulses of light along an almost invisible thread of glass, is effectively taking over the role of guided transmission by copper and, to a modest extent, the role of unguided free-space radio and infrared transmission.

If the medium itself is not strange enough, reflect on the form taken by the transmitter — the semiconductor laser diode. This device is usually no larger than a grain of sand and transmits milliwatts of infrared light — enough to send information at gigabits per second over long distances with an extremely low received bit-error rate.

Yet, looking a little closer at how this supposedly revolutionary technology has been used to date, you come away with a sense that vast new unexploited opportunities await. The bandwidth is 10 orders of magnitude greater than that of phone lines (25,000 gigahertz versus 3 kilohertz), and the raw bit-error rate on the link is 10 orders of magnitude lower. Design points of 10^{-15} are in place today (1 GHz = 10^9 Hz).

When you combine these numbers with the great potential for cost reductions, it becomes clear that fiber is much more promising than anything that network architects have ever had to play with before and that a great deal more can be done with photonic communication than is being done today.

The classic example of cost reduction of photonic technology is the short-wavelength laser diode used in compact disc players. Who would have thought during experiments with the first large, bulky lasers with their cryogenic cooling and all the other complications that lasers would eventually cost $5 apiece and be more plentiful than phonograph needles?

You sometimes get the impression that the only real network applications that have been found for this remarkable medium are those that substitute fiber for copper within the framework of some existing architecture. The performance is improved and the cost goes down, but the system is basically what it was before fiber optics came along.

The physical-level topologies, the layer structure, the protocols within the

0018-9162/91/0900-0078$01.00 © 1991 IEEE

layers, and the network-control functions in use today are all directly derived from the heritage of either voice-grade telephone lines or local-area coaxial cable. For example, today's DLC frame sizes are usually dictated by constraints on frame buffers that were imposed long ago by telephone-line bit rates and bit-error rates.

A number of aggressive activities in various research laboratories concern what I will arbitrarily call here third-generation lightwave networks. The purpose of this article (condensed from an upcoming book[1]) is to speculate about where these activities are taking us. These networks are also sometimes called all-optical networks because the entire path between end nodes is passive and optical. There are no conversions back and forth between electronics and photonics, except at the ends of a connection.

Many groups are contributing to the current state of such networks.[2] The active programs that have contributed the most are those at British Telecom Research Laboratories (Martlesham Heath in England), Bell Laboratories (Crawford Hill, N.J.), Bell Communication Research (Morristown and Red Bank, N.J.), NTT Laboratories (Yokosuka, Japan), Heinrich Herz Institute (Berlin), and IBM Research (Hawthorne, N.Y., and Zurich, Switzerland).

Recently, the US Defense Dept.'s Advanced Research Projects Agency (DARPA) announced[3] that it would sponsor one or more "precompetitive, generic, industrial" consortia to speed development of all-optical networks.

Three generations

Three generations of physical-level technology can be defined:

(1) *Fiber not used.* Some time in the early 1980s, fiber-optic technology reached the point where it could easily be exploited for computer networks. Examples of LANs and MANs that developed before this period include Ethernets and 802.5 token rings, 22-gauge copper local-exchange connections, and CATV connections. First-generation WANs would include such telephone company-based packet networks as ARPAnet, IBM's SNA, Digital Equipment's DNA, OSI, and TCP/IP from the American ARPAnet community.

Glossary

ANSI	American National Standards Institute
ATM	Asynchronous time multiplexing
BISDN	Broadband Integrated Services Digital Network
CATV	Coaxial cable community antenna television
CD	Collision detect
CDMA	Code-division multiple access
CSMA	Carrier-sense multiaccess
DLC	Data link control
DNA	Digital Network Architecture (DEC)
DQDB	Dual queue, dual bus
ERP	Error-recovery protocols
FC	Flow control
FDDI	Fiber data distribution interface
HDTV	High-definition television
Hippi	High-performance parallel interface
LAN	Local area network
MAN	Metropolitan-area network
OSI	Open Systems Interconnection
PKT	Packetizing/depacketizing
RISC	Reduced instruction-set computer
SNA	System Network Architecture (IBM)
SONet	Synchronous optical network
TCP/IP	Transport connection protocol/Internet protocol
TDM	Time-division multiplex
TDMA	Time-division multiple access
WAN	Wide area network
WDMA	Wavelength-division multiple access

(2) *Fiber used in traditional architectures.* The outstanding example of this generation is the point-to-point intercity trunk that is improved by upgrading single-copper or microwave-radio connections to fiber connections. It allows you to use higher and higher speed electronic time-division multiplexing and demultiplexing to try to keep up with the ever-increasing traffic demands; it does so by exploiting the speed of laser diodes and other photonic components and the low fiber attenuation and dispersion.

Other examples of the second generation include such LANs and MANs as the FDDI ring and the 802.6 DQDB bus.[4] These architectures substitute fiber for copper, but make only minor changes in topology and protocol-layer content. The electronic front end of each node must still be fast enough to handle the bits from all (or most) of the N nodes in the entire network, as had been the case with first-generation networks.

(3) *Fiber used for its unique properties.* Here, you might ask, "How can the nonclassical properties of fibers be employed to meet the needs of emerging applications by taking entirely new approaches?" How, for example, can you use 10 orders of magnitude improvement in error rate and bandwidth?

Applications

Of course, the answer to this question depends on the problem you are trying to solve with the network. As we approach the end of the century, an entire new set of applications is emerging that appears certain to consume very large bandwidth, probably up to 1 gigabit per second per node. In some cases, this 1-Gbps figure means a sustained rather than a burst bit rate.

These applications are quite variegated and depend on whether the physical topology is simply a point-to-point *link*, a *point-to-multipoint* structure, or a full *network* (any-to-any addressability available).

The most prominent link application is the long-haul common-carrier trunk. As installed today, such trunks can handle more than 1 Gbps of voice and data traffic. The challenge we face concerns how to upgrade long-haul links for the day when each user will want access to 1 Gbps of capacity.

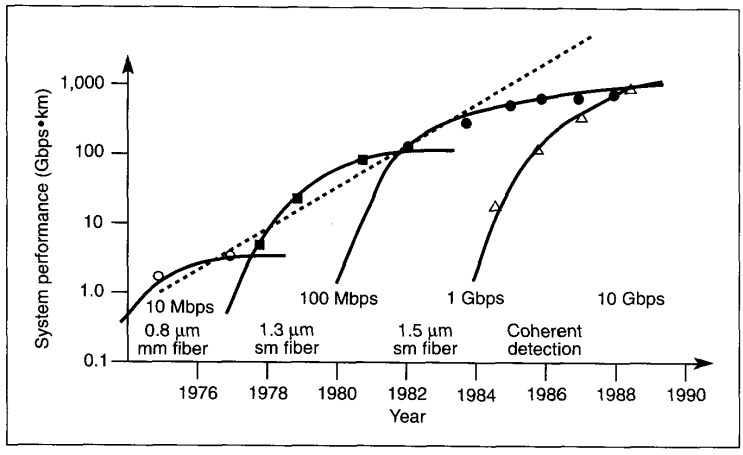


Figure 1. Progress in fiber-optic links, as measured by product of bit-rate times required inter-repeater distance.

To get a little more topologically complex, point-to-multipoint structures are expected to provide "fiber to the home" or "fiber to the office,"[5] carrying the multiple bit streams of HDTV, digital audio, and other emerging applications.

As for networks, the gigabit applications are appearing first in LAN and MAN environments. A network can be considered a topological structure, usually of more than one hierarchical level, that allows any node to reach any other node on a peer-to-peer basis.

There is already an urgent requirement[6] in university and other research environments for networks that will provide supercomputer support of hundreds or thousands of high-performance workstations with live-motion color graphics.

Medical imaging is another important application. The image-quality requirements are so exacting that the medical community is not sure it can trust image-compression techniques. Therefore, the images are usually transmitted in uncompressed form, and scanning one or two of these per second can generate bit rates approaching 1 Gbps.

Another pressing LAN/MAN application is the interconnection of computer mainframes to each other and to their storage peripherals within one center (the "glass house") and also the interconnection of glass houses. This requirement has led to the introduction of such second-generation systems as the ANSI 0.8-Gbps Hippi standard, the IBM Enterprise Connection at 0.2 Gbps,[7] and the current work in ANSI study group X3T9.3 to define a 1-Gbps serial standard (fiber channel). The two latter efforts are aimed at spanning MAN distances (tens of kilometers).

To satisfy all these emerging applications, it will be necessary to deal with limits on electronic speeds that appear to be effectively permanent. At 1-gigabit-per-node sustained demand, it is difficult to imagine that second-generation approaches will ever be sufficient to build a 1,000-node network. The 1,000-Gbps electronics that would be required to do this simply does not seem to be a realistic expectation.

The fiber links era

For most of the 25-year history of fiber-optic communication, the communication community has been preoccupied with sending the highest possible bit rate over a single link for the longest unrepeatered distance.[8] This has simply been a reflection of the priorities of the world's common carriers and their research laboratories. Not only has it been more in their interest to continue to upgrade the installed long-haul plant, but this has also been much easier to achieve — both technically and economically — than upgrading millions of individual copper subscriber loops.

Figure 1 shows the remarkable progress in bit-rate-distance product that was made over the past 15 years alone.[9] The mid-1970s' capability provided by multimode fiber driven by lasers operating in the near-infrared (0.85 micron wavelength) was soon superseded by that of single-mode fiber driven at 1.3-μ wavelength, where pulse smearing due to dispersion is small and attenuation is only 0.5 decibels per kilometer.

Systems at 1.5 μ, where this attenuation figure is only 0.2 dB per kilometer, soon followed. The introduction of coherent detection provided still more capability, because taking into account the frequency and phase of the light allows lower detectability limits to be achieved than can be achieved with much cruder direct detection.

In direct detection, the receiver registers only the amount of energy observed and ignores the sinusoidal character of the light. Recently, preoccupation with coherent detection has been challenged by the realization that error-rate performance that is almost as good can be achieved much more cheaply and simply by suitable use of direct detection working with photonic receiver preamplifiers. Such amplifiers have become truly commercially available only within the past year. Direct detection is very simple, because only the presence or absence of light is detected; phase, frequency, and polarization are ignored.

Until very recently, the single long-haul link has dominated thinking about lightwave communication. But now, an interesting change is taking place in the research agendas of the participating corporations. The emphasis is shifting from the single long-haul link to multipoints and networks. This is as true of component research as it is of system research.

Since long-haul transport, where bandwidth and distance are the imperatives, is in relatively good shape, attention is turning to providing connectivity rather than bandwidth, first over LAN distances and then over MAN distances. This shift is apparent in the research literature and in the advanced prototyping efforts of various leading-edge companies and carriers.[2]

Fiber paths in networks

Figure 2 is an attempt to capture the essential differences in the three network generations. In the first generation, all links are copper, and conversions between waveforms and bits take

place at every node along the path between end users.

In the second generation, all links are fiber, which improves the bit rate and bit-error rate but leaves propagation latency unchanged, as diagrammed in the figure by showing the links to be fatter but the same length as before. Conversion at each node still exists.

In third-generation networks, there are electrical-to-optical conversions only at the ends of a physical connection. Electronic bottlenecks stay out of the picture except at the ends.

The feasibility of all-optical communication at 1 Gbps per user is already being proven at LAN distances and to some extent for MANs, though it is only emerging slowly for WANs. This pattern matches the likely future availability of the dark fiber that will be required to build a third-generation network with a clear optical path connecting all N nodes.

The jargon term *dark fiber* refers to fiber onto which the user can directly attach his or her own optical terminations without having to go through someone else's electronic equipment with its restrictions on protocol, framing, bit rate, and so forth.

It is estimated that typically one-third of all the fiber that has been installed by the regional Bell operating companies, AT&T, and other interexchange carriers, and by the alternate-access carriers is dark.[10] This is not because third-generation networks were planned, but is the result of the usual practice of installing plenty of overcapacity for future growth.

When you realize that each fiber has a potential carrying capacity of 25,000 Gbps, equal to the volume of all US telephone calls at the peak busy hour (in the US, on Mother's Day), and that even those that are "lit" are using only one-tenth of 1 percent of this potential, it is hard not to wax enthusiastic about the possibilities.

Providing dark fiber is widely considered by the user community to be the responsibility of the common carriers, but the issue of whether they will readily do so is controversial. Instead of marketing capacity, the carriers traditionally prefer to market service. This involves interposing their own standardized TDM resources, whose main function is to support voice, not computer communication.

Today, only the alternate access vendors, serving mainly the larger American cities, are eager to provide dark fiber over tens of kilometers' distance.[10] Thus, before dark fiber is available nationally or internationally, we may see a repetition of the court battles of the 1970s over what constitutes a "transparent, highly usable" common-carrier offering. In fact, some litigation has already begun.[11] A more promising possibility is the growth of interest in providing dark fiber on the part of the alternate access carriers or the cable TV industry.

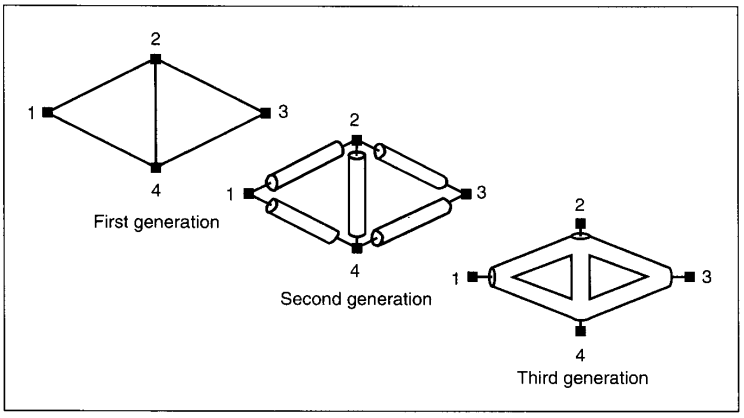


Figure 2. Schematic representation of the three generations for a toy network consisting of four nodes.

The limits on distance in third-generation networks do not appear to be technical. In particular, attenuation is not the problem that it used to be. Fiber attenuation is about 0.25 dB/km at 1.55 μ wavelength, and purely photonic amplifiers that have recently become commercially available have 20-30 dB of gain across several thousand gigahertz of bandwidth, while introducing only a modest amount of noise. Therefore, optical-to-electronic conversions involving traditional repeaters can be entirely avoided with technology available today.

In long, high-speed links, chromatic dispersion arises; the propagation velocity is slightly different across the signal bandwidth if the bit rate is too high. For third-generation LANs and MANs, this is unlikely to be the problem that it traditionally has been with long-haul telephone company links. This is because any single wavelength carries only the bit rate of one connection, not the aggregated bit rate of all connections as with second-generation links and networks.

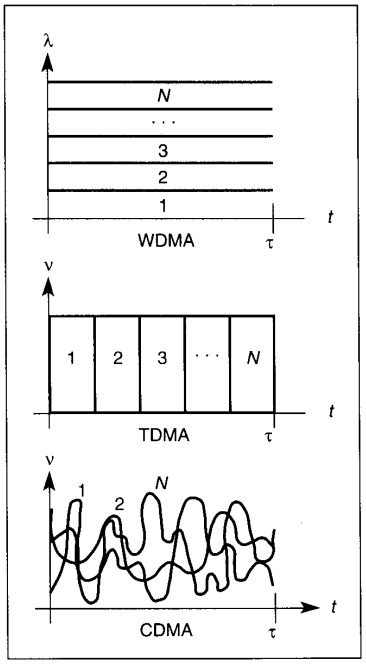


Figure 3. The three classes of addressing schemes: wavelength (frequency) division, time division, and code division (spread spectrum).

Forms of addressing

As Figure 3 shows, you can address messages from one node to another according to wavelength or frequency

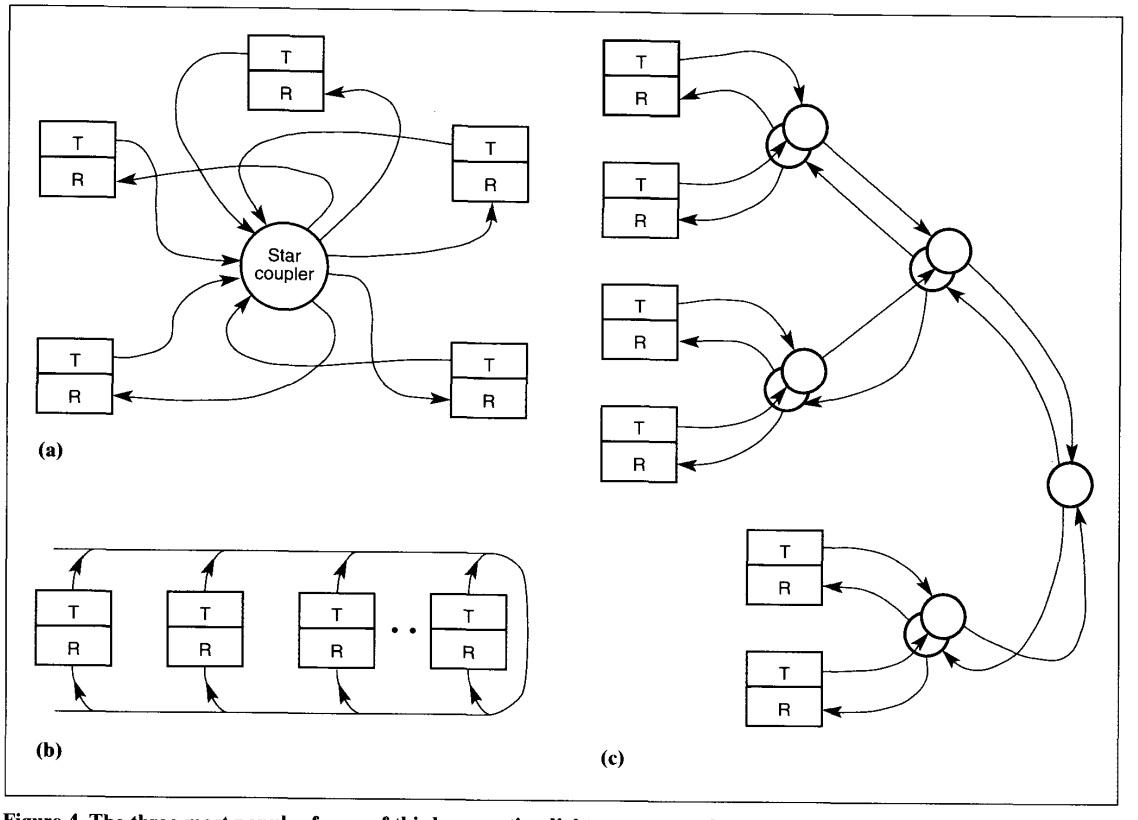


Figure 4. The three most popular forms of third-generation lightwave-network topology: (a) star connection, (b) reentrant bus, and (c) tree.

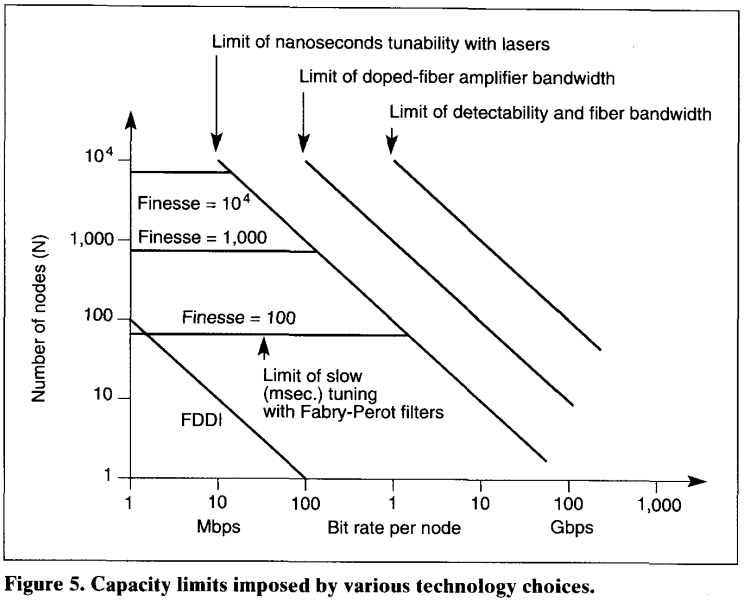


Figure 5. Capacity limits imposed by various technology choices.

(WDMA), time slot (TDMA), or waveshape (spread spectrum or CDMA). Of these, WDMA seems to be the current favorite.

One reason for this is that TDMA requires that there be many slots per bit time. CDMA goes even further by subdividing each bit time into hundreds or thousands of "chips" (waveform samples) per bit time. In fact, you can show that for about 10^{-10} bit-error rate, CDMA requires more than 40 chips per bit per node in the network.

All-optical TDMA and CDMA encoders and decoders are feasible enough, but when you think about 1,000-node networks at 1 Gbps each, the dispersion in the fiber, plus the requirement to synchronize to within one slot time (for TDMA) or one chip time (for CDMA), renders these two options relatively less attractive than WDMA, where nothing has to run any faster than the bit rate.

The space-division alternative

There is, of course, another way large numbers of nodes can be connected so that the electronics in each node runs no faster than that node's own bit rate. That way is to use a central space-division switch.[12]

Progress on large purely photonic (that is, third-generation) switches does not appear to be very promising so far.[13] Therefore, most current gigabit-per-port switch proposals involve the second-generation approach of converting from photons to electrons, doing the switching electronically with a high degree of parallelism (to minimize the amount of gigabit technology internal to the switch), and converting back to photons again.

Such switches might be developed rapidly and successfully, but the N-log-N internal connectivity complexity, plus the problem of call-queuing at a shared controller, could be important inhibitors. In addition, if the electronic speed bottleneck is to be avoided, the topology of the network is forced to be a single star with the switch at the center.

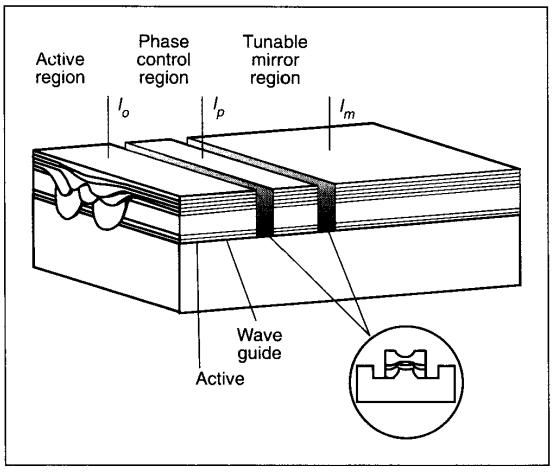


Figure 6. Experimental tunable laser diode (Nippon Electric Co.).

Overall network throughput capacity

Figure 4 shows — in its simplest form — what a WDMA all-optical network looks like. It can be a star connection to a central passive optical star coupler, a reentrant bus in which energy is taken from or added to a common bus by means of taps, or it can be a tree.

Work in the early 1980s on third-generation lightwave networking tended to emphasize the bus geometry. However, investigators soon realized that, since decibel loss accumulates linearly with the number of nodes along the bus but only logarithmically with a star, the latter would be preferable. Even the logarithmic loss is severe enough to constitute the principal limit on bit rate and number of nodes N in the network. For LANs or MANs, these losses are more important than any link attenuation that occurs due to network physical size.

Figure 5 shows the limit in number of nodes N and bit rate per node, assuming a star topology and a 10^{-10} bit-error rate. The rightmost diagonal line indicates the limit of 25,000-GHz bandwidth.

Interestingly, this limit on product of number of nodes and modulation bandwidth per node happens to be very close to that obtained by asking how many photons per bit will give a 10^{-10} bit-error rate.[14] (Commercial-grade photodetectors and 1-milliwatt laser diodes are assumed.) That is, first you take the number of photons per bit generated at the transmitter (divide the number of photons per second in 1 mW of power by the reciprocal of the bit rate) and split this across N nodes. At this stage, you can see how large N must get before the number of photons per bit becomes less than that required to develop a 10^{-10} bit-error rate. The answer is the same as that given by the fiber bandwidth.

The next diagonal line in Figure 5 shows the limit imposed by the finite bandwidth of today's commercially available photonic amplifiers. The third line shows the tuning limit of today's tunable laser diodes.

Technologies

If you plan to use wavelength (frequency) as the address parameter, it follows that some sort of tunability will be required — at each transmitter, at each receiver, or both. This requires either tunable lasers or tunable filters. The most convenient way to modulate digital information onto the laser light output is simply to turn the laser on and off ("on-off keying") with the two levels of the bit stream.

Lasers that will tune over the entire 200-nanometer wavelength range are clearly desirable, since they can provide transmitter tunability or receiver tunability — the latter by using the tunable laser as the local oscillator in a coherent receiver structure that resembles a radio receiver.

An even more important advantage of tunable lasers is that the speed of retuning the emitted wavelength of such a device can be as short as a few nanoseconds, since it is limited by device capacitance and carrier-transport times involving very tiny dimensions of less than a tenth of a millimeter. Unfortunately, the tunable laser diode, although the object of much research worldwide, still remains a laboratory device.

The currently favored design, the "three-section" laser diode (see Figure 6), is capable of tuning in nanoseconds, but only over about 9 nm (1,100 GHz); hence, the third diagonal line from the right in Figure 5, which gives the capacity limits for a network where nanosecond tuning times are a requirement. Such a network would use packet switching at gigabit rates, with the ability to re-address in a few hundred bit times appropriate to packet switching.

The three-section laser diode shown in Figure 6 operates by carefully controlling the currents injected into the three sections. The current into the leftmost section controls the intensity of the light emitted, and the other two currents control the way two resonances interact to define the radiated optical frequency.

The first of these resonances is a simple cavity resonance between two mirrors, the leftmost facet of the device and an effective mirror at the right formed by the tunable mirror region. The second resonance is due to the fact that the bottom of the tunable mirror region is

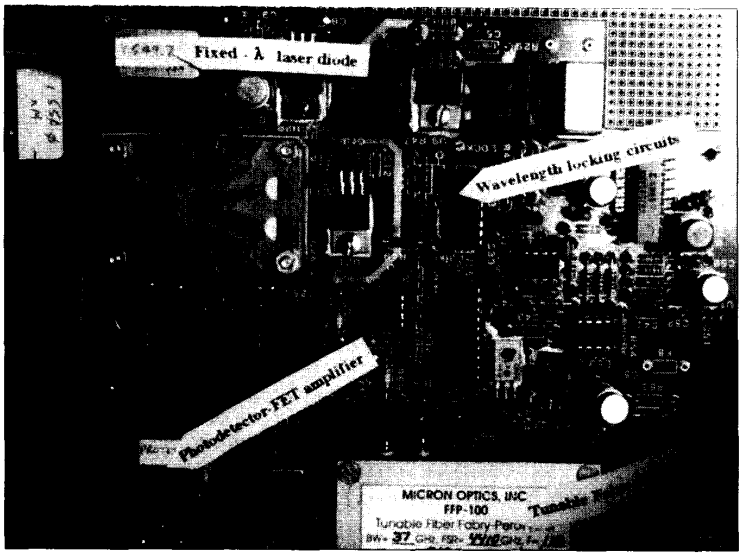


Figure 7. Optical portion of a plug-in card for an IBM PS/2 or RS-6000 computer, constituting one node in an all-optical network using wavelength addressing.

corrugated so as to form an optical grating whose resonant frequency depends on the refractive index of the material, which in turn depends on the current. This complex but widely understood device (and its descendents) represent today's best bet for nanosecond addressing speed in lightwave networks.

A technology commercially available today is the tunable Fabry-Perot filter, simply a resonant cavity consisting of two mirrors whose spacing can be tuned piezoelectrically. Such devices can be made tunable over almost any desired range, but the tuning speed is limited by mechanical inertia. The time to retune is at least hundreds of microseconds, which is fast enough for all circuit-switching applications, but not fast enough for packet switching of gigabit-per-second bit streams.

Commercially available Fabry-Perot filters have been used successfully in prototype third-generation networks. One of these networks is the Rainbow-I network developed at IBM Research. In Figure 7, an arrow points to a packaged commercial Fabry-Perot filter embodied in a Rainbow-I plug-in circuit card that runs on the Microchannel (system bus) of a PS/2 or RISC System 6000 workstation. In this network, up to 32 such nodes can be interconnected over 25-kilometer distances. Each node can operate at bit rates up to 300 Kbps.

In receivers using such devices, the light received from the central star coupler (and thus carrying all the bit streams at all the different wavelengths) is filtered in the Fabry-Perot device so as to pass only one node's transmission, and then is passed to a standard photodetector followed by low-noise amplifier stages and a threshold device.

The horizontal lines in Figure 5 show the capacity of networks using Fabry-Perot tunable filters. The limiting parameter is not tuning range, but the interchannel crosstalk that occurs when you place the channels too close together in frequency. Several different values of the maximum N (number of nodes) are shown for the finesse, or "Q," of these filters. Finesse values such as 100-200 are typical of such commercial devices as that of Figure 7.

Finesse is defined as the ratio of the spacing in optical frequency of the repetitive resonance peaks relative to their width. From this definition, it is clear that the amount of adjacent-channel crosstalk that leaks through to produce bit errors will be smaller when the finesse is higher. The finesse goes up with mirror parallelism, smoothness, and lossiness, and therefore the cost goes up with the finesse.

Several tricks can be exploited to make relatively inexpensive devices with finesses in the thousands, as would be required for networks with thousands of nodes. A number of researchers are also seeking ways to speed up tuning by electrically changing the index of the cavity material rather than the physical length of the cavity.

In the past year, purely photonic amplifiers having 20-30 dB of gain across many thousands of gigahertz of bandwidth have become commercially available. These use several meters of specially doped fiber that is spliced into the link at the appropriate point.

When the electrons in the dopant atoms are pumped to high energy levels by some 100 mW of light from a nearby small but powerful laser diode, any incident signal within the bandwidth of the device will be strongly amplified. Complete commercially available amplifiers that include the fiber, the pump-laser, and suitable couplers and isolators occupy only two or three cubic inches.

There are a number of techniques for locking the frequency of tunable devices to reference standards or to the wavelength of the incoming signal (as the circuitry in Figure 7 does).

Several of the approaches to third-generation network technology being addressed worldwide are so promising that experimental 1,000-node MANs (50 km in diameter) at 1 sustained gigabit per second per node will undoubtedly be built within the next three years.

Protocol layers before and after

Figure 8 presents a view of the protocol stack — the suite of architectural layers — for second- and third-generation lightwave networks. The left two diagrams represent first- and second-generation nodes, and the right two represent projected third-generation possibilities. The second and fourth diagrams indicate the traditional protocol layers:

- physical,
- multiaccess control (MAC),
- data link,
- network,
- transport,

- session,
- presentation, and
- application.

The activation, deactivation, and parameter setting for all the layers is presided over by the control point (usually an application), indicated by the shaded inverted **L** at the right of each layer diagram.

The content of each of the first four layers is dictated by the parameters of the network technology, and that of the upper three by the demands of the application user.

For WANs, the intermediate-node routing function (network layer) is important, but media access is simple, usually a matter of making a leased or dial-up connection (voice-grade dial-up, ISDN connection, and so forth). Conversely, for LANs and MANs, no routing decisions need be made, and the network layer is essentially absent, although the multiaccess protocol can become quite sophisticated.

With the arrival of third-generation networks, changes occur in the upper protocol levels, reflecting the session and presentation requirements of new applications. Some of the new applications are opened up by the great performance and ease-of-use advantages of third-generation networks.

The immediate changes, however, are occurring in the communication layers and in the control point.

Among the atomic protocol functions are packetizing/depacketizing, resequencing, splitting/merging, error recover, flow control, priority-class handling, and encryption/decryption. Various architectures handle them at various layers.

In the higher layers, such functions are intended to serve the needs of the application; in the lower layers, they cater to the peculiarities of the communication technology. As indicated in Figure 8, for first- and second-generation networks, many of these atomic functions were often repeated several times as the same data stream traversed the protocol stack on its way into or out of the node. Three of the atomic functions — error control, flow control, and packetizing/depacketizing — are illustrated alongside the left side of Figure 8, as done in a typical first-generation architecture (SNA).

The potential effect on this structure of introducing third-generation technologies and the corresponding protocols is to greatly simplify the communication layers, similarly simplify the control-point function, and execute many of the atomic functions only once on behalf of the application only. The diagram doesn't show the network layer, as is appropriate for a LAN or MAN, where there is only one hop between any two nodes (see Figure 4).

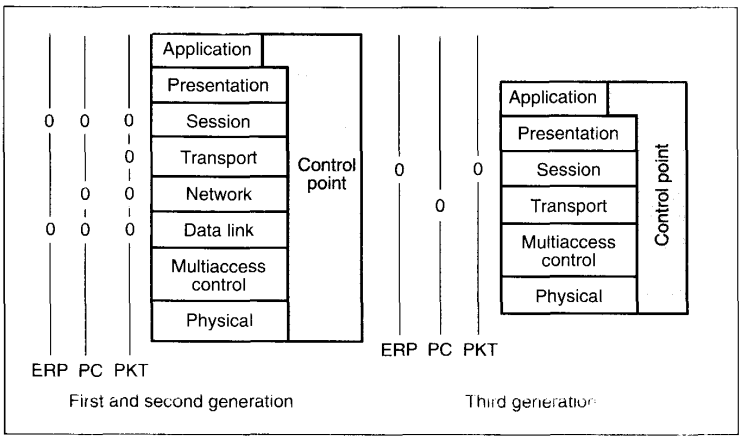


Figure 8. A representation of the potential for third-generation networks to realize several forms of improvement: (1) fewer and simpler protocol layers and their associated control point (layer diagrams at right), and (2) elimination of redundant instances of "atomic protocol functions" (circles at left). (ERP = error-recovery protocols, FC = flow control, and PKT = packetizing/depacketizing)

A number of lightweight protocols have already been implemented. These are all concerned with presenting, at the top of the transport layer, an interface that will have the greatest efficiency. This means at least

(1) removing redundant instances of the atomic functions,

(2) pruning the size and number of the finite-state machine representations of each individual protocol,

(3) designing toward an execution environment that makes the fewest possible memory moves and, if possible, no calls to the operating system, and

(4) implementing as much of the protocol stack as possible in hardware very near the physical port, without the system bus's intervening between the hardware and the node's CPU.

Sending the messages up and down through the protocol stack is a steady-state process in the sense that it takes place on a continuous basis as information is transmitted. Thus, it is extremely time-sensitive. (Note that in gigabit networks, even 1 millisecond of delay means 1 megabit of data delayed).

The network control functions, on the other hand, are also time-sensitive, although not so much so. A logical connection between application-layer entities at the two ends of the connection needs to be set up and taken down as fast as possible. Part of the lightweight protocol effort consists of pruning not only the steady-state processes within the layers but also the intermittent network-control functions within the control point (shaded inverted **L** in Figure 8).

Making the communication layers invisible

Work continues not only on making the protocol stack and its control more efficient, but also on making it serve the idiosyncrasies of the applications and hide from the application any dependencies on the underlying communica-

tion technology. Both jobs are made much easier by the new lightwave technology.

The agenda for network architects is to understand and exploit the revolutionary properties of optical fibers, particularly the 10 orders of magnitude greater bandwidth and smaller error rate, and to do this in the most effective way possible. With this in mind, I next discuss each of the communication layers.

Obviously, the LAN and MAN physical layer changes completely for the third generation. Commands entering the layer from the control point will lead to changing a wavelength, a time-division slot, or a CDMA waveform — most probably the first of these, if WDMA continues to be the format of choice.

WDMA third-generation lightwave networks that use circuit switching offer a high degree of protocol transparency, since different protocol stacks can be implemented on different physical-level connections at different wavelengths. This protocol transparency exists during the steady-state interval after the connection is set up and before it is taken down — just as with voice-grade phone lines.

The MAC layer changes completely. First-generation LAN protocols such as 802.3 carrier-sense multiaccess with collision detect (CSMA/CD) and 802.5 (token rings) assumed that the propagation time across the network was a small fraction of a packet duration. Waiting for a collision detection slot to expire or a token to arrive imposes an unacceptable performance hit as the network size grows larger than about 1 kilometer and the bit rate climbs above a few megabits.

In the second-generation FDDI and DQDB (802.6), some relief is obtained by allowing a limited amount of concurrency (more than one packet in progress somewhere on the network at any instant of time), but the multiple is not large — somewhere between 1 and 10. What is needed is for the amount of concurrency to approach N, the number of stations; this is what the all-optical technology provides.

Special circuit-switch and packet-switch protocols for third-generation networks are an active area of research.

First-order changes in the link layer and in the packetizing/depacketizing functions of the network layer are likely. Some argue that the achievable error rates of photonic technology (say 10^{-15}) are as good as the uncorrected error rates of existing DLCs and will be dominated by buffer overflow errors anyway, so why do any error detection and retransmission at all in third-generation networks?

Special circuit-switch and packet-switch protocols for third-generation networks are an active area of research.

Many applications — for example, certain graphics applications — do not require such heroic error-control in any event. Thus, it is possible to think of relegating all bit-error recovery to the transport layer or even the higher (application-driven) layers of the stack. If this proves workable, an essentially null link layer may be expected in the third generation.

For third-generation LANs and MANs that use one physical hop, the intermediate routing function of the network layer is no more necessary than it was for the first two generations. These are the cases in which nodes can reach each other directly rather than by going through intermediate nodes.

As for packetizing/depacketizing in this layer and the transport layer, the very low bit-error rate has some important consequences that recall the old arguments about the relative virtues of packet switching and circuit switching, and whether packet switching always has to mean short packets — for example, less than a thousand bytes.

With third-generation networks, packet size can be matched entirely to the application, including a packet length so long that you might normally consider it a circuit-switched connection.

If the application is credit-card checking, the packet length provided by the transport interface to the application instance can be that of the record, perhaps only a few hundred bytes. At the other extreme, if the application is to support an uninterrupted file transfer or a full-screen high-resolution workstation with continuous video refresh, the streaming type of service class can be provided in which a "packet" might be megabytes in length.

This should allow each of the many instances of the higher layers that sit on top of the transport layer to do its own error recovery and packetizing independently. Since all the instances would share the same lower layers through a common buffer, flow control at the transport layer is still required.

It is conceivable that eventually photonic technology will become so inexpensive that many physical ports per node would be feasible. In the case of WDMA, this means one fiber port per node, but many wavelengths at that port.

However, photonics cost levels are currently so high that it will not be easy to provide many physical ports per node using today's lightwave components, even though the fiber will often have more than enough bandwidth to serve all the ports in the network.

If costs ever drop low enough, the number of physical ports per node could grow until it becomes as large as the number of logical ports (communication-based application instances); that is, one protocol stack per application. You can compare this with today's situation in which the number of instances of the physical and data-link level (that is, the number of physical ports) is one or two orders of magnitude smaller than the number of instances of the higher layers.

Transport-level flow control to prevent shared-buffer overflow could then be done away with, since it might now take place entirely to serve individual applications. In today's packet-switching networks, buffers in the lower layers are not dedicated to individual applications but are instead dedicated to individual links that are shared across all applications.

Low-cost all-optical solutions will also mean that eventually this form of networking will be accessible to PCs and Macintoshes.

Network control

Imaginative simplifications should be possible in the control point, shown as the inverted **L** in Figure 8. Previously, network control consisted of the following steps (not necessarily in the or-

der listed) when some application required a logical connection to another application somewhere in the network, known only by name:

- The topology of the network changes when a new node or link joins or leaves (this new topology is made known to all nodes).
- The location of the named target application is learned by invoking some sort of directory function.
- The best route to the node containing the target application is determined.
- The logical connection (session) to that application is set up and this fact is confirmed at both ends.
- Use of the connection commences.

In principle, in a single one-hop third-generation network, where every node sees every other directly, as in Figure 4, all this could be collapsed into one exchange — an ask-and-go style of usage. The session request verb is broadcast and, if there is any node anywhere in the LAN or MAN containing an instance of the named resource, that node's address is returned in the session response and the connection is thus completed.

Interconnection of third-generation LANs and MANs

As mentioned previously, it is expected that third-generation networks will proliferate first as LANs, then as MANs, and later as WANs, partly because this is the order in which dark fiber will become freely available. The topic of extending third-generation LAN and MAN technology to wide area networking (for example, nationwide) is beginning to draw serious thought.

It appears that, for the foreseeable future, the only gigabit wide-area links that will be available are those second-generation point-to-point links that use specific TDM call-setup and framing conventions — for example, BISDN, SONet, ATM, and so forth.

Components are being developed that could be used for all-optical gateways between networks to do wavelength-swapping or wavelength-sensitive routing. However, these components are presently only at the research stage.

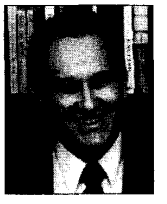
The National Research and Education Network is a testbed upon which many new ideas for gigabit networking could be tested, third-generation optical networks among them. The technology for interconnecting individual LAN or MAN "islands" is to use existing SONet and ATM long-haul transmission facilities, each having up to 2.49 Gbps aggregated capacity.

The state of today's lightwave technology is such that LANs and MANs providing 1-Gbps sustained capacity for each of 1,000 nodes will be practical within three to five years, certainly for circuit switching and probably for packet switching as well. If the opportunity is exploited with imagination, the positive effect on network architectures and on applications to be supported will be profound.

The public interest will best be served by enabling rather than inhibiting the diffusion of this exciting new generation of networking to the widest possible community of users. ■

References

1. P.E. Green, *Fiber-Optic Comm. Networks*, Prentice Hall, Englewood Cliffs, N.J., to be published in 1992.
2. Special issue on dense wavelength division for high-capacity and multiple-access communications systems, *IEEE J. Selected Areas in Comm.*, N.K. Cheung, K. Nosu, and G. Winzer, eds., Vol. 8, No. 6, Aug. 1990.
3. "Precompetitive Consortia for All-Optical Network Technology — SOL," DARPA Broad Agency Announcement 91-14, *Commerce Business Daily*, July 16, 1991.
4. R.M. Newman, Z.L. Budrikis, and J.L. Hullett, "The QPSX MAN," *IEEE Comm.*, Vol. 26, No. 4, Apr. 1988, pp 20-28.
5. P.W. Shumate, Jr., "Optical Fibers Reach into Homes," *IEEE Spectrum*, Vol. 26, No. 2, Feb. 1989, pp. 43-47.
6. Special issue on visualization in scientific computing," *Computer*, G.M. Nielson, ed., Vol. 22, No. 8, Aug. 1989.
7. J.C. Elliott and M.W. Sachs, "The IBM Enterprise Systems Connection Architecture," *IBM J. Research and Development*, to be published in 1992.
8. *Optical Fiber Communications-II*, S.E. Miller and I.P. Kaminow, eds., Academic Press, San Diego, Calif., 1988.
9. P.S. Henry, R.A. Linke, and A.H. Gnauch, "Introduction to Lightwave Systems," Chapter 21 of *Optical Fiber Telecommunications II*, S.E. Miller and I.P Kaminow, eds., Academic Press, San Diego, Calif, 1988.
10. Special issue on metropolitan-area networks, *Fiber Optics*, K. Kelly, ed., Information Gatekeepers Inc., Boston, Vol. 12, No. 2, 1990, pp. 1-32.
11. *FCC Dark Fiber Proc.*, CC Docket 88-136, Washington, DC, 1986.
12. J.Y. Hui, *Switching and Traffic Theory for Integrated Broadband Networks*, Kluwer, New York, 1990.
13. Special issue on photonic switching, *IEEE Comm.*, J. Baldini, ed., Vol. 25, No. 5, May 1987.
14. P.S. Henry, "High-Capacity Lightwave Local-Area Networks," *IEEE Comm.*, Vol. 27, No. 10, Oct. 1989, pp. 20-26.



Paul E. Green, Jr., is manager of advanced optical networking at the IBM T.J. Watson Research Center, Hawthorne, N.Y., where he has been employed since 1969. His research interests include applied communication theory and computer network architecture.

Green received degrees at the University of North Carolina and North Carolina State University and was awarded his PhD at the Massachusetts Institute of Technology in 1953. He is a member of the National Academy of Engineering and the IEEE Computer Society, is president-elect of the IEEE Communication Society, and has been an IEEE fellow since 1962. He received the IEEE Aerospace Society's Pioneer Award in 1981, the IEEE Communication Society's E.H. Armstrong Technical Achievement Award in 1989, and the IEEE Simon Ramo Medal in 1991. He was named a distinguished engineering alumnus of North Carolina State University in 1983.

Readers can write to the author at Rm. 3D54, IBM T.J. Watson Research Center, Hawthorne, NY 10532, or contact him by electronic mail at pgreen@watson.ibm.com.