Optical Satellite Networks

Vincent W. S. Chan, *Fellow, IEEE, Fellow, OSA*

Invited Paper

***Abstract*—With high-speed space optical crosslink being a reality, the construction of an optical satellite network as part of a larger integrated space-terrestrial network is now feasible. This paper explores the architecture implications of the invention of such a radical technology building block. Not only can the satellite network performance and cost undergo quantum-leap improvements but also such a network can have profound transforming effects on space system architectures and data network user applications.**

***Index Terms*—Optical space communications, satellite communications, satellite networks.**

I. INTRODUCTION

THIS paper builds on the premise that optical space communication at very high rates (>10 Gb/s) between satellites is now feasible [1]–[16]. It is reasonable to believe as more space packages are built and extensive on-orbit-operation experience develops in the next few years, the cost of high-rate optical crosslinks will be substantially lower than their microwave functional equivalent. A natural next step with such a powerful enabling technology is the realization of an optical satellite network of global extent. This optical satellite network can in turn revolutionalize space system architectures that may use the network as a critical subsystem. Examples of these space systems include those offering communication services or remote sensing. This paper summarizes briefly the state-of-the-art of optical crosslink technology, examines architectures (from the physical layer to the application layer) that combine other technologies to form an integrated space and terrestrial network, and finally explores the space of possible revolution in network performance and applications that are enabled by such a key technology innovation.

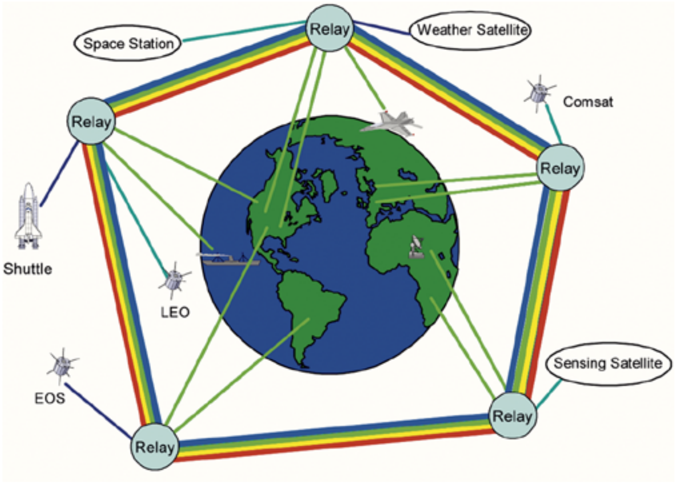


Fig. 1. General concept of an optical space network and its possible user community.

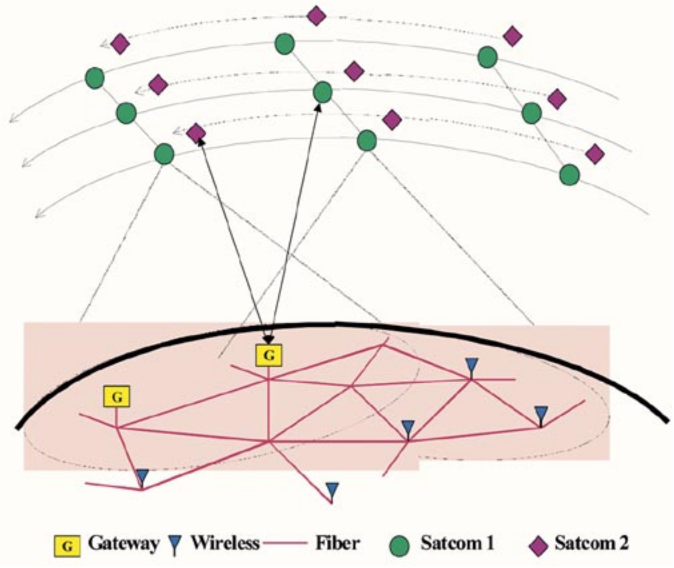


Fig. 2. Example of an integrated satellite and terrestrial network.

Fig. 1 illustrates the general concept of an optical space network and its possible user community. It can simultaneously serve a number of applications such as data readout of space sensors, support of space exploration, and science missions and act as the conduit for data communications for fixed ground terminals and mobile spaceborne, airborne, seaborne, and ground vehicles. The primary links through the atmosphere are assumed to be microwave due to poor optical propagation in bad weather conditions. However, we will describe technologies for special applications where optical downlinks from satellite to airborne and ground terminals are viable if absolute all-weather availability is not required. Note that this space backbone can be in high/geosynchronous (GEO; ~40 000 Km altitude), medium (MEO; ~5000–15 000 Km), or low (LEO; ~1000–2000 Km) earth orbits. Economics and applications will ultimately drive orbital choice and constellation configurations. Fig. 2 specifically gives an example of two LEO constellations interconnected by optical crosslinks and connected to terrestrial user ter-

---

Manuscript received February 5, 2003; revised August 19, 2003.

The author is with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science and Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: chan@mit.edu).

Digital Object Identifier 10.1109/JLT.2003.819534

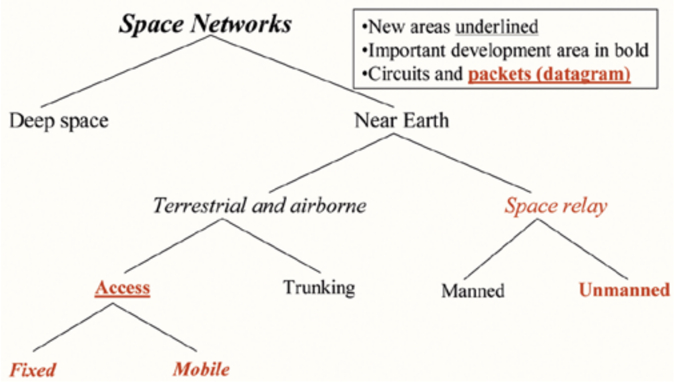


Fig. 3. New and important development area for space networking.

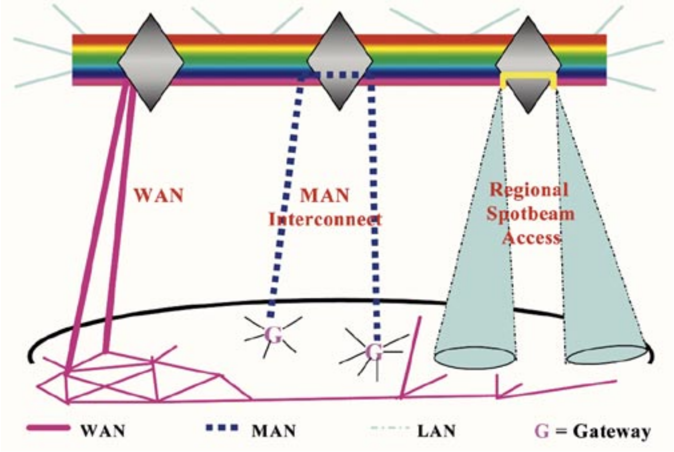


Fig. 4. Hierarchical space network architecture—wide/metro/local-area networks.

minals and interconnected with fiber and wireless networks via "gateways" forming a global heterogeneous network. Data networks based on such LEO, MEO, or GEO constellation topologies will be an important future alternative to provide global data networking services, especially in areas of poor or congested terrestrial infrastructure deployment, and in mobile and quick deployment application scenarios.

Space networking as an application area spans many different aspects of communication and relay services. Fig. 3 depicts these applications in a logical tree structure. The important forefronts as highlighted are data networking services for near-earth applications with emphasis on bursty computer traffic in both access and trunking. To serve a wide variety of usage, this network probably will have a hierarchical architecture, as shown in Fig. 4. Thus, the optical satellite network is a direct analog of the terrestrial network. The two wide-area networks (WANs) may interface at several gateway nodes, similar to interdomain connections in the current Internet. These interfaces will be via microwave links through the atmosphere or, when possible, use multiple optical links for diversity to increase availability and route around weather. A space network can also be used to interconnect metropolitan or local-area networks as a trunk service. This is an ongoing and viable market today in the business application of virtual private network interconnection via satellites. Perhaps the most aggressive application of a space network is for individual user bursty computer data access at high rates. This calls for radical changes in technology and architecture, as we will describe. To a large extent, this last application providing similar services of a classical local-area network (LAN) is not yet well explored but might well be the primary business area of the space network of the future. Finally, for space systems, it is interesting to examine the architecture implications of the advent of an economical and high-performance optical satellite network. Some consequences are mere linear extensions of current capabilities such as higher rates and cheaper services. However, the innovation can also lead to revolutionary space architectures as well as quantum leaps in performance, as we will describe.

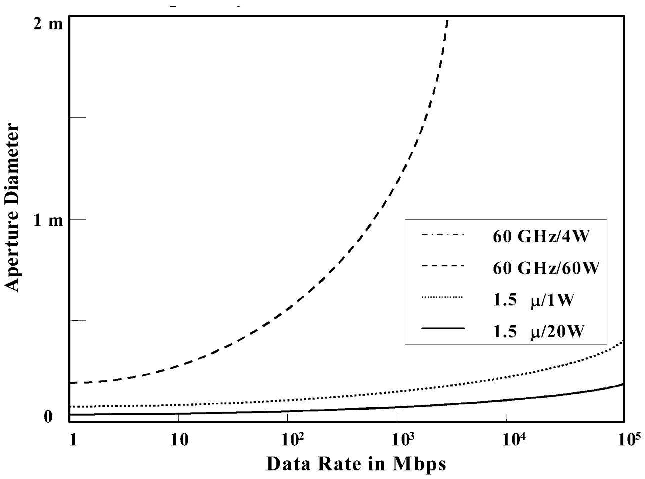


Fig. 5. Crosslink aperture size for RF and optical links with geosynchronous range.

II. Enabling Optical Crosslink Technology—An Overview

The first crosslinks are microwave systems. The geosynchronous-orbit MIT Lincoln Experimental Satellites (LES) 8 and 9, launched in 1976, were the first with a 38-GHz radio-frequency (RF) crosslink. Currently, the NASA Tracking and Data Relay Satellite System serves manned space exploration (e.g., space shuttle) and science experiments with both medium ($\sim$300 Mb/s) and low ($<$1 Mb/s) rate access links. The cost of integrating, launching, and deploying a communication antenna or telescope is a strong function of size and is a major design driver for satellite applications. This design issue is especially important for backbone relay satellites where there will be multiple apertures. Since the beam divergence of an RF or optical beam is roughly proportional to λ/D, where λ is the wavelength and D is the aperture diameter, optics have much higher antenna gains and can project the modest transmitter power into a smaller area at the receiving satellite, allowing much higher data rates. Fig. 5 compares crosslink aperture size for a link distance equal to one time synchronous orbit (44 000 Km), taking into account current and reasonable projection of transmitter power amplifier technologies. The crosslink cost function has a data-rate-independent term due to the need for spatial acquisition and tracking and an aperture-size-dependent term based on the cost of fabricating and integrating a telescope onto a spacecraft ($\sim$ diameter to the power of 2.6).

RF links require much larger aperture size than an optical link. Moreover, since the carrier frequencies of optics are very high (∼200 THz), each optical carrier can accommodate very high data rates (∼100 Gb/s) without the nasty dispersion effects of fiber, and there is the possibility of using wavelength-division multiplexing (WDM) to further increase the data rate per optical beam. There is no doubt that optical crosslink technology will greatly revolutionalize space system architectures.

In most applications, from space sensor readout to science and manned-space flight missions, the links are typically circuit-oriented (in the network sense). The timing for link setup and teardown is on the order of a second but no faster than a millisecond, which is not unlike the circuit setup requirement for the terrestrial optical networks being planned and deployed today. In this paper, until the discussion of optical multiple-access systems, we will treat the optical space communications problem to be one of providing high-speed circuit service with time-varying connectivity and setup and teardown times approximately the same as those for terrestrial fiber networks. We can summarize the distinguishing good properties of optical crosslinks with the following general characteristics, for high rates such as 10 Gb/s [16]:

1) small antenna sizes (<30 cm), compared to several feet for RF systems;
2) modest weight (∼100 lb) and power (∼100 W), compared to several hundred pounds and several hundred watts for RF systems;
3) continuous operations with the sun in or near the field of view;
4) easy multiplexing, demultiplexing, switching, and routing for network applications;
5) much lower cost than RF systems.

An optical space communication system is a truly “complex system” in every sense of the term, and especially in the engineering system context. The system has many high-precision subsystems that are intimately coupled and interacting (often not weakly). Not surprisingly, in many past instances, the system engineer could not adequately deal with these complex issues head-on in the creation of the architecture resulting in failures. This perhaps has been the biggest detractor of putting optical communications in space over the past decades. Though most of the failures usually manifest themselves in the shortcomings of particular subsystems, the culprit in actuality is due to the lack of balance and good engineering judgment in the design of the overall system.

A sensible approach to the design of a complex engineering system is to first try to understand the technology building blocks thoroughly and then break down the complex system into a small number of interacting logical subsystems. In this light, an optical space communication system can be partitioned into three interacting subsystems each with their separate critical design issues. These three subsystems are:

1) opto/mechanical/thermal subsystem;
2) spatial acquisition and tracking subsystem;
3) communication subsystem.

Note that these are logical partitions and not physical partitions. The subsystems, such as the tracking and optical subsystems, may share common physical hardware. Fig. 6 shows a high-level block diagram of an optical space communication system. The presence of tight coupling among subsystems is evident. We will briefly address the critical design issues and possible architecture solutions to these three subsystems.

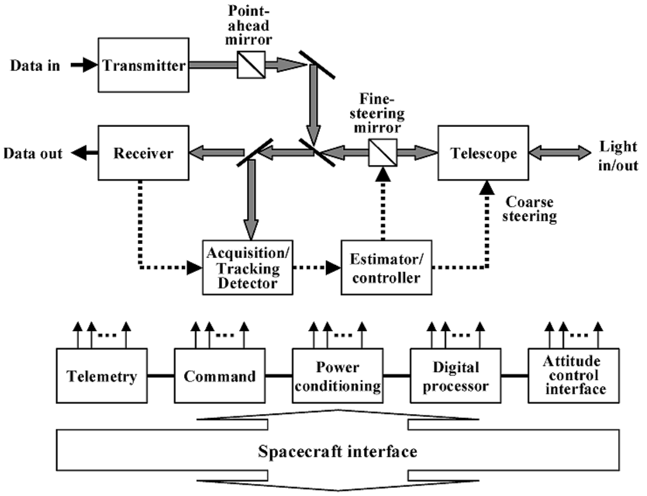


Fig. 6. High-level block diagram of an optical space communication system.

A. Opto/Mechanical/Thermal Subsystem

The opto/mechanical/thermal subsystem is perhaps the most difficult subsystem to design properly and yet also must be weight and power efficient. The housekeeping subsystems (bottom of Fig. 6), telemetry, command, power conditioning, digital processor, and attitude control system interfaces are conventional units and do not require any unusual treatment in opto/mechanical/thermal engineering. The optical part of the subsystem design however must minimize optical throughput losses (∼3 dB), provide high wavefront quality ($\sim\lambda/10$) under all operational scenarios, and maintain accurate beam pointing and alignment (∼1/20 beamwidth). The subsystem must also survive harsh launch loads and its performance must be maintained over on-orbit thermal environment and in the presence of mechanical disturbances from the spacecraft. A combination of clever mechanical engineering and system techniques must be used to arrive at a lightweight design. Submicrometer beam alignment is probably the most difficult requirement to meet through launch and on-orbit. After many failed approaches, the main principle that is now generally used to significantly lighten the structure is [16]:

> “Optically lock onto a beacon from the receiving platform, via sensors and fast steering mirrors and design, as much as possible, a common optical path for the transmit and receive beams while allowing the structure to flex with the mechanical disturbances and thermal distortions.”

After a high-level mechanical design is generated, detailed mechanical analysis is done via a “finite element” model approximating the system with a few thousand point-weight elements connected by “springs and dampers” of properties derived from the mechanical design of the components and materials used. A thermal and optical model should also be integrated with

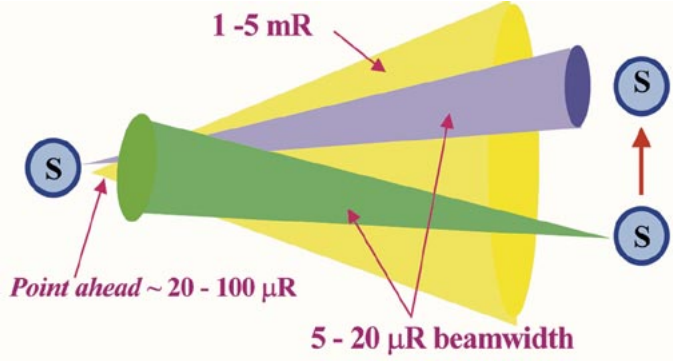


Fig. 7. Spatial acquisition geometry.

this mechanical model for the analysis of launch survivability and optical alignment integrity during operation. Refer to [16] for a more detailed description of this area. With advanced modulation and coding, great gains in the communication subsystem performance can be used to vastly improve the opto/mechanical/thermal subsystem. The biggest "quantum" jump will be to reduce the telescope to a small enough size ($\sim$10 cm) so that one can steer it with high enough speeds ($\sim$500 Hz) to track out all the disturbances, and eliminate the second high speed steering mirror (see Figs. 6 and 8). This substantially reduces the component count (such as a set of pupil relay optics for the fast steering mirror) of the optical train, and the smaller telescope yields easier, lighter, less power-consuming (for thermal control), and cheaper designs.

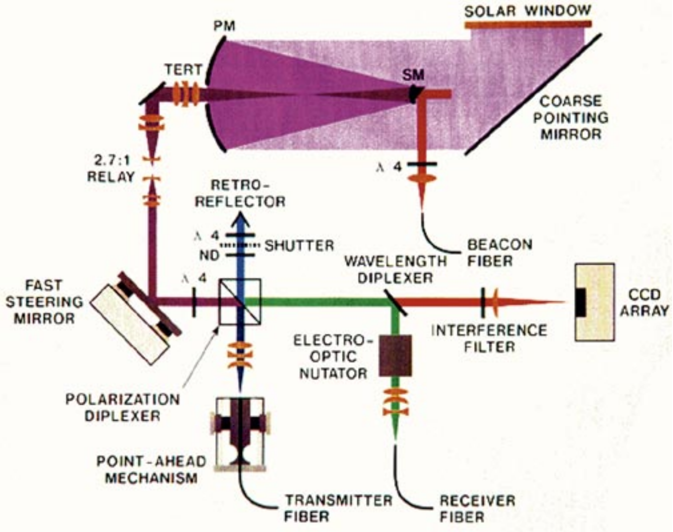


Fig. 8. Optical subsystem.

B. Spatial Acquisition and Tracking Subsystem

A typical optical space communication system points its transmit beam by tracking a beacon from the receiving satellite and then dial in a point ahead angle. Before this happens, each terminal must acquire the other satellites transmit beam or beacon by performing a spatial search of its angular uncertainty range to locate the other satellite. This spatial uncertainty is usually dominated by platform attitude control errors ($\sim$1 mR). The nature of this error can range from being a Gaussian random process, as contributed by a noisy earth or star spatial sensor, or a random amplitude and phase sinusoid, as induced by the movement of the solar array drive seen through the mechanical resonances of the platform. At low frequencies, the uncertainties can be tens of beamwidths wide. Fig. 7 illustrates the spatial acquisition geometry. The number of spatial cells that need to be searched is typically between 10^4 and 10^6. Due to the many mechanical disturbances that can yield slow drifts, it is prudent to try to acquire within the timescale of less than one second. Generally, acquisition strategies can be classified into serial and parallel searches, and there are many hybrids in between such as zooming. These strategies can be used for both illuminations by the beacon and search by the acquisition receiver. After acquisition, the system will enter the coarse and then fine tracking phases. Isolators are used to dampen jitter from the satellite platform as much as possible. The spatial error sensor detects the beacon and derives an estimate of the angle of arrival of the beacon. The signal is sent to a feedback controller, and any skew is tracked out by means of a slow outer loop ($\sim$10 Hz) with the telescope coarse pointing mirror and a fast inner loop ($\sim$1 KHz) with the high-speed fine tracking mirror. The transmit beam ($\sim$10 μR) shares the same optical train as the beacon, and thus transmit beam jitters are reciprocally tracked out by the steering mirrors (Fig. 8). A point-ahead angle ($\sim$20–60 μR) is dialed into the outgoing beam to compensate for the finite speed of light and receiver platform movements. The communication transmitter and receiver can be decoupled and remoted via the use of fiber couplers.

If one takes the simplifying assumption of the input disturbances as Gaussian and stationary plus a few random amplitude and phase sinusoids due to mechanical jitters, it is easy to analyze tracking performance using a stochastic control model. A linearized model can be used for the system under excellent tracking conditions (which is the usual condition for nominal link operations). A linear time-invariant system treatment can be used for stationary inputs such as nominal spacecraft jitters and spatial sensor noise, but the optimum time-varying linear (Kalman–Bucy) filter should be used for nonstationary inputs, such as control jets firing and sudden disturbances due to other spacecraft payload's nonperiodic movements. To predict capture and especially loss of lock behavior, nonlinear modeling should be used. It is not prudent to operate the link at tracking errors of more than 1/10 beamwidths. The linear and stationary noise model is only a gross approximation. Tracking errors resulting in, say, $>$3 dB average power loss will make the communication system exhibit long durations ($\sim$milliseconds) of poor bit error performance that will be extremely difficult to recover using communication systems techniques such as interleaving and error-correction coding. Current spatial tracking subsystem designs tend to be too conservative using linear system and stationary stochastic processes models, with plenty of margin to guard against model inaccuracies. This results in unnecessarily high structural weight for rigidity and unneeded high-speed (power consuming) mirrors. Moving to a) broader beams (as a consequence of an increase in communication performance and higher power amplifiers), b) use of more accurate

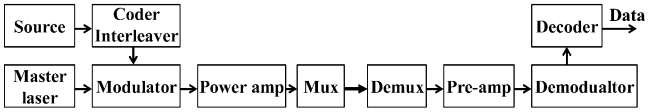


Fig. 9. Communication subsystem.

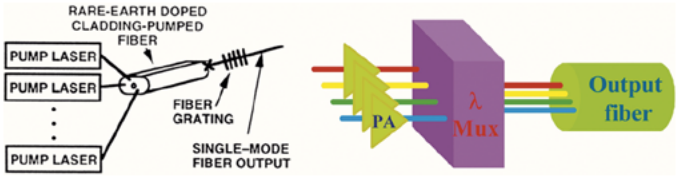


Fig. 10. Double-clad erbium amplifier and multiplexer.

nonstaionary characterization of disturbances, c) time-varying linear filtering at good quiescent tracking conditions, d) full nonlinear stochastic control system models at capture and loss of lock, and e) the additional use of hybrid digital hardware and software to realize these systems will substantially improve tracking performance and allow the design of the subsystem with much less weight and power (refer to [16] for a more detailed discussion of this area).

C. Communications Subsystem

The communication subsystem can be the Achilles' heel and source of great system performance enhancement at the same time. A typical block diagram for a space optical communication system operating at multigigabits/second is given in Fig. 9. Note the need for an optical power amplifier since the transmitted beam has to traverse a long distance with no intermediate boost-amplifier. Fortunately, unlike fibers, the vacuum medium does not have nonlinearities at these power levels (1–10 W) to degrade communication performance. Commercially available fiber communication components can be used for reliability and assured source of supply. Since putting midspan amplifiers in space is expensive, one needs the communication performance to be as close to the fundamental limit of quantum detection as possible. With these factors in mind, 1.5 μm is a sensible choice for the operating wavelength. There is a wealth of components available at this wavelengths, and there is also the high likelihood of continuous improvements driven by commercial system developments. The European Space Agency's SILEX system,[1] operating at 0.8 μm using an avalanche photodetector (APD) receiver at $\sim$50 Mb/s, represents an interesting alternative. Its subsystem design principles are similar to those described here.

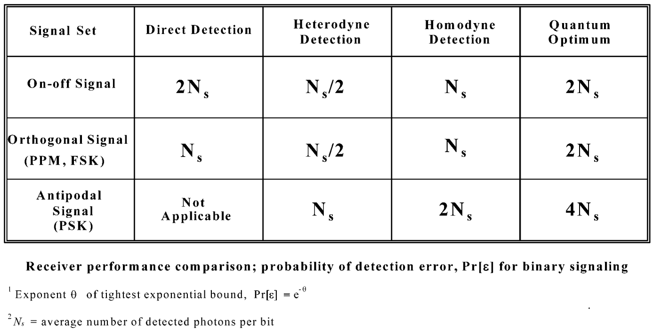
At the transmitter, in addition to picking a modulation scheme with a high theoretical efficiency, the design also has to use a configuration that maximizes hardware performance. Thus a low-power master laser is used to give the optical signal low phase and amplitude noise and good frequency control. A separate external modulator is used to reduce chirp and allow higher bandwidth modulation, reducing crosstalk. A 1–20 W optical power amplifier is used to reduce telescope size. Fig. 10 shows a design for very high power erbium-doped fiber amplifiers (EDFAs) by using a double clad fiber. The outer core is multimode and can be doped by an intermediary material such as ytterbium (Yb). Multiple semiconductor laser pumps at 0.8–0.9 μm are used to pump the cladding. Energy will then transfer from the Yb to the erbium ion in the single mode inner core. This configuration allows more pumps to be end-coupled or side-coupled into the multimode outer core. It is with this configuration that 20-W output power have been achieved [14]. With a 10-W amplifier, one can close a 40-Gb/s link over 44 000-Km distance with 10-cm apertures. Higher data rates can be realized by using WDM. The WDM coupler must be placed after the power amplifiers to benefit fully the additional output power due to multiplexing of multiple saturated amplfiers, making a low-loss ($<$1 dB) high-power-handling WDM coupler a critical component. With WDM, the cost of the optical crosslink, to first order, scales linearly with the data rate until the multiplexing loss becomes significant enough that telescope sizes have to be increased to compensate for the loss.

The faint optical fields at the receiving satellite exhibit significant quantum behavior (i.e., the energy levels are so small that a quantum system treatment of the electromagnetic field is necessary), and hence the optimum receiver is necessarily quantum-limited in nature. Quantum optimum receivers are often unrealizable with known techniques or their implementations, even if known, are very complicated [16]. Thus, simple receiver realizations, called "structured receivers," are used as near optimum compromises. The detection schemes used can be incoherent (direct) detection or coherent (heterodyne or homodyne) detection. In direct detection receivers, the received optical field is energy detected by means of a photodetector that usually provides gain. Examples are APDs, PIN-FET receivers, or photomultiplier tubes. Modulation schemes for direct detection systems are limited to intensity modulations such as on–off signaling and pulse position modulation. Phase modulations can be detected by using optical interferometry to convert the phase modulation to intensity modulation first. The ideal form of direct detection is a photon counting receiver, i.e., a receiver with enough electrical gain per photoelectron emitted by the detector surface such that individual photoevents are detected, timed, and counted by subsequent electronics. The direct detection photon-counting receiver can achieve theoretical optimum performance at a bit error rate of 10^{-12} of 28 detected photons per bit. Current state-of-the-art APD or PIN-FET receivers are 10–15 dB away, and the sensitivity will become progressively worse as the data rate increases due to the higher noise generated with a smaller resistive load at the output of the detectors. A new generation of APDs [17] can possibly achieve close to the photon counting receiver limit and may even be able to resolve individual photoevents. With a suitable low-noise optically preamplified direct detection receiver, many of the lost decibels can be recovered. This receiver has the same quantum mechanical model and detection limit as a heterodyne receiver Fig. 11. Near quantum-limited performance has also been achieved via the use of heterodyne detection [2]. Though the two types of receivers have similar performance, the EDFA preamplified direct detection receiver does lend itself to easier implementation.

In coherent detection, an optical local oscillator field is added to the received optical field and the sum is detected by a photodetector. The resulting signal is further processed at baseband (homodyne detection) or at an intermediate

---

[1] http://telecom.estec.esa.nl/telecom



Signal Set	Direct Detection	Heterodyne Detection	Homodyne Detection	Quantum Optimum
On-off Signal	$2N_s$	$N_s/2$	N_s	$2N_s$
Orthogonal Signal (PPM, FSK)	N_s	$N_s/2$	N_s	$2N_s$
Antipodal Signal (PSK)	Not Applicable	N_s	$2N_s$	$4N_s$

Receiver performance comparison; probability of detection error, Pr[ε] for binary signaling

[1] Exponent θ of tightest exponential bound, $\Pr[\varepsilon] = e^{-\theta}$

[2] N_s = average number of detected photons per bit

Fig. 11. Ideal receiver performance.

frequency (heterodyne detection). Phase and/or frequency tracking of the signal field by the local oscillator laser is required. Polarization matching can be implemented via the use of circular polarization and making sure the optical train does not have birefringence. The mixing of the weak signal field and the strong local oscillator field at the front-end of a coherent receiver provides linear amplification and converts the optical signal into an electrical output with gain (usually tens of decibels), raising the signal level well above the noise of subsequent electronics. A detector with gain is not required, and quantum-limited performance (for coherent detection, not quantum optimum detection) can be achieved usually with a dual-detector receiver [19], [20]. Coherent detection can be used on any type of modulations including those that imprint information on the phase of the carrier. Fig. 11 compares these structured receiver performances with that of the quantum optimum receiver for binary signaling. Note that the quantum model for preamplified direct detection is the same as heterodyne detection, and should be viewed as equivalent to heterodyne detection rather than standard direct detection.

For the future, there can be substantial improvement of communication performance with the use of error-correcting codes and higher symbol size signaling. Current implementable codes (such as turbo codes and low-density parity check codes) can approach the channel capacity of $1/\ln 2$ bits/detected-photon (for signaling schemes such as binary on–off signaling), yielding ~10 dB gains over the uncoded system. The capacity of the ideal direct detection system without bakground noise and restriction on the symbol size actually approaches infinity using increasing large symbol size M-ary pulse position modulation (PPM) signaling, albeit with the peak power of the pulse also increasing to infinity in the limit. For a channel with additive noise and an M-ary PPM system, it can be shown [18] that the computation cutoff rate (within a factor of ~2 of the channel capacity) is

$$R_{\text{comp}} = \log_2 M - \log_2 \Bigl[1 + (M-1) \times \exp - \left\{ \sqrt{\lambda_s + \lambda_n} - \sqrt{\lambda_n} \right\}^2 \Bigr]$$

where λ_s is the number of photons received in a single symbol and λ_n is the number of noise photons per time slot of the symbol. The modulation bandwidth of a space system is only limited by modulator speed and receiver bandwidth and not fiber dispersion. Thus, it is reasonable to expect a space system to be able to perform close to the limit of infinite modulation bandwidth. Under an average power constraint (realistic for EDFAs) of μ_s received photons per second, optimum communication efficiency is achieved as M, the symbol size, becomes unboundedly large, reaching the efficiency asymptote of

$$\left(1 - \frac{1}{M}\right) \mu_s / \ln 2 \text{ bits per second}$$
$$\left(1 - \frac{1}{M}\right) / \ln 2 \text{ bits per received photon}$$

which is basically the performance of a coded system. The advantage of such a large symbol size system over a small-symbol-size error-correction coded system is that the on-pulse has high enough signal power for a simple timing acquisition and tracking system to work properly. Thus for very high data rate services where the average received signal strength is high, a small symbol size coded system would be the choice, but for deep space systems where the signal strength is weak, a large symbol size system will yield much simpler system designs. This signaling scheme does make quite inefficient use of bandwidth, and modulator and detector bandwidth can be a problem for high-data-rate systems.

III. OPTICAL SATELLITE NETWORK ARCHITECTURE

There can be two reasons an optical satellite network is economically viable. The first is that for long-distance intercontinental transmissions, it can be cost-competitive with undersea fiber systems and can become an alternative for terrestrial networks. One interesting property space optical communications has is that the power attenuation due to free-space diffraction loss is only inversely proportional to the square of the link distance, whereas optical fiber attenuation is exponential in distance and amplifiers/repeaters at regular distances are required to maintain performance. Given long enough link distance, the total attenuation seen by the fiber link will become much larger than that seen by the space optical link. Perhaps the first generation of such a long-haul system will use RF up- and downlinks of up to a few Gb/s. Later generations may use optical links through the atmosphere together with multi-site diversity and a ground fiber network for interconnection of these gateways. Though the startup cost of an optical satellite network is higher, for long distances it can be more economical than (especially undersea) fiber systems.

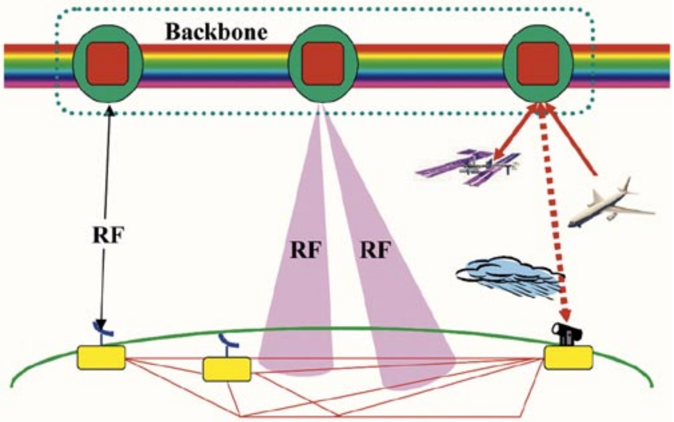


Fig. 12. Interfaces of the optical satellite network.

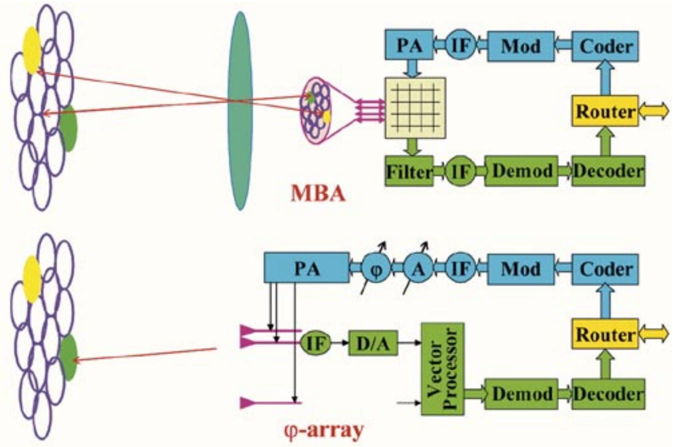


Fig. 13. Microwave access links for satellite systems—multiple beam antenna and phase array.

In [4], a simple life-cycle cost comparison of equal capacities in 10-Gb/s quanta was performed. The per-year operating costs are normalized with respect to system lifetime cost (∼10 years for space systems), included nonrecurrent engineering, production, deployment, and operating costs. The crossover between the two systems is estimated to be around 5000 Km (using undersea fiber systems as a benchmark and based on best guessed optical crosslink and RF up- and downlink technology costs) [16].

The second reason is that the optical satellite network may provide unique services to space missions and open-air-interface accesses such as voice and data communications over microwave satellite systems for mobile platforms and remote users with no broadband wired access, and satellite and terrestrial distributed sensing readout. The economical viability of such a network will heavily depend on the architecture. Since current satellite networks are mostly trunk-based with no particular attention paid to data communications, it is imperative that an entirely new space network architecture be developed. In this paper, we can only outline the necessary areas to be addressed and some possible solution pathways. The application section will simply invoke that optical crosslink-based satellite networks are feasible and economically viable and proceed to examine what new services and architectures are possible. We will first address the possible interfaces to the optical satellite network and then discuss other subsystems needed to support a new satellite network architecture.

A. Interfaces to RF and Other Access Links and to the Terrestrial Network

Fig. 12 depicts the various interfaces of the optical satellite network: microwave and optical trunk connections to the terrestrial WAN, metro-area network, and LAN interconnections via microwave up- and downlinks, and microwave and optical individual user accesses. Though limited in rates (∼ a small number of gigabits/second using high spectral efficiency modulation and coding), microwave links will be the primary connections to the terrestrial network. With this choke point, the optical satellite network cannot be considered architecturally as a simple extension of the terrestrial network but rather a separate network with only moderate rate gateway connections (<10 Gb/s) to the terrestrial WAN. Thus, direct cost comparison to the terrestrial WAN without taking into account the access network architecture and technology and the services the network provides is meaningless. Optical connections through the atmosphere to the ground can alleviate this choke point, but they can only operate in clear weather and need multipath diversity to provide any sensible network performance.

Microwave satellite data accesses at high rates require narrow spot antenna beams for high sensitivities and small user terminals. These can be realized via a multiple beam antenna system using a microwave lens or a microwave phase array, as shown in Fig. 13. To utilize communication resources, such as transmitters and receivers, onboard the satellite efficiently, a special data oriented layered-network architecture will be required. This includes an efficient media-access-control (MAC) protocol, designed especially for the large bandwidth-delay-product microwave satellite access links, for the allocation of satellite communication resources, switching and routing algorithms, and a transport layer protocol that deals with the special properties of the satellite medium. Optical accesses are simple for space platforms but only work for airborne and ground users when the weather is clear. For aircraft and ground-based nodes, the system has to contend with boundary layer and clear atmospheric turbulence perturbing the links and degrading link quality, as we will address later in the next section. Fig. 14 summarizes the nature of the user accesses to the satellite network. We will discuss briefly the design of such access networks in the next section.

B. Physical Network Topology

One of the most fundamental considerations in the design of a space backbone network is the physical topology of the backbone satellite constellation. The primary goal of a backbone constellation is to provide the coverage as required by the users; see Fig. 14. Users are in LEO, MEO, GEO, and the relevant parts of highly elliptical orbits, as well as airborne and on the ground. These coverage requirements can be met by a variety of constellations, with different altitudes, number of orbital planes, arrangement of the orbits, and arrangement of satellites within the orbits. All of these factors will influence the complexity and performance of the overall system. The complexity of the system

Connection Types	Sources in space	Destination node
Class 1: Long streams of 1-100 Gbps (minutes to always-on)	10-100 total	Class 1 accesses in worldwide WAN
	LEO (250-1,500 km)	mobile
Class 2: Unscheduled access of 1Mbps-1Gbps bursts (fractions of seconds to minutes)	MEO (5,000-15,000 km)	LEO, MEO, GEO spacecraft
	GEO (35,786 km)	Class 2-only access points
Class 3. Low rate random accesses	HEO (1,000-40,000 km) (can only access backbone when altitude >5,000 km)	Large # of Class 3

Fig. 14. Services provided by the backbone network for space or space/earth users.

Configuration	Altitude (km)	Total # of Sat	# of Orbital Planes	# of Sat per Plane	Orbit inclination
Polar LEO Space	1,550	12	3	4	90°
Polar LEO Earth	1,550	40	5	8	90°
Walker LEO Space	1,550	10	5	2	57.1°
Polar MEO Space	15,000	6	2	3	90°
Polar MEO Earth	15,000	8	2	4	90°
Walker MEO	15,000	5	5	1	43.7°
GEO	35,786	3	1	3	0°

Fig. 15. Satellite network constellation configurations for space or space/earth users.

can be quantified with various parameters: coverage requirements, constellation altitude, number of orbits, number of satellites required per orbit, and number of required ground gateway stations. These constellation parameters will in turn determine the complexity of the individual satellites, as each backbone satellite must maintain intrabackbone crosslinks within the constellation geometry as well as provide access connections for users in its coverage area. The complexity of each individual satellite can be quantified by the number of apertures required, the size required of each aperture, the slewing rate required of each aperture, and any obscuration issues that arise from the placement of apertures on the satellite. Fig. 15 provides a comparison of typical satellite network configurations (for a more detailed treatment of the subject, see [21] and [22]). These parameters can be used to compare possible backbone system designs and the derivation of a speculative cost model. GEO constellation has the drawback of the 1/4-s propagation delay that affects voice and video conferencing. However, though LEOs may alleviate the delay problem, its constellation is complex and requires all user terminals to have a tracking antenna (except for low rate users with omni-antennas). Since most of the traffic of a high-rate satellite network will be data-based, provided a suitable MAC protocol is used, the effects on user quality of service of the GEO delay can be minimized. In addition, passthrough traffic [22] in a satellite constellation, which increases crosslink capacities, is a strong function of constellation architecture. For example, for a LEO or MEO constellation with m planes and n satellites per plane and all satellites are connected to their neighbors, the ratio of passthrough traffic to add/drop local traffic in their coverage area is $\sim(m+n)/4-1$ [22]. This passthrough traffic can become a burden to the crosslink resources of the optical satellite network and drive the system cost up. Thus, it is prudent to minimize passthrough traffic with a small constellation, and a GEO constellation is most attractive. Given that GEO is the configuration of choice for the backbone of a satellite network, the capacities of the links in this backbone may reach 100 Gb/s and optical links are the only viable candidates. This is the reason why before optical crosslink technology matured, there were no high-rate satellite networks deployed. The traditional notion of connecting GEO nodes into a ring topology needs to be revised. To minimize passthrough traffic, a higher degree of connectivity (mesh) network topology should be used. An interesting benefit of a free-space link is that when the traffic load shifts, the physical connection topology can be easily changed via pointing the telescopes to a new satellites, and load balancing can be implemented more easily.

C. Spacecraft Node Switching Architecture

The backbone optical satellite network nodes must be designed to support necessary network functions just like their terrestrial network counterparts. The current terrestrial networks have backbone routers that deal with the long haul traffic and access routers for aggregation. These data network services may themselves co-exist on the same fiber plant with circuit-switched oriented services. Fig. 16 shows the connection architecture for an optical satellite network backbone node. A significant difference from a terrestrial backbone node is that the satellite node will also have to deal with accesses from individual users and thus have an aggregation function as well. Fig. 17 shows the different processing architecture options for a backbone satellite network node. The minimum configuration is to have transponders only and an analog switch to cross-connect input and output links. This can be done if the RF access links are subcarrier-modulated onto optical links at the expense of some link efficiency. Since space resources are precious, a better architecture will be to also have onboard demodulation and modulation together with switching and

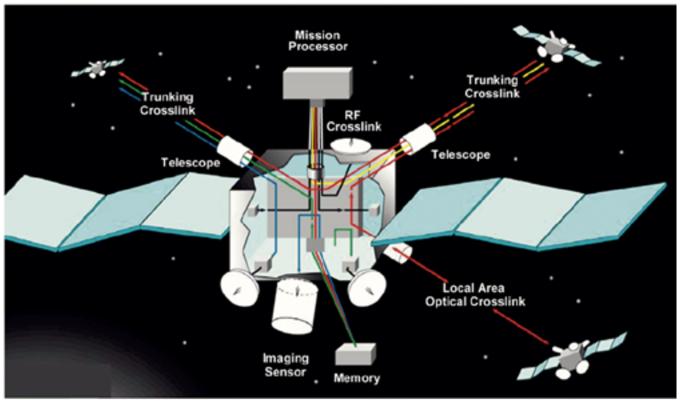


Fig. 16. Optical satellite node connection architecture.

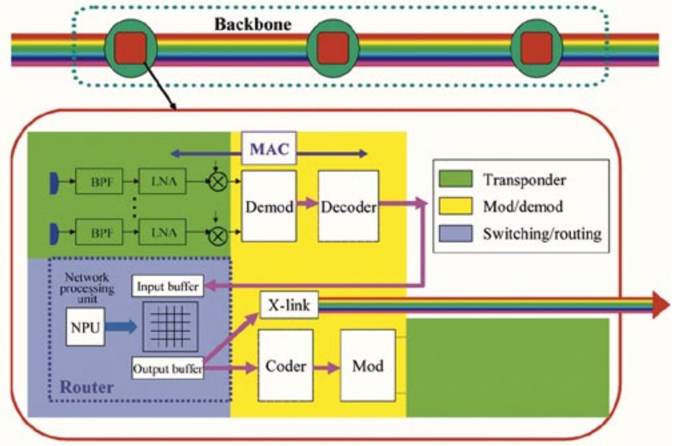


Fig. 17. Different processing architecture options for a satellite network node.

even routing, albeit not necessarily for all the traffic. The difference between simple switching and routing is that in routing, the presence of buffers allows more efficient statistical multiplexing of a large number of bursty computer users and is the right architecture for accesses.

Since a substantial fraction of the traffic at each node can be high-speed passthrough traffic, optical or electronic switching/“routing” via WDM techniques can be used to eliminate a significant fraction of expensive regeneration and electronic processing resources (such as packet-by-packet routing by a high-speed router) at the expense of losing some link performance (only if there is no regeneration). Fig. 18 depicts a backbone constellation with WDM access links and optical wavelength switching at the relay nodes. The node will also have to deal with RF access links. These can be demodulated and multiplexed onto an optical wavelength or can be subcarrier-modulated onto a wavelength, losing some power efficiency but gaining hardware simplification in the process.

D. Intraspacecraft Optical Network (Spacecraft-LAN)

The switching at the nodes may be configured as in Fig. 19, with trunk line switching as well as RF access link conversion to optical links and then switched. All-optical switching and routing at an optical satellite network node is format-insensitive to the traffic but will lose some link quality in the form of signal-to-noise ratio degradation.

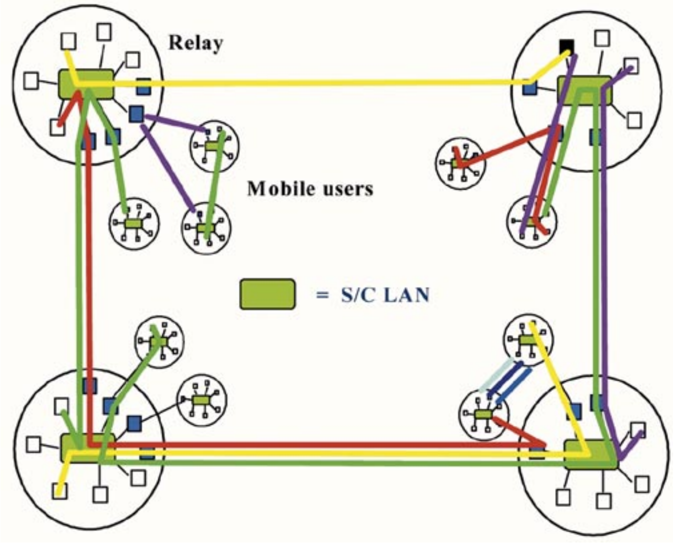


Fig. 18. WDM optical switching/routing of satellite network.

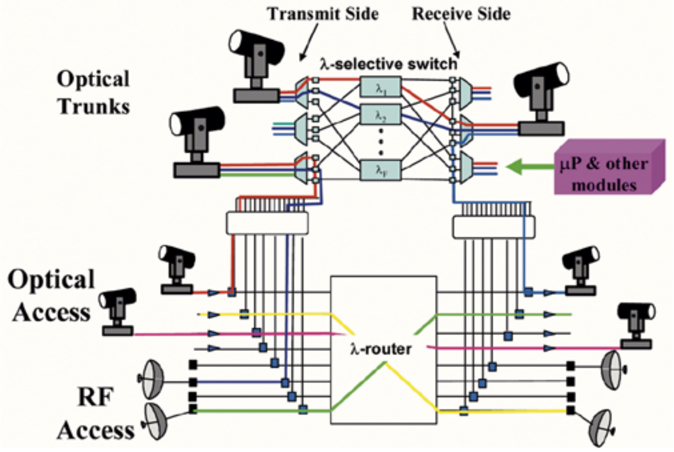


Fig. 19. Optical satellite network WDM switching node.

Format insensitivity may be a significant advantage for optical satellite systems since space systems are designed and deployed years ahead of user services. Even if demodulators and modulators are lightweight, the nonchangeable signaling format may make the network less adaptable to new services. Analog space links with high linearity and low distortions can be supported easily by all-optical switching onboard the spacecraft. This gain must be traded off against the loss of link performance over a fully regenerative relay node. In addition, when the source of the data is analog to begin with, such as the output of a space-borne sensing system, analog transmission at high fidelity may help substantially reduce system weight and power by eliminating analog-to-digital converters, data compression hardware and software, and digital regenerators at the mission satellites. A likely architecture for the satellite relay node is that the core backbone has circuit provisioning and switching and only accesses are routed and electronically processed. Thus, all relay nodes are logically one hop away from all the other nodes when viewed at Layer 3, the network layer. Some wavelength

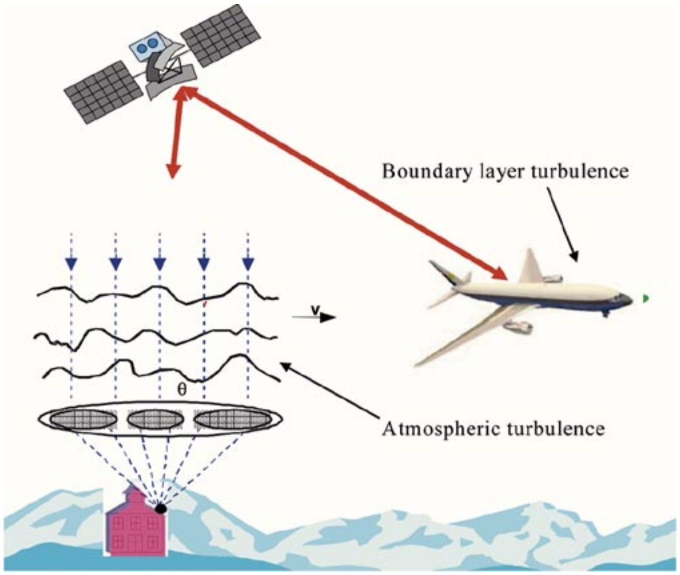


Fig. 20. Optical links over clear atmosphere and boundary layer turbulence.

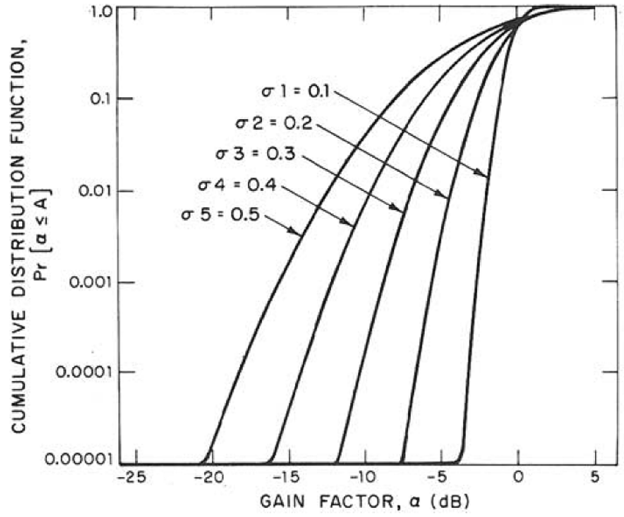


Fig. 21. Probability of fades as a function of turbulence strength σ: 0.1 light, 0.5 heavy.

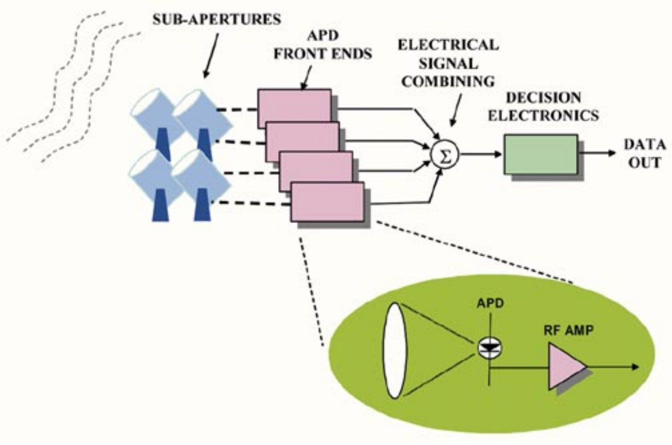


Fig. 22. Direct detection diversity receiver.

efficiency may be sacrificed, but it is a good trade for electronic processing hardware complexity.

E. Optical Links to the Ground and Aircraft—Clear Air Links

It is possible in clear weather conditions to communicate optically to a user terminal from space, especially if the terminal is located on a high-flying aircraft such as an airliner (see Fig. 20). It is also the only viable link (albeit at low rates due to significant phase front distortions as the optical beam propagate through the plasma generated) to the space shuttle during its RF blackout period during reentry.

In good weather conditions, the major impairments to line-of-sight laser beam propagation are due to atmospheric turbulence. If one terminal is on an aircraft, the boundary layer airflow or, in the case of supersonic flight, the bow-shock as well around the aircraft will also generate optical distortions. For communications at data rates ~ several gigabits/second using transmitter beamwidths and receiver fields of view of ~10 μrad, turbulence-induced fading due to refractive index fluctuations, which can be quite severe, is the primary cause of reduced communication performance. It is typical for deep fades (see Fig. 21) to last approximately 1–100 ms (directly caused by the speed of airflow across the beam), which, when operating at gigabits per second, results in the loss of a large number of consecutive bits. This motivates the need to consider schemes that minimize the probability that the receiver "sees" a significant fade. Most optical communication system design will be affected by events at 10^{-2} or lower probability of occurrence. Spatial, temporal, and frequency diversity architectures are candidates for fade mitigation of the received signal. A diversity system statistically guarantees that at least one of the independent paths has good quality. In contrast to spatial diversity for wireless systems, spatial diversity for atmospheric optical systems can be readily implemented since the intensity and phase coherence length is on the order of centimeters, i.e., multiple transmitters or receivers only need to be placed centimeters apart to see approximately independent channel fades. An error-correcting code, on the other hand, is not an attractive technique to mitigate fades since it would require an impractically large interleaver for a high-data-rate channel.

The question of gain of spatial diversity receivers has been analyzed with respect to average probability of error. However, this is not the correct performance metric. Even with a respectable long-term average bit error rate, the link can temporarily suffer outages of 1–100 ms due to deep fades. Thus, outage probability should be used to analyze the performance gain of using spatial diversity systems [23]. Outage probability is the probability that the bit error rate of the channel is higher than an outage threshold bit error rate. This is an important parameter for the link to be operated as part of a data communication network using automatic-repeat-request (ARQ) and forward error correction (FEC). Either direct (incoherent) detection receivers (Fig. 22) or coherent receivers (Fig. 23) can be used. Coherent receivers will provide high selectivity and excellent background noise rejection and the optimum gains in theory but they are harder to implement, with implementation losses offsetting the edge over direct detection receivers. Over 10 dB of gains can be realized for direct detection receivers in heavy turbulence. A substantial fraction of these gains can be achieved with around ten receivers

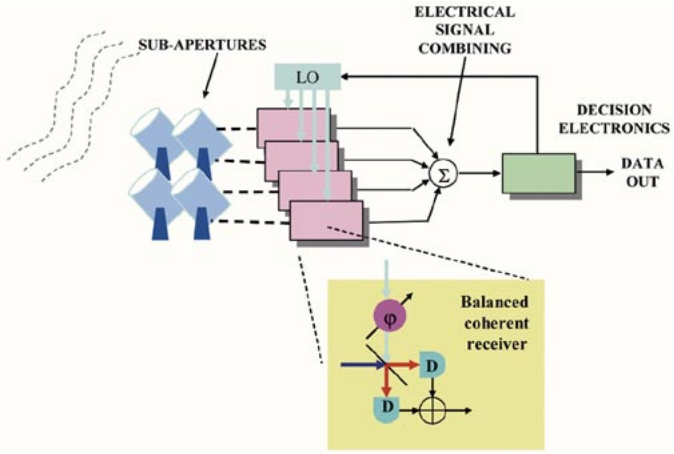


Fig. 23. Coherent combining diversity receiver.

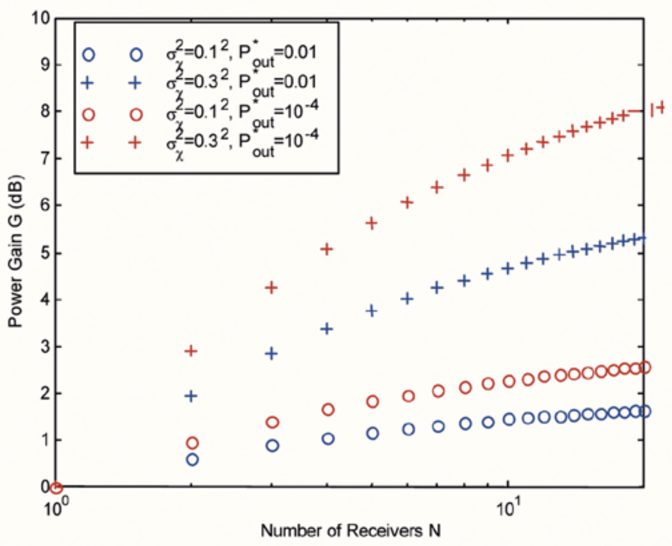


Fig. 24. Power gain as a function of number of receivers and turbulence strength log-amplitude variance σ_χ.

(Fig. 24). For small number of receivers N, the gain increases approximately as the square root of N. The readers are referred to [23] for a detailed analytic treatment of the topic.

F. Optical Multiple Access

In a number of specialized applications, multiple users can be within the same field of view of the satellite receiving telescope. From geosynchronous orbit, users more than 100 m apart can be resolved by multiple focal plane detectors of a modest telescope of ∼10 cm in diameter. However, for many lower rate random-access users, such as those on airliners, it is prohibitively costly to assign one optical receiver per user even if it is dynamically scheduled. Thus, an optical multiple-access receiver combining the signals of a larger group of users at the same detector makes sense (see Fig. 25).

The messages from the individual users can be extracted via decoding schemes implemented via electronics. Each user can gain the attention of decoding resources by sending a unique preamble [24]. Since each user laser will not be coherent with other users' lasers, the modulation scheme will have to be energy- and not phase-modulated. For lower rate users, there should be plenty of bandwidth available, and hence M-ary PPM is a good signaling scheme. It can be shown using random coding arguments [24] that if there are on the average K users active in the same optical channel, the capacity of the multiaccess channel is given by $C \sim \log M - \log K$ bits per use of the channel. As the number of users becomes large, the capacity approaches *ln2* bits per time slot. The above result assumes that the communication performance is limited only by multiple user interference and not by benign detection noise. This is a good assumption when the user population is moderate or large. The optimum receiver registers all time slots (out of M slots) that have signal pulses in them. Each user has its individual code that can be decoded by the receiver processor. The decoding process can be done at the receiver or the detection statistics can be shipped via crosslinks to a processing satellite or to the ground for decoding. When the demand is close to or exceeds the capacity of one single optical channel, multiple wavelengths can be used to increase the capacity of the system.

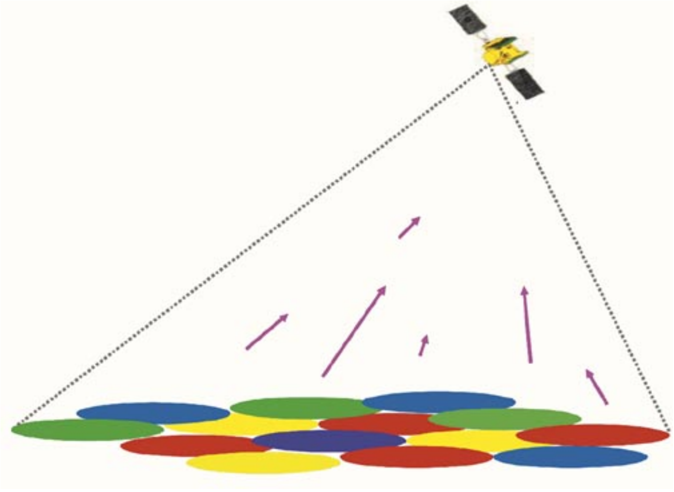


Fig. 25. Optical multiple access.

IV. NETWORK ARCHITECTURE ABOVE THE PHSYICAL LAYER

Networking over a satellite network is very different from terrestrial networking. At the physical layer, independent symbol errors occur due to RF and optical receiver noise and in the case of good optical receivers that are readily implementable, quantum uncertainties. In addition to the independent errors induced by the detection process, the satellite channels also can suffer time-varying degradations much longer in duration by comparison to a symbol time (e.g., during tracking system anomalies), and thus a large consecutive segment of data is affected. These channel effects include, in the case of a free-space link over the atmosphere, the turbulence-induced fades of the optical signal, as mentioned in the last section where diversity techniques can reduce the probability of occurrence of severe signal outage. In the case of high microwave frequencies (say, ∼20 GHz), turbulence-induced fades also occur, though not as deep. In addition, though microwave will propagate through bad weather, moisture does attenuate the signal significantly during moderate and heavy rain. The capacity of the microwave channel can vary over two orders of magnitude. Adaptive modulation and coding can keep the links operating at or close to capacity [25]. Thus, it is not necessary to waste the extra power reserved as link margin in most microwave systems today. For circuit-oriented services, this gain is of little use since the link capacity may change over seconds and thus cannot be conveniently allocated to additional users. For data-oriented services, these gains can be utilized effectively for bursty accesses. It will appear to the users as a less loaded network with faster response, not unlike a lightly loaded or faster Ethernet. The physical layer can adjust in the timescale of one second or longer [25]. However, current upper layer protocols such as the network layer routing protocols are designed to change slowly (as long as a timescale of minutes) to mitigate undesirable effects such as oscillations of network flows. Thus, new network layer protocols will have to be designed to avoid network instabilities due to fast adaptation and prevent network oscillations. Also, with the long link delays, especially over geosynchronous distances, and the high data rates possible with optical links, there will be typically many packets in flight. If traditional protocols such as transport control protocol (TCP) are used, the network will be very inefficient due to a number of well understood effects such as "timeouts" and "windows closing."

Satellite systems are an expensive investment. With the advancement of fiber, wireless and electronic technologies, terrestrial modalities have become very economical and price competitive. It will be impossible for satellite systems to compete by being inefficient. Thus, the overall network architecture must be optimized for satellite networks to have any chance of survival. This architecture must be satellite-technology specific and adaptable to the changing network conditions. The subject of higher layer network architecture for satellite systems is very complicated and will take more than the coverage allowable here to do it justice. For this audience, we will highlight the critical areas of a generic data satellite network architecture with emphasis on those that particularly affect the optical communication and network hardware design.

A. Data Link Control Layer (DLC)

The DLC will participate in adaptive rate adjustment of the microwave access links via changes to the modulation and code rates. It can also perform transmitter power adaptation in combination with the increase or decrease of the number of optical diversity receivers brought into action. An FEC will almost surely be used in most link designs. Most FECs being used in space communications today were conceived decades ago, and they are typically several to 10 dB away from the theoretical optimum channel capacity. Modern codes, such as low-density parity check code and turbo code, should be used to utilize the expensive links close to the capacity limit. For data packet communications, it is advisable to have an automatic repeat request (ARQ) in the DLC layer to guard against uncorrected errors by the FEC before these errors are presented to the transport layer and trigger unintended and detrimental effects such as window closing (a reaction by the transport layer misinterpreting erroneous packets as the consequence of network congestion and initiating a slowdown of the transmission rate). In any case, some form of ARQ will be used, if not at the DLC layer, then definitely at the transport layer. To jointly optimize the design of the bias point of the modulation/coding and the ARQ, the outage probability is the relevant metric, as we have alluded to in the last section. This is because uncorrected packet errors due to the link's falling below the outage threshold will trigger retransmissions and false triggering of congestion control mechanisms and window closing by Transmission Control Protocol (TCP), severely limiting throughput. Thus it is important to design the optical links to operate at the optimum bias point for the networking protocols used. An overdesign to achieve very low error rates will waste power and link capacity, and an underdesign will trigger frequent link errors lowering network throughput. This is currently an open problem. Some in the community do recognize the need to address the joint optimization across multiple layers, but the possible link designs and channel impairment statistics are just becoming known.

B. Network Layer

The network layer is responsible for routing. While there are many routing algorithms, including those used on the Internet today, there has been little development that is applicable to the satellite networks. The satellite network has the interesting property that it does not readily lend itself to incremental and responsive upgrades due to the long gestation period of the deployment of new space assets. Routing algorithms may have a significant effect on physical network architecture, such as constellation and link capacity that cannot be changed. A little foresight in the design of the network layer is necessary to arrive at the optimum hardware design up front.

We assume that the optical satellite network physically interfaces to the terrestrial WAN at multiple gateways via microwave or optics. The optical crosslinks in the satellite network can be sized to carry the expected traffic with some capacity margin to deal with statistical demand surges and uncertain future market growth. This margin can significantly oversize the design capacity and raised the network cost to the point of not being prize-competitive with alternative modalities. This problem is exacerbated by the long gestation period ($\sim$ years) for deploying extra capacities. Alternatively, the designer can size the crosslinks to carry the expected traffic plus a small margin. Any excess traffic beyond that can be routed through gateways to the terrestrial WAN for long-haul transmission on a per-unit cost basis, as depicted in Fig. 26. As the demand rises, the gateway links can be upgraded in a much more responsive manner by a combination of increasing the aperture size, receiver sensitivity, and transmitter power of the gateway terminals. This assumes that the transmitters on the satellites are designed to have variable rates to accommodate the increase in capacities, which they need to have to begin with if optimum transmission rate through the atmosphere is to be extracted via dynamic adaptation. The network layer design can be optimized jointly with the dimensioning of the optical crosslink capacities [26]. This is quite different from common practice in terrestrial networks.

It is possible to formulate satellite link dimensioning and routing as a two-stage stochastic programming problem. The first stage optimizes link capacities for a random input demand to minimize an effective system cost. The second stage assumes that link capacities are fixed and maximizes the utilization of

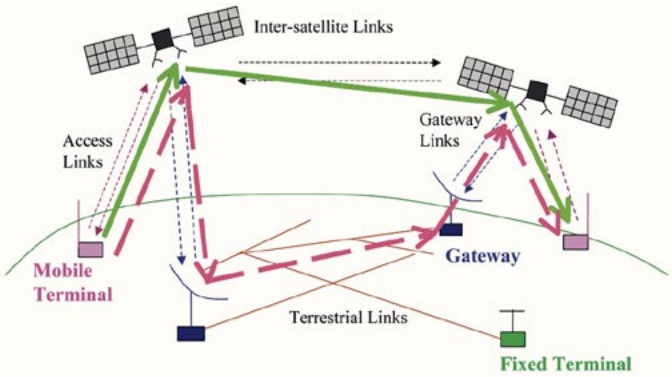


Fig. 26. Satellite routing over crosslinks and terrestrial network.

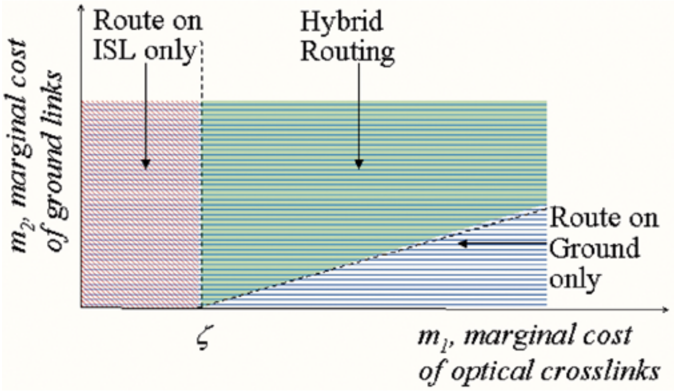


Fig. 27. Solution to the satellite network dimensioning problem.

satellite links through optimal routing. The objective is to minimize the sum of satellite network investment cost and an opportunity cost for rejecting excess input demands. One of the main advantages of using the stochastic programming technique for dimensioning satellite links is that the use of a random traffic model, instead of a fixed traffic model, better characterizes the uncertainties in traffic estimation and gives more explicit capacity and cost tradeoffs useful for network design and decision making. This is particularly important for satellite system planning due to the long planning horizon, typically on the order of five to ten years or more. Though the solution to this problem may need complex computations, the formulation is as straightforward as the following mathematical programming problem:

$$\min_R\{J(R)\} = \min_R\{c(R) + E\left[c_g(R,r) + D(R,r)\right]\}$$

where J is the overall cost including penalty of lost revenue due to lack of capacity, R is the capacity matrix of all the links, r is the stochastic traffic matrix, c is the total cost of optical crosslinks and gateway links, c_g is the cost of routing on the terrestrial links, D is the cost of rejecting traffic, and $E\{\cdot\}$ denotes taking expectation over the stochastic quantities inside the bracket.

Fig. 27 illustrates the spirit of this coupled network layer and physical layer topology design problem via a simple example of only using GEO satellites as backbones and access points. m_1 and m_2 are the marginal costs of optical crosslinks and ground links, respectively, and ς is a threshold based on a function of link and operating costs [26]. The result basically says that if the crosslink is expensive, one should use ground links only in the network, and vice versa; if the ground links are very expensive, only space crosslinks should be used. The interesting point is that there is a region where a hybrid strategy with both satellite crosslinks and terrestrial long-distance links yields a better overall system cost. The mathematical results have not included the additional possibility of capacity upgrades via changes to the gateway terminals and reprogrammable space transceivers. From the viewpoint of keeping the space investment cost down and retaining the flexibility for future upgrades, this hybrid approach should be used for the design of the optical satellite network.

C. *Transport Layer*

Standard transport layer protocols such as TCP are known to have problems over satellite networks. For example, errors in the link due to noise or multiple-access collisions may be interpreted as congestion in the network and lead to window closing, reducing throughput to a small fraction (~1%) of the channel capacity. The problems are likely to get much worse when the data rate of the links becomes higher as technology improves to multigigabits/second. Many of the "quick fixes" implemented today such as extra FECs, increase of TCP window size, insertion of ARQs at the DLC layer, and proxy services (at the application layer) at the terrestrial network to satellite network gateways may not work over a broad variety of links in the satellite network and compromise some of the necessary functions needed to perform congestion and flow control over the terrestrial network. There are more variables in the lower layers such as symbol errors, variability of link capacity due to weather, MAC delays, and congestion at routers for the current version of TCP to adapt in efficient ways. The right transport layer protocol should receive all these information as part of its observables list and optimize its congestion control function accordingly. In [28], Modiano considered a system consisting of two end nodes communicating over a multihop network using a higher layer (HL) protocol and a lower layer (LL) protocol. The HL at the two end nodes implements an additive-increase-multiplicative-decrease congestion control mechanism, similar to the operation of TCP; and the LL protocol implements a link layer retransmission mechanism (ARQ) over one of the links that experiences transmission errors. Both go-back-N and selective repeat ARQ protocols are considered. The study showed that the throughput of a jointly optimized protocol suite is higher, but specific tuning to the link properties is required. Recently, Katabi [27] proposed a new protocol called XCP that promises much better throughput and stability with minimal information requested from the physical layer. Adaptation to a detailed satellite network design is yet to be performed. Nonetheless, the protocol seems to have the right properties and performed well in simulations. The main attribute of this protocol is that the network layer requests at least some minimal information about the underlying physical layer and formulates its action accordingly. Thus, it is imperative and prudent to reexamine network architecture from the fundamental premises of efficient networking, breaking traditional views of layer boundaries and exploring what performance can be attained. This is also a new area of research that is yet to be fully developed.

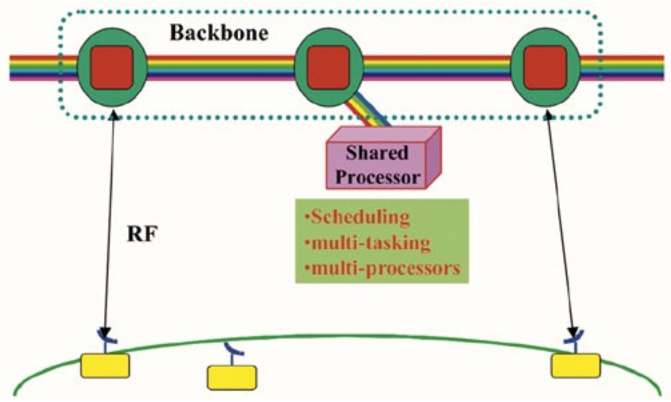


Fig. 28. Shared processing in space.

V. INNOVATIVE SPACE ARCHITECTURE IMPLICATIONS OF A HIGH-SPEED OPTICAL SATELLITE NETWORK

Optical crosslinks provide satellite networks with a quantum leap in capacity and lower cost. It has truly allowed the creation of a cost-competitive space network. However, the implication of an optical satellite network in space goes far beyond economics and rates. It is an enabling technology that permits the creation of new application architectures. Innovative ways of using this network may revolutionalize satellite communications and space missions such as remote sensing. Here we will suggest some promising avenues to explore.

A. Optical Satellite Network as an Enabling Technology for Shared Spaceborne Processing

In the near future, optical crosslinks will become much more economical and have higher rates than RF up- and downlinks. Space-borne sensors currently use RF downlinks to bring the data to processing centers on the ground. The high cost of high-rate RF downlinks from satellites, not the resolution of the sensors, will become the limiting factor on system resolution and sensed area rates. With an ultra-high-rate and economical optical satellite backbone in place, one can think about the concept of using shared space-borne processing (Fig. 28) to reduce the amount of data to be sent earth-bound and thus reduce the requirement for very costly high-RF downlink data rates.

Modest RF downlinks can then be used for the processed and reduced data, substantially lowering the overall system cost and raising the resolution and coverage rate of the sensor. With the long planning horizons and lifecycle time of space systems, space-borne processing technology tends to lag behind current technology by at least a few years if not more than a decade in many instances. Hence, not surprisingly, space-borne processing is not ubiquitous today. If a high-speed optical backbone satellite network is available, a more progressive concept is to locate a processing satellite somewhere close to or within the backbone, with the most advanced processor at the time of launch and periodically replaced and replenished as frequently as yearly. First the sharing allows much more efficient utilization of the processing system, as well as reduces the need for 100% redundancy. Current architectures use radiation-hard processors (equivalent in performance to five- to seven-year-old designs), and these processors have to function for seven to ten years, making the end-of-life processors very obsolete. With a short lifecycle agile replenishment on-demand technique, we can reduce the requirement on cumulative radiation hardening and thus use the most powerful processor available at the time of launch. Computing elements only have to live one to two years. For example, the current generation of Pentium processors can survive three to four years in GEO orbit. On the average, such an architecture will be using processors, data buses, memory technologies, and software that are ten to 15 years more advanced than current practices. Note that this replenishment and upgrade strategy is only sensible when several space systems share the processing. Otherwise individual replenishment will be unwieldy and costly and negate any benefit of the availability of more modern processors. This sharing in turn is only feasible if there is enough data rate in the space backbone to ship the data around and interconnect the multiple platforms that the processing satellite serves.

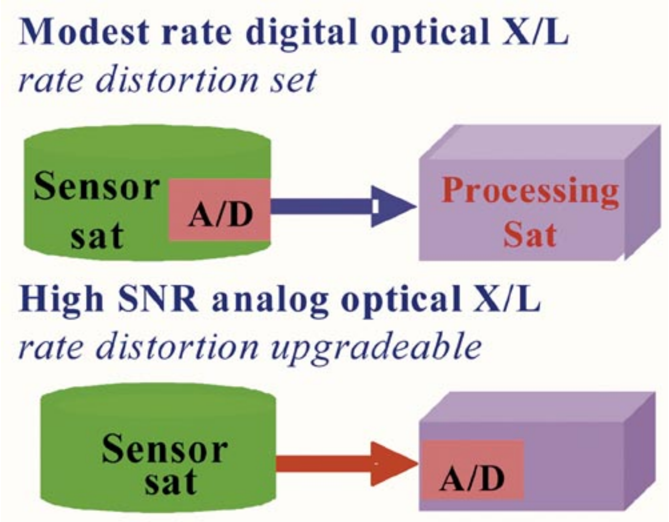


Fig. 29. Sensor resolution upgrades via analog links and upgradeable A/D technology.

B. Sensor Resolution Upgrade via Upgradeable A/D Technology

With a high-rate space backbone, processors may not be the only electronic component that can be replaced or upgraded at a much faster timescale than the lifetime of a satellite. Other electronic components and processors can also be decoupled from the mission satellite and relocated onto the processing satellite. For example, consider the placement of analog-to-digital converters (A/Ds) as illustrated in Fig. 29.

If A/Ds are built into the mission satellite, the raw data are digitized and transmitted in digital form to the processing satellites. If the quantization is lossy then, the distortion is set by rate distortion theory according to the available optical access link capacity. However, if A/Ds are implemented in the processing satellites, the raw data may be transmitted via analog links to the processing satellite, and then digitized and processed. Provided the analog link has been designed with enough

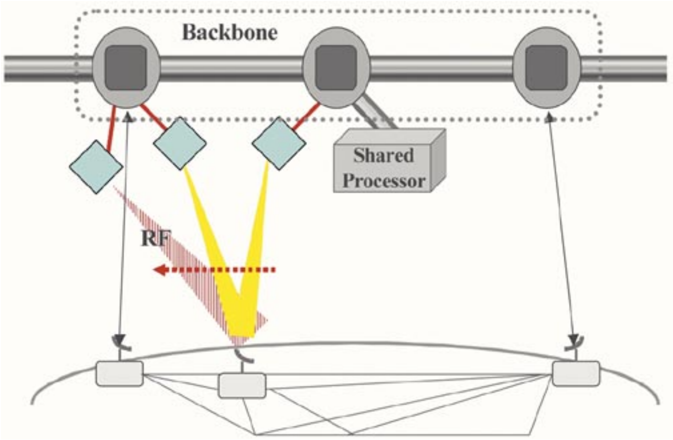


Fig. 30. High-resolution multiplatform distributed sensing.

signal-to-noise ratio, a new faster and finer A/D quantizer can be inserted into the processing satellite to reduce the distortion of the compressed signal when the technology has improved.

C. Multiplatform Multistatic Sensing

A sensor’s performance may be improved if the sensing function can be distributed over more than one satellite. For example, for geolocation, two satellites can form the arm of a long baseline interferometer provided the two sensed signals can be brought together for coherent processing (Fig. 30). This requires either fine quantization and significant data-rate transmissions to preserve phase information or coherent analog transmission at high fidelity, sometimes known as transmission transparency. An analog link of an optical satellite network can provide such service. In general, a distributed satellite system can substantially improve image-oriented sensing and object identification from space.

D. On-Orbit Upgradeable Satellite Communications

The RF section and the baseband processing section of a satellite transmitter and receiver have different lifecycle durations. The antenna and RF front-ends usually stay at near state-of-the-art longer than the processing segment. If, for example, the raw RF analog signal or the digitized waveform is sent to a processing satellite to perform the rest of the receiver function via software processing by general processors, both the processors and the software can be reprogrammed or upgraded to adopt new or better modulation, coding, and MAC protocols (Fig. 31). Similar changes can be made to the transmitter.

This is an empowering attribute for satellite communication, allowing the space network to parallel the rapid evolution of the terrestrial network.

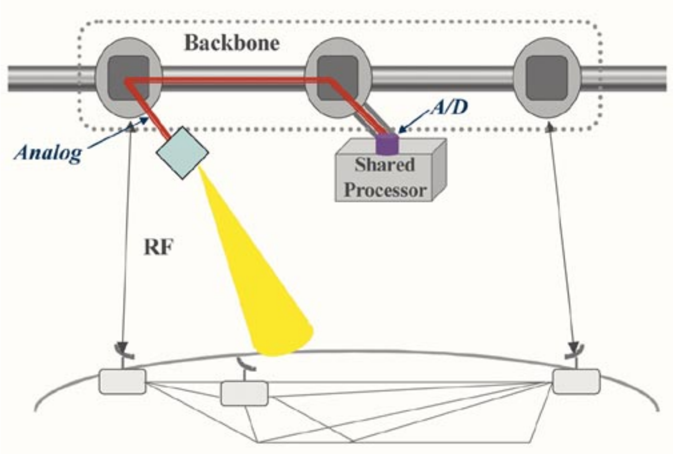


Fig. 31. Reconfigurable and upgradeable RF satellite access network.

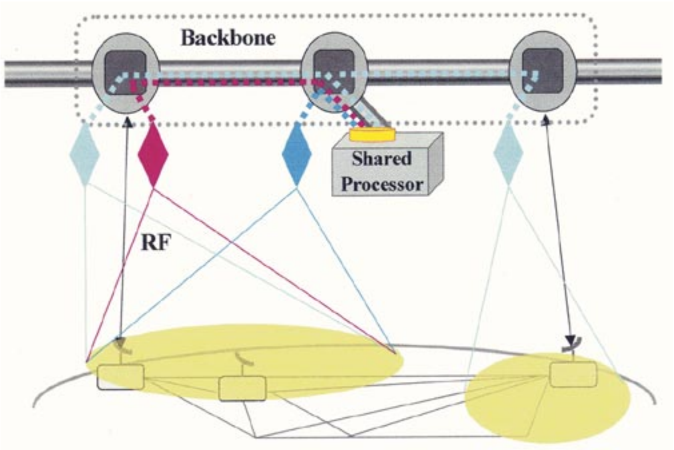


Fig. 32. Interoperable interconnected space communications.

E. Interoperable Satellite Communications

The terrestrial network ties together disparate modalities to form a single network. Satellite systems have not followed that paradigm. The main impediment has been each satellite systems have stovepipe designs that are not interoperable, and tying them together require gateways or teleports on the ground that not only are costly but also tie up a significant fraction of the up- and downlink resources for the connections to the conversion gateways. Moreover, interconnections at the application layer slow down the overall end-to-end network response. The high-speed optical satellite backbone and the processing satellite can be used to perform the conversion gateway function and tie the different satellite communication system in space rather than on the ground (Fig. 32). This may be the key technology that transforms the stove-piped satellite communication community to a data satellite network community serving many more users.

F. Multiplatform High-Performance Data Satellite Communications

Just as with multiplatform multistatic sensors, the optical satellite network enables the realization of a multiplatform satellite communication system, as in Fig. 33. The multiplatform system can be viewed as a traditional satellite communication system with a multi-element antenna array distributed over multiple satellites as a very large albeit “thinned’ (as in not filled) aperture. For example, this arrangement can improve communication performance to small and low-power terminals by forming an electronically antenna pattern with significant gain on the user and place nulls on strong interfering users to suppress their signals. With a powerful and modern processor farm on the processing satellite, this can be done dynamically in rapid response to bursty user demands and run

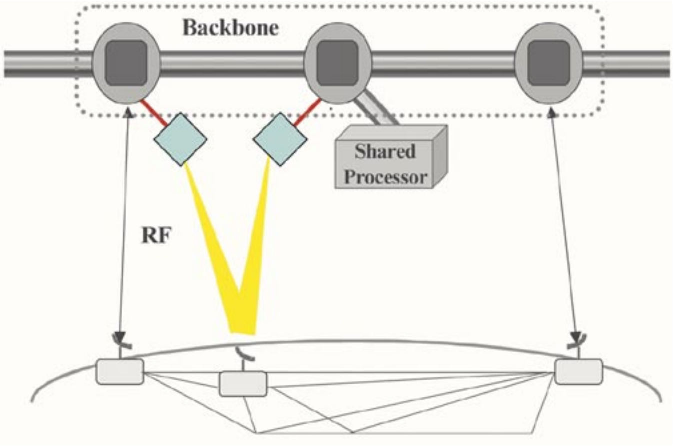


Fig. 33. Multiplatform distributed space communications.

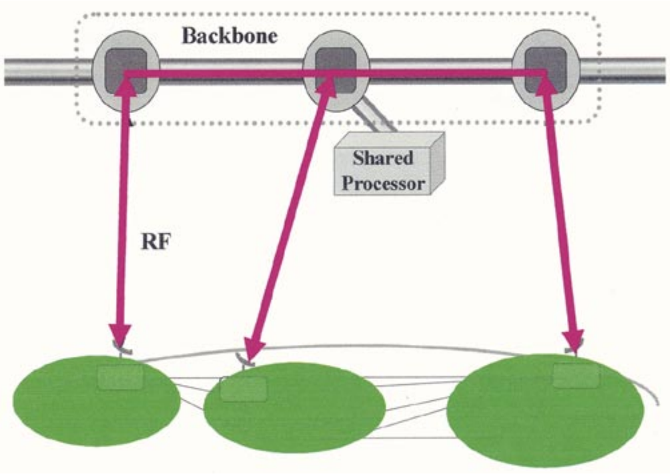


Fig. 34. Network for resonstitution, reconnection of disconnected terrestrial networks.

by a MAC protocol. Simultaneous demodulation of many users can be done in parallel.

G. Reconstitution of Disconnected Terrestrial Network

The terrestrial fiber network acts as a backbone to tie together subnets of different modalities and allow the Internet to behave as one single network. There may come a time when natural or man-made disasters disconnect part of this network from the rest of the network. The satellite network can serve as the alternate backbone to reconstitute a fully connected global network, as shown in Fig. 34. Architecturally, this notion requires the placement of gateways in sufficient density connecting the terrestrial network and satellite network, a network management and control strategy for the discovery of surviving resources, and performance of reconnection using the satellite network assets. A rich body of architecture has been developed for the restoration of IP-based terrestrial networks. The satellite network should adopt a similar architecture for interoperability and internetworking with the terrestrial network using Border Gateway Protocols (BGPs) and treat the satellite network as merely a different administrative domain.

VI. SUMMARY

Not very often in the history of communications and networking have there been truly transforming inventions that result in quantum leaps in the nature of services or costs to the end users. Examples that come to mind are the invention of the router that led to connectionless packet service and the Internet, and WDM technology's lowering long-haul costs significantly. Optical satellite communications will likely be classified as one such transforming technology if its architectural implications are fully exploited. Not only will the satellite network become economically viable, but also its deployment and the extraordinary services it can offer are capable of radically transforming space system architectures.

Not all optical space optical link designs are equal. In fact, their characteristics in terms of weight, power, size, and performance in terms of data rates and tracking robustness range over orders of magnitude. To make a space network cost competitive, it is imperative that the best designs be used. An unimaginative network will use the optical link as a backbone or a high-speed entrance link, which by itself may to some be transforming enough. However, the explosive potential is to design in a rich set of services, such as high-fidelity analog transmissions and spaceborne processing centers, to allow and stimulate new space system architectures and data network applications. It is in this second pathway that optical satellite networks should generate tremendous excitement and interests. This paper makes some suggestions for the new dimensions of space system architectures enabled by such a network. We are sure there are many others that are yet to be discovered. The opportunity is now to use imagination and creativity to bring home this new breakthrough in satellite networking and none too soon, given the state of stagnation in the aerospace community.

REFERENCES

[1] V. W. S. Chan, "Coherent optical space communications system: Architecture and technology issues," in *Proc. SPIE Control and Communication Technology in Laser Systems*, vol. 295, 1981, pp. 10–17.
[2] ——, "Space coherent optical communication systems—An introduction," *J. Lightwave Technol.*, vol. LT-5, pp. 633–637, Apr. 1987.
[3] V. W. S. Chan *et al.*, "Coherent intersatellite crosslink systems," in *Proc. SPIE Components for Fiber Optic Applications and Coherent Lightwave Communications*, vol. 988, 1988, pp. 325–335.
[4] V. W. S. Chan, "Optical space communications: Key building block for wide area space network," presented at the LEOS'99, San Francisco, CA, Nov. 1999.
[5] P. Van Hove and V. W. S. Chan, "Spatial acquisition algorithms and systems for optical ISL," in *IEEE Int. Conf. Communications*, Boston, MA, June 1983, pp. E1.6.1–E1.6.7.
[6] H. Van Trees, *Detection, Estimation and Modulation Theory I*. New York: Wiley, 1968. Section 4.2.3.
[7] E. A. Swanson *et al.*, "Heterodyne spatial tracking system for optical space communication," *IEEE Trans. Commun.*, vol. COM-34, pp. 118–126, Feb. 1986.
[8] T. E. Knibbe *et al.*, "An integrated heterodyne receiver and spatial tracker for binary FSK communication," in *Proc. SPIE Free Space Communication Technologies VI*, vol. 2123, 1994, pp. 188–199.
[9] M. Z. Win *et al.*, "Analysis of a spatial tracking subsystem for optical communications," *Proc. SPIE Free Space Communication Technologies IV*, vol. 1635, pp. 318–325, 1992.
[10] R. S. Bondurant, "High rate space laser communication," presented at the LEOS'99, San Francisco, CA, Nov. 1999.
[11] R. J. Glauber, "Coherent and incoherent states of the radiation field," *Phys. Rev.*, vol. 131, pp. 2766–2788, Sept. 15, 1963.

[12] C. W. Helstrom, J. Liu, and J. Gordon, "Quantum mechanical communicaton theory," *Proc. IEEE*, vol. 58, pp. 1578–1598, 1970.

[13] J. Livas, "High sensitivity optically preamplified 10 Gb/s receivers," in *OFC 1996*, San Jose, CA, Feb. 1996. Postdeadline Paper.

[14] V. Gapontsev, "High power optical amplifiers," presented at the LEOS'99, San Francisco, CA, Nov. 1999.

[15] J. Kaufmann, "Analog link design," in *Proc. SPIE Free Space Communication Technologies VIII*, vol. 22 699, 1996, pp. 38–49.

[16] V. W. S. Chan, "Optical space communication," *IEEE J. Select. Topics Quantum Electron.*, pp. 959–975, Nov./Dec. 2000.

[17] M. A. Saleh, M. M. Hayat, P. P. Sotirelis, A. L. Holmes, J. C. Campbell, B. E. A. Saleh, and M. C. Teich, "Impact-ionization and noise characteristics of thin III-V avalanche photodiodes," *IEEE Trans. Electron Dev.*, vol. 48, pp. 2722–2731, 2001.

[18] V. W. S. Chan, "Soft decision decoding for PPM optical channels with direct detection," in *IEEE National Telecommunications Conf.*, Nov. 1980.

[19] G. L. Abbas, V. W. S. Chan, and T. K. Yee, "A dyal-detector optical heterodyne receiver for local oscillator nosie suppression," *J. Lightwave Technol.*, vol. 3, pp. 1110–1122, Oct. 1985.

[20] H. P. Yuen and V. W. S. Chan, "Noise in homodyne and heterodyne detection," *Opt. Lett.*, Mar. 1983.

[21] J. Bau, "Topologies for Satellite Constellations in a Cross-Linked Space Backbone Network," M.Eng. Thesis, EECS MIT, 2002.

[22] K. C. H. Kwok, "Cost optimization and routing for satellite network constellations," S.M. thesis, Massachusetts Institute of Technology, 2001.

[23] E. Shin and V. Chan, "Optical communication over the turbulent atmospheric channel using spatial diversity," presented at the IEEE GLOBECOM, Nov. 2002.

[24] V. W. S. Chan, "A multiple-user random-access optical communication system," presented at the IEEE ICC, 1979.

[25] P. Choi and V. Chan, "Predicting and adapting satellite channels with weather-induced impairments," in *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, July 2002, pp. 2635–2639.

[26] L. Dai and V. Chan, "Capacity dimensioning and routing for hybrid satellite and terrestrial networks," *IEEE GLOBECOM 2000*, Nov. 2002.

[27] D. Katabi, M. Handley, and C. Rohrs, "Congestion control for high bandwidth-delay product networks," in *Proc. ACM SIGCOMM'02*, Aug. 2002.

[28] C. Liu and E. Modiano, "On the interaction of layered protocols: The case of window flow control and ARQ," presented at the CISS 2002, Princeton, NJ, Mar. 2002.

Vincent W. S. Chan (S'69–M'80–SM'92–F'94) received the B.S., M.S., E.E., and Ph.D. degrees in electrical engineering in the area of communications from the Massachusetts Institute of Technology (MIT), Cambridge, in 1971, 1971, 1972, and 1974, respectively.

From 1974 to 1977, he was an Assistant Professor with the School of Electrical Engineering, Cornell University, Ithaca, NY. He joined Lincoln Laboratory in 1977 as a Staff Member of the Satellite Communication System Engineering Group working on military communications and networking. In January 1981, he became the Assistant Leader of the Communication Technology Group, starting a research and development program on optical space communications. In July 1983, he formed and became Leader of the Optical Communication Technology Group and Manager of the LITE (Laser Intersatellite Transmission Experiment) Program. He became the Head of the Communications and Information Technology Division of Lincoln Laboratory until 1999. He also served as the Principal Investigator of a Next Generation Internet Consortium (ONRAMP) formed among AT&T, Cabletron, MIT, and JDS Fitel, and a Satellite Networking Research Consortium formed between MIT, Motorola, Teledesic and Globalstar. He is currently the Joan and Irwin Jacobs Professor of Aeronautics and Astronautics and Electrical Engineering and Computer Science, as well as Director of the Laboratory for Information and Decision Systems (LIDS) at MIT. He is a Member of the Board of Directors of Vitesse Semiconductor Corporation and the Chairman of its Technical Advisory Board. He also serves on the Technical Advisory Boards of Agility Communications, Axiowave Networks (TAB Chairman and as a Member of the Corporation of Draper Laboratory). His research interests are in optical communications, wireless communications, space communications, and networks.

He is a Fellow of the Optical Society of America (OSA).