Applied Technology Institute (ATIcourses.com)

Stay Current In Your Field • Broaden Your Knowledge • Increase Productivity

349 Berkshire Drive • Riva, Maryland 21140
888-501-2100 • 410-956-8805
Website: www.ATIcourses.com • Email: ATI@ATIcourses.com

Boost Your Skills With ATIcourses.com!

ATI Provides Training In:

- Acoustic, Noise & Sonar Engineering
- Communications and Networking
- Engineering & Data Analysis
- Information Technology
- Radar, Missiles & Combat Systems
- Remote Sensing
- Signal Processing
- Space, Satellite & Aerospace Engineering
- Systems Engineering & Professional Development

Check Our Schedule & Register Today!

The Applied Technology Institute (ATIcourses.com) specializes in training programs for technical professionals. Our courses keep you current in state-of-the-art technology that is essential to keep your company on the cutting edge in today's highly competitive marketplace. Since 1984, ATI has earned the trust of training departments nationwide, and has presented On-site training at the major Navy, Air Force and NASA centers, and for a large number of contractors. Our training increases effectiveness and productivity.

Learn From The Proven Best!

Via Satellite, July, 1998

A Primer on Satellite Communications

by Robert A. Nelson

In 1945 Arthur C. Clarke wrote an article entitled "The Future of World Communications" for the magazine Wireless World. This article, which the editors renamed "Extra-Terrestrial Relays", was published in the October issue. In it Clarke described the properties of the geostationary orbit, a circular orbit in the equatorial plane of the earth such that a satellite appears to hover over a fixed point on the equator. The period of revolution is equal to the period of rotation of the earth with respect to the stars, or 23 hours 56 minutes 4.1 seconds, and thus by Kepler's third law the orbital radius is 42,164 km. Taking into account the radius of the earth, the height of a satellite above the equator is 35,786 km. Clarke observed that only three satellites would be required to provide communications over the inhabited earth.

As a primary application of such a satellite system, Clarke proposed that satellites in geostationary orbit might provide direct broadcast television service similar to DBS systems like DirecTV -- a remarkable idea at a time when television was still in its infancy and it was not yet known whether radio signals could penetrate the ionosphere. He worked out a simple link budget, assuming a downlink frequency of 3 GHz, and estimated that the required transmitter output power for broadcast service to small parabolic antenna receivers would be about 50 watts. Electric power would be provided by steam generators heated by solar mirrors, but advances in technology might make it possible to replace them by arrays of photoelectric cells. Batteries would be used to provide uninterrupted service during eclipses, which occur in two seasons centered about the equinoxes. Clarke also estimated the mass ratio of a multistage launch vehicle necessary to deploy the satellite. However, he imagined the geostationary satellites to be outposts inhabited by astronauts to whom supplies would be ferried up on a regular basis, much like the Mir space station and the international space station now under construction.

Twenty years later, in his book *Voices from the Sky*, Clarke wrote a chapter entitled "A Short Pre-History of Comsats, Or: How I Lost a Billion Dollars in My Spare Time". For he did not patent the idea of a geostationary orbit and, believe it or not, orbits can and have been patented. (Recall the recent patent controversy between Odyssey and ICO.) However, despite the tongue-in-cheek subtitle, the famous author would not have profited from his idea for two reasons. First, arguably, prior art existed in the literature. In 1929 the Austrian engineer H. Noordwig observed that a satellite at an altitude of 35,786 km in the equatorial plane would appear motionless when viewed from earth (as cited by Bruno Pattan in *Satellite Systems: Principles and Technologies*). Second, had Clarke obtained a patent in 1945, it would have expired in 1962, 17 years after the concept was first disclosed and two years before the first geostationary satellite, *Syncom III*, was successfully launched. Nevertheless, Clarke can rightfully claim credit for the first detailed technical exposition of satellite communications with specific reference to the geostationary orbit. His vision was realized through the pioneering efforts of such scientists as John Pierce of the Bell Telephone Laboratories, head of the Telstar program and co-inventor of the traveling wave tube amplifier, and Harold Rosen of the Hughes Aircraft Company, who was the driving force behind the *Syncom* program.

Since 1964, approximately 265 satellites have been launched into geostationary orbit, of which approximately 185 are operational. Another 67 GEO satellites are presently on order. The majority of these satellites have been used for the traditional fixed satellite service in C- and Ku-band, but also include satellites in the direct broadcast service, digital audio radio service, and mobile satellite service. In addition, numerous nongeostationary systems are in the process of deployment or have been proposed for a variety of consumer services, including mobile telephony, data gathering and messaging, and broadband applications. In May, 1997, 73 new GEO satellites were licensed for broadband services at Ka-band and last September applications for a dozen more systems were submitted to the Federal Communications Commission (FCC) for geostationary, nongeostationary, and hybrid satellite systems to provide broadband services at V-band. The total number of planned new satellites exceeds 1300.

The design of a satellite communications system presents many interesting alternatives and tradeoffs. The characteristics include the choice of orbit, the method of multiple access, the methods of modulation and coding, and the tradeoff between power and bandwidth. In this article, these choices will be briefly described and hopefully a sense of why satellite engineers find this field of endeavor so fascinating will be conveyed.

ORBIT

The system design begins with the choice of orbit. The orbital altitude regimes have been conveniently classified as Low Earth Orbit (LEO), Medium Earth Orbit (MEO), and geostationary orbit (GEO). The altitude of LEO is about 1000 km, or above the atmosphere but below the first Van Allen radiation belt. The altitude of MEO is ten times greater, that is 10,000 km, which lies between the first and second Van Allen belts. The altitude of GEO is uniquely 35,786 km as stated above. A fourth category is High Earth Orbit (HEO), which is at about 20,000 km and is above the second Van Allen belt but below GEO. (The acronym HEO has also been used to mean "highly elliptical orbit"; can we find a new term for this category? The progression LEO, MEO, HEO, GEO is quite appealing.)

Besides altitude, two other important orbital parameters are inclination and eccentricity. The inclination may be selected on the basis of maximizing the level of multiple satellite coverage. Elliptical orbits may be used with eccentricities designed to maximize the dwell time over a particular region.

The appropriate orbit is often suggested by the nature of the service, the business plan, or the constraints of the communications link. These properties are well illustrated by the variety of satellite mobile telephony systems under construction. Iridium is designed for continuous global coverage. This is a LEO constellation of 66 satellites in polar orbits at an altitude of 780 km. The choice of LEO was dictated by the desire to minimize power in both the satellite and the mobile handset, minimize the satellite antenna size, minimize the time delay, or latency, for a two-way signal, and maximize the angle of elevation. The orbital period is 100 minutes and a given satellite is in view for only ten minutes before handover of a call to a following satellite. An Iridium satellite has extensive onboard processing and a telephone call is routed through the constellation via intersatellite links.

Globalstar employs a constellation of 48 satellites in orbits inclined at 52° at an altitude of 1406 km. This system concentrates coverage over the temperate regions of the earth from 70° S to 70° N latitude. A technique called spatial diversity is used, wherein signals received simultaneously from two satellites are combined in the receiver to mitigate losses due to blockage and multipath effects. Thus an inclined, nonpolar orbit constellation was chosen to ensure that at least two satellites are visible at all times. The Globalstar system uses nonprocessing, or “bent pipe” satellites.

The third major mobile telephony satellite entry is ICO. This system will consist of 10 operational satellites in MEO at an altitude of 10,355 km. (The acronym ICO derives from the term "intermediate circular orbit", a synonym for MEO.) MEO is an excellent compromise between LEO and GEO. The satellite antenna size and power are relatively modest and the latency is still small. Yet the number of satellites required for global coverage is significantly less than LEO and the dwell time is considerably longer. The ICO orbit has a period of revolution of 6 hours and the time a satellite is in view is on the order of two hours.

Other satellite mobile telephony systems include ECCO and Ellipso. ECCO is a circular orbit constellation in the equatorial plane designed for communications in tropical regions. Ellipso employs elliptical orbits to maximize coverage over the northern hemisphere.

There is, nevertheless, a valid geostationary alternative for a mobile telephony satellite. The primary advantage is that the system can be built up on a regional basis. With only one satellite, an entire country or geographical region can be served. Although the two-way time delay can be over a half second and is quite perceptible, this is a defect that a population may be willing to accept if it is underserved by a terrestrial telephony system. An example is the Asia Cellular Satellite system (Aces) that is being built by Lockheed Martin for service to the Pacific Rim. To provide the required cellular coverage, the satellite antennas are about 12 meters across.

MULTIPLE ACCESS

Multiple access refers to the method by which many users share a common satellite resource. There are three primary methods: Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), and Code Division Multiple Access (CDMA).

With FDMA the available spectrum is divided among all of the users. Each user obtains a dedicated portion of the spectrum. FDMA can be used for either analog or digital signals.

With TDMA each user is assigned a time slot in a repetitive time frame. Data bits are stored in a buffer and are burst to the satellite during the assigned time slot. The signal occupies the entire transponder bandwidth. Because bits are stored during the portion of the time frame not assigned to the user, TDMA is inherently digital.

CDMA is a method in which the signal to be transmitted is modulated by a pseudorandom noise (PRN) code. The code rate is usually several orders of magnitude greater than the information bit rate. Their ratio is called the processing gain. The code spreads the signal over the full bandwidth available (hence CDMA is also called "spread spectrum") and all users share the same spectrum. The receiver modulates the signals from all users simultaneously with a replica PRN code. The desired signal is obtained by autocorrelation, while all of the undesired signals are spread over the full bandwidth and appear as white noise.

Frequency Division Multiple Access is relatively simple both conceptually and in terms of the hardware required. However, a transponder is a nonlinear device. This means that the output power is not merely proportional to the input power, but rather is represented by a curve that can be approximated by a third order polynomial. For multiple carriers, this nonlinearity generates harmonics that produce intermodulation interference among neighboring channels. In order to mitigate this effect, the input power is reduced in order to operate in the linear portion of the transponder output vs. input power characteristic so that intermodulation is reduced to an acceptable level.

The reduction in power is called "backoff". At a typical backoff of 6 dB, the input power is only one fourth the maximum possible value at saturation and the output power is correspondingly less. Backoff is not required when only one carrier occupies the transponder, such as a typical FM video channel, a TDMA carrier, or several channels multiplexed onto a single carrier at the earth station.

A major advantage of TDMA is that backoff is not required, since at any given time a single user occupies the full bandwidth of the transponder. Thus the output power of the transponder is much higher than with FDMA. Another advantage of TDMA is that it is more flexible. User allocations can be changed with relatively simple changes to software.

CDMA offers the potential of greater capacity. However, the theory of CDMA assumes that all users appear to contribute equally to the overall noise. Because users are at different distances with respect to one another, this assumption implies the need for dynamic power control. Another advantage is that the population of users need not be known in advance. As users are added to the system, the signal quality degrades slowly. Other advantages are that CDMA mitigates interference and enhances data security.

The mobile telephony satellite systems illustrate these alternatives. Both Iridium and ICO use a combination of FDMA and TDMA. The available spectrum is divided into sub-bands and TDMA is used within each sub-band. The capacity per satellite for Iridium is approximately 1100 simultaneous users, whereas each ICO satellite is designed to support at least 4500 telephone channels. Globalstar uses a combination of FDMA and CDMA (channelized CDMA). The available spectrum is divided into 1.25 MHz sub-bands and multiple users simultaneously occupy each sub-band via CDMA.

BANDWIDTH

There are numerous measures of bandwidth and one must be careful to distinguish among them. The noise bandwidth is the bandwidth the noise power would have if it were contained in a rectangle whose height is the peak spectral power density. The noise bandwidth B is the bandwidth relating the thermal noise power N to the system temperature T, such that $N = k\ T B$, where k is Boltzmann's constant.

The occupied bandwidth is the bandwidth required for the signal to pass through a band limited filter. In an FDMA system, it is the occupied bandwidth that determines the system capacity. The occupied bandwidth is about 1.2 times greater than the noise bandwidth. The extra margin is the value of the rolloff in the pulse shaping, which is used to minimize intersymbol interference (ISI). This type of interference is caused when the tails of preceding and following pulses overlap the peak of the observed pulse. Nyquist showed that ISI could be eliminated if the pulses followed a sin x/x function. In practice, this is impossible to achieve and is approximated by raised cosine pulse shaping.

A third measure of bandwidth is the null-to-null bandwidth. This bandwidth is the width between the zeroes of the main spectral lobe. Other measures of bandwidth, such as the half-power bandwidth, are also used.

FREQUENCY

The frequency is chosen on the basis of maximizing the performance of the system and exploiting the portions of the electromagnetic spectrum that are available. One important relation is that the gain of an antenna increases with increasing frequency for a fixed antenna size. On the other hand, the antenna gain is determined by the area of coverage. Thus once the area of coverage is specified, the gain is determined and then for a specified frequency the size of the antenna is determined.

It can be shown that for fixed transmit antenna gain and fixed receive antenna gain, the received carrier power is maximum when the frequency is minimum. These conditions apply to mobile telephony, since the satellite antenna gain is fixed by the terrestrial cell size and the handset antenna gain is fixed by the condition that the antenna must cover the entire sky. Thus L-band (1.6 GHz) is used because it is the lowest practical frequency that is available.

Another factor is the availability of spectrum. Initially, C-band (6/4 GHz) was used exclusively for the fixed satellite service. Later, Ku-band (14/12 GHz) was used, both because it was a frequency domain that was available to expand capacity and because the higher frequency permits the use of smaller earth terminal antennas. However, more power is required to overcome the detrimental effects of rain.

As the frequency increases the effects of rain increase. Rain degrades a satellite communication link in two ways: by attenuating the signal over the signal path and by increasing the system noise temperature of the earth terminal. Attenuation is caused by scattering and absorption of the electromagnetic waves. As the frequency increases, the wavelength decreases. To the extent that the wavelength is comparable to the size of a typical rain drop (about 1.5 mm), the signal becomes more susceptible to scattering and absorption. The system noise temperature increases because the antenna sees the warm rain at room temperature instead of the cold sky.

At C-band (6 GHz) the wavelength is 50 mm (5.0 cm) and the rain attenuation per kilometer of path is about 0.1 dB/km for a maximum rain rate of 22 mm/h, corresponding to an availability of 99.95 percent in Washington, DC. At Ku-band (14 GHz), the wavelength is 21 mm (2.1 cm) and the rain attenuation is 1 dB/km under the same conditions.

New satellite systems for broadband applications are in various stages of development. These new systems will extend the frequency domain into Ka-band and V-band. Rain attenuation increases dramatically at these frequencies. At Ka-band (30 GHz) the wavelength is 10 mm and the attenuation is 5 dB/km for 99.95% availability in Washington. At V-band (50 GHz) the wavelength is only 6 mm and the corresponding attenuation is 9 dB/km. It will thus not be possible to achieve the same availability at Ka-band and at V-band as we are accustomed to achieving at C-band or even Ku-band. Without mitigating techniques, such as spatial diversity and switching to lower frequencies, the availabilities will be in the neighborhood of 98% for any reasonable rain attenuation allowance. Note that in addition to attenuating the signal, the rain also increases the system noise temperature. This contribution to the total system degradation can be comparable in magnitude to the attenuation itself.

MODULATION

A sinusoidal electromagnetic wave has three properties: amplitude, frequency, and phase. Any one of these parameters can be modulated to convey information. The modulation may be either analog or digital. In analog signals, the range of values of a modulated parameter is continuous. In terrestrial radio systems, for example, AM and FM channels represent amplitude and frequency modulation, respectively. In digital signals, the modulated parameter takes on a finite number of discrete values to represent digital symbols. The advantage of digital transmission is that signals can be regenerated without any loss or distortion to the baseband information.

A fundamental parameter in digital communication is the ratio of bit energy to noise density E_b/N_0. This parameter depends on three characteristics: the bit error ratio (BER); the method of modulation; and the method of coding.

By far the most common form of modulation in digital communication is *M*-ary phase shift keying (PSK). With this method, a digital symbol is represented by one of *M* phase states of a sinusoidal carrier. For binary phase shift keying (BPSK), there are two phase states, 0° and 180°, that represent a binary one or zero. With quaternary phase shift keying (QPSK), there are four phase states representing the symbols 11, 10, 01, and 00. Each symbol contains two bits. A QPSK modulator may be regarded as equivalent to two BPSK modulators out of phase by 90°.

For *M*-ary PSK, the noise bandwidth is the information bit rate divided by the number of bits per symbol. Thus for uncoded BPSK modulation, the noise bandwidth is equal to the information bit rate; for uncoded QPSK modulation the noise bandwidth is one-half the information bit rate. The null-to-null bandwidth is twice the noise bandwidth in each case.

QPSK is usually preferred over BPSK because for a given bit rate and BER it requires the same power, yet requires only half the bandwidth. The saving in bandwidth using QPSK instead of BPSK without any greater power is the digital communication equivalent of a "free lunch". The tradeoff is actually added complexity in the modulator, but QPSK modulators are commonplace and the distinction between a QPSK chip and a BPSK chip is comparable to the distinction between a Pentium computer chip and an 80-286 computer chip: the Pentium chip is much more complex, yet it is ubiquitous and inexpensive.

In some situations BPSK might be preferred, such as when sufficient bandwidth is available and it desired to minimize the spectral power flux density to meet a regulatory requirement. BPSK is also used in CDMA systems, in which the basic principle is maximizing the bandwidth.

Higher order PSK modulation schemes are also used, such as 8PSK. With 8PSK the required bandwidth is only one third the bandwidth of BPSK or two-thirds the bandwidth of QPSK. However, the phase states are 45° closer than QPSK, which makes it more difficult for the receiver to distinguish them. Thus for a given BER the required power is higher than that of either BPSK or QPSK. For example, at a BER of 10^{-8}, 8PSK requires about 4 dB more energy per bit.

In *M*-ary PSK, symbols are distinguished from one another by the carrier phase, but the amplitude remains the same. It is possible to modulate both the phase and the amplitude in order to increase the number of bits per symbol and reduce the bandwidth even further. For example, in 16QAM there are twelve phases and four amplitudes. There are four bits per symbol and the bandwidth is one-fourth the bandwidth of BPSK or one-half the bandwidth of QPSK. However, like 8PSK, this method requires more power because it is more vulnerable to transmission impairments. For a BER of 10^{-8} the required E_b/N_0 is about 4 dB more than QPSK. These higher order levels of carrier modulation are being developed in an effort to decrease the required bandwidth and thus increase the bandwidth efficiency of satellite communication systems.

In offset QPSK (OQPSK) and minimum shift keying (MSK), discontinuous phase transitions are avoided to suppress out-of-band interference. These two methods have a constant envelope and are attractive when the intermodulation effects of transponder nonlinearities are to be minimized. Another alternative is frequency shift keying (FSK). With this method of modulation the frequency of the carrier assumes one of a discrete number of frequencies during each bit period.

CODING

The amount of power, as represented by E_b/N_0, can be reduced through the use of forward error correction (FEC) coding. The reduction in the value of E_b/N_0 is called the coding gain. The code rate is the ratio of information bits to the number of coded bits.

Two types of codes are used: block codes and convolutional codes. In a block code a group of information bits are accepted as a block to the encoder and parity bits are added to form a code word. Names associated with this type of code include Hamming, Golay, BCH, and Reed-Solomon. In a convolutional code, bits are added to a shift register continuously and affect the formation of coded symbols over several bit periods. The number of bit periods that a given bit occupies the shift register is called the constraint length. The optimum method of decoding employs the Viterbi algorithm.

It is now becoming common in advanced communications systems to use concatenated coding, involving both an inner convolutional code and an outer Reed-Solomon block code. The Reed-Solomon code detects and corrects bursty type errors. Interleaving is sometimes also used to scramble the bits after coding and unscramble them before decoding so as to cause bursty errors that occur in transmission to be spread out in time and make them appear to be random. However, interleaving introduces an increase in the encoding delay.

Coding reduces power at the expense of increased bandwidth. For example, a rate 1/2 code doubles the required bandwidth. Thus the bandwidth of a rate 1/2 coded signal using QPSK modulation is equal to the bandwidth of an uncoded signal using BPSK modulation. A rate 1/2 coded 8PSK signal requires 2/3 the bandwidth of uncoded BPSK or 2/3 the bandwidth of rate 1/2 coded QPSK.

BIT RATE

The information bit rate R_b is determined by the service or activity to be supported by the communications link. The required carrier to noise density ratio C/N_0 is related to the energy per bit to noise density ratio E_b/N_0 through the fundamental relation $C/N_0 = R_b\, E_b/N_0$. Thus for a specified bit rate -- together with the specified BER, method of modulation, and method of coding -- the required C/N_0 is determined.

On the other hand, the available C/N_0 provided on either the uplink or the downlink is determined by the transmitter equivalent isotropic radiated power (EIRP), the receiver figure of merit *G/T*, the free space loss, impairments due to rain, any other losses, and various forms of interference. The transmitter EIRP and receiver *G/T* must be designed to achieve the desired bit rate, or conversely, the given EIRP and *G/T* determine the bit rate that the link can support.

As an example, we return to the paradigm of telephony. For standard pulse code modulation (PCM) to convert a baseband analog waveform to a digital signal, the analog signal must be sampled at the Nyquist rate, or twice the highest baseband frequency, and each sample is encoded by n bits to represent one of $2^n - 1$ levels. For a high quality voice channel, the highest baseband frequency is 4000 Hz, and if each sample is encoded by 8 bits to yield 255 levels, the required bit rate is $2 \times 4000 \times 8 = 64{,}000$ bps, or 64 kbps. This is the classic bit rate for a voice channel.

This recipe for PCM actually applies to the analog-to-digital conversion of any waveform without any knowledge of the nature of the signal. In the particular case of human speech, however, it is possible to drastically reduce the required bit rate by modelling speech patterns. In a vocoder (or voice coder), perceptually important parameters describing the pitch, phonetic envelope, and level of vowel sounds are transmitted instead of the full digital representation of the analog waveform. Thus 4.8 kbps or even 2.4 kbps bit rates are possible. Since bandwidth is at a premium, these are the rates that will used in the satellite mobile telephony systems.

CONCLUSION

The design of a satellite communication system involves a wide variety of alternatives and tradeoffs. Often a particular set of choices will reflect a particular design philosophy or experience in some other field of communication. The mobile telephony systems illustrate how different designs can be adopted to achieve similar objectives. For example, Iridium is a LEO satellite constellation with polar orbits providing global coverage using FDMA/TDMA. Globalstar is also a LEO constellation but uses inclined orbits for concentration of coverage in mid-latitudes and employs CDMA technology. ICO is an FDMA/TDMA MEO constellation. Aces is a regional system using a single geostationary satellite.

These various possibilities keep the satellite engineer busy. The work, fortunately, is also highly interesting.

Dr. Robert A. Nelson, P.E. is president of Satellite Engineering Research Corporation, a satellite engineering consulting firm in Bethesda, MD. He is *Via Satellite's* Technical Editor.