1

The Challenge of Increasing Broadband Capacity

By Dale N. Hatfield, University of Colorado at Boulder

The recent release of the National Broadband Plan by the Federal Communications Commission (FCC) has focused the attention of policymakers, industry leaders, academics, and ordinary citizens on the importance of having sufficient bandwidth available anytime and anyplace to support a growing array of broadband services.

Broadband services include both wireline and wireless access to the Internet and the delivery of high-definition, even 3-D, television. As popular as these two terms—bandwidth and broadband—have become, and as important as they are to our future as a nation, they are not always well understood. The purpose of this essay is to explain these terms in more technical detail and then to relate the explanations to the opportunities and challenges that are associated with increasing fixed and mobile broadband capacity as envisioned in the National Broadband Plan.

This essay is divided into three sections. The first section discusses the critical relationship between the digital transmission capacity of a communications channel (as expressed in binary digits or "bits" per second) and the amount of bandwidth associated with that channel (as expressed in analog terms). The second section, in turn, builds upon that discussion to explore the opportunities and challenges associated with increasing the capacity of the four primary transmission technologies used in the critical access portion of the network: namely, twisted-pair copper cable, coaxial cable, wireless links, and fiber optic cable. The third section provides a summary and offers some concluding thoughts.

Understanding Critical Relationships

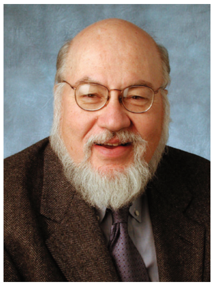
As discussed in the box on pages 8-9, both analog bandwidth as traditionally defined as well as digital bandwidth expressed as a bit rate determine how much information can be sent over a communications channel in a given amount of time. The two are related to one another by Shannon's Law. Shannon's Law is named after Claude Shannon, who is credited with being the founder of information theory, the basis of modern electronic communications. Expressed in words rather than in a mathematical formula, Shannon's Law states that the maximum amount of information that a circuit or channel can carry per unit of time (as measured in bits per second) depends upon the (analog) bandwidth and the strength of the desired signal relative to the strength of the accompanying undesired noise and interference, as measured at the receiving device.[1] For example, if the bandwidth of the channel is 1 MHz and the received power of the desired signal is 15 times as strong as the accompanying noise, then the maximum digital capacity of the 1 MHz channel would be 4 Mbps or 4 bps/Hz of bandwidth. Shannon's Law suggests two fundamental ways of increasing the digital capacity of a channel:

- Increasing the amount of bandwidth devoted to the channel, or
- Increasing the received signal level relative to the accompanying noise and interference.

Bandwidth increases, however, are often constrained by the technical characteristics of the transmission medium or, as in the case of wireless communications, by government regulation.

Increasing the digital capacity of a channel by increasing the transmitted power suffers from diminishing returns and from practical constraints. For instance, in the example given above, increasing the received power to 31 times as strong as the accompanying noise only increases the capacity to 5 bps/Hz. Moreover, such power increases are often impractical in the real world because of the increased interference that would be caused to other users of the same radio spectrum (i.e., the same channel), in the case of wireless communications. Moreover, many wireless devices today are battery-powered, and increasing their transmitted power can significantly decrease the length of time that the device can be operated without recharging the battery. While the base station to which the portable device communicates may be able to operate at a higher transmitter power, the power received at the base station is often limited by practical battery life considerations associated with the portable device.[2]

In the real world today, radio systems typically achieve efficiencies of between less than 1 bps/Hz to 7 bps/Hz or so depending upon the quality of the channel—for example, in terms of the received signal-to-noise/interference ratio. For example, today's comparatively high-powered digital television (DTV) systems achieve a transmission rate of almost 20 Mbps in a 6 MHz channel, for an efficiency of approximately 3.3 bps/Hz. In its recent report to Congress on the National Broadband Plan, the FCC reported that, over the years, the efficiencies of digital cellular radio systems have increased from much less than 0.1 bps/Hz at their inception to approximately 1.4 bps/Hz today. Modern cable systems using the DOCSIS 3.0 specification achieve a transmission rate of approximately 43 Mbps in a 6 MHz channel in the downstream direction (i.e., from the cable system "headend" to the subscriber's premises), for an efficiency of a little over 7 bps/Hz. The techniques used to improve efficiencies as



Dale N. Hatfield is the Executive Director of the Silicon Flatirons Center for Law, Technology, and Entrepreneurship and an Adjunct Professor in the Interdisciplinary Telecommunications Program—both at the University of Colorado at Boulder.

Prior to joining the University of Colorado, Hatfield was the chief of the Office of Engineering and Technology at the Federal Communications Commission (FCC) and, immediately before that, was chief technologist at the FCC. Mr. Hatfield retired from the Commission and government service in December 2000.

Before joining the Commission in December 1997, he was chief executive officer of Hatfield Associates, Inc., a Boulder, Colorado-based multidisciplinary telecommunications consulting firm. Before founding the consulting firm in 1982, Hatfield was acting assistant secretary of commerce for communications and information and acting administrator of the National Telecommunications and Information Administration (NTIA). Before moving to NTIA, Hatfield was chief of the Office of Plans and Policy at the FCC.

Mr. Hatfield has over four decades of experience in telecommunications policy and regulation, spectrum management, and related areas. Hatfield holds a B.S. in electrical engineering from the Case Institute of Technology and an M.S. in industrial management from Purdue University. Hatfield is currently serving as co-chairman of the Commerce Spectrum Management Advisory Committee.

Understanding Bandwidth

The term "bandwidth" originated in the analog world, where it is defined as a range ("band") of frequencies measured in cycles-per-second or Hertz (Hz). As a width, it represents the numerical difference between the upper and lower frequency limits of a channel of communications. In this context, a channel is a path used for the transmission of communications signals between two geographically separate points. For example, an ordinary telephone channel may have an upper frequency limit of 3.5 kHz and a lower frequency limit of .3 kHz and, hence, a audio frequency (AF) bandwidth of 3.2 kHz.[3] A high-fidelity audio amplifier, on the other hand, may have an upper frequency limit of 20 kHz and a lower frequency limit of 20 Hz—and thus a bandwidth of 19.98 kHz. In radio frequency (RF) communications, an ordinary television channel in the U.S. has a bandwidth of 6 MHz. For example, television Channel 2 occupies an RF range of 54 – 60 MHz and Channel 3 occupies an RF range of 60 – 66 MHz. In contrast, a single FM radio channel occupies an RF range of just 200 kHz.

In the analog world, signals and the channels they occupy are typically classified loosely and somewhat fluidly as narrowband, wideband, or broadband. A channel that is classified as broadband because of its greater width (say a television channel) can carry more information per unit of time than a channel classified as narrowband or wideband. A television signal made up of both video and sound (i.e., visual and aural) information contains more information than a simple audio signal associated with a telephone call. Hence, it requires more bandwidth to transmit in a given amount of time. Stated another way, a narrowband channel may be adequate to transmit an ordinary voice call but totally inadequate to transmit a television signal. In short, the more information it is desired to send in a given amount of time, the greater the analog bandwidth required.

Even though it is, strictly speaking, an analog expression, the term bandwidth has been carried over into the digital world. In the digital world, where information is carried as bits or "ones and zeros," the term bandwidth is also used to indicate how much information a channel can transmit in a given amount of time. However, in the digital world, bandwidth is measured in bits-per-second ("bps"). It is important to note that, in a digital network, the bandwidth is expressed as a *rate*—that is, how many events happen per unit of time. Other examples of rates include a pump that can discharge water at a rate of 10 gallons *per minute* or a bridge that can carry one thousand vehicles *per hour.* Stated again for emphasis, in the digital world, bandwidth refers to a transmission rate expressed in bits-per-second.

When digital networks are used to convey analog information, the analog signal is first converted to a digital signal at the originating end through a process known as analog-to-digital conversion and then, at the terminating end, the digital signal is converted back to an analog signal through a reverse process of digital-to-analog conversion. As in the case in the analog world, in the digital world, the digital signals and the channels they occupy are classified as narrowband, wideband, or broadband.

After the analog-to-digital conversion process described above, an ordinary voice signal requires a transmission rate on the order of a few tens of kilobits per second (kbps), while the transmission of a high quality still image in a reasonable amount of time may require a transmission rate of several hundreds kbps. A high-quality television signal at the other extreme may require on the order of several million bps (Mbps) for successful transmission in real time. Transmission rates in the tens-of-kbps range are typically categorized as narrowband, rates in the hundreds of kbps range are typically categorized as wideband, while rates in the millions of bps range are typically classified as broadband. So, to summarize in today's terms, in a digital network the term broadband is associated with a transmission rate of several Mbps or more.[4]

As emphasized earlier, in the world of digital communications, bandwidth is associated with a transmission *rate* but, confusingly, such rates are sometimes referred to as "speeds." That is, people often speak of "high-speed modems" or "high-speed networks" when they really mean high-bit rate modems or high-rate networks. "Speed" properly refers to the time it takes for an object—or, in the case of electronic communications, a signal—to travel from one point to another across an intervening space. In electronic communications, electromagnetic waves (e.g., RF signals) travel through space at the speed of light and via copper wires, coaxial cable or fiber optic cables, or other physical media at velocities that approach the speed of light.

In the digital world, some characteristic (or combination of the characteristics) of the transmitted electromagnetic/RF signal is rapidly changed to reflect whether the bit being sent is a one or a zero. The simplest digital transmission system to envision is one that sends a burst of electromagnetic/RF energy if the bit is a one and does not send a burst when the bit is a zero. In other words, the bursts of energy (or lack thereof) occur at regular intervals, representing a sequence of ones and zeros that correspond to the information being sent. In order to send information at a higher *rate* the intervals are shortened in time—that is, the ones and zeros are closer together in time and space—such that the transmission rate increases but the speed of transmitting individual bits remains the same, since the speed involved cannot exceed the speed of light. While the actual techniques used in the transmission of digital format are often much more sophisticated, the same basic principles apply.

Perhaps the distinction between speed and bit rate can be made clearer through an example drawn from the physical world. Consider, for example, a stream of semi-trailer trucks traveling down a highway at the speed limit of, say, 60 mph and assume that the trucks are physically spaced at intervals of one truck length. That would produce a "transmission" rate of some number of vehicles per hour. Now consider the same scenario, except that the trucks are replaced by Mini Cooper automobiles traveling at the same 60 mph and also spaced one vehicle-length apart. Because the vehicles, in this case the small Mini Coopers, are spaced much closer together than the semi-trailer trucks, the number of vehicles per hour would be much greater even though the speed has not changed. This is analogous to the digital transmission situation, where the bit rate—the number of bits-per-second or bps—is increased by spacing the ones and zeros closer together in time and space. The relationship between the maximum bit rate that a channel will support and the bandwidth of that channel (as that term is used in the analog world) will be explored later.

In the "network of networks" making up the Internet, the bits of information being transmitted are organized into sequences or units of bits called packets. In addition to the bits associated with the information being transmitted—say, a portion of an e-mail message or a voice telephone call—the packets include sequences of bits identifying the Internet address of the destination. These individually addressed packets of information are then routed from node to node between the origin and the destination in "store and forward" fashion. The time it takes for the packet to travel from the origin to the destination is known as latency. The irreducible or absolute minimum amount of time it takes for a packet to travel from the origin to the destination is constrained by the speed of light.

In addition to this irreducible minimum, the packets themselves may be processed several times at intermediary nodes along the way, thus adding to the latency. For example, an intermediary node, such as a packet switch or router, may store the packet briefly in order to read the address associated with the packet before forwarding it on to another node that is closer to the ultimate destination. This intermediary increases the latency. In addition, just as vehicles sometime encounter congestion on a highway during peak travel periods, the packets from different users and applications may encounter congestion delays as well, thus adding further to the total latency. To return to the highway analogy, a vehicle traveling at the speed limit may have to stop at a tollbooth and pay a toll. Moreover, there may not be enough individual tollbooths to handle all of the arriving vehicles, in which case additional delays may occur. The processing delay at the tollbooth and any congestion delays that occur upon arrival or departure from the tollbooth add to the total time it takes the truck to reach its destination.

Latency is an important measure of performance in some Internet applications, for example, in real-time voice applications and highly interactive games, and less important in other applications, e.g., e-mail or simple web browsing. The important point is that the minimum amount of latency *cannot* be decreased by increasing the bit rate, that is., by increasing the bandwidth. Stated another way, the minimum latency or delay between source and destination is constrained by the speed of light, and the transmission rate merely determines how fast the packet can be "unloaded" once it arrives. Thus two major performance measures associated with the Internet are speed/latency and bandwidth/data rate.[5]

measured in bps/Hz tend to increase the amount of transmitter energy that is emitted near the edges of the channel compared to the portion of the energy emitted near the center of the channel. Thus, just as higher transmitter power can cause increased interference to other (distant) systems operating on the same channel, increasing the capacity of a channel by increasing the efficiency (again, as measured in bps/Hz) can cause increased interference to other systems operating on adjacent channels. The key point is that the efficiencies and hence the total capacities of all of these communications systems are ultimately constrained by the Shannon limit.[6]

Before discussing the opportunities and challenges associated with increasing fixed and mobile broadband capacity, it may be useful to say a few words about the noise and interference that inevitably compete with the desired signal in a receiver and constrain the capacity in accordance with the Shannon limit just described. RF devices themselves internally generate a certain amount of electrical noise. This noise is produced by thermal agitation of electrons in a conductor. In an FM radio, this random noise would be heard as a hissing sound in the receiver, and since it is produced in the receiver itself, it would even be heard in a remote area where there were no broadcast stations operating on the channel.[7] This noise can be reduced by numerous techniques but it cannot be completely eliminated. In addition to this internal noise, the desired signal must compete with external sources of noise that typically enter the receiver through the same path (e.g., the antenna connection) as the desired signal. There are two broad classes of external noise—natural and man-made. Natural noise sources include, for example, electrical storms (lightning) which produce static in the AM band. Man-made noise in nearby radio channels can be produced by a host of sources including automobile ignition systems and rotating electrical machinery, and by devices like flourescent light fixtures. In effect, these devices act as miniature transmitters that produce unintended RF energy. In wireless systems, this external noise and interference generally decreases in importance at higher frequency bands in the radio spectrum.

In addition to this unintentional radiation, the receiver often must contend with the intentional radiation produced by other RF systems. In wireless systems, such interference may be produced by another transmitter operating on the same channel at some distance away or by "spillover" effects from transmitters operating on adjacent bands/channels as explained above. Also, from a practical point of view, receivers are not perfect in filtering out signals operating on adjacent bands/channels, thus allowing in additional interference from other systems. These unintended forms of man-made radiation can be looked at as a form of environmental pollution, and the FCC has adopted rules that seek to reduce the interference from such sources. Note that the policy objective should not be to eliminate all interference; indeed, that is impossible. Nor should the objective always be to minimize the probability of interference because that may be prohibitively expensive. Rather the objective, in principle at least, should be to choose the level of interference that optimizes the economic value of the resource.[8]

To summarize then, there are two primary "scarce" resources associated with the design of any digital transmission link:

- Received power
- "Analog" bandwidth

The underlying capacity of such a digital transmission link as measured in bps depends upon the analog bandwidth of the particular transmission media involved—twisted-pair copper cable, coaxial cable, wireless link, or fiber optic cable—and upon the strength of the desired signal relative to the noise and interference at the receiving device. The following section will review some of the opportunities and challenges associated with increasing fixed and mobile broadband capacity.

Opportunities and Challenges: Looking Ahead

It is generally agreed that, in many cases at least, the primary technical and economic bottleneck hindering greater availability of broadband facilities and services is in the access portion of the network—i.e., the portion of the network between the customer and a node in the network, which typically represents an aggregation point or point of traffic concentration. For example, in a wireless cellular system, this point might be a radio access node or base station; in the case of the telephone network, it might be a neighborhood terminal where the individual twisted-pair cables are terminated; and in the case of the cable network, it might be the location where the coaxial cables serving an individual neighborhood are terminated. Except in more remote, less-populated or other hard-to-serve areas, the challenges of connecting this access node to the balance of the network (i.e., to the core network) are typically less because upgrades in capacity are easier and there are greater opportunities to achieve economies of scale. For this reason, the rest of this essay will focus on this "last-mile" portion of the network—the portion between the access node just described and the customer premises.

The principle transmission media used to serve this portion of the network are those mentioned above:

- Twisted-pair copper cable
- Coaxial cable
- Wireless links
- Fiber optic cable

Since the underlying capacity of a digital transmission link depends upon the analog bandwidth of the particular transmission media and the signal-to-noise/interference ratio at the receiving device, it may be instructive to begin by focusing on these two fundamental factors in terms of each of the four types of transmission media just identified.

Twisted-Pair Cable

The original transmission media for both telegraph and telephone systems was open wirelines consisting of pairs of bare conductors that were tied to insulators attached to cross-arms on wooden poles. In the access part of the telephone network, these open wirelines eventually gave way to multipair cables consisting of individually insulated pairs of copper wires enclosed in a common protective sheath. Each wire pair is twisted together for technical reasons and, in a modern installation, these individual wire pairs are connected to the neighborhood access node mentioned above. Often a fiber optic cable is then used as the transmission media from the access node to the telephone company central office and the balance of the network. Because of the enormous bandwidth of the fiber optic cable (as explained in more detail below), it is the characteristics of the individual wire pairs in terms of their analog bandwidth and their noise/interference susceptibility that determine the digital bandwidth that can be delivered over the access portion of the network to the individual subscriber.

In terms of analog bandwidth, the individual wire pairs are capable of handling electrical signals in the AF range up to an RF frequency of perhaps one MHz or more. Stated another way, standard telephone wires have the bandwidth necessary to carry an ordinary analog voice conversation in one frequency range plus digital signals in the higher frequency range of the total available bandwidth. This sharing of the available bandwidth in the access portion of the network is the basis for the Digital Subscriber Line (DSL) service offerings of local telephone companies. The maximum bandwidth available on an individual wire pair depends upon a number of factors (e.g., the gauge of the wire) and is inversely proportional to distance. That is, the available analog bandwidth decreases with distance.

In terms of noise and interference, the individual wire pairs in an ordinary multipair telephone cable are not electrically shielded and thus are susceptible to picking up noise and interference from other nearby wire pairs ("cross-talk"), from other sources of man-made electrical noise and from RF sources such as nearby AM radio stations. While engineers and scientists have developed ingenious ways of dealing with these impairments, the combination of relatively limited analog bandwidth and a relatively hostile noise and interference environment has tended to limit DSL digital bandwidths to a few Mbps. Much higher data rates on the order of tens of Mbps are possible at shorter distances and by utilizing more than one wire pair from the neighborhood node to the individual subscriber.

Coaxial Cable

Early Community Antenna Television (CATV) systems used "twin lead" or "ladder lines" as the transmission medium for getting television signals from a central location (the "headend") to groups of subscribers. Early in the development of the industry, however, the transmission medium was changed to coaxial cable. A coaxial cable consists of a single wire conductor centered within a cylindrical, metallic outer conductor which serves as a shield. The two conductors are insulated from one another using various materials. As explained before, in a modern cable television network, coaxial cable is used as the transmission medium between a group of subscribers and a neighborhood access node. The connection between the access node and the headend typically utilizes a fiber optic cable, creating what is known as a Hybrid Fiber Coaxial (HFC) network.

In terms of analog bandwidth, the individual coaxial cables used in cable television networks are capable of handling electrical signals with an RF frequency of up to roughly one GHz. Thus one important advantage of coaxial cable as a transmission medium is its large analog bandwidth. However, the signals transmitted over the coaxial cables steadily weaken with distance and broadband analog amplifiers are required at regular intervals to boost the signals back up to suitable levels.

In terms of noise and interference, the shielded construction of the coaxial cable significantly reduces the susceptibility of the medium to noise and interference from external sources, including "over the air" wireless signals that utilize the same RF frequency range. While the shielded construction of the coaxial cable provides protection against external sources of noise and interference, the amplifiers that are required at regular intervals do produce some electrical noise and forms of self-interference (e.g., a form of interference known as "intermodulation") that must be dealt with by receivers at either end of the path. Because, to a great extent, only internal noise and interference must be dealt with, the environment is much more predictable and thus easier to mitigate. Moreover, the modern HFC architecture reduces the number of amplifiers required and, hence, the amount of internally generated noise and interference. Along with increasing internal noise and interference, the amplifiers employed also limit the available bandwidth. However, a modern HFC network typically has an upper RF frequency limit of 750 or 850 MHz.

In terms of digital bandwidth, it was noted above that modern (i.e., HFC) cable systems using the DOCSIS 3.0 specification are able to achieve an efficiency of slightly more than 7 bps/Hz. Thus, in gross terms, a modern cable system has an ultimate capacity of several Gbps. While the ultimate capacity of the coaxial cable is in the multiple-Gbps range even when constrained by the bandwidth of the associated amplifiers, two *caveats* are in order. *First*, reflecting the origins of the industry in retransmitting over-the-air broadcast signals, the bandwidth or spectrum associated with the cable is divided into scores of 6 MHz-wide channels and the division of those channels between the downstream (i.e., between the access node and the subscriber) and upstream channels (i.e., between the subscriber and the access node) is asymmetric. That is, the number of channels (and hence, the analog bandwidth) available in the upstream direction is considerably lower than the number of channels available in the downstream direction. Moreover, in the traditional architecture, a substantial fraction of the downstream channels is devoted to the delivery of entertainment television signals although, with advance digital signal compression techniques, the number of channels required to deliver such programming is significantly reduced. Over time, of course, these residual artifacts stemming from the origins of the industry can be removed and closer to the full capacity of the cable itself can be recovered.

Second, again reflecting the origins of the industry, the available capacity on the multiple coaxial cables emanating from a particular neighborhood node is shared among the various households served by that node. Sharing capacity made sense when a large fraction of the subscribers all wanted to view the same television program, say a championship sporting event like the Super Bowl game, at the same time. Said another way, it made sense to send a single copy of the program and have multiple subscribers tap into it rather than sending multiple but identical copies of the same program individually on separate facilities to each subscriber. When the HFC network is used to supply Internet access, however, the situation is much different. With the HFC architecture using DOCSIS modems, one or more of the shared 6 MHz channels is set aside for the provision of downstream Internet access. On this shared 6 MHz channel, different subscribers are independently receiving different digital content—aural, data, image, or video or combinations thereof—in individually addressed packets of information. The protocols or "rules of the road" prevent contending transmissions from interfering with one another. Since the transmission capacity is shared among multiple households in this manner, congestion and packet loss may result as traffic peaks and individual consumers contend for the available capacity.[9]

One fundamental way of reducing congestion in an HFC network is to decrease the number of households or customers served by each of the shared stretches of coaxial cable. This can be accomplished in two fundamental ways:

- First, imagine an access node that connects to three runs of coaxial cable collectively sharing the same capacity or channel and connected to a fiber optic cable link back to the headend. The capacity per household in this case can be increased by separating each of the three coaxial cable runs and connecting them to different fiber optic links running to the headend. As explained in more detail below, a fiber optic cable typically consists of multiple optical fibers, so such an approach may only require adding the necessary electronics to "light up" the additional individual fibers. In cable network parlance, this method of increasing capacity is known as "node splitting."
- Second, in new construction where it is easier to modify the physical layout of the network, the capacity of the HFC architecture can be increased by increasing the density of nodes and, hence, decreasing the number of households or customers served per coaxial cable run and access node. Note that this process moves the broadband fiber optic cable connection closer to the end users and allows more intense reuse of the RF bandwidth or spectrum carried within the individual coaxial cable runs.

Interestingly, each of these two techniques for reducing congestion and the associated latency has analogies in the wireless space and, to a lesser extent, the DSL space as well.

Wireless Links

It is generally agreed that the preferred frequency range for wireless RF communications used in the access portion of the network lies in the span of roughly 300 MHz to 3,000 MHz (3 GHz). Frequencies below this range are sometimes subject to long-range, highly variable interference due to atmospheric and ionospheric effects; require comparatively large antennas; and suffer to a greater extent from natural and man-made sources of radio noise and interference as explained earlier. Frequencies above this range tend to be blocked or attenuated more by intervening terrain and natural and man-made clutter such as foliage and buildings. At extremely high frequencies, above 10 GHz or so, additional attenuation due to rain and snow becomes an increasingly limiting factor in terms of readily achievable transmission distances.[10] Although gross generalizations about the suitability of various ranges of the spectrum for different uses are often suspect, it is widely agreed that, in terms of providing fixed and mobile access, the range from roughly 300 MHz to 3 GHz is the most desirable as indicated immediately above. Indeed it is so desirable that it is sometimes referred to as "beachfront property."

Of course this entire range of spectrum from 300 MHz to 3 GHz is not available for providing wireless broadband access. Large blocks of this spectrum range are already allocated to other important services and applications such as over-the-air radio and television broadcasting, the Global Positioning System (GPS), private and public safety mobile radio systems, weather radars, air navigation systems used to provide communications between pilots and air-traffic controllers on the ground, Digital Audio Radio Satellite (DARS) broadcasting, and Mobile Satellite Services (MSS) to name just a few. The net result is that only a fraction of the spectrum in the desirable 300 MHz to 3 GHz range is currently available for the provision of fixed and mobile broadband wireless access. In fact, in the recently released National Broadband Plan, the FCC indicates that only 547 MHz of spectrum is available for the provision mobile broadband services in the desirable frequency range below 3.7 GHz.[11]

The Obama administration has proposed, over the next 10 years, to nearly double the amount of spectrum available for commercial use—including for the provision of broadband wireless access—by reallocating some 500 MHz of spectrum currently held by the federal government and private companies. This proposal is consistent with recommendations contained in the National Broadband Plan and, presumably, the spectrum involved lies within the desirable range of 300 MHz to 3 GHz or so. If this proposal comes to fruition, it means that by the end of the decade there will be approximately 1 GHz of bandwidth available for the commercial provision of broadband wireless access within this desirable range. It is interesting and instructive to note that this net amount of spectrum is roughly comparable to the amount of analog bandwidth that is *currently* available on modern cable television systems that use the HFC architecture.[12]

Even though the net amount of analog bandwidth is roughly comparable (or will be if the Obama administration's proposal is implemented), there are important differences between the two methods of providing broadband access—coaxial cable and over-the-air wireless. Perhaps the foremost (and obvious) difference is the fact that a wireless network is able to offer mobility to its customers. However, there are significant other differences in terms of the quality and "usability" of the respective analog bandwidth and, as explained below, mobility increases the challenges associated with using the over-the-air spectrum.

Among the other differences is the fundamental one that the over-the-air spectrum available for the provision of wireless broadband access is scattered throughout different frequency ranges or bands and among different providers. That is, in contrast to the coaxial cable case, the spectrum available for broadband wireless access is neither contiguous (in the frequency dimension) nor controlled by a single entity. This, in turn, produces two related technical issues:

- First, as explained before, practical transmitters operating in one channel/band inevitably produce spillover interference into adjacent channels/bands and practical receivers are unable to totally reject radio signals that are emitted in adjacent channels/bands, even if the energy from those signals is contained entirely within those channels or bands. The lack of perfect transmitters and receivers inevitably results in the loss of some capacity between bands due to the need to, for example, provide a buffer or guard band between the two bands to supply the necessary isolation and thereby reduce the associated interference to an acceptable level. Excessively fragmenting the spectrum among different bands reduces the overall technical efficiency of spectrum utilization for this reason.

- Second, the problem is further compounded in the over-the-air case by what is known as the near-far problem. The problem arises when a user is attempting to receive a weak signal from a distant transmitter while in the immediate proximity of a transmitter operating on an adjacent (in frequency) channel. In such a situation, the signal from the nearby station may be so strong that it overwhelms the ability of the user's receiver to reject that signal and successfully receive the weaker, desired signal from the distant transmitter.

These types of adjacent channel/band interference problems are considerably reduced in the case of cable television systems because, being under the control of a single operator, adjacent channel signals can be adjusted by the cable operator to arrive at the receiver at close to the same signal strength; there is no near-far interference problem with which to deal.

In addition to the more benign adjacent channel/band interference environment associated with the spectrum on a cable television system, the shielding associated with the coaxial cable also protects desired signals internal to the cable from external sources of natural and man-made RF noise as well as interference from other users of the same frequency ranges that are external to the cable system. Such noise and interference can be a major factor in limiting the performance (i.e., the digital transmission capacity) of a given amount of spectrum.

Moreover, RF signals transmitted over-the-air—wireless signals—suffer from other impairments, the most fundamental of which is the steady weakening of the signal as the receiver is physically moved further away from the transmitter. Under ideal conditions (i.e., situations in which there is a totally clear, unobstructed path between the transmitter and the receiver), the strength of the transmitted signal decreases following what is known as the "inverse square rule." This rule, based upon the physics involved, predicts that doubling the distance decreases the received signal power by a factor of four. Under real-world conditions (i.e., in the presence of physical obstacles such as hills and buildings and clutter such as trees), the strength of the transmitted signal typically decreases with distance at a much faster rate. For example, in an urban area, doubling the distance may reduce the received signal power by a factor of 16 rather than four. The strength of the signals on a coaxial cable network also decrease with distance but, in a cable television network, the signals are boosted by broadband amplifiers at regular intervals so that the signal levels remain comparatively strong from the transmitter to the receiver.

This rapid falloff of signal strength with distance in the over-the-air environment has important network architecture implications. Recall once again that, according to Shannon's Law, the maximum data rate (digital bandwidth) that a channel will support depends upon the strength of the desired signal relative to the received noise/interference and the (analog) bandwidth of the channel. This means that, as the over-the-air signal rapidly weakens with distance, the maximum digital bandwidth that is achievable over the path decreases as well. If the transmitter is at a fixed location where primary power from the fixed electrical grid is readily available, the rapid falloff of signal strength can be compensated for, up to a point at least, by an increase in transmitter power.[13] In mobile applications with portable handsets, however, the transmitter power is typically constrained by battery size and life and other considerations such as limits on human exposure to nearby RF signals. Thus, at the edge of a coverage area, it is not the amount of bandwidth that limits the maximum digital transmission rate, but rather the strength of the transmitted signal from the handset as received at the base station receiver. This suggests that, in order to achieve broadband data rates while efficiently using the available bandwidth with portable handsets such as smartphones, the coverage areas must be kept small.

In addition to coping with adjacent channel/band interference problems, sources of external noise and interference, and rapid attenuation of signals with distance, over-the-air wireless systems must also contend with other RF transmission impairments found in an uncontrolled environment. These impairments include what is known as multipath (and its attendant consequences). Simply stated, multipath is a propagation phenomenon that results in RF signals arriving at a receiver by two or more paths.[14] One common cause of multipath is when the transmitted signal travels to the receiving antenna directly over what is known as a "line of sight" path and also via a reflection from a terrestrial object such as a mountain or building. Since the latter path is longer than the direct path, there is an additional delay in the amount of time it takes the reflected signal to arrive at the receiver. Depending upon the time spread between the direct and reflected signals, the copies of the transmitted signal arriving at the receiver may combine in a constructive or deconstructive fashion. As the two terms imply, when the combination is constructive, the total received signal gets stronger and when the combination is destructive the total received signal gets weaker. If the location of the transmitter and/or the receiver is changed, the geometry of the paths and, hence, the relative delay changes as well. This means that the combined signal at the receiver will vary over the immediate geographic coverage area. This variation in signal strength is known as fading and is familiar to users of wireless handsets when it manifests itself in the form of dropped calls.[15]

It should be noted that all of these problems tend to be exacerbated in a wireless mobile environment when, as is often the case, the end-user terminal or handset is in motion. For example, a user may go around the corner of a building and, in the process, move from a location where the signal from a base station is very weak (because the received signal is in the "shadow" of the building) to a location where the signal is very strong (because it is within the line of sight of the base station). Similarly, the co-channel and adjacent channel interference environment may change rapidly because of such movement and because of the changing multipath conditions as explained above. In short, mobility substantially increases the challenges associated with wireless communications and, while the effects of the associated impairments can be reduced (or, in the case of multipath, even exploited), there are typically penalties in terms of the added processing power required in the handset (with associated impact on battery life), added latency associated with the time it takes for the additional processing, and in advanced solutions requiring multiple antennas, added physical size.

So, to briefly summarize, over-the-air wireless systems face a significantly harsher signal environment compared to a "closed" coaxial cable-based system carrying RF signals and, in turn, over-the-air wireless systems serving mobile as opposed to fixed terminals face a still harsher signal environment. The effects of this progression can be seen in relative spectrum efficiencies (as measure in bps/Hz) achieved by cable television systems, over-the-air television broadcasting systems, and mobile wireless (cellular) systems reported earlier in this essay.

As noted, if the Obama administration's proposal to free up 500 Mhz of additional spectrum for wireless broadband uses is adopted, there will be approximately 1,000 MHz of analog wireless spectrum available for broadband wireless access in the desirable frequency range below 3 GHz or so. While, at first glance at least, this may seem like an abundance of spectrum, it may not be when viewed against the backdrop of the recent exponential increases in broadband wireless demand. To put this in perspective, assume for a moment that each wireless broadband user is consuming 1 Mbps of digital bandwidth and that the spectrum efficiency achieved in the mobile environment is 1 bps/Hz. Under these not unreasonable conditions, 1,000 MHz of bandwidth would only support 1,000 simultaneous users if, because of interference considerations, the spectrum could only be used once over a large geographic area. Clearly, this would be a woefully inadequate amount of capacity for a major metropolitan area like New York or Los Angeles, where there may be literally millions of subscribers. Indeed, in such heavily populated areas, the capacity would be inadequate even if the users were communicating using narrowband digital voice. While higher levels of spectrum efficiency can be achieved, the increases are constrained by Shannon's Law (the "laws of physics") and, in the upstream direction, even more so by the previously noted practical limits on handset transmitter power imposed by battery life and human RF exposure considerations.

Unlike traditional high-power, high-antenna-height wireless systems, which only allowed spectrum to be used once in a large geographic area, the more modern cellular mobile radio systems that emerged in the 1980s employ frequency reuse in a "cellular" configuration or architecture. Cellular mobile radio systems get their name from the notion of dividing the large geographic area into a series of small, geographically contiguous coverage areas called cells. Relatively low-power base station transmitters and receivers with relatively low-height antennas (towers) are placed in each cell and connected by wireline or microwave facilities to the balance of the network and, through interconnection arrangements, to other fixed and mobile networks as well. The relatively low-power/low-antenna heights of the base stations match the radio coverage to the area of these cells and the base stations communicate with the mobile units (e.g., cellular handsets or "smartphones") in their respective cells. Because of the low-power/low-antenna heights involved, co-channel interference is minimized and, hence, the same channel can be used simultaneously (i.e., reused) for different conversations or data sessions in different cells within the larger geographic area.

In contrast to the traditional systems of the 1970s and earlier, frequency reuse and the cellular architecture enables the scarce spectrum resource to be utilized in a much more intense manner. That is, rather than being used just once in a large geographic area, the assigned spectrum is reused many times, thus multiplying the capacity in that area. Another important feature of the cellular architecture is that, as demand develops, a larger number of still smaller cells (i.e., cells covering even smaller geographic areas) can be employed to increase the total capacity of the system. Or, stated another way, increased frequency reuse (and hence, increased efficiency) through "cell division" can be employed to increase capacity over time.

Such increases in capacity through increased frequency reuse are not insignificant; indeed, the opposite is true because the increases in capacity are exponential. For example, early cell sites (and those cell sites in more rural areas today where spectrum capacity is not yet a consideration) might have had a radius of ten miles. Decreasing the radius to 1 mile increases the capacity 100-fold and decreasing it to 0.1 miles increases it by 10,000.[16] As explained before, reducing the distance between the handset and the base station also allows much higher signal strengths at the receiver, thereby increasing the digital bandwidth available at the edge of the coverage area and/or reducing the power consumed by the handset. Thus, while further increases in broadband wireless capacity can be partially met by (a) doubling the amount of available bandwidth allocated to such services as recently proposed and (b) increasing the efficiency with which the available bandwidth is used by more closely approaching the Shannon limit as a practical matter, continued dramatic increases in broadband wireless demand will have to be met through the exponential increases in capacity associated with smaller cell sizes. Evidence suggests that the marketplace is already responding in exactly this way. Broadband wireless providers are not only seeking more spectrum and, with their vendors, further improving spectrum efficiency within the Shannon limits but also turning to microcells, pico cells and femto cells as well as to "smart antennas" and outdoor Distributed Antenna Systems—all of which permit much more intense frequency reuse.

It is interesting to note that enhancing wireless capacity through increased frequency reuse is analogous to how the cable television industry increases its capacity by node splitting or increasing the geographic density of the access nodes as explained earlier. In the wireless case and in the coaxial cable case, increasing the number of access nodes allows, respectively, the available over-the-air spectrum and the spectrum within the shielded coaxial cable to be used more intensely. In both the wireless and cable cases, shortening the access portion of the network—the distance between end user and the access node—implies the need for a denser fiber network to carry the traffic between the access nodes and the balance of the network.[17] Additionally, shortening the access portion of the network in the case of DSL would facilitate greater bandwidths for the reasons given earlier.

Fiber Optic Cable

In the descriptions of the three access technologies addressed so far, it was implicitly assumed that a fiber optic cable would be used to carry the broadband digital signals from the neighborhood node to the balance of the network.[18] In each case, the technical challenges associated with the three technologies in providing the connection between the neighborhood node and the individual subscriber were addressed. Although there are economic challenges in doing so, fiber optic cables can be used in this portion of the network in what is typically referred to as the Fiber-to-the-Home (FTTH) configuration or architecture.

Optical fibers consist of a very fine cylinder of glass, called the "core," surrounded by a concentric layer of glass, called the "cladding," and, as in the case of twisted-pair copper cable, multiple individual fibers are often grouped together to form a fiber optic cable. In terms of analog bandwidth, the individual fibers making up the cable are, depending upon the type of optic fiber deployed, capable of handling optical signals with a total analog bandwidth in the terahertz (THz)[19] or even tens of THz range. Although these rates are primarily associated with long-haul or core networks, the basic technology permits extremely high transmission rates in the local access network as well.

In addition to their enormous advantage in terms of analog bandwidth, optical fibers also have the benefit that, since they use optical signals rather than electrical signals, they do not suffer from the natural and man-made forms of electrical noise or interference described earlier. This immunity to electrical noise and interference results from the fact that the individual optical fibers are made of glass and do not conduct electricity. In terms of digital bandwidth, the combination of substantial analog bandwidth coupled with the immunity to electrical noise and interference means that, even in residential FTTH applications, an individual optical fiber is capable of transmission rates on the order of several Gbps. The actual transmission rate delivered to the subscriber depends upon, among other things, the details of the architecture. One popular architecture called a Passive Optical Network or "PON," divides the available transmission capacity among multiple subscribers—while another architecture, known as Point-to-Point (PtP), is capable of delivering the entire transmission capacity of the fiber to an individual subscriber. The former provides digital bandwidths in the tens of Mbps range while the latter provides digital bandwidths in the hundreds of Mbps range.

Summary and Conclusions

The recent release of the National Broadband Plan by the FCC has focused attention on the importance of having sufficient bandwidth available anytime and anyplace to support a growing array of broadband services that are critical to the nation's economic and social well being and to public safety and homeland security as well. This essay describes the relationship between the maximum digital transmission capacity of a channel and the two fundamental factors that determine that capacity. These fundamental factors are the analog bandwidth of the channel and the strength of the received signal relative to the noise. Building upon this understanding, this essay explored technical opportunities and challenges associated with increasing the capacity of four transmission technologies used in the critical access portion the network: namely, twisted-pair copper cable, coaxial cable, wireless links, and fiber optic cable.

The opportunities and challenges associated with increasing the capacity of each of the four technologies described above extend beyond the pure technical characteristics of the transmission media, associated electronics, and overall architecture to include not only cost but operational factors such as the ease with which the provider can gain access to the necessary public and private rights-of-way or access to added base station antenna sites in the case of wireless systems. It is beyond the scope of this essay to address these latter factors in any detail. Rather, in the balance of this section, the opportunities and challenges associated with the technical aspects of the four digital transmission technologies will be summarized and characterized.

Of the four technologies discussed in this essay, ordinary multipair copper cable used in the traditional telephone network is highly constrained in terms of analog bandwidth at longer distances and, because the wire pairs are unshielded, it is susceptible to both internally and externally generated electrical noise and interference. As Shannon's Law suggests, the resulting digital transmission bandwidth using DSL technology is limited to a few Mbps. While higher data rates are possible at shorter distances and by utilizing more than one wire pair to the customer, further increases in capacity are apt to be incremental and dramatic increases problematical because of the limited bandwidth and difficult noise and interference environment. An advantage of the DSL architecture, though, is that the entire capacity on a wire pair within the cable is available to a single customer—that is, it is unshared.

The second technology, coaxial cable, in contrast has a much larger analog bandwidth—on the order of 1 GHz—and its shielded construction reduces its susceptibility to externally generated electrical noise and interference. While the inherent bandwidth of the coaxial cable employed is large, in a traditional cable television network, amplifiers are required at regular intervals to maintain the transmitted signals at acceptable levels. These amplifiers generate electrical noise and forms of self-interference but, unlike the situation with wireless, over-the-air transmission, the noise and interference environment internal to the cable is much more predictable and, hence, easier to mitigate. Again, as Shannon's Law would suggest, with an analog bandwidth on the order of 1 GHz and a comparatively benign and controlled noise and interference environment, the resulting digital transmission bandwidth of a cable television network, optimally configured, is several Gbps. Artifacts stemming from the origins of the industry in delivering over-the-air broadcast television signals—e.g., the placement of the signals or channels within the available analog bandwidth and the need to continue to carry such signals in the analog format—reduce the current digital capacity. However, in the longer term it should be possible to exploit the full digital transmission capacity of the cable itself. A disadvantage of the cable television architecture is that numerous stretches of the coaxial cable—and hence, the digital transmission capacity—are shared among multiple customers. However, as explained in more detail above, the number of customers sharing the available capacity can be decreased and, accordingly, the digital transmission capacity per customer increased through node-splitting and shortening the distance spanned by any single stretch of cable.

The analog bandwidth available to the third technology, wireless links, depends upon the amount of spectrum allocated by the government for the provision of fixed and mobile wireless broadband access. In the United States, only 547 MHz of spectrum is currently available for such purposes.

Recently, though, the Obama administration proposed the reallocation of 500 MHz of additional spectrum which, if the reallocation is fully carried out, would bring the total analog bandwidth available to approximately 1 GHz. While wireless systems have the obvious advantage of offering mobility, their use of the over-the-air spectrum presents a myriad of challenges as described earlier. These challenges include the scattering of the available bandwidth across different bands within the desirable range and coping with adjacent channel/band interference problems, sources of external noise and interference, rapid attenuation of signals with distance, and over-the-air propagation impairments such as multipath. In short, wireless systems face a much harsher signal environment compared to "closed" coaxial cable-based and fiber optic cable-based systems. Over-the-air wireless systems serving mobile as opposed to fixed terminals face a still harsher signal environment because of the additional variability of the signal propagation conditions, and the net result is still lower spectrum efficiencies (as measured in bps/Hz). Because of the harsher signal environment, the rapid falloff of signal strength with distance, and the constraints on up-link transmitter power associated with mobile handsets, the overall digital bandwidth achievable in a mobile environment is significantly reduced compared to closed coaxial cable- and fiber optic, cable-based systems even when the available analog bandwidth is roughly the same.

The fourth technology, fiber optic cable, is generally regarded as the "gold standard" in terms of increasing broadband digital access capacity because of its enormous analog bandwidth and its immunity to natural and man-made forms of electrical noise and interference. The actual digital transmission rate delivered to or from a customer depends upon the details of the architecture employed, but the ultimate capacity is limited more by economic factors rather than by the inherent technical constraints on the underlying technology imposed by Shannon's Law. In this regard, fiber optic cable is often referred to as being "future-proof" because the maximum digital transmission rates are governed more by the electronic equipment attached to the cable rather than by the actual fiber itself. It is future-proof in the sense that the capacity can be increased by upgrading the associated electronic equipment rather than by taking the more expensive step of replacing the fiber itself.

With this background on the four technologies as summarized and characterized above, it seems apparent, from a technical standpoint at least, that (a) the exploding demand for wireless broadband capacity and (b) the constraints on the availability of the over-the-air wireless spectrum and the challenges associated with achieving dramatic improvements in spectrum efficiencies in the mobile wireless environment suggests a compelling need to meet a potentially large portion of that demand through even more intense frequency reuse in local geographic markets. More intense frequency reuse and greater edge-of-coverage area capacity can be accomplished with smaller cells and, in part, more directive antennas at the base station or access node. If seamless coverage is to be maintained, smaller cells translate directly into a requirement for a much higher density of access nodes. Moreover, both DSL and cable modem technology benefit from the shorter distances that are associated with a more dense deployment of their access nodes. This suggests the growing need to extend fiber optic cable capacity closer to the customer—either fixed or mobile—to minimize the distance between the customer and the access nodes.

This, in turn, further suggests that policymakers not only need to focus on the oft-stated long-term goal of encouraging FTTH but also on the more immediate need to bring fiber significantly closer to the customer to support a vastly increased number of access nodes. This is particularly important in the wireless case, where the capacity added through frequency reuse is critical to facilitating wireless competition with the two major suppliers of fixed broadband capacity—the incumbent telephone and cable television companies.

Endnotes

1. Certain types of man-made noise and interference may exhibit some regularity or predictability that permits them to be dealt with more effectively at the receiver. Strictly speaking, Shannon's Law assumes that the undesired signal is random rather than having some degree of predictability.
2. As discussed in more detail later, the lower transmitted power often associated with a wireless handset reduces the maximum transmission rate achievable when the device is at the edge of its coverage area and the signal received at the base station is the weakest.
3. Note that 1 kHz = 1,000 Hz, 1 MHz = 1,000,000 Hz, and 1 GHz = 1,000,000,000 Hz.
4. To put this transmission rate into perspective, the National Broadband Plan suggests a goal of having at least 100 million U.S. homes with affordable access to actual download speeds of at least 100 Mbps and actual upload speeds of at least 50 Mbps by the end of the decade.
5. A third important measure of performance in the Internet is packet loss. Just as vehicles sometimes get lost on their way to their destination, packets are sometimes lost as well. For example, a packet may be discarded at an intermediary node if it has already been delayed excessively or the amount of temporary storage available at the node is insufficient.
6. Another technique that can be used to increase the capacity of a link is digital signal compression. Compression is a digital signal processing technique that can reduce the transmission rate required to convey a signal in a given amount of time without unacceptable loss of quality. Digital compression techniques work, for example, by removing redundant information that may be present in the signal to be transmitted.
7. In a television receiver, such noise manifests itself as "speckles" or "snow" in the picture.
8. See Ronald H. Coase, *The Federal Communications Commission,* 2 J. L. & ECON. 1. 27 (1959) ("It is sometimes implied that the aim of regulation in the radio industry should be to minimize interference. But this would be wrong. The aim should be to maximize output.")
9. Although this discussion deals with possible congestion/latency in the downstream direction, similar considerations apply in the upstream direction as well.
10. The higher-frequency spectrum is useful for many purposes especially when there is a line-of-sight path between the two ends of a communications link. It is generally less suitable for longer-range communications, especially when the link is close to the ground and mobile devices are involved.
11. The FCC's National Broadband Plan recommended that 500 Mhz of spectrum be made available for broadband within 10 years, of which 300 Mhz should be made available for mobile use within five years.
12. As explained before, it is important to realize that, while both cable television systems and broadband wireless access systems both use RF spectrum, interference between the two uses of the spectrum is largely prevented by the metallic shielding associated with the coaxial cable transmission media used in cable television systems.
13. Directive antennas that concentrate the available transmitter power in the direction of the receiver can also be used to increase the received signal strength.
14. This definition of multipath is based upon Federal Standard 1037C: Glossary of Telecommunications Terms (available at http://www.its.bldrdoc.gov/fs-1037/).
15. The variation of signal strength or fading produces the familiar phenomenon of cell phone users moving around trying to find a good spot to make or maintain a cellular call.
16. This exponential increase is due to the fact that the area of a circle increases with the square of the radius. Thus, for example, decreasing the average cell radius by one-half decreases the coverage area of each cell by one-fourth and increases the amount of frequency reuse obtained by a factor of four. This increase may come at the expense of additional infrastructure investment, of course.
17. A recent Aspen Institute report reached a similar conclusion. See *Rethinking Spectrum Policy: A Fiber Intensive Wireless Architecture,* Mark MacCarthy Rapporteur, The Aspen Institute, Washington, DC (2010). ("A fiber intensive wireless network architecture should be considered, in parallel with the allocation of additional spectrum, as complementary long-term solutions to the problem of exploding demand for wireless services.")
18. In some situations, especially in more remote areas, broadband point-to-point microwave systems may continue to play an important role in carrying traffic from the access node to the balance of the network.
19. A terahertz is equal to 1,000 GHz or 1,000,000 MHz.