Chapter 1

Wireless Networks

Future wireless networks will enable people on the move to communicate with anyone, anywhere, at any time, using a range of multimedia services. The exponential growth of cellular telephone and paging systems coupled with the proliferation of laptop and palmtop computers indicate a bright future for such networks, both as stand-alone systems and as part of the larger networking infrastructure. In this chapter we provide an overview of wireless networks. We begin in Section 1.1 with a brief introduction to these networks, including their history, their future, and their technical challenges. The greatest technical challenge to wireless network design is the underlying wireless channel. In Section 1.2 we describe the main characteristics of the wireless channel and how these characteristics impact the link layer and network design. Link layer design techniques that overcome wireless channel impairments to deliver high data rates with low distortion are described in Section 1.3. The wireless channel is a limited resource that must be shared among many users. In Section 1.4 we describe channel access protocols for sharing this resource in an efficient and fair manner. Section 1.5 outlines design issues for wireless networks, including network architecture, user location and routing protocols, network reliability and QOS, internetworking between wireless and wired networks, and security issues. Current wireless network technology is described in Section 1.6 including cellular and cordless telephones, wireless LANs and wide-area data services, paging systems, and global satellite systems. Section 1.7 gives an overview of emerging systems and standards for future wireless networks. We conclude the chapter in Section 1.8 with a summary and discussion of the design trends and challenges for future wireless networks.

1.1 Introduction

Wireless communications is, by any measure, the fastest growing segment of the communications industry. As such, it has captured the attention of the media and the imagination of the public. Cellular phones, cordless phones, and paging services have experienced exponential growth over the last decade, and this growth continues unabated worldwide. Wireless communications has become a critical business tool and part of everyday life in most developed countries. In addition, wireless communication systems are rapidly replacing antiquated wireline systems in many developing countries. Will future wireless networks live up to their promise of multimedia communications anywhere and any time? In this introductory section we will briefly review the history of wireless networks, from the smoke signals of the Pre-industrial age to the cellular, satellite, and wireless data networks of today. We then discuss the wireless vision in more detail, including the technical challenges that must be overcome to make this vision a reality.

1.1.1 History of Wireless Networks

The first wireless networks were developed in the Pre-industrial age. These systems transmitted information over line-of-sight distances (later extended by telescopes) using smoke signals, torch signaling, flashing mirrors, signal flares, or semaphore flags. An elaborate set of signal combinations was developed to convey complex messages with these rudimentary signals. Observation stations were built on hilltops and along roads to relay these messages over large distances. These early communication networks were replaced first by the telegraph network (invented by Samuel Morse in 1838) and later by the telephone. In 1895, a few decades after the telephone was invented, Marconi demonstrated the first radio transmission from the Isle of Wight to a tugboat 18 miles away, and radio communications was born. Radio technology advanced rapidly to enable transmissions over larger distances with better quality, less power, and smaller, cheaper devices, thereby enabling public and private radio communications, television, and wireless networking.

Early radio systems transmitted analog signals. Today most radio systems transmit digital signals composed of binary bits, where the bits are obtained directly from a data signal or by digitizing an analog voice or music signal. A digital radio can transmit a continuous bit stream or it can group the bits into packets. The latter type of radio is called a *packet radio* and is characterized by bursty transmissions: the radio is idle except when it transmits a packet. The first packet radio network, ALOHANET, was developed at the University of Hawaii in 1971. This network enabled computer sites at seven campuses spread out over four islands to communicate with a central computer on Oahu via radio transmission. The network architecture used a star topology with the central computer at its hub. Any two computers could establish a bi-directional communications link between them by going through the central hub. ALOHANET incorporated the first set of protocols for channel access and routing in packet radio systems, and many of the underlying principles in these protocols are still in use today. The U.S. military was extremely interested in the combination of packet data and broadcast radio inherent to ALOHANET. Throughout the 70's and early 80's the Defense Advanced Research Projects Agency (DARPA) invested significant resources to develop packet radio networks for tactical communications in the battlefield. These activities peaked in the mid 1980's, but the resulting networks fell far short of expectations in terms of speed and performance. DARPA all but abandoned packet radio research by the end of the 1980's, and today packet radio networks are mostly used by commercial providers of wide-area wireless data services. These services, first introduced in the early 1990's, enable wireless data access (including email, file transfer, and web browsing) at fairly low speeds, on the order of 20 Kbps. The main players in this arena are ARDIS, RAM Mobile Data, CDPD, and Metricom. All of these service providers with the exception of Metricom provide coverage in most metropolitan regions of the U.S., while Metricom is limited to the San Francisco Bay Area, Seattle, and Washington D.C.. The market for these wide-area wireless data services is relatively flat, due mainly to their low data rates, high cost, and lack of "killer applications".

The introduction of wired Ethernet technology in the 1970's steered many commercial companies away from radio-based networking. Ethernet's 10 Mbps data rate far exceeded anything available using radio, and companies did not mind running cables within and between their facilities to take advantage of these high rates. In 1985 the Federal Communications Commission (FCC) enabled the commercial development of wireless LANs by authorizing the public use of the Industrial, Scientific, and Medical (ISM) frequency bands for wireless LAN products. The ISM band was very attractive to wireless LAN vendors since they did not need to obtain an FCC license to operate in this band. However, the wireless LAN systems could not interfere with the primary ISM band users, which forced them to use a low power profile and an inefficient signaling scheme. Moreover, the interference from primary users within this frequency band was quite high. As a result these initial LAN systems had very poor performance in terms of data rates and coverage. This poor performance, coupled with concerns about security, lack of standardization, and high cost (the first network adaptors listed for $1,400 as compared to a few hundred dollars for a wired Ethernet card) resulted in weak sales for these initial LAN systems. Few of these systems were actually used for data networking: they were relegated to low-tech applications like inventory control. The current generation of wireless LANS, based on the IEEE 802.11 standard, has better performance, although the data rates are still low (on the order of 2 Mbps) and the coverage area is still small (around 500 ft.). Wired Ethernets today offer data rates of 100 Mbps, and the performance gap between wired and wireless LANs is likely to increase over time without additional spectrum allocation. Thus, it is not clear if wireless LANs will ever compete with their wired counterparts except in applications where users are willing to sacrifice performance for mobility or when a wired infrastructure is not available.

By far the most successful application of wireless networking has been the cellular telephone system. Cellular telephones have approximately 200 million subscribers worldwide, and their growth continues at an exponential pace. The convergence of radio and telephony began in 1915, when wireless voice transmission between New York and San Francisco was first established. In 1946 public mobile telephone service was introduced in 25 cities across the United States. These initial systems used a central transmitter to cover an entire metropolitan area. This inefficient use of the radio spectrum coupled with the state of radio technology at that time severely limited the system capacity: thirty years after the introduction of mobile telephone service the New York system could only support 543 users.

A solution to this capacity problem emerged during the 50's and 60's when researchers at AT&T Bell Laboratories developed the cellular concept [34]. Cellular systems exploit the fact that the power of a transmitted signal falls off with distance. Thus, the same frequency channel can be allocated to users at spatially-separate locations with minimal interference between the users. Using this premise, a cellular system divides a geographical area into adjacent, non-overlapping, "cells". Different channel sets are assigned to each cell, and cells that are assigned the same channel set are spaced far enough apart so that interference between the mobiles in these cells is small. Each cell has a centralized transmitter and receiver (called a base station) that communicates with the mobile units in that cell, both for control purposes and as a call relay. All base stations have high-bandwidth connections to a mobile telephone switching office (MTSO), which is itself connected to the public-switched telephone network (PSTN). The handoff of mobile units crossing cell boundaries is typically handled by the MTSO, although in current systems some of this functionality is handled by the base stations and/or mobile units.

The original cellular system design was finalized in the late 60's. However, due to regulatory delays from the FCC, the system was not deployed until the early 80's, by which time much of the original technology was out-of-date. The explosive growth of the cellular industry took most everyone by surprise, especially the original inventors at AT&T, since AT&T basically abandoned the cellular business by the early 80's to focus on fiber optic networks. The first analog cellular system deployed in Chicago in 1983 was already saturated by 1984, at which point the FCC upped the cellular spectral allocation from 40MHz to 50MHz. As more and more cities became saturated with demand, the development of digital cellular technology for increased capacity and better performance became essential.

The current generation of cellular systems are all digital. In addition to voice communication these systems provide email, voice mail, and paging services. Unfortunately, the great market potential for cellular phones led to a proliferation of digital cellular standards. Today there are three different digital cellular phone standards in the U.S. alone, and other standards in Europe and Japan, none of which are compatible. The fact that different cities have different incompatible standards makes roaming throughout the U.S. using one digital cellular phone impossible. Most cellular phones today are dual-mode: they incorporate one of the digital standards along with the old analog standard, since only the analog standard provides universal coverage throughout the U.S. More details on today's digital cellular systems will be given in Section 1.6.1.

Radio paging systems are another example of an extremely successful wireless data network, with 50 million subscribers in the U.S. alone. However, their popularity is starting to wane with the widespread penetration and competitive cost of cellular telephone systems. Paging systems allow coverage over very wide areas by simultaneously broadcasting the pager message at high power from multiple base stations or satellites. These systems have been around for many years. Early radio paging systems were analog 1 bit messages signaling a user that someone was trying to reach him or her. These systems required callback over the regular telephone system to obtain the phone number of the paging party. Recent advances now allow a short digital message, including a phone number and brief text, to be sent to the pagee as well. In paging systems most of the complexity is built into the transmitters, so that pager receivers are small, lightweight, and have a long battery life. The network protocols are also very simple since broadcasting a message over all base stations requires no routing or handoff. The spectral inefficiency of these simultaneous broadcasts is compensated by limiting each message to be very short. Paging systems continue to evolve to expand their capabilities beyond very low-rate one-way communication. Current systems are attempting to implement "answer-back" capability, i.e. two-way communication. This requires a major change in the pager design, since it must now transmit signals in addition to receiving them, and the transmission distances can be quite large. Recently many of the major paging companies have teamed up with the palmtop computer makers to incorporate paging functions into these devices [48]. This development indicates that short messaging without additional functionality is no longer sufficient to meet today's communication needs.

Commercial satellite communication systems are now emerging as another major component of the wireless communications infrastructure. Satellite systems can provide broadcast services over very wide areas, and are also necessary to fill the coverage gap between high-density user locations. Satellite mobile communication systems follow the same basic principle as cellular systems, except that the cell base stations are now satellites orbiting the earth. Satellite systems are typically characterized by the height of the satellite orbit, low-earth orbit (LEOs), medium-earth orbit (MEO), or geosynchronous orbit (GEO). The geosynchronous orbits are seen as stationary from the earth, whereas the satellites with other orbits have their coverage area change over time. The disadvantage of high altitude orbits is that it takes a great deal of power to reach the satellite, and the propagation delay is typically too large for delay-constrained applications like voice. However, satellites at these orbits tend to have larger coverage areas, so fewer satellites (and dollars) are necessary to provide wide-area or global coverage. The concept of using geosynchronous satellites for communications was first suggested by the science fiction writer Arthur C. Clarke in 1945. However, the first deployed satellites, the Soviet Union's Sputnik in 1957 and the Nasa-Bell Laboratories Echo-1 in 1960, were not geosynchronous due to the difficulty of lifting a satellite into such a high orbit. The first GEO satellite was launched by Hughes and Nasa in 1963 and from then until very recently GEOs dominated both commercial and government satellite systems. The trend in current satellite systems is to use lower orbits so that lightweight handheld devices can communicate with the satellite [4]. Inmarsat is the most well-known GEO satellite system today, but most new systems use LEO orbits (e.g. Globalstar, Teledesic, and Iridium). These LEOs provide global coverage but the link rates remain low: less than 20kbps. Once deployed, these systems will allow calls any time and anywhere using a single communications device. The services provided by satellite systems include voice, paging, and messaging services, all at fairly low data rates[4, 6].

1.1.2 Wireless Data Vision

The vision of wireless communication providing high-speed high-quality information exchange between portable devices located anywhere in the world is the communications frontier of the next century. This vision will allow people to operate a virtual office anywhere in the world using a small handheld device - with seamless telephone, modem, fax, and computer communications. Wireless LANs will be used to connect together palmtop, laptop, and desktop computers anywhere within an office building or campus, as well as from the corner cafe. In the home these LANs will enable a new class of intelligent home electronics that can interact with each other and with the Internet. Video teleconferencing will take place between buildings that are blocks or continents apart, and these conferences can include travelers as well, from the salesperson who missed his plane connection to the CEO off sailing in the Carribean. Wireless video will be used to create remote classrooms, remote training facilities, and remote hospitals anywhere in the world.

The various applications described above are all components of the wireless vision. So what, exactly, is wireless communications? There are many different ways to segment this complex topic into different applications, systems, or coverage regions, as shown in Figure 1. Wireless applications include voice, Internet access, web browsing, paging and short messaging, subscriber information services, file transfer, and video teleconferencing. Systems include cellular telephone systems, wireless LANs, wide-area wireless data systems, and satellite systems. Coverage regions include in-building, campus, city, regional, and global. The question of how best to characterize wireless communications along these various segments has resulted in considerable fragmentation in the industry, as evidenced by the many different wireless products, standards, and services being offered or proposed. One reason for this fragmentation is that different wireless applications have different requirements. Voice systems have relatively low data rate requirements (around 20 Kbps) and can tolerate a fairly high probability of bit error (bit error rates, or BERs, of around 10^{-3}), but the total delay must be less than 100 msec or it becomes noticeable to the end user. On the other hand, data systems typically require much higher data rates (1-100 Mbps) and very small BERs (the target BER is 10^{-8} and all bits received in error must be retransmitted) but do not have a fixed delay requirement. Real-time video systems have high data rate requirements coupled with the same delay constraints as voice systems, while paging and short messaging have very low data rate requirements and no delay constraints. These diverse requirements for different applications make it difficult to build one wireless system that can satisfy all these requirements simultaneously. Wired networks are moving towards integrating the diverse requirements of different systems using a single protocol (e.g. ATM or SONET). This integration requires that the most stringent requirements for all applications be met simultaneously. While this is possible on wired networks, with data rates on the order of Gbps and BERs on the order of 10^{-12}, it is not possible on wireless networks, which have much lower data rates and higher BERs. Therefore, at least in the near future, wireless systems will continue to be fragmented, with different protocols tailored to support the requirements of different applications.

Will there be a large demand for all wireless applications, or will some flourish while others vanish? Companies are investing large sums of money to build multimedia wireless systems, yet the only highly profitable wireless application so far is voice. Experts have been predicting a huge market for wireless data services and products for the last 10 years, but the market for these products remains relatively small with sluggish growth. To examine the future of wireless data, it is useful to see the growth of various communication services over the last five years, as shown in Figure 2. In this figure we see that cellular and paging subscribers are growing exponentially. This growth is exceeded only by the growing demand for Internet access, driven by web browsing and email exchange. The number of laptop and palmtop computers is also growing steadily. These trends indicate that people want to communicate while on the move. They also want to take their computers wherever they go. It is therefore reasonable to assume that people want the same data communications capabilities on the move as they enjoy in their home or office. Yet the number of wireless data subscribers remains relatively flat. Why the discrepancy? We believe that the main reason for the lack of enthusiasm in wireless data is the high cost and poor performance of today's systems.

Consider the performance gap between wired and wireless networks for both local and wide-area networks, as shown in Figure 3. Wired local-area networks have data rates that are two orders of magnitude higher than their wireless counterparts. ATM is promising 100,000 Kbps for wired wide-area networks, while today's wide-area wireless data services provide only tens of Kbps. Moreover, the performance gap between wired and wireless networks appears to be growing. Thus, the most formidable obstacle to the growth of wireless data systems is their performance. Many technical challenges must be overcome to improve wireless network performance such that users will accept this performance in exchange for mobility.

1.1.3 Technical Challenges

The technical problems that must be solved to make the wireless vision a reality extend across all levels of the system design. At the hardware level the terminal must have multiple modes of operation to support the different applications and different media. Desktop computers currently have the capability to process voice, image, text, and video data, but breakthroughs in circuit design are required to implement multimode operation in a small, lightweight, handheld device. Since most people don't want to carry around a twenty pound battery, the signal processing and communications hardware of the portable terminal must consume very little power, which will impact higher levels of the system design. Many of the signal processing techniques required for efficient spectral utilization and networking demand much processing power, precluding the use of low power devices. Hardware advances for low power circuits with high processing ability will relieve some of these limitations. However, placing the processing burden on fixed location sites with large power resources has and will continue to dominate wireless system designs. The associated bottlenecks and single points-of-failure are clearly undesirable for the overall system. The finite bandwidth and random variations of the communication channel will also require robust compression schemes which degrade gracefully as the channel degrades.

The wireless communication channel is an unpredictable and difficult communications medium. First of all, the radio spectrum is a scarce resource that must be allocated to many different applications and systems. For this reason spectrum is controlled by regulatory bodies both regionally and globally. In the U.S. spectrum is allocated by the FCC, in Europe the equivalent body is the European Telecommunications Standards Institute (ETSI), and globally spectrum is controlled by the International Telecommunications Union (ITU). A regional or global system operating in a given frequency band must obey the restrictions for that band set forth by the corresponding regulatory body as well as any standards adopted for that spectrum. Spectrum can also be expensive, especially in the U.S., since the FCC began auctioning spectral allocations. In the recent spectral auctions at 2 GHz, companies spent over nine billion dollars for licenses. The spectrum obtained through these auctions must be used extremely efficiently to get a reasonable return on its investment, and it must also be reused over and over in the same geographical area, thus requiring cellular system designs with high capacity and good performance. At frequencies around several Gigahertz wireless radio components with reasonable size, power consumption, and cost are available. However, the spectrum in this frequency range is extremely crowded. Thus, technological breakthroughs to enable higher frequency systems with the same cost and performance would greatly reduce the spectrum shortage, although path loss at these higher frequencies increases, thereby limiting range.

As a signal propagates through a wireless channel, it experiences random fluctuations in time if the transmitter or receiver is moving, due to changing reflections and attenuation. Thus, the characteristics of the channel appear to change randomly with time, which makes it difficult to design reliable systems with guaranteed performance. Security is also more difficult to implement in wireless systems, since the airwaves are susceptible to snooping from anyone with an RF antenna. The analog cellular systems have no security, and you can easily listen in on conversations by scanning the analog cellular frequency band. To support applications like electronic commerce and credit card transactions, the wireless network must be secure against such listeners.

Wireless networking is also a significant challenge. The network must be able to locate a given user wherever it is amongst millions of globally-distributed mobile terminals. It must then route a call to that users as it moves at speeds of up to 100 mph. The finite resources of the network must be allocated in a fair and efficient manner relative to changing user demands and locations. There is also a tremendous infrastructure in place today of wired networks: the telephone system, the Internet, and fiber optic cable, which should be used to connect wireless systems together into a global network. However, wireless systems with mobile users will never be able to compete with wired systems in terms of data rate and reliability. The design of protocols to interface between wireless and wired networks with vastly different performance capabilities remains a challenging topic of research.

Perhaps the most significant technical challenge in wireless network design is an overhaul of the design process itself. Wired networks are mostly designed according to the layers of the OSI model: each layer is designed independent from the other layers with baseline mechanisms to interface between layers. This methodology greatly simplifies network design, although it leads to some inefficiency and performance loss due to the lack of a global design optimization. However, the large capacity and good reliability of wired network links make is easier to buffer high-level network protocols from the lower level protocols for link transmission and access, and the performance loss resulting from this isolated protocol design is fairly low. However, the situation is very different in a wireless network. Wireless links can exhibit very poor performance, and this performance along with user connectivity and network topology changes over time. In fact, the very notion of a wireless link is somewhat fuzzy due to the nature of radio propagation. The dynamic nature and poor performance of the underlying wireless communication channel indicates that high-performance wireless networks must be optimized for this channel and must adapt to its variations as well as to user mobility. Thus, these networks will require an integrated and adaptive protocol stack across all layers of the OSI model, from the link layer to the application layer. This new paradigm for wireless network protocol design will be discussed in Section 1.5.6.

In the remainder of this chapter we will provide more details about the technical challenges of wireless network design. We begin with a description of the wireless channel characteristics. We then describe the link layer design, which is responsible for sending and receiving bits over the wireless channel. Next we outline basic strategies for sharing the wireless channel among many users. We also describe the design challenges inherent to the overall wireless network. The chapter concludes with an in-depth description of current and future systems and standards, followed by the chapter summary.

1.2 The Wireless Channel

The wireless radio channel poses a severe challenge as a medium for reliable high-speed communication. Not only is it susceptible to noise, interference, blockage, and multipath, but these channel impediments change over time in unpredictable ways due to user movement. In this section we describe the underlying physics of the radio channel. These characteristics impose fundamental limits on the range, data rate, and reliability of communication over wireless links. These limits are determined by several factors, most significantly the propagation environment and the user mobility. For example, the radio channel for an indoor user at walking speeds typically supports higher data rates with better reliability than the channel of an outdoor user surrounded by tall buildings and moving at rapid speeds.

Wireless systems use the atmosphere as their transmission medium. Radio signals are sent across this media through an electromagnetic signal. Electromagnetic radiation is created by inducing a current of sufficient amplitude into an antenna whose dimensions are approximately the same as the wavelength of the generated signal, where the signal wavelength λ equals the signal's carrier frequency divided by the speed of light. We focus on radio waves in the UHF and SHF frequency bands, which occupy the .3-3GHz and 3-30GHz portions of the spectrum, respectively. Most terrestrial mobile communication systems use the UHF band, while satellite systems typically operate in the SHF band. The reason that we focus on these frequency bands is that most existing and emerging wireless communication system operate in these frequencies, and at these frequencies the earth's curvature and the ionosphere do not affect signal propagation.

A typical propagation scenario for a wireless system is shown in Figure 4. The transmitted signal has a direct-path component between the transmitter and receiver which may be attenuated or obstructed. Other components of the transmitted signal, refered to as multipath components, are reflected, scattered, or diffracted by surrounding objects, and arrive at the receiver shifted in amplitude, phase, and time relative to the direct-path signal. The received signal may also experience interference from other users in the same frequency band. Based on this model the wireless radio channel has four main characteristics: path loss, shadowing, multipath, and interference. Path loss determines how the average received signal power decreases with the distance between the transmitter and receiver. Shadowing characterizes the signal attenuation due to obstructions from buildings or other objects. Multipath fading is caused by constructive and destructive combining of the multipath signal components, which causes random fluctuations in the received signal amplitude (flat-fading) as well as self-interference (intersymbol interference or frequency-selective fading). Interference characterizes the effects of other users operating in the same frequency band either in the same system or in different systems. In the following sections we describe each of these characteristics in more detail, along with their impact on wireless system performance.

1.2.1 Path loss

Path loss is defined as the ratio of received power over transmitted power for a given propagation path, and is usually a function of propagation distance. The initial understanding of path loss associated with radio propagation goes back to the early work of Hertz in the 1880's, which showed that electromagnetic wave propagation was possible in free space. Free-space is the simplest propagation model for path loss. In this model there is a direct-path signal component between the transmitter and receiver, with no attenuating objects or multipath reflections. The received signal power in free space propagation is proportional to the transmitted power and the power gains from the transmit and receive antennas, and inversely proportional to the square of the signal carrier frequency and the square of the propagation distance [38]. Most electromagnetic waves in wireless systems propagate through environments more complex than free space, where they are reflected, scattered, and diffracted by walls, terrain, buildings, and other objects. The ultimate details of propagation in complex environments can be obtained by solving Maxwell's equations with boundary conditions that express the physical characteristics of the obstructing objects [43]. This requires calculation of the Radar Cross Section for large and complex structures. Since these calculations are difficult, and many times the necessary parameters are not available, approximations have been developed to characterize path loss in wireless systems without resorting to Maxwell's equations.

There are different, often complicated, models for path loss in different wireless environments [8, 35, 47, 11, 46, 27]. A simple exponential model for path loss that captures the key features of most of these complex models is that the received signal power is proportional to the transmitted power and inversely proportional to the square of the signal frequency and to the transmission distance raised to the power α, where α is the path loss exponent [44]. In free space propagation $\alpha = 2$, whereas in typical propagation environments α ranges between two and four.

For any path loss model the received signal-to-noise ratio (SNR=P_r/N) is the ratio of the received signal power to the noise power (noise is usually modeled as Gaussian and white, i.e. its power spectral density is constant). The BER of a wireless channel is a function of the channel's SNR. The SNR required to meet a given BER target depends on the data rate of the channel, the communication techniques used, and the channel characteristics. Since path loss impacts the received SNR it imposes limits on either the data rate or the signal range of a given communication system. Moreover, since the path loss exponent determines how quickly the signal power falls off with respect to distance, wireless channels with small path loss exponents will typically have larger coverage areas than those with large path loss exponents. Note that the path loss is inversely proportional to the square of the signal frequency so, for example, increasing the signal frequency by a factor of 10 reduces the received power and corresponding SNR by a factor of 100.

1.2.2 Shadow Fading

The transmission path between a transmitter and receiver is often blocked by hills or buildings outdoors and by furniture or walls indoors. As a result the received signal power is in fact a random variable that depends on the number and dielectric properties of the attenuating objects between the transmitter and receiver. Random signal variations due to these obstructing objects is called *shadow fading*. Measurements in many wireless environments indicate that the power, measured in decibles (dB), of a received signal subject to shadow fading follows a Gaussian (normal) distribution, with the mean determined by path loss and the standard deviation ranging from four to twelve dB, depending on the environment. When the shadow fading distribution for the average received power in dB is normal we call this *log-normal shadowing*, since the log of the received power follows a normal distribution. The random value of the shadow fading changes, or *decorrelates*, as the mobile unit moves past or around the obstructing object.

Based on path loss alone the received signal power at a fixed distance from the transmitter should be constant. However, shadow fading causes the received signal power at equal distances from the transmitter to be different, since some locations have more severe shadow fading than others. Thus, to ensure that the received SNR requirements are met at a given distance from the transmitter, the transmit power must be increased to compensate for severe shadow fading at some locations. This power increase imposes additional burdens on the transmitter battery and causes additional interference to other users in the same frequency band.

1.2.3 Multipath Flat-fading and Intersymbol Interference

Multipath gives rise to two significant channel impairments: *flat-fading* and *intersymbol interference*. Flat-fading describes the rapid fluctuations of the received signal power over short time periods or over short distances. Such fading is caused by the interference between different multipath signal components that arrive at the receiver at different times and hence are subject to constructive and destructive interference. This constructive and destructive interference generates a standing wave pattern of the received signal power relative to distance or, for a moving receiver, relative to time. Figure 5 shows a plot of the fading exhibited by the received signal power in dB as a function of time or distance. We see from this figure that the destructive interference causes the received signal power to fall more than 30 dB (three orders of magnitude) below its average value. We say that a channel is in a *deep fade* whenever its received signal power falls below that required to meet the link performance specifications. Thus, since communication links are designed with an extra power margin (link margin) of 10-20 dB to compensate for fading and other channel impairments, a channel is in a deep fade if its received power falls 10-20 dB below its average received power. We see from Figure 5 that flat-fading channels often experience deep fades. In addition, the changes in signal power are extremely rapid: the signal power changes drastically for distances of approximately half a signal wavelength. At a signal frequency of 900 MHz, this corresponds to ever .3 m or every microsecond for terminals moving at 30 mph. The variation in the received signal envelope of a flat-fading signal typically follows a Rayleigh distribution if the signal path between the transmitter and receiver is obstructed and a Ricean distribution if this signal path is not obstructed.

The combination of path loss, shadowing, and flat-fading is shown in Figure 6. We see that the power falloff with distance due to path loss is fairly slow, while the signal variation due to shadowing changes more quickly, and the variation due to flat-fading is very fast. Flat-fading has two main implications for wireless link design. First, in flat-fading the received signal power falls well below its average value. This causes a large increase in BER, which can be reduced somewhat by increasing the transmitted power of the signal. However, it is very wasteful of power to compensate for flat-fading in this manner, since deep fades occur rarely and over very short time periods. Thus, most systems do not increase their transmit power sufficiently to remove deep signal fades. For typical user speeds and data rates these fades affect many bits, thereby cause long strings of bit errors called *error bursts*. Error bursts are difficult to correct for using error-correction codes, since these codes can typically only correct for a few simultaneous bit errors. Other methods to compensate for error bursts due to flat-fading will be discussed in Section 1.3.3.

The other main impairment introduced by multipath is intersymbol interference (ISI). ISI becomes a significant problem when the maximum difference in the path delays of the different multipath components, called the *multipath delay spread*, exceeds a significant fraction of a bit time. This results in self-interference, since a multipath reflection carrying a given bit transmission will arrive at the receiver simultaneously with a different (delayed) multipath reflection carrying a previous bit transmission. In the frequency domain this self-interference corresponds to a non-flat frequency spectrum, so signal components at different frequencies are multiplied by different complex scale factors, thereby distorting the transmitted signal. For this reason ISI is also referred to as *frequency-selective fading*. A channel exhibits frequency-selective fading if the channel's *coherence bandwidth*, defined as the inverse of the channel's multipath delay spread, is less than the bandwidth of the transmitted signal. ISI causes a high BER that cannot be reduced by increasing the signal power, since that also increases the power of the self-interference. Thus, without compensation ISI forces a reduction in data rate such that the delay spread associated with the multipath components is less than a bit time. This imposes a stringent constraint on data rates, on the order of 100 Kbps for outdoor environments and 1 Mbps for indoor environments. Thus, some form of ISI compensation is needed to achieve high data rates. These compensation techniques will be discussed in Section 1.3.4.

1.2.4 Doppler Frequency Shift

Relative motion between the transmitter and receiver causes a shift in the frequency of the transmitted signal called the Doppler shift. The Doppler shift, f_D, is given by $f_D = v/\lambda$, where v is the relative velocity between the transmitter and receiver and λ is the wavelength of the transmitted signal. Since the mobile velocity varies with time, so does the Doppler shift. This variable Doppler shift introduces an FM modulation into the signal, causing the signal bandwidth to increase by roughly f_D. Assuming a transmit frequency of 900 MHz and a user speed of 60 mph, the Doppler shift is roughly 80 Hz. Since typical signal bandwidths are on the order of tens of kilohertz or more, bandwidth spreading due to Doppler is not a significant problem in most applications. However, Doppler does cause the signal to decorrelate over a time period roughly equal to $1/f_D$. For differential signal detection the Doppler imposes a lower bound on the channel BER that cannot be reduced by increasing the signal power. More details on the impact of channel Doppler for differential detection will be discussed in Section 1.3.1.

1.2.5 Interference

Wireless communication channels experience interference from various sources. The main source of interference in cellular systems is frequency reuse, where frequencies are reused at spatially-separated locations to increase spectral efficiency. Interference from frequency reuse can be reduced by multiuser detection [55], directional antennas [58], and dynamic channel allocation [31], all of which increase system complexity.

Other sources of interference in wireless systems include adjacent channel interference, caused by signals in adjacent channels with signal components outside their allocated frequency range, and narrowband interference, caused by users in other systems operating in the same frequency band. Adjacent channel interference can be mostly removed by introducing guard bands between channels, however this is wasteful of bandwidth. Narrowband interference can be removed through notch filters or spread spectrum techniques. Notch filters are simple devices however they require knowledge of the exact location of the narrowband interference. Spread spectrum is very effective at removing narrowband interference, as will be discussed in Section 1.3.4, however it requires significant spreading of the signal bandwidth as well as an increase in system complexity. For these reasons spread spectrum in not typically used just to remove narrowband interference. However, spread spectrum allows multiple users to share the same bandwidth, and is therefore also used for multiple access.

1.2.6 Infrared versus Radio

Infrared is a form of wireless communication where the frequency of the transmitted signal is much higher than typical radio frequencies, around 100 GHz [30]. Because the received signal power is inversely proportional to the square of the signal frequency, infrared transmission over large distances requires a very high transmit power or highly directional antennas. There are two main forms of infrared transmission: directive and nondirective. In directive transmission the transmit antenna is pointed directly at the receiver, whereas in nondirective transmission the signal is transmitted uniformly in all directions. Since directive transmission concentrates all its power in one direction, it typically achieves much higher data rates than nondirective transmission. However, these systems are severely degraded by obstructing objects and the corresponding shadow fading, which is difficult to avoid in most indoor environments with mobile users. Non-directed links have limited range due to path loss, typically in the tens of meters.

Infrared transmission enjoys a number of advantages over radio, most significantly the fact that spectrum in this frequency range is unregulated. Thus, infrared systems need not obtain an FCC license for operation. In addition, infrared systems are immune to radio interference. Infrared radiation will not penetrate walls and other opaque materials, so an infrared signal is confined to the room in which it originates. This makes infrared more secure against eavesdropping and it allows neighboring rooms to use the same infrared links without interference. These signals are not subject to flat-fading since variations in the received signal power are integrated out by the detector. However, ISI is a major problem for high-speed infrared systems, as it is for radio systems. Thus, high-speed infrared systems must use some form of ISI mitigation, typically equalization. Infrared systems are also significantly degraded by ambient light, since it radiates at roughly the same frequency, causing substantial noise. In summary, infrared systems are unlikely to be used for low-cost outdoor systems due to limited range. They have some advantages over radio systems in indoor environments, but must overcome the ambient light and range limitations to be successful in these applications. Due to these problems radio is still the dominant technology for both indoor and outdoor systems.

1.2.7 Capacity Limits of Wireless Channels

The pioneering work of Claude Shannon in 1949 determined the ultimate capacity limits of communication channels with white Gaussian noise [49]. For a channel without shadowing, fading, or ISI, Shannon determined that the maximum possible data rate on a given channel of bandwidth B is $R = B\log_2(1 + \mathrm{SNR})$ bps, where SNR is the received signal-to-noise power ratio. Shannon capacity is a theoretical limit that cannot be achieved in practice, however as link level design techniques improve data rates for these systems approach this theoretical bound. Since wireless communications is a relatively new field, the Shannon capacity is unknown for many wireless channels of interest, and depends not only on the channel but also on whether or not the transmitter and/or the receiver can track the channel variations [25, 14]. In addition, link level design for wireless channels is still relatively immature, hence there is typically a large gap between actual performance and Shannon capacity when this capacity is known. For example, one of the digital cellular standards has a 30 KHz bandwidth and a received SNR of approximately 20 dB (or more) after attenuation from shadow fading. The Shannon capacity of this channel with Rayleigh fading, assuming that the channel variation can be tracked, is on the order of 200 Kbps [25]. However, cellular systems only achieve data rates on the order of 20 Kbps per channel, an order of magnitude less than the Shannon limit. While some of this performance gap is due to channel impairments that are not incorporated into this simple model, most of the gap is due to the relatively inefficient signaling methods used on today's wireless channels. In wired channels, where technology is quite mature and the channel characteristics fairly benign, the data rates achieved in practice are quite close to the Shannon bound. For example, telephone lines have a capacity limit of between 30 and 60 Kbps, depending on the line quality, and modems sold today are achieving close to these data rates.

The simple formula given above for Shannon capacity is applicable to static channels with white Gaussian noise. Shannon capacity is also known for static channels with non-white noise and intersymbol interference [22]. However, determining the capacity limits of time-varying wireless channels with shadowing, multipath fading, and intersymbol interference is quite challenging, and depends on the channel characteristics, the channel rate of change, and the ability to track the channel variations. A relatively simple lower bound for the capacity of any channel is the Shannon capacity under the worst-case propagation conditions. This can be a good bound to apply in practice, since many communication links are designed to have acceptable performance even under the worst-case conditions. However, worst-case system design and capacity evaluation can be overly pessimistic, since typical operating conditions are generally much better than worst-case. For channels with severe multipath the channel capacity under worst-case fading conditions can be equal or close to zero. Fading compensation techniques can be used on these channels to increase both their Shannon capacity and their achievable data rates in practice. Some of these compensation techniques will be discussed in the next section.

1.3 Link Level Design

In this section we describe the link layer design for wireless channels. The goal of this design is to provide high data rates with low delays and BERs while using minimum bandwidth and transmit power. The link layer design must perform well in radio environments with fading, shadowing, multipath, and interference. Hardware constraints, such as imperfect timing and nonlinear amplifiers, must also be taken into consideration. Low-power implementations are needed, particularly for the mobile units, which have limited battery power. In addition, low-cost implementations for both transmitter and receiver are clearly desirable. Many of these properties are mutually exclusive, and induce tradeoffs in the choice of link level design techniques.

1.3.1 Modulation Techniques

Digital modulation is the process of encoding a digital information signal into the amplitude, phase, or frequency of the transmitted signal [59, 40, 57]. This encoding process impacts the bandwidth of the transmitted signal and its robustness to channel impairments. In general, a modulation technique encodes several bits into one symbol, and the rate of symbol transmission determines the bandwidth of the transmitted signal. Since the signal bandwidth is determined by the symbol rate, having a large number of bits per symbol generally yields a higher data rate for a given signal bandwidth. However, the larger the number of bits per symbol, the greater the required received SNR for a given target BER.

Digital modulation techniques typically fall in the category of linear or nonlinear. In linear modulation the amplitude and/or phase of the transmitted signal varies linearly with the digital modulating signal, whereas the transmitted signal amplitude is constant for nonlinear techniques. In general linear modulation techniques, which include all forms of quadrature-amplitude modulation (QAM) and phase-shift-keying (PSK), use less bandwidth than nonlinear techniques, which include various forms of frequency-shift-keying (FSK and MSK). However, since information is encoded into the amplitude and phase of linear modulation, this type of modulation is more susceptible to amplitude and phase fluctuations caused by multipath flat-fading. In addition, the amplifiers used for linear modulation must be linear, and these amplifiers are more expensive and less efficient than nonlinear amplifiers. Thus, the bandwidth efficiency of linear modulation is generally obtained at the expense of hardware cost, power, and higher BERs in fading. Linear modulation techniques are used in most wireless LAN products, whereas nonlinear techniques are used in most cellular and wide-area wireless data systems.

Linear modulation techniques can be detected coherently or differentially. Coherent detection requires the receiver to obtain a coherent phase reference for the transmitted signal. This is difficult to do in a rapidly fading environment, and also increases the complexity of the receiver. Differential detection uses the previously detected symbol as a phase reference for the current symbol. Because this detected symbol is a noisy reference, differential detection requires roughly double the power of coherent detection for the same BER. In addition, if the channel is changing rapidly then differential detection is not very accurate, since the channel phase may change considerably over one symbol time. As a result rapidly changing channels with differential detection have an irreducible error floor, i.e. the BER of the channel has a lower bound (error floor) that cannot be reduced by increasing the received SNR. This error floor increases as the rate of channel variation (the channel Doppler) increases and decreases as the data rate increases (since a higher data rate corresponds to a shorter bit time, so the channel phase has less time to decorrelate between bits). Thus, for high-speed wireless data (above 1 Mbps), the error floor is quite low at user speeds below 60 mph, but at lower data rates the error floor becomes significant, thereby preventing the use of differential detection.

1.3.2 Channel Coding and Link Layer Retransmission

Channel coding adds redundant bits to the transmitted bit stream, which are used by the receiver to correct errors introduced by the channel [33, 40, 57]. This allows for a reduction in transmit power to achieve a given target BER and also prevents packet retransmissions if all the bit errors in a packet can be corrected. Conventional forward error correction (FEC) codes use block or convolutional code designs to produce the redundant bits for FEC: the error correction capabilities of these code designs are obtained at the expense of increased signal bandwidth or a lower data rate. Trellis codes, invented in the early 1980's, use a joint design of the channel code and the modulation to provide FEC without bandwidth expansion or rate reduction [54]. The latest advance in coding technology is the family of Turbo codes, invented in 1993 [12]. Turbo codes, which achieve within a fraction of a dB of the Shannon capacity on certain channels, are complex code designs that combine an encoded version of the original bit stream with an encoded version of one or more permutations of this original bit stream [28]. The optimal decoder for the resulting code is very complex, but Turbo codes use an iterative technique to approximate the optimal decoder with reasonable complexity. While Turbo codes exhibit dramatically improved performance over previously-known coding techniques, and can be used with either binary or multilevel modulation [45], they generally have high complexity and large delays, which makes them unsuitable for many wireless applications.

Another way to compensate for the channel errors prevalent in wireless systems is to implement link layer retransmission (part of the protocol known as ARQ, for Automatic Repeat reQuest [13]). In ARQ data is collected into packets and each packet is encoded with a checksum. The receiver uses the checksum to determine if one or more bits in the packet were received in error. If so then the receiver requests a retransmission of the entire packet. Link layer retransmission is wasteful of system resources, since each retransmission requires additional power and bandwidth and also interferes with other users. In addition, packet retransmissions can cause data to be delivered to the receiver out of order as well as triggering duplicate acknowledgments or end-to-end retransmissions at the transport layer, further burdening the network. While ARQ has disadvantages, in applications that require error-free packet delivery at the link layer FEC is not sufficient, so ARQ is the only alternative.

1.3.3 Flat-Fading Countermeasures

The random variation in received signal power resulting from multipath flat-fading causes a very large increase in the average BER on the link. For example, in order to maintain an average BER of 10^{-3} (a typical requirement for the link design of voice systems), 60 times more power is required if flat-fading is present[1]. This required increase in power is even larger at the much lower BERs required for data transmission. Thus, countermeasure to combat the effects of flat-fading can significantly reduce the transmit power required on the link to achieve a target BER. The most common flat-fading countermeasures are diversity, coding and interleaving, and adaptive modulation.

In diversity, several separate, independently fading signal paths are established between the transmitter and receiver, and the received signals obtained from each of these paths are combined. Because there is a low probability of independent fading paths experiencing deep fades simultaneously, the signal obtained by combining several such fading paths is unlikely to experience large power variations, especially with four or more diversity paths. Independent fading paths can be achieved by separating the signal in time, frequency, space, or polarization. Time and frequency diversity are inefficient, since information is duplicated. Polarization diversity is of limited effectiveness because only two independent fading paths (corresponding to horizontal and vertical polarization) can be created, and the transmission power is usually divided between these two paths. That leaves space diversity as the most efficient diversity technique. In space diversity independent fading paths are obtained using an antenna array, where each antenna element receives an independent fading path. In order to obtain independent fading paths the antenna elements must be spaced at least one half signal wavelength apart, which may be difficult on small handheld devices, especially in the lower frequency bands ($f < 1$ GHz or, equivalently, $\lambda > .3$ m.).

Another technique to combat flat-fading effects is coding and interleaving. In general, flat-fading causes bit errors to occur in bursts corresponding to the times when the channel is in a deep fade. Low-complexity channel codes can correct at most a few simultaneous bit errors, and the code performance deteriorates rapidly when errors occur in large bursts. The basic idea of coding and interleaving is to spread burst errors over many codewords. Specifically, in the interleaving process adjacent bits in a single

[1]This calculation assumes BPSK modulation. Higher level modulations requires an even larger transmit power increase.

codeword are separated by bits from other codewords and then these scrambled bits are transmitted over the channel. Since channel errors occur in bursts, the scrambling prevents adjacent bits in the same codeword from being affected by the same error burst. At the receiver the bits are deinterleaved (descrambled) back to their original order and then passed to the decoder. If the interleaving process spreads out the burst errors, and burst errors do not occur too frequently, then the codewords passed to the decoder will have at most one bit error, and these errors are easily corrected with most FEC channel codes. The cost of coding and interleaving (increased delay and complexity) may be large if the fading rate relative to the data rate is slow, which is typically the case for high-speed data. For example, at a channel Doppler of 10 Hz and a data rate of 10 Mbps, an error burst will last around 300,000 bits. In this case the interleaver must be large enough to handle at least that much data and the application must be able to tolerate an interleaver delay of at least 30 msec.

When the channel can be estimated and this estimate sent back to the transmitter, the transmission scheme can be adapted relative to the channel conditions. In particular, the data rate, power, and coding scheme can be adapted relative to the channel fading to maximize the average data rate or to minimize the average transmit power or BER [23]. This adaptation allows the channel to be used more efficiently, since the transmission parameters are optimized to take advantage of favorable channel conditions. There are several practical constraints which determine when adaptive techniques should be used. If the channel is changing so fast that it cannot be accurately estimated or the estimate cannot be fed back to the transmitter before the channel changes significantly then adaptive techniques will perform poorly. Adaptive techniques also increase the complexity of both the transmitter and the receiver to account for the channel estimation and the adaptive transmission. Finally, a feedback path is required to relay the channel estimate back to the transmitter, which occupies a small amount of additional bandwidth on the return channel.

1.3.4 Intersymbol Interference Countermeasures

Techniques to combat ISI fall into two categories: signal processing and antenna solutions. Signal processing techniques, including equalization, multicarrier modulation, and spread spectrum, can either compensate for ISI at the receiver or make the transmitted signal less sensitive to ISI. Antenna solutions, including directive beams and smart antennas, change the propagation environment so that the delay between multipath components, and the corresponding ISI resulting from these delays, is reduced.

The goal of equalization is to cancel the ISI or, equivalently, to invert the effects of the channel [40]. Channel inversion can be achieved by passing the received signal through a linear equalizing filter with a frequency response that is the inverse of the channel frequency response. This method of channel inversion is called *zero forcing*, since the ISI is forced to zero. In linear zero-forcing equalization the noise is also passed through the inverse channel filter, and the noise is amplified over frequencies where the channel has low gain. This noise enhancement can significantly degrade the received SNR of systems with zero-forcing equalization. A better linear equalization technique uses an equalizing filter that minimize the average mean square error between the equalizer output and the transmitted bit stream. This type of linear equalizer is called a *minimum mean square error* equalizer. Both types of linear equalizers can be implemented in relatively simple hardware. Although linear equalizers work well on some channels, their performance can be quite poor on channels with a long delay spread or with large variations in the channel frequency response. For these channels the non-linear decision-feedback equalizer (DFE) tends to work much better than the linear techniques. A DFE determines the ISI from previously-detected symbols and subtracts it from the incoming symbols [10]. The DFE does not suffer from noise enhancement because it estimates the channel rather than inverting it. On most ISI channels the DFE has a much lower BER than a linear equalizer with a slightly higher complexity. Other forms of equalization include maximum-likelihood sequence estimation [21] and Turbo equalization [18], both of which usually outperform the DFE, but also have significantly higher complexity. All equalizer techniques require an accurate channel estimate, which is usually obtained by sending a training sequence over the channel. The equalizer must also track variations in the channel by periodic retraining and by adjusting the filter coefficients during data transmission based on the equalizer outputs [41, 42]. For this reason equalizers do not work well on channels changing so rapidly that an accurate channel estimate cannot be maintained without significant training overhead.

Multicarrier modulation is another technique to mitigate ISI. In multicarrier modulation the transmission bandwidth is divided into narrow subchannels, and the information bits are divided into an equal number of parallel streams. Each stream is used to modulate one of the subchannels, which are transmitted in parallel. Ideally the subchannel bandwidths are less than the coherence bandwidth of the channel, so that the fading on each subchannel is flat, not frequency-selective, thereby eliminating ISI. The simplest method for implementing multicarrier modulation is orthogonal (nonoverlapping) subchannels, but the spectral efficiency of multicarrier modulation can be increased by overlapping the subchannels. This is called orthogonal frequency division multiplexing (OFDM) [16, 15]. A big advantage of OFDM is that it can be efficiently implemented using the fast Fourier transform at both the transmitter and the receiver. Since the entire bandwidth of an OFDM signal experiences frequency-selective fading, some of the OFDM subchannels will have weak SNRs. The performance over the weak subchannels can be improved by coding across subchannels, frequency equalization, or adaptive loading (transmitting at a higher data rate on the subchannels with high SNRs). The advantage of multicarrier modulation over equalization is that frequency equalization requires less training than time equalization. However, flat-fading, frequency offset, and timing mismatch impair the orthogonality of the multicarrier subchannels, resulting in self-interference that can significantly degrade performance.

Spread spectrum is a technique that spreads the transmit signal over a wide signal bandwidth in order to reduce the effects of flat-fading, ISI, and narrowband interference [17, 56]. In spread spectrum the information signal is modulated by a wideband pseudo-noise (PN) signal, resulting in a much larger transmit signal bandwidth than in the original signal. Spread spectrum first achieved widespread use in military applications due to its ability to a hide a signal below the noise by spreading out its power over a wide bandwidth, its resistance to narrowband jamming, and its inherent ability to reduce multipath flat-fading and ISI. However, these advantages come with a significant complexity increase, especially in the receiver. There are two common forms of spread spectrum: direct sequence, in which the data sequence is multiplied by the PN sequence, and frequency hopping, in which the narrowband signal is "hopped" over different carrier frequencies based on the PN sequence. Both techniques result in a transmit signal bandwidth that is much larger than the original signal bandwidth, hence the name spread spectrum.

Spread spectrum demodulation occurs in two stages: first the received signal is despread by removing the PN sequence modulation, then the original information signal is demodulated to get the information bits. In direct sequence despreading is accomplished by multiplying the received signal with an exact copy of the PN sequence, perfectly synchronized in time. The synchronization process entails a great deal of receiver complexity, and can be degraded by interference, fading, and noise. Frequency-hopped signals are despread by synchronizing the carrier frequency at the receiver to the hopping pattern of the PN sequence. After despreading the data signal is demodulated at baseband in the conventional way. In the despreading process narrowband interference and delayed multipath signal components are modulated by the PN sequence, thereby spreading their signal power over the wide bandwidth of the PN sequence. The narrowband filter in the baseband demodulator then removes most of this power, which yields the narrowband interference and multipath rejection properties of spread spectrum. A RAKE receiver can also be used to coherently combine all multipath components, thereby providing improved performance through receiver diversity [53].

Antenna design can also reduce the effects of flat-fading and ISI. The most common wireless antenna is an omnidirectional antenna, where the power gain in all angular directions is the same. Directional antennas attenuate signals in all but a narrow angular range, and in this range signals are greatly magnified. Multipath signal components typically come from large range of angular directions. Thus, by using directional antennas at the transmitter and/or the receiver, the power in most of the multipath components can be greatly reduced, thereby eliminating most flat-fading and ISI. The ISI reduction directly translates into increased data rates. For example, data rates exceeding 600 Mbps have been obtained experimentally in an indoor environment with directional antennas at both the transmitter and the receiver [19]. Directional antennas can also greatly reduce interference from other users in a cellular system if these users are located outside the antenna's angular range of high power gain. However, directional antennas must be carefully positioned to point towards the user of interest, and this direction changes as users move around. This is a key motivation behind the development of steerable antennas, also called smart or adaptive antennas [2]. These antennas change the shape and direction of their transmission beams to accommodate changes in mobile position. A steerable transmitting antenna works by controlling the phases of the signals at each of its elements, which changes the angular locations of the antenna beams (angles with large gain) and nulls (angles with small gain). Using feedback control, an antenna beam can be steered to follow the movement of a mobile, greatly reducing flat-fading, and interference from other users. Smart antennas can also provide diversity gain.

1.4 Channel Access

Due to the scarcity of wireless spectrum, efficient techniques to share bandwidth among many heterogeneous users are needed. Applications requiring continuous transmission (e.g. voice and video) generally allocate dedicated channels for the duration of the call. Sharing bandwidth through dedicated channel allocation is called multiple access. In contrast to voice or video, transmission of data tends to occur in bursts. For example, a remote host will be idle when the user is not typing, so dedicated allocation of bandwidth for that user is inefficient. Bandwidth sharing for users with bursty transmissions generally use some form of random channel allocation which does not guarantee channel access. Bandwidth sharing using random channel allocation is called random access. In general, the choice of whether to use multiple access or random access, and which technique to use within each access type, will depend on the traffic characteristics of the system, the state of current access technology, and compatibility with other systems. This choice is further complicated when frequencies are reused to increase spectral efficiency, as in cellular system designs. In this section we describe the different multiple access and random access techniques along with their corresponding tradeoffs. We also describe spectral etiquette protocols for sharing bandwidth among different systems that do not coordinate with each other.

1.4.1 Multiple Access

Multiple access techniques assign dedicated channels to multiple users through bandwidth division. Methods to divide the spectrum include frequency-division (FDMA), time-division (TDMA), code-division (CDMA), and hybrid combinations of these methods [44]. In FDMA the total system bandwidth is divided into orthogonal channels nonoverlapping in frequency that are allocated to the different users. In TDMA time is divided into nonoverlapping time slots that are allocated to different users. In CDMA time and bandwidth are used simultaneously by different users, modulated by orthogonal or semi-orthogonal spreading codes. With orthogonal spreading codes the receiver can separate out the signal of interest from the other CDMA users with no residual interference between users. However, only a finite number of orthogonal spreading codes exist for any given signal bandwidth. With semi-orthogonal spreading codes the receiver cannot completely separate out signals from different users, so that after receiver processing there is some residual interference between users. However, there is no hard limit on how many semi-orthogonal codes exist within a given signal bandwidth. This property, known as *soft capacity*, has advantages and disadvantages in the overall system design which will be outlined in more detail below. Direct-sequence spread spectrum is often used to generate the semi-orthogonal CDMA signals. As discussed in Section 1.3.4, direct sequence spread spectrum has inherent benefits of multipath mitigation and narrowband interference rejection that are not inherent to either FDMA or TDMA, at a cost of somewhat increased complexity in the transmitter and receiver.

FDMA is the least complex of these multiple access techniques. TDMA is somewhat more complex, since it requires timing synchronization among all users. In addition, the orthogonality of the users in TDMA is significantly degraded by ISI, since a signal transmitted in one timeslot will interfere with subsequence timeslots due to the multipath delay spread. Semi-orthogonal CDMA is the most complex of the multiple access schemes due to the inherent complexity of spread spectrum in general and the additional complexity of separating out different semi-orthogonal CDMA users. Semi-orthogonal CDMA also requires stringent power control to prevent the near-far problem. The near-far problem arises from the non-orthogonality of the spreading codes, so that every user causes interference to all other users. Due to the power falloff with distance described in Section 1.2.1, the received signal power of a mobile unit located close to a receiver (or base station) is typically much larger than the received signal power of a mobile unit farther away. Thus, the interference power of this "near" mobile unit can be large in comparison with the received signal power of the "far" mobile unit. As a result the mobile units located far from the receiver typically experience high interference from other users and correspondingly poor performance. Power control mitigates the near-far problem by equalizing the received power (and the corresponding interference power) of all mobile units regardless of their distance from the receiver. However, this equalized power is difficult to maintain in a flat-fading environment, and is one of the major design challenges of semi-orthogonal CDMA.

Another interesting tradeoff in these multiple access methods is hard versus soft system capacity. TDMA and FDMA place a hard limit on the number of users sharing a given bandwidth, since the channel is divided into orthogonal time or frequency slots, each of which can only support one user. Orthogonal CDMA also has this hard limit. Conversely, semi-orthogonal CDMA has the advantage of soft capacity: there is no absolute limit on the number of users. However, since the semi-orthogonal codes interfere with each other, the performance of all users degrades as the number of users in the system increases: if the number of users is too large then no user will have acceptable performance. Interference from other CDMA users can be reduced using a range of sophisticated techniques including smart antennas [2], interference cancellation [20], and multiuser detection [55]. These techniques significantly increase the complexity of the system and can sometimes exhibit poor performance in practice.

The competing multiple access methods in the U.S. for cellular and PCS services are mixed FDMA/TDMA with three time slots per frequency channel (IS-54), semi-orthogonal CDMA (IS-95), and a combination of TDMA and slow frequency hopping (GSM). The debate among cellular and personal communication standards committees and equipment providers over which approach to use has led to numerous analytical studies claiming superiority of one technique over the other under different channel assumptions and operating scenarios. It is intractable to definitively analyze the performance of these different multiple access techniques in all real operating environments. Thus there is no common agreement as to which access technique is superior for any given system, especially for systems with frequency reuse or significant channel impairments.

1.4.2 Random Access

In most wireless data networks only a small, unpredictable, and dynamic subset of all the users in the network has data to send at any given time. For these systems it is inefficient to assign each user a dedicated channel. When dedicated channel access is not provided and access to the channel is not guaranteed, a random access protocol is required. Random access protocols are based on packetized data transmissions and typically fall in two categories: ALOHA techniques and reservation or demand-assignment protocols [3].

In pure ALOHA a transmitter will send data packets over the channel whenever data is available. This leads to a large number of data collisions at the receiver. A collision occurs when two or more packets are received simultaneously at the receiver and therefore none of the packets can be decoded correctly. Packets that collide must be resent at a later time. The throughput of an ALOHA channel, defined as the rate at which packets are correctly received, is low due to the high probability of collisions. In fact, under standard modeling assumptions the maximum throughput in an ALOHA channel is 18 percent of the channel data rate (the rate that a single user could achieve if it was not sharing the channel) [13]. Fewer collisions occur if time is divided into separate slots and packet transmissions are confined to these predetermined slots, since there is no partial overlap of packets. This modification of pure ALOHA is called slotted ALOHA. Although slotted ALOHA roughly doubles the maximum throughput relative to pure ALOHA, this throughput is still insufficient to support bandwidth-sharing among more than a handful of high-speed users. The number of collisions in ALOHA can be reduced by the capture effect, which is similar to the near-far effect described above for spread spectrum systems. Specifically, due to the power falloff with distance of the transmitted signal, a packet transmitted from a mobile that is far from the receiver typically causes just a small amount of interference to a packet transmitted from a closer location, so that despite a collision between these packets the latter packet can be "captured", i.e. received without error. Spread spectrum can also be combined with ALOHA to reduce collisions, since when packets modulated with PN spreading sequences collide, they can be separated out by a spread spectrum receiver. Since spread spectrum random access receivers must be able to demodulate the PN spreading sequences for a large number of users, these receivers typically have very high complexity.

ALOHA with carrier sensing is often used in wired networks (e.g. the Ethernet) to avoid packet collisions. In carrier sensing a transmitter senses the channel before transmission to determine if the channel is busy. If so then the transmission is delayed until the channel is free. Carrier sensing is often combined with collision detection, where the channel is monitored during packet transmission. If another user accesses the channel during this transmission, thereby causing a collision, then by detecting this collision the transmitter can resend the packet without waiting for a negative acknowledgement or timeout. Carrier sensing and collision detection require that a transmitting user can detect packet transmissions from other users to its intended receiver. This detection is often impossible in a wireless environment due to path loss. Specifically, suppose two transmitting users are on opposite sides of a receiver, so that the distance between each user and the receiver is half the distance between the two users. In this case, although the receiver may correctly decode a packet transmitted from either user in the absence of a collision, each user cannot detect a transmitted packet from the other user since the users are so far apart. Collision detection is also impaired by shadow fading, since the users may have an object obstructing the signal path between them. The difficulty of detecting collisions in a wireless environment is called the "hidden terminal problem" since, due to path loss and shadow fading, the signals from other users in the system may be hidden.

Since carrier sensing and collision detection are not very effective in wireless channels, the current generation of wireless LANs use collision avoidance. In collision avoidance the receiver notifies all nearby transmitters when it is receiving a packet by broadcasting a "busy tone". Transmitters with packets to send wait until some random time after the busy tone terminates to begin sending their packets. The random backoff prevents all users with packets to send from simultaneously transmitting as soon as the busy tone terminates. Collision avoidance significantly increases the throughput of ALOHA, and is currently part of several wireless LAN standards. However, the efficacy of collision avoidance can be degraded by the effects of path loss, shadowing, and multipath fading on the busy tone.

Reservation protocols assign channels to users on demand through a dedicated reservation channel [13]. The channel assignment is done by a central base station or by a common algorithm running in each terminal. In such a system the total channel bandwidth is divided into data channels and reservation channels, where typically the reservation channels only occupy a small fraction of the total bandwidth. When a user has data to transmit, he sends a short packet containing a channel request over the reservation channel. Assuming that this packet is correctly received and a data channel is available, a data channel is reserved for the user's data transmission and this channel assignment is conveyed back to the user. Demand-based assignment can be a very efficient means of random access since channels are only assigned when they are needed, as long as the required overhead traffic to assign channels is a small percentage of the message to be transmitted. If not then several problems arise. First there is the setup delay and overhead associated with the channel reservation request and assignment procedure. For long spurts of traffic this is not a serious limitation, but for networks with a considerable amount of short messaging this delay and overhead can significantly degrade network performance. Second, for heavily-loaded systems the reservation channel may become congested with reservation requests, basically shifting the multiple access problem from the data channel to the reservation channel. For networks serving a wide variety of data users where a considerable amount of the network traffic consists of small messages, reservation-based random access may not be the best choice of random access protocol.

Packet-Reservation Multiple Access (PRMA) is a relatively new random access technique that combines the advantages of reservation protocols and ALOHA [26]. In PRMA time is slotted and the time slots are organized into frames with N timeslots per frame. Active terminals with packets to transmit contend for free time slots in each frame. Once a packet is successfully transmitted in a time slot, the time slot is reserved for that user in each subsequent frame, as long as the user has packets to transmit. When the user stops transmitting packets in his reserved slot the reservation is forfeited, and the user must again contend for free time slots in subsequent packet transmissions. PRMA is well-suited to multimedia traffic with a mix of voice (or continuous stream) traffic and data. Once the continuous stream traffic has been successfully transmitted it maintains a dedicated channel for the duration of its transmission, while data traffic only uses the channel as long as it is needed. PRMA requires little central control and no reservation overhead, so it is superior to reservation-based protocols when there is a mix of voice and data traffic.

All random access protocols require packet acknowledgements to guarantee successful reception of packets. If a packet is not acknowledged within some predetermined time window then the packet is retransmitted. Since packet acknowledgements are also sent over a wireless channel, they are frequently delayed, lost, or corrupted due to channel impairments. This can result in unnecessary packet retransmission, which is inefficient, and packet duplication, which must be handled by the network protocol. Wireless networks can improve the likelihood of timely and uncorrupted packet acknowledgments by using more powerful link layer techniques to send these acknowledgements. However, this requires that the link layer differentiate between different types of data transmissions, which adds to the link layer complexity.

1.4.3 Spectral Etiquette

Channel access allows multiple users in the same system to share a given bandwidth allocation. However, in some cases multiple systems will share the same bandwidth without any coordination, and this requires interoperability of their different access techniques and communication designs. This co-existence can be accomplished through etiquette rules, which are a minimum set of rules that allow multiple systems to share the available bandwidth fairly. These techniques offer an alternative to standardization methods that require agreement on channel access and system design before systems can be built and deployed. WINForum, an association of companies developing wireless products, has defined a set of etiquette rules for the unlicensed 2 GHz PCS bands that has been adopted by the FCC [50]. The same set of rules are being considered for the 60 GHz spectrum allocation. The key elements of these etiquette rules are: (1) listen before transmitting, to insure that the transmitter is the only user of the spectrum while minimizing the possibility of interfering with other spectrum users; (2) limit transmission time in order to make it possible for other users to make use of the spectrum in a fair manner; and (3) limit transmitter power, so as not to interfere with users in nearby spectrum or reusing the same frequency spectrum some distance away.

1.5 Network Design

In this section we address some network design issues for wireless networks. We will first describe different network architectures and their relative tradeoffs. Distributed and centralized network control strategies with respect to these architectures will also be discussed. Next we consider protocols for mobility management, including the location of mobile users, user authentication, and call routing. Network reliability and quality-of-service guarantees for wireless networks will also be discussed. Some pitfalls of wired and wireless network interoperability using universal networking protocols like TCP/IP and ATM networks will be outlined, followed by a brief discussion of security issues in wireless networks.

1.5.1 Architecture

The choice of architecture for a two-way wireless network involves a host of issues dealing with the most fundamental aspects of network design. The three main types of network architectures are a star (central hub) topology, an ad-hoc or peer-to-peer structure, and a hierarchical or tree structure. These three architecture types are illustrated in Figure 7.

Hierarchical network architectures are usually only used for wireless networks spanning a range of coverage regions, as was shown in Figure 1. In this figure the lowest level of the hierarchy consists of indoor systems with small coverage areas, the next level of the hierarchy consists of cellular systems covering a city, followed by systems with regional and then global coverage. Since the coverage regions define a natural hierarchy of the overall network, a hierarchical network architecture along with hierarchical protocols for routing and identifying user locations are well-suited to this type of system. However, hierarchical architectures can be problematic for systems operating within a single coverage area where there is no natural hierarchy on which to base the architecture. In addition, hierarchical architectures tend to work best with hierarchical routing protocols, which are not very efficient for routing within a single coverage area. For these reasons systems designed for a single coverage region typically use either a star or a peer-to-peer architecture.

In a peer-to-peer architecture the nodes self-configure into an integrated network using distributed control, and the connection between any two nodes in the network consists of one or more peer-to-peer communication links. In a star architecture communication flows from network nodes to a central hub over one set of channels, and from the hub to the nodes over a separate set of channels. The choice of a peer-to-peer or a star network architecture depends on many factors. Peer-to-peer architectures require no existing infrastructure, are easily reconfigurable, and have no single points of failure, leading to very dynamic topologies. In addition, peer-to-peer architectures can use multiple hops for the end-to-end link, which has the advantage of extending the network range, and the disadvantage that if one of the hops fails, the entire end-to-end link is lost. However, this disadvantage is mitigated by the fact that each node may have connections to many other nodes, so there may be multiple ways to form an end-to-end connection with any other user. These advantages make peer-to-peer architectures the architecture of choice in military systems. Since star architectures have only one hop between a network node and the central hub, they tend to be more predictable and reliable, however if that connection is weak then there is no alternative connection. A big advantage of star architectures is that they can use centralized control functions at the hub for channel estimation, access, routing, and resource allocation. This centralized control usually results in a more efficient and reliable network, and for this reason many commercial wireless networks use the star architecture. Common examples of wireless systems with peer-to-peer architectures include packet radio networks and some wireless LANs, while the wireless star architecture is exemplified by cellular and paging systems.

1.5.2 Mobility Management

Mobility management consists of two related functions: location management and call routing. Location management is the process of identifying the physical location of user so that calls directed to that user can be routed to his location. Location management is also responsible for verifying the authenticity of users accessing the network. Routing consists of setting up a route through the network over which data directed to a particular user is sent, and dynamically reconfiguring the route as the user location changes. In cellular systems location management and routing is coordinated by the base stations or the central mobile telephone switching office (MTSO), whereas on the Internet these functions are handled by the Mobile Internetworking Routing Protocol (Mobile IP).

The location management and routing protocols in mobile IP and in cellular systems are somewhat different, but they both use local and remote data bases for user tracking, authentication, and call routing. In cellular systems location management and call routing are handled by the MTSO in each city. An MTSO is connected to all base stations in its city via high-speed communication links. The MTSO in each city maintains a home location database for local users and a visitor location database for visiting users. Calls directed to a particular mobile unit are routed through the public-switched telephone network to the MTSO in that mobile's home city. When a mobile unit in his home city turns on his handset, that signal is relayed by the local base station to the MTSO. The MTSO authenticates the ID number of the mobile and then registers that user in its home location database. After registration, any calls addressed to that user are sent to him by the MTSO via one of its base stations. If a mobile is roaming in a different city then, by turning on his handset (assuming he has roaming privileges in that city), the mobile registers with the MTSO in the visiting city. Specifically, the mobile's signal is picked up by a local base station in the visiting city, which relays the signal to the visiting city's MTSO. The visiting city's MTSO then sends a message to the MTSO in the mobile's home city requesting user authentication and call forwarding for that user. The MTSO in the mobile's home city authenticates the mobile's ID number, adds the location of the visiting city's MTSO to its home location database entry for the visiting mobile, and sends a confirmation message to the visiting city's MTSO. The visiting mobile is then registered in the visitor location database of the visiting city's MTSO. After this process is complete, when a call for a visiting mobile arrives at that mobile's home city, the home city MTSO sets up a circuit-switched connection with the visiting city's MTSO along which the call is routed. This method of call routing is somewhat inefficient, since a call must travel from its origin to the home city's MTSO and then be rerouted to the visiting city. The MTSO also coordinates handoffs between base stations by detecting when a mobile signal is becoming weak at its current base station and finding the neighboring base station with the best connection to that mobile. This handoff process will be discussed in more detail in Section 1.6.1.

Location management and routing on the Internet is handled by the Mobile IP protocol, an enhancement to the IP protocol for supporting user mobility. In the Internet, every node has a unique identifying address, its IP address. In the Mobile IP specification, every mobile host has a home network and a global IP home address. Software called the home agent resides on the home network, and this software is responsible for routing packets to each mobile host on this network. When a mobile host visits a different network, it must first discover a foreign agent (typically software residing on the visiting network) that will provide it with packet forwarding. The foreign agent provides the mobile host with a care-of-address, which is basically a temporary IP address where the mobile host can received packets on the visiting network. The mobile host registers its new care-of address with its home agent. All registration attempts must be carefully authenticated by the home agent, since otherwise a malicious user could hijack a mobile host's packets simply by furnishing its own IP address as this mobile's care-of address. After the registration process, if a sender directs packets to a visiting mobile at its home IP address, the home agent forwards these packets to the foreign agent by encapsulating them in a new packet with the mobile's care-of-address. The encapsulated packet is sent via conventional IP routing to the foreign agent, where it is deencapsulated and delivered to the mobile host. The mobile host can send packets back to the sender directly, but the sender's packets must always be routed through the home agent. This routing is inefficient because packets must be redirected by the home agent rather than routed directly to the mobile's visiting network. The mobile IP protocol does not support real-time handoff of a mobile between different networks: it is designed mainly for stationary users that occasionally move their computer from one network to another.

1.5.3 Network Reliability

An end-to-end connection in a wireless network is composed of one or more wireless and wired links, with at least one wireless link. These different links have widely varying data rates, BERs, and delays. Moreover, user mobility causes the characteristics of one or more of these links to change over time. These characteristics make it difficult to insure reliability of the end-to-end network connection. In particular reliability protocols like TCP that are designed for wired networks do not work well in wireless networks. That is because on today's wired networks link error rates are low and link data rates are high and easy to predict. Thus, network performance is largely determined by how the queues are managed within the switches. As a result, packet losses are due almost entirely to congestion-related queue overflows, and therefore TCP handles packets losses through congestion control. On wireless networks most packets losses are due to poor link quality and intermittent connectivity. Using the congestion control mechanisms of TCP to correct for these problems can cause large and variable end-to-end delays and low network throughput, as will be discussed in more detail in the next section. In addition, wireless channels have low data rates and high BERs, and the random characteristics of these channels make it difficult to guarantee or even predict end-to-end data rates, delay statistics, or packet loss probabilities.

Performance metrics such as data rates, end-to-end latency, and likelihood of packet loss are usually referred to as a connection's Quality of Service (QoS). The QoS requirements for a connection are based on the kind of data being transported over that connection. For example, voice has a high tolerance to packet loss but a low tolerance for delay, whereas data has the opposite requirements. QoS specifications are used by a network's admission control procedure to determine if a new connection can be initiated. A connection will not be initiated if its QoS requirements cannot be met throughout the duration of the connection. QoS is categorized by three types of service: guaranteed, predictive (or statistical), and best effort service. For guaranteed QoS the network determines that it can support the QoS requirements of the end-to-end connection throughout its duration. If so then the network's admission control procedure allows the connection to be established and guarantees its requirements will be met. Performance guarantees are very important for real-time, interactive applications, such as two-way audio or video. Guaranteed QoS is provided by the PSTN, which guarantees a certain voice quality and delay, as well as the asynchronous transfer mode (ATM) protocol. Both of these systems rely on actual or virtual circuits for the end-to-end connection to support the performance guarantees. In predictive QoS the network predicts the QoS available for a given connection and the application can choose to accept this predictive performance or not. For example, predictive QOS might offer a 12 msec round trip delay 90% of the time. No guarantees are given, but if the predictive performance is much better than the requirements of the application than the application will usually establish the connection. In best-effort service the network gives no information about the QoS on a given connection - it makes the best effort to transfer the data with minimum packet loss, minimum delay, and maximum data rate. Best effort services is appropriate for applications that do not demand real-time performance and that can gracefully adapt to varying performance in the network. In particular, best effort QoS is a good match for data that comes in interactive bursts, interactive bulk transfer, and asynchronous bulk transfer. However, the best effort model makes the implementation of real-time applications challenging. The inherent impairments and random variations of the wireless channel make it difficult to provide anything other than best effort service in wireless networks. This difficulty is the main challenge in supporting high-speed real-time applications like video teleconferencing over these networks. One method to compensate for the lack of QOS guarantees is to adapt at the application layer to the variable QOS offered by the network. This technique is described in more detail in Section 1.5.6.

1.5.4 Internetworking

In order to connect wired and wireless networks together they must share a common networking protocol. The two network protocols emerging today as standards are TCP/IP and ATM (Section 1.3). TCP/IP is compatible with most existing wired networks and has proven to work well over a range of wired subnets with vastly different performance. However, TCP has problems operating over wireless links, mainly due to its use of congestion control in response to packet delays. Specifically, in TCP a packet timeout occurs whenever a packet is delayed by more than a given threshold. In this case the packet is retransmitted, and the flow control window size is reduced. The window size is gradually increased if no more timeouts occur. This control mechanism works well when packet delays are due to congestion, since reducing the window size reduces flow. However, wireless links can experience large and variable delays, sporadic error bursts, and intermittent connectivity due to handoffs. Large and variable link delays cause large oscillations in the TCP sending rate, resulting in large and variable end-to-end delays. Error bursts can result in unnecessary retransmissions by TCP, since these errors are usually corrected at the link layer, and also cause significant throughput degradation, since flow is reduced in response to every error burst. The effect of intermittent connectivity on TCP is similar to that of error bursts, resulting in unnecessary retransmissions and throughput reduction. Various modifications to TCP have been proposed to address this issue, but none have emerged as a clear solution [9]. ATM is emerging as a standard for multimedia systems due to its QoS guarantees, which are required for these applications. However, ATM suffers from high overhead in its packet structure, and its not clear that the QoS guarantees inherent to ATM can be achieved in a wireless network. Eventually a dominant protocol will emerge for internetworking between wireless and wired networks. This protocol may be a modification to TCP/IP or ATM, a merging of these two protocols (e.g. TCP/IP over ATM), or something completely different.

1.5.5 Security

Wireless communication systems are inherently less private than wireline systems because the wireless link can be intercepted without any physical tap, and this interception cannot be detected by the transmitter or the receiver. This lack of link security also makes wireless networks more subject to usage fraud and activity monitoring than their wireline counterparts. Opportunities for fraudulent attacks will increase as services like wireless banking and commerce become available. Thus, security technology is an important challenge. Security issues can be broken down into three categories: network security, radio link security, and hardware security. Network security includes countermeasures to fraudulent access and monitoring of network activity, and end-to- end encryption. Radio link security entails preventing interception of the radio signal, ensuring privacy of user location information and, for military applications, anti-jam and low probability of interception and detection capabilities. Hardware security should prevent fraudulent use of the mobile terminal in the event of theft or loss, and user databases should also be secure against unauthorized access.

1.5.6 A New Paradigm for Wireless Network Design

Network design using the layered OSI architecture has worked fairly well for wired networks, especially as the communication links evolved to provide gigabit-per-second data rates and BERs of 10^{-12}. However, wireless channels typically have much lower data rates (tens or hundreds of Kbps for typical channels with high user mobility) and higher BERs (10^{-2}-10^{-6}) than wired channels, and also exhibit sporadic error bursts and intermittent connectivity. These performance characteristics also vary over time, as does the network topology and user traffic. This variability, coupled with the weak performance of wireless links in general, implies that good end-to-end wireless network performance will not be possible without a truly optimized, integrated, and adaptive network design.

The need for an integrated protocol design is most evident from the characteristics of the wireless link itself. The wireless link characteristics impact not only the link data rate, BER, and delay, but also user connectivity, multiple access, the overall network topology, and end-to-end delay and throughput. In other words, the wireless link characteristics impact all levels of the network protocol stack. Moreover, in order to optimize network performance each level in the protocol stack should adapt to wireless link variations in an appropriate manner, taking into account the adaptive strategies at the other layers. Therefore, an integrated adaptive design across all levels of the protocol stack is needed to best exploit interdependencies between protocol layers.

This integrated approach to adaptive protocol design entails two related questions: what performance metrics should be measured at each layer of the protocol stack, and what adaptive strategies should be developed for each layer of the protocol stack to best respond to variations in these local performance metrics. Network design based on the OSI model has looked at these two questions in isolation for each layer. But the best answer to both questions at a particular layer depends to a large extent on how and to what the network adapts at other layers of the protocol stack. In other words, the best overall network design requires that these questions be addressed at all layers of the protocol stack *simultaneously*.

The integrated adaptive protocol design should still be based on a hierarchical approach, since network variations take place on different time scales. Specifically, variations in link SNR (i.e. BER and connectivity) can be very fast, on the order of microseconds for vehicle-based users. Network topology changes more slowly, on the order of seconds, while variations of user traffic may change over tens to hundreds of seconds (although this may change as networks support more applications with short messaging). The different time scales of the network variations suggest a hierarchical approach for protocol design, since the rate at which a protocol can adapt to overall network changes is, to a large extent, determined by its location in the protocol stack. For example, suppose the link connectivity (link SNR) in the wireless link of an end-to-end network connection is weak. By the time this connectivity information is relayed to a higher level of the protocol stack (i.e. the network layer for rerouting or the application layer for reduced-rate compression), the link SNR may change. Therefore, it makes sense for each protocol layer to adapt to variations that are local to that layer. If this local adaptation is insufficient to compensate for the local performance degradation then the performance metrics at the next layer of the protocol stack will degrade as a result. Adaptation at this next layer may then correct or at least mitigate the problem that could not be fixed through local adaptation. For example, consider the weak link scenario. Link connectivity can be measured quite accurately and quickly at the link level. The link protocol can therefore respond to weak connectivity by increasing its transmit power or its error correction coding. This will correct for variations in connectivity due to, for example, multipath flat-fading. However, if the weak link is caused by something difficult to correct for at the link layer, e.g. the mobile unit is inside a tunnel, then it is better for a higher layer of the network protocol stack to respond by, for example, delaying packet transmissions until the mobile leaves the tunnel. However, real-time applications may not be able to tolerate an increase in packet delay, in which case the application can adapt by reducing its rate of packet transmission. This may entail using a lower rate compression scheme or sending only priority data (e.g. the voice component of a video stream or the low resolution components of an image). It is this integrated approach to adaptive networking - how each layer of the protocol stack should respond to local variations given adaptation at higher layers - that should be considered as a new paradigm in wireless network design.

1.6 Wireless Networks Today

In this section we give a brief overview of wireless networks in operation today, including cellular systems, cordless phones, wireless LANs, wide area wireless data systems, paging systems, and satellite systems. These systems are mainly differentiated by their application (voice or data), support for user mobility, and coverage areas.

1.6.1 Cellular Telephone Systems

Cellular telephone systems, also referred to as Personal Communication Systems (PCS), are extremely popular and lucrative worldwide: these systems have sparked much of the optimism about the future of wireless networks. Cellular telephone systems are designed to provide two-way voice communication at vehicle speeds with regional or national coverage. Cellular systems were initially designed for mobile terminals inside vehicles with antennas mounted on the vehicle roof. Today these systems have evolved to support lightweight handheld mobile terminals operating inside and outside buildings at both pedestrian and vehicle speeds.

The basic premise behind cellular system design is frequency reuse, which exploits path loss to reuse the same frequency spectrum at spatially-separated locations. Specifically, the coverage area of a cellular system is divided into nonoverlapping *cells* where some set of channels is assigned to each cell. This same channel set is used in another cell some distance away, as shown in Figure 8, where the shaded cells use the same channel set. Operation within a cell is controlled by a centralized base station, as described in more detail below. The interference caused by users in different cells operating on the same channel set is called intercell interference. The spatial separation of cells that reuse the same channel set, the *reuse distance*, should be as small as possible to maximize the spectral efficiency obtained by frequency reuse. However, as the reuse distance decreases, intercell interference increases, due to the smaller propagation distance between interfering cells. Since intercell interference must remain below a given threshold for acceptable system performance, reuse distance cannot be reduced below some minimum value. In practice it is quite difficult to determine this minimum value since both the transmitting and interfering signals experience random power variations due to path loss, shadowing, and multipath. In order to determine the best reuse distance and base station placement, an accurate characterization of signal propagation within the cells is needed. This characterization is usually obtained using detailed analytical models, sophisticated computer-aided modeling, or empirical measurements [44].

Initial cellular system designs were mainly driven by the high cost of base stations, approximation one million dollars apiece. For this reason early cellular systems used a relatively small number of cells to cover an entire city or region. The cell base stations were placed on tall buildings or mountains and transmitted at very high power with cell coverage areas of several square miles. These large cells are called macrocells. Signals propagated out from base stations uniformly in all directions, so a mobile moving in a circle around the base station would have approximately constant received power. This circular contour of constant power yields a hexagonal cell shape for the system, since a hexagon is the closest shape to a circle that can cover a given area with multiple nonoverlapping cells.

Cellular telephone systems are now evolving to smaller cells with base stations close to street level or inside buildings transmitting at much lower power. These smaller cells are called microcells or picocells, depending on their size. This evolution is driven by two factors: the need for higher capacity in areas with high user density and the reduced size and cost of base station electronics. A cell of any size can support roughly the same number of users if the system is scaled accordingly. Thus, for a given coverage area a system with many microcells has a higher number of users per unit area than a system with just a few macrocells. Small cells also have better propagation conditions since the lower base stations have reduced shadowing and multipath. In addition, less power is required at the mobile terminals in microcellular systems, since the terminals are closer to the base stations. However, the evolution to smaller cells has complicated network design. Mobiles traverse a small cell more quickly than a large cell, and therefore handoffs must be processed more quickly. In addition, location management becomes more complicated, since there are more cells within a given city where a mobile may be located. It is also harder to develop general propagation models for small cells, since signal propagation in these cells is highly dependent on base station placement and the geometry of the surrounding reflectors. In particular, a hexagonal cell shape is not a good approximation to signal propagation in microcells. Microcellular systems are often designed using square or triangular cell shapes, but these shapes have a large margin of error in their approximation to microcell signal propagation [24].

All base stations in a city are connected via a high-speed communications link to a mobile telephone switching office (MTSO). The MTSO acts as a central controller for the network, allocating channels within each cell, coordinating handoffs between cells when a mobile traverses a cell boundary, and routing calls to and from mobile users in conjunction with the public switched telephone network (PSTN). A new user located in a given cell requests a channel by sending a call request to the cell's base station over a separate control channel. The request is relayed to the MTSO, which accepts the call request if a channel is available in that cell. If no channels are available then the call request is rejected. A call handoff is initiated when the base station or the mobile in a given cell detects that the received signal power for that call is approaching a given minimum threshold. In this case the base station informs the MTSO that the mobile requires a handoff, and the MTSO then queries surrounding base stations to determine if one of these stations can detect that mobile's signal. If so then the MTSO coordinates a handoff between the original base station and the new base station. If no channels are available in the cell with the new base station then the handoff fails and the call is terminated. False handoffs may also be initiated if a mobile is in a deep fade, causing its received signal power to drop below the minimum threshold even though it may be nowhere near a cell boundary. User location management, authentication, and routing in cellular systems were described in Section 1.5.2.

Cellular telephone systems have also moved from analog to digital technology. Digital technology has many advantages over analog. The components are cheaper, faster, smaller, and require less power. Voice quality is improved due to error correction coding. Digital systems also have higher capacity than analog systems since they are not limited to FDMA multiple access, and they can take advantage of advanced compression techniques and voice activity factors. In addition, encryption techniques can be used to secure digital signals against eavesdropping. All cellular systems being deployed today are digital, and these systems provide voice mail, paging, and email services in addition to voice. Due to their lower cost and higher efficiency, service providers have used aggressive pricing tactics to encourage user migration from analog to digital systems. Since they are relatively new, digital systems do not always work as well as the old analog ones. Users experience poor voice quality, frequent call dropping, short battery life, and spotty coverage in certain areas. System performance will certainly improve as the technology and networks mature. However, it is unlikely that cellular phones will provide the same quality as wireline service any time soon. The great popularity of cellular systems indicates that users are willing to tolerate inferior voice communications in exchange for mobility.

Digital cellular systems can use any of the multiple access techniques described in Section 1.4.1 to divide up the signal bandwidth in a given cell. In the U.S. the standards activities surrounding the current generation of digital cellular systems provoked a raging debate on multiple access for these systems, resulting in several incompatible standards. In particular, there are two standards in the 900 MHz (cellular) frequency band: IS-54, which uses a combination of TDMA and FDMA, and IS-95, which uses semi-orthogonal CDMA. The spectrum for digital cellular in the 2 GHz (PCS) frequency band was auctioned off, so service providers could use an existing standard or develop proprietary systems for their purchased spectrum. The end result has been three different digital cellular standards for this frequency band: IS-136 (which is basically the same as IS-54 at a higher frequency), IS-95, and the European digital cellular standard GSM, which uses a combination of TDMA and slow frequency-hopping. The digital cellular standard in Japan in similar to IS-54 and IS-136 but in a different frequency band, and the GSM system in Europe is at a different frequency than the GSM systems in the U.S. This proliferation of incompatible standards in the U.S. and abroad makes it impossible to roam between systems nationwide or globally without using multiple phones (and phone numbers).

Efficient cellular system designs are *interference-limited*, i.e. the interference dominates the noise floor since otherwise more users could be added to the system. As a result, any technique to reduce interference in cellular systems leads directly to an increase in system capacity and performance. Some methods for interference reduction in use today or proposed for future systems include cell sectorization [44], directional and smart antennas [2], multiuser detection [55], and dynamic channel and resource allocation [31, 39].

1.6.2 Cordless Phones

Cordless telephones first appeared in the late 1970's and have experienced spectacular growth ever since. Roughly half of the phones in U.S. homes today are cordless. Cordless phones were originally designed to provide a low-cost low-mobility wireless connection to the PSTN, i.e. a short wireless link to replace the cord connecting a telephone base unit and its handset. Since cordless phones compete with wired handsets, their voice quality must be similar: initial cordless phones had poor voice quality and were quickly discarded by users. The first cordless systems allowed only one phone handset to connect to each base unit, and coverage was limited to a few rooms of a house or office. This is still the main premise behind cordless telephones in the U.S. today, although these phones now use digital technology instead of analog. In Europe and the Far East digital cordless phone systems have evolved to provide coverage over much wider areas, both in and away from home, and are similar in many ways to today's cellular telephone systems.

Digital cordless phone systems in the U.S. today consist of a wireless handset connected to a single base unit which in turn is connected to the PSTN. These cordless phones impose no added complexity on the telephone network, since the cordless base unit acts just like a wireline telephone for networking purposes. The movement of these cordless handsets is extremely limited: a handset must remain within range of its base unit. There is no coordination with other cordless phone systems, so a high density of these systems in a small area, e.g. an apartment building, can result in significant interference between systems. For this reason cordless phones today have multiple voice channels and scan between these channels to find the one with minimal interference. Spread spectrum cordless phones have also been introduced to reduce interference from other systems and narrowband interference.

In Europe and the Far East the second generation of digital cordless phones (CT-2, for cordless telephone, second generation) have an extended range of use beyond a single residence or office. Within a home these systems operate as conventional cordless phones. To extend the range beyond the home base stations, also called *phone-points* or *telepoints*, are mounted in places where people congregate, like shopping malls, busy streets, train stations, and airports. Cordless phones registered with the telepoint provider can place calls whenever they are in range of a telepoint. Calls cannot be received from the telepoint since the network has no routing support for mobile users, although some newer CT-2 handsets have built-in pagers to compensate for this deficiency. These systems also do not handoff calls if a user moves between different telepoints, so a user must remain within range of the telepoint where his call was initiated for the duration of the call. Telepoint service was introduced twice in the United Kingdom and failed both times, but these systems grew rapidly in Hong Kong and Singapore through the mid 1990's. This rapid growth deteriorated quickly after the first few years, as cellular phone operators cut prices to compete with telepoint service. The main complaint about telepoint service was the incomplete radio coverage and lack of handoff. Since cellular systems avoid these problems, as long as prices were competitive there was little reason for people to use telepoint services. Most of these services have now disappeared.

Another evolution of the cordless telephone designed primarily for office buildings is the European DECT system. The main function of DECT is to provide local mobility support for users in an in-building private branch exchange (PBX). In DECT systems base units are mounted throughout a building, and each base station is attached through a controller to the PBX of the building. Handsets communicate to the nearest base station in the building, and calls are handed off as a user walks between base stations. DECT can also ring handsets from the closest base station. The DECT standard also supports telepoint services, although this application has not received much attention, probably due to the failure of CT-2 services. There are currently around 7 million DECT users in Europe, but the standard has not yet spread to other countries.

The most recent advance in cordless telephone system design is the Personal Handyphone System (PHS) in Japan. The PHS system is quite similar to a cellular system, with widespread base station deployment supporting handoff and call routing between base stations. With these capabilities PHS does not suffer from the main limitations of the CT-2 system. Initially PHS systems enjoyed one of the fastest growth rates ever for a new technology. In 1997, two years after its introduction, PHS subscribers peaked at about 7 million users, and has declined slightly since then due mainly to sharp price cutting by cellular providers. The main difference between a PHS system and a cellular system is that PHS cannot support call handoff at vehicle speeds. This deficiency is mainly due to the dynamic channel allocation procedure used in PHS. Dynamic channel allocation greatly increases the number of handsets that can be serviced by a single base station, thereby lowering the system cost, but it also complicates the handoff procedure. It is too soon to tell if PHS systems will go the same route as CT-2. However, it is clear from the recent history of cordless phone systems that to extend the range of these systems beyond the home requires either the same functionality as cellular systems or a significantly reduced cost.

1.6.3 Wireless LANs

Wireless LANs provide high-speed data within a small region, e.g. a campus or small building, as users move from place to place. Wireless devices that access these LANs are typically stationary or moving at pedestrian speeds. Nearly all wireless LANs in the United States use one of the ISM frequency bands. The appeal of these frequency bands, located at 900 MHz, 2.4 GHz, and 5.8 GHz, is that an FCC license is not required to operate in these bands. However, this advantage is a double-edged sword, since many other systems operate in these bands for the same reason, causing a great deal of interference between systems. The FCC mitigates this interference problem by setting a stringent limit on the power per unit bandwidth for ISM-band systems. To satisfy this requirement wireless LANs use either direct sequence or frequency hopping spread spectrum so that their total power is spread over a wide bandwidth. Wireless LANs can have either a star architecture, with wireless access points or hubs placed throughout the coverage region, or a peer-to-peer architecture, where the wireless terminals self-configure into a network.

Dozens of wireless LAN companies and products appeared the early 1990's to capitalize on the "pent-up demand" for high-speed wireless data. These first wireless LANs were based on proprietary and incompatible protocols, although most operated in the 900 MHz ISM band using direct sequence spread spectrum with data rates on the order of 1-2 Mbps. Both star and peer-to-peer architectures were used. The lack of standardization for these products led to high development costs, low-volume production, and small markets for each individual product. Of these original products only a handful remain, including Proxim's RangeLAN, Lucent's WaveLAN, and Windata's FreePort. Only one of the first generation wireless LANs, Motorola's Altair, operated outside the 900 MHz ISM band. This system, operating in the licensed 18 GHz band, had data rates on the order of 6 Mbps. However, performance of Altair was hampered by the high cost of components and the increased path loss at 18 GHz. As a result Altair was recently discontinued.

The 900 MHz ISM band is not available in most parts of the world, so the new generation of wireless LANs operate in the 2.4 GHz ISM band, which is available worldwide. A wireless LAN standard for this frequency band, the IEEE 802.11 standard, was recently completed to avoid some of the problems with the proprietary first generation systems. The standard specifies frequency hopped spread spectrum with data rates of 1.6 Mbps and a range of approximately 500 ft. The network architecture can be either star or peer-to-peer. Many companies have developed products based on the 802.11 standard, and these products are constantly evolving to provide higher data rates and better coverage. Cabletron's Freelink is the only existing wireless LAN in the 5.8 GHz range: this system has slightly higher data rates and slightly lower range than the 802.11 systems. Because data rates are low and coverage is limited, the market for all wireless LANs has remained relatively flat (around $200 million, far below the billion dollar market of today's cellular systems.) Optimism remains high that the wireless LAN market is poised to take off, although this prediction has been made every year since the inception of wireless LANs yet the market has so far failed to materialize.

1.6.4 Wide Area Wireless Data Services

Wide area wireless data services in the U.S. provide low rate wireless data to high-mobility users over a very large coverage area. In this section we briefly describe a few of these services: a more in-depth discussion can be found in [37]. The initial two service providers for wireless data were the ARDIS network run by Motorola and RAM Mobile Data, which uses Ericcson's Mobitex technology. ARDIS and RAM Mobile Data provide service to most metropolitan areas within the U.S. In these systems a large geographical region is serviced by a few base stations mounted on towers, rooftops, or mountains and transmitting at high power. In ARDIS the base stations are connected to network controllers attached to a backbone network, whereas in RAM Mobile Data the base stations are at the bottom of a hierarchical network architecture. Both systems use a form of the ALOHA protocol for random access with collision reduction through either a busy tone transmission or carrier sensing. Initial data rates for these systems were low, 4.8 Kbps for ARDIS and 8 Kbps for RAM, but these rates have now increased to 19.2 Kbps for both systems.

Another wide area wireless data provider is Metricom, with systems operating in the San Francisco Bay Area, Seattle, and Washington D.C. The Metricom architecture is similar to that of microcell systems: a large network of small inexpensive base stations with small coverage areas are mounted close to street level. The increased efficiency of microcells allows for higher data rates in Metricom, 76 Kbps, than in the other wide-area wireless data systems. Metricom uses frequency hopped spread spectrum in the 900 MHz ISM band, with power control to minimize interference and improve battery life.

The cellular digital packet data (CDPD) system is a wide area wireless data service overlayed on the analog cellular telephone network. CDPD shares the FDMA voice channels of the analog systems, since many of these channels are idle due to the growth of digital cellular. The CDPD service provides packet data transmission at rates of 19.2 Kbps, and is available throughout the U.S.

All of these wireless data services have failed to grow as rapidly or to attract as many subscribers as initially predicted, especially in comparison with the rousing success of wireless voice systems. There is disagreement on why these systems have experienced such anemic growth. Data rates for these systems are clearly low, especially in comparison with their wireline counterparts. Pricing for these services also remains high. There is a perceived lack of "killer applications" for wireless data: while voice communication on the move seems essential for a large part of the population, most people can wait until they have access to a phone line or wired network for data exchange. This may change with the proliferation of laptop and palmtop computers and the explosive demand for constant Internet access and email exchange. Optimists point to these factors as the drivers for wireless data but, as with wireless LANs, wide area wireless data services have been the pot of gold around the corner for many years yet have so far failed to deliver on these high expectations.

1.6.5 Paging Systems

Paging systems provide very low rate one-way data services to highly mobile users over a very wide coverage area. Paging systems have experienced steady growth for many years and currently serve about 56 million customers in the United States. However, the popularity of paging systems is declining as cellular systems become cheaper and more ubiquitous. In order to remain competitive paging companies have slashed prices, and few of these companies are currently profitable. To reverse their declining fortunes, a consortium of paging service providers have recently teamed up with Microsoft and Compaq to incorporate paging functionality and Internet access into palmtop computers [48].

Paging systems broadcast a short paging message simultaneously from many tall base stations or satellites transmitting at very high power (hundreds of watts to kilowatts). Systems with terrestrial transmitters are typically localized to a particular geographic area, such as a city or metropolitan region, while geosynchronous satellite transmitters provide national or international coverage. In both types of systems no location management or routing functions are needed, since the paging message is broadcast over the entire coverage area. The high complexity and power of the paging transmitters allows low-complexity, low-power, pocket paging receivers with a long usage time from small and lightweight batteries. In addition, the high transmit power allows paging signals to easily penetrate building walls. Paging service also costs less than cellular service, both for the initial device and for the monthly usage charge, although this price advantage has declined considerably in recent years. The low cost, small and lightweight handsets, long battery life, and ability of paging devices to work almost anywhere indoors or outdoors are the main reasons for their appeal.

Some paging services today offer rudimentary (1 bit) answer-back capabilities from the handheld paging device. However, the requirement for two-way communication destroys the asymmetrical link advantage so well exploited in paging system design. A paging handset with answer-back capability requires a modulator and transmitter with sufficient power to reach the distant base station. These requirements significantly increase the size and weight and reduce the usage time of the handheld pager. This is especially true for paging systems with satellite base stations, unless terrestrial relays are used.

1.6.6 Satellite Networks

Satellite systems provide voice, data, and broadcast services with widespread, often global, coverage to high-mobility users as well as to fixed sites. Satellite systems have the same basic architecture as cellular systems, except that the cell base-stations are satellites orbiting the earth. Satellites are characterized by their orbit distance from the earth. There are three main types of satellite orbits: low-earth orbit (LEOs) at 500-2000 Kms, medium-earth orbit (MEO) at 10,000 Kms, and geosynchronous orbit (GEO) at 35,800 Kms. A geosynchronous satellite has a large coverage area that is stationary over time, since the earth and satellite orbits are synchronous. Satellites with lower orbits have smaller coverage areas, and these coverage areas change over time so that satellite handoff is needed for stationary users or fixed point service.

Since geosynchronous satellites have such large coverage areas just a handful of satellites are needed for global coverage. However, geosynchronous systems have several disadvantages for two-way communication. It takes a great deal of power to reach these satellites, so handsets are typically large and bulky. In addition, there is a large round-trip propagation delay: this delay is quite noticeable in two-way voice communication. Recall also from Section 1.6.1 that high-capacity cellular systems require small cell sizes. Since geosynchronous satellites have very large cells, these systems have small capacity, high cost, and low data rates, less than 10 Kbps. The main geosynchronous systems in operation today are the global Inmarsat system, MSAT in North America, Mobilesat in Australia, and EMS and LLM in Europe.

The trend in current satellite systems is to use the lower LEO orbits so that lightweight handheld devices can communicate with the satellites and propagation delay does not degrade voice quality. The best known of these new LEO systems are Globalstar, Iridium, and Teledesic. Both Globalstar and Iridium provide voice and data services to globally-roaming mobile users at data rates under 10 Kbps. These systems require roughly 50 satellites to maintain global coverage. Teledesic uses 288 satellites to provide global coverage to fixed-point users at data rates up to 2 Mbps. The cell size for each satellite in a LEO system is much larger than terrestrial macrocells or microcells, with the corresponding decrease in capacity associated with large cells. Cost of these satellites, to build, to launch, and to maintain, is also much higher than that of terrestrial base stations, so these new LEO systems are unlikely to be cost-competitive with terrestrial cellular and wireless data services. Although these LEO systems can certainly complement these terrestrial systems in low-population areas, and are also appealing to travelers desiring just one handset and phone number for global roaming, it remains to be seen if there are enough such users willing to pay the high cost of satellite services to make these systems economically viable.

1.6.7 Other Wireless Systems and Applications

Many other commercial systems using wireless technology are on the market today. Remote sensor networks which collect data from unattended sensors and transmit this data back to a central processing location are being used for both indoor (equipment monitoring, climate control) and outdoor (earthquake sensing, remote data collection) applications. Satellite systems that provide vehicle tracking and dispatching (OMNITRACs) are very successful. Satellite navigation systems (the Global Positioning System or GPS) are also widely used for both military and commercial purposes. A new wireless system for Digital Audio Broadcasting (DAB) has recently been introduced in Europe. New systems and standards that will become operational within the next few years are also emerging, as we discuss in more detail in the next section.

1.7 Future Systems and Standards

In this section we describe some of the wireless systems and standards that will emerge over the next few years. The wide range of activity described in this section indicates that today's systems are far from adequate to meet the expected demand for wireless, and that the capabilities and technologies for future wireless systems will continue to evolve and develop. The interest and activity in wireless networking is so intense that it is impossible to predict what systems will emerge more than a year or two into the future.

1.7.1 Wireless LANs

There are two major activities in future wireless LAN development: HIPERLAN Type 1 and Wireless ATM. HIPERLAN (for high performance radio LAN), being developed in Europe, is a family of wireless LAN standards tailored to different kinds of users [7]. The four members of the HIPERLAN family are classified as Types 1-4. HIPERLAN Type 1 is similar in terms of protocol support to the IEEE 802.11 wireless LAN standard, while HIPERLAN Types 1-3 support Wireless ATM. HIPERLAN Type 1 operates in the 5 GHz frequency band with data rates of 23 Mbps at a range of 150 feet. The network architecture is peer-to-peer, and the channel access mechanism uses a variation of ALOHA with prioritization based on the lifetime of packets. HIPERLAN Type 1 promises an order of magnitude data rate improvement over today's technology, making it competitive with 100 Mbps wired Ethernets.

Wireless ATM [7, 1] is a standard to extend ATM capabilities to wireless local access as well as wireless broadband services, a significant challenge given the difficultly of supporting ATM's QoS guarantees over a wireless medium. The standardization process for Wireless ATM is ongoing and involves, among others, the Wireless ATM working group of the ATM Forum and the Broadband Radio Access Networks group of ETSI. The Wireless ATM standard developed by this latter group will become the HIPERLAN Type 2 standard. Development of HIPERLAN standards for Types 2 and 3 has not yet started. The wireless ATM standard has two components: the radio access technology and enhancements to the existing ATM protocol for support of mobile terminals. There are different technologies and protocols currently under consideration for both of these components. A major factor in the eventual success of wireless ATM will be independent of its technological success: if wired networks of the future use ATM only for long-haul networks then wireless ATM cannot be supported at the local access point, however if these networks support end-to-end ATM then wireless access to these networks must also support the ATM protocol.

The FCC recently set aside 300 MHz of unlicensed spectrum around 5 GHz called the National Information Infrastructure (NII) frequency band. This allocation frees up a large amount of unlicensed spectrum for wireless LAN applications with data rates up to several tens of megabits per second. The NII band is divided into three 100 MHz blocks with different power restrictions (and corresponding coverage areas) in each block. The middle block is designated for campus-area wireless LANs and is compatible with the 5 GHz HIPERLAN spectral allocation in Europe so that products developed in both places can be used interchangeably. The lower block is restricted to indoor use and the higher block is designated for community networks. In all three blocks the FCC has imposed minimal technical restrictions to provide maximum flexibility in system design. There is currently much ongoing activity to develop standards and products for these frequency bands, but no consensus has yet emerged on the best choice of system design.

1.7.2 Ad-Hoc Wireless Networks

An ad-hoc wireless network is a collection of wireless mobile hosts forming a temporary network without the aid of any established infrastructure or centralized control. Ad-hoc wireless networks were traditionally of interest to the military, since these networks can be rapidly deployed and reconfigured, and do not have single points-of-failure due to their lack of infrastructure and decentralized control. Throughout the 70s and 80s DARPA funded much work in the design of ad-hoc packet radio networks, however the performance of these networks was somewhat disappointing [32]. Ad-hoc wireless networking is currently experiencing a resurgence of interest due to new applications and improved technology. These networks are now being considered for many commercial applications, including in-home networking, wireless LANs, nomadic computing, and short-term networking for disaster relief, public events, and temporary offices. Both the IEEE 802.11 and HIPERLAN Type 1 wireless LAN standards support ad-hoc wireless networking within a small area, and wider area networks are currently under development.

Ad-hoc networks require a peer-to-peer architecture, and the topology of the network depends on the location of the different users, which changes over time. In addition, since the propagation range of a given mobile is limited, the mobile may need to to enlist the aid of other mobiles in forwarding a packet to its final destination. Thus, the end-to-end connection between any two mobile hosts may consist of multiple wireless hops. It is a significant technical challenge to provide reliable high-speed end-to-end communications in ad-hoc wireless networks given their dynamic network topology, decentralized control, and multihop connections.

Much of the current research in ad-hoc wireless network design is focused on distributed routing. Every mobile host in a wireless ad-hoc network must operate as a router in order to maintain connectivity information and forward packets from other mobiles. Routing protocols designed for wired networks are not appropriate for this task, since they either lack the ability to quickly reflect the changing topology, or may require excessive overhead. Proposed approaches to distributed routing that quickly adapt to changing network topology without excessive overhead include dynamic source and associativity-based routing [29, 52]. Other protocols that address some of the difficulties in supporting multimedia applications over ad-hoc wireless networks include rate-adaptive compression, power control, and resource allocation through radio clustering [5].

1.7.3 IMT-2000

International Mobile Telecommunications 2000 (IMT-2000) is a worldwide standard sponsored by the International Telecommunications Union (ITU), a United Nations organization responsible for developing global telecommunication standards[2]. In 1992 the ITU set aside 230 MHz of global spectrum in the 2 GHz frequency band to provide wireless access to the global telecommunications infrastructure through both satellite and land-based systems. The initial goal of this spectral allocation was to facilitate a move

[2] The standard was initial called the Future Public Land Mobile Telecommunications (FPLMTS) standard.

from the worldwide collection of different second-generation wireless systems, each with its own set of standards, services, coverage areas, and spectral allocations, to a worldwide standard serving fixed and mobile users in both public and private networks.

IMT-2000 is designed to support a wide range of services including voice, high-rate and variable-rate data, and multimedia in both indoor and outdoor environments. A family of radio protocols suitable for a range of environments and applications is being developed to support these requirements. The goal for the protocol family is to maximize commonality within the family while maintaining flexibility to adapt to the different environments and applications.

The IMT-2000 standard specifies data rates of 2 Mbps for local coverage and 384 Kbps for wide-area coverage, and these rates can be variable depending on link and network conditions. It also supports both continuous stream and packet data. The requirement for packet data support is the main differentiator between IMT-2000 and digital cellular systems. IMT-2000 does not require backward compatibility with current wireless systems, so the design of the radio access and networking protocols can incorporate new technologies and innovations. A high majority of the proposals for the IMT-2000 radio transmission technology submitted from Asia, Europe, and North America, are based on wideband CDMA technology [36], and it appears that some form of this technology will be adopted as the radio access standard. Advantages of wideband CDMA include its capacity and coverage gain from frequency diversity, its ease of use in packet data transfer, its flexibility for different services, its asynchronous operation, and its built-in support for adaptive antenna arrays, multiuser detection, hierarchical cell structures, and transmitter diversity. Standardization of the IMT-2000 networking protocol is still several years away.

1.7.4 High Speed Digital Cellular

All of the digital cellular standards are undergoing enhancements to support high rate packet data transmission. The goal of these enhancements is for these systems to support the IMT-2000 data rate specification of 384 Kbps over wide areas. In the near term, GSM systems will provide data rates of up to 100 Kbps by aggregating all timeslots together for a single user. The enhancement to support 384 Kbps, called Enhanced Data Services for GSM Evolution (EDGE), increases the data rate further by using a high-level modulation format combined with FEC coding. This modulation is more sensitive to fading effects, and EDGE uses adaptive modulation and coding to mitigate this problem. Specifically, EDGE defines six different modulation and coding combinations, each optimized to a different value of received SNR. The received SNR is measured at the receiver and fed back to the transmitter, and the best modulation and coding combination for this SNR value is used.

The IS-54 and IS-136 systems currently provide data rates of 40-60 Kbps by aggregating time slots and using high-level modulation. These TDMA standards will support 384 Kbps by migrating to the GSM EDGE standard. This new TDMA standard is referred to as IS-136HS (high-speed). The IS-95 systems will support higher data rates by evolving to the wideband CDMA standard in IMT-2000, however it is not yet clear if that standard will be backward compatible with the IS-95 systems.

1.7.5 Fixed Wireless Access

Fixed wireless access provides wireless communications between a fixed access point and multiple terminals. These systems were initially proposed to support interactive video service to the home, but the application emphasis has now shifted to providing high speed data access (tens of Mbps) to the Internet, the WWW, and to high speed data networks for both homes and businesses. In the U.S. two frequency bands have been set aside for these systems: part of the 28 GHz spectrum is allocated for local distribution systems (local multipoint distribution systems or LMDS) and a band in the 2 GHz spectrum is allocated for metropolitan distribution systems (multichannel multipoint distribution services or MMDS). LMDS represents a quick means for new service providers to enter the already stiff competition among wireless and wireline broadband service providers. MMDS is a television and telecommunication delivery system with transmission ranges of 30-50 km. MMDS has the capability to deliver over one hundred digital video TV channels along with telephony and access to emerging interactive services such as the Internet. MMDS will mainly compete with existing cable and satellite systems.

1.7.6 HomeRF and Bluetooth

HomeRF is an RF standard in the 2 GHz frequency band for wireless home networking. The standard was initiated by Intel, HP, Microsoft, Compaq, and IBM to enable communications and Internet connectivity among different electronic devices in and around the home, including PCs, laptops, smart pads, and intelligent home appliances. The data rate for HomeRF is specified as 2 Mbps, with simultaneous support for voice and data, at a range of 50 meters. The HomeRF standard is expected to be finalized sometime in 1999, with products incorporating the standard introduced sometime in the year 2000.

Bluetooth is a cable-replacement RF technology for short range connections (less than 10 meters) between wireless devices. Its main application is to connect digital cellular phones, laptop and palmtop computers, portable printers and projectors, network access points, and other portable devices without the need to carry or connect cables. The Bluetooth standard was initiated by Ericsson, IBM, Intel, Nokia, and Toshiba, and has since been adopted by over 200 telecommunications and computer companies. Products compatible with Bluetooth should appear in late 1999. The system operates in the 2.4 GHz frequency band with data rates of 700 Kbps for data and up to three voice connections at 64 Kbps.

1.8 Summary

A desire for mobility coupled with the demand for voice, Internet, and multimedia services indicates a bright future for wireless networks. Digital cellular and paging systems have enjoyed enormous growth, yet current products and services for wireless data have not lived up to expectations. This is due mainly to their high cost and poor performance. New standards and systems are emerging worldwide to address these performance and cost issues. These system support a wide range of voice, data, and multimedia services for fixed and mobile users both indoors and out, in citys, rural areas, and remote regions.

There are many technical challenges to overcome in building high-performance wireless networks. The wireless channel is a difficult communications medium. Sophisticated techniques exist to compensate for many of the channel impairments, but these can entail significant cost and complexity. The spectrum must also be used extremely efficiently through advanced link layer, access, and cellular system design. Networking protocols to support roaming users and end-to-end QoS guarantees also pose a significant technical challenge. The limited size and battery life of mobile terminals impose significant complexity constraints, so complexity must be distributed throughout the network to compensate for this limitation. Finally, the unpredictable nature of the wireless channel requires adaptation across all levels of the wireless network design: the link layer, network layer, transport layer, and application layer. This requires interaction between these layers, which violates the traditional network design paradigm of designing each layer in the OSI model independently from the others. While this paradigm has worked well on wired networks, especially as wired technology has evolved to the high performance of today's networks, high-performance wireless networks will not possible without significant technical breakthroughs at all levels of the system design as well as an integrated and adaptable design for the overall network.

1.9 Notes

The wireless channel has been characterized in many books and articles over the last 30 years: [44, 38] and the references therein describe the basic models for both indoor and outdoor systems. A tutorial on wireless infrared communication systems can be found in [30]. Recent work on the capacity of wireless channels is summarized in [14]. Link level design for wireless channels is treated in many textbooks, including [40, 44, 51]. Spread spectrum for both ISI mitigation and multiple access is described in [17, 56]. A special issue of the *IEEE Personal Communications Magazine* is devoted to smart antennas and their use in wireless systems [2]. Several textbooks provide additional details on multiple and random access for wireless networks [44, 51, 37, 13]. Wireless network design is still an active area of research, and there is no definitive reference for this field. More details on the cellular phone systems and wide area wireless data services can be found in [37]. Emerging satellite systems are described in [4]. Future systems and standards are continually evolving: the best source of information in this area is the standards bodies and companies building the systems.

1.10 Problems

1. Consider a path loss model
$$\frac{P_r}{P_t} = \left(\frac{d_0}{d}\right)^{-\alpha},$$
where P_r is the received signal power, P_t is the transmitted signal power, $d_0 = 100$ meters is a propagation constant, d is the propagation distance, and α is the path loss exponent. If $P_t = 100$ milliwatts and $P_r = 1$ milliwatt is required for acceptable performance, what is the maximum transmission range of our system for a path loss exponent $\alpha = 2$? By how much would the transmission range decrease if the path loss exponent was $\alpha = 4$?

2. Shadow fading often follows a log-normal distribution, which means that the dB value of the received power $10\log_{10}(P_r)$ follows a Gaussian distribution. Suppose the received signal power follows this log-normal distribution. Assume the average value of $10\log_{10}(P_r/N)$ equals 10 dB and its variance equals 4 dB, where N is assumed to be constant. What is the probability that the received SNR is less than 5 dB.

3. Consider a channel with a multipath delay spread of 10 microseconds. Suppose a voice signal with a signal bandwidth of 30 KHz is transmitted over this channel. Will the channel exhibit flat or frequency-selective fading? How about for a data signal with a 1 MHz signal bandwidth?

4. The BER of binary phase-shift-keying with differential detection in white Gaussian noise (i.e. no fading, shadowing, or ISI) is $.5e^{-\gamma}$, where γ is the average received SNR. If there is also Rayleigh fading then the BER becomes $[2(1+\gamma)]^{-1}$. Consider a data system with a required BER of 10^{-6}. What average SNR is required to achieve this target BER both with and without Rayleigh fading?

5. The error floor due to channel Doppler for binary phase-shift-keying with differential detection is $.5(\pi f_D T_b)^2$, where f_D is the channel Doppler and T_b is the bit time (the inverse of the data rate). For $f_D = 80$ Hz and a data rate of 20 Kbps what is the error floor? How about for $f_D = 80$ Hz and a data rate of 1 Mbps? Why does the error floor decrease as the data rate increases?

6. How many independent diversity paths can be obtained using an antenna array mounted on a laptop of length .1 meter, assuming a carrier frequency of 5 GHz?

7. Consider a wireless system with total bandwidth of 3 MHz. How many users can share this channel using FDMA, where each user is assigned a 30 KHz channel? Suppose instead we use semi-orthgonal CDMA for multiple access. If each user has a received signal power of P and the interference power caused by that user to other users is $.01P$, for a required signal-to-interference power ratio of 10 how many users can be accommodated in this system?

8. Radio signals travel at the speed of light, equal to $3 * 10^8$ meters per second. What is the maximum orbit distance of a satellite such that the round trip propagation delay of a signal does not exceed the 100 microsecond delay constraint of voice systems. Based on this calculation determine which of the three satellite orbit distances, GEO, LEO, and MEO, can support voice services.

9. The IEEE 802.11 wireless LAN standard supports both star and peer-to-peer architectures. Describe a wireless LAN application that is well-suited to each type of architecture.

10. Discuss the differences between ad-hoc wireless networks and cellular networks in terms of network architecture and mobility management. What are the advantages and disadvantages of each network design? What applications are best suited to each type of network?

11. Spectrum for new wireless systems is being allocated by the FCC at higher frequencies (e.g. 5 GHz and 28 GHz) than in existing systems. What are the advantages and disadvantages of building systems at these higher frequencies?

Bibliography

[1] *IEEE Personal Communications Magazine: Special Issue on Wireless ATM*, August 1996.

[2] *IEEE Pers. Commun. Mag: Special Issue on Smart Antennas*, February 1998.

[3] N. Abramson. Wide-band random-access for the last mile. *IEEE Personal Communications Magazine*, pages 29–33, December 1996.

[4] F. Abrishamkar and Z. Siveski. PCS global mobile satellites. *IEEE Commun. Mag.*, pages 132–136, September 1996.

[5] A. Alwan, R. Bagrodia, N. Bambos, M. Gerla, L. Kleinrock, J. Short, and J. Villasenor. Adaptive mobile multimedia networks. *IEEE Personal Communications Magazine*, pages 34–51, April 1996.

[6] R. Ananasso and F. D. Priscoli. The role of satellites in personal communication services. *IEEE Journal on Selected Areas of Communications: Special Issue on mobile satellite communications for seamless PCS* pages 180-196, February 1995.

[7] G. Anastasi, L. Lenzini, and E. Mingozzi. Mac protocols for wideband wireless local access: evolution toward wireless atm. *IEEE Personal Communications Magazine*, pages 53–64, October 1998.

[8] J. B. Andersen, T. S. Rappaport, and S. Yoshida. Propagation measurements and models for wireless communications channels. *IEEE Commun. Mag.*, pages 42–49, January 1995.

[9] H. Balakrishnan, V.N. Padmanabhan, S. Seshan, and R.H. Katz. A comparison of mechanisms for improving tcp performance over wireless links. *IEEE/ACM Transactions on Networking*, pages 756–769, December 1997.

[10] C. A. Belfiore Jr. and J. H. Park. Decision-feedback equalization. *IEEE Proc.*, pages 1143–1156, August 1979.

[11] C. Bergljung and L. G. Olsson. Rigorous diffraction theory applied to street microcell propagation. In *Proc. Global Commun. Conf. (Globecom'91)*, pages 1292–1296, December 1991.

[12] C. Berrou, A. Glavieux, and P. Thitimajshima. Near shannon limit error-correcting coding: Turbo codes. In *Proc. IEEE Intl. Commun. Conf. (ICC'93)*, pages 1064–1070, May 1993.

[13] D. P. Bertsekas and R. G. Gallager. *Data Networks.* Englewood Cliffs, NJ: Prentice-Hall, 1992.

[14] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: information theoretic and communications aspects. *IEEE Transactions on Information Theory*, pages 2619–2692, October 1998.

[15] J. A. C. Bingham. Multicarrier modulation for data transmission: An idea whose time has come. *IEEE Commun. Mag.*, pages 5–14, May 1990.

[16] L. J. Cimini. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.*, pages 665–675, July 1985.

[17] R.C. Dixon. *Spread Spectrum Systems with Commercial Applications.* New York, NY: John Wiley & Sons, Inc., 3rd edition, 1994.

[18] C. Douillard, M. Jezequel, and C. Berrou. Iterative correction of intersymbol interference: Turbo equalization. *Europ. Trans. Telecommun.*, pages 507–511, September/October 1995.

[19] P. F. Driessen. Gigabit/s indoor wireless systems with directional antennas. *IEEE Trans. Commun.*, pages 1034–1043, August 1996.

[20] A. Duel-Hallen, J. Holtzman, and Z. Zvonar. Multiuser detection for cdma systems. *IEEE Personal Communications Magazine*, pages 46–58, April 1995.

[21] G.D. Forney Jr. Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference. *IEEE Trans. Inform. Theory*, pages 363–378, May 1972.

[22] R.G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[23] A. J. Goldsmith and S.-G Chua. Variable-rate variable-power M-QAM for fading channels. *IEEE Trans. Commun.*, pages 1218–1230, October 1997.

[24] A. J. Goldsmith and L.J. Greenstein. A measurement-based model for predicting coverage areas of urban microcells. *IEEE Journal on Selected Areas in Communication*, pages 1013–1023, September 1993.

[25] A. J. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. on Information Theory*, pages 1896–1992, November 1997.

[26] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi. Packet reservation multiple access for local wireless communications. *IEEE Transactions on Communications*, pages 885–890, August 1989.

[27] H. Hashemi. The indoor radio propagation channel. *Proc. IEEE*, pages 943–968, July 1993.

[28] C. Heegard and S. B. Wicker. *Turbo Coding.* Norwell, MA: Kluwer Academic Publishers, 1999.

[29] D. Johnson and D. Maltz. Protocols for adaptive wireless and mobile networking. *IEEE Personal Communications Magazine*, pages 34–42, February 1996.

[30] J.M. Kahn and J.R. Barry. Wireless infrared communications. *Proc. IEEE*, pages 265–298, February 1997.

[31] I. Katzela and M. Naghshineh. Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey. *IEEE Pers. Commun. Mag.*, pages 10–22, June 1996.

[32] B. Leiner, N. Nielson, and F. Tobagi. *IEEE Proceedings: Special Issue on Packet Radio Networks*, January 1987.

[33] S. Lin and D. J. Costello. *Error Control Coding : Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1983.

[34] V.H. McDonald. The Cellular Concept. *Bell Systems Technical Journal*, pages 15–49, January 1979.

[35] J. W. McKown Jr. and R. L. Hamilton. Ray tracing as a design tool for radio networks. *IEEE Network Mag.*, pages 27–30, November 1991.

[36] T. Ojanpera and R. Prasad. An overview of air interface multiple access for IMT-2000/UMTS. *IEEE Communications Magazine*, pages 82-95, September 1998.

[37] K. Pahlavan and A. H. Levesque. *Wireless Information Networks.* New York, NY: John Wiley & Sons, Inc., 1995.

[38] D. Parsons. *The Mobile Radio Propagation Channel.* New York, NY: John Wiley & Sons, 1992.

[39] G. Pottie. System design choices in personal communications. *IEEE Pers. Commun. Mag.*, pages 50–67, October 1995.

[40] J. G. Proakis. *Digital Communications.* New York, NY: McGraw-Hill, 3rd edition, 1995.

[41] J.G. Proakis. Adaptive equalization for tdma digital mobile radio. *IEEE Trans. Vehic. Technol.*, pages 333–341, May 1991.

[42] S.U. Qureshi. Adaptive equalization. *IEEE Proc.*, pages 1349–1387, September 1985.

[43] S. Ramo, J. R. Whinnery, and T. Van Duzer. *Fields and waves in communication electronics.* New York, NY: John Wiley & Sons, 3rd edition, 1994.

[44] T. S. Rappaport. *Wireless Communications: Principles and Practice.* Upper Saddle River, NJ: PTR Prentice Hall, 1996.

[45] P. Robertson and T. Worz. Bandwidth-efficient turbo trellis-coded modulation using punctured component codes. *IEEE J. Select. Areas Commun.*, pages 206–218, February 1998.

[46] A. A. M. Saleh and R. A. Valenzuela. A statistical model for indoor multipath propagation. *IEEE J. Select. Areas Commun.*, pages 128–137, February 1987.

[47] K. Schaubach, N. J. Davis IV, and T. S. Rappaport. A ray tracing method for predicting path loss and delay spread in microcellular environments. In *Proc. Vehicular Technology Conference (VTC'92)*, pages 932–935, May 1992.

[48] S. Schiesel. Paging allies focus strategy on the Internet. *New York Times*, April 19, 1999.

[49] C. E. Shannon and W. Weaver. *A Mathematical Theory of Communication.* Urbana, IL: Univ. Illinois Press, 1949, 1949.

[50] D. G. Steer. Wireless operation in the unlicensed PCS band. In *Proceedings of 1994 3rd IEEE International Conference on Universal Personal Communications*, pages 586–590, September 1994.

[51] G. Stuber. *Principles of mobile communication.* Boston, MA: Kluwer, 1996.

[52] C.-K. Toh. Associativity-based routing for ad-hoc mobile networks. *Wireless Personal Communications*, pages 103–139, March 1997.

[53] G.L. Turin. Introduction to spread spectrum antimultipath techniques and their application to urban digital radio. *IEEE Proc.*, pages 328–353, March 1980.

[54] G. Ungerboeck. Channel coding with multi-level/phase signals. *IEEE Trans. Inform. Theory*, pages 55–67, January 1982.

[55] S. Verdú. *Multiuser Detection.* Cambridge, U.K.: Cambridge University Press, 1998.

[56] A.J. Viterbi. *CDMA Principles of Spread Spectrum Communications.* New York, NY: Addison-Wesley, 1995.

[57] S. G. Wilson. *Digital Modulation and Coding.* Upper Saddle River, NJ: Prentice-Hall, 1996.

[58] J.H. Winters. Smart antennas for wireless systems. *IEEE Pers. Commun. Mag.*, pages 23–27, February 1998.

[59] R. E. Ziemer and W. H. Tranter. *Principles of Communications - Systems, Modulation, and Noise.* New York, NY: John Wiley & Sons, 4th edition, 1995.