



# Overview of Wireless Communications

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 systems have experienced exponential growth over the last decade and there are currently about two billion users worldwide. Indeed, cellular phones have become a critical business tool and part of everyday life in most developed countries, and they are rapidly supplanting antiquated wireline systems in many developing countries. In addition, wireless local area networks currently supplement or replace wired networks in many homes, businesses, and campuses. Many new applications – including wireless sensor networks, automated highways and factories, smart homes and appliances, and remote telemedicine – are emerging from research ideas to concrete systems. The explosive growth of wireless systems coupled with the proliferation of laptop and palmtop computers suggests a bright future for wireless networks, both as stand-alone systems and as part of the larger networking infrastructure. However, many technical challenges remain in designing robust wireless networks that deliver the performance necessary to support emerging applications. In this introductory chapter we will briefly review the history of wireless networks from the smoke signals of the pre-industrial age to the cellular, satellite, and other wireless networks of today. We then discuss the wireless vision in more detail, including the technical challenges that must still be overcome. We describe current wireless systems along with emerging systems and standards. The gap between current and emerging systems and the vision for future wireless applications indicates that much work remains to be done to make this vision a reality.

## 1.1 History of Wireless Communications

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 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 often characterized by bursty transmissions: the radio is idle except when it transmits a packet, although it may transmit packets continuously. The first network based on packet radio, 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 this combination of packet data and broadcast radio. Throughout the 1970s and early 1980s the Defense Advanced Research Projects Agency (DARPA) invested significant resources to develop networks using packet radios for tactical communications in the battlefield. The nodes in these ad hoc wireless networks had the ability to self-configure (or reconfigure) into a network without the aid of any established infrastructure. DARPA's investment in ad hoc networks peaked in the mid 1980s, but the resulting systems fell far short of expectations in terms of speed and performance. These networks continue to be developed for military use. Packet radio networks also found commercial application in supporting wide area wireless data services. These services, first introduced in the early 1990s, enabled wireless data access (including email, file transfer, and Web browsing) at fairly low speeds, on the order of 20 kbps. No strong market for these wide area wireless data services ever really materialized, due mainly to their low data rates, high cost, and lack of "killer applications". These services mostly disappeared in the 1990s, supplanted by the wireless data capabilities of cellular telephones and wireless local area networks (WLANs).

The introduction of wired Ethernet technology in the 1970s 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 attractive to wireless LAN vendors because they did not need to obtain an FCC license to operate in this band. However, the wireless LAN systems were not allowed to 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 wireless LANs 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

## 1.1 HISTORY OF WIRELESS COMMUNICATIONS

3

(the first wireless LAN access points listed for \$1400 as compared to a few hundred dollars for a wired Ethernet card) – resulted in weak sales. 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 family of IEEE 802.11 standards, have better performance, although the data rates are still relatively low (maximum collective data rates of tens of Mbps) and the coverage area is still small (around 100 m). Wired Ethernets today offer data rates of 1 Gbps, and the performance gap between wired and wireless LANs is likely to increase over time without additional spectrum allocation. Despite their lower data rates, wireless LANs are becoming the preferred Internet access method in many homes, offices, and campus environments owing to their convenience and freedom from wires. However, most wireless LANs support applications, such as email and Web browsing, that are not bandwidth intensive. The challenge for future wireless LANs will be to support many users simultaneously with bandwidth-intensive and delay-constrained applications such as video. Range extension is also a critical goal for future wireless LAN systems.

By far the most successful application of wireless networking has been the cellular telephone system. The roots of this system 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 support only 543 users.

A solution to this capacity problem emerged during the 1950s and 1960s as researchers at AT&T Bell Laboratories developed the cellular concept [1]. Cellular systems exploit the fact that the power of a transmitted signal falls off with distance. Thus, two users can operate on the same frequency at spatially separate locations with minimal interference between them. This allows efficient use of cellular spectrum, so that a large number of users can be accommodated. The evolution of cellular systems from initial concept to implementation was glacial. In 1947, AT&T requested spectrum for cellular service from the FCC. The design was mostly completed by the end of the 1960s; but the first field test was not until 1978, and the FCC granted service authorization in 1982 – by which time much of the original technology was out of date. The first analog cellular system, deployed in Chicago in 1983, was already saturated by 1984, when the FCC increased the cellular spectral allocation from 40 MHz to 50 MHz. The explosive growth of the cellular industry took almost everyone by surprise. In fact, a marketing study commissioned by AT&T before the first system rollout predicted that demand for cellular phones would be limited to doctors and the very rich. AT&T basically abandoned the cellular business in the 1980s to focus on fiber optic networks, eventually returning to the business after its potential became apparent. Throughout the late 1980s – as more and more cities saturated with demand for cellular service – the development of digital cellular technology for increased capacity and better performance became essential.

The second generation of cellular systems, first deployed in the early 1990s, was based on digital communications. The shift from analog to digital was driven by its higher capacity and the improved cost, speed, and power efficiency of digital hardware. Although second-generation cellular systems initially provided mainly voice services, these systems

gradually evolved to support data services such as email, Internet access, and short messaging. Unfortunately, the great market potential for cellular phones led to a proliferation of second-generation cellular standards: three different standards in the United States alone, other standards in Europe and Japan, and all incompatible. The fact that different cities have different incompatible standards makes roaming throughout the United States and the world with only one cellular phone standard impossible. Moreover, some countries have initiated service for third-generation systems, for which there are also multiple incompatible standards. As a result of this proliferation of standards, many cellular phones today are multimode: they incorporate multiple digital standards to facilitate nationwide and worldwide roaming and possibly the first-generation analog standard as well, since only this standard provides universal coverage throughout the United States.

Satellite systems are typically characterized by the height of the satellite orbit: low-earth orbit (LEOs at roughly 2000 km altitude), medium-earth orbit (MEOs, 9000 km), or geosynchronous orbit (GEOs, 40,000 km). The geosynchronous orbits are seen as stationary from the earth, whereas satellites with other orbits have their coverage area change over time. 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 owing to the difficulty of lifting a satellite into such a high orbit. The first GEO satellite was launched by Hughes and NASA in 1963; GEOs then dominated both commercial and government satellite systems for several decades.

Geosynchronous satellites have large coverage areas, so fewer satellites (and dollars) are necessary to provide wide area or global coverage. However, 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. These disadvantages caused a shift in the 1990s toward lower-orbit satellites [2; 3]. The goal was to provide voice and data service competitive with cellular systems. However, the satellite mobile terminals were much bigger, consumed much more power, and cost much more than contemporary cellular phones, which limited their appeal. The most compelling feature of these systems is their ubiquitous worldwide coverage, especially in remote areas or third-world countries with no landline or cellular system infrastructure. Unfortunately, such places do not typically have large demand or the resources to pay for satellite service either. As cellular systems became more widespread, they took away most revenue that LEO systems might have generated in populated areas. With no real market left, most LEO satellite systems went out of business.

A natural area for satellite systems is broadcast entertainment. Direct broadcast satellites operate in the 12-GHz frequency band. These systems offer hundreds of TV channels and are major competitors to cable. Satellite-delivered digital radio has also become popular. These systems, operating in both Europe and the United States, offer digital audio broadcasts at near-CD quality.

## 1.2 Wireless Vision

The vision of wireless communications supporting information exchange between people or devices is the communications frontier of the next few decades, and much of it already

## 1.2 WIRELESS VISION

5

exists in some form. This vision will allow multimedia communication from anywhere in the world using a small handheld device or laptop. Wireless networks will connect palm-top, laptop, and desktop computers anywhere within an office building or campus, as well as from the corner cafe. In the home these networks will enable a new class of intelligent electronic devices that can interact with each other and with the Internet in addition to providing connectivity between computers, phones, and security/monitoring systems. Such “smart” homes can also help the elderly and disabled with assisted living, patient monitoring, and emergency response. Wireless entertainment will permeate the home and any place that people congregate. 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 Caribbean. Wireless video will enable remote classrooms, remote training facilities, and remote hospitals anywhere in the world. Wireless sensors have an enormous range of both commercial and military applications. Commercial applications include monitoring of fire hazards, toxic waste sites, stress and strain in buildings and bridges, carbon dioxide movement, and the spread of chemicals and gasses at a disaster site. These wireless sensors self-configure into a network to process and interpret sensor measurements and then convey this information to a centralized control location. Military applications include identification and tracking of enemy targets, detection of chemical and biological attacks, support of unmanned robotic vehicles, and counterterrorism. Finally, wireless networks enable distributed control systems with remote devices, sensors, and actuators linked together via wireless communication channels. Such systems in turn enable automated highways, mobile robots, and easily reconfigurable industrial automation.

The various applications described here are all components of the wireless vision. So then what, exactly, is wireless communications? There are many ways to segment this complex topic into different applications, systems, or coverage regions [4]. Wireless applications include voice, Internet access, Web browsing, paging and short messaging, subscriber information services, file transfer, video teleconferencing, entertainment, sensing, and distributed control. Systems include cellular telephone systems, wireless LANs, wide area wireless data systems, satellite systems, and ad hoc wireless networks. 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 about 100 ms or else it becomes noticeable to the end user.<sup>1</sup> On the other hand, data systems typically require much higher data rates (1–100 Mbps) and very small BERs (a BER of  $10^{-8}$  or less, 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,

<sup>1</sup> Wired telephones have a delay constraint of  $\sim 30$  ms. Cellular phones relax this constraint to  $\sim 100$  ms, and voice over the Internet relaxes the constraint even further.

while paging and short messaging have very low data-rate requirements and no hard delay constraints. These diverse requirements for different applications make it difficult to build one wireless system that can efficiently satisfy all these requirements simultaneously. Wired networks typically satisfy the diverse requirements of different applications using a single protocol, which means that the most stringent requirements for all applications must be met simultaneously. This may be possible on some wired networks – with data rates on the order of Gbps and BERs on the order of  $10^{-12}$  – but it is not possible on wireless networks, which have much lower data rates and higher BERs. For these reasons, at least in the near future, wireless systems will continue to be fragmented, with different protocols tailored to support the requirements of different applications.

The exponential growth of cellular telephone use and wireless Internet access has led to great optimism about wireless technology in general. Obviously not all wireless applications will flourish. While many wireless systems and companies have enjoyed spectacular success, there have also been many failures along the way, including first-generation wireless LANs, the Iridium satellite system, wide area data services such as Metricom, and fixed wireless access (wireless “cable”) to the home. Indeed, it is impossible to predict what wireless failures and triumphs lie on the horizon. Moreover, there must be sufficient flexibility and creativity among both engineers and regulators to allow for accidental successes. It is clear, however, that the current and emerging wireless systems of today – coupled with the vision of applications that wireless can enable – ensure a bright future for wireless technology.

### 1.3 Technical Issues

Many technical challenges must be addressed to enable the wireless applications of the future. These challenges extend across all aspects of the system design. As wireless terminals add more features, these small devices must incorporate multiple modes of operation in order to support the different applications and media. Computers process voice, image, text, and video data, but breakthroughs in circuit design are required to implement the same multimode operation in a cheap, lightweight, handheld device. Consumers don’t want large batteries that frequently need recharging, so transmission and signal processing at the portable terminal must consume minimal power. The signal processing required to support multimedia applications and networking functions can be power intensive. Thus, wireless infrastructure-based networks, such as wireless LANs and cellular systems, place as much of the processing burden as possible on fixed sites with large power resources. The associated bottlenecks and single points of failure are clearly undesirable for the overall system. Ad hoc wireless networks without infrastructure are highly appealing for many applications because of their flexibility and robustness. For these networks, all processing and control must be performed by the network nodes in a distributed fashion, making energy efficiency challenging to achieve. Energy is a particularly critical resource in networks where nodes cannot recharge their batteries – for example, in sensing applications. Network design to meet application requirements under such hard energy constraints remains a big technological hurdle. The finite bandwidth and random variations of wireless channels also require robust applications that degrade gracefully as network performance degrades.

### 1.3 TECHNICAL ISSUES

7

Design of wireless networks differs fundamentally from wired network design owing to the nature of the wireless channel. This 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. 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. Spectrum can also be very expensive: in many countries spectral licenses are often auctioned to the highest bidder. In the United States, companies spent over \$9 billion for second-generation cellular licenses, and the auctions in Europe for third-generation cellular spectrum garnered around \$100 billion (American). The spectrum obtained through these auctions must be used extremely efficiently to receive a reasonable return on the 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. However, path loss at these higher frequencies is larger with omnidirectional antennas, thereby limiting range.

As a signal propagates through a wireless channel, it experiences random fluctuations in time if the transmitter, receiver, or surrounding objects are moving because of changing reflections and attenuation. Hence 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 by anyone with an RF antenna. The analog cellular systems have no security, and one can easily listen in on conversations by scanning the analog cellular frequency band. All digital cellular systems implement some level of encryption. However, with enough knowledge, time, and determination, most of these encryption methods can be cracked; indeed, several have been compromised. 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 among billions of globally distributed mobile terminals. It must then route a call to that user as it moves at speeds of up to 100 km/hr. The finite resources of the network must be allocated in a fair and efficient manner relative to changing user demands and locations. Moreover, there currently exists a tremendous infrastructure of wired networks: the telephone system, the Internet, and fiber optic cables – which could 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 rates and reliability. Interfacing between wireless and wired networks with vastly different performance capabilities is a difficult problem.

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 a layered approach, whereby protocols associated with different layers of the system operation are designed in isolation, with baseline mechanisms to interface between layers. The layers in

a wireless system include: the link or physical layer, which handles bit transmissions over the communications medium; the access layer, which handles shared access to the communications medium; the network and transport layers, which route data across the network and ensure end-to-end connectivity and data delivery; and the application layer, which dictates the end-to-end data rates and delay constraints associated with the application. While a layering methodology reduces complexity and facilitates modularity and standardization, it also leads to inefficiency and performance loss due to the lack of a global design optimization. The large capacity and good reliability of wired networks make these inefficiencies relatively benign for many wired network applications, although they do preclude good performance of delay-constrained applications such as voice and video. 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 owing to the nature of radio propagation and broadcasting. The dynamic nature and poor performance of the underlying wireless communication channel indicates that high-performance networks must be optimized for this channel and must be robust and adaptive to its variations, as well as to network dynamics. Thus, these networks require integrated and adaptive protocols at all layers, from the link layer to the application layer. This cross-layer protocol design requires interdisciplinary expertise in communications, signal processing, and network theory and design.

In the next section we give an overview of the wireless systems in operation today. It will be clear from this overview that the wireless vision remains a distant goal, with many technical challenges to overcome. These challenges will be examined in detail throughout the book.

## 1.4 Current Wireless Systems

This section provides a brief overview of current wireless systems in operation today. The design details of these system are constantly evolving, with new systems emerging and old ones going by the wayside. Thus, we will focus mainly on the high-level design aspects of the most common systems. More details on wireless system standards can be found in [5; 6; 7]. A summary of the main wireless system standards is given in Appendix D.

### 1.4.1 Cellular Telephone Systems

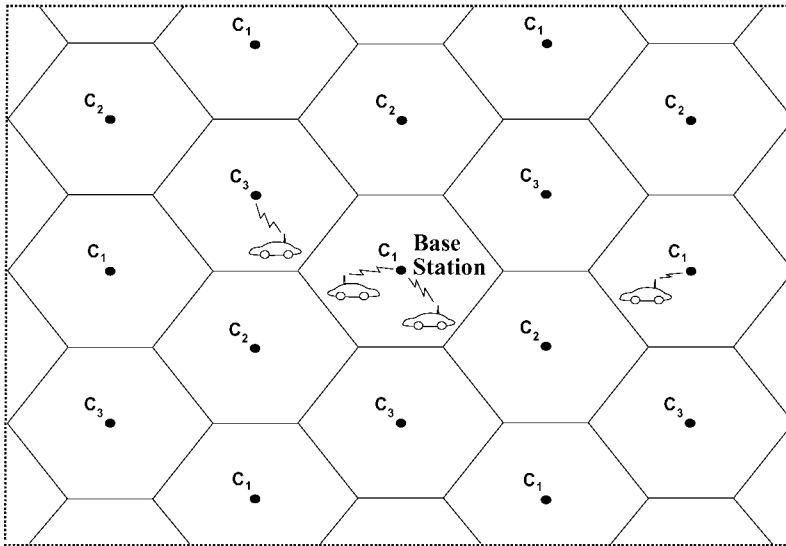
Cellular telephone systems are extremely popular and lucrative worldwide: these are the systems that ignited the wireless revolution. Cellular systems provide two-way voice and data communication with regional, national, or international coverage. Cellular systems were initially designed for mobile terminals 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 the fact that signal power falls off with distance 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 1.1, where  $C_i$  denotes the channel set used in a particular cell. Operation within a cell is controlled by a centralized



## 1.4 CURRENT WIRELESS SYSTEMS

9

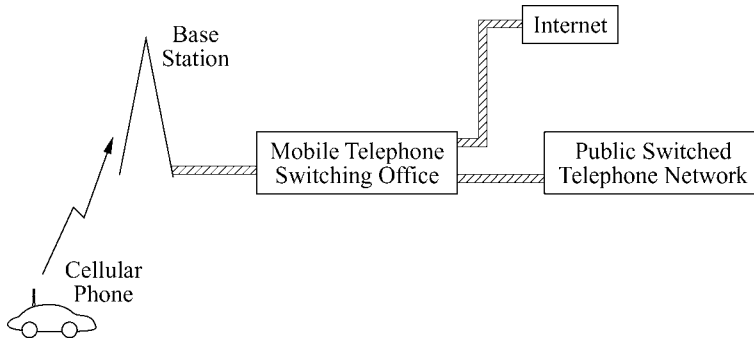


**Figure 1.1:** Cellular systems.

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 so that frequencies are reused as often as possible, thereby maximizing spectral efficiency. However, as the reuse distance decreases, intercell interference increases owing 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 the characteristics of wireless signal propagation. In order to determine the best reuse distance and base station placement, an accurate characterization of signal propagation within the cells is needed.

Initial cellular system designs were mainly driven by the high cost of base stations, approximately \$1 million each. 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*. Signal power radiated uniformly in all directions, so a mobile moving in a circle around the base station would have approximately constant received power unless the signal were blocked by an attenuating object. 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 systems in urban areas now mostly use smaller cells with base stations close to street level that are transmitting at much lower power. These smaller cells are called microcells or picocells, depending on their size. This evolution to smaller cells occurred for two reasons: 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



**Figure 1.2:** Current cellular network architecture.

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. 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, so handoffs must be processed more quickly. In addition, location management becomes more complicated, since there are more cells within a given area 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 generally 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 [8].

All base stations in a given geographical area are connected via a high-speed communications link to a mobile telephone switching office (MTSO), as shown in Figure 1.2. 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. The MTSO can route voice calls through the public switched telephone network (PSTN) or provide Internet access. 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. A call will also be dropped if the signal strength between a mobile and its base station falls below the minimum threshold needed for communication as a result of random signal variations.