CHAPTER 10

CELLULAR WIRELESS NETWORKS

Of all the tremendous advances in data communications and telecommunications, perhaps the most revolutionary is the development of cellular networks. Cellular technology is the foundation of mobile wireless communications and supports users in locations that are not easily served by wired networks. Cellular technology is the underlying technology for mobile telephones, personal communications systems, wireless Internet and wireless Web applications, and much more.

We begin this chapter with a look at the basic principles used in all cellular networks. Then we look at specific cellular technologies and standards, which are conveniently grouped into three generations. The first-generation is analog based and, while still widely used, is passing from the scene. The dominant technology today is the digital second-generation systems. Finally, third-generation high-speed digital systems have begun to emerge.

10.1 PRINCIPLES OF CELLULAR NETWORKS

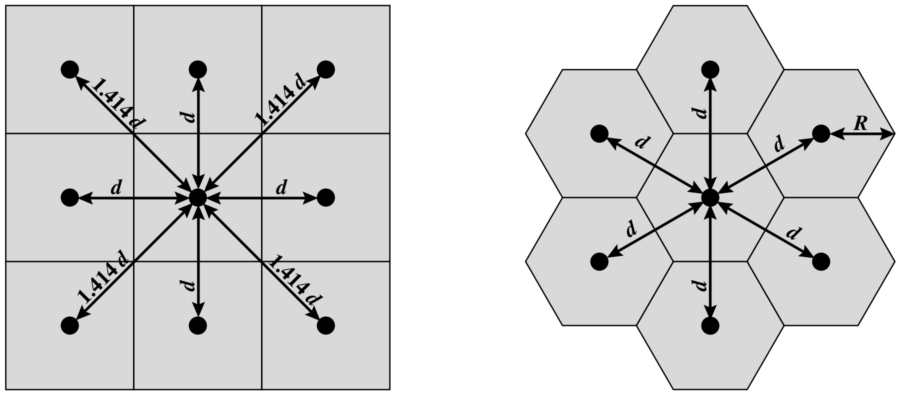
Cellular radio is a technique that was developed to increase the capacity available for mobile radio telephone service. Prior to the introduction of cellular radio, mobile radio telephone service was only provided by a high-power transmitter/receiver. A typical system would support about 25 channels with an effective radius of about 80 km. The way to increase the capacity of the system is to use lower-power systems with shorter radius and to use numerous transmitters/receivers. We begin this section with a look at the organization of cellular systems and then examine some of the details of their implementation.

Cellular Network Organization

The essence of a cellular network is the use of multiple low-power transmitters, on the order of 100 W or less. Because the range of such a transmitter is small, an area can be divided into cells, each one served by its own antenna. Each cell is allocated a band of frequencies and is served by a **base station**, consisting of transmitter, receiver, and control unit. Adjacent cells are assigned different frequencies to avoid interference or crosstalk. However, cells sufficiently distant from each other can use the same frequency band.

The first design decision to make is the shape of cells to cover an area. A matrix of square cells would be the simplest layout to define (Figure 10.1a). However, this geometry is not ideal. If the width of a square cell is d, then a cell has four neighbors at a distance d and four neighbors at a distance $\sqrt{2}d$. As a mobile user within a cell moves toward the cell's boundaries, it is best if all of the adjacent antennas are equidistant. This simplifies the task of determining when to switch the user to an adjacent antenna and which antenna to choose. A hexagonal pattern provides for equidistant antennas (Figure 10.1b). The radius of a hexagon is defined to be the radius of the circle that circumscribes it (equivalently, the distance from the center to each vertex; also equal to the length of a side of a hexagon). For a cell radius R, the distance between the cell center and each adjacent cell center is $d = \sqrt{3}R$

In practice, a precise hexagonal pattern is not used. Variations from the ideal are due to topographical limitations, local signal propagation conditions, and practical limitation on siting antennas.



(a) Square pattern

(b) Hexagonal pattern

Figure 10.1 Cellular Geometries

With a wireless cellular system, you are limited in how often you can use the same frequency for different communications because the signals, not being constrained, can interfere with one another even if geographically separated. Systems supporting a large number of communications simultaneously need mechanisms to conserve spectrum.

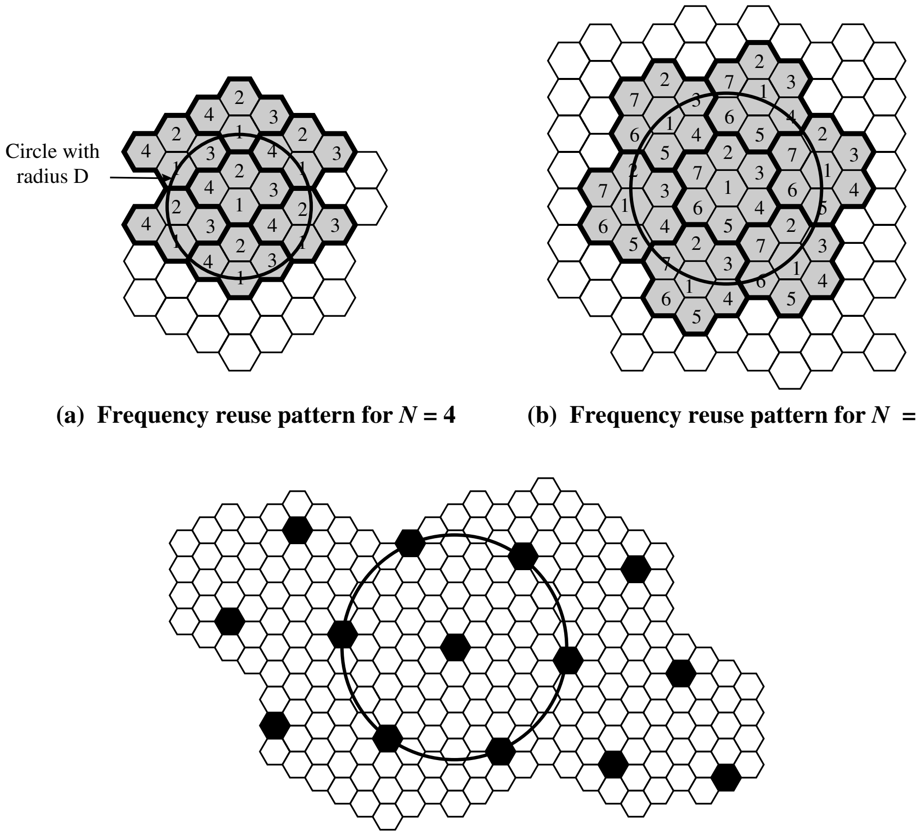
Frequency Reuse

In a cellular system, each cell has a base transceiver. The transmission power is carefully controlled (to the extent that it is possible in the highly variable mobile communication environment) to allow communication within the cell using a given frequency while limiting the power at that frequency that escapes the cell into adjacent ones. The objective is to use the same frequency in other nearby cells, thus allowing the frequency to be used for multiple simultaneous conversations. Generally, 10 to 50 frequencies are assigned to each cell, depending on the traffic expected.

The essential issue, of course, is to determine how many cells must intervene between two cells using the same frequency so that the two cells do not interfere with each other. Various patterns of frequency reuse are possible. Figure 10.2 shows some examples. If the pattern consists of N cells and each cell is assigned the same number of frequencies, each cell can have K/N frequencies, where K is the total number of frequencies allotted to the system. For AMPS, $K = 395$, and $N = 7$ is the smallest pattern that can provide sufficient isolation between two uses of the same frequency. This implies that there can be at most 57 frequencies per cell on average.

In characterizing frequency reuse, the following parameters are commonly used:

D = minimum distance between centers of cells that use the same band of frequencies (called cochannels)
R = radius of a cell
d = distance between centers of adjacent cells ($d = \sqrt{3}R$)
N = number of cells in a repetitious pattern (each cell in the pattern uses a unique band of frequencies), termed the **reuse factor**



(a) Frequency reuse pattern for $N = 4$

(b) Frequency reuse pattern for $N = 7$

(c) Black cells indicate a frequency reuse for $N = 19$

Figure 10.2 Frequency Reuse Patterns

In a hexagonal cell pattern, only the following values of N are possible:

$$N = I^2 + J^2 + (I \times J), \qquad I, J = 0, 1, 2, 3, \ldots$$

Hence, possible values of N are 1, 3, 4, 7, 9, 12, 13, 16, 19, 21, etc. The following relationship holds:

$$\frac{D}{R} = \sqrt{3N}$$

This can also be expressed as $D/d = \sqrt{N}$.

Increasing Capacity

In time, as more customers use the system, traffic may build up so that there are not enough frequencies assigned to a cell to handle its calls. A number of approaches have been used to cope with this situation, including the following:

- **Adding new channels:** Typically, when a system is set up in a region, not all of the channels are used, and growth and expansion can be managed in an orderly fashion by adding new channels.
- **Frequency borrowing:** In the simplest case, frequencies are taken from adjacent cells by congested cells. The frequencies can also be assigned to cells dynamically.
- **Cell splitting:** In practice, the distribution of traffic and topographic features is not uniform, and this presents opportunities of capacity increase. Cells in areas of high usage can be split into smaller cells. Generally, the original cells are about 6.5 to 13 km in size. The smaller cells can themselves be split; however, 1.5-km cells are close to the practical minimum size as a general solution (but see the subsequent discussion of microcells). To use a smaller cell, the power level used must be reduced to keep the signal within the cell. Also, as the mobile units move, they pass from cell to cell, which requires transferring of the call from one base transceiver to another. This process is called a *handoff*. As the cells get smaller, these handoffs become much more frequent. Figure 10.3 indicates schematically how cells can be divided to provide more capacity. A radius reduction by a factor of F reduces the coverage area and increases the required number of base stations by a factor of F^2.
- **Cell sectoring:** With cell sectoring, a cell is divided into a number of wedge-shaped sectors, each with its own set of channels, typically 3 or 6 sectors per cell. Each sector is assigned a separate subset of the cell's channels, and directional antennas at the base station are used to focus on each sector.
- **Microcells:** As cells become smaller, antennas move from the tops of tall buildings or hills, to the tops of small buildings or the sides of large buildings, and finally to lamp posts, where they form microcells. Each decrease in cell size is accompanied by a reduction in the radiated power levels from the base stations and the mobile units. Microcells are useful in city streets in congested areas, along highways, and inside large public buildings.

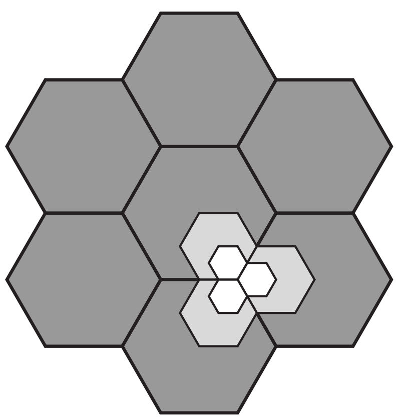


Figure 10.3 Cell Splitting

Table 10.1 Typical Parameters for Macrocells and Microcells [ANDE95]

	Macrocell	**Microcell**
Cell radius	1 to 20 km	0.1 to 1 km
Transmission power	1 to 10 W	0.1 to 1 W
Average delay spread	0.1 to 10 μs	10 to 100 ns
Maximum bit rate	0.3 Mbps	1 Mbps

Table 10.1 suggests typical parameters for traditional cells, called macrocells, and microcells with current technology. The average delay spread refers to multipath delay spread (i.e., the same signal follows different paths and there is a time delay between the earliest and latest arrival of the signal at the receiver). As indicated, the use of smaller cells enables the use of lower power and provides superior propagation conditions.

Example. [HAAS00]. Assume a system of 32 cells with a cell radius of 1.6 km, a total of 32 cells, a total frequency bandwidth that supports 336 traffic channels, and a reuse factor of $N = 7$. If there are 32 total cells, what geographic area is covered, how many channels are there per cell, and what is the total number of concurrent calls that can be handled? Repeat for a cell radius of 0.8 km and 128 cells.

Figure 10.4a shows an approximately square pattern. The area of a hexagon of radius R is $1.5R^2\sqrt{3}$. A hexagon of radius 1.6 km has an area of 6.65 km^2, and the total area covered is $6.65 \times 32 = 213$ km^2. For $N = 7$, the number of channels per cell is $336/7 = 48$, for a total channel capacity of $48 \times 32 = 1536$ channels. For the layout of Figure 10.4b, the area covered is $1.66 \times 128 = 213$ km^2. The number of channels per cell is $336/7 = 48$, for a total channel capacity of $48 \times 128 = 6144$ channels.

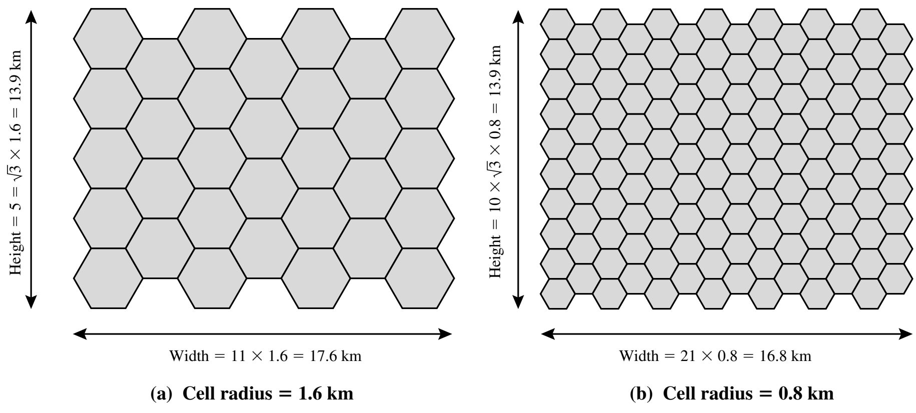


Figure 10.4 Frequency Reuse Example

Operation of Cellular Systems

Figure 10.5 shows the principal elements of a cellular system. In the approximate center of each cell is a base station (BS). The BS includes an antenna, a controller, and a number of transceivers, for communicating on the channels assigned to that cell. The controller is used to handle the call process between the mobile unit and the rest of the network. At any time, a number of mobile user units may be active and moving about within a cell, communicating with the BS. Each BS is connected to a mobile telecommunications switching office (MTSO), with one MTSO serving multiple BSs. Typically, the link between an MTSO and a BS is by a wire line, although a wireless link is also possible. The MTSO connects calls between mobile units. The MTSO is also connected to the public telephone or telecommunications network and can make a connection between a fixed subscriber to the public network and a mobile subscriber to the cellular network. The MTSO assigns the voice channel to each call, performs handoffs (discussed subsequently), and monitors the call for billing information.

The use of a cellular system is fully automated and requires no action on the part of the user other than placing or answering a call. Two types of channels are available between the mobile unit and the base station (BS): control channels and traffic channels. **Control channels** are used to exchange information having to do with setting up and maintaining calls and with establishing a relationship between a mobile unit and the nearest BS. **Traffic channels** carry a voice or data connection between users. Figure 10.6 illustrates the steps in a typical call between two mobile users within an area controlled by a single MTSO:

- **Mobile unit initialization:** When the mobile unit is turned on, it scans and selects the strongest setup control channel used for this system (Figure 10.6a). Cells with different frequency bands repetitively broadcast on different setup channels. The receiver selects the strongest setup channel and monitors that

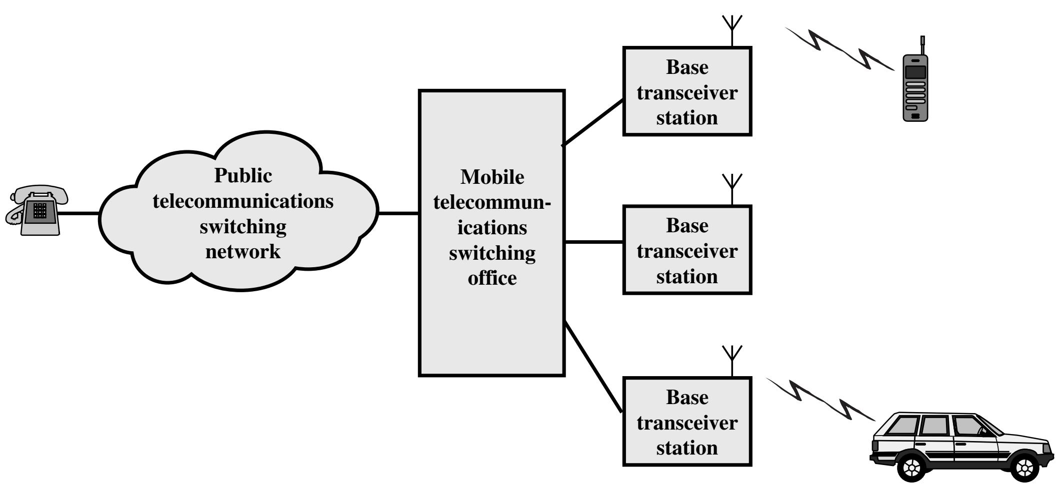


Figure 10.5 Overview of Cellular System

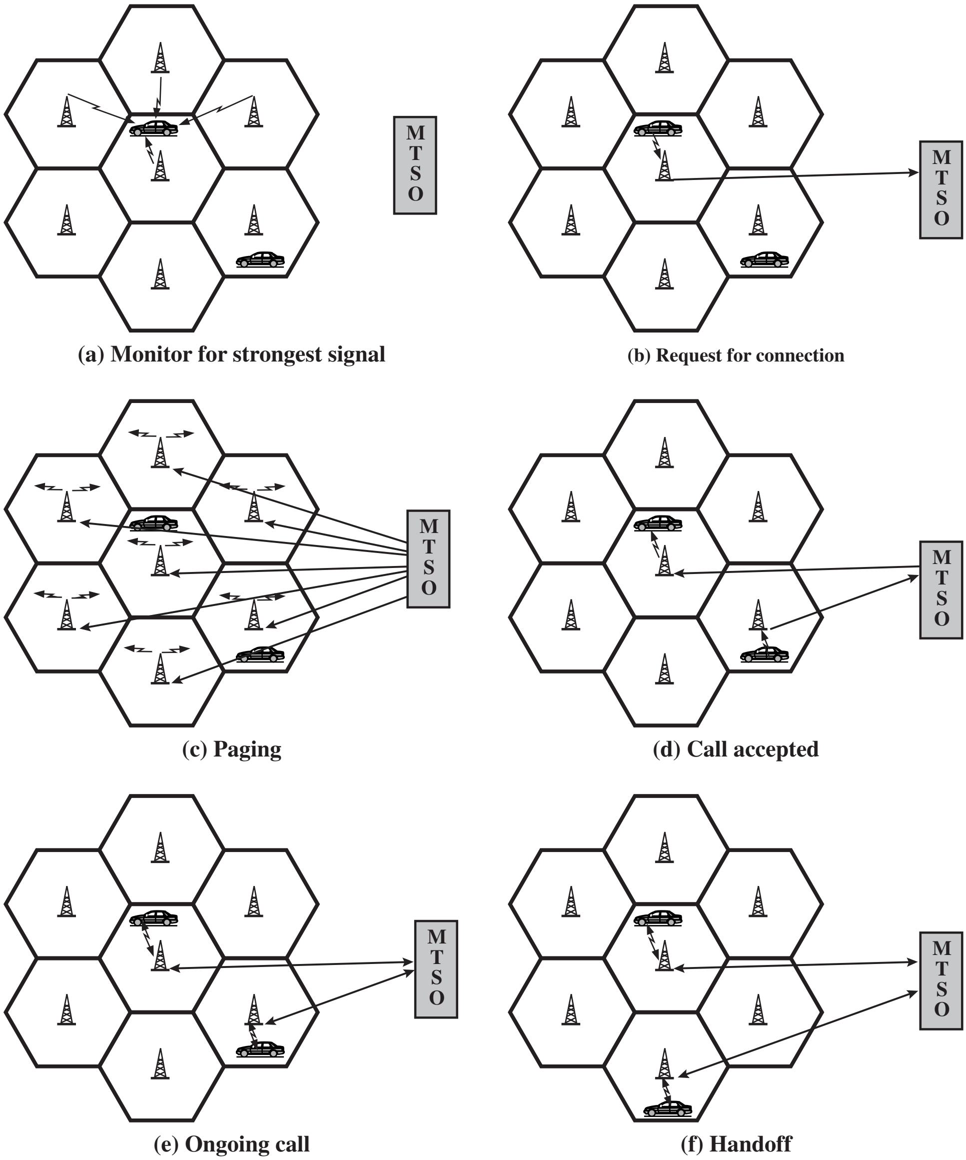


Figure 10.6 Example of Mobile Cellular Call

channel. The effect of this procedure is that the mobile unit has automatically selected the BS antenna of the cell within which it will operate.[1] Then a handshake takes place between the mobile unit and the MTSO controlling this cell, through the BS in this cell. The handshake is used to identify the user and

[1]Usually, but not always, the antenna and therefore the base station selected is the closest one to the mobile unit. However, because of propagation anomalies, this is not always the case.

register its location. As long as the mobile unit is on, this scanning procedure is repeated periodically to account for the motion of the unit. If the unit enters a new cell, then a new BS is selected. In addition, the mobile unit is monitoring for pages, discussed subsequently.

- **Mobile-originated call:** A mobile unit originates a call by sending the number of the called unit on the preselected setup channel (Figure 10.6b). The receiver at the mobile unit first checks that the setup channel is idle by examining information in the forward (from the BS) channel. When an idle is detected, the mobile may transmit on the corresponding reverse (to BS) channel. The BS sends the request to the MTSO.
- **Paging:** The MTSO then attempts to complete the connection to the called unit. The MTSO sends a paging message to certain BSs depending on the called mobile number (Figure 10.6c). Each BS transmits the paging signal on its own assigned setup channel.
- **Call accepted:** The called mobile unit recognizes its number on the setup channel being monitored and responds to that BS, which sends the response to the MTSO. The MTSO sets up a circuit between the calling and called BSs. At the same time, the MTSO selects an available traffic channel within each BS's cell and notifies each BS, which in turn notifies its mobile unit (Figure 10.6d). The two mobile units tune to their respective assigned channels.
- **Ongoing call:** While the connection is maintained, the two mobile units exchange voice or data signals, going through their respective BSs and the MTSO (Figure 10.6e).
- **Handoff:** If a mobile unit moves out of range of one cell and into the range of another during a connection, the traffic channel has to change to one assigned to the BS in the new cell (Figure 10.6f). The system makes this change without either interrupting the call or alerting the user.

Other functions performed by the system but not illustrated in Figure 10.6 include the following:

- **Call blocking:** During the mobile-initiated call stage, if all the traffic channels assigned to the nearest BS are busy, then the mobile unit makes a preconfigured number of repeated attempts. After a certain number of failed tries, a busy tone is returned to the user.
- **Call termination:** When one of the two users hangs up, the MTSO is informed and the traffic channels at the two BSs are released.
- **Call drop:** During a connection, because of interference or weak signal spots in certain areas, if the BS cannot maintain the minimum required signal strength for a certain period of time, the traffic channel to the user is dropped and the MTSO is informed.
- **Calls to/from fixed and remote mobile subscriber:** The MTSO connects to the public switched telephone network. Thus, the MTSO can set up a connection between a mobile user in its area and a fixed subscriber via the telephone network. Further, the MTSO can connect to a remote MTSO via the telephone

network or via dedicated lines and set up a connection between a mobile user in its area and a remote mobile user.

Mobile Radio Propagation Effects

Mobile radio communication introduces complexities not found in wire communication or in fixed wireless communication. Two general areas of concern are signal strength and signal propagation effects.

- **Signal strength:** The strength of the signal between the base station and the mobile unit must be strong enough to maintain signal quality at the receiver but no so strong as to create too much cochannel interference with channels in another cell using the same frequency band. Several complicating factors exist. Human-made noise varies considerably, resulting in a variable noise level. For example, automobile ignition noise in the cellular frequency range is greater in the city than in a suburban area. Other signal sources vary from place to place. The signal strength varies as a function of distance from the BS to a point within its cell. Moreover, the signal strength varies dynamically as the mobile unit moves.
- **Fading:** Even if signal strength is within an effective range, signal propagation effects may disrupt the signal and cause errors. Section 5.4 discussed fading and various countermeasures.

In designing a cellular layout, the communications engineer must take account of these various propagation effects, the desired maximum transmit power level at the base station and the mobile units, the typical height of the mobile unit antenna, and the available height of the BS antenna. These factors will determine the size of the individual cell. Unfortunately, as just described, the propagation effects are dynamic and difficult to predict. The best that can be done is to come up with a model based on empirical data and to apply that model to a given environment to develop guidelines for cell size. One of the most widely used models was developed by Okumura et al. [OKUM68] and subsequently refined by Hata [HATA80]. The original was a detailed analysis of the Tokyo area and produced path loss information for an urban environment. Hata's model is an empirical formulation that takes into account a variety of environments and conditions. For an urban environment, predicted path loss is

$$L_{dB} = 69.55 + 26.16 \log f_c - 13.82 \log h_t - A(h_r) + (44.9 - 6.55 \log h_t) \log d \quad (10.1)$$

where

f_c = carrier frequency in MHz from 150 to 1500 MHz
h_t = height of transmitting antenna (base station) in m, from 30 to 300 m
h_r = height of receiving antenna (mobile station) in m, from 1 to 10 m
d = propagation distance between antennas in km, from 1 to 20 km
$A(h_r)$ = correction factor for mobile antenna height

For a small or medium-sized city, the correction factor is given by

$$A(h_r) = (1.1 \log f_c - 0.7)\, h_r - (1.56 \log f_c - 0.8) \text{ dB}$$

and for a large city it is given by

$$A(h_r) = 8.29[\log(1.54\, h_r)]^2 - 1.1 \text{ dB} \quad \text{for } f_c \le 300 \text{ MHz}$$
$$A(h_r) = 3.2[\log(11.75\, h_r)]^2 - 4.97 \text{ dB} \quad \text{for } f_c \ge 300 \text{ MHz}$$

To estimate the path loss in a suburban area, the formula for urban path loss in Equation (10.1) is modified as

$$L_{\text{dB}}(\text{suburban}) = L_{\text{dB}}(\text{urban}) - 2[\log(f_c/28)]^2 - 5.4$$

For the path loss in open areas, the formula is modified as

$$L_{\text{dB}}(\text{open}) = L_{\text{dB}}(\text{urban}) - 4.78(\log f_c)^2 - 18.733(\log f_c) - 40.98$$

The Okumura/Hata model is considered to be among the best in terms of accuracy in path loss prediction and provides a practical means of estimating path loss in a wide variety of situations [FREE97, RAPP97].

Example. [FREE97]. Let f_c = 900 MHz, h_t = 40 m, h_r = 5 m, and d = 10 km. Estimate the path loss for a medium-size city.

$$\begin{aligned} A(h_r) &= (1.1 \log 900 - 0.7)\, 5 - (1.56 \log 900 - 0.8) \text{ dB} \\ &= 12.75 - 3.8 = 8.95 \text{ dB} \end{aligned}$$

$$\begin{aligned} L_{\text{db}} &= 69.55 + 26.16 \log 900 - 13.82 \log 40 - 8.95 + (44.9 - 6.55 \log 40)\log 10 \\ &= 69.55 + 77.28 - 22.14 - 8.95 + 34.4 = 150.14 \text{ dB} \end{aligned}$$

Handoff

Handoff[2] is the procedure changing the assignment of a mobile unit from one BS to another as the mobile moves from one cell to another. Handoff is handled in different ways in different systems and involves a number of factors. Here we give a brief overview.

Handoff may be network initiated, in which the decision is made solely by the network measurements of received signals from the mobile unit. Alternatively, mobile-assisted handoff schemes enable the mobile unit to participate in the handoff decision by providing feedback to the network concerning signals received at the mobile unit. In either case, a number of different performance metrics may be used to make the decision. [HAAS00] lists the following:

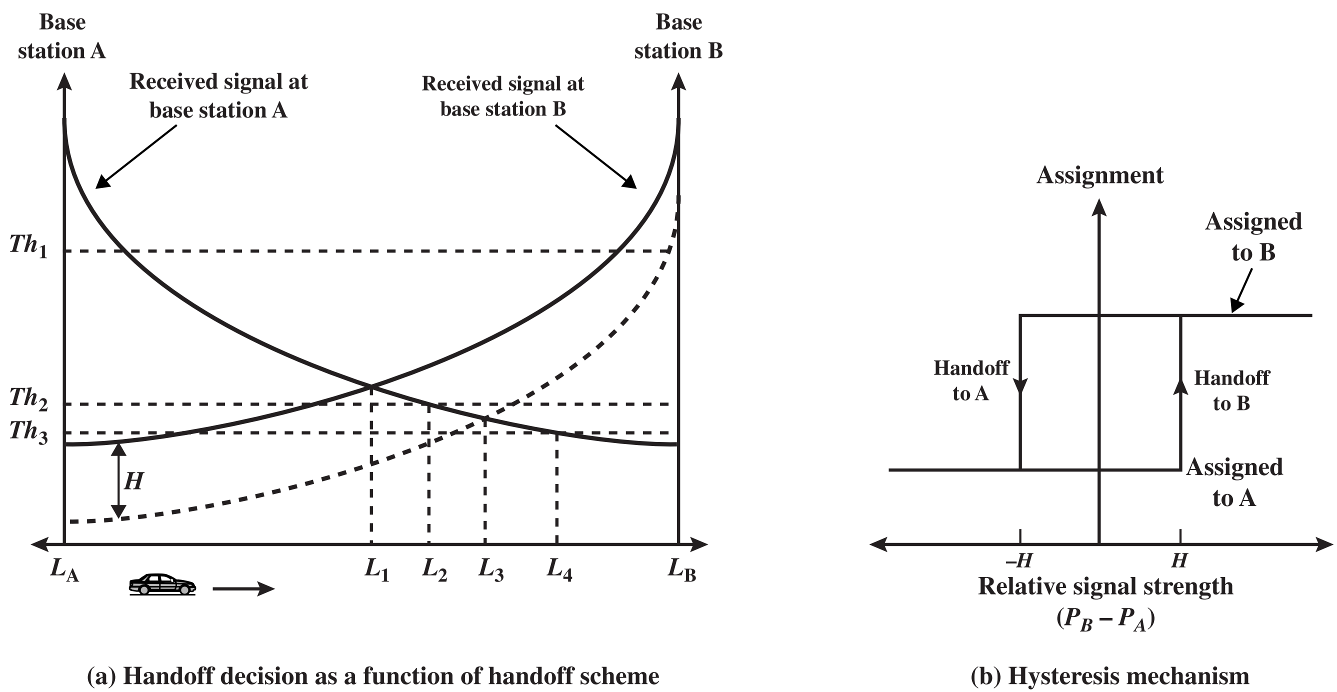
- **Cell blocking probability:** The probability of a new call being blocked, due to heavy load on the BS traffic capacity. In this case, the mobile unit is handed off to a neighboring cell based not on signal quality but on traffic capacity.

[2]The term *handoff* is used in U.S. cellular standards documents. ITU documents use the term *handover*, and both terms appear in the technical literature. The meanings are the same.

- **Call dropping probability:** The probability that, due to a handoff, a call is terminated.
- **Call completion probability:** The probability that an admitted call is not dropped before it terminates.
- **Probability of unsuccessful handoff:** The probability that a handoff is executed while the reception conditions are inadequate.
- **Handoff blocking probability:** The probability that a handoff cannot be successfully completed.
- **Handoff probability:** The probability that a handoff occurs before call termination.
- **Rate of handoff:** The number of handoffs per unit time.
- **Interruption duration:** The duration of time during a handoff in which a mobile is not connected to either base station.
- **Handoff delay:** The distance the mobile moves from the point at which the handoff should occur to the point at which it does occur.

The principal parameter used to make the handoff decision is measured signal strength from the mobile at the BS. Typically, the BS averages the signal over a moving window of time to remove the rapid fluctuations due to multipath effects. Figure 10.7a, based on one in [POLL96], shows the average received power level at two adjacent base stations as a mobile unit moves from BS A, at L_A, to BS B, at L_B. This figure is useful in explaining various handoff strategies that have been used to determine the instant of handoff:

- **Relative signal strength:** The mobile unit is handed off from BS A to BS B when the signal strength at B first exceeds that at A. If the signal strength at B subsequently falls below that of A, the mobile unit is handed back to A. In Figure 10.7a, handoff occurs at point L_1. At this point, signal strength to BS A is still adequate but is declining. Because signal strength fluctuates due to multipath effects, even with power averaging, this approach can lead to a ping-pong effect in which the unit is repeatedly passed back and forth between two BSs.
- **Relative signal strength with threshold:** Handoff only occurs if (1) the signal at the current BS is sufficiently weak (less than a predefined threshold) and (2) the other signal is the stronger of the two. The intention is that so long as the signal at the current BS is adequate, handoff is unnecessary. If a high threshold is used, such as Th_1, this scheme performs the same as the relative signal strength scheme. With a threshold of Th_2, handoff occurs at L_2. If the threshold is set quite low compared to the crossover signal strength (signal strength at L_1), such as Th_3, the mobile may move far into the new cell (L_4) before handoff. This reduces the quality of the communication link and may result in a dropped call. A threshold should not be used alone because its effectiveness depends on prior knowledge of the crossover signal strength between the current and candidate base stations.
- **Relative signal strength with hysteresis:** Handoff occurs only if the new base station is sufficiently stronger (by a margin H in Figure 10.7a) than the current one. In this case, handoff occurs at L_3. This scheme prevents the ping-pong



(a) Handoff decision as a function of handoff scheme

(b) Hysteresis mechanism

Figure 10.7 Handoff between Two Cells effect, because once handoff occurs, the effect of the margin H is reversed. The term *hysteresis* refers to a phenomenon known as relay hysteresis and can be appreciated with the aid of Figure 10.7b. We can think of the handoff mechanism as having two states. While the mobile is assigned to BS A, the mechanism will generate a handoff when the relative signal strength reaches or exceeds the H. Once the mobile is assigned to B, it remains so until the relative signal strength falls below $-H$, at which point it is handed back to A. The only disadvantage of this scheme is that the first handoff may still be unnecessary if BS A still has sufficient signal strength.

- **Relative signal strength with hysteresis and threshold:** Handoff occurs only if (1) the current signal level drops below a threshold, and (2) the target base station is stronger than the current one by a hysteresis margin H. In our example, handoff occurs at L_3 if the threshold is either Th_1 or Th_2 and at L_4 if the threshold is at Th_3.
- **Prediction techniques:** The handoff decision is based on the expected future value of the received signal strength.

The handoff decision is complicated by the use of power control techniques, which enable the BS dynamically to adjust the power transmitted by the mobile. This topic is discussed next.

Power Control

A number of design issues make it desirable to include a dynamic power control capability in a cellular system:

1. The received power must be sufficiently above the background noise for effective communication, which dictates the required transmitted power. As the mobile unit moves away from the transmitter, the received power declines due to normal attenuation. In addition, the effects of reflection, diffraction, and scattering can cause rapid changes in received power levels over small distances. This is because the power level is the sum from signals coming from a number of different paths and the phases of those paths are random, sometimes adding and sometimes subtracting. As the mobile unit moves, the contributions along various paths change.
2. At the same time, it is desirable to minimize the power in the transmitted signal from the mobile, to reduce cochannel interference (interference with channels on the same frequency in remote cells), alleviate health concerns, and save battery power.
3. In spread spectrum (SS) systems using code division multiple access (CDMA), it is desirable to equalize the received power level from all mobile units at the BS. This is crucial to system performance because all users have the same frequency allocation.

Figure 10.8, based on one in [PICH97], illustrates the two kinds of power control. **Open-loop power control** depends solely on the mobile unit, with no feedback

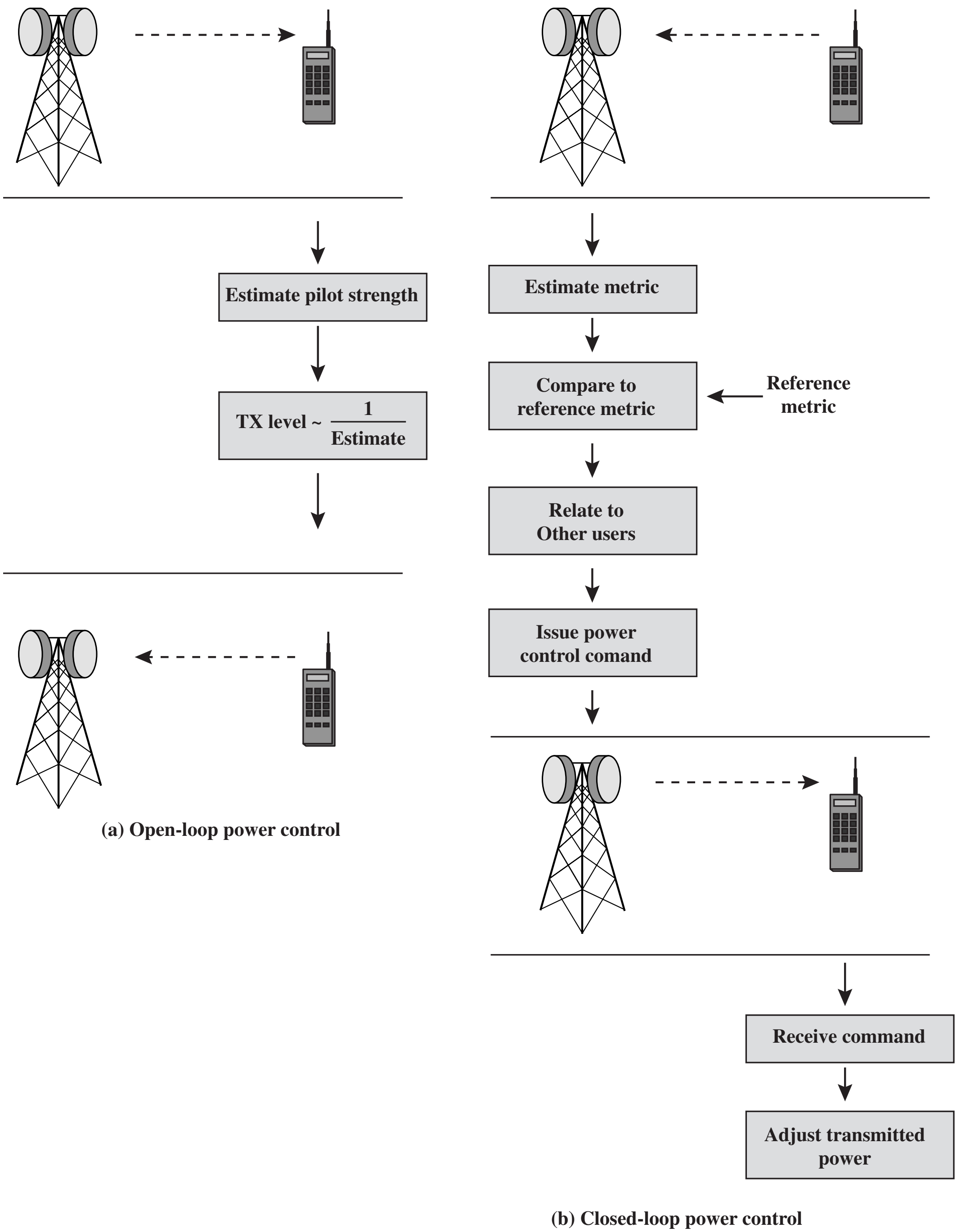


Figure 10.8 Reverse Link Power Control

from the BS, and is used in some SS systems. In SS systems, the BS continuously transmits an unmodulated signal, known as a pilot. The pilot allows a mobile unit to acquire the timing of the forward (BS to mobile) CDMA channel and provides a phase reference for demodulation. It can also be used for power control. The mobile monitors the received power level of the pilot and sets the transmitted power in the reverse (mobile to BS) channel inversely proportional to it. This approach assumes that the forward and reverse link signal strengths are closely correlated, which is generally the case. The open-loop approach is not as accurate as the closed-loop approach. However, the open-loop scheme can react more quickly to rapid fluctuations in signal strength, such as when a mobile emerges from behind a large building. This fast action is required in the reverse link of a CDMA system where the sudden increase in received strength at the BS may suppress all other signals.

Closed-loop power control adjusts signal strength in the reverse (mobile to BS) channel based on some metric of performance in that reverse channel, such as received signal power level, received signal-to-noise ratio, or received bit error rate. The BS makes the power adjustment decision and communicates a power adjustment command to the mobile on a control channel. Closed-loop power control is also used to adjust power in the forward channel. In this case, the mobile provides information about received signal quality to the BS, which then adjusts transmitted power.

Table 10.2 shows the power classes used in the GSM standard, which is a TDMA standard and is discussed in Section 10.3. GSM defines eight classes of base station channels and five classes of mobile stations, according to their power output. Adjustments in both directions are made using closed-loop power control.

Traffic Engineering

For an FDMA system, the capacity of a cell is equal to the number of frequency channels allocated to it. Ideally, the number of available channels in a cell would

Table 10.2 GSM Transmitter Classes

Power Class	Base station power (watts)	Mobile station power (watts)
1	320	20
2	160	8
3	80	5
4	40	2
5	20	0.8
6	10	
7	5	
8	2.5	

equal the total number of subscribers who could be active at any time. In practice, it is not feasible to have the capacity to handle any possible load at all times. Fortunately, not all subscribers place calls at the same time and so it is reasonable to size the network to be able to handle some expected level of load. This is the discipline of traffic engineering.

Traffic engineering concepts were developed in the design of telephone switches and circuit-switching telephone networks, but the concepts equally apply to cellular networks. Consider a cell able to handle N simultaneous users (capacity of N channels) that has L potential subscribers (L mobile units). If $L < N$, the system is referred to as *nonblocking*; all calls can be handled all the time. If $L > N$, the system is *blocking*; a subscriber may attempt a call and find the capacity fully in use and therefore be blocked. For a blocking system, the fundamental performance questions we wish to answer are as follows:

1. What is the degree of blocking; that is, what is the probability that a call request will be blocked? Alternatively, what capacity (N) is needed to achieve a certain upper bound on the probability of blocking?
2. If blocked calls are queued for service, what is the average delay? Alternatively, what capacity is needed to achieve a certain average delay?

In this subsection, we briefly introduce the relevant traffic engineering concepts and give an example of their use. Appendix B examines the subject in more detail.

Two parameters determine the amount of load presented to a system:

λ = the mean rate of calls (connection requests) attempted per unit time
h = the mean holding time per successful call

The basic measure of traffic is the **traffic intensity**, expressed in a dimensionless unit, the erlang:

$$A = \lambda h$$

A can be interpreted in several ways. It is a normalized version of λ: A equals the average number of calls arriving during the average holding period. We can also view the cell as a multiserver queuing system where the number of servers is equal to the channel capacity N. The average service time at a server is h. A basic relationship in a multiserver queue is $\lambda h = \rho N$, where ρ is server utilization, or the fraction of time that a server is busy. Therefore $A = \rho N$ and is a measure of the average number of channels required.

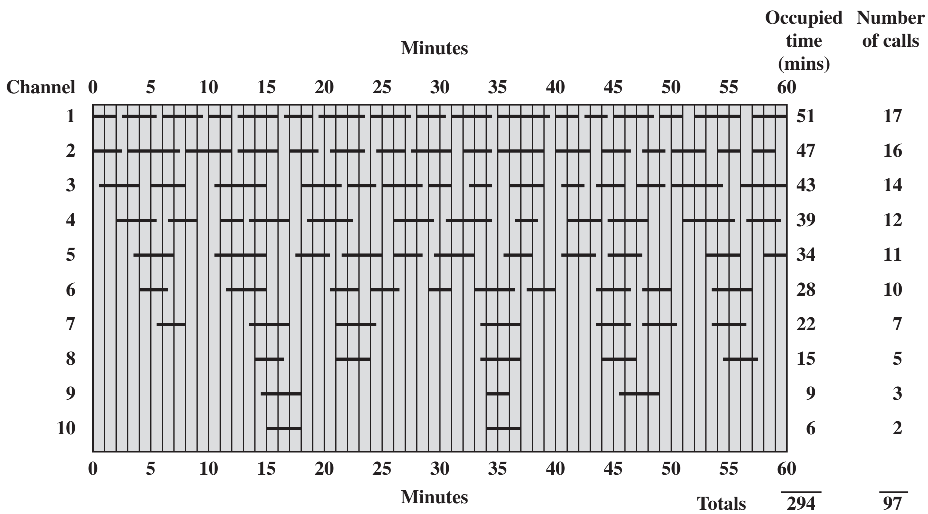
Example. If the calling rate averages 20 calls per minute and the average holding time is 3 minutes, then $A = 60$. We would expect a cell with a capacity of 120 channels to be about half utilized at any given time. A switch of capacity 50 would clearly be inadequate. A capacity of 60 would meet the average demand but, because of fluctuations around the mean rate A, this capacity would at times be inadequate.

Example. To clarify these concepts, consider Figure 10.9, which shows the pattern of activity in a cell of capacity 10 channels over a period of 1 hour. The rate of calls per minute is 97/60. The average holding time per call, in minutes, is 294/97. Thus $A = (97/60) \times (294/97) = 4.9$ erlangs. Another way of viewing the parameter A is that it is the mean number of calls in progress. Thus, on average, 4.9 channels are engaged. The latter interpretation, however, is true only in the nonblocking case. The parameter λ was defined as the rate of calls attempted, not carried traffic.

Typically, a blocking system is sized to deal with some upper limit of traffic intensity. It is generally thought unreasonable to size for the highest surge of traffic anticipated; rather, the common practice is to size the system to meet the average rate encountered during a busy hour. The busy hour is the 60-minute period during the day when the traffic is highest, in the long run. ITU-T recommends taking the average of the busy hour traffic on the 30 busiest days of the year, called the "mean busy-hour traffic," and using that quantity to size the system. The North American practice is to take the average over the 10 busiest days. These are typically measurements of carried rather than offered traffic and can only be used to estimate the true load.

The parameter A, as a measure of busy-hour traffic, serves as input to a traffic model. The model is then used to answer questions such as those posed in the beginning of this subsection. There are two key factors that determine the nature of the model:

- The manner in which blocked calls are handled
- The number of traffic sources



Note: **Horizontal lines indicate occupied periods to the nearest 1/2 minute**

Figure 10.9 Example Distribution of Traffic in a Cell with Capacity 10

Blocked calls may be handled in one of two ways. First, blocked calls can be put in a queue awaiting a free channel; this is referred to as **lost calls delayed** (LCD), although in fact the call is not lost, merely delayed. Second, a blocked call can be rejected and dropped. This in turn leads to two assumptions about the action of the user. If the user hangs up and waits some random time interval before another call attempt, this is known as **lost calls cleared** (LCC). If the user repeatedly attempts calling, it is known as **lost calls held** (LCH). For each of these blocking options, formulas have been developed that characterize the performance of the system. For cellular systems, the LCC model is generally used and is generally the most accurate.

The second key element of a traffic model is whether the number of users is assumed to be finite or infinite. For an infinite source model, there is assumed to be a fixed arrival rate. For the finite source case, the arrival rate will depend on the number of sources already engaged. In particular, if the total pool of users is L, each of which generates calls at an average rate of λ/L, then, when the cell is totally idle, the arrival rate is λ. However, if there are K users occupied at time t, then the instantaneous arrival rate at that time is $\lambda(L - K)/L$. Infinite source models are analytically easier to deal with. The infinite source assumption is reasonable when the number of sources is at least 5 to 10 times the capacity of the system.

Infinite Sources, Lost Calls Cleared

For an infinite source LCC model, the key parameter of interest is the probability of loss, or **grade of service**. Thus a grade of service of 0.01 means that, during a busy hour, the probability that an attempted call is blocked is 0.01. Values in the range 0.01 to 0.001 are generally considered quite good.

The equation of infinite source LCC, known as Erlang B, has the following form:

$$P = \frac{\dfrac{A^N}{N!}}{\displaystyle\sum_{x=0}^{N} \frac{A^x}{x!}}$$

where

A = offered traffic, erlangs
N = number of servers
P = probability of blocking (grade of service)

This equation is easily programmed, and tables of values are readily available. Table 10.3 is an extract from such a table. Given the offered load and number of servers, the grade of service can be calculated or determined from a table. More often, the inverse problem is of interest: determining the amount of traffic that can be handled by a given capacity to produce a given grade of service. Another problem is to determine the capacity required to handle a given amount of traffic at a given grade of service. For both these problems, tables or suitable trial-and-error programs are needed.

Two important points can be deduced from Table 10.3:

Table 10.3 Erlang B Table

	Capacity (erlangs) for grade of service of				
Number of servers (N)	**P = 0.02 (1/50)**	**P = 0.01 1/100)**	**P = 0.005 (1/200)**	**P = 0.002 (1/500)**	**P = 0.001 (1/1000)**
1	0.02	0.01	0.005	0.002	0.001
4	1.09	0.87	0.7	0.53	0.43
5	1.66	1.36	1.13	0.9	0.76
10	5.08	4.46	3.96	3.43	3.09
20	13.19	12.03	11.10	10.07	9.41
24	16.64	15.27	14.21	13.01	12.24
40	31.0	29.0	27.3	25.7	24.5
70	59.13	56.1	53.7	51.0	49.2
100	87.97	84.1	80.9	77.4	75.2

1. A larger-capacity system is more efficient than a smaller-capacity one for a given grade of service.
2. A larger-capacity system is more susceptible to reduction of the grade of service.

Example. To illustrate the first point, consider two cells, each with a capacity of 10 channels. They have a joint capacity of 20 channels and can handle a combined offered traffic intensity of 6.86 for a grade of service of 0.002. However, a single cell of capacity 20 channels will handle 10.07 erlangs at a grade of service of 0.002. To illustrate the second point, consider a cell of 10 channels giving a grade of service of 0.002 for a load of 3.43 erlangs. A 30% increase in traffic reduces the grade of service to 0.01. However, for a cell of capacity 70 channels, only a 10% increase in traffic reduces the grade of service from 0.002 to 0.01.

All of the preceding discussion deals with offered traffic. If sizing is done on the basis of system measurement, all that we are likely to have is carried traffic. A program can readily be developed that accepts carried traffic as input and then performs a seeking algorithm to work backward to offered traffic. The relationship between carried traffic C and offered traffic A is:

$$C = A(1 - P)$$

For small value of P, A is a good approximation of C.

Effect of Handoff

One complication in cellular traffic models not found in other such models is the effect of handoff. This is illustrated in Figure 10.10. The arrival rate of calls at a cell has two components: new calls placed by mobile units in the cell (λ_1), and calls

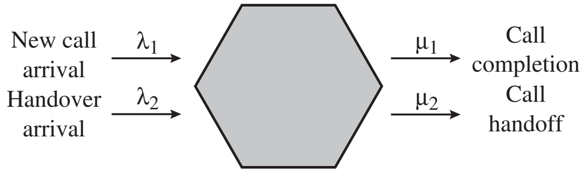


Figure 10.10 Cell Traffic Model

handed off to the cell for mobile units entering the cell while connected (λ_2). The total arrival rate is $\lambda = \lambda_1 + \lambda_2$. Similarly, the completion rate consists of calls being terminated and calls being handed off. The model must be adjusted accordingly to obtain overall arrival rates and holding times.

10.2 FIRST-GENERATION ANALOG

The original cellular telephone networks provided analog traffic channels; these are now referred to as first-generation systems. Since the early 1980s the most common first-generation system in North America has been the **Advanced Mobile Phone Service (AMPS)** developed by AT&T. This approach is also common in South America, Australia, and China. Although gradually being replaced by second-generation systems, AMPS is still in common use. In this section, we provide an overview of AMPS.

Spectral Allocation

In North America, two 25-MHz bands are allocated to AMPS (Table 10.4), one for transmission from the base station to the mobile unit (869–894 MHz), the other for transmission from the mobile to the base station (824–849 MHz). Each of these bands is split in two to encourage competition (i.e., so that in each market two operators can be accommodated). An operator is allocated only 12.5 MHz in each direction for its system. The channels are spaced 30 kHz apart, which allows a total of 416 channels per operator. Twenty-one channels are allocated for control, leaving 395 to carry calls. The control channels are data channels operating at 10 kbps. The conversation channels carry the conversations in analog using frequency modulation. Control information is also sent on the conversation channels in bursts as data. This number of channels is inadequate for most major markets, so some way must be found either to use less bandwidth per conversation or to reuse frequencies. Both approaches have been taken in the various approaches to mobile telephony. For AMPS, frequency reuse is exploited.

Operation

Each AMPS-capable cellular telephone includes a *numeric assignment module* (NAM) in read-only memory. The NAM contains the telephone number of the phone, which is assigned by the service provider, and the serial number of the phone,

Table 10.4 AMPS Parameters

Base station transmission band	869 to 894 MHz
Mobile unit transmission band	824 to 849 MHz
Spacing between forward and reverse channels	45 MHz
Channel bandwidth	30 kHz
Number of full-duplex voice channels	790
Number of full-duplex control channels	42
Mobile unit maximum power	3 watts
Cell size, radius	2 to 20 km
Modulation, voice channel	FM, 12-kHz peak deviation
Modulation, control channel	FSK, 8-kHz peak deviation
Data transmission rate	10 kbps
Error control coding	BCH (48, 36,5) and (40, 28,5)

which is assigned by the manufacturer. When the phone is turned on, it transmits its serial number and phone number to the MTSO (Figure 10.5); the MTSO maintains a database with information about mobile units that have been reported stolen and uses serial number to lock out stolen units. The MTSO uses the phone number for billing purposes. If the phone is used in a remote city, the service is still billed to the user's local service provider.

When a call is placed, the following sequence of events occurs [COUC01]:

1. The subscriber initiates a call by keying in the telephone number of the called party and presses the send key.
2. The MTSO verifies that the telephone number is valid and that the user is authorized to place the call; some service providers require the user to enter a PIN (personal identification number) as well as the called number to counter theft.
3. The MTSO issues a message to the user's cell phone indicating which traffic channels to use for sending and receiving.
4. The MTSO sends out a ringing signal to the called party. All of these operations (steps 2 through 4) occur within 10 s of initiating the call.
5. When the called party answers, the MTSO establishes a circuit between the two parties and initiates billing information.
6. When one party hangs up, the MTSO releases the circuit, frees the radio channels, and completes the billing information.

AMPS Control Channels

Each AMPS service includes 21 full-duplex 30-kHz control channels, consisting of 21 reverse control channels (RCCs) from subscriber to base station, and 21 forward channels base station to subscriber. These channels transmit digital data using FSK. In both channels, data are transmitted in frames.

Figure 10.11a shows the RCC frame structure. The frame begins with a 48-bit precursor, consisting of a 30-bit bit sync field of alternating ones and zeros, an 11 bit word sync field (11100010010), and a 7-bit digital color code (DCC). The DCC is used to distinguish transmissions in cochannel cells; it is a unique identifier of a base station and acts as a destination address for an RCC frame. Following the precursor, the frame contains from one to 6 words of data. Each word contains 36 data bits and is encoded using a shortened version of an $(n, k, t) = (63, 51, 5)$ BCH block code (see Table 8.4). In this shortened version, 12 check bits are added to the 36 data bits to form a 48-bit word. To further increase reliability, each word is transmitted five times in the same frame, and a majority logic is used to recover the word at the base station. When all the overhead is taken into account, the data rate is on the order of a few hundred bits per second. Examples of RCC messages include origination, page response, and order confirmation.

The FCC frame structure (Figure 10.11b) starts with a 10-bit bit sync and an 11-bit word sync. Each frame contains two words of data. Each word is encoded using BCH and contains 28 data bits and 12 check bits. Again, for reliability, each word is repeated five times. In addition, each FCC frame provides information about the status (idle or busy) of the corresponding RCC frame through the busy/idle bits that are inserted every tenth bit in the frame. This brings the total frame size to 463 bits. At the 10-kbps signaling rate, the data rate (excluding overhead) is about 1.2 kbps. FCC messages include paging messages and frequency assignment messages.

Finally, control information can be transmitted over a voice channel during a conversation. The mobile unit or the base station can insert a burst of data by turning off the voice FM transmission for about 100 ms and replacing it with an FSK-encoded message. These messages are used to exchange urgent messages, such as change power level and handoff.

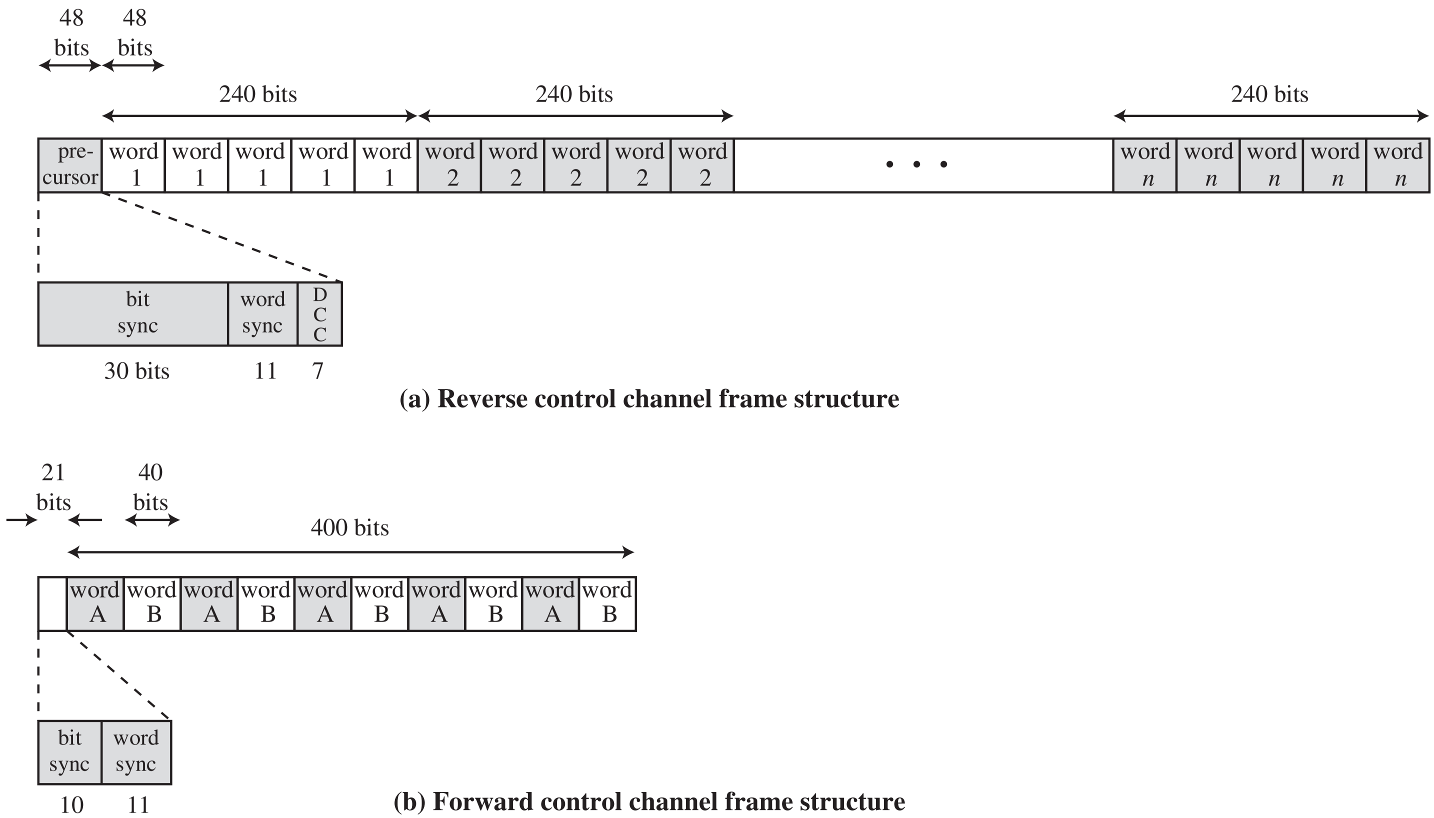
10.3 SECOND-GENERATION TDMA

This section begins our study of second-generation cellular systems. We begin with an overview and then look in detail at one type of second-generation cellular system.

First- and Second-Generation Cellular Systems

First-generation cellular networks, such as AMPS, quickly became highly popular, threatening to swamp available capacity. Second-generation systems have been developed to provide higher-quality signals, higher data rates for support of digital services, and greater capacity. [BLAC99] lists the following as the key differences between the two generations:

- **Digital traffic channels:** The most notable difference between the two generations is that first-generation systems are almost purely analog, whereas second-generation systems are digital. In particular, the first-generation systems are



(a) Reverse control channel frame structure

(b) Forward control channel frame structure

Figure 10.11 AMPS Control Channel Frame Formats designed to support voice channels using FM; digital traffic is supported only by the use of a modem that converts the digital data into analog form. Second-generation systems provide digital traffic channels. These readily support digital data; voice traffic is first encoded in digital form before transmitting. Of course, for second-generation systems, the user traffic (data or digitized voice) must be converted to an analog signal for transmission between the mobile unit and the base station (e.g., see Figure 6.14).

- **Encryption:** Because all of the user traffic, as well as control traffic, is digitized in second-generation systems, it is a relatively simple matter to encrypt all of the traffic to prevent eavesdropping. All second-generation systems provide this capability, whereas first-generation systems send user traffic in the clear, providing no security.
- **Error detection and correction:** The digital traffic stream of second-generation systems also lends itself to the use of error detection and correction techniques, such as those discussed in Chapter 8. The result can be very clear voice reception.
- **Channel access:** In first-generation systems, each cell supports a number of channels. At any given time a channel is allocated to only one user. Second-generation systems also provide multiple channels per cell, but each channel is dynamically shared by a number of users using time division multiple access (TDMA) or code division multiple access (CDMA). We look at TDMA-based systems in this section and CDMA-based systems in Section 10.4.

Beginning around 1990, a number of different second-generation systems have been deployed. Table 10.5 lists some key characteristics of three of the most important of these systems.

Time Division Multiple Access

First-generation cellular systems provide for the support of multiple users with frequency division multiple access (FDMA). FDMA was introduced in our discussion of satellite communications and the principle is the same here. FDMA for cellular systems can be described as follows. Each cell is allocated a total of $2M$ channels of bandwidth δ Hz each. Half the channels (the reverse channels) are used for transmission from the mobile unit to the base station: $f_c, f_c + \delta, f_c + 2\delta, \ldots, f_c + (M - 1)\delta$, where f_c is the center frequency of the lowest-frequency channel. The other half of the channels (the forward channels) are used for transmission from the base station to the mobile unit: $f_c, f_c + \delta + \Delta, f_c + 2\delta + \Delta, \ldots, f_c + (M - 1)\delta + \Delta$, where Δ is the spacing between the reverse and forward channels. When a connection is set up for a mobile user, the user is assigned two channels, at f and $f + \Delta$, for full-duplex communication. This arrangement is quite wasteful, because much of the time one or both of the channels are idle.

TDMA was also introduced in our discussion of satellite communications (e.g., see Figure 9.14). TDMA for cellular systems can be described as follows. As with FDMA, each cell is allocated a number of channels, half reverse and half forward. Again, for full-duplex communication, a mobile unit is assigned capacity on matching reverse and forward channels. In addition, each physical channel is further sub-

Table 10.5 Second-Generation Cellular Telephone Systems

	GSM	IS-136	IS-95
Year introduced	1990	1991	1993
Access method	TDMA	TDMA	CDMA
Base station transmission band	935 to 960 MHz	869 to 894 MHz	869 to 894 MHz
Mobile station transmission band	890 to 915 MHz	824 to 849 MHz	824 to 849 MHz
Spacing between forward and reverse channels	45 MHz	45 MHz	45 MHz
Channel bandwidth	200 kHz	30 kHz	1250 kHz
Number of duplex channels	125	832	20
Mobile unit maximum power	20 W	3 W	0.2 W
Users per channel	8	3	35
Modulation	GMSK	$\pi/4$ DQPSK	QPSK
Carrier bit rate	270.8 kbps	48.6 kbps	9.6 kbps
Speech coder	RPE-LTP	VSELP	QCELP
Speech coding bit rate	13 kbps	8 kbps	8, 4, 2, 1 kbps
Frame size	4.6 ms	40 ms	20 ms
Error control coding	Convolutional 1/2 rate	Convolutional 1/2 rate	Convolutional 1/2 rate forward; 1/3 rate reverse

divided into a number of logical channels. Transmission is in the form of a repetitive sequence of frames, each of which is divided into a number of time slots. Each slot position across the sequence of frames forms a separate logical channel. We saw an example of this in Figure 9.13.

Mobile Wireless TDMA Design Considerations

Before turning to the specific example of GSM, it will be useful to consider some general design guidelines by looking at a simple analysis, based on one in [JONE93]. This analysis motivates some of the design decisions made for GSM. The overall objective is to determine the length and composition of the traffic channel time slot that will provide effective speech and data transmission with efficient use of the radio spectrum. Let us consider the following set of requirements:

- **Number of logical channels (number of time slots in TDMA frame):** 8; this appears to be the minimum to justify the additional costs of multiplexing.
- **Maximum cell radius (R):** 35 km, to give a sufficiently high traffic level in rural areas.
- **Frequency:** Region around 900 MHz; this is commonly allocated to mobile radio applications.

- **Maximum vehicle speed (V_m):** 250 km/hr, or 69.4 m/s, to accommodate mobile units on high-speed trains.
- **Maximum coding delay:** Approximately 20 ms, to avoid adding unduly to delays within the fixed network, which may involve satellite links.
- **Maximum delay spread (Δ_m):** 10 (s (in mountainous regions); this is the difference in propagation delay among different multipath signals arriving at the same antenna.
- **Bandwidth:** Not to exceed 200 kHz, corresponding to 25 kHz per channel (the current spacing for analog FM cellular systems in Europe).

Figure 10.12 suggests the steps to be considered in designing the TDMA time slot. We use this as a guide in the following discussion.

The speech coder must provide satisfactory speech quality at minimum data rate. The traditional form of speech coding to produce a digital bit stream is pulse code modulation (PCM), which, as we saw in Section 6.4, results in a data rate of 64 kbps. This rate is undesirably high for use in cellular radio. With current technology, a data rate of 12 kbps is reasonable for producing good-quality speech reproduction.

If we restrict the coding delay to 20 ms, then it would be acceptable to form the encoded speech into blocks of 20 ms duration, or speech samples of 240 bits.

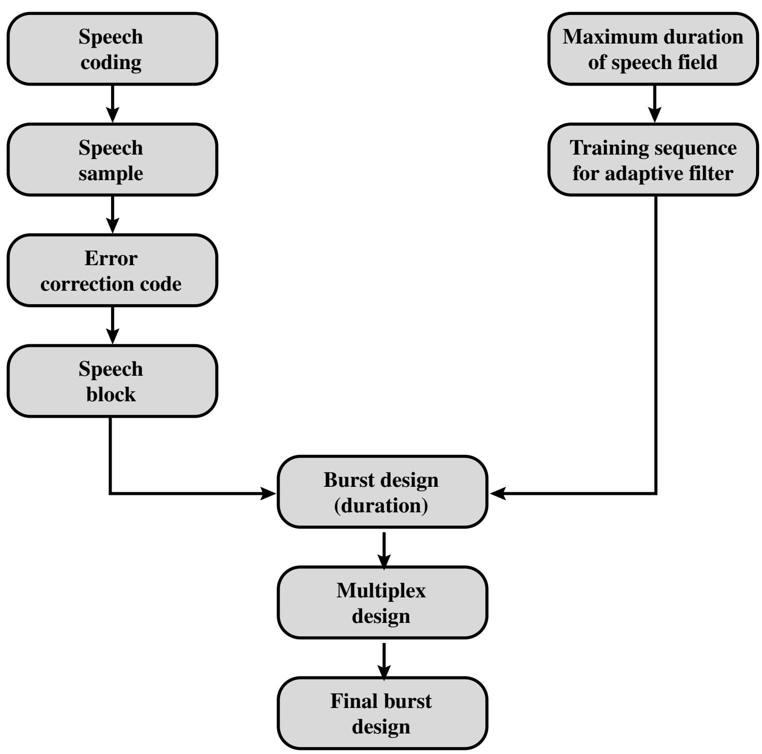


Figure 10.12 Steps in Design of TDMA Time Slot

Data at 12 kbps could also be blocked in 240-bit units. Error correction can then be applied to the 240-bit blocks.

For second-generation digital systems, convolutional error-correcting codes are commonly used with a code rate of 1/2. This overhead raises the number of bits in a block to 480. In addition, there is a constraint factor of 5, meaning that 4 bits must be added to the data block to account for the length of the shift register (see Section 8.3). This brings the speech block length to 488 bits.

With the parameters chosen so far, the minimum bit rate for an eight-channel system is

$$\frac{8 \text{ channels } \times 488 \text{ bits/channel}}{20 \times 10^{-3} \text{ s}} = 195.2 \text{ kbps}$$

In fact, the bit rate will be somewhat higher to take care of other design considerations, discussed subsequently. This means that a data rate of greater than 200 kbps will need to be carried in the available bandwidth of 200 kHz. In practice, such data rates cannot be achieved without the use of adaptive equalization. As was discussed in Section 5.4, in a mobile environment, adaptive equalization will require the inclusion of a new training sequence each time the mobile unit moves a distance sufficient to potentially cause changes in transmission path characteristics. Let us assume that a training sequence is included in each time slot. A rough criterion suggested in [JONE93] is that the phase angle of the carrier signal should be restricted to a change of 1/20th of a wavelength (an angle of $\pi/10$) after the training sequence. At 900 MHz, the wavelength is 0.333m. We can calculate

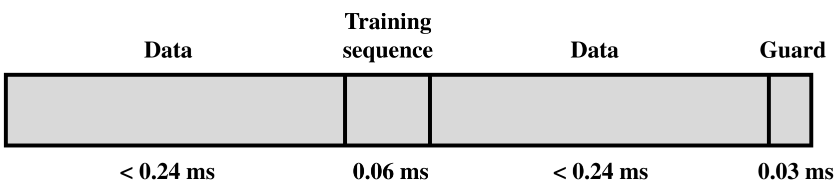
$$\text{Maximum transmission duration} = \frac{\lambda/20}{V_m} = \frac{0.333/20}{69.4} = 0.24 \text{ ms}$$

We can take better advantage of the training sequence by transmitting 0.24 ms of speech or data both before and after the training sequence and using the training sequence on the combined 0.48 ms of data.

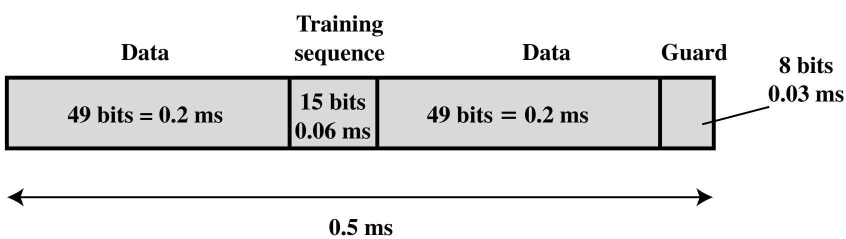
Next, we need to determine the length of the training sequence. In the design of an equalizer for a multipath signal whose bandwidth is about equal to the bit rate (200 kHz, 200 kbps), a rule of thumb is that the number of taps on the equalizer (Figure 5.14) should be equal to 6 times the number of bits transmitted in the maximum dispersal time ($\Delta_m = 0.01$ ms). Thus, the amount of time in the time slot devoted to the training sequence is 0.06 ms.

Now consider that a guard interval is needed at the end of each time slot to account for the differing amounts of delay between different mobile units and the base station. Because eight mobile units share the same TDMA frame, it is necessary to adjust the timing of the transmissions of the mobile units so that the transmission from one unit does not interfere with adjacent time slots. It is the responsibility of the base station to provide synchronization information to each mobile unit to adjust relative delays to enforce the time slot structure of the TDMA frame. However, the mobile units may be moving relative to the base station and relative to each other, so a guard time is inserted in each time slot to account for these discrepancies. When a mobile first makes a connection through the base station, the base station can provide the mobile unit with timing information based on the current propagation delay between the mobile unit and the base station. We would also like to add a guard time sufficient to avoid the need to frequently update this synchronization information. We can calculate the guard time as follows. The average telephone call is about 130 seconds [JONE93], so the radial distance toward or away from the base station that a mobile could cover is (130 s) $\times$ (69.4 m/s) = 9022 m. The change in propagation delay caused by a movement of this distance is $9022/(3 \times 10^8 \text{ m/s}) = 0.03$ ms.

Figure 10.13a shows the tentative time slot design. The next step is to fit a coded data block into a convenient number of time slots, together with the training sequence and guard bits. We have a maximum duration of a time slot of approximately 0.57 ms. With 8 time slots per frame, that gives a frame time of about 4.6 ms. We said that we wanted to send data with a coding delay of 20 ms, so if we round the frame time down to 4 ms (time slot = 0.5 ms), then we could conveniently send a block of speech in five successive slots on the same channel. A speech block consists of 488 bits, so each time slot would need to hold 488/5 or about 98 data bits. This yields a bit rate of 98/0.4 = 245 kbps. At this data rate, the minimum number of training bits required is (0.06 ms) $\times$ (245 kbps) = 14.7, which on rounding becomes 15 bits. Similarly, the minimum number of guard bits is (0.03 ms) $\times$ (245 kbps) = 7.35, which on rounding becomes 8 bits.



(a) Approximate field durations



(b) Approximate field sizes

Figure 10.13 TDMA Time Slot

The resulting frame structure is shown in Figure 10.13b. We have 121 bits transmitted in 0.5 ms for a channel bit rate of 242 kbps.

Global System for Mobile Communications

Before the Global System for Mobile Communications (GSM) was developed, the countries of Europe used a number of incompatible first-generation cellular phone technologies. GSM was developed to provide a common second-generation technology for Europe so that the same subscriber units could be used throughout the continent. The technology has been extremely successful and is probably the most popular standard, worldwide, for new implementations. GSM first appeared in 1990 in Europe. Similar systems have now been implemented in North and South America, Asia, North Africa, the Middle East, and Australia. The GSM Association claimed nearly three-quarters of a billion subscribers worldwide at the end of 2000, the bulk of these in Europe and Asia Pacific, but with over 8 million in North America.

GSM Network Architecture

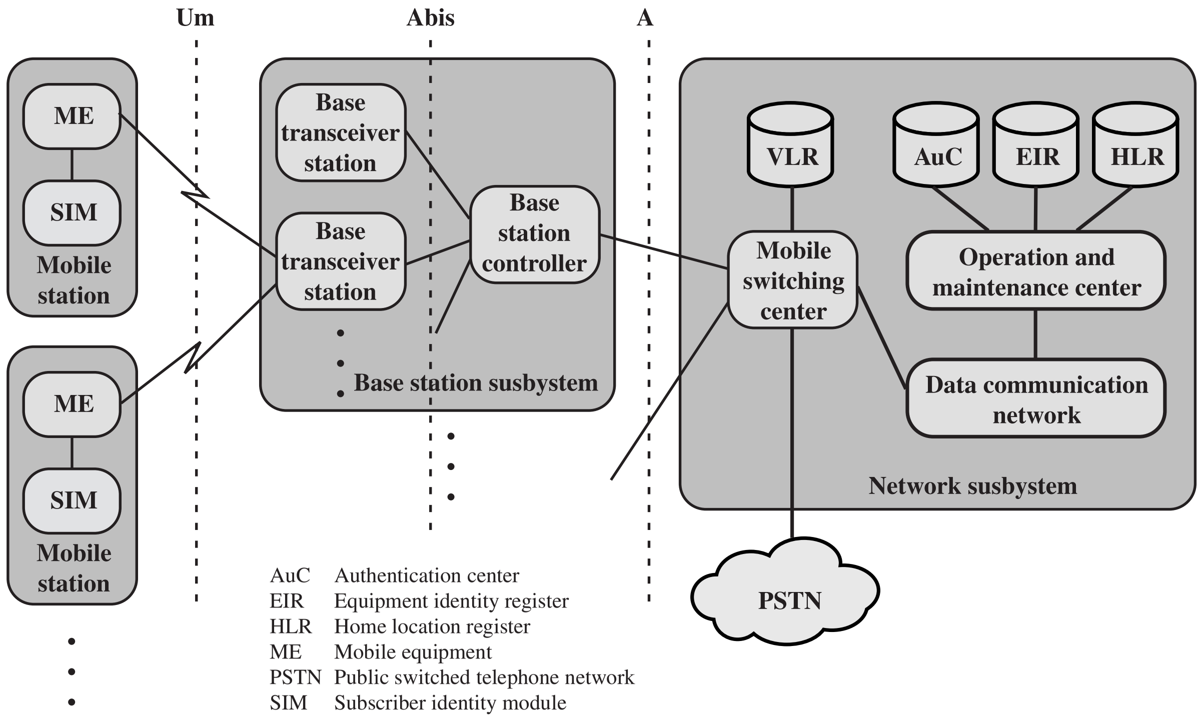
Figure 10.14 shows the key functional elements in the GSM system. The boundaries at Um, Abis, and A refer to interfaces between functional elements that are standardized in the GSM documents. Thus, it is possible to buy equipment from different vendors with the expectation that they will successfully interoperate. Additional interfaces are also defined in the GSM standards but need not concern us here.

Mobile Station

A mobile station communicates across the Um interface, also known as the **air interface**, with a base station transceiver in the same cell in which the mobile unit is located. The **mobile equipment** (ME) refers to the physical terminal, such as a telephone or PCS (personal communications service) device, which includes the radio transceiver, digital signal processors, and the **subscriber identity module** (SIM). The SIM is a portable device in the form of a smart card or plug-in module that stores the subscriber's identification number, the networks the subscriber is authorized to use, encryption keys, and other information specific to the subscriber. The GSM subscriber units are totally generic until an SIM is inserted. Therefore, a subscriber need only carry his or her SIM to use a wide variety of subscriber devices in many countries simply by inserting the SIM in the device to be used. In fact, except for certain emergency communications, the subscriber units will not work without a SIM inserted. Thus, the SIMs roam, not necessarily the subscriber devices.

Base Station Subsystem

A base station subsystem (BSS) consists of a base station controller and one or more base transceiver stations. Each **base transceiver station** (BTS) defines a single cell; it includes a radio antenna, a radio transceiver, and a link to a base station controller. A GSM cell can have a radius of between 100 m and 35 km, depending on the environment. A **base station controller** (BSC) may be collocated with a BTS or may control multiple BTS units and hence multiple cells. The BSC



AuC	Authentication center
EIR	Equipment identity register
HLR	Home location register
ME	Mobile equipment
PSTN	Public switched telephone network
SIM	Subscriber identity module
VLR	Visitor location register

Figure 10.14 Overall GSM Architecture

reserves radio frequencies, manages the handoff of a mobile unit from one cell to another within the BSS, and controls paging.

Network Subsystem

The network subsystem (NS) provides the link between the cellular network and the public switched telecommunications networks. The NS controls handoffs between cells in different BSSs, authenticates users and validates their accounts, and includes functions for enabling worldwide roaming of mobile users. The central element of the NS is the **mobile switching center** (MSC). It is supported by four databases that it controls:

- **Home location register (HLR) database:** The HLR stores information, both permanent and temporary, about each of the subscribers that "belongs" to it (i.e., for which the subscriber has its telephone number associated with the switching center).
- **Visitor location register (VLR) database:** One important, temporary piece of information is the location of the subscriber. The location is determined by the VLR into which the subscriber is entered. The visitor location register maintains information about subscribers that are currently physically in the region covered by the switching center. It records whether or not the subscriber is active and other parameters associated with the subscriber. For a call coming to the subscriber, the system uses the telephone number associated with the subscriber to identify the home switching center of the subscriber. This switching center can find in its HLR the switching center in which the subscriber is currently physically located. For a call coming from the subscriber, the VLR is used to initiate the call. Even if the subscriber is in the area covered by its home switching center, it is also represented in the switching center's VLR, for consistency.
- **Authentication center database (AuC):** This database is used for authentication activities of the system; for example, it holds the authentication and encryption keys for all the subscribers in both the home and visitor location registers. The center controls access to user data as well as being used for authentication when a subscriber joins a network. GSM transmission is encrypted, so it is private. A stream cipher, A5, is used to encrypt the transmission from subscriber to base transceiver. However, the conversation is in the clear in the landline network. Another cipher, A3, is used for authentication.
- **Equipment identity register database (EIR):** The EIR keeps track of the type of equipment that exists at the mobile station. It also plays a role in security (e.g., blocking calls from stolen mobile stations and preventing use of the network by stations that have not been approved).

Radio Link Aspects

The GSM spectral allocation is 25 MHz for base transmission (935–960 MHz) and 25 MHz for mobile transmission (890–915 MHz). Other GSM bands have also been defined outside Europe. Users access the network using a combination of frequency division multiple access (FDMA) and time division multiple access (TDMA) (both are discussed in the next section). There are radio-frequency carriers every 200 kHz, which provide for 125 full-duplex channels. The channels are modulated at a data rate of 270.833 kbps. As with AMPS, there are two types of channels, traffic and control.

TDMA Format

GSM uses a complex hierarchy of TDMA frames to define logical channels (Figure 10.15). Fundamentally, each 200-kHz frequency band is divided into 8 logical channels defined by the repetitive occurrence of time slots.

At the lowest level is the time slot, also called a burst period, which has a duration of 15/26 ms, or approximately 0.577 ms. With a bit rate of 270.833 kbps, each time slot has a length of 156.25 bits. The time slot includes the following fields:

- **Trail bits:** Allow synchronization of transmissions from mobile units located at different distances from the base station, as explained subsequently.
- **Encrypted bits:** Data is encrypted in blocks by conventional encryption of 114 plaintext bits into 114 ciphertext bits; the encrypted bits are then placed in two 57-bit fields in the time slot.
- **Stealing bit:** Used to indicate whether this block contains data or is “stolen” for urgent control signaling.
- **Training sequence:** Used to adapt the parameters of the receiver to the current path propagation characteristics and to select the strongest signal in case of multipath propagation. The training sequence is a known bit pattern that differs for different adjacent cells. It enables the mobile units and base stations to determine that the received signal is from the correct transmitter and not a strong interfering transmitter. In addition, the training sequence is used for multipath equalization, which is used to extract the desired signal from unwanted reflections. By determining how the known training sequence is modified by multipath fading, the rest of the signal is processed to compensate for these effects.
- **Guard bits:** Used to avoid overlapping with other bursts due to different path delays.

The time slot format shown in Figure 10.15 is called a normal burst and carries user data traffic (compare Figure 10.13b). Other burst formats are used for control signaling.

Moving up the frame format hierarchy, 8-slot TDMA frames are typically organized into a 26-frame multiframe. One of the frames in the multiframe is used for control signaling and another is currently unused, leaving 24 frames for data traffic. Thus, each traffic channel receives one slot per frame and 24 frames per 120-ms multiframe. The resulting data rate is

$$\frac{114 \text{ bits/slot} \times 24 \text{ slots/multiframe}}{120 \text{ ms/multiframe}} = 22.8 \text{ kbps}$$

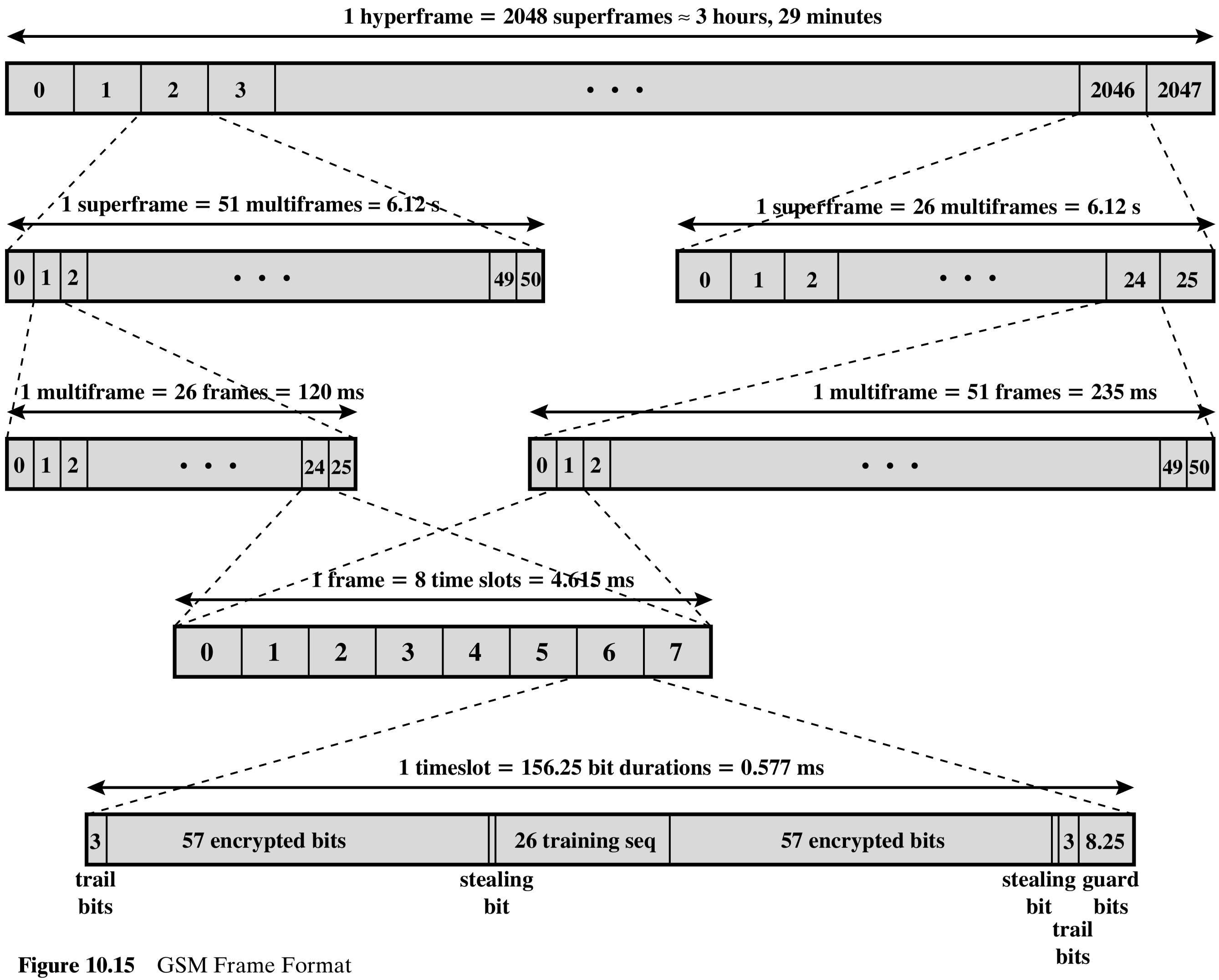


Figure 10.15 GSM Frame Format

The GSM specification also allows half-rate traffic channels, with two traffic channels each occupying one time slot in 12 of the 26 frames. With the use of half-rate speech coders, this effectively doubles the capacity of the system. There is also a 51-frame multiframe used for control traffic.

Speech coding

Figure 10.16 provides an overview of the processing of speech signals for transmission over a logical traffic channel. We look at each of these steps in turn.

The speech signal is compressed using an algorithm known as Regular Pulse Excited—Linear Predictive Coder (RPE-LPE) [KROO86]. In essence, data from previous samples are used to predict the current sample. Each sample is then encoded to consist of bits representing the coefficients of the linear combination of previous samples plus an encoded form of the difference between the predicted and actual sample. The result of the use of this code is to produce 260 bits every 20 ms, for a raw

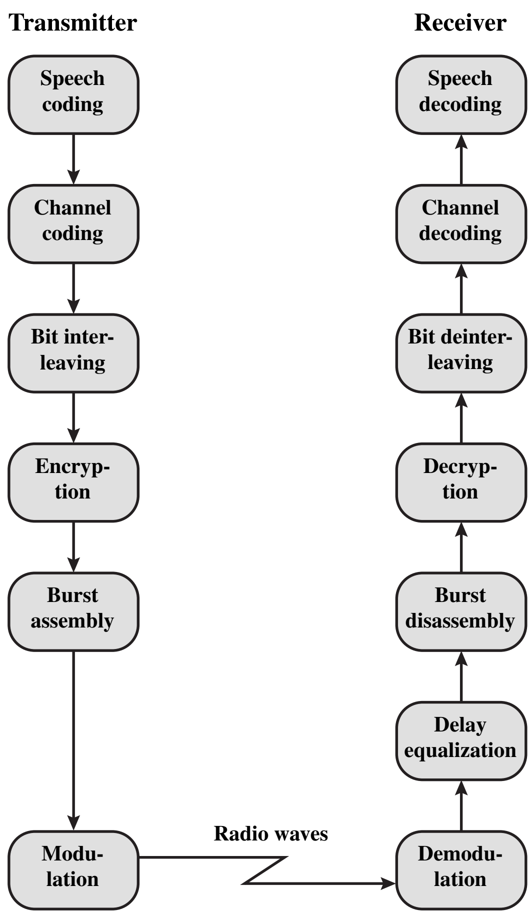


Figure 10.16 GSM Speech Signal Processing

data rate of 13 kbps. From the point of view of the quality of the speech produced by this encoding, the bits in the 260-bit block can be divided into three classes:

- **Class Ia:** 50 bits, most sensitive to bit errors
- **Class Ib:** 132 bits, moderately sensitive to bit errors
- **Class II:** 78 bits, least sensitive to bit errors

The first 50 bits are protected by a 3-bit cyclic redundancy check (CRC) error detection code. If an error is detected, the entire sample is discarded and replaced by a modified version of the preceding block. These 53 bits plus the 132 class 1b bits, plus a 4-bit tail sequence, are then protected by a convolutional (1, 2, 5) error-correcting code, resulting in $189 \times 2 = 378$ bits. The remaining 78 bits are unprotected and are appended to the protected bit to produce a block of 456 bits, with a resulting data rate of 456/20 ms = 22.8 kbps, which is the GSM traffic channel data rate.

To add protection against burst errors, each 456-bit block is divided into eight 57-bit blocks, which are transmitted in eight consecutive time slots. Because each time slot can carry two 57-bit blocks, each burst carries data from two different speech samples.

Following these steps, the speech data are encrypted 114 bits at a time, assembled into time slots (burst assembly), and finally modulated for transmission. The modulation scheme, Gaussian minimum shift keying (GMSK), is a form of frequency shift keying (FSK).

Data Encoding

Digital data are processed in a similar fashion as applied to speech signals. Data are processed in blocks of 240 bits every 20 ms, for a data rate of 12 kbps. Depending on the way logical channels are defined, the actual supported data rates are 9.6, 4,8, and 2.4 kbps. Each block is augmented by four tail bits. A (1, 2, 5) convolutional code is used to produce a block of $244 \times 2 = 488$ bits. Then 32 bits of this block are dropped (puncturing), leaving a block of 456 bits. A bit interleaving scheme is then used to spread the data over multiple bursts, again to reduce the effects of burst noise. The 488 bits are spread over 22 bursts in the following fashion:

- The 1st and 22nd bursts carry 6 bits each
- The 2nd and 21st bursts carry 12 bits each
- The 3rd and 20th bursts carry 18 bits each
- The 4th through 19th bursts carry 24 bits each

The result is that each burst carries information from 5 or 6 consecutive data blocks.

Slow Frequency Hopping

We have said that a given traffic channel is assigned a given frequency channel for transmission and reception. This is not strictly correct. GSM and many other cellular schemes use a technique known as slow frequency hopping to improve signal quality. Each successive TDMA frame in a given channel is carried on a different carrier frequency. Thus, the transmission frequency is changed once every 4.615 ms. Because multipath fading is dependent on carrier frequency, slow frequency hopping helps to compensate. Slow frequency hopping also reduces the effects of cochannel interference. Note that this is a form of spread spectrum communication.

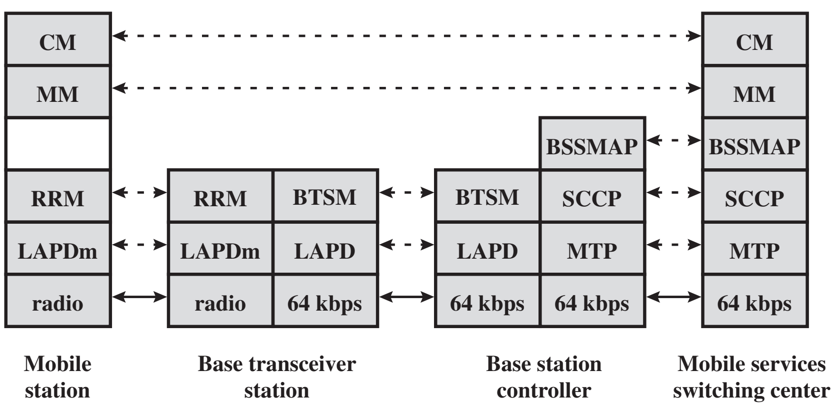
Delay Equalization

Because mobile units are at different distances from the base station within a cell, their transmissions suffer differing amounts of delay. This phenomenon creates a design issue, because up to eight mobile units share the same TDMA frame. Thus, the timing of frame slots is critical. The base station provides a control signal to synchronize the timing of the various mobile units. Within the slot format, the tail bits and guard bits provide a margin to prevent the overlap of data bits from one time slot to another. The base station can adjust the timing of any active mobile by control signals that instruct the mobile to increment or decrement its timing.

GSM Signaling Protocol Architecture

A number of control messages are exchanged between the key entities in Figure 10.14 that deal with mobility, radio resources, and connection management. A detailed look at the various message formats and semantics could fill a book. Here we give an overview of the structure, which suggests the complexity of second-generation design.

Figure 10.17 summarizes the protocols used between the main elements of the network architecture. The lowest layer of the architecture is tailored to the physical link between entities. Between the mobile station and the base transceiver station,



BSSMAP	=	BSS mobile application part	MM	=	mobility management
BTSM	=	BTS management	MTP	=	message transfer part
CM	=	connection management	RRM	=	radio resources management
LAPD	=	link access protocol, D channel	SCCP	=	signal connection control part

Figure 10.17 GSM Signaling Protocol Architecture

the radio link discussed in preceding subsections carries higher-level data inside the TDMA format. Between other entities a standard 64-kbps digital channel is used.

At the link layer, a data link control protocol (see Figure 4.3) known as LAPDm is used. This is a modified version of the LAPD protocol defined for the Integrated Services Digital Network (ISDN). The remaining links use the normal LAPD protocol. In essence, LAPD is designed to convert a potentially unreliable physical link into a reliable data link. It does this by using a cyclic redundancy check to perform error detection (discussed in Section 8.1) and automatic repeat request (ARQ) to retransmit damaged frames (discussed in Section 8.4).[3]

Above the link layer are a number of protocols that provide specific functions. These include the following:

- **Radio resource management:** Controls the setup, maintenance, and termination of radio channels, including handoffs.
- **Mobility management:** Manages the location updating and registration procedures, as well as security and authentication.
- **Connection management:** Handles the setup, maintenance, and termination of calls (connections between end users).
- **Mobile application part (MAP):** Handles most of the signaling between different entities in the fixed part of the network, such as between the HLR and VLR.
- **BTS management:** Performs various management and administrative functions at the base transceiver station, under the control of the base station controller.

The MAP does not run directly on top of the link layer but rather on top of two intermediate protocols, SCCP and MTP. These latter protocols are part of Signaling System Number 7, which is a set of protocols designed to provide control signaling within digital circuit-switching networks, such as digital public telecommunications networks. These protocols provide general functions used by various applications, including MAP.

10.4 SECOND-GENERATION CDMA

Code division multiple access (CDMA) is a spread spectrum–based technique for multiplexing, introduced in Section 7.4, that provides an alternative to TDMA for second-generation cellular networks. We begin this section with an overview of the advantages of the CDMA approach and then look at the most widely used scheme, IS-95.

Code Division Multiple Access

CDMA for cellular systems can be described as follows. As with FDMA, each cell is allocated a frequency bandwidth, which is split into two parts, half for reverse

[3]See Appendix D for a discussion of data link control protocols.

(mobile unit to base station) and half for forward (base station to mobile unit). For full-duplex communication, a mobile unit uses both reverse and forward channels. Transmission is in the form of direct-sequence spread spectrum (DS-SS), which uses a chipping code to increase the data rate of the transmission, resulting in an increased signal bandwidth. Multiple access is provided by assigning orthogonal chipping codes to multiple users, so that the receiver can recover the transmission of an individual unit from multiple transmissions.

CDMA has a number of advantages for a cellular network:

- **Frequency diversity:** Because the transmission is spread out over a larger bandwidth, frequency-dependent transmission impairments, such as noise bursts and selective fading, have less effect on the signal.
- **Multipath resistance:** In addition to the ability of DS-SS to overcome multipath fading by frequency diversity, the chipping codes used for CDMA not only exhibit low cross correlation but also low autocorrelation. Therefore, a version of the signal that is delayed by more than one chip interval does not interfere with the dominant signal as much as in other multipath environments.
- **Privacy:** Because spread spectrum is obtained by the use of noiselike signals, where each user has a unique code, privacy is inherent.
- **Graceful degradation:** With FDMA or TDMA, a fixed number of users can access the system simultaneously. However, with CDMA, as more users access the system simultaneously, the noise level and hence the error rate increases; only gradually does the system degrade to the point of an unacceptable error rate.

A number of drawbacks of CDMA cellular should also be mentioned:

- **Self-jamming:** Unless all of the mobile users are perfectly synchronized, the arriving transmissions from multiple users will not be perfectly aligned on chip boundaries. Thus the spreading sequences of the different users are not orthogonal and there is some level of cross correlation. This is distinct from either TDMA or FDMA, in which for reasonable time or frequency guardbands, respectively, the received signals are orthogonal or nearly so.
- **Near-far problem:** Signals closer to the receiver are received with less attenuation than signals farther away. Given the lack of complete orthogonality, the transmissions from the more remote mobile units may be more difficult to recover. Thus, power control techniques are very important in a CDMA system.
- **Soft handoff:** As is discussed subsequently, a smooth handoff from one cell to the next requires that the mobile acquires the new cell before it relinquishes the old. This is referred to as a soft handoff and is more complex than the hard handoff used in FDMA and TDMA schemes.

Mobile Wireless CDMA Design Considerations

Before turning to the specific example of GSM, it will be useful to consider some general design elements of a CDMA cellular system.

RAKE Receiver

In a multipath environment, which is common in cellular systems, if the multiple versions of a signal arrive more than one chip interval apart from each other, the receiver can recover the signal by correlating the chip sequence with the dominant incoming signal. The remaining signals are treated as noise. However, even better performance can be achieved if the receiver attempts to recover the signals from multiple paths and then combine them, with suitable delays. This principle is used in the RAKE receiver.

Figure 10.18 illustrates the principle of the RAKE receiver. The original binary signal to be transmitted is spread by the exclusive-OR (XOR) operation with the transmitter's chipping code. The spread sequence is then modulated for transmission over the wireless channel. Because of multipath effects, the channel generates multiple copies of the signal, each with a different amount of time delay (τ_1, τ_2, etc), and each with a different attenuation factors (a_1, a_2, etc.). At the receiver, the combined signal is demodulated. The demodulated chip stream is then fed into multiple correlators, each delayed by a different amount. These signals are then combined using weighting factors estimated from the channel.

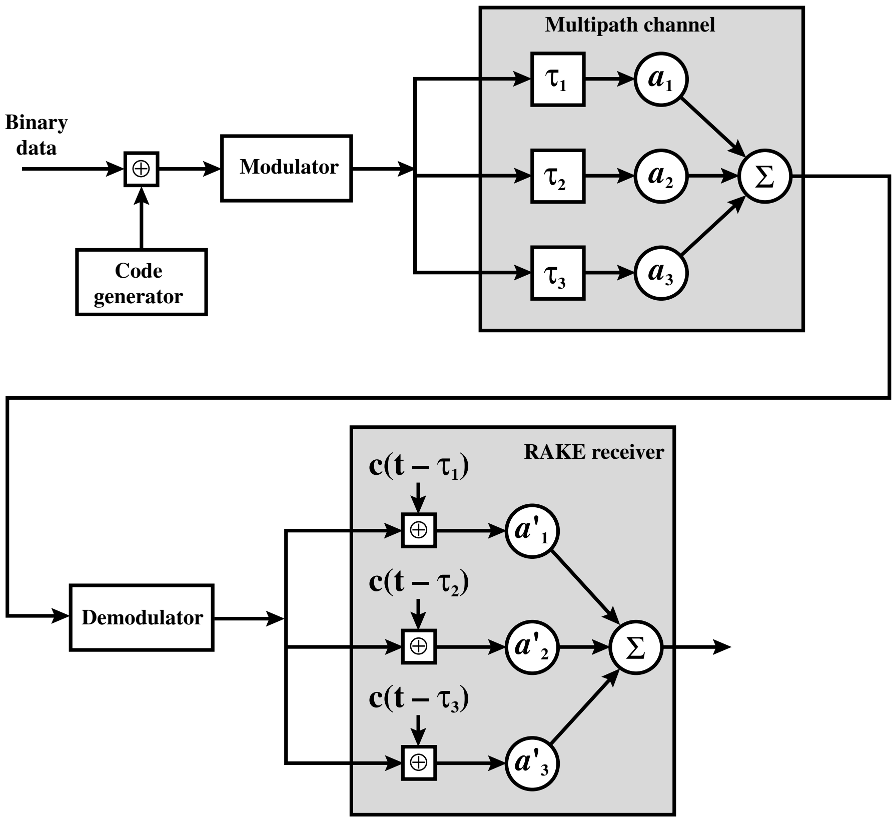


Figure 10.18 Principle of RAKE Receiver [PRAS98]

Soft Handoff

In an FDMA or TDMA system, neighboring cells use different portions of the available frequency spectrum (i.e., the frequency reuse factor N is greater than 1, typically 7). When the signal strength of a neighboring cell exceeds that of the current cell, plus a threshold, the mobile station is instructed to switch to a new frequency band that is within the allocation of the new cell. This is referred to as a **hard handoff**. In a typical CDMA cellular system, spatial separation of frequencies is not used (i.e., $N = 1$), because most of the time the interference from neighboring cells will not prohibit correct reception of a DS-SS signal.

In soft handoff, a mobile station is temporarily connected to more than one base station simultaneously. A mobile unit may start out assigned to a single cell. If the unit enters a region in which the transmissions from two base stations are comparable (within some threshold of each other), the mobile unit enters the soft handoff state in which it is connected to the two base stations. The mobile unit remains in this state until one base station clearly predominates, at which time it is assigned exclusively to that cell.

While in the soft handoff state, the transmissions from the mobile unit reaching the two base stations are both sent on to the mobile switching center, which estimates the quality of the two signals and selects one. The switch sends data or digitized speech signals to both base stations, which transmit them to the mobile unit. The mobile unit combines the two incoming signals to recover the information.

IS-95

The most widely used second-generation CDMA scheme is IS-95, which is primarily deployed in North America. Table 10.5 lists some key parameters of the IS-95 system. The transmission structures on the forward and reverse links differ and are described separately.

IS-95 Forward Link

Table 10.6 lists forward link channel parameters. The forward link consists of up to 64 logical CDMA channels each occupying the same 1228-kHz bandwidth (Figure 10.19a). The forward link supports four types of channels:

- **Pilot (channel 0):** A continuous signal on a single channel. This channel allows the mobile unit to acquire timing information, provides phase reference for the demodulation process, and provides a means for signal strength comparison for the purpose of handoff determination. The pilot channel consists of all zeros.
- **Synchronization (channel 32):** A 1200-bps channel used by the mobile station to obtain identification information about the cellular system (system time, long code state, protocol revision, etc.).
- **Paging (channels 1 to 7):** Contain messages for one or more mobile stations.
- **Traffic (channels 8 to 31 and 33 to 63):** The forward channel supports 55 traffic channels. The original specification supported data rates of up to 9600 bps. A subsequent revision added a second set of rates up to 14,400 bps.

Table 10.6 IS-95 Forward Link Channel Parameters

Channel	Sync	Paging		Traffic Rate Set 1				Traffic Rate Set 2			
Data rate (bps)	1200	4800	9600	1200	2400	4800	9600	1800	3600	7200	14400
Code repetition	2	2	1	8	4	2	1	8	4	2	1
Modulation symbol rate (sps)	4800	19,200	19,200	19,200	19,200	19,200	19,200	19,200	19,200	19,200	19,200
PN chips/ modulation symbol	256	64	64	64	64	64	64	64	64	64	64
PN chips/bit	1024	256	128	1024	512	256	128	682.67	341.33	170.67	85.33

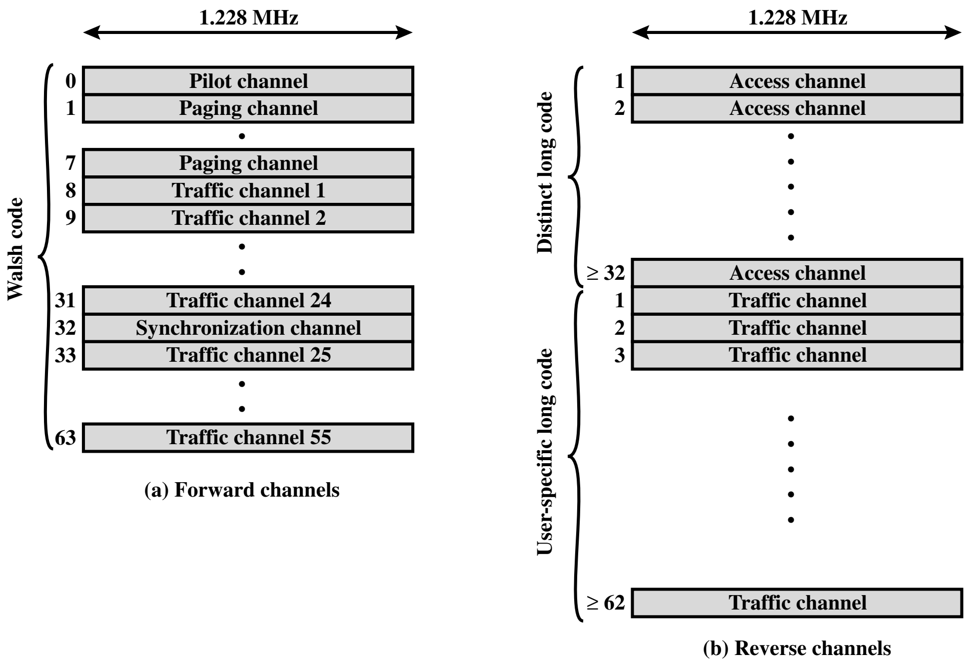


Figure 10.19 IS-95 Channel Structure

Note that all of these channels use the same bandwidth. The chipping code is used to distinguish among the different channels. For the forward channel, the chipping codes are the 64 orthogonal 64-bit codes derived from a 64×64 Walsh matrix (discussed in Section 7.5; see Figure 7.17).

Figure 10.20 shows the processing steps for transmission on a forward traffic channel using rate set 1. For voice traffic, the speech is encoded at a data rate of 8550 bps. After additional bits are added for error detection, the rate is 9600 bps. The full channel capacity is not used when the user is not speaking. During quiet periods the data rate is lowered to as low as 1200 bps. The 2400-bps rate is used to transmit transients in the background noise, and the 4800 bps rate is used to mix digitized speech and signaling data.

The data or digitized speech is transmitted in 20-ms blocks with forward error correction provided by a convolutional encoder with rate 1/2, thus doubling the effective data rate to a maximum of 19.2 kbps. For lower data rates, the encoder output bits (called code symbols) are replicated to yield the 19.2-kbps rate. The data are then interleaved in blocks to reduce the effects of errors by spreading them out.

Following the interleaver, the data bits are scrambled. The purpose of this is to serve as a privacy mask and also to prevent the sending of repetitive patterns, which in turn reduces the probability of users sending at peak power at the same time. The scrambling is accomplished by means of a long code that is generated as a pseudorandom number from a 42-bit long shift register. The shift register is initialized with the user's electronic serial number. The output of the long code generator is at a rate of 1.2288 Mbps, which is 64 times the rate of 19.2 kbps, so only one bit in 64 is selected (by the decimator function). The resulting stream is XORed with the output of the block interleaver.

The next step in the processing inserts power control information in the traffic channel. The power control function of the base station robs the traffic channel of bits at a rate of 800 bps. These are inserted by stealing code bits. The 800-bps channel carries information directing the mobile unit to increment, decrement, or keep stable its current output level. This power control stream is multiplexed into the 19.2 kbps by replacing some of the code bits, using the long code generator to encode the bits.

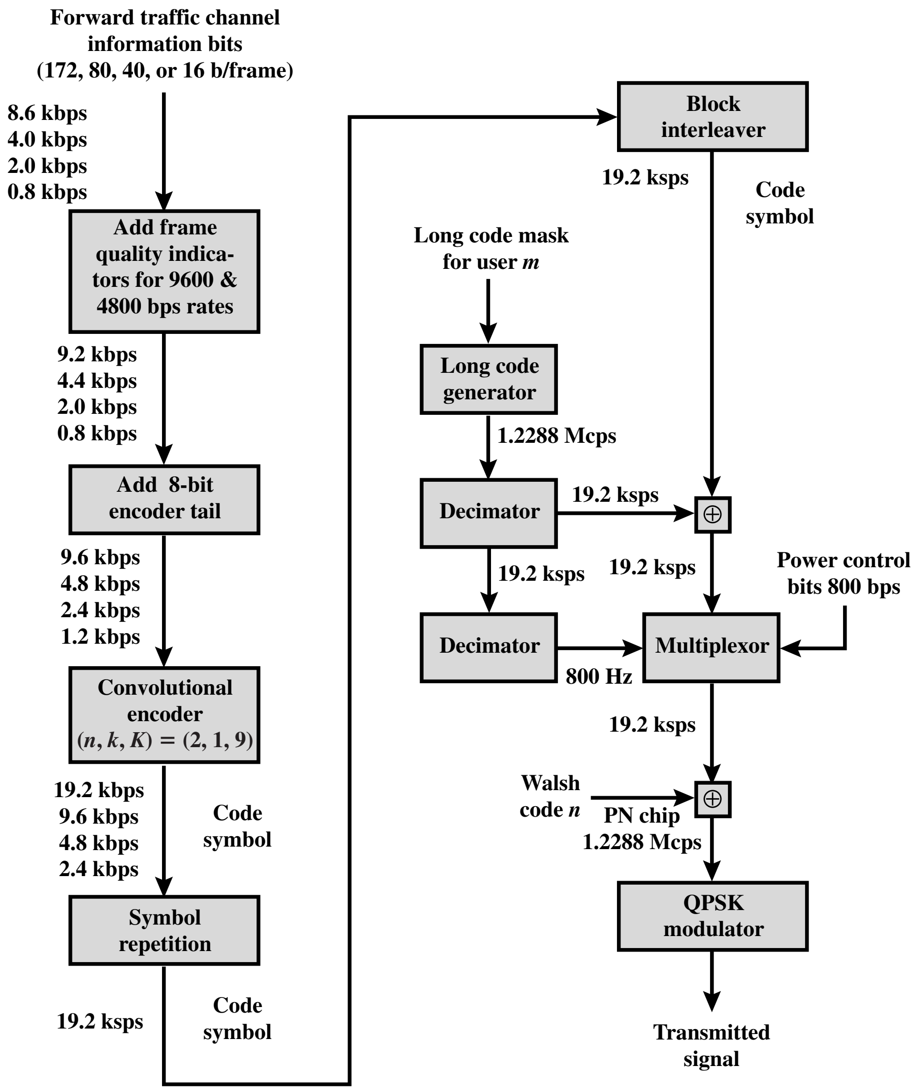


Figure 10.20 IS-95 Forward Link Transmission

The next step in the process is the DS-SS function, which spreads the 19.2 kbps to a rate of 1.2288 Mbps using one row of the 64×64 Walsh matrix. One row of the matrix is assigned to a mobile station during call setup. If a 0 bit is presented to the XOR function, then the 64 bits of the assigned row are sent. If a 1 is presented, then the bitwise XOR of the row is sent. Thus, the final bit rate is 1.2288 Mbps. This digital bit stream is then modulated onto the carrier using a QPSK modulation scheme. Recall from Chapter 6 that QPSK involves creating two bits streams that are separately modulated (see Figure 6.6). In the IS-95 scheme, the data are split into I and Q (in-phase and quadrature) channels and the data in each channel are XORed with a unique short code. The short codes are generated as pseudorandom numbers from a 15-bit long shift register.

IS-95 Reverse Link

Table 10.7 lists reverse link channel parameters. The reverse link consists of up to 94 logical CDMA channels each occupying the same 1228-kHz bandwidth (Figure 10.19b). The reverse link supports up to 32 access channels and up to 62 traffic channels.

The traffic channels in the reverse link are mobile unique. Each station has a unique long code mask based on its electronic serial number. The long code mask is a 42-bit number, so there are $2^{42} - 1$ different masks. The access channel is used by a mobile to initiate a call, to respond to a paging channel message from the base station, and for a location update.

Figure 10.21 shows the processing steps for transmission on a reverse traffic channel using rate set 1. The first few steps are the same as for the forward channel. For the reverse channel, the convolutional encoder has a rate of 1/3, thus tripling the effective data rate to a maximum of 28.8 kbps. The data are then block interleaved.

The next step is a spreading of the data using the Walsh matrix. The way in which the matrix is used, and its purpose, are different than that of the forward channel. In the reverse channel, the data coming out of the block interleaver are grouped in units of 6 bits. Each 6-bit unit serves as an index to select a row of the 64×64 Walsh matrix ($2^6 = 64$), and that row is substituted for the input. Thus the data rate is expanded by a factor of 64/6 to 307.2 kbps. The purpose of this encoding is to improve reception at the base station. Because the 64 possible codings are orthogonal, the block coding enhances the decision-making algorithm at the receiver and is also computationally efficient (see [PETE95] for details). We can view this Walsh modulation as a form of block error-correcting code with $(n, k) = (64, 6)$ and $d_{\min} = 32$. In fact, all distances are 32.

The data burst randomizer is implemented to help reduce interference from other mobile stations (see [BLAC99] for a discussion). The operation involves using the long code mask to smooth the data out over each 20-ms frame.

The next step in the process is the DS-SS function. In the case of the reverse channel, the long code unique to the mobile is XORed with the output of the randomizer to produce the 1.2288-Mbps final data stream. This digital bit stream is then modulated onto the carrier using an orthogonal QPSK modulation scheme. This differs from the forward channel in the use of a delay element in the modulator (Figure 6.6) to produce orthogonality. The reason the modulators are different is that in the forward channel, the spreading codes are orthogonal, all coming from

Table 10.7 IS-95 Reverse Link Channel Parameters

Channel	Access	Traffic-Rate Set 1				Traffic-Rate Set 2			
Data rate (bps)	4800	1200	2400	4800	9600	1800	3600	7200	14400
Code rate	1/3	1/3	1/3	1/3	1/3	1/2	1/2	1/2	1/2
Symbol rate before repetition (sps)	14,400	3600	7200	14,400	28,800	3600	7200	14,400	28,800
Symbol repetition	2	8	4	2	1	8	4	2	1
Symbol rate after repetition (sps)	28,800	28,800	28,800	28,800	28,800	28,800	28,800	28,800	28,800
Transmit duty cycle	1	1/8	1/4	1/2	1	1/8	1/4	1/2	1
Code symbols/modulation symbol	6	6	6	6	6	6	6	6	6
PN chips/ modulation symbol	256	256	256	256	256	256	256	256	256
PN chips/bit	256	128	128	128	128	256/3	256/3	256/3	256/3

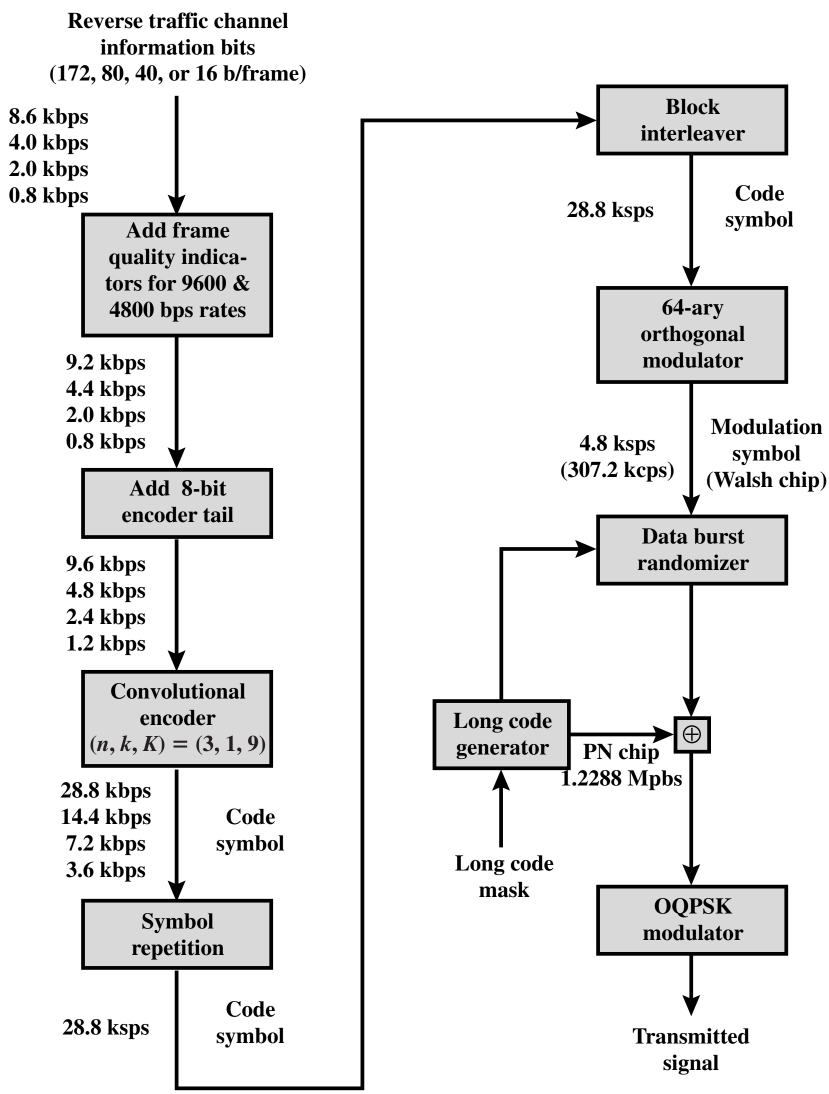


Figure 10.21 IS-95 Reverse Link Transmission

the Walsh matrix, whereas in the reverse channel, orthogonality of the spreading codes is not guaranteed.

10.5 THIRD-GENERATION SYSTEMS

The objective of the third-generation (3G) of wireless communication is to provide fairly high-speed wireless communications to support multimedia, data, and video in addition to voice. The ITU's International Mobile Telecommunications for the year 2000 (IMT-2000) initiative has defined the ITU's view of third-generation capabilities as follows:

- Voice quality comparable to the public switched telephone network
- 144 kbps data rate available to users in high-speed motor vehicles over large areas
- 384 kbps available to pedestrians standing or moving slowly over small areas
- Support (to be phased in) for 2.048 Mbps for office use
- Symmetrical and asymmetrical data transmission rates
- Support for both packet switched and circuit switched data services
- An adaptive interface to the Internet to reflect efficiently the common asymmetry between inbound and outbound traffic
- More efficient use of the available spectrum in general
- Support for a wide variety of mobile equipment
- Flexibility to allow the introduction of new services and technologies

More generally, one of the driving forces of modern communication technology is the trend toward universal personal telecommunications and universal communications access. The first concept refers to the ability of a person to identify himself or herself easily and use conveniently any communication system in an entire country, over a continent, or even globally, in terms of a single account. The second refers to the capability of using one's terminal in a wide variety of environments to connect to information services (e.g., to have a portable terminal that will work in the office, on the street, and on airplanes equally well). This revolution in personal computing will obviously involve wireless communication in a fundamental way. The GSM cellular telephony with its subscriber identity module, for example, is a large step toward these goals.

Personal communications services (PCSs) and personal communication networks (PCNs) are names attached to these concepts of global wireless communications, and they also form objectives for third-generation wireless.

Generally, the technology planned is digital using time division multiple access or code-division multiple access to provide efficient use of the spectrum and high capacity.

PCS handsets are designed to be low power and relatively small and light. Efforts are being made internationally to allow the same terminals to be used worldwide.

Alternative Interfaces

Figure 10.22 shows the alternative schemes that have been adopted as part of IMT-2000. The specification covers a set of radio interfaces for optimized performance in different radio environments. A major reason for the inclusion of five alternatives was to enable a smooth evolution from existing first- and second-generation systems.

The five alternatives reflect the evolution from the second-generation. Two of the specifications grow out of the work at the European Telecommunications Standards Institute (ETSI) to develop a UMTS (universal mobile telecommunications system) as Europe's 3G wireless standard. UMTS includes two standards. One of

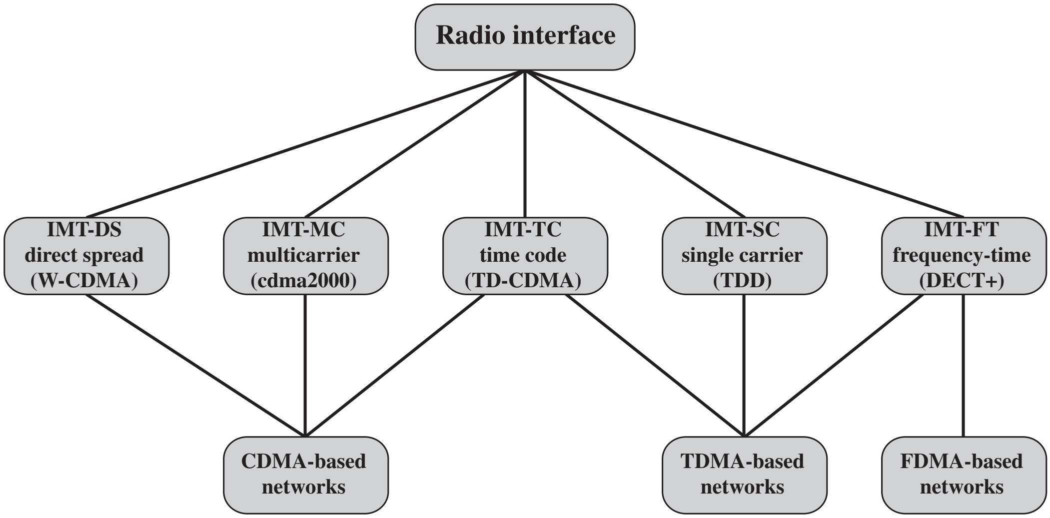


Figure 10.22 IMT-2000 Terrestrial Radio Interfaces

these is known as Wideband CDMA, or W-CDMA. This scheme fully exploits CDMA technology to provide high data rates with efficient use of bandwidth. Table 10.8 shows some of the key parameters of W-CDMA. The other European effort under UMTS is known as IMT-TC, or TD-CDMA. This approach is a combination of W-CDMA and TDMA technology. IMT-TC is intended to provide an upgrade path for the TDMA-based GSM systems.

Another CDMA-based system, known as cdma2000, has a North American origin. This scheme is similar to, but incompatible with, W-CDMA, in part because the standards use different chip rates. Also, cdma2000 uses a technique known as multicarrier, not used with W-CDMA.

We have looked at some of the technologies that appear in these 3G systems elsewhere in the book. These include turbo codes (Section 8.3), and Gold codes and variable-length spreading codes (Section 7.5).

Two other interface specifications are shown in Figure 10.22. IMT-SC is primarily designed for TDMA-only networks. IMT-FC can be used by both TDMA and FDMA carriers to provide some 3G services; it is an outgrowth of the Digital European Cordless Telecommunications (DECT) standard, discussed in Chapter 11.

CDMA Design Considerations

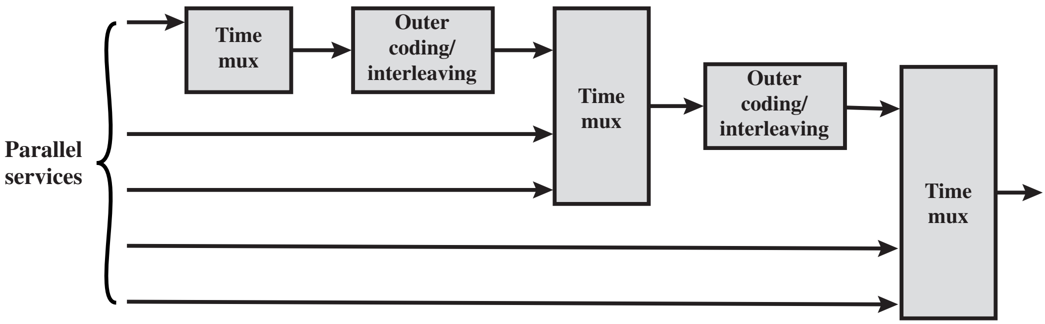
The dominant technology for 3G systems is CDMA. Although three different CDMA schemes have been adopted, they share some common design issues. [OJAN98] lists the following:

- **Bandwidth:** An important design goal for all 3G systems is to limit channel usage to 5 MHz. There are several reasons for this goal. On the one hand, a bandwidth of 5 MHz or more improves the receiver's ability to resolve multipath when compared to narrower bandwidths. On the other hand, available spectrum is limited by competing needs, and 5 MHz is a reasonable upper limit on what can be allocated for 3G. Finally, 5 MHz is adequate for supporting data rates of 144 and 384 kHz, the main targets for 3G services.

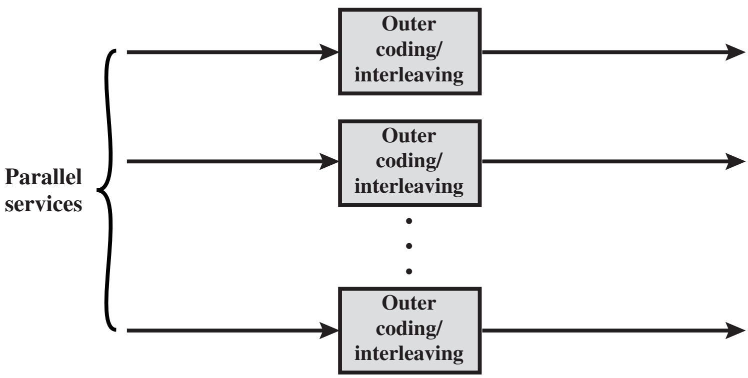
- **Chip rate:** Given the bandwidth, the chip rate depends on desired data rate, the need for error control, and bandwidth limitations. A chip rate of 3 Mcps or more is reasonable given these design parameters.
- **Multirate:** The term *multirate* refers to the provision of multiple fixed-data-rate logical channels to a given user, in which different data rates are provided on different logical channels. Further, the traffic on each logical channel can be switched independently through the wireless and fixed networks to different destinations. The advantage of multirate is that the system can flexibly support multiple simultaneous applications from a given user and can efficiently use available capacity by only providing the capacity required for each service.

Table 10.8 W-CDMA Parameters

Channel bandwidth	5 MHz
Forward RF channel structure	Direct spread
Chip rate	3.84 Mcps
Frame length	10 ms
Number of slots/frame	15
Spreading modulation	Balanced QPSK (forward) Dual channel QPSK (reverse) Complex spreading circuit
Data modulation	QPSK (forward) BPSK (reverse)
Coherent detection	Pilot symbols
Reverse channel multiplexing	Control and pilot channel time multiplexed. I and Q multiplexing for data and control channels
Multirate	Various spreading and multicode
Spreading factors	4 to 256
Power control	Open and fast closed loop (1.6 kHz)
Spreading (forward)	Variable length orthogonal sequences for channel separation. Gold sequences 2^{18} for cell and user separation.
Spreading (reverse)	Same as forward, different time shifts in I and Q channels
Handover	Soft handover



(a) Time multiplexing



(b) Code multiplexing

Figure 10.23 Time and Code Multiplexing Principles [OJAN98]

Multirate can be achieved with a TDMA scheme within a single CDMA channel, in which a different number of slots per frame are assigned to achieve different data rates. All the subchannels at a given data rate would be protected by error correction and interleaving techniques (Figure 10.23a). An alternative is to use multiple CDMA codes, with separate coding and interleaving, and map them to separate CDMA channels (Figure 10.23b).

10.6 RECOMMENDED READING AND WEB SITES

[BERT94] and [ANDE95] are instructive surveys of cellular wireless propagation effects. [POLL96] covers the handoff problem in depth. [EVER94] and [ORLI98] provide good accounts of cellular traffic analysis. [BLAC99] is one of the best technical treatments of second-generation cellular systems. A good survey of GSM concepts is [RAHN93]; for more detail see [GARG99].

[TANT98] contains reprints of numerous important papers dealing with CDMA in cellular networks. [DINA98] provides an overview of both PN and orthogonal spreading codes for cellular CDMA networks.

[OJAN98] provides an overview of key technical design considerations for 3G systems. Another useful survey is [ZENG00]. [PRAS00] is a much more detailed analysis.

ANDE95 Anderson, J.; Rappaport, T.; and Yoshida, S. "Propagation Measurements and Models for Wireless Communications Channels." *IEEE Communications Magazine*, January 1995.

BERT94 Bertoni, H.; Honcharenko, W.; Maciel, L.; and Xia, H. "UHF Propagation Prediction for Wireless Personal Communications." *Proceedings of the IEEE*, September 1994.

BLAC99 Black, U. *Second-generation Mobile and Wireless Networks.* Upper Saddle River, NJ: Prentice Hall, 1999.

DINA98 Dinan, E., and Jabbari, B. "Spreading Codes for Direct Sequence CDMA and Wideband CDMA Cellular Networks." *IEEE Communications Magazine*, September 1998.

EVER94 Everitt, D. "Traffic Engineering of the Radio Interface for Cellular Mobile Networks." *Proceedings of the IEEE*, September 1994.

GARG99 Garg, V., and Wilkes, J. *Principles and Applications of GSM.* Upper Saddle River, NJ: Prentice Hall, 1999.

OJAN98 Ojanpera, T., and Prasad, G. "An Overview of Air Interface Multiple Access for IMT-2000/UMTS." *IEEE Communications Magazine*, September 1998.

ORLI98 Orlik, P., and Rappaport, S. "Traffic Performance and Mobility Modeling of Cellular Communications with Mixed Platforms and Highly Variable Mobilities." *Proceedings of the IEEE*, July 1998.

POLL96 Pollini, G. "Trends in Handover Design." *IEEE Communications Magazine*, March 1996.

PRAS00 Prasad, R.; Mohr, W.; and Konhauser, W., eds. *Third-Generation Mobile Communication Systems.* Boston: Artech House, 2000.

RAHN93 Rahnema, M. "Overview of the GSM System and Protocol Architecture." *IEEE Communications Magazine*, April 1993.

TANT98 Tantaratana, S, and Ahmed, K., eds. *Wireless Applications of Spread Spectrum Systems: Selected Readings.* Piscataway, NJ: IEEE Press, 1998.

ZENG00 Zeng, M.; Annamalai, A.; and Bhargava, V. "Harmonization of Global Third-generation Mobile Systems. *IEEE Communications Magazine*, December 2000.

Recommended Web sites:

- **Cellular Telecommunications and Internet Association:** An industry consortium that provides information on successful applications of wireless technology
- **GSM World:** Lots of information and links concerning GSM
- **CDMA Development Group:** Information and links for IS-95 and CDMA generally
- **WOW-com:** Site of the Cellular Telecommunications Industry Association

10.7 KEY TERMS, REVIEW QUESTIONS, AND PROBLEMS

Key Terms

Advanced Mobil Phone Service (AMPS)	forward channel	power control
base station	frequency reuse	reuse factor
blocking network	Global System for Mobile Communications (GSM)	reverse channel
cellular network	handoff	second-generation (2G) network
closed-loop power control	handover	soft handoff
code division multiple access (CDMA)	hard handoff	third-generation (3G) network
first-generation (1G) network	mobile radio	time division multiple access (TDMA)
	nonblocking network	
	open-loop power control	

Review Questions

1 What geometric shape is used in cellular system design?
2 What is the principle of frequency reuse in the context of a cellular network?
3 List five ways of increasing the capacity of a cellular system.
4 Explain the paging function of a cellular system.
5 List and briefly define different performance metrics that may be used to make the handoff decision.
6 As a mobile unit in communication with base station moves, what factors determine the need for power control and the amount of power adjustment?
7 Explain the difference between open-loop and closed-loop power control.
8 What is the difference between traffic intensity and the mean rate of calls in a system?
9 What are the key differences between first- and second-generation cellular systems?
10 What are the advantages of using CDMA for a cellular network?
11 What are the disadvantages of using CDMA for a cellular network?
12 Explain the difference between hard and soft handoff.
13 What are some key characteristics that distinguish third-generation cellular systems from second-generation cellular systems?

Problems

1 Consider four different cellular systems that share the following characteristics. The frequency bands are 825 to 845 MHz for mobile unit transmission and 870 to 890 MHz for base station transmission. A duplex circuit consists of one 30-kHz channel in each direction. The systems are distinguished by the reuse factor, which is 4, 7, 12, and 19, respectively.

a. Suppose that in each of the systems, the cluster of cells (4, 7, 12, 19) is duplicated 16 times. Find the number of simultaneous communications that can be supported by each system.

b. Find the number of simultaneous communications that can be supported by a single cell in each system.

c. What is the area covered, in cells, by each system?

d. Suppose the cell size is the same in all four systems and a fixed area of 100 cells is covered by each system. Find the number of simultaneous communications that can be supported by each system.

2 Describe a sequence of events similar to that of Figure 10.6 for

a. a call from a mobile unit to a fixed subscriber.

b. a call from a fixed subscriber to a mobile unit.

3 In the discussion of the handoff procedure based on relative signal strength with threshold, it was pointed out that if the threshold is set quite low, such as Th_3, the mobile may move far into the new cell (L_4). This reduces the quality of the communication link and may result in a dropped call. Can you suggest another drawback to this scheme?

4 Hysteresis is a technique commonly used in control systems. As an example, describe the hysteresis mechanism used in a household thermostat.

5 A telephony connection has a duration of 23 minutes. This is the only connection made by this caller during the course of an hour. How much is the amount of traffic, in Erlangs, of this connection?

6 Using Table 10.3, approximate the answers to the following. Also, in each case, give a description in words of the general problem being solved. *Hint:* Straight-line interpolation is adequate.

a. Given $N = 20$, $A = 10.5$, find P.

b. Given $N = 20$, $P = 0.015$, find A.

c. Given $P = 0.005$, $A = 6$, find N.

7 An analog cellular system has a total of 33 MHz of bandwidth and uses two 25-kHz simplex (one-way) channels to provide full duplex voice and control channels.

a. What is the number of channels available per cell for a frequency reuse factor of (1) 4 cells, (2) 7 cells, and (3) 12 cells?

b. Assume that 1 MHz is dedicated to control channels but that only one control channel is needed per cell. Determine a reasonable distribution of control channels and voice channels in each cell for the three frequency reuse factors of part (a).

8 As was mentioned, the one-way bandwidth available to a single operator in the AMPS system is 12.5 MHz with a channel bandwidth of 30 kHz and 21 control channels. We would like to calculate the efficiency with which this system utilizes bandwidth for a particular installation. Use the following parameters:

- Cell area = 8 km^2
- Total coverage area = 4000 km^2
- Frequency reuse factor = 7
- Average number of calls per user during the busy hour = 1.2
- Average holding time of a call = 100 s
- Call blocking probability = 2%

a. How many voice channels are there per cell?

b. Use Table 10.3 and a simple straight-line interpolation to determine the total traffic carried per cell, in Erlangs/cell. Then convert that to Erlangs/km^2.

c. Calculate the number of calls/hour/cell and the number of calls/hour/km^2.

d. Calculate the number of users/hour/cell and the number of users/hour/channel.

e. A common definition of spectral efficiency with respect to modulation, or modulation efficiency, in Erlangs/MHz/km^2, is

$$\eta = \frac{(\text{Total traffic carried by the system})}{(\text{Bandwidth})(\text{Total coverage area})}$$

Determine the modulation efficiency for this system.

9 A cellular system uses FDMA with a spectrum allocation of 12.5 MHz in each direction, a guard band at the edge of the allocated spectrum of 10 kHz, and a channel bandwidth of 30 kHz. What is the number of available channels?

10 If 8 speech channels are supported on a single radio channel, and if no guard band is assumed, what is the number of simultaneous users that can be accommodated in GSM?

11 **a.** What is the duration of a bit in GSM?

b. If a user is allocated one time slot per frame, what is the delay between successive transmissions in successive frames?

12 If we consider the trailing bits, stealing bits, guard bits, and training bits in a GSM frame as overhead, and the rest of the bits as data, then what is the percentage overhead in a GSM frame?

13 Using the definition of slow frequency hopping from Chapter 7, demonstrate that GSM uses slow frequency hopping.

14 For a cellular system, FDMA spectral efficiency is defined as $\eta_a = \dfrac{B_c N_T}{B_w}$, where

B_c = channel bandwidth
B_w = total bandwidth in one direction
N_T = total number of voice channels in the covered area

a. What is an upper bound on η_a?

b. Determine η_a for the system of Problem 8.

15 Consider a 7-cell system covering an area of 3100 km^2. The traffic in the seven cells is as follows:

Cell number	1	2	3	4	5	6	7
Traffic (Erlangs)	30.8	66.7	48.6	33.2	38.2	37.8	32.6

Each user generates an average of 0.03 Erlangs of traffic per hour, with a mean holding time of 120 s. The system consists of a total of 395 channels and is designed for a grade of service of 0.02.

a. Determine the number of subscribers in each cell.

b. Determine the number of calls per hour per subscriber.

c. Determine the number of calls per hour in each cell.

d. Determine the number of channels required in each cell. *Hint:* You will need to extrapolate using Table 10.3.

e. Determine the total number of subscribers.

f. Determine the average number of subscribers per channel.

g. Determine the subscriber density per km^2.

h. Determine the total traffic (total Erlangs).

i. Determine the Erlangs per km^2.

j. What is the radius of a cell?