CHAPTER

4 Cellular systems: multiple access and interference management

4.1 Introduction

In Chapter 3, our focus was on *point-to-point* communication, i.e., the scenario of a single transmitter and a single receiver. In this chapter, we turn to a *network* of many mobile users interested in communicating with a common wireline network infrastructure.[1] This form of wireless communication is different from radio or TV in two important respects: first, users are interested in messages specific to them as opposed to the common message that is broadcast in radio and TV. Second, there is two-way communication between the users and the network. In particular, this allows feedback from the receiver to the transmitter, which is missing in radio and TV. This form of communication is also different from the all-wireless walkie-talkie communication since an access to a wireline network infrastructure is demanded. *Cellular systems* address such a multiuser communication scenario and form the focus of this chapter.

Broadly speaking, two types of spectra are available for commercial cellular systems. The first is *licensed*, typically nationwide and over a period of a few years, from the spectrum regulatory agency (FCC, in the United States). The second is unlicensed spectrum made available for experimental systems and to aid development of new wireless technologies. While licensing spectrum provides immunity from any kind of interference outside of the system itself, bandwidth is very expensive. This skews the engineering design of the wireless system to be as spectrally efficient as possible. There are no hard constraints on the power transmitted within the licensed spectrum but the power is expected to decay rapidly outside. On the other hand, unlicensed spectrum is very cheap to transmit on (and correspondingly larger

---

[1] A common example of such a network (wireline, albeit) is the public switched telephone network.

than licensed spectrum) but there is a maximum power constraint over the entire spectrum as well as interference to deal with. The emphasis thus is less on spectral efficiency. The engineering design can thus be very different depending on whether the spectrum is licensed or not. In this chapter, we focus on cellular systems that are designed to work on licensed spectrum. Such cellular systems have been deployed nationwide and one of the driving factors for the use of licensed spectrum for such networks is the risk of huge capital investment if one has to deal with malicious interference, as would be the case in unlicensed bands.

A cellular network consists of a number of fixed base-stations, one for each *cell*. The total coverage area is divided into cells and a mobile communicates with the base-station(s) close to it. (See Figure 1.2.) At the physical and medium access layers, there are two main issues in cellular communication: *multiple access* and *interference management*. The first issue addresses how the overall resource (time, frequency, and space) of the system is shared by the users in the same cell (intra-cell) and the second issue addresses the interference caused by simultaneous signal transmissions in different cells (inter-cell). At the network layer, an important issue is that of seamless connectivity to the mobile as it moves from one cell to the other (and thus switching communication from one base-station to the other, an operation known as *handoff*). In this chapter we will focus primarily on the physical-layer issues of multiple access and interference management, although we will see that in some instances these issues are also coupled with how handoff is done.

In addition to resource sharing between different users, there is also an issue of how the resource is allocated between the *uplink* (the communication from the mobile users to the base-station, also called the *reverse link*) and the *downlink* (the communication from the base-station to the mobile users, also called the *forward link*). There are two natural strategies for separating resources between the uplink and the downlink: *time division duplex* (TDD) separates the transmissions in time and *frequency division duplex* (FDD) achieves the separation in frequency. Most commercial cellular systems are based on FDD. Since the powers of the transmitted and received signals typically differ by more than 100 dB at the transmitter, the signals in each direction occupy bands that are separated far apart (tens of MHz), and a device called a *duplexer* is required to filter out any interference between the two bands.

A cellular network provides coverage of the entire area by dividing it into cells. We can carry this idea further by dividing each cell spatially. This is called *sectorization* and involves dividing the cell into, say three, *sectors*. Figure 4.1 shows such a division of a hexagonal cell. One way to think about sectors is to consider them as separate cells, except that the base-station corresponding to the sectors is at the same location. Sectorization is achieved by having a *directional antenna* at the base-station that focuses transmissions

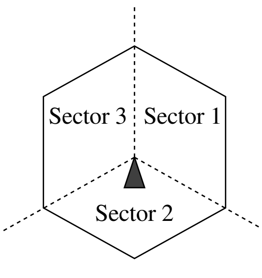

Figure 4.1 A hexagonal cell with three sectors.

into the sector of interest, and is designed to have a null in the other sectors. The ideal end result is an effective creation of new cells without the added burden of new base-stations and network infrastructure. Sectorization is most effective when the base-station is quite tall with few obstacles surrounding it. Even in this ideal situation, there is inter-sector interference. On the other hand, if there is substantial local scattering around the base-station, as is the case when the base-stations are low-lying (such as on the top of lamp posts), sectorization is far less effective because the scattering and reflection would transfer energy to sectors other than the one intended. We will discuss the impact of sectorization on the choice of the system design.

In this chapter, we study three cellular system designs as case studies to illustrate several different approaches to multiple access and interference management. Both the uplink and the downlink designs will be studied. In the first system, which can be termed a *narrowband system*, user transmissions within a cell are restricted to separate narrowband channels. Further, neighboring cells use different narrowband channels for user transmissions. This requires that the total bandwidth be split and reduces the *frequency reuse* in the network. However, the network can now be simplified and approximated by a collection of point-to-point *non-interfering* links, and the physical-layer issues are essentially point-to-point ones. The IS-136 and GSM standards are prime examples of this system. Since the level of interference is kept minimal, the point-to-point links typically have high signal-to-interference-plus-noise ratios (SINRs).[2]

The second and third system designs propose a contrasting strategy: all transmissions are spread to the entire bandwidth and are hence *wideband*. The key feature of these systems is *universal frequency reuse*: the same spectrum is used in every cell. However, simultaneous transmissions can now interfere with each other and links typically operate at low SINRs. The two system designs differ in how the users' signals are spread. The code division multiple access (CDMA) system is based on direct-sequence spread-spectrum. Here, users' information bits are coded at a very low rate and modulated by pseudonoise sequences. In this system, the simultaneous transmissions, intra-cell and inter-cell, cause interference. The IS-95 standard is the main example to highlight the design features of this system. In the orthogonal frequency division multiplexing (OFDM) system, on the other hand, users' information is spread by hopping in the time–frequency grid. Here, the transmissions within a cell can be kept orthogonal but adjacent cells share the same bandwidth and inter-cell interference still exists. This system has the advantage of the full frequency reuse of CDMA while retaining the benefits of the narrowband system where there is no intra-cell interference.

---

[2] Since interference plays an important role in multiuser systems, SINR takes the place of the parameter SNR we used in Chapter 3 when we only talked about point-to-point communication.

We also study the power profiles of the signals transmitted in these systems. This study will be conducted for both the downlink and the uplink to obtain an understanding of the peak and average power profile of the transmissions. We conclude by detailing the impact on power amplifier settings and overall power consumption in the three systems.

Towards implementing the multiple access design, there is an overhead in terms of communicating certain parameters from the base-station to the mobiles and vice versa. They include: authentication of the mobile by the network, allocation of traffic channels, training data for channel measurement, transmit power level, and acknowledgement of correct reception of data. Some of these parameters are one-time communication for a mobile; others continue in time. The amount of overhead this constitutes depends to some extent on the design of the system itself. Our discussions include this topic only when a significant overhead is caused by a specific design choice.

The table at the end of the chapter summarizes the key properties of the three systems.

4.2 Narrowband cellular systems

In this section, we discuss a cellular system design that uses naturally the ideas of reliable point-to-point wireless communication towards constructing a wireless network. The basic idea is to schedule all transmissions so that no two simultaneous transmissions interfere with each other (for the most part). We describe an identical uplink and downlink design of multiple access and interference management that can be termed narrowband to signify that the user transmissions are restricted to a narrow frequency band and the main design goal is to minimize all interference.

Our description of the narrowband system is the same for the uplink and the downlink. The uplink and downlink transmissions are separated, either in time or frequency. For concreteness, let us consider the separation to be in frequency, implemented by adopting an FDD scheme which uses widely separated frequency bands for the two types of transmissions. A bandwidth of W Hz is allocated for the uplink as well as for the downlink. Transmissions of different users are scheduled to be non-overlapping in time and frequency thus eliminating intra-cell interference. Depending on how the overall resource (time and bandwidth) is split among transmissions to the users, the system performance and design implications of the receivers are affected.

We first divide the bandwidth into N narrowband chunks (also denoted as *channels*). Each narrowband channel has width W/N Hz. Each cell is allotted some n of these N channels. These n channels are not necessarily contiguous. The idea behind this allocation is that all transmissions within this cell (in both the uplink and the downlink) are restricted to those n channels. To prevent interference between simultaneous transmissions in neighboring

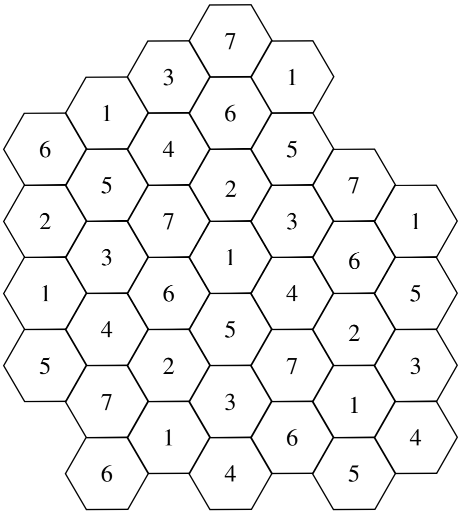

Figure 4.2 A hexagonal arrangements of cells and a possible reuse pattern of channels 1 through 7 with the condition that a channel cannot be used in one concentric ring of cells around the cell using it. The frequency reuse factor is 1/7.

cells, a channel is allocated to a cell only if it is not used by a few concentric rings of neighboring cells. Assuming a regular hexagonal cellular arrangement, Figure 4.2 depicts cells that can use the same channel simultaneously (such cells are denoted by the same number) if we want to avoid any neighboring cell from using the same channel.

The maximum number n of channels that a cell can be allocated depends on the geometry of the cellular arrangement and on the interference avoidance pattern that dictates which cells can share the same channel. The ratio n/N denotes how often a channel can be reused and is termed the *frequency reuse factor*. In the regular hexagonal model of Figure 4.2, for example, the frequency reuse factor is at least 1/7. In other words, $W/7$ is the effective bandwidth used by any base-station. This reduced spectral efficiency is the price paid up front towards satisfying the design goal of reducing all interference from neighboring base-stations. The specific reuse pattern in Figure 4.2 is ad hoc. A more careful analysis of the channel allocation to suit traffic conditions and the effect of reuse patterns among the cells is carried out in Exercises 4.1, 4.2, and 4.3.

Within a cell, different users are allocated transmissions that are non-overlapping, in both time and channels. The nature of this allocation affects various aspects of system design. To get a concrete feel for the issues involved, we treat one specific way of allocation that is used in the GSM system.

4.2.1 Narrowband allocations: GSM system

The GSM system has already been introduced in Example 3.1. Each narrowband channel has bandwidth 200 kHz (i.e. $W/N = 200\,\text{kHz}$). Time is divided into slots of length $T = 577\,\mu\text{s}$. The time slots in the different channels are the finest divisible resources allocated to the users. Over each slot, n simultaneous user transmissions are scheduled within a cell, one in each of the narrowband channels. To minimize the co-channel interference, these n channels have to be chosen as far apart in frequency as possible. Furthermore, each narrowband channel is shared among eight users in a time-division manner. Since voice is a fixed rate application with predictable traffic, each user is periodically allocated a slot out of every eight. Due to the nature of resource allocation (time and frequency), transmissions suffer no interference from within the cell and further see minimal interference from neighboring cells. Hence the network is stitched out of several point-to-point non-interfering wireless links with transmissions over a narrow frequency band, justifying our term "narrowband system" to denote this design paradigm.

Since the allocations are static, the issues of frequency and timing synchronization are the same as those faced by point-to-point wireless communication. The symmetric nature of voice traffic also enables a symmetric design of the uplink and the downlink. Due to the lack of interference, the operating received SINRs can be fairly large (up to 30 dB), and the communication scheme in both the uplink and the downlink is *coherent*. This involves learning the narrowband channel through the use of training symbols (or pilots), which are time-division multiplexed with the data in each slot.

Performance

What is the link reliability? Since the slot length T is fairly small, it is typically within the coherence time of the channel and there is not much time diversity. Further, the transmission is restricted to a contiguous bandwidth 200 kHz that is fairly narrow. In a typical outdoor scenario the delay spread is of the order of 1 μs and this translates to a coherence bandwidth of 500 kHz, significantly larger than the bandwidth of the channel. Thus there is not much frequency diversity either. The tough message of Chapter 3 that the error probability decays very slowly with the SNR is looming large in this scenario. As discussed in Example 3.1 of Chapter 3, GSM solves this problem by coding over eight consecutive time slots to extract a combination of time and frequency diversity (the latter via slow frequency hopping of the frames, each made up of the eight time slots of the users sharing a narrowband channel). Moreover, voice quality not only depends on the average frame error rate but also on how clustered the errors are. A cluster of errors leads to a far more noticeable quality degradation than independent frame errors even though the average frame error rate is the same in both the scenarios. Thus, the frequency hopping serves to break up the cluster of errors as well.

Signal characteristics and receiver design

The mobile user receives signals with energy concentrated in a contiguous, narrow bandwidth (of width (W/N), 200 kHz in the GSM standard). Hence the sample rate can be small and the sampling period is of the order of N/W

(5 μs in the GSM standard). All the signal processing operations are driven off this low rate, simplifying the implementation demands on the receiver design. While the sample rate is small, it might still be enough to resolve multipaths.

Let us consider the signals transmitted by a mobile and by the base-station. The *average* transmit power in the signal determines the performance of the communication scheme. On the other hand, certain devices in the RF chain that carry the transmit signal have to be designed for the *peak power* of the signal. In particular, the current bias setting of the power amplifier is directly proportional to the peak signal power. Typically class AB power amplifiers are used due to the linearity required by the spectrally efficient modulation schemes. Further, class AB amplifiers are very power inefficient and their cost (both capital cost and operating cost) is proportional to the bias setting (the range over which linearity is to be maintained). Thus an engineering constraint is to design transmit signals with reduced peak power for a given average power level. One way to capture this constraint is by studying the *peak to average power ratio* (PAPR) of the transmit signal. This constraint is particularly important in the mobile where power is a very scarce resource, as compared to the base-station.

Let us first turn to the signal transmitted by the mobile user (in the uplink). The signal over a slot is confined to a contiguous narrow frequency band (of width 200 kHz). In GSM, data is modulated on to this single-carrier using constant amplitude modulation schemes. In this context, the PAPR of the transmitted signal is fairly small (see Exercise 4.4), and is not much of a design issue. On the other hand, the signal transmitted from the base-station is a superposition of *n* such signals, one for each of the 200 kHz channels. The aggregate signal (when viewed in the time domain) has a larger PAPR, but the base-station is usually provided with an AC supply and power consumption is not as much of an issue as in the uplink. Further, the PAPR of the signal at the base-station is of the same order in most system designs.

4.2.2 Impact on network and system design

The specific division of resources here in conjunction with a static allocation among the users simplified the design complexities of multiple access and interference management in the network. There is however no free lunch. Two main types of price have to be paid in this design choice. The first is the physical-layer price of the inefficient use of the total bandwidth (measured through the frequency reuse factor). The second is the complexity of network planning. The orthogonal design entails a frequency division that has to be done up front in a global manner. This includes a careful study of the topology of the base-stations and shadowing conditions to arrive at acceptable interference from a base-station reusing one of the *N* channels. While Figure 4.2 demonstrated a rather simple setting with a suggestively simple design of reuse pattern, this study is quite involved in a real world system.

Further, the introduction of base-stations is done in an incremental way in real systems. Initially, enough base-stations to provide *coverage* are installed and new ones are added when the existing ones are overloaded. Any new base-station introduced in an area will require reconfiguring the assignment of channels to the base-stations in the neighborhood.

The nature of orthogonal allocations allows a high SINR link to most users, regardless of their location in the cell. Thus, the design is geared to allow the system to operate at about the same SINR levels for mobiles that are close to the base-stations as well as those that are at the edge of the cell. How does sectorization affect this design? Though sectored antennas are designed to isolate the transmissions of neighboring sectors, in practice, inter-sector interference is seen by the mobile users, particularly those at the edge of the sector. One implication of reusing the channels among the sectors of the same cell is that the dynamic range of SINR is reduced due to the intra-sector interference. This means that neighboring sectors cannot reuse the same channels while at the same time following the design principles of this system. To conclude, the gains of sectorization come not so much from frequency reuse as from an antenna gain and the improved capacity of the cell.

4.2.3 Impact on frequency reuse

How robust is this design towards allowing neighboring base-stations to reuse the same set of channels? To answer this question, let us focus on a specific scenario. We consider the uplink of a base-station one of whose neighboring base-stations uses the same set of channels. To study the performance of the uplink with this added interference, let us assume that there are enough users so that all channels are in use. Over one slot, a user transmission interferes directly with another transmission in the neighboring cell that uses the same channel. A simple model for the SINR at the base-station over a slot for one particular user uplink transmission is the following:

$$\mathsf{SINR} = \frac{P|h|^2}{N_0 + I}.$$

The numerator is the received power at the base-station due to the user transmission of interest with P denoting the average received power and $|h|^2$ the fading channel gain (with unit mean). The denominator consists of the background noise (N_0) and an extra term due to the interference from the user in the neighboring cell. I denotes the interference and is modeled as a random variable with a mean typically smaller than P (say equal to $0.2P$). The interference from the neighboring cell is random due to two reasons. One of them is small-scale fading and the other is the physical location of the user in the other cell that is reusing the same channel. The mean of I represents the average interference caused, averaged over all locations from which it could originate and the channel variations. But due to the fact that the interfering user can be at a wide range of locations, the variance of I is quite high.

We see that the SINR is a random parameter leading to an undesirably poor performance. There is an appreciably high probability of unreliable transmission of even a small and fixed data rate in the frame. In Chapter 3, we focused on techniques that impart channel diversity to the system; for example, antenna diversity techniques make the channel less variable, improving performance. However, there is an important distinction in the variability of the SINR here that cannot be improved by the diversity techniques of Chapter 3. The randomness in the interference I due to the interferer's location is inherent in this system and remains. Due to this, we can conclude that narrowband systems are unsuitable for universal frequency reuse. To reduce the randomness in the SINR, we would really like the interference to be *averaged* over several simultaneous lower-powered transmissions from the neighboring cell instead of coming from one user only. This is one of the important underlying themes in the design of the next two systems that have universal frequency reuse.

Summary 4.1 Narrowband systems

Orthogonal narrowband channels are assigned to users within a cell.

Users in adjacent cells cannot be assigned the same channel due to the lack of interference averaging across users. This reduces the frequency reuse factor and leads to inefficient use of the total bandwidth.

The network is decomposed into a set of high SINR point-to-point links, simplifying the physical-layer design.

Frequency planning is complex, particularly when new cells have to be added.

4.3 Wideband systems: CDMA

In narrowband systems, users are assigned disjoint time-frequency slots within the cell, and users in adjacent cells are assigned different frequency bands. The network is decomposed into a set of point-to-point non-interfering links. In a code division multiple access (CDMA) system design, the multiple access and interference management strategies are different. Using the direct-sequence spread-spectrum technique briefly mentioned in Section 3.4.3, each user spreads its signal over the entire bandwidth, such that when demodulating any particular user's data, other users' signals appear as pseudo white noise.

Thus, not only all users in the same cell share all the time-frequency degrees of freedom, so do the users in different cells. Universal frequency reuse is a key property of CDMA systems.

Roughly, the design philosophy of CDMA systems can be broken down into two design goals:

- First, the interference seen by any user is made as similar to white Gaussian noise as possible, and the power of that interference is kept to a minimum level and as consistent as possible. This is achieved by:
  - Making the received signal of every user as random looking as possible, via modulating the coded bits onto a long pseudonoise sequence.
  - Tight *power control* among users within the same cell to ensure that the received power of each user is no more than the minimum level needed for demodulation. This is so that the interference from users closer to the base-station will not overwhelm users further away (the so-called *near–far problem*).
  - *Averaging* the interference of many geographically distributed users in nearby cells. This averaging not only makes the aggregate interference look Gaussian, but more importantly reduces the randomness of the interference level due to varying locations of the interferers, thus increasing link reliability. This is the key reason why universal frequency reuse is possible in a wideband system but impossible in a narrowband system.
- Assuming the first design goal is met, each user sees a point-to-point wideband fading channel with additive Gaussian noise. Diversity techniques introduced in Chapter 3, such as coding, time-interleaving, Rake combining and antenna diversity, can be employed to improve the reliability of these point-to-point links.

Thus, CDMA is different from narrowband system design in the sense that all users share all degrees of freedom and therefore interfere with each other: the system is *interference-limited* rather than *degree-of-freedom-limited*. On the other hand, it is similar in the sense that the design philosophy is still to decompose the network problem into a set of independent point-to-point links, only now each link sees both interference as well as the background thermal noise. We do not question this design philosophy here, but we will see that there are alternative approaches in later chapters. In this section, we confine ourselves to discussing the various components of a CDMA system in the quest to meet the two design goals. We use the IS-95 standard to discuss concretely the translation of the design goals into a real system.

Compared to the narrowband systems described in the previous section, CDMA has several potential benefits:

- *Universal frequency reuse* means that users in all cells get the full bandwidth or degrees of freedom of the system. In narrowband systems, the number of degrees of freedom per user is reduced by both the number of users sharing the resources within a cell as well as by the frequency-reuse factor. This increase in degrees of freedom per user of a CDMA system however comes at the expense of a lower signal-to-interference-plus-noise ratio (SINR) per degree of freedom of the individual links.

- Because the performance of a user depends only on the *aggregate* interference level, the CDMA approach automatically takes advantage of the source variability of users; if a user stops transmitting data, the total interference level automatically goes down and benefits all the other users. Assuming that users' activities are independent of each other, this provides a *statistical multiplexing* effect to enable the system to accommodate more users than would be possible if every user were transmitting continuously. Unlike narrowband systems, no explicit re-assignment of time or frequency slots is required.
- In a narrowband system, new users cannot be admitted into a network once the time–frequency slots run out. This imposes a *hard* capacity limit on the system. In contrast, increasing the number of users in a CDMA system increases the total level of interference. This allows a more graceful degradation on the performance of a system and provides a *soft* capacity limit on the system.
- Since all cells share a common spectrum, a user on the edge of a cell can receive or transmit signals to two or more base-stations to improve reception. This is called *soft handoff*, and is yet another diversity technique, but at the network level (sometimes called *macrodiversity*). It is an important mechanism to increase the capacity of CDMA systems.

In addition to these network benefits, there is a further link-level advantage over narrowband systems: every user in a CDMA experiences a wideband fading channel and can therefore exploit the inherent frequency diversity in the system. This is particularly important in a slow fading environment where there is a lack of time diversity. It significantly reduces the *fade margin* of the system (the increased SINR required to achieve the same error probability as in an AWGN channel).

On the cons side, it should be noted that the performance of CDMA systems depends crucially on accurate power control, as the channel attenuation of nearby and cell edge users can differ by many tens of dBs. This requires frequent feedback of power control information and incurs a significant overhead per active user. In contrast, tight power control is not necessary in narrowband systems, and power control is exercised mainly for reducing battery consumption rather than managing interference. Also, it is important in a CDMA system that there be sufficient averaging of out-of-cell interference. While this assumption is rather reasonable in the uplink because the interference comes from many weak users, it is more questionable in the downlink, where the interference comes from a few strong adjacent base-stations.[3]

---

[3] In fact, the downlink of IS-95 is the capacity limiting link.

A comprehensive capacity comparison between CDMA and narrowband systems depends on the specific coding schemes and power control strategies, the channel propagation models, the traffic characteristics and arrival patterns of the users, etc. and is beyond the scope of this book. Moreover, many of the advantages of CDMA outlined above are qualitative and can probably be achieved in the narrowband system, albeit with a more complex engineering design. We focus here on a qualitative discussion on the key features of a CDMA system, backed up by some simple analysis to gain some insights into these features. In Chapter 5, we look at a simplified cellular setting and apply some basic information theory to analyze the tradeoff between the increase in degrees of freedom and the increase in the level of interference due to universal frequency reuse.

In a CDMA system, users interact through the interference they cause each other. We discuss ways to manage that interference and analyze its effect on performance. For concreteness, we first focus on the uplink and then move on to the downlink. Even though there are many similarities in their design, there are several differences worth pointing out.

4.3.1 CDMA uplink

The general schematic of the uplink of a CDMA system with K users in the system is shown in Figure 4.3. A fraction of the K users are in the cell and the rest are outside the cell. The data of the kth user are encoded into two BPSK sequences[4] $\{a_k^{\mathrm{I}}[m]\}$ and $\{a_k^{\mathrm{Q}}[m]\}$, which we assume to have equal amplitude for all m. Each sequence is modulated by a pseudonoise sequence, so that the transmitted complex sequence is

$$x_k[m] = a_k^{\mathrm{I}}[m]s_k^{\mathrm{I}}[m] + ja_k^{\mathrm{Q}}[m]s_k^{\mathrm{Q}}[m], \qquad m = 1, 2, \ldots, \tag{4.1}$$

where $\{s_k^{\mathrm{I}}[m]\}$ and $\{s_k^{\mathrm{Q}}[m]\}$ are pseudonoise sequences taking values ± 1. Recall that m is called a *chip time*. Typically, the chip rate is much larger than the data rate.[5] Consequently, information bits are heavily coded and the coded sequences $\{a_k^{\mathrm{I}}[m]\}$ and $\{a_k^{\mathrm{Q}}[M]\}$ have a lot of redundancy. The transmitted sequence of user k goes through a discrete-time baseband equivalent multipath channel $h^{(k)}$ and is superimposed at the receiver:

$$y[m] = \sum_{k=1}^{K} \left(\sum_{\ell} h_\ell^{(k)}[m] x_k[m-\ell] \right) + w[m]. \tag{4.2}$$

The fading channels $\{h^{(k)}\}$ are assumed to be independent across users, in addition to the assumption of independence across taps made in Section 3.4.3.

---

[4] Since CDMA systems operate at very low SINR per degree of freedom, a binary modulation alphabet is always used.

[5] In IS-95, the chip rate is 1.2288 MHz and the data rate is 9.6 kbits/s or less.

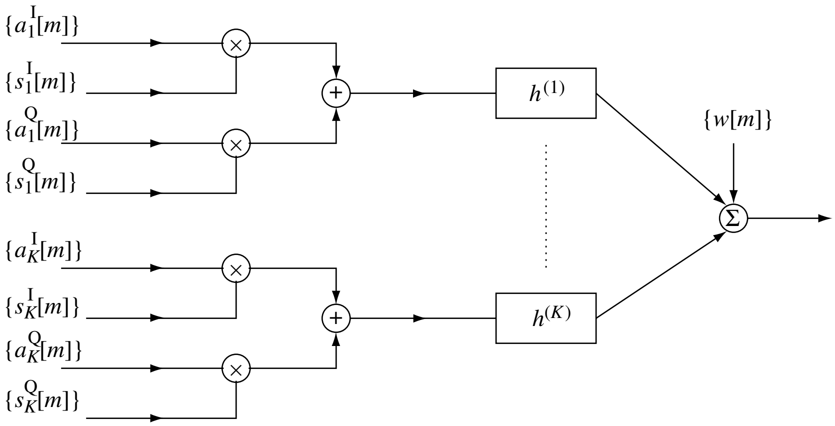

Figure 4.3 Schematic of the CDMA uplink.

The receiver for user k multiplies the I and Q components of the output sequence $\{y[m]\}$ by the pseudonoise sequences $\{s_k^{\mathrm{I}}[m]\}$ and $\{s_k^{\mathrm{Q}}[m]\}$ respectively to extract the coded streams of user k, which are then fed into a demodulator to recover the information bits. Note that in practice, the users' signals arrive asynchronously at the transmitter but we are making the idealistic assumption that users are *chip-synchronous*, so that the discrete-time model in Chapter 2 can be extended to the multiuser scenario here. Also, we are making the assumption that the receiver is already synchronized with each of the transmitters. In practice, there is a timing acquisition process by which such synchronization is achieved and maintained. Basically, it is a hypothesis testing problem, in which each hypothesis corresponds to a possible relative delay between the transmitter and the receiver. The challenge here is that because timing has to be accurate to the level of a chip, there are many hypotheses to consider and efficient search procedures are needed. Some of these procedures are detailed in Chapter 3 of [140].

Generation of pseudonoise sequences

The pseudonoise sequences are typically generated by *maximum length shift registers*. For a shift register of memory length r, the value of the sequence at time m is a linear function (in the binary field of $\{0, 1\}$) of the values at time $m-1, m-2, \ldots, m-r$ (its state). Thus, these binary 0–1 sequences are periodic, and the maximum period length is $p = 2^r - 1$, the number of non-zero states of the register.[6] This occurs when, starting from any non-zero state, the shift register goes through all possible $2^r - 1$ distinct non-zero states before returning to that state. Maximum length shift register (MLSR) sequences have this maximum periodic length, and they exist even for r very

---

[6] Starting from the zero state, the register will remain at the zero state, so the zero state cannot be part of such a period.

large. For CDMA applications, typically, r is somewhere between 20 and 50, thus the period is very long. Note that the generation of the sequence is a deterministic process, and the only randomness is in the initial state. An equivalent way to say this is that realizations of MLSR sequences are random shifts of each other.

The desired pseudonoise sequence $\{s[m]\}$ can be obtained from an MLSR sequence simply by mapping each value from 0 to $+1$ and from 1 to -1. This pseudonoise sequence has the following characteristics which make it look like a typical realization of a Bernoulli coin-flipped sequence ([52, 140]):

- $$\frac{1}{p}\sum_{m=1}^{p} s[m] = -\frac{1}{p}, \tag{4.3}$$

  i.e., the fraction of 0's and 1's is almost half-and-half over the period p.
- For all $\ell \neq 0$:

  $$\frac{1}{p}\sum_{m=1}^{p} s[m]s[m+\ell] = -\frac{1}{p}, \tag{4.4}$$

  i.e., the shifted versions of the pseudonoise sequence are nearly orthogonal to each other.

For memory $r = 2$, the period is 3 and the MLSR sequence is 110110110 … The states 11, 10, 01 appear in succession within each period. 00 does not appear, and this is the reason why the sum in (4.3) is not zero. However, this imbalance is very small when the period p is large.

If we randomize the shift of the pseudonoise sequence (i.e., uniformly chosen initial state of the shift register), then it becomes a random process. The above properties suggest that the resulting process is approximately like an i.i.d. Bernoulli sequence over a long time-scale (since p is very large). We will make this assumption below in our analysis of the statistics of the interference.

Statistics of the interference

In a CDMA system, the signal of one user is typically demodulated treating other users' signals as interference. The link level performance then depends on the statistics of the interference. Focusing on the demodulation of user 1, the aggregate interference it sees is

$$I[m] := \sum_{k>1}\left(\sum_{\ell} h_{\ell}^{(k)}[m]x_k[m-\ell]\right). \tag{4.5}$$

$\{I[m]\}$ has zero mean. Since the fading processes are circular symmetric, the process $\{I[m]\}$ is circular symmetric as well. The second-order statistics are then characterized by $\mathbb{E}[I[m]I[m+\ell]^*]$ for $\ell = 0, 1\ldots$ They can be computed as

$$\mathbb{E}[|I[m]|^2] = \sum_{k>1} \mathcal{E}_k^{\mathrm{c}}, \qquad \mathbb{E}[I[m]I[m+\ell]^*] = 0 \quad \text{for} \ell \neq 0, \tag{4.6}$$

where

$$\mathcal{E}_k^{\mathrm{c}} := \mathbb{E}[|x_k[m]|^2] \sum_{\ell} \mathbb{E}[|h_\ell^{(k)}[m]|^2] \tag{4.7}$$

is the total average energy received per chip from the kth user due to the multipath. In the above variance calculation, we make use of the fact that $\mathbb{E}[x_k[m]x_k[m+\ell]^*] = 0$ (for $\ell \neq 0$), due to the random nature of the spreading sequences. Note that in computing these statistics, we are averaging over both the data and the fading gains of the other users.

When there are many users in the network, and none of them contributes to a significant part of the interference, the Central Limit Theorem can be invoked to justify a Gaussian approximation of the interference process. From the second-order statistics, we see that this process is white. Hence, a reasonable approximation from the point of view of designing the point-to-point link for user 1 is to consider it as a multipath fading channel with white Gaussian noise of power $\sum_{k>1} \mathcal{E}_k^{\mathrm{c}} + N_0$.[7]

We have made the assumption that none of the users contributes a large part of the interference. This is a reasonable assumption due to two important mechanisms in a CDMA system:

- **Power control** The transmit powers of the users within the cell are controlled to solve the near–far problem, and this makes sure that there is no significant intra-cell interferer.
- **Soft handoff** Each base-station that receives a mobile's signal will attempt to decode its data and send them to the MSC (mobile switching center) together with some measure of the quality of the reception. The MSC will select the one with the highest quality of reception. Typically the user's power will be controlled by the base-station which has the best reception. This reduces the chance that some significant out-of-cell interferer is not power controlled.

We will discuss these two mechanisms in more detail later on.

Point-to-point link design

We have already discussed to some extent the design issues of the point-to-point link in a DS spread-spectrum system in Section 3.4.3. In the context

---

[7] This approach is by no means optimal, however. We will see in Chapter 6 that better performance can be achieved by recognizing that the interference consists of the data of the other users that can in fact be decoded.

of the CDMA system, the only difference here is that we are now facing the aggregation of both interference and noise.

The link level performance of user 1 depends on the SINR:

$$\mathsf{SINR}_\mathrm{c} := \frac{\mathcal{E}_1^\mathrm{c}}{\sum_{k>1} \mathcal{E}_k^\mathrm{c} + N_0}. \tag{4.8}$$

Note that this is the SINR *per chip*. The first observation is that typically the SINR per chip is very small. For example, if we consider a system with K perfectly power controlled users in the cell, even ignoring the out-of-cell interference and background noise, SINR_c is $1/(K-1)$. In a cell with 31 users, this is -15 dB. In IS-95, a typical level of out-of-cell interference is 0.6 of the interference from within the cell. (The background noise, on the other hand, is often negligible in CDMA systems, which are primarily interference-limited.) This reduces the SINR_c further to -17 dB.

How can we demodulate the transmitted signal at such low SINR? To see this in the simplest setting, let us consider an unfaded channel for user 1 and consider the simple example of BPSK modulation with coherent detection discussed in Section 3.4.3, where each information bit is modulated onto a pseudonoise sequence of length G chips. In the system discussed here which uses a *long* pseudonoise sequence $\{s[m]\}$ (cf. Figure 4.3), this corresponds to repeating every BPSK symbol G times, $a_1^I[Gi+m] = a_1^I[Gi]$, $m = 1, \ldots, G-1$.[8] The detection of the 0th information symbol is accomplished by projecting the in-phase component of the received signal onto the sequence $\mathbf{u} = [s_1^I[0], s_1^I[1], \ldots, s_1^I[G-1]]^t$, and the error probability is

$$p_\mathrm{e} = Q\left(\sqrt{\frac{2\|\mathbf{u}\|^2 \mathcal{E}_1^\mathrm{c}}{\sum_{k>1} \mathcal{E}_k^\mathrm{c} + N_0}}\right) = Q\left(\sqrt{\frac{2G\mathcal{E}_1^\mathrm{c}}{\sum_{k>1} \mathcal{E}_k^\mathrm{c} + N_0}}\right) = Q\left(\sqrt{\frac{2\mathcal{E}_\mathrm{b}}{\sum_{k>1} \mathcal{E}_k^\mathrm{c} + N_0}}\right) \tag{4.9}$$

where $\mathcal{E}_\mathrm{b} := G\mathcal{E}_1^\mathrm{c}$ is the received energy per bit for user 1. Thus, we see that while the SINR *per chip* is low, the SINR *per bit* is increased by a factor of G, due to the averaging of the noise in the G chips over which we repeat the information bits. In terms of system parameters, $G = W/R$, where W Hz is the bandwidth and R bits/s is the data rate. Recall that this parameter is called the *processing gain* of the system, and we see its role here as increasing the effective SINR against a large amount of interference that the user faces. As we scale up the size of a CDMA system by increasing the bandwidth W and the number of users in the system proportionally, but keeping the data rate of each user R fixed, we see that the total interference $\sum_{k>1} \mathcal{E}_k^\mathrm{c}$ and the

---

[8] As mentioned, a pseudonoise sequence typically has a period ranging from 2^{20} to 2^{50} chips, much larger than the processing gain G. In contrast, *short* pseudonoise sequences are used in the IS-95 downlink to uniquely identify the individual sector or cell.

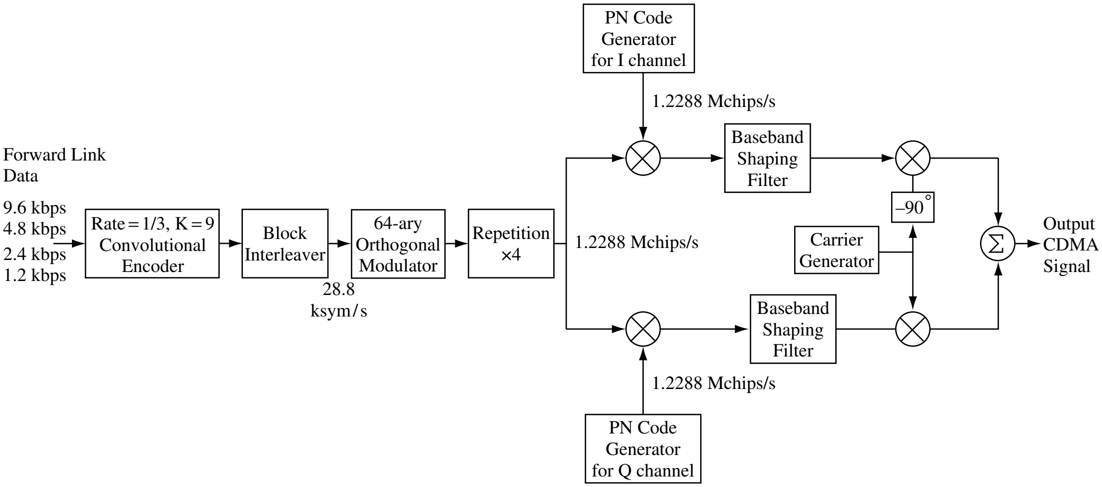

Figure 4.4 The IS-95 uplink.

processing gain G increase proportionally as well. This means that CDMA is an inherently scalable multiple access scheme.[9]

IS-95 link design

The above scheme is based on repetition coding. By using more sophisticated low-rate codes, even better performance can be achieved. Moreover, in practice the actual channel is a multipath fading channel, and so techniques such as time-interleaving and the Rake receiver are important to obtain time and frequency diversity respectively. IS-95, for example, uses a combination of convolutional coding, interleaving and non-coherent demodulation of M-ary orthogonal symbols via a Rake receiver. (See Figure 4.4.) Compressed voice at rate 9.6 kbits/s is encoded using a rate 1/3, constraint length 9, convolutional code. The coded bits are time-interleaved at the level of 6-bit blocks, and each of these blocks is mapped into one of $2^6 = 64$ orthogonal Hadamard sequences,[10] each of length 64. Finally, each symbol of the Hadamard sequence is repeated four times to form the coded sequence $\{a^{\mathrm{I}}[m]\}$. The processing gain is seen to be $3 \cdot 64/6 \cdot 4 = 128$, with a resulting chip rate of $128 \cdot 9.6 = 1.2288$ Mchips/s.

Each of the 6-bit blocks is demodulated non-coherently using a Rake receiver. In the binary orthogonal modulation example in Section 3.5.1, for each orthogonal sequence the non-coherent detector computes the correlation

---

[9] But note that as the bandwidth gets wider and wider, channel uncertainty may eventually become the bottleneck, as we have seen in Section 3.5.

[10] The Hadamard sequences of length $M = 2^J$ are the orthogonal columns of the M by M matrix $\mathbf{H}_M$, defined recursively as $\mathbf{H}_1 = [1]$ and for $M \geq 2$:

$$\mathbf{H}_M = \begin{bmatrix} \mathbf{H}_{M/2} & \mathbf{H}_{M/2} \\ \mathbf{H}_{M/2} & -\mathbf{H}_{M/2} \end{bmatrix}.$$

along each diversity branch (finger) and then forms the sum of the squares. It then decides in favor of the sequence with the largest sum (the square-law detector). (Recall the discussion around (3.147).) Here, each 6-bit block should be thought of as a coded symbol of an outer convolutional code, and we are not interested in hard decision of the block. Instead, we would like to calculate the branch metric for each of the possible values of the 6-bit block, for use by a Viterbi decoder for the outer convolutional code. It happens that the sum of the squares above can be used as a metric, so that the Rake receiver structure can be used for this purpose as well. It should be noted that it is important that the time-interleaving be done at the level of the 6-bit blocks so that the channel remains constant within the chips associated with each such block. Otherwise non-coherent demodulation cannot be performed.

The IS-95 uplink design employs non-coherent demodulation. Another design option is to estimate the channel using a pilot signal and perform coherent demodulation. This option is adopted for CDMA 2000.

Power control

The link-level performance of a user is a function of its SINR. To achieve reliable communication, the SINR, or equivalently the ratio of the energy per bit to the interference and noise per chip (commonly called $\mathcal{E}_\mathrm{b}/I_0$ in the CDMA literature), should be above a certain threshold. This threshold depends on the specific code used, as well as the multipath channel statistics. For example, a typical $\mathcal{E}_\mathrm{b}/I_0$ threshold in the IS-95 system is 6 to 7 dB. In a mobile communication system, the attenuation of both the user of interest and the interferers varies as the users move, due to varying path loss and shadowing effects. To maintain a target SINR, *transmit power control* is needed.

The power control problem can be formulated in the network setting as follows. There are K users in total in the system and a number of cells (base-stations). Suppose user k is assigned to base-station c_k. Let P_k be the transmit power of user k, and g_{km} be the attenuation of user k's signal to base-station m.

The received energy per chip for user k at base-station m is simply given by $P_k g_{km}/W$. Using the expression (4.8), we see that if each user's target $\mathcal{E}_\mathrm{b}/I_0$ is β, then the transmit powers of the users should be controlled such that

$$\frac{GP_k g_{k,c_k}}{\sum_{n\neq k} P_n g_{n,c_k} + N_0 W} \geq \beta, \qquad k = 1, \ldots, K. \tag{4.10}$$

where $G = W/R$ is the processing gain of the system. Moreover, due to constraints on the dynamic range of the transmitting mobiles, there is a limit of the transmit powers as well:

$$P_k \leq \hat{P}, \qquad k = 1, \ldots, K. \tag{4.11}$$

These inequalities define the set of all feasible power vectors $\mathbf{P} := (P_1, \ldots, P_K)^t$, and this set is a function of the attenuation of the users. If this set is empty, then the SINR requirements of the users cannot be simultaneously met. The system is said to be in *outage*. On the other hand, whenever this set of feasible powers is non-empty, one is interested in finding a solution which requires as little power as possible to conserve energy. In fact, it can be shown (Exercise 4.8) that whenever the feasible set is non-empty (this characterization is carried out carefully in Exercise 4.5), there exists a *component-wise* minimal solution $\mathbf{P}^*$ in the feasible set, i.e., $P_k^* \leq P_k$ for *every* user k in any other feasible power vector $\mathbf{P}$. This fact follows from a basic monotonicity property of the power control problem: when a user lowers its transmit power, it creates less interference and benefits all other users in the system. At the optimal solution $\mathbf{P}^*$, every user is at the minimal possible power so that their SINR requirements are met with equality and no more. Note that at the optimal point all the users in the same cell have the same received power at the base-station. It can also be shown that a simple distributed power control algorithm will converge to the optimal solution: at each step, each user updates its transmit power so that its own SINR requirement is just met with the current level of the interference. Even if the updates are done asynchronously among the users, convergence is still guaranteed. These results give theoretical justification to the robustness and stability of the power control algorithms implemented in practice. (Exercise 4.12 studies the robustness of the power update algorithm to inaccuracies in controlling the received powers of all the mobiles to be exactly equal.)

Power control in IS-95

The actual power control in IS-95 has an open-loop and a closed-loop component. The open-loop sets the transmit power of the mobile user at roughly the right level by inference from the measurements of the *downlink* channel strength via a pilot signal. (In IS-95, there is a common pilot transmitted in the downlink to all the mobiles.) However, since IS-95 is implemented in the FDD mode, the uplink and downlink channel typically differ in carrier frequency of tens of MHz and are not identical. Thus, open-loop control is typically accurate only up to a few dB. Closed-loop control is needed to adjust the power more precisely.

The closed-loop power control operates at 800 Hz and involves 1 bit feedback from the base-station to the mobile, based on measured SINR values; the command is to increase (decrease) power by 1 dB if the measured SINR is below (above) a threshold. Since there is no pilot in the uplink in IS-95, the SINR is estimated in a decision-directed mode, based on the output of the Rake receiver. In addition to measurement errors, the accuracy of power control is also limited by the 1-bit quantization. Since the SINR threshold β for reliable communication depends on the multipath channel statistics and is therefore not known perfectly in advance, there is also an outer loop which

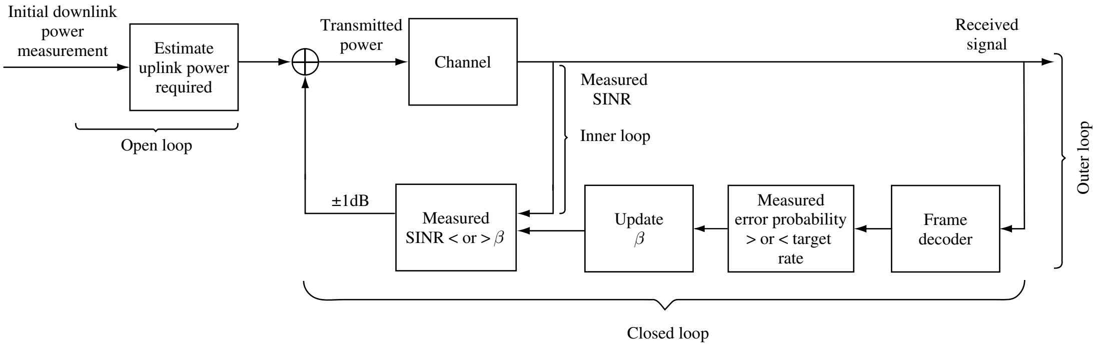

Figure 4.5 Inner and outer loops of power control.

adjusts the SINR threshold as a function of frame error rates (Figure 4.5). An important point, however, is that even though feedback occurs at a high rate (800 Hz), because of the limited resolution of 1 bit per feedback, power control does not track the fast multipath fading of the users when they are at vehicular speeds. It only tracks the slower shadow fading and varying path loss. The multipath fading is dealt with primarily by the diversity techniques discussed earlier.

Soft handoff

Handoff from one cell to the other is an important mechanism in cellular systems. Traditionally, handoffs are *hard*: users are either assigned to one cell or the other but not both. In CDMA systems, since all the cells share the same spectrum, *soft* handoffs are possible: multiple base-stations can simultaneously decode the mobile's data, with the switching center choosing

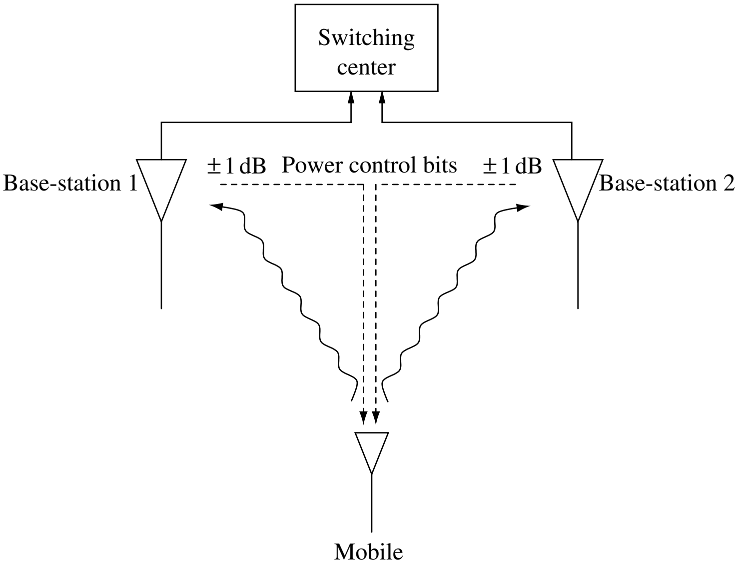

Figure 4.6 Soft handoff.

the best reception among them (Figure 4.6). Soft handoffs provide another level of diversity to the users.

The soft handoff process is mobile-initiated and works like this. While a user is tracking the downlink pilot of the cell it is currently in, it can be searching for pilots of adjacent cells (these pilots are known pseudonoise sequences shifted by known offsets). In general, this involves timing acquisition of the adjacent cell as well. However, we have observed that timing acquisition is a computationally very expensive step. Thus, a practical alternative is for the base-station clocks to be synchronized so that the mobile only has to acquire timing once. Once a pilot is detected and found to have sufficient signal strength relative to the first pilot, the mobile will signal the event to its original base-station. The original base-station will in turn notify the switching center, which enables the second cell's base-station to both send and receive the same traffic to and from the mobile. In the uplink, each base-station demodulates and decodes the frame or packet independently, and it is up to the switching center to arbitrate. Normally, the better cell's decision will be used.

If we view the base-stations as multiple receive antennas, soft handoff is providing a form of receive diversity. We know from Section 3.3.1 that the optimal processing of signals from the multiple antennas is *maximal-ratio combining*; this is however difficult to do in the handoff scenario as the antennas are geographically apart. Instead, what soft handoff achieves is *selection combining* (cf. Exercise 3.13). In IS-95, there is another form of handoff, called *softer handoff*, which takes place between sectors of the same cell. In this case, since the signal from the mobile is received at the sectored antennas which are co-located at the same base-station, maximal-ratio combining can be performed.

How does power control work in conjunction with soft handoff? Soft handoff essentially allows users to choose among several cell sites. In the power control formulation discussed in the previous section, each user is assumed to be assigned to a particular cell, but cell site selection can be easily incorporated in the framework. Suppose user k has an active set S_k of cells among which it is performing soft handoff. Then the transmit powers P_k *and* the cell site assignments $c_k \in S_k$ should be chosen such that the SINR requirements (4.10) are simultaneously met. Again, if there is a feasible solution, it can be shown that there is a component-wise minimal solution for the transmit powers (Exercise 4.5). Moreover, there is an analogous distributed asynchronous algorithm that will converge to the optimal solution: at each step, each user is assigned the cell site that will minimize the transmit power required to meet its SINR requirement, given the current interference levels at the base-stations. Its transmit power is set accordingly (Exercise 4.8). Put it another way, the transmit power is set in such a way that the SINR requirement is just met at the cell with the best reception. This is implemented in the IS-95 system as follows: all the base-stations in the soft handoff set will feedback power control bits to the mobile; the mobile will always decrease its transmit power by 1 dB if at least one of the soft handoff cell sites instructs it to do so. In other words, the minimum transmit power is always used. The advantages of soft handoff are studied in more detail in Exercise 4.10.

Interference averaging and system capacity

Power control and soft handoff minimize the transmit powers required to meet SINR requirements, if there is a feasible solution for the powers at all. If not, then the system is in outage. The *system capacity* is the maximum number of users that can be accommodated in the system for a desired outage probability and a link level $\mathcal{E}_b/I_0$ requirement.

The system can be in outage due to various random events. For example, users can be in certain configurations that create a lot of interference on neighboring cells. Also, voice or data users have periods of activity, and too many users can be active in the system at a given point in time. Another source of randomness is due to imperfect power control. While it is impossible to have a zero probability of outage, one wants to maintain that probability small, below a target threshold. Fortunately, the link level performance of a user in the uplink depends on the *aggregate* interference at the base-station due to many users, and the effect of these sources of randomness tends to average out according to the law of large numbers. This means that one does not have to be too conservative in admitting users into the network and still guarantee a small probability of outage. This translates into a larger system capacity. More specifically,

- **Out-of-cell interference averaging** Users tend to be in random independent locations in the network, and the fluctuations of the aggregate interference created in the adjacent cell are reduced when there are many users in the system.
- **Users' burstiness averaging** Independent users are unlikely to be active all the time, thus allowing the system to admit more users than if it is assumed that every user sends at peak rate all the time.
- **Imperfect power control averaging** Imperfect power control is due to tracking inaccuracy and errors in the feedback loop.[11] However, these errors tend to occur independently across the different users in the system and average out.

These phenomena can be generally termed *interference averaging*, an important property of CDMA systems. Note that the concept of interference averaging is reminiscent of the idea of *diversity* we discussed in Chapter 3: while diversity techniques make a point-to-point link more reliable by averaging over the channel fading, interference averaging makes the link more

---

[11] Since power control bits have to be fed back with a very tight delay constraint, they are usually uncoded which implies quite a high error rate.

reliable by averaging over the effects of different interferers. Thus, interference averaging can also be termed *interference diversity*.

To give a concrete sense of the benefit of interference averaging on system capacity, let us consider the specific example of averaging of users' burstiness. For simplicity, consider a single-cell situation with K users power controlled to a common base-station and no out-of-cell interference. Specializing (4.10) to this case, it can be seen that the $\mathcal{E}_b/I_0$ requirement β of all users is satisfied if

$$\frac{GQ_k}{\sum_{n\neq k} Q_n + N_0 W} \geq \beta, \qquad k = 1, \ldots, K, \tag{4.12}$$

where $Q_k := P_k g_k$ is the received power of user k at the base-station. Equivalently:

$$GQ_k \geq \beta \left(\sum_{n\neq k} Q_n + N_0 W \right) \qquad k = 1, \ldots, K. \tag{4.13}$$

Summing up all the inequalities, we get the following necessary condition for the Q_k:

$$[G - \beta(K-1)] \sum_{k=1}^{K} Q_k \geq K N_0 W \beta. \tag{4.14}$$

Thus a necessary condition for the existence of feasible powers is $G - \beta(K-1) > 0$, or equivalently,

$$K < \frac{G}{\beta} + 1. \tag{4.15}$$

On the other hand, if this condition is satisfied, the powers

$$Q_k = \frac{N_0 W \beta}{G - \beta(K-1)}, \qquad k = 1, \ldots, K \tag{4.16}$$

will meet the $\mathcal{E}_b/I_0$ requirements of all the users. Hence, condition (4.15) is a necessary and sufficient condition for the existence of feasible powers to support a given $\mathcal{E}_b/I_0$ requirement.

Equation (4.15) yields the *interference-limited system capacity* of the single cell. It says that, because of the interference between users, there is a limit on the number of users admissible in the cell. If we substitute $G = W/R$ into (4.15), we get

$$\boxed{\frac{KR}{W} < \frac{1}{\beta} + \frac{1}{G}.} \tag{4.17}$$

The quantity KR/W is the overall spectral efficiency of the system (in bits/s/Hz). Since the processing gain G of a CDMA system is typically large, (4.17) says that the maximal spectral efficiency is approximately $1/\beta$. In IS-95, a typical $\mathcal{E}_b/N_0$ requirement β is 6 dB, which translates into a maximum spectral efficiency of 0.25 bits/s/Hz.

Let us now illustrate the effect of user burstiness on the system capacity and the spectral efficiency in the single cell setting. We have assumed that all K users are active all the time, but suppose now that each user is active and has data to send only with probability p, and users' activities are independent of each other. Voice users, for example, are typically talking 3/8 of the time, and if the voice coder can detect silence, there is no need to send data during the quiet periods. If we let ν_k be the indicator random variable for user k's activity, i.e., $\nu_k = 1$ when user k is transmitting, and $\nu_k = 0$ otherwise, then using (4.15), the $\mathcal{E}_b/I_0$ requirements of the users can be met if and only if

$$\sum_{k=1}^{K} \nu_k < \frac{G}{\beta} + 1. \tag{4.18}$$

Whenever this constraint is not satisfied, the system is in outage. If the system wants to guarantee that no outage can occur, then the maximum number of users admissible in the network is $G/\beta + 1$, the same as the case when users are active all the time. However, more users can be accommodated if a small outage probability p_{out} can be tolerated: this number $K^*(p_{\text{out}})$ is the largest K such that

$$\Pr\left[\sum_{k=1}^{K} \nu_k > \frac{G}{\beta} + 1\right] \le p_{\text{out}}. \tag{4.19}$$

The random variable $\sum_{k=1}^{K} \nu_k$ is binomially distributed. It has mean Kp and standard deviation $\sqrt{Kp(1-p)}$, where $p(1-p)$ is the variance of ν_k. When $p_{\text{out}} = 0$, $K^*(p_{\text{out}})$ is $G/\beta + 1$. If $p_{\text{out}} > 0$, then $K^*(p_{\text{out}})$ can be chosen larger. It is straightforward to calculate $K^*(p_{\text{out}})$ numerically for a given p_{out}. It is also interesting to see what happens to the spectral efficiency when the bandwidth of the system W scales with the rate R of each user fixed. In this regime, there are many users in the system and it is reasonable to apply a Gaussian approximation to $\sum_{k=1}^{K} \nu_k$. Hence,

$$\Pr\left[\sum_{k=1}^{K} \nu_k > \frac{G}{\beta} + 1\right] \approx Q\left[\frac{G/\beta + 1 - Kp}{\sqrt{Kp(1-p)}}\right]. \tag{4.20}$$

The overall spectral efficiency of the system is given by

$$\rho := \frac{KpR}{W}, \tag{4.21}$$

since the mean rate of each user is pR bits/s. Using the approximation (4.20) in (4.19), we can solve for the constraint on the spectral efficiency ρ:

$$\boxed{\rho \leq \frac{1}{\beta}\left[1+Q^{-1}(p_{\text{out}})\sqrt{\frac{1-p}{pK}-\frac{1}{Kp}}\right]^{-1}.} \tag{4.22}$$

This bound on the spectral efficiency is plotted in Figure 4.7 as a function of the number of users. As seen in Eq. (4.17), the number $1/\beta$ is the maximum spectral efficiency if each user is non-bursty and transmitting at a constant rate equal to the mean rate pR of the bursty user. However, the *actual* spectral efficiency in the system with burstiness is different from that, by a factor of

$$\left(1+Q^{-1}(p_{\text{out}})\sqrt{\frac{1-p}{pK}-\frac{1}{Kp}}\right)^{-1}.$$

This loss in spectral efficiency is due to a need to admit fewer users to cater for the burstiness of the traffic. This “safety margin” is larger when the outage probability requirement p_{out} is more stringent. More importantly, for a given outage probability, the spectral efficiency approaches $1/\beta$ as the bandwidth W (and hence the number of users K) scales. When there are many users in the system, interference averaging occurs: the fluctuation of the aggregate interference is smaller relative to the mean interference level. Since the link level performance of the system depends on the aggregate interference, less excess resource needs to be set aside to accommodate the fluctuations. This is a manifestation of the familiar principle of *statistical multiplexing*.

In the above example, we have only considered a single cell, where each active user is assumed to be perfectly power controlled and the only source of interference fluctuation is due to the random number of active users. In a multicell setting, the level of interference from outside of the cell depends on the locations of the interfering users and this contributes to another source

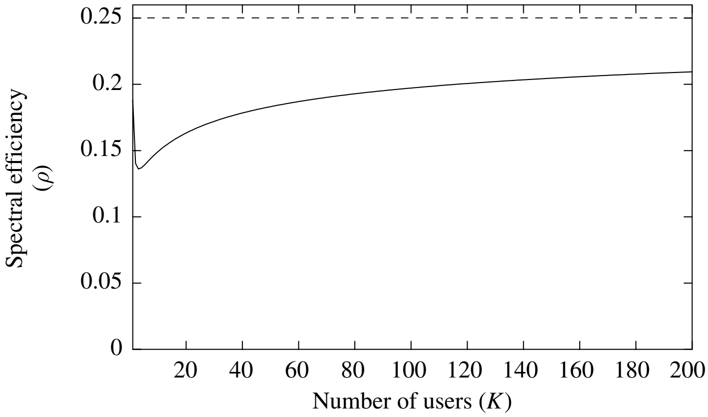

Figure 4.7 Plot of the spectral efficiency as a function of the number of users in a system with burstiness (the right hand side of (4.22)). Here, $p = 3/8$, $p_{\text{out}} = 0.01$ and $\beta = 6$ dB.

of fluctuation of the aggregate interference level. Further randomness arises due to imperfect power control. The same principle of interference averaging applies to these settings as well, allowing CDMA systems to benefit from an increase in the system size. These settings are analyzed in Exercises (4.11) and (4.12).

To conclude our discussion, we note that we have made an implicit assumption of *separation of time-scales* in our analysis of the effect of interference in CDMA systems. At a faster time-scale, we average over the pseudorandom characteristics of the signal and the fast multipath fading to compute the statistics of the interference, which determine the bit error rates of the point-to-point demodulators. At a slower time-scale, we consider the burstiness of user traffic and the large-scale motion of the users to determine the outage probability, i.e., the probability that the target bit error rate performance of users cannot be met. Since these error events occur at completely different time-scales and have very different ramifications from a system-level perspective, this way of measuring the performance of the system makes more sense than computing an overall average performance.

4.3.2 CDMA downlink

The design of the one-to-many downlink uses the same basic principles of pseudorandom spreading, diversity techniques, power control and soft handoff we already discussed for the uplink. However, there are several important differences:

- The near–far problem does not exist for the downlink, since all the signals transmitted from a base-station go through the same channel to reach any given user. Thus, power control is less crucial in the downlink than in the uplink. Rather, the problem becomes that of allocating different powers to different users as a function of primarily the amount of out-of-cell interference they see. However, the theoretical formulation of this power allocation problem has the same structure as the uplink power control problem. (See Exercise 4.13.)
- Since signals for the different users in the cell are all transmitted at the base-station, it is possible to make the users orthogonal to each other, something that is more difficult to do in the uplink, as it requires chip-level synchronization between distributed users. This reduces but does not remove intra-cell interference, since the transmitted signal goes through multipath channels and signals with different delays from different users still interfere with each other. Still, if there is a strong line-of sight component, this technique can significantly reduce the intra-cell interference, since then most of the energy is in the first tap of the channel.
- On the other hand, *inter-cell* interference is more poorly behaved in the downlink than in the uplink. In the uplink, there are many distributed

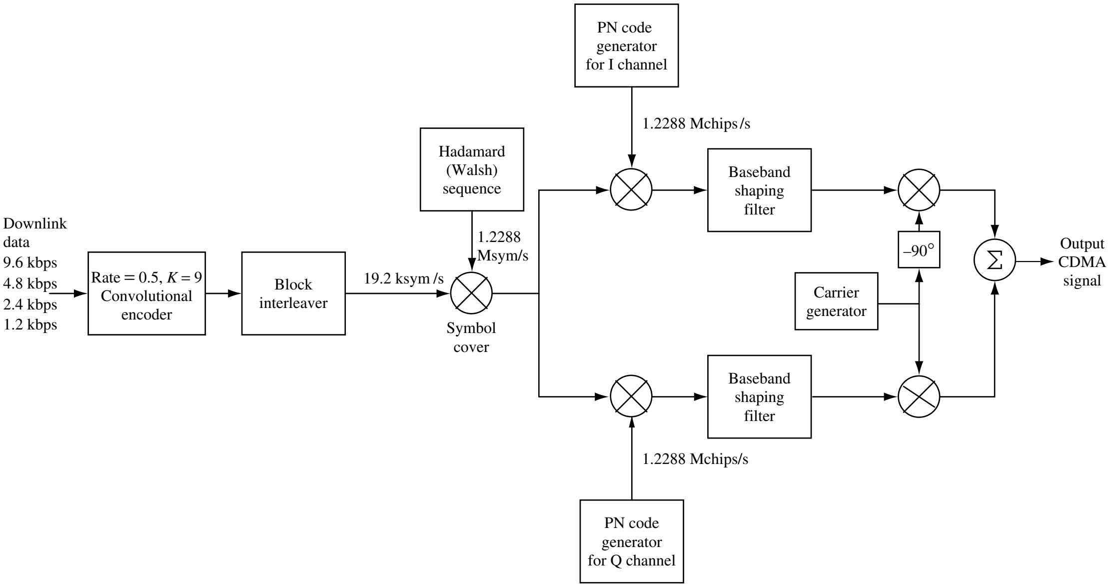

Figure 4.8 The IS-95 downlink.

users transmitting with small power, and significant interference averaging occurs. In the downlink, in contrast, there are only a few neighboring base-stations but each transmits at high power. There is much less interference averaging and the downlink capacity takes a significant hit compared to the uplink.

- In the uplink, soft handoff is accomplished by multiple base-stations listening to the transmitted signal from the mobile. No extra system resource needs to be allocated for this task. In the downlink, however, multiple base-stations have to simultaneously transmit to a mobile in soft handoff. Since each cell has a fixed number of orthogonal codes for the users, this means that a user in soft handoff is consuming double or more system resources. (See Exercise 4.13 for a precise formulation of the downlink soft handoff problem.)
- It is common to use a strong pilot and perform coherent demodulation in the downlink, since the common pilot can be shared by all the users. With the knowledge of the channels from each base-station, a user in soft handoff can also coherently combine the signals from the different base-stations. Synchronization tasks are also made easier in the presence of a strong pilot.

As an example, the IS-95 downlink is shown in Figure 4.8. Note the different roles of the Hadamard sequences in the uplink and in the downlink. In the uplink, the Hadamard sequences serve as an orthogonal modulation for each individual user so that non-coherent demodulation can be performed. In the downlink, in contrast, each user in the cell is assigned a *different* Hadamard sequence to keep them orthogonal (at the transmitter).

4.3.3 System issues

Signal characteristics

Consider the baseband uplink signal of a user given in (4.1). Due to the abrupt transitions (from $+1$ to -1 and vice versa) of the pseudonoise sequences s_n, the bandwidth occupied by this signal is very large. On the other hand, the signal has to occupy an allotted bandwidth. As an example, we see that the IS-95 system uses a bandwidth of 1.2288 MHz and a steep fall off after 1.67 MHz. To fit this allotted bandwidth, the signal in (4.1) is passed through a pulse shaping filter and then modulated on to the carrier. Thus though the signal in (4.1) has a perfect PAPR (equal to 1), the resulting transmit signal has a larger PAPR. The overall signal transmitted from the base-station is the superposition of all the user signals and this aggregate signal has PAPR performance similar to that of the narrowband system described in the previous section.

Sectorization

In the narrowband system we saw that all users can maintain high SINR due to the nature of the allocations. In fact, this was the benefit gained by paying the price of poor (re)use of the spectrum. In the CDMA system, however, due to the intra and inter-cell interferences, the values of SINR possible are very small. Now consider sectorization with universal frequency reuse among the sectors. Ideally (with full isolation among the sectors), this allows us to increase the system capacity by a factor equal to the number of sectors. However, in practice each sector now has to contend with inter-sector interference as well. Since intra-sector and inter-cell interference dominate the noise faced by the user signals, the additional interference caused due to sectorization does not cause a further degradation in SINR. Thus sectors of the same cell reuse the frequency without much of an impact on the performance.

Network issues

We have observed that timing acquisition (at a chip level accuracy) by a mobile is a computationally intensive step. Thus we would like to have this step repeated as infrequently as possible. On the other hand, to achieve soft handoff this acquisition has to be done (synchronously) for all base-stations with which the mobile communicates. To facilitate this step and the eventual handoff, implementations of the IS-95 system use high precision clocks (about 1 ppm (parts per million)) and further, synchronize the clocks at the base-stations through a proprietary wireline network that connects the base-stations. This networking cost is the price paid in the design to ease the handoff process.

Summary 4.2 CDMA

Universal frequency reuse: all users, both within a cell and across different cells, transmit and receive on the entire bandwidth.

The signal of each user is modulated onto a pseudonoise sequence so that it appears as white noise to others.

Interference management is crucial for allowing universal frequency reuse:

- Intra-cell interference is managed via power control. Accurate closed-loop power control is particularly important for combating the near–far problem in the uplink.
- Inter-cell interference is managed via averaging of the effects of multiple interferers. It is more effective in the uplink than in the downlink.

Interference averaging also allows statistical multiplexing of bursty users, thus increasing system capacity.

Diversity of the point-to-point links is achieved by a combination of low-rate coding, time-interleaving and Rake combining.

Soft handoff provides a further level of macrodiversity, allowing users to communicate with multiple base-stations simultaneously.

4.4 Wideband systems: OFDM

The narrowband system design of making transmissions interference-free simplified several aspects of network design. One such aspect was that the performance of a user is insensitive to the received powers of other users. In contrast to the CDMA approach, the requirement for accurate power control is much less stringent in systems where user transmissions in the same cell are kept orthogonal. This is particularly important in systems designed to accommodate many users each with very low average data rate: the fixed overhead needed to perform tight power control for each user may be too expensive for such systems. On the other hand there is a penalty of poor spectral reuse in narrowband systems compared to the CDMA system. Basically, narrowband systems are ill suited for universal frequency reuse since they do not average interference. In this section, we describe a system that combines the desirable features of both these systems: maintaining orthogonality of transmissions within the cell and having universal frequency reuse across cells. Again, the latter feature is made possible through interference averaging.

4.4.1 Allocation design principles

The first step in the design is to decide on the user signals that ensure orthogonality after passing through the wireless channel. Recall from the discussion of the downlink signaling in the CDMA system that though the *transmit signals* of the users are orthogonal, they interfere with each other at the receiver after passing through the multipath channel. Thus any orthogonal set of signals will not suffice. If we model the wireless channel as a linear time invariant multipath channel, then the only eigenfunctions are the *sinusoids*. Thus sinusoid inputs remain orthogonal at the receiver no matter what the multipath channel is. However, due to the channel variations in time, we want to restrict the notion of orthogonality to no more than a coherence time interval. In this context, sinusoids are no longer orthogonal, but the sub-carriers of the OFDM scheme of Section 3.4.4 with the cyclic *prefix* for the multipath channel provide a set of orthogonal signals over an OFDM block length.

We describe an allocation of sets of OFDM sub-carriers as the user signals; this description is identical for both the downlink and the uplink. As in Section 3.4.4, the bandwidth W is divided into N_c sub-carriers. The number of sub-carriers N_c is chosen to be as large as possible. As we discussed earlier, N_c is limited by the coherence time, i.e., the OFDM symbol period $N_c/W < T_c$. In each cell, we would like to distribute these N_c sub-carriers to the users in it (with say n sub-carriers per user). The n sub-carriers should be spread out in frequency to take advantage of frequency diversity. There is no interference among user transmissions within a cell by this allocation.

With universal frequency reuse, there is however inter-cell interference. To be specific, let us focus on the uplink. Two users in neighboring cells sharing the same sub-carrier in any OFDM symbol time interfere with each other directly. If the two users are close to each other, the interference can be very severe and we would like to minimize such overlaps. However, due to full spectral reuse, there is such an overlap at every OFDM symbol time in a fully loaded system. Thus, the best one can do is to ensure that the interference does not come solely from one user (or a small set of users) and the interference seen over a coded sequence of OFDM symbols (forming a *frame*) can be attributed to most of the user transmissions in the neighboring cell. Then the *overall* interference seen over a frame is a function of the average received power of *all* the users in the neighboring cells. This is yet another example of the *interference diversity* concept we already saw in Section 4.3.

How are the designs of the previous two systems geared towards harvesting interference diversity? The CDMA design fully exploits interferer diversity by interference averaging. This is achieved by every user spreading its signals over the entire spectrum. On the other hand, the orthogonal allocation of channels in the GSM system is poorly suited from the point of view of interferer diversity. As we saw in Section 4.2, users in neighboring cells that are close to each other and transmitting on the same channel over the same slot cause severe interference to each other. This leads to a very degraded performance and the reason for it is clear: interference seen by a user comes solely from one interferer and there is no scope to see an average interference from all the users over a slot. If there were no hopping and coding across the sub-carriers, the OFDM system would behave exactly like a narrowband system and suffer the same fate.

Turning to the downlink we see that now all the transmissions in a cell occur from the same place: at the base-station. However, the power in different sub-carriers transmitted from the base-station can be vastly different. For example, the pilots (training symbols) are typically at a much higher power than the signal to a user very close to the base-station. Thus even in the downlink, we would like to hop the sub-carriers allocated to a user every OFDM symbol time so that over a frame the interference seen by a mobile is a function of the average transmit power of the neighboring base-stations.

4.4.2 Hopping pattern

We have arrived at two design rules for the sub-carrier allocations to the users. Allocate the n sub-carriers for the user as spread out as possible and further, hop the n sub-carriers every OFDM symbol time. We would like the hop patterns to be as "apart" as possible for neighboring base-stations. We now delve into the design of *periodic* hopping patterns that meet these broad design rules that repeat, say, every N_c OFDM symbol intervals. As we will see, the choice of the period to be equal to N_c along with the assumption that N_c be prime (which we now make) simplifies the construction of the hopping pattern.

The periodic hopping pattern of the N_c sub-carriers can be represented by a square matrix (of dimension N_c) with entries from the set of *virtual channels*, namely $0, 1, \ldots, N_c - 1$. Each virtual channel hops over different sub-carriers at different OFDM symbol times. Each row of the hopping matrix corresponds to a sub-carrier and each column represents an OFDM symbol time, and the entries represent the virtual channels that use that sub-carrier in different OFDM symbol times. In particular, the (i, j) entry of the matrix corresponds to the virtual channel number the ith sub-carrier is taken on by, at OFDM symbol time j. We require that every virtual channel hop over all the sub-carriers in each period for maximal frequency diversity. Further, in any OFDM symbol time the virtual channels occupy different sub-carriers. These two requirements correspond to the constraint that each row and column of the hopping matrix contains every virtual channel number $(0, \ldots, N_c - 1)$, exactly once. Such a matrix is called a *Latin square*. Figure 4.9 shows hopping patterns of the 5 virtual channels over the 5 OFDM symbol times (i.e., $N_c = 5$). The horizontal axis corresponds to OFDM symbol times and the vertical axis denotes the 5 physical sub-carriers (as in Figure 3.25), and the sub-carriers the virtual channels adopt are denoted by darkened squares. The corresponding hopping pattern matrix is

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 0 & 1 \\ 4 & 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 & 0 \\ 3 & 4 & 0 & 1 & 2 \end{bmatrix}.$$

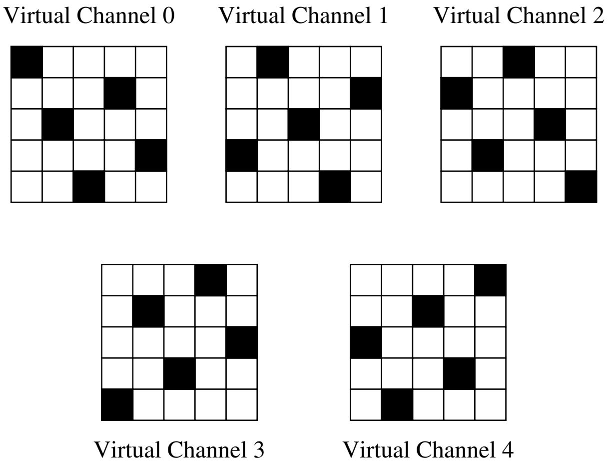

Figure 4.9 Virtual channel hopping patterns for $N_c = 5$.

For example, we see that the virtual channel 0 is assigned the OFDM symbol time and sub-carrier pairs (0, 0), (1, 2), (2, 4), (3, 1), (4, 3). Now users could be allocated n virtual channels, accommodating $\lfloor N_c/n \rfloor$ users.

Each base-station has its own hopping matrix (Latin square) that determines the physical structure of the virtual channels. Our design rule to maximize interferer diversity requires us to have minimal overlap between virtual channels of neighboring base-stations. In particular, we would like to have exactly one time/sub-carrier collision for every pair of virtual channels of two base-stations that employ these hopping patterns. Two Latin squares that have this property are said to be *orthogonal*.

When N_c is prime, there is a simple construction for a family of $N_c - 1$ mutually orthogonal Latin squares. For $a = 1, \ldots, N_c - 1$ we define an $N_c \times N_c$ matrix R^a with (i, j)th entry

$$R^a_{ij} = ai + j \quad \text{modulo } N_c. \tag{4.23}$$

Here we index rows and columns from 0 through $N_c - 1$. In Exercise 4.14, you are asked to verify that R^a is a Latin square and further that for every $a \neq b$ the Latin squares R^a and R^b are orthogonal. Observe that Figure 4.9 depicts a Latin square hopping pattern of this type with $a = 2$ and $N_c = 5$.

With these Latin squares as the hopping patterns, we can assess the performance of data transmission over a single virtual channel. First, due to the hopping over the entire band, the frequency diversity in the channel is harnessed. Second, the interference seen due to inter-cell transmissions comes from different virtual channels (and repeats after N_c symbol times). Coding over several OFDM symbols allows the full interferer diversity to be harnessed: coding ensures that no one single strong interference from a virtual channel can cause degradation in performance. If sufficient interleaving is permitted, then the time diversity in the system can also be obtained.

To implement these design goals in a cellular system successfully, the users within the cell must be synchronized to their corresponding base-station. This way, the simultaneous uplink transmissions are still orthogonal at the base-station. Further, the transmissions of neighboring base-stations also have to be synchronized. This way the design of the hopping patterns to average the interference is fully utilized. Observe that the synchronization needs to be done only at the level of OFDM symbols, which is much coarser than at the level of chips.

4.4.3 Signal characteristics and receiver design

Let us consider the signal transmission corresponding to a particular user (either in the uplink or the downlink). The signal consists of n virtual channels, which over a slot constitute a set of n OFDM sub-carriers that are hopped over OFDM symbol times. Thus, though the signal *information content* can be "narrow" (for small ratios n/N_c), the signal bandwidth itself is wide. Further, since the bandwidth range occupied varies from symbol to symbol, each (mobile) receiver has to be wideband. That is, the sampling rate is proportional to $1/W$. Thus this signal constitutes a (frequency hopped) spread-spectrum signal just as the CDMA signal is: the ratio of data rate to bandwidth occupied by the signal is small. However, unlike the CDMA signal, which spreads the energy over the entire bandwidth, here the energy of the signal is only in certain sub-carriers (n of a total N_c). As discussed in Chapter 3, fewer channel parameters have to be measured and channel estimation with this signal is superior to that with the CDMA signal.

The major advantages of the third system design are the frequency and interferer diversity features. There are a few engineering drawbacks to this choice. The first is that the mobile sampling rate is quite high (same as that of the CDMA system design but much higher than that of the first system). All signal processing operations (such as the FFT and IFFT) are driven off this basic rate and this dictates the processing power required at the mobile receiver. The second drawback is with respect to the transmit signal on the uplink. In Exercise 4.15, we calculate the PAPR of a canonical transmit signal in this design and observe that it is significantly high, as compared to the signal in the GSM and CDMA systems. As we discussed in the first system earlier, this higher PAPR translates into a larger bias in the power amplifier settings and a correspondingly lower average efficiency. Several engineering solutions have been proposed to this essentially engineering problem (as opposed to the more central communication problem which deals with the uncertainties in the channel) and we review some of these in Exercise 4.16.

4.4.4 Sectorization

What range of SINRs is possible for the users in this system? We observed that while the first (narrowband) system provided high SINRs to all the mobiles, almost no user was in a high SINR scenario in the CDMA system due to the intra-cell interference. The range of SINRs possible in this system is midway between these two extremes. First, we observe that the only source of interference is inter-cell. So, users close to the base-station will be able to have high SINRs since they are impacted less from inter-cell interference. On the other hand, users at the edge of the cell are interference limited and cannot support high SINRs. If there is a feedback of the received SINRs then users closer by the base-station can take advantage of the higher SINR by transmitting and receiving at higher data rates.

What is the impact of sectorization? If we universally reuse the frequency among the sectors, then there is inter-sector interference. We can now observe an important difference between inter-sector and inter-cell interference. While inter-cell interference affects mostly the users at the edge of the cell, inter-sector interference affects users regardless of whether they are at the edge of the cell or close to the base-station (the impact is pronounced on those at the edge of the sectors). This interference now reduces the dynamic range of SINRs this system is capable of providing.

Example 4.1 Flash-OFDM

A technology that partially implements the design features of the wideband OFDM system is Flash-OFDM, developed by Flarion Technologies [38]. Over 1.25 MHz, there are 113 sub-carriers, i.e., $N_c = 113$. The 113 virtual channels are created from these sub-carriers using the Latin square hopping patterns (in the downlink the hops are done every OFDM symbol but once in every 7 OFDM symbols in the uplink). The sampling rate (or equivalently, chip rate) is 1.25 MHz and a cyclic prefix of 16 samples (or chips) covers for a delay spread of approximately 11 μs. This means that the OFDM symbol is 128 samples, or approximately 100 μs long.

There are four *traffic channels* of different granularity: there are five in the uplink (comprising 7, 14, 14, 14 and 28 virtual channels) and four in the downlink (comprising 48, 24, 12, 12 virtual channels). Users are scheduled on different traffic channels depending on their traffic requirements and channel conditions (we study the desired properties of the scheduling algorithm in greater detail in Chapter 6). The scheduling algorithm operates once every *slot*: a slot is about 1.4 ms long, i.e., it consists of 14 OFDM symbols. So, if a user is scheduled (say, in the downlink) the traffic channel consisting of 48 virtual channels, it can transmit 672 OFDM symbols over the slot when it is scheduled. An appropriate rate LDPC (low-density parity check) code combined with a simple modulation scheme (such as

QPSK or 16-QAM) is used to convert the raw information bits into the 672 OFDM symbols.

The different levels of granularity of the traffic channels are ideally suited to carry bursty traffic. Indeed, Flash-OFDM is designed to act in a data network where it harnesses the statistical multiplexing gains of the user's bursty data traffic by its packet-switching operation.

The mobiles are in three different states in the network. When they are inactive, they go to a "sleep" mode monitoring the base-station signal every once in a while: this mode saves power by turning off most of the mobile device functionalities. On the other hand, when the mobile is actively receiving and/or sending data it is in the "ON" mode: this mode requires the network to assign resources to the mobile to perform periodic power control updates and timing and frequency synchronization. Apart from these two states, there is an in-between "HOLD" mode: here mobiles that have been recently active are placed without power control updates but still maintaining timing and frequency synchronization with the base-station. Since the intra-cell users are orthogonal and the accuracy of power control can be coarse, users in a HOLD state can be quickly moved to an ON state when there is a need to send or receive data. Flash-OFDM has the ability to hold approximately 30, 130 and 1000 mobiles in the ON, HOLD and sleep modes.

For many data applications, it is important to be able to keep a large number of users in the HOLD state, since each user may send traffic only once in a while and in short bursts (requests for http transfers, acknowledgements, etc.) but when they do want to send, they require short latency and quick access to the wireless resource. It is difficult to support this HOLD state in a CDMA system. Since accurate power control is crucial because of the near–far problem, a user who is not currently power-controlled is required to slowly ramp up its power before it can send traffic. This incurs a very significant delay.[12] On the other hand, it is very expensive to power control a large number of users who only transmit infrequently. In an orthogonal system like OFDM, this overhead can be largely avoided. The issue does not arise in a voice system since each user sends constantly and the power control overhead is only a small percentage of the payload (about 10% in IS-95).

Chapter 4 The main plot

The focus of this chapter is on multiple access, interference management and the system issues in the design of cellular networks. To highlight the

---

[12] Readers from the San Francisco Bay area may be familiar with the notorious "Fast Track" lanes for the Bay Bridge. Once a car gets on one of these lanes, it can cross the toll plaza very quickly. But the problem is that most of the delay is in getting to them through the traffic jam!

issues, we looked at three different system designs. Their key characteristics are compared and contrasted in the table below.

	Narrowband system	Wideband CDMA	Wideband OFDM
Signal	Narrowband	Wideband	Wideband
Intra-cell BW allocation	Orthogonal	Pseudorandom	Orthogonal
Intra-cell interference	None	Significant	None
Inter-cell BW allocation	Partial reuse	Universal reuse	Universal reuse
Inter-cell uplink interference	Bursty	Averaged	Averaged
Accuracy of power control	Low	High	Low
Operating SINR	High	Low	Range: low to high
PAPR of uplink signal	Low	Medium	High
Example system	GSM	IS-95	Flash-OFDM

4.5 Bibliographical notes

The two important aspects that have to be addressed by a wireless system designer are how resource is allocated within a cell among the users and how interference (both intra- and inter-cell) is handled. Three topical wireless technologies have been used as case studies to bring forth the tradeoffs the designer has to make. The standards IS-136 [60] and GSM [99] have been the substrate on which the discussion of the narrowband system design is built. The wideband CDMA design is based on the widely implemented second-generational technology IS-95 [61]. A succinct description of the the technical underpinnings of the IS-95 design has been done by Viterbi [140] with emphasis on a system view, and our discussion here has been influenced by it. The frequency hopping OFDM system based on Latin squares was first suggested by Wyner [150] and Pottie and Calderbank [94]. This basic physical-layer construct has been built into a technology (Flash-OFDM [38]).

4.6 Exercises

Exercise 4.1 In Figure 4.2 we set a specific reuse pattern. A channel used in a cell precludes its use in all the neighboring cells. With this allocation policy the reuse factor is at least 1/7. This is a rather ad hoc allocation of channels to the cells and the reuse ratio can be improved; for example, the four-color theorem [102] asserts that a planar graph can be colored with four colors with no two vertices joined by an edge sharing the same channel. Further, we may have to allocate more channels to cells which are crowded. In this question, we consider modeling this problem.

Let us represent the cells by a finite set (of vertices) $V := \{v_1, \ldots, v_C\}$; one vertex for each cell, so there are C cells. We want to be able to say that only a certain collection of vertices can share the same channel. We do this by defining an *allowable set* $S \subseteq V$ such that all the vertices in S can share the same channel. We are only interested in *maximal allowable sets*: these are allowable sets with no strict superset also an allowable set. Suppose the maximal allowable sets are M in number, denoted as $S_1, \ldots, S_M$. Each of these maximal allowable sets can be thought of as a *hyper-edge* (the traditional definition of *edge* means a pair of vertices) and the collection of V and the hyper-edges forms a *hyper-graph*. You can learn more about hyper-graphs from [7].

1. Consider the hexagonal cellular system in Figure 4.10. Suppose we do not allow any two neighboring cells to share the same channel and further not allow the same channel to be allocated to cells 1, 3 and 5. Similarly, cells 2, 4 and 6 cannot share the same channel. For this example, what are C and M? Enumerate the maximal allowable sets $S_1, \ldots, S_M$.
2. The hyper-edges can also be represented as an *adjacency matrix* of size $C \times M$: the (i, j)th entry is

$$a_{ij} := \begin{cases} 1 & \text{if } v_i \in S_j, \\ 0 & \text{if } v_i \notin S_j. \end{cases} \tag{4.24}$$

   For the example in Figure 4.10, explicitly construct the adjacency matrix.

Exercise 4.2 [84] In Exercise 4.1, we considered a graphical model of the cellular system and constraints on channel allocation. In this exercise, we consider modeling the *dynamic traffic and channel allocation algorithms*.

Suppose there are N channels to be allocated. Further, the allocation has to satisfy the reuse conditions: in the graphical model this means that each channel is mapped to one of the maximal allowable sets. The traffic comprises calls originating and terminating in the cells. Consider the following statistical model. The average number of overall calls in all the cells is B. This number accounts for *new* call arrivals and calls leaving the cell due to termination. The *traffic intensity* is the number of call arrivals per available channel, $r := B/N$ (in Erlangs per channel). A fraction p_i of these calls occur in cell i (so that $\sum_{i=1}^{C} p_i = 1$). So, the long-term average number of calls per channel to be handled in cell i is $p_i r$. We need a channel to service a call, so to meet this traffic we need on an average at least $p_i r$ channels allocated to cell i. We fix the traffic profile $p_1, \ldots, p_C$ over the time-scale for which the number of calls averaging is done. If a cell has used up all its allocated channels, then a new call cannot be serviced and is dropped.

A dynamic channel allocation algorithm allocates the N channels to the C cells to meet the instantaneous traffic requirements and further satisfies the reuse pattern. Let us focus on the average performance of a dynamic channel allocation algorithm: this is the sum of the average traffic per channel supported by each cell, denoted by $T(r)$.

1. Show that

$$T(r) \leq \max_{j=1\ldots M} \sum_{i=1}^{C} a_{ij}. \tag{4.25}$$

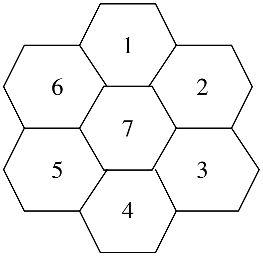

Figure 4.10 A narrowband system with seven cells. Adjacent cells cannot share the same channel and cells $\{1, 3, 5\}$ and $\{2, 4, 6\}$ cannot share the same channel either.

Hint: The quantity on the right hand side is the cardinality of the largest maximal allowable set.

2. Show that

$$T(r) \leq \sum_{i=1}^{C} p_i r = r, \tag{4.26}$$

i.e., the total arrival rate is also an upper bound.

3. Let us combine the two simple upper bounds in (4.25) and (4.26). For every fixed list of of C numbers $y_i \in [0, 1]$, $i = 1, \ldots, C$, show that

$$T(r) \leq \sum_{i=1}^{C} y_i p_i r + \max_{j=1\ldots M} \sum_{i=1}^{C} (1 - y_i) a_{ij}. \tag{4.27}$$

Exercise 4.3 This exercise is a sequel to Exercises 4.1 and 4.2. Consider the cellular system example in Figure 4.10, with the arrival rates $p_i = 1/8$ for $i = 1, \ldots, 6$ (all the cells at the edge) and $p_7 = 1/4$ (the center cell).

1. Derive a good upper bound on $T(r)$, the traffic carried per channel for any dynamic channel allocation algorithm for this system. In particular, use the upper bound derived in (4.27), but optimized over all choices of $y_1, \ldots, y_C$. *Hint*: The upper bound on $T(r)$ in (4.27) is *linear* in the variables $y_1, \ldots, y_C$. So, you can use software such as MATLAB (with the function `linprog`) to arrive at your answer.
2. In general, a channel allocation policy is dynamic: i.e., the number of channels allocated to a cell varies with time as a function of the traffic. Since we are interested in the average behavior of a policy over a large amount of time, it is possible that *static* channel allocation policies also do well. (Static policies allocate channels to the cells in the beginning and do not alter this allocation to suit the varying traffic levels.) Consider the following static allocation policy defined by the probability vector $\mathbf{x} := (x_1, \ldots, x_M)$, i.e., $\sum_{j=1}^{M} x_j = 1$. Each maximal allowable set S_j is allocated $\lfloor N x_j \rfloor$ channels, in the sense that each cell in S_j is allocated these $\lfloor N x_j \rfloor$ channels. Observe that cell i is allocated

$$\sum_{j=1}^{M} \lfloor N x_j \rfloor a_{ij}$$

channels. Denote $T_{\mathbf{x}}(r)$ as the carried traffic by using this static channel allocation algorithm.

If the incoming traffic is smooth enough that the carried traffic in each cell is the minimum of arrival traffic in that cell and the number of channels allocated to that cell,

$$\lim_{N\to\infty} T_{\mathbf{x}}(r) = \sum_{i=1}^{C} \min\left(r p_i, \sum_{j=1}^{M} x_j a_{ij} \right), \qquad \forall r > 0. \tag{4.28}$$

What are good static allocation policies? For the cellular system model in Figure 4.10, try out simple static channel allocation algorithms that you can think of. You can evaluate the performance of your algorithm numerically by simulating a smooth traffic arrival process (common models are uniform arrivals and independent and exponential inter-arrival times). How does your answer compare to the upper bound derived in part (1)?

In [84], the authors show that there exists a static allocation policy that can actually achieve (for large N, because the integer truncation effects have to be smoothed out) the upper bound in part (1) for every graphical model and traffic arrival rate.

Exercise 4.4 In this exercise we study the PAPR of the uplink transmit signal in narrowband systems. The uplink transmit signal is confined to a small bandwidth (200 kHz in the GSM standard). Consider the folowing simple model of the transmit signal using the idealized pulse shaping filter:

$$s(t) = \Re\left[\sum_{n=0}^{\infty} x[n]\,\mathrm{sinc}(t-nT)\exp(\mathrm{j}2\pi f_\mathrm{c}t)\right], \qquad t \geq 0. \tag{4.29}$$

Here T is approximately the inverse of the bandwidth (5 μs in the GSM standard) and $\{x[n]\}$ is the sequence of (complex) data symbols. The carrier frequency is denoted by f_c; for simplicity let us assume that $f_\mathrm{c}T$ is an integer.

1. The raw information bits are coded and modulated resulting in the data symbols $x[n]$. Modeling the data symbols as i.i.d. uniformly distributed on the complex unit circle, calculate the average power in the transmit signal $s(t)$, averaged over the data symbols. Let us denote the average power by P_av.
2. The statistical behavior of the transmit signal $s(t)$ is periodic with period T. Thus we can focus on the peak power within the time interval $[0, T]$, denoted as

$$PP(d) = \max_{0\leq t\leq T} |s(t)|^2. \tag{4.30}$$

The peak power is a random variable since the data symbols are random. Obtain an estimate for the average peak power. How does your estimate depend on T? What does this imply about the PAPR (ratio of PP to P_av) of the narrowband signal $s(t)$?

Exercise 4.5 [56] In this problem we study the uplink power control problem in the CDMA system in some detail. Consider the uplink of a CDMA system with a total of K mobiles trying to communicate with L base-stations. Each mobile k communicates with just one among a subset S_k of the L base-stations; this base-station assignment is denoted by c_k (i.e., we do not model diversity combining via soft handoff in this problem). Observe that by restricting S_k to have just one element, we are ruling out soft handoff as well. As in Section 4.3.1, we denote the transmit power of mobile k by P_k and the channel attenuation from mobile k to base-station m by g_{km}. For successful communication we require the $\mathcal{E}_\mathrm{b}/I_0$ to be at least a target level β, i.e., successful uplink communication of the mobiles entails the constraints (cf. (4.10)):

$$\frac{\mathcal{E}_\mathrm{b}}{I_0} = \frac{GP_k g_{k,c_k}}{\sum_{n\neq k} P_n g_{n,c_k} + N_0 W} \geq \beta_k, \qquad k = 1, 2, \ldots, K. \tag{4.31}$$

Here we have let the target level be potentially different for each mobile and denoted $G = W/R$ as the processing gain of the CDMA system. Writing the transmit powers as the vector $\mathbf{p} = (p_1, \ldots, p_K)^t$, show that (4.31) can be written as

$$(\mathbf{I}_K - \mathbf{F})\mathbf{p} \geq \mathbf{b}, \tag{4.32}$$

where $\mathbf{F}$ is the $K \times K$ matrix with strictly positive off-diagonal entries

$$f_{ij} = \begin{cases} 0 & \text{if } i = j, \\ \dfrac{g_{jc_i}\beta_i}{g_{ic_i}} & \text{if } i \neq j, \end{cases} \tag{4.33}$$

and

$$\mathbf{b} := N_0 W \left(\frac{\beta_1}{g_{1,c_1}}, \ldots, \frac{\beta_K}{g_{K,c_K}} \right)^t. \tag{4.34}$$

It can be shown (see Exercise 4.6) that there exist positive powers to make $\mathcal{E}_b/I_0$ meet the target levels, exactly when all the eigenvalues of $\mathbf{F}$ have absolute value strictly less than 1. In this case, there is in fact a *component-wise minimal* vector of powers that allows successful communication and is simply given by

$$\mathbf{p}^* = (\mathbf{I}_K - \mathbf{F})^{-1}\mathbf{b}. \tag{4.35}$$

Exercise 4.6 Consider the set of linear inequalities in (4.32) that correspond to the $\mathcal{E}_b/I_0$ requirements in the uplink of a CDMA system. In this exercise we investigate the mathematical constraints on the physical parameters of the CDMA system (i.e., the channel gains and desired target levels) which allow reliable communication.

We begin by observing that $\mathbf{F}$ is a non-negative matrix (i.e., it has non-negative entries). A non-negative matrix F is said to be *irreducible* if there exists a positive integer m such that $\mathbf{F}^m$ has all entries strictly positive.

1. Show that $\mathbf{F}$ in (4.33) is irreducible. (The number of mobiles K is at least two.)
2. Non-negative matrices also show up as the probability transition matrices of finite state *Markov chains*. An important property of irreducible non-negative matrices is the *Perron–Frobenius* theorem: There exists a strictly positive eigenvalue (called the Perron–Frobenius eigenvalue) which is strictly bigger than the absolute value of any of the other eigenvalues. Further, there is a unique right eigenvector corresponding to the Perron–Frobenius eigenvalue, and this has strictly positive entries. Recall this result from a book on non-negative matrices such as [106].
3. Consider the vector form of the $\mathcal{E}_b/I_0$ constraints of the mobiles in (4.32) with $\mathbf{F}$ a non-negative irreducible matrix and $\mathbf{b}$ having strictly positive entries. Show that the following statements are equivalent.
   (a) There exists $\mathbf{p}$ satisfying (4.32) and having strictly positive entries.
   (b) The Perron–Frobenius eigenvalue of $\mathbf{F}$ is strictly smaller than 1.
   (c) $(\mathbf{I}_K - \mathbf{F})^{-1}$ exists and has strictly positive entries.

The upshot is that the existence or non-existence of a power vector that permits successful uplink communication from all the mobiles to their corresponding base-stations (with the assignment $k \mapsto c_k$) can be characterized in terms of the Perron–Frobenius eigenvalue of an irreducible non-negative matrix $\mathbf{F}$.

Exercise 4.7 In this problem, a sequel to Exercise 4.5, we allow the assignment of mobiles to base-stations to be in our control. Let $\mathbf{t} := (\beta_1, \ldots, \beta_K)$ denote the vector of the desired target thresholds on the $\mathcal{E}_b/I_0$ of the mobiles. Given an assignment of mobiles to base-stations $k \mapsto c_k$ (with $c_k \in S_k$), we say that the pair $(c, \mathbf{t})$ is *feasible* if there is a power vector that permits successful communication from all the mobiles to their corresponding base-stations (i.e., user k's $\mathcal{E}_b/I_0$ meets the target level β_k).

1. Show that if $(c, \mathbf{t}^{(1)})$ is feasible and $\mathbf{t}^{(2)}$ is another vector of desired target levels such that $\beta_k^{(1)} \geq \beta_k^{(2)}$ for each mobile $1 \leq k \leq K$, then $(c, \mathbf{t}^{(2)})$ is also feasible.
2. Suppose $(c^{(1)}, \mathbf{t})$ and $(c^{(2)}, \mathbf{t})$ are feasible. Let $\mathbf{p}^{(1)*}$ and $\mathbf{p}^{(2)*}$ denote the corresponding minimal vectors of powers allowing successful communication, and define

   $$p_k^{(3)} := \min\left(p_k^{(1)*}, p_k^{(2)*}\right).$$

   Define the new assignment

   $$c_k^{(3)} := \begin{cases} c_k^{(1)} & \text{if } p_k^{(1)*} \leq p_k^{(2)*}, \\ c_k^{(2)} & \text{if } p_k^{(1)*} > p_k^{(2)*}. \end{cases}$$

   Define the new target levels

   $$\beta_k^{(3)} := \frac{g_{kc_k^{(3)}} p_k^{(3)*}}{N_0 W + \sum_{n \neq k} g_{nc_n^{(3)}} p_n^{(3)*}}, \qquad k = 1, \ldots, K,$$

   and the vector $\mathbf{t}^{(3)} = (\beta_1^{(3)}, \ldots, \beta_K^{(3)})$. Show that $(c^{(3)}, \mathbf{t}^{(3)})$ is feasible and further that $\beta_k^{(3)} \geq \beta_k$ for all mobiles $1 \leq k \leq K$ (i.e., $\mathbf{t}^{(3)} \geq \mathbf{t}$ component-wise).
3. Using the results of the previous two parts, show that if uplink communication is feasible, then there is a unique component-wise minimum vector of powers that allows for successful uplink communication of all the mobiles, by appropriate assignment of mobiles to base-stations allowing successful communication. Further show that for any other assignment of mobiles to base-stations allowing successful communication the corresponding minimal power vector is component-wise at least as large as this power vector.

Exercise 4.8 [56, 151] In this problem, a sequel to Exercise 4.7, we will see an adaptive algorithm that updates the transmit powers of the mobiles in the uplink and the assignment of base-stations to the mobiles. The key property of this adaptive algorithm is that it converges to the component-wise minimal power among all assignments of base-stations to the mobiles (if there exists some assignment that is feasible, as discussed in Exercise 4.7(3)).

Users begin with an arbitrary power vector $\mathbf{p}^{(1)}$ and base-station assignment $c^{(1)}$ at the starting time 1. At time m, let the transmit powers of the mobiles be denoted by (the vector) $\mathbf{p}^{(m)}$ and the base-station assignment function be denoted by $c^{(m)}$. Let us first calculate the interference seen by mobile n at each of the base-stations $l \in S_n$; here S_n is the set of base-stations that can be assigned to mobile n.

$$I_{nl}^{(m)} := \sum_{k \neq n} g_{kl} p_k^{(m)} + N_0 W. \tag{4.36}$$

Now, we choose *greedily* to assign mobile n to that base-station which requires the least transmit power on the part of mobile n to meet its target level β_n. That is,

$$p_n^{(m+1)} := \min_{l \in S_n} \frac{\beta_n I_{nl}^{(m)}}{G g_{nl}}, \tag{4.37}$$

$$c_n^{(m+1)} := \arg\min_{l \in S_n} \frac{\beta_n I_{nl}^{(m)}}{g_{nl}}. \tag{4.38}$$

Consider this greedy update to each mobile being done *synchronously*: i.e., the updates of transmit power and base-station assignment for every mobile at time $m+1$ is made based on the transmit powers of all other the mobiles at time m. Let us denote this greedy update algorithm by the map $I : \mathbf{p}^{(m)} \mapsto \mathbf{p}^{(m+1)}$.

1. Show the following properties of I. Vector inequalities are defined to be component-wise inequalities.
   (a) $I(\mathbf{p}) > 0$ for every $\mathbf{p} \geq \mathbf{0}$.
   (b) $I(\mathbf{p}) \geq I(\tilde{\mathbf{p}})$, whenever $\mathbf{p} \geq \tilde{\mathbf{p}}$.
   (c) $I(\alpha \mathbf{p}) \leq \alpha I(\mathbf{p})$ whenever $\alpha > 1$.
2. Using the previous part, or otherwise, show that if I has a fixed point (denoted by $\mathbf{p}^*$) then it is unique.
3. Using the previous two parts, show that if I has a fixed point then $\mathbf{p}^{(m)} \to \mathbf{p}^*$ component-wise as $m \to \infty$ where $\mathbf{p}^{(m)} := I(\mathbf{p}^{(m-1)})$ and $\mathbf{p}^{(1)}$ and $c^{(1)}$ are an arbitrary initial allocation of transmit powers and assignments of base-stations.
4. If I has a fixed point, then show that the uplink communication problem must be feasible and further, the fixed point $\mathbf{p}^*$ must be the same as the component-wise minimal power vector derived in Exercise 4.7(3).

Exercise 4.9 Consider the following *asynchronous* version of the update algorithm in Exercise 4.8. Each mobile's update (of power and base-station assignment) occurs asynchronously based on some previous knowledge of all the other users' transmit powers. Say the update of mobile n at time m is based on mobile k's transmit power at time $\tau_{nk}(m)$. Clearly, $\tau_{nk}(m) \leq m$ and we require that each user eventually has an update of the other users' powers, i.e., for every time m_0 there exists time $m_1 \geq m_0$ such that $\tau_{nk}(m) \geq m_0$ for every time $m \geq m_1$. We further require that each user's power and base-station assignment is allocated infinitely often. Then, starting from any initial condition of powers of the users, show that the asynchronous power update algorithm converges to the optimal power vector $\mathbf{p}^*$ (assuming the problem is feasible, so that $\mathbf{p}^*$ exists in the first place).

Exercise 4.10 Consider the uplink of a CDMA system. Suppose there is only a single cell with just two users communicating to the base-station in the cell.

1. Express mathematically the set of all feasible power vectors to support given $\mathcal{E}_b/I_0$ requirements (assumed to be both equal to β).
2. Sketch examples of sets of feasible power vectors. Give one example where the feasible set is non-empty and give one example where the feasible set is empty. For the case where the feasible set is non-empty, identify the component-wise minimum power vector.
3. For the example in part (2) where the feasible set is non-empty, start from an arbitrary initial point and run the power control algorithm described in Section 4.3.1 (and studied in detail in Exercise 4.8). Exhibit the trajectory of power updates and

how it converges to the component-wise minimum solution. (You can either do this by hand or use MATLAB.)

4. Now suppose there are two cells with two base-stations and each of the two users can be connected to either one of them, i.e. the users are in soft handoff. Extend parts (1) and (2) to this scenario.
5. Extend the iterative power control algorithm in part (3) to the soft handoff scenario and redo part (3).
6. For a general number of users, do you think that it is always true that, in the optimal solution, each user is always connected to the base-station to which it has the strongest channel gain? Explain.

Exercise 4.11 (Out-of-cell interference averaging) Consider a cellular system with two adjacent single-dimensional cells along a highway, each of length d. The base-stations are at the midpoint of their respective cell. Suppose there are K users in each cell, and the location of each user is uniformly and independently located in its cell. Users in cell i are power controlled to the base-station in cell i, and create interference at the base-station in the adjacent cell. The power attenuation is proportional to $r^{-\alpha}$ where r is the distance. The system bandwidth is W Hz and the $\mathcal{E}_b/I_0$ requirement of each user is β. You can assume that the background noise is small compared to the interference and that users are maintained orthogonal within a cell with the out-of-cell interference from each of the interferers spread across the entire bandwidth. (This is an approximate model for the OFDM system in the text.)

1. Outage occurs when the users are located such that the out-of-cell interference is too large. For a given outage probability p_{out}, give an approximate expression for the spectral efficiency of the system as a function of K, α and β.
2. What is the limiting spectral efficiency as K and W grow? How does this depend on α?
3. Plot the spectral efficiency as a function of K for $\alpha = 2$ and $\beta = 7$ dB. Is the spectral efficiency an increasing or decreasing function of K? What is the limiting value?
4. We have assumed orthogonal users within a cell. But in a CDMA system, there is intra-cell interference as well. Assuming that all users within a cell are perfectly power controlled at their base-station, repeat the analysis in the first three parts of the question. From your plots, what qualitative differences between the CDMA and orthogonal systems can you observe? Intuitively explain your observations. *Hint*: Consider first what happens when the number of users increases from $K = 1$ to $K = 2$.

Exercise 4.12 Consider the uplink of a single-cell CDMA system with N users active all the time. In the text we have assumed the received powers are controlled such that they are exactly equal to the target level needed to deliver the desired SINR requirement for each user. In practice, the received powers are controlled imperfectly due to various factors such as tracking errors and errors in the feedback links. Suppose that when the target received power level is P, the actual received power of user i is $\epsilon_i P$, where ϵ_i are i.i.d. random variables whose statistics do not depend on P. Experimental data and theoretical analysis suggest that a good model for ϵ_i is a log normal distribution, i.e., $\log(\epsilon_i)$ follows a Gaussian distribution with mean μ and variance σ^2.

1. Assuming there is no power constraint on the users, give an approximate expression for the achievable spectral efficiency (bits/s/Hz) to support N users for a given outage probability p_{out} and $\mathcal{E}_b/I_0$ requirement β for each user.

2. Plot this expression as a function of N for reasonable values of the parameters and compare this to the perfect power control case. Do you see any interference averaging effect?
3. How does this scenario differ from the users' activity averaging example considered in the text?

Exercise 4.13 In the downlink of a CDMA system, each users' signal is spread onto a pseudonoise sequence.[13] Uncoded BPSK modulation is used, with a processing gain of G. Soft handoff is performed by sending the same symbol to the mobile from multiple base-stations, the symbol being spread onto independently chosen pseudonoise sequences. The mobile receiver has knowledge of all the sequences used to spread the data intended for it as well as the channel gains and can detect the transmitted symbol in the optimal way. We ignore fading and assume an AWGN channel between the mobile and each of the base-stations.

1. Give an expression for the detection error probability for a mobile in soft handoff between two base-stations. You may need to make several simplifying assumptions here. Feel free to make them but state them explicitly.
2. Now consider a whole network where each mobile is already assigned to a set of base-stations among which it is in soft handoff. Formulate the power control problem to meet the error probability requirement for each mobile in the downlink.

Exercise 4.14 In this problem we consider the design of hopping patterns of neighboring cells in the OFDM system. Based on the design principles in Section 4.4.2, we want the hopping patterns to be Latin squares and further require these Latin squares to be orthogonal. Another way to express the orthogonality of a pair of Latin squares is the following. For the two Latin squares, the N_c^2 ordered pairs (n_1, n_2), where n_1 and n_2 are the entries (sub-carrier index) from the same position in the respective Latin squares, exhaust the N_c^2 possibilities, i.e., every ordered pair occurs exactly once.

1. Show that the $N_c - 1$ Latin squares constructed in Section 4.4.2 (denoted by R^a in (4.23)) are mutually orthogonal.
2. Show that there cannot be more than $N_c - 1$ mutually orthogonal Latin squares. You can learn more about Latin squares from a book on combinatorial theory such as [16].

Exercise 4.15 In this exercise we derive some insight into the PAPR of the uplink transmit signal in the OFDM system. The uplink signal is restricted to n of the N_c sub-carriers and the specific choice of n depends on the allocation and further hops from one OFDM symbol to the other. So, for concreteness, we assume that n divides N_c and assume that sub-carriers are uniformly separated. Let us take the carrier frequency to be f_c and the inter-sub-carrier spacing to be $1/T$ Hz. This means that the passband transmit signal over one OFDM symbol (of length T) is

$$s(t) = \Re\left[\frac{1}{\sqrt{N_c}} \sum_{i=0}^{n-1} \tilde{d}_i \exp\left(j2\pi\left(f_c + \frac{iN_c}{nT}\right)t\right)\right], \qquad t \in [0, T].$$

---

[13] Note that this is different from the downlink of IS-95, where each user is assigned an orthogonal sequence.

Here we have denoted $\tilde{d}_0, \ldots, \tilde{d}_{n-1}$ to be the data (constellation) symbols chosen according to the (coded) data bits. We also denote the product $f_c T$ by ζ, which is typically a very large number. For example, with carrier frequency $f_c = 2\,\mathrm{GHz}$ and bandwidth $W = 1\,\mathrm{MHz}$ with $N_c = 512$ tones, the length of the OFDM symbol is approximately $T = N_c/W$. Then ζ is of the order of 10^6.

1. What is the (average) power of $s(t)$ as a function of the data symbols $\tilde{d}_i$, $i = 0, \ldots, n-1$? In the uplink, the constellation is usually small in size (due to low SINR values and transmit power constraints). A typical example is equal energy constellation such as (Q)PSK. For this problem, we assume that the data symbols are uniform over the circle in the complex plane with unit radius. With this assumption, compute the average of the power of $s(t)$, averaged over the data symbols. We denote this average by P_{av}.
2. We define the peak power of the signal $s(t)$ as a function of the data symbols as the square of the largest absolute value $s(t)$ can take in the time interval [0, *T*]. We denote this by $PP(\tilde{\mathbf{d}})$, the peak power as a function of the data symbols $\mathbf{d}$. Observe that the peak power can be written in our notation as

$$PP(\tilde{\mathbf{d}}) = \max_{0 \le t \le 1} \left(\Re\left[\frac{1}{\sqrt{N_c}} \sum_{i=0}^{n-1} \tilde{d}_i \exp\left(\mathrm{j}2\pi \left(\zeta + \frac{iN_c}{n} \right) t \right) \right] \right)^2 .$$

   The peak to average power ratio (PAPR) is the ratio of $PP(\tilde{\mathbf{d}})$ to P_{av}.
   We would like to understand how $PP(\tilde{\mathbf{d}})$ behaves with the data symbols $\tilde{\mathbf{d}}$. Since ζ is a large number, $s(t)$ is wildly fluctuating with time and is rather hard to analyze in a clean way. To get some insight, let us take a look at the values of $s(t)$ at the sample times: $t = l/W, l = 0, \ldots, N_c - 1$:

$$s(l/W) = \Re[d[l]\exp(\mathrm{j}2\pi\zeta l)],$$

   where $(d[0], \ldots, d[N_c - 1])$ is the N_c point IDFT (see Figure 3.20) of the vector with ith component equal to

$$\begin{cases} \tilde{d}_l & \text{when } i = lN_c/n \text{ for integer } l, \\ 0 & \text{otherwise.} \end{cases}$$

   The worst amplitude of $s(l/W)$ is equal to the amplitude of $d[l]$, so let us focus on $d[0], \ldots, d[N_c - 1]$. With the assumption that the data symbols $\tilde{\mathbf{d}}_0, \ldots, \tilde{\mathbf{d}}_{n-1}$ are uniformly distributed on the circle in the complex plane of radius $1/\sqrt{N_c}$, what can you say about the marginal distributions of $d[0], \ldots, d[N_c - 1]$? In particular, what happens to these marginal distributions as $n, N_c \to \infty$ with n/N_c equal to a non-zero constant? The random variable $|d[0]|^2/P_{\mathrm{av}}$ can be viewed as a lower bound to the PAPR.
3. Thus, even though the constellation symbols were all of equal energy, the PAPR of the resultant time domain signal is quite large. In practice, we can tolerate some codewords having large PAPRs as long as the majority of the codewords (say a fraction equal to $1 - \eta$) have well-behaved PAPRs. Using the distribution

$|d[0]|^2/P_{\text{av}}$ for large n, N_c as a lower bound substitute for the PAPR, calculate $\theta(\eta)$ defined as

$$\mathbb{P}\left\{\frac{|d[0]|^2}{P_{\text{av}}} < \theta(\eta)\right\} = 1 - \eta.$$

Calculate $\theta(\eta)$ for $\eta = 0.05$. When the power amplifier bias is set to the average power times θ, then on the average 95% of the codewords do not get clipped. This large value of $\theta(\eta)$ is one of the main implementational obstacles to using OFDM in the uplink.

Exercise 4.16 Several techniques have been proposed to reduce the PAPR in OFDM transmissions. In this exercise, we take a look at a few of these.

1. A standard approach to reduce the large PAPR of OFDM signals is to restrict signals transmitted to those that have guaranteed small PAPRs. One approach is based on Golay's complementary sequences [48, 49, 50]. These sequences possess an extremely low PAPR of 2 but their rate rapidly approaches zero with the number of sub-carriers (in the binary case, there are roughly $n \log n$ Golay sequences of length n). A reading exercise is to go through [14] and [93] which first suggested the applicability of Golay sequences in multitone communication.
2. However, in many communication systems codes are designed to have maximal rate. For example, LDPC and Turbo codes operate very close to the Shannon limits on many channels (including the AWGN channel). Thus it is useful to have strategies that improve the PAPR behavior of *existing* code sets. In this context, [64] proposes the following interesting idea: Introduce fixed phase rotations, say $\theta_0, \ldots, \theta_{n-1}$, to each of the data symbols $\tilde{d}_0, \ldots, \tilde{d}_{n-1}$. The choice of these fixed rotations is made such that the overall PAPR behavior of the signal set (corresponding to the code set) is improved. Focusing on the worst case PAPR (the largest signal power at any time for any signal among the code set), [116] introduces a geometric viewpoint and a computationally efficient algorithm to find the good choice of phase rotations. This reading exercise takes you through [64] and [116] and introduces these developments.
3. The worst case PAPR may be too conservative in predicting the bias setting. As an alternative, one can allow large peaks to occur but they should do so with small probability. When a large peak does occur, the signal will not be faithfully reproduced by the power amplifier thereby introducing noise into the signal. Since communication systems are designed to tolerate a certain amount of noise, one can attempt to control the probability that peak values are exceeded and then ameliorate the effects of the additional noise through the error control codes. A probabilistic approach to reduce PAPR of existing codesets is proposed in [70]. The idea is to remove the worst (say half) of the codewords based on the PAPR performance. This reduces the code rate by a negligible amount but the probability (η) that a certain threshold is exceeded by the transmit signal can be reduced a lot (as small as η^2). Since the peak threshold requirement of the amplifiers is typically chosen so as to set this probability to a sufficiently small level, such a scheme will permit the threshold to be set lower. A reading exercise takes you through the unpublished manuscript [70] where a scheme that is specialized to OFDM systems is detailed.