Chapter 1

Introduction

1.1 Overview

The radio networks that are the subject of this thesis consist of a network of base stations, a large number of network users and two links of communication: user to base station (uplink) and base station to user (downlink). The objective of this thesis is to formulate and characterise various notions of network capacity and to investigate how the capacity can be allocated efficiently to the users. We concentrate almost exclusively on the uplink in this thesis, and the omission of downlink theory requires some justification. We leave the justifications to the introductions to Parts I and II, as the reason in Part I is different to the reason in Part II.

In Part I, we find fundamental constraints on the uplink capacity provided by the multi-user version of Shannon's theory of information applied to a channel consisting of a whole network of radio receivers (base stations). In Part II, we characterise the capacity of a communication model that is of interest to current applications. Part II contains an assumption to the effect that when one of the users in the channel is being decoded, the signals of the others are treated as random noise. This assumption is not made in the Shannon models in Part I. In Part II we consider the further problems of allocating the capacity equitably amongst the users in a dynamic environment in which users' demands on the network are random.

There are two ways to view the capacity problem in a radio network: one is to ask at what bit rates a group of users can transmit, and the other is to ask, given the rates, how many users can be accommodated in the network? We begin in Part I with a model in which the number of users and their positions in the network are fixed and we consider the first notion of capacity. Shannon theory imposes bounds on the bit rates that the users can obtain, taking into account their mutual interference. A conclusion drawn is that it is optimal for such networks to be spread spectrum in nature. By this we mean that each user should utilise the full radio bandwidth allocated to mobile radio so that all users are essentially sharing the same channel. We also use the theory to find spread spectrum schemes that approach the Shannon bounds, namely the interference cancellation schemes considered in Chapter 5.

Towards the end of Part I, in Section 6.3.1, we take the second approach to capacity and fix the bit rate requirements of the users. We use Shannon theory to bound the number of users in different positions in the network given that the bit rate requirements must be met.

In Part II we characterise the capacity in users of spread spectrum models in which interference from other users is treated as random noise (the information in the interference is not exploited). Interference fluctuations are mitigated by users independently controlling their transmitter power levels; by increasing or decreasing power, a user can increase or decrease its share of network capacity. Power control algorithms are modelled by ordinary differential equations. In spite of power control, there is inherent randomness in a radio network, but fortunately statistical multiplexing occurs in large scale spread spectrum systems. Large deviations theory is therefore applicable to the derivation of network capacity.

To discuss the issues in any more detail we must introduce some terminology and to this end we provide an overview of mobile radio as it exists today. In this overview we also make a case for why we should be
interested in spread spectrum in radio networks.

1.2 Mobile radio, cellular networks and macrodiversity

The current mobile radio systems use frequency division multiple access (FDMA) for their transmission scheme. That is, the frequency spectrum is divided into separate subchannels, each consisting of a fixed number of frequencies. The notions of bandwidth and frequency spectrum will be made precise in Chapter 3, but it is intuitively clear that the rate of information flow, in bits/sec, will be limited by the number of frequencies available. Each user occupies a subchannel which has enough frequencies for the user to achieve the desired data rate in bits per second. This type of communication is called narrowband communication because each subchannel has only a small fraction of the total frequency spectrum allocated to the whole channel. In contrast to this is spread spectrum communication in which each user has access to the full frequency bandwidth. The partitioning of users into separate narrowband subchannels is to avoid interference. However, signal strengths degrade with distance and with the presence of obstacles such as hills and buildings and so more than one user can occupy the same subchannel provided that they are sufficiently far apart. This degradation in signal strength is called “fading”. In the current system the entire geographical area is divided into cells large enough so that users can be allocated the same subchannel provided that they are not in the same or adjacent cells. Thus in cellular radio the frequency spectrum is divided up over a cluster of adjacent cells and then reused in the next cluster. Since FDMA is the basic multiple access technique here, and frequencies are reused at great enough distances, we call this scheme FDMA/reuse. Users that are close to each other are separated in bandwidth. We remark that it is only spread spectrum that has the potential for full frequency reuse in each cell (as we shall see). The FDMA/reuse cellular radio network is described in detail in Lee[37]. The fading characteristics of mobile radio are summarised in Cooper and Nettleton [9].

We remark that the term “fading” is often used for the self interference which results when various reflections of a signal interfere at a receiver. We shall refer to this as “multipath fading” in the present thesis.

In cellular radio each cell has a centrally located base station and each mobile in the cell sends to and receives from this station. The user is allocated two subchannels; one for the downlink (base station to mobile) and the other for the uplink (mobile to base station). The station is the interface between the mobile and the telephone network. When a mobile moves into another cell a process called handover takes place in which the user is allocated a new subchannel in the new cell.

FDMA/reuse requires nearby users to be separated in bandwidth and the reuse distance must be based on worst case considerations. To see why this is the case, consider the following hypothetical FDMA system in which the full bandwidth is used in each cell. The hexagonal cell structure is as depicted in Figure 1.1.

The interference at the base station in cell 0 to a user’s signal will come from six interferers, one located in each of the cells labelled 1, …, 6. All six interferers may be weak, but they may equally all be strong, and from the point of view of error control coding we must assume the latter. A consequence is that capacity is wasted when the interferers are weak. Apart from power fluctuations, the noise is clearly nonwhite; to take advantage of the dependencies inherent in the noise is an extremely complex decoding task. The cellular FDMA/reuse approach to interference wastes capacity in a similar way: the reuse partitioning distance must be based on worst case considerations and much of the time interferers could be closer together. The main advantage of the reuse approach is simplicity of decoding. In the hexagonal cell layout a standard reuse scheme is to give each cell one seventh of the total bandwidth and then reuse the bandwidth in clusters of seven cells over the entire network. Other reuse schemes can also be considered; it may be possible to bring the value down as low as four with strong error control coding. The danger in reducing the reuse distance is that it must be compensated by extra redundancy in the error control coding and it is likely that the capacity or the quality of service will suffer.

We remark that time division multiple access (TDMA) is another way to form narrowband channels. In TDMA, time is partitioned into slots in the same way that the frequency spectrum is partitioned in FDMA.
Each user is allocated a slot, and slots can be reused if the users are sufficiently far apart. From a Shannon-theoretic point of view, power constraints affect TDMA more severely than FDMA (Gallager [18]), but from a practical point of view they can be treated as interchangeable as far as most of the issues that we are concerned with in the present thesis.

An alternative to FDMA is to use spread spectrum communication in which each user has access to the full bandwidth rather than just a narrow sub-band. The main objective is to allow averaging of interference, so that worst case interference is not much worse than average interference. Spread spectrum has been used in military communications for many years in the context of communication in the presence of a jammer, but it was not until 1978 that it was first proposed for mobile radio. Many of the advantages of spread spectrum over narrowband FDMA are considered in the seminal paper by Cooper and Nettleton[9]. These are discussed in a less technical setting in Cooper et. al. [8].

There are two basic approaches to spread spectrum communication and both allow a user access to the full bandwidth in each cell. One is “fast frequency hopping” in which at any given time a user is transmitting in a narrowband, but over time the user rapidly jumps from subchannel to subchannel, thereby spreading its signal over the full bandwidth. Another approach is called direct sequence spread spectrum, or code division multiple access (CDMA). In this approach each user is assigned a codeword. The transmission is digital and so we can think of the codeword as a sequence of binary digits which are then modulated by the user’s data. We might think of a user as generating a relatively slow stream of 1s and 1s representing the voice information to be sent. At the same time it is generating a very fast stream of 1s and 1s representing the code bits. The length of the code is the number of code bits that are generated in the time of one data bit (the spreading factor); if the data bit is 1 then the code is transmitted directly and if the data bit is −1 then the code is inverted and transmitted (−1s flip to 1s and vice versa). In frequency terms, the fact the rate of code bits is much higher than the rate of data bits means that the bandwidth occupied is much higher than if we just tried to send the data bits alone. The increase in bandwidth is given precisely by the spreading factor.

It is not the purpose of the present section to make a case for or against spread spectrum, although the issues explored in the present thesis lead the author to the conclusion that spread spectrum is superior
to FDMA/reuse. Nevertheless, let us suppose that spread spectrum is chosen as the method of multiple access. The cellular structure can be retained for spread spectrum, and indeed we consider cellular spread spectrum models in Chapter 8. However, we do not assume that cells are partitioned in bandwidth in spread spectrum systems; each cell is allocated the full radio bandwidth. It follows that a user near a cell boundary creates much more interference at the adjacent cell sites than a user near the centre of a cell, and this interference can be very damaging (Hanly and Whiting [26]). It may be more appropriate to allow such a mobile to transmit to several cell sites simultaneously, to turn the damaging interference into a source of information. The term “macrodiversity” is used to describe the situation in which several base station receivers are combined to decode a signal from a mobile. In the most general macrodiversity model, each user is decoded by all the receivers in the network. In this case, the cellular structure has completely disappeared. In less extreme cases, the cellular structure is retained, but users at cell boundaries are put in “cell site diversity mode” (Salmasi and Gilhousem [48]) and are jointly decoded by several cell sites together.

Interference is a fundamental problem in any radio network. Its cost may be reflected in the reuse distance in FDMA/reuse, or in detrimental interference in a spread spectrum system. A characteristic of this interference is that it is random: the level of interference from a mobile depends upon its position in the network and on its transmitter power level. This level in turn depends upon the user's bit rate and bit error rate performance requirements, and if there are many different classes of user, it may be impossible to predict in advance what the bit rate and bit error rate requirements of a user will be. Voice users are not continuously active, but alternate between active and idle states, the proportion of activity being referred to as the voice activity factor (typically taken to be about 40% Brady [5]). In the future, traffic may be even more bursty. The search for efficient transmission schemes for mobile radio must take into account the inherent randomness, and the fact that as demand increases at a great rate, so does the variety of services offered and the stringency of the performance requirements of the users.

1.3 Summary of Part I

In Part I, Chapter 4, we measure the inefficiency of FDMA/frequency reuse schemes by comparing the frequency reuse capacity with the multi-user Shannon capacity of the network of receivers. The capacity considered is bit rate capacity; we consider a network of receivers to which a large number of users transmit, and we assume that all users require the same bit rate $R$ bits/sec. Shannon theory provides bounds on $R$, and the tightest bound gives us the Shannon multi-user capacity. Macrodiversity is assumed in this approach. We prove general theorems that show that frequency reuse partitioning is suboptimal. In specific examples we compute the Shannon capacity and compare it with the capacity under the additional constraint that frequency reuse is employed.

The result that FDMA/reuse is suboptimal adds impetus to the search for Shannon-optimal schemes. We prove in Chapter 4 that spread spectrum schemes exist which attain the Shannon bounds. It should be noted that in Part I we do not restrict ourselves to CDMA or fast frequency hopping, as described in Section 1.2. In Part I, “spread spectrum” shall refer to any scheme which has the power spectrum of each user spread over the full channel bandwidth. In such a scheme a user receives interference from all users in the system at all times. In Chapter 5 we find examples of Shannon-optimal schemes. In Chapter 6, the capacity of the network in terms of numbers of users is also addressed.

1.4 Summary of Part II

The models considered in Part II are Shannon suboptimal because information contained in the interference from other users is ignored. However, these schemes are simple to implement and are of current interest (both CDMA and fast frequency hopping are examples of such schemes). Moreover, they illustrate many of the advantages that spread spectrum accrues from statistical multiplexing. In these models, many users are multiplexed onto a wide bandwidth channel, and, as a result of economies of scale, stringent quality of
service requirements can be met. The main reason is that interference levels are averaged, and consequently
the total interference to a mobile's signal does not fluctuate to the extent that it does in a narrowband
system.

We consider two such models in Part II. In Chapter 8, we consider a cellular model and characterise its
capacity in terms of numbers of users. The capacity region turns out to be nonlinear and can be specified
in terms of a constraint on the eigenvalue of a random matrix. Large deviations theory is used to deal with
the randomness inherent in the model. For a configuration of users to be feasible, it is necessary for them
to find the right transmitter power levels. We use ordinary differential equations to model power control
algorithms, and in Chapter 8, the system turns out to be linear.

In Chapter 9 we consider a macrodiversity model of spread spectrum in which each user is decoded
by all the receivers in the network. In this case there is a well-defined capacity of the network, in the
sense that the capacity can be specified by a single number. If voice activity is neglected, then there is
no randomness to be accounted for. Randomness arises if we consider voice activity, or if we introduce
maximum transmitter power limitations on the users. Although the capacity region is linear in Chapter 9,
the power control algorithms turn out to be nonlinear; these are studied in Chapter 10. Insight is obtained
from the theory of nonlinear, compartmental systems of ordinary differential equations.

Most of the work in Parts I and II assumes that all users require the same bit rate \( R \) bits/sec, and this
keeps the notation reasonably simple. There is no fundamental reason why the model cannot be extended
to treat several classes of users, each with its own bit rate requirement. Indeed, we do take this approach
for the single cell case in Chapter 8, and conjecture the multi-receiver generalisation of these results in
Chapter 9. Large deviations techniques can easily be extended to deal with multi-class systems (Hanly
and Whiting [26]) and we indicate how the Chernoff bound can be applied to such systems in Chapter 9,
Section 9.7.