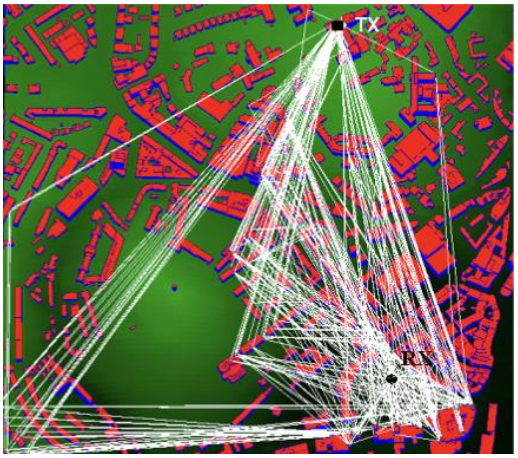
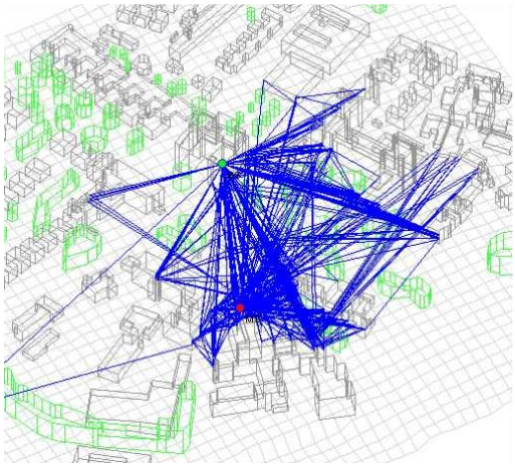
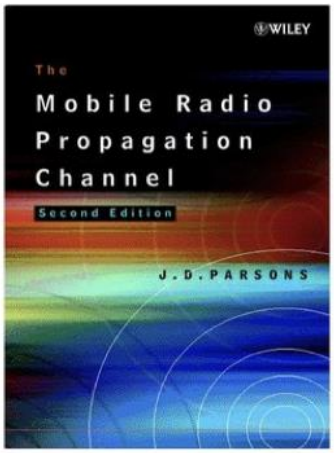
University of BRISTOL

Mobile Communication Systems





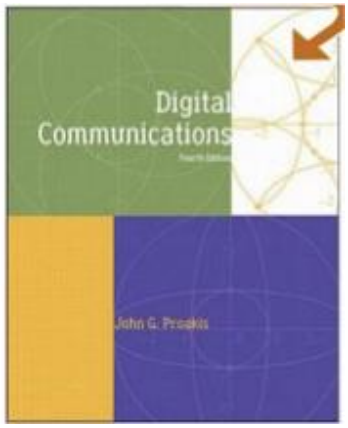
- Recommended Book List -



The Mobile Radio Propagation Channel, 2nd Edition J. D. Parsons, October 2000.

ISBN: 978-0-471-98857-1 (436 pages)

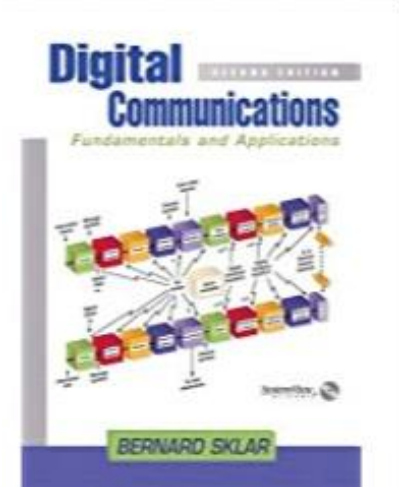
This book focuses on the mobile radio channel. It is very relevant for part 1 of this unit. It is probably best to borrow this book from the library as required.



Digital Communications, 4th Edition, McGraw-Hill, John Proakis, December 2000. (Also 5th edition-2007)

ISBN: 0071181830 (1024 pages)

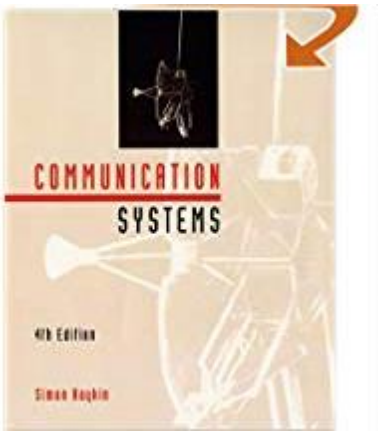
Covers material relevant to parts 1 and 2 of this unit. An excellent standard reference used all over the world. It is very mathematical and not recommended to initially understand new areas. Highly recommended for those planning a career in the mobile communications industry, or for those considering further postgraduate study.



Digital Communications: Fundamentals and Applications, 2nd Edition, Pentice Hall, Bernard Sklar, January 2001

ISBN: 0130847887 (1070 pages)

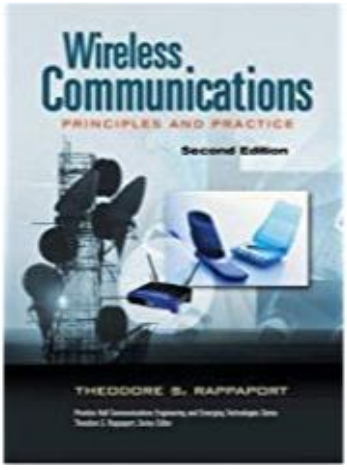
An excellent entry level book covering parts 1 and 2 of this unit. Much easier to understand than Proakis. Please note: the material from the propagation chapters is available on Blackboard.



Communication Systems, 4th Edition, John Wiley, Simon Haykin, June 2000. (Also 5th edition-2010)

ISBN: 0471178691 (816 pages)

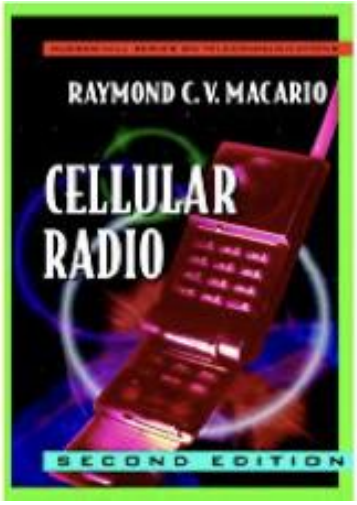
An excellent entry level book covering parts 1 and 2 of this unit. Strongly recommended for anyone interested in wireless communications. A very popular course book for US universities.



Wireless Communications: Principles and Practice, Prentice Hall, Theodore S. Rappaport, December 2001.

ISBN: 0130422320 (736 pages)

This book has been recommended by previous students.



Cellular Radio, 2nd Edition, R. C. V. MacArio, April 1997.

ISBN: 0070444331 (276 pages)

This book covers the material in part 3 of the unit. This book is not recommended for personal purchase unless you find this part of the unit difficult.

Aims

The aims and objectives of this handout can be broken down as follows:

Propagation Principles and Statistics

- To compute the received power density and received power in an urban environment
- To understand shadowing and to explain how suitable mathematical models are derived
- To explain the basic ideas behind urban multipath propagation
- To define Rayleigh and Rican amplitude and phase variations
- To introduce frequency selective fading, power delay profiles, rms delay spread, coherence bandwidth and intersymbol interference
- To understand how to compute the spaced-frequency, space-time and spaced-distance correlation function

Diversity

- To be able to list different methods for generating diversity signals
- To be able to explain the difference between switched and selection diversity
- To appreciate the advantages of equal gain and maximal ratio combining
- To be able to explain how diversity can improve performance

Cellular Design and Capacity

- To understand how continuous coverage can be achieved
- To be able to explain frequency re-use, handover and cluster size
- To understand the relationship between C/I and cluster size
- To be able to derive an expression linking the reuse ratio (D/R) to the cluster size (N)
- To be able to derive a simple expression linking the cluster size (N) to the C/I ratio
- To understand the definition of an Erlang and to be able to write down the equation for cellular capacity
- To understand the relationship between cellular capacity and cellular coverage
- To appreciate how microcells can increase capacity
- To be able to explain why microcells introduce new challenges

Radiowave Propagation

UNIVERSITY OF BRISTOL
Department of Electrical & Electronic Engineering

Mobile/Personal Communications

DEFINITIONS: - "communications with people on the move"
"phoning people *NOT* places"

All mobile services depend on the availability of radio spectrum. The radio frequency (RF) spectrum is a scarce resource, and the choice of operating frequency has a marked effect of the quality and type of service offered (as will be seen later in this unit). The spectrum is divided into a number of bands, the most common of which are listed below

Frequency Bands

DC	-	3 MHz	->	Low Frequency (LF)
3 MHz	-	75 MHz	->	High Frequency (HF)
75 MHz	-	225 MHz	->	Very High Frequency (VHF)
225 MHz	-	1 GHz	->	Ultra High Frequency (UHF)
1 GHz	-	Light	->	Microwave Frequencies and above

Examples of Wireless Communication Systems

MW Broadcast Radio	558-1611 kHz
FM Broadcast Radio	87-108 MHz
Analogue Television	470-790 MHz (Channels 21-68)
LTE	800 MHz
GSM	890-915 MHz (UL); 935-960 MHz (DL)
DECT	1885-1900 MHz
DCS1800	1710-1785 MHz (UL); 1805-1880 MHz (DL)
DCS1900	1850-1910 MHz (UL); 1930-1990 MHz (DL)
3G (W-CDMA)	1920-1980 MHz (UL); 2110-2170 MHz (DL)
WiFi	2.4-2483 GHz
LTE	2.6 GHz
5G (Microwave)	3.5 GHz (exact bands to be confirmed)
WiFi	5.15-5.35 GHz
Satellite TV	10.7-12.75 GHz
5G (mmWave)	24-28 GHz (57-64 GHz)

UL – uplink (communication from the handset to the basestation)
DL – downlink (communication from the basestation to the handset)

Some wireless communications systems operate with a single frequency allocation (i.e. WiFi), while others use two bands, one for the *uplink* and one for the *downlink* (i.e. GSM, W-CDMA, LTE). The spectrum is known as *paired* when two bands are supported. For single bands the allocation is *unpaired*. Paired allocations are used with systems based on *Frequency Division* Duplex (FDD). Unpaired allocations are used with systems based on *Time Division Duplex* (TDD).

(i) Full Duplex (ii) Half Duplex

Figure 1: Duplex Techniques

A channel that supports two way communications is known as *duplex*. A *full* duplex system is able to transmit and receive at the same time. A *half* duplex system can transmit and receive, but not at the same time. A one way communication system is known as *simplex*. FDD achieves two way communication by transmitting and receiving on different frequencies at the *same* time. The frequencies must be spaced sufficiently far apart (typically 40 MHz) to allow filters to separate the two bands (otherwise interference will occur). TDD achieves two way communication by transmitting and receiving at different times on the *same* frequency. A short UL burst is followed by a short DL burst. This pattern is repeated over and over again.

TDD and FDD should not be confused with FDMA and TDMA. These latter techniques describe how multiple access (MA) is achieved, i.e. how multiple users share the spectrum in time (TDMA) or in frequency (FDMA). There are other multiple access techniques; for example, the 3G standard makes use of Code Division Multiple Access (CDMA), however this is beyond the scope of this unit.

Multi-Band and Multi-Mode

Today, most mobile phones support multiple frequencies to enable operation on different networks and in different countries. These are typically referred to as multi-band phones. Recently multi-mode phones have become common, supporting systems such as LTE and WiFi.

Frequency Allocation

Three typical types of group make use of the radio spectrum:

Civilian
Broadcasters - for TV, radio

Military - Large bandwidths required to support anti-jamming and secure modulation systems

The military are often allocated large frequency bands. Traditionally, the military were amongst the first to control and regulate the radio spectrum, and many of their allocations are historical. In many countries (such as the UK), spectrum sharing with the military is becoming common. Military communications can occur at very high transmit powers, and in many cases, it is difficult to get clear information of the usage of the bands.

Broadcasters also enjoy large spectrum allocations, again for historical reasons. In particular, analogue terrestrial television was allocated 300 MHz of prime spectrum in the UK. Digital television allows a greater number of channels in a smaller allocation of radio spectrum. When analogue television was switched off in the UK (2012), a significant portion of the current television band was auctioned to support new mobile services.

Civilian systems must operate in bands made available by the regulators in each country. In many cases, large sums of money must be paid to the regulator to acquire a license. The famous 3G spectrum auction resulted in the successful UK operators paying approximately £4billlion pounds for the rights to operate a 3G cellular service. A further £4 billion pound was also required per operator to build and operate a network. Nowadays, spectrum licenses are vital since without them it is impossible to operate a network.

Civilian Wireless Communication Systems

Analogue

The analogue speech waveform modulates either the carrier amplitude, phase or frequency (AM, PM or FM). Examples shown in figure 2.

Cellular radio (Analogue cellular, known as Total Access Communications System, or TACS)
Cordless telephones (CT1, analogue cordless telephones)
Private Mobile Radio (PMR) (FM networks used by British Gas, Securicor, RAC etc.)

Analogue cellular radio was discontinued many years ago in the UK.


Figure 2: First generation portable phone, 'brick on a stick' mobile, Nokia TACS 101

Digital

Voice coders and decoders are used to convert the analogue waveform into a digital bit stream. The data is sent using a digital modulation format (for example QPSK).

Digital modulation was introduced to improve the spectrum efficiency of the network (and hence make more money for the operators). The digital decision process used to determine the most likely symbol at the receiver allows considerably more interference to be tolerated in a network. This allows channels to be reused over shorter distances, and in turn results in greater capacity (and more importantly revenue generation).

Digital modulation can also lead to better signal quality, since the decision process removes the influence of noise. However, if poor compression techniques are applied, or errors occur in the received data sequence, signal quality can degrade badly. It should be noted that many digital communication systems produce outputs that are inferior to their analogue counterparts. Taking digital television as an example, the picture quality depends of the type of content (movies, news, sports), the source data rate, the type of compression algorithm, and the quality of the communications link. Even today, football on digital television is often worse in terms of picture quality compared to analogue television. Subtle details such as grass patterns are poorly reproduced at low bit rates, especially for panning scenes.

Third Generation Systems

CDMA (IS-95, UMTS, W-CDMA, CDMA2000, EV-DO, HSPA)
GSM EDGE (Enhanced Data rates for GSM Evolution)

Third generation mobile systems were dominated by Code Division Multiple Access (CDMA) technology. This approach is also known as spread spectrum. This approach deliberately spread the users' voice and data signals over a large bandwidth to exploit the wideband properties of the mobile channel. CDMA is able to

efficiently support a large range of bearer rates, making flexible voice and data systems easier to implement. Supporters of CDMA claim the technology is able to manage interference more elegantly than earlier GSM systems. There are a number of variants of CDMA operating around the world.

Figure 3: Apple i-phone, *Figure4:Nokia N-95 (W-CDMA/GSM)*

Enhanced CDMA standards have also emerged, mainly for high speed data. Figure 3 shows several examples of 3G handsets.

GSM was also extended to support higher data rate services. Enhanced Data rates for GSM Evolution (EDGE) makes use of 8-PSK modulation and multi-time slot operation to increase data rates to 384 kbits/s.

Fourth Generation Systems

3G Long Term Evolution (LTE)

Fourth generation networks were introduced over the period 2008-2012. A new group, known as 3GPP2, was formed to develop fourth generation standards. The 3GPP2 standard is known as Long Term Evolution (LTE). This is based on OFDMA technology on the downlink, and Single Carrier Frequency Division Multiple Access (SC-FDMA) technology on the uplink.

Politics, Money, Regulations & Technology

In practice, technology tends to play a *small* role in the development of mobile communication systems. Since the majority of radio systems are regulated by Government, the role of politics cannot be underestimated. It can be a very time consuming and costly exercise to introduce a new type of communications system.

Radiowave Propagation

Traditionally, mobile radio services have been based on elevated base-stations (see figure 6) which communicate with a number of moving mobiles in the surrounding area (known as a point to multi-point network). Unfortunately, to support these highly mobile services, effects such as irregular terrain, user motion and operation in built-up areas must be considered. For indoor systems, factors such as transmission through walls, floors and ceilings also need to be taken into account. The following three propagation terms need to be carefully considered when designing a modern mobile radio communications system:

- Free Space Path Loss (or Attenuation)
- Multipath and Shadowing
- Time Delay Spread (Digital Transmissions)

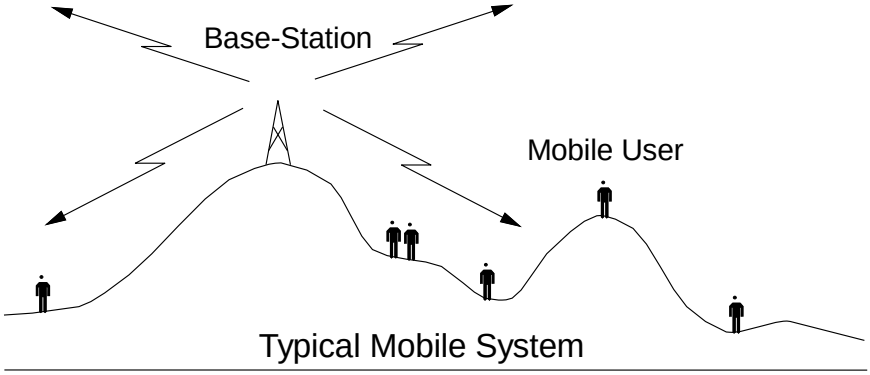


Figure 6: Current 'Macrocellular' Deployment

Free Space Path Loss (FSPL)

When a signal is transmitted from a point source it travels out in all directions on the surface of a sphere. As the radius of the sphere increases, so does its surface area. Since the transmit power is spread uniformly over the surface area, as the signal travels further from the source the power density (in terms of Watts per metre squared) reduces. Generally speaking, for a single path, the received power density is inversely proportional to the square of the receiver's distance from the transmitter (i.e. proportional to $1/d^2$). This dependency comes from the surface area of a sphere, which itself is proportional to $1/d^2$.

To compute the received power from the received power density we need to know the *effective area* of the receiving antenna. This is related to the operating frequency, and the physical construction of the antenna. The effective area of an antenna is

given by $G\lambda^2/4\pi$, where G represents the antenna gain (relative to an *isotropic* source), and λ represents the carrier wavelength. The carrier's frequency and wavelength are related by $c = f\lambda$, where f represents the carrier frequency and c the speed of light. The term isotropic means 'equal in all directions', and in this context an isotropic antenna is one that emits radiation equally in 3-D space. It should be noted that practical antennas tend to direct radiation in particular directions (such as the horizontal plane). When the transmit power is focused in particular directions, compared to an isotropic antenna, the received power density in those directions is higher. This results in an antenna gain G in those directions that is greater than one. In practice, the antenna gain is either provided by the antenna manufacturer or measured in an *anechoic* (a Greek work meaning no-echoes) chamber.

More precisely, the following equation shows the received power P_R at a distance *d* from the transmitter (A_i represents the effective area of an isotropic source (i.e. G =1) , λ represents the carrier wavelength and P_T the total transmit power in Watts).

$$P_R = \frac{P_T}{4\pi d^2} A_i = \frac{P_T}{4\pi d^2}\frac{\lambda^2}{4\pi} = P_T\left(\frac{\lambda}{4\pi d}\right)^2 \quad \text{(Watts)}$$

This equation assumes that the power is transmitted isotropically, i.e. 'equally in all directions'. In practice, isotropic antennas cannot be built and the above equation must be modified to take into account the real antenna gain patterns of the transmitter and receiver. We use the terms $G_T(\theta_T, \vartheta_T)$ and $G_R(\theta_R, \vartheta_R)$ to represent the antenna gains (relative to an isotropic source) as a function of the azimuth (θ) and elevation (ϑ) angle. Hence, the above equation can be rewritten

$$P_R = P_T G_T(\theta_T, \vartheta_T) G_R(\theta_R, \vartheta_R)\left(\frac{\lambda}{4\pi d}\right)^2 \quad \text{(Watts)}$$

The propagation path loss is defined as the ratio P_T / P_R.

$$\frac{P_T}{P_R} = 10\log\left[G_T(\theta_T, \vartheta_T) G_R(\theta_R, \vartheta_R)\left(\frac{\lambda}{4\pi d}\right)^2 \right]^{-1}$$

Free Space Path Loss (FSPL) is commonly quoted for isotropic antennas, in which case the antenna gain terms are omitted (since they are equal to unity).

$$FSPL = 10\log\left(\frac{4\pi d}{\lambda}\right)^2 = 10\log\left(\frac{4\pi df}{c}\right)^2$$

This is commonly quoted in decibels (dB).

$$FSPL = 20\log(d) + 20\log(f) + 20\log\left(\frac{4\pi}{c}\right)$$

$$FSPL = 20\log d + 20\log f - 147.56$$

In the above equation, d is in metres and f is in Hertz. Sometimes this equation is quoted for d in km, and f in MHz. In this case, we can write:

$$FSPL = 20\log d_{km} + 20\log f_{MHz} - 147.56 + 180$$

$$FSPL = 20\log d_{km} + 20\log f_{MHz} + 32.44$$

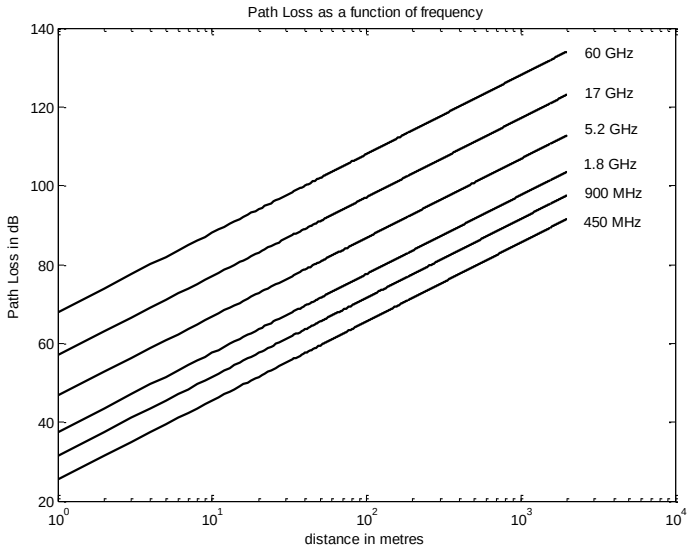


Figure 7: Path Loss as a function of Carrier Frequency and Distance

Example: A satellite transmitting at 8 GHz over a 36,000 km path will experience a free space path loss of approximately 200 dB. A cellular radio system operating at 900 MHz will experience a free space path loss of approximately 100 dB for a 2 km path.

The above equations are only useful for systems dominated by strong Line-Of Sight (LoS), i.e. satellite and microwave point-to-point communications; these equations cannot be applied when significant multipath components are present (i.e. mobile communications). For mobile radio applications, there is a large number of multipath components (see following section) generated by combinations of reflection, transmission, diffraction and scattering in the local environment.

Multipath Transmission, Reflection & Diffraction

The received signal at a mobile is made up from numerous attenuated, reflected, transmitted and diffracted versions of the original signal. This results in a phenomenon known as multipath fading, which can be divided into two categories, namely *slow* and *fast* fading.

The previous section described the signal attenuation with distance for line-of-sight (LoS) scenarios. Once LoS is lost, a phenomenon known as shadowing (or slow fading) becomes important.

Shadowing is the term given to the slow variations in received signal power as the user moves through the environment. Usually this can be attributed to the attenuation caused by large buildings, trees and/or nearby hills. As a result of the shadowing from urban clutter, the average signal strength is often assumed to follow an inverse fourth law rather than the inverse square normally seen in free space propagation.

Since shadowing reduces the received power at a given distance (or increases the path loss), it follows that an increase in the distance dependent exponent in the path loss and received power equations can achieve this effect. This modification is not based on any physical knowledge of the problem, but rather a pragmatic need to increase the path loss versus distance to take into account additional shadowing losses.

The value of the distance dependent exponent n is often determined by measuring the path loss as a function of distance (by driving around the cell) and by fitting the value of n that results in the lowest mean square error between the model and the observed data. The greater the shadowing, the higher the required value of the exponent. For LoS, the distance exponent takes a value of 2 (since there is no shadowing). For urban environments, n=3 is commonly used. For dense urban environments, n=4 is applied (i.e. the inverse fourth law mentioned previously).

For a distance dependent exponent n, the received power and path loss are given by

$$P_R = P_T G_T(\theta_T, \vartheta_T) G_R(\theta_R, \vartheta_R) \left(\frac{\lambda}{4\pi}\right)^2 \left(\frac{1}{d^n}\right) \quad \text{(Watts)}$$

$$PL = 10\log\left(\left(\frac{4\pi}{\lambda}\right)^2 d^n\right) = 10n\log d + 10\log\left(\frac{4\pi}{\lambda}\right)^2$$

Comparing the path loss equation to the straight line equation $y=mx+c$, we note that a graph of path loss in dB (y) versus *log d* (x) results in a straight line relationship with the gradient (m) equal to $10n$. Hence, to evaluate n we drive (or walk) around the cell of interest using GPS to record the mobile location, and hence the distance back to the basestation. We then produce a scatter plot of path loss (in dB) versus

log distance. Finally, by computing a line of best fit through this data, a value for the path loss exponent *n* can be evaluated. This equation models the mean path loss, and now incorporates the additional losses resulting from shadowing. From figure 8 we can see there is a significant variance around the straight line fit[1]. This is also modelled as an additional shadowing variance term. It can be shown that these variations (in their dB form) follow a zero-mean Normal distribution with variance σ_s^2 (typical values vary in the range 6-12 dB). This distribution is formally known as a log-Normal distribution.

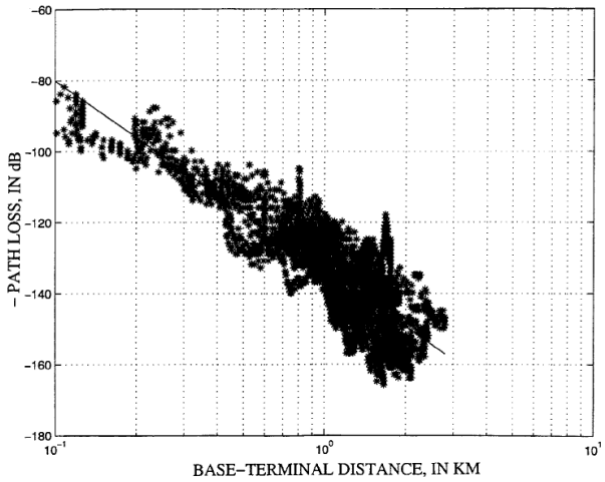


Figure 8: Example of path loss vs log distance scatter plot to determine n

Although shadowing can severely attenuate the received power level, its rate of change is generally low (<2 times per second). A second form of attenuation is known as *fast fading* and this can be attributed to the phasor addition of the various multipath signals. Figure 9 provides an example of fast and slow fading.

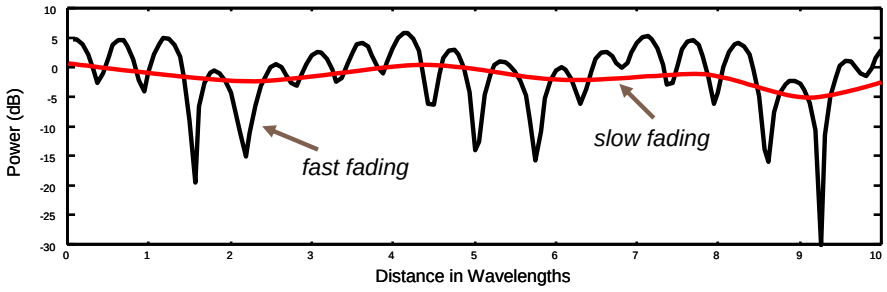


Figure 9: Fast and Slow Fading

If we assume that a sinusoidal signal is received through two independent paths, then the arrival phase for each of these rays will be different. Generally, the arrival phase

[1] Vinko Erceg et al, "An Empirically Based Path Loss Model for Wireless Channels in Suburban Environments, " IEEE Journal on Selected Areas in Communications, Vol. 17, No. 7, July 1999.

of a ray is determined by its electrical path length d and the carrier wavelength λ using the equation below, where θ_0 represents the initial phase of the ray.

$$\theta(d) = \theta_0 - \frac{2\pi d}{\lambda}$$

Generally, the equation for a given path is $y(d) = A(d)\cos(wt - \beta d)$, where $A(d)$ represents the distance dependent amplitude of the sinusoid, $w = 2\pi f$ represents the angular carrier frequency, $\beta = 2\pi/\lambda$ and t represents the elapsed time. As the user moves relative to the transmitter, the electrical path length varies and hence the phase of the ray alters. If the two rays arrive *in phase* then *constructive* interference will occur, resulting in a signal peak. However, if the phase of the two rays are in *anti-phase* (180 degrees apart) then *destructive* interference will occur and the two signals will cancel (or *fade*). In practice, there are usually many rays arriving at the receiver and the probability of total cancellation is small. A typical multipath scenario is shown in figure 10. The left-hand diagram shows a number of reflected and diffracted multipaths linking a transmitter (Tx) and receiver (Rx). The right-hand diagram shows the urban multipaths from a transmitter on the University's Wills building to a mobile at the bottom of Park Street. This output was generated from a prediction tool developed in the Department by a PhD student.

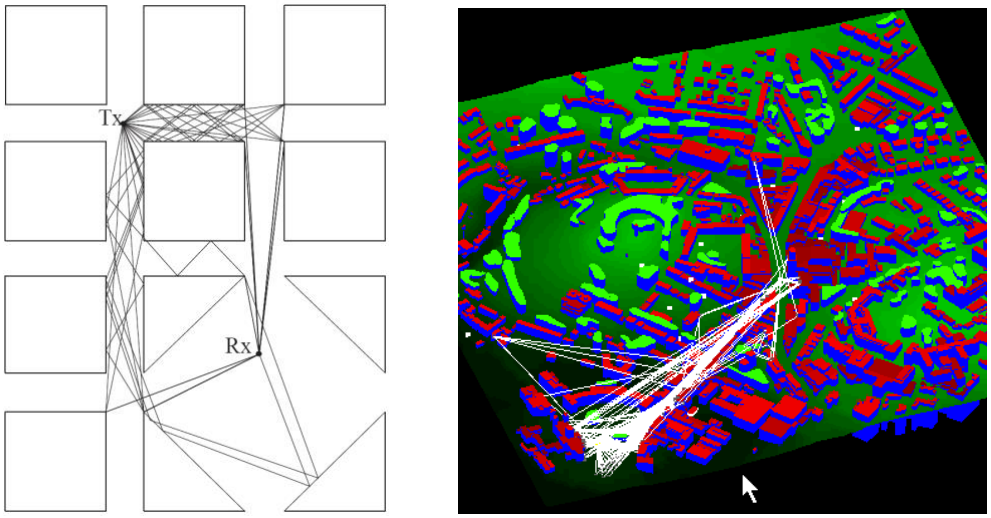
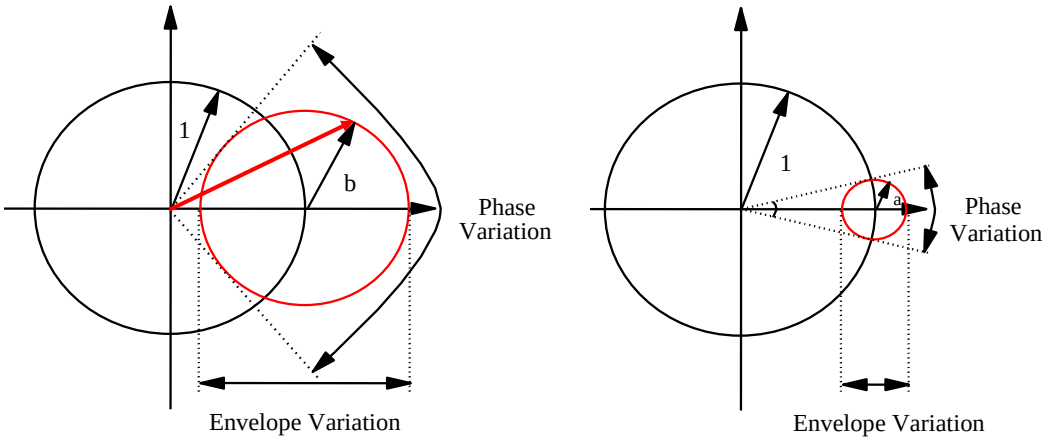


Figure 10: Example of Multipath Urban Multipath Propagation

The simplest multipath scenario is based on two arriving paths. If we consider the case where the two paths have a similar amplitude, then the resulting fading is representative of a *Rayleigh* scenario. In practice, to generate a good Rayleigh distribution at least 8 multipath must be modelled.

In many mobile scenarios one of the multipaths arriving at the receiver has far more power than the rest, and hence tends to dominate the resulting signal power. This is often the case in a LoS channel, although similar conditions can occur in NLoS, for example from a single strong reflector. In the case of a dominant path, the vector summation is far more deterministic (i.e. less variations) and the resulting envelope fading follows *a Rician* distribution. Figure 11 helps to illustrate the fading process.



(*i*) Rayleigh (*ii*) Rician

Figure 11: Rayleigh (Non-LoS) and Rician (LoS) Constructions

For the case of the Rayleigh distribution, the amplitude of the two rays are approximately equal (i.e. $b \approx 1$). The right hand circle (red in the electronic version of these notes) now represents the locus of the vector summation of the rays for all possible arrival phases. The resulting amplitude varies between the sum and difference of the two ray amplitudes. When the two rays have similar amplitudes, the resulting fading is very severe, with total cancellation possible when the two amplitudes are equal and the rays are in anti-phase. Interestingly, the phase variations of the resultant vector are also shown and can be seen to undergo a *maximum rate of change* during the bottom of a fade.

The second diagram shows a situation similar to a Rician fading channel. For this model, one of the rays is considerably larger than the other. The resulting amplitude and phase transients are now far less severe. As the amplitude ratio between the rays increases, the signal becomes more and more deterministic. For a Rician channel, the degree of variation is defined by the *K-factor*. Mathematically, the K-factor can be evaluated from the equation below, where A^2 and $2\sigma^2$ represent the envelope power of the dominant and random components respectively.

$$K = 10\log\left(A^2/2\sigma^2\right) \text{ dB}$$

For situations where no dominant component exists (Rayleigh), the K-factor will be small (and in some cases negative since we are using dB values). However, if a strong dominant path exists (Rician), the K-factor will normally take positive values between 5 and 10 dB.

The earlier two-ray models were used purely to explain the concepts of fading. In real channels, we tend to experience a large number of multipaths (normally well in excess of 8). Assuming an 8 ray model, typical fast fading variations are shown in figures 12 and 13 for Rayleigh and Rician Channels

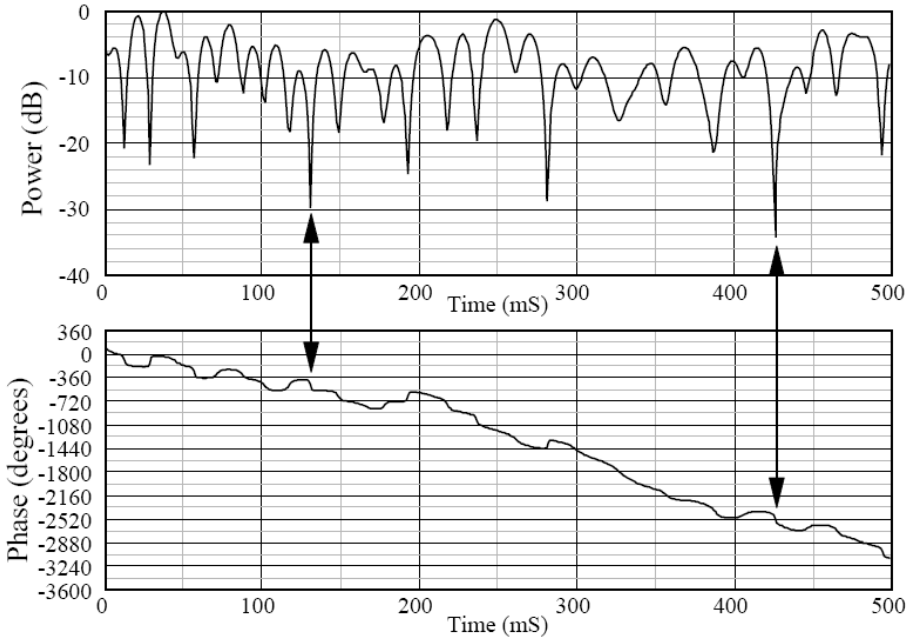


Figure 12: Rayleigh Envelope and Phase variations over time

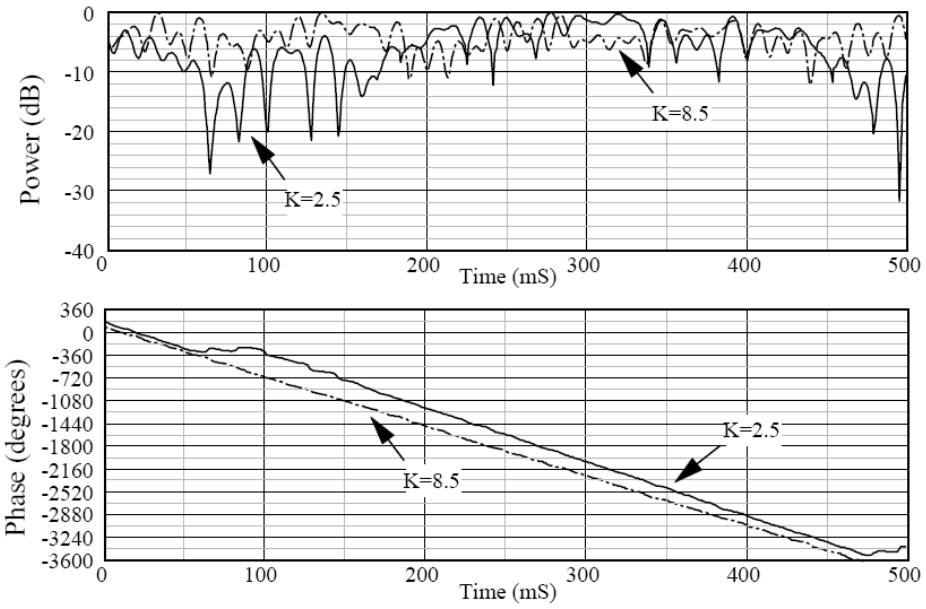


Figure 13: Rician Envelope and Phase variations over time (K=2.5 dB and K=8.5 dB)

In the Rician case, the severity of the fading is much less for the higher K-factor.

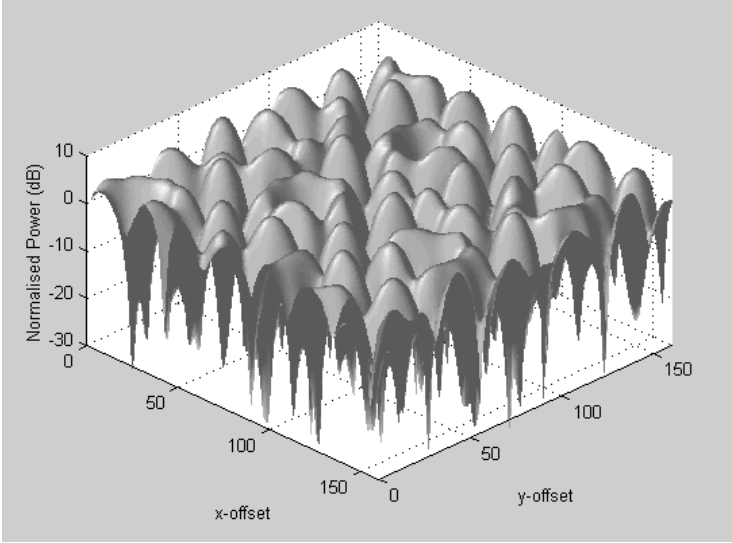


Figure 14: Rayleigh Fading over a rectangular surface

Figure 14 shows an example of the envelope fading that occurs over a rectangular area (such as a room). It is clear that small spatial movements result is massive changes to the signal envelope (and also significant phase rotations). Imagine trying to receive a 16QAM signal while walking around this environment. The signal's amplitude would vary violently with time. Typically, deep fades can be 40dB below the signal peaks. In these conditions, trying to determine the most likely transmit symbol is a major challenge.

Delay Dispersion

We have seen that the multipath signals arriving at a mobile receiver take unique paths, and as such they experience a unique time of flight. This time of flight will be different for each multipath, resulting in time dispersion between the multipaths at the receiver (see figure 15).

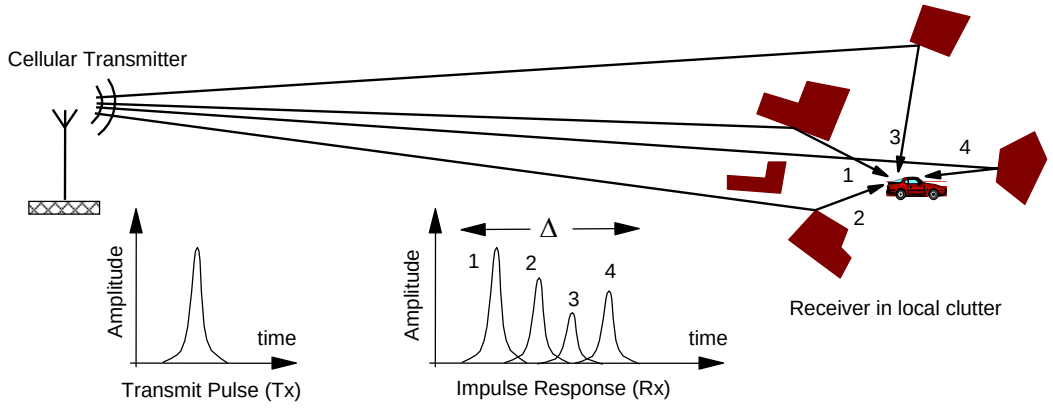


Figure 15: Time Dispersion between multipaths at Mobile

In the above figure, a single impulse is sent from a cellular transmitter. The signal bounces of a number of nearby buildings, resulting in four multipaths. The shortest path (multipath 1) has the shortest time of flight. The longest multipath (number 4) has the longest time of flight. As a result, four distinct pulses are seen at the receiver, resulting in a time spread denoted by Δ. If the time spread becomes large compared to the digital bit period then problems occur at the receiver. As the transmission rate increases, the bit period deceases, and at some point, the time dispersion in the channel becomes a limiting factor.

The time dispersive nature of a channel is often characterised by its Power Delay Profile (PDP). This is obtained by plotting the power of each multipath against its time of flight. In some cases, the time of flight is normalised relative to the time of the first arriving ray (in which case the time axis represents relative delay). A theoretic indoor PDP is shown in figure 16 (generated from a computer model for an indoor environment).

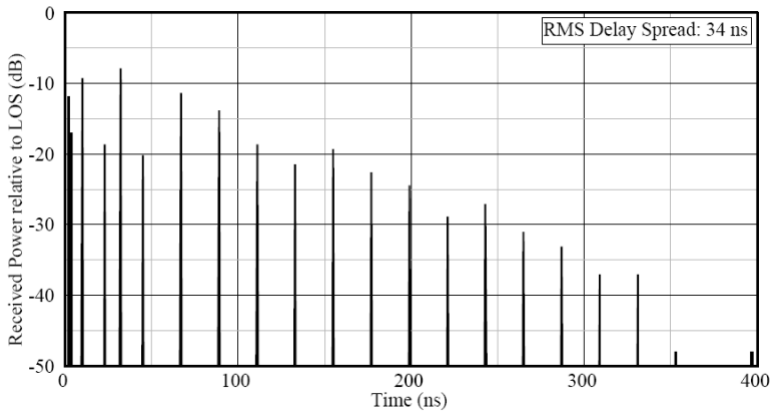


Figure 16: An example PDP (note the use of relative power and relative delay

Usually the first rays to arrive at the mobile have the largest amplitude (since they have travelled a shorter distance from the BS) and their magnitudes tend to fall away as the time delay increases (due to extra path loss and further reflection and diffraction loss). It is possible to quantify the degree of dispersion in a channel by evaluating its *Root Mean Square (RMS) delay spread*. This value can be calculated from the equations below, where α_k represents the received amplitude of the *k*-th ray after a time delay of τ_k seconds, and τ_a represents the time for half the power to arrive.

$$\tau_a = \frac{\sum_{k=1}^{N} \tau_k \alpha_k^2}{\sum_{k=1}^{N} \alpha_k^2}, \qquad \tau_{rms} = \sqrt{\frac{\sum_{k=1}^{N} [\tau_k - \tau_a]^2 \alpha_k^2}{\sum_{k=1}^{N} \alpha_k^2}}$$

This value incorporates all the information from a power delay profile and expresses it as a single value whose magnitude allows direct comparisons between differing channels and environments. In practice, since weak paths (which are not really significant at the receiver) with long delays can result in large rms delay spreads, it is common to discard ray paths below a power threshold relative to the strongest path. A threshold of 20dB is commonly used in practice. For digital transmissions, it is common to quote the value of the rms delay spread *normalised* to the transmission bit rate *R*, or the bit period T_b, as shown below.

$$d = \frac{\tau_{rms}}{\tau_n} = \tau_{rms} R$$

As a rule of thumb, an *unequalised* digital system generates a Bit Error Rate (BER) of around 1 in 1000 for a normalised rms delay spread of 0.1. The term *equalisation*

is used to describe the digital signal processing that must be performed to correct for time delay spread (or frequency selective fading). Digital modems that support high data rates always include some form of equalisation to overcome the harmful effects of delay spread. Equalisation is covered in other units.

The maximum bandwidth that can be supported *without the need for equalisation* in a time dispersive (or frequency selective) channel is referred to as the *coherence bandwidth*. This figure is loosely defined as the frequency separation required between two tones for the envelope fading correlation to fall to 0.5. Figure 17 shows the frequency domain channel response for the power delay profile described earlier.

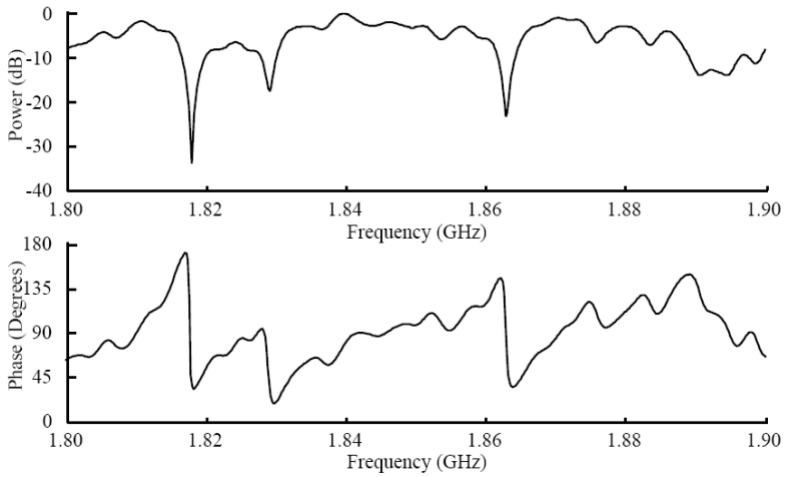


Figure 17: Channel Response (Power and Phase) as a function of Frequency

From figure 17 the frequency selective nature of the channel can be observed with deep notches with widths of approximately 5 to 7 MHz. The notch width is inversely proportional to the value of the rms delay spread. In the frequency domain, time delay spread creates deep frequency notches, whereas in the time domain it produces Inter Symbol Interference (ISI). Remember, if you apply a random notch filter to a digitally modulated waveform it is very likely that bit errors will result.

In some text books (such as Proakis), the PDP is assumed to follow an exponential profile. In this case, the coherence bandwidth is approximated by

$$B_c = \frac{1}{2\pi\tau_{rms}} \approx \frac{1}{6\tau_{rms}}$$

Inserting the earlier rms delay spread estimates into the above equation reveals that for 2-3 km GSM transmissions, the coherence bandwidth lies between 83-166 kHz (i.e. less than the GSM operating bandwidth). However, for shorter range transmissions (200m and 30m), the coherence bandwidth increases to approximately 833 kHz and 5.5 MHz respectively.

What is meant by a Wideband System?

Wideband transmission is the term used to describe a system where the channel bandwidth is far in excess of the channel's coherence bandwidth. If this is not the case, the system is known as narrowband.

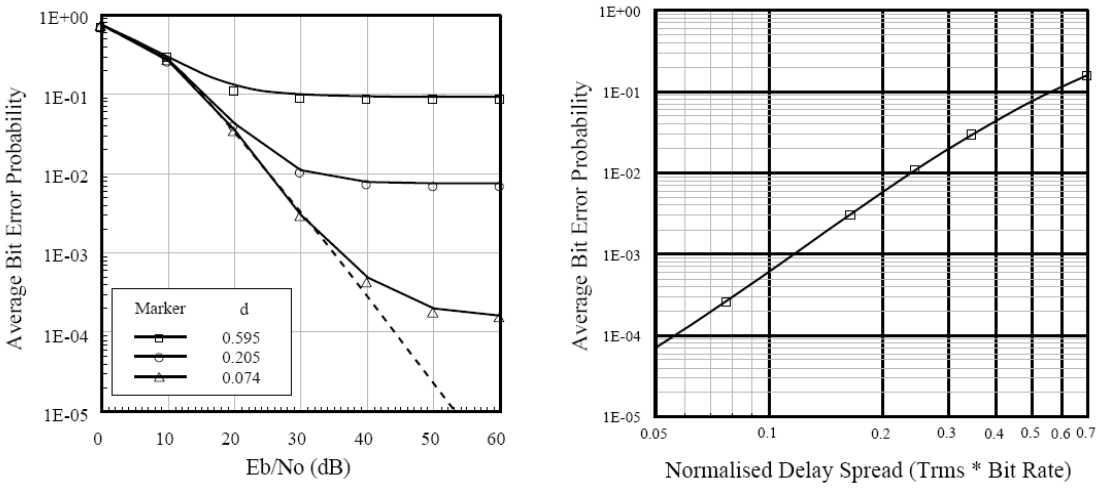


Figure 19: Modem Evaluation Graphs (see question below)

QUESTION

1. Assuming a maximum tolerable bit error rate of 1 in 1000, using the graphs in figure 19, calculate the maximum data rate that can be achieved for an rms delay spread of *i*) 50ns and *ii*) 2us.

2. If a data rate of *i*) 100kb/s, *ii*) 1Mb/s and *iii*) 10Mb/s is required for the previous system, what is the maximum tolerable rms delay spread?

Answers:

1. To determine the maximum data rate we assume a high SNR, such that errors are generated by delay dispersion and not thermal noise. Taking the 1 in 1000 BER threshold, and using the graph on the right, we find the maximum tolerable normalised rms delay spread in 0.11. We can now write

$$d_{nax} = T_{rms} R_{\max} = 0.11 \;\; => \;\; R_{\max} = \frac{0.11}{T_{rms}}$$

i) for $T_{rms} = 50ns$ we get $R_{\max} =$ 2.2 Mb/s,

ii) for $T_{rms} = 2us$ we get $R_{\max} =$ 55 kb/s.

2. We can write $T_{rms,\max} = \frac{0.11}{R_{\max}}$, and hence

i) for R_{max} = 100 kb/s we get T_{max} = 1.1 us,

ii) for R_{max} = 1 Mb/s we get T_{max} = 110 ns,

iii) for R_{max} = 10 Mb/s we get T_{max} = 11 ns.

Rayleigh Statistical Field Distribution

To create a Rayleigh distribution, we need to vector sum a number of paths where the amplitudes are similar, and the individual phases are uniformly distributed between 0 and 2π radians. Figure 20 shows a mobile receiver travelling from A to B at a constant speed *v*.

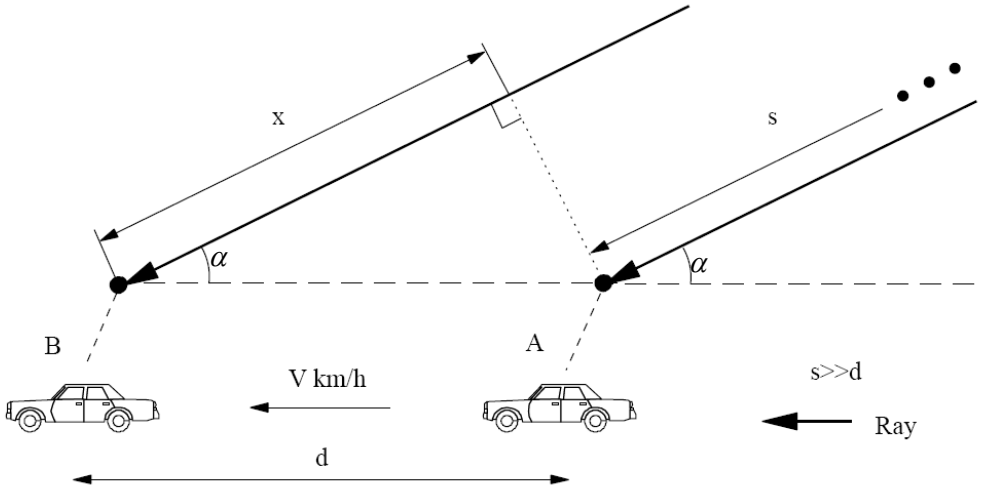


Figure 20: Mobile Reception of a Single Path

The distance moved by the mobile is represented by *d*, while *s* denotes the distance to the propagating source. If *s* is far greater than *d* then the angle of arrival in the azimuth plane for each ray can be assumed to remain constant throughout the motion (this results in a quasi-stationary model). For clarity, only one of the many rays arriving at the mobile is shown on the diagram. Furthermore, we ignore the elevation angle of the multipaths.

Assuming the transmission of a continuous wave (CW) sinusoidal signal, the received waveform *y(t)* at a mobile terminal takes the form:

$$y(t) = A\cos\left(2\pi ft + \phi + \vartheta_{Dop}(t)\right) \qquad (1)$$

where *A* represent the amplitude and ϕ the initial phase of the received waveform (i.e. at location *A*). The additional phase term, $\vartheta_{Dop}(t)$ represents the time varying phase that results as the terminal moves from A to B (this can also be written as a

distance based phase). For constant terminal velocity, $\vartheta_{Dop}(t)$ exhibits a constant rate of change of phase with time, (i.e. a fixed frequency shift). This frequency shift is better known as the *Doppler* frequency. As a terminal moves away from the basestation, the signal experiences a *negative* Doppler. As the terminal moves towards the basestation, it experiences a *positive* Doppler frequency. Next we derive a mathematical expression for this rate of change of phase (Doppler frequency) with time and/or distance.

At location B, the radio path back to the basestation has been extended in length (since the terminal is moving *away* from the BS) by a distance *x* (see diagram). From basic geometry, $x = d\cos\alpha$, where α represents the azimuth angle of arrival for a single path arriving at the terminal. The phase change for this single path as a function of the relative distance *d* (and hence *x*) from location A is given by

$$\theta_{Dop}(d) = -\beta x = \frac{-2\pi x}{\lambda} = \frac{-2\pi d\cos\alpha}{\lambda}$$

Assuming constant velocity, $v = d/t$ and assuming $c = f\lambda$ we can write

$$\theta_{Dop}(t) = \frac{-2\pi v f t\cos\alpha}{c}$$

The resulting frequency shift is given by

$$f_{Dop} = \frac{1}{2\pi}\frac{d\theta_{Dop}}{dt} = \frac{1}{2\pi}\left[\frac{-2\pi v f\cos\alpha}{c}\right] = -\left(\frac{vf}{c}\right)\cos\alpha$$

The above frequency shift is known as the Doppler frequency, and for a CW transmission the received signal is shifted by this value. The maximum value of the Doppler shift is given by

$$f_{\max} = \frac{vf}{c}$$

Note: Doppler frequency shifts apply to *single* multipaths. Each multipath is subject to a specific Doppler shift. The value of the shift can be positive or negative, depending on the direction of terminal motion compared to the azimuth arrival angle of the particular multipath. When multipaths arrive from different azimuth directions, for a given direction of motion, a range of different Doppler shifts are seen on the various multipaths, resulting in a frequency spread. The difference between the largest positive and largest negative shift amongst the multipaths is known as the Doppler frequency spread. It is also possible to compute the rms

Doppler spread in much the same way as the rms delay spread (i.e. by replacing the time delays with frequency shifts). Remember, a Doppler frequency spread applies to a *group* of multipaths that arrive at a specific point.

The concept of Doppler shift can be difficult to understand. Sometimes, an analogy can help. Imagine you are standing knee deep in water, with waves washing past you. If you stand still, you can measure the rate at which the waves pass by. This rate can be considered to be the transmission frequency. If you now run *towards* the waves, the rate at which the waves pass by will be slightly *higher* (a positive Doppler shift). If you run *away* from the waves, the rate at which the waves pass by will be slightly *lower* (a negative Doppler shift).

Another real world example of Doppler shift occurs with sound. Famously, the frequency (or pitch) of a whistle changes depending on the Doppler frequency. A train travelling towards you results in a higher pitched whistle (positive Doppler shift). A train travelling away from you results in a lower pitched whistle (negative Doppler shift).

To write a mathematical expression for Rayleigh fading over distance or time we must compute the vector sum of a number of multipaths with equal amplitudes and random instantaneous phases. If we assume there are L multipath components, we can write

$$r(d) = \sum_{n=1}^{L} A_n \exp j\left(\phi_n + \theta_n(d)\right) = \sum_{n=1}^{L} A_n \exp j\left(\phi_n - \frac{2\pi d \cos\alpha_n}{\lambda}\right)$$

$$r(t) = \sum_{n=1}^{L} A_n \exp j\left(\phi_n - \frac{2\pi v f t \cos\alpha_n}{c}\right) = \sum_{n=1}^{L} A_n \exp j\left(\phi_n - 2\pi f_{\max} t \cos\alpha_n\right)$$

To create a good *physical* Rayleigh fading model it is necessary for L to be greater than or equal to 8, for ϕ_n to be uniformly distributed between 0 and 360 degrees, for A_n to be equal for all n, and for α_n to be evenly distributed in the azimuth plane.

It is very important to be clear on the physical meaning of the various angles used in the equations above. For a given path, α represents the angle of arrival in the *azimuth* plane, ϕ represents the *initial phase* at the start of the motion, and θ represents the *additional phase shift* as a result of motion.

Figure 21 shows an example of the spatial fading generated from a simple Matlab model based on the above physical description.

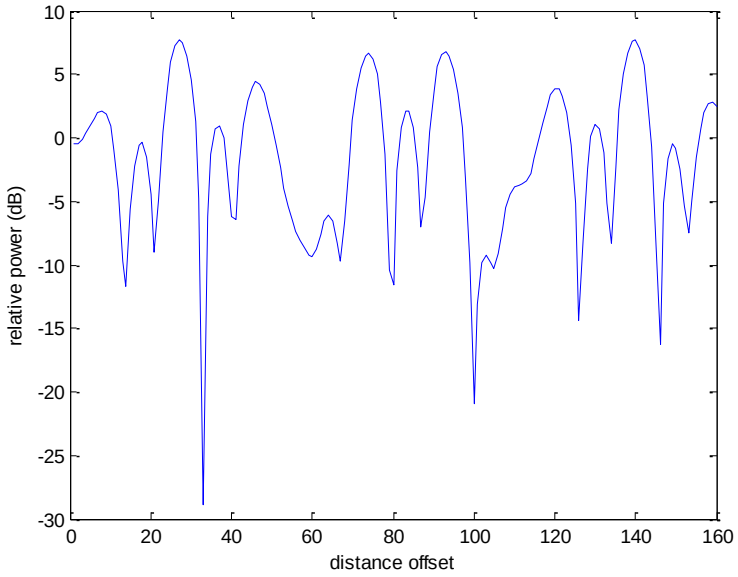


Figure 21: Example output from the Matlab Code

Listing

```
%   RAYLEIGH MODEL (EXAMPLE MATLAB CODE)
%   Andrew Nix, UoB, ©1997-2008

clear all;
pi=4*atan(1);                                % Define PI
raynum=8;                                    % Number of rays in model (L)
f=1800*1e6; c=3e8;                           % Frequency in Hz and Speed of light
length=10;                                   % fading simulation run in wavelengths
samps_per_wavelength=16;                     % Number of samples per wavelength
samp_num=length*samps_per_wavelength;        % Total number of samples

%   SET UP RAYLEIGH MODEL

for z=1:raynum;
    amp(z)=1;                                % Each ray defined with unity amplitude A
    arr_azimuth(z)=2*pi*((z-1)/raynum);      % Uniform azimuth distribution, alpha
    phase(z)=rand*2*pi;                      % Random phase angle, phi
end;

%   COMPUTE RAYLEIGH ENVELOPE

for x=1:samp_num
    vect=0;
    d=x/samps_per_wavelength;                % Offset distance in wavelengths
    for z=1:raynum;
       vect=vect+amp(z)*exp(j*(phase(z)-(2*pi*d*cos(arr_azimuth(z)))));
    end
    env(x)=abs(vect);
end

meanamp=mean(env);                           % Calculate mean envelope
for x=1:samp_num
    pow(x)=20*log10(env(x)/meanamp);         % Calculate power relative to mean
end;

figure(1); plot(pow);                        % Display power envelope
xlabel('distance offset'); ylabel('relative power (dB)');
```

Question: How would you modify the above to simulate a Rician Distribution?

Correlation Properties of the Fading Channel

Previously we computed the coherence bandwidth, which was based on the correlation of the fading envelope in the frequency domain. This parameter was shown to be inversely related to the channel's delay spread. Technically, the correlation of the fading envelope in the frequency domain is known as the spaced-frequency correlation function. As shown in figure 22, it can be computed from the Fourier Transform of the power delay profile (sometime known as the multipath intensity profile).

In addition to delay spread, we covered Doppler spread in the previous section. Doppler spread can be characterised by the Power Doppler Profile, which is also known as the Power Doppler Spectrum. Just as time spreading influences correlation in the frequency domain, frequency spreading influences correlation in the space/time domain (i.e. they are duals of one another). Correlation in the space/time domain is often characterised by the spaced-time correlation function, which can be computed as the Fourier Transform of the Power Doppler Spectrum (again a dual).

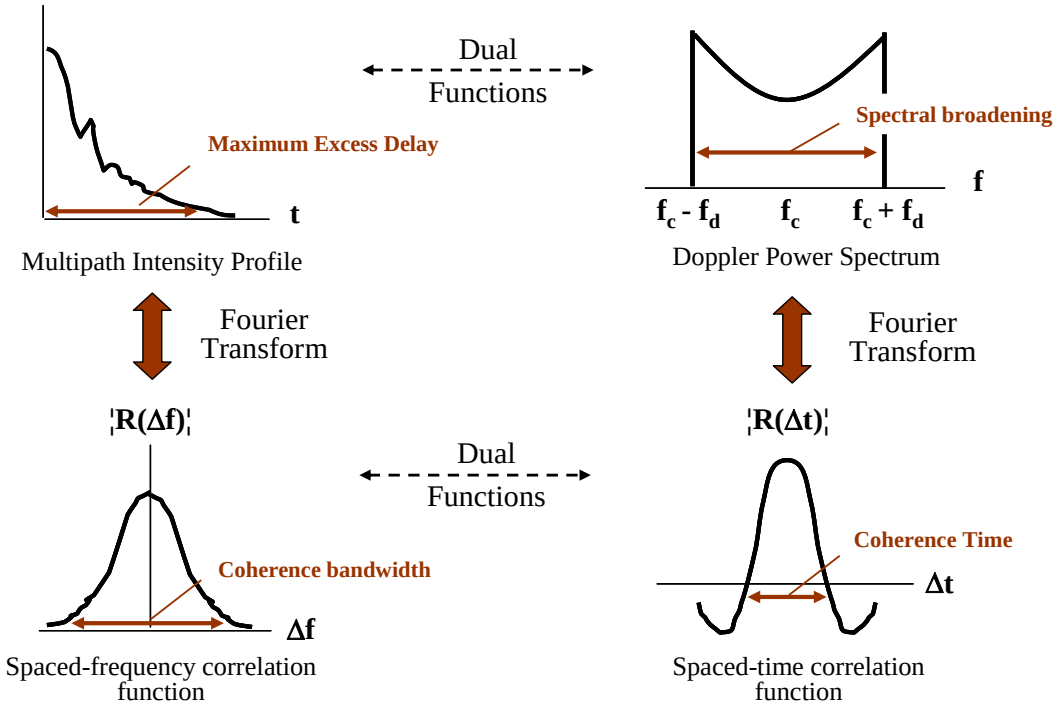


Figure 22: Correlation: Spaced-Frequency and Spaced-Time

The spaced-time correlation function determines how rapidly the envelope fading varies as a function of spatial, or for a fixed velocity temporal, offset. Since this correlation value is inversely related to the Doppler Spread, it follows from the earlier Doppler equations that it is also inversely related to the Azimuth Spread. The

spaced-time correlation function is used to compute both the coherence distance and the coherence time (for the latter it is necessary to define the terminal velocity).

In conclusion, we have introduced the coherence bandwidth, coherence time and coherence distance. The coherence bandwidth is inversely related to delay spread, and the coherence distance and time are inversely related to the Doppler spread (and also the Azimuth Spread).

Central Limit Theorem[2]

This theorem states that the Gaussian (or Normal) distribution is a good model for the cumulate effect of a large number of independent and identically distributed (i.i.d) random variables, regardless of the nature of their individual distributions. An expression for the Rayleigh random variable *r* can be derived *mathematically*, rather than *physically*, by invoking the central limit theorem.

Each multipath is a vector comprising a magnitude and a phase. It can thus be decomposed into its real and imaginary components. Via the Central Limit Theorem, the cumulative sum of the real components, and the cumulative sum of the imaginary components, can be modelled using the Normal distribution. This approximation improves as the number of multipath components *L* increases.

$$r(t) = \sum_{n=1}^{L} A_n \cos\left(\phi_n - 2\pi f_{\max} t \cos\alpha_n\right) + j \sum_{n=1}^{L} A_n \sin\left(\phi_n - 2\pi f_{\max} t \cos\alpha_n\right)$$

$$r(t) = T_c(t) + jT_s(t)$$

By application of the central limit theorem, $T_c(t)$ and $T_s(t)$ converge to independent Normal random processes. Let us denote T_c and T_s as random variables corresponding to the processes $T_c(t)$ and $T_s(t)$ at a given time *t*. Given the physical summations in the above equation, it is clear that both random processes should be identical in terms of their variance, and both should be zero mean (since $E\{\sum\cos\theta\} = E\{\sum\sin\theta\} = 0$ for uniformly distributed θ in the range 0-360 degrees). A Rayleigh sample can thus be generated as

[2] The first version of this theorem was postulated by the French-born English mathematician Abraham de Moivre, who, in a remarkable article published in 1733, used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin. This finding was far ahead of its time, and was nearly forgotten until the famous French mathematician Pierre-Simon Laplace rescued it in his work *Théorie Analytique des Probabilités*, which was published in 1812. Laplace's finding received little attention in his own time and it was not until 1901, when Russian mathematician Aleksandr Lyapunov defined it in general terms, that the full significance of this theorem was understood.

$$r = T_c + jT_s = N\left(0, \sigma^2\right) + jN(0, \sigma^2)$$

where $N(\text{mean}, \text{variance})$ represents the Normal process. The variance of the above Rayleigh process is $2\sigma^2$, however it should be noted that σ^2 represents the quadrature Normal (or Gaussian) variance. It is important when looking at other papers and books to confirm if the variable σ^2 is being used to represent the quadrature Normal variance, or the Rayleigh variance.

It is common to wish to produce a number of power *normalised* Rayleigh samples, where in this case the Rayleigh variance (or power) is equal to unity. This can be achieved as follows

$$r = \frac{1}{\sqrt{2}}\left(N(0,1) + jN(0,1)\right) = \frac{1}{\sqrt{2}}\left(\text{randn} + j\text{randn}\right)$$

where randn represents the Matlab command for a Normal random process (with zero mean and unity variance).

The Normal or Gaussian samples have a probability density as shown below

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \qquad (3)$$

where σ^2 represents the variance of the quadrature Gaussian processes and x represents the output random variable. The received signal envelope can now be evaluated (statistically) as below:

$$r = \sqrt{T_c^2 + T_s^2} \qquad (4)$$

It can be shown (via some significant mathematics!) that by substituting equation 3 into 4, the probability density function (pdf) for r can be expressed as:

$$P(r) = \left(\frac{r}{\sigma^2}\right) \exp\left(\frac{-r^2}{2\sigma^2}\right) = \left(\frac{2r}{\Omega}\right) \exp\left(\frac{-r^2}{\Omega}\right)$$

where r is greater than or equal to zero and $\Omega = E\{r^2\} = 2\sigma^2$. The above equation is often referred to as the Rayleigh pdf and it is shown together with the Gaussian pdf in figure 23.

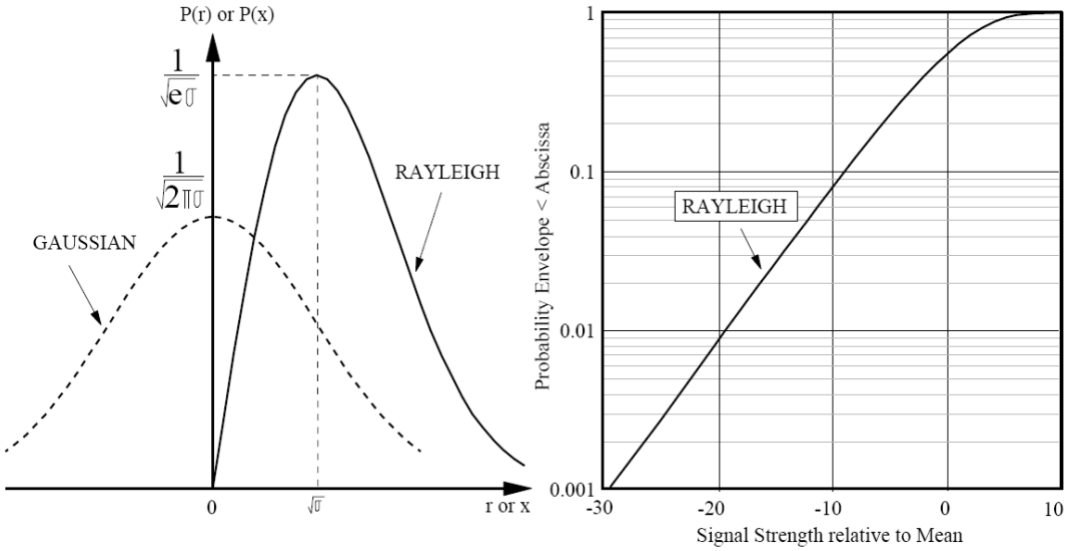


Figure 23: Gaussian and Rayleigh pdf (left), Rayleigh cdf (right)

Fading statistics are often presented in terms of the *cumulative probability density function*, i.e. the probability that the signal envelope does not exceed a specified level. The cumulative distribution for the Rayleigh process is shown below and is also plotted on the previous diagram.

$$p(r < R_s) = \int_{-\infty}^{R_s} p(r)dr = 1 - \exp\left(\frac{-R_s^2}{2\sigma^2}\right)$$

From the cdf graph it can be seen that for 99.9% of the time the signal envelope fluctuates with a dynamic range of approximately 40 dB.

Question: What is the probability of a Rayleigh envelope fading (relative to the local mean) by more than *i*) 10 dB and *ii*) 20 dB?

Answer: *i*) approximately 0.08 (or 8% of the time), *ii*) 0.009 (or just less than 1% of the time).

Rician Statistical Distribution

The determination of the statistics for the Rician fast fading distribution follows a similar process to that previously described for Rayleigh, although in this case a deterministic path is now included.

$$r(t) = \left(A_c \cos\alpha_d + \sum_{n=1}^{L} A_n \cos(\phi_n - 2\pi f_{\max} t \cos\alpha_n) \right) + j\left(A_c \sin\alpha_d + \sum_{n=1}^{L} A_n \sin(\phi_n - 2\pi f_{\max} t \cos\alpha_n) \right)$$

This equation is similar to the previous one although a known ray has now been added with a fixed magnitude A_c and at a known azimuth arrival angle α_d. Converting to polar co-ordinates and evaluating the appropriate integral (which is not trivial), the following Rician pdf can be obtained

$$p(r) = \left(\frac{r}{\sigma^2} \right) \exp\left(\frac{A_c^2 - r^2}{2\sigma^2} \right) I_o\left(\frac{rA_c}{2} \right)$$

where $I_o(z)$ represents the modified Bessel function of order zero. From this equation it can be seen that as the magnitude of the known component is reduced to zero the equation simplifies to the Rayleigh distribution quoted previously. In contrast, if the power of the known component dominates over the random power then the function peaks sharply about the amplitude of the known component and drops away rapidly from this point. The Rician pdf is plotted in figure 24 together with the Rayleigh function for various values of K-factor.

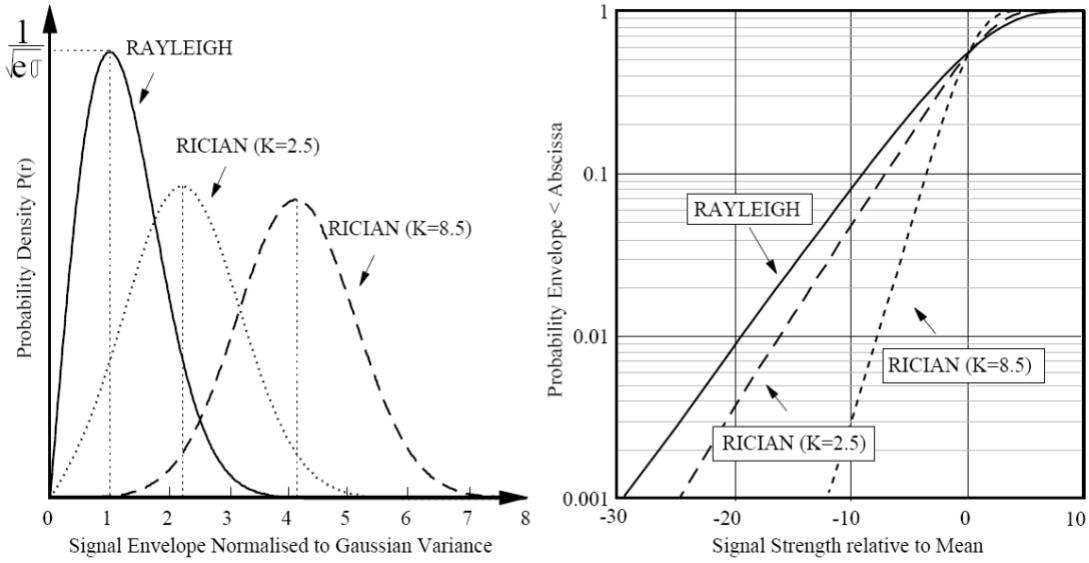


Figure 24: Rayleigh and Rician pdfs (left), cdfs (right)

The cumulative distribution clearly shows the more deterministic nature of the Rician channel. For the K=8.5 dB case, the signal fluctuates for 99.9% of the time within a dynamic range of just 24 dB (as opposed to 40 dB for the Rayleigh case).

Channel Variations and their effect on Data Transmission

As was seen in the previous sections, the received mobile signal suffers from rapid and severe envelope fades and phase shifts/flips that vary rapidly with time. Figure 27 shows the probability of error for coherent BPSK plotted against the energy per bit to noise power density (Eb/No). For BPSK, Eb/No is interchangeable with SNR. This graph assumes a White Gaussian Noise (AWGN) channel.

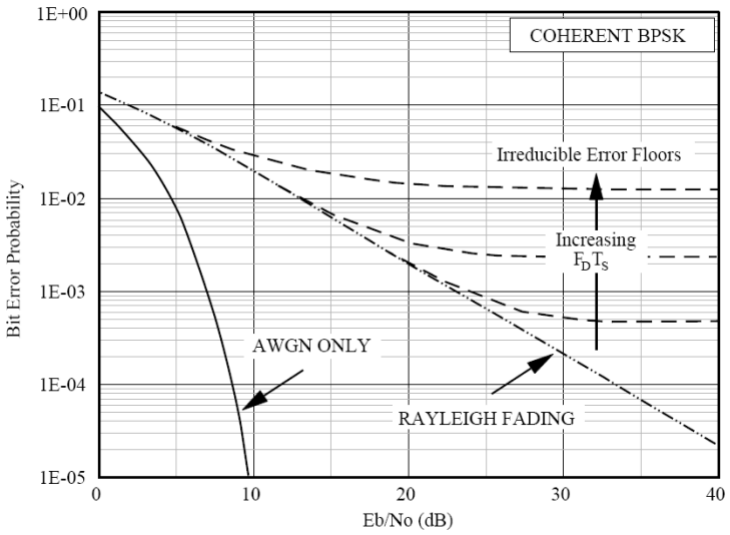


Figure 27: Bit Error Rate: AWGN vs Rayleigh Channels

For a fast fading channel, it is necessary to compute the expected (or average) probability of bit error. This must incorporate the fast fading statistics (more specifically the fading pdf relative to the local mean). For example, when the mean Eb/No is 20 dB, in a non-fading channel there will be no errors. However, for a fast fade of 15dB relative to the local mean, i.e. x=-15 dB, the instantaneous Eb/No will drop to 5dB, and errors will occur. This explains why the BER graph for a Rayleigh fading channel requires much higher average Eb/No values to reach a near error-free region. Since fades as deep at 30-40 dB are possible in a Rayleigh channel, even when the average Eb/No is 40 dB, there is a small possibility that instantaneous fades will drop the signal deep into the noise floor.

$$BER_{ff}(\bar{\gamma}) = \int_{-\infty}^{\infty} BER_{AWGN}(\bar{\gamma} + x)\, p(x)\, dx$$

The previous equation can be used to compute the BER in a fast fading channel (BER_{ff}) as a function of the average Eb/No (represented by $\bar{\gamma}$), the fast fading pdf (represented by $p(x)$), and the non-fading AWGN BER graph (represented by $BER_{AWGN}(.)$). This equation cannot be used to predict irreducible errors.

Using the previous diagram, we see that to maintain an error probability of 1 in 10,000 an Eb/No greater than 8.5 dB is required for an AWGN channel. For a Rayleigh fast fading channel the value of Eb/No must be increased to approximately 34 dB (hence the need for a fast fade margin in the earlier link budgets). For comparison, the error probability for a Rayleigh fast fading channel is also shown on the diagram.

In a Rayleigh channel, error floors are sometimes seen in the BER graph. These are known as irreducible error floors, since the error rate does not reduce as the Eb/No (or SNR) increases. These errors are caused by intersymbol interference.

In the previous diagram low value of $f_D T_S$ represents a slow moving mobile. Larger value of $f_D T_S$, can result in an irreducible error rate.

The above errors are described as irreducible since, unlike those generated by noise, these cannot be lowered by simply increasing the output power of the transmitter. This type of error is obviously very restrictive since, for a particular application, only so much error can be tolerated. This then results in a maximum velocity being imposed on the mobile users. For example, GSM at 900 MHz is designed for mobile velocities up to 300 km/h.

Question: Using the following graph, calculate the maximum vehicle speed assuming a 1 in a 1000 error threshold and the following system parameters: *i*) 150 MHz and 9.6kb/s, *ii*) 900 MHz and 72 kb/s, *iii*) 5.1 GHz and 72 kb/s, and *iv*) 5.1 GHz and 10 Mb/s.

Answer: We assume that the received Eb/No is sufficient to prevent errors as a result of thermal noise. Hence, from the right-hand graph in Figure 29 we find for an error threshold of 1 in 1000, the maximum normalised Doppler shift ($f_D T_S$) is 0.002. Assuming BPSK, the symbol period is given by the reciprocal of the data rate.

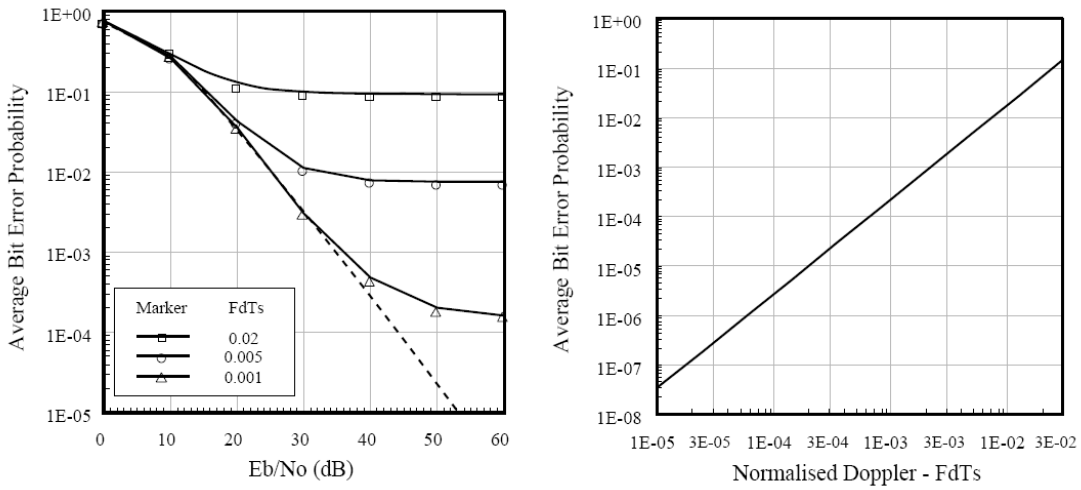


Figure 29: Modem Evaluation Graphs (see above question)

i) $\frac{v_{\max} f}{c} T_S = 0.002 \Rightarrow v_{\max} = \frac{0.002\,Rc}{f}$ = 38.4 m/s (138 km/h)

ii) $v_{\max} =$ = 48.0 m/s (173 km/h)

iii) $v_{\max} =$ = 8.5 m/s (31 km/h)

iv) $v_{\max} =$ = 1176 m/s (4235 km/h)

Propagation Summary

There are two main radio channel mechanisms that lead to the introduction of an irreducible error. Firstly, for the narrowband scenario, the amplitude and phase variations associated with the fading channel produce an error rate that is directly proportional to the users speed and the frequency of transmission. As the data rate is increased the problems due to Doppler reduce, although these are then replaced by the difficulties associated with delay spread. The data rate at which the channel can be considered wideband varies depending on the particular environment being considered. Generally, for a normalised delay spread less than 0.1 a system can be considered narrowband and the majority of the irreducible errors will arise as a result of the time variations in the channel. Once this normalised value exceeds 0.1 then the channel will start to exhibit wideband characteristics and some form of equalisation may be required to maintain a low overall error rate. This is summarised graphically in figure 30.

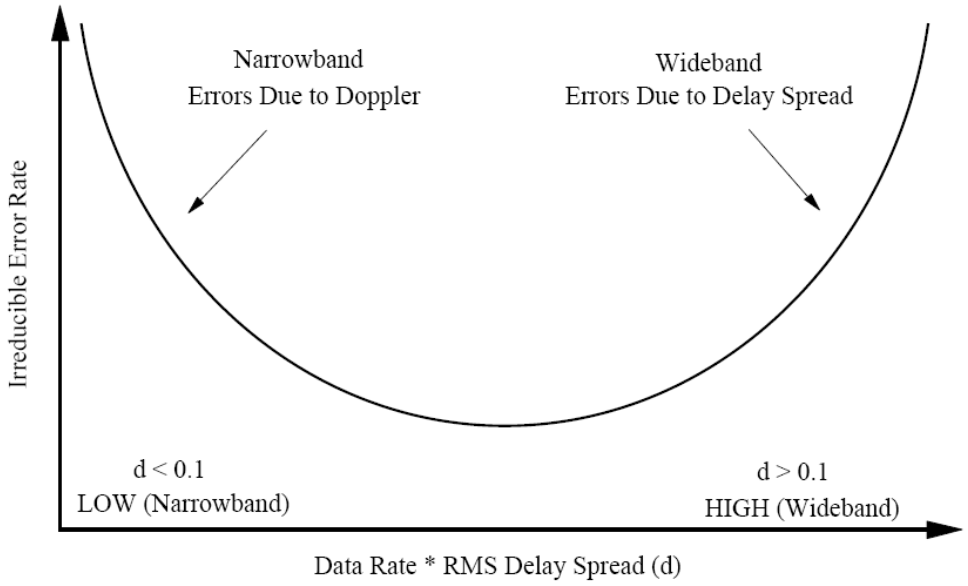
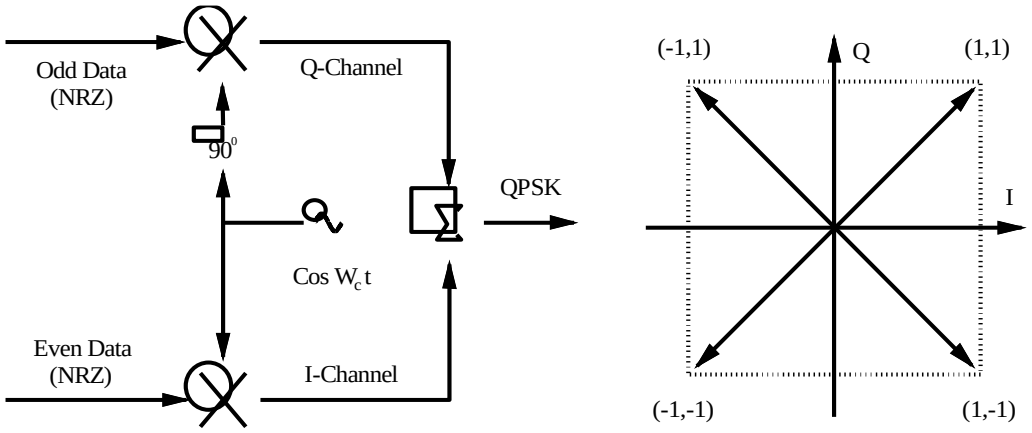


Figure 30: Summary of Irreducible Errors (Narrowband and Wideband)

Digital Modulation and Diversity

Quadrature Phase Shift Keying (QPSK)

The Figure below shows a simple block diagram for a *QPSK* modulator. The input binary data is first divided into odd and even data channels with each channel operating at a rate corresponding to half the original bit rate. These two data streams are then modulated onto the *in-phase* (I) and *quadrature* (Q) components of the carrier. The two quadrature transmissions can be separated via coherent reception since, although they occupy the same spectral range, they are modulated onto the in-phase and quadrature components of the carrier.



The Principles of Diversity

We have already seen that extreme and rapid signal variations are associated with typical mobile radio reception. A powerful method for reducing these fluctuations to a more acceptable range is now described. The technique is known as *diversity*, and it relies on the provision of two or more transmission paths, each of which carry the *same* message but suffer *different* fading statistics. Mathematically, different in this context means *independent* or *uncorrelated*.

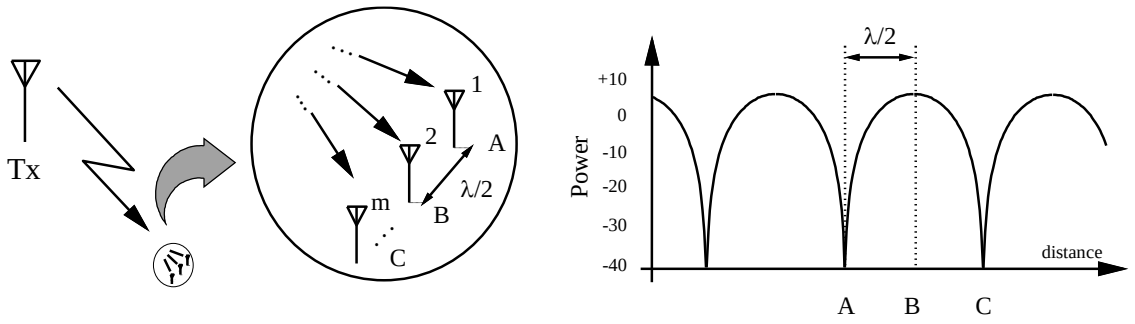
Generally, if two or more *uncorrelated* fading paths exist between the transmitter and the receiver then, by selecting or combining these signals it becomes possible to significantly reduce the impact of fast fading. This improvement arises since the probability of *both* the uncorrelated signals entering a fade at the same time is very low. Providing the branches are highly uncorrelated, when one encounters a signal fade the other will have a high probability of receiving a strong signal. If the signal strength is monitored at each of the branches then the resulting distortion can be minimised through suitable signal processing. The use of diversity reduces the likelihood of encountering deep signal fades and thus removes many of the *clicks* and signal outages traditionally encountered at a cell's boundary.

Generating Independent or Uncorrelated Paths

There are a number of methods for generating the uncorrelated paths required for diversity combining. Techniques such as *time, frequency, angle, and space diversity* have been studied over the years.

Space Diversity

The earliest forms of diversity required reception via a number of different antennas, each *spaced sufficiently far apart to ensure independent and uncorrelated fading*. This arrangement, known as *space diversity*, has found many applications due to its relative simplicity and the fact that no additional frequency spectrum is required. The basic requirement is for the spacing of the receiving antennas to be set such that the signals on adjacent branches appear uncorrelated (i.e. that the antennas are spaced greater than the *coherence distance*). This principle is shown diagrammatically in figure 13. The antenna spacings are normally a fraction of the carrier wavelength (commonly one quarter or one half of a wavelength) and this configuration is referred to as *microscopic diversity*.



(i) Space Diversity (ii) Power Variation with Distance

Figure 13: Space Diversity

Various antenna spacing have been tried over the years and it is generally accepted that for a handset, a value of around $\lambda/2$ offers the best compromise between the size of the antenna array and the system performance. However, it should be noted that the optimum spacing depends on the azimuth *angle spread* of the multipaths (which in turns influences the *Doppler spread*). The angle spread depends on the particular operating environment, for example for antennas mounted at roof height the majority of the energy arriving from a specific direction (i.e. the location of the terminal), and hence the angle and Doppler spread is low, and the required antenna spacing needs to larger (typically 2-3 wavelengths).

Figure 14: Antenna Space Diversity for 802.11

Figure 14 shows an example of spaced antenna diversity as applied to 802.11 (wireless LAN equipment).

Angle Diversity

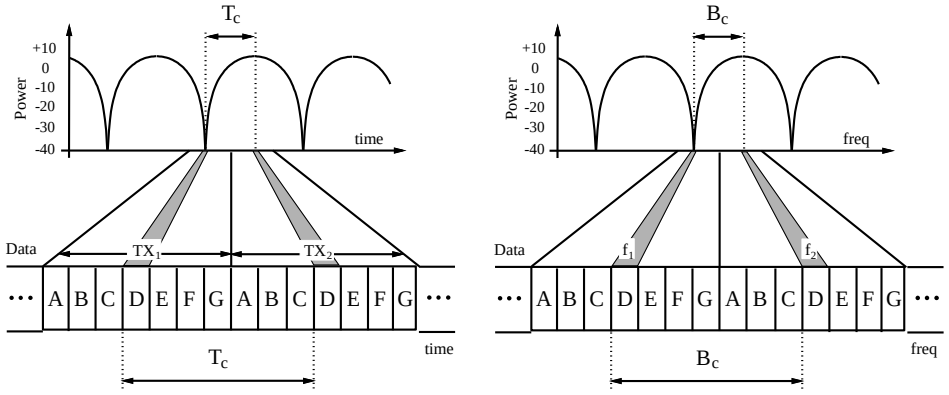
Following the introduction of space diversity, it was quickly realised that signals suitable for diversity combining could be obtained by methods other than spaced antennas. These alternative methods can be grouped into the following categories - *angle diversity, time diversity and frequency diversity*.

It has been observed that signals arriving from different directions are also uncorrelated and hence, directive antennas may be used to produce the required independently faded diversity signals. This technique, also benefits from the fact that directive antennas tend to restrict the angles over which rays may be received, therefore reducing the effects of *Doppler spread* and *time delay spread*. The use of sectorised antennas can also be used to improve the performance of wideband

systems, where the reduction in rms delay spread enables an increase in the unequalised data rate.

Time and Frequency Diversity

The remaining techniques, time and frequency diversity, are less desirable for mobile radio applications because of their requirement for extra signal bandwidth. Figure 16(i) shows the situation for time diversity, where the system operates by retransmitting the message signal after a suitable time delay. The required delay must be greater than the channel's *coherence time*, (which can be approximated from the reciprocal of the Doppler spread). Typically, for a coherence time in the order of milliseconds, this technique would require a reasonable degree of information storage at both the transmitter and receiver. The storage also introduces a delay into the received signal which may be unacceptable for certain type of applications (such as conversational audio). Finally, for time diversity there is no theoretical improvement if the user remains *stationary*. In practice some gain can still be achieved due to the time-variability associated with other motion in the channel.



(i) Time Diversity (ii) Frequency Diversity

Figure 16: Time and Frequency Diversity

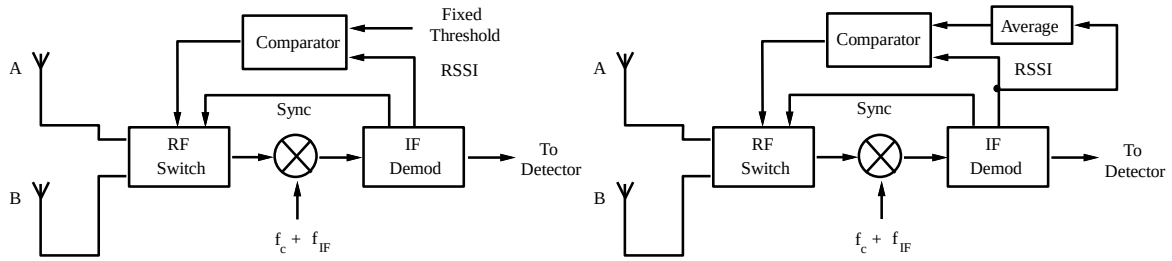
As can be seen from figure 16(ii), frequency diversity can be thought of as a dual to time diversity. Its main advantage is that there is no delay associated with the process and it can therefore be applied to all types of application. Frequency diversity operates by transmitting the required signal on more than one frequency. The separation of these frequencies must be enough to ensure uncorrelated fading, i.e. their separation must be greater than the *coherence bandwidth* (which can be approximated from the reciprocal of the rms Delay spread).

Diversity Combining Strategies

Once a number of uncorrelated paths have been obtained at the receiver they must be combined in order to improve the resulting signal to noise statistics. A number of competing strategies are now discussed in ascending order of complexity, cost and performance.

Switched (Scanning) Diversity

The simplest form of diversity combining is known as *switched* or *scanning* diversity. Compared with other forms of diversity it is inherently cheap to build since, irrespective of the number of branches, it requires just one circuit to measure the short term average signal power (i.e. to perform a *Received Signal Strength Indication or RSSI*). For scanning diversity the current branch remains selected until it *fails* a given metric, at this point the next branch in turn is *blindly* selected. If this new branch satisfies the required metric then it remains selected, otherwise the system moves on to the next branch (or back to the original for two branch systems). The most common configuration makes use of two branch space diversity combined with a simple metric based on signal strength. In this configuration the system simply monitors the signal strength on the current branch and then compares this with a fixed or pre-set threshold as shown in figure 17(i).



(i) Switch Diversity with Fixed Threshold (ii) Switch Diversity with Adaptive Threshold

Figure 17: Switched Diversity

Assuming a system based on two antennas, when the signal strength falls below the threshold the system simply switches to the second antenna (see figure 18). If the signal strength on this branch is *greater* than the threshold then the antenna remains selected until such times as the signal strength falls back below the threshold. If the signal strength on *both* antennas is below the threshold then the receiver continually switches between the two antennas until the signal on one of the branches rises above the threshold. Unfortunately, when applied to analogue modulation schemes, it is impossible to implement this switching in a manner transparent to the user. Each switch results in a brief *burst* of distortion and, as a result, an alternative strategy known as *switch and stay* may be more suitable. In this approach the switching still

occurs when the signal level falls below the threshold, however, unlike the previous approach the next branch remains selected irrespective of whether its signal strength proves acceptable or not. Further switching will only occur when the signal strength rises above, and then falls back below, the given threshold. Generally, for the first approach, which uses rapid antenna switching, the system results in a quicker return to an acceptable signal level. However, this is achieved at the expensive of rapid switching noise and, for this reason, for analogue systems the switch and stay approach is normally preferred. For digital systems, switching is forced to occur in the dead-time between data bursts.

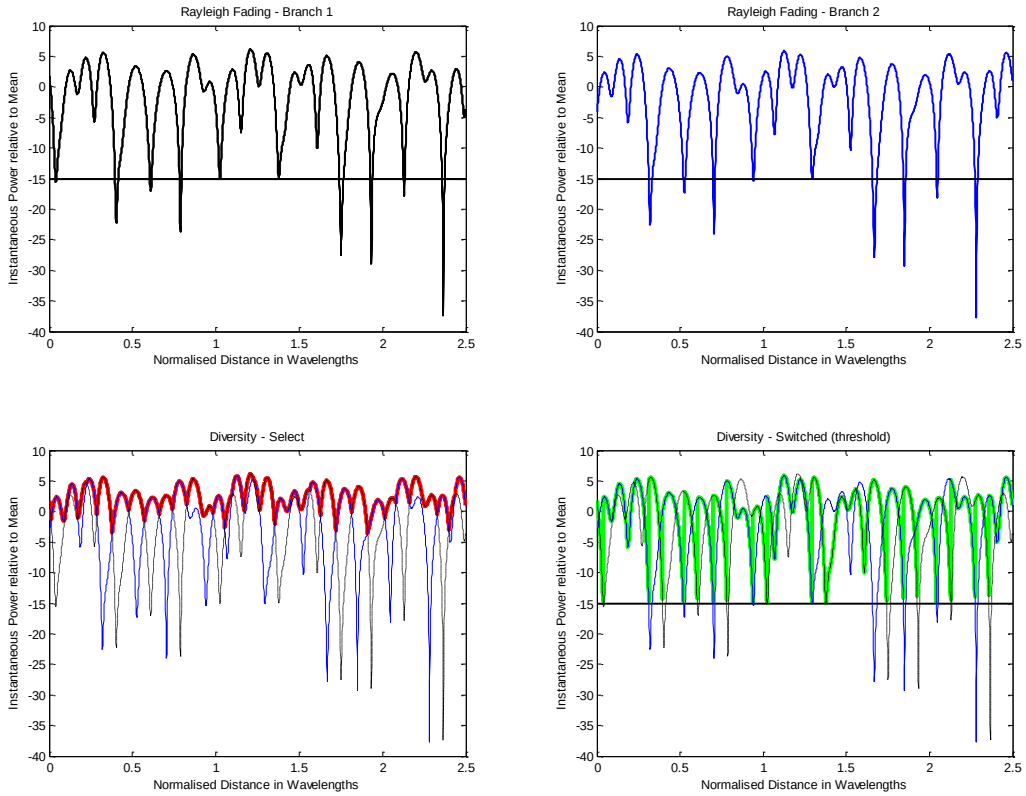


Figure 18: Example of Switched and Selection Diversity

In practice there is some advantage to be gained from the use of a *variable* threshold, since a setting that is satisfactory for one area may result in unnecessary switching when the user moves to another area where the mean signal strength is different. Figure 17(ii) shows a modified system where the threshold is obtained from a time average of the previous RSSI output. This approach allows fades relative to the *local mean* to be removed (as shown in figure 18, where the switched threshold is 15 dB below the estimated local mean) and provides superior performance since, in addition to errors due to AWGN, those caused by random FM and rms delay spread are also reduced. This occurs since these irreducible errors tend to occur at the bottom of deep fades, and diversity avoids detection at these points.

Selection Diversity

Selection diversity represents a slightly more complex combining strategy where the most appropriate branch or antenna is always chosen from those available. Unfortunately the system is more expensive to implement since the signal strength on each branch must now be estimated in parallel (figure 19).

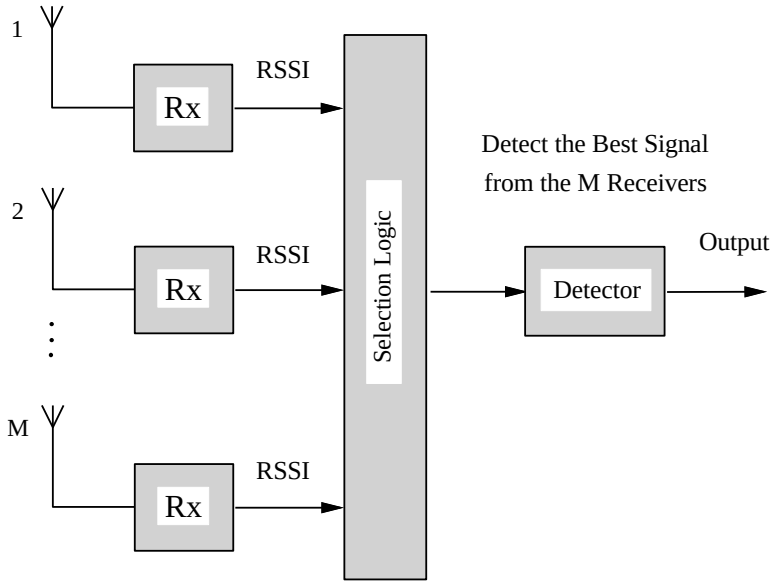


Figure 19: Selection Diversity

There are several techniques for selecting the most appropriate branch, the most straightforward simply chooses the branch with the largest RSSI. This method is used in the Matlab example demonstrated in figure 18. The selection combining approach ensures the stronger of the available fading envelopes is always chosen, and compared to switched diversity, the fading statistics are clearly superior (note the red envelope for selection diversity is clearly better than the green envelope for switched diversity). The Matlab source code allows the antenna spacing to be varied, thus altering the correlation of the fading envelope between the two branches. The envelope correlation is also computed by the Matlab software.

For systems suffering from *co-channel* or *adjacent-channel* interference, a direct measure of the signal strength may no longer be appropriate. A more reliable approach in such channels would be to evaluate the strength of the wanted signal on each branch through the use of a unique training or synchronisation sequence.

Although selection diversity is superior to scanning diversity, the difference is quite small and its greater complexity and cost makes it unattractive.

Maximal Ratio and Equal Gain Combining

Both techniques can be broadly classified as linear combiners, since the various signal inputs are now individually weighted and added. If addition takes place after detection then the system is known as *post-detection*, alternatively, if combining is implemented before detection, the process is referred to as *pre-detection* combining. The latter approach requires a process known as *co-phasing* to be applied before the signals are combined to prevent the possibility of destructive summation. This process is required to align the signal phasors and thus ensure constructive addition. Assuming pre-detection combining, the process is shown graphically in figure 20.

For maximal ratio combining each branch is weighted before summation in proportion to its own *signal to noise ratio*. The required cophasing process can represent a problem in certain systems. The resulting output from the combiner is therefore a weighted, cophased summation of all the diversity branches available at the receiver. Maximal ratio combining optimises the received signal to noise power ratio at the output from the combiner and therefore represents the optimal combining strategy (for a narrowband channel).

Rather than using maximal ratio combining, a slightly simpler approach is to implement *equal gain combining*. For this technique the gain of each branch is set to one, and the signals are simply cophased and summed.

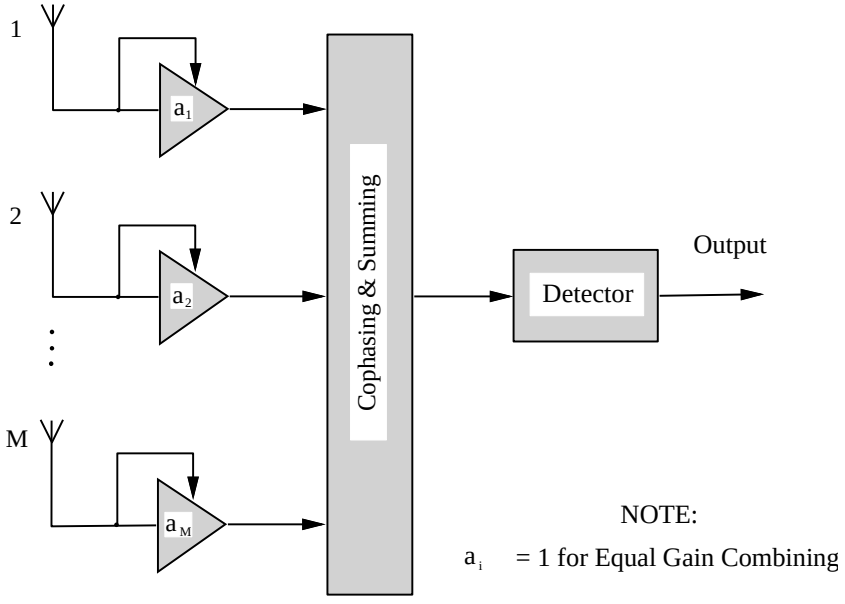


Figure 20: Maximal Ratio / Equal Gain Combining

Although EGC clearly represents a sub-optimal strategy, the avoidance of any signal to noise power estimation is *highly desirable*. The performance of equal gain combining is, in general, only marginally below that for maximal ratio combining

and therefore represents a useful compromise between cost, complexity and performance.

Diversity Improvement (Fading Channels)

The BER in a Rayleigh fading channel can be significantly *reduced* with the use of diversity. A common way to demonstrate the diversity gain is to analyse the cumulative distribution function of the fading envelope. Figure 21 shows an example of the output generated from the Matlab example ray_cdf.m. This code produces a Rayleigh fading envelope and then graphically demonstrates the computation of the envelop cdf. A threshold is generated from 10 dB above the mean to -30 dB below the mean. All the samples *below* the threshold are highlighted with a red circle. The software estimates the probability of the fading envelope being below the current threshold by computing the ratio of the number of circled points to the total number of points.

Figure 22 shows the fading cdf statistics with and without the use of selection diversity (diversity switching based on RSSI). From this graph, the probability of fading lower than -15 dB on the mean is 3% for a single antenna (the red circle), but just 0.06% with two branch selection diversity (the green circle). The technique can lower the irreducible error floors by around one to two orders of magnitude. Generally, at a BER threshold of 1 in 100, diversity can offer an 8-12 dB gain in a Rayleigh fading channel. Diversity combining can also increase the maximum possible unequalised bit rate in a time dispersive channel by a factor of two.

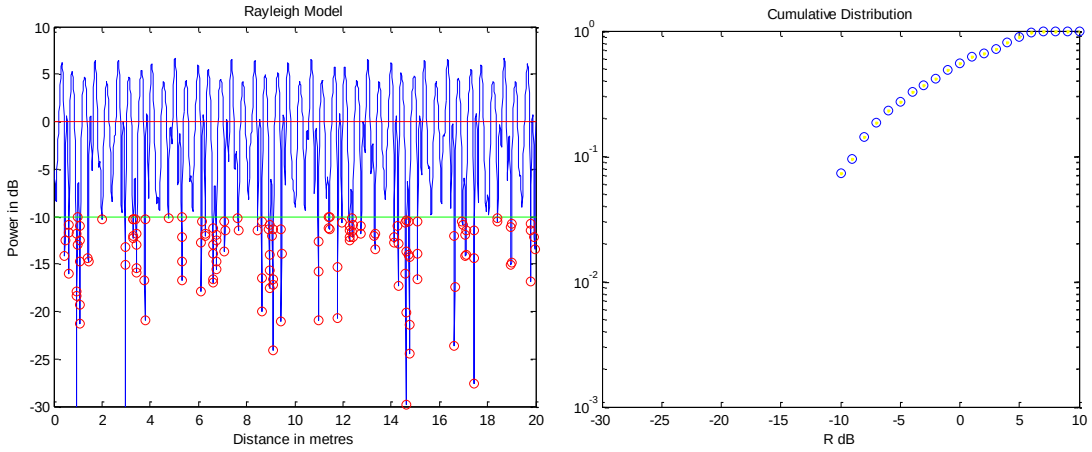


Figure 21: Graphical Demonstration of CDF

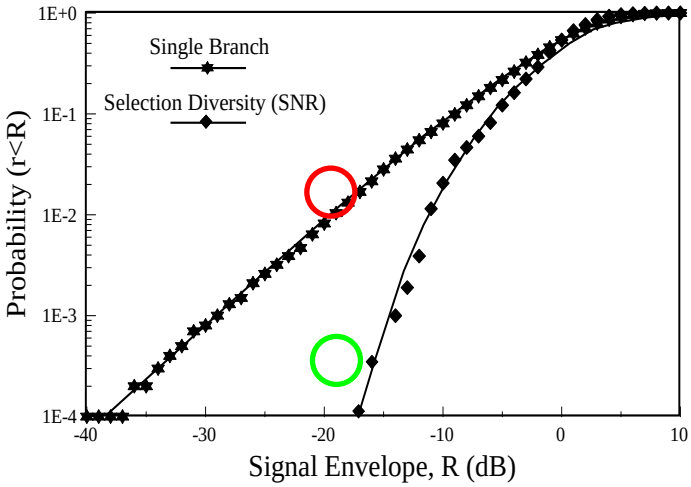
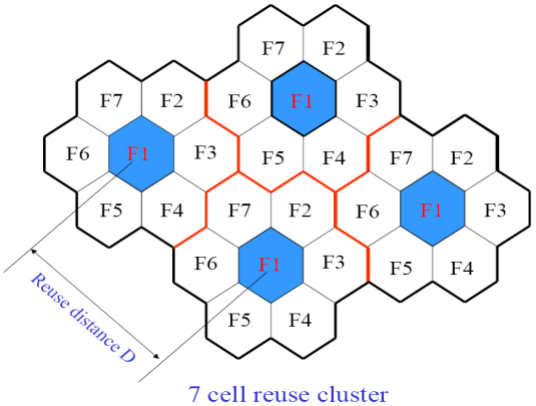


Figure 22: Diversity Improvement (Narrowband)

Cellular Principals and Design



7 cell reuse cluster

Traditional Cellular Design

Historically, cellular networks are built using a large number of basestations to achieve continuous street-level coverage. The early analogue systems placed basestations on tall buildings and hill tops to achieve a coverage radius of up to 20 km. Each basestation is capable of supporting a large number of simultaneous users, the exact number being a function of cost, complexity, bandwidth and modulation technique.

In the mid-1980's, when the first analogue systems were introduced into the UK, the expected number of subscribers was very low (hundreds of thousands) and almost entirely based on vehicular users. Large cells allowed rapid coverage to be easily achieved. For fast moving vehicles the large cells also reduced the problems of *handover* (the term used to describe transferring a call from one basestation to another). Figure 1 shows a rural basestation and the hexagonal planning used to achieve continuous coverage. For convenience, each cell is assumed to cover an ideal hexagon. Hexagons are used since they tessalate (i.e. cover an area without overlap) and also approximate cicular coverage better than a square or triangle.

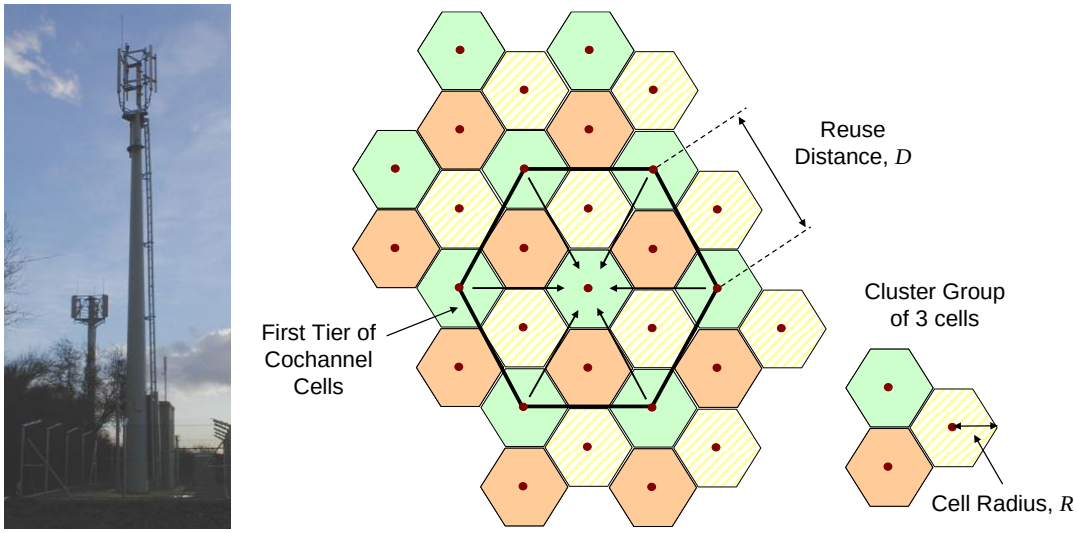


Figure 1. Rural base site (left), Traditional Cellular Deployment (right)

In a GSM network, adjacent cells cannot use the *same* set of frequencies. If this were allowed it would be possible for two users to interfere with one another (known as cochannel interference). It is worth noting that 3G networks use spread spectrum technology that allows all frequencies to be used in all cells. However, this approach is beyond the scope of this unit. In figure 1, the colour and shading patterns are used to represent unique frequency groups. We observe a central green cell surround by

an initial set of six cochannel cells (the centre of these cells forms a larger hexagon). These six cells form the first-tier of cochannel interferers. The pattern shown in Figure 1 is based on a *cluster group* of three cells. The cluster group (or *cluster size*) is normally denoted by the letter *N*. The available radio resources are shared equally across a group of *N* cells (with *N*=3 in this example). The resources are reused geography (denoted by cells of the same colour and shading). The shape of the three cell cluster group (also known as the *reuse pattern*) is shown in the far right-hand side of the figure. Later in the handout it will be shown that the distance to the first-tier of cochannel interferers is dependent on the size of the cluster group and the cell radius.

In any cellular design, the cluster group is simply stamped out again and again to achieve continuous spatial coverage. The size of the cluster group is calculated to ensure the interference generated between cochannel sites is acceptable. This will be analysed mathematically in more detail later in the handout. The distance between cochannel cells is referred to as the *cochanel reuse distance,* and this is normally denoted by the letter *D* (see figure 1 for an example). The cell radius is denoted by the letter *R*. The total available bandwidth allocation is split into *N* equal sections, and hence only one *N*-th of the total bandwidth is available to any one hexagonal cell.

The first analogue systems operated with a cluster group of twelve or more, however with the introduction of digital communications, the cluster size was reduced to four in most cases. This was made possible since the digital modulation format (GSM, LTE) is able to tolerate a higher level of cochannel interference. Obviously, the lower the cluster size the higher the systems' overall capacity (since more bandwidth is available in each cell). As mentioned above, the cluster size is governed by the radio system's sensitivity to cochannel interference. The higher the system's interference tolerance, the closer the same set of frequencies can be spatially reused.

From inspection of the reuse cells in figure 1 it can be seen that a first-tier of interfering basestations exist, and that this always consists of six cochannel cells. This is more clearly shown in figure 2, which also shows the second-tier of cochannel interferers (which comprises 12 cells). It is common to assume that the bulk of the of cochannel interference comes from the first-tier of cochannel cells (since these are much closer and we know the path loss increases significantly with separation distance). Later we will assume that the cochannel interference can be computed from just the first-tier of cochannel cells.

As more and more users began to use portable phones (not vehicular based), a problem with capacity began to occur. In general, the capacity of a network can be defined as the number of voice channels per MHz per kilometre squared. The term

Erlang is used to represent the concept of a single continuous voice call. Hence, capacity is often quoted as Erlangs per MHz per kilometre squared. A single user is often assumed to represent approximately 0.02 Erlangs (i.e. they use the phone for 2% of the time, or approximately 30 minutes per day). These ideas can be used to estimate the number of mobile subscribers that can be supported in a given area. For example, with mobile penetration now running at levels well beyond 50% of the population, very large capacities are required in city centres (for example, in Bristol there could easily be as many as 10,000 subscribers in the city centre during peak periods).

Cochannel Interference and Signal to Noise Tolerance

In a cochannel interference limited cellular system, each cell is surrounded by a number of reuse cells. As shown in figure 2, there are six reuse cells in the first-tier and twelve reuse cells in the second-tier, regardless of the cluster size *N*.

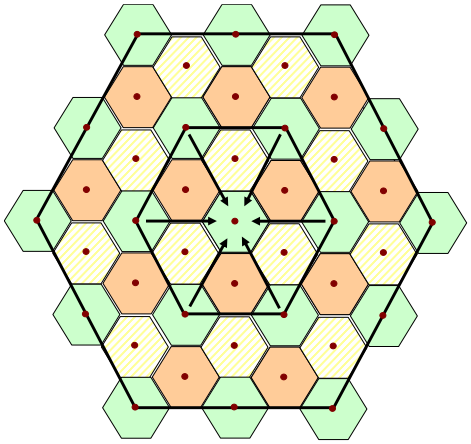


Figure 2: Cochannel Interference (6 in first tier, 12 in second Tier)

In practice, all the cochannel cells in a given tier are assumed to have identical statistical interference characteristics (i.e a common path loss model versus seperation distance). A mobile system is said to fail (or be in *outage*) when the threshold BER is exceeded for a certain period of time. The reuse distance *D* can be obtained by considering the statistics of the power versus distance curve. From the re-use distance and cell size, the system capacity can be estimated in terms of channels/MHz/km^2.

The left-hand side of Figure 3 shows a typical cellular arrangement (*N*=4), while the right-hand side of Figure 3 shows a user moving from point A to point B across three cells. Initially the user is served by BS1 on frequency 1, however as the mobile moves along the route from A to B there comes a point where the average received power from BS1 is less than the average received power from BS2 (of frequency 2). In this simple model, the signal levels from BS1 and BS2 are equal at the cell radius.

As the user moves beyond the cell radius of BS1, the power from BS2 becomes stronger than that from BS1. At this point the user undergoes a *handover* from BS1 (frequency 1) to BS2 (frequency 2). In GSM, the call is broken from BS1 and then re-established via BS2 (this is known as a *break-before-make* handover). As the user continues, a further handover occurs between BS2 and BS3. Depending on the reuse distance (*D*), BS3 may be able to spatially reuse the first frequency (as shown in the diagram). Alternatively, hand over to another frequency can occur if the interference from the first basestation is still too high. At the handover points in figure 3 the ratio between the wanted and unwanted powers is marked as the C/I *protection ratio*. This defines how much stronger the wanted signal must be compared to the unwanted interfering signals. The higher the required protection ratio the further the cochannel frequencies must be separated (i.e. the greater the value of *D* relative to the cell radius *R*).

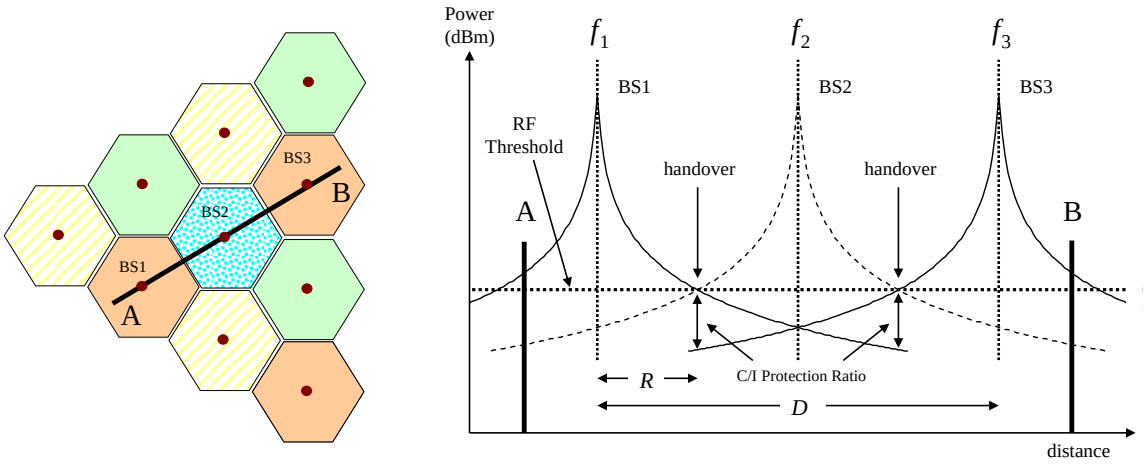


Figure 3: Frequency Reuse (N=4) and Principals of Cellular Handover

In a cochannel interference limited mobile radio system, adequate signal strength *and* signal to interference ratios (SIR) are essential for successful communications. Outage probability, defined as the probability of failing to simultaneously achieve a given signal to noise ratio (SNR) and a given signal to interference ratio, is an appropriate measure for evaluating the performance of mobile radio systems.

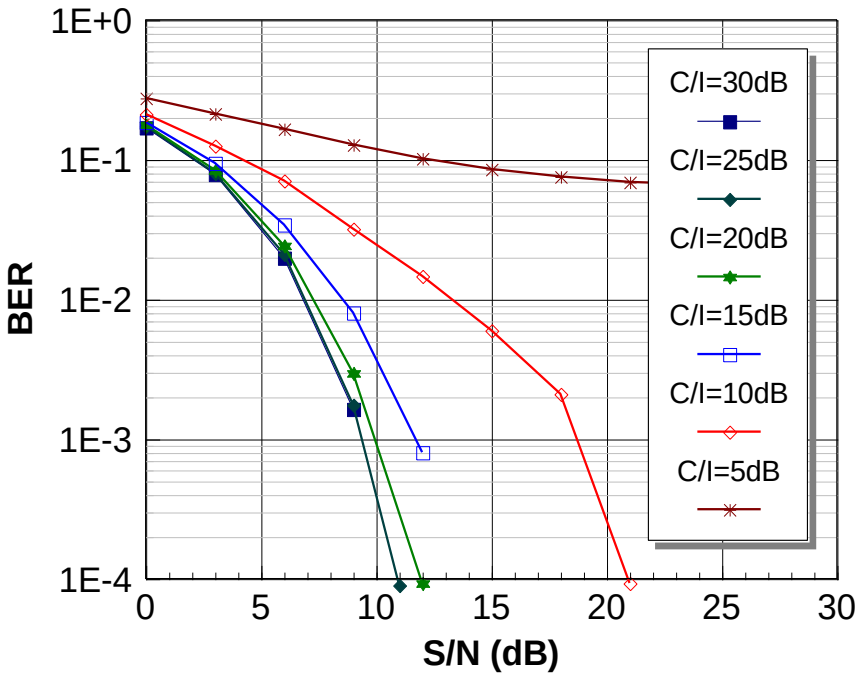
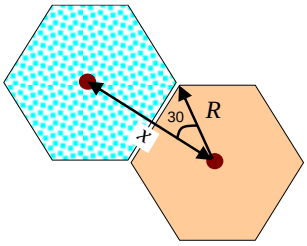


Figure 4: Typical SNR and C/I performance

Figure 4 shows the results of a simulation study for a typical digital modulation scheme. In the presence of cochannel interference and AWGN, it clearly shows that for a required BER and a required cochannel interference ratio, the desired signal to noise ratio can be determined. For example, assuming a BER of 1% and C/I of 15 dB, the required SNR would be approximately 8 dB. For higher cochannel interference ratios (say 10 dB), more signal power is required to meet the BER probability (SNR = 13 dB). It should be noted that these results apply to non-fading channels, as we have already seen, in a rapidly fading channel the average SNR required would be far higher. The above parameters are generally considered as fundamental factors in mobile radio system design and are often used in capacity evaluation. The results are interesting since they show that the optimum C/I performance is achieved by optimising the S/N and C/I ratios at the cell boundaries.

Cellular Planning and Capacity



It has already been seen that to ensure complete area coverage with no dead spots, a series of regular polygons can be used. For economic reasons, the hexagon is normally assumed (hexagons result in a more efficient design than squares or triangles). Based on simple geometry, the centre to centre distance between adjacent hexagons, x, is given by $\sqrt{3}R$, where R is the centre to vertex distance. In cellular design it is not normally possible to use the same set of frequencies in adjacent cells (there are exceptions such as direct sequence CDMA where such arrangements can be practical).

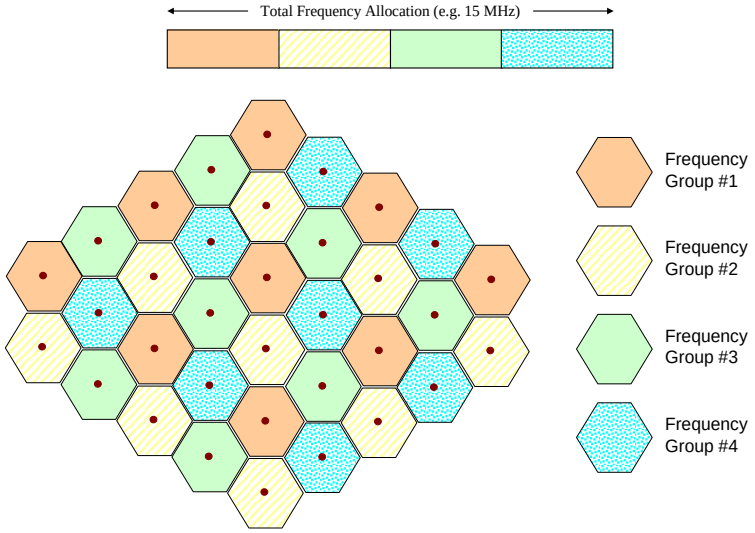


Figure 5: Frequency Planning and Re-use

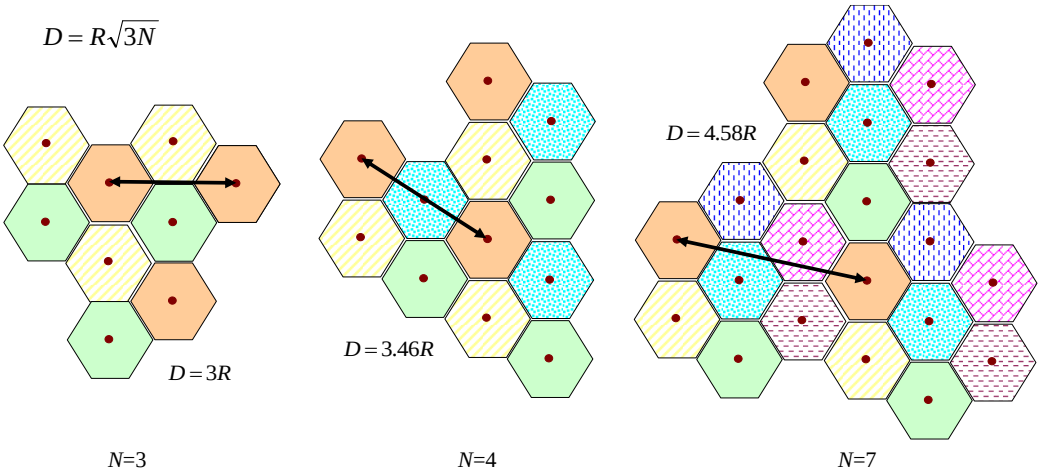


Figure 6: Frequency Reuse versus D/R

Figure 5 shows the method by which most systems are planned and deployed. An operator is allocated a certain bandwidth allocation, based on the type of radio system being used, this radio bandwidth is sub-divided into a number of frequency groups. As will be seen later, the exact number of groups is determined by the interference tolerance of the radio system and the type of basestation antenna arrangement employed. The number of frequency groups is defined by the cluster size, since cells are planned in clusters that use the entire frequency allocation. Figure 6 shows network examples based on a cluster size N of 3, 4 and 7. In figure 5 a cluster size of 4 is assumed, hence the total frequency allocation is divided into four groups (each indicated by a particular colour and shading pattern). Figure 6 shows how the cochannel reuse distance varies with cluster size. As the cluster group gets larger, so does the distance between adjacent cochannel cells. Later in this handout a mathematical relationship between these variables will be defined. It is common to normalised the reuse distance to the cell radius (known as the D/R ratio). Clearly, the more interference a system can tolerate, the lower the required D/R ratio and the smaller the resulting cluster size N.

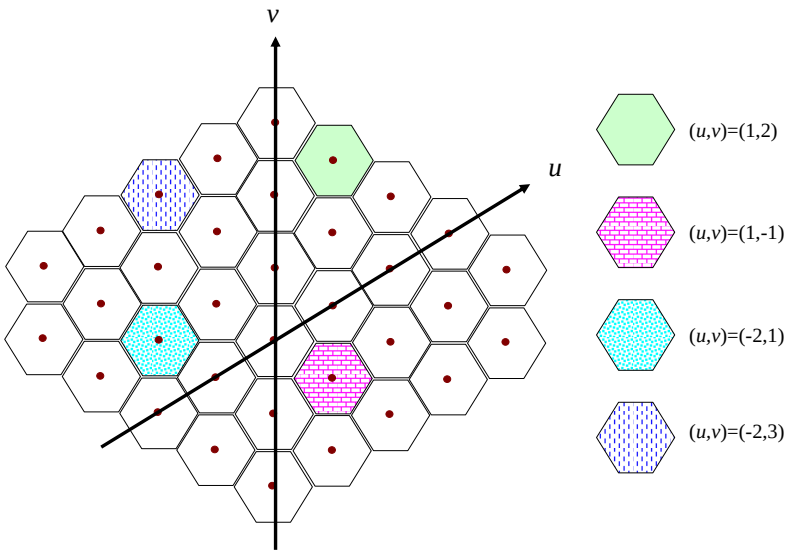


Figure 7: Hexagonal Axes and Co-ordinate System

As shown in Figure 7 above, the most convenient co-ordinate system for a hexagonal cell structure using axes inclined at 60 degrees. Assuming the origin to be at (0,0) and restricting u and v to integer values ($u=i$, $v=j$), the *normalised* distance to the cell centre is given by

$$D_{norm} = \sqrt{(i+j)^2 - ij}$$

Using the above equation, the normalised distance between any adjacent cell site is unity (i=1,j=0 or i=0 j=1). To compute the number of cells per cluster, an arrangement of the form shown in figure 8 is normally used. The cells designated by the letter A are the six nearest cochannel cells (i.e. the first-tier) to the centre cell. It can be seen that these cells are located at the vertices of the larger hexagonal cell of radius D. The radius of the larger cell D in figure 8 is given by $\sqrt{3}RD_{norm}$. We can write $D^2 = 3R^2 D_{norm}^2$, and expanding we get:

$$D^2 = 3R^2(i^2 + j^2 + ij)$$

Since the area of a hexagon is proportional to the square of its radius, the area of the larger hexagon is given by:

$$A_{large} = k(3R^2)(i^2 + j^2 + ij)$$

where k represents a fixed constant. Similarly, the area of the small hexagon is given by:

$$A_{small} = k(R^2)$$

Therefore, the ratio of the large hexagonal area to the small hexagonal area is given by:

$$A_{large} / A_{small} = 3(i^2 + j^2 + ij)$$

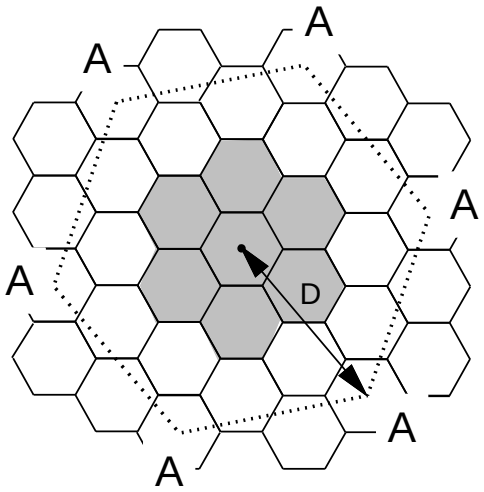


Figure 8: Cellular Geometry and Area

From the symmetry of the cellular geometry it can be seen that the larger hexagon incorporates the N shaded centre cells (N=7) and *one third* of the six surrounding N

cell clusters. Hence the total number of enclosed cells equals 3*N*. Since the enclosed area is proportional to the number of cells we can write:

$$A_{small} = k_2.1 \quad \text{hence} \quad k_2 = k(R^2)$$

$$A_{large} = k_2.(3N) = k(R^2)(3N) = k(3R^2)(i^2 + j^2 + ij)$$

$$A_{large} / A_{small} = 3(i^2 + j^2 + ij) = (k_2.3N)/(k_2.1)$$

$$\text{hence } N = (i^2 + j^2 + ij)$$

Since *i* and *j* are restricted to integer values, the cluster size *N* is restricted to a limited set of discontinuous vales. Hence it is not possible to plan a system with a cluster size of 2,5,6,8 ... etc. Combining the above equations for *D* and *N* we can show that:

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

The above equation is important since it relates the cell radius and the reuse distance to the cluster size. The term *D*/*R* is referred to as the normalised cochannel reuse ratio. For a cluster size of 7, the ratio of *D*/*R* is 4.6 (as shown in figure 6). For a cluster size of 21, the *D*/*R* ratio is 7.9.

i	*j*	*N*
1	0	1
1	1	3
2	0	4
2	1	7
3	0	9
2	2	12
3	1	13
4	0	16

Table 1: Cluster size as a function of i and j

Cluster size as a function of Cochannel Interference

The downlink of a cellular system can be modelled based on the wanted signal and the interference from the six surrounding 'first-tier' cochannel cells. For a worst case scenario, the user is assumed to lie at the cell boundary (a distance *R* from the centre basestation) and the six first-tier interfering cells are *all* assumed to lie at a distance of *D-R*. The signal power from the centre basestation is inversely proportional to the

n-th power of the distance, where *n* represents the propagation path loss index from the first handout (*n*=2 for line of sight communications). Mathematically

$$S \propto \frac{1}{R^n}$$

The interference power from each of the first-tier cells is inversely proportional to the *n*-th power of the distance (*D*-*R*). However, since *D*>>*R*, *D*-*R*$\approx$*D* and we can write

$$I_1 \propto \frac{1}{D^n} \qquad I_6 \propto \frac{6}{D^n}$$

Since there are six interfering tiers (assuming an omni-directional antenna) the actual interference is six times that of I_1. The resulting S/I ratio is then given by

$$\frac{S}{I} = \frac{(D/R)^n}{6} \qquad \text{hence} \qquad D/R = \left[6(S/I)\right]^{1/n}$$

Using the previous relationship between *D*/*R* and the cluster size *N*, the following equation can be written

$$N = \frac{1}{3}\left[6(S/I)\right]^{2/n}$$

Using the above equation, an estimate for the cluster size can be obtained from the S/I ratio required by the radio system. For networks with sectorised antennas, the number of interferers in the first-tier can be adjusted. The most common sectorisation strategy in to use the 120 degree sectors. In this case, only two interfering basesites would be visable, and the equation would be modified as below.

$$N_{3\text{sector}} = \frac{1}{3}\left[2(S/I)\right]^{2/n}$$

For systems such as GSM and DCS-1800, an S/I of 9 dB is required at the mobile. Inserting this value into the equation for an omni-directional antenna produces a normalised reuse distance of 4.38 (the path loss exponent is assumed to be 3). For a three sector antenna, the normalised resue distance drops to 2.1.

Systems such as TACS require an S/I of 17 dB, this results in a normalised reuse distance of 14.9 (omni antenna), and 7.2 (three sectors). Based on the reduced S/I alone, GSM offers just over three times the capacity of TACS. In practice, there are many other factors that effect the system capacity.

Capacity Evaluation

The capacity of a cellular system is a function of the total bandwidth, the channel bandwidth, the cluster size and the cell area. The total bandwidth for a cellular system is normally allocated by a regulatory body (such as Ofcom in the UK). Typically, this allocation is around 5-15 MHz per operator per network. The channel bandwidth is defined as the average bandwidth required to support a single communications channel. For TDMA systems this value can be obtained by dividing the carrier bandwidth by the number of user slots. As defined above, the cluster size is a function of the system's C/I tolerance and the cell sectorisation strategy. Finally, since capacity is often quoted 'per kilometre squared', the average cell radius is also used in the formulation. The general equation for capacity is therefore given by:

$$\text{Capacity} = \frac{\text{Total Number of channels}}{\text{Cluster size} * \text{Cell area} * \text{Total bandwidth}} \quad \text{Erlangs/MHz/km}^2$$

The channel bandwidth can vary significantly depending on technology. Early FM based TACS and AMPS systems used around 25-30 kHz per voice channel. This bandwidth is far in excess of the minimum required to support a voice channel and arises mainly due to the inefficiency of FM modulation. Newer systems such as GSM support 8 voice channels in a 200 kHz band, hence average channel bandwidths of around 25 kHz can be assumed. From the above capacity equation, it is clear that the total number of calls is directly related to the channel bandwidth. However, as mentioned earlier, this does not represent the full story since more bandwidth efficient schemes often require *larger* cluster sizes and reuse distances due to their greater sensitivity to cochannel interference.

Capacity is also a function of the cell area, not surprisingly higher capacity can be obtained if smaller cell sizes are used (although this requires a large number of cells, increasing the capital expenditure, or *capex*). These small cell systems, often referred to as microcells or picocells, represent a recent development to offer the high subscriber densities required in urban communication systems. Small cell systems offer higher capacity (in terms of subscribers per MHz per kilometre square), however larger numbers of basestations are required to achieve continuous large area coverage.

To determine the number of *subscribers* that can be supported, the previous capacity equation can be used in conjunction with the average traffic density per user. As mentioned previously, the average subscriber creates approximately 0.02 Erlangs of traffic. Hence, capacity can be estimated with the following expression

$$\text{Capacity} = \frac{\text{Total Number of channels}}{\text{Cluster size} * \text{Cell area} * \text{Total bandwidth} * \text{Erlangs / users}} \quad \text{subscribers/MHz/km}^2$$

The above is an estimation since it does not take into account the statistical *blocking probability* of a large number of users using their phones throughout the day.

Example 1:

An analogue cellular system has a total bandwidth of 5 MHz and operates with a cluster size of 7 and a cell radius of 12 km. The network comprises of 100 basestations and each voice channel requires 25 kHz. Assuming that each user represents a traffic load of 0.02 Erlangs. Calculate the total subscriber capacity in each cell. Next determine the total capacity in terms of subscribers per MHz per kilometre squared. Finally, what is the total subscriber capacity of the network?

Answer

Part 1

Calls per cell = 5/(0.025*7) = 28.57

Subscribers per cell = 28.57/0.02 = 1428.57

Part 2

Assuming a call bandwidth of 25 kHz, the 5 MHz system has 5/0.025 channels (200 calls)

These channels are used over a *cluster area* of 7 * 452 km^2 = 3164 km^2

Thus, total capacity = 200/(5*3164) = **0.0126 calls/MHz/km^2**

Subscriber capacity = (0.0126/0.02) = **0.632 subscribers/MHz/km^2**

Part 3

Total subscriber capacity = 1428.57*100 = **142,857**

or = 0.632*100*452*5 = 142,832

Total coverage area = 100 * 452 = **452,000 km^2**

Example 2:

A digital cellular system has a total bandwidth of 15 MHz and operates with a cluster size of 4 and a cell radius of 2 km. The network compromises 20 basestations and each 200 kHz carrier is made up of 8 TDMA voice channels. Assuming that each user represents a traffic load of 0.02 Erlangs. Calculate the total subscriber capacity in each cell. Next determine the total capacity in terms of subscribers per MHz per kilometre square. Finally, what is the total subscriber capacity of the network?

Answer

Part1

Calls per cell = 15/((0.200/8)*4) = 150

Subscribers per cell = 150/0.02 = 7500

Part 2

Assuming a 200 kHz call bandwidth, the 15 MHz system supports a total of 600 calls.

These channels are used over a cluster area of 4 * 12.6 km^2 = 50.4 km^2

Thus, total capacity = 600/(15*50.4) = **0.794** calls/MHz/ km^2

Subscriber capacity = (0.794/0.02) = **39.68** subscribers/MHz/ km^2

Part 3

Total subscriber capacity = 7500*20 = **150,000**

or = 39.68*20*12.6*15 = 150,000

Total coverage area = 20 * 12.6 = **252 km^2**

Microcellular Systems

The concept of a microcell has been introduced into cellular networks to provide increased capacity within a limited radio spectrum allocation. Although microcells are not formally defined, they generally refer to small cells (<1 km) in an urban area where the base station antennas are significantly lower than the surrounding roof tops. In Bristol, a high density of microcells can be found in Clifton. Unlike conventional large cell scenarios, where base station antennas are placed on the tallest buildings, a microcell's coverage area is largely governed by the effects of the surrounding buildings (i.e. by urban shielding). Microcells often operate with lower transmit powers at the base station, and this further reduces the coverage area.

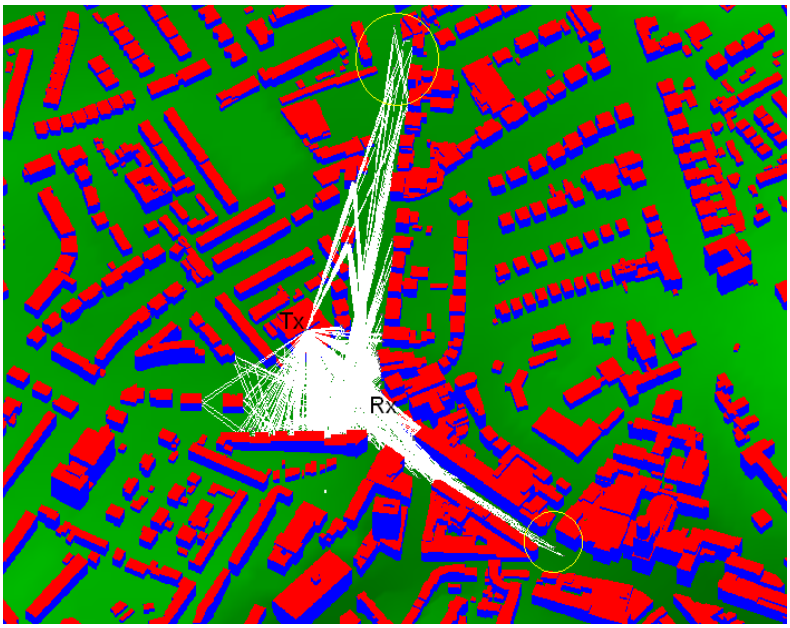


Figure 9: Deterministic Modelling for Microcells

The lower transmit power, combined with the effects of urban shielding, results in a shorter co-channel reuse distance, hence cell frequencies can be reused more often in a given area. A high density of reuse cells will naturally increase system capacity.

For microcellular studies, the propagation models must take into account the clutter introduced by individual buildings. Therefore, in a cochannel interference limited environment, the capacity can be deterministically evaluated by studying the effective coverage area of each individual cell. The propagation coverage for each cell site can be obtained either statistically (by fitting curves to measured results) or deterministically (see Figure 9) by methods such as ray-tracing (an active research area within the University of Bristol).

Microcell Capacity Evaluation

The first step in determining microcellular system capacity is to estimate the average cell coverage area, this is normally achieved by assuming the system to be noise limited and using measurements or models to determine how the average power varies with separation distance from the basestation. The average cell radius is defined as the average distance from the basestation where the required signal power and outage probabilities can be maintained.

For a cochannel limited system it is important to design the network to have the minimum practical separation distance between reuse frequencies. This minimum distance depends on many factors, such as the number of cochannel cells in the

vicinity of the centre cell, the type of geographic terrain, the antenna height, the antenna pattern, the sectorisation and the transmit power at each base site.

Figure 10 shows the situation where the wanted cell is receiving interference from six cochannel cells. With reference to the required signal to interference ratio, the reuse distance can be reduced until the C/I ratio is met. The diagram shows the received signal level relative to the cochannel interference for a given cochannel separation distance. As the cochannel cells are moved closer, there comes a point where the interference ratio and the outage probabilities are no longer met. This point represents the minimum practical reuse distance. Based on knowledge of the cell size *R* and the reuse distance *D* the cluster size can be determined.

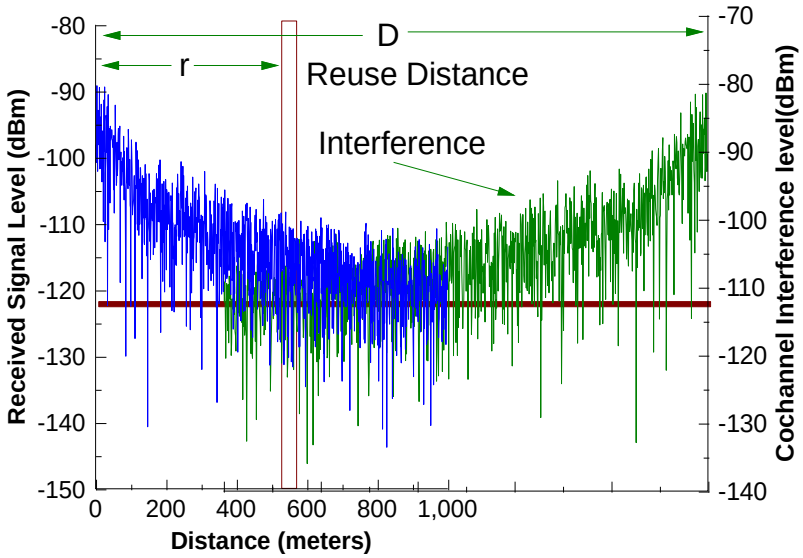


Figure 10: Determination Of Average Reuse Distance

Example 3:

A digital microcellular system has a total bandwidth of 15 MHz and operates with a cluster size of 7 and a cell radius of 160m. The network compromises 20 basestations and each 200 kHz carrier is made up of 8 TDMA voice channels. Assume that each user represents a traffic load of 0.02 Erlangs and that the average reuse distance is 600m. Calculate the total call and subscriber capacity in each cell and for the entire network.

Answer

Calls per cell = (15/(7*0.2))*8 = **85.71**
Subscribers per cell = **4,285.7**

Assuming a 200 kHz call bandwidth, the 15 MHz system supports a total of 600 calls.
These channels are used over a cluster area of 7 * 0.0804 km^2= 0.563 km^2
Thus, total capacity = 600/(15*0.563) = **71.04** calls/MHz/km^2
Subscriber capacity = (71.04/0.02) = **3,552** subscribers/MHz/km^2;
Total subscriber capacity = 4,285.7*20 = **85,714**
or = 3,552*20*0.0804*15 = 85,714
Total coverage area = 20 * 0.0804 = **1.608 km^2**

The above example illustrates how microcellular systems can achieve the very high traffic densities required in dense urban areas. A microcellular system can achieve very high user densities but these are normally limited to a small geographic coverage area. The above system supports just over 85,000 subscribers in a 1.6 km^2 area (the system has a capacity of over 3,500 subscribers/MHz/km^2). In contrast, the first two examples, which were more traditional cellular systems, achieved capacities of 0.63 and 39.68 subscribers/MHz/km^2. The first two examples offered much larger coverage areas (452,000 km^2 and 252 km^2 respectively). In practice, the choice of cell size and capacity must be accurately matched to the number of users in a given area. Modern mobile communication systems employ a mixture of small cells in the cities (often running a 3G service) and large cells for more rural areas (often running GSM). Large cells are also more appropriate for vehicular use to reduce the number of expected handovers.

It is interesting to note that some operators aim for high levels of geographic coverage (i.e. 90% of the country), while others aims for high levels of population coverage (i.e. 90% of the population). Since 90% of the population tend to live in cities, this can be achieved with a relatively low level of geographic coverage (for example, 15% of the population can be supported with a London network).