Project #: E-21-620  
Center #: N6478-0A0  
Contract#: AGR DTD 3/31/88  
Prime #:  
Subprojects #: N  
Main project #:  

Project unit: EE  
Project director(s): STUBER G L  

Sponsor/division names: BELLSouth ENTERPRISES  
Sponsor/division codes: 251 / 004  
Award period: 880329 to 910120 (performance) 910120 (reports)  

Sponsor amount  
Contract value 0.00  
Funded 0.00  
Cost sharing amount  

Does subcontracting plan apply #: N  
Title: CODING FOR DIGITAL CELLULAR RADIO  

- PROJECT ADMINISTRATION DATA -  
OCA contact: E. Faith Gleason 894-4820  
Sponsor technical contact  
NEALE HIGHTOWER  
(404)249-4510  
BELLSouth ENTERPRISES, INC  
Sponsor issuing office  
NEALE HIGHTOWER OR DOUG O'NEIL  
(404)249-4503  
BELL SOUTH ENTERPRISES, INC.  
SUITE 450  
1275 PEACHTREE STREET, NE  
ATLANTA, GA 30309  

Security class (U,C,S,TS) :  
Defense priority rating :  
Equipment title vests with: Sponsor X  

Onr resident rep. is ACO. (Y/N): N  
Supplemental sheet: GIT  

Administrative comments -  
* MODIFICATION NO. 6 PROVIDES A NO-COST EXTENSION TO JANUARY 31, 1991.
DIGITAL CELLULAR RADIO: EVALUATION SIMULATOR

Report for
OCA Project No. E-21-620

Prepared for
BellSouth Enterprises, Inc.

by

Gordon L. Stüber, Aubrey M. Bush, and Stephen B. Wicker

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia  30332

Sponsor Technical Contact

Neale Hightower

BellSouth Enterprises, Inc.
1155 Peachtree Street, NE
Atlanta, Georgia  30367

December 1988
DIGITAL CELLULAR RADIO: EVALUATION SIMULATOR

Report for
OCA Project No. E-21-620

Prepared for
BellSouth Enterprises, Inc.

by
Gordon L. Stüber, Aubrey M. Bush, and Stephen B. Wicker

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Sponsor Technical Contact

Neale Hightower

BellSouth Enterprises, Inc.
1155 Peachtree Street, NE
Atlanta, Georgia 30367

December 1988
Acknowledgements

We would like to thank Research Assistant Eric Long for his assistance in implementing the channel simulator and generating the errored data streams required for the voice coder experiments. Thanks are also given to Prof. Tom P. Barnwell for providing the facility to test the performance of various speech coders as well as the speech coders themselves.
1 Introduction

The following report describes a channel simulation package, developed within the School of Electrical Engineering at the Georgia Institute of Technology, that allows the evaluation of proposed digital evolutionary strategies for the North American cellular mobile phone system. The current version of the channel simulator uses SYSTID, a system simulation package donated to the School of Electrical Engineering by Hughes, along with a test station developed by Prof. T. P. Barnwell in the School of Electrical Engineering as part of an unrelated project. The channel simulator that has been developed is written in FORTRAN and is flexible enough to be incorporated into any system simulation package with minor modifications.

The usefulness of the channel simulator is discussed in the report. For any given transmission scheme, it is demonstrated how the channel simulator may be used to compare various voice coding options by using subjective voice tests. It is also shown how the same channel simulator may be used to compare different digital transmission schemes for a particular voice coder. The comparisons in this latter case are in terms of the tradeoffs between spectral efficiency, cell reuse patterns, cell sizes, and transmitter power requirements. By using the simulation facility described in the report, it is possible to estimate the performance of any proposed digital cellular system.

The report is organized as follows: Chapter 2 discusses the mobile radio propagation environment in detail. In particular, the connection between the channel propagation characteristics and the design of a cellular radio link is explained. The relationship between the required transmitter power and cell sizes is discussed, with the objective of maintaining a specified outage due to thermal noise. A new model is presented for computing the outage due to co-channel interference that is caused by interfering base stations, and the relationship between co-channel interference outage and spectral efficiency is discussed. The multipath-fading characteristics of a mobile radio channel are discussed. Finally, two channel simulators are proposed and tested. One simulator is based on Freudberg's approach along with a first order Markov model. The second is a simulator is one that has been suggested by Jakes.

Chapter 3 discusses various aspects of the SYSTID system simulation package. The
basic time domain analysis approach is described, along with the procedure by which various modules may be incorporated into the package. The flexibilities and limitations of the package are addressed.

Chapter 4 discusses the simulated performance of a particular digital cellular system. This system is based upon the 4-channel sub 30 kHz FDMA system being proposed by NEC. Because of the lack of available information, several "best guesses" had to be made concerning detailed system parameters. The results of the simulation experiments are summarized in terms of the probability of bit error vs. carrier-to-noise ratio and the probability of bit error vs. carrier-to-interference ratio characteristics. Such characteristics allow the objective comparison of different cellular systems.

Chapter 5 discusses the 16 kb/s subband speech coder (SBC) and the 4.8 kb/s self excited vocoder (SEV) developed by T.P. Barnwell, that were used in a subjective voice test experiment. It is important to note that these particular voice coders are not ideal for a mobile radio environment because they do not assume the use of any error correction coding, and in the case of the SEV vocoder, the rate is too low. However, this does not affect the basic methodology used to test voice coders.

Chapter 6 presents the results of the subjective voice experiment for the system described in chapter 4 and the SBC and SEV described in chapter 5, in the form of a demonstration tape. The cassette tape is included with, and is an integral part of, this report.
2 Mobile Radio Propagation

This chapter provides a discussion of the propagation characteristics of land mobile radio channels. The effect of path loss, shadowing, multipath-fading, and background noise on the design of a cellular mobile radio system is discussed in detail. Methods are proposed that allow the comparison of different cellular systems with respect to spectral efficiency and outage. These methods rely upon software simulations of proposed systems. A key component in the development of the required software simulation packages is an effective model of the cellular mobile radio channel.

2.1 Propagation and Radio Link Design

Propagation at UHF/VHF frequencies used in cellular mobile radio systems is largely influenced by three nearly independent factors; path loss variation with distance, slow log-normal shadowing, and fast multipath Rayleigh fading. Each of these phenomena is caused by a different underlying physical principle, and each must be accounted for when designing and evaluating the performance of a cellular system. Mobile radio systems typically exhibit a thresholding effect, where the transmission quality will be acceptable provided that both the received carrier-to-noise ratio (CNR) ($\Gamma$) and the received carrier-to-interference ratio (CIR) ($\Lambda$) exceed certain thresholds, denoted by $\Gamma_{th}$ and $\Lambda_{th}$, respectively. These thresholds depend upon the particular digital transmission methodology and the characteristics of the multipath-fading channel. Once $\Gamma_{th}$ and $\Lambda_{th}$ have been specified, shadowing and path loss determine the channel outage. Outage is defined as the fraction of the service area over which the transmission quality cannot be maintained.

There are two types of outage; that due to thermal noise (thermal noise outage), and that due to co-channel interference (co-channel interference outage). Both types of outage are caused by the combination of log-normal shadowing and the randomness of mobile locations within a cell.

When designing any cellular mobile radio system, it is necessary to specify two parameters; the minimum transmitter power and the co-channel reuse factor ($D/R$). Below, we discuss how these parameters are determined, how they depend upon the channel propagation characteristics, and how spectral efficiency and outages are computed from them.
2.2 Determining the Minimum Transmitter Power

Consider a cellular system for which the cell radius $R$ and the threshold CNR $\Gamma_{th}$ are known. To determine the minimum required transmitter power, the allowable thermal noise outage must be specified. Such a specification is necessary when designing a cellular system. Conversely, if the transmitter power is specified, the thermal noise outage may be calculated. This procedure is helpful when evaluating the performance of a cellular system.

Suppose that a mobile is at distance $r$ from a cell site. Let the average received CNR at distance $r$ be $\Gamma_m$ dB. The the probability density function of $\Gamma$, expressed in dB's, is

$$p(\Gamma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\Gamma - \Gamma_m)^2}{2\sigma^2}}. \quad (2.1)$$

The standard deviation $\sigma$ typically ranges from 5 to 12 dB and is independent of the length of the radio link. $\sigma = 8$ dB is a commonly used value.

Suppose that the cell has radius $R$ and let $\Gamma_m(R)$ be the value of $\Gamma_m$ on the cell fringe. The transmission quality will be acceptable provided that $\Gamma > \Gamma_{th}$ dB. Therefore, the outage of a mobile on a cell fringe is

$$F^1_j \equiv F^1_j(R) = P_r (\Gamma < \Gamma_{th}) = \int_{-\infty}^{\Gamma_{th}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\Gamma - \Gamma_m(R))^2}{2\sigma^2}} d\Gamma = Q \left( \frac{\Gamma_m(R) - \Gamma_{th}}{\sigma} \right). \quad (2.2)$$

In equation (2.2), $\Gamma_m(R) - \Gamma_{th}$ dB represents the minimum CNR margin on the cell fringe.

To obtain a relationship between the outage on a cell fringe and the outage within the whole cell, a propagation loss model must be specified. For this purpose the following simple model may be used.

$$\Gamma_m \equiv \Gamma_m(r) = 10 \log_{10} (Ar^{-\alpha}) \quad (2.3)$$

Typically, the value of $\alpha$ is about 3 to 4 in an urban area. By using this propagation model, the relation between the outage on a cell fringe and the average outage within the whole cell is

$$F^1_a = \frac{1}{\pi R^2} \int_0^R F^1_j(r) 2\pi r dr \quad (2.4)$$
where

\[ X_0 = \frac{\Gamma_m(R) - \Gamma_{th}}{\sigma}, \quad Y_o = \frac{2\sigma}{\alpha}. \]

The first term of this expression is equal to the outage at the cell fringe, \( F^1_f \), and the second term is the correction factor. The numerical results obtained by evaluating (2.4) are shown in Fig 2.1 for \( \alpha = 3.5 \).

By using Fig 2.1, \( F^1_a \) can be used to get \( F^1_f \). Then (2.2) can easily be solved for \( \Gamma_m(R) - \Gamma_{th} \). Fig. 2.2 is a plot of \( F^1_f \) versus \( \Gamma_m(R) - \Gamma_{th} \). If \( \Gamma_{th} \) is known from experiments or simulation, then \( \Gamma_m(R) \) can be obtained. Once \( \Gamma_m(R) \) is known, the minimum required transmitter power can be determined by using knowledge of the path loss and receiver noise power.
2.3 Determining the Minimum Co-Channel Reuse Distance

The co-channel reuse distance $D/R$ is the ratio of the minimum distance $D$ between base stations using a common set of channel frequencies and the cell radii $R$. By specifying the co-channel interference outage, the minimum required $D/R$ may be computed when designing a cellular system. Conversely, by specifying $D/R$, the co-channel interference outage may be determined when evaluating a cellular system. Accomplishing either of these tasks is much more difficult than determining the minimum required transmitter power as outlined in the previous section.

Many models have been proposed for computing the minimum required co-channel reuse distance. These models fall into two categories, geometric and statistical. Geometric models are constructed by considering the relative geographical locations of the transmitters and receivers. Although path loss is included, fading and shadowing are not included when the geometric models are developed. The general methodology with geometric models is
to include the effects of fading and shadowing when determining the threshold CIR $A_{th}$. This procedure, however, makes the determination of $A_{th}$ difficult because shadowing is a slowly changing stochastic process and therefore very large numbers of simulation runs are required.

The statistical models include all the propagation effects (path loss, fading, and shadowing) in a statistical fashion. The main problem with the statistical models is that they are mathematically cumbersome, if not intractable, when there are multiple interferers. Therefore, this approach is not acceptable.

With our methodology, the outage due to co-channel interference is determined by decoupling the effects of path loss and shadowing from the effects of fading. Specifically, this is accomplished as follows. For a given co-channel reuse distance, the outage is simply the probability that the threshold CIR $A_{th}$ exceeds the received CIR $A$. The threshold CIR is determined by performing subjective listening tests on a proposed cellular system assuming that the desired signal and the interfering signals are affected only by fading. The received CIR is a random variable that accounts for the effects of path loss and shadowing.

To compute the outage, the following assumptions are made concerning a cellular system:

1. There are six interfering base stations equidistant from the base station of the cell containing the intended receiver. The more distant interfering base stations are neglected in deference to the typically dominant effect of the first tier of interfering base stations.

2. A mobile can be anywhere within a cell with equal probability. Therefore, in the $x, y$ plane, the probability density function of the position of a mobile is $p(x, y) = \frac{1}{\pi R^2}$, where $R$ is the cell radius.

3. A common path loss model is used for the desired base-to-mobile channel and each of the interfering base-to-mobile channels. Several path loss models are described below.

4. The signal from the desired base station and each of the interfering base stations is affected by log-normal shadowing with independent and identical statistics. It is felt that independent shadowing is the worst case situation.
5. All base stations have a common height.

2.3.1 Path Loss Models

Path loss is the attenuation of a signal with distance, excluding shadowing and fading. Although, it may seem counterintuitive, path loss is necessary in cellular systems. The reason is that rapid attenuation with distance allows a small co-channel reuse distance and therefore good spectral efficiency, measured in \(\text{users/\(MHz/km^2\)}\). The path loss depends upon the distance between the base station and mobile. There are a multitude of technical reports concerned with path loss prediction methods for UHF/VHF land mobile radio in flat, urban, suburban, open, and hilly terrains. Most of these are empirical models; few of them are theoretical. Two useful empirical models are discussed here: Okumura's model and the area-to-area path loss prediction model.

Okumura's empirical model is probably the simplest to use, and can distinguish man-made structures. The empirical data for Okumura's model was collected in Tokyo. Be cautioned, however, that the path loss for Japanese suburban areas does not match North American suburban areas very well. The latter are more like the quasi-open areas in Japan, because the buildings in North America are more separated. Okumura's model is expressed in terms of the carrier frequency \(150 \leq f_c \leq 1500\) (in MHz), base-station antenna height \(30 \leq h_b \leq 300\) (in m), and the mobile-station antenna height \(1 \leq h_m \leq 10\) (in m). The empirical formula for the path loss is a function of the distance \(1 \leq r \leq 20\) (in km) between the base and mobile station and is accurate to within 1 dB for distances up to 20 km. With this model, the path loss in dB's is

\[
L_p = \begin{cases} 
A + B \log_{10}(r) & \text{for urban area} \\
A + B \log_{10}(r) - C & \text{for suburban area} \\
A + B \log_{10}(r) - D & \text{for open area}
\end{cases}
\]

where

\[
A = 69.55 + 26.16 \log_{10}(f_c) - 13.82 \log_{10}(h_b) - a(h_m) \\
B = 44.9 - 6.55 \log_{10}(h_b) \\
C = 5.4 + 2[\log_{10}(f_c/28)]^2
\]
\[ D = 40.94 + 4.78 \log_{10}(f_c) \\] 

and

\[ a(h_m) = \begin{cases} 
(1.1 \log_{10}(f_c) - 0.7)h_m - (1.56 \log_{10}(f_c) - 0.8) & \text{for medium or small city} \\
8.28(\log_{10}(1.54h_m))^2 - 1.1 & \text{for } f_c \geq 200 \text{ MHz} \\
3.2(\log_{10}(11.75h_m))^2 - 4.97 & \text{for } f_c \leq 400 \text{ MHz} 
\end{cases} \]

Another path loss model that is accurate and relatively easy to use is the area-to-area path loss prediction, as described by Lee [5]. The area-to-area prediction is generally used to predict a path loss over flat terrain, despite the fact that the terrain configuration over which the actual path loss is found may be unknown. Therefore, care must be exercised in its use. For example, if the actual terrain is hilly and an area-to-area path loss prediction is used, there will be a large difference between the actual and predicted path losses. Two parameters are required for the area-to-area path loss prediction model: (i) the power at the 1.6 km point of interception, \( P_{ro} \), and (ii) a path-loss slope, \( \gamma \). The field strength of the received signal can be expressed as

\[ P_r = P_{ro} \left( \frac{r}{r_o} \right)^{-\gamma} \left( \frac{f}{f_c} \right)^{-n} \alpha_o \]

\[ = P_{ro} - \gamma \log \left( \frac{r}{r_o} \right) - n \log \left( \frac{f}{f_c} \right) + \alpha_o \quad \text{dB} \quad (2.5) \]

where \( r \) is in kilometers and \( r_o = 1.6 \text{ km} \).

A 1.6 km point of interception is taken because within a 1.6 km radius there are very few streets available and, therefore, the statistical mean cannot be obtained. \( \alpha_o \) is an adjustment factor that is used to account for systems with different parameters, e.g. a different base station antenna height.

The following set of conditions is used to compute the path loss

- frequency \( f_c = 900 \text{ MHz} \)
- base-station antenna height = 30.48 m
- base-station power at the antenna = 10 watts
- base-station antenna gain = 6 dB above dipole gain
• mobile-unit antenna height = 3 m

• mobile-unit antenna gain = 0 dB above dipole gain

Then for a different set of conditions we have

\[
\begin{align*}
\alpha_1 &= \left( \frac{\text{new base-station antenna height (m)}}{30.48(m)} \right)^2 \\
\alpha_2 &= \left( \frac{\text{new mobile-unit antenna height (m)}}{3(m)} \right)^v \\
\alpha_3 &= \frac{\text{new transmitter power}}{10 \, \text{w}} \\
\alpha_4 &= \frac{\text{new base-station antenna gain with respect to } \lambda/2 \text{ dipole}}{4} \\
\alpha_5 &= \text{different antenna-gain correction factor at the mobile unit}
\end{align*}
\]

The value of \( \alpha_o \) is

\[
\alpha_o = \alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5 = \sum_{i=1}^{5} \alpha_i \, \text{dB}
\]

The parameters \( \gamma \) and \( P_{ro} \) are found from empirical measurements.

In free space:

\[
\begin{align*}
P_{ro} &= 10^{-4.5} \, \text{mwatts} \\
\gamma &= 2 \\
P_{ro} &= -45 \, dBm \\
\gamma &= 20 \, dB/dec
\end{align*}
\]

In an open area:

\[
\begin{align*}
P_{ro} &= 10^{-4.9} \, \text{mwatts} \\
\gamma &= 4.35 \\
P_{ro} &= -49 \, dBm \\
\gamma &= 43.5 \, dB/dec
\end{align*}
\]
In North American suburban areas:

\[ P_{ro} = 10^{-6.17} \text{ mwatts} \]  
\[ \gamma = 3.84 \]  
\[ P_{ro} = -61.7 \text{ dBm} \]  
\[ \gamma = 38.4 \text{ dB/dec} \]

In North American urban area (Philadelphia):

\[ P_{ro} = 10^{-7} \text{ mwatts} \]  
\[ \gamma = 3.68 \]  
\[ P_{ro} = -70 \text{ dBm} \]  
\[ \gamma = 36.8 \text{ dB/dec} \]

In North American urban area (Newark):

\[ P_{ro} = 10^{-6.4} \text{ mwatts} \]  
\[ \gamma = 4.31 \]  
\[ P_{ro} = -64 \text{ dBm} \]  
\[ \gamma = 43.1 \text{ dB/dec} \]

In Japanese urban area (Tokyo):

\[ P_{ro} = 10^{-8.4} \text{ mwatts} \]  
\[ \gamma = 3.05 \]  
\[ P_{ro} = -84 \text{ dBm} \]  
\[ \gamma = 30.5 \text{ dB/dec} \]

The value of \( n \) in equation (2.5) is

\[ 20 \text{ dB/dec} < n < 30 \text{ dB/dec}. \] (2.14)

This number is valid for ranges from 2 to 30 km. The value of \( n \) depends upon the carrier frequency and the geographical area. For \( f_c < 450 \text{ MHz} \) in a suburban or open area,
$n = 20\ dB/\text{dec}$ is recommended. In an urban area with $f_c > 450\ MHz$, $n = 30\ dB/\text{dec}$ is recommended.

The value of $v$ is also determined from empirical data.

$$v = \begin{cases} 
2 & \text{for a new mobile-unit antenna height} > 10\ m \\
1 & \text{for a new mobile-unit antenna height} < 3\ m
\end{cases} \quad (2.15)$$

The path loss $L_p$ is simply the difference between the transmitted and received field strengths, $L_p = P_t - P_r$. By using the above parameters for $P_{ro}$ and $\gamma$, the following path losses are obtained:

$$L_p = 80.8 + 20.0\log_{10} r + n\log_{10}(f/900) - \alpha_o \quad \text{dB} \quad \text{Free Space} \quad (2.16)$$

$$L_p = 80.0 + 43.5\log_{10} r + n\log_{10}(f/900) - \alpha_o \quad \text{dB} \quad \text{Open Area}$$

$$L_p = 93.7 + 38.4\log_{10} r + n\log_{10}(f/900) - \alpha_o \quad \text{dB} \quad \text{North American suburban}$$

$$L_p = 102.4 + 36.8\log_{10} r + n\log_{10}(f/900) - \alpha_o \quad \text{dB} \quad \text{Philadelphia}$$

$$L_p = 95.1 + 43.1\log_{10} r + n\log_{10}(f/900) - \alpha_o \quad \text{dB} \quad \text{Newark}$$

$$L_p = 117.7 + 30.5\log_{10} r + n\log_{10}(f/900) - \alpha_o \quad \text{dB} \quad \text{Tokyo}$$

These results along with those from Okumura's model are plotted in Figs. 2.3a and 2.3b for a base station height of 70 m, a mobile antenna height of 1.5 m, and a frequency of 900 MHz.

### 2.3.2 Computing the Co-Channel Interference Outage

By using the above path loss models, the co-channel interference outage can be computed by computer simulation. Such a task was undertaken, and typical results are summarized in Figs. 2.4 to 2.8. In all cases, a carrier frequency of 900 MHz, a base station antenna height of 70 m, and a mobile antenna height of 1.5 m was used when generating the simulation results.

Fig. 2.4 is a plot of the co-channel interference outage, $P_r(A < A_{th})$ assuming that the mobile is located on the fringe of a cell. The outage is plotted as a function of the co-channel reuse factor $D/R$. Okumura's large city path loss model was used. It was also assumed that both the desired signal and the interfering signals are affected by log-normal shadowing.
Fig. 2.3a Propagation Path Loss in Different Areas
Fig. 2.3b Propagation Path Loss in Different Areas
The area averaged co-channel interference outage was also obtained by simulation, where the mobile was assumed to be anywhere in the cell area with equal probability. The results are plotted in Fig. 2.5.

Fig 2.6 is a plot of the co-channel interference outage for various area-to-area path loss models with a $D/R$ of 4.3 ($D/R = 4.3$ is used in the North American AMPS system). Both the desired signal and the interfering signals are affected by shadowing. Referring to Fig. 2.6, the path loss models used are, from top to bottom, Tokyo, Philadelphia, North American suburban, Newark, and an open area. Fig 2.7 is the same plot, except that only the desired signal (not the interferers), is affected by shadowing. Note that the outage is worse when the interferers are affected by shadowing.

Finally, Fig. 2.8 is a plot of the outage using the area-to-area path loss model for Philadelphia with a $D/R$ of 4.3. In this figure, the standard deviation of the log normal shadowing is varied. Both the case of shadowed interferers and the case of unshadowed interferers is considered. Note how shadowing increases the outage, and the outage with shadowed interferers is greater than the outage with unshadowed interferers.

There are several ways that plots of the kind shown in Figs. 2.4 to 2.8 may be used when designing or evaluating cellular systems. For a given co-channel interference outage and carrier to interference threshold $\Lambda_{th}$, the required co-channel reuse factor $D/R$ may be computed. This in turn gives the number of cells per cluster as $N = (D/R)^2/3$, where $N$ takes on restricted values from the set {$1, 3, 4, 7, 9, 12, 13, \ldots$}. Alternatively, if $N$ and $\Lambda_{th}$ are specified, the co-channel interference outage may be computed.

The area of a cell $A \ km^2$, and hence its radius $R$ is determined by the transmitter power along with the allowable thermal noise outage. Alternatively, if the transmitter power and cell radii are specified, the thermal noise outage may be computed.

These relations can easily be used to compute the spectral efficiency. The spectral efficiency is

$$\eta = \frac{1}{B_c \cdot N \cdot A} \quad (2.17)$$

where $B_c$ is the carrier separation in $MHz$. Note that this equation implies one channel per carrier. The effect of using multiple channels per carrier, such as in narrow-band 30 $kHz$ TDMA schemes, can easily be accounted for by increasing $\eta$ by the multiplexing factor.
Fig. 2.4 Co-channel Interference Outage of a Mobile on the Cell Fringe
Fig. 2.5 Area Averaged Co-channel Interference Outage
Fig. 2.6 Co-channel Interference Outage with Various Path Loss Models. Interferers are Shadowed.
Fig. 2.7 Co-channel Interference Outage with Various Path Loss Models. Interferers not Shadowed.
Fig. 2.8 Co-channel Interference Outage with Varying Degrees of Shadowing
Finally, there is a tradeoff between outage and spectral efficiency. To see this, suppose that $D/R$ (and hence $N$) and $A$ are fixed. Then a larger outage means that the $A_{th}$ and $\Gamma_{th}$ can be larger. This in turn implies that $B_c$ can be smaller, and the spectral efficiency is increased.

### 2.4 Multipath-Fading

This section describes the multipath-fading channels commonly found in mobile radio environments. A good mathematical model of the mobile radio channel is necessary if meaningful simulations of proposed cellular systems are to be undertaken. Throughout this section, it will be assumed that the transmitted signals are vertically polarized and are sensed by a vertical whip (omnidirectional) antenna on the mobiles.

#### 2.4.1 Statistical Characterization of the Channel

Due to the combined effect of vehicle movement and multipath propagation, rapid and extreme phase and frequency modulation is introduced in the received signal, commonly referred to as fading. The statistical properties of multipath-fading in a mobile environment have been studied by a number of authors, including Jakes [3], Proakis [6], and Lee [4,5]. For our purpose, these statistical properties provide the theoretical foundation for the implementation of various multipath-fading channel simulators.

In multipath-fading channels, the received signal is the sum of many independently scattered components. Therefore, multipath-fading can be modeled as a complex-valued Gaussian random process. If the process has zero mean, then the envelope of the received signal at any time is Rayleigh distributed while the phase is uniformly distributed. However, if there also exists a specular component in the received signal, then the Gaussian process has non-zero mean and the envelope statistics are Rician. A system designed under the assumption of Rayleigh fading will always perform satisfactorily over a Rician fading channel. For this reason, most studies only consider Rayleigh fading.

Multipath-fading channels are characterized by two kinds of spreading: Doppler spread $B_d$ (spreading in frequency), and multipath spread $T_m$ (spreading in time). From these define $\Delta f_c = 1/T_m$ as the "coherence bandwidth" and $\Delta t_c = 1/B_d$ as the "coherence
time" of the channel. Roughly, two sinusoids having a frequency separation greater than \( \Delta f_c \) will be affected independently by the channel. Similarly, samples of the channel output having a time separation greater than \( \Delta t_c \) will be affected independently by the channel.

In digital systems, the medium characteristics are based on the signal duration \( T \) relative to \( \Delta t_c \), and signal bandwidth \( W \) relative to \( \Delta f_c \). Consequently, multipath-fading channels can be further classified as nondispersive (\( T \ll \Delta t_c, W \ll \Delta f_c \)), time-dispersive (\( W \ll \Delta f_c, T \gg \Delta t_c \)), frequency-dispersive (\( T \ll \Delta t_c, W \gg \Delta f_c \)), and doubly-dispersive (\( T \gg \Delta t_c, W \gg \Delta f_c \)). In many practical systems, \( W \) and \( T \) are chosen such that the channel is nondispersive. Sometimes, such a channel is referred to as being underspread.

### 2.4.2 Time Dispersion and the Coherence Bandwidth

The time dispersive channel characteristics are described by a "multipath intensity profile". The multipath intensity profile \( \phi(r) \) is just the average power output of the channel as a function of the time delay \( r \). For mobile radio, the interpretation of measured data has indicated that the following exponential multipath intensity profile is a good approximation:

\[
\phi(r) = \frac{1}{T_m} \exp \left( -\frac{r}{T_m} \right).
\]  

(2.18)

The channel will introduce intersymbol interference (ISI) unless \( T \gg T_m \). Typical values of \( T_m \) are included in Table 2.1. These values are valid for any operating frequency above 30 MHz. From Table 2.1, the coherence bandwidth of a typical urban channel is on the order of 100 KHz. Suburban and open areas have even a larger coherence bandwidth. Therefore, the effects of time dispersion can be ignored for both sub-30 kHz FDMA and 30 kHz narrow-band TDMA systems. For this reason, the time-dispersive characteristics of the mobile radio channel will not be discussed further. If further details are required, the interested reader is referred to Jakes [3] or Proakis [6].

### 2.4.3 Frequency Dispersion and the Coherence Time

Characterization of the frequency dispersive characteristics of the channel is very important for mobile radio applications. Doppler spread due to vehicle motion causes time variations in the envelope of the received signal. For digital systems, the coherence time of the channel
Table 2-1
Typical Delay Spreads

<table>
<thead>
<tr>
<th>Environment</th>
<th>Delay-Spread $T_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Area</td>
<td>$&lt; 2\mu s$</td>
</tr>
<tr>
<td>Suburban Area</td>
<td>$0.5\mu s$</td>
</tr>
<tr>
<td>Urban Area</td>
<td>$3 - 10\mu s$</td>
</tr>
</tbody>
</table>

is inversely proportional to the Doppler bandwidth, and hence vehicle speed, so that faster vehicle speeds result in faster fading. By using fundamental arguments, the power density spectra of the received signal from a CW transmission has the form

$$S(f) = \frac{3b_o}{\omega_m} \left[ 1 - \left( \frac{f - f_c}{f_m} \right)^2 \right]^{-1/2},$$

(2.19)

where $b_o$ is the mean signal power, and $f_m = v/\lambda$ with $v$ being the vehicle velocity and $\lambda$ being the wavelength of the transmitted tone. The power spectral density is illustrated in Fig. 2.9.

The autocorrelation of the in-phase and quadrature components of the received signal has the form

$$g(\tau) = b_o J_o(\omega_m \tau),$$

(2.20)

where $J_o(\cdot)$ is a zero order Bessel function of the first kind. It is desirable that a channel simulator maintain this autocorrelation characteristic. In this way, the envelope characteristics of the faded signal will be maintained. The envelope characteristics include the envelope autocorrelation, which can be expressed in terms of a hypergeometric function as follows:

$$R_r(\tau) = \frac{\pi}{2}b_o F \left[ \frac{1}{2}, -\frac{1}{2}; 1; \rho^2(\tau) \right]$$

(2.21)

$$= \frac{\pi}{2}b_o \left[ 1 + \frac{1}{4}\rho^2(\tau) \right],$$

(2.22)

where

$$\rho^2(\tau) = \frac{g^2(\tau)}{b_o^2}.$$

(2.23)
From this autocorrelation, the power spectral density of the envelope can be derived by taking the Fourier transform. It is given by

$$S_o(f) = \frac{b_o}{4\omega_m} K \left[ \sqrt{1 - \left( \frac{f}{2f_m} \right)^2} \right], \quad (2.24)$$

where $K(\cdot)$ is the complete elliptic integral of the first kind.

2.4.4 Level Crossing Rate and Fade Duration

Crossing rates, average fade durations, and fade duration distributions are all second-order statistics. These statistics are important for determining the effect of burst errors on digital communication systems. The level crossing rate and average fade duration for a Rayleigh fading channel is known, while the fade duration distribution is unavailable in closed form.

The level crossing rate (LCR) is just the rate at which the received signal power crosses a threshold relative to the average received signal power. The LCR increases with
Fig. 2.10 Normalized Level Crossing Rates of a Rayleigh Faded Signal, from [3]

the vehicle velocity. For Rayleigh fading, the level crossing rate is

\[ N_R = \sqrt{2\pi f_m \rho e^{-\rho^2}} \]  

(2.25)

where \( \rho = R / R_{rms} \). Fig. 2.10 shows the level crossing rates for mobile radio signals assuming Rayleigh fading. As an example, consider a signal at 1000 MHz received by a dipole at a mobile unit traveling 100 km/h. Suppose we wish to find the expected level crossing rate at \( \rho = 0 \) dB. In this case \( v = 27.8 \) m/s and \( \lambda = 0.3 \) m, so that \( f_m = 92.6 \) Hz. Then \( N_R = 85.4 \) crossing/s.

The average duration of Rayleigh fades is given by

\[ \tau = \frac{e^{\rho^2} - 1}{\rho f_m \sqrt{2\pi}} \]  

(2.26)

25
Fig. 2.11 Normalized Duration of Rayleigh Fades, from [3]

Fig. 2.11 shows the average duration of Rayleigh fades as a function of the received signal level with respect to the average received signal level. Typically, the fades tend to be rather short in duration relative to the frequency of their occurrence.

2.4.5 Random FM

Fast Rayleigh envelope fading is accompanied by fast phase changes that introduce random FM noise on the received carrier. The power spectral density of random FM noise extends to about twice the maximum Doppler frequency, i.e., \(2v/\lambda \text{ Hz}\), where \(v\) is the vehicle velocity and \(\lambda\) is the wavelength. For example, at 36 km/h and 900 MHz the random FM extends to about 75 Hz. The random FM noise usually presents no problem in analog FM systems.
because it can be filtered out with negligible effect on the speech waveform. For digital systems, it is necessary to use signaling waveforms that do not have energy concentrated in the random FM range. That is, the equivalent baseband signals must have low spectral concentrations around zero frequency.

2.5 Simulation of Multipath-Fading Channels

In this section, two multipath-fading simulators are presented. The first of these is simple to implement, but has some drawbacks. The second is of moderate complexity and models the channel quite accurately.

2.5.1 Freudberg's Method

Freudberg's [2] concept of a multipath-fading simulator is depicted in Fig. 2.12. Sometimes Freudberg's method is referred to as the first order Markov method. The in-phase and quadrature components are amplitude modulated with uncorrelated low-pass Gaussian noise sources. Freudberg's method provides uniform phase modulation and Rayleigh envelope fading. The noise sources must have the same power density spectrum to produce stationary fading, and the power spectrum of the fading signal will be the same as that of the noise sources. The main limitation with this approach is that only rational forms of the fading spectra can be produced. In reality, the fading spectra is nonrational as demonstrated by equation (2.18).

The choice of the low pass filters used to generate the fading signal is discretionary. The simplest is a first order low pass filter. Suppose that the fading is modeled as a discrete time Markov process. Let \( a_k \) and \( b_k \) represent the fade levels of the in-phase and quadrature components of the received signal. Then \( a_k \) and \( b_k \) are Gaussian random variables with the state equation

\[
(a_{k+1}, b_{k+1}) = \gamma(a_k, b_k) + (1 - \gamma)(w_{1k}, w_{2k}) ,
\]

where \( w_{1k} \) and \( w_{2k} \) are independent Gaussian random variables with zero mean and \( E[w_{ik}w_{ij}] = \sigma^2 \delta_{ki} \ i, j = 1, 2 \). The envelope \( \sqrt{a_k^2 + b_k^2} \) is Rayleigh distributed. It can be shown that the
autocorrelation of $a_k$ and $b_k$ is

\[ g(n) = E[a_k a_{k+n}] = E[b_k b_{k+n}] \]
\[ = (1 - \zeta^2 \sigma^2 \zeta^n) \quad n \geq 0. \]  \hspace{1cm} (2.28)

From equation (2.20), we see that the desired autocorrelation is

\[ g(n) = b_0 J_0(\omega_m T k) \quad , \]  \hspace{1cm} (2.29)

where $T$ is the sampling rate. Therefore, a zero order Bessel function of the first kind is being approximated by an exponential function. The remaining problem is the specification of $\sigma^2$ and $\zeta$.

To determine what $\sigma^2$ and $\zeta$ should be, first note that the noise spectral density at the output of the filter is

\[ S_n(\omega) = \frac{\sigma^2}{1 + \zeta^2 - 2\zeta \cos \omega T} . \]  \hspace{1cm} (2.30)
This is to be compared with the power spectral density of a faded carrier shown in Fig. 2.9.

From Fig. 2.9, observe that the spectrum is confined to the interval \( \omega_c - \omega_m \leq \omega \leq \omega_c + \omega_m \).

To approximate this spectrum, it is arbitrarily assumed that the corner frequency of the first order low pass filter is at \( \omega_m/4 \), where \( \omega_m = 2\pi v\lambda \). This gives the value of \( \zeta \) as

\[
\zeta = 2 - \cos(\omega_m T) - \sqrt{(2 - \cos \omega_m T)^2 - 1} .
\]

To ensure that the average power is normalized to 0 dB, we choose

\[
\sigma^2 = \frac{1 + \zeta}{1 - \zeta} .
\]

This simulator has been implemented and tested. Figs. 2.13 and 2.14 show sample outputs from the simulator assuming a carrier frequency of 900 MHz and vehicle velocities of 15 km/h and 45 km/h, respectively.

1.4.2 Jakes's Method

A method that replicates all the desired channel characteristics has been proposed by Jakes [3]. A diagram of Jakes's simulator is shown in Fig. 2.15. The mathematical development leading to this simulator can be found in [3]. Here, we only describe its realization and characteristics. The simulator consists of \( N_o \) low frequency oscillators with frequencies equal to \( \omega_m \cos(2\pi n/N_o) \), \( n = 1, 2, \ldots, N_o \), plus one with frequency \( \omega_m \), where \( \omega_m = 2\pi v/\lambda \) and \( N_o = \frac{1}{4}(N/2 - 1) \). Except for the oscillator of frequency \( \omega_m \), which has an amplitude of \( 1/\sqrt{2} \), all the oscillator amplitudes are unity. The proper oscillator phases are provided by amplifiers with gains set equal to \( 2 \cos \beta_n \) or \( 2 \sin \beta_n \), where \( \beta_n = \pi n/N_o \).

Jakes has tested the channel simulator in Fig. 2.15. The simulator provided excellent agreement with the theoretical properties described in section 2.4.

Jakes's method can be easily extended to provide up to \( N_o \) independently faded signals by using two quadrature low-frequency oscillators per offset in place of the single oscillators. The outputs of the quadrature oscillators are amplified with the modified gains as shown in Fig 2.16 for the \( n^{th} \) offset amplifier of the \( f^{th} \) simulator.

Jakes's simulator has been implemented and tested. Figs. 2.17 and 2.18 show sample outputs from the simulator assuming a carrier frequency of 900 MHz and vehicle velocities of 15 km/h and 45 km/h, respectively.
Fig. 2.13  Simulated Fading Signal Using Freudberg's Method, 900 MHz, 15 km/h
Fig. 2.14 Simulated Fading Signal Using Freudberg's Method, 900 MHz, 45 km/h
OFFSET OSCILLATORS

\[ \cos \omega_t \]

\[ 2 \sin \beta_1 \]

\[ 2 \cos \beta_1 \]

\[ \sqrt{2} \cos \omega_{\text{osc}} \]

\[ \cos \omega_{\text{osc}} \]

\[ x(t) \] and \( x(t) \) ARE APPROXIMATELY GAUSSIAN RANDOM PROCESSES.

IA IS RAYLEIGH-DISTRIBUTED.

**Fig. 2.15 Jakes’s Fading Simulator, from [3]**
Fig. 2.16 Jakes's Fading Simulator for Uncorrelated Fading Carriers, from [3]

2.6 Simulation Studies

As described in sections 2.2 and 2.3, the design and performance evaluation of cellular mobile radio systems requires the specification of the threshold CNR $\Gamma_{th}$ and the threshold CIR $\Lambda_{th}$. By using the multipath-fading channel simulators, $\Gamma_{th}$ and $\Lambda_{th}$ can be easily determined. To determine $\Gamma_{th}$, the end-to-end performance of the proposed system is evaluated when it is operated over a Rayleigh faded channel with additive white Gaussian noise. Likewise, to determine $\Lambda_{th}$, the end-to-end performance of the proposed system is evaluated when it is operated over a Rayleigh faded channel with six mutually noncoherent independently Rayleigh faded interferers. In either case, subjective listener tests are used to identify the thresholds.
Fig. 2.15 Simulated Fading Signal, 900 MHz, 15 km/h
Fig. 2.16 Simulated Fading Signal, 900 MHz, 45 km/h
References


3 The SYSTID Simulation Software Package

In February of 1988 the Hughes Aircraft Company made a substantial software donation to Georgia Tech through Professor Wicker. The bulk of this software was a communication system simulation package called SYSTID (SYSstem TIIme Domaiin simulation). SYSTID has been under continuous development within the System Laboratories of Hughes' Space and Communication Group for the past 20 years. It began as a sample data simulation library designed for use on specific NASA satellite programs. Over the years SYSTID has evolved into a powerful flexible simulator for virtually all communication systems. It is currently an integral part of the system engineering segment for several commercial and military satellite design efforts at Hughes Aircraft. If placed on the open market, SYSTID would have an estimated value of a quarter of a million dollars. In the following sections the SYSTID simulation methodology will be discussed. An example of a typical simulation will then be presented, followed by an overview of the simulation laboratory currently in place at Georgia Tech.

3.1 The SYSTID Simulation Methodology

SYSTID provides flexible topology simulations at speeds similar to the best of the fixed topology simulators through the integration of a large library of device models with a simulation language processor. It is in essence a block diagram driven simulator with multiple levels of flexibility. This can best be seen by following the simulation process from start to finish.

The first step in simulating a communication system is to break the system up into functional units, or “blocks”. These blocks (filters, modulators, encoders, etc.) are represented in the simulation by a series of models. SYSTID’s model library already contains representations for a large number of devices. An abbreviated listing is shown in Figure 3-1. If a unique application calls for a model that is not in the library, it is a simple matter to write a model in FORTRAN and incorporate it into the model library. This flexibility is one of SYSTID’s major strengths, and makes it a particularly good analysis tool for land mobile radio systems. Separate models may be written to represent multipath channels in urban, suburban, and rural environments. In addition, detailed models of specific modulators and
demodulators using measured data can be incorporated into the simulations. These specific models will be discussed later in this report.

The models are linked together into a topological description of the communication system using the SYSTID simulation language. This system description is sent to the language processor, which adds simulation control information and translates the result into FORTRAN source code [1]. During this process the library models are retrieved and integrated into the code. This simultaneous integration of the models and simulation control code into the FORTRAN source code allows for the use of reentrant unit descriptions. Reentrant code streamlines the simulation process significantly, obviating the need for duplication of some models. The resulting source code is combined with problem-specific FORTRAN routines and submitted to the host FORTRAN compiler.

The compiler generates object modules that are subsequently sent to the host loader for assembly into executable code. Precompiled code for certain models is added in this phase, completing the executable simulation code. The executable code is generally run in batch mode.

The operational flow of a SYSTID simulation is shown in Figure 3-2. The two-phase compilation and execution process allows for a high degree of flexibility and efficient use of the design engineer’s time [1]. Parametric studies may be run without having to execute the translation phase more than once. The run parameters may be varied during the compilation without altering the topology of the overall system.

A variety of outputs to a given simulation are available, depending on requests made in the SYSTID language description of the communication system. Data collection nodes can be defined between the blocks in the topological description. As signals are propagated in the time domain through the various blocks, samples will be saved in the designated nodes and recorded on the host disk. Various other printed and plotted outputs are possible as well. All of the outputs are saved for the postprocessing phase. The postprocessing package provides for the creation of time and frequency domain plots, histograms, eye diagrams, and a host of other representations of the information saved in the assorted nodes during the simulation. Examples of this capability will be shown in the sample simulation in the following section.
<table>
<thead>
<tr>
<th>SIGNAL GENERATORS</th>
<th>FILTERS</th>
<th>MISCELLANEOUS</th>
<th>ESTIMATORS</th>
<th>SOURCE ENCODERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHITE GAUSSIAN NOISE</td>
<td>BUTTERWORTH</td>
<td>LIMITERS</td>
<td>POWER SPECTRA</td>
<td>ANALOG TO DIGITAL</td>
</tr>
<tr>
<td>ARBITRARY PDF SOURCES</td>
<td>CHEBYCHEV</td>
<td>TWT</td>
<td>PROBABILITY OF BIT ERROR</td>
<td>DIGITAL TO ANALOG</td>
</tr>
<tr>
<td>PULSE GEN</td>
<td>BESSEL</td>
<td>ARBITRARY NONLINEARITIES</td>
<td>(MPSK, QASK)</td>
<td>MULTILEVEL PCM</td>
</tr>
<tr>
<td>TRANSCENDENTAL FUNCTIONS</td>
<td>BUTTERWORTH THOMPSON</td>
<td>TDA</td>
<td>SIGNAL DISTORTION</td>
<td>ORTHOGONAL</td>
</tr>
<tr>
<td>SQUARE WAVE</td>
<td>ELLIPTIC</td>
<td>UP/DOWNCONVERTERS</td>
<td>SIGNAL TO NOISE</td>
<td>FOURIER</td>
</tr>
<tr>
<td>ARBITRARY TABLES</td>
<td>QUASI ELLIPTIC</td>
<td>DIFFERENTIATOR</td>
<td>NOISE BANDWIDTH</td>
<td>HAAR</td>
</tr>
<tr>
<td>PN SEQ GEN</td>
<td>LEAD LAG</td>
<td>INTEGRATION</td>
<td>DELAY</td>
<td>MADAMARD</td>
</tr>
<tr>
<td>RANDOM BIT STREAM GEN</td>
<td>INTEGRATE AND DUMP</td>
<td>TIME DELAYS</td>
<td>CORRELATION</td>
<td>ORDERED MADAMARD</td>
</tr>
<tr>
<td>PN BIT STREAM GEN</td>
<td>QUADRATIC</td>
<td>PHASE DELAYERS</td>
<td>STATISTICS</td>
<td>DIGITAL INTERLEAVING AND</td>
</tr>
<tr>
<td>BPSK PN SIGNAL GEN</td>
<td>ARBITRARY POLES AND ZEROS</td>
<td>LATCHES AND GATES</td>
<td>EEVE PATTERN ANALYZER</td>
<td>DL INTERLEAVING</td>
</tr>
<tr>
<td>FM NOISE LOADED BASEBAND SIGNAL GEN</td>
<td>TRANSVERSAL</td>
<td>THRESHOLD DETECTORS</td>
<td>INTELLIGIBLE CROSSTALK</td>
<td>CONVOLUTIONAL</td>
</tr>
<tr>
<td>BIPHASE L PN GEN</td>
<td>ZONAL (IDEAL)</td>
<td>POWER METER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENB SIGNAL GEN</td>
<td>measured RESPONSE</td>
<td>AVERAGE METER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPULSE NOISE</td>
<td>FUNCTIONAL FORM</td>
<td>SIGNAL/LOADER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTINUOUS WAVE</td>
<td>ADAPTIVE TRANSVERSAL</td>
<td>EAVER/LOADER</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| MOD'S AND DEMOD'S                |                           |                             |                            |                          |
| AMPLITUDE                        |                           |                             |                            |                          |
| FREQUENCY                        |                           |                             |                            |                          |
| PHASE                            |                           |                             |                            |                          |
| DELTA MODULATOR                  |                           |                             |                            |                          |
| FSK                              |                           |                             |                            |                          |
| PSK                              |                           |                             |                            |                          |
| QASK, MPSK                       |                           |                             |                            |                          |
| DIGITAL ECHO MODULATOR           |                           |                             |                            |                          |
| ZF NO CROSSING DETECTOR          |                           |                             |                            |                          |

Figure 3-1
Model Listing [1]
Figure 3-2
SYSTID Operational Flow [1]
3.2 The Simulation of an FDMA Transponder

In this section a typical simulation is followed through to its completion. Figure 3-3 shows a detailed representation of an FDMA transponder. This transponder is typical of one that might be used in a digital cellular mobile phone base station. The design criterion are identical. The goal is to minimize distortion of the desired signal while also minimizing adjacent channel interference caused by intermodulation products [1]. The uplink in this system consists of eighteen 137 Mbps QPSK signals separated by 137 MHz. Even and odd channel carrier separation is maintained through the use of cross polarization. The block diagrams for the simulation are shown in Figures 3-4 and 3-5.

Figure 3-6 shows the input topological system description. The most important segments of the code can be seen directly under the "SIMULATE" heading. A QPSK source model is used to generate a time domain signal input for the multiplexers. The
Figure 3-4
Multiplexer Switch Assembly [1]
Figure 3-5
Individual Receiver and Signal Models [1]
FORTRAN code for this source model is shown in Figure 3-7. The multiplexer input signal "MUXIN" is then fed through three mux switch models to generate the inputs to the three travelling wave tube amplifiers. The upconverter, mux switch, and TWTA models are listed in Figures 3-8, 3-9, and 3-10 respectively. The QPSK demodulator model is listed in Figure 3-11.

Node save files were designated at the beginning of the listing in Figure 3-6. During the simulation data is saved at the multiplexer inputs (MUXIN), TWTA inputs (TWT1IN, TWT2IN, TWT3IN), TWTA outputs (TWT1O, TWT2O, TWT3O), mux outputs (MUX1, MUX2, MUX3), combined mux signals (DWNLNK), and receiver output (DEMOD). All of this information will now be available for post-simulation processing. These may all be represented in a variety of ways.

The postprocessor allows for bit error rate estimation. Figure 3-12 shows the transponder bit error rate as a function of the ratio of bit energy to noise density. Note that the simulation also allows one to vary the TWTA drive backoff level. Figure 3-13 shows the sensitivity of the bit error rate to the multiplexer bandwidth at two specific backoff levels.

Clearly this type of data is invaluable to the system designer. In later sections of this report it will be shown how models specific to the engineering of cellular mobile phones have been incorporated into SYSTID simulations.

3.3 Hardware Configuration

SYSTID is currently running on a MicroVAX 2 in the Electrical Engineering Building at Georgia Tech. It is set up in an absolute minimum configuration, with a single HP 150 serving as the graphic I/O terminal. Future plans call for the addition of a laser printer and several more graphics terminals. The most pressing hardware needs at the moment are for a multiport server, an Ethernet port, and a plotter. The multiport server will allow the communications faculty to conduct multiple simulations simultaneously. The current configuration allows for only a single user. The Ethernet port will provide for connection to the campus-wide Ethernet network currently under construction. This will also facilitate the creation of an undergraduate design laboratory based on the SYSTID facility. Finally, the plotter will provide for the accurate reproduction of postprocessing data obtained from
Figure 3-6
Topological System Description for FDMA Simulation [1]
MODEL: NULL > QPSK SOURCE (ISLOT) > OUTPUT

COMPLEX: OUTPUT, PILIN, PILOUT

STACK(1) ONCE

INTEGER UPSLOT, XPNDR
COMMON SINPAR, TB, NSWUPDATE, BSTMUX, BSTYPE, UPSLOT(9).

INTEGER IENTRY

DATA IENTRY/1/

CALCULATE

IF(ONCE.NE.0) GO TO 100
ONCE-ENTRY

WRITE(ISOUT,1000) IENTRY, BSTMUX

1000 FORMAT(//21X,' QPSK SOURCE ENTRY \
 11X.' B*T PRODUCT \
 0.110.4/) 

ISTIROD(7(IENTRY-1).NE)+1 

ISTOROD(IST14.25,NS)*1 

IENTRY.IENTRY+1

CONTINUE

SIMULATE

NULL > PNPULS(TB, NB, NR, ISTI) > P

NULL > PNPULS(TB, NB, NR, ISTD) > Q

PILIN=CMPLX(P, Q)

PILIN > CARTWH(3.1, 0., BSTMUX/10.4) > PILOUT

PILOUT > UP CONVERT (ISLOT) > OUTPUT

END

Figure 3-7

QPSK Signal Model Listing [1]
MODEL: \texttt{INSIG > UP CONVERT \{ISLOT\} > OUTSIG}

COMPLEX: \texttt{INSIG,OUTSIG,STACX(1) > COUNTR}

COMPLEX \texttt{SPINER(128)}

INTEGER \texttt{UPSLT}

COMMON /SIMPAR/ TB,NISP,BTXHIT,BTMUX,BTDUT,UPSLT(9),RIPPLE,ND,HR

DATA FIRST/.TRUE./

CALCULATE

\begin{verbatim}
IF (.NOT. FIRST) GO TO 100
NISP2=NISP/2
IF (NISP2.GT.128) WRITE (13OUT,1000) NISP2
FORMAT//(21X,’”UP CONVERTER ERROR”//21X,’NISP2 = ‘,I5,’ MAX SHOULD BE 256’)
IF (NISP2.GT.256) STOP

600 DPHI=2*PI/2*FLOAT(NISP2)
DO 10 I=1,NISP2
SPINER(I)=CEXP( COMPLX(0.,(I-1)*DPHI))
10  DPHI=1./(NISP2*DPHI)
WRITE (13OUT,2000) DPHI
FORMAT//(21X,’”UP CONVERTER SLOT SPACING = ‘,1PE12.4)
COUNTR=1
FIRST=.FALSE.
100 CONTINUE

SIMULATE

COUNTR=MOD(COUNT+ISLOT+NISP2-1,NISP2)+1
OUTSIG=INSIG*SPINER(COUNTR)
END
\end{verbatim}

Figure 3-8
Upconverter Model Listing [1]

MODEL: \texttt{INSIG > MUX SWITCH \{ISLOT\} > MUXOUT}

COMPLEX: \texttt{INSIG,BBSIG,FBBSIG,MUXOUT}

INTEGER \texttt{UPSLT,KMDR}

COMMON /SIMPAR/ TB,NISP,BTXHIT,BTMUX,BTDUT,UPSLT(9),RIPPLE,ND,HR

SIMULATE

\begin{verbatim}
INSIG > UP CONVERT \{-ISLOT\} > BBSIG
BBSIG > CHBESHEV (5.1,8.,BTMUX/TB,0.,1.,RIPPLE) > FBBSIG
FBBSIG > UP CONVERT \{ISLOT\} > MUXOUT
END
\end{verbatim}

Figure 3-9
Multiplexer Model Listing [1]
MODEL: ARPIN > TWTPAR (OPOINT) > AMPOUT

COMPLEX: ARPIN, AMPOUT

INTEGER UPSLOT,XPNDR

COMMON /SIRPAR/ TB,NEPB,BTXMIT,BTUX,BTDET,UPSLOT(9), $ RIPPLE,NB,NR

SIMULATE

ARPIN > RF POWER METER (10.*TB.POWER) > NULL
ARPIN > TW275H (POWER,OPOINT) > AMPOUT

--- MEASURE SIGNAL POWER
--- MEMORYLESS NONLINEARITY LIBRARY MODEL

Figure 3-10

TWTA Model Listing [1]

MODEL: RECIN > QPSK REC (ISLOT, EB) > DECOD

COMPLEX: RECIN,FILIN,FILOUT,DEMOD

INTEGER UPSLOT,XPNDR

COMMON /SIMPAR/ TB,NEPB,BTXMIT,BTUX,BTDET,UPSLOT(9), $ RIPPLE,NB,NR

DECLARE: ENO(3) = 2..6.
DECLARE: PHI(3) = 0..7.

INTEGER ENTRY

DATA ENTRY /1/

CALCULATE

IF(ONCE.NE.0) GO TO 100

ONCE=ENTRY

WRITE(1,1000) ENTRY,BTDET,ISLOT

1000 FORMAT(/2X,'** QPSK RECEIVER ENTRY *,12,**/',$ 21X,*** SLOT NUMBER = ',13/,$ 21X,*** B'T PRODUCT = ',F10.4/$

PRINTOUT HEADER INFORMATION
AND INITIALIZE PARAMETERS FOR
PN SEQUENCE REGENERATION

ISTO=MOD(IST1+[ENTRY-1],NB)+1

ISTQ=MOD(ISTI+25,NB)+1

ENTRY=ENTRY+1

CONTINUE

SIMULATE

NULL > PIWPULS(TD,HD,NR,IST1) > P
NULL > PIWPULS(TB,NB,NR,ISTQ) > Q

REC!! > UP CONVERT (-ISLOT) > FILIN
FILIN > BTWTH(5,1,TBDET/TB,.1,.1) > FILOUT

PD=REAL(FILOUT) + Q=AIMAG(FILOUT)

PD,TD,P,Q > BER EST (TB,HD/4,4,200,END,48*TD,ED,$ 6.32*TD,36*TD,48*2,PNOIS,DELAY,PHI,0) > NULL

DEMOD=CHPLX(0.,0.)

IF(ISLOT.EQ.0) DEMOD=CHPLX(0.,0.)

END

--- REGENERATE PN SEQUENCES
--- ORIGINALLY SENT

--- CONVERT TO BASEBAND
--- DETECTION FILTER

--- SIMULATION – ANALYTIC BIT
--- ERROR RATE ESTIMATOR

--- OUTPUT THE DEMODULATED
--- SIGNAL ONLY FOR THE
--- CENTER CHANNEL

Figure 3-11

QPSK Receiver Model Listing [1]
Figure 3-12

Bit Error Rate vs. $E_b/N_0$, 500 MHz Output MUX BW [1]
Figure 3-13
Performance vs. Output MUX BW [1]
the simulations.

References

4 Simulation of a Particular Mobile Radio System (NEC)

The evaluator simulator described in section 3 was used to obtain the simulated performance of the digital cellular system proposed by NEC. Because of the limited information available on the NEC system, some assumptions had to be made concerning its implementation.

A block diagram of the system under consideration, excluding the voice coder, is shown in Fig. 4.1. As shown in Fig. 4.1, the modulator consists of $\pi/4$ shifted differential QPSK with raised cosine pulse shaping. In the proposed NEC system 4 channels are time-division multiplexed in each 30 kHz channel. The transmission speed is 40 kb/s.

The channel was assumed to be characterized by Rayleigh fading. As discussed in section 2, there are two important parameters of interest, the carrier-to-noise threshold $\Gamma_{th}$ and the carrier-to-interference threshold $\Lambda_{th}$. To determine these thresholds, we require errored data streams for various carrier-to-noise ratios $\Gamma$ and several carrier-to-interference ratios $\Lambda$. These errored data streams can then be used to determine $\Gamma_{th}$ and $\Lambda_{th}$ for a particular voice coder/decoder. Each voice coder will have distinct values of $\Gamma_{th}$ and $\Lambda_{th}$.

Fig. 4.2 illustrates the technique used to generate the errored data streams. While generating the errored data streams, it is a simple procedure to determine the error statistics of the demodulated data as shown in Fig. 4.2. Fig. 4.3 shows the probability of received bit error as a function of $\Gamma_{th}$. The curve to the left represents the performance of the NEC system operating in a nonfading environment, i.e., an AWGN channel. The curve to the right shows the performance when the channel is affected by Rayleigh fading and AWGN. As is typically the case, the AWGN channel exhibits an exponential dependency of the bit error probability on $\Gamma$, while the Rayleigh fading channel exhibits an inverse linear one.

Fig. 4.4 shows the bit error probability as a function of $\Lambda$. The curve to the left shows the bit error probability for a nonfaded channel. In this case the received signal is corrupted by the presence of six nonfaded interferers. The carrier-to-noise ratio is 50 dB, which effectively means that the effect of background noise is negligible. By comparing the leftmost curves of Figs. 4.3 and 4.4, observe that the performance with nonfaded interferers is worse than the performance with an AWGN channel. This is due to the fact that the power spectral density of the AWGN is flat over the 30 kHz bandwidth, whereas the power spectral density of the interferers matches the non-flat power spectral density of the desired
signal. Once again, the rightmost curve of Fig. 4.4 shows that the bit error probability has an inverse linear dependency on $A$ for a Rayleigh fading environment.

In the next section, the errored data sequences will be used to determine $\Gamma_{th}$ and $A_{th}$ for a particular voice coder.

![Block diagram of $\frac{\pi}{4}$ shifted DQPSK simulation with co-channel interference and Rayleigh fading.](image-url)

**Figure 4.1:** Block diagram of $\frac{\pi}{4}$ shifted DQPSK simulation with co-channel interference and Rayleigh fading.

![Block diagram of post-simulation error processing.](image-url)

**Figure 4.2:** Block diagram of post-simulation error processing.
Fig. 4.3 Probability of Bit Error versus Carrier-to-Noise Ratio
Fig. 4.4 Probability of Bit Error versus Carrier-to-Interference Ratio

Rayleigh faded interferers

nonsanded interferers
5 Speech Coders

A general survey of speech coder technology suitable for application in the mobile cellular telephone environment was presented in an earlier report. [1] Here we describe briefly the speech coders used in this simulation effort to date. The coders used were real time coders which existed at Georgia Tech as the result of other studies undertaken by researchers in the Digital Signal Processing group.

First, the sub-band speech coder developed by the Digital Signal Processing Laboratory with the Atlanta Signal Processors, Inc. is described briefly here. [2] Other subband coders have been suggested by Motorola [7] and by AT&T [4].

The self-excited vocoder (SEV) coder developed for the Jet Propulsion Laboratory [3] has required some extensive modification to allow interface with the channel simulator; it will also be described very briefly below.

The SEV is a very efficient coder based on the code excited linear predictive coder (CELP) technique. The SEV has been designed to operate at 4800 bps for the mobile satellite channel; it does not give quite toll quality at this bit rate, but it is felt that a version operating at 6000 to 8000 might very well provide toll quality speech.

5.1 Sub-Band Coder

The subband coder used in the simulation is a real-time implementation of a subband coder using the TMS320C25 microprocessor chip as a basis. This digital signal processing chip allows a real time subband coder to be implemented on a single board. The implementation details have been given in the open literature. [2] In the version used in the simulations, the speech coder is constructed on a board which can be plugged into an AT compatible personal computer. Speech is captured by means of an input board and output to a filter and sound system for reconstruction.

For this application, input speech is captured in the first half of a buffer and processed to give output speech in the second half of the buffer. Data from the channel simulator implemented on the MicroVax is passed to the speech coder in the form of an errored data bit stream on floppy disk media. The speech coder in the AT personal computer combines the digital speech with the errored bit stream and reconstructs the speech signal from the
corrupted data. The personal computer then outputs the entire buffer which is clear text followed by corrupted text; the system will repeat this combination as long as desired.

The sub-band coder used here is shown in Figure 5-1 below. The tree structure used to develop the sub-band structure using a family of splitting filters and the use of conjugate quadrature mirror [5] [6] filters to eliminate distortion are notable features of the coder. The tree structure of filters first looks at a band of 4 kHz and divides it into two equal parts; then the upper part is divided into two again and only the lower half is retained. In the lower part of the initial split, the splitting continues. Finally the speech is seen to be divided into octave bands, beginning at the low end and continuing to the midpoint; the last portion is a final quarter band portion above the midpoint of the original band. A total of five bands is then coded for transmission. The filter outputs are assigned varying numbers of levels for PCM representation and then grouped for coding into bit streams. The final bit stream representation of the speech is at 16 kilobits/second. No error correction coding is provided in this coder. In the receiver the bit stream is appropriately demuxed to give the levels for each filter input and the speech is reconstructed.
5.2 Self-Excited Vocoder or SEV

The self excited vocoder is a member of the class of voice coders known as analysis-synthesis coders, which are distinguished by the method of coding the residual signal remaining after the short term linear predictor has removed all of the redundancy it can from the input speech signal. [1] The SEV forms its excitation by searching over the past of the long term predictor responses for the best representation of the residual. The CELP finds its representation of the residual by searching over a codebook of sample functions from a Gaussian random process. MPLPC and RPLPC coders form the residual representation form searching over a family of pulse signals. These systems can be illustrated as in Figure 5-2.

The version of the SEV used in the simulations here is implemented in real time using the WE-DSP32 microprocessor. The coder is implemented on a board which can be used in an AT class personal computer along with other boards to capture the speech and output the speech. The details of the implementation will be available in the final report.
of the development project for the MSAT-X 4800 bps SEV coder at the DSP Laboratory of the School of EE at Georgia Tech. This report is in final preparation at present.

The same errored data bit streams from the channel simulation were used to corrupt the transmitted bit stream for the SEV as were used for the subband coder.

References


6 Voice Coder Experiment

The voice coders described in chapter 5 were tested in conjunction with the system described in chapter 4. The accompanying tape is an actual record of the results of this test. When playing the tape, have Figs. 4.3 and 4.4 in the report at hand, as these will be referenced. Keep in mind that both of these voice coders were designed for other purposes and consequently are not well suited for the mobile telephone channel. They are both unprotected, and in the case of the 4.8 kb/s SEV, the rate is too low.

By using one's subjective option, it is possible to discern the threshold carrier-to-noise ratio and the threshold carrier-to-interference ratio. Once these thresholds have been obtained, please refer to chapter 2 for a discussion on how these threshold measurements may be used. A simple example is as follows. In a nonfaded environment, it may be claimed by listening to the voice coder experiments that the threshold carrier-to-interference ratio for the 16 kb/s SBC with six nonfaded interferers is about 11.0 dB. Suppose that the proposed NEC cellular system has \( D/R = 5 \). Then, by referring to Fig. 2.5, the area average co-channel interference outage is approximately 30%, i.e., any time acceptable communication cannot be established in 30% of the cell area. Now suppose that it is desirable to maintain a co-channel interference outage of only 20%. Then, from Fig. 2.5, \( D/R = 6 \) is required.
Security for Digital Cellular Mobile Phones

Professor Stephen B. Wicker
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
(404) 894-3129

December 5, 1989

Abstract

Private key cryptosystems such as the Data Encryption Standard offer security, speed, and flexibility of implementation for most applications. Unfortunately the problem of key distribution seriously limits their utility in a mobile environment. Public key cryptosystems provide a powerful solution to the key distribution problem, but suffer from a high level of computational complexity. The resulting delay and limitation on data rate preclude their use in most voice applications. In this paper both private and public key systems are discussed in detail. It is shown that the two can be combined in a manner that eliminates the drawbacks inherent in each approach. The resulting hybrid cryptosystem provides an ideal method for securing voice and data traffic over a digital cellular mobile phone network.
1. Introduction

There are several engineering issues that must be considered before a communication link in a cellular mobile radio network can be made secure. There are also perceptual issues that must be resolved if the consumer is to believe in the security of the system. The principal problem lies in the medium itself, for all radio transmissions are subject to interception. The specific interception threats with which the user must contend are a function of his or her particular application. The various forms of interception can be categorized as either passive or active. Passive interception consists of simple eavesdropping, while active interception involves placing spurious information onto the communication channel with the intention of fooling the recipient into accepting this information as valid. Within the context of cellular mobile telephony, passive interception threats include the following.

1.) Indiscriminate interception of simplex traffic. It is an extremely simple (though illegal) matter to purchase a radio receiver and to listen in on one-half of a mobile phone conversation.

2.) Indiscriminate interception of duplex traffic. Catching both ends of a mobile phone conversation is an extremely difficult task. Once one side of a conversation has been selected, acquiring the other side is only possible through an extensive search of several frequency bands and the subsequent identification of the appropriate voice or data signal. The effort involved essentially rules out the interceptor who is only interested in indiscriminate eavesdropping.

3.) Interception of a specific simplex transmission. The interception of a transmission from the mobile end of a specific conversation is relatively simple if the interceptor can get close enough to the mobile to identify the frequency band that has been assigned to the mobile transmitter by the base station. FAX transmissions of sales data, contracts, and other important commercial information are particularly susceptible to this threat. The interception of the base station end of a particular conversation requires an extensive frequency search and is probably beyond the means of all but the most persistent (and well financed) of adversaries.
4.) *Interception of a specific duplex transmission.* At least one extensive search of a number of frequency bands is necessary. Since this is an extremely difficult task, it can be claimed that the complexity of an urban cellular mobile radio network provides an adequate level of security for most commercial duplex applications. However, convincing the consumer of this fact may present a problem (hence the difference between actual and perceived security mentioned earlier).

5.) *Interception and processing of all transmissions within a network:* This is only of interest to those who view the governments of one or more entire countries as a possible threat. Though outrageously expensive, it is entirely possible to construct a computerized listening network whose function is to catch key words or phrases within the torrent of conversations underway in a telephone network at any given moment. If a key phrase is heard, the appropriate transmission is isolated and recorded.

Active interception threats in a cellular mobile phone network are in almost all cases some variant of the following:

1.) *Transmission of false data:* There are a variety of situations in which an interceptor may wish to emulate a particular mobile phone user and transmit false data. This would involve the duplication of appropriate control signals and considerable expense. Given the level of commercial espionage at which such activity might take place, however, it is reasonable to assume that the expense would be recovered in the form of damage to the competitor (the loss of a contract, etc.).

Both passive and active interception threats can be handled through appropriate types of voice and data encryption. In fact, all but one of the above threats can be met through the addition of some relatively simple but carefully selected cryptographic hardware (the threat of an enormously wealthy hostile government will not be countered by the work in this paper). In the next section this paper considers the impact of various aspects of mobile telephony on the selection of a cryptosystem. The next section discusses private key cryptosystems, and goes into some detail in the description of the data encryption standard (DES) and its possible use in a cellular mobile radio.
Public key cryptosystems are then introduced, and their potential utility in the commercial mobile radio environment is considered. The paper then concludes by presenting a combined private key/public key cryptosystem that is ideally suited for cellular mobile telephony.

2. Cryptosystem Requirements for Digital Cellular Mobile Radio

The land mobile radio channel is a particularly nasty environment in which to communicate. As transmitted signals propagate outward from their points of origin, they are reflected by trees, buildings, cars, and other objects [1-9,17]. These reflections form multiple delayed images of the signal that reinforce or interfere with one another at any given point in space. A moving receiver will thus find itself traveling through a series of fades and peaks that occur at a rate proportional to the receiver's rate of travel. From the point of view of the baseband digital data, the channel appears to be “bursty”, with bursts of bit errors occurring with varying frequency and duration. Unfortunately errors of any kind can be lethal to encrypted data. Most modern cryptosystems (e.g. the data encryption standard) use diffusion [14] to spread the effects of a single plaintext character throughout the entire ciphertext. Though security is greatly improved, a single ciphertext error will thus effect every letter in the plaintext, making it completely illegible.

The extreme sensitivity of encrypted data to channel errors can be handled in two ways. The first is to minimize the number of errors seen by the baseband digital data through the use of error control coding [17]. The best arrangement for encryption, error control coding, and source coding is shown in Figure 1. The source coding is placed in front of the encryption circuitry to minimize the amount of encrypted redundancy, minimizing the amount of information available to potential cryptanalysts.

The second method for dealing with the sensitivity of encrypted data to channel errors is the encryption of the data in relatively small blocks, limiting the range over which ciphertext errors can affect the plaintext. Unfortunately this presents a conflict in interest, for the encryption of small blocks limits the range of data
diffusion and reduces the security inherent in the system. A combination of error control and reduced block length is probably the best approach.

![Baseband Coding for Digital Voice/Data Transmission](image)

**Figure 1: Baseband Coding for Digital Voice/Data Transmission**

The encryption of data in small blocks is also suggested by the data frame formatting that will be used in commercial digital mobile phones. A standard for the data frame format is still under debate, but it is safe to assume that certain items will be necessary for its implementation. These items include small data frames to limit the duration of drop outs, adaptive equalization, frame synchronization, and sub-frame synchronization (for TDMA applications). A frame will look something like the one shown in Figure 2. The data sub-frames should be independently routable, limiting their length to the order of a few hundred bits. The selected cryptosystem must restrict its operations to single sub-frames, for their is no guarantee that adjacent sub-frames will be available at the same receiver.
### Frame Synchronization Sequence

The frame format for a digital cellular mobile phone is shown in Figure 2.

<table>
<thead>
<tr>
<th>AETS</th>
<th>SF SYNCH</th>
<th>DATA</th>
<th>AETS</th>
<th>SF SYNCH</th>
<th>DATA</th>
<th>AETS</th>
<th>SF SYNCH</th>
<th>DATA</th>
</tr>
</thead>
</table>

- **AETS** = adaptive equalizer training sequence
- **SF SYNCH** = sub-frame synchronization sequence

**Figure 2:** Frame Format for Digital Cellular Mobile Phone

---

### 3. Private Key Encryption: The Data Encryption Standard

Virtually all of the encryption schemes developed from the beginning of recorded history up through the late 1970's can be characterized as private key cryptosystems. In a private key system a single key is used for both encryption and decryption. Let us suppose that User A wishes to conduct a private conversation with user B. Both users must have agreed beforehand on a particular cryptosystem (e.g. simple letter substitution, DES, etc.). User A proceeds by generating a key for the selected cryptosystem and transmitting it to User B. Since anyone with access to this key can decrypt the information, the key must be protected, hence the term "private key". The key distribution function requires the use of a secure channel, as shown in Figure 3. The secure channel can be a separately encrypted line or a courier. In either case there is some risk that the key will fall into the wrong hands.

Once both users are in possession of the key User A may encrypt his information and transmit it to User B over an unsecure channel (e.g. a cellular mobile phone).
Figure 3: Private Key Cryptosystem

The security provided by a private key cryptosystem should not be a function of the cryptosystem itself. Security should depend exclusively on the particular key upon which the users have agreed [10,11,12,13,15]. The greater the number of possible keys for a given system, the greater the potential security.

Consider the case of a simple substitution cipher $C_k$ acting on the 26 letters of the plaintext alphabet $\Sigma = \{A, B, C, D, \ldots, Z\}$. The image of this transformation $C_k(\Sigma)$ is a permuted version of the original alphabet, say $C_k(\Sigma) = \{Q, R, C, \ldots, L\}$. The key to this cryptosystem identifies the particular image alphabet used in the substitution. The Caesar cipher (named after its originator, Julius Caesar [19]) restricted the image alphabets to simple shifts (e.g. $\{C, D, E, F, \ldots, A, B\}$ and $\{P, Q, R, S, \ldots, N, O\}$). There are thus 26 possible keys to Caesar's cryptosystem, each corresponding to a shifted version of the alphabet. The cryptanalysis of a piece of Caesar cipher encrypted text is a simple (and brief) matter of exhausting all possibilities.

If the image alphabets are allowed to take on all possible permutations, security is greatly increased. There are now $26! = 403291461126605635584000000$ possible keys to contend with, greatly increasing the difficulty of the task confronting the cryptanalyst.

Substitution ciphers mask information by changing the way each symbol in the plaintext is represented. Shannon referred to this approach as confusion [14]. If there is no redundancy in the plain text, this type of encryption is entirely sufficient for perfect security. If all possible patterns are valid, how can the
cryptanalyst tell when he has successfully decrypted the transmitted data? Unfortunately most written and spoken languages are extremely redundant. For example, written English text is estimated to be 67% redundant [11]. This is helpful when reading terrible handwriting, but can unfortunately lead to the breakdown of an otherwise excellent cryptosystem. Consider the simple substitution cipher that makes use of all possible permuted alphabets. Though the individual symbols have been changed, the relationships between symbols have not. Repeated patterns in the cipher text can be used to break the code without having to exhaust all 26! possible keys.

The relationships between symbols in the plain text can be masked through the use of various forms of transposition. For example, letters in the plaintext can be rearranged in the cipher text through an invertible transformation. Shannon called this approach "diffusion" [14]. Both confusion and diffusion are provided in great quantity in the Data Encryption Standard (DES).

3.1 The Data Encryption Standard

DES evolved out of the LUCIFER cipher developed by IBM in the early 1970's [18]. This cryptosystem is based on the fact that the determination of the existence of solutions for algebraic equations modulo 2 is NP-hard. Readers interested in the underlying number theory are referred to Koblitz [12], Welsh [13], and Feistel [18].

Figure 4 shows an outline of the basic DES algorithm. The algorithm begins by taking an input block of 64 plaintext bits, permuting them, splitting them into two equal portions, and placing them into two registers $L_0$ and $R_0$. The data in the registers then goes through 16 iterations of a bit selection and mapping procedure that is shown in the shaded portion of Figure 4. After the 16th iteration the contents of registers $L_{16}$ and $R_{16}$ are combined, permuted, and placed in the ciphertext output register.

---

1 An excellent example of this type of cryptanalysis can be found in "The Gold Bug", a short story by Edgar Allen Poe.
Figure 4: The Data Encryption Standard
The bit selection and mapping procedure is controlled by a series of 16 key vectors that are derived from a 64 bit key. 8 of the 64 key bits are redundant, leaving an effective key length of 56.

3.2 The Use of DES in a Cellular Mobile Phone System

From the user's perspective the principle points of interest are the implementation costs, speed, and the degree of security provided. The DES algorithm is commercially available at reasonable cost in both software and hardware. VLSI implementations of the algorithm are available for applications operating at T1 rates (1.544 Mbps), well above that needed for commercial mobile radio applications. The DES algorithm operates on 64 bit blocks, fitting in nicely with the constraints of the frame structure suggested in Figure 2. The data subfields can be encrypted individually so long as their length in bits is a multiple of 64.

The security provided by the DES algorithm can be further enhanced through cipher block chaining [11]. For example, suppose that the data sub-field consists of 384 bits. The simplest approach is to encrypt the sub-field as 6 distinct blocks of length 64. In cipher block chaining each successive data block is exclusive OR'd with the encrypted version of the previous block prior to encryption (see Figure 5). Cipher block chaining reduces the possibility of repeated patterns within the ciphertext, reducing the amount of information available to the eavesdropper. For reasons discussed in the previous section, the extent of the chaining should be limited to the boundaries of the data subfields.
The use of DES in cellular mobile phones is thus viable from both a functional and an economic point of view. The final consideration, the degree of security provided, has been a source of lively debate since its adoption as a commercial standard by the National Bureau of Standards. The most pointed criticism has come from the respected team of Diffie and Hellman [15], the inventors of public key cryptography (see the next section). Their principal argument is that the effective length of the key is too short. 56 effective bits provides approximately $7.2 \times 10^{16}$ different keys, allowing for the possibility of cryptanalysis by exhaustive search (trying all possible keys until the right one is found). Diffie and Hellman claim that a dedicated computer capable of breaking a DES encrypted message every 12 hours can be built for roughly $20$ million. The National Bureau of Standards strongly disagrees, claiming that the construction of such a machine is well beyond the current state of technology. NBS feels that the time required for a key search is somewhere between 91 and 2000 years given the current state of computer technology.

There is an additional objection to DES that is much more subtle. It has been noted that the S-boxes (see Figure 4) in the DES algorithm are not completely random. Critics believe that there may be a “trap door” within these boxes that allows for the decryption of encrypted data by those holding the secret to the structure of the boxes. It should be noted, however, that despite extensive effort no one has yet been able to discern a useful pattern within the S-boxes. In fact,
in the first 10 years of its existence there has yet to be a single valid claim that a DES encrypted message has been broken.

Based on the above it is safe to assume that a DES based cryptosystem for cellular mobile phones can counter most of the threats discussed in the introduction. The question of DES security against simplex passive interception is purely a question of how long it will take the interceptor to execute a key search (and how much money he or she is willing to spend). If Diffie and Hellman are correct, the interceptor will have to spend $20 million constructing an enormously complicated machine that will take 12 hours to break the code. Even if the extremely optimistic estimate of 12 hours is accepted, the need for security in some applications will no longer be necessary. If the National Bureau of Standards is correct, all parties involved will have long since ceased to care before a single key is broken. In either case an industrial spy is much more likely to use his $20 million in more profitable pursuits.

Interception of duplex transmissions is completely unrealistic because the search for the second half of the conversation cannot be executed without deciphering hundreds or thousands of different keys to identify the correct frequency band. Even if the key search takes the minimum estimate of 12 hours, the conversation will have concluded long before a single key can be obtained.

DES also provides security against active interception by requiring the would-be interceptor to know the correct key prior to the introduction of false information into the network. If an improperly encrypted message is introduced, it will be obliterated when the receiver attempts to decrypt it using the correct key.

3.3 Problems with the DES Approach

The primary drawback to using DES in a cellular mobile phone network is that all parties to any given secure conversation must have agreed on the key prior to the initiation of the conversation. This inflexibility can be a serious handicap in a mobile environment. Consider the following. A distinct DES key must be generated for all possible combinations of users. Copies of all appropriate keys
must be in the possession of the mobile operator prior to the inception of secure communication. All keys must therefore be recorded electronically within the mobile phone or written down in a secret directory. In both cases the generation, secure distribution, and updating of keys presents a complicated logistical problem. In both cases as well the probability that keys can be unknowingly compromised is intolerably high. The impact of this key distribution problem in previous applications has led to the development of public key cryptography, which is discussed in the next section.

4. Public Key Cryptography

Public key cryptography was first suggested by Diffie and Hellman in 1976 [16]. It is based on distinct encryption and decryption keys that have been generated in such a manner that one key cannot be derived from the other without an unreasonable amount of computational effort. Figure 6 shows how such a system can be used for data security. User B generates both keys and makes the encryption key public. The decryption key is kept secret and remains known only to User B. Any user wishing to communicate with User B need only obtain User B's encryption key from a public directory. Since only User B is capable of decrypting the message, the encrypted data can be transmitted over an unsecure channel. A comparison of Figure 6 to Figure 3 shows the clear advantage of public key cryptography: there is no need for a secure channel.

The encryption and decryption algorithms are based on the selection of a one-way trap-door function. A one-way function is a function that is easy to perform in one direction, but extremely difficult to perform in the other. Exponentiation and the taking of logarithms combine to form the typical example. A one-way trap-door function is a one-way function whose difficult direction is made significantly easier if certain side information is available. In the cryptosystem in Figure 6 the easy direction of the function is used for encryption. The difficult direction corresponds to decryption without the decryption key. The key is the side information that allows the user to access the trap door.
4.1 The Diffie-Hellman Public Key System

Diffie and Hellman suggested the use of exponentiation and the taking of logarithms in a finite field as the basis for a public key system. A finite field GF(p) is selected, where p is a prime number at least 100 bits in length. For this field, a primitive element \( \alpha \) is selected such that each of the elements in GF(p) can be represented as a distinct power of \( \alpha \). Each user \( U \) in the public key system then selects an integer \( m_U \) and computes \( e_U = \alpha^{m_U} \). The size of the finite field \( p \), the primitive element \( \alpha \), and the value of \( e_U \) for each user are all made public. The exponent \( m_U \) is kept secret by each user.

Suppose now that user A wishes to communicate with user B. User A first locates User B's encryption key \( e_B \) in the public key directory. He then computes the key

\[
k_{A,B} = e_B^{m_A} = \alpha^{m_A m_B}
\]  

---

\(^2\)A finite field of arbitrary order \( q = p^e \), \( p \) a prime and \( e \) a positive integer can be used in the Diffie-Hellman approach [11].
and uses it to encrypt the text. User B can arrive at the exact same key by raising User A's public encryption key to User B's secret exponent $m_B$.

$$k_{A,B} = e_A^{m_B} = a^{m_Bm_A} = a^{m_Am_B}$$

User A and User B are thus able to arrive at the same key without exchanging any secret information. No one else can arrive at the same key without access to the secret information held by either User A or User B.

The only known way of recovering the exponent $m_U$ from the encryption key $e_U$ is through the taking of discrete logarithms. If $p$ is 100 bits or more in length, then the operation is expected to take many years using a supercomputer. This system may thus be considered secure for all of the cellular mobile telephone security threats discussed earlier.

4.2 The Rivest-Shamir-Adelman (RSA) Public Key System

The RSA public key system involves operations similar to the Diffie-Hellman algorithm, but uses the factoring of large composite integers in the definition of the requisite one-way trap door function. The following description of the algorithm does not include proofs of the various theorems involved. The interested reader is referred to van Tilborg [11] and Koblitz [12].

The RSA system is based on an algorithm developed by Euclid and a function and theorem developed by Euler. Euclid's algorithm is a fast method for finding the greatest common divisor for a pair of integers $a$ and $b$. In the execution of this algorithm, two integers $s$ and $t$ are generated such that the greatest common divisor can be expressed as $\gcd(a,b) = as + bt$. Euler's totient $\phi$ function determines the number of integers relatively prime to a given integer. It can be computed as follows.
\[ \phi(n) = \prod_{p|n} (1 - \frac{1}{p}) \]

Euler's Theorem relates the greatest common divisor and the totient \( \phi \) function in the following manner.

*If the greatest common divisor of two integers \( a \) and \( b \) is 1, then \( a^{\phi(b)} \equiv 1 \) modulo \( b \).*

The RSA system uses these three number theoretic concepts to secure communication in the following manner. Each user \( U \) selects a pair of distinct primes \( p_U \) and \( q_U \), each over 100 bits in length, and computes their product \( n_U \). Equation (3) is then used to compute \( \phi(n_U) = (p_U - 1)(q_U - 1) \). The users then select an integer \( e_U \) such that \( 1 < e_U < \phi(n_U) \) and \( \gcd(e_U, \phi(n_U)) = 1 \). Euclid’s algorithm is used to compute \( d_U \) such that \( e_U \cdot d_U \equiv 1 \mod \phi(n_U) \), where \( 1 < d_U < \phi(n_U) \). The integers \( e_U \) and \( n_U \) are made public, while \( d_U \) is kept secret.

If user \( A \) wishes to communicate with User \( B \), the first step is to look up the values of \( e_B \) and \( n_B \) in the public key directory. User \( A \) then formats his message so that it consists of integers \( m \) in the range \( 0 < m < n_B \). The encryption process consists of raising each integer to the power \( e_B \) modulo \( n_B \):

\[ c = m^{e_B} \mod n_B. \]  

(4)

User \( B \) can decrypt the message by computing the following:

\[ c^{d_B} \equiv m^{e_Bd_B} \mod n_B \]

\[ \equiv m^{[1 + k\phi(n_B)]} \mod n_B \]

\[ \equiv m \mod n_B \]

\[ = m \]  

(5)

Anyone wishing to decrypt the ciphertext must have access to \( d_U \). To compute \( d_U \) one must know \( \phi(n_U) \), which in turn requires knowledge of the factorization of
nu. The security of the RSA system is thus predicated on the difficulty involved in factoring large composite numbers. Several factoring algorithms are known, but all would take several years of effort with a supercomputer to factor the product of two distinct prime numbers, each exceeding 100 bits in length. It is thus safe to conclude that given the current state of technology and number theory, the RSA system can provide adequate security against the aforementioned threats to cellular mobile communications.

4.3 The Use of Public Key Cryptography in a Cellular Mobile Phone System

The RSA system has proven somewhat easier to implement than the Diffie-Hellman approach and now enjoys extensive use in commercial applications. For the rest of this section the primary emphasis is placed on the RSA system.

There are four basic functions that must be performed by an RSA security system within a digital cellular mobile phone network:

1.) Computation of nu, eu, and du: The Solovay and Strassen or Cohen and Lenstra primality tests [11,12] can be used in conjunction with a random number generator to obtain two distinct 100 bit numbers of almost certain primality in a few moments computing time. nu, eu, and du can then be computed as noted earlier. The generation and verification of all numbers can be performed in a few minutes on a standard office personal computer. Once the keys are generated, the encryption key can be phoned or mailed in to a central public key distribution center. Dedicated circuitry for key generation can be placed inside the user's phone, but the relative infrequency with which keys need to be changed makes this unnecessary.

2.) Publication of encryption keys: The simplest approach is to publish the keys in a phonebook and allow the user to enter the information manually through his or her handset. Electronic accessing of a central database through the use of the signaling channel provides a faster but much more expensive alternative.

3.) Data Formatting: The digital data or compressed voice must be formatted to appear as a stream of integers mi with values in the range 0 < mi < nu. This
can be accomplished quite simply, though with some increased redundancy in
the transmitted data, by taking \( \lceil \log_2(nu) \rceil \) bits of data at a time and mapping
them onto the integer number line in an appropriate fashion.

4.) Encryption and decryption: The actual encryption and decryption
computations are rather difficult due to the size of the numbers involved. A
dedicated digital signal processing chip will probably provide the best solution.
The use of a software driven generic microprocessor will almost certainly be too
slow.

Public key cryptosystems eliminate the key distribution problem posed by the
private key approach. No one need know how to decrypt a user's message but
that user himself. Unfortunately there are two major drawbacks to the public key
approach that preclude their use in many applications. Both of the public key
systems discussed in this section are significantly more computationally
intensive than the DES private key algorithm. As a result, their operational
speed is much more restricted; current products have a maximum data
throughput rate on the order of ten Kilobits per second. Delay due to data
formatting requirements is also a significant problem. Given the current state of
technology it would be an extremely difficult task to use public key encryption to
secure a voice channel in a manner that would be tolerated by the consumer
and still provide a high level of security.

5. The Solution: A Combined Approach

Private key cryptography offers the desired speed and simplicity while public
key cryptography offers a secure key distribution system. The best approach to
securing a digital cellular mobile phone network lies in combining the two to
obtain the best of both worlds while avoiding the worst of either. Figure 7 shows
how this can be done.
Suppose that User A wishes to communicate with User B. The following sequence of events takes place:

1.) User A generates a random 56 bit vector and converts it into a 64 bit DES key by adding the appropriate redundancy bits.

2.) User A obtains user B's public encryption key and uses it to encrypt the DES key. This key is then sent to User B, who recovers it using his secret decryption key.

3.) Both users initialize their DES encryption/decryption circuitry using the DES key generated by User A. Secure communication can now commence.

This combined approach offers a fast and efficient means of providing secure voice and data communication over a digital cellular mobile phone network.
6. References


Downlink Outage Predictions in Cellular Radio Systems*

Gordon L. Stüber and Lih-Bor Yiin
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
(404) 894-2923
December 18, 1989

Abstract

This paper presents a general model for predicting downlink outage due to co-channel interference in cellular radio systems. The model accounts for the effects of path loss, shadowing, and Rayleigh fading. The outage predictions themselves are obtained by computer simulation. Various outage control techniques are considered, including cell sectoring, adaptive transmitter power control, and hand-offs. The outage predictions are applied to a digital cellular system that uses QDPSK signaling with Reed-Solomon coding.

*This research was supported by BellSouth Enterprises Inc.
I Introduction

Cellular mobile radio systems rely upon frequency reuse, where the same set of carrier frequencies is assigned to spatially separated cells [1,6,7,8]. Two of the most important performance measures in a cellular system are outage and spectral efficiency. Outage is defined here as the area averaged probability that a mobile cannot initiate a call with an acceptable transmission quality. Spectral efficiency is the number of channels per unit bandwidth per unit area. Outage and spectral efficiency are closely related because an increase in spectral efficiency can always be gained at the expense of an increase in outage.

Once a mobile has gained access to a channel two of the primary causes of outage in a cellular mobile radio system are thermal noise (thermal noise outage), and co-channel interference (co-channel interference outage) [1]. Thermal noise outage is dominant when large cell sizes are used, whereas co-channel interference outage is dominant when small cell sizes are used. From our perspective, the performance of a digital cellular system depends roughly upon the following quantities; the number of cells/cluster $N$, the probability that a channel is active $p$, the average carrier-to-noise ratio $\Theta$, the instantaneous carrier-to-noise ratio $\theta$, the average carrier-to-interference ratio $\Lambda$, and the instantaneous carrier-to-interference ratio $\lambda$. For rapidly moving vehicles $\Theta$ and $\Lambda$ are important, whereas for stationary vehicles $\theta$ and $\lambda$ are important.

Digital mobile radio systems commonly exhibit a thresholding effect where, for rapidly moving vehicles, the transmission quality will be acceptable provided that $\Theta > \Theta_{th}$ and $\Lambda > \Lambda_{th}$. For stationary vehicles these conditions become $\theta > \theta_{th}$ and $\lambda > \lambda_{th}$. The thresholds $\Theta_{th}$, $\Lambda_{th}$, $\theta_{th}$, and $\lambda_{th}$ depend upon the particular digital communication system, the measure of transmission quality, and the channel characteristics. For voice communication they even depend on the listener.

Suppose that a mobile has accessed a channel. Then in the case of rapidly
moving vehicles, path loss and shadowing determine the outage probability, $P_O$, once $\Theta_{th}$ and $\Lambda_{th}$ have been specified. Likewise, for stationary vehicles, path loss, shadowing, and fading determine the outage probability for a given $\theta_{th}$ and $\lambda_{th}$.

Since our interest will be directed to high capacity (densely populated) cellular systems, the effect of thermal noise outage will be neglected. Therefore, an outage occurs whenever $\lambda < \lambda_{th}$ or $\Lambda < \Lambda_{th}$, whichever is appropriate. In our analysis, the outage probability is computed for the downlink only. Previous studies have tended to suggest that the downlink and uplink outage probabilities are almost identical, although the channels are not exactly reciprocal [2,3]. It is unknown if a similar statement can be made for our analysis.

The remainder of the paper is organized as follows. In section II, we state the basic modeling assumptions used throughout the rest of the paper and demonstrate how the outage is computed for rapidly moving and stationary vehicles. Sections III, IV, and V investigate the use of cell sectoring, adaptive transmitter power control, and hand-offs, respectively. These results are then applied to a digital cellular system that is presented in section VI. Section VII discusses the use of Reed-Solomon codes for error correction. Results and discussion are presented in Section VIII.

II Outage Predictions

Similar to the model used in [11], the outage predictions in this paper are based upon the following assumptions:

1. The cellular layout is described by a uniform hexagonal grid. The cells are of radius $R$ and the minimum distance between cells using a common set of channel frequencies is $D$. The co-channel reuse factor is defined as the ratio $D/R$. For hexagonal cells, the number of cells per cluster $N$ is related to the
co-channel reuse factor by \( N = (D/R)^2/3 \).

2. Each channel is assumed to be independently active with probability \( p \), reflecting the busyness of the cellular system. If \( L \) is the total number of channels associated with a base-station, then the number of active channels has an \((L, p)\) binomial distribution.

3. Without cell sectoring, there are a maximum of six interfering base-stations equidistant from the base-station of the cell containing the intended receiver (mobile). The more distant base-stations are neglected in deference to the typically dominant effect of the first tier of interfering base-stations. This approximation has been shown to be very accurate [2]. With 120° cell sectoring, the maximum number of interfering base-stations is reduced from six to two.

4. The interfering signals are mutually noncoherent so that their power spectral densities add.

5. A mobile can be anywhere within a cell with equal probability. The probability density function of the mobile location, in cylindrical co-ordinates, is \( f(r, \theta) = r/\pi R^2 \), \( 0 \leq \theta \leq 2\pi \), \( 0 \leq r \leq R \), where \( R \) is the cell radius.

6. A common path loss model is used to describe the signal propagation from all base-stations. All base-stations have a common antenna height and antenna gain, although these can be varied.

7. The signals from the desired base-station and each of the interfering base-stations are affected by log-normal shadowing with independent and identical statistics. This will tend to make the results somewhat pessimistic, because the shadowing will probably be correlated and there is evidence suggesting that independent shadowing is the worst case situation [10,3].
8. The signals from the desired base-station and each of the interfering base-stations are affected by independent Rayleigh fading with identical statistics.

9. The shadowing is spatially uncorrelated. In reality, the shadowing will be spatially correlated. However, obtaining typical spatial correlation functions is quite difficult. Therefore, the outage probability at a particular spatial point can be interpreted as the outage probability averaged over a very large ensemble of cells.

Consider a mobile that is at distance $r$ from a central base-station, with probability $1 - p^L$ the mobile will gain access to a channel. Let the received field strength averaged over the shadowing and fading be denoted by $\Gamma_0(r)$.\footnote{The subscript 0 indicates reference to the central base-station.} Typically,

$$\Gamma_0(r) = Ar^{-\alpha},$$  \hspace{1cm} (1)

where $A$ is a constant and $\alpha$ is called the path loss exponent. For free space $\alpha = 2$, while for an urban environment $\alpha \approx 4$. If the background noise is neglected in deference to the dominant effect of co-channel interference, then the value of $A$ is irrelevant.

When shadowing is present the received field strength at distance $r$, denoted by $\Gamma_0(r)$, is log-normally distributed with probability density function

$$Pr_{\Gamma_0(r)}(x) = \exp \left\{ \frac{(x - \Gamma_0(r))^2}{2\sigma^2} \right\},$$  \hspace{1cm} (2)

where all units are in dBs. The standard deviation $\sigma$ ranges from 5 to 12 dB, with $\sigma = 8$ dB being a typical value. $\sigma$ has been observed to be independent of the length of the radio link [6] and this is assumed here also.

Finally, when Rayleigh fading is present the instantaneous received field strength at distance $r$, denoted by $\gamma_0(r)$, is exponentially distributed with mean
Now suppose that the mobile is at distance \( r \) from the central base-station, and at distances \( \{d_k, k = 1, 2, \ldots, K\} \) from the first tier of \( K \) potential interfering base-stations. With 120° cell sectoring \( K = 2 \) and without cell sectoring \( K = 6 \). The received CIR (average or instantaneous) depends not only upon how many interferers are active, but also upon which interferers are active. Let \( x_{0j} = (x_{j1}, x_{j2}, \ldots, x_{jK}) \) be a binary \( \{0,1\} \) vector representing the \( j^{th} \) interference pattern; \( x_{jk} = 1 \) if the \( k^{th} \) interferer is active in the \( j^{th} \) interference pattern, and \( x_{jk} = 0 \) otherwise. There are \( 2^K \) interference patterns and the probability of the \( j^{th} \) interference pattern is

\[
P \{x_{0j}\} = \prod_{k=1}^{K} p_{j,k}^0 (1 - p)^{1 - x_{jk}^0}.
\]

For the \( j^{th} \) interference pattern, the averaged received CIR is

\[
\Lambda(x_{0j})_{dB} = \Gamma_0(r)_{dB} - \left( \sum_{k=1}^{K} x_{jk}^0 \Gamma_0(d_k) \right)_{dB}.
\]

Likewise, the instantaneous received CIR is

\[
\lambda(x_{0j})_{dB} = \gamma_0(r)_{dB} - \left( \sum_{k=1}^{K} x_{jk}^0 \gamma_0(d_k) \right)_{dB}.
\]

II-A Rapidly Moving Vehicles

The probability of outage for a given mobile location \( (r, \theta) \), averaged over all \( 2^K \) interference patterns, is

\[
P_O(r, \theta) = p^L + (1 - p^L) \sum_{j=1}^{2^K} P \{\Lambda(x_{0j}) < \Lambda_{th}\} \prod_{k=1}^{K} p_{j,k}^0 (1 - p)^{1 - x_{jk}^0},
\]

where \( p^L \) is the probability that the central base-station does not have an available channel. Averaging \( P_O(r, \theta) \) over the random position of the mobile results in the
area averaged outage probability

\[ P_o = \int_0^R \int_0^{2\pi} P_o(r,\theta) f(r,\theta) r d\theta \]

\[ = \int_0^R \int_0^{2\pi} P_o(r,\theta) \frac{r}{\pi R^2} dr d\theta . \tag{8} \]

As an alternative to performing the integration in (8), the area averaged outage probability can also be computed by generating large sets of random uniformly distributed mobile positions and forming the empirical average

\[ P_o = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M P_o(r_i,\theta_i) . \tag{9} \]

Fig. 1 plots the probability of outage \( P_o \) against the threshold CIR \( \Lambda_{th} \) for an \( N = 7 \) cell reuse pattern and various busyness probabilities \( p \). A value of \( L = 135 \) channels/cell sector was chosen. The reason is that there are 999 two-way channels in the North American cellular system. Of these 42 are control channels leaving 957 channels for voice communication. These are assumed to be equally divided among the 7 cells in a cluster. This results in a minimum of 135 channels/cell.

Note that as \( p \rightarrow 1 \) the system becomes saturated so that no new channels are available. For a mobile that has gained access to a channel, the probability of outage in this case is

\[ \hat{P}_o(r,\theta) = P\{\Lambda(x_{0j}) < \Lambda_{th}\} , \tag{10} \]

where \( x_{0j} = (1, 1, \cdots, 1) \). The area averaged outage probability is

\[ \hat{P}_o = \int_0^R \int_0^{2\pi} \hat{P}_o(r,\theta) \frac{r}{\pi R^2} dr d\theta = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=1}^M \hat{P}_o(r_i,\theta_i) . \tag{11} \]

This is also shown in Fig. 1.

II-B Stationary Vehicles

Once again, suppose that a mobile is at distance \( r \) from the central base-station, and at distances \( \{d_k, k = 1, 2, \cdots, K\} \) from \( K \) interfering base-stations. If the
mobile has gained access to a channel, then from (6)

$$\lambda(x_{0j}) = \frac{\gamma_0(r)}{\sum_{k=1}^{K} x_{jk} \gamma_0(d_k)} .$$

(12)

To compute the probability density of $\lambda(x_{0j})$, first define $I(x_{0j}) := \sum_{k=1}^{K} x_{jk} \gamma_0(d_k)$.

If the $k^{th}$ interferer is the only one active, then

$$p_{I(x_{0j})}(y) = \frac{1}{\Gamma_0(d_k)} \exp\left\{-\frac{y}{\Gamma_0(d_k)}\right\} , \quad y \geq 0 .$$

(13)

If more than one interferer is active, then from [5, (7.5.26)]

$$p_{I(x_{0j})}(y) = \sum_{k=1}^{K} \frac{x_{0j}^{0} \pi_k}{\Gamma_0(d_k)} \exp\left\{-\frac{y}{\Gamma_0(d_k)}\right\} , \quad y \geq 0 ,$$

(14)

where

$$\pi_k = \prod_{i=1}^{K} \left(\frac{\Gamma_0(d_i)}{\Gamma_0(d_i) - \Gamma_0(d_k)}\right)^{x_{0j}^{0}} .$$

(15)

The density of $\lambda(x_{0j})$ is [9]

$$p_{\lambda(x_{0j})}(x) = \int_{0}^{\infty} p_{\gamma_0(r)}(xy)p_{I(x_{0j})}(y)dy$$

$$= \sum_{k=1}^{K} x_{jk}^{0} \pi_k \left(\frac{\Gamma_0(r)}{\Gamma_0(d_k)}\right)^{x_{0j}^{0}} \frac{1}{\Gamma_0(d_k) - \Gamma_0(r)} , \quad x \geq 0 .$$

(16)

The cumulative distribution function of $\lambda(x_{0j})$ is

$$F_{\lambda(x_{0j})}(x) = \sum_{k=1}^{K} x_{jk}^{0} \pi_k \left(1 - \frac{1}{1 + \frac{\Gamma_0(d_k)}{\Gamma_0(r)}}\right) , \quad x \geq 0 .$$

(17)

Hence, the outage probability as a function of $\Gamma_0 = \{\Gamma_0(r), \Gamma_0(d_k), k = 1, \ldots, K\}$ is

$$P_o(r, \theta, \Gamma_0) = p^L + (1 - p^L) \sum_{j=1}^{2^K} \prod_{k=1}^{K} p^{x_{jk}^{0}}(1 - p)^{1 - x_{jk}^{0}}$$

$$= p^L + (1 - p^L) \sum_{j=1}^{2^K} F_{\lambda(x_{0j})}(\lambda_{th}) \prod_{k=1}^{K} p^{x_{jk}^{0}}(1 - p)^{1 - x_{jk}^{0}} .$$

(18)

Averaging over the joint density of $\Gamma_0$ gives

$$P_o(r, \theta) = \int_{0}^{\infty} P_o(r, \theta, \Gamma_0)p(\Gamma_0)d\Gamma_0 .$$

(19)
on the value of $p$. Cell sectoring offers a similar improvement in $\lambda_{th}$ for stationary vehicles.

IV Power Control

Power control is a technique commonly used in cellular systems to prevent excessive hand-offs and to reduce co-channel interference. Power control may be implemented in several ways. One approach is to allow the base-stations and mobiles to adjust their transmitted power level as a function of the average or instantaneous received CIR. Another approach is to adjust the transmitted power level as a function of the length of the radio link between a mobile and a base-station.\footnote{This power control technique may be useful for systems where vehicle location is desired.} Both of these approaches are considered below.

IV-A CIR Driven Power Control

Suppose that the base-stations and mobiles adjust their transmitted power level as a function of $A(x_{0i})$ for rapidly moving vehicles, or $\lambda(x_{0j})$ for stationary vehicles. It is important to note that $A(x_{0j})$ is still given by (5) but, because power control is used, the $\{\Gamma_0(d_k), k = 1, \ldots, K\}$ are no longer log-normally distributed as in (2). Likewise, $\lambda(x_{0j})$ is still given by (6), but the $\{\gamma(d_k), k = 1, \ldots, K\}$ are no longer exponentially distributed as in (3). As discussed below these distributions have an extremely complex form.

With CIR driven power control, each base-station and each mobile has a dynamic range of $\pm \beta$ dB in its transmitted power level. In the case of rapidly moving vehicles, the average received CIR after power control and as a function of
Finally, the area averaged outage probability is given by (8) or (9). Fig. 2 plots the outage probability against the instantaneous threshold CIR $\lambda_{th}$ for $N = 7$ and various values of $p$. If $p = 1$ the probability of outage given that a mobile has gained access to a channel is

$$
\hat{P}_O(r, \theta) = \int_0^{\infty} \hat{P}_O(r, \theta, \Gamma_0)p(\Gamma_0) d\Gamma_0 ,
$$

where

$$
\hat{P}_O(r, \theta, \Gamma_0) = F_{\lambda(x_{0j})}(\lambda_{th}) ,
$$

and $x_{0j} = (1, 1, \ldots, 1)$. The area averaged outage is given by (11).

As evident from Figs. 1 and 2, an acceptable outage probability requires a transmission system having thresholds $\lambda_{th}$ and $\lambda_{th}$ that are too small to be practical. Methods are now discussed to mitigate the effect of co-channel interference (reduce outage or increase spectral efficiency).

III Cell Sectoring

Co-channel interference can be reduced significantly by using cell sectoring, where each cell is divided into multiple sectors by using directional antennae. Each sector employs a different set of channel frequencies. Here we consider $120^\circ$ cell sectors. Because of the use of directional antennae, the maximum number of principal interferers is reduced from $K = 6$ to $K = 2$ [8]. Furthermore, the effective co-channel reuse factor also increases [8]. The combination of these effects leads to a significant reduction in the outage probability.

Once again, the outage characteristics depend on the vehicle velocity. The effect of using cell sectoring is shown in Fig. 3 and 4, for rapidly moving and stationary vehicles, respectively. By comparing Figs. 1 and 3, observe that for a fixed $P_O$, $\lambda_{th}$ is about 5 to 10 dB larger with cell sectoring than without, depending
where all units are in dB. An outage occurs when \( \Lambda(x_{0j}) - \Lambda_{th} < -\beta \). Likewise, in the case of stationary vehicles the instantaneous CIR after power control and as a function of \( x_{0j} \) is

\[
\Lambda_F(x_{0j}) = \begin{cases} 
\Lambda(x_{0j}) - \beta, & \Lambda(x_{0j}) - \Lambda_{th} > \beta \\
\Lambda_{th}, & |\Lambda(x_{0j}) - \Lambda_{th}| \leq \beta \\
\Lambda(x_{0j}) + \beta, & \Lambda(x_{0j}) - \Lambda_{th} < -\beta
\end{cases}
\]

(22)

\[
\lambda_F(x_{0j}) = \begin{cases} 
\lambda(x_{0j}) - \beta, & \lambda(x_{0j}) - \lambda_{th} > \beta \\
\lambda_{th}, & |\lambda(x_{0j}) - \lambda_{th}| \leq \beta \\
\lambda(x_{0j}) + \beta, & \lambda(x_{0j}) - \lambda_{th} < -\beta
\end{cases}
\]

(23)

(all units in dB) and an outage occurs when \( \lambda(x_{0j}) - \lambda_{th} < -\beta \).

When computing the outage probability with power control, it must be assumed that the interfering base-stations are also using power control. This presents a problem because their interferers are also using power control and so on. Therefore, an approximation is required if any meaningful results are to be obtained. Consider a first-order approximation where the central base-station uses power control, but the interfering base-stations do not use power control. The resulting outage probabilities are plotted in Figs. 5 and 6 for rapidly moving and stationary vehicles, respectively, for \( \beta = 20 \text{ dB} \) and an \( N = 7 \) cell reuse pattern with 120° cell sectors. In either case, the probability of outage is averaged over all interference patterns, shadowing, and position of the intended mobile.

Now consider a second-order approximation where the central base-station uses power control, the interfering base-stations use power control, but their interfering base-stations do not use power control. In this case, the probability of outage must be averaged over all interference patterns, shadowing, and positions of the intended mobile and the mobiles receiving transmissions from the interfering base stations. The outage probabilities are plotted in Figs. 7 and 8 for rapidly
moving and stationary vehicles, respectively. Finally, a third-order approximation may be used, where the desired base-station and the first two levels of interfering base-stations use power control, but higher levels of interfering base-stations do not. The results are plotted in Figs. 9 and 10 for rapidly moving and stationary vehicles, respectively. Higher-order approximations are not computationally feasible.

Notice that there is a significant difference between the first and second order approximations, especially for stationary vehicles with small $p$. However, there is not much difference (only 2-3 dB for rapidly moving vehicles and 1-2 dB for stationary vehicles) between the second and third order approximations.

**IV-B Distance Driven Power Control**

With distance driven power control, the base-stations and mobiles adjust their transmitted power level as a function of the length of the radio link and have perfect knowledge of the path loss exponent $\alpha$. The basic concept is illustrated in Fig. 11. This diagram plots the average received field strength (1) as a function of the normalized length of the radio link $\delta$. Here the base-stations and mobiles have a dynamic range of $(-0$ and $+\beta$ dB) in their transmitted power levels. If the normalized length of the radio link exceeds $\delta_I$, then power control will be used to compensate for the path loss so that the average received field strength will be constant up to the normalized distance $\delta_F = \delta_I 10^{\beta/(10\alpha)}$. Thereafter, it will again decay with a path loss exponent $\alpha$. To use the full dynamic range capability, $\delta_I$ will be chosen so that $\alpha_F < 1$, and hence the allowable range of $\delta_I$ is $0 < \delta_I \leq 10^{-\beta/(10\alpha)}$.

Once again, when computing the outage with power control, it must be assumed that the interfering base-stations are also using power control. Therefore, the probability of outage must be averaged over the random position of the intended mobile, and the random positions of the mobiles receiving transmissions from the interfering base-stations. One analytical advantage that results from us-
ing the distance driven power control algorithm is that the distributions for \( \Gamma_0(r) \), \( \{\Gamma_0(d_k), k = 1, \cdots, K\} \), and \( \gamma_0(r) \), \( \{\gamma_0(d_k), k = 1, \cdots, K\} \) are still given by (5) and (6), respectively. The reason is that distance driven power control affects only \( \Gamma_0(r) \) and \( \{\Gamma_0(d_k), k = 1, \cdots, K\} \) in (1).

For a fixed dynamic range \((-\delta and + \beta)\) and outage probability \( P_o \), there is a value of \( \delta_f \) that maximizes \( \Lambda_{th} \) or \( \lambda_{th} \). For \( \beta = 40 \text{ dB} \) and an \( N = 7 \) cell reuse pattern with 120° cell sectors, \( \delta_f = 0.0818 \) is optimal for all \( \Lambda_{th} \) and \( \lambda_{th} \). This value of \( \delta_f \) results in \( \delta_F = 1 \). The corresponding optimal outage probabilities are plotted in Figs. 12 and 13 for rapidly moving and stationary vehicles, respectively. We have observed that very little additional improvement can be obtained in the average outage probability by increasing \( \beta \) beyond 20 dB.

Observe that for rapidly moving vehicles there is only 2-3 dB difference between CIR driven and distance driven power control. This reflects the notion that, for a reasonably large \( D/R \), variations in the average CIR are largely due to variations in the distance between a mobile and its base-station, rather than variations in the distances between a mobile and its interfering base-stations. Therefore, the outage characteristics with distance driven power control are not only less time consuming to compute, but are also representative of those with CIR driven power control.

For stationary vehicles CIR driven power control performs much better than distance driven power control. The reason is that CIR driven power control compensates for fades, whereas distance driven power control does not.

V Hand-offs

It would seem reasonable that a further outage reduction can be obtained by using hand-offs. For our purpose, it is assumed that distance driven power control is

\(^3\text{Hand-offs are also essential to allow roaming.}\)
being used with 120° cell sectors.

V-A Rapidly Moving Vehicles

Suppose that the mobile monitors the average received CIR $\Lambda_P(x_{0j})$ and initiates a hand-off to an adjacent cell when either $\Lambda_P(x_{0j}) < T$ or when the central base-station does not have an available channel. When hand-offs are used the busyness probability $p$ plays an important factor, especially as $p \rightarrow 1$. If $L$ denotes the total number of channels/cell sector, then with probability $p^L$ a base-station will not have an available channel for either a hand-off or call initiation. Let $z_n = \{z_{n0}, z_{n1}, z_{n2}, \ldots, z_{n6}\}$ be a binary $\{0, 1\}$ vector denoting the $n^{th}$ availability pattern of base-stations; $z_{ni} = 1$ if and only if the $i^{th}$ base-station has an available channel, and $z_{ni} = 0$ otherwise. It is important to note that $z_{n0} = 0$ or 1 if a call is being initiated. However, $z_{n0} = 1$ always if a hand-off is being made after a call has been established. The former case is considered here. Let $x_j = \{x_{0j}, x_{1j}, x_{2j}, \ldots, x_{6j}\}$ denote the vector of interference patterns associated with the central base-station and the six adjacent base-stations. If the $i^{th}$ base-station does not have an available channel, then the value of $x_{ij}$ is irrelevant.

Let $\{\Lambda_P(x_{ij}), 0 \leq i \leq 6\}$ denote the average received CIR for the central base-station and the six adjacent base-stations. To initiate a hand-off, the mobile is supplied with $\{\Lambda_P(x_{ij}), 0 \leq i \leq 6\}$. The mobile then hands-off to the base-station that has the largest average received CIR. After handing-off, the new average received CIR as a function of $z_n$ and $x_j$ is

$$L(x_j, z_n) = \max\{z_{n0}\Lambda_P(x_{0j}), z_{n1}\Lambda_P(x_{1j}), z_{n2}\Lambda_P(x_{2j}), \ldots, z_{n6}\Lambda_P(x_{6j})\}.$$ (24)

Usually, path loss has a dominating effect so that the mobile hands-off to the nearest

---

4 The subscript $P$ means that power control is being used.

5 When cell sectoring is used, hand-offs must be made to an appropriate cell sector.
base-station subject to availability. Sometimes $L(x_i, z_n) = \Lambda_P(x_{0i})$ so that a hand-off will not be successful when attempted.

The hand-off threshold $T$ is normally chosen to be greater than $\Lambda_{th}$. Otherwise, an outage may occur that could be prevented by performing a hand-off. This is the only case considered here. Note that the availability of a base-station channel $z_{ni}$ is independent of the interference pattern affecting a channel $x_{ij}$ associated with the base-station. Also, the outage depends not only upon how many base-stations are available, but also upon which base-stations are available. Let $w(z_n)$ be the number of available base-stations, i.e., the number of nonzero components of $z_n$. Then the outage probability at location $(r, \theta)$ due to shadowing, averaged over all availability vectors and all interference hit vectors, is

$$P_o(r, \theta) = (p^L)^7 + \sum_{n:z_n \neq 0} \sum_{j=1}^{(2^K)w(z_n)} P \{ L(x_j, z_n) < \Lambda_{th} \}$$

$$\times \prod_{i=0}^{6} \left\{ \prod_{k=1}^{K} p^{z_{ni}} (1 - p)^{1 - z_{ni}} \right\}^{z_{ni}} (p^L)^{1 - z_{ni}} (1 - p^L)^{z_{ni}}, \quad (25)$$

where the second sum is over all $(2^K)w(z_n)$ interference patterns corresponding to the $w(z_n)$ available base stations. The area averaged outage is given by (8) and (9).

The number of channels/cell sector depends upon the cluster size. Suppose that there are a total of $999 - 42 = 957$ channels and $120^\circ$ cell sectors are used. Assuming that the channels are equally divided among the cell sectors the minimum number of channels/cell sector is 45 for $N = 7$, 79 for $N = 4$, and 106 for $N = 3$.

Figs. 14 to 16 plot the outage probability against the threshold CIR $\Lambda_{th}$ with hand-offs for $N = 3, 4$ and 7. Table I summarizes typical results. When constructing Figs. 14 to 16, a hand-off threshold of $T = \infty$ was chosen so that the effectiveness of hand-offs could be demonstrated for the entire range of $\Lambda_{th}$. In practice, the hand-off threshold will be chosen to be somewhat greater than the $\Lambda_{th}$ corresponding to a desired outage probability. In this way, a hand-off will be attempted before an outage occurs, but excessive hand-offs will be prevented. Note that when $p = 1$
there are no hand-offs because the system is saturated. Therefore, the $P_o$ curves in Figs. 12 and 16 are identical. Further comparison of Figs. 12 and 16 shows that hand-offs result in a significant outage reduction.

V-B Stationary Vehicles

Since $\lambda$ is carrier dependent it is conceivable that a particular channel associated with a base-station is in a deep fade. To eliminate an unnecessary hand-off to an adjacent cell, the mobile is allowed to search a maximum of $N$ channels with the central base-station before a hand-off to an adjacent cell is attempted. All of these channels are assumed to be independently faded. If a hand-off to an adjacent base-station is necessary, the mobile monitors a maximum of $N$ channels for each base-station, all with distinct carrier frequencies and interference patterns. Define $w_n^i = \{z_{n1}^i, z_{n2}^i, \ldots, z_{nN}^i\}$ as the $n^{th}$ availability vector for the $i^{th}$ base-station.

Once again, if a mobile has already gained access to a channel and is attempting to hand-off, then $z_{n1}^0 = 1$. We consider the case of call initiation, where $z_{n1}^0 = 0$ or 1. The probability that the $i^{th}$ base-station has $n$ available channels is

$$p(n) = \begin{cases} \binom{k}{n} (1-p)^n p^{L-n}, & n = 0, 1, \ldots, N-1 \\ 1 - \sum_{n=0}^{N-1} p(n), & n = N \end{cases}$$

(26)

Let $y_j^i = \{x_{j1}^i, x_{j2}^i, \ldots, x_{jN}^i\}$ denote the $j^{th}$ interference pattern for the $i^{th}$ base-station. Then the instantaneous CIR for the central base-station, as a function of $y_j^0$ and $w_n^0$ is $E_0(y_j^0, w_n^0) = \max\{z_{n1}^0 \lambda_P(x_{j1}^0), z_{n2}^0 \lambda_P(x_{j2}^0), \ldots, z_{nN}^0 \lambda_P(x_{jN}^0)\}$. A mobile will initiate a hand-off whenever $E_0(y_j^0, w_n^0) < t$, where $t$ is the hand-off threshold. When a hand-off occurs, the mobile monitors the instantaneous received CIR for a maximum of $N$ channels from each of the six adjacent base-stations. These

---

6 Usually, $N = 2$ obtains most of the improvement from this strategy. Another effective strategy is to use antenna diversity, but this will not be considered here.

7 Note that $t \neq T$ in general, and this may present some implementation problems.
are denoted by \( \{\xi_i(y_j, w^i_n) = \max\{z_{i1}^i\lambda_F(x_{j1}^i), z_{i2}^i\lambda_F(x_{j2}^i), \ldots, z_{iM}^i\lambda_F(x_{jM}^i)\}, 1 \leq i \leq 6\} \). The instantaneous received CIR after a hand-off, as a function of \( y_j = \{y_j^0, y_j^1, \ldots, y_j^6\} \) and \( w_n = \{w_n^0, w_n^1, \ldots, w_n^6\} \), is

\[
\xi(y_j, w_n) = \max\{\xi_0(y_j^0, w_n^0), \xi_1(y_j^1, w_n^1), \xi_2(y_j^2, w_n^2), \ldots, \xi_6(y_j^6, w_n^6)\} \quad (27)
\]

The probability of outage for a fixed mobile location \((r, \theta), y_j, w_n, \) and received field strength vector \( \hat{\mathbf{h}} = \{h_0, h_1, \ldots, h_6\} \), is

\[
P_o(r, \theta, \hat{\mathbf{h}}, y_j, w_n) = \prod_{i=0}^{6} P\left\{\xi_i(y_j^i, w_n^i) < \lambda_{th}\right\} \quad (28)
\]

Here we have assumed that the hand-off threshold \( t \) is chosen so that \( \lambda_{th} < t \) to make the most effective use of the hand-off mechanism.

In the case of CIR driven power control, the analysis cannot be carried further. However, as seen in section IV, CIR driven power control performs about the same as distance driven power control. For distance driven power control, the analysis can be carried further by first recognizing that the cumulative distribution function of \( \xi_i(y_j^i, w_n^i) \) is

\[
F_{\xi_i(y_j^i, w_n^i)}(x) = \prod_{m=1}^{N} \left[ F_{\lambda(x_{jm}^m)}(x) \right]^{x_{jm}^m} \quad (29)
\]

where \( F_{\lambda(x_{jm}^m)}(x) \) is given by (17) with \( x_{j}^0 \) replaced by \( x_{jm}^0 \). Here, \( x_{jm}^m = 1 \) denotes that for the \( j^\text{th} \) interference pattern, the \( k^\text{th} \) interferer associated with the \( m^\text{th} \) channel of the \( i^\text{th} \) base-station is active. Otherwise, \( x_{jm}^m = 0 \). Therefore, (28) becomes

\[
P_o(r, \theta, \hat{\mathbf{h}}, y_j, w_n) = \prod_{i=0}^{6} \prod_{m=1}^{N} \left[ F_{\lambda(x_{jm}^m)}(\lambda_{th}) \right]^{x_{jm}^m} \quad (30)
\]

The outage probability averaged over all availability patterns and all interference patterns can be written as
\[ P_o(r, \theta, \hat{r}) = (p^L)^{2N} + \sum_{n:w_n \neq 0} \sum_{j=1}^{(2^K)w(w_n)} \prod_{i=0}^6 p(w(w_n^j)) \times \prod_{m=1}^{K} \left\{ F_{\lambda(x_{m_k})}(\lambda_{th}) \prod_{k=1}^{K} p(x_{m_k}^i (1 - p)^{1-x_{m_k}^i}) \right\} \]  

Once again, the second summation is over all \((2^K)w(w_n)\) interference patterns corresponding to the \(w(w_n)\) available channels. Finally, the average outage probability due to shadowing is given by (19) and the area averaged outage probability is given by (8) and (9). Figs. 17 to 19 plot the average outage probability against the instantaneous threshold CIR \(\lambda_{th}\) for \(N = 3, 4,\) and 7 and various busyness probabilities. Table II summarizes typical results.

VI Digital Cellular System

A block diagram of the modulation system under consideration is shown in Fig. 20. The modulator uses \(\pi/4\) shifted QDPSK and raised cosine pulse shaping with a rolloff factor of 0.8. The data rate is 40 kb/s, resulting in a quaternary symbol rate of 20 ks/s. Hence, the bandwidth of the transmitted signal is 32 kHz. The carrier separations are assumed to be 30 kHz.

The receiver front end uses a third-order Butterworth filter to reject adjacent channel interference and out-of-band noise. The cutoff frequency (3 dB bandwidth) of the filter at baseband is 15 kHz. The detector uses a sample and hold, where complex samples are taken at the middle of the symbol periods where minimal intersymbol interference occurs. The samples are processed with a differential phase detector. Finally, the \(\pi/4\) phase shift is removed, resulting in a decision variable. In the sequel we consider the use of Reed-Solomon codes with errors only decoding, so this decision variable is used to make symbol-by-symbol decisions.

The performance of the above system was analyzed by using SYSTID, a time
domain simulator [12]. It was reasonably assumed that the desired signal and each of the co-channel interferers were affected by independent Rayleigh fading. The fading was simulated by using the approach suggested by Jakes [4]. The channel was assumed to be free of intersymbol interference.

Fig. 21 plots the bit error probability against the CIR (λ or Λ). The curve to the left is the bit error probability for a nonfaded channel. In this case the received signal is corrupted by the presence of six nonfaded interferers and the background noise is assumed to be negligible. The probability of bit error is approximately

\[ P_b \approx 0.2238 \exp \left( -\frac{\lambda}{2.305} \right) . \]  

(32)

The curve to the right shows the performance when the desired signal and six interfering signals are affected by independent Rayleigh fading. The probability of bit error can be closely approximated as

\[ P_b \approx \frac{0.3980}{\Lambda^{0.7777}} . \]  

(33)

VII Performance of Reed-Solomon Codes

When analyzing the performance of Reed-Solomon (RS) codes, two cases are considered. The first assumes ideal interleaving so that the channel is memoryless. This corresponds to the case of rapidly moving vehicles. The second assumes that the channel is static, but random, during the reception of an entire code word. This corresponds to the case of stationary vehicles. For RS codes, symbol interleaving is more effective than bit interleaving, because symbol interleaving will tend to trap some of the error bursts. The effectiveness of using symbol interleaving will be discussed.
VII-A Rapidly Moving Vehicles

For QDPSK modulation with ideal bit interleaving, the probability of quaternary symbol error is \( P_Q = 1 - (1 - P_b)^2 \), where \( P_b \) is given by (33). With RS codes over \( GF(2^k) \), \( 2^{k-1} \) quaternary symbols are used to form one code symbol. Therefore, the probability of code symbol error is \( P_M = 1 - (1 - P_Q)^{k/2} \). For \( M \)-ary \( (M = 2^k) \) block codes with bounded distance decoding, the decoded symbol error probability can be approximated as [13]

\[
P_s \approx \frac{1}{n} \sum_{i=e+1}^{n} \binom{n}{i} P_M^i (1 - P_M)^{n-i},
\]

where \( e = \lfloor (d - 1)/2 \rfloor \) is the number of code symbol errors that can be corrected by the code \( (d \) is the minimum distance of the code, \( n \) is the block length, and \( \lfloor x \rfloor \) is the largest integer less than or equal to \( x \)). RS codes are maximum distance separable, so that for an \( (n, k) \) code \( d = n - k + 1 \), where \( n = 2^k - 1 \). An extended RS code is formed by adding an overall parity symbol, so that RS codes with an odd minimum distance will have their minimum distance increased by one. The probability of decoded bit error conditioned on the occurrence of a decoded symbol error is certainly no larger than \( \frac{1}{2} \). Hence \( P_b \approx \frac{1}{2} P_s \). The average received CIR \( A \) required to obtain a decoded bit error probability of \( P_b = 10^{-3} \) is shown in Fig. 22 for all extended RS codes of length \( 2^k \), \( k = 4, 6, 8 \).

To investigate the effect of symbol interleaving with RS codes, Fig. 23 plots the probability of symbol error against \( A \) with ideal symbol interleaving and ideal bit interleaving. Results are shown for 6 and 8 bit symbols, corresponding to RS \( (64, k) \) and RS \( (256, k) \) codes. A vehicle velocity of 100 km/hr and a carrier frequency of 900 MHz is assumed. The channel is Rayleigh faded with six Rayleigh faded interferers. Notice that symbol interleaving is about 2.5 dB and 3.0 dB more effective than bit interleaving with 6 bit and 8 bit symbols, respectively. Hence, the CIR values in Fig. 22 can be reduced by 2.5 dB for RS \( (64, k) \) codes and 3.0 dB for
RS \((256, k)\) codes.

Suppose that the digital cellular system described in section VI uses TDMA with three channels per carrier. Since the total bit rate per carrier is 40 kb/s, each channel is allocated a rate of 13.33 kb/s. The actual data rate will be less than 13.33 kb/s because some overhead will be required for synchronization and guard times. However, as a first approximation suppose that the full 13.33 kb/s is available to each user. The threshold CIR required to achieve a specified outage probability has been provided in Figs. 14 to 16 and Table I. By using Fig. 22, the largest code rate \(r_c\) that will achieve a specified decoded bit error probability can be determined. Since the total bit rate allocated to each channel is 13.33 kb/s, the maximum permissible voice coder rate is \(r_y = 13.33r_c\) kb/s. Table III tabulates values of \(r_c\) and \(r_y\) for various outage probabilities and \(p = 0.90\).

VII-B Stationary Vehicles

For stationary vehicles the channel attenuation is static, but random, during the reception of an entire code word. Therefore, no additional improvement will result from using symbol interleaving, because the errors occur randomly rather than in bursts. Consequently, the decoded symbol error probability is given by (34), where 
\[
P_M = 1 - (1 - P_Q)^{k/2} \quad \text{and} \quad P_Q = 1 - (1 - P_b)^2.
\]
However, in this case \(P_b\) is given by (32). The instantaneous received CIR \(\lambda_{th}\) required to achieve a decoded bit error probability of \(P_b = 10^{-3}\) is shown in Fig. 24 for all extended RS codes of length \(2^k, \ k = 4, 6, 8\). By using Fig. 24 and Figs. 17 to 19 (or Table II), it is apparent that coding is not necessary to achieve \(P_b = 10^{-3}\) with stationary vehicles. Therefore, stationary vehicles are much less prone to outage than rapidly moving vehicles.
VIII Concluding Remarks

This paper has presented simulation based results for predicting downlink outage in digital cellular radio systems. Obtaining the same results through analytical techniques is intractable due to the extreme complexity of the problem. The outage results were applied to a digital cellular system that uses QDPSK signaling with Reed-Solomon coding. One important conclusion is that the outage performance is limited by rapidly moving vehicles, the outage performance with stationary vehicles is always better.
References


Table I

$A_{th}$ Required to Achieve $P_O$ for Rapidly Moving Vehicles

<table>
<thead>
<tr>
<th>$P_O$</th>
<th>$N = 7$</th>
<th>$N = 4$</th>
<th>$N = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>27.6</td>
<td>28.4</td>
<td>26.3</td>
</tr>
<tr>
<td>2%</td>
<td>32.4</td>
<td>30.3</td>
<td>28.2</td>
</tr>
<tr>
<td>3%</td>
<td>34.1</td>
<td>31.5</td>
<td>29.4</td>
</tr>
<tr>
<td>4%</td>
<td>35.3</td>
<td>32.4</td>
<td>30.3</td>
</tr>
<tr>
<td>5%</td>
<td>36.2</td>
<td>33.2</td>
<td>31.1</td>
</tr>
</tbody>
</table>

Table II

$\lambda_{th}$ Required to Achieve $P_O$ for Stationary Vehicles

<table>
<thead>
<tr>
<th>$P_O$</th>
<th>$N = 7$</th>
<th>$N = 4$</th>
<th>$N = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>21.5</td>
<td>22.1</td>
<td>20.1</td>
</tr>
<tr>
<td>2%</td>
<td>26.6</td>
<td>24.0</td>
<td>21.9</td>
</tr>
<tr>
<td>3%</td>
<td>28.3</td>
<td>25.2</td>
<td>23.3</td>
</tr>
<tr>
<td>4%</td>
<td>29.5</td>
<td>26.2</td>
<td>24.4</td>
</tr>
<tr>
<td>5%</td>
<td>30.5</td>
<td>27.1</td>
<td>25.2</td>
</tr>
</tbody>
</table>

Table III

Error Correction Codes Rates and Voice Coder Rates for Various Outages and Cell Cluster Sizes with RS(64,k) Codes

<table>
<thead>
<tr>
<th>Outage Probability</th>
<th>$N = 7$</th>
<th>$N = 4$</th>
<th>$N = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$r_e$ kb/s</td>
<td>$k$</td>
</tr>
<tr>
<td>1%</td>
<td>54</td>
<td>11.25</td>
<td>54</td>
</tr>
<tr>
<td>2%</td>
<td>58</td>
<td>12.08</td>
<td>56</td>
</tr>
<tr>
<td>3%</td>
<td>60</td>
<td>12.50</td>
<td>58</td>
</tr>
<tr>
<td>4%</td>
<td>62</td>
<td>12.92</td>
<td>58</td>
</tr>
<tr>
<td>5%</td>
<td>62</td>
<td>11.25</td>
<td>60</td>
</tr>
</tbody>
</table>
Figure 1

Probability of Outage vs. Threshold CIR $\Lambda_{th}$; $\alpha = 3.68$, $\sigma = 8$
Figure 2
Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}; \alpha = 3.68, \sigma = 8$
Figure 3
Probability of Outage vs. Threshold CIR $A_{th}$ with Cell Sectoring; $\alpha = 3.68$, $\sigma = 8$
Figure 4
Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring;

$\alpha = 3.68, \sigma = 8$
Figure 5

Probability of Outage vs. Threshold CIR $A_{th}$ with Cell Sectoring and CIR Driven Power Control (first-order approximation); $\alpha = 3.68$, $\sigma = 8$
Figure 6

Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring and CIR Driven Power Control (first-order approximation); $\alpha = 3.68$, $\sigma = 8$
Figure 7
Probability of Outage vs. Threshold CIR $A_{th}$ with Cell Sectoring and CIR Driven Power Control (second-order approximation); $\alpha = 3.68$, $\sigma = 8$
Figure 8

Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring and CIR Driven Power Control (second-order approximation); $\alpha = 3.68$, $\sigma = 8$
Figure 9
Probability of Outage vs. Threshold CIR $A_{th}$ with Cell Sectoring and CIR Driven Power Control (third-order approximation); $\alpha = 3.68, \sigma = 8$
Figure 10
Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring and CIR Driven Power Control (third-order approximation); $\alpha = 3.68$, $\sigma = 8$
Figure 11
Concept of Distance Driven Power Control
Figure 12

Probability of Outage vs. Threshold CIR $A_{th}$ with Cell Sectoring and Distance Driven Power Control; $\alpha = 3.68, \sigma = 8$
Figure 13
Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring and Distance Driven Power Control; $\alpha = 3.68$, $\sigma = 8$
Figure 14

Probability of Outage vs. Threshold CIR \( \Lambda_{th} \) with Cell Sectoring, Distance Driven Power Control, and Hand-offs; \( N = 7, \alpha = 3.68, \sigma = 8 \)
Figure 15
Probability of Outage vs. Threshold CIR $\Lambda_{th}$ with Cell Sectoring, Distance Driven Power Control, and Hand-offs; $N = 4$, $\alpha = 3.68$, $\sigma = 8$
Figure 16
Probability of Outage vs. Threshold CIR $\Lambda_{th}$ with Cell Sectoring, Distance Driven Power Control, and Hand-offs; $N = 3$, $\alpha = 3.68$, $\sigma = 8$
Figure 17
Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring, Distance Driven Power Control, and Hand-offs; $N = 7$, $\alpha = 3.68$, $\sigma = 8$
Figure 18

Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring, Distance Driven Power Control, and Hand-offs; $N = 4, \alpha = 3.68, \sigma = 8$
Figure 19
Probability of Outage vs. Instantaneous Threshold CIR $\lambda_{th}$ with Cell Sectoring, Distance Driven Power Control, and Hand-offs; $N = 3$, $\alpha = 3.68$, $\sigma = 8$
Figure 20
QDPSK System
Figure 21

Probability of Bit Error vs. CIR
Figure 22
Average CIR Required for $P_b = 10^{-3}$ with RS Codes
Figure 23
Probability of Symbol Error with Ideal Bit and Symbol Interleaving
$P_b = 10^{-3}$

**Figure 24**

Instantaneous CIR Required for $P_b = 10^{-3}$ with RS Codes
MLSE Equalization and Decoding for Multipath-Fading Channels *

Wern-Ho Sheen and Gordon L. Stüber
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332
(404) 894-2923
December 18, 1989

Abstract

The performance of a receiver using a combined MLSE equalizer/decoder and $D$-diversity reception is analyzed for multipath Rayleigh fading channels. A new upper bound on decoded bit error probability is derived. Comparisons with simulation results show that this upper bound is very tight, especially when the signal-to-noise ratio is moderate or when diversity reception is used.

*This research was supported by BellSouth Enterprises Inc.
I Introduction

Many physical communication channels such as HF shortwave ionospheric, UHF troposcatter, VHF ionospheric forward scatter, UHF land mobile radio, and others [1,2,3] can be effectively modeled as time-variant multipath-fading channels. Three phenomena, namely time spread, Doppler spread, and multipath fading have been recognized as the main impairments to reliable communication with these types of channels. Time spread causes interference between adjacent symbols, known as InterSymbol Interference (ISI), Doppler spread necessitates a fast convergent algorithm when an adaptive receiver is employed, and multipath-fading results in a very low received signal-to-noise ratio when channel exhibits a deep fade.

It is well known that Maximum Likelihood Sequence Estimation (MLSE) is the optimum (maximum likelihood) detection technique for a digital signal corrupted by ISI and additive white noise. The Viterbi algorithm is a computationally efficient and practical method for implementing MLSE for moderate constraint length convolutional codes and channels with short impulse responses. In addition to using coding, multipath-fading can be combated by using diversity techniques where the receiver is provided with multiple independently faded replicas of the same information signal. Diversity is effective because the probability of simultaneously having two or more independently faded channels in a deep fade is small [1]. To overcome the effect of Doppler spread, a fast-convergent algorithm is usually employed in the receiver. For example, an adaptive equalizer using a fast Kalman or recursive Least Square Lattice algorithm may be needed [4,5].

The performance of MLSE has been analyzed thoroughly by Forney [6] and some others [7,8,9] for time-invariant channels. Very recently a new tighter upper bound on the bit error probability has been found by Verdú [10]. In this paper, we analyze the performance of MLSE with \( D \)-diversity reception for slowly time-varying multipath Rayleigh fading channels by developing a procedure similar to that in [6-
Since error correction coding might be advantageous for power-limited and/or bandwidth-limited channels [9,11], convolutional coding schemes are also included in our analysis. The primary result presented in this paper is a new upper bound on the bit error probability. Computer simulations show that the new upper bound is quite tight when diversity is used or when the signal-to-noise ratio is moderate.

The paper is organized as follows. Section II describes the system structure and channel model. Section III gives the structure of MLSE receiver. In Section IV, the upper bound is derived. In Section V, three example systems are analyzed and the analytical results are compared to computer simulations.

II System and Channel Model

A typical convolutionally coded digital communication system with \( D \)-diversity reception is shown in Fig. 1. \( \{I_k\} \) represents a sequence of \( r \)-bit symbols to be transmitted over the channel. \( \{U_k\} \) is the corresponding \( n \)-bit sequence out of the encoder which in turn is mapped into a \( N \)-symbol sequence \( \{X_k\} \) suitable for modulation, where \( X_k = (x_{kN}, x_{kN-1}, \ldots, x_{(k-1)N+1}) \). If we assume that \( x_j \) has an alphabet size equal to \( m \), then \( N = (\log_2 m)/n \). In practice, it is natural to assume that \( N \) is a positive integer. Since \( D \)-diversity is used in the system, the channel is modeled as \( D \) independent dispersive fading channels corrupted by additive white Gaussian noise \( \eta_i(t) \). A matched filter is employed to get the maximum signal-to-noise ratio at sampling instant and a noise-whitening filter is used to whiten the colored noise coming out of the matched filter for convenient system modeling. The received signal at epoch \( k \) is \( \{Y_k^d\} \), where \( Y_k^d = (y_{kN}^d, y_{kN-1}^d, \ldots, y_{(k-1)N+1}^d) \). \( \{\hat{I}_k\} \) is an estimate of the transmitted sequence \( \{I_k\} \).

Because all signals, channel and filter responses are written in equivalent low-pass form, \( X_k \) and \( Y_k^d \) are all complex-valued signals. As described in [1,6] it is convenient to model the transmitter filter, channel, matched filter, and noise-
whitening filter as a \((L + 1)\)-tap transversal filter with tap spacing equal to \(T_z\), the signal duration. (Sometimes it is called a discrete-time white noise model \([1]\)). The resulting system model is shown in Fig. 2. Since we are concerned with Rayleigh fading, the tap coefficients \(\{g_i^d\}\) are modeled as independent zero-mean complex-valued Gaussian random processes with variances \(\sigma_{g_i^d}^2 = \frac{1}{2}E[|g_i^d|^2]\) described by the channel multipath intensity profile (MIP) \([12]\). To model the time variations associated with the channel impulse response for simulation purposes, the \(\{g_i^d\}\) can be generated by passing complex white Gaussian noise through a lowpass filter having a cut off frequency equal to maximum Doppler frequency \([5]\). With this model

\[
y_{kN-j}^d = \sum_{m=0}^{L} g_m^d x_{kN-j-m} + \eta_{kN-j}^d, \quad j = 0, \ldots, N - 1.
\]

Since a noise-whitening filter has been built into system model, the terms \(\{\eta_i^d\}\) are i.i.d. zero-mean complex-valued Gaussian noise.

It is worthy to note that if the convolutional code has constraint length \(K\), then the overall channel is a finite state machine with \(M^J\) states, where \(M = 2^r\) and \(J = K - 1 + \lfloor L/N \rfloor\), and \(\lfloor x \rfloor\) denotes the smallest integer larger than \(x\). Hence, it can be described very conveniently by using an \(M^J\)-state trellis diagram \([6]\).

### III MLSE Receiver

Assume that \(k\) \(r\)-bit messages are transmitted over the channel. After receiving the sequence \(\{y_i\}_{i=1}^{k}\), where \(y_i = (Y_i^1, Y_i^2, \ldots, Y_i^D)\) is the collection of received signals for all diversity branches at epoch \(i\), the MLSE makes a decision in favor of the sequence \(\{I_i\}_{i=1}^{k}\) that maximizes the joint conditional probability density function

\[
P\{y_k, y_{k-1}, \ldots, y_1 \mid I_k, I_{k-1}, \ldots, I_1\}
\]
or, equivalently, the logarithm of this function

$$\ln[P\{Y_k, Y_{k-1}, \ldots, Y_1|I_k, I_{k-1}, \ldots, I_1\}]$$

(3)

Since the noise samples \(\{\eta_i^d\}\) are independent and \(Y_i\) only depends on the most recent \(J = K - 1 + [L/N]\) transmitted \(r\)-bit messages, (3) can be rewritten in the following form

$$\ln[P\{Y_k, Y_{k-1}, \ldots, Y_1|I_k, I_{k-1}, \ldots, I_1\}] = \ln[P\{Y_k|I_k, \ldots, I_{k-J}\}]$$

$$+ \ln[P\{Y_{k-1}|I_{k-1}, \ldots, I_1\}]$$

(4)

where \(I_{k-J} = 0\) for \(k - J \leq 0\). Because the second term has been calculated previously, only the first term, called branch metric, has to be computed for each incoming signal \(Y_k\). The algorithm that recursively calculates the joint conditional probability as described above is the well known Viterbi algorithm [13].

According to the model in Section II

$$\ln[P\{Y_k|I_k, \ldots, I_{k-J}\}] = C \cdot \left\{- \sum_{d=1}^{D} \sum_{j=0}^{N-1} |y_{kN-j}^d - \sum_{m=0}^{L} g_m^d x_{kN-j-m}|^2 \right\}$$

(5)

where \(C\) is a constant independent of transmitted sequence \(\{I_i\}\). Removing the common constant results in the simplified branch metric

$$\Pi_k = - \sum_{d=1}^{D} \sum_{j=0}^{N-1} |y_{kN-j}^d - \sum_{m=0}^{L} g_m^d x_{kN-j-m}|^2$$

(6)

From (6), the receiver needs to know the channel impulse response \(g_m^d, m = 0, \ldots, L\) to compute the metrics \(\Pi_k\). In our system the channel is time-varying so that an adaptive channel estimator has to be used for the MSLE receiver to work properly. Usually a transversal digital filter with an LMS algorithm is employed as a channel estimator for its hardware simplicity and very good performance [1,14].

As implied in (2), we may theoretically wait until the whole sequence \(\{Y_i\}_{i=1}^\infty\) has been received and then make a decision on the whole sequence. In practice such
a long delay (maybe infinite) is intolerable so very often a decision about $I_{k-Q}$ is made when $Y_k$ is received and processed. It has been reported that if $Q \geq 5J$, the performance degradation caused by path metric truncation is negligible [1].

Based on (6) and the discussions above, the proposed MLSE receiver structure is illustrated in Fig. 3. Input buffers store the received signals $Y_k^d$. The branch metrics $\Pi_k$ are computed by using the estimated channel impulse response $\hat{\theta}_i^d$ and channel input sequence corresponding to surviving paths. After this computation MLSE chooses the new set of surviving paths and makes a decision on $I_{k-Q}$. The sequence $\{\tilde{I}_{k-Q}\}$ is then transmitted to the digital sink and channel estimator for updating the estimated channel impulse response. This finishes up the work performed by the receiver, and receiver waits for another incoming signal. Note that it is possible to make decisions with delay $\hat{Q} < Q$ for the purpose of updating the channel estimator. This might be desirable if the channel is rapidly time-varying.

IV Performance Analysis

We analyze the system performance by developing a procedure similar to that in [6-9]. Recall that $\{I_i\}$ and $\{\tilde{I}_i\}$ represent a transmitted and an estimated sequence respectively. For every pair of sequences $\{I_i\}$ and $\{\tilde{I}_i\}$ the error sequence $e = \{e_i\}$ can be formed by defining $e_i = I_i - \tilde{I}_i$. Because we are interested in the bit error probability at epoch $k$, we assume $e_k \neq 0$ for all error sequences considered in this paper. For each error sequence $e$, we first define the following useful error events.

$\mathcal{E}'(e)$: The sequence $\{I_i\} - \{e_i\}$ is the maximum likelihood sequence.

$\mathcal{E}(e)$: The sequence $\{I_i\} - \{e_i\}$ has a larger path metric than the transmitted sequence $\{I_i\}$.

It is also convenient to define the events
where $F$ is the set of all possible error sequences and $E$ is the subset of $F$ consisting of error sequences that contain no more than $J - 1$ consecutive zeroes amid nonzero elements.

Following the above definitions, the probability that an error occurs at epoch $k$ is given by

$$P(I_k \neq \tilde{I}_k) = P(E'_F).$$

It can also be shown, although it is not self-evident, that

$$P(I_k \neq \tilde{I}_k) = P(E_E)$$

To see this consider the typical error-state trellis diagram in Fig. 4, where the error state at epoch $k$ is denoted by $S_k = (e_{k-1}, e_{k-2}, \cdots, e_{k-J})$, and the 0-state represents the zero-error state. Observe that for every $\varepsilon \in F \setminus E$ there exists an $\varepsilon \in E$. If the sequence $\{I_k\} - \varepsilon$ is the maximum likelihood sequence (i.e., the event $E'_F$ has occurred), then the sequence $\{I_k\} - \varepsilon \in E$ has a larger path metric than the sequence $\{I_k\}$ (i.e., the event $E_E$ has occurred). This means that $E'_F \Rightarrow E_E$. On the other hand, if $\varepsilon \in E$ and the sequence $\{I_k\} - \varepsilon \in E$ has a larger path metric than sequence $\{I_k\}$, then there exists a sequence $\varepsilon \in F$ such that the sequence $\{I_k\} - \varepsilon$ is the maximum likelihood sequence. Therefore, $E_E$ implies $E'_F$. Hence (10) follows.

$P(E_E)$ can be upper bounded by using a union bound

$$P(E_E) \leq \sum_{\varepsilon \in E} P(E(\varepsilon))$$

7
In fact, this is the upper bound used in [6-9]. Unfortunately, this upper bound is often very loose for multipath-fading channels. Hence, we use the following bound instead

\[ P(I_k \neq \tilde{I}_k) \leq \sum_{i=1}^{R} \sum_{\epsilon \in E_i} P(\epsilon) + \sum_{i=R+1}^{\infty} P(\bigcup_{\epsilon \in E_i} \epsilon) \]  

(12)

where \( R \) is some positive integer and \( E_i \subset E \) is the set of all error sequences having \( i \) bit errors. For multipath fading channels, most of the multi-error events occur when the channel exhibits a deep fade. In this case, it is very likely that event \( \epsilon_i \) implies event \( \epsilon_j \) for \( \epsilon_i, \epsilon_j \in E, \forall m \). The reason for this will be discussed after the development of (16). As a result the upper bound in (12) can be approximated as

\[ P(I_k \neq \tilde{I}_k) \leq \sum_{i=1}^{R} \sum_{\epsilon \in E_i} P(\epsilon) + \sum_{i=R+1}^{\infty} \max_{\epsilon \in E_i} \{ P(\epsilon) \} \]  

(13)

Although (15) is not an upper bound in all cases, it can be viewed as an upper bound for most applications for an appropriate choice of \( R \). Usually, \( R = 2 \) or 3 is sufficient.

From our definition about the event \( \epsilon \) we know (from now on we drop the letter \( \varepsilon \) for notation convenience)

\[ P(\epsilon) = P(\epsilon_1|\epsilon_{1})P(\epsilon_{1}) \]  

(14)

where the subevents \( \epsilon_1 \) and \( \epsilon_2 \) are defined as follows;

\( \epsilon_1 \): The set of information sequences \( \{I_i\} \) with the property that \( \{\tilde{I}_i\} = \{I_i\} - \{\epsilon_i\} \) is an allowable sequence.

\( \epsilon_2 \): The set of noise sequences such that the estimated sequence \( \{\tilde{I}_i\} \) has greater path metric than the correct sequence \( \{I_i\} \).

Note that \( \epsilon_1 \) depends only on the coding scheme and not on the channel, and \( P(\epsilon_1) \) can be calculated as shown later.
To compute $P(\xi_2|\xi_1)$ from (6)$^1 \times 2$

$$
P(\xi_2|\xi_1) = P\left\{ \sum_{d=1}^{D} \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} \left[ y_{iN-j}^d - \sum_{m=0}^{L} g_m^d \bar{z}_{iN-j-m} \right]^2 \right\}
$$

$$
< \sum_{d=1}^{D} \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} \left[ y_{iN-j}^d - \sum_{m=0}^{L} g_m^d x_{iN-j-m} \right]^2 \right\} \quad (15)
$$

where $\{\bar{z}_i\}$ is the channel input sequence corresponding to the estimated sequence $\{\hat{I_i}\}$. $H(e) + 1$ is the length in branches of the error path which diverges from the correct path in the $M^j$-state trellis diagram at epoch $k = 0$.

Substituting (1) into (15) results in

$$
P(\xi_2|\xi_1) = P\left\{ \sum_{d=1}^{D} \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} \left[ \sum_{m=0}^{L} g_m^d \bar{z}_{iN-j-m} + \eta_{iN-j}^d \right]^2 \right\}
$$

$$
< \sum_{d=1}^{D} \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} \left| \eta_{iN-j}^d \right|^2 \right\}
$$

$$
= P\left\{ \sum_{d=1}^{D} \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} 2Re \left[ (\theta_{iN-j}^d)^* \eta_{iN-j}^d \right] \right\}
$$

$$
< - \sum_{d=1}^{D} \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} \left| \theta_{iN-j}^d \right|^2 \right\} \quad (16)
$$

where$^3$

$$
\bar{e}_j = x_j - \bar{x}_j \quad (17)
$$

$^1$If the length of transmitted sequence is long enough it is reasonable to assume the error probability is independent of time index $k$ so that $k$ can be arbitrary set equal to zero.

$^2$It is assumed that the $\{g_i^d\}$ are known exactly.

$^3$For a coded system, the sequence $\{\bar{e}_i\}$ depends not only on the error sequence but also on the correct sequence $\{I_i\}$. However at this moment we only consider a particular correct sequence $\{I_i\}$ and later the probability will be averaged over all possible correct sequences.
and

\[ \theta_j^d = \sum_{m=0}^{L} g_m^d \bar{e}_{j-m}. \]  

(18)

The reason why event \( \mathcal{E}(e_i) \) implies event \( \mathcal{E}(e_j) \) for \( e_i, e_j, \in E_m \) for multi-error events can be seen from (16). When the channel is in a deep fade the \( \{g_m^d\} \) and hence the \( \theta_j^d \) are quite small. Therefore, \( P(\mathcal{E}_2|\mathcal{E}_1) \) is dominated by the noise sequence \( \{\eta_i^d\} \).

Because the \( \{\eta_i^d\} \) are i.i.d. zero-mean complex Gaussian random variables with variance

\[ \sigma_n^2 = \frac{1}{2} E[|\eta_i^d|^2] = N_o, \quad d = 1, \cdots, D \]  

(19)

it follows that

\[ P(\mathcal{E}_2|\mathcal{E}_1) = \frac{1}{2} \text{erfc} \left( \sqrt{\frac{1}{8N_o}} \sum_{d=1}^{D} \sum_{i=0}^{H(t)} \sum_{j=0}^{N-1} |\theta_{iN-j}^d|^2 \right). \]  

(20)

where

\[ \text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \, dt. \]  

(21)

Denote the new random variable \( \chi \) as

\[ \chi = \frac{1}{8N_o} \sum_{d=1}^{D} \sum_{i=0}^{H(t)} \sum_{j=0}^{N-1} |\theta_{iN-j}^d|^2 \]  

(22)

so that

\[ P(\mathcal{E}_2|\mathcal{E}_1) = \frac{1}{2} \text{erfc} (\sqrt{\chi}) \]  

(23)

and

\[ P(\mathcal{E}) = \frac{1}{2} \text{erfc} (\sqrt{\chi}) \, P(\mathcal{E}_1). \]  

(24)
Since the channel under consideration is a time-varying multipath-fading channel, the error probability must be obtained by averaging over the fading statistics. Recall that the channels we are interested in are slowly time-varying. Therefore, it is reasonable to assume that the channel impulse response remains fixed over several tens of symbol intervals. This observation greatly simplifies our analysis, because the error probability can be averaged over the ensemble of channel impulse responses. That is,

$$\bar{P}(\mathcal{E}) = P(\mathcal{E}_1) \cdot \int_0^\infty \frac{1}{2} \operatorname{erfc} (\sqrt{x}) \ f_x (x) \ dx$$  \hspace{1cm} (25)$$

where $f_x(x)$ is the probability density function of the random variable $x$.

Combining (13) and (25), the average bit error probability is bounded by

$$P_b \lesssim \frac{1}{r} \left\{ \sum_{i=1}^R \sum_{\mathbf{e} \in \mathcal{E}_i} <\bar{P}(\mathcal{E})> + \sum_{i=R+1}^\infty \max_{\mathbf{e} \in \mathcal{E}_i} \{ <\bar{P}(\mathcal{E})> \} \right\}$$  \hspace{1cm} (26)$$

where $<\bar{P}(\mathcal{E})>$ denotes the error event probability averaged over all possible correct sequences.

To evaluate the integral in (25) we proceed as follows. First, define the new random variables

$$\chi_d = \sum_{i=0}^{H(\mathcal{E})} \sum_{j=0}^{N-1} \left| g_{iN-j}^d \right|^2$$

$$= \sum_{i=0}^{H(\mathcal{E})} \sum_{j=0}^{N-1} \left| \sum_{m=0}^L g_{m}^d \tilde{e}_{iN-j-m} \right|^2 , \quad d = 1, \ldots, D$$  \hspace{1cm} (27)$$

Then from (22)

$$\chi = \frac{1}{8 N_o} \sum_{d=1}^D \chi_d$$  \hspace{1cm} (28)$$

Let $g_i^d = g_{i R}^d + j g_{i I}^d$ and let $\tilde{e}_i = \tilde{e}_{i R} + j \tilde{e}_{i I}$. Then it is shown in Appendix A that $\chi_d$ can be written in the following quadratic form

$$\chi_d = \mathbf{h}_d^T \mathbf{B} \mathbf{h}_d,$$  \hspace{1cm} (29)$$
where $T$ denotes vector or matrix transposition, $h_d = (h_1^d, \ldots, h_{2(L+1)}^d) = (g_0^R, g_1^R, \ldots, g_L^R, g_0^I, g_1^I, \ldots, g_L^I)^T$, and $B$ is a $2(L + 1)^{th}$ order real symmetric matrix of the form

$$B = \begin{bmatrix} \hat{B} & 0 \\ 0 & \hat{B} \end{bmatrix},$$

where $\hat{B}$ has elements

$$\hat{b}_{mn} = \sum_{i=0}^{H(e)} \sum_{j=0}^{N-1} (\bar{e}_{iN-j-m} R \bar{e}_{iN-j-n} R \bar{e}_{iN-j-m} I + \bar{e}_{iN-j-m} \bar{e}_{iN-j-n} I).$$

Since $B$ is a real symmetric matrix, there exists a linear transformation \cite{15} such that

$$\omega_d = C h_d$$

such that

$$\chi_d = \omega_d^T \Lambda \omega_d = \sum_{i=1}^{2(L+1)} \lambda_i (\omega_i^d)^2$$

where $\omega_d = (\omega_1^d, \omega_2^d, \cdots, \omega_{2(L+1)}^d)^T$, $C = [c_{ij}]$ is a $2(L + 1)^{th}$ order orthogonal matrix, and $\Lambda$ is a $2(L + 1)^{th}$ order diagonal matrix with diagonal elements $\lambda_1, \cdots, \lambda_{2(L+1)}$. \footnote{Actually, $\lambda_1, \cdots, \lambda_{2(L+1)}$ are the eigenvalues associated with matrix $B$.} From (32), we also know

$$\omega_i^d = \sum_{m=1}^{2(L+1)} c_{im} h_m^d, \quad i = 1, \cdots, 2(L + 1).$$

Without loss of generality, we can assume that the $h_m^d$, $m = 1, \cdots, 2(L+1)$ are \textit{i.i.d} zero-mean Gaussian random variables with variance $\sigma^2$. \footnote{In general, $\sigma_i^2 \neq \sigma_j^2$ for $i \neq j$. However, it is easy to normalize the \{g_i^d\} such that they have the same variance and $B$ still remains in real symmetric form.} In this case, $\omega_i^d$, $i = 1, \cdots, 2(L + 1)$ are also \textit{i.i.d.} zero-mean Gaussian random variables with variance $\sigma^2$. The proof is as follows. Since the $h_m^d$ are zero-mean Gaussian random variables, it is apparent that the $\omega_i^d$ are also zero-mean Gaussian random variables.
Since
\[
E \left[ \omega_i \omega_j \right] = \sum_{n=1}^{2(L+1)} \sum_{m=1}^{2(L+1)} c_{in} c_{jm} E \left[ h_n^d h_m^d \right]
\]
\[
= \sum_{m=1}^{2(L+1)} c_{im} c_{jm} E \left[ (h_m^d)^2 \right]
\]
\[
= \begin{cases} 
\sigma^2 & i = j \\
0 & i \neq j 
\end{cases}
\] 
(35)

it follows that the \( \omega_i^d \) are i.i.d. zero-mean Gaussian random variables with variance \( \sigma^2 \). In the final step, we have used the fact that the rows of matrix \( C \) are orthonormal to each other.

It is apparent from the form of \( B \) in (30) that the eigenvalues \( \{ \lambda_i \} \) of \( B \) always appear in pairs. Therefore, (33) can be rewritten as

\[
\chi_d = \sum_{j=1}^{L+1} \lambda_j \zeta_j^d ,
\] 
(36)

where \( \lambda_i \neq \lambda_j \) for \( i \neq j \), and the \( \{ \zeta_j^d \} \) are chi-square distributed with 2 degrees of freedom. Recall that all the received signals from the different diversity branches are independent, but have the same statistics. Hence, combining (28) and (36) results in

\[
\chi = \frac{1}{8N_0} \sum_{d=1}^{D} \sum_{j=1}^{L+1} \lambda_j \zeta_j^d
\]
\[
= \sum_{j=1}^{L+1} \gamma_j ,
\] 
(37)

where the \( \gamma_j \) are chi-square distributed with 2D degrees of freedom.

Let \( \Psi_{\gamma_j}(s) \) denote the characteristic function of random variable \( \gamma_j \). Then

\[
\Psi_{\gamma_j}(s) = \frac{1}{(1 - s\bar{\gamma}_j)^D} \quad , \quad j = 1, \ldots, L + 1
\] 
(38)
where

$$\tilde{\eta}_j = \frac{1}{4N_0} \lambda_j \sigma^2.$$  

(39)

Therefore,

$$\Psi_x(s) = \prod_{j=1}^{L+1} \frac{1}{(1 - s\tilde{\eta}_j)^D}$$

$$= \sum_{j=1}^{L+1} \sum_{d=1}^{D} \frac{A_{jd}}{(1 - s\tilde{\eta}_j)^d},$$  

(40)

where

$$A_{jd} = \frac{1}{(D - d)!(-\tilde{\eta}_j)^{D-d}} \left\{ \frac{d^{D-d}}{d\tilde{\eta}_j^D} (1 - s\tilde{\eta}_j)^D \Psi_x(s) \right\}_{s = \frac{1}{\tilde{\eta}_j}}.$$  

(41)

It follows that

$$f_x(x) = \sum_{j=1}^{L+1} \sum_{d=1}^{D} A_{jd} \frac{1}{(d - 1)!(\tilde{\eta}_j)^d} \chi^{d-1} e^{-x/\tilde{\eta}_j}.$$  

(42)

Let $\hat{P}(\xi)$ denote the integral in (25). Then by using (42) we can write [1]

$$\hat{P}({\xi}) = \sum_{j=1}^{L+1} \sum_{d=1}^{D} A_{jd} \left(1 - \mu_j\right)^d \sum_{m=0}^{d-1} \left(\frac{d - 1 + m}{m}\right) \left(\frac{1 + \mu_j}{2}\right)^m.$$  

(43)

where

$$\mu_j = \sqrt{\frac{\tilde{\eta}_j}{1 + \tilde{\eta}_j}}.$$  

(44)

and

$$\tilde{P}(\xi) = P(\xi) \cdot \hat{P}(\xi).$$  

(45)

To calculate (26), for each error sequence $\xi$, we need to know the values $P(\xi_1), \chi_d$ (hence $\hat{P}(\xi)$), and number of bit errors $w(\xi)$ associated with $\xi$.\footnote{From (26) it seems that we do not need to calculate $w(\xi)$, but this is not the case as will be shown later.} In
general, it is quite complicated to find these values. Here a systematic method is used which employs an error-state transition matrix \([16, 17]\).

To describe this method, it may be easier to start with an error-state diagram. Assume that we have a system with \(\nu\) error-states, \(v_0, v_1, \ldots, v_\nu\). As discussed in [6] a \((\nu + 1)\)-node error-state diagram can be constructed such that the initial node \(v_0\) and the final node \(v_\nu\) are both zero-error states, but the intermediate nodes are non-zero error states. Let \(t_{ij}\) denote the branch-weight associated with the \(v_j\) to \(v_i\) transition. We define \(t_{ij}\) as follows

\[
t_{ij} = p_{ij} Z_1 Z_2 Z_d(i, j)
\]  

(46)

where \(Z_1\) and \(Z_2\) are intermediate (dummy) variables.

\(p_{ij}\) is the fraction of correct symbols \(I_k\) such that the transition from \(v_j\) to \(v_i\) is possible.

\(u_{ij}\) is the number of bit errors associated with the transition from \(v_j\) to \(v_i\).

\(x_d(i, j)\) is defined as

\[
x_d(i, j) = \sum_{l=0}^{N-1} \left| \sum_{m=0}^{L} g_m^{l} \varepsilon_l(i, j) \right|^2
\]

(47)

and, as defined in (17), \(\varepsilon_l(i, j)\) is the difference between the correct and estimated channel input sequence. Here we emphasize that \(x_d(i, j)\) is a function of the \(v_j\) to \(v_i\) transition.

From (46) the path weight for a particular path in the error-state diagram is

\[
\prod_{\{(i, j)\}} p_{ij} Z_1^{u_{ij}} Z_2 x_d(i, j)
\]

(48)

where \(\{(i, j)\}\) denotes the set of state transitions associated with the path under consideration. Note that in the error-state diagram each path starting with the initial node and terminating at the final node represents an error sequence \(\varepsilon \in E\)
(this type of path is called an error sequence). In this case we have

$$P(\xi_1) = \prod_{\{i,j\}} p_{ij}$$  \hspace{1cm} (49)

$$w(\epsilon) = \sum_{\{i,j\}} u_{ij}$$  \hspace{1cm} (50)

and

$$\chi_d = \sum_{\{i,j\}} \hat{\chi}_d(i,j)$$  \hspace{1cm} (51)

by our definition of $t_{ij}$ in (46). Hence, to calculate $P(\xi_1)$, $w(\epsilon)$, and $\chi_d$ for an error sequence $\epsilon \in E$, we only have to calculate its corresponding path weight. These path weights of course can be found by determining the overall transfer function of the error-state diagram. But quite often this approach is very complicated. Instead, as described below, we can use a procedure which employs an error-state transition matrix for these path weight calculations.

As in [16], the error-state transition matrix is defined as

$$T = [t_{ij}] \quad 1 \leq i \leq \nu, \quad 1 \leq j \leq \nu$$  \hspace{1cm} (52)

Let $V_0 = (t_{10}, t_{20}, \cdots, t_{\nu,0}, 0)^T$. Then we know that $T^{H(\epsilon)}V_0$ is the vector whose $m^{th}$ element is the path weight sum of those paths spanning $H(\epsilon) + 1$ epochs and terminating at the error-state $v_m$. So the path weight sum of those error sequences spanning $H(\epsilon) + 1$ epochs is

$$U^T T^{H(\epsilon)} V_0 , \quad U = (0, 0, \cdots, 0, 1)^T ,$$  \hspace{1cm} (53)

the $\nu^{th}$ element of the vector $T^{H(\epsilon)}V_0$. Written explicitly (53) becomes

$$U^T T^{H(\epsilon)} V_0 = \sum_{\epsilon \in G_{H(\epsilon)}} P(\xi_1) Z_1^{w(\epsilon)} Z_2^{\chi_d}$$  \hspace{1cm} (54)

where $G_{H(\epsilon)} \subset E$ is the set of error sequences spanning $H(\epsilon) + 1$ epochs. Note that $\epsilon \in G_{H(\epsilon)}$ does not necessarily have $H(\epsilon) + 1$ bit errors even for the case $r = 1$. This is the reason why we need to keep the number of bit errors $w(\epsilon)$ associated
with \( \varepsilon \). By using (54), \( P(\varepsilon_1) \), \( w(\varepsilon) \), and \( \chi_d \) can be calculated for every \( \varepsilon \in E \) as \( H(\varepsilon) \) becomes larger. After \( P(\varepsilon_1) \), \( w(\varepsilon) \), and \( \chi_d \) have been calculated for all error sequences containing the same number of bit errors, i.e., all \( \varepsilon \in E_i \) for some \( i \), the terms

\[
\sum_{\varepsilon \in E_i} < P(\varepsilon) > 
\]

and

\[
\max_{\varepsilon \in E_i} \{ < P(\varepsilon) > \}
\]

can be formed and (26) can be calculated.

Observe from (26) that to calculate the bound we need to compute an infinite series. This is formidable. Therefore, we usually truncate the series at a point where the remainder can safely be assumed to be small. With this truncation, the whole procedure to calculate (26) can be outlined as in Fig. 5, where \( P_T \) is the threshold for truncating the calculation.

V Examples

Three systems have been analyzed by using the procedure developed in Section IV. Computer simulations have also been conducted for comparison with the analytical results. In our simulations, the discrete-time white noise channel model in Fig. 2 is adopted. The variances of the \( g_i^d \) are assumed to be equal for all \( i \) and \( d \). The tap coefficients are generated by passing independent complex white Gaussian noise through a digital Butterworth filter with a 3-dB cut off frequency equal to 0.2 Hz, the maximum Doppler frequency. This is representative of a typical HF channel [5]. The transmission rate in all cases is assumed to be 2400 symbols/sec.
V-A System 1

A three-tap channel with BPSK modulation is analyzed in this example. Since the system BPSK modulated, $X_k = x_k$ is a real number taking on the values ±1. Fig. 6 is the simplified error-state diagram for this system. Its error-state transition matrix and $V_0$ are

$$T = \begin{bmatrix}
0 & 0 & 0 & t_{14} + t_{14}' & 0 \\
t_{21} & t_{22} & t_{23} & 0 & 0 \\
t_{31} & t_{32} & t_{33} & 0 & 0 \\
t_{41} & t_{42} & t_{43} & 0 & 0 \\
0 & 0 & 0 & t_{54} & 0
\end{bmatrix}$$

(57)

$$V_0 = (t_{10}, 0, 0, 0, 0)^T$$

(58)

Fig. 7 is a comparison between analytical and simulated results. Note that in our analysis, the series in (26) is truncated at the value of i such that $\max_{\epsilon \in \mathcal{E}} \{< \check{P}(\epsilon) > \} < (0.01) \max \{\sum_{\epsilon \in \mathcal{E}_1} \{< \check{P}(\epsilon) > \}, \cdots, \sum_{\epsilon \in \mathcal{E}_R} \{< \check{P}(\epsilon) > \}\}$. $R = 2$ was chosen in this case. Observe that for $D = 1, 2$, the difference between analytical and simulated results is within 1 dB for the whole range of $E_b/N_o$, where

$$\frac{E_b}{N_o} (dB) = 10 \log \frac{\sum_{i=0}^{L} |g_i|^2}{2N_o} \cdot \frac{E[|x_k|^2]}{r}$$

(59)

In general, the bound is tighter for larger $E_b/N_o$. This phenomenon will be elaborated upon further.

V-B System 2

In this example, a system with QPSK modulation and a two-tap channel model is considered. The $X_k = x_k$ are complex taking on values $\frac{1}{\sqrt{2}} e^{j(\pm \frac{\pi}{4})}$ $k = 0, 1, 2, 3$. There are 8 different error symbols in this case, i.e., $\pm \sqrt{2}, \pm j\sqrt{2}, \pm \sqrt{2} \pm j\sqrt{2}$. 

18
Fig. 8 compares the analytical \((R = 2)\) and simulation results. For \(D = 1\) and 
\(E_b/N_o > 12.5\) dB, the difference is less than 1.25 dB. However, for \(E_b/N_o \approx 5\) dB 
there is a 2.5 dB difference. The reason behind this phenomenon is that at low 
\(E_b/N_o\) the term
\[
\sum_{i=1}^{R} i \sum_{\epsilon \in \mathcal{E}_i} < \tilde{P}(\epsilon) >
\]  
(60)
in (26), that represents a union bound on error events having \(R\) or fewer errors, 
is quite loose. This is because these error events in (60) have a more significant 
overlap. For \(D = 2\) the difference between the analytical and simulation results 
is kept within 1 dB, the reason being that for \(D = 2\) the channel is unlikely to 
experience a deep fade on both diversity branches.

In the previous example (System 1), the analytical bound remains uniformly 
tight over a larger range of \(E_b/N_o\). This is because there are only a small number 
of error paths containing \(R\) or fewer errors (in this case there are two 1-error paths 
and four 2-error paths).

V-C System 3

The third example consists of a convolutionally coded system with a QPSK modulator 
with a two tap channel model. A constraint length 3 rate 1/2 convolutional code 
was used having octal generators 5 and 7 [16]. The encoder and signal mapping 
functions are shown in Fig. 9. Fig. 10 is the comparison between analytical \((R = 3)\) 
and simulation results. Observe that for \(D = 1\) and \(E_b/N_o < 10\) dB the difference is 
about 2.5 dB, but for \(E_b/N_o \geq 12.5\) dB the difference is less than 1.5 dB. As noted 
previously the bound is very tight for \(D = 2\). In this case the difference is less than 
0.5 dB.

From the development of (26), the error event probability \(\tilde{P}(\epsilon)\) for a coded 
system has to be averaged over all possible correct sequences. This often invokes 
a computing difficulty especially for systems having a code with a long constraint
length. This can be alleviated by making the following observations. First, note from (39) that different correct sequences correspond to different values of \( \gamma_j \). Then, observe from (43) that \( \hat{P}(\varepsilon) \), which is the average value of \( P(\varepsilon) \) over all possible channel impulse responses, is not very sensitive to changes in the \( \gamma_j \). In fact for \( D = 1 \), \( \hat{P}(\varepsilon) \) has an inverse linear dependency on \( \gamma_j \). Therefore, it is expected that the average error event probability \( < \hat{P}(\varepsilon) > \) can be closely approximated by an error event probability \( \tilde{P}(\varepsilon) \) which corresponds to an arbitrary correct sequence \( \{ I_k \} \). To demonstrate this, Fig. 11 plots (26) by using different values for error event probability. For each set of curves the middle one is calculated by using the average error event probability \( < \hat{P}(\varepsilon) > \). The upper and lower ones use the maximum error event probability

\[
\max \{ \tilde{P}(\varepsilon) \} 
\]

and minimum error event probability

\[
\min \{ \tilde{P}(\varepsilon) \} 
\]

respectively. Since for each error sequence \( \varepsilon \)

\[
\min \{ \tilde{P}(\varepsilon) \} \leq \hat{P}(\varepsilon) \leq \max \{ \tilde{P}(\varepsilon) \}
\]

we note from Fig. 11 that the bound calculated by using an arbitrary \( \tilde{P}(\varepsilon) \) is at most 0.5 dB different from that calculated by using \( < \tilde{P}(\varepsilon) > \). Therefore, in practice an arbitrary \( \tilde{P}(\varepsilon) \) can be used for simplifying the calculation of (26).

VI Conclusions

The MLSE receiver has been analyzed thoroughly by Forney and some others [6-9] for a time-invariant ISI channel. However, for time-variant multipath fading channel, no similar analysis has been found in the literature. By developing a procedure similar to that in [6-9], we have analyzed the performance of a receiver
using a combined MSLE equalizer/decoder and $D$-diversity reception for multipath fading channels. A new upper bound on the bit error probability has been derived and compared to simulation results. These comparisons show that the bound is particularly tight for a system having a moderate signal-to-noise ratio or one that uses diversity.
A Quadratic Form for $\chi_d$

$$
\chi_d = \sum_{i=0}^{N-1} \sum_{j=0}^{L} \left| \sum_{m=0}^{L} g_m d \tilde{e}_{iN-j-m} \right|^2
$$

$$
= \sum_{i=0}^{N-1} \sum_{j=0}^{L} \sum_{m=0}^{L} \sum_{n=0}^{L} g_m d (g_n d)\tilde{e}_{iN-j-m} \tilde{e}_{iN-j-n}^*
$$

Let $g_m d = g_m d R + jg_m d I$ and $\tilde{e}_{iN-j-m} = \tilde{e}_{iN-j-m} R + j\tilde{e}_{iN-j-m} I$. Then

$$
\chi_d = \sum_{i=0}^{N-1} \sum_{j=0}^{L} \sum_{m=0}^{L} (g_m d R + jg_m d I)(g_n d R - jg_n d I)
$$

$$
\times (\tilde{e}_{iN-j-m} R + j\tilde{e}_{iN-j-m} I)(\tilde{e}_{iN-j-n} R - j\tilde{e}_{iN-j-n} I)
$$

$$
= \sum_{i=0}^{N-1} \sum_{j=0}^{L} \sum_{m=0}^{L} \sum_{n=0}^{L} (g_m d g_n d R + g_m d g_n d I)
$$

$$
\times (\tilde{e}_{iN-j-m} R\tilde{e}_{iN-j-n} R + \tilde{e}_{iN-j-m} I\tilde{e}_{iN-j-n} I)
$$

$$
= \hat{h}_d \begin{bmatrix} \hat{B} & 0 \\ 0 & \hat{B} \end{bmatrix} \hat{h}_d^T
$$

where $h = (g_0 R, g_1 R, \ldots, g_L R, g_0 I, g_1 I, \ldots, g_L I)$, and $\hat{B}$ is an $(L+1)^{th}$ order real symmetric matrix with elements

$$
\hat{b}_{mn} = \sum_{i=0}^{N-1} \sum_{j=0}^{L} \left( \tilde{e}_{iN-j-m} R\tilde{e}_{iN-j-n} R + \tilde{e}_{iN-j-m} I\tilde{e}_{iN-j-n} I \right)
$$
References


Fig. 1: Convolutionally Coded Digital Communication System with \( D \)-Diversity Reception
Fig. 2 Simplified System with Discrete-Time White Noise Channel Model
Fig. 3 MLSE Receiver
Fig. 4 A Typical Error-State Trellis Diagram
Fig. 5 A Procedure for Calculating Equation (26)
Fig. 6 Simplified Error-State Diagram for System 1
Fig. 7 Performance BPSK Signaling and MLSE Equalization over a Three-tap Multipath-Fading Channel

Bit Error Probability (Log)

Bit Energy-to-Noise Ratio $E_b/N_0$

- Analytical
- Simulated

$D = 1$

$D = 2$
Fig. 8 Performance QPSK Signaling and MLSE Equalization over a Two-tap Multipath-Fading Channel
Fig. 9 Convolutional Encoder and Signal Mapping

<table>
<thead>
<tr>
<th>$u_1$</th>
<th>$u_2$</th>
<th>Mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$\frac{1}{\sqrt{2}} (1+j)$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$\frac{1}{\sqrt{2}} (1-j)$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>$\frac{1}{\sqrt{2}} (-1-j)$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$\frac{1}{\sqrt{2}} (-1+j)$</td>
</tr>
</tbody>
</table>
Fig. 10 Performance of a Convolutional Code with QPSK Signaling and MLSE Equalization/Decoding over a Two-tap Multipath-Fading Channel
Fig. 11 Performance of Various Upper Bounds for a Convolutional Code with QPSK Signaling and MLSE Equalization/Decoding over a Two-tap Multipath-Fading Channel
CELLULAR RADIO SPECTRUM STUDY

Final Report
OCA Project No. E-21-620

Prepared for
BellSouth Enterprises, Inc.

by

Aubrey M. Bush, Gordon L. Stüber, and Stephen B. Wicker

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia  30332

Sponsor Technical Contact

Neale Hightower

BellSouth Enterprises, Inc.
1155 Peachtree Street, NE
Atlanta, Georgia  30367

June 1988
Contents

1 Introduction 1

2 Channel Characterization 3

2.1 Introduction .............................................. 3
2.2 Multipath-Fading ........................................... 3
   2.2.1 Envelope and Phase Characteristics ................. 5
   2.2.2 Crossing Rate and Fade Duration ..................... 7
2.3 Shadowing .................................................. 11
2.4 Path Loss .................................................. 11
2.5 Concluding Remarks ....................................... 12

3 Voice Coding Techniques 14

3.1 Introduction ............................................... 14
3.2 Classification of Voice Coding Techniques ............... 15
3.3 Waveform Coders .......................................... 16
3.4 Sub-Band Coders .......................................... 16
3.5 Analysis-Synthesis Coders ................................ 18
3.6 Analysis-by-Synthesis Voice Coders ....................... 22
3.7 Quality Testing of Voice Coders ......................... 26
3.8 Analysis-by-Synthesis Voice Coders in Cellular Radio ... 26
3.9 Conclusions and Recommendations ......................... 26

4 High Capacity Mobile Phone Systems with Analog Modulation Formats 32

4.1 FM based Systems ......................................... 32
4.2 Amplitude Modulation Formats ............................ 40
4.3 Technology Issues ......................................... 40
4.4 Compatibility Issues ....................................... 41

(ii)
Contents

5 Digital Modulation Schemes

5.1 Constant Envelope Modulation Formats .............................................. 44
   5.1.1 Phase Shift Keying and Frequency Shift Keying ........................ 45
   5.1.2 QPSK and Offset-QPSK ....................................................... 47
   5.1.3 MSK ................................................................. 51
   5.1.4 Gaussian MSK .............................................................. 53
   5.1.5 Tamed Frequency Modulation ............................................... 56
   5.1.6 Generalized Tamed Frequency Modulation ............................... 57
5.2 Nonconstant Envelope Modulation Formats ....................................... 60
5.3 Error Correcting Codes ............................................................... 62
5.4 Modem Technology ........................................................................... 63
   5.4.1 MSK Modems ....................................................................... 64
   5.4.2 TFM modems ....................................................................... 67
5.5 Other Advantages of the Digital Approach ...................................... 67

6 Spread-Spectrum and Time Division Multiple Access .......................... 73

6.1 Introduction ....................................................................................... 73
6.2 History of Spread-Spectrum in Land Mobile Radio ............................ 75
6.3 Characteristics and Advantages of Spread-Spectrum Mobile Radio .... 78
6.4 Particular Systems Using Spread-Spectrum ...................................... 82
   6.4.1 MATS-D System .................................................................... 82
   6.4.2 SFH900 System ..................................................................... 83
   6.4.3 The CD 900 System ............................................................... 83
   6.4.4 The DMS 90 System .............................................................. 86
   6.4.5 The GSM-System .................................................................. 87
6.5 Comparison of Spectral Efficiency .................................................... 88
6.6 Unresolved Issues .............................................................................. 89
6.7 Technology Requirements ............................................................... 90
6.8 Concluding Remarks .......................................................................... 90
1 Introduction

The use of cellular mobile radio telephones has achieved a high level of user acceptance with the current analog FM technology. The trade journals, the popular magazines, the business publications, and the theoretical communications journals all clearly reflect an extremely high level of interest in cellular systems.

In this report we survey the current and pending technologies for expanding the capacity of mobile cellular systems within the existing frequency band allocations and improving the quality and versatility of the service.

In Chapter 2, we treat the channel over which the cellular system must operate. Based on the theoretical and experimental evidence reported in the current technical literature, a mathematical model is developed for the channel environment.

Chapter 3 treats the voice coding problem. The voice coding alternatives are discussed from an intuitive point of view and the alternatives which look promising for the cellular environment are singled out. Conclusions and recommendations for a choice of the first and second generation voice coder properties are given.

Chapter 4 considers the possibility of expanding the capacity of the cellular system using analog technologies. There are several possibilities and the potential of these techniques is given. Some previous misconceptions are discussed.

Chapter 5 treats the selection of a modulation format for digital transmission over the cellular channel. The various alternatives are motivated and compared from the point of view of spectral occupancy and noise immunity, and some practical implementation issues are discussed.

Chapter 6 discusses the use of spread spectrum and time division multiplexing techniques on the cellular channel with examples taken from the technical literature describing the systems which have been proposed for a European standard.

Chapter 7 describes the spectral efficiency problem for cellular systems. This is a more difficult issue than for noncellular systems because of the frequency reuse factor. That is, bits/second/hertz is not the real measure of efficiency, but rather $Erlangs/Mhz/Km^2$ or $voicechannels/Mhz/Km^2$, which are discussed here. System comparisons of several recently recommended system formats are compared.
Chapter 8 gives some overall comparisons and additional considerations as well as overall system recommendations.

The final design choices will involve many tradeoffs and issues which are as much market oriented as engineering oriented. For example the choice of the voice encoding technique will depend on a judgment as to what kind of data handling and privacy features are to be included in the final system configuration.

The issue of switching systems for cellular connections is mentioned here only in connection with tandeming of voice coders. The issue as to whether to use common channel control or to use a separate control channel is not treated.

Bibliographic information is included at the end of each chapter.
2 Channel Characterization

2.1 Introduction

In many radio channels signals reflect off the surface of water, buildings, trees etc., causing multiple signal terms at the receiver. These multipath-fading channels are usually modeled by a transversal filter with randomly time-varying tap coefficients together with noise and interference. Extensive empirical studies have been undertaken to verify the theoretical models and to supply model parameters that are useful for system designs. There are many excellent tutorial references on the characterization of the mobile radio environment. Those by Stein [1], Proakis [2], Jakes [3], and Lee [4], [5] are among the best. This section provides a brief description of the most important aspects of the mobile radio environment; they must be considered when designing high capacity cellular systems.

Propagation at UHF/VHF frequencies used in land mobile radio is largely influenced by three nearly independent factors, path loss, shadowing, and multipath-fading. Each of these phenomena is the result of a different underlying physical principle, and each must be accounted for when designing a cellular system. In general, multipath-fading is short-term in nature and determines the required carrier-to-noise and carrier-to-interference ratios. This, in turn, is reflected in the cell size and cell reuse factor. Shadowing and path loss, on the other hand, are longer term variations that determine the availability of a channel.

2.2 Multipath-Fading

Multipath-fading is caused by multiple receptions of a transmitted waveform due to scattering in the medium. Multipath-fading channels are characterized by two kinds of spreading: Doppler spread $B_d$ (spreading in frequency), and multipath spread $T_m$ (spreading in time). From these define $\Delta f_c = 1/T_m$ as the “coherence bandwidth” and $\Delta t_c = 1/B_d$ as the “coherence time” of the channel. Roughly, two sinusoids having a frequency separation greater than $\Delta f_c$ will be affected independently by the channel. Similarly, samples of the channel output having a time separation greater than $\Delta t_c$ will be affected independently by the channel.

In digital systems, the medium characteristics are based on the signal duration $T$
Table 3-1

Typical Delay Spreads

<table>
<thead>
<tr>
<th>Environment</th>
<th>Delay-Spread $T_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Area</td>
<td>$&lt; 2 \mu s$</td>
</tr>
<tr>
<td>Suburban Area</td>
<td>$0.5 \mu s$</td>
</tr>
<tr>
<td>Urban Area</td>
<td>$3 - 10 \mu s$</td>
</tr>
</tbody>
</table>

relative to $\Delta t_c$, and signal bandwidth $W$ relative to $\Delta f_c$. Consequently, multipath-fading channels can be further classified as nondispersive ($T \ll \Delta t_c$, $W \ll \Delta f_c$), time-dispersive ($W \ll \Delta f_c$, $T > \Delta t_c$), frequency-dispersive ($T \ll \Delta t_c$, $W > \Delta f_c$), and doubly-dispersive ($T > \Delta t_c$, $W > \Delta f_c$). In many practical systems, $W$ and $T$ are chosen such that the channel is nondispersive. Sometimes, such a channel is referred to as being underspread.

The time dispersive channel characteristics are described by a “multipath intensity profile”. For mobile radio, the following exponential multipath intensity profile is commonly used;

$$\phi(r) = \frac{1}{T_m} \exp \left( - \frac{r}{T_m} \right) .$$

The multipath intensity profile $\phi(r)$ is just the average power output of the channel as a function of the time delay $r$. Therefore, the channel will exhibit intersymbol interference (ISI) unless $T \gg T_m$. Typical values of $T_m$ are included in Table 3.1. These values are valid for any operating frequency above 30 $MHz$.

The characteristics of coherence bandwidth in VHF and UHF urban land mobile radio channels have been experimentally studied by many authors. For narrow-band single channel per carrier (SCPC) digital land mobile radio systems with bit rates and channel spacings of approximately 16$Kb/s$ and 25 $KHz$, respectively, the effects of frequency-selective fading can be ignored. However, it must be taken into account with higher speed digital transmission systems such as direct-sequence (DS) spread-spectrum and time-division multiple-access (TDMA) systems. For example, a channel with a coherence bandwidth of 100 $KHz$ may be thought to limit the baud rate to 100 $Kbaud/s$, and the number
of channels per carrier to less than 6 if a 16 Kbps speech coder is used along with binary modulation, because of the effect of ISI. However, this number can be increased significantly by using adaptive equalizers.

Doppler spread due to vehicle motion causes time variations in the envelope of the received signal. For digital systems, the coherence time of the channel is inversely proportional to the Doppler bandwidth, and hence vehicle speed, so that faster vehicle speeds result in faster fading. This is discussed in more detail in the sequel.

A typical model of a multipath-fading channel is shown in Fig. 2.1, basically consisting of a transversal filter. The tap gains are modeled as complex Gaussian random processes. If the processes are zero mean, the magnitudes of the tap gains have Rayleigh distributions. Otherwise, the magnitudes have Rice distributions. Quite often, a simplified model is used when designing a system that consists of only two paths. Since the mobile radio channel has nonlinearities, simulations are required for system analysis. The utility of the two-path model is to reduce the resultant time in running the simulations, while providing representative results while the systems are being developed. More precise models may be used in further stages of the system development.

2.2.1 Envelope and Phase Characteristics

For a fading channel, the received signal is the sum of many independently scattered components which justifies the assumption that it is a Gaussian random process. Therefore, the envelope of the received signal at any time is Rayleigh distributed while the phase is uniformly distributed. If, however, there also exists a single dominant unfaded component in the received signal, the envelope statistics are Rician. A system designed under the assumption of Rayleigh fading will always perform satisfactorily with Rician fading and, for this reason, most studies consider Rayleigh fading only.

Fast Rayleigh envelope fading is accompanied by fast phase changes that introduce random FM noise on the received carrier. The power spectral density of random FM noise extends to about twice the maximum Doppler frequency, i.e., \( 2V/\lambda \) Hz, where \( V \) is the vehicle velocity and \( \lambda \) is the wavelength. For example, at 36 Kmph and 900 MHz the random FM extends to about 75 Hz. The random FM noise usually presents no problem
Fig. 2-1 Tapped Delay Line Model of a Multipath-fading Channel, from [2]
in analog FM systems because it can be filtered out with negligible effect on the speech waveform. For digital systems, it is necessary to use signaling waveforms that do not have energy concentrated in the random FM range. That is, the equivalent baseband signals must have low spectral concentrations around zero frequency.

2.2.2 Crossing Rate and Fade Duration

Crossing rates, average fade durations, and fade duration distributions are all second-order statistics. These statistics are important for determining the effect of burst errors on digital communication systems. The level crossing rate and average fade duration for a Rayleigh fading channel is known, while the fade duration distribution is unavailable in closed form. The level crossing rate (LCR) is just the rate at which the received signal power crosses a threshold relative to the average received signal power. The LCR increases with the vehicle velocity. Its effect is to limit the improvement in bit error rate that would result from increasing the carrier-to-noise ratio.

Fig. 2.2 shows the level crossing rates for $E_z$, $H_x$, and $H_y$ mobile radio signals assuming Rayleigh fading. $E_z$ is the main parameter of interest. As an example, assume a signal at 850 MHz received by a dipole at a mobile unit traveling 15 mph. We wish to find the expected level crossing rate at 10 dB below the average received power level. In this case $V = 22 \text{ ft/s}$ and $\lambda = 1.157 \text{ ft}$, so that $n_o = 47$. Then $n_R = 0.284$ and $n(R) = 13.35$ crossing/s.

Fig 2.3 shows the average duration of Rayleigh fades as a function of the received signal level with respect to the average received signal level. Typically, the fades tend to be rather short in duration relative to the frequency of their occurrence. This is why two-channel antenna diversity offers exceptional improvements in the performance. Also, deeper fades tend to have a shorter duration.

The distribution of the duration of fades $F_r(u, R)$ has been calculated by Rice for Rayleigh fading and is shown in Fig. 2.4, where $u = r/\bar{t}$. This diagram indicates the probability that $R(t) < R$ for an interval lasting longer than $r$. The average duration of the fades is $\bar{t}$, which can be obtained from Fig 2.3, and $R$ is the envelope with respect to its rms value.
Fig. 2-2 Level-Crossing Rates for $E_x$, $H_x$, and $H_y$ Mobile Radio Signals, from [5]
Fig. 2-3 Average Duration of Fades for $E_z$, $H_x$, and $H_y$ Mobile Radio Signals, from [5]
Fig. 2-4 Probability $F_r(u,R)$ that $R(t) < R$ for an Interval Greater than $r$, from [5]
2.3 Shadowing

Even when the multipath-fading is averaged out, nonselective shadowing still remains. Shadowing is caused by the effects of the terrain and the man-made environment. It imposes a slowly changing average on the Rayleigh (or Rician) fading statistics. Shadowing is usually modeled as having a log-normal distribution with a standard deviation of about 8 dB. That is, the average received signal power in dB's has a Gaussian distribution. This distribution closely fits experimental data in an urban environment. The 8 dB spread remains constant for almost all distances between base station and mobile.

2.4 Path Loss

Path loss is the average value of the log-normal shadowing. Path loss is actually necessary for cellular systems. This is because rapid attenuation with distance allows small frequency reuse distances. It depends upon the distance between the base station and mobile. There are a multitude of technical reports that are concerned with path loss prediction methods for UHF/VHF land mobile radio in flat urban, suburban and open, and hilly terrains. The empirical model by Okumura is probably the simplest to use, and can distinguish man-made structures. There are other path loss models and the reader is referred to the references by Lee [4], [5] for these. The empirical data for Okumura's model was collected in Tokyo. Be cautioned, however, that the path loss for Japanese suburban areas does not match North American suburban areas very well. The latter are more like the quasi-open areas in Japan, because the buildings are more spaced out.

The model is in terms of the carrier frequency $150 \leq f_c \leq 1500$ (in MHz), base-station antenna height $30 \leq h_b \leq 300$ (in m), and the mobile-station antenna height $1 \leq h_m \leq 10$ (in m). The empirical formula for the path loss is a function of the distance $1 \leq r \leq 20$ (in km) between the base and mobile station and is accurate to within 1 dB for distances up to 20 km. With this model, the path loss is

$$L_p = \begin{cases} A + B \log_{10}(r) & \text{for urban area} \\ A + B \log_{10}(r) - C & \text{for suburban area} \\ A + B \log_{10}(r) - D & \text{for open area} \end{cases}$$

11
where

\[ A = 69.55 + 26.16 \log_{10}(f_c) - 13.82 \log_{10}(h_b) - a(h_m) \]
\[ B = 44.9 - 6.55 \log_{10}(h_b) \]
\[ C = 5.4 + 2 (\log_{10} (f_c/28))^2 \]
\[ D = 40.94 + 4.78 (\log_{10}(f_c))^2 - 19.33 \log_{10}(f_c) \]

and

\[ a(h_m) = \begin{cases} 
(1.1 \log_{10}(f_c) - 0.7)h_m - (1.56 \log_{10}(f_c) - 0.8) & \text{for medium or small city} \\
8.28(\log_{10}(1.54h_m))^2 - 1.1 & \text{for } f_c \geq 200 \text{ MHz} \\
3.2(\log_{10}(11.75h_m))^2 - 4.97 & \text{for } f_c \leq 400 \text{ MHz} 
\end{cases} \]

2.5 Concluding Remarks

In this chapter, we have discussed the channel characteristics that are prevalent in a mobile radio environment, along with their physical causes and effects. Discussion of co-channel interference models is considered in section 7.3. Characterization of the channel in a cellular mobile radio environment is a well understood topic, but still claims much attention in the current literature. Almost every conference dealing with cellular mobile radio will devote one or more sessions to channel characterization. The reason is not to provide statistical models for describing the mobile radio channel, although sometimes a refined model is presented, but rather to gather empirical measurements and to determine parameters of existing models to fit such measurements. These studies commonly address both the multipath-fading and path loss characteristics.

References


3 Voice Coding Techniques

3.1 Introduction

The future of cellular mobile telephone systems depends on the ability of the technical community to increase the capacity of the system within the allocated bandwidths and to bring the system into some reasonable level of compatibility with the many new services which are available or are becoming available in the industry. Most of the new services, including the new generation of FAX equipment, will require digital formats. The expansion of the capacity of the existing cellular system and the upgrading of the system to include the new services and to interface with the coming ISDN service clearly dictate that the next generation of the cellular system will be a digital system rather than an analog system, even if an intermediate term expansion using analog technology is possible or desirable.

The choice of a voice coding strategy involves choosing a digitization technique which is efficient enough in its representation of the speech signal to allow transmission in less bandwidth than the analog system and at the same time is robust enough to allow transmission through the fading dispersive environment of the mobile radio channel. In addition the technique must allow graceful evolution both from the present analog system and into future generations of more efficient digitization techniques which are likely to be developed. In addition, the voice coding system must, at least for the present, interface with the existing switching systems industrywide; this means that tandem connections of the mobile voice coder with the PCM and possibly ADPCM voice coders followed by a mobile voice codec at the other end of the complete connection must give acceptable performance. A final difficulty is the handling of data modems and data signals through the mobile radio environment. It seems inevitable that special techniques will be required for the handling of data through a mobile connection.

In this section we will highlight the voice coding alternatives and attempt to place the available techniques in perspective. Our goal is to identify issues and questions and to focus on the design choices rather than to specify in detail a final choice. The final choice will necessarily involve overall system design issues and constraints. We will however compile a list of issues we feel are key in the final choice and which could serve as a checklist.
3.2 Classification of Voice Coding Techniques

The techniques available for the digitization of voice signals for transmission in a digital format followed by reconstruction for the user are many and varied. The most basic techniques are called Waveform Coding Techniques and the most advanced are Analysis-by-Synthesis Techniques listed by a variety of acronyms which we will describe shortly.

The Waveform Coding Techniques are PCM, Pulse Code Modulation, with some type of amplitude companding; ADPCM, Adaptive Differential Pulse Code Modulation; CVSD, Continuous Variable Slope Delta Modulation; and other variants of Delta Modulation. These techniques are called Waveform Techniques because they take very little or no advantage of the fact that the signal that they are digitizing is a voice signal. The source or voice signal could be any other signal as well as a voice signal. The digitization is on the basis of the waveform itself rather than a model of how the signal may have been generated.

The next level of techniques recognizes that the signal waveform is a speech signal and takes extensive advantage of this fact in its operation. These techniques are not based completely on a model of the physical mechanism of the production of human speech but do make use of the properties of the signal itself. The most successful of these techniques is the SBC, Sub-Band Coder. Although this technique basically follows the philosophy of the channel bank vocoder, it has been the subject of intense research in the past decade and has seen the application of very sophisticated digital filtering developments and a variety of quantization and coding schemes.

The more advanced techniques were originally called vocoder techniques but over the past decade have become known as Analysis-Synthesis Techniques. The most recent developments are called Analysis-by-Synthesis Techniques. These techniques model the speech signal as some excitation signal and a vocal tract filter model which processes the excitation. The vocal tract filter is developed by application of some form of LPC, Linear Predictive Coding, and a set of parameters chosen for the efficient transmission of the filter description. The excitation is handled by one of a variety of methods, leading to the names APC, Adaptive Predictive Coder; RELP, Residual Excited Linear Predictive Coders;
MPLPC, Multiple Pulse Excited Linear Predictive Coders; RPLPC, Regular Pulse Excited Linear Predictive Coders; CELP, Code Excited Linear Predictive Coders; SEV, Self Excited Voice Coders; and variants of this terminology chosen by each author who contributes to the field.

The bit rate requirements of the waveform techniques range from 64 kbits/s for the PCM, and 32 kbits/s for the ADPCM, to 16 kbits/s for the CVSD; quality, tandeming, and data modem handling capability decreases as the bit rate decreases. The SBC bit requirements range from 12 kbits/s to 16 kbits/s; tandeming has not been studied extensively for these coders and it is doubtful but not impossible that they will handle data modems acceptably. The SBC combined with appropriate channel protection will perform quite well on the mobile channel and seems to degrade fairly gracefully as the channel quality degrades. The analysis-synthesis techniques require bit rates from 1 kbit/s to 8 kbits/s and quality is better at higher bit rates; these techniques require a lower channel error rate, will probably not tandem very well, and do not tend to degrade gracefully in general. Substantial improvements in the performance and complexity of the analysis-synthesis techniques are being made currently.

3.3 Waveform Coders

The various waveform coders mentioned above are described in detail in the literature; the most convenient compilation of this material is a collection of papers edited by Jayant [1]. The PCM and ADPCM coders are not candidates because the bit rate required is too high to allow for any improvement in the cellular system capacity over the present analog system. The delta modulation systems and the CVSD systems are almost within the range of possibility as to bit rate but are still on the very edge of the high bit rate possibility and do not offer the voice quality of the more recently developed systems.

3.4 Sub-Band Coders

Sub-Band coders were introduced by Crochiere [2] and have been the subject of much recent research. Smith [3] and Barnwell [4] developed the quadrature mirror digital filters which serve to divide the speech into narrow bands. The most significant bands are selected,
Figure 3-1. Basic Sub-Band Coder Configuration. [6].

and the outputs of the corresponding filters are coded for transmission along with side information which identifies the transmitted bands and the gains of the individual filter outputs which are transmitted. Sub-Band coders tailored specifically for the cellular mobile radio environment have been described and demonstrated by Motorola [5], by A.T.&T. [6], and in Europe [7].

The basic configuration of the Sub-Band coder is shown in Figure 3-1 below.

Texas Instruments demonstrated a Sub-Band coder on a single board operating in real time at ICC'88 in June 1988. The algorithm was developed by Barnwell at Georgia Tech. This Sub-Band coder will very likely be available in a single chip implementation in the near future. AT&T also has demonstrated and implemented a Sub- Band coder; a tape of the coder signal played over a Rayleigh fading simulated channel was played at ICC'88. Both of these coders have a transmitted data rate of 16 kbits/s when the channel error protection is included.

The AT&T version uses 12 kbits/s for the speech and 4 kbits/s for the error protection; it uses a frame oriented K=5 family of punctured convolutional codes to provide a variable level of protection for the side information and the filter output representations, with a block oriented Viterbi decoder. No further error detection and interpolation or mut-
ing is done in this system. Interleaving bit by bit of two frames is done to protect against bursts of errors. It is said to degrade gracefully.

Detailed information as to the error control techniques or options in the TI system was not available.

A Sub-Band coder using \( K=6 \) self orthogonal convolutional coding and threshold decoding combined with error detection and single frame repetition or multiple frame muting was reported in [5]. Although the self orthogonal coding is simpler to decode, it does not offer the level of error protection that the Odenwalder codes used by the AT&T system offer; however, the added error detection and its application is not used in the AT&T system.

All of the above Sub-Band coders have been implemented using one or more of the special digital signal processor (DSP) chips currently available. The AT&T system uses the AT&T DSP32 chip, the TI version uses a TI chip, and the Motorola system uses NEC DSP chips supervised by Motorola microprocessor chips.

### 3.5 Analysis-Synthesis Coders

A good discussion of this and the Analysis-by-Synthesis found in the initial sections of Rose [8]. The following is an abbreviated and adapted summary from Rose. Early versions of the class of analysis-synthesis voice coders make use of the model of the human speech production process shown in Figure 3-2.

The resonator or filter is a linear filter model of the vocal tract. The excitation for the filter is either a periodic impulse train with the impulses at the pitch period, for voiced speech segment, or a "white" noise excitation signal for unvoiced speech segments. In the analysis phase the filter must be characterized, the speech must be classified voiced or unvoiced, and, if voiced, the pitch period must be determined. In addition a gain parameter must be determined. The excitation parameters and a robust description of the filter must be digitized and transmitted. The receiver uses the parameters to synthesize the speech at the receiver. Under ideal conditions these coders will operate at bit rates in the range of 2000 to 10000 bits/s. However, the synthesized speech quality is very sensitive to the pitch period and voiced/unvoiced, V/UV, decisions. Good V/UV determination and pitch period determination is possible under ideal laboratory conditions, but is not possible under
realistic operation with background noise and multiple speakers. The coders are quite vulnerable to transmission errors.

The analysis and the synthesis is treated as a stationary problem; speech is clearly nonstationary. Thus it is necessary to divide the speech into segments or windows within which the speech may be considered as stationary. Window lengths are typically on the order of 20 ms. The choice of the window and the degree of overlap of windows is specific to the particular version of the coder.

The particular representation of the filter parameters also varies with the specific system. Some representations are robust and others are extremely sensitive to quantization and transmission errors.

The determination of the filter parameters is a problem shared with later analysis-synthesis and analysis-by-synthesis voice coders.

The APC analysis-synthesis voice coder is represented, in its more recent form, by the block diagram of Figure 3-3.

This type coder was originally proposed by Atal [9]. The long term predictor portion of the system is a more recent feature. The short term predictor is an FIR digital filter with $P$ parameters; $P$ is typically equal to 10. This predictor attempts to predict the current
speech sample from the previous speech samples and subtracts its prediction from the speech to leave a residual error. The long term predictor then repeats the process by trying to predict the current value or the residual signal; the long term predictor is an FIR filter with typically 1, 2, or 3 parameters. In addition the long term predictor involves a delay of from 5 to 20 ms. The residual signal has a regular structure which shows correlations over delays much greater than a pitch period and the long term predictor attempts to remove this correlation to give a whitened residual output. The whitened residual signal and the parameters of the two predictors are quantized for digital transmission. The synthesis of the speech signal from this information is carried out as shown in Figure 3-4.

In a more recent residual coding techniques, the residual signal is low pass filtered, down sampled, and differentially coded for transmission. At the receiver a nonlinearity of some sort, for example a full wave rectifier, is used to regenerate high frequency harmonics. This results in a coder known as a RELP or residually excited linear predictor. [16]. RELP coders operate in the range from 6400 to 9600 bits/s. Of course the information lost by low pass filtering the residual signal is never recovered by the nonlinearity and quality is limited by this feature.

The determination of the short term predictor polynomial \( A(z) \) is performed as shown
in Figure 3-5.

The premphasis is a first order high pass filtering operation serving to remove the spectral tilt usually present and flatten the spectrum to permit greater numerical stability in the analysis procedures. The windowing may be done in several ways, generally in about 20 ms segments. A computationally efficient windowing technique is the recursive windowing procedure described by Barnwell [10]. This technique is illustrated in Figure 3-6.

The Toeplitz autocorrelation matrix is then used in the Levinson-Durbin algorithm to determine the P coefficients of the short term predictor. Other methods including Burg’s method [11] may also be used to find the filter coefficients in several forms of filter representations, for example the lattice representation.

Representations of the filter suitable for transmission include partial correlation coefficients, roots of the predictor polynomial, log area ratios, and line spectrum pair (LSP) representations. [12] [13] [14] [15]. The LSP technique offers quality, bit rate, and complexity advantages. The parameters of the long term predictor are typically determined by the use of the covariance method. Figure 3-7 shows the basic idea of the autocorrelation and the covariance methods of determining the predictor coefficients.

The long term predictor performs the same function as the V/UV and pitch detector
features of the pitch excited voice coder. The long term filter when excited by the whitened residual generates a pitch pulse type excitation when appropriate, a noise type excitation when appropriate, and a combination when neither of the V or UV options is a very good model. It is thus a more realistic model of the speech production process and gives better results.

3.6 Analysis-by-Synthesis Voice Coders

The bit rate required to transmit the residual signal in the Analysis-Synthesis coders is large, limiting the bit rate to which the operation of these systems can be lowered. The most recent developments are in the way that an excitation signal to the cascade of the two filters at the synthesis stage, Figure 3-4, is represented. This new class of coders is called Analysis-by-Synthesis to distinguish it from the the earlier coders. The operation of this class of coders is as shown in Figure 3-8.

The residual signal is not used directly. Instead, the long term and short term filters found as above are used, and an excitation signal is chosen from a limited class of excitation signals to minimize the weighted error between the new or synthetic speech and the actual speech signal. The coder is classified according to the class of input signals over which the error minimization is carried out. The restricted class of excitations can be communicated...
Figure 3-6. The Recursive Windowing Technique.
(a) The Recursive Window (b) The Sampled Speech
(c) The Reversed and Delayed Window (d) The Windowed Speech.
Figure 3-7

(a) The Autocorrelation Method (b) The Covariance Method
to the receiver much more easily than can the actual residual signal.

The Multiple Pulse Excited Linear Predictive Coder, MPLPC, also known by other similar acronyms, forms its excitation signal by selecting \( F \) impulses for each frame with the positions and weights of each impulse determined by the minimization and sent to the receiver. [17]

The Regular Pulse Excited Linear Predictive Coder, RPLPC, also frequently given other acronyms, uses regularly spaced impulses, spaced at \( M \) sample intervals, with only the weights of the pulses adjusted in the minimization and transmitted. [18]

The Code Excited Linear Predictive Coder, CELPC, again frequently given other similar acronyms, uses as its excitation family a limited, indexed ensemble of sample functions drawn from a Gaussian random ensemble. This is motivated by the fact that the whitened residual signal has been heavily filtered and spectrally flattened and thus is susceptible to a Gaussian model. If both the transmitter and the receiver have a code book containing the ensemble of possible excitation signals, then the index and gain is the only feature of the excitation which must be communicated. [19]

The Self-Excited-Vocoder, SEV, draws its excitation ensemble from the past of its
own whitened residual signal rather than an independent Gaussian process [20]

3.7 Quality Testing of Voice Coders

The evaluation of the quality of the synthetic speech signal produced by a voice coder is a matter of considerable difficulty. Because the speech is intended for a human listener, simple objective testing such as signal-to-noise ratio measurement is not very effective in comparing coders unless the coders have a very wide difference in performance. Extensive work has been done to develop objective tests [21]. It is necessary to perform careful subjective testing with human subjects in order to differentiate between coders which are at all close in performance. The subjective tests which have been used most are the Mean Opinion Score, MOS, and the Paired Acceptability Rating Method, PARM. [22].

3.8 Analysis-by-Synthesis Voice Coders in Cellular Radio

Evaluation of six medium bit rate voice coders intended for application in cellular radio systems was reported in February 1988. [7]. The six coders tested included four versions of SBC's and an MPE-LTP and an RPE-LTP from the analysis-by-synthesis class. All six operated at a final protected bit rate of 16 kbits/s. The SBC coders differed in the type of error protection provided, and in the number and type of filters used and the quantization technique. None of the coders include further error detection and final defense such as interpolation and muting. Following extensive subjective testing using the MOS test, the recommendation was for a new RPE-LTP coder operating at 13 kbits/s error-free; this coder is also referenced as the choice of the Pan-European group GSM. [24].

Kroon and DePrettere [23] indicate good performance of MPLPC and RPLPC coders down to 10 kbits/s with the CELP coders performing well down to 8 kbits/s.

3.9 Conclusions and Recommendations

The voice coder for a cellular mobile radio telephone system must have the following features:

- The bit rate of the basic voice coder should be as low as possible. Bit rates of 12 to 13 kbits/s are clearly possible now. Bit rates of 4800 to 8000 bits/s may become
possible in the future.

- The fading dispersive channel environment of the cellular channel is rather hostile and bit error probabilities of 0.01 to 0.03 or even higher must be tolerated as a basic raw channel error performance. The channel will also be subject to error bursts as well as random errors.

- The more advanced voice coders will not be useful if it is necessary to support voice-band data modem signals through the voice coder. Even the subband coders if optimized for speech quality will not handle data modem signals.

- Forward error correction will be required to protect the coded speech. Since not all of the parameters require equal protection, it is best to integrate the error protection and the voice coder.

- Since the coded speech will be subject to long error bursts under some conditions, it will be necessary to provide some error detection with interpolation and muting as a final defense against highly objectional failures of the coder. This will be particularly true in the case of the lower bit rate coders which do not degrade gracefully.

- The issue of voice privacy requires some study. If the coder is chosen to give a good combination of channel and source coding in order to operate over a channel with an error rate of 0.01 to 0.03 as the digital transmission literature indicates, then the issue of encryption is not simple since that level of errors will likely destroy any encryption scheme which simply operates on the digital data stream. It is generally essential that the encryption system see no errors; hence it must see the corrected channel rate rather the raw channel error rate. One possibility is to embed the encryption into the unified voice-coder/channel-coder error correction scheme. The alternative is to use a very low bit rate voice coder with a uniform channel coding scheme and accept the reduced level of quality which that may entail. A separate voice system for secure channels cannot be ruled out at this point.

- Special measures will need to be provided for data handling on the cellular radio telephone. The best technique is probably to sacrifice data rate to accuracy for the
data mode of operation. For example if the speech coder with its error protection is at 16 kbits/s with a raw channel error rate of 0.03, then the data mode might be at 8 or even perhaps 9.6 kbits/s with a rate 1/2 or slightly higher rate forward error correction code present to reduce the apparent error rate to acceptable levels. Even then the data mode user may need to provide outside ARQ error control.

- In the evolution of the cellular system a succession of voice coders with lower bit rate requirements may be anticipated. This is the case now if quality is sacrificed as bit rate decreases. Hence not just one step but at least two steps of evolution should be planned for in the digitization process.

- The time delay in the speech coder, including its error protection, must be small enough that the end-to-end delay is acceptable.

- As long as the cellular system is required to function within the existing network, tandeming of the cellular voice coder and PCM and ADPCM coders must result in acceptable speech quality. This aspect seems not to be reported on in the current literature, and will not be an issue in the future digital network format.
References


4 High Capacity Mobile Phone Systems with Analog Modulation Formats

The goal of this study is to highlight methods that show the greatest promise for increasing the spectral efficiency of the current mobile phone service. Spectral efficiency shall be defined as the maximum number of calls that can be handled over a given geographical area. There are a large number of candidates for consideration. They can be divided up according to the degree of modification of the current system that will be entailed. In this first section some basic analog schemes will be discussed. Several approaches for increasing capacity while retaining the current modulation format will be examined. Another promising analog format, amplitude companded single sideband, will also be considered. In later sections a variety of digital modulation based schemes will be examined.

4.1 FM based Systems

A new FM based system will have the advantage of a high degree of compatibility with the current system. If existing hardware and software can be used, a major cost savings will be entailed. All of this is clear, but one must be careful that one is not merely postponing a capacity crisis. A significant increase in capacity must be achieved if a system change is to be cost effective in the long run. For the purposes of this report a "significant" increase will be one that increases capacity by a factor of five or more. There do not appear to be any such schemes that retain an FM modulation format with no cell splitting. Without cell splitting a factor of 1.6 is about the best one can accomplish.

The most obvious approach is to reduce the modulator deviation ratio, reducing the spectral occupancy of the FM signal. The current deviation ratio is between 2.0 and 2.7. Unfortunately as the deviation ratio is decreased the susceptibility of the signal to noise is increased. Subjective voice quality tests have been conducted using 30 KHz and 15 KHz FM signals in an effort to determine the minimum allowable carrier-to-interference ratio [1]. The 30 KHz channel needed a C/I of 18 dB while the 15 KHz channel required a C/I of at least 24 dB. For fixed power levels, the 15 KHz channels cannot be reused as often geographically as the 30 KHz channels. The increase in the number of channels in the band
is offset by the increase in the required C/I. By reducing the deviation ratio, greater channel efficiency is acquired, but there is not a significant increase in spectral efficiency.

C. Y. Lee has suggested a method for getting around this problem [2]. 30 KHz channels can be combined in a single cell with 15 KHz and even 7.5 KHz channels to form a multiple channel bandwidth system. The two bandwidth case demonstrates the concept quite well. A hexagonal cell is divided into two concentric hexagons. The inner hexagon is serviced by the 15 KHz channels while the outer is serviced by the 30 KHz channels. A typical seven cell reuse pattern is shown in Figure 4-1 [2]. The 15 KHz channels achieve higher C/I's by having a larger cochannel reuse factor D/R = 6.3; the C/I and D/R are related as $C/I = (D/R)^4/6$ for a fourth law propagation loss model commonly used in cellular designs. The regions separating the 15 KHz cells are served by the less noise sensitive 30 KHz signals, requiring a D/R of only 4.6.

The C/I requirements previously mentioned determine the size of the inner ring. Assuming a fourth power loss model, the ratio of inner and outer cell radii, denoted as $R_1$
and $R_0$ respectively, must satisfy
\[
\frac{D/R_1}{D/R_0} = \frac{6.3}{4.6}.
\]

$R_1$ is thus seen to be equal to $0.70R_0$. The area of the inner cell is thus approximately equal to the area of the outer cell. Assuming that the number of channels in both the inner and the outer cells are equal, the amount of additional capacity attained can be readily calculated. Given 10 MHz of bandwidth divided into 30 KHz channels, there will be 333 available channels and $333/7$ or 47 channels per cell with an $N=7$ reuse pattern, where
\[
N = \frac{(D/R)^2}{3},
\]
and $D/R = 4.6$. In the two channel bandwidth system there are
\[
2 \times \frac{10 \text{ MHz}}{30 \text{ KHz}} + \frac{1}{3} \times \frac{10 \text{ MHz}}{15 \text{ KHz}} = 222 + 222 = 444 \text{ channels}.
\]

Given a seven cell reuse pattern, a quick calculation shows that the inner and outer cells will each have 31 channels. An increase in spectral efficiency of $62/47$ or 1.33 has been obtained.

A three channel bandwidth system with three concentric hexagons of equal area can be similarly shown to provide 777 channels. Spectral efficiency is thus increased by a factor of 2.33.

As the cell radius of the inner cells is decreased and as the number of the subcells increases, the problems involved in handing off between cells increases. In addition the change in bandwidth necessitates modification of each mobile and each base station.

S. W. Halpern [25] has suggested an underlay/overlay scheme to create a higher capacity system. In an underlay/overlay scheme an inner cell is created within the existing cells. Channels allocated to the inner cells are operated at the same power level, but are only used when the mobile is substantially closer to the base station. The reduction in cell radius for these channels allows for an increase in the geographical channel reuse rate. Performance is determined by the ratio $D/R$, all other things being equal. If $R$ is decreased, then $D$ may also be decreased. This scheme alone will provide an increase in capacity by a factor of 1.57. Consider the following example.
Maintain the ratio \( \frac{D}{R} \) at the required level of 4.6 for good quality in the inner and the outer cell regions, that is, \( \frac{D_1}{R_1} = \frac{D_0}{R_0} = 4.6 \). Use an \( N=3 \) reuse pattern for the inner cell and \( N=7 \) for the outer cell. Then \( \frac{D_1}{R_0} = 3 \) and

\[
\frac{D_1}{R_1} = \frac{4.6}{3}.
\]

Hence \( R_1 = 0.65R_0 \) is the relation between the inner and outer cell radii and the relation between the inner and the outer areas is \( A_1 = (0.65)^2 A_0 = 0.43A_0 \). With a total of 333 channels there will be \( 0.43(333) \) or 142 in the inner ring and the remainder or 191 in the outer ring. Then we have

\[
\text{channels/cell} = 191/7 + 142/3 = 27 + 47 = 74
\]

and an improvement factor of \( 74/47 = 1.57 \) is realized.

Lee [2] has suggested a power adjustment underlay/overlay technique. It does not appear that such a technique will provide any improvement since the C/I remains the same if the interference and the signal power are all reduced by the same factor.

Combining the underlay/overlay scheme of Halpern and the two channel bandwidth scheme as indicated in Figure 4-2 also does not provide improvement. In order to have adequate performance in the inner or low bandwidth ring we must have \( \frac{D_1}{R_1} = 6.3 \) corresponding to a C/I of 24 dB. Hence we have \( \frac{D_1}{R_1} = 6.3 \) and \( \frac{D_1}{R_0} = 3.0 \) or \( R_1 = 0.48R_0 \) and \( A_1 = 0.23A_0 \). If we assume that the channels are distributed uniformly with area then we have

\[
\frac{N_1 \times 10MHz_{15KHz}}{N_1 \times 10MHz_{15KHz} + (1 - N_1) \times 10MHz_{30KHz}} = 0.23
\]

and hence \( N_1 = 0.13 \). We then have the fraction 0.13 of the total bandwidth available for the narrowband inner ring. This implies that

\[
\text{channels/cell} = \frac{0.13 \times 10MHz_{15KHz}}{3} + \frac{0.87 \times 10MHz_{30KHz}}{7} = 70.
\]

The improvement is then \( 70/47 = 1.49 \) or less than that of the overlay/underlay technique used alone!

Lee suggests that a third method be incorporated with the first two to further increase system capacity [2]. The limiting factor in frequency reuse is cochannel interference. If
channels are offset in frequency as they are reused, cochannel interference is reduced in two ways:

1. cochannel interference energy contributions from nearby cells is reduced.

2. crosstalk in the form of intelligible speech is practically eliminated.

Figure 4-3 shows a one-third channel offset system using a seven cell reuse pattern. The closest cochannel interferers to the center cell 1 are three one-third offset cells 1' and three two-thirds offset cells 1''. The C/I formula, according to Lee, for this system is

\[
\frac{C}{I} = \frac{C}{\sum_{i=1}^{3}(I_1' + I_1'')} \geq 18 \text{ dB} \quad [2],
\]

where \(I_1'\) and \(I_1''\) are the interference components introduced by the one-third overlapping and two-third overlapping cochannel interferers respectively. As the noise energy introduced
is proportional to the spectral overlap, the preceding expression may be re-expressed as

\[ \frac{C}{I} = \frac{C}{3(0.333 + 0.667)I_i} \geq 18dB \]

\[ \Rightarrow \frac{C}{3I_i} \geq 63 \]

\[ \Rightarrow (\frac{D}{R})^4 = 189 \]

\[ \Rightarrow D = 3.71R \]

Since \( D/R = \sqrt{3N} \), \( N \) has been reduced from 7 to 4.58. A four cell reuse pattern may be selected if a small amount of additional cochannel interference suppression is obtained (e.g. by implementing each cell as three sectors with directional base antennas). The capacity improvement obtained through the one-third channel offset method alone is almost a factor of two, for the same number of channels distributed over seven cells are now distributed over four cells.

Unfortunately the formula above due to Lee does not take into account the fact that
the frequency offset also introduces interference from the other shifted groups of frequencies. For example, 2 and 1" are now interfering, where these cells are actually adjacent as seen in the Figure 4-3! The reduction in intelligible crosstalk may indeed be realized due to the FM capture effect.

Lee claims that the greatest capacity increases can be obtained by creating hybrid systems that use combinations of the aforementioned approaches [2]. Figure 4-4 shows the spectral efficiencies for a variety of such schemes. The best of the schemes offers an increase in capacity by a factor of 5.0. However since the multiple bandwidth and overlay/underlay techniques will not play together well and the one-third frequency offset scheme does not reduce actual interference levels, the best real gain is due to the overlay/underlay scheme alone and is 1.6 instead of 5.0.

Reiterating, there are, unfortunately, some inconsistencies in Lee's use of the three
aforementioned schemes for increasing capacity in an existing analog system. The multiple-bandwidth system will work as described, if it is used by itself. In this form it can provide a capacity increase by a factor of 1.33 in the two bandwidth case. Problems arise when one combines the multiple bandwidth scheme with the underlay/overlay scheme. Recall that the 15 KHz channels require a higher C/I than the 30 KHz channels. If the inner cells in an underlay/overlay scheme are assigned 15 KHz channels, the C/I levels will be too low. The radii of the inner cells must be further reduced by a factor of approximately 1.35 to compensate for the increased C/I requirement. Although more channels are made available by using 15 KHz channel spacings instead of 30 KHz, not enough of the channels are assigned to the 3 cell reuse pattern. Therefore, the combined increase in capacity will be slightly less than that for the underlay/overlay scheme alone! It is not clear at all how the one-third channel offset scheme offers any increase in capacity whatsoever. Though the energy contribution from the original interferers is reduced by one- or two-thirds, other cells which previously presented no problem become interferers. Given a continuous spectral allocation, the offset of the channel assignments merely moves part of the interfering energy from one cell to another. If the new interferer is closer, then the performance of the new system may even be worse than the non-offset system. The only possible benefit from the offset would be an elimination of crosstalk (the noise is now unintelligible, though of the same or greater intensity).

Of Lee's three suggestions, it is probably best to consider the use of the two-channel bandwidth scheme or the underlay/overlay scheme by themselves. These schemes offer modest improvements in capacity on the order of 1.3 or 1.57. The underlay/overlay scheme has the advantage that it apparently requires no change in the mobile units and no unreasonable changes in the base stations although the handoff problem is more severe.

The preceding methods for increasing capacity had the admirable trait of using existing base station locations. If one is willing to add new base stations, then it is much easier to obtain significant increases in capacity, though at greater cost. Ericsson and other European companies are currently considering the use of microcell cellular layouts [10]. With cell sizes down to 100 meters, capacity can be increased by an order of magnitude. The corresponding low transmitter power levels require the use of digital modulation formats,
which will be discussed in the next section. Significant capacity increases can still be achieved while retaining the current FM modulation format, though it is not clear what the minimum cell size would be (probably 500 - 1000 meters or perhaps one mile).

4.2 Amplitude Modulation Formats

One method for increasing the capacity of the current system is to adopt a new modulation format that allows SCPC channels with narrower bandwidths. Single-sideband AM has attracted a lot of attention because its spectral occupancy is approximately equal to the highest frequency component in the signal. Spectral occupancy can be further decreased through the use of frequency or amplitude companding.

Speech signals do not usually display high and low frequency components simultaneously. It is possible to fold the spectrum of a speech signal and reduce bandwidth by 60% [9]. If amplitude companding is used as well, channel spacings can be reduced to approximately 2.5 KHz, providing an order of magnitude increase in system capacity.

There are, however, some disadvantages to SSB-AM approaches in a mobile phone environment. The fading problem is as severe as that for FM signals, with the additional loss of any capture effect that might suppress cochannel interference in a cellular system [9]. Frequency stability problems also become an issue in 900 - 1000 MHz systems with channel bandwidths under 5 KHz. Finally, there has been a great deal of discussion in the literature on the effect of frequency companding on the intelligibility of speech. Frequency companded signals do not provide toll-quality speech. If the new system provides lower quality service than the old system, a great deal of user antipathy can be anticipated despite the increase in system capacity. Even with these handicaps, companded SSB-AM should be considered if a moderate increase (factor of 5) in capacity is all that is desired.

4.3 Technology Issues

All of the aforementioned techniques can be implemented using current technology. In fact, performance data from European simulations and field tests is already available for some of these schemes.
4.4 Compatibility Issues

Of the various methods for increasing capacity discussed in this report, the overlay/underlay of Halpern will have the least negative impact on the current users.

Multiple channel bandwidth schemes retain capacity for some of the original 30 KHz signals. Users with old equipment would be able to use it while new buyers could purchase equipment that supported the narrow band formats.

Channel offset schemes will require modifications of filtering and LO generation circuitry, but it is conceivable that existing equipment (both mobile and hub) could be modified without too much trouble. If this proved prohibitive, a compromise scheme could be devised in which some channels were included in the new offset scheme and others left in their original format. Of course, any such compromise would reduce the capacity increase realized from the change in format. These schemes are apparently of limited effectiveness, as discussed above.

Schemes that involve variations in transmitter power level (e.g. underlay/overlay as described by Lee) would require the least modification of existing equipment. It may be that the required power reduction is within the dynamic range of the current system’s power control mechanism. Again note that this scheme is apparently not effective.

References


5 Digital Modulation Schemes

There are many advantages inherent in digital modulation formats for cellular mobile phone systems [3,4,5,6,7,8,18,19,20]. In general a digital format allows for easier manipulation and control of the source data, creating opportunities for a variety of services hitherto impossible with the current analog format. Several possibilities for new services will be listed at the end of this section. The primary problem at hand, though, is capacity. Simply changing the modulation format from analog to digital will not increase available capacity by more than a factor of 1.5 to 2. Capacity increases of an order of magnitude or more can only be achieved through reductions in cell size and/or advanced spread-spectrum techniques in conjunction with a digital modulation format. The careful selection of a digital modulation format enhances the capacity increase realized through the other techniques.

In selecting a digital modulation format, the following requirements must be satisfied [19].

- **Compact Output Power Spectrum:** Adjacent channel interference should be kept below -60 dBc. Deep fades on the order of 40 dB are expected in the mobile environment. The -60 dBc ACI requirement leaves a 20 dB carrier-to-interference ratio (CIR).

- **Excellent Probability of Error Performance:** A high degree of noise immunity allows for operation at low SNR levels. This in turn allows for smaller cell sizes and high-density geographical channel reuse.

In the following section a variety of the most promising digital modulation schemes will be examined. An effort will be made to indicate the hardware complexity involved with these techniques.

5.1 Constant Envelope Modulation Formats

There are several points in a communication system in which it is advantageous to make use of nonlinear devices. For example, square law devices are quite useful in the design of detectors and phase locked loops. Other devices that are nominally linear can be forced to operate in a nonlinear region to improve their sensitivity or power efficiency. In a mobile phone system nonlinearities are most likely to be encountered in the design of power
amplifiers for the transmitter in the mobile. The use of Class C amplifiers is indicated by the limited power available in mobile or portable devices. Unfortunately the placement of nonlinearities in the signal path of a communication system creates certain problems for the designer. The nonlinearities convert amplitude modulation on the carrier into phase modulation. This AM/PM interference introduces phase jitter into the carrier and symbol recovery circuits, reducing their efficiency. One method for combating this problem is to minimize amplitude modulation on the carrier through the use of constant envelope modulation formats. In the following section the most promising of the constant envelope formats will be examined.

5.1.1 Phase Shift Keying and Frequency Shift Keying

There are essentially two different signaling techniques that provide a constant envelope: phase shift keying (PSK) and frequency shift keying (FSK). PSK formats translate a group of n bits in the message stream into one of \(2^n\) phase offsets in the carrier. FSK formats convert the n bit groups into carrier frequency offsets. Let us consider the binary cases (n=1) first.

In BPSK and BFSK a binary symbol is transmitted every T seconds. Their respective waveforms are shown in Figure 5-1. The phase difference between the two signals in the BPSK constellation is maximized at 180 degrees. This is equivalent to amplitude modulating the carrier with +1 or -1 every T seconds. The signals in this set are thus maximally separated, or antipodal. BPSK receivers recover a coherent carrier reference that is used to determine the instantaneous phase of the received signal. Receivers that perform carrier phase recovery are called coherent receivers. It can be shown that the coherent detection of antipodal signals in an AWGN environment offers the best possible bit error rate (BER) performance.

The BFSK modulation format uses the information stream to switch the carrier between two distinct frequencies. The simplest possible receiver structure for this modulation format consists of a pair of bandpass filters followed by envelope detectors and samplers operating at the bit rate. The basic idea is to determine which of the two filter passbands contains the most signal energy during a bit period. If the two signals are sufficiently sep-
arated in frequency, there will be no crosstalk. Under ideal conditions all of the carrier energy will fall in only one of the bandpass regions during a bit period. Signals that do not interfere with one another during detection are called orthogonal. Orthogonal signal sets are ideal for noncoherent receiver structures. An FSK signal set is noncoherently orthogonal if the constituent frequencies are pairwise separated by at least $1/T$ Hz, where $T$ is the bit period. Noncoherent receivers such as the one just described tend to be much less complicated than coherent receivers. Unfortunately their is also a significant reduction in bit error rate performance with respect to coherent BPSK. BFSK signals may also be received coherently, though their BER performance will still lag that of BPSK by 3 dB. For this reason BFSK and other FSK techniques are usually relegated to use in low data rate, low efficiency applications. However, as will be shown later, certain variants of FSK may prove ideal for use in mobile communication systems because of their associated spectral properties.

The BPSK signaling format provides the low power BER performance required by small-cell mobile communication systems. It does not, however, provide the necessary spectral containment. Figure 5-2 shows the power spectrum for a BPSK modulated carrier.
The main lobe has a one sided bandwidth of $1/T$. For this reason BPSK is said to provide a spectral efficiency of 1 bit/second/Hz. The sidelobes and a small portion of the main lobe can be removed by filtering without severely distorting the signal. This leaves a spectral occupancy on the order of 30 KHz for 16Kbps encoded voice. This is far too extravagant a use of bandwidth for cellular mobile communications. One must look towards the PSK formats that have higher spectral efficiencies.

5.1.2 QPSK and Offset-QPSK

In BPSK a single sinusoid was modulated by varying its phase in accordance with an incoming bit stream. This was seen to be equivalent to amplitude modulating with a +1 or -1. The resulting modulated carrier may be represented by

$$\frac{1}{\sqrt{2}} \alpha(t) \cos(2\pi f_c t),$$

where $f_c$ is the carrier frequency and $\alpha(t)$ is the modulating bit stream. Note that $\sin(2\pi f_c t)$ and $\cos(2\pi f_c t)$ are coherently orthogonal, i.e.

$$\int_0^{T_0} \sin(2\pi f_c t) \cos(2\pi f_c t) \, dt = 0$$
for integer values of n. A second BPSK signal may thus be placed in quadrature with the first without causing interference in the coherent detection of either signal. This is the basis for QPSK, or Quadrature Phase Shift Keying. The binary information sequence is separated into two modulating sequences, one \((\alpha_I(t))\) for the in-phase component and the other \((\alpha_Q(t))\) for the quadrature component. The resulting modulated carrier may be expressed as

\[
s(t) = \frac{1}{\sqrt{2}} \alpha_I(t) \cos(2\pi f_c t + \frac{\pi}{4}) + \frac{1}{\sqrt{2}} \alpha_Q(t) \sin(2\pi f_c t + \frac{\pi}{4})
\]

\[
= \cos(2\pi f_c t + \Theta(t)),
\]

where \(\Theta(t) \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}\). The possible values of \(\Theta\) correspond to the four possible combinations of values taken on by \(\alpha_I\) and \(\alpha_Q\) during a given symbol period. The signal constellation is shown in Figure 5-3.

The in-phase and quadrature information streams in a QPSK modulator are timed so that their bit transitions occur simultaneously. Figure 5-3 shows that the carrier may thus experience phase transitions of 0, ±90, or 180 degrees every 2T seconds. The range of these transitions is reduced to 0 and ±90 in offset-QPSK (OQPSK). In this variant of QPSK, the information streams are staggered so that the bit transitions in the in-phase stream occur half-way through the bit periods in the quadrature stream. Phase transitions in the OQPSK carrier occur twice as often as in the QPSK carrier, but with only half the intensity. This has absolutely no effect on the signal constellation or the signal power spectrum. The motivation for using this variant lies in the performance of filtered OQPSK signals as they pass through nonlinearities. If the spectrum is being shared with other users, the sidelobes of the signal are generally filtered out to reduce adjacent-channel interference. This will certainly be the case in a cellular mobile phone system. When a filtered QPSK signal passes through a nonlinearity, the sidelobes are partially reconstructed. This is a direct effect of the phase discontinuities in the signal. The filtered OQPSK signal suffers some sidelobe regeneration, but not anywhere near as much as the QPSK signal (see Figure 5-4). OQPSK also shows greater immunity to phase jitter in the receiver in the presence of additive white gaussian noise. OQPSK modulation is thus a stronger candidate for use in a mobile phone system.
Figure 5-3

QPSK Signal Constellation
Figure 5-4

QPSK and OQPSK Sidelobe Regeneration
The spectral efficiency of QPSK and OQPSK is 2 bits/second/Hz. They offer far better spectral containment than their binary relations and are thus used in many bandlimited applications (e.g. TDMA/FDMA commercial satellite channels). Unfortunately their performance is still not quite good enough for a mobile phone system. The distributed geometry of the mobile phone system creates a near/far problem in which mobiles at the center of a cell can drown out users in adjacent channels that are traveling at the periphery of the cell. For this reason severe adjacent channel interference specifications are needed. In general sidelobe leakage into adjacent channels should be restricted to -60 to -80dBc. QPSK and OQPSK cannot meet such specifications without extremely complex filtering.

There are two approaches to solving this problem. First we may again increase spectral efficiency by moving to 8-ary modulation formats. Unfortunately this will not work because of the BER requirements. BER performance deteriorates as the size of the signal set is increased. This makes sense intuitively for constant envelope modulation schemes because an increase in the number of signals implies a decrease in their separation at a fixed power level. The low SNR performance requirement (allowing for small cell size) restricts the selection of modulation formats to efficiencies on the order of 2 bits/second/Hz.

The second approach is to reduce sidelobe levels by minimizing the phase discontinuities in the modulated signal. A series of continuous phase modulation formats have been developed for just this purpose.

5.1.3 MSK

In QPSK and OQPSK the information streams were presented to the modulator as rectangular pulses. Since these pulses maintained a constant amplitude over a bit period, they appear as constants ($\alpha_I(t)$ or $\alpha_Q(t)$) in the signal representation and analysis. *Minimum Shift Keying*, or MSK, is an adaptation of OQPSK in which the modulating pulses are sinusoidal instead of rectangular [11]. The result of this modification is quite interesting and well worth pursuing.

The MSK signal can be expressed as

$$s(t) = \alpha_I(t)\cos\left(\frac{\pi t}{2T}\right)\cos(2\pi f_c t) + \alpha_Q(t)\sin\left(\frac{\pi t}{2T}\right)\sin(2\pi f_c t).$$
A few trigonometric substitutions yield

\[ s(t) = \cos[2\pi f_c t + b_k(t)\left(\frac{\pi t}{2T}\right) + \Phi_k], \]

where \( b_k \) is defined as 1 when the in-phase and quadrature information bits differ and -1 when they are the same. The function \( b_k(t) \) may thus be expressed as the product \(-a_1(t)a_Q(t)\).

Two points are immediately apparent. First, the signal has a constant envelope. Secondly, the phase is continuous at bit transitions. The phase of the MSK signal relative to the carrier may be represented by

\[ \Phi(t) = b_k(t)\frac{\pi t}{2T}. \]

The phase thus increases or decreases linearly during each bit period. What is not quite so obvious is that this signal can be viewed as a coherent FSK signal with frequency spacings of 1/2T Hz. This becomes clear when one combines the two time dependent terms in the argument of the cosine function. MSK derives its name from the fact that this is the minimum frequency spacing for coherently orthogonal sinusoids. MSK is sometimes referred to as Fast-FSK.

A comparison of the MSK spectrum with that of OQPSK and QPSK provides interesting results. The power spectra of these signals is a translated version of the Fourier transform of the modulating pulses. The normalized spectral density of the QPSK and OQPSK signals is thus

\[ \frac{G(f)}{T} = 2\left(\sin(2\pi f T)\right)^2 \]

while that for MSK is

\[ \frac{G(f)}{T} = 16\left(\frac{\cos(2\pi f T)}{1 - 16f^2 T^2}\right)^2. \]

These two expressions are plotted in Figure 5-5. There are two items to note that are of immediate interest. First, as expected, the sidelobes of the MSK signal are lower than that for the other two. Unfortunately this performance improvement is somewhat offset by an increase in the width of the main lobe. In some applications (e.g. narrowband FDMA satellite links), the wider main lobe eliminates MSK from consideration. The general rule of thumb is that when the product of channel spacing and symbol duration exceeds 2.3,
MSK will outperform QPSK and OQPSK in spectral containment tests. This is clearly the case for the spectral allocation in the current analog mobile phone service.

5.1.4 Gaussian MSK

Spectral containment may be further improved by introducing a controlled amount of intersymbol interference into the modulated signal [12,13,16,17,19,24]. This further reduces discontinuities in the carrier (in this case higher order derivatives of the time domain waveform) and increases the roll-off of the main lobe while further reducing sidelobe amplitude. The price for this improvement is the reduction in BER performance caused by the ISI. One particular method has attracted a lot of attention in the mobile communications field: Gaussian MSK (GMSK). In GMSK the sinusoidal modulation pulses are passed through a Gaussian filter prior to their use by the modulator, as shown in Figure 5-6. The result is the
set of spectra shown in Figure 5-7. As the normalized 3dB bandwidth $B_bT$ of the premodulation filter is decreased, the spectral occupancy of the GMSK signal is reduced. It has been determined that a normalized bandwidth value of 0.25 provides optimal performance in a GMSK cellular mobile phone system [19].

This is the first scheme discussed so far that provides enough out of band radiated power suppression to function efficiently in a cellular system, so a closer look is in order. The CCIR recommendations specify out of band power levels of -60dBc or less. Figure 5-8 shows the levels of adjacent channel interference as a function of normalized channel separation for several values of normalized premodulation filter bandwidth. A normalized channel separation of 1.5 corresponds to a 16 Kbps GMSK modulated signal in a 25 KHz frequency band. A normalized filter bandwidth of 0.25 is clearly adequate. Narrower bandwidths will allow more closely packed signals, increasing capacity by factors as high as 2. Unfortunately the penalty for the increased spectral efficiency of GMSK is exacted in the form of an increase in BER. A premodulation filter with a 0.25 normalized bandwidth reduces the normalized signal distance from 2.0 to 1.68. This corresponds to a 0.74 dB decrease in BER performance. This may be translated into an increase in required output
Figure 5-7

GMSK Power Spectra
power level at the mobile and base of 0.74 dB. This is a reasonable tradeoff, though it will set a lower limit on cell size. Further decrease in bandwidth cause an exponential decrease in signal separation, so substantial increases in system capacity cannot be achieved by simply adopting a GMSK modulation format. However, substantial increases in capacity (an order of magnitude or more) can be achieved by combining a GMSK modulation format with a microcell layout.

5.1.5 Tamed Frequency Modulation

Spectral efficiency may be further improved through the use of correlated (partial response) CPFSK modulation schemes [20,14,15]. Such schemes use modulation pulses that are wider than a single symbol period. Signal phase shifts are correlated so that the phase shift over a
bit interval is a function of the present bit as well as several previous bits. Improved spectral performance is thus provided at the expense of greater implementational complexity for the transmitter and receiver. There is one particular method that will more than likely find application in mobile phone systems: Tamed FM (TFM).

Recall that MSK improved on the spectral containment of QPSK and OQPSK by eliminating phase discontinuities in the time domain waveform. GMSK further reduced spectral containment by reducing discontinuities in the slope of the phase function; the injection of ISI "smoothed" the phase function at bit intervals. TFM takes this process a step further. The multiplexed data signal can be represented by

$$a(t) = \sum_{n=-\infty}^{\infty} \alpha_n \delta(t - nT), \, \alpha_n \in \{1, -1\}.$$  

The phase function for the MSK waveform must obey the difference equation

$$\Phi(mT + T) - \Phi(mT) = \alpha_m \frac{\pi}{2}$$

with initial conditions $\Phi(0) = 0$ if $\alpha_0 \alpha_1 = 1$ and $\Phi(0) = \frac{\pi}{4}$ if $\alpha_0 \alpha_1 = -1$. The phase function for TFM is further smoothed by spreading out the phase correlation over several bits. It must satisfy

$$\Phi(mT + T) - \Phi(mT) = \frac{\pi}{2} \left( \frac{\alpha_{m-1}}{4} + \frac{\alpha_m}{2} + \frac{\alpha_{m+1}}{4} \right)$$

with the same initial conditions. The maximum phase change over any bit interval is thus reduced to $\frac{\pi}{4}$. The phase function at bit transitions is smoothed through the use of a premodulation filter. Figure 5-9 shows the difference between the phase functions for MSK and TFM while Figure 5-10 shows the corresponding contrast in power spectral density functions. A moderate reduction in adjacent channel interference is acquired at the cost of a moderate increase in complexity.

5.1.6 Generalized Tamed Frequency Modulation

The spectral occupancy of TFM signals can be further reduced through the use of premodulation filters [15,19,20]. The premodulation filter is designed to smooth phase transitions while ensuring that the phase of the output signal settles at one of the desired values at the
Figure 5-9

Phase Behavior of MSK (—), MSK with Sinusoidal Smoothing (...), and TFM (—...).
Figure 5-10
Power Spectra of MSK (—), MSK with Sinusoidal Smoothing (...), and TFM (——)
end of a bit period [15]. A sufficient condition for such performance is the satisfaction of the third Nyquist criterion

\[ \int_{(2l-1)T/2}^{(2l+1)T/2} h(t) \, dt = \begin{cases} 1 & \text{for } l = 0 \\ 0 & \text{otherwise} \end{cases}, \]

where \( h(t) \) is the impulse response of the filter. The transform domain representation may be expressed as

\[ H(f) = W(f) \left[ \frac{\pi f T}{\sin(\pi f T)} \right], \]

where \( W(f) \) is a low pass filter that satisfies the first Nyquist criterion. \( W(f) \) is usually assumed to have a raised cosine characteristic

\[ W(f) = \begin{cases} 1, & 0 \leq |f| \leq \frac{(1-r)}{2T} \\ 0.5(1 - \sin\left(\frac{f T - 0.5\pi}{r}\right)), & \frac{(1-r)}{2T} \leq |f| \leq \frac{(1+r)}{2T} \\ 0, & \text{otherwise} \end{cases}, \]

Figure 5-11 shows a typical adaptation of the basic TFM modulator. The Nyquist-3 filter performs the premodulation filtering and an additional degree of freedom has been added in the form of variable tap coefficients \( a \) and \( B \) on the transversal filter. The transversal filter provides the correlative coding property for the TFM modulator. Its amplitude response can be represented as

\[ |S(f)| = B \left\{ 1 + \frac{2a}{B} \cos(2\pi f T) \right\}. \]

The resulting class of signals is referred to as Generalized-TFM (GTFM). Figure 5-12 compares power spectra for several different GTFM signals. Note that a premodulation filter with roll-off factor \( r=0 \) provides the basic TFM signal. GTFM offers a significant reduction in spectral occupancy at the expense of increased complexity and BER.

### 5.2 Nonconstant Envelope Modulation Formats

As mentioned earlier, one of the principle disadvantages of nonconstant envelope modulation schemes is the level of AM/PM noise they introduce at nonlinearities in the communication system. Their performance in fading environments is on the whole rather poor as
Figure 5-11

GTFM Modulator Block Diagram

Figure 5-12

The Power Spectra of Several GTFM signals. (a) MSK (reference) (b) B = 0.5, r = 0.0 (TFM), (c) B = 0.54, r = 0.2, (d) B = 0.58, r = 0.3, (e) B = 0.63, r = 0.36, (f) B = 0.8, r = 0.52, (g) B = 1.0, r = 0.5.
well. There is however one family of nonconstant envelope modulation formats that have been specifically developed for use with saturated HPA’s and TWTA’s. Through careful pulse shaping it has been shown that a high degree of spectral containment is possible at the expense of BER performance in low SNR conditions. This family of modulation formats is worthy of note here, and must be considered should an in-depth study of digital modulation schemes for mobile phone services be undertaken [19].

This family of schemes uses a raised cosine pulse

\[
p(t) = \begin{cases} 
\frac{1}{2}(1 + \cos \frac{\pi t}{T_s}) & |t| \leq T_s \\
0 & |t| > T_s 
\end{cases}
\]

to modulate the in-phase and quadrature components of the carrier. Note that the pulses occupy two signaling intervals, introducing a controlled amount of ISI. Both Quadrature-Overlapped-Raised-Cosine (QORC) and Staggered-Quadrature-Overlapped-Raised-Cosine (SQORC) have been shown to have less sidelobe regeneration after hard limiting than OQPSK and MSK. Their performance in a rapidly fading environment is not yet completely understood, but if it proves acceptable, QORC and SQORC will be prime candidates for use in densely packed FDMA cellular mobile phone systems.

5.3 Error Correcting Codes

One of the most important benefits acquired through the use of a digital modulation format is the availability of a host of error detection and correction schemes [21,22,23]. Such schemes are based on the introduction of redundancy into the information stream prior to or during the modulation process. The receiver uses this redundancy to detect or correct error patterns induced by the channel noise process. As the amount of redundancy is increased, the number of detectable or correctable error patterns is increased.

There are essentially two methods for inserting this redundancy. Additional bits can be added to the information stream, causing an increase in data rate and, for a fixed modulation format, an increase in bandwidth. Such methods include block and convolutional coding. Redundancy may also be inserted in the form of an increase in the size of the signaling alphabet. This method requires the use of a more spectrally efficient modulation
format. Trellis coded modulation (TCM) schemes are the most prominent examples from this set.

TCM-type coding schemes are difficult to incorporate into a mobile phone system. The fading environment coupled with the low SNR operational constraint limits spectral utilization to approximately 2 bits/second/Hz. Very little gain can be acquired from a TCM scheme at such rates.

Convolutional codes present a similar problem. They operate very well in low SNR conditions, but require an effective code rate of 1/2 to 3/4 to be useful. This translates into an increase in signal bandwidth by a factor of 1.5 to 2. Such code rates are clearly impractical in a mobile phone system.

Certain block codes are able to provide significant coding gains at high rates. Reed Solomon codes in particular are able to provide 1 to 5 dB coding gains at rates of 15/16 to 31/32. These coding gains allow for significantly lower transmitter power levels and smaller cell areas. A 31/32 Reed Solomon encoded GMSK signal would easily fit within the current SCPC bandwidth allocation. The technology for 16 Kbps Reed Solomon codecs currently exists and is readily available in chip form for integration into mobile transmitters and receivers. The cost impact would be small, while the benefit would be significant. Additional study should be made on this subject.

A second approach is to protect only the most significant bits of the encoded speech. If only a few of the bits are further encoded, low rate FEC codes may be used without significantly reducing the overall code rate. In this case it is best to use rate 1/2 block codes that provide significant protection with minimal implementational complexity. Golay codes (three error correcting, length 23) would be an excellent choice. The blending of speech coding with forward-error-correction coding has not been examined in sufficient detail. It may be possible to make 5 to 10 Kbps speech algorithms more robust, leading to significant reductions in spectral occupancy for mobile phone signals.

5.4 Modem Technology

The best way to get a feel for the specific technologies involved in the implementation of a digital modulation scheme is to take a close look at modem designs for appropriate
modulation formats. Two of the best candidates for implementation in a mobile phone system are GMSK and TFM. In the following pages we will take a look at block diagrams of modems for each of these formats. An effort will be made to point out where development trouble is most likely to arise.

5.4.1 MSK Modems

Figure 5-13 shows a typical MSK modulator. This block diagram can be modified for the GMSK case by prefiltering the in-phase and quadrature information signals. The premodulation Gaussian lowpass filters are simple devices and will not create any development problems. The bandpass filters in the in-phase and quadrature arms are equally simple. The two frequency components f+ and f- are created by isolating the two sidebands at the mixer output. LO suppression is not a problem. The mixer itself will provide 6 to 12 dB of rejection, making the conversion relatively power efficient. Any subsequent LO leakage can only aid in carrier recovery at the receiver. The filters design must focus on providing 20 to 40 dB of image rejection to minimize crosstalk on the received signal. In the 900 - 1000 MHz region this implies a filter with a Q factor on the order of 40,000. This is certainly well within the range of current technology. The summers can be realized with 3dB hybrids and the remaining multipliers with double-sideband mixers, all using simple existing technology.

A GMSK receiver is shown in Figure 5-14. This is a simple correlation receiver with staggered integrate-and-dump filters. The carrier recovery mechanism (Figure 5-15) will be the toughest segment of the entire development. This is not because of the technology: every item required is well within the bounds of current technology. The problem will lie in the sensitivity of the loop to phase transients and deep fades in the mobile phone environment. Cycle slippage and even loss of lock will be serious problems during deep fades. Robustness must be carefully integrated into the loop if it is to handle the environmental extremes. A significant development effort should be expected for the loop design. One should recall however that phase-locked FM tuners have been in use in automobile radios for some time. The effort under consideration would be an extension of that technology.

GMSK may also be demodulated noncoherently using discriminator or differential detection. A reduction in BER performance is entailed and, if used, it is not clear that there
Figure 5-13

MSK Transmitter
Figure 5-14
MSK Receiver

Figure 5-15
Carrier Recovery Circuit for MSK Receiver
would be any advantage over current analog formats. A good treatment of this approach can be found in Hirade [19] and Elnoubi [17]. Both coherent and discriminator approaches should be studied in greater detail before a decision is made.

5.4.2 TFM modems

A transmitter structure for TFM signaling is shown in Figure 5-16. Note the increase in complexity caused by the increased level of correlation in the modulating bit stream. A receiver structure is shown in Figure 5-17. Both of these devices may be readily adapted for use with GTFM. The primary source of concern in implementing this design will be the multiple delay sources in the receiver. These delays will vary with temperature, and will provide a significant level of loop stress (and hence phase jitter and increased possibility of cycle slip) in the carrier recovery system. Quite simply it will be easier to build a rugged GMSK receiver than to build a rugged TFM receiver. This factor should be weighed against the increase in spectral containment provided by the TFM format.

GTFM may also be demodulated using a discriminator approach [15]. This approach is much simpler and appears to be more promising than discriminator detection of GMSK. The discriminator approach should be considered, and may prove superior to the coherent approach when all things are considered. Further analysis and simulations are certainly in order.

5.5 Other Advantages of the Digital Approach

A digital modulation format will make possible several new user services at minimal additional cost. The most prominent among these possible services are high speed data transmission and secure voice transmission.

An extremely reliable high speed data communication link can be established using a 16 Kbps digitally modulated channel. Salesmen and other professionals that work out of their automobiles can be allowed direct access to computer data bases and test equipment, greatly enhancing their productivity. Sales orders and other documents can be FAX'ed to a central processing point in seconds instead of hours. The market for such services is already growing. FAX machines for analog mobile systems entered the market in California in May
Figure 5-16

TFM Transmitter
of this year. If the market is capable of supporting the slow (300/1200 baud) transmission of digital data over an analog channel, it will certainly support 9600 baud transmission over a digital channel at comparable prices.

Secure voice transmission over mobile phones has not been a great source of debate. However, as the amount of data transmission over mobile phone systems increases, data security will become more of an important issue. The fact of the matter is that the cellular system by its very nature offers a significant amount of protection against the casual eavesdropper. If such a person was somehow able to locate a particular transmission amidst the myriad other calls under way, he would lose the connection when the mobile crossed a cell boundary. It is also highly unlikely that he would catch both ends of the conversation. Such thoughts are comforting to those who understand the system and are given to careful reflection on the matter. Unfortunately the majority of the mobile phone market cannot be readily included in this group. Analog phone traffic can be scrambled, thwarting the most persistent eavesdropping threat. However this usually involves analog to digital conversion of one form or another. The price for protecting an analog link will therefore be significantly higher than that for protecting a digital link. The digital link will already have the
information in an easily manipulated form. A simple form of protection could be provided using linear shift registers with tapped delay lines. More sophisticated protection could be provided using chip sets based on the Data Encryption Standard (DES).

References


6 Spread-Spectrum and Time Division Multiple Access

6.1 Introduction

Spread-spectrum refers to a large class of modulation techniques where the bandwidth used to transmit a signal exceeds the minimum required for data communication and the band-spread is achieved by using a code that is independent of the data being transmitted. Because of the independence requirement, modulation techniques such as FM do not qualify as spread-spectrum. The receiver uses a synchronized version of the spreading sequence to despread the received signal allowing recovery of the original data sequence. Spread-spectrum systems have been traditionally used by the military to provide enemy jam-resistant communications. In the last decade, they have been proposed for land mobile radio systems. There are many excellent general references on spread-spectrum systems, the best to date being that by Simon, Omura, Scholtz and Levitt [1].

There are two basic forms of spread-spectrum modulation that have been proposed for use in land mobile radio, direct-sequence (DS) and frequency-hopped (FH). With DS spread-spectrum, spreading is accomplished by introducing rapid pseudo-random phase transitions in the carrier containing the data as shown in Fig 6.1. DS spread-spectrum reduces the effect of intentional and unintentional interference by averaging the effect of the interference over the spread-spectrum bandwidth.

FH systems achieve the spectral spread by pseudo-randomly changing the carrier frequency, as shown in Fig 6.2. FH systems reject interference by avoidance, where frequency hopping ensures that the signal does not remain in a section of bandwidth having a large amount of interference for an extended period. There are two basic types of FH systems, slow frequency hopped (SFH) and fast frequency hopped (FFH). SFH systems refer to those where the number of modulated symbols per hop frequency is greater than, or equal to, one. FFH systems, on the other hand, are those where the number of hops per modulated symbol is greater than one.

Regardless of the type of spreading used, the users in a spread-spectrum system share a common section of bandwidth. This type of multiple-accessing is sometimes referred to as spread-spectrum multiple-access (SSMA) or code division multiple-access (CDMA). In a
spread-spectrum network, each user is assigned a unique spreading code that is distinguishable from the spreading codes used by all the other users as well as cyclic shifts of itself. For the spreading codes to be distinguished from cyclic shifts of themselves, they must have a low off-peak autocorrelations. To be distinguishable from each other, the spreading codes must have low crosscorrelations. If the codes are chosen to meet these conditions, then the effect of all the other users on the receiver, matched to the desired signal, is additive interference. The statistical characteristics of this interference depends on the type of spreading, DS or FH, being used as well as the exact sequences being used. For DS spread-spectrum, the interfering users grossly appear as additive white Gaussian noise in the intended receiver. For FH systems, the statistical characteristics of the interference are more complex.

When SSMA is used in cellular land mobile radio systems, it is often used in conjunction with time-division multiple-access (TDMA). With TDMA, the channel is temporally divided into frames. Each frame consists of a fixed number of time slots as shown in Fig. 6.3. Each user is allowed to access the channel only during its designated time slot in each frame. Synchronization is required amongst the users so that they begin and end transmission at the beginning and end of their time slots. A guard-time is allowed between the slots within
a frame to relax the timing requirements in the system. Important parameters in a TDMA system are the slot duration, guard-time, number of slots per frame, and the transmission rate. TDMA is useful in cellular systems because it reduces the cost of the base station since the number of required base station receivers is reduced by the multiplexing factor.

Often the literature refers to narrow-band TDMA and wide-band TDMA. Narrow-band TDMA refers to a system with \( \approx 10 \) or less slots per frame, while wide-band TDMA has more than \( \approx 10 \) slots per frame.

6.2 History of Spread-Spectrum in Land Mobile Radio

Historically, spread-spectrum was first proposed for mobile radio in 1977 by Cooper and Nettleton [2]. They used FH spread-spectrum with differentially detected biphase modulation (FH/DPSK). Originally, they claimed a higher spectral efficiency than analog FDM/FM systems. However, subsequent analysis showed that the relative spectral efficiency was 8.4% as compared to the efficiency of 100% for an FDM/FM system with 30 \( K\)Hz channel spacings. Nettleton and Cooper provided a much more complete analysis of their system in 1981 [3]. Also, Matsumoto and Cooper have examined the effect of multiple narrow-band interferers on this system [4]. These interferers would be a result of overlaying the FH/DPSK
system on a frequency band that is occupied by conventional narrow-band systems. It was shown that the system can perform satisfactorily in the presence of this form of interference, with the expected result that an increase in the number of interferers gradually degrades the performance. Yue examined the performance of various receiver structures for FH/DPSK systems, and showed that a hard limiting receiver allowed for twice the capacity of the linear combining scheme used by Cooper and Nettleton [6]. In 1982, Yue summarized the state of spread-spectrum mobile radio [9].

In 1980, Goodman, Henry, and Prabhu proposed a FH spread-spectrum system with M-ary FSK modulation (FH/MFSK) [5]. The optimum receiver for FH multiple access systems with either noncoherent or differentially coherent detection was presented by Yue in 1982 for mobile-to-base transmission [7]. It was shown that with optimal receivers, the FH/MFSK system supported significantly more users that the FH/DPSK system. Similarly, in 1983 Viswanathan and Gupta compared the performance of maximum likelihood, hard-limiting and linear combining receivers for FH/MFSK base-to-mobile transmission [8]. For both base-to-mobile and mobile-to-base transmission, the conclusion was that the hard-limited receiver performed only slightly worse that the maximum likelihood receiver, while the linear combining receiver did not perform very well at all. As a result of these and other
studies, most of the subsequent literature considers FH/MFSK systems, examining various issues such as the interference characteristics. For example, in 1985, Agusti analyzed the effect of multiple tone interferers in a FH/MFSK system [13]. This analysis was similar to the one by Matsumoto and Cooper [4] for FH/DPSK systems and demonstrated the superior performance of the FH/MFSK system.

In 1983, Muammar and Gupta have examined the performance of space diversity for FH/MFSK systems in the presence of Rayleigh fading [10]. They showed that by using space diversity to lower the signal-to-noise ratio requirements, the capacity could be increased by about 25%. An analysis was presented by Viswanathan and Gupta in 1983 for the capacity of base-to-mobile transmission for FH/MFSK cellular systems [8]. For base-to-mobile transmission perfect synchronization can be assumed amongst the various users. They concluded that power control was necessary, because without it the number of simultaneous users the system can support drops significantly due to adjacent cell interference. They also concluded that when the users are uniformly distributed in the cells, the spectral efficiency of a cellular system is reduced by about 30 to 40 percent compared to an isolated cell. Unfortunately, a similar study is unavailable for mobile-to-base transmission.

Niyonizeye, Lecours, and Huynh have evaluated the effect of multiple-access interference in a FH binary FSK system [12]. They concluded that asynchronous FH/BFSK did not perform as well as FH/DPSK. However, a chip synchronous version of the system performed nearly as well as a FH/MFSK system.

The more recent spread-spectrum cellular mobile radio systems are second generation European systems that were very actively studied as part of an effort to propose to GSM of CEPT (Conférence Européene des Administrations des Postes et des Télécommunications) alternatives for the European standard to be specified in 1987 and fielded in 1991 [14]. These systems included MATS-D [15], studied at Philips, SFH900 [16] developed at LCT in France, and CD 900 also studied in France. Independent of these efforts, Ericsson of Sweden has developed the DMS 90 system [17], [18], similar to SFH 900, and has field tested it in 1986. All of these systems are highly sophisticated and in an advanced stage of development. The most recent spread-spectrum cellular system is the GSM-system which is very similar to the DMS 90 and SFH 900 systems developed by Ericsson and LCT.
respectively, in 1986. The GSM-system uses time-division multiple-access (TDMA) with a 200 kHz carrier separation. A brief description of these systems, along with others, is given in the sequel.

6.3 Characteristics and Advantages of Spread-Spectrum Mobile Radio

Most of the proposed spread-spectrum systems for land mobile radio are FH as opposed to DS systems. Regardless of this fact, there are many relative advantages and disadvantages of DS versus FH spread-spectrum systems. The most important of these with regard to cellular mobile radio are listed below.

- DS systems require a contiguous portion of bandwidth, while FH systems can operate over disjoint sections of bandwidth.

- The chip rate in a spread spectrum system is the rate of the fastest clock in the system. The processing gain is the ratio of the chip rate to the symbol (bit) rate, i.e., \( N = T_c/T_b \). In DS spread spread-spectrum systems, the processing gain is limited by the largest allowable chip rate, while the processing gain in a FH system is independent of the chip rate. Hence much larger processing gains can be achieved in FH systems.

- It is easier to obtain coarse synchronization (synchronization of the spreading sequence) in FH systems.

- Coherent detection is usually used in DS systems while noncoherent detection is commonly used in FH systems. This tends to favor DS systems, because coherent detection is inherently 3 dB more efficient than noncoherent detection.

- One disadvantage of FH spread-spectrum systems is that they are not suitable for overlays on existing single channel per carrier (SCPC) cellular systems. If FH systems were to be implemented, they would require a separate, but not necessarily contiguous, section of bandwidth from the SCPC systems. DS systems, on the other hand, would be suitable for an overlay on an existing SCPC system, although no one has attempted to do this.
DS systems are more susceptible to the near-far problem than FH systems. The near-far effect occurs in all systems that use a modulation technique which results in bandwidth expansion. The near-far effect is especially prevalent on the uplink where mobiles close to the base station capture the base station receiver and swamp out the mobiles farther away. To mitigate the near-far effect, it is necessary to use power control to compensate for the effects of path loss and shadowing. Power control ensures that the signals arriving at the base station from the mobiles have the same average received signal power. The near-far problem would not present on the downlink of a DS spread-spectrum system for reasons to be outlined in the discussion of the MATS-D system. Also, power control has been successfully implemented in SCPC cellular systems such as AMPS [19], so power control is not a technological problem. Furthermore, a positive side effect of using power control is that it makes the performance of the system less susceptible to changes in the propagation law (path loss).

In summary, the relative advantages of DS and FH systems is a mixed bag. However, from the above considerations, as well as others, FH systems are considered by many to be more suitable than DS systems for the land mobile radio environment. The main advantages of frequency hopping with regard to cellular systems can be summarized as follows.

- **Frequency Diversity** FH systems can operate quite easily within the coherence bandwidth of the channel. One method of obtaining frequency diversity with FH systems is to use fast frequency hopping (FFH), where each modulated symbol is transmitted at several hop frequencies. This is the same as using a simple repetition code with one code symbol transmitted at each hop. A diversity advantage is obtained because the fade levels at different hop frequencies are independent. A more effective method of improving the performance is to use an error correcting code that is more powerful than the simple repetition code. There are many well studied diversity combining techniques for FFH systems, and the reader is referred to [1] if necessary. Because FFH is just a repetition code, most of these diversity combining techniques can be easily altered for other error correcting codes.

The other method of obtaining a frequency diversity advantage is to use slow frequency hopping (SFH) with coding and symbol interleaving. Interleaving destroys the chan-
nel memory which is beneficial since most channel coding schemes are designed for memoryless channels.

- **Reduced Susceptibility to Interference** Collisions between two transmitters using the same frequency does not reoccur systematically at the next hop. The hop sequences are designed with low cross-correlations so that this does not occur.

- **Cell Reuse Factor** Because of the mitigation of co-channel interference from the use of frequency diversity and the reduced susceptibility to narrow-band interference, smaller reuse factors may be used resulting in a possible gain in spectral efficiency.

- **Channel-Access** There is no channel access delay since each user can access the channel at any time. There is no hard limit on the number of users the system can handle. There is a gradual degradation in performance as the number of users increases.

- **Privacy** Since each user is assigned a unique spreading sequence some degree of privacy is inherently provided.

- **Compatibility** Frequency-hopped systems may exist in the same frequency band as narrow-band systems without excessive mutual interference. In fact, frequency hopped systems do not even require a contiguous section of bandwidth so that conventional narrow-band channels could be dispersed amongst the hop frequencies. However, it would not be desirable to use frequency hopping as an overlay on a narrow-band system, especially for large user populations.

Regardless of the popularity of FH system, DS spread-spectrum is still a viable option. For example, MATS-D uses DS spread-spectrum on the downlink with narrow-band FDM/GTFM on the uplink [15]. A typical coherence bandwidth of an urban channel is in the neighborhood of 100 $kHz$. Direct-sequence systems would have a bandwidth considerably larger than this and could, in fact, have a bandwidth of 25 $MHz$. With DS it is possible to resolve the channel paths, as shown in Fig. 2.1, because of the orthogonality of the spreading sequences being used. In theory, this leads to a diversity advantage, sometimes labeled “spread-spectrum diversity” or “multipath diversity”, because each resolved channel path is independently faded. The receiver that does this is referred to as a “RAKE receiver”
Fig 6-4 Structure of a Typical RAKE Receiver, from [1]

, an example of which is shown in Fig 6.4. The term RAKE is not an acronym, but comes from the similarity of the action of a RAKE receiver and an ordinary garden rake.

A practical RAKE receiver could only process a few of these paths. The rest would be rejected resulting in a somewhat diminished performance. An advantage of DS code division multiplexing (CDM) over time division multiplexing (TDM), is that the symbol duration is maintained by the former, while the later reduces the symbol duration by the multiplexing factor. Therefore, TDM systems require an equalizer for baud rates above 100 KHz, the coherence bandwidth of a typical urban channel, while CDM systems may not.
6.4 Particular Systems Using Spread-Spectrum

In this section several specific systems are presented that use SSMA and TDMA. These examples provide good illustrations of the problems and solutions associated with spread-spectrum and TDMA systems.

6.4.1 MATS-D System

MATS-D uses SCPC FDMA with generalized tamed frequency modulation (GTFM) and a channel spacing of 25 KHz on the uplink. On the downlink, a combination of code-division and time-division multiple-access is used. The reason for using different multiplexing/modulation techniques on the uplink and downlink is that the interference characteristics are different in each direction. Consequently, compromises would have to be made in at least one direction if the same modulation and multiplexing technique were used in both directions.

Synchronous DS spread-spectrum is used on the downlink because the information streams directed to the mobiles can be synchronously multiplexed and modulated onto a common carrier. The signal-to-multiple-access-interference ratio of all the downlink channels that are code division multiplexed on a common carrier is constant and independent of the mobile location. This is convenient because it mitigates the near-far problem commonly plaguing DS spread-spectrum systems. In contrast, signals from the mobiles are basically asynchronous and arrive at the base station with different received power levels. Therefore, it would be more difficult to use DS spread-spectrum on the uplink.

The MATS-D system uses a spread bandwidth of 750 KHz with a 1.248 Mchips/s transmission rate. Within that bandwidth, 32 channels are simultaneously carried using quadrature modulation. Each channel consists of 16 Kbps speech with 2 Kbps coding redundancy on the most important bits in the source sequence. Any fraction of the 32 channels can be allocated to data transmission providing a variable bit rate ranging from 125 b/s to 576 Kbps.

In the receiver, space diversity is used (two antennas) with maximal ratio combining. Each of the two antenna outputs is processed by a separate RF receiver. With a spread-spectrum bandwidth of 750 KHz, there are as many as 8 or 9 resolvable channel paths for
a 100 KHz coherence bandwidth. A clever scheme is used where the channel path with the largest instantaneous signal-to-noise ratio is selected resulting in a form of time selective diversity combining. This form of diversity combining is a few dB less efficient than equal gain combining, but the receiver is much simpler.

In MATS-D, the expected delay profile characteristic of the channel is used to advantage in designing the spreading sequences. The codes have to be designed such that the autocorrelation and cross-correlation sidelobes are small in the time interval during which delayed signals with significant power levels are expected. Because all sequences transmitted by the base station are synchronous, only those sidelobes corresponding to shifts of ±10 chips are significant. For larger delays, the sequence correlation properties can be relaxed. This is a nice property since it reduces the length required for a given number of distinct spreading sequences. This in turn reduces the coarse acquisition time (time to acquire the phase of the spreading sequence) in the receiver.

At the base station, two-branch space diversity is used with post-detection switched diversity combining. The base station receiver uses discrimination detection followed by maximum likelihood sequence estimation.

Table 6.1 summarizes some of the parameters of the MATS-D system

6.4.2 SFH900 System

The SFH900 system uses a combination of SFH and TDMA, along with Viterbi quasi-coherent demodulation of GMSK and a concatenated coding scheme. A description of the system is provided in Table 6.2. The system uses a Viterbi equalizer with 16 states.

An experimental network was constructed in 1986 and demonstrated good performance. The spectral efficiency is about 1.42 channels/MHz/km².

6.4.3 CD 900 System

CD 900 is a broad-band system developed under contract by the French PTT. Similar to MATS-D, CD 900 uses direct-sequence spread-spectrum with a very small processing gain of about 2.67, resulting in a chip rate of 4 Mchips/s. The CD 900 receiver uses spread-spectrum diversity as in MATS-D. Although CD 900 uses about 100 KHz bandwidth
Table 6-1
Parameters of the MATS-D System

<table>
<thead>
<tr>
<th></th>
<th>uplink</th>
<th>downlink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier separation</td>
<td>25 KHz</td>
<td>1.25 MHz</td>
</tr>
<tr>
<td>Channels per carrier</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>Cells/cluster</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Frequency reuse distance</td>
<td>Not available</td>
<td>Not available</td>
</tr>
<tr>
<td>Channels per cell sector</td>
<td>Not available</td>
<td>Not available</td>
</tr>
<tr>
<td>Traffic density</td>
<td>Not available</td>
<td></td>
</tr>
<tr>
<td>Transmission data rate</td>
<td>19.5 Kb/s</td>
<td>1.248 Mc/s</td>
</tr>
<tr>
<td>Modulation</td>
<td>GTFM</td>
<td>QAM</td>
</tr>
<tr>
<td>Speech coder rate</td>
<td>16 Kb/s</td>
<td>16 Kb/s</td>
</tr>
<tr>
<td>Speech codec</td>
<td>RELP</td>
<td>RELP</td>
</tr>
<tr>
<td>Diversity methods</td>
<td>space diversity</td>
<td>spread-spectrum diversity</td>
</tr>
<tr>
<td></td>
<td>channel coding (rate 1/8)</td>
<td>channel coding (rate 1/8)</td>
</tr>
</tbody>
</table>

Table 6-2
Parameters of the SFH900 System

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier separation</td>
<td>143 KHz</td>
</tr>
<tr>
<td>Channels per carrier</td>
<td>3</td>
</tr>
<tr>
<td>Cells/cluster</td>
<td>3</td>
</tr>
<tr>
<td>Frequency reuse distance</td>
<td>Not available</td>
</tr>
<tr>
<td>Channels per cell sector</td>
<td>Not available</td>
</tr>
<tr>
<td>Traffic density</td>
<td>≈ 1.42 channels/MHz/km²</td>
</tr>
<tr>
<td>Transmission data rate</td>
<td>200 Kbps</td>
</tr>
<tr>
<td>Modulation</td>
<td>GMSK (BT₀ = 0.3)</td>
</tr>
<tr>
<td>TDMA time bursts</td>
<td>1.23 ms</td>
</tr>
<tr>
<td>Speech coder rate</td>
<td>16 Kbps</td>
</tr>
<tr>
<td>Speech codec</td>
<td>SBC</td>
</tr>
<tr>
<td>Diversity methods</td>
<td>adaptive Viterbi equalization channel coding (Reed-Solomon)</td>
</tr>
<tr>
<td></td>
<td>slow frequency hopping (250 hops/s)</td>
</tr>
</tbody>
</table>
carrier separation
Channels per carrier
6
Cells/cluster
Not available
Frequency reuse distance
Not available
Channels per cell sector
Not available
Traffic density
Not available
Transmission data rate
1.5 Mb/s
Modulation
QPSK
TDMA time bursts
Not available
Speech coder rate
12.8 Kb/s
Speech codec
SBC
Diversity methods
channel coding (Reed-Solomon)
spread-spectrum diversity

Table 6.3
Parameters of the CD 900 System

per 16 Kb/s digital channel, it is claimed that there is no penalty in frequency economy compared to an analog FDM/FM system [20]. This is verified by a spectral efficiency study, the results of which are included in Table 7-1. Further details of the CD 900 system are included in Table 6.3.

The CD 900 system requires an equalizer, because severe ISI is expected at the transmission rate being used. Also, because of the mobility of the network transceivers, the channel transfer function is constantly changing. It must be assumed that in the time interval between slots the channel characteristics change completely. Therefore, a separate measure of the channel transfer function must be made for each time slot. This implies a lot of computing. In the CD 900 system, the length of each time slot is 0.5 ms and the slots are repeated every 31.5 ms (63 slots per frame). Fortunately, the required computing can be carried out in the time between slot reoccurrences. Techniques for channel monitoring are discussed in [20].

It can be assumed that the channel transfer function remains constant for a slot duration. The validity of this assumption depends on the vehicle velocity. Recall that the coherence time of the channel is inversely proportional to the vehicle velocity. For example, consider a vehicle with a speed of 25 m/s (60 mph) and a radio frequency of 800 MHz. In
the slot period of 0.5 ms, the vehicle will move through 12.5 mm or 0.35 of a wavelength equivalent to a phase shift of 13°. This phase shift may or may not be significant, but there will be a significant phase shift at higher vehicle velocities.

6.4.4 DMS 90 System

DMS 90 is a TDMA system developed by Ericsson, and is similar to SFH 900. DMS 90 has 300 KHz carrier separations. Ten users are multiplexed on each carrier and transmit in 0.8 ms bursts every 8th ms. In each direction, there are 810 channels in the 25 MHz bandwidth. The data rate for each user is 16 Kb/s and speech is encoded using a 16 Kb/s subband codec. Further details on the DMS 90 system are given in Table 6.4 below.

GMSK modulation is used with $BT_b = 0.25$. Frequency hopping is used, as in the SFH900 system, with a hop rate of 125 Hops/s. A Reed-Solomon (12,8) code is used over $GF(2^4)$.

DMS 90 allows for the use of adaptive equalization of which there are two types:

- adaptive decision feedback equalizers (DFE).
- adaptive Viterbi (maximum likelihood sequence estimation) equalizers.
Table 6-5
Parameters of the GSM-System

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier separation</td>
<td>200 KHz</td>
</tr>
<tr>
<td>Channels per carrier</td>
<td>6</td>
</tr>
<tr>
<td>Cells/cluster</td>
<td>3</td>
</tr>
<tr>
<td>Frequency reuse distance</td>
<td>Not available</td>
</tr>
<tr>
<td>Traffic density</td>
<td>104</td>
</tr>
<tr>
<td>Channels per cell sector</td>
<td>Not available</td>
</tr>
<tr>
<td>Transmission data rate</td>
<td>270 Kbps</td>
</tr>
<tr>
<td>Modulation</td>
<td>GMSK</td>
</tr>
<tr>
<td>TDMA time bursts</td>
<td>0.58 ms</td>
</tr>
<tr>
<td>Speech coder rate</td>
<td>13Kbps</td>
</tr>
<tr>
<td>Speech codec</td>
<td>RPE</td>
</tr>
<tr>
<td>Diversity methods</td>
<td>adaptive equalization</td>
</tr>
<tr>
<td></td>
<td>channel coding (Convolutional)</td>
</tr>
<tr>
<td></td>
<td>slow frequency hopping (217 hops/s)</td>
</tr>
</tbody>
</table>

In an experimental DMS 90 system, a DFE was used and the performance from field trials was shown to be as good as a conventional 16 Kbps FDMA system. Even better performance can be achieved with the Viterbi equalizer because it can handle twice as much dispersion as the DFE.

6.4.5 GSM-System

In August 1987, thirteen European countries signed an agreement for a common cellular mobile radio system to be in operation in 1991. The Pan-European system is called the GSM-system [18], named after the CEPT group responsible for the specification, and will allow roaming throughout the whole of Europe.

In the GSM-system, eight users are multiplexed on each carrier. In addition, slow frequency hopping is used with interleaved channel coding and GMSK modulation. Frequency hopping is an add on feature intended to enhance the capacity of the system for reasons outlined in the next section. The GSM-System is similar to SFH 900 and DMS 90, and has been adopted as the European standard. The basic characteristics of this system are listed in Table 6.5.
The channel coder is a constraint length 4 convolutional code. The code is interleaved over 8 time slots (or 8 hops). More sensitive source bits are protected more than the less sensitive ones. For the most sensitive source bits, the code rate is 2/3. The GSM-system can operate at very low $E_b/N_0$ values, typically about 10 dB. Furthermore, a $C/I$ ratio of 10 dB is sufficient, which is about 8 dB lower than the analog AMPS system. This results in better capacity due to better channel reuse. In the GSM-system, a 3 cell per cluster structure is used as opposed to a 7-cell reuse pattern in AMPS. The number of channels per cell sector is increased from 45 to 104 over the AMPS system.

Since the GSM-system is very new, little additional information is present in the literature currently available. However, contributions in very recent (e.g. 38th IEEE Vehicular Technology Conference, 15-17 June 1988) technical conferences tend to indicate that there will be significant literature available on the GSM-system in the near future.

6.5 Comparison of Spectral Efficiency

Evaluation of the spectral efficiency of spread-spectrum systems is very complicated illustrated, for example, by the efficiency analysis of the SFH 900 system [16]. Furthermore, it is very difficult to find a fair evaluation of the spectral efficiency of spread-spectrum systems. There is no set standard in the literature for comparing these systems amongst themselves or other multiple-access schemes. Some authors, in particular Muilwijk [21], claim outright that spread-spectrum is less efficient because it is asynchronous. However, this is not always true. With MATS-D, the information transmitted from the base station to mobiles is available in trunked form and therefore can be synchronously transmitted on a common carrier. Also, a recent and detailed study of the SFH900 system tends to suggest that spread-spectrum can provide an improvement in spectral efficiency over conventional FDM/FM systems [16].

The measure of spectral efficiency commonly used in the literature for spread-spectrum systems is defined as the number of users per cell per 30 KHz of bandwidth, i.e.,

$$\eta_1 = U \frac{30 \text{ KHz}}{W}$$

where $m$ is the number of cells in the system, $U$ is the average number of users per cell, and $W$ is the one-way transmission bandwidth. In a multicell system $\eta_1$ is reduced by the
reuse factor. That is

\[ \eta_m = \frac{\eta_1}{F_r}. \]

As a word of caution, note that the above definition of spectral efficiency does not account for the spatial density of the users.

For a conventional FMD/FM system having a channel spacing of 25 KHz and with 21 cells/cluster, as proposed by Ericsson, \( \eta_m = 5.71\% \) [17]. Ericsson’s FMDA system with 25 KHz wide radio channels, 9 cells/cluster and 16 Kbps digital voice coding has an efficiency of \( \eta_m = 13.33\% \) [17]. Finally, the DMS 90 system developed by Ericsson has a spectral efficiency of \( \eta_m = 32.4\% \). It was shown by Yue [7], that the spectral efficiency of the system proposed by Cooper and Nettleton was 8.4%. Unfortunately, it is difficult to draw conclusions from these numbers. A more useful method of determining and comparing spectral efficiencies of various modulation and multiplexing schemes is presented in the next chapter.

6.6 Unresolved Issues

There still remain several open problems that must be solved to obtain efficient spread-spectrum cellular systems. The problem of designing spread-spectrum cellular systems can be divided into two parts; the first is the unique aspects associated with the cellular application, and the second is the design of the spread-spectrum multiple-access system within each of the cells. Both of these aspects require further research.

Within a cell, spread-spectrum cellular systems are essentially spread-spectrum packet radio networks. Spread-spectrum packet radio networks are used in a variety of applications other than land mobile radio, ranging from indoor wireless networks to satellite relay networks. Although single-user spread-spectrum systems are fairly well understood, the design of spread-spectrum networks is a new area of research with relatively few published results. It is still unclear, for example, how to optimize the tradeoff between the processing gain and code rate, and the best choice of retransmission policy. These issues are currently under investigation [23].

The design and analysis of effective protocols for voice-data transmission in spread-spectrum cellular systems is also important. Currently, cellular systems in North America
are only used for voice transmission. However, it may be desirable to provide (limited) ISDN capability in the future.

6.7 Technology Requirements

The biggest drawback of spread-spectrum systems has always been the cost. Typically the hardware cost is very expensive because hopping frequency synthesizers are required. The cost of these synthesizers, however, depends upon the hop rate. The SFH900 system alleviates this problem by hopping at a rate of only 250 hops/s. Settling times for current, state of the art, frequency synthesizers are on the order of 4 – 5 μs. The SFH900 system allows 100 μs for settling time. A prototype of this system indicated that fixed and mobile equipment can be produced at a low cost with today’s technology [16].

The MATS-D system uses DS spread-spectrum. Prototypes of this particular system have been constructed using, whenever possible, commercially available components. A technology study for the wide-band DS transmission scheme used in MATS-D has shown that the complexity of implementation will be comparable to that of the analog FDM/FM systems currently in use.

The DMS 90 and, obviously, the GSM-system are economically feasible. Both of these systems use slow frequency hopping, with 125 hops/s for the former and 217 hops/s for the latter. It has also been proposed by Ericsson that the GSM-system use a 16-state Viterbi equalizer [18]. The complexity of this equalizer is within the capability of CMOS technology for low power hand-held portables.

6.8 Concluding Remarks

In this chapter, spread-spectrum has been considered as a possible multiple accessing technique. Some basic conclusions concerning spread-spectrum are as follows:

- There are three basic multiple-access techniques, time division multiple-access (TDMA), frequency division multiple-access (FDMA) and spread-spectrum multiple-access (SSMA). SSMA by itself is less efficient than either TDMA or FDMA because SSMA is an asynchronous nonorthogonal multiple-access technique. However, SSMA may make up for
the reduction in multiple-access efficiency by reducing the $C/I$ requirements on the channel.

- Most spread-spectrum systems that are proposed for mobile radio system use frequency hopping as opposed to direct sequence spread-spectrum signaling. Several second generation European systems as well as the Pan-European GSM-System use frequency hopped spread-spectrum.

- Spread-spectrum is very useful as an add-on feature for mature cellular systems to increase capacity. The amount of capacity improvement that can be obtained depends upon the particular system under consideration. The main reason for achieving the capacity increase is the inherent frequency diversity that is present in frequency hopped spread-spectrum systems. Achieving this diversity advantage requires a digital modulation scheme with interleaving and forward error correction coding.

- FH spread-spectrum should not be used as an overlay on an existing analog FDM/FM cellular system. Unacceptable interference will result for the subscribers using analog FDM/FM.

- It is possible to use direct sequence spread-spectrum as an overlay on an existing cellular system. The introduction of such an overlay will essentially increase the level of background noise present on the channel. Each additional spread-spectrum user adds an equal amount of background noise. The power spectral density of the background noise is simply the transmitted power divided by the spread-spectrum bandwidth. Since most cellular mobile radio systems are interference limited, this increase in the level of background noise may not have a significant effect on the performance of the subscribers using FDMA or TDMA.

- Direct sequence spread-spectrum systems could use a RAKE receiver for effective performance. The RAKE essentially provides a diversity gain.

- Direct sequence spread-spectrum systems would require power control on the uplinks because of the near-far problem. Power control is not required on the downlink.
References


Evaluating Spectral Efficiencies

7.1 Introduction

To date, many methods have been proposed for evaluating the spectral efficiency of proposed cellular systems. These methods range from pure speculation, sophisticated mathematical derivations, simulations, to field trials. This chapter presents a thorough approach for evaluating the spectral efficiency of cellular land mobile radio systems. The approach accounts for all the pertinent factors within a cellular network and is based upon the results presented by Lee [1] and Hammuda et al. [2].

7.2 Measures of Spectral Efficiency

Two measures of spectral efficiency appear to be adequate and appropriate for cellular mobile radio systems. The first is

\[ \text{Erlangs/MHz/Km}^2. \] (7.1)

The second, and perhaps more conceptual measure, is in terms of

\[ \text{Voice Channels/MHz/Km}^2. \] (7.2)

These two measures of spectral efficiency are directly related. The conversion from one to the other is easily obtained given the blocking probability and holding time on the channels in the system. Only the second measure is discussed in detail in this report.

To compare the spectral efficiencies of different cellular systems, the quality of service must be included. The service quality basically consists of three categories.

- **Coverage** The percentage of the total area for which service is available.

- **Grade of Service** The blocking probability or waiting time for requested service.

- **Interference Levels** How much interference the system can sustain for a particular degree of voice quality.

Only the last two are pertinent, and therefore a reference must be established in terms of the grade of service and voice quality for comparing different systems.
In a cellular land mobile radio system two main parameters determine the spectral efficiency, the modulation technique and the multiple-access technique. The efficiencies of the modulation and multiple-accessing techniques can be decoupled and treated separately.

7.3 Efficiency of Modulation Techniques

Assume the following definitions:

\[
\eta_m := \text{Modulation efficiency in channels/MHz/Km}^2 \\
B_t := \text{Total bandwidth available to the system in MHz} \\
B_c := \text{Voice channel spacing in MHz} \\
N := \text{Number of cells per cluster} \\
A := \text{Cell area in Km}^2
\]

The spectral efficiency of a modulation scheme is given by

\[
\eta_m = \frac{\text{Total number of channels available to the system}}{\text{Total available bandwidth} \cdot \text{Cluster area}} = \frac{B_t}{B_t(N \cdot A)} \frac{B_t}{B_c(N \cdot A)} = \frac{1}{B_c(N \cdot A)}
\]

From equation (7.3), the efficiency \( \eta_m \) is the inverse product of two factors, the channel spacing \( B_c \) and the cluster area \( (N \cdot A) \). By using cell splitting the cluster area can be made very small, hence increasing efficiency. Cells as small as 1 Km may be possible with today’s technology. However, decreasing the cell size increases the cost of the infrastructure.

The relative efficiency between two modulation systems \( x \) and \( y \) is

\[
\eta_r = (B_c)_y N_y A_y \\
(7.4)
\]

Since a cellular system is interference limited, the cell area is independent of the amount of power transmitted and the modulation scheme used [2]. Therefore,

\[
\eta_r = (B_c)_y N_y \\
(7.5)
\]
Obviously, the number of cells per cluster is an important parameter. The co-channel reuse factor is defined as \( Q = D/R \), where \( D \) is the minimum co-channel cell separation and \( R \) is the cell radius. For a hexagonal geometry \( Q = \sqrt{3N} \) where \( N \) can only take on values 1, 3, 4, 7, 9, 12, 13 \( \cdots \) etc.

As mentioned, the spectral efficiency is dependent upon the subjective voice quality. The voice quality in turn specifies the required carrier-to-interference ratio \( (C/I) \). The value of \( C/I \) is a function of \( Q \). This functional dependency is dependent upon the interference model being used.

In general, however, two main categories of interference models are discussed in the literature. The first group is based solely on geographical considerations, while the second is based on statistical models that include the effects of path loss and shadowing. It was concluded by Hammuda et. al. [2] that a geometric model that accounts for the first tier of interferers (6 interferers) is the best for the following reasons.

- Geometric models are easier to use. The main drawback of the statistical models are their complexity and in the unrealistic assumption of a single interferer. Attempting to account for more than one interferer makes them even more difficult to use.

- The geometric model with six interferers (closest neighbors) is a good compromise between a model with one interferer and a model with an infinite number of interferers.

- The effect of path loss and shadowing can be included in the \( C/I \) value, by performing subjective measurements for various modulation schemes under realistic channel conditions.

- The statistical models tend to give pessimistic results.

Fig 7.1 compares the \( C/I \) requirements for the different co-channel interference models.

Using the geometric model with six interferers gives [1,2] \( C/I = Q^\alpha/6 \) resulting in \( C/I = (3N)^{\alpha/2}/6 \), where \( \alpha \) is the propagation law, and typically \( 3 < \alpha < 5 \). Hence, the efficiency is

\[
\eta_m = \frac{3}{B_c \cdot [6 C/I]^{\alpha/2} \cdot A}
\]

(7.6)
Fig. 7-1 Comparison of Various Co-channel Interference Models, from [2]
and the relative efficiency is

\[ \eta_r = \frac{(B_c)_v \cdot (C/I)_v^{2/\alpha}}{(B_c)_x \cdot (C/I)_x^{2/\alpha}}. \]  

(7.7)

Suppose that two systems have the same efficiency so that \( \eta_r = 1 \). Then, by using (7.7), we can write

\[ \frac{(B_c)_x}{(B_c)_v} = \left( \frac{(C/I)_x}{(C/I)_v} \right)^{\alpha/2}. \]  

(7.8)

With \( \alpha = 4 \), equation (7.8) states that on the basis of equal spectral efficiency, if the bandwidth is reduced by a factor of 2, the required \( C/I \) must increase by 4 times to maintain the same quality.

### 7.4 Efficiency of Multiple-Access Techniques

There are three basic kinds of multiple-access techniques, (i) frequency division multiple access (FDMA), (ii) time division multiple-access (TDMA) and, (iii) code division multiple-access (CDMA). All of these have been discussed earlier in the report. All of these techniques have an efficiency of unity, provided that the signals transmitted by the different users in the system are orthogonal. However, this will be difficult, if not impossible, to achieve. For example, with CDMA the spreading sequences for asynchronous transmission are not orthogonal. Also, FDMA and TDMA require guard bands and guard times, respectively.

For FDMA, the efficiency is

\[ \eta_t = \frac{B_c \cdot M_a}{B_t} \leq 1, \]  

(7.9)

where \( M_a \) is the total available number of voice channels.

For wide-band TDMA, the efficiency is

\[ \eta_t = \frac{r \cdot M_t}{T} \leq 1, \]  

(7.10)

where \( r \) is the slot duration, \( T \) is the frame duration, and \( M_t \) is the number of slots available for voice transmission in a frame. \( M_t \) may not be equal to the total number of slots in a frame, because some of them may have to be dedicated to synchronization sequences and other purposes.

For narrow-band TDMA,

\[ \eta_t = \frac{r \cdot M_t}{T} \cdot \frac{B_u \cdot M_u}{B_t} \leq 1, \]  

(7.11)

\[ 99 \]
Table 7-1 Spectral Efficiency of Some Cellular Schemes, from [2]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Channel occupancy (kHz)</th>
<th>required ( C/I ) (dB)</th>
<th>Modulation efficiency (Ch/Hz/Km)</th>
<th>Relative efficiency (to 25kHz/FM)</th>
<th>Trunking efficiency</th>
<th>Overall efficiency (E/Hz/Km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TACS/FM</td>
<td>25</td>
<td>18–19</td>
<td>1.82</td>
<td>1.0</td>
<td>&gt;0.95</td>
<td>1.47</td>
</tr>
<tr>
<td>AMPS/FM</td>
<td>30</td>
<td>17–18</td>
<td>1.52</td>
<td>0.84</td>
<td>&gt;0.95</td>
<td>1.19</td>
</tr>
<tr>
<td>ACSB/5kHz</td>
<td>5</td>
<td>20</td>
<td>7.10</td>
<td>3.90</td>
<td>&gt;0.95</td>
<td>6.57</td>
</tr>
<tr>
<td>CDMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDM/DMA/Digital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFH800</td>
<td>72</td>
<td>11–12</td>
<td>1.47</td>
<td>0.81</td>
<td>~0.80</td>
<td>0.94</td>
</tr>
<tr>
<td>Spread Spectrum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSS-90</td>
<td>10</td>
<td>7</td>
<td>1.42</td>
<td>0.78</td>
<td>~0.75</td>
<td>0.83</td>
</tr>
<tr>
<td>GTS/FDMA/Digital AT&amp;T Prop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDMA/Digital</td>
<td>10</td>
<td>15</td>
<td>7.96</td>
<td>4.37</td>
<td>&gt;0.95</td>
<td>6.21</td>
</tr>
<tr>
<td>DSS-90</td>
<td>30</td>
<td>10–12</td>
<td>3.54</td>
<td>1.95</td>
<td>~0.80</td>
<td>2.58</td>
</tr>
</tbody>
</table>

where \( B_u \) is the bandwidth available to a user during its slot time and \( M_u \) is the number of users sharing the same time slot, but having access to different frequency bands.

For CDMA, the efficiency is much more difficult to calculate because it depends upon the spreading sequences and spreading technique (direct-sequence or frequency-hopped) being used. However, the efficiency can be written as

\[
\eta_t = \frac{r \cdot M_t \cdot B_u \cdot M_u}{T \cdot B_t},
\]

(7.12)

The efficiency of CDMA may be as low as 0.70.

The overall spectral efficiency is simply given by the product of the modulation and multiple-accessing efficiencies

\[
\eta = \eta_m \eta_t.
\]

(7.13)

Table 7.1 compares the spectral efficiency of several existing and proposed cellular systems for a cluster area \( A = 3 \, \text{Km} \) and propagation law \( \alpha = 4 \).

### 7.5 Spectral Efficiency of Digital Cellular Systems

For a digital cellular system, two parameters must be specified, the bandwidth occupancy and the required \( C/I \). One way of doing this is to define an equivalent channel spacing as
follows:
\[
\text{Equiv. Ch. Spac.} = \frac{\text{Voice channel bit rate (Kb/s)}}{\text{Modern speed (b/s/Hz)}}. \tag{7.14}
\]
The required C/I is then subjectively determined. The values are then used in the equations in section 7.3.

For a digital cellular system, we have
\[
E_b R_b = E_s R_s = C, \tag{7.15}
\]
where \(E_b\) is the energy per bit, \(R_b\) is the bit rate, \(E_s\) is the energy per modulated symbol, \(R_s\) is the symbol rate, and \(C\) is the carrier power. Consider two systems having \(\eta_r = 1\).

Then using (7.15) in (7.8) gives
\[
\frac{(E_x R_x)_x}{(E_x R_x)_y} = \frac{(B_c^{\alpha/2})_y}{(B_c^{\alpha/2})_x}. \tag{7.16}
\]
However the interference level is the same independent of the different modulation schemes. Also, there is a linear relationship between \(R_s\) and \(B_c\). That is
\[
(R_s)_x = k (B_c)_x, \quad (R_s)_y = k (B_c)_y \tag{7.17}
\]
where \(k\) is a constant. Hence, using (7.17) in (7.16) gives
\[
\frac{(E_x)_x}{(E_x)_y} = \frac{(B_c^{\alpha/2+1})_y}{(B_c^{\alpha/2+1})_x}. \tag{7.18}
\]
When \(\alpha = 4\) equation (7.34) states that if \(B_c\) is reduced by half the required energy per symbol or bit increases by 8 times. This relationship is important when designing a cellular system.

### 7.6 Determining the Required C/I

The precise definition of the required C/I is [3]: "The level at which 75% of the users state that the voice quality is either good or excellent in 90% of the service area." All required values of C/I must be determined by using this standard, or any other common standard, if meaningful comparisons are to be made. There are basically three ways to determine the C/I requirement.
• **Mathematical Derivation:** This is very complicated and many simplifying assumptions have to be made to make the analysis tractable. It can be very costly.

• **Subjective Measurements:** This involves human perception of the voice quality. Measurements are performed under realistic operating conditions, but the fielding of an entire system is not required. Simulations may be useful for this technique.

• **Field Trials:** This involves an entire fielded system and can be very expensive. It is useful for systems already in operation. Subjective human tests are used to determine the required $C/I$ values.

Of the above three methods subjective measurements are probably the most useful when choosing future systems.

**References**


8 Conclusions and Recommendations

In the preceding chapters we have examined the design of an expanded and enhanced cellular mobile radio telephone system from the point of view of channel characterization, voice coding, analog and digital modulation, multiple access, and overall spectral efficiency. Consideration has been given as to how the system may evolve from the existing system into a future system that will meet the demands of an expanding market.

Here we will attempt to draw on what has been developed above to bring into focus the aspects of the system design which will dominate the development of the system.

First we examine the issue of overall spectral efficiency. It is clear from the results that have been published over the last several years that the dominant parameters in the determination of overall spectral efficiency for a cellular radio system are the signal-to-interference ratio, C/I, required for good operational quality, and the number of cells per cluster. The lower the C/I requirement and the smaller the number of cells per cluster, then the larger the overall spectral efficiency.

Improvement in efficiency can be considered from both the short term point of view and the long term point of view. In the short term, it may be appropriate to modify the existing analog system to gain a small improvement; this is particularly true in cities where the present system is approaching saturation. A short term analog system improvement may be needed for saturated systems prior to the appearance of digital systems. In the longer term it is clear that the existing analog system will be phased out and be replaced by a digital system. This will be necessary to allow the cellular system to be compatible with the ISDN system with which it will have to interact to provide advanced user services. In addition it is becoming clear that the digital cellular mobile system will be more spectrally efficient than the present analog system.

8.1 Short Term Improvements

Improvement techniques mentioned above for analog systems are as follows:

- Reduced Bandwidth Systems
- Overlay/Underlay Systems
Of these techniques the most suitable is the overlay/underlay technique. This approach is very promising since it does not require the modification of the mobile units. A pilot program to experimentally determine its effectiveness should be undertaken if this has not already been done. This overlay/underlay technique will require modification of the switching and handoff techniques and the frequency assignments, but it should be transparent to the subscribers. The factor of 1.57 improvement may justify the effort, and a cost/benefit analysis of this technique should be carried out.

A modification of the system to use spatial diversity is possible. It would be necessary for the mobile units to use two antennas with some form of combining; this may be effective, but no clear study of the ultimate effectiveness of a space diversity technique in the interference limited environment of the cellular mobile system is known. Detailed analysis of diversity improvements have been developed for fading environments. Since the cellular environment is a fading environment the use of diversity is perhaps indicated. However it is not clear how well diversity will perform in the cellular system, because the limiting feature of the cellular system performance is not fading but rather interference. If it were possible to obtain a 6 dB C/I gain with spatial diversity, then a reduced bandwidth system with 15 KHz channels would be possible. Note however that even if this is possible the entire system would essentially have to be reconstructed. For this reason it is not recommended that this be considered as a short term measure. It should be considered as a possibility in a digital implementation of the system along with other extensive improvements for the long term.

The other viable technique for short term improvement may be to proceed with cell splitting strategies originally proposed with the analog system. It is recommended that experimental efforts to determine the sequence of cell sizes suitable for application in cell splitting beginning with the current cell sizes and working down to the minimum usable cell radius. The minimum cell size is to be determined as well.
8.2 Long Term Improvements

For the long term a plan must be developed to move gracefully into a digital cellular mobile system.

8.2.1 Modulation Formats for Digital Transmission

There are a number of voice coder techniques which will provide good quality speech at bit rates of 16 kbits/s. With binary constant envelope modulation and either FDM or TDM multiple access techniques, this results in the equivalent of one digital channel in the same bandwidth as an existing analog channel.

If multilevel modulation is used, e.g. four level modulation, such as QDPSK or QPSK with good signal shaping, then two users may be fit into the equivalent of one analog voice channel.

If a more bandwidth efficient modulation technique based on trellis coding or another form of modulation utilizing memory and multiple bits per symbol can be adapted to the cellular channel environment, then perhaps three or four users may be fit into one analog voice channel. No discussion of such techniques is available in the current literature. It is recommended that these advanced modulation techniques be seriously studied for this application.

In addition, the existing analysis of modulation formats is mostly directed at performance in the presence of additive white gaussian noise and on fading dispersive channels, while the cellular system is limited by interference from the next channel with the same frequency assignment by the very design of the system. Some additional research is called for to assess the performance of the candidate systems in the combination of fading and cochannel interference. This problem will probably have to be dealt with via simulations and actual measurements. It is recommended that the aggressive study of this problem via computer simulation be undertaken.

It may be expected that more efficient speech coding algorithms will be adapted to the cellular environment as better digital signal processing power becomes available in smaller and more power efficient integrated circuits. If a rate of 8 kbits/s is achieved the capacity of a digital system will of course be doubled. It is recommended that the
development of such speech coders be anticipated in the overall system design and that this development be encouraged.

A comparison of existing system recommendations is given in the table on page 100, taken from reference [2] of Chapter 7. This table involves a number of assumptions but it seems to give a fair comparison of the techniques listed. The GSM-system is quite similar to the DMS-90 system listed. The factor of approximately 2 difference in the DMS-90 and the NB/FDMA/ Digital system is apparently due to the fact that the latter is using a multilevel modulation technique QDPSK instead of binary GMSK.

8.2.2 Voice Coding

There are several issues that must be considered in the choice of voice coding techniques for the cellular mobile environment:

- bit rate required for acceptable quality
- performance in the normal automotive background noise
- channel error protection required
- performance in tandem with itself and other voice coders
- provisions for privacy
- complexity
- compatibility with voiceband modems

As discussed in Chapter 3, the bit rate requirements currently range from about 12 kbits/s for a subband coder to around 6 or 8 kbits/s for a CELP coder. The possibility of achieving rates as low as 4800 bits/s is very clear.

The laboratory studies of voice coding may utilize an input signal which is clean compared to a signal obtained in a typical automobile interior. There will be a disruptive noise background due to the traffic and engine audio noise as well as the typical automobile electrical noise background. There will likely be multiple conversations among the occupants.
of the automobile which may contribute to this background condition. The voice coder will need to receive its input in a way that allow it to operate properly in spite of these conditions.

The voice coded systems are rather robust in the presence of channel bit errors for the high bit rate coders and less robust as the natural redundancy of the speech input is more effectively removed by the coder.

As detailed in Chapter 3, the error protection should be imbedded in the voice coder, with more protection assigned to the more important parameters. This is especially true for the current generation of voice coders. As the coders improve to allow more efficient representation of the speech, it will be necessary to provide more complete error protection and more of the final transmitted bit stream will be devoted to error protection than in the present coders; that is, if the coded speech is at a very low rate it will be necessary to protect the data stream almost as if it were an arbitrary data stream.

In any system a final error check and interpolation and muting of the output speech of the character now done in compact audio disc systems will potentially improve quality under fringe conditions and deep fades.

Since the coded speech is a digital data stream, it will be possible to utilize some form of privacy coding or encryption of the stream to assure confidentiality to the user. A very important point in this area is that the encryption is likely to be very sensitive to the bit error rate on the channel. If the voice coder utilizes imbedded error protection, it is not possible simply to tandem an encryption system, perhaps supplied by the user, for the encryption system would see the raw uncorrected channel error rate and degrade catastrophically. The encryption system must be placed in the system so that it sees the corrected or protected data stream. If the correction is imbedded in the voice coder the encryption must be imbedded also. It is recommended that a study be performed to find the most effective, flexible, and user friendly form of encryption for this particular environment.

It is also important to note that most of the proposed voice coders, particularly the more efficient ones, do not tolerate tandem connections with themselves or with other speech digitization systems. This will limit the incorporation of most systems into the existing telephone network. If it is necessary to convert the signal to analog or even 64 kbits/s PCM, and then back to analog or PCM or to 32 kbits/s ADPCM, perhaps several
times, and then eventually back to cellular digital format for a final mobile destination, then quality may be degraded to an unacceptable level. Application in the future ISDN environment may be easier, but even here unless the bit stream is treated as a data stream and its integrity maintained all the way to a final mobile unit some intermediate conversions will have to be tolerated. For a fixed to mobile call the voice digitizer will have accept a PCM or ADPCM (or whatever system is used eventually for fixed stations) input. For a mobile to fixed call the conversion to the fixed station format must be acceptable. If the call is mobile to mobile it either must be identified as such and not converted until it reaches is final destination or it must survive at least two conversions en route with acceptable degradation. None of the proposed systems seems to have addressed this issue at all.

Although complexity of the voice coder is an issue, it is presently possible to implement the coder and the error protection reasonably using one of the currently available special purpose digital signal processing integrated circuits. New error protection coders on single chips are announced frequently. It should be expected that if a volume equal to that of the present population of mobile users is anticipated, it is reasonable to expect most of the systems can be tailored to be implementable in VLSI form with some general purpose DSP chips and perhaps a few special purpose chips.

It is recommended that research address the above issues on a broad front, including the bit rate reduction issue, but not limited to bit rate efficiency, as the other issues listed above may restrict usable rates in themselves. The privacy issue deserves special attention as indicated above. Tandeming problems are also especially important.

As to compatibility with the present generation of voiceband modems it must be recognized that the efficient voice coders make extensive of the nature of the voice signal. The features which make the voice coder efficient also make it incompatible with present voiceband modems. Therefore it is not recommended that any attempt be made to retain this form of compatibility.

8.3 Multiple Access Techniques

The multiple access arrangements to be considered include:

- Time Division Multiple Access (TDMA)
• Frequency Division Multiple Access (FDMA)
• Code Division Multiple Access (CDMA) or Spread Spectrum Multiple Access (SSMA)
• Combined TDMA/FDMA or other combinations

From a purely theoretical or fundamental point of view all of these systems provide the same channel capacity. However, from the point of view of a well engineered, flexible, cost effective implementation which can evolve from the present system, interoperate with the present system, and handle future requirements as needed, there is substantial room for comparison.

Several different multiple access techniques have been proposed, as discussed in Chapter 7. These include a spread spectrum downlink (base to mobile) and FDMA uplink (mobile to base), narrowband FDMA, wideband TDMA, and narrowband TDMA. Past studies have treated wideband spread spectrum systems for the uplink and downlink.

A detailed consideration of overall situation shows that a 30 KHz channel bandwidth is desirable both to make the system more compatible with the present system and to keep the transmitted channel bandwidth well within the coherence bandwidth of the fading cellular channel so that adaptive equalizers will not be required. The narrowband TDMA system is more flexible in today's and tomorrow's environment, and will be less expensive to implement at the base station level and not more expensive in the mobile units. For these and other detailed considerations it is recommended that the narrowband TDMA format be adopted for the next generation of digital cellular systems. The advantages offered include:

• lower cost for the base stations and perhaps the mobiles.
• lower power consumption.
• more flexibility for future evolution with new technology.
• a very graceful transition from the current AMPS system to a digital system.
• flexibility for offering future ISDN-like services.

The following very long term consideration is noted. The cellular mobile system is by design an interference limited system. The effectiveness of spread spectrum systems
is most noted in the ability of the spread system to cope effectively with the intentional jamming interference of an adversary. The systems currently in use have no particular features or capabilities for dealing with interference effectively. There are many obstacles in the way of the application of spread spectrum techniques on a fading multipath mobile network. Nevertheless it would seem that some format should exist for a very effective spread spectrum system. It is recommended that fundamental research to determine this format be undertaken.

8.4 Implementation

There are several issues involved in the implementation of the digital cellular mobile system and in the implementation of any near term expansion of the current AMPS system. The issues of cost, reliability, ease of use, and customer acceptance will apply to any system. In addition, since the cellular system is part of the overall telephone network, future flexibility must also be considered.

The design of an expansion of the AMPS system should if possible be transparent to the mobile units to avoid the expense and customer annoyance of the conversion or replacement of these units. This is particularly true if the expansion is expected to be rapidly superseded by a further move to a digital system. Any expansion technique which would not enhance the next change into a digital format should be avoided.

Since multiple system changes are being considered it is essential that any changes be implemented in such a flexible way that the subsequent changes are anticipated and planned for in the first step. The change into an AMPS expansion should be a first step on the direction of a digital implementation. In order to accomplish this, it is necessary to provide for a radio frequency (RF) unit which could be used in either the expanded analog or the next generation digital system. In addition it would desirable to convert to a digital implementation format at the first point in the system where convenient sampling rates and analog to digital conversion are possible. From this point generic special purpose digital signal processing hardware should be used to allow future changes in this portion of the system to be implemented by software changes to the largest extent possible. Any design not exhibiting such flexibility should have a strong cost advantage, since its life will
be short.

A factor which has not been mentioned anywhere as yet is the possibility of utilizing the subcarrier data capability of the broadcast FM industry. For example, the paging function of the cellular network could equally well be accomplished via a paging channel on an FM broadcast station subcarrier. In addition, it is possible that a network timing signal could be distributed to both mobiles and base stations via a subcarrier transmission.

8.5 Evolution to a New System

In the evolution of the present analog system to a new digital system and the future evolution of the digital system to function efficiently within the ISDN network the following major issues will have to be treated:

- interoperability between any new system and the existing system.
- graceful phase-out and changeover from the existing system into the new system.
- standardization issues within the industry both in the same area and across area boundaries.

Interoperability ideally should mean that any existing mobile unit should continue to operate without prejudice as to quality or availability of service.

Phase-in of new systems can be accomplished by assigning small groups of channels in each cell to the new technique. The available channels are thus partitioned into two groups, new system channels and old system channels. This partitioning technique is frequently suggested but it does not meet the criterion stated above. Precise partitioning details would necessarily depend on the digital system characteristics.

In a nonpartitioned approach it is essential that each channel be operable in either mode. Both the base stations and the new mobile units would have to be capable of dual mode operation if this is to accomplished.

An obvious possibility is that an industry standard be set for both the new system and the phase-in approach.
A unit capable of dual mode operation can be designed in two ways: following a common RF unit, (1) a pair of separate systems would be required to perform the necessary functions of the analog FM set and the new digital format, or (2) one unit capable of performing either set of functions may be designed. The second alternative is possible utilizing the latest DSP chips and implementing the analog FM receiver with digital hardware. Essentially the same hardware could then be set up to perform either set of functions on command of the network. The same unit could also perform other functions related to data handling or other future requirements; it could even be arranged to accept new assignments via programs downloaded from the network.

If a system is designed with the level of flexibility described above the standards issue is also reduced to a handleable problem. Only two standards need be accommodated: a minimum radio-frequency interface standard and a minimum digital interface standard.

The radio-frequency interface would be defined by specifying potential channelization of the allocated frequency resource. The center frequency and bandwidth assignments chosen should be flexible enough to allow any system currently being proposed or reasonably anticipated to operate without prejudice. Some systems might use all of the bands and center frequencies, while other systems might use only a fraction of the set. It would be probably be desirable to specify allowable power levels and amplification classes as well. However the goal should be to provide a minimum RF interface standard such that minimum restrictions are placed on the industry, operators and manufacturers alike.

The digital interface standard should be specified so that a variety of components and digital implementations can interoperate. Again a minimum interface that the entire industry could design against for the forseeable future should be the goal. In the digital speech context, for example, different systems could be allowed to interoperate if appropriate arrangements are made to reconfigure the system for whatever area it happens to be operating into; the base station or the mobile could adapt as appropriate, under network control.

It is recommended that the appropriate industry bodies, including both manufacturers and operators and other interested parties, be organized to set such minimum interface standards at the RF and digital levels as soon as possible.