FRAME SYNCHRONIZATION FOR PILOT-SYMBOL-ASSISTED MODULATION

A Thesis Submitted to the College of Graduate Studies and Research in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Department of Electrical Engineering University of Saskatchewan Saskatoon

by

Qingyi Zhang

Spring 2002

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Electrical Engineering

University of Saskatchewan

Saskatoon, Saskatchewan, S7N 5A9

Abstract

Frame synchronization is a key function in wireless communications systems that employ pilot-symbol-assisted modulation (PSAM). Compared to traditional frame synchronization problems, frame synchronization for PSAM is more difficult due to channel fading and receiver carrier frequency error. In this study an innovative scheme of frame synchronization is proposed for pilot-symbol-assisted modulation. The approach features decomposition of the problem into reduced problems by employing non-coherent detection technique. The applications of minimum mean squared error (MMSE) estimation to channel fading and maximum likelihood estimation to pilot sequence detection provide a vehicle for design optimization. Simulation shows the scheme has improved immunity to receiver carrier frequency error than the only existing scheme in the public literature.

Acknowledgements

I thank Professor David Dodds and Professor Kunio Takaya for supervising my graduate study. Their ideas are critical to the success of my thesis, and their patience made it possible for me to finish it. I would like to acknowledge two professors who had profound influences on the development of my knowledge, which proved very helpful to my research work. Professor Eric Salt from University of Saskatchwen taught me probabilistic thinking, and Professor Fusheng Yang from Tsinghua University showed me the wonderful world of statistical signal processing.

I thank my wife, Ningli Zhang. It would not have been possible to complete this work without her support.

Financial assistance provided from Telecommunication Research Laboratories and Professor Dodds' NSERC operating grant is thankfully acknowledged.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AWGN	Additive White Gaussian Noise
DC	Direct Current
i.i.d.	Independently, Identically Distributed
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
p.d.f.	Probability Density Function
PSAM	Pilot Symbol Assisted Modulation
PTAM	Pilot Tone Assisted Modulation
QAM	Quadrature Amplitude Modulation
WSS	Wide Sense Stationary

1 Introduction

1.1 Wireless and Mobile Communications

Wireless communications refer to communications that do not rely on any type of wire connection between the transmitters and receivers. When the transmitters and receivers are in motion during their operation, the type of wireless communications is called mobile communications. As mobile communication is accomplished through radio wave propagation, we call it mobile radio communication. In recent years, mobile radio communications have seen the fastest growth.

Mobile radio communications can be further classified into narrow-band and wide-band mobile communications. This classification is different from the definitions for narrow-band and broadband signals, which are based on the bandwidth of the transmitted signal relative to the carrier frequency of the communications system. A narrow-band mobile communications system is one whose signal bandwidth is much narrower than the fading channel coherent bandwidth. The fading channel coherent bandwidth is the range of frequencies over which two frequency components have a strong potential for amplitude correlation. The synchronization scheme discussed in this study applies to narrow-band mobile communications system. This type of mobile communications systems finds its application in the areas of low data rate communication systems such as (TDMA cellular phone systems, narrow-band wireless Internet access and email pager systems, etc.)

1.2 Slow and Flat Fading

One of the most devastating phenomena associated with mobile communications is channel fading. Channel fading is the direct result of multipath propagation. Multipath propagation refers to the fact that radio wave can propagate along different paths that have different delays and arrival angles to the receiver antenna. The radio waves from different paths have different phase delay. When the differential phase between two paths is close to π, the two paths combine destructively to each other at the receiving antenna, and result in severe attenuation in received signal strength. This is called channel fading. When the transmitter, receiver and all objects in the multipaths are fixed, the channel fading is time invariant.

In mobile communications, the motion of transmitters, receivers and the moving objects in the multipaths introduces Doppler effect. Doppler effect is a phenomenon where the frequencies of radio waves observed by a receiver shift according to the velocity of the mobile unit. The amount of frequency shift is called Doppler shift. Because the radio waves from different paths have different arrival angles, they experience different Doppler shifts. These radio waves with different Doppler shifts can combine destructively to cancel at one time and constructively at another time, and thus produces time-variant channel fading.

Time-variant channel fading can be classified into fast and slow fading[1]. This classification is based on maximum fading rate. Maximum fading rate is defined as the ratio of maximum Doppler shift to signal bandwidth, or equivalently system symbol

[1] Slow fading and fast fading have other definitions in the literature. The definitions used in this thesis follow that of [1].

rate. It is calculated as

$$f_D = \frac{f_m}{B} = \frac{\frac{vCos\alpha}{\lambda}}{B} = \frac{f_c vCos\alpha}{Bc}, \quad (1)$$

where f_m is the maximum Doppler shift, B is the bandwidth of the baseband signal, v is the velocity of the mobile unit, λ is the wavelength of the carrier, f_c is the carrier frequency, α is the arrival angle of the path that has maximum Doppler shift, c is the constant of light speed. Slow fading refers to fading rate $f_D << 1$, which means the frequency shift of fading is much smaller than the bandwidth of the signal. Typical fading rate for slow fading is less than 5% [1].

Channel fading can also be classified as frequency selective fading and flat fading. Frequency selective fading has different magnitude attenuation and phase distortion for different frequencies within the signal bandwidth, while flat fading has approximately uniform magnitude and linear phase over the whole frequency range of a channel.

The two types of channel fading can also be specified in time domain using the system symbol rate and maximum differential delay. When maximum differential delay among major multipaths is much less than the symbol period, which is the inverse of the system symbol rate, the channel exhibits flat fading; otherwise the channel has frequency selective fading.

The type of fading channels to which this study applies is slow, flat fading. This is because this special type of fading can be conveniently modeled by the multiplicative distortion as shown in the following equation. The model is given as

$$r(t) = c(t)s(t) + n(t) \quad (2)$$

where $r(t)$ is the bandpass received signal, $s(t)$ is the transmitted signal, $n(t)$ is additive

noise, $c(t)$ is the multiplicative distortion that models the effect of slow, flat fading on bandpass received signal. The function $c(t)$ will be called the fading signal. The baseband version of (2) is derived in Section 2.1.

1.3 Channel Fading Estimation

Because the effect of flat fading is represented by $c(t)$, the channel fading signal, an effective way of combating this type of channel fading is to estimate $c(t)$ and use this information to reverse the effect of fading. The reversal of channel fading effect is accomplished by scaling and rotating the decision grid of the receiver in proportion to the estimated fading signal. A decision grid is made up of decision thresholds for the I and Q channels in a QAM receiver.

The estimation of $c(t)$, called channel fading estimation, can be achieved with or without the aid of a reference signal (or, a pilot signal). In pilot signal assisted channel fading estimation, a signal of fixed pattern is transmitted by the transmitter. The receiver stores an original copy of the pilot signal. It compares the received copy of the pilot signal with the original copy to calculate the distortion caused by channel fading. In the absence of a pilot signal, non-pilot signal assisted channel fading estimation must resort to complex algorithms to obtain estimation of the same accuracy as pilot signal assisted estimation. When the signal-to-noise ratio (SNR) is low, the outputs of these complex algorithms take a long time to converge from an initial value to a good estimate. Thus advantages of using pilot signal in channel fading estimation are low complexity and speed. The disadvantage of pilot signal assisted channel fading estimation is that these schemes sacrifice the channel bandwidth or power efficiencies, since the transmission of a pilot signal either consumes extra power of the transmitter or occupies extra time slots.

There are two types of pilot signals used for the purpose of channel fading estimation: pilot tone and pilot symbol [2]. Because transmission of a pilot signal is achieved by modulation, these two channel fading estimation schemes are given the name Pilot Tone Assisted Modulation (PTAM) and Pilot Symbol Assisted Modulation (PSAM). In PTAM, one or more pilot tones are inserted at nulls of the spectrum of the signal that carries the information[2]. After propagation over the multipath channel, each path has some Doppler shift and this generates frequency-shifted copies of the original pilot tone. These received copies of the original pilot tone are distinguished from each other by different frequency shifts and phase distortion, as each path has distinct angle of arrival and path delay. The collection of all frequency-shifted copies of the pilot tone forms a spectrum that reflects the effect of channel fading. The channel fading signal can be conveniently extracted from the received signal by using a narrow bandpass filter with the passband located around each pilot tone.

In PSAM, pilot symbols are periodically inserted into the data symbol sequence that carries the information. After propagation over the fading channel, each pilot symbol is distorted and the channel fading signal is then derived from the distortion on the pilot symbols. The effect of fading on data symbols is estimated using the channel fading signal determined from neighboring pilot symbols.

In this thesis, the practical problem of synchronization is addressed. Synchronization is essential for proper operation of PSAM. The thesis also includes a review of PTAM, since some research in this area is applicable to the design in this

[2] The nulls in the spectrum of the transmitted signal can be generated by some special modulation schemes such as the schemes given in [3].

thesis.

1.4 Frame Synchronization for PSAM

One practical problem in designing a PSAM receiver is how the receiver can discriminate between pilot and data symbols. This problem is first addressed by Gansman [4] in the public literature. Because the task of identifying pilot symbols can be accomplished by the techniques developed for frame synchronization in digital communications systems, the problem has been given the name "frame synchronization for PSAM".

The task of traditional frame synchronization is to recognize the frame structure of a sequence of received symbols. It is a relatively old problem discussed in many published literatures. However, most of the research addresses Gaussian channels. The research results cannot be applied to the design of PSAM receiver in a straightforward fashion. Gansman first proposed a frame synchronization scheme that explicitly addresses the effect of channel fading and is applicable specifically to PSAM. The research in this thesis was inspired by his pioneering work, but has taken a somewhat different approach, which emphasizes reduction of complexity and immunity to carrier frequency error.

1.5 Objective and Statement of Problem

The objective of this research is to develop a frame synchronization scheme for PSAM. The motivation originates from feasibility study for a project[3] at TRLabs,

[3]The project was proposed by Professor Kunio Takaya in the summer of 1997.

Saskatoon. The project was to implement a wireless communications system using Texas Instrument TMS320C30 DSP processors. The major difficulty was to find a low complexity scheme that tackles channel fading. PSAM was considered as the most promising candidate. However, after an exhaustive literature search, we could find only one frame synchronization scheme for PSAM, which was proposed by Gansman [4]. Because of the limited power of the TMS320C30 DSP processor, it was felt that a frame synchronization scheme of lower complexity was required. And also because 16-QAM was chosen for the wireless communications system, it would be beneficial to have a frame synchronization scheme tailored for that modulation format. At last, we felt that there was room for improvement with respect to the immunity to carrier frequency error.

In summary, the problem is to develop a frame synchronization scheme that should feature low complexity, immunity to carrier frequency error and suitability to 16-QAM.

1.6 Thesis Outline

In the rest of this thesis, the mathematical model for flat fading channel is presented in Chapter 2, then the detailed discussion on a frame synchronization scheme for PSAM is presented in Chapter 3. This scheme was evaluated by computer simulation and the simulation results are described in Chapter 4. Finally, the design process and result are summarized, and directions for future work are identified in Chapter 5.

2 Channel Model

The purpose of this chapter is to establish a mathematical framework for the transmission channel, i.e. a channel model. This model is the foundation for the development of a synchronization scheme. It is also important when we simulate the scheme to evaluate its performance.

The task of establishing mathematical models for physical systems is to obtain a minimum set of features that can be used to adequately describe the problem in hand. We will describe: (i) how the notion of flat fading is used to simplify the mathematical model in Section 2.1; (ii) how the model is further simplified by combining phase offset and phase distortion in one variable; and finally, (iii) how the numerical specification for all the parameters in our model are determined in Section 2.3.

2.1 Flat Fading Channel

As pointed out in Section 1.3, PSAM is a technique for narrow band flat fading channels, thus we focus on the flat fading channel model. The notion of a flat fading channel makes the mathematical model much simpler than for non-flat fading channels. To appreciate this simplification, the most powerful mathematical model, which is suitable for general fading channels, is used as the starting point. Then the properties of slow, flat fading and other approximation are used to simplify the general complex channel model into a very simple channel model. The purpose of the detailed derivation is to illustrate all the assumptions we have made. When designing a PSAM receiver

using the frame synchronization scheme developed based on the simplified channel model, all the assumptions must be satisfied to a certain extent. This general channel model is known as the Impulse Response Model, and is given as

$$r(d,k) = \sum_{i=-\infty}^{k} c(d,i,k-i)s(i) + n(k). \quad (3)$$

This equation describes the discrete time baseband signal as observed at the output of the symbol sampler of a receiver. The variable $r(d, k)$ represents the symbol received at position d and time k, assuming the symbol interval is normalized to 1 second; $s(i)$ represents the symbol transmitted at time i, $c(d, i, k-i)$ is the discrete time impulse response of the time variant fading channel, and $n(k)$ the additive noise. Because this model describes a Quadrature Amplitude Modulation (QAM) signal, all symbols in (3) have complex values. The real parts of $r(d, k)$, $s(k)$ and $n(k)$ represent the I-channel, imaginary parts represent the Q-channel.

A physical explanation for $c(d, i, j)$ is when a unit impulse is transmitted by the base station j seconds in the past, $c(d, i, j)$ is the signal received by the mobile unit at time i second, when it is at distance d from the base station. The variable j is a relative time measurement, i.e. it measures the delay between transmition of the unit impulse at the base station and observation of the unit impulse response, $c(d, i, j)$ at the mobile unit. The convolution means that the received signal (devoid of additive noise) is made up of contributions from all symbols that have been transmitted until the moment of the i^{th} second. The contribution from symbol $s(i)$ is weighted by the unit impulse response coefficient $c(d, i, k-i)$.

This form can be reduced to a form of time variant linear system by imposing the constraint that the mobile unit is moving at constant speed v m/s. Then d can be

expressed as $d = vi$, assuming i is 0 when the mobile is at the base station. Substituting this expression in (3) gives

$$r(vk,k) = \sum_{i=-\infty}^{k} c(vi,i,k-i)s(i) + n(k). \tag{4}$$

From (4) we see that $r(vk, k)$ is a function dependent only on k, hence is expressed as $r(k)$; $c(vk, i, k-i)$ is a function dependent of k and i, hence is expressed by $c'(i, k-i)$. Note that $c'(i, k-i)$ is not the derivative of $c(i, k-i)$; it is just a convenient symbol to differentiate different version of $c(d, i, k-i)$. The physical meaning of $c'(i, k-i)$ is redefined as the observation made at time i when a unit impulse is sent at time k. Now (4) can be rewritten as

$$r(k) = \sum_{i=-\infty}^{k} c'(i,k-i)s(i) + n(k) \tag{5}$$

This is the most concise form of the mathematical model for general fading channels. It is still a complicated model due to the summation of product between $c'(i, k-i)$ and $s(i)$. However, thanks to the notion of flat fading, this model can be further simplified. The simplification comes from an approximation to the unit impulse response for the fading channel. As introduced in Section 1.2, the differential delays among multiple paths are assumed to be much smaller than a symbol interval, and the Doppler shift is much smaller than the signal bandwidth. Based on these conditions, $c'(i, k-i)$ can be considered as a product of two factors:

$$c'(i,k-i) = c''(i)\delta_K(k-i), \tag{6}$$

where $\delta_K(\cdot)$ is the discrete Kronecker delta function. Besides the assumption of flat fading, $\delta_K(k-i)$ also implies that we have ignored delay in signal propagation. The variable $c''(k)$ is often referred to as channel state information (CSI) in the literature,

because the fading channel is often considered a linear finite state system. Substituting (6) into (5) gives

$$r(k) = c''(k)s(k) + n(k). \tag{7}$$

In (7), the effect of channel fading has been reduced to a multiplicative distortion to the transmitted signal. All frequency components of the signal undergo uniform attenuation (or gain) and linear phase shift.

2.2 Carrier Frequency and Phase Error

The model in (7) does not address an important practical problem. That is the local carrier frequency and phase offset. In a real communication system, the oscillator at the receiver cannot be tuned to exactly match its counterpart on the transmitter side due to that the transmitter and receiver oscillators can not be made to exactly generate one nominal frequency signal and both are also subject to different temperature and aging, which will affect the frequency of the signals they generate. The receiver carrier frequency and phase error makes it difficult to coherently demodulate the received signal. To take account of the effect of these errors on our frame synchronization scheme, we introduce a quantitative description of them in the channel model.

Carrier frequency and phase errors occur at bandpass stages of the receiver. In these stages, the received signal can be conveniently expressed as

$$r_p(t) = \mathrm{Re}[s_b(t)A(t)e^{j\theta(t)}e^{j\omega t}]. \tag{8}$$

where $r_p(t)$ is the real-valued bandpass received signal, as illustrated in (b) of Figure 2-1; $s_b(t)$ is the continuous time, complex, baseband signal, the real and imaginary components are illustrated in (a) of Figure 2-1; $A(t)$ is the channel gain, $\theta(t)$ is the time varying phase offset introduced by fading and Doppler shift, ω is the carrier frequency at

the transmitter.

After down conversion, the carrier frequency is replaced by a small value frequency $\Delta\omega$, which is the carrier frequency error at the receiver, and a constant phase offset α. Note that $\Delta\omega$ and α are slowly time-varying signals, they are approximated by constants because the variations are negligible compared to that of the channel fading. The baseband signal is

$$r_b(t) = s_b(t)A(t)e^{j\theta(t)}e^{j\Delta\omega t+\alpha} \tag{9}$$

The factor $A(t)e^{j\theta(t)}$ is the continuous time expression of the channel state information and is expressed in phasor notation. As we will discuss in Section 2.3, the statistical model for channel fading is usually considered as Rayleigh fading, which means that amplitude $A(t)$ has Rayleigh probability density distribution and phase $\theta(t)$ has uniform probability density distribution over $[0, 2\pi)$. The phase offset α is a random variable with uniform probability density over $[0, 2\pi)$. The sum of two random variables $\theta(t) + \alpha$ has the same uniform probability density distribution because the summation of phases is modulo of 2π. This suggests that if we combine $\theta(t)$ and α into only one random process, $\theta_c(t)$, the new model is equivalent to the old one, since the statistical properties of the ensembles represented by the two models are the same. Thus, the new model can be reduced to

$$r_b(t) = s_b(t)A(t)e^{j\theta_c(t)}e^{j\Delta\omega t}. \tag{10}$$

There is an important part missing from (10). That is the noise. We left it out in the foregoing discussion because inclusion of it would have made the equation unwieldy. Unlike the signal, noise does not necessarily propagate along multiple paths of a fading channel, therefore channel fading does not apply. Part of the noise comes into the system

at the antenna as unwanted electromagnetic radiation; and part of the noise comes at the baseband stages as thermal noise in electronic circuits. Noise in communication systems is considered to have a flat power spectrum (white noise) and a shift of central frequency does not change its statistical properties. For bandpass noise that enters the system before the down-conversion, carrier frequency error has no effect. It is generally assumed that this bandpass complex Gaussian noise has circular symmetry, i.e. the real and imaginary parts of random variable are independent with identical Gaussian distribution, so the carrier phase offset will not change its statistical properties. For the above reasons, noise can be expressed as an additive term in the expression for baseband received signal. The baseband channel noise is denoted by $n(t)$, and we use $c(t)$ instead of $A(t)e^{j\theta(t)}$ for channel fading. The received baseband signal is then

$$r_b(t) = c(t)e^{j\Delta\omega t}s_b(t) + n(t). \tag{11}$$

The continuous time baseband signal, $r_b(t)$, is filtered by a square root raised-cosine filter and sampled at the symbol rate. To derive the expression for the discrete time baseband signal, there are some assumptions and approximations. First, it is assumed that the receiver has perfect symbol timing, as symbol synchronization is beyond the scope of this study. Second, as discussed in Section 1.2, the channel fading signal $c(t)$ changes slowly and therefore is approximated as constant during one symbol period. Thus we can approximate the continuous time fading signal as a discrete time function, which has constant value during a symbol interval and changes its value at symbol boundaries. Third, the frequency offset represented by $\Delta\omega$ is small and within limits determined by the square root raised-cosine filter and symbol sampler [1]. When the offset frequency is a very small fraction of the symbol rate, it manifests itself as a

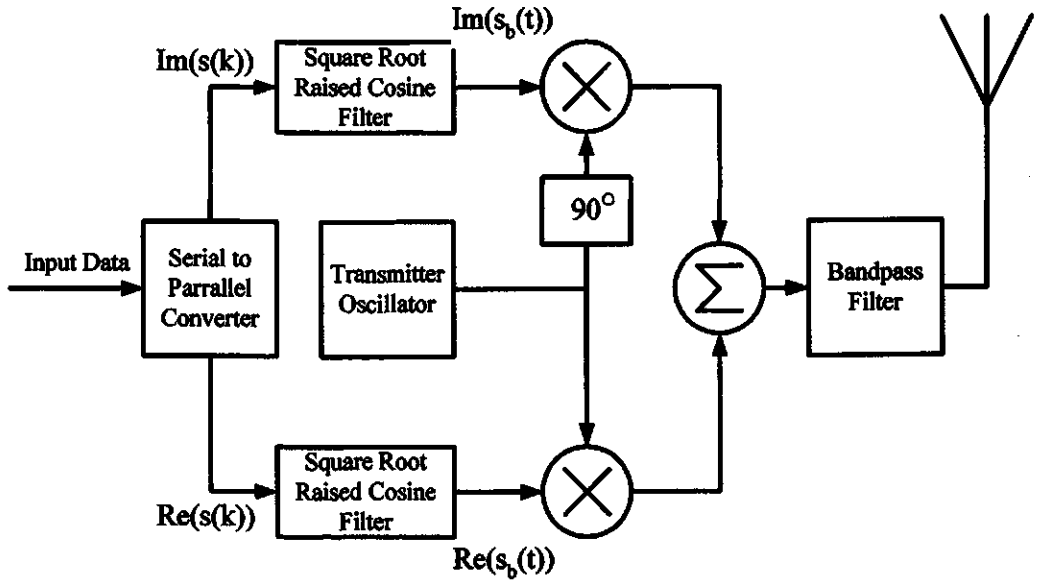
small amount of phase increase (or decrease) across successive symbols. During each symbol interval, the phase change is negligible. Thus we can approximate the frequency offset as discrete phase change that occurs at the boundaries between symbols. Based on these approximations, the discrete time channel model is

$$r(k) = c(k)e^{j\Delta\omega Tk}s(k) + n(k). \qquad t = Tk \tag{12}$$

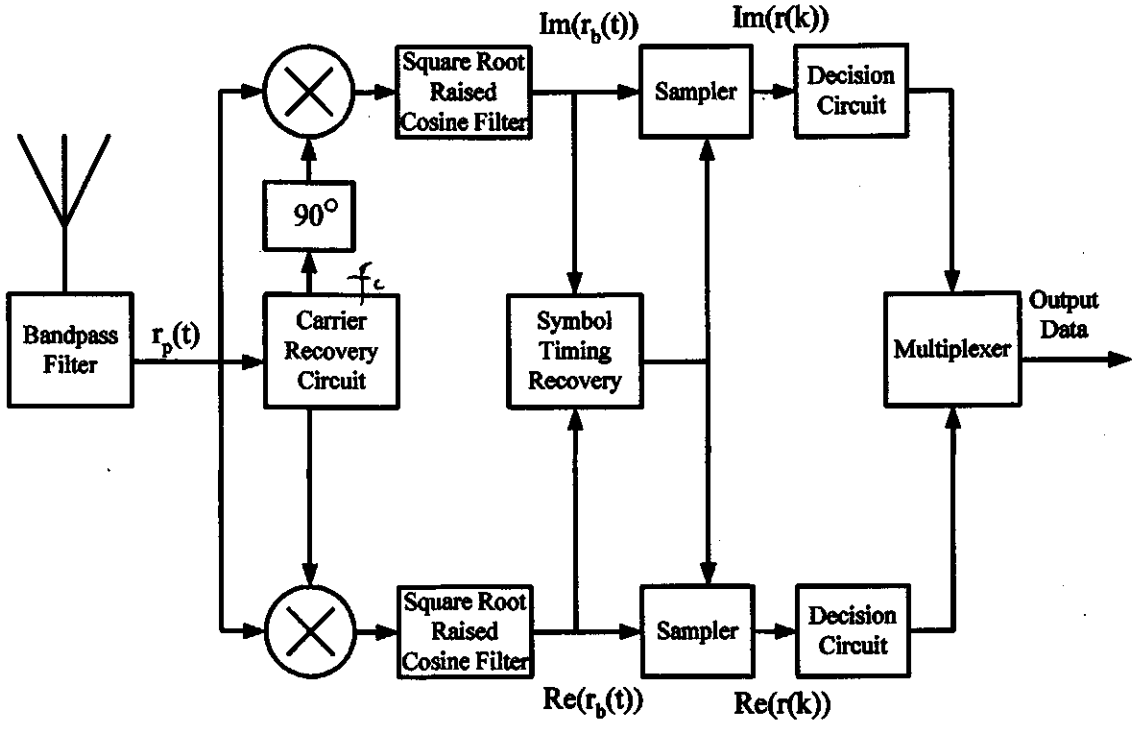
where $c(k)$ is the discrete time approximation to $c(t)$, T is the symbol interval. This is the mathematical model we will use in the development of a synchronization scheme in Chapter 3.

It shall be noted that even though the p.d.f. of $c(k)$ is assumed to be circular symmetric as will be discussed in the next section, and the p.d.f. of $c(k)e^{j\Delta\omega Tk}$ is also circular symmetric, however, we can not combine $e^{j\Delta\omega Tk}$ with $c(k)$ and simplify the product to $c(k)$ as we did to the carrier phase offset. This is because, when they are considered as random processes, $c(k)e^{j\Delta\omega Tk}$ is a frequency translated version of $c(k)$; it shifts the power spectrum of $c(k)$. Since the power spectrum of $c(k)$ is band limited and non-white, the frequency shifted spectrum represents another random process that is distinct from the random process represented by original spectrum. In [4], Gansman does not include the factor of carrier frequency shift in the channel model, but briefly discussed the effect of carrier frequency error in the simulation results. However, it is important for the model to accurately specify every imperfection of a physical channel, as this is the first step toward developing a complete solution to the problem.

At last, the signals used in the equations of the section are illustrated in Figure 2-1.



(a) QAM Transmitter



(b) QAM Receiver

Figure 2-1 Channel Model Signals in the QAM Transmitter and Receiver

2.3 Specification of Parameters

In the channel model of (12), the received signal $r(k)$ is expressed in terms of random processes $c(k)$, $s(k)$, $n(k)$ and random variable $\Delta\omega$. In this section we will give statistical models for each of the random processes and numerical specification for the parameters in the models.

The random process $s(k)$ is determined by the choice of modulation and coding schemes. The modulation format chosen for this study is Quadrature Amplitude Modulation with a 16-point square constellation (16-QAM). 16-QAM can double the amount of information transmitted by $\pi/4$-QPSK using the same bandwidth. The latter is widely used in cellular phone systems that employ TDMA technique. 16-QAM is a promising candidate to replace $\pi/4$-QPSK in these systems. 16-QAM is the simplest in the QAM family (excluding QPSK), and is most immune to noise; it is therefore best suited to severe wireless channels. To numerically specify the random process $s(k)$, we introduce $\sqrt{E}$ as the basic unit for the amplitude of 16-QAM symbols as shown in Figure 2-2. We assume the information is a stream of independent random binary bits. The stream is divided into 4-bit words, and these words are mapped into a 16-QAM constellation using Gray code shown in Figure 2-2. The Gray code mapping for QAM is commonly used in digital communication systems. One benefit of Gray code mapping is that when a received symbol is misinterpreted as one of its neighboring symbol in the QAM constellation, it will only result in one bit error. It should be noted that the validity of the result in this thesis does not depend on the choice of mapping scheme.

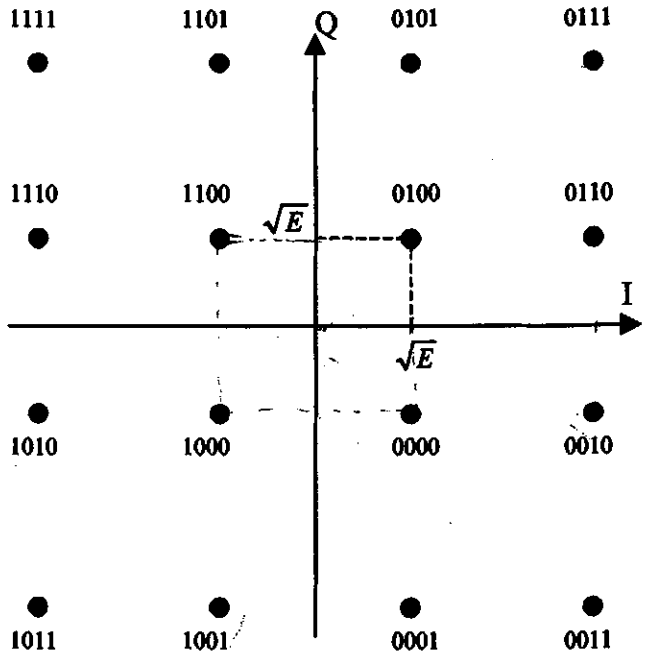


Figure 2-2 16-QAM Constellation and Gray Code Mapping

As a general study, we choose not to be committed to any particular coding scheme, rather we assume no channel coding at the transmitter. Given the random characteristics of the bit stream, each symbol in the symbol stream can fall on any of the 16 points with equal probability. The probability distribution functions for each symbol are independent, identically distributed (i.i.d.). Symbol stream $s(k)$ is a discrete time complex random process. The mean of $s(k)$ is

$$
\begin{aligned}
mean = & \sqrt{E}\left(\frac{4}{16}\right) - \sqrt{E}\left(\frac{4}{16}\right) + 2\sqrt{E}\left(\frac{4}{16}\right) - 2\sqrt{E}\left(\frac{4}{16}\right) + \\
& \left\{\sqrt{E}\left(\frac{4}{16}\right) - \sqrt{E}\left(\frac{4}{16}\right) + 2\sqrt{E}\left(\frac{4}{16}\right) - 2\sqrt{E}\left(\frac{4}{16}\right)\right\} i \\
& = 0 + 0i \\
& = 0.
\end{aligned} \tag{13}
$$

The variance of $s(k)$ is

$$
\begin{aligned}
var = & \left[\left(\sqrt{E}\right)^2 + \left(\sqrt{E}\right)^2\right]\left(\frac{4}{16}\right) + \left[\left(3\sqrt{E}\right)^2 + \left(\sqrt{E}\right)^2\right]\left(\frac{8}{16}\right) + \left[\left(3\sqrt{E}\right)^2 + \left(3\sqrt{E}\right)^2\right]\left(\frac{4}{16}\right) \\
& = 2E\left(\frac{1}{4}\right) + 10E\left(\frac{1}{2}\right) + 18E\left(\frac{1}{4}\right) \\
& = 10E.
\end{aligned} \tag{14}
$$

In frequency domain, the power spectrum of $s(k)$ has a constant value of $10E$ over $(-1/2T, 1/2T)$. The symbol T represents the symbol time.

The random process $c(k)$ can have one of several probability density functions. Some of the most common models in the literature are Rayleigh distribution and Rician distribution. The choice of probability distribution function is largely determined by the intended application of the communications system. For example, urban environments are often modeled by Rayleigh fading, since there is usually not a dominant direct path and the radio waves propagate along multiple paths that have similar attenuation. In rural areas, there is usually a direct path that is dominant over other indirect paths; therefore the fading channels are often modeled by Rician fading. Because Rayleigh fading can be considered as the worst case of Rician fading, it is selected for this study. In Rayleigh fading, the real and imaginary parts of $c(k)$ have independent Gaussian distributions with identical means and variances. The name of Rayleigh fading comes from the probability distribution function for $|c(k)|$. The power spectrum of the baseband

fading signal, $c(k)$, is derived in Appendix A. The derivation begins with the power spectrum of a bandpass fading signal as given by Jakes [5], and extends Jakes' result to the case of baseband discrete time signals. The power spectrum of a baseband fading signal is

$$S_c(f) = \frac{1}{\pi f_m \sqrt{1-\left(\frac{f}{f_m}\right)^2}}. \tag{15}$$

The notation, f_m, is the maximum Doppler frequency shift. This parameter can be accurately calculated from the speed of mobile unit and the carrier frequency. For convenience, the power spectrum for $c(k)$ is normalized so that the variance of the fading signal is unity. The normalization of the fading signal variance unifies the definition of signal-to-noise ratio (SNR) for AWGN and fading channels. In an AWGN channel specified by

$$r(k) = s(k) + n(k), \tag{16}$$

SNR is defined as $E\{|s(k)|^2\}/E\{|n(k)|^2\}$. For a fading channel specified by

$$r(k) = c(k)s(k) + n(k), \tag{17}$$

we can think of it as the signal $s(k)$ is first modulated by $c(k)$ and then transmitted over a virtual AWGN channel. This is a very useful interpretation of flat-fading channels, which is going to be exploited during the development of the frame synchronization scheme of this thesis. The SNR for this virtual AWGN channel is, by definition,

$$SNR = \frac{E\{|c(k)s(k)|^2\}}{E\{|n(k)|^2\}} = \frac{E\{|c(k)|^2\}E\{|s(k)|^2\}}{E\{|n(k)|^2\}}. \tag{18}$$

The right-hand side of (18) results from the fact that $c(k)$ and $s(k)$ are independent of

each other. Thanks to the normalization of the variance of $c(k)$, the SNR for the virtual AWGN channel is the same as the SNR for the fading channel proper, i.e. $E\{|s(k)|^2\}/E\{|n(k)|^2\}$. This is shown in (19).

$$SNR = \frac{E\{|c(k)|^2\}E\{|s(k)|^2\}}{E\{|n(k)|^2\}} = \frac{E\{|s(k)|^2\}}{E\{|n(k)|^2\}} \quad (19)$$

Noise $n(k)$ is defined to be white Gaussian noise. Mean of $n(k)$ is zero, variance σ_n^2 is calculated from the variance σ_s^2 of the transmitted signal and SNR, denoted by η, for the channel by $\sigma_n^2 = \sigma_s^2/\eta$. The power spectrum for $n(k)$ is constant and equal to σ_n^2 over the range of (-1/2T, 1/2T).

The symbol $\Delta\omega$ is determined by the carrier synchronization scheme chosen for a real design. Since we don't specify such a scheme in our study, the requirement for the range of $\Delta\omega$ is specified from the channel model's point of view. As a requirement of the discrete channel model, $\Delta\omega$ must be much smaller than the bandwidth of the square root raised cosine filter, otherwise the approximation used in transforming continuous time model to discrete time model is no longer valid.

3 Frame Synchronization Scheme

This chapter describes the development of a frame synchronization scheme for PSAM. The objectives of the design are first introduced, and followed by detailed description of how each objective is accomplished.

3.1 Design Goals

The purpose of frame synchronization for PSAM is to identify pilot symbols among the data symbols. If the received signal were a noiseless, distortion-free signal, as it is at the transmitter, then the only difficulty faced by the frame synchronizer would be how to tell the authentic pilot symbol sequence (pilot sequence for short) from the data symbol sequences (data sequence for short). This is particularly difficult when the data sequence happens to be similar or even identical to the pilot sequence. This is a problem that has well-established solutions in the literature. However, for mobile communication systems, the received signals are always distorted, noisy and frequency shifted. The frame synchronizer must be able to cope with all the difficulties caused by these imperfections. These imperfections are quantitatively represented as $c(k)$, $n(k)$ and $\Delta\omega$ in the channel model (12).

In customary systems, frame synchronization occurs after demodulation, which avoids the above difficulties. However, frame synchronization for PSAM must be achieved before the demodulation stage. In this case, the synchronization process must provide mechanisms for overcoming these problems. In the following, we discuss what

are the requirements for these functions that must be provided by the frame synchronizer.

First, consider the carrier frequency error $\Delta\omega$. It must be accepted that there is always some residual frequency error, which is a result of the carrier synchronizer. The focus of design should be on improving the frame synchronizer's immunity to the frequency error rather than building a superior carrier synchronization scheme within the frame synchronizer. The performance of the frame synchronizer should not degrade significantly in presence of small frequency error.

The purpose of PSAM is to estimate channel fading $c(k)$, thus the PSAM synchronizer must be able to operate during the fading. To improve the synchronizer's performance, designers must concentrate on counteracting the effect of fading through explicit estimation of channel fading or through the use of an optimal detection scheme. In either case, the inclusion of such mechanism results in considerable increase in computational complexity. Therefore, the emphasis of designing this anti-fading mechanism is on minimization of complexity.

In spite of the exhaustive research conducted in the past several decades, noise remains a difficult problem, because there is always not enough computation power to implement the most optimal algorithm. In practical systems, designers with limited computation resources rely on approximations, which result in sub-optimum performance. There are two types of approximation rules in the literature: one is more accurate for high SNR channels; the other is more accurate for low SNR channels. Because a 16-QAM receiver requires higher SNR than for a $\pi/4$-QPSK receiver, the design of the frame synchronizer should target channels that have high or moderate

SNRs. Thus the high SNR approximation rule can be exploited to reduce complexity.

In summary, the design goals are to achieve the following features for the frame synchronizer with the least complexity:

- The frame synchronizer is immune to a range of the frequency error.
- It can function properly in slow, flat fading.
- It works properly under high and moderate SNR conditions.[4]

3.2 Non-coherent Scheme

This section discusses the techniques that give the synchronizer immunity to carrier frequency error. The goal is to get rid of $\Delta\omega$ from the channel model in (12). As discussed in Section 3.1, it is not a practical option to further improve local carrier frequency using conventional carrier synchronization techniques, such as phase locked loop, within the frame synchronizer. Other techniques for improving frequency offset immunity, such as doubly differential modulation, not only demand major modifications from other parts of the system but also sacrifice channel bandwidth efficiency [7].

It is possible to design a statistically optimal algorithm, which can achieve the optimum performance for a given range of frequency shift. This approach leads to the most optimal solution in theory, but it is not a practical one. The more parameters we include in a parameter space (see Section 5.1 for definition of parameter space), the more complex an algorithm will become. Since complexity grows in proportion to the power of the number of parameters, complexity can grow very fast, which would make

[4] In order for a QAM receiver to function properly, the SNR must not be too low. For example, to achieve probability of symbol error $< 10^{-5}$ with 16-QAM, the SNR must be higher than 15 dB [6].

it impractical to implement the algorithm in real system design. In total, there are four random factors in the channel model, $\Delta\omega$, $c(k)$, $n(k)$ and the data part of $s(k)$, that obstruct correct detection of pilot sequence. If we deal with all of them at the same time, we are faced with a 4-dimensional parameter space, and the optimization problem becomes too complicated to handle. Therefore the statistical method is not the best choice if reducing complexity is an important concern.

Now, let us consider non-coherent communications systems. In these systems, information is transmitted independent of the absolute value of carrier phase, thus any carrier phase and frequency errors have no effect on the receiver's performance. If we apply this same idea to frame synchronization, it is possible to design a frame synchronizer that is immune to all kinds of phase and frequency distortions, including Doppler shift, carrier frequency offset and even random phase distortion of the fading channel[5]. To implement a non-coherent frame synchronizer, the pilot sequence must be transmitted completely in magnitudes. This is not a problem for 16-QAM. From Figure 3-1, we see that a symbol can take one of three different magnitudes; this allows us to define the pilot sequence using a pattern in magnitudes. In this study, we use symbols on the outer most circle for bit 1, and those on the inner most circle for bit 0. Thus, a pilot sequence, defined in bit sequence, can be easily mapped to a sequence of 16-QAM symbols.

[5] The application of a non-coherent scheme was first suggested by Professor Dodds.

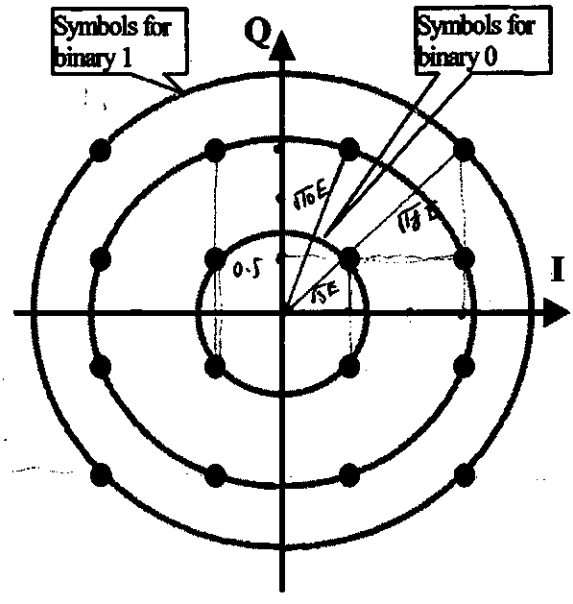


Figure 3-1 Mapping Binary Numbers to 16-QAM Symbols

One possible argument against this non-coherent scheme is that it is a waste of system's capability of bearing information in phase, and thus non-coherent schemes are inherently inferior to coherent schemes. Justification for its use in frame synchronization is twofold. First, the system does not compromise the use of phase information when it functions as a communication system, because each data symbol makes use of both magnitude and phase. Second, at the start of a receiver, a system is in a partially functioning state, i.e. only signal magnitude can be reliably transmitted and signal phase information can hardly be considered as reliable. Therefore, it is reasonable to use a reduced, yet reliable, function of the system to transmit the pilot sequence. The purpose of the pilot sequence is to then boost the receiver into a fully functioning state. ? How

As a final point in this section, it is worth noting that although the frame synchronizer only takes signal magnitude, the received signal magnitude is still a function dependent on carrier frequency error. This is because there is an additive noise

in the fading channel model, which results in a cross term containing $\Delta\omega$ in the expression for magnitude squared of the received signal. From (12), we can derive the magnitude squared of the received signal as

$$|r(k)|^2 = |c(k)|^2 |s(k)|^2 + 2\mathrm{Re}\left\{c(k)e^{j\Delta\omega kT} s(k)n(k)^*\right\} + |n(k)|^2. \quad (20)$$

From (20), it is obvious that for the same $c(k)$, $s(k)$ and $n(k)$, receivers with different carrier frequency errors (i.e. $\Delta\omega$), will observe different values of $|r(k)|^2$. Thus we cannot expect a non-coherent frame synchronization scheme to perform equally well on different values of $\Delta\omega$. Does this mean the scheme is not immune to carrier frequency error? The answer is no. If we combine $e^{j\Delta\omega kT}$ and $n(k)^*$ in the second term on the right side of (20), we can think of the product $e^{j\Delta\omega kT}n(k)^*$ as a frequency-shifted version of $n(k)^*$. Let $n'(k)$ denote $e^{-j\Delta\omega kT}n(k)$. This frequency shift transforms one random process into another random process. Since $n(k)$ is a zero-mean complex Gaussian noise with white spectrum, it can be proven that $n(k)$ and $n'(k)$ have the same correlation function (or power spectrum), i.e. they are statistically identical. Because $|n(k)|^2 = |n'(k)|^2$, substitution of $n'(k)$ in (20) gives

$$|r(k)|^2 = |c(k)|^2 |s(k)|^2 + 2\mathrm{Re}\left\{c(k)s(k)n'(k)^*\right\} + |n'(k)|^2. \quad (21)$$

From (21), we conclude that there always exists an equivalent combination of a receiver of zero carrier frequency error and a channel noise $e^{-j\Delta\omega kT}n(k)$ that can be used to replace the combination of a receiver with frequency shift $\Delta\omega$ and channel noise $n(k)$. Since $n(k)$ and $e^{-j\Delta\omega kT}n(k)$ are identical random process, the mapping between receivers implies that a scheme with a certain average performance on a receiver without frequency error will have the same average performance on a receiver of frequency error $\Delta\omega$ over the same

ensemble of $n(k)$, i.e. the scheme is statistically immune to carrier frequency error. It should be noted, however, that frequency error results in some signal loss in the receiver filters and this impacts the performance of the frame synchronizer.

In this section, we have shown reasons for choosing non-coherent receiver as a mechanism for improving immunity to carrier frequency error. Since the frame synchronizer is non-coherent, we will only concern ourselves with the magnitude of received signal in the following sections of this chapter.

3.3 Fading Estimation

Having eliminated the effect of $\Delta\omega$ by using a non-coherent frame synchronizer, we now attempt to eliminate $c(k)$ from the channel model in (12). As it has been shown in the previous section, for each receiver with carrier frequency error, there always is an equivalent receiver with zero frequency error. This allows us to use the simplified channel model in (7), which is repeated here for convenience:

$$r(k) = c(k)s(k) + n(k). \tag{22}$$

As in the case of carrier frequency error, there are basically two approaches to combating channel fading. One can consider $c(k)$ and $n(k)$ simultaneously, and design an optimal pilot sequence detection algorithm; or can attack them separately. Based on the same design strategy of not combining $\Delta\omega$ with $c(k)$ and $n(k)$, we choose not to combine $c(k)$ with $n(k)$, and therefore avoid complicating the detection algorithm.

To isolate the problem of channel fading from that of channel noise means we need to explicitly estimate channel fading. Channel fading estimation is usually considered a major task of the receiver, since it requires considerable amount of computation power of the receiver. It is by no means an easy task for a frame

synchronizer, which has only a small share of the total computation power of the receiver. We have mentioned in Chapter 1 that a receiver relies on a pilot sequence to accomplish channel fading estimation, and the pilot sequence needs to be detected by a frame synchronizer. However, a frame synchronizer needs to do channel estimation to accomplish its task. There seems to be a cycle in the line of reasoning, which leads us back to the original problem of channel fading estimation. This does not mean we have reached an impasse. The channel fading estimation for demodulation must provide information about magnitude attenuation and phase distortion. It turns out that the channel-fading information that the frame synchronizer needs is only the magnitude attenuation of channel fading. The reason why the frame synchronizer only needs magnitude attenuation is explained in Section 3.4.1. To estimate the magnitude of a fading channel is much simpler than to estimate both magnitude and phase of the fading channel. Therefore the channel fading estimator within the frame synchronizer is simple to implement, and the approach of explicit channel fading estimation leads to reduction of the original problem. In the rest of this section, we will explain the scheme for estimating channel fading from two different points of view: time domain explanation and frequency domain explanation.

3.3.1 Time Domain Interpretation

Because the original idea for the scheme[6] is conceived in time domain, it is natural to introduce the idea in the time domain. The time domain interpretation is based on the following scenario of a hypothetical experiment. Suppose we carry out an experiment in

[6] The idea was first suggested by Professor David E. Dodds [8].

order to measure the magnitude of a channel fading signal $c(k)$. A mobile unit moves along the same route at the same speed repeatedly. Each time a random signal $s(k)$ is transmitted. The receiver records one sample record for each run. The experiment has been repeated infinite times and finally we obtain an ensemble of all records $r(k)$. Now using the channel model in (22), we obtain the expression for the ensemble $\mathbf{r}(k)$ as

$$\mathbf{r}(k) = c(k)\mathbf{s}(k) + \mathbf{n}(k). \tag{23}$$

Note that, in this paragraph, the random processes $\mathbf{r}(k)$, $\mathbf{s}(k)$ and $\mathbf{n}(k)$ are presented in boldface, while the deterministic fading signal $c(k)$ is not. We identify $c(k)$ as a deterministic signal, because the mobile unit repeats the same route at the same speed. Furthermore, we distinguish $\mathbf{r}(k)$ from $\mathbf{s}(k)$ and $\mathbf{n}(k)$. While $\mathbf{s}(k)$ and $\mathbf{n}(k)$ are stationary random processes, $\mathbf{r}(k)$ is a non-stationary process. The mean squared value of $\mathbf{r}(k)$ is a function of the time index k, and can be expressed as

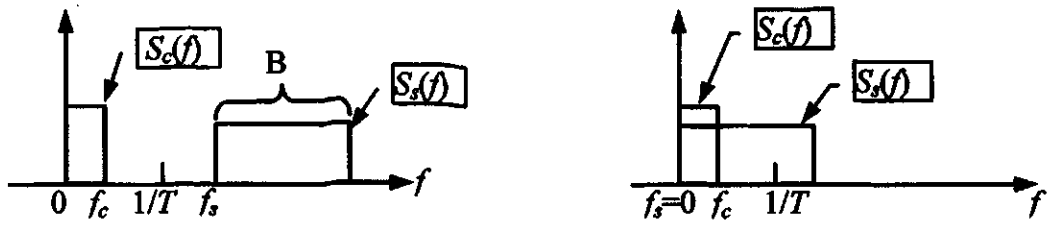
$$\begin{aligned} E\{|\mathbf{r}(k)|^2\} &= E\{|c(k)\mathbf{s}(k)|^2 + 2\mathrm{Re}(c(k)\mathbf{s}(k)\mathbf{n}(k)^*) + |\mathbf{n}(k)|^2\} \\ &= |c(k)|^2 E(|\mathbf{s}(k)|^2) + 2\mathrm{Re}(c(k)E(\mathbf{s}(k))\underbrace{E(\mathbf{n}(k)^*)}_{0}) + E(|\mathbf{n}(k)|^2) \\ &= |c(k)|^2 \sigma_s^2 + \sigma_n^2. \end{aligned} \tag{24}$$

The last equal sign in (24) results from the assumption that $\mathbf{s}(k)$ and $\mathbf{n}(k)$ are zero mean, stationary complex random processes. From(24), we can express the magnitude squared of the channel fading signal as

$$|c(k)|^2 = \frac{1}{\sigma_s^2}\left[E(|\mathbf{r}(k)|^2) - \sigma_n^2\right] \tag{25}$$

To obtain $|c(k)|^2$, we need σ_s^2, σ_n^2 and $E(|\mathbf{r}(k)|^2)$. The variance σ_s^2 is a design parameter for the transmitter. We normally assume a typical value of σ_n^2 according to system design specifications. Alternatively, σ_n^2 can be estimated by the receiver. The difficulty in evaluating $|c(k)|^2$ arises from the fact that the baseband equivalent power

spectrum of the fading signal, $S_c(f)$, overlaps the power spectrum, $S_s(f)$, of the transmitted signal as shown by graph (b) in Figure 3-2.



(a) desired channel (b) 16QAM channel

Figure 3-2 Conditions for Time Average Estimator

The graph (a) of Figure 3-2 shows the requirement for a running time estimator to accurately estimate $E(|r(k)|^2)$. The running-time average estimator for $E(|r(k)|^2)$ is

$$\hat{E}(|r(k)|^2) \approx \frac{1}{T}\sum_{i=k-1-T/2}^{k+T/2}|r(i)|^2. \quad (26)$$

It is desirable that the highest frequency in the fading channel, f_c, be less than the lowest frequency, f_s, in the transmitted signal $s(k)$. The bandwidth of $s(k)$ is denoted as B and the length of a running-time average estimator is denoted as T. If we assume random processes $s(k)$ and $n(k)$ are ergodic, and $c(k)$, $s(k)$ and $n(k)$ satisfy the following conditions[9]:

- $f_s > 1/T > f_c$,
- $1/T << B$,

we can use one experiment sample to calculate the mean squared value of $r(k)$, i.e. to use time average instead of an ensemble average. In the above conditions, f_s is the lowest

frequency in $s(k)$, while f_c is the highest frequency in $c(k)$. The first condition essentially says that the spectrum band of $s(k)$ must be on the higher side of that of $c(k)$, and both bands do not overlap. It specifies boundaries for the window length as being long enough to hold a whole period of the highest frequency component of $c(k)$, but no longer than the period of the lowest frequency component in $s(k)$. The second condition further requires the lowest frequency can be observed by the window must smaller than the bandwidth of $s(k)$. Strictly speaking, the flat fading channel model does not satisfy the first condition, because the band of $c(k)$ resides at the lower end of the band of s(k), as shown in Figure 3-2. The difficulty of evaluating $|c(k)|^2$ comes from the evaluation of $E(|r(k)|^2)$.

3.3.2 Frequency Domain Interpretation

The result from last subsection suggests that $|c(k)|^2$ can somehow be estimated from $|r(k)|^2$. Also the time-averaging operation in (26) basically can be considered as a lowpass filter which extracts the low frequency components of $|r(k)|^2$, thus it is of interest to investigate $|r(k)|^2$ in frequency domain. The random nature of $|r(k)|^2$ lends itself to statistical signal processing which dictates the evaluation of power spectra (or probability spectral density function) for the signals in our channel model.

In order to derive the power spectrum for $|r(k)|^2$, $r(k)$ must be a stationary signal. This can be achieved by adding another dimension of randomness to the experiment described in last subsection. By randomizing all factors having effects on fading signal $c(k)$, such as routes and vehicle speeds, $c(k)$ becomes a stationary random signal. Since $c(k)$, $s(k)$ and $n(k)$ all become stationary, $r(k)$ is stationary. This generalization of the hypothetical experiment described in Section 3.3.1 simply leads us back the original

channel model (7).

The power spectrum of $|r(k)|^2$, which is derived in Appendix B, is given as

$$S_{r^2}(f)=\sigma_s^4\left[S_{c^2}(f)-\delta(f)\right]+\left(2\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4\right)+\left(\sigma_s^2+\sigma_n^2\right)^2\delta(f). \qquad (27)$$

The power spectrum in (27) consists of three components: signal spectrum, noise spectrum and a DC component as shown in Figure 3-3. The first term is a scaled version of the power spectrum of $|c(n)|^2$, which is to be estimated, therefore it is called signal spectrum. The second term, $(2\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4)$, is called noise because it represents the variation in $|r(n)|^2$ caused by $s(n)$ and $n(n)$, and is unwanted. Finally, the third term, $(\sigma_s^2+\sigma_n^2)^2\delta_K(f)$, represents a DC value, which is the DC component of the squared magnitude of the fading signal, $|c(k)|^2$, in time domain.

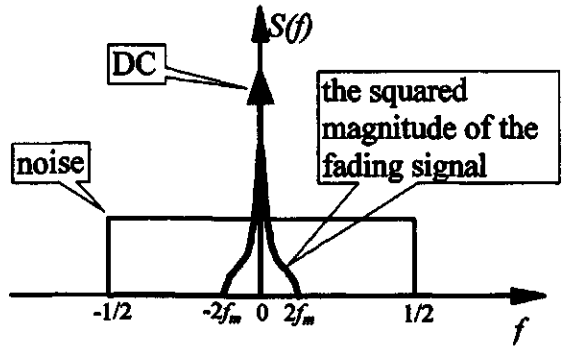


Figure 3-3 Components of the Power Spectrum for $|r(k)|^2$

From the power spectrum of $|r(n)|^2$, we can recognize two factors that explain why channel fading signal can be estimated by lowpass filtering $|r(n)|^2$. The first one is the additive relationship between signal and noise. That is the spectrum of $|c(n)|^2$ and that of the noise have an additive relationship in (27). This is only possible for the magnitude squared of the received signal $r(n)$, because, if we look at the power spectrum for $r(n)$ proper, the power spectrum of fading signal, $c(n)$, and that of "noise", $s(n)$, have a

multiplicative relationship. The additive relationship between the squared magnitude of the fading signal and the noise means that they can be separated by linear filtering.

There is an intuitive explanation for the existence of power spectrum of $|c(n)|^2$. Consider the pilot tone technique which is described in Section 1.3. In pilot tone assisted channel fading estimation, a single-frequency probing tone is transmitted along with the information-bearing signal. After propagation over fading channels, the probing tone becomes a randomly varying signal carrying the channel fading information, which manifests itself as a power spectrum of the fading signal located around the pilot tone in frequency domain. In the technique proposed in this thesis, the power spectrum of $|c(n)|^2$ in $|r(n)|^2$ is generated using the same mechanism but without sending a pilot tone explicitly. Because there is a DC component in $|s(n)|^2$, and since $|s(n)|^2$ is modulated by $|c(n)|^2$, the DC component of $|s(n)|^2$ simply serves as a pilot tone and generates a power spectrum of $|c(n)|^2$ around zero frequency. The advantage of this "pilot tone" is that it does not consume transmitter power; the disadvantage is that it is not as perfect as an explicit tone, because it may take a very long observation period for the average signal power to closely approach this DC value.

The second factor that makes lowpass filtering work is that the power spectrum of $|c(n)|^2$ has a narrow bandwidth and is located at low frequency. From Figure 3-3, the bandwidth of $|c(n)|^2$ is $4f_m$, while the bandwidth of the noise component occupies the whole frequency range. Thus the estimate of $|c(n)|^2$ is corrupted by "noise" from the transmitted signal, and there is an SNR associated with the fading signal estimate $|c(n)|^2$. For a typical maximum fading rate, which definition is given in Section 1.2, say 5%, the ratio of bandwidth of $|c(n)|^2$ to that of noise is 20% (4 times 5%). This means by simply extracting the bandwidth of $|c(n)|^2$, we can expect significant improvement in SNR. This

explains why a simple lowpass filter can be used to estimate $|c(n)|^2$ from $|r(n)|^2$.

3.3.3 Linear Minimum Mean Squared Error Estimator

A simple lowpass filter only takes advantage of the knowledge about the fading signal bandwidth. It improves SNR but is far from being optimum. In this section, we discuss the design of a linear minimum mean square error (MMSE) estimator using our knowledge about the power spectrum (or equivalently auto-correlation function) of the fading signal.

The MMSE estimator in our scheme has two modes of operation: block processing (or batch processing) and filtering. Block processing is the operation mode applied when the frame synchronizer starts to operate and has just collected the first full frame observation of the received symbols. At this time, the function of channel fading estimation is to provide the best initial estimates based on the first frame of received symbols, so that the frame synchronizer can immediately start detecting the pilot sequence based on the fading signal made available by block processing. After the fading estimator has processed the first frame of received symbols, initial values necessary for the state variables in an estimation filter are obtained. The estimation filter can be an FIR or IIR implementation. The fading estimator can properly work in filtering mode, as the filter is tuned to the fading signal. Compared to block processing mode, filtering mode has less computational complexity and is fast.

The block processing estimator is virtually a bank of FIR Wiener filters, and is expressed as

$$\mathbf{c} = \mathbf{H}(\mathbf{r} - \boldsymbol{\mu}_r) + \boldsymbol{\mu}_c, \tag{28}$$

where $\mathbf{c}$ is a full frame of channel fading estimates, $\mathbf{c} = [\,|c_0|^2, |c_1|^2, \ldots |c_{N-1}|^2\,]^T$; H is N×N

matrix of filter coefficients; **r** is a full frame of received symbols (magnitude squared), **r** = [$|r_0|^2$, $|r_1|^2$, ... $|r_{N-1}|^2$]^T; μ_r and μ_c are vectors of means of received symbols and channel fading estimates, i.e. $\mu_r = E\{\mathbf{r}\}$and $\mu_c = E\{\mathbf{c}\}$. Taking ensemble average on both sides of (20), we have $\mu_r = (\sigma_s^2 + \sigma_n^2)[\ 1\ 1\ \ldots\ 1\ ]^T$. By the assumption of unit variance for channel fading signal, we have $\mu_c = [\ 1\ 1\ \ldots 1\ ]^T$. H is calculated by

$$H = C_{rr}^{-1} C_{cr}, \tag{29}$$

where C_{rr} is a N×N covariance matrix of a frame of received symbols, $C_{rr} = [c_{rr}(i, j)]$ = [$E\{|r_i|^2|r_j|^2\} - \sigma_r^4$]; C_{cr} is a N×N cross-covariance matrix of a frame of channel fading signal and a frame of received symbols, $C_{cr} = [c_{cr}(i, j)] = [\ E\{(|c_i|^2-\sigma_c^2)(|r_j|^2-\sigma_r^2)\}\ ]$.

The element $c_{rr}(i, j)$ can be expressed in auto-correlation of $|r|^2$ as

$$c_{rr}(i,j) = R_{r^2}(i-j) - \sigma_r^4 \tag{30}$$

Inserting (104) of Appendix B into (30) gives

$$c_{rr}(i,j) = \sigma_s^4\left[R_{c^2}(i-j)-1\right] + \left(2\sigma_s^4 + 2\sigma_s^2\sigma_n^2 + \sigma_n^4\right)\delta_k(i-j). \tag{31}$$

From Appendix A we know

$$\begin{aligned} R_{c^2}(k) &= 4R_{C_I}^2(k) + 1 \\ &= 4J_0^2(2\pi k f_m) + 1. \end{aligned} \tag{32}$$

Inserting (32) into (31) gives

$$c_{rr}(i,j) = 4\sigma_s^4 J_0^2(2\pi(i-j)f_m) + \left(2\sigma_s^4 + 2\sigma_s^2\sigma_n^2 + \sigma_n^4\right)\delta_k(i-j). \tag{33}$$

The element $c_{cr}(i, j)$ of C_{cr} is derived as follows. First the cross-covariance can be expressed in cross-correlation as

$$
\begin{aligned}
c_{cr}(i,j) &= E\left\{\left(|c_i|^2-\sigma_c^2\right)\left(|r_j|^2-\sigma_r^2\right)\right\} \\
&= E\left\{|c_i|^2|r_j|^2\right\}-\sigma_c^2\sigma_r^2.
\end{aligned} \tag{34}
$$

The cross-correlation between $|c_i|^2$ and $|r_j|^2$ can be expanded using (7) as

$$
\begin{aligned}
R_{c^2r^2}(i,j) &= E\left\{|c_i|^2|r_j|^2\right\} \\
&= E\left\{|c_i|^2\left[|c_j|^2|s_j|^2+|n_j|^2+2\mathrm{Re}\left(c_j s_j n_j^*\right)\right]\right\} \\
&= R_{c^2}(i-j)\sigma_s^2+\sigma_c^2\sigma_n^2.
\end{aligned} \tag{35}
$$

Inserting (35) into (34) and gives

$$
\begin{aligned}
c_{cr}(i,j) &= \sigma_s^2 R_{c^2}(i-j)+\sigma_n^2-\sigma_r^2 \\
&= \sigma_s^2 R_{c^2}(i-j)+\sigma_n^2-\left(\sigma_s^2+\sigma_n^2\right) \\
&= \sigma_s^2\left[R_{c^2}(i-j)-1\right].
\end{aligned} \tag{36}
$$

In deriving (36) we assumed that $\sigma_c^2 = 1$. Inserting (32) into (36) gives

$$
c_{cr}(i,j) = 4\sigma_s^2 J_0^2\left(2\pi(i-j)f_m\right). \tag{37}
$$

The vector of Mean Squared Errors (MSE) for the estimates is expressed as

$$
\mathbf{e}_{\mathrm{MS}} = [1\ldots1]\times \mathrm{I}\times\left[\mathrm{C}_{cc}-\mathrm{H}^{\mathrm{T}}\mathrm{C}_{cr}\right]. \tag{38}
$$

where $\mathbf{e}_{\mathrm{MS}}$ is a vector, which elements are the MSE for each estimate in **c**. The first factor on the right side of (38) is a row vector which elements are unit value. The matrix I is the identity matrix. The row vector and unit matrix serves to extract the diagonal elements of the rest expression on the right of (38). The matrix C_{cc} is a matrix of auto-covariance of $|c|^2$, the element is defined as

$$
c_{cc}(i,j) = R_{c^2}(i-j)-\sigma_c^4 = 4R_{c_I}^2(i-j) = 4J_0^2(i-j).
$$

C_{cr} is a matrix of cross-covariance between $|c|^2$ and $|r|^2$; the element is defined as in (37).

The error vector is a function of filter length (i.e. the length of observation vector), fading rate and SNR of the channel. The closed form expression of the dependence is

very difficult to derive and will not be further pursued in this thesis. However, for design purposes, we can numerically evaluate the effect of filter length, fading rate and SNR on MSE.

In order to gain some insight into the above equations, we give an example fading estimator that uses 2 received symbols and the closed expression of mean squared error for this simple case. From (29), the weights of fading estimator is calculated as

$$
\begin{aligned}
\mathbf{H}_{2\times 2} &= \begin{bmatrix} 4\sigma_s^4 J_0^2(0)+\left(2\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4\right) & 4\sigma_s^4 J_0^2(-2\pi f_m) \\ 4\sigma_s^4 J_0^2(2\pi f_m) & 4\sigma_s^4 J_0^2(0)+\left(2\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4\right) \end{bmatrix}^{-1} \\
&\times \begin{bmatrix} 4\sigma_s^2 J_0^2(0) & 4\sigma_s^2 J_0^2(-2\pi f_m) \\ 4\sigma_s^2 J_0^2(2\pi f_m) & 4\sigma_s^2 J_0^2(0) \end{bmatrix} \\
&= \begin{bmatrix} 6\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4 & 4\sigma_s^4 J_0^2(2\pi f_m) \\ 4\sigma_s^4 J_0^2(2\pi f_m) & 6\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4 \end{bmatrix}^{-1} \\
&\times 4\sigma_s^2 \begin{bmatrix} 1 & J_0^2(2\pi f_m) \\ J_0^2(2\pi f_m) & 1 \end{bmatrix} \\
&= \frac{4}{\sigma_s^2\left\{\left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)^2-16J_0^4(2\pi f_m)\right\}} \\
&\times \begin{bmatrix} 6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4} & -4J_0^2(2\pi f_m) \\ -4J_0^2(2\pi f_m) & 6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4} \end{bmatrix} \times \begin{bmatrix} 1 & J_0^2(2\pi f_m) \\ J_0^2(2\pi f_m) & 1 \end{bmatrix} \\
&= \frac{4}{\sigma_s^2\left\{\left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)^2-16J_0^4(2\pi f_m)\right\}} \\
&\times \begin{bmatrix} 6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}-4J_0^4(2\pi f_m) & \left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)J_0^2(2\pi f_m) \\ \left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)J_0^2(2\pi f_m) & 6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}-4J_0^4(2\pi f_m) \end{bmatrix}
\end{aligned}
\tag{39}
$$

In reduction of (39) we used the following properties of the Bessel function

$$J_0(0)=1,$$
$$J_0(x)=J_0(-x).$$

The error vector is calculated from (38) as

$$\begin{aligned}
\mathbf{e}_{\mathrm{MS}}^{\mathrm{T}} &= \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\left\{4\begin{bmatrix} J_0^2(0) & J_0^2(-2\pi f_m)\\ J_0^2(2\pi f_m) & J_0^2(0)\end{bmatrix}-\mathrm{H}_{2\times 2}^{\mathrm{T}}\times 4\sigma_s^2\begin{bmatrix} J_0^2(0) & J_0^2(-2\pi f_m)\\ J_0^2(2\pi f_m) & J_0^2(0)\end{bmatrix}\right\}\\
&= \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\left\{4\begin{bmatrix} 1 & J_0^2(2\pi f_m)\\ J_0^2(2\pi f_m) & 1\end{bmatrix}-\mathrm{H}_{2\times 2}^{\mathrm{T}}\times 4\sigma_s^2\begin{bmatrix} 1 & J_0^2(2\pi f_m)\\ J_0^2(2\pi f_m) & 1\end{bmatrix}\right\}\\
&= \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\left\{4\begin{bmatrix} 1 & J_0^2(2\pi f_m)\\ J_0^2(2\pi f_m) & 1\end{bmatrix}-\right.\\
&\quad \frac{16}{\left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)^2-16J_0^4(2\pi f_m)}\begin{bmatrix} 6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}-4J_0^4(2\pi f_m) & \left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)J_0^2(2\pi f_m)\\ \left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)J_0^2(2\pi f_m) & 6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}-4J_0^4(2\pi f_m)\end{bmatrix} \qquad (40)\\
&\quad \left.\times\begin{bmatrix} 1 & J_0^2(2\pi f_m)\\ J_0^2(2\pi f_m) & 1\end{bmatrix}\right\}\\
&= \left[\frac{\left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)\left(8+8\frac{\sigma_n^2}{\sigma_s^2}+4\frac{\sigma_n^4}{\sigma_s^4}\right)-\left(32-2\frac{\sigma_n^2}{\sigma_s^2}-\frac{\sigma_n^4}{\sigma_s^4}\right)J_0^4(2\pi f_m)}{\left(6+2\frac{\sigma_n^2}{\sigma_s^2}+\frac{\sigma_n^4}{\sigma_s^4}\right)^2-16J_0^4(2\pi f_m)}\right]\begin{bmatrix}1 & 1\end{bmatrix}.
\end{aligned}$$

From (40) it can be seen that the estimation error is a functional dependent of the channel fading rate f_m and SNR σ_s^2/σ_n^2. What is not so obvious is that this error is also dependent of the filter length, i.e. the length of observation. As the filter length increases, more received symbols are used to calculate the fading estimate, the estimation error decreases.

The filtering mode of fading estimator can be implemented as FIR filter or IIR filter. [The FIR filter is simply the bottom row of the weight matrix H as defined in (39).] An IIR filter can take advantages of all available observations, i.e. the length of

observation, upon which an IIR filter calculates the estimate, increases as more and more data is received. Therefore, it is superior to a FIR filter. However, the design of an IIR Wiener filter requires factorization of the power spectrum of received signal $|r(k)|^2$. Because there is no closed form expression for $|r(k)|^2$, as is shown in Appendix B, it increases the difficulty of IIR filter design. Therefore, IIR filter is not further discussed in this thesis.

3.4 Pilot Sequence Detector

For a fading channel there are three factors that obstruct correct detection of pilot sequence. These are channel fading, channel noise and the data portion of transmitted signal. Using the scheme described in last section, the receiver can obtain an estimate of the channel fading signal; therefore, it is possible for the receiver to first eliminate the effect of channel fading and then only deal with the effects of noise and the random data portion of the transmitted signal. Further simplification is possible when using only pilot-spaced observations of the received signal in detecting the pilot sequence. It will become obvious why these simplifications are valid when the following questions are answered in this section:

- Why can a fading channel be converted approximately into a Gaussian noise channel under high SNR using the knowledge of $|c(n)|^2$?
- Why can the detection be achieved using only pilot-spaced observations?

3.4.1 Reducing to Additive Noise Channels

The benefit of explicit estimation of channel fading signal is that given channel state information, detection of pilot sequence on fading channels can be achieved using

traditional detection methods developed for Gaussian noise channels. We shall prove this point by showing that the p.d.f. of received signal (conditioned on the transmitted signal) approximates to a Gaussian distribution.

Assuming perfect knowledge of the channel fading signal, i.e. ignoring the error induced by channel fading estimation, we can think of the fading channel as an equivalent combination of a hypothetical transmitter and a Gaussian noise channel as shown in Figure 3-4. The hypothetical transmitter first modulates the transmitted signal $s(n)$ with a fading signal $c(n)$, then transmits the modulated signal $c(n)s(n)$ over a Gaussian noise channel. For simplicity of notation, we shall first drop $c(n)$ in deriving p.d.f. for $|r(n)|^2$. After we get the final expression, substitution of $c(n)s(n)$ for $s(n)$ gives the suitable result for a fading channel.

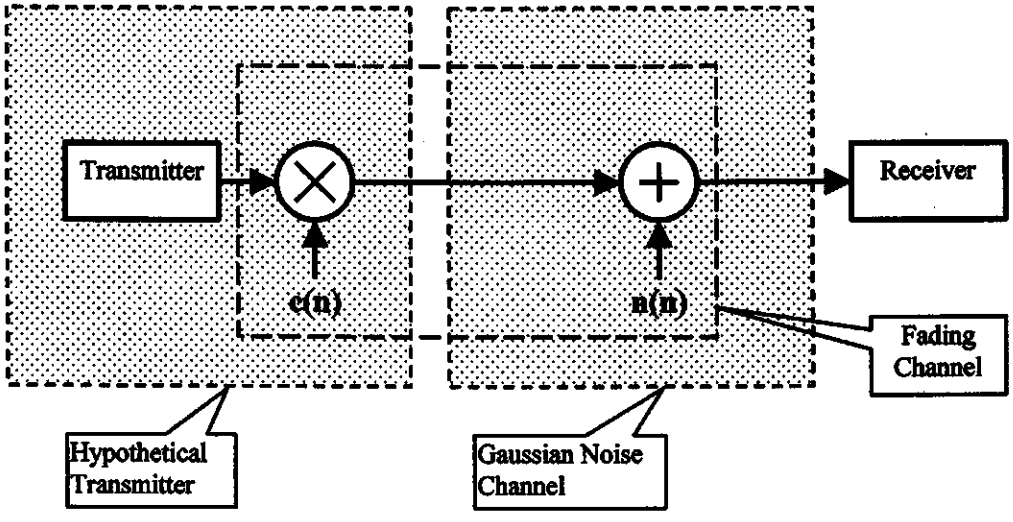


Figure 3-4 Hypothetical Transmitter and Gaussian Noise Channel

It is well known that, for a Gaussian noise channel, the p.d.f. of received signal is

$$f(r \mid s) = \frac{1}{\pi\sigma^2} \exp\left[-\frac{(r-s)(r-s)^*}{\sigma^2} \right], \qquad (41)$$

where σ^2 is the covariance for r and s is defined as $E[(r-s)(r-s)^*]$. By expressing r and s in their phasor forms ($|r|$, θ) and ($|s|$,α), we obtain

$$f((|r|,\theta)|(|s|,\alpha)) = \frac{|r|}{\pi\sigma^2} \exp\left(\frac{-(|r|^2 + |s|^2)}{\sigma^2} \right) \exp\left(\frac{2|r||s|\cos(\theta-\alpha)}{\sigma^2} \right). \qquad (42)$$

As discussed in Section 3.2, the pilot sequence is expressed in signal magnitude, therefore, we are only interested in the p.d.f. of $|r|$ conditioned on $|s|$. This p.d.f. $f(|r| \mid |s|)$ is obtained in the following steps.

First, the p.d.f. in (42) is integrated with respect to θ over $[0, 2\pi)$ for a specific value of α, which gives

$$f\left(|r|\left|\left(|s|,\alpha=\frac{\pi}{4}\right)\right.\right)=\int_{-\pi}^{\pi} f\left(\left(|r|,\theta\right)\left|\left(|s|,\alpha=\frac{\pi}{4}\right)\right.\right)d\theta$$

$$=\int_{-\pi}^{\pi}\frac{|r|}{\pi\sigma^2}\exp\left(\frac{-\left(|r|^2+|s|^2\right)}{\sigma^2}\right)\exp\left(\frac{2|r||s|\cos\left(\theta-\frac{\pi}{4}\right)}{\sigma^2}\right)d\theta$$

$$=\frac{|r|}{\pi\sigma^2}\exp\left(\frac{-\left(|r|^2+|s|^2\right)}{\sigma^2}\right)\int_{-\pi}^{\pi}\exp\left(\frac{2|r||s|\cos\left(\theta-\frac{\pi}{4}\right)}{\sigma^2}\right)d\theta \qquad (43)$$

$$=\frac{|r|}{\pi\sigma^2}\exp\left(\frac{-\left(|r|^2+|s|^2\right)}{\sigma^2}\right)\int_{-\pi}^{\pi}\exp\left(\frac{2|r||s|\cos(\theta)}{\sigma^2}\right)d\theta$$

$$=\frac{2|r|}{\sigma^2}\exp\left(\frac{-\left(|r|^2+|s|^2\right)}{\sigma^2}\right)I_0\left(\frac{2|r||s|}{\sigma^2}\right).$$

The third line in (43) results because the kernel of the integral is a periodic function with a period of 2π and the interval of the integral is also 2π. The fourth line is obtained using the following equation from [10]

$$I_0(x)=\int_{-\pi}^{\pi}\exp(x\cos\theta)d\theta. \qquad (44)$$

where $I_0(x)$ is the modified Bessel function of the first kind and order zero. The p.d.f. for cases where the phase α is $3/4\pi$, $-\pi/4$, and $-3/4\pi$ is the same as (43).

The second step is to derive the p.d.f. $f(|r| \mid |s|)$. This requires taking into account all the symbols with the same magnitude but different phase in the 16-QAM constellation. Based on the assumption that all the symbol in the 16-QAM constellation is equally likely to be sent, the probability of sending a symbol with a specific phase

conditioned on $|s|$ is equal. For example, when $|s|^2$ is $2E$, the conditional probabilities for each possible phase are

$$\begin{aligned} P\left(\alpha=\frac{\pi}{4}\,||s|=\sqrt{2E}\right) &= P\left(\alpha=\frac{3\pi}{4}\,||s|=\sqrt{2E}\right) \\ &= P\left(\alpha=-\frac{\pi}{4}\,||s|=\sqrt{2E}\right) \\ &= P\left(\alpha=-\frac{3\pi}{4}\,||s|=\sqrt{2E}\right) \\ &= \frac{1}{4}. \end{aligned} \tag{45}$$

The p.d.f. $f(|r|\,|\,|s|)$ for the case $|s|^2 = 2E$ is

$$\begin{aligned} f\left(|r|\,||s|=\sqrt{2E}\right) &= f\left(|r|\,|\left(|s|=\sqrt{2E},\alpha=\frac{\pi}{4}\right)\right)P\left(\alpha=\frac{\pi}{4}\right) \\ &+ f\left(|r|\,|\left(|s|=\sqrt{2E},\alpha=\frac{3\pi}{4}\right)\right)P\left(\alpha=\frac{3\pi}{4}\right) \\ &+ f\left(|r|\,|\left(|s|=\sqrt{2E},\alpha=-\frac{\pi}{4}\right)\right)P\left(\alpha=-\frac{\pi}{4}\right) \\ &+ f\left(|r|\,|\left(|s|=\sqrt{2E},\alpha=-\frac{3\pi}{4}\right)\right)P\left(\alpha=-\frac{3\pi}{4}\right) \\ &= f\left(|r|\,|\left(|s|=\sqrt{2E},\alpha=\frac{\pi}{4}\right)\right). \end{aligned} \tag{46}$$

Similarly, the same result can be derived for other values of $|s|$. In summary, for each of the three rings in the 16-QAM constellation, the p.d.f. $f(|r|\,|\,|s|)$ is

$$f\left(|r|\,||s|\right)=\frac{2|r|}{\sigma^2}\exp\left(\frac{-\left(|r|^2+|s|^2\right)}{\sigma^2}\right)I_0\left(\frac{2|r||s|}{\sigma^2}\right). \tag{47}$$

Under high SNR, i.e. $\frac{2|r||s|}{\sigma^2} >> 1$, the Bessel function can be approximated as

$\sigma^2 = \sigma_{rs}^2$ covariance of r and s.

$$I_0(x) \approx \frac{\exp\left(|x| - \frac{1}{2}\ln|x|\right)}{\sqrt{2\pi}} \approx \frac{\exp(|x|)}{\sqrt{2\pi}}. \tag{48}$$

Substitution of (48) into (47) gives

$$\begin{aligned} f(|r|\,||s|) &\approx \frac{2|r|}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-\left(|r|^2+|s|^2\right)}{\sigma^2}\right) \exp\left(\frac{2|r||s|}{\sigma^2}\right) \\ &\approx \frac{2|r|}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(|r|-|s|)^2}{\sigma^2}\right). \end{aligned} \tag{49}$$

Finally, we can obtain the conditional p.d.f. of $|r|^2$ from (49) as

$$f\left(|r|^2 \,||s|^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(|r|-|s|)^2}{\sigma^2}\right). \tag{50}$$

Equation (50) implies that, under high SNR, the conditional p.d.f. of magnitude squared $|r|^2$ of received signal is a Gaussian distribution in terms of $|r|$. Note that this Gaussian function is not expressed in $|r|^2$. Therefore, the p.d.f. of $|r|^2$ is not a Gaussian distribution. However, this Gaussian function in $|r|$ makes it very easy to simplify the computation of the conditional p.d.f. for a sequence of symbols. As mentioned earlier in this section, to make (50) suitable for a fading channel, we replace $|s|$ with $|cs|$ and obtain

$$f\left(|r|^2 \,||s|^2, |c|^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(|r|-|c||s|)^2}{\sigma^2}\right). \tag{51}$$

Equation (51) is the p.d.f. of squared magnitude for a single received symbol. It is the building block for expressions to be derived in the following section, which are used to calculate conditional p.d.f. for a sequence of received symbols. Thus the variables on the right side of (51) represent the information that the receiver needs to know to carry out frame synchronization. Since the channel fading signal appears as $|c|$, it is

straightforward to conclude that the frame synchronizer does not require full information about channel fading (phase and amplitude) but rather only $|c|^2$. This is a conclusion used in Section 3.3.

3.4.2 Detection of the Pilot Sequence

There are two types of frame formats commonly used in traditional frame synchronization: block format and interleaved format. Pilot symbol assisted modulation uses an interleaved frame format. In this format, pilot symbols are interleaved with data symbols in a fashion shown in Figure 3-5. In Figure 3-5, pilot symbols are represented by P_i, where i is the index of pilot symbol into the pilot sequence. P_0, P_1, ...P_{N-1} form the pilot sequence. Data symbols are represented by D_j, where j is the index of data symbol into the subframe. A subframe consists of a pilot symbol followed by M-1 data symbols, i.e. P_i, D_1, D_2,...D_{M-1} form one subframe. One frame contains N subframes. N is the length of a pilot sequence. One subframe has M symbols. Therefore, one frame has total M×N symbols. One frame of received symbols is called a *full frame observation*. If we select one symbol at a fixed index from each subframe and concatenate them into an order-preserving sequence, we obtain a *pilot-spaced observation* of a full frame observation. There are M different pilot-spaced observations in a full frame observation.

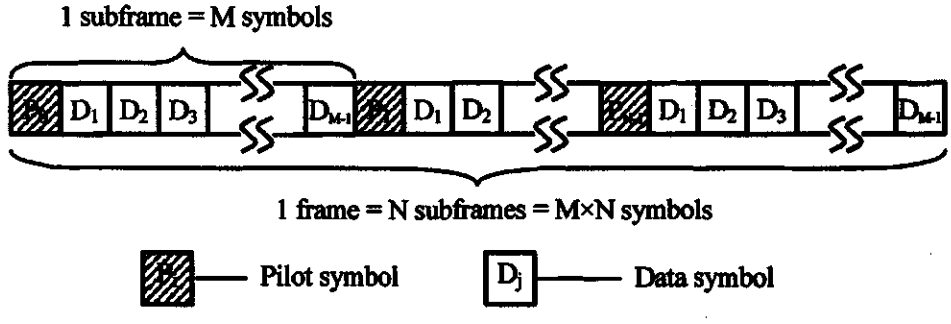
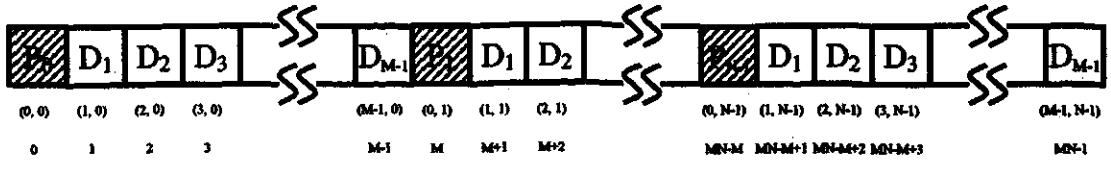


Figure 3-5 Frame Format for PSAM

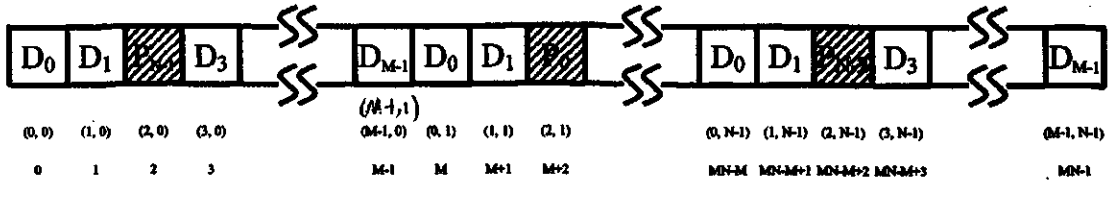
Frame synchronization for PSAM is to determine the relative position of the first pilot symbol P_0. If the frame synchronization happens after the hard decision stage of a receiver, the input symbols to a frame synchronizer are hard decisions, and free of channel fading and noise except for an acceptable symbol error rate, then the task can be easily accomplished by selecting the pilot-spaced observation that produces maximum correlation with the pilot sequence. However, frame synchronization for PSAM must be achieved before the receiver can make any hard decisions on received symbols, therefore the input to the frame synchronizer is a series of raw received symbols. These raw symbols are distorted by channel fading and noise. The correlation method can only produce optimal results for the case where the noise is additive Gaussian noise and the transmitted signals are equally possible. Therefore it is not competent for this type of received signals. To design a better frame synchronization scheme, the effects of channel fading and noise must be accounted for. In the literature, the effect of noise was addressed much earlier than that of the channel fading. Massey[11] introduced the optimum frame synchronization for binary Gaussian channels using Maximum Likelihood detection theory. Lui and Tan[12] extended Massey's work to general M-ary modulation formats. Because both Massey and Lui's schemes targeted at block frame format, in which the synchronization word (equivalent to the pilot sequence) is inserted in a contiguous form at the front of each frame, their results do not apply to PSAM frame synchronization in a straightforward fashion. In this subsection, we extend Maximum Likelihood frame synchronization to interleaved frame format.

Let μ be the index of pilot symbol P_0 within a full frame, the boldface indicates it

is a random variable; let μ be the value taken by μ, μ is an integer ranging over [0, M×N-1]. Maximum Likelihood estimation of the location of the pilot sequence is to search for the value of μ that maximizes the likelihood function $f_x(x|\mu=\mu)$, where x is a full frame observation. For the interleaved frame format, μ can be equivalently expressed as a pair of indices (μ_1, μ_2). The index μ_1 specifies the pilot-spaced observation that is made up of pilot symbols, and ranges over [0, M-1]. The index μ_2 is the position of the first pilot symbol, P_0, within this pilot-spaced observation, and ranges over [0, N-1]. The relation between μ and (μ_1, μ_2) is $\mu = \mu_2 \times M + \mu_1$. Figure 3-6 gives two examples that illustrate the definition of (μ_1, μ_2). Under each frame bar are the indices of each symbol, expressed using two-index system (upper row) and one-index system (lower row). The likelihood function $f_x(x|\mu=\mu)$ can be equivalently expressed as $f_x(x \mid (\mu_1,\mu_2) = (\mu_1,\mu_2))$.



P_0 index: 2 index system $\mu_1 = 0$, $\mu_2 = 0$; one index system $\mu = 0$.



P_0 index: 2 index system $\mu_1 = 2$, $\mu_2 = 1$; one index system $\mu = M+2$.

Figure 3-6 Two Examples Illustrating μ_1, μ_2.

To derive the likelihood function, we define the following notations:

- The pilot-spaced observation $x_p(i)$ is the collection of symbols within a full

frame that starts at the i^{th} position and M symbols apart from each other. Subscript p stands for pilot-spaced.

- The shifted pilot sequence $p(i)$ is the original pilot sequence cyclically right shifted by i symbols.
- The pilot spaced data symbol sequence d_p is any of the possible data sequences. The data symbols in these sequences are M symbols apart from each other when they appear in a full frame of observation. All possible data symbol sequences form a set of data sequences, D.

To facilitate concise expression, we permute the symbols within one full frame observation, such that symbols from the same pilot-spaced subframe concatenate back to back, and the pilot-spaced subframe that containing pilot sequence is placed at the beginning of the permuted frame. This is illustrated in Figure 3-7.

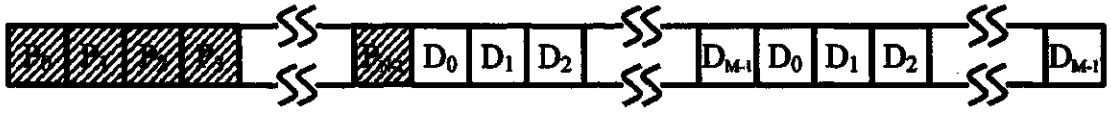


Figure 3-7 Illustration of Symbol Permutation within a Full Frame

The likelihood function for the full frame observation in Figure 3-7 can be expressed as

$$f(x \mid \mu_1, \mu_2) = f(x_p(\mu_1) \mid p(\mu_2)) \prod_{i=0,\, i \neq \mu_1}^{N-1} \sum_{d_p \in D} f(x_p(i) \mid d_p) P(d_p) \qquad (52)$$

The likelihood function in (52) can be simplified by dividing both sides with the expression $\prod_{i=0}^{N-1} \sum_{d_p \in D} f(x_p(i) \mid d_p) P(d_p)$, which is independent of μ_1 and μ_2. Note that this will cancel the part on the right side of (52) that is a functional of x_p(i)'s for i from 0 to

N-1 except i = μ_1. The resultant expression, which is no longer a likelihood function proper and is called decision metric from now on, is

$$\text{Metric} = \frac{f(x_p(\mu_1)|\, p(\mu_2))}{\sum_{d_p \in D} f(x_p(\mu_1)|\, d_p) P(d_p)}. \tag{53}$$

From the procedure for deriving the metric in (53), it is obvious that the search for (μ_1, μ_2) that maximizes the likelihood function in (52) can be equivalently done by the search for (μ_1, μ_2) that maximizes the decision metric in (53). Because the metric only requires computation for a pilot-spaced subframe of observation, using the metric instead of the likelihood function saves computation.

The denominator of the expression in (53) requires calculation over the set of all possible data symbol sequences, which can be a formidable task for large frames. To further reduce computation, we notice that contributions from each term in the summation are not equal. The data symbol sequence that is closest to the pilot-spaced observation x_p will contribute most significantly to the sum. We call this term the dominant term, and use the dominant term as an approximation to the sum. This approximation gives good performance in high SNR cases. Thus the decision metric is further reduced to

$$\text{Metric} = \frac{f(x_p(\mu_1)|\, p(\mu_2))}{f(x_p(\mu_1)|\, d_{dom}) P(d_{dom})}. \tag{54}$$

where d_{dom} is the data sequence that is closest to the received pilot-spaced observation $x_p(\mu_1)$.

From (50), $f(x_p(\mu_1)|p(\mu_2))$ can be expressed as

$$f\left(x_p(\mu_1)\mid p(\mu_2)\right)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left\{-\frac{\left[x_p(\mu_1)^{1/2}-p(\mu_2)^{1/2}\right]^T\left[x_p(\mu_1)^{1/2}-p(\mu_2)^{1/2}\right]}{\sigma^2}\right\} \tag{55}$$

where σ^2 is the variance of received symbol, $x_p(\mu_1)^{1/2}$ is the vector which elements are square roots of those in $x_p(\mu_1)$, $p(\mu_2)^{1/2}$ is the vector which elements are square roots of those in $p(\mu_2)$. The denominator of (54), $f(x_p(\mu_1)|d_{dom})$, can be expanded as

$$f\left(x_p(\mu_1)\mid d_{dom}\right)=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^2 \exp\left\{-\frac{\left[x_p(\mu_1)^{1/2}-d_{dom}^{1/2}\right]^T\left[x_p(\mu_1)^{1/2}-d_{dom}^{1/2}\right]}{\sigma^2}\right\} \tag{56}$$

Inserting (55) and (56) into (54) and taking logarithm gives

$$\begin{aligned}\text{Metric} &= \left[x_p(\mu_1)^{1/2}-d_{dom}^{1/2}\right]^T\left[x_p(\mu_1)^{1/2}-d_{dom}^{1/2}\right] \\ &\quad -\left[x_p(\mu_1)^{1/2}-p(\mu_2)^{1/2}\right]^T\left[x_p(\mu_1)^{1/2}-p(\mu_2)^{1/2}\right]-\ln P(d_{dom}) \\ &= 2\left(x_p(\mu_1)^{1/2}\right)^T\left(p(\mu_2)^{1/2}\right)-\left(p(\mu_2)^{1/2}\right)^T\left(p(\mu_2)^{1/2}\right) \\ &\quad -2\left(x_p(\mu_1)^{1/2}\right)^T\left(d_{dom}{}^{1/2}\right)+\left(d_{dom}{}^{1/2}\right)^T\left(d_{dom}{}^{1/2}\right) \\ &\quad -\ln P(d_{dom})\end{aligned} \tag{57}$$

The second term on the right side of (57) is expanded as

$$\left(p(\mu_2)^{1/2}\right)^T\left(p(\mu_2)^{1/2}\right)=\sum_{i=0}^{N-1}|P_i|^2. \tag{58}$$

Because for a given pilot sequence, its norm is constant for all shifted versions, thus the expression in(58) is not a functional of either μ_2 or μ_1. We can drop the second term from the expression. Note that d_{dom} is related to the pilot-spaced observation $x_p(\mu_1)$, so d_{dom} is dependent on μ_1. Including these two considerations, the expression of the decision metric is modified as

$$\begin{aligned}\text{Metric} &= 2\left(x_p(\mu_1)^{1/2}\right)^T\left(p(\mu_2)^{1/2}\right)-2\left(x_p(\mu_1)^{1/2}\right)^T\left(d_{dom}(\mu_1)^{1/2}\right) \\ &\quad +\left(d_{dom}(\mu_1)^{1/2}\right)^T\left(d_{dom}(\mu_1)^{1/2}\right)-\ln P(d_{dom}(\mu_1)).\end{aligned} \tag{59}$$

There still remains a question about the expression in (59): how do we determine

the data sequence d_{dom} of the dominant term. Recall from (54) that d_{dom} generates the maximum value for the expression $f(x_p(\mu_1)|\, d)P(d)$ among all possible data sequences. Since the elements in d is assumed independent random variables and the channel noise is white, we can expand the foregoing expression as

$$f(x_p(\mu_1)|\, d_{dom}(\mu_1))P(d_{dom}(\mu_1)) = \prod_{i=0}^{N-1} f(x_p(\mu_1,i)|\, d_{dom}(\mu_1,i))P(d_{dom}(\mu_1,i)), \qquad (60)$$

where $x_p(\mu_1,i)$ is the ith element of $x_p(\mu_1)$, $d_{dom}(\mu_1,i)$ is the ith element of $d_{dom}(\mu_1)$. To maximize the left-hand side of (60) is equivalent to maximize each factor on the right-hand side. The later can be achieved by [Maximum A Posteriori] (MAP) decision on each symbol of the pilot-spaced observation vector $x_p(\mu_1)$ assuming that data symbols are transmitted.

From Figure 3-1 we noticed that there are 3 levels of data symbol magnitude for a square 16-QAM constellation. These levels of symbol magnitude squared are $2E$, $10E$, and $18E$. The MAP decision device for magnitude detection can be implemented as a two-threshold slicer with the thresholds set somewhere in-between these levels. The decision thresholds are derived as follows.

The MAP decision problem is modeled as three-hypothesis problem. The hypotheses are defined as

H_{2E} : $|s| = \sqrt{2E}$ was transmitted,

H_{10E} : $|s| = \sqrt{10E}$ was transmitted,

H_{18E} : $|s| = \sqrt{18E}$ was transmitted.

Let x denote the magnitude squared of the observed symbol. The *a posterior* probability $P(H_{2E}|\, x)$ is obtained as

$\sqrt{x}$ is the original symbol magnitude $|r| = \sqrt{x}$

$$P(H_{2E}|x)=\frac{f_{X|H_{2E}}(x|H_{2E})P(H_{2E})}{f_X(x)}$$
$$=\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(\sqrt{|x|}-\sqrt{2E}\right)^2}{\sigma^2}\right)P(H_{2E})}{f_X(x)}. \quad (61)$$

The result is obtained using equation (50). The *a posteriori* probabilities for hypotheses H_{2E} and H_{18E} are obtained similarly as

$$P(H_{10E}|x)=\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(\sqrt{|x|}-\sqrt{10E}\right)^2}{\sigma^2}\right)P(H_{10E})}{f_X(x)}. \quad (62)$$

$$P(H_{18E}|x)=\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(\sqrt{|x|}-\sqrt{18E}\right)^2}{\sigma^2}\right)P(H_{18E})}{f_X(x)}. \quad (63)$$

Assuming the 16 symbols in 16-QAM constellation are equally likely to be transmitted, the probabilities of the three hypotheses are: $P(H_{2E}) = P(H_{18E}) = 1/4$, $P(H_{10E}) = 1/2$. Since $f_X(x)$ is a common denominator of the three probabilities, and we are only interested in their relative values, the three a posteriori probabilities can be reduced to

$$f_X(x \mid H_{2E})P(H_{2E}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(\sqrt{x}-\sqrt{2E}\right)^2}{\sigma^2}\right)\left(\frac{1}{4}\right),$$
$$f_X(x \mid H_{10E})P(H_{10E}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(\sqrt{x}-\sqrt{10E}\right)^2}{\sigma^2}\right)\left(\frac{1}{2}\right), \quad (64)$$
$$f(x \mid H_{18E})P(H_{18E}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\left(\sqrt{x}-\sqrt{18E}\right)^2}{\sigma^2}\right)\left(\frac{1}{4}\right).$$

Under high SNR condition, the *a posteriori* probabilities diminish quickly when the observation is far from the hypothesis value (i.e. the transmitted value). This means that when an observed value is between two hypotheses values, the third hypothesis contributes very little to the probability of detection error. This allows us to approximate the three-hypothesis test with two binary hypothesis tests, which lead to very simple solutions. The first binary hypothesis test is between H_{2E} and H_{10E} when the observation is less than 10E. The second binary hypothesis is between H_{10E} and H_{18E} when the observation is more than 10E. The case where the observation equals 10E can be arbitrarily combined with either hypothesis. Applying the MAP decision rule to the equations in (64), we can obtain two thresholds that divide the range of $\sqrt{x}$ into three regions corresponding to three data symbol magnitudes, $\sqrt{2E}$, $\sqrt{10E}$, and $\sqrt{18E}$. These thresholds are

$$\begin{cases} \text{Threshold1} = \dfrac{\sigma^2 \ln\left(\frac{1}{2}\right) + 8E}{2\left(\sqrt{10E}-\sqrt{2E}\right)}, \\ \text{Threshold2} = \dfrac{\sigma^2 \ln(2) + 8E}{2\left(\sqrt{18E}-\sqrt{10E}\right)}. \end{cases} \quad (65)$$

Note that the thresholds apply to the magnitude instead of magnitude squared of the observation. This is because the independent variables in the a posteriori probabilities

are expressed in magnitude of the observation. To illustrate the use of (65), let us consider a simple example. Assuming E is 1, and SNR is 20dB, the thresholds in (65) evaluate to $SNR = \frac{E[|h|^2]}{E[|n|^2]} = 10\log\frac{\tilde{E}}{\tilde{N}} = 10\log\frac{1}{\sigma_n^2} = 20 \Rightarrow \sigma_n^2 = \frac{1}{10}$.

$$\text{Threshold1} = \frac{\frac{1}{10}\ln\left(\frac{1}{2}\right)+8}{2\left(\sqrt{10}-\sqrt{2}\right)} \approx 2.27,$$
$$\text{Threshold2} = \frac{\frac{1}{10}\ln(2)+8}{2\left(\sqrt{18}-\sqrt{10}\right)} \approx 3.73. \tag{66}$$

The thresholds and hypothesis values are illustrated in Figure 3-8.

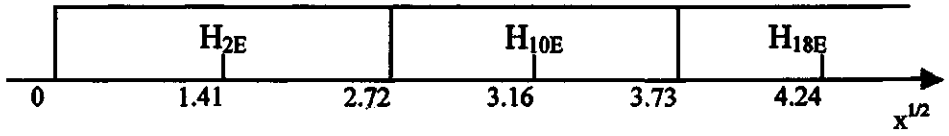


Figure 3-8 Illustration of the Decision Thresholds

The expressions in (65) need modification for fading channels. This is achieved by replacing $\sqrt{E}$ with $|c|\sqrt{E}$. This will cause $|c|$ appear in denominators. Because $|c|$ is only an estimate of the real fading signal, estimation error can well cause it to take on negative or zero values. Therefore, for negative estimation, we set $|c|$ to be zero, since true value for $|c|$ cannot be lower than zero. This means $|c|$ cannot appear in denominators. To avoid this problem, we use $|c|x_p(\mu_1,i)$ instead of $x_p(\mu_1,i)$ and the thresholds are modified to

$$\text{Threshold1} = \frac{\sigma^2 \ln\left(\frac{1}{2}\right) + 8E|c|^3}{2\left(\sqrt{10E} - \sqrt{2E}\right)},$$
$$\text{Threshold2} = \frac{\sigma^2 \ln(2) + 8E|c|^3}{2\left(\sqrt{18E} - \sqrt{10E}\right)}. \tag{67}$$

The above thresholds are MAP decision thresholds for the signal $|c|x_p(\mu_1,i)$. When $|c|x_p(\mu_1,i)$ is less than Threshold1 the decision is $d_{dom}(\mu_1,i) = 2E$; when larger than *Threshold1* but less than Threshold2 than $d_{dom}(\mu_1,i) = 10E$; when larger than Threshold2 then $d_{dom}(\mu_1,i) = 18E$.

Another question associated with the decision metric in (59) is how to search for the indices μ_1, μ_2. It is obvious that we can choose from two different search orders: (i) We first fix μ_1 and search for the μ_2 that maximize the decision metric for a given μ_1. Then using only the μ_2's that locally maximize the decision metric, we search over μ_1 for the pair (μ_1,μ_2) that globally maximizes the decision metric. (ii) Alternatively, we first fix μ_2 and search for μ_1 that locally maximizes the decision metric for a given μ_2. The properly chosen search sequence can save on computation cost. From (59) we observe that only the first term is dependent of the index μ_2. Therefore, in order to search for the μ_2 that locally maximize the decision metric for a fixed μ_1, it is only necessary to calculate the first term of (59). This suggests that μ_2 can be estimated by choosing the shifted version of pilot sequence that produces the maximum correlation with the pilot-spaced observation $x_p(\mu_1)$. For a given μ_1, the decision metric for computing μ_2 is

$$\text{Metric} = \left(x_p(\mu_1)^{1/2}\right)^T \left(p(\mu_2)^{1/2}\right). \tag{68}$$

In conclusion, we propose the following implementation of the pilot sequence detection algorithm as described by the block diagram in Figure 3-9. A full frame of received symbols is the input. First the full frame observation forks into two branches. One branch is fed into the "Full Frame To Pilot-spaced Subframe" block. The output from this block are M pilot spaced subframes, $x_p(i)$, i = 0 to M-1. The other branch of full frame observation is first modulated by the magnitude of channel fading estimate, and then fed into a slicer which thresholds are given by (67). The output of the slicer is also converted into pilot spaced subframes, $d_{dom}(i)$, i = 0 to M-1. The pilot spaced subframes of observation forks again. One branch goes to the μ_1 selector which decision metric is given by (59). The other branch goes to the μ_2 selector, which decision metric is given by (68). The output from the μ_2 selector is a vector of M elements, with the i^{th} element indicating which shifted version of the pilot sequence has maximum correlation with $x_p(i)$. At last the μ_1 selector takes all the outputs from other blocks in the diagram and outputs the pair (μ_1,μ_2) that maximizes (59).

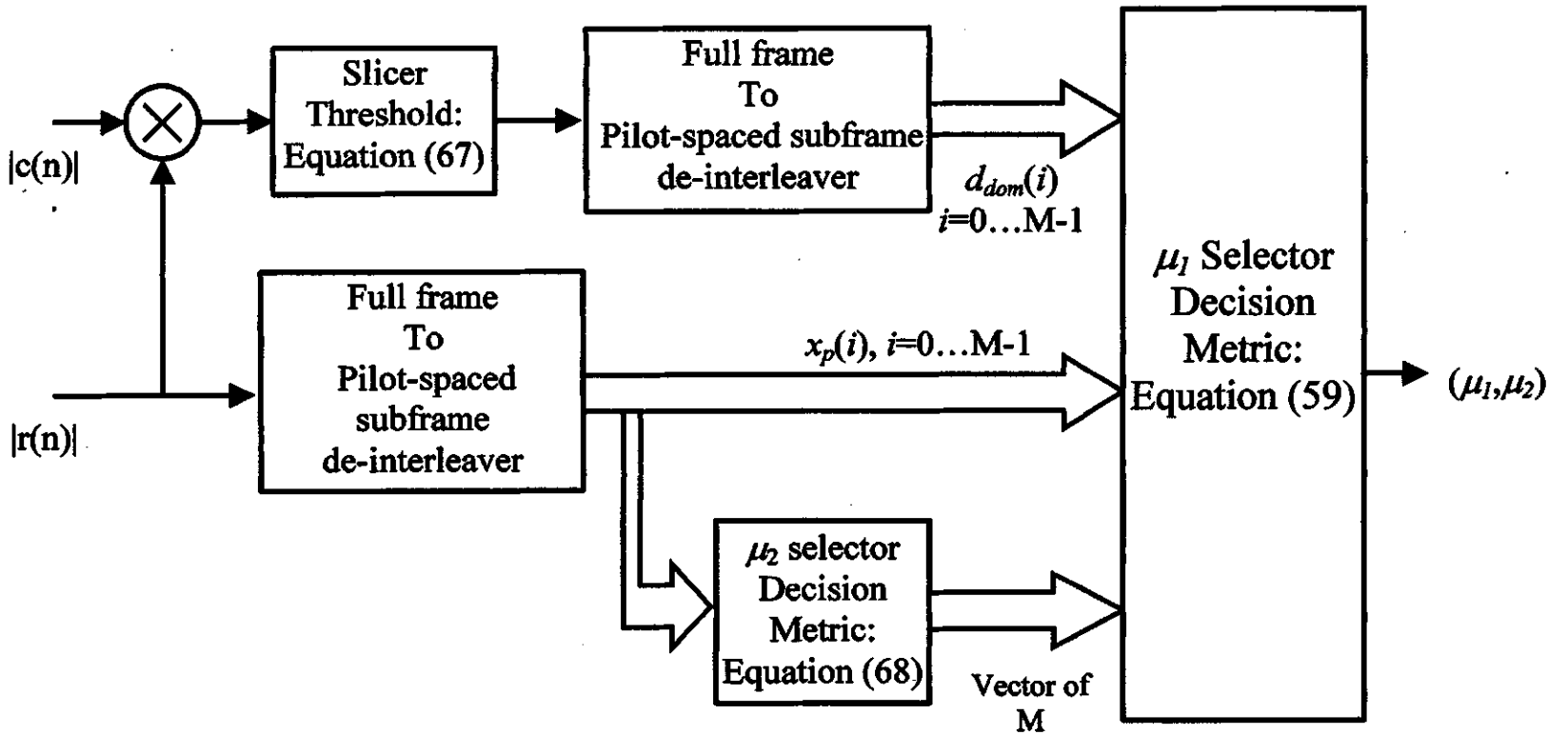


Figure 3-9 Block Diagram for Pilot Sequence Detection Scheme

3.4.3 Search Scheme

The decision made by the pilot sequence detector (Figure 3-9) is processed by a search/verification scheme. The purpose of a search/verification scheme is to guarantee that the decision from pilot sequence detector is not a false lock. A false lock can happen in two types of scenarios: (i) A pilot-spaced random data sequence has the exact magnitude profile as that of the pilot sequence; (ii) Channel fading and noise cause the pilot sequence detector to make wrong decision. In either case, the situation can be detected by a verification scheme.

The search-verification scheme is usually depicted by a Markov chain flow graph as show in Figure 3-10. The scheme has two modes. The process starts with the acquisition mode. During the acquisition mode, the pilot sequence detector makes the decision about the value of indices μ_1 and μ_2. Then the scheme enters verification mode. If the decision from acquisition mode passes the verification mode, frame synchronization is declared. Otherwise, the scheme goes back to acquisition mode.

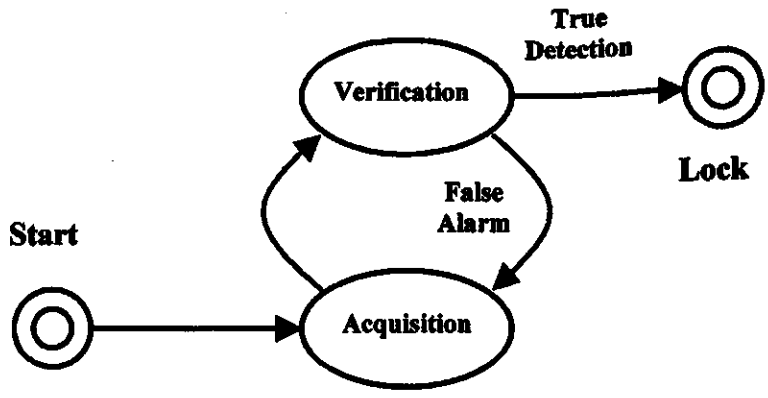


Figure 3-10 Search Scheme Flow Graph

Note that, in the literature, the acquisition mode is usually implemented as serial search, because serial search scheme requires much less hardware complexity than parallel search, which is described in Section 3.4.2. However, due to the special form of the pilot sequence detection metric in this scheme, i.e. index μ_2 can be calculated with much less effort than index μ_1, the advantage of serial search is significantly compromised. To illustrate this point, let us compare the complexity of decision metric (68) and (59) for a frame format of 110 symbols and a pilot sequence of 11 symbols. Assuming that the costs of various numerical operations are equal, we can represent the costs of (59) as four units and (68) one unit. To search through the two-dimensional uncertainty space spanned by μ_1 and μ_2, the serial search costs 4×110 units of operations, while the parallel search using the search sequencing in Section 3.4.2 costs 1×11×10+3×11=143 units. We therefore can implement the parallel search in a serial fashion, i.e. only one hardware unit for (59) is implemented and the results for μ_1 and μ_2

first stored in memory and compared after all values are available. Thus a true ML estimation can be achieved using same hardware complexity but only a fraction of the time that a serial search scheme needs. The parallel search is fast, and it is also more robust than the serial search because it is more immune to the time variation of channel fading and noise.

The verification mode can also be implemented in serial or parallel fashions. The verification in this study is implemented in parallel fashion. Using the decision from the acquisition mode, the correct indices μ_1 and μ_2 for the next full frame observation are calculated. If, according to the decision from the acquisition mode, the pilot spaced observation that corresponds to the pilot sequence maximizes the two decision metrics (59) and (68) among all pilot spaced observations, then the verification passes. A successful verification is that the decision from acquisition mode passes B out of A tests, where A, B are integers and A>B. The values of A and B can be chosen according the desired synchronization speed and the tolerable probability of false lock by numerical simulation.

4 Simulation and Results

In this chapter we first briefly describe the software we have chosen for the simulation. Then we describe the simulation model and discuss some of the key technicalities at a system level. Finally, there are simulation results and discussion.

4.1 MATLAB and Simulink

There are several commercial software tools for simulation in the market; however, the choice of simulation software is largely determined by its availability. The primary reason we have chosen Simulink (part of the MATLAB suite) as the simulation software is because of its wide acceptance in both industry and academic societies.

MATLAB is a suite of software packages developed by MathWorks. Although it is meant to be general mathematical software, its strength lies in numerical calculations (as opposed to symbolic calculation), especially for engineering oriented calculations. It is one of the most widely used simulation tools in industry. MATLAB consists of two major components: MATLAB compiler and Simulink. The MATLAB compiler serves as a numerical calculation environment and an interpreter for the MATLAB programming language. Simulink has a block diagram language interpreter and many libraries (or toolboxes) of building blocks for different areas of engineering applications.

With the MATLAB suite we have two options in building a simulation model: MATLAB compiler, or Simulink. To build a model in MATLAB language is much like implementing a model in C language. The system model is usually divided into static

parts according to the system functions. These parts of system are implemented as programs or procedures. The system is simulated by processing a large ensemble of random data through these static procedures. Simulink, on the other hand, provides a block diagram language. Modeling a system in this language is much like drawing block diagrams. Thus the dynamic features of a system as reflected in system block diagrams proper can be easily modeled by Simulink. The advantages and disadvantages associated with each option are listed in Table 1

Table 1 Advantages of MATLAB vs. Simulink

	MATLAB compiler	Simulink
Advantages	• Requires minimum calculation. Fast simulation. • The simulation process can be automated by a control script.	• Easy to build models. • Easy to model dynamic features.

Because dynamic features, such as system delay, are very important in simulating communications system, we have chosen Simulink. The choice certainly means more machine time and more demands on the power of computers. The latter factor has significant impact on the simulation results when there are limits on computer's memory. However, the decision also saved much human time on building and debugging the model that would have been required if MATLAB had been chosen. Since the cost of human time is high, and machines will become cheaper and more powerful, the decision is appropriate for future investigations.

4.2 System Level Description

The model has three modules on the top level. These modules and their

interconnection are illustrated in Figure 4-1.

Figure 4-1 System Block Diagram (Top Level)

The three big blocks in Figure 4-1 represent the top-level system modules. Ideally, these modules should be conceptual modules and their only purpose is to promote modular design. However, due to limited computer memory, the top-level modules are implemented as physically separated Simulink files. Thus the data flow between these modules must be accomplished using data files, i.e., the output from the first module is stored as a data file and the data file is then used as input to the second module. We use the two small blocks to represent the data files.

The first decision to be made when modeling a communications system is whether it is for a discrete time baseband channel or for a continuous time baseband channel. A discrete time baseband channel is a conceptual communication channel as observed at the input of square root raised cosine filter at the transmitter and the output of the symbol sampler at the receiver. A continuous-time baseband channel is a conceptual channel as observed at the input of a modulator and the output of a demodulator. The most important difference between the two conceptual channels is the amount of computation required for simulating them. To simulate a continuous-time system on digital computer means significant workload. Because our channel model in (7) describes a discrete-time channel, most simulation is conducted on a discrete-time

baseband channel. However, the continuous-time channel model is used when investigating the effect of carrier frequency offset, this is because that the square root raised cosine filter must be included in the model.

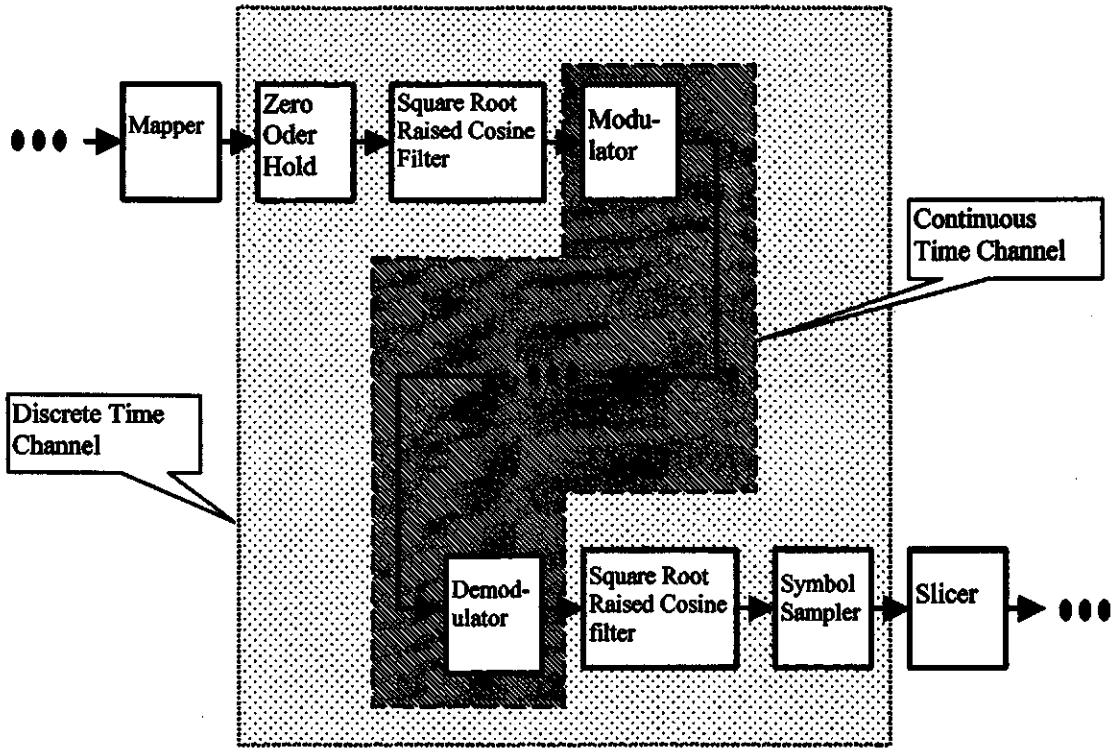


Figure 4-2 Illustration of Discrete Time Channel and Continuous Time Channel

When modeling multipath fading channel, we are also faced with more than one option. There are fundamentally two approaches to simulating the baseband effect of flat-fading: (i) implementing multipath to produce the fading effects (to be called the multipath approach) or (ii) generating a fading signal to model the flat-fading model in (7) (to be called the fading signal approach). Most variants of flat-fading simulation can be classified into these two categories or a combination of the two. Features of these two approaches are listed in Table 2.

Table 2 Merits of the Multipath Approach and the Fading Signal Approach

	Multipath Approach	Fading Signal Approach
Advantages	• Can specify delays • Less computation	• Can specify spectrum for the fading signal

A combination of both approaches can be used when both delays and spectra of fading signals must be exactly specified. However, the power of modeling comes at the price of computation. In our simulation, we experimented with both the basic approaches but not the combined approach due to limitation on computers.

To model a fading channel using the fading signal approach, we first generate a white Gaussian noise sequence, and then we shape the spectrum of the random sequence into that of (15). This spectrum shaping can be accomplished in frequency domain or in time domain. In the frequency domain method, white random sequence is first transformed to its spectrum by the FFT. The spectrum is then multiplied by a mask specified by (15). The resulting spectrum is then transformed back to time domain signal. Although the frequency domain method can easily accomplish almost any shape of spectrum, it only provides batch processing of data, which is difficult to integrate into Simulink diagrams, which features sequential processing of data. For this reason we have to choose the time domain method in which the fading signal is spectrum shaped using FIR or IIR filters.

As in hardware testing, a test bench is required to facilitate and automate the simulation. The test bench in our simulation provides the following functions:

- Generating the input sequence for the transmitter.
- Generating a reference signal to be compared with the output of frame

synchronizer.

- Processing the frame synchronizer output to generate the desired simulation statistics.

In order to generate a reference signal, we choose to implement a reference transmission system. The primary purpose of the reference transmission system is to accurately take account of all the effects of system delays. The approach could result in redundant computation, but it features easy modeling, because to build this model all we need to do is to copy and paste the existing model and to simplify it by some deletion. In the reference transmission system, there is a simplified version of the transmitter and receiver, and the channel is an ideal channel with delays that match those of the fading channel.

As mentioned at the beginning of this section, the whole system is physically divided and stored in three files. This division also applies to the test bench and is shown in Figure 4-3.

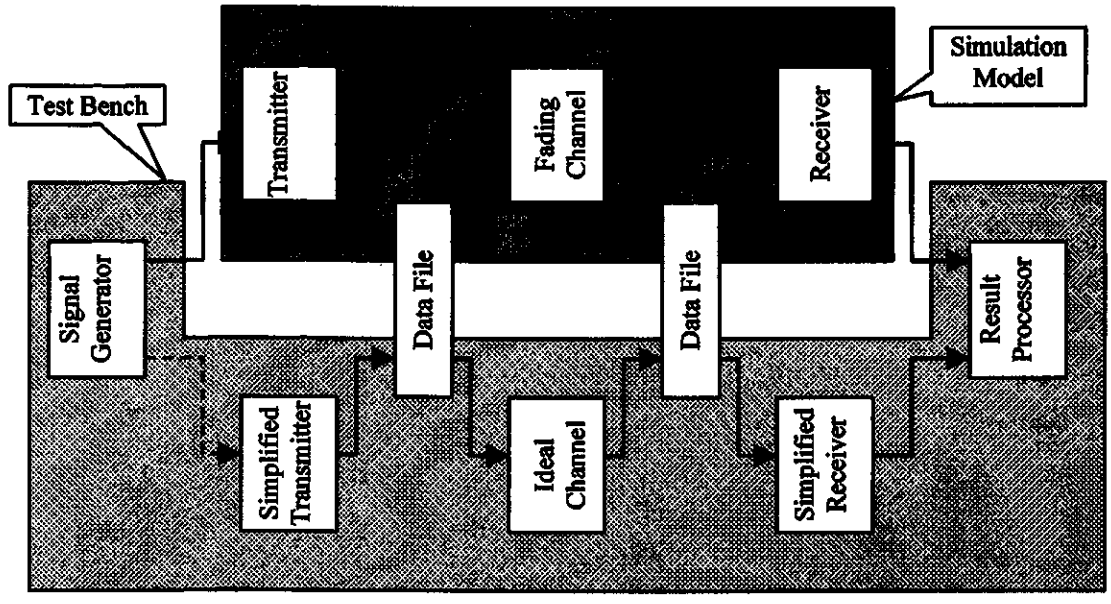


Figure 4-3 Test Model and Test Bench

4.3 Results

4.3.1 Description of the Simulated System

The communication system simulated by the test model used the following specification. The frame format was 110-symbol full frame and 10-symbol subframe. The pilot sequence is 11-symbol barker sequence, which, expressed in binary, is [1 1 1 0 0 0 1 0 0 1 0]. The fading rate was 1% of system symbol rate. The channel fading signal was zero mean complex Gaussian random sequences, with unit variance.

4.3.2 Probabilities of Detection for the Acquisition Mode

The probabilities of detection vs. channel SNR are plotted in Figure 4-4. The measurements were obtained using the following method. First, the frame synchronizer in the receiver and the simplified receiver of Figure 4-3 keep working in search mode. That is, for each full frame observation, the frame synchronizer always generates an

estimate of the index of pilot symbol P_0, and never enters the verification mode. Second, the signal generator continuously generates pseudo-random sequence until 1,000 full frame observations is received at the receiver. Third, in Figure 4-3, the transmitter of the simulation model and the simplified transmitter of the test bench work in a synchronized fashion, so that they both insert the same pilot symbol in exactly the same location of the symbol stream.

While the communication system in the Simulation Model works in a condition that mimics a real mobile communication system, the communication system contained in the test bench works in an ideal condition. First, the Simplified Transmitter only transmits a stream of value zero and one. Value one indicates a pilot symbol P_0 is transmitted; value zero means either a data symbol or a pilot symbol other than P_0. The timing of zero-one symbol stream is synchronized to the pseudo-random symbol stream fed into the transmitter of the simulation model. This synchronization is indicated by the dotted line from the signal generator to the simplified transmitter in the test bench. Second, the ideal channel is just a delay line that delays the transmitted signal by the same amount of time as the fading channel in the simulation model does; it does not add noise neither fading effects. Thus, the simplified receiver can accurately detect the location of pilot symbol P_0.

The result processor compares the output of the receiver to that of the simplified receiver, and counts the number of P_0 symbols that are correctly detected during 1,000 full observations. The ratio of the number of correctly detected frames to the number of total simulated frames is taken as the probability of detection.

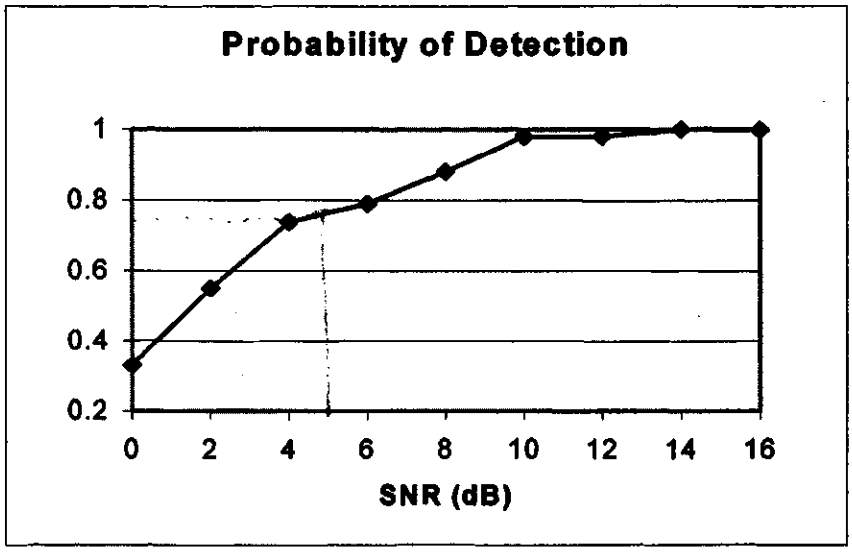


Figure 4-4 Probabilities of Detection with Fading Rate of 1%

4.3.3 Mean Time to Synchronization

The mean time to synchronization, T_{SYNC}, can be calculated using the probability of detection and the expression is derived as follows. Let N be the length (in number of symbols) of the pilot sequence, M be the length of a subframe. The length of a full frame observation is therefore M×N symbols. The mean time to synchronization is defined as

$$T_{SYNC} = \sum_{k=1}^{\infty} T(k)p(k). \tag{69}$$

In the above equation, $T(k)$ represents the time that the frame synchronizer uses to achieve true synchronization at the k^{th} time it transits from acquisition mode to verification mode. A cycle of acquisition mode to verification mode and back to acquisition mode is illustrated in Figure 3-10. The probability that the frame synchronizer comes out of acquisition mode with the correct estimate of μ_1 and μ_2 at the k^{th} round is

$$p(k) = P_D(1-P_D)^{k-1}. \tag{70}$$

where P_D is the probability of correct detection of pilot symbol P_0. The time $T(k)$ consists of two parts: the time consumed in the acquisition mode, $T_{ACQ}(k)$, and the time consumed in the verification mode, $T_{VER}(k)$. We can express $T(k)$ in terms of $T_{ACQ}(k)$ and $T_{VER}(k)$, and reorganize the equation (69) as

$$\begin{aligned} T_{SYNC} &= \sum_{k=1}^{\infty} T_{ACQ}(k)p(k) + \sum_{k=2}^{\infty} T_{VER}(k)p(k) \\ &= T_{ACQ} + T_{VER}. \end{aligned} \tag{71}$$

Let us first consider the part of the mean time to synchronization that used by the acquisition mode, T_{ACQ}. A ML estimation of the indices μ_1 and μ_2 consumes one full frame observation. If we ignore the time it takes to evaluate the metric (59) on the last pilot spaced observation, then the time to finish one search in acquisition mode takes MN symbol times. The mean time in acquisition mode is k times one full frameobservation weighted by $p(k)$,

$$\begin{aligned} T_{ACQ} &= \sum_{k=1}^{\infty} MNkp(k) \\ &= MN\sum_{k=1}^{\infty} kP_D(1-P_D)^{k-1}. \end{aligned} \tag{72}$$

Since, by definition, $p(k)$ is a probability density function, then

$$\sum_{k=1}^{\infty} p(k) = \sum_{k=1}^{\infty} P_D(1-P_D)^{k-1} = 1. \tag{73}$$

Differentiating both sides of (73) with respect to P_D and simplifying gives [13]

$$\sum_{k=1}^{\infty} kP_D(1-P_D)^{k-1} = \frac{1}{P_D}. \tag{74}$$

Substituting (74) into (72) gives

$$T_{ACQ} = \frac{MN}{P_D}. \tag{75}$$

The time used in the verification mode to find a false synchronization can vary from A×(M×N) symbol times to B×(M×N) symbol times. To simplify the calculation of T_{VER}, we use the worst case of detecting an incorrect synchronization in the verification mode, i.e. it always takes B×(M×N) symbol times to discover a false synchronization. This approximation produces a tight upper bound when A is close to B (A, B are defined in Section 3.4.3). Then the mean time spent in the verification mode is

$$\begin{aligned} T_{VER} &= \sum_{k=1}^{\infty} BMN(k-1)p(k) \\ &= BMN\left[\sum_{k=1}^{\infty} kp(k) - \sum_{k=1}^{\infty} p(k)\right]. \end{aligned} \tag{76}$$

Substituting (69), (73), and (74) into (76) and simplifying gives

$$\begin{aligned} T_{VER} &= BMN\left[\sum_{k=1}^{\infty} kp(k) - \sum_{k=1}^{\infty} p(k)\right] \\ &= BMN\left(\frac{1}{P_D} - 1\right). \end{aligned} \tag{77}$$

Combining (75) and (77) gives

$$\begin{aligned} T_{SYNC} &= T_{ACQ} + T_{VER} \\ &= MN\left(\frac{B+1}{P_D} - B\right). \end{aligned} \tag{78}$$

For the simulation described in Section 4.3.1, M=10, N=11, (M, N are defined in Section 3.4.2) and B=3 (B is defined in Section 3.4.3). The value of B is chosen to be consistent with that of [4]. Using the probabilities of detection in Figure 4-4, the mean times to synchronization vs. different SNR are calculated and given in Figure 4-5.

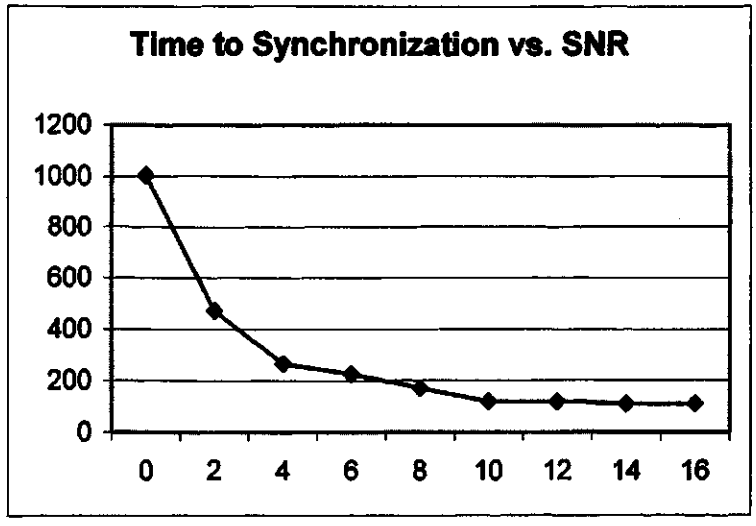


Figure 4-5 Mean Time To Synchronization vs. SNR

4.3.4 Performance Degradation in Presence of Carrier Frequency Error

The effect of local carrier frequency error is simulated by a continuous time channel model as defined in Section 4.2. This is required since the discrete time channel does not include the receiving square root raised-cosine filter. Performance degradation is a result of the mismatch between shifted signal frequency and the band limits of the receiving square root raised-cosine filter. Because the continuous time channel model has a much higher sampling rate than a discrete time channel model, this means we have to simplify the fading signal generation mechanism to avoid excessive simulation time. For this reason, we implemented the fading channel by the multipath approach as described in Section 4.2. This choice means that the fading signal generated by this approach falls into the category of regular random process, and thus ensemble average cannot be obtained from time average. However, the choice is justified because our concern in this simulation is the effect of carrier frequency error, not channel fading. Thus we can choose to fix the fading factor while validity of the result is not

compromised.

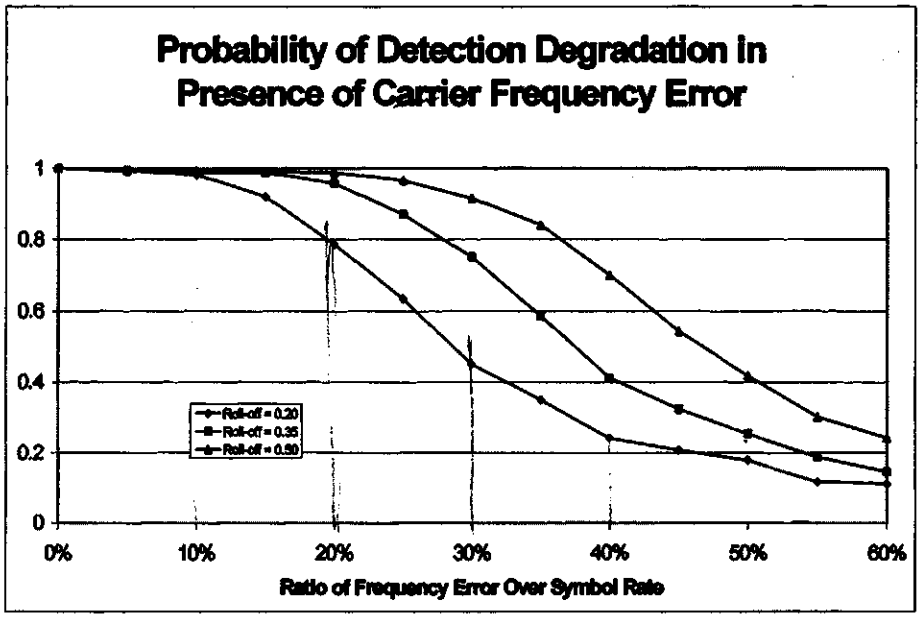


Figure 4-6 Probability of Detection Degradation in Presence of Carrier Frequency Error (Fading rate = 1%, SNR = 10dB)

In Figure 4-6, the performance of frame synchronizer is normalized by dividing the probability of detection under different carrier frequency errors to the probability of detection under zero carrier frequency error. From Figure 4-6 we conclude that performance degradation depends on the roll-off factor of the square root raised cosine filter. Large roll-off factor allows operation with larger local carrier frequency error.

4.4 Comparison with Existing Scheme

By the time the research work presented in the thesis was finished, the only published PSAM frame synchronization scheme was that of Gansman [4]. By comparing the performance of the scheme in this thesis and the one in [4], we can identify the following two points.

First, the scheme in this thesis is more immune to carrier frequency error. Despite

the fact that Gansman provides no detail data on the degradation of performance in presence of carrier frequency error, it is clear the performance will degrade significantly when the frequency offset is larger than 5%. For the scheme in this thesis, Figure 4-6 indicates that the performance has 1% degradation for a frequency offset as high as 15%. This is due to the non-coherent feature of the scheme.

Second, the performance of the scheme of this thesis degrades more quickly than the Gansman's scheme as the channel SNR falls. This is because our scheme is based on the high SNR approximation rule, while Gansman's scheme is based on low SNR approximation rule. Since a communication system using 16-QAM requires SNR greater than 10 dB, this degradation is not considered to be a problem.

5 Conclusion

In this chapter, the key points of the design process and the simulation result are first summarized. Then some areas of possible improvement for the scheme are suggested.

5.1 Review of Design Philosophy and Simulation Results

In this work, the design philosophy is to take advantage of the fact that the random variables in channel model (12) are independent. Because of the independency, it is possible to decompose the task of a frame synchronizer into isolated stages, with each stage dealing with fewer random variables. It is not just to factor the p.d.f. into four factors, rather it is to hide one or more random variables, so that a subspace of less dimensionality is extracted from the 4-dimensional space.

The following steps were taken to reduce dimensionality. As a first step, by taking magnitude-squared of the received symbols, we effectively reduced 4 dimensions to 3 dimensions by removing $\Delta\omega$. The problem is then decomposed into two problems of 2-dimensional spaces: channel fading estimation and frame synchronization in noisy channel. In channel fading estimation, the fading signal is considered the desired signal; transmitted symbols (pilot and data symbols) and channel noise are considered together as noises. Thus at this stage, the probability space reduces to a 2-dimensional one. After the fading signal estimation, the problem of fading channel is reduced to a problem of additive noisy channel, and we are left with the second problem of detecting pilot

sequence among random data symbols over a noisy channel. At this stage, channel noise and data symbols are considered unwanted factors, and the probability space is 2-dimensional. Because the two problems can fit very well in two well-developed theoretical frameworks, linear mean squared error estimation and maximum likelihood estimation, the expressions in our design can be more clearly interpreted from theoretical point of view.

The performance of the scheme is measured by means of simulation. From the simulation results, synchronization was obtained with good probability at high to moderate SNR. The mean time to synchronization of this scheme is similar to published work [4].

5.2 Suggestion for Future Work

In this section, we suggest three major aspects concerning future research on this scheme. The first aspect is about some implementation details. In deriving the mathematical model for the fading channel, we have imposed the assumption that the variance of fading signal must be unit valued. This assumption suggests some automatic gain control circuit at a stage prior to frame synchronization. However, typical automatic gain control circuit targets the power of received signal rather than the power of fading signal. This means a new automatic gain control scheme is required. Because the fading signal estimator, (39), contains the term σ_s^2/σ_n^2, for the fading estimator to work properly, knowledge about the SNR is required. Although there are various algorithms dealing with this problem, they usually are too complex to implement in frame synchronizer. Since this knowledge of SNR is not crucial to receiving data, it is possible to compromise on the accuracy for low complexity and to optimize the design trade-off.

The second important aspect that is worth serious investigation is the alternative approximation rule, low SNR approximation rule. As the simulation results suggest, high SNR approximation rule has poor performance for a fading signal with power spectrum described by (15). We have already indicated in Section 4.3.1 that this poor performance is due to the scheme's incapability of handling long deep fading periods. However, major difficulty in deploying low SNR approximation rule is that the resultant expressions are not similar to Gaussian distribution, which prevents effective simplification in the following stages. One promising approach could be that we first build an approximate expression from the symmetry to the results we have derived using high SNR approximation rule, and optimize the parameters for this approximate expression through numerical simulation.

The effect of imperfect symbol timing on the probability of detection needs further investigation by simulation and mathematical analysis. In this thesis, symbol timing of the symbol sampler is assumed to be perfect. However, in real implementation of a receiver, symbol timing error always exist, and can become worse under low SNR.

A final important analytical work missing from this study is an expression, which can relate probability of correct detection to the MSE of channel fading estimator. As we have indicated, the mean time to synchronization is mainly composed of system delays. Among these delays, that of the fading estimator plays a key role; thus minimization of the length of fading estimator is the most important factor in improving the speed of frame synchronizer. Since the length of fading estimator determines the MSE of the fading estimate, and the MSE affects the probability of correct detection, an equation that expresses the probability of correct detection in terms of the length of the fading estimation filter is very helpful in calculating the optimal filter length. The optimal filter

length, of course, can be achieved by exhaust numerical simulation. However, the approach is brute force. Additional benefit of the analytical approach is that the fading estimation filter can be designed to adapt its length according to the SNR and fading rate of a fading channel. An adaptive-length fading estimator is obviously a very desirable feature for applications that require the highest possible speed.

References

[1] J. K. Cavers, On the validity of the slow and moderate fading models for matched filter detection of Rayleigh fading signals, *Can. J. Elect. Comput. Eng.*, vol. 17(10), pp.183-189, 1992.

[2] J. K. Cavers, A comparison of pilot tone and pilot symbol techniques for digital mobile communication, in *Proc. Global Telecommunications Conference*, vol. 2 pp.915 –921, 1992.

[3] Andrew Bateman and Warren Yates, "*Digital Signal Processing Design*," W. H. Freeman, 1989.

[4] J. A. Gansman, M. P. Fitz, And J. V. Krogmeier, Optimum and Suboptimum frame Synchronization for Pilot-symbol-Assisted Modulation, *IEEE Trans. Commun.*, vol. COM-45, pp.1327-1337, 1997.

[5] W. C. Jakes, "*Microwave Mobile Communications,*" IEEE Press, 1974.

[6] John G. Proakis, "*Digital Communications*," 3rd ed. McGraw-Hill, 1995.

[7] A. Aghamohammadi and H. Meyr, A new method for phase synchronization and automatic gain control for linearly modulated signals in frequency flat fading, *IEEE Trans, Commun.*, vol. 38, pp.25-29, 1991.

[8] D. E. Dodds, K. Takaya and Q. Zhang, Frame synchronization for pilot symbol assisted modulation, in *Proc. 1999 IEEE Canadian Conference on Electrical and Computer Engineering*, vol.1 pp.52 – 58, 1999.

[9] J. S. Bendat and P. G. Allan, "*Engineering Applications of Correlation and*

Spectral Analysis," John Wiley & Sons, Inc., 1993.

[10] Milton Abramowitz and Irene A Stegun, "*Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables,*" National Bureau of Standards, June 1964.

[11] J. L. Massey, Optimum Frame synchronization, *IEEE Trans Commun.*, vol. COM-20(2), pp.115-119, 1971.

[12] G. L. Lui and H. H.Tan, Frame Synchronization for Gaussian Channels, *IEEE Trans. Commun.*, vol. COM-35, pp. 818-829, Aug. 1987.

[13] Simon, Marvin K., Omura, Jim K., Scholtz, Robert A. and Levitt, Barrry K., "*Spread Spectrum Communications,"* Volume III. Computer Science Press, 1985.

[14] Athanasios Papoulis, "*Probability, Random Variables, and Stochastic Processes,"* McGraw-Hill, 1991.

Apendix A Power Spectra of c(n) and $|c(n)|^2$

In [5], Jakes gives the power spectrum for the pre-envelope, $c_+(t)$, of a bandpass fading signal, which is observed by a vertical whip antenna with antenna gain pattern $G(\alpha) = 1.5$ as

$$S_{c_+}(f) = \frac{3b}{\omega_m \sqrt{1 - \left(\frac{f - f_c}{f_m}\right)^2}}. \tag{79}$$

where b is the average power which would be received by an isotropic antenna assuming power gain pattern is uniform over $[0, 2\pi)$, f_c is the carrier frequency, f_m is the maximum Doppler frequency shift, and $\omega_m = 2\pi f_m$. It is worth to mention that equation (79) is only valid for the positive frequency range $[fc\text{-}fm,\ fc+fm]$, thus the square root in the denominator is always nonnegative. The value of $Sc_+(f)$ becomes infinite at $fc - fm$ and $fc + fm$, this is not an accurate model of the reality and it is difficult to conduct numerical evaluation of the equation. Nevertheless, this is still the most cited model in the literature as it is elegant from mathematical perspective, and also gives reasonable approximation to simulation results.

The above power spectrum is for a bandpass fading signal. But what is needed in this study is the baseband power spectrum of the discrete time baseband fading signal, which appears in the channel model in (7). The bandpass fading signal is related to its baseband complex envelope by

$$c_+(t) = c(t)e^{j2\pi f_c t}, \tag{80}$$

where $c_+(t)$ is the complex-valued pre-envelope of the bandpass fading signal, $c(t)$ is the continuous time baseband complex envelop, f_c is the carrier frequency. The baseband continuous time complex signal is related to the discrete time baseband complex signal by

$$c(t) = \sum_{k=-\infty}^{\infty} c(kT)\,\mathrm{sinc}\left(\frac{t-kT}{T}\right), \tag{81}$$

where $c(kT)$ is discrete time sample of the complex-valued baseband fading signal, $c(t)$; T is the symbol period, and will take unit value in much of this appendix for simplicity. The function sinc(x) is defined as $\frac{\sin(\pi x)}{\pi x}$. Substituting (81) into (80) gives the relationship between $c_+(t)$ and c(k) as

$$c_+(t) = \sum_{k=-\infty}^{\infty} c(k)\,\mathrm{sinc}(t-k)e^{j2\pi f_c t}. \tag{82}$$

In this appendix, the power spectrum of $c(k)$ is derived from that of $c_+(t)$. The relationship between the power spectra of $c_+(t)$ and $c(k)$ can be derived from the relationship between the corresponding correlation functions. The correlation function of $c_+(t)$ can be expressed in terms of correlation function of $c(k)$ as

$$
\begin{aligned}
R_{c+}(\tau) &= \frac{1}{2}E\left[c_+(t)c_+^*(t-\tau)\right] \\
&= \frac{1}{2}E\left[c(t)e^{j2\pi f_c t}c^*(t-\tau)e^{-j2\pi f_c(t-\tau)}\right] \\
&= \frac{1}{2}E\left[c(t)c^*(t-\tau)\right]e^{j2\pi f_c\tau} \\
&= \frac{1}{2}E\left\{\left[\sum_{m=-\infty}^{\infty}c(m)\mathrm{sinc}(t-m)\right]\left[\sum_{n=-\infty}^{\infty}c(n)\mathrm{sinc}(t-\tau-n)\right]\right\}e^{j2\pi f_c\tau} \\
&= \frac{1}{2}E\left\{\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}c(m)c^*(n)\mathrm{sinc}(t-m)\mathrm{sinc}(t-\tau-n)\right\}e^{j2\pi f_c\tau} \\
&= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{1}{2}E\left[c(m)c^*(n)\right]\mathrm{sinc}(t-m)\mathrm{sinc}(t-\tau-n) \\
&= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}R_c(m-n)\mathrm{sinc}(t-m)\mathrm{sinc}(t-\tau-n)e^{j2\pi f_c\tau}
\end{aligned}
\tag{83}
$$

The last line in equation (83) shows that the correlation function $R_{c+}(\tau)$ is dependent on time variable t. This suggests that the random process denoted by $c_+(t)$ is a non-stationary process. Of course, the fading signal is stationary. This is an artifact caused by the use of the sinc(t) function to bridge the gap between the discrete time baseband fading signals to the continuous time baseband fading signal, i.e. the use of equation (81). Since for discrete time function the correlation function is also discrete time, so t and τ can only take integer values. The artifact vanishes when only discrete time value of t and τ are considered. Note that, the function sinc(t) has value 1 when $t = 0$, and value 0 when t is a non-zero integer. The only case that the product $\mathrm{sinc}(t-m)\mathrm{sinc}(t-\tau-n)$ will have nonzero value is when $\tau = m - n$ and $t = m$, and the nonzero value is 1. The product produces value zero in all other cases. So when only consider discrete time value kT ($T = 1$) for the variable τ, equation (83) can be reduced to

$$
R_{c+}(k) = R_c(k)e^{j2\pi f_c k}. \tag{84}
$$

Note that the dependency on variable t of the last line in equation (83) is removed

because the summation $\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}$ ensures that there is always one and only one item satisfying the requirement $t = m$, $\tau = m - n$. Taking discrete time Fourier transform on both sides of(84), we obtain the power spectrum relationship

$$S_c(f) = S_{c+}(f + f_c) \tag{85}$$

From (79) and (85), we obtain the equation for $S_c(f)$ as

$$S_c(f) = \frac{3b}{\omega_m \sqrt{1-\left(\frac{f}{f_m}\right)^2}}. \tag{86}$$

Since $S_{c+}(f)$ is defined on $[f_c-f_m, f_c+f_m]$, f in (86) is defined on $[-f_m, +f_m]$.

By means of gain control, the variance of a fading signal can be normalized to unit value. Thus we can determine the value of b for the normalized fading signal through the following equation

$$\begin{aligned} Var(c) = E\left\{|c(k)|^2\right\} = R_c(0) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_c(f)df \\ &= \frac{3b}{2\pi}\int_{-f_m}^{f_m} \frac{df}{f_m\sqrt{1-\left(\frac{f}{f_m}\right)^2}} \\ &= \frac{1.5b}{\pi}\int_{-1}^{1}\frac{dx}{\sqrt{1-x^2}} \\ &= \frac{1.5b}{\pi}\sin^{-1}(x)\Big|_{-1}^{1} \\ &= 1.5b \end{aligned} \tag{87}$$

In deriving (87), the first line integrates the power spectrum over the full range of f. In the second line, the integration range is refined to $[-f_m, +f_m]$, as $S_c(f)$ is defined on it. The third line results from substitution of $\frac{f}{f_m}$ with x. Using the integral formula

$\int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}} = \sin^{-1}(x)\Big|_{-1}^{1}$ gives the fourth line. Since the variance of the fading signal has been adapted to unit value by means of a gain control, $Var(c)=1$ and b = 2/3. Substitution of the value in (86) gives the power spectrum of normalized $c(k)$ as

$$S_c(f) = \frac{1}{\pi f_m \sqrt{1-\left(\frac{f}{f_m}\right)^2}}. \tag{88}$$

From (88) we can develop the power spectrum of the envelope squared of a baseband fading signal, i.e. the power spectrum of $|c(k)|^2$. We first derive the power spectra for the real and imaginary components of the complex fading signal. The complex envelope of fading signal can be expressed in its real and imaginary components as

$$c(n) = c_I(n) + jc_Q(n), \tag{89}$$

where $c_I(n)$ is the in-phase component, $c_Q(n)$ is the quadrature component. The autocorrelation of $c(n)$ can in turn be expressed as

$$\begin{aligned} R_c(k) &= \frac{1}{2}E\left\{\left[c_I(n)+jc_Q(n)\right]\left[c_I(n-k)-jc_Q(n-k)\right]\right\} \\ &= \frac{1}{2}\left\{R_{c_I}(k)+R_{c_Q}(k)+j\left[R_{c_Ic_Q}(k)-R_{c_Qc_I}(k)\right]\right\} \\ &= R_{c_I}(k), \end{aligned} \tag{90}$$

where $Rc_I(k)$, $Rc_Q(k)$ are auto-correlation functions of $c_I(k)$ and $c_Q(k)$ respectively; $Rc_Ic_Q(k)$ and $Rc_Qc_I(k)$ are cross-correlation. The last line of (90) results from the assumption that $c_I(k)$ and $c_Q(k)$ are independent, identically distributed, zero mean Gaussian random processes. From (90), the power spectra of real component is

$$S_{c_I}(f) = S_c(f) = \frac{1}{\pi f_m \sqrt{1-\left(\frac{f}{f_m}\right)^2}}. \tag{91}$$

The inverse Fourier transform of (91) is [4]

$$R_{c_I}(k) = J_0(2\pi k f_m), \tag{92}$$

where $J_0(2\pi k f_m)$ is 0^{th}-order Bessel function.

Next, we derive the power spectrum for $|c(n)|^2$ and express it in terms of $S_c(f)$. But first we need to develop equation for the auto-correlation function of $|c(n)|^2$.

$$\begin{aligned} R_{c^2}(k) &= E\left\{|c(n)|^2 |c(n-k)|^2\right\} \\ &= E\left\{\left[c_I^2(n)+c_Q^2(n)\right]\left[c_I^2(n-k)+c_Q^2(n-k)\right]\right\} \\ &= 2E\left\{c_I^2(n)c_I^2(n-k)\right\} + E\left\{c_I^2(n)c_Q^2(n-k)\right\} + E\left\{c_Q^2(n)c_I^2(n-k)\right\} \\ &= 2E\left\{c_I^2(n)c_I^2(n-k)\right\} + 2E\left\{c_I^2(n)\right\}E\left\{c_Q^2(n)\right\} \\ &= 2R_{c_I^2}(k) + 2\left(\frac{1}{2}\sigma_c^2\right)^2 \\ &= 2R_{c_I^2}(k) + \frac{1}{2}. \end{aligned} \tag{93}$$

In the above equation we used the assumption that $\sigma_c^2 = 1$, and in-phase and quadrature components $c(k)$ are identical, independent real random processes.

It is known that, for zero mean Gaussian random signals, the auto-correlation of a square law device's output is related to the auto-correlation function of its input by the following equation [14]

$$R_{c_I^2}(k) = R_{c_I}^2(0) + 2R_{c_I}^2(k). \tag{94}$$

Substitution of (94) into (93) gives

$$R_{c^2}(k) = 4R_{c_I}^2(k) + \frac{1}{2}. \tag{95}$$

Evaluating $R_{c^2}(k)$ at $k = 0$, we obtain $R_{c^2}(0) = 2$. This value is used in Appendix B when deriving expression for auto-correlation function of $|r(t)|^2$.

Taking discrete time Fourier transform of (95), we obtain the power spectrum of $|c(n)|^2$

$$\begin{aligned} S_{c^2}(f) &= 4S_{c_I}(f) * S_{c_I}(f) + \delta(f) \\ &= 4S_c(f) * S_c(f) + \delta(f). \end{aligned} \tag{96}$$

The symbol * in (96) represents convolution operation in frequency domain. From (96) we can get some insight about the power spectrum of $|c(n)|^2$. First the power spectrum is made of two components, covariance spectrum and a spectral line at $f = 0$. Covariance spectrum represents the variation in $|c(n)|^2$. The covariance spectrum is a self-convolution of the power spectrum for $c(n)$. Due to the convolution operation, the bandwidth of covariance spectrum $S_c(f)*S_c(f)$ is double the bandwidth of $S_c(f)$. Physical explanation of the bandwidth doubling is that the highest frequency components in $|c(n)|^2$ is two times of the maximum Doppler shift, and this is due to the nonlinear effect of the square law device. The spectral line at DC frequency indicates the mean value of $|c(n)|^2$. The magnitude of the DC component in $|c(n)|^2$ is unit value because the variance of the fading signal has been normalized to one.

Unfortunately, it is impossible to give explicit expression for $|c(n)|^2$ based on the fading model of Jakes. This is because the correlation function of $c(n)$, which is the inverse Fourier transform of $S_c(f)$ in (91), is 0^{th}-order Bessel function $J_0(2\pi n f_m)$, and its square does not satisfy the following condition for convergence of Fourier transform

$$\sum_{n=-\infty}^{+\infty} J_0^2(2\pi n f_m) < \infty. \tag{97}$$

Fortunately, lack of a closed form for the power spectrum of $|c(n)|^2$ does not necessarily

render the notion of power spectrum useless in design practice. Actually, our design can benefit from the correlation of $|c(n)|^2$ as described in Section 3.3.

Apendix B Power spectrum of $|r(k)|^2$

This appendix gives the derivation of the expression for the power spectrum of $|r(k)|^2$. The assumptions used in this derivation are summarized as follows:

- $c(k)$, $s(k)$ and $n(k)$ are stationary complex random processes. They can be represented in the forms $c_I(k)+jc_Q(k)$, $s_I(k)+js_Q(k)$, and $n_I(k)+jn_Q(k)$. The in-phase and quadrature components of a baseband complex random process are independent, identically distributed.
- $c(k)$, $s(k)$ and $n(k)$ have variances 1, σ_s^2 and σ_n^2 respectively; $c(k)$, $s(k)$ and $n(k)$ are zero mean processes. This means that each component random process of the complex processes is zero mean and $E(c_I^2) = E(c_Q^2) = 0.5$, $E(s_I^2) = E(s_Q^2) = 0.5\sigma_s^2$, and $E(n_I^2) = E(n_Q^2) = 0.5\sigma_n^2$.
- $s(k)$ and $n(k)$ are independent identically distributed random variables for different values of k. This means their power spectra are constants over the frequency range of baseband.

Besides the above assumptions, the following symbols are defined to simplify notation in our derivation

$$
\begin{aligned}
&R_{r^2}(k) = E\left\{|r(m)|^2 |r(m-k)|^2\right\},\\
&R_{c^2}(k) = E\left\{|c(m)|^2 |c(m-k)|^2\right\},\\
&R_{s^2}(k) = E\left\{|s(m)|^2 |s(m-k)|^2\right\},\\
&R_{n^2}(k) = E\left\{|n(m)|^2 |n(m-k)|^2\right\},\\
&\sigma_c^2 = E(|c(k)|^2), \sigma_s^2 = E(|s(k)|^2), \sigma_n^2 = E(|n(k)|^2).
\end{aligned}
$$

We start with the expression for $|r(k)|^2$. From (7), $|r(k)|^2$ is expanded as

$$|r(k)|^2 = |c(k)|^2|s(k)|^2 + |n(k)|^2 + 2\mathrm{Re}\left[c(k)s(k)n^*(k)\right]. \tag{98}$$

The auto-correlation of $|r(k)|^2$ expands to

$$\begin{aligned}
R_{r^2}(k) &= E\left\{|r(m)|^2|r(m-k)|^2\right\} \\
&= E\left\{\left\{|c(m)|^2|s(m)|^2 + |n(m)|^2 + 2\mathrm{Re}\left[c(m)s(m)n^*(m)\right]\right\}\right. \\
&\times\left\{|c(m-k)|^2|s(m-k)|^2 + |n(m-k)|^2\right. \\
&\left.\left.+ 2\mathrm{Re}\left[c(m-k)s(m-k)n^*(m-k)\right]\right\}\right\} \\
&= E\left\{|c(m)|^2|c(m-k)|^2|s(m)|^2|s(m-k)|^2\right\} + \left\{|n(m)|^2|n(m-k)|^2\right\} \\
&+ \left\{|c(m)|^2|s(m)|^2|n(m-k)|^2\right\} + \left\{|c(m-k)|^2|s(m-k)|^2|n(m)|^2\right\} \\
&+ 4\mathrm{Re}\left[c(m)s(m)n^*(m)\right]\mathrm{Re}\left[c(m-k)s(m-k)n^*(m-k)\right] \\
&+ 2\mathrm{Re}\left[c(m)s(m)n^*(m)\right]\left\{|c(m-k)|^2|s(m-k)|^2 + |n(m-k)|^2\right\} \\
&+ 2\mathrm{Re}\left[c(m-k)s(m-k)n^*(m-k)\right]\left\{|c(m)|^2|s(m)|^2 + |n(m)|^2\right\}.
\end{aligned} \tag{99}$$

Terms in the last two lines on the right of the above equation all evaluate to zero after being taken ensemble average, because the random processes $s(m)$ and $n(m)$ are zero mean. The third line from the bottom is expanded using the real and imaginary parts of random signals as

$$\begin{aligned}
&E\left\{4\mathrm{Re}\left[c(m)s(m)n^*(m)\right]\mathrm{Re}\left[c(m-k)s(m-k)n^*(m-k)\right]\right\} \\
&= 4E\left\{\left\{c_I(m)s_I(m)n_I(m) - c_Q(m)s_Q(m)n_I(m) + c_Q(m)s_I(m)n_Q(m) + c_I(m)s_Q(m)n_Q(m)\right\}\right. \\
&\times\left\{c_I(m-k)s_I(m-k)n_I(m-k) - c_Q(m-k)s_Q(m-k)n_I(m-k)\right. \\
&\left.\left.+ c_Q(m-k)s_I(m-k)n_Q(m-k) + c_I(m-k)s_Q(m-k)n_Q(m-k)\right\}\right\} \\
&= 4E\left\{c_I(m)s_I(m)n_I(m)c_I(m-k)s_I(m-k)n_I(m-k)\right. \\
&+ c_Q(m)s_Q(m)n_I(m)c_Q(m-k)s_Q(m-k)n_I(m-k) \\
&+ c_Q(m)s_I(m)n_Q(m)c_Q(m-k)s_I(m-k)n_Q(m-k) \\
&\left.+ c_I(m)s_Q(m)n_Q(m)c_I(m-k)s_Q(m-k)n_Q(m-k)\right\} \\
&= 4\left\{R_{c_I}(k)R_{s_I}(k)R_{n_I}(k) + R_{c_Q}(k)R_{s_Q}(k)R_{n_I}(k)\right. \\
&\left.+ R_{c_Q}(k)R_{s_I}(k)R_{n_Q}(k) + R_{c_I}(k)R_{s_Q}(k)R_{n_Q}(k)\right\} \\
&= 16R_{c_I}(k)R_{s_I}(k)R_{n_I}(k).
\end{aligned}$$

In the above equation, $c_I(m)$, $s_I(m)$ and $n_I(m)$ are in-phase components of $c(m)$, $s(m)$ and

$n(m)$ respectively; $c_Q(m)$, $s_Q(m)$ and $n_Q(m)$ are quadrature components. $R_{XY}(k)$ is the autocorrelation function for the y components of x signal, where X can be one of c, s or n, y can be either I or Q. For example, $Rc_I(k)$ is the autocorrelation function for the in-phase component of channel fading signal $c(k)$. In deriving the second equal sign in the above equation, we eliminated 12 cross terms using the following assumptions: (1) in-phase and quadrature components of the same complex random process are independent, (2) $s(k)$ and $n(k)$ are zero mean random processes. For example, the product of $c_I(m)s_I(m)n_I(m)$ and $c_Q(m-k)s_Q(m-k)n_I(m-k)$ will evaluate to zero after being taking ensemble average, because $s_I(m)$ and $s_Q(m-k)$ are independent and zero mean. In deriving the last equal sign, we used the assumption that in-phase and quadrature components of $c(m)$, $s(m)$ and $n(m)$ are identical real random processes. If we further assume $s(m)$ and $n(m)$ are white random processes with variance σ_s^2 and σ_n^2 respectively, and $c(m)$ has unit variance, then the result can be further reduced to

$$\begin{aligned} & E\{4\mathrm{Re}[c(m)s(m)n^*(m)]\mathrm{Re}[c(m-k)s(m-k)n^*(m-k)]\} \\ &= 16R_{c_I}(k)R_{s_I}(k)R_{n_I}(k) \\ &= 2\sigma_s^2\sigma_n^2\delta_K(k). \end{aligned} \tag{100}$$

Another simplification about (99) is

$$\begin{aligned} & E\{\{|c(m)|^2|s(m)|^2|n(m-k)|^2\}+\{|c(m-k)|^2|s(m-k)|^2|n(m)|^2\}\} \\ &= 2E\{|c(m)|^2\}E\{|s(m)|^2\}E\{|n(m)|^2\} \\ &= 2\sigma_s^2\sigma_n^2. \end{aligned} \tag{101}$$

The first equal sign in (101) makes use of the assumption that $c(m)$, $s(m)$ and $n(m)$ are independent stationary random processes. The second equal sign is based on the assumption that $c(m)$ has unit variance.

The auto-correlation functions for $|s(k)|^2$ can be expressed as

$$\begin{aligned} R_{s^2}(k) &= E\left\{|s(m)|^2 |s(m-k)|^2\right\} \\ &= E\left\{\left[s_I^2(m)+s_Q^2(m)\right]\left[s_I^2(m-k)+s_Q^2(m-k)\right]\right\} \\ &= E\left\{s_I^2(m)s_I^2(m-k)+s_Q^2(m)s_Q^2(m-k)\right. \\ &\quad \left.+s_I^2(m)s_Q^2(m-k)+s_Q^2(m)s_I^2(m-k)\right\} \\ &= 2R_{s_I^2}(k)+2\left(\tfrac{1}{2}\sigma_s^2\right)^2 \\ &= 2\left[R_{s_I}^2(0)+2R_{s_I}^2(k)\right]+\tfrac{1}{2}\sigma_s^4 \\ &= \tfrac{1}{2}\sigma_s^4+4\left[\tfrac{1}{2}\sigma_s^2\delta_K(k)\right]^2+\tfrac{1}{2}\sigma_s^4 \\ &= \sigma_s^4\delta_K(k)+\sigma_s^4. \end{aligned} \tag{102}$$

Similarly, $|n(k)|^2$ can be expressed as

$$\begin{aligned} R_{n^2}(k) &= E\left\{|n(m)|^2 |n(m-k)|^2\right\} \\ &= \sigma_n^4\delta_K(k)+\sigma_n^4. \end{aligned} \tag{103}$$

Substitution of (100), (101), (102) and (103) into (99), and using the notation $R_{c^2}(k) = E\left\{|c(m)|^2 |c(m-k)|^2\right\}$ gives

$$\begin{aligned} R_{r^2}(k) &= R_{c^2}(k)\left[\sigma_s^4\delta_K(k)+\sigma_s^4\right]+\left[\sigma_n^4\delta_K(k)+\sigma_n^4\right]+2\sigma_s^2\sigma_n^2+2\sigma_s^2\sigma_n^2\delta_K(k) \\ &= R_{c^2}(k)\sigma_s^4+\left[R_{c^2}(0)\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4\right]\delta_K(k)+2\sigma_s^2\sigma_n^2+\sigma_n^4 \\ &= \left[R_{c^2}(k)-1\right]\sigma_s^4+\left(2\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4\right)\delta_K(k)+\left(\sigma_s^2+\sigma_n^2\right)^2. \end{aligned} \tag{104}$$

Taking Fourier transform on both sides of (104), we obtain the power spectrum of $|r(n)|^2$ as

$$S_{r^2}(f) = \sigma_s^4\left[S_{c^2}(f)-\delta(f)\right]+\left(2\sigma_s^4+2\sigma_s^2\sigma_n^2+\sigma_n^4\right)+\left(\sigma_s^2+\sigma_n^2\right)^2\delta(f) \tag{105}$$

This equation plays a key role in development of the fading signal estimation scheme in Section 3.3.2.