1972

A learning receiver for communication in three-component multipath channels

Richard Paul Brueggemann

Department: Electrical and Computer Engineering

A LEARNING RECEIVER FOR COMMUNICATION

IN THREE-COMPONENT MULTIPATH CHANNELS

by

RICHARD PAUL BRUEGGEMANN, 1939-

A DISSERTATION

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI-ROLLA

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

1972

R. E. Ziemer Advisor	Glen Haddock
David Cunningham	Lee J. Bain
William H. Tranter	John L. Rhea

ABSTRACT

An adaptive receiver is designed for transmissions through a time-varying multipath channel which may include both specular and diffuse components. The design is based on the theory of unsupervised learning machines and the receiver is a recursive structure which does not grow in complexity with each new observation, but is Bayes' optimal at each instant of time. The multipath medium is modelled as an aggregate of L conditionally independent transmission paths, each consisting of random and/or fixed reflections, and is identified in terms of three components: (1) indirect diffuse scatter, (2) indirect specular reflection, and (3) direct transmission. The channel parameters are time-varying and either independent from one signaling interval to the next or at most M-th order Markov dependent. A review of machines that learn without a teacher is presented and the learning receiver for three-component multipath is designed and modelled on the digital computer. A Monte Carlo simulation is used to estimate the performance when the channel is either Rician or nonfading. This performance, in terms of probability of error, is shown to be consistent with the existing coherent receivers and improves on their performance when the correlation between observations is increased.

ACKNOWLEDGEMENT

The author wishes to express his appreciation to Dr. R. E. Ziemer of the Electrical Engineering Department for suggesting the problem and providing the guidance necessary for its solution.

TABLE OF CONTENTS

Table of Contents (continued)

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION AND SUMMARY

A. Introduction

It is well known that the propagation phenomena one encounters in long-distance radio communication are statistical in nature. Whether the transmission of the signal from transmitter to receiver is attributable to refraction in the ionosphere or scattering in the troposphere or by the surface of the earth, unpredictable flucuations in the transmission medium cause random perturbations in the received signal. These perturbations are, for the most part, non-additive disturbances of the signal transmission and the analysis of their effects must be handled statistically, as must the design of systems to cope with the disturbances.

In many communication channels the signal that is received is a combination of direct transmission and one or more additional components received via reflections from objects or conditions within the channel. The totality of the transmission paths is termed multipath and may often be described by a combination of three components: (1) Direct; (2) Indirect specular reflections; (3) Diffuse scatter within the channel. In general, the statistical description of these components has been identified as a narrowband gaussian process for the diffuse component and extended to the Rician probability density when specular reflections are included [1-8]. This choice of statistics indeed determines the channel model, and consequently the resulting design of the receiver that is optimum in some sense.

Two explicit types of multipath channels that have received considerable attention are termed "frequency selective" and "frequency non-selective". The frequency selective channel is characterized by constructive interference at some frequencies in the transmission band, and destructive interference at others. The individual paths in such a channel are separated in time by their respective delays and as such are resolvable. The non-selective channel is frequency flat and consequently results in unresolvable paths such that the total multipath return appears to be one path in a fading channel.

The condition of multipath interference is encountered in various situations associated with terrestrial, airborne, and spaceborne communications. In the latter two cases, the multipath channel consists of a line-of-sight transmission path and possibly multiple extraneous reflected paths with well-defined differential delays. The nature of the reflecting surface determines, to a large extent, the character of these components. An example of a spaceborne communication environment is a data-relay satellite system consisting of several user satellites and a series of data-relay satellites in orbits such that there is always at least one in position to relay data to the earth station and commands to the users. A multipath scatter channel exists between each user and the relay due to reflections from the earth's surface and also transmissions through the ionosphere. When the reflecting surface is relatively smooth, the reflected ray is likely to be of a specular nature. On the other hand, when encountering a rough reflecting surface, the reflected ray is found to have a highly diffuse nature.

Based on the choice of a mathematical model of the transmission channel, receiver designs exist for each of the types of multipath disturbances encountered [1-4]. All of these designs, however, are governed by the statistical description of the channel being at most Rician. As such they are either simplifications or extensions of the probability computing receiver [1]. While this receiver is in fact optimum (in the minimum probability of error sense) for the Rician channel it does have the limitation of not accounting for the correlation between observations for the slowly fading phenomena.

In order to account for and use this correlation, a receiver must be able to adapt its decision function as it "learns" more about the channel from the observations. It is the purpose of this dissertation to demonstrate the applicability of self-learning machines [9] to the problem of communication through a multipath/fading channel.

B. Statement of the Problem

A typical binary detection problem can be stated as follows: Choose with minimum cost (Bayes' optimal decision) between the hypotheses

H_1: One of a given class of signals was transmitted

H_0: No signal was transmitted

based on observing the receiver output at a given instant of time. The N-ary decision problem, where an attempt is made to determine which signal was transmitted, is defined by the set of hypotheses

H_i: The i-th signal was transmitted, $i = 0,1,..,N$

where $i = 0$ corresponds to no signal.

The signal is assumed to be transmitted through a channel which is modelled as a collection of L conditionally independent transmission paths, each consisting of one or both of a fixed and a random component of gain (or more exactly, attenuation) defined by amplitude and phase. The fixed component can be considered the specular reflection (or the direct transmission in one case) and the random component the diffuse scattering. The channel gain components are taken to be time-varying with a value dependence between observations that is at most M-th order Markov.

The problem is stated as follows: Design a receiver for transmissions through a time-varying multipath/fading channel that is adaptive to the changing environment and is Bayes' optimal at each observation instant. The approach to this design is based on the unsupervised learning machine of Fralick [9] as modified by Hilborn and Lainiotis [10].

C. Summary

The multipath channel is modelled using in-phase and quadrature components (complex notation) and is developed following Turin's early development [4]. Using this model the probability computer is derived and presented as a basis for comparison. A review of the development of the unsupervised learning machine is presented and the optimum receiver, in the minimum probability of error sense, is shown to be one that calculates the <u>a posteriori</u> message probabilities, given all prior observations, and chooses that signal for which it is maximized. The following assumptions are used:

1. The channel delays are known.
2. The gains are slowly time-varying, i.e., a change may occur on each new observation.
3. The gains are value dependent between observations according

to a (known) M-th order Markov process.

4. The gains are independent of the transmitted signal and of each other.

5. The transmitted signals are independent with known a priori probabilities.

With the exception of Assumption 3 these restrictions also apply to previously derived receivers with the addition of another limiting assumption, viz., the channel statistics are either known or are measurable with a given distribution. In this sense the unsupervised learning receiver developed in this dissertation is essentially distribution free. The only two physical requirements on this system are that the probability density of the additive receiver noise be known and the Markov transition mechanism is known and can be implemented. This latter requirement implies that the ranges of the gain variations are also known.

The derived learning receiver is simulated on the digital computer for the purpose of investigating its performance. Monte Carlo techniques are employed and the probability of error is determined for a binary frequency-shift-keyed (FSK) transmission. For the purpose of comparing with published optimum designs, the channel is simulated as conditional Rician. Some specific cases of selective and non-selective two-path channels are analyzed and compared with existing curves [5-8] and after matching parameter values the results are shown to be consistent.

CHAPTER II

REVIEW OF THE LITERATURE

A. Multipath Channels and Receivers

Probably the first application of the probability computing receivers to the scattering channel was presented by Price [1] for the Rayleigh fading channel. He derived the statistical model of the channel as a narrowband process with known parameters and, using the maximum *a posteriori* decision criterion, developed a discrete system which computed these probabilities for each of the possible transmitted messages. A small signal-to-noise ratio approximation was also included.

Price extended this work [2] to include additive white gaussian receiver noise, and showed that the optimum receiver would operate on the received waveforms with filter functions and biasing constants determined by pairs of inhomogeneous and homogeneous integral equations, respectively. He concluded that the filter functions could be physically realizable and that for a single scatter path, the optimum receiver may be interpreted as the combination of a correlator with an optimum estimator of the Wiener type.

Later, Price and Green [3] applied communication methods to derive the RAKE receiver. This technique uses wide band transmissions and isolates, at the receiver, those portions of the transmitted signal arriving with different delays by using correlation detection techniques. Before being recombined by addition, these separated signals are processed by weighting coefficients and delays to bring them back into time coincidence. The appropriate weighting coefficients are shown to be measured by the system.

In an earlier paper Turin [4] applied statistical methods of communication theory to develop the probability computer for the conditional Rician channel. He first established both an *a priori* and *a posteriori* channel model and, using these, developed the operational form of the receiver. A few special cases were analyzed for the probability of error. This modelling procedure forms the basis for the multipath channel examined in this dissertation and is developed in detail in Chapters III and V.

Using the models developed in his early paper [4] Turin presented extensive curves [5] showing performance estimates for the non-selective coherent and non-coherent receivers. He further demonstrated similar estimates for the selective channel receiver [6]; however, this was limited to either the Rayleigh fading or nonfading channels. In both papers binary transmission was assumed.

Lindsey [7] further investigated the Rician fading multichannel reception problem where the modes were a mixture of nonfading, Rayleigh fading, and Rician fading components. Some results presented in his paper are used for comparison in Chapter VI with the learning receiver.

Jones [8] considered the three component multipath channel for non-coherent FSK and differentially coherent PSK systems for slow non-selective fading. The three components consisted of two specular components and one scatter (diffuse) component. Of interest here is the diversity combining technique he used for non-coherent FSK. The system analyzed was square-law envelope addition which is suboptimum. Some of the curves presented in this paper are used for comparison with the learning receiver performance.

B. Learning Machines

The original concept of learning machines was developed for the purpose of solving pattern recognition problems. It was only after much research was done in this respect that adaptive communication receivers were examined on this basis.

The learning machine of interest in this dissertation is classed as "learning without a teacher" and one of the first to publish a good treatise on its development was Fralick [9]. He obtained a general solution which includes the solutions to the problems of learning without a teacher, learning with a teacher, and no learning. The solution was extended to include problems in which the unknown parameter is time-varying. The resulting systems were shown to be stable and to have performance which converges to the performance of systems which have a priori knowledge of the unknown parameters being learned.

Hilborn and Lainiotis [10] derived the unsupervised learning machine for time-varying parameters that are M-th order Markov dependent between observations. This paper was written as a correction to the similar development given by Fralick. These two papers form the basis for the learning receiver that is reviewed in Chapter IV and used in Chapter V.

One of the earliest papers that dealt with learning machines was by Abramson and Braverman [11] in which the optimal use of a sequence of prior observations was made in order to recognize patterns. This was the classic "learning with a teacher" paper. Spragins [12] presented a review of the unsupervised learning machine by comparing the different approaches.

Daly [13], Keehn [14], and Scudder [15] each applied the learning procedures to problems associated with communications. Application of these learning techniques to solve the three-component multipath communication problem explicitly, has not yet been publicized.

CHAPTER III

DEVELOPMENT OF THE THEORY: MULTIPATH MODEL

In order to design a receiver that is optimum is some sense, a model describing the transmission channel is desired. This model should be in the form of statistical knowledge of the channel available to the receiver (and transmitter). The one discussed in this chapter consists of identifying discrete parameters associated with probability density functions which can be used to describe a variety of physical phenomena [4]. The model is general enough to allow investigation of both frequency selective and non-selective channels. The following development is essentially the approach presented originally by Turin [4] with three alterations: (1) Three-component multipath is explicit; (2) Quadrature component representations are used; and (3) Vector representation via time-domain sampling is used.

A. The Composite Channel Model

The transmission channel consists of an additive random disturbance and a non-additive disturbance in the form of multipath interference. For the purposes of developing the model define a transmitted "sounding signal" by

$$s(t) = \mathrm{Re}[X(t)\exp(j2\pi f_0 t)], \tag{1}$$

where $X(t)$ is the complex modulating waveform representing a possible message. The total received waveform is given by

$$v(t) = \mathrm{Re}[Z(t)\exp(j2\pi f_0 t)], \tag{2}$$

where Z(t) is the complex envelope. This waveform consists of two components, namely an additive noise component, n(t), and the multipath medium output, u(t), such that

$$v(t) = u(t) + n(t). \tag{3}$$

The additive noise is assumed to be a stationary, gaussian, white process, independent of the multipath medium and bandlimited to W_N (Hz), with a power spectral density of N_0 (watts/Hz). The noise bandwidth is considered to at least cover the transmission bandwidth, W. Using complex representation this precess is represented by

$$n(t) = \mathrm{Re}[N(t)\exp(j\pi f_0 t)]. \tag{4}$$

According to the sampling theorem for complex waveforms [16] N(t) can be completely specified by its complex time samples, N_i, taken at intervals of $1/W_N$, i.e., $N_i = N(i/W_N)$. Since n(t) is a gaussian process with a flat power spectral density over W_N (the autocorrelation function has zeros every $1/W_N$ seconds) the components of $N_i = \hat{N}_i - j\tilde{N}_i$ are independent, as are the samples. Hence the joint probability density function (pdf) of the complex samples in a T-second interval ($T >> 1/W_N$) is*

$$p(\underline{\hat{N}},\underline{\tilde{N}}) = (2\pi W_N N_0)^{-TW_N} \exp[-\frac{\underline{\hat{N}}^t\underline{\hat{N}} + \underline{\tilde{N}}^t\underline{\tilde{N}}}{2W_N N_0}], \tag{5}$$

where $\underline{\hat{N}}$ and $\underline{\tilde{N}}$ are vectors whose rows are the TW_N samples of the components of N(t) and the superscript t denotes the transpose.

*A waveform cannot be simultaneously of finite bandwidth and finite time duration; however, for $T>>1/W_N$ the approximation is very good!

The multipath medium is described in terms of elementary "sub-paths" which group together to form "paths". When the sounding signal of Equation (1), with a bandwidth less than or equal to W, is applied to the channel, the complex output from the k-th sub-path of the ℓ-th path, defined by strength $b_{\ell k}$ and delay $t_{\ell k}$, is given by

$$U_{\ell k}(t) = b_{\ell k} X(t-t_{\ell k}) \exp[j2\pi f_0 (t-t_{\ell k})]. \tag{6}$$

This assumes that the multipath medium is linear and that its physical properties do not vary appreciably across the transmission band.

The ℓ-th path is defined as a group of sub-paths whose delays differ from one another by amounts much less that the reciprocal of the bandwidth, W, i.e.,

$$|t_{\ell k} - t_{\ell j}| << 1/W. \tag{7}$$

This is the condition of "frequency non-selective" sub-paths. The ℓ-th path output is found by summing Equation (6) over all k satisfying Equation (7):

$$U_{\ell}(t) = \sum_{k} b_{\ell k} X(t-t_{\ell k}) \exp[j2\pi f_0 (t-t_{\ell k})]. \tag{8}$$

Equation (7) implies that $X(t-t_{\ell k}) \simeq X(t-\tau_{\ell})$, where τ_{ℓ} may be set equal to any one of the $t_{\ell k}$'s. By defining a path gain, a_{ℓ}, and phase, θ_{ℓ}, according to

$$a_{\ell} \exp(-j\theta_{\ell}) = \sum_{k} b_{\ell k} \exp(-j2\pi f_0 t_{\ell k}), \tag{9}$$

the complex envelope of the total L-path output is given by

$$Y(t) = \sum_{\ell=1}^{L} a_\ell \exp(-j\theta_\ell) X(t-\tau_\ell). \tag{10}$$

The different types of multipath to be considered are determined by the characteristics of a_ℓ, θ_ℓ and τ_ℓ, which, in general will contain random time-varying quantities. The individual paths are taken to be "frequency selective", i.e., the modulation delays differ by amounts greater than 1/W. This is defined by Turin as the "resolvability condition".

$$|\tau_\ell - \tau_m| > 1/W \ , \ \ell \neq m \tag{11}$$

It should be pointed out that this is not too restrictive in that the frequency non-selective case can be considered a priori as one path.

The three types of multipath channels which are to be considered include:

1. Single Component (Diffuse Scatterers). The a_ℓ and θ_ℓ are random variables that are Rayleigh and uniformly distributed, respectively.

2. Two Component (Diffuse plus Indirect Specular Reflectors). The terms in Equation (9) consist of two types: fixed and randomly time-varying. Thus

$$a_\ell \exp(-j\theta_\ell) = \alpha_\ell \exp(-j\delta_\ell) + \mu_\ell \exp(-j\varepsilon_\ell), \tag{12}$$

where α_ℓ and δ_ℓ are the fixed quantities corresponding to the specular components. The μ_ℓ and ε_ℓ are Rayleigh and uniform, respectively. This is sometimes called the "Rician Channel".

3. Three Component (Diffuse plus Indirect Specular Reflectors plus a Direct Path). This model is a direct extension of Equation (12)

by adding an additional specular component, i.e., $\alpha_0 \exp(-j\delta_0)$.

To expand Equation (10) into quadrature components, define the real and imaginary channel parameters by

$$\hat{a}_\ell - j\tilde{a}_\ell = (\hat{\alpha}_\ell + \hat{\mu}_\ell) - j(\tilde{\alpha}_\ell + \tilde{\mu}_\ell) \ , \ \ell = 0,.,\ldots,L. \tag{13}$$

Equation (10) is then

$$Y(t) = \sum_{\ell=0}^{L} [\hat{a}_\ell \hat{X}(t-\tau_\ell) + \tilde{a}_\ell \tilde{X}(t-\tau_\ell)]$$

$$+ j \sum_{\ell=0}^{L} [\hat{a}_\ell \tilde{X}(t-\tau_\ell) - \tilde{a}_\ell \hat{X}(t-\tau_\ell)], \tag{14}$$

which clearly defines $\hat{Y}(t)$ and $\tilde{Y}(t)$. In terms of vectors with the time samples for rows, Equation (14) is written as

$$\underline{\hat{Y}} + j\underline{\tilde{Y}} = \sum_{\ell=0}^{L} [\hat{a}_\ell \underline{\hat{X}}_\ell + \tilde{a}_\ell \underline{\tilde{X}}_\ell] + j \sum_{\ell=0}^{L} [\hat{a}_\ell \underline{\tilde{X}}_\ell - \tilde{a}_\ell \underline{\hat{X}}_\ell]. \tag{15}$$

The subscript ℓ on the X's denotes the signal delayed by τ_ℓ. In view of the resolvability condition, Equation (11), the time duration of $Y(t)$ will be greater than that of $X(t)$. Calling the channel output time span T' ($>T$) the total number of samples in each of $\underline{\hat{Y}}$ and $\underline{\tilde{Y}}$ must be at least $T'W$. Since $W_N > W$ then, to accurately represent the entire received waveform, the components of $Z(t)$ must have $T'W_N$ samples in their vector representation. The complete received vector of samples is

$$\underline{\hat{Z}} + j\underline{\tilde{Z}} = (\underline{\hat{Y}} + \underline{\hat{N}}) + j(\underline{\tilde{Y}} + \underline{\tilde{N}}). \tag{16}$$

The description of the ℓ-th path is now reduced to that of three parameters: $\hat{a}_\ell$, $\tilde{a}_\ell$, and τ_ℓ. These are generally random processes and are described in terms of joint probability density functions. For the purposes of this analysis the medium will be completely described by the joint first-order distribution of the three sets of characteristics: $(\hat{a}_\ell)$, $(\tilde{a}_\ell)$, and (τ_ℓ) with $(\cdot)$ denoting a vector. The joint pdf on (τ_ℓ) will be factored out to be considered separately:

$$p[(\hat{a}_\ell),(\tilde{a}_\ell),\ (\tau_\ell)] = p[(\tau_\ell)]p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell)]. \tag{17}$$

It is further assumed that all paths are conditionally independent

$$p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell)] = \prod_{\ell=0}^{L} p(\hat{a}_\ell,\tilde{a}_\ell|\tau_\ell). \tag{18}$$

The output of a three-path medium described by Equation (10) and satisfying Equation (11) is illustrated in Figure 1.

Knowledge of the channel may be divided into two types: _a priori_ and _a posteriori_. The former type may be based on a physical model of the channel; however it may reflect only ignorance of the channel. The latter is based on measurements of the channel parameters. The _a priori_ knowledge is essentially the complete knowledge of the first-order distribution of Equation (17). The _a posteriori_ knowledge is associated with the computation of the joint first-order pdf conditioned on the received waveform and the knowledge of the signal transmitted:

$$p[(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell)\,|\underline{Z},\underline{X}] = p[(\tau_\ell)\,|\underline{Z},\underline{X}]p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell),\underline{Z},\underline{X}]. \tag{19}$$

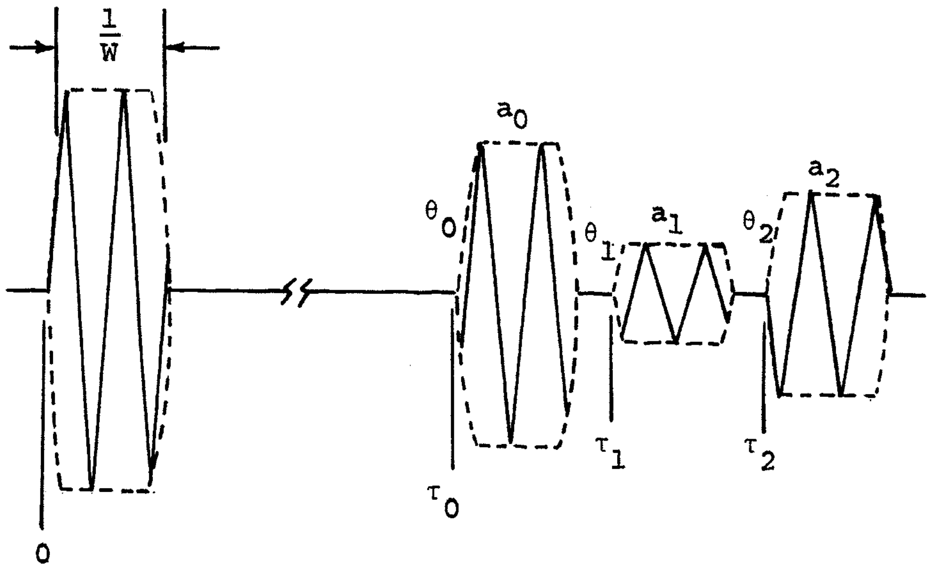


Figure 1: Resolvable Three-Path Channel

Using Baye's Rule,

$$p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell),\underline{Z},\underline{X}] = \frac{p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell)]\,p[\underline{Z}\,|\,(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}]}{p[\underline{Z}\,|\,(\tau_\ell),\underline{X}]} . \tag{20}$$

The first term in the numerator is the <u>a priori</u> distribution. The denominator becomes a normalizing factor insuring that the integral of the expression is unity. The remaining factor is evaluated from Equation (5) with $\underline{N}$ replaced by $\underline{Z}-\underline{Y}$ for fixed values of $(\hat{a}_\ell)$, $(\tilde{a}_\ell)$, and (τ_ℓ):

$$p[\underline{Z}\,|\,(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}] = p[\hat{\underline{N}} = \hat{\underline{Z}}-\hat{\underline{Y}},\ \tilde{\underline{N}} = \tilde{\underline{Z}}-\tilde{\underline{Y}}\,|\,(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}] . \tag{21}$$

This is called the conditional likelihood function of $\underline{Z}$.

B. A Priori Distribution

For the channel parameters defined in Equation (12), hence Equation (13), the a priori quadrature components, $\hat{a}_\ell$ and $\tilde{a}_\ell$, are independent gaussian random variables with variance σ_ℓ^2 and respective means $\hat{\alpha}_\ell$ and $\tilde{\alpha}_\ell$. The joint conditional pdf of Equation (18) is then

$$p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell)] = \prod_{\ell=0}^{L} \frac{1}{2\pi\sigma_\ell^2} \exp[-\frac{(\hat{a}_\ell-\hat{\alpha}_\ell)^2 + (\tilde{a}_\ell-\tilde{\alpha}_\ell)^2}{2\sigma_\ell^2}], \tag{22}$$

which factors into $p[(\hat{a}_\ell)\,|\,(\tau_\ell)]p[(\tilde{a}_\ell)\,|\,(\tau_\ell)]$. This identifies the a priori knowledge. The a priori pdf associated with each multipath channel model considered is determined from Equation (22) as follows:

1. Diffuse Multipath; $\hat{\alpha}_\ell = \tilde{\alpha}_\ell = 0$, all ℓ.
2. Diffuse plus Indirect Specular Multipath; Eliminate $\ell = 0$ term.
3. Diffuse plus Indirect Specular Multipath plus Direct Path; $\sigma_0 = 0$.

C. A Posteriori Distribution

Using the resolvability condition, Equation (11), the conditional likelihood function of Equation (21), derived in Appendix A, is

$$p[\underline{Z}|\,(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}] = (2\pi W_N N_0)^{-T'W_N} \exp[-\frac{\hat{\underline{Z}}^t\hat{\underline{Z}} + \tilde{\underline{Z}}^t\tilde{\underline{Z}}}{2W_N N_0}]$$

$$\cdot \prod_{\ell=0}^{L} \exp[\frac{\hat{a}_\ell}{W_N N_0}(\hat{\underline{Z}}^t\hat{\underline{X}}_\ell+\tilde{\underline{Z}}^t\tilde{\underline{X}}_\ell) + \frac{\tilde{a}_\ell}{W_N N_0}(\hat{\underline{Z}}^t\tilde{\underline{X}}_\ell-\tilde{\underline{Z}}^t\hat{\underline{X}}_\ell) - (\hat{a}_\ell^2+\tilde{a}_\ell^2)\frac{WE}{W_N N_0}]. \tag{23}$$

The *a posteriori* pdf is found by substituting Equations (22) and (23) into Equation (20). To be classified as a pdf it must integrate to unity. This operation is performed in Appendix B with the following result:

$$p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell),\underline{Z},\underline{X}] = \prod_{\ell=0}^{L} \frac{1}{2\pi(\sigma_\ell')^2} \exp\left[-\frac{(\hat{a}_\ell-\hat{\alpha}_\ell')^2 + (\tilde{a}_\ell-\tilde{\alpha}_\ell')^2}{2(\sigma_\ell')^2}\right], \tag{24}$$

where the primed parameters are given by

$$\begin{aligned} (\sigma_\ell')^2 &= \left(\frac{2WE}{W_N N_0} + \frac{1}{\sigma_\ell^2}\right)^{-1} \\ \hat{\alpha}_\ell' &= (\sigma_\ell')^2 \left(\frac{\hat{G}_\ell}{N_0} + \frac{\hat{\alpha}_\ell}{\sigma_\ell^2}\right) \\ \tilde{\alpha}_\ell' &= (\sigma_\ell')^2 \left(\frac{\tilde{G}_\ell}{N_0} + \frac{\tilde{\alpha}_\ell}{\sigma_\ell^2}\right). \end{aligned} \tag{25}$$

The $\hat{G}_\ell$ and $\tilde{G}_\ell$ are the quadrature components of the complex cross-correlation of $\underline{Z}$ with $\underline{X}_\ell$ as defined by Equation (A.8).

It is observed from Equation (25) that the *a posteriori* parameters reflect the *a priori* knowledge as well as the measurement. The *a priori* ignorance is identified by σ_ℓ^2 in that the larger a particular σ_ℓ is, the more uncertain is the *a priori* knowledge of the complex path gain. In the limit ($\sigma_\ell \to \infty$) the *a posteriori* parameters are defined solely by measured quantities.

It remains to determine $p[(\tau_\ell)\,|\,\underline{Z},\underline{X}]$ in Equation (19). Using Baye's Rule and recognizing that (τ_ℓ) and $\underline{X}$ are independent

$$p[(\tau_\ell)|\underline{Z},\underline{X}] = p[(\tau_\ell)] \frac{p[\underline{Z}|(\tau_\ell),\underline{X}]}{p[\underline{Z}|\underline{X}]} . \tag{26}$$

The conditional pdf $p[\underline{Z}|(\tau_\ell),\underline{X}]$ is determined from Appendix B, Equations (B.3) and (B.5), to be

$$p[\underline{Z}|(\tau_\ell),\underline{X}] = (2\pi W_N N_0)^{-T'W_N} \exp[-\frac{\hat{\underline{Z}}^t\hat{\underline{Z}} + \tilde{\underline{Z}}^t\tilde{\underline{Z}}}{2W_N N_0}]$$

$$\cdot \prod_{\ell=0}^{L} (\frac{\sigma'_\ell}{\sigma_\ell})^2 \exp\{\frac{1}{2}[(\frac{\hat{\alpha}'_\ell}{\sigma'_\ell})^2 - (\frac{\hat{\alpha}_\ell}{\sigma_\ell})^2 + (\frac{\tilde{\alpha}'_\ell}{\sigma'_\ell})^2 - (\frac{\tilde{\alpha}_\ell}{\sigma_\ell})^2]\}. \tag{27}$$

The a posteriori pdf on (τ_ℓ) is then determined by the a priori knowledge of (τ_ℓ) and the channel measurements.

Since the derivation of Equation (24) is based on the resolvability condition of Equation (11) the a priori pdf, $p[(\tau_\ell)]$, cannot be an arbitrary distribution. Turin [4] points out, however, that if the total number of paths is small, or the range of values of the τ_ℓ is large, most cases of interest will not be seriously affected by the contradiction of assuming the delays uniformly distributed and independent. Thus

$$p[(\tau_\ell)] = \prod_{\ell=1}^{L} p(\tau_\ell), \tag{28}$$

and

$$p(\tau_\ell) = \frac{1}{T_B - T_A} \tag{29}$$

for $T_A \le \tau_\ell \le T_B$ and zero elsewhere.

It should be pointed out that each of the joint pdf's as well as the conditional likelihood functions derived in this chapter can be converted to the form given by Turin [4] via a simple probability density transformation.

CHAPTER IV

DEVELOPMENT OF THE THEORY: LEARNING RECEIVERS

The material presented in this chapter is by no means original but is included for the purpose of making the dissertation complete and self-contained. The developments that follow closely adhere to the original work of Fralick [9] with certain corrections attributed to Hilborn and Lainiotis [10].

A. Systems with Fixed Parameters

Consider the multiple-hypothesis problem in which one of N possible signals, s_1, $s_2, \ldots,$ s_N, is transmitted through a channel which corrupts it by some means that is represented by a parameter vector $\underline{\theta}_i$, $i = 1,2,\ldots,N$ and by additive noise, represented by the sample function $n(t)$. The parameter vector is assumed to be fixed, but unknown. After making a sequence of K observations, each of length T, of the received waveform, $v(t)$, the receiver will be required to decide, with minimum probability of error, which of the N signals was transmitted in the K-th interval. Restated, the receiver must choose among the hypotheses:

$$H_i\colon\ v_K(t) = s_{Ki}(t,\underline{\theta}_i) + n_K(t)\ ;\ i = 1,2,\ldots,N\ , \tag{30}$$

for $(K-1)T \leq t \leq KT$.

Assuming a signal bandwidth of W, $s_{Ki}(t,\underline{\theta}_i)$ can be represented by the column vector $\underline{s}_{Ki}(\underline{\theta}_i)$, $i = 1,2,\ldots,N$, which has for its rows the 2TW samples [16] in the K-th interval. Using this notation the hypotheses are written as

$$H_i: \ \underline{V}_K = \underline{s}_{Ki}(\underline{\theta}_i) + \underline{n}_K, \ i = 1,2,\ldots,N. \tag{31}$$

Then, if the parameter vectors, $\underline{\theta}_i$, the a priori signal probabilities, P_i, and the noise statistics were known the optimum system would compute the weighted a posteriori probability density functions of $\underline{V}_K$ conditioned on $\underline{\theta}_i$ and H_i and choose H_i corresponding to the largest; i.e., choose the largest of [17]

$$P_i p(\underline{V}_K|\underline{\theta}_i, H_i) = P_i p_i(\underline{V}_K|\underline{\theta}_i) \ , \ i = 1,2,\ldots,N. \tag{32}$$

If the parameters were random with known distribution, $p(\underline{\theta}_i)$, the Bayes optimum system would average Equation (32) over each $\underline{\theta}_i$. If the distribution on $\underline{\theta}_i$ is unknown or if $\underline{\theta}_i$ is not random but unknown, then one standard procedure is to treat it as random and use the "least favorable distribution" for $\underline{\theta}_i$ and average [18].

In order to take advantage of all priori information define the sequence of all previous (K-1) observations as the matrix of column vectors:

$$\lambda_{K-1} = \underline{V}_{K-1}, \underline{V}_{K-2}, \ldots, \underline{V}_1. \tag{33}$$

The optimum system then computes the a posteriori probability density function conditioned on H_i and λ_{K-1} and weighted by P_i. (This is shown in Appendix C). In the notation of Equation (32) this is

$$P_i p_i(\underline{V}_K|\lambda_{K-1}) \ , \ i = 1,2,\ldots,N. \tag{34}$$

This is computed from Equation (32) using the conditional expectation:

$$p_i(\underline{V}_K|\lambda_{K-1}) = \int_{\underline{\theta}_i} p_i(\underline{V}_K|\underline{\theta}_i) p(\underline{\theta}_i|\lambda_{K-1}) d\underline{\theta}_i \tag{35}$$

(See Appendix D). The underlying assumption is conditional independence of the $\underline{V}_K$. The synthesis of a system which will compute $p_i(\underline{V}_K|\underline{\theta}_i)$ is a standard problem of detection theory (assuming the statistics of $\underline{n}$ are known). The problem here is to compute $p(\underline{\theta}_i|\lambda_{K-1})$.

Using Bayes' rule

$$p(\underline{\theta}_i|\lambda_{K-1}) = \frac{p(\underline{V}_{K-1}|\lambda_{K-2},\underline{\theta}_i)}{p(\underline{V}_{K-1}|\lambda_{K-2})}\, p(\underline{\theta}_i|\lambda_{K-2}). \tag{36}$$

The denominator of Equation (36) can be written in terms of the N conditional densities

$$p(\underline{V}_{K-1}|\lambda_{K-2}) = \sum_{j=1}^{N} P_j p_j(\underline{V}_{K-1}|\lambda_{K-2}). \tag{37}$$

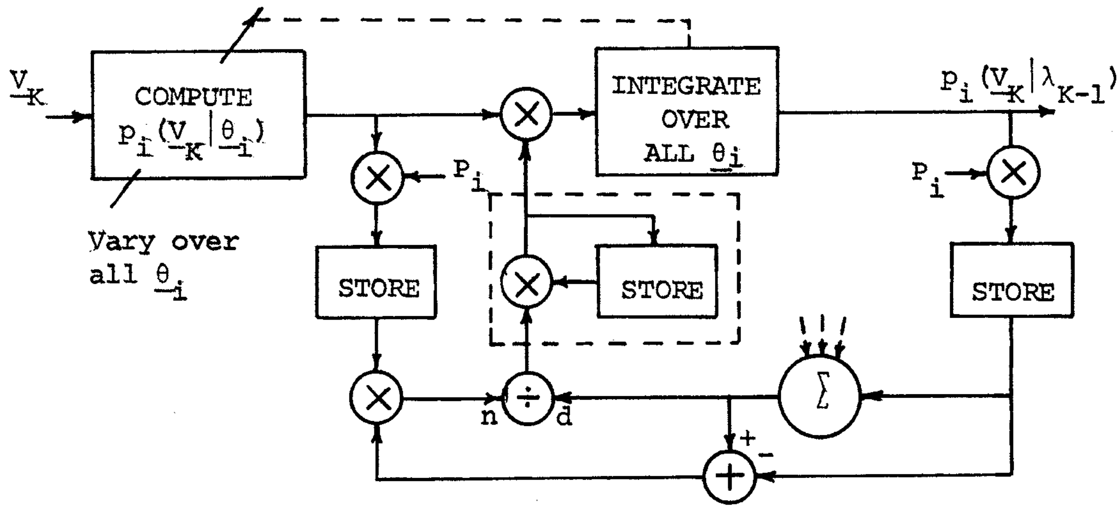
The numerator can be expanded in a similar way; however, using the conditional independence assumption, the term corresponding to H_i is free of λ_{K-2} (knowing $\underline{\theta}_i$ precludes necessity of λ_{K-2}) while the other N-1 terms do not need $\underline{\theta}_i$. The following equation results:

$$p(\underline{V}_{K-1}|\lambda_{K-2},\underline{\theta}_i) = P_i p_i(\underline{V}_{K-1}|\underline{\theta}_i) + \sum_{\substack{j=1\\ \neq i}}^{N} P_j p_j(\underline{V}_{K-1}|\lambda_{K-2}). \tag{38}$$

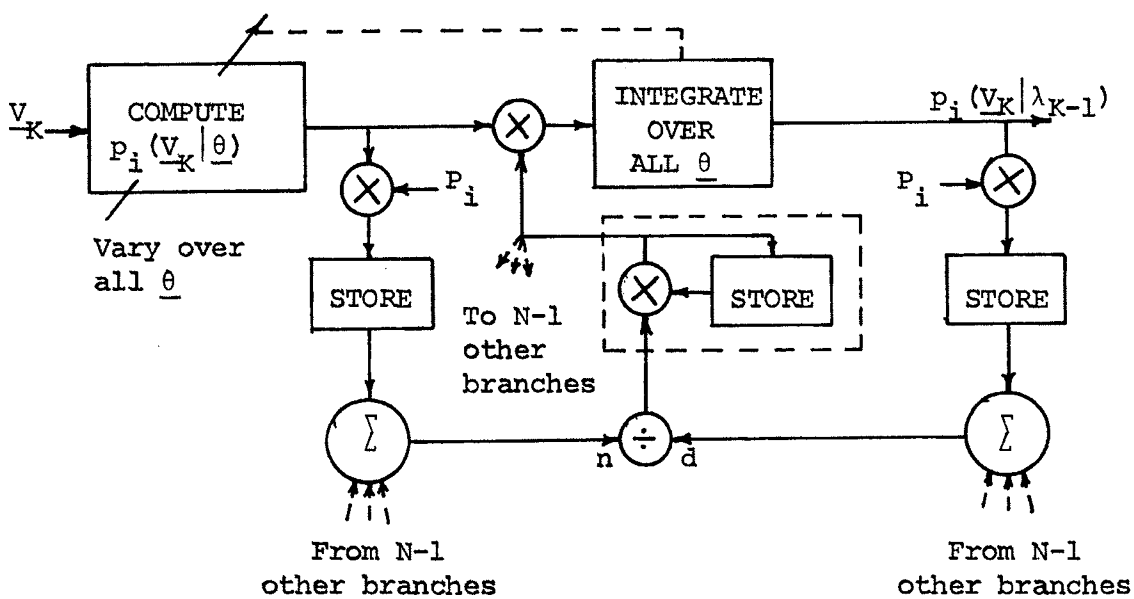
Combining Equations (36), (37) and (38), the necessary recursive relation obtains. The complete system is synthesized in Figure 2(a).

In the event that the parameter vector is independent of the hypothesis; i.e., $\underline{\theta}_i = \underline{\theta}$ for all i, then Equation (38) becomes

$$p(\underline{V}_{K-1}|\lambda_{K-2},\underline{\theta}) = \sum_{i=1}^{N} P_i p_i(\underline{V}_{K-1}|\underline{\theta}). \tag{39}$$



(a) Parameter Vector Dependent on H_i



(b) Parameter Vector Independent of H_i

Figure 2: An N-ary Learning Machine

In this case knowledge of $\underline{\theta}$ means that nothing is learned from λ_{K-2}. This system is shown in Figure 2(b).

The recursive nature of these systems implies that each decision is based on the knowledge gained from all of the previous observations. The λ_{K-1} is defined as the "learning sequence" and, since the correct classification of each member of the sequence is not given to the machine, it is said to "learn without a teacher." In order to "start" the machine some initial probability, $p_0(\underline{\theta})$ must be given. This distribution may be uniform over $\underline{\theta}$ or it may have any convenient form consistent with *a priori* knowledge of $\underline{\theta}$. The two major assumptions used were: (i) the observations are conditionally independent (requiring independent noise samples) and (ii) the *a priori* signal probabilities were known.

B. Systems with Time-Varying Parameters

The multiple-hypothesis problem of the last section is modified to account for time varying parameters. These parameters are assumed to vary at a rate commensurate with the signal bandwidth previously established. To account for the possibility of more than one parameter, a vector is used with elements corresponding to each parameter; thus each signal sample is dependent on a parameter vector possibly unique to that sample.

The i-th hypothesis on the K-th observation with parameter vector $\underline{\theta}_{Ki}$ is

$$H_i: \ \underline{V}_K = \underline{s}_{Ki}(\underline{\theta}_{Ki}) + \underline{n}_K \ , \ i = 1,2,\ldots,N. \tag{40}$$

As for the fixed parameter case, an optimum system is desired which will decide which of the N signals is contained in the K-th observation by making use of the learning sequence λ_{K-1}. Assuming that the statistical nature of the additive noise is known, a statistical model of the signal-parameter variations from observation to observation is required. This model should include a "value dependence" and a "time dependence." The former describes the way in which the current values depend on the past values while the latter is a description of the statistics of the times of occurrence of changes. For the physical problem considered in this dissertation, it is assumed that a change can take place at the start of each observation. This is designated the "general random walk."

The value dependence will be described by the probability density of the K-th realization of the parameter vector conditioned on all of the past realizations, $p(\underline{\theta}_K|\underline{\theta}_{K-1},\ldots,\underline{\theta}_1)$. Using the entire past, as this suggests, leads to a system which grows in size with K. For this reason the value dependence will be restricted to be at worst M-th order Markov; i.e.,

$$p(\underline{\theta}_K|\underline{\theta}_{K-1},\ldots,\underline{\theta}_1) = p(\underline{\theta}_K|\underline{\theta}_{K-1},\ldots,\underline{\theta}_{K-M}). \tag{41}$$

The a posteriori probability density upon which a decision will be based is again given by Equations (34) and (35) but with subscript K included on the parameter vector. Now $p(\underline{\theta}_{Ki}|\lambda_{K-1})$ can be found from the joint density of the parameter vectors on the K observations conditioned on λ_{K-1} by integrating out all $\underline{\theta}_{ki}$'s for $k < K$. Using the Markov-M dependence, this is written as

$$p(\underline{\theta}_{Ki}|\lambda_{K-1}) = \int \cdots \int p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1}|\lambda_{K-1}) d\underline{\theta}_{K-1} \cdots d\underline{\theta}_{K-M+1} \tag{42}$$

Assuming conditional independence of the observation vectors, $\underline{V}_K$, a recursive relationship is derived in Appendix E to be

$$p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1}|\lambda_{K-1}) = \frac{p(\underline{V}_{K-1}|\underline{\theta}_{K-1,i})}{p(\underline{V}_{K-1}|\lambda_{K-2})}$$

$$\cdot \int p(\underline{\theta}_K|\underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M}) p(\underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M}|\lambda_{K-2}) d\underline{\theta}_{K-M}. \tag{43}$$

For N-ary signalling

$$\frac{p(\underline{V}_{K-1}|\underline{\theta}_{K-1,i})}{p(\underline{V}_{K-1}|\lambda_{K-2})} = \frac{P_i p_i(\underline{V}_{K-1}|\underline{\theta}_{K-1,i}) + \sum_{\substack{j=1 \\ \neq i}}^{N} P_j p_j(\underline{V}_{K-1}|\lambda_{K-2})}{\sum_{j=1}^{N} P_j p_j(\underline{V}_{K-1}|\lambda_{K-2})}. \tag{44}$$

If the parameter vectors are independent of the signals Equation (44) becomes

$$\frac{p(\underline{V}_{K-1}|\underline{\theta}_{K-1})}{p(\underline{V}_{K-1}|\lambda_{K-2})} = \frac{\sum_{j=1}^{N} P_j p_j(\underline{V}_{K-1}|\underline{\theta}_{K-1})}{\sum_{j=1}^{N} P_j p_j(\underline{V}_{K-1}|\lambda_{K-2})}. \tag{45}$$

An N-ary learning receiver for Markov-M time-varying parameters is constructed as shown in Figure 2 with the sections inside of the dashed lines replaced by the system shown in Figure 3. This figure clearly shows why the parameter value dependence must be limited to the M-th order.

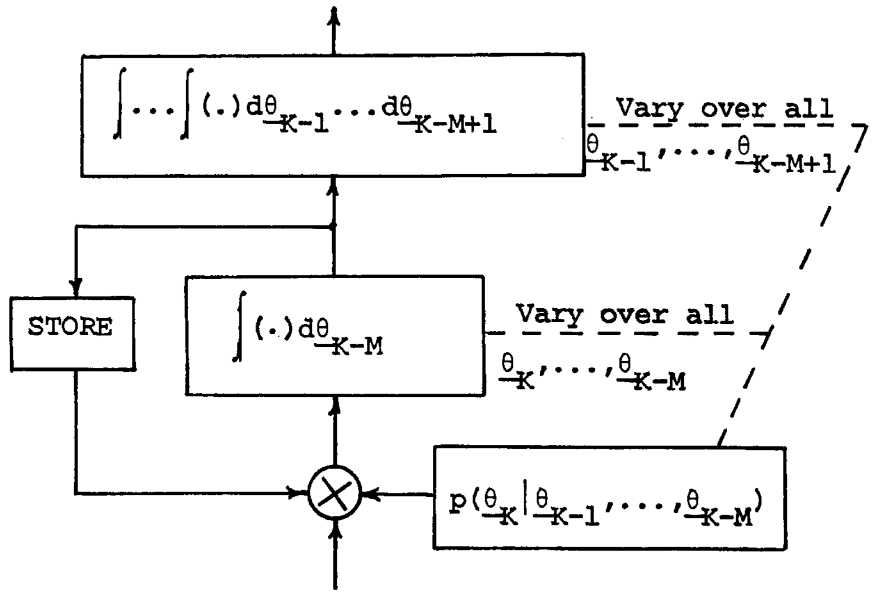


Figure 3: Time-Varying Markov-M Modification

For the special case, $M = 1$, the multiple integration is removed with the following recursive relation resulting:

$$p(\underline{\theta}_K | \lambda_{K-1}) = \int p(\underline{\theta}_K | \underline{\theta}_{K-1}) \frac{p(\underline{V}_{K-1} | \underline{\theta}_{K-1})}{p(\underline{V}_{K-1} | \lambda_{K-2})} p(\underline{\theta}_{K-1} | \lambda_{K-2}) d\underline{\theta}_{K-1}, \tag{46}$$

with Equations (44) and (45) applying accordingly.

CHAPTER V

THE UNSUPERVISED LEARNING RECEIVER

In this chapter the multipath model presented in Chapter III is combined with the unsupervised learning machine developed in Chapter IV to derive the receiver which learns the *a posteriori* probability density of the channel parameters conditioned on all of the previous received data. To lay necessary groundwork and because it was probably the first adaptive system to be used as a multipath receiver, the probability computer [4] is first discussed for quadrature channel reception. The learning receiver is then derived which removes some of the statistical restrictions imposed by the probability computer at the expense of increased complexity.

The complexity of the learning receiver is greatly reduced by limiting the observation dependence of the parameters to be first-order Markov. A storage and integration time problem is discussed and is considerably relaxed via a simplifying assumption, which, while not mathematically rigorous is rather appealing.

The chapter is concluded with a description of a digital computer simulation of the quadrature channel unsupervised learning receiver. Some of the simulation results are discussed in Chapter VI.

A. The Probability Computer

The ideal receiver, according to Woodward and Davies [19], uses its knowledge of the transmitted signal and channel to derive from the received waveform the *a posteriori* probabilities of the possible transmitted message-waveform sequences. The probability computer discussed here is restricted to per-waveform operation. That is, the receiver considers each waveform as an event which is independent of each other waveform. This independence does not in fact exist, for although the transmitted waveforms may be independent, the perturbed waveforms of the received sequence are not. This follows from the fact that the characteristics of the multipath medium have been assumed to change very slowly from one signaling baud to the next. (This restriction is removed in the learning receiver developed in Chapter IV.) The per-waveform operation assumption implies two other assumptions: that all message waveforms have the same duration and that enough time is allowed between the transmission of successive message waveforms so that no overlap of waveforms takes place at the multipath channel output. An additional restriction is that the message-waveform durations are small enough so that the multipath characteristics are essentially fixed during a signaling baud.

The two restrictions just discussed allow the multipath medium to be completely described in terms of first-order joint distributions of the parameters.

The problem is stated as follows: The transmitter transmits a sequence of message waveforms chosen independently with probabilities P_n from a set of N message waveforms

$$s_n(t) = \mathrm{Re}[X_n(t)\exp(j2\pi f_0 t)] \tag{47}$$

$n = 1,2,\ldots, N$. These waveforms and probabilities are known to the receiver. The receiver receives a signal

$$v(t) = \mathrm{Re}[Z(t)\exp(j2\pi f_0 t)], \tag{48}$$

where Z(t) is the complex envelope and is the sum of a noise waveform, N(t), and the multipath output, Y(t). The probability computer is asked to operate on Z(t), using its knowledge of the channel and a priori probabilities, P_n, in such a way as to obtain a posteriori probabilities of the possible transmitted messages, $P[X_n|Z]$, n = 1,2, ...,N.

From Bayes' theorem

$$P[X_n|Z] = \frac{P_n p[Z|X_n]}{p(Z)} . \tag{49}$$

The P_n are known and p(Z) is just a normalizing factor independent of n, so the problem reduces to that of computing the likelihoods, $p[Z|X_n]$. Using vector notation these are

$$p[\underline{Z}|\underline{X}_n] = \int\cdots\int p[\underline{Z}|(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}_n]p[(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell)]d(\hat{a}_\ell)d(\tilde{a}_\ell)d(\tau_\ell). \tag{50}$$

The conditional likelihood in the integrand is given by Equation (23) with the subscript n appropriately placed.

Using the factorization of the probability densities given by Equations (17), (18), (22), (23) and (A.9), the likelihood function becomes

$$p[\underline{Z}|\underline{X}_n] = \prod_{\ell=0}^{L} \int p(\tau_\ell)\Big[\int_{-\infty}^{\infty} \hat{p}(\underline{Z}|\tau_\ell,\hat{a}_\ell,\underline{X}_n)p(\hat{a}_\ell|\tau_\ell)d\hat{a}_\ell$$

$$\cdot \int_{-\infty}^{\infty} \tilde{p}(\underline{Z}|\tau_\ell,\tilde{a}_\ell,\underline{X}_n)p(\tilde{a}_\ell|\tau_\ell)d\tilde{a}_\ell\Big]d\tau_\ell. \tag{51}$$

If it is assumed that the (τ_ℓ) are known then Equation (51) is reduced to

$$p[\underline{Z}|(\tau_\ell),\underline{X}_n] = \hat{p}[\underline{Z}|(\tau_\ell),\underline{X}_n]\tilde{p}[\underline{Z}|(\tau_\ell),\underline{X}_n], \tag{52}$$

where

$$\hat{p}[\underline{Z}|(\tau_\ell),\underline{X}_n] = \prod_{\ell=0}^{L} \int_{-\infty}^{\infty} \hat{p}(\underline{Z}|\tau_\ell,\hat{a}_\ell,\underline{X}_n)p(\hat{a}_\ell|\tau_\ell)d\hat{a}_\ell, \tag{53}$$

and similarly for $\tilde{p}[\underline{Z}|(\tau_\ell),\underline{X}_n]$. The channel parameter pdf's are given by the a priori pdf of Equation (22) or the a posteriori pdf of Equation (24). Using the unprimed parameters for convenience the integration in Equation (53) is performed as in Appendix B. The following factors for the likelihood function result:

$$\hat{p}[\underline{Z}|(\tau_\ell),\underline{X}_n] = (2\pi W_N N_0)^{-\frac{T'W_N}{2}} \exp[-\frac{\hat{\underline{Z}}^t\hat{\underline{Z}}}{2W_N N_0}]$$

$$\cdot \prod_{\ell=0}^{L} \frac{1}{1+\beta_{n\ell}} \exp[\frac{\frac{\sigma_\ell^2}{N_0}\hat{G}_{n\ell}^2 + 2\hat{\alpha}_\ell\hat{G}_{n\ell} - 2\hat{\alpha}_\ell^2\frac{W}{W_N}E_n}{2N_0(1+\beta_{n\ell})}]$$

$$\tag{54}$$

$$\tilde{p}[\underline{Z}|(\tau_\ell),\underline{X}_n] = (2\pi W_N N_0)^{\frac{-T'W_N}{2}} \exp[-\frac{\tilde{\underline{Z}}^t\tilde{\underline{Z}}}{2W_N N_0}]$$

$$\cdot \prod_{\ell=0}^{L} \frac{1}{1+\beta_{n\ell}} \exp[\frac{\frac{\sigma_\ell^2}{N_0}\tilde{G}_{n\ell} + 2\tilde{\alpha}_\ell\tilde{G}_{n\ell} - 2\tilde{\alpha}_\ell^2\frac{W}{W_N}E_n}{2N_0(1+\beta_{n\ell})}].$$

The $\hat{G}_{n\ell}$ and $\tilde{G}_{n\ell}$ are given by

$$\hat{G}_{n\ell} = \frac{1}{W_N} (\hat{\underline{Z}}^t \hat{\underline{X}}_{n\ell} + \tilde{\underline{Z}}^t \tilde{\underline{X}}_{n\ell})$$

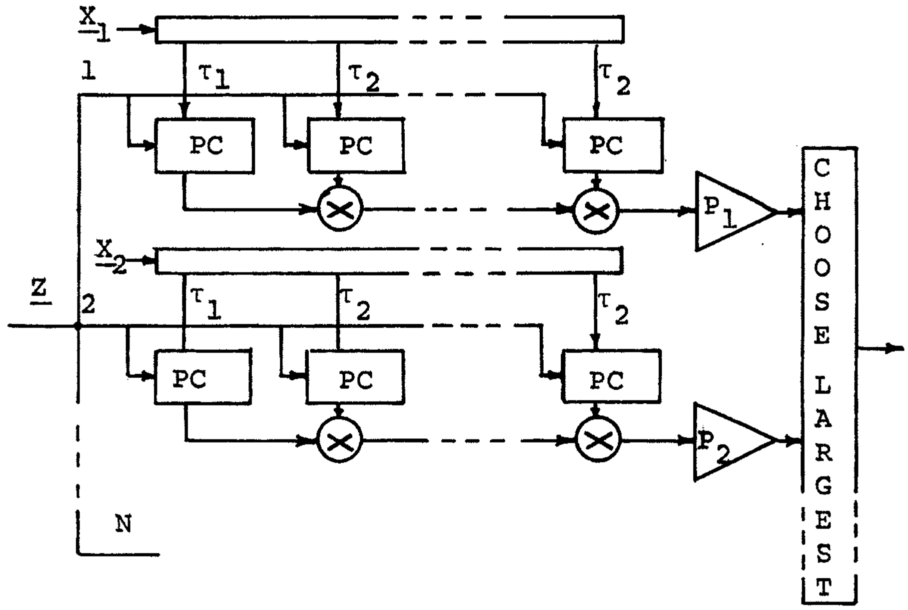
$$\tilde{G}_{n\ell} = \frac{1}{W_N} (\hat{\underline{Z}}^t \tilde{\underline{X}}_{n\ell} - \tilde{\underline{Z}}^t \hat{\underline{X}}_{n\ell}) \tag{55}$$

and $\beta_{n\ell} = \frac{2WE_n}{W_N N_0} \sigma_\ell^2$ with E_n the energy in the n-th message waveform and N_0 the power spectral density of the white noise.

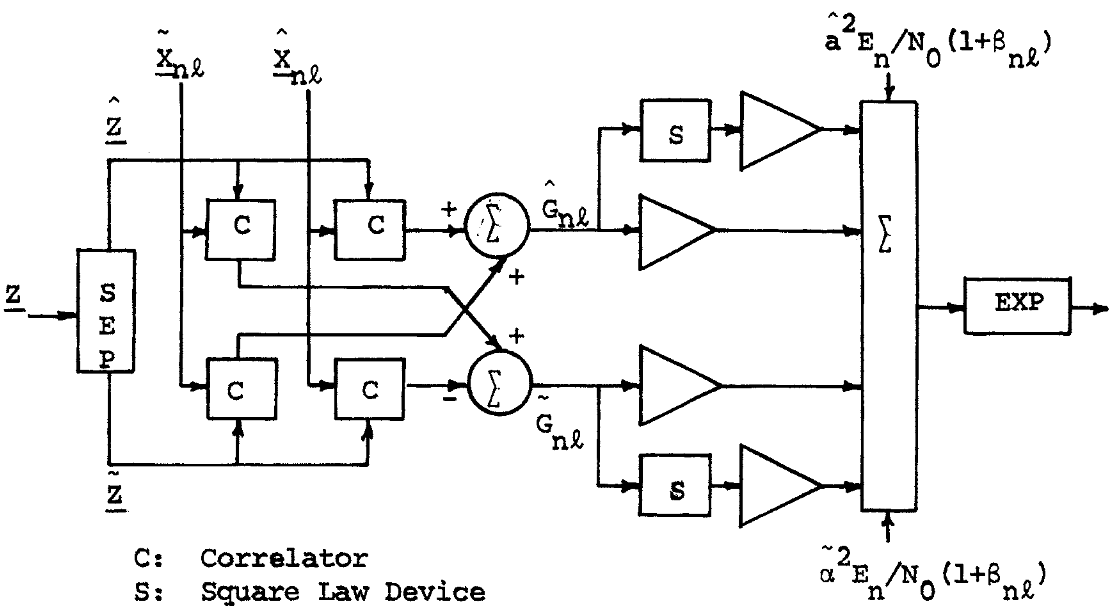
From Equations (54) and (55) it is observed that the operations performed on the received signal by the probability computer consist in 1) the cross-correlation of this waveform with the N (known) message waveforms, 2) sampling these correlations at (known) delays τ_ℓ, and 3) the sampling of the envelopes of the correlations at delays τ_ℓ. A digitized representation of this machine is illustrated by the block diagram of Figure 4. The boxes marked PC in Figure 4(a) are illustrated in Figure 4(b). The boxes marked C are the individual correlators for the quadrature components. The unmarked amplifiers have gains consistent with the constants in Equations (54) and (55) and are determined either by the a priori knowledge of the channel or the measurements indicated by Equations (25).

The form of the receiver in Figure 4(a) is essentially that of the delayed reference version of the RAKE receiver [3]. While not explicitly carried out by Price and Green in their original paper [3], this derivation was indicated in a footnote.

Two significant observations are apparent from inspection of Equations (54) and Figure 4(b). If the medium contains no random path components or the receiver has exact a posteriori knowledge of the medium then $\tau_\ell = 0$, all ℓ, and the samples of the envelope of the



(a) Delayed Reference Receiver



(b) Correlator/PDF Computer

Figure 4: The Probability Computer

cross-correlations disappear. This makes sense as complete knowledge of the quadrature channel parameters (implying no phase uncertainity) precludes the necessity of envelope sampling. On the other hand if the receiver knows a priori that the channel contains no fixed-path components and no channel measurements are made then the α_ℓ's are all zero and only the envelope sampling remains.

For the case of large additive noise ($N_0 \to \infty$) the receiver converges to the fixed-path case. This implies that, in the noise limited case, the information transferred through the channel is conveyed exclusively by the fixed-path components.

B. The Learning Receiver

It is clear from inspection of Equations (50) through (54) that some knowledge of the channel parameters is necessary a priori in order to design the probability computer, the least of which is the form of the joint probability density function of the parameters. Based on a known (or assumed) form, the parameters are then measured prior to the observation upon which a decision is based. In a sense this is adaptive and the probability computer and the RAKE each exhibit this characteristic.

The learning machine derived in Chapter IV, however, is designed to make a Bayes' optimal decision on each observation while retaining and using the information learned about the channel from all previous observations. What's more, the prior knowledge as to the form of the parameter vector pdf is not necessary so long as the initially assumed pdf encompasses the range of values of the parameters. With this (not too serious) restriction satisfied, the machine will adapt its structure as it learns the correct probability density function of the channel parameters conditioned on past observations.

Comparing Equation (49) with Equation (34) the learning receiver bases its decision on the weighted likelihood conditioned on the entire sequence of past observations rather than the present received waveform only. Using the vector envelope notation this is

$$P_n p_n(\underline{Z}_K|\lambda_{K-1}) = P_n p(\underline{Z}_K|\lambda_{K-1},\underline{X}_n). \tag{56}$$

The observation learning sequence is

$$\lambda_{K-1} = \underline{Z}_{K-1}\ ,\ \underline{Z}_{K-2},\ldots,\ \underline{Z}_1\ , \tag{57}$$

where each $\underline{Z}$ is complex. The parameter vector used in Equation (35) is, for the multipath channel, taken to be independent of the transmitted signal and is slowly time-varying in the sense that it can change from one observation interval to the next but not during a given interval. It is further assumed that the parameters are value dependent from observation to observation and that the process is homogeneous Markov of order M (finite) [20]. Define the parameter vector by

$$\underline{\theta}_K = (\hat{a}_\ell)_K,\ (\tilde{a}_\ell)_K,\ (\tau_\ell)_K. \tag{58}$$

The conditional likelihood of interest here is then, from Equation (35),

$$p_n(\underline{Z}_K|\lambda_{K-1}) = \int p_n(\underline{Z}_K|\underline{\theta}_K)p(\underline{\theta}_K|\lambda_{K-1})d\underline{\theta}_K. \tag{59}$$

The integration in Equation (59) is of multiplicity 3L. The conditional pdf to be learned is $p(\underline{\theta}_K|\lambda_{K-1})$.

In order to design this learning receiver, it is required to have a priori knowledge of the form of $p_n(\underline{Z}_K|\underline{\theta}_K)$ and the Markov-M transition mechanism. Assuming these are known then, from Equation (42),

$$p(\underline{\theta}_K|\lambda_{K-1}) = \int \cdots \int p(\underline{\theta}_K, \underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M+1}|\lambda_{K-1}) d\underline{\theta}_{K-1} \cdots d\underline{\theta}_{K-M+1} \tag{60}$$

and from Equation (43) the recursive relationship is

$$p(\underline{\theta}_K, \underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M+1}|\lambda_{K-1})$$

$$= \frac{p(\underline{Z}_{K-1}|\underline{\theta}_{K-1})}{p(\underline{Z}_{K-1}|\lambda_{K-2})} \int p(\underline{\theta}_K|\underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M}) \cdot p(\underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M}|\lambda_{K-2}) d\underline{\theta}_{K-M}. \tag{61}$$

The integration in Equation (60) has multiplicity 3(M-1)L and in Equation (61) has 3L. The total number of integrations in Equations (59), (60) and (61) is then 3(M+1)L, so it is easily seen why M is restricted. For the purposes of designing a receiver, no loss of generality will occur if M is chosen as 1. The recursive conditional pdf to be learned is then given by Equation (46) and repeated here:

$$p(\underline{\theta}_K|\lambda_{K-1}) = \int p(\underline{\theta}_K|\underline{\theta}_{K-1}) \frac{p(\underline{Z}_{K-1}|\underline{\theta}_{K-1})}{\bar{p}(\underline{Z}_{K-1}|\lambda_{K-2})} p(\underline{\theta}_{K-1}|\lambda_{K-2}) d\underline{\theta}_{K-1}. \tag{62}$$

It will be assumed that the parameters given in Equation (58) are conditionally independent, i.e.,

$$p(\underline{\theta}_K|\underline{\theta}_{K-1}) = p[(\hat{a}_\ell)_K|(\hat{a}_\ell)_{K-1}]p[(\tilde{a}_\ell)_K|(\tilde{a}_\ell)_{K-1}]p[(\tau_\ell)_K|(\tau_\ell)_{K-1}]. \tag{63}$$

Using the factorization of $p(\underline{Z}_{K-1}|\underline{\theta}_{K-1})$ shown in Equation (A.9) the conditional parameter pdf is then

$$p[(\hat{a}_\ell)_K,(\tilde{a}_\ell)_K,(\tau_\ell)_K|\lambda_{K-1}] = \int p[(\tau_\ell)_K|(\tau_\ell)_{K-1}]$$

$$\cdot \Big(\int p[(\tilde{a}_\ell)_K|(\tilde{a}_\ell)_{K-1}] \Big\{ \int p[(\hat{a}_\ell)_K|(\hat{a}_\ell)_{K-1}] \qquad (64)$$

$$\cdot \frac{\sum_{n=1}^{N} P_n \hat{p}_n[\underline{Z}_{K-1}|(\tau_\ell)_{K-1},(\hat{a}_\ell)_{K-1}]\tilde{p}_n[\underline{Z}_{K-1}|(\tau_\ell)_{K-1},(\tilde{a}_\ell)_{K-1}]}{\sum_{n=1}^{N} P_n p_n[\underline{Z}_{K-1}|\lambda_{K-2}]}$$

$$\cdot\, p[(\hat{a}_\ell)_{K-1},(\tilde{a}_\ell)_{K-1},(\tau_\ell)_{K-1}|\lambda_{K-2}]d(\hat{a}_\ell)_{K-1}\Big\}\; d(\tilde{a}_\ell)_{K-1}\Big)d(\tau_\ell)_{K-1}.$$

When the path delays are assumed known (or estimated) a priori the recursive conditional pdf of the parameters simplifies. The integrations over the (τ_ℓ) shown in Equation (64) and implied in Equation (59) are eliminated. The receiver will now be designed to learn only the quadrature gain parameters keeping in mind that learning the delay characteristics involves only the additional L-fold integration over the range of delays. The (τ_ℓ) will be dropped in the succeeding equations with the knowledge of its values understood. The recursive conditional pdf of the parameters is now

$$p[(\hat{a}_\ell)_K,(\tilde{a}_\ell)_K|\lambda_{K-1}] = \int p[(\tilde{a}_\ell)_K|(\tilde{a}_\ell)_{K-1}] \Big\{ \int p[(\hat{a}_\ell)_K|(\hat{a}_\ell)_{K-1}]$$

$$\cdot \frac{\sum_{n=1}^{N} P_n \hat{p}_n[\underline{Z}_{K-1}|(\hat{a}_\ell)_{K-1}]\tilde{p}_n[\underline{Z}_{K-1}|(\tilde{a}_\ell)_{K-1}]}{\sum_{n=1}^{N} P_n p_n[\underline{Z}_{K-1}|\lambda_{K-2}]}\, p[(\hat{a}_\ell)_{K-1},(\tilde{a}_\ell)_{K-1}|\lambda_{K-2}] \qquad (65)$$

$$\cdot d(\hat{a}_\ell)_{K-1}\Big\}\; d(\tilde{a}_\ell)_{K-1}.$$

The conditional likelihood to be used in the decision process is

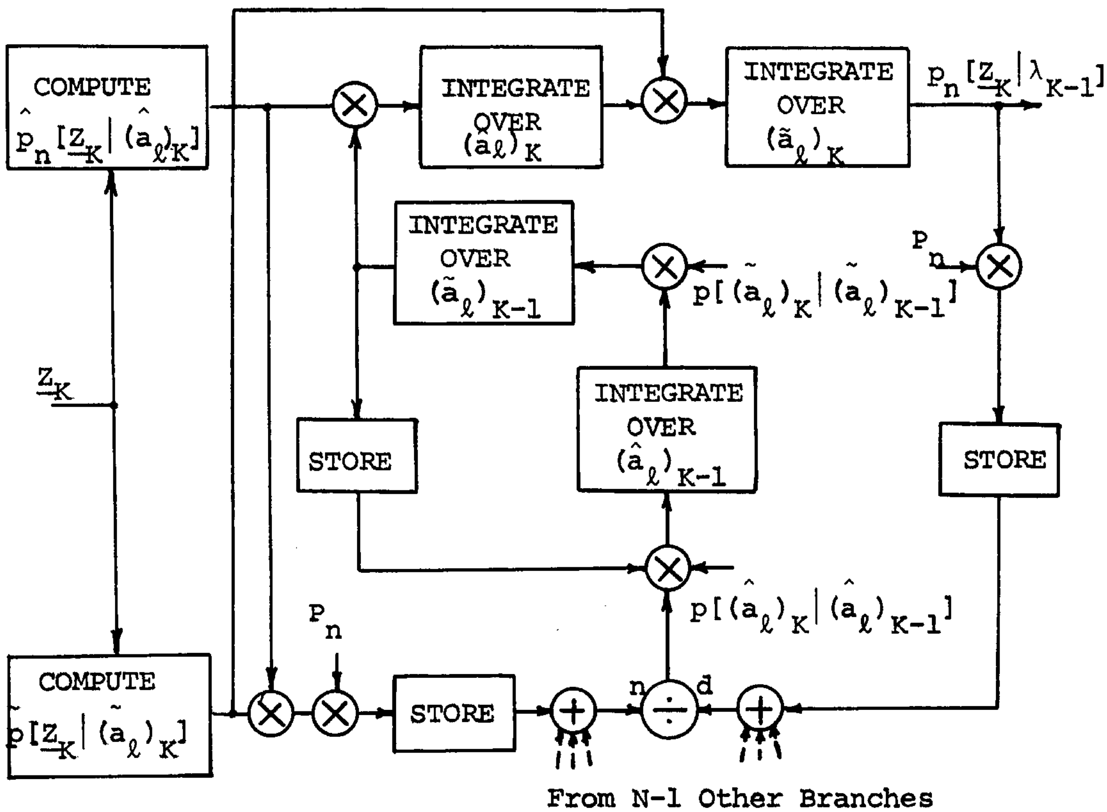
$$p_n[\underline{Z}_K|\lambda_{K-1}] = \int \tilde{p}_n[\underline{Z}_K|(\tilde{a}_\ell)_K] \; \{ \int \hat{p}_n[\underline{Z}_K|(\hat{a}_\ell)_K] \quad (66)$$

$$p[(\hat{a}_\ell)_K, (\tilde{a}_\ell)_K|\lambda_{K-1}]d(\hat{a}_\ell)_K\} \; d(\tilde{a}_\ell)_K,$$

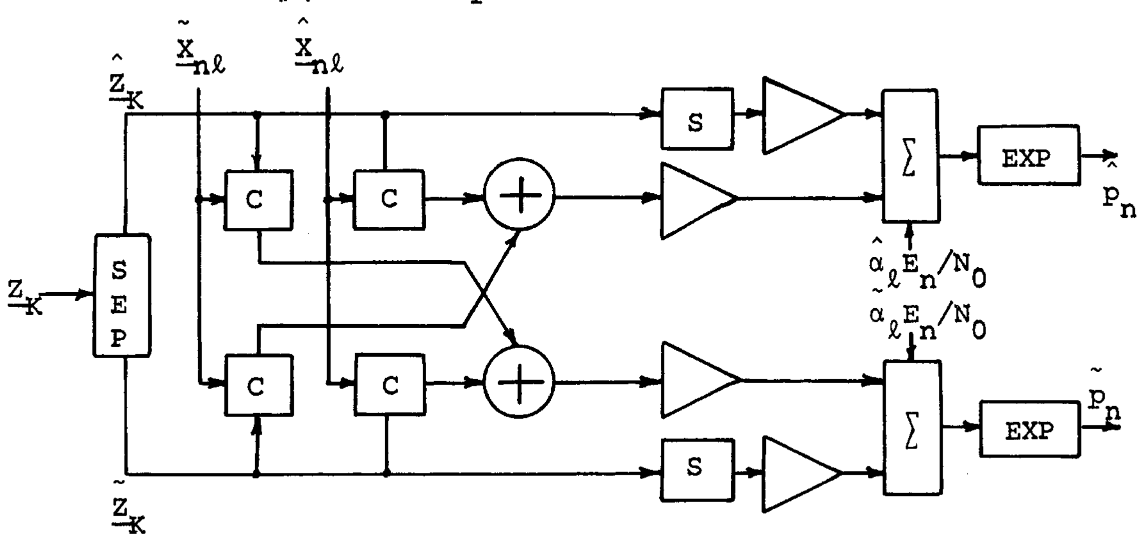
where $\tilde{p}_n[\underline{Z}_K|(\tilde{a}_\ell)_K]$ and $\hat{p}_n[\underline{Z}_K|(\hat{a}_\ell)_K]$ are given by the factors of Equation (23) (See Equations A.9) with τ_ℓ implied. The learning receiver described by Equations (65) and (66) is shown in Figure 5. The computation of Equation (23) shown in Figure 5(b) is similar to the probability computer counterpart of Figure 4(b). The main difference lies in the absence of the computation of the sampled envelope of the cross-correlation from the learning receiver.

In a similar problem associated with the Rayleigh fading channel Fralick [9] indicated (via a short proof) that the joint conditional parameter pdf that is learned can be factored, implying conditional independence. While an inspection of Equations (65) and (66) clearly indicates that this is not the case here, it nevertheless is a condition which, if assumed true, will greatly simplify the receiver structure by reducing the amount of storage and the number of integrations necessary. Assuming digital operation these requirements are determined by (1) the number of signals to be stored, N, (2) the number of time samples of each signal, N_s, and received waveform, N_z, (3) the number of paths, L and (4) the number of possible values of the parameters to be considered, N_T. The storage budget is as follows:

Markov Transition Mechanism	N_T^2
Samples of Stored Signals:	$2NN_s$



(a) The Composite Structure



(b) Computation of Equation (23)

Figure 5: The Known Delay Learning Receiver

Samples of Received Waveform: $2\ N_Z$

Values of Parameters: $N_T L$

Conditional PDF of Parameters: $(N_T)^{2L}$

Conditional Likelihoods: $2NN_T L$

The number of integrations is 2L each for the learned conditional pdf on the parameters and the computed conditional likelihoods. It is the inner L integrations performed in each of Equations (65) and (66) that require the most computations in a digital processor. In performing this integration digitally a total of $(N_T)^{2L}$ computations are performed for Equation (65) and $(N_T)^{4L}$ for Equation (66). If it is assumed that the conditional parameter pdf can be factored, then

$$p[(\hat{a}_\ell)_K, (\tilde{a}_\ell)_K | \lambda_{K-1}] = p[(\hat{a}_\ell)_K | \lambda_{K-1}] p[(\tilde{a}_\ell)_K | \lambda_{K-1}]. \tag{67}$$

The simplified learning receiver is then described by the following equations.

$$p[(\hat{a}_\ell)_K | \lambda_{K-1}] = \int p[(\hat{a}_\ell)_K | (\hat{a}_\ell)_{K-1}] \frac{\sum_{n=1}^{N} P_n \hat{p}_n [\underline{Z}_{K-1} | (\hat{a}_\ell)_{K-1}]}{\sum_{n=1}^{N} P_n \hat{p}_n [\underline{Z}_{K-1} | \lambda_{K-2}]} \cdot p[(\hat{a}_\ell)_{K-1} | \lambda_{K-2}] d(\hat{a}_\ell)_{K-1} \tag{68}$$

$$\hat{p}_n [\underline{Z}_K | \lambda_{K-1}] = \int \hat{p}_n [\underline{Z}_K | (\hat{a}_\ell)_K] p[(\hat{a}_\ell)_K | \lambda_{K-1}] d(\hat{a}_\ell)_K,$$

with similar equations for the quadrature component. The reduction in storage occurs in the conditional pdf of the parameters (which is the largest). The storage requirement changes from $(N_T)^{2L}$ to $2(N_T)^L$

which is a substantial reduction for $N_T > 2$. The reduction in the digital integration is similar: $(N_T)^{2L}$ and $(N_T)^{4L}$ become $2(N_T)^L$ and $4(N_T)^L$, respectively.

Another assumption is made which, while not as restrictive, does simplify the processing slightly. The transmitted modulation envelope $X_n(t)$ is considered to be purely real. By making this assumption the problem simplifies to a multipath channel consisting of two quadrature components each operating independently on the transmitted signal. The Bayes' optimum learning receiver then consists of two quadrature channel processors, operating independently, and computing conditional likelihoods that are then weighted by the a priori signal probabilities, multiplied together and compared for the decision.

The net result of these assumptions is illustrated by the following equations for the quadrature channel learning receiver:

$$p[(\hat{a}_\ell)_K|\hat{\lambda}_{K-1}] = \int p[(\hat{a}_\ell)_K|(\hat{a}_\ell)_{K-1}] \frac{\sum_{n=1}^{N} P_n p_n[\underline{\hat{Z}}_{K-1}|(\hat{a}_\ell)_{K-1}]}{\sum_{n=1}^{N} P_n p_n[\underline{\hat{Z}}_{K-1}|\hat{\lambda}_{K-2}]} \cdot p[(\hat{a}_\ell)_{K-1}|\hat{\lambda}_{K-2}]\, d(\hat{a}_\ell)_{K-1} \tag{69}$$

$$p_n[\underline{\hat{Z}}_K|\hat{\lambda}_{K-1}] = \int p_n[\underline{\hat{Z}}_K|(\hat{a}_\ell)_K]\, p[(\hat{a}_\ell)_K|\hat{\lambda}_{K-1}]\, d(\hat{a}_\ell)_K .$$

Similar equations can be written for the quadrature channel. A more usable form can be written by taking advantage of the factorization permitted by the independence of the L paths (see Equation (23)). Equations (69) are then written

$$p(\hat{a}_{1K},\ldots,\hat{a}_{LK}|\hat{\lambda}_{K-1}) = \frac{1}{p[\hat{\underline{Z}}_{K-1}|\hat{\lambda}_{K-2}]} \sum_{n=1}^{N} P_n$$

$$\cdot \int p(\hat{a}_{LK}|\hat{a}_{L,K-1})p_n(\hat{\underline{Z}}_{K-1}|\hat{a}_{L,K-1})\ldots\int p(\hat{a}_{1K}|\hat{a}_{1,K-1})p_n(\hat{\underline{Z}}_{K-1}|\hat{a}_{1,K-1})$$

$$\cdot\, p(\hat{a}_{1,K-1},\ldots,\hat{a}_{L,K-1}|\hat{\lambda}_{K-2})d\hat{a}_{1,K-1}\ldots d\hat{a}_{L,K-1} \tag{70}$$

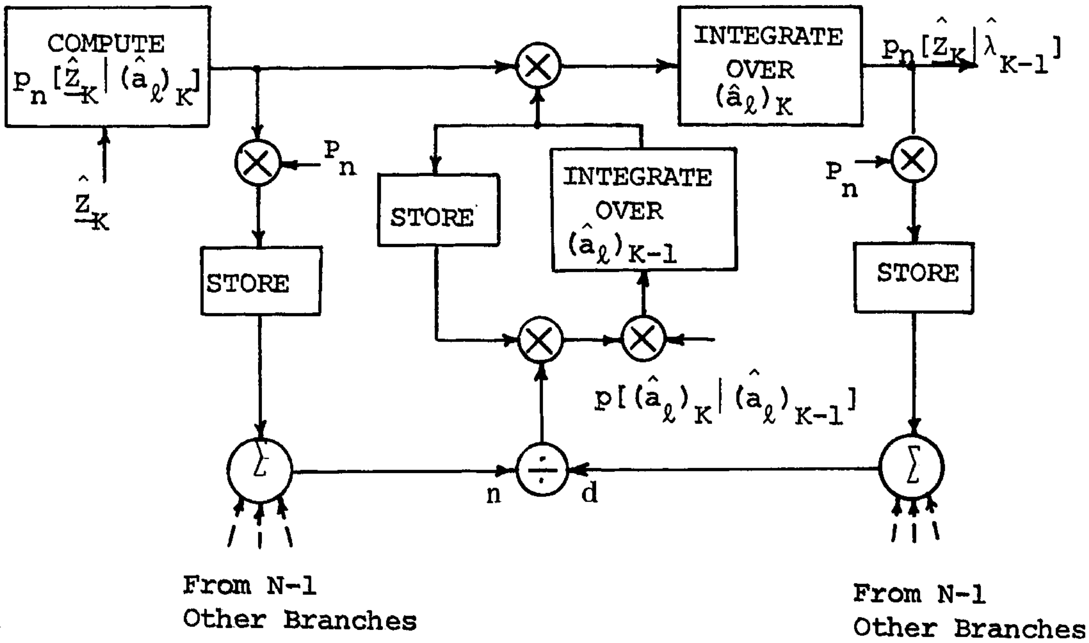
$$p_n[\hat{\underline{Z}}_{K-1}|\hat{\lambda}_{K-1}] = \int p_n(\hat{\underline{Z}}_K|\hat{a}_{LK})\ldots\int p_n(\hat{\underline{Z}}_K|\hat{a}_{1K})p(\hat{a}_{1K},\ldots\hat{a}_{LK}|\hat{\lambda}_{K-1})$$

$$\cdot\, d\hat{a}_{1K}\ldots d\hat{a}_{LK}\,.$$

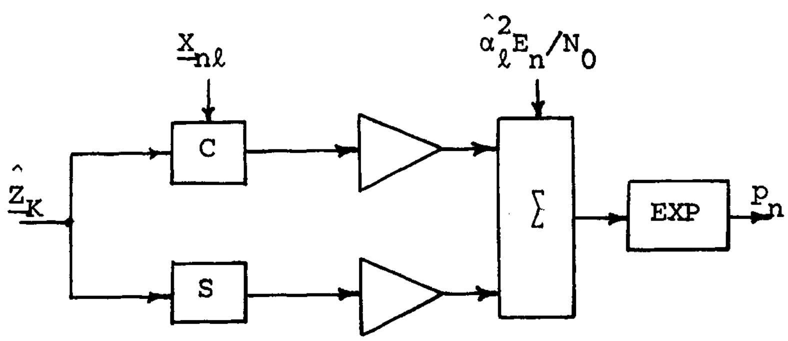
The receiver represented by these equations is shown in Figure 6.

C. Digital Simulation of a Binary Learning Receiver

In order to demonstrate the capabilities of the learning receiver a digital computer program has been developed to simulate the machine described by Equations (70). The special case of binary signaling is implemented using the historical representation of X_1 being a Mark and X_2 a Space. The program is flexible enough that the form of X_1 and X_2 is variable according to choice. For the purpose of comparing the performance of this machine with those reported in the literature, the channel is modelled as conditional Rician. While the computation time required by Equations (70) is not extensive for each observation, the total time required to perform the computations for the order of 100 observations in enough to require that some simplification



(a) Real Channel Receiver



(b) Computation of $p_n[\hat{\underline{Z}}_K \mid (\hat{a}_\ell)_K]$

Figure 6: Known Delay Real Channel
Learning Receiver

be employed. To avoid any further approximations on the machine itself the simplification used is to examine the multipath medium as a two-path frequency selective channel where each path can be any combination of the three components described in Chapter III. The frequency non-selective paths are then modelled as just one path consisting of from one to three components.

The binary decision process is given by

$$P_2 p_2(\hat{\underline{Z}}_K|\hat{\lambda}_{K-1}) p_2(\tilde{\underline{Z}}_K|\tilde{\lambda}_{K-1}) \underset{H_1}{\overset{H_2}{\gtrless}} P_1 p_1(\hat{\underline{Z}}_K|\hat{\lambda}_{K-1}) p_1(\tilde{\underline{Z}}_K|\tilde{\lambda}_{K-1}), \tag{71}$$

with the probability of error, P_e, being given by the total probability of an incorrect decision. Due to the recursive nature of the learning procedure, the bit error rate computation is intractable in closed form. This necessitates the use of Monte Carlo techniques in the simulation; that is, the transmitted signal is chosen randomly with equal probabilities between X_1 and X_2. The bit error probability is then approximated by the total number of incorrect decisions divided by the number of trials.

Using two paths for the channel results in learned pdf's of the channel gain quadrature components that are each two dimensional arrays. In order to monitor the learning procedure the program is directed to output these pdf's at pre-specified observations. The decision variables as well as the decisions and transmission selections are printed at each observation to keep track of the errors as they occur.

The first-order Gauss-Markov dependence between adjacent observations of the channel gain components is given by

$$\hat{a}_K = b\,\hat{a}_{K-1} + \hat{\varepsilon}_K$$

$$0 \leq b \leq 1, \tag{72}$$

$$\tilde{a}_K = b\,\tilde{a}_{K-1} + \tilde{\varepsilon}_K$$

with the mutual independence of $\hat{a}_{K-1}$, $\tilde{a}_{K-1}$, $\hat{\varepsilon}_K$, and $\tilde{\varepsilon}_K$. The ε_K's are random perturbations in the gain components and are distributed as $N(\mu_\varepsilon, \sigma_\varepsilon)$. The transition pdf's are then of the form

$$p(a_K|a_{K-1}) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp[-\frac{(a_K - ba_{K-1} - \mu_\varepsilon)^2}{2\sigma_\varepsilon^2}], \tag{73}$$

with the circumflex's appropriately placed.

From Equations (72) the parameters of the random perturbations' pdf are easily found to be

$$\mu_\varepsilon = \alpha_K - b\,\alpha_{K-1}$$

$$\tag{74}$$

$$\sigma_\varepsilon^2 = V(a_K) - b^2 V(a_{K-1}),$$

where the α's are mean values of the a's and $V(\cdot)$ represents the variance. If the mean and variance of the channel gains are constant at α and σ^2 respectively, then

$$\mu_\varepsilon = (1 - b)\alpha$$

$$\tag{75}$$

$$\sigma_\varepsilon^2 = (1 - b^2)\sigma^2.$$

Also, under this condition the correlation coefficient between observations is simply b.

In illustrating the performance of the learning receiver via the graphs presented in Chapter VI the following parameters are defined

$$\gamma^2 = \frac{\hat{\alpha}^2 + \tilde{\alpha}^2}{\sigma^2}$$

$$= \frac{\text{Twice the power in the specular path component}}{\text{Average power in the random path component}}\,, \tag{76}$$

where the subscript ℓ has been dropped because of the restricted number of paths examined. Also

$$\beta = \frac{\sigma^2}{\sigma_N^2}$$

$$= \frac{\text{Average power in the random path component}}{\text{Average power in the additive noise}}\,, \tag{77}$$

where $\sigma_N^2 = W_N N_0$. The underlying normalization implied in Equations (76) and (77) is the unit power in the signals.

$$\underline{x}_1^t\, \underline{x}_1 = \underline{x}_2^t\, \underline{x}_2 = 1. \tag{78}$$

CHAPTER VI

PERFORMANCE ESTIMATES

As with any communications receiver design, a good measure of quality, aside from its relative complexity, is the probability of error as a function of the signal-to-noise ratio. In the case of multipath interference, a trade-off between complexity and the ability of the receiver to utilize the entire received waveform in its decision process is necessary before selecting a design. The learning receiver discussed in this dissertation, while being rather complex in its structure, makes complete use of the total channel output. This quality is only realized if the performance of the learning receiver is at least as good as the non-learning optimum systems heretofore reported [1-4] when operated under similar conditions.

It is the purpose of this chapter to present some results of a Monte Carlo simulation of the learning receiver when receiving signals at the output of a Rician channel and a specular reflective channel. To simplify the computation a binary symmetric FSK transmission is used and a slow fading channel is assumed. Both selective and nonselective channels are considered.

A. The Learned Probability Density Function

The first of Equations (70) is the joint pdf that the learning receiver must learn in order to make the Bayes' optimal decision. For the two-path case modelled here, this joint density can be represented as a two dimensional array of its samples. For the two types of channels analyzed, the most interesting cases are those for which the channel is frequency selective, giving resolvable paths.

For the selective Rician channel, the direct specular component is assumed to be resolvable from the indirect diffuse component. Figures 7(a) and (b) illustrate the center cuts in the joint pdf of the real components of the path gains. The parameters used for these graphs imply unity signal to receiver noise in each path. The machine is initiated with a uniform pdf containing the channel gains in its range [9], shown as $K = 0$, and the learned pdf is shown after the first and tenth observations. The reason for the apparent speed with which the machine "locks" onto the true pdf is that the channel is modelled as a Gauss-Markov process with $b = .1$ which suggests that the receiver's first estimate will be gaussian-like in shape. Of significance also is the relatively good estimate that is made of the standard deviation, the true value of which is 0.2 in this calculation.

The receiver's learning ability is further illustrated by the learned pdf's for the two ray specular channel. The principal axis cuts are shown in Figures 8(a) and (b) for the equal path gain situation. The tendency toward the gaussian shape is still prevalent, and the variance is rapidly decreasing with K. This channel is modelled with correlation coefficient $b = .95$. The reason for selecting the parameters such that $\beta\gamma^2 = 2$ will become evident in the next section.

B. Error Probability

Before proceeding with the comparison of error rates for the various receivers some discussion of the literature is necessary. Turin originally defined the parameters β and γ similarly to

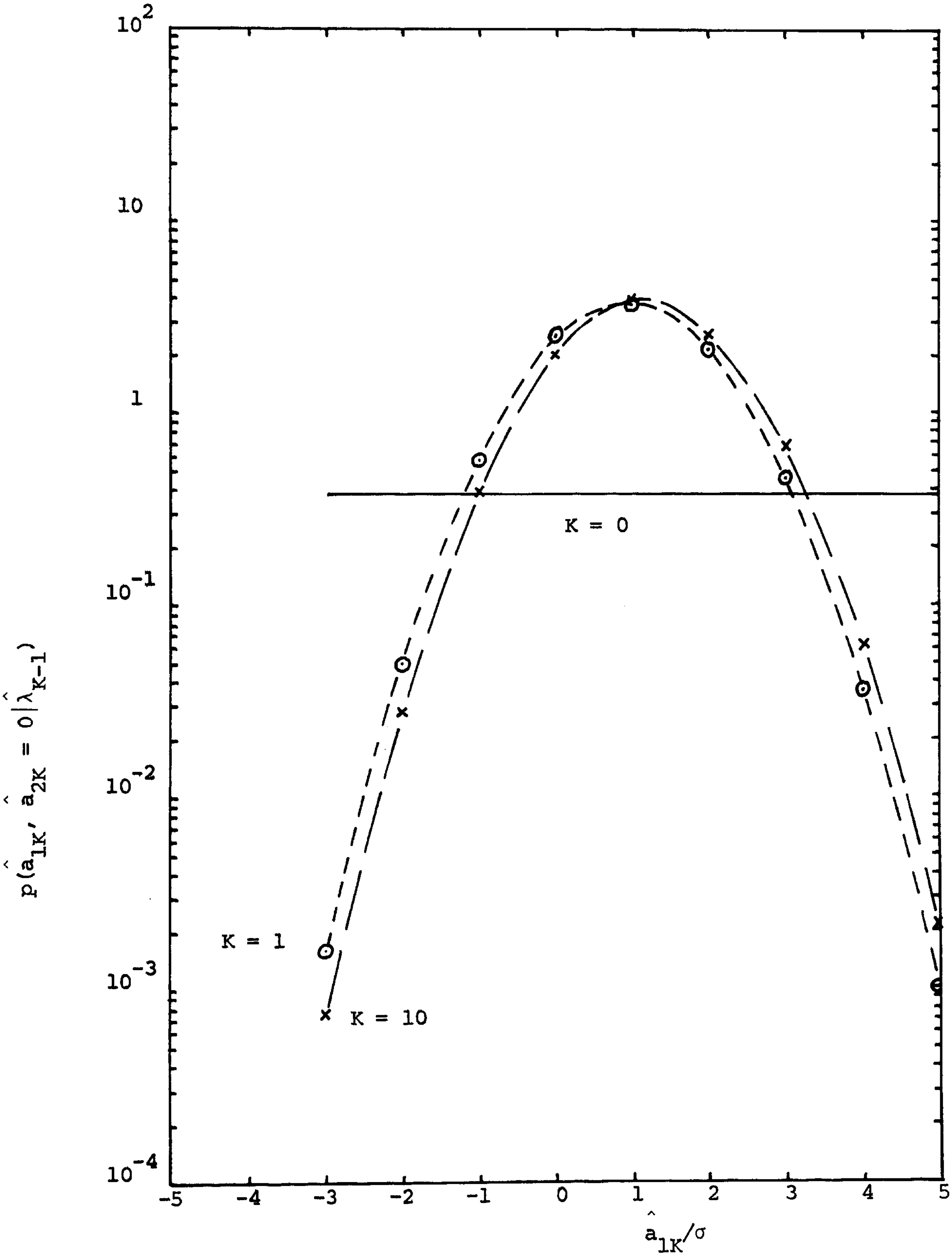


Figure 7(a): Learned PDF - Selective Rician Channel. $a_{2K} = 0$, $\gamma = 1$, $\beta = 1$

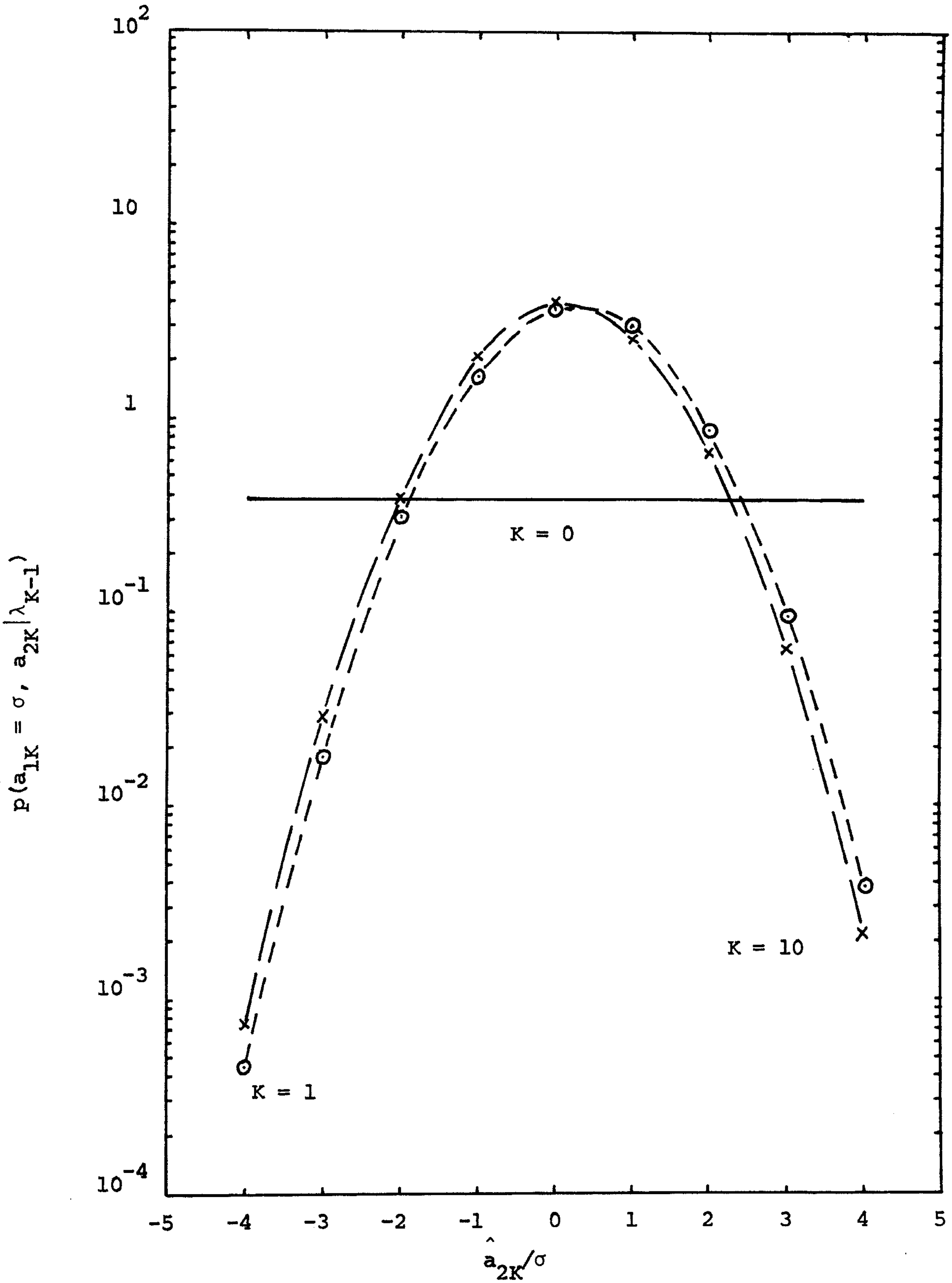


Figure 7(b): **Learned PDF - Selective Rician Channel.** $\hat{a}_{1K} = \sigma,\ \gamma = 1,\ \beta = 1$

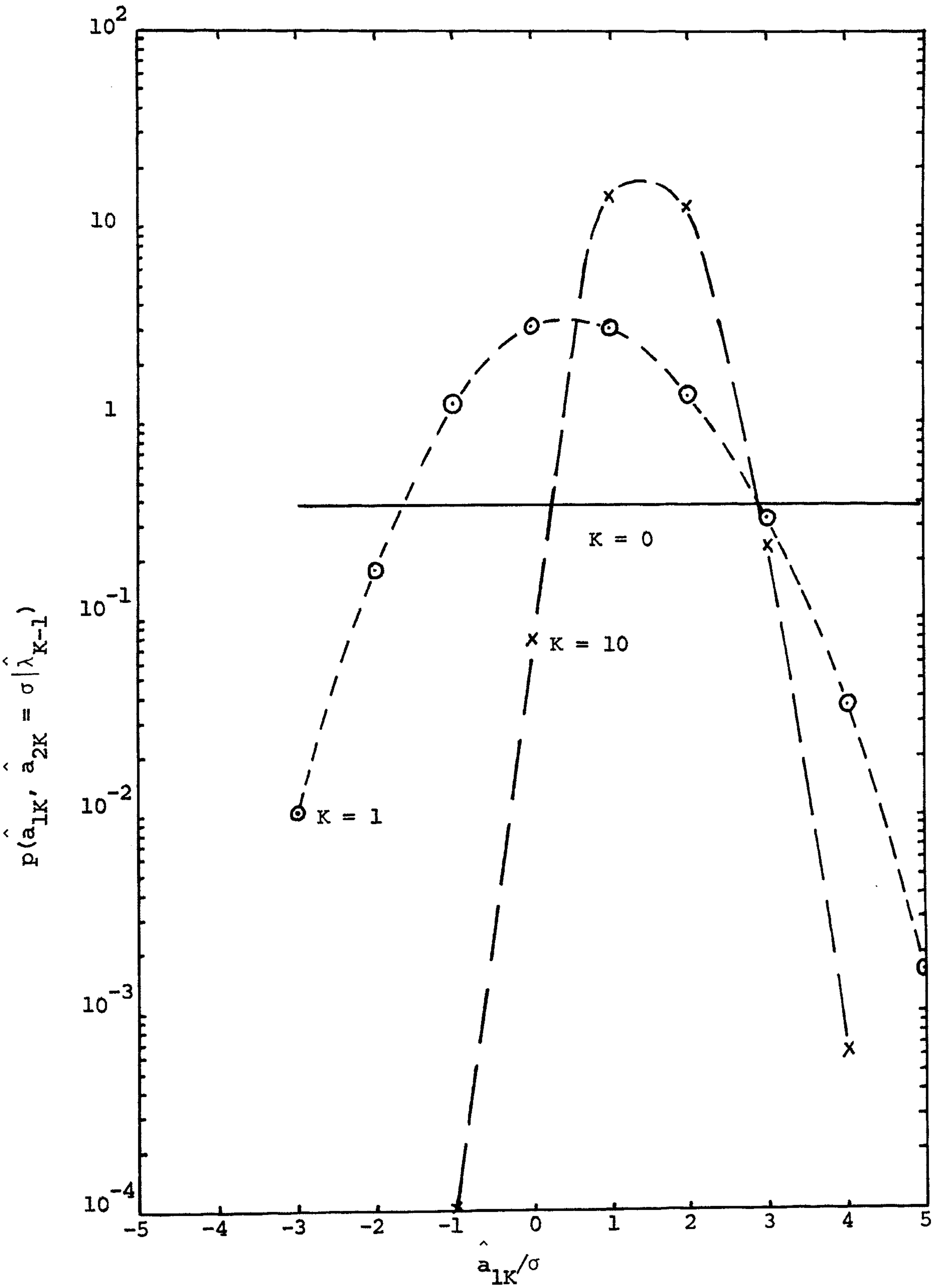


Figure 8(a): Learned PDF - Selective Specular Channel. $L = 2$, $\alpha_1 = \alpha_2$, $a_{2K} = \sigma$, $\beta\gamma^2 = 2$

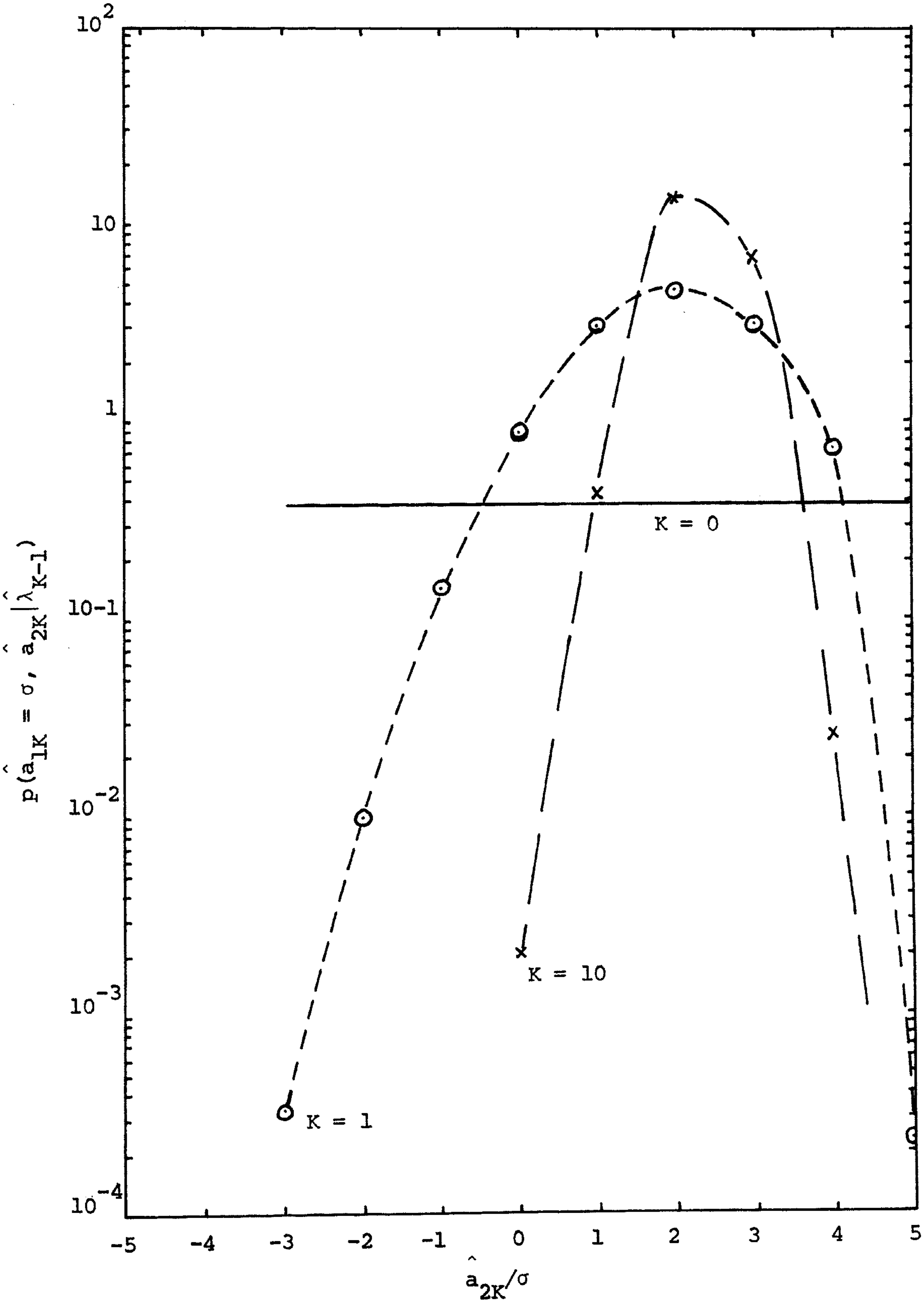


Figure 8(b): Learned PDF - Selective Specular Channel. $L = 2$, $\alpha_1 = \alpha_2$, $\hat{a}_{1K} = \sigma$, $\beta\gamma^2 = 2$

Equations (76) and (77) [4]. His error probability plots in this and later papers [5, 6] employed the quantity $\gamma^2/2$. This may have been what led Lindsey [7] to redefine γ^2 as one-half of the original quantity when he analyzed the noncoherent and coherent Rician channel receiver. Van Trees [17], however, even though referencing Lindsey, reverted back to Turin's definition of γ, but still plotted versus $\gamma^2/2$. His curves of error probability, incidentally, are mislabelled on the abscissa as β when, in fact, it should be $\beta(1 + \gamma^2/2)$.

Figure 9 illustrates the learning receivers performance in a non-selective Rician channel as compared with the optimum coherent system (solid lines). The learning receiver (dashed lines) is seen to improve on what is already optimum! This can be explained by pointing out that the coherent receiver is designed to be optimum for a channel whose parameters are essentially independent from one observation to the next. The learning receiver, on the other hand, makes use of any knowledge it can gain as it receives each observation. When the observations are partially correlated (b = .1 here) the receiver must be redesigned to account for it. The curve labeled $\gamma = \infty$ is the non-fading case.

Results for the selective Rician channel, in which the direct specular and indirect diffuse components are in separate paths, are given in Figure 10. No solid curve is shown as the writer was unable to find any published performance estimates for the coherent-diversity Rician-channel receiver. A comparison of the selective and non-selective performance is shown. The improved performance with channel diversity is well known [7, 8] and the learning receiver is no exception. A curve for b = .707 is also shown which indicates the learning

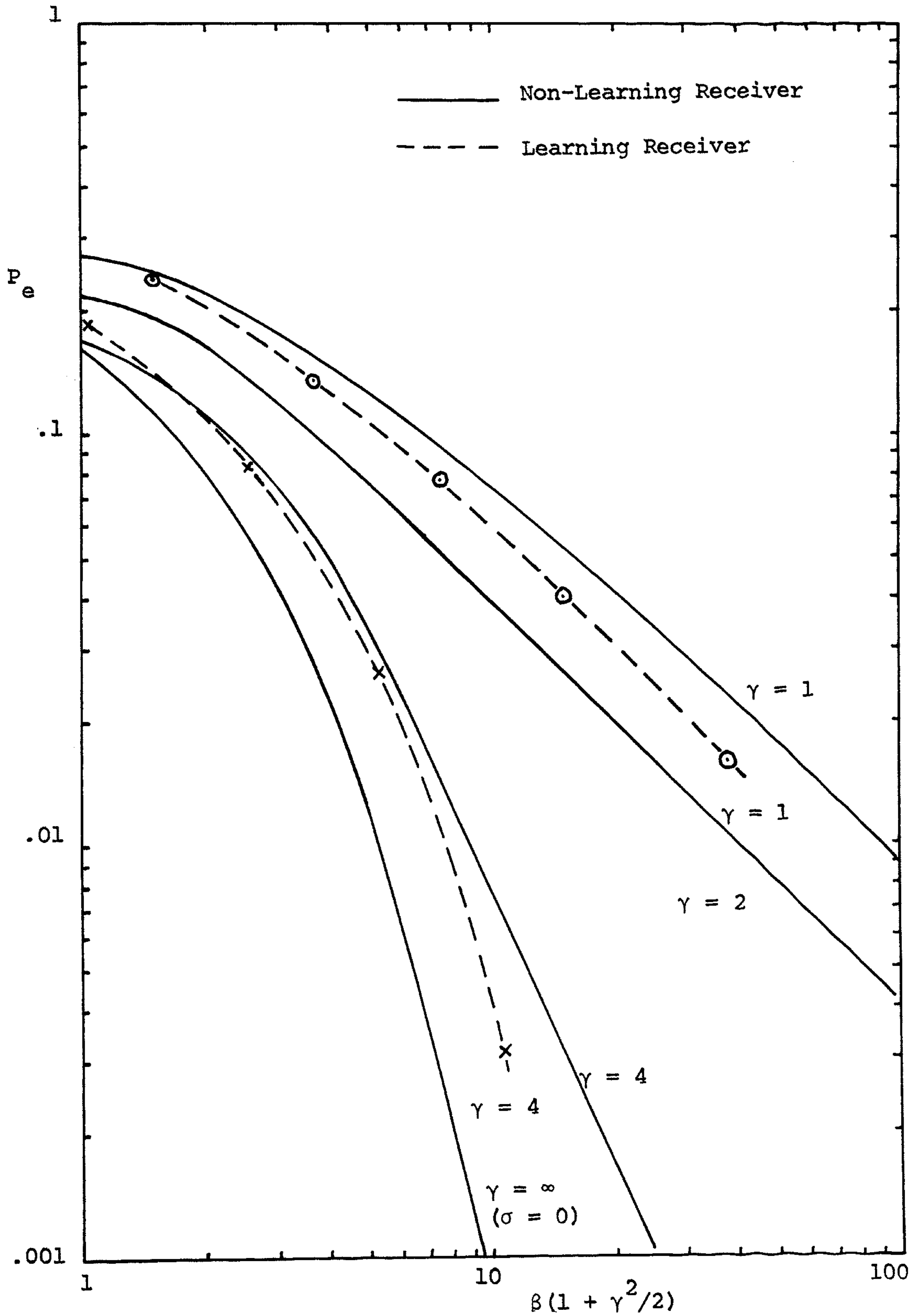


Figure 9: Nonselective Rician Channel Performance
b = 0.1

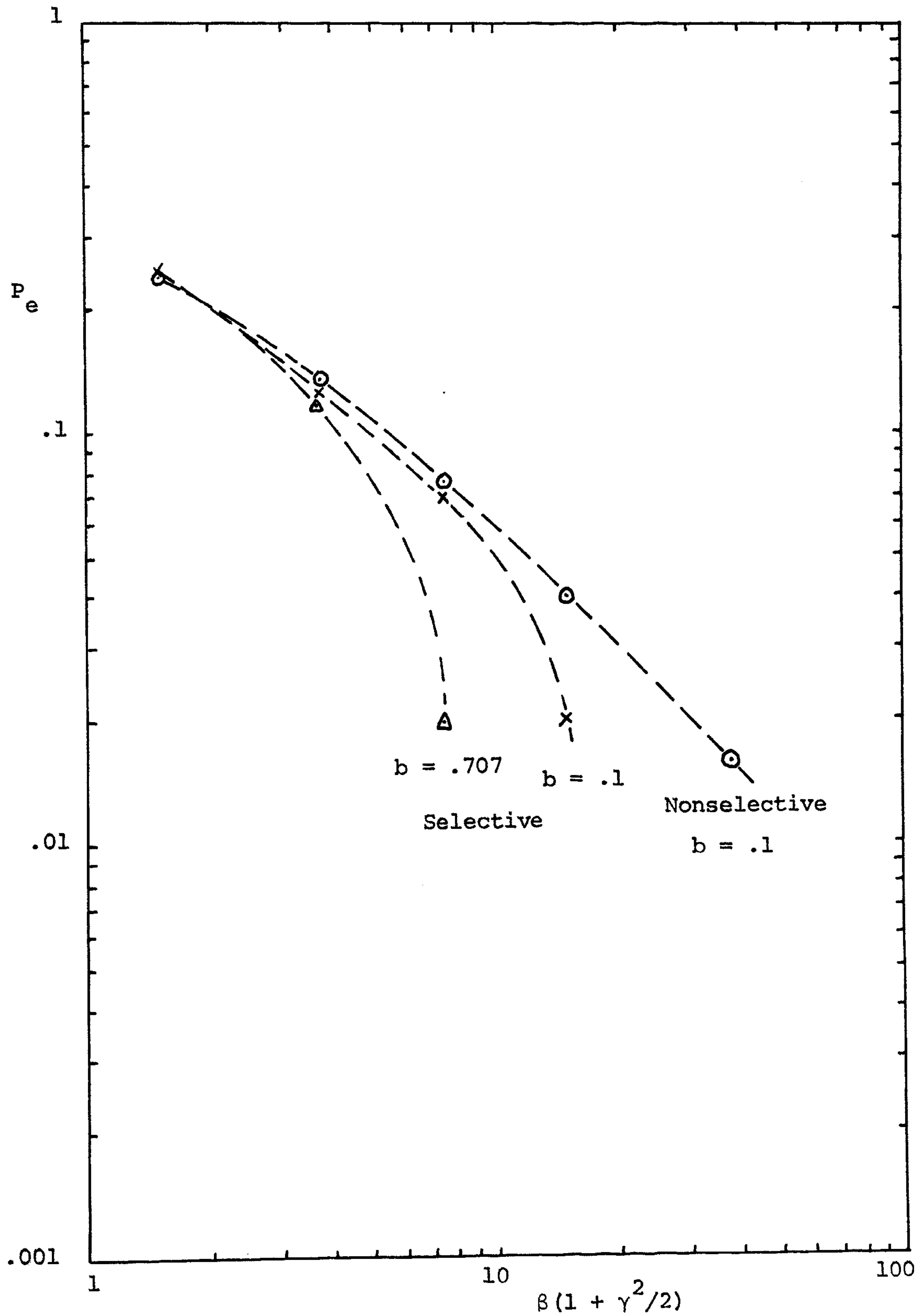


Figure 10: Performance Comparison - Selective and Nonselective Rician Channels $\gamma = 1$

improvement with increased observation correlation.

The relative performance for the specular reflective channel is shown in Figures 11 and 12. The performance of the learning receiver is generally between that of a coherent and a noncoherent system. The coherent system, in this case, implies a completely known signal. This would imply that $b = 1$. For a Gauss-Markov dependent channel, however, values of b less than 1 suggest a slight fading component which will degrade performance. A value of b close to 1 was run in the simulation. This curve is shown ($b = .999$) and it seems to indicate an improvement in performance over the optimum system. This slight discrepancy may be accounted for by the limited number of observations used in determining the error probability for the specular channel. In any Monte Carlo simulation the number of trials determines the accuracy of the results. The data presented here is merely for the purpose of indicating the trend in performance. Naturally, had the results been in the other direction they would have been less appealing.

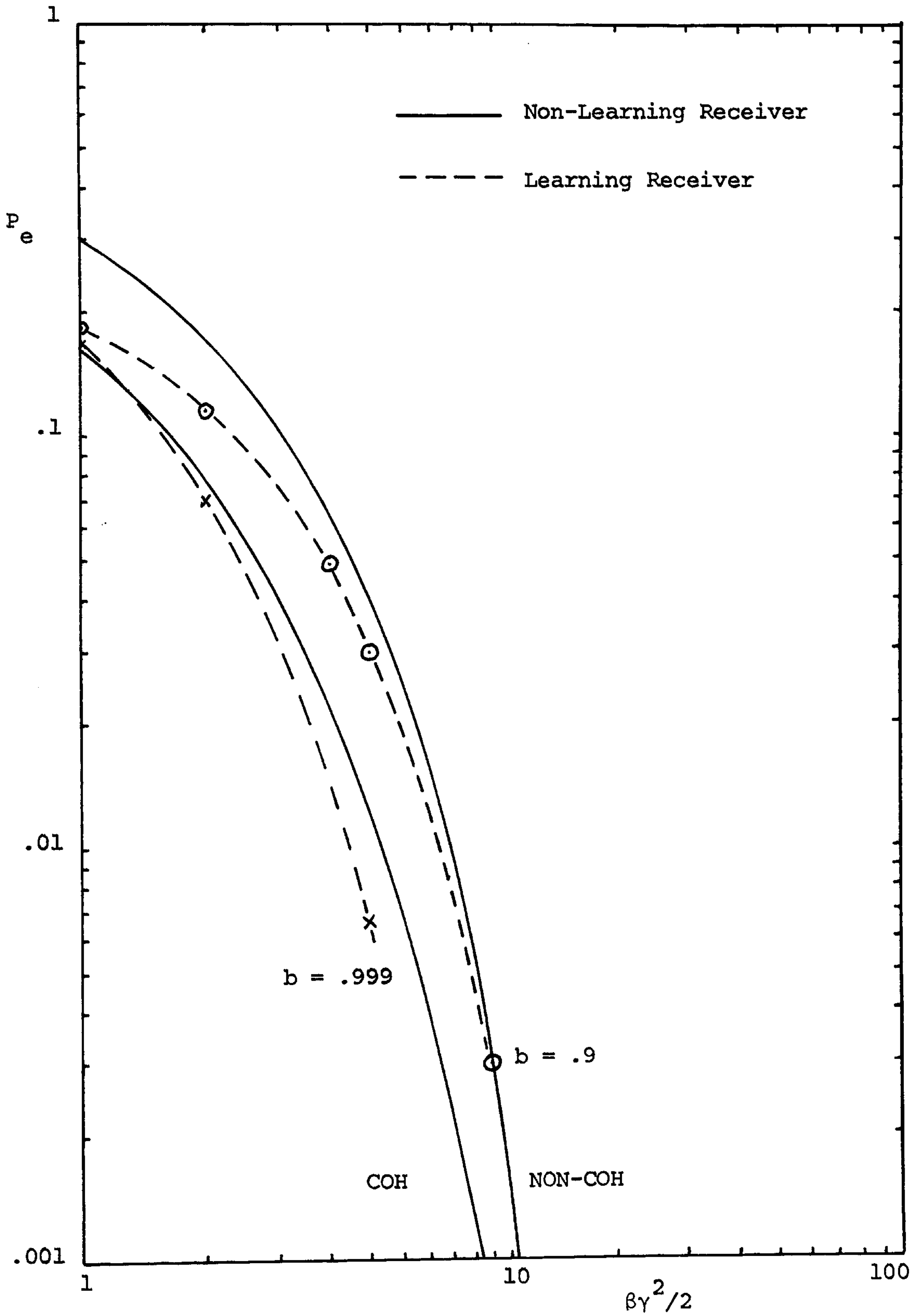


Figure 11: Nonselective Specular Channel Performance

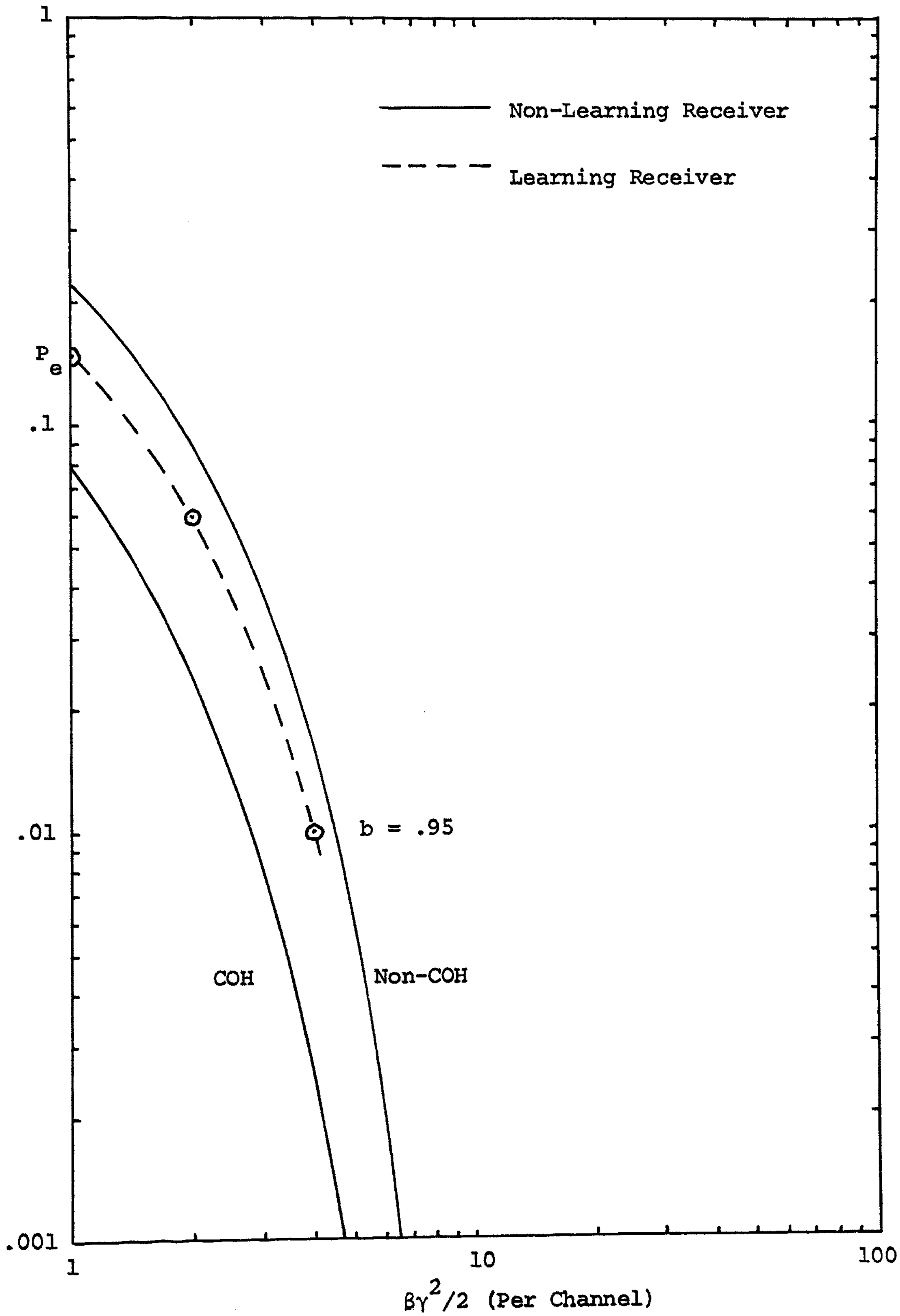


Figure 12: Two Ray Selective Specular Channel Performance. $\alpha_1 = \alpha_2$

CHAPTER VII

CONCLUSIONS AND EXTENSIONS

A. Conclusions

A receiver has been designed for communications in three-component multipath channels based on the theory of machines that learn without a teacher. As a step toward this design the multipath channel was modelled in terms of quadrature gain components. The only knowledge required by the receiver is the value dependence of the channel from one observation to the next (Gauss-Markov assumed), the possible signals transmitted, their prior probabilities, and the receiver noise statistics. Based on certain simplifying assumptions, this unsupervised learning receiver was modelled on the digital computer and a Monte Carlo simulation was performed to determine an estimate of its error rate performance. It was then compared with the published performance curves of some previously designed coherent and noncoherent receivers for Rician and nonfading channels.

Both frequency selective and nonselective channels were analyzed. The learning receiver appears to improve on the performance of the "optimum" systems as the observation correlation increases. This is a reasonable result as the optimum designs are based on independent observations. According to the theory of unsupervised learning machines [9], the receiver that learns without a teacher should converge in performance to the optimum system (which is designed for the given conditions) as the number of observations increases.

While the learning receiver appears to improve on performance of existing systems, its principal advantage is that it is not dependent on statistical knowledge of the channel, as are presently designed systems. Whatever type of channel model is employed, this receiver will learn the probability density functions of its parameters, conditioned on past observations, if the *a priori* probability density function does not exclude possible values of the parameters.

B. Suggestions for Further Work

The original concept of learning machines was primarily oriented toward the pattern recognition problem. This dissertation extends the application of unsupervised learning systems to the well studied problem of multipath interference. The particular channel models analyzed are Rician and nonfading. Further study could include such non-gaussian applications as laser communications.

One important problem which requires considerable research is the application of the techniques described herein to the design of clutter rejection radar systems. With the advent of the Kalman filter, adaptive radar systems have recently come into existence. An unsupervised learning radar would be an original research topic worthy of investigation.

BIBLIOGRAPHY

1. Price, R., "The Detection of Signals Perturbed by Scatter and Noise," Trans. IRE, PGIT-4 (Sept., 1954), 163-170.

2. Price, R., "Optimum Detection of Random Signals in Noise, with Applications to Scatter-Multipath Communication, I," Trans. IRE, Vol. IT-2 (Dec., 1956), 125-135.

3. Price, R. and Green, P. E., Jr., "A Communication Technique for Multipath Channels," Proc. IRE, Vol. 46 (Mar., 1958), 555-570.

4. Turin, G. L., "Communication Through Noisy, Random-Multipath Channels," IRE Convention Record, Vol. 4, Pt. 4 (1956), 154-166.

5. Turin, G. L., "Error Probabilities for Binary Symmetric Ideal Reception through Nonselective Slow Fading and Noise, "Proc. IRE, Vol. 46 (Sept., 1958), 1603-1619.

6. Turin, G. L., "Some Computations of Error Rates for Selectively Fading Multipath Channels," Proc. NEC, Vol. 15 (1959), 431-440.

7. Lindsey, W. C., "Error Probabilities for Rician Fading Multi-Channel Reception of Binary and N-ary Signals," IEEE Trans., Vol. IT-10 (Oct., 1964), 339-350.

8. Jones, J. Jay, "Multipath FSK and DPSK Reception with Three-Component Multipath," IEEE Trans., Vol, COM-16 (Dec., 1968), 808-821.

9. Fralick, S. C., "Learning to Recognize Patterns without a Teacher," Stanford Electronics Lab Rept. SEL-65-011 (TR6103-10), Stanford University, Stanford, Calif. (Mar., 1965)

10. Hilborn, C. G. and Lainiotis, D. G., "Recursive Computations for the Optimal Tracking of Time-Varying Parameters," IEEE Trans., Vol. IT-14 (May, 1968), 514-515.

11. Abramson, N. and Braverman, D. J., "Learning to Recognize Patterns in a Random Environment," Trans. IRE, Vol. IT-8 (Sep., 1962), 58-63.

12. Spragins, J., "Learning Without a Teacher," IEEE Trans., Vol. IT-12 (Apr., 1966), 223-230.

13. Daly, R. F., "Adaptive Binary Detectors," Stanford Electronics Lab., Standord, Calif. TR2003-2 (Jun., 1961).

14. Keehn, D. G., "A Note on Learning for Gaussian Properties," IEEE Trans., Vol. IT-11 (Jan., 1965), 126-132.

15. Scudder, H. J., "Adaptive Communication Receivers," IEEE Trans. Vol. IT-11 (Apr., 1965), 167-174.

16. Woodward, P. M., Probability and Information Theory With Applications to Radar, New York: Pergamon, 1964.

17. Van Trees, H. L., Detection, Estimation, and Modulation Theory, Part I, New York: John Wiley, 1968.

18. Helstrom, C. W., Statistical Theory of Signal Detection, 2 Ed., New York: Pergamon, 1968.

19. Woodward, P. M. and Davies, I. L., Proc. IEEE, Pt. III 99, 37 (1952).

20. Sage, A. P. and Melsa, J. L., Estimation Theory with Applications to Communications and Control, New York: McGraw-Hill, 1971.

VITA

Richard Paul Brueggemann was born on March 11, 1939 in St. Louis, Missouri. He received his primary and secondary education in St. Louis. He received his college education from the University of Missouri-Rolla, in Rolla, Missouri and Washington University in St. Louis. He received the Bachelor of Science and Master of Science degrees in Electrical Engineering from the University of Missouri-Rolla in January 1961 and May 1964, respectively.

Mr. Brueggemann has been employed by Emerson Electric Company in St. Louis and McDonnell Aircraft Company in St. Louis in the field of radar systems analysis. He has also been employed as a Graduate Instructor of Electrical Engineering by the University of Missouri-Rolla.

He has been enrolled in the Graduate School of the University of Missouri-Rolla since September 1969. He is a member of IEEE, Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi, and an associate member of Sigma Xi.

APPENDIX A

DERIVATION OF THE LIKELIHOOD FUNCTION FOR $\underline{Z}$

For Equations (5) and (21) it is desired to compute the inner products $(\hat{\underline{Z}}-\hat{\underline{Y}})^t(\hat{\underline{Z}}-\hat{\underline{Y}})$ and $(\tilde{\underline{Z}}-\tilde{\underline{Y}})^t(\tilde{\underline{Z}}-\tilde{\underline{Y}})$. Using Equation (15) this is done as follows:

$$(\hat{\underline{Z}}-\hat{\underline{Y}})^t(\hat{\underline{Z}}-\hat{\underline{Y}}) = \{\hat{\underline{Z}}-\sum_{\ell=0}^{L}[\hat{a}_\ell\hat{\underline{X}}_\ell+\tilde{a}_\ell\tilde{\underline{X}}_\ell]\}^t\{\hat{\underline{Z}}-\sum_{\ell=0}^{L}[\hat{a}_\ell\hat{\underline{X}}_\ell+\tilde{a}_\ell\tilde{\underline{X}}_\ell]\}$$

$$= \hat{\underline{Z}}^t\hat{\underline{Z}} - 2\hat{\underline{Z}}^t\sum_{\ell=0}^{L}[\hat{a}_\ell\hat{\underline{X}}_\ell+\tilde{a}_\ell\tilde{\underline{X}}_\ell]$$

$$+\sum_{\ell=0}^{L}\sum_{m=0}^{L}[\hat{a}_\ell\hat{a}_m\hat{\underline{X}}_\ell^t\hat{\underline{X}}_m+\hat{a}_\ell\tilde{a}_m\hat{\underline{X}}_\ell^t\tilde{\underline{X}}_m+\tilde{a}_\ell\hat{a}_m\tilde{\underline{X}}_\ell^t\hat{\underline{X}}_m+\tilde{a}_\ell\tilde{a}_m\tilde{\underline{X}}_\ell^t\tilde{\underline{X}}_m]. \tag{A.1}$$

Performing a similar operation for the quadrature term and then combining with Equation (A.1) results in

$$([\underline{Z}-\underline{Y}],[\underline{Z}-\underline{Y}]) = (\hat{\underline{Z}}-\hat{\underline{Y}})^t(\hat{\underline{Z}}-\hat{\underline{Y}}) + (\tilde{\underline{Z}}-\tilde{\underline{Y}})^t(\tilde{\underline{Z}}-\tilde{\underline{Y}})$$

$$= \hat{\underline{Z}}^t\hat{\underline{Z}} + \tilde{\underline{Z}}^t\tilde{\underline{Z}} - 2\sum_{\ell=0}^{L}[\hat{a}_\ell(\hat{\underline{Z}}^t\hat{\underline{X}}_\ell + \tilde{\underline{Z}}^t\tilde{\underline{X}}_\ell) + \tilde{a}_\ell(\hat{\underline{Z}}^t\tilde{\underline{X}}_\ell - \tilde{\underline{Z}}^t\hat{\underline{X}}_\ell)]$$

$$+\sum_{\ell=0}^{L}\sum_{m=0}^{L}(\hat{a}_\ell\hat{a}_m + \tilde{a}_\ell\tilde{a}_m)(\hat{\underline{X}}_\ell^t\hat{\underline{X}}_m + \tilde{\underline{X}}_\ell^t\tilde{\underline{X}}_m)\ . \tag{A.2}$$

The last set of inner products in Equation (A.2) can be shown to be the real autocorrelation function of the sounding signal as follows:

$$(\underline{X}_\ell,\underline{X}_m) = W\int_0^{T'} X^*(t-\tau_\ell)X(t-\tau_m)\,dt$$

$$= W\int_0^{T'} [\hat{X}(t-\tau_\ell)\hat{X}(t-\tau_m) + \tilde{X}(t-\tau_\ell)\tilde{X}(t-\tau_m)]\,dt$$

$$+ jW\int_0^{T'} [\hat{X}(t-\tau_\ell)\tilde{X}(t-\tau_m) - \tilde{X}(t-\tau_\ell)\hat{X}(t-\tau_m)]\,dt. \tag{A.3}$$

Equation (A.3) is seen to be the complex autocorrelation of $X(t)$ evaluated at $\tau_m-\tau_\ell$, defined by $F(\tau_m-\tau_\ell) = \hat{F}(\tau_m-\tau_\ell) + j\tilde{F}(\tau_m-\tau_\ell)$, so

$$\underline{\hat{X}}_\ell^t\underline{\hat{X}}_m + \underline{\tilde{X}}_\ell^t\underline{\tilde{X}}_m = \mathrm{Re}[(\underline{X}_\ell,\underline{X}_m)] = W\,\hat{F}(\tau_m-\tau_\ell) \;. \tag{A.4}$$

Now, since $\hat{F}(0) = 2E$, and using the resolvability condition, Equation (11), it is seen that

$$\hat{F}(\tau_m-\tau_\ell) << 2E \quad, m \neq \ell. \tag{A.5}$$

Therefore the off-diagonal terms in Equation (A.2) are negligible. Using this result and Equation (A.2) in Equations (21) and (5) the likelihood function for $\underline{Z}$ is written

$$p[\underline{Z}|(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}] = (2\pi W_N N_0)^{-T'W_N}\exp[\;-\frac{\underline{\hat{Z}}^t\underline{\hat{Z}} + \underline{\tilde{Z}}^t\underline{\tilde{Z}}}{2W_N N_0}\;]$$

$$\cdot\prod_{\ell=0}^{L}\exp\{\;\frac{\hat{a}_\ell}{W_N N_0}[\underline{\hat{Z}}^t\underline{\hat{X}}_\ell + \underline{\tilde{Z}}^t\underline{\tilde{X}}_\ell] + \frac{\hat{a}_\ell}{W_N N_0}[\underline{\hat{Z}}^t\underline{\tilde{X}}_\ell - \underline{\tilde{Z}}^t\underline{\hat{X}}_\ell] - (\hat{a}_\ell^2+\tilde{a}_\ell^2)\frac{W\,E}{W_N N_0}\;\}. \tag{A.6}$$

This can be factored according to the quadrature components of the channel gain.

$$p[\underline{Z}|(\hat{a}_\ell),(\tilde{a}_\ell),(\tau_\ell),\underline{X}] = \hat{p}[\underline{Z}|(\hat{a}_\ell),(\tau_\ell),\underline{X}]\tilde{p}[\underline{Z}|(\tilde{a}_\ell),(\tau_\ell),\underline{X}] \qquad (A.7)$$

Defining the complex cross-correlation, G, between $\underline{Z}$ and $\underline{X}$ by the inner product

$$G = (\underline{Z},\underline{X})/W_N$$

$$= \frac{1}{W_N}[\hat{\underline{Z}}^t\hat{\underline{X}} + \tilde{\underline{Z}}^t\tilde{\underline{X}}) + j(\hat{\underline{Z}}^t\tilde{\underline{X}} - \tilde{\underline{Z}}^t\hat{\underline{X}})]$$

$$= \hat{G} + j\tilde{G}, \qquad (A.8)$$

the factors in Equation (A.7) are written as

$$\hat{p}[\underline{Z}|(\hat{a}_\ell),(\tau_\ell),\underline{X}] = (2\pi W_N N_0)^{-T'W_N/2}\exp[-\frac{\hat{\underline{Z}}^t\hat{\underline{Z}}}{2W_N N_0}]\prod_{\ell=0}^{L}\exp[\frac{\hat{G}_\ell}{N_0}\hat{a}_\ell - \frac{W_N E}{W_N N_0}\hat{a}_\ell^2] \qquad (A.9)$$

$$\tilde{p}[\underline{Z}|(\tilde{a}_\ell),(\tau_\ell),\underline{X}] = (2\pi W_N N_0)^{-T'W_N/2}\exp[-\frac{\tilde{\underline{Z}}^t\tilde{\underline{Z}}}{2W_N N_0}]\prod_{\ell=0}^{L}\exp[\frac{\tilde{G}_\ell}{N_0}\tilde{a}_\ell - \frac{W_N E}{W_N N_0}\tilde{a}_\ell^2].$$

APPENDIX B

DERIVATION OF THE A POSTERIORI PDF

Substituting Equations (22) and (23) into Equation (20) gives

$$p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell),\underline{Z},\underline{X}] = \frac{(2\pi W_N N_0)^{-T'W_N}}{p[\underline{Z}\,|\,(\tau_\ell),\underline{X}]} \exp\left[-\frac{\hat{\underline{Z}}^t\hat{\underline{Z}} + \tilde{\underline{Z}}^t\tilde{\underline{Z}}}{2W_N N_0}\right]$$

$$\cdot \prod_{\ell=0}^{L} \frac{1}{2\pi\sigma_\ell^2} \exp\left[-\frac{(\hat{a}_\ell-\hat{\alpha}_\ell)^2+(\tilde{a}_\ell-\tilde{\alpha}_\ell)^2}{2\sigma_\ell^2} - \frac{2WE(\hat{a}_\ell^2+\tilde{a}_\ell^2)-2\hat{a}_\ell W_N\hat{G}_\ell-2\tilde{a}_\ell W_N\tilde{G}_\ell}{2W_N N_0}\right], \tag{B.1}$$

where the definition of Equation (A.8) has been used. In order that Equation (B.1) be a pdf it must be shown that

$$\int_{(\hat{a}_\ell)} \int_{(\tilde{a}_\ell)} p[(\hat{a}_\ell),(\tilde{a}_\ell)\,|\,(\tau_\ell),\underline{Z},\underline{X}]\; d(\hat{a}_\ell)\; d(\tilde{a}_\ell) = 1. \tag{B.2}$$

The integrations over the gradrature components can be performed independently. The following integral is evaluated:

$$\int_{-\infty}^{\infty} \exp\left[-\frac{(a_\ell-\alpha_\ell)^2}{2\sigma_\ell^2} - \frac{2WEa_\ell^2-2a_\ell W_N G_\ell}{2W_N N_0}\right] da_\ell$$

$$= \int_{-\infty}^{\infty} \exp\left[-\frac{a_\ell^2}{2}\left(\frac{2WE}{W_N N_0} + \frac{1}{\sigma_\ell^2}\right) + \left(\frac{G_\ell}{N_0} + \frac{\alpha_\ell}{\alpha_\ell^2}\right) a_\ell - \frac{\alpha_\ell^2}{2\alpha_\ell^2}\right] da_\ell$$

$$= \sqrt{2\pi(\sigma_\ell')^2}\; \exp\left[\frac{1}{2}\left(\frac{\alpha_\ell'}{\sigma_\ell'}\right)^2\right]\; \exp\left[-\frac{\alpha_\ell^2}{2\sigma_\ell^2}\right], \tag{B.3}$$

where

$$(\sigma_\ell')^2 = \left(\frac{2WE}{W_N N_0} + \frac{1}{\sigma_\ell^2}\right)^{-1}$$

$$\alpha_\ell' = (\sigma_\ell')^2\left(\frac{G_\ell}{N_0} + \frac{\alpha_\ell}{\sigma_\ell^2}\right). \tag{B.4}$$

Defining

$$C = \frac{(2\pi W_N N_0)^{-T' W_N}}{p[\underline{Z}|(\tau_\ell),\underline{X}]} \exp\left[-\frac{\underline{\hat{Z}}^t\underline{\hat{Z}} + \underline{\tilde{Z}}^t\underline{\tilde{Z}}}{2W_N N_0}\right](2\pi\sigma_\ell^2)^{-\frac{L+1}{2}}, \tag{B.5}$$

and setting the integral of Equation (B.1) to unity, it is easily shown using Equation (B.3) that

$$p[(\hat{a}_\ell),(\tilde{a}_\ell)|(\tau_\ell),\underline{Z},\underline{X}] = \prod_{\ell=0}^{L} [2\pi(\sigma_\ell')^2]^{-1} \exp\left[-\frac{1}{2}\left(\frac{2WE}{W_N N_0} + \frac{1}{\sigma_\ell^2}\right)(\hat{a}_\ell^2+\tilde{a}_\ell^2)\right.$$

$$\left. + \left(\frac{\hat{G}_\ell}{N_0} + \frac{\hat{\alpha}_\ell}{\sigma_\ell^2}\right)\hat{a}_\ell + \left(\frac{\tilde{G}_\ell}{N_0} + \frac{\tilde{\alpha}_\ell}{\sigma_\ell^2}\right)\tilde{a}_\ell - \frac{1}{2}\left(\frac{\hat{\alpha}_\ell'}{\sigma_\ell'}\right)^2 - \frac{1}{2}\left(\frac{\tilde{\alpha}_\ell'}{\sigma_\ell'}\right)^2\right]. \tag{B.6}$$

Using the definitions in Equation (25) and factoring Equation (B.6) gives

$$p[(\hat{a}_\ell)|(\tau_\ell),\underline{Z},\underline{X}] = \prod_{\ell=0}^{L} \frac{1}{\sqrt{2\pi(\sigma_\ell')^2}} \exp\left[-\frac{(\hat{a}_\ell-\hat{\alpha}_\ell')^2}{2(\sigma_\ell')^2}\right]$$

$$p[(\tilde{a}_\ell)|(\tau_\ell),\underline{Z},\underline{X}] = \prod_{\ell=0}^{L} \frac{1}{\sqrt{2\pi(\sigma_\ell')^2}} \exp\left[-\frac{(\tilde{a}_\ell-\tilde{\alpha}_\ell')^2}{2(\sigma_\ell')^2}\right]. \tag{B.7}$$

APPENDIX C

CONDITIONAL MAP TEST: MULTIPLE OBSERVATIONS

The average cost or risk for the N hypotheses - single measurement case is

$$\overline{C} = \sum_{i=1}^{N} \sum_{j=1}^{N} P_j C_{ij} \int_{R_i} p_j(\underline{V})\,d\underline{V}\ , \tag{C.1}$$

where P_j is the a priori signaling probability, C_{ij} is the cost of choosing hypothesis H_i when actually H_j is true and R_i is the region in the domain of $\underline{V}$ where H_i is considered to be true. The Bayes optimum test simply varies the R_i to minimize $\overline{C}$. When there are multiple observations, to take advantage of all previous experience the observation sequence $\lambda_K = \underline{V}_K, \underline{V}_{K-1}, \ldots, \underline{V}_1$ is used in place of $\underline{V}$ in Equation (C.1). The integration, then, is taken over a region defined by $[R_i]_K$, a matrix extension of R_i. Interchanging the integration and summation over j, $\overline{C}$ becomes

$$\overline{C} = \sum_{i=1}^{N} \int_{[R_i]_K} \{ \sum_{j=1}^{N} C_{ij} P_j p_j(\lambda_K) \} d(\lambda_K), \tag{C.2}$$

where $d(\lambda_K) \equiv d\underline{V}_K d\underline{V}_{K-1} \cdot \cdot \cdot d\underline{V}_1$.

If the cost assignment is determined by $C_{ij} = 1 - \delta_{ij}$ where δ_{ij} is the Kronecker delta, then $\overline{C}$ is the probability of error

$$P_e = \sum_{i=1}^{N} \int_{[R_i]_K} \{ \sum_{j \neq i}^{N} P_j p_j(\lambda_K) \} d(\lambda_K). \tag{C.3}$$

The summation is easily seen to be

$$\sum_{j\neq i}^{N} P_j p_j(\lambda_K) = p(\lambda_K) - P_i p_i(\lambda_K). \tag{C.4}$$

Now the error probability is

$$P_e = 1 - \sum_{i=1}^{N} \int_{[R_i]_K} P_i p_i(\lambda_K)\, d(\lambda_K). \tag{C.5}$$

The λ_K will be included in only one integral, therefore it should be assigned to the region $[R_i]$ where it will make the smallest contribution to P_e. This is done by choosing the largest $P_i p_i(\lambda_K)$. Maximizing this quantity as it stands implies waiting for all of the data in the sequence to be received and then performing the computation, followed by a decision. A computation can be performed on each new observation and a decision made which is Bayes optimal. Note that

$$\begin{aligned} p_i(\lambda_K) &= p_i(\underline{V}_K, \lambda_{K-1}) \\ &= p_i(\underline{V}_K | \lambda_{K-1}) p(\lambda_{K-1}). \end{aligned} \tag{C.6}$$

Substituting this into Equation (C.5)

$$P_e = 1 - \sum_{i=1}^{N} \int_{[R_i]_{K-1}} [\int_{R_i} P_i p_i(\underline{V}_K | \lambda_{K-1}) d\underline{V}_K] p(\lambda_{K-1}) d(\lambda_{K-1}). \tag{C.7}$$

It is seen from this that P_e is minimized by choosing the R_i for which $P_i p_i(\underline{V}_K | \lambda_{K-1})$ is largest since $p(\lambda_{K-1})$ is independent of H_i.

If the P_i are unknown then Equation (C.6) is replaced by

$$P_i p_i(\lambda_K) = p(\lambda_K, H_i)$$

$$= p_i(\underline{V}_K|\lambda_{K-1}) P(H_i|\lambda_{K-1}) p(\lambda_{K-1}), \tag{C.8}$$

and the Bayes optimum system will compute $P(H_i|\lambda_{K-1}) p_i(\underline{V}_K|\lambda_{K-1})$ and choose H_i for which it is the largest.

APPENDIX D

DERIVATION OF EQUATION (35)

According to the rules of conditional probability:

$$p_i(\underline{V}_K|\lambda_{K-1}) = \frac{p_i(\underline{V}_K,\lambda_{K-1})}{p(\lambda_{K-1})} . \tag{D.1}$$

The conditional joint density in the numerator can be found by integrating $\underline{\theta}_i$ out on the conditional joint pdf $p_i(\underline{V}_K,\lambda_{K-1},\underline{\theta}_i)$. Hence

$$p_i(\underline{V}_K|\lambda_{K-1}) = \frac{1}{p(\lambda_{K-1})} \int_{\underline{\theta}_i} p_i(\underline{V}_K,\lambda_{K-1},\underline{\theta}_i)d\underline{\theta}_i . \tag{D.2}$$

The integrand can be written

$$p_i(\underline{V}_K,\lambda_{K-1},\underline{\theta}_i) = p_i(\underline{V}_K|\lambda_{K-1},\underline{\theta}_i)p(\lambda_{K-1},\underline{\theta}_i) . \tag{D.3}$$

When divided by $p(\lambda_{K-1})$ there results

$$p_i(\underline{V}_K|\lambda_{K-1}) = \int_{\underline{\theta}_i} p_i(\underline{V}_K|\lambda_{K-1},\underline{\theta}_i)p(\underline{\theta}_i|\lambda_{K-1})d\underline{\theta}_i . \tag{D.4}$$

Now $\underline{\theta}_i$ is assumed to be the only unknown parameter and assuming that the $\underline{V}_K$ are independent conditioned on $\underline{\theta}_i$ then

$$p_i(\underline{V}_K|\lambda_{K-1},\underline{\theta}_i) = p_i(\underline{V}_K|\underline{\theta}_i) . \tag{D.5}$$

Substituting Equation (D.5) into (D.4) results in Equation (35).

APPENDIX E

DERIVATION OF EQUATION (43)

The conditional probability density function in the integrand of Equation (42) can be modified as follows:

$$p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1} | \lambda_{K-1}) = \frac{p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1}, \lambda_{K-1})}{p(\lambda_{K-1})}$$

$$= \frac{p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1}, \underline{V}_{K-1} | \lambda_{K-2}) p(\lambda_{K-2})}{p(\underline{V}_{K-1} | \lambda_{K-2}) p(\lambda_{K-2})}$$

$$= \frac{p(\underline{V}_{K-1} | \underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1}, \lambda_{K-2}) p(\underline{\theta}_K, \ldots, \theta_{K-M+1} | \lambda_{K-2})}{p(\underline{V}_{K-1} | \lambda_{K-2})} . \tag{E.1}$$

Now, invoking the conditional independence of the $\underline{V}_K$ (conditioned on $\underline{\theta}_K$) and recognizing that, given the parameter vector the previous observations are unnecessary, the following is true:

$$p(\underline{V}_{K-1} | \underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1}, \lambda_{K-2}) = p(\underline{V}_{K-1} | \underline{\theta}_{K-1}) . \tag{E.2}$$

So Equation (E.1) becomes

$$p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1} | \lambda_{K-1}) = \frac{p(\underline{V}_{K-1} | \underline{\theta}_{K-1})}{p(\underline{V}_{K-1} | \lambda_{K-2})} p(\underline{\theta}_K, \ldots, \underline{\theta}_{K-M+1} | \lambda_{K-2}) . \tag{E.3}$$

This is not yet a recursive relation as the right hand side needs $p(\underline{\theta}_{K-1}, \ldots, \underline{\theta}_{K-M} | \lambda_{K-2})$ in it.

Observe that

$$p(\underline{\theta}_K, \ldots, \theta_{K-M+1} | \lambda_{K-2}) = \int p(\underline{\theta}_K, \ldots, \theta_{K-M+1}, \underline{\theta}_{K-M} | \lambda_{K-2}) d\underline{\theta}_{K-M} . \tag{E.4}$$

The integrand here is, according to Bayes' rule,

$$p(\underline{\theta}_K,\ldots,\underline{\theta}_{K-M+1},\underline{\theta}_{K-M}|\lambda_{K-2})$$

$$= p(\underline{\theta}_{K-1},\ldots,\underline{\theta}_{K-M}|\lambda_{K-2})p(\underline{\theta}_K|\underline{\theta}_{K-1},\ldots,\underline{\theta}_{K-M}). \qquad \text{(E.5)}$$

This is the Markov-M dependence relationship needed. Substituting Equation (E.5) into (E.4) and this result into Equation (E.3) results in Equation (43).