ADAPTIVE FILTERING TECHNIQUES
FOR
TONE AIDED TRANSMISSION SYSTEMS

by

Henry W. H. Li
B.A.Sc., University of Waterloo, 1985

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF APPLIED SCIENCE

in the School
of
Engineering Science

October 1989

ISBN 0-315-66177-1

APPROVAL

NAME : Henry Wai Hung Li

DEGREE : Master of Applied Science (Engineering Science)

TITLE OF THESIS : Adaptive Filtering Techniques For Tone Aided Transmission Systems

EXAMINING COMMITTEE:

Chairman : Dr. Vladimir Cuperman

Dr. James K. Cavers
Senior Supervisor

Dr. Paul Ho

Dr. Mehrdad Saif

DATE APPROVED : [illegible] 13th 1989

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Project/Extended Essay

ADAPTIVE FILTERING TECHNIQUES FOR TONE AIDED

TRANSMISSION SYSTEMS

Author: ______________________

(signature)

HENRY W. H. LI

(name)

OCT 17th 1989

(date)

ABSTRACT

Pilot tones have been used as means of phase and amplitude reference for demodulation in a fading environment, such as that of mobile communication. The receiver in a pilot tone aided transmission system extracts the reference with a narrowband pilot filter. The filter bandwidth has an optimum value, which represents a compromise between the amount of distortion and additive noise present in the received pilot. This optimum bandwidth is a function of the doppler frequency, which in turn varies with vehicle speed. The thesis investigates the use of adaptive filtering algorithms for extracting the pilot, so that the pilot filter bandwidth varies automatically in response to changes in the vehicle speed. It is the first time in which the issue of adjusting the pilot filter bandwidth with speed changes has been addressed.

Three algorithms have been investigated, of which two are commonly used, namely the Stochastic Gradient Transversal filter and the Stochastic Gradient Lattice Joint Processor Estimator. The third algorithm is novel. With this algorithm, the receiver selects the optimum member from a pre-calculated bank of stored filters. The technique is found to be very robust, and its bit error rate performance is superior to that of the other two algorithms investigated. It can provide up to a 2.0 dB improvement over a non-adaptive filtering scheme. Derivations and analysis for the adaptation schemes investigated are presented, accompanied by simulation results.

To Mom and Dad

ACKNOWLEDGEMENT

I would like to express my sincere appreciation for the support and guidance which my senior supervisor, Dr. James Cavers, has given me. The financial support provided by the British Columbia Advanced Systems Institute during the course of this work is greatly appreciated.

I would also like to thank Ms. Susan Lui. The completion of this work would not be possible without her continued encouragement.

Special thanks are due to Mobile Data International (MDI) Inc. for its support and for providing the printing facilities for generating this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. General

The popularity of mobile data communication systems have been on the rise as the technology slowly penetrates into the various markets such as taxi, courier, public utilities, safety and law enforcement. Mobile public access data networks are beginning to spring up. One such example is the system developed by Mobile Data International (MDI) in Hong Kong for Hutchison Mobile Data Limited (HMDL) which has a capability of handling a maximum of 5000 users. These networks make mobile data communication available to the smaller private user groups. Moreover, with the introduction of technologies such as MSAT, Mobile SATellite, in 1992, global coverage will be possible for private users which will increase the popularity of mobile data communication even further. With increasing popularity comes the increasing need for higher data rates and this demands more stringent requirement on the received signal fidelity.

One of the major problems which affects received signal fidelity in a mobile environment, especially land mobile, is signal fading. There are other problems such as adjacent channel interference, co-channel interference and amplifier non-linearity which play less important roles. The effects of fading need to be considered not only in the physical link level but also in the error control and higher system protocol levels.

To overcome the effect of fading, past focus of research on modulation and demodulation techniques for use in the mobile environment has been on the development of constant envelope signalling schemes demodulated non-coherently. Some examples are the developments of

Generalized Tamed FM (GTFM) [1], Gaussian MSK (GMSK) [2], and various other continuous phase modulation techniques. These techniques provide reasonably good performance in a fading environment with data rates up to 19.2 kbps for a 25 kHz channel. Today, the most active area of research in modulation technique for the mobile environment is in the use of multi-level modulations combined with trellis coding [3]. These modulation techniques invariably require good phase and amplitude reference for high performance. Multipath fading prevents the reliable acquisition of these references from the modulated data. One solution to this problem is by transmitting a pilot along with the modulated signal to be used as reference. This pilot-based reference technique is commonly known as tone-calibrated transmission (TCT) [4]. Most recently, another technique which uses reference symbols embedded in the data stream has been introduced. Data rates of up to 64 kbps have been reported for a Quadrature-Amplitude Modulated (QAM) system using either of these approaches [5].

1.2. Fading Channel Characteristics [6,7]

A fading channel generally consists of both a fast and a slow component. The slow fading component affects the median signal level as a result of shadowing by terrain features or man-made obstructions. This component can be modeled as having a log-normal amplitude probability distribution. Fast fading (or commonly known as multipath fading) affects the instantaneous signal amplitude and is caused by interference of electromagnetic wave resulting from reflections through different paths (and hence the name multipath fading). The single most important parameter in determining the statistics of a fading channel is the doppler frequency which is a function of the carrier frequency and vehicle velocity. The carrier frequency is generally fixed for a particular system. However, the vehicle velocity is subjected to change as the vehicle accelerates and decelerates. This means that the doppler frequency and therefore the fade statistics are also subjected to change.

Two of the most popular statistical distributions for use in modeling multipath fading are the Rice and Rayleigh distributions. The Rice distribution applies to the case when there is a strong line of sight component in the received signal such as in satellite communication. Rayleigh distribution, a special case of the Rice distribution, is often used to model the land mobile channel where line of sight component is very weak or not present at all. The fading model used in this thesis assumes Rayleigh fading. One of the most serious side effects of fast fading is random FM. The effect of random FM on bit error rate (BER) is small at low received signal energy to noise density ratio (E_b/N_0) where additive white Gaussian noise is the dominant cause of bit error. At high E_b/N_0, bit errors due to random FM dominate. The result is an irreducible error floor such that no matter how high E_b/N_0 is, the BER levels out at a particular value which is a function of the doppler frequency. Use of a pilot or pilot symbols provides a solution to this problem.

1.3. Tone-Aided Error Floor Suppression Techniques

Systems which involve the use of a pilot tone have appeared in many forms. The original pilot based calibration system was first proposed by Davarian [4] and was named Tone-Calibrated Transmission (TCT) system. A number of variations have emerged since the introduction of TCT. These include Dual Tone-Calibrated Transmission (DTCT) [8] and Phase-Locked Transparent Tone-In-Band (PL-TTIB) [9]. All these systems invariably involve transmitting a tone (or tones) either along side or in the center of the transmission band. A pilot based system can reduce or even eliminate the irreducible error floor and, at the same time, allow coherent detection to be used. One disadvantage is that it produces a non-constant envelope signal and thus requires amplifier with highly linear transfer characteristics for transmission. However, this does not pose an addition problem for a QAM system because QAM is a non-constant envelope signalling scheme. Each variation of pilot based calibration system has its own advantages and disadvantages. For example, TCT places the pilot in the center of the transmission band. Therefore, the reference provided by a TCT system gives the best representation of the channel distortion and is least susceptible to adjacent channel interference. However, this scheme requires a zero d.c. signal level and thus places restrictions on the modulation and coding schemes. A DTCT system uses two pilot tones, one placed at either side of the transmission band. Obviously, this scheme does not require zero d.c. level in the data spectrum but it suffers from susceptibility to adjacent channel interference and a 3 dB loss in BER performance due to the need for differential encoding. TTIB splits the data spectrum in half and moves them apart to create a null in the center of band where the pilot is placed. This scheme also does not require the data spectrum to have zero d.c. Its main disadvantage is in the increased processing complexity.

1.4. Thesis Objectives

One of the major design tradeoffs in any pilot-based system is in the pilot filter bandwidth. This issue has been investigated thoroughly by Cavers for TCT systems [10]. If the bandwidth is too narrow, then the filter cannot follow channel fluctuations. If it is too wide, then the filter admits too much noise and the result is degradation in BER. The optimal bandwidth is one which is just wide enough to cover the fade spectrum. In most literature of pilot tone techniques, the pilot filter is usually assumed to be a unity gain rectangular filter with bandwidth equal to the maximum doppler frequency plus frequency offset expected during operation. The actual doppler frequency and frequency offset during system operation is normally below the expected maxima. This suggests that there is more noise allowed into the pilot filter than is necessary. The purpose of the research leading to this thesis is to investigate the use of various adaptive filtering techniques for implementing the pilot filter so that the filter response can adjust to changes in doppler frequency and thereby minimize the system bit error rate. The idea of making the pilot filter adaptive is new. No previous work of a similar nature has been reported prior to the completion of this thesis. Three different algorithms were investigated: (1) the Stochastic Gradient Transversal Joint-Process Estimator or otherwise known simply as the Stochastic Gradient Transversal (SGT) filter, (2) the Stochastic Gradient Lattice Joint-Processor Estimator (SGL-JPE) and (3) the Filter Switching Algorithm (FSA). The first two algorithms are well known and has been used extensively in areas such as reduction of intersymbol interference and echo cancellation [11, 12]. The filter switching algorithm is a novel adaptive filtering technique which has been developed explicitly for the present application. In this thesis, we will show that the filter switching algorithm is a more suitable choice for implementing an adaptive pilot filter when compared to the two stochastic gradient algorithms.

1.5. Thesis Outline

The overall system model is presented in chapter 2 where the basic TCT structure is used. The signal flow along with all expressions describing the signals at the input and output of each system block are given in this chapter.

In order to enhance understanding of the analysis to be presented in subsequent chapters, we reproduced some of the analytical results from the paper *"Performance of Tone Calibration with Frequency Offset and Imperfect Pilot Filter"* by J. K. Cavers [10] in chapter 3. These results form a starting point for the derivations of the three adaptive filtering algorithms.

Chapter 4 examines the performance of the traditional non-adaptive approach and discusses some of the considerations in the evaluation of the performance of a system using an adaptive pilot filter. This chapter also explores some of the advantages and disadvantage of the proposed adaptive schemes.

Derivations and convergence analysis of the SGT pilot filter are presented in chapter 5. It is shown here that the minimum mean square error solution also gives a solution which minimizes the system BER with respect to the pilot filter coefficients. The SGT filter is found to perform adequately well under low signal energy to noise density ratio. However, it suffers from slow convergence when the signal energy to noise density ratio is high due to the large input eigenvalue spread. Simulations results are also presented showing the average BER performance and the convergence behavior.

Chapter 6 provides brief derivations and BER sensitivity analysis of the SGL-JPE for use in the pilot filtering application. It is shown here that the BER is extremely sensitive to fluctuations in the filter coefficients which leads to poor BER performance. Simulations results are given to support the analysis.

Detailed derivations and analysis of the filter switching algorithm are presented in chapter 7. A Markov model is introduced which enables the computation of the average BER and convergence time. It is shown from the computed results that using the FSA can improve the BER performance of up to 2 dB at a BER of 10^{-2} when compared to a non-adaptive system. With the FSA, it is possible to trade off computational complexity for convergence speed. Moreover, even the simplest implementation of the FSA is shown to converge fast enough for tracking changes in fade statistics introduced by changes in vehicle speed. A simple scheme for estimating the frequency offset between the transmitter and receiver oscillators is also presented in chapter 7.

Chapter 8 gives the conclusions and provides some recommendations for future work in the research area of this thesis.

2. SYSTEM MODEL

The system model of a pilot based calibration system is shown in figure 2.1. All signals described in this thesis are assumed to be in complex envelope representations.

2.1. Transmitted Signal

For simplicity, the transmitted signal is assumed to be BPSK and Manchester coded to create a spectral null for the pilot. The transmitted power is split between the data signal, s(t) and the pilot having amplitude, a. The transmitted complex envelope is given by:

$$z(t) = s(t) + a \tag{2,1}$$

where the data signal is defined by:

$$s(t) = A \sum_{i=-\infty}^{\infty} b_i \, p(t-iT) \tag{2.2}$$

p(t) is assumed to be an unit energy pulse such that $\int |p(t)|^2 \, dt = 1$. b_i is the binary data which can assume the values: +1 or -1. We define the ratio of pilot tone power to data signal power as r, given by:

$$r = \frac{a^2}{A^2 R_b} \tag{2.3}$$

R_b, in (2.3), denotes the bit rate of the system.

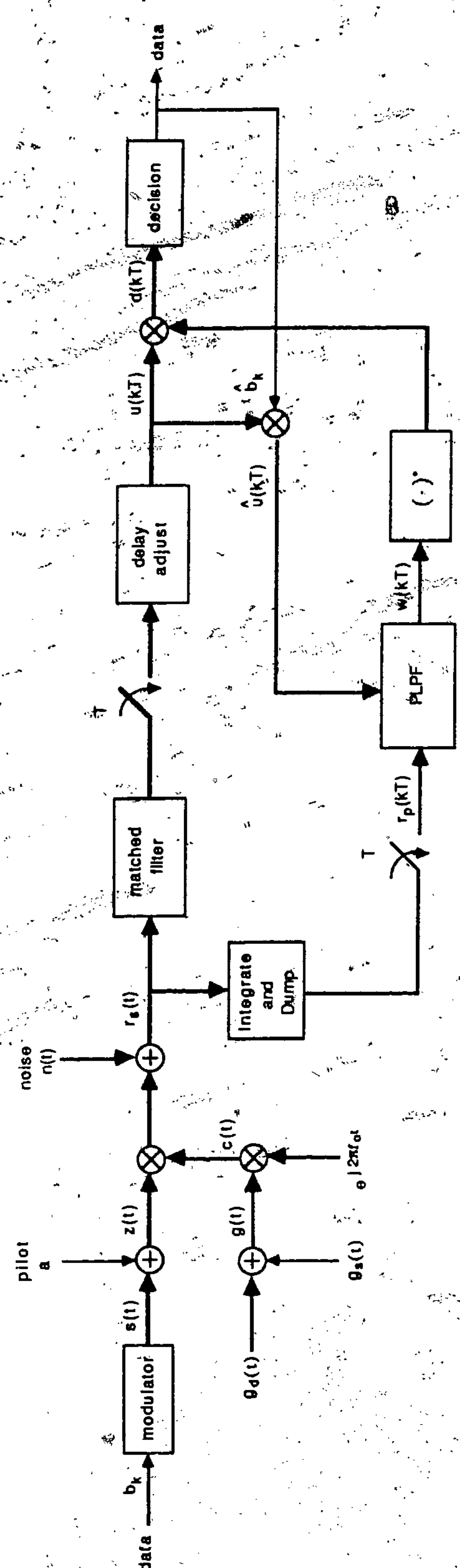


Figure 2.1 - System Model

2.2. Fading Channel

The pilot-added data signal z(t), is multiplied by a time-varying complex gain c(t). Complex white Gaussian noise n(t), with power spectral density (PSD) N_o, is then added to the modified data signal to form the received signal $r_s(t)$ given by:

$$r_s(t) = c(t)\, z(t) + n(t) \tag{2.4}$$

The complex gain can be written as:

$$c(t) = g(t) \exp(j2\pi f_o t) \tag{2.5}$$

where g(t) is a time-varying gain representing the effect of multipath fading. g(t) is modeled by a zero mean complex Gaussian process with doppler bandwidth $f_D = v/\lambda$, where v is the vehicle speed and λ is the wavelength of the carrier. f_o denotes the frequency offset between the transmitter and receiver oscillators.

The spectrum of c(t), denoted by S_c, can be expressed in terms of the spectrum of g(t) as:

$$S_c(f) = S_g(f-f_o) \tag{2.8}$$

$S_g(f)$ can be written as:

$$S_g(f) = \sigma_g^2\, \tilde{S}_g(f) \tag{2.9}$$

where σ_g^2 is the total power gain and $\tilde{S}_g(f)$ is the fade spectrum normalized to unit power.

The scattering due to multipath reflection is assumed to be isotropic such that the normalized fade spectrum is given by:

$$\tilde{S}_g(f) = \frac{1}{\pi\sqrt{f_D^2 - f^2}} \tag{2.10}$$

for which the autocorrelation function is:

$$\tilde{R}_g(\tau) = F^{-1}\{\tilde{S}_g(f)\} = J_o(2\pi f_D \tau) \tag{2.11}$$

$J_o(\bullet)$ in (2.11) is the Bessel function of the first kind with order zero. Figure 2.2 shows the shape of the fade spectrum.

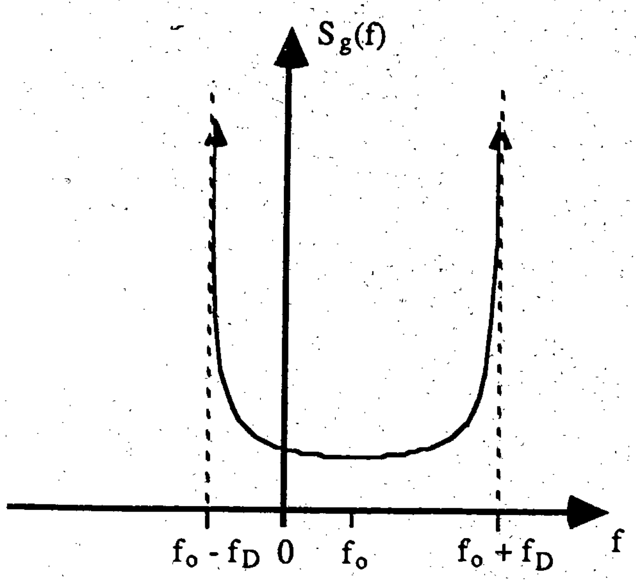


Figure 2.2 - Isotropic Fade Spectrum

2.3. Received Signal

The received signal is split into two branches, one for processing the data signal and the other for processing the pilot.

2.3.1. Data Signal Processing Branch

At the data processing branch, the received signal r(t) passes through a unit energy filter which is matched to the modulating pulse shape p(t). The resulting signal is then sampled at rate 1/T which is assumed here to equal the bit rate, R_b. Sampling time is assumed to be in perfect alignment with the bit boundary. Delay is added to the matched filter output in order to equalize the extra delay in the pilot branch. Assuming that the doppler frequency is much less than the bit rate, the delayed filter output is given by:

$$u(kT) \approx c(kT)\, A\, b_k + n_u(kT) \tag{2.12}$$

where $n_u(kT)$ is additive white Gaussian noise (AWGN) with variance N_0. The delayed matched filter output is phase-corrected by the pilot filter output w(kT) to form the data decision variable:

$$d(kT) = \mathrm{Re}\left[u(kT)\, w^*(kT)\right] \tag{2.13}$$

Since the modulation is assumed to be BPSK, the decision device merely consists of a threshold comparator.

In order to remove the data dependence in delayed matched filter output u(kT), the demodulated data, $\hat{b}_k$, is multiplied with u(kT). The resulting data removed signal, $\hat{u}$(kT), is then used as a performance reference for the adaptive pilot filter. The technique of removing data dependence in u(kT) by using demodulated data, is known as decision direction.

2.3.2. Pilot Signal Processing Branch

The pilot processing branch is responsible for producing a pilot reference with as little distortion as possible. To accomplish this task, the data modulation in the received signal is first removed by the use of an integrate and dump filter. The integrate and dump filter output is then sampled and filtered by the adaptive pilot filter. The pilot filter output is conjugated and multiplied with the delayed matched filter output for phase-correction. In figure 2.1, conjugation has been denoted by $(\bullet)^*$ as will be throughout the rest of the thesis where applicable.

The sampled integrate and dump filter output is:

$$r_p(kT) = a\, c(kT) + n_p(kT) \tag{2.14}$$

where $n_p(kT)$ is additive white Gaussian noise (AWGN) with variance N_o. $n_p(kT)$ is assumed to be uncorrelated with the noise term, $n_u(kT)$, present in the received data signal.

One should note that the integrate and dump filter cannot remove all of the data modulation due to the distortions caused by the fading channel. The result is the presence of data dependent or self noise in the received pilot. The effect of self-noise is investigated using simulations for the SGT filter and filter implemented with the FSA in sections 5.5.2 and 7.4 respectively.

3. BER OF A TONE-CALIBRATED TRANSMISSION SYSTEM USING BPSK

Analytical results related to the transmission of BPSK in a Rayleigh fading channel using pilot tone has been derived by Cavers [10]. Some of these results are reproduced here, with permission, because they form the basis of many of the analysis presented in this thesis.

One of the most important results from [10] is the analytical expression relating BER and the complex correlation coefficient between matched filter output and pilot filter output, i.e.

$$P_e = \frac{\left[\sqrt{[1-\rho_i^2]} - \rho_r\right]}{2\sqrt{1-\rho_i^2}} \tag{3.1}$$

In this relation, ρ_r and ρ_i are the real and imaginary parts of the complex correlation coefficient ρ where ρ itself is given by:

$$\rho = \frac{1}{\sqrt{1+(1+r)\frac{N_o}{E_b}}} \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{S}_g(e^{j(\omega-\omega_0)})\, H_p(e^{j\omega})^* \, d\omega}{\sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{S}_g(e^{j(\omega-\omega_0)})\, |H_p(e^{j\omega})|^2 \, d\omega + (1+r)\frac{N_o}{E_b}\frac{B_n}{rR_b}}} \tag{3.2}$$

In (3.2), $H_p(e^{j\omega})$ and B_n denote the frequency response and the noise equivalent bandwidth of the pilot filter. For a pilot filter with a real frequency response, ρ_i equals zero; so, the expression for BER reduces to:

$$P_e = \frac{(1-\rho)}{2} \tag{3.3}$$

For a rectangular pilot filter, the correlation coefficient is given by:

$$\rho = \frac{1}{\sqrt{1+(1+r)\frac{N_o}{E_b}}} \frac{P_d}{\sqrt{P_d + (1+r)\frac{N_o}{E_b}\frac{2B_p}{rR_b}}} \tag{3.4}$$

where

$$P_d = \frac{1}{\pi}\left\{ \arcsin\left[\frac{\min[B_p, f_D+f_o] - f_o}{f_D}\right] - \arcsin\left[\frac{\max[-B_p, -f_D+f_o] - f_o}{f_D}\right]\right\} \tag{3.5}$$

and B_p is the baseband pilot filter bandwidth such that the frequency response of a unit energy rectangular pilot filter is defined by the following:

$$H_p(e^{j\omega}) = \begin{cases} \frac{1}{\sqrt{2B_p}} & -2\pi B_p \leq \omega \leq 2\pi B_p \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

4. PILOT FILTER IMPLEMENTATIONS

This chapter examines the non-adaptive and adaptive approaches for pilot filtering. In section 4.1, we present some results obtained using the traditional non-adaptive filtering technique. Section 4.2 discusses some of the considerations in using an adaptive implementation in general and also some of the pros and cons of each of the three adaptive schemes which are investigated in the later chapters.

4.1. Non-Adaptive Case

Conventional pilot filter implementations has assumed an ideal rectangular filter [4,8]. Because doppler frequency ~~is~~ subjected to change and oscillator frequency offset is normally unknown, the bandwidth of a non-adaptive pilot filter needs to be wide enough to accommodate the largest doppler frequency and oscillator frequency offset which are to be expected during operation. The optimum bandwidth for a symmetric rectangular pilot filter, i.e. rectangular filter with a symmetric frequency response, is approximately equal to the sum of the doppler frequency and frequency offset. The BER performance of a rectangular pilot filter, with bandwidth wider than the optimal is degraded due to excess noise allowed into the pilot. Figure 4.1 shows the difference in BER performance at various doppler frequencies between a rectangular pilot filter with optimum bandwidth and one whose bandwidth is fixed at 150Hz. The performance penalty for using a wider filter is 2.0, 1.4 and 0.8 dB for doppler of 10, 50 and 100Hz respectively.

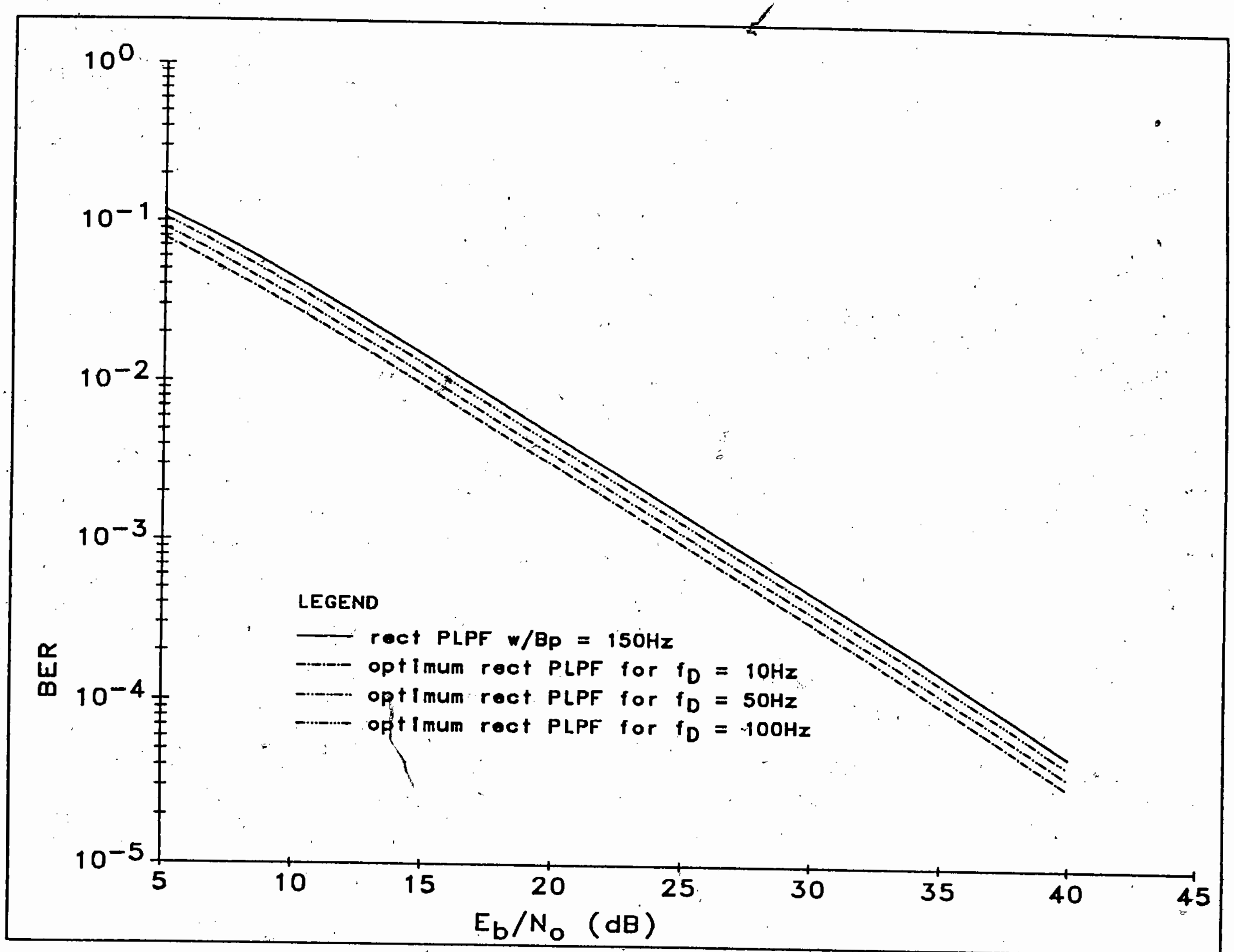


Figure 4.1 - BER Performance of a Non-Adaptive Rectangular Pilot Filter and a Rectangular Pilot Filter with Optimum Bandwidth

4.2. Adaptive Case

In the adaptive implementations, the central idea is to make the pilot filter adapt to the frequency offset and changes in doppler frequency. An adaptive filter, by definition, is a filter whose coefficients vary with time according to some performance criterion. An important consideration in adaptive pilot filter implementations is the convergence time, especially for burst systems, because the BER is often poor while the filter is adapting. The situation is worst for the case when the doppler is increasing because when the filter bandwidth cannot widen fast enough, part of the pilot energy is filtered out. The result is a serious distortion in the filtered pilot which can lead to large degradation in BER. Moreover, because of the particular shape of the fade spectrum, there is more energy in the two "horns" of the spectrum for a given frequency range near the doppler frequency (approximately 15% energy in 5% of the bandwidth) than in the center section of the spectrum. For example, using a rectangular pilot filter with bandwidth of 50Hz at a doppler of 50Hz, the BER at E_b/N_0 of 20dB is 3.59×10^{-3}, assuming no frequency offset. When the doppler takes a sudden increase to 60Hz, the BER now becomes 1.07×10^{-1}. In reality, the change in doppler is gradual and slow so that an adaptive pilot filter does not need to have a very fast convergence speed to track these changes and maintain a low BER during convergence.

There are three adaptive schemes investigated in this thesis. Each has different advantages and disadvantages; the following sections discusses some of them.

4.2.1. Stochastic Gradient Transversal Joint Process Estimator

The SGT filter is simple to implement. It uses a minimum mean square error (MMSE) criterion which will be shown in section 5.2 to be equivalent to minimizing the BER with respect to the filter coefficients. As a characteristic of all stochastic gradient adaptive algorithms, the resulting average mean square error, or in this case, average BER, is higher than optimal due to the use of the stochastic gradient approximation. The major disadvantage of this adaptive filtering scheme is the dependence of convergence speed on the eigenvalue spread of the input correlation matrix [13, 14, 15]. As will be shown later, this results in a very slow convergence at high E_b/N_0 due to the high eigenvalue spread under this condition.

4.2.2. Stochastic Gradient Lattice Joint Process Estimator

The SGL-JPE is slightly more complex than the SGT filter because it has an additional lattice structure. It also uses MMSE criterion and the stochastic gradient approximation so that there is some degradation in BER compared to the optimal filter. One of the main advantages of the SGL-JPE over the SGT filter is that its convergence behavior is not sensitive to the eigenvalue spread. Its major problem in the present application lies in the high sensitivity of the BER to changes in filter coefficients.

4.2.3. Filter Switching Algorithm

The filter switching algorithm is the most complex in terms of computation and memory requirements. However, the algorithm provides a mean of trading off complexity for convergence speed. Recall from section 4.1 that the maximum gain for using an adaptive filter over a fixed rectangular filter is only about 0.8 dB for 100Hz doppler. This suggests that BER performance is extremely important in determining the usefulness of an adaptive pilot filter. It will be shown later that the filter switching algorithm can provide a better BER performance than the other two schemes which makes it a good candidate for an adaptive pilot filter implementation.

5. STOCHASTIC GRADIENT TRANSVERSAL PILOT TONE FILTERING

The transversal or tapped-delay line is one of the most widely used structures for implementing an adaptive filter. The main reason for its popularity is because of its non-recursive structure which makes analysis of its behavior simple. A popular algorithm for adapting the transversal filter coefficients is the Stochastic Gradient (SG), also called the Least Mean Square, algorithm. This algorithm is designed to minimize the mean square error (MSE) between a desired response and the filter output. We will show in this section that the stochastic gradient transversal filter can adapt itself to minimize the BER in a pilot tone aided transmission system for a given set of parameters. Although the adaptation process is slow under certain conditions, it has its merits in that the adaptation is unaffected by frequency offset in the fade spectrum and changes in fade statistics due to shifts in doppler frequency. Section 5.1 gives a derivation of the SG algorithm. In section 5.2, we establish the equivalence between minimizing BER and the minimization of the MSE. Section 5.3 deals with the analysis of the mean convergence behavior and section 5.4 utilizes these results for computing mean convergence curves for BER. Section 5.5 provides some simulation results and lastly, section 5.6 gives a brief summary of the results obtained using the SGT pilot filter.

5.1. Filter Derivation

Detailed derivation of the SGT filter can be found in many references [13, 14, 15]. A brief derivation has been included here for completeness. Before we proceed, some notations need to be clarified. Letters in bold will be used to denote vectors or matrices. $(\bullet)^{\mathbf{T}}$ denotes transposition and $(\bullet)^{\mathbf{H}}$ denotes hermitian (or conjugate) transpose. Sampling period T will be dropped from all expressions where applicable.

Figure 5.1 shows the basic structure of the adaptive transversal pilot filter. Input to the pilot filter is the data-removed received signal, $r_p(k)$. Expression for $r_p(k)$ is given by (2.14) and is reproduced here for convenience:

$$r_p(k) = a\,c(k) + n_p(k) \tag{5.1}$$

$n_p(k)$ is AWGN with variance N_0.

Note that a Moving Window Averager (MWA) has been included as part of the filter structure which is not found in a conventional SGT filter. The MWA helps reduce the required number of coefficients by enabling the overall impulse response to cover a longer time span. Averaged samples are delayed and multiplied by filter coefficients $\mathbf{h}(k)$ where $\mathbf{h}(k)$ is the column vector defined by:

$$\mathbf{h}(k) = [h_{-L}, ..., h_{-1}, h_0, h_1, ..., h_L]^{\mathbf{T}} \tag{5.2}$$

$\mathbf{h}(k)$ are adapted to reduce the MSE between the pilot filter output and reference symbol sequence $\hat{u}(k)$ where $\hat{u}(k)$ is the decision corrected matched filter output. Although the ultimate goal for the present application is to minimize BER, it will be shown later that the minimum MSE solution is also one which gives minimum BER.

The pilot filter output is given by:

$$w(k) = \mathbf{h}^{\mathbf{H}}\,\mathbf{r}(k) \tag{5.3}$$

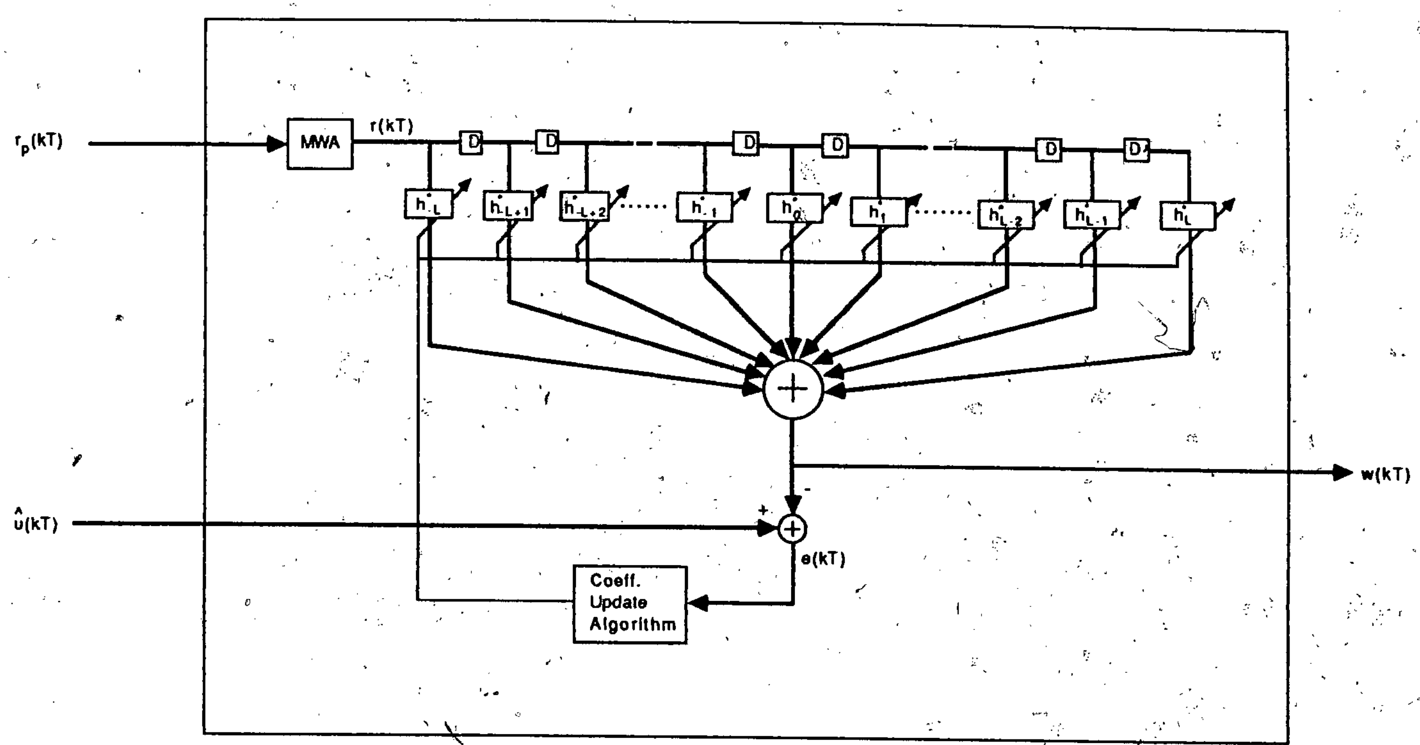


Figure 5.1 - Stochastic Gradient Transversal Filter

with $\mathbf{r}(k)$ denoting the output vector of the MWA which is defined by:

$$\mathbf{r}(k) = [r(k\text{-}L), \ldots, r(k\text{-}1), r(k), r(k+1), \ldots, r(k+L)]^{\mathbf{T}} \tag{5.4}$$

$r(k)$ is the MWA output given by:

$$r(k) = a \sum_{i=-(K-1)/2}^{(K-1)/2} c(k-i) \; + n(k) \tag{5.5}$$

$n(k)$ is AWGN with variance N_0K where K is the length of the MWA which is assumed to be odd. Error $e(k)$ is formed by taking the difference between desired symbol and pilot filter output such that:

$$\begin{aligned} e(k) &= \hat{u}(k) - w(k) \\ &= \hat{b}_k \, u(k) - \mathbf{h}^{\mathbf{H}} \, \mathbf{r}(k) \end{aligned} \tag{5.6}$$

$u(k)$ is given by (2.12) and is reprinted here for convenience:

$$u(k) = c(k) \, A \, b_k + n_u(k) \tag{5.7}$$

From (5.6) and (5.7), the expression for MSE can be shown to be:

$$\begin{aligned} J(\mathbf{h}) &= E[|e(k)|^2] \\ &= \sigma_u^2 - \mathbf{h}^{\mathbf{H}} \; E[\hat{b}_k \, u^*(k) \, \mathbf{r}(k)] \\ &\quad - E[\hat{b}_k \, u(k) \, \mathbf{r}^{\mathbf{H}}(k)] \, \mathbf{h} + \mathbf{h}^{\mathbf{H}} \, E[\mathbf{r}(k) \, \mathbf{r}^{\mathbf{H}}(k)] \, \mathbf{h}(k) \end{aligned} \tag{5.8}$$

where σ_u^2 denotes the variance of u(k). This equation reveals that if $\mathbf{h}_o$ is the optimal coefficient vector given that $\hat{b}_k = b_k$, then another solution which minimizes $J(\mathbf{h})$ is $\mathbf{h} = -\mathbf{h}_o$ given that $\hat{b}_k = -b_k$. In other words, there is a two-fold ambiguity associated with the minimum MSE solution vector caused by decision direction. One way to resolve this ambiguity is to employ differential encoding or to initialize the adaptation process with a short training sequence at the start of the algorithm. It has been found using simulations that 20 training bits are sufficient to avoid misadaptation. Differential encoding is not recommended because it imposes a 3-dB penalty in a Rayleigh fading environment.

Terms involving $\hat{b}_k$ can be expanded and simplified in the following manner:

$$\begin{aligned} E[\hat{b}_k\, u^*(k)\, \mathbf{r}(k)] &= E[\hat{b}_k\, (A\, b_k\, c^*(k) + n_u^*(k))\, \mathbf{r}(k)] \\ &= E[\hat{b}_k\, b_k]\, E[A\, c^*(k)\, \mathbf{r}(k)] + E[\hat{b}_k]\, E[n_u^*(k)\, \mathbf{r}(k)] \end{aligned} \tag{5.9}$$

Here, $\hat{b}_k$ and b_k are assumed to be independent of c(k), r(k) and $n_u(k)$. When the BER is small, $\hat{b}_k \approx b_k$. So $E[\hat{b}_k\, b_k] \approx 1$, $E[\hat{b}_k] \approx 1$ and (5.9) can be approximated by:

$$E[\hat{b}_k\, u^*(k)\, \mathbf{r}(k)] \approx E[(A\, c^*(k) + n_u^*(k))\, \mathbf{r}(k)] \tag{5.10}$$

Let us denote the cross-correlation vector between $\hat{u}(k)$ and $\mathbf{r}(k)$ by $\mathbf{p}$ and the correlation matrix of $\mathbf{r}(k)$ by $\mathbf{R}$ such that:

$$\mathbf{p} = E[\hat{u}^*(k)\, \mathbf{r}(k)] \approx E[(A\, c^*(k) + n_u^*(k))\, \mathbf{r}(k)] \tag{5.11}$$

and

$$\mathbf{R} = E[\mathbf{r}(k)\,\mathbf{r}^H(k)] \tag{5.12}$$

J(**h**) can then be written as:

$$J(\mathbf{h}) = \sigma_u^2 - \mathbf{h}^H \mathbf{p} - \mathbf{p}^H \mathbf{h} + \mathbf{h}^H \mathbf{R}\,\mathbf{h} \tag{5.13}$$

To find the minimum MSE solution, we differentiate the MSE with respect to the coefficient vector **h**. Differentiation of the terms $\mathbf{h}^H \mathbf{p}$, $\mathbf{p}^H \mathbf{h}$, and $\mathbf{h}^H \mathbf{R}\,\mathbf{h}$ with respect to **h** results in the following [15]:

$$\frac{d(\mathbf{h}^H \mathbf{p})}{d\mathbf{h}} = \mathbf{0} \tag{5.14}$$

$$\frac{d(\mathbf{p}^H \mathbf{h})}{d\mathbf{h}} = 2\,\mathbf{p} \tag{5.15}$$

and

$$\frac{d(\mathbf{h}^H \mathbf{R}\,\mathbf{h})}{d\mathbf{h}} = 2\,\mathbf{R}\,\mathbf{h} \tag{5.16}$$

where **0** is a null vector with the same dimension as **h**. Using (5.15) and (5.16), the gradient of J(**h**) is thus given by:

$$\frac{dJ(\mathbf{h})}{d\mathbf{h}} = -2\,\mathbf{p} + 2\,\mathbf{R}\,\mathbf{h} \tag{5.17}$$

Next, we equate the gradient to the null vector giving:

$$\mathbf{R}\,\mathbf{h}_o = \mathbf{p} \tag{5.18}$$

with $\mathbf{h}_o$ denoting the optimum coefficient vector. The optimum filter represented by $\mathbf{h}_o$ is called Wiener filter. Equation 5.18 is known as the *normal equation* . The reason for this name is because when the optimum filter is used, the estimation error vector, $\mathbf{e}_o(k)$, is *normal* to the filter output vector, $\mathbf{w}(k)$.

Solving (5.18) by estimating $\mathbf{R}$ and $\mathbf{p}$, and inverting $\mathbf{R}$ can be computationally difficult. An alternative is to find $\mathbf{h}_o$ in successive steps by making correction to the coefficient vector in a direction opposite to the gradient vector (i.e. direction of the steepest descent). This procedure is known as the method of steepest descent [14]. The steepest descent algorithm is represented by the update equation:

$$\mathbf{h}(k+1) = \mathbf{h}(k) - \frac{\Delta}{2}\,\frac{dJ(\mathbf{h})}{d\mathbf{h}} \tag{5.19}$$

where Δ controls the step size. The factor of two is introduced for convenience only. The gradient vector given by (5.17) is a statistically averaged quantity. In practice, it is common to use an instantaneous estimate of the gradient (hence the name stochastic gradient) formed by removing the expectation in the expressions defining $\mathbf{R}$ and $\mathbf{p}$. The resulting update equation can be shown to be:

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \Delta\, e^*(k)\, \mathbf{r}(k) \tag{5.20}$$

This update equation completely describes the stochastic gradient algorithm.

Although use of the stochastic gradient greatly simplifies the coefficient update algorithm, a price has to be paid in the form of increased average MSE and hence increased BER. The *excess* average MSE (defined as J_∞ - J_{min}) is due to random fluctuations of the gradient estimate. An analytical expression exists for computing the excess average MSE [14]. However, similar expression for *excess* BER is difficult to derive due to the complex dependence of BER on **h**. Nevertheless, effect of various parameters on the excess BER will be investigated in detail using simulation results.

Since the pilot filter implementation is digital, the delay mismatch between pilot filter output and matched filter output is negligible if an appropriate delay compensation is used. If we further assume that the fading process is stationary (so that $R_g(k)$ is conjugate symmetric), then the resulting optimum coefficient vector will be conjugate symmetric. This means that, at any given time instant, only positive time samples are necessary in determining signal statistics while negative time samples give the same stochastic information. One can take advantage of this redundant information to reduce the amount of noise present in the gradient estimate and therefore decrease the excess BER by averaging the positive and negative time samples. We do this by changing the update equation to the following:

$$h_i(k+1) = h_i(k) + \frac{\Delta}{2}\,(e(k)\,r^*(k-i) + e^*(k)\,r(k+i)) \qquad i = 0,\ldots,L \qquad (5.21)$$

One can easily show that the resulting coefficient vector **h** described by (5.21) is now conjugate symmetric so that $h_{-i}(k+1) = h_i^*(k+1)$.

5.2. Equivalence between Minimum BER and Minimum Mean Square Error Solutions

In this section, we will show that the use of coefficient vector which minimizes the mean square error between $\hat{u}(k)$ and $w(k)$ also result in minimum BER. To begin, we first make the assumption that the filter is forced to be conjugate symmetric. From Fourier transform theory [16], this implies that the resulting filter frequency response will be real. As discussed in section 3, the correlation coefficient between $\hat{u}(k)$ and $w(k)$, under this condition, will also be real and the expression for the BER is given by (3.3).

The correlation coefficient between $\hat{u}(k)$ and $w(k)$ is defined by:

$$\rho = \frac{\sigma_{\hat{u}w}{}^2}{\sigma_{\hat{u}}\,\sigma_w} \tag{5.22}$$

where $\sigma_{\hat{u}w}{}^2$ denotes the covariance between $\hat{u}(k)$ and $w(k)$ and $\sigma_w{}^2$ denotes the variance of $w(k)$. Using (5.3) and (5.1), the terms $\sigma_{\hat{u}w}{}^2$ and σ_w can be expanded in matrix notation as:

$$\sigma_{\hat{u}w}{}^2 = E[\hat{u}(k)\,w^*(k)] = \mathbf{p}^{\mathbf{H}}\,\mathbf{h} \tag{5.23}$$

and

$$\sigma_w = \sqrt{E[w(k)\,w^*(k)]} = \sqrt{\mathbf{h}^{\mathbf{H}}\;\mathbf{R}\;\mathbf{h}} \tag{5.24}$$

Equation 5.22 may then be written as:

$$\rho = \frac{1}{\sigma_{\hat{u}}} \frac{\mathbf{p}^H \mathbf{h}}{\sqrt{\mathbf{h}^H \mathbf{R} \mathbf{h}}} \tag{5.25}$$

From (3.3), we see that the solution which gives minimum BER is one which maximizes ρ. We maximize ρ by differentiating (5.25) with respect to **h** and equating the result to the null vector (see ref. [15] for a review in vector differentiation). The resulting equation can be shown to be:

$$\frac{d\sigma_{\hat{u}w}^2}{d\mathbf{h}} \sigma_w = \frac{d\sigma_w}{d\mathbf{h}} \sigma_{\hat{u}w}^2 \tag{5.26}$$

Since $\sigma_{\hat{u}w}^2$ is real (because ρ is real), it can be written as:

$$\begin{aligned} \sigma_{\hat{u}w}^2 &= \frac{1}{2} (\sigma_{\hat{u}w}^2 + \sigma_{\hat{u}w}^{2*}) \\ &= \frac{1}{2} (\mathbf{p}^H \mathbf{h} + \mathbf{h}^H \mathbf{p}) \end{aligned} \tag{5.27}$$

Using (5.27), we can evaluate $\frac{d\sigma_{\hat{u}w}^2}{d\mathbf{h}}$ to give:

$$\frac{d\sigma_{\hat{u}w}^2}{d\mathbf{h}} = \mathbf{p} \tag{5.28}$$

The next step is to obtain $\frac{d\sigma_w}{d\mathbf{h}}$. Differentiating (5.24) with respect to **h** gives:

$$\begin{aligned} \frac{d\sigma_w}{d\mathbf{h}} &= \frac{1}{2\sqrt{\mathbf{h}^H \mathbf{R} \mathbf{h}}} \frac{d(\mathbf{h}^H \mathbf{R} \mathbf{h})}{d\mathbf{h}} \\ &= \frac{1}{2\sigma_w} 2 \mathbf{R} \mathbf{h} \\ &= \frac{1}{\sigma_w} \mathbf{R} \mathbf{h} \end{aligned} \tag{5.29}$$

Substituting (5.28) and (5.29) into (5.26) gives the following result:

$$\mathbf{p}\,\sigma_w^2 = \mathbf{R}\,\mathbf{h}_o\,\sigma_{\hat{u}w}^2 \tag{5.30}$$

To see if the solution to the normal equation also solves (5.30), we substitute the normal equation into (5.30) and we get:

$$\sigma_w^2 = \sigma_{\hat{u}w}^2 \tag{5.31}$$

From (5.23) and (5.24), equation 5.32 can be expanded to give:

$$\mathbf{h}_o^H\,\mathbf{R}\,\mathbf{h}_o = \mathbf{p}^H\,\mathbf{h}_o \tag{5.32}$$

Since $\mathbf{p}^H\,\mathbf{h}_o$ is real, $\mathbf{p}^H\,\mathbf{h}_o$ can be replaced by $\mathbf{h}_o^H\,\mathbf{p}$ so that (5.32) becomes:

$$\mathbf{h}_o^H\,\mathbf{R}\,\mathbf{h}_o = \mathbf{h}_o^H\,\mathbf{p} \tag{5.33}$$

Finally, if we again substitute the normal equation into (5.33), we see that the RHS of (5.33) now equals the LHS. This completes the proof. Closer examination of (5.33) reveals that the use of any multiple of $\mathbf{h}_o$ will give rise to a minimum BER meaning that the optimum coefficient vector (for minimum BER) is gain independent. This fact is evident from (5.25).

Expression for the minimum BER can be obtained by substituting the normal equation into (5.25) to find the maximum correlation coefficient and then make use of (3.3). The maximum correlation coefficient is given by:

$$
\begin{aligned}
\rho_{max} &= \frac{\mathbf{p}^H \mathbf{h}_o}{\sqrt{\sigma_{\hat{u}}^2\, \mathbf{h}_o^H \mathbf{R}\, \mathbf{h}_o}} \\
&= \frac{\mathbf{h}_o^H \mathbf{R}^H \mathbf{h}_o}{\sqrt{\sigma_{\hat{u}}^2\, \mathbf{h}_o^H \mathbf{R}\, \mathbf{h}_o}} \\
&= \frac{\mathbf{h}_o^H \mathbf{R}\, \mathbf{h}_o}{\sqrt{\sigma_{\hat{u}}^2\, \mathbf{h}_o^H \mathbf{R}\, \mathbf{h}_o}} \\
&= \sqrt{\frac{\mathbf{h}_o^H \mathbf{R}\, \mathbf{h}_o}{\sigma_{\hat{u}}^2}} \\
&= \sqrt{\frac{\mathbf{p}^H \mathbf{h}_o}{\sigma_{\hat{u}}^2}}
\end{aligned} \tag{5.34}
$$

5.3. Convergence Analysis

All recursive algorithms go through a transient or convergence period before a steady state can be reached. These algorithms invariably involve feedback and therefore are subjected to instability. Thus, we need to first examine the conditions which make the algorithm stable. For simplicity, we will assume that the fading process is stationary.

Defining the coefficient-error vector as:

$$\mathbf{c}(k) = \mathbf{h}(k) - \mathbf{h}_o, \tag{5.35}$$

we now perform a coordinate transformation by substituting (5.6) and (5.35) into the update equation 5.20. The result is:

$$\mathbf{c}(k+1) = (\mathbf{I} - \Delta\, \mathbf{r}(k)\, \mathbf{r}^H(k))\, \mathbf{c}(k) + \Delta\, (\mathbf{r}(k)\, \hat{b}_k\, u^*(k) - \mathbf{r}(k)\, \mathbf{r}^H(k)\, \mathbf{h}_o) \tag{5.36}$$

where $\mathbf{I}$ is the identity matrix. If we now take expectation of $\mathbf{c}(k+1)$ and replace various terms with $\mathbf{p}$ and $\mathbf{R}$, we have:

$$E[\mathbf{c}(k+1)] = (\mathbf{I} - \Delta\,\mathbf{R})\,E[\mathbf{c}(k)] + \Delta\,(\mathbf{p} - \mathbf{R}\,\mathbf{h}_o) \tag{5.37}$$

Using the normal equation, (5.37) can be reduced to:

$$E[\mathbf{c}(k+1)] = (\mathbf{I} - \Delta\,\mathbf{R})\,E[\mathbf{c}(k)] \tag{5.38}$$

We now diagonalize $\mathbf{R}$ so that it can be written as [15, 17]:

$$\mathbf{R} = \mathbf{M}\,\Lambda\,\mathbf{M}^{\mathbf{H}} \tag{5.39}$$

Here, $\mathbf{M}$ is a matrix with its columns consisting of eigenvectors of $\mathbf{R}$, and Λ is a diagonal matrix with eigenvalues of $\mathbf{R}$ as its diagonal elements. Using (5.39), $E[\mathbf{c}(k+1)]$ can be expressed as:

$$E[\mathbf{c}(k+1)] = (\mathbf{I} - \Delta\,\mathbf{M}\,\Lambda\,\mathbf{M}^{\mathbf{H}})\,E[\mathbf{c}(k)] \tag{5.40}$$

Changing variables so that $\mathbf{q}(k) = \mathbf{M}^{\mathbf{H}}\,E[\mathbf{c}(k)]$, we rewrite (5.40) as:

$$\mathbf{q}(k+1) = (\mathbf{I} - \Delta\,\Lambda)\,\mathbf{q}(k) \tag{5.40}$$

The separate dimensions of $E[\mathbf{c}(k+1)]$ are now decoupled into their natural "modes" so that we can write:

$$q_i(k+1) = (1 - \Delta \lambda_i)\, q_i(k) \qquad i = -L,\ldots,L \tag{5.42}$$

where λ_i is the i^{th} eigenvalue of the correlation matrix **R**. Equation 5.42 can be rewritten in a more convenient form:

$$q_i(k+1) = (1 - \Delta \lambda_i)^k\, q_i(0) \tag{5.43}$$

Recognizing that (5.43) represents a geometric series, we obtain the stability condition:

$$| 1 - \Delta \lambda_i | < 1 \tag{5.44}$$

Thus, for stability, the step size Δ must satisfy the following:

$$0 < \Delta < \frac{2}{\lambda_i} \qquad \text{for all } i \tag{5.45}$$

Since the fading process is assumed stationary, the correlation matrix **R** is positive definite [15] and its eigenvalues are all real and positive. It is sufficient that the following is true for stability:

$$0 < \Delta < \frac{2}{\lambda_{max}} \tag{5.46}$$

Equation 5.46 describes the condition for "mean" convergence. In reality, in order to avoid divergence due to statistical variations, it is common [15] to use a more restrictive bound for selecting the step size Δ:

$$0 < \Delta < \frac{2}{(2L+1)\, E[|r|^2]} \tag{5.47}$$

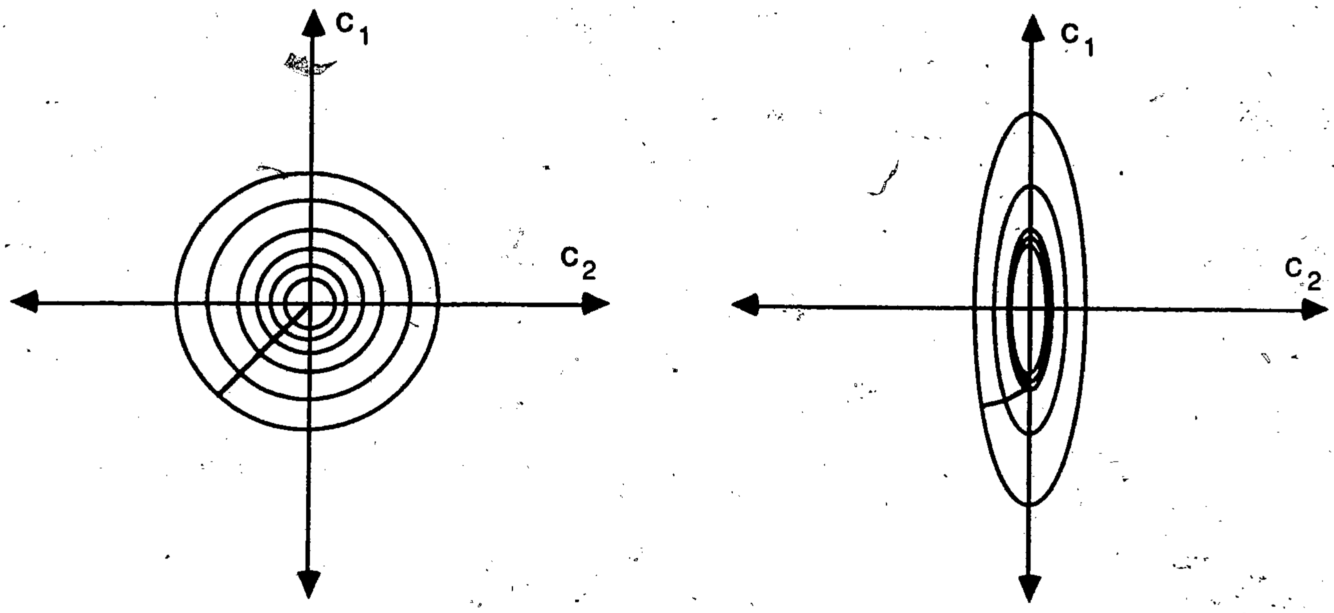
Here, we have made use of the following inequality:

$$\lambda_{max} \le \sum_{i=-L}^{L} \lambda_i \tag{5.48}$$

and the fact that [15]:

$$\sum_{i=-L}^{L} \lambda_i = \text{sum of mean-square values of all tap inputs} \tag{5.49}$$

Having established the stability condition, we proceed to analyze the convergence properties. In terms of convergence, we require that all modes of the algorithm to converge before steady state can be reached. From (5.43), it is clear that convergence speed increases with the step size. However, step size is limited by (5.46). Step size which is small enough to ensure stability can make convergence slow for modes with small eigenvalues. Assuming that we set Δ to its upper limit, $2/\lambda_{max}$, then convergence speed is limited by $\lambda_{max} / \lambda_{min}$, i.e. the eigenvalue spread. The effect of eigenvalue spread on convergence can be visualized by plotting contours of equi-MSE as function of c(k). Figure 5.2 shows equi-MSE contours for a second order system with small and large eigenvalue spreads.



(a) Small Eignevalue Spread (b) Large Eigenvalue Spread

Figure 5.2 - Contours of MSE as a function of **c**(k)

When eigenvalue spread is small (≈ 1), the resulting contour is circular. This means that the direction of negative gradient vector is always in the direction of minimum MSE as illustrated in figure 5.2a. For the case of a large eigenvalue spread, the contour is elliptical [13] as shown in figure 5.2b; each step does not go directly toward the minimum. So the number of steps required to reach the minimum increases. It should also be clear from figure 5.2 that convergence speed is also highly dependent on the initial coefficient vector as one can choose an initial vector which is arbitrarily close to the minimum.

Aside from the eigenvalues, the corresponding eigenvectors of **R** also play an important role in determining convergence behavior. From the definition of **q**(k), we see that the coefficient error vector is simply a linear combination of the eigenvectors such that:

$$c(k) = \sum_{i=-L}^{i=L} q_i(k)\, \xi_i \tag{5.49}$$

where ξ_i is the i^{th} eigenvector corresponding to λ_i. Thus, each eigenvector shapes the filter response independently and the amount of shaping or weighting is determined by $q_i(k)$. In other words, the overall convergence behavior is affected by the amount of contribution of each mode to the MSE or BER as well as how fast the individual mode converges.

In order to further understand the convergence behavior of the SGT filter, we need to understand some of the physical significance of eigenvalue spread. First consider the case when the successive samples of r are uncorrelated. With the assumption that the input process is stationary, this implies that **R** is a multiple of **I**. In this case, the eigenvalue spread is at its minimum, i.e. equal to one. Conversely, when r is completely correlated (correlation coefficient = 1), then all elements of **R** are identical. It can be shown that, in this case, at least one of the eigenvalues of **R** must be zero meaning that the eigenvalue spread is infinite. So, eigenvalue spread can be considered as a measure of the correlatedness between time samples of a stochastic process, in this case, r. We can therefore expect that when samples of r are highly correlated, convergence will be slow. This is intuitively satisfying because if we consider correlation as a measure of information content, high correlation in the input samples implies low information content. This means that it will take many samples in order to characterize the process which generated these samples.

5.4. Computed Results

In this section, we attempt to predict the effect of various parameters on the convergence of the SGT by evaluating the input eigenvalue spread. We will present some computed results showing the convergence of BER as a function of time in the form of learning curves [14]. We proceed by first deriving expressions for computing the correlation matrix **R** and the cross-correlation vector **p**.

The correlation matrix **R** can be defined as a matrix with elements:

$$R_{nm}(k) = E[r^*(k-n)\, r(k-m)] \tag{5.51}$$

Expanding r using (5.5) and remembering that input samples to the transversal filter is separated in time by KT, we get:

$$\begin{aligned} R_{nm}(k) &= E\Big[\Big(a \sum_{i=-(K-1)/2}^{(K-1)/2} c^*((k-n)K-i) + n^*((k-n)K)\Big)\Big(a \sum_{j=-(K-1)/2}^{(K-1)/2} c((k-m)K-j) + n((k-m)K)\Big)\Big] \\ &= a^2 \sum_{i=-(K-1)/2}^{(K-1)/2} \sum_{j=-(K-1)/2}^{(K-1)/2} E[c^*((k-n)K-i)\, c((k-m)K-j)] + E[n^*((k-n)K)\, n((k-m)K)] \\ &= 2a^2 \sum_{i=-(K-1)/2}^{(K-1)/2} \sum_{j=-(K-1)/2}^{(K-1)/2} R_g((n-m)K+(i-j)) + 2N_oK\delta_{nm} \end{aligned} \tag{5.52}$$

where δ_{nm} is the Kronecker delta function. In (5.52), we have made use of the assumption that n(i) and c(j) are uncorrelated. Now, defining $l = i - j$ and $q(l) = K - |l|$ and making use of the conjugate symmetry of R_g, we can combine the double summation in (5.52) into a single one and get:

$$R_{nm}(k) = 2a^2 \sum_{l=-(K-1)}^{K-1} q(l)\, R_g((n-m)K+l) + 2N_oK\delta_{nm} \tag{5.53}$$

Equation 5.53 can be expressed in terms of the received energy per bit E_b. With P_r and P_t denoting the total transmit and receive power, E_b is given by:

$$E_b = \frac{P_r}{R_b} = \sigma_g^2 \frac{P_t}{R_b} = \sigma_g^2 A^2 (1+r) = \sigma_g^2 \frac{a^2}{r R_b} (1+r) \qquad (5.54)$$

where (2.3) has been used. Using (5.53), expression for the correlation becomes:

$$\begin{aligned} R_{nm}(k) &= 2N_oK \left\{ \frac{a^2\sigma_g^2}{N_oK} \sum_{l=-(K-1)}^{K-1} q(l)\, \tilde{R}_g((n-m)K+l) + \delta_{nm} \right\} \\ &= 2N_oK \left\{ \frac{E_b}{N_o} \frac{r}{1+r} \frac{R_b}{K} \sum_{l=-(K-1)}^{K-1} q(l)\, \tilde{R}_g((n-m)K+l) + \delta_{nm} \right\} \qquad (5.55) \end{aligned}$$

Expression for the cross-correlation vector elements can be derived in a similar fashion to obtain the following result:

$$p_n(k) = 2\, E_b \frac{\sqrt{rR_b}}{1+r} \sum_{i=-(K-1)/2}^{(K-1)/2} \tilde{R}_g(nK-i) \qquad (5.56)$$

For a given set of channel and filter parameters, we can compute **R** and **p** using 5.55 and 5.56. With **R** and **p**, we can compute the optimum coefficient vector $\mathbf{h_o}$ by solving the normal equation and all eigenvalues and eigenvectors associated with **R** by using standard routines [18]. The eigenvalues and eigenvectors are needed to compute the time evolution of the filter coefficient vector. From the coefficient vector, we can compute the BER at each time step using (5.25) and (3.3).

5.4.1. Input Eigenvalues

The input eigenvalue spread has been computed as a function of various parameters in order to predict the convergence performance of the algorithm under various conditions. The results are summarized below:

- E_b/N_o - Eigenvalue spread is directly proportional to E_b/N_o. In terms of the correlation matrix **R**, decreasing noise level decreases values of the diagonal elements of **R** which directly decreases the eigenvalues. Since E_b/N_o can vary significantly (10-30dB typically), the eigenvalue spread may vary up to 100 fold.
- f_DT - Eigenvalue spread is a decreasing function with increasing doppler frequency. Increasing f_DT has the effect of narrowing the autocorrelation function of the fading process, R_g. Since R_g has a decreasing envelope with time, successive samples of the fading process appear less correlated with increasing doppler. Change in eigenvalue spread as a function of f_DT is small compared to E_b/N_o. For f_DT varied from 0.5% to 4%, the eigenvalue spread is approximately halved.
- MWA - Increasing MWA length also increases eigenvalue spread. Recall from section 5.1 that the length of the MWA determines the effective time separation between successive samples of the input of the transversal filter. Hence, as the length of the MWA increases, the time separation also increases and successive samples become less correlated. However, the averaging has the opposite effect of increasing the correlation between samples. Even at moderate doppler frequency, the correlation function of the fading process has a relatively wide main lobe. As such, the decrease in correlation due to the increase in time separation is small compared to the increase in correlation due to averaging. The result is an overall increase in eigenvalue spread. Like f_DT, the change

is small as compared to change due to E_b/N_0. For an increase of MWA length from 1 to 5, the increase in eigenvalue spread is approximately 2.5 times.

- pilot to signal power ratio r - change in eigenvalue spread due to r is small. Computation shows a 1.5 fold increase in eigenvalue spread for r increased from 0.2 to 0.5.
- filter length - longer filter tends to increase the eigenvalue spread. However, the change is negligibly small compared to changes due to other parameters.

The following conclusion can be drawn from these results. Convergence will be slow at low doppler frequency (and hence low vehicle speed). It will be VERY slow for a system operating at high E_b/N_0 such as 30-40dB. Fortunately, most mobile communication systems operate at the vicinity of 20dB where eigenvalue spread is not a problem.

5.4.2. Convergence of BER in the Mean

Traditionally, a learning curve is defined as a plot of MSE versus the number of iterations. For the present application, we will use it to represent a plot of BER versus number of iterations. Procedures for evaluating mean convergence of the coefficient vector and BER at each iteration has been discussed earlier. Here, we present some learning curves in an attempt to gain more insight into the convergence behavior before proceeding with simulations.

Figure 5.3 shows the BER learning curve for various f_DT. The following parameters were used:

E_b/N_0 = 20dB; filter length = 5; MWA length = 3; and step size = 0.1. Note the two clearly defined sections evident in the two learning curves with f_DT equal to 0.0208 and 0.0417. The flat portion of the curves were due to slow convergence of modes with small eigenvalues.

Figure 5.4 shows learning curves for different E_b/N_o with f_DT of 0.0208 and same filter parameters as used for the previous figure. The set of curves illustrates heavy dependence of convergence speed on E_b/N_o. Note that even though convergence is slow for E_b/N_o of 30dB, the BER is already below 10^{-3} after about 300 iterations.

One of the most important advantage of an adaptive pilot filter over a non-adaptive one is the ability of the adaptive pilot filter to adjust its bandwidth when the vehicle speed changes. Convergence speed during vehicle deceleration is not a problem because this only means that the filter bandwidth is too wide, resulting in more noise being admitted than is necessary. At a typical E_b/N_o, the deterioration of BER in this case is small. However, during vehicle acceleration, the bandwidth of the adapting filter will be too narrow to cover the entire fade spectrum. The result is a large increase in BER while the filter tries to adapt its coefficients to increase its bandwidth. This phenomenon is illustrated by figure 5.5 which shows the learning curves for increasing doppler frequency at various E_b/N_o. f_DT for each of the curves are increased by 10/2400 every 3000 bits (corresponding to a stepwise acceleration of 10 kmph/sec for a 2400 bps system) from 0.00417 to 0.0417. For the initial 6000 bits, the E_b/N_o has been set at 15dB for all curves. This is necessary to enable fast convergence so that the SGT filter is very close to being converged at each E_b/N_o, before the doppler is stepped up. For E_b/N_o of 10 and 20dB, the jumps in BER are not as evident because convergence at these signal to noise ratios is relatively fast. The large "bump" in BER for E_b/N_o of 40dB shows high sensitivity of BER to vehicle acceleration at high E_b/N_o. For comparison, the BER at various E_b/N_o for a ideal rectangular pilot filter with normalized bandwidth of 0.0625 are also shown in figure 5.5. At $E_b/N_o \geq 20$dB, rectangular filter is better than SGT filter except during initial part of the acceleration period. For $E_b/N_o < 20$dB, SGT filter is better through most part of the acceleration.

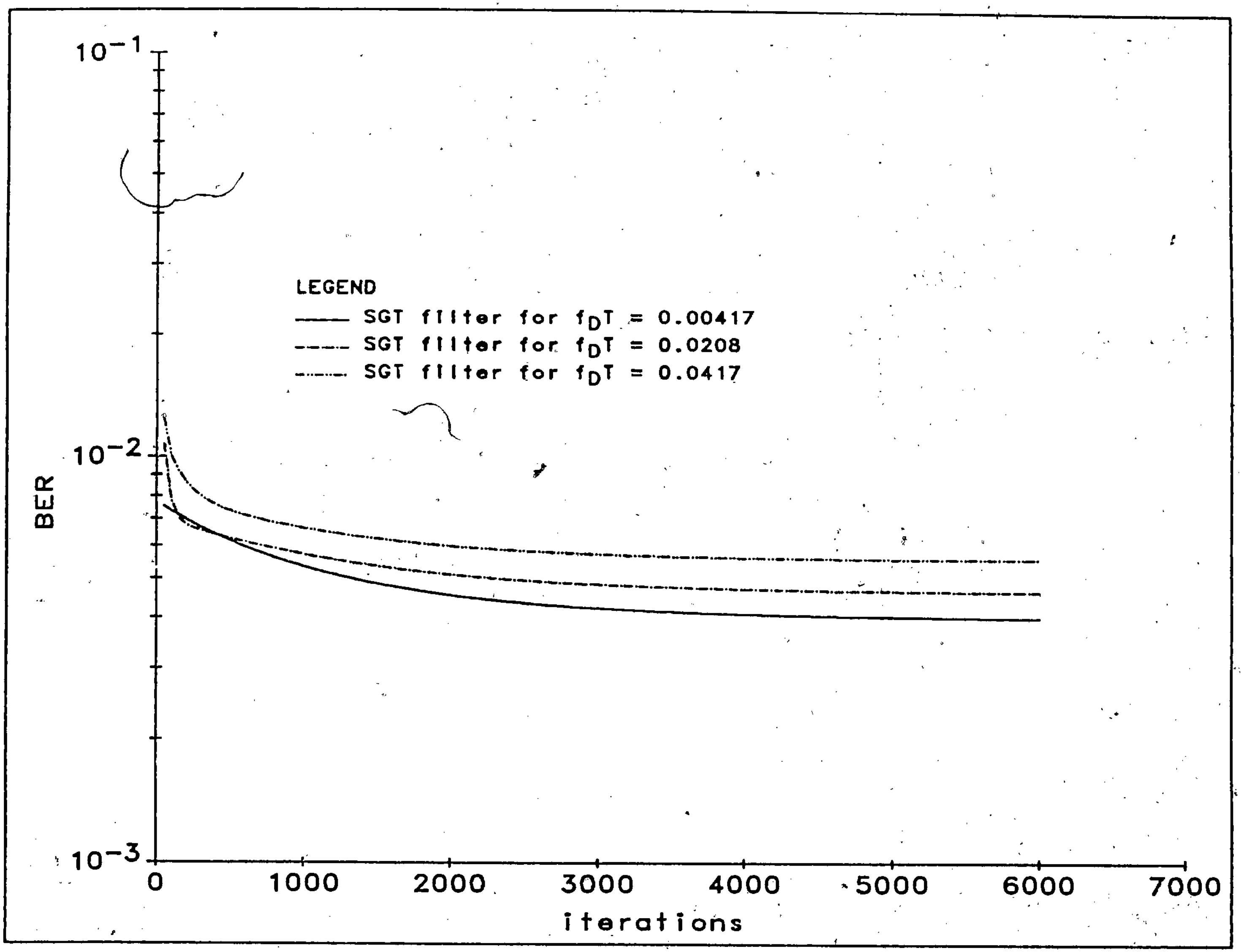


Figure 5.3 - Mean Convergence Curves of a SGT Pilot Filter at Various Doppler Frequencies

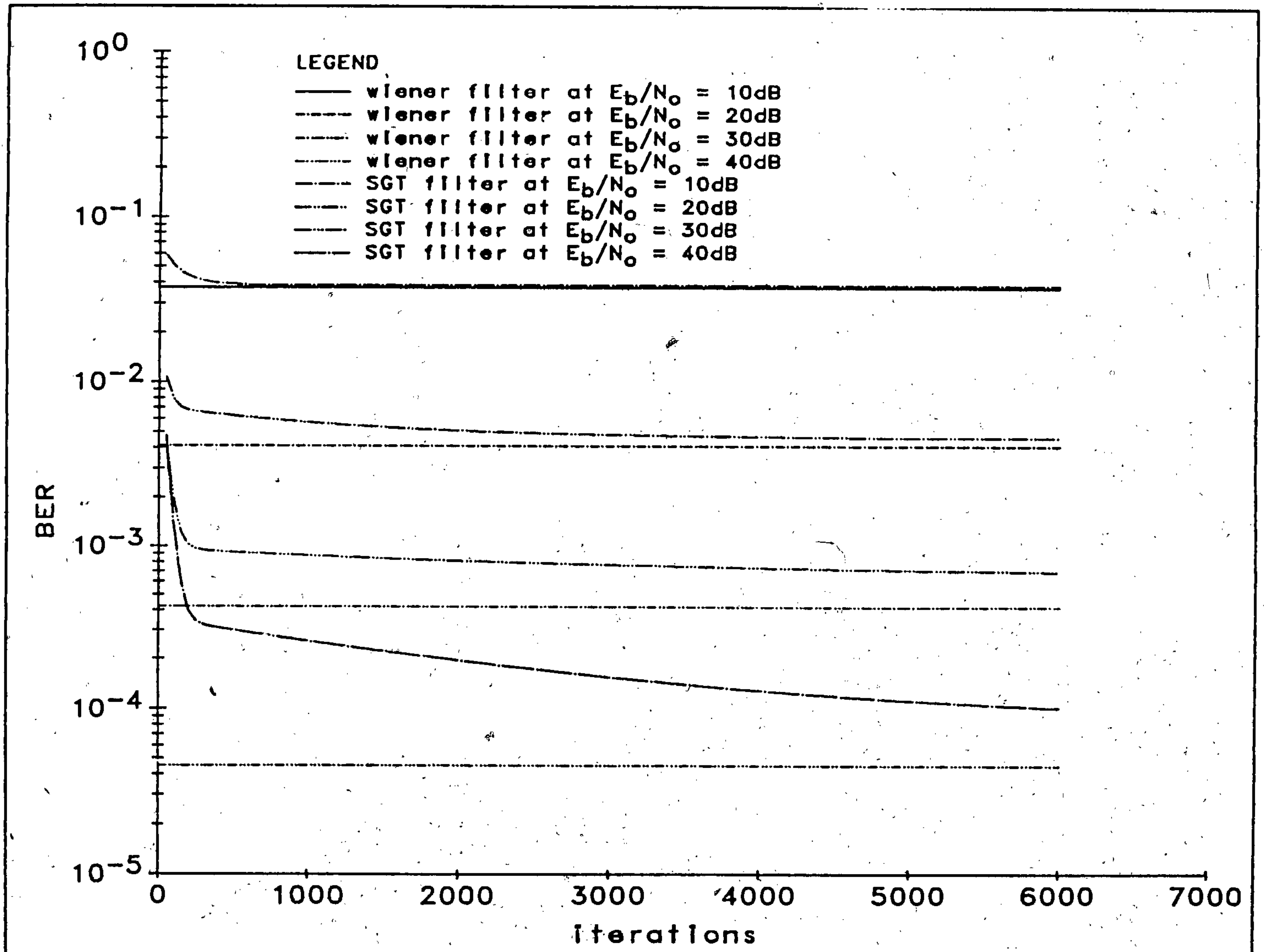


Figure 5.4 - Mean Convergence Curves of a SGT Pilot Filter at Various E_b'/N_o

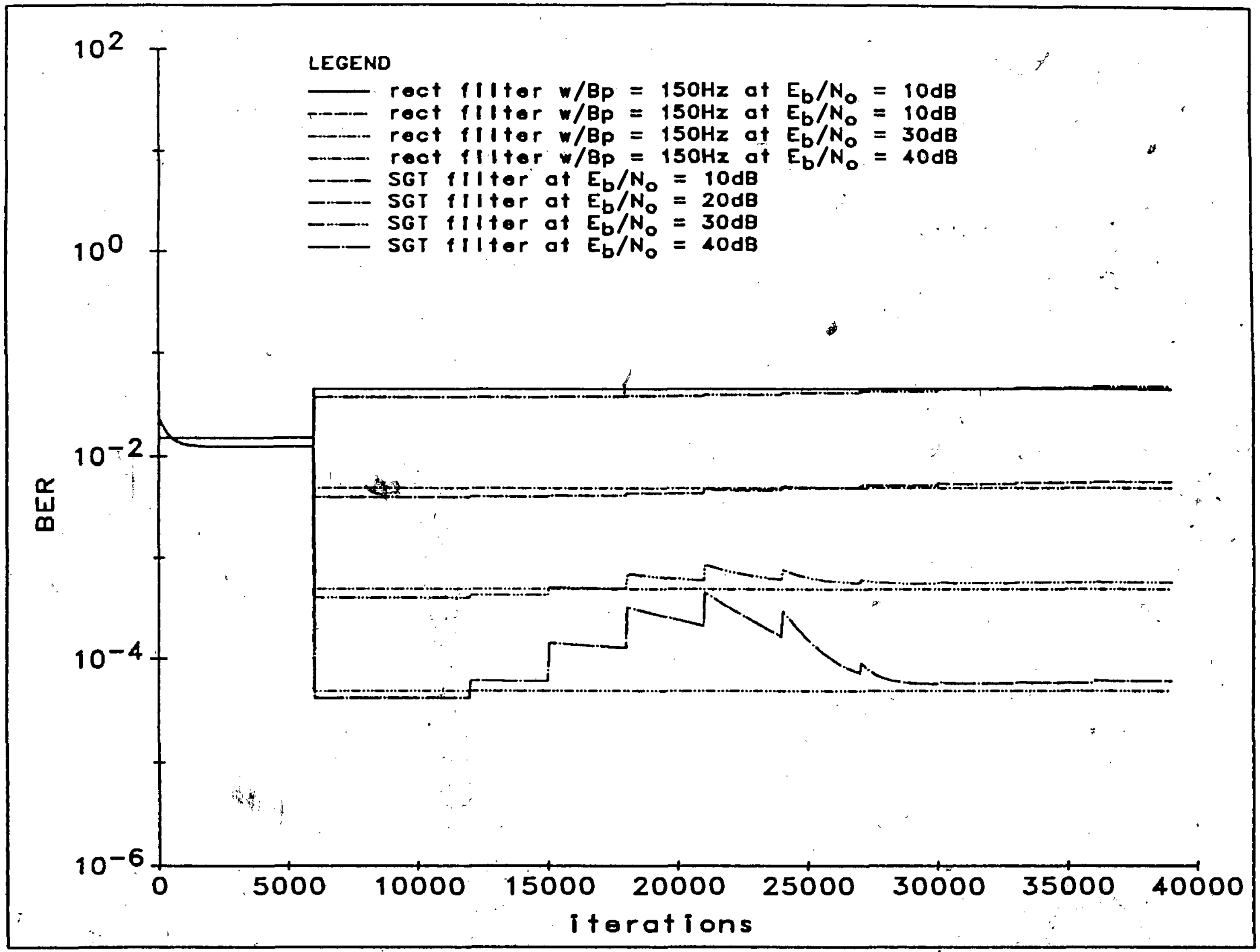


Figure 5.5 - Mean Convergence Curves of a SGT Pilot Filter at Various E_b/N_o With Stepwise Increase in Doppler Frequency

5.5. Simulation Results

Monte Carlo simulations were performed as a final part in the investigation of the use of a SGT pilot filter. The main purposes of the simulations were (1) to verify convergence behavior as predicted by results found in section 5.4.2, (2) to obtain the steady state average BER under various conditions, and (3) to investigate the effect of self-noise and decision direction.

Gain data of the fading process was generated by passing white Pseudo-Noise (PN) sequence through a FIR filter for which the magnitude squared of the frequency response approximated the fade spectrum, S_g. Another PN sequence, which was made independent of the sequence used to generate the gain data, was used to represent additive noise. Received samples were then formed and processed according to the model given in figure 5.6 with 8 samples used to represent each data bit. Note that this figure differs from figure 2.1 in that the removal of data dependence in the reference signal, u(k), has been made perfect. Also, the pilot tone and data signal were transmitted and processed separately so that the results obtained in the simulations were not affected by self-noise. All of the simulations followed the configuration illustrated in figure 5.6 except where the effects of decision direction and self noise were being investigated.

Unless specified otherwise, the following parameters were used for all simulation results given:

E_b/N_o = 20dB, f_DT = 0.0208, f_oT = 0, power split ratio r = 0.2, step size = 0.05, filter length = 5 and MWA length = 3. The SGT filter was forced to be conjugate symmetric using update equation 5.21.

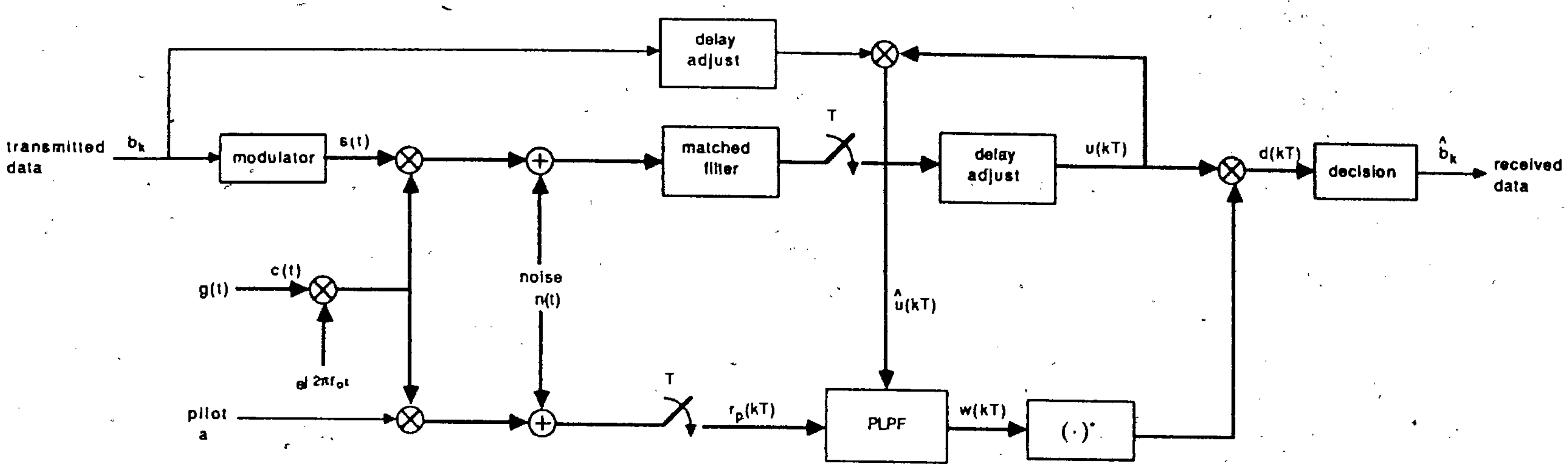


Figure 5.6 - Simulation Model (with Perfect Adaptation Reference and no Self-Noise)

5.5.1. Steady State Average Bit Error Rate

The steady state average BER was obtained by running simulation for a sufficient number of iterations to reach convergence (or close to convergence) and time averaging the post-convergence BER. The BER at each iteration was computed using (3.3) and (5.25).

Effect of step size

Figure 5.7 shows the average BER versus E_b/N_0 for various step sizes, Δ. The average BER increases with increasing Δ as expected because Δ determines the size of the fluctuations of the coefficient vector from the optimum. Clearly, the larger the fluctuations, the higher is the average BER. The amount of excess power required to compensate for the increase in average BER due to the use of noisy gradient, which we will refer to as excess loss, increases with E_b/N_0. For Δ of 0.025 and 0.05, the excess loss was approximately 0.1dB at a BER of 10^{-2} and 0.3dB at 10^{-4}.

Effect of f_DT

The effect of the doppler frequency on the average BER is illustrated in figure 5.8. Here, we can observe that the excess loss increased with increasing f_DT. This makes intuitive sense because variations in the gain are more rapid at a higher f_DT, which causes the MSE gradient estimate to also wander more rapidly. Excess loss at f_DT of 0.00417 was found to be negligible; whereas for f_DT of 0.0208 and 0.0417, the excess losses were 0.2 and 0.5dB respectively. Compared to an ideal rectangular pilot filter with normalized bandwidth of 0.0625, the SGT filter was inferior at all E_b/N_0 for f_DT of 0.0417. For f_DT of 0.0208, SGT filter was better for E_b/N_0 less than 20dB. SGT filter was better for all E_b/N_0 at f_DT of 0.00417. The breakeven doppler frequency was at approximately 0.02.

Effect of f_0T

Frequency offset was found to have negligible effect on the steady state average BER for f_0T of up to 10%. The ability of the SGT filter to compensate for f_0T is only limited by the MWA and the matched filter.

Effect of Filter Length and Moving Window Averager

Figure 5.9 gives the average BER versus E_b/N_0 using various filter and MWA lengths. We can see from figure 5.9 that the use of filter length greater than 11 should be avoided. The best combination of filter length and MWA overall was 5 and 5 respectively. Longer MWA produced better results. However, it should be remembered that a longer MWA reduces average BER at the expense of decreased correction range for frequency offset.

Effect of Decision Direction and Self-Noise

Previous results had been obtained without the effects of decision direction and self-noise. We investigated the effects of decision direction by using demodulator decisions to remove the data dependence in u(k). The effects of self-noise had also been selectively included in the simulation by combining the pilot tone and data signal in the transmitter as illustrated by the system model given in figure 2.1. The results are illustrated in figure 5.10 which shows the BER performance with and without the effect of decision direction and self-noise. A reference training sequence of 20 bits was used for the simulations using decision direction. Effect of decision direction is not discernible in figure 5.10. However, examination of BER values revealed a slight increase in average BER at low E_b/N_o. This is to be expected as there are more decision errors at low E_b/N_o. Self-noise had little effect at low E_b/N_o. But at high E_b/N_o, self-noise caused a large increase in average BER. The amount of self-noise is not dependent on E_b/N_o so that bit errors at low E_b/N_o are dominated by additive noise whereas at high E_b/N_o, bit errors are dominated by self-noise. The excess loss due to self-noise at a BER of 10^{-4} was approximately 3dB. It should be noted that this numerical results is for the use of Manchestor coding in creating the spectral null required for placing the pilot. Other techniques such as phase-locked TTIB [9] can provide a much smaller excess loss due to self-noise.

Comparison with Complex Filter

The benefit of forcing conjugate symmetry on the coefficient vector is illustrated by figure 5.11 which compares average BER for filters using update equations 5.20 and 5.21. A forced conjugate symmetric filter was found to be better at all E_b/N_o. Additional excess loss for a complex filter is negligible at BER higher than 10^{-3}. For BER of 10^{-4}, the additional excess loss is 0.5dB. It was also found that the additional excess loss was greater for longer filter.

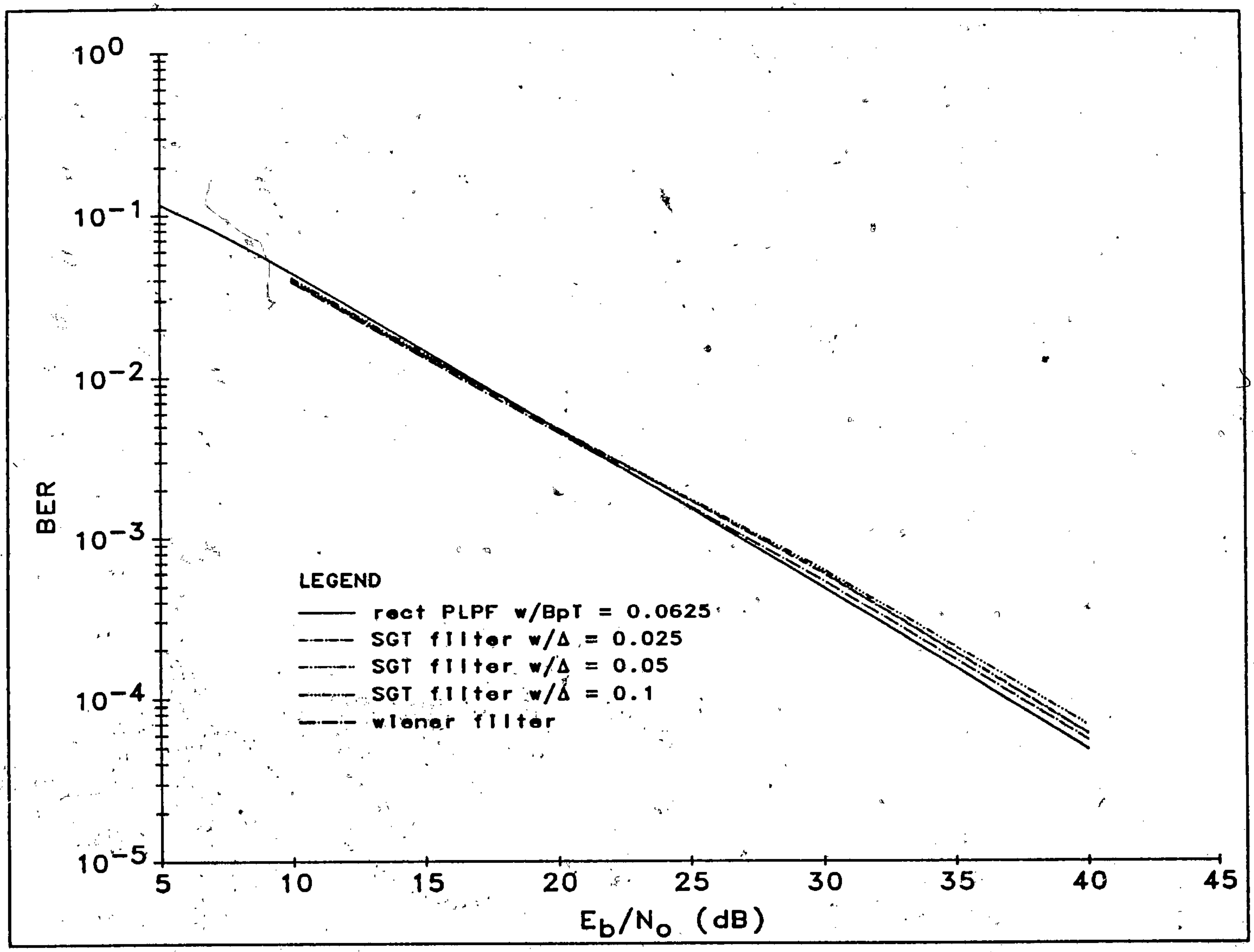


Figure 5.7 - Effect of Step Size on the Average BER For the SGT Pilot Filter

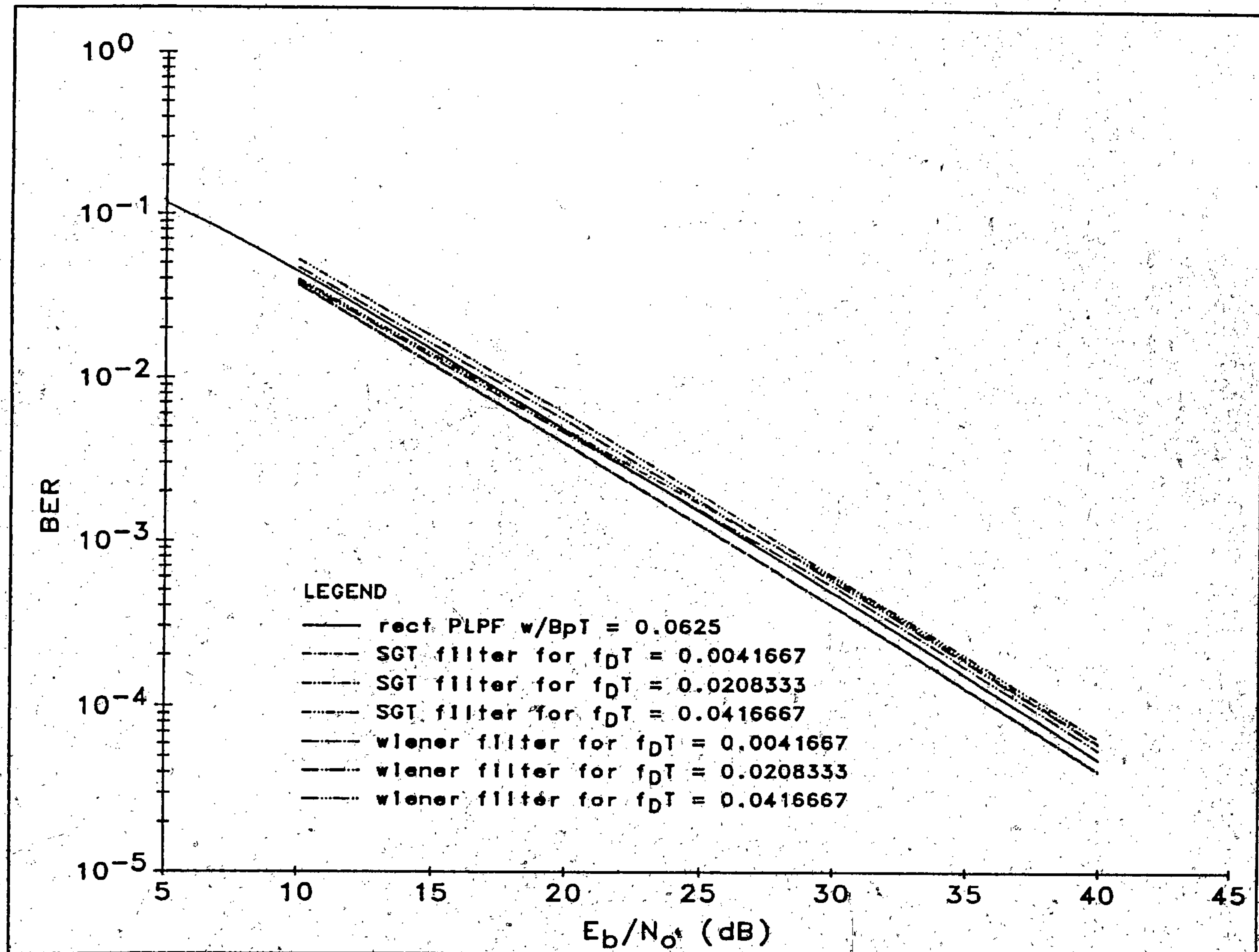


Figure 5.8 - Average BER Performance of the SGT Pilot Filter at Various Doppler Frequencies

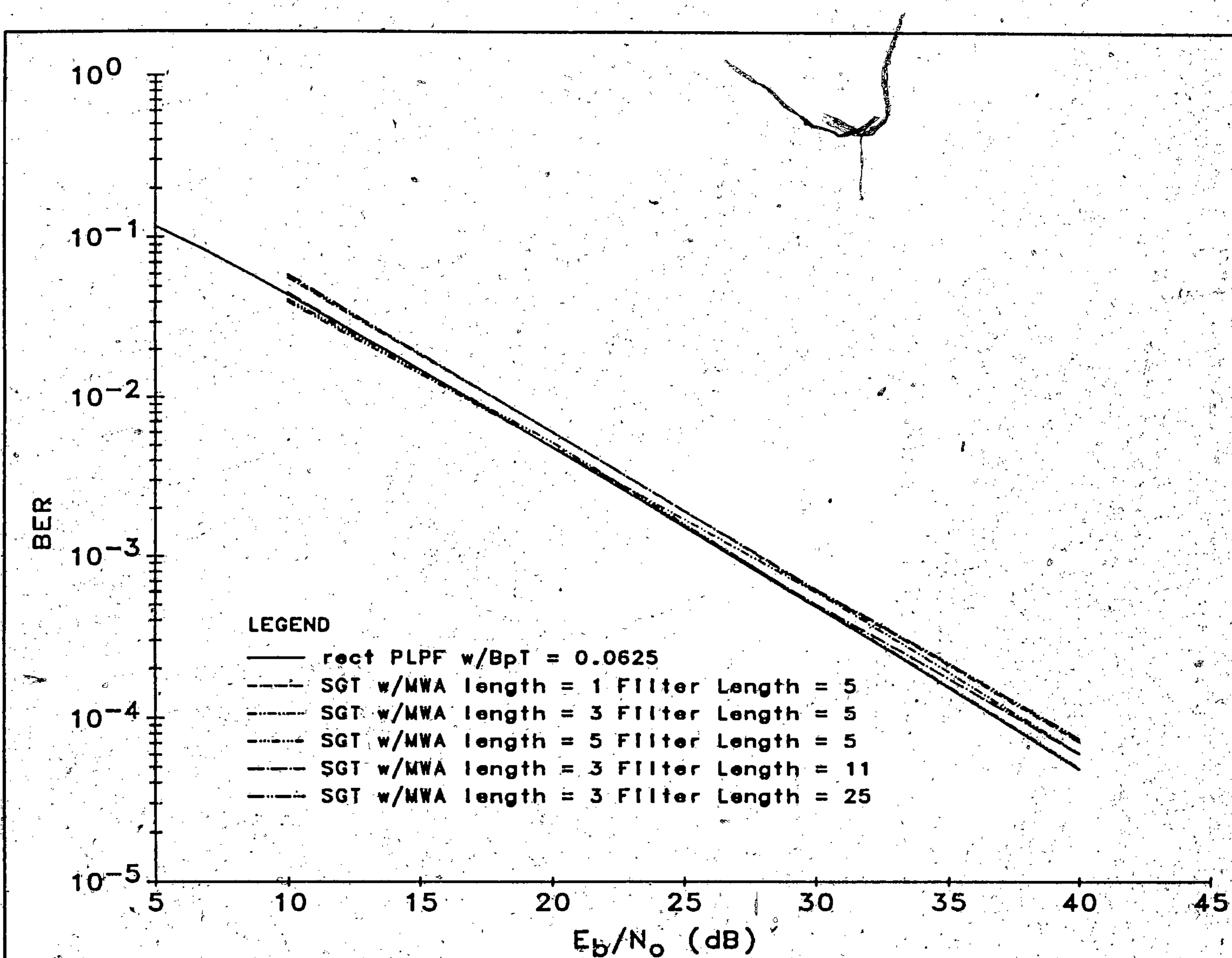


Figure 5.9 - Average BER Performance of the SGT Pilot Filter with Various Filter Lengths and Moving Window Averager Lengths

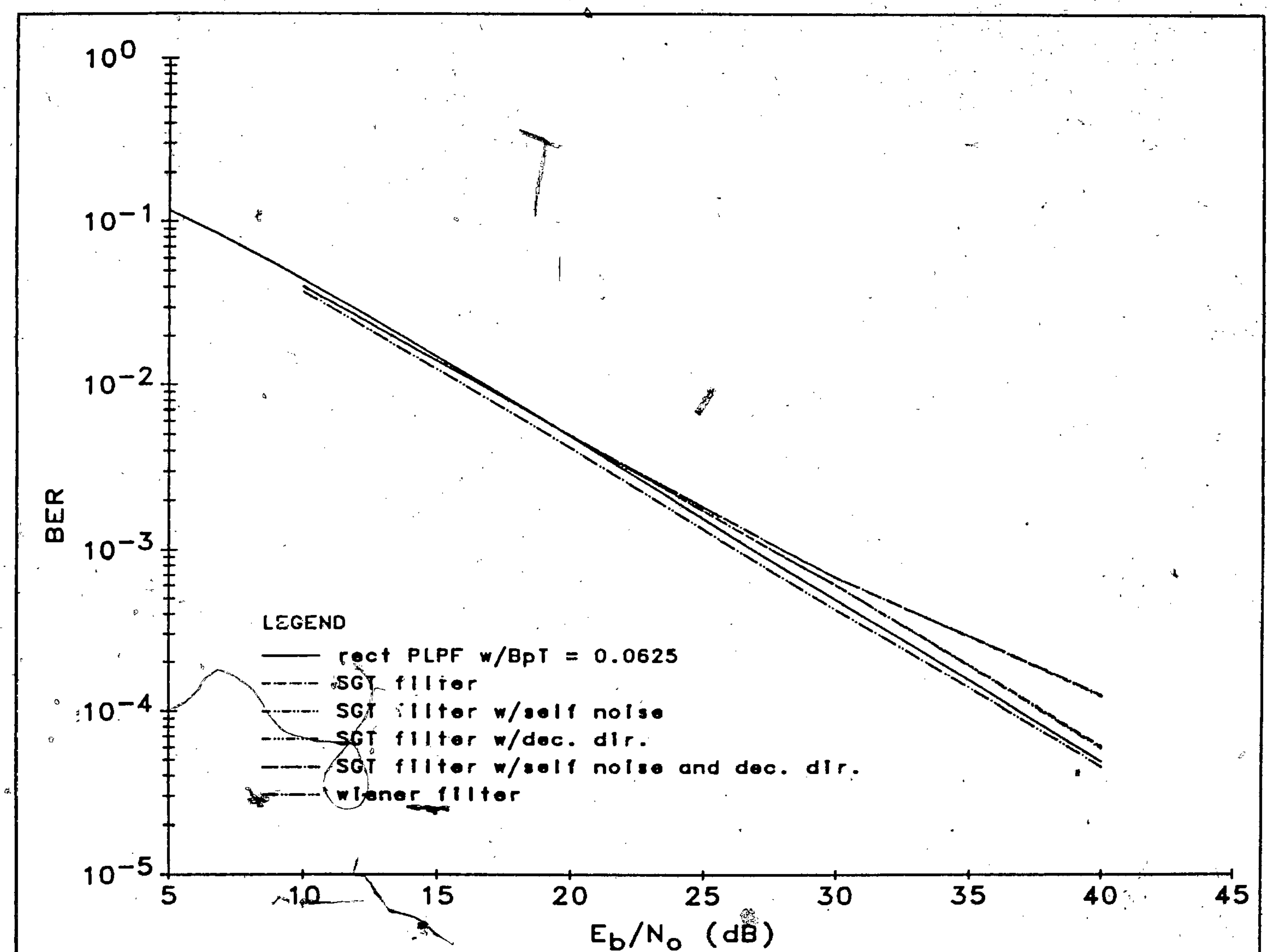


Figure 5.10 - Effect of Self-noise and Decision Direction on the Average BER For the SGT Pilot Filter

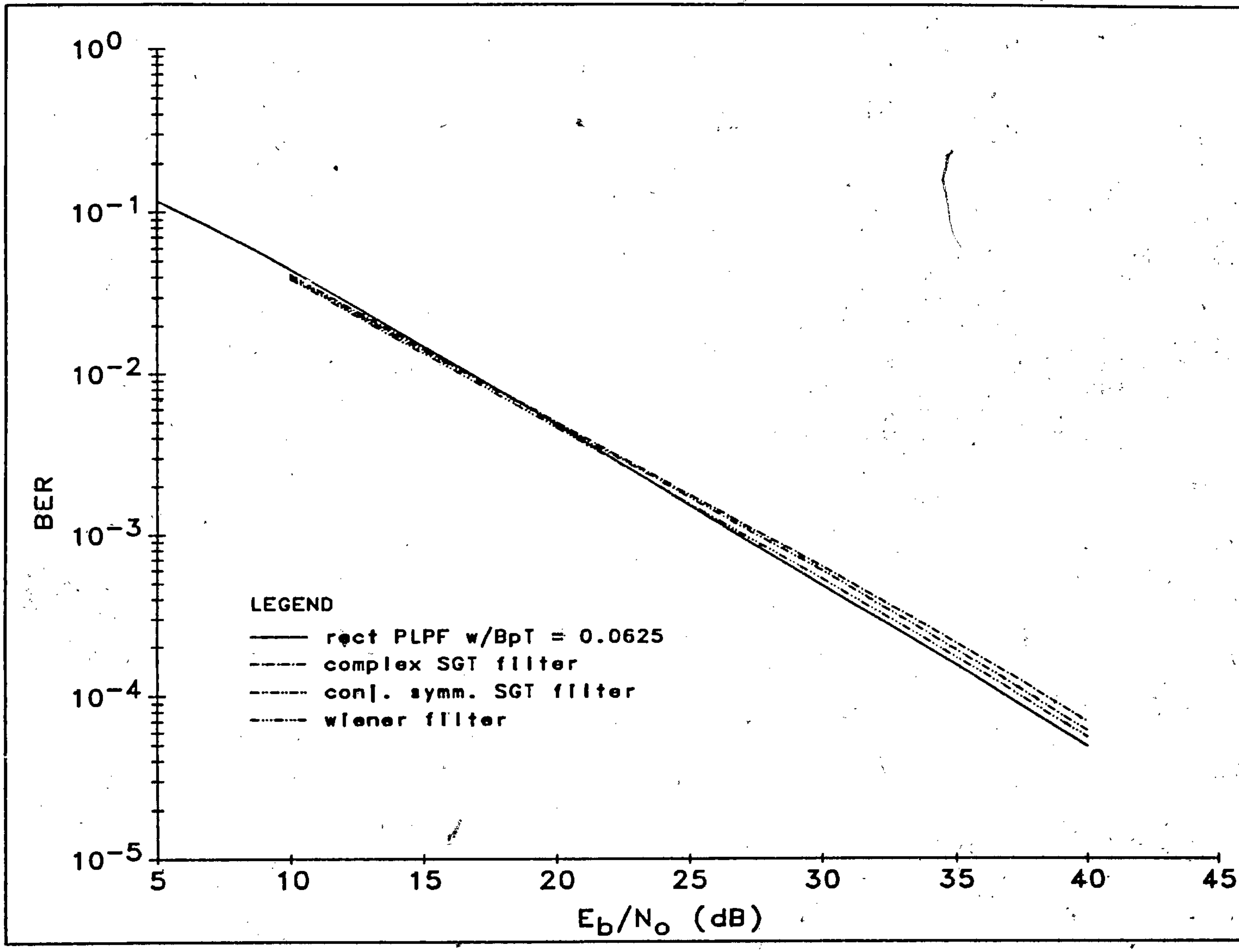


Figure 5.11 - Comparison of Average BER Performance Between a Complex SGT Pilot Filter and a Conjugate Symmetric SGT Pilot Filter

5.5.2. Convergence Time

In comparing convergence performance, it is convenient to use convergence time. For the present application, we define convergence time as the number of iterations required for the instantaneous BER to drop below the steady state average BER for the first time. It should be noted that convergence time depends heavily on the choice of initial coefficient vector, especially at high E_b/N_0. The results obtained in this section utilized an asymmetric initial vector of [1+j, 0, 0, -1+j, 0].

Effect of step size

Figure 5.12 shows the BER learning curve for different step sizes. The convergence time as a function of various step sizes, Δ, are tabulated in table 5.1. Convergence time decreased with Δ as expected. One should note from figure 5.12 that the magnitude of the random fluctuations also increased with Δ.

Δ	Conv. time
0.025	1000
0.05	1850
0.1	3600

Table 5.1 - Convergence Time of the SGT Filter as a function of the Step Size

Effect of E_b/N_o

From results found in section 5.4, convergence speed is expected to increase significantly with increased E_b/N_o. This was found to be true as illustrated by figure 5.13. Table 5.2 gives the convergence time as a function of E_b/N_o. Note the slow convergence for E_b/N_o of 30 and 40dB. For comparison, the computed learning curves for mean convergence are also shown in figure 5.13. The computed and simulation results are indeed very close.

E_b/N_o	Conv. time
10	400
20	1850
30	16100
40	>100000

Table 5.2 - Convergence time of the SGT Filter as a function of E_b/N_o

Effect of f_DT

The effect of f_DT is illustrated in figures 5.14 and 5.15. Figure 5.14 shows the learning curves for various doppler frequencies. The convergence time as a function of f_DT is tabulated in table 5.3. Convergence time decreases with f_DT as expected. From figure 5.14, it is also evident that the amount of random fluctuations of BER increases as f_DT is increased. Figure 5.15 shows the simulated and computed (mean) convergence curves for stepwise increase of f_DT simulating a vehicle acceleration of 10 kmph/sec at 40dB. Simulated results were indeed found to be very close to the computed results. One point to remember is that 40dB is an unrealistically high E_b/N_o value. This value is used here only to accentuate the effect of changing f_DT.

f_DT	Conv. time
0.00417	2600
0.0208	1850
0.0417	1500

Table 5.3 - Convergence time of the SGT Filter as a function of Normalized Doppler Frequency

Effect of f_0T

Figure 5.16 shows learning curves for various frequency offsets. f_0T has no noticeable effect on convergence except for f_0T of 0.1667 where BER is deteriorated by distortion of the received pilot by the MWA.

Effect of Filter Length

Increasing the filter length has the effect of increasing the amount of fluctuations of BER as shown in figure 5.17. This is because long filter increases the susceptibility to gradient fluctuations. Convergence time was also found to increase with increasing filter length. So, in terms of convergence behavior, a shorter filter is better.

Effect of Moving Window Averager Length

There was an increase in random fluctuations in BER due to increase in MWA length. However, change in convergence time was small as the length of the MWA was increased.

Effect of Self-Noise

The effects of self-noise on convergence behavior are shown in figures 5.18 and 5.19 for filter lengths of 5 and 11 respectively. Convergence time was found to change little with the presence of self-noise. However, self-noise caused a greater fluctuations of BER at high E_b/N_o and hence deteriorated the steady state average BER as found earlier. Comparison between figures 5.18 and 5.19 also showed that self-noise had a much greater impact on longer filter.

Effect of Decision Direction

Figure 5.20 shows learning curve with decision direction at various E_b/N_o. There was some degradation in the BER when E_b/N_o was low. Overall, changes in the convergence time due to decision direction were negligible.

Comparison with Complex Filter

Convergence was found to be unaffected by enforcing conjugate symmetry in the coefficient vector.

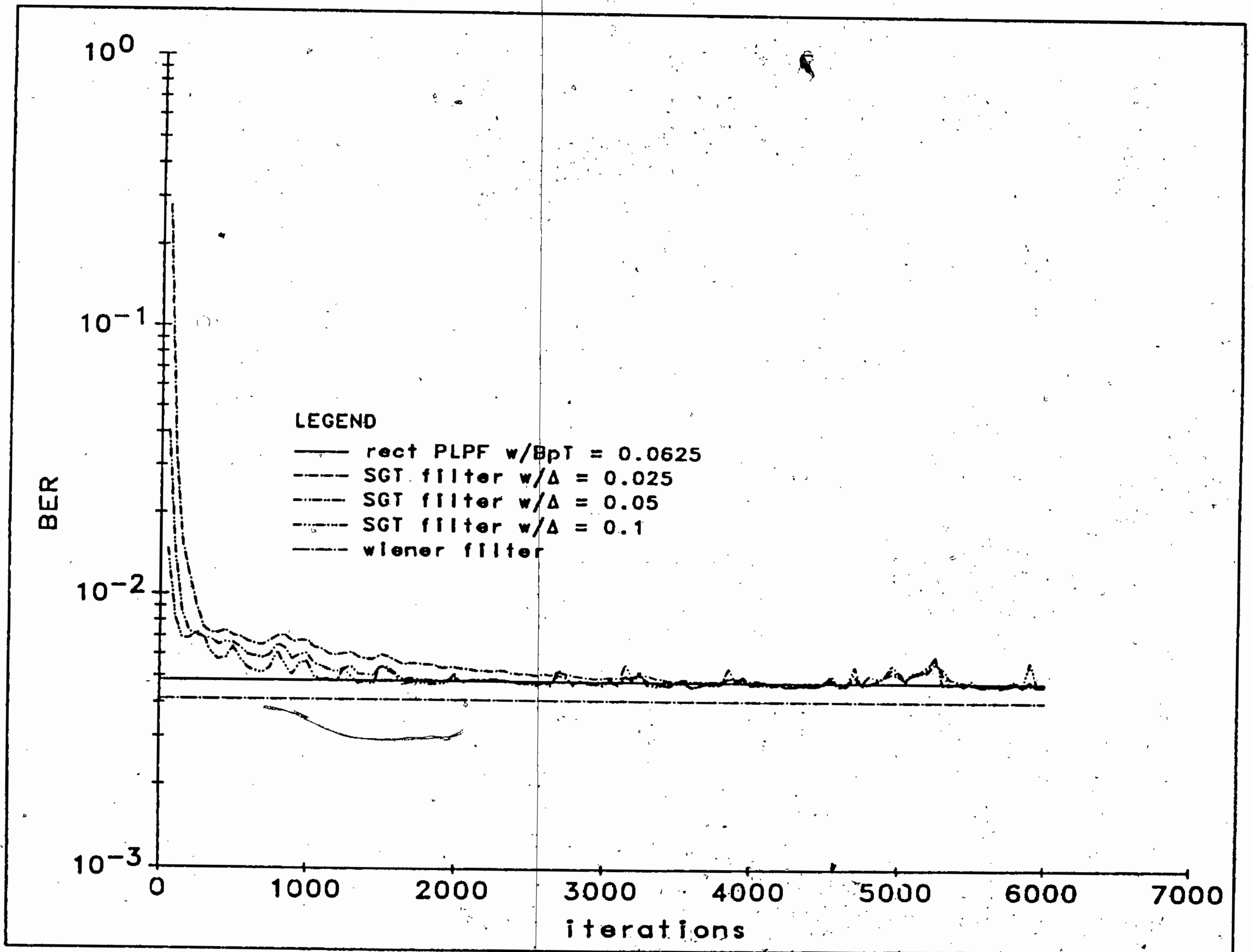


Figure 5.12 - Learning Curves of the SGT Pilot Filter with Various Step Sizes

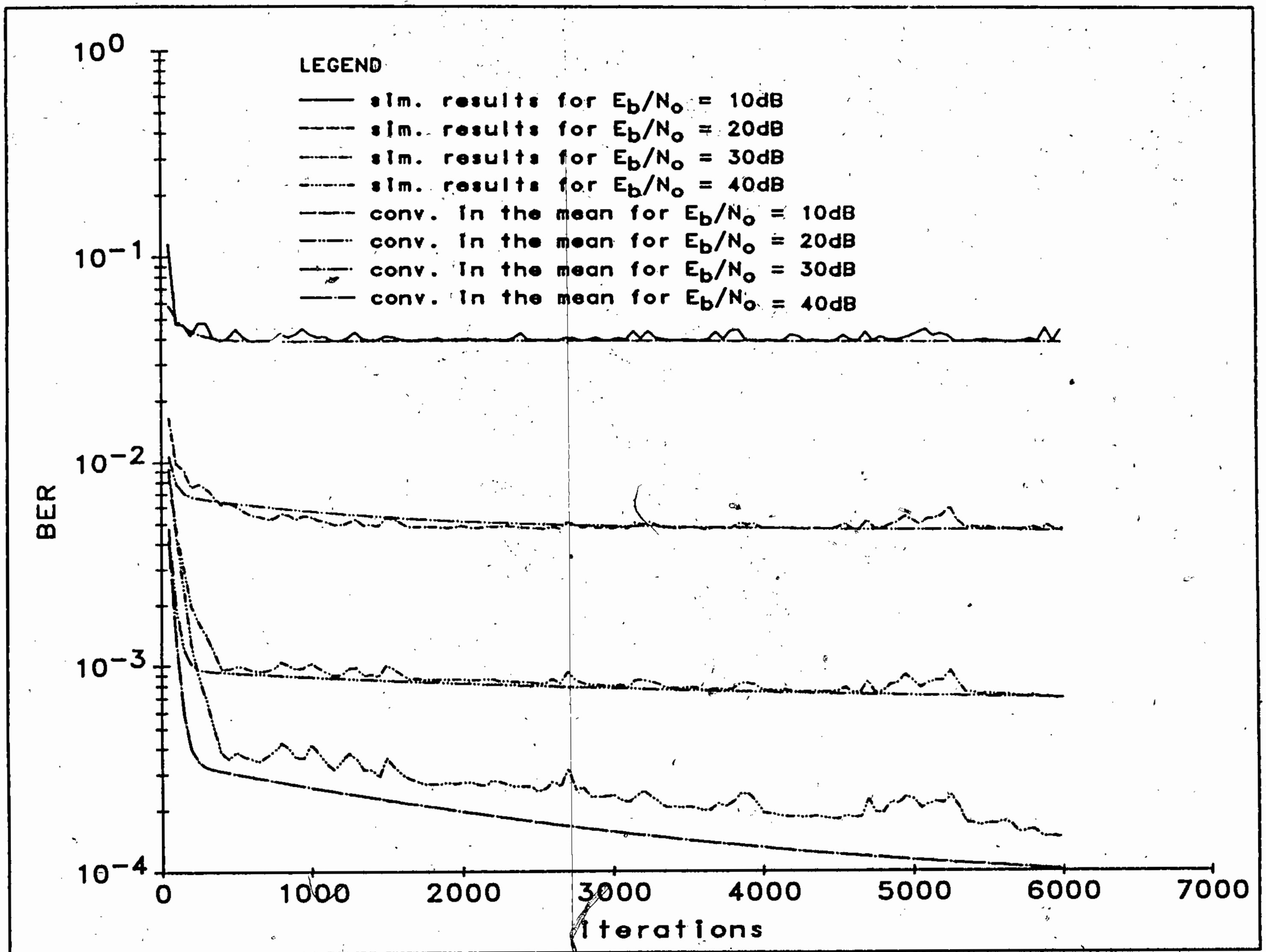


Figure 5.13 - Learning Curves of the SGT Pilot Filter at Various E_b/N_o

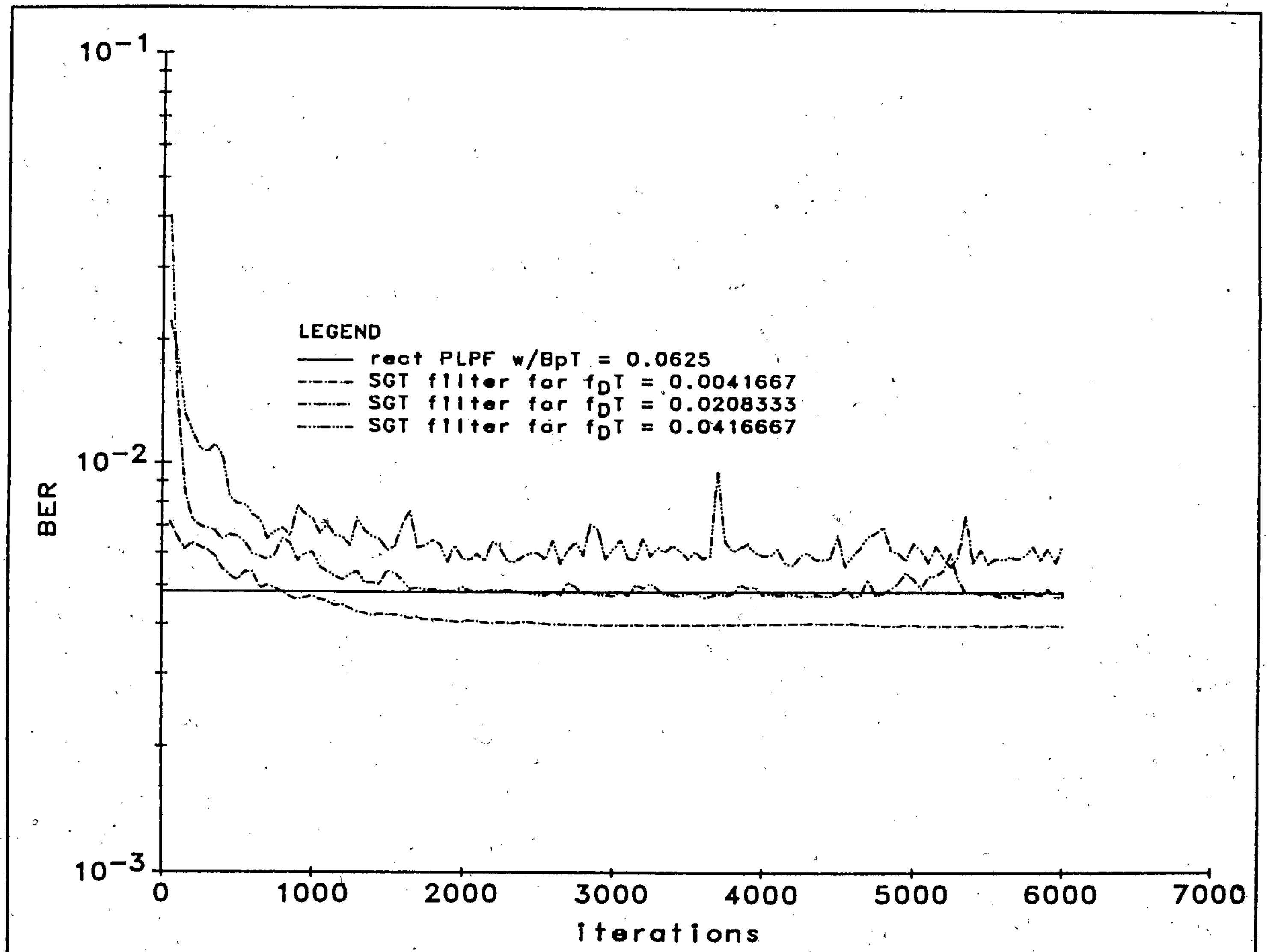


Figure 5.14 - Learning Curves of the SGT Pilot Filter at Various Doppler Frequencies

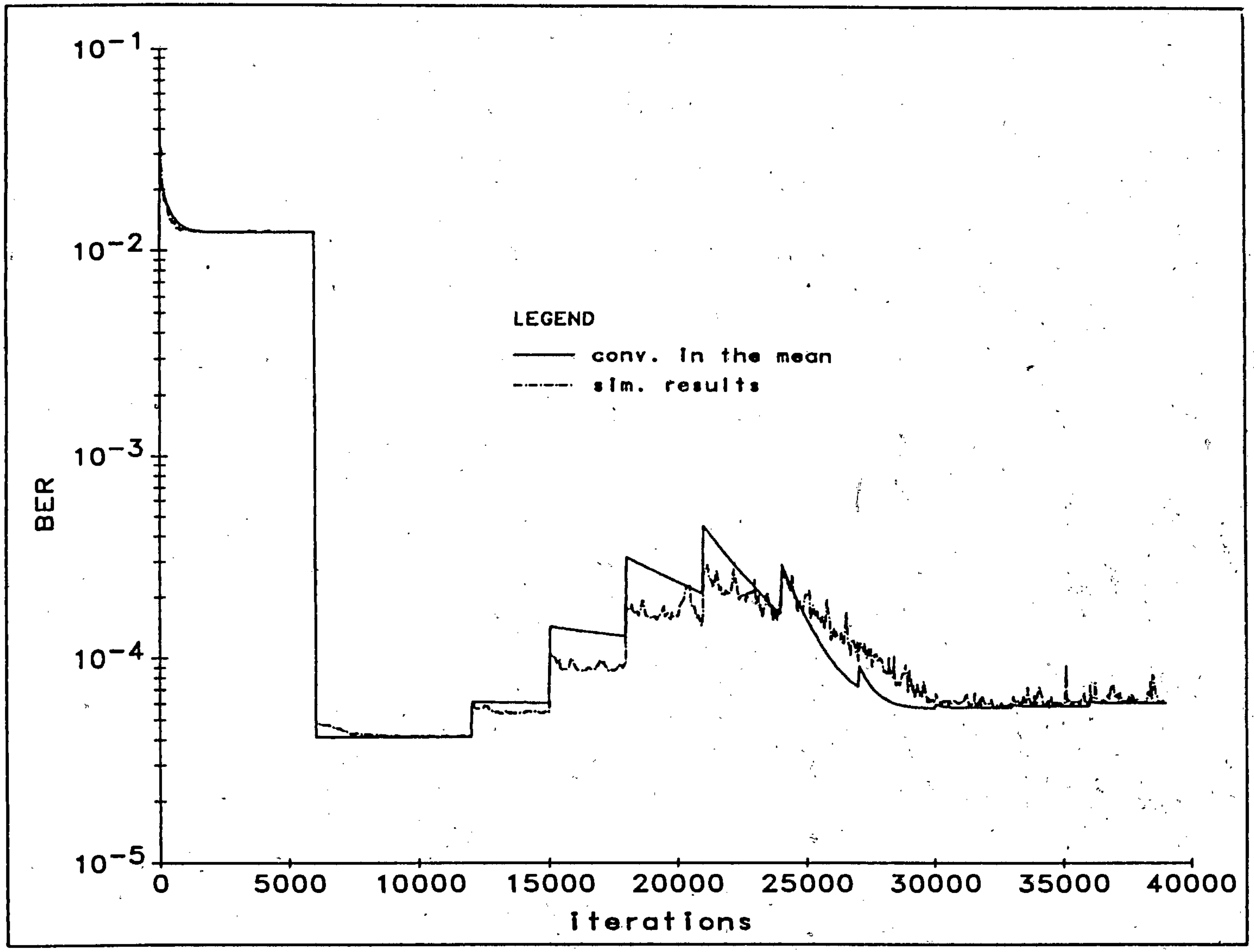


Figure 5.15 - Learning Curves of the SGT Pilot Filter With Stepwise Increase in Doppler Frequencies

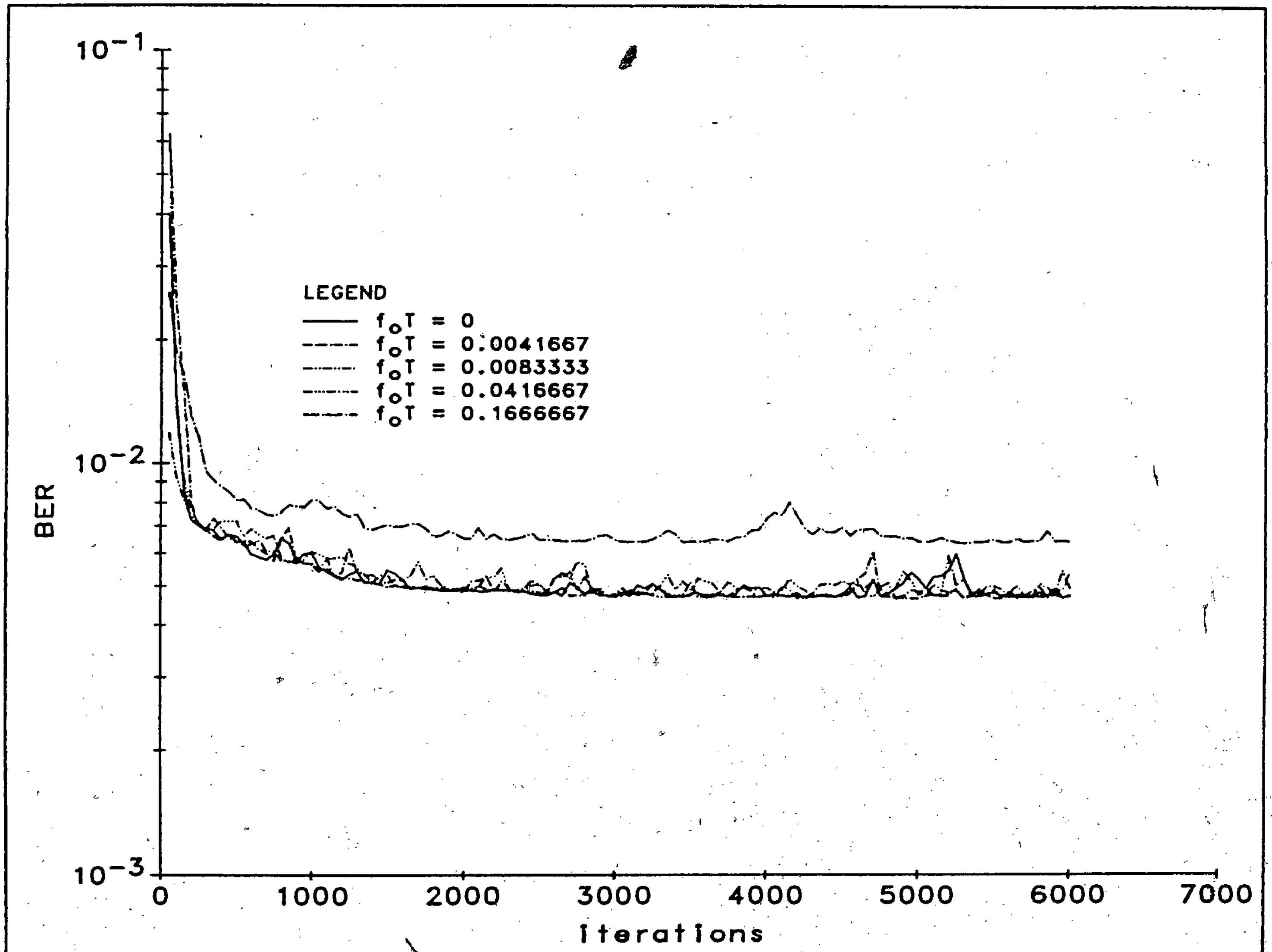


Figure 5.16 - Learning Curves of the SGT Pilot Filter with Various Frequency Offsets

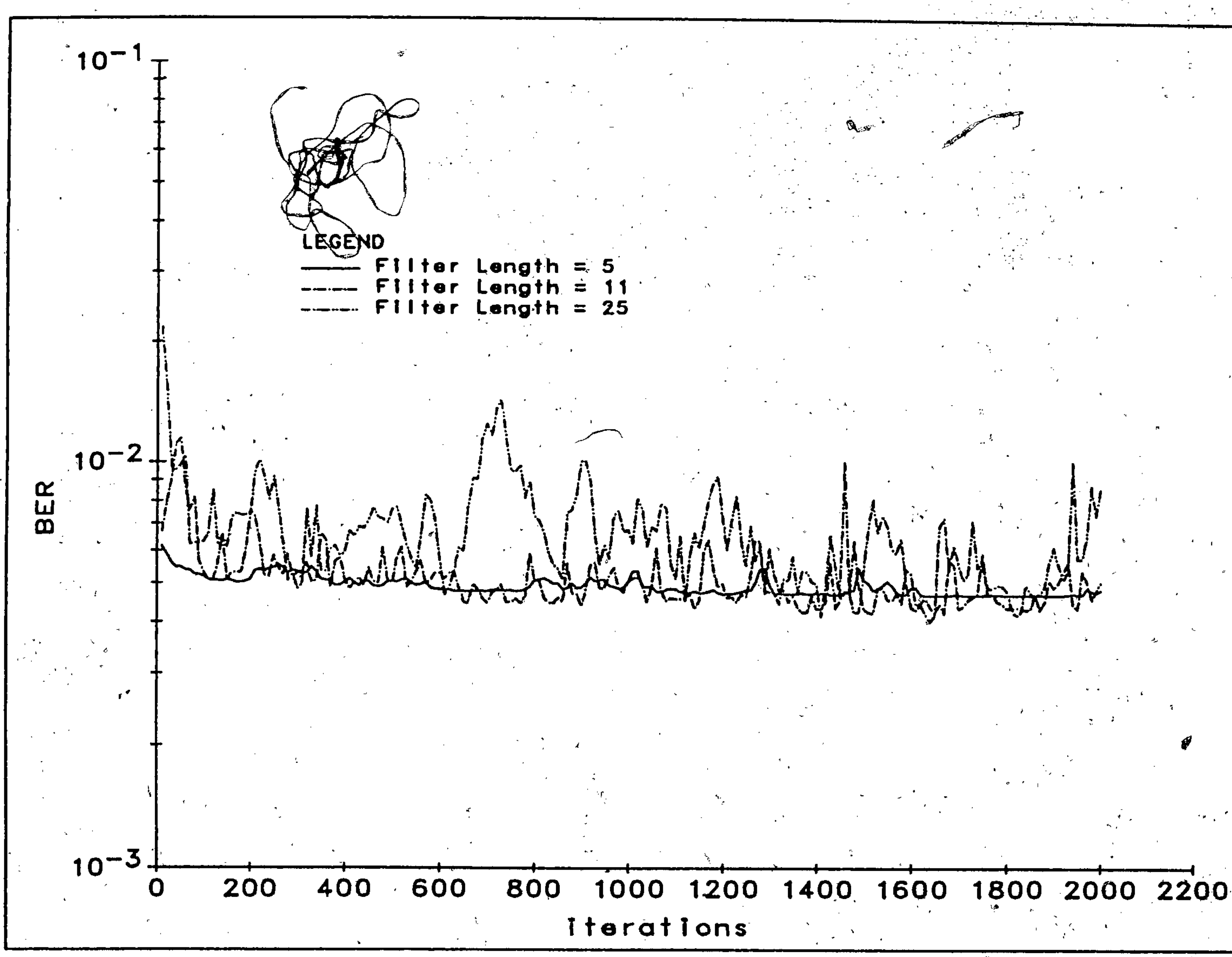


Figure 5.17 - Learning Curves of the SGT Pilot Filter with Various Filter Lengths

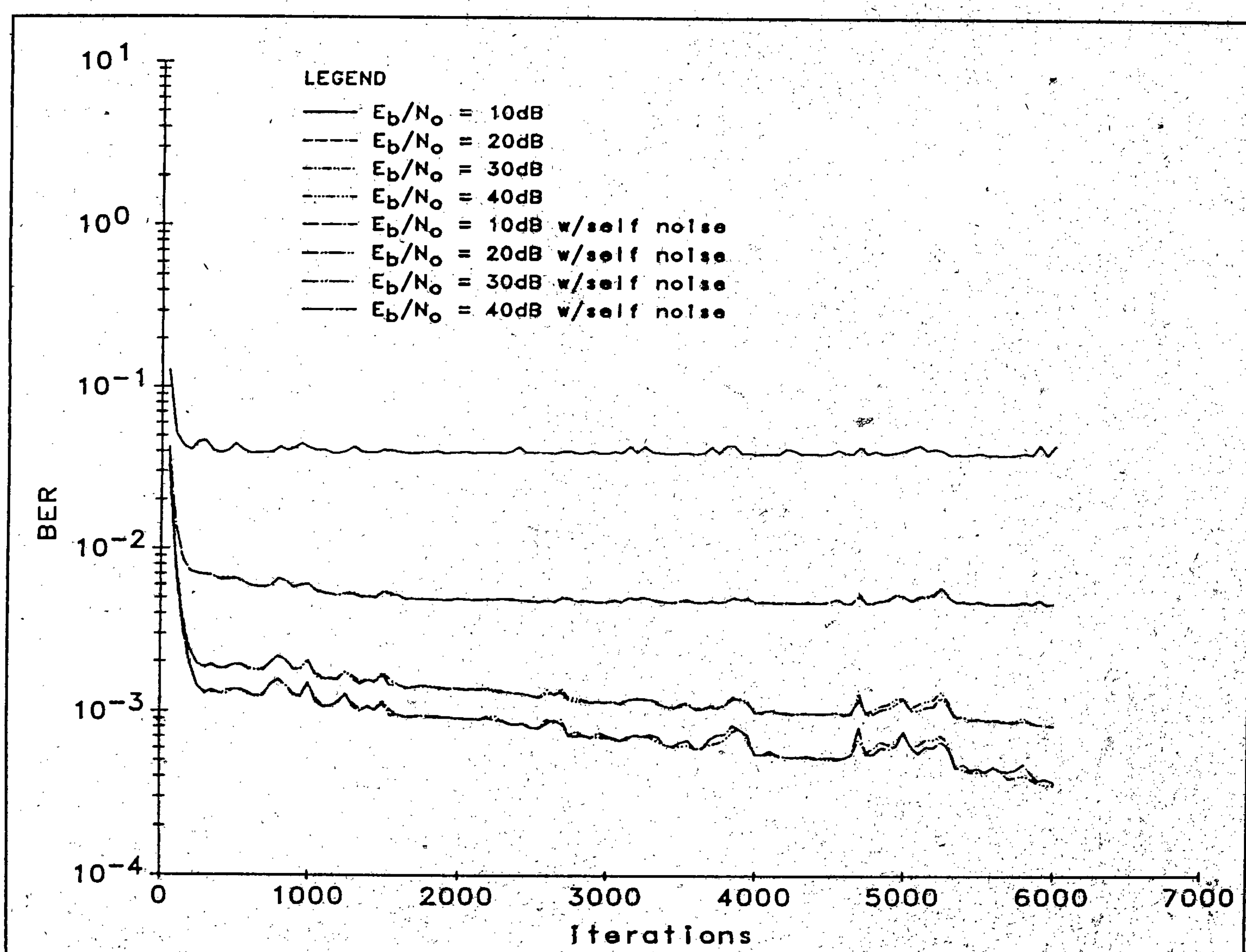


Figure 5.18 - Effect of Self-Noise on the Convergence of the SGT Pilot Filter with a Filter Length of 5

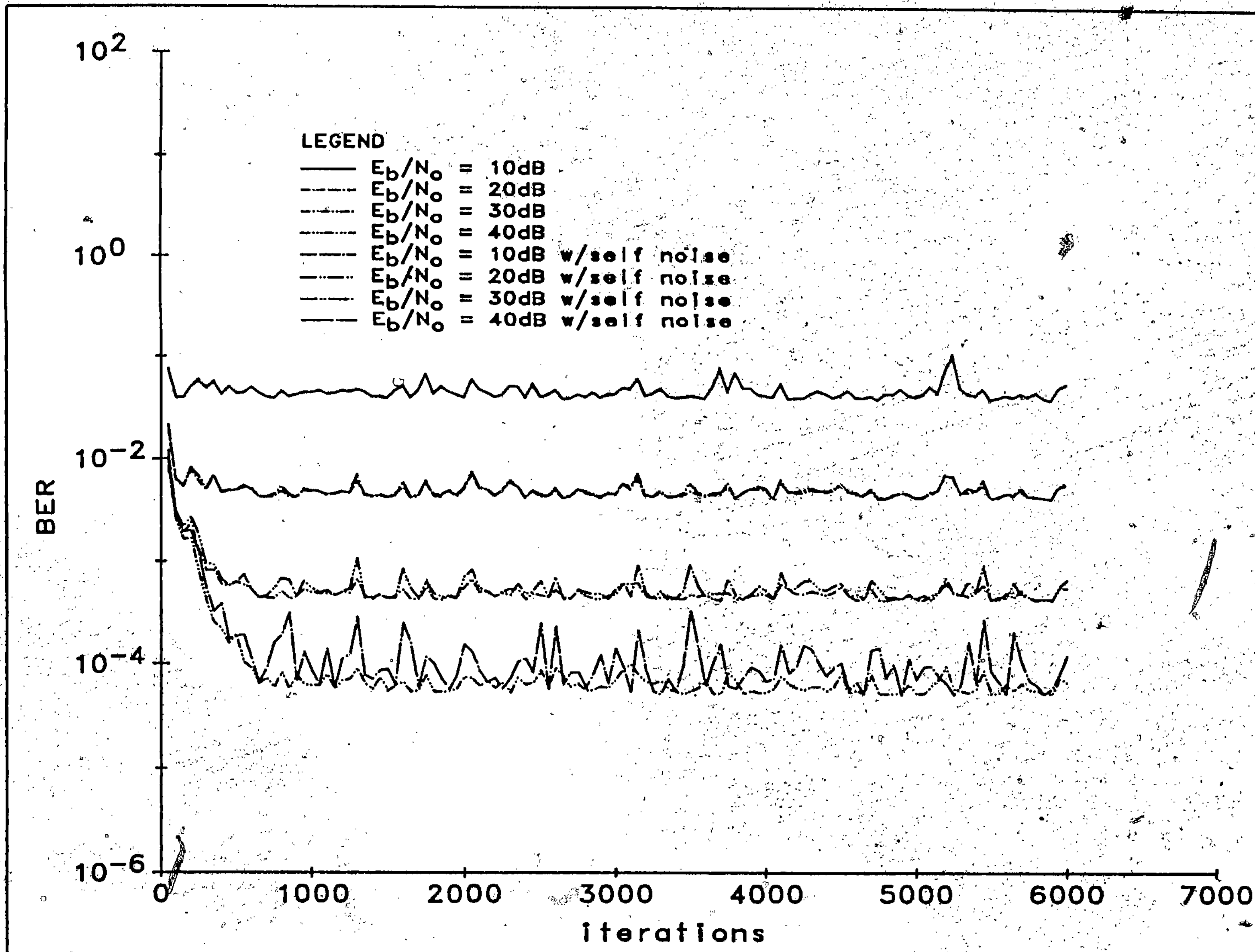


Figure 5.19 - Effect of Self-Noise on the Convergence of the SGT Pilot Filter with a Filter Length of 11

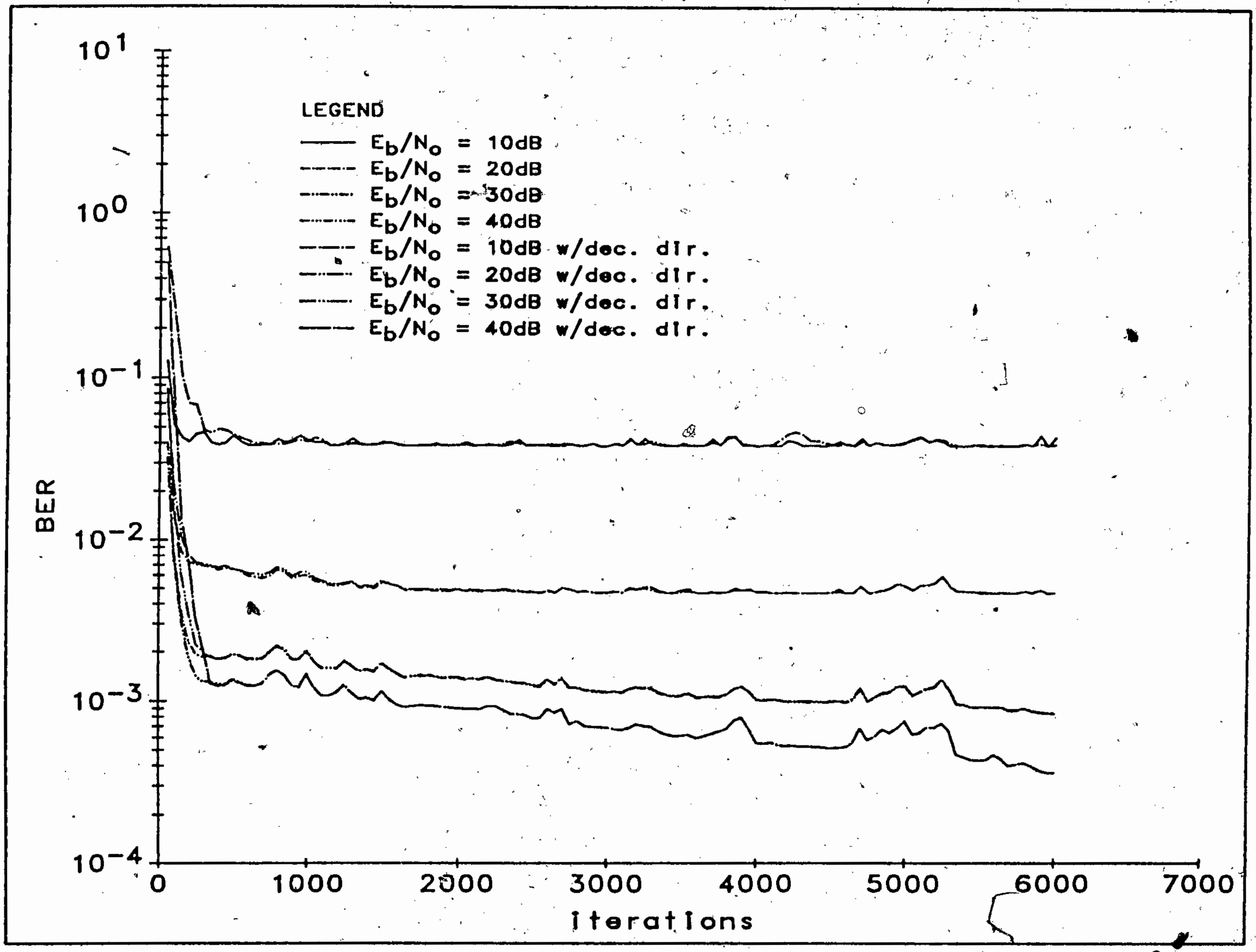


Figure 5.20 - Effect of Decision Direction on the Convergence of the SGT Pilot Filter

5.6. Summary and Comments

For E_b/N_0 less than 20dB, the SGT filter was found to be able to adapt itself quickly enough to follow changes in vehicle speed assuming that data is continuously transmitted. When the normalized doppler frequency is less than 2% (corresponding to a vehicle speed of 60 kmph for a 2400bps system operating at 850Mhz), the use of a SGT pilot filter can provide a lower BER than using an ideal rectangular filter with a normalized bandwidth of 0.0625. So, the SGT filter is suitable for use in an urban environment where vehicle speed is expected to be low. However, if the doppler is greater than 2%, then using a fixed filter will give better BER performance. Frequency offset, decision direction and self-noise were found to have negligible effect on convergence.

6. STOCHASTIC GRADIENT LATTICE PILOT TONE FILTERING

Recall that the main drawback of the SGT filter is the dependence of convergence speed on the eigenvalue spread of the input correlation matrix. There are other algorithms which have been demonstrated to have overcome this problem [13]. One such algorithm is the stochastic gradient lattice joint process estimator (SGL-JPE) which uses a lattice filter in addition to a transversal structure. This chapter explores the possible use of the SGL-JPE as pilot filter.

The SGL-JPE was first proposed by Makhoul [19] and Griffiths [12] in the context of noise cancellation. The main idea behind the SGL-JPE is to overcome the problem of eigenvalue spread of the SGT filter by preceding it with a decorrelator. The lattice filter is ideal for use as a decorrelator because the backward prediction error outputs of each of its stages are orthogonal. Thus, instead of forming the LMS estimate using delayed samples of the input signal directly, we first pass the input signal through a lattice filter and then form the LMS estimate using a linear combination of the backward prediction errors. Because of the orthogonality of the backward prediction errors, their eigenvalues are easy to estimate. (The eigenvalues are simply equal to the corresponding prediction error powers.) An appropriate step size can then be used to update each coefficient used to form the LMS estimate, resulting in a vast improvement in convergence speed. Unfortunately, it will be shown later than this improvement in convergence speed is at the expense of large degradation in BER due to the sensitivity of the BER to the filter coefficients.

The lattice filter is closely associated with linear prediction theory. Hence, it is natural to begin this section with a discussion of linear prediction theory. Following this, brief derivations of the important expressions related to the SGL-JPE will be given. Procedure for computing the

BER given the lattice and JPE coefficients are discussed. Some numerical analysis of the sensitivity of the BER to the lattice and JPE coefficients will then be presented followed by simulation results.

6.1. Linear Prediction Theory

Linear prediction deals with by the use of a linear filter on past samples of a stochastic process to predict a future value. In one form, the prediction problem can be formulated as:

$$\hat{x}(k) = \sum_{n=1}^{M} y_{M,n}^{*}\ x(k-n) \tag{6.1}$$

This formulation is called forward prediction. $x(k)$ is the sample of the stochastic process at time kT and $y_{M,n}$ is the n-th forward prediction coefficient for an order M predictor. Another form, called backward prediction, uses values $x(k)$ ... $x(k-M-1)$ to make prediction of the sample $x(k-M)$ such that:

$$\hat{x}(k-M) = \sum_{n=1}^{M} c_{M,n}^{*}\ x(k-n+1) \tag{6.2}$$

where $c_{M,n}$ is the backward prediction coefficient. Defining $f_M(k)$ and $b_M(k)$ as the forward and backward prediction errors of order M, then $f_M(k)$ and $b_M(k)$ are given by:

$$\begin{aligned} f_M(k) &= x(k) - \hat{x}(k) \\ &= \sum_{n=0}^{M} g_{M,n}\ x(k-n) \end{aligned} \tag{6.3}$$

and

$$b_M(k) = x(k-M) - \hat{x}(k-M) = \sum_{n=0}^{M} a_{M,n}\ x(k-n+1) \tag{6.4}$$

where

$$g_{M,n} = \begin{cases} 1 & n = 0 \\ -y_{M,n}^{*} & n = 1, \ldots, M \end{cases} \tag{6.5}$$

and

$$a_{M,n} = \begin{cases} -c_{M,n}^{*} & n = 0, \ldots, M-1 \\ 1 & n = M \end{cases} \tag{6.6}$$

$g_{M,n}$ and $a_{M,n}$ are the n-th forward and backward prediction error coefficients of order M. The FIR filters in which they represent are called forward and backward prediction error filters.

Equations 6.3 and 6.4 can be solved to minimize the mean square forward and backward prediction errors. The results are two normal equations identical in form to (5.18) from wiener filter theory and are given by:

$$\mathbf{R}\ \mathbf{y}_o = \mathbf{s}_f \tag{6.7}$$

and

$$\mathbf{R}\ \mathbf{c}_o = \mathbf{s}_b \tag{6.8}$$

where **R** is the correlation matrix of the input process; $\mathbf{y_o}$ and $\mathbf{c_o}$ are the optimal forward and backward prediction coefficient vectors; $\mathbf{s_f}$ and $\mathbf{s_b}$ are the forward and backward correlation vectors. Denoting $\mathbf{x}(k-1)$ as the input vector such that:

$$\mathbf{x}(n-1) = [x(k-1), x(k-2), ..., x(k-M)]^{\mathbf{T}} \tag{6.9}$$

then **R**, $\mathbf{s_f}$ and $\mathbf{s_b}$ are given by:

$$\mathbf{R} = E[\mathbf{x}(k)\ \mathbf{x}^{\mathbf{H}}(k)] \tag{6.10}$$

$$\mathbf{s_f} = E[\mathbf{x}(k-1)\ x^{*}(k)] = [s(-1), s(-2), ..., s(-M)]^{\mathbf{T}} \tag{6.11}$$

$$\mathbf{s_b} = E[\mathbf{x}(k)\ x^{*}(k-M)] = [s(M), s(M-1), ..., s(1)]^{\mathbf{T}} \tag{6.12}$$

where s(k) is the autocorrelation function of x with a lag of kT.

Recall that for the optimum filter, the estimation error vector is normal to the filter output vector. Using this fact and along with (6.4) and (6.6), one can show that the sequence of backward prediction errors $b_0, b_1, ..., b_M$ are all orthogonal to one another when the optimal prediction coefficients are used [15]; i.e.:

$$E[b_m(k)\ b_i^{*}(k)] = \begin{cases} 0 & m \neq i \\ P_m & m = i \end{cases} \tag{6.13}$$

where P_m is the prediction error power of order m. This orthogonality makes the convergence speed of the SGL-JPE insensitive to eigenvalue spread.

An efficient technique for solving the normal (6.7) and (6.8) exists and is known as the Levinson-Durbin recursion [15]. The Levinson-Durbin algorithm makes use of the Toeplitz property of the correlation matrix **R** to recursively compute the solution to the normal equation starting from order 1 through to the final order of the filter. The Levinson-Durbin algorithm can be summarized by the following equations:

$$a_{m,n} = a_{m-1,n} + \Gamma_m a_{m-1,m-n}^* \qquad n = 0, 1, \ldots, m \tag{6.14}$$

$$P_m = P_{m-1} (1 - |\Gamma_m|^2) \tag{6.15}$$

where Γ_m is known as the reflection coefficient of order m.

6.2. Filter Derivation

The SGL-JPE is a well known algorithm and detailed derivations can be found in many literature [13, 15]. A summary of the derivations is presented here.

Figure 6.1 shows the structure of the SGL-JPE.

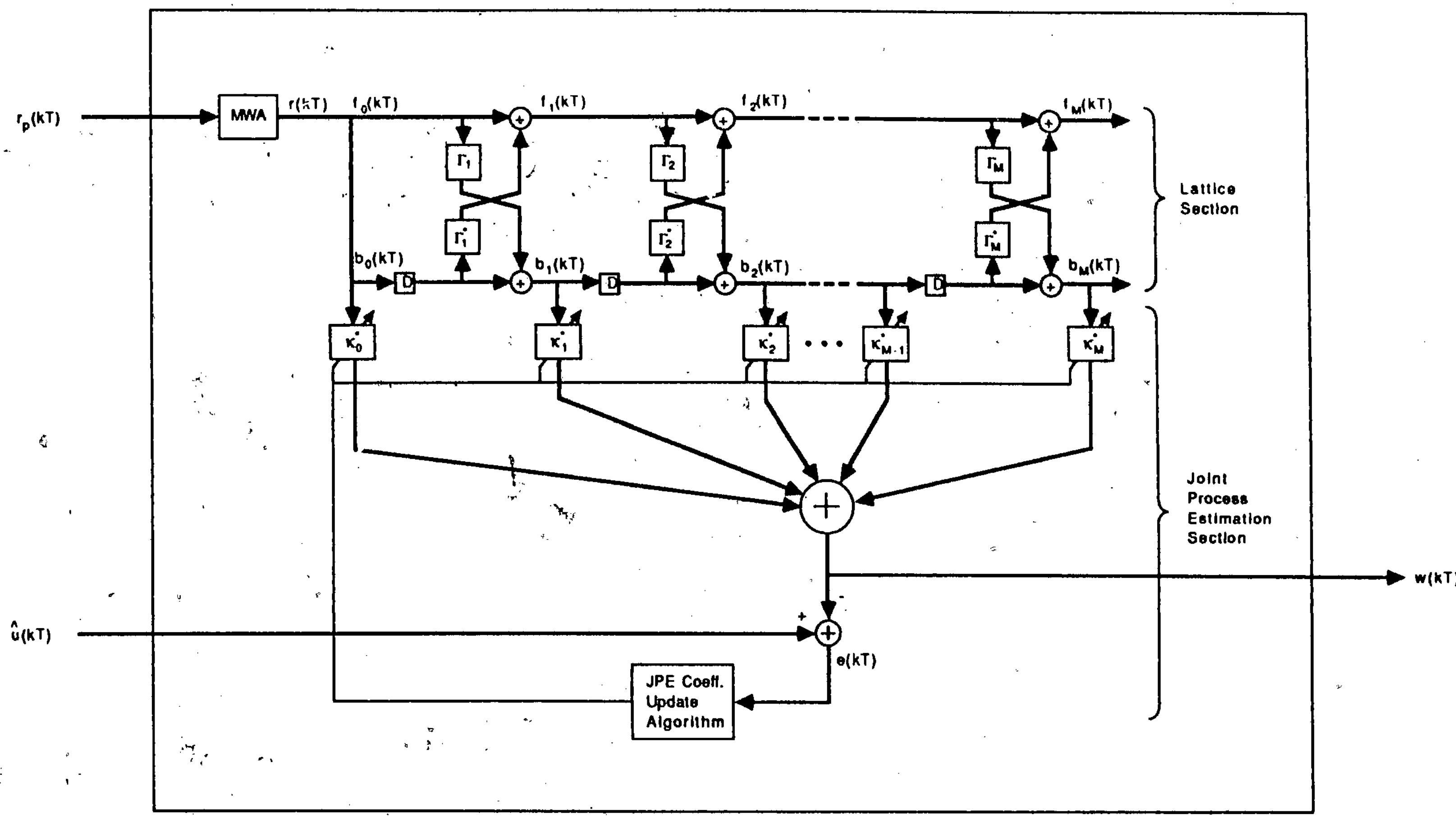


Figure 6.1 - Stochastic Gradient Lattice Joint-Process Estimator

The lattice filter section is described by the pair of equations:

$$f_m(k) = f_{m-1}(k) + \Gamma_m^* \, b_{m-1}(k-1) \qquad m = 1, \ldots, M \tag{6.16}$$

$$b_m(k) = b_{m-1}(k-1) + \Gamma_m \, f_{m-1}(k) \qquad m = 1, \ldots, M \tag{6.17}$$

where f_m and b_m are the m-th order forward and backward prediction errors as discussed in the previous section. M is the order of the SGL-JPE which is assumed to be odd. The lattice filter section is preceded by a MWA as in the SGT case. In the most common form, the reflection coefficients, Γ_m, are chosen to minimize the sum of mean squared forward and backward prediction errors defined by:

$$e_m(k) = E[|f_m(k)|^2] + E[|b_m(k)|^2] \tag{6.18}$$

We can perform the minimization adaptively by using the stochastic gradient algorithm as before. In this case, the gradient of e_m is given by:

$$\frac{de_m(k)}{d\Gamma_m(k)} = 2\,\Gamma_m(k) \left\{ E[|f_{m-1}(k)|^2] + E[|b_{m-1}(k-1)|^2] \right\} + 4\,E[f_{m-1}^*(k)\, b_{m-1}(k-1)] \tag{6.19}$$

To reduce the mean square error, we take steps in direction opposite to the direction of the gradient vector so that:

$$\Gamma_m(k+1) = \Gamma_m(k) - \frac{1}{2}\mu_m(k) \frac{de_m(k)}{d\Gamma_m(k)} \tag{6.20}$$

where μ_m is the step size for order m update. Replacing the gradient of e_m with an instantaneous estimate, we get:

$$\Gamma_m(k+1) = \Gamma_m(k) - \mu_m(k) \left\{ \Gamma_m(k) \left[|f_{m-1}{}^2(k)| + |b_{m-1}(k-1)|^2 \right] + 2\, f_{m-1}{}^*(k)\, b_{m-1}(k-1) \right\} \tag{6.21}$$

Next, we choose the step size to be the reciprocal of an estimate of the prediction error power, E_{m-1}, yielding [20]:

$$\Gamma_m(k+1) = \Gamma_m(k) - \frac{1}{E_{m-1}(k)} \left\{ \Gamma_m(k) \left[|f_{m-1}{}^2(k)| + |b_{m-1}(k-1)|^2 \right] + 2\, f_{m-1}{}^*(k)\, b_{m-1}(k-1) \right\} \tag{6.22}$$

Expanding this and making use of (6.16) and (6.17), we can simplify (6.22) to give:

$$\Gamma_m(k+1) = \Gamma_m(k) - \frac{1}{E_{m-1}(k)} \left\{ f_{m-1}{}^*(k)\, b_m(k) + b_{m-1}(k-1)\, f_m{}^*(k) \right\} \tag{6.23}$$

The prediction error power estimate is computed as:

$$E_{m-1}(k) = \lambda\, E_{m-1}(k-1) + |f_{m-1}(k)|^2 + |b_{m-1}(k-1)|^2 \tag{6.24}$$

where λ is an aging coefficient introduced to allow tracking of changing input statistics.

The joint-process estimation section is identical in structure to the SGT filter with input to each stage replaced by the backward prediction errors. The JPE coefficients are adapted using the stochastic gradient algorithm as are the SGT filter coefficients. The difference in this case is in the choice of step size. When the lattice is converged, the backward prediction errors are orthogonal. This means that the backward prediction error powers are equal to the eigenvalues of the respective stages. Furthermore, it can be shown that $E[|f_{m-1}(k)|^2] = E[|b_{m-1}(k-1)|^2]$ [15]. So we can make use of (6.24) to obtain an estimate for the eigenvalue of stage m. For stage m, we use a step size $\Delta(m)$ such that:

$$\Delta(m) = \frac{2}{E_{m-1}(k)} \tag{6.25}$$

It follows that the JPE coefficient update equation is:

$$\kappa_m(k+1) = \kappa_m(k) + \frac{2}{E_{m-1}(k)} \{e^*(k)\, b_m(k)\} \tag{6.26}$$

where e(k) is the estimation error given by:

$$e(k) = \hat{u}(k) - \boldsymbol{\kappa}^H \mathbf{b}(k) \tag{6.27}$$

6.3. BER Computation

To compute the BER, we need to transform the sets of reflection and joint-process estimator coefficients into an equivalent set of FIR coefficients. To accomplish this, we make use of (6.4) which relates the backward prediction error of a particular order to the input vector. The set of equations describing the backward prediction error for each order can be grouped together to form the following matrix equation:

$$\mathbf{b}(k) = \mathbf{L}\,\mathbf{r}(k) \tag{6.28}$$

where

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{1,1} & 1 & 0 & \cdots & 0 \\ a_{2,2} & a_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{M,M} & a_{M,M-1} & a_{M,M-2} & \cdots & 1 \end{bmatrix} \tag{6.29}$$

and

$$\mathbf{r}(k) = [r(k-\tfrac{M-1}{2}), \ldots, r(k-1), r(k), r(k+1), \ldots, r(k+\tfrac{M-1}{2})]^{\mathbf{T}} \tag{6.30}$$

Output of the joint-process estimator is:

$$w(k) = \kappa^{\mathbf{H}}\, \mathbf{b}(k) \tag{6.31}$$

Substituting (6.28) into (6.31) yields:

$$w(k) = \kappa^{\mathbf{H}}\, \mathbf{L}\, \mathbf{r}(k) \tag{6.32}$$

Comparing this equation wiιn (5.3), the set of equivalent FIR coefficients can be readily recognized to be:

$$\mathbf{h} = \mathbf{L}^{\mathbf{H}}\, \kappa \tag{6.33}$$

Given the set of equivalent FIR coefficients, the bit error rate can now be computed using the procedure outlined in section 5.4. The only remaining item required for the computation is the matrix of prediction coefficients, **L**, which can be easily obtained from the reflection coefficients by using the Levinson-Durbin recursion.

6.4. BER Sensitivity to Filter Coefficients

The SGL-JPE is defined by two sets of coefficients which together determine the filter response. The filter response is a highly non-linear function of the reflection coefficients making any type of analysis difficult. In order to determine the sensitivity of BER to changes in the two sets of coefficients, we have pursued a computational approach. The BER was computed as a function of changes in the filter coefficients for the following parameters: $E_b/N_o = 40\text{dB}$, $f_DT = 0.0208$, $r = 0.2$, $M = 4$ and a MWA of length 3. Using the optimal sets of reflection and joint-process estimator coefficients for the above parameter, the BER was found to be 5.573 x 10^{-5}. Results of BER calculated as a function of percentage changes in the reflection and joint-process estimator coefficients are given in table 6.1. In table 6.1, sensitivity is defined as:

$$\eta = \frac{\partial \text{BER}}{\partial z}\frac{z}{\text{BER}} \approx \frac{\Delta \text{BER}}{\Delta z}\frac{z}{\text{BER}} \tag{6.34}$$

where $z \in \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \kappa_0, \kappa_1, \kappa_2, \kappa_3, \kappa_4\}$.

Coeff. changed	% change	BER	% incr in BER	Sensitivity η
Γ_1	-1	1.047×10^{-4}	88	88
	-2	2.39×10^{-4}	329	
	-10	2.973×10^{-3}	5230	
	-50	1.764×10^{-2}	31550	
Γ_2	-1	5.73×10^{-5}	2.8	2.8
	-2	6.105×10^{-5}	9.5	
	-10	1.91×10^{-4}	243	
	-50	3.25×10^{-3}	5730	
Γ_3	-10	5.573×10^{-5}	~0	0
	-50	5.573×10^{-5}	~0	
Γ_4	-10	5.573×10^{-5}	~0	0
	-50	5.602×10^{-5}	.5	
κ_0	-1	6.016×10^{-5}	8	8
	-2	7.39×10^{-5}	33	
	-10	5.697×10^{-4}	922	
	-50	2.416×10^{-2}	43250	
κ_1	-1	6.049×10^{-5}	8.5	8.5
	-2	7.455×10^{-5}	34	
	-10	5.408×10^{-4}	870	
	-50	1.493×10^{-2}	25960	
κ_2	-1	5.575×10^{-5}	0.04	0.04
	-2	5.591×10^{-5}	0.3	
	-10	6.236×10^{-5}	12	
	-50	2.354×10^{-4}	322	
κ_3	-10	5.594×10^{-5}	0.3	0.03
	-50	6.267×10^{-5}	12	
κ_4	-10	5.582×10^{-5}	0.2	0.02
	-50	5.724×10^{-5}	3	

Table 6.1 - BER Sensitivity to Lattice and Joint Process Estimator Coefficients

From this table, we can observe that the BER was extremely sensitive to the two lowest order reflection and JPE coefficients, particularly the lowest order reflection coefficient. Sensitivity also increased with increased percentage change in coefficients. It is clear from these observations that any sizeable fluctuations in the lower order coefficients will be detrimental to the performance of the SGL-JPE. This hypothesis will be further demonstrated to be true in the following simulation results.

6.5. Simulation Results

Monte Carlo simulations were performed with the SGL-JPE algorithm using a similar procedure as for the SGT case. Figure 6.2 shows the learning curves for different values of E_b/N_o with aging coefficient $\lambda = 0.995$, $f_DT = 0.0208$ and $r = 0.2$. In terms of convergence speed, the SGL-JPE is significantly faster than SGT especially at high E_b/N_o. The learning curve is smooth at low E_b/N_o. At high E_b/N_o, there are large fluctuations in BER. The presence of the fluctuations can be explained by examining figure 6.3 which shows the evolution of the imaginary part of the lowest order reflection coefficient. The fluctuations in the reflection coefficient match almost perfectly with those in the BER curves. The reason for the increase in BER fluctuations with E_b/N_o was because at low E_b/N_o, the received signal was dominated by additive white Gaussian noise thus masking out the effect of the fluctuations in the reflection coefficient. At higher E_b/N_o, fluctuations in the reflection coefficient become dominant and are manifested as fluctuations in BER. From these results, we can conclude that although the SGL-JPE provides more rapid convergence over SGT, it is not suitable for use in the pilot filter application because of the high sensitivity of the BER to the filter coefficients.

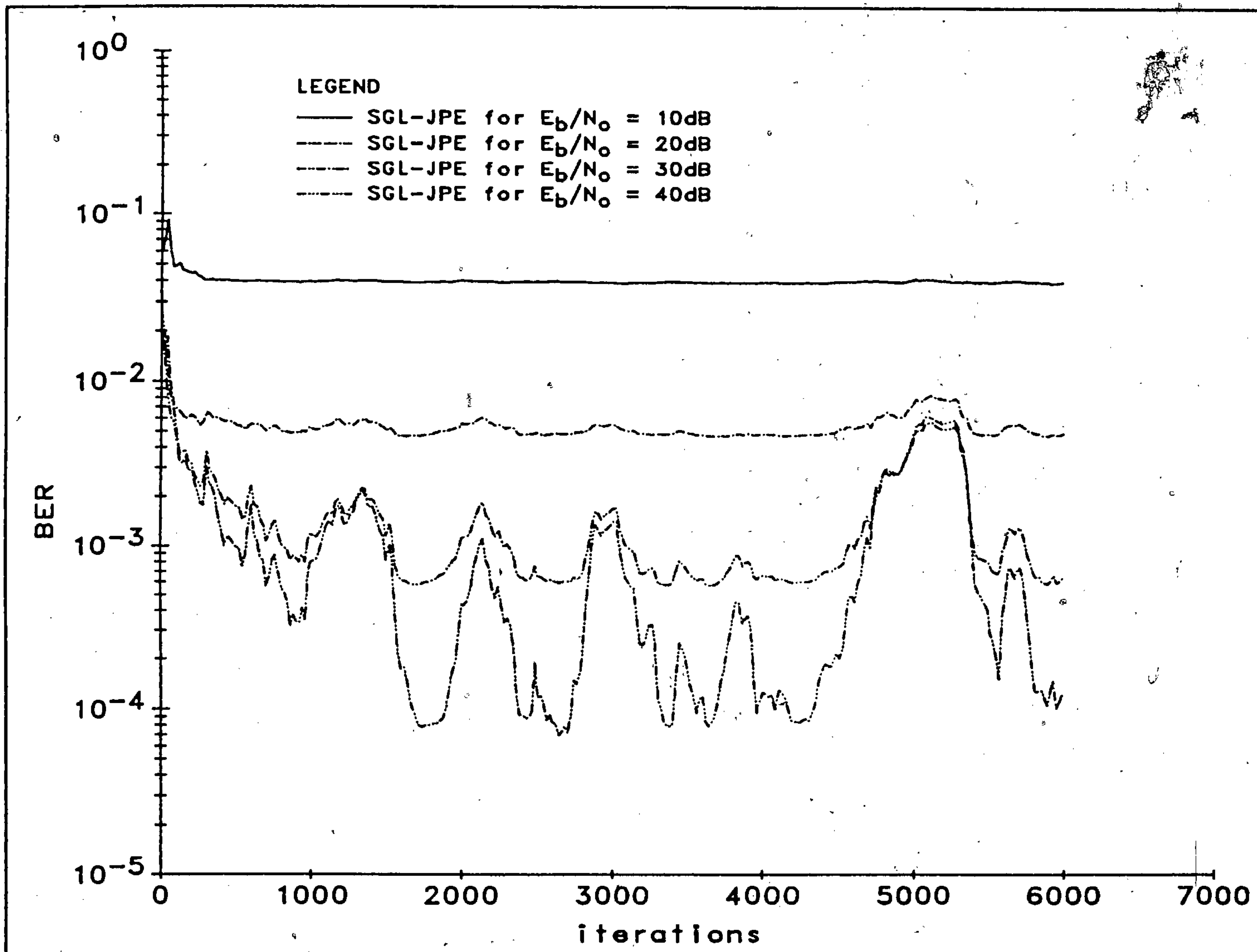


Figure 6.2 - Learning Curves of a Pilot Filter Implemented with the SGL-JPE at Various E_b/N_o

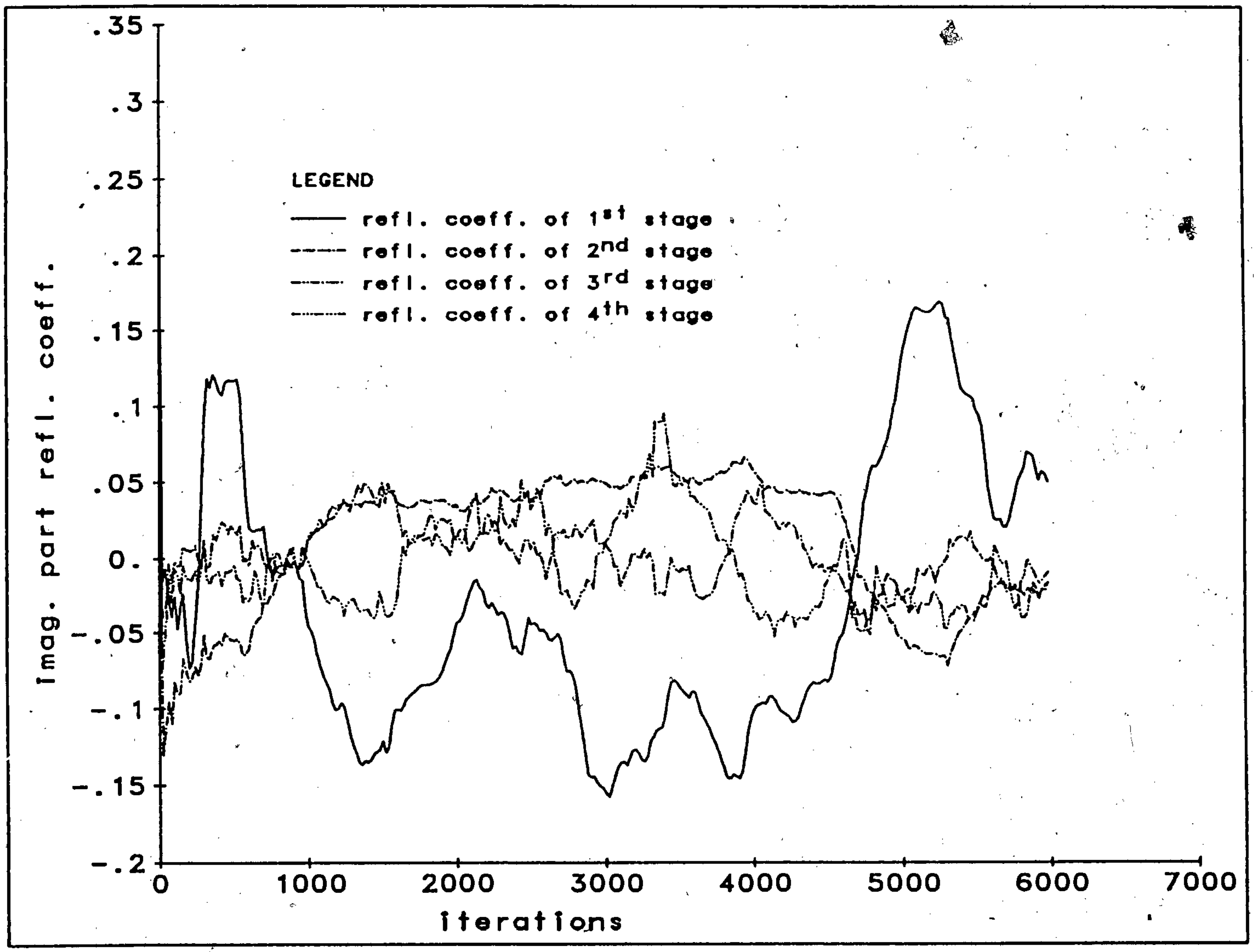


Figure 6.3 - Imaginary Parts of the Reflection Coefficients of a Pilot Filter Implemented with the SGL-JPE

7. FILTER SWITCHING ALGORITHM

We have seen in chapter 5 that the use of the stochastic gradient transversal filter for extracting pilot tone has the problem of converging very slowly when the received signal energy to noise density ratio is high. We have also demonstrated in chapter 6 that the gradient adaptive lattice joint-process estimator is not suitable for use as a pilot filter because of the high sensitivity of the resulting BER to changes in the filter coefficients. In this chapter, we present a novel technique for extracting the pilot by storing a pre-calculated bank of filters and simply selecting one as pilot filter. The technique has been given the name, filter switching algorithm (FSA).

One of the fundamental drawbacks of the SGT, SGL-JPE and other popular adaptive filtering algorithms is the need of these algorithms to update each and every filter coefficient. In a sense, they all perform a multi-dimensional adaptation in one form or another. In the pilot filter application, we have found that the optimum filter is the Wiener filter. An adaptive pilot filter needs only to adjust for changes in doppler frequency and frequency offset. This means that the adaptation process can be reduced to two dimensions. The FSA is formulated based on this idea. It will be shown that this approach can provide significant improvement in BER and convergence speed over the SGT and SGL-JPE algorithms. The FSA discussed in this chapter provides adjustment only for the doppler frequency. The problem of frequency offset compensation is considered separately in section 7.8 where a method for estimating the frequency offset is presented.

This chapter begins with a detailed description of the filter operation and derivations of the switching algorithm. In section 7.2, we present an analytical model which enabled us to compute the average BER and convergence speed of the new adaptation scheme. Sections 7.3 and

7.4 provides some numerical results computed using the analytical model just mentioned, followed by simulated results demonstrating the accuracy of the model given in section 7.5. Section 7.6 deals with some of the factors which need to be considered when implementing the FSA. Section 7.7 gives a summary of findings and compares the performance of a pilot filter implemented using the FSA with that of a SGT pilot filter. The final section discusses the problem of frequency offset compensation.

7.1. FILTER DESCRIPTION

7.1.1. Signal Flow

The structure of a pilot filter using the filter switching algorithm is shown in figure 7.1. The input samples, $r_p(k)$, are first averaged by a moving window averager (MWA). The MWA reduces the storage and computational requirements of the algorithm. The averaged samples are then split into two branches. The top branch consists of a set of reduced coefficient filters, one of which is selected to perform the actual filtering of the pilot. The reduced coefficient filter has non-zero coefficients spaced by the length of the MWA. We denote the overall time response of the combined MWA and reduced coefficient filter by $h_p(i, k)$, where i is the index of a particular filter in the filter bank and k is the time index. The pilot filter output can be written in terms of $h_p(i, k)$ as:

$$w(k) = a\,[c(k) * h_p(i, k)] + n_w(k) \tag{7.1}$$

where * denotes convolution.

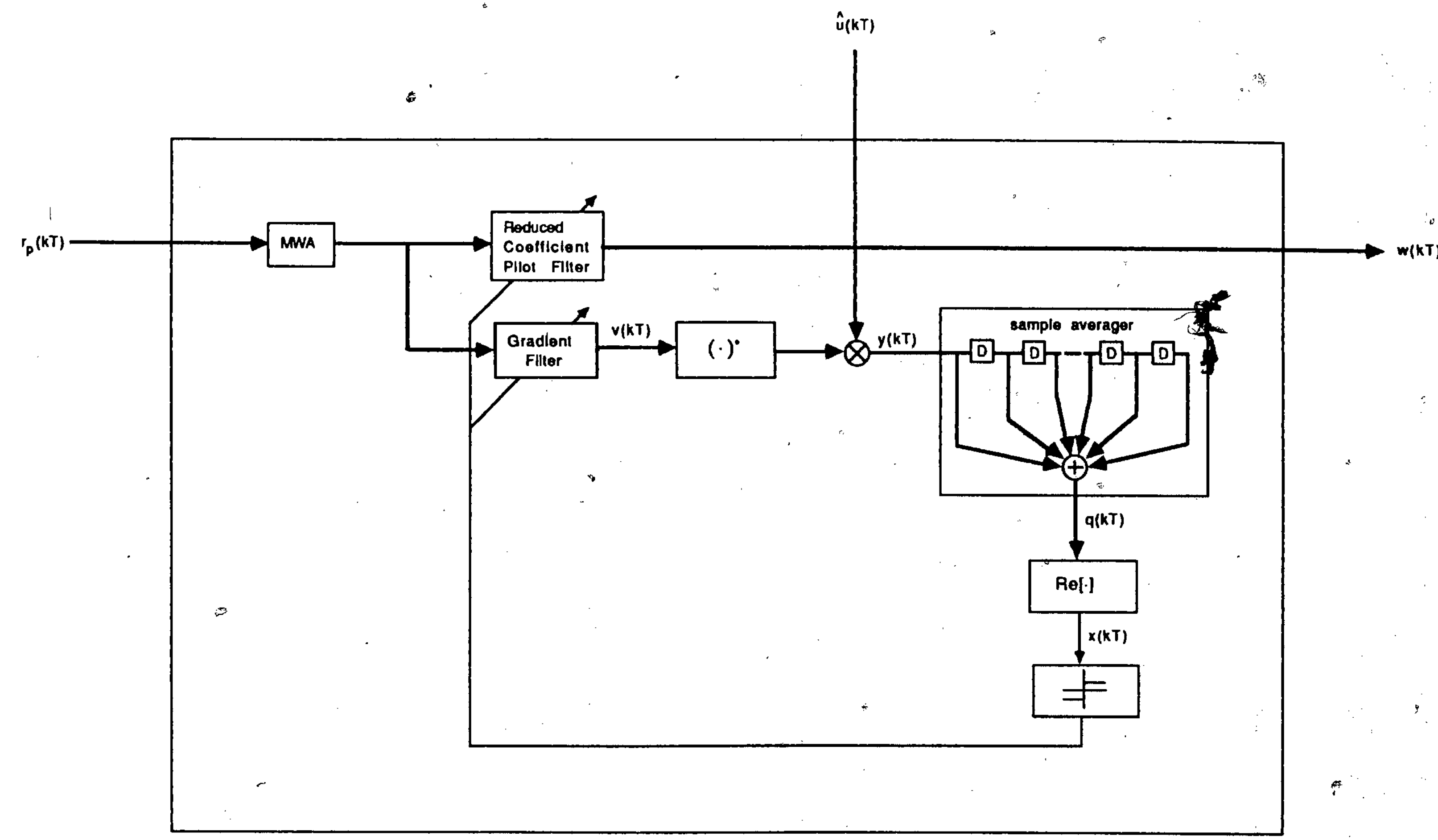


Figure 7.1 - Adaptive Pilot Filter using Filter Switching Algorithm

The lower branch of pilot filter comprises an adaptation loop which provides the mechanism for selecting the appropriate reduced coefficient filter from the filter ensemble. The adaptation loop itself consists of a bank of gradient filters and a sample averager. Each gradient filter has an impulse response which equals the difference between impulse responses of two reduced coefficient filters and there is one corresponding gradient filter for each reduced coefficient filter in the upper branch. Let us denote the time response of the combined MWA and gradient filter by:

$$\Delta h_p(i,k) = h_p(i+1,k) - h_p(i,k) \tag{7.2}$$

The gradient filter output is then given by:

$$v(k) = a\,[c(k) * \Delta h_p(i,k)] + n_v(k) \tag{7.3}$$

v(k) is conjugated and multiplied by the decision corrected data signal $\hat{u}(k)$ to form an instantaneous estimate of the cross-correlation y(k). y(k) is averaged by a sample averager with length N to give a sampled cross-correlation:

$$q(k) = \frac{1}{N}\sum_{m=0}^{N-1} y(k-m) = \frac{1}{N}\sum_{m=0}^{N-1} \hat{u}(k-m)\, v(k-m)^* \tag{7.4}$$

The decision variable Re[q(k)], denoted by x(k), determines the next filter to be used.

7.1.2. Filter Operation

The bank of reduced coefficient filters consists of filters optimized for different doppler frequencies and is arranged in order of increasing doppler. During initial operation, the filter optimized for the largest doppler is selected. A performance index is then evaluated at some fixed time interval which determines whether to switch to a filter optimized for a larger doppler or one which is optimized for a smaller doppler frequency. At a particular doppler frequency, it can be shown that as one increases the length of the optimum (Wiener) filter, its frequency response approaches that of a rectangular low pass filter and this approximation improves with increased signal to noise ratio. If we approximate the set of optimum filters with rectangular filters, then we can accommodate changes in doppler frequency by simply varying the pilot filter bandwidth. This approximation eases the analysis and understanding of the effects that various parameters have on the FSA. The effect of using sets of rectangular filters vs optimum FIR filters will be discussed later in section 7.3.2.

Assuming that the frequency responses of the combinations of MWA and reduced coefficient filters are ideal rectangular, then after convergence, the pilot filter bandwidth jitters about the optimal value equal to $(f_D + f_o)$ [10]. If at a particular time filter i is selected, it is possible to switch from filter i to filter i+n where i+n is less than or equal to the total number of filters in the set. To simplify analysis, we restrict n to only take on values +1 or -1.

Some of the parameters important to the performance of the algorithm are: filter shape, filter length and bandwidth spacing. The widest bandwidth filter is set to equal to $(f_{D\,max} + f_{o\,max})$. Overall shape of the filters affects performance. As shown by the results in section 7.3.2, rectangular filters do not give the best performance. Selection of filter length involves the usual tradeoff between performance, computation complexity and the amount of delay. Smaller bandwidth spacing provides some improvement which will be discussed in section 7.3.1. However, the number of filters and hence the amount of storage space required is inversely proportional to the bandwidth spacing. Again some tradeoff needs to be made.

7.1.3. Performance Index Selection

Integral to the design of all adaptive algorithms is the need for a performance index; for the LMS algorithms, mean square error is used. The performance index needs to be well behaved. This means that the performance index as a function of the filter bandwidth must not have any local peaks. In the present application, we would like to minimize the BER. For a pilot filter with a real frequency response, this corresponds to maximizing the correlation coefficient ρ between the matched filter and pilot filter output as evident from (3.3). However, ρ is not a suitable performance index because it is difficult to compute and analyze. An alternative is to simply employ the covariance which, for Rayleigh fading, is equivalent to the cross-correlation since $E[u(k)] = E[w(k)] = 0$. There are two advantages in using the cross-correlation as performance index. One advantage is that it is easier to compute than correlation coefficient because there is no need to obtain the variances of the filters' input and output as in the case of correlation coefficient. The second advantage is related to the fact that the cross-correlation is linear with respect to the pilot filter response (see equation 3.2). This means that computing the difference in cross-correlation for two filter output is equivalent to computing the cross-correlation for a filter whose impulse response is the difference in impulse response between the two filters. This third filter is the gradient filter referred to earlier. For a pilot filter ensemble made up of rectangular filters, each gradient filters will have a frequency response as shown in figure 7.2.

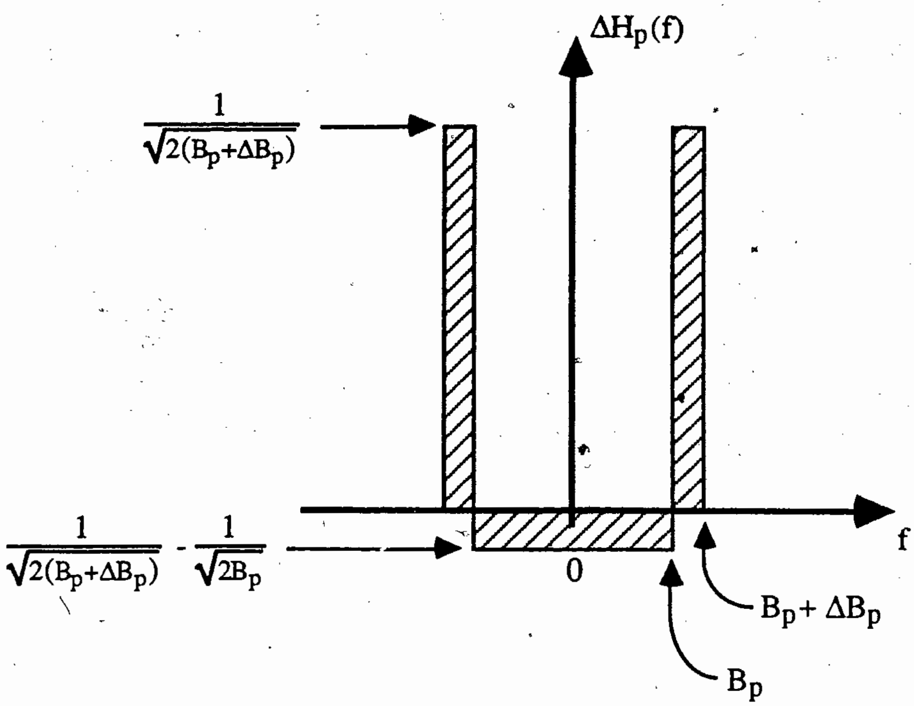


Figure 7.2 - Gradient Filter Frequency Response (for rectangular pilot filters)

At this time, we should point out that rectangular filters with identical gain cannot be utilized if cross-correlation is to be used as performance index. This is because the cross-correlation for rectangular filters having the same gain are identical if their bandwidths are greater than (f_D + f_o). Using identical gain rectangular filters, the switching algorithm will select ANY of the filters with bandwidths greater than $f_D + f_o$ at random. Obviously, only the filter with bandwidth closest to (and greater than) $f_D + f_o$ is optimal.

Denoting the cross-correlation by C, two questions remain in determining whether C is suitable as a performance index. The first question is whether the pilot filter bandwidth corresponding to maximum C, denoted by $B_{p,max}$, also gives minimum BER; or equivalently, whether $B_{p,max}$ gives maximum ρ. The second question is whether ΔC is well-behaved. The first question can be answered by differentiating C and ρ with respect to the pilot filter bandwidth B_p, setting the derivatives to zero and solving for $B_{p,max}$. It can be shown that in both of these cases, $B_{p,max}$

$= f_D + |f_o|$. (See Appendix 1 for the derivations). From the same analysis, it can also be shown that C has no local peaks.

7.1.4. Derivation of the Filter Switching Algorithm

The filter switching algorithm can be summarized as taking a step in a direction opposite to the gradient of C. This is the steepest descent algorithm. The update equation is given by:

$$B_p(k+1) = B_p(k) - \mu \frac{\delta C}{\delta B_p} \tag{7.5}$$

where B_p is the filter bandwidth, μ is a positive scalar constant and k is the time index. Approximating $\frac{\delta C}{\delta B_p}$ by $\frac{\Delta C}{\Delta B_p}$ and letting $\mu' = \frac{\mu}{\Delta B_p}$, then:

$$B_p(k+1) = B_p(k) - \mu' \Delta C \tag{7.6}$$

ΔC is defined as:

$$\Delta C = E[\hat{u}(k)\, v(k)^*] \tag{7.7}$$

We now quantize B_p so that only a discrete number of filters is required. To simplify the algorithm, we further replace ΔC with sign(ΔC) and approximate ΔC by the real part of a sample mean so that the decision variable becomes $x(k) = \mathrm{Re}[q(k)]$ where $q(k)$ is defined by (7.4). The resulting algorithm is then to calculate x(k):

$$B_p(k+1) = \begin{cases} B_p(k) + \Delta B_p & x(k) > 0 \\ B_p(k) - \Delta B_p & \text{otherwise} \end{cases}$$

Assuming that the i^{th} filter is used at time k, i.e. $B_p(k) = B_p(i,k)$, the algorithm can be rewritten as:

$$B_p(k+1) = \begin{cases} B_p(i+1,k+1) & x(k) > 0 \\ B_p(i-1,k+1) & \text{otherwise} \end{cases}$$

7.2. FILTER SWITCHING ANALYSIS

This section provides an analytical model for the filter switching algorithm from which the convergence speed and the average BER can be computed. We first begin with a discussion of some of the assumptions made.

7.2.1. Assumptions

In order to simplify analysis, there are three major assumptions made.

(1) Time between each adaptation step, or adaptation period, is equal to the time spanned by the sample size of q(k), denoted by N, so that values of q(k) used for each adaptation step contain no overlapping samples.

(2) Values of q(k) taken N samples apart are uncorrelated.

(3) There are enough independent samples in q(k) such that the sum is Gaussian. Some justification for this and the previous assumption is the fact that the cross-correlation is a function of $J_0(2\pi f_D k)$ which has an envelope that decreases with time.

7.2.2. Markov Chain Model

x(k) is a random variable. Since the selection of filters depends on x(k), the index associated with each filter in the filter banks is a discrete time random variable. From assumptions (2) and (3), successive values of x(k) used in the switching decision are independent. This implies that the filter index i at the next adaptation step depends only on the present value of i. So the process describing i forms a Markov chain. Associated with every Markov chain is a transition probability matrix and an initial state probability vector. For the present application, the transition probabilities are determined by the probability density function (pdf) of x(k). The initial state probability vector has the value one for the highest state M corresponding to the index of filter with the widest bandwidth and zero for all others. Given the transition probabilities and the initial state probability vector, a set of steady state probabilities can be computed if the Markov chain is irreducible[1]. For a particular f_D and pilot filter used, BER can be computed using (3.2) and (3.3). The average BER is simply the sum of BER given each filter, weighted by the state probabilities such that:

$$E[P_e] = \sum_{i=1}^{M} \frac{1-\rho(i)}{2} v_i \tag{7.8}$$

where v_i denotes the steady state probability for filter i.

An illustration of the Markov chain model representing the filter switching algorithm is given in figure 7.3.

[1] A markov chain is said to be irreducible if every state can be reached from every other state in a finite number of steps.

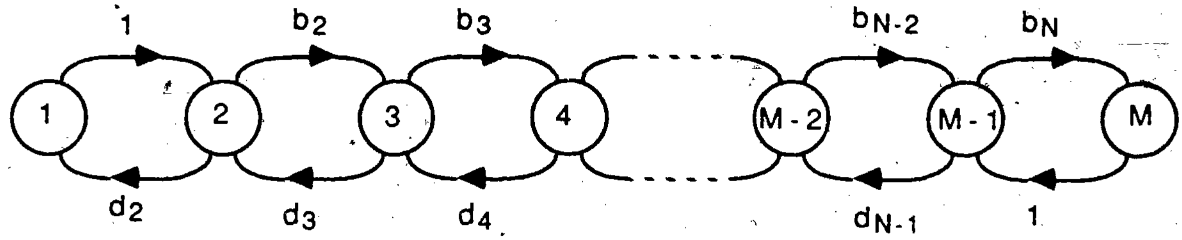
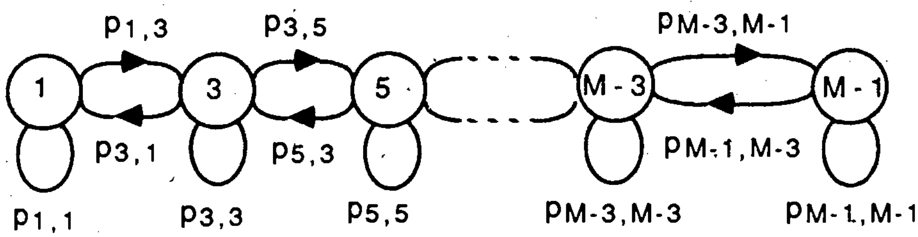


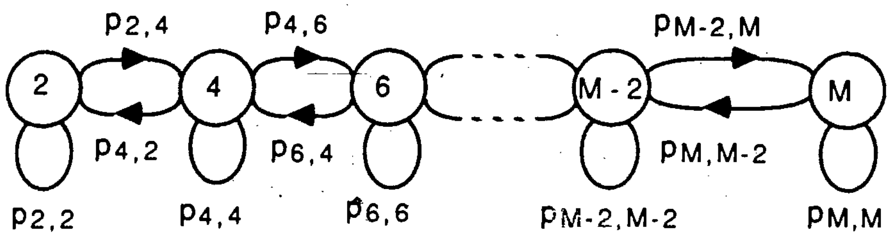
Figure 7.3 - Markov Chain Model of the Filter Switching Process

In queueing theory, this is the Markov chain of the birth-death process [21] with transitions restricted to neighboring states only. The transition probabilities of the Markov chain are represented by b_i and d_i denoting the birth and death probabilities at state i respectively. As mentioned earlier, in order to obtain the average BER, the steady state probabilities are needed. Unfortunately, the model as shown in figure 7.3 strictly does not have a steady state because no state is allowed to jump back to itself in a single transition, i.e. there is no self-loop. However, the model can be modified by considering the following. Assuming that the number of time steps taken is odd, then every odd state can be reached if the initial state is even. The chain will be in an even state only at the start. Similarly, if the number of steps taken is even, then the process will be able to reach every even state. If the initial state is odd, then every even state can be reached in an odd number of steps. If the number of steps taken is even, then the process will be able to reach every odd state given that the initial state is odd. The above observations indicate that the original chain can be split into two, an even and an odd states chain. If we group every two transitions on the original chain into one transition on one of the even or odd state chains and consider even and odd step transitions separately, then this model is identical to that shown in figure 7.3.

The new model is called dual Markov chain (DMC) model and is shown in figure 7.4 for the case when M is even.



a. Odd State Chain



b. Even State Chain

Figure 7.4 - Dual Markov Chain Model of the Filter Switching Process for M even

The transition probability matrices of the DMC model, $\mathbf{P}_{odd}$ and $\mathbf{P}_{even}$ can be obtained from the transition probability matrix of the original chain $\mathbf{P}$ as given by:

$$\mathbf{P}_{odd} = \mathbf{P}_{even} = \mathbf{P}^2 \tag{7.9}$$

It can be shown that the even and odd states chains are irreducible so that steady state probabilities exist. Details of the derivation of the steady state probability vector ν for an irreducible Markov chain can be found in [21]. In general, the procedure is to make use of the matrix equation:

$$\nu = [\nu_1, \nu_2, \nu_3, .] = \nu \ \mathbf{P} \tag{7.10}$$

where $\sum_{i} v_i = 1$ (7.11)

In terms of the state indices of the original chain, the steady state probability vector of the odd states chains can be shown to be:

$$v_i = \frac{v_{i-2}\, p_{i-2,i}(2)}{p_{i,i-2}(2)} = \prod_{j=3,odd}^{i} \frac{p_{j-2,j}(2)}{p_{j,j-2}(2)} v_1 \qquad i = 3 \text{ to } M-1 \qquad (7.12)$$

where

$$v_1 = \frac{1}{1 + \sum_{i=3,odd}^{M-1} \left\{ \prod_{j=3,odd}^{i} \frac{p_{j-2,j}(2)}{p_{j,j-2}(2)} \right\}} \qquad (7.13)$$

and $p_{i,j}(n)$ denotes the n-step transition probability from state i to state j. Similarly, for the even states chain, the steady state probabilities are given by:

$$v_i = \frac{v_{i-2}\, p_{i-2,i}(2)}{p_{i,i-2}(2)} = \prod_{j=4,even}^{i} \frac{p_{j-2,j}(2)}{p_{j,j-2}(2)} v_2 \qquad i = 4, M \qquad (7.14)$$

where

$$v_2 = \frac{1}{1 + \sum_{i=4,even}^{M-1} \left\{ \prod_{j=4,even}^{i} \frac{p_{j-2,j}(2)}{p_{j,j-2}(2)} \right\}} \qquad (7.15)$$

At any time, the probabilities of having taken an odd or an even number of steps are the same and are equal to 1/2. This means that the process on the average spends half its time in the odd state chain and half its time in the even state chain. Consequently, the average steady state probability is simply given by:

$$\nu_{i,avg} = \frac{1}{2}[\nu_i] \tag{7.16}$$

The transition probabilities b_i is derived from the pdf of the decision variable, x(k) which equals Re[q(k)], as follows. The adaptation algorithm dictates that if the process is at state i, it should switch to the next higher state if x(k) > 0; otherwise switch to the next lower state. Thus the probability of switching from state i to state i+1 is simply given by:

$$b_i = p_{i,i+1} = \Pr(x(k) > 0) \tag{7.17}$$

With the Gaussian approximation of x(k), all that are required are the first and second order statistics of x(k) which are derived next.

7.2.3. Statistics of the Sampled Cross-Correlation

First Order Statistics

The mean of x(k) is given by:

$$\begin{aligned} E[x(k)] &= E\Big[\frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\,[\hat{u}(k)\, v(k)^*]\Big] \\ &= \frac{1}{N}\sum_{k=0}^{N-1} E\big[\hat{u}_I(k)\, v_I(k) + \hat{u}_Q(k)\, v_Q(k)\big] \\ &= R_{\hat{u}_I v_I}(0) + R_{\hat{u}_Q v_Q}(0) \end{aligned} \tag{7.18}$$

where $R_{\hat{u}_I v_I}(k)$ and $R_{\hat{u}_Q v_Q}(k)$ are the cross-correlation functions between $\hat{u}$ and v as given in Appendix 3.

Second Order Statistics

The variance of x(k) can be expressed as:

$$
\begin{aligned}
\mathrm{Var}[x(k)] &= \mathrm{Var}\Big[\frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\,[\hat{u}(k)\, v(k)^*]\,\Big] \\
&= E\Big[\Big\{\frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\,[\hat{u}(k)\, v(k)^*]\,\Big\}^2\Big] - \Big\{E\Big[\frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\,[\hat{u}(k)\, v(k)^*]\,\Big]\Big\}^2 \\
&= E[\{\mathrm{Re}\,[q(k)]\}^2] - \{E[x(k)]\}^2
\end{aligned}
\qquad (7.19)
$$

where

$$
E[\{\mathrm{Re}\,[q(k)]\}^2] = E\Big[\Big\{\frac{1}{N}\sum_{k=0}^{N-1} \mathrm{Re}\,[\hat{u}(k)\, v(k)^*]\Big\}^2\Big]
$$

$$
\begin{aligned}
= \frac{1}{N^2} E\Big[\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} \{\, &\hat{u}_I(i)\, v_I(i)\, \hat{u}_I(k)\, v_I(k) + \hat{u}_I(i)\, v_I(i)\, \hat{u}_Q(k)\, v_Q(k) \\
+\ &\hat{u}_Q(i)\, v_Q(i)\, \hat{u}_I(k)\, v_I(k) + \hat{u}_Q(i)\, v_Q(i)\, \hat{u}_Q(k)\, v_Q(k)\}\Big]
\end{aligned}
$$

(7.20)

and E[x(k)] is given by (7.18). Using the result on high order joint moments of Gaussian random variables [22] and after some simple algebraic manipulations, the second moment of Re[q(k)] can be shown to be:

$$E[\{Re[q(k)]\}^2] = \frac{2}{N^2}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1}\Big\{2\,R_{\hat{u}_I v_I}(0)^2 + R_{\hat{u}_I \hat{u}_I}(i-k)\,R_{v_I v_I}(i-k) + R_{\hat{u}_I v_I}(i-k)^2$$
$$+ R_{\hat{u}_I \hat{u}_Q}(k-i)\,R_{v_I v_Q}(k-i) - R_{\hat{u}_I v_Q}(i-k)^2\Big\} \qquad (7.21)$$

This equation can be further simplified by converting the double sum into a single one by using the substitution l = i-k. For any function f, it can be shown that:

$$\sum_{i=0}^{N-1}\sum_{k=0}^{N-1}\Big\{f(i-k)\Big\} = \sum_{l=-(N-1)}^{N-1}\Big\{e(l)\,f(l)\Big\} \qquad (7.22)$$

where $e(l) = N - |l|$.

The resulting expression for Var[x(k)] is:

$$Var[x(k)] = \frac{2}{N^2}\sum_{l=-(N-1)}^{N-1}\Big\{e(l)\Big[2\,R_{\hat{u}_I v_I}(0)^2 + R_{\hat{u}_I \hat{u}_I}(l)\,R_{v_I v_I}(l) + R_{\hat{u}_I v_I}(l)^2$$
$$+ R_{\hat{u}_I \hat{u}_Q}(-l)\,R_{v_I v_Q}(-l) - R_{\hat{u}_I v_Q}(l)^2\Big]\Big\} - \Big\{R_{\hat{u}_I v_I}(0) + R_{\hat{u}_Q v_Q}(0)\Big\}^2 \qquad (7.23)$$

where $e(l) = N - |l|$ as defined previously.

Expressions for the correlation functions $R_{\hat{u}_I \hat{u}_I}(l)$, $R_{v_I v_I}(l)$, $R_{\hat{u}_I \hat{u}_Q}(l)$, $R_{v_I v_Q}(l)$, $R_{\hat{u}_I v_I}(l)$ and $R_{\hat{u}_I v_Q}(l)$ are given in Appendix 3.

7.2.4. Convergence time

Convergence time of a Markov chain is defined here as the expected number of steps required to reach a particular (destination) state for the first time. The method of deriving this convergence time is by first changing the destination state to an absorbing state and renumbering its state index to that of the last state. The transition probability of the modified chain, **P'**, is given by:

$$\mathbf{P'} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,i-1} & p_{1,i+1} & \cdots & p_{1,M} & p_{1,i} \\ p_{2,1} & \cdots & p_{2,i-1} & p_{2,i+1} & \cdots & p_{2,M} & p_{2,i} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ p_{i-1,1} & \cdots & p_{i-1,i-1} & p_{i-1,i+1} & \cdots & p_{i-1,M} & p_{i-1,i} \\ p_{i+1,1} & \cdots & p_{i+1,i-1} & p_{i+1,i+1} & \cdots & p_{i+1,M} & p_{i+1,i} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \\ p_{M,1} & \cdots & p_{M,i-1} & p_{M,i+1} & \cdots & p_{M,M} & p_{M,i} \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{7.24}$$

We then partition **P'** into the following form:

$$\mathbf{P'} = \left[\begin{array}{cccc|c} p'_{1,1} & p'_{1,2} & \cdots & p'_{1,M-1} & p'_{1,M} \\ p'_{2,1} & p'_{2,2} & \cdots & p'_{2,M-1} & p'_{2,M} \\ p'_{3,1} & p'_{3,2} & \cdots & p'_{3,M-1} & p'_{3,M} \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ p'_{M-1,1} & p'_{M-1,2} & \cdots & p'_{M-1,M-1} & p'_{M-1,M} \\ \hline 0 & 0 & \cdots & 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{0} & 1 \end{array}\right] \tag{7.25}$$

Let Y_N be the number of steps required to reach state N, the probability that the process is at state N after n steps be denoted by $p'_N(n)$ and $\mathbf{p}'(n)$ be the row vector with elements $p'_i(n)$, then:

$$p'_N(n) = \Pr[Y_N \leq n] \tag{7.26}$$

and

$$\mathbf{p}'(n) = [\, \tilde{\mathbf{p}}'(n) \mid p'_N(n) \,]$$

$$= [\, \tilde{\mathbf{p}}'(n-1) \mid p'_N(n-1) \,]\, \mathbf{P}'$$

$$= [\, \tilde{\mathbf{p}}'(n-1)\, \mathbf{A} \mid \tilde{\mathbf{p}}'(n-1)\, \mathbf{b} + p'_N(n-1) \,] \tag{7.27}$$

where $\tilde{\mathbf{p}}'(n-1) = [\, p_1(n-1), p_2(n-1), p_3(n-1), \ldots, p_{N-1}(n-1) \,]$

From (7.27), we get the following relations:

$$\tilde{\mathbf{p}}'(n) = \tilde{\mathbf{p}}'(n-1)\, \mathbf{A}$$

$$= \tilde{\mathbf{p}}'(0)\, \mathbf{A}^n \tag{7.28a}$$

and

$$p'_N(n) = \tilde{\mathbf{p}}'(n-1)\, \mathbf{b} + p'_N(n-1) \tag{7.28b}$$

The pdf of Y_N can be obtained from (7.28):

$$\Pr[Y_N = n] = \Pr[Y_N \le n] - \Pr[Y_N \le n-1]$$

$$= p'_N(n) - p'_N(n-1)$$

$$= \tilde{\mathbf{p}}'(0)\, \mathbf{A}^{n-1}\, \mathbf{b} \tag{7.29}$$

The moment generating function of Y_N can be shown to be:

$$P_{Y_N}(z) = \tilde{\mathbf{p}}'(0) \left[\frac{1}{z\mathbf{I}-\mathbf{A}} \right] \mathbf{b} \tag{7.30}$$

The convergence time is then given by:

$$E[Y_N] = -\frac{d}{dz} P_{Y_N}(z) \,\big|_{z=1} = \tilde{\mathbf{p}}'(0) \left[\frac{1}{\mathbf{I}-\mathbf{A}} \right]^2 \mathbf{b} \tag{7.31}$$

7.2.5. Variations of Algorithm Implementations

Step margin

It has been found that using a PLPF which is too narrow, i.e. less than ($f_D + f_o$), can raise the error floor [10]. This indicates that some margin of safety should be added to the pilot filter bandwidth chosen because the index will fluctuate about the optimum value. This idea has been incorporated into the filter switching algorithm by selecting a reduced coefficient filter whose bandwidth is a number of steps wider than the bandwidth of the filter corresponding to the gradient filter in use. The margin of safety is called the step margin. Section 7.3.1 will discuss the effect of step margin on BER performance.

Exponential Bandwidth Increment

Discussion thus far has assumed that the bandwidth spacing in the ensemble of filters is constant. However, this spacing arrangement may not provide the best overall performance for reasons which will be explained later. One alternative arrangement is to employ exponential bandwidth increment, i.e. to have the bandwidth increment arranged so that successive bandwidths follow an exponential function. As will be shown in section 7.3.1, this scheme introduces some tradeoff.

Dual Threshold

Recall from the Markov model presented earlier, steady state probabilities do not exist because no self-loop is allowed. This leads to instability as there is a tendency for the algorithm to jitter about the optimal state after convergence. One solution to this problem is to add self-loop probability. To do this, we alter the switching algorithm as follow:

If $x(k) > T1$,

$$B_p(k+1) = B_p(k) + \Delta B_p$$

else if $x(k) \leq T1$ and $x(k) \leq T2$,

$$B_p(k+1) = B_p(k)$$

else,

$$B_p(k+1) = B_p(k) - \Delta B_p$$

T1, T2 are transition thresholds. We can see from this new switching algorithm that there is now a finite probability of not changing state which is equal to $Pr(T1 \leq x(k) \leq T2)$. The Markov chain describing the new algorithm is one which represents a pure birth-death process with finite number of states[21]. It can be shown that the new Markov chain is irreducible and the steady state probability vector can be easily computed by solving (7.10) and (7.11). The steady state BER follows from (7.8). One should note that, for this algorithm, the transition thresholds need to be made a function of E_b (or E_b/N_o since N_o is not expected to change) because, as are evident from (7.18) and (7.23) the mean and variance of the decision variable x(k) are functions of E_b. This means that either E_b or E_b/N_o needs to be estimated by the algorithm.

7.3. COMPUTED BER PERFORMANCE BASED ON MARKOV MODEL

The effects of various parameters on the BER were investigated. BER curves were computed based on the procedure discussed in section 7.3. All results obtained in this section assumed the following unless stated otherwise: ideal rectangular pilot filters, R_b = 2400bps, r' = 0.2, f_o = 0, number of filters = 40, cross-correlation sample size = 299 and bandwidth increment ΔB_p = 5Hz.

7.3.1. General Results Using Ideal Rectangular Pilot Filters

Figure 7.5 shows the upward transition probabilities and average steady state probabilities as functions of the filter number (or state index) for f_D of 100Hz and E_b/N_o of 20dB with no step margin.

Transition probabilities depend on the amount of area in the fade spectrum covered by ΔH_p (see equations 7.18, A3.38 and A3.41). Because of the shape of the assumed U-shaped fade spectrum, one expects an increase in upward transition probability with state index which indicates an increasing tendency to move upward (i.e. to a higher state) for filters with $B_p < f_D$. This is shown by figure 7.5 to be true. An interesting observation is the fact that upward transition probability is nearly zero for $B_p > f_D$. The large tendency to move downward is caused by the negative gain of the gradient filter response in the range $(-B_p, B_p)$. (See figure 7.2 for an illustration of the gradient filter frequency response.) The small values of the upward transition probabilities are due to the more rapid decay of the correlation functions R_{vv} and $R_{\hat{u}v}$ for $B_p >$ f_D. As expected, the steady state probability curve indicates an increase as B_p increases to f_D. Another point worth noting is the fact that the steady state probabilities had significant values for $B_p \geq 80$Hz (state 16) and $B_p < 100$Hz (state 20). One expects a high error floor as a result because high steady state probabilities for $B_p < f_D$ suggest high probabilities of using filters which were too narrow. Fortunately, the high error floor can be reduced by using step margin as will be shown later in this section.

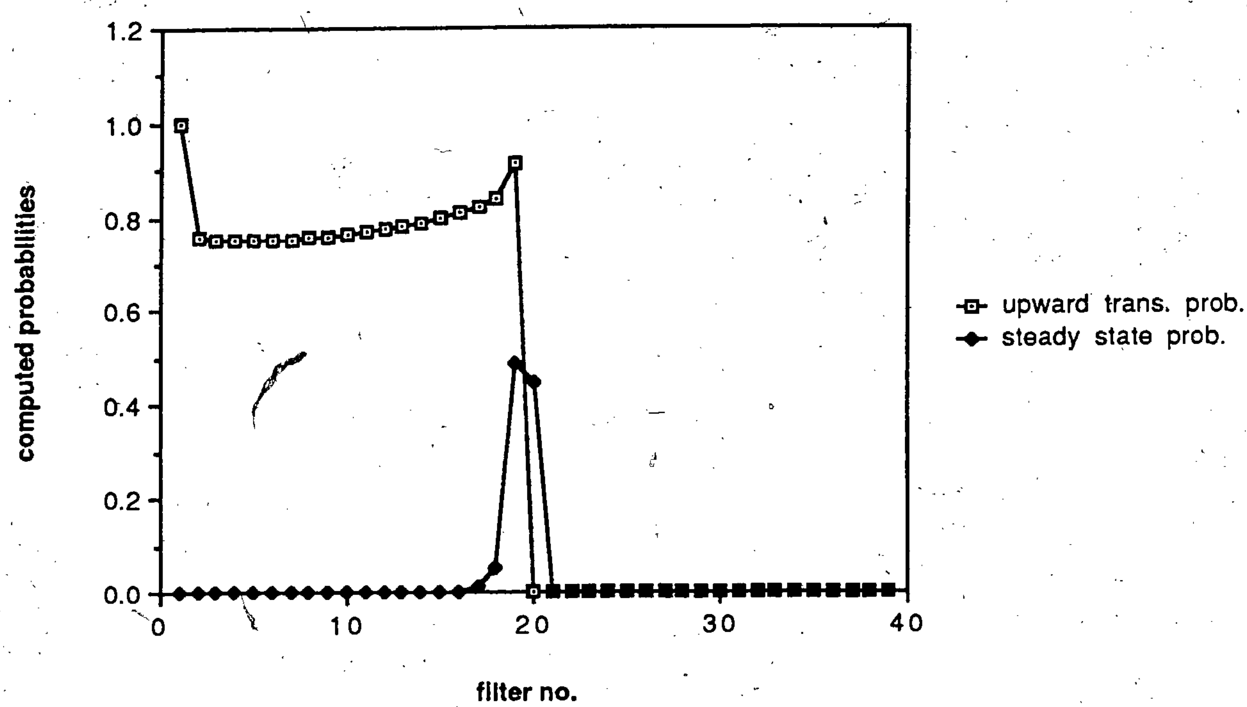


Figure 7.5 - Computed Probabilities for Filter Switching Algorithm using Ideal Rectangular Filters

Convergence time was 21 steps from the last state (largest B_p) and 32 steps from the first state (smallest B_p). Because the transition probabilities are nearly 0 for states with $B_p > f_D$, the process spends almost no time in these states. 32 steps is the worst case convergence time corresponding to the situation of a vehicle accelerating from standstill to 127 kmph in zero time (assuming 850 MHz carrier). In terms of f_D/s, this is 25 Hz/sec. (One step = 125 msec.) Realistically, the fastest change in f_D which can be expected is only about 15.2 Hz/sec (corresponding to at an acceleration of 0-100 kmph in 5 sec). So the convergence speed of the filter switching algorithm is more than adequate to accommodate changes in f_D from vehicle acceleration. As discussed in section 5.4.2, convergence during vehicle deceleration is not a problem. One should also note that because downward transition probability is nearly 1 for $B_p > f_D$, tracking is faster for a decrease in f_D than for an increase.

Figure 7.6 shows the average BER vs E_b/N_0. Note the enormous error floor due to transitions into states with too narrow a bandwidth. Also shown in figure 7.6 is the BER curve for a non-adaptive system using an ideal rectangular filter with B_p of 150Hz. Although it seems like the FSA performed very poorly here against the non-adaptive scheme, it will be shown later that this situation can be turned around by introducing step margin.

The optimum power split ratio r was relatively insensitive to changes in E_b/N_0 and was found to be 0.33.

Effect of Step Margin

The effect of varying step margin was investigated and the results are shown in figure 7.7 for f_D of 50Hz. The irreducible error floor dropped as the step margin was increased. For a margin of 6, the BER curve is almost parallel to that of the non-adaptive case and we can observe some significant improvement of the FSA over the non-adaptive case in which a rectangular pilot with 150Hz bandwidth was assumed. The improvement at a BER of 10^{-2} was about 0.8dB. A point which is worth noting is that as the margin was increased, the low E_b/N_0 portion of the BER curve moved upward as a result of increase in the average bandwidth. This shows a tradeoff between BER in the low and high E_b/N_0 regions when the step margin is varied.

Effect of Number of Filters

The effect of varying the number of filters used was negligible for a f_D of 100Hz. Recall that the steady state probabilities were nearly zero for $B_p > 100$Hz. This suggests that filters with $B_p \geq 100$Hz were rarely used and adding filters with $B_p > 100$Hz would have very little effect. Also, using a step margin of 5 had the effect of shifting the steady state probabilities up 5 states. This means that steady state probabilities could then be non-zero for B_p up to 125Hz and nearly zero for B_p greater than 125Hz. So, filters with B_p up to 125Hz were needed in this case. The important point to realize here is that, depending on the bandwidth increment chosen, we only need to use enough filters so that $B_{p\,max}$ is greater than ($f_{D\,max}$ + frequency margin).

Effect of f_D

Figure 7.8 shows BER for various f_D with a step margin of 5. BER for a lower f_D was smaller than for a higher f_D. This is because on the average at lower f_D, narrower filters were used more frequently than at higher f_D causing less noise to appear at the pilot filter output. At a BER of 10^{-2}, improvement was about 1.0dB for 50Hz doppler and 0.3dB for 100Hz. The error floor was also found to be higher for a lower f_D due to the increase in the low frequency component (or flat portion) of S_g as f_D was decreased. It can be shown that this low frequency component increases as $1/f_D$ for small f_D. So, for small f_D, larger part of S_g was covered by the negative gain portion of the gradient filter frequency response than for large f_D with the same gradient filter. This indicates that more margin was needed for small f_D.

Effect of Bandwidth Increment

Figure 7.9 shows the BER for two sets of filters with different ΔB_p for f_D of 50Hz. Different step margins were used so that the margin in frequency remained the same. Performance at low E_b/N_0 was about the same for both ΔB_p but the error floor nearly disappeared for the smaller ΔB_p. This is intuitively satisfying because d.c. gain of the gradient filter decreases with decreasing ΔB_p. A smaller (and negative) d.c. gain means a higher cross-correlation which would result in larger upward transition probabilities and smaller steady state probabilities for $B_p < f_D$. Since the d.c. gain has inverse dependence on B_p, steady state probabilities at smaller B_p get affected more. Overall result is that variations in B_p is less, which means that one can use a smaller step margin and get a corresponding improvement in BER. The cost of using a smaller ΔB_p is the increased storage requirement for more filters and the increase in convergence time.

Effect of f_o

The effect of f_o on BER performance is illustrated by figure 7.10 which shows BER for various f_o with a step margin of 6. The degradation in BER performance was primarily in the increase of the error floor as f_o was increased. When f_o was increased, one of the "horns" of the U-shaped fade spectrum moved closer to the low frequency portion of the gradient filter response which had a negative gain. From (7.18), (A3.38) and (A3.41), we see that $E[x(k)]$ would decrease with f_o which means that the upward transition probability had to decrease as well. When the upward transition probability became smaller, the variations in B_p became larger thus spreading out the steady state probabilities The spreading of the steady state probabilities was the primary cause of the increase in the error floor. In order to compensate for the spreading, we could increase the step margin at the expense of degrading the BER performance at low E_b/N_o. The spreading of steady state probability is illustrated by figure 7.11 which shows the steady state probabilities for system operating at E_b/N_o of 20 dB with various f_o. It can be observed from figure 7.10 that the maximum tolerable frequency offset was 10Hz for a step margin of 6. However, the step margin could be increased in order to accommodate a larger offset.

Effect of Cross-Correlation Sample Size

The cross-correlation sample size affects BER performance and convergence time as it determines the accuracy in the estimation of the cross-correlation. Figure 7.12 shows the effect of sample size on the BER for f_D of 100Hz. The decrease in the sample size produced a corresponding increase in the error floor. Fewer samples means a shorter time span covering the fading process, leading to greater uncertainty in the cross-correlation estimate. This in turn causes a larger steady state probability spread and as a result, a larger error floor. The response time was also proportionally increased. Figure 7.12 indicates that a minimum of 150 samples were needed in order to keep the error floor low enough. One should note that it is the total length of the time spanned by the cross-correlation samples which is important in determining BER. One could reduce the sample size by increasing the time between samples, or by using a longer MWA, and obtain a similar BER performance. This is true as long as the reciprocal of the time between samples is greater than the Nyquist rate of the fading process.

Exponential Bandwidth Increment

The results of using exponential bandwidth increment are shown in figures 7.13 and 7.14 for f_D of 50 and 100Hz. Successive filter bandwidths are fitted using an exponential function such that the bandwidths at states 1 (the lowest state) and 40 (the highest state) are 5 and 200Hz respectively. For 50Hz, exponential bandwidth increment performed better than linear increment at small step margin. With step margin greater than 1, linear increment was better. For f_D of 100Hz, exponential increment had a slight advantage over linear increment at high E_b/N_o (> 20dB) while the reverse was true at lower E_b/N_o. The above observations can be explained by considering step margin as frequency steps. At small f_D, the algorithm operates with small B_p; ΔB_p is small for exponential bandwidth increment which means that the frequency margin above the optimal bandwidth is smaller for exponential increment than for linear increment scheme using the *same* step margin. As a result, the exponential increment scheme gave better performance for small step margin. At larger f_D, the frequency margin is larger for exponential increment. So for the same step margin using exponential increment provides a lower irreducible error floor. However, this also means that the average bandwidth is also larger which causes inferior BER performance at low E_b/N_o. In general, we expect f_D to be small for urban driving and we need a step margin of 5 in order to keep the error floor low enough. Under these circumstances, using a linear bandwidth increment will provide better performance.

Dual Threshold

Figure 7.15 gives the results of using the dual threshold algorithm for f_D of 100Hz with various thresholds and no step margin. For simplicity, we had set the two thresholds to be constant multiples of E_b/N_o and they were made symmetrical such that T2 = -T1 where T1 > 0. One striking feature as evident from figure 7.15 is the improvement in error floor as the symmetrical thresholds are increased. With thresholds of ±1.0 E_b/N_o, we could almost do away with step margin. The drastic improvement was due to the increase in the ratios between upward and downward transition probabilities since these ratios determined the steady state probabilities. When $B_p < f_D$ and frequency offset is absent, we would like the steady state probabilities to be small for all states with $B_p < f_D$ in order to obtain a small error floor. This means that we need the upward to downward transition probability ratio to be large. Clearly, the transition probability ratio increases with increasing threshold values. Hence, increasing the transition thresholds had the effect of decreasing the error floor. Unfortunately, this reduction in error floor also resulted in the drastic increase in convergence time, as is indicated by figure 7.16 which shows the convergence time (in number of steps) as a function of transition threshold (in multiple of E_b/N_0). The increase in convergence time was caused by the increased self-loop probability. In practice, BER performance will also be degraded by noise present in the E_b/N_0 estimate. Thus, with all considerations, using a single threshold of zero is superior to using a dual threshold scheme.

7.3.2. Optimum FIR Filter

Ideal rectangular pilot filters were used to obtain all of the previous results. For the remainder of the discussion, we investigate the effect of using FIR filters optimized for various f_D at a fixed E_b/N_0 of 40dB. The optimum filters have been designed for increasing f_D at 5Hz increment starting at 5Hz and ending at 125Hz. The filters are numbered in order of increasing f_D so that filter number 1 corresponds to a filter optimized for 5Hz doppler and filter number 2 corresponds to a filter optimized for 10Hz doppler, etc. As shown in Appendix 2, the cross-correlation for a set of optimum FIR filters is not as "well behaved" as its rectangular counterpart. However, it will be shown later in this section that using a set of optimum FIR filters can offer BER performance comparable to that using a set of ideal rectangular filters. First, we investigate the effect of varying the length of the MWA and shaping filter in a non-adaptive environment.

Effect of MWA and Shaping Filter Length

Figures 7.17, 7.18 and 7.19 shows BER for f_D of 100 with various shaping filter length and MWA length of 1, 3 and 5 respectively. MWA length of 1 is equivalent to no MWA. The filter lengths are defined such that overall filter length equals MWA length x shaping filter length. For the three MWA lengths used, shaping filter lengths greater than or equal to 51 were close in performance. However, an increase in filter length means a corresponding increase in both the amount of computation and memory space. With a MWA length of 5, shaping filter length as low as 11 gave reasonable performance. One should note, however, that in actual operation with the FSA, a filter length below 21 is undesirable. The reason is because the passband ripples in the frequency response for a short filter can cause the upward transition probabilities to decrease, so that a larger step margin is required in order to maintain a low error floor. For the best compromise between BER performance and complexity, a MWA length of 5 and filter length of 51 should be used.

Ideal Rectangular vs Optimum FIR Filter

Figure 7.20 shows the BER using various step margins for f_D of 50Hz, shaping filter length of 51 and MWA length of 3. The error floor was negligibly small for step margin greater than 4. This is in contrast with results found using ideal rectangular filters as given in figure 7.7 which shows error floor still present with a margin of 6. The main reason for this is because of the gradual roll-off of the FIR filter low pass frequency response which allows for more of the fade spectrum to be covered.

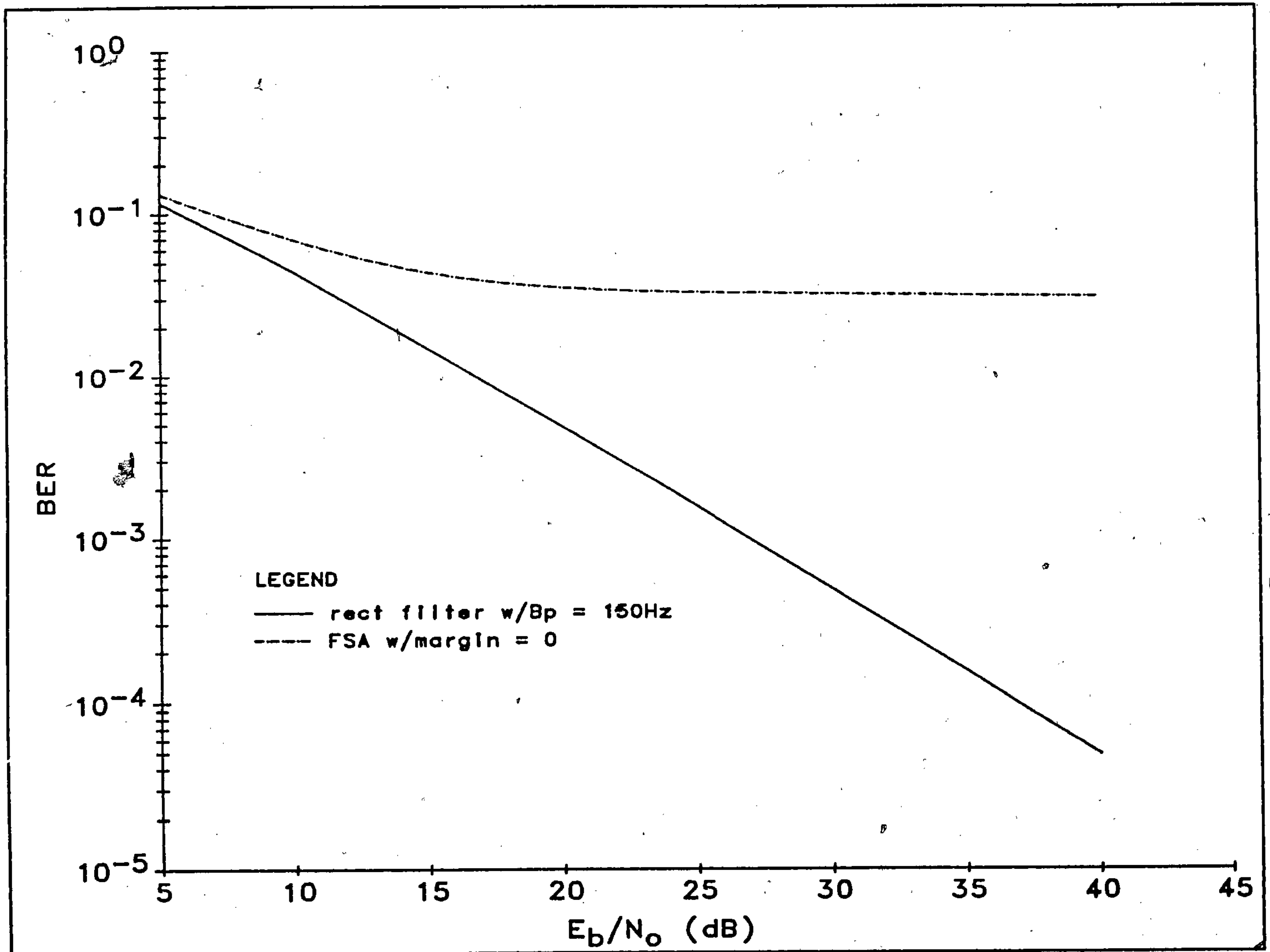


Figure 7.6 - Average BER Performance of a Pilot Filter Using FSA with no Step Margin

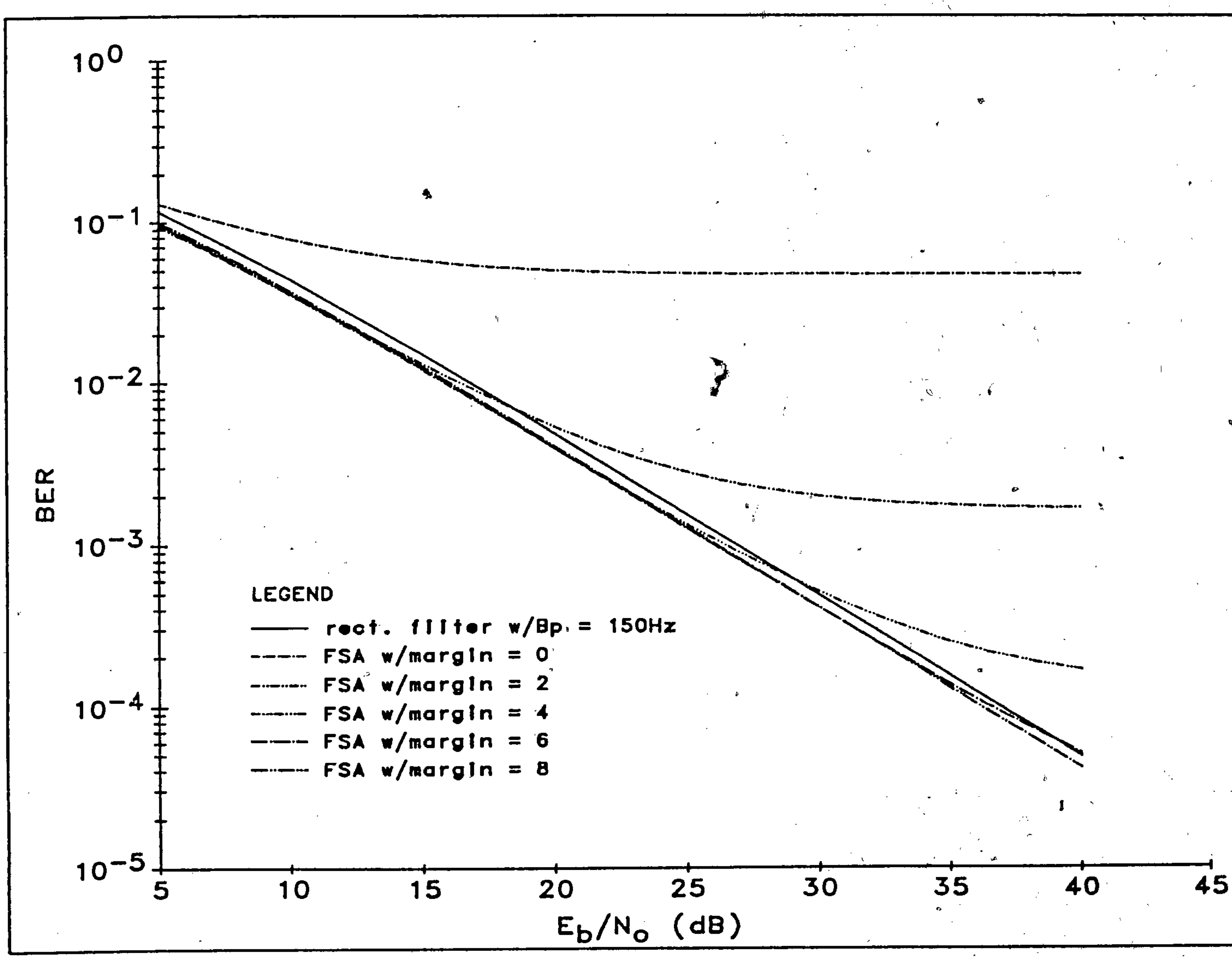


Figure 7.7 - Effect of Using Step Margin on the Average BER Performance of a Pilot Filter Implemented with the FSA

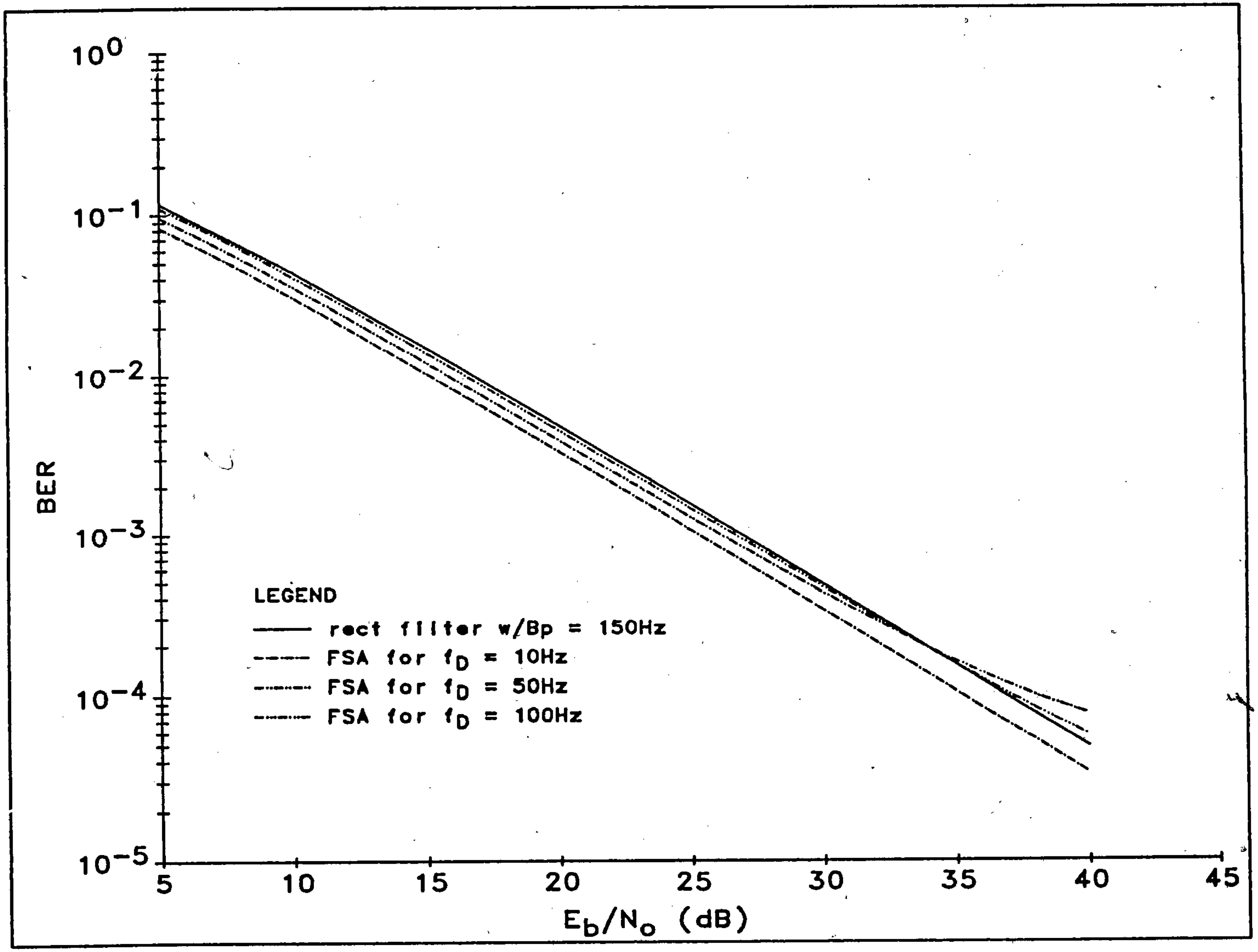


Figure 7.8 - Average BER Performance of a Pilot Filter Using FSA at Various Doppler Frequencies

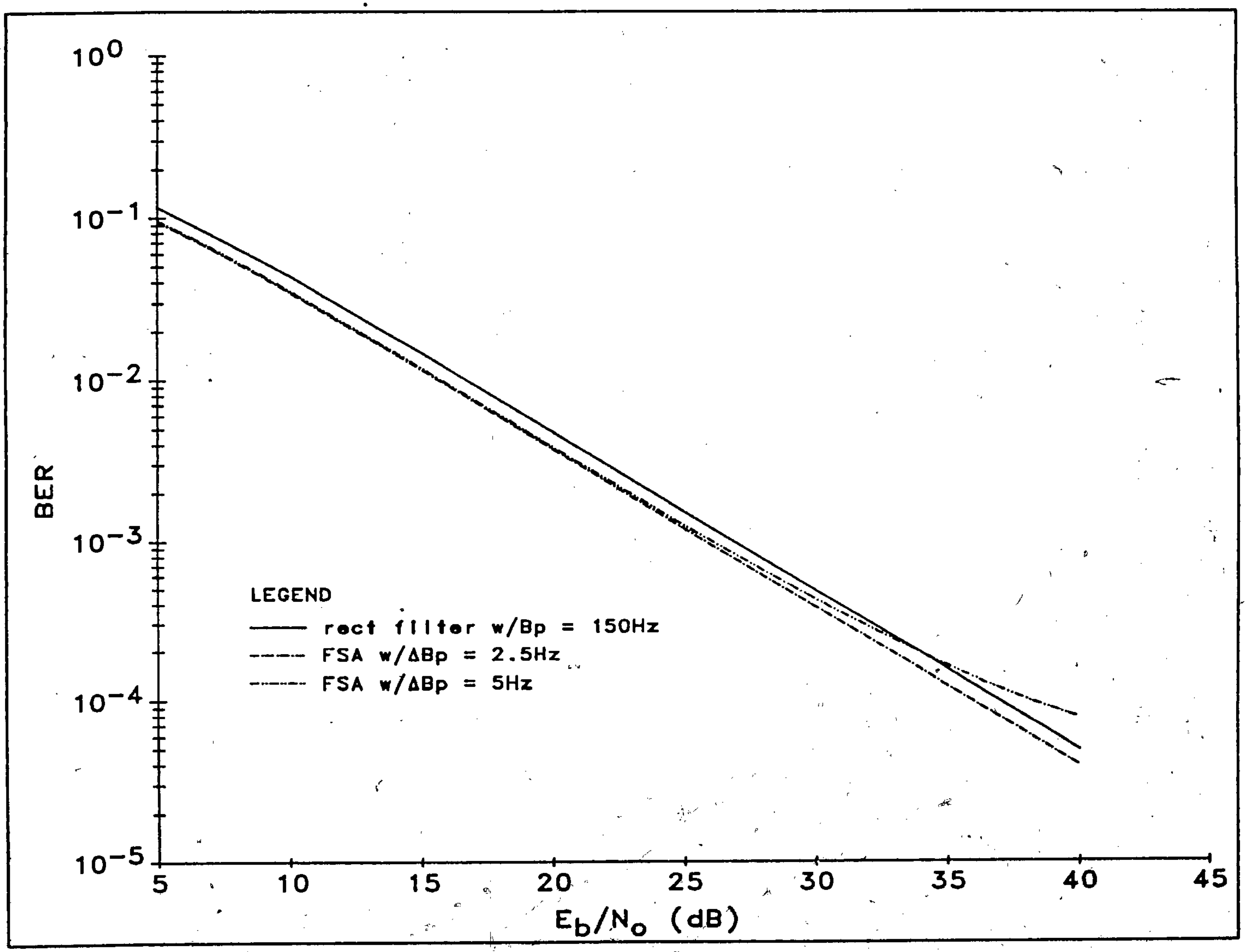


Figure 7.9 - Effect of Bandwidth Increment on the Average BER Performance of a Pilot Filter Implemented with the FSA

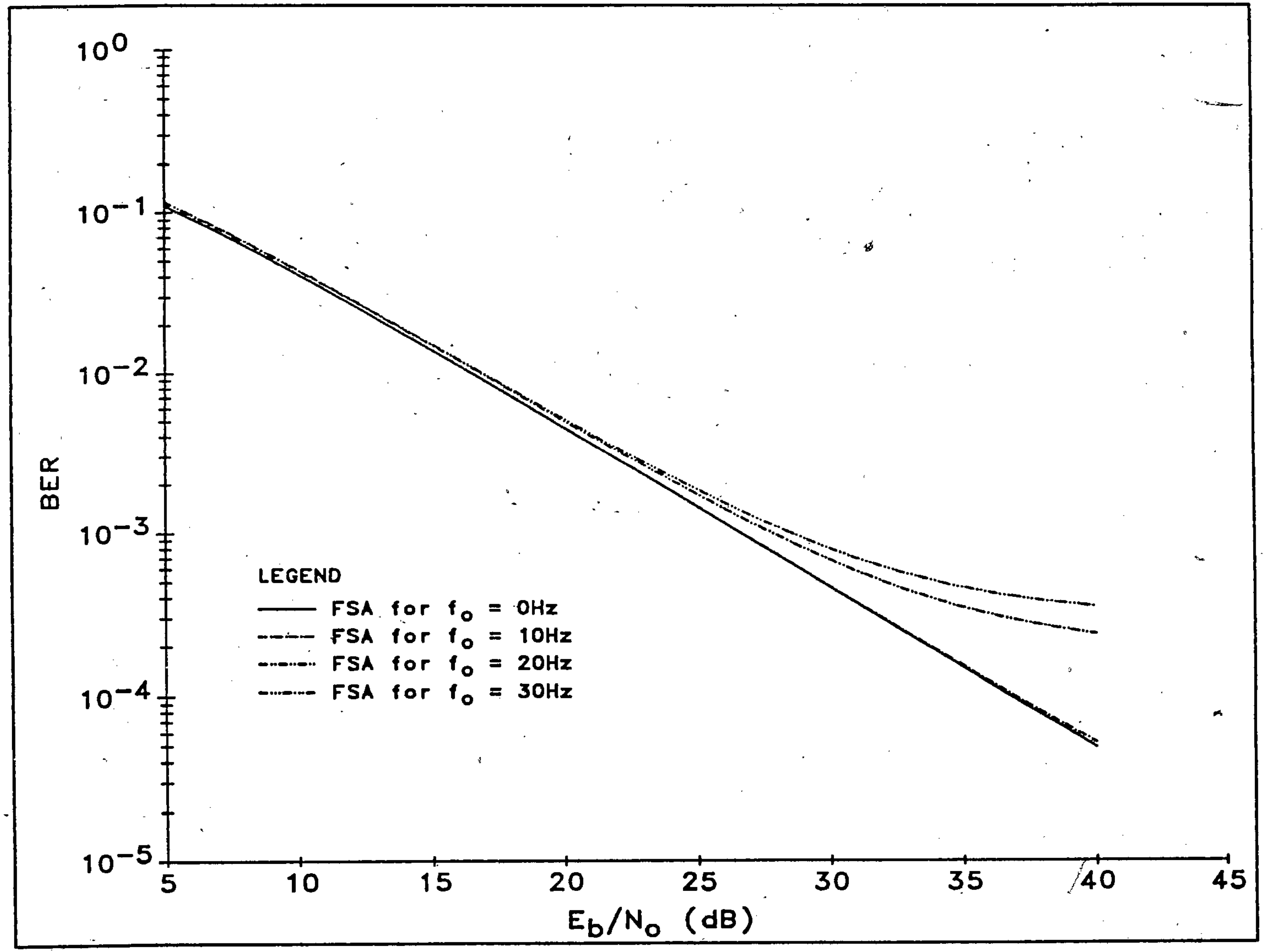
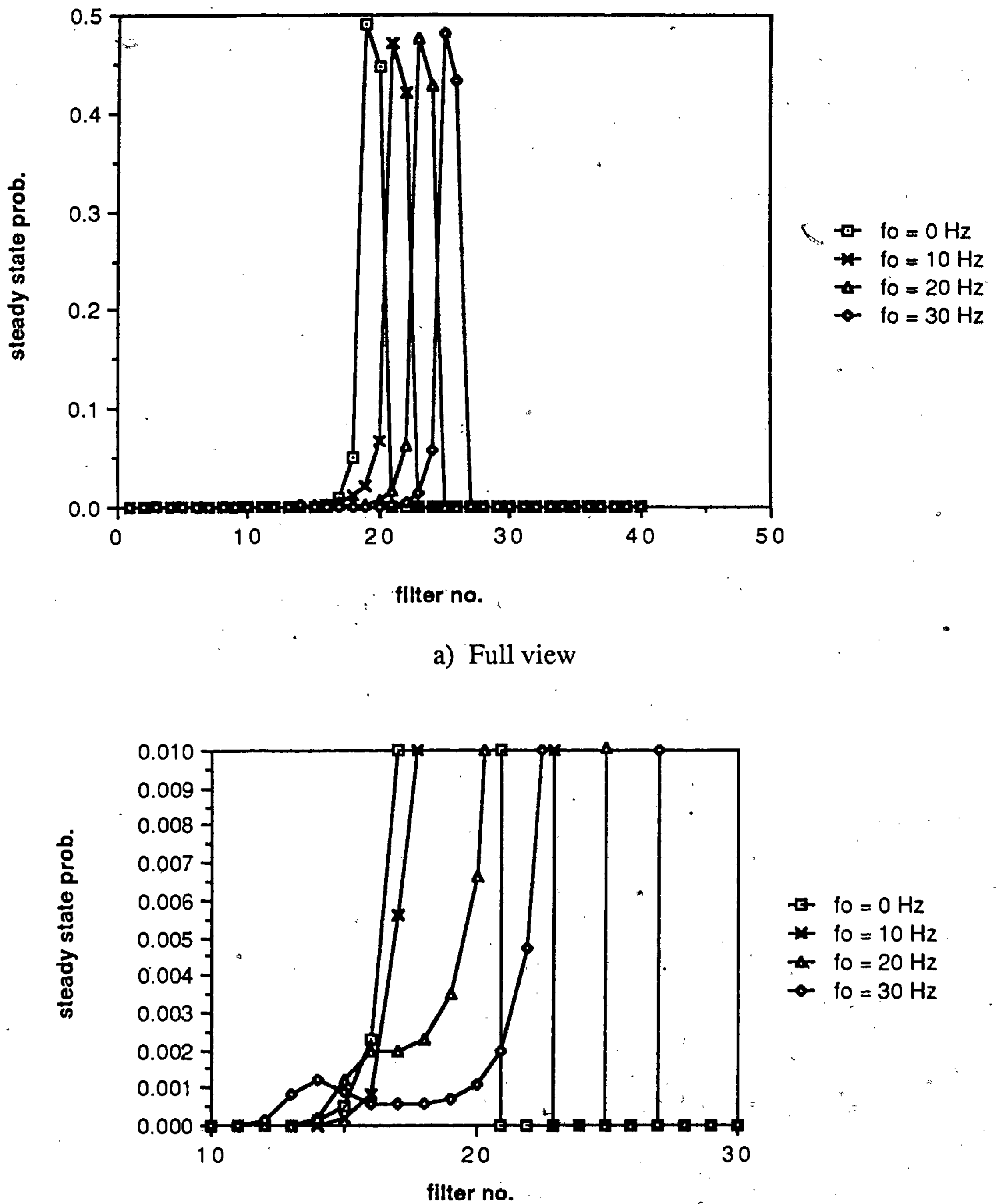


Figure 7.10 - Effect of Frequency Offset on the Average BER Performance of a Pilot Filter Implemented with the FSA



a) Full view

b) Expanded view of (a)

Figure 7.11 - Effect of f_o on Steady State Probabilities of FSA at E_b/N_o of 20 dB

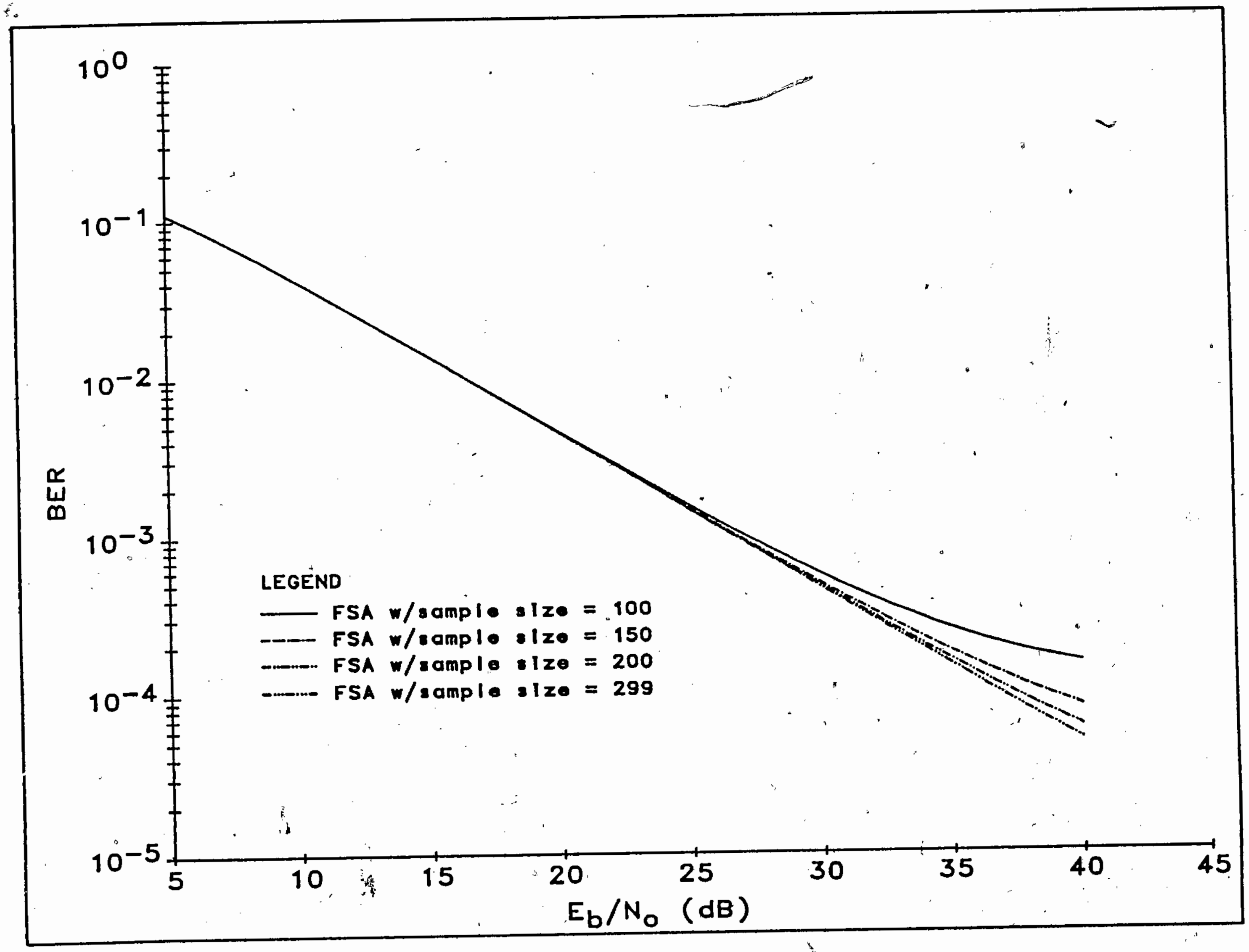


Figure 7.12 - Effect of Varying the Cross-Correlation Sample Size on the Average BER Performance of a Pilot Filter Implemented with the FSA

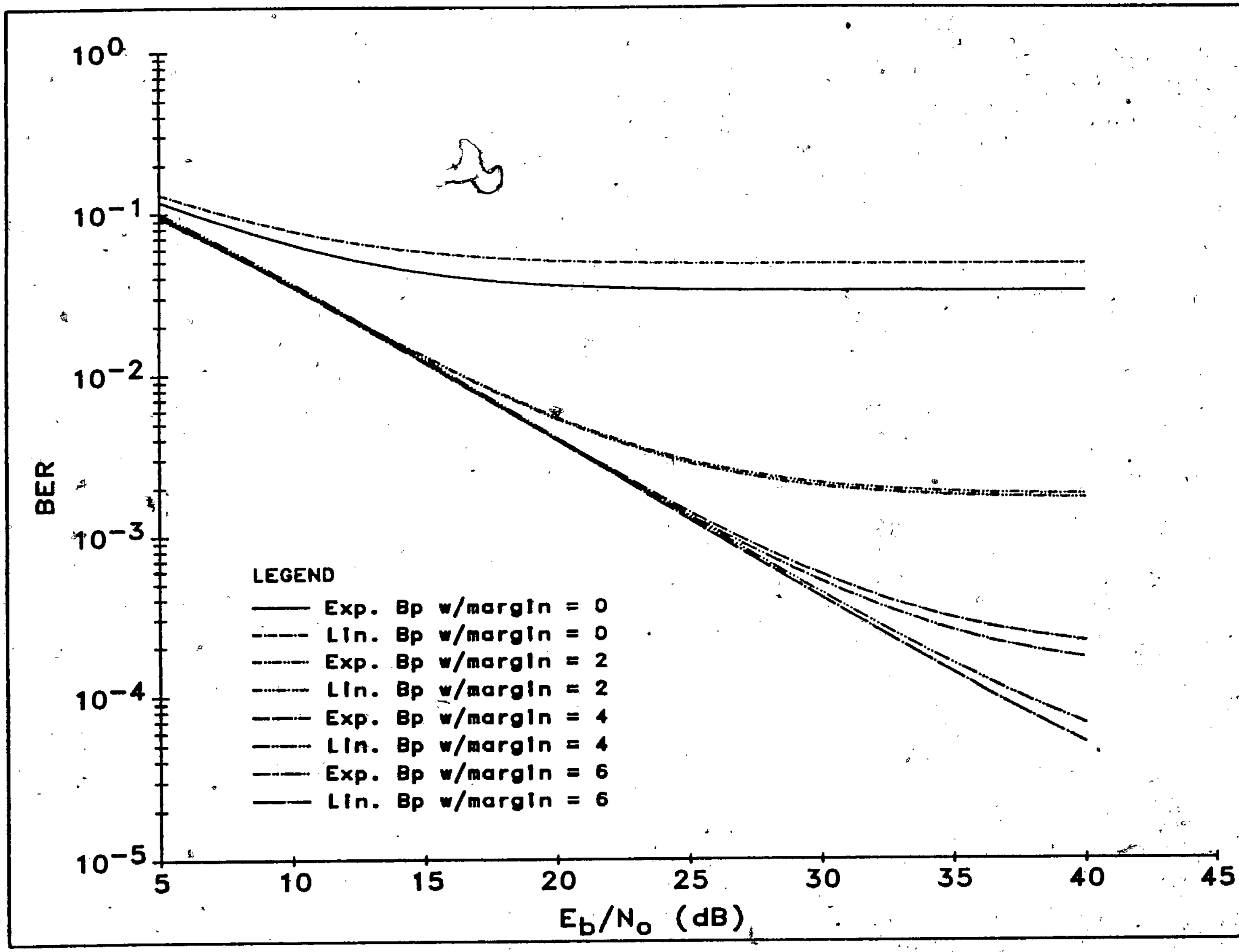


Figure 7.13 - Effect of Using Exponential Bandwidth Increment on the Average BER Performance of a Pilot Filter Implemented with the FSA at 50Hz Doppler

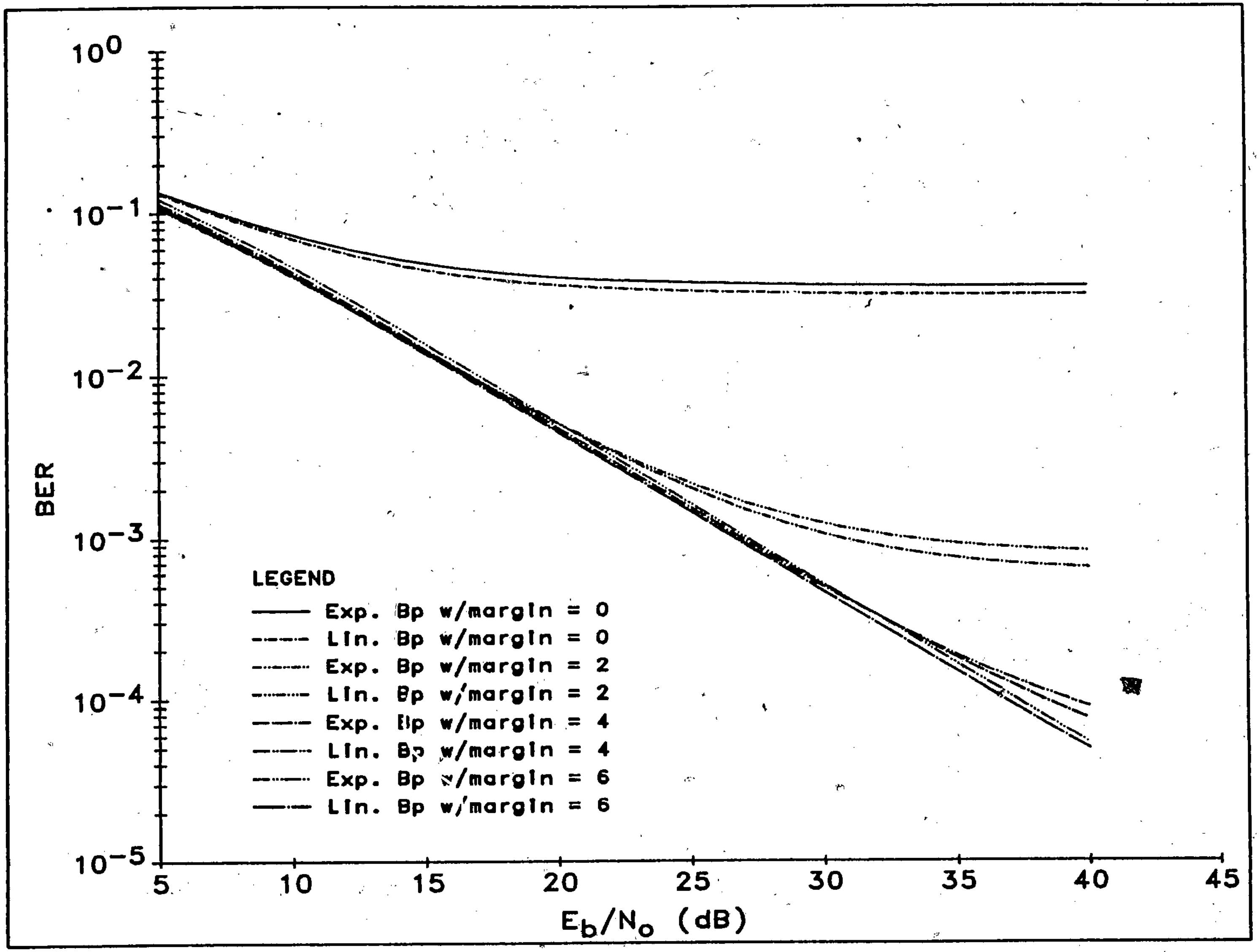


Figure 7.14 - Effect of Using Exponential Bandwidth Increment on the Average BER Performance of a Pilot Filter Implemented with the FSA at 100Hz Doppler

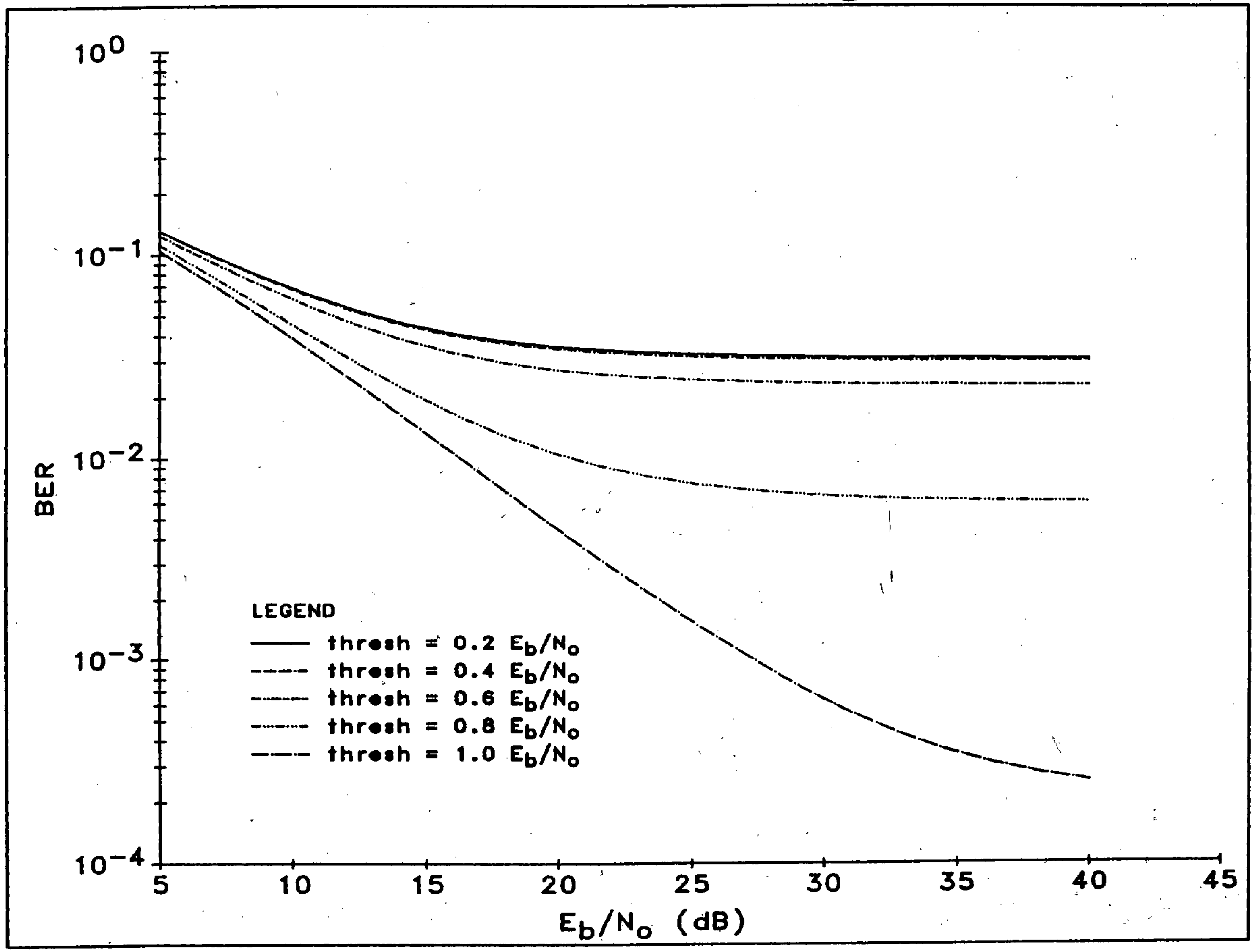


Figure 7.15 - Average BER Performance of a Pilot Filter Using FSA with Dual Thresholds

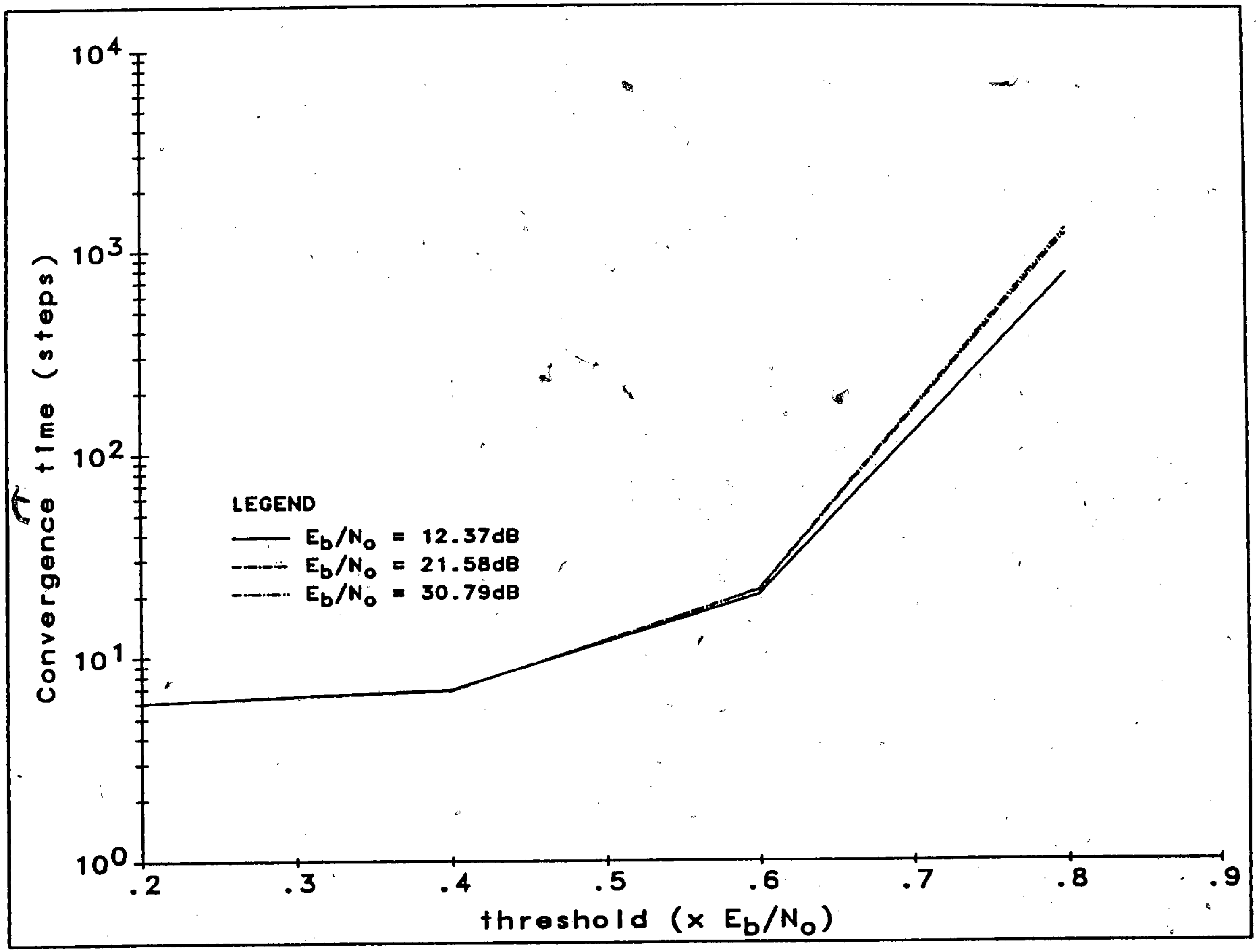


Figure 7.16 - Mean Convergence Time of a Pilot Filter Using FSA with Dual Thresholds

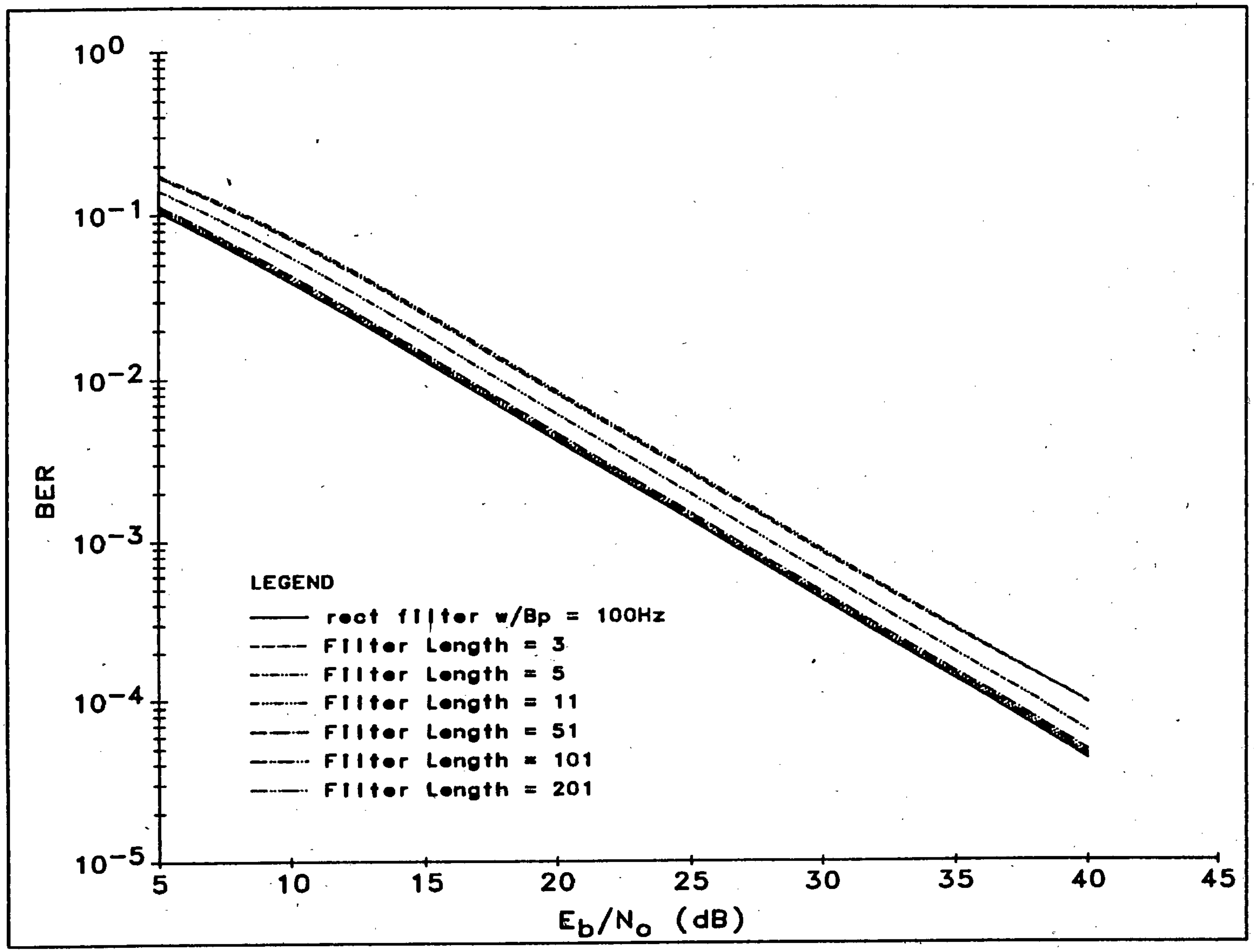


Figure 7.17 - Average BER Performance of a Pilot Filter Using FSA with Various Filter Lengths and a Moving Window Averager Lenth of 1

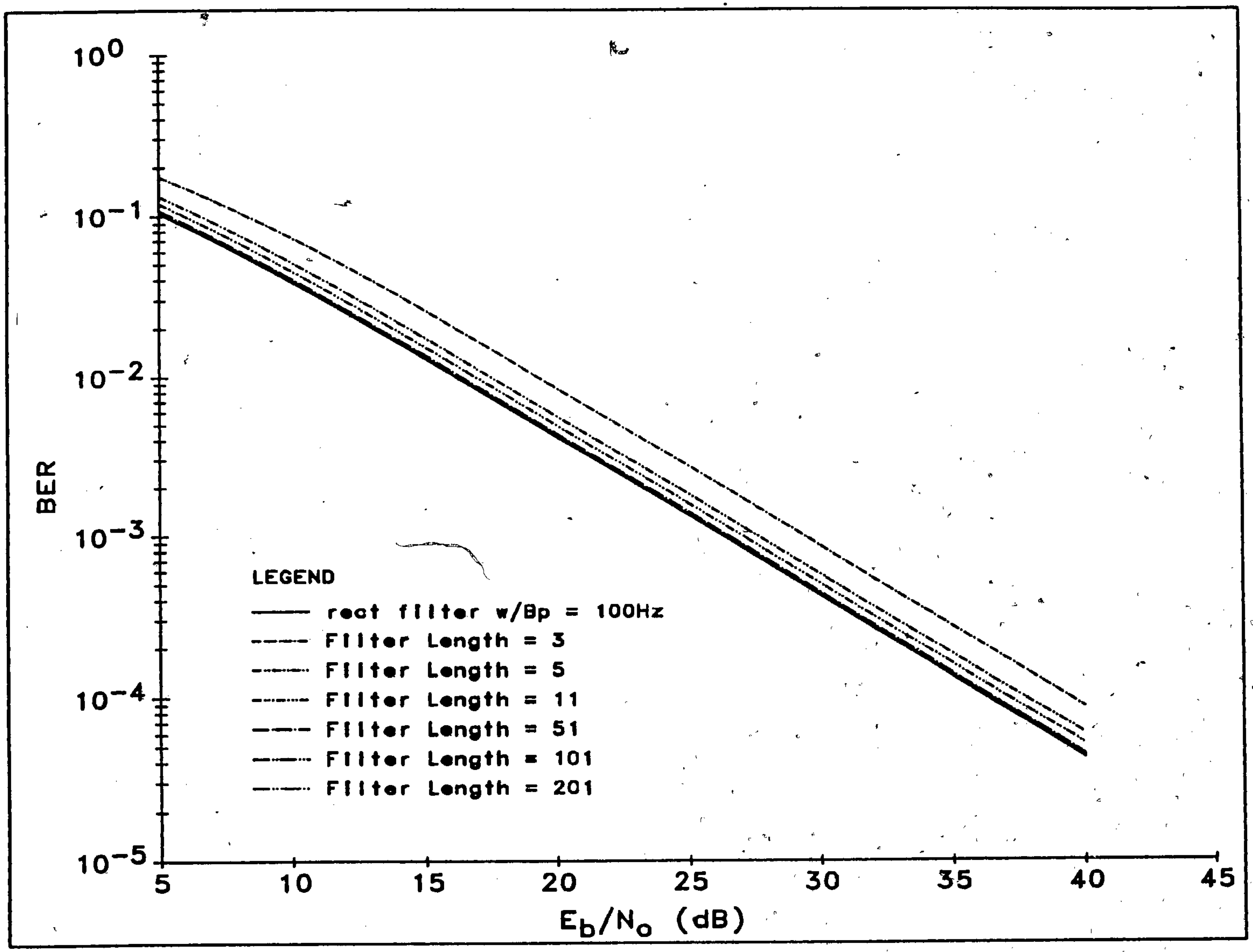


Figure 7.18 - Average BER Performance of a Pilot Filter Using FSA with Various Filter Lengths and a Moving Window Averager Lenth of 3

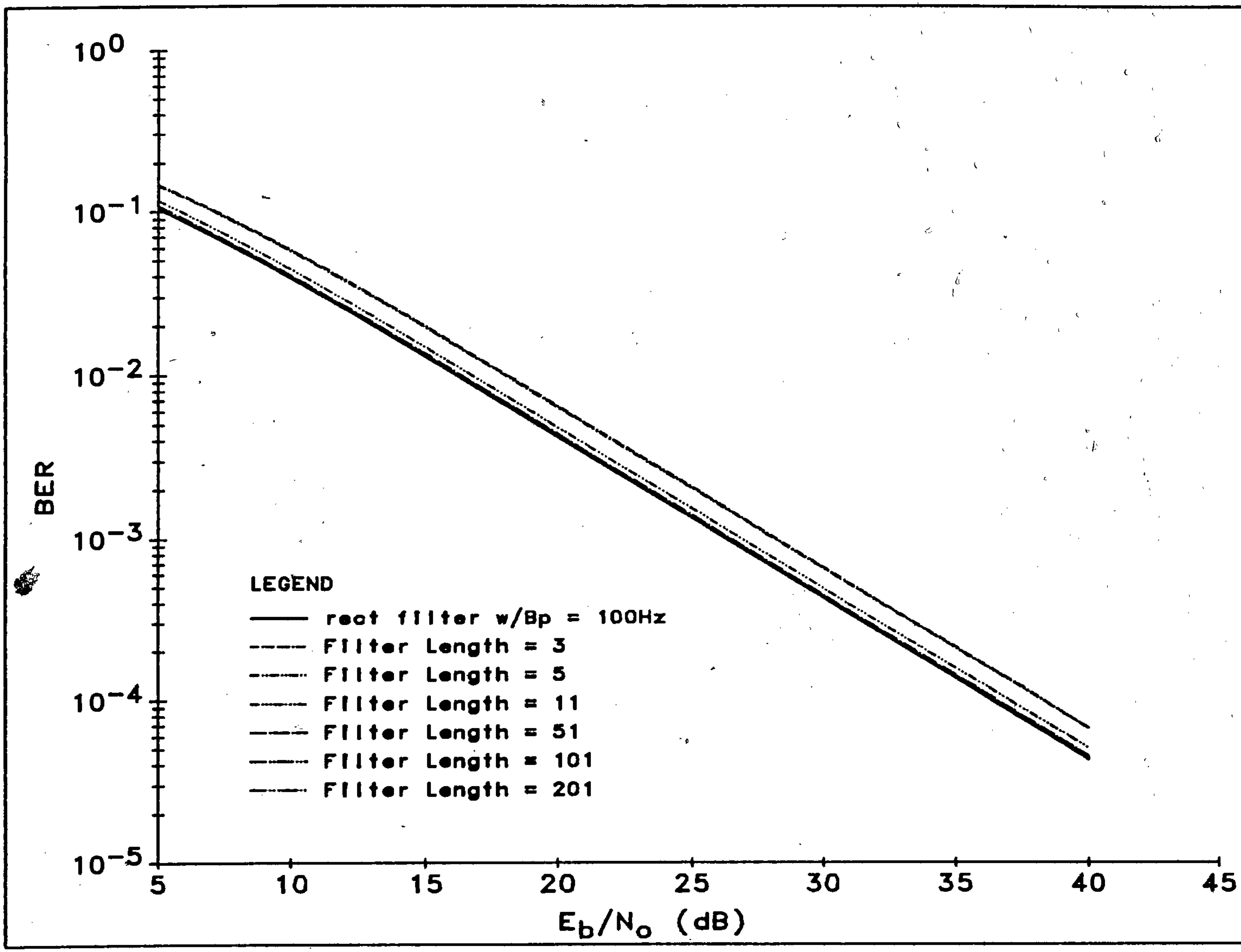


Figure 7.19 - Average BER Performance of a Pilot Filter Using FSA with Various Filter Lengths and a Moving Window Averager Lenth of 5

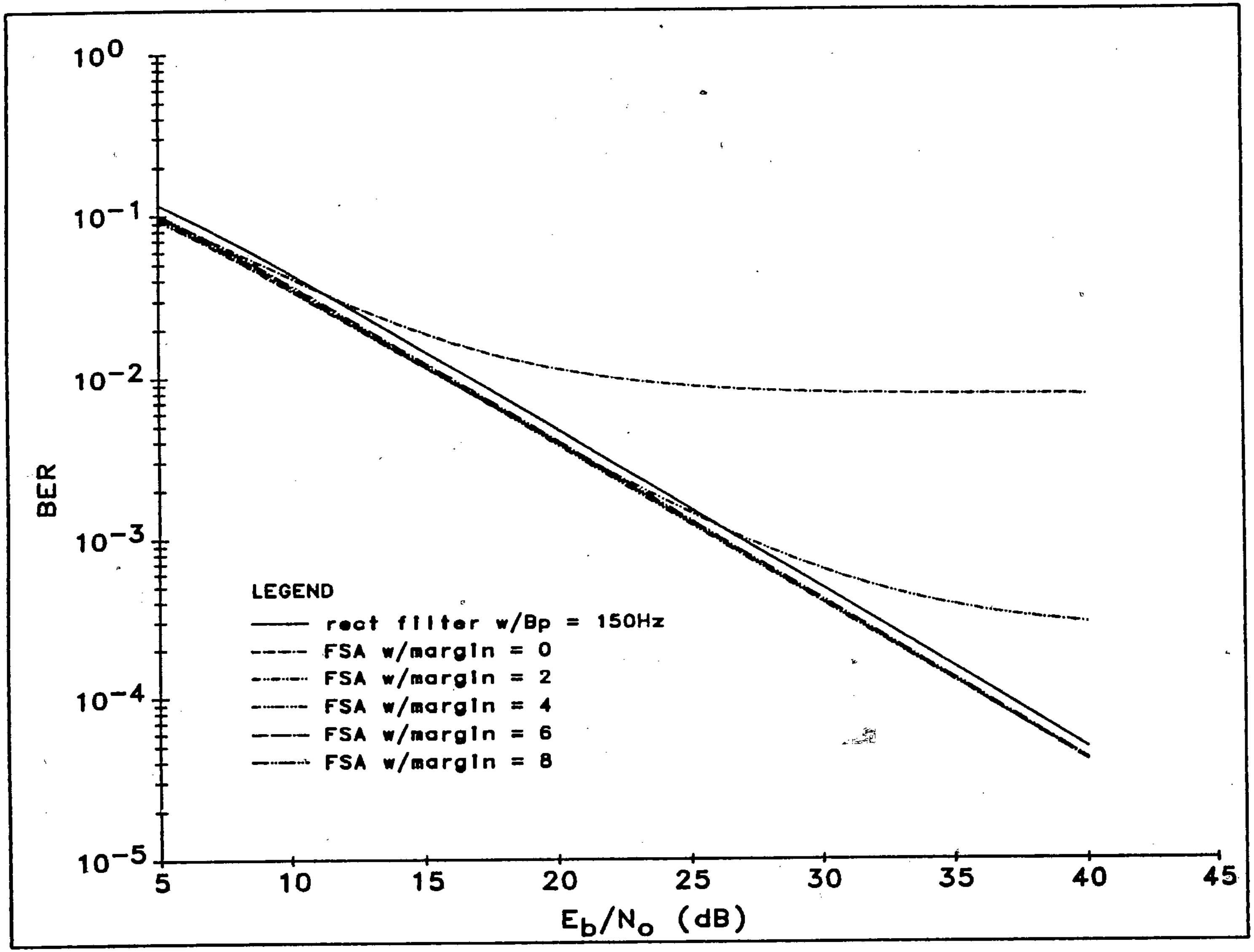


Figure 7.20 - Average BER Performance of a Pilot Filter Using FSA Implemented with a Set of Optimum FIR Filters

7.4. COMPUTED CONVERGENCE TIME BASED ON MARKOV MODEL

The effects of various parameters on the mean convergence time of the FSA are investigated in this section. The mean convergence time has been computed using (7.31) assuming that the first state (corresponding to B_p of 5Hz) is the starting state. The first state convergence time has been used here because, as found in section 7.3., convergence to a higher state is slower than to a lower state. As in section 7.3., all results obtained in this section assumed the following unless stated otherwise: ideal rectangular pilot filters, R_b = 2400bps, r = 0.2, f_o = 0, number of filters = 40, cross-correlation sample size = 299 and bandwidth increment ΔB_p = 5Hz.

7.4.1. Effect of E_b/N_o

E_b/N_o was found to have negligible effect on the convergence time. This is in contrast with the results found for pilot filter implemented using the SGT filter algorithm where convergence speed was found to be highly sensitive to E_b/N_o.

7.4.2. Effect of f_D

The effect of doppler frequency on the convergence time is shown in figure 7.21. The convergence time increases with increasing doppler. This is as expected because a wider bandwidth pilot filter is required to cover the fade spectrum at higher doppler which means that FSA needs to traverse more states before reaching the optimal state (or bandwidth). Since each transition is restricted to one step only, it therefore takes more iteration to arrive at the optimal state at high f_D. The convergence "speed", defined here as the average number of state traversed per iteration, remains nearly the same for 50 and 100Hz doppler.

7.4.3. Effect of Bandwidth Increment

Figure 7.22 shows the convergence time as a function of bandwidth increment. Based on this figure, we observed that increasing the bandwidth increment had the effect of decreasing the convergence time. This is of no surprise because, for a larger bandwidth increment, less states need to be traversed before reaching a given bandwidth. The convergence speed was faster for larger bandwidth increment.

7.4.4. Effect of Cross-Correlation Sample Size

Increasing the cross-correlation sample size decreases the variance in the cross-correlation estimate. The decrease in variance increases the upward transition probabilities and thus results in a decrease in convergence time in number of iterations. Recall that the time between iterations in number of bit periods is equal to the sample size. Hence, when we consider the convergence time in number of bits, increasing the cross-correlation sample size has the opposite effect of increasing the convergence time. Overall, the convergence time in number of bits increases almost linearly with increasing sample size. Figure 7.23 shows the convergence time as a function of cross-correlation sample size which illustrates this.

7.4.5. Effect of f_o

The effect of f_o on the convergence time is shown in figure 7.24. The convergence time was found to increase with increasing f_o; the amount of increase was most significant for f_o greater than 20Hz. The increase in convergence time was due to the decrease in upward transition probabilities, as explained in section 7.3.1., and to the increase in the optimal bandwidth as f_o was increased.

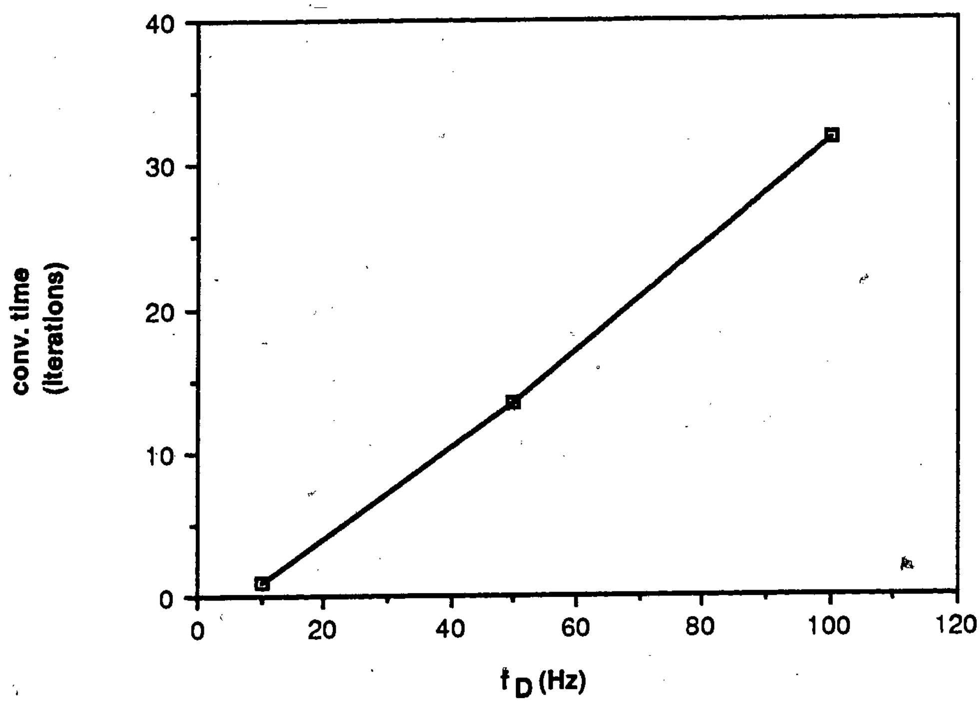


Figure 7.21 - Convergence time of the FSA vs Doppler Frequency

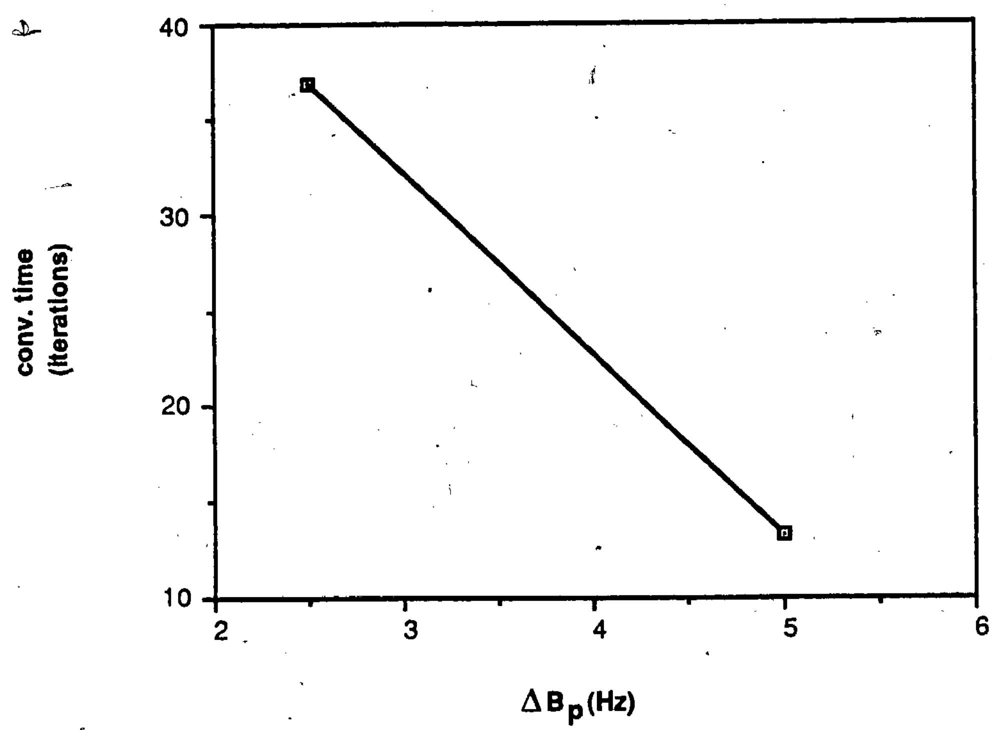


Figure 7.22 - Convergence time of the FSA vs. Bandwidth Increment

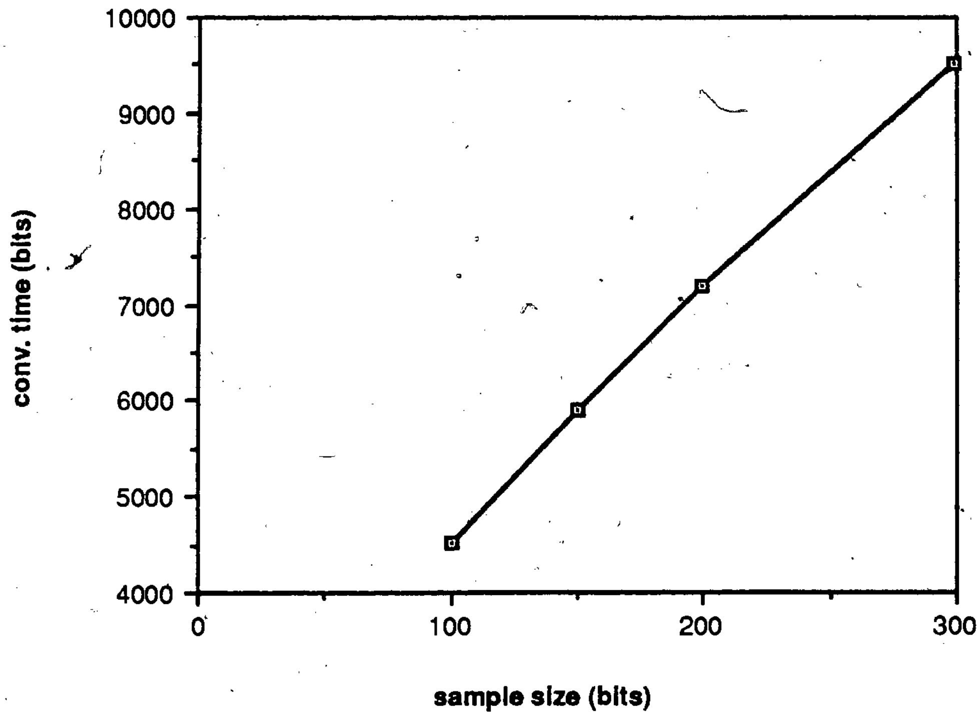


Figure 7.23 - Convergence time of the FSA vs. Cross Correlation Sample Size

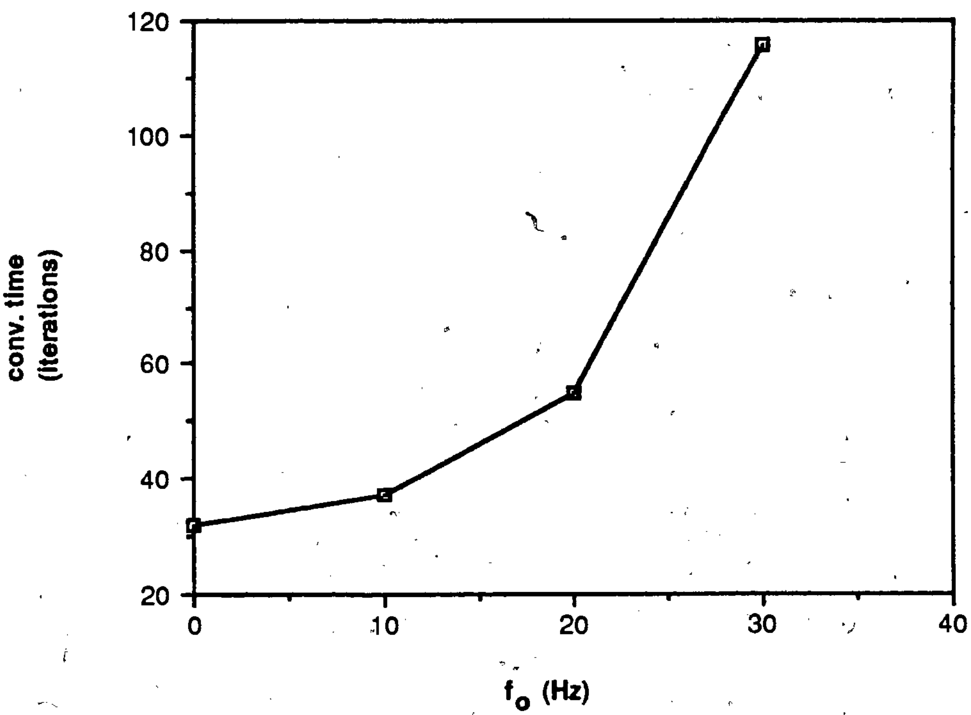


Figure 7.24 - Convergence time of the FSA vs. Frequency Offset

7.5. SIMULATION RESULTS

Simulations were performed to determine the accuracy of the model and to assess the effects of self-noise and decision direction. The simulation method used to obtain the results given in this section was the same as that outlined at the beginning of section 5.5.

7.5.1. Accuracy of the Markov Model

Figure 7.25 shows the steady state probabilities from the simulated and computed values for f_D of 50Hz and a bandwidth increment of 5Hz.

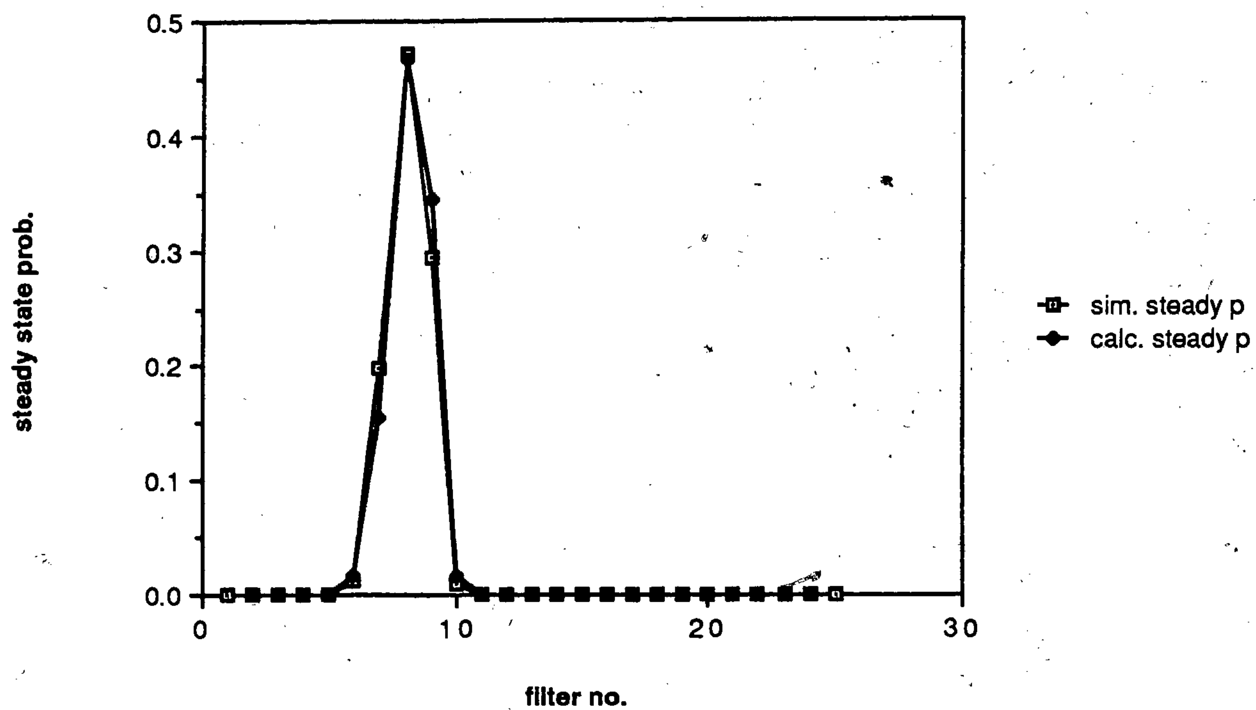


Figure 7.25 - Calculated vs. Simulated Steady State Probabilities of the FSA

Some discrepancies can be noted for filters which are one step wider and one step narrower than the optimum. These have been attributed to the deviation of the sampled cross-correlation from the Gaussian approximation. One should note that the filter bandwidth with the highest steady state probability indicated by figure 7.25 is 40Hz (state 8), not 50Hz (state 10) as would be expected if rectangular filters are used. The particular distribution of steady state probability was again a result of the gradual roll-off of the FIR filter frequency response.

7.5.2. Effect of Decision Direction and Self-Noise

Results obtained in section 7.4 assumed perfect removal of data dependence in the reference signal, $\underline{u}(k)$. In order to investigate the effect of decision direction, simulations were performed for f_D = 50Hz, E_b/N_o = 10dB, r = 0.2 using demodulator decision to remove the data dependence in $\hat{u}(k)$. (See figure 2.1). Figure 7.26 shows the results which indicates that decision direction has very little effect on the average steady state probabilities even at low E_b/N_o. The simulated average steady state probabilities were very close to the calculated value.

Previous results had also assumed that the pilot tone and data signal were transmitted separately so that the results obtained were not affected by self-noise. We investigated the effects of self-noise by transmitting the pilot and data signal over the same (simulated) channel as shown in the system model given in figure 2.1. The following parameters were used in the simulations: f_D = 100Hz, E_b/N_o = 20dB, r = 0.2. Fast fading was chosen because it introduced larger spectral spread so that if the effect of self-noise is small for large f_D, then the effect will be even less at smaller f_D. Results of the simulation are summarized in figure 7.27. Here, it shows that the effect of self-noise on the average steady state probabilities is again small but there is larger discrepancy between the values predicted by the Markov model and the simulated results. One should note from figure 7.27 that this discrepancy is not related to self-noise, but is due to the increased statistical variations as a result of a larger doppler frequency.

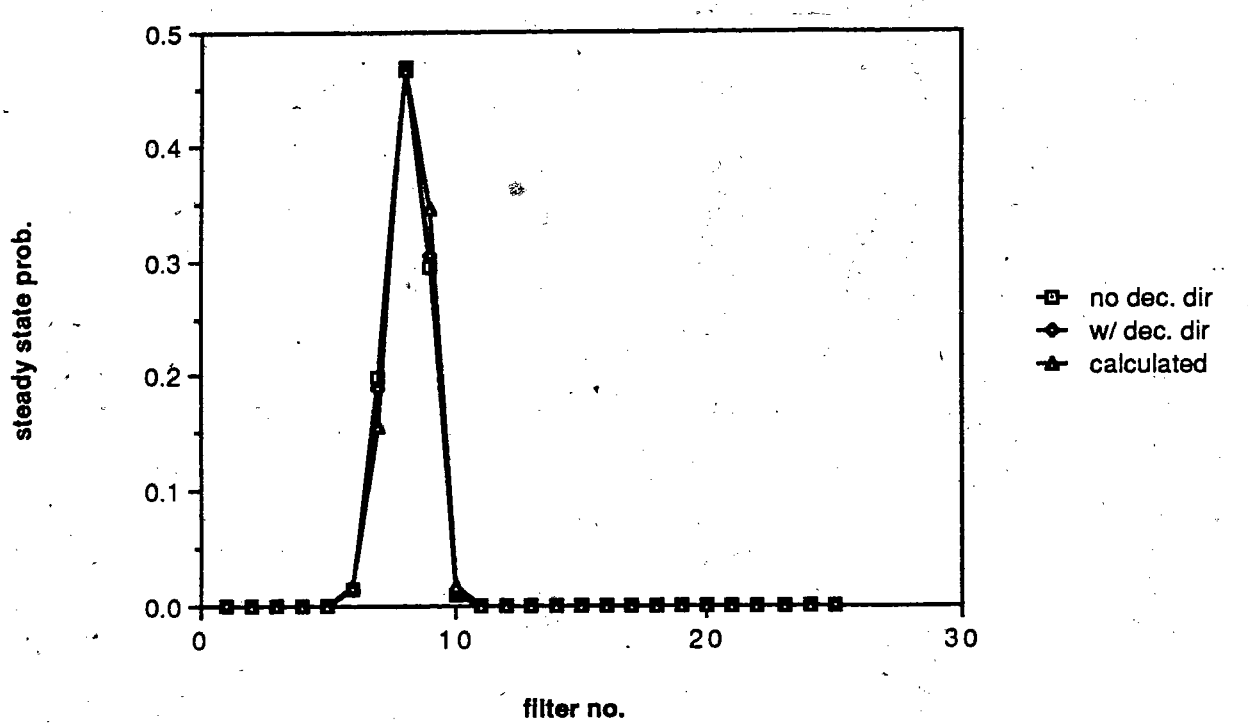


Figure 7.26 - Simulated Steady State Probabilities of the FSA with Decision Direction

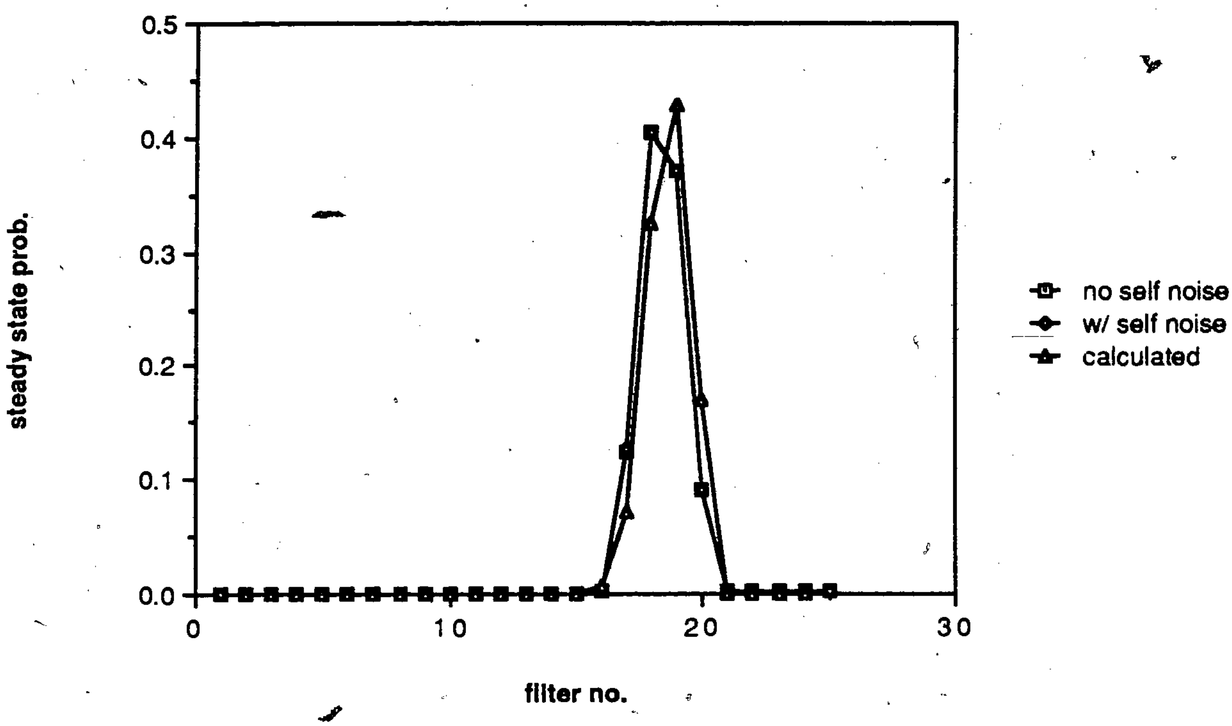


Figure 7.27 - Simulated Steady State Probabilities of the FSA with Self-Noise

7.6. SOME IMPLEMENTATION CONSIDERATIONS

Recall that the filter switching algorithm described thus far assumes that the adaptation period is equal to the sample size of the sampled cross-correlation. This implementation is the simplest in terms of analysis and complexity. By changing the adaptation period, tradeoffs between complexity and convergence speed can be made. One variation is to take an adaptation step every bit period, which is equivalent to the use of an adaptation period of one. Intuitively, this scheme should offer great improvement in convergence speed because the algorithm can now make transitions more often. Successive values of the sampled cross-correlation in this case will be highly correlated thus making the Markov model invalid. One will have to rely on computer simulations to determine the BER and convergence behaviors. The difference in computational complexity between the two algorithms is significant. For sample size N and filter length M, the original algorithm requires approximately M multiply-add operations per bit to compute the sampled cross-correlation whereas the second algorithm requires M x N multiply-adds per bit. Other values of the adaptation period will provide different tradeoffs between complexity and convergence speed.

When implementing the FSA in practice, it is necessary to assign different weights to each cross-correlation sample in forming the sample mean because the fading process is non-stationary in general. It is possible to use an exponential decay averaging scheme for computing the cross-correlation estimate. The cross-correlation estimate can be computed as:

$$x(k+1) = \lambda\, x(k) + \mathrm{Re}[\hat{u}(k)\, v^*(k)] \tag{7.32}$$

where λ is the aging coefficient. This method is simpler to implement than a sample averager and may provide satisfactory results, but only simulations can determine its performance.

7.7. SUMMARY OF RESULTS

Results obtained for the filter switching algorithm can be summarized as follow.

- Compared with a non-adaptive pilot tone calibration system using a rectangular filter with bandwidth of 150Hz, the FSA provided an improvement in average BER which was a function of the doppler frequency. At a BER of 10^{-2}, the improvement was about 0.3dB for 100Hz doppler, 1.0dB for 50Hz doppler and almost 2.0dB for a 10Hz doppler.

- Increasing the step margin had the effect of reducing the error floor but at the same time, increasing the average BER at low E_b/N_o. A step margin corresponding to a frequency margin of 60Hz gave a good compromise.
- The bandwidth increment and the number of reduced coefficient filters required should be chosen such that $B_{p\,max} > (f_{D\,max} + \text{frequency margin})$.
- The maximum frequency offset in the fade spectrum which could be tolerated was 10Hz. Larger offset had the effect of raising the error floor. Increase in the error floor could be compensated for by increasing the step margin, at the expense of degrading the BER at low E_b/N_o.
- For a 2400bps system, a minimum cross-correlation sample size of 150 was necessary to give a reasonably low error floor. Decreasing the sample size had the effect of increasing the error floor.
- Using a constant bandwidth spacing of 5Hz was found to be effective. For smaller spacing, BER improved slightly but more filters were required and convergence time also increased. The reverse was true for larger spacing.
- The set of reduced coefficient filters could be designed effectively using mean squared error optimization. Compared to rectangular filters, the set of optimum filters required less step margin to give the same error floor.
- The optimal combination of MWA and shaping filter length was 5 and 51 respectively.
- Worst case convergence time was 32 steps for a cross-correlation sample size of 299. With a 2400bps system, this corresponds to 4 sec.
- E_b/N_o has very little effect on convergence time. However, it increases significantly with increasing f_o.
- Effects of decision direction and self-noise were negligible.

7.7.1. Comparison with SGT Filter

Figure 7.28 shows the BER vs E_b/N_0 curves for pilot filter implemented using the FSA and the SGT algorithm at various doppler frequencies. The filter length was 11 and the MWA length was 3 for both cases. Except for f_DT of 0.00417 (10Hz at 2400bps), the FSA performed better than the SGT filter. The difference in performance at a BER of 10^{-2} was -0.5, 0.7 and 1.1 dB for f_DT of 0.00417, 0.0208 and 0.0417 respectively. One should keep in mind that the BER performance of the SGT filter in the presence of self-noise is significantly deteriorated whereas for the FSA, self-noise has very little effect on BER. Also, recall that the BER performance of the SGT filter deteriorates with increasing filter length beyond 11 whereas the performance of the FSA increases with filter length. Figure 7.29 shows the BER curve for FSA with length 51 filters and SGT filter of length 11. The FSA out-performed the SGT filter for all f_DT. The difference in BER performance in this case was 0.2, 0.8 and 1.2 dB at 10^{-2} BER for dopplers of 0.00417, 0.0208 and 0.0417.

In terms of convergence behavior, pilot filter using the FSA is also superior than a SGT pilot filter. Figure 7.30 shows the learning curves from simulations of stepwise increase in f_DT from 0.00417 to 0.0417 at E_b/N_0 of 40 dB for both implementations. Length 51 filters were used for the FSA and length 5 filter was used for the SGT algorithm. We can see that once converged, the FSA had no difficulty in tracking increases in the doppler frequency. As for SGT filter, there were jumps in BER due to the inability of the SGT algorithm to follow the changes in doppler at high E_b/N_0. The difference in convergence performance between the two algorithms would be smaller at low E_b/N_0.

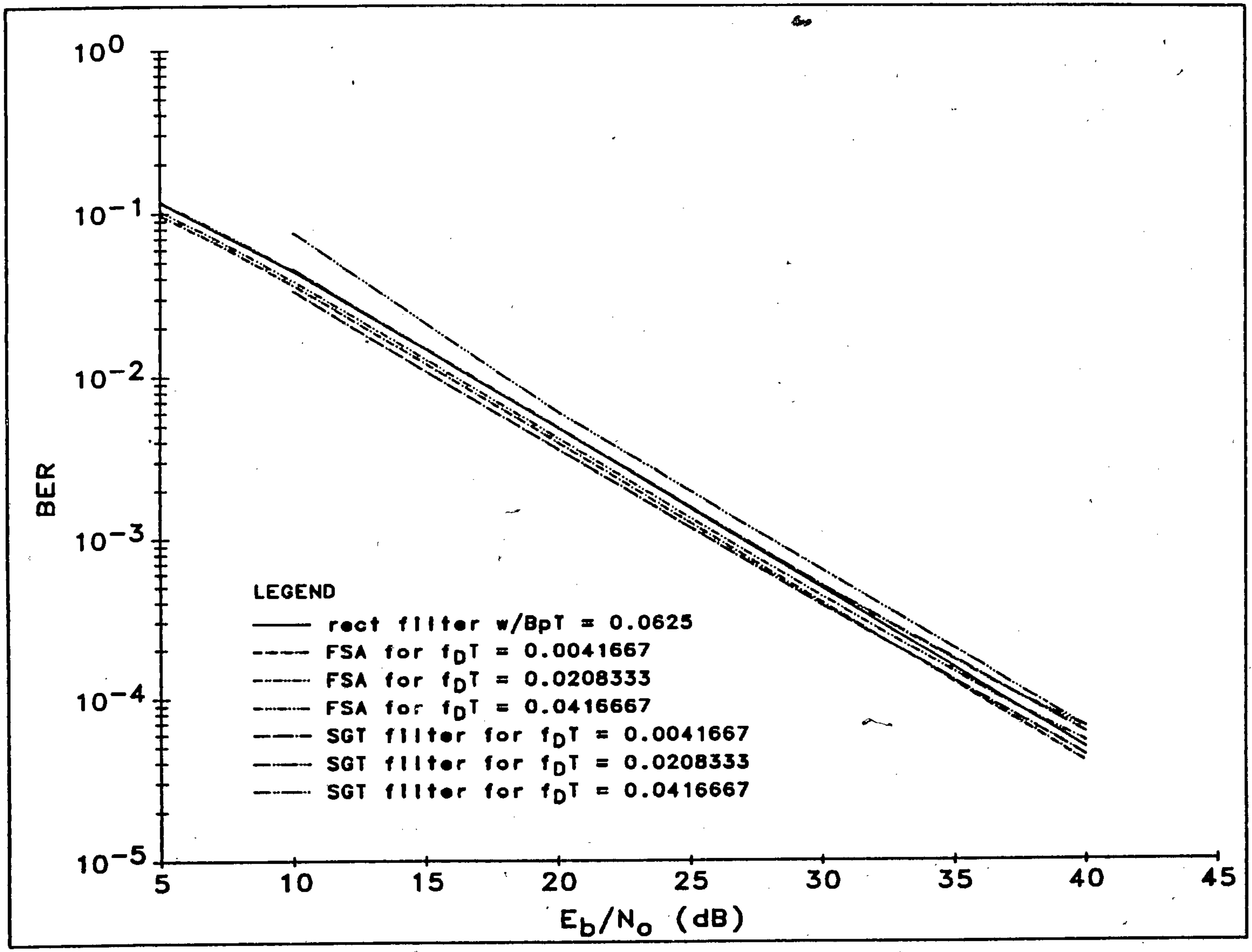


Figure 7.28 - Comparison of Average BER Performance Between a Length 11 SGT Pilot Filter and a Pilot Filter Using FSA Implemented with Length 11 Filters.

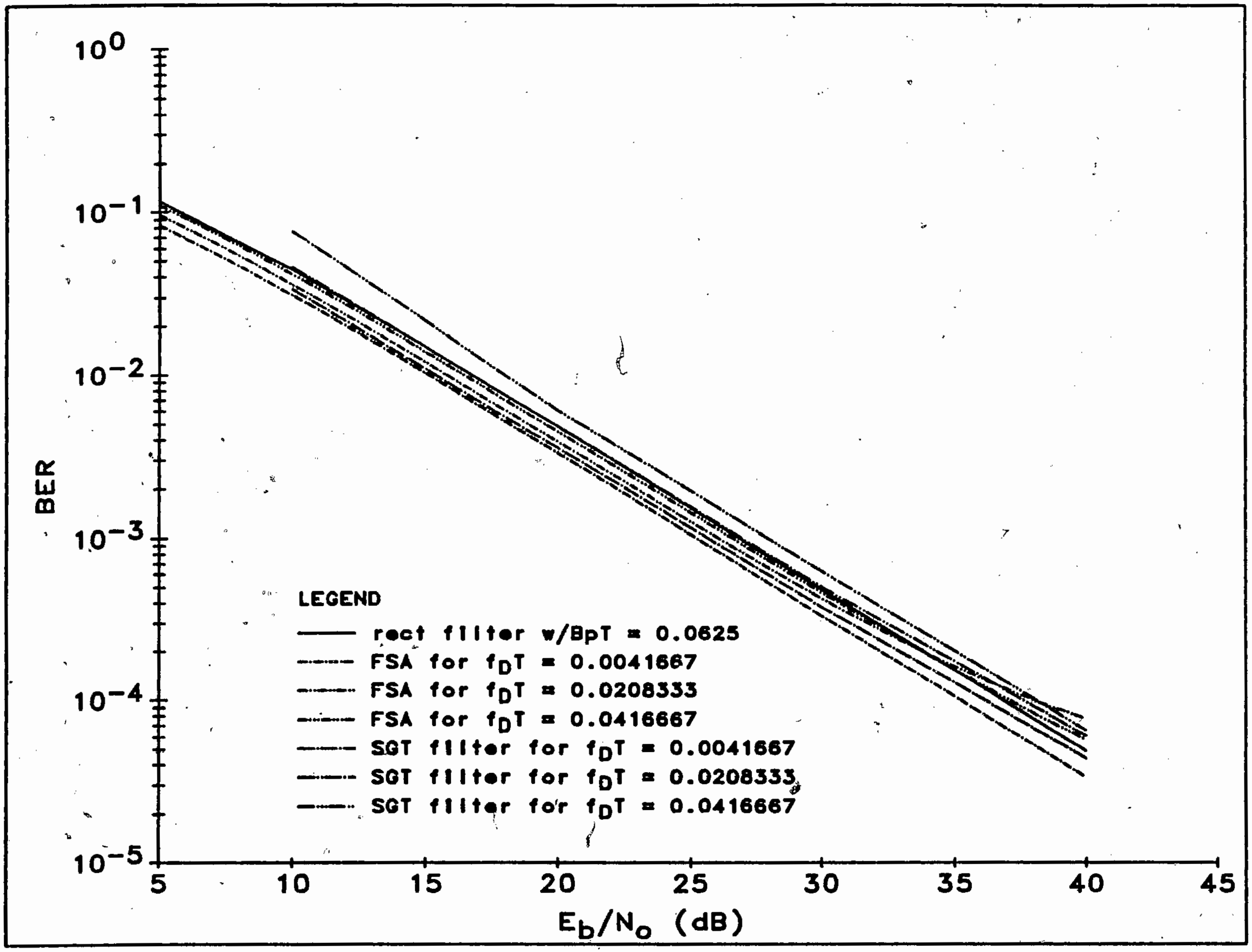


Figure 7.29 - Comparison of Average BER Performance Between a Length 11 SGT Pilot Filter and a Pilot Filter Using FSA Implemented with Length 51 Filters.

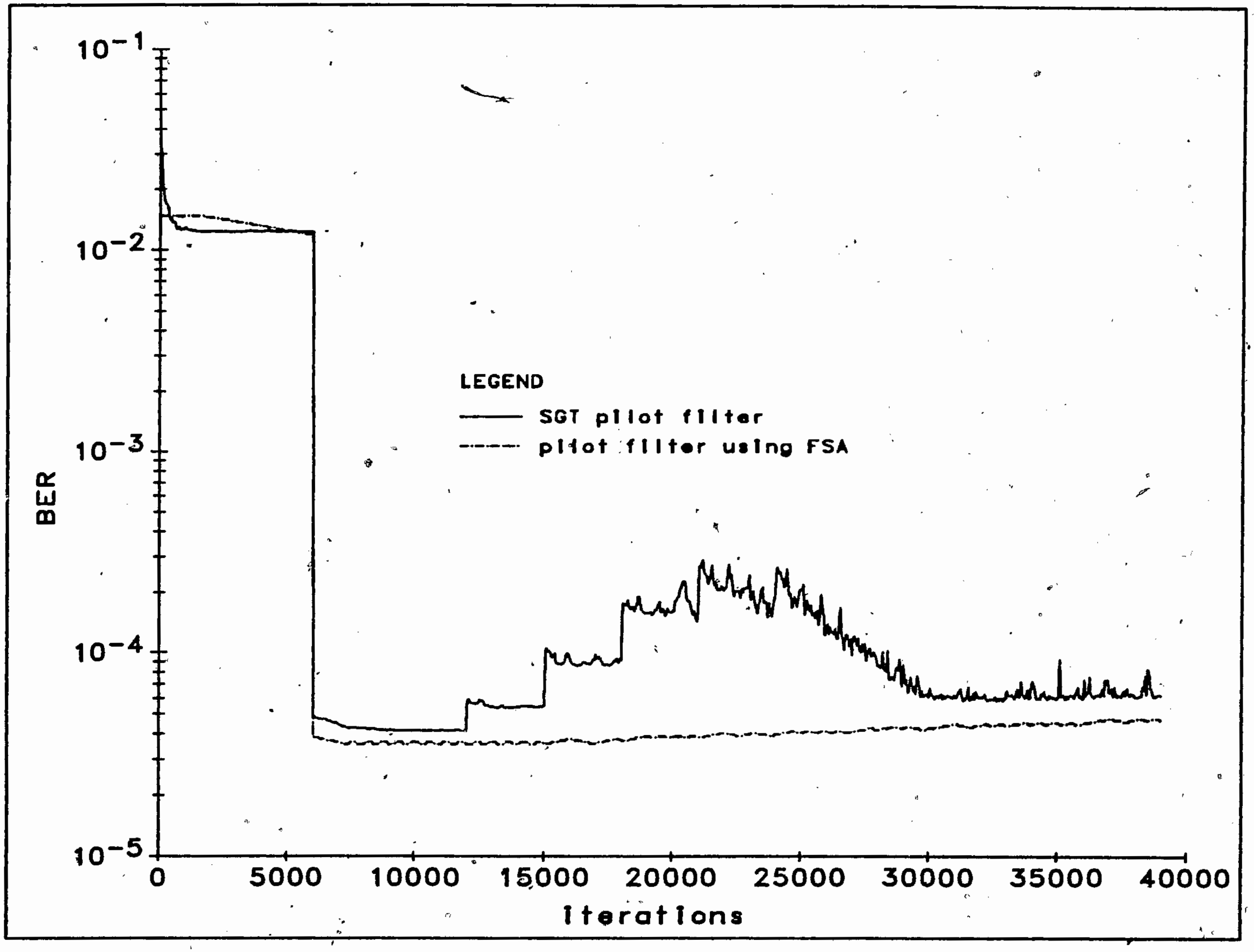


Figure 7.30 - Comparison of Convergence Behavior Between a Length 5 SGT Pilot Filter and a Pilot Filter Using FSA Implemented with Length 51 Filters

7.8. FREQUENCY OFFSET ADJUSTMENT

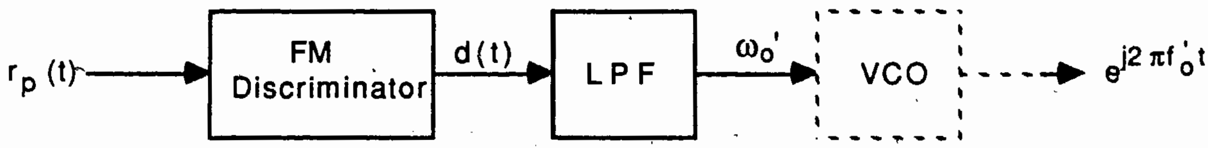
The frequency offset which exists in the spectrum of the fading process is a direct result of the difference in frequency between the transmit and receive oscillators. This offset can be considered constant for a particular transmitter-receiver pair. Therefore, in order to provide an appropriate correction for this offset, one needs only to estimate it once during system initialization. The offset estimate, denoted by f_o', can then be used to de-rotate the received samples during subsequent operation by multiplying the samples with $e^{-j2\pi f_o' kT}$. This realization suggests that there is no need for dynamic tracking of the frequency offset. It is one of the reasons why frequency offset adjustment has not been incorporated into the FSA as a two-dimensional adaptation. In this section, we consider a method for estimating the frequency offset by the use of a FM discriminator.

Figure 7.31 shows block diagram and the associated model of the subsystem for estimating the frequency offset using a FM discriminator.

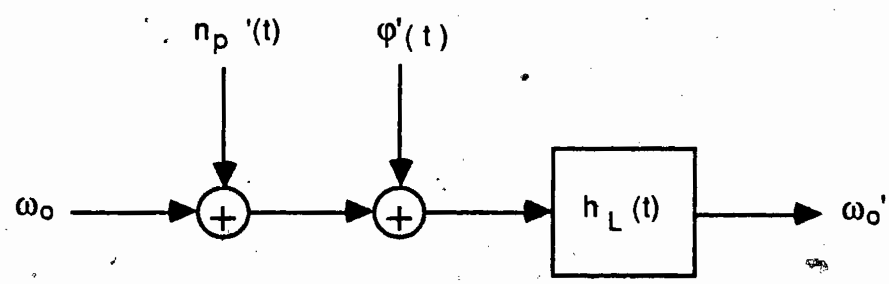
The discriminator is preceded by a limiter which limits the amplitude fluctuation of the input signal. The input to the limiter, $r_p(t)$, is given in discrete time by (2.14). In continuous time, $r_p(t)$ can also be written as:

$$r_p(t) = a\, A(t) \exp\,(j\omega_o t + \varphi(t)) + n_p(t) \qquad (7.33)$$

where the complex gain c(t) has been represented by its amplitude and frequency components. The use of continuous time representation is for simplicity only. Results obtained are also applicable for discrete time implementation.



(a) Block Diagram



(b) Model

Figure 7.31 - Block Diagram and Model of Frequency Offset Estimation Subsystem with FM Discriminator

The frequency discriminator, assumed to be ideal with unity gain, is followed by a low pass filter (LPF) for noise reduction. Output of the discriminator is:

$$\begin{aligned} d(t) &= \omega_o + \frac{d\varphi(t)}{dt} + \frac{dn_p(t)}{dt} \\ &= \omega_o + \varphi'(t) + n_p'(t) \end{aligned} \tag{7.34}$$

Estimate of the frequency offset in radian is given by:

$$\omega_o' = d(t) * h_L(t)$$
$$= \omega_o H_L(0) + \varphi'(t) * h_L(t) + n_p'(t) * h_L(t) \qquad (7.35)$$

where $h_L(t)$ is the impulse response of the LPF and $H_L(0)$ is the d.c. value of the Fourier transform of $h_L(t)$. From (7.35), we see that the output of the subsystem has a mean value which is proportional to the desired frequency offset but it is perturbed by two noise terms. To investigate the amount of variations expected from the frequency offset estimate, we need to calculate its variance. Assuming that $\varphi(t)$ and $n_p(t)$ have zero means, the variance of ω_o' is given by:

$$\sigma_{\omega_o'}^2 = \int_{-\infty}^{\infty} N_p'(\omega)\, |H_L(\omega)|^2\, d\omega + \int_{-\infty}^{\infty} \phi'(\omega)\, |H_L(\omega)|^2\, d\omega \qquad (7.36)$$

where $N_p'(\omega)$ and $\phi'(\omega)$ are the Fourier transforms of $n_p'(t)$ and $\varphi'(t)$ respectively. Statistics of the random FM component $\varphi'(t)$ has been studied elsewhere [6, 7, 23]. Its spectrum has been numerically computed and has the shape which is shown in figure 7.32 [23].

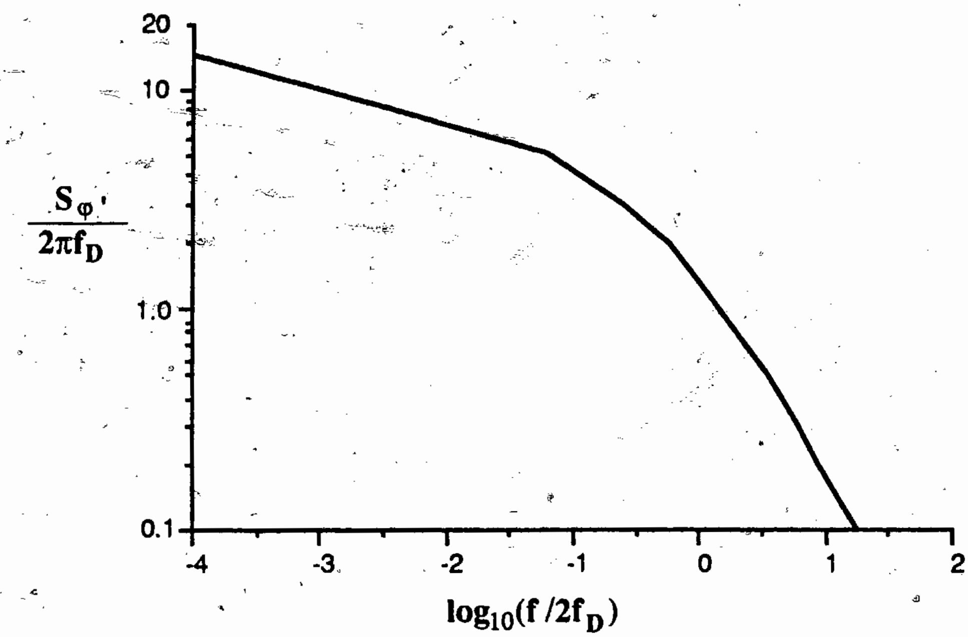


Figure 7.32 - Power Spectrum of Random FM (taken from [23])

$n_p'(t)$ is AWGN passed through a discriminator. From FM communication theory [24], we know that $n_p'(t)$ has a parabolic spectrum given by:

$$N_p'(\omega) = \frac{\omega^2}{4\pi^2 r_{cn}} \tag{7.37}$$

where r_{cn} is the carrier to noise ratio.

If we assume that the LPF is ideal with unity gain and bandwidth W, then the two integrals in (7.36) can be evaluated using figure 7.32 and (7.37) to give:

$$\sigma_{\omega_o'}{}^2 \approx 10\,\pi\, f_D\, W \tag{7.38}$$

A typical carrier to noise ratio of 10dB is assumed in obtaining (7.38). The average amount of variations in the frequency offset can be expressed as:

$$\sigma_{f_o'} \approx 0.892 \sqrt{f_D \ B} \qquad (7.39)$$

where B is the single sided bandwidth of LPF in Hz.

From (7.39), we see that the deviation in f_o is approximately $8.92 \sqrt{B}$ for 100Hz doppler. In order to maintain a small deviation, we need to have a very narrow filter. For example, to obtain a deviation of less than 10Hz, which is about the maximum frequency offset tolerable by the FSA, we will need a LPF with bandwidth smaller than 2.5Hz. Although a filter with this narrow a bandwidth will have a long delay (~ 60ms), this delay will not pose a problem because the LPF is not in the data processing path.

8. CONCLUSIONS AND RECOMMENDATIONS

This thesis has addressed the issue of introducing adaptivity into the pilot filter for a tone aided transmission system. Adaptivity is introduced to allow the pilot filter to operate at or near optimum bandwidth under changing vehicle speed. The present study investigated the performance of various adaptive filtering schemes. The general approach taken in this study is to first derive the various adaptation algorithms in the context of pilot filtering. The BER and convergence performance are then analyzed wherever possible, and supported with simulation results.

Two conventional adaptive filtering algorithms have been investigated: the Stochastic Gradient Transversal filter and the Stochastic Gradient Lattice Joint-Process Estimator. Although the two stochastic gradient algorithms use minimum mean square error as criterion, it has been shown here that minimizing the mean square error between the pilot filter output and the data derived reference is equivalent to the minimization of BER.

Analysis has shown that the convergence speed of the SGT filter is highly sensitive to the eigenvalue spread of the input correlation matrix. The smaller is the eigenvalue spread, the slower is the convergence speed. The most important parameter affecting input eigenvalue spread is E_b/N_0 because of its large dynamic range. Input eigenvalue spread is found to be directly proportional to the E_b/N_0. At high E_b/N_0, the eigenvalue spread is very large, which means that convergence is very slow. Fortunately, practical systems generally operate at the vicinity of 20dB where eigenvalue spread has been shown not to be a problem.

Because of the stochastic gradient approximation, there is an excess loss associated with the BER performance of the SGT filter so that its BER is always higher than that achievable with the optimal filter. This is also true for the SGT-JPE. However, it is found that if one forces the SGT filter coefficients to be conjugate symmetric, then the excess loss can be reduced without affecting convergence. The excess loss is an increasing function with increasing doppler and it can be as much as 0.5 dB at 4% doppler. This makes the use of a SGT pilot filter suitable only for low vehicle speed because the maximum performance gain of an adaptive system over a non-adaptive system is already very small at high doppler.

For most applications, the SGL-JPE is expected to perform better than the SGT filter, especially with respect to convergence speed, because of the decorrelating property of the lattice section used in the SGL-JPE. However, both numerical analysis and simulation results have shown that the SGL-JPE is not suitable to be used in the pilot filtering application because of the high sensitivity of the BER to fluctuations in the SGL-JPE coefficients.

Due to the various problems associated with the two stochastic gradient algorithms, a novel approach has been taken, leading to the development of a new adaptive filtering algorithm. By reducing the adaptivity to one dimension, namely bandwidth, the filter switching algorithm is capable of achieving high convergence speed at the expense of processing complexity.

In order to analyze the performance of the algorithm, a dual Markov model has been introduced. This model enables the computation of both the average BER and the convergence speed. Simulations have been performed which showed that the model has a high degree of accuracy. Although the model is valid only for the FSA implemented in its simplest form, it provides insights into the effect of various parameters on the algorithm performance. Using the Markov model, the FSA, even in its simplest and slowest form, is found to converge fast enough to track the changes in doppler during vehicle acceleration. BER performance of the FSA has been compared with that of a non-adaptive system using a rectangular pilot filter with bandwidth of 150Hz. For a 2400 bps system operating at a BER of 10^{-2}, the improvement is 0.3dB for 100Hz doppler, 1.0dB for 50Hz doppler and 2.0dB for 10Hz doppler. In addition to these improvements, the FSA is very robust in the sense that self-noise and decision direction have very little effect on its performance. Convergence time is also found to be insensitive to changes in E_b/N_0.

Compared to the two stochastic gradient algorithms, the FSA provides better performance in both BER and convergence speed. The FSA can be used in other filtering applications where bandwidth adaptivity is important. However, one must evaluate the suitability of using cross-correlation as performance index in the particular application.

A simple scheme of utilizing a FM discriminator for estimating the transmitter-receiver oscillator frequency offset has been presented. This scheme is found to require a very narrowband low pass filter in order to keep the variance of the frequency offset estimation small. The long filter delay of the low pass filter is not a problem because the frequency offset is relatively stationary with time.

Future work in the area of this thesis may include the following:

- Investigation into the use of recursive least square algorithms such as the fast transversal filter [15] can be undertaken. The convergence speeds of these algorithms have been demonstrated to be insensitive to input eigenvalue spread and they have been known to provide a lower average mean square error than stochastic gradient algorithms. Most of these algorithms, however, are susceptible to numerical instability. Algorithms which utilize the lattice joint-process structure should also be avoided.

- Investigation into the use of adaptive pilot/interpolation filter for a pilot symbol aided transmission system is recommended. Similar performance improvements are expected because of its functional similarity to tone aided systems.

APPENDIX 1 : Derivations of C_{max} and ρ_{max} for a Rectangular Pilot Filter

In this appendix, we will demonstrate that for a rectangular pilot filter, the bandwidth which gives maximum cross-correlation (C_{max}) also gives maximum correlation coefficient (ρ_{max}). At the same time, we will show that the cross-correlation function, C, has no local peaks other than C_{max}. To accomplish this, a combination of analytical and graphical techniques will be used.

Analysis of Cross-Correlation

We begin by first differentiating C with respect to the bandwidth, B_p, and then compute B_p for which $\frac{dC}{dB_p}$ equals 0.

The pilot filter is assumed to be unit energy so that its frequency response is given by:

$$H_p(f) = \begin{cases} \frac{1}{\sqrt{2B_p}} & |f| \le B_p \\ 0 & \text{otherwise} \end{cases} \tag{A1.1}$$

with B_p assumed to be positive.

The cross-correlation, C, can be shown to be [10]:

$$C = \sigma_{uw}^2 = \frac{E_b\sqrt{rR_b}}{1+r} \int_{-\infty}^{\infty} \tilde{S}_g(f-f_o)\, H_p^*(f)\, df \tag{A1.2}$$

For a rectangular pilot filter, this reduces to:

$$C = \frac{N_o}{\gamma}\sqrt{\frac{rR_b}{2B_p}} \int_{\max[-B_p,\, -f_D+f_o]}^{\min[B_p,\, f_D+f_o]} \tilde{S}_g(f-f_o)\, df$$

$$= \frac{N_o}{\gamma}\sqrt{\frac{rR_b}{2B_p}}\, P_d \qquad \text{(A1.3)}$$

where γ and P_d are given by:

$$\gamma = \frac{N_o}{E_b}(1+r) \qquad \text{(A1.4)}$$

and:

$$P_d = \int_{\max[-B_p,\, -f_D+f_o]}^{\min[B_p,\, f_D+f_o]} \tilde{S}_g(f-f_o)\, df$$

$$= \frac{1}{\pi}\left\{ \arcsin\left[\frac{\min[B_p,\, f_D+f_o] - f_o}{f_D}\right] - \arcsin\left[\frac{\max[-B_p,\, -f_D+f_o] - f_o}{f_D}\right]\right\} \qquad \text{(A1.5)}$$

Because of discontinuities in $S_g(f)$ and $H_p(f)$, we need to examine two cases: (i) when f_o is negative and (ii) when f_o is non-negative.

Case (i) : $f_o < 0$

We consider B_p in three non-overlapping regions:

(a) $B_p < f_o+f_D$

(b) $f_o+f_D \le B_p \le f_D-f_o$

(c) $B_p > f_D-f_o$

Region (a) : $B_p < f_o+f_D$

We differentiate C by first finding the derivative of P_d.

$$\frac{\delta P_d}{\delta B_p} = \frac{1}{2\pi}\left[\frac{1}{[f_D^2-(B_p-f_o)^2]^{1/2}} + \frac{1}{[f_D^2-(-B_p-f_o)^2]^{1/2}}\right] \tag{A1.6}$$

From (A1.6), we can see that $\frac{\delta P_d}{\delta B_p}$ is positive in region (a).

$$\frac{\delta^2 P_d}{\delta^2 B_p} = \frac{1}{2\pi}\left[\frac{B_p-f_o}{[f_D^2-(B_p-f_o)^2]^{3/2}} + \frac{-B_p-f_o}{[f_D^2-(-B_p-f_o)^2]^{3/2}}\right] \tag{A1.7}$$

In (A1.7), the denominator of the first term is less than that of the second term and the numerator of the first term is greater than the numerator of the second term. Therefore, the first term must be greater than the second term meaning that $\frac{\delta^2 P_d}{\delta^2 B_p}$ is also positive in region (a).

The derivative of C with respect to B_p is given by:

$$\begin{aligned}\frac{\delta C}{\delta B_p} &= \frac{N_o}{\gamma}\sqrt{\frac{rR_b}{2}}\left[\frac{\delta P_d}{\delta B_p}\frac{1}{\sqrt{B_p}} - \frac{P_d}{2(B_p)^{3/2}}\right] \\ &= \frac{N_o}{\gamma}\sqrt{\frac{rR_b}{2}}\frac{1}{(B_p)^{3/2}}\left[B_p\frac{\delta P_d}{\delta B_p} - \frac{P_d}{2}\right]\end{aligned} \tag{A1.8}$$

Let $f(B_p) = \left[B_p\frac{\delta P_d}{\delta B_p} - \frac{P_d}{2}\right]$, then

$$\frac{\delta f(B_p)}{\delta B_p} = B_p\frac{\delta^2 P_d}{\delta^2 B_p} + \frac{\delta P_d}{\delta B_p} - \frac{1}{2}\frac{\delta P_d}{\delta B_p} = B_p\frac{\delta^2 P_d}{\delta^2 B_p} + \frac{1}{2}\frac{\delta P_d}{\delta B_p} \tag{A1.9}$$

which is positive because, as shown earlier, both the first and second derivatives of P_d are positive. Consequently, $f(B_p)$ is monotonically increasing in region (a). Since $f(0) = 0$, $f(B_p)$ must necessarily be positive. This implies that $\frac{\delta C}{\delta B_p}$ is positive and that $C(B_p)$ is monotonically increasing in the region $B_p < f_o+f_D$.

Region (b) : $f_o+f_D \le B_p \le f_D-f_o$

In this region,

$$P_d = \frac{1}{\pi}\left\{\frac{\pi}{2} - \arcsin\left[\frac{-B_p-f_o}{f_D}\right]\right\} \tag{A1.10}$$

Its first and second derivatives with respect to B_p are:

$$\frac{\delta P_d}{\delta B_p} = \frac{1}{2\pi}\left[\frac{1}{[f_D^2-(-B_p-f_o)^2]^{1/2}}\right] \tag{A1.11}$$

and

$$\frac{\delta^2 P_d}{\delta^2 B_p} = \frac{1}{2\pi}\left[\frac{-B_p-f_o}{[f_D^2-(-B_p-f_o)^2]^{3/2}}\right] \tag{A1.12}$$

Clearly, one cannot determine from (A1.8), (A1.10) and (A1.11) whether any local extremum exists. Instead, we have made use of graphically techniques in order to demonstrate that no local extremum exists in this region. Figures A1.1, A1.2, A1.3 and A1.4 show the cross-correlation as a function of the filter bandwidth for various doppler frequencies at 20dB E_b/N_o with 0, -10, -50 and -100Hz offset respectively. Based on these figures, we can observe that $C(B_p)$ is monotonically increasing and that no local extremum exists for $f_o+f_D \le B_p \le f_D-f_o$.

Region (c) : $B_p < f_D-f_o$

In this region,

$$P_d = 1.$$

The derivative of C with respect to B_p, with P_d equal unity, is:

$$\frac{\delta C}{\delta B_p} = -\frac{N_o}{\gamma}\sqrt{\frac{rR_b}{2}}\frac{1}{2(B_p)^{3/2}} \qquad \text{(A1.13)}$$

which is negative. $C(B_p)$ is, thus, monotonically decreasing in the region $B_p < f_D - f_o$. Since $C(B_p)$ is monotonically increasing in regions (a) and (b) but is monotonically decreasing in region (c), this implies that there is a unique global maximum which must occur at the boundary between regions (b) and (c); i.e.

$$B_{p,max} = f_D - f_o \qquad \text{for} \quad f_o < 0 \qquad \text{(A1.14)}$$

Figures A1.1, A1.2, A1.3, and A1.4 confirm this conclusion.

Case (ii) : $f_o \geq 0$

We again divide the analysis into three non-overlapping regions:

(a) $B_p < f_D - f_o$

(b) $f_D - f_o \leq B_p \leq f_o + f_D$

(c) $B_p > f_o + f_D$

By symmetry, the results from the corresponding regions for case (i) also apply for case (ii), i.e. $C(B_p)$ is monotonically increasing in regions (a) and (b) but it is monotonically decreasing in region (c). We can, therefore, draw a similar conclusion as for case (i), i.e. there is a unique global maximum for C as a function of B_p which occurs at:

$$B_{p,max} = f_D + f_o \qquad \text{for} \quad f_o \geq 0 \qquad \text{(A1.15)}$$

We can conlude from the above analysis that, for a rectangular pilot filter, the cross-correlation function is "well behaved" and it attains maximum value for

$$B_{p,max} = f_D + |f_o|. \qquad \text{(A1.16)}$$

Analysis of Correlation Coefficient

Following a similar approach as used for the analysis of the cross-correlation, we differentiate the correlation coefficient ρ with respect to the filter bandwidth B_p by first reducing the expression for ρ into a simpler form.

From (3.2), the expression for the correlation coefficient, ρ, is given by:

$$\rho = \frac{\int_{-\infty}^{\infty} \tilde{S}_g(f-f_o)\, H_p^*(f)\, df}{\left[(1+\gamma)\left\{\int_{-\infty}^{\infty} \tilde{S}_g(f-f_o)\, |H_p(f)|^2\, df + \frac{\gamma B_n}{rR_b}\right\}\right]^{\frac{1}{2}}} \tag{A1.17}$$

For a rectangular pilot filter, this reduces to:

$$\rho = \frac{\int_{\max[-B_p,\, f_D+f_o]}^{\min[B_p,\, f_D+f_o]} \tilde{S}_g(f-f_o)\, df}{\left[(1+\gamma)\left\{\int_{\max[-B_p,\, f_D+f_o]}^{\min[B_p,\, f_D+f_o]} \tilde{S}_g(f-f_o)\, df + \frac{\gamma 2B_p}{rR_b}\right\}\right]^{\frac{1}{2}}}$$

$$= \frac{P_d}{\left[(1+\gamma)\left\{P_d + \frac{\gamma 2B_p}{rR_b}\right\}\right]^{\frac{1}{2}}} \tag{A1.18}$$

Again, we examine the cases where f_o is negative and where f_o is non-negative, separately.

Case (i) : $f_o < 0$

We consider B_p in three non-overlapping regions:

(a) $B_p < f_o+f_D$

(b) $f_o+f_D \leq B_p \leq f_D-f_o$

(c) $B_p > f_D-f_o$

Region (a) : $B_p < f_o+f_D$

The derivative of ρ with respect to B_p is:

$$\begin{aligned}\frac{\delta\rho}{\delta B_p} &= \frac{1}{\sqrt{1+\gamma}}\left[\frac{\delta P_d}{\delta B_p}\left(P_d+\frac{\gamma 2B_p}{rR_b}\right)^{-\frac{1}{2}} - \frac{1}{2}\left(P_d+\frac{\gamma 2B_p}{rR_b}\right)^{-\frac{3}{2}}\left(\frac{\delta P_d}{\delta B_p}+\frac{2\gamma}{rR_b}\right)P_d\right] \\ &= \frac{1}{\sqrt{1+\gamma}}\,\frac{\frac{1}{2}P_d\frac{\delta P_d}{\delta B_p}+\frac{\gamma 2B_p}{rR_b}\frac{\delta P_d}{\delta B_p} - P_d\frac{\gamma}{rR_b}}{\left(P_d+\frac{2\gamma}{rR_b}\right)^{\frac{3}{2}}}\end{aligned} \tag{A1.19}$$

Denoting the numerator in (A1.19) by $g(B_p)$, then the derivative of $g(B_p)$ is given by:

$$\frac{\delta g}{\delta B_p} = \frac{1}{2}\left(\frac{\delta P_d}{\delta B_p}\right)^2 + \frac{1}{2}P_d\frac{\delta^2 P_d}{\delta^2 B_p} + \frac{\gamma}{rR_b}\frac{\delta P_d}{\delta B_p} + \frac{2\gamma}{rR_b}B_p\frac{\delta^2 P_d}{\delta^2 B_p} \tag{A1.20}$$

Since both the first and second derivatives of P_d with respect to B_p are positive, it follows that $\frac{\delta g}{\delta B_p}$ is also positive in region (a). With $g(0) = 0$, $g(B_p)$ must be postive as well meaning that $\rho(B_p)$ is monotonically increasing in the region $B_p < f_o+f_D$.

Region (b) : $f_o+f_D \leq B_p \leq f_D-f_o$

As with cross-correlation, graphical techniques have been used to determine whether ρ is monotonically increasing in this region. Figures A1.5, A1.6, A1.7 and A1.8 show the bit error rate as a function of the filter bandwidth for various doppler frequencies at 20dB E_b/N_o with 0, -10, -50 and -100Hz offset respectively. From (3.3), ρ is related to the bit error rate for a rectangular pilot filter simply as:

$$\rho = 1 - 2\,P_e \tag{A1.21}$$

BER has been used here for convenience only. All of the figures show that BER is monotonically decreasing for $f_o+f_D \leq B_p \leq f_D-f_o$ which means that ρ is monotonically increasing in this region.

Region (c) : $B_p < f_D-f_o$

In this region,

$$P_d = 1$$

and

$$\frac{\delta\rho}{\delta B_p} = -(1+\gamma)\,\frac{\gamma}{rR_b}\left[(1+\gamma)\left(1+\frac{\gamma 2B_p}{rR_b}\right)\right]^{\frac{3}{2}} < 0 \tag{A1.22}$$

ρ is, therefore, monotonically decreasing for $B_p < f_D-f_o$. With the results found for regions (a) and (b), we conclude that ρ has a unique global maximum at:

$$B_{p,max}' = f_D-f_o \qquad \text{for } f_o < 0 \tag{A1.23}$$

Figures A1.5, A1.6, A1.7 and A1.8 confirm this result.

Case (ii) : $f_o \geq 0$

Again, by symmetry, we can conclude that there is a unique global maximum for ρ which occurs at:

$$B_{p,max}' = f_D + f_o \qquad \text{for } f_o \geq 0 \tag{A1.24}$$

Combining with (A1.23), we can express the optimum bandwidth at which maximum correlation coefficient occurs as:

$$B_{p,max}' = f_D + |f_o|. \tag{A1.25}$$

This equation is of course identical to (A1.16) which gives the optimal bandwidth for peak cross-correlation. We have therefore shown that for a rectangular pilot filter, both the cross-correlation and correlation coefficient attain their maxima for the same filter bandwidth as given by (A1.16).

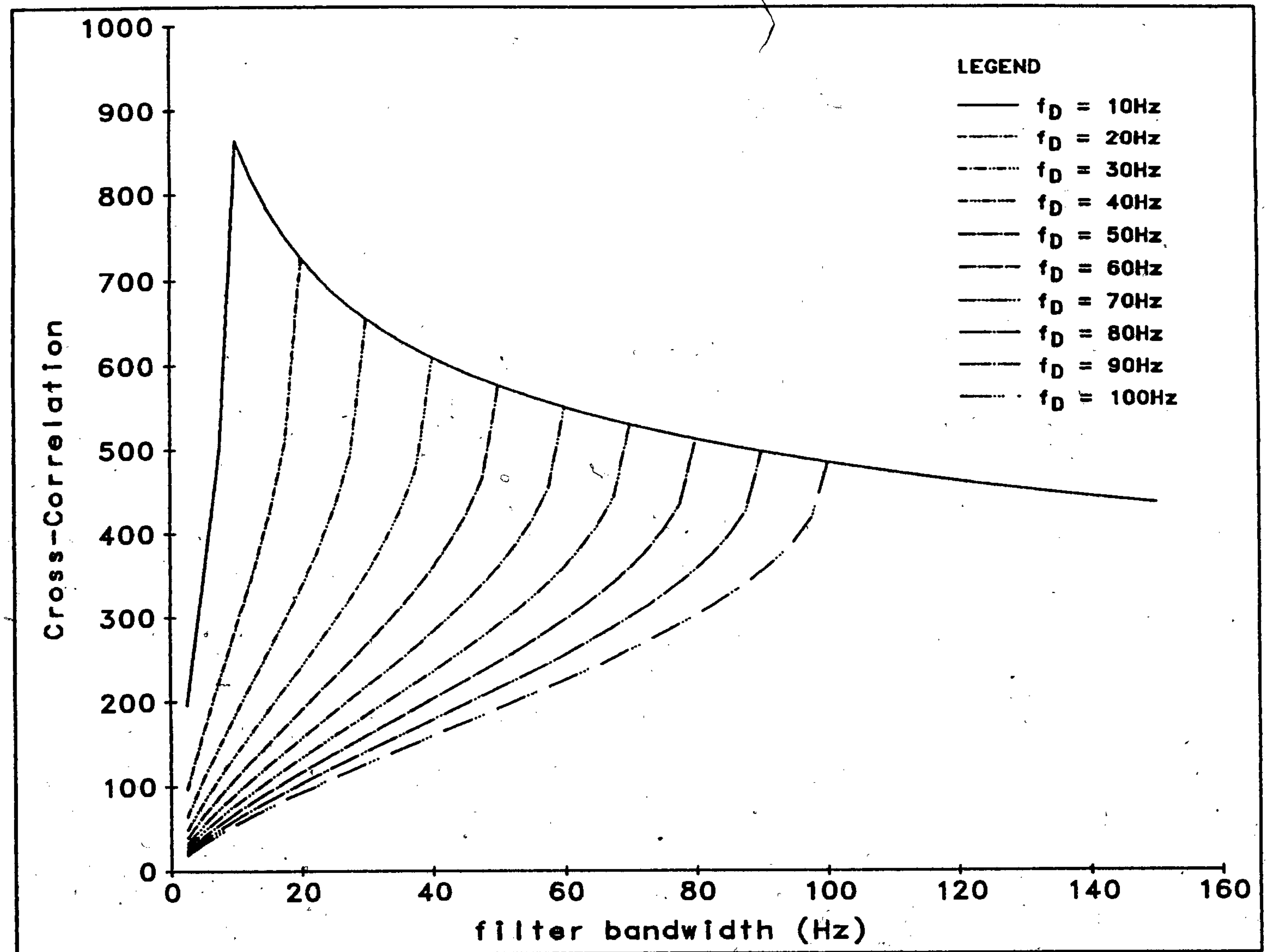


Figure A1.1 - Cross-Correlation as a Function of Filter Bandwidth with no Frequency Offset

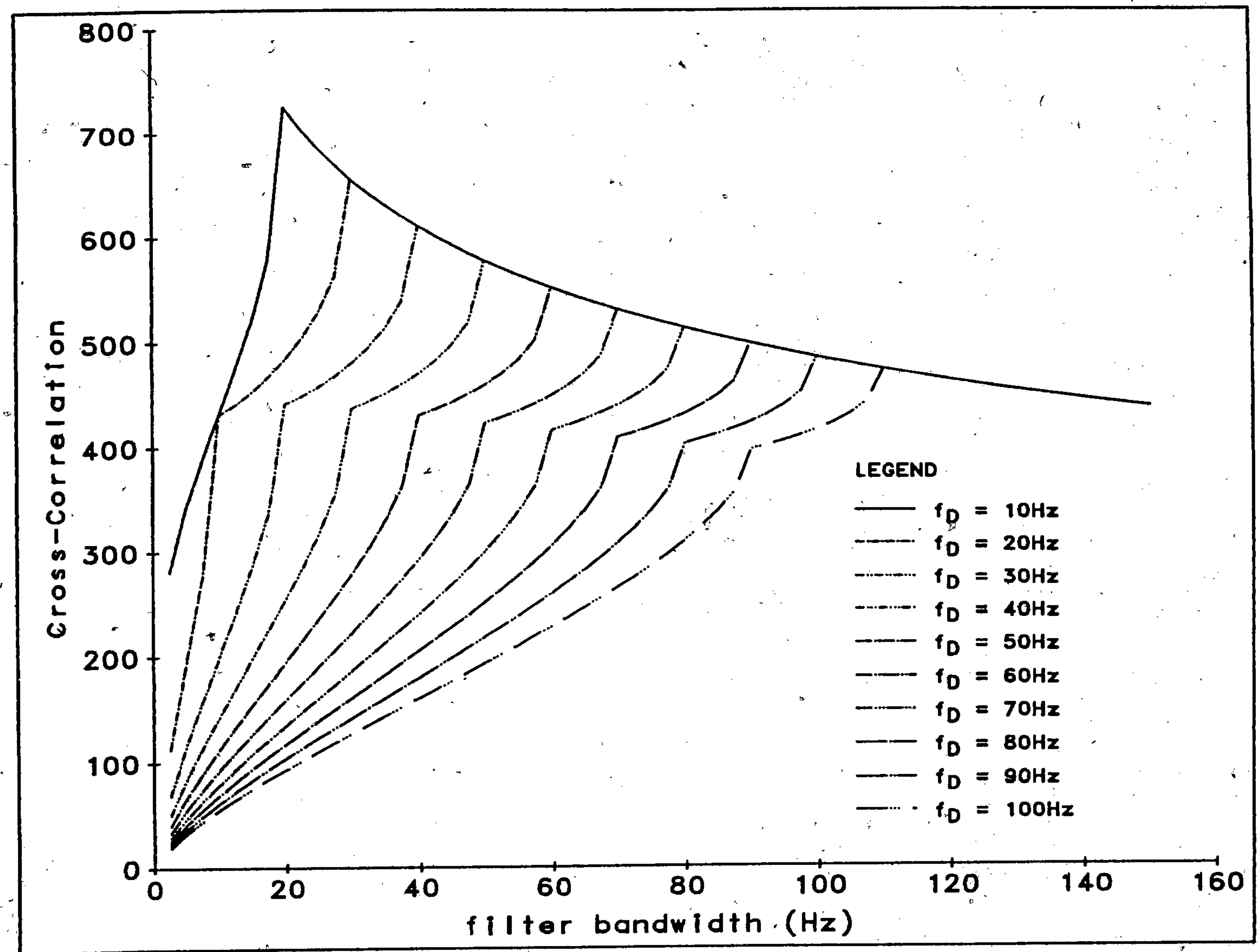


Figure A1.2 - Cross-Correlation as a Function of Filter Bandwidth Under 10Hz Frequency Offset

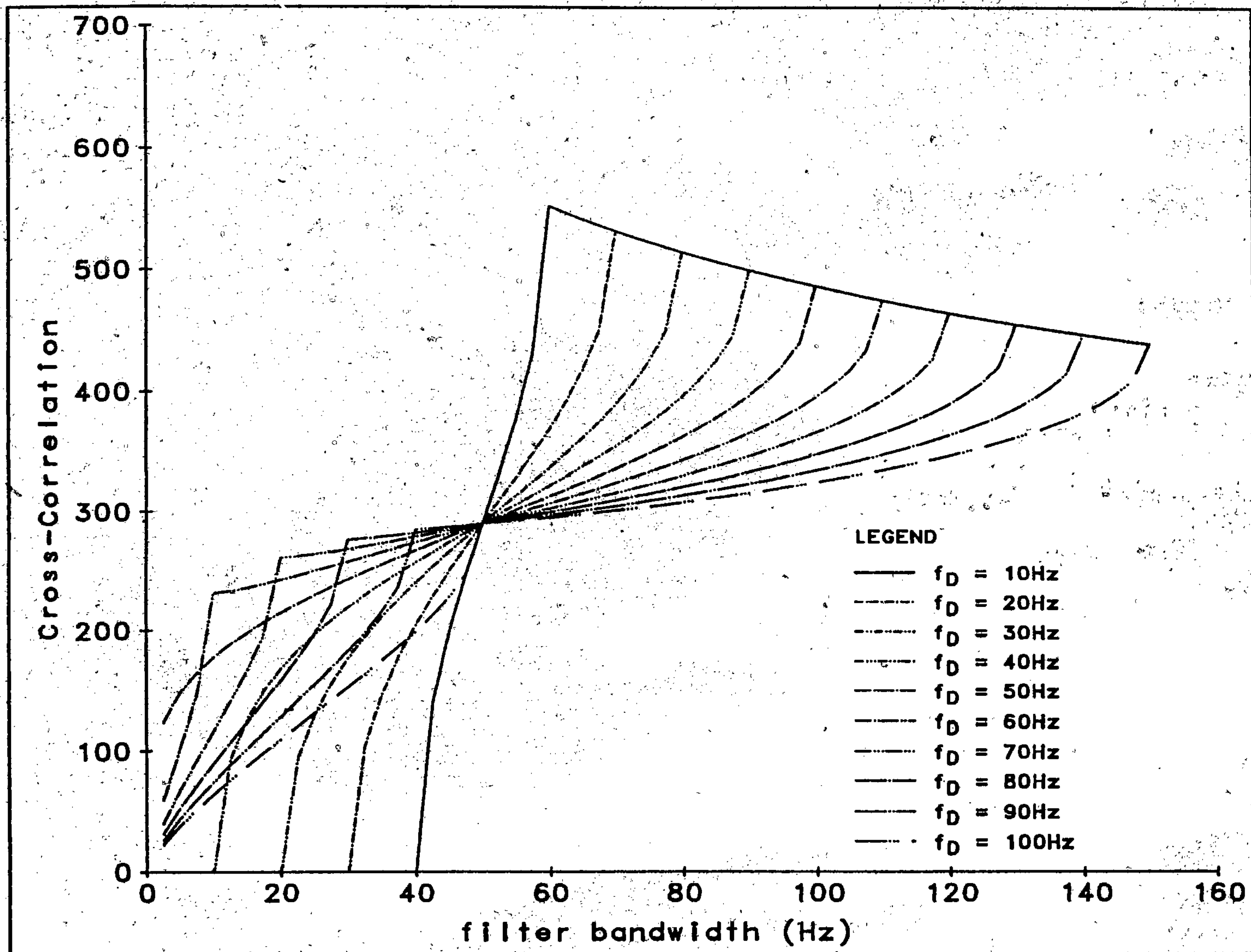


Figure A1.3 - Cross-Correlation as a Function of Filter Bandwidth Under 50Hz Frequency Offset

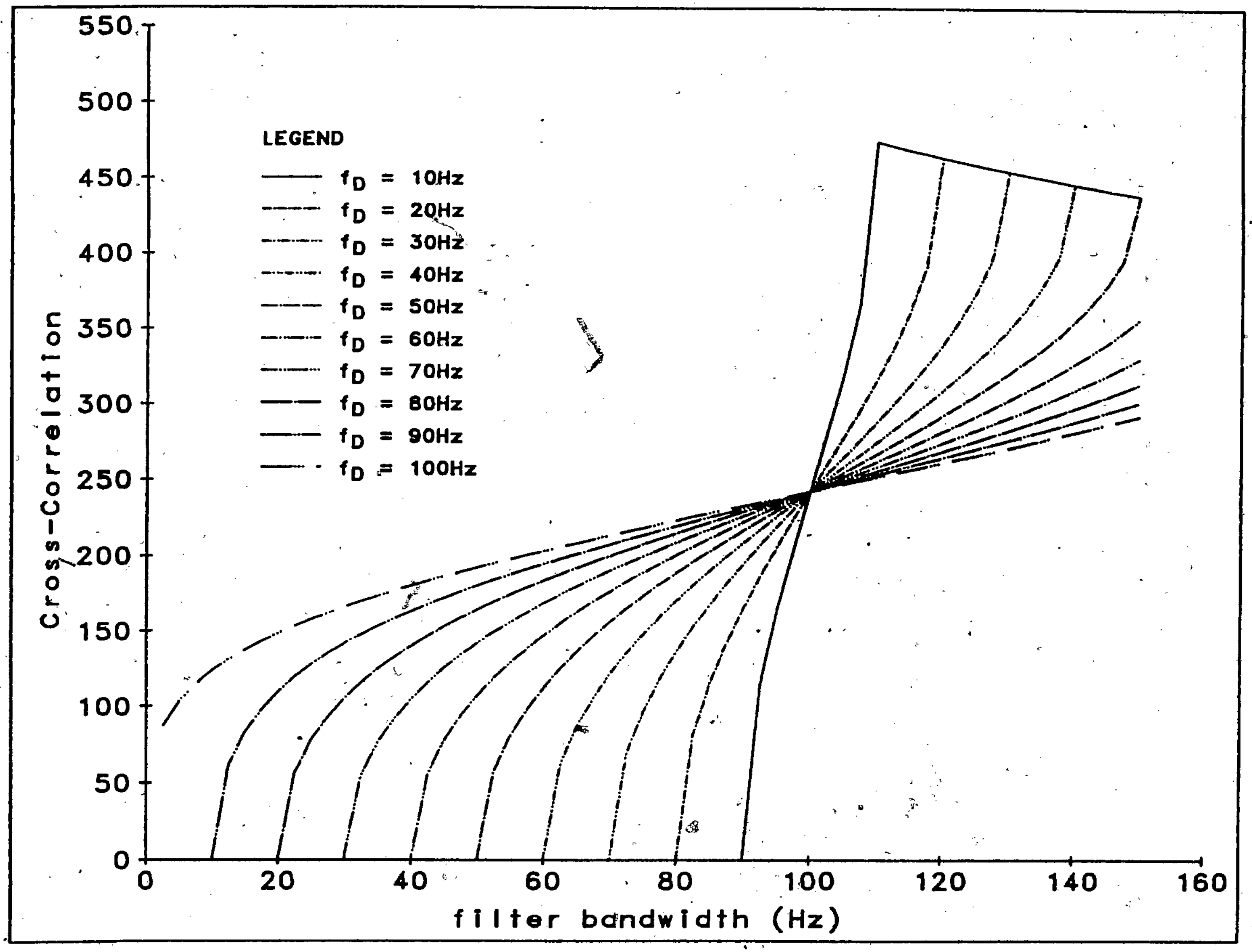


Figure A1.4 - Cross-Correlation as a Function of Filter Bandwidth Under 100Hz Frequency Offset

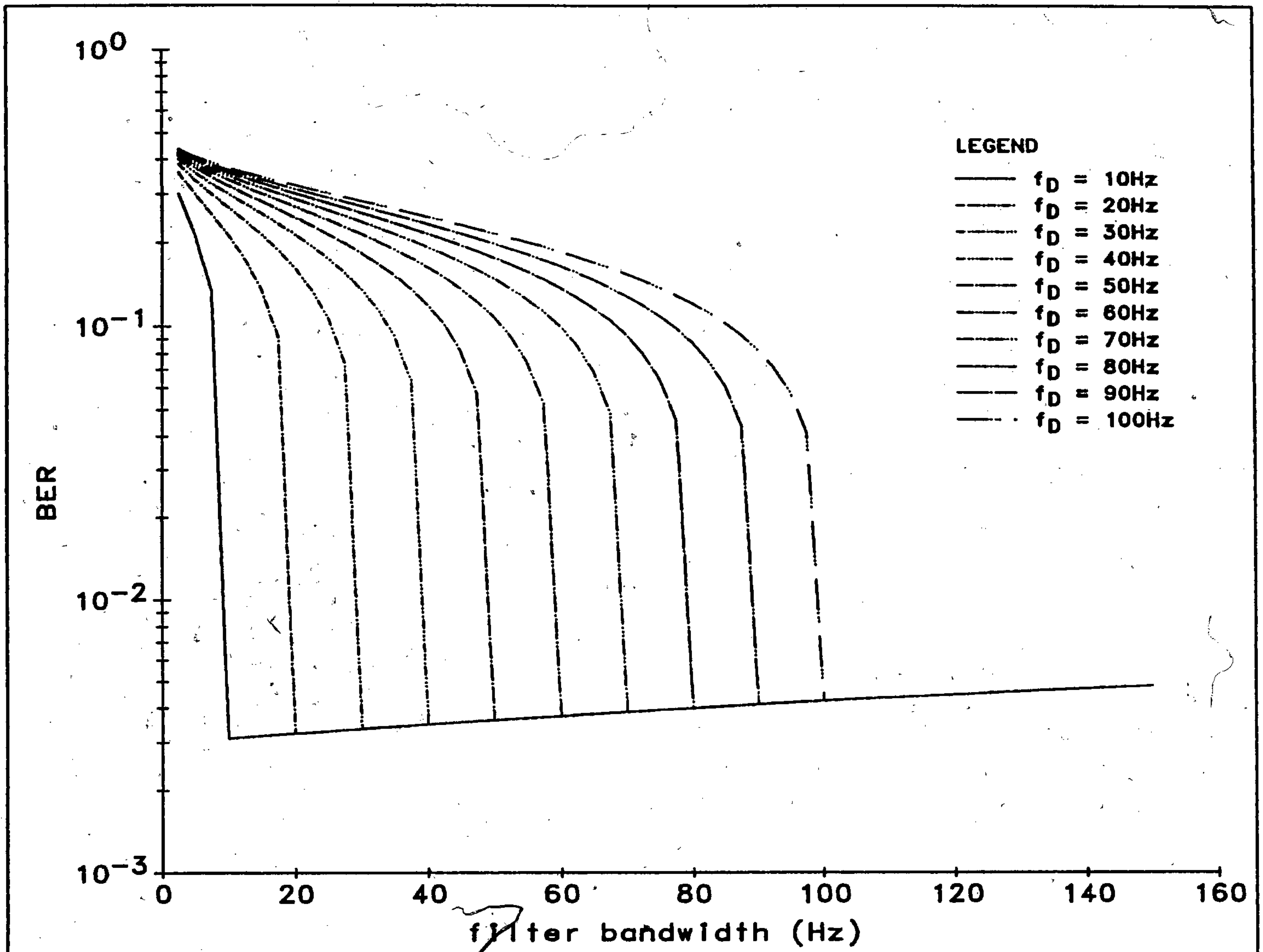


Figure A1.5 - BER as a Function of Filter Bandwidth with no Frequency Offset

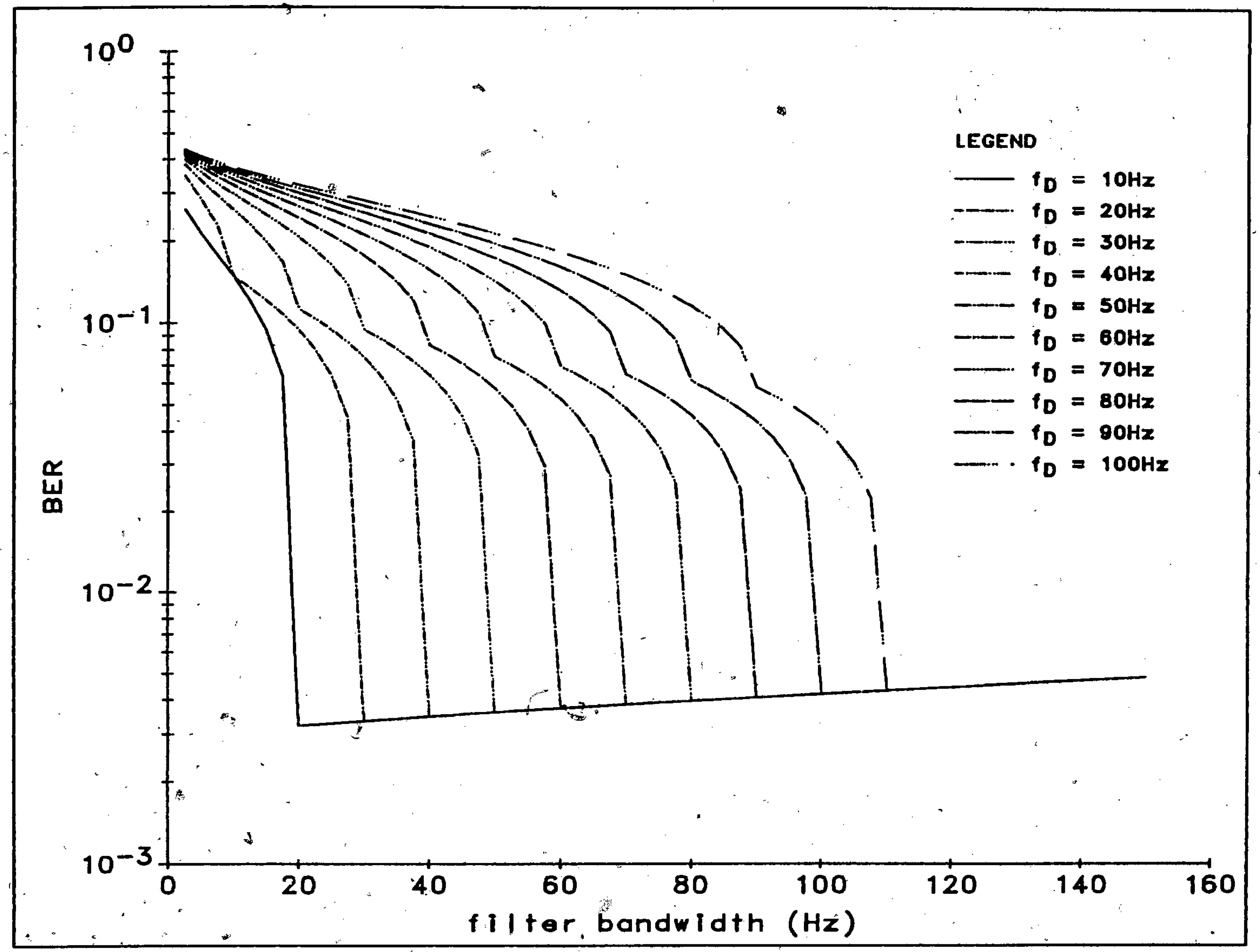


Figure A1.6 - BER as a Function of Filter Bandwidth Under 10Hz Frequency Offset

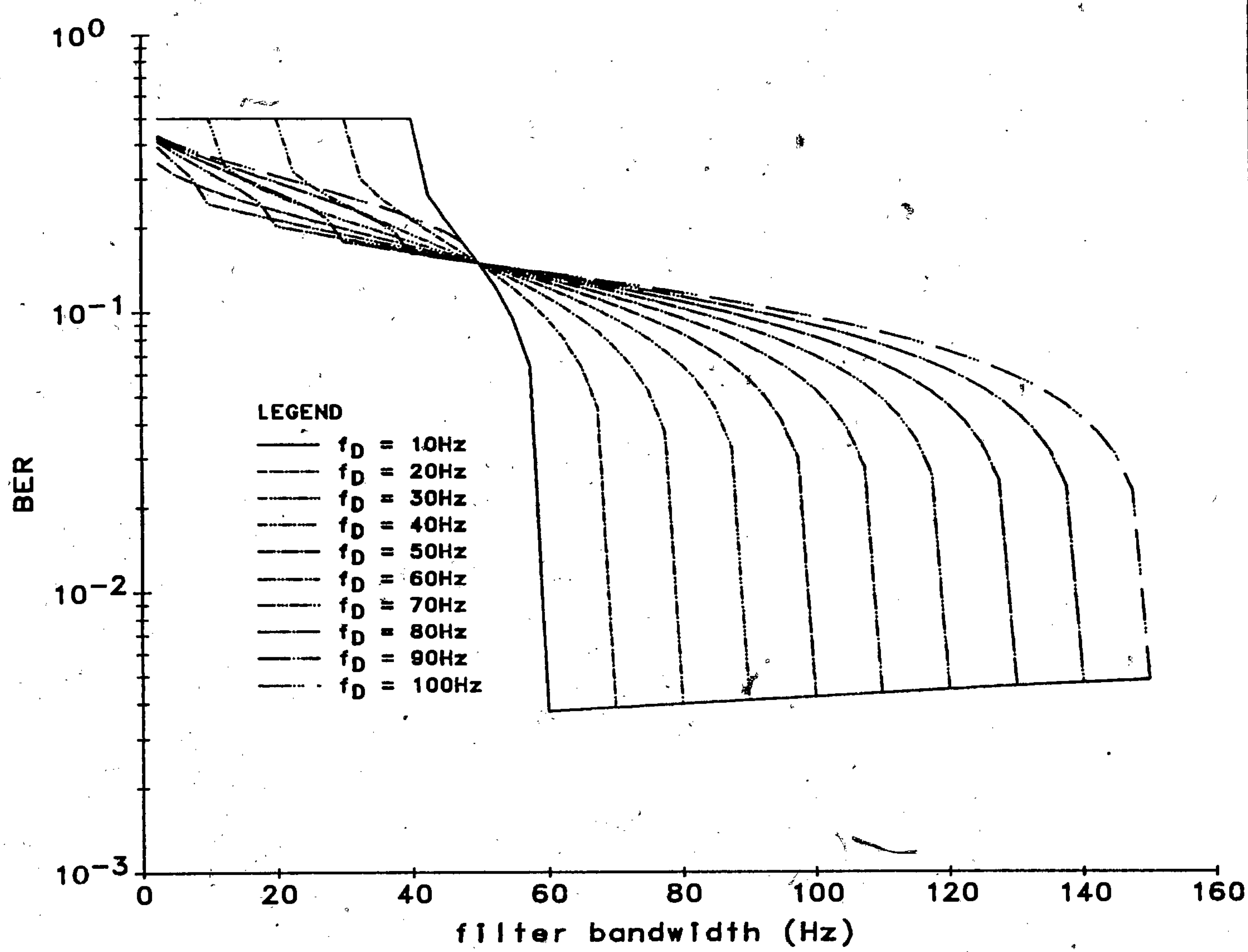


Figure A1.7 - BER as a Function of Filter Bandwidth Under 50Hz Frequency Offset

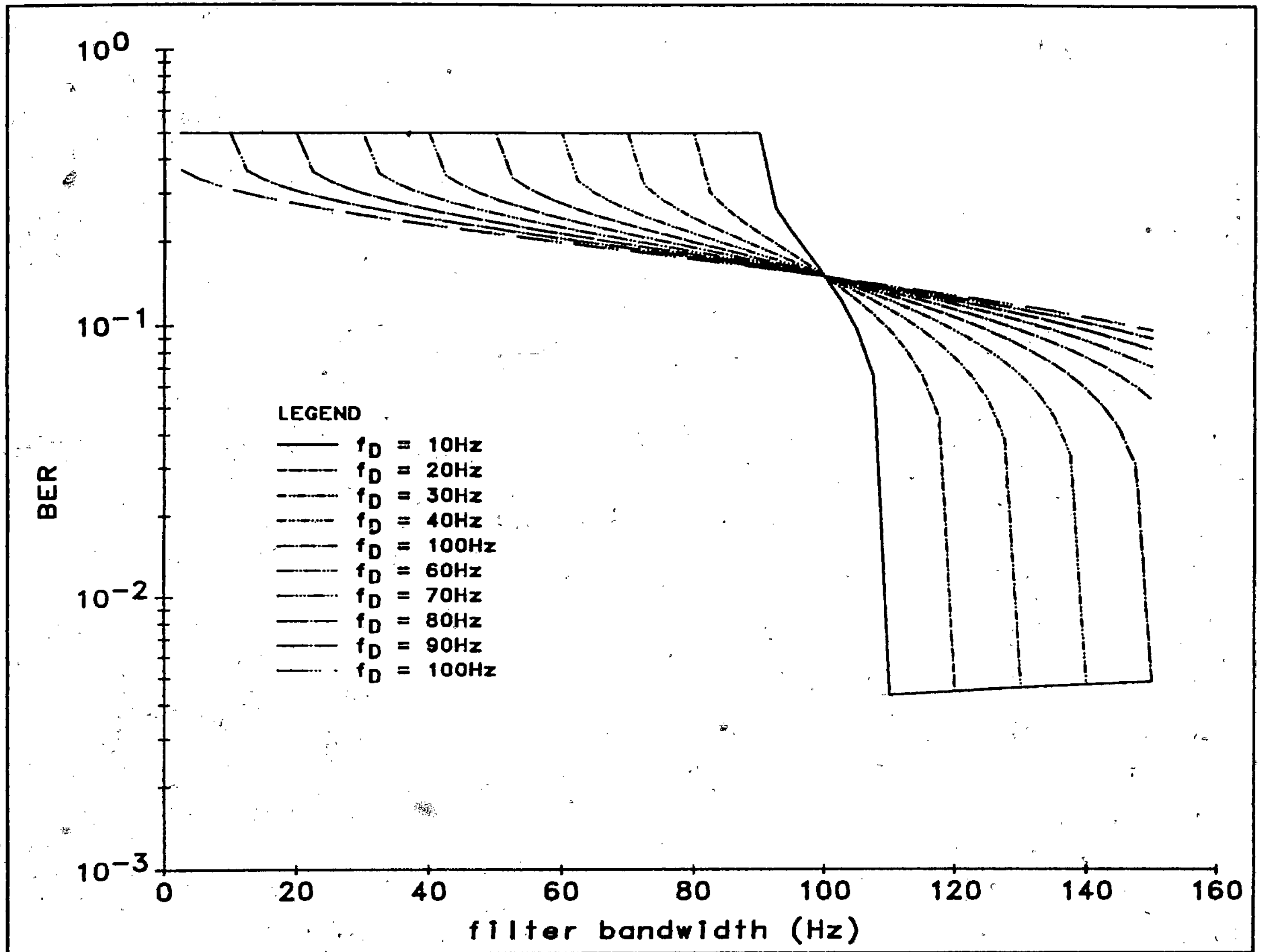


Figure A1.8 - BER as a Function of Filter Bandwidth Under 100Hz Frequency Offset

APPENDIX 2 : Behavior of the Cross-Correlation for a Set of Optimum FIR Filters

It has been shown in Appendix 1 that the cross-correlation is "well behaved" for a set of rectangular pilot filters. Here, we examine, graphically, the behavior of the cross-correlation when a set of optimum FIR filters is used. The set of optimum FIR filters refers to filters which have been optimized for different doppler at fixed E_b/N_0 and tone-to-signal power ratio.

Figures A2.1 and A2.2 show the cross-correlation and BER as functions of the filter number (or state index) at various doppler for E_b/N_0 of 20dB. Filter length of 51 and MWA length of 3 have been used. The optimum filters have been designed for increasing f_D at 5Hz increment starting at 5Hz and ending at 125Hz. The filters are numbered in order of increasing f_D so that filter number 1 corresponds to a filter optimized for 5Hz doppler and filter number 2 corresponds to a filter optimized for 10Hz doppler, etc.

We can observe from figure A2.1 that the cross-correlation function contains a number of "plateaus", especially at high doppler. These plateaus are due to the ripples and gradual roll-off of the filter frequency responses. The peaks of the cross-correlation do not occur at the expected filter number as for a set of rectangular filters. For example, using a set rectangular filters at f_D of 50Hz with a similar bandwidth arrangement (5Hz increment), we expect the cross-correlation to peak at filter number 10. For the set of optimum FIR filters used, the cross-correlation peaked at filter number 8. The difference is again due to the gradual roll-off of the filter frequency response.

Similar phenomena can be observed for the BER as a function of the filter number in figure A2.2. Careful examination of figures A2.1 and A2.2 will reveal that the filter numbers which give minimum BER (and hence maximum correlation coefficient for the set of filters used) do not correspond to those which give maximum cross-correlation. Figure A2.3 shows the locations of the respective peaks of the cross-correlation and correlation coefficient at different doppler frequencies. The locations of the peaks differ by only one state except for states 1 and 2. These differences can be easily compensated for, by using appropriate numbers of step margin.

It should be noted from figure A2.1, as well as in A2.2, that there is a small local maximum for f_D of 70Hz. Fortunately, because of the stochastic nature of the fading process, there is no danger of the algorithm being "trapped" at the particular state where the local maximum is situated. In general, the cross-correlation for a set of optimum FIR filters is not as "well behaved" as for a set of rectangular filters. However, as shown by the results given in section 7.3.2, the BER performance of the FSA using an ensemble of optimum FIR filters is comparable to that using a set of rectangular filters.

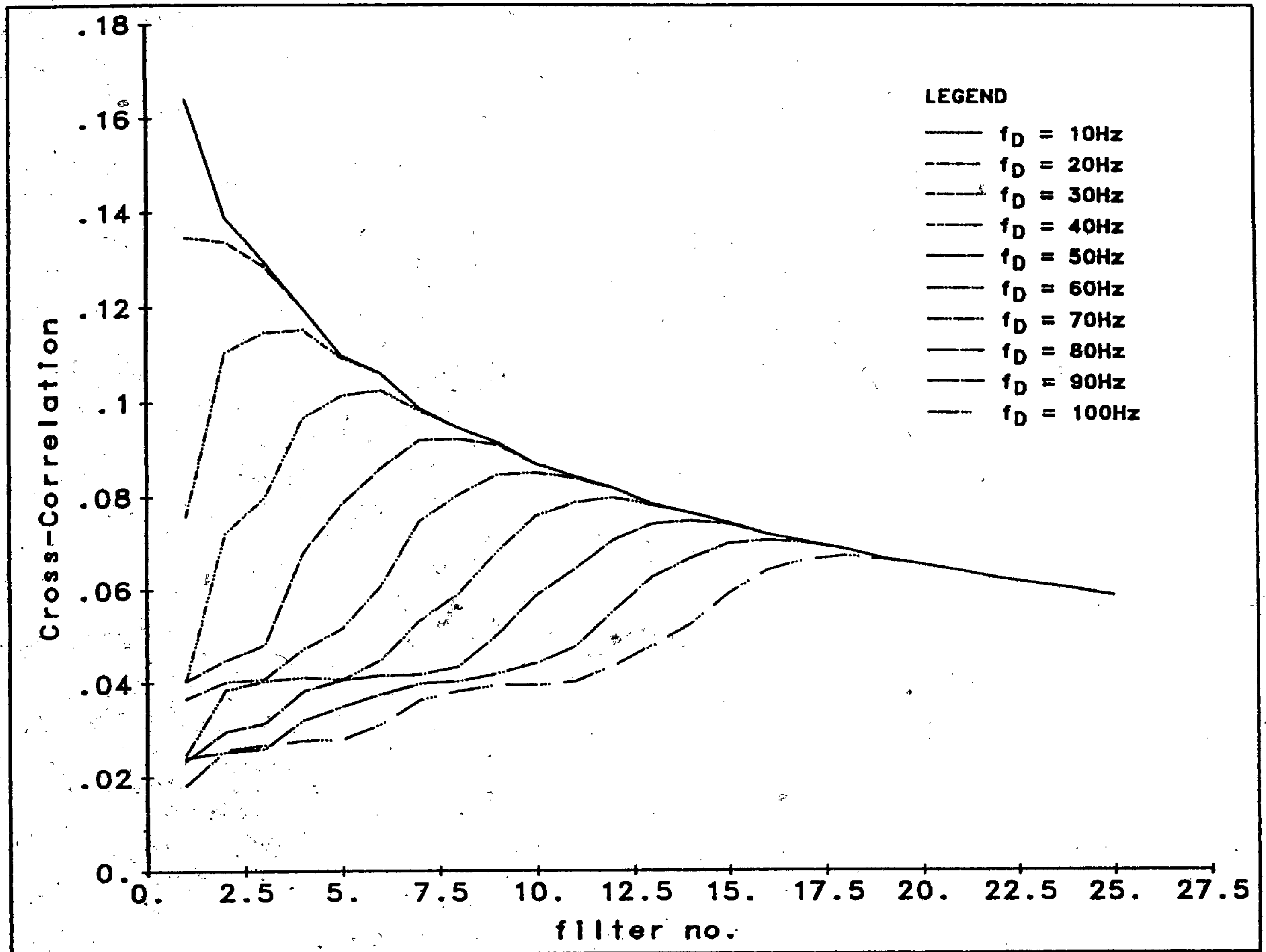


Figure A2.1 - Cross-Correlation as a Function of Filter Number in a Set of Optimum FIR Filter

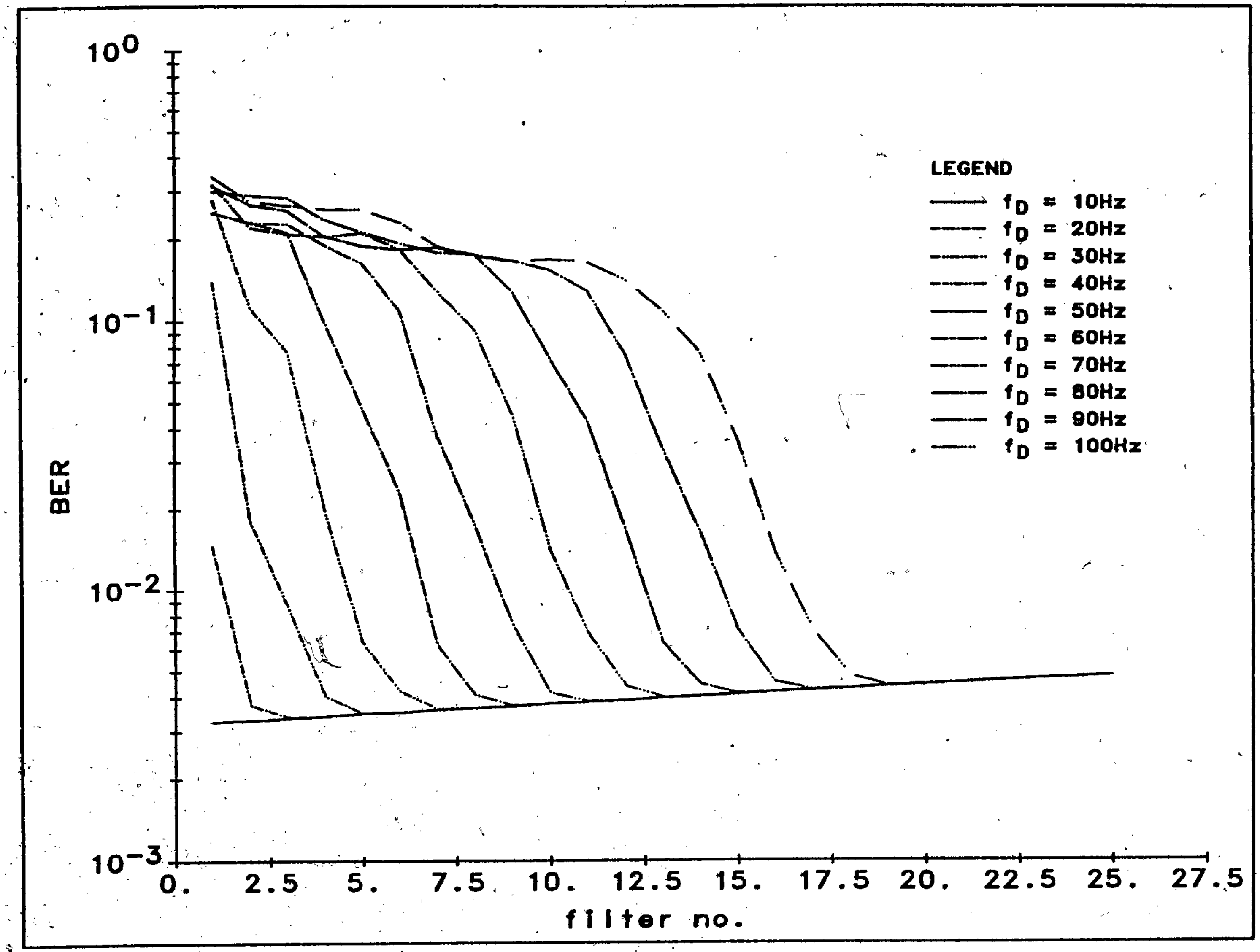


Figure A2.2 - BER as a Function of Filter Number in a Set of Optimum FIR Filter

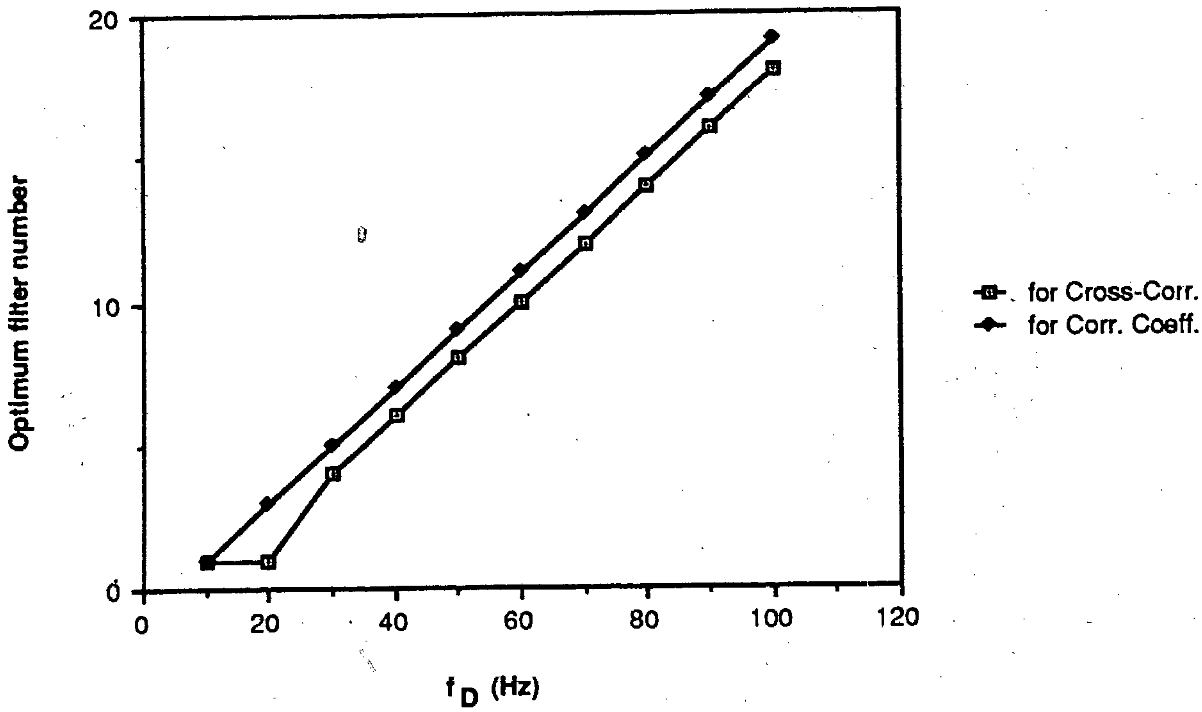


Figure A2.3 - Optimal Filter Number at Different Doppler Frequencies

APPENDIX 3 : Derivations of the Correlation Functions of $\hat{u}(k)$ and $v(k)$

This appendix presents detail derivations of the correlation functions between the in-phase and quadrature components of $\hat{u}(k)$ and $v(k)$. These correlation functions are:

1) $R_{\hat{u}_I\hat{u}_I}(i-k) = E[\hat{u}_I(i)\, \hat{u}_I(k)]$ (A3.1)

2) $R_{\hat{u}_I\hat{u}_Q}(i-k) = E[\hat{u}_I(i)\, \hat{u}_Q(k)]$ (A3.2)

3) $R_{\hat{u}_Q\hat{u}_Q}(i-k) = E[\hat{u}_Q(i)\, \hat{u}_Q(k)]$ (A3.3)

4) $R_{v_Iv_I}(i-k) = E[v_I(i)\, v_I(k)]$ (A3.4)

5) $R_{v_Iv_Q}(i-k) = E[v_I(i)\, v_Q(k)]$ (A3.5)

6) $R_{v_Qv_Q}(i-k) = E[v_Q(i)\, v_Q(k)]$ (A3.6)

7) $R_{\hat{u}_Iv_I}(i-k) = E[\hat{u}_I(i)\, v_I(k)]$ (A3.7)

8) $R_{\hat{u}_Iv_Q}(i-k) = E[\hat{u}_I(i)\, v_Q(k)]$ (A3.8)

9) $R_{\hat{u}_Qv_I}(i-k) = E[\hat{u}_Q(i)\, v_I(k)]$ (A3.9)

10) $R_{\hat{u}_Qv_Q}(i-k) = E[\hat{u}_Q(i)\, v_Q(k)]$ (A3.10)

with $\hat{u}_I$, $\hat{u}_Q$,v_I , v_Q denoting the in-phase and quadrature components of $\hat{u}$ and v.

The given derivations assume that the in-phase and quadrature components of the fading process, g, and the noise process in the decision corrected reference, n_u, are uncorrelated. Furthermore, the auto-correlation functions of their in-phase and quadrature components are assumed to be the same. These assumptions lead to the following:

$$E[g_I(i)g_Q(k)] = E[n_{uI}(i)n_{uQ}(k)] = 0 \quad (A3.11)$$

$$E[g_I(i)g_I(k)] = E[g_Q(i)g_Q(k)] = R_g(i-k) \quad (A3.12)$$

$$E[n_{uI}(i)n_{uI}(k)] = E[n_{uQ}(i)n_{uQ}(k)] \tag{A3.13}$$

The expressions for $\hat{u}(i)$ and $v(i)$ can be expanded to give their in-phase and quadrature components as given below:

$$\begin{aligned}
\hat{u}(i) &= A\, g(i) \exp(2\pi f_o i) + n_u(i) \\
&= A \{g_I(i)+jg_Q(i)\} \{\cos(2\pi f_o i)+j\sin(2\pi f_o i)\} + \{n_{vI}(i)+jn_{vQ}(i)\} \\
&= A \{g_I(i) \cos(2\pi f_o i) - g_Q(i) \sin(2\pi f_o i)\} + jA \{g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i)\} \\
&\quad + \{n_{vI}(i)+jn_{vQ}(i)\}
\end{aligned} \tag{A3.14}$$

$$\hat{u}_I(i) = A \{g_I(i) \cos(2\pi f_o i) - g_Q(i) \sin(2\pi f_o i)\} + n_{vI}(i) \tag{A3.15}$$

$$\hat{u}_Q(i) = A \{g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i)\} + n_{vQ}(i) \tag{A3.16}$$

$$\begin{aligned}
v(i) &= a\,[c(i) * \Delta h_p(i)] + n_v(i) \\
&= a \sum_{n=-\infty}^{\infty} \{g(i-n) \exp(j2\pi f_o(i-n))\, \Delta h_p(n)\} + n_v(i) \\
&= a \sum_{n=-\infty}^{\infty} \{(g_I(i-n)+jg_Q(i-n))\, (\cos(2\pi f_o(i-n))+j\sin(2\pi f_o(i-n)))\, (\Delta h_{pI}(n)+j\Delta h_{pQ}(n))\} \\
&\quad + (n_{vI}(i)+jn_{vQ}(i)) \\
&= a \sum_{n=-\infty}^{\infty} \{g_I(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pI}(n) - g_Q(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pQ}(n) \\
&\qquad - g_I(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pQ}(n) - g_Q(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pI}(n)\} \\
&\quad + j\, \{g_I(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pI}(n) - g_Q(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pQ}(n) \\
&\qquad + g_I(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pQ}(n) + g_Q(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pI}(n)\} \\
&\quad + (n_{vI}(i)+jn_{vQ}(i))
\end{aligned} \tag{A3.17}$$

$$v_I(i) = \sum_{n=-\infty}^{\infty} \{g_I(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pI}(n) - g_Q(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pQ}(n)$$

$$- g_I(i-n)\sin(2\pi f_o(i-n))\,\Delta h_{pQ}(n) - g_Q(i-n)\sin(2\pi f_o(i-n))\,\Delta h_{pI}(n)\Big\} + n_{vI}(i) \qquad \text{(A3.18)}$$

$$v_Q(i) = \sum_{n=-\infty}^{\infty} \Big\{ g_I(i-n)\sin(2\pi f_o(i-n))\,\Delta h_{pI}(n) - g_Q(i-n)\sin(2\pi f_o(i-n))\,\Delta h_{pQ}(n)$$
$$+ g_I(i-n)\cos(2\pi f_o(i-n))\,\Delta h_{pQ}(n) + g_Q(i-n)\cos(2\pi f_o(i-n))\,\Delta h_{pI}(n)\Big\} + n_{vQ}(i) \qquad \text{(A3.19)}$$

The variances of the in-phase and quadrature components of the gradient filter output noise process, $n_{vI}(i)$ and $n_{vQ}(i)$, have been derived separately in Appendix 4.

1) The auto-correlation function, $R_{\hat{u}_I\hat{u}_I}(i-k)$, is given by

$$\begin{aligned}
R_{\hat{u}_I\hat{u}_I}(i-k) &= E[\hat{u}_I(i)\,\hat{u}_I(k)] \\
&= E\Big[\Big\{A\,\{g_I(i)\cos(2\pi f_o i) - g_Q(i)\sin(2\pi f_o i)\} + n_{uI}(i)\Big\} \\
&\quad \times \Big\{A\,\{g_I(k)\cos(2\pi f_o k) - g_Q(k)\sin(2\pi f_o k)\} + n_{uI}(k)\Big\}\Big] \\
&= E\Big[A^2\,\{g_I(i)\cos(2\pi f_o i)\,g_I(k)\cos(2\pi f_o k)\} \\
&\quad + A^2\,\{g_Q(i)\sin(2\pi f_o i)\,g_Q(k)\sin(2\pi f_o k)\} + n_{uI}(i)\,n_{uI}(k)\Big] \\
&= E[g_I(i)\,g_I(k)]\,A^2\cos(2\pi f_o(i-k)) + E[n_{uI}(i)\,n_{uI}(k)] \\
&= A^2 R_g(i-k)\cos(2\pi f_o(i-k)) + \sigma^2_{n_{uI}}\,\delta_{ik} \\
&= A^2 \sigma^2_g\,\tilde{R}_g(i-k)\cos(2\pi f_o(i-k)) + \sigma^2_{n_{uI}}\,\delta_{ik} \\
&= \frac{E_b}{1+r}\,\tilde{R}_g(i-k)\cos(2\pi f_o(i-k)) + \sigma^2_{n_{uI}}\,\delta_{ik} \qquad \text{(A3.20)}
\end{aligned}$$

In (A3.20) and all subsequent derivations, terms involving $E[g_I(\bullet)\,g_Q(\bullet)]$, $E[n_{uI}(\bullet)\,g_I(\bullet)]$, $E[n_{uI}(\bullet)\,g_Q(\bullet)]$, $E[n_{uQ}(\bullet)\,g_I(\bullet)]$, $E[n_{uQ}(\bullet)\,g_Q(\bullet)]$ and $E[n_{uI}(\bullet)\,n_{uQ}(\bullet)]$ will not be shown as they equal 0. Also, equation 5.54 will be used extensively to relate the terms $a^2 \sigma_g^2$, $A^2 \sigma_g^2$, and $A\, a\, \sigma_g^2$ to E_b, as it has been used to obtain (A3.20).

2) The auto-correlation function, $R_{\hat{u}_I\hat{u}_Q}(i\text{-}k)$, is given by

$$
\begin{aligned}
R_{\hat{u}_I\hat{u}_Q}(i-k) &= E[\hat{u}_I(i)\, \hat{u}_Q(k)] \\
&= E\Big[\Big\{ A\{ g_I(i) \cos(2\pi f_o i) - g_Q(i) \sin(2\pi f_o i)\} + n_{uI}(i) \Big\} \\
&\quad \times \Big\{ A\{ g_Q(k) \cos(2\pi f_o k) + g_I(k) \sin(2\pi f_o k)\} + n_{uQ}(k) \Big\}\Big] \\
&= E\Big[A^2 \{ g_I(i) \cos(2\pi f_o i)\, g_I(k) \sin(2\pi f_o k) \} \\
&\quad - A^2 \{ g_Q(i) \sin(2\pi f_o i)\, g_Q(k) \cos(2\pi f_o k) \}\Big] \\
&= -\, E[g_I(i)\, g_I(k)]\, A^2 \sin(2\pi f_o(i-k)) \\
&= -\, A^2 R_g(i-k) \sin(2\pi f_o(i-k)) \\
&= -\frac{E_b}{1+r} \tilde{R}_g(i-k) \sin(2\pi f_o(i-k))
\end{aligned} \qquad \text{(A3.21)}
$$

3) The auto-correlation function, $R_{\hat{u}_Q\hat{u}_Q}(i\text{-}k)$, is given by

$$
\begin{aligned}
R_{\hat{u}_Q\hat{u}_Q}(i-k) &= E[\hat{u}_Q(i)\, \hat{u}_Q(k)] \\
&= E\Big[A\{ g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i)\} + n_{uQ}(i) \\
&\quad \times A\{ g_Q(k) \cos(2\pi f_o k) + g_I(k) \sin(2\pi f_o k)\} + n_{uQ}(k)\Big] \\
&= E\Big[A^2 \{ g_Q(i) \cos(2\pi f_o i)\, g_Q(k) \cos(2\pi f_o k) \} \\
&\quad + A^2 \{ g_I(i) \sin(2\pi f_o i)\, g_I(k) \sin(2\pi f_o k) \} + n_{uQ}(i)\, n_{uQ}(k)\Big] \\
&= E[g_Q(i)\, g_Q(k)]\, A^2 \cos(2\pi f_o(i-k)) + E[n_{uQ}(i)\, n_{uQ}(k)] \\
&= A^2 R_{gQ}(i-k) \cos(2\pi f_o(i-k)) + \sigma^2_{n_{uQ}}\, \delta_{ik}
\end{aligned}
$$

$$= \frac{E_b}{1+r} \tilde{R}_g(i-k) \cos(2\pi f_o(i-k)) + \sigma^2_{n_{uQ}} \delta_{ik}$$

$$= E[\hat{u}_I(i)\, \hat{u}_I(k)] \tag{A3.22}$$

4) The auto-correlation function, $R_{v_Iv_I}(i-k)$, is given by:

$$R_{v_Iv_I}(i-k) = E[v_I(i)\, v_I(k)]$$

$$= E\Big[\{a\, \mathrm{Re}[c(i) * \Delta h_p(i)] + n_{v_I}(i)\}\ \{a\, \mathrm{Re}[c(k) * \Delta h_p(k)] + n_{v_I}(k)\}\Big]$$

$$= E\Big[\Big\{a \sum_{n=-\infty}^{\infty} \mathrm{Re}\{g(i-n) \exp(j2\pi f_o(i-n))\, \Delta h_p(n)\} + n_{v_I}(i)\Big\}$$
$$\times \Big\{a \sum_{m=-\infty}^{\infty} \mathrm{Re}\{g(k-m) \exp(j2\pi f_o(k-m))\, \Delta h_p(m)\} + n_{v_I}(k)\Big\}\Big]$$

$$= E\Big[\Big\{a \sum_{n=-\infty}^{\infty} \Big\{g_I(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pI}(n) - g_Q(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pQ}(n)$$
$$- g_I(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pQ}(n) - g_Q(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pI}(n) + n_{v_I}(i)\Big\}\Big\}$$
$$\times \Big\{a \sum_{m=-\infty}^{\infty} \Big\{g_I(k-m) \cos(2\pi f_o(k-m))\, \Delta h_{pI}(m) - g_Q(k-m) \cos(2\pi f_o(k-m))\, \Delta h_{pQ}(m)$$
$$- g_I(k-m) \sin(2\pi f_o(k-m))\, \Delta h_{pQ}(m) - g_Q(k-m) \sin(2\pi f_o(k-m))\, \Delta h_{pI}(m)\Big\} + n_{v_I}(k)\Big\}\Big]$$

$$= a^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Big\{E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pI}(m) \cos(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$
$$- E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m) \cos(2\pi f_o(i-n)) \sin(2\pi f_o(k-m))$$
$$+ E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m) \cos(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$
$$+ E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \cos(2\pi f_o(i-n)) \sin(2\pi f_o(k-m))$$
$$+ E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m) \sin(2\pi f_o(i-n)) \sin(2\pi f_o(k-m))$$
$$- E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \sin(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$

$$+ \; E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pI}(m)\, \sin(2\pi f_o(i-n))\, \sin(2\pi f_o(k-m))$$

$$+ \; E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m)\, \sin(2\pi f_o(i-n))\, \cos(2\pi f_o(k-m)) \Big\}$$

$$+ \; E[n_{vI}(i)\, n_{vI}(k)]$$

$$= a^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Big\{ R_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pI}(m)$$

$$+ \; R_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m)$$

$$+ \; R_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m)$$

$$- \; R_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \Big\} + \sigma^2_{n_{vI}}\, \delta_{ik}$$

$$= \frac{r}{1+r} E_b R_b \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Big\{ \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pI}(m)$$

$$+ \; \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m)$$

$$+ \; \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m)$$

$$- \; \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \Big\}$$

$$+ \; \sigma^2_{n_{vI}}\, \delta_{ik} \qquad \text{(A3.23)}$$

5) The auto-correlation function, $R_{v_Iv_Q}(i-k)$, is given by:

$$R_{v_Iv_Q}(i-k) = E[v_I(i)\, v_Q(k)]$$

$$= E\Big[\{a\, \mathrm{Re}[c(i) * \Delta h_p(i)] + n_{vI}(i)\}\, \{a\, \mathrm{Im}[c(k) * \Delta h_p(k)] + n_{vQ}(k)\} \Big]$$

$$= E\Big[\Big\{ a \sum_{n=-\infty}^{\infty} \mathrm{Re}\{g(i-n) \exp(j2\pi f_o(i-n))\, \Delta h_p(n)\} + n_{vI}(i) \Big\}$$

$$\times \; \Big\{ a \sum_{m=-\infty}^{\infty} \mathrm{Im}\{g(k-m) \exp(j2\pi f_o(k-m))\, \Delta h_p(m)\} + n_{vQ}(k) \Big\} \Big]$$

$$= E\Big[\Big\{ a \sum_{n=-\infty}^{\infty} \{ g_I(i-n)\cos(2\pi f_o(i-n))\,\Delta h_{pI}(n) - g_Q(i-n)\cos(2\pi f_o(i-n))\,\Delta h_{pQ}(n)$$

$$- g_I(i-n)\sin(2\pi f_o(i-n))\,\Delta h_{pQ}(n) - g_Q(i-n)\sin(2\pi f_o(i-n))\,\Delta h_{pI}(n) \} + n_{vI}(i) \Big\}$$

$$\times \Big\{ a \sum_{n=-\infty}^{\infty} \{ g_I(k-m)\sin(2\pi f_o(k-m))\,\Delta h_{pI}(m) - g_Q(k-m)\sin(2\pi f_o(k-m))\,\Delta h_{pQ}(m)$$

$$+ g_I(k-m)\cos(2\pi f_o(k-m))\,\Delta h_{pQ}(m) + g_Q(k-m)\cos(2\pi f_o(k-m))\,\Delta h_{pI}(m) \} + n_{vQ}(k) \Big\}\Big]$$

$$= a^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{ E[g_I(i-n)\, g_I(k-m)]\,\Delta h_{pI}(n)\,\Delta h_{pI}(m)\cos(2\pi f_o(i-n))\sin(2\pi f_o(k-m))$$

$$+ E[g_I(i-n)\, g_I(k-m)]\,\Delta h_{pI}(n)\,\Delta h_{pQ}(m)\cos(2\pi f_o(i-n))\cos(2\pi f_o(k-m))$$

$$+ E[g_Q(i-n)\, g_Q(k-m)]\,\Delta h_{pQ}(n)\,\Delta h_{pQ}(m)\cos(2\pi f_o(i-n))\sin(2\pi f_o(k-m))$$

$$- E[g_Q(i-n)\, g_Q(k-m)]\,\Delta h_{pQ}(n)\,\Delta h_{pI}(m)\cos(2\pi f_o(i-n))\cos(2\pi f_o(k-m))$$

$$- E[g_I(i-n)\, g_I(k-m)]\,\Delta h_{pQ}(n)\,\Delta h_{pI}(m)\sin(2\pi f_o(i-n))\sin(2\pi f_o(k-m))$$

$$- E[g_I(i-n)\, g_I(k-m)]\,\Delta h_{pQ}(n)\,\Delta h_{pQ}(m)\sin(2\pi f_o(i-n))\cos(2\pi f_o(k-m))$$

$$+ E[g_Q(i-n)\, g_Q(k-m)]\,\Delta h_{pI}(n)\,\Delta h_{pQ}(m)\sin(2\pi f_o(i-n))\sin(2\pi f_o(k-m))$$

$$- E[g_Q(i-n)\, g_Q(k-m)]\,\Delta h_{pI}(n)\,\Delta h_{pI}(m)\sin(2\pi f_o(i-n))\cos(2\pi f_o(k-m)) \}$$

$$= a^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{ - R_g(i-n-(k-m))\sin(2\pi f_o(i-n-(k-m)))\,\Delta h_{pI}(n)\,\Delta h_{pI}(m)$$

$$+ R_g(i-n-(k-m))\cos(2\pi f_o(i-n-(k-m)))\,\Delta h_{pI}(n)\,\Delta h_{pQ}(m)$$

$$- R_g(i-n-(k-m))\sin(2\pi f_o(i-n-(k-m)))\,\Delta h_{pQ}(n)\,\Delta h_{pQ}(m)$$

$$- R_g(i-n-(k-m))\cos(2\pi f_o(i-n-(k-m)))\,\Delta h_{pQ}(n)\,\Delta h_{pI}(m) \}$$

$$= \frac{r}{1+r} E_b R_b \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{ - \tilde{R}_g(i-n-(k-m))\sin(2\pi f_o(i-n-(k-m)))\,\Delta h_{pI}(n)\,\Delta h_{pI}(m)$$

$$+ \tilde{R}_g(i-n-(k-m))\cos(2\pi f_o(i-n-(k-m)))\,\Delta h_{pI}(n)\,\Delta h_{pQ}(m)$$

$$- \tilde{R}_g(i-n-(k-m))\sin(2\pi f_o(i-n-(k-m)))\,\Delta h_{pQ}(n)\,\Delta h_{pQ}(m)$$

$$- \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m))) \Delta h_{pQ}(n) \Delta h_{pI}(m) \Big\}$$

(A3.24)

6) The auto-correlation function, $R_{v_Q v_Q}(i-k)$, is given by:

$$R_{v_Q v_Q}(i-k) = E[v_Q(i)\, v_Q(k)]$$

$$= E\Big[\{a\, \mathrm{Im}[c(i) * \Delta h_p(i)] + n_{vQ}(i)\}\ \{a\, \mathrm{Im}[c(k) * \Delta h_p(k)] + n_{vQ}(k)\}\Big]$$

$$= E\Big[\Big\{a \sum_{n=-\infty}^{\infty} \mathrm{Im}\{g(i-n) \exp(j2\pi f_o(i-n))\, \Delta h_p(n)\} + n_{vQ}(i)\Big\}$$

$$\times \Big\{a \sum_{m=-\infty}^{\infty} \mathrm{Im}\{g(k-m) \exp(j2\pi f_o(k-m))\, \Delta h_p(m)\} + n_{vQ}(k)\Big\}\Big]$$

$$= E\Big[\Big\{a \sum_{n=-\infty}^{\infty} \Big\{g_I(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pI}(n) - g_Q(i-n) \sin(2\pi f_o(i-n))\, \Delta h_{pQ}(n)$$

$$+ g_I(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pQ}(n) + g_Q(i-n) \cos(2\pi f_o(i-n))\, \Delta h_{pI}(n) + n_{vQ}(i)\Big\}\Big\}$$

$$\times \Big\{a \sum_{n=-\infty}^{\infty} \Big\{g_I(k-m) \sin(2\pi f_o(k-m))\, \Delta h_{pI}(m) - g_Q(k-m) \sin(2\pi f_o(k-m))\, \Delta h_{pQ}(m)$$

$$+ g_I(k-m) \cos(2\pi f_o(k-m))\, \Delta h_{pQ}(m) + g_Q(k-m) \cos(2\pi f_o(k-m))\, \Delta h_{pI}(m)\Big\} + n_{vQ}(k)\Big\}\Big]$$

$$= a^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Big\{E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pI}(m) \sin(2\pi f_o(i-n)) \sin(2\pi f_o(k-m))$$

$$+ E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m) \sin(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$

$$+ E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m) \sin(2\pi f_o(i-n)) \sin(2\pi f_o(k-m))$$

$$- E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \sin(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$

$$+ E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m) \cos(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$

$$+ E[g_I(i-n)\, g_I(k-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \cos(2\pi f_o(i-n)) \sin(2\pi f_o(k-m))$$

$$+ E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pI}(m) \cos(2\pi f_o(i-n)) \cos(2\pi f_o(k-m))$$

$$- \; E[g_Q(i-n)\, g_Q(k-m)]\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m) \cos(2\pi f_o(i-n)) \sin(2\pi f_o(k-m)) \Big\}$$

$$+ E[n_{vQ}(i)\, n_{vQ}(k)]$$

$$= a^2 \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Big\{ R_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pI}(m)$$

$$+ \; R_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m)$$

$$+ \; R_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m)$$

$$- \; R_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \Big\} + \sigma^2_{n_{vQ}} \delta_{ik}$$

$$= \frac{r}{1+r} E_b R_b \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Big\{ \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pI}(m)$$

$$+ \; \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pI}(n)\, \Delta h_{pQ}(m)$$

$$+ \; \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m)$$

$$- \; \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))\, \Delta h_{pQ}(n)\, \Delta h_{pI}(m) \Big\}$$

$$+ \sigma^2_{n_{vQ}} \delta_{ik}$$

$$= E[v_I(i)\, v_I(k)] \qquad \text{(A3.25)}$$

7) The auto-correlation function, $R_{\hat{u}_I v_I}(i-k)$, is given by

$$R_{\hat{u}_I v_I}(i-k) = E[\hat{u}_I(i)\, v_I(k)]$$

$$= E\Big[\Big\{ A\{g_I(i) \cos(2\pi f_o i) - g_Q(i) \sin(2\pi f_o i)\} + n_{uI}(i) \Big\}$$

$$\times \; \Big\{ a\, \mathrm{Re}[c(k) * \Delta h_p(k)] + n_{vI}(k) \Big\} \Big]$$

$$= E\Big[\Big\{ A\{g_I(i) \cos(2\pi f_o i) - g_Q(i) \sin(2\pi f_o i)\} + n_{uI}(i) \Big\}$$

$$\times \Big\{ a \sum_{n=-\infty}^{\infty} \mathrm{Re}\{g(k-n) \exp(j2\pi f_o(k-n)\, \Delta h_p(n)\} + n_{vI}(k) \Big\} \Big]$$

$$= E\Big[\big\{A\{g_I(i)\cos(2\pi f_o i) - g_Q(i)\sin(2\pi f_o i)\} + n_{uI}(i)\big\}$$

$$\times \Big\{a \sum_{n=-\infty}^{\infty} \big\{g_I(k-n)\cos(2\pi f_o(k-n))\,\Delta h_{pI}(n) - g_I(k-n)\sin(2\pi f_o(k-n))\,\Delta h_{pQ}(n)$$

$$- g_Q(k-n)\cos(2\pi f_o(k-n))\,\Delta h_{pQ}(n) - g_Q(k-n)\sin(2\pi f_o(k-n))\,\Delta h_{pI}(n)\big\} + n_{vI}(k)\Big\}$$

$$= a\,A \sum_{n=-\infty}^{\infty} \big\{E[\,g_I(i)\,g_I(k-n)]\cos(2\pi f_o i)\cos(2\pi f_o(k-n))\,\Delta h_{pI}(n)$$

$$-\ E[\,g_I(i)\,g_I(k-n)\,]\cos(2\pi f_o i)\sin(2\pi f_o(k-n))\,\Delta h_{pQ}(n)$$

$$+\ E[\,g_Q(i)\,g_Q(k-n)\,]\sin(2\pi f_o i)\cos(2\pi f_o(k-n))\,\Delta h_{pQ}(n)$$

$$+\ E[\,g_Q(i)\,g_Q(k-n)\,]\sin(2\pi f_o i)\sin(2\pi f_o(k-n))\,\Delta h_{pI}(n)\ \big\}$$

$$= a\,A \sum_{n=-\infty}^{\infty} \big\{R_g(i-(k-n))\cos(2\pi f_o(i-(k-n)))\,\Delta h_{pI}(n)$$

$$+ R_g(i-(k-n))\sin(2\pi f_o(k-n))\,\Delta h_{pQ}(n)\ \big\}$$

$$= E_b \frac{\sqrt{rR_b}}{1+r} \sum_{n=-\infty}^{\infty} \big\{\ \tilde{R}_g(i-(k-n))\cos(2\pi f_o(i-(k-n)))\,\Delta h_{pI}(n)$$

$$+ \tilde{R}_g(i-(k-n))\sin(2\pi f_o(k-n))\,\Delta h_{pQ}(n)\ \big\} \qquad \text{(A3.26)}$$

8) The auto-correlation function, $R_{\hat{u}_I v_Q}(i-k)$, is given by

$$R_{\hat{u}_I v_Q}(i-k) = E[\hat{u}_I(i)\,v_Q(k)]$$

$$= E\Big[\big\{A\{\,g_I(i)\cos(2\pi f_o i) - g_Q(i)\sin(2\pi f_o i)\} + n_{uI}(i)\big\}$$

$$\times\ \big\{a\,\mathrm{Im}[c(k) * \Delta h_p(k)] + n_{vQ}(k)\big\}\Big]$$

$$= E\big[A\{\,g_I(i)\cos(2\pi f_o i) - g_Q(i)\sin(2\pi f_o i)\} + n_{uI}(i)$$

$$\times \left\{ a \sum_{n=-\infty}^{\infty} \mathrm{Im}\{g(k-n) \exp(j2\pi f_o(k-n))\, \Delta h_p(n)\} + n_{vQ}(k) \right\} \Big]$$

$$= E\Big[A\{ g_I(i) \cos(2\pi f_o i) - g_Q(i) \sin(2\pi f_o i)\} + n_{uI}(i)$$

$$\times \left\{ a \sum_{n=-\infty}^{\infty} \Big\{ g_I(k-n) \sin(2\pi f_o(k-n))\, \Delta h_{pI}(n) + g_I(k-n) \cos(2\pi f_o(k-n))\, \Delta h_{pQ}(n) \right.$$

$$\left. - g_Q(k-n) \sin(2\pi f_o(k-n))\, \Delta h_{pQ}(n) + g_Q(k-n) \cos(2\pi f_o(k-n))\, \Delta h_{pI}(n) + n_{vQ}(k) \Big\} \right\}$$

$$= a\,A \sum_{n=-\infty}^{\infty} \Big\{ E[\, g_I(i)\, g_I(k-n)\,] \cos(2\pi f_o i) \sin(2\pi f_o(k-n))\, \Delta h_{pI}(n)$$

$$+ E[\, g_I(i)\, g_I(k-n)\,] \cos(2\pi f_o i) \cos(2\pi f_o(k-n))\, \Delta h_{pQ}(n)$$

$$+ E[\, g_Q(i)\, g_Q(k-n)\,] \sin(2\pi f_o i) \sin(2\pi f_o(k-n))\, \Delta h_{pQ}(n)$$

$$- E[\, g_Q(i)\, g_Q(k-n)\,] \sin(2\pi f_o i) \cos(2\pi f_o(k-n))\, \Delta h_{pI}(n) \Big\}$$

$$= a\,A \sum_{n=-\infty}^{\infty} \Big\{ - R_g(i-(k-n)) \sin(2\pi f_o(i-(k-n)))\, \Delta h_{pI}(n)$$

$$+ R_g(i-(k-n)) \cos(2\pi f_o(k-n))\, \Delta h_{pQ}(n) \Big\}$$

$$= E_b \frac{\sqrt{r R_b}}{1+r} \sum_{n=-\infty}^{\infty} \Big\{ - \tilde{R}_g(i-(k-n)) \sin(2\pi f_o(i-(k-n)))\, \Delta h_{pI}(n)$$

$$+ \tilde{R}_g(i-(k-n)) \cos(2\pi f_o(k-n))\, \Delta h_{pQ}(n) \Big\} \qquad \text{(A3.27)}$$

9) The auto-correlation function, $R_{\hat{u}_Q v_I}(i-k)$, is given by

$$R_{\hat{u}_Q v_I}(i-k) = E[\hat{u}_Q(i)\, v_I(k)]$$

$$= E\Big[\big\{ A\,\{g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i)\} + n_{uQ}(i) \big\}$$

$$\times \big\{ a\,\mathrm{Re}[c(k) * \Delta h_p(k)] + n_{vI}(k) \big\} \Big]$$

$$= E\Big[A\{g_Q(i)\cos(2\pi f_o i) + g_I(i)\sin(2\pi f_o i)\} + n_{uQ}(i)$$

$$\times \Big\{a \sum_{n=-\infty}^{\infty} \mathrm{Re}\{g(k-n)\exp(j2\pi f_o(k-n)\,\Delta h_p(n)\} + n_{vI}(k)\Big\}\Big]$$

$$= E\Big[A\{g_Q(i)\cos(2\pi f_o i) + g_I(i)\sin(2\pi f_o i)\} + n_{uQ}(i)$$

$$\times \Big\{a \sum_{n=-\infty}^{\infty} \Big\{g_I(k-n)\cos(2\pi f_o(k-n))\,\Delta h_{pI}(n) - g_I(k-n)\sin(2\pi f_o(k-n))\,\Delta h_{pQ}(n)$$

$$- g_Q(k-n)\cos(2\pi f_o(k-n))\,\Delta h_{pQ}(n) - g_Q(k-n)\sin(2\pi f_o(k-n))\,\Delta h_{pI}(n) + n_{vI}(k)\Big\}\Big\}$$

$$= a\,A \sum_{n=-\infty}^{\infty} \Big\{- E[g_Q(i)\,g_Q(k-n)]\cos(2\pi f_o i)\cos(2\pi f_o(k-n))\,\Delta h_{pQ}(n)$$

$$- E[g_Q(i)\,g_Q(k-n)]\cos(2\pi f_o i)\sin(2\pi f_o(k-n))\,\Delta h_{pI}(n)$$

$$+ E[g_I(i)\,g_I(k-n)]\sin(2\pi f_o i)\cos(2\pi f_o(k-n))\,\Delta h_{pI}(n)$$

$$- E[g_I(i)\,g_I(k-n)]\sin(2\pi f_o i)\sin(2\pi f_o(k-n))\,\Delta h_{pQ}(n)\Big\}$$

$$= a\,A \sum_{n=-\infty}^{\infty} \Big\{- R_g(i-(k-n))\cos(2\pi f_o(i-(k-n)))\,\Delta h_{pQ}(n)$$

$$+ R_g(i-(k-n))\sin(2\pi f_o(k-n))\,\Delta h_{pI}(n)\Big\}$$

$$= E_b \frac{\sqrt{rR_b}}{1+r} \sum_{n=-\infty}^{\infty} \Big\{- \tilde{R}_g(i-(k-n))\cos(2\pi f_o(i-(k-n)))\,\Delta h_{pQ}(n)$$

$$+ \tilde{R}_g(i-(k-n))\sin(2\pi f_o(k-n))\,\Delta h_{pI}(n)\Big\}$$

$$= - E[\hat{u}_I(i)\,v_Q(k)] \qquad \text{(A3.28)}$$

10) The auto-correlation function, $R_{\hat{u}_Q v_Q}(i-k)$, is given by

$$
\begin{aligned}
R_{\hat{u}_Q v_Q}(i-k) &= E[\hat{u}_Q(i)\, v_Q(k)] \\
&= E\Big[\big\{ A\, \{ g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i) \} + n_{uQ}(i) \big\} \\
&\quad \times \big\{ a\, \mathrm{Im}[c(k) * \Delta h_p(k)] + n_{vQ}(k) \big\} \Big]
\end{aligned}
$$

$$
\begin{aligned}
&= E\Big[\big\{ A\, \{ g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i) \} + n_{uQ}(i) \big\} \\
&\quad \times \Big\{ a \sum_{n=-\infty}^{\infty} \mathrm{Im}\{ g(k-n) \exp(j2\pi f_o (k-n))\, \Delta h_p(n) \} + n_{vQ}(k) \Big\} \Big]
\end{aligned}
$$

$$
\begin{aligned}
&= E\Big[A\{ g_Q(i) \cos(2\pi f_o i) + g_I(i) \sin(2\pi f_o i) \} + n_{uQ}(i) \\
&\quad \times \Big\{ a \sum_{n=-\infty}^{\infty} \big\{ g_I(k-n) \sin(2\pi f_o (k-n))\, \Delta h_{pI}(n) + g_I(k-n) \cos(2\pi f_o (k-n))\, \Delta h_{pQ}(n) \\
&\quad - g_Q(k-n) \sin(2\pi f_o (k-n))\, \Delta h_{pQ}(n) + g_Q(k-n) \cos(2\pi f_o (k-n))\, \Delta h_{pI}(n) + n_{vQ}(k) \big\} \Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= a\, A \sum_{n=-\infty}^{\infty} \big\{ - E[\, g_Q(i)\, g_Q(k-n)\,] \cos(2\pi f_o i) \sin(2\pi f_o (k-n))\, \Delta h_{pQ}(n) \\
&\qquad + E[\, g_Q(i)\, g_Q(k-n)\,] \cos(2\pi f_o i) \cos(2\pi f_o (k-n))\, \Delta h_{pI}(n) \\
&\qquad + E[\, g_I(i)\, g_I(k-n)\,] \sin(2\pi f_o i) \sin(2\pi f_o (k-n))\, \Delta h_{pI}(n) \\
&\qquad + E[\, g_I(i)\, g_I(k-n)\,] \sin(2\pi f_o i) \cos(2\pi f_o (k-n))\, \Delta h_{pQ}(n) \big\}
\end{aligned}
$$

$$
\begin{aligned}
&= a\, A \sum_{n=-\infty}^{\infty} \big\{ R_g(i-(k-n)) \sin(2\pi f_o (i-(k-n)))\, \Delta h_{pQ}(n) \\
&\qquad + R_g(i-(k-n)) \cos(2\pi f_o (k-n))\, \Delta h_{pI}(n) \big\}
\end{aligned}
$$

$$
\begin{aligned}
&= E_b \frac{\sqrt{r R_b}}{1+r} \sum_{n=-\infty}^{\infty} \big\{ \tilde{R}_g(i-(k-n)) \sin(2\pi f_o (i-(k-n)))\, \Delta h_{pQ}(n) \\
&\qquad + \tilde{R}_g(i-(k-n)) \cos(2\pi f_o (k-n))\, \Delta h_{pI}(n) \big\} \qquad \text{(A3.29)}
\end{aligned}
$$

Each summation term in the correlation functions involving the in-phase and/or quadrature components of v is in the form:

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \left\{ R(i-n-(k-m))\ \Delta h_{p1}(n)\ \Delta h_{p2}(m) \right\}$$

where $R(i-n-(k-m)) = \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m)))$

or $\tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m)))$

$\Delta h_{p1}(n) = \Delta h_{pI}(n)$ or $\Delta h_{pQ}(n)$

$\Delta h_{p2}(m) = \Delta h_{pI}(m)$ or $\Delta h_{pQ}(m)$.

For convenience, define $\Delta H_{p1}(e^{j\omega}) = \mathcal{F}\ \{\Delta h_{p1}(n)\}$,

$\Delta H_{p2}(e^{j\omega}) = \mathcal{F}\ \{\Delta h_{p2}(n)\}$ and

$S(e^{j\omega}) = \mathcal{F}\ \{R(n)\}$.

Let $l = n-m$, then

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \left\{ R(i-n-(k-m))\ \Delta h_{p1}(n)\ \Delta h_{p2}(m) \right\}$$

$$= \sum_{l=-\infty}^{\infty} R(i-k-l) \sum_{m=-\infty}^{\infty} \Delta h_{p1}(l+m)\ \Delta h_{p2}(m)$$

$$= \sum_{l=-\infty}^{\infty} R(i-k-l)\ h(l) \qquad \text{where } h(l) = \sum_{m=-\infty}^{\infty} \Delta h_{p1}(l+m)\ \Delta h_{p2}(m) \text{ , and}$$

$$\mathcal{F}\ \{h(l)\} = [\Delta H_{p1}(e^{j\omega})]^* \ \Delta H_{p2}(e^{j\omega})$$

$$= \mathcal{F}^{-1}\left\{ \mathcal{F}\ \{ \sum_{l=-\infty}^{\infty} R(i-k-l)\ h(l) \} \right\}$$

$$= F^{-1}\left\{ S(e^{j\omega}) \, [\Delta H_{p_1}(e^{j\omega})]^* \, \Delta H_{p_2}(e^{j\omega}) \right\}$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} S(e^{j\omega}) \, [\Delta H_{p_1}(e^{j\omega})]^* \, \Delta H_{p_2}(e^{j\omega}) \, e^{j\omega(i-k)} \, d\omega \qquad \text{(A3.30)}$$

Similarly, each summation term in the correlation functions involving the in-phase and/or quadrature components of $\hat{u}$ and v is in the form:

$$\sum_{n=-\infty}^{\infty} \left\{ R(i-(k-n)) \;\; \Delta h_{p_1}(n) \right\} \qquad \text{where } R(i-(k-n)) = \tilde{R}_g(i-(k-n)) \sin(2\pi f_o(i-(k-n)))$$

$$\text{or } \tilde{R}_g(i-(k-n)) \cos(2\pi f_o(i-(k-n)))$$

$$\Delta h_{p_1}(n) = \Delta h_{p_I}(n) \text{ or } \Delta h_{p_Q}(n) \text{ as before.}$$

$$\sum_{n=-\infty}^{\infty} \left\{ R(i-(k-n)) \;\; \Delta h_{p_1}(n) \right\} = F^{-1}\left\{ F\left\{ \sum_{n=-\infty}^{\infty} R(i-(k-n)) \;\; \Delta h_{p_1}(n) \right\} \right\}$$
$$= F^{-1}\left\{ S(e^{j\omega}) \, \Delta H_{p_1}(e^{j\omega}) \right\}$$
$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} S(e^{j\omega}) \, \Delta H_{p_1}(e^{j\omega}) \, e^{j\omega(i-k)} \, d\omega \qquad \text{(A3.31)}$$

(A3.30) and (A3.31) can be used to derive expressions for $R_{v_Iv_I}(i-k)$, $R_{v_Iv_Q}(i-k)$, $R_{v_Qv_Q}(i-k)$, $R_{\hat{u}_Iv_I}(i-k)$, $R_{\hat{u}_Iv_Q}(i-k)$, $R_{\hat{u}_Qv_I}(i-k)$, and $R_{\hat{u}_Qv_Q}(i-k)$ in the frequency domain.

The results are summarized as follows:

1) $$R_{\hat{u}_I\hat{u}_I}(i-k) = \frac{E_b}{1+r} \tilde{R}_g(i-k) \cos(2\pi f_o(i-k)) + \sigma^2_{n_{uI}} \delta_{ik} \qquad \text{(A3.32)}$$

2) $$R_{\hat{u}_I\hat{u}_Q}(i-k) = -\frac{E_b}{1+r} \tilde{R}_g(i-k) \sin(2\pi f_o(i-k)) \qquad \text{(A3.33)}$$

3) $$R_{\hat{u}_Q\hat{u}_Q}(i-k) = R_{\hat{u}_I\hat{u}_I}(i-k) \qquad \text{(A3.34)}$$

4) $$R_{v_I v_I}(i-k) = \frac{r}{1+r} E_b R_b \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{ \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m))) \Delta h_{pI}(n) \Delta h_{pI}(m)$$

$$+ \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m))) \Delta h_{pI}(n) \Delta h_{pQ}(m)$$

$$+ \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m))) \Delta h_{pQ}(n) \Delta h_{pQ}(m)$$

$$- \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m))) \Delta h_{pQ}(n) \Delta h_{pI}(m) \}$$

$$+ \sigma^2_{n_{vI}} \delta_{ik}$$

$$= \frac{r}{1+r} E_b R_b \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) + \tilde{S}_g(e^{j(\omega+\omega o)})}{2} [\Delta H_{pI}(e^{j\omega})]^* \Delta H_{pI}(e^{j\omega}) e^{j\omega(i-k)} d\omega \right.$$

$$+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) - \tilde{S}_g(e^{j(\omega+\omega o)})}{2j} [\Delta H_{pI}(e^{j\omega})]^* \Delta H_{pQ}(e^{j\omega}) e^{j\omega(i-k)} d\omega$$

$$+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) + \tilde{S}_g(e^{j(\omega+\omega o)})}{2} [\Delta H_{pQ}(e^{j\omega})]^* \Delta H_{pQ}(e^{j\omega}) e^{j\omega(i-k)} d\omega$$

$$\left. - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) - \tilde{S}_g(e^{j(\omega+\omega o)})}{2j} [\Delta H_{pQ}(e^{j\omega})]^* \Delta H_{pI}(e^{j\omega}) e^{j\omega(i-k)} d\omega \right\}$$

$$+ \sigma^2_{n_{vI}} \delta_{ik} \qquad \text{(A3.35)}$$

5) $$R_{v_I v_Q}(i-k) = \frac{r}{1+r} E_b R_b \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{ - \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m))) \Delta h_{pI}(n) \Delta h_{pI}(m)$$

$$+ \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m))) \Delta h_{pI}(n) \Delta h_{pQ}(m)$$

$$- \tilde{R}_g(i-n-(k-m)) \sin(2\pi f_o(i-n-(k-m))) \Delta h_{pQ}(n) \Delta h_{pQ}(m)$$

$$- \tilde{R}_g(i-n-(k-m)) \cos(2\pi f_o(i-n-(k-m))) \Delta h_{pQ}(n) \Delta h_{pI}(m) \}$$

$$= \frac{r}{1+r} E_b R_b \left\{ - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) - \tilde{S}_g(e^{j(\omega+\omega o)})}{2j} [\Delta H_{pI}(e^{j\omega})]^* \Delta H_{pI}(e^{j\omega}) e^{j\omega(i-k)} d\omega \right.$$

$$+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) + \tilde{S}_g(e^{j(\omega+\omega o)})}{2} [\Delta H_{pI}(e^{j\omega})]^* \Delta H_{pQ}(e^{j\omega}) e^{j\omega(i-k)} d\omega$$

$$- \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega o)}) - \tilde{S}_g(e^{j(\omega+\omega o)})}{2j} [\Delta H_{pQ}(e^{j\omega})]^* \Delta H_{pQ}(e^{j\omega}) e^{j\omega(i-k)} d\omega$$

$$- \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega_o)}) + \tilde{S}_g(e^{j(\omega+\omega_o)})}{2} [\Delta H_{pQ}(e^{j\omega})]^* \Delta H_{pI}(e^{j\omega})\, e^{j\omega(i-k)}\, d\omega \Big\} \tag{A3.36}$$

6) $$R_{v_Q v_Q}(i-k) = R_{v_I v_I}(i-k) \tag{A3.37}$$

7) $$R_{\hat{u}_I v_I}(i-k) = E_b \frac{\sqrt{rR_b}}{1+r} \sum_{n=-\infty}^{\infty} \Big\{ \tilde{R}_g(i-(k-n)) \cos(2\pi f_o(i-(k-n)))\, \Delta h_{pI}(n) + \tilde{R}_g(i-(k-n)) \sin(2\pi f_o(k-n))\, \Delta h_{pQ}(n) \Big\}$$

$$= E_b \frac{\sqrt{rR_b}}{1+r} \Big\{ \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega_o)}) + \tilde{S}_g(e^{j(\omega+\omega_o)})}{2} \Delta H_{pI}(e^{j\omega})\, e^{j\omega(i-k)}\, d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega_o)}) - \tilde{S}_g(e^{j(\omega+\omega_o)})}{2j} \Delta H_{pQ}(e^{j\omega})\, e^{j\omega(i-k)}\, d\omega \tag{A3.38}$$

8) $$R_{\hat{u}_I v_Q}(i-k) = E_b \frac{\sqrt{rR_b}}{1+r} \sum_{n=-\infty}^{\infty} \Big\{ - \tilde{R}_g(i-(k-n)) \sin(2\pi f_o(i-(k-n)))\, \Delta h_{pI}(n) + \tilde{R}_g(i-(k-n)) \cos(2\pi f_o(k-n))\, \Delta h_{pQ}(n) \Big\}$$

$$= E_b \frac{\sqrt{rR_b}}{1+r} \Big\{ \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega_o)}) + \tilde{S}_g(e^{j(\omega+\omega_o)})}{2} \Delta H_{pQ}(e^{j\omega})\, e^{j\omega(i-k)}\, d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{S}_g(e^{j(\omega-\omega_o)}) - \tilde{S}_g(e^{j(\omega+\omega_o)})}{2j} \Delta H_{pI}(e^{j\omega})\, e^{j\omega(i-k)}\, d\omega \tag{A3.39}$$

9) $$R_{\hat{u}_Q v_I}(i-k) = - R_{\hat{u}_I v_Q}(i-k) \tag{A3.40}$$

10) $$R_{\hat{u}_Q v_Q}(i-k) = R_{\hat{u}_I v_I}(i-k) \tag{A3.41}$$

Symmetries in the Correlation Functions of $\hat{u}(k)$ and $v(k)$

Since Δh_p is conjugate symmetric, Δh_{pI} has even symmetry and Δh_{pQ} has odd symmetry. This, in addition to the fact that R_g is real and even, means that the following symmetries exist:

1) $R_{\hat{u}_I\hat{u}_I}(i-k) = R_{\hat{u}_I\hat{u}_I}(k-i)$ (A3.42)

2) $R_{\hat{u}_I\hat{u}_Q}(i-k) = -R_{\hat{u}_I\hat{u}_Q}(k-i)$ (A3.43)

3) $R_{\hat{u}_Q\hat{u}_Q}(i-k) = R_{\hat{u}_Q\hat{u}_Q}(k-i)$ (A3.44)

4) $R_{v_Iv_I}(i-k) = R_{v_Iv_I}(k-i)$ (A3.45)

5) $R_{v_Iv_Q}(i-k) = -R_{v_Iv_Q}(k-i)$ (A3.46)

6) $R_{v_Qv_Q}(i-k) = R_{v_Qv_Q}(k-i)$ (A3.47)

7) $R_{\hat{u}_Iv_I}(i-k) = R_{\hat{u}_Iv_I}(k-i)$ (A3.48)

8) $R_{\hat{u}_Iv_Q}(i-k) = -R_{\hat{u}_Iv_Q}(k-i)$ (A3.49)

9) $R_{\hat{u}_Qv_I}(i-k) = -R_{\hat{u}_Qv_I}(k-i)$ (A3.50)

10) $R_{\hat{u}_Qv_Q}(i-k) = R_{\hat{u}_Qv_Q}(k-i)$ (A3.51)

APPENDIX 4 : Derivation of Noise Variance at the Gradient Filter Output

The Gaussian noise at the gradient filter output can be expressed in terms of the in-phase and quadrature components of the input noise and filter impulse response as follows:

$$\begin{aligned} n_v(i) &= \sum_{n=-\infty}^{\infty} n(i-n)\ \Delta h_p(n) \\ &= \sum_{n=-\infty}^{\infty} \{n_I(i-n)+jn_Q(i-n)\}\ \{\Delta h_{pI}(n)+j\Delta h_{pQ}(n)\} \\ &= \sum_{n=-\infty}^{\infty} \{n_I(i-n)\ \Delta h_{pI}(n) - n_Q(i-n)\ \Delta h_{pQ}(n)\} + j\ \{n_Q(i-n)\ \Delta h_{pQ}(n) + n_I(i-n)\ \Delta h_{pI}(n)\} \end{aligned} \tag{A4.1}$$

where n(i) is the input noise assumed to be white Gaussian.

The in-phase and quadrature components of n_v are thus given by:

$$n_{vI}(i) = \sum_{n=-\infty}^{\infty} \{n_I(i-n)\ \Delta h_{pI}(n) - n_Q(i-n)\ \Delta h_{pQ}(n)\} \tag{A4.2}$$

$$n_{vQ}(i) = \sum_{n=-\infty}^{\infty} \{n_Q(i-n)\ \Delta h_{pQ}(n) + n_I(i-n)\ \Delta h_{pI}(n)\} \tag{A4.3}$$

The variances of n_{vI} and n_{vQ} can be derived as follows:

$$\begin{aligned} \sigma^2_{n_{vI}} &= E[n_{vI}(i)n_{vI}(i)] \\ &= E\Big[\sum_{n=-\infty}^{\infty} \{n_I(i-n)\ \Delta h_{pI}(n) - n_Q(i-n)\ \Delta h_{pQ}(n)\} \\ &\quad \times \sum_{m=-\infty}^{\infty} \{n_I(i-m)\ \Delta h_{pI}(m) - n_Q(i-m)\ \Delta h_{pQ}(m)\}\Big] \end{aligned}$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{E[n_I(i-n)\, n_I(i-m)]\, \Delta h_{pI}(n)\, \Delta h_{pI}(m)$$

$$+ E[n_Q(i-n)\, n_Q(i-m)]\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m)\}$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \{R_{n_I n_I}(m-n)\, \Delta h_{pI}(n)\, \Delta h_{pI}(m) + R_{n_Q n_Q}(m-n)\, \Delta h_{pQ}(n)\, \Delta h_{pQ}(m)\}$$

(A4.4)

Terms involving $E[n_I(\bullet)n_Q(\bullet)]$ are omitted as they equal 0.

Since samples of n_I and n_Q are uncorrelated, $R_{n_I n_I}(m-n) = R_{n_Q n_Q}(m-n) = \begin{cases} N_o & m = n \\ 0 & m \neq n \end{cases}$

where N_o is the variance of the input noise. Substituting for $R_{n_I n_I}(m-n)$ and $R_{n_Q n_Q}(m-n)$ and using Parseval's theorem gives:

$$\sigma^2_{n_{vI}} = N_o \sum_{n=-\infty}^{\infty} \{\Delta h_{pI}(n)^2 + \Delta h_{pQ}(n)^2\}$$

$$= \frac{N_o}{2\pi} \int_{-\pi}^{\pi} \{|\Delta H_{pI}(e^{j\omega})|^2 + |\Delta H_{pQ}(e^{j\omega})|^2\}\, d\omega \qquad \text{(A4.5)}$$

The variance of the quadrature component of n_v, $\sigma^2_{n_{vQ}}$, can be derived in a similar fashion and can be shown to be equal to $\sigma^2_{n_{vI}}$.

LIST OF REFERENCES

1. K. Murota, K. Hirade, "GMSK Modulation for Mobile Radio Telephony," IEEE Trans. on Comm., Vol. COM-29, pp. 1104-1050, July 1981.

2. K. S. Chung, L. E. Zegers, "Generalized Tamed Frequency Modulation," Proc. IEEE Int. Conf. on ASSP, Vol. 3, pp. 1805-1808, Paris, May 1982.

3. D. Divsalar, M. K. Simon, "Trellis-Coded Modulation for 4800-9600 bits/s Transmission Over a Fading Mobile Satellite Channel," IEEE Journal on Selected Areas in Comm., Vol. SAC-5, No. 2, pp. 162-175, Feb. 1987.

4. Faramaz Davarian, "Mobile Digital Communication via Tone Calibration," IEEE Trans. on Veh. Technol., Vol. VT-36, No. 2, pp. 55-62, May 1987.

5. S. Sampei, T. Sunaga, "Rayleigh Fading Compensation Method for 16QAM in Digital Land Mobile Radio Channels," Proc. IEEE Veh. Technol. Conf., Vol 2, pp. 640-646, San Francisco, May 1988.

6. William C. Y. Lee, *Mobile Communications Engineering* , McGraw Hill, 1982.

7. W. C. Jakes, *Microwave Mobile Communications* , Wiley, 1974.

8. M. K. Simon, "Dual-Pilot Tone Calibration Technique," IEEE Trans. on Veh. Technol., Vol. VT-35, No. 2, pp. 63-70, May 1986.

9. J. P. McGeehan, A. J. Bateman, "Phase-Locked Transparent Tone-in-Band (TTIB): A New Spectrum Configuration Particularly Suited to the Transmission of Data Over SSB Mobile Radio Networks," IEEE Trans. on Comm. Vol. COM-32, No. 1, pp. 81-87, Jan. 1984.

10. J. K. Cavers, "Performance of Tone Calibration With Frequency Offset and Imperfect Pilot Filter," CSS/LCCR TR 88-15, Simon Fraser University, Burnaby, B.C., Canada.

11. J. G. Proakis, J. H. Miller, "An Adaptive Receiver for Digital Signaling Through Channels With Intersymbol Interference," IEEE Trans. on Inform. Th. VOL. IT-15, No. 4, pp. 484-497, July 1969.

12. L. J. Griffiths, "An Adaptive Lattice Structure for Noise-Cancellation Applications," in Proc. IEEE Int. Conf. on ASSP. (Tulsa, Oklahoma), pp. 87-90, 1978.

13. M. L. Honig, D. G. Messerschmitt, *Adaptive Filter: Structures, Algorithms, and Applications*, Kluwer Academic Publishers, 1984.

14. B. Widrow, S. D. Stearns, *Adaptive Signal Processing* , Prentice-Hall, 1985.

15. Simon Haykin, *Adaptive Filter Theory* , Prentice-Hall, 1986.

16. A. V. Oppenheim, R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, 1975.

17. H. Anton, *Elementary Linear Algebra*, 2nd Ed., Wiley & Sons, 1977.

18. IMSL MATH/LIBRARY: FORTRAN Subroutines for Mathematical Applications.

19. J. Makhoul, "A Class of All-Zero Lattice Digital Filters: Properties and Applications," IEEE Trans. on ASSP., vol. ASSP-26, pp. 304-314, Aug. 1978.

20. S. T. Alexander, *Adaptive Signal Processing: Theory and Applications*, Springer-Verlag, 1986.

21. K. S. Trivedi, *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*, Prentice-Hall, 1982.

22. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, 1965.

23. M. J. Gans, "A Power-Spectral Theory of Propagation in the Mobile Radio Environment," IEEE Trans. Veh. Techn. VT-21, pp. 27-38, Feb. 1972.

24. Simon Haykin, *Communication Systems*, 2nd Ed., Wiley & Sons, 1978.