Fall 2008

Multiple-input multiple-output wireless communications with imperfect channel knowledge

Christopher Potter

Department: Electrical and Computer Engineering

Recommended Citation

Potter, Christopher, "Multiple-input multiple-output wireless communications with imperfect channel knowledge" (2008). *Doctoral Dissertations*. 1926.
https://scholarsmine.mst.edu/doctoral_dissertations/1926

MULTIPLE-INPUT MULTIPLE-OUTPUT WIRELESS COMMUNICATIONS WITH IMPERFECT CHANNEL KNOWLEDGE

by

CHRISTOPHER GENE POTTER

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2008

Approved by:

K. Kosbar, Advisor
S. Grant
R. Moss
R. Zheng
I. Morgan

PUBLICATION DISSERTATION OPTION

Pages 4-24 will be submitted for publication in "*Elsevier Journal of Signal Processing*". Pages 25-58 will be submitted for publication in "*IEEE Transactions on Wireless Communications*". Pages 59-74 have been accepted for publication in "*International Telemetry Conference*" in 2008. Pages 75-88 have been accepted for publication in "*International Telemetry Conference*" in 2007. Pages 89-106 have been accepted for publication in "*International Telemetry Conference*" in 2006.

ABSTRACT

In the first work a recurrent neural network (RNN) is employed for MIMO channel prediction. A novel PSO-EA-DEPSO off-line training algorithm is presented and is shown to outperform PSO, PSO-EA, and DEPSO. This predictor is shown to be robust to varying channel scenarios. New expressions for the received SNR, array gain, average probability of error, and diversity gain are derived.

Next, a new expression for the outage capacity of a MIMO system with no CSI at the transmitter and an estimate at the receiver is presented. Since the outage capacity is a function of the first and second moments of the mutual information, new closed form approximations are derived at low and high effective SNR. Also at low effective SNR a new result for the outage capacity is presented. Finally, the outage capacity for a frequency selective channel is derived.

This is followed by a MIMO RNN predictor that operates online. A single RNN is constructed to predict all of the MIMO sub-channels instantaneously. The extended Kalman filter (EKF) and real-time recurrent learning (RTRL) algorithms are applied to compare the MSE of the prediction error.

A new expression for the channel estimation error of a continuously varying MIMO channel is derived next. The optimal amount of time to send training pilots is investigated for different channel scenarios. Special cases of the new expression for the channel estimation error lead to previously established results.

The last work investigates the performance of a MIMO aeronautical system in a two-ray ground reflection scenario. The ergodic capacity is analyzed when the altitude, horizontal displacement, antenna separation, and aircraft velocity are varied.

ACKNOWLEDGMENT

I would like to thank Dr. Kosbar for his colorful discussions, honesty, availability, and intelligence.

I would like to thank my committee for the knowledge I have gained. Each one of you has provided me with a unique challenge to strive and achieve.

A special thanks to Sree Hari Rao Vadrevu for your countless hours of discussion and enthusiasm along with your willingness to critique and edit my work with no questions asked.

Thank you Dr. Venayagamoorthy for inspiring me to take my work to the next level.

To my family. Thanks Mom and Dad for putting me in situations to succeed and always telling me I can do anything I put my mind to. It was from you I developed the drive to strive and be the best. Thanks Jason and Aaron (in no particular order) for keeping things light. Its great to come home and just be a brother. Also to the rest of my relatives. I love you all.

Most importantly I'd like to thank you Jennifer. Without you I simply did not stand a chance. The phone calls at night, the support, the honesty. It was the words you said and how you said them. Other times it was the ones you did not speak of but I understood. *I love you.*

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. SUMMARY OF WORK

In this section a brief overview of each section of the dissertation is presented.

1.1. MIMO BEAM-FORMING WITH NEURAL NETWORK CHANNEL PREDICTION TRAINED BY A NOVEL PSO-EA-DEPSO-ALGORITHM

A new hybrid algorithm based on particle swarm optimization (PSO), evolutionary algorithm (EA), and differential evolution (DE) is presented for training a recurrent neural network (RNN) for multiple-input multiple-output (MIMO) channel prediction. The hybrid algorithm is shown to outperform PSO, DEPSO, and PSO-EA. To explore the effects of channel prediction error at the receiver, new expressions for the received SNR, array gain, average probability of error, and diversity order are derived and analyzed.

1.2. MULTIPLE-INPUT MULTIPLE-OUTPUT RAYLEIGH FADING OUTAGE CAPACITY WITH CHANNEL UNCERTAINTY

New analytical bounds for the outage capacity of a multiple-input multiple-output (MIMO) Rayleigh flat fading wireless system are derived for the case when the channel is unknown at the transmitter while the receiver has access to channel state information (CSI). To aid in the understanding of these bounds, new accurate closed-form approximations for the mean and variance of the mutual information are derived when the system is operating in the low and high effective signal-to-noise ratio (SNR) regimes. Also, at low effective SNR new accurate closed form approximations for the outage capacity bounds are presented. To show how the bounds are affected by a doppler shift, a MIMO wireless system employing pilot symbol assisted modulation (PSAM) is considered. To conclude, the established results for the flat fading case are shown to hold for frequency selective channels.

1.3. MIMO CHANNEL PREDICTION USING RECURRENT NEURAL NETWORKS

Adaptive modulation is a communication technique capable of maximizing throughput while guaranteeing a fixed symbol error rate (SER). However, this technique requires instantaneous channel state information at the transmitter. This can be obtained by predicting channel states at the receiver and feeding them back to the trasnmitter. Existing algorithms used to predict single-input single-output (SISO) channels with recurrent neural networks (RNN) are extended to multiple-input multiple-output (MIMO) channels for use with adaptive modulation and their performance is demonstrated in several examples.

1.4. MODELING CHANNEL ESTIMATION ERROR IN CONTINUOUSLY VARYING MIMO CHANNELS

The accuracy of channel estimation plays a crucial role in the demodulation of data symbols sent across an unknown wireless medium. In this work a new analytical expression for the channel estimation error of a multiple-input multiple-output (MIMO) system is obtained when the wireless medium is continuously changing in the temporal domain. Numerical examples are provided to illustrate our findings.

1.5. SINGLE BOUNCE AIR TO GROUND CHANNEL CAPACITY FOR MIMO SYSTEMS

This paper addresses the air-to-ground communication problem, where multiple transmit antennas are used on the aircraft to combat multi-path interference. The channel is assumed to have a line-of-sight component and a single ground reflection. Multiple-input multiple-output (MIMO) techniques can be used in this situation, to increase the reliability and data rate. In this paper we discuss how the MIMO channel capacity changes with the aircraft antenna configuration, altitude, velocity, range, and a number of other parameters. For comparison, the MIMO results are compared to systems which have single antennas at the transmitter, at the receiver, or at both ends.

2. MIMO BEAM-FORMING WITH NEURAL NETWORK CHANNEL PREDICTION TRAINED BY A NOVEL PSO-EA-DEPSO

2.1. INTRODUCTION

Multiple-input multiple-output (MIMO) systems have been shown to provide significant gains in both spectral efficiency and reliability [1]. These results, however, are based on the assumption that the transmitter and/or receiver have perfect knowledge of the channel state information (CSI). An alternative to this is to estimate the channel at the receiver and send the channel state information (CSI) back to the transmitter. This approach suffers from outdated CSI upon arrival at the transmitter when the channel varies with time. Performance can be improved in this case by sending back a *prediction* of the CSI.

Unlike the use of classical prediction techniques [2], a recurrent neural network (RNN) is proposed for prediction. In [3] an extended Kalman filter (EKF) was employed for training a RNN for time series prediction. A hybrid particle swarm optimization evolutionary algorithm (PSO-EA) was utilized in [4] for time series. In this work, a new hybrid algorithm composed of PSO, EA, and differential evolution (DE) is proposed for MIMO channel prediction. It is shown that this hybrid algorithm outperforms PSO, DEPSO, PSO-EA, and a Levinson-Durbin linear predictor. New expressions for the received SNR, array gain, average probability of error, and diversity order are derived to analyze the performance degradation due to prediction error.

The rest of this paper is organized as follows. The next section describes the MIMO beam-forming input/output model. This is followed by the channel model in Section 3. After this the RNN used for MIMO channel prediction is introduced. Section 5 provides a brief review of related off-line training algorithms and the proposal of a novel PSO-EA-DE hybrid training algorithm. The training results for the different training algorithms are presented next. The received SNR, array gain, average probability of error, and diversity order are derived and analyzed in Section 7 which is followed by the concluding remarks.

2.2. MIMO RECEIVED MODEL

A MIMO wireless flat fading baseband communication system with N_t transmit antennas and N_r receive antennas is modeled at discrete time k by

$$\boldsymbol{y}(k) = \boldsymbol{H}(k)\boldsymbol{x}(k) + \boldsymbol{n}(k),$$

where $\boldsymbol{y}(k)$ is the $N_r \times 1$ received vector, $\boldsymbol{x}(k)$ is the $N_t \times 1$ transmitted symbol vector with $x_i(k), i = 1, \ldots, N_t$, belonging to constellation $\mathcal{C}$, and $\boldsymbol{n}(k)$ is the white noise vector of size $N_r \times 1$ with $n_i \overset{iid}{\sim} \mathcal{CN}(0, N_o)^1, i = 1, \ldots, N_r$. The $N_r \times N_t$ channel matrix $\boldsymbol{H}(k) = \{h_{mn}(k)\}$ describes the complex channel gain between the m^{th} receiver antenna and n^{th} transmit antenna. Denoting P_t as the total transmit power, the transmitted symbols must satisfy

$$\mathbb{E}||\boldsymbol{x}(k)||_2^2 = P_t \tag{2.1}$$

The beam-formed symbols are expressed in two scenarios, when the transmitter and receiver have full channel state information (CSI) and when they only have access to the prediction.

Let $\boldsymbol{H}(k) = \boldsymbol{U}(k)\boldsymbol{D}(k)\boldsymbol{V}(k)^H$ be the singular value decomposition (SVD) of the MIMO channel matrix. The i^{th} column of the unitary matrices $\boldsymbol{U}(k)$ and $\boldsymbol{V}(k)$, denoted by $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ respectively, are the left and right singular vectors corresponding to the i^{th} singular value $\sigma_{\boldsymbol{H}_i}(k)$ satisfying $0 \leq \sigma_{\boldsymbol{H}_{i-1}}(k) \leq \sigma_{\boldsymbol{H}_i}(k) \leq \sigma_{\boldsymbol{H}_{i+1}}(k) \leq \sigma_{\boldsymbol{H}_{r(\boldsymbol{H}(k))}}(k)$, where $r(\boldsymbol{H}(k)) = \text{rank}(\boldsymbol{H}(k))$. If $\tilde{\boldsymbol{x}}(k) = \boldsymbol{v}_1(k)x(k)$ the received beam-formed symbol is

$$\boldsymbol{u}_1(k)^H\boldsymbol{y}(k) = \sigma_{\boldsymbol{H}_1}(k)x(k) + \boldsymbol{u}_1(k)^H\boldsymbol{n}(k).$$

---

[1] This notation stands for a complex normal random variable with mean zero and variance N_o.

When the transmitter and receiver only have a prediction of the channel matrix $\widehat{\boldsymbol{H}}(k) = \widehat{\boldsymbol{U}}(k)\widehat{\boldsymbol{D}}(k)\widehat{\boldsymbol{V}}(k)^H$, the received symbols are

$$\widehat{\boldsymbol{u}}_1(k)^H\boldsymbol{y}(k) = \widehat{\boldsymbol{u}}_1(k)^H\boldsymbol{U}(k)\boldsymbol{D}(k)\boldsymbol{V}(k)^H\widehat{\boldsymbol{v}}_1(k)x(k) + \widehat{\boldsymbol{u}}_1(k)^H\boldsymbol{n}(k).$$

To supply the transmitter with the dominant left singular vector a feedback link must be established. Let t_d be defined as the total delay arising from processing and transmission latencies. To prevent the CSI from becoming stale [5]

$$t_d \ll \frac{.423}{f_d},$$

where f_d is the maximum doppler frequency. This seems discouraging since for a maximum doppler frequency of 100 Hz the total delay must be less than 4.23 ms. However, the accurate estimation of the doppler frequency has been a well researched topic [6, 7]. Assuming a worst case approximation error of .5 Hz, the maximum delay increases to .846 sec which will be adequate for most wireless latencies [8].

2.3. CHANNEL MODEL

The MIMO sub-channels are represented by [9]

$$g_{mn}(k) = f\Big(g^I_{mn}(k) + jg^Q_{mn}(k)\Big),$$

where $f : \mathbb{C} \to \mathbb{C}$ is bounded,

$$g^I_{mn}(k) = \sqrt{\frac{2}{M}}\sum_{n=1}^{M}\cos(2\pi f_d k T_s \cos(\alpha_n) + \phi_n)$$

is the in-phase component,

$$g^Q_{mn}(k) = \sqrt{\frac{2}{M}}\sum_{n=1}^{M}\cos(2\pi f_d k T_s \sin(\alpha_n) + \psi_n)$$

is the quadrature component, T_s is the sampling period, and

$$\alpha_n = \frac{2\pi n - \pi + \theta}{4M}.$$

The parameters ϕ_n, ψ_n, and θ are $U[-\pi, \pi)$, where $U[a, b]$ is a uniform random variable between a and b. To accommodate the possibility of spatial correlation between both transmit and receive antennas we model the overall channel by

$$\boldsymbol{H}(k) = \boldsymbol{\Phi}_{Rx}(k)^{1/2}\boldsymbol{G}(k)\boldsymbol{\Phi}_{Tx}(k)^{1/2}, \tag{2.2}$$

where

$$\begin{aligned} \boldsymbol{\Phi}_{Tx} &\triangleq \frac{1}{N_r}\sum_{r=1}^{N_r} \mathbb{E}\{\boldsymbol{h}_{r,\cdot}(k)^H \boldsymbol{h}_{r,\cdot}(k)\} \\ \boldsymbol{\Phi}_{Rx} &\triangleq \frac{1}{N_t}\sum_{r=1}^{N_t} \mathbb{E}\{\boldsymbol{h}_{r}(k) \boldsymbol{h}_{r}^H(k)\} \end{aligned}$$

with $\boldsymbol{h}_{i,\cdot}$ denoting the i^{th} row of $\boldsymbol{H}(k)$. These correlation matrices are dependent on many parameters, including the angle of arrival, transmit and receiver antenna distances, and angle spread [10]. Since the focus of this work will not be emphasized on any particular antenna geometry, we will maintain spatial generality and consider a simple Toeplitz stucture [11]

$$\boldsymbol{\Phi}_{Tx} = \begin{bmatrix} 1 & \gamma_t & \gamma_t^4 & \cdots & \gamma_t^{(N_t-1)^2} \\ \gamma_t & 1 & \gamma_t & \ddots & \vdots \\ \gamma_t^4 & \gamma_t & 1 & \ddots & \gamma_t^4 \\ \vdots & \ddots & \ddots & \ddots & \gamma_t \\ \gamma_t^{(N_t-1)^2} & \cdots & \gamma_t^4 & \gamma_t & 1 \end{bmatrix}$$

$$\boldsymbol{\Phi}_{Rx} = \begin{bmatrix} 1 & \gamma_r & \gamma_r^4 & \dots & \gamma_r^{(N_r-1)^2} \\ \gamma_r & 1 & \gamma_r & \ddots & \vdots \\ \gamma_r^4 & \gamma_r & 1 & \ddots & \gamma_r^4 \\ \vdots & \ddots & \ddots & \ddots & \gamma_r \\ \gamma_r^{(N_r-1)^2} & \dots & \gamma_r^4 & \gamma_r & 1 \end{bmatrix}.$$

It should be noted that the simplicity of this model does not prevent accurate approximations for a variety of antenna configurations.

2.4. CHANNEL ESTIMATION

For a reasonable prediction it is necessary for the RNN to learn the statistics of the fading process. Following the same procedure in [12], [13], and [14] the channel is written in terms of its minimum mean squared error (MMSE) estimate $\widetilde{\boldsymbol{H}}(k)$ by

$$\widetilde{\boldsymbol{H}}(k) = \boldsymbol{H}(k) + \boldsymbol{W}(k),$$

where $\widetilde{\boldsymbol{H}}(k)$ and $\boldsymbol{W}(k)$ are uncorrelated with $\widetilde{h}_{mn}(k) \sim \mathcal{CN}(0, \sigma_{\widetilde{h}}(k)^2)$ and $w_{mn}(k) \sim \mathcal{CN}(0, \sigma_w(k)^2)$. The channel estimation error between the m^{th} receiver and n^{th} transmitter is thus

$$\sigma_w^2(k) = |\sigma_h^2(k) - \sigma_{\widetilde{h}}^2(k)|.$$

2.5. A RECURRENT NEURAL NETWORK FOR CHANNEL PREDICTION

The recurrent neural network used for prediction is shown in Figure 2.1. The output of the activation functions are

$$d_j(k) = \phi_j\left(\sum_{i=1}^{N_p+m} a_{ji} s_i(k)\right), \quad j = 1, \dots, m,$$

where

$$\boldsymbol{s}(k) = [d_1(k-1) \cdots d_m(k-1), \widetilde{h}_{mn}(k) \; \dots \; \widetilde{h}_{mn}(k - N_p)]^T.$$

is the RNN input and

$$\phi_j(x) = \tanh(x), \quad j = 1, \ldots, m$$

are the non-linear activation functions. The RNN predictor is implemented in two stages. During the first $d_1(k-1), \ldots, d_m(k-1) = 0$. The outputs are then fed back to form $\boldsymbol{s}(k)$. Finally the output of the second stage $d_1(k)$ is the RNN prediction. Unless otherwise stated, $N_p = 5$ and $m = 2$ will be chosen.[2]

2.6. TRAINING ALGORITHMS FOR RNN PREDICTOR

Before directly proceeding to the proposed hybrid algorithm PSO, EA, PSO-EA, and DEPSO are briefly summarized.

2.6.1. PSO.

PSO is a evolutionary computation technique developed by Kennedy and Eberhart in 1995 [15]. It was inspired by swarm intelligence where a collection of unsophisticated individuals (particles) can solve complex problems by interacting with one another. Some examples of this behavior are a flock of birds or a school of fish. PSO has been shown to be an algorithm that when used properly can perform multi-parameter optimization [16]. Some applications of PSO include artificial life, social psychology, engineering, and computer science.

Let P be the number of particles in the swarm and D the number of particle dimensions. At each epoch the velocity and positions for the i^{th} particle and d^{th} dimension are respectively governed by

$$\begin{aligned} v_{id} &= wv_{id} + c_1U(0,1)(pbest_{id} - x_{id}) + c_2U(0,1)(gbest_d - x_{id}) \qquad (2.3) \\ x_{id} &= x_{id} + v_{id}. \end{aligned}$$

The parameter w is the inertia constant while c_1 and c_2 are the cognitive and social acceleration constants respectively.

---

[2] This value of m was found by simulation to reduce the complexity of the RNN while producing satisfactory results.

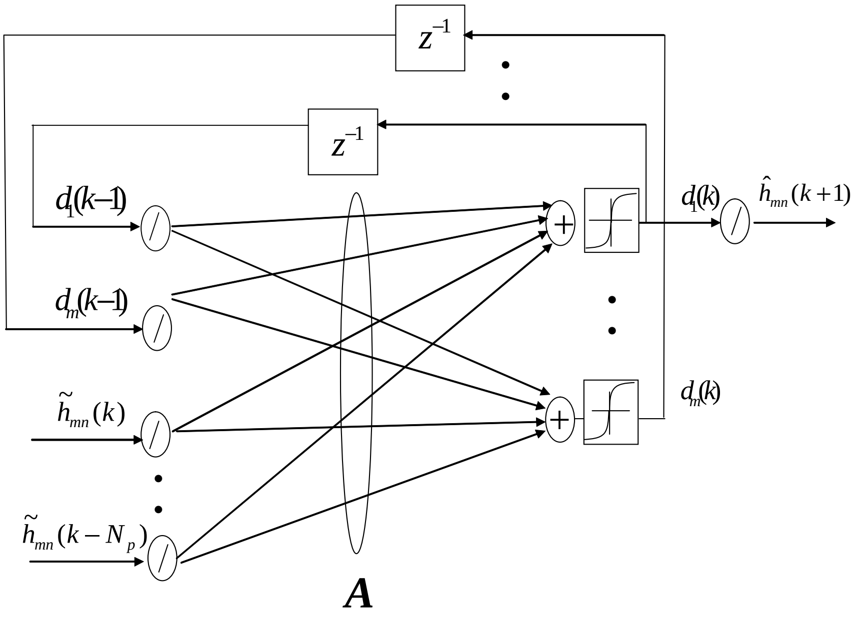

Figure 2.1 RNN for channel prediction.

2.6.2. EA. Each parent or offspring in the EA algorithm is represented as a chromosome made up of genes. Given a population N of chromosomes, for every generation each parent is described by its genes $\boldsymbol{w}_i$ and self adaptive parameter $\boldsymbol{\beta}_i, i = 1, \ldots, N$, both of which are $N_w \times 1$ vectors. Each parent/chromosome generates an offspring $\acute{P}_i$ with $\acute{\boldsymbol{\beta}}_i$ and $\acute{\boldsymbol{w}}_i$ updated according to [4]

$$\acute{\beta}_{ij} = \beta_{ij} \exp(\tau \mathcal{N}(0,1)) \tag{2.4}$$

$$\acute{w}_{ij} = w_{ij} + \acute{\beta}_{ij} \mathcal{N}(0,1), \tag{2.5}$$

where $\tau = 1/\sqrt{2\sqrt{N_w}}$. The parents are then put through a selection process where the losers are eliminated and replaced by the winners offspring.

2.6.3. PSO-EA. The hybrid PSO-EA algorithm combines the swarm behavior with increased diversity through mutation of the "fitter" particles. For every iteration, the particles with the lowest fitness values are discarded and the survivors produce offspring with velocities and positions calculated by (2.4) and (2.5) respectively. For more details, the reader is referred to [4].

2.6.4. DEPSO. DEPSO is a hybrid of DE and PSO which provides diversity on the population while keeping the swarm searching capabilities intact [17]. Letting $\mathcal{P} = \{1, \ldots, P\}$ be the swarm alphabet, the *pbest* of each particle is updated by

$$\begin{aligned}&\textbf{IF}(\text{rand}(\cdot) < P_c)\textbf{OR}(i == \xi)\\&\textbf{THEN}\, pbest_{id} = gbest_d + \Delta_L\end{aligned}$$

where

$$\Delta_L = \frac{1}{L}\sum_{j=1}^{L} pbest_{Aj} - pbest_{Bj}$$

and $\xi, A, B \in \mathcal{P}$ are randomly chosen.

2.6.5. Hybrid PSO-EA-DEPSO Algorithm. In this work, a new algorithm is proposed that is a hybrid version of PSO, EA, and DEPSO. The block diagram is displayed in Figure 2.2. The idea behind the algorithm is to alternate between PSO, EA, and DEPSO to continually provide diversity for the particles/parents. This in theory prevents the particles/parents from reaching a premature convergence. The PSO algorithm is implemented for one fourth of the total iterations to converge quickly on the potential solution. The PSO-EA and DEPSO algorithms then alternate for the remaining iterations. For PSO, the velocity, position, and p_{best} values are all initialized from a uniform distribution.

2.7. TRAINING RESULTS

The RNN in Figure 2.1 is trained off-line with data generated from (2.2). This is less restrictive than online training [3, 18, 19] where it was assumed therein that the instantaneous value of the error (and hence full channel knowledge) is known at the receiver. We also propose a more robust training method than in [20], where the RNN weights had to be retrained whenever the channel changed. The fitness function is defined to be

$$MSE(k) \triangleq \frac{1}{N_{train}} \sum_{k=1}^{N_{train}} \left(h^A_{mn}(k) - \widehat{h}^A_{mn}(k)\right)^2 \tag{2.6}$$

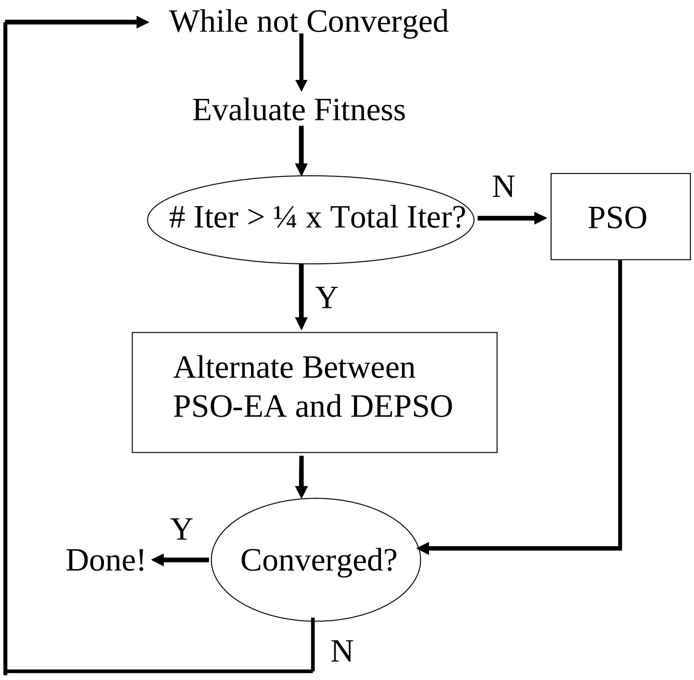

Figure 2.2 Hybrid PSO-EA-DEPSO training algorithm.

where $A \in \{I, Q\}$ and N_{train} is the number of training samples. When a specified number of iterations have occurred or the fitness function has reached a desired value, the RNN weights are frozen and brought online.

To provide a comparison of the training algorithms described in the previous section, a 2×2 spatially uncorrelated MIMO channel with $f(x) = x$, $\gamma_t = \gamma_r = 0, \sigma_w^2(k) = .001$, and normalized doppler frequency $f_d T_s = 0.05$ is predicted using PSO, DEPSO, PSO-EA, and the new hybrid PSO-EA-DEPSO algorithm. Since the training algorithms use a stochastic search for the optimum RNN weights, convergence in the mean squared error is investigated. The experiment consists of fifty independent trials. For each training algorithm, the mean of the lowest value of (2.6) that has currently been found for $h_{11}(k)$ is plotted versus the number of iterations in Figures 2.3 and 2.4. For the first fifty iterations the proposed hybrid algorithm is operating in PSO mode and thus the performance is comparable to that of PSO. For the remaining iterations the hybrid algorithm uses its diversity to outperform the competition.

Although accurate training of the weights is important, the robustness to different channel conditions is also critical for the RNN predictor. To investigate this the weights are trained with the new hybrid PSO-EA-DEPSO training algorithm when $f(x) = x, \gamma_t = \gamma_r = 0, \sigma_w^2(k) = 0.01$ and $f_d T_s = 0.1$ and are brought online to predict a channel with a time varying $f_d T_s$. The in-phase and quadrature channel coefficients along with their predictions are illustrated in Figures 2.5 and 2.6 respectively. The accuracy of these predictions verify the robustness of the RNN equipped with the new PSO-EA-DEPSO training algorithm.

The RNN using the new hybrid algorithm is compared to a fifth order feed-forward linear predictor using the Levinson-Durbin Algorithm [21] for $f(x) = \exp(x)x, f_d T_s = 0.5$ and $\sigma_w^2(k) = 0.01$. Observing Figures 2.7 and 2.8 the RNN outperforms the linear predictor. This example suggests the RNN trained with the PSO-EA-DEPSO algorithm is capable of predicting certain non-linear, non-stationary channels better than the Levinson-Durbin linear predictor.

2.8. PERFORMANCE MEASURES OF RNN PREDICTOR

Now that good training performance has been established, the impact of prediction error at the receiver is investigated. In this section, new expressions for the received SNR, array gain, probability of error, and diversity gain are derived for a MIMO RNN channel predictor. All of these measures are dependent on the received SNR which will be the starting point of the derivation. For convenience the discrete time dependence will become implicit.

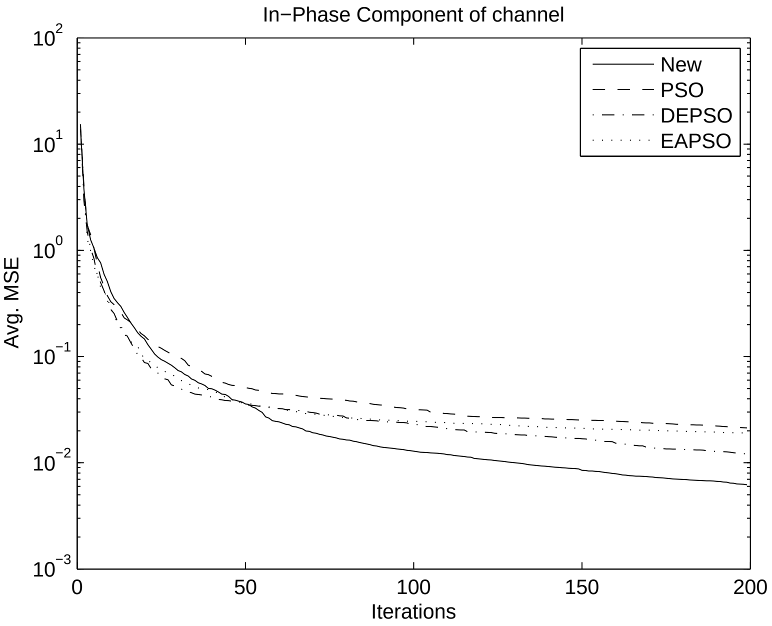

Figure 2.3 Mean squared error comparison for several training algorithms of in-phase component when $f_dT_s = 0.05$.

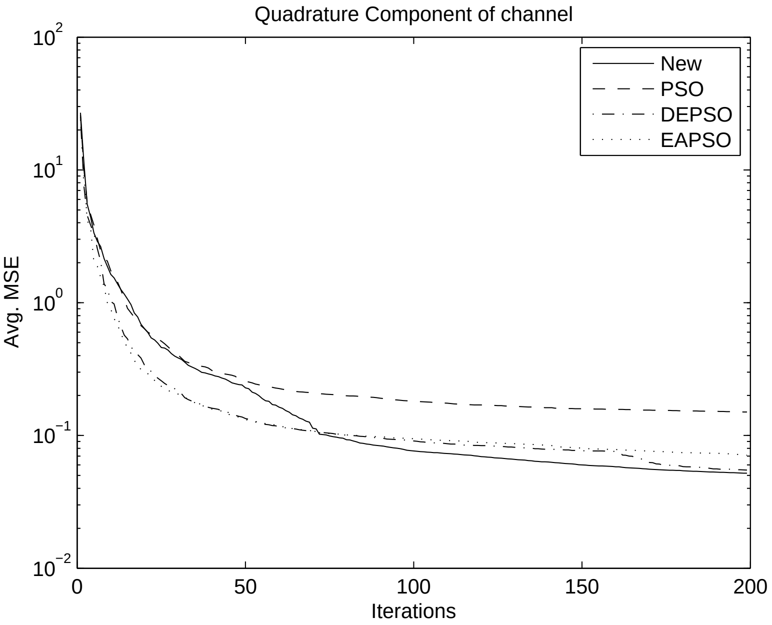

Figure 2.4 Mean squared error comparison for several training algorithms of quadrature component when $f_dT_s = 0.05$.

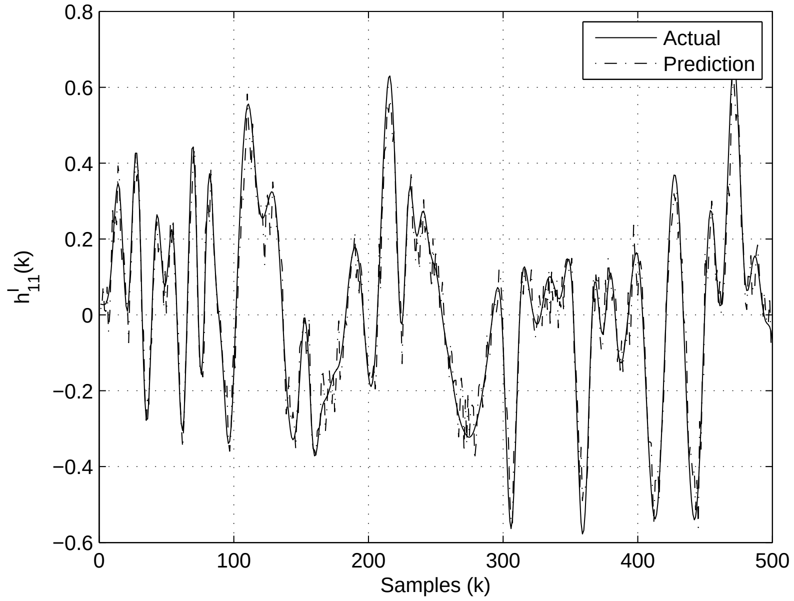

Figure 2.5 Comparison of new hybrid PSO-EA-DEPSO with the actual in-phase channel coefficients for varying f_dT_s.

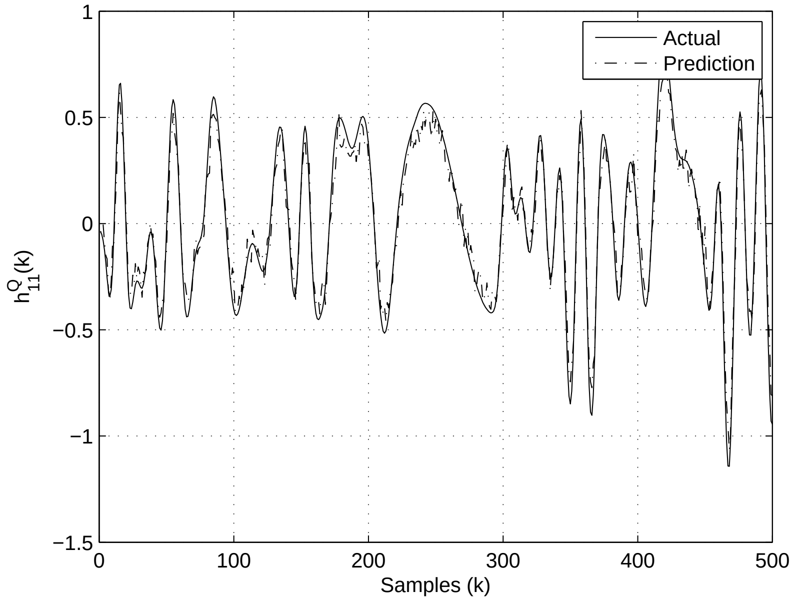

Figure 2.6 Comparison of new hybrid PSO-EA-DEPSO with the actual quadrature channel coefficients for varying f_dT_s.

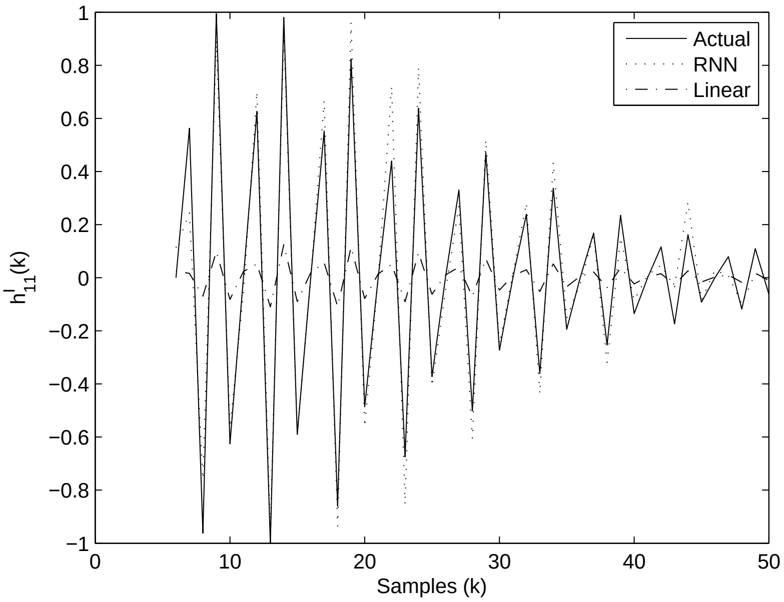

Figure 2.7 MSE of RNN predictor.

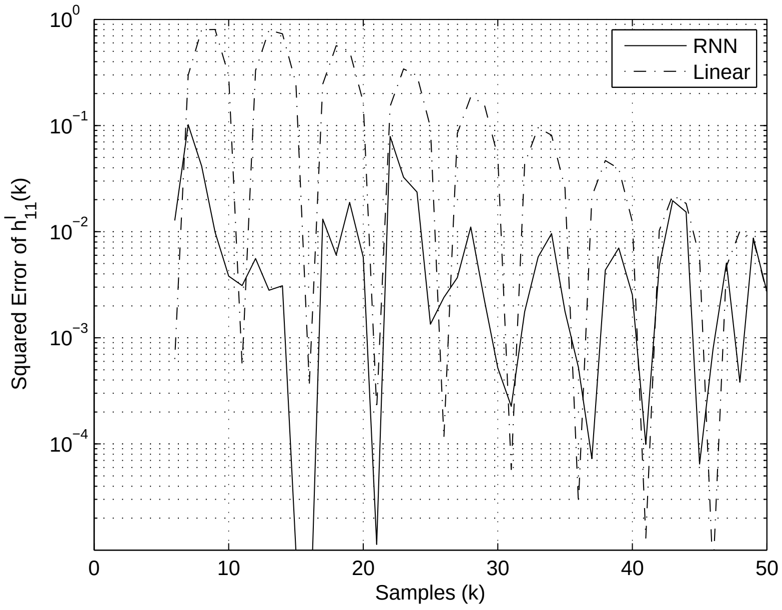

Figure 2.8 MSE of linear predictor.

2.8.1. Received SNR. The instantaneous received SNR for the MIMO beamformer when the receiver knows the channel is defined as

$$\rho_{bf} \triangleq \rho \sigma^2_{\boldsymbol{H}_1}.$$

where $\rho \triangleq \frac{E_s}{N_o}$ is the SNR of a SISO AWGN channel (see (**??**)). Since the RNN is equipped with non-linear activation functions, the exact distribution of the predictor is difficult and cumbersome. Writing $\widehat{\boldsymbol{v}_1} = \boldsymbol{v}_1 + \Delta \boldsymbol{v}_1$ and $\widehat{\boldsymbol{u}}_1 = \boldsymbol{u}_1 + \Delta \boldsymbol{u}_1$ the received SNR for the MIMO RNN predictor is

$$\rho_{bf} = \frac{\sigma^2_{\boldsymbol{H}_1}}{\beta + \rho^{-1}}, \tag{2.7}$$

where

$$\beta \triangleq \left| \sigma_{\boldsymbol{H}_1} \boldsymbol{v}_1^H \Delta \boldsymbol{v}_1 + \sigma_{\boldsymbol{H}_1} \Delta \boldsymbol{u}_1^H \boldsymbol{u}_1 + \Delta \boldsymbol{u}_1^H \boldsymbol{H} \Delta \boldsymbol{v}_1 \right|^2. \tag{2.8}$$

2.8.2. Array Gain. The array gain measures the increase in average received SNR compared to a SISO system and is defined as

$$\eta \triangleq \frac{\mathbb{E}\{\rho_{bf}\}}{\rho}.$$

For a MIMO RNN predictor, the array gain becomes

$$\eta = \frac{\mathbb{E}\{\sigma^2_{\boldsymbol{H}_1}\}}{\rho \mathbb{E}\{\beta\} + 1} \tag{2.9}$$

which indicates the prediction error will have a negative influence on the array gain with increasing SNR. In other words, the prediction error introduces an irreducible noise floor. To analyze the increase in performance note that

$$\frac{\|\boldsymbol{H}\|_F^2}{r(\boldsymbol{H})} \leq \sigma^2_{\boldsymbol{H}_1} \leq \|\boldsymbol{H}\|_F^2, \tag{2.10}$$

where $\|\cdot\|_F$ is the Frobenius norm [22]. Let $\boldsymbol{\mathcal{H}} = \text{vec}(\boldsymbol{H})$, where $\text{vec}(\cdot)$ stacks the columns of its argument from left to right. Then [23]

$$\boldsymbol{\mathcal{H}} = \left(\boldsymbol{\Phi}_{Tx}^{T/2} \otimes \boldsymbol{\Phi}_{Rx}^{1/2}\right)\text{vec}(\boldsymbol{G})$$

where $\otimes$ denotes the kronecker product. It follows after several manipulations and various kronecker product properties that

$$\mathbb{E}\|\boldsymbol{H}\|_F^2 = \mathbb{E}\|\boldsymbol{\mathcal{H}}\|_F^2 = \text{trace}(\boldsymbol{\Phi}_{Tx}^2)\text{trace}(\boldsymbol{\Phi}_{Rx}^2).$$

Therefore

$$\frac{\text{trace}(\boldsymbol{\Phi}_{Tx}^2)\text{trace}(\boldsymbol{\Phi}_{Rx}^2)}{r(\boldsymbol{H})} \leq \mathbb{E}\{\sigma_{\boldsymbol{H}_1}^2\} \leq \text{trace}(\boldsymbol{\Phi}_{Tx}^2)\text{trace}(\boldsymbol{\Phi}_{Rx}^2)$$

which after substitution into (2.9) yields the following new bounds for the array gain

$$\frac{\text{trace}(\boldsymbol{\Phi}_{Tx}^2)\text{trace}(\boldsymbol{\Phi}_{Rx}^2)}{r(\boldsymbol{H})(\rho\mathbb{E}\{\beta\}+1)} \leq \eta \leq \frac{\text{trace}(\boldsymbol{\Phi}_{Tx}^2)\text{trace}(\boldsymbol{\Phi}_{Rx}^2)}{\rho\mathbb{E}\{\beta\}+1}.$$

Although the array gain is potentially higher for spatially correlated channels, when exact knowledge of the correlation matrices is not available, we can expect the array gain to fall off more dramatically than the spatially uncorrelated case. This will occur if the RNN is trained using a spatially uncorrelated channel when in fact the antennas are spatially dependent.

To support these observations the array gain is plotted for a 2 × 2 MIMO system with $f(x) = x, f_dT_s = 0.1$ for the uncorrelated case in Figure 2.9 and with $\gamma_t = \gamma_r = 0.7$ in Figure 2.10 for various channel estimation errors. For both cases, the RNN was trained using a spatially uncorrelated channel with $f_dT_s = 0.1$ whose estimation error matched the online channel of interest. As discussed, the array gain is larger for the spatial correlation case but deteriorates faster.

2.8.3. Average Probability of Error. The probability of error was approximated for SISO systems in [24] by

$$P_e \approx LQ\left(\sqrt{\frac{\rho d^2}{2}}\right), \tag{2.11}$$

where $Q(\cdot)$ is the Gaussian Q function [25] while L and d are respectively the number of nearest neighbors and minimum Euclidean distance in the normalized constellation[3]. Inserting (2.7) into (2.11), the average probability of error for a MIMO system is

$$\mathbb{E}\{P_e\} \approx L\mathbb{E}\left\{Q\left(\sqrt{\frac{\sigma^2_{\boldsymbol{H}_1} d^2}{2(\beta+\rho^{-1})}}\right)\right\}. \tag{2.12}$$

Observing this expression, the average probability will behave similar to the perfect CSI case until the noise floor becomes saturated by β.

2.8.4. Diversity Order. The diversity order of a system is the slope of the average probability of error at high SNR. Applying the Chernoff bound [25] and (2.10) to (2.12) the probability of error is bounded by

$$L\exp\left(-\frac{\sigma^2_{\boldsymbol{H}_1} d^2}{4\left(\beta+\rho^{-1}\right)}\right) \leq P_e \leq L\exp\left(-\frac{\sigma^2_{\boldsymbol{H}_1} d^2}{4r(\boldsymbol{H})\left(\beta+\rho^{-1}\right)}\right).$$

---

[3] This is just $\mathcal{C}$ normalized such that the average symbol energy is unity.

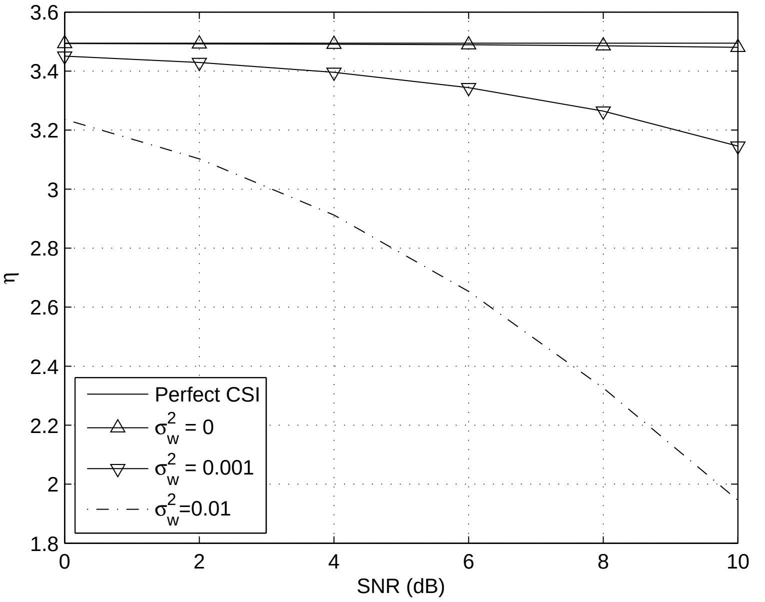

Figure 2.9 Array gain for a 2 × 2 MIMO beam-forming fast fading uncorrelated channel.

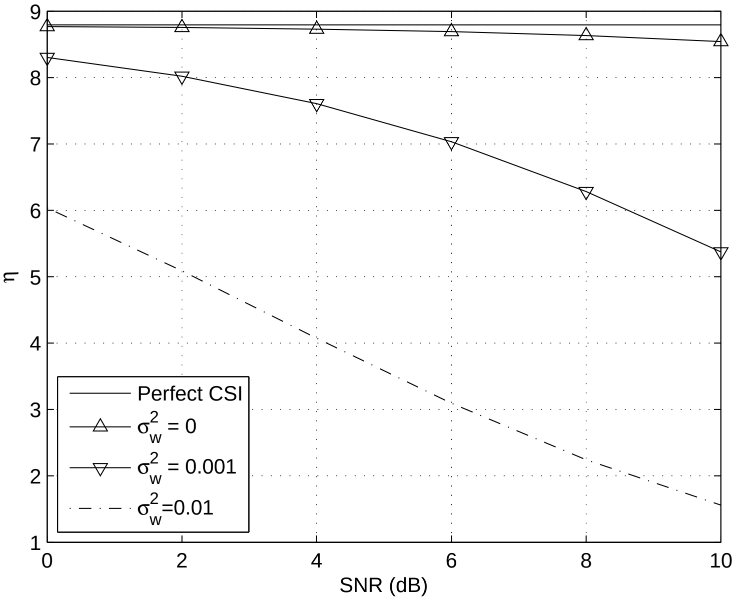

Figure 2.10 Array gain for a 2 × 2 MIMO beam-forming fast fading spatially correlated channel.

Applying the moment generating function in [26] it follows after several manipulations that the bounds on the average probability of error $\mathbb{E}\{P_e\}$ at high SNR are

$$L\left(\frac{d^2}{4\Big(\beta+\rho^{-1}\Big)}\right)^{-r(\boldsymbol{R})}\prod_{i=1}^{r(\boldsymbol{R})}(\lambda_{\boldsymbol{R}_i})^{-1}\leq\mathbb{E}\{P_e\}\leq L\left(\frac{d^2}{4r(\boldsymbol{H})\Big(\beta+\rho^{-1}\Big)}\right)^{-r(\boldsymbol{R})}\prod_{i=1}^{r(\boldsymbol{R})}(\lambda_{\boldsymbol{R}_i})^{-1}$$

where $\boldsymbol{R}\triangleq\mathbb{E}\{\text{vec}(\boldsymbol{H}\boldsymbol{H}^H)\}$. To find the rank of this matrix, it follows using the same techniques when deriving the array gain that

$$\boldsymbol{R}=\big(\boldsymbol{\Phi}_{Tx}^T\otimes\boldsymbol{\Phi}_{Rx}\big).$$

Using the rank property of kronecker products

$$r(\boldsymbol{R})=r(\boldsymbol{\Phi}_{Tx})r(\boldsymbol{\Phi}_{Rx}). \tag{2.13}$$

This result indicates that the diversity order depends on both the spatial correlation and channel prediction error. Unlike the array gain, the diversity order is larger when the antennas are spatially uncorrelated [26]. As long as the prediction error does not dominate the noise floor the diversity order will match the perfect CSI case.

The average probability of error was simulated for a 2 × 2 MIMO system when $f(x)=x, f_dT_s=0.1, \gamma_t=\gamma_r=0$ using 10^5 BPSK symbols with symbol period $T=T_s$ for various channel estimation errors in Figure 2.11 and with $\gamma_t=\gamma_r=0.7$ in Figure 2.12. For both cases, the RNN was trained with $f(x)=x, \gamma_t=\gamma_r=0$, and $f_dT_s=0.1$ whose estimation error matched the online channel of interest. When the antennas are spatially uncorrelated the diversity order is clearly larger as discussed earlier. In both configurations when the channel estimation error increases, the average probability of error begins to saturate due to the irreducible noise floor caused by the prediction error. To show the accuracy of the

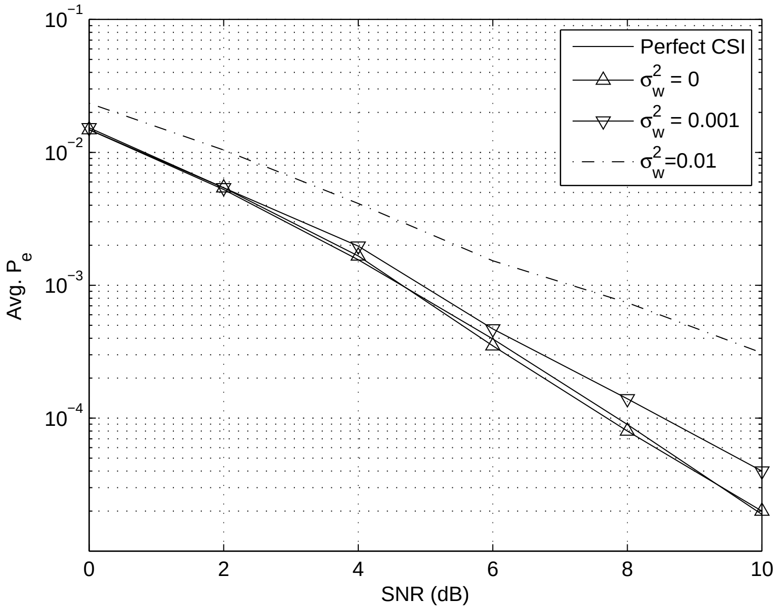

Figure 2.11 BER comparison for a 2 × 2 MIMO beam-forming fast fading spatially uncorrelated channel.

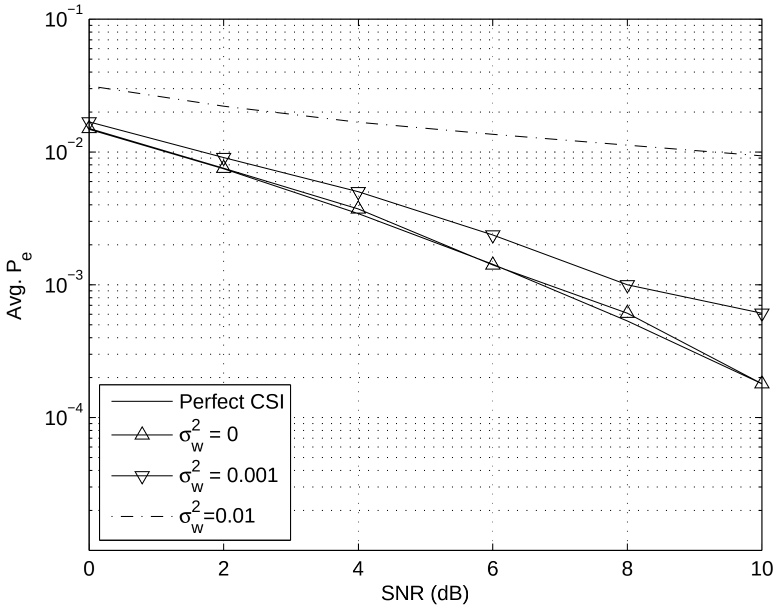

Figure 2.12 BER comparison for a 2 × 2 MIMO beam-forming fast fading spatially correlated channel.

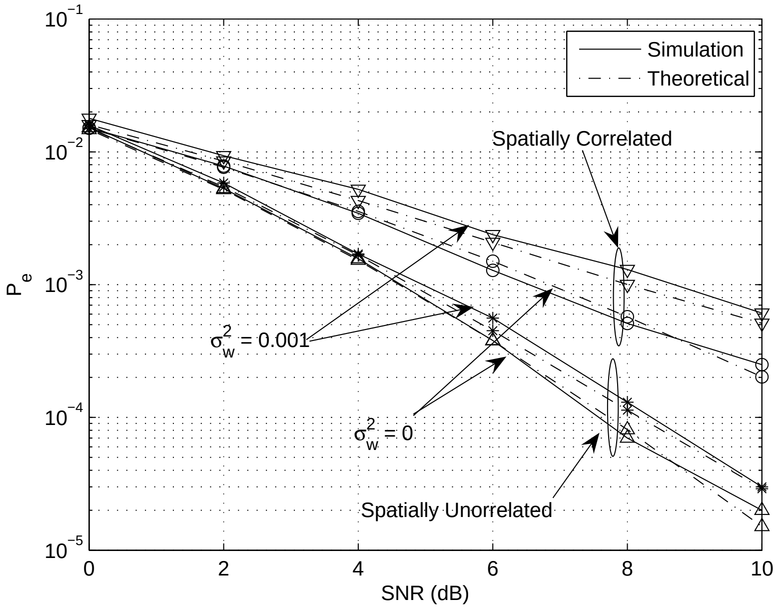

Figure 2.13 BER comparison for a 2 × 2 MIMO beam-forming fast fading spatially correlated channel.

new analytic expression for the average probability of error, (2.12) is plotted versus the simulated results in Figure 2.13 and is seen to be in excellent agreement.

2.9. CONCLUSION

A recursive neural network trained off-line by a novel PSO-EA-DEPSO was used to predict a MIMO channel. This training algorithm was shown to be superior to PSO, PSO-EA, and DEPSO for different fast fading scenarios. The RNN predictor was then shown to outperform a linear predictor trained by the Levinson-Durbin algorithm. New expressions for the received SNR, array gain, average probability of error, and diversity order for the MIMO RNN predictor were then derived. The array gain for spatially correlated systems was shown to be higher and decay sharper than spatially uncorrelated systems. The diversity order for the RNN predictor was shown to equal the perfect CSI case for small estimation errors and gradually worsen.

The application of the hybrid PSO-EA-DEPSO algorithm to frequency selective channels remains to be investigated. Also, the use of perturbation theory to analyze the degradation in performance due to prediction error is being explored.

3. MULTIPLE-INPUT MULTIPLE-OUTPUT RAYLEIGH FADING OUTAGE CAPACITY WITH CHANNEL UNCERTAINTY

3.1. INTRODUCTION

To achieve the capacity promised in Shannon's channel coding theorem, one may need to use codewords which grow exponentially with respect to the uncertainty of the channel. When finite decoding constraints are placed on a wireless receiver that is subject to deep fades, the ergodic capacity may not be the best measure of system potential. In particular, it was shown in [27, 28] that when the decoding time is constrained, there exists a non-zero probability (independent of the code length and consequently the data rate) for which the probability of error did not exponentially decrease with the code length, thus having an ergodic capacity of zero.

An alternative measure that was first considered in [29] is the outage capacity, which is defined to be the maximum achievable rate that can be maintained subject to an outage probability. This definition can be understood as follows. Since the transmitter does not know the channel it must choose an achievable code with a data rate independent of the instantaneous received SNR and hence capacity. When the data rate is lower than the capacity there will be no errors. Otherwise, some bits may be decoded incorrectly, for which case an outage event is declared.

The channel coding theorem also assumes that the receiver has perfect knowledge of channel state information (CSI). In time-varying mobile fading environments, this assumption may not be reasonable, and one should include the resulting decrease in capacity calculations. This was considered for a SISO system in [30], where upper and lower bounds for the instantaneous mutual information were obtained under imperfect CSI. These bounds were used in [31] to develop upper and lower bounds on the outage capacity of a SISO Rayleigh flat fading channel.

Multiple-input multiple-output (MIMO) systems have been shown to provide an increase in capacity while maintaining the same operating bandwidth [1]. In this work the results of the SISO case in [31] are extended to arrive at new analytical bounds for the outage capacity of a Rayleigh fading MIMO system under channel uncertainty. Since these bounds are a function of the mean and variance of the instantaneous mutual information, the exact values for these moments are derived. To gain insight into these expressions, new accurate closed form approximations of the mean and variance for systems operating in the low and high effective signal-to-noise ratio (SNR) regimes are derived. Also, new accurate closed form approximations of the outage capacity are found at low effective SNR. As an application the bounds are analyzed for a MIMO system using pilot symbol assisted modulation (PSAM). To conclude, these results are shown to hold for frequency selective channels.

The rest of this work is organized is follows. The next section mathematically describes the received symbol vectors as well as the MIMO channel model. We end the section with previously attained upper and lower bounds for the instantaneous mutual information. In Section 3.3 we derive new analytic bounds for the outage capacity of a MIMO Rayleigh flat fading system and derive accurate approximations for systems operating in the high and low effective SNR regimes. The analytic bounds are then applied to PSAM for channel estimation in Section 3.4. The outage capacity for a MIMO frequency selective channel is considered next in Section 3.5. This is followed by our concluding remarks in Section 3.6.

3.2. SYSTEM MODELS/PREVIOUS RESULTS

A MIMO baseband system with N_t transmit antennas and N_r receive antennas at discrete time η is described by

$$\boldsymbol{y}(\eta) = \sum_{i=0}^{L-1} \boldsymbol{H}(\eta, i)\boldsymbol{x}(\eta - i) + \boldsymbol{n}(\eta) \tag{3.1}$$

where $\boldsymbol{y}(\eta)$ is the $N_r \times 1$ complex valued received symbol vector, $\boldsymbol{x}(\eta)$ is the $N_t \times 1$ transmitted vector with $\mathbb{E}\|\boldsymbol{x}(\eta)\|_2^2 = P$, $\boldsymbol{n}(\eta)$ is the noise vector of size $N_r \times 1$ with $n_i \overset{iid}{\sim} \mathcal{CN}(0, N_o), i = 1, \ldots, N_r$, and L is the number of resolvable delay paths (see Appendix B for more details). The $N_r \times N_t$ MIMO channel matrix satisfies $h_{mn}(\eta, i) \sim \mathcal{CN}(0,1)$ with sub-channel autocorrelations

$$R_{h_{mn}h_{mn}}(\tau, i) = J_o(2\pi f_{d_i} T_s \tau), \quad \tau = 0, 1, \ldots$$

where f_{d_i} is the maximum doppler frequency for the i^{th} delay path, T_s is the sampling period, and $J_o(\cdot)$ is the zeroth order Bessel function of the first kind. It is assumed that the sub-channels are uncorrelated both spatially and with respect to different path delays, i.e.,

$$\mathbb{E}\{\boldsymbol{H}(\eta, i)^H \boldsymbol{H}(\eta, j)\} = N_r \delta_k(i-j) \boldsymbol{I}_{N_t} \tag{3.2}$$

$$\mathbb{E}\{\boldsymbol{H}(\eta, i) \boldsymbol{H}(\eta, j)^H\} = N_t \delta_k(i-j) \boldsymbol{I}_{N_r}. \tag{3.3}$$

where $\delta_k(\cdot)$ denotes the kronecker delta.

Before investigating the frequency selective channel, we will consider first a Rayleigh flat fading channel (L = 0). In particular it will be shown that the results obtained in the flat fading case will be directly applicable to frequency selective channels.

Following the same reasoning as [12], [13], and [14], the channel is written in terms of its minimum mean squared error (MMSE) estimate $\widehat{\boldsymbol{H}}(\eta)$

$$\boldsymbol{H}(\eta) = \widehat{\boldsymbol{H}}(\eta) + \boldsymbol{E}(\eta)$$

where $\widehat{\boldsymbol{H}}(\eta)$ and $\boldsymbol{E}(\eta)$ are uncorrelated with $\widehat{h}_{mn}(\eta) \sim \mathcal{CN}(0, \sigma_{\widehat{h}}^2)$ and $e_{mn}(\eta) \sim \mathcal{CN}(0, \sigma_e^2)$. The channel estimation error between the m^{th} receiver and n^{th} transmitter is thus

$$\sigma_e^2 = 1 - \sigma_{\widehat{h}}^2.$$

3.2.1. Exact mutual information. The exact mutual information for imperfect CSI at the receiver was shown in [30] to be difficult. In particular, it was shown if the distribution of the channel estimation error is unknown, then mismatch issues need to be considered [32]. When the distribution error is known, the calculation of the conditional entropy is possible but may be impractical. The focus is now turned to obtaining upper and lower bounds for the instantaneous mutual information.

3.2.2. Bounds on mutual information. The upper and lower bounds for the instantaneous mutual information $\mathcal{I}(\boldsymbol{x}(\eta); \boldsymbol{y}(\eta))$ were extended in [12] to multiple antennas at the transmitter and receiver to yield for $N_r \leq N_t$

$$
\begin{aligned}
\mathcal{I}_{lb}(\eta) &= \log_2 \det \left(\boldsymbol{I}_{N_r} + \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)} \underline{\widehat{\boldsymbol{H}}}(\eta)\ \underline{\widehat{\boldsymbol{H}}}(\eta)^H \right) \\
\mathcal{I}_{ub}(\eta) &= \mathcal{I}_{lb} + N_r \mathbb{E}_{\boldsymbol{x}} \left\{ \log_2 \frac{\rho\sigma_e^2 + 1}{\frac{\rho}{N_t}\sigma_e^2 \|\boldsymbol{x}\|_2^2 + 1} \right\}
\end{aligned}
$$

where $\underline{\widehat{\boldsymbol{H}}}(\eta) \triangleq \widehat{\boldsymbol{H}}(\eta)/\sigma_{\widehat{h}}$, $\boldsymbol{x}$ is a N_r x 1 vector with $x_n \overset{iid}{\sim} \mathcal{CN}(0,1)$, and $\underline{\widehat{\boldsymbol{H}}}(\eta)\ \underline{\widehat{\boldsymbol{H}}}(\eta)^H$ has a Wishart distribution, and $\rho \triangleq \frac{P}{N_o}$ is the SNR. At this stage it is convenient to define another measure of SNR which will be used throughout the remainder of this work. Let the *effective SNR*, also denoted by ρ_{eff}, be defined as the SNR due to the presence of channel estimation error at the receiver. This is expressed mathematically as

$$
\rho_{eff} \triangleq \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}
$$

For $N_r > N_t$ we have

$$
\begin{aligned}
\mathcal{I}_{lb}(\eta) &= \log_2 \det \left(\boldsymbol{I}_{N_t} + \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)} \underline{\widehat{\boldsymbol{H}}}(\eta)^H\ \underline{\widehat{\boldsymbol{H}}}(\eta) \right) && (3.4) \\
\mathcal{I}_{ub}(\eta) &= \mathcal{I}_{lb} + N_r \mathbb{E}_{\boldsymbol{x}} \left\{ \log_2 \frac{\rho\sigma_e^2 + 1}{\frac{\rho_\tau}{N_t}\sigma_e^2 \|\boldsymbol{x}\|_2^2 + 1} \right\}. && (3.5)
\end{aligned}
$$

In either case the upper and lower bounds are the same due to the identity

$$\det\left(\boldsymbol{I}_{N_r} + \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\underline{\widehat{\boldsymbol{H}}}(\eta)\ \underline{\widehat{\boldsymbol{H}}}(\eta)^H\right) = \det\left(\boldsymbol{I}_{N_t} + \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\underline{\widehat{\boldsymbol{H}}}(\eta)^H\ \underline{\widehat{\boldsymbol{H}}}(\eta)\right).$$

For the remainder of this work the time dependence will be assumed autonomous for convenience unless explicitly stated.

3.3. OUTAGE CAPACITY FOR A MIMO RAYLEIGH FLAT FADING CHANNEL

An outage event is defined to occur when the instantaneous mutual information is strictly less than some fixed data rate r. The outage probability is defined to be [31]

$$P_o(\mathcal{I}) \triangleq P(\mathcal{I}(\boldsymbol{x};\boldsymbol{y}) \leq r). \tag{3.6}$$

Due to the monotonicity of a cumulative distribution function, the outage capacity can be defined as

$$\mathcal{C}(P_o(\mathcal{I})) \triangleq \underset{r}{\operatorname{argmax}}\{r > 0 : P_o(\mathcal{I}) = P\}$$

where P is the desired outage probability. Observing the inequalities

$$\begin{aligned} P_o(\mathcal{I}_{ub}) &\leq P_o(\mathcal{I}) \\ P_o(\mathcal{I}) &\leq P_o(\mathcal{I}_{lb}) \end{aligned}$$

it follows that

$$\begin{aligned} \mathcal{C}_{lb}(P_o(\mathcal{I}_{lb})) &\triangleq \underset{r}{\operatorname{argmax}}\{r > 0 : P_o(\mathcal{I}_{lb}) = P\} && (3.7) \\ \mathcal{C}_{ub}(P_o(\mathcal{I}_{ub})) &\triangleq \underset{r}{\operatorname{argmax}}\{r > 0 : P_o(\mathcal{I}_{ub}) = P\} && (3.8) \end{aligned}$$

are the lower and upper bounds for the outage capacity. In order to solve for r in (3.7) and (3.8), the probability density function (pdf) for (3.4) and (3.5) must be obtained. The mutual information can be written as

$$\begin{aligned} \mathcal{I}_{lb} &= \sum_{i=1}^{N} \log_2 \left(1 + \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)} \lambda_i \right) && (3.9) \\ \mathcal{I}_{ub} &= \mathcal{I}_{lb} + N_r \mathbb{E}_{\boldsymbol{x}} \left\{ \log_2 \frac{\rho\sigma_e^2 + 1}{\frac{\rho_\tau}{N_t}\sigma_e^2 \|\boldsymbol{x}\|_2^2 + 1} \right\} && (3.10) \end{aligned}$$

where $N = \text{rank}(\boldsymbol{H})$ and λ_i is the i^{th} unordered eigenvalue of a Wishart matrix. Letting $k \triangleq \min(N_r, N_t)$, $l \triangleq \max(N_r, N_t)$ and $d \triangleq l - k$, the pdf of the unordered eigenvalue is [1]

$$f_\lambda(\lambda) = \frac{\exp(-\lambda)}{k} \sum_{i=0}^{k-1} \frac{i!}{(i+d)!} \left(L_i^d(\lambda) \right)^2 \lambda^d \tag{3.11}$$

where $L_i^d(x)$ is the associated Laguerre polynomial defined by

$$L_i^d(x) \triangleq \sum_{p=0}^{i} (-1)^p \begin{pmatrix} i+d \\ i-p \end{pmatrix} \frac{x^p}{p!}. \tag{3.12}$$

Observing (3.11) and (3.12), it is clear that finding the pdf of (3.9) and (3.10) is rather involved compared to the SISO case in [31] where simple pdf transformations were performed. They can, however, be utilized for calculating (3.7) and (3.8) numerically, which will be known as the *numerical* upper and lower bounds on the outage capacity. This method is not attractive when the solution requires a lot of computational time[4] or when an analytical expression is desired. A contribution of this work is to find new *analytic* upper and lower bounds to (3.7) and (3.8).

The derivation will start with [33], where it is shown for $N_t, N_t \geq 2$ that $\mathcal{I}(\boldsymbol{x};\boldsymbol{y})$ can be well approximated by a Gaussian random variable. The lower and upper bounds satisfy

---

[4] It is also possible the numerical solution may oscillate and never converge.

$\mathcal{I}_{lb} \sim \mathcal{N}(\mu_{\mathcal{I}_{lb}}, \sigma_{\mathcal{I}}^2)$ and $\mathcal{I}_{ub} \sim \mathcal{N}(\mu_{\mathcal{I}_{ub}}, \sigma_{\mathcal{I}}^2)$ where

$$\begin{aligned}
\mu_{\mathcal{I}_{lb}} &= \int_0^\infty \log_2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda + 1\right) K(\lambda,\lambda)d\lambda && (3.13)\\
\mu_{\mathcal{I}_{ub}} &= \mu_{\mathcal{I}_{lb}} + N_r \mathbb{E}_{\boldsymbol{x}}\left\{ \log_2 \frac{\rho\sigma_e^2+1}{\frac{\rho_\tau}{N_t}\sigma_e^2\|\boldsymbol{x}\|_2^2+1}\right\} && (3.14)\\
\sigma_{\mathcal{I}}^2 &= \int_0^\infty \log_2^2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda + 1\right) K(\lambda,\lambda)d\lambda && (3.15)\\
&- \int_0^\infty \int_0^\infty \log_2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda_1 + 1\right)\\
&\quad \times \log_2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda_2 + 1\right) K^2(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2
\end{aligned}$$

and

$$K(x,y) \triangleq \sum_{i=0}^{k-1} \frac{i!}{(i+d)!} L_i^d(x) L_i^d(y) x^{d/2} y^{d/2} \exp\left(-(x/2+y/2)\right).$$

This allows us the write the desired probabilities in (3.7) and (3.8) as

$$\begin{aligned}
\mathrm{P}(\mathcal{I}_{lb} \le r) &= \int_0^r \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{I}_{lb}}} \exp\left(\frac{-(x-\mu_{\mathcal{I}_{lb}})^2}{2\sigma_{\mathcal{I}_{lb}}^2}\right)dx && (3.16)\\
\mathrm{P}(\mathcal{I}_{ub} \le r) &= \int_0^r \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{I}_{ub}}} \exp\left(\frac{-(x-\mu_{\mathcal{I}_{ub}})^2}{2\sigma_{\mathcal{I}_{ub}}^2}\right)dx. && (3.17)
\end{aligned}$$

which leads to the following new result for analytic upper and lower bounds of the MIMO outage capacity with channel uncertainty.

Theorem 3.1. *The outage capacity for a Rayleigh flat fading MIMO system with no CSI at the transmitter and channel estimation at the receiver is bounded by*

$$\mathcal{C}_{lb}(P) = \begin{cases}
\mu_{\mathcal{I}_{lb}} - \sigma_{\mathcal{I}}\sqrt{W\left(\frac{1}{2\pi\alpha_{lb}^2}\right)} & \textit{for } 0 < P < 0.159\\
\mu_{\mathcal{I}_{lb}} - \sigma_{\mathcal{I}}\sqrt{2\pi}\left(Q\left(\frac{\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) + P - \frac{1}{2}\right) & \textit{for } 0.159 < P < 0.5\\
\mu_{\mathcal{I}_{lb}} + \sigma_{\mathcal{I}}\sqrt{2\pi}\left(Q\left(\frac{\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) + P - \frac{1}{2}\right) & \textit{for } 0.5 < P < 0.841\\
\mu_{\mathcal{I}_{lb}} + \sigma_{\mathcal{I}}\sqrt{W\left(\frac{1}{2\pi\beta_{lb}^2}\right)} & \textit{for } 0.841 < P < 1
\end{cases} \qquad (3.18)$$

and

$$\mathcal{C}_{ub}(P) = \begin{cases} \mu_{\mathcal{I}_{ub}} - \sigma_{\mathcal{I}}\sqrt{W\left(\frac{1}{2\pi\alpha_{ub}^2}\right)} & \text{for } 0 < P < 0.159 \\ \mu_{\mathcal{I}_{ub}} - \sigma_{\mathcal{I}}\sqrt{2\pi}\left(Q\left(\frac{\mu_{\mathcal{I}_{ub}}}{\sigma_{\mathcal{I}}}\right) + P - \frac{1}{2}\right) & \text{for } 0.159 < P < 0.5 \\ \mu_{\mathcal{I}_{ub}} + \sigma_{\mathcal{I}}\sqrt{2\pi}\left(Q\left(\frac{\mu_{\mathcal{I}_{ub}}}{\sigma_{\mathcal{I}}}\right) + P - \frac{1}{2}\right) & \text{for } 0.5 < P < 0.841 \\ \mu_{\mathcal{I}_{ub}} + \sigma_{\mathcal{I}}\sqrt{W\left(\frac{1}{2\pi\beta_{ub}^2}\right)} & \text{for } 0.841 < P < 1 \end{cases} \tag{3.19}$$

where

$$\begin{aligned} \alpha_{lb} &\triangleq Q\left(\frac{\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) + P \\ \alpha_{ub} &\triangleq Q\left(\frac{\mu_{\mathcal{I}_{ub}}}{\sigma_{\mathcal{I}}}\right) + P \\ \beta_{lb} &\triangleq -Q\left(\frac{\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) + 1 - P \\ \beta_{ub} &\triangleq -Q\left(\frac{\mu_{\mathcal{I}_{ub}}}{\sigma_{\mathcal{I}}}\right) + 1 - P \end{aligned}$$

and $Q(\cdot)$ and $W(\cdot)$ are respectively the Gaussian Q function and Lambert's W function [34].

Proof. Without loss of generality only the lower bound will be considered. Writing (3.16) in terms of Q functions we obtain

$$Q\left(\frac{r - \mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) = Q\left(\frac{\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) + (1 - P). \tag{3.20}$$

The following approximations for the Q function are employed

$$\begin{aligned} Q(x) &\approx 1/2\left(1 - \frac{2}{\sqrt{2\pi}}x\right), -1 < x < 1 & (3.21) \\ Q(x) &\approx \frac{1}{\sqrt{2\pi}x}\exp\left(-x^2/2\right), x > 1 & (3.22) \\ Q(x) &\approx 1 - \frac{1}{\sqrt{2\pi}x}\exp\left(-x^2/2\right), x < -1. & (3.23) \end{aligned}$$

There are four distinct cases to consider.

(I) $\frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} < -1$. Using (3.23) in (3.20) we obtain after several several manipulations

$$\left(\frac{r-\mu_{\mathcal{I}_{lb}}}{-\sigma_{\mathcal{I}}}\right)^2 \exp\left(\frac{r-\mu_{\mathcal{I}_{lb}}}{-\sigma_{\mathcal{I}}}\right)^2 = \frac{1}{2\pi\alpha_{lb}^2}.$$

Noting the definition of the lambert W function we obtain

$$\left(\frac{r-\mu_{\mathcal{I}_{lb}}}{-\sigma_{\mathcal{I}}}\right)^2 = W\left(\frac{1}{2\pi\alpha_{lb}^2}\right)$$

from which after taking the square root and solving for r the result follows.

(II) $-1 < \frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} < 0$. Using (3.21) in (3.20) we obtain

$$\frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} = \frac{\sqrt{2\pi}}{2}\left(-2Q\left(\frac{\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}\right) + 1 - 2P\right)$$

and obtain the result after solving for r.

(III) $0 < \frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} < 1$. Using (3.21) in (3.20) and following the same method as Case II, the result follows.

(IV) $\frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} > 1$. Using (3.22) in (3.20) and following the same procedure in Case I yields the desired result.

It remains to show the outage probabilities corresponding to the values of $\frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}}$. For a Gaussian random variable, it is well known that

$$\begin{aligned} P(|x-\mu| < \sigma) &\leq 0.682 \\ P(|x-\mu| < 2\sigma) &\leq 0.95 \\ P(|x-\mu| < 3\sigma) &\leq 0.997. \end{aligned}$$

Thus

$$\begin{aligned}
\frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} &< -1 \quad \text{implies} \quad 0 < P < 0.159 \\
-1 < \frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} &< 0 \quad \text{implies} \quad 0.159 < P < 0.5 \\
0 < \frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} &< 1 \quad \text{implies} \quad 0.5 < P < 0.841 \\
\frac{r-\mu_{\mathcal{I}_{lb}}}{\sigma_{\mathcal{I}}} &> 1 \quad \text{implies} \quad 0.841 < P < 1
\end{aligned}$$

which completes the proof. ☐

At this point it is difficult to gain insight on (3.18) and (3.19) due to their complicated structure. It is therefore appropriate to show by illustration how the new analytic bounds behave with respect to various parameters.

First let $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.01$. The analytic and numerical bounds are plotted for comparison in Figure 3.1 for various P. The two methods of calculating the bounds are observed to be in close agreement. It is also noted that increasing the outage probability

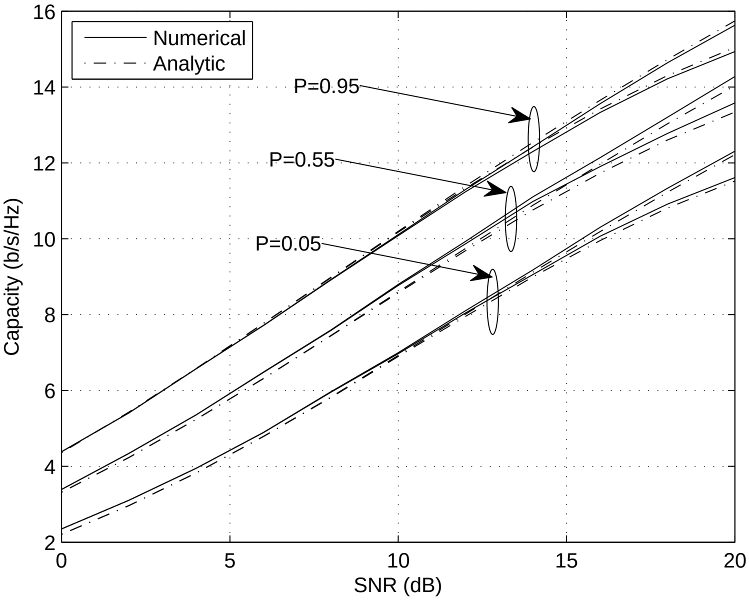

Figure 3.1 Comparison of numerical and analytical bounds for the MIMO outage capacity with channel uncertainty for various P when $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.01$.

increases the outage capacity bounds. This follows immediately from the definition of outage probability in (3.6).

The number of transmit and receive antennas are now varied while fixing $P = 0.05$ and $\sigma_e^2 = 0.01$ in Figure 3.2. The outage capacity bounds appear logarithmic with increasing received SNR and linear with increasing antenna elements which are fundamental results for MIMO capacity [1].

The importance of accurate channel estimation is demonstrated in Figure 3.3 for $N_t = 2, N_r = 5, P = 0.05$, and $\sigma_e^2 = 0.01$. At high SNR the outage capacity bounds saturate and loosen as the estimation error increases. The saturation can be explained by inaccurate channel estimation effectively adding noise, or equivalently uncertainty, to the system which prevents the differential entropy from decreasing at high SNR. The looseness in the bounds follows by visualizing the upper and lower bounds of the mutual information as two extremes, the former is when the estimation error behaves as noise, the latter when it presents itself as transmit power.

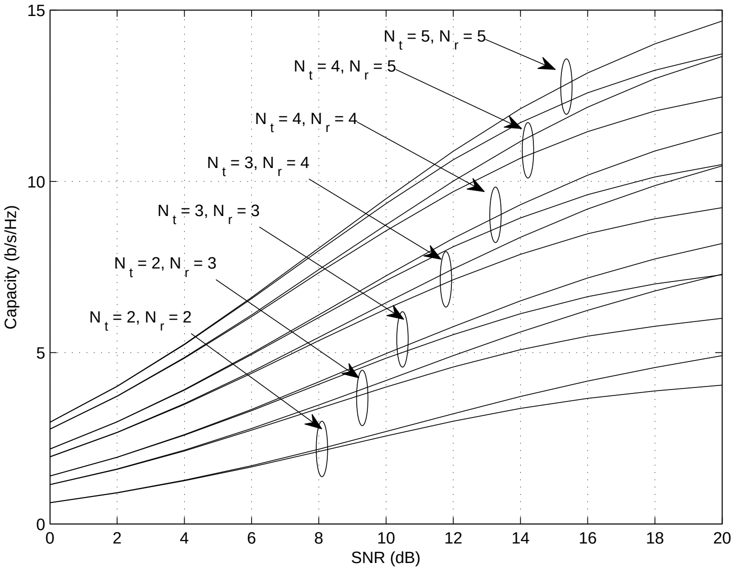

Figure 3.2 Analytic bounds for the MIMO outage capacity with channel uncertainty for various N_t and N_r with $P = 0.05$ and $\sigma_e^2 = 0.01$.

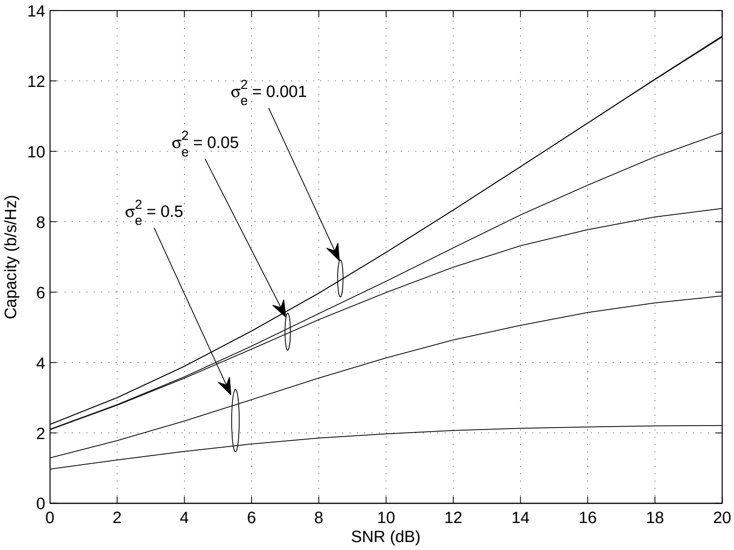

Figure 3.3 Analytic bounds for the MIMO outage capacity with channel uncertainty for various σ_e^2 when $N_t = 2, N_r = 5$, and $P = 0.05$.

Noting that the outage capacity is a function of the mean and variance of the mutual information, a new exact expression for these moments is now presented. Although cumbersome, they can be implemented using numerical integration techniques which avoid the inaccuracies that can result from Monte Carlo simulations.

Theorem 3.2. *The mean and variance of the instantaneous mutual information for a MIMO Rayleigh flat fading system with no CSI at the transmitter and channel estimation at the receiver are*

$$\begin{aligned}
\mu_{\mathcal{I}_{lb}} &= \sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}(-1)^{p_1+p_2}\frac{i!\binom{i+d}{i-p_1}\binom{i+d}{i-p_2}}{\ln 2\ (i+d)!p_1!p_2!} \\
&\quad \times G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\ 0,0,p_1+p_2+d\end{matrix}\right) && (3.24)\\
\mu_{\mathcal{I}_{ub}} &= \mu_{\mathcal{I}_{lb}} + N_r\log_2(N_t\sigma_e^2+1) \\
&- N_r\int_0^{\infty}\frac{1}{2^{N_t/2}\Gamma(N_t/2)}x^{N_t/2-1}e^{-x/2}\log_2(x\sigma_e^2+1)d\lambda && (3.25)\\
\sigma_{\mathcal{I}}^2 &= \sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\sum_{j=0}^{p_1+p_2+d}(-1)^{d-j} \\
&\quad \times\frac{2e^{\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}}i!\binom{p_1+p_2+d}{j}\binom{i+d}{i-p_1}\binom{i+d}{i-p_2}}{(\ln 2)^2(i+d)!p_1!p_2!\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right)^{p_1+p_2+d+1}} \\
&\quad \times G_{40}^{34}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}-j,-j,-j\\ 0,-j-1,-j-1,-j-1\end{matrix}\right) \\
&- \left[\sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}(-1)^{p_1+p_2}\frac{i!\binom{i+d}{i-p_1}\binom{i+d}{i-p_2}}{(\ln 2)^2(i+d)!p_1!p_2!}\right. \\
&\quad \left.\times G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\ 0,0,p_1+p_2+d\end{matrix}\right)\right]^2 && (3.26)
\end{aligned}$$

Proof. We begin by noting that

$$\begin{aligned} K(\lambda,\lambda) &= \sum_{i=0}^{k-1} \Psi_i L_i^d(\lambda) L_i^d(\lambda)\lambda^d e^{-\lambda} \\ &= \sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i} \Psi_i \Phi_{p_1}\Phi_{p_2}\lambda^{p_1+p_2+d}e^{-\lambda} \end{aligned}$$

and substituting in (3.13) yields

$$\begin{aligned} \mu_{\mathcal{I}_{lb}} &= \frac{1}{\ln 2}\sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i} \Psi_i \Phi_{p_1}\Phi_{p_2}\int_0^\infty \ln\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda+1\right)\lambda^{p_1+p_2+d}e^{-\lambda}d\lambda \\ &= \frac{1}{\ln 2}\sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i} \Psi_i \Phi_{p_1}\Phi_{p_2} G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\0,0,p_1+p_2+d\end{matrix}\right) \\ &= \sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}(-1)^{p_1+p_2}\frac{i!\begin{pmatrix}i+d\\i-p_1\end{pmatrix}\begin{pmatrix}i+d\\i-p_2\end{pmatrix}}{\ln 2\ (i+d)!p_1!p_2!} \\ &\quad\times G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\0,0,p_1+p_2+d\end{matrix}\right) \end{aligned}$$

where the second equality is obtained from (11),(21) in [35] and $G(\cdot)$ is the Meijer G function [34, 35]. The variance evaluation begins by noting that

$$
\begin{aligned}
&\int_0^\infty \log_2^2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda+1\right)K(\lambda,\lambda)d\lambda\\
=\;& \frac{1}{(\ln 2)^2}\sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\Psi_i\Phi_{p_1}\Phi_{p_2}\int_0^\infty \ln^2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda+1\right)\lambda^{p_1+p_2+d}e^{-\lambda}d\lambda\\
=\;& \sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\Psi_i\Phi_{p_1}\Phi_{p_2}\frac{2e^{\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}}}{(\ln 2)^2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right)^{p_1+p_2+d+1}}\\
&\times\sum_{j=0}^{p_1+p_2+d}\begin{pmatrix} p_1+p_2+d\\ j\end{pmatrix}(-1)^{p_1+p_2+d-j}\\
&\times G_{34}^{40}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}-j,-j,-j\\ 0,-j-1,-j-1,-j-1\end{matrix}\right)\\
=\;& \sum_{i=1}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\sum_{j=0}^{p_1+p_2+d}(-1)^{d-j}\frac{i!}{(i+d)!p_1!p_2!}\begin{pmatrix} p_1+p_2+d\\ j\end{pmatrix}\begin{pmatrix} i+d\\ i-p_1\end{pmatrix}\\
&\times\begin{pmatrix} i+d\\ i-p_2\end{pmatrix}\frac{2e^{\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}}}{(\ln 2)^2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right)^{p_1+p_2+d+1}}\\
&\times G_{34}^{40}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}-j,-j,-j\\ 0,-j-1,-j-1,-j-1\end{matrix}\right)
\end{aligned}
\tag{3.27}
$$

where the third equality is obtained from [36]. The second term in (3.15) is obtained by first writing

$$
\begin{aligned}
K^2(\lambda_1,\lambda_2) &= \left(\sum_{i=0}^{k-1}\Psi_i L_i^d(\lambda_1)L_i^d(\lambda_2)(\lambda_1\lambda_2)^{d/2}e^{-(\lambda_1+\lambda_2)/2}\right)\\
&\quad\times\left(\sum_{j=0}^{k-1}\Psi_i L_i^d(\lambda_1)L_i^d(\lambda_2)(\lambda_1\lambda_2)^{d/2}e^{-(\lambda_1+\lambda_2)/2}\right)\\
&= (\lambda_1\lambda_2)^d e^{-(\lambda_1+\lambda_2)}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\Psi_i\Psi_j L_i^d(\lambda_1)L_i^d(\lambda_2)L_j^d(\lambda_1)L_j^d(\lambda_2).
\end{aligned}
$$

Substituting this into (3.15) yields

$$
\begin{aligned}
&\int_0^\infty\int_0^\infty \log_2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda_1+1\right)\log_2\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda_2+1\right)K^2(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2\\
=\;& \frac{1}{(\ln 2)^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\sum_{p_3=0}^{j}\sum_{p_4=0}^{j}\Psi_i\Psi_j\Phi_{p_1}\Phi_{p_2}\Phi_{p_3}\Phi_{p_4}\\
&\times\int_0^\infty \ln\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda_1+1\right)e^{-\lambda_1}\lambda_1^{p1+p3+d}d\lambda_1\\
&\times\int_0^\infty \ln\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\lambda_2+1\right)e^{-\lambda_2}\lambda_2^{p1+p3+d}d\lambda_2\\
=\;& \sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\sum_{p_3=0}^{j}\sum_{p_4=0}^{j}(-1)^{p_1+p_2+p_3+p_4}\frac{i!}{(\ln 2)^2(i+d)!p_1!p_2!p_3!p_4!}\begin{pmatrix} i+d\\ i-p_1\end{pmatrix}\\
&\times\begin{pmatrix} i+d\\ i-p_2\end{pmatrix}\begin{pmatrix} i+d\\ i-p_3\end{pmatrix}\begin{pmatrix} i+d\\ i-p_4\end{pmatrix}G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\0,0,p_1+p_3+d\end{matrix}\right)\\
&\times G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\0,0,p_2+p_4+d\end{matrix}\right)\\
=\;& \left[\sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}(-1)^{p_1+p_2}\frac{i!\begin{pmatrix} i+d\\ i-p_1\end{pmatrix}\begin{pmatrix} i+d\\ i-p_2\end{pmatrix}}{(\ln 2)^2(i+d)!p_1!p_2!}\right.\\
&\left.\times G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix}0,1\\0,0,p_1+p_2+d\end{matrix}\right)\right]^2
\end{aligned}
\tag{3.28}
$$

Subtracting (3.27) from (3.28) yields the variance

$$
\begin{aligned}
\sigma_{\mathcal{I}}^2 \;=\; & \sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}\sum_{j=0}^{p_1+p_2+d}(-1)^{d-j}\frac{2e^{\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}}i!\begin{pmatrix} p_1+p_2+d \\ j \end{pmatrix}\begin{pmatrix} i+d \\ i-p_1 \end{pmatrix}\begin{pmatrix} i+d \\ i-p_2 \end{pmatrix}}{(\ln 2)^2(i+d)!p_1!p_2!\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right)^{p_1+p_2+d+1}} \\
& \times G_{40}^{34}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix} -j,-j,-j \\ 0,-j-1,-j-1,-j-1 \end{matrix}\right) \\
- & \left[\sum_{i=0}^{k-1}\sum_{p_1=0}^{i}\sum_{p_2=0}^{i}(-1)^{p_1+p_2}\frac{i!\begin{pmatrix} i+d \\ i-p_1 \end{pmatrix}\begin{pmatrix} i+d \\ i-p_2 \end{pmatrix}}{(\ln 2)^2(i+d)!p_1!p_2!}\right. \\
& \left.\times G_{23}^{31}\left(\frac{N_t(1+\rho\sigma_e^2)}{\rho(1-\sigma_e^2)}\middle|\begin{matrix} 0,1 \\ 0,0,p_1+p_2+d \end{matrix}\right)\right]^2
\end{aligned}
$$

which is the desired result. □

To provide additional insight and relieve the computational burden of numerical integration, the next topic will be the derivation of new accurate closed form approximations for these moments. The approximations are driven by the asymptotic behavior of the logarithm.

3.3.1. Low Effective SNR.

The following approximation for the logarithm is valid in the low effective SNR regime

$$
\log(1+x) \approx \frac{x}{\ln 2}, \tag{3.29}
$$

and is utilized for the following new result.

Theorem 3.3. *The mean and variance of the mutual information of a MIMO Rayleigh flat fading system with no CSI at the transmitter and channel estimation at the receiver are well* *approximated at low effective SNR by*

$$\begin{aligned} \mu_{\mathcal{I}_{lb}} &\approx \frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)} k(k+d) && (3.30) \\ \mu_{\mathcal{I}_{lb}} &\approx \mu_{\mathcal{I}_{ub}} + M_{ub} && (3.31) \\ \sigma_{\mathcal{I}}^2 &\approx \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\right)^2 \frac{N_r}{N_t} && (3.32) \end{aligned}$$

where

$$M_{ub} = \begin{cases} 0 & \text{for } \rho\sigma_e^2 < 0.5 \\ \frac{N_r}{\ln 2}\left(\ln\frac{N_t}{2} - \psi\left(\frac{N_t}{2}\right)\right) & \text{for } \rho\sigma_e^2 > 2 \\ \frac{N_r}{\ln 2}\frac{(\rho-\rho^-)}{(\rho^+-\rho^-)}\left(\ln\frac{N_t}{2} - \psi\left(\frac{N_t}{2}\right)\right) & \text{otherwise} \end{cases}$$

and

$$\rho^- = \max\{\rho \in \mathbb{R} : \rho\sigma_e^2 < 0.5\}$$

$$\rho^+ = \min\{\rho \in \mathbb{R} : \rho\sigma_e^2 > 2\}$$

with $\psi(\cdot)$ *being the psi function [34].*

Proof. We begin by using (3.29) to approximate (3.13) by

$$\begin{aligned} \mu_{\mathcal{I}_{lb}} &= \int_0^\infty \frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)} \lambda K(\lambda,\lambda) d\lambda \\ &= \sum_{i=0}^{k-1} \frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)} \frac{i!}{(i+d)!} \int_0^\infty L_i^d(\lambda) L_i^d(\lambda) \lambda^{d+1} \exp(-\lambda) d\lambda. && (3.33) \end{aligned}$$

Using the identities [34]

$$\begin{aligned} L_i^{d-1}(\lambda) &= L_i^d(\lambda) - L_{i-1}^d(\lambda) && (3.34) \\ \int_0^\infty \exp(-x) x^\alpha L_n^\alpha(x) L_m^\alpha(x) dx &= \begin{cases} 0 & , m \neq n \\ \frac{\Gamma(\alpha+n+1)}{n!} & , m = n \end{cases} && (3.35) \end{aligned}$$

we obtain after several manipulations

$$\mu_{\mathcal{I}_{lb}} = \frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)} k(k+d) \tag{3.36}$$

Noting that at low effective SNR the approximation $\sigma_e \approx 1$ holds and there are three cases to consider for M_{ub}.

1. $\rho\sigma_e^2 < 0.5$. This renders the approximation

   $$\ln\left(\frac{\rho_\tau}{N_t}\sigma_e^2\|\boldsymbol{x}\|_2^2 + 1\right) \approx \frac{\rho\sigma_e^2\|\boldsymbol{x}\|_2^2}{\ln 2\, N_t}. \tag{3.37}$$

   Since $||\boldsymbol{x}||_2^2 \sim \chi^2_{N_t}$ it follows that

   $$M_{ub} = 0. \tag{3.38}$$

2. $\rho\sigma_e^2 \geq 2$. Then

   $$\ln\left(\frac{\rho\sigma_e^2 + 1}{\frac{\rho_\tau}{N_t}\|\boldsymbol{x}\|_2^2\sigma_e^2 + 1}\right) \approx \ln\left(\frac{N_t}{\|\boldsymbol{x}\|_2^2}\right). \tag{3.39}$$

   Using the identity [34]

   $$\int_0^\infty \frac{\mu^\nu x^{\nu-1}\exp(-\mu x)\ln x}{\Gamma(\nu)} dx = \psi(\nu) + \ln 2 \tag{3.40}$$

   we obtain

   $$M_{ub} = \frac{N_r}{\ln 2}\left(\ln\frac{N_t}{2} - \psi\left(\frac{N_t}{2}\right)\right).$$

3. $0.5 < \rho\sigma_e^2 < 2$. In this region it is difficult to find a simple logarithmic approximation. Observing from Figure 3.4 that the upper bound of the mean increases in a linear fashion, a first order linear interpolation is employed to obtain

   $$M_{ub} = \frac{N_r}{\ln 2}\frac{(\rho-\rho^-)}{(\rho^+-\rho^-)}\left(\ln\frac{N_t}{2} - \Psi\left(\frac{N_t}{2}\right)\right) \tag{3.41}$$

where

$$\begin{aligned}\rho^{-} &= \max\{\rho \in \mathbb{R} : \rho\sigma_e^2 < 0.5\} \\ \rho^{+} &= \min\{\rho \in \mathbb{R} : \rho\sigma_e^2 > 2\}.\end{aligned}$$

Using the assumption now in (3.29), the variance can be expressed as

$$\sigma_{\mathcal{I}}^2 = \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 (I_1 - I_2) \tag{3.42}$$

where

$$\begin{aligned} I_1 &= \int_0^\infty \lambda^2 K(\lambda,\lambda) d\lambda \\ I_2 &= \int_0^\infty \int_0^\infty \lambda_1\lambda_2 K^2(\lambda_1,\lambda_2) d\lambda_1 d\lambda_2. \end{aligned}$$

Expanding I_1 first gives

$$I_1 = \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 \sum_{i=0}^{k-1} \frac{i!}{(i+d)!} \int_0^\infty L_i^d(\lambda) L_i^d(\lambda) \lambda^{d+2} \exp(-\lambda) d\lambda.$$

After applying (3.35) we obtain after several manipulations

$$I_1 = \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 k(k+d)(d+2k).$$

Focusing now on I_2 we have

$$\begin{aligned} I_2 &= \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 \int_0^\infty \int_0^\infty \lambda_1^{d+1}\lambda_2^{d+1} \exp(-\lambda_1 - \lambda_2) \\ &\times \left(\sum_{i=0}^{k-1} \frac{i!}{(i+d)!} L_i^d(\lambda_1) L_i^d(\lambda_2)\right) \left(\sum_{j=0}^{k-1} \frac{j!}{(j+d)!} L_j^d(\lambda_1) L_j^d(\lambda_2)\right) d\lambda_1 d\lambda_2 \\ &= \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 \left(\sum_{i,j\in\mathcal{A}} I_{\mathcal{A}} + \sum_{i,j\in\mathcal{A}^c} I_{\mathcal{A}^c}\right) \end{aligned} \tag{3.43}$$

where $\mathcal{A} = \{(i,j) \in \{0,\ldots,k-1\} \times \{0,\ldots,k-1\} : i = j\}$ and

$$\begin{aligned} I_{\mathcal{A}} &= \left(\frac{i!}{(i+d)!}\int_0^\infty \lambda^{d+1}\exp(-\lambda)L_i^d(\lambda)L_i^d(\lambda)d\lambda\right)^2 \\ &= (2i+d+1)^2 \qquad (3.44) \\ I_{\mathcal{A}^c} &= \frac{i!}{(i+d)!}\frac{(i+1)!}{(i+1+d)!}\left(\int_0^\infty \lambda^{d+1}\exp(-\lambda)L_i^d(\lambda)L_j^d(\lambda)d\lambda\right)^2 \\ &= (i+1)(i+d+1). \qquad (3.45) \end{aligned}$$

Substituting (3.44) and (3.45) into (3.43) yields

$$I_2 = \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 \left(\frac{k}{3}\big(3d^2+4k^2+6dk-1\big) + \frac{k}{3}(k-1)(2k-1+3d)\right).$$

Subtracting I_2 from I_1 yields the variance

$$\sigma_{\mathcal{I}}^2 = \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\, N_t(1+\rho\sigma_e^2)}\right)^2 \Big(k(k+d)\Big) = \left(\frac{\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\right)^2 \frac{N_r}{N_t}.$$

□

The approximations are examined by plotting the effective received SNR along with (3.30)-(3.32) for $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.95$ in Figure 3.4. A comparison with (3.13)-(3.15) indicates there is close agreement. The simplification of these moments lead to new closed form estimates for the outage capacity.

Theorem 3.4. *The outage capacity bounds of a MIMO Rayleigh flat fading system operating at low effective SNR with no CSI at the transmitter and channel estimation at the receiver*

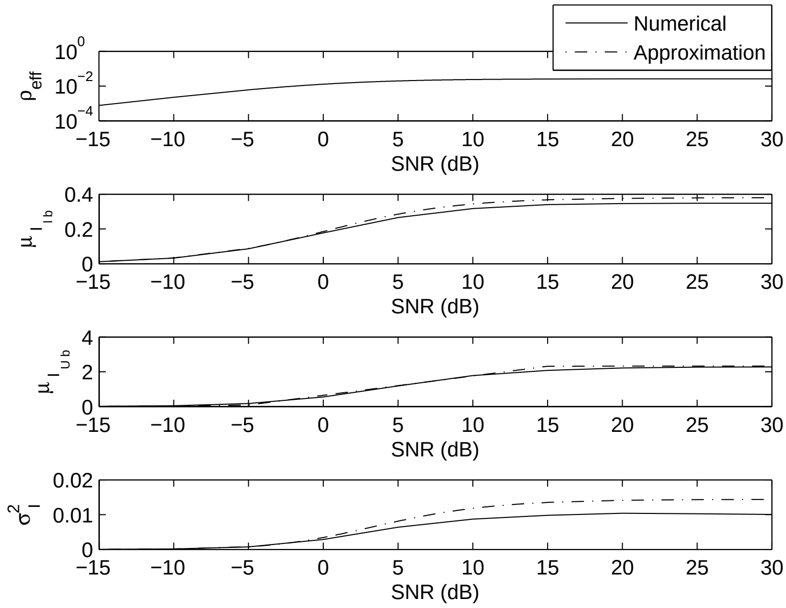

Figure 3.4 Comparison of numerical and approximate moments for the mutual information at low effective SNR when $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.95$.

are well approximated by

$$\mathcal{C}_{lb}(P) \approx \begin{cases} \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1-\sqrt{\frac{W\left(\frac{1}{2\pi\alpha_{lb}^2}\right)}{kl}}\right) & \textit{for}\ 0 < P < 0.159 \\ \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1-\sqrt{\frac{2\pi}{kl}}\left(Q\left(\sqrt{kl}\right)+P-\frac{1}{2}\right)\right) & \textit{for}\ 0.159 < P < 0.5 \\ \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1+\sqrt{\frac{2\pi}{kl}}\left(Q\left(\sqrt{kl}\right)+P-\frac{1}{2}\right)\right) & \textit{for}\ 0.5 < P < 0.841 \\ \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1+\sqrt{\frac{W\left(\frac{1}{2\pi\beta_{lb}^2}\right)}{kl}}\right) & \textit{for}\ 0.841 < P < 1 \end{cases}$$

and

$$\mathcal{C}_{ub}(P) \approx \begin{cases} \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1-\sqrt{\frac{W\left(\frac{1}{2\pi\alpha_{lb}^2}\right)}{kl}}\right)+M_{ub} & \textit{for}\ 0 < P < 0.159 \\ \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1-\sqrt{\frac{2\pi}{kl}}\left(Q\left(\sqrt{kl}+\delta\right)+P-\frac{1}{2}\right)\right)+M_{ub} & \textit{for}\ 0.159 < P < 0.5 \\ \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1+\sqrt{\frac{2\pi}{kl}}\left(Q\left(\sqrt{kl}+\delta\right)+P-\frac{1}{2}\right)\right)+M_{ub} & \textit{for}\ 0.5 < P < 0.841 \\ \frac{N_r\rho(1-\sigma_e^2)}{\ln 2\,(1+\rho\sigma_e^2)}\left(1+\sqrt{\frac{W\left(\frac{1}{2\pi\beta_{lb}^2}\right)}{kl}}\right)+M_{ub} & \textit{for}\ 0.841 < P < 1 \end{cases}$$

where

$$\begin{aligned}
\alpha_{lb} &\triangleq Q\left(\sqrt{kl}\right)+P\\
\alpha_{ub} &\triangleq Q\left(\sqrt{kl}+\delta\right)+P\\
\beta_{lb} &\triangleq -Q\left(\sqrt{kl}\right)+1-P\\
\beta_{ub} &\triangleq -Q\left(\sqrt{kl}+\delta\right)+1-P
\end{aligned}$$

and

$$\delta \triangleq \frac{\ln 2M_{ub}N_t(1+\rho\sigma_e^2)}{\sqrt{kl}\rho(1-\sigma_e^2)}$$

along with

$$M_{ub}=\begin{cases} 0 & \text{for } \rho\sigma_e^2<0.5\\ \frac{N_r}{\ln 2}\left(\ln\frac{N_t}{2}-\Psi\left(\frac{N_t}{2}\right)\right) & \text{for } \rho\sigma_e^2>2\\ \frac{N_r}{\ln 2}\frac{(\rho-\rho^-)}{(\rho^+-\rho^-)}\left(\ln\frac{N_t}{2}-\Psi\left(\frac{N_t}{2}\right)\right) & \text{otherwise}\end{cases}$$

where

$$\begin{aligned}
\rho^- &= \max\{\rho\in\mathbb{R}:\rho\sigma_e^2<0.5\}\\
\rho^+ &= \min\{\rho\in\mathbb{R}:\rho\sigma_e^2>2\}
\end{aligned}$$

Proof. This follows at once upon inserting (3.30)-(3.32) into (3.18) and (3.19). □

We begin the discussion with a qualitative analysis. Starting with the lower bound, for each case the outage capacity is depicted by $K(1\pm a)$. We now show $a\to 0$ when the number of antennas are sufficiently large. For $.159<P<.841$ this is straightforward since the Q function is bounded. For the remaining two scenarios, observe that the estimate $W(x)\le x$ holds. Therefore

$$0\le\frac{W\left(\frac{1}{2\pi(Q(\sqrt{kl})+P)^2}\right)}{kl}\le\frac{1}{2\pi P^2kl}$$

which is squeezed to zero as the number of antenna elements are increased. The factored term K shows that the outage capacity increases linearly with N_r and saturates as the estimation error worsens. Note that no linear increase occurs when increasing N_t because the transmit covariance matrix allocates equal power on each antenna [25] when the transmitter does not know the channel. When the error is small, the capacity increases linearly with received SNR, consistent with [1].

The behavior of the upper bound is similar to the lower bound minus two exceptions. The first is the inclusion of δ into the Q functions. By close inspection δ is non-negative, causing a to vanish quicker. The second is the presence of M_{ub}, which is composed of strictly increasing functions. Observing that [34]

$$\lim_{N_t \to \infty} \psi(N_t) = \ln N_t \tag{3.46}$$

it is clear that M_{ub} will vanish for a sufficiently large number of transmit antennas. This suggests at low effective SNR that increasing the number of transmit antennas does not benefit the outage capacity.

Numerical examples are now provided to validate our observations. Letting $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.95$, the outage capacity using the approximations in Theorem 3.4 are compared to the actual outage capacity in Figure 3.5. The approximations are seen to be in close agreement over the range of received SNR.

The impact on the outage capacity approximations when the number of transmit antennas are varied for $N_r = 2, 4$ along with $P = 0.05$ and $\sigma_e^2 = 0.95$ are illustrated in Figure 3.6. As indicated in the analysis, increasing the number of receive elements increases the outage capacity while increasing the number of transmit antennas has little impact.

3.3.2. High Effective SNR.

When operating in the high effective SNR regime the approximation

$$\log_2(1 + x) \approx \log_2 x$$

is valid. The upper and lower bounds for the mutual information are well approximated for $N_r \leq N_t$ by

$$\begin{aligned}
\mathcal{I}_{lb} &= \log_2 \det\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)} \widehat{\boldsymbol{H}}\, \widehat{\boldsymbol{H}}^H\right) \\
\mathcal{I}_{ub} &= \mathcal{I}_{lb} + N_r \mathbb{E}_{\boldsymbol{x}}\left\{ \log_2 \frac{N_t\sigma_e^2+1}{\|\boldsymbol{x}\|_2^2\sigma_e^2+1} \right\}
\end{aligned}$$

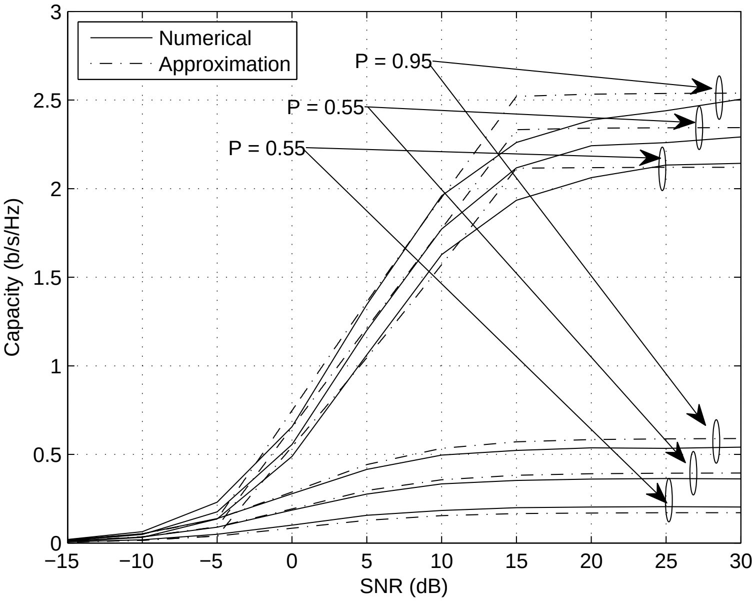

Figure 3.5 Comparison of numerical bounds and their approximations for the MIMO outage capacity with channel uncertainty at low effective SNR for various P when $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.95$.

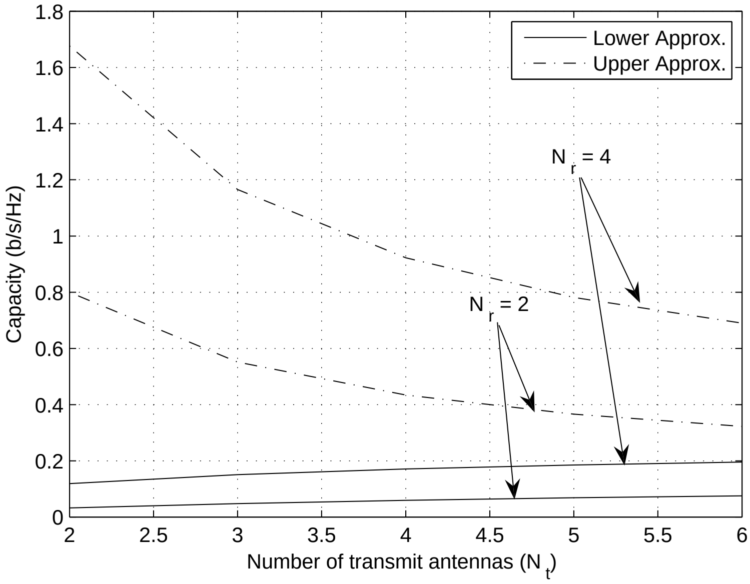

Figure 3.6 Approximations for the MIMO outage capacity with channel uncertainty at low effective SNR for various N_t and $N_r = 2, 4$ with $P = .05$ and $\sigma_e^2 = .95$.

and for $N_r > N_t$

$$\begin{aligned}\mathcal{I}_{lb} &= \log_2 \det\left(\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\widehat{\boldsymbol{H}}^H\,\widehat{\boldsymbol{H}}\right)\\ \mathcal{I}_{ub} &= \mathcal{I}_{lb} + N_r\mathbb{E}_{\boldsymbol{x}}\left\{\log_2\frac{N_t\sigma_e^2+1}{\|\boldsymbol{x}\|_2^2\sigma_e^2+1}\right\}\end{aligned}$$

which leads to new closed form approximate moments at high SNR.

Theorem 3.5. *The mean and variance of the instantaneous mutual information for a MIMO Rayleigh flat fading system with no CSI at the transmitter and channel estimation at the receiver are well approximated at high effective SNR by*

$$\mu_{\mathcal{I}_{lb}} \approx \frac{1}{\ln 2}\left(\sum_{i=0}^{k-1}\psi(k+d-i) + k\ln\frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right) \tag{3.47}$$

$$\mu_{\mathcal{I}_{ub}} \approx \mu_{\mathcal{I}_{lb}} + M_{ub} \tag{3.48}$$

$$\sigma_{\mathcal{I}}^2 \approx \frac{1}{(\ln 2)^2}\sum_{i=0}^{k-1}\zeta(2,k+d-i) \tag{3.49}$$

where

$$M_{ub} = \begin{cases} 0 & \text{for } \rho\sigma_e^2 < 0.5\\ \frac{N_r}{\ln 2}\left(\ln\frac{N_t}{2} - \psi\left(\frac{N_t}{2}\right)\right) & \text{for } \rho\sigma_e^2 > 2\\ \frac{N_r}{\ln 2}\frac{(\rho-\rho^-)}{(\rho^+-\rho^-)}\left(\ln\frac{N_t}{2} - \psi\left(\frac{N_t}{2}\right)\right) & \text{otherwise}\end{cases}$$

and

$$\rho^- = \max\{\rho\in\mathbb{R} : \rho\sigma_e^2 < 0.5\}$$

$$\rho^+ = \min\{\rho\in\mathbb{R} : \rho\sigma_e^2 > 2\}$$

with $\zeta(\cdot,\cdot)$ being the Riemann zeta function [34].

Proof. Assume that $N_r \le N_t$. We begin by noting for iid complex normal matrices at high effective SNR that [37]

$$\frac{\det \boldsymbol{H}\boldsymbol{H}^H}{2^{-k}} \sim \prod_{i=1}^{k}\chi^2_{2(k+d-i+1)}.$$

Since the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^H$ and $\boldsymbol{H}^H\boldsymbol{H}$ are the same, the result holds as well for $N_r > N_t$. It is routine to show that

$$\mathcal{I}_{lb} \sim \sum_{i=0}^{k-1} \ln(\chi^2_{2(k+d-i)}) + k \ln \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)} \tag{3.50}$$

and therefore

$$\mu_{\mathcal{I}_{lb}} = \frac{1}{\ln 2}\left(\mathbb{E}\left\{\sum_{i=0}^{k-1} \ln(\chi^2_{2(k+d-i)})\right\} + k \ln \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right).$$

Invoking the fact that the chi squared random variables are independent

$$\mu_{\mathcal{I}_{lb}} = \frac{1}{\ln 2}\left(\sum_{i=0}^{k-1} \mathbb{E}\left\{\ln(\chi^2_{2(k+d-i)})\right\} + k \ln \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right).$$

Utilizing (3.40) we obtain

$$\mu_{\mathcal{I}_{lb}} = \frac{1}{\ln 2}\left(\sum_{i=0}^{k-1} \psi(k+d-i) + k \ln \frac{\rho(1-\sigma_e^2)}{N_t(1+\rho\sigma_e^2)}\right).$$

The derivation of M_{ub} and hence of $\mu_{\mathcal{I}_{ub}}$ is precisely the same as in the proof of Theorem 3.3. The variance is found by utilizing the chi squared independence again along with $Var(x+a) = Var(x)$ to yield

$$\begin{aligned}
\sigma_{\mathcal{I}}^2 &= \sum_{i=0}^{k-1} Var\left\{\ln(\chi^2_{2(k+d-i)})\right\} \\
&= \sum_{i=0}^{k-1} \left(\mathbb{E}\ln(\chi^2_{2(k+d-i)})^2 - \mu^2_{\mathcal{I}_{lb}}\right).
\end{aligned}$$

Observing that [34]

$$\int_0^\infty \frac{\mu^\nu x^{\nu-1} \exp(-\mu x) \ln x^2}{\Gamma(\nu)} dx = (\psi(\nu) - \ln 2)^2 + \zeta(2,\nu)$$

we obtain

$$\sigma_{\mathcal{I}}^2 = \frac{1}{(\ln 2)^2} \sum_{i=0}^{k-1} \zeta(2, k+d-i)$$

which is the desired result. □

Since (3.47)-(3.49) are difficult to simplify, expressions for the outage capacity are omitted. It is worth mentioning that with the aid of (3.46) the estimate

$$\sum_{i=0}^{k-1} \psi(k+d-i) \leq k \ln l$$

holds and when inserted into (3.47) yields

$$\mu_{\mathcal{I}_{lb}} \leq k \left(\ln \left(\frac{l}{N_t} \right) + \ln \rho \right)$$

which implies that the outage capacity increases linearly with the minimum number of antenna elements and is logarithmic with an increase in SNR, consistent with Figure 3.2.

To illustrate these results (3.13)-(3.15) are plotted against the high effective SNR approximations for $N_t = 2, N_r = 5$ and $\sigma_e = 0.001$ in Figure 3.7. Comparing these new approximations with the exact moments, there is close agreement. These moments are then used to calculate the outage capacity bounds for various P in Figure 3.8 which again are in close proximity.

3.4. PILOT SYMBOL ASSISTED MODULATION

In this section the estimation error of a Rayleigh fast fading channel is analyzed for a MIMO system equipped with PSAM [38, 39]. Each transmit antenna sends length T_b symbol packets, N_t of which are pilots. Let $D \triangleq T_b - N_t$ denote the number of data symbols. For each transmit antenna, a spreading code of length N_t is defined by [39]

$$\boldsymbol{c}_n \triangleq [c_n(0) \ldots c_n(N_t - 1)]^T, \quad n = 1, \ldots, N_t$$

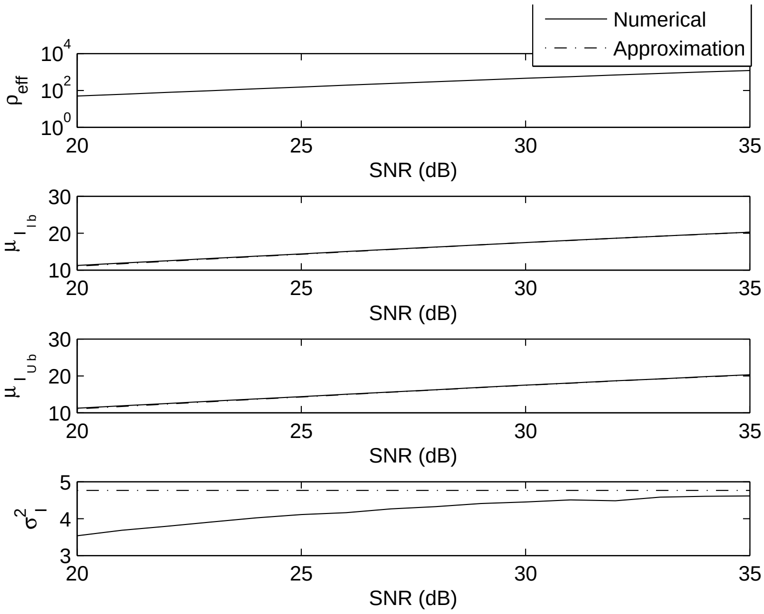

Figure 3.7 Comparison of numerical and approximate moments for the mutual information at high effective SNR when $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.0001$.

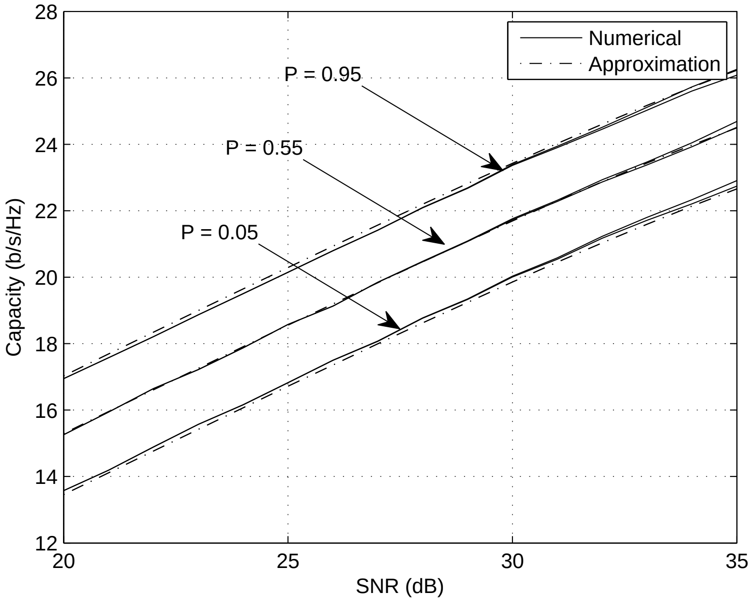

Figure 3.8 Approximations for the MIMO outage capacity with channel uncertainty at high effective SNR for various P when $N_t = 2, N_r = 5$, and $\sigma_e^2 = 0.0001$.

with the property

$$\boldsymbol{c}_n^H \boldsymbol{c}_m = \delta_k(n-m)$$

During the i^{th} packet the pilot symbols are spread by

$$\boldsymbol{x}_n(i) = x_n(i)\boldsymbol{c}_n, \quad n = 1, \ldots, N_t.$$

from which

$$\boldsymbol{y}_m(i) = \sum_{n=1}^{N} \boldsymbol{x}_n(i) h_{mn}(i) + \boldsymbol{n}, \quad m = 1, \ldots, N_r$$

where $\boldsymbol{y}_m(i) = [y_m(iT_b) \ldots y_m(iT_b + (N_t - 1))]^T$ and $\boldsymbol{n}$ is a noise vector with iid $\mathcal{CN}(0, N_o)$ entries. The receiver calculates the crude estimates

$$\widetilde{h}_{mn}(i) = \boldsymbol{c}_n^H \boldsymbol{y}_m = h_{mn}(i) + \boldsymbol{c}_n^H \boldsymbol{n}/x_n(i). \tag{3.51}$$

and feeds them to a M^{th} order Wiener filter to estimate the channel during data transmission [38].

Letting $P = 0.05$, $f_d T_s = 0.1$, $N_t = 2, N_r = 5$, $T = 10$, and $M = 10$ the analytic bounds are plotted in Figure 3.9 for various sampling periods after channel estimation. The outage capacity deviates substantially from the case of perfect CSI when the elapsed time exceeds one sample period. At elapsed time $6T_s$, the outage capacity saturates over the SNR range, indicating a high estimation error. This is because for fast fading systems the symbol period should be much less than the coherence time which is inversely proportional to the doppler spread. A popular rule of thumb for modern wireless systems is [5]

$$f_d T_s \ll .423. \tag{3.52}$$

Since this condition is not satisfied for $f_d T_s = 0.1$, it is not surprising to see sub-par performance. Next the outage capacity bounds when $f_d T_s = 0.005$ are plotted in Figure

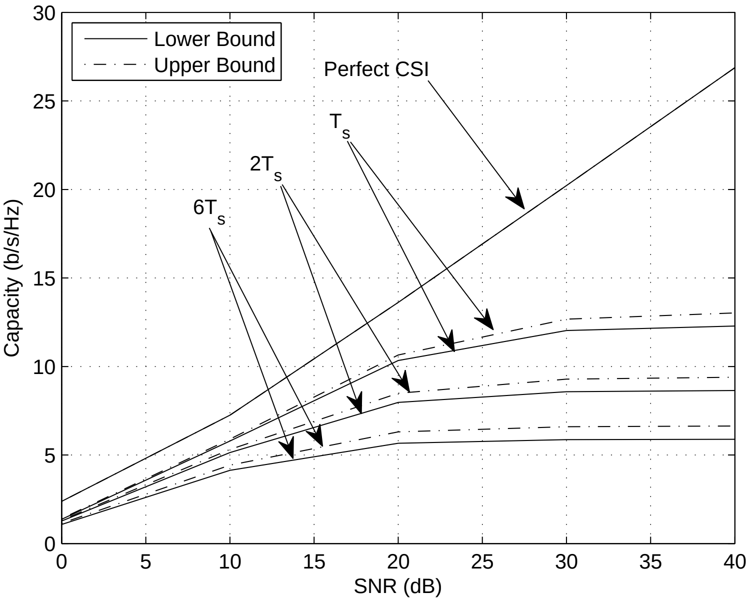

Figure 3.9 Analytic bounds for the outage capacity using PSAM when $N_t = 2, N_r = 5$, $P = 0.05$ and $f_d T_s = 0.1$.

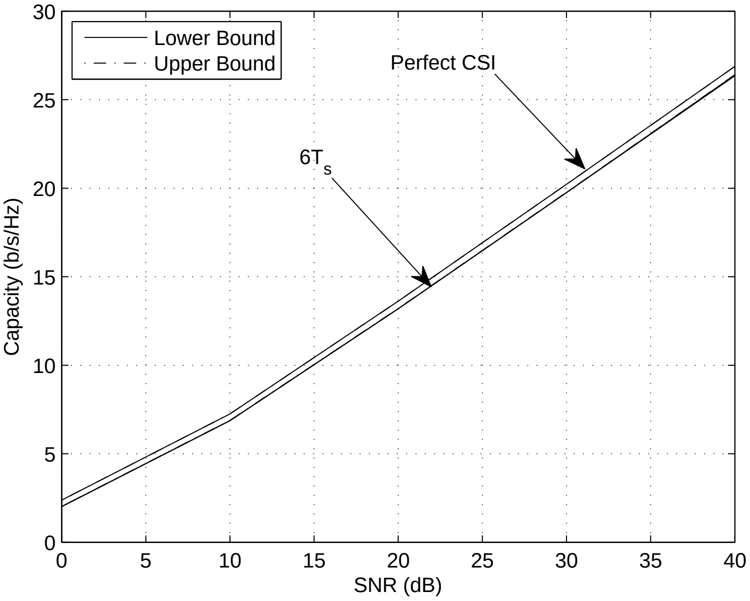

Figure 3.10 Analytic bounds for the outage capacity using PSAM when $N_t = 2, N_r = 5$, $P = 0.05$ and $f_d T_s = 0.005$.

3.10. The outage capacity remains in close proximity to the case of perfect CSI up until a SNR of 30 dB. This is expected since (3.52) is now satisfied.

3.5. FREQUENCY SELECTIVE CHANNEL

The analytical bounds for the outage capacity of the frequency selective Rayleigh MIMO channel described in (4.1) can be derived by utilizing OFDM with B carriers to obtain the frequency flat channels

$$\boldsymbol{\mathcal{H}}(\eta,\omega) \triangleq \sum_{i=0}^{L-1} \boldsymbol{H}(i)\exp(-j2\pi i(\omega/B)),\ \omega = 1,\ldots,B.$$

The bandwidth of each sub-band is chosen to be the coherence bandwidth of $\{\boldsymbol{H}(\eta)\}_{\eta=1}^{L}$. It follows from the linearity of the DFT that the each channel is complex Gaussian. Direct calculation of the covariance matrices corresponding to each column of $\boldsymbol{\mathcal{H}}(\eta,\omega)$ with (3.3) in mind verifies the sub-channels of $\boldsymbol{\mathcal{H}}(\eta,\omega)$ are $\mathcal{CN}(0,1)$. Similar to the time domain analysis, the channel estimation error for a MMSE estimate retains its orthogonality[40]

$$\sigma_{\mathcal{E}}^2 = 1 - \sigma_{\widehat{\mathcal{H}}}^2.$$

The upper and lower bounds of the mutual information are

$$\begin{aligned}
\mathcal{I}_{lb}(\eta) &= \frac{1}{B}\sum_{\omega=1}^{B}\log_2\det\left(\boldsymbol{I}_{N_r} + \frac{\rho(1-\sigma_{\mathcal{E}}^2)}{N_t(1+\rho\sigma_{\mathcal{E}}^2)}\underline{\widehat{\boldsymbol{H}}}(\eta,\omega)\ \underline{\widehat{\boldsymbol{H}}}(\eta,\omega)^H\right)\\
\mathcal{I}_{ub} &= \mathcal{I}_{lb} + N_r\mathbb{E}_{\boldsymbol{x}}\left\{\log_2\frac{\rho\sigma_{\mathcal{E}}^2+1}{\frac{\rho_\tau}{N_t}\sigma_{\mathcal{E}}^2\|\boldsymbol{x}\|_2^2+1}\right\}
\end{aligned}$$

Since the bins are chosen so that each channel sub-band is independent, it immediately follows that $\mathcal{I}_{lb} \sim \mathcal{N}(\mu_{\mathcal{I}_{lb}},\sigma_{\mathcal{I}}^2)$ and $\mathcal{I}_{ub} \sim \mathcal{N}(\mu_{\mathcal{I}_{ub}},\sigma_{\mathcal{I}}^2)$, which verifies that the outage capacity for a frequency selective channel maintains the same form as the flat fading case. The claim that application of the results in Sections 3 and 4 to frequency selective channels is hereby established.

3.6. CONCLUSION

Analytic upper and lower bounds for the outage capacity of a MIMO Rayleigh flat fading wireless channel with no CSI at the transmitter and channel estimation at the receiver were derived. Numerical examples illustrated that the outage capacity increased with the number of antennas, received SNR, and outage probability. They also became saturated at high received SNR when the estimation error became significant. To solidify these behaviors with analysis, the bounds were approximated for systems operating in the low and high effective SNR regimes. For low effective SNR, new accurate approximations for the mean, variance, and outage capacity were derived and were shown to behave similar to the MIMO ergodic capacity with channel estimation at the receiver. For the high effective SNR case, new accurate closed form expressions for the mean and variance were derived and the outage capacity was shown to be logarithmic with increased received SNR. We next considered how the outage capacity was affected by the doppler frequency using PSAM. The outage capacity was shown for a fast fading channel to significantly degrade, while for a slow fading channel it was shown that the outage capacity remained close to the perfect CSI case for a wide range of received SNR.

The outage capacity was then investigated for a frequency selective fading channel. Since the bandwidth of each frequency sub-band was chosen to be the coherence bandwidth, the bounds had the same form as a frequency flat channel.

4. MIMO CHANNEL PREDICTION USING RECURRENT NEURAL NETWORKS

4.1. INTRODUCTION

Multiple-input multiple-output (MIMO) wireless communication systems have recently received a considerable amount of attention [1, 41]. There have been numerous papers which report promising gains in capacity and diversity [26]. These results all stem from the assumption that the receiver and/or transmitter have perfect knowledge of the channel. In a typical telemetry application this is not the case and the receiver must settle for an estimate of the channel. Example channel estimation techniques include sending pilot symbols in bursts [42], periodically, [39], or in space time block codes [43].

Accurate channel estimation is often crucial for satisfactory decoding performance at the receiver. Additional performance improvements can be obtained using adaptive modulation techniques if channel state information (CSI) is available at the transmitter. If full CSI is available, the water-filling algorithm can be applied to achieve capacity [1] or pre-coding can be performed on transmitted symbols to achieve a better symbol error rate (SER) [44] . With partial CSI, the transmitter can adapt the modulation scheme to maximize spectral efficiency given a minimum SER specification [45].

A feedback link between the receiver and transmitter is commonly used to obtain CSI at the transmitter. When a channel undergoes fast fading (for example, from doppler shifts due to transmitter/receiver motion), CSI may become outdated quickly, prohibiting the transmitter from adapting correctly. Performance can be improved in this case by sending back a *prediction* of the CSI.

Previous works have used linear prediction techniques to predict CSI. In many telemetry applications, however, the received signal is corrupted by non-linear effects such as distortion from power amplification. In these situations non-linear prediction techniques become appropriate. This was originally done in [46] for multi-layer perceptron (MLP) neural networks. This was then extended to recurrent neural networks (RNN) for channel prediction in [19]. All of these approaches are valid for single-input single-output (SISO) systems. The contribution of this work is to extend these prediction results to the MIMO case.

The next section describes the input-output MIMO signal model. Section 4.3 provides the channel estimate/prediction model. The RNN used for channel estimation and the RNN weight update equations are provided in Sections 4.4 and 4.5 respectively. Finally, MIMO channel prediction results using the proposed RNN are presented, followed by a conclusion summarizing the work.

4.2. INPUT-OUTPUT DESCRIPTION

A MIMO wireless baseband communication system with N_t transmit antennas and N_r receive antennas is described at discrete time k by

$$\boldsymbol{r}(k) = \boldsymbol{H}(k)\boldsymbol{s}(k) + \boldsymbol{n}(k),$$

where $\boldsymbol{r}$ is the $N_r \times 1$ received vector, $\boldsymbol{s}$ is the $N_t \times 1$ transmitted symbol vector with each s_i belonging to constellation $\mathcal{C}$., and $\boldsymbol{n}$ is the white noise vector of size $N_r \times 1$ with $n_i \overset{iid}{\sim} \mathcal{CN}(0, N_o)$. The $N_r \times N_t$ channel matrix $\boldsymbol{H}(k)$ is modeled by

$$\boldsymbol{H}(k) = \boldsymbol{f}\Big(\sqrt{\alpha(k)}\boldsymbol{G}(k) + \sqrt{1-\alpha(k)}\boldsymbol{W}(k)\Big), \tag{4.1}$$

where $\boldsymbol{f} : \mathbb{C}^{N_r} \longrightarrow \mathbb{C}^{N_r}$ is a bounded function, $0 < \alpha(k) < 1$, $\boldsymbol{W}(k)$ is a noise matrix with $w_{mn}(k) \overset{iid}{\sim} \mathcal{CN}(0, \sigma_w^2)$, and $g_{mn}(k)$ is the gain between the m^{th} receiver and n^{th} transmitter with temporal correlation modeled by Jakes [3]. Specifically, each sub-channel autocorrelation function satisfies

$$R_{g_{mn}g_{mn}}(\tau) = J_0(2\pi f_d T_s \tau), \quad \tau = 0, 1, \dots$$

where f_d is the maximum doppler frequency and T_s is the symbol period. Denoting P_s as the total transmit power, the transmitted symbols must satisfy

$$\mathbb{E}||\boldsymbol{s}(k)|| \tag{4.2}$$

4.3. CHANNEL ESTIMATION

Accurate prediction of future CSI requires knowledge of channel fading statistics. Following the same procedure in [12], [13], and [14], the minimum mean squared error (MMSE) estimate obtained at discrete time $i \triangleq N_e k$ is modeled as

$$\widetilde{\boldsymbol{H}}(i) = \boldsymbol{H}(i) + \widetilde{\boldsymbol{E}}(i), \tag{4.3}$$

where $\widetilde{\boldsymbol{H}}(i)$ and $\widetilde{\boldsymbol{E}}(i)$ are uncorrelated with $\widetilde{h}_{mn}(k) \sim \mathcal{CN}(0, \sigma_{\widetilde{h}}(k)^2)$ and $\widetilde{e}_{mn}(k) \sim \mathcal{CN}(0, \sigma_w(k)^2)$. The channel estimation error between the m^{th} receiver and n^{th} transmitter is thus

$$\sigma_{\widetilde{e}}^2 = |1 - \sigma_{\widetilde{h}}^2(k)|.$$

Taking the vec$(\cdot)$ of both sides of (4.3) yields

$$\widetilde{\boldsymbol{h}}(i) = \boldsymbol{h}(i) + \widetilde{\boldsymbol{e}}(i).$$

This form will be convenient when discussing the operation of the RNN predictor. Observing the block diagram in Figure 4.1, the N_p most recent estimates are used for the prediction.

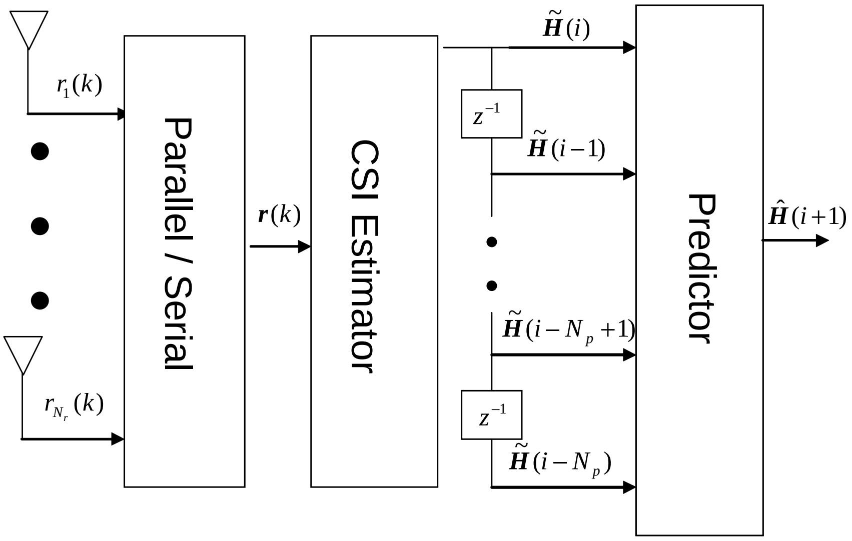

Figure 4.1 Block diagram of MIMO channel predictor.

4.4. A RECURRENT NEURAL NETWORK

A RNN with n input neurons, m hidden neurons, and r output neurons shown in Figure 4.2 is constructed for non-linear MIMO channel prediction. All outputs of the hidden layers are fed back to the input layer. The activation functions are described by the mapping $\boldsymbol{\Phi} : \mathbb{C}^m \longrightarrow \mathbb{C}^m$ with real and imaginary components defined by

$$\boldsymbol{\Phi}(\cdot) = \boldsymbol{\phi}(\cdot) + j\boldsymbol{\phi}(\cdot) = \tanh(\cdot) + j\tanh(\cdot).$$

Since the neural network is implemented on a signal processing unit (with sampling period possibly different than the receiver analog to digital converter (ADC)), the discrete time index n (not to be confused with the number of input neurons) will be employed. Using the equivalent state space representation model in [47] the input-output relationship can be extended to MIMO systems by

$$\begin{aligned} \boldsymbol{y}(n+1) &= \boldsymbol{C}\boldsymbol{x}(n+1) \\ \boldsymbol{x}(n+1) &= \boldsymbol{\Phi}(\boldsymbol{W}_x(n)\boldsymbol{x}(n) + \boldsymbol{W}_u(n)\boldsymbol{u}(n)), \end{aligned}$$

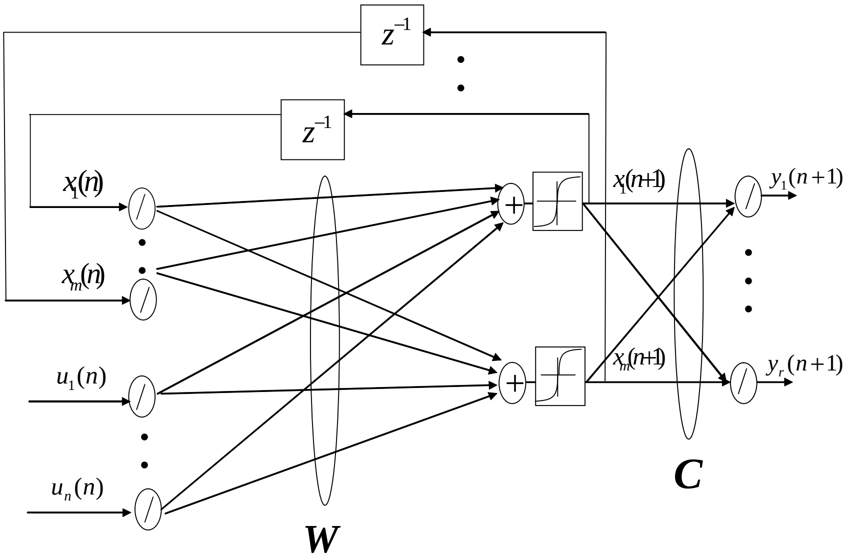

Figure 4.2 State space representation of a recurrent neural network.

where $\boldsymbol{C}$ is a $r\times m$ real valued matrix and $\boldsymbol{W}_x(n)$ and $\boldsymbol{W}_u(n)$ are complex valued neural network weight matrices of size $m\times m$ and $m\times n$ respectively. Letting $\boldsymbol{W}(n) = [\boldsymbol{W}_x(n)\boldsymbol{W}_u(n)]$ and $\boldsymbol{z}(n) = [\boldsymbol{x}(n)\,\boldsymbol{u}(n)]^T$ the real (I) and imaginary (Q) components of the state vector are

$$\boldsymbol{x}(n+1) = \boldsymbol{\phi}\Big(\boldsymbol{W}^I(n)\boldsymbol{z}^I(n) - \boldsymbol{W}^Q(n)\boldsymbol{z}^Q(n)\Big) + j\boldsymbol{\phi}\Big(\boldsymbol{W}^I(n)\boldsymbol{z}^Q(n) + \boldsymbol{W}^Q(n)\boldsymbol{z}^I(n)\Big),$$

which is the desired form for deriving the RNN weight updates.

4.5. MIMO RNN WEIGHT UPDATES

For accurate channel prediction, the neural network weights should minimize an appropriate cost function. The squared error between the neural network prediction and the most recent channel estimate is proposed as

$$J(n) = \frac{1}{2}\boldsymbol{e}^H(n+1)\boldsymbol{e}(n+1), \tag{4.4}$$

where

$$\boldsymbol{e}(n+1) \triangleq \widetilde{\boldsymbol{h}}(n) - \boldsymbol{C}\boldsymbol{x}(n+1). \tag{4.5}$$

Although this is not the actual prediction error, it provides a realistic way of measuring the performance, as opposed to [19] which assumes the instantaneous channel is available when calculating this error. The real time recurrent learning (RTRL) and Extended Kalman Filter (EKF) algorithms are considered for training the RNN weights. The RTRL seeks to minimize the instantaneous value of (4.4) whereas the EKF minimizes the expected value of (4.4).

Let $\boldsymbol{w}_i$ be the i^{th} column of $\boldsymbol{W}^T$. The weight updates for the RTRL algorithm are [18, 47]

$$\begin{aligned}\Delta_{\boldsymbol{w}_i}(n+1) &= \gamma_g[(\boldsymbol{e}^I(n))^T\boldsymbol{C} \quad (\boldsymbol{e}^Q(n))^T\boldsymbol{C}]\begin{bmatrix}\boldsymbol{\Lambda}_i^{II}(n)+j\boldsymbol{\Lambda}_i^{IQ}(n)\\ \boldsymbol{\Lambda}_i^{QI}(n)+j\boldsymbol{\Lambda}_i^{QQ}(n)\end{bmatrix}+\gamma_m\boldsymbol{w}_i(n)\\ \boldsymbol{w}_i(n+1) &= \Delta_{\boldsymbol{w}_i}(n+1)+\boldsymbol{w}_i(n),\end{aligned}$$

where γ_g is the learning gain, γ_m is the momentum gain, and Λ^{AB} satisfies the following recursion

$$\begin{bmatrix}\boldsymbol{\Lambda}_i^{II}(n+1) & \boldsymbol{\Lambda}_i^{IQ}(n+1)\\ \boldsymbol{\Lambda}_i^{QI}(n+1) & \boldsymbol{\Lambda}_i^{QQ}(n+1)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\phi} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\phi}\end{bmatrix}\begin{bmatrix}\boldsymbol{W}_x^I(n) & -\boldsymbol{W}_x^Q(n)\\ \boldsymbol{W}_x^Q(n) & \boldsymbol{W}_x^I(n)\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_i^{II}(n) & \boldsymbol{\Lambda}_i^{IQ}(n)\\ \boldsymbol{\Lambda}_i^{QI}(n) & \boldsymbol{\Lambda}_i^{QQ}(n)\end{bmatrix} + \begin{bmatrix}\boldsymbol{z}_i^I(n) & -\boldsymbol{z}_i^Q(n)\\ \boldsymbol{z}_i^Q(n) & \boldsymbol{z}_i^I(n)\end{bmatrix}$$

with the initialization

$$\begin{bmatrix}\boldsymbol{\Lambda}_i^{II}(0) & \boldsymbol{\Lambda}_i^{IQ}(0)\\ \boldsymbol{\Lambda}_i^{QI}(0) & \boldsymbol{\Lambda}_i^{QQ}(0)\end{bmatrix} = \boldsymbol{0}.$$

Before the weight updates for the EKF are given, the complex Jacobian in [18] is corrected to

$$\begin{aligned}\boldsymbol{\Lambda}(n+1) &= \frac{\partial(\boldsymbol{x}(n+1))}{\partial(\boldsymbol{w}^I(n))}+j\frac{\partial(\boldsymbol{x}(n+1))}{\partial(\boldsymbol{w}^Q(n))}\\ &= \boldsymbol{\Lambda}_i^{II}(n)-\boldsymbol{\Lambda}_i^{QQ}(n)+j\Big(\boldsymbol{\Lambda}_i^{QI}(n)+\boldsymbol{\Lambda}_i^{IQ}(n)\Big).\end{aligned}$$

This is applied to the EKF weight updates

$$\begin{aligned}
\boldsymbol{\Gamma}(n) &= [\boldsymbol{\Lambda}(n)\boldsymbol{P}(n)\boldsymbol{\Lambda}^H(n) + \boldsymbol{R}(n)]^{-1} \\
\boldsymbol{K}(n) &= \boldsymbol{P}(n)\boldsymbol{\Lambda}^H(n)\boldsymbol{\Gamma}(n) \\
\boldsymbol{w}_i(n+1) &= \boldsymbol{w}_i(n) + \boldsymbol{K}(n)\boldsymbol{e}(n) \\
\boldsymbol{P}(n+1) &= \boldsymbol{P}(n) - \boldsymbol{K}(n)\boldsymbol{\Lambda}(n)\boldsymbol{P}(n) + \boldsymbol{Q}(n)
\end{aligned}$$

with the following initializations

$$\begin{aligned}
\boldsymbol{R}(0) &= \mu^{-1}(\boldsymbol{I} + j\boldsymbol{I}) \\
\boldsymbol{Q}(0) &= \rho(\boldsymbol{I} + j\boldsymbol{I}) \\
\boldsymbol{P}(0) &= \epsilon^{-1}(\boldsymbol{I} + j\boldsymbol{I}).
\end{aligned}$$

A contribution of this work is the ability to predict all $N_r N_t$ sub-channels simultaneously by letting

$$\begin{aligned}
\boldsymbol{y}(n+1) &= \widehat{\boldsymbol{h}}(n+1) \\
\boldsymbol{u}(n) &= \begin{bmatrix} \widetilde{\boldsymbol{h}}(n) & \widetilde{\boldsymbol{h}}(n-1) & \cdots & \widetilde{\boldsymbol{h}}(n-N_p) \end{bmatrix}^T.
\end{aligned}$$

4.6. DELAYED PREDICTION

In certain situations it may be unrealistic to assume that the RNN predictor has instantaneous access to the MMSE estimates. In this case the prediction error and RNN inputs are only updated every N_s samples and the RNN output is the N_s sample delayed prediction.

4.7. PREDICTION RESULTS

This section presents numerical examples to compare the tracking capability of the EKF and RTRL algorithms. Parameters held constant are given in Table 4.1. The MSE(SSE) of the EKF and RTRL algorithms is shown in Figure 4.3 when $N_s = 1$. The EKF outperforms the RTRL algorithm and reaches an average MSE that is two orders of magnitude smaller. The comparison between the actual channel coefficients and their predictions for each MIMO sub-channel are shown in Figures 4.4-4.7. These plots verify the EKF predictions match the actual channel much better than the RTRL predictions.

Next, the two training algorithms are compared in a delayed prediction scenario for $N_s = 10$. Once again the MSE indicates the EKF performance is superior to the RTRL as seen in Figure 4.8. The comparison between the actual channel coefficients and their predictions for each MIMO sub-channel are shown in Figures 4.9-4.12. These plots verify the EKF predictions are able to track the channel reasonably well in comparison to the RTRL algorithm when the RNN weights are only updated every 10 sampling periods.

An example comparing the EKF neural network predictor to a linear predictor using the Levinson-Durbin recursion is considered next. The fixed parameters are given in Table 4.2. The non-linear channel is described by (4.1) when $\boldsymbol{f}(\cdot) = [\tanh(\cdot) \quad \cdots \quad \tanh(\cdot)]^T$, $f_d T_s = 0.1$, and $\alpha(k)$ takes on the values zero or one for random time intervals. The noise variance is $\sigma_w^2 = 0.1$. The MSE comparison in Figure 4.13 indicates that the EKF predictor slightly outperforms the L-D approach. The real and imaginary components of the

Table 4.1 Parameters Values for channel tracking example

Parameter	Value	Description
$f_j(a)$	a	No Distortion
N_r	2	Receive Antennas
N_t	2	Transmit Antennas
N_p	5	Previous Channel Estimates
N_e	1	Samples Between Estimation
σ_e^2	.001	Estimation Error
f_d	500 Hz	Doppler Frequency
T_s	10 msec	Symbol Period
T_h	1	Prediction step
μ	.1	EKF measurement noise parameter
ρ	.1	EKF process noise parameter
ϵ	10	EKF error covariance parameter
γ_g	.001	RTRL learning gain
γ_m	.3	RTRL momentum gain

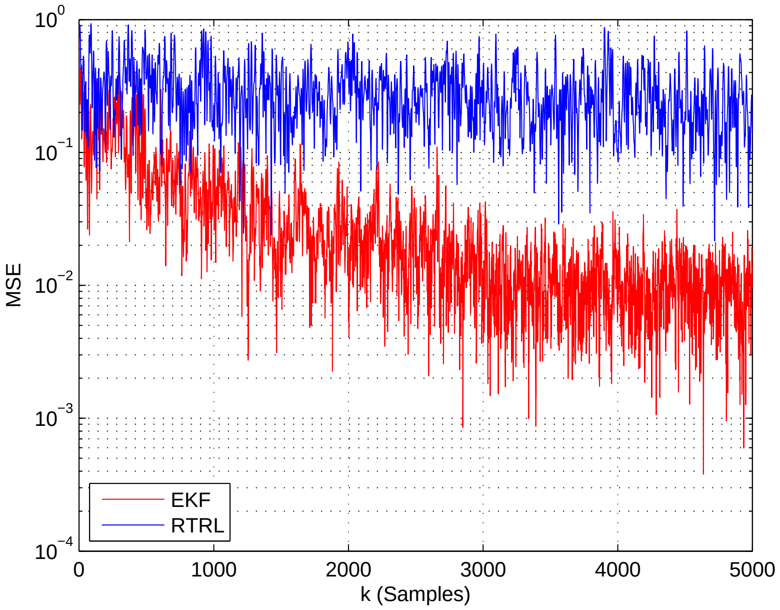

Figure 4.3 Mean squared error between EKF and RTRL algorithms when $N_s = 1$.

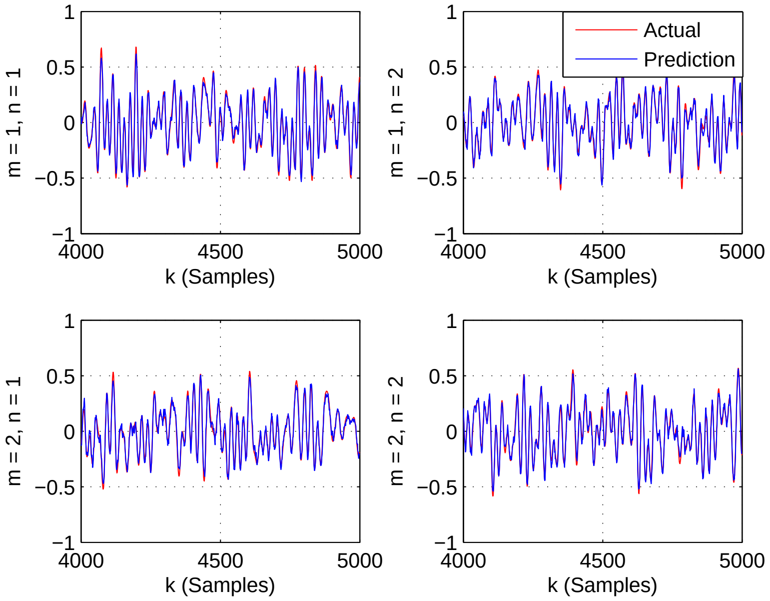

Figure 4.4 Real component of MIMO channel coefficients using EKF algorithm when $N_s = 1$.

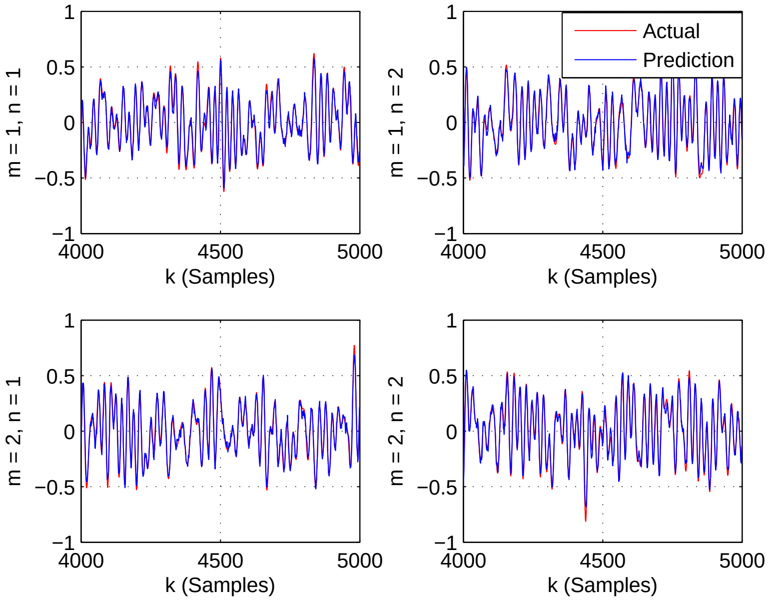

Figure 4.5 Imaginary component of MIMO channel coefficients using EKF algorithm when $N_s = 1$.

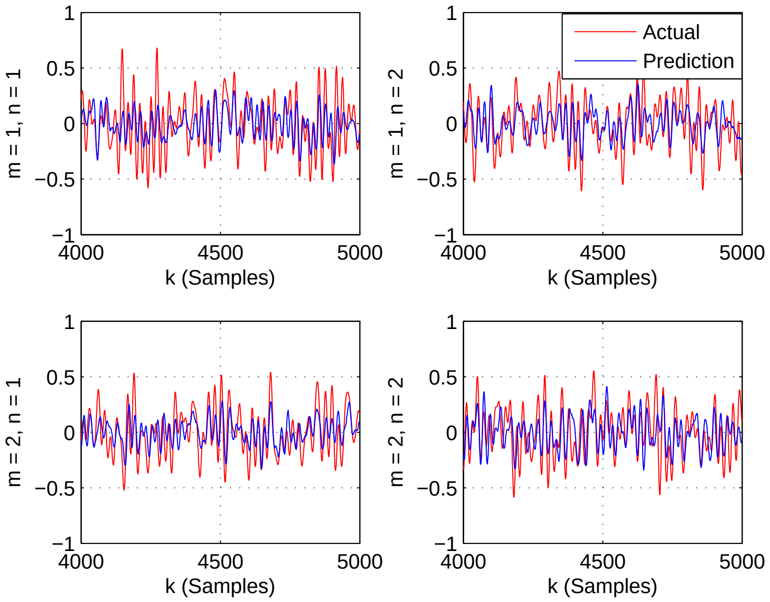

Figure 4.6 Real component of MIMO channel coefficients using RTRL algorithm when $N_s = 1$.

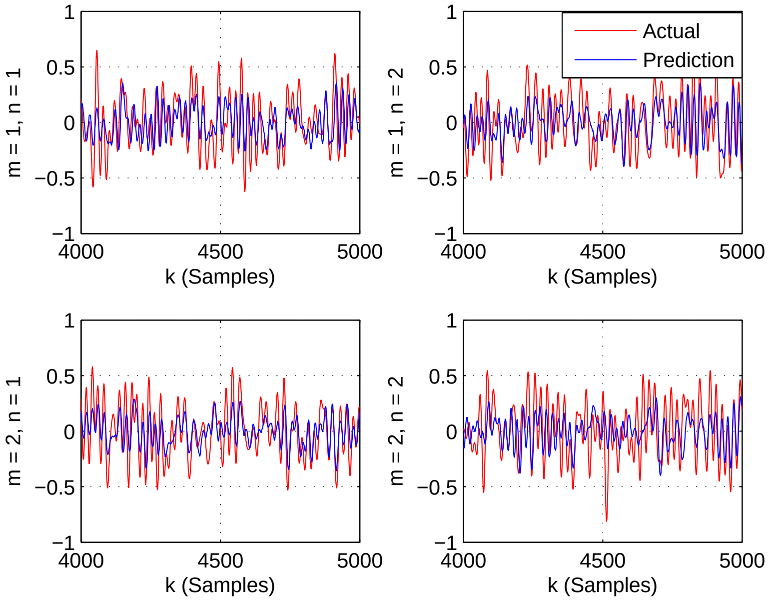

Figure 4.7 Imaginary component of MIMO channel coefficients using RTRL algorithm when $N_s = 1$.

Table 4.2 Parameters Values for non-linear channel example

Parameter	Value	Description
$f_j(a)$	$\tanh(a)$	Non-linear Distortion
N_r	2	Receive Antennas
N_t	2	Transmit Antennas
N_p	5	Previous Channel Estimates
N_e	1	Samples Between Estimation
σ_e^2	.001	Estimation Error
T_s	10 msec	Symbol Period
μ	.1	EKF measurement noise parameter
ρ	.1	EKF process noise parameter
ϵ	10	EKF error covariance parameter

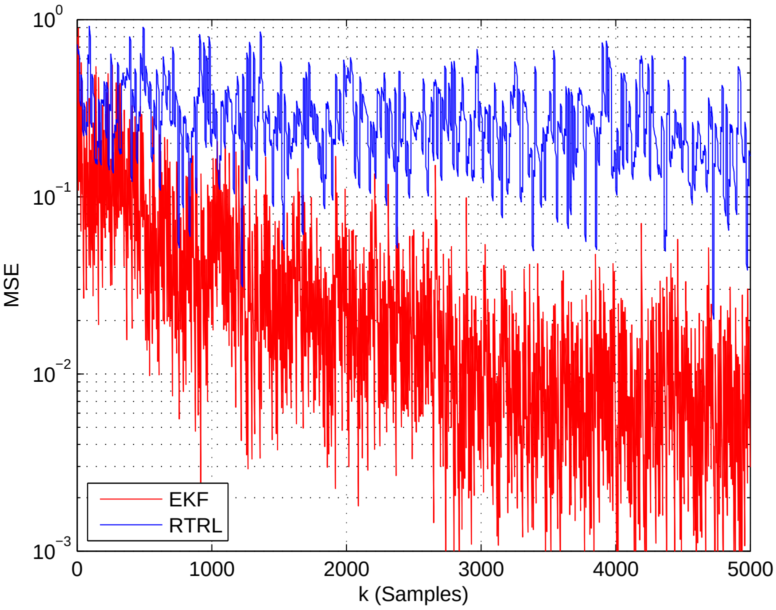

Figure 4.8 Mean squared error between EKF and RTRL algorithms when $N_s = 1$.

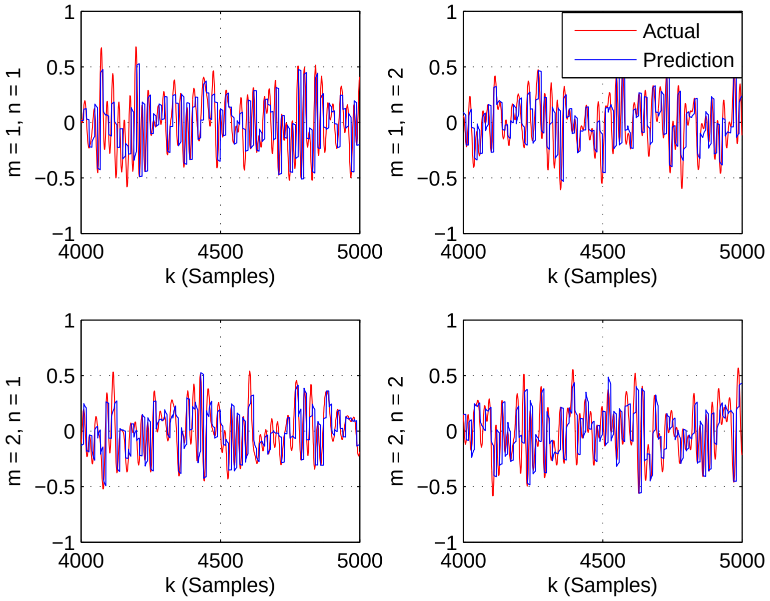

Figure 4.9 Real component of MIMO channel coefficients using EKF algorithm when $N_s = 10$.

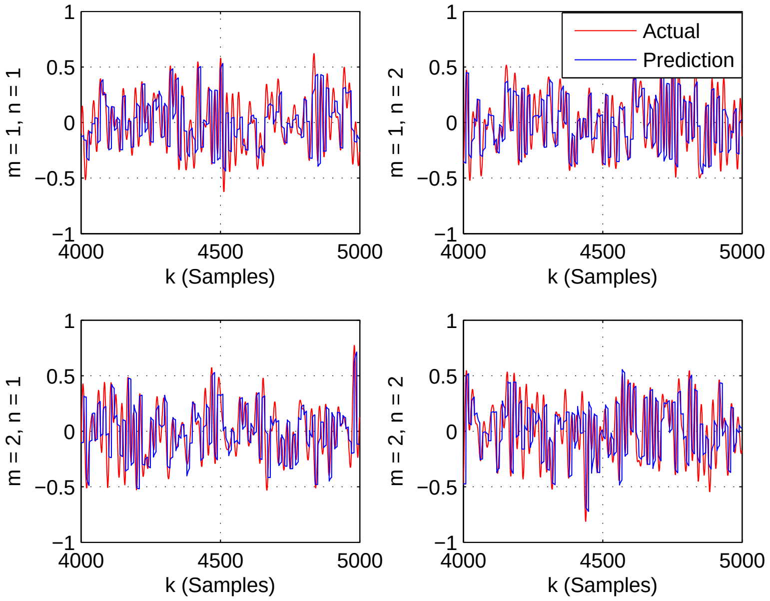

Figure 4.10 Imaginary component of MIMO channel coefficients using EKF algorithm when $N_s = 10$.

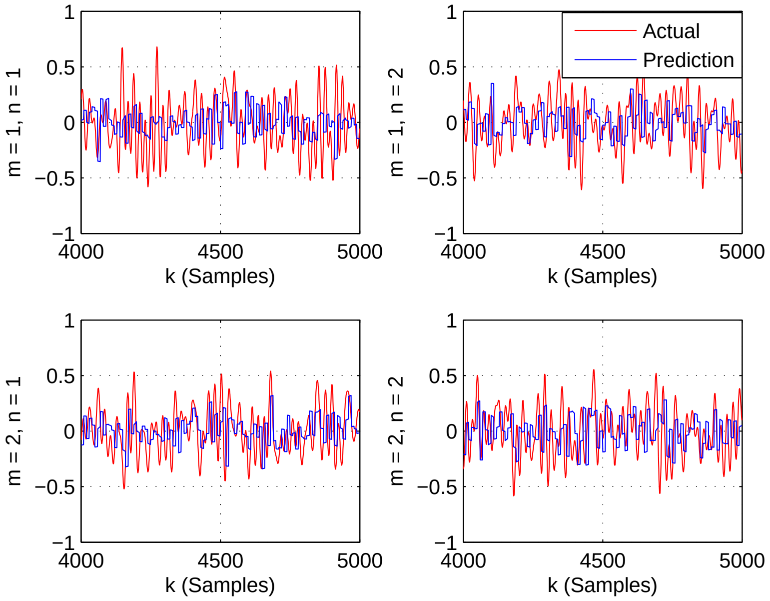

Figure 4.11 Real component of MIMO channel coefficients using RTRL algorithm when $N_s = 10$.

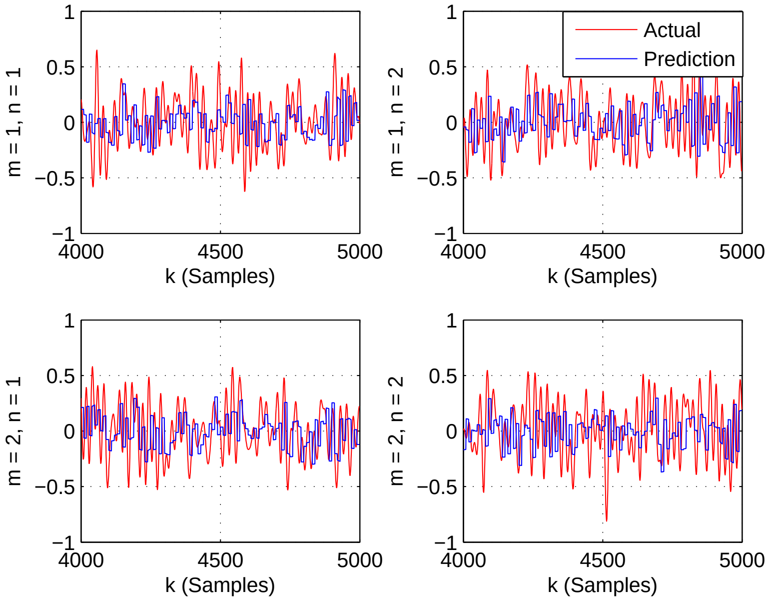

Figure 4.12 Imaginary component of MIMO channel coefficients using RTRL algorithm when $N_s = 10$.

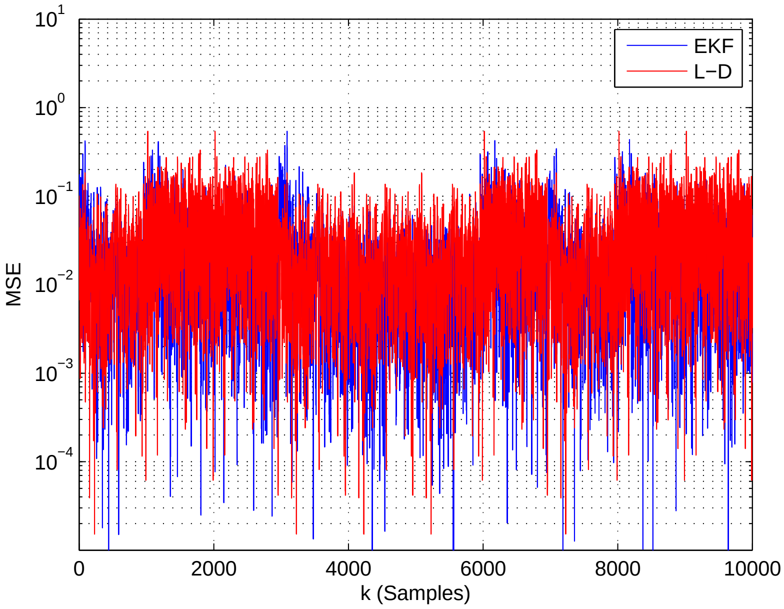

Figure 4.13 Mean squared error between EKF and L-D algorithm for a non-linear channel.

sub-channel corresponding to the first transmit and receive antennas are displayed in Figure 4.14. Unlike the linear predictor, the EKF neural network predictor is able to explore the non-linear correlations present in the channel.

4.8. CONCLUSION

Existing SISO prediction algorithms were extended to the MIMO case for the EKF and RTRL algorithms. Examples presented show the EKF algorithm outperformed the RTRL algorithm in both single and multi step prediction. When channel non-linearities are present, the neural network predictor using the EKF algorithm outperformed the Levinson-Durbin linear predictor.

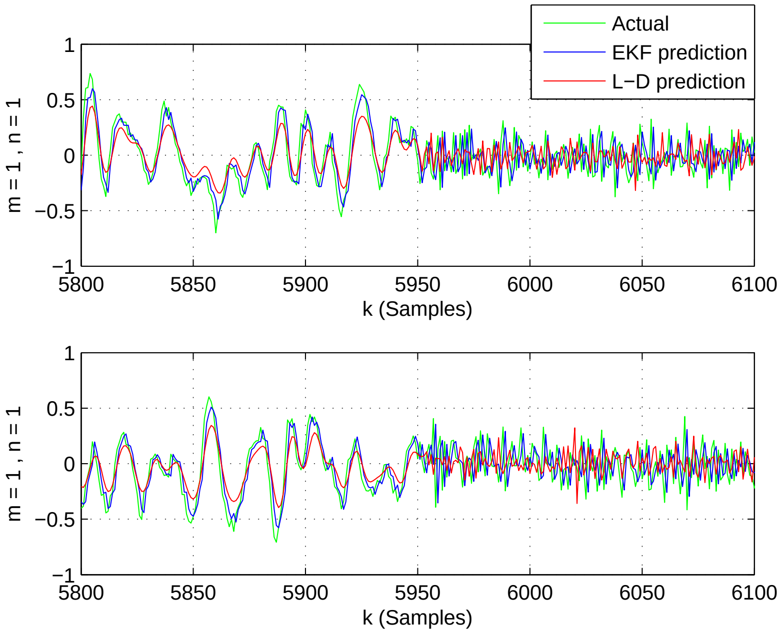

Figure 4.14 Real and imaginary components of a MIMO non-linear sub-channel.

5. MODELING CHANNEL ESTIMATION ERROR IN CONTINUOUSLY VARYING MIMO CHANNELS

5.1. INTRODUCTION

Multiple-input multiple-output (MIMO) wireless communication systems have the capability of providing a substantial increase in data rate [1, 48]. For optimal performance, one must assume that the receiver has perfect channel state information (CSI). When this is not the case, incorrect decoding at the receiver may result.

This situation motivates the use of pilot symbols to estimate the CSI. One can measure how close this estimate is to the actual channel coefficients by measuring the mean squared error (MSE) between the true channel and its estimate. In [42] the MSE was found for the Rayleigh block fading scenario, where the channel is assumed to remain constant for a block of transmitted symbols. This was extended to an arbitrary channel distribution in [49]. These works did not take into account different channel models or consider a channel that is time varying during the payload. Time variation was considered in [50], but the channel distribution for the diffuse component was assumed to be Rayleigh and no expression for the MSE was found.

In this work, a new expression for the MSE is derived for an arbitrary channel distribution that is continuously varying in time. From this result, optimal values for the training time are found and are shown to vary according to the temporal autocorrelation between channel coefficients. The rate of change of the MSE with received SNR and number of transmit and receive antennas is investigated. [5] Particular cases of the new MSE expression are shown to match previous results. The case of Rayleigh fading is presented to show how the channel estimation error can increase dramatically as time between training increases.

---

[5] Knowing how these parameters affect the MSE can be beneficial to the transmitter in a smart antenna environment or when the receiver is utilizing a low rate feedback loop to the transmitter.

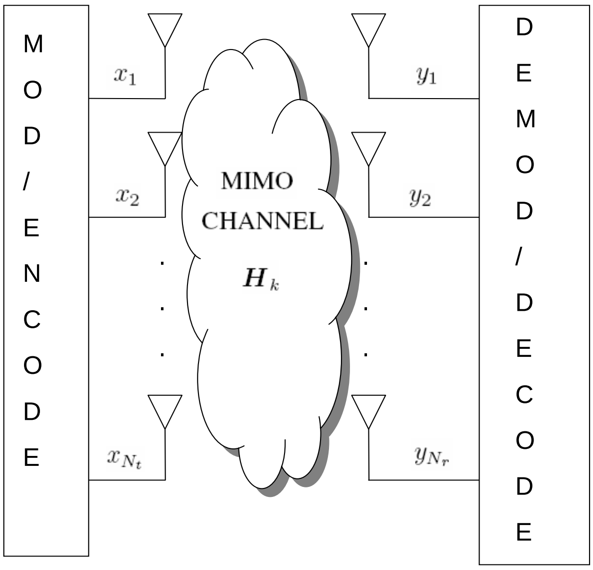

Figure 5.1 Block diagram for a MIMO system.

The rest of this work is presented as follows. The next section describes the mathematical models that are used to derive the main result. This is followed by previous results that relate to our work. A new expression for the channel estimation error for a continuously changing MIMO channel is next derived. Optimal values for the amount of time to spend training are found and the number of antennas and received SNR are varied to show how the rate of change of the MSE is affected. Special cases of our result are shown to yield expressions that have occurred in the literature. This is followed by our concluding remarks.

5.2. MATHEMATICAL MODELS

5.2.1. Received symbols. Let the received symbols of a baseband MIMO channel with N_t transmitters and N_r receivers be described at discrete time index k by

$$\boldsymbol{y}(k) = \sqrt{\frac{\rho}{N_t}}\boldsymbol{H}(k)\boldsymbol{x}(k) + \boldsymbol{n}(k) \tag{5.1}$$

where $\rho = \frac{E_s}{N_o}$ is the SNR with average symbol energy E_s and noise variance N_o, $\boldsymbol{y}(k)$ is the $N_r \times$ 1 received symbol vector, $\boldsymbol{H}(k)$ is the $N_r \times N_t$ flat fading MIMO channel matrix, $\boldsymbol{x}(k)$ is the $N_t \times$ 1 transmitted symbol vector with $\mathbb{E}||\boldsymbol{x}(k)||_2^2 = N_t$, and $\boldsymbol{n}(k)$ is the N_r

x 1 noise vector with elements $n_i(k) \overset{iid}{\sim} \mathcal{CN}(0,1), i = 1, \ldots, N_r$. A block diagram of the overall MIMO system is illustrated in Figure 5.1. Throughout this work assume that the antennas are spaced sufficiently apart at the transmitter and receiver so that $\boldsymbol{H}(k)$ is spatially uncorrelated. The channel matrix also satisfies

$$\frac{\mathbb{E}||\boldsymbol{H}(k)||_F^2}{N_t N_r} = 1 \tag{5.2}$$

where $||\cdot||_F$ is the Frobenius norm.

5.2.2. Training Phase. Suppose that each transmitter sends T_τ consecutive pilot symbols at the beginning of a T symbol block to form the $N_t \times T_\tau$ matrix $\boldsymbol{X}(\tau)$. As seen in Figure 5.2, the receiver obtains the noise corrupted symbols according to (5.1) which are sent to a buffer to obtain the $N_r \times T_\tau$ matrix $\boldsymbol{Y}(\tau)$. To make training feasible, assume that the channel, denoted by $\boldsymbol{H}(\tau)$, does not change throughout the training interval. The received symbols during the training period can be written as

$$\boldsymbol{Y}(\tau) = \sqrt{\frac{\rho_\tau}{N_t}}\boldsymbol{H}(\tau)\boldsymbol{X}(\tau) + \boldsymbol{N}(\tau)$$

where ρ_τ is the SNR during training and $\boldsymbol{N}(\tau)$ is the $N_r \times T_\tau$ noise matrix. The training matrix is restricted such that

$$||\boldsymbol{X}(\tau)||_F = T_\tau N_t.$$

5.2.3. Channel Variation. During the data phase assume the channel varies k symbols in advance from $\boldsymbol{H}(\tau)$ by [50]

$$\boldsymbol{H}(\tau+k) = \sqrt{\alpha_k}\boldsymbol{H}(\tau) + \sqrt{1-\alpha_k}\boldsymbol{W}(\tau+k)\,, \quad k = T_\tau + 1, \ldots, T.\,^{6} \tag{5.3}$$

The parameter α_k is deterministic but unknown and $\boldsymbol{W}(\tau+k)$ is a N_r x N_t matrix with $w_{mn}(\tau+k) \overset{iid}{\sim} \mathcal{CN}(0,1)$. Assuming that $\boldsymbol{W}(\tau+k)$, $\boldsymbol{N}(\tau)$, and $\boldsymbol{H}(\tau)$ are all uncorrelated

---

[6] The combining of τ and k is a slight abuse of notation but is effective at describing the channel variation once confusion is avoided.

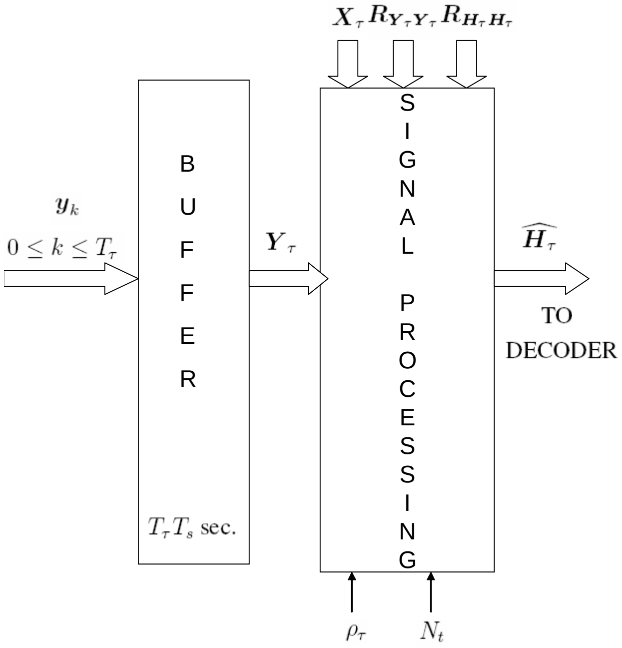

Figure 5.2 Block diagram of training scheme.

along with recognizing (5.3) as $N_r N_t$ independent k^{th} order autoregressive processes, it follows that

$$\alpha_k = \left[\frac{r_{hh}(kT_s)}{r_{hh}(0)}\right]^2$$

where T_s is the symbol period and $r_{hh}(\cdot)$ is the sub-channel autocorrelation function.

5.3. PREVIOUS RESULTS

In [42] the authors assumed *Rayleigh block fading*, where $\boldsymbol{H}(k) = \boldsymbol{H}(\tau)$ throughout the block and $h_{mn}(\tau) \overset{iid}{\sim} \mathcal{CN}(0,1)$. The MSE between $\boldsymbol{H}(\tau)$ and the linear minimum mean square error (LMMSE) estimate $\widehat{\boldsymbol{H}}(\tau)$ was shown to be

$$\sigma_{bf}^2 = \frac{1}{1 + \frac{\rho_\tau T_\tau}{N_t}}. \tag{5.4}$$

In [49] the MSE for the linear minimum mean squared error (LMMSE) estimate for an arbitrary channel distribution under block fading was found to be

$$\sigma_{bf}^2 = \frac{1}{\frac{\rho_\tau T_\tau}{N_t} + \frac{N_r}{N_t}\mathrm{trace}(\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}^{-1})}. \tag{5.5}$$

where

$$\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau} \triangleq \mathbb{E}\{\boldsymbol{H}(\tau)^H \boldsymbol{H}(\tau)\}. \tag{5.6}$$

These results do not take into account a general channel distribution that is time varying during the payload. A contribution of this work fulfills that task.

5.4. CHANNEL ESTIMATION ERROR OF CONTINUOUSLY VARYING MIMO CHANNEL

The MSE for a continuously varying MIMO channel is defined by

$$\sigma_{cf}^2(k) \triangleq \mathbb{E}\{\|\boldsymbol{H}(\tau+k) - \widehat{\boldsymbol{H}}(\tau)\|_F^2\} \tag{5.7}$$

where $\widehat{\boldsymbol{H}}_\tau$ is the LMMSE expressed by

$$\widehat{\boldsymbol{H}}_\tau = \sqrt{\frac{\rho_\tau}{N_t}} \boldsymbol{Y}_\tau \boldsymbol{R}_{\boldsymbol{Y}_\tau \boldsymbol{Y}_\tau} \boldsymbol{X}_\tau^H \boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}.$$

Expanding (5.7), we have

$$\begin{aligned}
\sigma_{cf}^2(k) &= \mathrm{trace}(\alpha_k \boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}) \\
&+ \mathrm{trace}((1-\alpha_k)N_r \boldsymbol{I}_{N_t}) + \mathrm{trace}(\frac{\rho_\tau}{N_t}(1-2\sqrt{\alpha_k})\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau} \boldsymbol{X}(\tau) \boldsymbol{R}_{\boldsymbol{Y}_\tau \boldsymbol{Y}_\tau}^{-1} \boldsymbol{X}(\tau)^H \boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}) \\
&= (\alpha_k + 1 - 2\sqrt{\alpha_k})\mathrm{trace}(\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}) + (1-\alpha_k)N_r N_t \\
&- (1-2\sqrt{\alpha_k})\mathrm{trace}([\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}^{-1} + \frac{\rho_\tau}{N_t N_t}\boldsymbol{X}(\tau)\boldsymbol{X}(\tau)^H]^{-1}).
\end{aligned} \tag{5.8}$$

We now find find a training matrix such that $\boldsymbol{X}_\tau \boldsymbol{X}_\tau^H$ is non-singular which minimizes (5.8). Since only the third term depends on $\boldsymbol{X}_\tau$, the minimization can be stated as

$$\arg\min_{\boldsymbol{X}_\tau^H} \operatorname{trace}([\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}^{-1} + \frac{\rho_\tau}{N_t N_t} \boldsymbol{X}(\tau) \boldsymbol{X}(\tau)^H]^{-1}) \tag{5.9}$$

such that

$$\operatorname{trace}(\boldsymbol{X}(\tau)\boldsymbol{X}(\tau)^H) = N_t T_\tau. \tag{5.10}$$

Setting the derivative of (5.9) with respect to $\boldsymbol{X}(\tau)^H$ equal to zero it can be shown that

$$\left[\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}^{-1} + \frac{\rho_\tau}{N_t N_r} \boldsymbol{X}(\tau) \boldsymbol{X}(\tau)^H\right]^{-2} = \lambda \boldsymbol{I}_{N_t}. \tag{5.11}$$

Invoking the eigenvalue decomposition of matrices we may write

$$\boldsymbol{X}(\tau)\boldsymbol{X}(\tau)^H = \boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^H - \frac{N_t N_r}{\rho_\tau} \boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}^{-1} \tag{5.12}$$

where $\boldsymbol{Q}$ is a unitary matrix and Γ is a diagonal matrix. Substituting (5.12) into (5.11) gives

$$\left[\frac{\rho_\tau}{N_r N_t} \boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^H\right]^{-2} = \lambda \boldsymbol{I}_{N_t}.$$

Solving for $\boldsymbol{\Gamma}$ gives

$$\boldsymbol{\Gamma} = \frac{N_t N_r}{\rho_\tau \sqrt{\lambda}} \boldsymbol{I}_{N_t} \tag{5.13}$$

Substituting (5.13) into (5.12) gives

$$\boldsymbol{X}(\tau)\boldsymbol{X}(\tau)^H = \frac{N_t N_r}{\rho_\tau \sqrt{\lambda}} \boldsymbol{I}_{N_t} - \frac{N_t N_r}{\rho_\tau} \boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}^{-1} \tag{5.14}$$

We now take the trace of both sides of (5.14) and invoke (5.10) to get

$$\frac{N_tN_r}{\rho_\tau\sqrt{\lambda}} = \frac{N_r}{\rho_\tau}\mathrm{trace}(\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau\boldsymbol{H}_\tau}) + T_\tau$$

which gives

$$\boldsymbol{X}(\tau)\boldsymbol{X}(\tau)^H = \left(\frac{N_r}{\rho_\tau}\mathrm{trace}(\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau\boldsymbol{H}_\tau}) + T_\tau\right)\boldsymbol{I}_{N_t} - \frac{N_tN_r}{\rho_\tau}\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau\boldsymbol{H}_\tau}. \qquad (5.15)$$

Plugging (5.15) into (5.8), simplifying, and taking (5.2) into account, the normalized MSE can be expressed by

$$\widetilde{\sigma}^2_{cf}(k) \triangleq \frac{\sigma^2_{cf}(k)}{N_rN_t} = 2(1-\sqrt{\alpha_k}) - \frac{(1-2\sqrt{\alpha_k})}{\frac{\rho_\tau T_\tau}{N_t} + \frac{N_r}{N_t}\mathrm{trace}(\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau\boldsymbol{H}_\tau})}\,. \qquad (5.16)$$

Note that the additional noise term used to model the channel variation in (5.3) results in $0 \le \widetilde{\sigma}^2_{cf}(k) \le 2$.

5.4.1. Optimal Training Length.

The optimal training length is now investigated. The partial derivative of (5.16) with respect to T_τ can be written as

$$\begin{aligned}\frac{\partial\widetilde{\sigma}^2_{cf}(k)}{\partial T_\tau} &= -\frac{\partial}{\partial T_\tau}\left[\frac{(1-2\sqrt{\alpha}_k)}{\frac{\rho_\tau T_\tau}{N_t} + \frac{N_r}{N_t}\mathrm{trace}\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau\boldsymbol{H}_\tau}}\right]\\ &= \frac{(1-2\sqrt{\alpha_k})}{(\frac{T_\tau N_t}{\rho_\tau} + \frac{N_rN_t}{\rho_\tau^2}\mathrm{trace}\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau\boldsymbol{H}_\tau})^2}\,. \end{aligned} \qquad (5.17)$$

Observe that since the denominator in (5.17) is positive, the sign of the derivative is completely dependent on α_k. This yields three possible cases.

1. $\alpha_k = 1/4$: Clearly $\frac{\partial\widetilde{\sigma}^2_{cf}(k)}{\partial T_\tau} = 0$ which means $\widetilde{\sigma}^2_{cf}(k)$ is constant for every T_τ, which suggests choosing $T_\tau = 0$. The channel coefficients in this case do not possess enough correlation to warrant training.

2. $\alpha_k < 1/4$: Since the derivative is positive, $\widetilde{\sigma}^2_{cf}(k)$ is strictly increasing, inferring the choice of $T_\tau = 0$. This is justified by noting that when α_k is decreased, the channel coefficients becomes less correlated and hence deviate substantially from a training estimate.

3. $\alpha_k > 1/4$: Through similar reasoning as Case 2, $\widetilde{\sigma}^2_{cf}(k)$ is strictly decreasing which justifies making T_τ as large as possible. When the channel coefficients are highly correlated, they are less susceptible to variation, making it beneficial to spend sufficient time learning the channel

Looking at (5.17), it is clear that increasing N_r or N_t would result in decreasing the rate of change of the MSE, whereas increasing ρ_τ will have the opposite effect.

To illustrate these results with a simple example (5.16) and (5.17) have been plotted respectively in Figures 5.3 and 5.4 for various values of ρ_τ, α_k, and N_t under the assumption of iid Rayleigh fading. The graphs validate the optimal training values for the different values of α_k and also exemplify how the rate of change of the MSE varies with ρ_τ, N_t, and N_r. We now summarize these results into the following proposition.

Proposition 5.1. *Given a MIMO continuously varying channel with parameters* $(\alpha_k, \rho_\tau, N_r, N_t, T_\tau)$ *the optimal time to spend training is*

1. $T_\tau = 0$ *when* $\alpha_k \leq 1/4$.

2. $T_\tau = D$ *when* $\alpha_k > 1/4$.

Furthermore, larger N_t *and* N_r *reduce the rate of change of the MSE, whereas a larger* ρ_τ *accelerates the rate of change.*

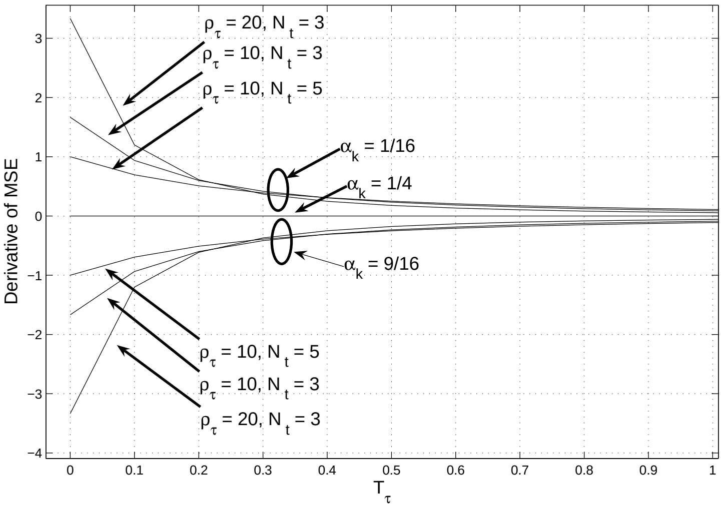

Figure 5.3 Variation in the partial derivative of the mean squared error with respect to T_τ for various α_k, N_t, and ρ_τ.

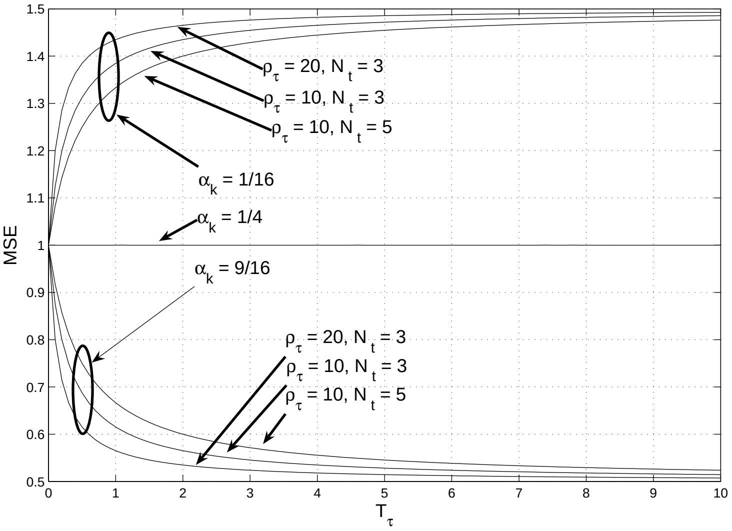

Figure 5.4 Variation in MSE with respect to T_τ for various α_k, N_t, and ρ_τ.

5.5. SPECIAL CASES

In this section special cases of our new result for the channel estimation error of a continuously varying MIMO channel are presented which yield previous forms in the literature. These are followed by the uncorrelated block fading and Rayleigh time varying cases.

5.5.1. Rayleigh Block Fading. Let $\alpha_k = 1$ for all k and $\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau} = N_r \boldsymbol{I}_{N_t}$. Then (5.16) reduces to

$$\widetilde{\sigma}^2_{cf}(k) = \frac{1}{1 + \frac{\rho_\tau T_\tau}{N_t}}\ , \quad k = T_\tau + 1, \ldots, T$$

which is precisely (5.4).

5.5.2. Arbitrary Block Fading. Suppose $\alpha_k = 1$ for all k. Then (5.16) becomes

$$\widetilde{\sigma}^2_{cf}(k) = \frac{1}{\frac{\rho_\tau T_\tau}{N_t} + \frac{N_r}{N_t}\mathrm{trace}(\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau})}\ , \quad k = T_\tau + 1, \ldots, T$$

which is exactly (5.5).

5.5.3. Uncorrelated Block Fading. Let $\alpha_k = 0$ for all k. Then (5.16) can be written as

$$\widetilde{\sigma}^2_{cf}(k) = 2 - \frac{1}{\frac{\rho_\tau T_\tau}{N_t} + \frac{N_r}{N_t}\mathrm{trace}(\boldsymbol{R}^{-1}_{H_\tau H_\tau})}\ , \quad k = T_\tau + 1, \ldots, T\ . \tag{5.18}$$

We now seek the $\boldsymbol{H}_\tau$ that minimizes (5.18). It can be shown [51] that

$$\mathrm{trace}(\boldsymbol{R}^{-1}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}) \geq \sum_i \frac{1}{[\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}]_{ii}}$$

with equality if and only if $\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}$ is diagonal. We now seek to find the optimal diagonal matrix that satisfies

$$\arg \min_{\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}} \operatorname{trace}(\boldsymbol{R}^{-1}_{H_\tau H_\tau})$$

such that

$$\mathbb{E}||\boldsymbol{H}(\tau)||_F = N_r N_t \ .$$

Differentiating with respect to $\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau}$, setting the result equal to zero and applying the constraint we obtain

$$\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau} = N_r \boldsymbol{I}_{N_t}$$

which shows that the MSE is minimized under time-uncorrelated block fading when the channel coefficients are uncorrelated. The MSE is

$$\widetilde{\sigma}^2_{cf}(k) = \frac{2}{1 + \frac{N_t}{\rho_\tau T_\tau}} + \frac{1}{\frac{\rho_\tau T_\tau}{N_t} + 1} \ , \quad k = T_\tau + 1, \ldots, T. \tag{5.19}$$

5.5.4. Rayleigh Time Varying. Letting $\boldsymbol{R}_{\boldsymbol{H}_\tau \boldsymbol{H}_\tau} = N_r \boldsymbol{I}_{N_t}$, the channel estimation error can be written as

$$\widetilde{\sigma_{cf}}^2(k) = \frac{2(1 - \sqrt{\alpha_k})}{1 + \frac{N_t}{\rho_\tau T_\tau}} + \frac{1}{\frac{\rho_\tau T_\tau}{N_t} + 1}.$$

The MSE is expressed as the sum of two terms, the first is the error due to the time varying nature of the channel and the second is the Rayleigh block fading MSE previously mentioned. Unlike the block fading term, an increase in ρ_τ or T_τ adversely affects the time varying term, necessitating a tradeoff between the two components. Note that when $\alpha_k = 0$ this simplifies to (5.19) which happens to be the worst case scenario for this case. This is not surprising since highly correlated channel coefficients will fluctuate less and thus remain close to their estimate. A concrete example is now presented to illustrate this behavior.

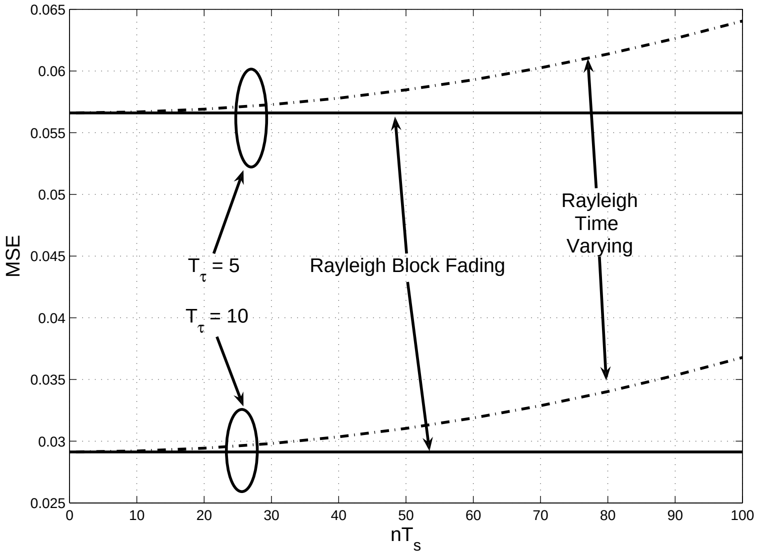

Figure 5.5 Mean squared error comparison for the Rayleigh block fading case and the Rayleigh time varying case as a function of symbol period for various T_τ when $N_r = N_t = 3$ and $f_d T_s = .0002$.

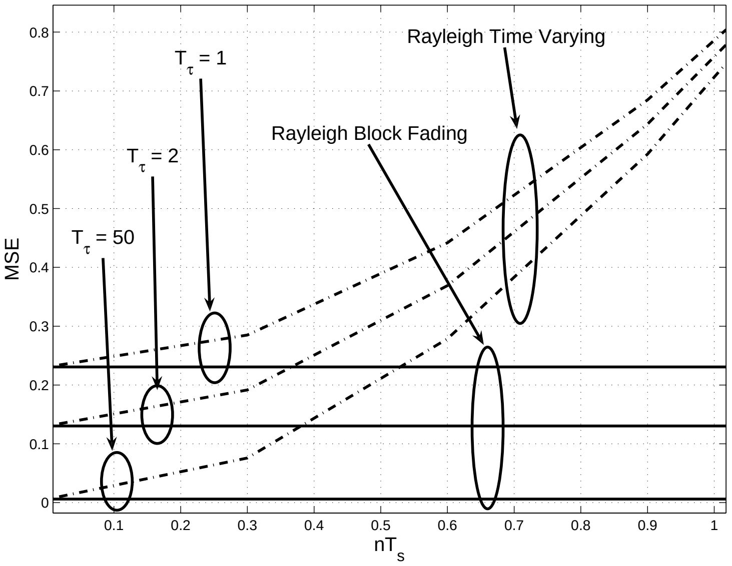

Figure 5.6 Mean squared error comparison for the Rayleigh block fading case and the Rayleigh time varying case as a function of symbol period for various T_τ when $N_r = N_t = 3$ and $f_d T_s = .2$.

Assume that the temporal variation in the channel obeys Jake's model resulting in [3]

$$\alpha_k = J_o(2\pi k f_d T_s)^2$$

where f_d is the maximum doppler frequency and $J_o(\cdot)$ is the zero order bessel function of the first kind. When f_d is not accurately tracked, frequency dispersion can corrupt the spectrum of the transmitted signal. The coherence time is the time domain dual for doppler frequency and is interpreted to be the maximum time a signal can pass undistorted (due to frequency dispersion) through the channel. A common expression for the coherence time is [5]

$$T_c = \frac{.423}{f_d}.$$

Therefore choosing a normalized frequency that satisfies

$$f_d T_s \ll .423 \tag{5.20}$$

will result in a static channel (with respect to doppler shift) throughout a symbol period. The MSE will now be investigated for different values of $f_d T_s$.

Let $f_d T_s = .0002$ and $N_r = N_r = 3$. Looking at Figure 5.5 one can see that the continuously fading MSE does not deviate substantially from the block fading case for 100 symbol periods. This reinforces the constraint on the normalized frequency in (5.20) which prevents α_k from impacting the time varying component of the MSE.

Suppose now that $f_d T_s = .2$. The results in Figure 5.6 indicate that the continuously fading MSE varies drastically from the block fading case. The normalized frequency no longer satisfies (5.20), thereby causing α_k to fluctuate and negatively impact the MSE during the payload.

5.6. CONCLUSION

Previous expressions for the channel estimation error did not take into account an arbitrary channel that varies temporally during the payload. The main topic of this work accomplished that task. Optimal values for the training time were found and it was shown how different values for the received SNR and the number of antennas affected the rate of change of the channel estimation error. Special cases of this new expression were shown to yield previous results in the literature. For Rayleigh flat fading it was shown that varying the normalized frequency impacted the accuracy of the channel estimate and hence the channel estimation error.

6. SINGLE BOUNCE AIR TO GROUND CHANNEL CAPACITY FOR MIMO SYSTEMS

6.1. INTRODUCTION

Since the inception of multiple-input multiple-output (MIMO) antennas for digital communications, dramatic increases in capacity have been have been reported [1, 48]. The main topic of this work is to investigate how the capacity of an air to ground aeronautical communication link is affected by various physical parameters.

This paper is organized as follows. In the first section, the continuous time channel model is defined. Next, the discretized channel model is presented. An analysis is then presented that describes mathematically how the capacity will be affected by varying numerous parameters. This is followed by our concluding remarks.

6.2. CHANNEL MODEL

An aircraft flying with velocity $v(t)$ at an altitude of h_{tx} toward a base station is displayed in Figure 6.1. The aircraft and base station are separated by a distance d_{hor}. Let the baseband channel impulse response for a single reflection be represented by

$$h_b(t) = e^{-j2\pi f_c \tau_1(t)}\delta(\tau - \tau_1(t)) + \Gamma e^{-j2\pi f_c \tau_2(t)}\delta(\tau - \tau_2(t)) \tag{6.1}$$

where f_c is the carrier frequency and $\tau_i(t)$ is the delay of the i^{th} incoming wave. Noting that the wavelength $\lambda = \frac{c}{f_c}$ where c is the speed of light we have

$$h_b(t) = e^{\frac{-j2\pi d(t)}{\lambda}}\delta\Big(\tau - \frac{d(t)}{c}\Big) + \Gamma e^{\frac{-j2\pi \widehat{d}(t)}{\lambda}}\delta\Big(\tau - \frac{\widehat{d}(t)}{c}\Big) \tag{6.2}$$

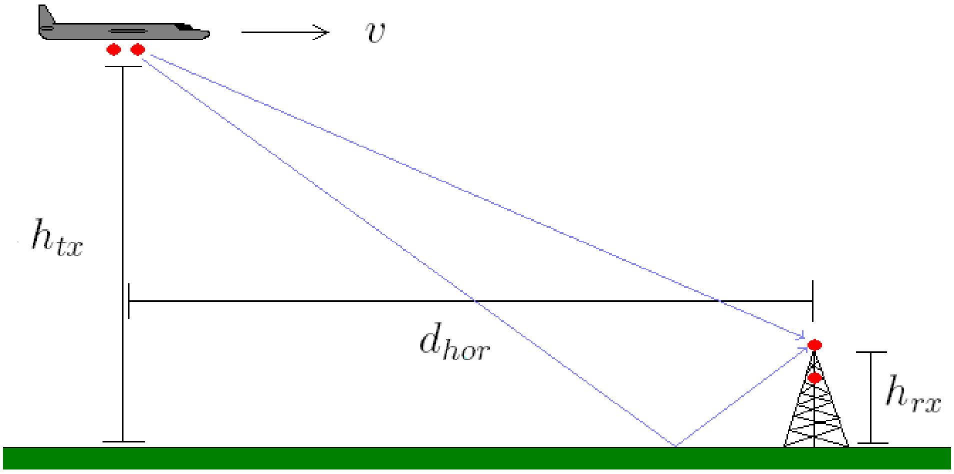

Figure 6.1 Scenario of an aircraft flying toward the base station.

where

$$\begin{aligned} d(t) &= \sqrt{(h_{tx} - h_{rx})^2 + (d_{hor}(t))^2} && (6.3) \\ &= \sqrt{(h_{tx} - h_{rx})^2 + (x_o - v(t)t)^2} && (6.4) \\ \widehat{d}(t) &= \sqrt{(h_{tx} + h_{rx})^2 + (x_o - v(t)t)^2} && (6.5) \end{aligned}$$

are respectively the LOS and reflected distances while x_o is the initial horizontal displacement.

The presence of velocity in (6.9) and (6.10) suggests the possibility of channel distortion due to doppler shift if the carrier frequency and aircraft velocity are not sufficiently tracked. The two paths are assumed to be unresolvable at the receiver if (Appendix A)

$$|\tau_1(t) - \tau_2(t)| \ll B^{-1} \tag{6.6}$$

or equivalently

$$\left| \frac{\sqrt{(h_{tx} - h_{rx})^2 + (v(t)t)^2} - \sqrt{(h_{tx} + h_{rx})^2 + (v(t)t)^2}}{c} \right| \ll B^{-1} \tag{6.7}$$

To find what signal bandwidths satisfy this constraint, we present a concrete example. A Lockheed Martin/Boeing F-22 Raptor is traveling at its supercruise velocity of Mach 1.82 (1,220 mph, 542.2 m/s) at an altitude of 50,000 ft (15,240 m). The aircraft is initially 10,000 m from the base station whose antenna height is 100 ft (30.48 m). The two path delays are plotted in Figure 6.2 along with their difference in absolute value in Figure 6.3. The dip in Figure 6.2 occurs when the aircraft is directly over the base station. Observing Figure 6.3, the signal bandwidth can be in the MHz range and still satisfy (6.7). It should be noted that the narrow-band assumption for a two ray model has also been employed in [52, 53, 54, 55].

For simulation purposes, the wireless channel must be digitized. To simplify the notation, the subscript in (6.1) will be dropped and the base-band assumption will be used throughout the remainder of the paper. Using (A.10) for $N_p = 2, b_1(k) = 1, b_2(k) = \Gamma$, the channel at discrete time kT_s is

$$\begin{aligned} h(k) &= e^{-j2\pi f_c \tau_1(k)} + \Gamma e^{-j2\pi f_c \tau_2(k)} \\ h(k) &= e^{\frac{-j2\pi d(k)}{\lambda}} + \Gamma e^{\frac{-j2\pi \widehat{d}(k)}{\lambda}} \end{aligned} \tag{6.8}$$

where

$$d(k) = \sqrt{(h_{tx} - h_{rx})^2 + (x_o - v(k)kT_s)^2} \tag{6.9}$$

$$\widehat{d}(k) = \sqrt{(h_{tx} + h_{rx})^2 + (x_o - v(k)kT_s)^2} \tag{6.10}$$

and T_s is the sampling period. For MIMO systems, we extend (6.8) to account for multiple antenna elements at the transmitter and receiver and express the channel coefficient between the m^{th} receiver and n^{th} transmitter as

$$h_{mn}(k) = e^{\frac{-j2\pi d_{mn}(k)}{\lambda}} + \Gamma e^{\frac{-j2\pi \widehat{d}_{mn}(k)}{\lambda}}. \tag{6.11}$$

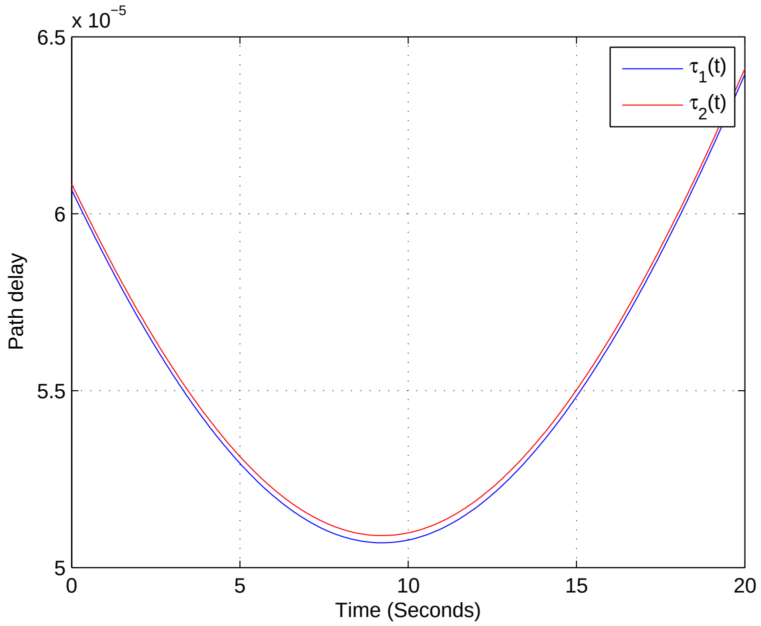

Figure 6.2 Time delay of two the paths.

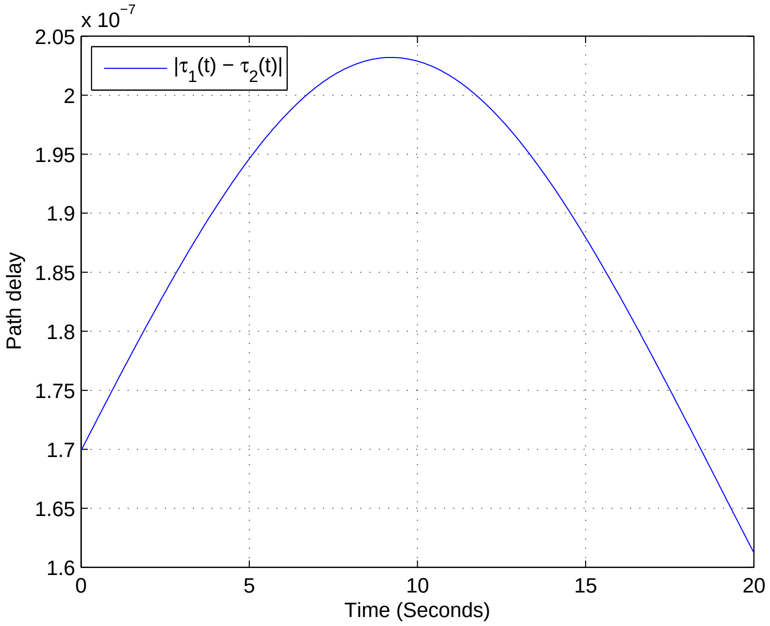

Figure 6.3 Difference in time delay of the two paths.

Thus the channel matrix can be expressed at the k^{th} sample as

$$\boldsymbol{H}(k) = \begin{bmatrix} h_{11}(k) & \dots & h_{1N_t}(k) \\ \vdots & \ddots & \vdots \\ h_{N_r1}(k) & \dots & h_{N_rN_t}(k) \end{bmatrix} \tag{6.12}$$

where N_t and N_r are respectively the number of transmit and receive antennas. For the remainder of this paper the dependence on k will be suppressed.

6.3. CAPACITY

The performance measure chosen in this work to compare the different scenarios is the ergodic capacity. For our discussion, we will look at the following scenarios which are illustrated in Figure 6.4:

1. one transmitter and one receiver antenna
2. two horizontally spaced transmitter antennas and two vertically spaced receiver antennas
3. four rectangular spaced transmitter antennas and four rectangular receiver antennas

For the first case, the capacity is

$$C_{siso} = \log_2(1 + \rho|h|^2) \tag{6.13}$$

where ρ is the SNR. For the remaining geometries, the MIMO capacity is

$$C_{mimo} = \log_2 \det\left(\boldsymbol{I} + \frac{\rho}{N_t}\boldsymbol{H}\boldsymbol{H}^H\right). \tag{6.14}$$

We are now ready to state the following theorem.

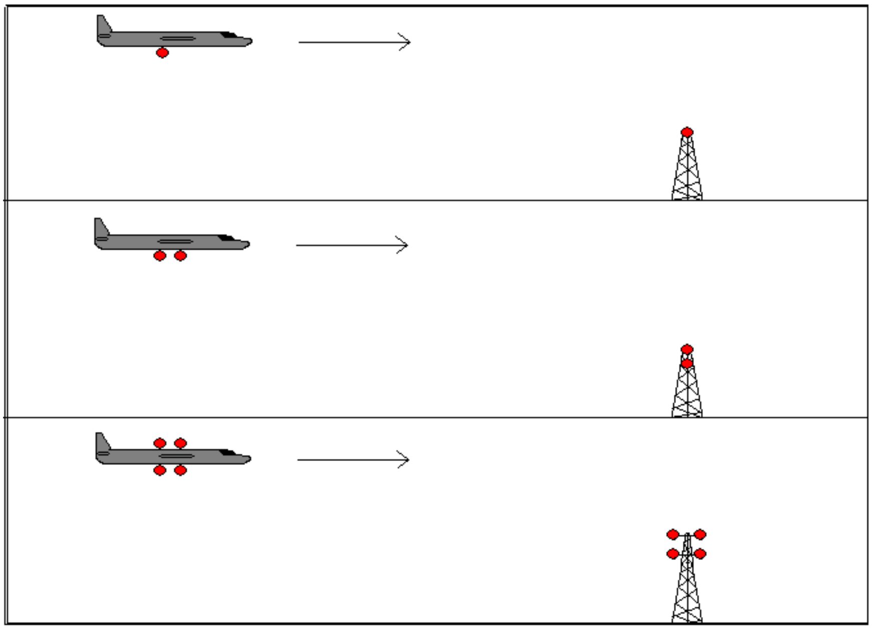

Figure 6.4 Three different antenna scenarios.(Top) One antenna at the transmitter and receiver. (Middle) Two horizontally spaced antennas at the transmitter and two vertically spaced antennas at the receiver. (Bottom) Four rectangularly spaced antennas at the transmitter and four rectangularly spaced antennas at the receiver.

Theorem 6.1. *If $-1 < \Gamma < 1$ then the capacity is always greater than zero.*

Proof. Expanding (6.11) into real and imaginary parts, the channel coefficient between the m^{th} receiver and n^{th} transmitter is

$$h_{mn} = \left(\cos\left(\frac{2\pi d_{mn}}{\lambda}\right) + \Gamma\cos\left(\frac{2\pi \widehat{d}_{mn}}{\lambda}\right)\right) - j\left(\sin\left(\frac{2\pi d_{mn}}{\lambda}\right) + \Gamma\sin\left(\frac{2\pi \widehat{d}_{mn}}{\lambda}\right)\right). \quad (6.15)$$

Clearly, for C_{mimo} to equal zero

$$\cos\left(\frac{2\pi d_{mn}}{\lambda}\right) = -\Gamma\cos\left(\frac{2\pi \widehat{d}_{mn}}{\lambda}\right) \quad (6.16)$$

$$\sin\left(\frac{2\pi d_{mn}}{\lambda}\right) = -\Gamma\sin\left(\frac{2\pi \widehat{d}_{mn}}{\lambda}\right) \quad (6.17)$$

for all m and n. Observing (6.16) and (6.17), the following condition must be satisfied

$$\frac{\cos\left(\frac{2\pi d_{mn}}{\lambda}\right)}{\cos\left(\frac{2\pi \widehat{d}_{mn}}{\lambda}\right)} = \frac{\sin\left(\frac{2\pi d_{mn}}{\lambda}\right)}{\sin\left(\frac{2\pi \widehat{d}_{mn}}{\lambda}\right)} \tag{6.18}$$

$$\tag{6.19}$$

or equivalently

$$\sin\left(\frac{2\pi}{\lambda}\left(\widehat{d}_{mn} - d_{mn}\right)\right) = 0 \tag{6.20}$$

where the appropriate trigonometric identity has been used. There are only two solutions modulo 2π for (6.20),which are $\frac{2\pi}{\lambda}(\widehat{d}_{mn} - d_{mn}) = 0, \pi$. For the first case we easily find that $\Gamma = -1$. For the second case

$$\frac{\cos\left(\frac{2\pi}{\lambda} d_{mn} + \pi\right)}{\cos\left(\frac{2\pi}{\lambda} d_{mn}\right)} = -\Gamma \tag{6.21}$$

$$\tag{6.22}$$

which implies $\Gamma = 1$. □

An alternate expression for the capacity of a MIMO channel can be written as

$$C_{mimo} = \log_2 \prod_{i=1}^{M} (1 + \frac{\rho}{N_t}\sigma_i^2) \tag{6.23}$$

$$= \sum_{i=1}^{M} \log_2 (1 + \frac{\rho}{N_t}{\sigma_i}^2) \tag{6.24}$$

where $M = \text{rank}(\boldsymbol{H})$ and σ_i^2 is the i^{th} singular value of $\boldsymbol{H}$ satisfying $0 \leq \sigma_{i-1} \leq \sigma_i \leq \sigma_{i+1} \leq \sigma_M$. Thus, increasing the number of high strength eigen-channels will have a positive effect on the ergodic capacity.

6.4. SPATIAL VARIATION

For MIMO systems, the separation between antennas can play a significant role in performance. Suppose we have a 4x4 MIMO system with rectangular spacing at both the transmitter and receiver, as illustrated in Figure 6.4. To separate the effect on performance due to spatial variation and frequency dispersion, it is assumed in this subsection that the doppler frequency is perfectly estimated and removed at the receiver. Let the antennas be uniformly spaced and separated by d_{sep}. The LOS and reflected distances for two of the sub-channels are

$$\begin{aligned} d_{11} &= \sqrt{(x_{rx} - x_{tx})^2 + (h_{rx} - h_{tx} - d_{sep})^2} && (6.25) \\ \widehat{d}_{11} &= \sqrt{(x_{rx} - x_{tx})^2 + (h_{rx} + h_{tx} + d_{sep})^2} && (6.26) \\ d_{12} &= \sqrt{(x_{rx} - x_{tx} - d_{sep})^2 + (h_{rx} - h_{tx} - d_{sep})^2} && (6.27) \\ \widehat{d}_{12} &= \sqrt{(x_{rx} - x_{tx} - d_{sep})^2 + (h_{rx} + h_{tx} + d_{sep})^2} . && (6.28) \end{aligned}$$

Note that increasing the separation between antennas causes a larger discrepancy between the LOS and reflected distances. This discrepancy will prevent the phases in (6.11) from matching and consequently cancel the transmitted signal. Thus it is expected that increasing the antenna separation will have a positive effect on the ergodic capacity. This performance gain is compromised at high altitudes and/or large horizontal displacements since h_{tx} and d_{hor} will dominate (6.25)-(6.28) and thereby weaken the ability to differentiate between the reflected and LOS distances.

6.5. DOPPLER EFFECT

For certain applications accurate tracking of the doppler frequency may not be possible. The approach taken here is to lump the channel estimation error into an effective SNR. The capacity is then lower bounded as [56]

$$C_{mimo} \geq \log_2 \det \left(\boldsymbol{I} + \frac{\rho_{eff}}{N_t} \boldsymbol{H}\boldsymbol{H}^H \right) \tag{6.29}$$

where the effective SNR is

$$\rho_{eff} = \frac{\rho(1-\beta) + \rho\beta\alpha_k}{1 + \frac{\alpha_k N_r}{T_\tau} + (1-\alpha_k)\rho\beta}. \tag{6.30}$$

The parameter $0 < \beta < 1$ determines the line of site strength (0 is pure LOS), $\alpha_k = J_o(2\pi f_d k T_s)$ is the zeroth order bessel function of the first kind, $f_d = \frac{v}{\lambda}$ is the maximum doppler frequency, and T_τ is the number of training symbols. Observing (6.30), as the channel becomes uncorrelated ($\alpha_k \rightarrow 0$) the effective SNR, and consequently the capacity, decrease.

6.6. NUMERICAL RESULTS

To verify the results in the preceding two subsections, we present some numerical examples. For all simulations the carrier frequency is 1 GHz, the base station is at a height of 30.48 m (100 ft), the initial horizontal difference is 10000 m (30000 ft), the SNR was 10 dB, and the number of samples taken are 20000. We decided to use the three different antenna geometries illustrated in Figure 6.4.

The first set of experiments consisted of varying the altitude of the aircraft as well as the antenna separation (in terms of the wavelength λ) for $\Gamma = 1$. This value for the

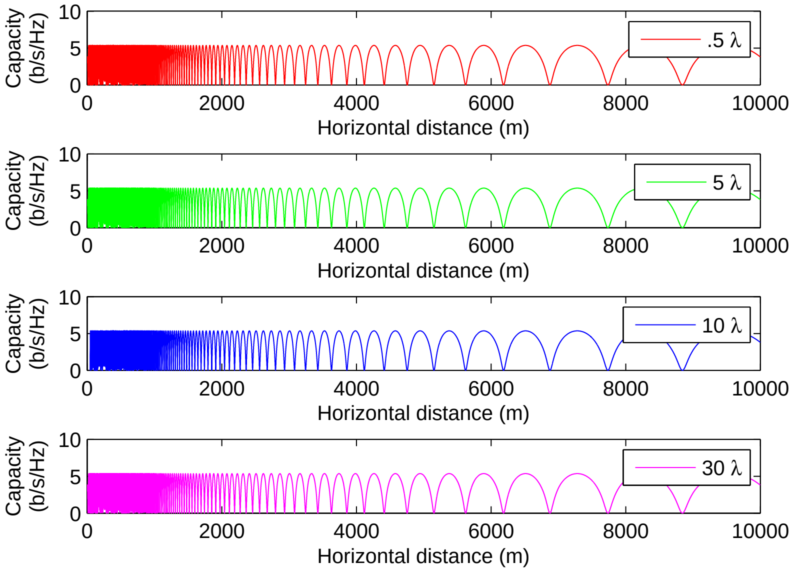

Figure 6.5 Effect on capacity due to different antenna separations for a 1x1 scenario with aircraft flying at an altitude of 304.8 m (1000 ft).

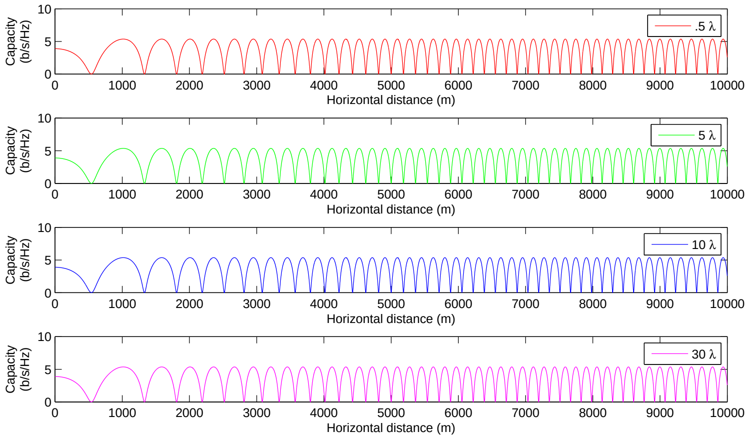

Figure 6.6 Effect on capacity due to different antenna separations for a 1x1 scenario with aircraft flying at an altitude 12,192 m (40000 ft).

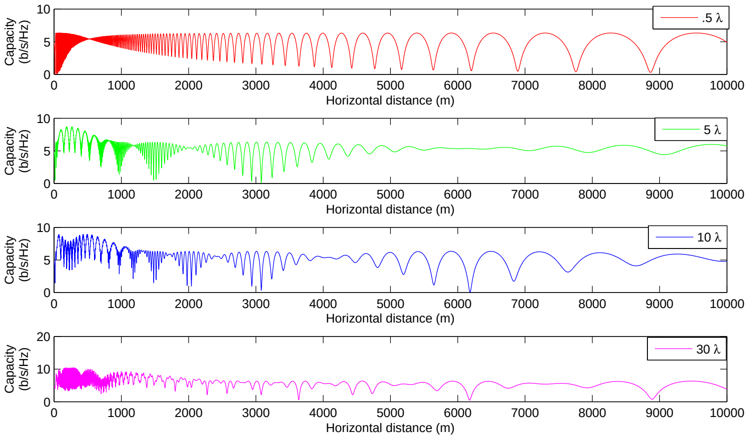

Figure 6.7 Effect on capacity due to different antenna separations for a 2x2 scenario with aircraft flying at an altitude of 304.8 m (1000 ft)

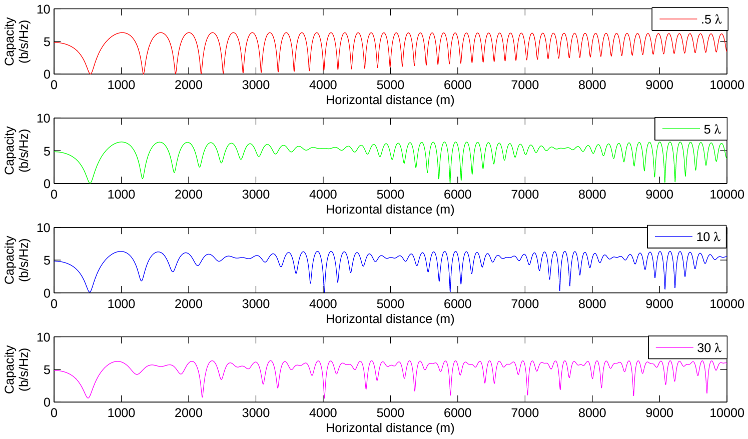

Figure 6.8 Effect on capacity due to different antenna separations for a 2x2 scenario with aircraft flying at an altitude of 12,192 m (40000 ft).

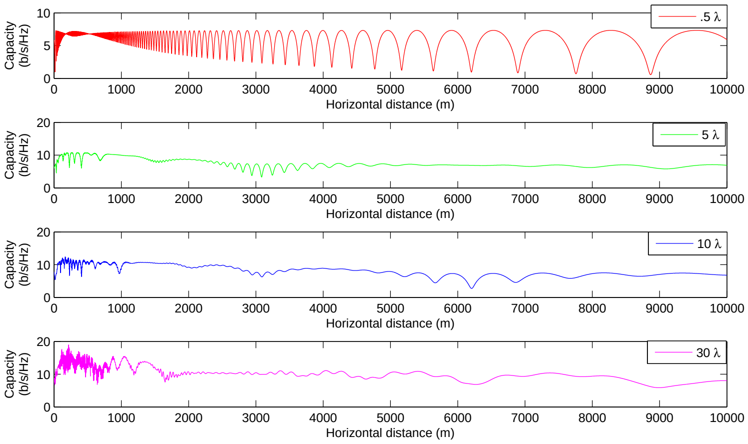

Figure 6.9 Effect on capacity due to different antenna separations for a 4x4 scenario with aircraft flying at an altitude of 304.8 m (1000 ft)

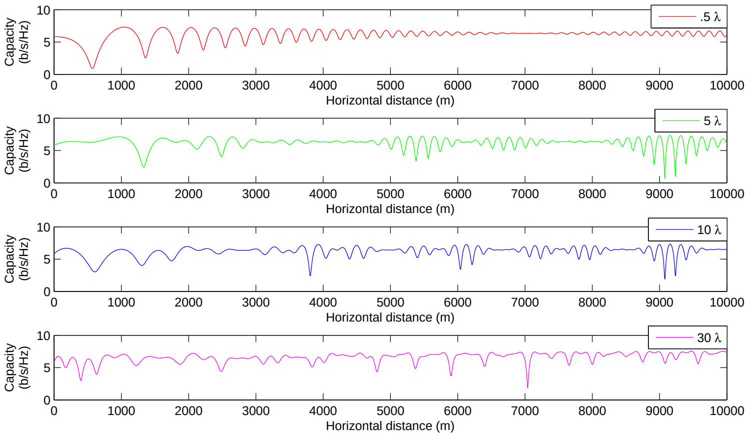

Figure 6.10 Effect on capacity due to different antenna separations for a 4x4 scenario with aircraft flying at an altitude of 12,192 m (40000 ft).

reflection coefficient is chosen to provide a worst case scenario for null occurence. These results are reported in Figures 6.5-6.9. It is clear that increasing the antenna separation is beneficial to the performance which supports our previous observation. Note also that increasing the number of antennas decreases the number of nulls. This is not surprising, since more antennas create additional paths for the transmitted signals to reach the receiver. To further illustrate this the singular values of the 4x4 case are plotted in Figures 6.11 and 6.12. As the antenna spacing is increased, more significant non-zero singular values become apparent. When the aircraft is flying at higher altitudes and/or is far away from the base station, a change in antenna spacing has virtually no impact on obtaining more significant non-zero singular values. This verifies our discussion in the previous subsections.

In the next simulation, the reflection coefficient as well as the altitude are varied while keeping the antenna spacing at 5λ to see how the performance is affected. We observe in Figures 6.13-6.16 that for $0 < \Gamma < 1$ the capacity never goes to zero. This justifies Lemma 6.1 in the previous section. We see in Figure 6.8 that a decrease in absolute value in Γ corresponds to a smaller capacity variation around the $\Gamma = 0$ scenario. When observing Figures 6.15 and 6.16, there are distances where the variation in capacity never goes below the $\Gamma = 0$ line. Moreover, the capacity varies the same for both attenuation factors and is actually better for higher absolute values of Γ. This can be explained by recalling that multiple antennas at the transmitter and receiver provide more paths for the transmitted signals to reach the receiver. Since the spacing is 5λ, the performance benefits due to antenna spacing illustrated in the previous section apply here as well. This added robustness to null occurrence will allow better performance for higher values of Γ. Unlike SISO systems, MIMO systems having “destructive” interference can be desirable for performance.

In the final simulation both velocity and altitude were varied and (6.29) and (6.30) were used to evaluate the performance. The channel was estimated using pilot symbols for $T_\tau = 10$. Observing Figure 6.17 and 6.18, at low altitude the performance is greatly affected by a change in velocity opposed to a higher altitude. This is due to the presence of

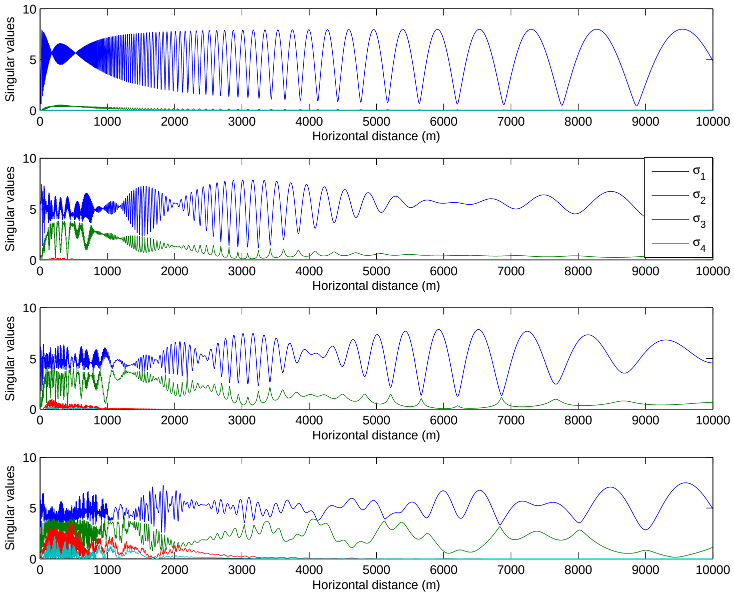

Figure 6.11 Singular values for a 4x4 scenario with an aircraft flying at altitude of 304.8 m (1000 ft).

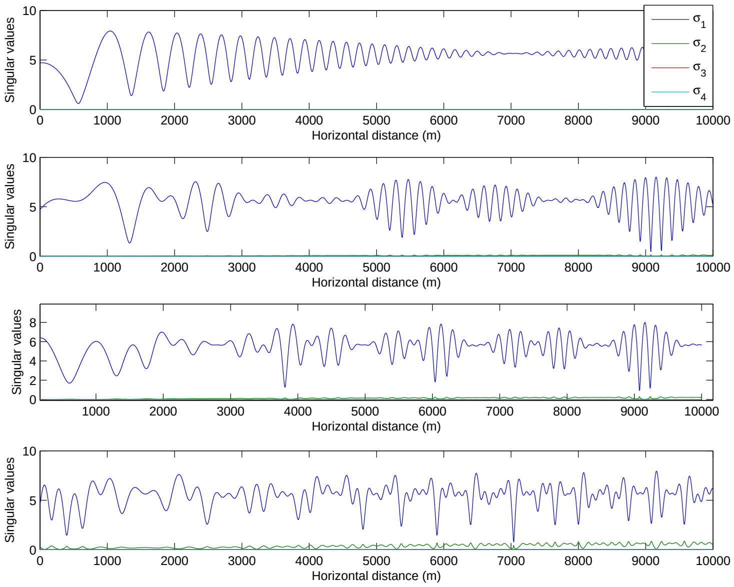

Figure 6.12 Singular values for a 4x4 scenario with aircraft flying at 12,192 m (40000 ft). For both scenarios, the antenna separation is (from top to bottom) .5λ, 5λ, 10λ, and 30λ.

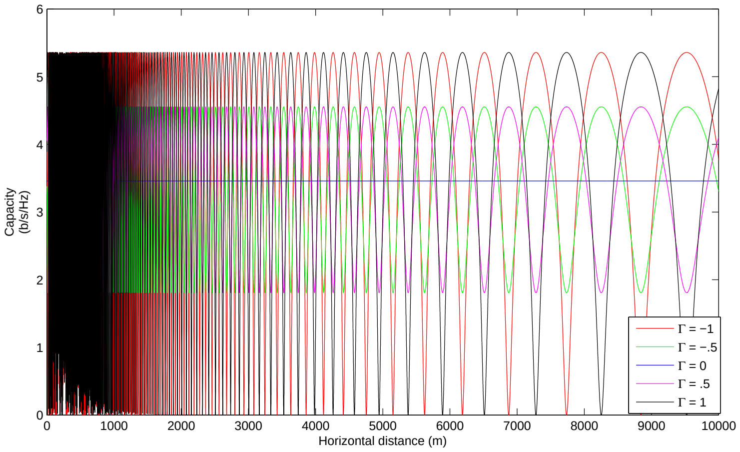

Figure 6.13 Effect on capacity for different reflection coefficients for a 1x1 scenario with aircraft flying at an altitude of 304.8 m (1000 ft)

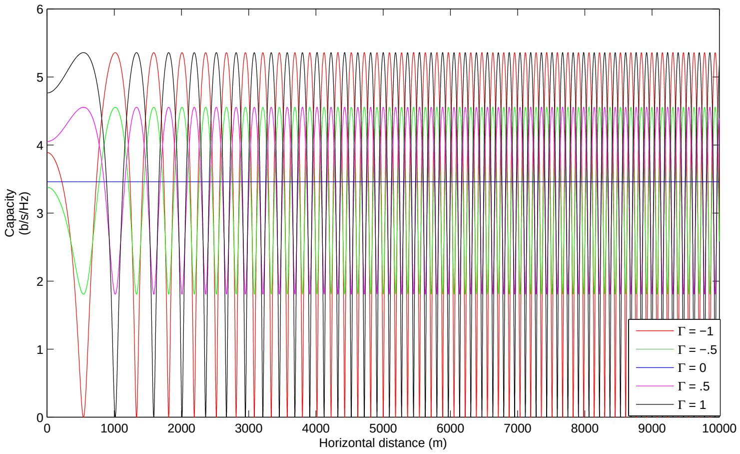

Figure 6.14 Effect on capacity for different reflection coefficients for a 1x1 scenario with aircraft flying at an altitude of 12,192 m (40000 ft).

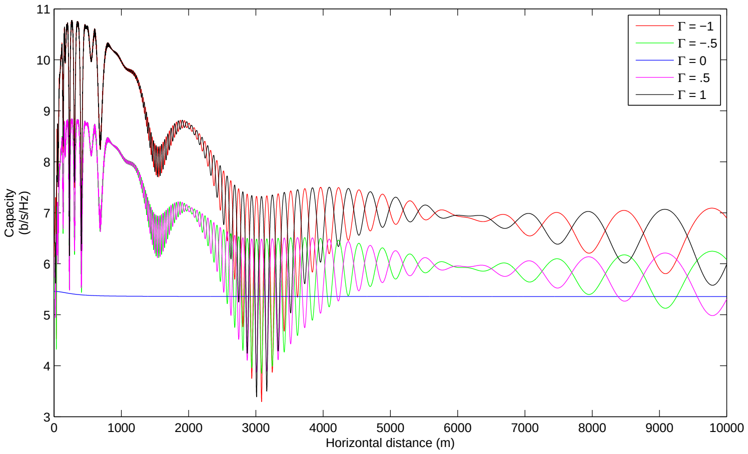

Figure 6.15 Effect on capacity for different reflection coefficients for a 4x4 scenario with aircraft flying at an altitude of 304.8 m (1000 ft)

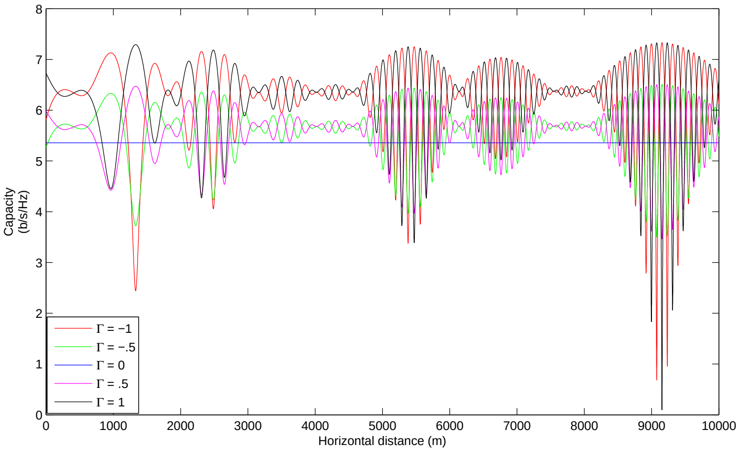

Figure 6.16 Effect on capacity for different reflection coefficients for a 4x4 scenario with aircraft flying at an altitude of 12,192 m (40000 ft).

more resolvable paths and hence more non-zero singular values (Figure 6.12) which follows from the previous discussion on altitude and horizontal displacement. This corresponds to a larger value of β. This places more emphasis on the second term in the numerator in (6.30) as well as the third term in the denominator. Clearly α_t will have a significant impact on the capacity: the faster the velocity, the faster the bessel function will go to zero which will result in ρ_{eff} going to zero. When the aircraft is flying at a higher altitude, we have shown that the channel paths become unresolvable which yields a smaller value for β. Since $\rho\beta$ becomes negligible and $N_r/T_\tau < 1$, it is clear that ρ_{eff} is less sensitive to a doppler shift. Hence, the capacity is severely attenuated due to a doppler shift at low altitudes, but only slightly affected at high altitudes.

6.7. CONCLUSION

A channel model that takes into consideration the altitude, antenna separation, and doppler shift was obtained. An aircraft flying at low altitudes and small horizontal displacement was found to have more desirable performance when the doppler shift is accurately tracked. When the number of antennas and their separation were increased, the capacity had a significant decrease in null occurrence and was less sensitive to a change in the reflection coefficient. In fact it was shown that for MIMO systems a higher value for the reflection coefficient was beneficial to the performance.

An expression for the capacity in terms of a effective SNR was used to account for the error in channel estimation due to a doppler shift. From this we were able to conclude that the velocity of an aircraft affected the performance more significantly at lower altitudes.

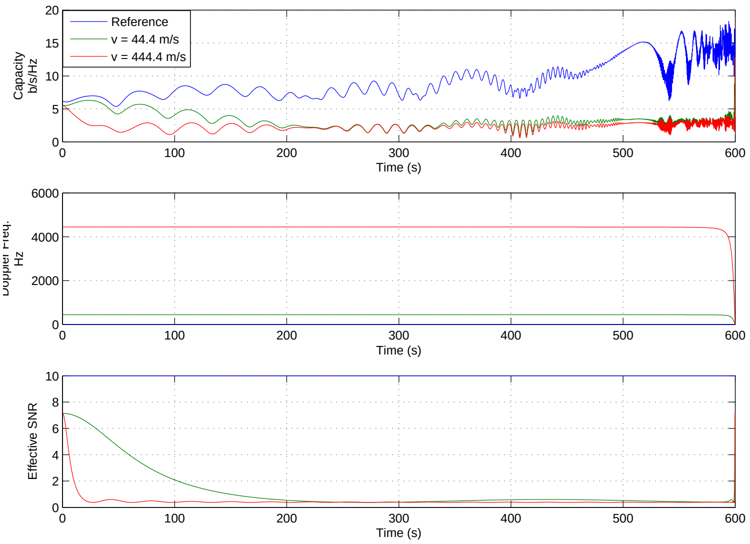

Figure 6.17 Effect on capacity, doppler shift, and effective SNR for an aircraft flying at various velocities at an altitude of 304.8 m (1000 ft)

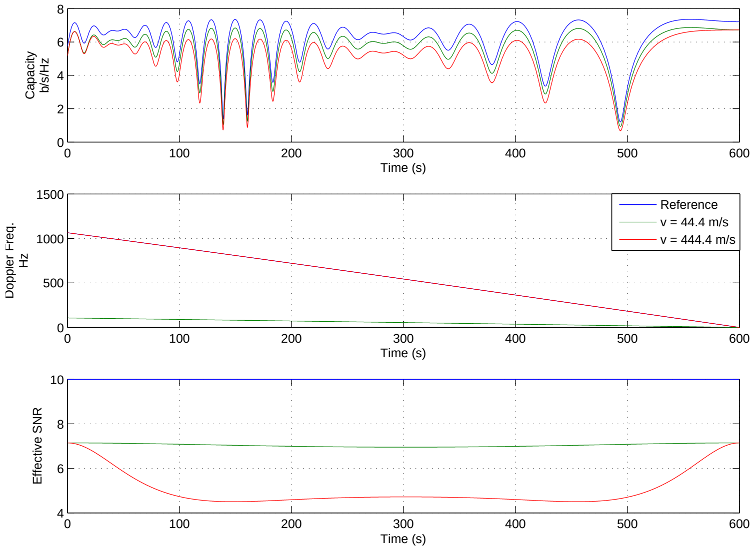

Figure 6.18 Effect on capacity, doppler shift, and effective SNR for an aircraft flying at various velocities at an altitude of 12,192 m (40000 ft).

APPENDIX
A LOW PASS DISCRETE TIME MIMO MODEL

In this section a low pass discrete time channel model is derived for both SISO and MIMO systems.

SISO Systems. Recall that the output for a SISO noise free wireless baseband system can be represented by

$$y_b(t) = \sum_{i=1}^{N_p} a_i(t) e^{-j2\pi f_c \tau_i(t)} x_b(t - \tau_i(t)) \tag{A.1}$$

where N_p is the number of delay paths, $a_i(t)$ is a real number describing the channel gain, f_c is the carrier frequency, $\tau_i(t)$ is the i^{th} path delay, and $x_b(t)$ is the baseband transmitted signal. Assuming that $x_b(t)$ is band-limited to B Hz, application of the sinc interpolation theorem yields [57]

$$x_b(t) = \frac{B}{f_s} \sum_{n=-\infty}^{\infty} x_b(nT_s) \mathrm{sinc}(B(t - nT_s)) \tag{A.2}$$

where T_s is the sampling rate of the signal. For simplicity let $T_s = \frac{1}{B}$. Substituting (A.2) into (A.1) we obtain

$$y_b(t) = \sum_{n=-\infty}^{\infty} x_b(nT_s) \sum_{i=1}^{N_p} a_i(t) e^{-j2\pi f_c \tau_i(t)} \mathrm{sinc}(B(t - \tau_i(t)) - n)) \tag{A.3}$$

If we now sample $y_b(t)$ at integer of multiples of T_s the discrete time baseband output is

$$y_b(kT_s) = \sum_{n=-\infty}^{\infty} x_b(nT_s) \sum_{i=1}^{N_p} a_i(kT_s) e^{-j2\pi f_c \tau_i(kT_s)} \mathrm{sinc}(k - n - B\tau_i(kT_s)))) \tag{A.4}$$

Letting $m = k - n$ and defining $y(k) \triangleq y_b(kT_s), x(n) \triangleq x_b(bT_s), b_i(k) \triangleq a_i(kT_s)$, and $\tau_i(k) \triangleq \tau_i(kT_s)$[7] we may write

$$y(k) = \sum_{m=-\infty}^{\infty} x(k-m)h(k,m) \tag{A.5}$$

where

$$h(k,m) \triangleq \sum_{i=1}^{N_p} b_i(k)e^{-j2\pi f_c \tau_i(k)}\text{sinc}(m - B\tau_i(k)) \tag{A.6}$$

Looking at (A.6) the channel paths are resolved into temporal components of length $\frac{m}{B}$. Since $\tau_i(k) > 0$, we must have $m \geq 0$. If σ_τ is the root mean squared (RMS) delay spread [53], the number of resolvable multi-path components M must satisfy

$$\frac{M}{B} \approx \sigma_\tau \tag{A.7}$$

which yields

$$M \approx \lfloor B\sigma_\tau \rfloor \tag{A.8}$$

Therefore (A.6) can be written as

$$y(k) = \sum_{m=0}^{M-1} x(k-m)h(k,m) \tag{A.9}$$

Flat Fading. Suppose that all path delays are in the interval $\left[\frac{m}{B} - \frac{1}{2B}, \frac{m}{B} + \frac{1}{2B}\right]$. Then it follows directly that

$$h(k) \approx \sum_{i=1}^{N_p} b_i(k)e^{-j2\pi f_c \tau_i(k)}. \tag{A.10}$$

The channel has been reduced to a complex gain and only affects the current symbol.

---

[7] This is a convenient slight abuse of notation.

MIMO Systems. Utilizing (A.9), the input-output relationship for a discrete time noise free MIMO system with N_t transmit and N_r receive antennas is described by

$$\boldsymbol{y}(k) = \sum_{m=0}^{M-1} \boldsymbol{H}(k,m)\boldsymbol{x}(k-m) \tag{A.11}$$

where the sub-channel corresponding to transmit antenna i and receiver antenna j is

$$h_{ij}(k,m) \triangleq \sum_{i=1}^{N_p} b_i(k) e^{-j2\pi f_c \tau_i(k)} \mathrm{sinc}(m - B\tau_i(k)) \tag{A.12}$$

BIBLIOGRAPHY

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," *AT&T Bell Laboratories Internal Tech. Memo.*, June 1995.

[2] A.Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill Science, 1990.

[3] W. Jakes, *Microwave Mobile Communications.* New York: IEEE Press, 1993.

[4] X. Cai, N. Zhang, G. K. Venayagamoorthy, and D. C. Wunsch, "Time series prediction with recurrent neural networks trained by a hybrid pso-ea algorithm," *Neurocomputing*, vol. 70, pp. 2342–2353, Aug. 2007.

[5] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Upper Saddle River,NJ: Prentice Hall, 2002.

[6] S. Rice, "Statistical properties of a sine wave plus random noise," *Bell Syst. Tech. J.*, vol. 27, pp. 109–157, 1948.

[7] A. Sampath and J. Holtzman, "Estimation of maximum doppler frequency for handoff decisions," *Proc. IEEE VTC*, pp. 859–862, May 1993.

[8] Y. Bitran, "Broadband data, video, voice, and mobile convergence-extending the triple play," *Texas Inst. White Paper*, 2004.

[9] Y. Zheng and C. Xiao, "Improved models for the generation of multiple uncorrelated rayleigh fading waveforms," *IEEE Commun. Lett.*, vol. 6, no. 6, 2002.

[10] D. Shiu, G. Goschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, 2000.

[11] A. van Zelst and J. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple-output (mimo) radio channels," *Proc. of Gen. Assembly of the Int. Union of Radio Science*, Aug. 2002.

[12] T. Yoo and A. Goldsmith, "Capacity of fading mimo channels with channel estimation error," *IEEE Comm. Soc.*, pp. 808–813, 2004.

[13] ——, "Capacity and power allocation for fading mimo channels with channel estimation error," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[14] T. Yoo, E. Yoon, and A. Goldsmith, "Mimo capacity with channel uncertainty: Does feedback help?" *Proc. of IEEE GlobeCom*, pp. 96–100, 2004.

[15] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE Int. Conf. Neural Networks*, vol. 4, pp. 1942–1948, Dec. 1995.

[16] Y. Valle *et al.*, "Particle swarm optimization: Basic concepts, variants and applications in power systems," *IEEE Trans. Evol. Comput.*, vol. 12, no. 2, pp. 171–195, 2008.

[17] R. Xu, G. K. Venayagamoorthy, and D. C. Wunsch, "Modeling of gene regulatory networks with hybrid differential evolution and particle swarm optimization," *Elsevier Journal on Neural Networks*, vol. 21, pp. 917–927, Oct. 2007.

[18] J. Choi, M. Bouchard, and T. Yeap, “Decision feedback recurrent neural equalization with fast convergence rate,” *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 699–708, May 2005.

[19] W. Liu, L.-L. Yang, and L. Hanzo, “Recurrent neural network based narrowband channel prediction,” *IEEE Veh. Tech. Conf.*, vol. 5, pp. 2173–2177, May 2006.

[20] C. Potter, “Mimo beam-forming with neural network channel prediction trained by a novel pso-ea-depso algorithm,” *IEEE World Congress on Comp. Intell.*, 2008.

[21] L. Jacson, *Digital Filters and Signal Processing.* Kluwer, 1989.

[22] J. M. Ortega, *Matrix Theory-A Second Course.* New York, NY: Plenum, 1987.

[23] J. Brewer, “Kronecker products and matrix calculus in system theory,” *IEEE Transactions on Circuits and Systems*, vol. 25, pp. 772–781, Oct. 2004.

[24] J. Proakis, *Digital Communications.* New York,NY: McGraw-Hill, 1995.

[25] A. Goldsmith, *Wireless Communications.* New York: Cambridge University Press, 2005.

[26] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications.* Cambridge University.

[27] R. Gray, *Entropy and Information Theory.* New York: Springer-Verlag, 1990.

[28] J. Wolfitz, *Coding Theorems of Information Theory.* Berlin: Springer-Verlag, 1978.

[29] L. Ozarow, S. Shamai, and A. Wyner, “Information-theoretic considertations in cellular mobile radio,” *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[30] M. Medard, “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[31] J. Diaz, Z. Latinovic, and Y. Bar-Ness, “Impact of imperfect channel state information upon the outage capacity of rayleigh fading channels,” *IEEE Globecom*, pp. 887–892, 2004.

[32] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communications aspects,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[33] Z. Wang and G. Giannakis, “Outage mutual information of space-time mimo channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 4, pp. 657–662, Apr. 2004.

[34] S. Gradshteyn, I. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products, 5th ed.* San Diego, CA: Academic.

[35] V. Adamchik and O. Marichev, “The algorithm for calculating integrals of hypergeometric type functions and its realization in reduce systems.” *Proc. Int. Conf. Symbolic Algebraic Computation*, pp. 212–224, 1990.

[36] O. Oyman *et al.*, “Characterizing the statistical properites of mutual information in mimo channels,” *submitted to IEEE Trans. on Inform. Theory*, vol. 51, Dec. 2005.

[37] N. Goodman, “The distribution of a complex wishart distributed matrix,” *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 178–180, Mar. 1963.

[38] J. Cavers, “An analysis of pilot symbol assisted modulation for rayleigh fading channels,” *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[39] S. Zhou and G. B. Giannakis, “How accurate channel prediction needs to be for transmit beamforming with adaptive modulation over rayleigh mimo channels?” *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1285–1294, 2004.

[40] S. Sun, I. Wiemer, C. Ho, and T. Tjhung, “Training sequence assisted channel estimation for mimo ofdm,” *IEEE WCNC*, vol. 1, pp. 38–43, Mar. 2003.

[41] G. Foschini and M. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[42] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[43] S. Alamouti, “A simple transmit diverstiy technique for wireless communications,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Aug. 1998.

[44] L. Collin *et al.*, “Optimal minimum distance based precoder for mimo spatial multiplexing systems,” *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 617–627, Mar. 2004.

[45] Z. Zhou *et al.*, “Mimo systems with adaptive modulation,” *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1828–1842, Sept. 2005.

[46] B. Visweswaran and T. Kiran, “Channel prediction based power control in w-cdma systems,” *First Int. Conf. on 3G Mobile Comm. Techn.*, pp. 41–45, Mar. 2000.

[47] P. Coelho, “A new state space model for a complex rtrl neural network,” *Proc IJCNN*, vol. 3, pp. 439–441, 2001.

[48] G. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Syst. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[49] M. Biguesh and A. Gershman, “Training-based mimo channel estimation: A study of estimator tradeoffs and optimal training signals,” *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1053–1058, Mar. 2006.

[50] C. Peel and A. Swindlehurst, “Effective snr for space-time modulation over a time-varying rician channel,” *IEEE Trans. Wireless Commun.*, vol. 52, no. 1, pp. 17–23, Jan. 2004.

[51] S. Kay, *Fundamentals of Statistical Signal Processing:Estimation Theory.* Upper Saddle River,NJ: Prentice Hall, 1993.

[52] C. Bianchi and K. Sivaprasad, “A channel model for multipath interference on terrestrial line-of-sight digital radio,” *IEEE Trans. Antennas Propagat.*, vol. 46, no. 6, pp. 891–901, 1998.

[53] M. Rice, “Narrowband channel model for aeronautical telemetry,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1371–1377, Oct. 2000.

[54] D. Matolak and J. Rodenbaugh, “Optimum detection of differentially-encoded mpsk in a dispersive aeronautical channel,” *Proc of the 33rd Southeastern Symp. on System Theory*, pp. 47–51, 2001.

[55] W. Yang, Z. Naitong, Z. Qinyu, and Y. Jun, “A 2-ray path loss model for i-uwb signals transmission,” *Int. Conf. on Wireless Comm., Networking, and Mob. Comput.*, pp. 554–556, Sept. 2007.

[56] C. Potter, A. Panagos, and K. Kosbar, “Optimally training parameters for continuously varying mimo channels,” *Proceedings of the International Telemetering Conference*, Oct. 2005.

[57] G. E. Carlson, *Signal and Linear System Analysis.* New York, NY: John Wiley and Sons, Inc, 1987.

VITA

Christopher Gene Potter was born on May 30, 1981 in St. Charles, Missouri. He received his Bachelors of Science at the University of Missouri-Rolla in May 2004 and his Masters of Science in May 2006 in Electrical Engineering. He received his Ph.D. in Electrical Engineering at the Missouri University of Science and Technology in December 2008. His current research interests are MIMO communications, channel estimation, channel prediction, signal processing, swarm optimization, and radar.