AFRL-RI-RS-TR-2016-126

AVERAGE LIKELIHOOD METHODS OF CLASSIFICATION OF CODE DIVISION MULTIPLE ACCESS (CDMA)

MAY 2016

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW, Wright-Patterson AFB Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2016-126 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
SCOTT S. SHYNE
Chief, Cyber Operations Branch

/ S /
WARREN H. DEBANY JR.
Technical Advisor, Information
Exploitation and Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
May 2016	FINAL TECHNICAL REPORT	Nov 2013 – Nov 2015

4. TITLE AND SUBTITLE

AVERAGE LIKELIHOOD METHODS OF CLASSIFICATION OF CODE DIVISION MULTIPLE ACCESS (CDMA)

5a. CONTRACT NUMBER
IN-HOUSE / R193

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)

Alfredo Vega Irizarry

5d. PROJECT NUMBER
GCIH

5e. TASK NUMBER
OM

5f. WORK UNIT NUMBER
IC

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RIGB
525 Brooks Road
Rome NY 13441-4505

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RIGB
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2016-126

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. PA# 88ABW-2016-2048
Date Cleared: 22 Apr 2016

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This final report summarizes the most significant findings under the in-house effort titled “Average Likelihood For CDMA”. The effort included refinements to the proposed average likelihood method discussed in AFRL-RI-RS-TR-2014-09-093. The refinements include a Taylor’s approximation of the log-likelihood function the development of algorithms to extract features, the inclusion of unbalanced CDMA cases and the extension of the average likelihood to complex cases of CDMA. The performance of the classifier is validated by the use of accuracy versus SNR curves. Classification of CDMA under a hypothesis expressed in terms of code length and active number of user is feasible by performing a simple preprocessing stage at the beginning of the classifier.

15. SUBJECT TERMS

Classification, CDMA, Average Likelihood

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**ALFREDO VEGA IRIZARRY**
U	U	U	UU	111	19b. TELEPHONE NUMBER (*Include area code*) **N/A**

Table of contents

List of figures

List of tables

1 Summary

Signal classification or automatic modulation classification is an area of research that has been studied for many years, originally motivated by military applications and in current years motivated by the development of cognitive radios. Its functions may include the surveillance of signals of interest and providing information to blind demodulation systems.

The problem of classifying Code Division Multiple Access (CDMA) signals in the presence of Additive White Gaussian Noise (AGWN) is explored using Decision Theory. Prior state-of-the-art has been limited to single channel digital signals such as MPSK and QAM, with few limited attempts to develop a CDMA classifiers. Such classifiers make use of the cyclic-correlation spectrum for single user and feature-based neural network approach for multiple user CDMA. Other approaches have focused on blind detection, which could be used for classification in an indirect manner.

The discussion is focused on the development of classifiers using the average likelihood function. This approach will ensure that the development is optimal in the sense of minimizing the error in classification when compared with any other types of classification techniques. However, this approach has a challenging problem: it requires averaging over many unknown parameters and can become an intractable problem.

This research was successful in reducing some of the complexity of this problem. Starting with the definition of the probability of the code matrix and the development of the likelihood of MPSK signals, it was possible to find an analytical solution for CDMA signals with a small code length. Averaging over matrices with the lowest Total Squared Correlation (TSC) allowed simplifying the equations for higher code lengths. The resulting algorithm was tested using Receiver Operating Characteristic Curves and Accuracy versus Signal-to-Noise Ratio (SNR). The algorithm that classifies CDMA in terms of code length and number of active users was extended to different complex types of CDMA under the assumptions of full-loaded, underloaded, balanced and unbalanced CDMA, for orthogonal or quasi-orthogonal codes, and chip-level synchronization.

2 Introduction

2.1 Modulation Classification

As the meaning suggests, *modulation classification* is the determination of the modulation type of unknown signals for the purpose of identification, as it is often the case in military applications, or for channelizing the signal through automating signal processing algorithm, as it is the case of cognitive radio applications. Modulation classification is part of a broader problem known as blind or uncooperative demodulation the goal of which is the extraction the information contents of unknown signals.

An important consideration in solving this problem is the selection of the *approach*. The literature presents three common types of classification methods: ad-hoc, feature-based and decision theoretic classification. Ad-hoc methods are of problematic because they are based on intuition, offering no guarantees in the performance of the classifier and thus becoming questionable for the development of robust telecommunication systems. The feature-based approaches attempt to classify signals based on the extraction of ad-hoc features by using clustering algorithm such as neural networks. The method tries to infer the underlying probabilities of each class during a training session and then assigns classes by measuring some distance metric to the clusters in a test session. The use of these methods becomes relevant in problems where the system's stochastic model is either incomplete or too complex to be described in mathematical terms. Feature based methods often provide an acceptable performance; however, they present two subjective problems: the selection of features and the selection of data sets for training. These weaknesses also leave the uncertainty whether optimal classification is achieved because these methods often deal with a performance that depends on the training data set selected. A third method for classification is model-based decision theoretic. The method makes use of the Bayes Criterion for developing mathematical rules that guarantees optimal performance in noise, i.e., rules that guarantee the lowest error in classification. The method is suitable in problems where models are available and have low complexity. Its main disadvantage is the development of rules due to the mathematical complexity, especially when problems deal with many unknown variables.

The research presented herein is dedicated to the study of classical decision theoretic approaches for the classification of CDMA signals. The *motivation* of this study includes: 1. the need for developing applications that identify and exploit unknown signals, 2. the lack of study done in applying average likelihood techniques to CDMA, and 3. the benefit of implementing optimal classification method that provides reliable classification in a noisy channel.

2.2 Research Objectives

The *research objective* is to detect or classify CDMA signals in the presence of noise. The *scope* of this research includes: the classification of CDMA under various scenarios such as

fully-loaded, underloaded, balanced and unbalanced CDMA. It also covers the classification of four types of CDMA that arise from a combination of BPSK and QPSK signals. For this study, a decision theoretic approach was chosen because of the lack of methods for the classification of CDMA, this method would serve as a future benchmark for comparing alternative classification methods.

An *initial assessment* of the problem made evident that classification of CDMA is extremely challenging. The first major problem is trying to detect CDMA signals that exhibit noise-like characteristics. Such detection would need to rely on the ability to identify distinctive statistical features between CDMA and noise signals. A second problem deals with the complexity of the CDMA model. In CDMA, the large number of unknowns turns the classical decision theory into a difficult task. In the pursuit of this goal, the research will consider the development of MPSK decision theoretic classifier as a model to follow.

The prior *state-of-the-art* in modulation classification lacks of modulation classification techniques for CDMA. A survey on modulation classification methods shows that prior decision theoretic approaches have been applied to single user digital signals. The few approaches related to CDMA are based on the concept of cyclostationary features. Two of these approaches are limited to single user CDMA. A third method is not a classification method in itself, but a blind demodulation method that can be used for parameter estimation in a Generalized Likelihood Ratio Test (GLRT) classifier. The algorithm has been tested up to code lengths of 32 with 8 active users. The idea of performing brute force demodulation prior to the detection will be ruled out as an alternative due to its overwhelming computational costs and convergence problems.

The classification of CDMA problem started off the well-known MPSK classifier proposed in [4]. The intuition suggest that a CDMA classifier is some sort of convoluted form of an BPSK classifier because CDMA is based on BPSK signals. This reasoning proved to be somewhat correct because all CDMA classification rules result in a BPSK rule when the number of active users is one. The development of MPSK classifier also served as a starting point to other optimal signal classifiers such as QAM and FSK.

One of the most obvious challenges in the developing an *Average Likelihood Classifier for CDMA* is averaging over unknown variables. The number of unknowns grows proportional to the dimensions of the code matrix used for generating CDMA; however, typical code matrices such as Hadamard matrices occurs at code lengths of powers of 2. For these codes, the dimensions grows also exponentially, which clearly presents a problem.

A key finding in this research was the proposition of a weight for averaging CDMA codes. This weighting function is referred in this discussion as the probability of the code matrix. This weight is based on the Total Squared Correlation and was a key factor in the development of a simplified decision rule. Other consideration is the form of the simplified average likelihood function. It was found that this function can be expressed as a product of hyperbolic cosine functions and decaying exponentials. Many of these decaying exponentials are a function of an

introduced parameter referred as the precision of the probability of the code. Increasing the precision of our classifier eliminated many terms of the average likelihood. This simplification is important because the likelihood function can easily become intractable or show numerical problems.

The major *contribution* of this study is the development of the *first average likelihood classifier for multiuser CDMA*. The hypothesis under test is the code length and the number of users, which in some aspect is analogous to the classification of M-ary PSK with M as the hypothesis under test. The simplified average likelihood of CDMA was developed using a standard procedure, so the likelihood function can be used against any other signal such as MPSK or QAM. In this research, CDMA was tested against BPSK and QPSK, which are special cases of CDMA under the assumption the the code length equals one. The same average likelihood for CDMA made of BPSK symbols can be reused and applied to all types of CDMA generated from combinations of BPSK and QPSK symbols without adding more complexity.

2.3 Roadmap

The research path is provided in Figure 1. It shows the main step taken to solve the problem. The objective is to detect/classify CDMA signals in AWGN, Box A in the diagram. The study begins with a brief discussion of Detection Theory and how this theory was applied to the previous state-of-the-art as discussed in Chapter 3. In the case of single user signals, modulation classification associates a digital signal to a constellation. The hypothesis is the signal type and in the case of MPSK and QAM signals, the type is associated to the size of the constellation. In the case of CDMA, the modulation type will be associated to the code length and the number of active users.

The based model for developing and simplifying the average likelihood function of CDMA signals is the BPSK, Box B in the diagram. Both signals are based on binary antipodal variables and, at the end of the development, their likelihoods share some similarities. The likelihood for one BPSK symbol is very simple. It is expressed as a product of an exponential function in terms of the signal-to-noise (SNR) ratio. However, the likelihood of a CDMA is extremely complicated for a CDMA. This problem was solved by using symbolic algebra algorithms (Box C) for small code lengths as discussed in Chapter 4.

After analysing code lengths of 2, 3 and 4, the research turns its attention to the development of a mathematical formulation CDMA (Box D) with the following assumptions:

- chip-synchronous CDMA,
- frame-asynchronous CDMA,
- type 1 (BPSK-code/BPSK-data),
- unknown orthogonal code matrix (unknown permutations),

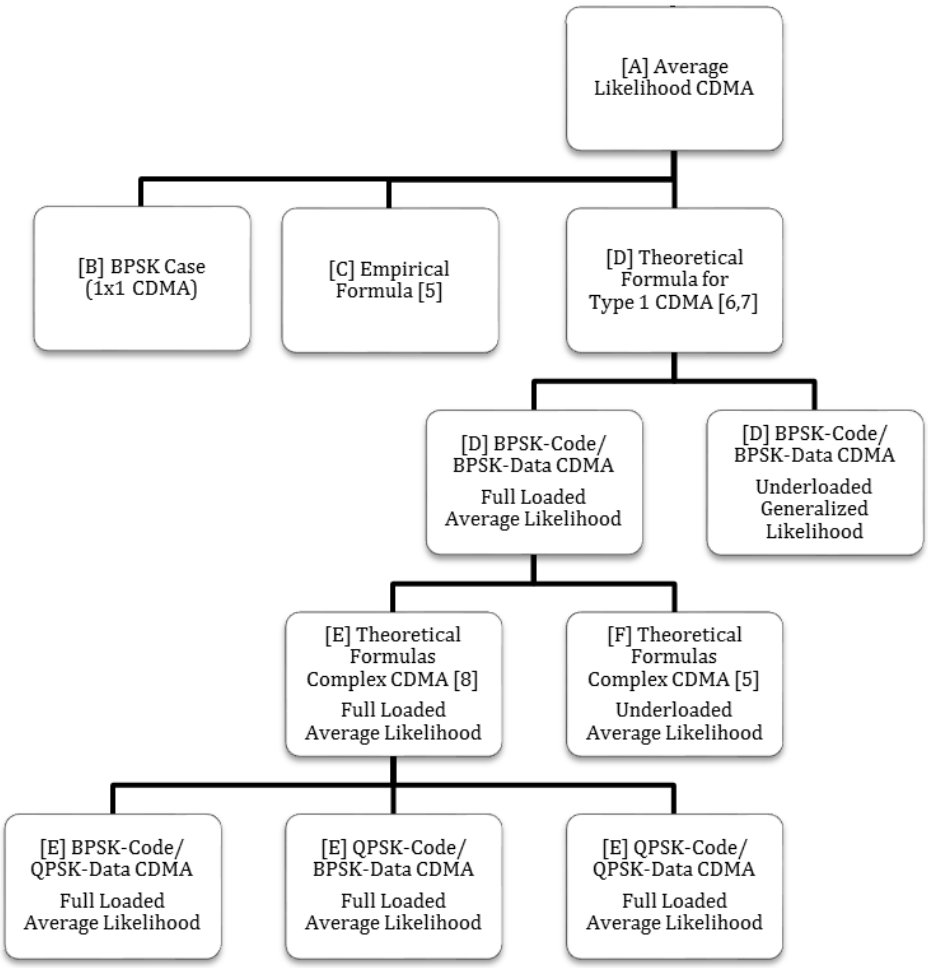

Fig. 1 Research Path

- unknown data vector,
- unknown code length and number of users (hypothesis),
- fully-loaded.

The case of unbalanced CDMA and averaging over Gaussian Distributed amplitude deviations over a nominal energy per symbol (Box F). This work has been presented in MILCOM 2014. [5] The same development was slightly modified and applied to other types of CDMA generated from a combination of QPSK and BPSK symbols (Box E). A key approach was to represent complex models in terms of real block matrices and apply an approach similar to type 1 CDMA. A technical paper has been submitted to MILCOM 2016. [6]

2.4 Organization

This publication is organized in six sections. This introduction covers the problem of modulation classification and discusses briefly the research objectives. In section 3, the discussion covers some technical background with the prior state-of-the-art; discusses some essential concepts found in decision theory; introduces the potential approaches; and provides and insight of how decision theory is applied to the classification of MPSK signals. section 4, considers the classification problem applied to CDMA and provide the unique mathematical framework. The average likelihood is simplified in section 5. In section 6 the concept is extended to other types of CDMA derived from combinations of BPSK and QPSK symbols. Finally, section 7 provides a summary of the findings, future work and conclusions.

3 Modulation Classification Methods

This section provides a brief review of the concepts found in decision theory and how they are applied to the problem of modulation classification. Existing methods in modulation classification can be grouped into three main categories which are discussed in the following sections. Feature-based and decision theoretic provide a strong mathematical framework and become the main subjects in the discussion. The decision theoretic method will be the method of choice for developing CDMA classification rules after evaluating the prior state-of-the-art and their weaknesses. Selecting a decision theoretic approach is risky research due to the complexities in averaging the CDMA model, but this risk is compensated by developing a new algorithm based on optimization principles.

3.0.1 Ad Hoc

As mentioned before, Ad Hoc classifiers consist of deriving intuitive rules for classification. Several of these methods have been discussed in [7] and applied to military applications. Some of them are simple as: thresholding signals, counting samples, generating histograms [8] or applying mathematical rules such and the well-known M^{th}-Power Classifier for MPSK signals. The construction of these rules do not allow for analytically predicting the performance. Although simple rules such as the M^{th} power law could be approximations of optimal rules, they are commonly considered of limited significance unless their performance is supported by some theoretical framework.

3.0.2 Feature Based Methods

A more scientific approach consists of identifying important features of a signal and try to cluster the features in a multidimensional space. The different clusters will be associated to a modulation type by using a clustering technique such as neural network. The approach is a good choice when the there is little known about the stochastic model. Examples of these methods are the Statistical-Moment Based Classifier presented by [9]. The rationale is that statistical moments derived from signal parameters such as frequency, amplitude or phase can provide useful classification features for a neural network classifier.

A reference of this technique is the method proposed by [1] for classifying single user CDMA using cyclostationary features, i.e., the peaks found in the cyclic correlation spectrum. These features are obtained by applying a Fast Fourier Transform (FFT) and averaging over frequencies as illustrated in Figure 2. Frequency smoothing is used to find significant peaks which are fed into a neural network.

The selection of peaks as features is motivated by the signal's cyclostationary properties of a single CDMA user. The referenced method is limited to a single spreading sequence and classifying multiuser CDMA signals is out of the scope of this particular development. Under the proposed example, the sequence resulting from the modulation of a QPSK spreading and

a BPSK modulation is another QPSK signal. So in reality this neural network classifier is processing periodic QPSK signals for ultrawide band radar applications rather than for multi-user CDMA for telecommunications. The report tested the classification of two pseudo noise sequences with code lengths 15 and 31. The classifier was able to achieve accuracies between 86.5 and 100 percent for $SNR = -3dB$ and $SNR = \infty$ respectively.

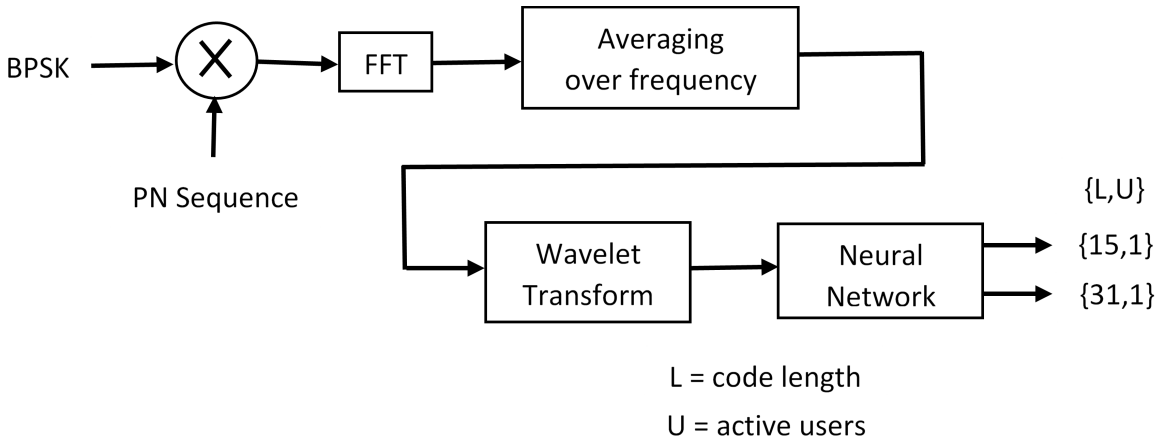

Fig. 2 Feature-Based Neural Network Classifier

The performance of a neural network depends on these features and so there is no guarantee that the features selected is a fairly complete representation of all the features needed for a correct classification. In addition to this fact, the performance of clustering methods is known to depend on the training data.

3.0.3 Decision Theoretic Methods

Decision theoretic methods deal with the problem of optimization of metrics: the minimization of a cost function or the maximization of the accuracy. The theory is based on classical detection theory [10]. Bayes Theorem provides the statistical model that characterizes the source, the channel and the observation space as shown in Figure 4. In our problem, the source generates a finite set of known classes $\{\mathscr{H}_i\}$ representing the modulation types. Each class is associated to probability $p(\mathscr{H})$ referred as the prior probability. The source generates an element of a class and sends it through a noisy channel which is characterized by a likelihood probability $p(r|\mathscr{H})$ or the probability of the observation when a class in a noisy channel. The observation r is fed into a detector that contains a classification rule based on optimality principles.

Table 1 shows four types of outcomes may occur in a binary classification. Similar concepts can be extended to multiple classes by constructing an error probability matrix of multiple dimensions. The sum of non-diagonal elements represents the total probability of error in classification while the sum of diagonal elements represents the accuracy. The construction of

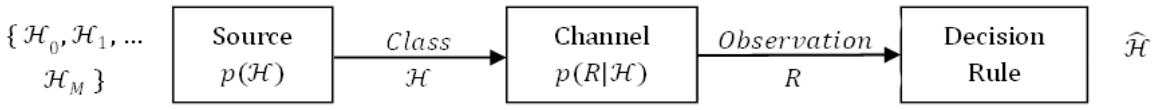

Fig. 3 Bayesian Model

Table 1 Error Probability Matrix

True Class:	Class $\mathscr{H}_0$	Class $\mathscr{H}_1$
Decide for $\mathscr{H}_0$	True Negative $P(Decide\ \mathscr{H}_0 \mid \mathscr{H}_0)$	False Negative $P(Decide\ \mathscr{H}_0 \mid \mathscr{H}_1)$
Decide for $\mathscr{H}_1$	False Positive $P(Decide\ \mathscr{H}_1 \mid \mathscr{H}_0)$	True Positive $P(Decide\ \mathscr{H}_1 \mid \mathscr{H}_1)$

an optimal decision rule shall maximize the sum of the diagonal elements and minimize the non-diagonal elements.

The decision process requires establishing thresholds or decision regions in the observation space of r. A cost function can be assigned to each decision, but these costs are commonly ignored in classification problems. The optimization of the decision regions for achieving maximum accuracy is equivalent to maximizing the posterior probability according to (1).

$$\begin{aligned} \text{Decide for } \mathscr{H} &= \underset{\mathscr{H}\in\{\mathscr{H}_0,\mathscr{H}_1,\ldots,\mathscr{H}_N\}}{\arg\max} p(\mathscr{H}|r)p(r) \\ &= \underset{\mathscr{H}\in\{\mathscr{H}_0,\mathscr{H}_1,\ldots,\mathscr{H}_N\}}{\arg\max} p(r|\mathscr{H})p(\mathscr{H}) \end{aligned} \tag{1}$$

When the priors are not available, a decision is taken by choosing the maximum likelihood as shown in (2). Ignoring the priors is equivalent to assuming that all prior probabilities are equal.

$$\text{Decide for } \mathscr{H} = \underset{\mathscr{H}\in\{\mathscr{H}_0,\mathscr{H}_1,\ldots,\mathscr{H}_N\}}{\arg\max} p(r|\mathscr{H}) \tag{2}$$

In binary classification, the same criterion can be expressed in a form of a Likelihood Ratio Test (LRT) as shown in (3). The determination of the likelihood function and its simplification is an important step in deriving optimal rules for modulation classification. Computing the likelihood becomes more complicated when the likelihood involves multiple unknown

Table 2 Hypotheses in Various Modulation Classification Problems

Problem	Hypothesis	Cluster
Detection of BPSK	$\mathscr{H} \in \{\pm 1\}$	point in the $\mathbb{C}$ plane
Classification of MPSK	$\mathscr{H} \in \{M = 2, 4, ...\}$	constellation in the $\mathbb{C}$ plane
Classification of CDMA	$\mathscr{H} \in \{L, U\}$	vectors $\vec{a} \in \mathbb{R}^L$ $\vec{a}^T \vec{a} = L \cdot U$

parameters.

$$\frac{p(r|\mathscr{H}_1)}{p(r|\mathscr{H}_0)} \underset{Decide\ \mathscr{H}_0}{\overset{Decide\ \mathscr{H}_1}{\gtrless}} \frac{p(\mathscr{H}_0)}{p(\mathscr{H}_1)} \tag{3}$$

Modulation classification deals with signals with unknown parameters in noise. The resulting likelihood is conditioned on a parameter vector $\vec{\theta}$ that is unknown to the classifier and has to be supplied in some way prior to classification. Often, the parameters are treated as random variables and the expectation must be taken over all the possible values of $\vec{\theta}$ according to (4) in order to remove the condition according to the Bayes Theorem.

$$p(r|\mathscr{H}) = \mathbb{E}_{\vec{\theta}}\{p(r|\mathscr{H}, \vec{\theta})\} \tag{4}$$

A hypothesis can be interpreted as the label of a region belonging to a class in the observation space. (See Table 5) In detection problems, the hypothesis is the symbol. In blind demodulation of single user signals, the hypothesis can be visualized as a label associated to a constellation in the complex plane. In the case of CDMA, the hypothesis can be visualized as a label to associated to a collection of particular vectors.

3.1 Decision Theory for Modulation Classification

Stochastic processes are characterized by probabilities that evolve over parameters such as time t which are not associated to a random variable. A stochastic processes $n(t)$ is characterized by the statistical moments. Modulation classification often deals with wide-sense stationary

processes characterized by having a zero mean and constant autocorrelation function:

$$\begin{aligned} \mathbb{E}\{n(t)\} &= 0, \\ \mathbb{E}\{n(t)n^*(t+\Delta t)\} &= \frac{N_0}{2}\delta(\Delta t). \end{aligned} \tag{5}$$

Modulation classification also deals with cyclostationary processes, i.e., processes that have a periodic nature in the correlation function $\mathscr{R}$ as shown in (6). The parameter α in this equation represents the cyclic-correlation period.

$$\begin{aligned} \mathscr{R}(\Delta t) &= \mathbb{E}\{r(t)r^*(t+\Delta t)\} \\ \mathscr{R}(\Delta t+\alpha) &= \mathscr{R}(\Delta t) \end{aligned} \tag{6}$$

A common channel found in telecommunications is the Additive White Gaussian channel given by (7). The modulated signal is transmitted through a noisy channel with noise $n(t)$. At the receiver, we have the observation in the form of a process $r(t)$. A classification rule developed from (1) or (4) requires using random variables. It would be impractical to characterize the likelihood using an infinite number of observations generated by $r(t)$ at every instant.

$$\begin{aligned} r(t) &= s(t,\vec{\theta}) + n(t) \\ n(t) &\sim \mathscr{N}(\mu,\sigma^2) \end{aligned} \tag{7}$$

Parametrizing the likelihood in terms of t is inconvenient because the observation $r(t)$ has infinite dimensions and the likelihood of a given ensemble of $r(t)$ is zero. A practical way of handling processes is by approximating the observation to a finite set of observations by using some finite approximation. A finite set of observations can be achieved by defining a complete orthonormal (CON) set $\{\psi_i(t)\}$ in (8) and approximating $r(t)$ in terms of a finite set of independent coefficients $\{r_k(\vec{\theta})\}$.

$$\begin{aligned} CON &= \{\psi_k(t)\}^{k=0:\infty} \\ \langle \psi_i(t), \psi_j(t) \rangle &= \int_{-\infty}^{\infty} \psi_i(t)\psi_j^*(t)dt = \delta_{i,j} \end{aligned} \tag{8}$$

In modulated sequences, a convenient choice for our basis is a set of normalized, non-overlapping pulses with a width T delayed by kT according to (9).

$$\psi_k(t) = \begin{cases} \dfrac{1}{T} & \text{for } T\ k \le t < T(k+1) \text{ and k Integer} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The reduction in dimensionality is achieved by truncating the series to a finite number of coefficients ($K < \infty$) according to (10). The likelihood $p(\{r_k\}^K|\mathscr{H})$ will be constructed from the statistics of each random variables r_k. For a linear Gaussian process $r(t)$, the statistics of these random variables are also Gaussian.

$$\begin{aligned} r_k(\vec{\theta}) &= \langle r(t), \psi_k(t) \rangle \\ r(t) &\approx \sum_{k=0}^{K-1} r_k(\vec{\theta}) \psi_k(t) \end{aligned} \tag{10}$$

Eliminating the dependency of the observation parameters requires either an estimate of the parameter vector or performing the expectation over all the unknowns as shown in (4). If the last method is selected, the expectation can be calculated in several ways. The exact computation of the expectation results in average likelihood function and the classifier is an Average Likelihood Ratio Test (ALRT). An approximation of the expectation results in Quasi-Average Likelihood Ratio Test (QALRT). The substitution of $\vec{\theta}$ by estimated parameters results in a Generalized Likelihood Ratio Test (GLRT). A combination of any of these methods results in a Hybrid Likelihood Ratio Test (HLRT).

3.2 Modulation Classification in the Literature

Table 3 shows a list of known methods adapted from [2]. Likelihood functions are specific to modulation signals and their respective channel models which may be seen as a disadvantage because changing the model will require a new computation of the likelihood function. The most popular classifiers are based on single channel BPSK, QPSK, MPSK and QAM signals.

Decision theoretic methods are less abundant in the literature when compared to feature-based approaches. This can be attributed to the complexity in the development of such algorithms. Decision theoretic methods based on likelihood functions require good signal models and few unknown parameters. To overcome some of the difficulties in calculating the average likelihood function, authors recur to sub-optimal approaches such as QLRT, GLRT and HLRT. The case of CDMA has simple models, but as the number of unknowns increases, the averaging process translates into higher computational costs.

From the survey of decision theoretical methods, it is evident that there was a lack of likelihood methods for classifying of CDMA signals and it is speculated that the absence of CDMA methods is related to the complexity of the averaging process. A contributions of this research include the development and publication of classifier for BPSK-code/BPSK-signals in 2014 and the submission of a similar paper on complex CDMA classification in 2016.

Table 3 Survey of Likelihood-Based Classifiers

Authors	Classifier	Modulation	Unknowns	Channel
Sills	ALRT	BPSK, QPSK, 16QAM, V29, 32QAM, 64QAM	carrier phase	AWGN
Wei, Mendel	ALRT	16QAM, V29	-	AWGN
Kim, Polydoros	QLRT	BPSK, QPSK	carrier phase	AWGN
Sapiano, Martin	ALRT	UW, BPSK, QPSK, 8PSK, 16PSK	carrier phase	AWGN
Long	QLRT	UW, BPSK, QPSK, 8PSK	carrier phase, timing offset	AWGN
Hong, Ho	ALRT	BPSK, QPSK	symbol level	AWGN
Beidas, Weber	QLRT	32FSK, 64FSK	phase jitter	AWGN
Beidas, Weber	QLRT	32FSK, 64FSK	phase jitter, timing	AWGN
Panagiotu	GLRT, HLRT	16PSK, 16QAM	carrier phase	AWGN
Chugg	HLRT	BPSK, QPSK, OQPSK	carrier phase, signal power, PSD	AWGN
Hong, Ho	HLRT	BPSK, QPSK	signal level	AWGN
Hong, Ho	HLRT	BPSK, QPSK	angle of arrival	AWGN
Dobre	HLRT	BPSK, QPSK, 16QAM, V29, 32QAM, 64QAM	channel amplitude, phase	flat fading
Abdi	ALRT, QLRT	16QAM, 32QAM, 64QAM	Channel amplitude and phase	flat fading
Vega-Irizarry, Fam	ALRT	CDMA type 1	code, data vector	AWGN
Vega-Irizarry, Fam	ALRT	CDMA types 2-4	code, data vector	AWGN

3.2.1 Generalized Likelihood Methods

A generalized classification method for CDMA may consist of making a estimate of the spreading matrix C and data vector $\vec{b}$ with code length L, number of users U and energy per

symbol E, as defined in the CDMA model of (11). If an accurate estimate is available, a generalized classification rule would outperform the results of an average likelihood classifier. However, in the case of CDMA, such estimation would be at the expense of high computational costs and therefore the average approach would seem as a more viable approach. For comparison purposes, we consider the development of a GLRT using Maximum A Posteriori (MAP) or Expectation Maximization for estimating unknown parameters prior to classification.

$$\begin{gathered} \vec{y}_k = \sqrt{\frac{E}{L}}\, C\, \vec{b}_k + \vec{n}_k \\ C \in \{\pm 1\}^{L \times U} \\ \vec{y}_k \in \{\pm 1\}^{L} \\ \vec{n}_k \sim \mathscr{N}(0, N_0/2\, I) \\ \text{for } k = 0, 1, ..., K-1 \end{gathered} \tag{11}$$

The performance of generalized methods depends on the accuracy of the parameter estimate, so this approach can be wasteful in computational resources when estimating the parameters of wrong classes. For example, matrices of size 4×4 have a total of 2^{16} distinct realizations using BPSK symbols. An algorithm would need to be smart enough to discard many useless code matrices that would never be used in CDMA transmissions. It would have to know that from a total of 2^{16}, only 768 are orthogonal matrices suitable for CDMA, all of them permutations of one single matrix. As the dimensions of the code matrix increase, this kind of processing becomes infeasible.

3.2.1.1 Maximum Likelihood Estimator

The parameters C and $\vec{b}$ can be estimated using a maximum likelihood estimator. Given a likelihood function of a multivariate Gaussian stochastic process (12), one can assume the values L and U and try to estimate the parameters. Because we are dealing with observations $\vec{y}_k$ of different code lengths, this would require reformatting the received vector in different code lengths.

$$p(\{\vec{y}_k\}^K | \mathscr{H}, C, \{\vec{b}\}^K) = \prod_{k=0}^{K-1} \frac{1}{(\pi N_0)^{L/2}} exp(\|\vec{y}_k - \sqrt{E/L}\, C\, \vec{b}_k\|^2 / N_0) \tag{12}$$

Under unknown code matrix and data vector samples, the problem of maximizing the likelihood of a CDMA signal (13) does not have a closed form solution. In addition, its large number of variables makes the computation of the parameters extremely difficult to estimate due to the limitation of numerical methods use to solve the equations. Using an average likelihood estimate of the parameters for classification purposes will not solve the classification problem

in an efficient way.

$$\begin{aligned} \nabla_{\vec{b}} p(\vec{y}_k | L \times U \; CDMA, C, \vec{b}) &= 0 \\ \rightarrow \sum_k C^H (\vec{y}_k - \sqrt{E/L}\, C\, \vec{b}_k) &= 0 \\ \nabla_C p(\vec{y}_k | L \times U \; CDMA, C, \vec{b}_k) &= 0 \\ \rightarrow \sum_k \vec{b}_k^H (\vec{y}_k - \sqrt{E/L}\, C\, \vec{b}_k) &= 0 \end{aligned} \tag{13}$$

3.2.1.2 Expectation Maximization Approach

An Expectation Maximization (EM) approach for CDMA for blind detection [11] provides an iterative way of estimating the code matrix C as an updatable parameter and the data vector $\vec{b}$ as a hidden random variable of the EM algorithm. EM belongs to the group of statistical inference methods in which the parameter is adjusted such that it minimizes the distance between an empirical distribution and a model distribution [12].

The proposed algorithm consists of four major steps.

1. Project $\vec{y}$ into a signal space.
2. Define Ω as the updatable code matrix given by (16).
3. Estimate $\vec{b}$ using a MMSE estimator.
4. Maximize the expectation to obtain an update of Ω.
5. Repeat the last two steps until convergence is achieved.

The algorithm assumes that the code length L is known and the number of users U is determined by setting a threshold on the singular values of the correlation matrix $\mathscr{R}$.

$$\mathscr{R} = \begin{bmatrix} V_s & V_n \end{bmatrix} \begin{bmatrix} \Lambda_s & 0 \\ 0 & \Lambda_n \end{bmatrix} \begin{bmatrix} V_s & V_n \end{bmatrix}^H = \mathbb{E}\{\vec{y}\,\vec{y}^T\} \tag{14}$$

The eigenvalues of the signal space given by Λ_s contain the energy of the signal. The eigenvalues of the noise space given by Λ_n contain only noise energy as shown in (15). The eigenvectors in V_s associated with the signal space contain information about the code matrix used to generate the CDMA signal.

$$(\Lambda_s)_{i,i} = \begin{cases} E_i/L + N_0/2 & \text{for signal present} \\ 0 & \text{for noise only} \end{cases}, \; (\Lambda_n)_{i,i} = \begin{cases} 0 & \text{for signal present} \\ N_0/2 & \text{for noise only} \end{cases} \tag{15}$$

The observation $\vec{r}$ is the projection of $\vec{y}$ into the signal space. This is an estimate value because it requires a threshold to separate the signal from the noise space in (14).

$$\begin{aligned} \vec{r} &= V_s^H \, \vec{y} \\ \Omega &= V_s^H \, C \end{aligned} \tag{16}$$

The updatable matrix Ω is the product of the projection and the code C. Both Ω and $\vec{r}$ are parameters of the likelihood function given below.

$$p(\vec{r}, \vec{b}(\vec{y})|\Omega) = \frac{1}{(\pi \, N_0)^{U/2}} \, e^{-\|\vec{r} - \sqrt{E/L}\,\Omega\,\vec{b}\|^2/N_0} \tag{17}$$

The data vector $\vec{b}$ can be calculated using a Minimum Mean Squared Error (MMSE) detector [13] as shown in (18).

$$\begin{aligned} \hat{\vec{b}} &= sign(\vec{w}^T \, \vec{r}) \\ \vec{w} &= \arg\min_{\vec{w}} \mathbb{E}\{\vec{b} - \hat{\vec{b}}\} \end{aligned} \tag{18}$$

The updated Ω is calculated from the EM algorithm in (19).

$$\hat{\Omega} = \arg\max_{\Omega} p(\{\vec{r}_k\}^K, \{\vec{b}_k\}^K|\Omega) \tag{19}$$

The computation of the priors $p(\{\vec{b}_k\}^K)$ presents a problem in the algorithm because computing $\vec{b}_k$ for possible 2^U values becomes computationally expensive. Instead, the author has decided to simplify the prior probability by forcing it to be constant $p(\{\vec{b}_k\}^K) = 1/2^{UK}$.

As it was mentioned before, estimating the code matrix using (19) would be impractical for modulation classification because any attempt to demodulate a wrong hypothesis results in a waste of computational resources and processing time. Also, the blind demodulation algorithm has been tested up to code lengths of 32 and 8 active users. Extending the algorithm to higher code lengths would be unusable because the convergence of the EM would degrade as the number of clusters increases. The number of clusters would be equivalent to the number of elements in the code matrix. EM is a method of choice when the number of clusters is relatively small for quick convergence, which is not the case for classifying higher code lengths. So in conclusion, the possibility of using EM blind demodulation method for constructing GLRT modulation classifier can be ruled out as a viable alternative.

3.2.2 The Average Likelihood Method

In this section, the average likelihood for CDMA approach will be studied by following the development of the MPSK classifier. First, it would be assessed if constructing a simplified likelihood function is feasible and meaningful for classification purposes. The strategy, if correctly implemented and tested, is expected to reduce to a BPSK likelihood function for code lengths of 1.

3.2.2.1 MPSK Model

Ideal MPSK signals in (20) are constructed from a sequence of orthonormal pulses (9) with amplitude $\sqrt{E}$ and a parameter vector $\vec{\theta} = \{\varepsilon, \{b_k\}^K\}$ where εT is a time delay expressed as a fraction ε of the pulse width T. The set $\{b_k\}^K$ contain the signal's symbols.

$$\begin{gathered} x(t) = x(t,\vec{b}) = \begin{cases} \sqrt{E}\sum_{k=0}^{K-1} b_k\ \psi_k(t) & \text{for } 0 \le t < KT \\ 0 & \text{otherwise} \end{cases} \\ b_k\ \in\ \{e^{i2\pi m/M} \mid m = 0,1,...,M-1\} \\ \text{for } k = 0:(K-1) \end{gathered} \tag{20}$$

Figure 4 shows the block diagram of the source, channel and classification process. The source generates a symbol at each time interval. The variable K represents the total number of symbols in the MPSK signal. The modulation is a form of encoding represented by b_k which can take M possible phase values. The encoding produces the signal $x(t)$ which is corrupted with AWGN $n(t)$ and produces the received signal $r(t)$. The received signal is decorrelated using the pulses which forms an orthonormal set. A delay ε is added at the input of the correlator to represent the chip asynchronous nature of the correlator. The delay takes the value within the range $[0,1)$. The coefficients $\{r_k\}^K$ become a finite set of observations which is processed by the classifier.

The CON representation in (21) allow us to deal with a set of K random variables which are the decorrelated symbols at the receiver. The concept is also applied to the unnormalized transmit signal $s(t)$ and the noise $n(t)$ due to the linearity of the process.

$$\begin{aligned} y_k(\varepsilon) &= \langle y(t-\varepsilon T), \psi_k(t)\rangle \\ s_k(\varepsilon) &= \langle \sum_{k=0}^{K-1} b_k\ \psi_k(t-\varepsilon T), \psi_k(t)\rangle \\ n_k(\varepsilon) &= \langle n(t-\varepsilon T), \psi_k(t)\rangle \\ y_k(\varepsilon) &= s_k(\varepsilon) + n_k(\varepsilon) \end{aligned} \tag{21}$$

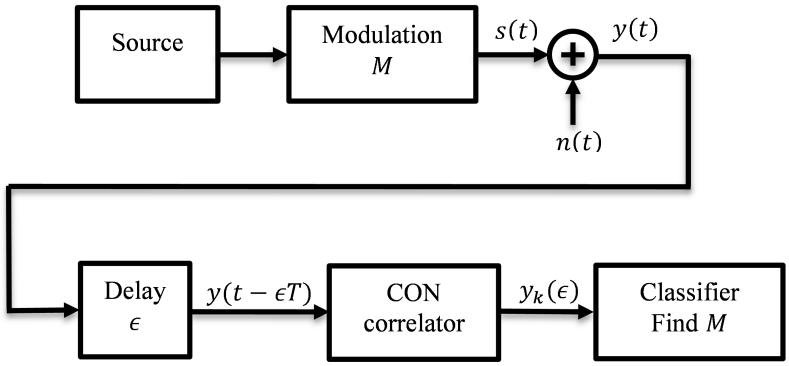

Fig. 4 Block Diagram of Preprocessing Stage

The statistics of the noise coefficients in (22) can be easily obtained by calculating (5) from the series representation. The autocorrelation formula of $n_k(\varepsilon)$ is obtained from the scalar product $\langle N_0/2\ \delta(\tau), \psi_l(\tau)\rangle$.

$$\begin{aligned} n(t) &= \sum_{k=0}^{K-1} n_k(\varepsilon)\psi_k(t) \\ \mathbb{E}\{n_k(\varepsilon)\} &= 0 \\ \mathbb{E}\{n_k(\varepsilon)\ n_l^*(\varepsilon)\} &= \frac{N_0}{2}\delta_{k,l} \end{aligned} \tag{22}$$

The output of the correlator is normalized so its variance is one. The conditional likelihood is given by (23). The SNR γ is assumed be known prior to the classification.

$$\begin{aligned} p(r_k(\varepsilon)|\mathscr{H}, b_k, \varepsilon) &= \frac{1}{\sqrt{\pi\, N_0}}\, e^{-\|\vec{r}_k(\varepsilon)\|^2/2 - \sqrt{2\gamma}\, Re\{\vec{r}_k(\varepsilon)\ b_k^*\} - \gamma} \\ \vec{r}_k(\varepsilon) &= \frac{1}{\sqrt{N_0/2}}\langle y(t), \psi_k(t - \varepsilon T)\rangle \\ \gamma &= \frac{E}{N_0} \end{aligned} \tag{23}$$

This model assumes phase synchronization for simplicity. The conditional likelihood of the MPSK signal in (23) is calculated by assuming independent symbols. The average likelihood function is calculated in the next step.

3.2.2.2 Averaging over Symbols

Under the assumption of independent symbols, the conditional likelihood function is given by (24).

$$\lambda(\{r\}^K|\mathscr{H}=M,\{b_k\}^K,\varepsilon)=\prod_{k=0}^{K-1}p(r_k|\mathscr{H}=M,b_k,\varepsilon) \tag{24}$$

The condition on the symbols $\{b_k\}^K$ is eliminated by taking the expectation (4) over all possible MPSK symbol values and results in a product of sums of hyperbolic cosine functions given by:

$$\lambda(\{r\}^K|\mathscr{H}=M,\varepsilon)=e^{-\|r_k(\varepsilon)\|^2-K\,\gamma}\prod_{k=0}^{K-1}\sum_{m=0}^{M/2-1}cosh(\sqrt{2\gamma}\,Re\{r_k(\varepsilon)\,e^{-i2\pi m/M}\}). \tag{25}$$

The development of the average likelihood in [4] resulted in a familiar solution (26) when comparing hypothesis M against $M/2$. The average likelihood over $\{b_k\}$ contains the power-law classifier derived from an Ad-Hoc rule. This classifier can be interpreted as follows: if a signal belongs to class M, then r_k raised to the M^{th} power results in a constant value; if a signal belongs to class $M/2$, then the operation would results in noise.

$$log\left(\frac{\lambda(\{r_k\}^K|M,\varepsilon)}{\lambda(\{r_k\}^K|M/2,\varepsilon)}\right)\approx log(\mathbb{E}_\varepsilon\{exp(\frac{2}{M}(\frac{\gamma}{2})^{M/2}Re\{\sum_{k=0}^{K-1}r_k(\varepsilon)^M\})\}) \tag{26}$$

3.2.2.3 CDMA Model in Time Domain

A time-domain CDMA signal is constructed from a set mutually orthogonal or quasi-orthogonal BPSK waveforms $p_u(t)$ for $u\in\{0,1,...,L-1\}$ given by (27).

$$\begin{gathered}p_u(t)=\sum_{l=0}^{L-1}c_{l,u}\,\psi_l(t)\text{ for }u=0,1,...,L\\ c_{k,u}\in\{\pm1\}\end{gathered} \tag{27}$$

The coefficients $c_{i,u}$ are chosen to ensure the orthogonality of the set of waveforms. Each waveform will serve as a separable coding channel.

$$\langle p_u(t),p_v(t)\rangle=\begin{cases}L^2 & u=v\\ \approx 0 & u\neq v\end{cases} \tag{28}$$

An ideal $L \times U$ CDMA signal $x(t)$ is defined in terms of the sum of modulated versions of these waveforms with a set of modulation parameters $\{b_{u,k}\}^{U \times K}$ where the variable K now represents the total number of chip intervals in a single transmission. Each waveform u is normalized with a factor $1/\sqrt{L}$ such that the waveform energy is 1. The product of the data vector and normalized waveform has energy E_u.

$$x(t) = x(t, C, \{\vec{b}\}^{K/L}) = \begin{cases} \sum_{k=0}^{K/L-1} \sum_{u=0}^{U-1} \sqrt{\frac{E_u}{L}} b_{u,k}\, p_u(t - kLT) & \text{for } 0 \le t < KT \\ 0 & \text{otherwise} \end{cases}$$
$$b_{k,u} \in \{\pm 1\} \tag{29}$$

The channel is characterized by an AWGN process. Similar CON pulses to (9) are used at the correlator. Each coefficient has two subscripts i and k that identify the location of the i^{th} pulse within the k^{th} symbol interval. Each coefficient is a function of ε which now characterizes a frame asynchronous version of the CDMA.

$$\begin{aligned} r_{i,k}(\varepsilon) &= \frac{1}{\sqrt{N_0/2}} \langle y(t - \varepsilon T), \psi_i(t - kLT) \rangle \\ s_{i,k}(\varepsilon) &= \langle \sum_{k=0}^{K/L-1} \sum_{u=0}^{U-1} b_{u,k}\, p_u(t - kLT - \varepsilon T), \psi_i(t - kLT) \rangle \\ n_{i,k}(\varepsilon) &= \langle n(t - \varepsilon T), \psi_i(t - kLT) \rangle \\ \varepsilon &\in 0, 1, ..., L-1 \end{aligned} \tag{30}$$

The observation and noise statistics are obtained in the same manner as those in (22).

$$\begin{aligned} \mathbb{E}\{n_{i,k}(\varepsilon)\} &= 0 \\ \mathbb{E}\{n_{i,k}(\varepsilon) n^*_{l,m}(\varepsilon)\} &= \frac{N_0}{2} \delta_{i,l} \delta_{k,m} \end{aligned} \tag{31}$$

For balanced CDMA, we define a chip signal-to-noise ratio (33) in terms of a nominal value of the energy per symbol, noise power value N_0 and code length as:

$$\gamma_c = \frac{E}{N_0 L}. \tag{32}$$

The chip-level SNR is not a good choice for a parameter because of its dependency on the hypothesis. A better choice is the ratio of the total energy density (E_T/K) and the noise power density value N_0. For simplicity, the discussion will refer to this parameter as the density ratio. The formula is derived assuming that the total energy E_T equals the energy per symbol times

the number of users ($E \cdot U$) multiplied by the total number of symbol intervals (K/L).

$$\text{Density Ratio} = \frac{E_T}{KN_0} = \gamma_s \frac{L}{U} \tag{33}$$

All this information is sufficient for expressing the conditional likelihood of the $L \times U$ CDMA signal per symbol frame.

$$\begin{gathered}
\lambda(\vec{r}_k|\mathscr{H} = \{L,U\}, C, \vec{b}, \varepsilon) = \\
\frac{1}{(\pi N_0)^{L/2}} e^{-\vec{r}_k^H(\varepsilon)\vec{r}_k(\varepsilon)/2 + \sqrt{2\gamma_c}\,\mathrm{Re}\{\vec{r}_k^H(\varepsilon)\}\vec{s}_k(\varepsilon) - \gamma_c \vec{s}_k^T(\varepsilon)\vec{s}_k(\varepsilon)} \\
\vec{r}_k(\varepsilon) = \{r_{i,k}(\varepsilon)\}^{i=0:L-1} \\
\vec{s}_k(\varepsilon) = \{s_{i,k}(\varepsilon)\}^{i=0:L-1} \\
\vec{s}_k(0) = \{\sum_{u=0}^{U-1} c_{i,u} b_{u,k}\}^{i=0:L-1}
\end{gathered} \tag{34}$$

This theoretical development summarizes the existing prior knowledge on classification of CDMA signals that was derived from MPSK classifiers. The next section will discuss the steps necessary for developing a compact form of the average likelihood (35) after computing the expectation over all the unknowns.

$$\lambda(\vec{r}_k|\mathscr{H}) = \mathbb{E}_\varepsilon\{\mathbb{E}_{c_{0,0},c_{0,1},\ldots,c_{L-1,U-1}}\{\mathbb{E}_{b_{0,k},b_{1,k},\ldots,b_{U-1,k}}\{\lambda(\vec{r}_k|\mathscr{H} = \{L,U\}, C, \vec{b}_k, \varepsilon)\}\}\} \tag{35}$$

4 Modulation Classification Assumptions and Procedures

This section discusses the strategy for finding a simplified form of the CDMA likelihood function. The first approach consists on solving the average likelihood for the simplest case of CDMA, i.e., $\mathscr{H} = \{2,2\}$ and then extend the solution to code lengths of 3, 4 and higher code lengths. For a hypothesis $\mathscr{H} = \{2,2\}$, the averaging process will produce $2^{LU+L} = 64$ exponential terms that would require simplification. This research found that a pattern can be obtained with the assistance of a symbolic algebra algorithm shown in Appendix A. Finding a rigorous mathematical proof followed such discovery and made it very simple to develop. The success of this approach was validated using ROC curves. Such curves ensured that the classification of simple CDMA signal can be achieved with the proposed procedure.

4.1 Uniformly Distributed Codes and Data

The first choice for averaging over the unknown codes assumesthat each code coefficient of a spreading matrix C is a uniformly distributed (36) binary antipodal symbol. Each coefficient is assumed to be completely independent from all the other coefficients. A problem with this assumption is the implication that any random code could be used for CDMA, which is a wrong assumption. Although code coefficients appear to be random, a random selection of code coefficients does not necessarily produce a CDMA code with its characteristic low correlation properties.

$$\begin{aligned} P(C) &= \frac{1}{2^{LU}} \\ P(\vec{b}) &= \frac{1}{2^{U}} \end{aligned} \tag{36}$$

Under the assumptions of perfect symbol synchronization ($\varepsilon = 0$), balanced energy ($E = constant$) and full load ($L = U$) CDMA, the conditional likelihood of one symbol is given by (37). If ε is unknown, then the probability over the variable can be assumed to be uniform: $P(\varepsilon) = 1/L$.

$$\begin{aligned} &\lambda(\vec{r}_k|\mathscr{H} = \{L,U\}, C, \vec{b}) = \\ &\frac{1}{(\pi N_0)^{L/2}} exp(-\vec{r}_k(\varepsilon)^H \vec{r}_k(\varepsilon)/2 + \sqrt{2\gamma_c}\,\mathrm{Re}\{\vec{r}_k(\varepsilon)^H\}\vec{s}_k(\varepsilon) - \gamma_c \vec{s}_k^T(\varepsilon)\vec{s}_k(\varepsilon)) \end{aligned} \tag{37}$$

Several key propositions and definition were established for averaging over C and $\vec{b}$. First, we would deal with the expectation over the data vector and average over U unknown users. Second, we would deal with the expectation of the code matrix by splitting the sum over C in three terms that provide simplification.

Proposition 4.1 *The average of a function $f(C\,\vec{b})$ over $C \in \{\pm 1\}^{L\times U}$ and $\vec{b} \in \{\pm 1\}^{U}$ is given by:*

$$\frac{1}{2^{LU}}\sum_{C}\frac{1}{2^{U}}\sum_{\vec{b}} f(C\,\vec{b}) = \frac{1}{2^{LU}}\sum_{C} f(C\,\vec{1}). \tag{38}$$

Proof 4.1 *Defining $q_{i,j} = c_{i,j}b_j$ and substituting in the original equation gives the desired result:*

$$\begin{aligned}\frac{1}{2^{LU}}\sum_{C}\frac{1}{2^{U}}\sum_{\vec{b}} f(C\,\vec{b}) &= \frac{1}{2^{LU}}\sum_{Q}\frac{1}{2^{U}}\sum_{\vec{b}} f(Q\,\vec{1}) \\ &= \frac{1}{2^{LU}}\sum_{Q} f(Q\,\vec{1})\frac{1}{2^{U}}\sum_{\vec{b}} 1 \\ &= \frac{1}{2^{LU}}\sum_{C} f(C\,\vec{1}).\end{aligned} \tag{39}$$

■

The next step requires averaging the likelihood over the code matrix. The summation over all possible combinations of is split in two summation terms. The first one is the summation over a set of matrices $\mathbb{S}_{\vec{a}}$. A second summation is a set of amplitude vectors $\vec{a}$ that have a specific construction. Vectors $\vec{a}$ have non-negative coefficients a_i and can take odd or even values. The vector represents all the possible amplitude levels (40) that can be generated from the product $C\,\vec{b}$.

$$a_i = \begin{cases} 0,2,4,...,U & \text{for } U \text{ even} \\ 1,3,5,...,U & \text{for } U \text{ odd} \end{cases} \tag{40}$$

The space of all possible code matrices will be partitioned into cells defined by the set $\mathbb{S}_{\vec{a}}$.

Definition 4.1 *The set $\mathbb{S}_{\vec{a}}$ is defined in terms of the amplitude vectors $\vec{a}$ as:*

$$\mathbb{S}_{\vec{a}} = \left\{ C \mid C \in \{\pm 1\}^{L\times U} \text{ and } a_i = \left| \sum_{j=0}^{U-1} c_{i,j} \right| \right\} \tag{41}$$

In order to obtain from $\vec{a}$ all the possible values generated by the product $C\,\vec{b}$ it would be necessary to multiply the amplitude vector by a diagonal matrix G that restores the signs as shown in (42).

$$\begin{aligned} G &= diag(\vec{g}) \\ \vec{g} &\in \{\pm 1\}^{L} \end{aligned} \tag{42}$$

The average over C and $\vec{b}$ is replaced by the average over the product of G and $\vec{a}$. A new parameter ρ can be interpreted as the probability of the sign vector $\vec{g}$.

$$\lambda(\vec{r}_k|\mathscr{H}=\{L,U\}) = \frac{1}{2^{LU}}\frac{1}{(\pi N_0)^{L/2}}\sum_{\vec{a}}\rho\sum_{\vec{g}}\sum_{C\in\mathbb{S}_{\vec{a}}} e^{-\vec{r}_k^H\vec{r}_k/2+\sqrt{2\gamma_c}Re\{\vec{r}_k^H\, GC\,\vec{1}\}-\gamma_c\|GC\vec{1}\|^2} \qquad (43)$$
$$\rho = \frac{1}{2^L}$$

The averaging over all $\vec{g}$ is accomplished by using Proposition 2.

Proposition 4.2 *The average of the function $exp(\vec{g}^T\vec{a})$ over $\vec{g}\in\{\pm 1\}$ is given by:*

$$\frac{1}{2^L}\sum_{\vec{g}} e^{\vec{g}^T\,\vec{a}} = \prod_{i=0}^{L-1} cosh(a_i) \qquad (44)$$

Proof 4.2 *This proposition is proven by induction. Consider vectors $\vec{g}_1=[g_0,...,g_{L-1}]^T$, $\vec{a}_1=[a_0,...,a_{L-1}]^T$, $\vec{g}_2=[g_0,...,g_{L-1},g_L]^T$ and $\vec{a}_2=[a_0,...,a_{L-1},a_L]^T$ with $g_i\in\{\pm 1\}$. The proposition should hold true when increasing the dimensionality of $\vec{g}_1$ to $\vec{g}_2$.*

$$\textit{If}\quad \frac{1}{2^L}\sum_{\vec{g}_1} e^{\vec{g}_1^T\,\vec{a}_1} = \prod_{i=0}^{L-1} cosh(a_i)\ \textit{is true,}$$
$$\textit{then}\quad \frac{1}{2^{L+1}}\sum_{\vec{g}_2} e^{\vec{g}_2^T\,\vec{a}_2} = \prod_{i=0}^{L} cosh(a_i)\ \textit{is true.}$$
$$\textit{That is:}$$
$$\begin{aligned}\frac{1}{2^{L+1}}\sum_{\vec{g}_2} e^{\vec{g}_2^T\,\vec{a}_2} &= \frac{1}{2^L}\sum_{\vec{g}_1} e^{\vec{g}_1^T\,\vec{a}_1}\,\frac{1}{2}\sum_{g_L} e^{g_L a_L} \\ &= \frac{1}{2^L}\sum_{\vec{g}_1} e^{\vec{g}_1^T\,\vec{a}} cosh(a_L) \\ &= \prod_{i=0}^{L} cosh(a_i)\end{aligned} \qquad (45)$$

■

The average likelihood in (43) is expressed as a summation of hyperbolic cosine function, similar to the likelihood of MPSK signals.

$$\lambda(\vec{r}_k|\mathscr{H}=\{L,U\}) = \frac{1}{2^{LU}}\frac{1}{(\pi N_0)^{L/2}}\sum_{\vec{a}}\sum_{C\in\mathbb{S}_{\vec{a}}} e^{-\vec{r}_k^H\vec{r}_k/2-\gamma_c\|\vec{a}\|^2}\prod_{i=0}^{L-1} cosh(\sqrt{2\gamma_c}Re\{r_{i,k}^*\}a_i) \qquad (46)$$

Equation (46) was found to be identical to the solution provided by the symbolic algebra algorithm presented in Appendix A by setting a precision parameter $\beta = 0$. The implementation offers reduced performance when classifying between BPSK and 2×2 CDMA. The performance of the classifier in Figure 5) has a weaker performance when compared to a classifier that was developed using a non-uniform probability as shown in Figure 7. The equation (46) is difficult to implement due to the large number of terms generated by all the possible combinations that $\vec{a}$. Each vector coefficient can take $U/2$ values according to (40) and there are L coefficients, which translate to $(U/2)^L$ different amplitude vectors. For a 4×4 CDMA, the equation is extremely complex and the implementation produced numerical problems. A more practical approach can be obtained from weighting the CDMA matrices according to their low correlation properties.

4.2 Averaging over a Non-Uniform Probability of the Code

When considering potential approaches for classifying CDMA matrices, some consideration must be given to the code design. CDMA matrices are designed in a way that they exhibit low cross correlation properties between their column vectors. The Total Squared Correlation [14] defined in (47) is a parameter that measures the correlation between column vectors of the code matrix.

$$\begin{gathered} \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |\vec{c}_i^T \ \vec{c}_j|^2 \geq L^2 \\ \vec{c}_i = \{c_{i,j}\}^{j=0:L-1} \end{gathered} \tag{47}$$

A modification of this metric (48) will be used for constructing a weight referred as the probability of a CDMA code.

$$\begin{gathered} \tau(C) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |\vec{c}_i^T \ \vec{c}_j|^2 - L^2 \\ \tau(C) = \|C^T \ C\|_F^2 - L^2 \geq 0 \end{gathered} \tag{48}$$

Only full rank matrices will be considered when designing a CDMA classifier. If the condition of fully-loaded ($L = U$) is not enforced, then it would be possible to find matrices such that the TSC is minimum; however, such matrices would not qualify as CDMA codes because extending the matrices to a full rank code matrix gives no guarantee of achieving a minimum TSC. Therefore, any classification of underloaded CDMA signals must be constrained to full rank code matrices that are highly uncorrelated.

The weighted average can be seen as a filtering process where some code matrices are emphasized over others. Averaging the likelihood over a uniform probability of codes means that no particular code matrix is preferred over others. This assessment is incorrect because

the likelihood in (46) considers codes that do not exhibit low correlation properties. The idea of using the TSC norm for constructing a probability of code was developed with significant success. Norms such as the Maximum Squared Correlation (MSC) may be used; however, for the purpose of achieving correct classification, the usage of the TSC proved to be sufficient.

A desirable property of the TSC is that it is invariant to permutations of rows and columns. This is easily verified by expressing the TSC as a Frobenius Norm, which is also invariant to permutations and arbitrary rotations. This invariance allows us to remove the energy term of the likelihood out of the summation over the sign vector $\vec{g}$.

4.2.1 Design of Code Matrices using TSC

The design of orthogonal matrices for CDMA is a vast and complex field of study out of the scope of our research goals. A brief overview of the design theory of orthogonal will be provided to provide the idea of how the code matrices were selected in our classification problems. The discussion starts with the definition of Hadamard matrices.

Definition 4.2 *Hadamard matrix H of size $L \times L$ is a matrix of elements $|h_{i,j}| = 1$ that satisfies the following conditions:*

$$H^T H = L I_{L\times L} \tag{49}$$

Hadamard matrices are constrained to $\{\pm 1\}^{L\times L}$ in the generation of CDMA signals using BPSK symbols. There are many way to construct Hadamard matrices, but the most common one is the so called Sylvester's construction.

Definition 4.3 *The Sylvester's construction is given by the following recursive formula using the Kronecker product [15]:*

$$\begin{aligned} H_2 &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ H_{2^k} &= H_2 \otimes H_{2^{k-1}} \end{aligned} \tag{50}$$

Hadamard matrices have been a subject of study of mathematicians for years. Several code constructions such as Paley and Williamson constructions exist at the present time. An interesting conjecture states that Hadamard matrices can be designed for code lengths divisible by 4. Unfortunately, there is no theorem that can prove the existence for any arbitrary code length divisible by four [16].

Conjecture 4.1 *A real Hadamard matrix with must be of code length (order) one, two or a multiple of four. [17]*

The following theorem is a modification of the relationship between the Frobenius Norm and the Kronecker product [15].

Theorem 4.3 *The Total Squared Correlation of a matrix generated from a Kronecker product is given by: [15]*

$$\tau(H_a \otimes H_b) = \tau(H_a)\tau(H_b) + L_a^2\tau(H_b) + L_b^2\tau(H_a) \tag{51}$$

Equation (51) leads to the following corollary.

Corollary 4.4 *If the TSC of two matrices are zero, then the TSC of the Kronecker product must also be zero.*

4.2.2 Probability of the Code Matrix

The proposed probability of a code matrix is a function of the TSC such that the weight for small values of TSC are high while the weight for high TSC values are low. The first choice is to construct the exponentially decaying probability function shown in (52).

$$\begin{aligned} P(C) &= \frac{1}{W} w(C) e^{-\beta\tau(C)} \\ W &= \sum_C w(C) e^{-\beta\tau(C)} \\ w(C) &= \begin{cases} 1 & \text{for } \tau(C) \le \tau_{min} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{52}$$

This probability resembles a sampled version of an exponential probability function with a precision parameter β that controls the selection of code matrices. This definition seems to be a unique contribution of this effort, because this concept does not appear the after a review of the technical literature. Other probabilities can be derived using similar concepts, but the interest in this particular one is the preservation of the exponential nature of the entire likelihood function. The probability can be tailored for a limited set of TSC values least or equal than $\tau(C) < \tau_{max}$. We assume that $\tau_{max} \to \infty$.

4.2.3 Expectation over a Non-Uniform Probability of Code

The development of the likelihood under the assumption of non-uniform probability of code starts with Proposition 4.5 and results in (53) assuming that a frame synchronous signal is provided.

$$\lambda(\vec{r}_k | \mathscr{H} = \{L,U\}) = \frac{1}{(\pi N_0)^{L/2}} \sum_{\vec{a}} \sum_{\tau} \rho \sum_{\vec{g}} \sum_{C \in \mathbb{S}_{\vec{a},\tau}} P(C) e^{-\vec{r}_k^H \vec{r}_k / 2 + \sqrt{2\gamma_c}\, Re\{\vec{r}_k^H\, GC\, \vec{1}\} - \gamma_c \| GC\vec{1} \|^2} \tag{53}$$

The space of all possible code matrices will be partitioned in non-overlapping sets similar to the methodology used for uniformly distributed codes. These sets contain matrices such that the absolute value of the sum of the columns equals some vector $\vec{a}$ and a code matrix in the set has a TSC value τ according to 54.

Definition 4.4 *The set $\mathbb{S}_{\vec{a},\tau}$ is defined as:*

$$\mathbb{S}_{\vec{a},\tau} = \left\{ C \mid C \in \{\pm 1\}^{L\times U}, a_i = \left| \sum_{j=0}^{U-1} c_{i,j} \right| \text{ and } \tau(C) = \tau \right\} \tag{54}$$

The summation over the code coefficients is split into three summation terms, the summation over $\vec{g}$, $C \in \mathbb{S}_{\vec{a},\tau}$, and τ as shown in (53). Then, we proceed to rearrange the summations as follows:

$$\lambda(\vec{r}_k|\mathscr{H} = \{L,U\}) = \frac{e^{-\vec{r}_k^H \vec{r}_k/2}}{(\pi N_0)^{L/2}} \sum_{\vec{a}} e^{-\gamma_c \|\vec{a}\|^2} \sum_{\tau} \sum_{C\in\mathbb{S}_{\vec{a},\tau}} \frac{e^{-\beta\tau(C)}}{W} \rho \sum_{\vec{g}} e^{\sqrt{2\gamma_c}\, Re\{\vec{r}_k^H\, G\, \vec{a}\}}. \tag{55}$$

Averaging over G is performed by swapping the diagonal terms in the matrix with the observation vector $\vec{r}$ in the correlation term.

$$\vec{r}^T\, G\, \vec{a} = \vec{g}^T\, diag(\vec{r})\, \vec{a} \tag{56}$$

Proposition 4.2 can be applied to obtain the average likelihood function in the form of products of hyperbolic cosine functions given by (57).

$$\lambda(\vec{r}_k|\mathscr{H} = \{L,U\}) = \frac{e^{-\vec{r}_k^H \vec{r}_k/2}}{(\pi N_0)^{L/2}} \sum_{\vec{a}} e^{-\gamma_c \|\vec{a}\|^2} \alpha(\vec{a},\beta) \prod_{i=0}^{L-1} cosh(\sqrt{2\gamma_c}\, Re\{r_{i,k}^*\}\, a_i) \tag{57}$$

$$\alpha(\vec{a},\beta) = \sum_{\tau} \sum_{C\in\mathbb{S}_{\vec{a},\tau}} \frac{e^{-\beta\tau(C)}}{W} \tag{58}$$

This equation can be divided in four terms: the energy of the received signal, the energy of the model, a term that depends on the code and a product of hyperbolic cosine functions that depends on the correlation between the received signal and the model.

4.2.4 Analytical Average Likelihood

The empirical likelihood obtained from the symbolic-algebra algorithm is consistent with the derived likelihood (57) and (58) for code lengths between 1 through 4. The generality of the development give no reason to suspect that this average likelihood formula would not hold for higher code lengths.

4.2.4.1 BPSK versus 2×2 CDMA

The likelihood (57) of a single symbol interval with $\mathscr{H} = \{L, U\}$ provides an exact solution (60) applicable to both: uniformly distributed code $\beta = 0$ and the non-uniform probability of the code $\beta \neq 0$. It also provides the formula for BPSK signals by just setting $L = 1$. A binary classification between BPSK and 2×2 CDMA requires comparing pairs independent BPSK symbols against single 2×2 CDMA symbols. We obtain the formula for two BPSK independent symbols.

$$\lambda(\{r_k\}^2 | \mathscr{H} = \{1,1\}) = e^{-\gamma_c}\, cosh(\sqrt{2\gamma_c}\, Re\{r_0\}) \cdot e^{-\gamma_c}\, cosh(\sqrt{2\gamma_c}\, Re\{r_1\}) \tag{59}$$

The exact likelihood of a 2×2 CDMA provides means to explore the effect of the precision parameter on the overall likelihood function.

$$\begin{aligned} \lambda(\vec{r}_k | \mathscr{H} = \{2,2\}) = \frac{1}{2(1+e^{-8\beta})}(e^{-8\beta} + e^{-4\gamma_c}\, cosh(2\sqrt{2\gamma_c}\, Re\{r_{0,k}\}) + \\ e^{-4\gamma_c}\, cosh(2\sqrt{2\gamma_c}\, Re\{r_{1,k}\}) + e^{-8\beta-4\gamma_c}\, cosh(2\sqrt{2\gamma_c}\, Re\{r_{0,k}\}) cosh(2\sqrt{2\gamma_c}\, Re\{r_{1,k}\})) \end{aligned} \tag{60}$$

Figures 5 through 7 show the Receiver Operating Characteristic (ROC) Curves for a LRT for 2×2 CDMA vs. BPSK. These curves were based on 2000 tests, each test uses a sequence of 128 chips, and plotted for different total-SNR (total energy over N_0). Each graph is for a single value of β. The best performance is obtained when $\beta \to \infty$. Due to the exponential probability (52) the coefficients α decays quickly for low TSC values.

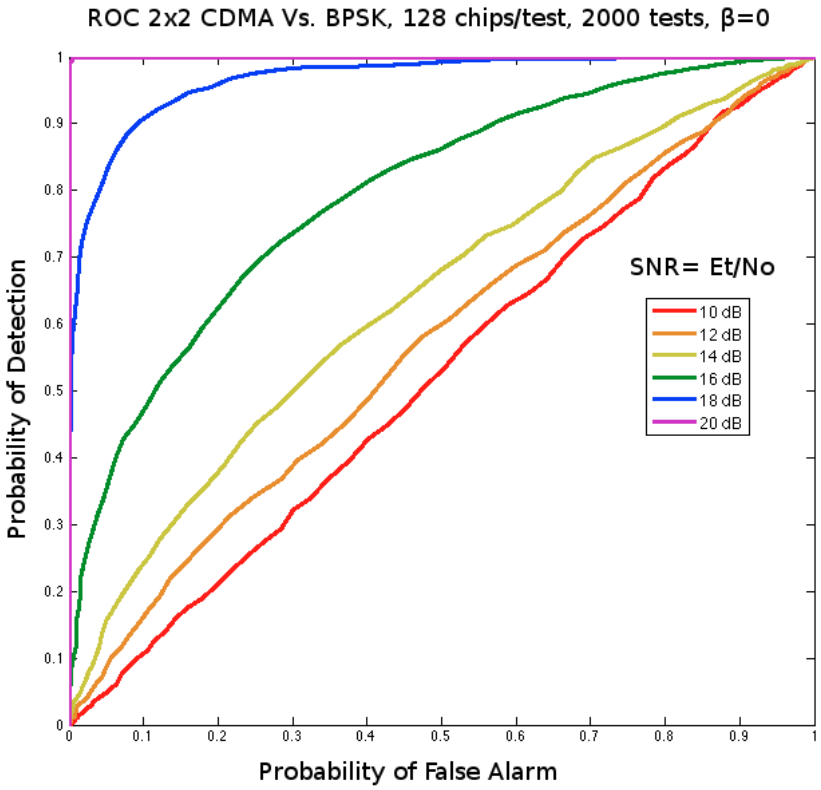

Fig. 5 ROC $\mathscr{H}_1 = \{2,2\}$ versus $\mathscr{H}_0 = \{1,1\}$, $\beta = 0$

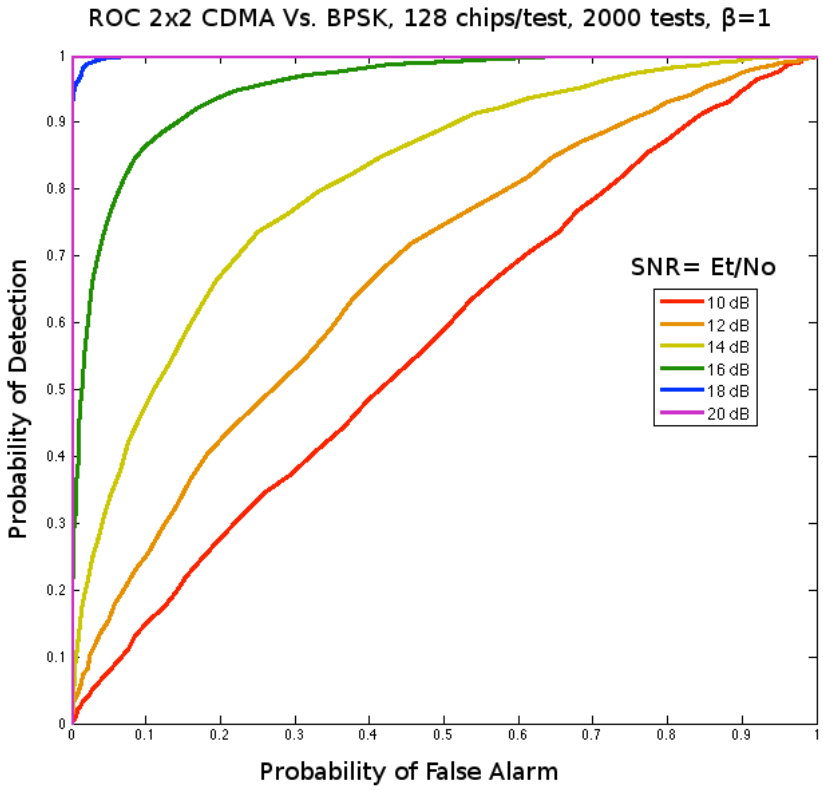

Fig. 6 ROC $\mathscr{H}_2 = \{2,2\}$ versus $\mathscr{H}_1 = \{1,1\}$, $\beta = 1$

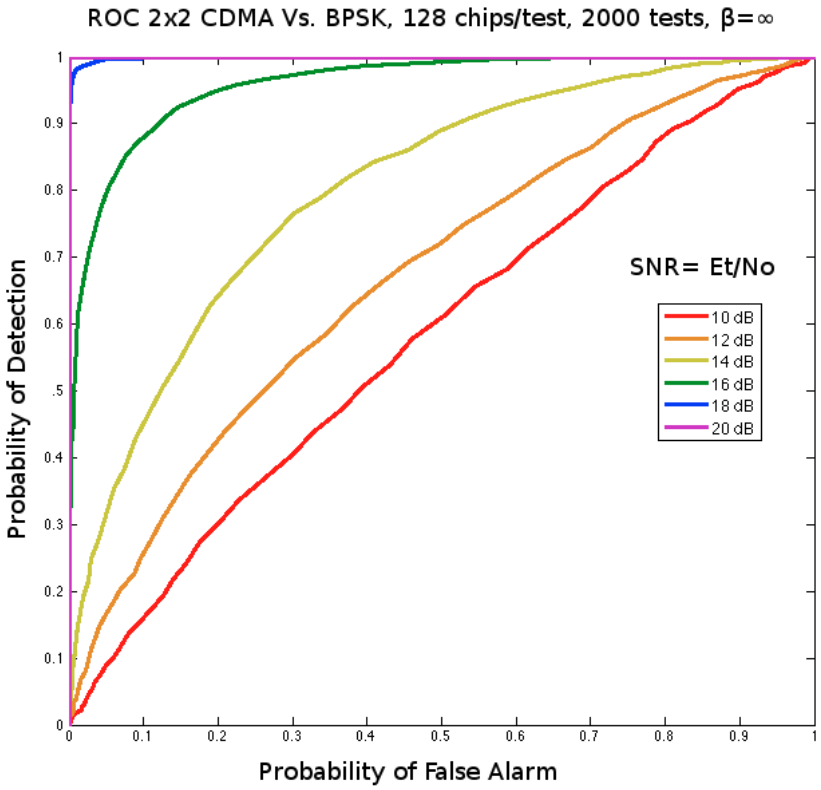

Fig. 7 ROC $\mathscr{H}_1 = \{2,2\}$ versus $\mathscr{H}_1 = \{1,1\}$, $\beta = \infty$

4.3 Simplification of the Average Likelihood

The likelihood of a CDMA can be further decomposed in two terms: those that quickly decay to zero and those that persist when $\beta \to \infty$. The dependency of each $\alpha(\vec{a},\beta)$ coefficient for a 2×2 CDMA is shown in Figure 8 and 9. In this specific case, $\alpha([0,0]^T,\beta)$ and $\alpha([2,2]^T,\beta)$ converge to zero while $\alpha([0,2]^T,\beta)$ and $\alpha([2,0]^T,\beta)$ converge to $1/2$.

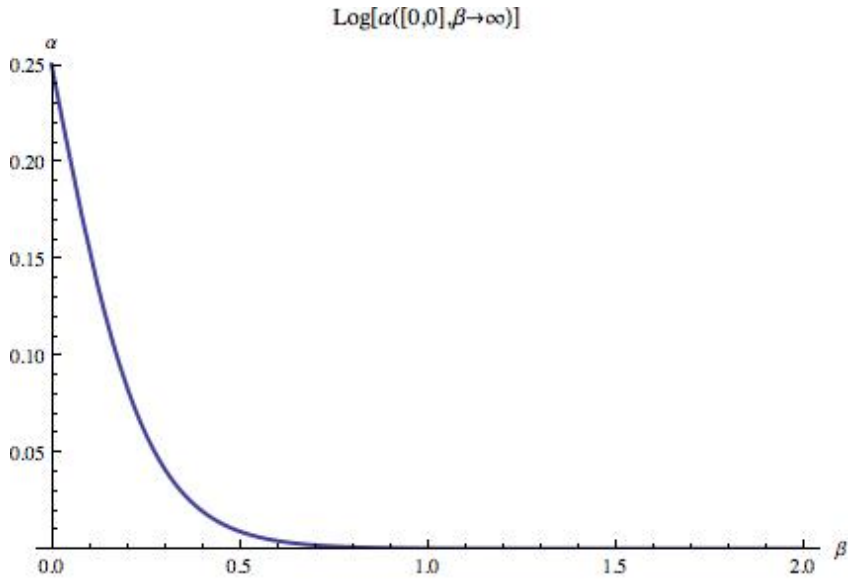

Fig. 8 2×2 CDMA Vanishing Coefficients α versus β, $\vec{a}\in\{[0,0]^T,[2,2]^T\}$

The computation of these coefficients become extremely difficult even for simple cases like a 2×2 CDMA. A logical approach must consider the limit as $\beta\to\infty$ using the definition of α provided in 58.

Definition 4.5 *A feature vector $\vec{a}^\star=\{a_i^\star\}$ for a given hypothesis $\mathscr{H}=\{L,U\}$is defined in terms of minimum TSC matrices $C^\star$ such that:*

$$\begin{aligned}\tau(C)\geq\tau(C^\star)&=\tau_{min}\\ a_i^\star&=|\sum_{j=0}^{U-1}c_{i,j}^\star|.\end{aligned}\tag{61}$$

Proposition 4.5 *The limit of $\alpha(\vec{a},\beta)$ as β approaches infinity depends on the feature vectors as follows:*

$$\lim_{\beta\to\infty}\alpha(\vec{a},\beta)=\begin{cases}0 & for\ \vec{a}\neq\vec{a}^\star\\ \dfrac{|\mathbb{S}_{\vec{a},\tau_{min}}|}{\sum_{\vec{a}\in\{\vec{a}^\star\}}|\mathbb{S}_{\vec{a},\tau_{min}}|} & for\ \vec{a}=\vec{a}^\star\end{cases}\tag{62}$$

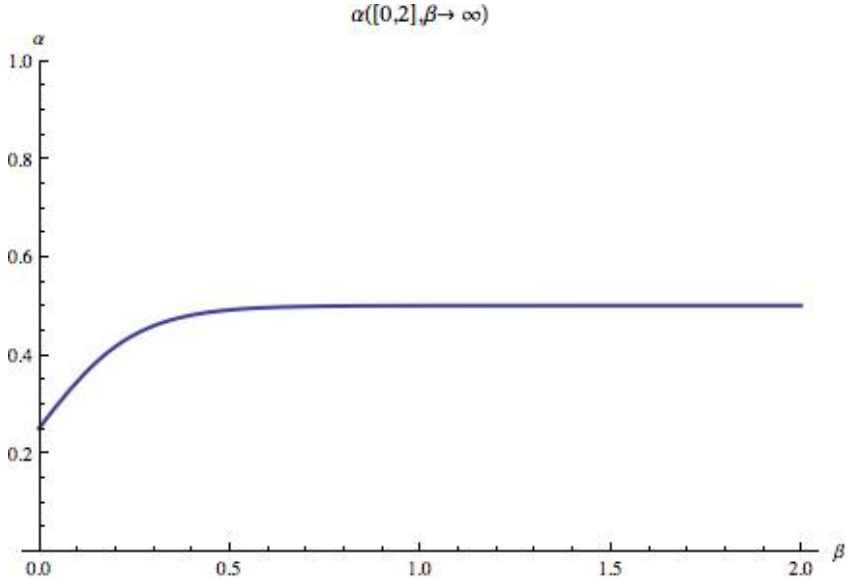

Fig. 9 2×2 CDMA Non-Vanishing Coefficients α versus β, $\vec{a} \in \{[0,2]^T, [2,0]^T\}$

Proof 4.3 *From a set of all possible vectors* $\{\vec{a}_j\}$ *and all possible TSC values* $\{\tau_i\}$ *with* $0 \leq \tau_0 < \tau_i < \tau_{i+1}$ *for* $\mathscr{H} = \{L,U\}$*, we define* $d_{i,j} = |\mathbb{S}_{\vec{a}_j,\tau_i}|$ *and* $d_i = \sum_j d_{i,j}$*. The coefficients can be expressed as a ratio of exponential functions.*

$$\lim_{\beta\to\infty} \alpha(\vec{a},\beta) = \frac{d_{i,k}e^{-\beta\tau_k}}{\sum_i d_i e^{-\beta\tau_i}} \tag{63}$$

Then, we multiply by $e^{\beta\tau_0}$ *on both the numerator and denominator. Any term with* $\tau_k > \tau_0$ *vanishes as* $\beta \to \infty$ *leaving only the terms where* $\tau_k = \tau_0$*.*

$$\lim_{\beta\to\infty} \alpha(\vec{a},\beta) = \frac{d_{i,k}e^{-\beta(\tau_k-\tau_0)}}{\sum_i d_i e^{-\beta(\tau_i-\tau_0)}} \tag{64}$$

Therefore, the coefficients depend on the feature vector set $\{\vec{a}^\star\}$*.*

$$\lim_{\beta\to\infty} \alpha(\vec{a},\beta) = \begin{cases} 0 & \text{for } \vec{a} \neq \vec{a}^\star \\ \dfrac{d_{0,j}}{d_0} & \text{for } \vec{a} = \vec{a}^\star \end{cases} \tag{65}$$

■

By using an infinite precision, many terms vanishes from the average likelihood function leaving only the terms that are contributions of the lowest TSC code matrix and allowing to perform the classification in a computationally efficient way. The question remains whether

computing the likelihood coefficients would be affordable after this simplification. The Matlab algorithm shown in Appendix B was implemented for computing for small code lengths. Unfortunately, the code becomes useless for processing code lengths greater than 4. A computationally efficient approach for estimating the coefficients can be constructed from interpreting the meaning of $\alpha(\vec{a},\infty)$.

4.3.1 Interpretation of the CDMA Likelihood Coefficients

The values of $\alpha(\vec{a},\beta)$ can be calculated exactly for code lengths of 2 as follows. First, the feature vectors $\vec{a}$ are identified.

$$\begin{aligned} \vec{a}_1^\star &= [0,2]^T \\ \vec{a}_2^\star &= [2,0]^T \end{aligned} \tag{66}$$

The code matrices of interest in the sets $\mathbb{S}_{\vec{a},\tau}$ are:

$$\begin{aligned} \mathbb{S}_{\vec{a}_1^\star,\tau} = \{C_1^\star\} \;\; \mathbb{S}_{\vec{a}_2^\star,\tau} = \{C_2^\star\} \\ C_1^\star = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \;\; C_2^\star = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \end{aligned} \tag{67}$$

With each set of cardinality 1, it is easy to calculate the values of the likelihood coefficients as follows:

$$\begin{aligned} \alpha(\vec{a}_1^\star,\infty) &= \frac{|\mathbb{S}_{\vec{a}_1^\star,0}|}{|\mathbb{S}_{\vec{a}_1^\star,0}| + |\mathbb{S}_{\vec{a}_2^\star,0}|} = \frac{1}{2} \\ \alpha(\vec{a}_2^\star,\infty) &= \frac{|\mathbb{S}_{\vec{a}_2^\star,0}|}{|\mathbb{S}_{\vec{a}_1^\star,0}| + |\mathbb{S}_{\vec{a}_2^\star,0}|} = \frac{1}{2} \end{aligned} \tag{68}$$

A closer look at (68) reveals that the likelihood coefficients α are a ratio of the occurrence of particular CDMA vectors $abs(C\,\vec{b})$ over the total number of feature vectors that exists for given hypothesis $\mathscr{H} = \{L,U\}$. A more practical way of calculating the coefficients comes from (69).

$$\alpha(\vec{a}^\star,\infty) \approx \sum_{\vec{b}} \frac{count(abs(C^\star\,\vec{b}) = \vec{a}^\star)}{2^U} \tag{69}$$

This limit is an approximation because for a given code length, there may exist several orthogonal matrices $C^\star$ that are not related by a permutation of columns and rows. These code matrices are referred as non-equivalent matrices [18]. Unfortunately, despite the numerous code constructions that exist in the literature, there is no specific theorem that predicts the number of non-equivalent matrices for a specific code length. It is known that for code lengths of $L = 4,8,16,32,64$ and 128 there are $1,1,5,3,60$ and 487 non-equivalent matrices respectively [19]. This limits the construction of the average likelihood to our knowledge on codes. It was found that extracting the feature vectors of each of the five 16×16 CDMA non-equivalent

matrices resulted in the same values of $\vec{a}^\star$ with slightly different coefficients values. An assumption that non-equivalent Hadamard matrices generate the same set of feature vector will be use in our classification problems. The algorithm for generating the likelihood coefficients and the feature vectors is provided in Appendix D.

The following procedure summarizes the construction of the simplified likelihood function given a hypothesis $\{L,U\}$. The procedure is applied to find the likelihood function of a 4×4 CDMA.

1. Find $C^\star$, a code matrix with minimum TSC;
2. Extract the feature vector from $C^\star\,\vec{b}$ by varying $\vec{b}$;
3. Compute the likelihood coefficients using (69);
4. Using the coefficients of the feature vector, construct the product of hyperbolic cosines in (57).

The 4×4 Hadamard matrix is given by the Sylvester's Construction (50). There are two equivalent matrices of interest that produce two important feature vectors. All other feature vectors can be derived from the permutation of these vectors.

$$C_1^\star = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad C_2^\star = \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \tag{70}$$

The 4×4 case has two feature vectors which can be obtained from adding the columns of the previous matrices. The likelihood coefficients do not change when a $\vec{a}^\star$ is permuted. This invariance of α is due to the exchangeability property [20] of the model that makes the likelihood invariant to row permutations. Because the transmission model admits CDMA code matrix C with any arbitrary permutation of rows, the final likelihood does not depend on the order of the vector coefficients of $\vec{r}$.

The cardinality of the sets is computed as follows. For $C_1^\star$, the rows that add up to zero can be permuted in 3! ways. Also there are 2^4 ways of changing the sign. The value $|\mathbb{S}_{\vec{a}_1^\star,0}|$ is $3!\cdot 2^4 = 96$. For $C_2^\star$, the rows that add up to 2 can be permuted in 4! ways. Also there are 2^4 ways of changing the sign. The value $|\mathbb{S}_{\vec{a}_2^\star,0}|$ is $4!\cdot 2^4 = 384$. At the time of computing the coefficients α, one needs to account for 4 possible permutations the $\vec{a}_1^\star$. The total number of Hadamard matrices for a 4×4 code matrix is the sum of the cardinalities of these sets, that is

768. This result is in agreement with the experimental computations.

$$\begin{aligned}
\vec{a}_1^{\star} &= [4,0,0,0]^T \\
\vec{a}_2^{\star} &= [1,1,1,1]^T \\
\alpha(\vec{a}_1^{\star},\infty) &= \frac{|\mathbb{S}_{\vec{a}_1^{\star},0}|}{4|\mathbb{S}_{\vec{a}_1^{\star},0}|+|\mathbb{S}_{\vec{a}_2^{\star},0}|} = \frac{1}{8} \\
\alpha(\vec{a}_2^{\star},\infty) &= \frac{|\mathbb{S}_{\vec{a}_2^{\star},0}|}{4|\mathbb{S}_{\vec{a}_1^{\star},0}|+|\mathbb{S}_{\vec{a}_2^{\star},0}|} = \frac{1}{2}
\end{aligned} \tag{71}$$

Finally, the 4×4 average likelihood for a single symbol is constructed by forming the hyperbolic cosine products.

$$\begin{aligned}
\lambda(\vec{r}_k|\mathscr{H}=\{4,4\}) = \frac{e^{-\langle \vec{r}_k,\vec{r}_k\rangle/2}}{(\pi N_0)^{L/2}} \Bigg(&\sum_{i=0}^{3} \frac{1}{8} e^{-4^2\gamma_c} cosh(4\sqrt{2\gamma_c}\,\mathrm{Re}\{r_{i,k}\}) \\
&+\frac{1}{2}\prod_{i=0}^{3} e^{-4^2\gamma_c} cosh(2\sqrt{2\gamma_c}\,\mathrm{Re}\{r_{i,k}\}) \Bigg)
\end{aligned} \tag{72}$$

As the code length and number of user grow, the average likelihood becomes a complicated expression and further simplification is needed to avoid numerical errors or any excessive computational burden.

4.4 Discussion and Results

Figure 10 through 11 show ROC curves for several simple cases of spreading codes. The length of the signal under test is expressed in number of chips. A test sample of 128 and 192 chips was chosen for code lengths multiple of 2 or 3 respectively. The plots are based on 2000 tests. The total SNR is the total energy E_T divided by the noise power N_0. Few observations can be made from these simulations. The conclusion is that the average likelihood classifier for CDMA is capable of classifying small code lengths. Figure 14 shows that it is possible to distinguish between BPSK and CDMA with no surprise because the proposed development follows a standard procedure that can be extended to other hypothesis such as MPSK or QAM modulation.

4.4.1 Average Likelihood for the Underloaded Case

The underloaded CDMA model assumes that the data vector $\vec{b}$ is given by (73). The equation introduces new variables Δ_i that represent whether the user i is active (1) or inactive (0). Defining a scenario of $U < L$ users can be expressed in $L!/((L-U)!U!)$ different permutations, but all

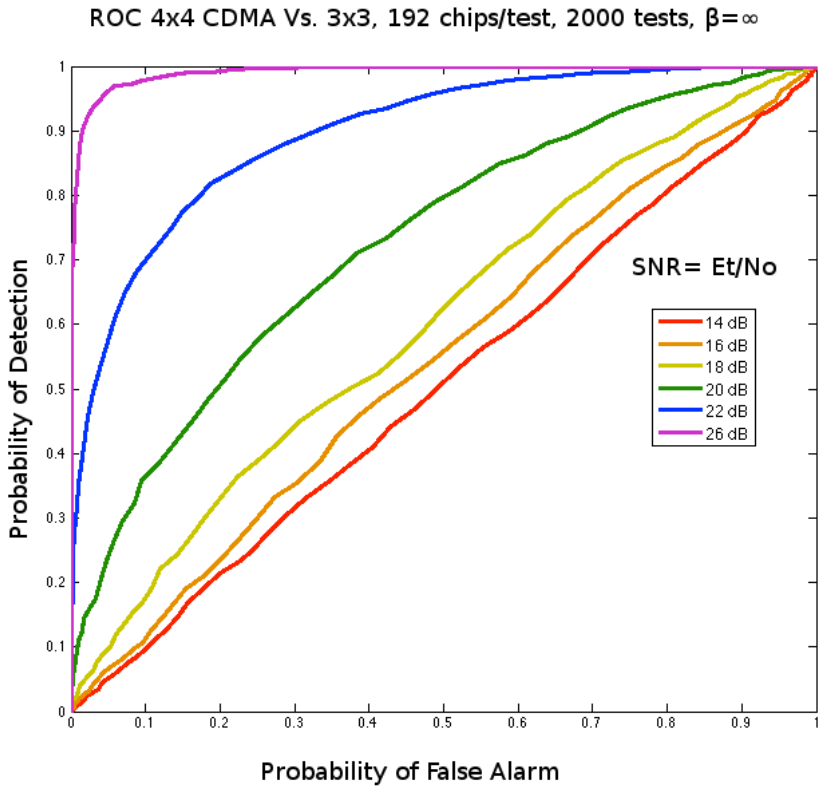

Fig. 10 ROC Curve for $\mathscr{H} = \{4,4\}$ vs. $\{3,3\}$ CDMA

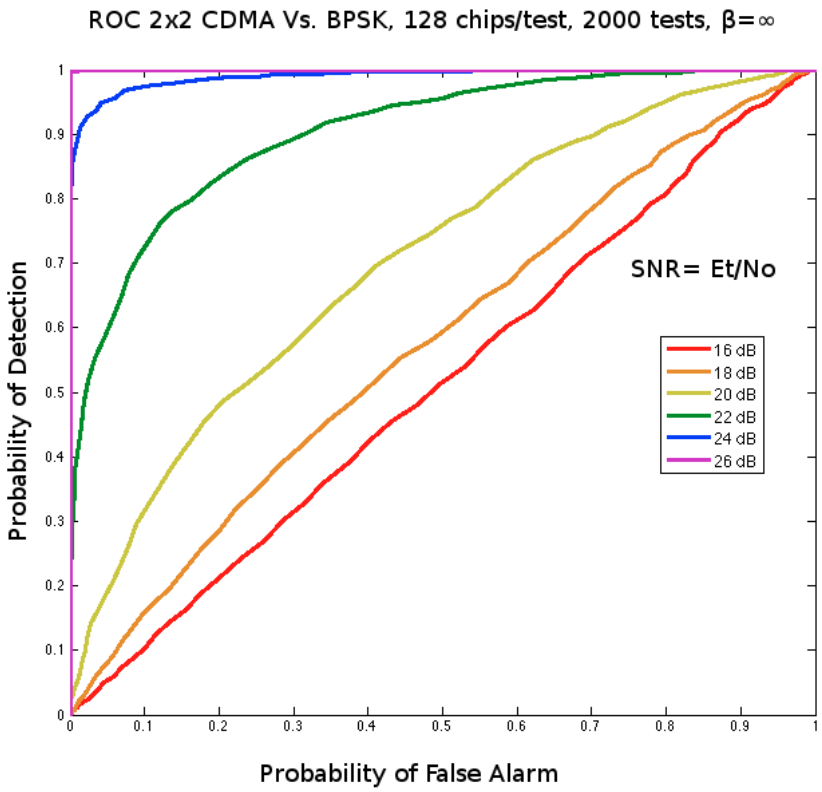

Fig. 11 ROC Curve for $\mathscr{H} = \{2,2\}$ CDMA vs. BPSK

these results are equivalent if one introduces an arbitrary permutation matrix $\mathscr{P}$ between Q and $\vec{b}$ which leaves the same model if $Q\mathscr{P}$ redefined as Q'. It is not necessary to average over all possible combinations of Δ_i, because its average will be equivalent to the likelihood of any arbitrary column permutation.

$$\begin{gathered}\vec{b} = [b_0\Delta_0, b_1\Delta_1, ..., b_L\Delta_L]^T \\ \Delta_i \in \{0,1\}\end{gathered} \tag{73}$$

From the point of view of (57) the form of the average likelihood should remain the same, except for the definitions of the feature vectors (69). They must be calculated assuming that $L-U$ arbitrary users are inactive.

An interesting case of underloaded CDMA is shown in Figure 12 where a 4×2 CDMA fails to be differentiated from a 2×2 CDMA. The behavior of the classifier is explained by understanding the Sylvester's Construction that was use to generate the code. A 4×2 CDMA can be seen as two appended CDMA vectors from a 2×2 matrix. Other underloaded cases are shown in Figures 13 and 14. Both figures validates the possibility of classifying underloaded CDMA for small code lengths.

Another interesting construction is the case of $L\times 1$ CDMA. The $L\times 1$ CDMA likelihood function reduces to the likelihood of L independent BPSK symbols; therefore, it will be impossible to differentiate between those two classes given the provided CDMA model. In such cases, the classification would have to rely on the cyclostationary concepts.

The next section provides a simplification of the likelihood function as well as an extension to much larger code length values.

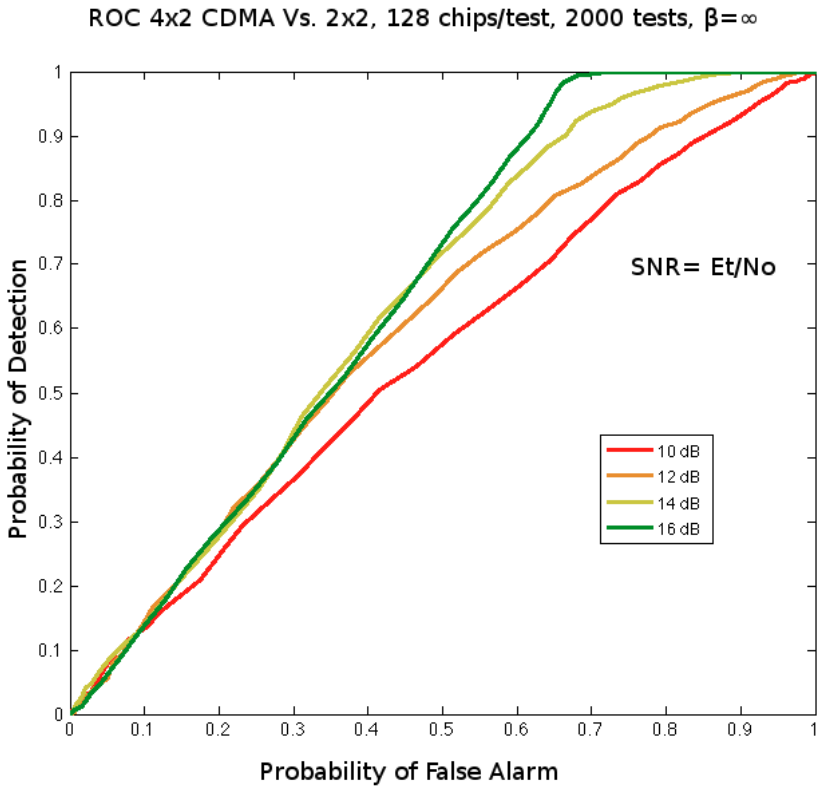

Fig. 12 ROC Curve for $\mathscr{H} = \{4,2\}$ vs. $\{2,2\}$ CDMA

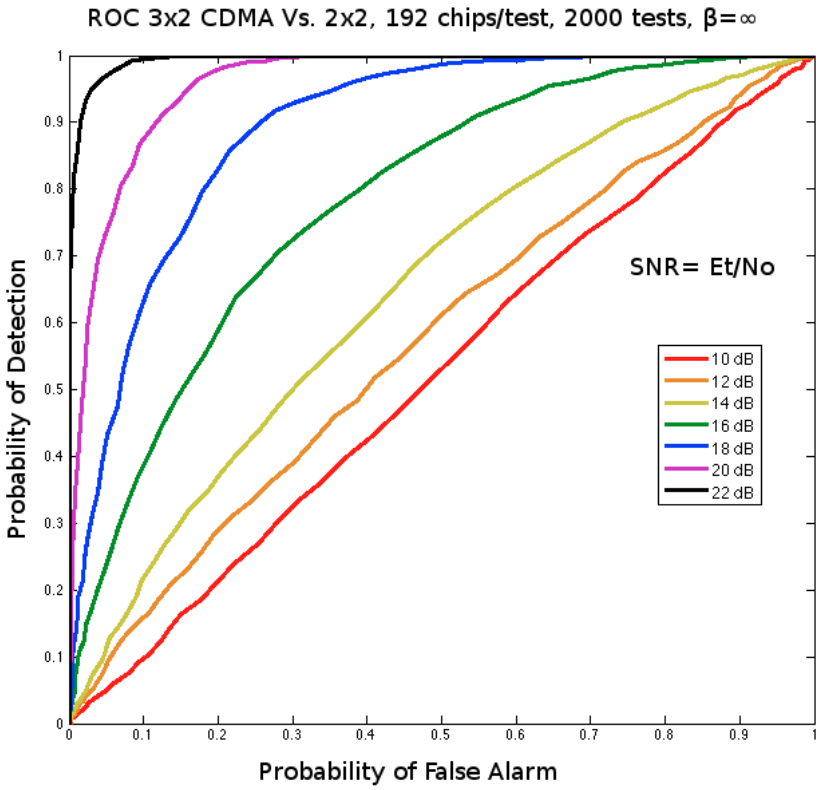

Fig. 13 ROC Curve for $\mathscr{H} = \{3,2\}$ vs. $\{2,2\}$ CDMA

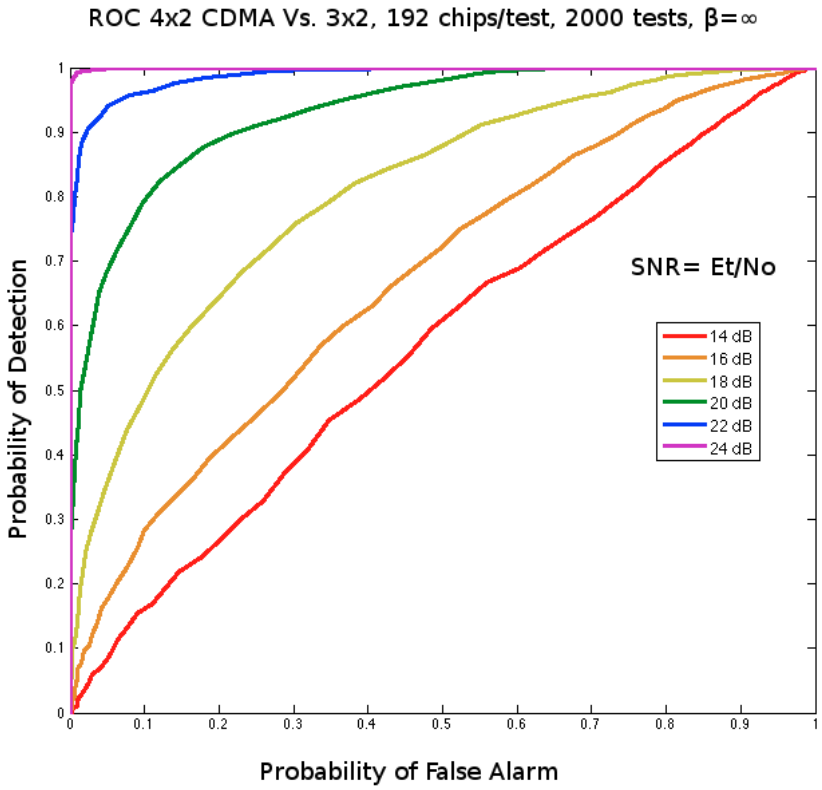

Fig. 14 ROC Curve for $\mathscr{H} = \{4,2\}$ vs. $\{3,2\}$ CDMA

5 Method Simplification, Assumptions and Procedures

The goal of this section is to approximate of the average likelihood function developed in the previous section. The approximation is needed for processing high values of code length that could quickly produce numerical problems and computational inefficiencies due to the large number of variables. The approximation is achieved by either reducing the set of the feature vectors, using Taylor series and taking advantage of the properties of the likelihood function which is an even and exchangeable function.

5.1 Revisiting the CDMA Model

The CDMA model will be expressed in terms of the matrix equation shown in (74). For simplicity, we will assume that all the variable Q, $\vec{g}$ and $\vec{b}$ are real and binary antipodal. The system is assumed to be balanced, with energy per symbol E. The code matrix is an $L \times U$ matrix. The noise $\vec{n}_k$ is now a real Gaussian random vector with zero mean and correlation matrix $N_0/2\, I$. The observation $\vec{r}_k = \vec{y}_k / \sqrt{N_0/2}$ is also defined as a real vector.

$$\begin{aligned} \vec{y}_k &= \sqrt{\frac{E}{L}} G\, Q\, \vec{b}_k + \vec{n}_k \\ G &= diag(\vec{g}) \end{aligned} \tag{74}$$

These variables are averaged using the following weights for $\vec{g}$, $\vec{b}$ and Q.

$$\begin{aligned} \rho &= \frac{1}{2^L} \\ P(\vec{b}) &= \frac{1}{2^L}, \\ P(Q) &= \frac{e^{-\beta\,\tau(Q)}}{\sum_C e^{-\beta\,\tau(C)}} \end{aligned} \tag{75}$$

Using the key propositions and definitions discussed in section 4, we find that our likelihood model is given by equations (76) through (78).

$$\lambda(\vec{r}_k | \mathscr{H} = \{L,U\}) = \frac{e^{-\vec{r}_k^H \vec{r}_k/2}}{(\pi\, N_0)^{L/2}} \sum_{\vec{a}} e^{-\gamma_c \|\vec{a}\|^2} \alpha(\vec{a},\beta) \prod_{i=0}^{L-1} cosh(\sqrt{2\gamma_c}\, Re\{r_{i,k}\}\, a_i) \tag{76}$$

$$\alpha(\vec{a},\beta) = \sum_{\tau} \sum_{C \in \mathbb{S}_{\vec{a},\tau}} \frac{e^{-\beta\,\tau(C)}}{W} \tag{77}$$

$$\lim_{\beta\to\infty}\alpha(\vec{a},\beta)=\begin{cases}0 & \text{for } \vec{a}\neq\vec{a}^{\star}\\ \dfrac{|\mathbb{S}_{\vec{a},\tau_{min}}|}{\sum_{\vec{a}\in\{\vec{a}^{\star}\}}|\mathbb{S}_{\vec{a},\tau_{min}}|} & \text{for } \vec{a}=\vec{a}^{\star}\end{cases} \tag{78}$$

5.2 Classification Using a Single Feature Vector

A distinctive characteristic of the Sylvester's construction (79) is having a row and column of ones. A feature vector derived from the product $C\,\vec{1}$ (79) provides a simple way to classify CDMA signals because it reduces the complexity of equation (76).

$$\begin{aligned}\vec{a}^{\star}&=[L,0,0,...0]^{T}\\ \alpha(\vec{a}^{\star},\infty)&=p_{\vec{a}^{\star}}\end{aligned} \tag{79}$$

The average likelihood function of such scheme reduces to the following expression:

$$\lambda(\vec{r}_k|\mathscr{H}=\{L,U\})=\frac{e^{-\vec{r}_k^{H}\vec{r}_k/2}}{(\pi\,N_0)^{L/2}}\,e^{-\gamma_c L^2}\,p_{\vec{a}^{\star}}\sum_{i=0}^{L-1}cosh(L\,\sqrt{2\gamma_c}\,r^{*}_{i,k}) \tag{80}$$

The construction of the likelihood ratio between hypotheses $\mathscr{H}_1=\{2L,2L\}$ and $\mathscr{H}_2=\{L,L\}$ is a simple ratio of hyperbolic cosine functions. The terms containing the energy of the vectors and N_0 are cancelled out. The formula requires to give a special attention should be given to the format of the observation as either $2L\times 2L$ or $L\times L$ CDMA.

$$\frac{\lambda(\vec{r}_k|2L)}{\lambda(\vec{r}'_k,\vec{r}'_{k+1}|L)}=\frac{\sum_{i=0}^{2L-1}cosh(\sqrt{2\gamma_{c1}}\,2\,L\,r_{i,k})}{\sum_{i=0}^{L-1}cosh(\sqrt{2\gamma_{c0}}\,L\,r'_{i,k})\sum_{i=0}^{L-1}cosh(\sqrt{2\gamma_{c0}}\,L\,r'_{i,k+1})}\underset{H_0}{\overset{H_1}{\gtrless}}\eta \tag{81}$$

Figure 15 through 17 show the ROC curves for detecting CDMA of various code lengths using the provided simplification. The plots are based on 2000 tests, with 128 chips/test, using frame asynchronous symbols and plotted for several Total-SNR values. For small code lengths, the feature vector (79) appears to be relevant in the classification of classes, but as the code length increases, the performance of the classifier is degraded under the same amount of samples. This degradation occurs presumably for two reasons: 1. classifying higher code lengths at the same total-SNR requires more samples and 2. This feature (79) becomes less significant as the code length increases.

The CDMA classifier represented by the rule in (81) can be interpreted as follows. Within a single frame of L chips, the classifier must find a single spike of amplitude proportional to L and probability $p_{\vec{a}^{\star}}$. This rule is only valid under the assumption of fully-loaded, small code lengths CDMA.

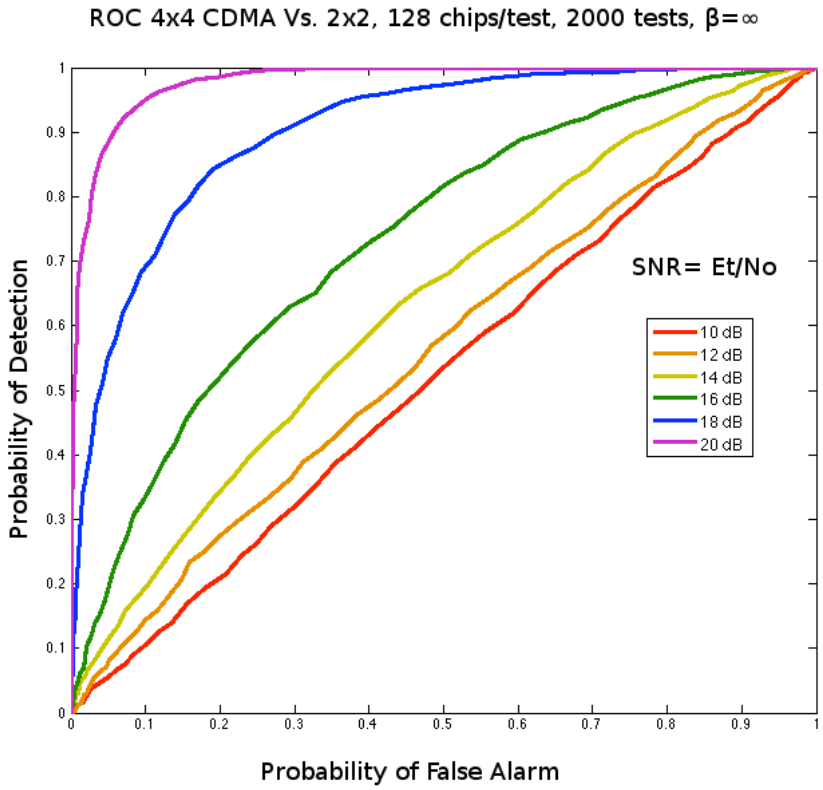

Fig. 15 Single Feature Detection: ROC $\{4,4\}$ versus $\{2,2\}$

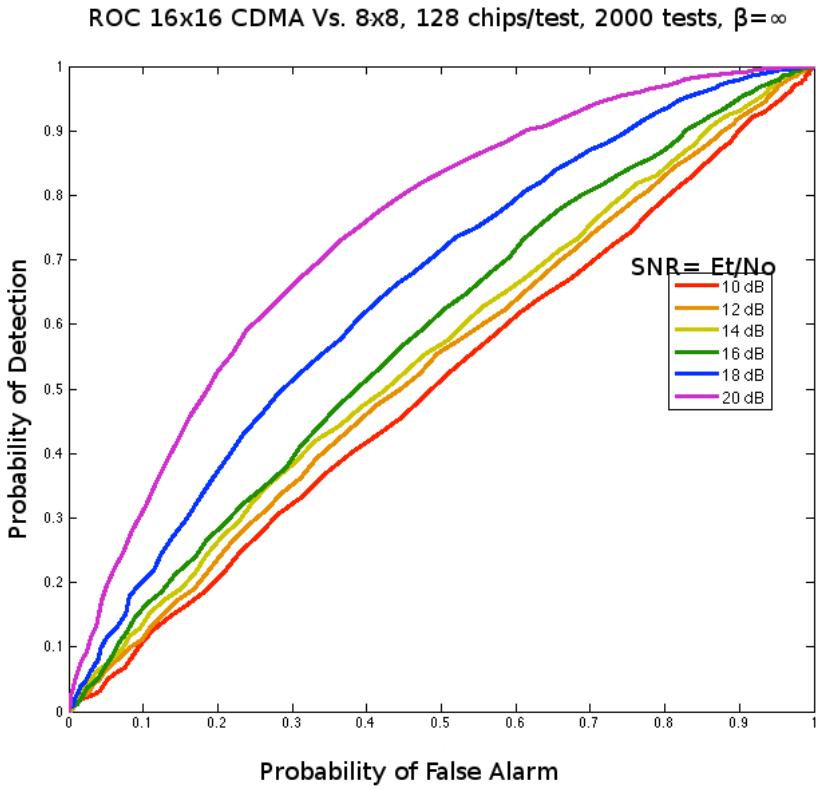

Fig. 16 Single Feature Detection: ROC $\{16, 16\}$ versus $\{8, 8\}$

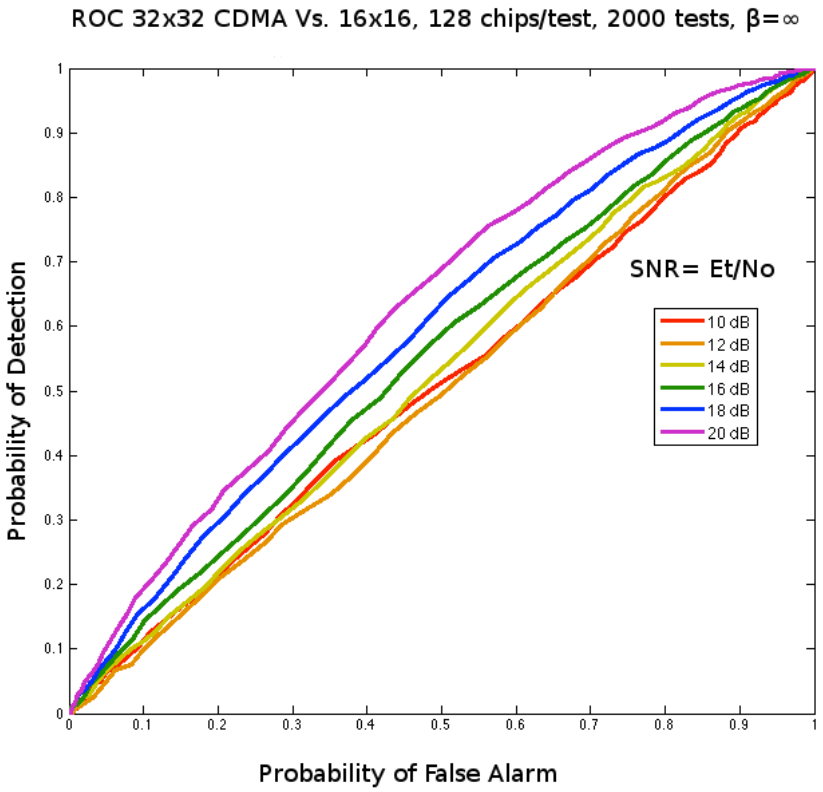

Fig. 17 Single Feature Detection: ROC $\{32, 32\}$ versus $\{16, 16\}$

5.3 Taylor's Approximation

The first consideration in approximating the average likelihood is to select the proper sample size. A sample size of K chips must be divisible by the value of the code length of all the hypotheses, so each hypothesis contains an integer number of symbols K/L. Under this assumption, the terms $\prod_{k=0}^{K/L-1} e^{-\langle \vec{r}_k, \vec{r}_k \rangle/2}$ and $\prod_{k=0}^{K/L-1} (\pi N_0)^{L/2}$ are common factors to each hypothesis and can be neglected if necessary. For convenience, the energy term of the observation is kept.

$$\prod_{k=0}^{K/L-1} \lambda(\vec{r}_k|\mathscr{H}) = \prod_{k=0}^{K/L-1} \sum_{\vec{a}\in\{\vec{a}_l^\star\}} \alpha(\vec{a},\infty) e^{\|\vec{r}_k\|^2 - \gamma_c \|\vec{a}\|^2} \prod_{i=0}^{L-1} cosh(\sqrt{2\gamma_c}\, r_{i,k}\, a_i) \tag{82}$$

The product of the hyperbolic cosine functions can be approximated using (83). The likelihood expression results in exponential terms that depends on the absolute value of each observation $r_{i,k}$. In both approximations, there is a $e^{\pm 1}$ term that can be neglected because it is a common factor to all the likelihood functions.

$$\prod_{i=0}^{L-1} e^{log(cosh(\sqrt{2\gamma_c}\, r_{i,k}\, a_i))} \approx \begin{cases} e^{\sum_{i=0}^{L-1} \sqrt{2\gamma_c}\, abs(r_{i,k})\, a_i - 1} & \text{for } \gamma_c \gg 1 \\ e^{\sum_{i=0}^{L-1} \gamma_c r_{i,k}^2\, a_i^2 + 1} & \text{for } \gamma_c \ll 1 \end{cases} \tag{83}$$

The approximation for low values of γ_c did not have a significant improvement on the classification, so only the approximation for $\gamma_c \gg 1$ was considered regardless of the value of γ_c. Combining (82) and (83) under the assumption of $\gamma_c \gg 1$ result in:

$$\prod_{k=0}^{K/L-1} \lambda(\vec{r}_k|\mathscr{H}) = \prod_{k=0}^{K/L-1} \sum_{\vec{a}\in\{\vec{a}_l^\star\}} e^{-\|abs(\vec{r}_k) - \sqrt{2\gamma_c}\vec{a}\|^2/2 - log(\alpha(\vec{a},\infty))}. \tag{84}$$

Using the log-likelihood instead of the likelihood provides further simplification. The log-sum of exponents in (84) can be replaced by choosing the maximum value of all the exponents in the sum.

$$\begin{gathered} log(\sum_l e^{\omega_l}) \approx max(\omega_l) \\ \omega_l = -\|abs(\vec{r}_k) - \sqrt{2\gamma_c}\vec{a}_l\|^2/2 + log(\alpha(\vec{a}_l,\infty)) \end{gathered} \tag{85}$$

With $\|\vec{a}\|^2 = L\,U$ and $\|\vec{r}_k\|^2$ constant, the maximum value of ω_i depends on the correlation of $\vec{a}$ and $abs(\vec{r}_k)$ and the coefficient $\alpha(\vec{a},\infty)$. Not much simplification can be applied to the likelihood coefficient other than investigating its behavior through actual calculations. This behavior is shown in Figure 18 and shows the values of $\alpha(\vec{a},\infty)$ bounded by the red and the blue graphs as the code length increases. Our approach ignores the effects of these coefficients

because bounds are difficult to assess. The approach also relying exclusively on the correlation term for approximating the likelihood function.

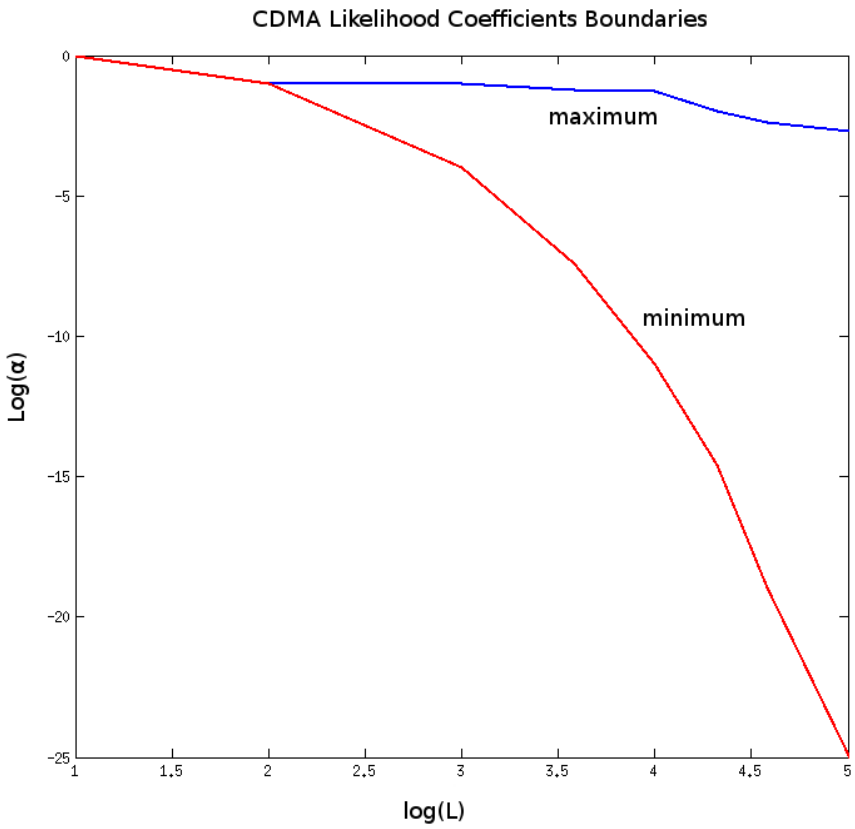

Fig. 18 Boundaries of Minimum and Maximum Coefficients vs. $log(L)$

A given feature vector $\vec{a}^{\star}$ and all its permutations qualify as features for classification. This fact presents a problem because the number of permutations becomes large for higher code lengths causing the algorithm to be computationally inefficient when computing (95). For example, $\vec{a}^{\star} = [4,4,0,0,0,0,0,0]^T$ is a feature vector under hypothesis $\mathscr{H} = \{8,8\}$, but also all its $8!/((8-2)!2!) = 28$ permutations. In this case, there are two distinct vector coefficients. In order to deal with this problem, we take advantage of a Computer Sciences theorem known as the Maximum Scalar Product.

Proposition 5.1 *If $\vec{x} = \{x_i\}^L$ and $\vec{y} = \{x_i\}^L$, and $\mathscr{P}$ is an arbitrary $L \times L$ permutation matrix, then the maximum scalar product between $\vec{x}$ and $\mathscr{P}\ \vec{y}$ is given by:*

$$\begin{aligned} max(\vec{x}^T\ \mathscr{P}\ \vec{y}) &= \sum_{i=0}^{L-1} x_{\sigma(i)}\, y_{\pi(i)} \\ x_{\sigma(i)} &\leq x_{\sigma(i+1)} \\ y_{\pi(i)} &\leq y_{\pi(i+1)} \end{aligned} \tag{86}$$

where $\sigma(i)$ and $\pi(i)$ represent permutations of the i^{th} element with $i \in \{0,1,2,...,L-2\}$.

Proof 5.1 *Let's assume that the sorted vectors are given by $\vec{x}_{(\sigma)} = \{x'_i\}^L$ and $\vec{y}_{(\pi)} = \{y'_i\}^L$. We claim that:*

$$\vec{x}^T_{(\sigma)}\ \vec{y}_{(\pi)} \geq \vec{x}^T\ \vec{y} \tag{87}$$

We define $c_{[k_1,k_2,...,k_n]}$ as:

$$c_{[k_1,k_2,...,k_n]} = \sum_{\substack{i=0 \\ i \neq k_1,k_2,...,k_n}}^{L-1} x'_i\, y'_i \tag{88}$$

Then, it can be shown that:

$$x'_{k_1} y'_{k_1} + x'_{k_1} y'_{k_1} + c_{[k_1,k_2]} \geq x'_{k_1} y'_{k_2} + x'_{k_2} y'_{k_1} + c_{[k_1,k_2]} \tag{89}$$

is true since

$$x'_{k_1}(y'_{k_1} - y'_{k_2}) + c_{[k_1,k_2]} \geq x'_{k_1}(y'_{k_2} - y'_{k_1}) + c_{[k_1,k_2]} \tag{90}$$

For distinct indexes k_1 and k_2, if $y'_{k_1} \geq y'_{k_2}$, the inequality above is true because:

$$\begin{aligned} x'_{k_1} &\geq x'_{k_2} \\ y'_{k_1} - y'_{k_2} &\geq 0 \end{aligned} \tag{91}$$

If $y'_{k_2} \geq y'_{k_1}$, the inequality also remains true because:

$$x'_{k_1} y'_{k_1} + x'_{k_2} y'_{k_2} + c_{[k_1,k_2]} \leq x'_{k_1} y'_{k_2} + x'_{k_2} y'_{k_1} + c_{[k_1,k_2]} \tag{92}$$

and

$$\begin{aligned} x'_{k_2} &\geq x'_{k_1} \\ y'_{k_2} - y'_{k_1} &\geq 0. \end{aligned} \tag{93}$$

Once a permutation of two arbitrary indexes is verified, we can proceed with a permutation of two arbitrary indexes k_3 and k_4 contained in $c_{[k_1,k_2]}$.

$$x'_{k_1}y'_{k_1}+x'_{k_2}y'_{k_2}+x'_{k_3}y'_{k_3}+x'_{k_4}y'_{k_4}+c_{[k_1,k_2,k_3,k_4]}\geq x'_{k_1}y'_{k_2}+x'_{k_3}y'_{k_4}+x'_{k_4}y'_{k_3}+x'_{k_2}y'_{k_1}+c_{[k_1,k_2,k_3,k_4]} \tag{94}$$

The inequality also holds whether $x_{k_3}\geq x_{k_4}$ or $y_{k_3}\geq y_{k_4}$ or $x_{k_4}\geq x_{k_3}$ or $y_{k_4}\geq y_{k_3}$. The process of permuting indexes is continued until the original sorted sequences become two arbitrary permuted ones: $\vec{x}=\{x_i\}^L$ and $\vec{y}=\{y_i\}^L$. ■

The use of Proposition 5.1 allows estimating the maximum value of (95). The process is depicted in Figure 19. A sequence of decorrelated samples from the CON decorrelator retains only the absolute value of the real part. The samples are formatted as an $L\times 1$ vector. The vector components are sorted and sent into a classifier that computes the likelihood (95) for a given hypothesis $\{L,U\}$. The same process is repeated for multiple hypotheses. Finally, a MAP classifier determine the class by computing the maximum log-likelihood of all the hypotheses.

$$\begin{aligned} max(\sum_i |r_{i,k}|a_i^\star) &= \vec{R}^T\,\vec{A} \\ \vec{R} &= sort(|\vec{r}_k|) \\ \vec{A} &= sort(\vec{a}_k^\star) \end{aligned} \tag{95}$$

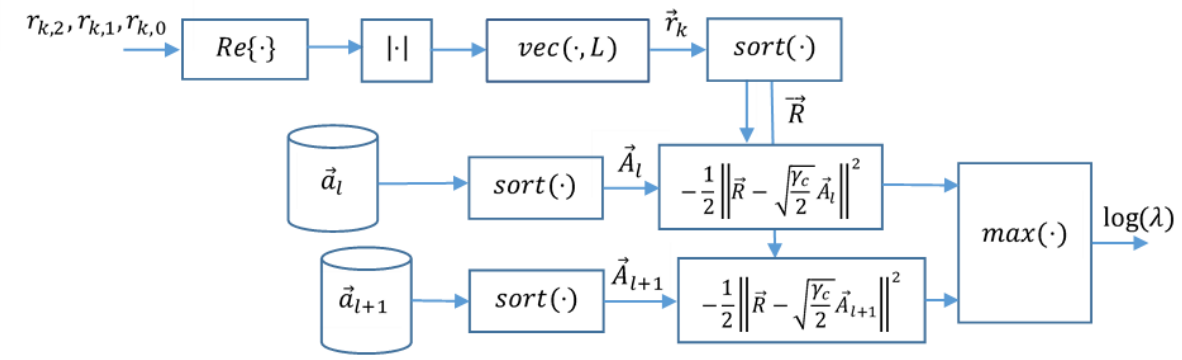

Fig. 19 Block Diagram of the Computation of the CDMA Log-Likelihood

5.3.1 Extraction of Features and Coefficients

The feature vectors for a given hypothesis must be extracted and stored in a database prior to the classification. The extraction process is obtained by using the algorithm in Appendix D. Sorting the vectors and computing the rate of occurrence reduces the complexity that comes from dealing with different permutations of the feature vector.

In the case of perfectly orthogonal codes, the feature vectors could be computed by solving the Diophantine equations given by (96). The features can take values of $a_i \in \{2n\}^{n=0:2:L}$ and f_n is the number of occurrences of a_i as given by (40), in other words, $\|\vec{a}\|^2 = L\,U$. The solution can be obtained by using the Lenstra–Lenstra–Lovász lattice basis reduction algorithm [21].

$$\begin{aligned} \sum_{i=0}^{U-1} (2n)^2 f_n &= L\,U \\ \sum_{i=0}^{L-1} f_n &= U \\ 0 \leq f_n \leq L,\ f_n &\in \mathbb{Z}^* \end{aligned} \tag{96}$$

5.4 Averaging over Unbalanced Energy

The expectation over the energies of different users can be computed under the following assumptions. A nominal energy E per symbol is assigned to each user. The energy variation from the nominal value is represented as a continuous random variable Δ_i such that the energy per symbol of the user i is given by:

$$E_i = E\ (1+\Delta_i)^2. \tag{97}$$

The model of the CDMA is constructed such that the likelihood can be separated into two terms: a contribution of a balanced CDMA and a deviation $\vec{\delta}$.

$$\vec{y} = \sqrt{E/L}\,C\,\vec{b} + \vec{\delta} + \vec{n} \tag{98}$$

Under the assumption that Δ is a diagonal matrix with uniform or normally distributed coefficients, the vector $\vec{\delta} = \sqrt{E/L}\,C\,D\,\vec{b}$ can be represented as Gaussian random vector (using the Central Limit Theorem) with zero mean and covariance matrix $\sigma^2\,I$. The expected value of the conditional likelihood on $\vec{\delta}$ is integrated over the interval $(-\infty, +\infty)$ as shown in (99). For simplicity, the integral consider only the terms that depends on $\vec{\delta}$.

$$\mathbb{E}_{\delta_i}\{e^{-\gamma_c \delta_i^2} cosh(\sqrt{2\,\gamma_c} r_{i,k} a_i)\} = \frac{e^{\frac{-\gamma_c\,\sigma^2}{\sqrt{1+2\,\gamma_c\,\sigma^2}}}}{\sqrt{1+2\,\gamma_c\,\sigma^2}} cosh(\sqrt{2\,\gamma_c} r_{i,k} a_i) \tag{99}$$

Finally, the average likelihood function of the unbalance CDMA is the product of the average likelihood of the balance CDMA and a factor χ given by (100). The addition of any additional users with unbalanced energy decreases the likelihood and degrades the accuracy of

the classifier.

$$\chi(\gamma_c\,\sigma^2) = \frac{e^{\frac{-U\gamma_c\,\sigma^2}{\sqrt{1+2\,\gamma_c\,\sigma^2}}}}{(1+2\,\gamma_c\,\sigma^2)^{U/2}} \tag{100}$$

5.5 Discussion and Results

Our simulations are based on the code presented in Appendices C to F. Each plot assumes that the code can appear in any form of permutation, the observation is frame asynchronous with $\varepsilon \in [0, L-1]$. The tests consist of binary classifications using a MAP approach using log-likelihood functions. The probability of the classes is uniform, which implies that the same number of test samples is the same for each class. A samples of K chips is used to compute the likelihood function. The sample contains K/L symbols.

Figure 20 and 21 show the accuracy versus total-SNR. It was found that a energy density to noise density ratio, referred as density ratio, serves as a better parameter for all code lengths. Usage of the density ratio parameter used in Figures 22 through 26 reveals the existence of a minimum required density ratio value for accurate classification between classes.

The plots are based on 2000 tests with chips lengths that varies between 2^6 to 2^{16} depending on the hypothesis. The chip length must ensure that there is enough number of symbols in a class for make a classification. The required chip length appears to be a function of the symbol length of the highest class.

The simplified likelihood function does not assume that the codes are perfectly orthogonal. As long as the code matrix achieves a minimum TSC, the equation should remain valid. The likelihood function for CDMA can be used to classify CDMA from other signals, given that the log-likelihood functions of the other signals are provided.

The simulations shown in Figures 27 to 31 demonstrate that the concept can be extended to underloaded CDMA cases. We may find examples where the classification is ambiguous in underloaded CDMA cases. A number of active users less that the code length, $U \leq L$ can be confused with a $L/2 \times U$ CDMA especially when the signals are constructed using (50).

A particular case of interest is the classification between orthogonal and non-orthogonal CDMA matrices. Figure 32 shows the classification between classes $\{63, 63\}$ and $\{64, 64\}$. The feature vectors were obtained from a 64×64 Sylvester Hadamard matrix on both cases. The feature vectors of the 63×63 CDMA were obtained by removing an arbitrary column and row of the orthogonal matrix. At the moment of testing classes, the Hadamard with 63 rows was substituted with a Gold Code matrix made from the following polynomials: $x^8 + x + 1$ and $x^8 + x^7 + x^2 + x^1 + 1$. Because the Gold Code shares similar features, it was possible to classify between both classes.

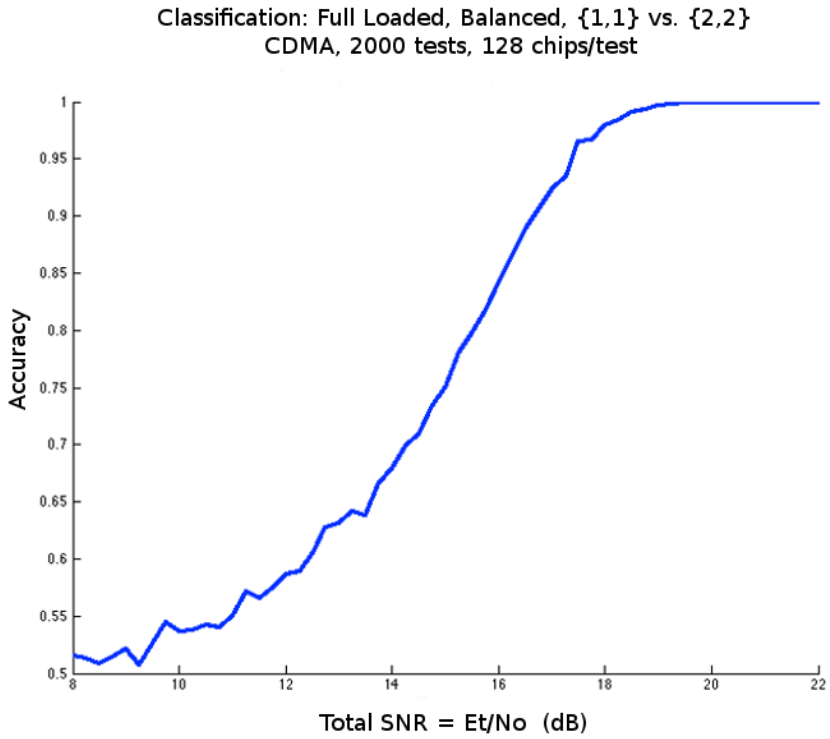

Fig. 20 Classification $\{1,1\}$ versus $\{2,2\}$ Using Symbol SNR

5.5.1 Execution Times

The simulations were conducted in a cloud using Matlab parallel programming. A single test was assigned to one of 48 available nodes. The estimated time of execution is per test is given in Table 4. The execution times depend on several variables, such as the data transfer times between nodes, the number of chips/test and the total number of feature vectors.

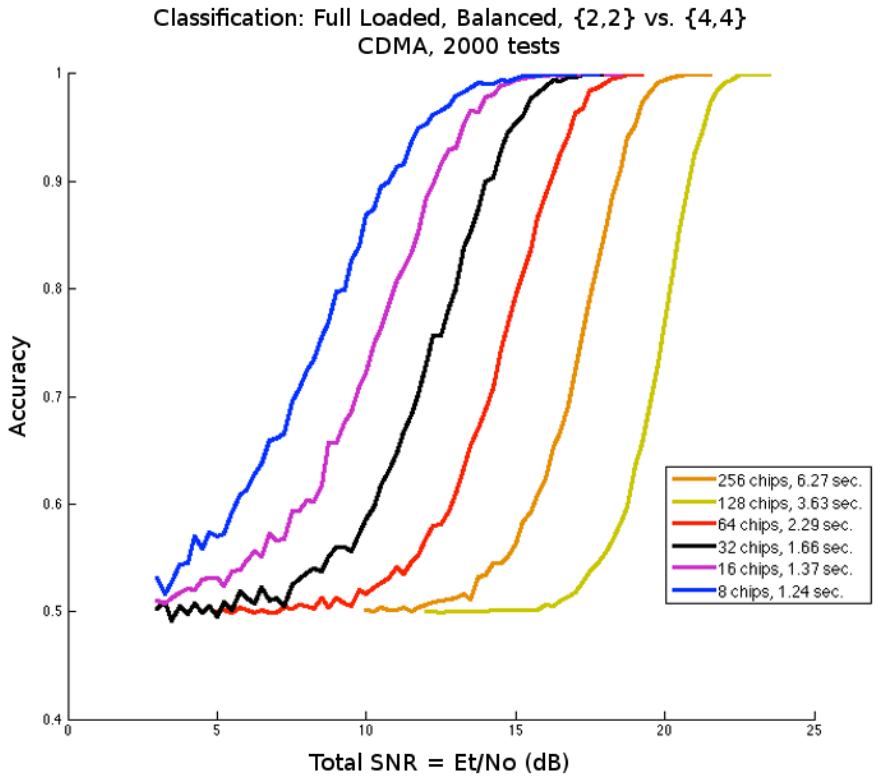

Fig. 21 Classification $\{2,2\}$ versus $\{4,4\}$ Using Symbol SNR

Table 4 Execution Times per Test

Hypothesis $\mathscr{H}_0$	Hypothesis $\mathscr{H}_1$	chips/test	features	Execution Time (sec.)
$\{1,1\}$	$\{2,2\}$	128	2	3.3
$\{2,2\}$	$\{4,4\}$	128	3	3.6
$\{4,4\}$	$\{8,8\}$	128	5	4.0
$\{8,8\}$	$\{16,16\}$	512	21	30.3
$\{16,16\}$	$\{32,32\}$	2048	56	39.5
$\{63,63\}$	$\{64,64\}$	4032	98337	15268

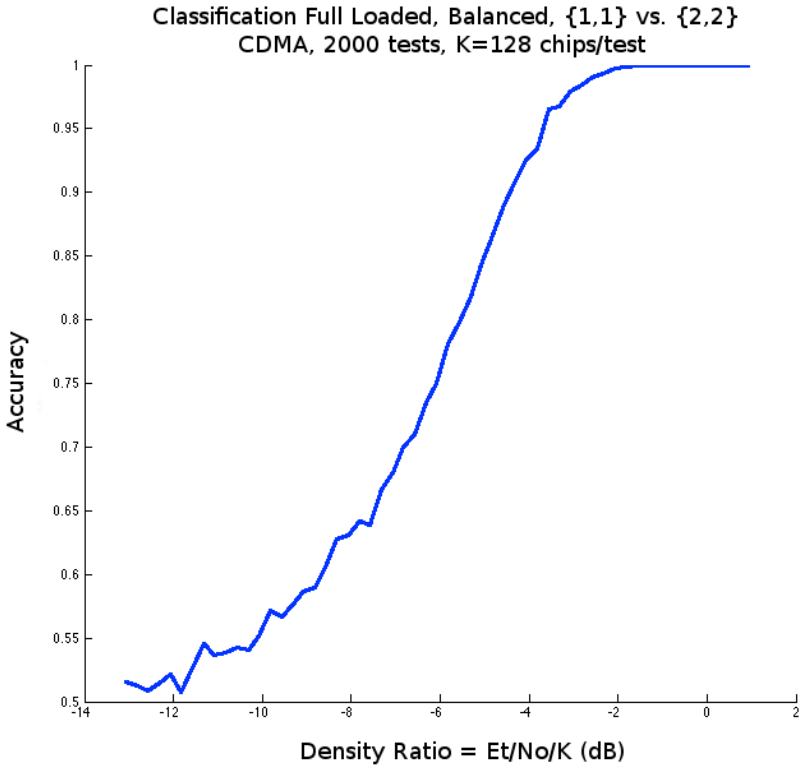

Fig. 22 Classification: $\{1,1\}$ versus $\{2,2\}$ Using Density Ratio

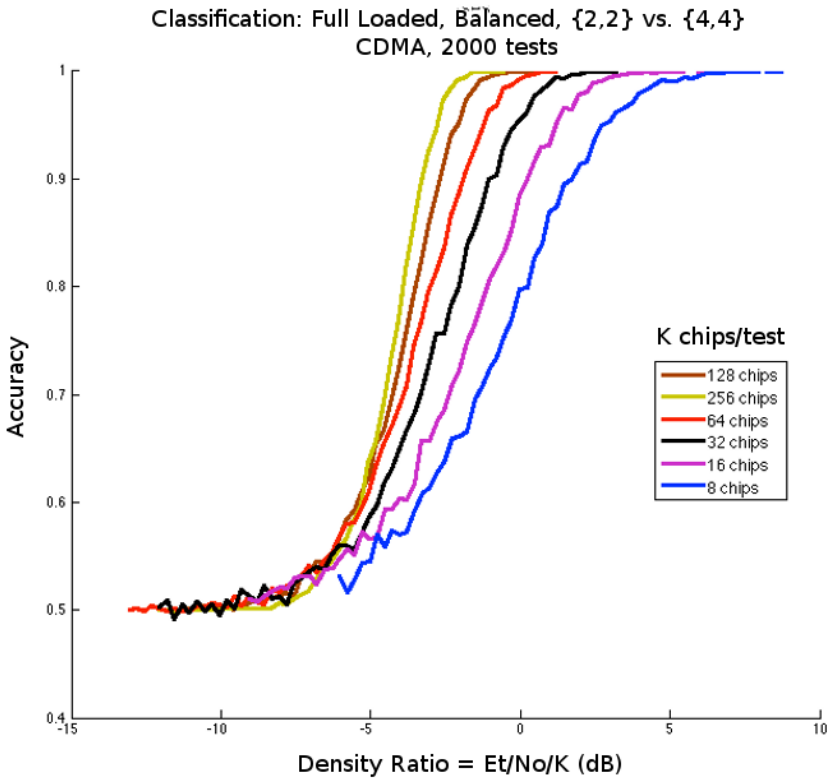

Fig. 23 Classification: $\{2,2\}$ versus $\{4,4\}$ Using Density Ratio

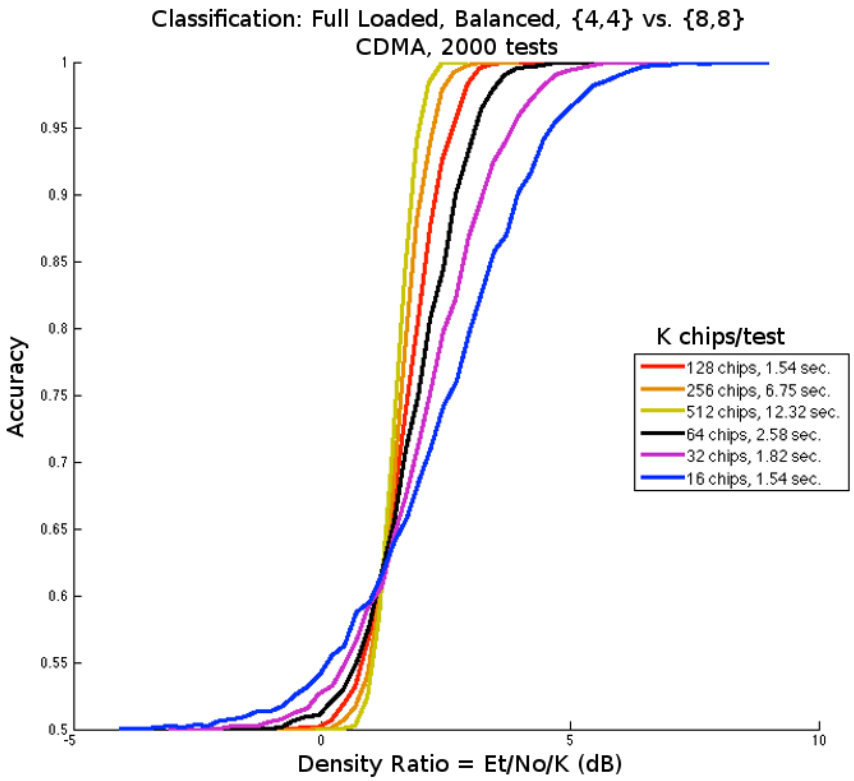

Fig. 24 Classification $\{4,4\}$ versus $\{8,8\}$

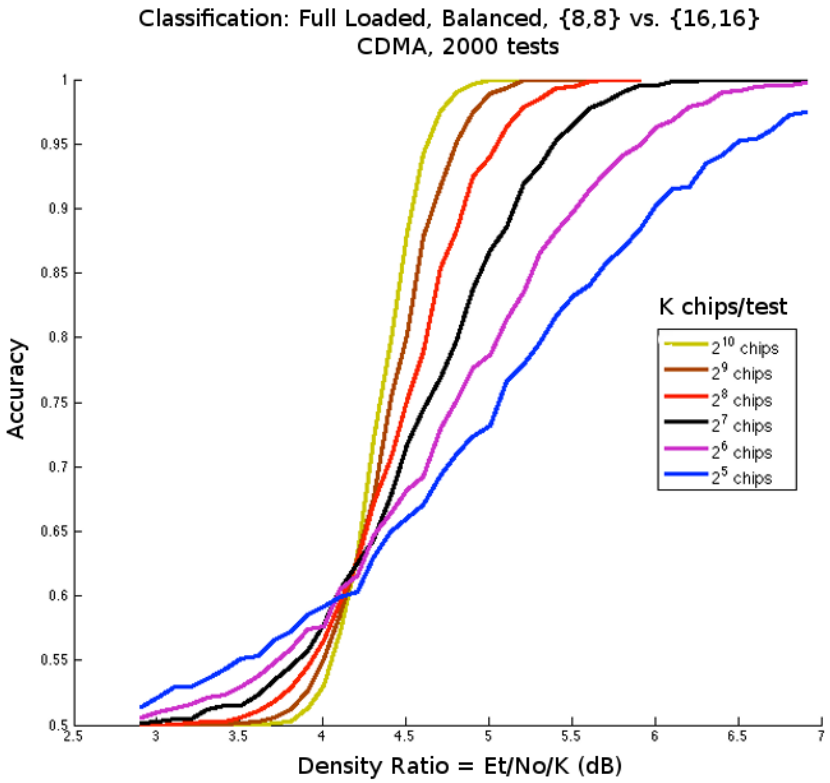

Fig. 25 Classification: $\{8,8\}$ versus $\{16,16\}$

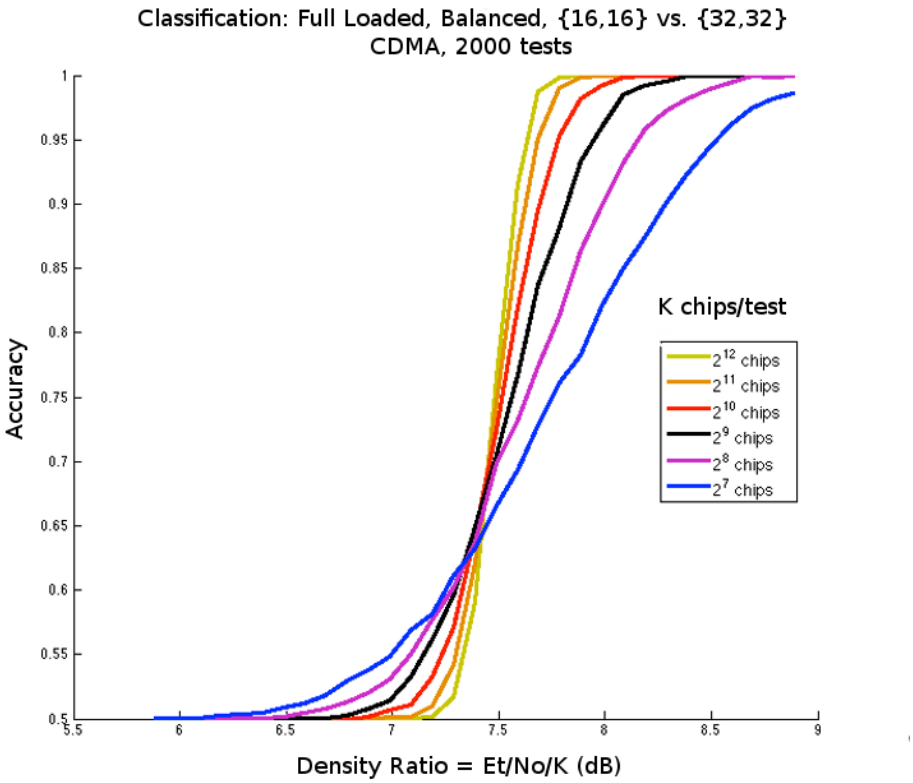

Fig. 26 Classification: $\{16, 16\}$ versus $\{32, 32\}$

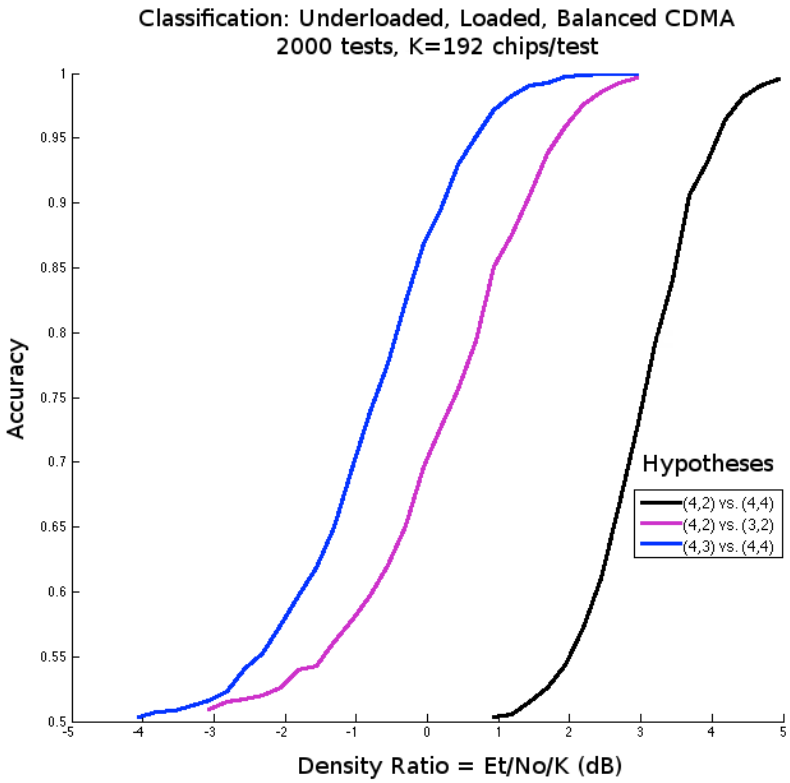

Fig. 27 Classification of Underloaded CDMA

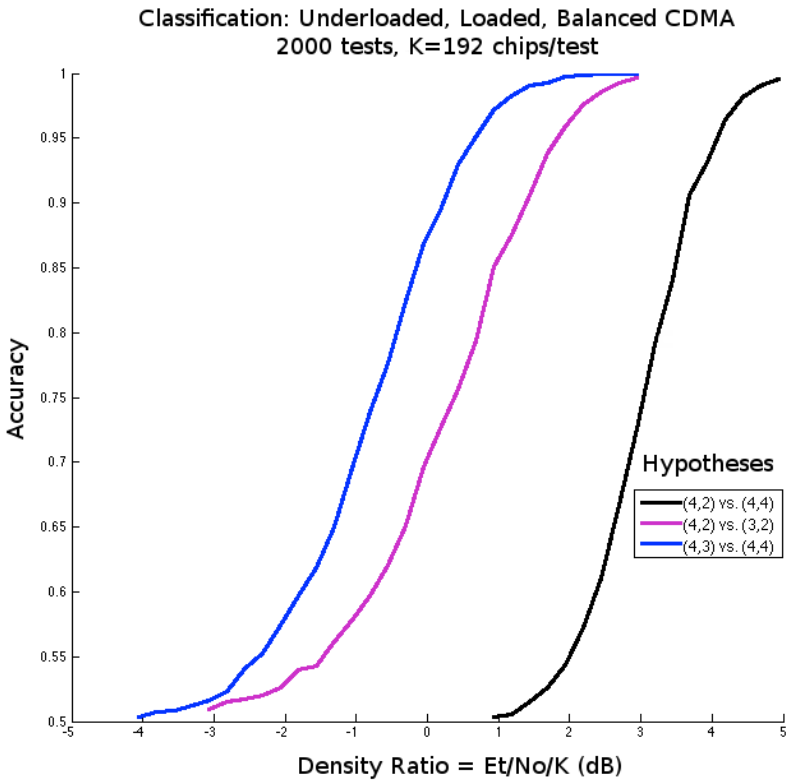

Fig. 28 Classification of Underloaded CDMA for Code Length $L = 4$

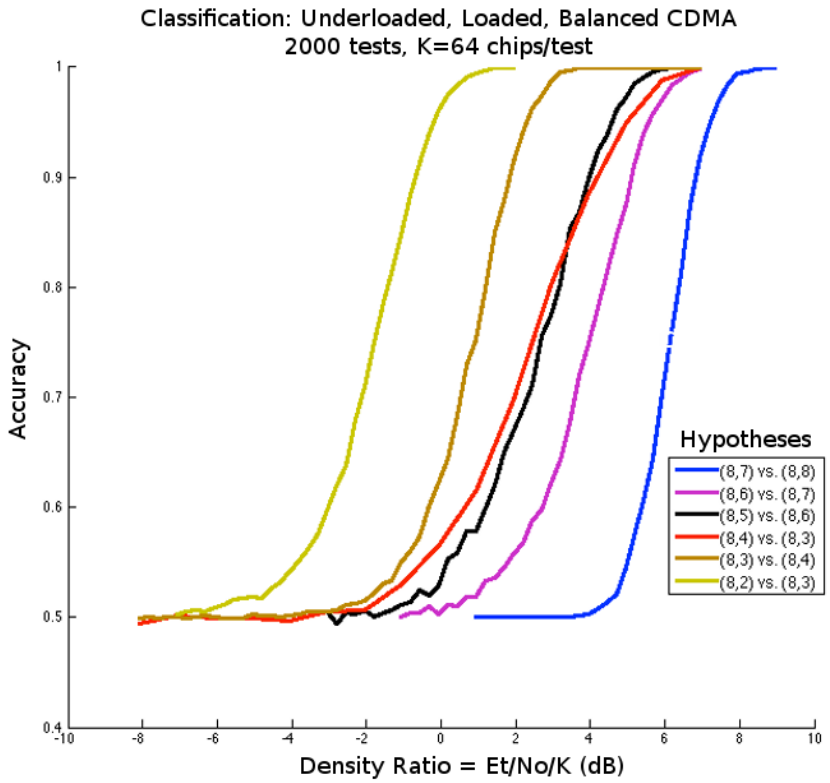

Fig. 29 Classification of Underloaded CDMA for Code Length $L = 8$

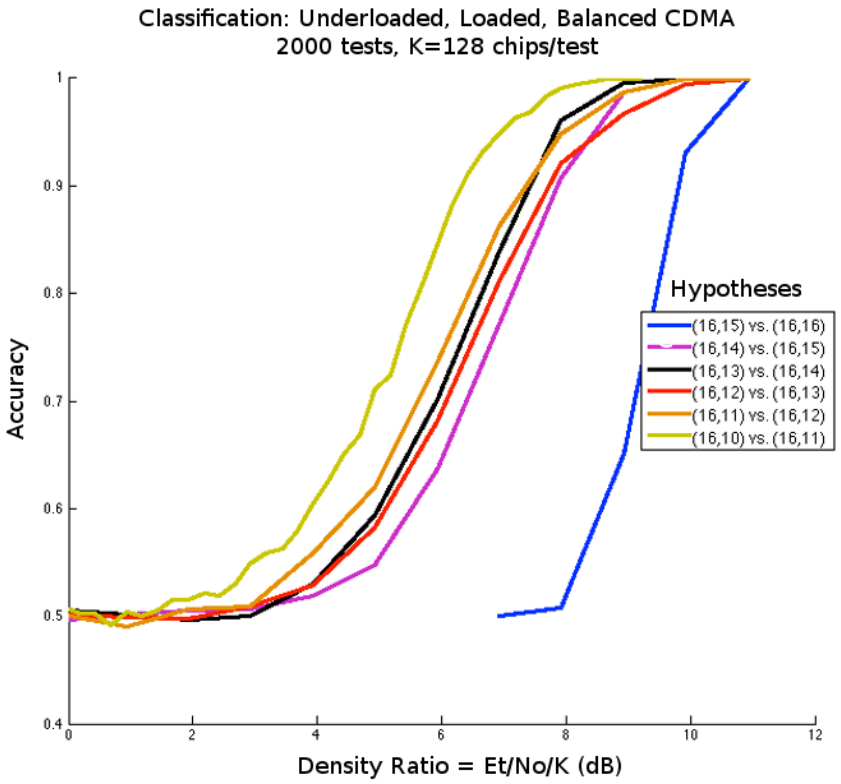

Fig. 30 Classification of Underloaded CDMA for Code Length $L = 16$

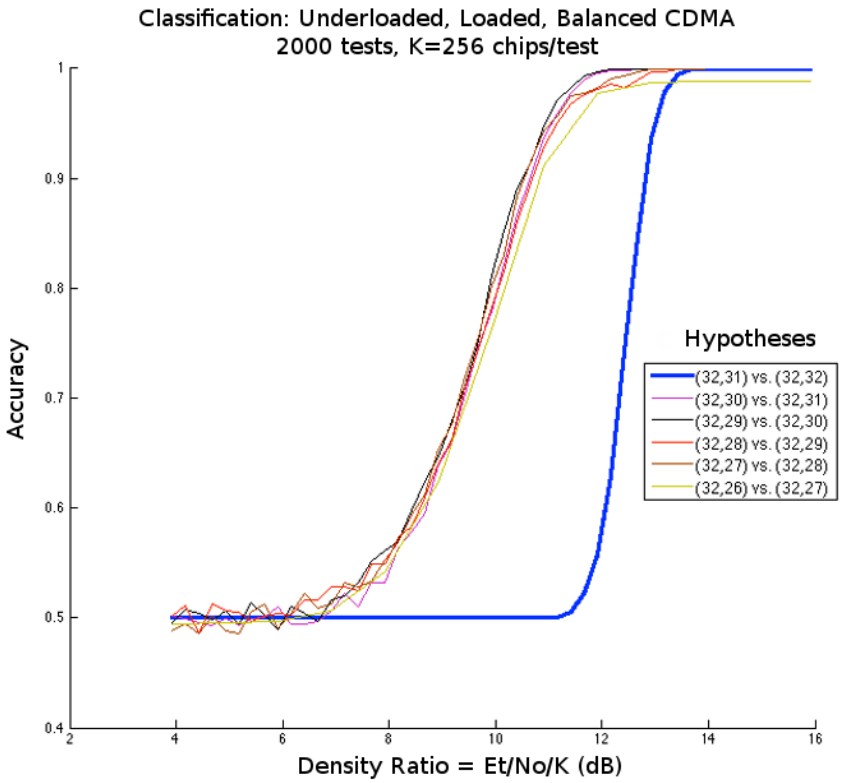

Fig. 31 Classification of Underloaded CDMA for Code Length $L = 32$

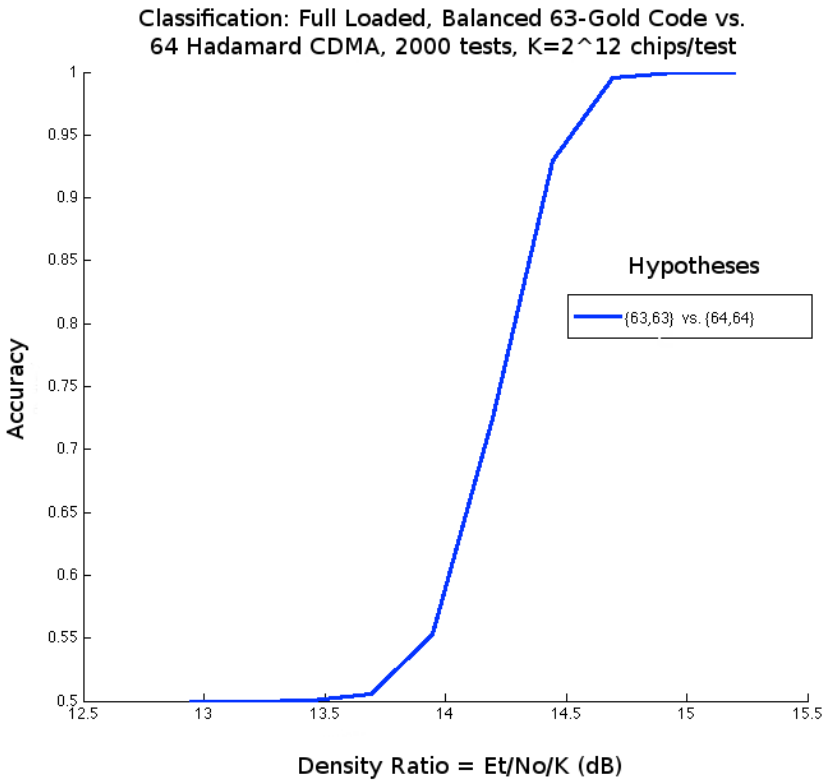

Fig. 32 Classification of $\{63, 63\}$-Gold Code versus $\{64, 64\}$-Hadamard

Table 5 CDMA Study Cases

CDMA Schemes	BPSK-data	QPSK-data
BPSK-code (spreading)	Type 1 $C \in \{\pm 1\}^{L\times U}$ $\vec{b} \in \{\pm 1\}^{U\times 1}$	Type 3 $C \in \{\pm 1\}^{L\times U}$ $\vec{b} \in \{\pm 1, \pm i\}^{U\times 1}$
QPSK-code (spreading)	Type 2 $C \in \{\pm 1 \pm i\}^{L\times U}$ $\vec{b} \in \{\pm 1\}^{U\times 1}$	Type 4 $C \in \{\pm 1 \pm i\}^{L\times U}$ $\vec{b} \in \{\pm 1, \pm i\}^{U\times 1}$

6 Extension of the Average Likelihood Method, Assumptions and Procedures

The development of the likelihood function for $L \times U$ CDMA was presented in the previous section and now it is time to extend the algorithm to all types of CDMA [3] generated from a combinations of BPSK and QPSK symbols as shown in Table 5. An alternative definition can be obtained by applying 90 degree rotations to types 2-4, but because C and $\vec{b}$ are unknowns, the selection of the model is a preference rather than a necessity. A 90 degree rotation only means that the real and imaginary parts of the model can be exchanged without affecting the modulation type.

The new developments start from reformulating the models in terms of real, binary antipodal matrices. The three terms of the likelihood function will be computed: the correlation term $\vec{r}^T\ \vec{s}$, the energy term $\vec{s}^T\ \vec{s}$ of the conditional likelihood and the covariance matrix of each CDMA model $\vec{n}^T\ \vec{n}$. The observation and noise vectors are given by a combination of the real and imaginary parts of the CDMA vectors.

6.1 General CDMA Model for Complex Types

The conversion of a complex model to a real model is possible when we use the homomorphic transformation (101) of complex matrices in $\mathbb{C}^{L\times L}$ to $\mathbb{R}^{2L\times 2L}$ real matrices.

$$\begin{aligned} \vec{r}_R + i\vec{r}_I \to \begin{bmatrix} \vec{r}_R \\ \vec{r}_I \end{bmatrix},\ \vec{n}_R + i\vec{n}_I \to \begin{bmatrix} \vec{n}_R \\ \vec{n}_I \end{bmatrix} \\ Q_R + iQ_I \to \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \end{aligned} \tag{101}$$

For complex codes, we define a new version of the TSC in (102). The variable L_Q represents the effective code length, i.e., the actual code length of the complex model times a factor of 2. The probability of the code matrix (48) in terms of the TSC retains its form. Also, we may consider extending the sets $\mathbb{S}_{\vec{a}}$ and $\mathbb{S}_{\vec{a},\tau}$ to include complex codes in the form of block matrices.

$$\tau(Q) = \|Q^T \cdot Q\|_F^2 - L_Q^2 \cdot rank(Q) \tag{102}$$

Definition 6.1 *The subset $\mathbb{S}_{\vec{a}_+,\vec{a}_-}$ of matrices $Q = Q_+ + iQ_-$ with Q_+ and $Q_- \in \{\pm 1\}^{L\times U}$ is defined as:*

$$\mathbb{S}_{\vec{a}_+,\vec{a}_-} = \{Q \mid \vec{a}_+ = Q_+\vec{1} \text{ and } \vec{a}_- = Q_-\vec{1} \text{ with } (\vec{a}_+)_i \geq 0 \text{ and } (\vec{a}_-)_i \geq 0\ \}. \tag{103}$$

Definition 6.2 *The subset $\mathbb{S}_{\vec{a}_+,\vec{a}_-,\tau}$ is defined as:*

$$\mathbb{S}_{\vec{a}_+,\vec{a}_-,\tau} = \{Q \mid Q \in \mathbb{S}_{\vec{a}_+,\vec{a}_-} \text{ and } \tau(Q) = \tau\}. \tag{104}$$

The next two properties will allow simplifying the CDMA model in terms of binary antipodal matrices. Both properties are used in the development the average likelihood for types 2 and 4.

$$X = \begin{bmatrix} I & I \\ -I & I \end{bmatrix},\ X^{-1} = \frac{1}{2}X^T \tag{105}$$

$$X \cdot \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \cdot X^{-1} = \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \tag{106}$$

6.2 Type 3 CDMA Average Likelihood

The type 3 CDMA Average Likelihood is very simple. We start with the block matrix model in (107). There are two orthogonal equations of interest, one for $\vec{y}_R$ and the other for $\vec{y}_I$. By treating the real and imaginary part as independent equations, the problem can be divided in two sets of orthogonal Type 1 CDMA under the assumption of uncorrelated noise. Matrices G_{R1} and G_{R2} are two diagonal, binary-antipodal matrices that has the possible sign variations of the real and imaginary parts of the transmit CDMA signal.

$$\begin{bmatrix} \vec{y}_R \\ \vec{y}_I \end{bmatrix} = \sqrt{\frac{E_s}{L}} \begin{bmatrix} G_{R1} & \mathbf{0} \\ \mathbf{0} & G_{R2} \end{bmatrix} \begin{bmatrix} Q_R & \mathbf{0} \\ \mathbf{0} & Q_R \end{bmatrix} \begin{bmatrix} \vec{b}_R \\ \vec{b}_I \end{bmatrix} + \begin{bmatrix} \vec{n}_R \\ \vec{n}_I \end{bmatrix}$$
$$I:\ \vec{y}_R = \sqrt{\frac{E}{L}} G_{R1}\, Q_R\, \vec{b}_R + \vec{n}_R \tag{107}$$
$$II:\ \vec{y}_I = \sqrt{\frac{E}{L}} G_{R2}\, Q_R\, \vec{b}_I + \vec{n}_I$$

The average likelihood function for type 3 can be approximated by the product of the likelihood in (107)-I and (107)-II. Although the approximation assumes that $G_{R1} \neq G_{R2}$, the simulations show that the assumption is reasonable for detecting type 3 without increasing the complexity of the average likelihood function.

$$\lambda_{T2}(\vec{r}_R + i\vec{r}_I|\mathscr{H}) = \lambda_{T1}(\vec{r}_R|\mathscr{H}) \cdot \lambda_{T1}(\vec{r}_I|\mathscr{H}) \tag{108}$$

An empirical calculation of the likelihood using symbolic algebra reveals that the true likelihood is a product of hyperbolic cosine terms that depend on $\vec{r}_R^T\vec{a}_R$, $\vec{r}_I^T\vec{a}_I$, $\vec{r}_I^T\vec{a}_R$ and $\vec{r}_I^T\vec{a}_I$. The exact equation ensures that $G_{R1} = G_{R2}$. This relationship is not significant at the time of classifying CDMA in terms of the code length and number of users for type 3, so it can be relaxed to $G_{R1} \neq G_{R2}$.

6.2.1 Discussion and Results

The experiments assume that the code matrices appear in any form of column and row permutation and the correlator process frame asynchronous samples. The complex noise power is $N_0/2$, so the real and imaginary components contribute with half of the noise $N_0/4$. Type 3 CDMA uses the same features for Type 1 CDMA. The average likelihood is the product of two orthogonal Type 1 CDMA, so the performance is expected to be similar. Figure 33 shows a comparison between the classification of the classes $\{2,2\}$ and $\{4,4\}$. The accuracy of type 3 has a similar profile, but shifted backward by approximately 3 dB.

The plots in Figure 35 were based on a variable chips lengths that varies between 2^6 to 2^{16} depending on the hypothesis. The chip length must ensure that there is enough number of symbols in a class for make a classification.

A characteristic feature of these plots is the sudden break in the performance of the classifier. The minimum density ratio appears to be a function of the highest code length hypothesis in the classifier. Figure 34 shows an approximate relationship between the 75 percent accuracy point and versus the highest code length in a binary classification. (The data was interpolated from the markers in 35 with the code length is shown in the legend.) An interpolation on the graph suggests that a $\{16,16\}$ versus $\{24,24\}$ classifier requires a minimum density ratio of 6.14 dB. The experimental value agrees with this prediction.

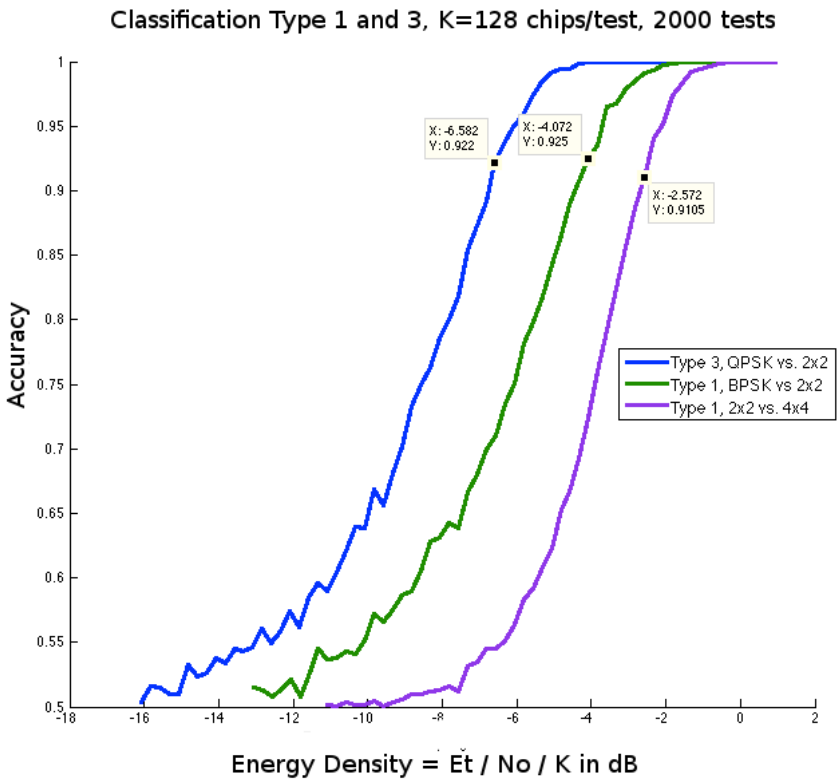

Fig. 33 Classification of $\{1,1\}$ versus $\{2,2\}$ for Types 1 and 3

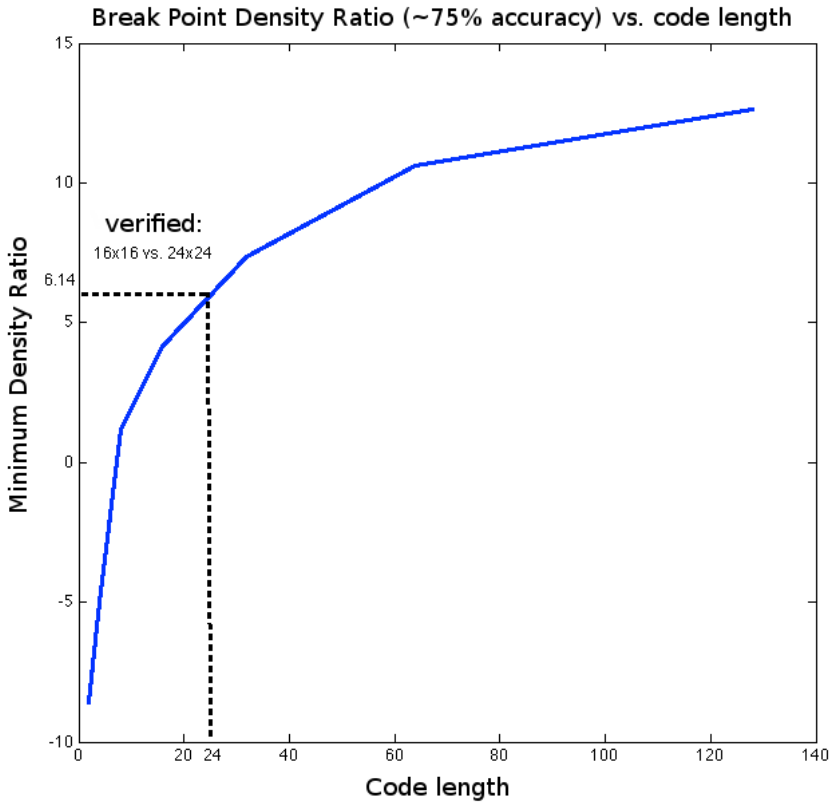

Fig. 34 Classifier Breaking Points for Type 3

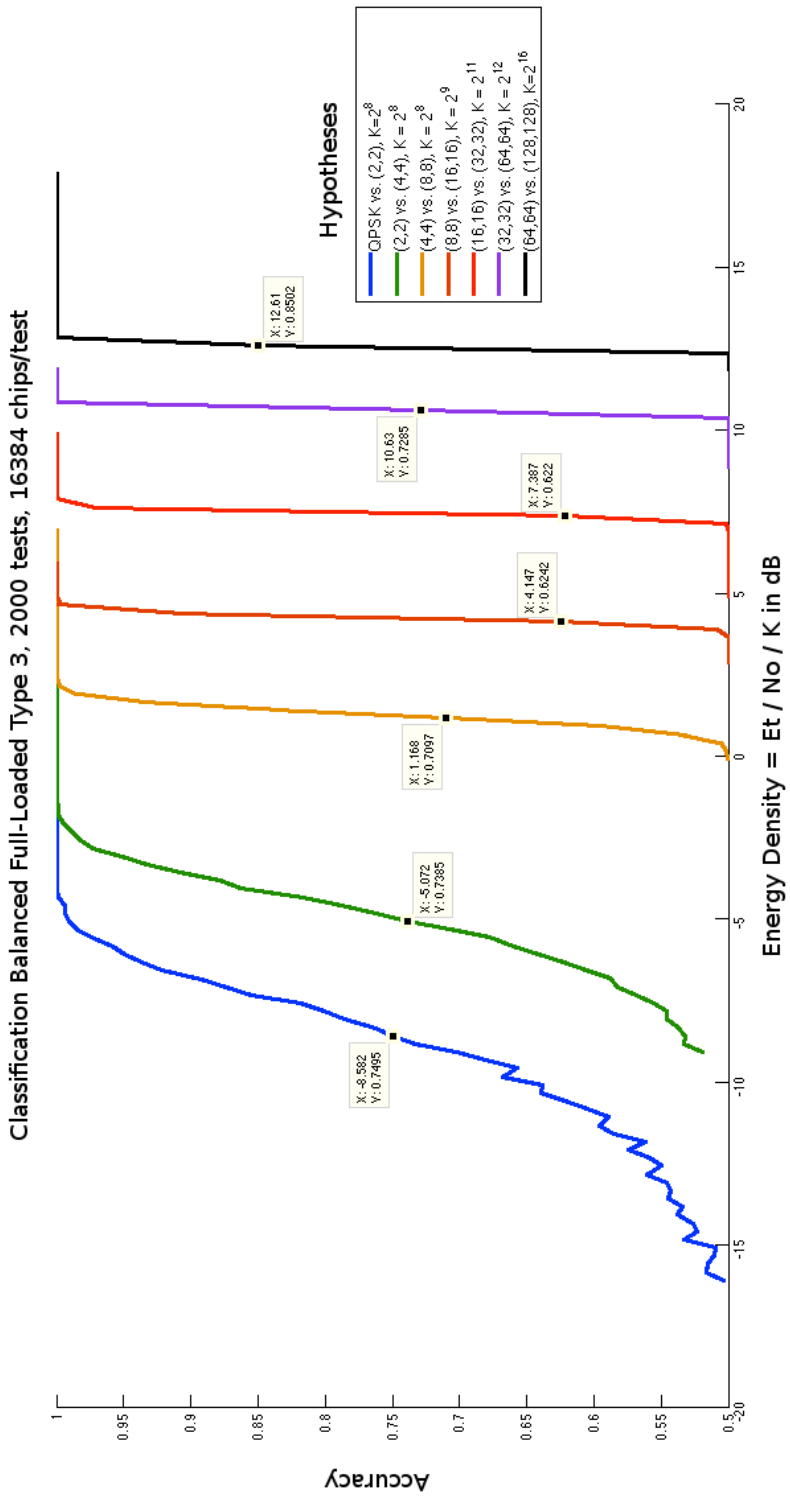

Fig. 35 Classification of $\{L, L\}$ versus $\{L/2, L/2\}$ for $log_2(L)$ between 2 and 7

6.3 Type 2 CDMA Average Likelihood

The average likelihood of a type 2 CDMA can also be expressed in terms of the type 1 likelihood under simplifying assumptions. The model in (109) resembles a type 1 CDMA with the hypothesis $\mathscr{H} = \{2L, U\}$ where U represents either the number of variables in the real or imaginary part. Its block matrix G consists of four diagonal matrices with symbols $(G_R)_{i,i}$ and $(G_I)_{i,i} \in \{0, \pm 1\}$.

$$\begin{bmatrix} \vec{y}_R \\ \vec{y}_I \end{bmatrix} = \sqrt{\frac{E_s}{2L}} \begin{bmatrix} G_{R1} & -G_{I1} \\ G_{I2} & G_{R2} \end{bmatrix} \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \begin{bmatrix} \vec{b}_R \\ \vec{0} \end{bmatrix} + \begin{bmatrix} \vec{n}_R \\ \vec{n}_I \end{bmatrix} \tag{109}$$

The complex code matrices in this study were derived from Hadamard codes. The construction was made by selecting a Hadamard matrix and multiplying it by a permutation involutory matrix to form the imaginary part. A 90 degree rotation in the complex plane might be required in order to express the code according to the definition in Table (5).

Definition 6.3 *A permutation involutory matrix has the following properties:*

$$\begin{aligned} P \cdot P &= I, \\ P\,\vec{1} &= \vec{1}, \\ P_{i,j} &\in \{0, 1\}. \end{aligned} \tag{110}$$

Proposition 6.1 *If H is a real Hadamard matrix and P is a permutation involutory matrix with $P \neq I$, then the code $Q = H + iP\,H$ is a complex Hadamard matrix.*

Proof 6.1 *This can be verified by calculating $Q^H\,Q = (2L)I$.* ■

Based on the previous proposition, one can derive matrices Q_R and Q_I satisfying the condition imposed in Table 5. Setting the sign of the amplitude vectors to $G_{I1} = G_{I2} = \mathbf{0}$, $G_{R1} \neq G_{R2} = \mathbf{0}$ results in two orthogonal sets of equations given by: (111).

$$\begin{aligned} I &: \vec{y}_R = \sqrt{\frac{E}{2L}} G_{R1}\, Q_R\, \vec{b}_R + \vec{n}_R \\ II &: \vec{y}_I = \sqrt{\frac{E}{2L}} G_{R2}\, Q_I\, \vec{b}_R + \vec{n}_I. \end{aligned} \tag{111}$$

Without entering in details, the average likelihood approximation of type 2 is identical to the type 3 already discussed.

$$\lambda_{T3}(\vec{r}_R + i\vec{r}_I | \mathscr{H}) = \lambda_{T2}(\vec{r}_R + i\vec{r}_I | \mathscr{H}) \tag{112}$$

A computation using symbolic algebra reveals that the true likelihood function of type 2 is a hyperbolic product of the terms: $\vec{r}_R^T \vec{a}_R$, $\vec{r}_I^T \vec{a}_I$, $\vec{r}_I^T \vec{a}_R$ and $\vec{r}_I^T \vec{a}_I$. The exact equation characterizes a true type 2 signals that satisfies the condition $G_{R1} = G_{R2}$. This likelihood would be useful for differentiating between CDMA types; however, the relationship is not essential for classifying CDMA in terms of the code length and number of users.

6.3.1 Discussion and Results

Figures 36 to 38 show the classification of type 2 signals using 200 tests, and chip/test that varies from 2^8 to 2^{16}. The tests assume that the code matrices comes in any form of column and row permutation and the correlator process frame asynchronous samples. The complex noise power is $N_0/2$, so the real and imaginary components contribute with half of the noise $N_0/4$. The graphs are similar to the type 3 CDMA classification that was already discussed.

Figure 37 shows several classifiers and the characteristic breaking points for fully-loaded balanced CMDA. Figures 39 to 38 shows the classification of unbalanced CDMA with variances in the amplitudes $\sigma^2 = \{0, .1, .25\}$. The effect on amplitude unbalance on the accuracy is more evident in the classification of higher code lengths since its effect grows exponentially as a function of the number of users according to (100).

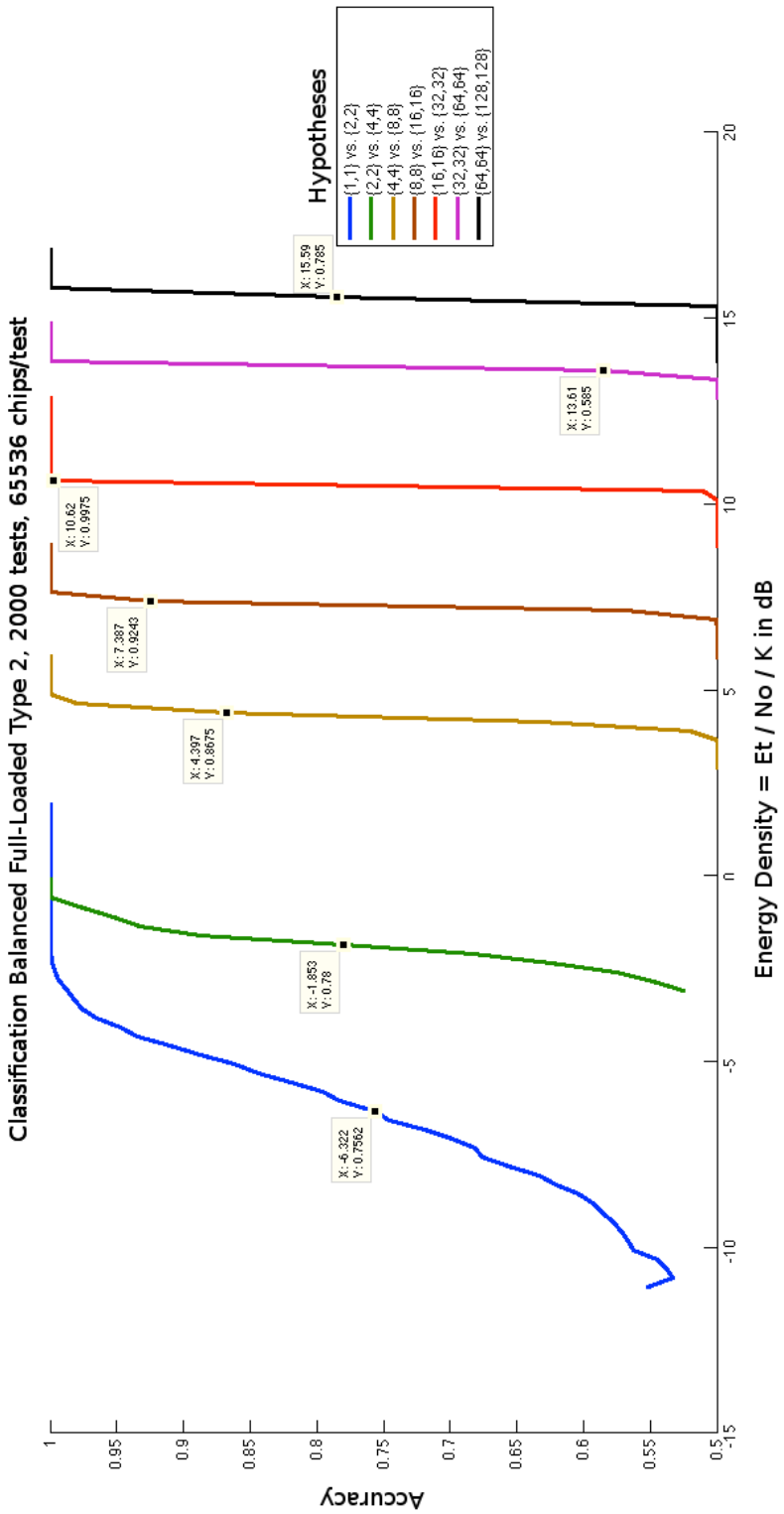

Fig. 36 Classification of $\{L, L\}$ versus $\{L/2, L/2\}$ for $log_2(L)$ between 2 and 7

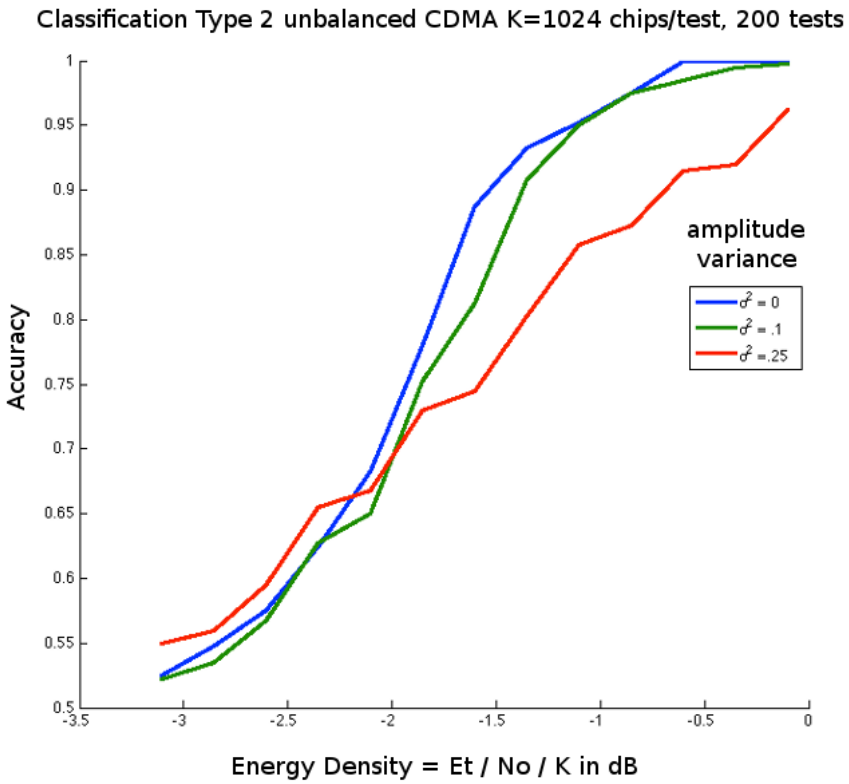

Fig. 37 Classification of $\{1,1\}$ versus $\{2,2\}$ for Types 1 and 3

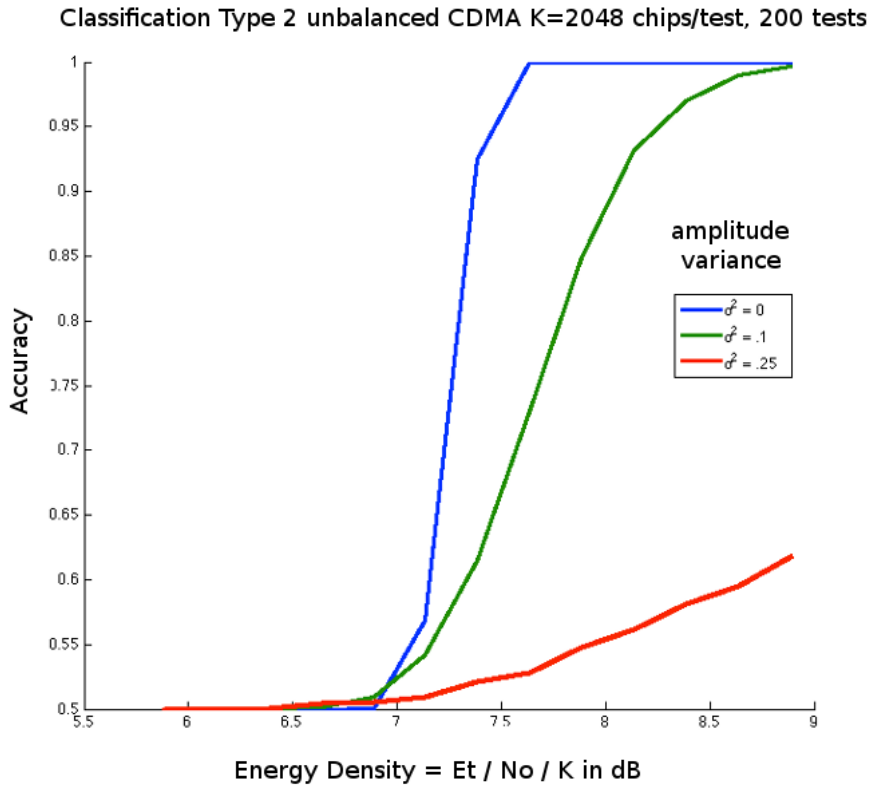

Fig. 38 Classification of $\{1,1\}$ versus $\{2,2\}$ for Types 1 and 3

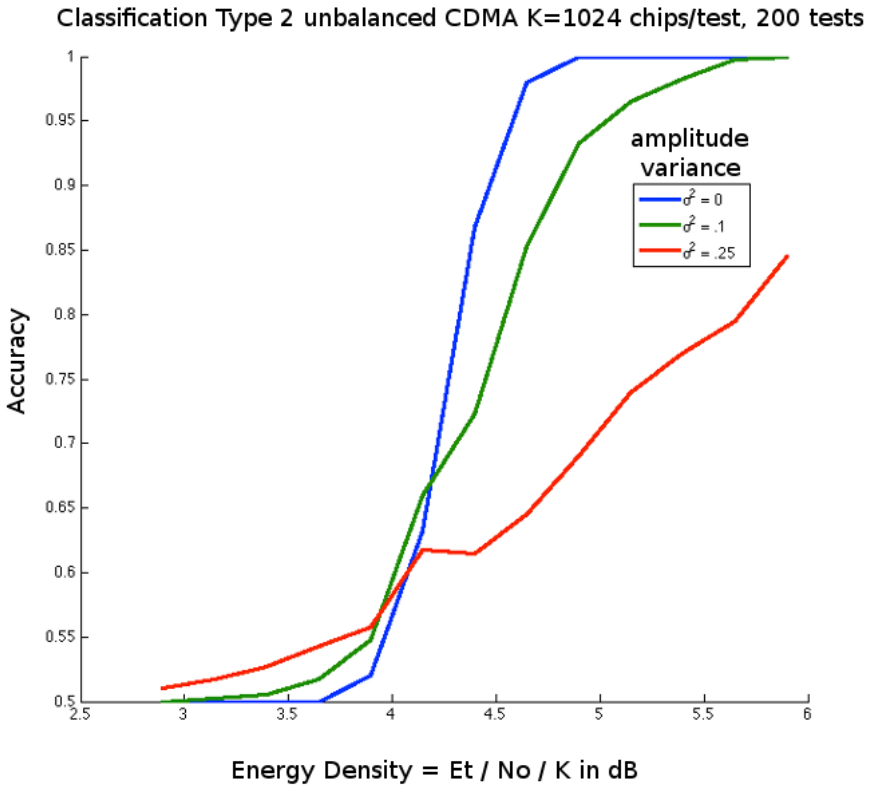

Fig. 39 Classification of $\{1,1\}$ versus $\{2,2\}$ for Types 1 and 3

6.4 Type 4 CDMA Average Likelihood

The type 4 CDMA model can be viewed as a type 1 CDMA with twice the code length and number of users. The model is subject to code matrices that follows the structure given by (113).

$$\begin{bmatrix}\vec{y}_R\\ \vec{y}_I\end{bmatrix} = \sqrt{\frac{E_s}{2L}}\begin{bmatrix}G_{R1} & -G_{I1}\\ G_{I2} & G_{R2}\end{bmatrix}\begin{bmatrix}Q_R & -Q_I\\ Q_I & Q_R\end{bmatrix}\begin{bmatrix}\vec{b}_R\\ \vec{b}_I\end{bmatrix} + \begin{bmatrix}\vec{n}_R\\ \vec{n}_I\end{bmatrix} \tag{113}$$

If the assumptions $G_{R1} = G_{R2}$, $G_{I1} = G_{I2}$ are enforced, then the sets $\mathbb{S}_{\vec{a}_R,\vec{a}_I,\tau}$ need to be considered in the development of the likelihood. The feature vectors $\vec{a}_R = Q_R\vec{b}_R - Q_I\vec{b}_I$ and $\vec{a}_I = Q_I\vec{b}_R + Q_R\vec{b}_I$ have positive coefficients by definition. The product over the block matrix G would restore the signature ± 1 to each feature vector, which is a 0 or a 180 degree rotation in the complex plane. We should also allow these block matrices to implement a 90 and -90 degree rotations because such rotations do not change the modulation type. In all these cases, the diagonal matrices G_{R1} and G_{I1} can take values $\{0,+1,-1\}$ according to the rotation in the complex plane such that any combination given by $\pm(\vec{a}_R)_j \pm i(\vec{a}_I)_j$ or $\pm(\vec{a}_I)_j \pm i(\vec{a}_R)_j$ is included in our averaging process.

The conversion to a model made of binary antipodal variables (114) is obtained by multiplying the matrix X in equation (105) and applying the property discussed in (106). Vectors $[\vec{b}_R;\vec{b}_I]$, $[\vec{n}_R;\vec{n}_I]$ and $[\vec{r}_R;\vec{r}_I]$ are redefined using the following convention: $\vec{x}_- = -\vec{x}_R + \vec{x}_I$, $\vec{x}_+ = \vec{x}_R + \vec{x}_I$ to produce the binary antipodal vectors in (115). The simplest model is obtained by setting $G_{I1} = G_{I2} = \mathbf{0}$.

$$X\begin{bmatrix}\vec{y}_R\\ \vec{y}_I\end{bmatrix} = \sqrt{\frac{E_s}{2L}}X\begin{bmatrix}G_{R1} & \mathbf{0}\\ \mathbf{0} & G_{R1}\end{bmatrix}X^{-1}X\begin{bmatrix}Q_R & -Q_I\\ Q_I & Q_R\end{bmatrix}X^{-1}X\begin{bmatrix}\vec{b}_R\\ \vec{b}_I\end{bmatrix} + X\begin{bmatrix}\vec{n}_R\\ \vec{n}_I\end{bmatrix} \tag{114}$$

$$\begin{bmatrix}\vec{y}_+\\ \vec{y}_-\end{bmatrix} = \sqrt{\frac{E_s}{2L}}\begin{bmatrix}G_{R1} & \mathbf{0}\\ \mathbf{0} & G_{R1}\end{bmatrix}\begin{bmatrix}Q_R & -Q_I\\ Q_I & Q_R\end{bmatrix}\begin{bmatrix}\vec{b}_+\\ \vec{b}_-\end{bmatrix} + \begin{bmatrix}\vec{n}_+\\ \vec{n}_-\end{bmatrix} \tag{115}$$

The average likelihood for type 4 CDMA (116) is a special case of type 1 CDMA with twice the code length and number of users:

$$\lambda_{T4}(\vec{r}_R + i\vec{r}_I|\mathscr{H} = \{L,U\},N_0/2) \approx \lambda_{T1}(\vec{R}|\mathscr{H} = \{2L,2U\},N_0), \tag{116}$$

where $\vec{R} = [\vec{r}_+;\vec{r}_-]$, $\vec{N} = [\vec{n}_+;\vec{n}_-]$, $\mathbb{E}\{\vec{N}\vec{N}^T\} = N_0$. The real component of the noise is assumed to satisfy $\mathbb{E}\{\vec{n}_R\vec{n}_R^T\} = \mathbb{E}\{\vec{n}_I\vec{n}_I^T\} = N_0/2$.

6.4.1 Alternative Model

We define $\vec{g}_+$ and $\vec{g}_-$ to be binary antipodal vectors $\{\pm 1\}^L$ satisfying the following relationship:

$$\begin{aligned} \vec{g}_+ &= \vec{g}_R + \vec{g}_I \\ \vec{g}_- &= -\vec{g}_R + \vec{g}_I \\ G_{R1} &= diag(\vec{g}_R) \\ G_{I1} &= diag(\vec{g}_I). \end{aligned} \tag{117}$$

We can calculate the correlator term of the conditional likelihood using (113) with $G_{R1} = G_{R2}$ and $G_{I1} = G_{I2}$ and take advantage of the Proposition 4.2.

$$\begin{bmatrix} \vec{r}_+ \\ \vec{r}_- \end{bmatrix}^T \begin{bmatrix} G_{R1} & -G_{I1} \\ G_{I1} & G_{R2} \end{bmatrix} \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \begin{bmatrix} \vec{b}_+ \\ \vec{b}_- \end{bmatrix} \tag{118}$$

The following form of the correlation provides what is necessary to compute the average over the block matrix G.

$$\begin{bmatrix} \vec{g}_+ \\ \vec{g}_- \end{bmatrix}^T \begin{bmatrix} diag(-\vec{r}_I) & diag(\vec{r}_R) \\ diag(\vec{r}_R) & diag(\vec{r}_I) \end{bmatrix} \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \begin{bmatrix} \vec{b}_+ \\ \vec{b}_- \end{bmatrix} \tag{119}$$

Since the block matrix G is orthonormal, the energy term remains the same:

$$\left\| \begin{bmatrix} G_{R1} & -G_{I1} \\ G_{I1} & G_{R2} \end{bmatrix} \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \begin{bmatrix} \vec{b}_+ \\ \vec{b}_- \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} Q_R & -Q_I \\ Q_I & Q_R \end{bmatrix} \begin{bmatrix} \vec{b}_+ \\ \vec{b}_- \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \vec{a}_R \\ \vec{a}_I \end{bmatrix} \right\|^2. \tag{120}$$

The difference between the type 4 model in (119) and the type 1 correlation term in (56) is the substitution of $diag(\vec{r})$ with the following matrix:

$$diag(\vec{r}_R) \rightarrow D = \begin{bmatrix} diag(-\vec{r}_I) & diag(\vec{r}_R) \\ diag(\vec{r}_R) & diag(\vec{r}_I) \end{bmatrix}. \tag{121}$$

Using a development a similar treatment used in type 1 CDMA, we can arrive to the average likelihood function shown in equations (122) and (123).

$$\lambda_{T4}(\vec{R}|\mathscr{H} = \{L,U\}) = \frac{e^{-\vec{R}^T\vec{R}/2}}{(\pi N_0)^{L/2}} \sum_{\vec{a}} e^{-\gamma_c \|\vec{a}\|^2} \alpha(\vec{a},\beta) \prod_{i=0}^{2L-1} cosh(\sqrt{2\gamma_c} \sum_j D^*_{i,j}\, a_j) \tag{122}$$

with $\vec{R} = [\vec{r}_+; \vec{r}_-]$, $\vec{a} = [\vec{a}_R; \vec{a}_I]$ and

$$\begin{aligned} \alpha(\vec{a}, \beta) &= \sum_{\tau} \sum_{Q \in \mathbb{S}_{\vec{a}_R, \vec{a}_I, \tau}} \frac{e^{-\beta \tau(Q)}}{W} \\ \gamma_c &= \frac{E}{2LN_0}. \end{aligned} \tag{123}$$

The same formula can be use for Type 2 CDMA, which can be seen as a special form of type 4 CDMA with $\vec{b}_I = 0$.

6.4.2 Discussion and Results

The first task in generating simulations was the extraction of the feature vectors defined in (104). It was found that the composite feature vector $\vec{A} = [\vec{a}_R; \vec{a}_I]$ has the same features for a corresponding Type 1 of size $\{2L, 2U\}$. This finding supports our approximation of a $L \times U$ Type 4 CDMA likelihood to a Type 1 $2L \times 2U$ CDMA. The experiments also assume that the code matrices comes in any form of column and row permutation and the correlator process frame asynchronous samples. The noise power of the model is $N_0/2$ and consists of the contribution of the real and imaginary components under the assumption of uncorrelated noise. Only the a short range of density ratios is considered due to the long execution time.

Figure 40 resembles the same characteristic features of the previously discussed classifiers. It seems that all complex types of CDMA can be viewed as a different way of encoding binary antipodal symbols. The average likelihood under the unbalanced energy assumption degrades the performance of the classifier as expected.

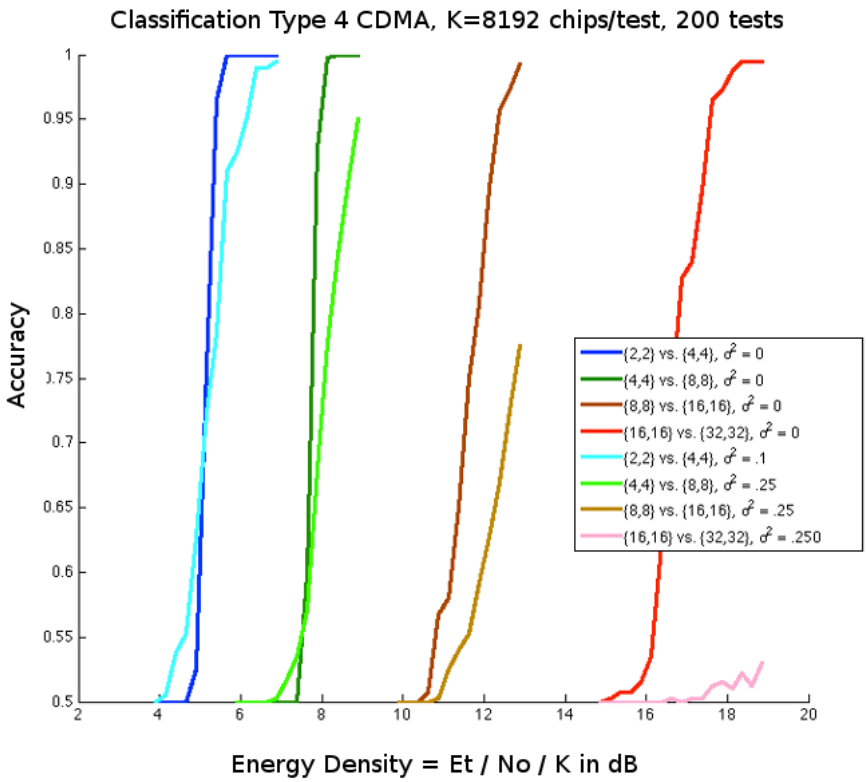

Fig. 40 Classification of $\{L,L\}$ versus $\{L/2,L/2\}$ for $log_2(L)$ between 2 and 7

7 Conclusion

The knowledge on modulation classification applied to CDMA was limited the use of cyclostationary features in neural network prior to this research. Prior state of the art was limited to single user CDMA signals and relied entirely on the cyclostationary properties to detect code lengths by using a feature based neural network with multiple layers. Such theory cannot be easily extended to the classification of multiuser CDMA signals.

A generalized likelihood approach was considered as a potential method for classifying CDMA. The main weakness of this approach is the costly estimation of the code matrices and the data vectors, especially if one consider the estimation of the parameters under the wrong hypothesis.

Th main contribution of the research presented in this document was able to provide a classification algorithm for multiuser CDMA using the average likelihood. Due to the model limitations, the algorithm cannot be applied to single user CDMA. However, by developing decision theoretic approach, we ensure that the classification of multiuser CDMA is optimal by virtue of the theory. Although the approximation of the average likelihood may be turn the classifier in a suboptimal method, it is the only available method at this moment.

The construction of the average likelihood function was based on the following assumptions:

- chip-synchronous,
- frame-asynchronous,
- arbitrary codes achieving lower bound TSC,
- unknown data vector,
- unknown permutations,
- fully-loaded,
- underloaded,
- balanced, and
- unbalanced energies.

The development of the average likelihood for CDMA was possible when key concepts such as the probability of the code matrix and key properties such as the average of the exponential functions were formulated. Averaging over a weight that depends on the TSC behaves as a filtering process where matrices with low TSC are preferred over all other matrices. By defining a precision parameter β, we are able to control the width of the probability of code matrices.

Taking the limit of β to infinity allows finding a simplified likelihood function in terms of the feature vectors, which are unsigned versions of the product between the code matrix and the data vector. This simplification eliminates many negligible terms that have no significant contribution to the average likelihood function. Furthermore, a Taylor simplification reduces the likelihood to a sum of exponentials. Also log-likelihood expression of a sum of exponentials can be simplified to the selection of the maximum value of dot product of the sorted versions of the feature vectors $\vec{a}^{\star}$ and the amplitude of the observation vectors $abs(\vec{r}_k)$.

The proposed method produced desirable results for classifying CDMA with code lengths between 2 and 128. The proposed approximation avoids multiple problems such as large number of unknowns, numerical overflow, and intractability issues derived from an increasingly large number of row permutations in the code matrix. Extending the classifier to higher code lengths is always possible; however, the implementation will require longer processing times for acquiring the feature vectors and computing the average likelihood.

The existence of a minimum energy-density to noise-density ratio is a characteristic feature of the newly developed CDMA classifiers. The classifier's performance below this limit approaches to 50 percent accuracy, with a quick transition to 100 percent accuracy. The slope of the accuracy versus density ratio varies depending on the number of samples per test. The curve is steeper with higher samples per test. In machine learning terms, this could be mean that the slope of a sigmoid function depends on the number of test samples. In information theoretic terms, this behavior could imply that the capacity of the channel has been exceeded to a point where it is impossible to identify the modulation type of the signal. These hypotheses would require verification.

These concepts can be extended complex types of CDMA. The entire complex model can be mapped into a pure real CDMA system that has binary antipodal symbols. Types 2 and 3 CDMA turn into two orthogonal sets of type 1 CDMA. Type 4 CDMA, under the assumption of perfectly orthogonal signals can be represented as a type 1 CDMA with twice the code length and twice the number of users. Because of this, the same average likelihood function can be applied to all types of CDMA with minor modifications.

CDMA signals may resemble Gaussian noise. However, framing the signal at the proper code length, removing the polarity and sorting the vector coefficients of the observation remove the random nature of CDMA and make the signal completely detectable. In the case of one user, the model becomes equivalent to an ordinary BPSK or QPSK signal. In CDMA cases where the number of users less or equal than half of the code length, the classification may fail when compared to a signal with half of the code length due to way that Hadamard codes of code length L can be constructed from codes with code lengths of $L/2$.

7.1 Future Work

There are many scenarios for detecting CDMA that can be explored. Several scenarios such as multipath or chip asynchronous CDMA can be studied and tested in the field. A more ambitious development includes the estimation of the code matrix using a average likelihood and the automation of a demodulation process for CDMA.

References

[1] S. Khamy, A. Helw and A. Mahdy. Classification of Multirate CDMA Signals using Compressed Cyclostationary Features. *Progress in Electromagnetics Research Proceedings*, 2012.

[2] O. Dobre, A. Abdi and W. Su. Survey of Automatic Modulation Classification Techniques: Classical Approaches and New Trends. *IET Communications*, 2007.

[3] S. Kim et Al. A Coherent Dual-Channel QPSK Modulation for CDMA Systems. *Vehicular Technology Conference*, 1996.

[4] C. Huang and A. Polydoros. Likelihood Methods for MPSK Modulation Classification. *IEEE Transactions on Communications*, 1995.

[5] A. Vega Irizarry and A. Fam. Average Likelihood Method for Classification of CDMA. *IEEE Military Communications Conference (MILCOM)*, 2014.

[6] A. Vega Irizarry and A. Fam. Average Likelihood Method for Classification of CDMA. *Submitted to: IEEE Military Communications Conference (MILCOM)*, 2016.

[7] E. Azzouz and A. Nandi. *Automatic Modulation Recognition of Communication Signals*. Springer, 1996.

[8] F. Lifdte. Computer Simulation of an Automatic Classification Procedure for Digitally Modulated Signals. *Signal Processing*, 6-4, 1984.

[9] S. Soliman and S. Hsue. Signal Classification Using Statistical Moments. *IEEE Transactions on Communications*, 40-5, 1992.

[10] H. L. V. Trees. *Detection, Estimation and Modulation Theory*. John Wiley & Sons, Inc., 2001.

[11] Y. Yao and V. Poor. Eavesdropping in the Synchronous CDMA Channel: An EM-Based Approach. *IEEE Transactions on Signal Processing*, 2001.

[12] A. Basu. *Statistical Inference, The Minimum Distance Approach*. Chapman & Hall/CRC, 2011.

[13] S. Verdu. *Multiuser Detection*. Cambridge University Press, 1988.

[14] P. Cotae. Distributive Iterative Method for Minimizing Generalized Total Square Correlation of CDMA Systems. *IEEE ITW*, 2003.

[15] P. Lancaster, M. Tesmenetsky. *Theory of Matrices with Applications*. Academic Press, 1985.

[16] E. Assmus, J. Key. *Hadamard Matrices and their Designs*, volume 330-1. 1992.

[17] P. Lancaster, H. Farahat. Norms on Direct Products and Tensor Products. *Mathematics of Computation*, 26, 1972.

[18] K. Horadam. *Hadamard Matrices and Their Applications*. Princeton University Press, 2007.

[19] Wolfram. *Wolfram's, A New Kind of Science*. Wolfram Media, Inc., 2002.

[20] A. Pallini. Estimating Probabilities from Invariant Permutations Distributions. *Journal of the Italian Statistical Society*, pages 77–91, 1994.

[21] A. K. Lenstra, H. W. Lenstra and L. Lovasz. Solving a System of Diophantine Equations with Lower and Upper Bonds on the Variables. 1998.

Appendix A Mathematica Code for Computing the Average Likelihood Function

```
(* CDMA LIKELIHOOD FUNCTION *)
(* 1. Code Length and Number of Users *)
L = 2; U = L;

(* 2. Code Matrix Defined *)
H = Array[Subscript[c, #1 - 1, #2 - 1] &, {L, L}];

(* 3. Data Vector Defined *)
B = Array[Subscript[b, # - 1, K]&, {L, 1}];

(* 4. Observation Vector Defined *)
R = Array[Subscript[r, #1 - 1, K]&, {L, 1}];

(* 5. Correlation Term *)
X = R.H.B;
X = Expand[X[[1, 1]]]

(* 6. TX Signal and Energy *)
Y = (H.B);
ET = (Y.Y);
ET = Expand[ET];
ET = ET/.v_^2->1;
ET = ET[[1, 1]]

 (* 7. Total Squared Correlation *)
H2 = H.H;
V2 = 0;
For[i1 = 1, i1 ≤ L, i1++,
For[i2 = 1, i2 ≤ L, i2++,
If[i1 ≠ i2,
V2 = V2 + H2[[i1, i2]]^2
]
];
];
w1 = Expand[V2]/.{_^2 → 1}/.{_^4 → 1};
```

```
FML = w1

(* 8. Normalization Constant of Probability of the CDMA Code *)
Z = Join[Flatten[H]];
If[L ≠ 1,
W1 = Exp[-β * w1];
For[i = 1, i ≤ Length[Z], i++,
If[!NumberQ[Z[[i]]],
rl = {Z[[i]] → z};
W1 = W1/.rl;
W1 = Sum[W1, {z, {-1, 1}}]
];
],
W1 = Exp[-β];


(* 9. Conditional likelihood *)
CL = Exp[Sqrt[2γ] * X - γ * ET - β * FML]

(*10. Averaging the Conditional Likelihood*)
AL = CL;
Z = Join[Sqrt[1] * Flatten[H], Flatten[B]]
For[i = 1, i ≤ Length[Z], i++,
If[!NumberQ[Z[[i]]],
rl = {Z[[i]] → z};
AL = AL/.rl;
If [NumberQ [Z[[i]]/. {c_,_ → 0}] ,
AL = Sum[AL, {z, {-1, 1}}],
AL = 1/2 * Sum[AL, {z, {-1, 1}}]
];
];
];


(* 11. Expansion of the Exponents *)
ALS = AL/.Exp[aaa_] :→ Exp[Expand[aaa]];
rule1 = {Exp [Plus[ex1__] + ex2__ * r_{s_,K}] :→ Exp[Plus[ex1]] * Subscript[P, s]^Times[ex2]} ;
ALS = ALS//.rule1;
rule2 = {Exp [ex2__ * r_{s_,K}] :→ Subscript[P, s]^Times[ex2]} ;
ALS = ALS//.rule2;
rule3 = {Subscript[P, s_]^Times[ex2__] :→ Cosh[Times[ex2] * Subscript[r, s, K]]
```

```
+Sinh[Times[ex2] * Subscript[r, s, K]]};
ALS = ALS//.rule3;
ALS = Expand[ALS]

(* 12.GENERAL FORM OF THE 2x2 CDMA Likelihood Function *)
ALSS = Expand[ALS/W1]
```

$$\frac{4e^{-8\beta}}{8+8e^{-8\beta}} + \frac{4e^{-4\gamma}\text{Cosh}\left[2\sqrt{2}\sqrt{\gamma}r_{0,K}\right]}{8+8e^{-8\beta}} + \frac{4e^{-4\gamma}\text{Cosh}\left[2\sqrt{2}\sqrt{\gamma}r_{1,K}\right]}{8+8e^{-8\beta}} + \frac{4e^{-8\beta-8\gamma}\text{Cosh}\left[2\sqrt{2}\sqrt{\gamma}r_{0,K}\right]\text{Cosh}\left[2\sqrt{2}\sqrt{\gamma}r_{1,K}\right]}{8+8e^{-8\beta}}$$

```
(* 13. Precision to Infinity *)
LF1 = ALSS/.{β → ∞}
```

$$\frac{1}{2}e^{-4\gamma}\text{Cosh}\left[2\sqrt{2}\sqrt{\gamma}r_{0,K}\right] + \frac{1}{2}e^{-4\gamma}\text{Cosh}\left[2\sqrt{2}\sqrt{\gamma}r_{1,K}\right]$$

Appendix B Computation of the CDMA Average Likelihood Coefficients

```
%-------------------------------------------------------------------------
% Calculation of the numerator of the alpha parameters using formula.
% A. Vega, Last Modified: October 8, 2013
% inputs: a = vector vector, U = number of users
% output: matrix = [ TSC, occurrences ]
% formatted output = occurrences * e^(beta*TSC)
%-------------------------------------------------------------------------
function output = alphaParam(a,U)
% Ensure positive coefficients
a = abs(a);
% Code length of the spreading matrix
L = length(s);
% Calculate possible column permutations such that sum(cols) >= 0
colPerm = [];
colPerm(1).x(1,:) = ones(1,U);
for k1 = 1:floor(U/2)
    x = ones(1,U);
    k2 = k1;
while( k2 > 0 )
x(1,k2) = -1;
        k2 = k2 - 1;
end
colPerm(k1+1).x = unique(perms(x),'rows')
end
% Display possible row vectors
if( 0 )
for k3 = 1:size(a(:),1)
squeeze(a(k3).x(:,:))
end
end
% Construct matrices using row vectors
group = abs(U-a)/2+1;
elems = zeros(1,U);
for k1 = 1:L
elems(1,k1) = size(colPerm(group(k1)).x,1);
end
% This algorithm is inefficient for L>6
tic
stack = []; mtx = zeros(L,U);
start = ones(1,L);
while( all(start<=elems) )
        % Generate Matrix
for k1 = 1:L
mtx(k1,:) = colPerm(group(k1)).x(start(1,k1),:);
```

```
end
        % Calculate Total Square Correlation
tau = 0;
        h2 = mtx'*mtx;
for k1 = 1:L
for k2 = 1:L
if( k1 ~= k2 )
tau = tau + h2(k1,k2)^2;
end
end
end
        % Store value
stack = [stack,tau];
    %end
    % This algorithm explores all possible combinations of a row in a matrix
    % The sum of the row >= 0
    k2 = 1;
while( k2 <= L )
        c = start(1,k2) + 1;
if( c >elems(1,k2) )
if( k2 < L )
start(1,k2) = 1;
                k2 = k2 + 1;
else
start(1,k2) = c;
                k2 = k2 + 1;
end
else
start(1,k2) = c;
            k2 = L+1;
end
end
end
toc
% Count all unique values
vals = unique(stack','rows')';
svals = zeros(size(vals));
for k1 = 1:length(vals)
svals(1,k1) = sum(stack==vals(k1));
end
% Display results
[vals',svals'*2^L/2^(sum(s==0))]
```

Appendix C Generation of Type 1 CDMA Signal

```
%--------------------------------------------------------------------------
% Matlab Code Snippet
% Generation of CDMA signals Type 1
% Inputs:
%   L = code length, power of 2, L > 1
%   U = active users
%   chips = signal length in chips, multiple of L
%   delta_var = variance of the amplitude
% Outputs:
%   s = CDMA signal samples
%--------------------------------------------------------------------------
% Generate Code Matrix Type 1
h = hadamard(L)/sqrt(L);
% Implement Row and Column Permutation
r = randperm(L);
c = randperm(L);
h = h(r,c);
h = h(:,1:U);
% Generate Data Vector
symbols = randsrc(U,chips/L);
% GenerateAmplitudevariation
if( deltavar ~= 0 )
    AA = diag(delta_var/U*randn(U,1));
else
    AA = zeros(U);
end
%  CDMA signal
s = h*(eye(U)+AA)*symbols;
%  reformat as a stream
 s = reshape(s,[size(s,1)*size(s,2),1]);
```

Appendix D Feature Extraction Algorithm

```
%-------------------------------------------------------------------------
% Matlab Code Snippet
% Feature Vector Extraction
% Inputs:
%   L = code length
%   N = size of the histogram
% Outputs:
%   aa = feature vector array
%   alpha = alpha coefficients
%-------------------------------------------------------------------------
function[ aa, alpha ] = surveyfeat(L,U,N)
hL = hadamard(L);
fv = zeros(L,0);
alpha= zeros(1,0);
for k = 1:N
    % Compute CDMA
    if( U == L )
        s = hL*randsrc(L,1);
    else
        u = randperm(L);
        u = u(1:U);
        s = hL(:,u)*randsrc(U,1);
    end
    % Sort entries by absolute value
    ss = sort(abs(s));
    % Enter first element
    if(isempty(fv) )
        fv = [ ss ];
        alpha= [ 1 ];
    else
        % search closest match
        done = 0; rindex = 1; cindex = 1;
        while( done == 0 )
            if(fv(rindex,cindex) == ss(rindex,1) )
                % Smart sorting to save time
                if(rindex == L )
                    % found match, go to next column
                    F(1,cindex) = F(1,cindex) + 1;
                    done = 1;
                elseif(rindex< L )
                    % go next row and compare
                    rindex = rindex + 1;
                else
                    % reached end of row
                    if(cindex< size(fv,2) )
                    % go to next column
```

```
                        cindex = cindex + 1;
                        rindex = 1;
                    else
                        % reached last entry, append to end
                        fv = [ fv, ss ];
                        alpha= [ F, 1 ];
                        done = 1;
                    end
                end
            elseif(fv(rindex,cindex) >ss(rindex,1) )
                % found bigger entry
                if(cindex == 1 )
                    % append at beginning
                    fv = [ ss, fv ];
                    alpha= [ 1, alpha];
                    done = 1;
                else
                    % append in before the bigger entry
                    fv = [ fv(:,1:(cindex-1)), ss, fv(:,cindex:end) ];
                    alpha= [ F(1,1:(cindex-1)),  1, F(1,cindex:end) ];
                    done = 1;
                end
            else
                % found smaller vector
                if(cindex == size(fv,2) )
                    % no more vectors found, append at end
                    fv = [ fv, ss ];
                    alpha= [ F, 1 ];
                    done = 1;
                else
                    % more vectors available , increase indexes
                    cindex = cindex + 1;
                    rindex = 1;

                end
            end
        end
    end
end
aa = fv;
alpha = alpha/sum(alpha);
```

Appendix E Calculation of the Average Likelihood

```
%-------------------------------------------------------------------------
% Computation of the log-likelihood per signal
% Inputs:
%   rx = unformatted observation vector with proper length
%   gam = total energy to noise ratio
%   L = code length
%   U = active users
%   delta_var = variance of the amplitude
% Outputs:
%   llkh = total log-likelihood
%-------------------------------------------------------------------------
% Load feature vectors and alpha coefficients
[aa,alpha] = loadfeatures(L,U);
% Number of possible delays, frame-asynchronous
delayavg = L;
% Number of samples assuming multiple of L
chips = length(rx);
% factor for amplitude variations
gammac = gam/chips/U;
% log-likelihood per delay per sample
dist = zeros(chips/L,delayavg);
% factor for amplitude variations
logchi = unbalanced_factor(delta_var,L,U,gammac);
% Calculate likelihood for different delays
for delay = 0:delayavg-1
    % implement a given delay
    r_delay = circshift(rx,delay);
    % For each sample
        cnt = 1;
    for k1 = 1:L:lim
        % Format as a vector assuming code length = L
        r_vec = r_delay(k1:k1+L-1,1);
    dist2 = loglikelihood(r_vec,aa,gammac) +log(alpha) +logchi;
        % feature vector, log sum of exponents
    dist(cnt,k2+1) = log_sum_exp( dist2 );
        cnt = cnt + 1;
    end
end
% Compute likelihood per delay
llkh_delay = sum(dist);
% Average over delays
llkh_symbols = log_sum_exp(llkh_delay)-log(delayavg);
% Sum of log likelihood of all independent symbols
llkh = sum(lkh_symbols);
```

```
%-------------------------------------------------------------------------
% Matlab Code Snippet
% Computation of the distance term of the likelihood per feature vector
% Inputs:
%   r = unformatted observation vector with length L
%   gammac = chip level SNR
%   aa = collection of feature vectors
% Outputs:
%   llkh = log-likelihood per sample
%-------------------------------------------------------------------------
functiondist = loglikelihood( r, aa, gammac )
    % assuming that feature vector is already sorted
    % aa = sort(r);
    rs = sort(abs(real(r)));
    dist = zeros(1,size(aa,2));
    for k=1:size(aa,2)
        dist(1,k) = -1/2*sum((r - sqrt(2*gammac)*aa(:,k)).^2);
    end
```

Appendix F Error Probability Matrix Code

```
%-------------------------------------------------------------------------
% Matlab Code Snippet
% Computation of the error probability matrix
% Inputs:
%   HT = array of hypothesis, column-wise pair { L, U }
%   Et = total energy, No = noise power
%   chips = chip length of the signal
%   L = true code length, U = true active users
%   delta_var = variance of the amplitude
%   ntests = number of tests
% Outputs:
%   confmatrix = error probability matrix
%-------------------------------------------------------------------------
% Matlab parallel for loop goes here
parfor tests = 1:ntest
    % Generate a stream of CDMA signals
    s = randBaseBandSignal('type1',chips,L,U,delta_var);
    % Transmitted signal (baseband)
    Es = Et/(chips/L)/U;
    tx = sqrt(Es)*s;
    % choose a random delay between 0 and L-1 chips
    delay = randi([0,L-1],1,1);
    tx = circshift(tx,delay);
    % AWGN channel
    n = noise(NoReal, length(tx) );
    rx = tx + n;
    % For all the hypothesis test in array HT
    for k2 = 1:length(HT)
        Ltest = HT(1,k2);
        Utest = HT(2,k2);
        gammac = Et/NoReal;
        rr = real(rx)/sqrt(NoReal/2);
        % Store the true class and the loglikelihood in a structure
        lkhTest(tests).class = k1;
        lkhTest(tests).lkh(k2) = cdmaLogLikelihood(rr,gamma,Ltest,Utest,delta_var);
end
end
% Compute Error Probability Matrix using MAP, in this case the same as LRT
for test = 1:ntest
    trueClass = lkhTest(test).c;
    [~,decision] = max(lkhTest(test).lkh);
    confmatrix(trueClass,decision) = confmatrix(trueClass,decision)+1;
end
```

Nomenclature

Acronyms

ALRT	.....	Average Likelihood Ratio Test
AGWN	.....	Additive White Gaussian Noise
BPSK	.....	Binary Phase Shift Keying
CDMA	.....	Code Division Multiple Access
CON	.....	Complete Orthonormal Set
dB	.....	Decibels
EM	.....	Expectation Maximization
FFT	.....	Fast Fourier Transform
GLRT	.....	Generalized Likelihood Ratio Test
HLRT	.....	Hybrid Likelihood Ratio Test
LRT	.....	Likelihood Ratio Test
MMSE	.....	Minimum Mean Squared Error
MPI	.....	Multiple Programming Interface
MSC	.....	Maximum Squared Correlation
PN	.....	Pseudo noise
QAM	.....	Quadrature Amplitude Modulation
QLRT	.....	Quasi-Average Likelihood Ratio Test
QPSK	.....	Quaternary Phase Shift Keying
ROC	.....	Receiver's Operating Characteristic
SNR	.....	Signal-to-Noise Ratio
TSC	.....	Total Squared Correlation

Symbols

$\{\}^L$	.....	set of L elements or column vector
$\{\}^{L\times U}$	.....	matrix of L rows and U columns
$\mathbb{R}^L$	.....	real domain of dimension L
$\mathbb{C}$	.....	complex domain
$\mathbb{Z}^*$	.....	domain of real non-negative integers
$Re\{\cdot\}$	.....	real part operator
$Im\{\cdot\}$	.....	imaginary part operator
$\otimes$	.....	Kronecker product
$\langle\cdot,\cdot\rangle$	.....	inner product
$\|\cdot\|_F$	.....	Frobenius norm
$\|\cdot\|$	.....	vector norm
$\delta(t)$	.....	Dirac delta function
$\delta_{i,j}$	.....	Kronecker delta
$(\cdot)^H$	.....	Hermitian operator
$(\cdot)^T$	.....	transpose operator
$abs(\cdot)$	.....	absolute value
!	.....	factorial symbol
$\vec{r}_R$	.....	real part of a complex column vector
$\vec{r}_I$	.....	imaginary part of a complex column vector
$(\vec{r}_R)_i$	.....	i^{th} component of a real column vector
$\vec{0}$	.....	null column vector
$\vec{1}$	.....	column vector of 1 entries
$\mathbf{0}$	.....	null matrix
I	.....	identity matrix
H	.....	Hadamard matrix
$I_{L\times L}$	.....	$L\times L$ identity matrix
Q_R	.....	real matrix
Q_I	.....	imaginary part of a matrix
$c_{i,j}$	.....	matrix element, row i and column j
$\vec{a}$	.....	CDMA amplitude vector
$\vec{g}$	.....	CDMA sign vector
$diag(\vec{g})$	.....	diagonal matrix construction based on $\vec{g}$
β	.....	Precision parameter
E	.....	Energy per symbol
E_T	.....	Total energy

$N_0/2$	.....	Noise power density
γ	.....	symbol-level SNR
γ_c	.....	chip-level SNR
$\vec{b}$	.....	data vector
C, Q	.....	code matrices
$s(t)$	.....	transmit signal
$y(t), \vec{y}$	.....	received signal
$r(t), \vec{r}$	.....	observation
$n(t), \vec{n}$	.....	noise signal
$\mathbb{E}\{\cdot\}$	.....	expectation operator
$\mathscr{R}$	.....	Autocorrelation matrix
$\mathscr{R}(\Delta t)$	.....	Autocorrelation function
$\mathscr{N}$	.....	Normal distribution
$\mathscr{H}$	.....	hypothesis
R	.....	Observation
L	.....	code length
U	.....	active users
K	.....	number of chips
$\lambda(\vec{r}\|\mathscr{H})$	.....	average likelihood
$\lambda(\vec{r}\|\mathscr{H}, C, \vec{b})$	.....	conditional likelihood