

國立交通大學
National Chiao Tung University

出國報告（出國類別：☐ A 類、考察訪問
☒ B 類、出國短期研究
☐ C 類、國際會議

題目：Code Design for Combined Channel
Estimation and Error Correct under MIMO
Block Fading Channels

服務機關：電信工程研究所

姓名職稱：吳佳龍 博士後研究員

前往國家：瑞典 斯德哥爾摩 皇家理工學院

出國期間：民國 99/4/1~6/21

報告日期：民國 98/7/8

撰 寫 人	審 核 人	初 閱	複 閱
吳佳龍		陳伯寧	教授兼研發長 李鎮宜

備註：出國報告書審核程序如下

- 一、初閱：各學院教師 A、B、C 類及其他行政單位 A 類由單位主管，研究生由指導教授；中心計畫及學群 A、B、C 類由各中心計畫主持人。
- 二、複閱：經費所屬之一級單位；中心計畫及學群 A、B、C 類由頂尖計畫執行長。

一、摘要

本次赴瑞典短期研究是與瑞典皇家理工學院 Skoglund 教授合作，討論如何在接收機(receiver)僅具有部分通道狀態資訊(Partial CSI)的情況下，設計好的通道編碼。我們這項合作主要分為兩個部份來進行，第一部分是推導在僅有部分通道狀態資訊情況下之最大概度解碼量度(maximum-likelihood decoding metric)，第二部份則是利用最大概度解碼量度進一步推得成對錯誤機率(pair-wise error probability)的聯合上界(union bound)，並以此上界做為找最佳碼的準則。

關於第一部分，本研究所考慮的通道是多路徑區塊時變衰減通道(multi-path block fading channels)。在此通道模型下，通道參數在區塊時間內對接收端與傳送端皆是未知常數(unknown constant)，我們發現：最大概度解碼量度可以分解為兩個部份，一部分是傳統同調檢測(coherent detection)的解碼量度，另一部分則是非同調檢測的解碼量度。經由模擬顯示，與傳統僅單純利用同調檢測的解碼量度所得的效能比較，以新的最大概度解碼量度所得的效能最少可以有 0.5 dB 的改善。

其次，在第二部分，使用最大概度解碼量度所得到的成對錯誤率上限作為找尋最佳碼的準則，再利用隨機搜尋演算法 Simulated Annealing 來搜索最佳編碼方法。所得的編碼將會是針對多路徑衰減通道下，接收機僅具有部分通道狀態資訊的最佳編碼設計。利用同樣的編碼設計技術，可以進而設計多輸入多輸出(multiple-input multiple-output)系統的最佳編碼，達到本研究考慮在擁有部分通道資訊的情況下，結合通道估計的多輸入多輸出的通道編碼設計。

二、目次

一、摘要.....	2
二、目次.....	3
三、本文.....	4
(一) 目的.....	4
(二) 過程.....	5
1. 斯德哥爾摩生活.....	5
2. 研究合作.....	6
(三) 建議及心得.....	7
1. 建議.....	7
2. 心得.....	7
四、附錄.....	7

三、本文

(一) 目的

瑞典皇家理工學院通信理論(Communication Theory)實驗室是由 Skoglund 教授所主持，在大斯德哥爾摩地區，Skoglund 教授的實驗室是唯一專注於通信理論的大學研究室，其團隊尤其專精於非同調通訊(Non-coherent Communications)技術與合作式分集(Cooperation Diversity)的分析與使用，近年來 Skoglund 教授所領導的團隊更開始涉足 Network Coding 等相關前瞻領域。本次赴瑞典皇家理工學院短期訪問的主要目的，就是要與 Skoglund 教授學習非同調通訊的相關技術並建立合作關係。此外，我們也相信透過 Skoglund 教授發展的編碼設計方法可以成為未來非同調通訊研究上有效之工具。在這三個月與 Skoglund 教授學習的項目，可以分為兩大項，

(1) 接收機具有部分通道狀態資訊之理論探討：目前通訊系統約略可以分為兩大類，同調通訊與非同調通信。同調通訊是指接收機利用傳送與接收雙方都已知的訓練序列(Training Sequence)來估計通道參數，在假設該通道參數是完美估計(Perfect Estimation)並以此估計結果用於通道解碼；而非同調通訊則完全相反，在接收端完全不需要通道估計裝置，一般會直接使用接收訊號進行完全搜尋(Exhausted check)解碼，所以解碼複雜度非常的高。其實，真正的系統情況，應該是介於上述兩種極端的情況，此次拜訪瑞典皇家理工學院，正是想學習該實驗室用於非同調通訊系統分析與設計的技巧，由於結合通道估計與錯誤更正的通訊架構之效能 再相關的文獻中已經確認其確實更優於傳統將通道估計單元獨立設計的方式，如何考量現實通訊系統設計的限制來進一步改善其效能，也是本次短期研究與 Skoglund 教授要交換的研究心得。

(2) 在 relay 通道上的分散式編碼：近來，Skoglund 教授研究的方向多是朝向合作式分集的通訊系統設計，其中最被關注的有 Network Coding 與分散式編碼(Distributed Coding)。以分散式編碼為例，系統設計的方法主要是聚焦在如何利用其他的 relay nodes 來幫助效能的提昇。Skoglund 教授已經發表相當數量的論文在 IEEE Transaction on Information Theory，皆為非常具有原創性的研究成果。此次的短期訪問，也對這項研究進行討論，希望未來能對這項題目進一步的合作。

Skoglund 教授的團隊與各國都有合作，在訪問期間，除了有來自挪威的交換研究生外，還有來自德國與美國新墨西哥大學的訪問教授。透過這些來訪的學者提供的學術演說，也讓我可以了解目前通訊理論研究的最近進展，並與這些講者討教，甚至也談及未來進行合作的可能。

在瑞典三個月的期間，我也利用機會至斯德哥爾摩大學圖書館收集研究資料，主要是關於電腦語言學相關的資料，以作為未來新研究題目的參考。斯德哥爾摩大學擁有豐富的社會與人文館藏，電腦語言學作為語言學的一個分支，也是斯德哥爾摩大學所注重。近年來，由於 Google 等網路搜尋技術的發展，電腦語言學技術也被廣泛的使用在資料挖礦(Data Mining)和網頁搜尋等應用。

(二) 過程

1. 斯德哥爾摩生活

這是本人第二次前往斯德哥爾摩，前一次是取得國科會千里馬計畫的補助，赴瑞典進行為期一年的博士論文訪問研究。這次的短期訪問，也有幸能舊地重遊，再次訪問 Skoglund 教授。斯德哥爾摩的租屋市場是求過於供，虧有安娜(Anna Löfstrand)的幫助，才在距離斯德哥爾摩市區約半小時車程的小鎮 Vallentuna 租到低價舒適的房間，使我可以很快的讓生活步上軌道，馬上可以開始合作研究的事宜。Skoglund 教授也總是在百忙之中固定抽出時間與我討論研究上的問題，才能讓我此行能圓滿的結束。也由於是第二次訪問該研究團隊，許多前一次認識的成員都還在該實驗室，所以與大家相處起來更是融洽。

Skoglund 教授個性沉默和善，由於目前兼任皇家理工學院的副院長，必須行政與研究兩邊忙碌，往往是每天最晚離開研究室的人。在這三個月短期參訪期間，也定期參加該研究團隊的論文討論，了解該團隊目前著手的研究題目與進行進度。而 Skoglund 教授實驗室是一個相當龐大的研究團隊，除了 Skoglund 教授之外，還有 Lar 教授以及兩位助理教授，並聘有五位博士後研究員，以及 13 位的博士研究生，目前也持續招聘新的博士後研究員及博士生。也因為如此，該團隊的研究包括了非常豐富的通訊理論主題，方向非常多元，實驗室的學生與研究員可以依照自己的專長與興趣，選擇研究的題目。此外，Skoglund 教授跟我也利用這次機會，對目前兩邊實驗室未來可能的合作題目交換意見。



(圖一) 瑞典 Skansen 室外博物館一景。可以遠眺斯德哥爾摩市區

而在生活上，我也利用假日，了解斯德哥爾摩的生活環境，雖然是第二次到訪斯

德哥爾摩，還是有許多瑞典特色的文化傳統讓我覺得新奇。例如 Skansen 是瑞典一個著名的戶外博物館，裡面有瑞典各地收集而來的歷史性建築物，例如北方原住民的帳篷還有從前貴族的度假別墅。在傳統節慶時也會舉辦各式各樣的清除活動。另外一個讓我印象深刻是瑞典的王儲 Victoria 公主結婚的日子。她與他的夫婿 Daniel 在斯德哥爾摩舉辦了一場豪華的婚禮。婚禮老城的大教堂裡舉行，來自世界各國的王公貴族都到場觀禮，包括有挪威國王與荷蘭女王以及挪威、丹麥、日本和荷蘭等國的王子與公主。典禮結束之後新娘與新郎搭上馬車繞行斯德哥爾摩市區接受人民的歡呼，氣氛熱烈讓我印象深刻。

另外，許多瑞典人待人接物的方式也是值得學習。瑞典的環境整潔，而且很容易親近大自然，可以感覺出瑞典人對環境保護的用心。瑞典都市的空間不會擁擠。以斯德哥爾摩 200 萬人口來說都市的規模與台北市相當，但是行走在都市的當中不會有人車爭道的情形。行人都可以悠閒的漫步在街道上，與台灣的各大都市那種車水馬龍的景象截然不同。瑞典的居住環境確實讓人羨慕。

2. 研究合作

本次赴瑞典皇家理工學院與 Skoglund 教授的短期合作，是探討結合通道估計與錯誤更正的通道在多輸入多輸出系統環境下的編碼設計。事實上，關於這項題目日本實驗室早先已經獲得幾項具體的突破：

- I. 首先，對於單一輸入單一輸出(Single-Input Single-Output)的通道，本實驗室已經有成果發表在 2009 年 9 月的 IEEE Transaction on Information Theory 期刊。在這項成果中，我們提出一個系統化的建碼方法，以及一個加速的循序解碼演算法。
- II. 對於多重輸入多重輸出(Multiple-Input Multiple -Output)的通道，Skoglund 教授與本實驗室共同發表結果在韓國首爾舉辦的 ISIT2009 研討會。在這篇研討會論文中，我們針對傳送天線數為二的情況下，證明只要信號矩陣滿足一個特定的結構，就可以避免成對錯誤率有最差的情況，進一步確保系統的效能。

基於上述的結果，本次前往研究之目的主要討論當接收機有部分通道狀態資訊下，並且遭遇區塊衰減通道，結合通道估量與錯誤更正的系統設計。我們首先由單一輸入單一輸出的系統出發，而後進一步討論多重輸入多重輸出的系統設計。我們這項合作主要分為兩個部份來進行，第一部分是推導在僅有部分通道狀態資訊情況下之最大概度解碼量度(maximum-likelihood decoding metric)，第二部份則是利用最大概度解碼量度進一步推得成對錯誤機率(pair-wise error probability)的聯合上界(union bound)，並以此上界做為找最佳碼的準則。

關於第一部分，本研究所考慮的通道是多路徑區塊時變衰減通道(multi-path block fading channels)。在此通道模型下，通道參數在區塊時間內對接收端與傳送端皆是未知常數(unknown constant)，我們發現：最大概度解碼量度可以分解為兩個部份，一部分是傳統同調檢測(coherent detection)的解碼量度，另一部分則是非同調檢測的非同調檢測的解碼量度。

經由模擬顯示，與傳統僅單純利用同調檢測的解碼量度所得的效能比較，以新的最大概度解碼量度所得的效能最少可以有 0.5 dB 的改善。其次，在第二部分，使用最大概度解碼量度所得到的成對錯誤率上限作為找尋最佳碼的準則，再利用隨機搜尋演算法 Simulated Annealing 來搜索最佳編碼方法。所得的編碼將會是針對多路徑衰減通道下，接收機僅具有部分通道狀態資訊的最佳編碼設計。利用同樣的編碼設計技術，可以進而設計多輸入多輸出(multiple-input multiple-output)系統的最佳編碼，達到本研究考慮在擁有部分通道資訊的情況下，結合通道估計的多輸入多輸出的通道編碼設計。關於這三個月的研究成果，Skoglund 教授與我們實驗室正在整理，預計將會投稿至明年在日本京都舉辦的 2011 International Conference of Communications。

(三) 建議及心得

這三個月的建議與心得可以用以下幾點來總結：

1. 建議

- ✓ 除了英語之外，應加強第二外語能力。在瑞典，學生一般會在中學時選擇第二外語，比較受歡迎的選擇有德語、法語和西班牙語。雖然一般的學術訪問，英語就已經足夠應付溝通的需求，然而若要建立私人的情誼，若能知道或使用彼此間的語言，更能幫助交流的進行，否則似乎只能僅於官方的討論。加上對彼此語言的了解，在學術交流中也是一種尊重的表現，學習當地的語言是有必要的。
- ✓ 鼓勵台灣與瑞典之間的互訪交流。目前台灣與瑞典之間的學術交流似乎沒有台灣與美國間的交流熱切，原因或許是台灣的研究一般是以美國馬首是瞻。然而，以瑞典在某些特定領域前瞻研究，確實有許多值得合作之處，例如瑞典的卡車工業，如 SCANIC 公司，通信產業，例如 Ericson 公司等等，有許多科技與技術的競爭力是與美國並駕齊驅。

2. 心得

- ✓ 瑞典上班制度與台灣有明顯的不同。在台灣如果在私人企業工作往往需要加班到很晚，甚至在某些公司工程師在每個週末都需要到公司趕工。但是在瑞典這是很不可思議的事情。瑞典人重視休閒生活，強調工作與私人生活能保持平衡。例如一般勞工一般每年都有四個星期的有薪假。即使員工們享有這麼多的假日也不影響瑞典企業的競爭力。希望台灣未來能夠往這個方向發展。
- ✓ 培養外語能力，尤其是英語。這次短期訪問的國家為瑞典，瑞典國民大多數能講流利的英語，跟德國或日本英語僅能在大學校園之內有用的情況很不一樣。在大學裡由於實驗室的成員來自數個不同的國家，英語更是唯一的選擇。

四、附錄

附上目前的研究成果，將會整理後投稿至於日本京都舉行的 ICC 2011。

Joint Design of Pilots and Code for Partial CSI at the Receiver

July 8, 2010

1 Motivation

In packet-switched systems, frame-synch and channel estimation is often done using the same training sequence or pilot signals. The system model in [2, 4, 5] assumes that the receiver can synchronize the codeword margins exactly. Without the information of the codeword margins, decoding of the joint channel estimation and error correction codes would become technically infeasible. So, at this stage, the codes may only be able to perform in an initial-sync, or circuit-switched, or TDD-based system environment.

In this work, we consider the practical packet-switched systems with training sequence or pilots for frame-synch and channel estimation. Then, the available channel state information (CSI) in receivers would not be perfect because of additive noise. The code and decoder for the receivers with partial CSI would be discussed here.

Several notations appear in this work: $\det[\mathbb{X}]$ is the determinant of a matrix \mathbb{X} and $\text{tr}(\mathbb{X})$ is the trace of the matrix. Notation \mathbb{X}^T and \mathbb{X}^H denote transpose and Hermitian transpose of \mathbb{X} , respectively. Another superscript \mathbb{X}^* means the conjugate matrix of \mathbb{X} , i.e., all element of \mathbb{X}^* is taking the complex conjugate of corresponding element of \mathbb{X} . The maximum and minimum eigenvalue of \mathbb{X} are represented as $\lambda_{\max}(\mathbb{X})$ and $\lambda_{\min}(\mathbb{X})$, respectively. In addition, for a vector \mathbf{x} and a matrix \mathbb{X} , we define $\|\mathbf{x}\| \triangleq \sqrt{\mathbf{x}^H \mathbf{x}}$ and $\|\mathbb{X}\| \triangleq \sqrt{\text{tr}(\mathbb{X}^H \mathbb{X})}$, respectively. The notation of a random vector \mathbf{x} that has circular symmetric complex Gaussian distribution with \mathbf{m} mean and covariance \mathbb{S} is $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbb{S})$, and $\mathcal{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$ which is the area under the tail of a real-valued zero-mean Gaussian probability density function with variance 1.

2 System Model

In packet-switched systems, a signal $\mathbf{b} = [b_1, \dots, b_N]^T$ is transmitted over a frequency-selective block fading (specifically, quasi-static fading) channel of memory order $P - 1$. For $1 \leq i \leq N$, we limit that b_i is belong to a constant-amplitude 2^M -PSK modulation, i.e., $|b_i|^2 = 1$, where $M > 0$

and is a integer. By letting the codeword matrix be

$$\mathbb{B} \triangleq \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ \vdots & b_1 & \ddots & \vdots \\ b_N & \vdots & \ddots & 0 \\ 0 & b_N & \ddots & b_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & b_N \end{bmatrix}_{L \times P},$$

where $L = N + P - 1$ and with power constraint $\text{tr}(\mathbb{B}^H \mathbb{B}) = P \cdot N$, the received signal \mathbf{y} is

$$\mathbf{y} = \mathbb{B} \mathbf{h} + \mathbf{n} \quad (1)$$

where \mathbf{n} is zero-mean circular symmetric complex Gaussian with correlation matrix $\sigma_n^2 \mathbb{I}$, and \mathbf{h} is a unknown constant during L symbol intervals and independently changes between intervals.

After using the pilot signal in a part of \mathbb{B} , we assume that the perfect frame-synch and knowledge of the multipath parameter P are available in the system model. We also assume both transmitter and receiver know nothing about the channel coefficients \mathbf{h} . Finally, adequate guard periods are between consecutive encoding blocks in order to guarantee zero interblock interference.

2.1 On the arrangement of training sequence

For frame-synch and channel estimation, a transmitted signal \mathbb{B} consists of

$$\mathbb{B} = \begin{bmatrix} \mathbb{B}_P \\ \mathbb{B}_D \end{bmatrix},$$

where

$$\mathbb{B}_P \triangleq \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ b_2 & b_1 & \ddots & \vdots \\ b_3 & b_2 & \ddots & 0 \\ b_4 & b_3 & \ddots & b_1 \\ \vdots & \ddots & \ddots & \vdots \\ b_T & b_{T-1} & \cdots & b_{T-P+1} \end{bmatrix}_{T \times P} \quad \text{and} \quad \mathbb{B}_D \triangleq \begin{bmatrix} b_{T+1} & b_T & \cdots & b_{T-P+2} \\ \vdots & b_{T+1} & \ddots & b_{T-P+3} \\ b_N & \vdots & \ddots & \vdots \\ 0 & b_N & \ddots & b_T \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & b_N \end{bmatrix}_{(L-T) \times P}$$

are a pilot signal that is known to both transmitter and receiver and the signal transmitted information, respectively. Base on the system model (1), the received signals of \mathbb{B}_P and \mathbb{B}_D would be

$$\mathbf{y}_P = \mathbb{B}_P \mathbf{h} + \mathbf{n}_P$$

and

$$\mathbf{y}_D = \mathbb{B}_D \mathbf{h} + \mathbf{n}_D,$$

where $\mathbf{n}^H = [\mathbf{n}_P^H \quad \mathbf{n}_D^H]$. Due to additive white Gaussian noise and unknown constant \mathbf{h} , ML estimator of \mathbf{h} should be least square estimation

$$\hat{\mathbf{h}} = (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{y}_P$$

under the condition $T \geq P$. Let least-square error $\tilde{\mathbf{h}} = \mathbf{h} - \hat{\mathbf{h}}$, we can know that $\tilde{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 (\mathbb{B}_P^H \mathbb{B}_P)^{-1})$ because

$$\begin{aligned}\tilde{\mathbf{h}} = \mathbf{h} - \hat{\mathbf{h}} &= \mathbf{h} - (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{y}_P \\ &= \mathbf{h} - (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H (\mathbb{B}_P \mathbf{h} + \mathbf{n}_P) \\ &= -(\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P.\end{aligned}$$

with $E[-(\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P] = \mathbf{0}$ and

$$E[\tilde{\mathbf{h}} \tilde{\mathbf{h}}^H] = E[(\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P \mathbf{n}_P^H \mathbb{B}_P (\mathbb{B}_P^H \mathbb{B}_P)^{-1}] = \sigma_n^2 (\mathbb{B}_P^H \mathbb{B}_P)^{-1}.$$

For the case of $P = 1$, the Theorem 1 in [1] had showed that this system model we are considering makes the worst capacity when the estimate error of a training-based system is zero-mean Gaussian with variance σ_n^2 and independent to channel estimation $\hat{\mathbf{h}}$.

A well-known lower bound of mean square error $E[\|\tilde{\mathbf{h}}\|^2]$ [4, 8] is

$$E[\|\tilde{\mathbf{h}}\|^2] = \text{tr} \left(E[\tilde{\mathbf{h}} \tilde{\mathbf{h}}^H] \right) = \sigma_n^2 \text{tr}((\mathbb{B}_P^H \mathbb{B}_P)^{-1}) \geq \frac{\sigma_n^2 P}{T}, \quad (2)$$

the equality is hold if $\mathbb{B}_P^H \mathbb{B}_P = T\mathbf{I}$. Perhaps the modulation signals make \mathbb{B}_P not to satisfy the condition $\mathbb{B}_P^H \mathbb{B}_P = T\mathbf{I}$. We can use a method in [5] to relax \mathbb{B}_P for satisfying $\mathbb{B}_P^H \mathbb{B}_P = \mathbb{S}_P$, where \mathbb{S}_P is a matrix that has the least distance $\|\mathbb{S}_P - T\mathbf{I}\|$. Obviously, to place the pilot signal in the beginning of \mathbb{B} is one of arrangements to make the biggest T in the lower bound (2).

3 Decoding for Partial CSI at Receivers

3.1 ML Decoding Metric

The channel model (1) can be rewritten as

$$\mathbf{y}_D = \mathbb{B}_D \hat{\mathbf{h}} + \mathbb{B}_D (\mathbf{h} - \hat{\mathbf{h}}) + \mathbf{n}_D. \quad (3)$$

We know that $\Pr(\mathbf{y}_D | \mathbb{B}_D, \hat{\mathbf{h}})$ is complex gaussian with mean $E[\mathbf{y}_D] = \mathbb{B}_D \hat{\mathbf{h}}$ and covariance

$$\mathbf{C} = \sigma_n^2 \mathbf{I} + \mathbb{B}_D \mathbf{C}_{\tilde{\mathbf{h}}} \mathbb{B}_D^H = \sigma_n^2 (\mathbf{I} + \mathbb{B}_D (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D^H).$$

By Sylvester's determinant theorem,

$$\det |\mathbf{C}| = \sigma_n^{2P} \det |\mathbf{I} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D^H \mathbb{B}_D|.$$

Since

$$\mathbf{C}^{-1} = \sigma_n^{-2} (\mathbf{I} - \mathbb{B}_D (\mathbb{B}_D^H \mathbb{B}_D + \mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D^H),$$

the ML decoding criterion for the receiver with partial CSI should be

$$\begin{aligned}\hat{\mathbf{b}}_{ML} &= \arg \max_{\mathbb{B}_D} \Pr(\mathbf{y}_D | \mathbb{B}_D, \hat{\mathbf{h}}) \\ &= \arg \max_{\mathbb{B}_D} \left\{ \frac{1}{\pi^P \det |\mathbf{C}|} \exp\{-(\mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}})^H \mathbf{C}^{-1} (\mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}})\} \right\} \\ &= \arg \min_{\mathbb{B}_D} \{(\mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}})^H \mathbf{C}^{-1} (\mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}}) - \log \det |\mathbf{C}|\} \\ &= \arg \min_{\mathbb{B}_D} \left\{ \|\mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}}\|^2 - \mathbf{y}_D^H \mathbb{B}_D (\mathbb{B}_D^H \mathbb{B}_D + \mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D^H \mathbf{y}_D \right. \\ &\quad \left. - \log \det |\mathbf{I} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D^H \mathbb{B}_D| \right\}. \quad (4)\end{aligned}$$

Obviously, ML metric consists of two part, the coherent part $\|\mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}}\|^2$ and the non-coherent part which is the rest of the metric.

We take a look at two extreme cases at the receiver, no CSI (non-coherent receiver) and prefect CSI (coherent receiver). If there is no CSI at the receiver, it means that \mathbb{B}_P is a $T \times P$ all-zero matrix and (4) can be reduced as well-known GLRT criterion

$$\begin{aligned}\hat{\mathbf{b}}_{non-coh} &= \arg \min_{\mathbb{B}_D} \left\{ -\mathbf{y}_D^H \mathbb{B}_D (\mathbb{B}_D^H \mathbb{B}_D)^{-1} \mathbb{B}_D^H \mathbf{y}_D + \log \det |\mathbb{C}^{-1}| \right\} \\ &= \arg \min_{\mathbb{B}_D} \left\{ -\mathbf{y}_D^H \mathbb{B}_D (\mathbb{B}_D^H \mathbb{B}_D)^{-1} \mathbb{B}_D^H \mathbf{y}_D + \log \det |\mathbb{I} - \mathbb{B}_D (\mathbb{B}_D^H \mathbb{B}_D)^{-1} \mathbb{B}_D^H| \right\} \\ &= \arg \min_{\mathbb{B}_D} \left\| \mathbb{P}_D^\perp \mathbf{y}_D \right\|^2,\end{aligned}$$

where $\mathbb{P}_D^\perp = \mathbb{I} - \mathbb{B}_D (\mathbb{B}_D^H \mathbb{B}_D)^{-1} \mathbb{B}_D^H$. Otherwise, if the receiver has perfect CSI, i.e., perfect channel estimation, we have $\tilde{\mathbf{h}} = \mathbf{0}$ and $E[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H]$ is a all-zero matrix. It would make $\mathbb{C} = \sigma_n^2 \mathbb{I}$ and reduce (4) to be

$$\hat{\mathbf{b}}_{coh} = \arg \min_{\mathbb{B}_D} \left\{ \left\| \mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}} \right\|^2 \right\}. \quad (5)$$

3.2 Numerical Simulations

The channel parameters \mathbf{h} used in our simulations is zero-mean complex-Gaussian distributed with $E[\mathbf{h}\mathbf{h}^H] = P\mathbb{I}$ and $P = 2$. The definition of SNR is

$$SNR = \frac{\text{tr}(E[\mathbf{h}\mathbf{h}^H])}{\sigma_n^2} = \frac{E[\mathbf{h}^H \mathbf{h}]}{\sigma_n^2} = \frac{1}{\sigma_n^2}.$$

We examined the system consist of 7-bit pilot signals $\mathbb{B}_P = [0, 0, 0, 0, 1, 0, 1]$ and (15, 11) Hamming codes. The performance of receivers using traditional coherent coding criterion of (5) and the ML decoding of (4) are compared in Fig. 1. It shows that ML decoding can outperform the tradition coherent decoding and the performance gap is about 0.5dB.

4 Code Design

By the inequality in Section III-A of [5], if each \mathbb{B}_D can satisfy the self-orthogonal condition

$$\mathbb{B}_D^H \mathbb{B}_D = N\mathbb{I}, \quad (6)$$

the average SNR at receivers can be maximized regardless of channel statistics. Due to the identical reason mentioned in 2.1, the limitation of modulation symbols, sometimes it is impossible to make every codeword \mathbb{B}_D have the property in (6). Therefore, we can relax the condition and make \mathbb{B}_D to satisfy $\mathbb{B}_D^H \mathbb{B}_D = \mathbb{S}_D$, where \mathbb{S}_D is one of matrices that has the least distance $\|\mathbb{S}_D - N\mathbb{I}\|$ based on the modulation symbol of \mathbb{B}_D . Our work on the code design for partial CSI at receivers would only focus on the signals satisfied $\mathbb{B}_D^H \mathbb{B}_D = \mathbb{S}_D$. Then, ML decoding metric (4) could be reduced as

$$\hat{\mathbf{b}}_{ML} = \arg \min_{\mathbb{B}_D} \left\{ \left\| \mathbf{y}_D - \mathbb{B}_D \hat{\mathbf{h}} \right\|^2 - \mathbf{y}_D^H \mathbb{B}_D (\mathbb{S}_D + \mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D^H \mathbf{y}_D \right\}. \quad (7)$$

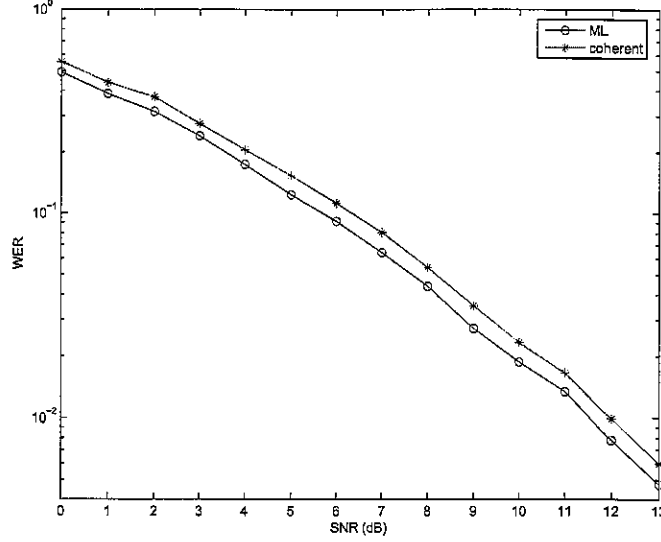


Figure 1: Performance of the ML and traditional coherent decoding for the system with 7-bits training sequence and (15, 11) Hamming codes. The traditional coherent decoding adopts $\hat{\mathbf{h}}$ as the true \mathbf{h} and use the criterion for perfect CSI to decode \mathbb{B}_D .

4.1 Analysis of Pairwise Error Probability in very High SNR Regime

We define $\mathbb{B}_D(i)$ is the i -th codeword in a code \mathcal{C} for $i = 0, \dots, 2^K - 1$. Let $\mathbf{S}_i = \mathbb{B}(i)^H \mathbb{B}(i) + \mathbb{B}_P^H \mathbb{B}_P$, so that both \mathbf{S}_i and \mathbf{S}_i^{-1} are Hermitian matrices. For certain channel realization \mathbf{h} , if $\mathbb{B}_D(i)$ is transmitted, the pairwise error probability is

$$\begin{aligned}
 P_{j|i} &\triangleq \Pr \left(\|\mathbf{y}_D - \mathbb{B}_D(i)\hat{\mathbf{h}}\|^2 - \mathbf{y}_D^H \mathbb{B}_D(i) \mathbf{S}_i^{-1} \mathbb{B}_D(i)^H \mathbf{y}_D - \log(\det[\mathbf{I} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D(i)^H \mathbb{B}_D(i)]) \right. \\
 &\quad \left. - \|\mathbf{y}_D - \mathbb{B}_D(j)\hat{\mathbf{h}}\|^2 + \mathbf{y}_D^H \mathbb{B}_D(j) \mathbf{S}_j^{-1} \mathbb{B}_D(j)^H \mathbf{y}_D + \log(\det[\mathbf{I} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D(j)^H \mathbb{B}_D(j)]) \geq 0 \right) \\
 &= \Pr(2\Re(\mathbf{y}_D^H \mathbf{L}_{i,j} \hat{\mathbf{h}}) + \hat{\mathbf{h}}^H \mathbf{M}_{i,j} \hat{\mathbf{h}} + \mathbf{y}_D^H \mathbf{K}_{i,j} \mathbf{y}_D \geq \delta_{i,j}),
 \end{aligned}$$

where $\mathbf{K}_{i,j} = (\mathbb{B}_D(j) \mathbf{S}_j^{-1} \mathbb{B}_D(j)^H - \mathbb{B}_D(i) \mathbf{S}_i^{-1} \mathbb{B}_D(i)^H)$, $\mathbf{M}_{i,j} = (\mathbb{B}(i)^H \mathbb{B}(i) - \mathbb{B}(j)^H \mathbb{B}(j))$, $\mathbf{L}_{i,j} = \mathbb{B}_D(j) - \mathbb{B}_D(i)$ and $\delta_{i,j} = \log \left(\frac{\det[\mathbf{I} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D(i)^H \mathbb{B}_D(i)]}{\det[\mathbf{I} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_D(j)^H \mathbb{B}_D(j)]} \right)$. Due to $\mathbf{y}_D = \mathbb{B}_D(i)\mathbf{h} + \mathbf{n}_D$ and $\hat{\mathbf{h}} = \mathbf{h} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P$, we have

$$\begin{aligned}
 2\Re(\mathbf{y}_D^H \mathbf{L}_{i,j} \hat{\mathbf{h}}) &= 2\Re((\mathbb{B}_D(i)\mathbf{h} + \mathbf{n}_D)^H \mathbf{L}_{i,j} (\mathbf{h} + (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P)) \\
 &= 2\Re(\mathbf{h}^H \mathbb{B}_D(i)^H \mathbf{L}_{i,j} \mathbf{h} + \mathbf{n}_D^H \mathbf{L}_{i,j} \mathbf{h} + \mathbf{h}^H \mathbb{B}_D(i)^H \mathbf{L}_{i,j} (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P \\
 &\quad + \mathbf{n}_D^H \mathbf{L}_{i,j} (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P),
 \end{aligned}$$

$$\begin{aligned}
 \hat{\mathbf{h}}^H \mathbf{M}_{i,j} \hat{\mathbf{h}} &= \mathbf{h}^H \mathbf{M}_{i,j} \mathbf{h} + \mathbf{n}_P^H \mathbb{B}_P (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbf{M}_{i,j} (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P \\
 &\quad + 2\Re(\mathbf{h}^H \mathbf{M}_{i,j} (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P),
 \end{aligned}$$

and

$$\begin{aligned}
 \mathbf{y}_D^H \mathbf{K}_{i,j} \mathbf{y}_D &= (\mathbb{B}_D(i)\mathbf{h} + \mathbf{n}_D)^H \mathbf{K}_{i,j} (\mathbb{B}_D(i)\mathbf{h} + \mathbf{n}_D) \\
 &= \mathbf{h}^H \mathbb{B}_D(i)^H \mathbf{K}_{i,j} \mathbb{B}_D(i)\mathbf{h} + 2\Re(\mathbf{h}^H \mathbb{B}_D(i)^H \mathbf{K}_{i,j} \mathbf{n}_D) + \mathbf{n}_D^H \mathbf{K}_{i,j} \mathbf{n}_D.
 \end{aligned}$$

Let

$$\begin{aligned}
\chi &= -2\Re(\mathbf{h}^H \mathbb{B}_D(i)^H \mathbb{L}_{i,j} \mathbf{h}) - \mathbf{h}^H \mathbb{B}_D(i)^H \mathbb{K}_{i,j} \mathbb{B}_D(i) \mathbf{h} - \mathbf{h}^H \mathbb{M}_{i,j} \mathbf{h} \\
&= -\mathbf{h}^H (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{L}_{i,j}^H \mathbb{B}_D(i) + \mathbb{B}_D(i)^H \mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{M}_{i,j}) \mathbf{h} \\
&= -\mathbf{h}^H \mathbb{V}_{i,j} \mathbf{h},
\end{aligned}$$

where $\mathbb{V}_{i,j} = (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{L}_{i,j}^H \mathbb{B}_D(i) + \mathbb{B}_D(i)^H \mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{M}_{i,j})$ and let

$$x = 2\Re(\mathbf{n}_D^H (\mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{L}_{i,j}) \mathbf{h} + \mathbf{h}^H (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{M}_{i,j}) (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H \mathbf{n}_P).$$

Obviously, x is zero-mean Gaussian with variance

$$\begin{aligned}
\sigma_x^2 &= 2\sigma_n^2 \|(\mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{L}_{i,j}) \mathbf{h}\|^2 + 2\sigma_n^2 \|\mathbf{h}^H (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{M}_{i,j}) (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H\|^2 \\
&\quad + 2\mathbf{h}^T (\mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{L}_{i,j})^T E[\mathbf{n}_D^* \mathbf{n}_D^H] (\mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{L}_{i,j}) \mathbf{h} \\
&\quad + 2\mathbf{h}^H (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{M}_{i,j}) (\mathbb{B}_P^H \mathbb{B}_P)^{-1} \mathbb{B}_P^H E[\mathbf{n}_P \mathbf{n}_P^T] \mathbb{B}_P^H (\mathbb{B}_P^H \mathbb{B}_P)^{-1} (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{M}_{i,j})^T \mathbf{h}^*.
\end{aligned}$$

Because both \mathbf{n}_D and \mathbf{n}_P are zero-mean circular symmetric complex Gaussian, their pseudo-covariance matrix $E[\mathbf{n}_D^* \mathbf{n}_D^H]$ and $E[\mathbf{n}_P \mathbf{n}_P^T]$ are zero matrices. Then, we know that

$$\sigma_x^2 = 2\sigma_n^2 \mathbf{h}^H \mathbb{U}_{i,j} \mathbf{h},$$

where

$$\mathbb{U}_{i,j} = ((\mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{L}_{i,j})^H (\mathbb{K}_{i,j} \mathbb{B}_D(i) + \mathbb{L}_{i,j}) + (\mathbb{B}_D(i)^H \mathbb{L}_{i,j} + \mathbb{M}_{i,j}) (\mathbb{B}_P^H \mathbb{B}_P)^{-1} (\mathbb{L}_{i,j}^H \mathbb{B}_D(i) + \mathbb{M}_{i,j}^H)).$$

By the skill in [6, 7], the quadratic terms can be ignored in very high SNR. Finally, we have

$$P_{j|i} \approx \Pr(x \geq \chi) = \mathcal{Q}\left(\frac{\chi + \sigma_{i,j}}{\sigma_x}\right).$$

Because both $\mathbb{U}_{i,j}$ and $\mathbb{V}_{i,j}$ are Hermitian matrices,

$$\chi \geq -\lambda_{\max}(\mathbb{V}_{i,j}) \|\mathbf{h}\|^2 \text{ and } \sigma_x^2 \leq 2\sigma_n^2 \lambda_{\max}(\mathbb{U}_{i,j}) \|\mathbf{h}\|^2,$$

the equalities, are hold if it happens $\mathbb{V}_{i,j} = \mathbb{I}$ and $\mathbb{U}_{i,j} = \mathbb{I}$, respectively. We obtain a upper bound in very high SNR

$$\mathcal{Q}\left(\frac{\chi}{\sigma_x}\right) \leq \mathcal{Q}\left(\frac{-\lambda_{\max}(\mathbb{V}_{i,j}) \|\mathbf{h}\|^2 + \sigma_{i,j}}{\sigma_n \sqrt{2\lambda_{\max}(\mathbb{U}_{i,j})} \|\mathbf{h}\|}\right).$$

We design the codes based on $\|\mathbf{h}\| = 1$ that is the setting is used in simulations for evaluating conventional systems [4]. Then, our code \mathcal{C} can be designed via the cost function

$$f(\mathcal{C}) = \min_{i \neq j} \left\{ \frac{\lambda_{\max}(\mathbb{V}_{i,j}) - \sigma_{i,j}}{\sqrt{\lambda_{\max}(\mathbb{U}_{i,j})}} \right\}, \quad (8)$$

this optimization is regardless to signal-to-noise ratio $1/\sigma_n^2$.

4.2 Iterative Joint Design Algorithm

We provide an iterative algorithm to design the training sequence and code jointly. For convenience, some notations will be used to mention the algorithm. The signal \mathbb{B}_D can be divided to two parts

$$\mathbb{B}_D(i) = \overline{\mathbb{B}}_P + \overline{\mathbb{B}}_D(i),$$

where

$$\overline{\mathbb{B}}_P \triangleq \begin{bmatrix} 0 & b_T & \cdots & b_{T-P+2} \\ 0 & 0 & \ddots & \vdots \\ 0 & \vdots & \ddots & b_T \\ 0 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{(L-T) \times P} \quad \text{and} \quad \overline{\mathbb{B}}_D(i) \triangleq \begin{bmatrix} b_{T+1}(i) & 0 & \cdots & 0 \\ \vdots & b_{T+1}(i) & \ddots & \vdots \\ b_N(i) & \vdots & \ddots & 0 \\ 0 & b_N(i) & \ddots & b_{T+1}(i) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & b_N(i) \end{bmatrix}_{(L-T) \times P},$$

where $\overline{\mathbb{B}}_P$ is decided by \mathbb{B}_P completely.

For our code design algorithm, after a initial training sequence $\mathbb{B}_P^{(0)}$ is generated randomly, Simulated Annealing Algorithm [2] is used for searching a good code $\mathcal{C}^{(0)}$. Based on the $\mathcal{C}^{(0)}$, a better $\mathbb{B}_P^{(1)}$ can be obtained via Simulated Annealing Algorithm again or exhausted search. If the updated \mathbb{B}_P satisfies the stop condition, output the code \mathcal{C} and \mathbb{B}_P . Otherwise, go back Simulated Annealing Algorithm to continue refining the code and training sequence. We list the details of our algorithm in the following:

Step 1. Initialize $i = 0$. Generate $\mathbb{B}_P^{(i)}$ randomly.

Step 2. Perform Simulated Annealing Algorithm to search a good code $\mathcal{C}^{(i)}$ based on the cost function (8) with $\mathbb{B}_P^{(i)}$.

Step 3. Set $i = i + 1$.

Perform Simulated Annealing Algorithm or exhausted search to look for a good pilot signals $\mathbb{B}_P^{(i)}$ based on the cost function (8) with $\mathcal{C}^{(i-1)}$.

Step 4. If $\mathbb{B}_P^{(i)H} \mathbb{B}_P^{(i)} = \mathbb{B}_P^{(i-1)H} \mathbb{B}_P^{(i-1)}$ and $\overline{\mathbb{B}}_P^{(i)} = \overline{\mathbb{B}}_P^{(i-1)}$, output $\mathcal{C}^{(i-1)}$ and $\mathbb{B}_P^{(i)}$, and the algorithm stops; otherwise, go to Step 2.

References

- [1] B. Hassibi and B. M. Hochward, "How much Training is needed in multiple-Antenna Wireless links," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951-963, Apr. 2003.
- [2] M. Skoglund, J. Giese and S. Parkvall, "Code design for combined channel estimation and error protection," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1162-1171, May 2002.
- [3] J. Giese and M. Skoglund, "Space-Time Constellation Design for Partial CSI at the Receiver," *IEEE Trans. Inform. Theory*, vol. 53, no. 8, pp. 2715-2731, August 2007.

- [4] O. Coskun and K. M. Chugg, "Combined coding and training for unknown ISI channels," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1310-1322, August 2005.
- [5] C.-L. Wu, P.-N. Chen, Yunghsiang S. Han and M.-H. Kuo, "Maximum-likelihood priority-first search decodable codes for combined channel estimation and error correction," accepted by *IEEE Trans. Inform. Theory*, vol. 55, no. 9, Sept 2009.
- [6] M. Beko, J. Xavier and V. A. N. Barroso, "Noncoherent communication in multiple-antenna system: receiver design and codebook construction," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5703-5715, Dec. 2007.
- [7] E. G. Larsson and P. Stoica, *Space-Time Block Codes for Wireless Communications.*, Cambridge University Press, 2003.
- [8] C. Fragouli, N. Al-Dhahir and W. Turin, "Finite-Alphabet Constant-Amplitude Training Sequences for Multiple-Antennas," *Proc. IEEE International Conference on Communications (ICC 2002)*, New York, USA, April 28-May 2, 2002.