Code Construction for Gauss-Markov Fading Channels

by

Chin-Hui Cheng

Thesis submitted to the Institute of Communications Engineering
National Chiao Tung University, Taiwan
in partial fulfillment of the requirements
for the degree of Master of Science
2005

Advisory Committee:

Professor Po-Ning Chen, Advisor
Professor Yunghsiang S. Han
Professor Chung-Chin Lu
Professor Chung-Hsuan Wang

ABSTRACT

In 2002, Skoglund, Giese and Parkvall introduced a novel concept of combining channel estimation, equalization and decoding. They optimized the block error rate by designing a block code which can provide channel estimation and error protection to the receiver at the same time. However it is not clear that their approach will perform well on a fast-varying channel. In this thesis, We extend their approach in order to adjust well to Gauss-Markov channels. By deriving the pairwise error probability as a function of the criterion for the simulated annealing algorithm, Our designed codes provide a coding gain of about 3.5 dB and 6 dB on the Gauss-Markov channels with channel lengths 1 and 2 respectively over those given in [7] at WER= 10^{-2}.

ACKNOWLEDGMENTS

I would like to thank everyone who helps me complete this thesis. To Professor Chen and Professor Han, it is your patient guidance and encouragement that makes this thesis possible. I am really grateful to be led by such great teachers. To my labmates, all of you always sincerely listen to my problems and give me helpful advice. To my dear family and Ching-Hui, I have a joyful school life because of your supports.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

Recently, wireless communication has become a popular domain because of the great demand in market [1]. Not only the speed but also the quality are important issues for wireless communication applications. For wireless communications, the main design challenge arises from the harsh propagation environment determined by channel fading parameters. It may be resulted from reflex and diffuse multipath loss, and cochannel interference, and then makes reliable transmission more difficult. Multipath propagation and limited bandwidth are the two main causes of signal distortion that lead to intersymbol interference (ISI). ISI may lead to higher error rates in symbol detection at the receiver. Moreover, the more obstructions in the communication path, the faster the channel varies. The general methods to combat this problem are channel coding, channel estimation and equalization [2].

Channel estimation scheme in the receiver estimates channel parameters at present by a known training sequence, and passes these parameters to equalizer to equalize the effect on the received signal induced by channel fading [3]. Since the training sequence dose not carry any information data and is a waste of channel usage, an alternative approach, a blind method, transmitting no training data but only the channel output is used for channel estimation [4]. Another hybrid approach, called semiblind, utilizes both training data and input information to perform channel estimation [5, 6].

Many surveys emphasized on channel coding (or error control coding) had been proposed. In 2002, Skoglund, Giese and Parkvall introduced a novel concept of combining channel estimation, equalization and decoding, where they focused on the design of an encoder rather than on schemes employed at the receiver that can improve the performance of the parameter estimation [7]. They tried to optimize the block error rate by designing a block code which can provide channel estimation and error protection to the receiver at the same time. The channel model under consideration in their approach is given as

$$\mathbf{y} = \begin{bmatrix} h_1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ h_p & \cdots & h_1 & 0 & \cdots & 0 \\ 0 & h_p & \cdots & h_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & h_P & \cdots & h_1 \\ \vdots & \ddots & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & h_p \end{bmatrix} \mathbf{b} + \mathbf{n} = \mathbf{H}\mathbf{b} + \mathbf{n}$$

or equivalently,

$$\mathbf{y} = \mathbf{B}\mathbf{h} + \mathbf{n}, \mathbf{B} = \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ b_2 & b_1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ b_p & b_{p-1} & \cdots & b_1 \\ \vdots & & \vdots & \vdots \\ b_N & b_{N-1} & \cdots & b_{N-p+1} \\ 0 & b_N & \cdots & b_{N-p+2} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & b_N \end{bmatrix},$$

where $\mathbf{b} \in \{\pm 1\}^N$ is the transmitted codeword of a block code, $\mathbf{n}$ is zero-mean complex Gaussian noise with correlation $E[\mathbf{n}\mathbf{n}^H] = \sigma_n^2 \mathbf{I}_L$, and $\mathbf{h} = [h_1, \ldots, h_p]^T$ is the channel coefficients and is assumed constant over the transmission of one block $\mathbf{b}$, but varied between each blocks. The optimal joint maximum-likelihood (ML) decoder estimating channel coefficients $\mathbf{h}$ and the transmitted codeword $\mathbf{b}$ is given as

$$(\hat{\mathbf{h}}, \hat{\mathbf{b}}) = \arg\min_{\mathbf{b},\mathbf{h}} \{\|\mathbf{y} - \mathbf{B}\mathbf{h}\|^2\}.$$

Since for each fixed $\mathbf{b}$, $\hat{\mathbf{h}}$ can be expressed as

$$\hat{\mathbf{h}} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T \cdot \mathbf{y}$$

we can remove the dependence on the channel of the receiver. Then, the decision of transmitted bits is

$$\begin{aligned} \hat{\mathbf{b}} &= \arg\min_{\mathbf{b} \in \{\pm 1\}^N} \{\|\mathbf{y} - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y}\|^2\} \\ &= \arg\min_{\mathbf{b} \in \{\pm 1\}^N} \{\|\mathbf{y} - \mathbf{P}_B\mathbf{y}\|^2\} \\ &= \arg\min_{\mathbf{b} \in \{\pm 1\}^N} \{\|\mathbf{P}_B^{\perp}\mathbf{y}\|^2\}, \end{aligned} \tag{1.1}$$

where $\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ and $\mathbf{P}_{\boldsymbol{B}}^{\perp} = \boldsymbol{I} - \mathbf{P}_{\boldsymbol{B}}$. The receiver then decide which codeword is transmitted according to (1.1). By comparing their designed code to a Hamming code working with optimal joint estimation, equalization and soft decoding, they found that the designed code outperformed the above Hamming code significantly.

The code designed in [7] has been proved to have excellent performance over the block Rayleigh-fading channel. However, it is not indicated how it will perform on a more critical channel which is not block fading. Usually the channel coefficients are changing during the period of transmission of codeword in practice. Thus, we are interested in finding a code adequate to a fast time-varying channel. In this thesis, we will focus on the code design over the Gauss-Markov channel [8, 9]. The reason that we are enthusiastic about Gauss-Markov channel is that it can imitate any fast time-varying channel as long as the order of Markov factor is large enough.

1.2 Outline of Thesis

The remaining part is organized into four parts. Chapter 2 describes the Gauss-Markov channel. The simulated annealing algorithm that will be used in next section and the code design problem are also given. The code design problem is further discussed in chapter 3, with newly derived criterion and a complete search algorithm over the Gauss-Markov channel. Our simulation results and comparisons to the results given in [7] are given in chapter 4. Chapter 5 is the summary and the conclusions of this work.

2. PRELIMINARIES

Since what we concern in this work is to find a good code reaching low error rate in the specific Gauss-Markov channel, in this chapter, we briefly give the characteristics of the channel and the search algorithm we will use for code design.

2.1 An upper bound on block error rate

For a time-varying channel with noise and inter-symbol interference, the received data at time k is

$$r_k = \mathbf{a}_k^T \mathbf{h}_k + n_k,$$

where $\mathbf{a}_k^T = [a_k, a_{k-1}, \cdots, a_{k-M+1}]$ and M is the channel length indicating how far the past transmitted bits affect the received bit in this time instance. Furthermore, $\mathbf{a}_k^T$ is a complex row vector containing M transmitted data from time k to time k-M+1, $\mathbf{h}_k$ is a complex column vector with

M channel impulse response coefficients at time k, and n_k is a Gaussian complex noise with mean zero, variance σ^2. Let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_N]$ be the matrix of channel coefficient vectors representing by columns, and let $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_N]^T$ represent the matrix of transmitted data representing by rows. Also let $\mathbf{r} = [r_1, \cdots, r_N]^T$ represent the received vector, while $\mathbf{n} = [n_1, \cdots, n_N]^T$ be the vector of noise from time 1 to time N. If channel impulse response coefficients follow the Gauss-Markov distribution, then it is called a Gauss-Markov channel. Moreover, the channel impulse response $\mathbf{h}_k = \alpha \mathbf{h}_{k-1} + \mathbf{v}_k$ at time k, where $\mathbf{v}_k$ is a complex white Gaussian with mean $\mathbf{d}$ and covariance $\mathbf{C}$, and α is a complex first order Markov factor whose absolute value $|\alpha| = e^{-\omega T}$ with ω/π the Doppler spread and T the sampling period. As we can see, Gauss-Markov channel is a channel whose time-varying behavior is represented by Markov and Gauss random variables. With this relation, $f(\mathbf{H})$ can be written as

$$f(\mathbf{H}) = f(\mathbf{h}_1) \prod_{k=2}^{N} f(\mathbf{h}_k | \mathbf{h}_{k-1}). \tag{2.1}$$

Therefore the conditional probability density function $f(\mathbf{r}|\mathbf{H}, \mathbf{A})$ can be expressed as

$$f(\mathbf{r}|\mathbf{H}, \mathbf{A}) = \frac{1}{(\pi\sigma^2)^N} \prod_{k=1}^{N} e^{-\frac{|r_k - \mathbf{a}_k^T \mathbf{h}_k|^2}{\sigma^2}}, \tag{2.2}$$

and also the likelihood function be

$$f(\mathbf{r}|\mathbf{A}) = E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] = \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A}) f(\mathbf{H}) d\mathbf{H}. \tag{2.3}$$

Substituting (2.2), (2.1) into (2.3), and assuming the mean of $\mathbf{v}_k$ is zero and initial channel coefficient $\mathbf{h}_0$ is known, we can derive the likelihood function of Gauss-Markov channel as

$$\begin{aligned} f(\mathbf{r}|\mathbf{A}) &= E[f(\mathbf{r}|\mathbf{H},\mathbf{A})] \\ &= \int_{\mathbf{H}} f(\mathbf{r}|\mathbf{H},\mathbf{A}) f(\mathbf{H}) d\mathbf{H} \\ &= \int_{\mathbf{H}} \left(\prod_{k=1}^{N} e^{-\frac{|r_k - \mathbf{a}_k^T \mathbf{h}_k|^2}{\sigma^2}} \right) f(\mathbf{h}_1) \prod_{k=2}^{N} f(\mathbf{h}_k|\mathbf{h}_{k-1}) d\mathbf{H}. \end{aligned} \tag{2.4}$$

According to maximum-likelihood decoding rule, our goal is to find an $\hat{A}$ to maximize the likelihood function. That is,

$$\hat{A} = \arg\max_{\boldsymbol{A}} \; f(\mathbf{r}|\mathbf{A}). \tag{2.5}$$

We then marginalize the channel coefficient distribution in (2.4) to

$$\begin{aligned} f(\mathbf{r}|\mathbf{A}) &= \int_{\boldsymbol{h}_N} \int_{\boldsymbol{h}_{N-1}} \cdots \int_{\boldsymbol{h}_1} \left(\prod_{k=1}^{N} e^{-\frac{|r_k - \boldsymbol{a}_k^T \boldsymbol{h}_k|}{\sigma^2}} \right) \prod_{k=1}^{N} e^{-(\boldsymbol{h}_k - \alpha \boldsymbol{h}_{k-1})^H \boldsymbol{C}^{-1} (\boldsymbol{h}_k - \alpha \boldsymbol{h}_{k-1})} d\boldsymbol{h}_1 d\boldsymbol{h}_2 \cdots d\boldsymbol{h}_N \\ &= \int_{\boldsymbol{h}_N} e^{-\boldsymbol{E}_N} \left(\int_{\boldsymbol{h}_{N-1}} \cdots \left(\int_{\boldsymbol{h}_2} e^{-\boldsymbol{E}_2} \left(\int_{\boldsymbol{h}_1} e^{-\boldsymbol{E}_1} d\boldsymbol{h}_1 \right) d\boldsymbol{h}_2 \right) \cdots d\boldsymbol{h}_{N-1} \right) d\boldsymbol{h}_N, \end{aligned}$$

where $\boldsymbol{E}_1$ depends on $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$, $\boldsymbol{E}_2$ depends on $\boldsymbol{h}_2$ and $\boldsymbol{h}_3$,. . .,$\boldsymbol{E}_{N-1}$ depends on $\boldsymbol{h}_{N-1}$ and $\boldsymbol{h}_N$, and $\boldsymbol{E}_N$ depends only on $\boldsymbol{h}_N$. Then,

$$\begin{aligned}
\tilde{\boldsymbol{E}}_1 &= \frac{|r_1 - \boldsymbol{a}_1^T \boldsymbol{h}_1|^2}{\sigma^2} + (\boldsymbol{h}_1 - (\alpha \boldsymbol{h}_0))^H \boldsymbol{C}^{-1} (\boldsymbol{h}_1 - (\alpha \boldsymbol{h}_0)) + (\boldsymbol{h}_2 - \alpha \boldsymbol{h}_1)^H \boldsymbol{C}^{-1} (\boldsymbol{h}_2 - \alpha \boldsymbol{h}_1) \\
&= \frac{1}{\sigma^2}(r_1^* - \boldsymbol{h}_1^H \boldsymbol{a}_1^*)(r_1 - \boldsymbol{a}_1^T \boldsymbol{h}_1) + (\boldsymbol{h}_1^H - \alpha^* \boldsymbol{h}_0^H) \boldsymbol{C}^{-1} (\boldsymbol{h}_1 - \alpha \boldsymbol{h}_0) \\
&\quad + (\boldsymbol{h}_2^H - \alpha^* \boldsymbol{h}_1^H) \boldsymbol{C}^{-1} (\boldsymbol{h}_2 - \alpha \boldsymbol{h}_1) \\
&= \frac{1}{\sigma^2} \left[|r_1|^2 - r_1^* \boldsymbol{a}_1^T \boldsymbol{h}_1 - r_1 \boldsymbol{h}_1^H \boldsymbol{a}_1^* + \boldsymbol{h}_1^H \boldsymbol{a}_1^* \boldsymbol{a}_1^T \boldsymbol{h}_1 \right] \\
&\quad + \left[\boldsymbol{h}_1^H \boldsymbol{C}^{-1} \boldsymbol{h}_1 - \alpha^* \boldsymbol{h}_0^H \boldsymbol{C}^{-1} \boldsymbol{h}_1 - \alpha \boldsymbol{h}_1^H \boldsymbol{C}^{-1} \boldsymbol{h}_0 + |\alpha|^2 \boldsymbol{h}_0^H \boldsymbol{C}^{-1} \boldsymbol{h}_0 \right] \\
&\quad + \left[\boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 - \alpha \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_1 - \alpha^* \boldsymbol{h}_1^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 + |\alpha|^2 \boldsymbol{h}_1^H \boldsymbol{C}^{-1} \boldsymbol{h}_1 \right].
\end{aligned}$$

$|r_1|^2/\sigma^2$ and $|\alpha|^2 \boldsymbol{h}_0^H \boldsymbol{C}^{-1} \boldsymbol{h}_0$ can be removed in $\hat{\boldsymbol{A}}$ derivation because of their irrelevance to the maximization operation. $\boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2$ can be moved out from

$\tilde{\boldsymbol{E}}_1$ since it can be combined into $\boldsymbol{E}_2$. This simplifies $\tilde{\boldsymbol{E}}_1$ to

$$
\begin{aligned}
& \frac{1}{\sigma^2}\left[-r_1^*\boldsymbol{a}_1^T\boldsymbol{h}_1 - r_1\boldsymbol{h}_1^H\boldsymbol{a}_1^* + \boldsymbol{h}_1^H\boldsymbol{a}_1^*\boldsymbol{a}_1^T\boldsymbol{h}_1\right] \\
& +\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha^*\boldsymbol{h}_0^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_0 \\
& -\alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha^*\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_2 + |\alpha|^2\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 \\
= \; & -\frac{1}{\sigma^2}r_1^*\boldsymbol{a}_1^T\boldsymbol{h}_1 - \frac{1}{\sigma^2}r_1\boldsymbol{h}_1^H\boldsymbol{a}_1^* + \left(\frac{1}{\sigma^2}\boldsymbol{h}_1^H\boldsymbol{a}_1^*\boldsymbol{a}_1^T\boldsymbol{h}_1 + |\alpha|^2\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 + \boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_1\right) \\
& -\alpha^*\boldsymbol{h}_0^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_0 \\
& -\alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha^*\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_2 \\
= \; & -\frac{1}{\sigma^2}r_1^*\boldsymbol{a}_1^T\boldsymbol{h}_1 - \frac{1}{\sigma^2}r_1\boldsymbol{h}_1^H\boldsymbol{a}_1^* + \boldsymbol{h}_1^H\left(\frac{\boldsymbol{a}_1^*\boldsymbol{a}_1^T}{\sigma^2} + (1+|\alpha|^2)\boldsymbol{C}^{-1}\right)\boldsymbol{h}_1 \\
& -\alpha^*\boldsymbol{h}_0^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_0 \\
& -\alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1}\boldsymbol{h}_1 - \alpha^*\boldsymbol{h}_1^H\boldsymbol{C}^{-1}\boldsymbol{h}_2 \\
= \; & \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1 - \boldsymbol{h}_1^H\left(\alpha^*\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \alpha\boldsymbol{C}^{-1}\boldsymbol{h}_0 + \frac{r_1\boldsymbol{a}_1^*}{\sigma^2}\right) \\
& -\left(\alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1} + \alpha^*\boldsymbol{h}_0^H\boldsymbol{C}^{-1} + \frac{r_1^*\boldsymbol{a}_1^T}{\sigma^2}\right)\boldsymbol{h}_1
\end{aligned}
$$

(Assume that $((\boldsymbol{C})^{-1})^H = (\boldsymbol{C})^{-1}$ and $\mathbf{G}_1^{-1} = \boldsymbol{a}_1^*\boldsymbol{a}_1^T/\sigma^2 + (1+|\alpha|^2)\boldsymbol{C}^{-1}$.)

$$= \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1 - \boldsymbol{h}_1^H\left(\alpha^*\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \boldsymbol{q}_1\right) - \left(\alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1} + \boldsymbol{q}_1^H\right)\boldsymbol{h}_1$$

(For notation convenience, let $\boldsymbol{g}_1 = \boldsymbol{G}_1(\alpha^*\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \boldsymbol{q}_1)$ and $\mathbf{q}_1 = r_1\boldsymbol{a}_1^*/\sigma^2 + \alpha\boldsymbol{C}^{-1}\boldsymbol{h}_0$.

Also note that $(\boldsymbol{G}_1^{-1})^H = \boldsymbol{G}_1^{-1}$.)

$$= \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1 - \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1 - \boldsymbol{g}_1^H(\boldsymbol{G}_1^{-1})^H\boldsymbol{h}_1$$

$$= \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1 - \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1 - \boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1$$

$$= \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1 - \boldsymbol{h}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1 - \boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{h}_1 + \boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1 - \boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1$$

$$= (\boldsymbol{h}_1 - \boldsymbol{g}_1)^H\boldsymbol{G}_1^{-1}(\boldsymbol{h}_1 - \boldsymbol{g}_1) - \boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1.$$

As $\boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1$ does not depend on $\boldsymbol{h}_1$, we obtain

$$\boldsymbol{E}_1 = (\boldsymbol{h}_1 - \boldsymbol{g}_1)^H\boldsymbol{G}_1^{-1}(\boldsymbol{h}_1 - \boldsymbol{g}_1).$$

Note that $\boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1$ is equal to

$$\begin{aligned}
\boldsymbol{g}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{g}_1 &= (\alpha^*\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \boldsymbol{q}_1)^H\boldsymbol{G}_1^H\boldsymbol{G}_1^{-1}\boldsymbol{G}_1(\alpha^*\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \boldsymbol{q}_1) \\
&= (\alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1} + \boldsymbol{q}_1^H)\boldsymbol{G}_1(\alpha^*\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \boldsymbol{q}_1) \\
&\quad \text{(since } (\boldsymbol{G}_1^H\boldsymbol{G}_1^{-1})^H = (\boldsymbol{G}_1^{-1})^H\boldsymbol{G}_1 = \boldsymbol{G}_1^{-1}\boldsymbol{G}_1 = \boldsymbol{I}) \\
&= |\alpha|^2\boldsymbol{h}_2^H\boldsymbol{C}^{-1}\boldsymbol{G}_1\boldsymbol{C}^{-1}\boldsymbol{h}_2 + \alpha\boldsymbol{h}_2^H\boldsymbol{C}^{-1}\boldsymbol{G}_1\boldsymbol{q}_1 + \alpha^*\boldsymbol{q}_1^H\boldsymbol{G}_1\boldsymbol{C}^{-1}\boldsymbol{h}_2 \\
&\quad + \boldsymbol{q}_1^H\boldsymbol{G}_1\boldsymbol{q}_1.
\end{aligned}$$

$\boldsymbol{q}_1^H\boldsymbol{G}_1\boldsymbol{q}_1$ in the above equation should be removed in $\boldsymbol{E}_2$ derivation because of their irrelevance to the integration operation (but it is relevant to the maximization operation).

Finally, we have

$$\begin{aligned} I_1 &= \int_{\boldsymbol{h}_1} e^{-\left(\boldsymbol{E}_1 - \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{q}_1\right)} d\boldsymbol{h}_1 \\ &= \int_{\boldsymbol{h}_1} e^{-\boldsymbol{E}_1} \cdot e^{\boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{q}_1} d\boldsymbol{h}_1 \\ &= \int_{\boldsymbol{h}_1} e^{-(\boldsymbol{h}_1 - \boldsymbol{g}_1)^H \boldsymbol{G}_1^{-1} (\boldsymbol{h}_1 - \boldsymbol{g}_1)} d\boldsymbol{h}_1 \cdot e^{\boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{q}_1} \\ &= e^{\boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{q}_1} |\boldsymbol{G}_1|. \end{aligned}$$

Now, consider $\boldsymbol{E}_2$ with the terms moved from the derivation of $\boldsymbol{E}_1$ that are $\boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 - |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 - \alpha \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1 - \alpha^* \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2$. Then

$$\begin{aligned} \tilde{\boldsymbol{E}}_2 &= \frac{|r_2 - \boldsymbol{a}_2^T \boldsymbol{h}_2|^2}{\sigma^2} + (\boldsymbol{h}_3 - \alpha \boldsymbol{h}_2)^H \boldsymbol{C}^{-1} (\boldsymbol{h}_3 - \alpha \boldsymbol{h}_2) + \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 - |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\ &\quad - \alpha \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1 - \alpha^* \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\ &= \frac{1}{\sigma^2} \left[|r_2|^2 - r_2^* \boldsymbol{a}_2^T \boldsymbol{h}_2 - r_2 \boldsymbol{h}_2^H \boldsymbol{a}_2^* + \boldsymbol{h}_2^H \boldsymbol{a}_2^* \boldsymbol{a}_2^T \boldsymbol{h}_2 \right] \\ &\quad + (\boldsymbol{h}_3^H - \alpha^* \boldsymbol{h}_2^H) \boldsymbol{C}^{-1} (\boldsymbol{h}_3 - \alpha \boldsymbol{h}_2) + \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 - |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\ &\quad - \alpha \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1 - \alpha^* \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\ &= \frac{1}{\sigma^2} |r_2|^2 - \frac{1}{\sigma^2} r_2^* \boldsymbol{a}_2^T \boldsymbol{h}_2 - \frac{1}{\sigma^2} r_2 \boldsymbol{h}_2^H \boldsymbol{a}_2^* + \frac{1}{\sigma^2} \boldsymbol{h}_2^H \boldsymbol{a}_2^* \boldsymbol{a}_2^T \boldsymbol{h}_2 \\ &\quad + \boldsymbol{h}_3^H \boldsymbol{C}^{-1} \boldsymbol{h}_3 - \alpha^* \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_3 - \alpha \boldsymbol{h}_3^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 + |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\ &\quad + \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 - |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 - \alpha \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1 - \alpha^* \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2. \end{aligned}$$

$|r_2|^2/\sigma^2$ can be removed in $\hat{\boldsymbol{A}}$ derivation because of their irrelevance to the maximization operation. $\boldsymbol{h}_3^H \boldsymbol{C}^{-1} \boldsymbol{h}_3$ can be moved out from $\tilde{\boldsymbol{E}}_2$ since it will be combined in $\boldsymbol{E}_3$. This simplifies $\tilde{\boldsymbol{E}}_2$ to

$$\begin{aligned}
& -\frac{1}{\sigma^2} r_2^* \boldsymbol{a}_2^T \boldsymbol{h}_2 - \frac{1}{\sigma^2} r_2 \boldsymbol{h}_2^H \boldsymbol{a}_2^* + \frac{1}{\sigma^2} \boldsymbol{h}_2^H \boldsymbol{a}_2^* \boldsymbol{a}_2^T \boldsymbol{h}_2 - \alpha^* \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_3 - \alpha \boldsymbol{h}_3^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\
& + |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 + \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{h}_2 - |\alpha|^2 \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 - \alpha \boldsymbol{h}_2^H \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1 - \alpha^* \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{C}^{-1} \boldsymbol{h}_2 \\
= \; & \boldsymbol{h}_2^H \left(\frac{1}{\sigma^2} \boldsymbol{a}_2^* \boldsymbol{a}_2^T + |\alpha|^2 \boldsymbol{C}^{-1} + \boldsymbol{C}^{-1} - |\alpha|^2 \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1} \right) \boldsymbol{h}_2 \\
& - \left(\alpha \boldsymbol{h}_3^H \boldsymbol{C}^{-1} + \frac{1}{\sigma^2} r_2^* \boldsymbol{a}_2^T + \alpha^* \boldsymbol{q}_1^H \boldsymbol{G}_1 \boldsymbol{C}^{-1} \right) \boldsymbol{h}_2 \\
& - \boldsymbol{h}_2^H \left(\alpha^* \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \frac{1}{\sigma^2} r_2 \boldsymbol{a}_2^* + \alpha \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1 \right).
\end{aligned}$$

Let

$$\begin{aligned}
\boldsymbol{G}_2^{-1} &= \boldsymbol{a}_2^* \boldsymbol{a}_2^T / \sigma^2 + |\alpha|^2 \boldsymbol{C}^{-1} + \boldsymbol{C}^{-1} - |\alpha|^2 \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{C}^{-1}, \\
\boldsymbol{q}_2 &= \frac{r_2 \boldsymbol{a}_2^*}{\sigma^2} + \alpha \boldsymbol{C}^{-1} \boldsymbol{G}_1 \boldsymbol{q}_1,
\end{aligned}$$

and $\boldsymbol{g}_2 = \boldsymbol{G}_2 \left(\alpha^* \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \boldsymbol{q}_2 \right)$. This reduces the above equation to

$$\boldsymbol{h}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{h}_2 - \boldsymbol{g}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{h}_2 - \boldsymbol{h}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{g}_2 = (\boldsymbol{h}_2 - \boldsymbol{g}_2)^H \boldsymbol{G}_2^{-1} (\boldsymbol{h}_2 - \boldsymbol{g}_2) - \boldsymbol{g}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{g}_2.$$

As $\boldsymbol{g}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{g}_2$ does not depend on $\boldsymbol{h}_2$, we obtain

$$\boldsymbol{E}_2 = (\boldsymbol{h}_2 - \boldsymbol{g}_2)^H \boldsymbol{G}_2^{-1} (\boldsymbol{h}_2 - \boldsymbol{g}_2).$$

Note that $\boldsymbol{g}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{g}_2$ is equal to

$$\begin{aligned}
\boldsymbol{g}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{g}_2 &= (\alpha^* \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \boldsymbol{q}_2)^H \boldsymbol{G}_2^H \boldsymbol{G}_2^{-1} \boldsymbol{G}_2 (\alpha^* \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \boldsymbol{q}_2) \\
&= (\alpha \boldsymbol{h}_3^H \boldsymbol{C}^{-1} + \boldsymbol{q}_2^H) \boldsymbol{G}_2 (\alpha^* \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \boldsymbol{q}_2) \\
&\quad \text{(Because } (\boldsymbol{G}_2^H \boldsymbol{G}_2^{-1})^H = (\boldsymbol{G}_2^{-1})^H \boldsymbol{G}_2 = \boldsymbol{G}_2^{-1} \boldsymbol{G}_2 = \boldsymbol{I}.) \\
&= |\alpha|^2 \boldsymbol{h}_3^H \boldsymbol{C}^{-1} \boldsymbol{G}_2 \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \alpha \boldsymbol{h}_3^H \boldsymbol{C}^{-1} \boldsymbol{G}_2 \boldsymbol{q}_2 + \alpha^* \boldsymbol{q}_2^H \boldsymbol{G}_2 \boldsymbol{C}^{-1} \boldsymbol{h}_3 + \boldsymbol{q}_2^H \boldsymbol{G}_2 \boldsymbol{q}_2.
\end{aligned}$$

$\boldsymbol{q}_2^H \boldsymbol{G}_2 \boldsymbol{q}_2$ in the above equation should be removed in $\boldsymbol{E}_3$ derivation because of their irrelevance to the integration operation (but it is relevant to the maximization operation). Therefore,

$$\begin{aligned}
I_2 &= \int_{\boldsymbol{h}_2} e^{-\left(\boldsymbol{E}_2 - \boldsymbol{q}_2^H \boldsymbol{G}_2 \boldsymbol{q}_2\right)} d\boldsymbol{h}_2 \\
&= \int_{\boldsymbol{h}_2} e^{-(\boldsymbol{h}_2 - \boldsymbol{g}_2)^H \boldsymbol{G}_2^{-1} (\boldsymbol{h}_2 - \boldsymbol{g}_2)} d\boldsymbol{h}_2 \cdot e^{\boldsymbol{q}_2^H \boldsymbol{G}_2 \boldsymbol{q}_2} \\
&= e^{\boldsymbol{q}_2^H \boldsymbol{G}_2 \boldsymbol{q}_2} |\boldsymbol{G}_2|.
\end{aligned}$$

Similarly, after we integrate over $\mathbf{h}_k$, for $2 \le k \le n$, we have

$$I_k = e^{\mathbf{q}_k^H \mathbf{G}_k \mathbf{q}_k} |\mathbf{G}_k|,$$

where

$$\begin{aligned}
\mathbf{G}_k^{-1} &= \frac{\mathbf{a}_k^* \mathbf{a}_k^T}{\sigma^2} + \mathbf{C}^{-1} + |\alpha|^2 (\mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{G}_{k-1} \mathbf{C}^{-1}) \text{ and} \\
\mathbf{q}_k &= \frac{r_k \mathbf{a}_k^*}{\sigma^2} + \alpha \mathbf{C}^{-1} \mathbf{G}_{k-1} \mathbf{q}_{k-1}.
\end{aligned}$$

Finally, we obtain the likelihood function as

$$f(\boldsymbol{r}|\mathbf{A}) = \prod_{k=1}^{n} e^{\mathbf{q}_k^H \mathbf{G}_k \mathbf{q}_k} |\mathbf{G}_k|. \tag{2.6}$$

As mentioned before, we are devoted to finding a block code that has minimum average error rate over the Gauss-Markov channel. Because the exact analysis of P_e is difficult, we choose an upper bound on P_e as the criterion instead. Let the code be with length N and rate $R = K/N$ information bits per code bits. That is, $\mathbf{S} = \{\mathbf{x}(1), \mathbf{x}(2), \cdots, \mathbf{x}(2^K)\} \subset \{0,1\}^N$ is the set of all codewords. Then, the average block error rate can be upper-bounded by union bound as

$$\begin{aligned} P_e &\triangleq \Pr(\hat{\mathbf{x}} \neq \mathbf{x}) \\ &= 2^{-K} \sum_i \Pr(\hat{\mathbf{x}} \neq \mathbf{x}(i) | \mathbf{x}(i) \text{ is transmitted}) \\ &\leq 2^{-K} \sum_i \sum_{j \neq i} p_{j|i}, \end{aligned} \tag{2.7}$$

where $\mathbf{x}$ is the transmitted codeword, $\hat{\mathbf{x}}$ is the decision at the receiver, and $p_{j|i}$ is the pairwise error probability of mistaken codeword $\mathbf{x}(j)$ when $\mathbf{x}(i)$ was transmitted. According to the maximum-likelihood decoding rule, we have

$$p_{j|i} = \Pr\left[\log \frac{\Pr[\mathbf{r}|\mathbf{x}(j)]}{\Pr[\mathbf{r}|\mathbf{x}(i)]} > 0\right] + \frac{1}{2} \Pr\left[\log \frac{\Pr[\mathbf{r}|\mathbf{x}(j)]}{\Pr[\mathbf{r}|\mathbf{x}(i)]} = 0\right]. \tag{2.8}$$

Now we are enable to derive the pairwise error probability on the condition that $\mathbf{x}(i)$ was transmitted and $\mathbf{x}(j)$ is received from substituting (2.6) into (2.8),

$$\begin{aligned} p_{j|i} &= \Pr\left[\log \frac{\prod_{k=1}^{N} e^{\mathbf{q}_k(j)^H \mathbf{G}_k(j)\mathbf{q}_k(j)} |\mathbf{G}_k(j)|}{\prod_{k=1}^{N} e^{\mathbf{q}_k(i)^H \mathbf{G}_k(i)\mathbf{q}_k(i)} |\mathbf{G}_k(i)|} > 0\right] \\ &\quad + \frac{1}{2}\Pr\left[\log \frac{\prod_{k=1}^{N} e^{\mathbf{q}_k(j)^H \mathbf{G}_k(j)\mathbf{q}_k(j)} |\mathbf{G}_k(j)|}{\prod_{k=1}^{N} e^{\mathbf{q}_k(i)^H \mathbf{G}_k(i)\mathbf{q}_k(i)} |\mathbf{G}_k(i)|} = 0\right] \\ &= \Pr\left[\sum_{k=1}^{N} \left(\mathbf{q}_k(j)^H \mathbf{G}_k(j)\mathbf{q}_k(j) - \mathbf{q}_k(i)^H \mathbf{G}_k(i)\mathbf{q}_k(i)\right) > \sum_{k=1}^{N} \left(\log \frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right] \\ &\quad + \frac{1}{2}\Pr\left[\sum_{k=1}^{N} \left(\mathbf{q}_k(j)^H \mathbf{G}_k(j)\mathbf{q}_k(j) - \mathbf{q}_k(i)^H \mathbf{G}_k(i)\mathbf{q}_k(i)\right) = \sum_{k=1}^{N} \left(\log \frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right]. \end{aligned} \tag{2.9}$$

Here we adopt the simulated annealing algorithm to search for good codes based on the above criterion. The reason is due to the fact that researchers have shown great successes of using simulated annealing algorithms to construct good source codes, error-correcting codes, spherical codes [10], and also codes combining channel estimation and error protection [7] by optimizing the cost functions like distortion of source codes, minimum distance, minimum separating angle, and union bound of block error probability, and so on.

2.2 *Simulated annealing algorithm*

"Anneal" means to heat and then to cool steel or glass. The cooling stage is slowly so that the potential energy stored in the molecular configuration can be minimized. A code structure is similar to the molecular constitution in the aspect of the analogy between good distance properties of a code and low potential energy of a molecular configuration. The outline of simulated annealing algorithm used to find good codes is given below.

```
Choose initial temperature T_0, code C
Do{
    Do{
        Choose C', a random perturbation of C
        Let Δe=energy(C')-energy(C)
        If (Δe<0) then C←C'
        Else with probability exp(-Δe/T) C←C'
        }
    Until (several energy drops or too many iterations)
    Lower temperature T← αT
    }
```

Until (stable code configuration is obtained or running time is up)

There are two loops in this algorithm. The outer one decreases the temperature by a factor α, $0 < \alpha < 1$, and terminates when the code achieves a stable configuration or a pre-specified running time is up. The inner one perturbs the code until a prior assigned number of energy drops (usually $3 \sim 5$) is reached or no change happens in long enough time so that the loop will finish in finite time. To perturb a code, we randomly select one or two bits in a codeword and flip them. In general, we use all-zero code as initial code since it gets randomized more quickly than other codes at high temperature. Next we compare the energy (or cost function) of C and C'. If energy becomes lower, we substitute C' for C. If energy does not decrease, we still take C' with probability exp(-Δe/T) to avoid becoming trapped in a local minimum. Therefore, temperature T is a control parameter for accepting the perturbed code. At high temperatures, exp(-Δe/T) closes to one, which means that we accept the new code with higher probability. On the other hand, at low temperatures exp(-Δe/T) depends more on the value of Δe. When Δe is large, exp(-Δe/T) approaches to zero, and we are not likely to accept the worse code. Since the energy function decides whether the code is adopted or not, it is important to choose a good one. The energy function should be designed to allow the code follow the direction of reducing its objective function; as a result, during the time when simulated annealing algorithm proceeds, the code has a tendency to decrease the energy. For example, for a source code we may set the energy function to be the distortion:

$$E = \frac{1}{2^n} \sum_{\mathbf{y}} \min_{\mathbf{x}} d_H(\mathbf{x}, \mathbf{y}).$$

And for a constant-weight code we might choose

$$E = \sum_{\mathbf{x} \neq \mathbf{y}} [d_H(\mathbf{x}, \mathbf{y})]^{-k}.$$

Being related to the union bound on the block error rate, the energy function of this work is defined as a function of $2^{-K} \sum_{i} \sum_{j \neq i} p_{j|i}$. In next chapter, we will start to investigate this criterion further, and then present the set of designed parameters in the algorithm in order to search for good code.

3. CODE DESIGN METHODOLOGY

Based on the simulated annealing algorithm introduced in Chapter 2, the main objective of this chapter is to present a feasible code design algorithm to search for a good code. We first present the approach to transform the pairwise error probability into a quadratic forms, then apply it as the criterion using in the simulated annealing algorithm.

3.1 Computing the Pairwise Error Probabilities

Given i, j, with $i \neq j$, we reform the q-matrix and G-matrix as

$$\mathbf{q}(j,i) = \begin{bmatrix} \mathbf{q}_1(j) \\ \mathbf{q}_2(j) \\ \vdots \\ \mathbf{q}_N(j) \\ \mathbf{q}_1(i) \\ \mathbf{q}_2(i) \\ \vdots \\ \mathbf{q}_N(i) \end{bmatrix}, \mathbf{G}(j,i) = \begin{bmatrix} \mathbf{G}_1(j) & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \mathbf{G}_2(j) & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \dots & 0 \\ 0 & 0 & \dots & \mathbf{G}_N(j) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & -\mathbf{G}_1(i) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & -\mathbf{G}_N(i) \end{bmatrix}.$$

Hence, (2.9) can be reformulated as

$$\begin{aligned} p_{j|i} &= \Pr\left[\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i) > \sum_{k=1}^{N}\left(\log\frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right] \\ &\quad +\frac{1}{2}\Pr\left[\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i) = \sum_{k=1}^{N}\left(\log\frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right], \end{aligned} \quad (3.1)$$

where $\mathbf{q}(j,i)$ is complex and Gaussian with mean $\mathbf{m}_q(j,i)$ and covariance

$\mathbf{S}_q(j,i)$. The mean and variance can be obtained as follows.

$$\begin{aligned}
\mathbf{m}_q(j,i) &= E\left[\mathbf{q}(j,i)\right] \\
&= \begin{bmatrix}
\frac{E[r_1]\mathbf{a}_1^*(j)}{\sigma^2} + \alpha\mathbf{C}^{-1}E[\mathbf{h}_0] \\
\frac{E[r_2]\mathbf{a}_2^*(j)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_1(j)E[\mathbf{q}_1(j)] \\
\vdots \\
\frac{E[r_N]\mathbf{a}_N^*(j)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{N-1}(j)E[\mathbf{q}_{N-1}(j)] \\
\frac{E[r_1]\mathbf{a}_1^*(i)}{\sigma^2} + \alpha\mathbf{C}^{-1}E[\mathbf{h}_0] \\
\frac{E[r_2]\mathbf{a}_2^*(i)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_1(i)E[\mathbf{q}_1(i)] \\
\vdots \\
\frac{E[r_N]\mathbf{a}_N^*(i)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{N-1}(i)E[\mathbf{q}_{N-1}(i)]
\end{bmatrix},
\end{aligned}$$

and

$$\mathbf{S}_q(j,i) = E\left[\mathbf{q}(j,i)\mathbf{q}^H(j,i)\right] - \mathbf{m}_q(j,i)\mathbf{m}_q^H(j,i).$$

Note that $f(\mathbf{h}_k|\mathbf{h}_{k-1}) \sim CN(\alpha\mathbf{h}_{k-1}, C)$,where $CN(\mathbf{m},\mathbf{c})$ denotes a complex Gaussian vector with mean $\mathbf{m}$ and covariance $\mathbf{c}$. When $\mathbf{h}_0$ is a constant vector, $\mathbf{h}_k$ is complex Gaussian with mean μ_k and covariance $\mathbf{S}_k$, where

$$\mu_k = E[\mathbf{h}_k] = E_{\mathbf{h}_{k-1}}[E_{\mathbf{h}_k}[\mathbf{h}_k|\mathbf{h}_{k-1}]] = E_{\mathbf{h}_{k-1}}[\alpha\mathbf{h}_{k-1}] = \alpha\mu_{k-1} = \alpha^k\mathbf{h}_0,$$

and

$$
\begin{aligned}
\mathbf{S}_k &= E[(\mathbf{h}_k - \mu_k)(\mathbf{h}_k - \mu_k)^H] \\
&= E[\mathbf{h}_k\mathbf{h}_k^H] - \mu_k E[\mathbf{h}_k^H] - E[\mathbf{h}_k]\mu_k^H + \mu_k\mu_k^H \\
&= E[(\alpha\mathbf{h}_{k-1} + \mathbf{v}_k)(\alpha\mathbf{h}_{k-1} + \mathbf{v}_k)^H] - |\alpha|^2\mu_{k-1}E[\mathbf{h}_{k-1}^H] - |\alpha|^2 E[\mathbf{h}_{k-1}]\mu_{k-1}^H \\
&\quad +|\alpha|^2\mu_{k-1}\mu_{k-1}^H (\text{because } \mathbf{h}_k = \alpha\mathbf{h}_{k-1} + \mathbf{v}_k, \mu_k = \alpha\mu_{k-1}) \\
&= |\alpha|^2 E[\mathbf{h}_{k-1}\mathbf{h}_{k-1}^H] + E[\mathbf{v}_k\mathbf{v}_k^H] - |\alpha|^2\mu_{k-1}E[\mathbf{h}_{k-1}^H] - |\alpha|^2 E[\mathbf{h}_{k-1}]\mu_{k-1}^H + |\alpha|^2\mu_{k-1}\mu_{k-1}^H \\
&= |\alpha|^2 E[(\mathbf{h}_{k-1} - \mu_{k-1})(\mathbf{h}_{k-1} - \mu_{k-1})^H] + \mathbf{C} \\
&= (1 + |\alpha|^2 + |\alpha|^4 + \cdots + |\alpha|^{2(k-1)})\mathbf{C} \\
&= \frac{1 - |\alpha|^{2k}}{1 - |\alpha|^2}\mathbf{C}.
\end{aligned}
$$

Because $|\alpha| < 1$, the steady-state probability of $\mathbf{h}_k$ (i.e. k$\rightarrow \infty$) is

$$f(\mathbf{h}_k) \sim CN\left(0, \frac{1}{1 - |\alpha|^2}\mathbf{C}\right).$$

Hence the mean vector of $\mathbf{q}(j, i)$ is given as

$$\mathbf{m}_q(j, i) = \mathbf{0},$$

and

$$
\begin{aligned}
& \mathbf{S}_q(j,i) \\
= \;& E\left[\mathbf{q}(j,i)\mathbf{q}^H(j,i)\right] \\
= \;& E\begin{bmatrix}
\mathbf{q}_1(j)\mathbf{q}_1^H(j) & \mathbf{q}_1(j)\mathbf{q}_2^H(j) & \dots & \mathbf{q}_1(j)\mathbf{q}_N^H(j) & \mathbf{q}_1(j)\mathbf{q}_1^H(i) & \dots & \dots & \mathbf{q}_1(j)\mathbf{q}_N^H(i) \\
\mathbf{q}_2(j)\mathbf{q}_1^H(j) & \mathbf{q}_2(j)\mathbf{q}_2^H(j) & \dots & \dots & \mathbf{q}_2(j)\mathbf{q}_1^H(i) & \dots & \dots & \mathbf{q}_2(j)\mathbf{q}_N^H(i) \\
\vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\mathbf{q}_N(j)\mathbf{q}_1^H(j) & \dots & \dots & \dots & \mathbf{q}_N(j)\mathbf{q}_N^H(i) & \dots & \dots & \vdots \\
\mathbf{q}_1(i)\mathbf{q}_1^H(j) & \mathbf{q}_1(i)\mathbf{q}_2^H(j) & \dots & \mathbf{q}_1(i)\mathbf{q}_N^H(j) & \mathbf{q}_1(i)\mathbf{q}_1^H(i) & \dots & \dots & \mathbf{q}_1(i)\mathbf{q}_N^H(i) \\
\mathbf{q}_2(i)\mathbf{q}_1^H(j) & \mathbf{q}_2(i)\mathbf{q}_2^H(j) & \dots & \dots & \mathbf{q}_2(i)\mathbf{q}_1^H(i) & \dots & \dots & \mathbf{q}_2(i)\mathbf{q}_N^H(i) \\
\vdots & \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\
\mathbf{q}_N(i)\mathbf{q}_1^H(j) & \dots & \dots & \dots & \mathbf{q}_N(i)\mathbf{q}_N^H(i) & \dots & \dots & \mathbf{q}_N(i)\mathbf{q}_N^H(i)
\end{bmatrix}.
\end{aligned}
$$

The general form of the (k,l) element in $\mathbf{S}_q(j,i)$ is given as

$$
\begin{aligned}
& E[\mathbf{q}_k(n)\mathbf{q}_l^H(m)] \\
= \; & E\left[\left(\frac{r_k\mathbf{a}_k^*(n)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{k-1}(n)\mathbf{q}_{k-1}(n)\right)\left(\frac{r_l\mathbf{a}_l^*(m)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{l-1}(m)\mathbf{q}_{l-1}(m)\right)^H\right] \\
= \; & E[r_k r_l^*]\frac{\mathbf{a}_k^*(n)\mathbf{a}_l^T(m)}{\sigma^4} + \frac{\alpha}{\sigma^2}\mathbf{C}^{-1}\mathbf{G}_{k-1}(n)E\left[\mathbf{q}_{k-1}(n)r_l^*\right]\mathbf{a}_l^T(m) \\
& + \frac{\alpha^*}{\sigma^2}\mathbf{a}_k^*(n)E[r_k\mathbf{q}_{l-1}^H(m)]\mathbf{G}_{l-1}(m)\mathbf{C}^{-1} \\
& + |\alpha^2|\mathbf{C}^{-1}\mathbf{G}_{k-1}(n)E[\mathbf{q}_{k-1}(n)\mathbf{q}_{l-1}^H(m)]\mathbf{G}_{l-1}(m)\mathbf{C}^{-1} \\
= \; & \left(\mathbf{a}_k^T(n)E[\mathbf{h}_k\mathbf{h}_l^H]\mathbf{a}_l^*(m) + E[n_k n_l]\right)\frac{\mathbf{a}_k^*(n)\mathbf{a}_l^T(m)}{\sigma^4} \\
& + \frac{\alpha}{\sigma^2}\mathbf{C}^{-1}\mathbf{G}_{k-1}(n)E[\mathbf{q}_{k-1}(n)\mathbf{h}_l^H]\mathbf{a}_l^*(m)\mathbf{a}_l^T(m) \\
& + \frac{\alpha^*}{\sigma^2}\mathbf{a}_k^*(n)\mathbf{a}_k^T(n)E[\mathbf{h}_k\mathbf{q}_{l-1}^H(m)]\mathbf{G}_{l-1}(m)\mathbf{C}^{-1} \\
& + |\alpha|^2\mathbf{C}^{-1}\mathbf{G}_{k-1}(n)E[\mathbf{q}_{k-1}(n)\mathbf{q}_{l-1}^H(m)]\mathbf{G}_{l-1}(m)\mathbf{C}^{-1}. \qquad (3.2)
\end{aligned}
$$

It is noticeable that channel impulse response $\mathbf{h}_k$ and $\mathbf{q}_k$ are recursively correlated with their previous states, and therefore so are $E[\mathbf{h}_k\mathbf{h}_l^H]$, $E[\mathbf{q}_{k-1}(n)\mathbf{h}_l^H]$ and $E[\mathbf{h}_k\mathbf{q}_{l-1}^H(m)]$. After recursive substitutions,

$$
\begin{aligned}
E[\mathbf{h}_k\mathbf{h}_l^H] &= |\alpha|^2E[\mathbf{h}_{k-1}\mathbf{h}_{l-1}^H] + E[\mathbf{v}_k\mathbf{v}_l^H] \\
&= |\alpha|^2\left(|\alpha|^2E[\mathbf{h}_{k-2}\mathbf{h}_{l-2}^H] + E[\mathbf{v}_{k-1}\mathbf{v}_{l-1}^H]\right) + E[\mathbf{v}_k\mathbf{v}_l^H] \\
&= |\alpha|^{2k}E[\mathbf{h}_0\mathbf{h}_{l-k}^H] + |\alpha|^{2(k-1)}E[\mathbf{v}_1\mathbf{v}_{l-k+1}^H] + |\alpha|^{2(k-2)}E[\mathbf{v}_2\mathbf{v}_{l-k+2}^H] + \ldots
\end{aligned}
$$

$$= \begin{cases} |\alpha|^{2k}\mathbf{h}_0\mathbf{h}_0^H + \left(|\alpha|^{2(k-1)} + |\alpha|^{2(k-2)} + \ldots + |\alpha|^0\right)\mathbf{C} & \text{, if k=l} \\ |\alpha|^{2k}\mathbf{h}_0 E[\mathbf{h}_{l-k}^H] & \text{, if k}\neq\text{ l} \end{cases}$$

$$= \begin{cases} |\alpha|^{2k}\mathbf{h}_0\mathbf{h}_0^H + \frac{1-|\alpha|^{2k}}{1-|\alpha|^2}\mathbf{C} & \text{, if k=l} \\ |\alpha|^{k+l}\mathbf{h}_0\mathbf{h}_0^H & \text{, if k}\neq\text{ l} \end{cases} \tag{3.3}$$

and by replacing $\mathbf{q}_{k-1}$ with $\frac{r_{k-1}\mathbf{a}_{k-1}^*}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{G}_{k-2}\mathbf{q}_{k-2}$, r_{k-1} with $\mathbf{a}_{k-1}^T\mathbf{h}_{k-1} + n_{k-1}$ we have

$$\begin{aligned}
& E[\mathbf{q}_{k-1}(n)\mathbf{h}_l^H] \\
= \; & \frac{\mathbf{a}_{k-1}^*(n)\mathbf{a}_{k-1}^T(n)}{\sigma^2}E[\mathbf{h}_{k-1}\mathbf{h}_l^H] + \alpha\mathbf{C}^{-1}\mathbf{G}_{k-2}(n)E[\mathbf{q}_{k-2}(n)\mathbf{h}_l^H] \\
= \; & \frac{\mathbf{a}_{k-1}^*(n)\mathbf{a}_{k-1}^T(n)}{\sigma^2}E[\mathbf{h}_{k-1}\mathbf{h}_l^H] \\
& +\alpha\mathbf{C}^{-1}\mathbf{G}_{k-2}(n)\left(\frac{\mathbf{a}_{k-2}^*(n)\mathbf{a}_{k-2}^T(n)}{\sigma^2}E[\mathbf{h}_{k-2}\mathbf{h}_l^H] + \alpha\mathbf{C}^{-1}\mathbf{G}_{k-3}(n)E[\mathbf{q}_{k-3}(n)\mathbf{h}_l^H]\right) \\
= \; & \frac{\mathbf{a}_{k-1}^*(n)\mathbf{a}_{k-1}^T(n)}{\sigma^2}E[\mathbf{h}_{k-1}\mathbf{h}_l^H] + \alpha\mathbf{C}^{-1}\mathbf{G}_{k-2}(n)\frac{\mathbf{a}_{k-2}^*(n)\mathbf{a}_{k-2}^T(n)}{\sigma^2}E[\mathbf{h}_{k-2}\mathbf{h}_l^H] \\
& +(\alpha\mathbf{C}^{-1})^2\mathbf{G}_{k-2}(n)\mathbf{G}_{k-3}(n)\frac{\mathbf{a}_{k-3}^*(n)\mathbf{a}_{k-3}^T(n)}{\sigma^2}E[\mathbf{h}_{k-3}\mathbf{h}_l^H] + \ldots \\
& +(\alpha\mathbf{C}^{-1})^{k-2}\mathbf{G}_{k-2}(n)\mathbf{G}_{k-3}(n)\ldots\mathbf{G}_1(n)E[\mathbf{q}_1(n)\mathbf{h}_l^H]
\end{aligned} \tag{3.4}$$

with

$$\begin{aligned}
E[\mathbf{q}_1(n)\mathbf{h}_l^H] &= E\left[\left(\frac{r_1\mathbf{a}_1^*(n)}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{h}_0\right)\mathbf{h}_l^H\right] \\
&= E\left[\frac{\mathbf{a}_1^T(n)\mathbf{a}_1^*(n)\mathbf{h}_1\mathbf{h}_l^H}{\sigma^2} + \frac{\mathbf{a}_1^*(n)\mathbf{n}_1\mathbf{h}_l^H}{\sigma^2} + \alpha\mathbf{C}^{-1}\mathbf{h}_0\mathbf{h}_l^H\right] \\
&= \frac{\mathbf{a}_1^T(n)\mathbf{a}_1^*(n)}{\sigma^2}\left(|\alpha|^{1+l}\mathbf{h}_0\mathbf{h}_0^T + \frac{1-|\alpha|^l}{1-|\alpha|}\mathbf{C}\right) + |\alpha|^{1+l}\mathbf{h}_0\mathbf{h}_0^T
\end{aligned}$$

and $E[\mathbf{h}_k\mathbf{q}_{l-1}^H(m)]$ is intuitively equal to the Hermitian of $E[\mathbf{q}_{k-1}(n)\mathbf{h}_l^H]$ with some notation changes.

Since $\mathbf{S}_q(j,i)$ is real and symmetric, it can be represented by two real and symmetric square roots $\mathbf{S}_q^{1/2}(j,i)$:

$$\mathbf{S}_q(j,i) = \mathbf{S}_q^{1/2}(j,i)\mathbf{S}_q^{1/2}(j,i).$$

We first concentrate on the 2MN×2MN matrix

$$\mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i).$$

Since it is also real and symmetric it can be expressed by its eigenvalues λ_n and eigenvectors $\mathbf{k}_n$ as

$$\mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i) = \sum_{n=1}^{2MN} \lambda_n \mathbf{k}_n \mathbf{k}_n^T, \tag{3.5}$$

where the eigenvalues are all real and the eigenvectors can always be chosen to be orthonormal.

From (3.5), $\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i)$ in (3.1) can be further specified as

$$\begin{aligned}
\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i) &= \left(\mathbf{S}_q^{-1/2}(j,i)\mathbf{q}(j,i)\right)^H \mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i)\left(\mathbf{S}_q^{-1/2}(j,i)\mathbf{q}(j,i)\right) \\
&= \sum_{n=1}^{2MN} \lambda_n |\mathbf{k}_n^T \mathbf{S}_q^{-1/2}(j,i)\mathbf{q}(j,i)|^2 \\
&= \sum_{n=1}^{2MN} \lambda_n |X_n|^2, \qquad (3.6)
\end{aligned}$$

where $X_n = \mathbf{k}_n^T \mathbf{S}_q^{-1/2}(j,i)\mathbf{q}(j,i)$ is complex Gaussian with mean

$$\mathbf{k}_n^T \mathbf{S}_q^{-1/2}(j,i) E[\mathbf{q}(j,i)] = \mathbf{0},$$

variance

$$\mathbf{k}_n^T \mathbf{S}_q^{-1/2}(j,i) E[\mathbf{q}(j,i)\mathbf{q}(j,i)^H] \mathbf{S}_q^{-1/2}(j,i)\mathbf{k}_n = \mathbf{k}_n^T \mathbf{k}_n = 1,$$

and is independent in n. Hence,

$$p_{j|i} = \Pr\left[\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i) > \sum_{k=1}^{n}\left(\log \frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right]$$

is the probability that a linear combination of a set of independent chi-square random variables with two degrees of freedom is greater than the constant $\sum_{k=1}^{n}\left(\log \frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)$. Assuming that there are $\bar{L}$ different and nonzero eigenvalues $\bar{\lambda}_l$ of $\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i)$, and indicating each orders of multiplicity by κ_l, $l = 1, \ldots, \bar{L}$, the sum in (3.6) is the same as

$$\sum_{l=1}^{\bar{L}} \bar{\lambda}_l \chi^2(2\kappa_l), \tag{3.7}$$

where $\chi^2(2\kappa_l)$ denotes a central χ^2-variable with $2\kappa_l$ degrees of freedom.

So far we have found that to calculate the pairwise error probability is relevant to eigenvalues of $\mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i)$. However, since

$$\begin{aligned}
\mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i)\mathbf{k}_n &= \lambda_n \mathbf{k}_n \\
\mathbf{S}_q^{1/2}(j,i)\mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i)\mathbf{k}_n &= \lambda_n \mathbf{S}_q^{1/2}(j,i)\mathbf{k}_n \\
\mathbf{S}_q(j,i)\mathbf{G}(j,i)\left(\mathbf{S}_q^{1/2}(j,i)\mathbf{k}_n\right) &= \lambda_n \left(\mathbf{S}_q^{1/2}(j,i)\mathbf{k}_n\right)
\end{aligned}$$

we can obtain the eigenvalues from $\mathbf{S}_q(j,i)\mathbf{G}(j,i)$ instead of $\mathbf{S}_q^{1/2}(j,i)\mathbf{G}(j,i)\mathbf{S}_q^{1/2}(j,i)$ and do not need to calculate the square root $\mathbf{S}_q^{1/2}(j,i)$ of $\mathbf{S}_q(j,i)$ anymore.

Now since those 2MN χ^2-variables in (3.7) are independent to each others, the characteristic function $\phi(t)$ of $Y = \mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i)$ is

$$\begin{aligned}
\phi(t) &= E\left[e^{jtY}\right] \\
&= \prod_{l=1}^{\bar{L}} (1-2j\bar{\lambda}_l t)^{-\kappa_l/2}.
\end{aligned}$$

By taking inverse Fourier transform, the corresponding probability density function is

$$\begin{aligned}
f(Y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jtY}\phi(t)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-jtY}\prod_{l=1}^{\bar{L}} (1-2j\bar{\lambda}_l t)^{-\kappa_l/2}dt,
\end{aligned}$$

and the pairwise error probability $p_{j|i} = \Pr\left(Y > \sum_{k=1}^{n}\left(\log\frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right)$ can be obtained as

$$p_{j|i} = \frac{1}{2\pi}\int_{\sum_{k=1}^{n}\left(log\frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)}^{\infty}\int_{-\infty}^{\infty} e^{-jtY}\prod_{l=1}^{\bar{L}} (1-2j\bar{\lambda}_l t)^{-\kappa_l/2}dt. \tag{3.8}$$

Next we adopt a closed-form expression given by Imhof [11] as the solution of (3.8) by assuming that the eigenvalues have been ordered as

$$\bar{\lambda}_1 > \bar{\lambda}_2 > \ldots > \bar{\lambda}_p > 0 > \bar{\lambda}_{p+1} > \bar{\lambda}_{p+2} > \ldots > \bar{\lambda}_{\bar{L}}.$$

Let $n = \sum_{l=1}^{\bar{L}} \kappa_l$, then

$$\begin{aligned} p_{j|i} &= \Pr\left[\mathbf{q}^H(j,i)\mathbf{G}(j,i)\mathbf{q}(j,i) > \sum_{k=1}^{n}\left(\log\frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)\right] \\ &= \sum_{l=1}^{p}\frac{1}{(\kappa_l-1)!}\left[\frac{\partial^{\kappa_l-1}}{\partial x^{\kappa_l-1}}F_l(x)\right]_{x=\bar{\lambda}_l}, \end{aligned} \tag{3.9}$$

where

$$F_l(x) = x^{q-1}\exp\left[\frac{-\sum_{k=1}^{n}\left(\log\frac{|\mathbf{G}_k(i)|}{|\mathbf{G}_k(j)|}\right)}{2x}\right]\prod_{r=1,r\neq l}^{\bar{L}}(x-\bar{\lambda}_r)^{-\kappa r}. \tag{3.10}$$

3.2 Simulated Annealing Algorithm for Code design

In the above section, the formula of the pairwise error probability is derived. Substituting (3.9) and (3.10) into (2.7), it turns out to be our criterion for the simulated annealing algorithm. We define the energy function as

$$E = 2^{-K}\sum_{i=1}^{2^K}\sum_{j=1,j\neq i}^{2^K} p_{j|i},$$

which results in searching for a code having minimum average block error rate. The algorithm begins at temperature $T_0 = 10^3$ and terminates at

$T_f = 10^{-7}$ with the decreasing factor $\alpha = 0.995$. Each inner loop ends when five successive perturbations are accepted or 300 successive perturbations are not accepted. These parameters are chosen to allow the cooling stage be slow enough to achieve a stable configuration. To speed up the process of randomization at low temperature, we randomly flip three to five bits in a codeword in each perturbation rather than only one or two as at high temperatures. It seems to be a reasonable acceleration for code search because a code is always easily accepted when the temperature is high. Moreover, noting that perturbation of u_{th} codeword will only affect the pairwise error probabilities of those u_{th}-related pairs, i.e. $(1, u), (2, u), \ldots, (u-1, u), (u+1, u), \ldots, (2^k, u)$ and $(u, 1), (u, 2), \ldots, (u, u-1), (u, u+1), \ldots, (u, 2^k)$. we have no need to calculate all $2^K \cdot (2^K - 1)$ pairwise error probabilities. It saves about $1 - \frac{2 \cdot (2^K - 1)}{2^K \cdot (2^K - 1)} = 1 - 2^{-K+1}$ portion of running time, but the cost is that all pairwise error probabilities must be recorded. With this condition, the energy function is now defined as

$$\begin{aligned} E(C') = E(C) \quad - \quad & 2^{-K} \sum_{i=1\ldots 2^K, i \neq u, (i,u) \in C} (p_{i|u} + p_{u|i}) \\ & + 2^{-K} \sum_{i'=1\ldots 2^K, i' \neq u', (i',u') \in C'} (p_{i'|u'} + p_{u'|i'}), \quad (3.11) \end{aligned}$$

where C represents the best code found so far, and C’ represents the resultant code after perturbation of C.

4. THE DESIGNED CODES AND THEIR PERFORMANCE

We have simulated two designed codes for different channel lengths of the Gauss-Markov channel. One is for M=1, a singular path model, and the other for M=2, a channel with two fading paths. The codes were first encoded as BPSK signals, suffered fading through the Gauss-Markov channel, and then decoded according to the optimal rule

$$\hat{\mathbf{a}}(\mathbf{r}) = \mathbf{a}(i'), \text{ where } i' = \arg\max_{i \in 2^K} \left(\sum_{k=1}^{N} \left(\mathbf{q}_k(i)^H \mathbf{G}_k(i) \mathbf{q}_k(i) + \log |\mathbf{G}_k(i)| \right) \right). \tag{4.1}$$

The initial channel coefficient vector $\mathbf{h}_0$ assumed to be known at transmitter and receiver is given as $\mathbf{1}$, and $\mathbf{h}_k = 0.9\mathbf{h}_{k-1} + \mathbf{v}_k$, where $\mathbf{v}_k$ has mean $\mathbf{d}$=$\mathbf{0}$ and covariance

$$\mathbf{C} = v \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0.001 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Let the design SNR be at 10 dB. Then the variance σ^2 of n_k can be derived from

$$
\begin{aligned}
\overline{SNR} &= \frac{\sum_{k=1}^{N} E\left[|\mathbf{h}_k|^2\right]}{\sum_{k=1}^{N} E\left[|n_k|^2\right]} \\
&= \frac{E\left[\sum_{k=2}^{N}\left(|h_{k,1}|^2+|h_{k,2}|^2\right)+|h_{1,1}|^2\right]}{N\sigma^2} \\
&= \frac{\sum_{k=2}^{N}\left(\frac{1-|\alpha|^{2k}}{1-|\alpha|^2}v+|\alpha|^{2k}|h_{0,1}|^2\right)+\sum_{k=2}^{N}\left(\frac{1-|\alpha|^{2k}}{1-|\alpha|^2}v+|\alpha|^{2k}|h_{0,2}|^2\right)+\frac{1-|\alpha|^2}{1-|\alpha|^2}v+|\alpha|^2|h_{0,1}|^2}{N\sigma^2} \\
&= \frac{\sum_{k=1}^{N}\left(\frac{1-0.9^{2k}}{1-0.9^2}0.001+0.9^{2k}\right)+\sum_{k=2}^{N}\left(\frac{1-0.9^{2k}}{1-0.9^2}0.001+0.9^{2k}\right)}{N\sigma^2},
\end{aligned}
\tag{4.2}
$$

where N is the length of the code.

We compared the performance of our designed codes to the code presented in [7]. Figure 4.1 and 4.2 shows the word error rate (WER) for the designed $(10,5)$ code with N=10, K=5 when the channel length M=1 and M=2, respectively. Both codes use the optimal decoding rule given in (4.1). It can be observed that our designed code indeed outperforms the code given in [7] on Gauss-Markov channels. The coding gain is about 3.5 dB at WER $= 10^{-2}$ when M=1 and it is enlarged to 6 dB when M=2. Hence, when the channel becomes multipath, the performance gain of the designed code is almost double to that of the code given in [7] even though both codes perform worse than on the single path channel (M=1).

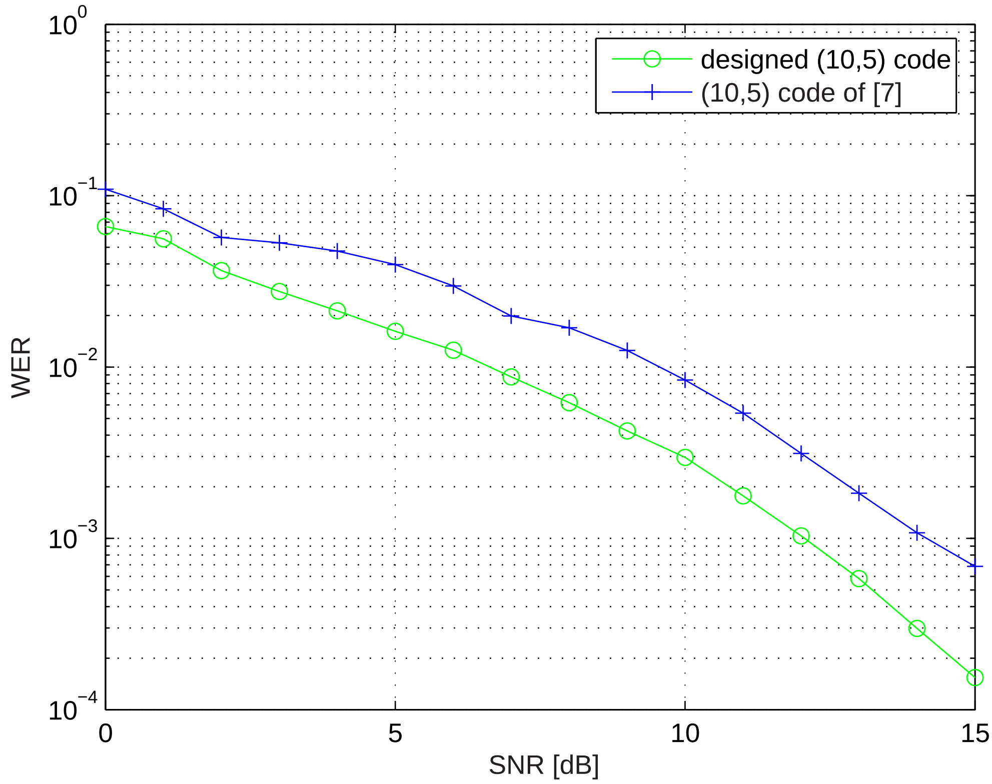

Fig. 4.1: Word error rate for (10,5) code on the Gauss-Markov channel (M=1), where SNR is calculated by (4.2).

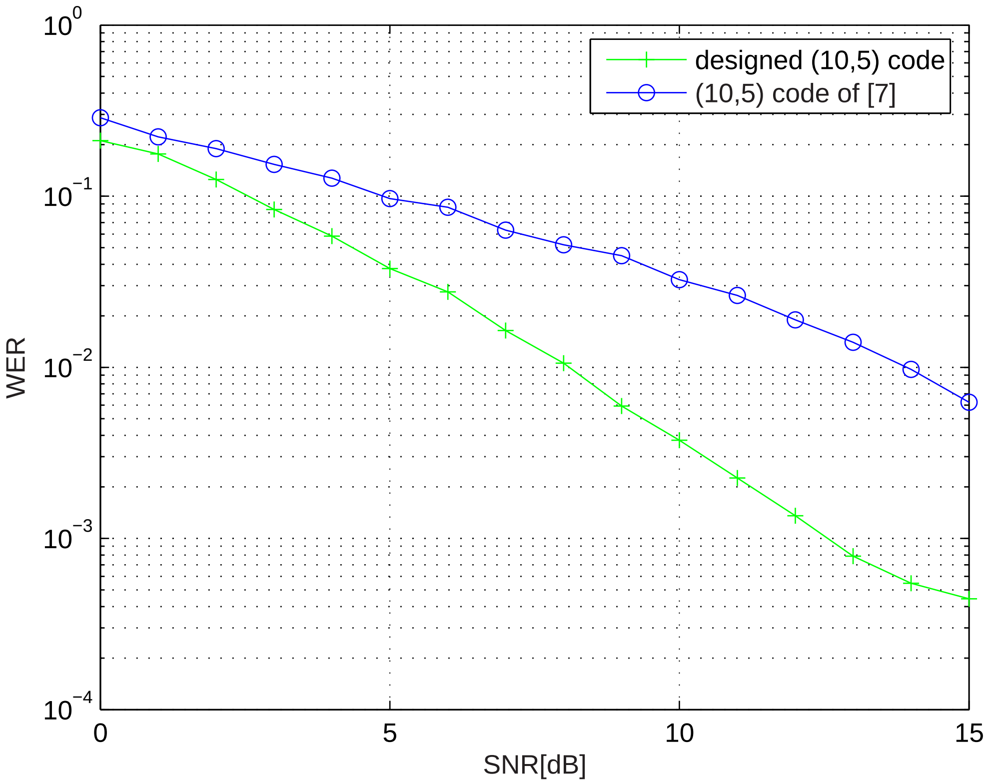

Fig. 4.2: Word error rate for (10,5) code on the Gauss-Markov channel (M=2).

Tab. 4.1: comparison of complexity

Criterion for Gauss-Markov channel			Criterion of [7]		
	number of multiplication	number of addition		number of multiplication	number of addition
$\mathbf{Q}(j,i)$	24N-16	8N-6	$\mathbf{P}_B(i)$	$2N^2$+12N+13	N^2+8N+4
$\mathbf{G}(j,i)$	54N-40	18N-12	$\mathbf{Q}(j,i)$	0	N^2+2N+1
$\mathbf{S}_q(j,i)$	$\frac{4}{3}N^3+\frac{187}{6}N^2$ $-\frac{241}{6}N+23$	$4N^2$+8N+2	$\mathbf{S}_y(i)$	$2N^2$+9N+7	$2N^2$+6N+4
$\mathbf{S}_q(j,i)\cdot\mathbf{G}(j,i)$	$64N^3$	$64N^3-16N^2$	$\mathbf{S}_y(i)\cdot\mathbf{Q}(j,i)$	$(N+1)^3$	$N(N+1)^2$

The cost of the lower error rate is the tradeoff of higher complexity of the system. In order to make the code be adapted to the fast-varying characteristic of channel, the code search algorithm is designed to deal with this characteristic such that the algorithm is more complicated. Table 4.1 lists the comparison of complexities between the major computations of our criterion and that given in [7]. The complexity is counted by how many multiplications and additions are taken in the process. Table 4.1 records the computation complexity needed of once energy function calculation in simulation annealing algorithm for a code with length N. We may take an example of $N = 10$, the codes demonstrated in both figures. The total numbers of multiplications and additions are 68795 and 63124 respectively for our criterion, and 1961 and 1779 respectively for that given in [7].

5. CONCLUSION AND FUTURE WORK

In this work, we describe the Gauss-Markov channel and its likelihood function which is a recursive form of every prior state. Then we derive the pairwise error probability as a function of the criterion for the simulated annealing algorithm. By the simulated annealing algorithm we found some codes that have better WER performance than those obtained from [7]. Our designed codes provide a coding gain of about 3.5 dB and 6 dB on the Gauss-Markov channels with channel lengths 1 and 2 respectively over those given in [7] at WER= 10^{-2}. During the process to obtain this criterion, we found that the complicate characteristic of the channel increases the operations needed for code search drastically. Even though We have performed some reductions on the criterion to speed up the algorithm, it still takes a lot of time to search for good codes. The nature future work will be to further speed up the algorithm by further simplifying the code search criterion.

BIBLIOGRAPHY

[1] ANSI/IEEE Std 802.11, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1999 Edition.

[2] Theodore S. Rappaport, *Wireless Communications: Principle and Practice*, Prectice Hall, 1996.

[3] L. Ljung, *System Identification–Theory for the user*, Englewood Cliffs, NJ: Prentice Hall, 1987.

[4] J. G. Proakis, "Adaptive algorithms for blind channel equalization," *Linear Algebra for Signal Processing*, A. Bojanczyk and G. Cybenko,Eds. New York: Springer-Verlag, 1995.

[5] E. de Carvalho and D. Slock, "Semi-blind maximum-likelihood multichannel estimation with Gaussian prior for the symbols using soft decisions," *Proc. IEEE Vehicular Technology Conf.*, pp. 1563-1567, May 1998.

[6] B. C. Ng, D. Gesbert and A. Paulraj, "A semi-blind approach to structured channel equalization," *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Seattle, WA, pp. 3385-3388, 1998.

[7] J. Giese, S. Parkvall and M. Skoglund, "Code design for combined channel estimation and error protection," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1162-1171, May 2002.

[8] K. Buckley, H. Chen and R. Perry, "Time-recursive maximum likelihood based sequence estimation for unknown ISI channels," *Signals, Systems and Computers*, vol. 2, pp. 1005-1009, Nov. 2000.

[9] W. A. Berger, K. Buckley and R. Perry, "EM algorithm for sequence estimation over Gauss-Markov ISI channels," *Communications, IEEE International Conference*, vol. 1, pp. 18-20, June 2000.

[10] A. E. Gamal, L. Hemachandra, I. Shperling and V. Wei, "Using simulated annealing to design good codes," *Information Theory, IEEE Transactions*, vol. IT-33, pp. 116-123, Jan. 1987.

[11] J. P. Imhof, "Computing the distribution of quadratic forms in normal variables," *Biometrica*, vol. 48, no. 3-4, pp. 419-426, 1961.