Doctoral Dissertation

Good Universal Codes Can Be Obtained by Concatenated Codes

Tomohiko UYEMATSU

Supervisor: Associate Professor Tomohiko Uyematsu

School of Information Science
Japan Advanced Institute of Science and Technology

February 15, 1997

Abstract

This paper investigates the error correcting capabilities of concatenated codes employing maximum distance separable (MDS) codes as outer codes and time-varying randomly selected constant composition inner codes, used on discrete memoryless channels with modified maximum mutual information decoding. It is proved that such code can achieve Gallager's random coding error exponent for all rates, while both encoding and decoding of the codes do not depend on the channel.

Acknowledgments

The author wishes to express his sincere gratitude to his principal advisor Associate Professor Tomohiko Uyematsu of Japan Advenced Institute of Science and Techology for his constant encouragement and kind guidance during this work. The author would like to thank his advisor Professor Eiji Okamoto of Japan Advanced Institute of Science and Technology for his helpful discussions and suggestions.

The author also wishes to express his thanks to Associate Professor Kunihiko Hiraishi and Mineo Kaneko of Japan Advanced Institute of Science and Technology for their suggestions and continuous encouragements.

The author is grateful to Professor Yoji Kajitani of Tokyo Institute of Technology for their helpful suggestions and discussions.

The author is grateful to all who have affected or suggested his areas of research. Professor Te Sun Han of Electro-communication university inspired the author through his constant activities on information theory, and gave the author valuable suggestions and kind encouragements.

The author devotes his sincere thanks and appreciation to all of them, and his colleagues.

Contents

Chapter 1

Introduction

The problem of specifying the structure of codes which achieve Gallager's error exponent remains as a major problem in information theory and coding theory. On this subject, Thommesen[8] investigated the concatenated codes with maximum distance separable (MDS) outer codes and time-varying inner codes used on DMC with maximum likelihood decoding. He has proved that Gallager's error exponents are asymptotically obtained for all rates by such codes. On the other hand, Ahlswede and Dueck[9] investigated the code called "permutation code" obtained by permutating coordinates of a single codeword. They have proved that Gallager's error exponents are asymptotically obtained for all rates less than capacity, by employing maximum mutual information (MMI) decoding. Since both encoding and decoding of their codes do not depend on particular channel, and their codes yield the best asymptotic performance, they are usually called as "universal codes". However, it is not known whether universal codes can be obtained by other structures of codes.

This paper extends the results of Thommesen, and shows that there exist good universal codes in a class of concatenated codes. We investigate the decoding error probability of concatenated codes employing MDS codes as outer codes and time-varying randomly selected constant composition codes as inner codes, used on arbitrary DMC with modified MMI decoding. It is shown that Gallager's error exponents are universally obtained for all rates by the proposed codes, provided that the length of the code is sufficiently large.

Chapter 2

Preliminaries

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite set. P will denote a distribution on $\mathcal{X}$, and V and W will denote discrete memoryless channels (DMC's), that is, stochastic matrices with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. In order to specify the input and output alphabets of a DMC, we shall use the shorthand notation $V : \mathcal{X} \to \mathcal{Y}$.

The *type* of a sequence $\boldsymbol{x} \in \mathcal{X}^n$ is the distribution $P_{\boldsymbol{x}}$ on $\mathcal{X}$, where $P_{\boldsymbol{x}}(a)$ is given by

$$P_{\boldsymbol{x}}(a) = \frac{1}{n} \cdot (\text{number of occurrences of } a \in \mathcal{X} \text{ in } \boldsymbol{x}). \tag{2.1}$$

We shall write $\mathcal{P}_n$ for the set of types of sequences in $\mathcal{X}^n$. The *joint type* $P_{\boldsymbol{x},\boldsymbol{y}}$ of the two sequences $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$ is the distribution on $\mathcal{X} \times \mathcal{Y}$ defined similarly. The set of sequences of type P in $\mathcal{X}^n$ is denoted by T_P^n or T_P. Further, for every $\boldsymbol{x} \in \mathcal{X}^n$ and $\boldsymbol{y} \in \mathcal{Y}^n$, if $\boldsymbol{x}$ and $\boldsymbol{y}$ has joint type

$$P_{\boldsymbol{x},\boldsymbol{y}}(a,b) = P_{\boldsymbol{x}}(a)V(b|a), \tag{2.2}$$

then we shall say that $\boldsymbol{y}$ has *conditional type* V given $\boldsymbol{x}$. The set of such $\boldsymbol{y}$ will be denoted by $T_V(\boldsymbol{x})$. We shall say that P is a *type* of sequences in $\mathcal{X}$ if $T_P \neq \emptyset$, and Y is a *conditional type of sequences in $\mathcal{Y}^n$ given* $\boldsymbol{x}$ if $T_V(\boldsymbol{x}) \neq \emptyset$. We shall denote by $\mathcal{V}(P) = \mathcal{V}_n(P)$ the set of stochastic matrices $V : \mathcal{X} \to \mathcal{Y}$ such that $T_V(P) \neq \emptyset$ for the sequence $\boldsymbol{x}$ of type P.

We shall write $I(P,V)$ and $H(V|P)$ for the mutual information $I(X \wedge Y)$ and conditional entropy $H(Y|X)$, respectively, of random variables (RV's) X and Y such that X has distribution P and Y is connected with X by the channel V. Further, for arbitrary distribution P and Q on $\mathcal{X}$ and channels $V : \mathcal{X} \to \mathcal{Y}$, $W : \mathcal{X} \to \mathcal{Y}$, we denote by $D(P \parallel Q)$ and $D(V \parallel W|P)$, the information divergence

$$D(P \parallel Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}, \tag{2.3}$$

and the conditional information divergence

$$D(V \parallel W|P) \triangleq \sum_{x \in \mathcal{X}} P(x) D(V(\cdot|x) \parallel W(\cdot|x)), \tag{2.4}$$

respectively. From now on, all logarithms and exponentials are to the base two.

Denoting by W^n the n-th memoryless extension of W, we have

$$\begin{aligned} W^n(\boldsymbol{y}|\boldsymbol{x}) &= \exp\{-n[D(V \parallel W|P) + H(V|P)]\}, \\ &\qquad \text{if } \boldsymbol{x} \in T_P \text{ and } \boldsymbol{y} \in T_V(\boldsymbol{x}). \end{aligned} \tag{2.5}$$

For any distribution P over $\mathcal{X}$, Gallager defined the random coding error exponent[7] given by

$$E_r(R, P, W) = \max_{0 \leq \rho \leq 1} \left[E_o(\rho, P, W) - \rho R \right], \tag{2.6}$$

where $E_o(\rho, P, W)$ is the Gallager function

$$E_o(\rho, P, W) = -\log \sum_{y \in \mathcal{Y}} \left[\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \tag{2.7}$$

On the other hand, in case of fixed composition codes with type P, Csiszár, Körner and Marton gave another form of the random coding error exponent[10]

$$E_r(R, P, W) = \min_{V} (D(V \parallel W|P) + |I(P, V) - R|^{+}), \tag{2.8}$$

where minimum is taken over all stochastic matrices $V : \mathcal{X} \to \mathcal{Y}$, and $|x|^{+} = \max(x, 0)$. In what follows, we only consider fixed composition codes and use the above two definitions of the random coding error exponent interchangeably.

Chapter 3

Structure of Codes

3.1 Inner Encoders

The outer codes to be considered are maximum distance separable (MDS) codes. Let Γ denote an MDS code over $GF(q)$, where q is a prime power. Let K denotes its dimension and N its block length. Then the minimum distance D of Γ is given by $D = N - K + 1$. When we refer to the rate τ of the outer code, we mean the dimensionless rate $\tau = K/N$.

The symbols of the codewords in Γ are encoded into sequences in T_P^n by inner encoders, where P is a specified type of sequences. Suppose that all inner encoders have the same code length n, the inner encoders $g_1, \cdots, g_N$ are defined by the mappings

$$g_i : GF(q) \to T_P^n \qquad (1 \le i \le N),$$

where the inner encoders are not necessary to be one to one. The rate r of the inner encoders is defined by $r = \ln q/n$ (nats/symbol).

In what follows, the inner codes are selected randomly from the specified ensemble. Especially, we deal with the following P-ensemble as an ensemble of inner codes.

Definition 1 (P-ensemble) For a given outer code, a given type $P \in \mathcal{P}_n$, and a given block length n of the inner code, we select Nq sequences $g_i(u) \in T_P^n$, where $1 \le i \le N$ and $u \in GF(q)$, independently and uniformly over T_P^n.

3.2 Construction of Concatenated Codes

We employ Forney's concatenated encoding scheme[1] to construct codes. In the first stage, a message from the message set $\{1, 2, \cdots, q^K\}$ is mapped into the codeword $\boldsymbol{u} \in \Gamma$ by an outer encoder, and in the second stage $\boldsymbol{u}$ is mapped into the channel codeword $\boldsymbol{x} = g(\boldsymbol{u})$, where the mapping $g : [GF(q)]^N \to X^{nN}$ is defined by

$$g(\boldsymbol{u}) = (g_1(u_1), g_2(u_2), \cdots, g_N(u_N))$$

for $\boldsymbol{u} = (u_1, u_2, \cdots, u_N)$ $(\in [GF(q)]^N)$. Since the outer encoder can be assumed to be one to one, the receiver only has to estimate which codeword $\boldsymbol{u} \in \Gamma$ is sent. The resulting concatenated code is a block code of length $N_o = nN$ and rate $R_o = \tau r$ (nats/symbol).

3.3 Decoding Scheme

The encoder encodes all the messages into constant composition codewords with type P, we can employ maximum mutual information (MMI) decoding[10]. In MMI decoding, the channel output $\boldsymbol{y}$ is decoded as the message $\boldsymbol{x}$, if the corresponding codeword $\boldsymbol{x}$ maximizes the mutual information function

$$I(\boldsymbol{x} \wedge \boldsymbol{y}) \overset{\triangle}{=} I(P, V) \quad \text{if } \boldsymbol{y} \in T_V(\boldsymbol{x}) \text{ and } \boldsymbol{x} \in T_P. \tag{3.1}$$

However, for the proposed code, we employ the following modified MMI decoding. Let $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)$ $(y_i \in \mathcal{Y}^n)$ be a received word when a codeword $g(\boldsymbol{u}) = (g_1(u_1), g_2(u_2), \cdots, g_N(u_N))$ is transmitted. In this representation, we assume that the output y_i corresponds to the input $g_i(u_i)$. In modified MMI decoding, the channel output $\boldsymbol{y}$ is decoded as the message $\boldsymbol{u} \in \Gamma$, if the corresponding codeword $g(\boldsymbol{u})$ maximizes the sum of the mutual information between $g_i(u)$ and y_i, i.e.

$$\sum_{i=1}^{N} I(g_i(u_i) \wedge y_i). \tag{3.2}$$

Especially, for an encoder g, let the decoder φ be any function $\varphi : \mathcal{Y}^{nN} \longrightarrow \Gamma$ such that $\varphi(\boldsymbol{y}) = \boldsymbol{u}$ satisfies

$$\sum_{i=1}^{N} I(g_i(u_i) \wedge y_i) = \max_{\hat{\boldsymbol{u}} \in \Gamma} \sum_{i=1}^{N} I(g_i(\hat{u}_i) \wedge y_i) \tag{3.3}$$

where we assume that an ambiguous estimate is considered to be erroneous. It is easy to see that this decoding rule is independent of the channel.

For a DMC $W : \mathcal{X} \to \mathcal{Y}$, if $\boldsymbol{y} \in \mathcal{Y}^{nN}$ leads to an erroneous decoding of the message $\boldsymbol{u} \in \Gamma$, then both

$$y_i \in T_{V_i}(g(u_i)) \cap T_{\hat{V}_i}(g(\hat{u}_i)) \quad (i = 1, 2, \cdots, N) \tag{3.4}$$

and

$$\sum_{i=1}^{N} I(P, \hat{V}_i) \geq \sum_{i=1}^{N} I(P, V_i) \tag{3.5}$$

must hold for some $\hat{\boldsymbol{u}}(\neq \boldsymbol{u})$ in Γ, $\hat{V}_i$ and V_i belonging to $\mathcal{V}_n(P)$ $(1 \leq i \leq N)$. Hence, the probability of the decoding error $\Phi(\boldsymbol{u})$ can be given by

$$\begin{aligned} \Phi(\boldsymbol{u}) &= W^{nN}(\{\boldsymbol{y} : \varphi(\boldsymbol{y}) \neq \boldsymbol{u}\} | g(\boldsymbol{u})) \\ &\leq \sum_{\substack{V_1, \cdots, V_N, \hat{V}_1, \cdots, \hat{V}_N \in \mathcal{V}(P) \\ \sum_{i=1}^{N} I(P, \hat{V}_i) \geq \sum_{i=1}^{N} I(P, V_i)}} \sum_{\substack{\hat{\boldsymbol{u}} \in \Gamma \\ \hat{\boldsymbol{u}} \neq \boldsymbol{u}}} \prod_{i=1}^{N} W^n(T_{V_i}(g(u_i)) \cap T_{\hat{V}_i}(g(\hat{u}_i)) | g(u_i)), \end{aligned} \tag{3.6}$$

3.4 Preliminaries from Coding Theory

In this subsection, we shall introduce some results of coding theory.

Let $S(N) = \{1, 2, \ldots, N\}$, then for every nonempty subset I of $S(N)$ with

$$I = \{i_1, i_2, \cdots, i_{|I|}\}, \qquad (1 \leq i_1 < i_2 < \cdots < i_{|I|} \leq N),$$

we define $g_i(\boldsymbol{u})$ as

$$g_I(\boldsymbol{u}) = (g_{i_1}(u_{i_1}), g_{i_2}(u_{i_2}), \cdots, g_{i_{|I|}}(u_{i_{|I|}})).$$

Let $\Gamma(I)$ indicates the subset of codewords in Γ which have nonzeros in the positions specified by I and zeros outside these positions. By using $\Gamma(I)$, nonzero codewords in Γ can be written by

$$\Gamma \backslash \{\mathbf{0}\} = \bigcup_{\substack{I \subset S(N) \\ D \leq |I| \leq N}} \Gamma(I). \tag{3.7}$$

Thommesen [8] gives an upper bound on the number of codewords in $\Gamma(I)$, which is given by

$$|\Gamma(I)| \leq q^{|I|-D+1} \qquad (D \leq |I|). \tag{3.8}$$

Chapter 4

Conclusion

This paper evaluates the error correcting capabilities of concatenated codes employing MDS codes as outer codes and time-varying randomly selected inner codes, used on discrete memoryless channels with modified MMI decoding. It is proved that Gallager's random coding error exponent can be obtained for all rates by such codes.

Bibliography

[1] G. D. Forney, Jr. : "Concatenated Codes", MIT Press (1966).

[2] J. Justesen : "A Class of Constructive Asymptotically Good Algebraic Codes", IEEE Trans. on Inform. Theory, vol.IT-18, no.5, pp.652-656 (Sep. 1972).

[3] P. Delsarte and P. Piret :"Algebraic Constructions of Shannon Codes for Regular Channel", IEEE Trans. on Inform. Theory, vol.IT-28, no.4, pp.593-599 (July 1982).

[4] M. Steiner : "Constructive Codes for Arbitrary Discrete Memoryless Channels", IEEE Trans. on Inform. Theory, vol.IT-40, no.3, pp.929-934 (May 1994).

[5] V. D. Goppa : "Codes on Algebraic Curves", Soviet Math. Dokl. vol.24, pp.170-172 (1981).

[6] T. Uyematsu and E. Okamoto: "A Construction of Codes with Exponential Error Bounds on Arbitrary Discrete Memoryless Channels", submitted to IEEE Trans. on Inform. Theory.

[7] R. G. Gallager : "Information Theory and Reliable Communication", Wiley (1968).

[8] C. Thommesen : "Error-Correcting Capabilities of Concatenated Codes with MDS Outer Codes on Memoryless Channels with Maximum-Likelihood Decoding", IEEE Trans. Inform. Theory, vol. IT-33, no.5, pp. 632-640 (Sep. 1987).

[9] R. Ahlswede and G. Dueck: "Good Codes Can Be Produced by a Few Permutations", IEEE Trans. on Inform. Theory, vol.IT-28, no.3, pp.430-443 (May 1982).

[10] I. Csiszár and J. Körner: Information Theory, Coding Theorems for Discrete Memoryless Systems, Academic Press (1981).

Publications

[1] T. Uyematsu, K. Kikuchi and K. Sakaniwa: "Trellis Coded Modulation for Multilevel Photon Communication System," Proc. of Inter. Symp. on Inform. Theory and Its Applications '92, pp.582-587 (Nov. 1992).

[2] T. Uyematsu: "Efficient Maximum Likelihood Decoding Algorithms for Linear Codes over Z-Channel," IEICE Trans. Fundamentals, vol.E76-A, no.9, pp.1430-1436 (Sep. 1994).

[3] E. Okamoto, T. Uyematsu, M. Mambo: "Permutation Cipher Scheme Using Polynomials over a Field," IEICE Trans. on Information and Systems, E78-D, no.2, pp.138-142 (Feb. 1995).

[4] T. Uyematsu and E. Okamoto: "A Construction of Codes with Exponential Error Bounds on Arbitrary Discrete Memoryless Channels," submitted to IEEE Trans. on Information Theory.