Performance of Multistage Universal Decoder for Given Linear Codes over Additive Noise Channels

Tomohiko UYEMATSU[†] and Fumio KANAYA[‡]

[†] Dept. of Communications and Integrated Systems,
Tokyo Institute of Technology,
Ookayama, Meguro-ku, Tokyo 152-8552, Japan.
E-mail: uyematsu@ieee.org

[‡] Dept. of Information Science,
Shonan Institute of Technology,
Fujisawa-shi, Kanagawa 251-0046, Japan.
E-mail: fkanaya@info.shonan-it.ac.jp

Abstract

We propose a multilevel coding and multistage decoding scheme which is universal for a class of additive noise channels. Then, for a given pair of binary linear codes, we show an upper bound of the decoding error probability. Further, we clarify the condition that the proposed coding and decoding scheme achieves the capacity of the channel and that the decoding error probability vanishes exponentially with code length.

1. INTRODUCTION

A universal channel code [1] is a coding scheme whose encoding and decoding do not depend on the particular channel but can achieve the capacity of channel. Universal channel codes have been studied by many researchers [1, 2, 3]. These researches picked up a particular universal decoder for a class of channels (i.e. memoryless channels or finite-state channels) and then showed the existence of the (universal) encoder such that the pair of encoder and decoder can achieve the optimum error exponent asymptotically. Namely, they first fix a universal decoder, and then find a universal encoder specialized for the decoder. In this paper, we explore another direction in the research of universal channel codes, i.e. we first fix an encoder and universal decoder pair and then analyze the performance of this coding scheme. Especially, we propose a multilevel coding (MLC) and multistage decoding (MSD) scheme which is universal for a class of additive noise channels. Then, for a given pair of binary linear codes, we show an upper bound of the decoding error probability. Further, we clarify the condition that the proposed MLC/MSD scheme achieves the capacity of the channel and that the decoding error probability vanishes exponentially with code length. We believe that the obtained condition is not significant and almost all good binary codes can satisfy this condition.

2. PRELIMINARIES

Consider a discrete memoryless channel (DMC) with a common finite alphabet $\mathcal{X}$ for both input and output, where we assume that $\mathcal{X}$ is an additive group. A DMC can be characterized by the probability distribution function $W(y|x)$, where $x \in \mathcal{X}$ denotes the input to the channel, and $y \in \mathcal{X}$ denotes the output of the channel. We assume an additive noise channel which can be characterized by the transition probability $W(y|x) = Q(y-x)$, where Q is a distribution over $\mathcal{X}$. $W_N(\boldsymbol{y}|\boldsymbol{x})$ is the probability distribution for sequences of length N. Since the channel is memoryless, we have

$$W_N(\boldsymbol{y}|\boldsymbol{x}) = \prod_{n=1}^{N} Q(y_n - x_n),$$

where $\boldsymbol{y} = y_1 y_2 \cdots y_N \in \mathcal{X}^N$ and $\boldsymbol{x} = x_1 x_2 \cdots x_N \in \mathcal{X}^N$.

Let $C(\subset \mathcal{X}^N)$ denote the set of codewords of a linear code which is represented by its parity-check matrix A with L rows and N columns, $N > L$, such that $\boldsymbol{x} \in \mathcal{X}^N$ is a codeword if and only if $\boldsymbol{x}A^T = 0$. Let M denotes the number of different codewords in the code C, and the rate of the code is $R = (\log M)/N$. In what follows, we assume that the base of logarithm and exponent is two.

For a probability distribution P over the alphabet $\mathcal{X}$, the entropy of P is defined by

$$H(P) \triangleq -\sum_{a \in \mathcal{X}} P(a) \log P(a).$$

Similarly, for a distribution P over $\mathcal{X}$ and a conditional distribution $V(b|a)$, $(a, b \in \mathcal{X})$, the conditional entropy is defined by

$$H(V|P) = \sum_{a \in \mathcal{X}} P(a) H(V(\cdot|a)).$$

This work has been supported in part by Japan Society for the Promotion of Science under Grant-in-Aid for Scientific Research (C)(2)(No.15560319).

Further, for two probability distribution P_1 and P_2 over the finite set $\mathcal{Z}$, the divergence of distributions is defined by

$$D(P_1\|P_2) \overset{\triangle}{=} \sum_{a\in\mathcal{Z}} P_1(a)\log\frac{P_1(a)}{P_2(a)}.$$

The type [2] of a sequence $\boldsymbol{x} \in \mathcal{X}^N$ and the joint type of $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathcal{X}^N$ are the probability distributions $P_{\boldsymbol{x}}$ and $P_{\boldsymbol{xy}}$ defined by

$$\begin{aligned} P_{\boldsymbol{x}}(a) &= \frac{1}{N}|\{n : x_n = a,\ 1 \le n \le N\}|,\ \forall a \in \mathcal{X} \\ P_{\boldsymbol{xy}}(a,b) &= \frac{1}{N}|\{n : (x_n, y_n) = (a,b),\ 1 \le n \le N\}|, \\ &\quad \forall (a,b) \in \mathcal{X}^2. \end{aligned}$$

The set of possible types of sequences in $\boldsymbol{x} \in \mathcal{X}^N$ is denoted by $\mathcal{P}_N$.

For $P \in \mathcal{P}_N$, the type class $\{\boldsymbol{x} \in \mathcal{X}^N : P_{\boldsymbol{x}} = P\}$ will be denoted by T_P. Similarly, For any conditional distribution V, we define $T_V(\boldsymbol{x})$ as

$$\begin{aligned} T_V(\boldsymbol{x}) &= \{(\boldsymbol{x},\boldsymbol{y}) : \boldsymbol{x},\boldsymbol{y} \in \mathcal{X}^N, \\ &\qquad P_{\boldsymbol{xy}}(a,b) = V(b|a)P_{\boldsymbol{x}}(a),\ \forall (a,b) \in \mathcal{X}^2\}. \end{aligned}$$

3. DECODING ERROR PROBABILITY FOR ME DECODER

In this section, we shall show the performance bound on the minimum entropy decoder for additive channels.

The minimum entropy (ME) decoder is defined as follows.

Definition 1 (The ME decoder): [8] For a linear code C with length N and parity check matrix A, and given a received sequence $\boldsymbol{y} \in \mathcal{X}^N$, define the minimum entropy (ME) decoder $u : \mathcal{X}^N \to C$ by

$$u(\boldsymbol{y}) \overset{\triangle}{=} \boldsymbol{y} + \arg \min_{\substack{\boldsymbol{e}\in\mathcal{X}^N: \\ \boldsymbol{e}A^T=\boldsymbol{y}A^T}} H(P_{\boldsymbol{e}}), \qquad (1)$$

where $H(P_{\boldsymbol{e}})$ denotes the entropy of the type of $\boldsymbol{e} \in \mathcal{X}^N$.

The next theorem shows that for a given linear code, the performance of ME decoder for additive channels.

Theorem 1: Consider a linear code C with rate R and length N. Denote the decoding error probability by P_{me}, when a codeword of C is transmitted through the additive noise channel $W(y|x)$, and decoded by the ME decoder. Then, we have

$$P_{me} \le (N+1)^{2|\mathcal{X}|}\exp\left\{-N\,E_u(R + (\log\alpha)/N)\right\}, \quad (2)$$

where

$$\begin{aligned} E_u(R) &\overset{\triangle}{=} \min_P[D(P\|Q) \\ &\qquad +|\log|\mathcal{X}| - H(P) - R|^+], \qquad (3) \\ |x|^+ &= \max\{0, x\}, \\ \alpha &\overset{\triangle}{=} \max_{P\in\mathcal{P}_N : P(0)\neq 1}\frac{|C\cap T_P|}{|T_P|}|\mathcal{X}|^N\exp\{-NR\}, \\ &\qquad (4) \end{aligned}$$

and the minimum in (3) is taken over the probability distribution P over $\mathcal{X}$.

This theorem is a natural extension of the result obtained by Shulman and Feder [7] for binary linear codes. It should also be noted that the error exponent $E_u(R)$ coincides with the random coding error exponent [2] of the additive noise channel. This implies that as long as $(\log\alpha)/N$ is negligible small, the ME decoder is as good as maximum likelihood (ML) decoder for additive noise channels.

Proof of Theorem 1: We use the technique developed by Shulman and Feder [7] in order to bound the probability of decoding error.

For a given linear code C, we construct an ensemble $\mathcal{C}$ of linear codes as follows. Generate an ensemble $\mathcal{C}$ from the given code by including all possible permutations σ of the order of the symbols in the codewords, where the permutations are uniformly distributed. The probability of decoding error of each code in $\mathcal{C}$ is obviously equal to the probability of decoding error of the original code C (the occurrence probability of error vector is invariant to codeword and symbol permutation σ, since the channel is memoryless and the occurrence probability of error vector only depends on its type). Hence, the averaged probability of error over the ensemble $\mathcal{C}$ is equal to the probability of error of the original code C.

Over the ensemble $\mathcal{C}$, it is easy to see that for any nonzero vector $\boldsymbol{z} \in \mathcal{X}^n$ with its type P, we have

$$\begin{aligned} \Pr(\boldsymbol{z}A^T = \boldsymbol{O}) &\le \frac{\prod_{a\in\mathcal{X}}(NP(a))!}{N!}\times|C\cap T_P| \\ &= \frac{|C\cap T_P|}{|T_P|} \\ &\le \max_{P\in\mathcal{P}_N : P(0)\neq 1}\frac{|C\cap T_P|}{|T_P|} \\ &= \alpha|\mathcal{X}|^{-N}\exp\{NR\}, \qquad (5) \end{aligned}$$

where α is defined in (4). It should be noted that this bound does not depend on the type P of $\boldsymbol{z}$.

When the ME decoder is employed and the error $\boldsymbol{e} \in \mathcal{X}^n$ occurs, the probability of decoding error for the ensemble $\mathcal{C}$ is given by

$$P_{error}(\boldsymbol{e}) \quad = \sum_{\substack{\boldsymbol{e}'\neq\boldsymbol{e}: \\ H(P')\leq H(P)}} \Pr(\boldsymbol{e}A^T = \boldsymbol{e}'A^T),$$

where P and P' denote the type of $\boldsymbol{e}$ and $\boldsymbol{e}'$, respectively. Then, by using (5) and a standard method of type (see e.g. [2]), we have

$$\begin{aligned}
&P_{error}(\boldsymbol{e}) \\
&\leq \sum_{\substack{\boldsymbol{e}'\neq\boldsymbol{e}: \\ H(P')\leq H(P)}} \alpha|\mathcal{X}|^{-N}\exp\{NR\} \\
&\leq \alpha|\mathcal{X}|^{-N}\exp\{NR\} \sum_{P'\in\mathcal{P}_N:H(P')\leq H(P)} |T(P')| \\
&\leq \alpha|\mathcal{X}|^{-N}\exp\{NR\} \sum_{P'\in\mathcal{P}_N:H(P')\leq H(P)} \exp\{NH(P')\} \\
&\leq \alpha|\mathcal{X}|^{-N}\exp\{NR\} \sum_{P'\in\mathcal{P}_N:H(P')\leq H(P)} \exp\{NH(P)\} \\
&\leq \alpha|\mathcal{X}|^{-N}\exp\{NR\}(N+1)^{|\mathcal{X}|}\exp\{NH(P)\} \\
&= (N+1)^{|\mathcal{X}|}\exp\{-N(-H(P)+\log|\mathcal{X}| \\
&\quad -R-(\log\alpha)/N)\}. \qquad (6)
\end{aligned}$$

On the other hand, $P_{error}(\boldsymbol{e}) \leq (N+1)^{|\mathcal{X}|}$ is obvious. Hence, we have

$$\begin{aligned}
P_{error}(\boldsymbol{e}) \quad \leq \quad & (N+1)^{|\mathcal{X}|}\exp\{-N|\log|\mathcal{X}|-H(P) \\
& -R-(\log\alpha)/N|^+\}.
\end{aligned}$$

By using this inequality, the probability of decoding error for the ensemble $\mathcal{C}$ is given by

$$\begin{aligned}
&P_{me} = \sum_{\boldsymbol{e}\in\mathcal{X}^N} Q(\boldsymbol{e})P_{error}(\boldsymbol{e}) \\
&\leq (N+1)^{|\mathcal{X}|}\sum_{P\in\mathcal{P}_N}\exp\{-ND(P\|Q)\} \\
&\quad \times\exp\{-N|\log|\mathcal{X}|-H(P)-R-(\log\alpha)/N|^+\} \\
&\leq (N+1)^{2|\mathcal{X}|}\exp\{-N\min_{P\in\mathcal{P}_N}(D(P\|Q) \\
&\quad +|\log|\mathcal{X}|-H(P)-R-(\log\alpha)/N|^+)\} \\
&\leq (N+1)^{2|\mathcal{X}|}\exp\{-N\,E_u(R+(\log\alpha)/N)\}.
\end{aligned}$$

This completes the proof.

4. MULTILEVEL CODING AND MULTISTAGE UNIVERSAL DECODING

Usually, the complexity of the ME decoder is as high as that of the ML decoder. Here, in order to reduce the computational complexity of the decoder, we introduce the multistage decoding. We shall analyze the performance of the multistage decoding for a given set of binary codes, and clarify its error exponent.

For simplicity, assume that $\mathcal{X}$ has a structure of two dimensional vector space over F_2, i.e. $\mathcal{X} = F_2 \times F_2$. It should be noted that the result can be extended to the case where $\mathcal{X}$ is a higher dimensional vector space over F_2. Let C_i $(i=1,2)$ be a pair of binary linear codes with rate R_i and the common length N. Now consider the following multilevel coding (MLC) and multistage decoding (MSD).

At the transmitter, let $(x_1, x_2, \cdots, x_N) \in F_2^N$ and $(x'_1, x'_2, \cdots, x'_N) \in F_2^N$ be codewords of the code C_1 and C_2, respectively. Then, $((x_1,x'_1),(x_2,x'_2),\cdots,(x_N,x'_N)) \in \mathcal{X}^N$ is fed into the channel. The rate of this MLC is $R_o = R_1 + R_2$. For the received word

$$\boldsymbol{y} = ((\hat{y}_1,\hat{y}'_1),(\hat{y}_2,\hat{y}'_2),\cdots,(\hat{y}_N,\hat{y}'_N)) \in \mathcal{X}^N,$$

the receiver decodes the pair of codewords as follows. We first estimate the codeword $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$ of C_1 from a binary vector $\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N)$ by using the ME decoder.

Next, we estimate the codeword $(x'_1, x'_2, \cdots, x'_N)$ of C_2 from a binary vector $\boldsymbol{y}'' = (\hat{y}'_1, \hat{y}'_2, \cdots, \hat{y}'_N)$ and the error vector $\boldsymbol{e}$ estimated by the first decoder. The decoding of the code C_2 is done by the following decoding rule.

Definition 2 (The conditional ME decoder): For a linear code C_2 with length N and parity check matrix A_2, given a received sequence $\hat{\boldsymbol{y}}' \in F_2^N$ and the estimated error vector $\boldsymbol{e} \in F_2^N$ of the first stage decoder, define the conditional minimum entropy (ME) decoder $\tilde{u} : F_2^N \times F_2^N \to C_2$ by

$$\tilde{u}(\hat{\boldsymbol{y}}', \boldsymbol{e}) \overset{\triangle}{=} \hat{\boldsymbol{y}}' + \arg \min_{\substack{\boldsymbol{e}'\in F_2^N: \\ \boldsymbol{e}'A_2^t = \hat{\boldsymbol{y}}'A_2^t}} H(V_{\boldsymbol{e}\boldsymbol{e}'}|P_{\boldsymbol{e}}), \qquad (7)$$

where $V_{\boldsymbol{e}\boldsymbol{e}'}$ denotes the conditional probability defined by the joint type of $(\boldsymbol{e}, \boldsymbol{e}')$, i.e.

$$V_{\boldsymbol{e}\boldsymbol{e}'}(a|b) \overset{\triangle}{=} \frac{P_{\boldsymbol{e}\boldsymbol{e}'}(b,a)}{\sum_{a'\in F_2} P_{\boldsymbol{e}\boldsymbol{e}'}(b,a')}, \quad \forall(a,b)\in F_2^2.$$

The next theorem is our main result.

Theorem 2: Let $P_{me,1}$ be the decoding error probability for the code C_1, when the ME decoder is employed. Then, we have

$$P_{me,1} \leq (N+1)^4 \exp\left\{-N\,E_u(R_1 + (\log\alpha(C_1))/N)\right\}, \qquad (8)$$

where

$$\begin{aligned}E_u(R) &\stackrel{\triangle}{=} \min_P [D(P\|Q_o) + |1 - H(P) - R|^+],\\ \alpha(C_1) &\stackrel{\triangle}{=} \max_{0<w\leq N} \frac{S(C_1,w)}{\binom{N}{w}} \exp\{N(1-R_1)\},\\ Q_o(a) &\stackrel{\triangle}{=} Q(a,0)+Q(a,1), \quad \forall a \in F_2.\end{aligned}$$

and $S(C_1, w)$ denotes the number of codewords with the weight w. Further, denote the decoding error probability by $P_{me,2}$, when the code C_2 is decoded by the conditional ME decoder. Then, assuming that the first stage of decoding is successful, we have

$$P_{me,2} \leq (N+1)^6 \exp\left\{-N\, E'_u(R_2 + (\log \alpha(C_2))/N)\right\}, \tag{9}$$

where

$$E'_u(R) \stackrel{\triangle}{=} \min_{P,V} [D(P \cdot V\|Q) + |1 - H(V|P) - R|^+], \tag{10}$$

and minimum is taken over any pair of distribution P over F_2 and conditional distribution $V(b|a)$ $(a, b \in F_2)$.

This theorem implies that the total decoding error probability can be bounded by

$$P_{MSD} \leq (N+1)^6 \exp\{-NE(R_1,R_2)\},$$

where

$$\begin{aligned}E(R_1,R_2) &= \min[E_u(R_1 + (\log\alpha(C_1))/N),\\ &\qquad E'_u(R_2 + (\log\alpha(C_2))/N)].\end{aligned}$$

Hence, the decoding error probability vanishes exponentially, if a pair of linear codes satisfies

$$\lim_{N\to\infty} \frac{1}{N} \log \alpha(C_1) = \lim_{N\to\infty} \frac{1}{N} \log \alpha(C_2) = 0. \tag{11}$$

As is well known [9], the weight distribution of many codes C with length N and rate R are well approximated by the binomial distribution, i.e.

$$S(C,w) \approx \exp\{-NR\}\binom{N}{w}.$$

Further, many BCH codes satisfy

$$S(C,w) = \exp\{-NR\}\binom{N}{w}(1 + O(N^{-1/10})).$$

Hence, We believe that the condition (11) is not significant and almost all BCH codes can satisfy this condition.

Further, for the occurrence probability Q of error in the channel, let

$$W_o(b|a) \stackrel{\triangle}{=} \frac{Q(a,b)}{Q_o(a)} \quad \forall (a,b) \in F_2^2.$$

Then, we have $E_u(R_1) > 0$ whenever $0 \leq R_1 < 1 - H(Q_o)$ and $E'_u(R_2) > 0$ whenever $0 \leq R_2 < 1 - H(W_o|Q_o)$. Since $H(W_o|Q_o) + H(Q_o) = H(Q)$, for a given rate $R_o < 1 - H(Q)$, we can choose a pair of rate such that $R_o = R_1 + R_2$ and

$$R_1 < 1 - H(Q_o) \quad \text{and} \quad R_2 < 1 - H(W_o|Q_o).$$

This implies that a combination of multistage codes and MSD achieves the capacity of the channel as long as a pair of linear codes satisfies the condition (11).

Lastly, we investigate the complexity of the decoder. The decoding complexity of MSD is at most $\exp\{NR_1\} + \exp\{NR_2\}$, and is much smaller than $\exp\{N(R_1 + R_2)\}$ of ML decoder or ME decoder. Hence, MSD is more efficient than both ML decoder and ME decoder, although the error exponent of MSD may be smaller than that for ML decoder.

Proof of Theorem 2: Since the channel for the code C_1 is additive, (8) can be immediately obtained by Theorem 1. So, we shall only prove (9).

Assume that $\boldsymbol{e}_1 \in T_P$ is the error vector estimated by the first stage decoding. Then, by using a similar technique used in the proof of Theorem 1, the probability that the error $\boldsymbol{e} \in T_V(\boldsymbol{e}_1)$ cannot be estimated correctly is upper bounded by

$$\begin{aligned}&P_{error}(\boldsymbol{e}|\boldsymbol{e}_1)\\ &\leq \sum_{\substack{\boldsymbol{e}'\neq\boldsymbol{e}:\\ H(V'|P)\leq H(V|P)}} \alpha(C_2)\exp\{N(R_2-1)\}\\ &\leq \alpha(C_2)\exp\{N(R_2-1)\} \sum_{\substack{\boldsymbol{e}'\neq\boldsymbol{e}:\\ H(V'|P)\leq H(V|P)}} |T_{V'}(\boldsymbol{e}_1)|\\ &\leq \alpha(C_2)\exp\{N(R_2-1)\}\\ &\qquad \times \sum_{V':H(V'|P)\leq H(V|P)} \exp\{N\,H(V'|P)\}\\ &\leq \alpha(C_2)\exp\{N(R_2-1)\}\\ &\qquad \times \sum_{V':H(V'|P)\leq H(V|P)} \exp\{N\,H(V|P)\}\\ &\leq (N+1)^4\alpha(C_2)\exp\{N(R_2-1)\}\exp\{N\,H(V|P)\}\\ &\leq (N+1)^4\exp\{-N(1-H(V|P)\\ &\qquad -R_2 - (\log\alpha_2)/N)\}.\end{aligned}$$

Again, $P_{error}(\boldsymbol{e}|\boldsymbol{e}_1) \leq (N+1)^4$ is obvious. Hence,

$$\begin{aligned}P_{error}(\boldsymbol{e}|\boldsymbol{e}_1) &\leq (N+1)^4\exp\{-N|1-H(V|P)\\ &\qquad -R_2 - (\log\alpha_2)/N|^+\}.\end{aligned}$$

By using this inequality, assuming that the first stage decoding is successful, the probability of the second stage decoder can be bounded by

$$\begin{aligned}
&P_{me,2}\\
&= \sum_{\boldsymbol{e}\in F_2^N, \boldsymbol{e}_1\in F_2^N} Q(\boldsymbol{e},\boldsymbol{e}')P_{error}(\boldsymbol{e}|\boldsymbol{e}_1)\\
&\le \sum_{P\in\mathcal{P}, V\in\mathcal{V}(P)} \exp\{-D(P\cdot V\|Q)\}\\
&\quad \times(N+1)^4\exp\{-N|-H(V|P)-1+R_2\\
&\quad -(\log\alpha_2)/N|^+\}\\
&\le (N+1)^6\exp\{-N\min_{P,V}(D(P\cdot V\|Q)\\
&\quad +|1-H(V|P)-R_2-(\log\alpha_2)/N|^+)\}\\
&= (N+1)^6\exp\{-N\,E_u''(R_2+(\log\alpha_2)/N)\}.
\end{aligned}$$

This completes the proof of Theorem 2.

5. GENERALIZATION OF THE RESULT

In this section, we generalize Theorem 1 and Theorem 2 for more wide class of universal decoders. We first define the decoding function.

Definition 3: A function $f_N : \mathcal{X}^N \to R$ is said to be a *decoding function*, provided that the following two conditions are satisfied.
(Condition 1) For any $\boldsymbol{x}\in\mathcal{X}^N$,

$$\frac{1}{N}f_N(\boldsymbol{x}) \le H(P_{\boldsymbol{x}}).$$

(Condition 2) There exists a constant $\beta\ge 0$ (which may depend on N) such that $\lim_{N\to\infty}\beta/N = 0$ and

$$\sum_{\boldsymbol{x}\in\mathcal{X}^N}\exp\{-f_N(\boldsymbol{x})-\beta\}\le 1.$$

Each decoding function induces a corresponding universal decoder as follows.

Definition 4: For a decoding function f_N, and a linear code C with length N and parity check matrix A, and a received sequence $\boldsymbol{y}\in\mathcal{X}^N$, define the universal decoder $\hat{f}_N : \mathcal{X}^N \to C$ by

$$\hat{f}_N(\boldsymbol{y}) \overset{\triangle}{=} \boldsymbol{y} + \arg\min_{\substack{\boldsymbol{e}\in\mathcal{X}^N:\\ \boldsymbol{e}A^T=\boldsymbol{y}A^T}} f_N(P_{\boldsymbol{e}}).$$

Example 1: $f_N(\boldsymbol{x}) = NH(P_{\boldsymbol{x}})$ is a decoding function with $\beta = |\mathcal{X}|\log(N+1)$. Hence the ME decoder is a special case of the universal decoder.

Example 2: Ziv's complexity $f_Z(\boldsymbol{x})$ defined by

$$f_Z(\boldsymbol{x}) = c(\boldsymbol{x})\log c(\boldsymbol{x}),$$

is a decoding function, where $c(\boldsymbol{x})$ denotes the number of distinct phrases in the parsing of $\boldsymbol{x}$ induced by the incremental parsing [4].

Example 3: The length function of Rissanen's adaptive arithmetic code [10] defined by

$$f_R(\boldsymbol{x}) = NH(P_{\boldsymbol{x}}) + \log(N+1),$$

is a decoding function.

These examples shows that almost all length functions of universal lossless source codes satisfy Conditions 1 and 2, and can be used as a decoding function.

The next theorem is a generalization of Theorem 1.

Theorem 3: Consider a linear code C with rate R and length N. Denote the decoding error probability by $P_{\hat{f}}$, when a codeword of C is transmitted through the additive noise channel $W(y|x)$, and decoded by the universal decoder $\hat{f}_N$. Then, we have

$$P_{\hat{f}} \le (N+1)^{2|\mathcal{X}|}\exp\{-N\,E_u(R+(\beta+\log\alpha)/N)\}, \tag{12}$$

where $E_u(R)$ is defined in (3).

Proof of Theorem 3: This theorem can be proved in a similar manner as the proof of Theorem 1. According to Condition 2 of the decoding function, we have

$$\sum_{\boldsymbol{x}\in\mathcal{X}^N}\exp\{-f_N(\boldsymbol{x})-\beta\}\le 1,$$

which shows that there exists a prefix code such that $\boldsymbol{x}$ can be encoded into $\beta + f_N(\boldsymbol{x})$ bits. Hence, we have

$$\begin{aligned}
|\{\boldsymbol{e}'\in\mathcal{X}^N : f_N(\boldsymbol{e}')\le f_N(\boldsymbol{e})\}| &\le \exp\{\beta+f_N(\boldsymbol{e})\}\\
&\le \exp\{\beta+NH(P_{\boldsymbol{e}})\},
\end{aligned}$$

where the last inequality comes from Condition 1 of the decoding function. When the universal decoder $\hat{f}_N$ is employed and the error $\boldsymbol{e}\in T_P$ occurs, the average probability of decoding error over the ensemble $\mathcal{C}$ defined in the proof of Theorem 1 is given by

$$\begin{aligned}
&P_{error}(\boldsymbol{e})\\
&= \sum_{\substack{\boldsymbol{e}'\ne\boldsymbol{e}:\\ f_N(\boldsymbol{e}')\le f_N(\boldsymbol{e})}} \Pr(\boldsymbol{e}A^T=\boldsymbol{e}'A^T)\\
&\le \sum_{\substack{\boldsymbol{e}'\ne\boldsymbol{e}:\\ f_N(\boldsymbol{e}')\le f_N(\boldsymbol{e})}} \alpha|\mathcal{X}|^{-N}\exp\{NR\}\\
&\le \alpha|\mathcal{X}|^{-N}\exp\{NR\}\sum_{P'\in\mathcal{P}_N:H(P')\le H(P)}\exp\{\beta+NH(P)\}\\
&\le (N+1)^{|\mathcal{X}|}\exp\{-N(-H(P)+\log|\mathcal{X}|\\
&\quad -R-(\beta+\log\alpha)/N)\}.
\end{aligned}$$

By using this inequality instead of (6), we can prove Theorem 3 in a similar manner as Theorem 1.

Next, we consider a generalization of the conditional ME decoder.

Definition 5: A function $f_N : F_2^N \times F_2^N \to R$ is said to be a *conditional decoding function*, provided that the following two conditions are satisfied.
(Condition 1) For any $(\boldsymbol{x},\boldsymbol{y}) \in F_2^N \times F_2^N$,

$$\frac{1}{N} f_N(\boldsymbol{x},\boldsymbol{y}) \leq H(P_{\boldsymbol{x}\boldsymbol{y}}),$$

where $P_{\boldsymbol{x}\boldsymbol{y}}$ denotes the joint type of $(\boldsymbol{x},\boldsymbol{y})$.
(Condition 2) There exists a constant $\beta \geq 0$ such that $\lim_{N\to\infty} \beta/N = 0$ and

$$\sum_{\boldsymbol{y}\in F_2^N} \exp\{-f_N(\boldsymbol{x},\boldsymbol{y}) + H(P_{\boldsymbol{x}}) - \beta\} \leq 1, \quad \forall \boldsymbol{x} \in F_2^N.$$

Each conditional decoding function induces a corresponding universal decoder as follows.

Definition 6: In the MSD, for a conditional decoding function f_N, a linear code C_2 with length N and parity check matrix A_2, a received sequence $\hat{\boldsymbol{y}}' \in F_2^N$ and the estimated error vector $\boldsymbol{e} \in F_2^N$ of the first stage decoder, define the conditional universal decoder $\tilde{f}_N : F_2^N \times F_2^N \to C_2$ by

$$\tilde{f}_N(\boldsymbol{e},\hat{\boldsymbol{y}}') \overset{\triangle}{=} \hat{\boldsymbol{y}}' + \arg \min_{\substack{\boldsymbol{e}'\in F_2^N:\\ \boldsymbol{e}'A_2^t=\hat{\boldsymbol{y}}'A_2^t}} f(\boldsymbol{e},\boldsymbol{e}'). \tag{13}$$

Example 4: $f_N(\boldsymbol{x},\boldsymbol{y}) = N(H(V_{\boldsymbol{x}\boldsymbol{y}}|P_{\boldsymbol{x}}) + H(P_{\boldsymbol{x}}))$ is a decoding function with $\beta = 4\log(N+1)$. Since $H(P_{\boldsymbol{x}})$ does not depend on $\boldsymbol{y}$, for any given $\boldsymbol{x}$, $\boldsymbol{y}$ minimizing $f_N(\boldsymbol{x},\boldsymbol{y})$ also minimizes $H(V_{\boldsymbol{x}\boldsymbol{y}}|P_{\boldsymbol{x}})$. Hence, the conditional ME decoder is a special case of the conditional universal decoder.

Example 5: Let $c(\boldsymbol{x},\boldsymbol{y})$ denote LZ complexity [4] of the joint sequence $(\boldsymbol{x},\boldsymbol{y})$, that is, the number of phrases obtained by incremental parsing for the joint sequence. Then,

$$f_N(\boldsymbol{x},\boldsymbol{y}) = c(\boldsymbol{x},\boldsymbol{y}) \log c(\boldsymbol{x},\boldsymbol{y})$$

is a conditional decoding function.

The next theorem is a generalization of Theorem 2.

Theorem 4: In the MSD, denote the decoding error probability at the second stage by $P_{\hat{f},2}$, when the code C_2 with rate R_2 is decoded by the conditional universal decoder $\hat{f}_N$. Then, assuming that the first stage of decoding is successful, we have

$$P_{\hat{f},2} \leq (N+1)^6 \exp\left\{-N\, E'_u(R_2 + (\beta + \log\alpha(C_2))/N)\right\}, \tag{14}$$

where $E'_u(R)$ is defined in (10).

This theorem can be proved in a similar manner as Theorem 2, by using a similar argument in the proof of Theorem 3. So, we omit the proof.

References

[1] I. Csiszár, J. Körner and K. Marton, “A new look at the error exponent of discrete memoryless channels,” presented at the IEEE Int. Symp. Information Theory, Cornell Univ., Ithaca, NY, 1977.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

[3] J. Ziv, “Universal decoding for finite-state channels,” *IEEE Trans. on Inform. Theory,* vol.IT-31, pp.453-460, July 1985.

[4] J. Ziv and A. Lempel, “Compression of individual sequences via variable rate coding,” *IEEE Trans. on Inform. Theory*, vol.IT-24, pp.530-536, 1978.

[5] A. Lapidoth and J. Ziv, “On the universality of the LZ-based decoding algorithm,” *IEEE Trans. on Inform. Theory,* vol.44, pp.1746-1755, Sep. 1998.

[6] I. Csiszár and J. Körner, “Graph decomposition: a new key to coding theorems,” *IEEE Trans. on Inform. Theory,* vol. IT-27, no.1, pp.5-12, Jan. 1981.

[7] N. Shulman and M. Feder, “Random coding techniques for nonrandom codes,” *IEEE Trans. on Inform. Theory*, vol.45, no.6, pp.2101-2104, 1999.

[8] I. Csiszár, “Linear codes for source and source networks: error exponents, universal coding,” *IEEE Trans. on Inform. Theory,* vol.IT-28, pp.585-592, 1982.

[9] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes.* North-Holland, 1977.

[10] J. Rissanen, “Complexity of strings in the class of Markov sources,” *IEEE Trans. on Inform. Theory*, vol.IT-32, pp.526-532, 1986.