Coding Theorems for Point-to-Point Communication Systems using Sparse Matrix Codes

疎行列を用いた2端子通信系における符号化定理

Shigeki Miyake

三宅 茂樹

Contents

Acknowledgments

It is my pleasure to thank my advisor, Prof. Hirosuke Yamamoto, for his guidance, encouragement, and support. I am grateful to Dr. Fumio Kanaya for introducing this fascinating research area, Information Theory, to me. The results in this thesis are the product of joint work with Dr. Jun Muramatsu, who has been an invaluable colleague during the past years. Constructive comments and suggestions by the anonymous reviewers of [29], [30], [31], [32], and [33] have significantly improved the presentation of our results.

This research was supported by Nippon Telephone and Telegraph (NTT) Corporation, and I would like to acknowledge here the generosity of the organization. I would also like to thank Dr. Mitsuru Maruyama of NTT for his encouragement.

Finally, I wish to thank my family, who have always been the most important part of my life, and who have supported me throughout.

Chapter 1

Introduction

1.1 Background

Shannon showed the theoretical limit of coding performance and the existence of code that can attain this limit in his seminal paper that founded Information Theory [40]. However, since the random coding technique was used for the existence proof of a code, it was not clear how to construct the code. In the late 1950's, BCH code [3] [19] and Reed-Solomon code [38] were invented as channel codes. These codes can be implemented efficiently, but in the limit of long block length n, an asymptotically positive transmission rate ($R > 0$) and decoding error $\rightarrow 0$ ($as\ n \rightarrow \infty$) are not compatible. The first code with which the above properties are compatible was invented by Justesen in 1972 [20]. However, this code also could not attain the optimal rate shown by Shannon.

In the 1990's Turbo code was invented [1] and Low Density Parity Check (LDPC) code was rediscovered [15] [24] [23]. It was experimentally shown that Turbo code and LDPC code could be implemented efficiently and the coding rate could approach the optimal rate with an arbitrary small decoding error. Research on the construction of codes that can be implemented efficiently and, at the same time, approach an optimal rate shown by coding theorems have been an important problem in Information Theory.

Information Theory treats the problem of source coding and channel coding. For each problem, a point-to-point system, where there is one sender and one receiver of a message, and a multi-terminal system, where there is more than one sender or receiver, can be considered. A point-to-point sys-

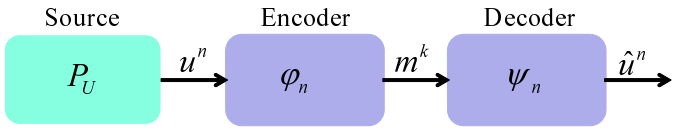

Figure 1.1: Lossless source coding system

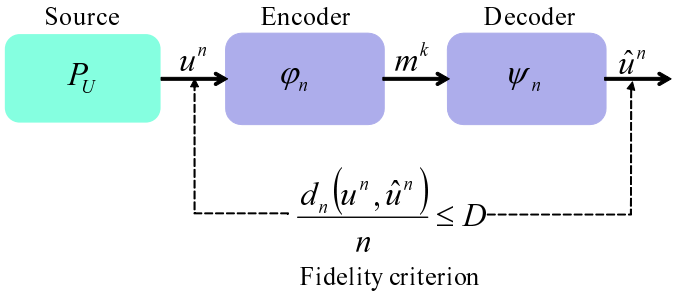

Figure 1.2: Lossy source coding system

tem is more fundamental. Figures 1.1 and 1.2 show lossless and lossy source coding problems of point-to-point system, respectively.

In the lossless source coding problem, the infimum value of the compression rate that asymptotically makes the decoding error to be 0 has been investigated. The value is known to be the entropy of the source $H(U)$ defined by $\sum_a P_U(a) \log \frac{1}{P_U(a)}$, where the source P_U is assumed to be stationary memoryless. In the lossy source coding problem, the infimum value of the compression rate that makes distortion between the original message sequence and the reproduction message sequence to be less than a given value D with high probability has been investigated. The value is known as the rate-distortion function $R(D)$ defined by $\min_{P_{V|U}:\sum_{a,b} P_{UV}(a,b)d(a,b)\le D} I(P_U, P_{V|U})$, where the source P_U is assumed to be stationary memoryless. Figure 1.3 shows the channel coding problem of a point-to-point system. In this problem, the supremum value of the transmission rate that asymptotically makes decoding error to be 0 has been investigated. The value is the channel capacity $C(W)$ defined by $\max_{P_X} I(P_X, W_{Y|X})$, where the channel $W_{X|Y}$ is assumed to be stationary memoryless. When the code that satisfying the condition that the decoding error asymptotically decrease to 0 for the lossless source

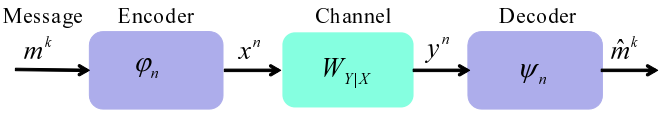

Figure 1.3: Channel coding system

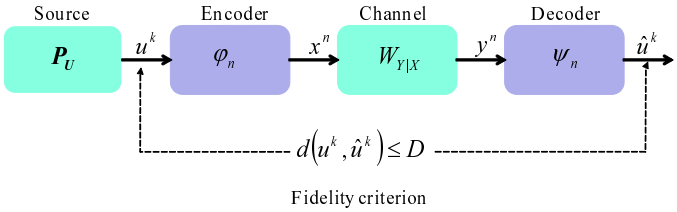

Figure 1.4: Joint source-channel coding system

coding or channel coding problem, or the condition that the distortion between the original message sequence and the reproduction message sequence is less than a given constant for the lossy source coding problem, achieves the optimal rate asymptotically, we say that the code has asymptotic optimality.

In real communication systems, we do not always know the statistical properties of the source or the channel. In this framework, the problem whether the code that asymptotically makes decoding error to be 0 exists, or if such code exists, how fast the error approaches 0, is called the problem of universal coding. For the universal source coding problem, the speed at which the decoding error approaches 0 (" error exponent") is strictly obtained, on the other hand, for the universal channel coding problem, the upper and lower bounds of the error exponent are known to have a gap (e.g. [7]).

When both the source and the channel constitute a communication system, and k outputs from the source are transmitted through n channel usage (Figure 1.4), the problem investigating the infimum value of n/k (Limit of the Minimum Transmission Ratio: LMTR), that makes the distortion between the original message sequence and the reproduction message sequence to be less than a given constant D with high probability, is called the joint (lossy) source-channel coding problem. LMTR is known to be $R(D)/C(W)$ when both the source and channel are assumed to be stationary memoryless.

For multi-terminal systems, there are various types of source coding and

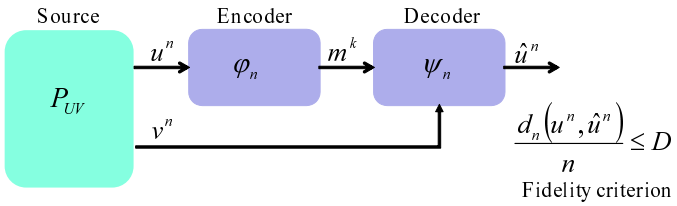

Figure 1.5: Wyner-Ziv system

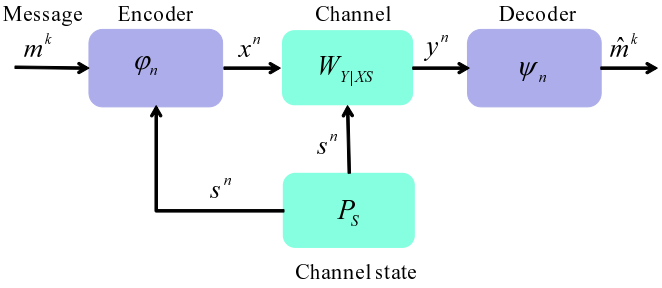

Figure 1.6: Gel'fand-Pinsker system

channel coding problems. For examples, the Wyner-Ziv system (Figure 1.5) shows the lossy source coding problem with side information at the decoder, and the Gel'fand-Pinsker system (Figure 1.6) shows the channel coding problem where the encoder knows the channel state. The former is expected to be applied to the sensor network [43], and the latter can be regarded as a model of steganography [26]. Although there exist many kinds of multi-terminal systems [10] [21] besides these systems, details of multi-terminal systems are omitted since the theme of this thesis is on point-to-point systems.

For various communication systems, existence of code with asymptotic optimality has been proven. Since the proofs are non-constructive using random coding arguments, construction of the code that has asymptotic optimality, and at the same time, can be implemented using an efficient algorithm has been an unsolved problem (note that for the lossless source coding problem, the variable length code that has both the above properties has been constructed (e.g. [48] [49])).

LDPC code was first proposed as a linear code using LDPC matrices for a given channel. Since LDPC matrices have $O(n)$ non-zero elements,

where n denotes the block length of the code, there exist efficient algorithms that approximate the maximum-likelihood (ML) decoding [15] [24]. Caire, Shamai, and Verdú [5] proposed a lossless universal source code that can be applied to sources with memory using multiple LDPC matrices and selecting the most favorable one by MDL (Minimum Description Length) [39] and showed high compression performance.

In a theoretical aspect, Miller and Burshtein [28] proved the asymptotic optimality of LDPC codes for channels whose noise distribution is symmetric such as the binary symmetric channel (BSC). For the lossy source coding problem, Matsunaga and Yamamoto [27] showed the asymptotic optimality of LDPC codes under the assumption of uniform distribution over the binary alphabet and using the Hamming distance as a distortion measure, and Miyake [29] extended their results to a non-binary alphabet case. Martinian and Wainwright [25] also showed the asymptotic optimality of LDPC codes under the same assumption, and extended their results to multi-terminal systems such as the Wyner-Ziv and the Gel'fand-Pinsker systems [26]. Note that all the results for the asymptotic optimality of LDPC codes were obtained under the assumption that the stochastic properties governing the system considered has symmetric properties such as uniform distribution over the alphabet or additive noise on the channel. To show the asymptotic optimality of LDPC codes for arbitrary stationary memoryless channels, which is not assumed to have symmetric properties, Bennatan [2] obtained the desired LDPC code after considering uniform distribution over a sufficiently large alphabet and constructing a map ("quantization map") from the large alphabet to the alphabet of the system under consideration. For lossy source coding problem with respect to an arbitrary stationary memoryless source, Gupta and Verdú [18] constructed a heuristic coding algorithm and showed asymptotic optimality.

1.2 Contributions

We constructed a lossless universal source code, lossy source code, and channel code using sparse matrices for stationary memoryless systems, and showed their error exponent (lossless universal code) or asymptotic optimality (lossy source code, channel code). It should be noted that the sparse matrix code constructed in this thesis can be applied to any discrete stationary memoryless sources and channels, which are not assumed to have symmetric prop-

erties of probability distributions of the communication models. Here we distinguish the word "sparse matrix" from "LDPC matrix". Since the number of non-zero elements for sparse matrices is $O(n \log n)$ compared to $O(n)$ for LDPC matrices, we need a distinction. Therefore, it can be seen that under the same condition, while implementing algorithms such as the sum-product algorithm takes $O(n)$ execution time for code using LDPC matrices, the polynomial order of n will be necessary for implementing code using sparse matrices.

In the lossless universal source coding problem, we showed the universality of sparse matrices that construct the encoder and decoder, and showed that by using the decoder that does not depend on the statistical properties of the source, the decoding error asymptotically approaches 0. The fact that the obtained error exponent is similar to that of the ordinary linear code is remarkable.

The lossy source code constructed using sparse matrices is shown to have asymptotic optimality for arbitrary stationary memoryless sources with bounded and additive distortion measures. Simulation experiments are carried out by implementing the sparse matrix code using the linear programming method proposed by Feldman [12], and show that the code attains high compression performance that goes beyond the time-sharing bound.

The channel code constructed using sparse matrices is shown to have asymptotic optimality for arbitrary stationary memoryless channels. Note that the code constructed here is simpler than the code proposed by Bennatan and Burshtein [2] who used a "quantization map" over a sufficiently large virtual alphabet. While they assumed the decoder was ML decoder, we can show the universality of the code using minimum entropy decoding as the decoding operation [33]. The duality of encoder and decoder between the lossy source code and the channel code seems interesting.

For joint source-channel coding systems, while the code approaching LMTR is ordinarily constructed by serially combining the optimal lossy source code and the optimal channel code, we show that by taking output from the vector quantizer of the lossy source code as the channel codeword, code construction becomes much simpler.

We showed that fundamental problems of point-to-point communication systems can be analyzed using simple common concepts and techniques. Point-to-point systems are so fundamental in Information Theory that coding theorems of many multi-terminal systems are derived by combining results or techniques of point-to-point systems. We can expect in sparse matrix coding,

multi-terminal systems can be investigated similarly, and so far, Muramatsu and Miyake [35] have constructed sparse matrix codes for the Wyner-Ziv and the Gel'fand-Pinsker systems and shown their asymptotic optimality. In the near future, we are looking forward to seeing implementations of efficient and optimal video coding or radio communication using the results of the multi-terminal source or channel coding problems.

1.3 Composition

In Chapter 2, we show the fundamental lemmas, which we will often use in the following chapters. In Chapters 3, 4, and 5, a lossless universal source code, lossy source code, and channel code are constructed using sparse matrices and their asymptotic optimality are proven. We conclude in Chapter 6 by describing future works.

1.4 Notations

A list of the commonly used notation is as follows. The first entry is the symbol, followed by its meaning.

$\mathbf{R}$	set of real numbers
$\mathbf{R}^+$	set of positive real numbers
$[1\mathbf{:}\ n\ ]$	a set of integers $\{1, 2, \cdots, n\}$
$[a\mathbf{,}\ b\ ]$	a closed interval including both endpoints $a,\ b \in \mathbf{R}$
$(a\mathbf{,}\ b)$	an open interval which does not include both endpoints $a,\ b \in \mathbf{R}$
$GF(q)$	a finite field constituted by $[0 : q-1]$, where q is a prime number
$\lvert\mathbf{set}\rvert$	a cardinality of the set; otherwise, $\lvert$number$\rvert$ denotes the absolute value of the number
x^n	n-dimensional row vector $(x_1, x_2, ..., x_n)$
$\boldsymbol{A}\mathbf{,}\ \boldsymbol{B}$	sparse matrices

1[(logical) equation] indicator function: if (logical) equation is satisfied, the value is 1; otherwise, 0

$\mathbf{E}_X$ expectation operation with respect to the random variable X

$d_H(a,b)$ Hamming distance between $a, b \in GF(q)$: if $a = b$, the value is 0; otherwise, 1

$w(z^n)$ Hamming weight of the sequence z^n: number of non-zero components of z^n

R coding rate defined by $R = k/n$ for source coding shown in Figure 1.1 and 1.2, and for channel coding shown in Figure 1.3.

z^* operation of taking value 1 if $z \neq 0$; otherwise, 0

$|t|^+$ $\max\{0, t\}$

$S(w)$ weight spectrum defined by (2.5) in Lemma 2.2

$\{\alpha_n(R,w)\}$ a sequence which satisfies (2.8)

$\{\beta_n(R)\}$ a sequence which satisfies (2.6)

e Napier's constant

ln logarithmic function with base e

log logarithmic function with base q specified in the context

$h(p)$ binary entropy function defined by $p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$ for $p \in [0,1]$

$H(U)$ entropy function of random variable U defined by $\sum_a P_U(a) \log \frac{1}{P_U(a)}$
another notation $H(P_U)$ is used depending on the context

$H(V|U)$ conditional entropy function of random variable V conditioned on U, which is defined by $\sum_{a,b} P_{UV}(a,b) \log \frac{1}{P_{V|U}(b|a)}$
another notation $H(P_{V|U}|P_U)$ is used depending on the context

$I(U;V)$ mutual information between random variables U and V, which is defined by $\sum_{a,b} P_{UV}(a,b) \log \frac{P_{V|U}(b|a)}{P_V(b)}$
another notation $I(P_U, P_{V|U})$ is used depending on the context

$D(P||Q)$ Kullback Libler divergence between probability distributions P and Q defined by $\sum_a P(a) \log \frac{P(a)}{Q(a)}$

$||P - Q||$ variational distance between probability distributions P and Q defined by $\sum_a |P(a) - Q(a)|$

T^n_Q type set (set of sequences with type Q)

$T^n_W(x^n)$ conditional type set (set of sequences with conditional type W for a given x^n)

$T^n_{Q\varepsilon}$ (jointly) typical sequence set defined by (2.49) or (2.51)

$T^n_{W\varepsilon}(x^n)$ conditionally typical sequence set defined by (2.52)

Chapter 2

Preliminaries for Sparse Matrix Coding

In this chapter, concepts and lemmas which are used in the following chapters are shown.

In each of Sections 2.1 and 2.2, first, we describe the construction of a sparse matrix and correspondence between the syndrome constraint $x^n \boldsymbol{A} = c^k$ and a random walk. The lemma of the weight spectrum is fundamental for applying the random walk formula to the proof of the sparse matrix coding theorem. A second moment lemma is necessary for evaluating Chebyshev's inequality by using the weight spectrum lemma. In Section 2.3, concepts and some lemmas of the type method [7], which is often used throughout the thesis, are shown.

In Section 2.1, issues are described in the case of binary alphabet for readability, and in Section 2.2, general non-binary alphabet cases are investigated.

2.1 Sparse Matrix with Binary Alphabet

In this section, alphabet is taken as $GF(2)$, where $GF(2)$ is a finite field constructed by $\{0, 1\}$. First, we show the construction of a sparse matrix that will be used in coding and decoding. After showing the correspondence between the sparse matrix coding operation and random walk, some useful lemmas of random walk follow, which will be used in proofs of coding theorems.

2.1.1 Construction of Sparse Matrix

$n \times k$ sparse matrix $\boldsymbol{A}$ is constructed as follows. Let a sparse matrix parameter be t that is an appropriate even natural number. Then, the construction is

Step 1: Set all elements of $\boldsymbol{A}$ to be 0.
In each row, the following operation (Step 2) is carried out.

Step 2: Take number $a \in [1:k]$ uniformly at random. Add number 1 to the a-th column, where the addition is modulo 2. Repeat this step t times.

Note that throughout the thesis, for a positive integer k, we define $[1:k] \stackrel{\text{def}}{=} \{1, 2, \cdots, k\}$.

From the above construction, $\boldsymbol{A}$ can be regarded as a random variable. We use notations $P_{\boldsymbol{A}}[\mathcal{E}_{\boldsymbol{A}}]$ and $\mathrm{E}_{\boldsymbol{A}} f(\boldsymbol{A})$ as a probability for an event $\mathcal{E}_{\boldsymbol{A}}$ and an expectation of function $f(\boldsymbol{A})$, respectively.

Remark 2.1
Sparse matrix parameter t is specified for the formulas stated in the following subsections to hold.

2.1.2 Random Walk Correspondence

In sparse matrix coding, a probability $P_{\boldsymbol{A}}\left[x^n \boldsymbol{A} = c^k\right]$ is often evaluated for given $x^n \in GF(2)^n$ and $c^k \in GF(2)^k$. When we define the weight function $w : GF(2)^n \to [0:n]$ as

$$w(x^n) \stackrel{\text{def}}{=} \sum_{i=1}^{n} d_H(x_i, 0), \tag{2.1}$$

where d_H is Hamming distance:

$$d_H(a, b) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } a = b \\ 1, & \text{otherwise.} \end{cases} \tag{2.2}$$

Then $x^n \boldsymbol{A} = c^k$ can be interpreted as random walk on hypercube $\{0,1\}^k$ starting from the origin 0^k and attaining the point c^k in $w(x^n) \times t$ steps. Therefore, lemmas for random walk are available for analyzing properties of sparse matrix coding. Figure 2.1 shows an example of random walk correspondence. At each step, a particle on a vertex of hyper cube $\{0,1\}^k$ moves

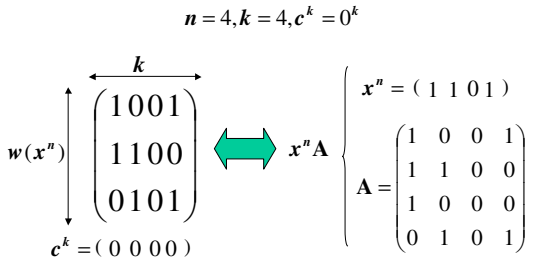

Figure 2.1: Example of random walk correspondence

to the vertex connected by an edge. $w(x^n) \times t$ corresponds to the number of steps, and c^k corresponds to the point where the particle starting from the origin stops after $w(x^n) \times t$ steps.

Note that when the weight of c^k is even, if $w(x^n) \times t$ is odd, then the probability $P_{\boldsymbol{A}}\left[x^n \boldsymbol{A} = c^k\right]$ is apparently 0. We assume that c^k is even weight and the sparse matrix parameter t is an even number when we consider a binary alphabet.

2.1.3 Random Walk Lemmas

The next lemma for random walk is well known.

Lemma 2.1 (Random Walk Formula [22])
A random walk on a k-dimensional hyper cube $\{0,1\}^k$ is considered. A particle moves from one vertex to another, which are connected by an edge, with uniform probability $1/k$.
The probability that the particle starting at origin 0^k stays at the vertex, whose Hamming weight is w after m steps, is

$$\frac{1}{2^k}\sum_{j=0}^{k} a_{wj}\left(1-\frac{2j}{k}\right)^m, \tag{2.3}$$

where

$$a_{wj} \stackrel{\text{def}}{=} \sum_{\nu=0}^{j}(-1)^{\nu}\binom{w}{\nu}\binom{k-w}{j-\nu} \tag{2.4}$$

■

Using the above lemma and considering that $w = 0$, the next lemma holds.

Lemma 2.2 ([28][31][35])
Let the weight spectrum $S(w)$ be

$$S(w) \stackrel{\text{def}}{=} \sum_{z^n : w(z^n)=w} \mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[z^n \boldsymbol{A} = 0^k\right], \tag{2.5}$$

where $\mathbf{1}[\cdot]$ denotes an indicator function. Then for a fixed $R = k/n$, there exists a positive number $\gamma_R \in (0, 1]$ satisfying the following statements:
1) *There exists a sequence $\{\beta_n(R)\}_{n=1}^{\infty}$ which satisfies*

$$\sum_{w=1}^{n\gamma_R} S(w) < \beta_n(R), \tag{2.6}$$

where $\beta_n(R) \to 0$ $(n \to \infty)$.
2) *For any integer $w \in [n\gamma_R + 1 : n]$, if the sparse matrix parameter t satisfies the conditions that t is an even number and that*

$$t \geq \max\left\{\frac{10}{3}, \frac{\ln n}{\gamma_R}\right\} \tag{2.7}$$

then there exists a sequence $\{\alpha_n(R; w)\}_{n=1}^{\infty}$ which satisfies

$$P_{\boldsymbol{A}}\left[z^n \boldsymbol{A} = 0^k \;\middle|\; w(z^n) = w\right] = \frac{\alpha_n(R; w)}{2^{k-1}}, \tag{2.8}$$

and there exists a positive number sequence $\{\delta_n(R)\}_{n=1}^{\infty}$ satisfying

$$|\alpha_n(R; w) - 1| \leq \delta_n(R) \text{ for any } w \in [n\gamma_R + 1 : n], \tag{2.9}$$

and

$$\delta_n(R) \to 0 \ (n \to \infty). \tag{2.10}$$

■

Remark 2.2
Note that Lemma 2.2 represents an original form of the hash property proposed in [35].

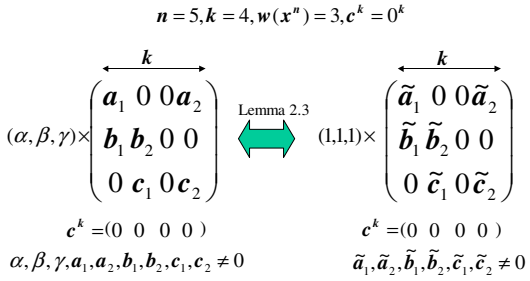

Figure 2.2: Random walk correspondence with non-binary alphabet

2.2 Sparse Matrix with Non-Binary Alphabet

In this section, an alphabet is taken as $GF(q)$ for a prime number q, where $GF(q)$ is a finite field constructed by $[0 : q-1]$, and discuss the extension to a non-binary alphabet. Throughout this thesis, the base of log is q.

2.2.1 Construction of Sparse Matrix

$n\times k$ sparse matrix $\boldsymbol{A}$ is constructed as follows. Similar to the binary alphabet case, let t be a sparse matrix parameter that is an appropriate even natural number.

Step 1: Set all elements of $\boldsymbol{A}$ to be 0.
In each row, the following operation (Step 2) is carried out.

Step 2: Take numbers $a \in [1 : k]$ and $b \in [1 : q-1]$ uniformly at random. Add the number b to the a-th column, where the addition is modulo q. Repeat this step t times.

2.2.2 Random Walk Correspondence

In a non-binary alphabet, the correspondence between $x^n\boldsymbol{A} = c^k$ and random walk is not obvious. The left figure in Figure 2.2 shows an example. Since a non-zero number, which can be different from 1, is multiplied with each raw, to establish the correspondence, some precautions are needed.

The next lemma shows the equality between the probability of the left figure and that of the right figure in Figure 2.2. Note that the correspondence between the right figure in Figure 2.2 and random walk on k-dimensional lattice is obvious.

Lemma 2.3 (Random Walk Correspondence in Non-Binary Alphabet [11])

Let x^n be some q-ary n-tuple, and let x^{*n} be defined as follows:

$$x_i^* \stackrel{\text{def}}{=} \begin{cases} 0, & x_i = 0 \\ 1, & x_i \neq 0. \end{cases} \tag{2.11}$$

Then we have

$$P_{\boldsymbol{A}}\left[x^n \boldsymbol{A} = c^k\right] = P_{\boldsymbol{A}}\left[x^{*n} \boldsymbol{A} = c^k\right]. \tag{2.12}$$

■

2.2.3 Random Walk Lemmas

The next lemma corresponds to Lemma 2.1.

Lemma 2.4 (Random Walk Formula [31])
A random walk on a k-dimensional hyper lattice $[0, q-1]^k$ is considered. A particle moves from one vertex to another, which is different in only one coordinate, and the coordinate and its value are selected uniformly at random with probability $\frac{1}{k(q-1)}$.
The probability that the particle starting at origin 0^k stays at the vertex, whose Hamming weight is w, after m steps is

$$\frac{1}{q^k} \sum_{j=0}^{k} \tilde{a}_{wj} \left(1 - \frac{qj}{(q-1)k}\right)^m, \tag{2.13}$$

where

$$\tilde{a}_{wj} \stackrel{\text{def}}{=} \sum_{\nu=0}^{j} (-1)^\nu \binom{w}{\nu} \binom{k-w}{j-\nu} (q-1)^{j-\nu} \tag{2.14}$$

■

Note that each term in the summation of $\tilde{a}_{wj}$ in (2.14) can be a negative number.

[Proof of Lemma 2.4]

Let $P[\boldsymbol{h}]$ be the probability that a particle moves along with the vector $\boldsymbol{h} \in [0:q-1]^k$ in the unit step. Note that from the definition of random walk on $[0:q-1]^k$, the weight of the vector $\boldsymbol{h}$ is 1, and $P[\boldsymbol{h}] = \frac{1}{k(q-1)}$. Let

$$\hat{\mu}(\boldsymbol{\xi}) \stackrel{\text{def}}{=} \sum_{\boldsymbol{h}\in[0:q-1]^k : w(\boldsymbol{h})=1} P[\boldsymbol{h}] \mathrm{e}^{\frac{2\pi i \boldsymbol{\xi}\cdot\boldsymbol{h}}{q}}, \tag{2.15}$$

where $w(\boldsymbol{h})$ is a Hamming weight of $\boldsymbol{h}$.

Then, noting that the random walk is an additive process and by using the property of Fourier transformation, we have

$$P_{\boldsymbol{g}} = \frac{1}{q^k} \sum_{\boldsymbol{\xi}\in[0:q-1]^k} \hat{\mu}(\boldsymbol{\xi})^m \mathrm{e}^{-\frac{2\pi i \boldsymbol{\xi}\cdot\boldsymbol{g}}{q}}, \tag{2.16}$$

where $P_{\boldsymbol{g}}$ is the probability that a particle that started at origin 0^k stays at $\boldsymbol{g} \in [0:q-1]^k$ whose Hamming weight is w.

The Fourier coefficient $\hat{\mu}(\boldsymbol{\xi})$ can be computed as follows.

$$\hat{\mu}(\boldsymbol{\xi}) \stackrel{\text{(a)}}{=} \frac{\omega^{\xi_1} + .. + \omega^{(q-1)\xi_1} + .. + \omega^{\xi_k} + .. + \omega^{(q-1)\xi_k}}{(q-1)k}, \tag{2.17}$$

where at (a), ω is a primitive q-th root of 1.

Note that for an integer $\xi \neq 0$, since it holds that

$$\omega^{\xi} + \omega^{2\xi} + \cdots + \omega^{(q-1)\xi} = -1, \tag{2.18}$$

we obtain

$$\begin{aligned} \hat{\mu}(\boldsymbol{\xi}) &= \frac{(k - w(\boldsymbol{\xi}))(q-1) - w(\boldsymbol{\xi})}{(q-1)k} \\ &= 1 - \frac{q w(\boldsymbol{\xi})}{(q-1)k}. \end{aligned} \tag{2.19}$$

By substituting the above equation into (2.16), it holds that

$$\begin{aligned} P_{\boldsymbol{g}} &= \frac{1}{q^k} \sum_{\boldsymbol{\xi}\in[0:q-1]^k} \left(1 - \frac{q w(\boldsymbol{\xi})}{(q-1)k}\right)^m \omega^{-\boldsymbol{\xi}\cdot\boldsymbol{g}} \\ &= \frac{1}{q^k} \sum_{j=0}^{k} \left(1 - \frac{qj}{(q-1)k}\right)^n \sum_{\boldsymbol{\xi}: w(\boldsymbol{\xi})=j} \omega^{-\boldsymbol{\xi}\cdot\boldsymbol{g}}. \end{aligned} \tag{2.20}$$

When we set $w(\boldsymbol{g}) = w$, it can be shown that

$$\begin{aligned}\sum_{\boldsymbol{\xi}:w(\boldsymbol{\xi})=j} \omega^{-\boldsymbol{\xi}\cdot\boldsymbol{g}} &= \sum_{\nu=0}^{j} \binom{w}{\nu}\binom{l-w}{j-\nu}(-1)^{\nu}(q-1)^{j-\nu} \\ &= \tilde{a}_{wj}. \end{aligned} \tag{2.21}$$

By substituting (2.21) into (2.20), we obtain the desired formula.

[End of Proof of Lemma 2.4]

The next lemma corresponds to Lemma 2.2.

Lemma 2.5 ([31][35])
For the weight spectrum $S(w)$ defined in the same way as (2.5) *of Lemma* 2.2 *except $z^n \in [0 : q-1]^n$, then for a fixed $R = k/n$, there exists a positive number $\gamma_R \in (0, 1]$ satisfying the following statements:*
1) *There exists a sequence $\{\beta_n(R)\}_{n=1}^{\infty}$ that satisfies*

$$\sum_{w=1}^{n\gamma_R} S(w) < \beta_n(R), \tag{2.22}$$

where $\beta_n(R) \to 0$ $(n \to \infty)$.
2) *For any integer $w \in [n\gamma_R + 1 : n]$, if the sparse matrix parameter t satisfies the conditions that t is an even number and that*

$$t \geq \max\left\{\frac{10}{3}, \frac{2}{\log(q-1)}, \frac{\ln n}{\gamma_R}\right\} \tag{2.23}$$

then there exists a sequence $\{\alpha_n(R; w)\}_{n=1}^{\infty}$ that satisfies

$$P_{\boldsymbol{A}}\left[z^n\boldsymbol{A} = 0^k \;\middle|\; w(z^n) = w\right] = \begin{cases} \frac{\alpha_n(R;w)}{2^{k-1}}, & q = 2 \\ \frac{\alpha_n(R;w)}{q^k}, & q \geq 3, \end{cases} \tag{2.24}$$

and there exists a positive number sequence $\{\delta_n(R)\}_{n=1}^{\infty}$ satisfying

$$|\alpha_n(R; w) - 1| \leq \delta_n(R) \text{ for any } w \in [n\gamma_R + 1 : n] \tag{2.25}$$

and

$$\delta_n(R) \to 0 \ (n \to \infty). \tag{2.26}$$

■

[Proof of Lemma 2.5]
From the definition, it holds that

$$\begin{aligned}
S(w) &= \sum_{z^n\in[0:q-1]^n:w(z^n)=w} \mathrm{E}_{\boldsymbol{A}}\mathbf{1}\left[z^n\boldsymbol{A}=0^k\right] \\
&\overset{\text{(a)}}{=} \sum_{z^n\in[0:q-1]^n:w(z^n)=w} \mathrm{E}_{\boldsymbol{A}}\mathbf{1}\left[z^{*n}\boldsymbol{A}=0^k\right] \\
&= (q-1)^w \sum_{z^{*n}\in\{0,1\}^n:w(z^{*n})=w} \mathrm{E}_{\boldsymbol{A}}\mathbf{1}\left[z^{*n}\boldsymbol{A}=0^k\right] \\
&\overset{\text{(b)}}{=} (q-1)^w\binom{n}{w}\frac{1}{q^k}\sum_{j=0}^{k}\binom{k}{j}(q-1)^j\left(1-\frac{qj}{(q-1)k}\right)^{wt}, \qquad (2.27)
\end{aligned}$$

where (a) is from Lemma 2.3 and $z_i^* \overset{\text{def}}{=} \begin{cases} 0, & z_i=0 \\ 1, & z_i\neq 0. \end{cases}$, (b) is from Lemma 2.4 by setting $w=0$ of $\tilde{a}_{wj}$ in (2.13).
1) Using (2.27), it is sufficient to show

$$\lim_{n\to\infty}\sum_{w=1}^{n\gamma}\binom{n}{w}\frac{(q-1)^w}{q^k}\sum_{j=0}^{k}\binom{k}{j}(q-1)^j\left(1-\frac{qj}{(q-1)k}\right)^{wt}=0 \qquad (2.28)$$

for some $\gamma>0$, which we specify as γ_R later.

The left hand side of the above equation can be transformed as follows.

$$\begin{aligned}
&\sum_{w=1}^{n\gamma}\binom{n}{w}\frac{(q-1)^w}{q^k}\sum_{j=0}^{k}\binom{k}{j}(q-1)^j\left(1-\frac{qj}{(q-1)k}\right)^{wt} \\
&=\sum_{w=1}^{n\gamma}\binom{n}{w}\frac{(q-1)^w}{q^k} \\
&\quad\cdot\left(\sum_{j=0}^{\lfloor\frac{(q-1)k}{q}\rfloor}+\sum_{j=\lfloor\frac{(q-1)k}{q}\rfloor+1}^{k}\right)\binom{k}{j}(q-1)^j\left(1-\frac{qj}{(q-1)k}\right)^{wt} \qquad (2.29) \\
&\le\sum_{w=1}^{n\gamma}\binom{n}{w}\frac{(q-1)^w}{q^k}\sum_{j=0}^{\lfloor\frac{(q-1)k}{q}\rfloor}\binom{k}{j}(q-1)^j\mathrm{e}^{-\frac{qwtj}{(q-1)k}} \\
&\quad+\sum_{w=1}^{n\gamma}\binom{n}{w}\frac{(q-1)^w}{q^k}\sum_{j=\lfloor\frac{(q-1)k}{q}\rfloor+1}^{k}\binom{k}{j}(q-1)^j\left(\frac{1}{q-1}\right)^{wt} \qquad (2.30)
\end{aligned}$$

Note that when $q = 2$, the second term of (2.29) is equal to the first term. The first term of (2.30) is derived using $1 - x \leq \mathrm{e}^{-x}$, and the second term is followed by noting that the sparse matrix parameter t is assumed to be an even number and the fact that the k-th term of the summation attains the maximum value.

The first term of (2.30) is upper bounded as follows.

$$\begin{aligned}
&\sum_{w=1}^{n\gamma} \binom{n}{w} \frac{(q-1)^w}{q^k} \sum_{j=0}^{\lfloor \frac{(q-1)k}{q} \rfloor} \binom{k}{j} (q-1)^j \mathrm{e}^{-\frac{qwtj}{(q-1)k}} \\
&\leq \sum_{w=1}^{n\gamma} \binom{n}{w} \frac{(q-1)^w}{q^k} \sum_{j=0}^{k} \binom{k}{j} (q-1)^j \mathrm{e}^{-\frac{qwtj}{(q-1)k}} \\
&= \left(\sum_{w=1}^{\frac{bn}{\ln n}} + \sum_{w=\frac{bn}{\ln n}+1}^{n\gamma} \right) \binom{n}{w} (q-1)^w \left(\frac{1 + (q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}}{q} \right)^k, \qquad (2.31)
\end{aligned}$$

where at the last equality, the binomial theorem is used, and $b > 0$ is a positive constant specified later.

By substituting $k = nR, t = \xi \ln n$, the second term of (2.31) is evaluated as follows.

$$\begin{aligned}
&\sum_{w=\frac{bn}{\ln n}+1}^{n\gamma} \binom{n}{w} (q-1)^w \left(\frac{1 + (q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}}{q} \right)^k \\
&\leq n\gamma q^{n(h(\gamma)+\gamma)} \left(\frac{1 + (q-1)\mathrm{e}^{-\frac{q\xi \ln n \frac{bn}{\ln n}}{(q-1)nR}}}{q} \right)^{nR} \\
&= n\gamma q^{n(h(\gamma)+\gamma)} \left(\frac{1 + (q-1)\mathrm{e}^{-\frac{qb\xi}{(q-1)R}}}{q} \right)^{nR} \\
&= n\gamma \left(\frac{q^{\frac{h(\gamma)+\gamma}{R}} \left(1 + (q-1)\mathrm{e}^{-\frac{qb\xi}{(q-1)R}}\right)}{q} \right)^{nR} \qquad (2.32)
\end{aligned}$$

At the above inequality, $h(p)$ is a binary entropy function defined by $p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$, and using the technique of the type method, the upper bound

is derived (see (2.62) of Lemma 2.7 shown later). From the above evaluation, if $\frac{b\xi}{R} \geq 1$ holds, then there exist γ_R satisfying

$$q^{\frac{h(\gamma_R)+\gamma_R}{R}} \left(1 + (q-1)\mathrm{e}^{-\frac{qb\xi}{(q-1)R}}\right) < q, \tag{2.33}$$

and this fact shows that the second term of (2.31) vanishes exponentially when $n \to \infty$.

As the next step, we evaluate the first term of (2.31). Each term of the summation can be evaluated as

$$\begin{aligned}
&\binom{n}{w}(q-1)^w \left(\frac{1+(q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}}{q}\right)^k \\
&\leq (nq)^w \left(\frac{1+(q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}}{q}\right)^k \\
&\overset{\text{(a)}}{\leq} (nq)^w \left(\frac{1+(q-1)\left(1-\frac{0.3qwt}{(q-1)k}\right)}{q}\right)^k \\
&= (nq)^w \left(1-\frac{0.3wt}{k}\right)^k \\
&\leq (nq)^w \mathrm{e}^{-0.3wt} \\
&= \left(qn^{1-0.3\xi}\right)^w,
\end{aligned} \tag{2.34}$$

where at (a), $\mathrm{e}^{-x} \leq 1 - 0.3x$ $(0 \leq x \leq 3)$ is used under the condition $\frac{b\xi}{R} \leq \frac{3}{2}$. (Note that if $\frac{b\xi}{R} \leq \frac{3}{2}$, $\frac{qwt}{(q-1)k}$ is upper bounded by 3.) By substituting the above evaluation, we have

$$\begin{aligned}
&\sum_{w=1}^{\frac{bn}{\ln n}} \binom{n}{w}(q-1)^w \left(\frac{1+(q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}}{q}\right)^k \\
&\leq \frac{qn^{1-0.3\xi}}{1-qn^{1-0.3\xi}} \to 0 \ (n \to \infty, \text{ if } \xi > \tfrac{10}{3}).
\end{aligned} \tag{2.35}$$

As the final step, the second term of (2.30) is evaluated as follows. Noticing the statement following (2.30), we assume $q \geq 3$.

$$
\begin{aligned}
&\sum_{w=1}^{n\gamma_R} \binom{n}{w} \frac{(q-1)^w}{q^k} \sum_{j=\lfloor \frac{(q-1)k}{q} \rfloor + 1}^{k} \binom{k}{j} (q-1)^j \left(\frac{1}{q-1}\right)^{wt} \\
&= \sum_{w=1}^{n\gamma_R} \binom{n}{w} (q-1)^{w(1-t)} \\
&\quad \sum_{j=\lfloor \frac{(q-1)k}{q} \rfloor + 1}^{k} \binom{k}{j} \left(\frac{1}{q}\right)^{k-j} \left(1-\frac{1}{q}\right)^{j} \\
&\leq \sum_{w=1}^{n\gamma_R} \binom{n}{w} (q-1)^{w(1-t)} \\
&\leq \sum_{w=1}^{n\gamma_R} q^{nh\left(\frac{w}{n}\right) - w(t-1)\log(q-1)} \\
&= n\gamma_R q^{-n \min_{\frac{1}{n} \leq x \leq \gamma} \{x(t-1)\log(q-1) - h(x)\}} \qquad (2.36)
\end{aligned}
$$

Note that by differentiating $x(t-1)\log(q-1) - h(x)$ by x, if $q \leq n^{\xi \log(q-1)-1}$, then the minimum value of $\min_{\frac{1}{n} \leq x \leq \gamma} \{x(t-1)\log(q-1) - h(x)\}$ is attained at $x = \frac{1}{n}$. If $\xi > \frac{2}{\log(q-1)}$, then $q \leq n$ is sufficient for the above minimum value being attained at $x = \frac{1}{n}$. Then we have

$$
\begin{aligned}
&\sum_{w=1}^{n\gamma_R} \binom{n}{w} \frac{(q-1)^w}{q^k} \sum_{j=\lfloor \frac{(q-1)k}{q} \rfloor + 1}^{k} \binom{k}{j} (q-1)^j \left(\frac{1}{q-1}\right)^{wt} \\
&\leq n\gamma_R q^{-n\left\{\frac{1}{n}(t-1)\log(q-1) - h\left(\frac{1}{n}\right)\right\}} \\
&= n\gamma_R q^{-(t-1)\log(q-1) + n\left(-\frac{1}{n}\log\frac{1}{n} - \left(1-\frac{1}{n}\right)\log\left(1-\frac{1}{n}\right)\right)} \\
&= n\gamma_R q^{-(t-1)\log(q-1) + \log n + \log\left(1+\frac{1}{n-1}\right)^{n-1}} \\
&\overset{(a)}{\leq} n\gamma_R q^{-(t-1)\log(q-1) + \log n + \frac{1}{\ln q}} \\
&= n\gamma_R q^{\log(q-1) + \frac{1}{\ln q} - (\xi \ln(q-1) - 1)\log n} \\
&= \gamma_R q^{\log(q-1) + \frac{1}{\ln q}} n^{-(\xi \ln(q-1) - 2)} \to 0 \ (n \to \infty, \text{ if } \xi > \tfrac{2}{\log(q-1)}). \qquad (2.37)
\end{aligned}
$$

At (a), $(1 + 1/n)^n \leq e$ is used.

From (2.32), (2.35), and (2.37), by setting

$$\beta_n(R) = n\gamma \left(\frac{q^{\frac{h(\gamma)+\gamma}{R}} \left(1 + (q-1)\mathrm{e}^{-\frac{qb\xi}{(q-1)R}}\right)}{q} \right)^{nR} + \frac{qn^{1-0.3\xi}}{1 - qn^{1-0.3\xi}} + \gamma_R q^{\log(q-1)+\frac{1}{\ln q}} n^{-(\xi \ln(q-1)-2)}, \tag{2.38}$$

the desired inequality is obtained.

2) It is sufficient to show that

$$\sum_{j=0}^{k} \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \to 1 \ (n \to \infty) \tag{2.39}$$

First, it holds that

$$\begin{aligned} &\sum_{j=0}^{k} \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \\ &= \left(\sum_{j=0}^{\lfloor \frac{(q-1)k}{q} \rfloor} + \sum_{j=\lfloor \frac{(q-1)k}{q} \rfloor + 1}^{k} \right) \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt}. \end{aligned} \tag{2.40}$$

Note that when $q = 2$, the first term of (2.40) is equal to the second term of (2.40).

The proof is completed after showing

$$\sum_{j=0}^{\lfloor \frac{(q-1)k}{q} \rfloor} \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \to 1 \ (n \to \infty) \tag{2.41}$$

and

$$\sum_{j=\lfloor \frac{(q-1)k}{q} \rfloor + 1}^{k} \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \to 0 \ (n \to \infty). \tag{2.42}$$

Noticing that the term of $j = 0$ in (2.41) is equal to 1, we have

$$\begin{aligned}
1 &\leq \sum_{j=0}^{\lfloor \frac{(q-1)k}{q} \rfloor} \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \\
&\leq \sum_{j=0}^{\lfloor \frac{(q-1)k}{q} \rfloor} \binom{k}{j} (q-1)^j \mathrm{e}^{-\frac{qwt}{(q-1)k} j} \\
&\leq \sum_{j=0}^{k} \binom{k}{j} \left((q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}\right)^j \\
&= \left(1 + (q-1)\mathrm{e}^{-\frac{qwt}{(q-1)k}}\right)^k \\
&\overset{\text{(a)}}{\leq} \mathrm{e}^{(q-1)k\mathrm{e}^{-\frac{q}{q-1}\ln k}} \\
&= \mathrm{e}^{(q-1)k^{\frac{-1}{q-1}}} \to 1 \ (n \to \infty),
\end{aligned} \tag{2.43}$$

where at (a), $1+x \leq \mathrm{e}^x$ and $wt \geq k \ln k$ is used. Note that since $w \in [n\gamma_R+1 : n]$ and $t \geq \frac{\ln n}{\gamma_R}$ from (2.23), we have $wt \geq n\gamma_R \cdot \frac{\ln n}{\gamma_R} = n \ln n \geq k \ln k$.

On the other hand, using the assumption that t is an even number, it holds that

$$\begin{aligned}
&\sum_{j=\lfloor \frac{(q-1)k}{q} \rfloor + 1}^{k} \binom{k}{j} (q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \\
&\overset{\text{(b)}}{\leq} \sum_{j=\lfloor \frac{2k}{3} \rfloor + 1}^{k} \binom{k}{j} (q-1)^j (q-1)^{-wt} \\
&\overset{\text{(c)}}{\leq} (q-1)^{-k\ln k} q^{k(1+h(1/3))} \to 0 \ (n \to \infty),
\end{aligned} \tag{2.44}$$

where at (b), $q \geq 3$ is used, and at (c), $wt \geq k \ln k$ is used. From (2.43) and (2.44), by setting

$$\delta_n(R) = \mathrm{e}^{(q-1)k^{\frac{-1}{q-1}}} - 1 + (q-1)^{-k\ln k} q^{k(1+h(1/3))}, \tag{2.45}$$

the proof is completed.

[End of Proof of Lemma 2.5]

Remark 2.3
Throughout the thesis, when we use Lemma 2.5*, we assume that the condition for* t (2.23) *is satisfied. When* n *is large, the condition* $t = O(\log n)$ *is essential. This condition corresponds to the fact that after* $O(n \log n)$ *steps, the probability that the random walking particle stays at a vertex converges to a uniform distribution.*

Remark 2.4
Let k*-dimensional row vector* $c^k \in [0 : q-1]^k$ *be given, and* $w(c^k) = c$*, then from Lemma* 2.4*, we have*

$$\begin{aligned}
&P_{\boldsymbol{A}}\left[z^n\boldsymbol{A} = c^k \;\middle|\; w(z^n) = w\right] \\
&\le \frac{1}{q^k}\sum_{j=0}^{k} |\tilde{a}_{cj}| \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \\
&\le \frac{1}{q^k}\sum_{j=0}^{k} \binom{k}{j}(q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \\
&= P_{\boldsymbol{A}}\left[z^n\boldsymbol{A} = 0^k \;\middle|\; w(z^n) = w\right], \qquad (2.46)
\end{aligned}$$

where at the last inequality, the relationship $\sum_{\nu=0}^{j}\binom{w}{\nu}\binom{k-w}{j-\nu} = \binom{k}{j}$ *is used.* (2.46) *shows that we can use* 1) *of Lemma* 2.5 *for the case of* $z^n\boldsymbol{A} = c^k$ *with* $c^k \neq 0^k$*. On the other hand, from* (2.13) *it can be shown that*

$$\begin{aligned}
&\left|\sum_{j=0}^{k} \tilde{a}_{wj}\left(1 - \frac{qj}{(q-1)k}\right)^{wt} - 1\right| = \left|\sum_{j=1}^{k} \tilde{a}_{wj}\left(1 - \frac{qj}{(q-1)k}\right)^{wt}\right| \\
&\le \sum_{j=1}^{k} |\tilde{a}_{wj}| \left(1 - \frac{qj}{(q-1)k}\right)^{wt} \\
&\le \frac{1}{q^k}\sum_{j=1}^{k} \binom{k}{j}(q-1)^j \left(1 - \frac{qj}{(q-1)k}\right)^{wt}. \qquad (2.47)
\end{aligned}$$

Therefore, by evaluating (2.47) *using the same arguments for the evaluation of* (2.43) *and* (2.44)*, we can use* 2) *of Lemma* 2.5 *for the case of* $z^n\boldsymbol{A} = c^k$ *with* $c^k \neq 0^k$*.*

2.3 Types and Second Moment Lemma

So far, we considered lemmas whose statements involve formulas like $\mathrm{E}_{\boldsymbol{A}}\mathbf{1}[f(\boldsymbol{A})]$. Some analyses, such as Chebyshev's inequality, need formulas like $\mathrm{E}_{\boldsymbol{A}}\mathbf{1}[f(\boldsymbol{A})]\mathbf{1}[g(\boldsymbol{A})]$. This section shows lemmas available in that situation.

Before stating the lemmas, we show the definitions of the types and typical sequence sets that will be used throughout this thesis.

Definition 2.1 *[Type, Type set, Typical sequence set [7]]*
For a sequence $x^n \in \mathcal{X}^n$, where $\mathcal{X}$ is a finite set, a type of x^n is defined as P_{x^n}:

$$P_{x^n}(a) \stackrel{\text{def}}{=} \frac{N(a|x^n)}{n} \tag{2.48}$$

for any $a \in \mathcal{X}$, where

$$N(a|x^n) \stackrel{\text{def}}{=} \left|\{i \in [1:n] | x_i = a\}\right|,$$

and $|set|$ denotes the cardinality of the set. A type set $T^n_Q \subset \mathcal{X}^n$ is the set of sequences whose type is Q. A typical sequence set of a probability distribution Q, $T^n_{Q\varepsilon} \subset \mathcal{X}^n$, is defined as

$$T^n_{Q\varepsilon} \stackrel{\text{def}}{=} \left\{x^n \in \mathcal{X}^n \;\middle|\; |P_{x^n}(a) - Q(a)| \le \varepsilon Q(a) \text{ for any } a \in \mathcal{X}\right\}. \tag{2.49}$$

For sequences $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite, a joint type of (x^n, y^n) is defined as $P_{x^n y^n}$:

$$P_{x^n y^n}(a,b) \stackrel{\text{def}}{=} \frac{N(a,b|x^n y^n)}{n} \tag{2.50}$$

for any $(a,b) \in \mathcal{X} \times \mathcal{Y}$, where

$$N(a,b|x^n y^n) \stackrel{\text{def}}{=} \left|\{i \in [1:n] | x_i = a,\ y_i = b\}\right|.$$

A joint type set $T^n_Q \subset \mathcal{X}^n \times \mathcal{Y}^n$ is the set of sequences whose joint type is Q. A jointly typical sequence set of a joint probability distribution Q on $\mathcal{X} \times \mathcal{Y}$, $T^n_{Q\varepsilon} \subset \mathcal{X}^n \times \mathcal{Y}^n$, is defined as

$$\begin{aligned} T^n_{Q\varepsilon} \stackrel{\text{def}}{=} \Big\{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \;\Big|\; \\ |P_{x^n y^n}(a,b) - Q(a,b)| \le \varepsilon Q(a,b) \text{ for any } (a,b) \in \mathcal{X} \times \mathcal{Y}\Big\}. \end{aligned} \tag{2.51}$$

For a conditional probability distribution $W_{Y|X}$ on $y \in \mathcal{Y}$ given $x \in \mathcal{X}$, a conditional type set conditioned on x^n, $T^n_W(x^n) \subset \mathcal{Y}^n$, is the set of sequences satisfying

$$N(a,b|x^n,y^n) = N(a|x^n)W(b|a)$$

for any $(a,b) \in \mathcal{X} \times \mathcal{Y}$. A conditional typical sequence set of a conditional distribution W conditioned on x^n is defined as $T^n_{W\varepsilon}(x^n) \subset \mathcal{Y}^n$:

$$T^n_{W\varepsilon}(x^n) \stackrel{\rm def}{=} \Big\{ y^n \in \mathcal{Y}^n \Big| \, |\frac{N(a,b|x^n,y^n)}{n} - \frac{N(a|x^n)}{n}W(b|a)| \\ \le \varepsilon \frac{N(a|x^n)}{n} W(b|a) \text{ for any } (a,b) \in \mathcal{X} \times \mathcal{Y} \Big\}. \tag{2.52}$$

Lemma 2.6 *[Second Moment Lemma]*
Let a probability distribution P over $[0:q-1]$ be given. If for a positive number ε, $z^n \in T^n_{P\varepsilon}$ and $\min_{a:P(a)>0} P(a)(1-\varepsilon) > 2\gamma_R$, then the following statements hold.
1)

$$\mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[z^{*n}\boldsymbol{A} = c^k\right] \sum_{w=1}^{n\gamma_R} \sum_{\tilde{z}^n : w(\tilde{z}^{*n} - z^{*n}) = w} \mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})\boldsymbol{A} = 0^k\right] \\ \le \begin{cases} \frac{(1+\delta_n(R))\beta_n(R)}{2^{k-1}}, & q = 2 \\ \frac{(1+\delta_n(R))\beta_n(R)}{q^k}, & q \ge 3, \end{cases} \tag{2.53}$$

where $z_i^* \stackrel{\rm def}{=} \begin{cases} 0, & z_i = 0 \\ 1, & z_i \neq 0 \end{cases}$, and when $q = 2$, the Hamming weight of c^k is taken to be even.
2) *For ε satisfying $2P^*(1)\varepsilon \le \frac{\gamma_R}{2}$, and $\tilde{z}^n \in T^n_{P\varepsilon}$ and $w(\tilde{z}^{*n} - z^{*n}) > n\gamma_R$,*

$$\mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[z^{*n}\boldsymbol{A} = c^k\right] \mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})\boldsymbol{A} = 0^k\right] \\ = \begin{cases} \frac{\alpha_n(R;w(z^{*n}))}{2^{k-1}} \frac{\alpha_n(R;w(\tilde{z}^{*n}-z^{*n}))}{2^{k-1}}, & q = 2 \\ \frac{\alpha_n(R;w(z^n))}{q^k} \frac{\alpha_n(R;w(\tilde{z}^{*n}-z^{*n}))}{q^k}, & q \ge 3, \end{cases} \tag{2.54}$$

where P^* is a probability distribution over $\{0,1\}$ that satisfies $P^*(Z^* = 0) = P(Z = 0)$. ■

[Proof of Lemma 2.6*]*
Let $I \subseteq [1:n]$ be a set of indices, and $\boldsymbol{A}_I$ be a submatrix of $\boldsymbol{A}$ whose indices of rows are included in I, and row vector $(z^n)_I$ is defined in a similar way. Note that $\boldsymbol{A}_I$ is a $|I| \times k$ matrix, and $(z^n)_I$ is a $|I|$-dimensional row vector. In the following proof, $q \geq 3$ is assumed. When $q = 2$, a similar argument holds.
1) When sequences z^n and $\tilde{z}^n$ are given, define

$$I_c \stackrel{\text{def}}{=} \left\{ i \in [1:n] \;\middle|\; z_i = \tilde{z}_i \right\} \tag{2.55}$$

and

$$I_d \stackrel{\text{def}}{=} \left\{ i \in [1:n] \;\middle|\; z_i \neq \tilde{z}_i \right\}. \tag{2.56}$$

Then

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[z^{*n}\boldsymbol{A} = c^k \right] \sum_{w=1}^{n\gamma_R} \sum_{\tilde{z}^n : w(\tilde{z}^{*n} - z^{*n}) = w} \mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})\boldsymbol{A} = 0^k \right] \\
&= \sum_{w=1}^{n\gamma_R} \sum_{\tilde{z}^n : w(\tilde{z}^{*n} - z^{*n}) = w} \mathrm{E}_{\boldsymbol{A}} \left\{ \mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})_{I_d} \boldsymbol{A}_{I_d} = 0^k \right] \right. \\
&\qquad \left. \mathrm{E}_{\boldsymbol{A}_{I_c}} \mathbf{1}\left[(z^n)_{I_c} \boldsymbol{A}_{I_c} = c^k - (z^n)_{I_d} \boldsymbol{A}_{I_d} \right] \;\middle|\; (\tilde{z}^{*n} - z^{*n})_{I_d} \boldsymbol{A}_{I_d} = 0^k \right\} \\
&\stackrel{\text{(a)}}{\leq} \sum_{w=1}^{n\gamma_R} \sum_{\tilde{z}^n : w(\tilde{z}^{*n} - z^{*n}) = w} \mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})_{I_d} \boldsymbol{A}_{I_d} = 0^k \right] \frac{1 - \delta_n(R)}{q^k} \\
&\stackrel{\text{(b)}}{\leq} \beta_n(R) \frac{1 - \delta_n(R)}{q^k}
\end{aligned} \tag{2.57}$$

where at (a), since $z^n \in T^n_{P\varepsilon}$ and $\min_{a:P(a)>0} P(a)(1-\varepsilon) > 2\gamma_R$,

$$w\left((z^n)_{I_C}\right) \geq n(1 - P(0))(1-\varepsilon) - w \geq 2n\gamma_R - n\gamma_R = n\gamma_R, \tag{2.58}$$

and 2) of Lemma 2.5, and the independence of rows between indices in I_c and I_d are used. At (b), 1) of Lemma 2.5 is used.
2) For a given z^n, let

$$I(z^n) \stackrel{\text{def}}{=} \left\{ i \in [1:n] \;\middle|\; z^* = 1 \right\}, \tag{2.59}$$

and note that if $z^n \in T^n_{P\varepsilon}$, then $z^{*n} \in T^n_{P^*\varepsilon}$.

Since $\tilde{z}^n \in T^n_{P_\varepsilon}$ and $w(\tilde{z}^{*n} - z^{*n}) > n\gamma_R$,

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}}\mathbf{1}\left[z^{*n}\boldsymbol{A} = c^k\right]\mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})\boldsymbol{A} = 0^k\right] \\
&= \mathrm{E}_{\boldsymbol{A}}\mathbf{1}\left[z^{*n}\boldsymbol{A} = c^k\right] \\
&\quad \cdot \mathrm{E}_{\boldsymbol{A}_{I_d\backslash I(z^n)}}\left\{\mathbf{1}\left[(\tilde{z}^{*n} - z^{*n})_{I_d\backslash I(z^n)}\boldsymbol{A}_{I_d\backslash I(z^n)} = (\tilde{z}^{*n} - z^{*n})_{I(z^n)}\boldsymbol{A}_{I(z^n)}\right]\right. \\
&\qquad\qquad \left. \Big| \; z^{*n}\boldsymbol{A} = c^k\right\} \\
&\overset{\text{(c)}}{=} \mathrm{E}_{\boldsymbol{A}}\mathbf{1}\left[z^{*n}\boldsymbol{A} = c^k\right]\frac{\alpha_n(R; w(\tilde{z}^{*n} - z^{*n}))}{q^k} \\
&= \frac{\alpha_n(R; w(z^{*n}))}{q^k}\frac{\alpha_n(R; w(\tilde{z}^{*n} - z^{*n}))}{q^k},
\end{aligned} \tag{2.60}$$

where at (c), after noting the fact that $z^{*n}, \tilde{z}^{*n} \in T^n_{P^*\varepsilon}$ and $w(\tilde{z}^{*n} - z^{*n}) = w$, and $2P^*(1)\varepsilon \le \frac{\gamma_R}{2}$ imply

$$|I_d\backslash I(z^{*n})| \ge n\frac{\gamma_R - 2P^*(1)\varepsilon}{2} \ge n\frac{\gamma_R}{4}, \tag{2.61}$$

we used 2) of Lemma 2.5 and the independence of rows between indices in $I_d\backslash I_{z^n}$ and I_{z^n}.

[End of Proof of Lemma 2.6*]*

Useful lemmas often used in the following chapters are stated below.

Lemma 2.7 *[[7] [17]]*
1) *For* $0 \le \gamma \le 1/2$,

$$\sum_{i=0}^{n\gamma}\binom{n}{i} \le q^{nh(\gamma)}. \tag{2.62}$$

2) *Let a set of types which is constructed by sequences over* $[0 : q-1]$ *with length* n *be* $\mathcal{P}^{(1)}_n$*, and a set of joint types which is constructed by sequences over* $[0 : q-1] \times [0 : q-1]$ *with length* n *be* $\mathcal{P}^{(2)}_n$*, then*

$$|\mathcal{P}^{(1)}_n| \le (n+1)^q, \tag{2.63}$$

and

$$|\mathcal{P}^{(2)}_n| \le (n+1)^{q^2}, \tag{2.64}$$

3) *For a type set* T^n_Q*,*

$$\frac{q^{nH(Q)}}{(n+1)^q} \le |T^n_Q| \le q^{nH(Q)}. \tag{2.65}$$

For a conditional type set conditioned on x^n, $T_W^n(x^n)$, if $T_W^n(x^n)$ is non-void,

$$\frac{q^{nH(W|P_{x^n})}}{(n+1)^{q^2}} \leq |T_W^n(x^n)| \leq q^{nH(W|P_{x^n})}. \tag{2.66}$$

■

Chapter 3

Lossless Universal Source Coding

In practical communication systems, the source statistics often are not known to both the encoder and decoder. In this situation, a universal code can attain optimal performance in the following sense. A universal code has two types of coding scheme: variable length coding and fixed length coding. In the former type, the compression rate can attain the compression limit or entropy; redundancy, which is the difference between the compression rate and entropy, has been analyzed by many researchers [39][42]. In the latter type, a universal code attains the optimal decoding error exponent, which governs the decreasing speed of the decoding error, while the compression rate remains constant. The Lempel-Ziv code [48][49] is one of the most famous variable length universal codes and has practical encoding and decoding computing times. Caire, Shamai and Verdú proposed a variable length universal code that combines linear code and the MDL principle [4][5].

On the other hand, little is currently known about fixed length universal codes, which have practical encoding and decoding computing times. While Coleman, Médard, and Effros [6] recently proposed an efficient universal decoding algorithm using linear programming or expander code, they did not discuss the efficiency of the algorithm in practical use.

In this chapter, we construct a fixed length code using sparse matrices with minimum entropy decoding [8] as a decoding scheme because it does not depend on the statistical properties of the information source, and show that the code has universal properties in the sense of fixed length coding. We also show that error exponents similar to those obtained in [8] can be

obtained using a sparse matrix random coding technique.

Using sparse matrices for code construction enables us to use existing efficient decoding algorithms such as the sum-product decoding algorithm (e.g. [14]). Fixed length universal codes constructed by sparse matrices provide a computationally practical encoding and decoding scheme whose error probability decreases optimally without the statistical properties of its sources being known.

We consider only a class of i.i.d. sources to simplify the arguments specific to sparse matrix coding. To show that a linear code can be used as a universal code, when analyzing decoding error, evaluation of a supremum with respect to an objective class of sources is inserted into the expectation operation of sparse matrix random coding. When we handle random coding of sparse matrices, the probability distribution is no longer uniform, and the evaluation problem becomes harder. The upper bound of the decoding error probability will be shown in exponential form using the expurgated ensemble technique, which Miller and Burshtein [28] and Erez and Miller [11] adopted to evaluate the exponent of the decoding error probability of a linear channel code using sparse matrices.

3.1 Preliminaries and Problem Setting

We focus on our problem in the non-binary alphabet framework, where alphabet $\mathcal{U} = GF(q)$ for a given prime number q. In the following, the base of ln is e and the base of log is q.

Let a class of probability distribution on alphabet $\mathcal{U}$ be $\mathcal{P}$. A random variable denoting output from the source $P_U \in \mathcal{P}$ is U. Assume that the stochastic process U^n is i.i.d.:

$$P_{U^n}[U^n = u^n] = \prod_{i=1}^{n} P_U[U_i = u_i]. \tag{3.1}$$

For a positive number $\tau > 0$, $\mathcal{P}(\tau)$ is a set of stationary memoryless probability distributions on $\mathcal{U}$ defined as

$$\mathcal{P}(\tau) \stackrel{\text{def}}{=} \left\{ P \in \{\text{i.i.d. sources on } \mathcal{U}\} \;\middle|\; \min_{a \in [0:q-1]} P(a) > \tau \right\}. \tag{3.2}$$

Note that from the definition, $\inf_{P \in \mathcal{P}(\tau)} \min_{a \in [0:q-1]} P(a) > \tau$, and the positive constant τ is not necessarily known to both the encoder and decoder.

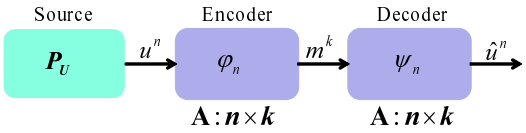

Figure 3.1: Lossless source coding system

Figure 3.1 shows the lossless source coding system considered in this chapter.

For a sequence u^n from the source, the encoder $\varphi_n : GF(q)^n \to GF(q)^k$ is defined as

$$\varphi_n(u^n) \stackrel{\text{def}}{=} u^n \boldsymbol{A}, \tag{3.3}$$

where $\boldsymbol{A}$ is an $n \times k$ sparse matrix constructed following the description in Section 2.2.1. The compression rate R is defined as k/n.

As a decoding operation, minimum entropy decoding [8] is adopted. For a given codeword $m^k \in GF(q)^k$, the decoder $\psi_n : GF(q)^k \to GF(q)^n$ is defined as

$$\psi_n(m^k) \stackrel{\text{def}}{=} \arg \min_{u^n : u^n \boldsymbol{A} = m^k} H(P_{u^n}), \tag{3.4}$$

where P_{u^n} is a type of u^n (see Definition 2.1), and $H(Q)$ is an entropy function defined as

$$H(Q) \stackrel{\text{def}}{=} \sum_{a \in [0:q-1]} Q(a) \log \frac{1}{Q(a)}. \tag{3.5}$$

The problem addressed in this chapter is as follows:

[Problem Setting]

When the compression rate is given as R, and a sparse matrix $\boldsymbol{A}$ is given and fixed, evaluate the upper bound of the expectation value of the decoding error probability for any $P \in \mathcal{P}(\tau)$:

$$\sum_{u^n} P(u^n) \mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right]. \tag{3.6}$$

Remark 3.1

Since the decoder ψ_n adopts minimum entropy decoding (3.4) *and does not*

depend on the specific probability distribution $P \in \mathcal{P}(\tau)$, obviously the decoder has the universal property. To show that the encoder φ_n has the universal property, i.e., the sparse matrix $\boldsymbol{A}$ in (3.3) *does not depend on the specific probability distribution $P \in \mathcal{P}(\tau)$, it should be proven that the decoding error that is averaged over $\boldsymbol{A}$ after taking the operation $\sup_{P\in\mathcal{P}(\tau)}$ becomes* 0 *asymptotically with the block length n shown as* (3.6).

In this chapter, we study issues of the fixed length universal codes denoted above, which are problems of constructing universal codes with the optimal error exponent under the constant compression rate condition. Refer to Csiszár [8] for studies on the optimal error exponent of linear codes.

3.2 Main Theorem and Proof

3.2.1 Main Theorem

Before stating the main theorem of this chapter, we define a set $\mathcal{P}_\eta(\tau)$ for $0 < \eta < 1$, which is a subset of $\mathcal{P}(\tau)$:

$$\mathcal{P}_\eta(\tau) \stackrel{\text{def}}{=} \left\{ \tilde{P} \in \mathcal{P}(\tau) \;\middle|\; H(\tilde{P}) < \eta \right\}. \tag{3.7}$$

Note that from the definition of $\mathcal{P}(\tau)$, $0 < H(P) < 1$ for $P \in \mathcal{P}(\tau)$, and since the base of log is q, $H(P) \leq 1$.

Theorem 3.1 *[Error exponent of source code constructed by sparse matrices]*
For any $\eta > 0$ and sufficiently large n, there exists a sparse matrix $\boldsymbol{A}$ that constructs φ_n and ψ_n satisfying for any $P \in \mathcal{P}(\tau)$

$$\begin{aligned} &\sum_{u^n} P(u^n)\mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right] \\ &\leq q^{-n \inf_{\tilde{P}\in\mathcal{P}_{H(P)}(\tau)} \min_Q \left[D(Q||\tilde{P})+|R-H(Q)-\eta|^+\right]} \end{aligned} \tag{3.8}$$

with high probability. ■

Note that $\mathcal{P}_{H(P)}(\tau)$ is obtained by replacing η of $\mathcal{P}_\eta(\tau)$ by $H(P)$.

When $R \leq H(P)$, the upper bound of the decoding error shown in Theorem 3.1 is trivial. On the other hand, when $R > H(P)$, it is a nontrivial evaluation.

Remark 3.2
The upper bound of the decoding error shown in Theorem 3.1 *is identical to the one obtained for the linear code constructed using non-sparse matrices [8, Theorem* 2*] except the outer* inf *operation. When the compression rate* R *is in the interval* $(H(P), R_{cr})$ *for a* R_{cr}*, the obtained exponent can be shown to be optimal for the same reason as discussed by Csiszár [8]. Also for the same reason discussed there,* $\mathcal{P}(\tau)$ *can be extended to the class of* k*-th Markov sources without difficulty.*

When $q = 2$ (binary alphabet), $\tilde{P}$ satisfying $H(\tilde{P}) = H(P)$ is $\tilde{P} = P$ or $\tilde{P} = 1 - P$. Therefore, since by interchanging inf and min it holds that

$$\inf_{\tilde{P}\in\mathcal{P}_{H(P)}(\tau)} \min_{Q} \left[D(Q||\tilde{P}) + |R - H(Q)|^+\right] = \min_{Q} \left[D(Q||P) + |R - H(Q)|^+\right],$$

the next corollary is obtained. Note that when $q \geq 3$, the type of $\tilde{P}$ satisfying $H(\tilde{P}) = H(P)$ cannot be determined.

Corollary 3.1
When $q = 2$ *(binary alphabet), for any* $\eta > 0$ *and sufficiently large* n*, there exists a sparse matrix* $\boldsymbol{A}$ *that constructs* φ_n *and* ψ_n *satisfying for any* $P \in \mathcal{P}(\tau)$

$$\sum_{u^n} P(u^n)\mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right] \leq 2^{-n\min_Q\left[D(Q||P)+|R-H(Q)-\eta|^+\right]}$$

with high probability.

The next corollary simplifies the problem setting. Note that for any i.i.d. probability distribution P with $\min_{a\in[0:q-1]} P(a) > 0$, if $\tau_n \to \infty$ $(n \to \infty)$, then $\min_{a\in[0:q-1]} P(a) > \tau_n$ holds for sufficiently large n.

Corollary 3.2
Theorem 3.1 *and Corollary* 3.1 *hold even in the case of* $\tau = 0$.

For the proof of Corollary 3.2, we consider two cases:
Case 1. For all $a \in [0 : q-1]$, $P(a) > 0$.

The proof of Theorem 3.1 below also holds when we replace τ with τ_n, where $1/\tau_n$ grows with polynomial order of n.
Case 2. For some $a \in [0 : q-1]$, $P(a) = 0$.

In 2) of Lemma 3.2 below, q, an exponent of the upper bound of $|\hat{\mathcal{P}}_\eta(\tau)|$, is replaced with $|\{a \in [0 : q-1]|P(a) > 0\}|$, which is smaller than q, and the proof of Theorem 3.1 also holds.

3.2.2 Some Lemmas

In proving the theorem, the random coding expectation is taken over the expurgated ensemble. Since we delete the "bad code set" from the original ensemble, the expectation of decoding error can be decreased. The set from which the "bad code subset" is removed is called the expurgated ensemble [11] [28]. The next lemma shows that the bad code subset which will be removed has small probability. While in [11] the lemma was shown for sparse matrices constructed using bipartite graph correspondence, the following lemma is for sparse matrices constructed using the manner described in Section 2.2.1.

Lemma 3.1 *[Expurgation Lemma]*
Let $\boldsymbol{A}$ *be an* $n \times k$ *sparse matrix constructed using the manner described in Section* 2.2.1*,* $R = k/n$ *be fixed, and*

$$d_{\min}(\boldsymbol{A}) \stackrel{\text{def}}{=} \min_{x^n \neq 0^n : x^n \boldsymbol{A} = 0^k} w(x^n). \tag{3.9}$$

Then, for any $d \geq \gamma_R$*, there exist* $\delta_n(R)$ $(\delta_n(R) \to 0 \ (n \to \infty))$ *and* $\beta_n(R)$ $(\beta_n(R) \to 0 \ (n \to \infty))$ *satisfying*

$$P_{\boldsymbol{A}}\left[d_{\min}(\boldsymbol{A}) < nd\right] \leq \begin{cases} \beta_n(R) + 2(1+\delta_n(R))2^{-n\left(\frac{k}{n} - h(d)\right)}\mathbf{1}[d > \gamma_R], & q = 2 \\ \beta_n(R) + (1+\delta_n(R))q^{-n\left(\frac{k}{n} - h(d) - d\right)}\mathbf{1}[d > \gamma_R], & q \geq 3, \end{cases} \tag{3.10}$$

where $h(x)$ *is a binary entropy function defined as*

$$h(x) \stackrel{\text{def}}{=} -x\log(x) - (1-x)\log(1-x). \tag{3.11}$$

■

[Proof of Lemma 3.1*]*
From the definition of $d_{\min}$, it holds that

$$\begin{aligned}
&\mathbf{1}\left[d_{\min}(\boldsymbol{A}) \leq nd\right] = \mathbf{1}\left[\min_{z^n \neq 0^n : z^n \boldsymbol{A} = 0^k} w(z^n) \leq nd\right] \\
&= \mathbf{1}\left[\exists z^n \neq 0^n \text{ s.t. } z^n \boldsymbol{A} = 0^k \text{and } z^n \in \bigcup_{w=1}^{nd} \{w(z^n) = w\}\right] \\
&\leq \sum_{z^n : z^n \boldsymbol{A} = 0^k} \mathbf{1}\left[z^n \in \bigcup_{w=1}^{nd} \{w(z^n) = w\}\right] \\
&\leq \sum_{z^n : z^n \boldsymbol{A} = 0^k} \sum_{w=1}^{nd} \mathbf{1}\left[w(z^n) = w\right] \\
&\leq \sum_{w=1}^{nd} \sum_{z^n \in [0:q-1]^n} \mathbf{1}\left[w(z^n) = w\right] \mathbf{1}\left[z^n \boldsymbol{A} = 0^k\right]. \qquad (3.12)
\end{aligned}$$

Using this, we obtain

$$\begin{aligned}
&P_{\boldsymbol{A}}\left[d_{min}(\boldsymbol{A}) < nd\right] \\
&= \mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[d_{min}(\boldsymbol{A}) < nd\right] \\
&\leq \left\{\sum_{w=1}^{n\gamma_R} + \sum_{w=n\gamma_R+1}^{nd}\right\} \mathrm{E}_{\boldsymbol{A}} \sum_{z^n \in [0:q-1]^n} \mathbf{1}\left[w(z^n) = w\right] \mathbf{1}\left[z^n \boldsymbol{A} = 0^k\right] \\
&= \sum_{w=1}^{n\gamma_R} S(w) + \sum_{w=n\gamma_R+1}^{nd} \mathrm{E}_{\boldsymbol{A}} \sum_{z^n \in [0:q-1]^n} \mathbf{1}\left[w(z^n) = w\right] \mathbf{1}\left[z^n \boldsymbol{A} = 0^k\right]. \qquad (3.13)
\end{aligned}$$

Note that if $nd = n\gamma_R$, the second term of (3.13) vanishes. The first term of (3.13) is upper bounded by $\beta_n(R)$ in Lemma 2.5. On the other hand,

$$\begin{aligned}
&\sum_{w=n\gamma_R+1}^{nd} \mathrm{E}_{\boldsymbol{A}} \sum_{z^n\in[0:q-1]^n} \mathbf{1}\left[w(z^n)=w\right]\mathbf{1}\left[z^n\boldsymbol{A}=0^k\right] \\
&\overset{\text{(a)}}{=} \sum_{w=n\gamma_R+1}^{nd} \mathrm{E}_{\boldsymbol{A}} \sum_{z^n\in[0:q-1]^n:\ w(z^n)=w} \mathbf{1}\left[z^{*n}\boldsymbol{A}=0^k\right] \\
&= \sum_{w=n\gamma_R+1}^{nd} (q-1)^i \mathrm{E}_{\boldsymbol{A}} \sum_{z^{*n}\in\{0,1\}^n:\ w(z^{*n})=w} \mathbf{1}\left[z^{*n}\boldsymbol{A}=0^k\right] \\
&\overset{\text{(b)}}{=} \sum_{w=n\gamma_R+1}^{nd} (q-1)^w \frac{\binom{n}{w}\alpha_n(R;w)}{q^k} \\
&\overset{\text{(c)}}{\le} (1+\delta_n(R))q^{-n\left(\frac{k}{n}-d-h(d)\right)},
\end{aligned} \tag{3.14}$$

where at (a), Lemma 2.3 is used and $z_i^* \overset{\text{def}}{=} \begin{cases} 0, & z_i = 0 \\ 1, & z_i \neq 0 \end{cases}$, at (b), Lemma 2.5 (b) is used, and at (c), 1) of Lemma 2.7 and 2) of Lemma 2.5 are used.

[End of Proof of Lemma 3.1*]*

To prove the theorem, we will evaluate

$$\mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right] \tag{3.15}$$

for a fixed η $(0 < \eta < 1)$. After evaluating (3.15) and setting the upper bound as error(η), we will show that there exists the sparse matrix $\boldsymbol{A}$ that satisfies

$$\sup_{P\in\mathcal{P}_{i/L_n}(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right] \le L_n^2 \mathrm{error}(i/L_n) \tag{3.16}$$

for all $i \in [1 : L_n - 1]$ with high probability. When we set L_n as a polynomial of n that satisfies $L_n \to \infty$ $(n \to \infty)$, the proof is completed [44].

To evaluate the $\sup_{P\in\mathcal{P}_\eta(\tau)}$ part of (3.15), we use an approximation set of $\mathcal{P}_\eta(\tau)$. Let $\hat{\mathcal{P}}_\eta(\tau)$ be a set of positive functions $\hat{P} : [0 : q-1] \to \mathbf{R}^+$. Then $\hat{\mathcal{P}}_\eta(\tau)$ is an approximating set of $\mathcal{P}_\eta(\tau)$ if, for a given $\varepsilon > 0$ and for any $P \in \mathcal{P}_\eta(\tau)$, there exists $\hat{P} \in \hat{\mathcal{P}}_\eta(\tau)$ satisfying

$$\left|P(a) - \hat{P}(a)\right| \le \varepsilon P(a) \text{ for any } a \in [0 : q-1]. \tag{3.17}$$

The next lemma shows the existence of $\hat{\mathcal{P}}_\eta(\tau)$ and evaluation of $|\hat{\mathcal{P}}_\eta(\tau)|$.

Lemma 3.2 *[Approximating Set]*
For a given positive number $\varepsilon > 0$, there exists an approximating set $\hat{\mathcal{P}}_\eta(\tau)$ that has the following properties:
1) For any $P \in \mathcal{P}_\eta(\tau)$, there exists $\hat{P} \in \hat{\mathcal{P}}_\eta(\tau)$ satisfying

$$\left| P(a) - \hat{P}(a) \right| \leq \varepsilon P(a) \text{ for any } a \in [0 : q-1],$$

and
2)

$$|\hat{\mathcal{P}}_\eta(\tau)| \leq \left(\frac{1}{\tau\varepsilon} \right)^q,$$

where $\tau > 0$ is a positive number specified in the definition of the probability distribution class $\mathcal{P}(\tau)$ in (3.2). ■

[Proof of Lemma 3.2*]*
Consider a line $[0, 1]$ in $\mathbf{R}$ and divide it by width $\tau\varepsilon$. Make a correspondence between $P \in \mathcal{P}_\eta(\tau)$ and $\hat{P} \in \hat{\mathcal{P}}_\eta(\tau)$ by

$$\hat{P}(a) = \left\lfloor \frac{P(a)}{\tau\varepsilon} \right\rfloor \tau\varepsilon \text{ for any } a \in [0 : q-1],$$

and then it is straightforward, by using the definition of the considered class of probability distribution (3.2), to show that properties 1) and 2) above hold.

[End of Proof of Lemma 3.2*]*
Note that $\hat{P} \in \hat{\mathcal{P}}_\eta(\tau)$ is not necessarily a probability distribution on $\mathcal{U}$.

3.2.3 Evaluation of Decoding Error Probability

Let an expurgated ensemble $\mathcal{D}$ be

$$\mathcal{D} \stackrel{\text{def}}{=} \left\{ \text{Sparse matrix } \boldsymbol{A} \;\middle|\; d_{\min}(\boldsymbol{A}) > n\gamma_R \right\}. \tag{3.18}$$

Denote an expectation operation on $\mathcal{D}$ as $\mathrm{E}_{\boldsymbol{A}}^{(ex)}$, then since, from Lemma 3.1, we have $P_{\boldsymbol{A}}(\mathcal{D}) \geq 1 - \beta_n(R)$, we can obtain

$$\mathrm{E}_{\boldsymbol{A}}^{(ex)} \mathbf{1}[\cdot] \leq \frac{1}{1 - \beta_n(R)} \mathrm{E}_{\boldsymbol{A}} \mathbf{1}[\cdot] \mathbf{1}[\mathcal{D}]. \tag{3.19}$$

In the following, we evaluate

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}}^{(ex)} \sup_{P\in\mathcal{P}_\eta(\tau)} P\left[\text{Decoding error}\right] \\
&= \mathrm{E}_{\boldsymbol{A}}^{(ex)} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right]
\end{aligned} \tag{3.20}$$

instead of (3.6).

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}}^{(ex)} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[u^n \neq \psi_n\left(\varphi_n(u^n)\right)\right] \\
&\leq \mathrm{E}_{\boldsymbol{A}}^{(ex)} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[\exists \tilde{u}^n \neq u^n \ s.t.\ \tilde{u}^n\boldsymbol{A} = u^n\boldsymbol{A}\right. \\
&\qquad\qquad\qquad\qquad \left.\text{and } H(P_{\tilde{u}^n}) \leq H(P_{u^n})\right] \\
&\overset{\text{(a)}}{\leq} \frac{1}{1-\beta_n(R)} \\
&\qquad \cdot \mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[\exists \tilde{u}^n \neq u^n \ s.t.\ \tilde{u}^n\boldsymbol{A} = u^n\boldsymbol{A}\right. \\
&\qquad\qquad\qquad\qquad \left.\text{and } H(P_{\tilde{u}^n}) \leq H(P_{u^n})\right]\mathbf{1}\left[\mathcal{D}\right],
\end{aligned} \tag{3.21}$$

where (a) comes from (3.19).

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n)\mathbf{1}\left[\exists \tilde{u}^n \neq u^n \ s.t.\ \tilde{u}^n\boldsymbol{A} = u^n\boldsymbol{A} \ and \ H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \\
&\qquad \cdot \mathbf{1}\left[\mathcal{D}\right] \\
&\le \mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n) \sum_{\tilde{u}^n:\tilde{u}^n\neq u^n} \mathbf{1}\left[(\tilde{u}^n - u^n)\boldsymbol{A} = 0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \\
&\qquad \cdot \mathbf{1}\left[\mathcal{D}\right] \\
&= \mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n) \left(\sum_{w=1}^{n\gamma_R} + \sum_{w=n\gamma_R+1}^{n} \right) \\
&\qquad \cdot \sum_{\tilde{u}^n:w(\tilde{u}^{*n}-u^{*n})=w} \mathbf{1}\left[(\tilde{u}^{*n} - u^{*n})\boldsymbol{A} = 0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \mathbf{1}\left[\mathcal{D}\right] \\
&\overset{\text{(b)}}{=} \mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n) \sum_{w=n\gamma_R+1}^{n} \\
&\qquad \cdot \sum_{\tilde{u}^n:w(\tilde{u}^{*n}-u^{*n})=w} \mathbf{1}\left[(\tilde{u}^{*n} - u^{*n})\boldsymbol{A} = 0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \mathbf{1}\left[\mathcal{D}\right] \\
&\le \mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n) \sum_{w=n\gamma_R+1}^{n} \\
&\qquad \cdot \sum_{\tilde{u}^n:w(\tilde{u}^{*n}=u^{*n})=w} \mathbf{1}\left[(\tilde{u}^{*n} - u^{*n})^*\boldsymbol{A} = 0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right],
\end{aligned}
\tag{3.22}
$$

where (b) comes from the definition of $\mathcal{D}$ and Lemma 2.3 in which $u_i^* \overset{\text{def}}{=} \begin{cases} 0, & u_i = 0 \\ 1, & u_i \neq 0 \end{cases}$.

To evaluate (3.22), we use Approximating Set Lemma (Lemma 3.2) as

follows:

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}} \sup_{P\in\mathcal{P}_\eta(\tau)} \sum_{u^n} P(u^n) \\
&\qquad \cdot \sum_{w=n\gamma_R+1}^{n} \sum_{\tilde{u}^n : w(\tilde{u}^{*n}-u^{*n})=w} \mathbf{1}\left[(\tilde{u}^{*n}-u^{*n})\boldsymbol{A}=0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \\
&\overset{\text{(c)}}{\le} \mathrm{E}_{\boldsymbol{A}} \sup_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \frac{\hat{P}(u^n)}{(1-\varepsilon)^n} \\
&\qquad \cdot \sum_{w=n\gamma_R+1}^{n} \sum_{\tilde{u}^n : w(\tilde{u}^{*n}-u^{*n})=w} \mathbf{1}\left[(\tilde{u}^{*n}-u^{*n})\boldsymbol{A}=0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \\
&\overset{\text{(d)}}{\le} \mathrm{e}^{2n\varepsilon} \mathrm{E}_{\boldsymbol{A}} \sup_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \hat{P}(u^n) \\
&\qquad \cdot \sum_{w=n\gamma_R+1}^{n} \sum_{\tilde{u}^n : w(\tilde{u}^{*n}-u^{*n})=w} \mathbf{1}\left[(\tilde{u}^{*n}-u^{*n})\boldsymbol{A}=0^k\right] \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \\
&\le \mathrm{e}^{2n\varepsilon} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \hat{P}(u^n) \\
&\qquad \cdot \sum_{w=n\gamma_R+1}^{n} \sum_{\tilde{u}^n : w(\tilde{u}^{*n}-u^{*n})=w} \mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[(\tilde{u}^{*n}-u^{*n})\boldsymbol{A}=0^k\right] \\
&\qquad\qquad \cdot \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \\
&\overset{\text{(e)}}{\le} \mathrm{e}^{2n\varepsilon} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \hat{P}(u^n) \sum_{w=n\gamma_R+1}^{n} \sum_{\tilde{u}^n} \mathbf{1}\left[H(P_{\tilde{u}^n}) \le H(P_{u^n})\right] \frac{1+\delta_n(R)}{q^k} \\
&\overset{\text{(f)}}{\le} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \hat{P}(u^n) n(n+1)^q (1+\delta_n(R)) q^{-n(R-H(P_{u^n})-2\varepsilon)}, \qquad (3.23)
\end{aligned}
$$

where at (c), we use Lemma 3.2 while noting that type P_{u^n} in the indicator function $\mathbf{1}[\cdot]$ does not depend on the probability distribution $P(u^n)$ nor $\hat{P}(u^n)$, at (d) we assume $\varepsilon < 1/2$ and use the fact $\frac{1}{1-\varepsilon} < 1+2\varepsilon$ and $1+2\varepsilon < \mathrm{e}^{2\varepsilon}$, (e) comes from 2) of Lemma 2.5, and (f) is derived using 2) and 3) of Lemma 2.7.

Using the above result, we can evaluate (3.20) as

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}}^{(ex)} \sup_{P\in\mathcal{P}_\eta(\tau)} P\left[\text{Decoding error}\right] \\
&\le \frac{1}{1-\beta_n(R)} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \hat{P}(u^n) n(n+1)^q (1+\delta_n(R)) q^{-n|R-H(P_{u^n})-2\varepsilon|^+},
\end{aligned} \tag{3.24}$$

where $|x|^+ \stackrel{\text{def}}{=} \max\{x, 0\}$. The final step is as follows, by using ordinary techniques of types:

$$\begin{aligned}
&\frac{n(n+1)^q(1+\delta_n(R))}{1-\beta_n(R)} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{u^n} \hat{P}(u^n) q^{-n|R-H(P_{u^n})-2\varepsilon|^+} \\
&= \frac{n(n+1)^q(1+\delta_n(R))}{1-\beta_n(R)} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{Q\in\mathcal{Q}_n} \sum_{u^n\in T_Q^n} q^{-n\left(H(Q)+D(Q||\hat{P})+|R-H(P_{u^n})-2\varepsilon|^+\right)} \\
&= \frac{n(n+1)^q(1+\delta_n(R))}{1-\beta_n(R)} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} \sum_{Q\in\mathcal{Q}_n} |T_Q^n| q^{-n\left(H(Q)+D(Q||\hat{P})+|R-H(P_{u^n})-2\varepsilon|^+\right)} \\
&\stackrel{\text{(g)}}{\le} \frac{n(n+1)^{2q}(1+\delta_n(R))}{1-\beta_n(R)} \sum_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)} q^{-n\min_Q\left(D(Q||\hat{P})+|R-H(Q)-2\varepsilon|^+\right)} \\
&\stackrel{\text{(h)}}{\le} \frac{n(n+1)^{2q}(1+\delta_n(R))}{1-\beta_n(R)} \left(\frac{1}{\tau\varepsilon}\right)^q \\
&\qquad\qquad \cdot q^{-n\inf_{P\in\mathcal{P}_\eta(\tau)}\min_Q\left(D(Q||P)-\log(1+\varepsilon)+|R-H(Q)-2\varepsilon|^+\right)},
\end{aligned} \tag{3.25}$$

where (g) comes from using 2) and 3) of Lemma 2.7, and (h) is from using 2) of Lemma 3.2 and $D(Q||\hat{P}) \ge D(Q||P) - \log(1+\varepsilon)$ from 1) of Lemma 3.2. Since $\varepsilon > 0$ can be made arbitrarily small, $\inf_{\hat{P}\in\hat{\mathcal{P}}_\eta(\tau)}$ can be replaced by $\inf_{P\in\mathcal{P}_\eta(\tau)}$ in (3.25).

In the following part of the proof, we take η as i/L_n for $i \in [1 : L_n - 1]$ with a large number L_n specified later. By setting the right hand side of

(3.25) as error(i/L_n) and using Markov's inequality,

$$\begin{aligned}
&\mathrm{P}_{\boldsymbol{A}}^{(ex)}\left[\left\{\sup_{P\in\mathcal{P}_{1/L_n}(\tau)} P[\text{Decoding error}] > L_n^2\text{error}(1/L_n)\right\}\cup\right.\\
&\qquad\left.\cdots\cup\left\{\sup_{P\in\mathcal{P}_{(L_n-1)/L_n}(\tau)} P[\text{Decoding error}] > L_n^2\text{error}((L_n-1)/L_n)\right\}\right]\\
&\le\sum_{i=1}^{L_n-1}\mathrm{P}_{\boldsymbol{A}}^{(ex)}\left[\sup_{P\in\mathcal{P}_{i/L_n}(\tau)} P[\text{Decoding error}] > L_n^2\text{error}(i/L_n)\right]\\
&\le\sum_{i=1}^{L_n-1}\frac{\mathrm{E}_{\boldsymbol{A}}\sup_{P\in\mathcal{P}_{i/L_n}(\tau)} P[\text{Decoding error}]}{L_n^2\text{error}(i/L_n)}\\
&\le\frac{1}{L_n}. \qquad (3.26)
\end{aligned}$$

When we set L_n as a polynomial of n satisfying $L_n\to\infty$ $(n\to\infty)$, we can take the sparse matrix $\boldsymbol{A}$ that satisfies

$$\sup_{P\in\mathcal{P}_{i/L_n}(\tau)} P[\text{Decoding error}] \le L_n^2\text{error}(i/L_n) \qquad (3.27)$$

for all $i\in[1:L_n-1]$ with high probability, and the proof is completed. ∎

3.3 Simulation Results

In this section we discuss simulation results that show the universal property of the code constructed by sparse matrices in Theorem 3.1 and compare the error exponent obtained experimentally with the theoretical error exponent. Note that we adopt sum-product decoding instead of minimum entropy decoding in the interest of computational efficiency.

In simulations, after sequences from the i.i.d. source are encoded into codewords, the codewords are decoded by sum-product decoding and the decoding error is computed. We conducted simulation experiments for a binary alphabet case ($q=2$). Note that when $q=2$, the minimum entropy decoding is equivalent to the maximum likelihood (ML) decoding, and the sum-product algorithm approximately implements the ML decoding.

Figure 3.2 shows the plots of decoding errors for various sources. The sources are i.i.d., and probability $P_U(U=1)$ is taken as a probability parame-

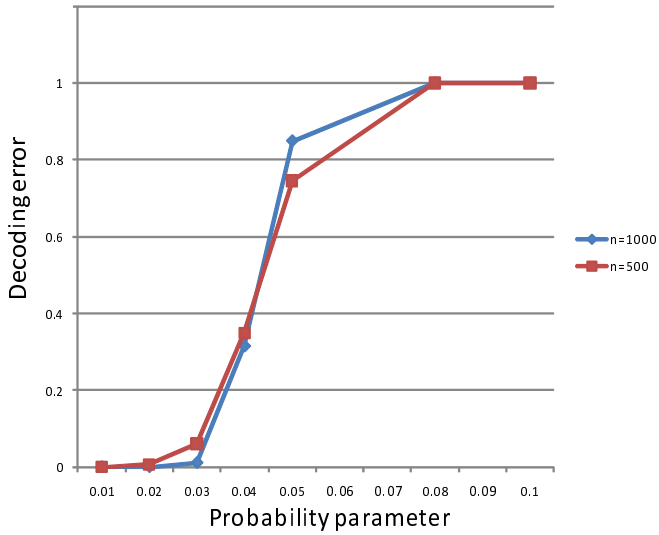

Figure 3.2: Simulation of universal code: $R = 0.5$

ter, which is represented by the horizontal axis. Decoding error is represented by the vertical axis.

The plot labeled "$n = 1000$" shows decoding error for the code with which the sparse matrix and sum-product decoding parameters were fixed during simulation when block length $n = 1000$ and compression rate $R = 0.5$. For each source, 1000 sample sequences were encoded and decoded, and decoding error was computed as the relative frequency of the error for the samples. For the plot labeled "$n = 500$", the block length condition was $n = 500$ and other conditions were the same as in the case of $n = 1000$. Both plots show that decoding error tends to decrease as source parameters decrease. In both cases, decoding error drops drastically around $P_U(U = 1) = 0.03$. Since the compression rate R was 0.5 and $h(0.11) \simeq 0.5$, if we used an idealistic encoder and decoder, under the parameter range less than $P_U(U = 1) = 0.1$, decoding errors of the simulation would approach 0. The difference between the experiments and theory seems to come from the approximate implementation of the ML decoding by the sum-product algorithm we adopted.

Figure 3.3 shows decoding errors under conditions of varying block length with a fixed compression rate ($R = 0.5$) and probability parameter ($p =$

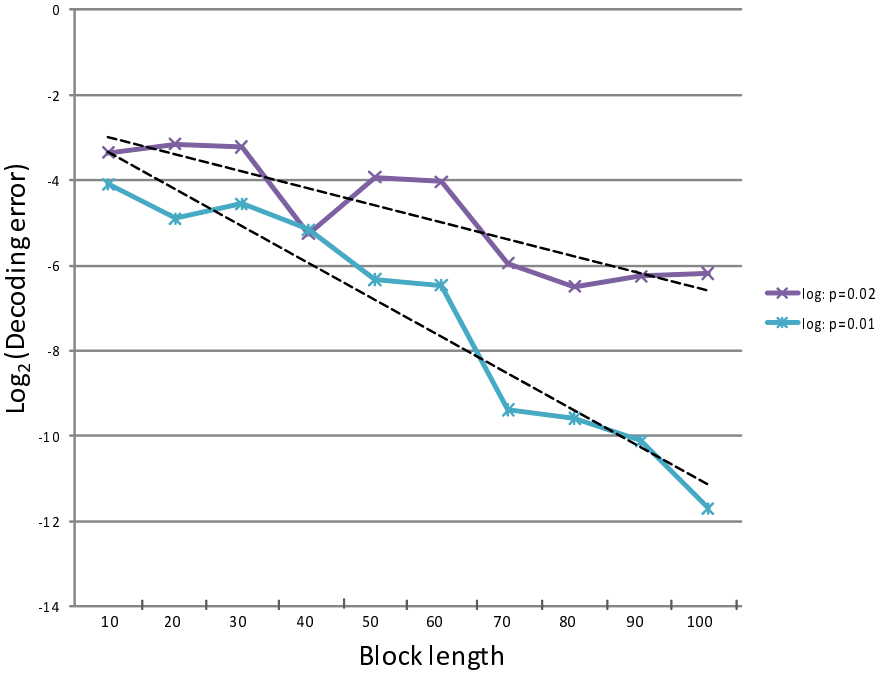

Figure 3.3: Simulation of error exponent: $R = 0.5$

0.01, 0.02). In each block length, 10,000 samples were simulated. For data of $p = 0.01$ and $p = 0.02$, the computed exponents were 0.087 and 0.040, respectively. The theoretical values computed by the formula

$$\min_{Q} \left\{ D(Q||P) + |R - H(Q)|^{+} \right\} \tag{3.28}$$

for $R = 0.5$ and $P_U(U = 1) = 0.01$ or $P_U(U = 1) = 0.02$ are 0.238 and 0.147, respectively. The exponent obtained in the simulation is about $1/3 \sim 1/4$ the size of the one obtained theoretically. The difference also seems to come from the approximate implementation of the ML decoding by the sum-product algorithm we adopted.

3.4 Concluding Remarks

A universal code for a class of i.i.d. sources was constructed using sparse matrices, and expectation of decoding error using minimum entropy decoding

was upper bounded by an exponential function of block length n. A simple simulation showed the universality of the code constructed by sparse matrices and sum-product decoding with a fixed parameter for a class of binary i.i.d. sources.

A construction of an efficient decoding scheme with a non-binary alphabet was recently proposed by Coleman, Médard and Effros [6]. Studying application of their decoding scheme to our universal coding scheme and construction of an efficient universal decoding algorithm with a non-binary alphabet will be the subject of our next study.

In this chapter, to elucidate methods for analyzing sparse matrix coding properties, we focused on the simplest $\mathcal{P}(\tau)$ that is the set of i.i.d. sources. For the case that $\mathcal{P}(\tau)$ takes other sets such as an arbitrarily varying source (AVS [7]) or correlated sources, similar results were obtained in [30].

Chapter 4

Lossy Source Coding

Matsunaga and Yamamoto [27] constructed a lossy source code that asymptotically attains optimality of the rate-distortion function under the conditions of a binary alphabet, uniform distribution, and Hamming distortion measure. Similar results were obtained by Martinian and Wainwright [25]. Miyake [29] extended Matsunaga and Yamamoto's results to a non-binary alphabet condition. Zamir, Shamai and Erez [46] proposed nested linear/lattice codes.

Note that the results referred to above hold under limited conditions, such as a uniform distribution of information sources, and that some results do not attain a theoretically optimal coding rate. In this chapter, we construct a source code with a fidelity criterion using sparse matrices for arbitrary discrete stationary memoryless sources that is not necessarily uniform, and we demonstrate the asymptotic optimality of the code. In the construction of sparse matrices, "sparse" matrices have 1's of $O(n \log n)$ like those constructed by Matsunaga and Yamamoto [27]. By combining a coding scheme constructed using these sparse matrices with an efficient algorithm such as the LP decoding algorithm [12] or the sum-product algorithm, the computing time of coding or decoding processes can be estimated by the polynomial order of block length n, but not by the linear of n.

Bennatan [2] and Erez [11] have already investigated the theory of sparse matrices over $GF(q)$. They constructed a channel code by extending the bipartite graph method, through which it is not easy to strictly evaluate the the probability that a sequence with a given weight becomes a codeword. In this chapter, constructing sparse matrix code that corresponds to a random walk makes accurate evaluation of the above probability possible, leading to

a proof of the lossy source coding theorem for arbitrary discrete stationary memoryless sources.

An efficient algorithm, such as the LP decoding or the sum-product algorithm, is available using matrices for encoding and decoding. The results obtained here will make good algorithms of, for example, vector quantization, for which only a heuristic algorithm is currently available for the case of a large block size [16].

4.1 Preliminaries and Problem Setting

The alphabet focused on here is a set $[0 : q-1]$, where q is a prime number and the set is also considered as a field $GF(q)$. The basis of ln is e and of log is q.

Notations are defined as follows.

Information Sources:

A probability distribution of the source is assumed to be stationary and memoryless and is denoted as P_U with the generic random variable U.

Encoder and Decoder:

The encoder and decoder are denoted as $\varphi_n : GF(q)^n \to GF(q)^k$ and $\psi_n : GF(q)^k \to GF(q)^n$, respectively. In this setting, the compression rate R is defined as $R \stackrel{\text{def}}{=} k/n$.

Fidelity Criterion:

Let the bounded and additive distortion measure be $d_n : GF(q)^n \times GF(q)^n \to \boldsymbol{R}$. For a given constant D, the fidelity criterion adopted here is

$$\lim_{n\to\infty} P_{U^n}\left[\frac{d_n\left(U^n, \psi_n\left(\varphi_n(U^n)\right)\right)}{n} > D\right] = 0. \tag{4.1}$$

A typical example of a distortion measure is Hamming measure $d_n(u^n, v^n) \stackrel{\text{def}}{=} \sum_{i=1}^n d_H(u_i, v_i)$, where $d_H(a,b) \stackrel{\text{def}}{=} \begin{cases} 0, & a = b, \\ 1, & a \neq b. \end{cases}$

Rate-Distortion Function:

The minimum value of the compression rate with which φ_n and ψ_n

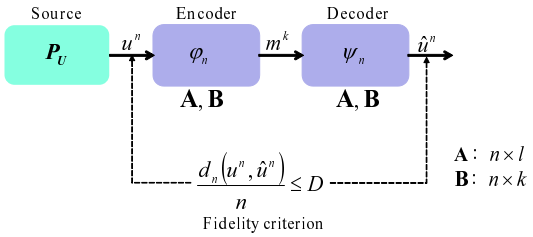

Figure 4.1: Lossy source coding problem

satisfy the fidelity criterion is given by the rate-distortion function, $R(D)$ (e.g., [7], [9]). $R(D)$ is obtained by the formula

$$R(D) = \min_{P_{V|U}:\sum_{a,b} P_U(a)P_{V|U}(b|a)d_1(a,b)\le D} I(U;V). \tag{4.2}$$

When $q = 2$, if we use a Hamming measure as d_1 in the formula, $R(D)$ can be explicitly computed as $R(D) = h(p) - h(D)$ where $p \stackrel{\text{def}}{=} P_U(U = 1)$, and $h(x)$ is the binary entropy function defined as $h(x) \stackrel{\text{def}}{=} -x\log(x) - (1-x)\log(1-x)$.

With the above preparation, the problem considered is as follows.
[**Problem**:]
When a fidelity criterion is given for a stationary memoryless source P_U, construct the encoder φ_n and the decoder ψ_n using sparse matrices, of which the compression rate approaches the rate-distortion function $R(D)$ asymptotically with block length n.

Figure 4.1 shows a block diagram of the lossy source coding problem.

4.2 Main Theorem and Proofs

4.2.1 Main Theorem

Theorem 4.1
Let a conditional probability distribution $P_{V|U}$ be given and fixed. If there exists a positive number δ that satisfies $\frac{l+k}{n} > H(V)+\delta^{1/3}$ and $\frac{l}{n} < H(V|U)-\delta$ for sufficiently large n, l, and k, then an encoder φ_n and a decoder ψ_n can

be constructed by an $n \times l$ sparse matrix $\boldsymbol{A}$ and an $n \times k$ sparse matrix $\boldsymbol{B}$, which satisfy the fidelity criterion

$$P_{U^n}\left[\frac{d_n\left(U^n, \psi_n\left(\varphi_n(U^n)\right)\right)}{n} > D\right] \to 0 \quad (n \to \infty). \tag{4.3}$$

■

Remark 4.1
Assume that the conditions of Theorem 4.1, $\frac{l+k}{n} > H(V) + \delta^{1/3}$ *and* $\frac{l}{n} < H(V|U) - \delta$, *are satisfied for a sufficiently small* δ. *If we set* $\frac{l+k}{n} = H(V) + 2\delta^{1/3}$ *and* $\frac{l}{n} = H(V|U) - 2\delta$, *then the coding rate* $\frac{k}{n}$ *is equal to* $I(U;V) + 2(\delta + \delta^{1/3})$. *This observation shows that if* $I(U;V)$ *is equal to the rate-distortion function* $R(D)$, *the coding rate can asymptotically approach the rate-distortion limit.*

Note that it is well known that the conditional probability, which makes $I(U;V)$ *equal to* $R(D)$, *can be computed by the Arimoto-Blahut algorithm (e.g., [9], [7]).*

4.2.2 Construction of Encoder φ_n and Decoder ψ_n

In this subsection, construction of an encoder and a decoder using sparse matrices is shown. For the construction, $n \times l$ sparse matrix $\boldsymbol{A}$, $n \times k$ sparse matrix $\boldsymbol{B}$, and l-dimensional row vector c^l are used. Both matrices are constructed following the manner described in Section 2.1.1 ($q = 2$) or 2.2.1 ($q \geq 3$). c^l is taken as a non-zero row vector, and especially when $q = 2$, the Hamming weight of c^l is taken to be even.

Note that random variables $\boldsymbol{A}$, $\boldsymbol{B}$ are independent of each other, and we use $\mathrm{P}_{\boldsymbol{AB}}[\cdot]$ and $\mathrm{E}_{\boldsymbol{AB}}[\cdot]$ as the probability distribution and expectation operation over the random variables $\boldsymbol{A}$, $\boldsymbol{B}$, respectively. Assume that $\boldsymbol{A}$, $\boldsymbol{B}$, the realization value of corresponding random variables, and a fixed row vector c^l are known to both encoder and decoder.

Construction of Encoder φ_n

Figure 4.2 shows an outline of the encoding process. An encoder φ_n consists of a quantization part and a compression part. Each part is constructed with sparse matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as follows.

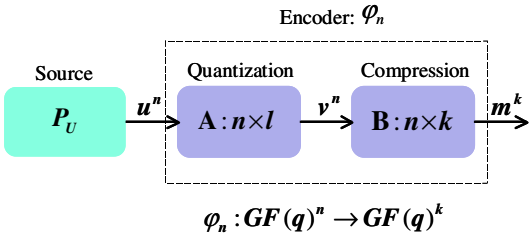

Figure 4.2: Construction of encoder

Let the output sequence of the source be u^n and the output of the quantization part be v^n defined as

$$v^n \stackrel{\text{def}}{=} \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} P_{V^n|U^n}(\tilde{v}^n|u^n), \tag{4.4}$$

where $c^l \in GF(q)^l$ is a fixed row vector. Note that when $q = 2$, the Hamming weight of c^l must be even, since the parity of the weight $w(c^l)$ is equal to that of $w(\tilde{v}^n) \times t$ and t is even from the assumption of the sparse matrix parameter.

Codeword $m^k = \varphi_n(u^n)$ is given by

$$m^k \stackrel{\text{def}}{=} v^n \boldsymbol{B}, \tag{4.5}$$

where m^k can be regarded as the output of the compression part.

Construction of Decoder ψ_n

Figure 4.3 shows an outline of the decoding process. When codeword $m^k \in GF(q)^k$ is given, decoding sequence $\hat{u}^n = \psi_n(m^k)$ is obtained by

$$\hat{u}^n \stackrel{\text{def}}{=} \arg \max_{\substack{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l \\ \tilde{v}^n \boldsymbol{B} = m^k}} P_{V^n}(\tilde{v}^n). \tag{4.6}$$

Note that if decoding is performed correctly, $v^n = \hat{u}^n$ holds.

Remark 4.2
When $q = 2$ (binary alphabet) and $P_U(U = 1) = 0.5$ (uniform distribution)

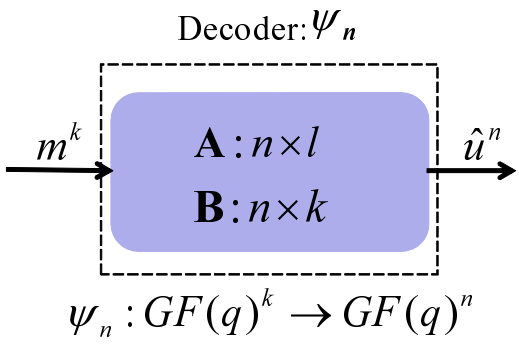

Figure 4.3: Construction of decoder

with the Hamming distance as the distortion measure, conditions in Theorem 4.1 *are stated as*

$$\frac{l+k}{n} > 1 + \delta^{1/3}$$

and

$$\frac{l}{n} < h(D) - \delta.$$

Since P_V *also becomes the uniform distribution, the decoder* (4.6) *obtains the decoding sequence* $\hat{u}^n$ *by solving linear equations* $\tilde{v}^n \boldsymbol{A} = c^l$ *and* $\tilde{v}^n \boldsymbol{B} = m^k$. *From the condition* $\frac{l+k}{n} > 1 + \delta^{1/3}$, *it holds that the rank of the* $n \times (l+k)$ *concatenated matrix of* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *becomes* n *with high probability.*

Remark 4.3
When we set $D = 0$, *since*

$$P_{V|U}(v|u) = \begin{cases} 1, & \text{if } u = v \\ 0, & \text{otherwise,} \end{cases}$$

conditions of Theorem 4.1 *are stated as*

$$\frac{l+k}{n} > H(U) + \delta^{1/3}$$

and

$$\frac{l}{n} < -\delta.$$

In this case, since l and $\boldsymbol{A}$ are void, the first condition $\frac{k}{n} > H(U) + \delta^{1/3}$ becomes the existence condition of the lossless source encoder and decoder constructed by the sparse matrix $\boldsymbol{B}$.

4.3 Proof of Theorem

In this section, it is shown that encoder φ_n and decoder ψ_n constructed above satisfy optimality of the compression rate and the fidelity criterion simultaneously.

For source output sequence u^n, if the encoder outputs v^n, which is a jointly typical sequence with respect to probability distribution P_{UV}, it can be shown that the distortion between u^n and its quantization v^n, $d_n(u^n, v^n)/n$, becomes asymptotically upper bounded by the constant D in the fidelity criterion as follows: Throughout the proof, $\varepsilon > 0$ is a given positive constant and fixed. Since $(u^n, v^n) \in T^n_{P_{UV}\varepsilon}$,

$$\begin{aligned}
\frac{1}{n} d_n(u^n, v^n) &= \frac{1}{n} \sum_{i=1}^{n} d_1(u_i, v_i) \\
&= \sum_{a,b} P_{u^n v^n} d_1(a, b) \\
&\leq \sum_{a,b} \{P(a, b) + |P(a, b) - P_{u^n v^n}(a, b)|\} \, d_1(a, b) \\
&\leq \sum_{a,b} P(a, b)(1 + \varepsilon) d_1(a, b) \\
&\leq (1 + \varepsilon) D. \qquad (4.7)
\end{aligned}$$

By use of the above observation and the fact that the output of the source is in a typical sequence set with high probability, to prove the theorem it is sufficient to show that the probability that the output of the decoder is not jointly typical with the input of the encoder becomes asymptotically 0.

In the remainder of this section, it will be shown that the output of "quantizer" in the encoder, v^n (see Figure 4.2), which is a jointly typical sequence with the input of the encoder, u^n, coincides with the output of the decoder with high probability, and that, at the same time, compression rate $R = k/n$ asymptotically approaches the rate-distortion function $R(D) = I(U; V)$.

In the proof of Theorem 4.1, we use a random coding technique over sparse matrices $\boldsymbol{A}$, $\boldsymbol{B}$.

Note that

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}} P_U^n \left[(U^n, \psi_n(\varphi_n(U^n)) \notin T^n_{P_{UV}\hat{\delta}(\varepsilon)} \text{ for } \exists \hat{\delta}(\varepsilon) \to 0 \ (\varepsilon \to 0) \right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_1 \bigcup \mathcal{E}_2 \bigcup \mathcal{E}_3 \right] \\
&\leq P_U^n \left[\mathcal{E}_1 \right] + \mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_2 \bigcap \mathcal{E}_1^c \right] + \mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_3 \bigcap \mathcal{E}_2^c \bigcap \mathcal{E}_1^c \right], \qquad (4.8)
\end{aligned}$$

where

$$\mathcal{E}_1 \stackrel{\text{def}}{=} \left\{ U^n \notin T^n_{P_{U\varepsilon}} \right\}, \qquad (4.9)$$

$$\mathcal{E}_2 \stackrel{\text{def}}{=} \left\{ V^n \notin T^n_{P_{V|U}\delta'(\varepsilon)}(U^n) \text{ for } \exists \delta'(\varepsilon) \to 0 \ (\varepsilon \to 0) \right\}, \qquad (4.10)$$

$$\mathcal{E}_3 \stackrel{\text{def}}{=} \left\{ V^n \neq \psi_n \left(\varphi_n(U^n) \right) \right\}, \qquad (4.11)$$

and in (4.10) and (4.11), V^n is the output of the "quantizer" part of the encoder (see Figure 4.2). $\mathcal{E}_2$ and $\mathcal{E}_3$ refer to "encoding error" and "decoding error", respectively.

From the law of large numbers,

$$P_U^n \left[\mathcal{E}_1 \right] \to 0 \ (n \to \infty). \qquad (4.12)$$

In the next sections, evaluations of $\mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_2 \bigcap \mathcal{E}_1^c \right]$ and $\mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_3 \bigcap \mathcal{E}_2^c \bigcap \mathcal{E}_1^c \right]$ are described.

4.3.1 Evaluation of $\mathbf{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_2 \bigcap \mathcal{E}_1^c \right]$

Let

$$G_{\boldsymbol{A}}(c^l) \stackrel{\text{def}}{=} \left\{ v^n \in [0 : q-1]^n \ \middle| \ v^n \boldsymbol{A} = c^l \right\}, \qquad (4.13)$$

$$\mathcal{E}_{21} \stackrel{\text{def}}{=} \left\{ G_{\boldsymbol{A}}(c^l) \bigcap T^n_{P_{V|U}\varepsilon}(u^n) = \emptyset \right\}, \qquad (4.14)$$

and

$$\mathcal{E}_{22} \stackrel{\text{def}}{=} \left\{ V^n \notin G_{\boldsymbol{A}}(c^l) \bigcap T^n_{P_{V|U}\delta'(\varepsilon)}(u^n) \text{ for } \exists \delta'(\varepsilon) \to 0 \ (\varepsilon \to 0) \right\}, \qquad (4.15)$$

then note that

$$\mathcal{E}_2 \subset \mathcal{E}_{21} \bigcup \mathcal{E}_{22}. \tag{4.16}$$

Therefore, we have

$$\mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_2 \bigcap \mathcal{E}_1^c \right] \leq \mathrm{E}_{\boldsymbol{A}} P_U^n \left[\mathcal{E}_{21} \bigcap \mathcal{E}_1^c \right] + \mathrm{E}_{\boldsymbol{AB}} P_U^n \left[\mathcal{E}_{22} \bigcap \mathcal{E}_{21}^c \bigcap \mathcal{E}_1^c \right]. \tag{4.17}$$

Note that $\mathcal{E}_{21}$ does not depend on $\boldsymbol{B}$.

For $\mathrm{E}_{\boldsymbol{A}} P_U^n [\mathcal{E}_{21} \bigcap \mathcal{E}_1^c]$, the next lemma holds.

Lemma 4.1
For any $u^n \in T^n_{P_{U\varepsilon}}$, if there exists a positive number δ that satisfies $l/n < H(V|U) - \delta$ for sufficiently large n and l, then it holds that

$$\lim_{n\to\infty} \mathrm{E}_{\boldsymbol{A}} P_U^n \left[\mathcal{E}_{21} \bigcap \mathcal{E}_1^c \right] = 0. \tag{4.18}$$

■

[Proof of Lemma 4.1*]*
We prove the case $q \geq 3$. The case $q = 2$ can be proven similarly.

Let $R' = l/n$ and set $\gamma_{R'} < \min_{b:P_V(b)>0} P_V(b)(1-\varepsilon)$. Then, note that for any $v^n \in T^n_{P_{V|U^\varepsilon}}(u^n)$ with $u^n \in T^n_{P_{U^\varepsilon}}$, the Hamming weight of v^n satisfies $w(v^n) > n\gamma_{R'}$.

To prove the lemma, it is sufficient to show that for any $u^n \in T^n_{P_{U^\varepsilon}}$,

$$P_{\boldsymbol{A}} \left\{ \sum_{v^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \mathbf{1}\left[v^n \boldsymbol{A} = c^l \right] = 0 \right\} \to 0 \quad (n \to \infty), \tag{4.19}$$

where $\mathbf{1}$["logical equation"] is an indicator function for the logical equation.

$$
\begin{aligned}
&P_{\boldsymbol{A}}\left\{\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\mathbf{1}\left[v^n\boldsymbol{A}=c^l\right]=0\right\}\\
&=P_{\boldsymbol{A}}\left\{\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\left(\mathbf{1}\left[v^n\boldsymbol{A}=c^l\right]-\frac{\alpha_n(R';w(v^n))}{q^l}\right)\right.\\
&\qquad\qquad\left.=-\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\frac{\alpha_n(R';w(v^n))}{q^l}\right\}\\
&\le P_{\boldsymbol{A}}\left\{\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\left(\frac{\alpha_n(R';w(v^n))}{q^l}-\mathbf{1}\left[v^n\boldsymbol{A}=c^l\right]\right)\right.\\
&\qquad\qquad\left.\ge\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\frac{\alpha_n(R';w(v^n))}{q^l}\right\}\\
&\overset{\text{(a)}}{\le}\frac{\mathrm{E}_{\boldsymbol{A}}\left|\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\left(\frac{\alpha_n(R';w(v^n))}{q^l}-\mathbf{1}\left[v^n\boldsymbol{A}=c^l\right]\right)\right|^2}{\left|\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\frac{\alpha_n(R';w(v^n))}{q^l}\right|^2}\\
&\overset{\text{(b)}}{\le}\frac{\mathrm{E}_{\boldsymbol{A}}\left|\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\left(\frac{\alpha_n(R';w(v^n))}{q^l}-\mathbf{1}\left[v^n\boldsymbol{A}=c^l\right]\right)\right|^2}{\left|\frac{(1-\delta_n(R'))|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right|^2},
\end{aligned}
\tag{4.20}
$$

where (a) comes from Chebyshev's inequality, and at (b), 2) of Lemma 2.5 with Remark 2.4 is used for the denominator.

On the other hand,

$$
\begin{aligned}
&\text{(Numerator of (4.20))}\\
&= \mathrm{E}_{\boldsymbol{A}} \sum_{v^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \left| \left(\frac{\alpha_n(R'; w(v^n))}{q^l} - \mathbf{1}\left[v^n \boldsymbol{A} = c^l\right] \right) \right|^2 \\
&\quad + \mathrm{E}_{\boldsymbol{A}} \sum_{v^n \neq \tilde{v}^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \left\{ \frac{\alpha_n(R'; w(v^n))}{q^l} \frac{\alpha_n(R'; w(\tilde{v}^n))}{q^l} \right. \\
&\qquad - \frac{\alpha_n(R'; w(v^n)) \mathbf{1}\left[\tilde{v}^n \boldsymbol{A} = c^l\right]}{q^l} \\
&\qquad \left. - \frac{\alpha_n(R'; w(\tilde{v}^n)) \mathbf{1}\left[v^n \boldsymbol{A} = c^l\right]}{q^l} + \mathbf{1}\left[v^n \boldsymbol{A} = c^l\right] \mathbf{1}\left[\tilde{v}^n \boldsymbol{A} = c^l\right] \right\} \\
&\overset{\text{(c)}}{=} \sum_{v^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \left\{ \frac{\alpha_n(R'; w(v^n))}{q^l} - \left(\frac{\alpha_n(R'; w(v^n))}{q^l} \right)^2 \right\} \\
&\quad - \sum_{v^n \neq \tilde{v}^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \frac{\alpha_n(R'; w(v^n))}{q^l} \frac{\alpha_n(R'; w(\tilde{v}^n))}{q^l} \\
&\quad + \mathrm{E}_{\boldsymbol{A}} \sum_{v^n \neq \tilde{v}^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \mathbf{1}\left[v^n \boldsymbol{A} = c^l\right] \mathbf{1}\left[\tilde{v}^n \boldsymbol{A} = c^l\right], \qquad (4.21)
\end{aligned}
$$

where at (c), 2) of Lemma 2.5 with Remark 2.4 is used.

The third term of (4.21) can be evaluated as follows:

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}} \sum_{v^n \neq \tilde{v}^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \mathbf{1}\left[v^n \boldsymbol{A} = c^l\right] \mathbf{1}\left[\tilde{v}^n \boldsymbol{A} = c^l\right] \\
&\overset{\text{(d)}}{=} \mathrm{E}_{\boldsymbol{A}} \sum_{v^n \in T^n_{P_{V|U^\varepsilon}}(u^n)} \left(\sum_{w=1}^{n\gamma_{R'}} + \sum_{w=n\gamma_{R'}+1}^{n} \right) \\
&\qquad \sum_{\tilde{v}^n \in T^n_{P_{V|U^\varepsilon}}(u^n): w(\tilde{v}^{*n} - v^{*n}) = w} \mathbf{1}\left[v^{*n} \boldsymbol{A} = c^l\right] \mathbf{1}\left[(\tilde{v}^{*n} - v^{*n}) \boldsymbol{A} = 0^l\right] \\
&\overset{\text{(e)}}{\leq} \frac{(1 + \delta_n(R'))\beta_n(R') |T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l} + (1 + \delta_n(R'))^2 \left(\frac{|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l} \right)^2, \qquad (4.22)
\end{aligned}
$$

where (d) comes from Lemma 2.3, and (e) comes from Lemma 2.6 and (4.25).

Note that

$$\frac{|T^n_{P_{V|U}\varepsilon}(u^n)|}{q^l} \geq q^{n\left(H(V|U)-\frac{l}{n}-\delta(\varepsilon)\right)}, \tag{4.23}$$

where (4.23) comes from 3) of Lemma 2.7, and the left hand side of (4.23) approaches ∞ ($n \to \infty$). This comes from the assumption $\frac{l}{n} < H(V|U) - \delta$ by taking ε that satisfies

$$\varepsilon \log \frac{1}{\varepsilon} + 3\varepsilon < \frac{\delta}{2}, \tag{4.24}$$

for sufficiently large n to hold

$$\frac{q^2 \log(n+1)}{n} < \frac{\delta}{2}, \tag{4.25}$$

where $\delta(\varepsilon) \stackrel{\text{def}}{=} \varepsilon \log \frac{1}{\varepsilon} + 3\varepsilon + \frac{q^2 \log(n+1)}{n}$.

By substituting (4.22) into (4.21) and noting that

$$\delta_n(R') \to 0 \ (n \to \infty), \tag{4.26}$$

which comes from 2) of Lemma 2.5, it can be shown that

$$
\begin{aligned}
&\frac{\mathrm{E}_{\boldsymbol{A}}\left|\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\left(\frac{\alpha_n(R';w(v^n))}{q^l}-\mathbf{1}\left[v^n\boldsymbol{A}=c^l\right]\right)\right|^2}{\left|\frac{(1-\delta_n(R'))|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right|^2}\\
&\le \frac{\sum_{v^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\left\{\frac{\alpha_n(R';w(v^n))}{q^l}-\left(\frac{\alpha_n(R';w(v^n))}{q^l}\right)^2\right\}}{\left|\frac{(1-\delta_n(R'))|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right|^2}\\
&\quad-\frac{\sum_{v^n\neq\tilde{v}^n\in T^n_{P_{V|U^\varepsilon}}(u^n)}\frac{\alpha_n(R';w(v^n))}{q^l}\frac{\alpha_n(R';w(\tilde{v}^n))}{q^l}}{\left|\frac{(1-\delta_n(R'))|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right|^2}\\
&\quad+\frac{\frac{(1+\delta_n(R'))\beta_n(R')|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}}{\left|\frac{(1-\delta_n(R'))|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right|^2}+\frac{(1+\delta_n(R'))^2\left(\frac{|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right)^2}{\left|\frac{(1-\delta_n(R'))|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right|^2}\\
&\overset{\text{(f)}}{\le}\frac{1+\delta_n(R')}{(1-\delta_n(R'))^2}\left(\frac{|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right)^{-1}-1\\
&\quad+\frac{1+\delta_n(R')}{(1-\delta_n(R'))^2}\beta_n(R')\left(\frac{|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right)^{-1}+\frac{(1+\delta_n(R'))^2}{(1-\delta_n(R'))^2}\\
&=\frac{1+\delta_n(R')}{(1-\delta_n(R'))^2}\left(\frac{|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right)^{-1}\\
&\quad+\frac{1+\delta_n(R')}{(1-\delta_n(R'))^2}\beta_n(R')\left(\frac{|T^n_{P_{V|U^\varepsilon}}(u^n)|}{q^l}\right)^{-1}+\frac{4\delta_n(R')}{(1-\delta_n(R'))^2}\\
&\to 0\ (n\to\infty),
\end{aligned}
\tag{4.27}
$$

where at (f), we use 2) of Lemma 2.5. From the above arguments, (4.18) holds.

[End of Proof of Lemma 4.1*]*

Remark 4.4
In [35], c^l is taken as a random variable. Here, with the help of Lemma 2.6, c^l can be taken as a fixed vector.

Before analyzing the second term of (4.17), we show a lemma that has a dual meaning with Lemma 4.1.

Lemma 4.2
For any $u^n \in T^n_{P_{U^\epsilon}}$ and for any $\delta > 0$, let

$$\tilde{G}_{\boldsymbol{A}}(c^l) \stackrel{\text{def}}{=} \left\{ v^n \in [0:q-1]^n \;\middle|\; v^n \boldsymbol{A} = c^l \text{ and } H(W|P_{u^n}) < \frac{l}{n} - \delta \right\}, \quad (4.28)$$

where W appearing in the above definition is a conditional type of v^n conditioned on u^n.

Then it holds that

$$\lim_{n\to\infty} P_{\boldsymbol{A}} \left[\tilde{G}_{\boldsymbol{A}}(c^l) \neq \emptyset \right] = 0. \quad (4.29)$$

■

[Proof of Lemma 4.2]

$$\begin{aligned}
&P_{\boldsymbol{A}} \left[\tilde{G}_{\boldsymbol{A}}(c^l) \neq \emptyset \right] \\
&= \mathrm{E}_{\boldsymbol{A}} \mathbf{1} \left[\sum_{v^n \in [0:q-1]^n} \mathbf{1}\left[v^n \boldsymbol{A} = c^l \right] \mathbf{1} \left[v^n \in \bigcup_{W: H(W|P_{u^n}) < \frac{l}{n} - \delta} T^n_W(u^n) \right] \geq 1 \right] \\
&\stackrel{\text{(a)}}{\leq} \mathrm{E}_{\boldsymbol{A}} \sum_{v^n \in [0:q-1]^n} \mathbf{1}\left[v^n \boldsymbol{A} = c^l \right] \mathbf{1} \left[v^n \in \bigcup_{W: H(W|P_{u^n}) < \frac{l}{n} - \delta} T^n_W(u^n) \right], \quad (4.30)
\end{aligned}$$

where (a) comes from Markov's inequality.

Since the assumption is $c^l \neq 0^l$, we obtain

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}} \sum_{v^n \in [0:q-1]^n} \mathbf{1}\left[v^n \boldsymbol{A} = c^l\right] \mathbf{1}\left[v^n \in \bigcup_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} T_W^n(u^n)\right] \\
&\overset{\text{(b)}}{=} \mathrm{E}_{\boldsymbol{A}} \left(\sum_{w=1}^{n\gamma_{R'}} + \sum_{w=n\gamma_{R'}+1}^{n}\right) \\
&\qquad \cdot \sum_{v^n : w(v^{*n})=w} \mathbf{1}\left[v^{*n}\boldsymbol{A} = c^l\right] \mathbf{1}\left[v^n \in \bigcup_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} T_W^n(u^n)\right] \\
&\leq \sum_{w=1}^{n\gamma_{R'}} \sum_{v^n : w(v^{*n})=w} \mathrm{E}_{\boldsymbol{A}} \mathbf{1}\left[v^{*n}\boldsymbol{A} = c^l\right] \\
&\qquad + \sum_{w=n\gamma_{R'}+1}^{n} \mathrm{E}_{\boldsymbol{A}} \sum_{v^n : w(v^{*n})=w} \mathbf{1}\left[v^{*n}\boldsymbol{A} = c^l\right] \mathbf{1}\left[v^n \in \bigcup_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} T_W^n(u^n)\right] \\
&\overset{\text{(c)}}{\leq} \beta_n(R') + (1+\delta_n(R')) \sum_{v^n \in [0:q-1]^n} \frac{\mathbf{1}\left[v^n \in \bigcup_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} T_W^n(u^n)\right]}{q^l} \\
&\overset{\text{(d)}}{\leq} \beta_n(R') + (1+\delta_n(R')) \sum_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} \frac{|T_W^n(u^n)|}{q^l} \\
&\leq \beta_n(R') + (1+\delta_n(R')) \sum_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} \frac{q^{nH(W|P_{u^n})}}{q^l} \\
&\leq \beta_n(R') + (1+\delta_n(R')) \sum_{W:H(W|P_{u^n})<\frac{l}{n}-\delta} \frac{q^{n(\frac{l}{n}-\delta)}}{q^l} \\
&\leq \beta_n(R') + (1+\delta_n(R'))(n+1)^{q^2} q^{-n\delta}, \qquad (4.31)
\end{aligned}
$$

where (b) is from Lemma 2.3, (c) comes from Lemma 2.5 with Remark 2.4, and (d) is derived using ordinary type techniques [7].

By substituting (4.31) into (4.30), Lemma 4.2 is proved.

[End of Proof of Lemma 4.2]

Remark 4.5

The duality between Lemma 4.1 and Lemma 4.2 is remarkable. Lemma 4.1

shows that if $\frac{l}{n} < H(V|P_{u^n}) - \delta$, *then* $\left\{v^n \mid v^n \boldsymbol{A} = c^l\right\} \bigcap T^n_{V\epsilon}(u^n) \neq \emptyset$ *holds with high probability. On the other hand, Lemma* 4.2 *shows that if* $\frac{l}{n} > H(V|P_{u^n}) + \delta$, *then it holds that* $\left\{v^n \mid v^n \boldsymbol{A} = c^l\right\} \bigcap T^n_{V\epsilon}(u^n) = \emptyset$ *with high probability. These similar lemmas will be used for the proof of the channel coding theorem in Chapter* 5.

Definition 4.1 *[Variational Distance] For given probability distributions* P *and* Q *on a set* $\mathcal{U}$, *the variational distance between* P *and* Q *is defined as*

$$||P - Q|| \stackrel{\text{def}}{=} \sum_{a \in \mathcal{U}} |P(a) - Q(a)|. \tag{4.32}$$

Lemma 4.3
For given probability distributions P *and* Q *on a set* $\mathcal{U}$, *if the following two conditions hold,*
1) $||P - Q|| \leq \varepsilon$,
and
2) when $P(a) = 0$, *then* $Q(a) = 0$,
then it holds that

$$|P(a) - Q(a)| \leq \frac{\varepsilon}{\min_{b:P(b)>0} P(b)} P(a) \text{ for any } a \in \mathcal{U}.$$

■

[Proof of Lemma 4.3*]*
From the above assumptions,

$$\begin{aligned}
\varepsilon \geq ||P - Q|| &= \sum_a |P(a) - Q(a)| \\
&\geq \sum_{a:P(a)>0} \frac{\min_{b:P(b)>0} P(b)}{P(a)} |P(a) - Q(a)| \\
&\geq \frac{\min_{b:P(b)>0} P(b)}{P(a)} |P(a) - Q(a)|.
\end{aligned} \tag{4.33}$$

[End of Proof of Lemma 4.3*]*

To prove

$$\mathrm{E}_{\boldsymbol{AB}} P^n_U \left[\mathcal{E}_{22} \bigcap \mathcal{E}^c_{21} \bigcap \mathcal{E}^c_1\right] \to 0 \ (n \to \infty), \tag{4.34}$$

it is sufficient to show that for any $u^n \in T^n_{P_{U\varepsilon}}$, the output of "vector quantization" part of the encoder, $\hat{v}^n$, is jointly typical with u^n with high probability, which means that

$$||P_{u^n\hat{v}^n} - P_{UV}|| < \delta'(\varepsilon) \text{ for } \exists\delta'(\varepsilon) \text{ that satisfies } \delta'(\varepsilon) \to 0 \ (\varepsilon \to 0) \tag{4.35}$$

with high probability, where

$$\hat{v}^n \stackrel{\text{def}}{=} \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} P^n_{V|U}(\tilde{v}^n|u^n), \tag{4.36}$$

by using Lemma 4.3.

Assume that $u^n \in T^n_{P_{U\epsilon}}$. Note that

$$\begin{aligned}
&\arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} P_{V|U}(\tilde{v}^n|u^n) = \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} \log P_{V|U}(\tilde{v}^n|u^n) \\
&= \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} \sum_{i=1}^{n} \log P_{V|U}(\tilde{v}_i|u_i) \\
&= \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} n \sum_{a,b} P_{u^n}(a) W(b|a) \log P_{V|U}(b|a) \\
&= \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} \left\{ \sum_{a,b} P_{u^n}(a) W(b|a) \log \frac{P_{V|U}(b|a)}{W(b|a)} \right. \\
&\qquad\qquad \left. - \sum_{a,b} P_{u^n}(a) W(b|a) \log \frac{1}{W(b|a)} \right\} \\
&= \arg \min_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} \left\{ D(W||P_{V|U}|P_{u^n}) + H(W|P_{u^n}) \right\},
\end{aligned} \tag{4.37}$$

where W is a conditional type of the argument $\tilde{v}^n$ conditioned on u^n in the above optimization. Let $\hat{W}$ be the corresponding conditional type of $\hat{v}^n$ attaining the minimum of the right hand side in (4.37). From Lemma 4.1, there exists a v^n satisfying $v^n \boldsymbol{A} = c^l$ and $v^n \in T^n_{P_{V|U^\varepsilon}}(u^n)$ with high

probability. With this v^n, since $||P_{u^n v^n} - P_{UV}|| < 2\varepsilon$, we obtain

$$\begin{aligned}
& D(\hat{W}||P_{V|U}|P_{u^n}) + H(\hat{W}|P_{u^n}) \\
& \stackrel{\text{(a)}}{\leq} -\frac{1}{n}\log P_{V^n|U^n}(v^n|u^n) \\
& = \sum_{a,b} P_{u^n v^n}(a,b)\log\frac{1}{P_{V|U}(b|a)} \\
& \leq \sum_{a,b} |P_{u^n v^n}(a,b) - P_{UV}(a,b)|\log\frac{1}{P_{V|U}(b|a)} \\
& \quad + \sum_{a,b} P_{UV}(a,b)\log\frac{1}{P_{V|U}(b|a)} \\
& \stackrel{\text{(b)}}{\leq} 2\eta\varepsilon + H(P_{V|U}|P_U), \qquad (4.38)
\end{aligned}$$

where at (a), the definition of $\hat{v}^n$, and at (b), $\eta \stackrel{\text{def}}{=} \max_{a,b:P_{V|U}(b|a)>0} \log\frac{1}{P_{V|U}(b|a)}$ and the fact that $||P_{u^n v^n} - P_{UV}|| < 2\varepsilon$ are used, respectively.

If $D(\hat{W}||P_{V|U}|P_{u^n}) > (2\eta + 10\log\frac{1}{\varepsilon})\varepsilon$, then

$$H(\hat{W}|P_{u^n}) \leq H(P_{V|U}|P_U) - 10\varepsilon\log\frac{1}{\varepsilon} \qquad (4.39)$$

is derived. If

$$10\varepsilon\log\frac{1}{\varepsilon} > \delta, \qquad (4.40)$$

then there exists $\frac{l}{n}$ satisfying both $\frac{l}{n} < H(V|U) - \delta$ and $H(\hat{W}|P_{u^n}) < \frac{l}{n} - (10\varepsilon\log\frac{1}{\varepsilon} - \delta)$. Therefore, from Lemma 4.2, the probability that $\hat{v}^n$ is included in $G_{\boldsymbol{A}}(c^l)$ approaches 0 for sufficiently large n.

Considering the above fact, it is sufficient to investigate only the case of $D(\hat{W}||P_{V|U}|P_{u^n}) \leq (2\eta + 10\log\frac{1}{\varepsilon})\varepsilon$. Then we have

$$\begin{aligned}
(2\eta + 10\log\frac{1}{\varepsilon})\varepsilon & \geq D(\hat{W}||P_{V|U}|P_{u^n}) \\
& \stackrel{\text{(c)}}{\geq} \frac{1}{2\ln 2}\sum_a P_{u^n}(a)||\hat{W}(*|a) - P_{V|U}(*|a)||^2 \\
& \stackrel{\text{(d)}}{\geq} \frac{1}{2\ln 2}\left\{\sum_a P_{u^n}(a)||\hat{W}(*|a) - P_{V|U}(*|a)||\right\}^2. \qquad (4.41)
\end{aligned}$$

At (c), we use the formula $D(P||Q) \geq \frac{1}{2\ln 2}||P-Q||^2$ [9, Lemma 11.6.1]. At (d), Jenzen's inequality [9, Theorem 2.6.2] is applied.

As a result, we obtain

$$\sum_a P_{u^n}(a)||\hat{W}(*|a) - P_{V|U}(*|a)|| < \sqrt{4(\eta + 5\log\frac{1}{\varepsilon})\varepsilon \ln 2}, \tag{4.42}$$

which means

$$\begin{aligned}
&\sqrt{4(\eta + 5\log\frac{1}{\varepsilon})\varepsilon \ln 2} \\
&> \sum_a P_{u^n}(a)||\hat{W}(*|a) - P_{V|U}(*|a)|| \\
&= \sum_{a,b} \left| P_{u^n}(a)\hat{W}(b|a) - P_{u^n}(a)P_{V|U}(b|a) \right| \\
&\geq ||P_{u^n}\hat{W} - P_U P_{V|U}|| \\
&\qquad - \sum_{a,b} \left| P_U(a)P_{V|U}(b|a) - P_{u^n}(a)P_{V|U}(b|a) \right| \\
&= ||P_{u^n}\hat{W} - P_{UV}|| - \sum_a |P_U(a) - P_{u^n}(a)| \\
&\geq ||P_{u^n}\hat{W} - P_{UV}|| - \varepsilon = ||P_{u^n\hat{v}^n} - P_{UV}|| - \varepsilon.
\end{aligned} \tag{4.43}$$

From Lemma 4.3, when we set $\delta'(\varepsilon) \stackrel{\rm def}{=} \frac{\varepsilon + \sqrt{4(\eta+5\log\frac{1}{\varepsilon})\varepsilon\ln 2}}{\min_{a,b:P_{UV}(a,b)>0} P_{UV}(a,b)}$ in (4.8), (4.35) is proved. Therefore, we obtain (4.34).

4.3.2 Evaluation of $\mathbf{E}_{\boldsymbol{AB}} P_U^n [\mathcal{E}_3 \bigcap \mathcal{E}_2^c \bigcap \mathcal{E}_1^c]$

It is sufficient to show that if $\frac{l+k}{n} > H(V) + \delta^{1/3}$ for a positive number $\delta > 0$, then for any $u^n \in T^n_{P_U\varepsilon}$ we obtain

$$\lim_{n\to\infty} P_{\boldsymbol{AB}} \left[\hat{v}^n \neq \arg \max_{\substack{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l \\ \tilde{v}^n \boldsymbol{B} = m^k}} P_V(\tilde{v}^n) \right] = 0, \tag{4.44}$$

where $\hat{v}^n = \arg\max_{v^n : v^n \boldsymbol{A} = c^l} P_{V|U}(v^n|u^n)$ and $m^k = \hat{v}^n \boldsymbol{B}$.

From the results in Section 4.3.1, it can be assumed that $\hat{v}^n \in T^n_{P_V\delta'(\varepsilon)}$. Then it holds that

$$
\begin{aligned}
&P_{\boldsymbol{AB}}\left[\hat{v}^n \neq \arg\max_{\substack{\tilde{v}^n:\tilde{v}^n\boldsymbol{A}=c^l\\ \tilde{v}^n\boldsymbol{B}=m^k}} P_{V^n}(\tilde{v}^n)\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\exists\tilde{v}^n \neq \hat{v}^n \text{ s.t. } P_{V^n}(\tilde{v}^n) \geq P_{V^n}(\hat{v}^n), \tilde{v}^n\boldsymbol{A} = c^l, \tilde{v}^n\boldsymbol{B} = m^k\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\sum_{\tilde{v}^n\neq\hat{v}^n}\mathbf{1}\left[P_{V^n}(\tilde{v}^n) \geq P_{V^n}(\hat{v}^n), \tilde{v}^n\boldsymbol{A} = c^l, \tilde{v}^n\boldsymbol{B} = m^k\right] \\
&\overset{\text{(a)}}{\leq} \mathrm{E}_{\boldsymbol{AB}}\sum_{\tilde{v}^n\neq\hat{v}^n}\mathbf{1}\left[P_{V^n}(\tilde{v}^n) \geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right] \\
&\qquad\qquad \cdot\mathbf{1}\left[\tilde{v}^n\boldsymbol{A} = c^l\right]\mathbf{1}\left[(\tilde{v}^n - \hat{v}^n)\,\boldsymbol{B} = 0^k\right] \\
&= \mathrm{E}_{\boldsymbol{AB}}\left(\sum_{w=1}^{n\gamma_R} + \sum_{w=n\gamma_R+1}^{n}\right)\sum_{\tilde{v}^n:w(\tilde{v}^{*n}-\hat{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n) \geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right] \\
&\qquad\qquad \cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A} = c^l\right]\mathbf{1}\left[(\tilde{v}^{*n} - \hat{v}^{*n})\,\boldsymbol{B} = 0^k\right] \\
&\leq \mathrm{E}_{\boldsymbol{B}}\sum_{w=1}^{n\gamma_R}\sum_{\tilde{v}^n:w(\tilde{v}^{*n}-\hat{v}^{*n})=w}\mathbf{1}\left[(\tilde{v}^{*n} - \hat{v}^{*n})\,\boldsymbol{B} = 0^k\right] \\
&\quad + \mathrm{E}_{\boldsymbol{AB}}\sum_{w=n\gamma_R+1}^{n}\sum_{\tilde{v}^n:w(\tilde{v}^{*n}-\hat{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n) \geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right] \\
&\qquad\qquad \cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A} = c^l\right]\mathbf{1}\left[(\tilde{v}^{*n} - \hat{v}^{*n})\,\boldsymbol{B} = 0^k\right], \qquad (4.45)
\end{aligned}
$$

where at (a) $\hat{v}^n \in T^n_{P_V\delta'(\varepsilon)}$ and $m^k = \hat{v}^n\boldsymbol{B}$ are used.

By 1) of Lemma 2.5, the first term of (4.45) is upper bounded by $\beta_n(R')$. On the other hand, the second term of (4.45) is evaluated as follows:

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}}\sum_{w=n\gamma_R+1}^{n}\sum_{\tilde{v}^n:w(\tilde{v}^{*n}-\hat{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n) \geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right] \\
&\qquad\qquad \cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A} = c^l\right]\mathbf{1}\left[(\tilde{v}^{*n} - \hat{v}^{*n})\,\boldsymbol{B} = 0^k\right] \\
&\overset{\text{(b)}}{\leq} \mathrm{E}_{\boldsymbol{A}}\sum_{w=n\gamma_R+1}^{n}\sum_{\tilde{v}^n:w(\tilde{v}^{*n}-\hat{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n) \geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right] \\
&\qquad\qquad \cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A} = c^l\right]\frac{1+\delta_n(R)}{q^k}, \qquad (4.46)
\end{aligned}
$$

where at (b), 2) of Lemma 2.5 is applied to random variable $\boldsymbol{B}$. Note that in (4.46), since both $\hat{v}^{*n}$ in summation and $\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]$ depend on $\boldsymbol{A}$, we cannot directly apply 2) of Lemma 2.5 to $\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]$. To apply Lemma 2.5, we have to decorrelate these variables as follows:

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{A}}\sum_{w=n\gamma_R+1}^{n}\sum_{\tilde{v}^n:w(\tilde{v}^{*n}-\hat{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n)\geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right]\\
&\qquad\cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]\frac{1+\delta_n(R)}{q^k}\\
&\leq \mathrm{E}_{\boldsymbol{A}}\sum_{\tilde{v}^n}\mathbf{1}\left[P_V(\tilde{v}^n)\geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right]\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]\frac{1+\delta_n(R)}{q^k}\\
&=\mathrm{E}_{\boldsymbol{A}}\left(\sum_{w=1}^{n\gamma_{R'}}+\sum_{w=n\gamma_{R'}+1}^{n}\right)\sum_{\tilde{v}^n:w(\tilde{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n)\geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right]\\
&\qquad\cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]\frac{1+\delta_n(R)}{q^k}\\
&\leq \mathrm{E}_{\boldsymbol{A}}\sum_{w=1}^{n\gamma_{R'}}\sum_{\tilde{v}^n:w(\tilde{v}^{*n})=w}\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]\frac{1+\delta_n(R)}{q^k}\\
&\quad+\mathrm{E}_{\boldsymbol{A}}\sum_{w=n\gamma_{R'}+1}^{n}\sum_{\tilde{v}^n:w(\tilde{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n)\geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right]\\
&\qquad\cdot\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A}=c^l\right]\frac{1+\delta_n(R)}{q^k}\\
&\overset{\text{(c)}}{\leq}\frac{(1+\delta_n(R))\beta_n(R')}{q^k}\\
&\quad+\mathrm{E}_{\boldsymbol{A}}\sum_{w=n\gamma_{R'}+1}^{n}\sum_{\tilde{v}^n:w(\tilde{v}^{*n})=w}\mathbf{1}\left[P_V(\tilde{v}^n)\geq q^{-n(H(V)+\delta(\delta'(\varepsilon)))}\right]\\
&\qquad\cdot\frac{(1+\delta_n(R))(1+\delta_n(R'))}{q^{l+k}}\\
&\overset{\text{(d)}}{\leq}\frac{(1+\delta_n(R))\beta_n(R')}{q^k}+\frac{(1+\delta_n(R))(1+\delta_n(R'))q^{n(H(V)+\delta(\delta'(\varepsilon)))}}{q^{l+k}}.
\end{aligned}
\tag{4.47}
$$

At (c), Lemma 2.5 with Remark 2.4 is used, and at (d), ordinary techniques of types (e.g., [7]) are used.

With the above argument, if we take ε and n, l, k satisfying

$$\delta'(\varepsilon) \log \frac{1}{\delta'(\varepsilon)} + 3\delta'(\varepsilon) < \frac{\delta^{1/3}}{2} \tag{4.48}$$

and

$$\frac{q^2 \log(n+1)}{n} < \frac{\delta^{1/3}}{2} \tag{4.49}$$

and $\frac{l+k}{n} > H(V) + \delta^{1/3}$, then it is shown that the second term of (4.45) approaches 0 with $n \to \infty$.

Note that with the results in Sections 4.3.1 and 4.3.2, by setting $\hat{\delta}(\varepsilon) \stackrel{\text{def}}{=} \varepsilon + \delta'(\varepsilon)(1+\varepsilon)$, we can prove Theorem 4.1 through the evaluation of (4.8). ∎

Remark 4.6
δ and ε must satisfy the following relationships: From (4.40)*,*

$$10\varepsilon \log \frac{1}{\varepsilon} > \delta, \tag{4.50}$$

from (4.24)*,*

$$8\varepsilon \log \frac{1}{\varepsilon} < \delta, \tag{4.51}$$

and from (4.48)*,*

$$8\delta'(\varepsilon) \log \frac{1}{\delta'(\varepsilon)} < \delta^{1/3}. \tag{4.52}$$

By the straight forward calculation, if

$$\varepsilon \left(\log \frac{1}{\varepsilon} \right)^3 \frac{(160 \ln 2)^3}{(\min_{a,b:P_{UV}(a,b)>0} P_{UV}(a,b))^6} < 1 \tag{4.53}$$

is satisfied, which is a sufficient condition of $\delta'(\varepsilon)^3 < \varepsilon$, it can be shown that there exists δ satisfying (4.50)*,* (4.51)*, and* (4.52)*.*

Remark 4.7
Note that the statement of Theorem 4.1 *can be replaced by*

$$\lim_{n\to\infty} \mathrm{E}_{\boldsymbol{AB}} P_{U^n}\left[\frac{d_n\left(U^n, \psi_n\left(\varphi_n(U^n)\right)\right)}{n} > D\right] = 0. \tag{4.54}$$

Using the above fact and Markov's inequality, it is shown that we can get sparse matrices $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, which have desired properties, with high probability as follows: for any positive number* $\varepsilon > 0$*,*

$$\begin{aligned} &P_{\boldsymbol{AB}}\left[P_{U^n}\left[\frac{d_n\left(U^n, \psi_n\left(\varphi_n(U^n)\right)\right)}{n} > D\right] > \varepsilon\right] \\ &\le \frac{\mathrm{E}_{\boldsymbol{AB}} P_{U^n}\left[\frac{d_n(U^n, \psi_n(\varphi_n(U^n)))}{n} > D\right]}{\varepsilon} \to 0 \ (n \to \infty) \end{aligned} \tag{4.55}$$

4.4 Simulation Experiments

Linear code linear programming (LCLP) proposed by Feldman [12] is a promising method for implementing the lossy source coding process efficiently. In the encoding and decoding of a lossy source code constructed by sparse matrices, we propose auxiliary methods to obtain approximate optimal values in the case where the output of LCLP takes non-integer values. In a simulation, the proposed method [32] attains smaller average distortions than the time-sharing bound.

4.4.1 LCLP Algorithm and Its Property

In this subsection, a binary alphabet ($q = 2$) is assumed, and we show how the LCLP algorithm can be used to implement the vector-quantization part of the encoding process:

$$\max_{v^n : v^n \boldsymbol{A} = c^l} P_{V^n|U^n}(v^n|u^n). \tag{4.56}$$

Note that the LCLP algorithm can also be applied to the decoding process in the same way as to the encoding process.

Let $\boldsymbol{A} = (a_{ij})_{n\in[1:n], j\in[1:l]}$ and $c^l = (c_j)_{j\in[1:l]}$. The objective function to be minimized is

$$\sum_{i=1}^{n} v_i \log \frac{P_{V|U}(0|u_i)}{P_{V|U}(1|u_i)} \to \min, \tag{4.57}$$

where $(v_i)_{i\in[1:n]}$ are variables to be determined, by noting that

$$\begin{aligned}
&\arg\max_{v^n:v^n\boldsymbol{A}=c^l} P_{V^n|U^n}(v^n|u^n) = \arg\max_{v^n:v^n\boldsymbol{A}=c^l} \log P_{V^n|U^n}(v^n|u^n)\\
&= \arg\max_{v^n:v^n\boldsymbol{A}=c^l} \sum_{i=1}^{n} \log P_{V|U}(v_i|u_i)\\
&= \arg\max_{v^n:v^n\boldsymbol{A}=c^l} \sum_{i=1}^{n} \left\{(1-v_i)\log P_{V|U}(0|u_i) + v_i \log P_{V|U}(1|u_i)\right\}\\
&= \arg\min_{v^n:v^n\boldsymbol{A}=c^l} \sum_{i=1}^{n} v_i \log \frac{P_{V|U}(0|u_i)}{P_{V|U}(1|u_i)}. \qquad (4.58)
\end{aligned}$$

For $j \in [1:l]$, let

$$\mathcal{N}(j) \stackrel{\text{def}}{=} \left\{i \in [1:n] \;\middle|\; a_{ij} = 1\right\} \qquad (4.59)$$

and

$$\mathcal{E}_j \stackrel{\text{def}}{=} \left\{S \subseteq \mathcal{N}(j) \;\middle|\; |S| + c_j = 1 \pmod 2\right\}, \qquad (4.60)$$

where $|A|$ denotes the cardinality of set A. Then the "relaxed" linear constraints associated with the original linear constraints $v^n\boldsymbol{A} = c^l$ are as follows:

$$0 \le v_i \le 1 \text{ for } i \in [1:n], \qquad (4.61)$$

and for any $j \in [1:l]$ and $S \subseteq \mathcal{E}_j$,

$$\sum_{i\in S}(1-v_i) + \sum_{i\in\mathcal{N}(j)\setminus S} v_i \ge 1. \qquad (4.62)$$

Feldman [12] showed that when LCLP is applied to channel coding, the output becomes the maximum likelihood estimator if all of the output values are integers. This property is called Maximum Likelihood Certificate Property (MLCP). When LCLP is applied to source coding problems, while row vector c^l of the parity check condition $v^n\boldsymbol{A} = c^l$ is a fixed vector whose Hamming weight is even, the algorithm described here can also be shown to have MLCP using similar arguments to [12]. The next proposition summarizes this fact.

Proposition[MLCP]

The LCLP algorithm described here has the MLCP.

4.4.2 Auxiliary Methods

As mentioned above, LCLP does not always output integers. In an ordinary process, when non-integer output is produced, the output must be treated as an error. In this situation, the LCLP process is said to have failed.

In channel coding with the sum-product decoding algorithm, some trials have been investigated to improve the case of failures in decoding. Fossorier, Lin and Snyders [13] proposed an approximation process to improve the convergence property of the sum-product algorithm in channel coding.

In this subsection, we propose auxiliary methods that enable LCLP to be used for lossy source coding problems.

When the LCLP outputs v^n with non-integer values, we scalar-quantize v_i to 0 or 1 by a given rule if it corresponds to a redundant row of the parity-check matrix. After this scalar-quantization, we solve linear equations $v^n \boldsymbol{A} = c^l$ to determine the remaining part of v^n. Then, we obtain the output $\tilde{v}^n$ that consists of integers and satisfies the parity-check condition. The block diagram of this auxiliary algorithm (AA) is shown in Figure 4.4, and the details of AA are given in the following part of this subsection.

Although the AA can output a solution with all integers, the output cannot always be a near-optimal solution. Hence, we propose another auxiliary method called the 2nd optimization to attain a smaller distortion using the output of AA, which is described following the description of AA.

Auxiliary Algorithm

The AA consists of three steps: 1. sweep-out, 2. scalar-quantization, 3. solving of the linear equation. In the second step, scalar-quantization is applied to $\tilde{v}_i$'s, which correspond to redundant parts of a solution in the indeterminate equation $v^n \boldsymbol{A} = c^l$. In the third step, the remaining parts of the solution are determined by solving l linear equations with l variables. The details of each step are given as follows. For simplicity, the rank of $\boldsymbol{A}$ is assumed to be l.

[1. Sweep-out step]

Under the above assumption, by application of the sweep-out row operation to $\boldsymbol{A}$, the transformed matrix of $\boldsymbol{A}$ contains l unit row vectors $(1, 0, .., 0), ..., (0, .., 0, 1)$. After elimination of l unit vectors from the transformed matrix, the set of residual row indices is denoted as *IndX*. An example of the process is shown in Figure 4.4.

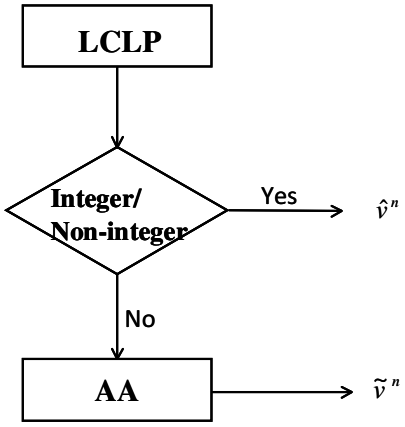

Figure 4.4: LCLP and AA

[2. Scalar-quantization step]

By use of a scalar-quantizer $Q : [0,1] \to \{0,1\}$, each value of $\tilde{v}_i (i \in IndX)$ is determined. Let LCLP output be $v^n \in [0,1]^n$, then, $\tilde{v}_i = Q(y_i)$ for $i \in IndX$. For example, we set

$$Q(v_i) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } v_i \leq 0.5 \\ 1, & \text{otherwise,} \end{cases} \tag{4.63}$$

which is used in our simulation.

[3. Solving of the linear equation step]

By solving of l linear equations with l variables, each value of $\tilde{v}_i (i \in [1:n] \backslash IndX)$ is determined.

2nd Optimization

While the output of AA, $\tilde{v}^n$, satisfies the parity-check condition $\tilde{v}^n \boldsymbol{A} = c^l$, the value of the objective function derived from the output is not always near the optimum. In this subsection, we propose an algorithm called the 2nd Optimization (2nd Opt) whose output satisfies the parity-check condition and, at the same time, can be a value closer to the optimum. The 2nd Opt carries out the procedures of LCLP and AA by fixing v_i in (4.64) for each $i \in [1:n]$ (Figure 4.6).

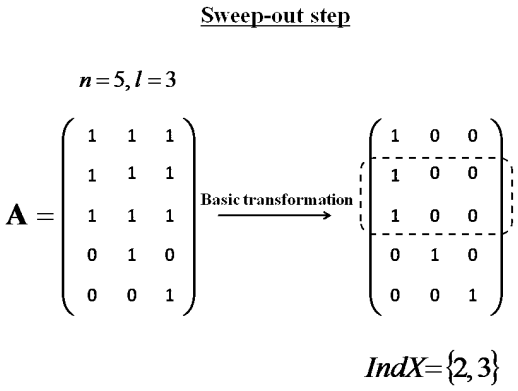

Figure 4.5: Sweep-out step

The algorithm is as follows.
Let $\tilde{v}^n$ be the output of AA. For $i \in [1:n]$,

$$\arg \max_{\substack{v^n : v^n \boldsymbol{A} = c^l \\ v_i = \tilde{v}_i}} P_{V^n|U^n}(v^n|u^n) \tag{4.64}$$

is carried out by LCLP and AA, and let $z^{(i)n}$ be the output of the i-th loop of LCLP and AA. The final output of 2nd Opt is obtained by

$$\arg \max_{z^{(i)n}:1\le i\le n} P_{V^n|U^n}(z^{(i)n}|u^n). \tag{4.65}$$

Remark 4.8
Let T_n be a computational complexity of LCLP, where n denotes block length. Since the computational complexity of the sweep out process of a matrix is $O(n^3)$, that of AA is estimated by $O(n^3)$, and that of LCLP and AA by $T_n + O(n^3)$. Therefore, the computational complexity of 2nd Opt, which repeats LCLP and AA n times, is $n \times (T_n + O(n^3))$. This complexity is apparently higher than that of the sum-product algorithm whose complexity is estimated by $O(n)$.

Hence, our next goal in future work is to contrive an efficient algorithm with low computational complexity, which can output a near-optimal v^n satisfying a given parity-check condition.

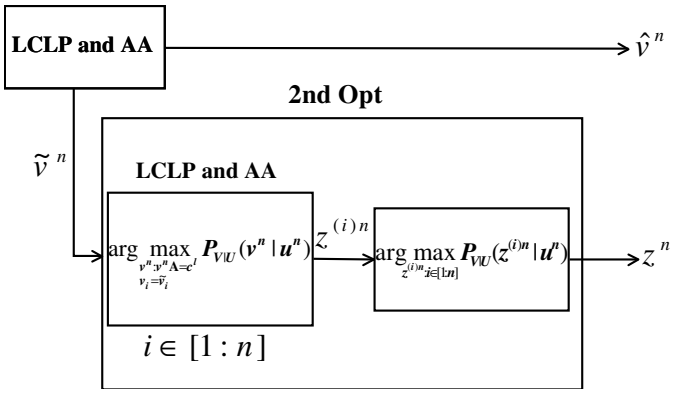

Figure 4.6: LCLP, AA, and 2nd Opt

4.4.3 Simulation Results

In this subsection, we first denote the conditions of simulation, and then we show some simulation results.

[Conditions]

The probability distribution of the source is set to be uniform, s.t. $P_U(U = 0) = P_U(U = 1) = 0.5$, and the distortion measure is the Hamming measure. Block length n is fixed to be 100. An $n \times l$ sparse matrix $\boldsymbol{A}$ and an $n \times k$ sparse matrix $\boldsymbol{B}$ are generated using the bipartite graph method (Figure 4.7; see e.g. [24]) with parameters l and k shown in Table 1. Note that the values of l and k in Table 1 satisfy the conditions stated in Theorem 4.1. For a given compression rate, each simulation consists of generation of sparse matrices, 100 samplings of the source message with block length $n = 100$, and lossy encoding and decoding using LCLP, AA, and 2nd Opt.

[Simulation Results]

Average distortions calculated by 100 encodings and decodings for each compression rate are shown in Figure 4.8. The horizontal and vertical axes denote compression rate and average distortion, respectively.

The distortion-rate curve and time-sharing bound are denoted by the solid curve and dotted line, respectively. In the figure, "LCLP-success" means that no auxiliary methods (AA, 2nd Opt) were used in the vector-quantization process, and the average distortion was calculated only for the case that the LCLP succeeds in encoding to generate $\hat{v}^n$ with all integer components. "LCLP and AA" and "LCLP and AA and 2nd Opt" mean that the corre-

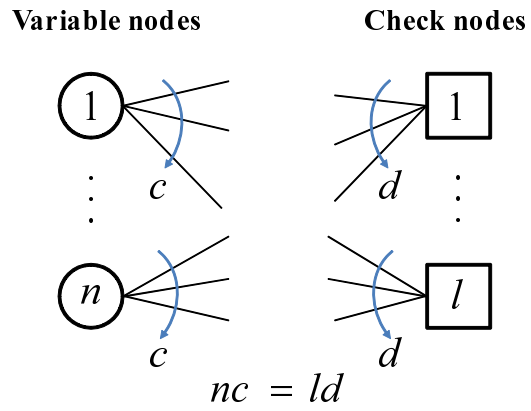

Figure 4.7: Bipartite graph method for generating sparse matrices

sponding auxiliary methods (AA only or AA and 2nd Opt) were used in the process. At decoding, "LCLP and AA" was applied to all cases. The figure shows that "LCLP and AA and 2nd Opt" attained the smallest average distortions over the entire range of coding rate. Note that the 2nd Opt was not used at the rate of 0.8 since LCLP almost always succeeds for the vector-quantization process (Table 2). "LCLP-success" and "LCLP and AA and 2nd Opt" attained average distortion smaller than the time-sharing bound. However, in the case that only LCLP was used for the vector-quantization process, the success rate was very low when the compression rate was low as shown in Table 4.2. Thus, "LCLP-success" is plotted only for a compression rate higher than 0.5. These findings show the effectiveness of the auxiliary methods.

Remark 4.9
We note in Table 4.2 *that the success rate of the vector-quantization process* (4.56) *increases as the compression rate* k/n *increases, and there is a gap between compression rates 0.5 and 0.6. It might be an interesting open problem to clarify the relation among parameters* k*,* l*,* n *and success rate.*

4.5 Concluding Remarks

We considered a source coding problem with a fidelity criterion and constructed a lossy source code that can achieve the optimal rate, the rate-

Table 4.1: Parameters for matrices: $n = 100$

Rate	l	k
0.25	80	25
0.40	75	40
0.50	60	50
0.60	50	60
0.75	40	75
0.80	25	80

Table 4.2: Success Rate of LCLP for Vector-Quantization

Compression rate	0.25	0.40	0.50	0.60	0.75	0.80
LCLP success rate (%)	0	0	0	50	64	88

distortion function. While some simulation experiments using the LP relaxation method were presented, further improvement of the proposed algorithm will be needed. Since Murayama [37] showed good performance for longer block length cases by improving the sum-product algorithm, improving or modifying the sum-product algorithm is also needed to be investigated.

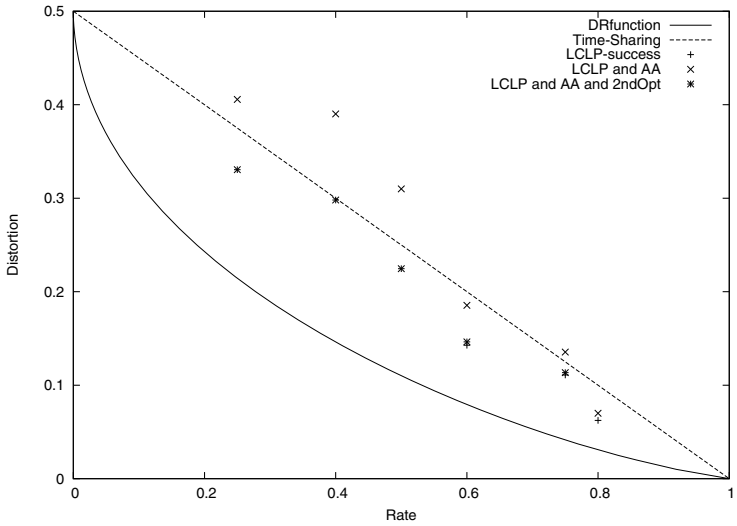

Figure 4.8: Simulation results: $n = 200$, $P_U(U = 1) = 0.5$.

Chapter 5

Channel Coding

Miller and Burshtein [28] showed that the Shannon limit of the transmission rate for the binary symmetric channel (BSC) can be attained by using a low density parity check (LDPC) code. When the channel is a general stationary memoryless channel, which is not necessarily a BSC, a quantization mapping method [2] developed by Bennatan and Burshtein can be used. Using the method, they evaluated an error exponent and obtained various interesting simulation results. However, since they assumed the decoder was an ML decoder, the universal channel code was not easy to construct. Moreover, the properties of joint source-channel coding systems constructed by LDPC codes or sparse matrices have not yet been studied much.

It is well known that for a joint source-channel system, optimal coding in the sense of the limit of the minimum transmission ratio (LMTR) [7] can be attained using the independently optimized source code channel code. An evaluation of an error exponent [47] and an analysis of general sources and channels [45] have recently been reported, whereas the advantages of using sparse matrix coding for the joint source-channel system have not yet been investigated.

In this chapter, we will present a channel coding theorem for arbitrary stationary memoryless channels, in which code is constructed in a manner dual to that in the lossy source coding theorem. By combining the results for the lossy source coding and channel coding, we will obtain a simpler result for the construction of joint source-channel coding system. The result shows that we can obtain a simpler code using sparse matrix coding for the system.

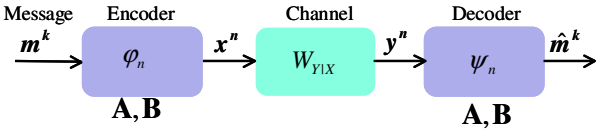

Figure 5.1: Channel Coding System

5.1 Preliminaries and Problem Setting

The alphabet treated here is a set $[0 : q-1]$, where q is a prime number and the set is also considered as a field $GF(q)$. The basis of ln is e and of log is q.

Let a transmitted message be $m^k \in GF(q)^k$ and an encoder be $\varphi_n : GF(q)^k \to GF(q)^n$. A channel is assumed to be stationary and memoryless and is described by a conditional probability distribution $W_{Y|X}$, where X and Y denote random variables describing channel input and output, respectively. Both X and Y are elements in $GF(q)$. A decoder is denoted as $\psi_n : GF(q)^n \to GF(q)^k$, and the decoded message is $\hat{m}^k \in GF(q)^k$. φ_n and ψ_n are constructed by $n \times l$ sparse matrix $\boldsymbol{A}$ and $n \times k$ sparse matrix $\boldsymbol{B}$ as shown in a later section. Note that the transmission rate R is defined as k/n and the channel capacity $C(W)$ can be described as (e.g. [7] [9])

$$C(W) = \max_P I(P, W). \tag{5.1}$$

With the above preparation, the problem considered is as follows.

[**Problem**:]

When a stationary memoryless channel $W_{Y|X}$ is given, construct the encoder φ_n and the decoder ψ_n using sparse matrices, for which the decoding error vanishes and the transmission rate approaches the channel capacity $C(W)$ asymptotically with block length n.

Figure 5.1 shows a block diagram of the channel coding problem.

5.2 Main Theorem and Proofs

5.2.1 Main Theorem

Theorem 5.1

Let a probability distribution P_X be given and fixed. If there exists a positive

number δ that satisfies $\frac{l+k}{n} < H(X)-\delta$ and $\frac{l}{n} > H(X|Y)+\delta^{1/3}$ for sufficiently large n, l, and k, then an encoder φ_n and a decoder ψ_n can be constructed by an $n \times l$ sparse matrix $\boldsymbol{A}$ and an $n \times k$ sparse matrix $\boldsymbol{B}$ that satisfy

$$W_{Y^n|X^n}\left(\left(\psi_n^{-1}(m^k)\right)^c|\varphi_n(m^k)\right) \to 0 \ (n \to \infty).$$

■

Remark 5.1
Assume that the conditions of Theorem 5.1, $\frac{l+k}{n} < H(X) - \delta$ *and* $\frac{l}{n} > H(X|Y) + \delta^{1/3}$, *are satisfied for a sufficiently small* δ. *If we set* $\frac{l+k}{n} = H(X) - 2\delta$ *and* $\frac{l}{n} = H(X|Y) + 2\delta^{1/3}$, *then the transmission rate* $\frac{k}{n}$ *is equal to* $I(X;Y) - 2(\delta + \delta^{1/3})$.

This observation shows that if $I(X;Y)$ is equal to the channel capacity, the transmission rate can approach the capacity asymptotically.

Note that the input probability distribution, which makes $I(X;Y)$ equal to the channel capacity, can be computed by the Arimoto-Blahut algorithm (e.g. [7] [9]).

5.2.2 Construction of Encoder φ_n and Decoder ψ_n

In this subsection, construction of an encoder and a decoder using sparse matrices is shown. For the construction, $n \times l$ sparse matrix $\boldsymbol{A}$, $n \times k$ sparse matrix $\boldsymbol{B}$, and l-dimensional row vector c^l are used. Both matrices are constructed in the manner described in Section 2.1.1 ($q = 2$) or 2.2.1 ($q \geq 3$). c^l is taken as a non-zero vector. Especially when $q = 2$, the Hamming weight of c^l is taken to be even, which comes from the same reason as stated in Section 4.2.2.

Assume that $\boldsymbol{A}$, $\boldsymbol{B}$, the realization value of corresponding random variables, and a fixed row vector c^l are known to both encoder and decoder.

Construction of Encoder φ_n

Figure 5.2 shows an outline of the encoding process.

Assume that a probability distribution P_X is given. Let the message sequence be $m^k \in GF(q)^k$. The codeword $x^n = \varphi_n(m^k)$ is then obtained by

$$\varphi_n(m^k) \stackrel{\text{def}}{=} \arg \max_{\substack{\tilde{x}^n : \tilde{x}^n \boldsymbol{A} = c^l \\ \tilde{x}^n \boldsymbol{B} = m^k}} P_{X^n}(\tilde{x}^n). \tag{5.2}$$

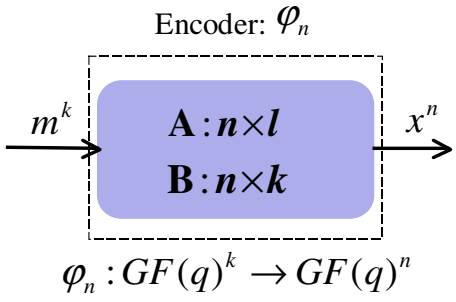

Figure 5.2: Construction of encoder

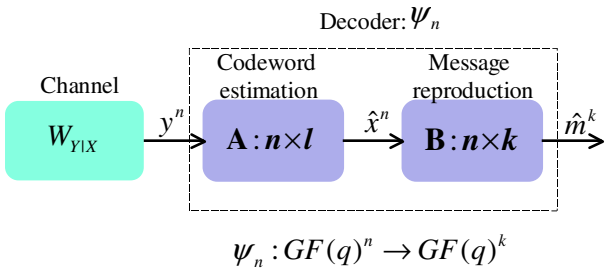

Figure 5.3: Construction of decoder

Construction of Decoder ψ_n

Figure 5.3 shows an outline of the decoding process. The decoder ψ_n consists of an estimation part and a reproduction part. Both parts are constructed of sparse matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

When a channel output $y^n \in GF(q)^n$ is given, the output of the estimation part, $\hat{x}^n$, is defined as

$$\hat{x}^n \stackrel{\text{def}}{=} \arg \max_{\tilde{x}^n : \tilde{x}^n \boldsymbol{A} = c^l} P_{X^n|Y^n}(\tilde{x}^n | y^n), \tag{5.3}$$

where $P_{X^n|Y^n} = P_{X^n} W_{Y^n|X^n} / P_{Y^n}$. Reproduction message $\hat{m}^k = \psi_n(y^n)$ is obtained by

$$\hat{m}^k \stackrel{\text{def}}{=} \hat{x}^n \boldsymbol{B}. \tag{5.4}$$

If decoding is correctly performed, $\hat{m}^k = m^k$ holds.

In the above definition, matrix $\boldsymbol{A}$ provides the syndrome that is necessary for decoding, and matrix $\boldsymbol{B}$ is introduced to make a correspondence between a message and a codeword. Note that the channel decoding process (5.3) and (5.4) stated above seems to be dual to the encoding process of lossy source coding (4.3) and (4.4).

Remark 5.2
When $q = 2$ (binary alphabet) and the channel is binary symmetric with the crossover probability ξ, the optimal probability distribution is $P_X(X = 1) = 0.5$. Then, conditions in Theorem 5.2 *are stated as*

$$\frac{l+k}{n} < 1 - \delta$$

and

$$\frac{l}{n} > h(\xi) + \delta^{1/3}.$$

Since P_X is uniform, the encoder (4.2) *obtains the codeword x^n by solving linear equations $\tilde{x}^n \boldsymbol{A} = c^l$ and $\tilde{x}^n \boldsymbol{B} = m^k$. From the condition $\frac{l+k}{n} < 1 - \delta$, the rank of concatenated matrix of $\boldsymbol{A}$ and $\boldsymbol{B}$ becomes $l + k$ with high probability. Therefore, by taking extra $n - (l + k)$ values of $\tilde{x}^n$ as arbitrary fixed values, we can increase the number of constraints in the decoding process* (4.3) *from l to $l + n - (l + k) = n - k$.*

5.3 Proof of Theorem

Since when the channel capacity $C(W)$ is 0, reliable transmission over the channel is impossible, $C(W) > 0$ is assumed throughout this section. Therefore, for the input probability distribution P_X in lemmas below, $H(P_X) > 0$ is assumed. Throughout the proof, $\varepsilon > 0$ is a given positive constant and fixed, the conditions of which are specified later in this section.

Note that from the construction described in Section 2.1 or 2.2, random variables $\boldsymbol{A}$, $\boldsymbol{B}$ are independent of each other, and we use $\mathrm{P}_{\boldsymbol{AB}}[\cdot]$ and $\mathrm{E}_{\boldsymbol{AB}}[\cdot]$ as the probability distribution and expectation operation over the random variables $\boldsymbol{A}$, $\boldsymbol{B}$, respectively.

In the proof of Theorem 4.1, we use a random coding technique over sparse matrices $\boldsymbol{A}$, $\boldsymbol{B}$.

Note that for a fixed message m^k,

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}} W_{Y^n|X^n}\left[\psi_n(Y^n) \neq \varphi_n(m^k) \mid \varphi_n(m^k)\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1\right] + \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] W_{Y^n|X^n}\left[\mathcal{E}_2 \bigcup \mathcal{E}_3 \mid \varphi_n(m^k)\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1\right] + \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] W_{Y^n|X^n}\left[\mathcal{E}_2 \mid \varphi_n(m^k)\right] \\
&\qquad + \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] W_{Y^n|X^n}\left[\mathcal{E}_3 \bigcap \mathcal{E}_2^c \mid \varphi_n(m^k)\right], \qquad (5.5)
\end{aligned}$$

where

$$\mathcal{E}_1 \stackrel{\mathrm{def}}{=} \left\{\varphi_n(m^k) \notin T^n_{P_X \delta''(\varepsilon)} \text{ for } \exists \delta''(\varepsilon) \to 0 \ (\varepsilon \to 0)\right\}, \qquad (5.6)$$

$$\mathcal{E}_2 \stackrel{\mathrm{def}}{=} \left\{Y^n \notin T^n_{W\varepsilon}(\varphi_n(m^k))\right\}, \qquad (5.7)$$

$$\mathcal{E}_3 \stackrel{\mathrm{def}}{=} \left\{\hat{X}^n \neq \varphi_n(m^k)\right\}, \qquad (5.8)$$

and $\hat{X}^n$ in (5.8) is the output of the "codeword estimation" part of the decoder (see Figure 5.3). $\mathcal{E}_1$ and $\mathcal{E}_3$ refer to "encoding error" and "decoding error", respectively.

From the law of large numbers,

$$W_{Y^n|X^n}\left[\mathcal{E}_2 \mid \varphi_n(m^k)\right] \to 0 \ (n \to \infty). \qquad (5.9)$$

In the remainder of this section, we will evaluate

$$\mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1\right]$$

and

$$\mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] W_{Y^n|X^n}\left[\mathcal{E}_3 \bigcap \mathcal{E}_2^c \mid \varphi_n(m^k)\right].$$

5.3.1 Evaluation of $\mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1\right]$

This subsection proceeds almost parallel to Section 4.3.1.

Let

$$G_{\boldsymbol{AB}}(c^l m^k) \stackrel{\mathrm{def}}{=} \left\{x^n \in [0:q-1]^n \mid x^n(\boldsymbol{AB}) = (c^l m^k)\right\}, \qquad (5.10)$$

$$\mathcal{E}_{11} \stackrel{\mathrm{def}}{=} \left\{G_{\boldsymbol{AB}}(c^l m^k) \bigcap T^n_{P_X \varepsilon} = \emptyset\right\}, \qquad (5.11)$$

and

$$\mathcal{E}_{12} \stackrel{\text{def}}{=} \left\{ \varphi_n(m^k) \notin G_{\boldsymbol{AB}}(c^l m^k) \bigcap T^n_{P_X \delta''(\varepsilon)}(u^n) \text{ for } \exists \delta''(\varepsilon) \to 0 \ (\varepsilon \to 0) \right\}, \tag{5.12}$$

where $(\boldsymbol{AB})$ denotes juxtaposition of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and similarly $(c^l m^k)$ denotes juxtaposition of row vectors c^l and m^k. Then, note that

$$\mathcal{E}_1 \subset \mathcal{E}_{11} \bigcup \mathcal{E}_{12}. \tag{5.13}$$

Therefore, it holds that

$$\mathrm{E}_{\boldsymbol{AB}} \mathbf{1}\left[\mathcal{E}_1\right] \leq \mathrm{E}_{\boldsymbol{AB}} \mathbf{1}\left[\mathcal{E}_{11}\right] + \mathrm{E}_{\boldsymbol{AB}} \mathbf{1}\left[\mathcal{E}_{12} \bigcap \mathcal{E}_{11}^c\right]. \tag{5.14}$$

For $\mathrm{E}_{\boldsymbol{AB}} \mathbf{1}\left[\mathcal{E}_{11}\right]$, the next lemma holds.

Lemma 5.1
For a fixed m^k, if there exists a positive number δ that satisfies $\frac{l+k}{n} < H(X) - \delta$ for sufficiently large n, l, and k, then we have

$$\lim_{n \to \infty} \mathrm{E}_{\boldsymbol{AB}} \mathbf{1}\left[\mathcal{E}_{11}\right] = 0. \tag{5.15}$$

■

[Proof of Lemma 5.1*]*
We prove the case $q \geq 3$. The case $q = 2$ can be proven similarly.

Let $\tilde{R} = \frac{l+k}{n}$, and set $\gamma_{\tilde{R}} < \min_{a: P_X(a) > 0} P_X(a)(1-\varepsilon)$. Then, note that for any $x^n \in T^n_{P_X \varepsilon}$, the Hamming weight of x^n satisfies $w(x^n) > n\gamma_{\tilde{R}}$.

To prove the lemma, it is sufficient to show

$$P_{\boldsymbol{AB}} \left\{ \sum_{x^n \in T^n_{P_X \varepsilon}} \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] = 0 \right\} \to 0 \quad (n \to \infty), \tag{5.16}$$

where $\mathbf{1}$["logical equation"] is an indicator function for the logical equation.

$$
\begin{aligned}
&P_{\boldsymbol{AB}}\left\{\sum_{x^n\in T^n_{P_X\varepsilon}}\mathbf{1}\left[x^n(\boldsymbol{AB})=(c^lm^k)\right]=0\right\}\\
&=P_{\boldsymbol{AB}}\left\{\sum_{x^n\in T^n_{P_X\varepsilon}}\left(\mathbf{1}\left[x^n(\boldsymbol{AB})=(c^lm^k)\right]-\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}\right)\right.\\
&\qquad\qquad\left.=-\sum_{x^n\in T^n_{P_X\varepsilon}}\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}\right\}\\
&\le P_{\boldsymbol{AB}}\left\{\sum_{x^n\in T^n_{P_X\varepsilon}}\left(\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}-\mathbf{1}\left[x^n(\boldsymbol{AB})=(c^lm^k)\right]\right)\right.\\
&\qquad\qquad\left.\ge\sum_{x^n\in T^n_{P_X\varepsilon}}\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}\right\}\\
&\overset{\text{(a)}}{\le}\frac{\mathrm{E}_{\boldsymbol{AB}}\left|\sum_{x^n\in T^n_{P_X\varepsilon}}\left(\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}-\mathbf{1}\left[x^n(\boldsymbol{AB})=(c^lm^k)\right]\right)\right|^2}{\left|\sum_{x^n\in T^n_{P_X\varepsilon}}\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}\right|^2}\\
&\overset{\text{(b)}}{\le}\frac{\mathrm{E}_{\boldsymbol{AB}}\left|\sum_{x^n\in T^n_{P_X\varepsilon}}\left(\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}-\mathbf{1}\left[x^n(\boldsymbol{AB})=(c^lm^k)\right]\right)\right|^2}{\left|\frac{(1-\delta_n(\tilde{R}))|T^n_{P_X\varepsilon}|}{q^{l+k}}\right|^2},
\end{aligned}
\tag{5.17}
$$

where (a) comes from Chebyshev's inequality and at (b), 2) of Lemma 2.5 is used for the denominator.

On the other hand,

$$
\begin{aligned}
&\text{(Numerator of (5.17))}\\
&= \mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \in T^n_{P_X \varepsilon}} \left| \left(\frac{\alpha_n(\tilde{R}; w(x^n))}{q^{l+k}} - \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \right) \right|^2 \\
&\quad + \mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \neq \tilde{x}^n \in T^n_{P_X \varepsilon}} \left\{ \frac{\alpha_n(\tilde{R}; w(x^n))}{q^{l+k}} \frac{\alpha_n(\tilde{R}; w(\tilde{x}^n))}{q^{l+k}} \right. \\
&\qquad\qquad - \frac{\alpha_n(\tilde{R}; w(x^n)) \mathbf{1}\left[\tilde{x}^n(\boldsymbol{AB}) = (c^l m^k)\right]}{q^{l+k}} \\
&\qquad\qquad - \frac{\alpha_n(\tilde{R}; w(\tilde{x}^n)) \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right]}{q^{l+k}} \\
&\qquad\qquad \left. + \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[\tilde{x}^n(\boldsymbol{AB}) = (c^l m^k)\right] \right\} \\
&\overset{\text{(c)}}{=} \sum_{x^n \in T^n_{P_X \varepsilon}} \left\{ \frac{\alpha_n(\tilde{R}; w(x^n))}{q^{l+k}} - \left(\frac{\alpha_n(\tilde{R}; w(x^n))}{q^{l+k}} \right)^2 \right\} \\
&\quad - \sum_{x^n \neq \tilde{x}^n \in T^n_{P_X \varepsilon}} \frac{\alpha_n(\tilde{R}; w(x^n))}{q^{l+k}} \frac{\alpha_n(\tilde{R}; w(\tilde{x}^n))}{q^{l+k}} \\
&\quad + \mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \neq \tilde{x}^n \in T^n_{P_X \varepsilon}} \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[\tilde{x}^n(\boldsymbol{AB}) = (c^l m^k)\right], \qquad (5.18)
\end{aligned}
$$

where at (c), 2) of Lemma 2.5 with Remark 2.4 is used.

The third term of (5.18) can be evaluated as follows:

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \neq \tilde{x}^n \in T^n_{P_X \varepsilon}} \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[\tilde{x}^n(\boldsymbol{AB}) = (c^l m^k)\right] \\
&\overset{\text{(d)}}{=} \mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \in T^n_{P_X \varepsilon}} \left(\sum_{w=1}^{n\gamma_{\tilde{R}}} + \sum_{w=n\gamma_{\tilde{R}}+1}^{n} \right) \\
&\qquad \sum_{\tilde{x}^n \in T^n_{P_X \varepsilon} : w(\tilde{x}^{*n} - x^{*n}) = w} \mathbf{1}\left[x^{*n}(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[(\tilde{x}^{*n} - x^{*n})(\boldsymbol{AB}) = 0^{l+k}\right] \\
&\overset{\text{(e)}}{\leq} \frac{(1 + \delta_n(\tilde{R}))\beta_n(\tilde{R}) |T^n_{P_X \varepsilon}|}{q^{l+k}} + (1 + \delta_n(\tilde{R}))^2 \left(\frac{|T^n_{P_X \varepsilon}|}{q^{l+k}} \right)^2, \qquad (5.19)
\end{aligned}
$$

where (d) comes from Lemma 2.3, and (e) comes from Lemma 2.6.

Note that

$$\frac{|T^n_{P_X\varepsilon}|}{q^{l+k}} \geq q^{n\left(H(X)-\frac{l+k}{n}-\delta(\varepsilon)\right)}, \tag{5.20}$$

where (5.20) comes from 3) of Lemma 2.7, and the left hand side of (5.20) approaches ∞ ($n \to \infty$). This comes from the assumption $\frac{l+k}{n} < H(X) - \delta$ by taking ε that satisfies

$$\varepsilon \log \frac{1}{\varepsilon} + 3\varepsilon < \frac{\delta}{2}, \tag{5.21}$$

with sufficiently large n to hold

$$\frac{q^2 \log(n+1)}{n} < \frac{\delta}{2}, \tag{5.22}$$

where $\delta(\varepsilon) \stackrel{\text{def}}{=} \varepsilon \log \frac{1}{\varepsilon} + 3\varepsilon + \frac{q^2 \log(n+1)}{n}$.

By substituting (5.19) into (5.18) and noting that

$$\delta_n(\tilde{R}) \to 0 \ (n \to \infty), \tag{5.23}$$

which comes from 2) of Lemma 2.5, it can be shown that

$$
\begin{aligned}
&\frac{\mathrm{E}_{\boldsymbol{AB}}\left|\sum_{x^n\in T^n_{P_{X^\varepsilon}}}\left(\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}-\mathbf{1}\left[x^n(\boldsymbol{AB})=(c^lm^k)\right]\right)\right|^2}{\left|\frac{(1-\delta_n(\tilde{R}))|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right|^2}\\
&\le \frac{\sum_{x^n\in T^n_{P_{X^\varepsilon}}}\left\{\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}-\left(\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}\right)^2\right\}}{\left|\frac{(1-\delta_n(\tilde{R}))|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right|^2}\\
&\quad-\frac{\sum_{x^n\neq\tilde{x}^n\in T^n_{P_{X^\varepsilon}}}\frac{\alpha_n(\tilde{R};w(x^n))}{q^{l+k}}\frac{\alpha_n(\tilde{R};w(\tilde{x}^n))}{q^{l+k}}}{\left|\frac{(1-\delta_n(\tilde{R}))|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right|^2}\\
&\quad+\frac{\frac{(1+\delta_n(\tilde{R}))\beta_n(\tilde{R})|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}}{\left|\frac{(1-\delta_n(\tilde{R}))|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right|^2}+\frac{(1+\delta_n(\tilde{R}))^2\left(\frac{|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right)^2}{\left|\frac{(1-\delta_n(\tilde{R}))|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right|^2}\\
&\overset{\text{(f)}}{\le}\frac{1+\delta_n(\tilde{R})}{(1-\delta_n(\tilde{R}))^2}\left(\frac{|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right)^{-1}-1\\
&\quad+\frac{1+\delta_n(\tilde{R})}{(1-\delta_n(\tilde{R}))^2}\beta_n(\tilde{R})\left(\frac{|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right)^{-1}+\frac{(1+\delta_n(\tilde{R}))^2}{(1-\delta_n(\tilde{R}))^2}\\
&=\frac{1+\delta_n(\tilde{R})}{(1-\delta_n(\tilde{R}))^2}\left(\frac{|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right)^{-1}\\
&\quad+\frac{1+\delta_n(\tilde{R})}{(1-\delta_n(\tilde{R}))^2}\beta_n(\tilde{R})\left(\frac{|T^n_{P_{X^\varepsilon}}|}{q^{l+k}}\right)^{-1}+\frac{4\delta_n(\tilde{R})}{(1-\delta_n(\tilde{R}))^2}\\
&\to 0\ (n\to\infty),
\end{aligned}
\tag{5.24}
$$

where at (f), we use 2) of Lemma 2.5. From the above arguments, (5.15) holds.

[End of Proof of Lemma 5.1*]*

Before analyzing the second term of (5.14), we show a lemma that has a dual meaning with Lemma 5.1.

Lemma 5.2
For a fixed m^k, and for any positive number $\delta > 0$, let

$$\tilde{G}_{\boldsymbol{AB}}(c^l m^k) \stackrel{\text{def}}{=} \left\{ x^n \in [0:q-1]^n \;\middle|\; x^n(\boldsymbol{AB}) = (c^l m^k) \text{ and } H(Q) < \frac{l+k}{n} - \delta \right\}, \tag{5.25}$$

where Q appearing in the above definition is a type of x^n. Then we have

$$\lim_{n\to\infty} P_{\boldsymbol{AB}}\left[\tilde{G}_{\boldsymbol{AB}}(c^l m^k) \neq \emptyset\right] = 0. \tag{5.26}$$

■

[Proof of Lemma 5.2]

$$\begin{aligned}
&P_{\boldsymbol{AB}}\left[\tilde{G}_{\boldsymbol{AB}}(c^l m^k) \neq \emptyset\right] \\
&= \mathrm{E}_{\boldsymbol{AB}} \mathbf{1}\left[\sum_{x^n \in [0:q-1]^n} \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[x^n \in \bigcup_{Q: H(Q) < \frac{l+k}{n} - \delta} T_Q^n \right] \geq 1 \right] \\
&\stackrel{\text{(a)}}{\leq} \mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \in [0:q-1]^n} \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[x^n \in \bigcup_{Q: H(Q) < \frac{l+k}{n} - \delta} T_Q^n \right],
\end{aligned} \tag{5.27}$$

where (a) comes from Markov's inequality.

Since the assumption $c^l \neq 0^l$, it holds that

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}} \sum_{x^n \in [0:q-1]^n} \mathbf{1}\left[x^n(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[x^n \in \bigcup_{Q: H(Q) < \frac{l+k}{n} - \delta} T_Q^n \right] \\
&\stackrel{\text{(b)}}{=} \mathrm{E}_{\boldsymbol{AB}} \left(\sum_{w=1}^{n\gamma_{\tilde{R}}} + \sum_{w=n\gamma_{\tilde{R}}+1}^{n} \right) \\
&\qquad \cdot \sum_{x^n : w(x^{*n}) = w} \mathbf{1}\left[x^{*n}(\boldsymbol{AB}) = (c^l m^k)\right] \mathbf{1}\left[x^n \in \bigcup_{Q: H(Q) < \frac{l+k}{n} - \delta} T_Q^n \right]
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{w=1}^{n\gamma_{\tilde{R}}} \sum_{x^n:w(x^{*n})=w} \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[x^{*n}(\boldsymbol{AB}) = (c^l m^k)\right] \\
&\quad + \sum_{w=n\gamma_{\tilde{R}}+1}^{n} \mathrm{E}_{\boldsymbol{AB}} \sum_{x^n:w(x^{*n})=w} \mathbf{1}\left[x^{*n}(\boldsymbol{AB}) = (c^l m^k)\right] \\
&\qquad\qquad \cdot\mathbf{1}\left[x^n \in \bigcup_{Q:H(Q)<\frac{l+k}{n}-\delta} T_Q^n\right] \\
&\overset{\text{(c)}}{\leq} \beta_n(\tilde{R}) + (1+\delta_n(\tilde{R})) \sum_{x^n\in[0:q-1]^n} \frac{\mathbf{1}\left[x^n \in \bigcup_{Q:H(Q)<\frac{l+k}{n}-\delta} T_Q^n\right]}{q^{l+k}} \\
&\overset{\text{(d)}}{\leq} \beta_n(\tilde{R}) + (1+\delta_n(\tilde{R})) \sum_{Q:H(Q)<\frac{l+k}{n}-\delta} \frac{|T_Q^n|}{q^{l+k}} \\
&\leq \beta_n(\tilde{R}) + (1+\delta_n(\tilde{R})) \sum_{Q:H(Q)<\frac{l+k}{n}-\delta} \frac{q^{nH(Q)}}{q^{l+k}} \\
&\leq \beta_n(\tilde{R}) + (1+\delta_n(\tilde{R})) \sum_{Q:H(Q)<\frac{l+k}{n}-\delta} \frac{q^{n(\frac{l+k}{n}-\delta)}}{q^{l+k}} \\
&\leq \beta_n(\tilde{R}) + (1+\delta_n(\tilde{R}))(n+1)^q q^{-n\delta}, \qquad (5.28)
\end{aligned}
$$

where (c) comes from Lemma 2.5, and (d) is derived using the union bound.

By substituting (5.28) into (5.27), Lemma 5.2 is proved.

[End of Proof of Lemma 5.2*]*

Using the argument similar to that in the proof of the lossy source coding theorem, to prove

$$\mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_{12}\bigcap\mathcal{E}_{11}^c\right] \to 0 \ (n\to\infty), \qquad (5.29)$$

it is sufficient to show that the output of the encoder, $\hat{x}^n$, is a P_X-typical sequence with high probability, which means that

$$||P_{\hat{x}^n} - P_X|| < \delta''(\varepsilon) \text{ for a function } \delta''(\varepsilon) \text{ which satisfies } \delta''(\varepsilon)\to 0 \ (\varepsilon\to 0), \qquad (5.30)$$

where

$$\hat{x}^n \overset{\text{def}}{=} \arg \max_{\tilde{x}^n:\tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} P_X^n(\tilde{x}^n), \qquad (5.31)$$

by using Lemma 4.3.

Note that

$$\begin{aligned}
&\arg \max_{\tilde{x}^n : \tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} P_{X^n}(\tilde{x}^n) = \arg \max_{\tilde{x}^n : \tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} \log P_{X^n}(\tilde{x}^n) \\
&= \arg \max_{\tilde{x}^n : \tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} \sum_{i=1}^{n} \log P_X(\tilde{x}_i) \\
&= \arg \max_{\tilde{x}^n : \tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} n \sum_a Q(a) \log P_X(a) \\
&= \arg \max_{\tilde{x}^n : \tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} n \left\{ \sum_a Q(a) \log \frac{P_X(a)}{Q(a)} - \sum_a Q(a) \log \frac{1}{Q(a)} \right\} \\
&= \arg \min_{\tilde{x}^n : \tilde{x}^n(\boldsymbol{AB})=(c^l m^k)} \{ D(Q||P_X) + H(Q) \}, \qquad (5.32)
\end{aligned}$$

where Q is a type of the argument $\tilde{x}^n$ in the above optimization. Let $\hat{Q}$ be the corresponding type of $\hat{x}^n$ attaining the minimum of the right hand side in (5.30). From Lemma 5.1, there exists x^n satisfying $x^n(\boldsymbol{AB}) = (c^l m^k)$ and $x^n \in T^n_{P_X \varepsilon}$ with high probability. With this x^n, since $||P_{x^n} - P_X|| < \varepsilon$ holds, we obtain

$$\begin{aligned}
&D(\hat{Q}||P_X) + H(\hat{Q}) \\
&\overset{(a)}{\leq} -\frac{1}{n} \log P_{X^n}(x^n) \\
&= \sum_a P_{x^n}(a) \log \frac{1}{P_X(a)} \\
&\leq \sum_a |P_{x^n}(a) - P_X(a)| \log \frac{1}{P_X(a)} \\
&\quad + \sum_a P_X(a) \log \frac{1}{P_X(a)} \\
&\overset{(b)}{\leq} \eta\varepsilon + H(P_X), \qquad (5.33)
\end{aligned}$$

where at (a), the definition of $\hat{x}^n$, and at (b), $\eta \overset{\text{def}}{=} \max_{a : P_X(a) > 0} \log \frac{1}{P_X(a)}$ and the fact $||P_{x^n} - P_X|| < \varepsilon$ are used, respectively.

If $D(\hat{Q}||P_X) > (\eta + 10 \log \frac{1}{\varepsilon})\varepsilon$, then

$$H(\hat{Q}) \leq H(P_X) - 10\varepsilon \log \frac{1}{\varepsilon} \qquad (5.34)$$

is derived. If

$$10\varepsilon \log \frac{1}{\varepsilon} > \delta, \tag{5.35}$$

then there exists $\frac{l+k}{n}$ satisfying both $\frac{l+k}{n} < H(X) - \delta$ and $H(\hat{Q}) < \frac{l+k}{n} - (10\varepsilon \log \frac{1}{\varepsilon} - \delta)$. Therefore, from Lemma 5.2, the probability approaches 0 for sufficiently large n that $\hat{x}^n$ satisfying the above inequality (5.34) is included in $G_{\boldsymbol{AB}}(c^l m^k)$.

Considering the above fact, it is sufficient to focus on the case of $D(\hat{Q}||P_X) \leq (\eta + 10 \log \frac{1}{\varepsilon})\varepsilon$. It holds that

$$\begin{aligned} (\eta + 10 \log \frac{1}{\varepsilon})\varepsilon &\geq D(\hat{Q}||P_X) \\ &\overset{\text{(c)}}{\geq} \frac{1}{2 \ln 2} ||\hat{Q} - P_X||^2. \end{aligned} \tag{5.36}$$

At (c), we use the formula $D(P||Q) \geq \frac{1}{2 \ln 2}||P - Q||^2$ [9, Lemma 11.6.1].

As a result, since we obtain

$$||\hat{Q} - P_X|| < \sqrt{2(\eta + 10 \log \frac{1}{\varepsilon})\varepsilon \ln 2}, \tag{5.37}$$

(5.30) is derived. From Lemma 4.3, when we set $\delta''(\varepsilon) \overset{\text{def}}{=} \frac{\sqrt{2(\eta+10 \log \frac{1}{\varepsilon})\varepsilon \ln 2}}{\min_{a:P_X(a)>0} P_X(a)}$ in (5.6), (5.29) is proved.

5.3.2 Evaluation of $\mathbf{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] W_{Y^n|X^n}\left[\mathcal{E}_3 \bigcap \mathcal{E}_2^c \;\middle|\; \varphi_n(m^k)\right]$

Assume that the message m^k is fixed. It is sufficient to show that if $\frac{l}{n} > H(X|Y) + \delta$ holds for sufficiently large n, l and for a positive number $\delta > 0$, then we have

$$\begin{aligned} \lim_{n\to\infty} \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \mathbf{1}\left[\hat{x}^n \neq \varphi_n(m^k)\right] \\ \cdot W_{Y^n|X^n}\left[Y^n \in T^n_{W\varepsilon}(\varphi_n(m^k)) \;\middle|\; \varphi_n(m^k)\right] = 0, \end{aligned} \tag{5.38}$$

where $\hat{x}^n = \arg\max_{x^n:x^n\boldsymbol{A}=c^l} P_{X|Y}(x^n|y^n)$ with channel output y^n.

Let $R' = l/n$. Then it can be shown that

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right]\mathbf{1}\left[\hat{x}^n \neq \varphi_n(m^k)\right] \\
&\qquad \cdot W_{Y^n|X^n}\left[Y^n \in T^n_{W\varepsilon}(\varphi_n(m^k)) \;\middle|\; \varphi_n(m^k)\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \\
&\qquad \cdot \mathbf{1}\left[\exists \tilde{x}^n \neq \varphi_n(m^k) \text{s.t. } P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq P_{X^n|Y^n}(\varphi_n(m^k)|y^n), \tilde{x}^n\boldsymbol{A} = c^l\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \\
&\qquad \cdot \sum_{\tilde{x}^n:\tilde{x}^n \neq \varphi_n(m^k)} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq P_{X^n|Y^n}(\varphi_n(m^k)|y^n), \tilde{x}^n\boldsymbol{A} = c^l\right] \\
&\overset{(a)}{\leq} \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \\
&\qquad \cdot \sum_{\tilde{x}^n:\tilde{x}^n \neq \varphi_n(m^k)} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\mathbf{1}\left[\tilde{x}^n\boldsymbol{A} = c^l\right] \\
&= \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \left(\sum_{w=1}^{n\gamma_{R'}} + \sum_{w=n\gamma_{R'}+1}^{n}\right) \\
&\qquad \cdot \sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right] \\
&\leq \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \sum_{w=1}^{n\gamma_{R'}} \sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w} \mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right] \\
&\quad + \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[\mathcal{E}_1^c\right] \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \sum_{w=n\gamma_{R'}+1}^{n} \\
&\qquad \cdot \sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right] \\
&\leq \mathrm{E}_{\boldsymbol{A}} \sum_{w=1}^{n\gamma_{R'}} \sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w} \mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right] \\
&\quad + \mathrm{E}_{\boldsymbol{AB}} \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \sum_{w=n\gamma_{R'}+1}^{n} \\
&\qquad \cdot \sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right],
\end{aligned}
\tag{5.39}
$$

where at (a), $\varphi_n(m^k) \in T^n_{P_X \delta''(\varepsilon)}$ is used.

Using 1) of Lemma 2.5, the first term of (5.39) is upper bounded by $\beta_n(R')$. Note that in the second term of (5.39), since both $\varphi_n(m^k)$ and $\mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right]$ depend on $\boldsymbol{A}$, we cannot directly apply 2) of Lemma 2.5 to $\mathbf{1}\left[\tilde{v}^{*n}\boldsymbol{A} = c^l\right]$. To apply Lemma 2.5, we have to decorrelate these variables using the permutation group method developed in [41] and [34].

Definition 5.1 *[Permutation Group]*
The set of all n-th order permutations $\sigma : [1:n] \to [1:n]$ is denoted as $\mathcal{S}_n$, where for $x^n = x_1, x_2, ..., x_n$ and $\boldsymbol{A} = (a_{ij})$, the operation of σ is defined as

$$\sigma(x^n) = x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)} \tag{5.40}$$

and

$$\sigma(\boldsymbol{A}) = \left(a_{\sigma(i)j}\right). \tag{5.41}$$

Note that from the construction of sparse matrices in Sections 2.1 and 2.2,

$$\begin{aligned} \mathrm{E}_{\sigma(\boldsymbol{A})\sigma(\boldsymbol{B})}\mathbf{1}\left[f\left(\sigma(\boldsymbol{A}),\sigma(\boldsymbol{B})\right)\right] &\overset{\text{def}}{=} \sum_{\boldsymbol{A},\boldsymbol{B}} P_{\boldsymbol{A}}\left(\sigma(\boldsymbol{A})\right) P_{\boldsymbol{B}}\left(\sigma(\boldsymbol{B})\right) \mathbf{1}\left[f\left(\sigma(\boldsymbol{A}),\sigma(\boldsymbol{B})\right)\right] \\ &= \mathrm{E}_{\boldsymbol{AB}}\mathbf{1}\left[f\left(\sigma(\boldsymbol{A}),\sigma(\boldsymbol{B})\right)\right], \end{aligned} \tag{5.42}$$

and from the definition of the encoding process (5.2), it holds that

$$\begin{aligned} \varphi_n^{\sigma}(m^k) &\overset{\text{def}}{=} \arg \max_{\substack{\tilde{x}^n : \tilde{x}^n\sigma(\boldsymbol{A})=c^l \\ \tilde{x}^n\sigma(\boldsymbol{B})=m^k}} P_{X^n}(\tilde{x}^n) \\ &= \sigma\left(\varphi_n(m^k)\right). \end{aligned} \tag{5.43}$$

The second term of (5.39) can be evaluated as follows:

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{AB}} \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \sum_{w=n\gamma_{R'}+1}^{n} \\
&\quad \cdot \sum_{\tilde{x}^n : w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right] \mathbf{1}\left[\tilde{x}^{*n}\boldsymbol{A} = c^l\right] \\
&= \frac{1}{n!} \sum_{\sigma\in\mathcal{S}_n} \mathrm{E}_{\sigma(\boldsymbol{A})\sigma(\boldsymbol{B})} \sum_{y^n \in T^n_{W\varepsilon}(\sigma(\varphi_n(m^k)))} W(y^n|\sigma(\varphi_n(m^k))) \sum_{w=n\gamma_{R'}+1}^{n} \\
&\quad \cdot \sum_{\tilde{x}^n : w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right] \mathbf{1}\left[\tilde{x}^{*n}\sigma(\boldsymbol{A}) = c^l\right] \\
&\overset{\text{(b)}}{=} \frac{1}{n!} \sum_{\sigma\in\mathcal{S}_n} \mathrm{E}_{\boldsymbol{AB}} \sum_{\sigma^{-1}(y^n) \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(\sigma^{-1}(y^n)|\varphi_n(m^k)) \sum_{w=n\gamma_{R'}+1}^{n} \\
&\quad \cdot \sum_{\tilde{x}^n : w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|\sigma^{-1}(y^n)) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right] \\
&\qquad\qquad \cdot \mathbf{1}\left[\tilde{x}^{*n}\sigma(\boldsymbol{A}) = c^l\right] \\
&= \frac{1}{n!} \sum_{\sigma\in\mathcal{S}_n} \mathrm{E}_{\boldsymbol{AB}} \sum_{y^n \in T^n_{W\varepsilon}(\varphi_n(m^k))} W(y^n|\varphi_n(m^k)) \sum_{w=n\gamma_{R'}+1}^{n} \\
&\quad \cdot \sum_{\tilde{x}^n : w(\tilde{x}^{*n})=w} \mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n) \geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right] \\
&\qquad\qquad \mathbf{1}\left[\tilde{x}^{*n}\sigma(\boldsymbol{A}) = c^l\right]. \qquad (5.44)
\end{aligned}
$$

At (b), (5.42) is used.

To proceed with the analysis, we use the fact that

$$
\begin{aligned}
&\mathbf{1}\left[\tilde{x}^{*n}\sigma(\boldsymbol{A}) = c^l\right] \\
&= \mathbf{1}\left[\sigma^{-1}(\tilde{x}^{*n})\boldsymbol{A} = c^l\right] \\
&= \mathbf{1}\left[\exists z^n \in \{0,1\}^n \text{ s.t. } \sigma(z^n) = \tilde{x}^{*n} \text{ and } z^n\boldsymbol{A} = c^l\right] \\
&\leq \sum_{z^n\in\{0,1\}^n : z^n\boldsymbol{A}=c^l} \mathbf{1}\left[\sigma(z^n) = \tilde{x}^{*n}\right]. \qquad (5.45)
\end{aligned}
$$

Then

$$\begin{aligned}
&\frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\mathrm{E}_{\boldsymbol{AB}}\sum_{y^n\in T^n_{W\varepsilon}(\varphi_n(m^k))}W(y^n|\varphi_n(m^k))\sum_{w=n\gamma_{R'}+1}^{n}\\
&\quad\cdot\sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w}\mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n)\geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\mathbf{1}\left[\tilde{x}^{*n}\sigma(\boldsymbol{A})=c^l\right]\\
&\overset{\text{(c)}}{\leq}\mathrm{E}_{\boldsymbol{AB}}\sum_{y^n\in T^n_{W\varepsilon}(\varphi_n(m^k))}W(y^n|\varphi_n(m^k))\sum_{w=n\gamma_{R'}+1}^{n}\\
&\quad\cdot\sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w}\mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n)\geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\\
&\quad\cdot\sum_{z^n\in\{0,1\}^n:z^n\boldsymbol{A}=c^l}\frac{\sum_{\sigma\in\mathcal{S}_n}\mathbf{1}\left[\sigma(z^n)=\tilde{x}^{*n}\right]}{n!}\\
&=\mathrm{E}_{\boldsymbol{AB}}\sum_{y^n\in T^n_{W\varepsilon}(\varphi_n(m^k))}W(y^n|\varphi_n(m^k))\sum_{w=n\gamma_{R'}+1}^{n}\\
&\quad\cdot\sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w}\mathbf{1}\left[P_{X^n|Y^n}(\tilde{x}^n|y^n)\geq q^{-n(H(X|Y)+\delta(\delta''(\varepsilon)))}\right]\\
&\qquad\cdot\frac{\sum_{z^n\in\{0,1\}^n}\mathbf{1}\left[z^n\boldsymbol{A}=c^l\right]}{\binom{n}{w}}\\
&\leq\mathrm{E}_{\boldsymbol{AB}}\sum_{y^n\in T^n_{W\varepsilon}(\varphi_n(m^k))}W(y^n|\varphi_n(m^k))\sum_{w=n\gamma_{R'}+1}^{n}\\
&\quad\cdot\sum_{\tilde{x}^n:w(\tilde{x}^{*n})=w}P_{X^n|Y^n}(\tilde{x}^n|y^n)q^{n(H(X|Y)+\delta(\delta''(\varepsilon)))}\frac{\sum_{z^n\in\{0,1\}^n}\mathbf{1}\left[z^n\boldsymbol{A}=c^l\right]}{\binom{n}{w}}\\
&\leq\mathrm{E}_{\boldsymbol{AB}}\sum_{y^n\in T^n_{W\varepsilon}(\varphi_n(m^k))}W(y^n|\varphi_n(m^k))\sum_{w=n\gamma_{R'}+1}^{n}\\
&\quad\cdot q^{n(H(X|Y)+\delta(\delta''(\varepsilon)))}\frac{\sum_{z^n\in\{0,1\}^n}\mathbf{1}\left[z^n\boldsymbol{A}=c^l\right]}{\binom{n}{w}}\\
&\leq\mathrm{E}_{\boldsymbol{AB}}\sum_{w=n\gamma_{R'}+1}^{n}q^{n(H(X|Y)+\delta(\delta''(\varepsilon)))}\frac{\sum_{z^n\in\{0,1\}^n}\mathbf{1}\left[z^n\boldsymbol{A}=c^l\right]}{\binom{n}{w}}\\
&\overset{\text{(d)}}{\leq}(1+\delta_n(R'))nq^{-n\left(\frac{l}{n}-H(X|Y)-\delta(\delta''(\varepsilon))\right)},
\end{aligned}\tag{5.46}$$

where at (c), (5.45), and at (d), 2) of Lemma 2.5 with Remark 2.4 are used, respectively.

With the above argument, if we take ε and n, l, k satisfying

$$\delta''(\varepsilon)\log\frac{1}{\delta''(\varepsilon)} + 3\delta''(\varepsilon) < \frac{\delta^{1/3}}{2}, \tag{5.47}$$

$$\frac{q^2\log(n+1)}{n} < \frac{\delta^{1/3}}{2}, \tag{5.48}$$

and $\frac{l}{n} > H(X|Y) + \delta^{1/3}$, then it is shown that the second term of (5.39) approaches 0 with $n \to \infty$.

Note that with the results in Sections 5.4.1 and 5.4.2, we can prove Theorem 5.1 through the evaluation of (5.5). ∎

Remark 5.3
δ and ε must satisfy the following relationships: from (5.35)*,*

$$10\varepsilon\log\frac{1}{\varepsilon} > \delta, \tag{5.49}$$

from (5.21)*,*

$$8\varepsilon\log\frac{1}{\varepsilon} < \delta, \tag{5.50}$$

and from (5.47)*,*

$$8\delta''(\varepsilon)\log\frac{1}{\delta''(\varepsilon)} < \delta^{1/3}. \tag{5.51}$$

By a straightforward calculation, if

$$\varepsilon\left(\log\frac{1}{\varepsilon}\right)^3\frac{(40\ln 2)^3}{(\min_{a:P_X(a)>0}P_X(a))^6} < 1 \tag{5.52}$$

is satisfied, which is a sufficient condition of $\delta''(\varepsilon)^3 < \varepsilon$, it can be shown that there exists δ satisfying (5.49)*,* (5.50)*, and* (5.51)*.*

Remark 5.4
In the proof of lossy source coding theorem, "decorrelation" of two random variables was also needed. (See (4.46)*.) Note that* (4.46) *and* (5.39) *were "decorrelated" using different techniques.*

Remark 5.5
Note that the decoding error criterion adopted here is not the maximum decoding error but the average decoding error. The maximum decoding error is defined by

$$\max_{m^k \in [0:q-1]^k} W_{Y^n|X^n} \left[\psi_n(Y^n) \neq m^k \mid \varphi_n(m^k) \right], \tag{5.53}$$

and the average decoding error is defined by

$$\frac{1}{q^k} \sum_{m^k \in [0:q-1]^k} W_{Y^n|X^n} \left[\psi_n(Y^n) \neq m^k \mid \varphi_n(m^k) \right]. \tag{5.54}$$

From Theorem 5.1, it can be easily shown that

$$\lim_{n\to\infty} \mathrm{E}_{\boldsymbol{AB}} \frac{1}{q^k} \sum_{m^k \in [0:q-1]^k} W_{Y^n|X^n} \left[\psi_n(Y^n) \neq m^k \mid \varphi_n(m^k) \right] = 0. \tag{5.55}$$

By applying Markov's inequality to the result of Theorem 5.1*, we can see that, with high probability, sparse matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *can be taken, which makes the average decoding error arbitrarily small.*

5.4 Joint Source-Channel Coding

Combining Theorem 4.1 of a lossy source coding with Theorem 5.1, we can consider the joint source-channel coding (JSCC) problem. In a point-to-point communication system, which has one sender and one receiver, we can obtain a code optimal in the sense of the limit of the minimum transmission ratio (LMTR [7]) by serially using the optimal source code and the optimal channel code.

Approaches to the JSCC problem that are different from the LMTR criterion have recently been proposed. Some studies evaluated the error exponent or analyzed the transmissibility condition for general sources and channels [45] [47].

In this section, we take another approach to the problem. The issue examined here is how we can simplify the encoding and decoding process while keeping transmissibility within a given distortion criterion. We are not concerned with the LMTR criterion here. In other words, using sparse matrix code, we want to simplify the conventional JSCC system described in Figure

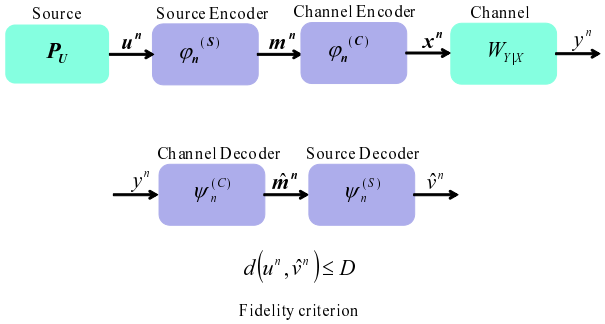

Figure 5.4: Joint source-channel coding system

5.4. For the joint source-channel coding system, it has been shown that by concatenating source encoder (decoder) and channel encoder (decoder) serially as shown in Figure 5.4, optimal performance of the code can be achieved ("source-channel separation theorem" e.g. [9]). However, the optimal system obtained by block codes need optimization processes in each encoder and/or decoder. We will show that using sparse matrix coding, the number of optimization processes can be reduced, and more efficient code than the ordinary block code in the sense of number of optimization is obtained.

For simplicity, the block length of the message from the source and that of the channel code are assumed to be equal. In this section, sufficient conditions for the following points are clarified.

- Sparse matrices of lossy source code and of channel code can the same while keeping transmissibility within a given distortion criterion.
- A quantized message of the source output becomes a channel codeword as it is.

Let a sequence from the source be U^n and that of a reproduction message be V^n. Other notations are the same as those used so far. The distortion criterion is D, and $d_n : U^n \times V^n \to [0, \infty)$ is a bounded and additive distortion measure.

A corollary about constructing the JSCC encoder and decoder shown in Figure 5.4 using encoders and decoders obtained in Theorems 4.1 and 5.2 is derived straightforwardly as follows (Figure 5.5).

By comparing conditions for $\frac{l}{n}$ and $\frac{l+k}{n}$ in Theorems 4.1 and 5.1, we can see that conditions for the existence of common sparse matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

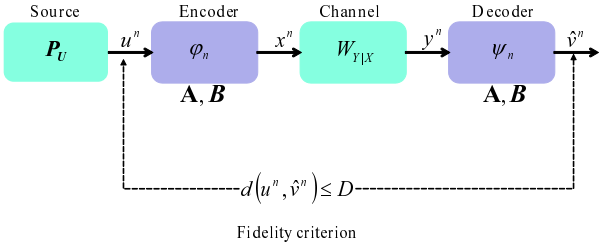

Figure 5.5: Joint source-channel coding system constructed by sparse matrix code

between lossy source code and channel code are

$$H(X|Y) < H(V|U) \tag{5.56}$$

and

$$H(V) < H(X). \tag{5.57}$$

Note that from the above conditions, the inequality $I(X;Y) > I(U;V)$ is derived.

Then, we can set the JSCC encoder $\varphi_n^{(J)} : GF(q)^n \to GF(q)^n$ as

$$x^n = \varphi_n^{(J)}(u^n) \stackrel{\text{def}}{=} \arg \max_{\substack{\tilde{x}^n : \tilde{x}^n \boldsymbol{A} = c^l \\ \tilde{x}^n \boldsymbol{B} = \left(\arg\max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} P_{V^n|U^n}(\tilde{v}^n|u^n)\right) \cdot \boldsymbol{B}}} P_{X^n}(\tilde{x}^n) \tag{5.58}$$

and the JSCC decoder $\psi_n^{(J)} : GF(q)^n \to GF(q)^n$ as

$$\hat{v}^n = \psi_n^{(J)}(y^n) \stackrel{\text{def}}{=} \arg \max_{\substack{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l \\ \tilde{v}^n \boldsymbol{B} = \left(\arg\max_{\tilde{x}^n : \tilde{x}^n \boldsymbol{A} = c^l} P_{X^n|Y^n}(\tilde{x}^n|y^n)\right) \cdot \boldsymbol{B}}} P_{V^n}(\tilde{v}^n). \tag{5.59}$$

In the above definition, if there does not exist $\tilde{x}^n$ or $\tilde{v}^n$ that satisfies conditions of the maximization, then the encoder or decoder output a fixed sequence specified in advance.

$\varphi_n^{(J)}$ maps the message from the source to the channel input, and $\psi_n^{(J)}$ maps the channel output to the reproduction message.

The next corollary follows the above discussion.

Corollary 5.1 *[Joint Source-Channel Coding]*
Source P_U, a conditional probability distribution $P_{V|U}$, an input probability distribution P_X, and channel $W_{Y|X}$ are given and fixed. Assume that

$$H(X|Y) < H(V|U)$$

and

$$H(V) < H(X)$$

are satisfied.

If there exists a positive number δ that satisfies

$$H(X|Y) + \delta^{1/3} < \frac{l}{n} < H(V|U) - \delta \tag{5.60}$$

and

$$H(V) + \delta^{1/3} < \frac{l+k}{n} < H(X) - \delta \tag{5.61}$$

for sufficiently large n, l, and k, then an encoder $\varphi_n^{(J)}$ and a decoder $\psi_n^{(J)}$ can be constructed by an $n \times l$ sparse matrix $\boldsymbol{A}$ and an $n \times k$ sparse matrix $\boldsymbol{B}$, which satisfy

$$\sum_{u^n} P_{U^n}(u^n) \sum_{y^n} W_{Y^n|X^n}\left(y^n \mid \varphi_n^{(J)}(u^n)\right) \mathbf{1}\left[\frac{d_n\left(u^n, \psi_n^{(J)}(y^n)\right)}{n} > D\right] \to 0$$
$$(n \to \infty).$$

Note from the above corollary that, while we can use common sparse matrices for lossy source coding and channel coding, in the encoding and decoding processes, lossy source encoding/decoding and channel encoding/decoding are performed separately, which seems to require heavy computation (see Figure 5.5). In contrast, the following corollary shows that the JSCC encoder and decoder can be constructed in a simpler manner. In the encoder of Corollary 5.1, after compressing the output of quantization part of the lossy encoder, the channel codeword is obtained by regarding the compressed output of the lossy encoder as the input of the channel encoder. If the output of quantization part of the lossy encoder has enough redundancy, it seems possible that the output of quantization part becomes the channel codeword as it is.

Reconsidering the proof of Theorem 5.1, we can summarize conditions for channel codewords to be correctly decoded as follows.

1. Channel codewords are included in $T^n_{P_{X\varepsilon}}$.
2. Channel codewords satisfy "syndrome condition": $\tilde{x}^n \boldsymbol{A} = c^l$

On the other hand, in Section 4.3.1, we can see that the output of the quantizer

$$v^n = \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} P_{V^n|U^n}(\tilde{v}^n | u^n) \tag{5.62}$$

is included in $T^n_{P_{V\varepsilon}}$ with high probability and, at the same time, satisfies the "syndrome condition"

$$v^n \boldsymbol{A} = c^l.$$

Therefore, we can easily conjecture that if we set the JSCC encoder $\varphi_n^{(JL)} : GF(q)^n \to GF(q)^n$, which we call the Joint Source-Channel Linear Coding (JSCLC) encoder hereafter, as

$$v^n = \varphi_n^{(JL)}(u^n) \stackrel{\text{def}}{=} \arg \max_{\tilde{v}^n : \tilde{v}^n \boldsymbol{A} = c^l} P_{V^n|U^n}(\tilde{v}^n | u^n), \tag{5.63}$$

and the JSCLC decoder $\psi_n^{(JL)} : GF(q)^n \to GF(q)^n$ as

$$\hat{v}^n = \psi_n^{(JL)}(y^n) \stackrel{\text{def}}{=} \arg \max_{\tilde{x}^n : \tilde{x}^n \boldsymbol{A} = c^l} P_{X^n|Y^n}(\tilde{x}^n | y^n), \tag{5.64}$$

then the code constructed here can satisfy the distortion criterion over the given channel. In the above definition, if there does not exist $\tilde{v}^n$ or $\tilde{x}^n$ that satisfies the optimization conditions, then the encoder or decoder output a fixed sequence specified in advance.

Here, we can see $\varphi_n^{(JL)}$ as a quantizer of lossy source coding and $\psi_n^{(JL)}$ as a channel decoder (see Figure 5.6). Note that the above construction of the code includes n and l through the sparse matrix $\boldsymbol{A}$ and does not include k nor the sparse matrix $\boldsymbol{B}$.

When we check the conditions for $\frac{l}{n}$ in Theorems 4.1 and 5.1, we can see that the condition for the existence of the sparse matrix $\boldsymbol{A}$, which constructs $\varphi_n^{(JL)}$ and $\psi_n^{(JL)}$, is

$$H(V|Y) < H(V|U)$$

by identifying V in Theorem 4.1 with X. The next corollary follows the above discussion.

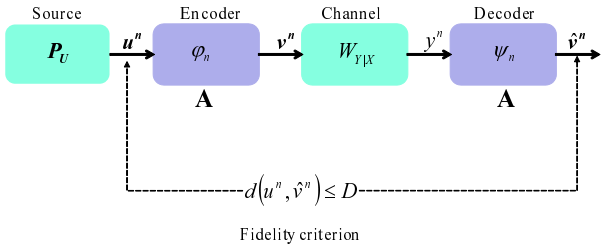

Figure 5.6: Joint source-channel linear coding system

Corollary 5.2 *[Joint Source-Channel Linear Coding]*
Source P_U, conditional probability distribution $P_{V|U}$, and channel $W_{Y|X}$, where X is identified with V, are given and fixed. Assume that

$$H(V|Y) < H(V|U)$$

is satisfied.

If there exists a positive number δ that satisfies

$$H(V|Y) + \delta^{1/3} < \frac{l}{n} < H(V|U) - \delta \tag{5.65}$$

for sufficiently large n and l, then an encoder $\varphi_n^{(JL)}$ and a decoder $\psi_n^{(JL)}$ can be constructed by an $n \times l$ sparse matrix $\boldsymbol{A}$, which satisfy

$$\sum_{u^n} P_{U^n}(u^n) \sum_{y^n} W_{Y^n|X^n}\left(y^n \mid \varphi_n^{(JL)}(u^n)\right) \mathbf{1}\left[\frac{d_n\left(u^n, \psi_n^{(JL)}(y^n)\right)}{n} > D\right] \to 0$$
$$(n \to \infty).$$

In the above corollary, not only can common sparse matrices in lossy source coding and channel coding be used, but encoding and decoding processes can also be replaced by a single operation of lossy encoding (vector quantization) and channel decoding, respectively.

Remark 5.6
Note in the above corollary that only the condition on l, and not the condition on k, appears. The right inequality of (5.65) *is a condition of the existence of*

a "good" reproduction message in the lossy source coding framework, and the left inequality of (5.65) *is a condition of obtaining a correct codeword from the channel output in the channel coding framework.*

There are two reasons the condition on k is unnecessary. One is that, from the viewpoint of lossy source coding, we do not need to compress the reproduction message in the above JSCLC framework. The other is that, from the viewpoint of channel coding, we need only a codeword and not a transmitted message in the channel coding problem.

Remark 5.7
When $q = 2$ (binary alphabet), $P_U(U = 1) = 0.5$, the channel is binary symmetric with crossover $\xi \leq 0.5$, and the distortion measure is the Hamming distance, then the condition in Corollary 5.2 *$H(V|Y) < H(V|U)$ is equivalent with $\xi < D$ by taking $P_{V|U}$ as a binary symmetric channel with crossover D.*

Corollary 5.2 shows that $I(U;V) < I(Y;V)$ is a sufficient condition for transmitting a message while keeping a distortion less than D with high probability in the JSCLC framework. Conversely, by combining proofs of converse theorems for lossy source coding and channel coding, it can be shown that when a message is transmissible with a distortion less than D, there exists a random variable V that satisfies $I(U;V) \leq I(Y;V)$.

In the remainder of this section, after defining the quantity that corresponds to the optimal distortion, we state and prove a corollary about the optimality of the quantity.

Definition 5.2 *[JSCLC-Distortion]*
Let

$$D(P_U, W_{Y|X}) \stackrel{\text{def}}{=} \min_{P_{V|U}: I(U;V) \leq I(Y;V)} E_{UV} d_1(U, V) \tag{5.66}$$

be called JSCLC-Distortion.

The next corollary holds.

Corollary 5.3
$D(P_U, W_{Y|X})$ is the minimum distortion with which a message from source P_U can be recovered after transmitting channel $W_{Y|X}$.

[Proof of Corollary 5.3*]*
With the help of the proof of an ordinary converse coding theorem (e.g. [9]), we will show that if a message can be transmitted within a certain distortion level, there exists a random variable V that satisfies $I(U;V) \leq I(Y;V)$. Random variables U^n, V^n, Y^n, and $\hat{V}^n$ are shown in Figure 5.6. At first, note that the system described in Figure 5.4 achieves the optimal performance by the source-channel separation theorem. Using above fact and the correspondence between Figure 5.4 and 5.6, it should be noted that $\hat{V}^n$ can be regarded as an estimator of "channel codeword" V^n.

Now we will make a desirable probability distribution of random variable V. From the data processing lemma [9][7],

$$I(V^n;\hat{V}^n) \leq I(V^n;Y^n) \tag{5.67}$$

holds. In the following, the left and right hand sides of (5.67) are evaluated separately.

$$\begin{aligned}
I(V^n;\hat{V}^n) &= H(V^n) - H(V^n|\hat{V}^n) \\
&\overset{\text{(a)}}{\geq} H(V^n) - n\varepsilon \\
&\geq H(V) - H(V^n|U^n) - n\varepsilon \\
&= I(U^n;V^n) - n\varepsilon \\
&\overset{\text{(b)}}{\geq} \sum_{i=1}^{n} I(U_i;V_i) - n\varepsilon \\
&\overset{\text{(c)}}{=} n \cdot \frac{1}{n}\sum_{i=1}^{n} I\left(P_U, P_{V_i|U_i}\right) - n\varepsilon \\
&\overset{\text{(d)}}{\geq} nI\left(P_U, \frac{1}{n}\sum_{i=1}^{n} P_{V_i|U_i}\right) - n\varepsilon,
\end{aligned} \tag{5.68}$$

where ε is a positive number. At (a), Fano's inequality [9][7], at (b), the i.i.d. property of U_i, at (c), another description of mutual information $I(X;Y) = I(P_X, P_{Y|X})$, and at (d), convexity of $I(P,W)$ with respect to W, are used.

On the other hand, it holds that

$$\begin{aligned} I(V^n;Y^n) &\overset{\text{(e)}}{\leq} \sum_{i=1}^{n} I(V_i;Y_i) \\ &= n \cdot \frac{1}{n} \sum_{i=1}^{n} I(P_{V_i}, W_{Y|X}) \\ &\overset{\text{(f)}}{\leq} nI\left(\frac{1}{n}\sum_{i=1}^{n} P_{V_i}, W_{Y|X}\right), \end{aligned} \tag{5.69}$$

where at (e), a memoryless property of channel $W_{Y|X}$, and at (f), a concave property of $I(P,W)$ with respect to P, are used. Combining the above evaluations, we obtain

$$I\left(P_u, \frac{1}{n}\sum_{i=1}^{n} P_{V_i|U_i}\right) \leq I\left(\frac{1}{n}\sum_{i=1}^{n} P_{V_i}, W_{Y|X}\right). \tag{5.70}$$

When we define the probability distribution of (U,V) as

$$P_U(u) \cdot \frac{1}{n}\sum_{i=1}^{n} P_{V_i|S_i}(v|u) \tag{5.71}$$

and note that

$$\begin{aligned} P_V(v) &\overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\sum_{u} P_U(u)P_{V_i|U_i}(v|u) \\ &= \frac{1}{n}\sum_{i=1}^{n} P_{V_i}(v), \end{aligned} \tag{5.72}$$

then the inequality $I(U;V) \leq I(Y;V)$ is shown to hold for V defined here.

[End of Proof of Corollary 5.3*]*

Remark 5.8

Under the same setting of Remark 5.7*, by using the symmetric property of probability distributions, it can be seen that $P_{V|U}$, which attains the minimum value in the definition of $D(P_U, W_{Y|X})$, becomes a binary symmetric channel. Using this fact, $D(P_U, W_{Y|X}) = \xi$ is obtained.*

5.5 Concluding Remarks

A channel code for a stationary discrete memoryless channel, which is not necessarily a BSC, was constructed by sparse matrices, and coding theorem was proved.

We applied sparse matrix coding results to the joint source-channel coding problem. Combining channel code with lossy source code, both of which are constructed by sparse matrices, a simpler joint source-channel code can be constructed than that constructed by the ordinary block code. The concept of JSCLC-distortion is defined as an optimal distortion keeping transmissibility.

To examine whether the proposed coding scheme can be efficiently implemented, simulation experiments that combine it with an efficient algorithm, such as the sum-product algorithm or the LCLP algorithm [12], are necessary and are future works.

There is also the universal channel coding problem. In this problem, the decoding error exponent under a fixed transmission rate is of interest. However, since we adopt the average decoding error criterion to deal with the term $\mathrm{E}_{\boldsymbol{A}\mathbf{1}}[\mathcal{E}_1]$, which depends on each message m^k, in Section 5.3, it is not straightforward to evaluate the error exponent using the expurgation technique as in Chapter 3. This problem is also a future work.

Chapter 6

Conclusion and Future Works

In fundamental point-to-point communication systems, we constructed lossless universal source code, lossy source code, and channel code using sparse matrices for stationary memoryless systems, and showed their error exponent (lossless universal code) or asymptotic optimality (lossy source code, channel code).

In Chapter 3 focusing on the lossless universal source coding problem, we showed the universality of sparse matrices that construct the encoder and decoder. We also showed that by using a decoder that does not depend on the statistical properties of the source, the decoding error approaches 0 asymptotically. The fact that the error exponent obtained in the sparse matrix coding framework is the same form as that obtained in the ordinary linear coding framework is remarkable.

In Chapter 4, the lossy source code constructed using sparse matrices was shown to have asymptotic optimality for arbitrary stationary memoryless sources with bounded and additive distortion measures. Simulation experiments were carried out by implementing the sparse matrix code using the linear programming method proposed by Feldman [12]. The experiments showed the code attains high compression performance that goes beyond the time-sharing bound.

In Chapter 5, the channel code constructed using sparse matrices was shown to have asymptotic optimality for arbitrary stationary memoryless channels. Note that the code constructed is simpler than the code proposed by Bennatan and Burshtein [2] that used a "quantization map" over a sufficiently large virtual alphabet, and while they assumed the decoder was an ML decoder, we can show the universality of our code by using the mini-

mum entropy decoding as the decoding operation [33]. We also showed the interesting duality of the encoder and decoder between the lossy source code and the channel code. For the joint source-channel coding system, while the code approaching LMTR is ordinarily constructed by serially combining the optimal lossy source code and the optimal channel code, we showed that by taking the output from the vector quantizer of the lossy source code as the channel codeword, the code construction becomes much simpler.

By constructing codes using sparse matrices, we may be able to implement the codes with efficient algorithms such as the sum-product algorithm or linear programming. In this thesis, while we showed some simulation experiments of sparse matrix codings in a simple setting, the following issues are left as future works:

1. Increase implementable block length
2. Obtain higher success rate of encoding and decoding
3. Speed up the algorithm

While we used sparse matrices, whose number of non-zero elements is $O(n \log n)$, many papers concerning LDPC codes use matrices, whose number of non-zero elements is $O(n)$. This difference appears in the execution time of the coding algorithm. In the former case, the time can be evaluated as a polynomial order of n, while in the latter case, as a linear order of n. When a sparse matrix code has asymptotic optimality, whether it is necessary that the number of non-zero elements is $O(n \log n)$ is left as another future work.

Lossy source coding systems and channel coding systems are the most fundamental constituents in information theory. Therefore, combining these coding techniques, we can expect that codes of many multi-terminal communication systems can be constructed using sparse matrices. Muramatsu and Miyake [35] [36] constructed codes for the Wyner-Ziv and the Gel'fand-Pinsker systems using the hash property that is an extended concept of sparse matrix coding. The existence theorems of many multi-terminal coding systems have been proved using random block coding. However, whether or not they can be proved using sparse matrix codes is still not clear.

Bibliography

[1] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," *Proc. 1993 Int. Conf. Commun.*, pp. 1064–1070, 1993.

[2] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete-memoryless channels," *IEEE Trans. Inform. Theory*, **vol. 50**, no. 3, pp. 417–438, 2004.

[3] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, **vol. 3**, no. 3, pp. 68–79, 1960.

[4] G. Caire, S. Shamai and S. Verdú, "Lossless data compression with error correction codes," *Proc. 2003 IEEE Int. Symp. Inform. Theory*, pp. 22, 2003.

[5] G. Caire, S. Shamai and S. Verdú, "Noiseless data compression with low density parity check codes," *in DIMACS Series in Discrete Mathematics and Theoretical Computer Science: Advances in Network Information Theory*, **vol. 66**, pp. 263–284, American Mathematical Society, 2004.

[6] T. P. Coleman, M. Médard and M. Effros, "Towards practical minimum-entropy universal coding," *Proc. 2005 Data Compression Conference (DCC'05)*, pp. 33–42, 2005.

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

[8] I. Csiszár, "Linear codes for sources and source networks: error exponents, universal coding," *IEEE Trans. Inform. Theory*, **vol. 28**, no. 4, pp. 585–592, 1982.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory: 2nd edition*, John Wiley and Sons, Inc., 2006.

[10] A. El Gamal and Y. H. Kim, *Lecture Notes on Network Information Theory*, available at arXiv:1001.3404v3 [cs.IT], 2010.

[11] U. Erez and G. Miller, "The ML decoding performance of LDPC ensembles over Z_q," *IEEE Trans. Inform. Theory*, **vol. 51**, no. 5, pp. 1871–1879, 2005.

[12] J. Feldman, *Decoding error-correcting codes via linear programming*, Ph. D. dissertation, MIT, Cambridge, 2003.

[13] M. P. C. Fossorier, S. Lin and J. Snyders, "Reliability-based syndrome decoding of linear block codes," *IEEE Trans. Inform. Theory*, **vol. 44**, no. 1, pp. 388–398, 1998.

[14] B. J. Frey, *Graphical Models for Machine Learning and Digital Communication*, Cambridge, MA: MIT Press, 1998.

[15] R. G. Gallager, *Low Density Parity Check Codes*, no. 21 in Research Monograph Series. Cambridge, MA: MIT Press, 1963.

[16] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[17] R. M. Gray, *Entropy and Information Theory*, Springer-Verlag, 1990.

[18] A. Gupta and S. Verdú, "Nonlinear sparse-graph codes for lossy compression," Proceedings of Information Theory Workshop 2007 (ITW'07), pp. 541–546, 2007.

[19] A. Hochquenghem, "Codes correcteurs d'erreurs," *Chiffres*, **vol. 2**, pp. 147–156, 1959.

[20] J. Justesen, "A class of constructive asymptotically good algebraic code," *IEEE Trans. Inform. Theory*, **vol. 18**, no. 5, pp. 652–656, 1972.

[21] G. Kramer, *Topics in Multi-User Information Theory*, Now Publishers, 2008.

[22] G. Letac and L. Takacs, "Random walk on the m-dimensional cube," *J. Reine Angrew. Math.*, **vol. 310**, pp. 187–195, 1979.

[23] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low-density parity-check codes," *Electron. Lett.*, pp. 1645–1646, 1997.

[24] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, **vol. 45**, no. 2, pp. 399–431, 1999.

[25] E. Martinian and M. J. Wainwright, "Low density codes achieve the rate-distortion bound," *Proc. Data Compression Conference, DCC 2006*, pp. 153–162, 2006.

[26] E. Martinian and M. J. Wainwright, "Low-density constructions can achieve the Wyner-Ziv and Gelfand-Pinsker bounds," *Proc. 2006 IEEE Int. Symp. Inform. Theory*, pp. 484–488, 2006.

[27] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Trans. Inform. Theory*, **vol. 49**, no. 9, pp. 2225–2229, 2003.

[28] G. Miller and D. Burshtein, "Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Inform. Theory*, **vol. 47**, no. 7, pp. 2696–2710, 2001.

[29] S. Miyake, "Lossy data compression over $\boldsymbol{Z}_q$ by LDPC code," *Proc. 2006 IEEE Int. Symp. Inform. Theory*, pp. 813–816, 2006.

[30] S. Miyake and M. Maruyama, "Construction of universal codes using LDPC matrices and their error exponents," *IEICE Trans. Fundamentals*, **vol. E90-A**, no. 9, pp. 1830–1839, 2007.

[31] S. Miyake and J. Muramatsu, "A construction of lossy source code using LDPC matrices," *IEICE Trans. Fundamentals*, **vol. E91-A**, no. 6, pp. 1488–1501, 2008.

[32] S. Miyake, J. Honda and H. Yamamoto, "Application of LCLP to lossy source coding," *Proc. Int. Symp. Inform. Theory and Its Applications (ISITA2008)*, pp. 589–594, 2008.

[33] S. Miyake and J. Muramatsu, "A construction of channel code, joint source-channel code, and universal code for arbitrary stationary memoryless channels using sparse matrices," *IEICE Trans. Fundamentals*, **vol. E92-A**, no. 9, pp. 2333–2344, 2009.

[34] J. Muramatsu, T. Uyematsu and T. Wadayama, "Low density parity check matrices for coding of correlated sources," *IEEE Trans. Inform. Theory*, **vol. 51**, no. 10, pp. 3645–3654, 2005.

[35] J. Muramatsu and S. Miyake, "Hash property and coding theorems for sparse matrices and maximum-likelihood coding," *IEEE Trans. Inform. Theory*, **vol. 56**, no. 5, pp. 2143–2167, 2010.

[36] J. Muramatsu and S. Miyake, "Hash property and fixed-rate universal coding theorems," *IEEE Trans. Inform. Theory*, **vol. 56**, no. 6, pp. 2688–2698, 2010.

[37] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Physical Review E*, **vol. E69**, pp. 035105(1)–035105(4), 2004.

[38] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Indust. Appl. Math.*, **vol. 8**, pp. 300–304, 1960.

[39] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inform. Theory*, **vol. 30**, no. 4, pp. 629–636, 1984.

[40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, **vol. 27**, pp. 379–423, pp. 623–656, 1948.

[41] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. Inform. Theory*, **vol. 45**, no. 6, pp. 2101–2104, 1999.

[42] K. Visweswariah, S. R. Kulkarni and S. Verdú, "Universal variable-to-fixed length source codes," *IEEE Trans. Inform. Theory*, **vol. 47**, no. 4, pp. 1461–1472, 2001.

[43] Z. Xiong, A. D. Liveris and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Proc. Mag.*, **vol. 21**, pp. 80–94, 2004.

[44] H. Yamamoto, *Private communication*, 2010.

[45] S. Yang and P. Qiu, "On the performance of lossless joint source-channel coding based on linear codes," *Proc. 2006 IEEE Inform. Theory Workshop*, pp. 160–164, 2006.

[46] R. Zamir, S. Shamai and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, **vol. 48**, no. 6, pp. 1250–1276, 2002.

[47] Y. Zhong, F. Alajaji and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inform. Theory*, **vol. 52**, pp. 1450–1468, 2006.

[48] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inform. Theory*, **vol. 23**, no. 3, pp. 337–343, 1977.

[49] J. Ziv and A. Lempel, "Compression of individual sequences by variable rate coding," *IEEE Trans. Inform. Theory*, **vol. 24**, no. 5, pp. 530–536, 1978.