Channel Code
Using a Constrained Random Number Generator

Jun Muramatsu
NTT Communication Science Laboratories, NTT Corporation, E-mail: muramatsu.jun@lab.ntt.co.jp

***Abstract*—A stochastic encoder for channel coding is introduced with a rate close to the channel capacity, where the only restriction is that the channel input alphabet is finite. Random numbers, which satisfy a condition specified by a function and its value, are used to construct the stochastic encoder. The proof of the theorem is based on the hash property of an ensemble of functions, where the results are extended to a general channel by deriving an alternative formula for the capacity. Since an ensemble of sparse matrices has a hash property, we can construct a code by using sparse matrices.**

I. Introduction

The aim of this paper is to introduce a code for general channels including additive Gaussian, Markov, and non-stationary channels. The only assumption is that the channel input alphabet is finite. We prove that the channel capacity is achievable with the proposed code. We introduce a *stochastic* encoder for constructing a practical code. Let $\mathcal{X}^n$ be the cartesian power of a set $\mathcal{X}$, and $\boldsymbol{x}$ denotes an element of $\mathcal{X}^n$. To construct a stochastic encoder, we use a sequence of random numbers subject to a distribution ν on $\mathcal{X}^n$ defined as

$$\nu(\boldsymbol{x}) \equiv \begin{cases} \frac{\mu(\boldsymbol{x})}{\mu(\{\boldsymbol{x}:A\boldsymbol{x}=\boldsymbol{c}\})} & \text{if } A\boldsymbol{x}=\boldsymbol{c} \\ 0 & \text{if } A\boldsymbol{x}\neq\boldsymbol{c} \end{cases}$$

for a given probability distribution μ on $\mathcal{X}^n$, a function $A : \mathcal{X}^n \to \{A\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}^n\}$, and $\boldsymbol{c} \in \{A\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}^n\}$. Let us call a generator for this type of random numbers a *constrained random number generator*. It should be noted that there is a practical algorithm [15] for the constrained random number generator by using the sum-product algorithm [1][10].

One contribution of this paper is to extend the results of [12] to general channels In [12], the direct part of the channel coding theorem for a discrete stationary memoryless channel is shown based on the hash property of an ensemble of functions. In this paper, an alternative general formula for the channel capacity is derived and the achievability of the proposed code is proved based on a stronger version of hash property introduced in [13][14]. Since an ensemble of sparse matrices has a hash property, we can construct a code by using sparse matrices.

Another contribution of this paper is that we introduce a practical code for a (continuous, asymmetric) channel by using sparse matrices and the constrained random number generator instead of the apparently intractable deterministic encoder presented in [12]. There are many ways to construct channel codes [2][7] by using sparse matrices. These approaches assume that a channel is stationary memoryless and symmetric, or a quantization map [6, Section 6.2] is used for an asymmetric channel. On the other hand, the only requirement for the proposed code is that the channel input alphabet is finite.

It should be noted that a similar idea has appeared in [16][19], where they introduced random bin coding (privacy amplification) and Slepian-Wolf decoding[1] (information reconciliation) for the construction of codes, and their proofs are based on the fact that the output statistics of random binning are uniformly distributed. This paper describes the explicit construction of encoding function and theorem is proved simply and rigorously based on the technique reported in [14], where it is proved that we can use sparse matrices for the construction of a code.

II. General Formulas for Channel Capacity

This section provides a formal description of the problem and formulas for the channel capacity. All the results in this paper are presented by using the information spectrum method introduced in [8][9][18], where the consistency and stationarity of channels/sources are not assumed. It should be noted that all the results reported in this paper can be applied to stationary ergodic channels and stationary memoryless channels. Throughout this paper, we denote the probability of an event by $P(\cdot)$ and denote the probability distribution of a random variable U by μ_U.

We call a sequence $\boldsymbol{U} \equiv \{U^n\}_{n=1}^{\infty}$ of random variables a *general source*, where $U^n \in \mathcal{U}^n$. For a general source $\boldsymbol{U}$, we define the spectral inf-entropy rate $\underline{H}(\boldsymbol{U})$ as

$$\underline{H}(\boldsymbol{U}) \equiv \sup\left\{\theta : \lim_{n\to\infty} P\left(\frac{1}{n}\log\frac{1}{\mu_{U^n}(U^n)} < \theta\right) = 0\right\}.$$

For a pair $(\boldsymbol{U},\boldsymbol{V}) = \{(U^n,V^n)\}_{n=1}^{\infty}$ of general sources, we define the spectral conditional sup-entropy rate $\overline{H}(\boldsymbol{U}|\boldsymbol{V})$, and the spectral inf-mutual information rate $\underline{I}(\boldsymbol{U};\boldsymbol{V})$ as

$$\overline{H}(\boldsymbol{U}|\boldsymbol{V}) \equiv \inf\left\{\theta : \lim_{n\to\infty} P\left(\frac{1}{n}\log\frac{1}{\mu_{U^n|V^n}(U^n|V^n)} > \theta\right) = 0\right\}$$

$$\underline{I}(\boldsymbol{U};\boldsymbol{V}) \equiv \sup\left\{\theta : \lim_{n\to\infty} P\left(\frac{1}{n}\log\frac{\mu_{U^nV^n}(U^n,V^n)}{\mu_{U^n}(U^n)\mu_{V^n}(V^n)} < \theta\right) = 0\right\},$$

where $\mu_{U^nV^n}$ is the joint probability distribution corresponding to (U^n, V^n).

In the following, we introduce the definition of the channel capacity for a general channel. Let $\mathcal{X}^n$ and $\mathcal{Y}^n$ be the alphabets

[1] It should be noted that the idea of using Slepian-Wolf decoding has already been mentioned in [11][12].

of a channel input X^n and a channel output Y^n, respectively. A sequence $\boldsymbol{W} \equiv \{\mu_{Y^n|X^n}\}_{n=1}^{\infty}$ of conditional probability distributions is called a *general channel*.

Definition 1: For a general channel $\boldsymbol{W}$, we call a rate R *achievable* if for all $\delta >$ and all sufficiently large n there is a pair consisting of an encoder $\varphi_n : \mathcal{M}_n \to \mathcal{X}^n$ and a decoder $\psi_n : \mathcal{Y}^n \to \mathcal{M}_n$ such that

$$\frac{1}{n}\log|\mathcal{M}_n| \geq R$$
$$P(\psi_n(Y^n) \neq M_n) \leq \delta,$$

where $\mathcal{M}_n$ is a set of messages, $[1/n]\log|\mathcal{M}_n|$ represents the rate of the code, M_n is a random variable of the message corresponding to the uniform distribution on $\mathcal{M}_n$ and the joint distribution $\mu_{M_nY^n}$ is given as

$$\mu_{M_nY^n}(\boldsymbol{m},\boldsymbol{y}) \equiv \frac{\mu_{Y^n|X^n}(\boldsymbol{y}|\varphi_n(\boldsymbol{m}))}{|\mathcal{M}_n|}.$$

The *channel capacity* $C(\boldsymbol{W})$ is defined by the supremum of the achievable rate.

For a general channel $\boldsymbol{W}$, the channel capacity $C(\boldsymbol{W})$ is derived in [18] as

$$C(\boldsymbol{W}) = \sup_{\boldsymbol{X}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}), \tag{1}$$

where the supremum is taken over all general sources $\boldsymbol{X} = \{X^n\}_{n=1}^{\infty}$ and the joint distribution $\mu_{X^nY^n}$ is given as

$$\mu_{X^nY^n}(\boldsymbol{x},\boldsymbol{y}) \equiv \mu_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\mu_{X^n}(\boldsymbol{x}). \tag{2}$$

We introduce the following lemma, which will be proved in Section V-A. It should be noted that this capacity formula is a straightforward generalization of that obtained by Shannon [17].

Lemma 1: For a general channel $\boldsymbol{W}$,

$$C(\boldsymbol{W}) = \sup_{\boldsymbol{X}} \left[\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})\right], \tag{3}$$

where the supremum is taken over all general sources $\boldsymbol{X}$ and the joint distribution of $(\boldsymbol{X},\boldsymbol{Y})$ is given by (2).

In this paper, we construct a channel code whose rate is close to the channel capacity given by (3). The constructed code is given by a pair consisting of a stochastic encoder $\Phi_n : \mathcal{M}_n \to \mathcal{X}^n$ and a decoder $\psi_n : \mathcal{Y}^n \to \mathcal{M}_n$. It should be noted that the capacity formulas (1) and (3) are satisfied when a stochastic encoder is allowed. In fact, by considering the average over stochastic encoders and using the random coding argument, we can construct a deterministic encoder from a stochastic encoder. Thus the rate of the stochastic encoder should be upper bounded by the channel capacity. On the other hand, the channel capacity is achievable with a stochastic encoder because a deterministic encoder is one type of stochastic encoder.

III. $(\boldsymbol{\alpha},\boldsymbol{\beta})$-HASH PROPERTY

In this section, we introduce the hash property[2] introduced in [13][14] and its implications. We use the following definitions and notations. The set $\mathcal{U}^c$ denotes the complement of $\mathcal{U}$ and the set $\mathcal{U}\setminus\mathcal{V} \equiv \mathcal{U}\cap\mathcal{V}^c$ denotes the set difference. Let $A\boldsymbol{u}$ denote a value taken by a function $A : \mathcal{U}^n \to \overline{\mathcal{U}}$ at $\boldsymbol{u} \in \mathcal{U}^n$, where $\mathcal{U}^n$ is the domain of A and $\overline{\mathcal{U}}$ is the range of A. It should be noted that A may be nonlinear. When A is a linear function expressed by an $l \times n$ matrix, we assume that $\mathcal{U} \equiv \mathrm{GF}(q)$ is a finite field and the range of functions is $\mathcal{U}^l$. For a set $\mathcal{A}$ of functions, we define $\mathrm{Im}\mathcal{A} \equiv \bigcup_{A\in\mathcal{A}}\{A\boldsymbol{u} : \boldsymbol{u} \in \mathcal{U}^n\}$. We define $\mathcal{C}_A(\boldsymbol{c}) \equiv \{\boldsymbol{u} : A\boldsymbol{u} = \boldsymbol{c}\}$ and $\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}) \equiv \{\boldsymbol{u} : A\boldsymbol{u} = \boldsymbol{c}, B\boldsymbol{u} = \boldsymbol{m}\}$, where they are called cosets in the context of linear codes. The random variables of a function A and a vector $\boldsymbol{c} \in \mathrm{Im}A$ are denoted by the sans serif letters A and c, respectively. It should be noted that the random variable of a n-dimensional vector $\boldsymbol{u} \in \mathcal{U}^n$ is denoted by the Roman letter U^n that does not represent a function, which is the way it has been used so far. Here, we introduce the hash property for an ensemble of functions. It requires stronger conditions than those introduced in [12].

Definition 2: Let $\boldsymbol{\mathcal{A}} \equiv \{\mathcal{A}_n\}_{n=1}^{\infty}$ be a sequence of sets such that $\mathcal{A}_n$ is a set of functions $A : \mathcal{U}^n \to \mathrm{Im}\mathcal{A}_n$. For a probability distribution $p_{\mathsf{A},n}$ on $\mathcal{A}_n$, we call a sequence $(\boldsymbol{\mathcal{A}},\boldsymbol{p}_{\mathsf{A}}) \equiv \{(\mathcal{A}_n, p_{\mathsf{A},n})\}_{n=1}^{\infty}$ an *ensemble*. Then, $(\boldsymbol{\mathcal{A}},\boldsymbol{p}_{\mathsf{A}})$ has an $(\boldsymbol{\alpha}_{\mathsf{A}},\boldsymbol{\beta}_{\mathsf{A}})$*-hash property* if there are two sequences $\boldsymbol{\alpha}_{\mathsf{A}} \equiv \{\alpha_{\mathsf{A}}(n)\}_{n=1}^{\infty}$ and $\boldsymbol{\beta}_{\mathsf{A}} \equiv \{\beta_{\mathsf{A}}(n)\}_{n=1}^{\infty}$, depending on $\{p_{\mathsf{A},n}\}_{n=1}^{\infty}$, such that

$$\lim_{n\to\infty} \alpha_{\mathsf{A}}(n) = 1 \tag{H1}$$

$$\lim_{n\to\infty} \beta_{\mathsf{A}}(n) = 0 \tag{H2}$$

$$\sum_{\substack{\boldsymbol{u}'\in\mathcal{U}^n\setminus\{\boldsymbol{u}\}: \\ p_{\mathsf{A},n}(\{A:A\boldsymbol{u}=A\boldsymbol{u}'\}) > \frac{\alpha_{\mathsf{A}}(n)}{|\mathrm{Im}\mathcal{A}_n|}}} p_{\mathsf{A},n}\left(\{A : A\boldsymbol{u} = A\boldsymbol{u}'\}\right) \leq \beta_{\mathsf{A}}(n) \tag{H3}$$

for any n and $\boldsymbol{u} \in \mathcal{U}^n$. Throughout this paper, we omit the dependence of $\mathcal{A}$, p_{A}, α_{A} and β_{A} on n.

Let us remark on the condition (H3). This condition requires the sum of the collision probabilities $p_{\mathsf{A}}\left(\{A : A\boldsymbol{u} = A\boldsymbol{u}'\}\right)$, which is greater than $\alpha_A/|\mathrm{Im}\mathcal{A}|$, to be bounded by β_A, where the sum is taken over all $\boldsymbol{u}'$ except $\boldsymbol{u}$. It should be noted that this condition implies

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{T} \\ \boldsymbol{u}'\in\mathcal{T}'}} p_{\mathsf{A}}\left(\{A : A\boldsymbol{u} = A\boldsymbol{u}'\}\right) \leq |\mathcal{T}\cap\mathcal{T}'| + \frac{|\mathcal{T}||\mathcal{T}'|\alpha_{\mathsf{A}}}{|\mathrm{Im}\mathcal{A}|} + \min\{|\mathcal{T}|,|\mathcal{T}'|\}\beta_{\mathsf{A}} \tag{H3'}$$

for any $\mathcal{T},\mathcal{T}' \subset \mathcal{U}^n$, which is introduced in [12].

It should be noted that when $\mathcal{A}$ is a two-universal class of hash functions [5] and p_{A} is the uniform distribution on $\mathcal{A}$, then $(\boldsymbol{\mathcal{A}},\boldsymbol{p}_{\mathsf{A}})$ has a $(\boldsymbol{1},\boldsymbol{0})$-hash property, where $\boldsymbol{1}$ and $\boldsymbol{0}$ denote the constant sequences of 1 and 0, respectively. Random bin coding [3] and the set of all linear functions [4] are examples of the two-universal class of hash functions. It is proved in [13, Section III-B] that an ensemble of sparse matrices has a (strong) hash property.

Lemma 2 ([13, Lemma 4]): Let $(\boldsymbol{\mathcal{A}},\boldsymbol{p}_{\mathsf{A}})$ and $(\boldsymbol{\mathcal{B}},\boldsymbol{p}_{\mathsf{B}})$ be ensembles satisfying an $(\boldsymbol{\alpha}_{\mathsf{A}},\boldsymbol{\beta}_{\mathsf{A}})$-hash property and an $(\boldsymbol{\alpha}_{\mathsf{B}},\boldsymbol{\beta}_{\mathsf{B}})$-hash property, respectively. Let $\mathcal{A} \in \boldsymbol{\mathcal{A}}$ (resp. $\mathcal{B} \in \boldsymbol{\mathcal{B}}$) be a set of functions $A : \mathcal{U}^n \to \mathrm{Im}\mathcal{A}$ (resp.

[2] In [13] [14], it is called the 'strong hash property.' Throughout this paper, we call it simply the 'hash property.'

$B : \mathcal{U}^n \to \mathrm{Im}\mathcal{B}$). Let $(A,B) \in \mathcal{A} \times \mathcal{B}$ be a function defined as $(A,B)\boldsymbol{u} \equiv (A\boldsymbol{u}, B\boldsymbol{u})$ for each $\boldsymbol{u} \in \mathcal{U}^n$. Let p_{AB} be a joint distribution on $\mathcal{A} \times \mathcal{B}$ defined as $p_{\mathsf{AB}}(A,B) \equiv p_{\mathsf{A}}(A)p_{\mathsf{B}}(B)$ for each $(A,B) \in \mathcal{A} \times \mathcal{B}$. Then the ensemble $(\boldsymbol{\mathcal{A}} \times \boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{AB}})$ has an $(\boldsymbol{\alpha}_{\mathsf{AB}}, \boldsymbol{\beta}_{\mathsf{AB}})$-hash property, where $(\alpha_{\mathsf{AB}}, \beta_{\mathsf{AB}})$ is defined as $\alpha_{\mathsf{AB}} \equiv \alpha_{\mathsf{A}}\alpha_{\mathsf{B}}$ and $\beta_{\mathsf{AB}} \equiv \beta_{\mathsf{A}} + \beta_{\mathsf{B}}$.

Lemma 3 ([12, Lemma 1]): If $(\mathcal{A}, p_{\mathsf{A}})$ satisfies (H3'), then

$$p_{\mathsf{A}}\left(\{A : [\mathcal{G} \setminus \{\boldsymbol{u}\}] \cap \mathcal{C}_A(A\boldsymbol{u}) \neq \emptyset\}\right) \leq \frac{|\mathcal{G}|\alpha_{\mathsf{A}}}{|\mathrm{Im}\mathcal{A}|} + \beta_{\mathsf{A}}$$

for all $\mathcal{G} \subset \mathcal{U}^n$ and $\boldsymbol{u} \in \mathcal{U}^n$.

Lemma 4 ([14, Lemma 4]): If $(\mathcal{A}, p_{\mathsf{A}})$ satisfies (H3), then

$$E_{\mathsf{A}}\left[\sum_{\boldsymbol{c}} \left| \frac{Q(\mathcal{T} \cap \mathcal{C}_{\mathsf{A}}(\boldsymbol{c}))}{Q(\mathcal{T})} - \frac{1}{|\mathrm{Im}\mathcal{A}|} \right| \right] \leq \sqrt{\alpha_{\mathsf{A}} - 1 + \frac{[\beta_{\mathsf{A}} + 1]|\mathrm{Im}\mathcal{A}| \max_{\boldsymbol{u} \in \mathcal{T}} Q(\boldsymbol{u})}{Q(\mathcal{T})}}$$

for any function $Q : \mathcal{U}^n \to [0, \infty)$ and $\mathcal{T} \subset \mathcal{U}^n$, where $Q(\mathcal{T}) \equiv \sum_{\boldsymbol{u} \in \mathcal{T}} Q(\boldsymbol{u})$.

IV. Construction of Channel Code

This section introduces a channel code. The idea for the construction is drawn from [11][12][14]. It should be noted that we assume that the channel input alphabet $\mathcal{X}^n$ is a finite set but allow the channel output alphabet $\mathcal{Y}^n$ to be an arbitrary (infinite, continuous) set.

For given $r >$ and $R > 0$, let $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}})$ and $(\boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{B}})$ be ensembles of functions $A : \mathcal{X}^n \to \mathrm{Im}\mathcal{A}$ and $B : \mathcal{X}^n \to \mathrm{Im}\mathcal{B}$ satisfying

$$r = \frac{1}{n}\log|\mathrm{Im}\mathcal{A}|$$
$$R = \frac{1}{n}\log|\mathrm{Im}\mathcal{B}|,$$

respectively, where we define $\mathcal{M}_n \equiv \mathrm{Im}\mathcal{B}$ and R represents the rate of the code. We fix functions $A \in \mathcal{A}$, $B \in \mathcal{B}$, and a vector $\boldsymbol{c} \in \mathrm{Im}\mathcal{A}$ so that they are shared by an encoder and a decoder. We assume that the distribution $\mu_{Y^n|X^n}$ of a channel and a channel input distribution μ_{X^n} are given. Let $\mu_{X^n|Y^n}$ be a conditional distribution given as

$$\mu_{X^n|Y^n}(\boldsymbol{x}|\boldsymbol{y}) \equiv \sum_{\boldsymbol{y}} \mu_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\mu_{X^n}(\boldsymbol{x}),$$

where the summention is replaced by the integral when $\mathcal{Y}^n$ is a continuous set.

We use a constraint random number generator to construct an encoder. Let $\widetilde{X}^n \equiv \widetilde{X}^n_{AB}(\boldsymbol{c}, \boldsymbol{m})$ be a random variable corresponding to the distribution

$$\nu_{\widetilde{X}^n}(\boldsymbol{x}) \equiv \begin{cases} \frac{\mu_{X^n}(\boldsymbol{x})}{\mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}, & \text{if } \boldsymbol{x} \in \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}), \\ 0, & \text{if } \boldsymbol{x} \notin \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}), \end{cases}$$

where the eoncder generates $\boldsymbol{x}$ satisfying $A\boldsymbol{x} = \boldsymbol{c}$ and $B\boldsymbol{x} = \boldsymbol{m}$ with probability $\nu_{\widetilde{X}^n}(\boldsymbol{x})$. We define the stochastic encoder $\Phi_n : \mathrm{Im}\mathcal{B} \to \mathcal{X}^n$ as

$$\Phi_n(\boldsymbol{m}) \equiv \begin{cases} \widetilde{X}^n_{AB}(\boldsymbol{c},\boldsymbol{m}) & \text{if } \mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})) > 0 \\ \text{encoding error} & \text{if } \mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})) = 0. \end{cases}$$

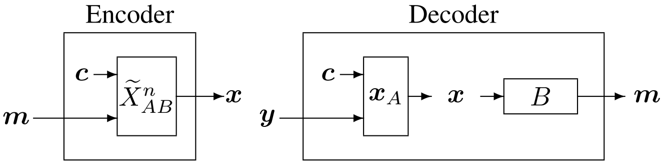

Fig. 1. Construction of Channel Code

Let $\boldsymbol{y} \in \mathcal{Y}^n$ be a channel output. The decoder guesses $\boldsymbol{x}$ satisfying $A\boldsymbol{x} = \boldsymbol{c}$ by using maximum-likelihood decoding and obtains a reproduced message by $\boldsymbol{m} = B\boldsymbol{x}$. We define the decoder $\psi_n : \mathcal{Y}^n \to \mathrm{Im}\mathcal{B}$ as

$$\psi_n(\boldsymbol{y}) \equiv B\boldsymbol{x}_A(\boldsymbol{c}|\boldsymbol{y}),$$

where $\boldsymbol{x}_A$ is defined as

$$\boldsymbol{x}_A(\boldsymbol{c}|\boldsymbol{y}) \equiv \arg\max_{\boldsymbol{x}' \in \mathcal{C}_A(\boldsymbol{c})} \mu_{X^n|Y^n}(\boldsymbol{x}'|\boldsymbol{y}).$$

The flow of vectors is illustrated in Fig. 1. It should be noted that the stochastic encoder is different from the conventional random coding. In the conventional random coding, the construction of a decoder depends on a randomly constructed *deterministic* encoding function. On the othe hand, in the stochastic encoding, the construction of a decoder depends only on A, B, and $\boldsymbol{c}$. It is unnecessary for the decoder to know which $\boldsymbol{x} \in \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})$ is assigned[3] corresponding to a message $\boldsymbol{m}$.

The error probability $\mathrm{Error}(A, B, \boldsymbol{c})$ is given by

$$\mathrm{Error}(A,B,\boldsymbol{c}) \equiv \sum_{\boldsymbol{m}:\mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))=0} \frac{1}{|\mathcal{M}_n|} + \sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y}: \\ \mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0 \\ \boldsymbol{x} \in \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}) \\ \psi_n(\boldsymbol{y}) \neq \boldsymbol{m}}} \frac{\mu_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\mu_{X^n}(\boldsymbol{x})}{|\mathcal{M}_n|\mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}.$$

We have the following theorem, which is shown in Section V-B.

Theorem 1: Assume that for $r, R > 0$ satisfying

$$r > \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) \tag{4}$$
$$r + R < \underline{H}(\boldsymbol{X}) \tag{5}$$

ensembles $(\boldsymbol{\mathcal{A}}, p_{\mathsf{A}})$ and $(\boldsymbol{\mathcal{B}}, p_{\mathsf{B}})$ have a hash property. Then for any $\delta > 0$ and all sufficiently large n there are functions $A \in \mathcal{A}$, $B \in \mathcal{B}$, and a vector $\boldsymbol{c} \in \mathrm{Im}\mathcal{A}$ such that $\mathrm{Error}(A, B, \boldsymbol{c}) \leq \delta$. The channel capacity is achievable with the proposed code by letting $\boldsymbol{X}$ be a source that attains the supremum on the right hand side of (3).

[3] The encoder could select $\boldsymbol{x} \in \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})$ deterministically. In fact, from the random coding argument, there is a good deterministic function which asigns $\boldsymbol{x} \in \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})$ corresponding to a message $\boldsymbol{m}$. It should be noted that we consider a random assignment here.

V. Proofs

A. Proof of Lemma 1

Since $\underline{I}(\boldsymbol{X};\boldsymbol{Y}) \ge \underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})$ for any $(\boldsymbol{X},\boldsymbol{Y})$, we have

$$C(\boldsymbol{W}) \ge \sup_{\boldsymbol{X}} \left[\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})\right].$$

In the following, we prove that

$$C(\boldsymbol{W}) \le \sup_{\boldsymbol{X}} \left[\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})\right], \tag{6}$$

which completes the proof of the lemma.

From the definition of $C(\boldsymbol{W})$, we have the fact that for any $\delta > 0$ and sufficiently large n there is a pair consisting an encoder $\varphi_n : \mathcal{M}_n \to \mathcal{X}^n$ and a decoder $\psi_n : \mathcal{Y}^n \to \mathcal{M}_n$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{M}_n| \ge C(\boldsymbol{W}) - \delta \tag{7}$$

$$\lim_{n\to\infty} P(\psi_n(Y^n) \neq M_n) = 0. \tag{8}$$

We can assume[4] that $\mathcal{M}_n \subset \mathcal{X}^n$ without loss of generality. Since the distribution μ_{M_n} of M_n is uniform on $\mathcal{M}_n$, we have the fact that

$$\begin{aligned}\frac{1}{n}\log\frac{1}{\mu_{M_n}(\boldsymbol{x})} &= \frac{1}{n}\log|\mathcal{M}_n| \\ &\ge \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{M}_n| - \delta\end{aligned} \tag{9}$$

for all $\boldsymbol{x} \in \mathcal{M}_n$, $\delta > 0$, and sufficiently large n. Since

$$\frac{1}{n}\log\frac{1}{\mu_{M_n}(\boldsymbol{x})} = \infty$$

for every $\boldsymbol{x} \notin \mathcal{M}_n$, we have the fact that

$$\frac{1}{n}\log\frac{1}{\mu_{M_n}(\boldsymbol{x})} \ge \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{M}_n| - \delta$$

for every $\boldsymbol{x} \in \mathcal{X}^n$, $\delta > 0$ and sufficiently large n. This implies that

$$\lim_{n\to\infty} P\left(\frac{1}{n}\log\frac{1}{\mu_{M_n}(M_n)} < \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{M}_n| - \delta\right) = 0. \tag{10}$$

Let $\boldsymbol{M} \equiv \{M_n\}_{n=1}^{\infty}$ be a general source. Then we have

$$\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{M}_n| - \delta \le \underline{H}(\boldsymbol{M}) \tag{11}$$

from (10) and the definition of $\underline{H}(\boldsymbol{M})$. We have

$$\begin{aligned}C(\boldsymbol{W}) &\le \liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{M}_n| + \delta \\ &\le \underline{H}(\boldsymbol{M}) + 2\delta \\ &= \underline{H}(\boldsymbol{M}) - \overline{H}(\boldsymbol{M}|\boldsymbol{Y}) + 2\delta \\ &\le \sup_{\boldsymbol{X}}\left[\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})\right] + 2\delta,\end{aligned} \tag{12}$$

where the first inequality comes from (7), the second inequality comes from (11), and the equality comes from the fact that $\overline{H}(\boldsymbol{X}|\boldsymbol{Y}) = 0$ obtained from (8). We have (6) by letting $\delta \to 0$. ■

B. Proof of Theorem 1

We omit dependence on n of X and Y when they appear in the subscript of μ.

From (4) and (5), we have the fact that there is $\varepsilon > 0$ satisfying

$$r > \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) + \varepsilon \tag{13}$$

$$r + R < \underline{H}(\boldsymbol{X}) - \varepsilon. \tag{14}$$

Let $\underline{\mathcal{T}}_X \subset \mathcal{X}^n$ and $\overline{\mathcal{T}}_{X|Y} \subset \mathcal{X}^n \times \mathcal{Y}^n$ be defined as

$$\underline{\mathcal{T}}_X \equiv \left\{\boldsymbol{x} : \frac{1}{n}\log\frac{1}{\mu_X(\boldsymbol{x})} \ge \underline{H}(\boldsymbol{X}) - \varepsilon\right\}$$

$$\overline{\mathcal{T}}_{X|Y} \equiv \left\{(\boldsymbol{x},\boldsymbol{y}) : \frac{1}{n}\log\frac{1}{\mu_{X|Y}(\boldsymbol{x}|\boldsymbol{y})} \le \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) + \varepsilon\right\}.$$

Assume that $(\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}$ and $\boldsymbol{x}_A(A\boldsymbol{x}|\boldsymbol{y}) \neq \boldsymbol{x}$. Then we have the fact that there is $\boldsymbol{x}' \in \mathcal{C}_A(A\boldsymbol{x})$ such that $\boldsymbol{x}' \neq \boldsymbol{x}$ and

$$\mu_{X|Y}(\boldsymbol{x}'|\boldsymbol{y}) \ge \mu_{X|Y}(\boldsymbol{x}|\boldsymbol{y}) \ge 2^{-n[\overline{H}(\boldsymbol{X}|\boldsymbol{Y})+\varepsilon]}.$$

This implies that $\left[\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y}) \setminus \{\boldsymbol{x}\}\right] \cap \mathcal{C}_A(A\boldsymbol{x}) \neq \emptyset$, where $\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y}) \equiv \left\{\boldsymbol{x} : (\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}\right\}$. We have

$$\begin{aligned}&E_{\mathsf{A}}\left[\chi(\boldsymbol{x}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}|\boldsymbol{y}) \neq \boldsymbol{x})\right] \\ &\le p_{\mathsf{A}}\left(\left\{A : [\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y}) \setminus \{\boldsymbol{x}\}] \cap \mathcal{C}_A(A\boldsymbol{x}) \neq \emptyset\right\}\right) \\ &\le \frac{|\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\alpha_{\mathsf{A}}}{|\mathrm{Im}\mathcal{A}|} + \beta_{\mathsf{A}} \\ &\le 2^{-n[r-\overline{H}(\boldsymbol{X}|\boldsymbol{Y})-\varepsilon]}\alpha_{\mathsf{A}} + \beta_{\mathsf{A}}\end{aligned} \tag{16}$$

for all $(\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}$, where $\chi(\cdot)$ is the indicator function, the second inequality comes from Lemma 3, and the third inequality comes from the fact that $|\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y})| \le 2^{n[\overline{H}(\boldsymbol{X}|\boldsymbol{Y})+\varepsilon]}$. We have the fact that

$$\begin{aligned}&E_{\mathsf{A}}\left[\sum_{\boldsymbol{x},\boldsymbol{y}}\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\chi(\boldsymbol{x}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}|\boldsymbol{y}) \neq \boldsymbol{x})\right] \\ &= \sum_{(\boldsymbol{x},\boldsymbol{y})\in\overline{\mathcal{T}}_{X|Y}}\mu_{XY}(\boldsymbol{x},\boldsymbol{y})E_{\mathsf{A}}\left[\chi(\boldsymbol{x}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}|\boldsymbol{y}) \neq \boldsymbol{x})\right] \\ &\quad + \sum_{(\boldsymbol{x},\boldsymbol{y})\notin\overline{\mathcal{T}}_{X|Y}}\mu_{XY}(\boldsymbol{x},\boldsymbol{y})E_{\mathsf{A}}\left[\chi(\boldsymbol{x}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}|\boldsymbol{y}) \neq \boldsymbol{x})\right] \\ &\le 2^{-n[r-\overline{H}(\boldsymbol{X}|\boldsymbol{Y})-\varepsilon]}\alpha_{\mathsf{A}} + \beta_{\mathsf{A}} + \mu_{XY}([\overline{\mathcal{T}}_{X|Y}]^c),\end{aligned} \tag{17}$$

where the last inequality comes from (16). We also have the fact that

$$\begin{aligned}&E_{\mathsf{AB}}\left[\sum_{\boldsymbol{c},\boldsymbol{m}}\left|\mu_X(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m})) - \frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\right|\right] \\ &\le E_{\mathsf{AB}}\left[\sum_{\boldsymbol{c},\boldsymbol{m}}\left|\mu_X(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m}) \cap \underline{\mathcal{T}}_X) - \frac{\mu_X(\underline{\mathcal{T}}_X)}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\right|\right] \\ &\quad + E_{\mathsf{AB}}\left[\sum_{\boldsymbol{c},\boldsymbol{m}}\left[\mu_X(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m}) \cap [\underline{\mathcal{T}}_X]^c) + \frac{\mu_X([\underline{\mathcal{T}}_X]^c)}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\right]\right]\end{aligned}$$

[4] This assumption is used merely so that $\boldsymbol{M} \equiv \{M_n\}_{n=1}^{\infty}$ is a general source satisfying $M_n \in \mathcal{X}^n$. It should be noted that $\mathcal{M}_n$ and $\{\varphi_n(\boldsymbol{m}) : \boldsymbol{m} \in \mathcal{M}_n\}$ are different subsets of $\mathcal{X}^n$ in general. We could define a channel code by a subset $\mathcal{M}_n$ of $\mathcal{X}^n$ as defined in [9][18] instead of introducing an encoder φ_n. We introduce an encoder φ_n to consider a stochastic encoder.

$$
\begin{aligned}
&E_{\mathsf{AB}\mathbf{c}}\left[\mathrm{Error}(\mathsf{A},\mathsf{B},\mathbf{c})\right]\\
&= E_{\mathsf{AB}}\left[\sum_{\substack{\boldsymbol{c},\boldsymbol{m}:\\ \mu_{X^n}(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m}))=0}} \frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|} + \sum_{\substack{\boldsymbol{c},\boldsymbol{m},\boldsymbol{x},\boldsymbol{y}:\\ \mu_{X^n}(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m}))>0\\ \boldsymbol{x}\in\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m})\\ \boldsymbol{x}_{\mathsf{A}}(\boldsymbol{c}|\boldsymbol{y})\neq\boldsymbol{x}}} \mu_{XY}(\boldsymbol{x},\boldsymbol{y})\left[1+\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m}))}-1\right]\right]\\
&\leq E_{\mathsf{A}}\left[\sum_{\boldsymbol{x},\boldsymbol{y}}\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\chi(\boldsymbol{x}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}|\boldsymbol{y})\neq\boldsymbol{x})\right] + E_{\mathsf{AB}}\left[\sum_{\boldsymbol{c},\boldsymbol{m}}\left|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))-\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\right|\right]\\
&\leq 2^{-n[r-\overline{H}(\boldsymbol{X}|\boldsymbol{Y})-\varepsilon]}\alpha_{\mathsf{A}}+\beta_{\mathsf{A}}+\mu_{XY}([\overline{\mathcal{T}}_{X|Y}]^c)+\sqrt{\alpha_{\mathsf{AB}}-1+[\beta_{\mathsf{AB}}+1]2^{-n[\underline{H}(\boldsymbol{X})-r-R-\varepsilon]}}+2\mu_X([\underline{\mathcal{T}}_X]^c) \qquad (19)
\end{aligned}
$$

$$
\begin{aligned}
&= \mu_X(\underline{\mathcal{T}}_X)E_{\mathsf{AB}}\left[\sum_{\boldsymbol{c},\boldsymbol{m}}\left|\frac{\mu_X(\mathcal{C}_{\mathsf{AB}}(\boldsymbol{c},\boldsymbol{m})\cap\underline{\mathcal{T}}_X)}{\mu_X(\underline{\mathcal{T}}_X)}-\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\right|\right]\\
&\quad + 2\mu_X([\underline{\mathcal{T}}_X]^c)\\
&\leq \mu_X(\underline{\mathcal{T}}_X)\sqrt{\frac{\alpha_{\mathsf{AB}}-1+[\beta_{\mathsf{AB}}+1]|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|\max_{\boldsymbol{x}\in\underline{\mathcal{T}}_X}\mu_X(\boldsymbol{x})}{\mu_X(\underline{\mathcal{T}}_X)}}\\
&\quad + 2\mu_X([\underline{\mathcal{T}}_X]^c)\\
&\leq \sqrt{\alpha_{\mathsf{AB}}-1+[\beta_{\mathsf{AB}}+1]2^{-n[\underline{H}(\boldsymbol{X})-r-R-\varepsilon]}}+2\mu_X([\underline{\mathcal{T}}_X]^c), \qquad (18)
\end{aligned}
$$

where the second inequality comes from Lemmas 2, 4. Then we have (19), which appears on the top of this page, where $\mathbf{c}$ is a random variable corresponding to the uniform distribution on $\mathrm{Im}\mathcal{A}$, the first inequality comes from the fact that

$$
\begin{aligned}
&\sum_{\substack{\boldsymbol{c},\boldsymbol{m},\boldsymbol{x},\boldsymbol{y}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0\\ \boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})}} \mu_{XY}(\boldsymbol{x},\boldsymbol{y})\\
&\quad\cdot\left[\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}-1\right]\\
&\leq \sum_{\substack{\boldsymbol{c},\boldsymbol{m}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0}}\left|\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}-1\right|\\
&\quad\cdot\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))\\
&= \sum_{\boldsymbol{c},\boldsymbol{m}}\left|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))-\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\right|\\
&\quad - \sum_{\substack{\boldsymbol{c},\boldsymbol{m}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))=0}}\frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}, \qquad (20)
\end{aligned}
$$

and the second inequality comes from (17), (18). From (13), (14), (19) and the fact that $\alpha_{\mathsf{A}}\to 1$, $\beta_{\mathsf{A}}\to 0$, $\alpha_{\mathsf{AB}}\to 1$, $\beta_{\mathsf{AB}}\to 0$, $\mu_X([\underline{\mathcal{T}}_X]^c)\to 0$, $\mu_{XY}([\underline{\mathcal{T}}_{XY}]^c)\to 0$ as $n\to\infty$, we have the fact that there are functions $A\in\mathcal{A}$, $B\in\mathcal{B}$, and a vector $\boldsymbol{c}\in\mathrm{Im}\mathcal{A}$ satisfying $\mathrm{Error}(A,B,\boldsymbol{c})\leq\delta$. ■

Acknowledgements

The author thanks Dr. S. Miyake, Prof. K. Iwata, and Prof. T. Ogawa for helpful discussions.

References

[1] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 325–343, Mar. 2000.

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

[3] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.

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

[5] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *J. Comput. Syst. Sci.*, vol. 18, pp. 143–154, 1979.

[6] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., 1968.

[7] R. G. Gallager, *Low Density Parity Check Codes*, Cambridge, MA:M.I.T Press, 1963.

[8] T.S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. IT-39, no. May, pp. 752–772, May 1993.

[9] T.S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2003.

[10] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[11] J. Muramatsu, T. Uyematsu, and T. Wadayama, "Low density parity check matrices for coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 10, pp. 3645–3653, Oct. 2005.

[12] J. Muramatsu and S. Miyake, "Hash property and coding theorems for sparse matrices and maximal-likelihood coding," *IEEE Trans. Inform. Theory*, vol. IT-56, no. 5, pp. 2143–2167, May 2010. Corrections: vol. IT-56, no. 9, p. 4762, Sep. 2010.

[13] J. Muramatsu and S. Miyake, "Construction of Slepian-Wolf source code and broadcast channel code based on hash property," available at `arXiv:1006.5271[cs.IT]`, 2010.

[14] J. Muramatsu and S. Miyake, "Construction of strongly secure wiretap channel code based on hash property," *Proc. of 2011 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 31–Aug. 5, 2011, pp. 612–616.

[15] J. Muramatsu, "Algorithms for constrained random number generation," submitted to *Proc. of 2013 IEEE Information Theory Workshop*, 2013.

[16] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inform. Theory*, vol. IT-57, pp. 7377–7385, Nov. 2011.

[17] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[18] S. Verdú and T.S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. IT-40, no. 4, pp. 1147–1157, Jul. 1994.

[19] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *Proc. 2012 IEEE Int. Symp. Inform. Theory*, Cambridge, MA, USA, Jul. 1–6, 2012, pp. 1049–1053.