Channel Code Using Constrained-Random-Number Generator Revisited

Jun Muramatsu and Shigeki Miyake

***Abstract*— A construction of a channel code by using a source code with decoder side information is introduced. The encoder and decoder pair of any source code can be used for the construction. Constrained-random-number generators, which generate random numbers satisfying a condition specified by a function and its value, are used to construct stochastic encoders and decoders. The result suggests that we can divide the channel coding problem into the problems of channel encoding and source decoding with side information.**

***Index Terms*— Shannon theory, channel coding, source code with decoder side information, constrained-random-number generator.**

I. Introduction

THIS paper revisits the channel code using constrained-random-number generator introduced in [28] from the viewpoint of the source code with decoder side information. Our contributions are summarized as follows.

- It is shown that we can construct a stochastic decoder for a source code with decoder side information by using a constrained-random-number generator.
- It is shown that we can construct a channel code (Fig. 1) from a given source code of $\boldsymbol{X}$ with decoder side information $\boldsymbol{Y}$ (Fig. 2), where the channel input and output are given by $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.
- We can construct a code that achieves the capacity by letting $\boldsymbol{X}$ be an optimum channel input random variable and using a source code achieving the limit $\overline{H}(\boldsymbol{X}|\boldsymbol{Y})$. It should be noted that [28] shows only the fact that there are a pair of functions with which the code achieves the capacity.
- The above facts imply that both encoding and decoding functions of a channel code can be constructed by using constrained-random-number generators. It should be noted that, by assuming that a channel is memoryless, we can use the sum-product algorithm or the Markov-Chain-Monte-Carlo method to implement a tractable constrained-random-number generator [28], [29], where 'tractable' means that there is an iterative approximation algorithm with polynomial computational complexity with respect to the block length by assuming the constant number of iterations.

Manuscript received April 16, 2017; revised July 23, 2018; accepted September 24, 2018. Date of publication October 26, 2018; date of current version December 19, 2018. This paper was presented in part at the 2016 International Symposium on Information Thoeory and its Applications (ISITA2016) and in part at the 2017 IEEE International Symposium on Information Thoeory (ISIT2017).

J. Muramatsu is with NTT Communication Science Laboratories, NTT Corporation, Kyoto 619-0237, Japan (e-mail: muramatsu.jun@lab.ntt.co.jp).

S. Miyake is with NTT Network Innovation Laboratories, NTT Corporation, Kanagawa 239-0847, Japan (e-mail: miyake.shigeki@lab.ntt.co.jp).

Communicated by V. Vaishampayan, Associate Editor for Source Coding.
Digital Object Identifier 10.1109/TIT.2018.2878217

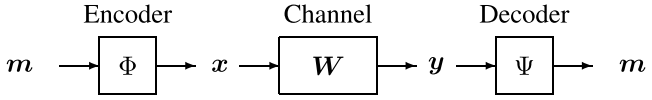

Fig. 1. Channel Coding: An encoder sends a codeword $\boldsymbol{x}$ obtained from a message $\boldsymbol{m}$ by using a (possibly stochastic) encoding function Φ. A decoder receives an output $\boldsymbol{y}$ of a channel $\boldsymbol{W}$ and reproduces $\boldsymbol{m}$ from $\boldsymbol{y}$ by using a (possibly stochastic) decoding function Ψ.

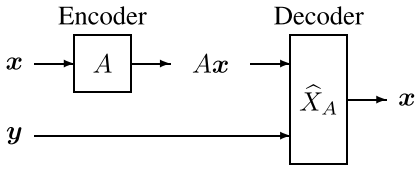

Fig. 2. Source Coding with Decoder Side Information: An encoder sends a codeword $A\boldsymbol{x}$ obtained from a source output $\boldsymbol{x}$ by using an encoding function A. A decoder reproduces $\boldsymbol{x}$ from the codeword and side information $\boldsymbol{y}$ by using a (possibly stochastic) decoding function $\widehat{X}_A$.

A. Relation to Previous Results

The source coding with decoder side information is a special case of the distributed coding of correlated sources introduced by Slepian and Wolf [41]. Let (X^n, Y^n) be a pair of correlated sources. We consider a source code where an encoder transmits a codeword obtained from a source output X^n and a decoder reproduces X^n from the codeword and the side information Y^n, where it is expected that the decoding error probability is close to zero. From the Slepian-Wolf theorem [41], the asymptotically optimum encoding rate for stationary memoryless sources is given by the conditional entropy $H(X|Y)$. The result is extended to general correlated sources $(\boldsymbol{X}, \boldsymbol{Y})$ in [25] and [43], where conditions such as stationarity and ergodicity are not assumed and the fundamental limit is given by the conditional spectral sup-entropy rate $\overline{H}(\boldsymbol{X}|\boldsymbol{Y})$. This paper considers a pair of general correlated sources $(\boldsymbol{X}, \boldsymbol{Y})$, and the results can be applied to the stationary memoryless case.

Historically, the Slepian-Wolf codes are constructed by using channel codes. In [41], the code is given by using a set of randomly-generated channel codewords that covers the conditionally typical set of X for a given Y. Wyner [45] introduced the Slepian-Wolf code by using parity check matrices, where it is shown by Elias [13] that the capacity of a binary symmetric channel is achievable by using a linear or a convolutional code. In accordance with this idea, the Slepian-Wolf codes are constructed in [3], [17], [21], and [38] from turbo codes [5],

polar codes [2], and low density parity check (LDPC) codes [16] with practical decoding algorithms, where the correlation of two sources are assumed to be binary-symmetric. Codes for an asymmetric channel can be constructed by using the channel-input alphabet extension [4], [15, Sec. 6.2], the chaining construction [26], or polar codes [20].

On the other hand, Cover [9] introduced the random binning method for constructing the Slepian-Wolf code, where conditions such as symmetric correlations are not assumed for two sources. Following this idea, Csiszár [11] proved that the fundamental limit is achievable by using a linear code. In [37], it is proved that the fundamental limit is achievable by using an LDPC code. These results are unified by introducing the notion of hash property [30], [31], which is the extension of the 2-universal class of hash functions [7]. It should be noted that the use of a typical-set decoder or a maximum-likelihood decoder is assumed in these results.

Based on the concept of hash functions, in this paper we adopt an approach where we construct a channel code from a source code with decoder side information. This approach was originated from [27, Sec. III], [37, Th. 4] and is categorized in the integrated scheme [26]. A similar approach is investigated in the context of the linear codebook-level duality of channel codes and the Slepian-Wolf codes [8], where the symmetric correlation of two sources (channel input and output) is assumed. This paper does not assume such correlations. It should be noted that this approach is investigated in [24], [28], [30], and [46], where these papers prove that there is a pair consisting of a source code with decoder side information and a encoding map to construct a channel code. However, a maximal-likelihood decoder is assumed, and it is unknown whether for an arbitrary given source code with decoder side information there is a good encoding map with which to construct a channel code. In [37], it is proved by assuming a stationary memoryless condition that for a given arbitrary linear source code with decoder side information there is a good encoding map with which to construct a channel code, where the encoding map is intractable. In contrast, this paper introduces a tractable encoding map by using a constrained-random-number generator [28]. We can use *any* source code with decoder side information, where it is confirmed theoretically or empirically that the decoding error probability is small. Neither a typical-set decoder nor a maximal-likelihood decoder is assumed for the source code with decoder side information. Our result suggests that we can divide channel coding problem into the problems of channel encoding and source decoding with decoder side information. It should be noted that the similar results have been appeared in [39, Remark 2], [46] when the output distribution of the encoder with side information is close to a uniform distribution. In contrast, this paper clarifies that such an assumption is unnecessary.

B. Paper Outline

This paper is organized as follows. In Section II, we review the constrained-random-number generator introduced in [28]. In Section III, we construct a source code with decoder side information, where a stochastic decoder is constructed by using a constrained-random-number generator. In Section IV, we introduce the construction of a channel code by using an arbitrary source code with decoder side information, where a stochastic encoder is introduced by using another constrained-random-number generator. Based on these results, we show that the channel capacity is achievable with these codes using constrained-random-number generators. Proof of theorems is presented in Section V.

C. Definitions and Notations

Throughout this paper, we use the following definitions and notations. The complement of $\mathcal{U}$ is denoted by $\mathcal{U}^c$ and the set difference is defined as $\mathcal{U}\setminus\mathcal{V} \equiv \mathcal{U}\cap\mathcal{V}^c$. Let $F\boldsymbol{x}$ denote a value taken by a function F at $\boldsymbol{x}\in\mathcal{X}^n$, where F may be nonlinear. When F is a linear function expressed by an $l\times n$ matrix, we assume that $\mathcal{X}\equiv \mathrm{GF}(q)$ is a finite field and the range of functions is $\mathcal{X}^l$. For a function F and a set $\mathcal{F}$ of functions, let $\mathrm{Im}F$ and $\mathrm{Im}\mathcal{F}$ be defined as

$$\mathrm{Im}F \equiv \{F\boldsymbol{x} : \boldsymbol{x}\in\mathcal{X}^n\}$$
$$\mathrm{Im}\mathcal{F} \equiv \bigcup_{F\in\mathcal{F}} \mathrm{Im}F.$$

We define a set $\mathcal{C}_F(\boldsymbol{v})$ as

$$\mathcal{C}_F(\boldsymbol{v}) \equiv \{\boldsymbol{x} : F\boldsymbol{x}=\boldsymbol{v}\}.$$

The random variables of a function F and a vector $\boldsymbol{v}$ are denoted by the sans serif letters F and $\mathbf{v}$, respectively. It should be noted that the random variable of an n-dimensional vector $\boldsymbol{x}\in\mathcal{X}^n$ is denoted by the Roman letter X^n that does not represent a function. The symbol E denotes the expectation. For example, $E_{\mathsf{F}\mathbf{v}}[\cdot]$ denotes the expectation with respect to random variables F and $\mathbf{v}$.

Let $\mathcal{F}_n$ be a set of functions on $\mathcal{X}^n$ and $p_{\mathsf{F},n}$ be a probability distribution on $\mathcal{F}_n$. We call a pair $(\mathcal{F}_n, p_{\mathsf{F},n})$ an *ensemble*. We sometimes omit the dependence of $\mathcal{F}$ and p_{F} on n.

All the results in this paper are presented by using the information spectrum method introduced in [18], [19], and [44], where the consistency and stationarity are not assumed.

II. Constrained-Random-Number Generator

Here, we review the constrained-random-number generator introduced in [28]. It is used in the construction of stochastic encoder and decoder.

For a given probability distribution μ_{X^n} of X^n, a constrained-random-number generator generates random sequence $\check{X}^n\in\mathcal{X}^n$ subject to a distribution

$$\mu_{\check{X}^n|V_n}(\boldsymbol{x}|\boldsymbol{v}) \equiv \frac{\mu_{X^n}(\boldsymbol{x})\chi(F\boldsymbol{x}=\boldsymbol{v})}{\sum_{\boldsymbol{x}'}\mu_{X^n}(\boldsymbol{x}')\chi(F\boldsymbol{x}'=\boldsymbol{v})} = \frac{\mu_{X^n}(\boldsymbol{x})\chi(F\boldsymbol{x}=\boldsymbol{v})}{\mu_{X^n}(\mathcal{C}_F(\boldsymbol{v}))} \tag{1}$$

for a given function F on $\mathcal{X}^n$ and a vector $\boldsymbol{v}\in\mathrm{Im}F$, where $\chi(\cdot)$ is a support function defined as

$$\chi(\mathrm{S}) \equiv \begin{cases} 1, & \text{if the statement S is true} \\ 0, & \text{if the statement S is false.} \end{cases} \tag{2}$$

The constrained-random-number generator generates $\boldsymbol{x}$ that satisfies $F\boldsymbol{x}=\boldsymbol{v}$ with probability $\mu_{\check{X}^n|V_n}(\boldsymbol{x}|\boldsymbol{v})$.

When F is a q-ary matrix and X^n is memoryless, we can use the sum-product algorithm or the Markov-Chain-Monte-Carlo method to implement the constrained-random-number generator [28], [29]. When we use the sum-product algorithm, the computational complexity is $O(\iota[n-l]lw_{\mathrm{row}}^2 q)$, where F is assumed to be an $l \times n$-ary matrix and ι denotes the number of iterations of the sum-product algorithm and w_{row} denotes the maximum row weight of F [29].[1] When we use the Markov-Chain-Monte-Carlo method, the computational complexity is $O(\kappa w_{\mathrm{col}} q)$, where κ denotes the number of iterations of Markov chain and w_{col} denotes the maximum column weight of F [29]. It should be noted that κ depends on the size and density of F to obtain good approximation.

III. Source Code With Decoder Side Information

In this section, we consider a source code with decoder side information illustrated in Fig. 2. The fundamental limit for this problem is given as the conditional spectral sup-entropy rate $\overline{H}(\boldsymbol{X}|\boldsymbol{Y})$ for a general source $(\boldsymbol{X}, \boldsymbol{Y})$, which is specified by a sequence $\{\mu_{X^n Y^n}\}_{n=1}^{\infty}$ of joint probability distributions.

The achievability of this problem is proved via the Slepian-Wolf theorem using random binning [9], [25], [43] or the ensemble of all q-ary matrices [11]. The construction of an encoder using sparse matrix is studied in [38] and [40] and the achievability is proved in [31] and [37] by using a maximum-likelihood or minimum-divergence decoding. We obtain the coding theorem based on the collision-resistance property as a corollary of [30, Th. 7] for stationary memoryless sources.

A. Code Construction

First, we construct a source code with decoder side information using the constrained-random-number generator introduced in [28]. The construction is analogous to the syndrome encoding/decoding when an encoder is a linear function.

We assume that the alphabet $\mathcal{X}^n$ of X^n is a finite set but allow the alphabet $\mathcal{Y}^n$ of Y^n to be an arbitrary (infinite, continuous) set. For a given encoding rate r, let $(\mathcal{A}, p_{\mathsf{A}})$ be an ensemble of functions on the set $\mathcal{X}^n$ satisfying

$$r = \frac{1}{n} \log |\mathrm{Im}\mathcal{A}|. \tag{3}$$

We fix an encoding function $A : \mathcal{X}^n \to \mathrm{Im}\mathcal{A}$ generated at random subject to the distribution p_{A}. The codeword $\boldsymbol{c}$ of $\boldsymbol{x} \in \mathcal{X}^n$ is given as $\boldsymbol{c} \equiv A\boldsymbol{x}$.

Here, we use a constrained-random-number generator to construct a stochastic decoder $\widehat{X}_A^n : \mathrm{Im}A \times \mathcal{Y}^n \to \mathcal{X}^n$. Let $C_n \equiv AX^n$. For given codeword $\boldsymbol{c} \in \mathrm{Im}A$ and side information $\boldsymbol{y} \in \mathcal{Y}^n$, the reproduction $\widehat{X}^n \equiv \widehat{X}_A^n(\boldsymbol{c}, \boldsymbol{y}) \in \mathcal{X}^n$ is determined at random subject to the distribution

$$\mu_{\widehat{X}^n|C_n Y^n}(\widehat{\boldsymbol{x}}|\boldsymbol{c}, \boldsymbol{y}) \equiv \frac{\mu_{X^n|Y^n}(\widehat{\boldsymbol{x}}|\boldsymbol{y})\chi(A\widehat{\boldsymbol{x}} = \boldsymbol{c})}{\mu_{X^n|Y^n}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})}, \tag{4}$$

where $\mu_{X^n|Y^n}$ be defined as

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

This constrained-random-number generator generates $\widehat{\boldsymbol{x}}$ that satisfies $A\widehat{\boldsymbol{x}} = \boldsymbol{c}$ with probability $\mu_{\widehat{X}^n|C_n Y^n}(\widehat{\boldsymbol{x}}|\boldsymbol{c}, \boldsymbol{y})$.

The decoding error probability $\mathrm{Error}(A)$ is given as

$$\mathrm{Error}(A) \equiv \sum_{\substack{\boldsymbol{x}, \boldsymbol{y}, \widehat{\boldsymbol{x}} \\ \widehat{\boldsymbol{x}} \neq \boldsymbol{x}}} \mu_{\widehat{X}^n|C_n Y^n}(\widehat{\boldsymbol{x}}|A\boldsymbol{x}, \boldsymbol{y}) \mu_{X^n Y^n}(\boldsymbol{x}, \boldsymbol{y}). \tag{5}$$

B. $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-Collision-Resistance Property

To state the theorem, we introduce a variant of the hash property [28], [30]. We revisit [30, Remark 1], which mentions that some ensembles of sparse matrices satisfy the weaker condition $\lim_{n\to\infty}[1/n] \log \alpha_A(n) = 0$. We introduce the collision-resistance property as follows.

Definition 1: Let $\mathcal{A}_n$ be a set of functions on $\mathcal{X}^n$ and $p_{\mathsf{A},n}$ be a probability distribution on $\mathcal{A}_n$. Then a sequence $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}}) \equiv \{(\mathcal{A}_n, p_{\mathsf{A},n})\}_{n=1}^{\infty}$ has an $(\boldsymbol{\alpha}_{\mathsf{A}}, \boldsymbol{\beta}_{\mathsf{A}})$-*collision-resistance 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

$$\limsup_{n\to\infty} \frac{1}{n} \log \alpha_{\mathsf{A}}(n) = 0 \tag{CR1}$$

$$\limsup_{n\to\infty} \beta_{\mathsf{A}}(n) = 0 \tag{CR2}$$

and

$$\sum_{\substack{\boldsymbol{x}' \in \mathcal{X}^n \setminus \{\boldsymbol{x}\}: \\ p_{\mathsf{A},n}(\{A : A\boldsymbol{x} = A\boldsymbol{x}'\}) > \frac{\alpha_{\mathsf{A}}(n)}{|\mathrm{Im}\mathcal{A}_n|}}} p_{\mathsf{A},n}\left(\{A : A\boldsymbol{x} = A\boldsymbol{x}'\}\right) \le \beta_{\mathsf{A}}(n) \tag{CR3}$$

for all sufficiently large n and all $\boldsymbol{x} \in \mathcal{X}^n$. In the following, we omit the dependence of α_{A}, and β_{A} on n.

Here, let us introduce examples satisfying the collision-resistance property. When $\mathcal{A}$ is a two-universal class of hash functions [7] and p_{A} is the uniform distribution on $\mathcal{A}$, then $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}})$ has a $(\mathbf{1}, \mathbf{0})$-collision-resistance property, where $\mathbf{1}$ and $\mathbf{0}$ denote the constant sequences of 1 and 0, respectively. Random binning [9] and a set of all linear functions [11] are examples of the two-universal class of hash functions. Since an ensemble of sparse matrices has a hash property [28, Sec. III-B], we have the fact that this ensemble also has a collision-resistance property. Furthermore, we can show that some expurgated ensembles of sparse matrices have a collision-resistance property [4], [14].

The following lemma is related to the collision-resistance property, that is, if the number of bins is greater than the number of items then there is an assignment such that every bin contains at most one item.

Lemma 1 ([28, Lemma 4], [30, Lemma 1]): If $(\mathcal{A}, p_{\mathsf{A}})$ satisfies (CR3), then

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

for all $\mathcal{G} \subset \mathcal{X}^n$ and $\boldsymbol{x} \in \mathcal{X}^n$.

[1] In [29], the computational complexity is given as $O(\iota[n-l]lw_{\mathrm{row}}^2 q^{w_{\mathrm{row}}})$. However, we can reduce the computational complexity by using (Fast-) Fourier-Transform to compute convolutions in the sum-product algorithm. We assume $q \le w_{\mathrm{row}}$ when q is a prime number and $\log_p q \le w_{\mathrm{row}}$ when q is a power of a prime number p to obtain the computational complexity $O(\iota[n-l]lw_{\mathrm{row}}^2 q)$.

C. Existence of Good Function

Now, we are in position to state the theorem, where the proof is given in Section V-A. It should be noted that $\mathcal{Y}$ is allowed to be an infinite/continuous set and the correlation of the two sources is allowed to be asymmetric.

Theorem 1: Let $(\boldsymbol{X}, \boldsymbol{Y})$ be a pair of correlated general sources. Assume that an ensemble $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}})$ has an $(\boldsymbol{\alpha}_{\mathsf{A}}, \boldsymbol{\beta}_{\mathsf{A}})$-collision-resistance property for a given r satisfying (3) and

$$r > \overline{H}(\boldsymbol{X}|\boldsymbol{Y}). \tag{6}$$

Then for any $\delta > 0$ and all sufficiently large n there is a function (sparse matrix) $A \in \mathcal{A}$ such that

$$\mathrm{Error}(A) \le \delta.$$

Remark 1: We can use the maximum a posteriori probability decoder $\widehat{\boldsymbol{x}}_A : \mathrm{Im}\mathcal{A} \times \mathcal{Y}^n \to \mathcal{X}^n$ defined as

$$\widehat{\boldsymbol{x}}_A(\boldsymbol{c}, \boldsymbol{y}) \equiv \arg \max_{\boldsymbol{x} \in \mathcal{C}_A(\boldsymbol{c})} \mu_{X^n|Y^n}(\boldsymbol{x}|\boldsymbol{y}) \tag{7}$$

instead of the stochastic decoder in the above coding scheme. When the maximum a posteriori probability decoder is used, the theorem is implicitly proved in [28, eq. (58)] from Lemma 1. Although the maximum a posteriori probability decoder minimizes the decoding error probability, it may be intractable. On the other hand, our theorem asserts that the stochastic decoding by using the constrained-random-number generator is *sufficient* and *tractable* to achieve the fundamental limit.

IV. Channel Code by Using Arbitrary Source Code With Decoder Side Information

In this section, we construct a channel code by using an arbitrary source code with decoder side information. Because the theorem depends only on the performance of a source code with decoder side information but does not depend on its construction, it is not mandatory to use the code introduced in the previous section.

A. Construction of Channel Code

This section introduces a channel code. The idea for the construction is drawn from [30], [33], and [37]. 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.

We consider a general source and a general channel. A general source $\boldsymbol{X}$ is defined by a sequence $\boldsymbol{X} \equiv \{\mu_{X^n}\}_{n=1}^{\infty}$ of probability distributions and a general channel is defined by a sequence $\boldsymbol{W} \equiv \{\mu_{Y^n|X^n}\}_{n=1}^{\infty}$ of conditional probability distributions. We assume that the channel distribution $\mu_{Y^n|X^n}$ and the input distribution μ_{X^n} are given. Then the joint distribution of (X^n, Y^n) is given as

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

Here, we assume that an arbitrary source code $(A, \widehat{X}_A^n)$ with decoder side information (Fig. 2) is given, where $A : \mathcal{X}^n \to \mathrm{Im}A$ is an encoding function and $\widehat{X}_A^n : \mathrm{Im}A \times \mathcal{Y}^n \to \mathcal{X}^n$ is a (possibly stochastic) decoding function specified by a conditional probability distribution $\mu_{\widehat{X}^n|C_nY^n}$. Then the coding rate[2] r of the code is given as

$$r \equiv \frac{1}{n} \log |\mathrm{Im}A|. \tag{8}$$

The decoding error probability $\mathrm{Error}(A)$ of this code is given by (5), where $\mu_{\widehat{X}^n|C_nY^n}$ is allowed to be an arbitrary conditional probability distribution and need not to be defined by (4). It should be noted that we can use any source code as well as that constructed in the previous section. The condition $\lim_{n\to\infty} \mathrm{Error}(A) = 0$ is not assumed for this code, and this code may be sub-optimal in the sense that the coding rate r is not close to the fundamental limit $\overline{H}(\boldsymbol{X}|\boldsymbol{Y})$.

For a given rate $R > 0$ of the channel code, let $(\mathcal{B}, p_{\mathsf{B}})$ be an ensemble of functions on the set $\mathcal{X}^n$ satisfying

$$R = \frac{1}{n} \log |\mathrm{Im}\mathcal{B}|, \tag{9}$$

where $\mathrm{Im}\mathcal{B}$ represents the set of all messages. We obtain a function $B \in \mathcal{B}$ and a vector $\boldsymbol{c} \in \mathrm{Im}A$ generated at random subject to the distribution p_{B} and $\{\mu_{X^n}(\mathcal{C}_A(\boldsymbol{c}))\}_{\boldsymbol{c}\in\mathrm{Im}A}$, respectively. It should be noted that we can obtain $\boldsymbol{c} \equiv A\boldsymbol{x}$ generated at random subject to the distribution $\{\mu_{X^n}(\mathcal{C}_A(\boldsymbol{c}))\}_{\boldsymbol{c}\in\mathrm{Im}A}$ by generating $\boldsymbol{x}$ at random subject to the distribution μ_{X^n} and operating A on $\boldsymbol{x}$.

We fix B and $\boldsymbol{c}$ so that they are shared by the channel encoder and the channel decoder. To summarize, the channel encoder has functions A, B, and a vector $\boldsymbol{c}$, and the channel decoder has functions $\widehat{X}_A^n$, B, and a vector $\boldsymbol{c}$.

We use a constrained-random-number generator to construct a stochastic encoder. For a given message $\boldsymbol{m} \in \mathrm{Im}\mathcal{B}$, let $\widetilde{X}^n \equiv \widetilde{X}_{AB}^n(\boldsymbol{c}, \boldsymbol{m})$ be a random variable corresponding to the distribution

$$\mu_{\widetilde{X}^n|C_nM_n}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{m}) \equiv \frac{\mu_{X^n}(\boldsymbol{x})\chi(\boldsymbol{x} \in \mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m}))}{\mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m}))},$$

where $\mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m}) \equiv \mathcal{C}_A(\boldsymbol{c}) \cap \mathcal{C}_B(\boldsymbol{m})$. The encoder generates $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \boldsymbol{c}$ and $B\boldsymbol{x} = \boldsymbol{m}$ with probability $\mu_{\widetilde{X}^n|C_n,M_n}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{m})$. We define the stochastic channel encoder $\Phi_n : \mathrm{Im}\mathcal{B} \to \mathcal{X}^n$ as

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

Let $\boldsymbol{y} \in \mathcal{Y}^n$ be a channel output. We define the (possibly stochastic) channel decoder $\Psi_n : \mathcal{Y}^n \to \mathrm{Im}\mathcal{B}$ as

$$\Psi_n(\boldsymbol{y}) \equiv B\widehat{X}_A^n(\boldsymbol{c}, \boldsymbol{y}),$$

[2]It should be noted that (8) is different from (3). Eq. (8) is a condition for a fixed encoding function A while (3) is a condition for an ensemble $\mathcal{A}$ of encoding functions.

$$\mathrm{Error}(A, B, \boldsymbol{c}) \equiv \sum_{\substack{\boldsymbol{m}: \\ \mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))=0}} \frac{1}{|\mathrm{Im}\mathcal{B}|} + \sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}: \\ \mu_{X^n}(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0 \\ \boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}) \\ B\widehat{\boldsymbol{x}}\neq\boldsymbol{m}}} \frac{\mu_{\widehat{X}^n|C_nY^n}(\widehat{\boldsymbol{x}}|\boldsymbol{c}, \boldsymbol{y})\mu_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\mu_{\widetilde{X}^n|C_nM_n}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{m})}{|\mathrm{Im}\mathcal{B}|}, \tag{10}$$

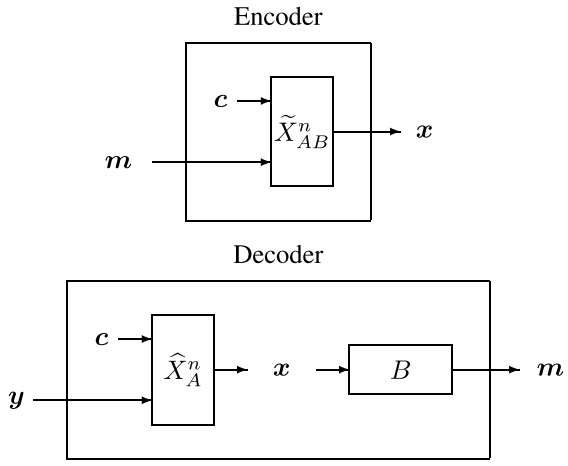

Fig. 3. Construction of Channel Code

where the decoder reproduces $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \boldsymbol{c}$ by using arbitrary (possibly stochastic) decoding function $\widehat{X}^n_A$ and obtains a reproduced message $\boldsymbol{m} = B\boldsymbol{x}$. The flow of vectors is illustrated in Fig. 3.

The error probability $\mathrm{Error}(A, B, \boldsymbol{c})$ is given by (10), shown at the bottom of the previous page, where the first term corresponds to the encoding error probability and the second term corresponds to the decoding error probability.

B. $(\boldsymbol{\alpha}, \boldsymbol{\beta})$-Balanced-Coloring Property

To state the theorem, we review the balanced-coloring property [29], which is a variant of the hash property. It requires weaker conditions than the hash property introduced in [28], [32], and [33].

Definition 2 [29]: Let $\mathcal{B}_n$ be a set of functions on $\mathcal{X}^n$ and $p_{\mathsf{B},n}$ be a probability distribution on $\mathcal{B}_n$. Then a sequence $(\boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{B}}) \equiv \{(\mathcal{B}_n, p_{\mathsf{B},n})\}_{n=1}^{\infty}$ has an $(\boldsymbol{\alpha}_{\mathsf{B}}, \boldsymbol{\beta}_{\mathsf{B}})$-*balanced-coloring property* if there are two sequences $\boldsymbol{\alpha}_{\mathsf{B}} \equiv \{\alpha_{\mathsf{B}}(n)\}_{n=1}^{\infty}$ and $\boldsymbol{\beta}_{\mathsf{B}} \equiv \{\beta_{\mathsf{B}}(n)\}_{n=1}^{\infty}$, depending on $\{p_{\mathsf{B},n}\}_{n=1}^{\infty}$, such that

$$\limsup_{n\to\infty} \alpha_{\mathsf{B}}(n) = 1 \tag{BC1}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log(\beta_{\mathsf{B}}(n) + 1) = 0 \tag{BC2}$$

and

$$\sum_{\substack{\boldsymbol{x}' \in \mathcal{X}^n \setminus \{\boldsymbol{x}\}: \\ p_{\mathsf{B},n}(\{B : B\boldsymbol{x} = B\boldsymbol{x}'\}) > \frac{\alpha_{\mathsf{B}}(n)}{|\mathrm{Im}\mathcal{B}_n|}}} p_{\mathsf{B},n}\left(\left\{B : B\boldsymbol{x} = B\boldsymbol{x}'\right\}\right) \le \beta_{\mathsf{B}}(n) \tag{BC3}$$

for all sufficiently large[3] n and all $\boldsymbol{x} \in \mathcal{X}^n$, where (BC3) is the same as (CR3). In the following, we omit the dependence of α_{B} and β_{B} on n.

Here, let us introduce examples satisfying the balanced-coloring property. When $\mathcal{B}$ is a two-universal class of hash functions [7] and p_{B} is the uniform distribution on $\mathcal{B}$, then $(\boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{B}})$ has a $(\mathbf{1}, \mathbf{0})$-balanced-coloring property. Random binning [9] and a set of all linear functions [11] are examples of the two-universal class of hash functions. It is proved in [29] that an ensemble of systematic sparse matrices has a balanced-coloring property, where a matrix has an identity sub-matrix with the same number of rows.

The following lemma is an extension of the leftover hash lemma [22], the balanced-coloring lemma [1, Lemma 3.1], [12, Lemma 17.3], and the output statistics of random binning [46]. This lemma implies that there is a function B such that $\mathcal{T}$ is almost equally partitioned by B with respect to a measure Q.

Lemma 2 ([28, Lemma 5], [33, Lemma 4]): If $(\mathcal{B}, p_{\mathsf{B}})$ satisfies (BC3), then

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

for any function $Q : \mathcal{X}^n \to [0, \infty)$ and $\mathcal{T} \subset \mathcal{X}^n$, where

$$Q(\mathcal{T}) \equiv \sum_{\boldsymbol{x}\in\mathcal{T}} Q(\boldsymbol{x}).$$

C. Existence of Good Function

Now, we are in position to state the theorem, where the proof is given in Section V-B.

Theorem 2: Let $(A, \widehat{X}^n_A)$ be an arbitrary source code with decoder side information, where the encoding rate and the decoding error probability are given by (8) and (5), respectively. Assume that $(\boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{B}})$ has an $(\boldsymbol{\alpha}_{\mathsf{B}}, \boldsymbol{\beta}_{\mathsf{B}})$-balanced-coloring property for a given R satisfying (9) and

$$r + R < \underline{H}(\boldsymbol{X}), \tag{11}$$

where $\underline{H}(\boldsymbol{X})$ is the spectral inf-entropy rate. Then for any $\delta > 0$ and all sufficiently large n there is a function $B \in \mathcal{B}$, and a vector $\boldsymbol{c} \in \mathrm{Im}A$ such that

$$\mathrm{Error}(A, B, \boldsymbol{c}) \le \mathrm{Error}(A) + \delta. \tag{12}$$

Remark 2: In [39, Remark 2], [46], inequality (12) is shown by assuming that the output distribution of an encoding function A is close to a uniform distribution. In contrast, such an assumption is not assumed in the above theorem.

D. Achievablity to Channel Capacity

In this section, we show that the channel capacity is achievable by combining proposed codes.

First, we review the definition of the capacity of a general channel. Let $\mathcal{X}$ and $\mathcal{Y}$ be the alphabets of a channel input and output, respectively. Then product sets $\mathcal{X}^n$ and $\mathcal{Y}^n$ are the alphabets of a channel input vector X^n and a channel output vector Y^n, respectively. It should be noted that $\mathcal{X}$ and $\mathcal{Y}$ are allowed to be infinite/uncountable/continuous sets on condition that probability distributions/measures μ_{X^n} and $\mu_{Y^n|X^n}(\cdot|\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}^n$ are well-defined.

Here, we define the operational channel capacity with a channel input constraint specified by a set $\mathcal{P}_n$ of probability

[3] In [29], an ensemble is required to satisfy (10) for all n and all $\boldsymbol{x} \in \mathcal{X}^n$. However, it is sufficient to assume that an ensemble satisfies (10) for sufficiently large n and all $\boldsymbol{x} \in \mathcal{X}^n$ because we finally let $n \to \infty$.

distributions on $\mathcal{X}^n$. A typical example of a channel input constraint is the cost constraint, where any distribution $\mu \in \mathcal{P}_n$ satisfies

$$\int c_n(\boldsymbol{x})\mu(\boldsymbol{x})d\boldsymbol{x} < C$$

for a given cost function $c_n : \mathcal{X}^n \to [0, \infty)$ and $C \in [0, \infty)$.

Definition 3: Let $\boldsymbol{\mathcal{P}} \equiv \{\mathcal{P}_n\}_{n=1}^{\infty}$ be a sequence of the set of probability distributions on $\mathcal{X}^n$. For a general channel $\boldsymbol{W}$, we call a rate R *achievable* if for all $\delta > 0$ and all sufficiently large n there is a pair consisting of a (possibly stochastic) encoder $\Phi_n : \mathcal{M}_n \to \mathcal{S}^n$ and a (possibly stochastic) decoder $\Psi_n : \mathcal{Y}^n \to \mathcal{M}_n$ such that

$$\frac{1}{n}\log|\mathcal{M}_n| \geq R$$
$$\mu_{X^n} \in \mathcal{P}_n$$
$$\mathrm{Prob}(\widehat{M}_n \neq M_n) \leq \delta,$$

where we call a subset $\mathcal{S}$ of $\mathcal{X}$ a *signaling alphabet*,[4] $\mathcal{M}_n$ 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 $\mathcal{M}_n$, Y^n is the random variable of a channel output with an input $X^n \equiv \Phi_n(M_n)$, $\widehat{M}_n \equiv \Psi_n(Y^n)$ is the random variable of a reproduction, and the joint distribution $\mu_{M_n X^n Y^n \widehat{M}_n}$ of $(M_n, X^n, Y^n, \widehat{M}_n)$ is given as

$$\mu_{M_n X^n Y^n \widehat{M}_n}(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{y}, \widehat{\boldsymbol{m}}) \equiv \frac{\mu_{\widehat{M}_n|Y^n}(\widehat{\boldsymbol{m}}|\boldsymbol{y})\mu_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\mu_{X^n|M_n}(\boldsymbol{x}|\boldsymbol{m})}{|\mathcal{M}_n|}$$

by using conditional distributions $\mu_{X^n|M_n}$ and $\mu_{\widehat{M}_n|Y^n}$ characterizing Φ_n and Ψ_n, respectively. The *channel capacity* $C_{\mathcal{S}}(\boldsymbol{W})$ is defined by the supremum of the achievable rate, where the signaling alphabet $\mathcal{S}$ is specified.

It should be noted that the standard definition of channel capacity can be denoted by $C_{\mathcal{X}}(\boldsymbol{W})$, where the signaling alphabet $\mathcal{S}$ is equal to $\mathcal{X}$. It should also be noted that we can let $\mathcal{P}_n$ be the set of all probability distributions on $\mathcal{X}^n$ when it is assumed that there is no channel input constraint.

In the following, we review the formulas of channel capacity for a general channel $\boldsymbol{W}$, where encoders and decoders are assumed to be deterministic. Let $\widetilde{C}_{\mathcal{X}}(\boldsymbol{W})$ be the channel capacity for deterministic encoders and decoders. The information theoretic expression of $\widetilde{C}_{\mathcal{X}}(\boldsymbol{W})$ is derived in [44], [19, Th. 3.6.1][5] as

$$\widetilde{C}_{\mathcal{X}}(\boldsymbol{W}) = \sup_{\boldsymbol{X}\in\boldsymbol{\mathcal{P}}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}),$$

where the supremum is taken over all general sources $\boldsymbol{X} = \{\mu_{X^n}\}_{n=1}^{\infty}$ such that $\mu_{X^n} \in \mathcal{P}_n$ for every n, and the joint distribution $\mu_{X^nY^n}$ is given as

$$\mu_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y}) \equiv \mu_{Y^n|X^n}(\boldsymbol{y}|\boldsymbol{x})\mu_{X^n}(\boldsymbol{x}). \tag{13}$$

Furthermore, similarly to the proof in [28], we can show the formula

$$\widetilde{C}_{\mathcal{X}}(\boldsymbol{W}) = \sup_{\boldsymbol{X}\in\boldsymbol{\mathcal{P}}} \left[\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})\right]$$

when $\mathcal{X}$ is finite, where the supremum is taken over all general sources $\boldsymbol{X}$ and the joint distribution of $(\boldsymbol{X}, \boldsymbol{Y})$ is given by (13). We can show by using random coding argument that $C_{\mathcal{X}}(\boldsymbol{W}) = \widetilde{C}_{\mathcal{X}}(\boldsymbol{W})$, which implies the capacity formula as

$$C_{\mathcal{X}}(\boldsymbol{W}) = \sup_{\boldsymbol{X}\in\boldsymbol{\mathcal{P}}} \left[\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})\right]. \tag{14}$$

This implies that the capacity does not increase by allowing stochastic encoders and decoders.

Here, let us assume that $\mathcal{X}$ is a finite set. Then, from Theorems 1 and 2, we have the following corollary, which is an improvement of [28, Th. 1]. It should be noted that the conditions for $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}})$ and $(\boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{B}})$ are weaker than those in [28, Th. 1]. Furthermore, both encoding and decoding functions can be constructed by using constrained-random-number generators.

Corollary 3: Assume that $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}})$ and $(\boldsymbol{\mathcal{B}}, \boldsymbol{p}_{\mathsf{B}})$ have an $(\boldsymbol{\alpha}_{\mathsf{A}}, \boldsymbol{\beta}_{\mathsf{A}})$-collision-resistance property and an $(\boldsymbol{\alpha}_{\mathsf{B}}, \boldsymbol{\beta}_{\mathsf{B}})$-balanced-coloring property, respectively, for given r and R satisfying

$$r = \frac{1}{n}\log|\mathrm{Im}\mathcal{A}|$$
$$R = \frac{1}{n}\log|\mathrm{Im}\mathcal{B}|$$
$$r > \overline{H}(\boldsymbol{X}|\boldsymbol{Y})$$
$$r + R < \underline{H}(\boldsymbol{X}).$$

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 the decoding error probability is less than δ. 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 (14), $\mathcal{M}_n \equiv \mathrm{Im}\mathcal{B}$, $r \to \overline{H}(\boldsymbol{X}|\boldsymbol{Y})$, $R \to \underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})$, and $\delta \to 0$.

Finally, let us apply our results to the case when $\mathcal{X}$ is an infinite/continuous set. Let us define the capacity $C_{\mathcal{X}}^q(\boldsymbol{W})$ of a channel with a finite signaling alphabet as

$$C_{\mathcal{X}}^q(\boldsymbol{W}) \equiv \sup_{\mathcal{S}\subset\mathcal{X}:|\mathcal{S}|\leq q} C_{\mathcal{S}}(\boldsymbol{W}).$$

Since $\{\mathcal{S} : |\mathcal{S}| \leq q\} \subset \{\mathcal{S} : |\mathcal{S}| \leq q+1\}$, we have the fact that $C_{\mathcal{X}}^q(\boldsymbol{W})$ is a non-decreasing function of q. We have the following lemma.

Lemma 3 ([34, Theorem 2][6]): When $C_{\mathcal{X}}(\boldsymbol{W}) < \infty$, we have

$$\begin{aligned} C_{\mathcal{X}}(\boldsymbol{W}) &= \lim_{q\to\infty} C_{\mathcal{X}}^q(\boldsymbol{W}) \\ &= \lim_{q\to\infty} \sup_{\substack{\mathcal{S}\subset\mathcal{X}:\\ |\mathcal{S}|\leq q}} \sup_{\substack{\boldsymbol{X}\in\boldsymbol{\mathcal{P}}:\\ X^n\in\mathcal{S}^n \text{ for all } n}} \underline{I}(\boldsymbol{X};\boldsymbol{Y}) \\ &= \lim_{q\to\infty} \sup_{\substack{\mathcal{S}\subset\mathcal{X}:\\ |\mathcal{S}|\leq q}} \sup_{\substack{\boldsymbol{X}\in\boldsymbol{\mathcal{P}}:\\ X^n\in\mathcal{S}^n \text{ for all } n}} [\underline{H}(\boldsymbol{X}) - \overline{H}(\boldsymbol{X}|\boldsymbol{Y})], \end{aligned} \tag{15}$$

where the condition $X^n \in \mathcal{S}^n$ implies that the support of the probability distribution of a channel input is a subset of $\mathcal{S}^n$.

From the above lemma, we have the fact that the capacity of a channel with an uncountable channel input alphabet is

[4]This terminology comes from [6].

[5]In [19, Th. 3.6.1], it is assumed that $\boldsymbol{\mathcal{P}}$ is a cost constraint. However, we can easily extend the result to an arbitrary channel input constraint.

[6]In [34, Th. 2], deterministic encoders and decoders are allowed. We can show the lemma by using the relation $C_{\mathcal{X}}(\boldsymbol{W}) = \widetilde{C}_{\mathcal{X}}(\boldsymbol{W})$.

achievable with the code by optimizing the finite signaling alphabet $\mathcal{S}$ and letting $|\mathcal{S}| \to \infty$.

Remark 3: For many channels it is known that an optimal input distribution in a channel coding has a discrete support, where a support is defined as the set of all elements with positive measure. For example, for an additive white Gaussian noise (AWGN) channel, it is shown in [42] that the optimal input distribution has a discrete and finite support under the maximum power constraint. It should be noted that the above lemma implies that we can approach the capacity with a sufficiently large signaling alphabet for any channel with an uncountable/continuous channel input alphabet (e.g. AWGN channel under the average power constraint).

Remark 4: In [6, Sec. 7.8], the optimal signaling alphabet $\mathcal{S} \in \mathcal{X}$ is derived for an additive white Gaussian noise channel, where it is assumed that all the symbols in $\mathcal{S}$ are used equally often, that is, the input distribution is uniform on $\mathcal{S}$. This assumption is natural when we use conventional linear codes. On the other hand, it is unnecessary to assume that the input distribution is uniform on $\mathcal{S}$ in our code construction, where the encoding rate may increase.

V. Proof of Theorems

In the following proofs, we omit the dependence on n of X, Y, C, and $\widehat{X}$ when they appear in the subscripts of μ, $\overline{\mathcal{T}}$, and $\underline{\mathcal{T}}$. The integral over the alphabet $\mathcal{Y}^n$ is denoted by $\sum$.

A. Proof of Theorem 1

From (6), we have the fact that there is $\varepsilon > 0$ satisfying

$$r > \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) + \varepsilon. \tag{16}$$

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

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

Assume that $(\boldsymbol{x}, \boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}$ and $\widehat{\boldsymbol{x}}_A(A\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x}$, where $\widehat{\boldsymbol{x}}_A(A\boldsymbol{x}, \boldsymbol{y})$ is defined by (7). 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 $\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y}) \equiv \{\boldsymbol{x} : (\boldsymbol{x}, \boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}\}$. We have

$$\begin{aligned}
&E_{\mathsf{A}}\left[\chi(\widehat{\boldsymbol{x}}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x})\right] \\
&\quad \le p_{\mathsf{A}}\left(\left\{A : [\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y}) \setminus \{\boldsymbol{x}\}] \cap \mathcal{C}_A(A\boldsymbol{x}) \neq \emptyset\right\}\right) \\
&\quad \le \frac{|\overline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\alpha_{\mathsf{A}}}{|\mathrm{Im}\mathcal{A}|} + \beta_{\mathsf{A}} \\
&\quad \le 2^{-n[r-\overline{H}(\boldsymbol{X}|\boldsymbol{Y})-\varepsilon]}\alpha_{\mathsf{A}} + \beta_{\mathsf{A}}
\end{aligned} \tag{17}$$

for all $(\boldsymbol{x}, \boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}$, where $\chi(\cdot)$ is defined by (2), the second inequality comes from Lemma 1, and the third inequality comes from (6) and 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

$$\begin{aligned}
&E_{\mathsf{A}}\left[\mu_{XY}\left(\{(\boldsymbol{x}, \boldsymbol{y}) : \widehat{\boldsymbol{x}}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x}\}\right)\right] \\
&\quad = E_{\mathsf{A}}\left[\sum_{\boldsymbol{x},\boldsymbol{y}} \mu_{XY}(\boldsymbol{x}, \boldsymbol{y})\chi(\widehat{\boldsymbol{x}}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x})\right] \\
&\quad = \sum_{(\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{T}}_{X|Y}} \mu_{XY}(\boldsymbol{x}, \boldsymbol{y}) E_{\mathsf{A}}\left[\chi(\widehat{\boldsymbol{x}}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x})\right] \\
&\qquad + \sum_{(\boldsymbol{x},\boldsymbol{y}) \notin \overline{\mathcal{T}}_{X|Y}} \mu_{XY}(\boldsymbol{x}, \boldsymbol{y}) E_{\mathsf{A}}\left[\chi(\widehat{\boldsymbol{x}}_{\mathsf{A}}(\mathsf{A}\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x})\right] \\
&\quad \le 2^{-n[r-\overline{H}(\boldsymbol{X}|\boldsymbol{Y})-\varepsilon]}\alpha_{\mathsf{A}} + \beta_{\mathsf{A}} + \mu_{XY}([\overline{\mathcal{T}}_{X|Y}]^c),
\end{aligned} \tag{18}$$

where the last inequality comes from (17). From (CR1),(CR2), (16) and the fact that $\mu_{XY}([\underline{\mathcal{T}}_{XY}]^c) \to 0$ as $n \to \infty$, we have the fact that there is a function $A \in \mathcal{A}$ satisfying

$$\mu_{XY}\left(\{(\boldsymbol{x}, \boldsymbol{y}) : \widehat{\boldsymbol{x}}_A(A\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x}\}\right) \le \frac{\delta}{2} \tag{19}$$

for all $\delta > 0$ and all sufficiently large n.

Since the joint distribution of (X^n, Y^n, C_n) is given as

$$\mu_{XYC}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{c}) = \mu_{XY}(\boldsymbol{x}, \boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c}), \tag{20}$$

we have

$$\begin{aligned}
\mu_{X|CY}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{y}) &= \frac{\mu_{XY}(\boldsymbol{x}, \boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c})}{\sum_{\boldsymbol{x}'} \mu_{XY}(\boldsymbol{x}', \boldsymbol{y})\chi(A\boldsymbol{x}' = \boldsymbol{c})} \\
&= \frac{\mu_{X|Y}(\boldsymbol{x}|\boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c})}{\mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})} \\
&= \mu_{\widehat{X}|CY}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{y}),
\end{aligned} \tag{21}$$

that is, the constrained-random-number generator defined by (4) is the stochastic decision with $\mu_{X|CY}$. Furthermore, we have

$$\begin{aligned}
\widehat{\boldsymbol{x}}_A(\boldsymbol{c}, \boldsymbol{y}) &= \arg\max_{\widehat{\boldsymbol{x}} \in \mathcal{C}_A(\boldsymbol{c})} \mu_{X|Y}(\widehat{\boldsymbol{x}}|\boldsymbol{y}) \\
&= \arg\max_{\widehat{\boldsymbol{x}}} \mu_{X|Y}(\widehat{\boldsymbol{x}}|\boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c}) \\
&= \arg\max_{\widehat{\boldsymbol{x}}} \frac{\mu_{X|Y}(\widehat{\boldsymbol{x}}|\boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c})\mu_Y(\boldsymbol{y})}{\mu_{CY}(\boldsymbol{c}, \boldsymbol{y})} \\
&= \arg\max_{\widehat{\boldsymbol{x}}} \mu_{X|CY}(\widehat{\boldsymbol{x}}|\boldsymbol{c}, \boldsymbol{y}),
\end{aligned} \tag{22}$$

that is, $\widehat{\boldsymbol{x}}_A(\boldsymbol{c}, \boldsymbol{y})$ is a maximum a posteriori decision for a given joint distribution μ_{XYC}, where the last equality comes from (20). Then we have

$$\begin{aligned}
&\mathrm{Error}(A) \\
&= \sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{c},\widehat{\boldsymbol{x}}} \mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\boldsymbol{c}, \boldsymbol{y})\mu_{XY}(\boldsymbol{x}, \boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c})\chi(\widehat{\boldsymbol{x}} \neq \boldsymbol{x}) \\
&= \sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{c},\widehat{\boldsymbol{x}}} \mu_{X|CY}(\widehat{\boldsymbol{x}}|\boldsymbol{c}, \boldsymbol{y})\mu_{X|CY}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{y})\mu_{CY}(\boldsymbol{c}, \boldsymbol{y})\chi(\widehat{\boldsymbol{x}} \neq \boldsymbol{x}) \\
&\le 2\sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{c}} \mu_{X|CY}(\boldsymbol{x}|\boldsymbol{c}, \boldsymbol{y})\mu_{CY}(\boldsymbol{c}, \boldsymbol{y})\chi(\widehat{\boldsymbol{x}}_A(\boldsymbol{c}, \boldsymbol{y}) \neq \boldsymbol{x}) \\
&= 2\sum_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{c}} \mu_{XY}(\boldsymbol{x}, \boldsymbol{y})\chi(A\boldsymbol{x} = \boldsymbol{c})\chi(\widehat{\boldsymbol{x}}_A(\boldsymbol{c}, \boldsymbol{y}) \neq \boldsymbol{x}) \\
&= 2\sum_{\boldsymbol{x},\boldsymbol{y}} \mu_{XY}(\boldsymbol{x}, \boldsymbol{y})\chi(\widehat{\boldsymbol{x}}_A(A\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x}) \\
&= 2\mu_{XY}\left(\{(\boldsymbol{x}, \boldsymbol{y}) : \widehat{\boldsymbol{x}}_A(A\boldsymbol{x}, \boldsymbol{y}) \neq \boldsymbol{x}\}\right) \\
&\le \delta
\end{aligned} \tag{23}$$

for all $\delta > 0$ and all sufficiently large n, where the second equality comes from (20) and (21), the first inequality comes from Lemma 4 in Appendix, and the last inequality comes from (19). ■

B. Proof of Theorem 2

From (11), we have the fact that there is $\varepsilon > 0$ satisfying

$$r + R < \underline{H}(\boldsymbol{X}) - \varepsilon. \tag{24}$$

Let $\underline{\mathcal{T}}_X \subset \mathcal{X}^n$ be defined as

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

First, for all A, we have

$$\begin{aligned}
&E_{\mathsf{Bc}}\left[\sum_{\boldsymbol{m}} \left| \frac{\mu_X(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))}{\mu_X(\mathcal{C}_A(\mathbf{c}))} - \frac{1}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&= E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \left| \mu_X(\mathcal{C}_{A\mathsf{B}}(\boldsymbol{c},\boldsymbol{m})) - \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}))}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&\leq E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \left| \mu_X(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}) \cap \underline{\mathcal{T}}_X) - \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&\quad + E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \left| \mu_X(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}) \cap [\underline{\mathcal{T}}_X]^c) \right| \right] \\
&\quad + E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \left| -\frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap [\underline{\mathcal{T}}_X]^c)}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&= E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \left| \mu_X(\mathcal{C}_{A\mathsf{B}}(\boldsymbol{c},\boldsymbol{m}) \cap \underline{\mathcal{T}}_X) - \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&\quad + E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \mu_X(\mathcal{C}_{A\mathsf{B}}(\boldsymbol{c},\boldsymbol{m}) \cap [\underline{\mathcal{T}}_X]^c) \right] \\
&\quad + E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap [\underline{\mathcal{T}}_X]^c)}{|\mathrm{Im}\mathcal{B}|} \right],
\end{aligned} \tag{25}$$

where the first equality comes from the fact that the distribution of the random variable $\mathbf{c}$ is $\{\mu_X(\mathcal{C}_A(\boldsymbol{c}))\}_{\boldsymbol{c} \in \mathrm{Im}A}$, the first inequality comes from the triangular inequality and the fact that

$$\begin{aligned}
\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})) &= \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}) \cap \underline{\mathcal{T}}_X) \\
&\quad + \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}) \cap [\underline{\mathcal{T}}_X]^c) \\
\mu_X(\mathcal{C}_A(\boldsymbol{c})) &= \mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X) + \mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap [\underline{\mathcal{T}}_X]^c).
\end{aligned}$$

Since $\{C_{AB}(\boldsymbol{c},\boldsymbol{m})\}_{\boldsymbol{c},\boldsymbol{m}}$ and $\{C_A(\boldsymbol{c})\}_{\boldsymbol{c}}$ form a partition, the second and the third terms on the right hand side of (25) are evaluated as

$$E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \mu_X(\mathcal{C}_{A\mathsf{B}}(\boldsymbol{c},\boldsymbol{m}) \cap [\underline{\mathcal{T}}_X]^c)\right] = \mu_X([\underline{\mathcal{T}}_X]^c) \tag{26}$$

$$E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap [\underline{\mathcal{T}}_X]^c)}{|\mathrm{Im}\mathcal{B}|}\right] = \mu_X([\underline{\mathcal{T}}_X]^c). \tag{27}$$

On the other hand, the first term on the right hand side of (25) is evaluated as (28), shown at the bottom of this page, where the fist inequality comes from Lemma 2 and the fact that $\mu_X(\boldsymbol{x}) \leq 2^{-n[\underline{H}(\boldsymbol{X})-\varepsilon]}$ for all $\boldsymbol{x} \in \underline{\mathcal{T}}_X$, the second inequality comes from the Jensen inequality, the third equality comes from (8), (9), and the last inequality comes from the fact that $\mu_X(\underline{\mathcal{T}}_X) \leq 1$.

Next, we have

$$\begin{aligned}
&E_{\mathsf{Bc}}\left[\sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}: \\ \mu_{X^n}(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))>0 \\ \boldsymbol{x} \in \mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}) \\ B\widehat{\boldsymbol{x}} \neq \boldsymbol{m}}} \frac{\mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\mathbf{c},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})}{\mu_X(\mathcal{C}_A(\mathbf{c}))}\right] \\
&\leq E_{\mathsf{Bc}}\left[\sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}: \\ \boldsymbol{x} \in \mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}) \\ \widehat{\boldsymbol{x}} \neq \boldsymbol{x}}} \frac{\mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\mathbf{c},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})}{\mu_X(\mathcal{C}_A(\mathbf{c}))}\right]
\end{aligned}$$

$$\begin{aligned}
&E_{\mathsf{B}}\left[\sum_{\boldsymbol{m},\boldsymbol{c}} \left| \mu_X(\mathcal{C}_{A\mathsf{B}}(\boldsymbol{c},\boldsymbol{m}) \cap \underline{\mathcal{T}}_X) - \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&= \sum_{\boldsymbol{c}} \mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X) E_{\mathsf{B}}\left[\sum_{\boldsymbol{m}} \left| \frac{\mu_X(\mathcal{C}_{\mathsf{B}}(\boldsymbol{m}) \cap \mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)} - \frac{1}{|\mathrm{Im}\mathcal{B}|} \right| \right] \\
&\leq \sum_{\boldsymbol{c}} \mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X) \sqrt{\alpha_{\mathsf{B}} - 1 + \frac{[\beta_{\mathsf{B}}+1]|\mathrm{Im}\mathcal{B}|2^{-n[\underline{H}(\boldsymbol{X})-\varepsilon]}}{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}} \\
&\leq \mu_X(\underline{\mathcal{T}}_X) \sqrt{\sum_{\boldsymbol{c}} \frac{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}{\mu_X(\underline{\mathcal{T}}_X)} \left[\alpha_{\mathsf{B}} - 1 + \frac{[\beta_{\mathsf{B}}+1]|\mathrm{Im}\mathcal{B}|2^{-n[\underline{H}(\boldsymbol{X})-\varepsilon]}}{\mu_X(\mathcal{C}_A(\boldsymbol{c}) \cap \underline{\mathcal{T}}_X)}\right]} \\
&= \mu_X(\underline{\mathcal{T}}_X) \sqrt{\alpha_{\mathsf{B}} - 1 + \frac{[\beta_{\mathsf{B}}+1]|\mathrm{Im}A||\mathrm{Im}\mathcal{B}|2^{-n[\underline{H}(\boldsymbol{X})-\varepsilon]}}{\mu_X(\underline{\mathcal{T}}_X)}} \\
&= \sqrt{[\alpha_{\mathsf{B}} - 1]\,\mu_X(\underline{\mathcal{T}}_X)^2 + [\beta_{\mathsf{B}}+1]2^{-n[\underline{H}(\boldsymbol{X})-r-R-\varepsilon]}\mu_X(\underline{\mathcal{T}}_X)} \\
&\leq \sqrt{\alpha_{\mathsf{B}} - 1 + [\beta_{\mathsf{B}}+1]2^{-n[\underline{H}(\boldsymbol{X})-r-R-\varepsilon]}}
\end{aligned} \tag{28}$$

$$= E_{\mathsf{B}}\left[\sum_{\substack{\boldsymbol{c},\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})\\ \widehat{\boldsymbol{x}}\neq\boldsymbol{x}}} \mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\boldsymbol{c},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\right]$$

$$= E_{\mathsf{B}}\left[\sum_{\substack{\boldsymbol{c},\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})\\ \widehat{\boldsymbol{x}}\neq\boldsymbol{x}}} \mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|A\boldsymbol{x},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\right]$$

$$= \sum_{\substack{\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \widehat{\boldsymbol{x}}\neq\boldsymbol{x}}} \mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|A\boldsymbol{x},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y}), \tag{29}$$

where the first equality comes from the fact that the distribution of the random variable $\mathbf{c}$ is $\{\mu_X(\mathcal{C}_A(\boldsymbol{c}))\}_{\boldsymbol{c}\in\mathrm{Im}A}$, the second equality comes from the fact that $\boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})$ is equivalent to $A\boldsymbol{x}=\boldsymbol{c}$ and $B\boldsymbol{x}=\boldsymbol{m}$, and the third equality comes from the fact that $\{C_{AB}(\boldsymbol{c},\boldsymbol{m})\}_{\boldsymbol{c},\boldsymbol{m}}$ forms a partition.

Finally, we have (30), shown at the bottom of this page, where the first inequality comes from (29) and the fact that

$$\begin{aligned}
&\sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0\\ \boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})\\ B\widehat{\boldsymbol{x}}\neq\boldsymbol{m}}} \mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\boldsymbol{c},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\\
&\quad\cdot\left[\frac{1}{|\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}-\frac{1}{\mu_X(\mathcal{C}_A(\boldsymbol{c}))}\right]\\
&\le \sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0\\ \boldsymbol{x}\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})\\ B\widehat{\boldsymbol{x}}\neq\boldsymbol{m}}} \mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\boldsymbol{c},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\\
&\quad\cdot\left|\frac{1}{|\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}-\frac{1}{\mu_X(\mathcal{C}_A(\boldsymbol{c}))}\right|\\
&\le \sum_{\substack{\boldsymbol{m}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0}} \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))\\
&\quad\cdot\left|\frac{1}{|\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}-\frac{1}{\mu_X(\mathcal{C}_A(\boldsymbol{c}))}\right|\\
&= \sum_{\substack{\boldsymbol{m}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))>0}} \left|\frac{\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}{\mu_X(\mathcal{C}_A(\boldsymbol{c}))}-\frac{1}{|\mathrm{Im}\mathcal{B}|}\right|\\
&= \sum_{\boldsymbol{m}}\left|\frac{\mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))}{\mu_X(\mathcal{C}_A(\boldsymbol{c}))}-\frac{1}{|\mathrm{Im}\mathcal{B}|}\right|-\sum_{\substack{\boldsymbol{m}:\\ \mu_X(\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}))=0}}\frac{1}{|\mathrm{Im}\mathcal{B}|},
\end{aligned} \tag{31}$$

and the second inequality comes from (25)–(28). From (BC1), (BC2), (24), (30), and the fact that $\mu_X([\underline{\mathcal{T}}_X]^c)\to 0$ as $n\to\infty$, we have the fact that there is a pair consisting of a function $B\in\mathcal{B}$ and a vector $\boldsymbol{c}\in\mathrm{Im}A$ that satisfy (12). ■

VI. Concluding Remarks

It is shown that we can construct a channel code from a given source code of $\boldsymbol{X}$ with decoder side information $\boldsymbol{Y}$, where the channel input and output are given by $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. We can construct a code that achieves the capacity by letting $\boldsymbol{X}$ be an optimum channel input random variable and using a source code achieving the limit $\overline{H}(\boldsymbol{X}|\boldsymbol{Y})$. This idea can be extended to multiple-input-multiple-output channels [36] including multiple-access channels and broadcast channels.

Constrained-random-number generators provide building blocks for code constructions.[6] We can implement tractable stochastic encoding and decoding functions by assuming a memoryless channel and using constrained-random-number generator with the sum-product algorithm or the Markov-Chain-Monte-Carlo method [28], [29]. Finding good

[6]We would like to call this type of codes CoCoNuTS (Codes based on Constrained Numbers Theoretically-achieving the Shannon limit).

$$\begin{aligned}
&E_{\mathsf{Bc}}\left[\mathrm{Error}(A,\mathsf{B},\mathbf{c})\right]\\
&= E_{\mathsf{Bc}}\left[\sum_{\substack{\boldsymbol{m}:\\ \mu_{X^n}(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))=0}}\frac{1}{|\mathrm{Im}\mathcal{B}|}+\sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \mu_{X^n}(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))>0\\ \boldsymbol{x}\in\mathcal{C}_{AB}(\mathbf{c},\boldsymbol{m})\\ \mathsf{B}\widehat{\boldsymbol{x}}\neq\boldsymbol{m}}}\frac{\mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\mathbf{c},\boldsymbol{y})\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})\mu_X(\boldsymbol{x})}{|\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))}\right]\\
&= E_{\mathsf{Bc}}\left[\sum_{\substack{\boldsymbol{m}:\\ \mu_{X^n}(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))=0}}\frac{1}{|\mathrm{Im}\mathcal{B}|}+\sum_{\substack{\boldsymbol{m},\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \mu_{X^n}(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))>0\\ \boldsymbol{x}\in\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m})\\ \mathsf{B}\widehat{\boldsymbol{x}}\neq\boldsymbol{m}}}\mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|\mathbf{c},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\left[\frac{1}{\mu_X(\mathcal{C}_A(\mathbf{c}))}+\frac{1}{|\mathrm{Im}\mathcal{B}|\mu_X(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))}-\frac{1}{\mu_X(\mathcal{C}_A(\mathbf{c}))}\right]\right]\\
&\le \sum_{\substack{\boldsymbol{x},\boldsymbol{y},\widehat{\boldsymbol{x}}:\\ \widehat{\boldsymbol{x}}\neq\boldsymbol{x}}}\mu_{\widehat{X}|CY}(\widehat{\boldsymbol{x}}|A\boldsymbol{x},\boldsymbol{y})\mu_{XY}(\boldsymbol{x},\boldsymbol{y})+E_{\mathsf{Bc}}\left[\sum_{\boldsymbol{m}}\left|\frac{\mu_X(\mathcal{C}_{A\mathsf{B}}(\mathbf{c},\boldsymbol{m}))}{\mu_X(\mathcal{C}_A(\mathbf{c}))}-\frac{1}{|\mathrm{Im}\mathcal{B}|}\right|\right]\\
&\le \mathrm{Error}(A)+\sqrt{\alpha_{\mathsf{B}}-1+[\beta_{\mathsf{B}}+1]2^{-n[\underline{H}(X)-r-R-\varepsilon]}}+2\mu_X([\underline{\mathcal{T}}_X]^c)
\end{aligned} \tag{30}$$

matrices and improving of these algorithms are left as a future research.

APPENDIX

This appendix reviews the result of [35], which investigates the error probability of the stochastic decision. It should be noted that stochastic decoding is a stochastic decision in the context of a coding scheme.

Let $\mathcal{U}$ and $\mathcal{V}$ be the alphabets of random variable U and V, respectively. We assume that the joint distribution p_{UV} of (U, V) is given.

Let us assume the situation where a decoder make a stochastic decision of the invisible state U after the observation V. We use a random number generator $\widehat{U} \in \mathcal{U}$ after observing V and let $\widehat{U}$ be a decision (guess) about the state U. Formally, we generate $\widehat{U}$ subject to the conditional distribution $q_{\widehat{U}|V}(\cdot|V)$ on $\mathcal{U}$ depending on an observation V and let an output be a decision of U, where U and $\widehat{U}$ are conditionally independent for a given V, that is, $U \leftrightarrow V \leftrightarrow \widehat{U}$ forms a Markov chain. The joint distribution $p_{UV\widehat{U}}$ of $(U, V, \widehat{U})$ is given as

$$p_{UV\widehat{U}}(u, v, \widehat{u}) = q_{\widehat{U}|V}(\widehat{u}|v)p_{U|V}(u|v)p_V(v).$$

Let us call $q_{\widehat{U}|V}$ a *stochastic decision rule*. As a special case, when $q_{\widehat{U}|V}$ is given by using a function $f : \mathcal{U} \to \mathcal{V}$ and is defined as

$$q_{\widehat{U}|V}(\widehat{u}|v) = \begin{cases} 1 & \text{if } \widehat{u} = f(v) \\ 0 & \text{if } \widehat{u} \neq f(v), \end{cases} \tag{32}$$

we call $q_{\widehat{U}|V}$ or f a *deterministic decision rule*.

Let χ be defined by (2). Then the error probability $\mathrm{Error}(q_{\widehat{U}|V})$ of a (stochastic) decision rule $q_{\widehat{U}|V}$ is given as

$$\begin{aligned} &\mathrm{Error}(q_{\widehat{U}|V}) \\ &= \sum_v p_V(v) \sum_u p_{U|V}(u|v) \sum_{\widehat{u}} q_{\widehat{U}|V}(\widehat{u}|v)\chi(\widehat{u} \neq u) \\ &= \sum_v p_V(v) \sum_u p_{U|V}(u|v)[1 - q_{\widehat{U}|V}(u|v)]. \end{aligned} \tag{33}$$

In the last equality, $1 - q_{\widehat{U}|V}(u|v)$ corresponds to the error probability of the decision rule $q_{\widehat{U}|V}$ after the observation $v \in \mathcal{V}$, and $\mathrm{Error}(q_{\widehat{U}|V})$ corresponds to the average of this error probability. When $q_{\widehat{U}|V}$ is defined by using $f : \mathcal{V} \to \mathcal{U}$ and (32), the decision error probability $\mathrm{Error}(f)$ of a deterministic decision rule f is given as

$$\begin{aligned} \mathrm{Error}(f) &\equiv \sum_v p_V(v) \sum_u p_{U|V}(u|v)\chi(f(v) \neq u) \\ &= \sum_v p_V(v)[1 - p_{U|V}(f(v)|v)]. \end{aligned} \tag{34}$$

It should be noted that the right hand side of the first equality can be derived directly from (33) and the fact that

$$\begin{aligned} q_{\widehat{U}|V}(u|v) &= \chi(f(v) = u) \\ &= 1 - \chi(f(v) \neq u). \end{aligned} \tag{35}$$

That is, we have $\mathrm{Error}(f) = \mathrm{Error}(q_{\widehat{U}|V})$ when f and $q_{\widehat{U}|V}$ satisfy (32).

It is well-known fact (see [35, Lemma 1]) that an optimal strategy for guessing the state U is finding $\widehat{u}$ which maximize the conditional probability $p_{U|V}(\widehat{u}|v)$ depending on a given observation v. Formally, by taking $\widehat{u}$ that maximizes $p_{U|V}(\widehat{u}|v)$ for each $v \in \mathcal{V}$, we can define the function $f_{\mathrm{MAP}} : \mathcal{V} \to \mathcal{U}$ as

$$f_{\mathrm{MAP}}(v) \equiv \arg\max_{\widehat{u}} p_{U|V}(\widehat{u}|v) \tag{36}$$

$$= \arg\max_{\widehat{u}} p_{UV}(\widehat{u}, v). \tag{37}$$

We call (36) and (37) a *maximum a posteriori decoder* and a *maximum likelihood decoder*, respectively. It should be noted that the discussion does not depend on the choice of states with the same maximum probability.

Here, let us consider the case $q_{\widehat{U}|V}(\widehat{u}|v) = p_{U|V}(\widehat{u}|v)$ for all $(\widehat{u}, v)$, that is, we make a stochastic decision with the conditional distribution $p_{U|V}$ of a state U for a given observation V. It should be noted that the joint distribution $p_{UV\widehat{U}}$ of $(U, V, \widehat{U})$ is given as

$$p_{UV\widehat{U}}(u, v, \widehat{u}) = p_{U|V}(\widehat{u}|v)p_{U|V}(u|v)p_V(v).$$

We call this type of decision rule a *stochastic decision with the a posteriori distribution*.

We have the following lemma. It should be noted that this lemma is presented in [23, Lemma 3][7] in the context of stochastic decoding of channel code.

Lemma 4 ([10, Eq. (29)] [35, Lemma 3]): Let (U, V) be a pair consisting of a state U and an observation V and p_{UV} be the joint distribution of (U, V). When we make a stochastic decision with $p_{U|V}$, the decision error probability of this rule is at most twice the decision error probability of the maximum a posteriori decision rule f_{MAP}. That is, we have

$$\mathrm{Error}(p_{U|V}) \leq 2\mathrm{Error}(f_{\mathrm{MAP}}).$$

Proof: In this proof, we assume that $\mathcal{U}$ and $\mathcal{V}$ are finite sets. It should be noted that the result does not change when $\mathcal{V}$ is an infinite/continuous set, where the summation should be replaced with the integral. We have

$$\begin{aligned} &\mathrm{Error}(p_{U|V}) \\ &= \sum_v p_V(v) \sum_u p_{U|V}(u|v)[1 - p_{U|V}(u|v)] \\ &= \sum_v p_V(v) \left[1 - \sum_u p_{U|V}(u|v)^2\right] \\ &\leq \sum_v p_V(v) \left[1 - p_{U|V}(f_{\mathrm{MAP}}(v)|v)^2\right] \\ &= \sum_v p_V(v)[1 - p_{U|V}(f_{\mathrm{MAP}}(v)|v)][1 + p_{U|V}(f_{\mathrm{MAP}}(v)|v)] \\ &\leq 2\sum_v p_V(v)[1 - p_{U|V}(f_{\mathrm{MAP}}(v)|v)] \\ &= 2\mathrm{Error}(f_{\mathrm{MAP}}), \end{aligned} \tag{38}$$

where the second inequality comes from the fact that $p_{U|V}(f(v)|v) \leq 1$ and the fourth equality comes from (34). ■

ACKNOWLEDGMENT

The authors thank the anonymous reviewers and Associate Editor Prof. V. Vayshampayan for constructive comments and suggestions. They have significantly improved the presentation of the results.

[7]Lemma 3 indicates the lemma not in the conference proceedings version but in the online version.

REFERENCES

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography—Part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.

[2] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[3] J. Bajcsy and P. Mitran, "Coding for the Slepian–Wolf problem with turbo codes," in *Proc. IEEE Globecom*, Nov. 2001, pp. 1400–1404.

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

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

[6] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA, USA: Addison-Welsey, 1987.

[7] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *J. Comput. Syst. Sci.*, vol. 18, no. 2, pp. 143–154, Apr. 1979.

[8] J. Chen, D.-K. He, A. Jagmohan, L. A. Lastras-Montaño, and E.-H. Yang, "On the linear codebook-level duality between Slepian–Wolf coding and channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5575–5590, Dec. 2009.

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

[10] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

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

[12] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[13] P. Elias, "Coding for noisy channels," *IRE Int. Conv. Rec.*, vol. 3, pp. 37–46, 1955.

[14] U. Erez and G. Miller, "The ML decoding performance of LDPC ensembles over $\mathbb{Z}_q$," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1871–1879, May 2005.

[15] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.

[16] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA, USA: MIT Press, 1963.

[17] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, Oct. 2001.

[18] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[19] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer, 2003.

[20] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[21] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, Jun./Jul. 2009, pp. 1488–1492.

[22] R. Impagliazzo and D. Zuckerman, "How to recycle random bits," in *Proc. 30th IEEE Symp. Fund. Comput. Sci.*, Oct./Nov. 1989, pp. 248–253.

[23] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, and R. Urbanke, "Comparing the bit-MAP and block-MAP decoding thresholds of Reed–Muller codes on BMS channels," [Online]. Available: https://arxiv.org/abs/1602.06048

[24] S. Miyake and J. Muramatsu, "A construction of channel code, JSCC and universal code for discrete memoryless channels using sparse matrices," *IEICE Trans. Fundam.*, vol. E92-A, no. 9, pp. 2333–2344, Sep. 2009.

[25] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. Fundamentals*, vols. E78-A, no. 9, pp. 1063–1070, 1995.

[26] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "How to achieve the capacity of asymmetric channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3371–3393, May 2018.

[27] J. Muramatsu, "Applications of Slepian–Wolf source coding," in *Proc. Workshop Concepts Inform. Theory*, Viareggio, Italy, Oct. 2004, pp. 20–23.

[28] J. Muramatsu, "Channel coding and lossy source coding using a generator of constrained random numbers," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2667–2686, May 2014.

[29] J. Muramatsu, "Variable-length lossy source code using a constrained-random-number generator," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3574–3592, Jun. 2015.

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

[31] J. Muramatsu and S. Miyake. (Jun. 2010). "Construction of Slepian–Wolf source code and broadcast channel code based on hash property." [Online]. Available: https://arxiv.org/abs/1006.5271

[32] J. Muramatsu and S. Miyake, "Construction of broadcast channel code based on hash property," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 575–579.

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

[34] J. Muramatsu and S. Miyake, "Fundamental limits are achievable with countable alphabet," in *Proc. Int. Symp. Inf. Theory Appl.*, Monterey, CA, USA, Oct./Nov. 2016, pp. 573–577.

[35] J. Muramatsu and S. Miyake, "On the error probability of stochastic decision and stochastic decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 1643–1647. [Online]. Available: https://arxiv.org/abs/1701.04950

[36] J. Muramatsu and S. Miyake, "Multi-terminal codes using constrained-random-number generators," in *Proc. Int. Symp. Inf. Theory Appl.*, Singapore, Oct. 2018, pp. 612–616. [Online]. Available: https://arxiv.org/abs/1801.02875

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

[38] T. Murayama, "Statistical mechanics of the data compression theorem," *J. Phys. A, Math. Gen.*, vol. 35, pp. L95–L100, Feb. 2002.

[39] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7377–7385, Nov. 2011.

[40] D. Schonberg, S. S. Pradhan, and K. Ramchandran, "LDPC codes can approach the Slepian Wolf bound for general binary sources," in *Proc. 40th Annu. Allerton Conf. Commun., Control, Comput.* Monticello, IL, USA: Allerton House, Oct. 2002, pp. 576–585.

[41] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[42] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, Apr. 1971.

[43] Y. Steinberg and S. Verdú, "Channel simulation and coding with side information," *IEEE Trans. Inf. Theory*, vol. 40, no. 3, pp. 634–646, May 1994.

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

[45] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 1, pp. 2–10, Jan. 1974.

[46] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, Nov. 2014.

Jun Muramatsu received the B.S. and M.S. degrees in mathematics and the Ph.D. degree from Nagoya University, Aichi, Japan, in 1990, 1992, and 1998, respectively. He joined NTT Laboratories in 1992. At NTT, he has been engaged in research on information theory. From Feb. 2007 to Feb. 2008, he was a visiting researcher in ETH, Zurich, Switzerland. He is currently a Research Scientist in NTT Communication Science Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and the IEEE Information Theory Society. During 2006–2010, he was an associate editor of IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. He received the Young Researcher Award of SITA (the Society of Information Theory and Its Application) in 2003 and the 63rd Best Paper Award of IEICE in 2007.

Shigeki Miyake received his B.E. and M.E. degrees in Physical Engineering and Ph.D. degree in science from Tokyo University, Tokyo, Japan, in 1987, 1989, and 2010, respectively. He joined NTT Laboratories in 1989. He has been engaged in research on information theory and its application except from 1998 to 2004 when he worked in business division of NTT. He is currently a Research Engineer in NTT Network Innovation Laboratories, Kanagawa Japan. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and the IEEE Information Theory Society.