ISIT'2013 1569717873

Lossy Source 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 lossy source coding is introduced with a rate-distortion pair close to the boundary of the rate-distortion region, where the only restriction is that the reproduction 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 rate-distortion region. 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 lossy code for general sources including additive Gaussian, Markov, and non-stationary sources. The only assumption is that the reproduction alphabet is finite. We prove that the rate-distortion region 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 [13] for the constrained random number generator by using the sum-product algorithm [1][8].

One contribution of this paper is to extend the results of [10] to general sources. In [10], the direct part of the rate-distortion theorem for a discrete stationary memoryless source is shown based on the hash property of an ensemble of functions. In this paper, an alternative general formula for the rate-distortion region is derived and the achievability of the proposed code is proved based on a stronger version of hash property introduced in [11][12]. 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) source by using sparse matrices and the constrained random number generator instead of the apparently intractable deterministic encoder presented in [10]. There are many ways to construct lossy source codes [5][14][9][16] by using sparse matrices. These approaches assume that a source is discrete stationary memoryless and symmetric, or a quantization map is used for an asymmetric source. On the other hand, the only requirement for the proposed code is that the reproduction alphabet is finite.

It should be noted that a similar idea has appeared in [17], where they introduced random bin coding and Slepian-Wolf decoding[1] 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 a practical construction of encoding function and theorem is proved simply and rigorously based on the technique reported in [12], where it is proved that we can use sparse matrices for the construction of the code.

II. General Formulas for Rate-Distortion Region

This section provides a formal description of the problem and formulas for the rate distortion region. All the results in this paper are presented by using the information spectrum method introduced in [6][7], where the consistency and stationarity of sources are not assumed. It should be noted that all the results reported in this paper can be applied to stationary ergodic sources and stationary memoryless sources. 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 sup-entropy rate $\overline{H}(\boldsymbol{U})$ as

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

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

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

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

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

In the following, we introduce the achievable rate-distortion region for a general source. Let $\mathcal{Y}^n$ be a source alphabet and $\mathcal{X}^n$ be a reproduction alphabet[2]. Let $d_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, \infty)$

[1] It should be noted that the idea of using Slepian-Wolf decoding has already been mentioned in [10].

[2] It should be noted that the roles of $\mathcal{X}^n$ and $\mathcal{Y}^n$ are the reverse of those in the conventional definition of the rate-distortion theory.

be a distortion function.

Definition 1 ([7, Def. 5.3.1]): We call a pair (R, D) consisting of a rate R and a distortion D *achievable* if for all $\delta > 0$ and all sufficiently large n there is a pair consisting of an encoder $\varphi_n : \mathcal{Y}^n \to \mathcal{M}_n$ and a decoder $\psi_n : \mathcal{M}_n \to \mathcal{X}^n$ such that

$$\frac{1}{n} \log |\mathcal{M}_n| \le R \tag{1}$$

$$P\left(d_n(\psi_n(\varphi_n(Y^n)), Y^n) > D\right) \le \delta, \tag{2}$$

where $\mathcal{M}_n$ is a set of codewords. The *achievable rate-distortion region* $\mathcal{R}(\boldsymbol{Y})$ is defined by the set of all achievable pairs (R, D).

For a pair $(\boldsymbol{X}, \boldsymbol{Y})$ of general sources, let $\overline{D}(\boldsymbol{X}, \boldsymbol{Y})$ be defined as

$$\overline{D}(\boldsymbol{X}, \boldsymbol{Y}) \equiv \inf\left\{\theta : \lim_{n\to\infty} P\left(d_n(X^n, Y^n) > \theta\right) = 0\right\}.$$

For a general source $\boldsymbol{Y}$, the rate-distortion region $\mathcal{R}(\boldsymbol{Y})$ is derived in [15][7, Theorem 5.4.1][3] as

$$\mathcal{R}(\boldsymbol{Y}) = \bigcup_{\boldsymbol{W}} \left\{(R, D) : \begin{array}{c} \overline{I}(\boldsymbol{X}; \boldsymbol{Y}) \le R \\ \overline{D}(\boldsymbol{X}, \boldsymbol{Y}) \le D \end{array}\right\}, \tag{3}$$

where the union is taken over all general channels $\boldsymbol{W} \equiv \{\mu_{X^n|Y^n}\}_{n=1}^\infty$ and the joint distribution $\mu_{X^nY^n}$ is given as

$$\mu_{X^nY^n}(\boldsymbol{x}, \boldsymbol{y}) \equiv \mu_{X^n|Y^n}(\boldsymbol{x}|\boldsymbol{y})\mu_{Y^n}(\boldsymbol{y}). \tag{4}$$

We introduce the following lemma, which is proved in Section V-A.

Lemma 1: For a general source $\boldsymbol{Y}$,

$$\mathcal{R}(\boldsymbol{Y}) = \bigcup_{\boldsymbol{W}} \left\{(R, D) : \begin{array}{c} \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}|\boldsymbol{Y}) \le R \\ \overline{D}(\boldsymbol{X}, \boldsymbol{Y}) \le D \end{array}\right\}, \tag{5}$$

where the union is taken over all channels $\boldsymbol{W}$ and the joint distribution of $(\boldsymbol{X}, \boldsymbol{Y})$ is given as (4).

In this paper, we construct a fixed-rate lossy source code, where (R, D) is close to the boundary of the region given by the right hand side of (5). The constructed code is given by a pair consisting of a stochastic encoder $\Phi_n : \mathcal{Y}^n \to \mathcal{M}_n$ and a decoder $\psi_n : \mathcal{M}_n \to \mathcal{X}^n$. It should be noted that formulas (3) and (5) are satisfied when a stochastic encoder is allowed. In fact, by considering the average over the stochastic encoders and using the random coding argument, we can construct a deterministic encoder from a stochastic encoder without any loss of encoding rate. Thus the rate-distortion pair of the stochastic encoder should be in the rate-distortion region. On the other hand, the rate-distortion region is achievable with a stochastic encoder because a deterministic encoder is one type of stochastic encoder. It should also be noted that we have a similar result that is obtained in this paper by assuming $\max_{n,\boldsymbol{x},\boldsymbol{y}} d_n(\boldsymbol{y}, \boldsymbol{x}) < \infty$, where (2) is replaced by the average distortion criterion

$$E_{Y^n}\left[d_n(\psi_n(\varphi_n(Y^n)), Y^n)\right] \le D + \delta.$$

[3]The rate-distortion function, which is the infimum of R such that (R, D) is achievable for a given D, is derived in [15][7, Theorem 5.4.1].

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

In this section, we introduce the hash property[4] introduced in [11][12] 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 $\mathbf{c}$, 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 [10].

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

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

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

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

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

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

for any $\mathcal{T}, \mathcal{T}' \subset \mathcal{U}^n$, which is introduced in [10].

It should be noted that when $\mathcal{A}$ is a two-universal class of hash functions [4] and p_{A} is the uniform distribution on $\mathcal{A}$, then $(\boldsymbol{\mathcal{A}}, \boldsymbol{p}_{\mathsf{A}})$ has a $(\mathbf{1}, \mathbf{0})$-hash property, where $\mathbf{1}$ and $\mathbf{0}$ denote the constant sequences of 1 and 0, respectively. Random bin coding [2] and the set of all linear functions [3] are examples of the two-universal class of hash functions. It is proved in

[4]In [11] [12], it is called the 'strong hash property.' Throughout this paper, we call it simply the 'hash property.'

[11, Section III-B] that an ensemble of sparse matrices has a (strong) hash property.

Lemma 2 ([11, 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. $B : \mathcal{U}^n \to \mathrm{Im}\mathcal{B}$). 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 ([10, 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 ([12, Lemma 4]): If $(\mathcal{A}, p_{\mathsf{A}})$ satisfies (H3), then

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

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 Lossy Source Code

This section introduces a lossy source code. We assume that a reproduction alphabet $\mathcal{X}^n$ is finite set but a source alphabet $\mathcal{Y}^n$ is allowed to be arbitrary (infinite, continuous) set.

For given $r >$ and $R > 0$, let $(\boldsymbol{\mathcal{A}}, p_{\mathsf{A}})$ and $(\boldsymbol{\mathcal{B}}, p_{\mathsf{B}})$ 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 $\mathcal{M}_n \equiv \log|\mathrm{Im}\mathcal{B}|$. 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 μ_{Y^n} of a source Y^n and a conditional distribution $\mu_{X^n|Y^n}$ are given. Let μ_{X^n} be defined as

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

We use a constrained random number generator to construct an encoder. Let $\widetilde{X}^n \equiv \widetilde{X}^n_A(\boldsymbol{c}|\boldsymbol{y})$ be a random variable corresponding to the distribution

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

where the encoder generates $\boldsymbol{x}$ satisfying $A\boldsymbol{x} = \boldsymbol{c}$ with probability $\nu_{\widetilde{X}^n|Y^n}(\boldsymbol{x}|\boldsymbol{y})$ for a given source output $\boldsymbol{y}$. We define the stochastic encoder $\Phi_n : \mathcal{Y}^n \to \mathrm{Im}\mathcal{B}$ as

$$\Phi_n(\boldsymbol{y}) \equiv \begin{cases} B\widetilde{X}^n_A(\boldsymbol{c}|\boldsymbol{y}) & \text{if } \mu_{X^n|Y^n}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y}) > 0 \\ \text{encoding error} & \text{if } \mu_{X^n|Y^n}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y}) = 0. \end{cases}$$

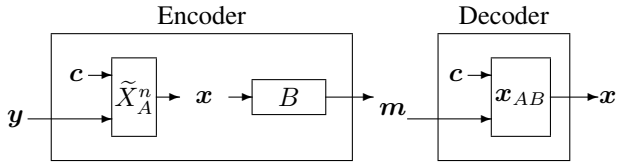

Fig. 1. Construction of Lossy Source Code

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

$$\psi_n(\boldsymbol{m}) \equiv \boldsymbol{x}_{AB}(\boldsymbol{c}, \boldsymbol{m}),$$

where $\boldsymbol{x}_{AB}$ is defined as

$$\boldsymbol{x}_{AB}(\boldsymbol{c}, \boldsymbol{m}) \equiv \arg\max_{\boldsymbol{x}' \in \mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m})} \mu_{X^n}(\boldsymbol{x}').$$

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 $\boldsymbol{x} \in \mathcal{C}_A(\boldsymbol{c})$ is assigned[5] corresponding to a source output $\boldsymbol{y}$.

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

$$\mathrm{Error}(A, B, \boldsymbol{c}, D) \equiv P\left(d_n(\psi_n(\Phi_n(Y^n)), Y^n) > D\right), \quad (6)$$

where we define $d_n(\psi_n(\Phi_n(\boldsymbol{y})), \boldsymbol{y}) \equiv \infty$ when $\mu_{X^n|Y^n}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y}) = 0$. We have the following theorem, which is shown in Section V-B.

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

$$r < \underline{H}(\boldsymbol{X}|\boldsymbol{Y}) \quad (7)$$
$$r + R > \overline{H}(\boldsymbol{X}) \quad (8)$$

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

$$\mathrm{Error}(A, B, \boldsymbol{c}, D) \leq P\left(d_n(X^n, Y^n) > D\right) + \delta. \quad (9)$$

By assuming that $\{\mu_{X^n|Y^n}\}_{n=1}^{\infty}$ satisfies $\overline{D}(\boldsymbol{X}, \boldsymbol{Y}) < D$, we have the fact that $\lim_{n\to\infty} P\left(d_n(X^n, Y^n) > D\right) = 0$ from the definition of $\overline{D}(\boldsymbol{X}, \boldsymbol{Y})$. Then, by letting $n \to \infty$, $\delta \to 0$, and $r \to \underline{H}(\boldsymbol{X}|\boldsymbol{Y})$, we have the fact that for any (R, D) close to the boundary of $\mathcal{R}(\boldsymbol{Y})$ there is a sequence of proposed codes such that $\lim_{n\to\infty} \mathrm{Error}(A, B, \boldsymbol{c}, D) = 0$.

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

V. PROOFS

A. Proof of Lemma 1

We use the following lemma in the proof.

Lemma 5 ([7, Lemma 2.6.2]): If a general source $\widehat{\boldsymbol{X}}$ satisfies $|\{\boldsymbol{x} \in \mathcal{X}^n : \mu_{\widehat{X}^n}(\boldsymbol{x}) > 0\}| \le M_n$ for a positive number M_n, we have

$$P\left(\frac{1}{n}\log\frac{1}{\mu_{\widehat{X}^n}(\widehat{X}^n)} \ge \frac{1}{n}\log M_n + \varepsilon\right) \le 2^{-n\varepsilon}$$

for an arbitrary constant $\varepsilon > 0$.

Now we prove Lemma 1. Since $\overline{I}(\boldsymbol{X};\boldsymbol{Y}) \le \overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}|\boldsymbol{Y})$ for any $(\boldsymbol{X},\boldsymbol{Y})$, we have

$$\begin{aligned}\mathcal{R}(\boldsymbol{Y}) &= \bigcup_{\boldsymbol{W}}\left\{(R,D) : \begin{array}{c}\overline{I}(\boldsymbol{X};\boldsymbol{Y}) \le R \\ \overline{D}(\boldsymbol{X},\boldsymbol{Y}) \le D\end{array}\right\} \\ &\supset \bigcup_{\boldsymbol{W}}\left\{(R,D) : \begin{array}{c}\overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}|\boldsymbol{Y}) \le R \\ \overline{D}(\boldsymbol{X},\boldsymbol{Y}) \le D\end{array}\right\}.\end{aligned}$$

In the following, we prove that

$$\mathcal{R}(\boldsymbol{Y}) \subset \bigcup_{\boldsymbol{W}}\left\{(R,D) : \begin{array}{c}\overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}|\boldsymbol{Y}) \le R \\ \overline{D}(\boldsymbol{X},\boldsymbol{Y}) \le D\end{array}\right\}, \quad (10)$$

which completes the proof of the lemma.

Assume that $(R,D) \in \mathcal{R}(\boldsymbol{Y})$. From (3), we have the fact that for all $\delta > 0$ and all sufficiently large n, there is a pair consisting an encoder φ_n and a decoder ψ_n satisfying (1) and (2). Let $\widehat{X}^n \equiv \psi_n(\varphi_n(Y^n)) \in \mathcal{X}^n$. Then we have

$$\begin{aligned}&P\left(\frac{1}{n}\log\frac{1}{\mu_{\widehat{X}^n}(\widehat{X}^n)} > R + \varepsilon\right) \\ &\le P\left(\frac{1}{n}\log\frac{1}{\mu_{\widehat{X}^n}(\widehat{X}^n)} \ge \frac{1}{n}\log|\mathcal{M}_n| + \varepsilon\right) \\ &\le 2^{-n\varepsilon} \end{aligned} \quad (11)$$

for any $\varepsilon > 0$, where the first inequality comes from (1), and the second inequality comes from Lemma 5 and the fact that the cardinality of the range of $\widehat{X}^n$ is at most $|\mathcal{M}_n|$. By letting $n \to \infty$, we have the fact that a general source $\widehat{\boldsymbol{X}} \equiv \{\psi_n(\varphi_n(Y^n))\}_{n=1}^{\infty}$ satisfies

$$\begin{aligned}\overline{H}(\widehat{\boldsymbol{X}}) - \underline{H}(\widehat{\boldsymbol{X}}|\boldsymbol{Y}) &\le \overline{H}(\widehat{\boldsymbol{X}}) \\ &\le R + \varepsilon. \end{aligned} \quad (12)$$

By letting $\varepsilon \to 0$, we have

$$\overline{H}(\widehat{\boldsymbol{X}}) - \underline{H}(\widehat{\boldsymbol{X}}|\boldsymbol{Y}) \le R.$$

On the other hand, we have

$$\lim_{n\to\infty} P\left(d_n(\widehat{X}^n, Y^n) > D\right) = 0$$

from (2) by letting $n \to \infty$ and $\delta \to 0$. This implies that

$$\overline{D}(\widehat{\boldsymbol{X}},\boldsymbol{Y}) \le D.$$

Then we have

$$(R,D) \in \bigcup_{\boldsymbol{W}}\left\{(R,D) : \begin{array}{c}\overline{H}(\boldsymbol{X}) - \underline{H}(\boldsymbol{X}|\boldsymbol{Y}) \le R \\ \overline{D}(\boldsymbol{X},\boldsymbol{Y}) \le D\end{array}\right\},$$

which implies (10). ■

B. Proof of Theorem 1

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

From (7) and (8), we have the fact that there is $\varepsilon > 0$ satisfying

$$r < \underline{H}(\boldsymbol{X}|\boldsymbol{Y}) - \varepsilon \quad (13)$$

$$r + R > \overline{H}(\boldsymbol{X}) + \varepsilon. \quad (14)$$

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

$$\begin{aligned}\overline{\mathcal{T}}_X &\equiv \left\{\boldsymbol{x} : \frac{1}{n}\log\frac{1}{\mu_X(\boldsymbol{x})} \le \overline{H}(\boldsymbol{X}) + \varepsilon\right\} \\ \underline{\mathcal{T}}_{X|Y} &\equiv \left\{(\boldsymbol{x},\boldsymbol{y}) : \frac{1}{n}\log\frac{1}{\mu_{X|Y}(\boldsymbol{x}|\boldsymbol{y})} \ge \overline{H}(\boldsymbol{X}|\boldsymbol{Y}) - \varepsilon\right\}.\end{aligned}$$

Assume that $\boldsymbol{x} \in \overline{\mathcal{T}}_X$ and $\boldsymbol{x}_{AB}(A\boldsymbol{x}, B\boldsymbol{x}) \ne \boldsymbol{x}$. Then we have the fact that there is $\boldsymbol{x}' \in \mathcal{C}_{AB}(A\boldsymbol{x}, B\boldsymbol{x})$ such that $\boldsymbol{x}' \ne \boldsymbol{x}$ and

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

This implies that $\left[\overline{\mathcal{T}}_X \setminus \{\boldsymbol{x}\}\right] \cap \mathcal{C}_{AB}(A\boldsymbol{x}, B\boldsymbol{x}) \ne \emptyset$. Then we have

$$\begin{aligned}&E_{\mathsf{AB}}\left[\chi(\boldsymbol{x}_{\mathsf{AB}}(\mathsf{A}\boldsymbol{x},\mathsf{B}\boldsymbol{x}) \ne \boldsymbol{x})\right] \\ &\le p_{\mathsf{AB}}\left(\left\{(A,B) : \left[\overline{\mathcal{T}}_X \setminus \{\boldsymbol{x}\}\right] \cap \mathcal{C}_{AB}(A\boldsymbol{x}, B\boldsymbol{x}) \ne \emptyset\right\}\right) \\ &\le \frac{|\overline{\mathcal{T}}_X|\alpha_{\mathsf{AB}}}{|\mathrm{Im}\mathcal{A}|} + \beta_{\mathsf{AB}} \\ &\le 2^{-n[r-\overline{H}(\boldsymbol{X})-\varepsilon]}\alpha_{\mathsf{AB}} + \beta_{\mathsf{AB}}, \end{aligned} \quad (15)$$

where $\chi(\cdot)$ is the indicator function, the second inequality comes from Lemmas 2, 3, and the last inequality comes from the fact that $|\overline{\mathcal{T}}_X| \le 2^{n[\overline{H}(\boldsymbol{X})+\varepsilon]}$. We have the fact that

$$\begin{aligned}&E_{\mathsf{AB}}\left[\sum_{\boldsymbol{x}} \mu_X(\boldsymbol{x})\chi(\boldsymbol{x}_{\mathsf{AB}}(\mathsf{A}\boldsymbol{x},\mathsf{B}\boldsymbol{x}) \ne \boldsymbol{x})\right] \\ &= \sum_{\boldsymbol{x}\in\overline{\mathcal{T}}_X} \mu_X(\boldsymbol{x}) E_{\mathsf{AB}}\left[\chi(\boldsymbol{x}_{\mathsf{AB}}(\mathsf{A}\boldsymbol{x},\mathsf{B}\boldsymbol{x}) \ne \boldsymbol{x})\right] \\ &\quad + \sum_{\boldsymbol{x}\notin\overline{\mathcal{T}}_X} \mu_X(\boldsymbol{x}) E_{\mathsf{AB}}\left[\chi(\boldsymbol{x}_{\mathsf{AB}}(\mathsf{A}\boldsymbol{x},\mathsf{B}\boldsymbol{x}) \ne \boldsymbol{x})\right] \\ &\le 2^{-n[r+R-\overline{H}(\boldsymbol{X})-\varepsilon]}\alpha_{\mathsf{AB}} + \beta_{\mathsf{AB}} + \mu_X([\overline{\mathcal{T}}_X]^c), \end{aligned} \quad (16)$$

where the last inequality comes from (15). We also have the fact that

$$\begin{aligned}&E_{\mathsf{A}}\left[\sum_{\boldsymbol{c},\boldsymbol{y}} \mu_Y(\boldsymbol{y})\left|\mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})|\boldsymbol{y}) - \frac{1}{|\mathrm{Im}\mathcal{A}|}\right|\right] \\ &\le E_{\mathsf{A}}\left[\sum_{\boldsymbol{c},\boldsymbol{y}} \mu_{X|Y}(\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})\mu_Y(\boldsymbol{y}) \right. \\ &\qquad \left. \cdot \left|\frac{\mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c}) \cap \underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})}{\mu_{X|Y}(\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})} - \frac{1}{|\mathrm{Im}\mathcal{A}|}\right|\right] \\ &\quad + E_{\mathsf{A}}\left[\sum_{\boldsymbol{c},\boldsymbol{y}} \mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c}) \cap [\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})]^c|\boldsymbol{y})\mu_Y(\boldsymbol{y})\right] \\ &\quad + E_{\mathsf{A}}\left[\sum_{\boldsymbol{c},\boldsymbol{y}} \frac{\mu_{X|Y}([\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})]^c|\boldsymbol{y})\mu_Y(\boldsymbol{y})}{|\mathrm{Im}\mathcal{A}|}\right]\end{aligned}$$

$$
\begin{aligned}
&E_{\mathsf{ABc}}\left[\mathrm{Error}(\mathsf{A},\mathsf{B},\mathbf{c},D)\right] \\
&\le E_{\mathsf{AB}}\left[\sum_{\substack{\boldsymbol{c},\boldsymbol{y}:\\ \mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})|\boldsymbol{y})=0}} \frac{\mu_Y(\boldsymbol{y})}{|\mathrm{Im}\mathcal{A}|} + \sum_{\substack{\boldsymbol{c},\boldsymbol{x},\boldsymbol{y}:\\ \boldsymbol{x}\in\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})\\ \mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})|\boldsymbol{y})>0\\ d_n(\boldsymbol{x},\boldsymbol{y})>D \text{ or } \boldsymbol{x}_{\mathsf{AB}}(\mathsf{A}\boldsymbol{x},\mathsf{B}\boldsymbol{x})\neq\boldsymbol{x}}} \mu_{XY}(\boldsymbol{x},\boldsymbol{y})\left[1+\frac{1}{|\mathrm{Im}\mathcal{A}|\mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})|\boldsymbol{y})}-1\right]\right] \\
&\le P(d_n(X^n,Y^n)>D) + E_{\mathsf{AB}}\left[\sum_{\boldsymbol{x}}\mu_X(\boldsymbol{x})\chi(\boldsymbol{x}_{\mathsf{AB}}(\mathsf{A}\boldsymbol{x},\mathsf{B}\boldsymbol{x})\neq\boldsymbol{x})\right] + E_{\mathsf{A}}\left[\sum_{\boldsymbol{c},\boldsymbol{y}}\mu_Y(\boldsymbol{y})\left|\mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})|\boldsymbol{y})-\frac{1}{|\mathrm{Im}\mathcal{A}|}\right|\right] \\
&\le P(d_n(X^n,Y^n)>D) + 2^{-n[r+R-\overline{H}(\boldsymbol{X})-\varepsilon]}\alpha_{\mathsf{AB}} + \beta_{\mathsf{AB}} + \mu_X([\overline{\mathcal{T}}_X]^c) \\
&\quad + \sqrt{\alpha_{\mathsf{A}}-1+[\beta_{\mathsf{A}}+1]2^{-n[\underline{H}(\boldsymbol{X}|\boldsymbol{Y})-r-\varepsilon]}} + 2\mu_{XY}([\underline{\mathcal{T}}_{X|Y}]^c)
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
&= \sum_{\boldsymbol{y}}\mu_{X|Y}(\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})\mu_Y(\boldsymbol{y}) \\
&\quad\cdot E_{\mathsf{A}}\left[\sum_{\boldsymbol{c}}\left|\frac{\mu_{X|Y}(\mathcal{C}_{\mathsf{A}}(\boldsymbol{c})\cap\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})}{\mu_{X|Y}(\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})}-\frac{1}{|\mathrm{Im}\mathcal{A}|}\right|\right] \\
&\quad + 2\mu_{XY}([\underline{\mathcal{T}}_{X|Y}]^c) \\
&\le \sum_{\boldsymbol{y}}\mu_{X|Y}(\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})\mu_Y(\boldsymbol{y}) \\
&\quad\cdot\sqrt{\frac{\alpha_{\mathsf{A}}-1+[\beta_{\mathsf{A}}+1]|\mathrm{Im}\mathcal{A}|\max_{\boldsymbol{x}\in\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})}\mu_{X|Y}(\boldsymbol{x}|\boldsymbol{y})}{\mu_{X|Y}(\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y})|\boldsymbol{y})}} \\
&\quad + 2\mu_X([\underline{\mathcal{T}}_X]^c) \\
&\le \sqrt{\alpha_{\mathsf{A}}-1+[\beta_{\mathsf{A}}+1]2^{-n[\underline{H}(\boldsymbol{X}|\boldsymbol{Y})-r-\varepsilon]}} + 2\mu_{X|Y}([\underline{\mathcal{T}}_{X|Y}]^c),
\end{aligned}
\tag{17}
$$

where $\underline{\mathcal{T}}_{X|Y}(\boldsymbol{y}) \equiv \{\boldsymbol{x} : (\boldsymbol{x},\boldsymbol{y}) \in \underline{\mathcal{T}}_{X|Y}\}$ and the second inequality comes from Lemma 4. Then we have (18), 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 second inequality comes from the fact that

$$
\begin{aligned}
&\sum_{\substack{\boldsymbol{c},\boldsymbol{x},\boldsymbol{y}:\\ \boldsymbol{x}\in\mathcal{C}_A(\boldsymbol{c})\\ \mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})>0}}\mu_{XY}(\boldsymbol{x},\boldsymbol{y})\left[\frac{1}{|\mathrm{Im}\mathcal{A}|\mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})}-1\right] \\
&\le \sum_{\substack{\boldsymbol{c},\boldsymbol{y}:\\ \mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})>0}}\mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})\mu_Y(\boldsymbol{y}) \\
&\quad\cdot\left|\frac{1}{|\mathrm{Im}\mathcal{A}|\mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})}-1\right| \\
&= \sum_{\boldsymbol{c},\boldsymbol{y}}\mu_Y(\boldsymbol{y})\left|\mu_X(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})-\frac{1}{|\mathrm{Im}\mathcal{A}|}\right| \\
&\quad - \sum_{\substack{\boldsymbol{c},\boldsymbol{y}:\\ \mu_{X|Y}(\mathcal{C}_A(\boldsymbol{c})|\boldsymbol{y})=0}}\frac{\mu_Y(\boldsymbol{y})}{|\mathrm{Im}\mathcal{A}|},
\end{aligned}
\tag{19}
$$

and the third inequality comes from (16), (17). From (13), (14), (18) and the fact that $\alpha_{\mathsf{A}} \to 1$, $\beta_{\mathsf{A}} \to 0$, $\alpha_{\mathsf{AB}} \to 1$, $\beta_{\mathsf{AB}} \to 0$, $\mu_X([\underline{\mathcal{T}}_X]^c) \to 0$, $\mu_{XY}([\underline{\mathcal{T}}_{X|Y}]^c) \to 0$ 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 (9). ■

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] 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.

[3] 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.

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

[5] A. Gupta and S. Verdú, "Nonlinear sparse-graph codes for lossy compression," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 1961–1975, May 2009.

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

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

[14] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Phys. Rev. E*, vol. 69, no. 035105(R), 2004.

[15] Y. Steinberg and S. Verdú, "Simulation of random process and rate-distortion theory," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 63–86, Jan. 1996.

[16] M. Wainwright, and E. Martinian, "Low density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inform. Theory*, vol. IT-55, no. 3, pp. 1061–1079, Mar. 2009.

[17] 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.