PAPER Special Section on Information Theory and Its Applications

Upper Bounds on the Error Probability for the Ensemble of Linear Block Codes with Mismatched Decoding*

Toshihiro NIINOMI[†a)], **Hideki YAGI**[††], ***Members***, ***and*** **Shigeichi HIRASAWA**[†††], ***Fellow***

SUMMARY In channel decoding, a decoder with suboptimal metrics may be used because of the uncertainty of the channel statistics or the limitations of the decoder. In this case, the decoding metric is different from the actual channel metric, and thus it is called mismatched decoding. In this paper, applying the technique of the DS2 bound, we derive an upper bound on the error probability of mismatched decoding over a regular channel for the ensemble of linear block codes, which was defined by Hof, Sason and Shamai. Assuming the ensemble of random linear block codes defined by Gallager, we show that the obtained bound is not looser than the conventional bound. We also give a numerical example for the ensemble of LDPC codes also introduced by Gallager, which shows that our proposed bound is tighter than the conventional bound. Furthermore, we obtain a single letter error exponent for linear block codes.

key words: *DS2 bound, ensemble of linear block codes, mismatched decoding, regular channel, single letter exponent*

1. Introduction

In practical situations, we need to use a decoder with suboptimal metrics because of the uncertainty of the channel statistics or the limitations of the decoder. In such a case, the decoding metric is different from the actual channel metric. Such decoding is called mismatched decoding, and several studies have been conducted form various perspectives.

Among these studies, Shamai and Sason [2] derived an upper bound on the error probability of the mismatched decoding for an ensemble of linear codes. In their paper, they considered the linear code ensemble which was generated from uniformly interleaved turbo codes. Then, the upper bound on the error probability was derived using the average weight distribution and Duman and Salehi's type 2 (DS2) upper bounding method.

On the other hand, Hof, Sason and Shamai [3], [4] derived upper bounds on the error probability for non-binary linear block codes. In contrast to Shamai and Sason [2], they considered another ensemble of linear codes and the regular channels introduced in [5]. In their derivation, the DS2 bounding in [6] is also applied as a key technique.

In this paper, we derive an upper bound on the error probability on mismatched decoding for the ensemble of linear block codes defined in [3]. Due to the property of this ensemble and the idea of the DS2 bounding, that is, choosing a nonnegative function $g(\cdot)$, we derive the new upper bound. For the ensemble of random linear block codes defined by Gallager [7], we show that the optimal function $g(\cdot)$ in the sense of minimizing the upper bound depends on the mismatched decoding metric, but not the actual channel, and the obtained bound is not looser than the conventional bound [2]. With the newly derived bound, we show a numerical example of the bound for the ensemble of LDPC codes introduced by Gallager, indicating that the new bound is tighter than the conventional bound. We also derive exponential upper bounds similar to those derived by Shulman and Feder [17].

This paper is organized as follows: in Sect. 2 we define the regular channel and the ensemble of linear block codes which is called the HSS ensemble. We review the conventional upper bound on the error probability for linear block codes in Sect. 3. Then, the proposed upper bounds and its numerical examples are presented in Sect. 4. In Sect. 5, the conclusions are presented.

2. Preliminaries

2.1 Regular Channel

In this paper, a class of memoryless symmetric channels, introduced by Delsarte and Piret [5] and extended by Hof et al. [3], is assumed. Let $x \in A$ be a channel input, and $y \in B$ be a channel output, where A and B are assumed to be finite**. Let $P = \{p(y|x) \mid x \in A,\ y \in B\}$ be the channel matrix of a discrete memoryless channel, where $p(y|x)$ is the transition probability from x to y. We assume that the channel input alphabet $A = \{0, 1, \cdots, q-1\}$ forms an Abelian group under the addition operation, and there exists a function $\tau : B \times A \to B$ satisfying the following conditions:

1) For every $x \in A$, the function $\tau(\cdot, x) : B \to B$ is bijective.

Manuscript received January 26, 2021.
Manuscript revised June 16, 2021.
Manuscript publicized October 8, 2021.

†The author is with Tokyo City University, Tokyo, 158-8557 Japan.
††The author is with the University of Electro-Communications, Chofu-shi, 182-8585 Japan.
†††The author is with Waseda University, Tokyo, 169-8555 Japan.

*The contents of this paper is partly presented at International Symposium on Information Theory and its Applications (ISITA 2020) [1]. This work is supported in part by JSPS KAKENHI Grant Numbers, JP17K0020, JP18H01438, JP20K04462.

a) E-mail: niinomi@cs.tcu.ac.jp

DOI: 10.1587/transfun.2021TAP0001

**This assumption is just due to the notational simplicity. All the results of this paper can be extended to the case of general B as is argued in [3].

2) For every $x_1, x_2 \in A$ and $y \in B$, $p(y|x_1) = p(\tau(y, x_2 - x_1)|\ x_2)$ holds.

We call such a discrete memoryless channel the regular channel [4]. Any memoryless binary input output symmetric (MBIOS) channel and memoryless additive noise channel are regular channels. We consider another regular channel $W = \{w(y|x) \mid x \in A,\ y \in B\}$ which indicates a mismatched metric. Although a mismatched metric is not necessarily a probability measure nor a regular channel, we assume that W is a regular channel.

2.2 Weight Distribution and Ensembles of Block Codes

Complete Weight Distribution and Hamming Weight Distribution

We consider a block code C consisting of M code words whose block length is N and rate is $R = \frac{\ln M}{N}$ (nat/symbol). Let $S_{\mathbf{t}}$ be the number of code words whose type is $\mathbf{t} = (t_0, t_1, \cdots, t_{q-1})$ with $t_i \geq 0$ for $i = 0, 1, \cdots, q-1$ and $\sum_{i=0}^{q-1} t_i = N$, where t_i denotes the number of code symbols equal to $i \in A$. We call $\{S_{\mathbf{t}}\}$ the complete weight distribution of the block code C. Let a_ℓ be the number of code words whose Hamming weight is ℓ. We call $\{a_\ell\}$ the Hamming weight distribution of the block code C. We consider the ensemble of block codes as below.

The Ensemble of Linear Block Codes (HSS ensemble)

In [3], Hof, Sason and Shamai considered an ensemble of (N, k) linear block codes. For this ensemble, they assumed the probability that a sequence is a code word depends only on its Hamming weight. Then, since the number of sequences with Hamming weight ℓ is $\begin{pmatrix} N \\ \ell \end{pmatrix}(q-1)^\ell$,

$$\mathbb{E}[a_\ell] = P(\ell)\begin{pmatrix} N \\ \ell \end{pmatrix}(q-1)^\ell \tag{1}$$

holds, where $\mathbb{E}[\cdot]$ denotes the expected value over the ensemble and $P(\ell)$ denotes the probability that a sequence of Hamming weight ℓ is a codeword in a randomly selected codebook from the ensemble. So, all sequences of a same type are chosen as code words with an equal probability, because sequences of the same type have the same Hamming weight. That is, if Hamming weight of a sequence of type $\mathbf{t}$ is ℓ,

$$\mathbb{E}[S_{\mathbf{t}}] = P(\ell)\begin{pmatrix} N \\ \mathbf{t} \end{pmatrix}, \tag{2}$$

holds, where we define

$$\begin{pmatrix} N \\ \mathbf{t} \end{pmatrix} = \frac{N!}{t_0! t_1! \cdots t_{q-1}!}$$

and

$$\mathbb{E}[a_\ell] = \sum_{\mathbf{t}: N - t_0 = \ell} \mathbb{E}[S_{\mathbf{t}}].$$

We call this ensemble the HSS ensemble. We can give two examples for the HSS ensemble. One is the ensemble of random linear block codes [7], and the other is the ensemble of regular LDPC codes by Gallager [7].

Later, we will show numerical examples for the ensemble of regular LDPC codes. In order to obtain a numerical example, we use the relationship between the average complete weight distribution for the ensemble of regular LDPC codes and $P(\ell)$. It was first derived by Hof et al [4]. The ensemble of the (c, h) regular LDPC codes by Gallager [7] can be defined as the ensemble in which the parity check matrix is filled with c nonzero elements in each column and h nonzero elements in each row. This ensemble is obtained by the following operations:

1. Construct a matrix D (which is called the base matrix) whose Hamming weight of each row is h and that of column is one. Then non-zero elements are randomly selected from $q-1$ input letters other than 0.
2. Create $c-1$ matrices $H_1, \cdots, H_{c-1}$ by randomly permuting the columns of D, and replacing non-zero elements with $q-1$ input letters other than 0.
3. Allocate matrices $H_1, \cdots, H_{c-1}$ to the bottom of D to make a parity check matrix

$$H = \begin{bmatrix} D \\ H_1 \\ \vdots \\ H_{c-1} \end{bmatrix}.$$

For this code ensemble, the following lemma has been derived:

Lemma 2.1 (Hof et al. [3]) We consider the q-ary (c, h) regular LDPC ensemble of Gallager. Then, $P(\ell)$ is given by

$$P(\ell) = \left[\frac{V_\ell}{\begin{pmatrix} N \\ \ell \end{pmatrix}(q-1)^\ell}\right]^c,\ 2 \leq \ell \leq N,$$

where each V_ℓ satisfies

$$\sum_{2 \leq \ell \leq N} V_\ell X^\ell = (V^*(X))^{\frac{N}{h}},$$

$$V^*(X) = 1 + \frac{1}{q}\sum_{i=2}^{h}\left((q-1)^i + (q-1)(-1)^i\right)\begin{pmatrix} h \\ i \end{pmatrix}X^i.$$

From Lemma 2.1 and (1), the average Hamming weight distribution for the ensemble of regular LDPC codes $\{\ \mathbb{E}[a_\ell]\ \}$ can be obtained. In addition, from Lemma 2.1 and (2), the average complete weight distribution $\{\ \mathbb{E}[S_{\mathbf{t}}]\ \}$ can be also obtained.

We also use the average complete weight distribution of

which the code words of small Hamming weight are expurgated as in the same manner in [4]. We review this average complete weight distribution of the expurgated ensemble in the following lemma:

Lemma 2.2 (Hof et al. [3]) Consider an expurgation of the codebooks whose minimum Hamming distance is not larger than D_N. Assume that the average complete weight distribution of the non-expurgated ensemble satisfies

$$\sum_{\mathbf{t}:N-t_0\le D_N} \mathbb{E}[S_{\mathbf{t}}] \le \epsilon_N$$

for some $\epsilon_N > 0$. Letting $\mathbb{E}[S_{\mathbf{t}} \mid d_{min} > D_N]$ be the average complete weight distribution of the expurgated ensemble whose minimum Hamming distance is lager than D_N, it is upper bounded by

$$\mathbb{E}[S_{\mathbf{t}} \mid d_{min} > D_N] \le \frac{\mathbb{E}[S_{\mathbf{t}}]}{1-\epsilon_N}.$$

3. The Conventional Upper Bound on Error Probability for Linear Block Codes

Let a received word be denoted by $\mathbf{y} = (y_1, y_2, \cdots, y_N) \in B^N$, let the code word for message m be denoted by $\mathbf{x}_m = (x_{m1}, x_{m2}, \cdots, x_{mN}) \in A^N$, $m = 0, 1, \cdots, M-1$ and let the likelihood of $\mathbf{x}_m$ be denoted by $P(\mathbf{y}|\mathbf{x}_m) = \prod_{i=1}^N p(y_i|x_{mi})$. In mismatched decoding, the decoder operates with a mismatched metric $W(\mathbf{y}|\mathbf{x}_m) = \prod_{i=1}^N w(y_i|x_{mi})$. So the decision region Λ_m for $\mathbf{x}_m$ can be defined as

$$\Lambda_m = \Big\{\mathbf{y} \in B^N \Big|\, W(\mathbf{y}|\mathbf{x}_m) > W(\mathbf{y}|\mathbf{x}_i),\ i < m,\ W(\mathbf{y}|\mathbf{x}_m) \ge W(\mathbf{y}|\mathbf{x}_i),\ i > m \Big\}. \tag{3}$$

Equation (3) represents the decoding region of maximum likelihood decoding (MLD) with the mismatched metric. For equi-probable messages, the MLD is equivalent to the maximum a posteriori probability decoding which minimize the error probability when $W = P$.

For a given transmitted message m, let $P_e(m)$ be the error probability. Assuming equi-probable messages, the average of the error probabilities is denoted by

$$P_e = \frac{1}{M}\sum_m P_e(m).$$

Then the standard Gallager bound [9] on $P_e(m)$ can be written as

$$P_e(m) \le \sum_{\mathbf{y}\in B^N} P(\mathbf{y}|\mathbf{x}_m)\left\{\sum_{m'\ne m}\left(\frac{W(\mathbf{y}|\mathbf{x}_{m'})}{W(\mathbf{y}|\mathbf{x}_m)}\right)^{\lambda}\right\}^{\rho}, \tag{4}$$
$$\lambda \ge 0,\ \rho \ge 0.$$

We apply the Duman and Salehi's type 2 bounding technique [6] to (4). Introducing some probability mass function (pmf) $\Phi_N^m(\mathbf{y})$, the right hand side (RHS) of (4) can be rewritten as

$$\begin{aligned} P_e(m) &\le \sum_{\mathbf{y}\in B^N} \Phi_N^m(\mathbf{y})\Phi_N^m(\mathbf{y})^{-1}P(\mathbf{y}|\mathbf{x}_m)\left\{\sum_{m'\ne m}\left(\frac{W(\mathbf{y}|\mathbf{x}_{m'})}{W(\mathbf{y}|\mathbf{x}_m)}\right)^{\lambda}\right\}^{\rho} \\ &= \sum_{\mathbf{y}\in B^N}\Phi_N^m(\mathbf{y})\left\{\sum_{m'\ne m}\Phi_N^m(\mathbf{y})^{-\frac{1}{\rho}}P(\mathbf{y}|\mathbf{x}_m)^{\frac{1}{\rho}}\cdot\left(\frac{W(\mathbf{y}|\mathbf{x}_{m'})}{W(\mathbf{y}|\mathbf{x}_m)}\right)^{\lambda}\right\}^{\rho} \\ &\le \left\{\sum_{\mathbf{y}\in B^N}\sum_{m'\ne m}\Phi_N^m(\mathbf{y})^{1-\frac{1}{\rho}}P(\mathbf{y}|\mathbf{x}_m)^{\frac{1}{\rho}}\left(\frac{W(\mathbf{y}|\mathbf{x}_{m'})}{W(\mathbf{y}|\mathbf{x}_m)}\right)^{\lambda}\right\}^{\rho}, \\ &\qquad \lambda \ge 0,\ 0 \le \rho \le 1, \end{aligned} \tag{5}$$

where the last inequality holds by invoking Jensen's inequality. Hereafter, we consider a q-ary linear block code transmitted over a regular channel with input alphabet $A = \{0, 1, \cdots, q-1\}$. We also assume that $\Phi_N^m(\mathbf{y})$ for $\mathbf{y} = (y_1, y_2, \cdots, y_N)$ does not depend on message m and expressed in a product form as

$$\Phi_N^m(\mathbf{y}) = \prod_{i=1}^N \phi(y_i)$$

with some pmf $\phi(\cdot)$ on B. Without loss of generality, we can assume that the actual transmitted code word is all zero $\mathbf{x}_0 = (0, 0, \cdots, 0)$ with $m = 0$. Then the term inside the brackets $\{\cdot\}^\rho$ in (5) can be rewritten as

$$\begin{aligned} &\sum_{y_1\in B}\cdots\sum_{y_N\in B}\sum_{m\ne 0}\prod_{i=1}^N \phi(y_i)^{1-\frac{1}{\rho}}p(y_i|0)^{\frac{1}{\rho}}\cdot w(y_i|0)^{-\lambda}w(y_i|x_{m'i})^{\lambda} \\ &= \sum_{m'\ne 0}\prod_{i=1}^N\left[\sum_{y_i\in B}\phi(y_i)^{1-\frac{1}{\rho}}p(y_i|0)^{\frac{1}{\rho}}w(y_i|0)^{-\lambda}w(y_i|x_{m'i})^{\lambda}\right] \\ &= \sum_{\mathbf{t}\ne 0}S_{\mathbf{t}}\prod_{x\in A}\left[\sum_{y\in B}\phi(y)^{1-\frac{1}{\rho}}p(y|0)^{\frac{1}{\rho}}w(y|0)^{-\lambda}w(y|x)^{\lambda}\right]^{t_x}, \end{aligned} \tag{6}$$

where $S_{\mathbf{t}}$ is dependent on a code. Since we assumed a linear code, the error probabilities for all codewords equals P_e, and $P_e = P_e(0)$, which is the error probability for message $m = 0$. Then, from (5) and (6), P_e can be upper bounded as

$$P_e \le \left\{\sum_{\mathbf{t}\ne 0}S_{\mathbf{t}}\prod_{x\in A}\left[\sum_{y\in B}\phi(y)^{1-\frac{1}{\rho}}p(y|0)^{\frac{1}{\rho}}\cdot w(y|0)^{-\lambda}w(y|x)^{\lambda}\right]^{t_x}\right\}^{\rho}. \tag{7}$$

Note that, for any $a_{\mathbf{t}} \ge 0$ indexed by $\mathbf{t}$,

$$\left(\sum_{\mathbf{t}} a_{\mathbf{t}}\right)^{\rho} \le \sum_{\mathbf{t}} a_{\mathbf{t}}{}^{\rho},\ \ 0 \le \rho \le 1 \tag{8}$$

holds. Setting

$$a_{\mathbf{t}} = S_{\mathbf{t}} \prod_{x\in A} \left[\sum_{y\in B} \phi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} w(y|0)^{-\lambda} w(y|x)^{\lambda} \right]^{t_x} \tag{9}$$

on the RHS of (7), we can derive the following upper bound:

$$P_e \le \sum_{\mathbf{t}\neq 0} (S_{\mathbf{t}})^{\rho} \prod_{x\in A} \left[\sum_{y\in B} \phi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} \cdot w(y|0)^{-\lambda} w(y|x)^{\lambda} \right]^{t_x \rho}. \tag{10}$$

If we consider the MBIOS channel, (10) is rewritten as

$$P_e \le \sum_{d=d_{\min}}^{N} (S_d)^{\rho} \left[\sum_{y\in B} \phi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} \right]^{(N-d)\rho} \cdot \left[\sum_{y\in B} \phi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} w(y|0)^{-\lambda} w(y|1)^{\lambda} \right]^{d\rho}, \tag{11}$$

where S_d is the number of code words whose Hamming weight is d and $d_{\min}$ is the minimum Hamming weight of the code. Equation (11) coincides with [2, Eqs. (37), (94) and (95)]. To obtain tight upper bounds for (11), one must choose good $\phi(y)$ and parameters $\lambda \ge 0$, $0 \le \rho \le 1$. Shamai and Sason [2] determine $\phi(y)$ for given λ and ρ as

$$\phi(y) = \frac{p(y|0)\left(1+\gamma(\frac{w(y|1)}{w(y|0)})^{\lambda}\right)^{\rho}}{\sum_{y\in B} p(y|0)\left(1+\gamma(\frac{w(y|1)}{w(y|0)})^{\lambda}\right)^{\rho}}, \tag{12}$$

where γ is the solution of the equation

$$\frac{\sum_{y\in B} p(y|0)\left(1+\gamma(\frac{w(y|1)}{w(y|0)})^{\lambda}\right)^{\rho-1}}{\sum_{y\in B} p(y|0)\left(1+\gamma(\frac{w(y|1)}{w(y|0)})^{\lambda}\right)^{\rho}} = 1 - \frac{d}{N}. \tag{13}$$

The existence and the uniqueness of γ are proved in [2]. After that, λ and ρ are optimized so as to minimize the RHS of (11).

For the ensemble of binary codes, we use the average weight distribution $\mathbb{E}[S_d]$ instead of S_d. Invoking Jensen's inequality $\mathbb{E}[X^{\rho}] \le (\mathbb{E}[X])^{\rho}$, $0 \le \rho \le 1$, (11) is immediately written as

$$\mathbb{E}[P_e] \le \sum_{d=d_{\min}}^{N} (\mathbb{E}[S_d])^{\rho} \left[\sum_{y\in B} \phi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} \right]^{(N-d)\rho} \cdot \left[\sum_{y\in B} \phi(y)^{1-\frac{1}{\rho}} p(y|0)^{\frac{1}{\rho}} w(y|0)^{-\lambda} w(y|1)^{\lambda} \right]^{d\rho}. \tag{14}$$

4. The Proposed Upper Bounds for Mismatched Decoding

In Sect. 3, Eq. (10) is derived by applying (8) to (7). As a result, Eq. (10) is a union bound over all subcodes with a constant Hamming weight, due to the channel symmetry and the property of the linear codes. Unfortunately, taking union bounds for each subcode with (8) leads to a looser bound than (7). Nevertheless, the technique of deriving the union bound with each subcode is standard and used in many analyzes of the upper bound on the error probability. For example, it was used for the convolutional code by Forney [10], [11], and for the generalized Viterbi algorithm by Hashimoto [12]. Though it is such a standard technique for analyzing the upper bound of error probability, we avoid using it in this section. That is, assuming the HSS ensemble, we derive the so-called direct bound and the Shulman-Feder type bound without using (8).

4.1 The DS2 Bound for Specific Linear Codes

We derive the upper bound for (7) without using (8). To begin this derivation, we set

$$\phi(y) = \frac{g(y)p(y|0)}{\sum_{y'\in B} g(y')p(y'|0)}, \tag{15}$$

where $g(\cdot)$ is an arbitrary non-negative function which satisfies

$$\sum_{y'\in B} g(y')p(y'|x) > 0, \quad x \in A. \tag{16}$$

Then (7) is rewritten as

$$P_e \le \left(\sum_{y\in B} g(y)p(y|0) \right)^{N(1-\rho)} \left(\sum_{\mathbf{t}\neq 0} S_{\mathbf{t}} \Gamma_g(\mathbf{t}) \right)^{\rho}, \tag{17}$$

$$\Gamma_g(\mathbf{t}) = \prod_{x\in A} \left[\sigma_g(x)\right]^{t_x}, \tag{18}$$

$$\sigma_g(x) = \sum_{y\in B} g(y)^{1-\frac{1}{\rho}} p(y|0) w(y|0)^{-\lambda} w(y|x)^{\lambda}. \tag{19}$$

4.2 The DS2 Bound for HSS Ensemble

In this section, we derive the DS2 bound [13] for the HSS ensemble from (7) and discuss how to determine reasonable $g(y)$ via the ensemble of random linear codes.

For the HSS ensemble, we use the average complete weight distribution $\mathbb{E}[S_{\mathbf{t}}]$ instead of $S_{\mathbf{t}}$. The first term in (17) does not depend on a code. So, invoking Jensen's inequality $\mathbb{E}[X^{\rho}] \le (\mathbb{E}[X])^{\rho}$, $0 \le \rho \le 1$ to the term inside the brackets $(\cdot)^{\rho}$ in (17), the average error probability over the ensemble can be expressed as

$$\mathbb{E}[P_e] \le \left(\sum_{y\in B} g(y)p(y|0) \right)^{N(1-\rho)} \left(\sum_{\mathbf{t}\neq 0} \mathbb{E}[S_{\mathbf{t}}] \Gamma_g(\mathbf{t}) \right)^{\rho}. \tag{20}$$

Assuming the HSS ensemble, we can show the following lemma.

Lemma 4.1 (The new DS2 bound for mismatched decoding) For any given q-ary input regular channel and the HSS ensemble $\overline{C_{N,k}}$, there exists a q-ary (N,k) linear block code which satisfies:

$$P_e \le D_g(\lambda, \rho, \overline{C_{N,k}}) \tag{21}$$

where $\lambda \ge 0$, $0 \le \rho \le 1$ and

$$D_g(\lambda, \rho, \overline{C_{N,k}}) = \left[\sum_{y\in B} g(y)p(y|0)\right]^{N(1-\rho)} \left[F_g(s, \rho, \overline{C_{N,k}})\right]^{\rho}, \tag{22}$$

$$F_g(\lambda, \rho, \overline{C_{N,k}}) = \sum_{\ell=1}^{N} P(\ell) \begin{pmatrix} N \\ \ell \end{pmatrix} \cdot \left\{\sigma_g(0)\right\}^{N-\ell} \left\{\sum_{x\in A\setminus 0} \sigma_g(x)\right\}^{\ell}. \tag{23}$$

Proof: By using (2), the term inside the brackets $(\cdot)^{\rho}$ on the RHS of (20) can be rewritten as

$$\sum_{\mathbf{t}\neq 0} \mathbb{E}[S_{\mathbf{t}}]\Gamma_g(\mathbf{t}) = \sum_{\ell=1}^{N} P(\ell) \sum_{\mathbf{t}:N-t_0=\ell} \begin{pmatrix} N \\ \mathbf{t} \end{pmatrix} \Gamma_g(\mathbf{t}), \tag{24}$$

where we use the assumption that the probability that each sequence of type $\mathbf{t}$ is a code word is the same if the Hamming weight of $\mathbf{t}$ is equal to ℓ. Using $\Gamma_g(\mathbf{t})$ in (18) and $\sigma_g(x)$ in (19), the RHS of (24) can be expressed as

$$\begin{aligned}
&\sum_{\ell=1}^{N} P(\ell) \sum_{\mathbf{t}:N-t_0=\ell} \begin{pmatrix} N \\ \mathbf{t} \end{pmatrix} \prod_{x\in A} \left[\sigma_g(x)\right]^{t_x} \\
&= \sum_{\ell=1}^{N} P(\ell) \begin{pmatrix} N \\ \ell \end{pmatrix} \left[\sigma_g(0)\right]^{N-\ell} \\
&\qquad \cdot \sum_{t_1+t_2+\cdots+t_{q-1}=\ell} \begin{pmatrix} \ell \\ t_1, \cdots, t_{q-1} \end{pmatrix} \prod_{x\in A\setminus 0} \left[\sigma_g(x)\right]^{t_x} \\
&= \sum_{\ell=1}^{N} P(\ell) \begin{pmatrix} N \\ \ell \end{pmatrix} \left[\sigma_g(0)\right]^{N-\ell} \left[\sum_{x\in A\setminus 0} \sigma_g(x)\right]^{\ell},
\end{aligned} \tag{25}$$

where the multinomial theorem yields the last equality. Thus Lemma 4.1 can be proved. □

From Lemma 4.1, our task is to find $g(y)$ that minimizes the RHS of (21). However, it seems difficult to find the optimal $g(y)$ for an arbitrary average complete weight distribution described in (2). In the next section, although the ensemble of random linear block codes [7] is a special subclass of the HSS ensemble, we derive the optimal $g(y)$ for the ensemble of random linear block codes, whose average complete weight distribution can be simply described.

4.3 Optimal $g(y)$ for the Ensemble of Random Linear Block Codes

In this section, we derive the optimal $g(y)$ for the ensemble of random linear block codes [7], which is a subclass of the HSS ensemble. It is known that the average complete weight distribution for the ensemble of random linear block codes is an ideal complete weight distribution which maximizes the single letter exponent derived by Shulman and Feder [17]. We consider the q-ary (N,k) linear block code. For the ensemble of random linear block codes, the average complete weight distribution satisfies $P(\ell) = q^{-(N-k)}$, which does not depend on ℓ. Then, the following theorem, which indicates that the optimal $g^*(y)$ depends only on decoding metric $W = \{w(y|x), x \in A, y \in B\}$, can be shown.

Theorem 4.1 (The optimal $g(y)$ for the ensemble of random linear block codes) For the ensemble of random q-ary (N,k) linear block codes, the optimal $g^*(y)$, which minimizes the RHS of (21), is given by

$$g^*(y) = w(y|0)^{-\lambda\rho} \left(\sum_{x\in A} w(y|x)^{\lambda}\right)^{\rho}. \tag{26}$$

Proof: For the ensemble of random linear block codes which is q-ary (N,k) linear code, $P(\ell) = q^{-(N-k)}$ holds, which does not depend on ℓ. From the binomial theorem, (23) is rewritten as

$$\begin{aligned}
&F_g(\lambda, \rho, \overline{C_{N,k}}) \\
&= q^{-(N-k)} \sum_{\ell=1}^{N} \begin{pmatrix} N \\ \ell \end{pmatrix} \left[\sigma_g(0)\right]^{N-\ell} \left[\sum_{x\in A\setminus 0} \sigma_g(x)\right]^{\ell} \\
&= q^{-(N-k)} \left[\sum_{x\in A} \sigma_g(x)\right]^{N}.
\end{aligned} \tag{27}$$

Then (22) is rewritten as

$$\begin{aligned}
&D_g(\lambda, \rho, \overline{C_{N,k}}) = q^{-(N-k)\rho} [\, I(g) \,]^{N}, \\
&I(g) = \left[\sum_{y\in B} g(y)p(y|0)\right]^{(1-\rho)} \\
&\qquad \cdot \left[\sum_{y\in B} g(y)^{1-\frac{1}{\rho}} p(y|0)w(y|0)^{-\lambda} \sum_{x\in A} w(y|x)^{\lambda}\right]^{\rho}.
\end{aligned} \tag{28}$$

By setting

$$\begin{aligned}
a(y) &= p(y|0)^{(1-\rho)}, \\
b(y) &= \left[g(y)^{-\frac{1}{\rho}} p(y|0)w(y|0)^{-\lambda} \sum_{x\in A} w(y|x)^{\lambda}\right]^{\rho},
\end{aligned} \tag{29}$$

$I(g)$ can be expressed as

$$I(g) = \left[\sum_{y\in B} g(y)a(y)^{\frac{1}{(1-\rho)}}\right]^{(1-\rho)} \left[\sum_{y\in B} g(y)b(y)^{\frac{1}{\rho}}\right]^{\rho} \geq \sum_{y\in B} g(y)a(y)b(y)$$

with equality if $a(y)^{\frac{1}{(1-\rho)}} = b(y)^{\frac{1}{\rho}}$, where the last inequality is due to Holder's inequality [9]. Hence, $a(y)^{\frac{1}{(1-\rho)}} = b(y)^{\frac{1}{\rho}}$ leads to the optimal $g(y) = g^*(y)$, which is denoted in (26). □

Substituting $g^*(y)$ noted in (26) to $g(y)$ in (28), the following corollary can be derived in the same manner as the proof of [13, Theorem 3.2].

Corollary 4.1 (Upper bound for mismatched decoding with random linear code ensemble) For the ensemble of random (N, k) linear block codes, there exists a code which satisfies

$$P_e \leq e^{-N[E_o(\lambda,\rho)-\rho R]}, \tag{30}$$

$$E_o(\lambda,\rho) = -\ln\left[\sum_{y\in B}\left(\sum_{x\in A}\frac{1}{q}p(y|x)w(y|x)^{-\lambda\rho}\right)\cdot\left(\sum_{x\in A}\frac{1}{q}w(y|x)^{\lambda}\right)^{\rho}\right], \quad \lambda \geq 0,\ 0 \leq \rho \leq 1. \tag{31}$$

Proof: Substituting $g^*(y)$ noted in (26) to $g(y)$ in (28), we obtain the following inequality from (21).

$$P_e \leq q^{k\rho}[\,\theta\,]^N,$$
$$\theta = q^{-\rho}\cdot I(g^*) = \sum_{y\in B} p(y|0)w(y|0)^{-\lambda\rho}\left(\frac{1}{q}\sum_{x\in A} w(y|x)^{\lambda}\right)^{\rho}. \tag{32}$$

Since the regular channels P and W are symmetric,

$$\theta = \sum_{y'\in B} p(y'|z)w(y'|z)^{-\lambda\rho}\left(\frac{1}{q}\sum_{x\in A} w(y'|x)^{\lambda}\right)^{\rho}$$

holds for any $z \in A$. In view of $|A| = q$, we obtain

$$q\theta = \sum_{z\in A}\sum_{y'\in B} p(y'|z)w(y'|z)^{-\lambda\rho}\left(\frac{1}{q}\sum_{x\in A} w(y'|x)^{\lambda}\right)^{\rho}. \tag{33}$$

Corollary 4.1 can be immediately derived by dividing both sides of (33) by q and exchanging the order of $\sum_{z\in A}$ and $\sum_{y'\in B}$. □

The properties of $g^*(y)$ are summarized in the following remarks.

Remark 4.1 In the derivation of Shamai and Sason's DS2 bounds (11)–(13), Inequality (8) is used. On the other hand, we avoid using (8) and use Jensen's inequality instead. However, by choosing $g^*(y)$ specified in (26) and the special structure of the ensemble such that $P(\ell) = q^{-(N-k)}$, the effect of Jensen's inequality disappears, and indeed the RHS of (30) coincides with the one in (7). Noticing that the RHS of (7) is also the intermediate step during deriving Shamai and Sason's DS2 bounds, we can conclude that our upper bound in Corollary 4.1 is not looser than those previous bounds (11)–(13).

Remark 4.2 It is of interest to see the function $g^*(y)$, which maximizes the RHS of (21), depends only on the decoding metric W, but not the actual channel P. On the other hand, in the upper bound in Eqs. (11)–(13), the optimal α in $\phi(y)$ depends not only on the mismatched metric W but also on the actual channel matrix P. By substituting the obtained $g^*(y)$ into (21), the tightest upper bound on the decoding error probability (21) can be calculated. The RHS of (21) is given by the mismatched metric W, the actual channel matrix P, and optimizing parameters λ, ρ.

4.4 The Direct Bound for HSS Ensemble

In this section, we derive the direct bound [13] from the DS2 bound for mismatched decoding.

Though it is sub-optimal, $g^*(y)$ specified in (26) can be applied to Lemma 4.1 with other HSS ensembles with arbitrary $P(\ell)$ satisfying (2). For example, we can apply it to the ensemble of LDPC codes introduced by Gallager [7]. Substituting $g^*(y)$ to $g(y)$ in (19), a new upper bound on the error probability can be derived. We call this bound the direct bound for mismatched decoding.

Theorem 4.2 (The direct bound for mismatched decoding) For any given q-ary input regular channel and the HSS ensemble $\overline{C_{N,k}}$, there exists a q-ary (N, k) linear block code which satisfies:

$$P_e \leq D^*(\lambda,\rho,\overline{C_{N,k}}) \tag{34}$$

where $\lambda \geq 0$, $0 \leq \rho \leq 1$ and

$$D^*(\lambda,\rho,\overline{C_{N,k}}) = \left[\sum_{y\in B} f(y)\right]^{N(1-\rho)}\left[F^*(\lambda,\rho,\overline{C_{N,k}})\right]^{\rho},$$
$$F^*(\lambda,\rho,\overline{C_{N,k}}) = q^N\sum_{\ell=1}^{N} P(\ell)\begin{pmatrix} N \\ \ell \end{pmatrix}\cdot\left\{\sum_{y\in B} f(y)\xi(y)\right\}^{N-\ell}\cdot\left\{\sum_{y\in B} f(y)\big(1-\xi(y)\big)\right\}^{\ell}, \tag{35}$$
$$f(y) = p(y|0)w(y|0)^{-\lambda\rho}\left(\frac{1}{q}\sum_{x\in A} w(y|x)^{\lambda}\right)^{\rho},$$
$$\xi(y) = \frac{w(y|0)^{\lambda}}{\sum_{x\in A} w(y|x)^{\lambda}}.$$

Proof: Setting $g(y) = g^*(y)$ in (23), we have

$$\sigma_{g^*}(0) = q \sum_{y \in B} f(y)\xi(y),$$

$$\sum_{x \in A \setminus 0} \sigma_{g^*}(x) = q \sum_{y \in B} f(y)\big(1 - \xi(y)\big).$$

On the other hand, setting $g(y) = g^*(y)$ for the term inside the parentheses $[\cdot]^{N(1-\rho)}$ in (22), we have

$$\sum_{y \in B} g^*(y)p(y|0) = \sum_{y \in B} f(y).$$

Thus, Theorem 4.2 can be proved. □

In contrast to Theorem 4.2, the upper bound on the decoding error probability with the actual channel metric was shown by Hof et al. [3, Eq. (16)]. Under the conditions written in the following remark, upper bound (34) obtained in Theorem 4.2 coincides with the upper bound [3, Eq. (16)] of Hof et al. This fact indicates that we have generalized [3, Eq. (16)] to mismatched decoding for the HSS ensemble in this paper.

Remark 4.3 If we set $\lambda = \frac{1}{1+\rho}$ and replace $w(y|x)$ with $p(y|x)$ for the RHS of (34), this upper bound coincides with the bound given in [3, Eq. (16)] for sufficiently small ϵ_N.

In the rest of this section, numerical examples of the bound shown in Theorem 4.2 are given. We consider the $(6, 12)$ LDPC code ensemble of length $N = 2004$. To calculate the upper bounds, we find $P(\ell)$ by Lemma 2.1 For the expurgated ensemble described in Lemma 2.2, we take $D_N = 160$, $\epsilon_N = 10^{-13}$. For the actual channel, we consider the channel with the additive white Gaussian noise (AWGN) and using BPSK modulation ($q = 2$) in Fig. 1, which sends the channel input of {0,1} to the channel as {1, −1} by the binary bipolar conversion. The channel output y is subject to the AWGN. On the other hand, we set the channel with mismatched metric, also described in Fig. 1. The receiving side determines that "0" is received if $y \geq T$ ($y \in r_0$), determines erasure "ε" if $-T \leq y < T$ ($y \in r_\varepsilon$), and determines that "1" is received if $y < -T$ ($y \in r_1$), where $T \geq 0$ denotes threshold. Then

$$w(0|0) = w(1|1) = \int_T^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y-1)^2}{2\sigma^2}}\, dy,$$

$$w(\varepsilon|0) = w(\varepsilon|1) = \int_{-T}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y-1)^2}{2\sigma^2}}\, dy,$$

$$w(1|0) = w(0|1) = \int_{-\infty}^{-T} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y-1)^2}{2\sigma^2}}\, dy,$$

$$\sigma^2 = E_s/N_o, \tag{36}$$

where E_s/N_o is the ratio of the signal power per channel input to noise.

Now we have prepared an example of the actual channel metric and the mismatched metric. Then we compare the numerical example of the proposed upper bound denoted in (34) with the conventional upper bound denoted in (14). We also compare these bounds with [3, Eq. (16)] of non-mismatched decoding, which uses the actual metric. In Fig. 2, each bound is depicted by optimizing λ and ρ within $\lambda \geq 0$ and $0 \leq \rho \leq 1$.

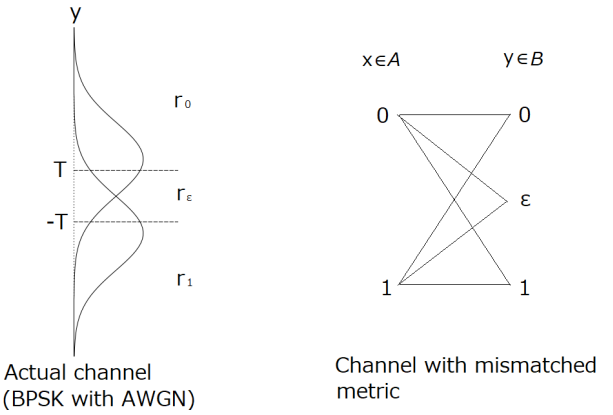

Fig. 1 Actual channel model and channel with mismatched metric.

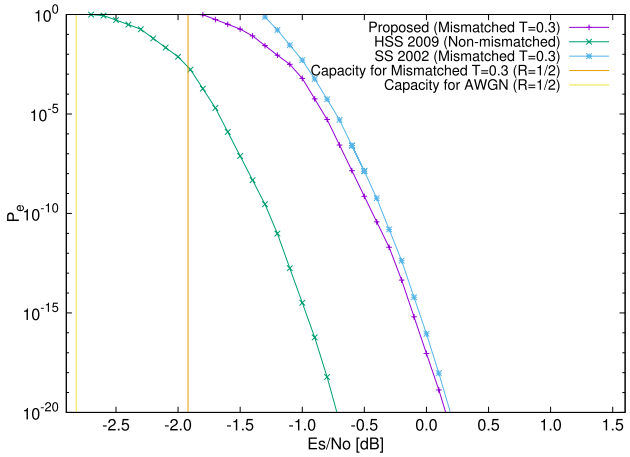

Fig. 2 P_E for the (6,12) LDPC code ensemble of length $N = 2004$ over the AWGN channel with BPSK ($q = 2$), $T = 0.3$. The curve "Proposed" depicts the RHS of (34), "HSS 2009" depicts the RHS of [3, Eq. (16)] and "SS 2002" depicts the RHS of (14), respectively.

Figure 2 shows that upper bound (34) shown in Theorem 4.2 is tighter than the conventional upper bound denoted in (14). In $E_s/N_o = -0.50$ [dB], the value on the RHS of (34) is 7.25×10^{-10}, while the value on the RHS of (14) is 1.32×10^{-8}.

The capacity for mismatched decoding has not yet been known accurately, and lower bounds on the the capacity for various discrete channels are discussed in [14]–[16]. However, the regular channels are assumed for both the actual channel and the channel with mismatched metric in this paper, and therefore the optimum distribution on the input alphabet is uniform. Then the capacity can be numerically calculated for the example shown in Fig. 1. In Fig. 2, we plot the value of E_s/N_o such that the capacity is $R = 1/2$ for the actual channel and that for the chan-

nel with mismatched metric. By numerical calculations, these values are $E_s/N_o = -2.82$ [dB] for actual channel and $E_s/N_o = -1.92$ [dB] for the channel with mismatched metric, respectively. It is interesting to see that these values are close to E_s/N_o at which the the decoding error occurs with probability almost one for the ensemble of the LDPC codes.

As the value of E_s/N_o is increased, the values on the RHS of (34) approaches that of (14). This fact indicates that the optimal parameter ρ, which makes the bound tightest, approaches one by increasing the value of E_s/N_o. It is because, as ρ approaches one in (8), the value of the RHS becomes equal to that of the LHS.

4.5 The Shulman-Feder Type Bound for HSS Ensemble

Next, following Shulman and Feder [17], we introduce $\alpha_q(\overline{C_{N,k}})$ to simplify the bound at the sacrifice of tightness of the bound noted in Theorem 4.2, and $\alpha_q(\overline{C_{N,k}})$ is given by

$$\alpha_q(\overline{C_{N,k}}) = \max_{\mathbf{t}\neq\mathbf{0}} \frac{\mathbb{E}[S_\mathbf{t}]}{q^{-(N-k)}\begin{pmatrix} N \\ \mathbf{t} \end{pmatrix}}. \tag{37}$$

For type $\mathbf{t}$ whose Hamming weight equals ℓ, the probability that each sequence of type $\mathbf{t}$ is a code word is equi-probable [4]. That is,

$$\max_{\mathbf{t}:N-t_0=\ell} \frac{\mathbb{E}[S_\mathbf{t}]}{\begin{pmatrix} N \\ \mathbf{t} \end{pmatrix}} = P(\ell).$$

Let ℓ^* be the minimum integer which satisfies

$$P(\ell^*) = \max_{1\le\ell\le N} P(\ell) = q^{-(N-k)}\alpha_q(\overline{C_{N,k}}). \tag{38}$$

Replacing $P(\ell)$ in (35) with $P(\ell^*)$ for all $1 \le \ell \le N$, we can derive the Shulman and Feder (SF) type bound via the binomial theorem. Thus the following theorem can be established.

Theorem 4.3 (The SF type bound for mismatched decoding) For the HSS ensemble $\overline{C_{N,k}}$, there exists an (N,k) linear block code which satisfies

$$P_e \le e^{-N\left[E_o(\lambda,\rho) - \rho\left(R + \frac{\ln \alpha_q(\overline{C_{N,k}})}{N}\right)\right]}, \tag{39}$$

$$E_o(\lambda,\rho) = -\ln\left[\sum_{y\in B}\left(\sum_{x\in A}\frac{1}{q}p(y|x)w(y|x)^{-\lambda\rho}\right)\cdot\left(\sum_{x\in A}\frac{1}{q}w(y|x)^{\lambda}\right)^{\rho}\right],$$

$$\lambda \ge 0,\ 0 \le \rho \le 1.$$

Proof: Replacing $P(\ell)$ in (35) with $P(\ell^*)$ given in (38) for all $1 \le \ell \le N$, (34) can be rewritten via the binomial theorem as

$$P_e \le q^{N\rho}\cdot P(\ell^*)^{\rho}\cdot[\theta]^N, \tag{40}$$

where θ is described in (32). Since θ satisfies (33), Theorem 4.3 can be proved in the same manner as Corollary 4.1. □

Thus the obtained upper bound leads to a single-letter expression of the lower bound on the error exponent. It shows the relation between the random coding exponent of block codes [18] and that of linear block codes.

Remark 4.4 In [18, Eq. (14)], let us suppose that both the actual channel and the channel with mismatched metric are regular channels, and the input assignment is equi-probable. Setting $\lambda = 1$ and replace $R + \frac{\ln \alpha_q(\overline{C_{N,k}})}{N}$ with R in (39), these two bounds coincide.

5. Conclusion

Conventionally, the upper bound on the error probability for linear codes over a regular channel with mismatched decoding was discussed in [2]. In this paper, we derive the new DS2 bound for the HSS ensemble. Then, for the ensemble of random linear codes, which is a subclass of the HSS ensemble, we minimize the new upper bound by choosing optimal $g^*(y)$. It is of interest to see that the optimal $g^*(y)$ depends only on the mismatched metric, but not on the actual one. On the other hand, the optimization of $g(y)$ for the HSS ensemble, whose average complete weight distribution is arbitrary, remains for future work. We also give the numerical example for the ensemble of LDPC codes also introduced by Gallager, which shows that our proposed bound is tighter than the conventional bound. Furthermore, we obtain the Shulman and Feder type bound for mismatched decoding, which enables us to obtain a single letter error exponent for linear block codes.

Acknowledgments

The authors would like to thank the anonymous reviewers and the editor for their helpful comments.

References

[1] T. Niinomi, H. Yagi, and S. Hirasawa, "Upper bounds on the error probability for the ensemble of linear block codes with mismatched decoding," B01-06, 2020 International Symposium on Information Theory and its Applications (ISITA2020), pp.151–155, Oct. 2020.
[2] S. Shamai (Shitz) and I. Sason, "Variations on the Gallager bounds, connections, and applications," IEEE Trans. Inf. Theory, vol.IT-48, no.12, pp.3029–3051, Dec. 2002.
[3] E. Hof, I. Sason, and S. Shamai (Shitz), "Performance bounds for nonbinary linear block codes over memolyless symmetric channels," IEEE Trans. Inf. Theory, vol.IT-55, no.3, pp.977–996, March 2009.
[4] E. Hof, I. Sason, and S. Shamai (Shitz), "Performance bounds for erasure, list, and decision feedback schemes with linear block codes," IEEE Trans. Inf. Theory, vol.IT-56, no.8, pp.3754–3778, Aug. 2010.
[5] P. Delsarte and P. Piret, "Algebraic constructions of Shannon codes for regular channels," IEEE Trans. Inf. Theory, vol.IT-28, no.4, pp.593–599, July 1982.

[6] T.M. Duman and M. Salehi, "New performance bounds for turbo codes," IEEE Trans. Commun., vol.COM-46, no.6, pp.717–723, June 1998.
[7] R.G. Gallager, Low-Density Parity-Check Codes, MIT Press, Cambridge, MA, 1963.
[8] T. Niinomi, H. Yagi, and S. Hirasawa, "Decision feedback scheme with criterion LR+Th for the ensemble of linear block codes," IEICE Trans. Fundamentals, vol.E103-A, no.1, pp.334–345, Jan. 2020.
[9] R.G. Gallager, Information Theory and Reliable Communication, Wiley, NY, 1968.
[10] G.D. Forney, Jr., "Convolutional codes II: Maximum likelihood decoding," Inform. Control, vol.25, pp.222–266, July 1974.
[11] G.D. Forney, Jr., "Convolutional codes III: Sequential decoding," Inform. Control, vol.25, no.3, pp.267–297, July 1974.
[12] T. Hashimoto, "A list-type reduced-constraint generalization of the Viterbi algorithm," IEEE Trans. Inf. Theory, vol.IT-33, no.6, pp.866–876, Nov. 1987.
[13] T. Niinomi, H. Yagi, and S. Hirasawa, "On the DS2 bound for Forney's generalized decoding using non-binary linear block codes," IEICE Trans. Fundamentals, vol.E101-A, no.8, pp.1223–1234, Aug. 2018.
[14] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," IEEE Trans. Inf. Theory, vol.IT-40, no.6, pp.1953–1967, Nov. 1994.
[15] A. Ganti, A. Lapidoth, and I.E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," IEEE Trans. Inf. Theory vol.IT-46, no.7, pp.2315–2328, Nov. 2000.
[16] A. Somekh-Baruch, "A general formula for the mismatch capacity," Proc. 2014 International Symposium on Inf. Theory (ISIT2014), pp.3067–3071, 2014.
[17] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," IEEE Trans. Inf. Theory, vol.IT-45, no.6, pp.2101–2104, Sept. 1999.
[18] I.G. Stiglitz, "Coding for a class of unknown channels," IEEE Trans. Inf. Theory, vol.IT-12, no.2, pp.189–195, April 1966.
[19] I. Sason and S. Shamai (Shitz), "On improved bounds on the decoding error probability of block codes over interleaved fading channels, with applications to turbo-like codes," IEEE Trans. Inf. Theory, vol.IT-47, no.6, pp.2275–2299, Sept. 2001.

Toshihiro Niinomi received the B.E. degree, the M.E. degree, and the Ph.D. degree in industrial and management systems engineering from Waseda University, Tokyo, Japan in 1988, 1990, and 2006, respectively. He was with Department of industrial and management systems engineering, School of Science and Engineering, Waseda University as a Research Associate from 1993 to 1996. He was a part time Lecturer at Shonan Institute of Technology and Polytechnic University from 1996 to 1999. He was with Department of Information Network Engineering, Kanagawa Institute of Technology from 1999 to 2007. He is currently an Associate Professor at the Department of Computer Science, Faculty of Information Technology, Tokyo City University (renamed form Musashi Institute of Institute of Technology), Tokyo, Japan from 2007. He is a member of the Electrical and Electronics Engineers (IEEE). His research interests include information and coding theory.

Hideki Yagi received the B.E. degree, the M.E. degree, and the Ph.D. degree in industrial and management systems engineering from Waseda University, Tokyo, Japan in 2001, 2003, and 2005, respectively. He is currently an Associate Professor at the Department of Computer and Network Engineering, University of Electro-Communications, Tokyo, Japan. He was with Media Network Center, Waseda University as a Research Associate from 2005 to 2007, and an Assistant Professor from 2007 to 2008. In the winter of 2008 and from July, 2010 to January, 2011, he was a Visiting Fellow at Princeton University. He is a member of the Electrical and Electronics Engineers (IEEE). His research interests include information and coding theory and information theoretic security.

Shigeichi Hirasawa was born in Kobe, Japan, on Oct. 2, 1938. He received the B.S. degree in mathematics and the B.E. degree in electrical communication engineering from Waseda University, Tokyo, Japan, in 1961 and 1963, respectively, and the Dr.E. degree in electrical communication engineering from Osaka University, Osaka, Japan, in 1975. From 1963 to 1981, he was with the Mitsubishi Electric Corporation, Hyogo, Japan. From 1981 to 2009, he was a professor of the School of Science and Engineering, Waseda University, Tokyo, Japan. He is currently a professor emeritus, and a researcher emeritus at the Waseda Research Institute for Science and Engineering, Waseda University. In 1979, he was a Visiting Scholar in the Computer Science Department at the University of California, Los Angeles (CSD, UCLA), CA. He was a Visiting Researcher at the Hungarian Academy of Science, Hungary, in 1985, and at the University of Trieste, Italy, in 1986. In 2002, he was again a Visiting Faculty at CSD, UCLA. From 1987 to 1989, he was the Chairman of the Technical Group on Information Theory of IEICE. He received the 1993 Achievement Award and the 1993 Kobayashi-Memorial Achievement Award from IEICE. In 1996, he was the President of the Society of Information Theory and Its Applications (Soc. of ITA). His research interests are information theory and its applications, and information processing systems. He is an IEEE Life Fellow, and a member of IPSJ, and JASMIN.