Successive-Cancellation Decoding of Linear Source Code

Jun Muramatsu
NTT Communication Science Laboratories, NTT Corporation
Hikaridai 2-4, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan

Abstract—This paper investigates the error probability of several decoding methods for a source code with decoder side information, where the decoding methods are: 1) symbol-wise maximum a posteriori decoding, 2) successive-cancellation decoding, and 3) stochastic successive-cancellation decoding. The proof of the effectiveness of a decoding method is reduced to that for an arbitrary decoding method, where ‘effective’ means that the error probability goes to zero as the block length goes to infinity. Furthermore, we revisit the polar source code showing that stochastic successive-cancellation decoding, as well as successive-cancellation decoding, is effective for this code.

I. INTRODUCTION

Successive-cancellation (SC) decoding is one of the elements constituting the polar codes introduced by Arıkan [1]. This paper investigates the error probability of SC decoding for a source code with decoder side information by extending the results in [2], [11] to general linear source codes [4], [9], [10]. It is shown that if for a given encoder there is a decoder such that the block error probability is $o(1/n)$ for the block length $n$, then the block error probability of an SC decoder for the same encoder is $o(1)$. Furthermore, we introduce stochastic successive-cancellation (SSC) decoding and show that it is equivalent to the constrained-random-number generator introduced in [5]. It is shown that if for a given encoder there is a decoder such that the block error probability is $o(1)$, then the block error probability of an SC decoder for the same encoder is $o(1)$. It is also shown that the error probability of the symbol-wise maximum a posteriori decoding of a linear source code and the SSC decoder of the polar source code goes to zero as $n$ goes to infinity. It should be noted that the results of this paper can be applied to the channel coding as introduced in [2], [9], [11].

II. SYMBOL-WISE MAXIMUM A POSTERIORI DECODING

First, we revisit symbol-wise maximum a posteriori (SMAP) decoding, which is used for the conventional decoding of a low density parity check code. Although the symbol error rate (the Hamming distance between a source output and its reproduction divided by the block length $n$) is discussed with symbol-wise maximum a posteriori decoding, we focus on the block error probability (an error occurs when a source output and its reproduction are different, that is, the Hamming distance is positive) throughout this paper.

Let $(A, \phi)$ be a pair consisting of a source encoder $A : \mathcal{X}^n \rightarrow \mathcal{X}^l$ and a decoder $\phi : \mathcal{X}^l \times \mathcal{Y}^n \rightarrow \mathcal{X}^n$ with side information. Let $c_1 \equiv Ax$ be the codeword of a source output $x \in \mathcal{X}^n$. The decoder $\hat{\phi} \equiv \{\hat{\phi}_i\}_{i=1}^n$ is constructed by using functions reproducing the $i$-th coordinate as

$$\hat{\phi}_i(c_1, y) \equiv \arg \max_{x_i \in \mathcal{X}} \mu_{X_i|c_1, y}(x_i|c_1, y).$$

It should be noted that when $(X, Y) \equiv (X^n, Y^n)$ is memoryless and $A$ is a sparse matrix we can use the sum-product algorithm to obtain an approximation of $\mu_{X_i|c_1, y}(x_i|c_1, y)$.

We have the following theorem, where the proof is given in the extended version [7].

**Theorem 1:** The error probability of the code $(A, \hat{\phi})$ is bounded as

$$\text{Prob}(\hat{\phi}(AX, Y) \neq X) \leq n \text{Prob}(\phi(AX, Y) \neq X),$$

where the right hand side of this inequality goes to zero as $n \rightarrow \infty$ when $\text{Prob}(\phi(AX, Y) \neq X) = o(1/n)$.

It is known that, when $l/n > H(X^n|Y^n)/n$ there is an encoding function $A : \mathcal{X}^n \rightarrow \mathcal{X}^l$ such that error probability $\text{Prob}(\phi(AX, Y) \neq X)$ is close to zero for all sufficiently large $n$ [4], [10], where we can use one of the following decoders:

- the typical set decoder
- the maximum a posteriori probability decoder defined as

$$\phi(c_1, y) \equiv \arg \max_{x} \mu_{X|c_1, Y}(x|c_1, y).$$

The following sections show upper bounds of the error probability for several decoders in terms of the error probability of a code $(A, \phi)$, where $\phi$ is an arbitrary decoder. It should be noted that we can use one of the decoders mentioned above. We can reduce the effectiveness of the decoder to that of an arbitrary decoder, where ‘effective’ means that the error probability goes to zero as $n$ goes to infinity. For example,
III. DECODING EXTENDED CODEWORD

Let $A : X^n \rightarrow X^l$ be an encoder of a source code with decoder side information. Here, we assume that, for a given $A$ there is a function $B : X^n \rightarrow X^{n-l}$ and a bijection $Q : X^n \rightarrow X^n$ such that $Q(Ax, Bx) = x$ for all $x \in X^n$. In particular, this condition is satisfied when $A$ is a full-rank matrix. We define the bijection $[A, B] : X^n \rightarrow X^n$ as $[A, B]x = (Ax, Bx)$.

Let $I_0$ and $I_1$ be a partition of $N = \{1, \ldots, n\}$, that is, they satisfy $I_0 \cap I_1 = \emptyset$ and $I_0 \cup I_1 = N$. We call $I_1$ and $I_0$ ordered when $I_1 = \{1, \ldots, l\}$ and $I_0 = \{l+1, \ldots, n\}$. For a vector $c \equiv (c_1, \ldots, c_n) \in X^n$, define $c_0 \in X^{n-l}$ and $c_1 \in X^l$ so that $c_i$ is a symbol in $c_0$ when $i \in I_0$ for every $b \in \{0, 1\}$. In the following, we assume that $Ax = c_1$ and $Bx = c_0$, where corresponding index sets $I_1$ and $I_0$ may not be ordered. We call $(c_0, c_1)$ the extended codeword of $c_1$. In the following, we denote $c = (c_0, c_1)$ omitting the dependence on $(I_0, I_1)$.

Let $f : X^l \times Y^n \rightarrow X^n$ be a function that reproduces the extended codeword by using the side information. For a codeword $c_1 \in X^l$ and side information $y \in Y^n$, the source decoder $\psi$ with side information is defined as

$$\psi(c_1, y) \equiv Q(f(c_1, y)).$$

In the context of the polar source codes, $c_0$ corresponds to unfrozen symbols and $Q$ corresponds to the final step of SC decoding. We have the following lemma for a general case, where the proof is given in the extended version [7].

**Lemma 1:** Let $C_1 \equiv AX$ and $C_0 \equiv BX$. Then we have

$$\Pr(\psi(AX, Y) \neq X) = \Pr(f(C_1, Y) \neq (C_0, C_1)).$$

In the following, we investigate the decoding error probability for an extended codeword.

IV. SUCCESSIVE-CANCELLATION DECODING

This section investigates the error probability of the (deterministic) SC decoding. For a source encoder $A : X^n \rightarrow X^l$, let $B$, $Q$, $C_0$, and $C_1$ be defined as in the previous section. For a codeword $c_1 \in X^l$ and side information $y \in Y^n$, the output $\hat{c}_i \equiv f(c_1, y)$ of an SC decoder $f$ is defined recursively as

$$\hat{c}_i \equiv \begin{cases} f_i(c_1^{-1}, y) & \text{if } i \in I_0 \\ c_i & \text{if } i \in I_1 \end{cases}$$

by using functions $\{f_i\}_{i \in I_0}$ defined as

$$f_i(c_1^{-1}, y) \equiv \arg \max_{c_i} \mu_{C_i|C_1^{-1}Y}(c_i|c_1^{-1}, y),$$

which is known as the maximum a posteriori decision rule after the observation $(c_1^{-1}, y)$, where $\mu_{C_i|C_1^{-1}Y}$ is the conditional probability defined as

$$\mu_{C_i|C_1^{-1}Y}(c_i|c_1^{-1}, y) \equiv \frac{\sum_{c_{n-l+1}^{n-1}} \mu_{C_0C_iC_1Y}(c_0, c_1, c_i, y)}{\sum_{c_{n-l+1}^{n-1}} \mu_{C_0C_1Y}(c_0, c_1, y)}.$$  

To simplify the notation, we define $f_i(c_1^{-1}, y) \equiv c_i$ when $i \in I_1$ although $c_i$ does not depend on $c_1^{-1}$. We have the following lemma.

**Lemma 2:**

$$\Pr(f(C_1, Y) \neq (C_0, C_1)) \leq \sum_{i \in I_0} \Pr(f_i(C_1^{-1}, Y) \neq C_i).$$

**Proof:** As with the proof in [1], we can express the block error events $f(c_1, y) \neq (c_0, c_1)$ as $E \equiv \bigcup_{i=1}^n E_i$, where

$$E_i \equiv \left\{ \begin{array}{l} (c, y) : f_j(c_1^{-1}, y) = c_j \text{ for all } j \in \{1, \ldots, i-1\} \\ f_i(c_1^{-1}, y) \neq c_i \end{array} \right\}$$

is an event where the first decision error in SC decoding occurs at stage $i$. The decoding error probability for an extended codeword is evaluated as

$$\Pr(f(C_1, Y) \neq (C_0, C_1)) = \Pr((C_0, C_1, Y) \in \mathcal{E}) \leq \sum_{i \in I_0} \Pr(f_i(C_1^{-1}, Y) \neq C_i),$$

where the first inequality comes from the union bound and the fact that $f_i(C_1^{-1}, Y) = C_i$ when $i \in I_1$, and the last inequality comes from the fact that $(c_0, c_1, y) \in \mathcal{E}_i$ implies $f_i(c_1^{-1}, y) \neq c_i$.

When the index sets $I_1$ and $I_0$ are not ordered like the polar source codes [2], [11], $f_i$ defined by (2) may not use the full information of a codeword $c_1 \equiv \{c_i\}_{i \in I_1}$. Borrowing words from [1], $f_i$ treats future symbols as random variables rather than as known symbols. In other words, $f_i$ ignores the future symbols in a codeword $c_1$. This implies that $\{f_i\}_{i=1}^n$ is different from the optimum maximum a posteriori decoder defined as $f_{\text{MAP}}(c_1, y) \equiv \arg \max_{c_i} \mu_{C_i|C_1Y}(c_i|c_1, y)$.

The following investigates the error probability of the SC decoding by assuming that the index sets $I_1$ and $I_0$ are ordered, that is, $I_1 = \{1, \ldots, l\}$ and $I_0 = \{l+1, \ldots, n\}$. This implies that for every $i \in I_0$, $f_i$ defined by (2) uses the full information of a codeword $c_1$.

**Lemma 3:** For a source encoder $A : X^n \rightarrow X^l$ and decoder $\phi : X^l \times Y^n \rightarrow X^n$ with side information, let $B$, $Q$, $C_0$, and $C_1$ be defined as in the previous section, where it is assumed that the index sets $I_0$ and $I_0$ are ordered. Then we have

$$\Pr(f_i(C_1^{-1}, Y) \neq C_i) \leq \Pr(\phi(AX, Y) \neq X).$$

**Proof:** For $i \in I_0$, let $f_i(c_1, y)$ be the $i$-th coordinate of the extended codeword of $Q^{-1}(\phi(c_1, y))$. Then we have
the fact that $f'_i(c_i, y) \neq c_i$ implies $\phi(Ax, y) \neq x$ for all $x$ satisfying $Ax = c_i$ and $Bx = 0$. Then we have

$$\begin{align*}
\text{Prob}(f_i(C_1^{-1}, Y) \neq C_1) &= \text{Prob}(\arg \max_{c_i} \mu_{C_1}C_1^{-1}Y(c_i|C_1^{-1}, Y) \neq C_1) \\
&\leq \text{Prob}(\arg \max_{c_i} \mu_{C_1}C_1Y(c_i|C_1, Y) \neq C_1) \\
&\leq \text{Prob}(f'_i(C_1, Y) \neq C_1) \\
&\leq \text{Prob}(\phi(AX, Y) \neq X),
\end{align*}$$

(5)

where the first inequality comes from [7, Lemma 7] and the fact that $C_1 = C_1^T$, and the second inequality comes from the fact that the maximum a posteriori decision rule minimizes the decision error probability.

From Lemmas 1–3 and the fact that $C_0 = X_1^{-1}$, we replace $C_1^{-1}Y(c_i|C_1^{-1}, y)$, where

$$\psi(AX, Y, y) = \mu_{C_0}C_0^{-1}Y(c_i|C_0^{-1}Y, y)$$

and $C_0^{-1}$ is the identity matrix and the right part of (5) is memoryless and reduce the conditional probability

$$\text{Prob}(\psi(AX, Y) \neq X) = \text{Prob}(\phi(AX, Y) \neq X),$$

where the right hand side of this inequality goes to zero as $n \to \infty$ when $\text{Prob}(\phi(AX, Y) \neq X) = o(1/n)$.

It should be noted again that the index sets $I_1$ and $I_0$ are ordered, while they are not ordered in the original polar source code. In contrast, we can use an arbitrary function $B$ that satisfies the assumption and rearrange the index sets $I_1$ and $I_0$ so that they are ordered, while they are fixed in the original polar source code.

V. STOCHASTIC SUCCESSIVE-CANCELLATION DECODING

This section introduces stochastic successive-cancellation (SSC) decoding, which is known as randomized rounding in the context of polar codes.

When $i \in I_0$, we replace $f_i$ defined in (2) by the stochastic decision rule generating $c_i$ randomly subject to the probability distribution $\{\mu_{C_1}C_1^{-1}Y(c_i|C_1^{-1}, y)\}_{y \in X}$ for a given $(c_i^{-1}, y)$. Let $F_i$ be the stochastic decision rule described above. Let $F$ be the stochastic decoder by using $F_i$ instead of $f_i$ when $i \in I_0$. We denote the stochastic decoder corresponding to (1) by $\Psi$. An analysis of the error probability will be presented in the next section.

VI. IMPLEMENTATION OF SUCCESSIVE-CANCELLATION DECODING

In this section, we assume that $A$ is a full-rank $l \times n$ (sparse) matrix. Without loss of generality, we can assume that the right part of $A$ is an invertible $l \times l$ matrix. This condition is satisfied for an arbitrary full-rank matrix $A$ by using a permutation matrix $S$, where $AS$ satisfies the condition, and the codeword can be obtained as $Ax = AS[S^{-1}x]$.

Let $B$ be an $[n-l] \times n$ matrix and assume that we obtain the invertible $n \times n$ matrix $[A, B]$ by concatenating row vectors of $B$ to $A$, that is, $[A, B]$ is bijective. By using $A$ and $B$, we can construct a successive-cancellation decoder that reproduces an extended codeword with $I_1 = \{1, ..., l\}$ and $I_0 = \{l+1, ..., n\}$.

Here, let us assume that the left part of $B$ is the $[n-l] \times [n-l]$ identity matrix and the right part of $B$ is the $[n-l] \times l$ zero matrix. It should be noted that a similar discussion is possible when the identity matrix is replaced by a permutation matrix.

From the assumptions of $B$, we have the fact that for all $j \in \{1, ..., n-l\}$ the $(l+j, j)$-element of $[A, B]$ is 1, which is the only positive element in $(l+j)$-th row of $[A, B]$. Then we have the fact that $C_{l+j} = X_j$ for all $j \in \{1, ..., n-l\}$, which implies $C_0 = X_1^{-1}$.

First, we reduce the conditional probability

$$\mu_{C_1}C_1^{-1}Y(c_i|C_1^{-1}, y)$$

defined by (3). For $i \in \{l+1, ..., n\}$ and $j \equiv i - l$, we have

$$\mu_{C_1}C_1^{-1}Y(c_i|C_1^{-1}, y) = \mu_{C_1}C_1^{-1}C_1Y(c_{l+i}|C_{l+i}^{-1}, c_1, y)$$

$$= \mu_{X_j}X_j^{-1}C_1Y(c_{l+i}|C_{l+i}^{-1}, c_1, y),$$

(6)

where the first equality comes from the fact that $I_1 = \{1, ..., l\}$ and the second equality comes from the fact that $C_{l+j} = X_j$ for all $j \in \{1, ..., n-l\}$. By substituting $c_{l+i} = x_{l+i}^1$, we have

$$\mu_{C_1}C_1^{-1}Y(c_{l+i}|x_{l+i}^1, y)$$

$$= \sum_{x_{l+i}^2} \mu_{X_j}X_j^{-1}Y(x_{l+i}^1|y) \chi(Ax = c_1)$$

$$= \sum_{x_{l+i}^2} \mu_{X_j}X_j^{-1}Y(x_{l+i}^1|y) \chi(Ax = c_1),$$

(7)

for $i \in \{l+1, ..., n\}$ and $j \equiv i - l$, where $x_{l+i}^1$ is the concatenation of $c_1$ and $x_{l+i}^1$. It should be noted that the right hand side of the second equality appears in the constrained-random-number generation algorithm [5, Eq. (41)]. This implies that the constrained-random-number generator can be considered as an SSC decoding $\Psi$ of the extended codeword specified in the previous section, where we have assumed that this algorithm uses the full information of the codeword $c_1$ for every $i \in \{l+1, ..., n\}$.

Next, we assume that $(X^n, Y^n)$ is memoryless and reduce the condition $Ax = c_1$ to improve the algorithm. This idea has already been presented in [6]. Let $a_j$ be the $j$-th column vector of $A$. Let $A_1^{-1}$ be the sub-matrix of $A$ obtained by using $\{a_j\}_{j=1}^{l-1}$ and $A_0^n$ be that obtained by using $\{a_j\}_{j=l}^n$. At the computation of (7) for $j \in \{1, ..., n-l\}$, we can assume that $x_{l+i}^{-1}$ has already been determined. Furthermore, we have the fact that the condition $Ax = c_1$ is equivalent to $A_0^n x_{l+i}^{-1} = c_1 - A_0^n x_{l+i}^{-1}$. Then, by letting $c_j(j) \equiv c_1 - A_0^n x_{l+i}^{-1}$, we can reduce (7) as follows:

$$\mu_{C_1}C_1^{-1}Y(x_{l+i}^1|c_1^{-1}, y)$$

$$= \frac{\sum_{x_{l+i}^2} \mu_{X_j}X_j^{-1}Y(x_{l+i}^1|y) \chi(Ax = c_1)}{\sum_{x_{l+i}^2} \mu_{X_j}X_j^{-1}Y(x_{l+i}^1|y) \chi(Ax = c_1)},$$

1 Rigorous proof is given in [7, Eq. (12)].
2 In [5, Eq. (41)], $\mu_{X_j}X_j^{-1}$ should be replaced by $\mu_{X_j}X_j^{-1}Y$.
we obtain

\[ \sum_{x_{n-l+1}^n} \left[ \prod_{k=j}^n \mu_{X_k|Y_k}(x_k|y_k) \right] \chi(A^n_j x_j^n = c'_1(j)) \]
\[ \sum_{x_{n-l+1}^n} \left[ \prod_{k=j}^n \mu_{X_k|Y_k}(x_k|y_k) \right] \chi(A^n_j x_j^n = c_1(j)) \]  

(8)

It should be noted that we can obtain \( A_{n-l} \) recursively by deleting the left-end column vector of \( A_{n-l+1} \). We can obtain the vector \( c'_1(j) \) recursively by using the relations

\[ c'_1(j) \equiv c_1(j) \equiv c'_1(j-1) - x_{j-1}a_{j-1} \text{ for } j \in \{2, \ldots, n-l\}. \]

These operations reduce the computational complexity of the algorithm.

Next, we convert the reproduction of a extended codeword to the reproduction of a source output. When \( j = n - l \), we have obtained the extended codeword \((c_1, c_0)\), where \( c_0 \equiv x_{n-l+1} = x_{l-1}^{-1}. \) We can reproduce the source output \( x \) by using the relation \( x \equiv [A, B]^{-1}c \), where \([A, B]^{-1}\) is the inverse of the concatenation of \( A \) and \( B \). From the assumptions of \( A \) and \( B \), we have the relations

\[ c_1 = A^{-1}x_{n-l} + A^n_{n-l+1}x_{n-l+1} \]
\[ c_0 = x_{n-l+1}^{-1}. \]

Since

\[ c'_1(n - l + 1) = c_1 - A^{-1}x_{n-l} \]

we obtain \( x_{n-l+1} \) as

\[ x_{n-l+1} = [A_{n-l+1}]^{-1} c'_1(n - l + 1), \]

where \([A_{n-l+1}]^{-1}\) is the inverse of \( A_{n-l+1} \).

Finally, we summarize the decoding algorithm. We assume that \((X^n, Y^n)\) is memoryless, \( A \) is an \( l \times n \) (sparse) matrix satisfying that \( A_{n-l+1}^{-1} \) is an \( l \times l \) invertible matrix, and \( B \) is an \([n - l] \times n\) matrix satisfying that \( B_{n-l+1}^{-1} \) is an \([n - l] \times [n - l]\) identity matrix and \( B_{n-l+1}^{-1} \) is the \([n - l] \times [n - l]\) zero matrix. It should be noted that the sum-product algorithm can be employed to obtain an approximation in line 4. The function \( \text{RNG}(\mu) \) in line 6 generates a random number subject to a distribution \( \mu \).

**SC/SSC Decoding Using Sum-Product Algorithm**

1: \( c'_1 \leftarrow c_1 \)
2: for \( j \in \{1, \ldots, n - l\} \) do
3: for \( x_j \in X \) do
4: \( \mu_{C_{l+j}|C_1^{l+j-1}Y}(x_j|x_1^{j-1}, y) \)
\[ \sum_{x_{j+1}^{n}} \prod_{k=j}^{n} \mu_{X_k|Y_k}(x_k|y_k) \chi(A_j^n x_j^n = c'_1) \]
\[ \sum_{x_j} \prod_{k=j}^{n} \mu_{X_k|Y_k}(x_k|y_k) \chi(A_j^n x_j^n = c_1) \]
5: end for
6: (SC decoding)
\[ x_j \leftarrow \arg \max_{x_j} \mu_{C_{l+j}|C_1^{l+j-1}Y}(x_j|x_1^{j-1}, y) \]
(SC decoding)
\[ x_j \leftarrow \text{RNG}(\mu_{C_{l+j}|C_1^{l+j-1}Y}(x_j|x_1^{j-1}, y)) \]
7: \( c'_1 \leftarrow c'_1 - x_j a_j \)
8: end for
9: \( x_{n-l+1}^{n} \leftarrow [A_{n-l+1}]^{-1}c'_1 \)
10: return \( x_{n-l+1}^{n} \)

Since the SSC decoder is equivalent to a constrained-random-number generator generating a random sequence subject to the a posteriori probability distribution \( \mu_{X|C_1, Y} \) [5, Theorem 5], we have the following theorem from the fact that the error probability of a stochastic decision with an a posteriori probability distribution is at most twice that of any decision rule [8, Lemma 3].

**Theorem 3:** For a linear source code \((A, \phi)\) with decoder side information, the decoding error of the SSC decoding algorithm is bounded as

\[ \text{Prob}(\Psi(AX, Y) \neq X) \leq 2 \text{Prob}(\phi(AX, Y) \neq X), \]

where the right hand side of this inequality goes to zero as \( n \to \infty \) when \( \text{Prob}(\phi(AX, Y) \neq X) = o(1) \).

**VII. ANALYSIS WHEN INDEX SETS ARE NOT ORDERED**

In the previous sections, it was assumed that the index sets \( I_1 \) and \( I_0 \) corresponding to \( c_1 = Ax \) and \( c_0 = Bx \) are ordered, that is, \( I_1 = \{1, \ldots, l\} \) and \( I_0 = \{l+1, \ldots, n\} \). This section investigates the case when they are not ordered. The following lemma asserts that the effectiveness of the decoder is reduced to a condition where the sum of the conditional entropies corresponding to the complement of the codeword goes to zero as \( n \to \infty \). The proof is given in the extended version [7].

**Lemma 4:** Let \( \psi \) and \( \Psi \) be the SC and SSC decoding functions, respectively. Then

\[ \text{Prob}(\psi(AX, Y) \neq X) \leq \frac{1}{2 \log 2} \sum_{i \in I_0} H(C_i|C_i^{-1}, Y) \]
\[ \text{Prob}(\Psi(AX, Y) \neq X) \leq \frac{1}{2 \log 2} \sum_{i \in I_0} H(C_i|C_i^{-1}, Y). \]

The above lemma implies that the error probability of SC/SSC decoding is small when \( \sum_{i \in I_0} H(C_i|C_i^{-1}, Y) \) is small. The following lemma introduces quasi-polarization where the both (9) and (10) are satisfied for all \( \delta > 0 \) and sufficiently large \( n \). It should be noted here that (9) implies that \( H(C_i|C_i^{-1}) \) is close to 0 but (10) may not imply that \( H(C_i|C_i^{-1}) \) is close to 1. The proof is given in the extended version [7].

**Lemma 5:** The condition

\[ \sum_{i \in I_0} H(C_i|C_i^{-1}, Y) \leq \delta \]  

(9)

is equivalent to the condition

\[ \sum_{i \in I_0} H(C_i|C_i^{-1}, Y) \geq H(X|Y) - \delta. \] 

(10)

**Remark 1:** It is mentioned in [3, “Polarization is commonplace”] that a random permutation of the set \( \{0, 1\}^n \) is a good polarizer with a high probability. We can show a similar fact
regarding a good source code \((A, \phi)\) and a matrix \(B\) that introduces the extended codeword, where the index sets \(I_0\) and \(I_1\) are ordered. Let \(\varepsilon\) be the decoding error probability of \((A, \phi)\). Then we have

\[
\sum_{i=t+1}^{n} H(C_i|C_{i-1}^{n-1}, Y) = H(BX|AX, Y) \\
\leq H(X|AX, Y) \\
\leq H(X|\phi(AX, Y)) \\
\leq \varepsilon \log |X|^n + h(\varepsilon) \\
\leq \varepsilon \left[ n + \log \frac{1}{\varepsilon} + \log e \right],
\]

(11)

where \(e\) is the base of the natural logarithm, the third inequality comes from the Fano inequality, and the fourth inequality comes from the fact that

\[
h(\varepsilon) = \varepsilon \log \frac{1}{\varepsilon} + (1 - \varepsilon) \log \left( \frac{1 + \varepsilon}{1 - \varepsilon} \right) \\
\leq \varepsilon \log \frac{1}{\varepsilon} + \varepsilon \log e \\
= \varepsilon \left[ \log \frac{1}{\varepsilon} + \log e \right],
\]

(12)

by using the relation \(\log(1 + \theta) \leq \theta \log e\). This means that we have quasi-polarization when \(\varepsilon = o(1/n)\). In particular, when \(\varepsilon\) goes to zero exponentially as \(n \to \infty\), \(\sum_{i=t+1}^{n} H(C_i|C_{i-1}^{n-1}, Y)\) also goes to zero exponentially.

**VIII. Stochastic Successive-Cancellation Decoding of Polar Source Code**

In this section, we revisit the polar source codes introduced in [2], [11]. For simplicity, we assume that \(|X|\) is a prime number. For a given positive integer \(k\), let \(n \equiv 2^k\). The source polarization transform \(G\) is defined as

\[
G \equiv \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^\otimes k S_{BR},
\]

where \(\otimes k\) denotes the \(k\)-th Kronecker power and \(S_{BR}\) is the bit-reversal permutation matrix defined in [1]. Then the extended codeword \(e \in X^n\) of a source output \(x \in X^n\) is defined as \(e \equiv t S_{BR}^T G x\), where both \(e\) and \(x\) are column vectors.

From [11, Theorem 4.10], we have

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i \in N} \mathbb{P}\left( Z(C_i|C_{i-1}^{n-1}, Y) \leq 2^{-n^{\alpha}} \right) = 1 - H(X|Y)
\]

for all \(\beta \in (0, 1/2)\), where \(Z\) is the source Bhattacharyya parameter defined as

\[
Z(U|V) = \frac{1}{|X|-1} \sum_{u,v} \sum_{u',v'} \sqrt{\mu_{uv}(u, v)\mu_{uv}(u', v)}.
\]

Let \(I_0\) and \(I_1\) be defined as

\[
I_0 \equiv \left\{ i \in N : Z(C_i|C_{i-1}^{n-1}, Y) \leq 2^{-n^{\alpha}} \right\}
\]

\[
I_1 \equiv \left\{ i \in N : Z(C_i|C_{i-1}^{n-1}, Y) > 2^{-n^{\alpha}} \right\}.
\]

Then, from (13), we have the fact that the encoding rate \(|I_1|/n\) approaches \(H(X|Y)\) as

\[
\lim_{n \to \infty} \frac{|I_1|}{n} = \lim_{n \to \infty} \frac{n - |I_0|}{n} = 1 - \lim_{n \to \infty} \frac{|I_0|}{n} = H(X|Y).
\]

Furthermore, from Lemma 4, we have

\[
\lim_{n \to \infty} \frac{|I_0|}{n} \mathbb{P}(\psi AX, Y) \neq X \leq \lim_{n \to \infty} \frac{1 - \frac{1}{2} \log 2}{n} \sum_{i \in I_0} H(C_i|C_{i-1}^{n-1}, Y) \\
\leq \frac{1}{2} \log 2 \lim_{n \to \infty} \sum_{i \in I_0} |X| - 1 Z(C_i|C_{i-1}^{n-1}, Y) \\
\leq \frac{1}{2} \log 2 \lim_{n \to \infty} n 2^{-n^{\alpha}} \\
= 0
\]

(14)

for all \(\beta \in (0, 1/2)\), where the second inequality comes from the relation

\[
H(U|V) \leq \log(1 + |X| - 1|Z(U|V)|) \leq |X| - 1|Z(U|V)|
\]

shown in [2, Eq. (5)], [11, Eq. (4.11)]. This implies the well-known fact that SC decoding of the polar source code is effective [2], [11].

Similarly, we have the following theorem, which implies the effectiveness of the SSC decoding of the polar source code.

**Theorem 4:**

\[
\lim_{n \to \infty} \mathbb{P}(\psi AX, Y) \neq X \leq \frac{|X| - 1}{\log 2} \lim_{n \to \infty} n 2^{-n^{\alpha}} = 0
\]

for all \(\beta \in (0, 1/2)\).

**References**

[1] E. Arıkan, “Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inform. Theory*, vol. IT-55, no. 7, July 2009.

[2] E. Arıkan, “Source polarization,” *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, U.S.A., June 13–18, 2010, pp. 899–903.

[3] E. Arıkan, “Polar coding — status and prospects,” *Plenary Lecture at the 2010 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Aug. 1, 2011.

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

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

[6] J. Muramatsu, “Variable-length lossy source code using a constrained-random-number generator,” *IEEE Trans. Inform. Theory*, vol. IT-61, no. 6, Jun. 2015.

[7] J. Muramatsu, “Successive-cancellation decoding of linear source code,” available at arXiv:1903.11787[cs.IT], 2019.

[8] J. Muramatsu and S. Miyake, “On the error probability of stochastic decision and stochastic decoding,” *Proc. 2017 IEEE Int. Symp. Inform. Theory*, Aachen, Germany, June 25–30, 2017, pp. 1643–1647. Extended version available at arXiv:1701.04950[cs.IT], 2017.

[9] J. Muramatsu and S. Miyake, “Channel code using constrained-random-number generator revisited,” *IEEE Trans. Inform. Theory*, vol. IT-65, no. 1, pp. 500–508, Jan. 2019.

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

[11] E. Şaşıoğlu, “Polarization and polar codes,” *Fund. Trends Commun. Inform. Theory*, vol. 8, no. 4, pp. 259–381, Oct. 2012.