Channel Polarization on \(q\)-ary Discrete Memoryless Channels by Arbitrary Kernels

Ryuhei Mori
Graduate School of Informatics
Kyoto University
Kyoto, 606–8501, Japan
Email: rmori@sys.i.kyoto-u.ac.jp

Toshiyuki Tanaka
Graduate School of Informatics
Kyoto University
Kyoto, 606–8501, Japan
Email: tt@i.kyoto-u.ac.jp

Abstract—A method of channel polarization, proposed by Arıkan, allows us to construct efficient capacity-achieving channel codes. In the original work, binary input discrete memoryless channels are considered. A special case of \(q\)-ary channel polarization is considered by Şaşoğlu, Telatar, and Arıkan. In this paper, we consider more general channel polarization on \(q\)-ary channels. We further show explicit constructions using Reed-Solomon codes, on which asymptotically fast channel polarization is induced.

I. INTRODUCTION

Channel polarization, proposed by Arıkan, is a method of constructing capacity achieving codes with low encoding and decoding complexities [1]. Channel polarization can also be used to construct lossy source codes which achieve rate-distortion trade-off with low encoding and decoding complexities [2]. Arıkan and Telatar derived the rate of channel polarization [3]. In [4], a more detailed rate of channel polarization which includes coding rate is derived. In [1], channel polarization is based on a \(2 \times 2\) matrix. Korada, Şaşoğlu, and Urbanke considered generalized polarization phenonemon which is based on an \(\ell \times \ell\) matrix and derived the rate of the generalized channel polarization [5]. In [6], a special case of channel polarization on \(q\)-ary channels is considered. In this paper, we consider channel polarization on \(q\)-ary channels which is based on arbitrary mappings.

II. PRELIMINARIES

Let \(u_{\ell-1}^{\ell-1}\) and \(u_j^j\) denote a row vector \((u_0, \ldots, u_{\ell-1})\) and its subvector \((u_i, \ldots, u_j)\). Let \(\mathcal{F}\) denote the complement of a set \(\mathcal{F}\), and \(|\mathcal{F}|\) denotes cardinality of \(\mathcal{F}\). Let \(\mathcal{X}\) and \(\mathcal{Y}\) be an input alphabet and an output alphabet, respectively. In this paper, we assume that \(\mathcal{X}\) is finite and that \(\mathcal{Y}\) is at most countable. A discrete memoryless channel (DMC) \(W\) is defined as a conditional probability distribution \(W(y | x)\) over \(\mathcal{Y}\) where \(x \in \mathcal{X}\) and \(y \in \mathcal{Y}\). We write \(W : \mathcal{X} \rightarrow \mathcal{Y}\) to mean a DMC \(W\) with an input alphabet \(\mathcal{X}\) and an output alphabet \(\mathcal{Y}\). Let \(q\) be the cardinality of \(\mathcal{X}\). In this paper, the base of the logarithm is \(q\) unless otherwise stated.

Definition 1: The symmetric capacity of \(q\)-ary input channel \(W : \mathcal{X} \rightarrow \mathcal{Y}\) is defined as

\[
I(W) := \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{q} W(y | x) \log \frac{W(y | x)}{\sum_{y' \in \mathcal{Y}} W(y' | x)}. \]

Note that \(I(W) \in [0, 1]\).

Definition 2: Let \(\mathcal{D}_x := \{y \in \mathcal{Y} | W(y | x) > W(y | x'), \forall x' \in \mathcal{X}, x' \neq x\}\). The error probability of the maximum-likelihood estimation of the input \(x\) on the basis of the output \(y\) of the channel \(W\) is defined as

\[
P_e(W) := \frac{1}{q} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{D}_x} W(y | x).
\]

Definition 3: The Bhattacharyya parameter of \(W\) is defined as

\[
Z(W) := \frac{1}{q(q-1)} \sum_{x \in \mathcal{X}, x' \in \mathcal{X}, x \neq x'} Z_{x,x'}(W)
\]

where the Bhattacharyya parameter of \(W\) between \(x\) and \(x'\) is defined as

\[
Z_{x,x'}(W) := \sum_{y \in \mathcal{Y}} \sqrt{W(y | x) W(y | x')}.
\]

The symmetric capacity \(I(W)\), the error probability \(P_e(W)\), and the Bhattacharyya parameter \(Z(W)\) are interrelated as in the following lemmas.

Lemma 4:

\[
P_e(W) \leq (q-1)Z(W).
\]

Lemma 5: [6]

\[
I(W) \geq \log \frac{q}{1 + (q-1)Z(W)} \quad I(W) \leq \log(q/2) + (\log 2) \sqrt{1 - Z(W)^2} \quad I(W) \leq 2(q-1)(\log e) \sqrt{1 - Z(W)^2}.
\]

Definition 6: The maximum and the minimum of the Bhattacharyya parameters between two symbols are defined as

\[
Z_{\text{max}}(W) := \max_{x \in \mathcal{X}, x' \in \mathcal{X}, x \neq x'} Z_{x,x'}(W) \quad Z_{\text{min}}(W) := \min_{x \in \mathcal{X}, x' \in \mathcal{X}} Z_{x,x'}(W).
\]

Let \(\sigma : \mathcal{X} \rightarrow \mathcal{X}\) be a permutation. Let \(\sigma^i\) denote the \(i\)th power of \(\sigma\). The average Bhattacharyya parameter of \(W\) between \(x\) and \(x'\) with respect to \(\sigma\) is defined as the average of
as $D_{MC}$s with transition probabilities $x \rightarrow X$

III. CHANNEL POLARIZATION ON $q$-ARY DMC INDUCED BY NON-LINEAR KERNEL

We consider a channel transform using a one-to-one onto mapping $g : \mathcal{X}^e \rightarrow \mathcal{X}^e$, which is called a kernel. In the previous works [11, 13], it is assumed that $q = 2$ and that $g$ is linear. In [6], $\mathcal{X}$ is arbitrary but $g$ is restricted. In this paper, $\mathcal{X}$ and $g$ are arbitrary.

Definition 7: Let $W : \mathcal{X}^e \rightarrow \mathcal{Y}$ be a DMC. Let $W^g_\ell(x_0^{n\ell} : x_0^{n\ell−1}) = \frac{1}{q^{\ell+1}} \sum_{u_0^{n\ell−1}} W^{(g^\ell(x_0^{n\ell−1}) | g(u_0^{n\ell−1}))}.$

Definition 8: Let $\{B_i\}_{i=0, 1, \ldots}$ be independent random variables such that $B_i = k$ with probability $\frac{1}{n\ell}$, for each $k = 0, 1, \ldots, \ell − 1$.

In probabilistic channel transform $W \rightarrow W^{(B_i)}$, expectation of the symmetric capacity is invariant due to the chain rule for mutual information. The following lemma is a consequence of the martingale convergence theorem.

Lemma 9: There exists a random variable $I_\infty$ such that $I(W^{(B_i)} \cdots (B_{i+1}))$ converges to $I_\infty$ almost surely as $n \rightarrow \infty$.

When $q = 2$ and $g(u_i) = (u_0 + u_1, u_1)$, Arkan showed that $P(I_\infty \in \{0, 1\}) = 1$ [11]. This result is called channel polarization phenomenon since subchannels polarize to noiseless channels and pure noise channels. Korada, Sapoglu, and Urbanke consider channel phenomenon when $q = 2$ and $g$ is linear [5].

From Lemma 3 $I(W)$ is close to 0 and 1 when $Z(W)$ is close to 1 and 0, respectively. Hence, it would be sufficient to prove channel polarization if one can show that $Z(W^{(B_i)} \cdots (B_{i+1}))$ converges to $Z_\infty \in \{0, 1\}$ almost surely. Here we instead show a weaker version of the above property in the following lemma and its corollary.

Lemma 10: Let $\{\mathcal{Y}_n\}_{n \in \mathbb{N}}$ be a sequence of discrete sets. Let $\{W_n : \mathcal{X} \rightarrow \mathcal{Y}_n\}_{n \in \mathbb{N}}$ be a sequence of $q$-ary DMCs. Let $\sigma$ and $\tau$ be permutations on $\mathcal{X}$. Let $W_n^\sigma, W_n^\tau : \mathcal{X}^e \rightarrow \mathcal{Y}_n^e$ be defined as DMCs with transition probabilities $x \rightarrow X$

Proof: Let $Z_\infty Z_{x, x'}(W)$ over the subset $\{(z, z') = (\sigma^i(x), \sigma^i(x')) \in \mathcal{X}^2$ for $i = 0, 1, \ldots, q^n−1\}$

$$Z_{x, x'}(W) := \frac{1}{q^n} \sum_{i=0}^{q^n−1} Z_{\sigma^i(x), \sigma^i(x')}(W).$$

Since $I(W_n^g) = I(Z(Y_1, Y_2))$ and $I(W_n^h) = I(Z(Y_1)$

$$I(Z|Y_1, Y_2) - I(Z|Y_1) = I(Z|Y_2 | Y_1)$$

tends to 0 by the assumption. Since the mutual information is lower bounded by the cut-off rate, one obtains

$$P(z, y) = 0$$

and

$$P(z, y) = 1$$

for any $z, x \in \mathcal{X}^2$. It consequently implies that for any $\delta \in (0, 1/2), there$ exists $m$ such that

$$P_{\infty}Z_{x, x'}(W) \leq (\delta, 1 - \delta)$$

for any $x, x' \in \mathcal{X}$ and $n \geq m$.
Variables ranging over a permutation \( v \) such that \( \sigma(\cdot) \) and \( \tau(\cdot) \) are invariant under an operation \( G \) we assume that \( G \) is a lower triangle and, perturbation matrices, without loss of generality we assume that \( G \) is a lower triangle matrix and that \( G_{kk} = 1 \) where \( k \in \{0, \ldots, \ell - 1\} \) is the largest number such that the number of non-zero elements in \( k \)-th row of \( G \) is greater than 1, and where \( G_{ij} \) denotes \( (i, j) \) element of \( G \).

**Theorem 13:** Assume that \( \mathcal{X} \) is a field of prime cardinality, and that linear kernel \( G \) is not diagonal. Then, \( P(I_{\infty} \in \{0, 1\}) = 1 \).

**Proof:** It holds

\[
W(k)(y_0^{\ell-1}, u_k^{\ell-1} | u_k) = \frac{1}{q^{\ell-1}} \prod_{j=1}^{\ell-1} \left( \sum_{x \in \mathcal{X}} W(y_j | x) \right) \times \prod_{j \in S_0} \prod_{j \in S_1} W(y_j | G_{kj} u_k + x_j)
\]

where \( S_0 := \{ j \in \{0, \ldots, \ell - 1\} | G_{kj} = 0 \} \), \( S_1 := \{ j \in \{0, \ldots, \ell - 1\} | G_{kj} \neq 0 \} \), and \( x_j \) is \( j \)-th element of \( (u_0^{\ell-1}, 0^{\ell-1}) G \) where \( G \) is the zero matrix of length \( \ell \).

Let \( m \in \{0, \ldots, \ell - 1\} \) be such that \( G_{km} \neq 0 \). Since each \( k^{-1} \) occurs with positive probability \( 1/q^\ell \), we can apply Lemma\ref{lem:coro10} with \( \sigma(x) = x \) and \( \tau(x) = G_{km} x + z \) for arbitrary \( z \in \mathcal{X} \). Hence, for sufficiently large \( m \), \( \mathcal{Z}_{x,x'}^m(W(B_1) \circ \cdots \circ (B_m)) \) is close to 0 or almost surely where \( \mu(x) = G_{km} x + z \) for any \( x \in \mathcal{X} \) such that \( |Z| \) is close to 1 when \( \mu_0(z) = z + x' - x \) for \( x \neq x' \), \( Z_{x,x'}^m(W(B_1) \circ \cdots \circ (B_m)) \) is close to 0 and only if \( Z(W(B_1) \circ \cdots \circ (B_m)) \) is close to 0 or 1, respectively.

This result is a simple generalization of the special case considered by Šaşoğlu, Telatar, and Arikan. For a prime power \( q \) and a finite field \( \mathcal{X} \), we show a sufficient condition for channel polarization in the following corollary.

**Corollary 14:** Assume that \( \mathcal{X} \) is a field and that a linear kernel \( G \) is not diagonal. If there exists \( j \in \{0, \ldots, k-1\} \) such that \( G_{kj} \) is a primitive element. Then, \( P(I_{\infty} \in \{0, 1\}) = 1 \).

**Proof:** By applying Lemma\ref{lem:coro10} one sees that for almost every sequence \( b_1, \ldots, b_m \) of \( 0, \ldots, \ell - 1 \), and for any \( \delta \in (0, 1/2) \), there exists \( m \) such that \( |Z_{x,x'}^m(W(B_1) \circ \cdots \circ (B_m))| \neq |\delta, 1 - \delta| \) for any \( x \in \mathcal{X}, x' \in \mathcal{X} \) and \( n \geq m \).

If a kernel is linear, a more detailed condition is obtained.

**Definition 12:** Assume \((\mathcal{X}, +, \cdot)\) be a commutative ring. A kernel \( g : \mathcal{X}^d \to \mathcal{X}^d \) is said to be linear if \( g(a x + b z) = a g(x) + b g(z) \) for all \( a, b \in \mathcal{X} \), \( x, z \in \mathcal{X}^d \), and \( \mathcal{X}^d \in \mathcal{X}^d \).

If \( g \) is linear, \( g \) can be represented by a square matrix \( G \) such that \( g(u_0^{\ell-1}) = u_0^{\ell-1} G \). Let \( u_0^{\ell-1}, X_0^{\ell-1} \) and \( Y_0^{\ell-1} \) denote random variables taking values on \( \mathcal{X}^d, \mathcal{X}^d \) and \( \mathcal{X}^d \), respectively, and obeying distribution

\[
P(U_0^{\ell-1} = u_0^{\ell-1}, X_0^{\ell-1} = x_0^{\ell-1}, Y_0^{\ell-1} = y_0^{\ell-1}) = \frac{1}{q^{2\ell}} W_{\ell, \ell-1}(y_0^{\ell-1} | u_0^{\ell-1} G) \mathbb{1}\{x_0^{\ell-1} V = u_0^{\ell-1}\}
\]

where \( V \) denotes an \( \ell \times \ell \) full-rank upper triangle matrix. There exists a one-to-one correspondence between \( X_0^{\ell-1} \) and \( U_0^{\ell-1} \) for all \( i \in \{0, \ldots, \ell - 1\} \). Hence, statistical properties of \( W^{(i)} \) are invariant under an operation \( G \to VG \). Further, a permutation of columns of \( G \) does not change statistical properties of \( W^{(i)} \) either. Since any full-rank matrix can be decomposed to the form \( VLP \) where \( V, L, \) and \( P \) are upper triangle, lower triangle, and permutation matrices, without loss of generality we assume that \( G \) is a lower triangle and that \( G_{kk} = 1 \) where \( k \in \{0, \ldots, \ell - 1\} \) is the largest number such that the number of non-zero elements in \( k \)-th row of \( G \) is greater than 1, and where \( G_{ij} \) denotes \( (i, j) \) element of \( G \).

**IV. Speed of Polarization**

Ankara and Telatar showed the speed of polarization \cite{Art}. Korada, Šaşoğlu, and Urbanke generalized it to any binary linear kernels \cite{KOR}.
Proposition 15: Let \( \{\hat{X}_n \in (0, 1)\}_{n\in\mathbb{N}} \) be a random process satisfying the following properties.

1) \( \hat{X}_n \) converges to \( \hat{X}_\infty \) almost surely.
2) \( \hat{X}_{n+1} \leq \hat{c} \hat{X}_{D_n} \) where \( \{D_n \geq 1\}_{n\in\mathbb{N}} \) are independent and identically distributed random variables, and \( \hat{c} \) is a constant.

Then,
\[
\lim_{n\to\infty} P(\hat{X}_n < 2^{-2^{\beta n}}) = P(\hat{X}_\infty = 0)
\]
for \( \beta < \mathbb{E}[\log_2 D_1] \) where \( \mathbb{E}[\cdot] \) denotes an expectation. Similarly, let \( \{\tilde{X}_n \in (0, 1)\}_{n\in\mathbb{N}} \) be a random process satisfying the following properties.

1) \( \tilde{X}_n \) converges to \( \tilde{X}_\infty \) almost surely.
2) \( \tilde{X}_{n+1} \geq \tilde{c} \tilde{X}_{D_n} \) where \( \{D_n \geq 1\}_{n\in\mathbb{N}} \) are independent and identically distributed random variables, and \( \tilde{c} \) is a constant.

Then,
\[
\lim_{n\to\infty} P(\tilde{X}_n < 2^{-2^{\beta n}}) = 0
\]
for \( \beta > \mathbb{E}[\log_2 \tilde{D}_1] \).

Note that the above proposition can straightforwardly be extended to include the rate dependence [4].

In order to apply Proposition 15 to \( Z_{\max}(W(B_1)\ldots(B_n)) \) and \( Z_{\min}(W(B_1)\ldots(B_n)) \) as \( \hat{X}_n \) and \( \tilde{X}_n \), respectively, the second conditions have to be proven. In the argument of [5], partial distance of a kernel corresponds to the random variables \( \hat{D}_n \) and \( \tilde{D}_n \) in Proposition 15.

Definition 16: Partial distance of a kernel \( g : \mathcal{X}^\ell \to \mathcal{X}^\ell \) is defined as
\[
D^{(i)}_{x,x'}((i-1)) := \min_{v_{i-1},w_{i-1}} d(g(u_{i-1}^{(i-1)}, x, v_{i-1}^{(i-1)}), g(u_{i-1}^{(i-1)}, x', w_{i-1}^{(i-1)}))
\]
where \( d(a, b) \) denotes the Hamming distance between \( a \in \mathcal{X}^\ell \) and \( b \in \mathcal{X}^\ell \).

We also use the following quantities.
\[
D^{(i)}_{x,x'} := \min_{v_{i-1}} D^{(i)}_{x,x'}(v_{i-1}) \quad D^{(i)}_{x,x'}(v_{i-1}) := \max_{x \in \mathcal{X}, x' \in \mathcal{X}} D^{(i)}_{x,x'}(v_{i-1}) \quad D^{(i)} := \min_{x \not= x'} D^{(i)}_{x,x'}.
\]

When \( g \) is linear, \( D^{(i)}_{x,x'}(u_{i-1}^{(i-1)}) \) does not depend on \( x, x' \) or \( u_{i-1}^{(i-1)} \), in which case we will use the notation \( D^{(i)} \) instead of \( D^{(i)}_{x,x'}(u_{i-1}^{(i-1)}) \).

From Lemma 24 in the appendix, the following lemma is obtained.

Lemma 17: For \( i \in \{0, \ldots, \ell - 1\} \),
\[
\frac{1}{q^{2\ell - 2 - \ell}} Z_{\min}(W) D^{(i)}_{x,x'} \leq Z_{x,x'}(W_{\ell}^{(i)}) \leq q^{-1-\ell} Z_{\max}(W) D^{(i)}_{x,x'}.
\]

Corollary 18: For \( i \in \{0, \ldots, \ell - 1\} \),
\[
\frac{1}{q^{2\ell - 2 - 1}} Z_{\min}(W) D^{(i)}_{\min} \leq Z_{\min}(W_{\ell}^{(i)}) \leq q^{-1-\ell} Z_{\max}(W) D^{(i)}_{\min}.
\]

From Proposition 15 and Corollary 18 the following theorem is obtained.

Theorem 19: Assume \( P(I_\infty(W) \in \{0, 1\}) = 1 \). It holds
\[
\lim_{n\to\infty} P(Z(W(B_1)\ldots(B_n)) < 2^{-2^{\alpha n}}) = 0
\]
for \( \beta > (1/\ell) \sum_i \log_q D^{(i)}_{\max} \).

When \( \hat{Q} \) is a linear kernel represented by a square matrix \( \hat{G} \), \( (1/\ell) \sum_i \log_q D^{(i)}_{\max} \) is called the exponent of \( \hat{G} \).

Example 20: Assume that \( \mathcal{X} \) is a field and that \( \alpha \in \mathcal{X} \) is a primitive element. For a non-zero element \( \gamma \in \mathcal{X} \), let
\[
G_{RS}(q) = \begin{bmatrix}
1 & 1 & \cdots & 1 & 0 \\
\alpha^{q-2} & \alpha^{q-3} & \cdots & \alpha^{q-2} & 1 \\
\alpha^{q-2} & \alpha^{q-3} & \cdots & \alpha^{q-3} & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\alpha^{q-2} & \alpha^{q-3} & \cdots & \alpha & 1 \\
1 & 1 & \cdots & 1 & \gamma
\end{bmatrix}
\]

Since \( G_{RS}(q) \) can be regarded as a generalization of Arikan’s original matrix. The relation between binary polar codes and binary Reed-Muller codes also holds for \( q \)-ary polar codes using \( G_{RS}(q) \) and \( q \)-ary Reed-Muller codes. From Theorem 13 the channel polarization phenomenon occurs on \( G_{RS}(q) \) for any \( \gamma \neq 0 \) when \( q \) is a prime. When \( \gamma \) is a primitive element, from Corollary 14 the channel polarization phenomenon occurs on \( G_{RS}(q) \) for any prime power \( q \). We call \( G_{RS}(q) \) the Reed-Solomon kernel since the submatrix which consists of \( i \)-th row to \( (q - 1) \)-th row of \( G_{RS}(q) \) is a generator matrix of a generalized Reed-Solomon code, which is a maximum distance separable code i.e., \( D^{(i)} = i + 1 \). Hence, the exponent of \( G_{RS}(q) \) is \( 1 \sum \log_q (i + 1) \) where \( \ell = q \). Since
\[
\frac{1}{\ell} \sum_{i=0}^{\ell-1} \log_q (i + 1) \geq \frac{1}{\ell \log_q \ell} \int_1^\ell \log_q x dx = 1 - \frac{\ell - 1}{\ell \log_q \ell}
\]
the exponent of the Reed-Solomon kernel tends to 1 as \( \ell = q \) tends to infinity. When \( q = 2^2 \), the exponent of the Reed-Solomon kernel is \( \log_2 24/(4 \log_2 4) = 0.57312 \). In Arikan’s original work, the exponent of the \( 2 \times 2 \) matrix is 0.5 [3]. In [5], Korada, Şaşoglu, and Urbanke showed that by using large kernels, the exponent can be improved, and found a matrix of size 16 whose exponent is about 0.51828. The above-mentioned Reed-Solomon kernel with \( q = 2^2 \) is reasonably small and simple but has a larger exponent than binary linear...
kernels of small size. This demonstrates the usefulness of considering $q$-ary rather than binary channels. For $q$-ary DMC where $q$ is not a prime, it can be decomposed to subchannels of input sizes of prime numbers \cite{korada2009polar} by using the method of multilevel coding \cite{imai1993new}. The above example shows that when $q$ is a power of a prime, without the decomposition of $q$-ary DMC, asymptotically better coding scheme can be constructed by using $q$-ary polar codes with $G_{RS}(q)$.

V. CONCLUSION

The channel polarization phenomenon on $q$-ary channels has been considered. We give several sufficient conditions on kernels under which the channel polarization phenomenon occurs. We also show an explicit construction with a $q$-ary linear kernel $G_{RS}(q)$ for $q$ being a power of a prime. The exponent of $G_{RS}(q)$ is $\log_e(q!)/(q \log_q q)$ which is larger than the exponent of binary matrices of small size even if $q = 4$. Our discussion includes channel polarization on non-linear kernels as well. It is known that non-linear binary codes may have a larger minimum distance than linear binary codes, e.g., the Nordstrom-Robinson codes \cite{1996arXiv0901.0536A}. This implies possibility that there exists a non-linear kernel with a larger exponent than any linear kernel of the same size.

APPENDIX

**Lemma 21:**

$$\frac{1}{Z(x,x')W(x,x')} \leq Z_{x,x'}(W_{u_0^{(i)}}) \leq q(q-1)^{-i} Z_{x,x'}(W_{u_0^{(i)}})$$

**Proof:** For the second inequality, one has

$$Z_{x,x'}(W_{u_0^{(i)}}) = \sum_{y_0^{(i)}} \sqrt{W_{u_0^{(i)}}(y_0^{(i)} | x)W_{u_0^{(i)}}(y_0^{(i)} | x')}$$

$$= q^i \sum_{y_0^{(i)}} \sqrt{W(i)(y_0^{(i)}, u_0^{(i)} | x)W(i)(y_0^{(i)}, u_0^{(i)} | x')}$$

$$= \frac{1}{q(q-1)^{-i}} \sum_{y_0^{(i)}} \sum_{v_0^{(i)}} W(y_0^{(i)} | u_0^{(i)}, x, v_0^{(i)})W(y_0^{(i)} | u_0^{(i)}, x', v_0^{(i)})$$

$$\leq \frac{1}{q(q-1)^{-i}} \sum_{y_0^{(i)}} \sum_{v_0^{(i)}} W(y_0^{(i)} | u_0^{(i)}, x, v_0^{(i)})W(y_0^{(i)} | u_0^{(i)}, x', v_0^{(i)})$$

The first inequality is obtained as follows.

$$Z_{x,x'}(W_{u_0^{(i)}}) = \sum_{y_0^{(i)}} \sqrt{W(i)(y_0^{(i)} | x)W(i)(y_0^{(i)} | x')}$$

$$= q^i \sum_{y_0^{(i)}} \sqrt{W(i)(y_0^{(i)}, u_0^{(i)} | x)W(i)(y_0^{(i)}, u_0^{(i)} | x')}$$

$$= \sum_{y_0^{(i)}} \left( \sum_{v_0^{(i)}} \left( \frac{1}{q(q-1)^{-i}} \times W(y_0^{(i)} | u_0^{(i)}, u_0^{(i)} | x)W(y_0^{(i)} | u_0^{(i)}, u_0^{(i)} | x') \right) \right)$$

$$\geq \frac{1}{q^2(q-1)^{-i+1}} Z_{x,x'}(W_{u_0^{(i)}}).$$

ACKNOWLEDGMENT

TT acknowledges support of the Grant-in-Aid for Scientific Research on Priority Areas (No. 18079010), MEXT, Japan.

REFERENCES

[1] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, July 2009.
[2] S. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” 2009. [Online]. Available: [http://arxiv.org/abs/0903.0307](http://arxiv.org/abs/0903.0307)
[3] E. Arıkan and E. Telatar, “On the rate of channel polarization,” in Proc. 2009 IEEE International Symposium on Information Theory, June 28-July 3 2009, pp. 1493–1495.
[4] T. Tanaka and R. Mori, “Refined rate of channel polarization,” 2010. [Online]. Available: [http://arxiv.org/abs/1001.2067](http://arxiv.org/abs/1001.2067)
[5] S. Korada, E. Şaşoğlu, and R. Urbanke, “Polar codes: characterization of exponent, bounds, and constructions,” 2009. [Online]. Available: [http://arxiv.org/abs/0901.0556](http://arxiv.org/abs/0901.0556)
[6] E. Şaşoğlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” 2009. [Online]. Available: [http://arxiv.org/abs/0908.0302](http://arxiv.org/abs/0908.0302)
[7] E. Şaşoğlu, E. Telatar, and E. Arıkan, “Polarization for arbitrary discrete memoryless channels,” in Proc. 2009 IEEE Information Theory Workshop, Taormina, Italy, 11–16 Oct. 2009, pp. 144–148.
[8] H. Imai and S. Hiraoka, “A new multilevel coding method using error-correcting codes,” Information Theory, IEEE Transactions on, vol. 23, no. 3, pp. 371–377, may 1977.
[9] F. MacWilliams and N. Sloane, The Theory of Error-Correcting Codes. North-Holland Amsterdam, 1977.