Introduction to Polar Codes

Ryuhei Mori

Postdoctoral Fellow of ELC P≠NP and
Tokyo Institute of Technology

Workshop on Modern Error Correcting Codes
August 30, 2013
Tokyo Japan
Polar codes [Arıkan 2008]

Capacity achieving codes with efficient encoding and decoding algorithms for any symmetric channels

- $O(N \log N)$ encoding and decoding complexity for blocklength $N$.
- $o(e^{-N^{1/2-\epsilon}})$ error probability for any $R < I(W)$ and $\epsilon > 0$. 
The $2 \times 2$ matrix

\[
G_2 = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
\end{bmatrix}
\]

\[
[u_1 \quad u_2] \begin{bmatrix}
1 & 0 \\
1 & 1 \\
\end{bmatrix} = [x_1 \quad x_2]
\]
Kronecker product

\[ G_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \otimes^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \]
Kronecker product

\[ G_{2n} = (I_{2^{n-1}} \otimes G_2) R_{2^n} (I_2 \otimes G_{2^{n-1}}) = B_{2^n} G_2^{\otimes n} \]
Encoding of polar codes

∅: frozen bit  ≠ check nodes = N \log N
Encoding of polar codes

∅: frozen bit  ≠ check nodes = $N \log N$
Encoding of polar codes

Encoding complexity $\propto \# \text{ check nodes} = N \log N$
Decoding of polar codes

Successive cancellation (SC) decoding: Sequential decoding from the top to the bottom

$F$: The index set of frozen bits

- If $i \in F$, $\hat{u}_i = 0$.
- If $i \notin F$, $\hat{u}_i = \arg \max_{0,1} W^{(i)}(\hat{u}_0^{i-1}, y_0^{N-1} | u_i)$. 
Decoding of polar codes
Decoding of polar codes

ML on a tree $\iff$ belief propagation (BP)
Question

- Why do polar codes achieve symmetric capacity with SC decoding?
- Which bits should be chosen as information bits?

Channel polarization
Channel polarization

- $U_0^{N-1}$: Uniform random variable on $\{0, 1\}^N$.
- $X_0^{N-1} := U_0^{N-1} G_N$.
- $Y_0^{N-1}$: Random variable corresponding output of $W^n$

\[NI(W) = I(U_0^{N-1}; Y_0^{N-1}) = \sum_{i=0}^{N-1} I(U_i; Y_0^{N-1}, U_0^{i-1}) = \sum_{i=0}^{N-1} I(W^{(i)})\]

where $W^{(i)} : U_i \mapsto Y_0^{N-1}, U_0^{i-1}$.

**Theorem (Channel polarization [Arıkan 2008])**

\[
\frac{\left| \left\{ i \in \{0, 1, \ldots, N - 1\} \mid I(W^{(i)}) > 1 - \epsilon \right\} \right|}{N} = I(W)
\]

\[
\frac{\left| \left\{ i \in \{0, 1, \ldots, N - 1\} \mid I(W^{(i)}) < \epsilon \right\} \right|}{N} = 1 - I(W)
\]
Channel polarization

$W = \text{BEC}(0.4), I(W) = 0.6, N = 1024.$
Process of subchannels
Binary erasure channel

\[ P_e(W^{(1)}) = \varepsilon^2, \quad P_e(W^{(0)}) = 1 - (1 - \varepsilon)^2 \]

\[ \frac{P_e(W^{(0)}) + P_e(W^{(1)})}{2} = \varepsilon = P_e(W) \]
Martingale convergence theorem

\[ B_1, \ldots, B_n: \text{Uniform i.i.d. 0-1 random variables.} \]

\[ Z_n := P_e(W(B_1)(B_2)\cdots(B_n)). \]

The random process \( Z_n \) is a martingale, i.e.,

\[
\mathbb{E}[Z_{n+1} \mid B_1, \ldots B_n] = \frac{Z_n^2 + 1 - (1 - Z_n)^2}{2} = Z_n
\]

**Theorem (Martingale convergence theorem)**

*If \( \sup_n \mathbb{E}[X_n] < \infty \) then a martingale \( X_n \) converges almost surely.*
Convergence to 0-1 random variable

\[
Z_n = \begin{cases} 
  Z_{n-1}^2, & \text{w.p. } \frac{1}{2} \\
  1 - (1 - Z_{n-1})^2, & \text{w.p. } \frac{1}{2}
\end{cases}
\]

\[Z_0 = \epsilon\]

\[Z_\infty \text{ takes 1 with prob. } \epsilon \text{ and takes 0 with prob. } 1 - \epsilon.\]
From BEC to general channel

For general channel, \( P_e(W^{(B_1)\cdots(B_n)}) \) is not a martingale.

But, \( I(W^{(B_1)\cdots(B_n)}) \) is a martingale.

Theorem (Channel polarization [Arıkan 2008])

\[
\frac{\left| \left\{ i \in \{0, 1, ..., N - 1\} \mid I(W^{(i)}) > 1 - \epsilon \right\} \right|}{N} = I(W)
\]

\[
\frac{\left| \left\{ i \in \{0, 1, ..., N - 1\} \mid I(W^{(i)}) < \epsilon \right\} \right|}{N} = 1 - I(W).
\]
Construction and error

Bits associated to high $I(W^{(i)})$ subchannels are chosen as information bits.

Bits associated to low $I(W^{(i)})$ subchannels are chosen as frozen bits.

$I(W^{(i)})$ (and also $P_e(W^{(i)})$) can be evaluated by density evolution [Mori and Tanaka 2009].

\[ B_i := \{ u_1^N, y_1^N \mid \hat{u}_1^{i-1} = u_1^{i-1}, \hat{U}_i(y_1^N, \hat{u}_1^{i-1}) \neq u_i \} \]
\[ \subseteq \{ u_1^N, y_1^N \mid \hat{U}_i(y_1^N, u_1^{i-1}) \neq u_i \} =: A_i. \]

\[ P_{\text{error}}(F) = \Pr \left( \bigcup_{i \in F^c} B_i \right) = \sum_{i \in F^c} \Pr(B_i) \leq \sum_{i \in F^c} \Pr(A_i) = \sum_{i \in F^c} P_e(W_N^{(i)}) \]
Table of contents

- Speed of polarization [Arıkan and Talatar 2008]
- $\ell \times \ell$ construction [Korada, Şaşoğlu, and Urbanke 2009]
- Detailed speed of polarization [Tanaka and Mori 2010], [Hassani and Urbanke 2010], [Hassani, Mori, Tanaka, and Urbanke 2012]
- Scaling of polarization [Koorada, Montanari, Telatar and Urbanke 2010]
- Compound capacity [Hassani, Korada and Urbanke 2009]
- Non-binary polar codes [Şaşoğlu, Telatar and Arıkan 2009], [Mori and Tanaka 2010]
Speed of polarization

[Arıkan and Telatar 2008] For any $\epsilon > 0$

$$\lim_{n \to \infty} \Pr \left( Z_n < 2^{-N^\frac{1}{2} - \epsilon} \right) = I(W)$$

$$\lim_{n \to \infty} \Pr \left( Z_n < 2^{-N^\frac{1}{2} + \epsilon} \right) = 0$$

Hence, error probability of SC decoding for polar codes is $o \left( 2^{-N^\frac{1}{2} - \epsilon} \right)$ and $\omega \left( 2^{-N^\frac{1}{2} + \epsilon} \right)$ for any $\epsilon > 0$. 

Table of contents

- Speed of polarization [Arıkan and Talatar 2008]
- $\ell \times \ell$ construction [Korada, Şaşoğlu, and Urbanke 2009]
- Detailed speed of polarization [Tanaka and Mori 2010], [Hassani and Urbanke 2010], [Hassani, Mori, Tanaka, and Urbanke 2012]
- Scaling of polarization [Koarada, Montanari, Telatar and Urbanke 2010]
- Compound capacity [Hassani, Korada and Urbanke 2009]
- Non-binary polar codes [Şaşoğlu, Telatar and Arıkan 2009], [Mori and Tanaka 2010]
Polar codes using an $\ell \times \ell$ matrix $G$ instead of $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. [Korada, Şaşoğlu, and Urbanke 2009]

$E(G) \in (0, 1)$ is called an exponent of $G$ if it holds

$$\lim_{n \to \infty} \Pr \left( Z_n < 2^{-N^{E(G)-\epsilon}} \right) = I(W)$$

$$\lim_{n \to \infty} \Pr \left( Z_n < 2^{-N^{E(G)+\epsilon}} \right) = 0$$

for any $\epsilon > 0$.

Obviously, $E \left( \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \right) = \frac{1}{2}$. 
Exponent of matrix

[Korada, Şaşoğlu and Urbanke 2009]

\[ E(G) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log_\ell D_i \]

For \( \ell \times \ell \) matrix \( G \), partial distance \( D_i \) is defined as

\[ D_i := d(g_i, \langle g_{i+1}, \ldots, g_\ell \rangle), \quad i = 1, \ldots, \ell - 1 \]

\[ D_\ell := d(g_\ell, 0) \]

Here, \( g_i \) is \( i \)-th row of \( G \) and \( \langle g_i, \ldots, g_\ell \rangle \) is the subcode spanned by \( g_i, \ldots, g_\ell \).

Example

\[
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 1 \\
1 & 1 & 0 \\
\end{bmatrix}
\]

\( D_1 = 1, \ D_2 = 1, \ D_3 = 3 \)

\( D_1 = 1, \ D_2 = 2, \ D_3 = 2 \)
Exponent of matrix

\[ E(G) = \frac{1}{\ell} \sum_{i=1}^{\ell} \log \ell D_i \]

For

\[ E_\ell := \max_{G \in \{0,1\}^{\ell \times \ell}} E(G) \quad \text{it holds} \quad \lim_{\ell \to \infty} E_\ell = 1 \]

\[ E_\ell \leq \frac{1}{2} \text{ for } \ell \leq 14 \text{ and } E_{16} = 0.51828 > \frac{1}{2}. \]

[Korada, Şaşoğlu, and Urbanke 2009]
Table of contents

- Speed of polarization [Arıkan and Talatar 2008]
- $\ell \times \ell$ construction [Korada, Şaşoğlu, and Urbanke 2009]
- Detailed speed of polarization [Tanaka and Mori 2010], [Hassani and Urbanke 2010], [Hassani, Mori, Tanaka, and Urbanke 2012]
- Scaling of polarization [Koarada, Montanari, Telatar and Urbanke 2010]
- Compound capacity [Hassani, Korada and Urbanke 2009]
- Non-binary polar codes [Şaşoğlu, Telatar and Arıkan 2009], [Mori and Tanaka 2010]
Detailed speed of polarization

[Tanaka and Mori 2010] [Hassani and Urbanke 2010]
[Hassani, Mori, Tanaka and Urbanke 2011]

For \( R \in (0, 1) \),

\[
\lim_{n \to \infty} \Pr \left( Z(W_n) \leq 2^{-\ell n E(G) + \sqrt{n V(G)} Q^{-1}(R/I(W)) + f(n)} \right) = R
\]

for any \( f(n) = o(\sqrt{n}) \) where

\[
Q(x) := \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx
\]

\[
E(G) := \frac{1}{\ell} \sum_{i=1}^{\ell} \log_{\ell} D_i(G)
\]

\[
V(G) := \frac{1}{\ell} \sum_{i=1}^{\ell} \left( \log_{\ell} D_i(G) - E(G) \right)^2.
\]
Table of contents

- Speed of polarization [Arıkan and Talatar 2008]
- $\ell \times \ell$ construction [Korada, Şaşoğlu, and Urbanke 2009]
- Detailed speed of polarization [Tanaka and Mori 2010], [Hassani and Urbanke 2010], [Hassani, Mori, Tanaka, and Urbanke 2012]
- Scaling of polarization [Koarada, Montanari, Telatar and Urbanke 2010]
- Compound capacity [Hassani, Korada and Urbanke 2009]
- Non-binary polar codes [Şaşoğlu, Telatar and Arıkan 2009], [Mori and Tanaka 2010]
Scaling of polar codes $\text{BEC}(\epsilon = 0.5)$

$N = 2^{10}, 2^{12}, 2^{14}, 2^{16}, 2^{18}$
Scaling of polar codes $\text{BEC}(\epsilon = 0.5)$

$N \to \infty$ while $R$ fixed $\iff$ Shannon–Gallager type analysis
Scaling of polar codes $\text{BEC}(\epsilon = 0.5)$

$N \to \infty$ while $P_e$ fixed $\iff$ Weiss–Dobrushin–Strassen type analysis
Scaling of polarization

[Korada, Montanari, Telatar, and Urbanke 2010]
[Hassani, Alishahi, and Urbanke 2010]

For arbitrary fixed $a \in (0, 1)$

$$F_N(\epsilon) := \Pr(\epsilon \leq Z(W_n) \leq a).$$

Scaling Assumption:
There exists $\mu > 0$ (called a scaling parameter) such that for any $\epsilon \in (0, a]$,

$$F(\epsilon) := \lim_{N \to \infty} N^{\frac{1}{\mu}} F_N(\epsilon) \in (0, \infty).$$

If the scaling assumption holds,

$$F^{-1}(N^{\frac{1}{\mu}} (I(W) - R)) \leq P_e.$$
Scaling parameter of polar codes

[Korada, Montanari, Telatar, and Urbanke 2010]
[Hassani, Alishahi, and Urbanke 2010]

\[ P_e \geq F^{-1}(N^\frac{1}{\mu}(I(W) - R)) \]
\[ NR \leq NI(W) - N^{1-\frac{1}{\mu}} F(P_e) \]

This is Weiss–Dobrushin–Strassen type analysis.

From the scaling assumption

\[ -\frac{1}{\mu} = \lim_{n \to \infty} \frac{1}{n} \log \Pr(a \leq Z(W_n) \leq b) \]

It can be evaluated for BEC, \( 1/\mu \approx 0.2757 \).
By Gaussian approximation, for AWGN channel, \( 1/\mu \approx 0.2497 \).
Random codes and LDPC codes have \( 1/\mu = 0.5 \).
Table of contents

- Speed of polarization [Arıkan and Talatar 2008]
- $\ell \times \ell$ construction [Korada, Şasoğlu, and Urbanke 2009]
- Detailed speed of polarization [Tanaka and Mori 2010], [Hassani and Urbanke 2010], [Hassani, Mori, Tanaka, and Urbanke 2012]
- Scaling of polarization [Koarada, Montanari, Telatar and Urbanke 2010]
- Compound capacity [Hassani, Korada and Urbanke 2009]
- Non-binary polar codes [Şasoğlu, Telatar and Arıkan 2009], [Mori and Tanaka 2010]
Compound capacity of polar codes

[Hassani, Korada, and Urbanke 2009]

\( \mathcal{W} \): A set of channels

\[
C(\mathcal{W}) = \max_{P_X} \inf_{W \in \mathcal{W}} I(X; Y)
\]

The compound capacity of polar codes with SC decoding is

\[
\lim_{N \to \infty} \sum_{i=1}^{N} \inf_{W \in \mathcal{W}} I(W^{(i)}_N)
\]

For BEC(0.5) and BSC(0.11002), the compound capacity is about 0.4816.
Table of contents

- Speed of polarization [Arıkan and Talatar 2008]
- $\ell \times \ell$ construction [Korada, Şaşoğlu, and Urbanke 2009]
- Detailed speed of polarization [Tanaka and Mori 2010], [Hassani and Urbanke 2010], [Hassani, Mori, Tanaka, and Urbanke 2012]
- Scaling of polarization [Koirada, Montanari, Telatar and Urbanke 2010]
- Compound capacity [Hassani, Korada and Urbanke 2009]
- Non-binary polar codes [Şaşoğlu, Telatar and Arıkan 2009], [Mori and Tanaka 2010]
Polarization by non-binary matrix

The matrix
\[
\begin{bmatrix}
1 & 0 \\
1 & 1
\end{bmatrix}
\]
on the commutative ring \(\mathbb{Z}/q\mathbb{Z}\) polarizes any channel if and only if \(q\) is a prime.
[Şaşoğlu, Telatar, and Arıkan 2009]

The matrix
\[
\begin{bmatrix}
1 & 0 \\
1 & \gamma
\end{bmatrix}
\]
on the field \(\mathbb{F}_q\) polarizes any channel if and only if \(\mathbb{F}_p(\gamma) = \mathbb{F}_q\).
[Mori and Tanaka 2010]

Necessary and sufficient condition for \(\ell \times \ell\) matrix on \(\mathbb{F}_q\) is also obtained in [Mori and Tanaka 2012].
Let $\alpha$ be a primitive element of $\mathbb{F}_q$.

A Reed-Solomon matrix $G_{RS}(q)$ is defined as

$$
\begin{pmatrix}
\alpha^{q-2} & \alpha^{q-3} & \cdots & \alpha & 1 & 0 \\
X^{q-1} & 1 & 1 & \cdots & 1 & 1 & 0 \\
X^{q-2} & \alpha^{(q-2)(q-2)} & \alpha^{(q-3)(q-2)} & \cdots & \alpha^{q-2} & 1 & 0 \\
X^{q-3} & \alpha^{(q-2)(q-3)} & \alpha^{(q-3)(q-3)} & \cdots & \alpha^{q-3} & 1 & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\
X & \alpha^{q-2} & \alpha^{q-3} & \cdots & \alpha & 1 & 0 \\
1 & 1 & 1 & \cdots & 1 & 1 & 1
\end{pmatrix}
$$

Submatrix which consists of $i$th row to the last row is a generator matrix of extended Reed-Solomon code.

The size $\ell$ of RS matrix is $q$.

Since $G_{RS}(2) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, RS matrix can be regarded as a generalization of Arıkan’s binary matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Since $D_i = i + 1$, $E(G_{RS}(q)) = \frac{\log(q!)}{q \log q}$.
Exponent of Reed-Solomon matrix

\[ E(G_{RS}(q)) = \frac{\log(q!)}{q \log q} \]

<table>
<thead>
<tr>
<th>( q )</th>
<th>2</th>
<th>4</th>
<th>16</th>
<th>64</th>
<th>256</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(G_{RS}(q)) )</td>
<td>0.5</td>
<td>0.573120</td>
<td>0.691408</td>
<td>0.770821</td>
<td>0.822264</td>
</tr>
</tbody>
</table>

\[ \lim_{q \to \infty} E(G_{RS}(q)) = 1 \]

The exponent of binary matrix of size smaller than 32 is smaller than 0.55

[Korada, Şaşoğlu, and Urbanke 2009]

Reed-Solomon matrix is useful for obtaining large exponent!

How about the performance for finite blocklength?
Simulation result on BAWGNC \((I(W) = 0.5)\)

- Error probability vs. rate
- Binary polar codes for different values of \(N\):
  - \(N = 2^7, 2^9, 2^{11}, 2^{13}\)
- 4-ary polar codes
Polar codes and Reed-Muller codes: binary case

[Arıkan 2009]

\[
X : \begin{bmatrix} 1 & 0 \end{bmatrix} \quad (X_2, X_1) : (1, 1)(1, 0)(0, 1)(0, 0)
\]

\[
X_2X_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \quad 00
\]

\[
X_2 \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \quad 01
\]

\[
X_1 \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \quad 10
\]

\[
1 \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \quad 11
\]

Polar rule: \( i \in \{0, \ldots, 2^n - 1\} \mid P_e(W^{(i_1)\cdots(i_n)}) < \epsilon \) \\
Reed-Muller rule: \( i \in \{0, \ldots, 2^n - 1\} \mid i_1 + \cdots + i_n > k \) \\
Binary polar codes using \( \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \) and binary Reed-Muller codes are similar.

Reed-Muller rule maximizes the minimum distance.
Polar codes using RS matrix and Reed-Muller codes: \( q \)-ary case

\[
(X_2, X_1) : (2, 2) (2, 1) (2, 0) (1, 2) (1, 1) (1, 0) (0, 2) (0, 1) (0, 0)
\]

\[
\begin{bmatrix}
X_2^2 X_1^2 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
X_2^2 X_1 & 2 & 1 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\
X_2 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
X_2 X_1^2 & 2 & 2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
X_2 X_1 & 1 & 2 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\
X_2 & 2 & 2 & 2 & 1 & 1 & 1 & 0 & 0 & 0 \\
X_1^2 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\
X_1 & 2 & 1 & 0 & 2 & 1 & 0 & 2 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\]

Polar rule: \( \{ i \in \{0, \ldots, q^n - 1\} \mid P_e(W^{(i_1)\cdots(i_n)}) < \epsilon \} \)

Reed-Muller rule: \( \{ i \in \{0, \ldots, q^n - 1\} \mid i_1 + \cdots + i_n > k \} \)

\( Q \)-ary polar codes using \( G_{RS}(q) \) and \( q \)-ary Reed-Muller codes are also similar.

Hyperbolic rule: \( \{ i \in \{0, \ldots, q^n - 1\} \mid (i_1 + 1) \cdots (i_n + 1) > k \} \)

Hyperbolic rule maximizes the minimum distance
(Massey–Costello–Justesen codes, hyperbolic cascaded RS codes).
Summary

- Polar code is provably capacity-achieving codes with efficient decoding algorithm.
- Error probability of polar codes decays slowly if rate is close to the capacity.
- Asymptotic performance of polar codes can be improved by $\ell \times \ell$ matrix and non-binary matrix.
- Polar code is suitable for many problems, e.g., lossless and lossy source coding, problems with side information, multiple access channel, etc.