Lecture 6
Polar Coding

I-Hsiang Wang

Department of Electrical Engineering
National Taiwan University

ihwang@ntu.edu.tw

December 5, 2016
Since 1948, Shannon's theory has drawn the sharp boundary between the possible and the impossible in data compression and data transmission.

Once fundamental limits are characterized, the next natural question is:

How to achieve these limits with acceptable complexity?

For **lossless source coding**, it did not take us too long to find optimal schemes with low complexity:

- Huffman Code (1952): optimal for memoryless source

On the other hand, for **channel coding** and **lossy source coding**, it turns out to be much harder. It has been the holy grail for coding theorist to find codes that achieve Shannon's limit with low complexity.
In Pursuit of Capacity-Achieving Codes

Two barriers in pursuing low-complexity capacity-achieving codes:

1. **Lack of explicit construction.** In Shannon's proof, it is only proved that there exists coding schemes that achieve capacity.

2. **Lack of structure to reduce complexity.** In the proof of coding theorems, complexity issues are often neglected, while codes with structures are hard to prove to achieve capacity.

Since 90's, several practical codes were found to approach capacity – turbo code, low-density parity-check (LDPC) code, etc. They perform well empirically, but lack rigorous proof of optimality.

The first provably capacity-achieving coding scheme with acceptable complexity is polar code, introduced by Erdal Arıkan in 2007.

Later in 2012, spatially coupled LDPC codes were also shown to achieve capacity (Shrinivas Kudekar, Tom Richardson, and Rüdiger Urbanke).
Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels

Erdal Arikan, Senior Member, IEEE

The paper wins the 2010 Information Theory Society Best Paper Award.
Overview

When Arıkan introduced polar codes in 2007, he focus on achieving capacity for the general binary-input memoryless symmetric channels (BMSC), including BSC, BEC, etc.

Later, polar codes are shown to be optimal in many other settings, including lossy source coding, non-binary-input channels, multiple access channels, channel coding with encoder side information (Gelfand-Pinsker), source coding with side information (Wyner-Ziv), etc.

Instead of giving a comprehensive introduction, we shall focus on polar coding for channel coding. The outline is as follows:

1. First we introduce the concept of channel polarization.
2. Second we explore polar coding for binary input channels.
3. Finally we briefly talk about polar coding for source coding (source polarization).
In channel coding, we use the DMC $N$ times where $N$ is the *blocklength* of the coding scheme.

Since the channel is the main focus, we use the following notations throughout this lecture:

- $W$ to denote the channel $P_{Y|X}$
- $P$ to denote the input distribution $P_X$
- $I(P; W)$ to denote $I(X; Y)$.

Since we focus on BMSC, and $X \sim \text{Ber} \left( \frac{1}{2} \right)$ achieves the channel capacity of any BMSC, we shall use $I(W)$ (slight abuse of notation) to denote $I(P; W)$ when the input $P$ is $\text{Ber} \left( \frac{1}{2} \right)$.

In other words, the channel capacity of a BMSC $W$ is $I(W)$. 
1 Polarization
- Basic Channel Transformation
- Channel Polarization

2 Polar Coding
- Encoding and Decoding Architectures
- Performance Analysis
Single Usage of Channel $W$

$X \rightarrow W \rightarrow Y$

$N$ Usage of Channel $W$

$M \rightarrow \text{ENC} \rightarrow W \rightarrow \text{DEC} \rightarrow \hat{M}$

$X_1 \rightarrow W \rightarrow Y_1$

$X_2 \rightarrow W \rightarrow Y_2$

$\vdots$

$X_N \rightarrow W \rightarrow Y_N$
Arikan's Idea

Apply special transforms to both input and output
Arikan's Idea

\[ U_1 \rightarrow W_1 \rightarrow V_1 \]

\[ U_2 \rightarrow W_2 \rightarrow V_2 \]

\[ \vdots \]

\[ U_N \rightarrow W_N \rightarrow V_N \]
Arikan's Idea

Roughly $N\log(W)$ channels with capacity $\approx 1$

$U_1 \rightarrow W_1 \rightarrow V_1$

$U_2 \rightarrow W_2 \rightarrow V_2$

$\vdots$

$U_N \rightarrow W_N \rightarrow V_N$
Arıkan's Idea

Roughly $N \{W\}$ channels with capacity $\approx 1$

$U_1 \rightarrow W_1 \rightarrow V_1$

$U_2 \rightarrow W_2 \rightarrow V_2$

$\vdots$

Roughly $N \{1 - I(W)\}$ channels with capacity $\approx 0$

$U_N \rightarrow W_N \rightarrow V_N$

Equivalently some perfect channels and some useless channels $\rightarrow$ Polarization

**Coding becomes extremely simple:** simply use those perfect channels for uncoded transmission, and throw those useless channels away.
1 Polarization
   ▪ Basic Channel Transformation
   ▪ Channel Polarization

2 Polar Coding
   ▪ Encoding and Decoding Architectures
   ▪ Performance Analysis
Arıkan's Basic Channel Transformation

Consider two channel uses of $W$: 

\[ X_1 \xrightarrow{W} Y_1 \]
\[ X_2 \xrightarrow{W} Y_2 \]
Arıkan's Basic Channel Transformation

Consider two channel uses of $W$:

Apply the pre-processor: $X_1 = U_1 \oplus U_2$, $X_2 = U_2$, where $U_1 \perp U_2$, $U_1, U_2 \sim \text{Ber} \left( \frac{1}{2} \right)$.

We now have two synthetic channels induced by the above procedure:

- $W^- : U_1 \rightarrow V_1 \triangleq (Y_1, Y_2)$
- $W^+ : U_2 \rightarrow V_2 \triangleq (Y_1, Y_2, U_1)$

The above transform yields the following two crucial phenomenon:

- $I(W^-) \leq I(W) \leq I(W^+)$  \hspace{1cm} (Polarization)
- $I(W^-) + I(W^+) = 2I(W)$  \hspace{1cm} (Conservation of Information)
Example: Binary Erasure Channel

Example 1

Let $W$ be a BEC with erasure probability $\varepsilon \in (0, 1)$, and $I(W) = 1 - \varepsilon$. Find the values of $I(W^-)$ and $I(W^+)$, and verify the above properties.

**sol:** Intuitively $W^-$ is worse than $W$ and $W^+$ is better than $W$:

- For $W^-$, input is $U_1$, output is $(Y_1, Y_2)$: Only when both $Y_1$ and $Y_2$ are not erased, one can figure out $U_1! \implies W^-$ is BEC with erasure probability $1 - (1 - \varepsilon)^2 = 2\varepsilon - \varepsilon^2$.

- For $W^+$, input is $U_2$, output is $(Y_1, Y_2, U_1)$: As long as one of $Y_1$ and $Y_2$ are not erased, one can figure out $U_2! \implies W^+$ is BEC with erasure probability $\varepsilon^2$.

Hence, $I(W^-) = 1 - 2\varepsilon + \varepsilon^2$ and $I(W^+) = 1 - \varepsilon^2$. 

Example 2

Let $W$ be a BSC with crossover probability $p \in (0, 1)$, and $I(W) = 1 - H_b(p)$. Find the values of $I(W^-)$ and $I(W^+)$. 
Basic Properties

Theorem 1

For any BMSC $W$ and the induced channels \{${W^{-}, W^{+}}$\} from Arikan's basic transformation, we have

- $I(W^{-}) \leq I(W) \leq I(W^{+})$ with equality iff $I(W) = 0$ or $1$.
- $I(W^{-}) + I(W^{+}) = 2I(W)$

pf: We prove the conservation of information first:

\[
I(W^-) + I(W^+) = I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1) = I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2 | U_1) \\
= I(U_1, U_2; Y_1, Y_2) = I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1) + I(X_2; Y_2) = 2I(W).
\]

$I(W^+) = I(X_2; Y_1, Y_2, U_1) \geq I(X_2; Y_2) = I(W)$, and hence the first property holds.

(Proof of the condition for equality is left as exercise.)
The clue is now to realize that we have equality in (12.27) if, and only if, $Y_1$ is conditionally independent of $X_1$ given $U_1$ and $Y_2$. This can happen only in exactly two cases: Either $W$ is useless, i.e., $Y_1$ is independent of $X_1$ and any other quantity related with $X_1$ such that all conditioning disappears and we have $H(Y_1) \in (12.25)$ (this corresponds to the situation when $I(W) = 0$). Or $W$ is perfect so that from $Y_2$ we can perfectly recover $U_2$ and — with the additional help of $U_1$ — also $X_1$ (this corresponds to the situation when $I(W) = 1$ bit).

It can be shown that a BEC yields the largest difference between $I(W) + I(W)$ and $I(W)$ and the BSC yields the smallest difference. Any other DMC will yield something in between. See Figure 12.4 for the corresponding plot.

If we plot the "information stretch" $I(W^+) - I(W^-)$ vs. the original $I(W)$, it turns out among all BMSC:

- BEC maximizes the stretch
- BSC minimizes the stretch

Lower boundary: $2H_b(2p(1-p)) - 2H_b(p)$, where $p = H_b^{-1}(1 - I(W))$.

Upper boundary: $2I(W)(1 - I(W))$. 

(Taken from Chap. 12.1 of Moser[4].)
1. Polarization
   - Basic Channel Transformation
   - Channel Polarization

2. Polar Coding
   - Encoding and Decoding Architectures
   - Performance Analysis
Recursive Application of Arikan's Transformation

Duplicate $W$, apply the transformation, and get $W^-$ and $W^+$. 
Duplicate $W$, apply the transformation, and get $W^-$ and $W^+$. 

Duplicate $W^-$ (and $W^+$). 

Diagram: 

- Duplicate $W$ 
- Apply transformation 
- Get $W^-$ and $W^+$ 
- Repeat for $W^-$ (and $W^+$)
Recursive Application of Arikan's Transformation

Duplicate $W$, apply the transformation, and get $W^{-}$ and $W^{+}$.

Duplicate $W^{-}$ (and $W^{+}$).

Apply the transformation on $W^{-}$, and get $W^{--}$ and $W^{+-}$. 
Recursive Application of Arikan's Transformation

Duplicate $W$, apply the transformation, and get $W^-$ and $W^+$.  
Duplicate $W^-$ (and $W^+$).
Apply the transformation on $W^-$, and get $W^{--}$ and $W^{-+}$.  
Apply the transformation on $W^+$, and get $W^{+-}$ and $W^{++}$.  

Diagram:

- Duplicate $W$, apply the transformation, and get $W^-$ and $W^+$.  
- Duplicate $W^-$ (and $W^+$).
- Apply the transformation on $W^-$, and get $W^{--}$ and $W^{-+}$.  
- Apply the transformation on $W^+$, and get $W^{+-}$ and $W^{++}$.  

Diagram:

- Duplicate $W$, apply the transformation, and get $W^-$ and $W^+$.  
- Duplicate $W^-$ (and $W^+$).
- Apply the transformation on $W^-$, and get $W^{--}$ and $W^{-+}$.  
- Apply the transformation on $W^+$, and get $W^{+-}$ and $W^{++}$.  

Diagram:
Recursive Application of Arikan's Transformation

Duplicate $W$, apply the transformation, and get $W^-$ and $W^+$. 

Duplicate $W^-$ (and $W^+$). 

Apply the transformation on $W^-$, and get $W^{--}$ and $W^{-+}$. 

Apply the transformation on $W^+$, and get $W^{+-}$ and $W^{++}$. 

\[ \vdots \]

We can keep going and going, until the desired blocklength is reached.
After one recursion and getting $W^{-}$ and $W^{+}$, let us duplicate them.
Polarized Channels after Recursive Application

Apply the transformation on $W^-$:

\[
\begin{align*}
W^-: U_1 &\rightarrow ((Y_1, Y_2), (Y_3, Y_4)) = Y_4 \\
W^+: U_2 &\rightarrow ((Y_1, Y_2), (Y_3, Y_4), U_1) = (Y_4, U_1)
\end{align*}
\]
Polarized Channels after Recursive Application

Apply the transformation on $W^+$:

\[
W^{+-} : U_3 \rightarrow ((Y_1, Y_2, U_1 \oplus U_2), (Y_3, Y_4, U_2)) = (Y^4, U^2)
\]

\[
W^{++} : U_4 \rightarrow ((Y_1, Y_2, U_1 \oplus U_2), (Y_3, Y_4, U_2), U_3) = (Y^4, U^3)
\]
Polarized Channels after Recursive Application

Putting things together, we have:

\[ W^{--} : U_1 \rightarrow (Y^4, \emptyset) \]
\[ W^{-+} : U_2 \rightarrow (Y^4, U^1) \]
\[ W^{+-} : U_3 \rightarrow (Y^4, U^2) \]
\[ W^{++} : U_4 \rightarrow (Y^4, U^3) \]
Recursive Application of Arikan's Transformation

With proper naming of the inputs (do it yourself), \( \ell \)-times recursion generates a system with \( N = 2^\ell \) channel uses.

The \( N \) polarized channels are \( W^{s_1,\ldots,s_\ell} \), \( s_j \in \{+,-\} \), where

\[
W^{s_1,\ldots,s_\ell} : U_i \rightarrow (Y^N, U^{i-1})
\]

If we set \(- \leftrightarrow 0 \) and \(+ \leftrightarrow 1\), then the index \( i \) of channel \( W^{s_1,\ldots,s_\ell} \) is

\[ i = 1 + \sum_{j=1}^{\ell} s_j 2^{\ell-j}, \]

one plus the number with binary representation of \((s_1, \ldots, s_\ell)\) (MSB \( \rightarrow \) LSB).

In the following we use \( W^{(i)}_N \) to denote the \( W^{s_1,\ldots,s_\ell} \) generated above, for \( i = 1, \ldots, N, N = 2^\ell \).
Theorem 2 (Channel Polarization)

For any BMSC $W$, the polarized channels $\left\{ W_N^{(i)} \left| i = 1, \ldots, N \right. \right\}$ ($N = 2^\ell$) satisfy the following:

For all $a, b$ such that $0 < a < b < 1$,

\[
\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \in [0, a) \right\} \right| = 1 - I(W) \quad \text{and}
\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \in (b, 1] \right\} \right| = I(W)
\]

\[
\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \in [a, b] \right\} \right| = 0
\]

Interpretation: When $N$ is sufficiently large, roughly $N I(W)$ of them are noiseless (capacity = 1), and $N (1 - I(W))$ of them are useless (capacity = 0).
\[ \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right| \]

\[ N = 2^0 \]
\[ \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right| \]

\[ N = 2^1 \]
\[
\frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right|
\]

\[N = 2^2\]
\[ \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right| \]

\[ N = 2^4 \]
\[ \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right| \]

\[ N = 2^8 \]
\frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right| \leq \frac{1}{2^{12}}
\[
\frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right|
\]

\[
N = 2^{20}
\]
\[ \frac{1}{N} \left| \left\{ i : I \left( W_N^{(i)} \right) \leq \varepsilon \right\} \right| \]

\[ N = \infty \]
Proof of Channel Polarization

pf: Define the averaged first and second moment of \( \{ W_{2^\ell}^{(i)} \mid i = 1, \ldots, 2^\ell \} \) as follows: \((2^\ell = N)\)

\[
\mu_\ell \equiv \frac{1}{2^\ell} \sum_{i=1}^{2^\ell} I \left( W_{2^\ell}^{(i)} \right), \quad \nu_\ell \equiv \frac{1}{2^\ell} \sum_{i=1}^{2^\ell} \left( I \left( W_{2^\ell}^{(i)} \right) \right)^2
\]

Due to the conservation of information of Arikan's transformaion (Theorem 1), \( \mu_\ell = I(W) \) for all \( \ell \).

As for the averaged second moment, note that

\[
\frac{1}{2} \left( \left( I(W^+) \right)^2 + \left( I(W^-) \right)^2 \right) = \left( \frac{1}{2} \left( I(W^+) + I(W^-) \right) \right)^2 + \left( \frac{1}{2} \left( I(W^+) - I(W^-) \right) \right)^2 = I(W)^2 + \Delta(W)^2,
\]

where \( \Delta(W) \equiv \frac{1}{2} \left( I(W^+) - I(W^-) \right) \).
\[ \nu_{\ell+1} = \frac{1}{2^\ell} \sum_{i=1}^{2^\ell} I\left( W_{2^\ell}^{(i)} \right)^2 + \Delta \left( W_{2^\ell}^{(i)} \right)^2 \geq \nu_\ell + \kappa(a, b)^2 \theta_\ell(a, b), \]  

(1)

where

\[ \kappa(a, b) \triangleq \min \{ \Delta(W_{BSC_a}), \Delta(W_{BSC_b}) \} \]

\[ \theta_\ell(a, b) \triangleq \frac{1}{2^\ell} \left\lfloor \left\{ i : I\left( W_{2^\ell}^{(i)} \right) \in [a, b] \right\} \right\rfloor. \]

Hence, \( \{\nu_\ell\} \) form a non-decreasing sequence.

Meanwhile, since all channels are binary-input, \( I\left( W_{2^\ell}^{(i)} \right) \leq 1 \), and therefore \( \nu_\ell \leq 1 \).

(Modified from Chap. 12.1 of Moser[4].)
Hence, $\nu_0 \leq \nu_1 \leq \ldots \leq \nu_\ell \leq \ldots \leq 1 \implies \lim_{\ell \to \infty} \nu_\ell$ exists.

By (1), we have
\[
\theta_\ell(a, b) \leq \frac{\nu_{\ell+1} - \nu_\ell}{\kappa(a, b)^2} \implies \lim_{\ell \to \infty} \theta_\ell(a, b) = 0. \quad \text{(since } \lim_{\ell \to \infty} \nu_\ell \text{ exists.)}
\]

Finally, define $\alpha_\ell(a) \triangleq \frac{1}{2\ell} \left| \left\{ i : I(W^{(i)}_{2\ell}) \in [0, a) \right\} \right|$ and $\beta_\ell(b) \triangleq \frac{1}{2\ell} \left| \left\{ i : I(W^{(i)}_{2\ell}) \in (b, 1] \right\} \right|$.

Observe that
\[
I(W) = \mu_\ell \leq a \cdot \alpha_\ell(a) + b \cdot \theta_\ell(a, b) + 1 \cdot \beta_\ell(b) = a + (b - a)\theta_\ell(a, b) + (1 - a)\beta_\ell(b)
\]
\[
1 - I(W) = 1 - \mu_\ell \leq 1 - 0 \cdot \alpha_\ell(a) - a \cdot \theta_\ell(a, b) - b \cdot \beta_\ell(b) = (1 - b) + (b - a)\theta_\ell(a, b) + b\alpha_\ell(a)
\]

It is then not hard to show that $\liminf_{\ell \to \infty} \beta_\ell(b) \geq I(W)$ and $\liminf_{\ell \to \infty} \alpha_\ell(b) \geq 1 - I(W)$.

Proof is complete by sandwich principle.\hfill \Box
From Channel Polarization to Polar Coding

Recall the original goal:

Caveat:

What we have done, however, is the following: we created $N = 2^\ell$ polarized channels

$$W_N^{(i)} : U_i \rightarrow V_i = (Y^N, U^{i-1}) .$$

However, we cannot obtain the true $U^{i-1}$ from the channel output $Y^N$.

This issue can be fixed by successive decoding, where $V_i \triangleq (Y^N, \hat{U}^{i-1})$ instead of $(Y^N, U^{i-1})$.

Encoding is based on those "synthetic" polarized channels $\{W_N^{(i)}\}$, and the $i$-th synthetic channel is a good approximation as long as $\hat{U}^{i-1} = U^{i-1}$ with high probability.
1 Polarization
   - Basic Channel Transformation
   - Channel Polarization

2 Polar Coding
   - Encoding and Decoding Architectures
   - Performance Analysis
1 Polarization
- Basic Channel Transformation
- Channel Polarization

2 Polar Coding
- Encoding and Decoding Architectures
- Performance Analysis
Overview of Polar Coding Architecture

1 Preparation

Generate the $N = 2^\ell$ synthetic polarized channels $\{W^{(i)}_N \mid i = 1, 2, \ldots, N\}$.

2 Encoding

To encode $K$ information bits into an $N$-bit codeword, the encoder picks a subset $\mathcal{A} \subseteq \{1 : N\}$ of synthetic polarized channels from the $N$ channels above, based on the qualities of them:

- For each $i \in \mathcal{A}$, use $U_i$ to sent an information bit.
- For each $i \in \mathcal{F} \triangleq \mathcal{A}^c$, fix $U_i$ to a dummy bit $u_i^*$ (frozen bits).

3 Decoding is based on successive cancellation, where

- the decoded $\hat{U}_i$ is determined by $(Y^N, \hat{U}^{i-1})$ if $i \in \mathcal{A}$.
- the decoded $\hat{U}_i = u_i^*$, the pre-fixed dummy frozen bit, if $i \in \mathcal{F}$.
For the $i$-th synthesized channel, its input is $U_i$, output is $(Y^N, U^{i-1})$, and the channel law is

$$W_N^{(i)} (y^N, u^{i-1} | u_i) = \frac{1}{2^{N-1}} \sum_{u_i+1=0}^1 \ldots \sum_{u_N=0}^1 P (y^N | u^N),$$

where $P (y^N | u^N) = \prod_{i=1}^N W (y_i | x_i)$. The relationship between $x^N$ and $u^N$ is described in the next slide.
Relation between $U^N$ and $X^N$

As mentioned before, with a proper "bit-reversal permutation" of indices of $U_i$'s, one can obtain $\tilde{U}_i$'s, where the relationship between $X^N$ and $\tilde{U}^N$ can be characterized by the $\ell$-times Kronecker product:

$$X^N = \tilde{U}^N \cdot G_N,$$

where $G_N = G_2^\otimes \ell \triangleq G_2 \otimes \ldots \otimes G_2$, $N = 2^\ell$, and $G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Easy to check: $(G_N)^{-1} = G_N$.

The "bit-reversal permutation" $\sigma_{br}$ is described as follows: for $i = 1 + \sum_{j=1}^\ell s_j 2^{\ell-j}$,

$$\sigma_{br}(i) = 1 + \sum_{j=1}^\ell s_j 2^{j-1}.$$

In other words, the binary representation of $\sigma_{br}(i) - 1$ is the reverse of that of $i - 1$, and vice versa. We shall use $R_N$ to denote the matrix representation of $\sigma_{br}$. Easy to check: $(R_N)^{-1} = R_N$. Hence,

$$X^N = U^N \cdot R_N G_N \quad \text{and} \quad U^N = X^N \cdot G_N R_N.$$
Encoding

Two things to be specified for polar encoding:

1. Determine the **active set** $\mathcal{A}$ and the **frozen set** $\mathcal{F}$.

2. Determine what to send on the indices of the frozen set.

**Selection of the Frozen Set**

Let $K$ denote the number of information bits to be delivered. Then, in principle, one should choose $\mathcal{A}$ and $\mathcal{F}$ such that $|\mathcal{A}| = K$ and $\forall i \in \mathcal{A}, j \in \mathcal{F}$, channel $W_N^{(i)}$ has "better quality" than channel $W_N^{(j)}$.

*How to evaluate "quality" of the synthetic polarized channels $\{W_N^{(i)}\}$? Discussed later.

**Setting Values of the Frozen Bits**

The values of the frozen bits are known to both encoder and decoder – part of the codebook design.
Encoding Architecture

\[ \mathbf{x} = \mathbf{u} \mathbf{R}_N \mathbf{G}_N \]

where

- \( \mathbf{u} \triangleq [u_1 \ u_2 \ \ldots \ u_N] \) denotes the uncoded bits (union of information and frozen bits).
- \( \mathbf{x} \triangleq [x_1 \ x_2 \ \ldots \ x_N] \) denotes the coded bits (codeword).
- \( \mathbf{G}_N \triangleq \mathbf{G}_2 \otimes \ldots \otimes \mathbf{G}_2 \) is the encoding matrix \( \ell \) times
  \( (N = 2^\ell \text{ and } \mathbf{G}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} ). \)
- \( \mathbf{R}_N \) is the bit-reversal permutation matrix.

Rate \( R = \frac{|A|}{N} \).
Encoding Architecture

\[ x = u R_N G_N \]

where

- \( u \triangleq [u_1 \ u_2 \ \ldots \ u_N] \) denotes the uncoded bits (union of information and frozen bits).
- \( x \triangleq [x_1 \ x_2 \ \ldots \ x_N] \) denotes the coded bits (codeword).
- \( G_N \triangleq G_2 \otimes \ldots \otimes G_2 \) is the encoding matrix \( \ell \) times
  \((N = 2^\ell \text{ and } G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix})\).
- \( R_N \) is the bit-reversal permutation matrix.

\[ \text{Rate } R = \frac{|A|}{N}. \]
Decoding

**Successive Cancellation Decoding (SC Decoding)**
Upon receiving $y^N$, the decoder starts to decode $u_i$ from $i = 1$ to $i = N$ in a sequential manner, following the rule below:

- $\hat{u}_i = u_i^*$ if $i \in \mathcal{F}$.
- $\hat{u}_i = \arg\max_{u \in \{0,1\}} W_N^{(i)}(y^N, \hat{u}^{i-1}|u)$ if $i \in \mathcal{A}$.

In words, the decoder performs **bit-wise sequential decoding**.

**Note**: For $i \in \mathcal{A}$, the decoding rule for that bit is not maximum likelihood decoding because it does not make use of all frozen bits. In particular, $\{u_j^*: j \in \mathcal{F}, j > i\}$ are not harnessed when decoding $U_i$. 
1 Polarization
- Basic Channel Transformation
- Channel Polarization

2 Polar Coding
- Encoding and Decoding Architectures
- Performance Analysis
**Probability of Error**

Under SC decoding, probability of error of the proposed polar coding scheme depends on (1) channel $W$, (2) blocklength $N$, (3) code rate $\frac{K}{N}$, (4) frozen set $\mathcal{F} \subset [1 : N]$, (5) frozen bits $u_{\mathcal{F}}$.

**Notation:** here we use $u_{\mathcal{F}}$ and $u_{\mathcal{A}}$ to denote the frozen bits and information bits respectively.

First, define the average (over all codewords) probability of error with given frozen bits $u_{\mathcal{F}}$:

$$P_e^{(N)} \left( \frac{K}{N}, \mathcal{F}, u_{\mathcal{F}} \right) \triangleq \mathbb{P} \left\{ U^N \neq \hat{U}^N \right\} = \sum_{u_{\mathcal{A}} \in \{0,1\}^K} 2^{-K} \cdot \mathbb{P} \left\{ \exists i \in \mathcal{A} \text{ s.t. } \hat{U}_i \neq U_i \mid U_{\mathcal{A}} = u_{\mathcal{A}} \right\}.$$

Next, we further average over uniformly randomly chosen frozen bits $u_{\mathcal{F}}$ and define

$$P_e^{(N)} \left( \frac{K}{N}, \mathcal{F} \right) \triangleq \sum_{u_{\mathcal{F}} \in \{0,1\}^{N-K}} 2^{-(N-K)} \cdot P_e^{(N)} \left( \frac{K}{N}, \mathcal{F}, u_{\mathcal{F}} \right).$$
Upper Bounding the Probability of Error of Polar Coding (1)

Observe that

\[ P_{e}^{(N)} \left( \frac{K}{N}, \mathcal{F} \right) = \mathbb{P} \left\{ \bigcup_{i \in A} \left\{ \hat{U}_i \neq U_i, \hat{U}^{i-1} = U^{i-1} \right\} \right\} \]

\[ \leq \sum_{i \in A} \mathbb{P} \left\{ \hat{U}_i(Y^N, \hat{U}^{i-1}) \neq U_i, \hat{U}^{i-1} = U^{i-1} \right\} , \]

where \( U_i \overset{i.i.d.}{\sim} \text{Ber} \left( \frac{1}{2} \right) \), for all \( i = 1, 2, \ldots, N \). The inequality is due to union bound.

Recall that decoding function \( \hat{U}_i(Y^N, \hat{U}^{i-1}) = \arg \max_{u \in \{0,1\}} W_N^{(i)} \left( Y^N, \hat{U}^{i-1} \middle| u \right) \). Hence,

\[ \mathbb{P} \left\{ \hat{U}_i(Y^N, \hat{U}^{i-1}) \neq U_i, \hat{U}^{i-1} = U^{i-1} \right\} = \mathbb{P} \left\{ \hat{U}_i(Y^N, U^{i-1}) \neq U_i, \hat{U}^{i-1} = U^{i-1} \right\} \]

\[ \leq \mathbb{P} \left\{ \hat{U}_i(Y^N, U^{i-1}) \neq U_i \right\} . \]

(2)
Now things boil down to upper bounding $\mathbb{P}\left\{ \hat{U}_i(Y^N, U^{i-1}) \neq U_i \right\}$, where

$$\hat{U}_i(y^N, u^{i-1}) = \arg\max_{u \in \{0,1\}} W_N^{(i)}(y^N, u^{i-1}|u).$$

(4)

**Key Observation:** $\mathbb{P}\left\{ \hat{U}_i(Y^N, U^{i-1}) \neq U_i \right\}$ is the optimal error probability of error of a binary detection problem, since the bitwise decoder (4) above is the corresponding MAP/ML detection rule!

Next, we introduce $Z(W)$ as an error probability upper bound for a binary detection problem with input $X \sim \text{Ber}\left(\frac{1}{2}\right)$ and observation $Y$, following the probability transition law $W(y|x)$. Naturally it measures the **reliability** of a channel $W$. 
Lemma 1

For a binary detection problem with input $X \sim \text{Ber} \left( \frac{1}{2} \right)$ and observation $Y$ following the probability transition law $W(y|x)$, the optimal probability of error (note: ML is optimal)

$$\mathbb{P}\left\{ \hat{X}_{ML}(Y) \neq X \right\} \leq Z(W), \text{ where } Z(W) \triangleq \sum_{y \in Y} \sqrt{W(y|0) \cdot W(y|1)}.$$  \hspace{1cm} (5)

pf: Recall that the ML detection rule: $\hat{X}_{ML}(y) = x$ if $W(y|x) \geq W(y|x \oplus 1)$. Hence,

$$\mathbb{P}\left\{ \hat{X}_{ML}(Y) \neq X \right\} = \mathbb{E}_{X,Y} \left[ 1 \{ W(Y|X) < W(Y|X \oplus 1) \} \right] \leq \mathbb{E}_{X,Y} \left[ \sqrt{\frac{W(Y|X+1)}{W(Y|X)}} \right].$$

It is not hard to verify that $Z(W) = \mathbb{E}_{X,Y} \left[ \sqrt{\frac{W(Y|X+1)}{W(Y|X)}} \right]$ (left as exercise). \hfill \square
Properties of the Reliability Function $Z(\cdot)$

Proofs of the following properties are neglected here.

1. **Range of $Z$:** $0 \leq Z(W) \leq 1$.  
   (By Cauchy-Schwarz)

2. **Polarization:** under Arıkan's transformation,
   - $Z(W^+) = (Z(W))^2$, $Z(W^-) \leq 2Z(W) - (Z(W))^2$  
   - $Z(W^+) + Z(W^-) \leq 2Z(W)$.  
   - $Z(W^+) \leq Z(W) \leq Z(W^-)$.  
   (Reliability is improved after the polarization)

3. **Relation with $I(W)$:** $1 - Z(W) \leq I(W) \leq 1 - (Z(W))^2$.
   - $I(W) \approx 1 \iff Z(W) \approx 0$
   - $I(W) \approx 0 \iff Z(W) \approx 1$

Hence, one can expect that channel polarization (Theorem 2) still holds if we change the measure of "goodness" from capacity to reliability function.
Combining (2), (3), and Lemma 1, we arrive at a nice upper bound on $P_e(N; K/N, F)$:

$$P_e(N; K/N, F) \leq \sum_{i \in A} Z(W_N^{(i)}) .$$

\[ (6) \]

**Implications of Upper Bound (6):**

1. **How to choose the frozen set $F$?** If we would like to minimize (6), we should choose $A$ and $F$ such that $Z(W_N^{(i)}) \leq Z(W_N^{(j)})$ for all $i \in A$ and $j \in F$. In other words, we use $Z(\cdot)$ to evaluate the quality of the synthetic polarized channels.

2. **Suppose we can compute the asymptotic limit of the proportion** of synthetic polarized channels whose $Z(\cdot)$ is smaller than some $\delta_N = o(N^{-1})$. If $R$ is less than this limit, for sufficiently large $N$, we can further upper bound (6) by $NR \cdot \delta_N$ which will vanish as $N \to \infty$. 
Speed of Channel Polarization

In other words, we would like to have some theorem which gives us the following result:

$$\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : Z \left( \mathcal{W}_{N}^{(i)} \right) < \delta_{N} \right\} \right| = I(W).$$

This is a stronger version of channel polarization than Theorem 2.

To see this, note that we can easily replace $I \left( \mathcal{W}_{N}^{(i)} \right)$ by $1 - Z \left( \mathcal{W}_{N}^{(i)} \right)$ in Theorem 2 and the results remain to hold for constants $a$ and $b$, where $a, b = \Theta(1)$, invariant to $N$.

However, the desired theorem requires replacing $a$ and $b$ by $\delta_{N}$ and $1 - \delta_{N}$ respectively, where $\delta_{N} = o \left( N^{-1} \right)$. The proof of Theorem 2 presented before cannot be extended to this case.
Nevertheless, Arıkan and Telatar proved an even stronger result, where $\delta_N = 2^{-N^\beta}, \beta \in (0, \frac{1}{2})$.

Below we present this result without proving it.

**Theorem 3 (Rate of Channel Polarization [Arıkan-Telatar ISIT09])**

**Direct Part:** For $\beta \in (0, \frac{1}{2})$,

\[
\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : Z \left( W_N^{(i)} \right) < 2^{-N^\beta} \right\} \right| = I(W) \tag{7}
\]

\[
\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : Z \left( W_N^{(i)} \right) > 1 - 2^{-N^\beta} \right\} \right| = 1 - I(W) \tag{8}
\]

**Converse Part:** For $\beta > \frac{1}{2}$, if $I(W) < 1$,

\[
\lim_{N \to \infty} \frac{1}{N} \left| \left\{ i : Z \left( W_N^{(i)} \right) < 2^{-N^\beta} \right\} \right| = 0.
\]
Coding Theorem for Polar Coding

**Theorem 4 (Polar Coding Achieves Capacity of BMSC)**

Suppose the frozen set $\mathcal{F}$ is chosen such that $Z \left( W_N^{(i)} \right) \leq Z \left( W_N^{(j)} \right)$ for all $i \in \mathcal{A}$ and $j \in \mathcal{F}$. Then,

$$\lim_{N \to \infty} P_e^{(N)}(R, \mathcal{F}) \cdot 2^{N\beta} = 0$$

for any rate $R < I(W)$ and $\beta \in (0, \frac{1}{2})$. In other words, $P_e^{(N)}(R, \mathcal{F}) = o \left( 2^{-N\beta} \right)$.

**Note:** (9) guarantees that the probability of error vanishes as $N \to \infty$ for some choice of frozen bits $u_{\mathcal{F}}$, as long as $R < I(W)$, the channel capacity. Hence, it shows that polar code can achieve the capacity of the channel $W$.

**Remark:** In fact, for symmetric channels, it can be shown that (9) remains true even if we replace $P_e^{(N)}(R, \mathcal{F})$ by $P_e^{(N)}(R, \mathcal{F}, u_{\mathcal{F}})$ for any $u_{\mathcal{F}} \in \{0, 1\}^{N(1-R)}$. This will be explored in HW4.
pf: Fix some $\beta' \in (\beta, \frac{1}{2})$. Since $R < I(W)$, by Lemma 3, for $N$ sufficiently large,

$$\left| \left\{ i : Z\left(W_N^{(i)}\right) < 2^{-N\beta'} \right\} \right| > NR.$$ 

Since we pick $\mathcal{A}$ and $\mathcal{F}$ such that $|\mathcal{A}| = NR$ and all synthetic polarized channels with indices in $\mathcal{A}$ have smaller $Z(\cdot)$ than those in $\mathcal{F}$, we have

$$Z\left(W_N^{(i)}\right) < 2^{-N\beta'}, \ \forall i \in \mathcal{A}.$$

Hence, by the upper bound (6), we conclude that

$$P_e^{(N)}(R, \mathcal{F}) < NR \cdot 2^{-N\beta'} \implies P_e^{(N)}(R, \mathcal{F}) \cdot 2^{N\beta} < NR \cdot 2^{-(N\beta' - N\beta)}.$$ 

Proof is complete by observing $\lim_{N \to \infty} NR \cdot 2^{-(N\beta' - N\beta)} = 0$. 

$\square$