An Introduction to Polar Codes

T. Ganesan

gana@ti.com

SPC Lab

Aug 11th, 2012

Outline

Channel Coding- I

- Shannon's channel coding theorem proves that reliable communication is possible when $R < C$.
  - Channel capacity is achievable only when code length approaches infinity.
- Practical channel codes fall into two categories:
  - Algebraic codes and
  - Iteratively decodable codes
- Coding and decoding complexity increases exponentially with length of the codes.

Types of Encoders and Decoders- I

- Encoding
  - Linear block codes
  - Convolutional codes
  - Modern codes (Turbo, LDPC, IRA etc.)
- Decoding
  - ML decoder
  - List decoder
  - ML sequence detection decoding
  - Reduced state decoders (e.g. Fano)
  - Iterative decoders (e.g., SPA, MAP)

Types of Encoders and Decoders- II

- Issues:
  - Encoding complexity (for Large block codes)
  - Decoder complexity (ML decoding, iterative decoding, MAP)

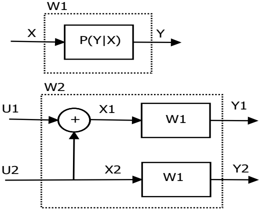
Is there any simple way of achieving capacity with less complex codes ?

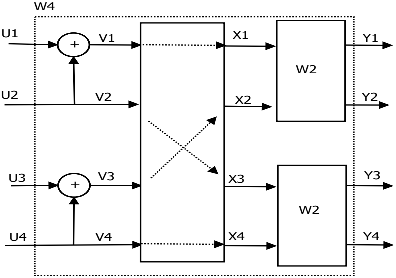
Yes! Polar codes of length N provably achieve capacity with $\mathcal{O}(N \log N)$ encoding and decoding complexity.

Channel Translation Function- I

- Consider a discrete memoryless channel (DMC) described by the conditional probability function $W = P(Y|X)$ with $X \in \mathcal{X}$ as input, $Y \in \mathcal{Y}$ as output.
- Since it is memoryless, we can say $W^N = P(Y^N|X^N) = \prod_i P(Y_i|X_i)$.
- 'Translate' the W^N channel as in the following diagram.

Channel Translation Function- II





where N is 2^n.

- The equivalent channel representation

$$\mathbf{x} = \mathbf{u} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \mathbf{u}\mathbf{G}_4$$

Channel Translation (Contd.)

- That is, the *transition probabilities* for the two channels are related as

$$W_N(\mathbf{y}|\mathbf{u}) = W^N(\mathbf{y}, \mathbf{u}\mathbf{G}_N) \; \forall \; \mathbf{y} \in \mathcal{Y}^N, \mathbf{u} \in \mathcal{X}^N.$$

- It can be generalized that $\mathbf{G}_N = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}}_{\mathbf{G}_2} \otimes \underbrace{\mathbf{G}_2}_{2} \otimes \ldots \otimes \underbrace{\mathbf{G}_2}_{n}$

Channel Splitting- I

- The translated channel W_N can be split into N binary input *coordinate channels*

$$W_N^i(\mathbf{y}, \mathbf{u}_1^{i-1}|u_i) \triangleq \sum_{\mathbf{u}_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N(\mathbf{y}|\mathbf{u})$$

  where $(\mathbf{y}, \mathbf{u}_1^{i-1})$ denotes the output of W_N^i when u_i is its input.

- That is, the i^{th} channel output is y_i with past channel inputs $\mathbf{u}_1^{i-1}$ as side information and u_i as its input.

Channel Polarization- I

- For large N, W_N^i channels polarize its output to be either close to the output of an ideal channel or worst channel.
  - In fact, the fraction of channels which polarize to ideal channel is equal to the *capacity of the underlying DMC*.
  - That is, for $\delta \in (0, 1)$, as $N \to \infty$, $I(u_i, \mathbf{y}) \in (1 - \delta, 1]$ or $I(u_i, \mathbf{y}) \in [0, \delta)$.
  - The fraction of indices $i \in [1, 2, \ldots, N]$ for which $I(u_i, \mathbf{y}) \in (1 - \delta, 1]$ goes to $I(X^N, Y^N)$.

Channel Polarization- II

- As an example, consider a BEC, $I(X^2, Y^2) = 2(1-\epsilon)$, where ϵ is the erasure probability and W_2

$\mathbf{y}$	$W(\mathbf{y}\|u_1 = 0)$	$W(\mathbf{y}\|u_1 = 1)$
00	$\frac{(1-\epsilon)^2}{2}$	0
01	0	$\frac{(1-\epsilon)^2}{2}$
$0E$	$\frac{\epsilon(1-\epsilon)}{2}$	$\frac{\epsilon(1-\epsilon)}{2}$
10	0	$\frac{(1-\epsilon)^2}{2}$
11	$\frac{(1-\epsilon)^2}{2}$	$\frac{\epsilon(1-\epsilon)}{2}$
$1E$	$\frac{\epsilon(1-\epsilon)}{2}$	$\frac{\epsilon(1-\epsilon)}{2}$
$E0$	$\frac{\epsilon(1-\epsilon)}{2}$	$\frac{\epsilon(1-\epsilon)}{2}$
$E1$	$\frac{\epsilon(1-\epsilon)}{2}$	$\frac{\epsilon(1-\epsilon)}{2}$
EE	ϵ^2	ϵ^2

Channel Polarization- III

- One can compute $W(\mathbf{y}, u_1 | u_2 = 0)$ and $W(\mathbf{y}, u_1 | u_2 = 1)$ similarly assuming u_1 is known accurately.
- Thus, the channel W_N can be split into N channels whose transition probabilities are specific to a given channel type W.

Encoding : Coset codes - I

- Let $\mathbf{G}_N$ be the generator matrix and $\mathcal{A} \subset [1, 2, \ldots N]$ is an index set with K elements. Let $\mathcal{A}^c$ denote the complement of $\mathcal{A}$.
- The channel input vector $\mathbf{x}$ can be written as a sum of 2 vectors. i.e.,

$$\mathbf{x} = \mathbf{u}_{\mathcal{A}} \mathbf{G}_{\mathcal{A}} \oplus \mathbf{u}_{\mathcal{A}^c} \mathbf{G}_{\mathcal{A}^c}$$

where $\mathbf{G}_{\mathcal{A}}$ denotes the sub-matrix of $\mathbf{G}$ formed by the rows with indices in $\mathcal{A}$.

Encoding : Coset codes - II

- If suppose $\mathbf{u}_{\mathcal{A}^c}$ is known to both encoder and decoder, then the various codewords output by the encoder are cosets with the coset index denoted by $\mathbf{u}_{\mathcal{A}^c}\mathbf{G}_{\mathcal{A}^c}$.
- The coding rate of this code is $\frac{K}{N}$.

Successive Cancellation Decoding - I

- Using its knowledge of $\mathcal{A}^c$, the decoder estimates the u_i by computing

$$\hat{u}_i = \begin{cases} u_i, & \text{if } i \in \mathcal{A}^c \\ h_i(\mathbf{y}, \hat{\mathbf{u}}_1^{i-1}), & \text{if } i \in \mathcal{A} \end{cases}$$

where

$$h_i(\mathbf{y}, \hat{\mathbf{u}}_1^{i-1}) = \begin{cases} 0, & \frac{W(\mathbf{y}, \mathbf{u}_1^{i-1} | u_i = 0)}{W(\mathbf{y}, \mathbf{u}_1^{i-1} | u_i = 1)} \geq 1 \\ 1, & \text{otherwise} \end{cases}$$

for all $\mathbf{y} \in \mathcal{Y}$, $\hat{\mathbf{u}}_1^{i-1} \in \mathcal{X}^{i-1}$.

Probability of decoding Error - I

- Let $P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ denote the probability of error for the SC-decoder when $\mathbf{u}_{\mathcal{A}} \in \mathcal{X}^K$ is selected randomly with equal probability.

$$P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c}) = \frac{1}{2^K} \sum_{\mathbf{u}_{\mathcal{A}} \in \mathcal{X}^K} \sum_{\mathbf{y} \in \mathcal{Y}^N : \hat{\mathbf{u}}(\mathbf{y}) \neq \mathbf{u}} W_N(\mathbf{y}|\mathbf{u})$$

- Averaging this over all choices of $\mathbf{u}_{\mathcal{A}^c}$ we get

$$P_e(N, K, \mathcal{A}) = \frac{1}{2^{N-K}} \sum_{\mathbf{u}_{\mathcal{A}^c} \in \mathcal{X}^{N-K}} P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$$

Probability of decoding Error - II

- It can be shown that for a symmetric binary input DMC W and any choice of parameter $(N, K, \mathcal{A})$ the above can be upper bounded as

$$P_e(N, K, \mathcal{A}) \leq \sum_{i \in \mathcal{A}} Z(W_N^i)$$

where $Z(W_N^i) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}^N} \sqrt{W_N(\mathbf{y}|u_i = 0) W_N(\mathbf{y}|u_i = 1)}$.

- Moveover, there exists $\mathbf{u}_{\mathcal{A}^c}$ such that the

$$P_e(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c}) \leq \sum_{i \in \mathcal{A}} Z(W_N^i)$$

Definition of Polar Code - I

- Given a BI-DMC W, a $\mathbf{G}_N$ coset code with parameters $(N, K, \mathcal{A}, \mathbf{u}_{\mathcal{A}^c})$ is called a *Polar code* for W if the information set $\mathcal{A}$ is chosen as a K-element subset of $[1, \ldots, N]$ such that $Z(W_N^i) \leq Z(W_N^j), \ \forall i \in \mathcal{A}, \ \text{and } j \in \mathcal{A}^c$.
- Polar codes are channel specific designs and they achieve capacity.
- Unlike conventional codes, a polar code designed for one channel may not be optimal for another channel.
    - That is, if the *SNR* changes, the code has to be changed. Hmmm!

References- I

1. **E. Arikan**, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels", *IEEE Trans. on Info. Theory*, Vol.55, No. 7, Jul. 2009.
2. **S. B. Korada**, "Polar Codes for Channel and Source Coding", *Ph.D. Thesis*, EPFL, Jul. 2009.
3. **E. Sasoglu**, "Polar Coding Theorems for Discrete Systems", *Ph.D. Thesis*, EPFL, Nov. 2011.