Chap. 7: Introduction to Polar Codes

Spring 2022

Kyeongcheol Yang

Pohang University of Science and Technology (POSTECH)

Contents

Preliminaries

Polar Codes

- A new class of source and channel codes invented by E. Arikan (2007)
- Provably capacity-achieving in many scenarios
- Complexity of implementation is $N \log_2 N$ in code length N
- Error probability goes down as $O\left(2^{-\sqrt{N}}\right)$ (Arikan's construction).
- Theoretically interesting (practically ?)

Preliminaries

Polarization

– Polar codes are based on the polarization phenomenon.

– Polarization uses multiple independent copies of a given channel and manufactures three classes of channels:

(a) Good Channels: Almost noiseless

(b) Bad Channels: Almost pure noise

(c) Ugly channels: neither too good nor too bad, but not too many.

– Coding problems trivialize, once polarization is achieved.

Preliminaries

Binary-Input Discrete Memoryless Channels (BI-DMC)

$$X \rightarrow \boxed{W : \mathcal{X} \rightarrow \mathcal{Y}} \rightarrow Y$$

- Input alphabet : $\mathcal{X} = \{0,1\}$
- Output alphabet : $\mathcal{Y}$ (arbitrary)
- Transition probabilities : $W(y \mid x)$ for $x \in \mathcal{X},\ y \in \mathcal{Y}$
- N uses of channel W :

$$W^N : \ \mathcal{X}^N \rightarrow \mathcal{Y}^N \ \text{ with } \ W^N\left(y_1^N \mid x_1^N\right) = \prod_{n=1}^{N} W\left(y_i \mid x_i\right)$$

Preliminaries

Mutual Information

– For a uniformly distributed input X, the symmetric capacity $I(W)$ is given by

$$I(W) := \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \frac{1}{2} W(y \mid x) \log \frac{W(y \mid x)}{\frac{1}{2} W(y \mid 0) + \frac{1}{2} W(y \mid 1)}.$$

– It is the highest rate at which reliable communication is possible across W.

– Noiseless channels : $I(W) = 1$

vs.

Purely noisy channels : $I(W) = 0$

Preliminaries

Bhattacharyya Parameter

$$Z(W) := \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$$

- It is an upper bound on the probability of maximum-likelihood (ML) decision error.

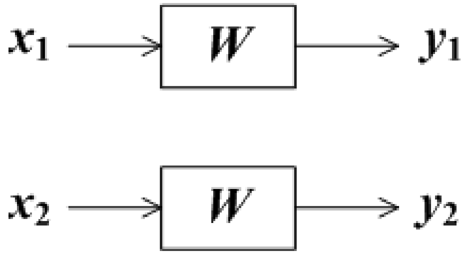
- $I(W) \approx 1$ iff $Z(W) \approx 0$, and $I(W) \approx 0$ iff $Z(W) \approx 1$.

Contents

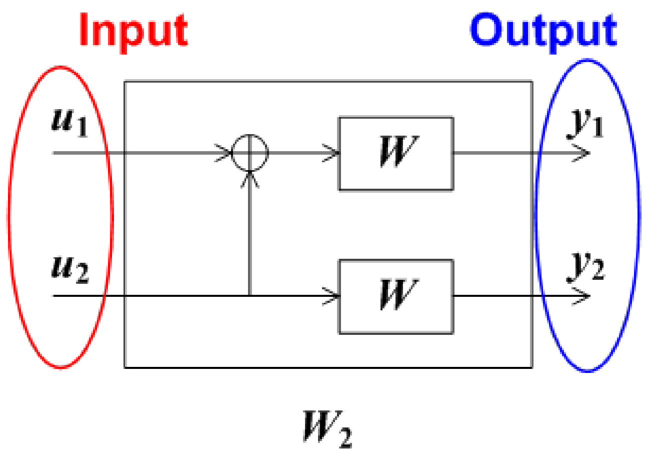
Channel Polarization

Channel Combining

– Given two copies of a binary input channel $W : \mathcal{X} = \{0,1\} \to \mathcal{Y}$,



combine two channels and obtain the channel $W_2 : \mathcal{X}^2 \to \mathcal{Y}^2$ as follows:

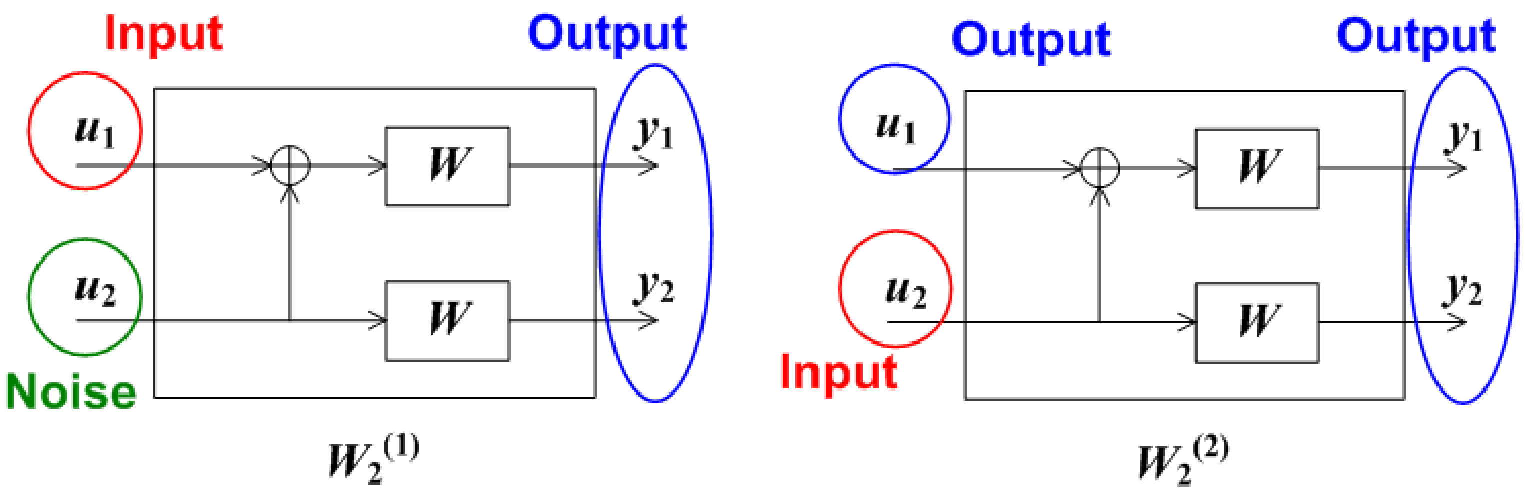


Channel Polarization

Channel Splitting

– The next step is to split W_2 into a set of two binary input channels

$$W_2^{(1)} : \mathcal{X} \to \mathcal{Y}^2 \quad \text{and} \quad W_2^{(2)} : \mathcal{X} \to \mathcal{Y}^2 \times \mathcal{X}.$$



Channel Polarization

Polarized Channels

– Transition probability of $W_2^{(1)}$ and $W_2^{(2)}$

$$W_2^{(1)}(y_1, y_2 \mid u_1) = \sum_{u_2} \frac{1}{2} W(y_1 \mid u_1 \oplus u_2)\, W(y_2 \mid u_2),$$

$$W_2^{(2)}(y_1, y_2, u_1 \mid u_2) = W(y_1 \mid u_1 \oplus u_2)\, W(y_2 \mid u_2).$$

– Channel transformation manufactures two classes of channels

$W_2^{(1)}$: Bad Channel

$W_2^{(2)}$: Good Channel

Change of Channel Reliability

– For a set of binary-input channels,

$$Z(W_2^{(1)}) \leq 2Z(W) - Z(W)^2, \qquad (1)$$

$$Z(W_2^{(2)}) = Z(W)^2$$

where the equality in (1) holds iff W is a binary erasure channel (BEC).

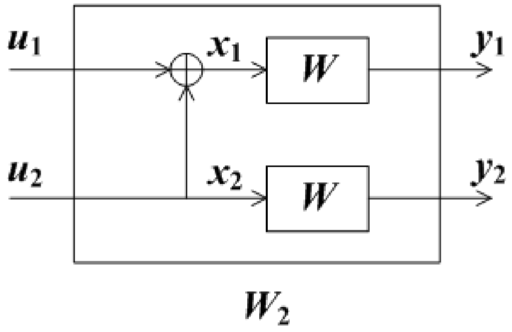
– For a set of binary-input channels,

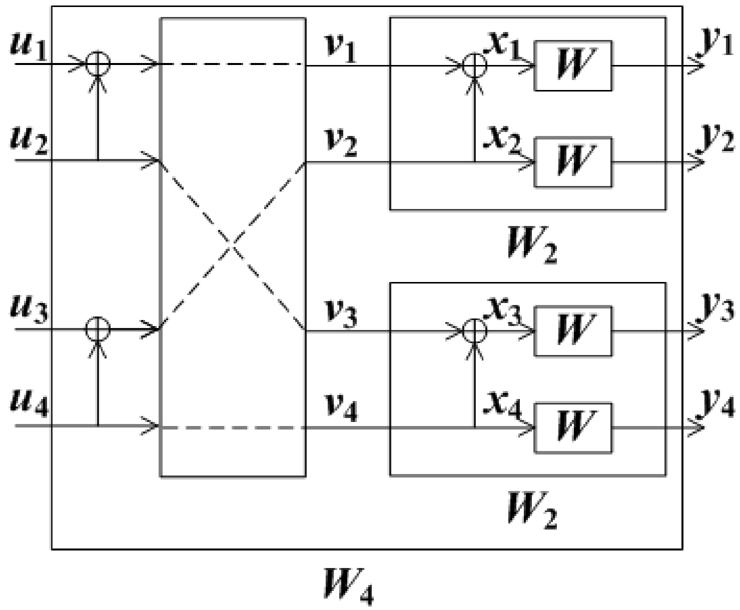
$$I(W_2^{(1)}) + I(W_2^{(2)}) = 2I(W),$$

$$I(W_2^{(1)}) \leq I(W) \leq I(W_2^{(2)})$$

with equality iff $I(W)$ equals 0 or 1.

Recursive Construction (1)





$$W_2^{(1)} : \mathcal{X} \to \mathcal{Y}^2$$

$$W_2^{(2)} : \mathcal{X} \to \mathcal{Y}^2 \times \mathcal{X}$$

$$W_4^{(1)} : \mathcal{X} \to \mathcal{Y}^4$$

$$W_4^{(2)} : \mathcal{X} \to \mathcal{Y}^4 \times \mathcal{X}$$

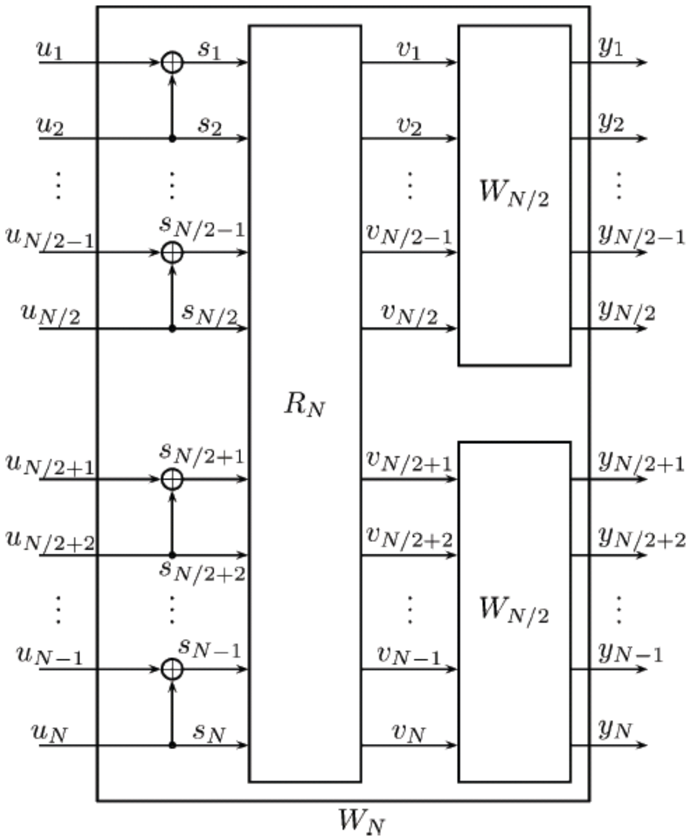
$$W_4^{(3)} : \mathcal{X} \to \mathcal{Y}^4 \times \mathcal{X}^2$$

$$W_4^{(4)} : \mathcal{X} \to \mathcal{Y}^4 \times \mathcal{X}^3$$

Channel Polarization

Recursive Construction (2)

– Recursive construction of W_N from two copies of $W_{N/2}$



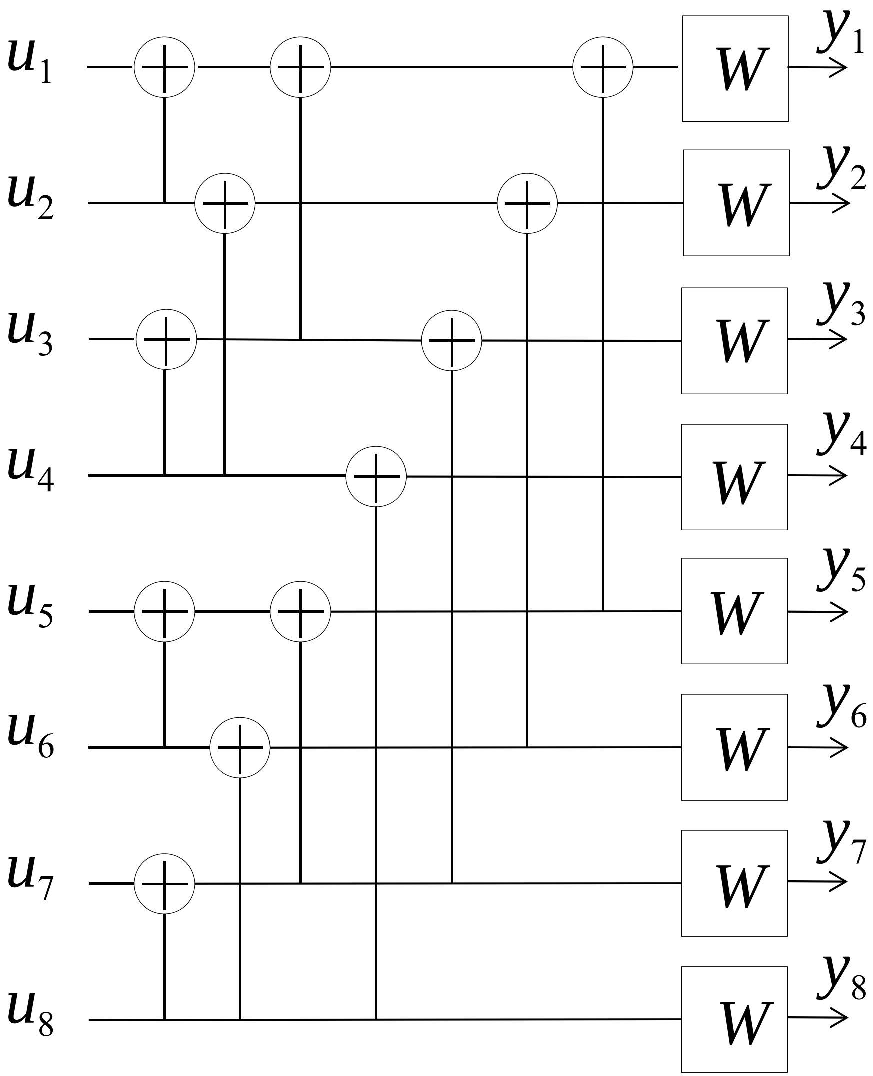
$$\left(W_{N/2}^{(1)}, W_{N/2}^{(2)}, \ldots, W_{N/2}^{(N/2-1)}, W_{N/2}^{(N/2)}\right)$$

$$\Downarrow$$

$$\left(W_N^{(1)}, W_N^{(2)}, \ldots, W_N^{(N-1)}, W_N^{(N)}\right)$$

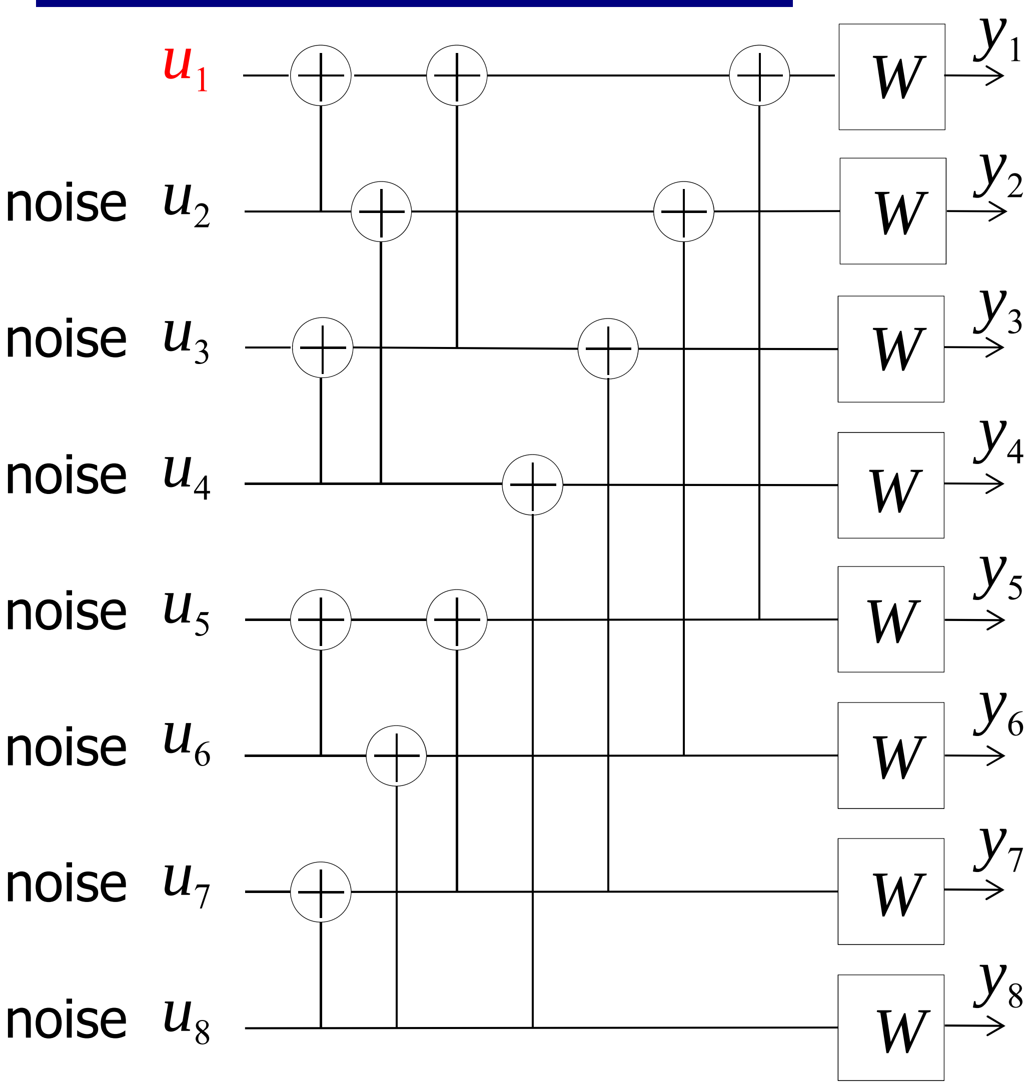
Channel Polarization

Recursive Construction (3)



Channel Polarization

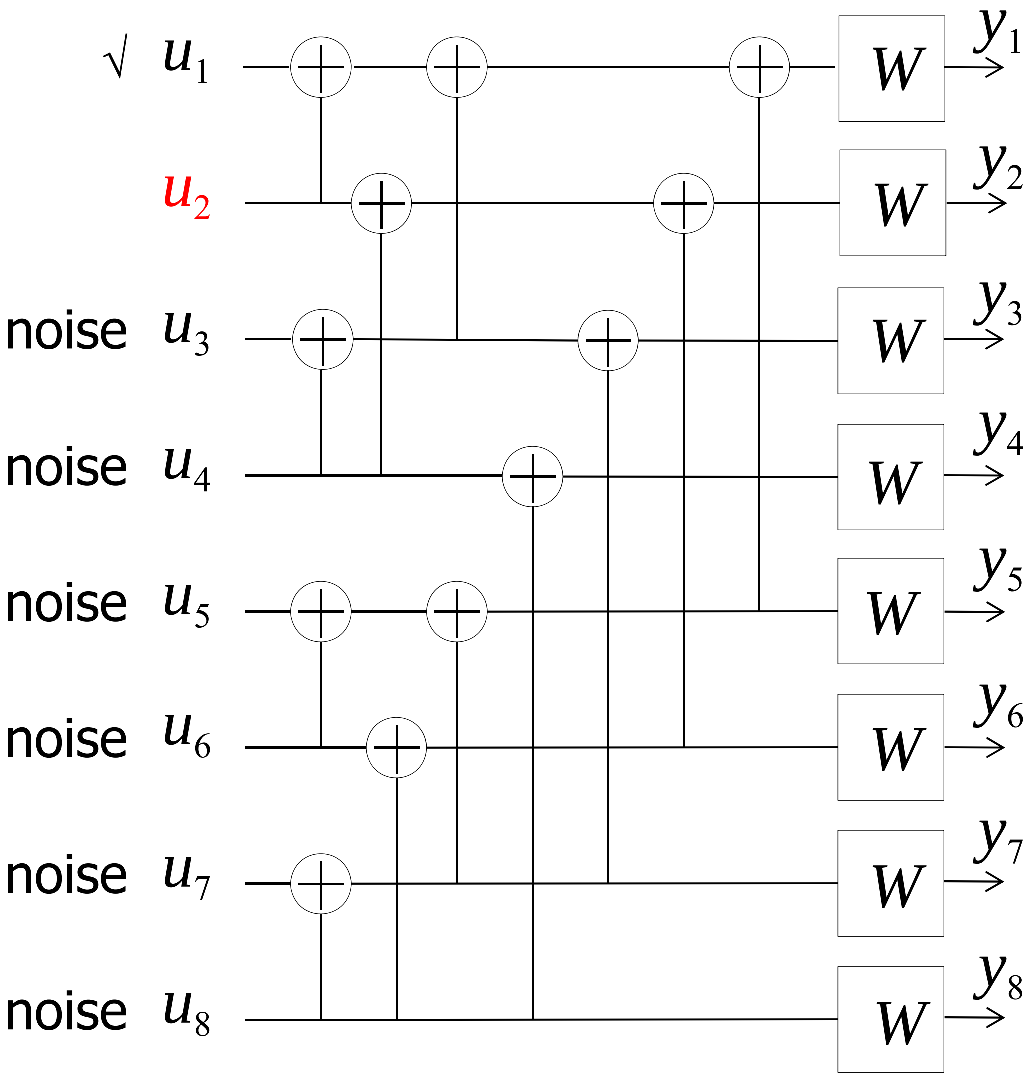
Recursive Construction (3)



Bit channel	Type
$u_1 \rightarrow W_8^{(1)} \rightarrow y_1^8$	BBB

Channel Polarization

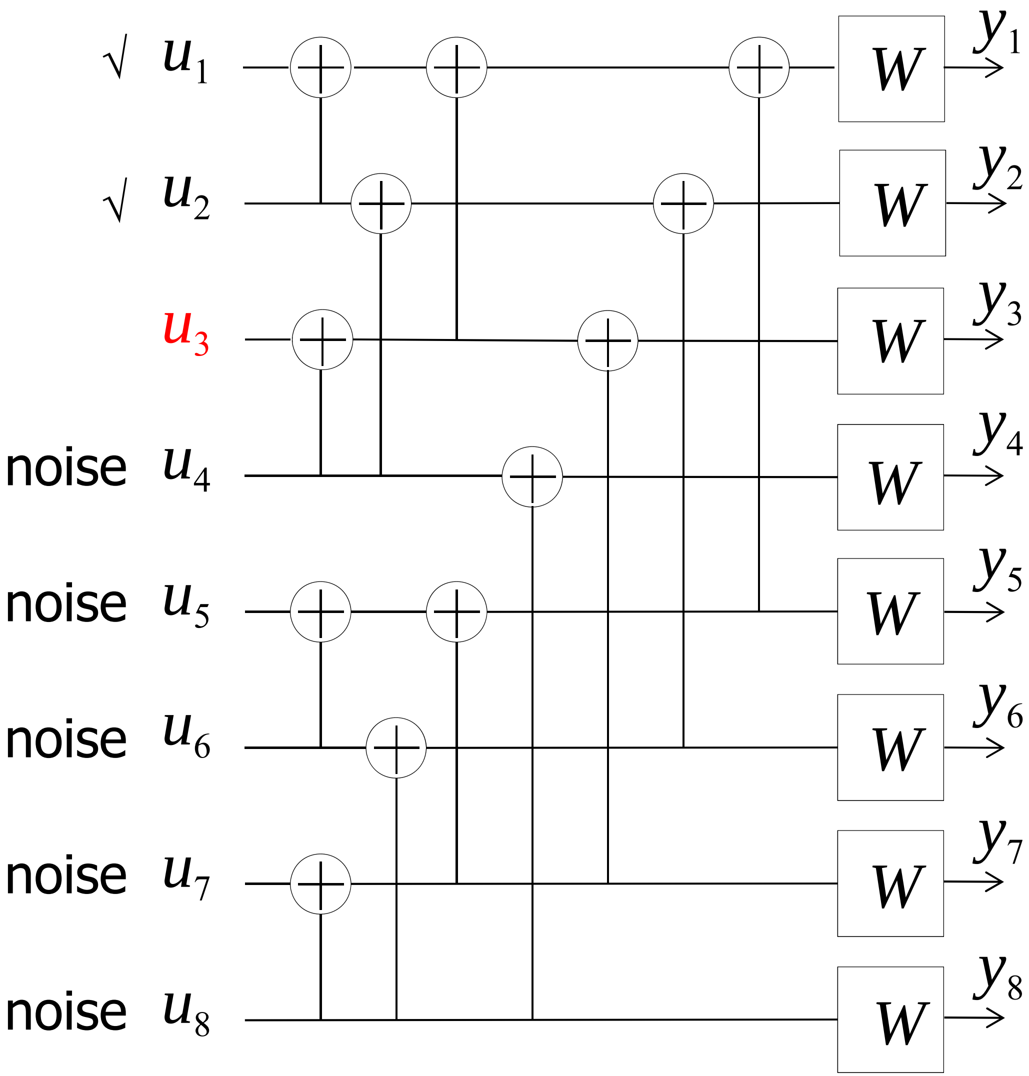
Recursive Construction (3)



Bit channel	Type
$u_2 \longrightarrow \boxed{W_8^{(2)}} \longrightarrow y_1^8, u_1$	BBG

Channel Polarization

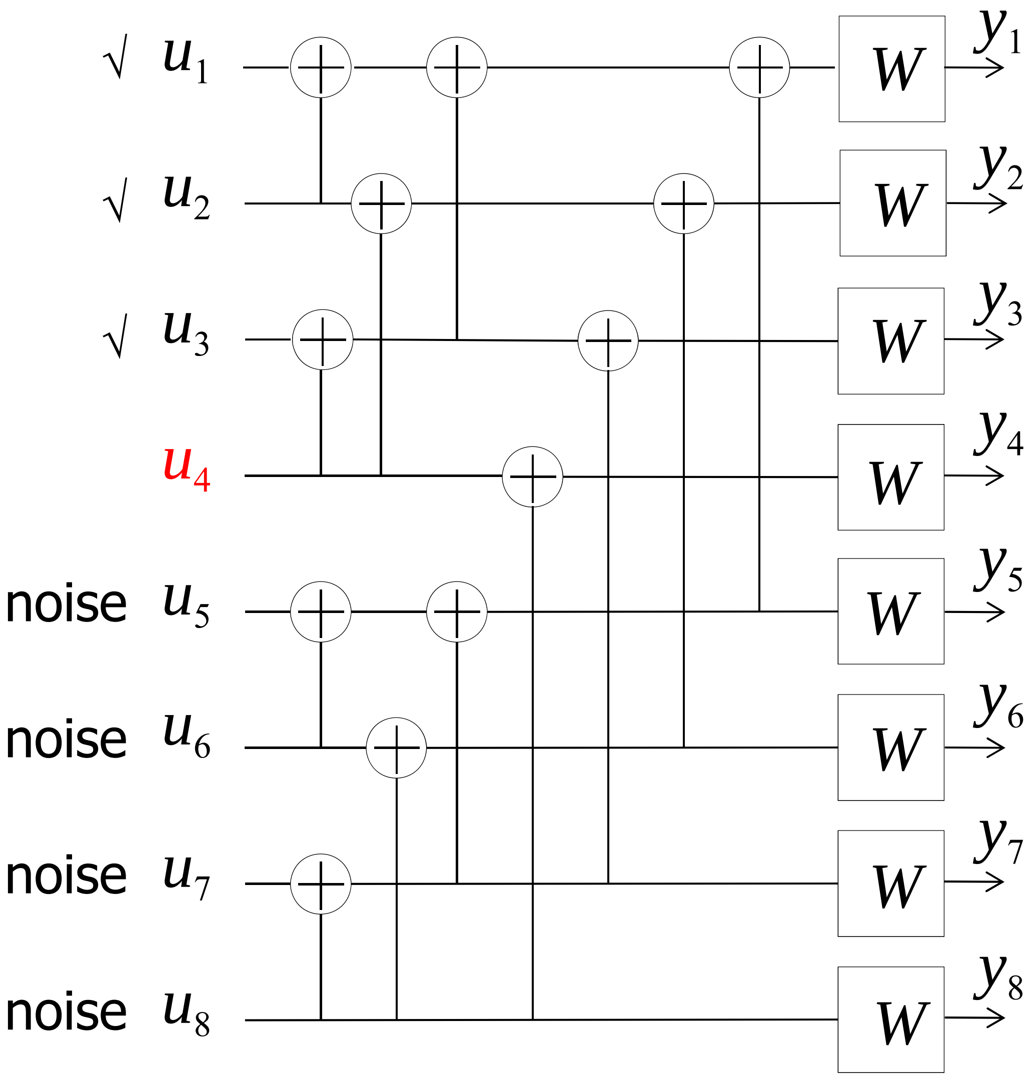
Recursive Construction (3)



Bit channel	Type
$u_3 \rightarrow W_8^{(3)} \rightarrow y_1^8, u_1^2$	BGB

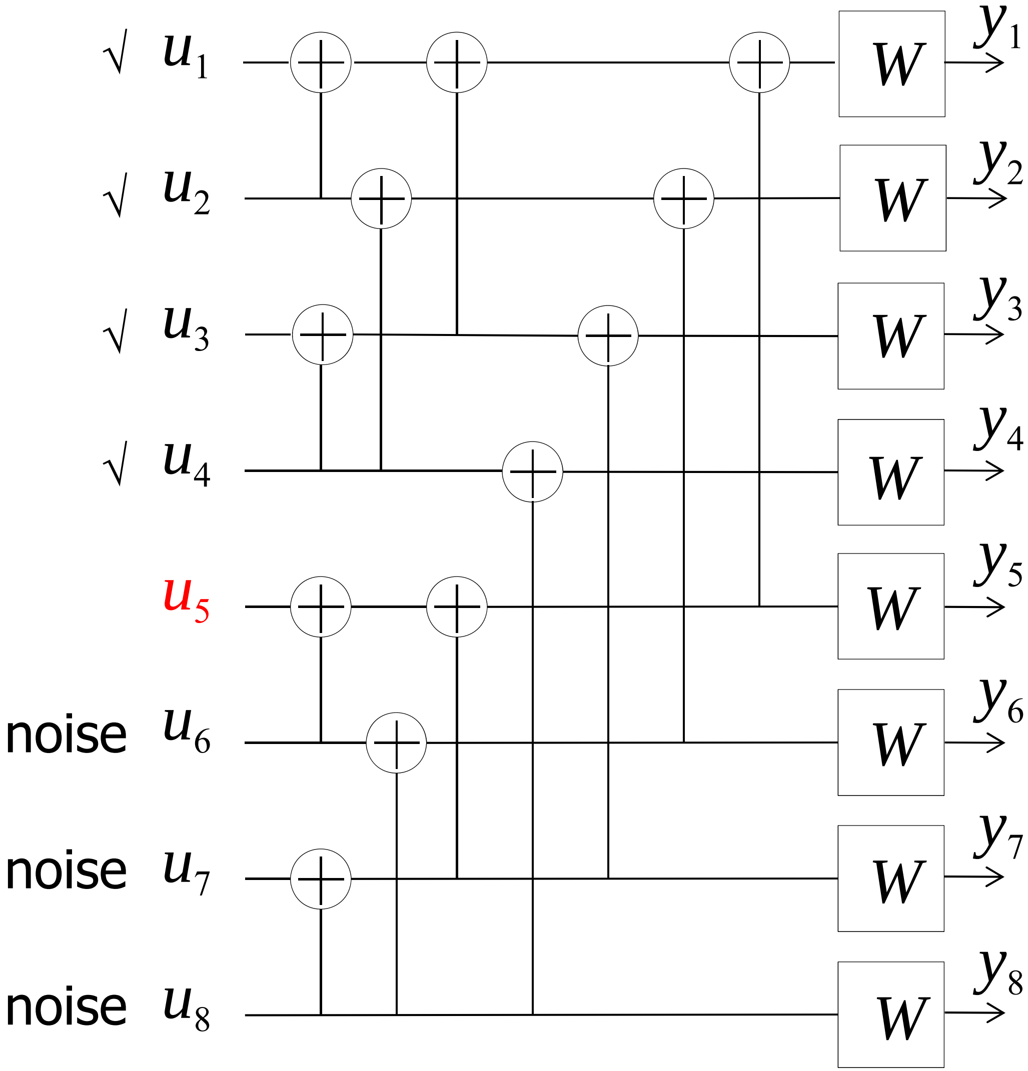
Channel Polarization

Recursive Construction (3)



Bit channel	Type
$u_4 \longrightarrow \boxed{W_8^{(4)}} \rightarrow y_1^8, u_1^3$	BGG

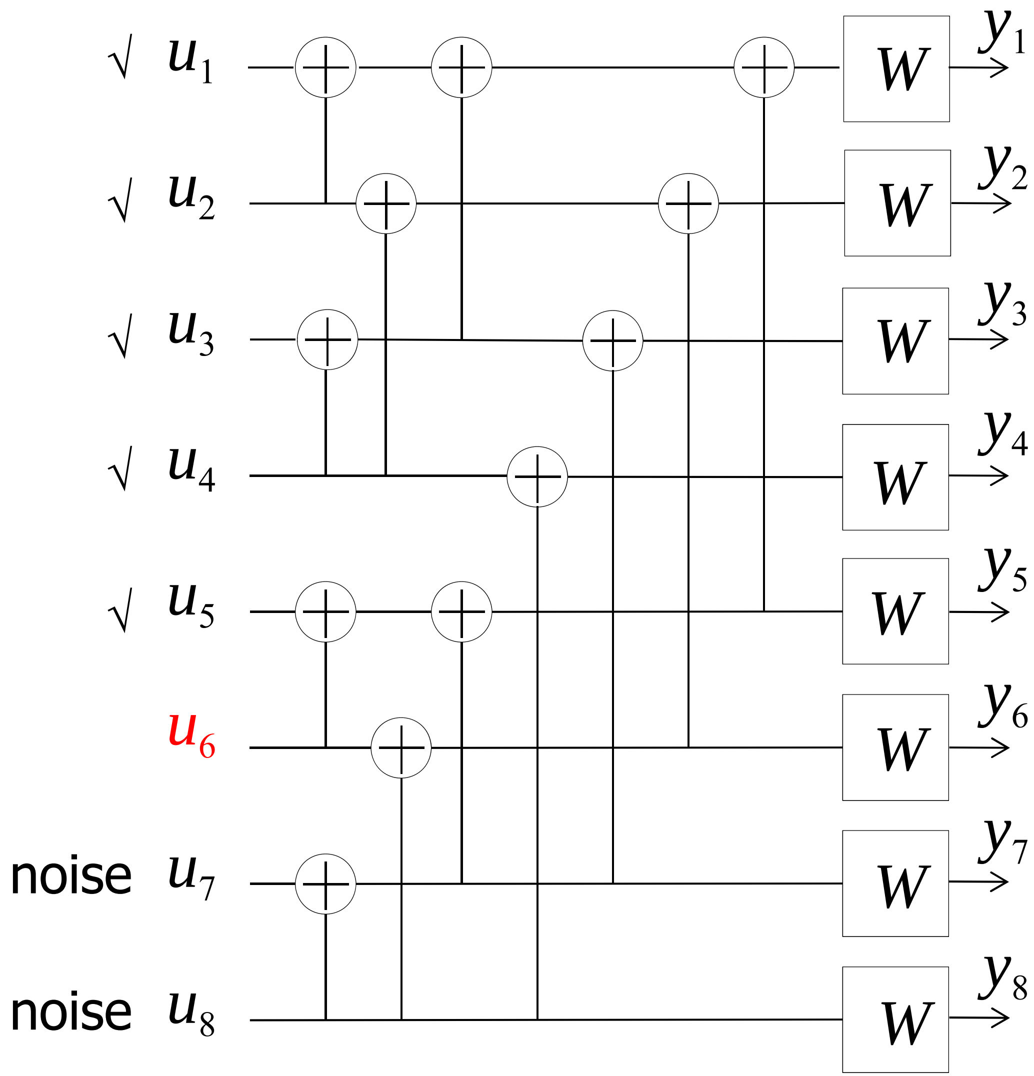
Recursive Construction (3)



Bit channel	Type
$u_5 \rightarrow W_8^{(5)} \rightarrow y_1^8, u_1^4$	GBB

Channel Polarization

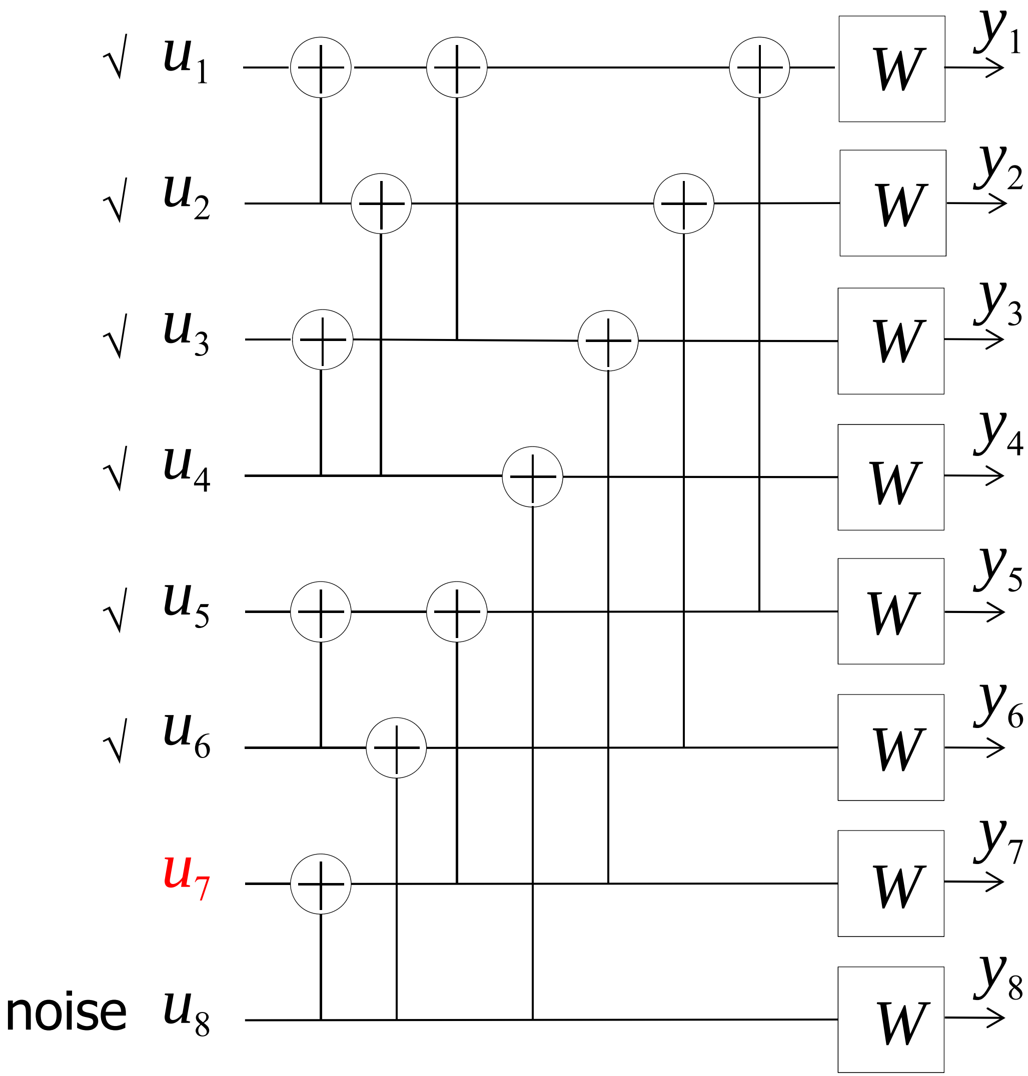
Recursive Construction (3)



Bit channel	Type
$u_6 \text{ — } \boxed{W_8^{(6)}} \to y_1^8, u_1^5$	GBG

Channel Polarization

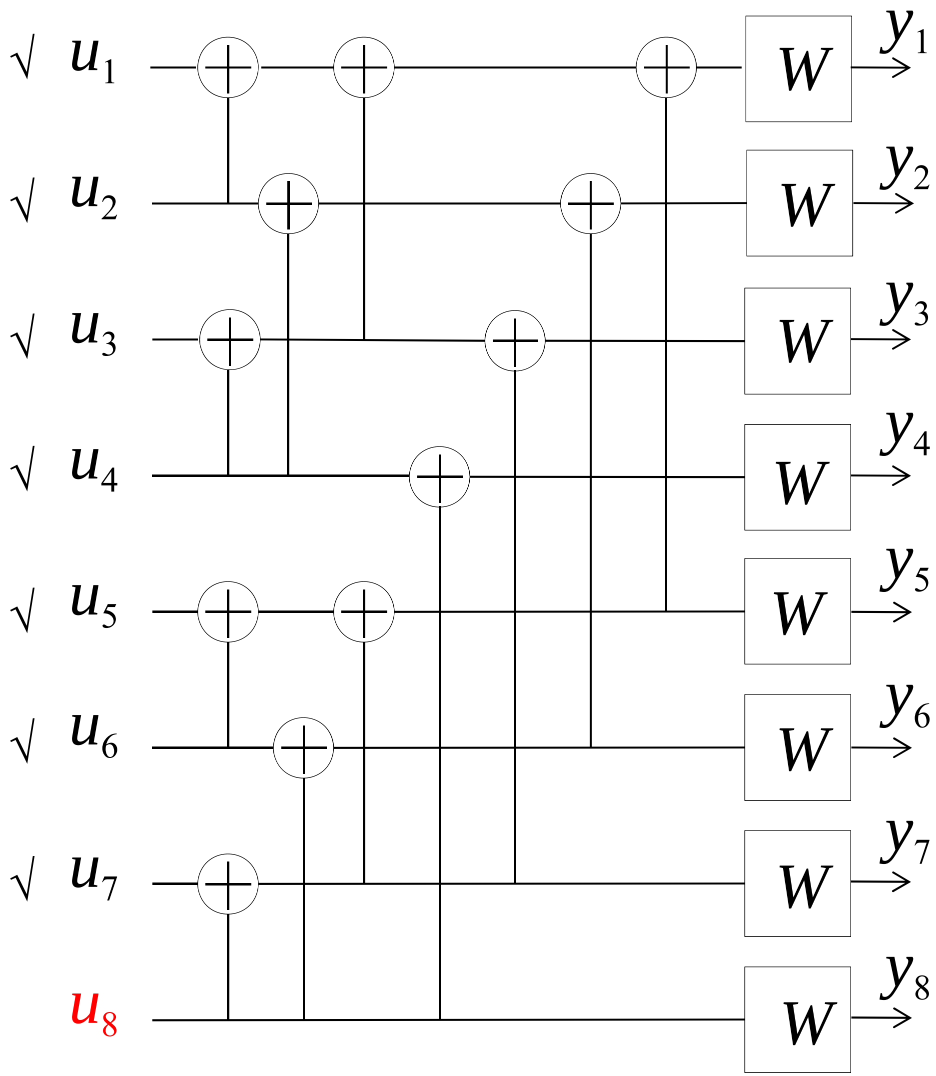
Recursive Construction (3)



Bit channel	Type
$u_7 \longrightarrow \boxed{W_8^{(7)}} \longrightarrow y_1^8, u_1^6$	GGB

Channel Polarization

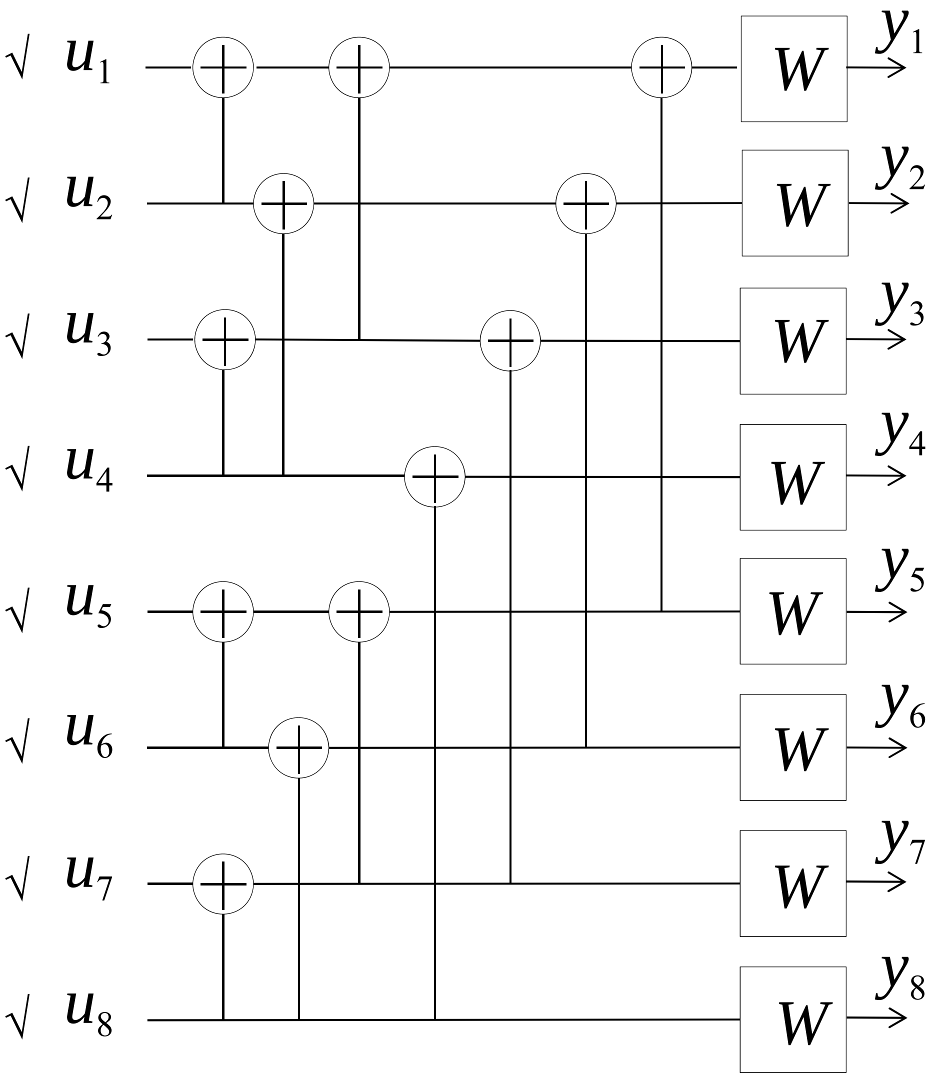
Recursive Construction (3)



Bit channel	Type
u_8 — $W_8^{(8)}$ → y_1^8, u_1^7	GGG

Channel Polarization

Recursive Construction (3)



	Bit channel		Type
u_1	$W_8^{(1)}$	y_1^8	BBB
u_2	$W_8^{(2)}$	y_1^8, u_1	BBG
u_3	$W_8^{(3)}$	y_1^8, u_1^2	BGB
u_4	$W_8^{(4)}$	y_1^8, u_1^3	BGG
u_5	$W_8^{(5)}$	y_1^8, u_1^4	GBB
u_6	$W_8^{(6)}$	y_1^8, u_1^5	GBG
u_7	$W_8^{(7)}$	y_1^8, u_1^6	GGB
u_8	$W_8^{(8)}$	y_1^8, u_1^7	GGG

Arikan's Result

Theorem : For any fixed $\delta \in (0,1)$, as N goes to infinity, the fraction of indices $i \in \{1, 2, ..., N\}$ converge to extremal values in the sense that

$$\frac{\left|\left\{ i : I(W_N^{(i)}) > 1-\delta \right\}\right|}{N} \rightarrow I(W)$$ "Capacity-achieving"

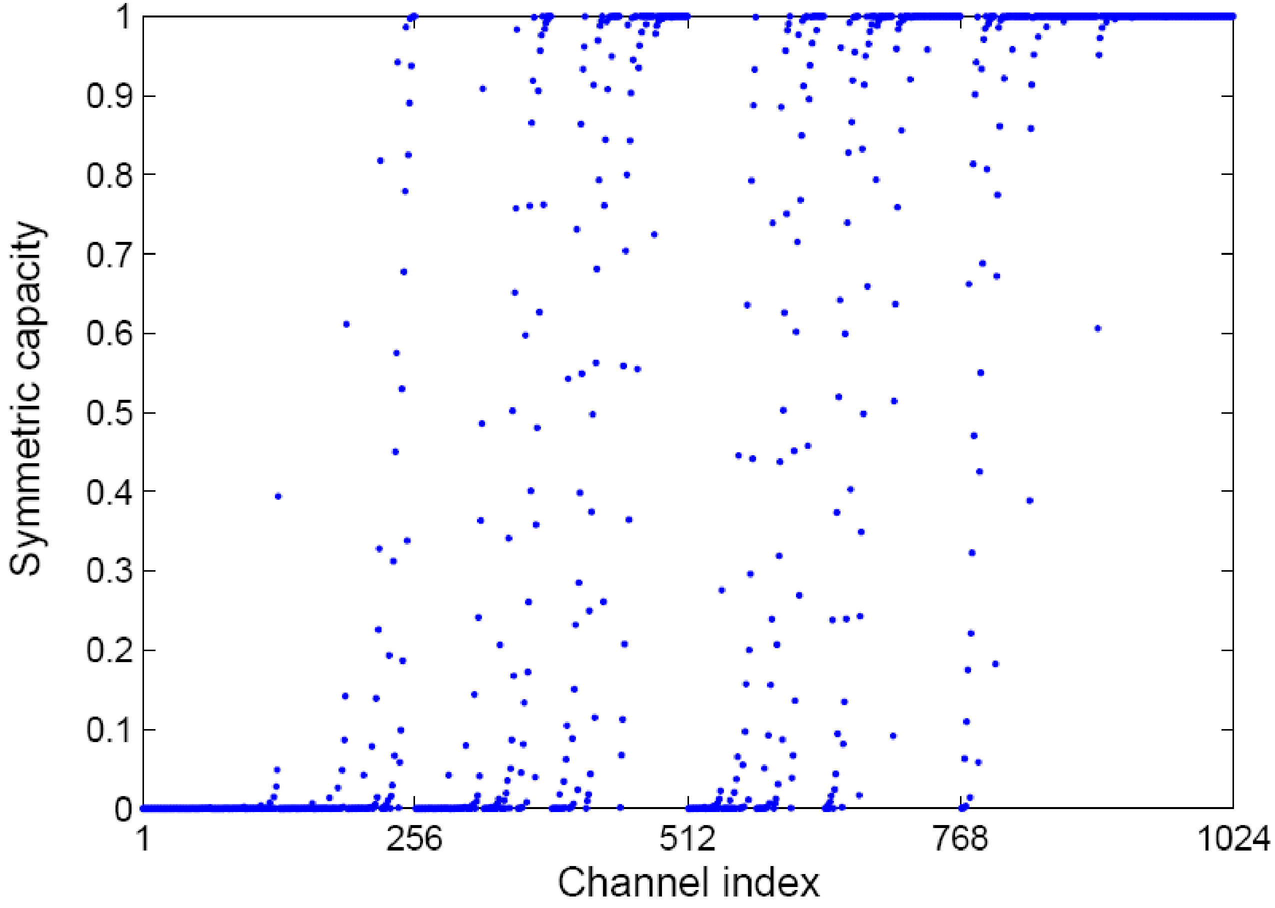
and

$$\frac{\left|\left\{ i : I(W_N^{(i)}) < \delta \right\}\right|}{N} \rightarrow 1 - I(W).$$

Channel Polarization

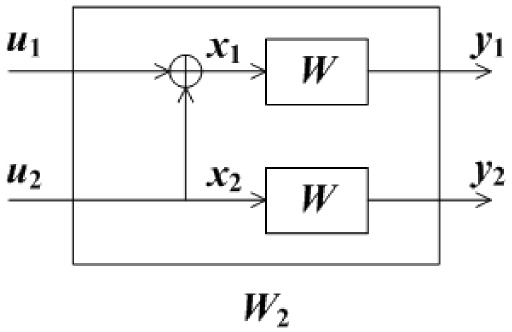
Channel Reliability

– BEC, $\varepsilon = 0.5$, Code length $N = 1024$

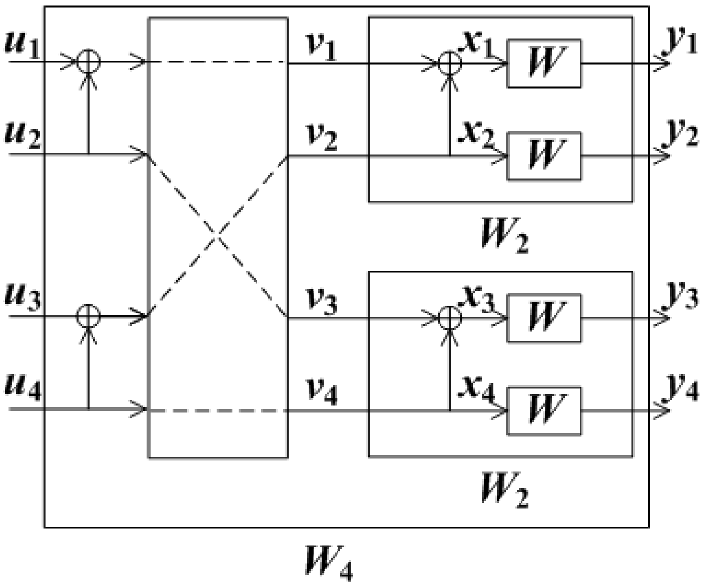


Contents

Polarizing Matrix (1)



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



$$[x_1 \ x_2 \ x_3 \ x_4] = [u_1 \ u_2 \ u_3 \ u_4]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Polarizing Matrix (2)

- Recursive construction :

$$G_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \Rightarrow G_2^{\otimes 2} := G_2 \otimes G_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\Rightarrow G_4 = B_4 G_2^{\otimes 2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

where $\otimes$ denotes the Kronecker product, and B_4 is the row permutation matrix.

Selection of Frozen Bits

- Pick $K = NR$ good indices i such that

$$Z(W_N^{(i)}) \text{ is low}$$

or equivalently, $I(W_N^{(i)})$ is high.

where R is a code rate.

- Information bits are assigned to good indices.

  The bits corresponding to the remaining indices are set to be frozen.

Frozen Bits & Code Design

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

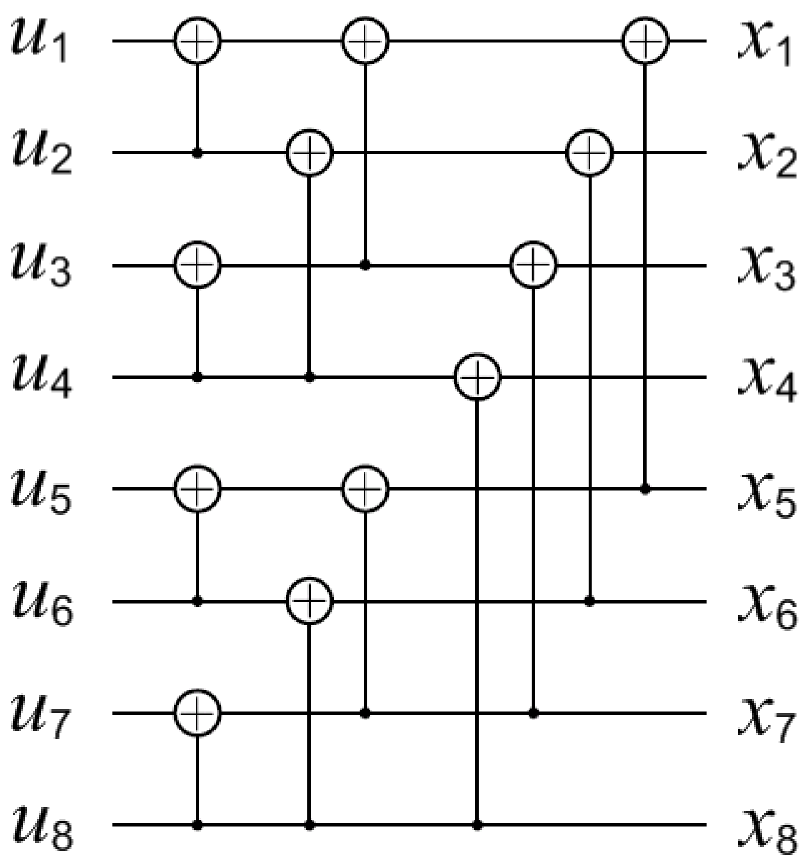
Frozen bits
$u_1 = u_2 = 0$

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} = \begin{bmatrix} u_3 & u_4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Generator matrix of polar code

Encoding of Polar Codes

– Encoding complexity is $O(N \log_2 N)$.



Successive Cancellation (SC) Decoding (1)

– Decoding complexity is $O(N \log_2 N)$.

– Polar codes with SC decoding achieve the channel capacity.

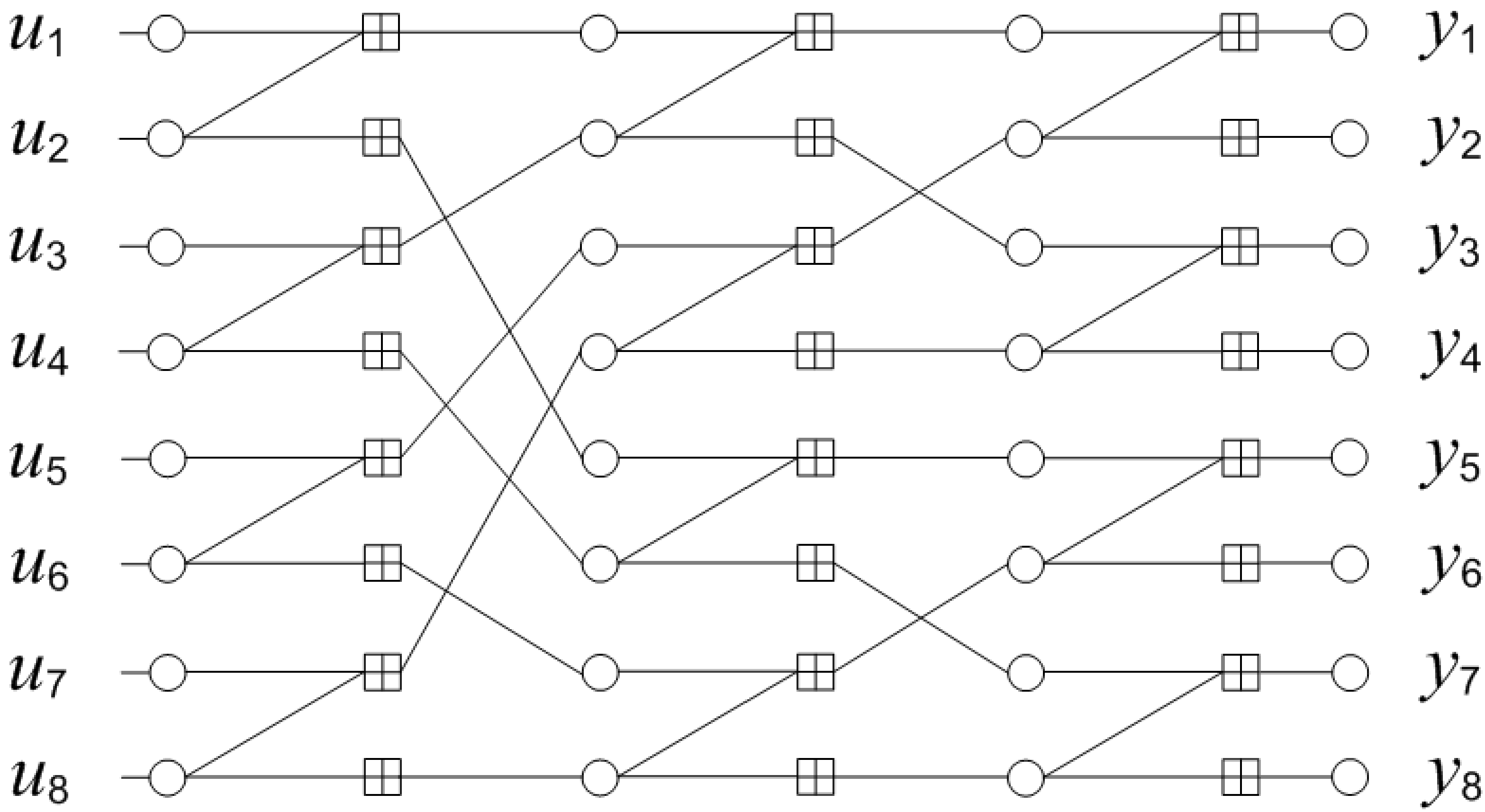
– Characteristics of SC decoding :

(a) Before decoding the i th information bit, the j th bit information has been already determined for all $j < i$.

(b) For all $j > i$, the j th information bit is considered to be unknown. (it is regarded as a noise)

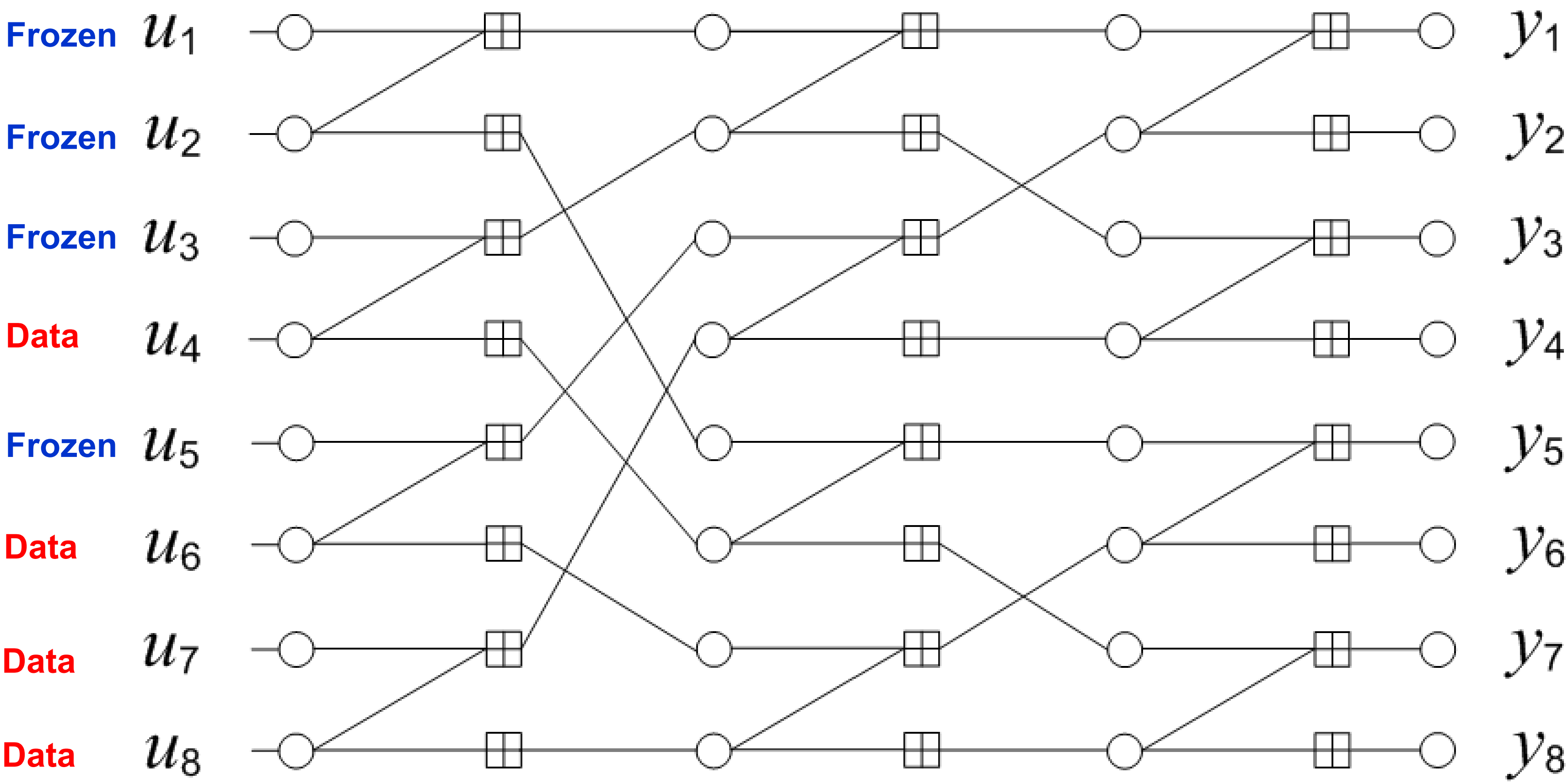
Successive Cancellation (SC) Decoding (2)

– **Example** : Graph for decoding the 4th bit when $N=8$ and $K=4$



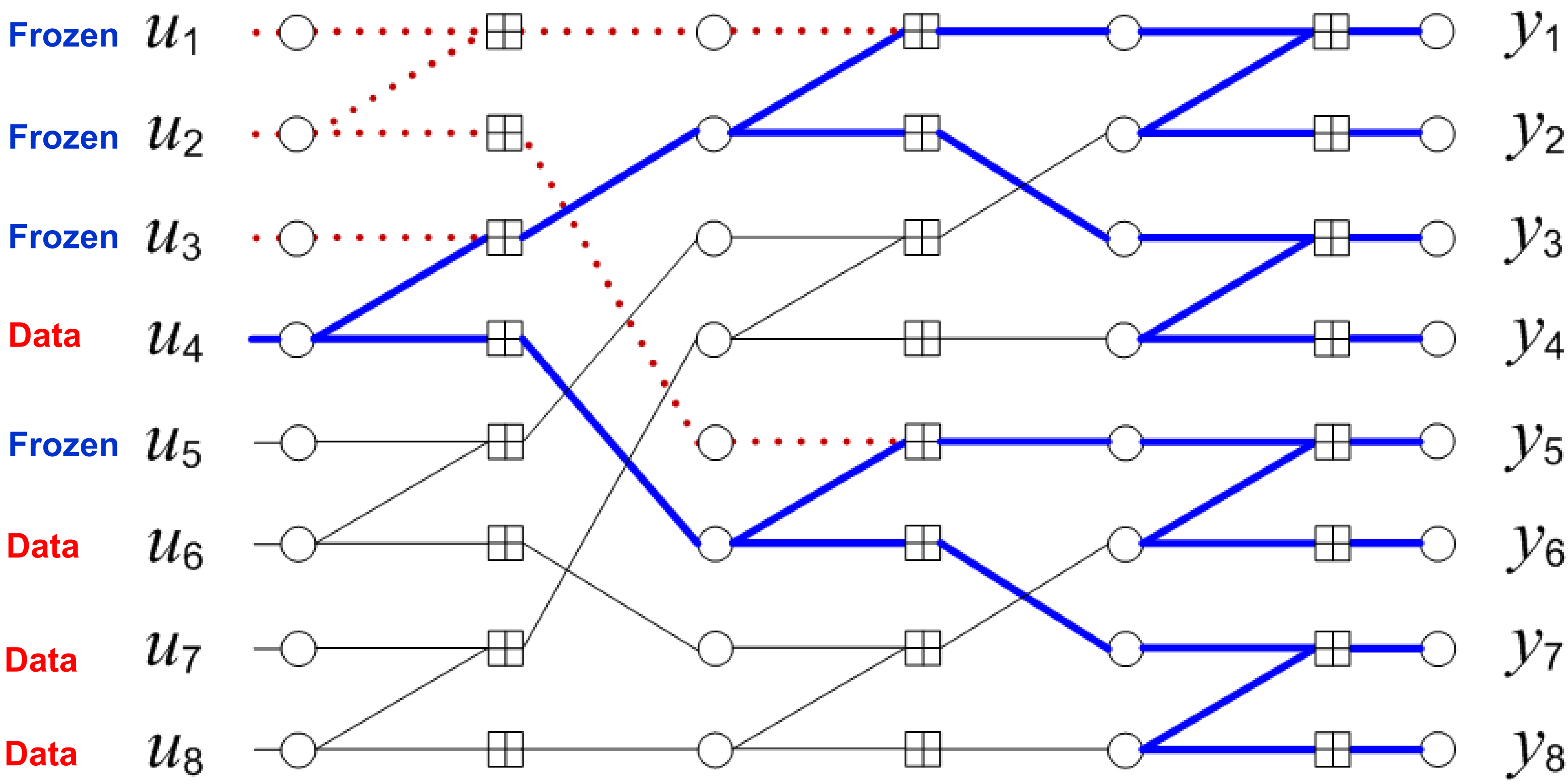
Successive Cancellation (SC) Decoding (2)

– **Example** : Graph for decoding the 4th bit when $N=8$ and $K=4$



Successive Cancellation (SC) Decoding (2)

– **Example** : Graph for decoding the 4th bit when $N=8$ and $K=4$



Generalized Channel Polarization

– Let G be an invertible $l \times l$ matrix.

– A necessary and sufficient condition for G to be polarizing:
The matrix G is not upper triangular.

Design of Polar Codes of Different Lengths

– Let G_i be a polarizing matrix of size $l_i \times l_i$.

– The matrix $G = \bigotimes_{i=1}^{n} G_i^{\otimes m_i}$ is a generator matrix for a polar code of length $l = l_1^{m_1} l_2^{m_2} \cdots l_n^{m_n}$.

Contents

Asymptotic Performance of Polar Codes

Rate of Polarization

For any B-DMC W with $0 < I(W) < 1$, we say that an $l \times l$ matrix G has rate of polarization (or exponent) $E(G)$ if

i) For any fixed $\beta < E(G)$,

$$\liminf_{n\to\infty} \Pr\left[Z_n \le 2^{-l^{n\beta}}\right] = I(W).$$

ii) For any fixed $\beta > E(G)$,

$$\liminf_{n\to\infty} \Pr\left[Z_n > 2^{-l^{n\beta}}\right] = 1.$$

Behavior of Block Error Probability under SC Decoding

- $P_e(G,n)$: Block error probability of a polar code constructed by $G^{\otimes n}$ under SC decoding
- For sufficiently large n,

$$P_e(G,n) \le 2^{-l^{nE(G)}}.$$

Asymptotic Performance of Polar Codes

Partial Distance of a Polarizing Matrix

– Given $G = \left[g_1^T, \ldots, g_l^T \right]^T$, the partial distances of G are defined as

$$D_n = d_H \left(g_n, \langle g_{n+1}, \ldots, g_l \rangle \right), \quad n = 1, \ldots, l-1,$$
$$D_l = d_H \left(g_l, 0 \right),$$

where $\langle g_{n+1}, \ldots, g_l \rangle$ denotes the linear code generated by the vectors $g_{n+1}, \ldots, g_l$ and d_H is the Hamming distance.

– Example)

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \Rightarrow \quad D_1 = 1, \; D_2 = 1, \; D_3 = 3$$

Asymptotic Performance of Polar Codes

Exponent from Partial Distances

– For any B-DMC and any $l \times l$ polarizing matrix G with partial distances $\{D_i\}_{i=1}^{l}$, the rate of polarization $E(G)$ is given by

$$E(G) = \frac{1}{l} \sum_{i=1}^{l} \log_l D_i.$$

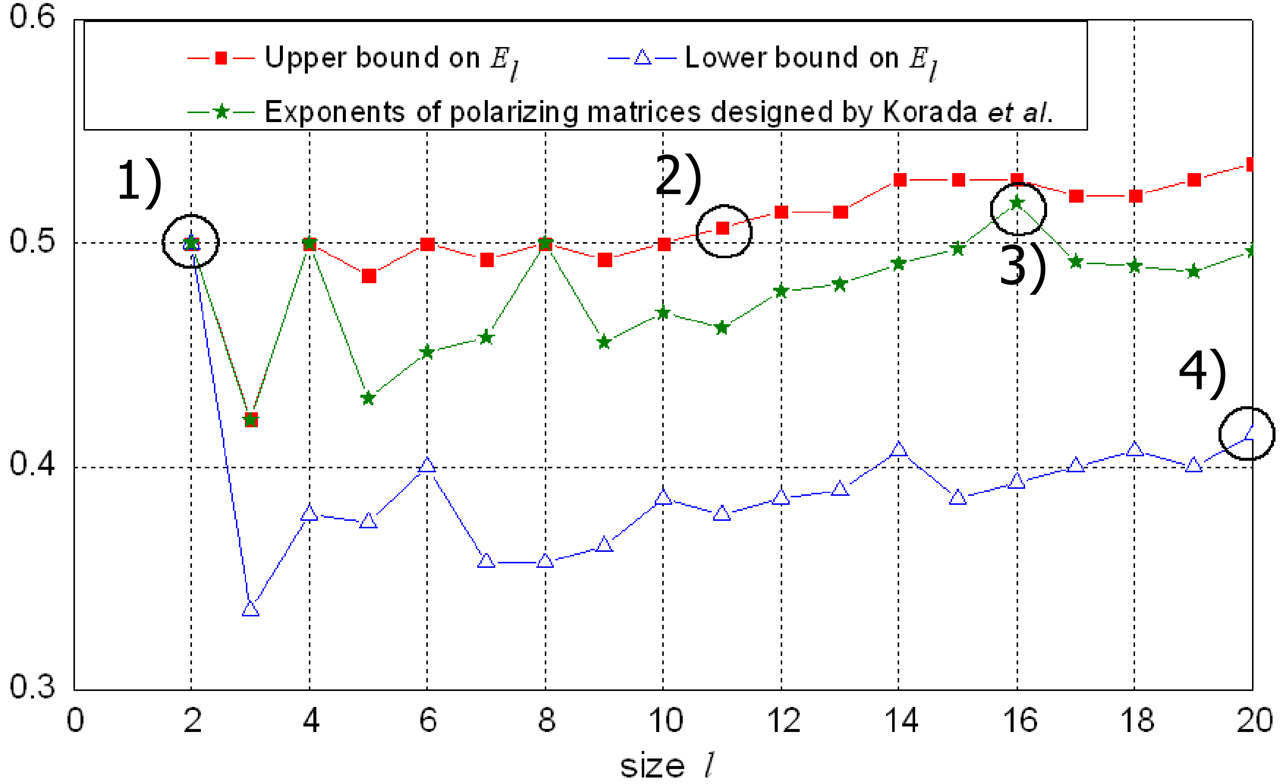
– Example) $F = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \Rightarrow E(F) = \frac{1}{3}\left(\log_3 1 + \log_3 1 + \log_3 3\right) = \frac{1}{3}$

Maximum Exponent for the $l \times l$ Polarizing Matrices

$$E_l \triangleq \max_{G \in \{0,1\}^{l \times l}} E(G)$$

Asymptotic Performance of Polar Codes

Bounds on the Exponent



1) Exponent of Arikan's construction = $1/2$

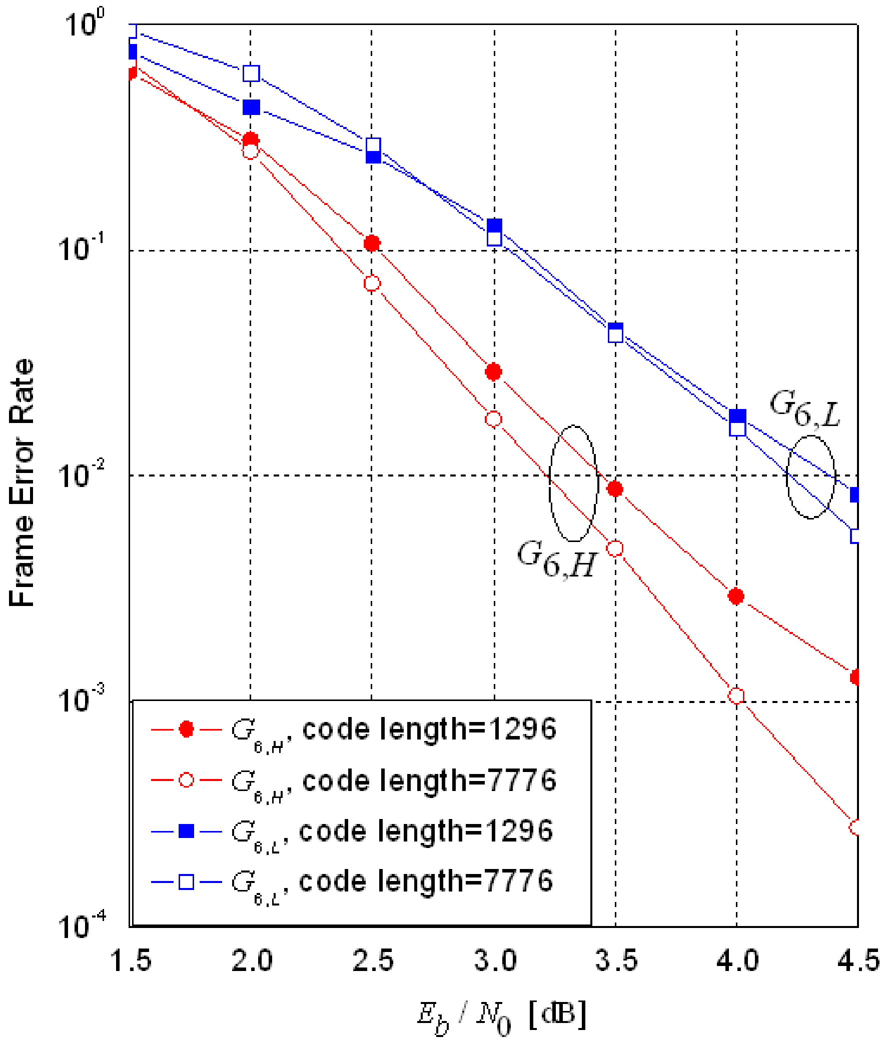
2) No matrix with exponent $> 1/2$ can be found for $l < 11$.

3) Korada *et al.* designed polarizing matrices with exponent $> 1/2$.

4) Exponent 1 is achievable : $\lim_{l \to \infty} E_l = 1$.

Asymptotic Performance of Polar Codes

Numerical Example : Relation between FER and Exponent



– Four rate-1/2 polar codes are constructed from polarizing matrices $G_{6,H}$ and $G_{6,L}$ such that

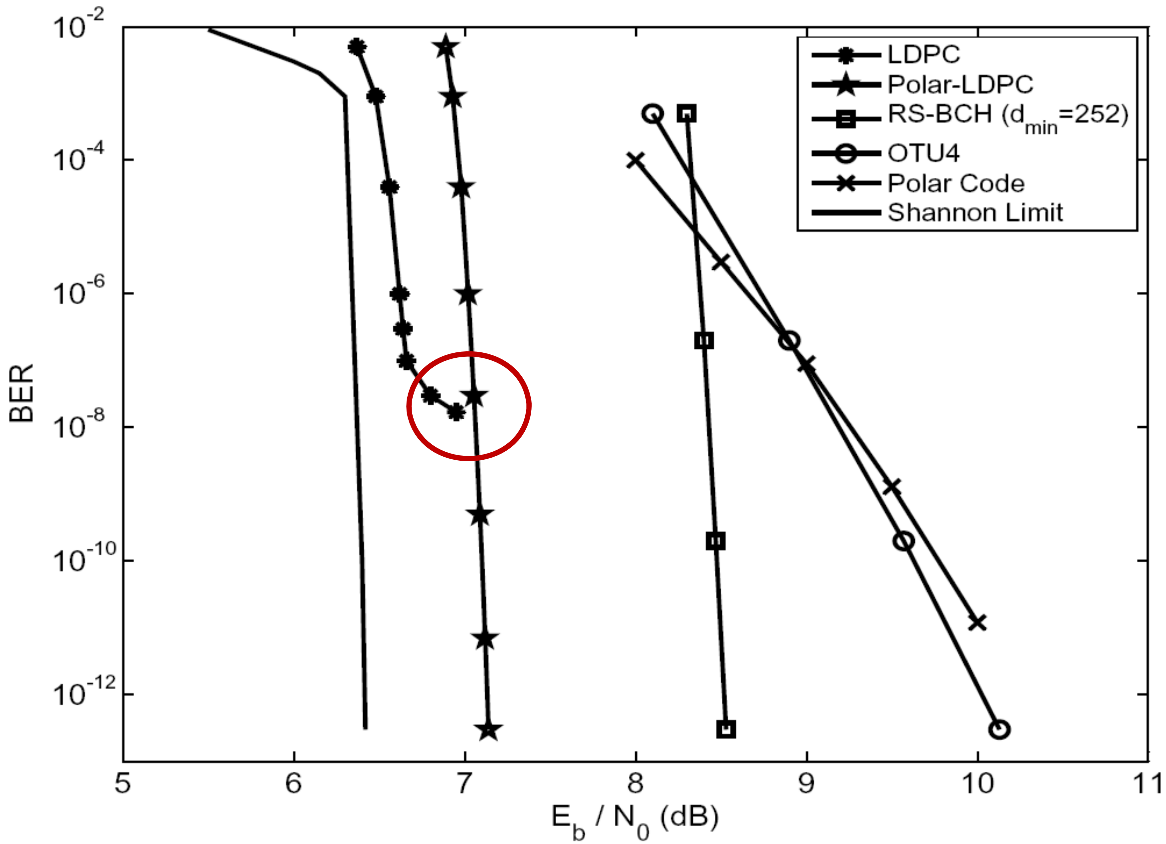
$$E(G_{6,H}) > E(G_{6,L}).$$

– The polar code with higher rate of polarization has better FER performance.

Contents

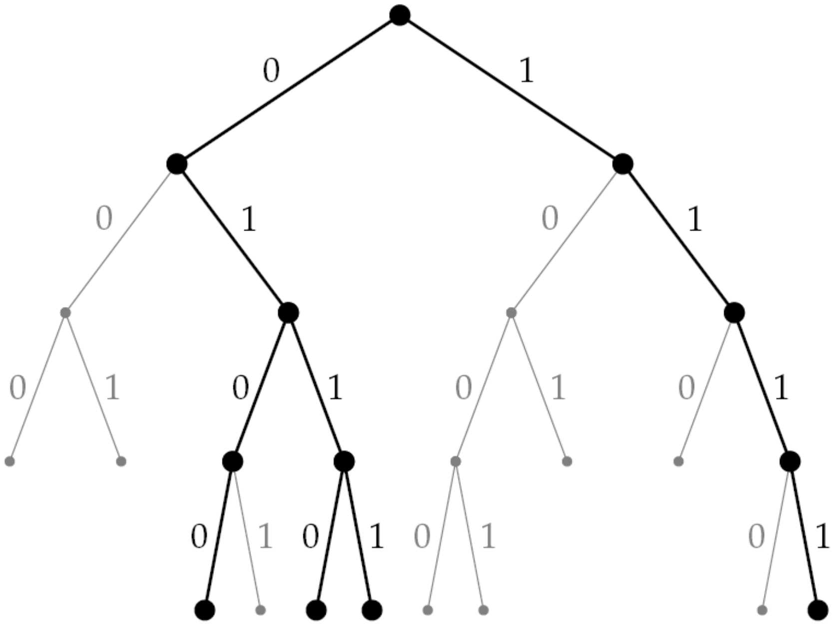
Concatenation of LDPC Codes and Polar Codes [3]

Information source → Polar Encoder → LDPC Encoder → Channel

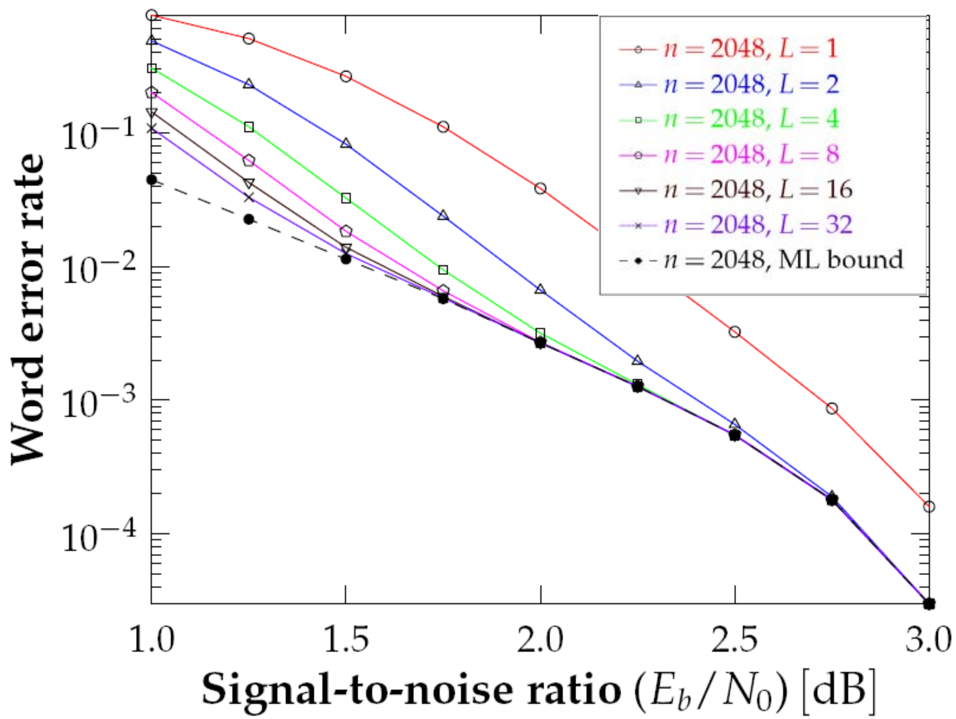


- Effective code rate $= 0.93$
- The polar-LDPC scheme shows no sign of error floor down to BER of 10^{-13}.
- Code length $= 2^{16}$ bits
- Eslami and Pishro-Nik (ISIT 2011)

SC List Decoding Algorithm (1) [4]

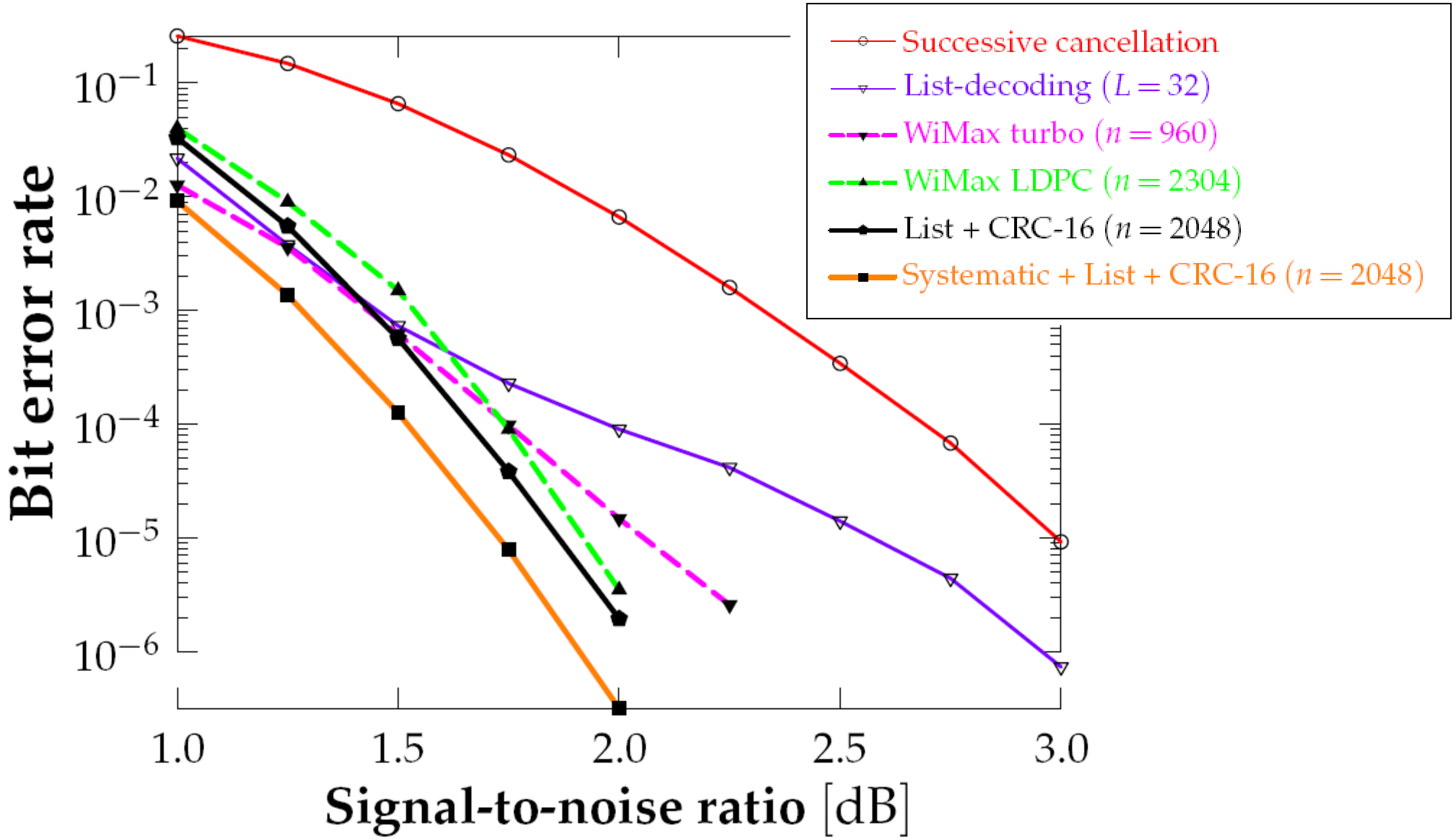


[Decoding path, list size L=4]



– The SC list decoding shows performance improvement as compared with SC decoding. (Tal and A. Vardy, ISIT 2011)

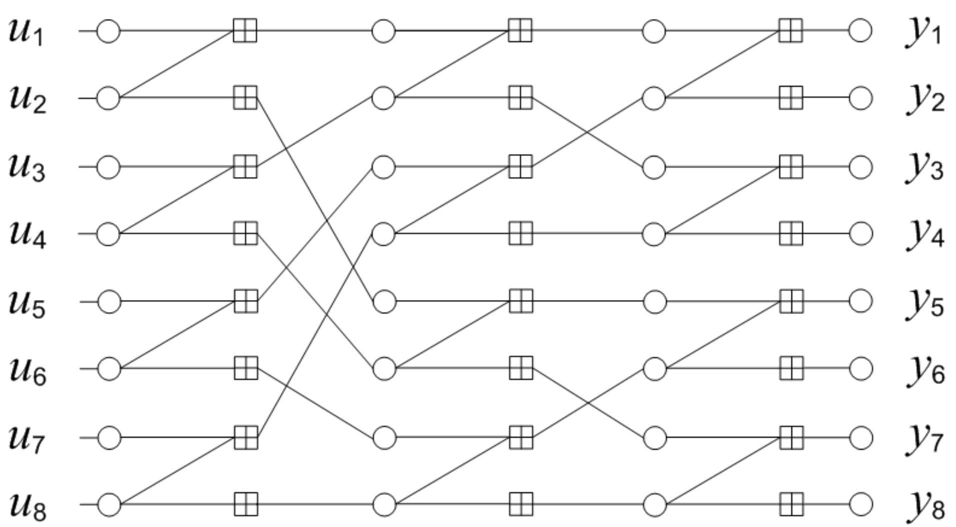
SC List Decoding Algorithm (2) [4]



– Polar codes (+CRC) under list decoding are competitive with the best LDPC codes (Code rate : 0.5)

Exact Selection of Frozen Bits [5], [6]

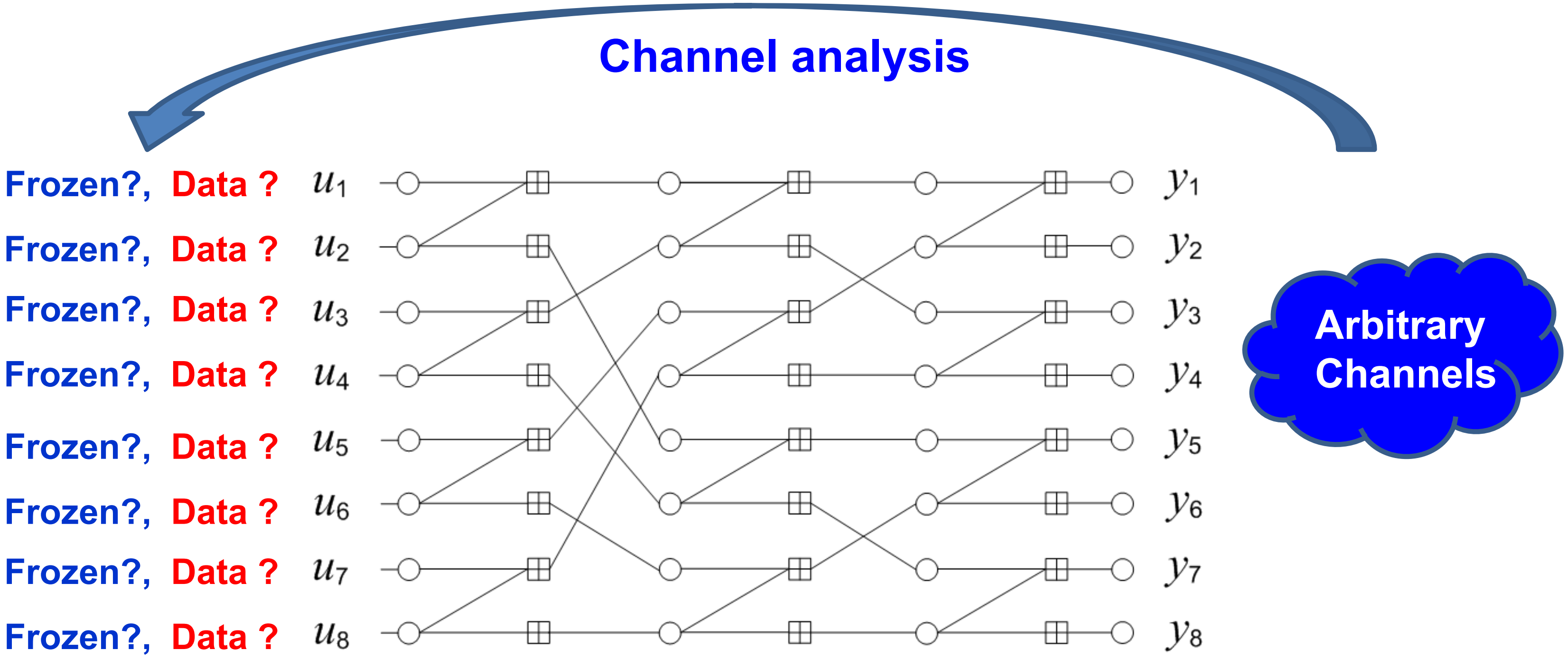
- The estimation of reliability of polarized channels can be performed by ‘Density Evolution’ or ‘Channel Upgrading & Degrading’.
- The performance of a polar code depends on the selection of frozen bits



Exact Selection of Frozen Bits [5], [6]

– The estimation of reliability of polarized channels can be performed by 'Density Evolution' or 'Channel Upgrading & Degrading'.

– The performance of a polar code depends on the selection of frozen bits



Refined Rate of Polarization [7]

– Rate-independent bound on block error probability:

For sufficient large n,

$$P_e(G,n) \le 2^{-l^{nE(G)}}.$$

– Rate-dependent bound on block error probability:

Let code rate $R \in \left(0, I(W)\right)$ be fixed. For sufficient large n,

$$P_e(G,n) \le 2^{-l^{(n+t\sqrt{n})E(G)}}$$

for any t satisfying $t < Q^{-1}\left(R / I(W)\right)$.

Other Issues

- Puncturing patterns of polar codes
- Construction of systematic polar codes
- Design of nonbinary polar codes
- Polar codes for high-order modulations
- Efficient decoding of polar codes

Contents

Summary

- Polar codes are capacity-achieving.
- Encoding complexity is $O(N \log N)$.
- Decoding complexity is $O(N \log N)$.
- Probability of decoding error decays roughly like $2^{-\sqrt{N}}$.
- There are many interesting research problems.

Reference

[1] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, July 2009.

[2] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: characterization of exponent, bounds, and construction," *IEEE Trans. Inf. Theory*, Dec. 2010.

[3] A. Eslami and H. Pishro-Nik, "A practical approach to polar codes," in *Proc. IEEE ISIT,* 2011.

[4] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE ISIT,* 2011.

[5] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, July 2009.

[6] I. Tal and A. Vardy, "How to construct polar codes," arXiv:1105.6164v1, May 2011.

[7] S. H. Hassani and R. Urbanke, "On the scaling of polar codes: I. the behavior of polarized channels," in *Proc. IEEE ISIT,* 2010.