Polar Coding

Part 1 - Background

Erdal Arıkan

Electrical-Electronics Engineering Department,
Bilkent University, Ankara, Turkey

Algorithmic Coding Theory Workshop
June 13 - 17, 2016
ICERM, Providence, RI

Outline

Sequential decoding and the cutoff rate

Guessing and cutoff rate

Boosting the cutoff rate

Pinsker's scheme

Massey's scheme

Polar coding

Sequential decoding and the cutoff rate

Guessing and cutoff rate

Boosting the cutoff rate

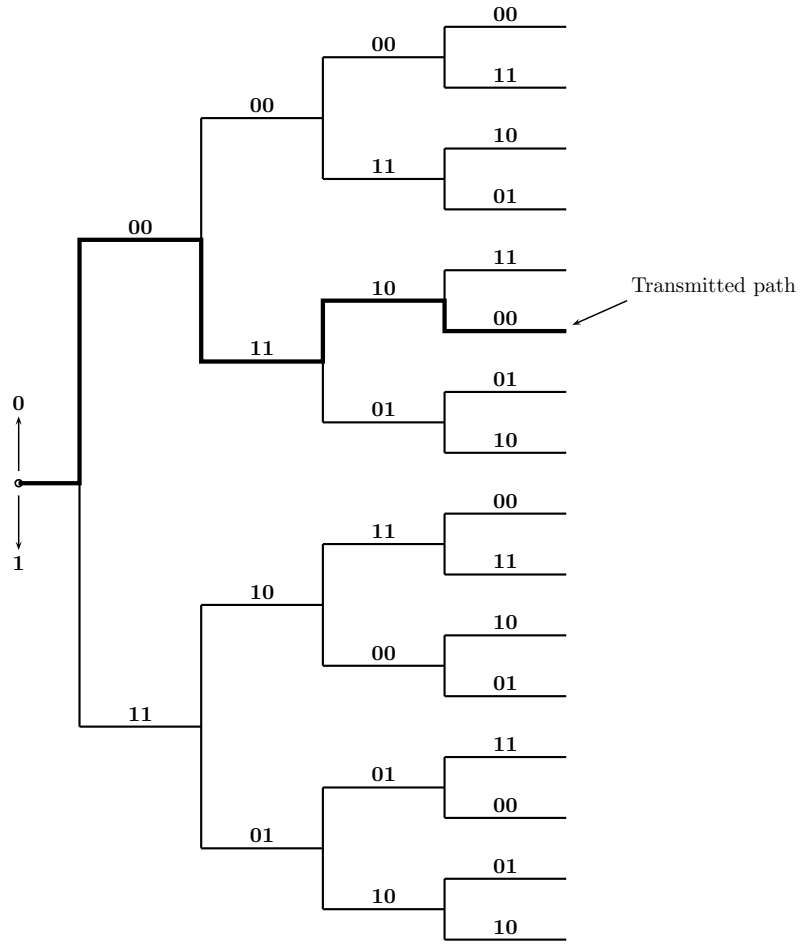
Pinsker's scheme

Massey's scheme

Polar coding

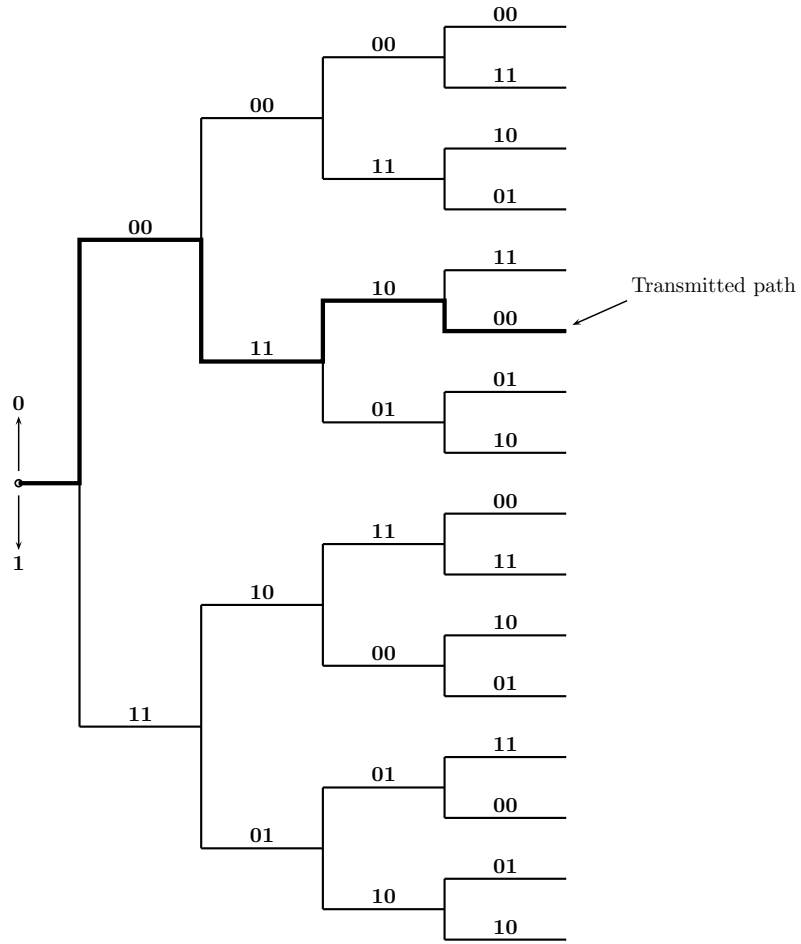
Tree coding and sequential decoding (SD)

- Consider a tree code (of rate 1/2)
- A path is chosen and transmitted
- Given the channel output, search the tree for the correct (transmitted) path
- The tree structure turns the ML decoding problem into a tree search problem
- A depth-first search algorithm exists called **sequential decoding** (SD)



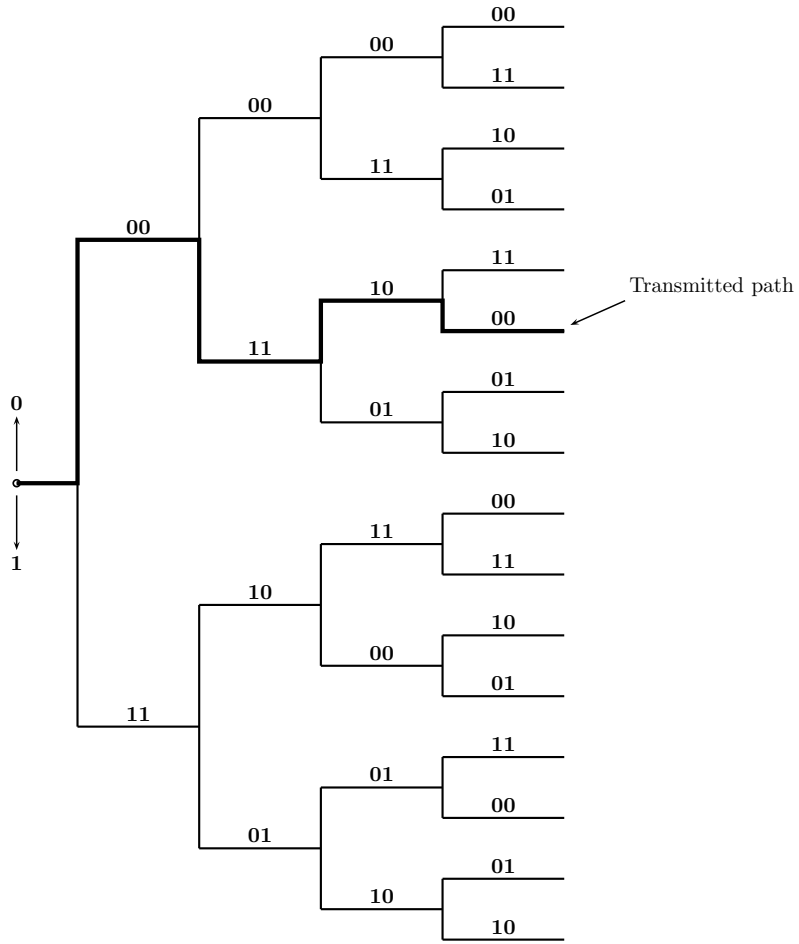
Tree coding and sequential decoding (SD)

- Consider a tree code (of rate 1/2)
- A path is chosen and transmitted
- Given the channel output, search the tree for the correct (transmitted) path
- The tree structure turns the ML decoding problem into a tree search problem
- A depth-first search algorithm exists called **sequential decoding** (SD)



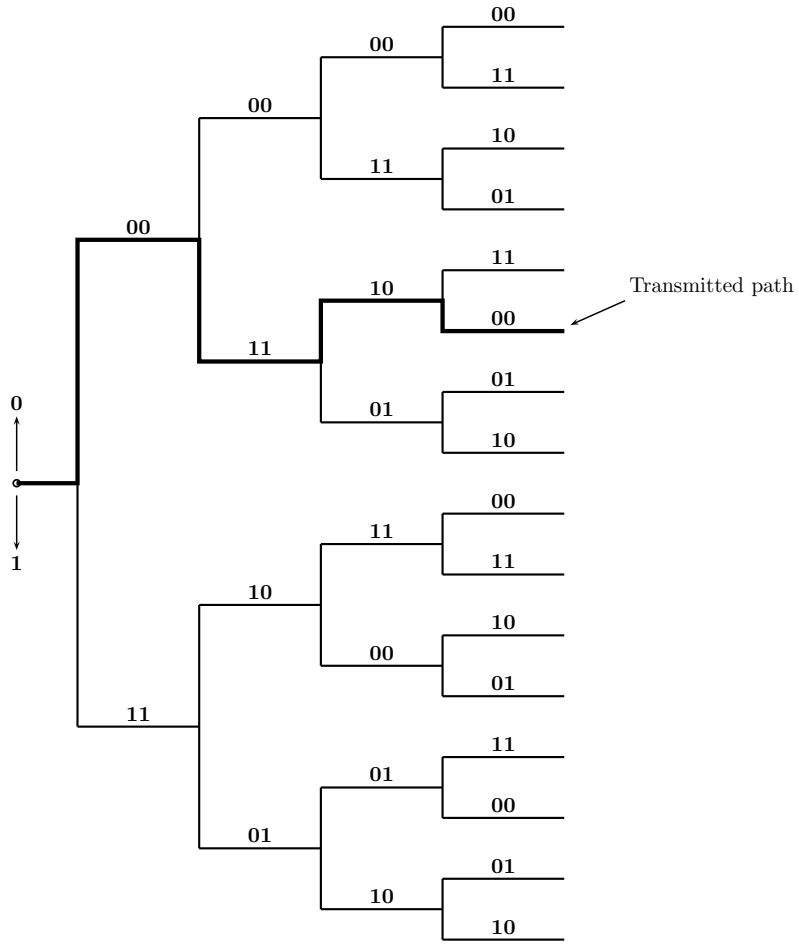
Tree coding and sequential decoding (SD)

- Consider a tree code (of rate 1/2)
- A path is chosen and transmitted
- Given the channel output, search the tree for the correct (transmitted) path
- The tree structure turns the ML decoding problem into a tree search problem
- A depth-first search algorithm exists called **sequential decoding** (SD)



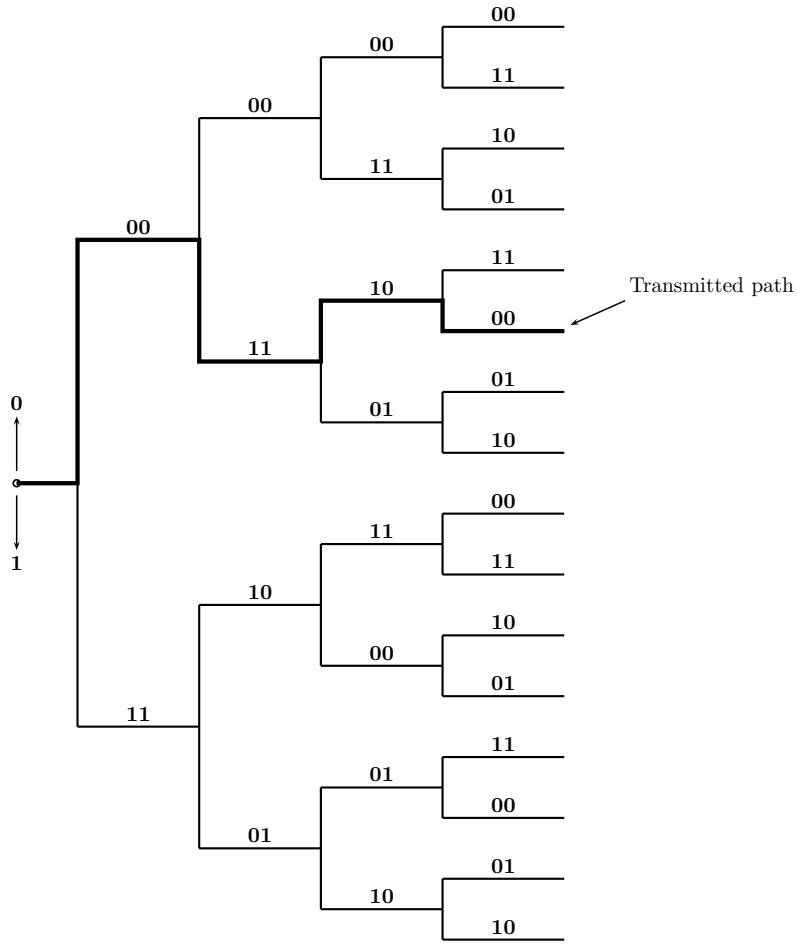
Tree coding and sequential decoding (SD)

- Consider a tree code (of rate 1/2)
- A path is chosen and transmitted
- Given the channel output, search the tree for the correct (transmitted) path
- The tree structure turns the ML decoding problem into a tree search problem
- A depth-first search algorithm exists called **sequential decoding** (SD)



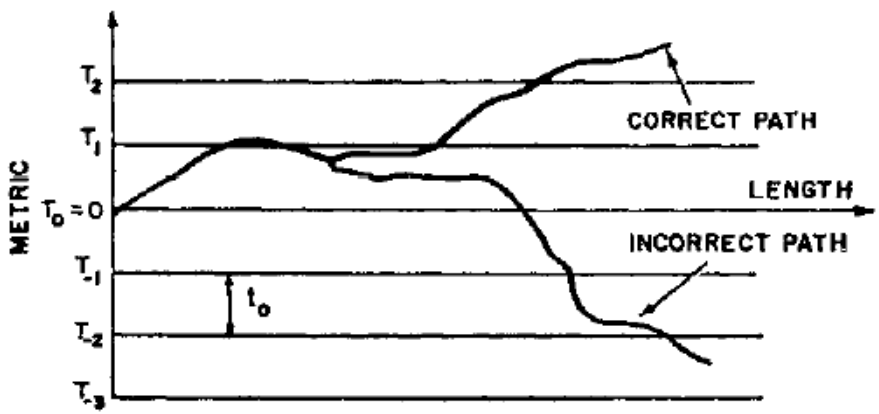
Tree coding and sequential decoding (SD)

- Consider a tree code (of rate 1/2)
- A path is chosen and transmitted
- Given the channel output, search the tree for the correct (transmitted) path
- The tree structure turns the ML decoding problem into a tree search problem
- A depth-first search algorithm exists called **sequential decoding** (SD)



Search metric

SD uses a "metric" to distinguish the correct path from the incorrect ones



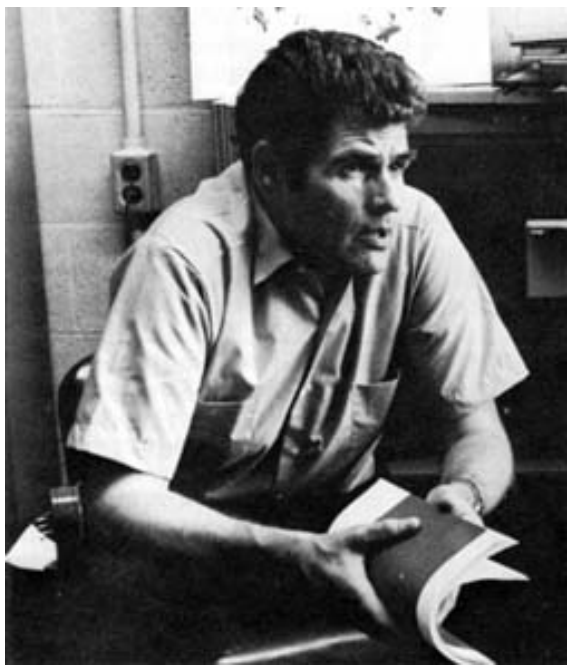
Fano's metric:

$$\Gamma(y^n, x^n) = \log \frac{P(y^n|x^n)}{P(y^n)} - nR$$

path length	n
candidate path	x^n
received sequence	y^n
code rate	R

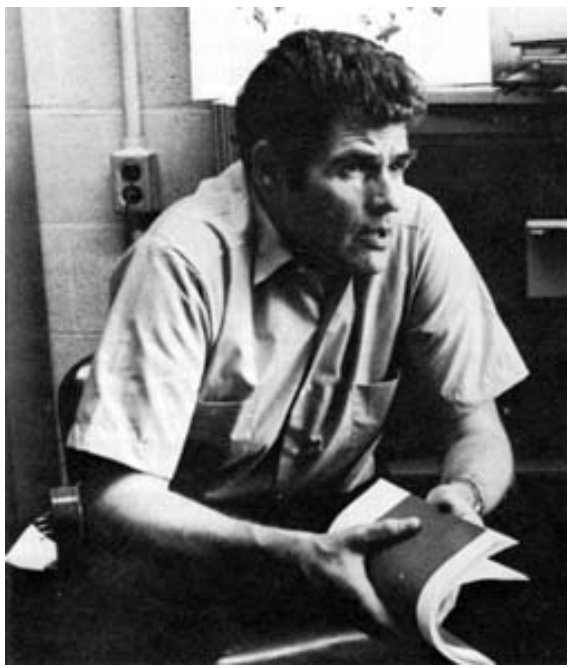
History

- Tree codes were introduced by Elias (1955) with the aim of reducing the complexity of ML decoding (the tree structure makes it possible to use search heuristics for ML decoding)
- Sequential decoding was introduced by Wozencraft (1957) as part of his doctoral thesis
- Fano (1963) simplified the search algorithm and introduced the above metric



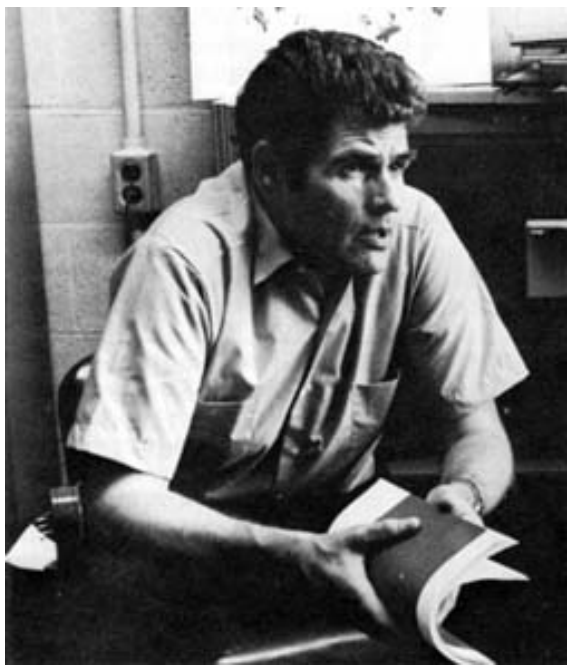
History

- Tree codes were introduced by Elias (1955) with the aim of reducing the complexity of ML decoding (the tree structure makes it possible to use search heuristics for ML decoding)
- Sequential decoding was introduced by Wozencraft (1957) as part of his doctoral thesis
- Fano (1963) simplified the search algorithm and introduced the above metric



History

- Tree codes were introduced by Elias (1955) with the aim of reducing the complexity of ML decoding (the tree structure makes it possible to use search heuristics for ML decoding)
- Sequential decoding was introduced by Wozencraft (1957) as part of his doctoral thesis
- Fano (1963) simplified the search algorithm and introduced the above metric



Drift properties of the metric

- On the correct path, the expectation of the metric per channel symbol is

$$\sum_{y,x} p(x,y)\left[\log \frac{p(y|x)}{P(y)} - R\right] = I(X;Y) - R.$$

- On any incorrect path, the expectation is

$$\sum_{x,y} p(x)p(y)\left[\log \frac{p(y|x)}{p(y)} - R\right] \leq -R$$

- A properly designed SD scheme – given enough time – identifies the correct path with probability one at any rate $R < I(X;Y)$.

Drift properties of the metric

- On the correct path, the expectation of the metric per channel symbol is

$$\sum_{y,x} p(x,y)\left[\log \frac{p(y|x)}{P(y)} - R\right] = I(X;Y) - R.$$

- On any incorrect path, the expectation is

$$\sum_{x,y} p(x)p(y)\left[\log \frac{p(y|x)}{p(y)} - R\right] \leq -R$$

- A properly designed SD scheme – given enough time – identifies the correct path with probability one at any rate $R < I(X;Y)$.

Drift properties of the metric

- On the correct path, the expectation of the metric per channel symbol is

$$\sum_{y,x} p(x,y)\left[\log \frac{p(y|x)}{P(y)} - R\right] = I(X;Y) - R.$$

- On any incorrect path, the expectation is

$$\sum_{x,y} p(x)p(y)\left[\log \frac{p(y|x)}{p(y)} - R\right] \leq -R$$

- A properly designed SD scheme – given enough time – identifies the correct path with probability one at any rate $R < I(X;Y)$.

Computation problem in sequential decoding

- Computation in sequential decoding is a random quantity, depending on the code rate R and the noise realization
- Bursts of noise create barriers for the depth-first search algorithm, necessitating excessive backtracking in the search
- Still, the average computation per decoded digit in sequential decoding can be kept bounded provided the code rate R is below the *cutoff rate*

$$R_0 \triangleq -\log \sum_y \left(\sum_x Q(x) \sqrt{W(y|x)} \right)^2$$

- So, SD solves the coding problem for rates below R_0
- Indeed, SD was the method of choice in space communications, albeit briefly

Computation problem in sequential decoding

- Computation in sequential decoding is a random quantity, depending on the code rate R and the noise realization
- Bursts of noise create barriers for the depth-first search algorithm, necessitating excessive backtracking in the search
- Still, the average computation per decoded digit in sequential decoding can be kept bounded provided the code rate R is below the *cutoff rate*

$$R_0 \triangleq -\log \sum_y \left(\sum_x Q(x) \sqrt{W(y|x)} \right)^2$$

- So, SD solves the coding problem for rates below R_0
- Indeed, SD was the method of choice in space communications, albeit briefly

Computation problem in sequential decoding

- Computation in sequential decoding is a random quantity, depending on the code rate R and the noise realization
- Bursts of noise create barriers for the depth-first search algorithm, necessitating excessive backtracking in the search
- Still, the average computation per decoded digit in sequential decoding can be kept bounded provided the code rate R is below the *cutoff rate*

$$R_0 \stackrel{\Delta}{=} -\log \sum_y \left(\sum_x Q(x) \sqrt{W(y|x)} \right)^2$$

- So, SD solves the coding problem for rates below R_0
- Indeed, SD was the method of choice in space communications, albeit briefly

Computation problem in sequential decoding

- Computation in sequential decoding is a random quantity, depending on the code rate R and the noise realization
- Bursts of noise create barriers for the depth-first search algorithm, necessitating excessive backtracking in the search
- Still, the average computation per decoded digit in sequential decoding can be kept bounded provided the code rate R is below the *cutoff rate*

$$R_0 \stackrel{\Delta}{=} -\log \sum_y \left(\sum_x Q(x) \sqrt{W(y|x)} \right)^2$$

- So, SD solves the coding problem for rates below R_0
- Indeed, SD was the method of choice in space communications, albeit briefly

Computation problem in sequential decoding

- Computation in sequential decoding is a random quantity, depending on the code rate R and the noise realization
- Bursts of noise create barriers for the depth-first search algorithm, necessitating excessive backtracking in the search
- Still, the average computation per decoded digit in sequential decoding can be kept bounded provided the code rate R is below the *cutoff rate*

$$R_0 \stackrel{\Delta}{=} -\log \sum_y \left(\sum_x Q(x) \sqrt{W(y|x)} \right)^2$$

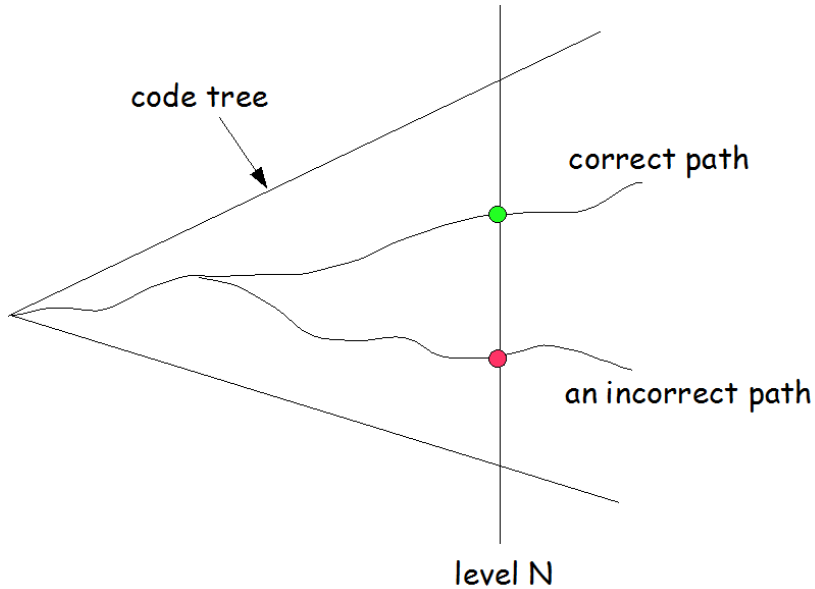
- So, SD solves the coding problem for rates below R_0
- Indeed, SD was the method of choice in space communications, albeit briefly

References on complexity of sequential decoding

- Achievability: Wozencraft (1957), Reiffen (1962), Fano (1963), Stiglitz and Yudkin (1964)
- Converse: Jacobs and Berlekamp (1967)
- Refinements: Wozencraft and Jacobs (1965), Savage (1966), Gallager (1968), Jelinek (1968), Forney (1974), Arıkan (1986), Arıkan (1994)

A computational model for sequential decoding

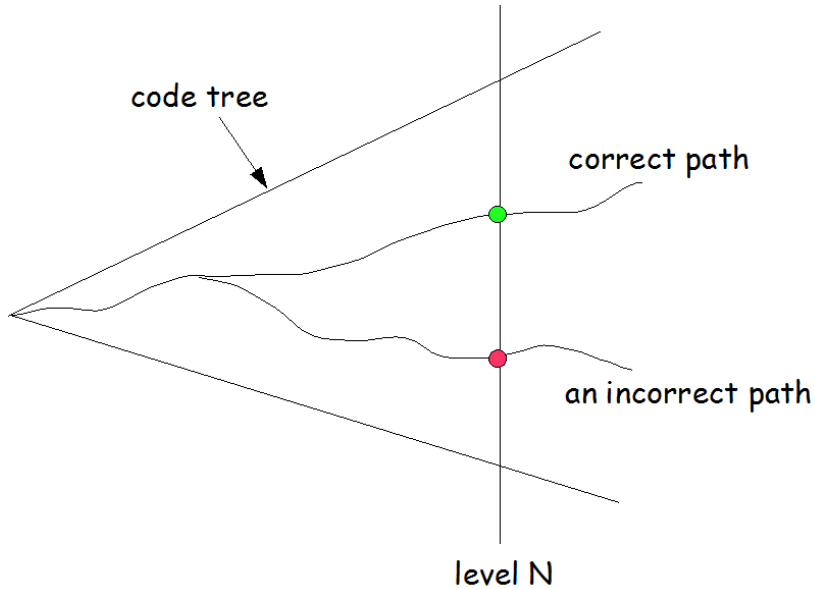
- SD visits nodes at level N in a certain order



- No "look-ahead" assumption: SD forgets what it saw beyond level N upon backtracking
- Complexity measure G_N: The number of nodes searched (visited) at level N until the correct node is visited for the first time

A computational model for sequential decoding

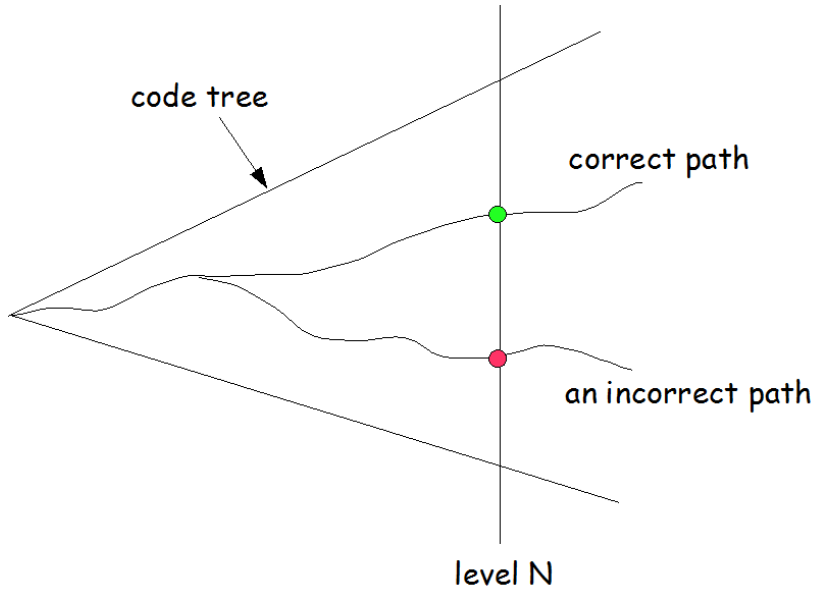
- SD visits nodes at level N in a certain order



- No "look-ahead" assumption: SD forgets what it saw beyond level N upon backtracking
- Complexity measure G_N: The number of nodes searched (visited) at level N until the correct node is visited for the first time

A computational model for sequential decoding

- SD visits nodes at level N in a certain order



- No "look-ahead" assumption: SD forgets what it saw beyond level N upon backtracking
- Complexity measure G_N: The number of nodes searched (visited) at level N until the correct node is visited for the first time

A bound of computational complexity

- Let R be a fixed code rate.
- There exist tree codes of rate R such that

$$E[G_N] \leq 1 + 2^{-N(R_0 - R)}.$$

- Conversely, for any tree code of rate R,

$$E[G_N] \gtrsim 1 + 2^{-N(R_0 - R)}$$

A bound of computational complexity

- Let R be a fixed code rate.
- There exist tree codes of rate R such that

$$E[G_N] \leq 1 + 2^{-N(R_0 - R)}.$$

- Conversely, for any tree code of rate R,

$$E[G_N] \gtrsim 1 + 2^{-N(R_0 - R)}$$

A bound of computational complexity

- Let R be a fixed code rate.
- There exist tree codes of rate R such that

$$E[G_N] \leq 1 + 2^{-N(R_0 - R)}.$$

- Conversely, for any tree code of rate R,

$$E[G_N] \gtrsim 1 + 2^{-N(R_0 - R)}$$

The Guessing Problem

- Alice draws a sample of a random variable $X \sim P$.
- Bob wishes to determine X by asking questions of the form

  "Is X equal to x ?"

  which are answered truthfully by Alice.
- Bob's goal is to minimize the expected number of questions until he gets a YES answer.

The Guessing Problem

- Alice draws a sample of a random variable $X \sim P$.
- Bob wishes to determine X by asking questions of the form

  "Is X equal to x ?"

  which are answered truthfully by Alice.
- Bob's goal is to minimize the expected number of questions until he gets a YES answer.

The Guessing Problem

- Alice draws a sample of a random variable $X \sim P$.
- Bob wishes to determine X by asking questions of the form

  "Is X equal to x ?"

  which are answered truthfully by Alice.
- Bob's goal is to minimize the expected number of questions until he gets a YES answer.

Guessing with Side Information

- Alice samples $(X, Y) \sim P(x, y)$.
- Bob observes Y and is to determine X by asking the same type of questions

  "Is X equal to x ?"

- The goal is to minimize the expected number of quesses.

Guessing with Side Information

- Alice samples $(X, Y) \sim P(x, y)$.
- Bob observes Y and is to determine X by asking the same type of questions

  "Is X equal to x ?"

- The goal is to minimize the expected number of quesses.

Guessing with Side Information

- Alice samples $(X, Y) \sim P(x, y)$.
- Bob observes Y and is to determine X by asking the same type of questions

  "Is X equal to x ?"

- The goal is to minimize the expected number of quesses.

Optimal guessing strategies

- Let G be the number of guesses to determine X.
- The expected no of guesses is given by

$$\mathbb{E}[G] = \sum_{x \in \mathcal{X}} P(x) G(x)$$

- A guessing strategy minimizes $\mathbb{E}[G]$ if

$$P(x) > P(x') \Longrightarrow G(x) < G(x').$$

Optimal guessing strategies

- Let G be the number of guesses to determine X.
- The expected no of guesses is given by

$$\mathbb{E}[G] = \sum_{x \in \mathcal{X}} P(x)G(x)$$

- A guessing strategy minimizes $\mathbb{E}[G]$ if

$$P(x) > P(x') \Longrightarrow G(x) < G(x').$$

Optimal guessing strategies

- Let G be the number of guesses to determine X.
- The expected no of guesses is given by

$$\mathbb{E}[G] = \sum_{x \in \mathcal{X}} P(x)G(x)$$

- A guessing strategy minimizes $\mathbb{E}[G]$ if

$$P(x) > P(x') \Longrightarrow G(x) < G(x').$$

Upper bound on guessing effort

For any *optimal* guessing function

$$\mathbb{E}[G^*(X)] \leq \left[\sum_x \sqrt{P(x)}\right]^2$$

Proof.

$$G^*(x) \leq \sum_{\text{all } x'} \sqrt{P(x')/P(x)} = \sum_{i=1}^{M} i p_G(i)$$

$$\mathbb{E}[G^*(X)] \leq \sum_x P(x) \sum_{x'} \sqrt{P(x')/P(x)} = \left[\sum_x \sqrt{P(x)}\right]^2 .$$

Lower bound on guessing effort

For any guessing function for a target r.v. X with M possible values,

$$\mathbb{E}[G(X)] \geq (1 + \ln M)^{-1} \left[\sum_x \sqrt{P(x)} \right]^2$$

For the proof we use the following variant of Hölder's inequality.

Lemma

Let a_i, p_i be positive numbers.

$$\sum_i a_i p_i \geq \left[\sum_i a_i^{-1}\right]^{-1} \left[\sum_i \sqrt{p_i}\right]^2 .$$

Proof. Let $\lambda = 1/2$ and put $A_i = a_i^{-1}$, $B_i = a_i^{\lambda} p_i^{\lambda}$, in Hölder's inequality

$$\sum_i A_i B_i \leq \left[\sum_i A_i^{1/(1-\lambda)}\right]^{1-\lambda} \left[\sum_i B_i^{1/\lambda}\right]^{\lambda} .$$

Proof of Lower Bound

$$
\begin{aligned}
\mathbb{E}[G(X) &= \sum_{i=1}^{M} i p_G(i) \\
&\geq \left(\sum_{i=1}^{M} 1/i \right)^{-1} \left(\sum_{i=1}^{M} \sqrt{p_G(i)} \right)^2 \\
&= \left(\sum_{i=1}^{M} 1/i \right)^{-1} \left(\sum_{x} \sqrt{P(x)} \right)^2 \\
&\geq (1 + \ln M)^{-1} \left(\sum_{x} \sqrt{P(x)} \right)^2
\end{aligned}
$$

Essense of the inequalities

For any set of real numbers $p_1 \geq p_2 \geq \cdots \geq p_M > 0$,

$$1 \geq \frac{\sum_{i=1}^{M} i\, p_i}{\left[\sum_{i=1}^{M} \sqrt{p_i}\right]^2} \geq (1 + \ln M)^{-1}$$

Guessing Random Vectors

- Let $\mathbf{X} = (X_1, \ldots, X_n) \sim P(x_1, \ldots, x_n)$.
- Guessing $\mathbf{X}$ means asking questions of the form

  $$\text{“Is } \mathbf{X} = \mathbf{x} \text{ ?”}$$

  for possible values $\mathbf{x} = (x_1, \ldots, x_n)$ of $\mathbf{X}$.
- Notice that coordinate-wise probes of the type

  $$\text{“Is } X_i = x_i \text{ ?”}$$

  are not allowed.

Complexity of Vector Guessing

Suppose X_i has M_i possible values, $i = 1, \ldots, n$. Then,

$$1 \;\geq\; \frac{\mathbb{E}[G^*(X_1, \ldots, X_n)]}{\left[\sum_{x_1,\ldots,x_n} \sqrt{P(x_1, \ldots, x_n)}\right]^2} \;\geq\; [1 + \ln(M_1 \cdots M_n)]^{-1}$$

In particular, if $X_1, \ldots, X_n$ are i.i.d. $\sim P$ with a common alphabet $\mathcal{X}$,

$$1 \;\geq\; \frac{\mathbb{E}[G^*(X_1, \ldots, X_n)]}{\left[\sum_{x \in \mathcal{X}} \sqrt{P(x)}\right]^{2n}} \;\geq\; [1 + n \ln |\mathcal{X}|]^{-1}$$

Guessing with Side Information

- (X, Y) a pair of random variables with a joint distribution $P(x, y)$.
- Y known. X to be guessed as before.
- $G(x|y)$ the number of guesses when $X = x$, $Y = y$.

Lower Bound

For any guessing strategy and any $\rho > 0$,

$$\mathbb{E}[G(X|Y)] \geq (1 + \ln M)^{-1} \sum_y \left[\sum_x \sqrt{P(x,y)} \right]^2$$

where M is the number of possible values of X.

Proof.
$$\mathbb{E}[G(X|Y)] = \sum_y P(y) \mathbb{E}[G(X|Y=y)]$$
$$\geq \sum_y P(y)(1 + \ln M)^{-1} \left[\sum_x \sqrt{P(x|y)} \right]^2$$
$$= (1 + \ln M)^{-1} \sum_y \left[\sum_x \sqrt{P(x,y)} \right]^2$$

Upper bound

Optimal guessing functions satisfy

$$\mathbb{E}[G^*(X|Y)] \leq \sum_y \left[\sum_x \sqrt{P(x,y)} \right]^2 .$$

Proof.

$$\begin{aligned}
\mathbb{E}[G^*(X|Y)] &= \sum_y P(y) \sum_x P(x|y) G^*(x|y) \\
&\leq \sum_y P(y) \left[\sum_x \sqrt{P(x|y)} \right]^2 \\
&= \sum_y \left[\sum_x \sqrt{P(x,y)} \right]^2 .
\end{aligned}$$

Generalization to Random Vectors

For optimal guessing functions, for $\rho > 0$,

$$\begin{aligned} 1 &\geq \frac{\mathbb{E}[G^*(X_1, \ldots, X_k | Y_1, \ldots, Y_n)]}{\sum_{y_1,\ldots,y_n} \left[\sum_{x_1,\ldots,x_k} \sqrt{P(x_1, \ldots, x_k, y_1, \ldots, y_n)}\right]^2} \\ &\geq [1 + \ln(M_1 \cdots M_k)]^{-1} \end{aligned}$$

where M_i denotes the number of possible values of X_i.

A "guessing" decoder

- Consider a block code with M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ of block length N.
- Suppose a codeword is chosen at random and sent over a channel W
- Given the channel output $\mathbf{y}$, a "guessing decoder" decodes by asking questions of the form

  "Is the correct codeword the mth one?"

  to which it receives a truthful YES or NO answer.
- On a NO answer it repeats the question with a new m.
- The complexity C for this decoder is the number of questions until a YES answer.

A "guessing" decoder

- Consider a block code with M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ of block length N.
- Suppose a codeword is chosen at random and sent over a channel W
- Given the channel output $\mathbf{y}$, a "guessing decoder" decodes by asking questions of the form

  "Is the correct codeword the mth one?"

  to which it receives a truthful YES or NO answer.
- On a NO answer it repeats the question with a new m.
- The complexity C for this decoder is the number of questions until a YES answer.

A "guessing" decoder

- Consider a block code with M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ of block length N.
- Suppose a codeword is chosen at random and sent over a channel W
- Given the channel output $\mathbf{y}$, a "guessing decoder" decodes by asking questions of the form

  "Is the correct codeword the mth one?"

  to which it receives a truthful YES or NO answer.
- On a NO answer it repeats the question with a new m.
- The complexity C for this decoder is the number of questions until a YES answer.

A "guessing" decoder

- Consider a block code with M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ of block length N.
- Suppose a codeword is chosen at random and sent over a channel W
- Given the channel output $\mathbf{y}$, a "guessing decoder" decodes by asking questions of the form

  "Is the correct codeword the mth one?"

  to which it receives a truthful YES or NO answer.
- On a NO answer it repeats the question with a new m.
- The complexity C for this decoder is the number of questions until a YES answer.

A "guessing" decoder

- Consider a block code with M codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ of block length N.
- Suppose a codeword is chosen at random and sent over a channel W
- Given the channel output $\mathbf{y}$, a "guessing decoder" decodes by asking questions of the form

  "Is the correct codeword the mth one?"

  to which it receives a truthful YES or NO answer.
- On a NO answer it repeats the question with a new m.
- The complexity C for this decoder is the number of questions until a YES answer.

Optimal guessing decoder

An optimal guessing decoder is one that minimizes the expected complexity $E[C]$.
Clearly, $E[C]$ is minimized by generating the guesses in decreasing order of likelihoods $W(\mathbf{y}|\mathbf{x}_m)$.

$\mathbf{x}_{i_1} \leftarrow$ 1st guess (the most likely codeword given $\mathbf{y}$)

$\mathbf{x}_{i_2} \leftarrow$ 2nd guess (2nd most likely codeword given $\mathbf{y}$)

$\vdots$

$\mathbf{x}_L \leftarrow$ correct codeword obtained; guessing stops

Complexity C equals the number of guesses L

Application to the guessing decoder

- A block code $\mathcal{C} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ with $M = e^{NR}$ codewords of block length N.
- A codeword $\mathbf{X}$ chosen at random and sent over a DMC W.
- Given the channel output vector $\mathbf{Y}$, the decoder guesses $\mathbf{X}$. A special case of guessing with side information where

$$P(\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}) = e^{-NR} \prod_{i=1}^{N} W(y_i|x_i), \quad \mathbf{x} \in \mathcal{C}$$

Cutoff rate bound

$$\mathbb{E}[G^*(\mathbf{X}|\mathbf{Y})] \geq [1+NR]^{-1} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} \sqrt{P(\mathbf{x},\mathbf{y})} \right]^2$$

$$= [1+NR]^{-1}\ e^{NR} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) \sqrt{W_N(\mathbf{x},\mathbf{y})} \right]^{2N}$$

$$\geq [1+NR]^{-1}\ e^{N(R-R_0(W))}$$

where

$$R_0(W) = \max_Q \left\{ -\ln \sum_y \left[\sum_x Q(x) \sqrt{W(y|x)} \right]^2 \right\}$$

is the channel *cutoff rate*.

Boosting the cutoff rate

- It was clear almost from the beginning that R_0 was at best shaky in its role as a limit to practical communications
- There were many attempts to boost the cutoff rate by devising clever schemes for searching a tree
- One striking example is Pinsker's scheme that displayed the strange nature of R_0

Boosting the cutoff rate

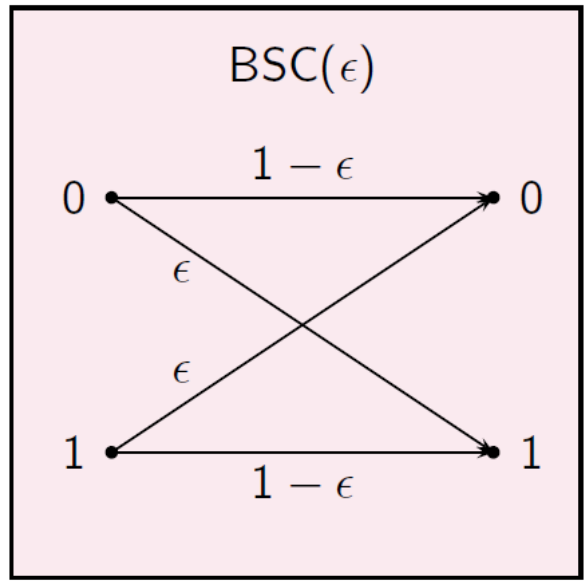
- It was clear almost from the beginning that R_0 was at best shaky in its role as a limit to practical communications
- There were many attempts to boost the cutoff rate by devising clever schemes for searching a tree
- One striking example is Pinsker's scheme that displayed the strange nature of R_0

Boosting the cutoff rate

- It was clear almost from the beginning that R_0 was at best shaky in its role as a limit to practical communications
- There were many attempts to boost the cutoff rate by devising clever schemes for searching a tree
- One striking example is Pinsker's scheme that displayed the strange nature of R_0

Binary Symmetric Channel

We will describe Pinsker's scheme using the BSC example:



- Capacity

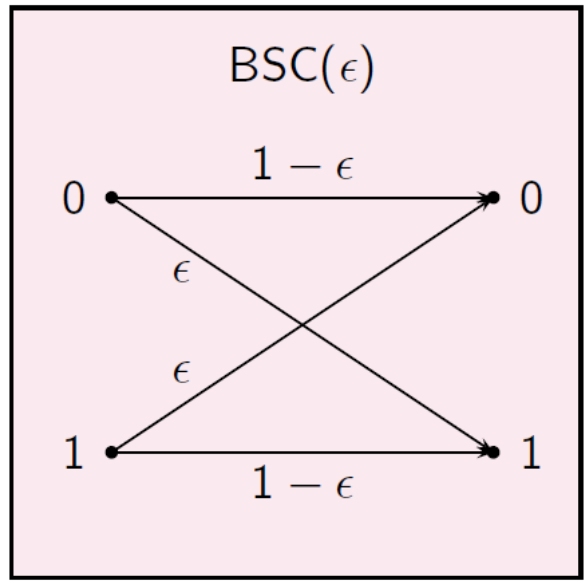
$$C = 1 + \epsilon \log_2(\epsilon) + (1-\epsilon)\log_2(1-\epsilon)$$

- Cutoff rate

$$R_0 = \log_2 \frac{2}{1 + 2\sqrt{\epsilon(1-\epsilon)}}$$

Binary Symmetric Channel

We will describe Pinsker's scheme using the BSC example:



- Capacity

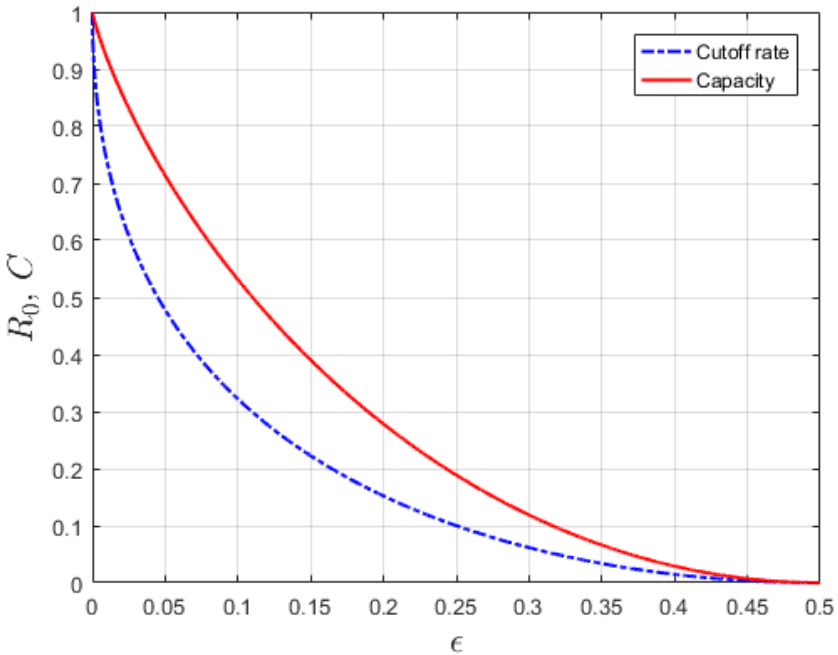
$$C = 1 + \epsilon \log_2(\epsilon) + (1-\epsilon)\log_2(1-\epsilon)$$

- Cutoff rate

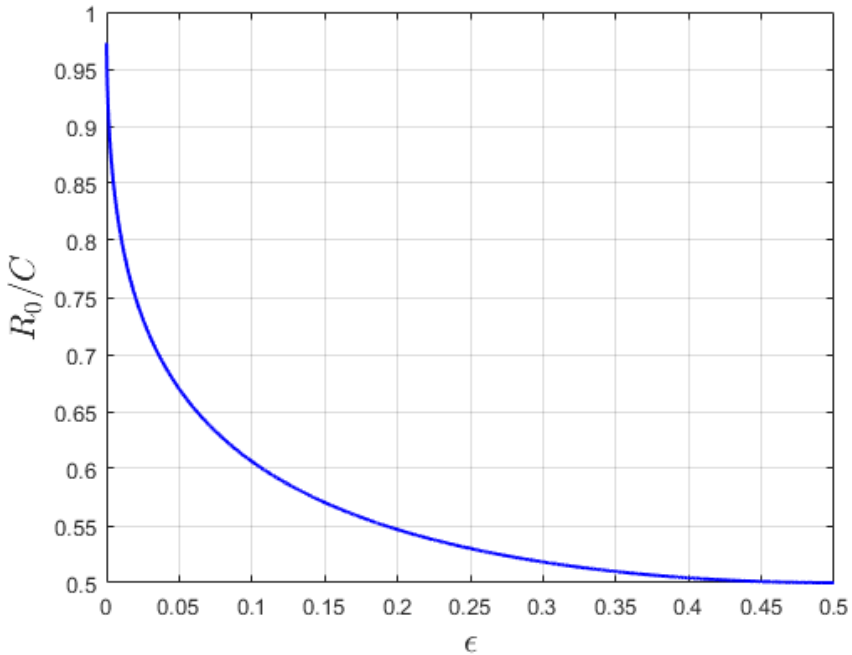
$$R_0 = \log_2 \frac{2}{1 + 2\sqrt{\epsilon(1-\epsilon)}}$$

Capacity and cutoff rate for the BSC

R_0 and C



R_0/C



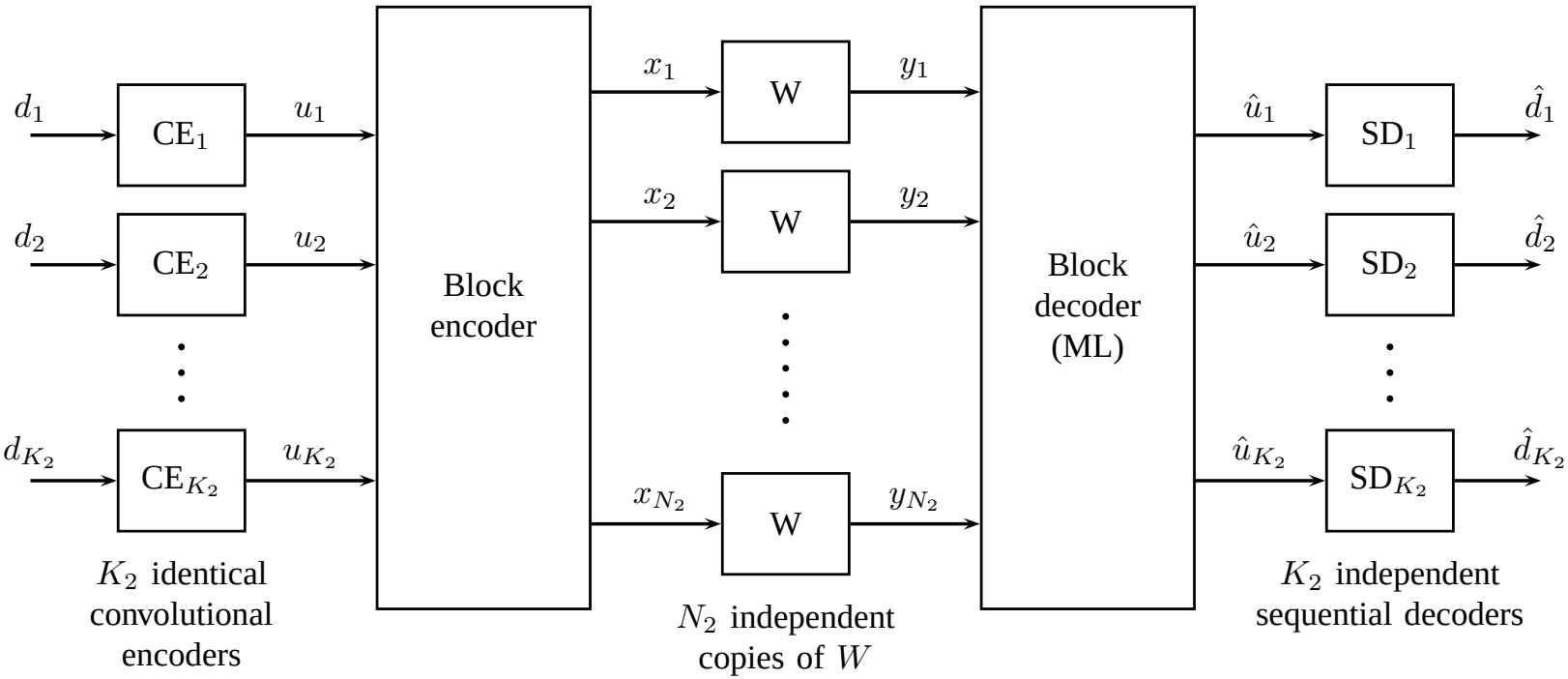
Pinsker's scheme

Based on the observations that as $\epsilon \to 0$

$$\frac{R_0(\epsilon)}{C(\epsilon)} \to 1 \quad \text{and} \quad R_0(\epsilon) \to 1,$$

Pinsker (1965) proposed concatenation scheme that achieved capacity within constant average cost per decoded bit irrespective of the level of reliability

Pinsker's scheme



The inner block code does the initial clean-up at huge but finite complexity; the outer convolutional encoding (CE) and sequential decoding (SD) boosts the reliability at little extra cost.

Discussion

- Although Pinsker's scheme made a very strong theoretical point, it was not practical.
- There were many more attempts to go around the R_0 barrier in 1960s:
  - D. Falconer, "A Hybrid Sequential and Algebraic Decoding Scheme," Sc.D. thesis, Dept. of Elec. Eng., M.I.T., 1966.
  - I. Stiglitz, Iterative sequential decoding, IEEE Transactions on Information Theory, vol. 15, no. 6, pp. 715721, Nov. 1969.
  - F. Jelinek and J. Cocke, "Bootstrap hybrid decoding for symmetrical binary input channels," Inform. Contr., vol. 18, no. 3, pp. 261-298, Apr. 1971.
- It is fair to say that none of these schemes had any practical impact

Discussion

- Although Pinsker's scheme made a very strong theoretical point, it was not practical.
- There were many more attempts to go around the R_0 barrier in 1960s:
  - D. Falconer, "A Hybrid Sequential and Algebraic Decoding Scheme," Sc.D. thesis, Dept. of Elec. Eng., M.I.T., 1966.
  - I. Stiglitz, Iterative sequential decoding, IEEE Transactions on Information Theory, vol. 15, no. 6, pp. 715721, Nov. 1969.
  - F. Jelinek and J. Cocke, "Bootstrap hybrid decoding for symmetrical binary input channels," Inform. Contr., vol. 18, no. 3, pp. 261-298, Apr. 1971.
- It is fair to say that none of these schemes had any practical impact

Discussion

- Although Pinsker's scheme made a very strong theoretical point, it was not practical.
- There were many more attempts to go around the R_0 barrier in 1960s:
  - D. Falconer, "A Hybrid Sequential and Algebraic Decoding Scheme," Sc.D. thesis, Dept. of Elec. Eng., M.I.T., 1966.
  - I. Stiglitz, Iterative sequential decoding, IEEE Transactions on Information Theory, vol. 15, no. 6, pp. 715721, Nov. 1969.
  - F. Jelinek and J. Cocke, "Bootstrap hybrid decoding for symmetrical binary input channels," Inform. Contr., vol. 18, no. 3, pp. 261-298, Apr. 1971.
- It is fair to say that none of these schemes had any practical impact

Discussion

- Although Pinsker's scheme made a very strong theoretical point, it was not practical.
- There were many more attempts to go around the R_0 barrier in 1960s:
  - D. Falconer, "A Hybrid Sequential and Algebraic Decoding Scheme," Sc.D. thesis, Dept. of Elec. Eng., M.I.T., 1966.
  - I. Stiglitz, Iterative sequential decoding, IEEE Transactions on Information Theory, vol. 15, no. 6, pp. 715721, Nov. 1969.
  - F. Jelinek and J. Cocke, "Bootstrap hybrid decoding for symmetrical binary input channels," Inform. Contr., vol. 18, no. 3, pp. 261-298, Apr. 1971.
- It is fair to say that none of these schemes had any practical impact

Discussion

- Although Pinsker's scheme made a very strong theoretical point, it was not practical.
- There were many more attempts to go around the R_0 barrier in 1960s:
  - D. Falconer, "A Hybrid Sequential and Algebraic Decoding Scheme," Sc.D. thesis, Dept. of Elec. Eng., M.I.T., 1966.
  - I. Stiglitz, Iterative sequential decoding, IEEE Transactions on Information Theory, vol. 15, no. 6, pp. 715721, Nov. 1969.
  - F. Jelinek and J. Cocke, "Bootstrap hybrid decoding for symmetrical binary input channels," Inform. Contr., vol. 18, no. 3, pp. 261-298, Apr. 1971.
- It is fair to say that none of these schemes had any practical impact

Discussion

- Although Pinsker's scheme made a very strong theoretical point, it was not practical.
- There were many more attempts to go around the R_0 barrier in 1960s:
  - D. Falconer, "A Hybrid Sequential and Algebraic Decoding Scheme," Sc.D. thesis, Dept. of Elec. Eng., M.I.T., 1966.
  - I. Stiglitz, Iterative sequential decoding, IEEE Transactions on Information Theory, vol. 15, no. 6, pp. 715721, Nov. 1969.
  - F. Jelinek and J. Cocke, "Bootstrap hybrid decoding for symmetrical binary input channels," Inform. Contr., vol. 18, no. 3, pp. 261-298, Apr. 1971.
- It is fair to say that none of these schemes had any practical impact

R_0 as practical capacity

- The failure to beat the cutoff rate bound in a meaningful manner despite intense efforts elevated R_0 to the status of a "realistic" limit to reliable communications
- R_0 appears as the key figure-of-merit for communication system design in the influential works of the period:
  - Wozencraft and Jacobs, *Principles of Communication Engineering*, 1965
  - Wozencraft and Kennedy, "Modulation and demodulation for probabilistic coding," IT Trans.,1966
  - Massey, "Coding and modulation in digital communications," Zürich, 1974
- Forney (1995) gives a first-hand account of this situation in his Shannon Lecture "Performance and Complexity"

R_0 as practical capacity

- The failure to beat the cutoff rate bound in a meaningful manner despite intense efforts elevated R_0 to the status of a "realistic" limit to reliable communications
- R_0 appears as the key figure-of-merit for communication system design in the influential works of the period:
  - Wozencraft and Jacobs, *Principles of Communication Engineering*, 1965
  - Wozencraft and Kennedy, "Modulation and demodulation for probabilistic coding," IT Trans.,1966
  - Massey, "Coding and modulation in digital communications," Zürich, 1974
- Forney (1995) gives a first-hand account of this situation in his Shannon Lecture "Performance and Complexity"

R_0 as practical capacity

- The failure to beat the cutoff rate bound in a meaningful manner despite intense efforts elevated R_0 to the status of a "realistic" limit to reliable communications
- R_0 appears as the key figure-of-merit for communication system design in the influential works of the period:
  - Wozencraft and Jacobs, *Principles of Communication Engineering*, 1965
  - Wozencraft and Kennedy, "Modulation and demodulation for probabilistic coding," IT Trans.,1966
  - Massey, "Coding and modulation in digital communications," Zürich, 1974
- Forney (1995) gives a first-hand account of this situation in his Shannon Lecture "Performance and Complexity"

R_0 as practical capacity

- The failure to beat the cutoff rate bound in a meaningful manner despite intense efforts elevated R_0 to the status of a "realistic" limit to reliable communications
- R_0 appears as the key figure-of-merit for communication system design in the influential works of the period:
  - Wozencraft and Jacobs, *Principles of Communication Engineering*, 1965
  - Wozencraft and Kennedy, "Modulation and demodulation for probabilistic coding," IT Trans.,1966
  - Massey, "Coding and modulation in digital communications," Zürich, 1974
- Forney (1995) gives a first-hand account of this situation in his Shannon Lecture "Performance and Complexity"

R_0 as practical capacity

- The failure to beat the cutoff rate bound in a meaningful manner despite intense efforts elevated R_0 to the status of a "realistic" limit to reliable communications
- R_0 appears as the key figure-of-merit for communication system design in the influential works of the period:
  - Wozencraft and Jacobs, *Principles of Communication Engineering*, 1965
  - Wozencraft and Kennedy, "Modulation and demodulation for probabilistic coding," IT Trans.,1966
  - Massey, "Coding and modulation in digital communications," Zürich, 1974
- Forney (1995) gives a first-hand account of this situation in his Shannon Lecture "Performance and Complexity"

R_0 as practical capacity

- The failure to beat the cutoff rate bound in a meaningful manner despite intense efforts elevated R_0 to the status of a "realistic" limit to reliable communications
- R_0 appears as the key figure-of-merit for communication system design in the influential works of the period:
  - Wozencraft and Jacobs, *Principles of Communication Engineering*, 1965
  - Wozencraft and Kennedy, "Modulation and demodulation for probabilistic coding," IT Trans.,1966
  - Massey, "Coding and modulation in digital communications," Zürich, 1974
- Forney (1995) gives a first-hand account of this situation in his Shannon Lecture "Performance and Complexity"

Other attempts to boost the cutoff rate

Efforts to beat the cutoff rate continues to this day

- D. J. Costello and F. Jelinek, 1972.
- P. R. Chevillat and D. J. Costello Jr., 1977.
- F. Hemmati, 1990.
- B. Radosavljevic, E. Arıkan, B. Hajek, 1992.
- J. Belzile and D. Haccoun, 1993.
- S. Kallel and K. Li, 1997.
- E. Arıkan, 2006
- ...

Other attempts to boost the cutoff rate

Efforts to beat the cutoff rate continues to this day

- D. J. Costello and F. Jelinek, 1972.
- P. R. Chevillat and D. J. Costello Jr., 1977.
- F. Hemmati, 1990.
- B. Radosavljevic, E. Arıkan, B. Hajek, 1992.
- J. Belzile and D. Haccoun, 1993.
- S. Kallel and K. Li, 1997.
- E. Arıkan, 2006
- ...

In fact, polar coding originates from such attempts.

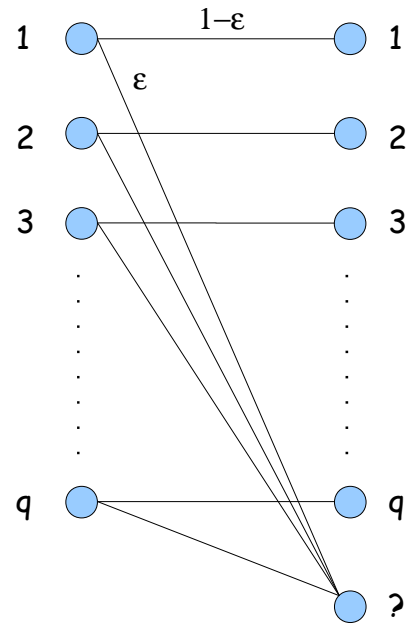
The R_0 debate

A case study by McEliece (1980) cast a big doubt on the significance of R_0 as a practical limit

- McEliece's study was concerned with a Pulse Position Modulation (PPM) scheme, modeled as a q-ary erasure channel
- Capacity: $C(q) = (1 - \epsilon) \log q$
- Cutoff rate: $R_0(q) = \log \frac{q}{1+(q-1)\epsilon}$
- As the bandwidth (q) grew,

$$\frac{R_0(q)}{C(q)} \to 0$$

- Algebraic coding (Reed-Solomon) scored a big win over probabilistic coding!



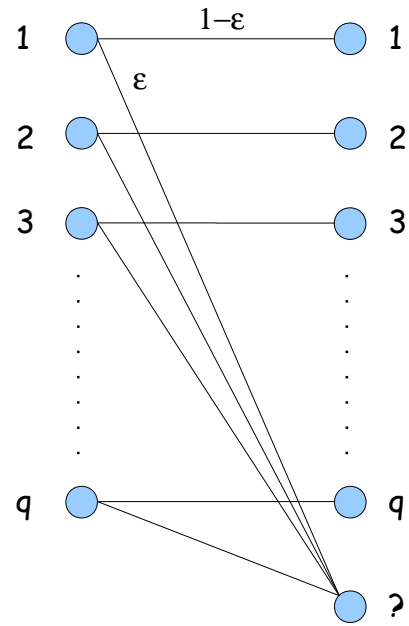
The R_0 debate

A case study by McEliece (1980) cast a big doubt on the significance of R_0 as a practical limit

- McEliece's study was concerned with a Pulse Position Modulation (PPM) scheme, modeled as a q-ary erasure channel
- Capacity: $C(q) = (1 - \epsilon) \log q$
- Cutoff rate: $R_0(q) = \log \frac{q}{1+(q-1)\epsilon}$
- As the bandwidth (q) grew,

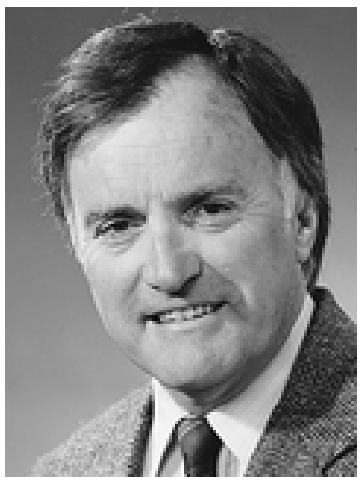
$$\frac{R_0(q)}{C(q)} \to 0$$

- Algebraic coding (Reed-Solomon) scored a big win over probabilistic coding!



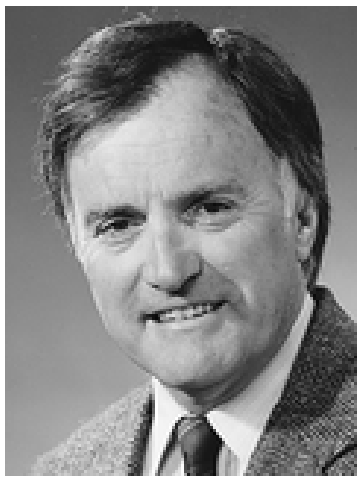
Massey meets the challenge

- Massey (1981) showed that there was a different way of doing coding and modulation on a q-ary erasure channel that boosted R_0 effortlessly
- Paradoxically, as Massey restored the status of R_0, he exhibited the "flaky" nature of this parameter

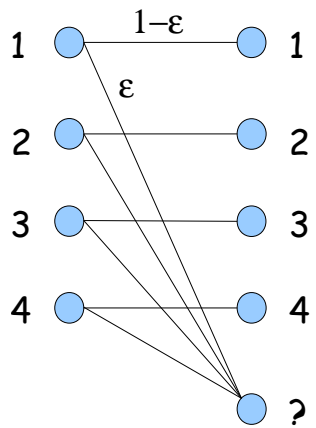


Massey meets the challenge

- Massey (1981) showed that there was a different way of doing coding and modulation on a q-ary erasure channel that boosted R_0 effortlessly
- Paradoxically, as Massey restored the status of R_0, he exhibited the "flaky" nature of this parameter

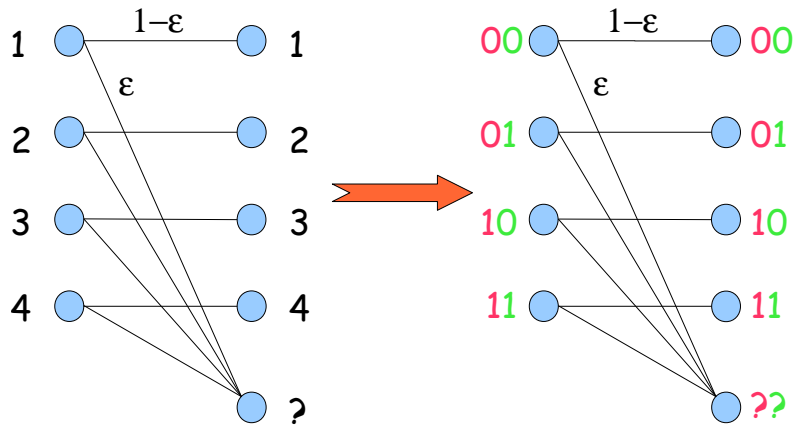


Channel splitting to boost cutoff rate (Massey, 1981)



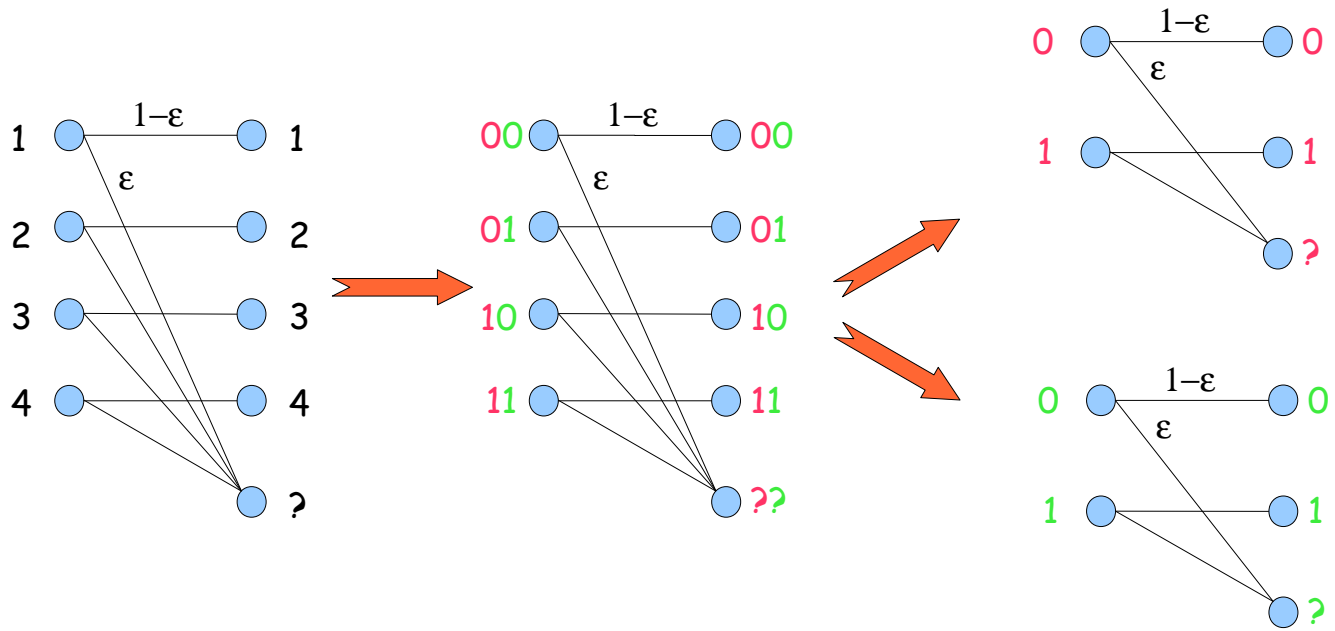
- Begin with a quaternary erasure channel (QEC)

Channel splitting to boost cutoff rate (Massey, 1981)



- Relabel the inputs

Channel splitting to boost cutoff rate (Massey, 1981)



- Split the QEC into two binary erasure channels (BEC)
- BECs fully correlated: erasures occur jointly

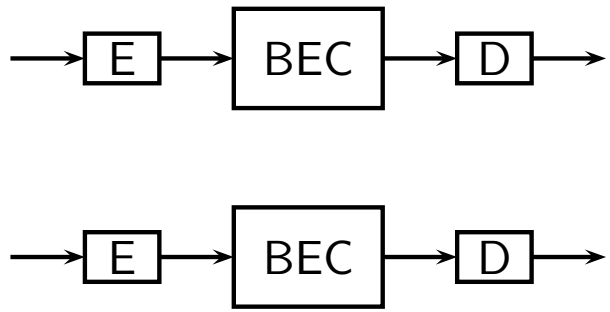
Capacity, cutoff rate for one QEC vs two BECs

Ordinary coding of QEC

$C(\text{QEC}) = 2(1-\epsilon)$

$R_0(\text{QEC}) = \log \frac{4}{1+3\epsilon}$

Independent coding of BECs



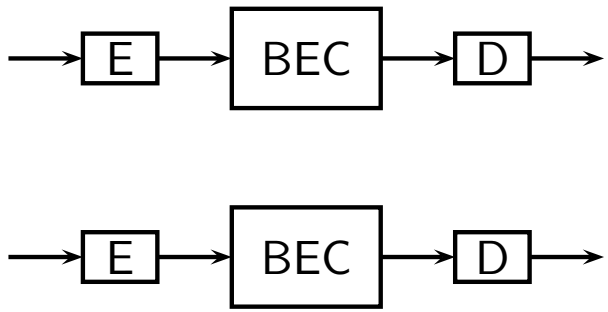
$C(\text{BEC}) = (1-\epsilon)$

$R_0(\text{BEC}) = \log \frac{2}{1+\epsilon}$

Capacity, cutoff rate for one QEC vs two BECs

Ordinary coding of QEC

Independent coding of BECs



$C(\text{QEC}) = 2(1-\epsilon)$

$R_0(\text{QEC}) = \log \frac{4}{1+3\epsilon}$

$C(\text{BEC}) = (1-\epsilon)$

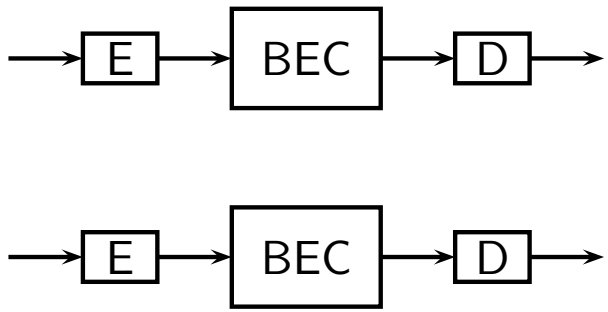
$R_0(\text{BEC}) = \log \frac{2}{1+\epsilon}$

- $C(\text{QEC}) = 2 \times C(\text{BEC})$

Capacity, cutoff rate for one QEC vs two BECs

Ordinary coding of QEC

Independent coding of BECs



$C(\text{QEC}) = 2(1-\epsilon)$

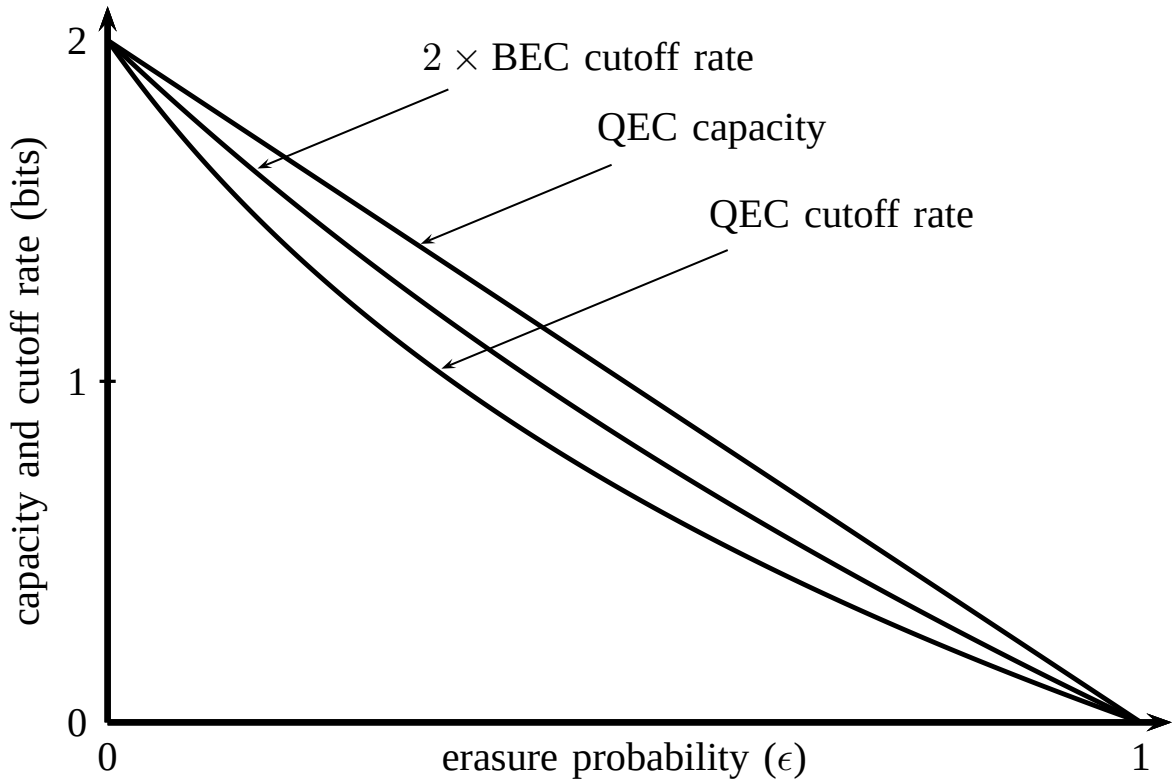
$R_0(\text{QEC}) = \log \frac{4}{1+3\epsilon}$

$C(\text{BEC}) = (1-\epsilon)$

$R_0(\text{BEC}) = \log \frac{2}{1+\epsilon}$

- $C(\text{QEC}) = 2 \times C(\text{BEC})$
- $R_0(\text{QEC}) \leq 2 \times R_0(\text{BEC})$ with equality iff $\epsilon = 0$ or 1.

Cutoff rate improvement by splitting



Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

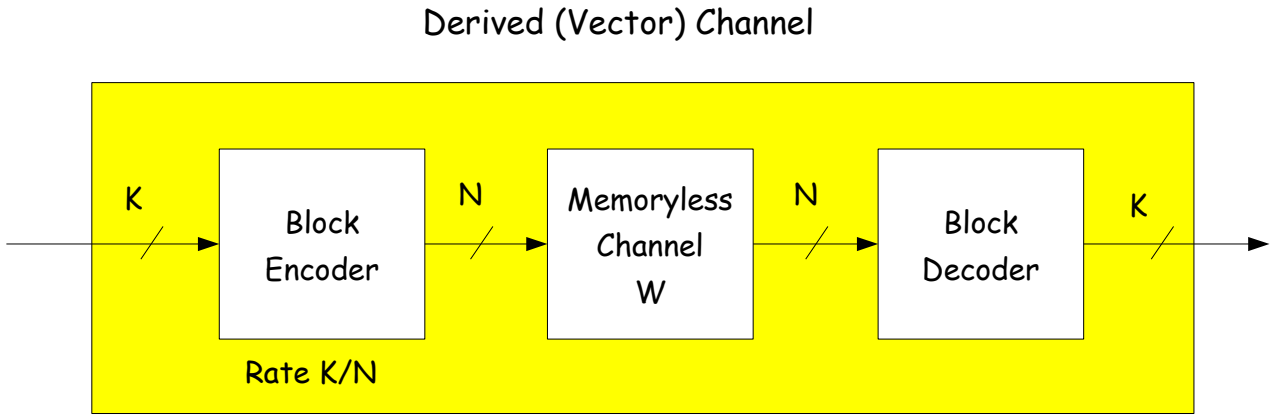
Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

Comparison of Pinsker's and Massey's schemes

- Pinsker
  - Construct a superchannel by combining independent copies of a given DMC W
  - Split the superchannel into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Can be used universally
  - Can achieve capacity
  - Not practical
- Massey
  - Split the given DMC W into correlated subchannels
  - Ignore correlations between the subchannels, encode and decode them independently
  - Applicable only to specific channels
  - Cannot achieve capacity
  - Practical

A conservation law for the cutoff rate

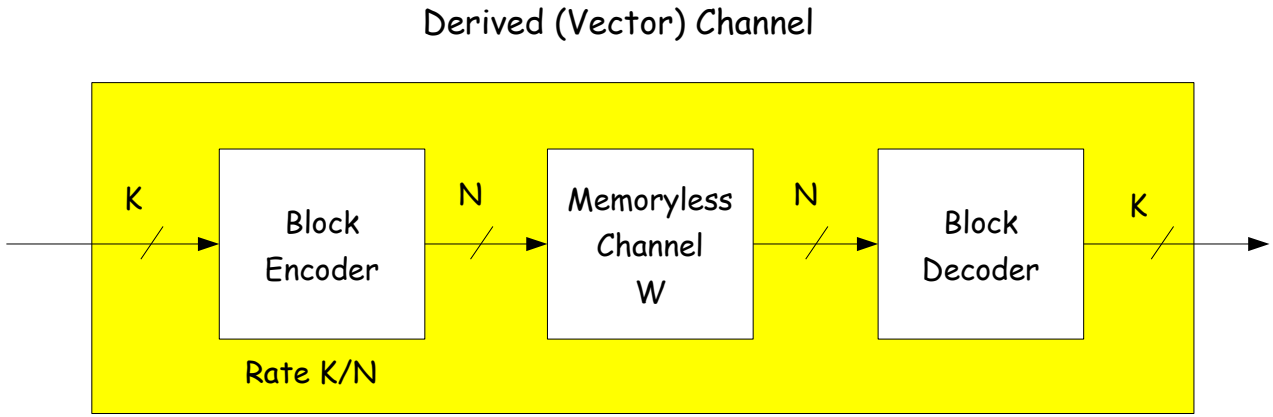


- "Parallel channels" theorem (Gallager, 1965)

$$R_0(\text{Derived vector channel}) \leq N\,R_0(W)$$

- "Cleaning up" the channel by pre-/post-processing can only hurt R_0
- Shows that boosting cutoff rate requires more than one sequential decoder

A conservation law for the cutoff rate

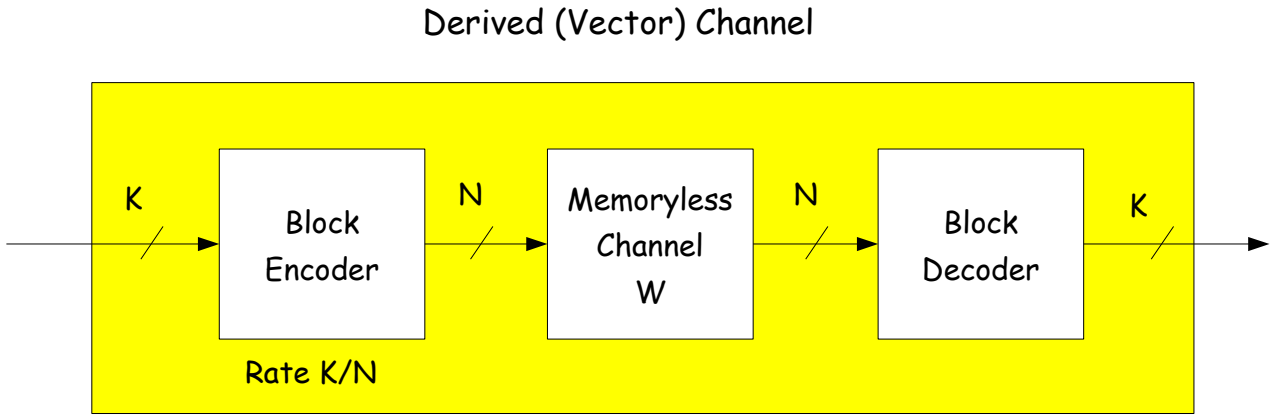


- "Parallel channels" theorem (Gallager, 1965)

$$R_0(\text{Derived vector channel}) \leq N\,R_0(W)$$

- "Cleaning up" the channel by pre-/post-processing can only hurt R_0
- Shows that boosting cutoff rate requires more than one sequential decoder

A conservation law for the cutoff rate

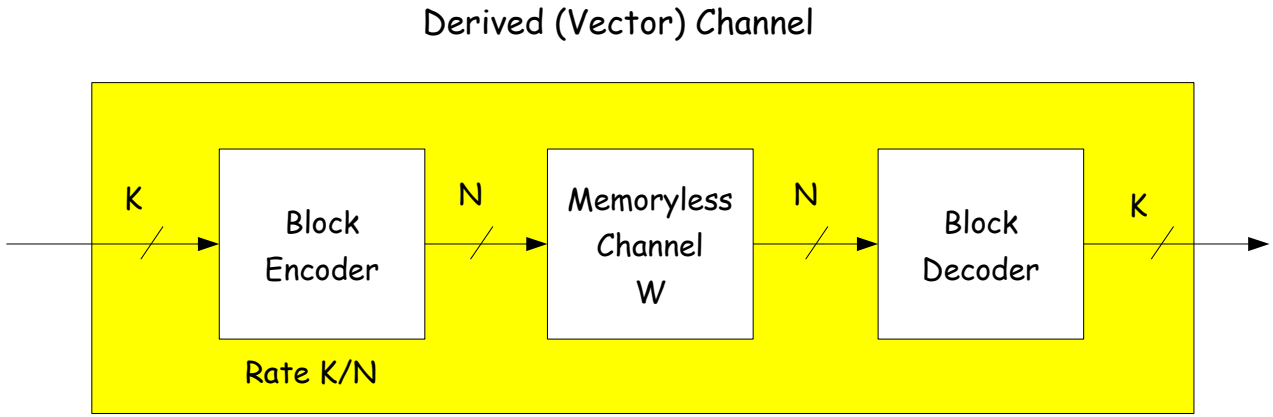


- "Parallel channels" theorem (Gallager, 1965)

$$R_0(\text{Derived vector channel}) \leq N\, R_0(W)$$

- "Cleaning up" the channel by pre-/post-processing can only hurt R_0
- Shows that boosting cutoff rate requires more than one sequential decoder

A conservation law for the cutoff rate



- "Parallel channels" theorem (Gallager, 1965)

$$R_0(\text{Derived vector channel}) \leq N\,R_0(W)$$

- "Cleaning up" the channel by pre-/post-processing can only hurt R_0
- Shows that boosting cutoff rate requires more than one sequential decoder

Sequential decoding and the cutoff rate

Guessing and cutoff rate

Boosting the cutoff rate

Pinsker's scheme

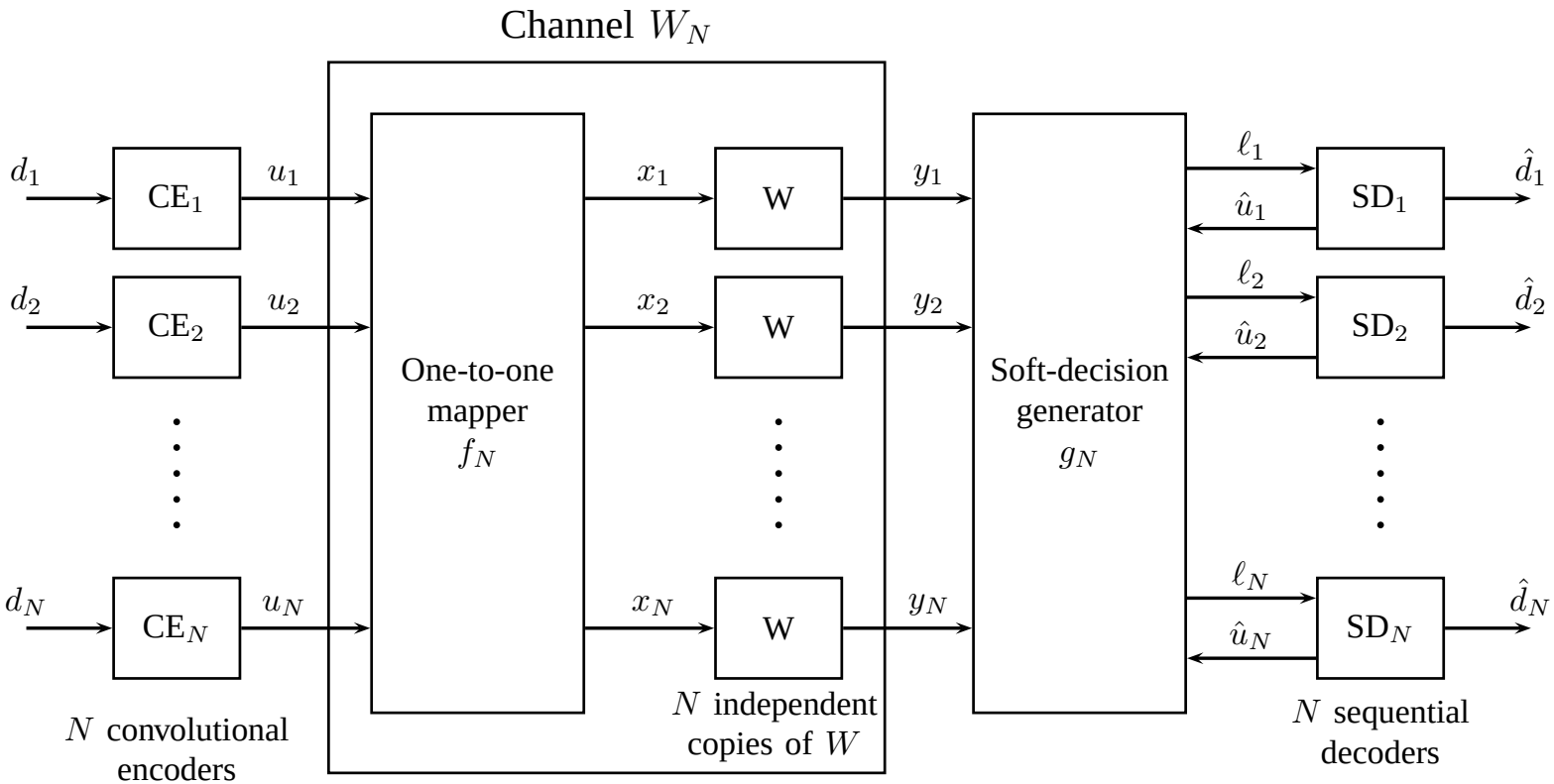
Massey's scheme

Polar coding

Prescription for a new scheme

- Consider small constructions
- Retain independent encoding for the subchannels
- Do not ignore correlations between subchannels at the expense of capacity
- This points to multi-level coding and successive cancellation decoding

Multi-stage decoding architecture



Prescription for a new scheme

- Consider small constructions
- Retain independent encoding for the subchannels
- Do not ignore correlations between subchannels at the expense of capacity
- This points to multi-level coding and successive cancellation decoding

Notation

- Let $V : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$ be an arbitrary binary-input memoryless channel
- Let (X, Y) be an input-output ensemble for channel V with X uniform on $\mathbb{F}_2$
- The (symmetric) capacity is defined as

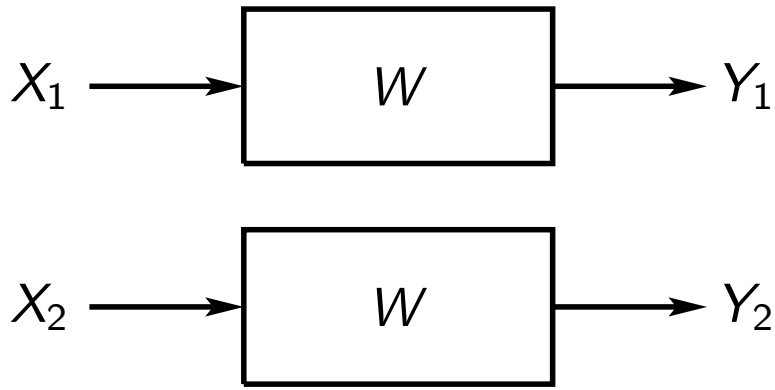
$$I(V) \stackrel{\Delta}{=} I(X;Y) \stackrel{\Delta}{=} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathbb{F}_2} \tfrac{1}{2} V(y|x) \log \frac{V(y|x)}{\frac{1}{2}V(y|0) + \frac{1}{2}V(y|1)}$$

- The (symmetric) cutoff rate is defined as

$$R_0(V) \stackrel{\Delta}{=} R_0(X;Y) \stackrel{\Delta}{=} -\log \sum_{y \in \mathcal{Y}} \left[\sum_{x \in \mathbb{F}_2} \tfrac{1}{2} \sqrt{V(y|x)} \right]^2$$

The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$



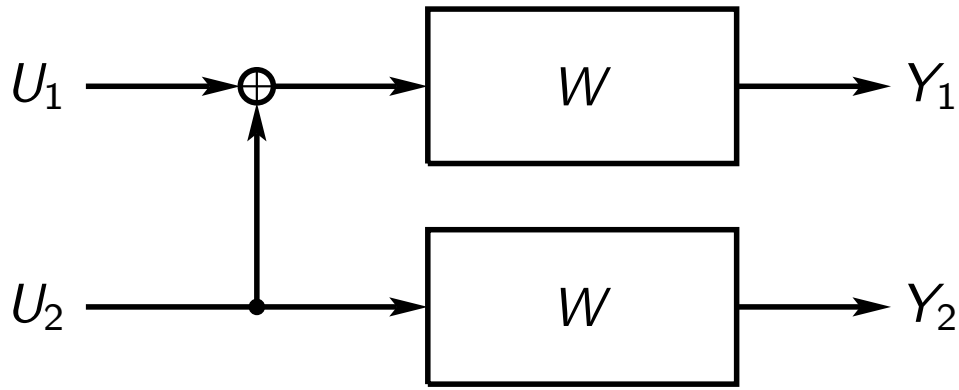
consider the transformation above to generate two channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ with

$$W^-(y_1 y_2|u_1) = \sum_{u_2} \tfrac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2)$$

$$W^+(y_1 y_2 u_1|u_2) = \tfrac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2)$$

The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$



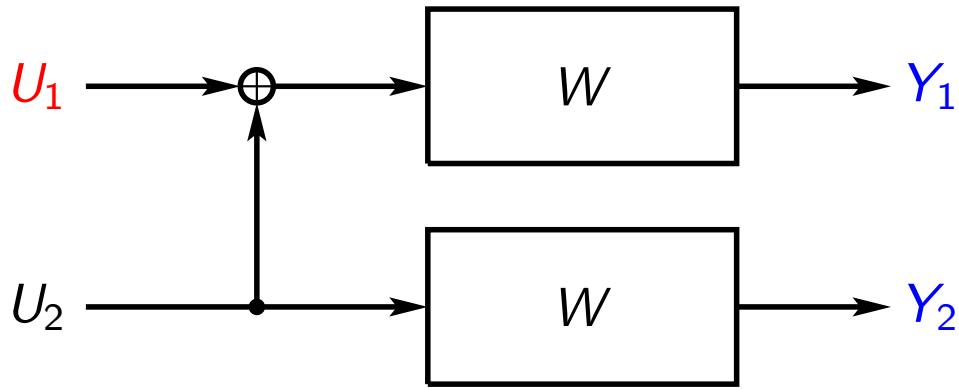
consider the transformation above to generate two channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ with

$$W^-(y_1 y_2|u_1) = \sum_{u_2} \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

$$W^+(y_1 y_2 u_1|u_2) = \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$



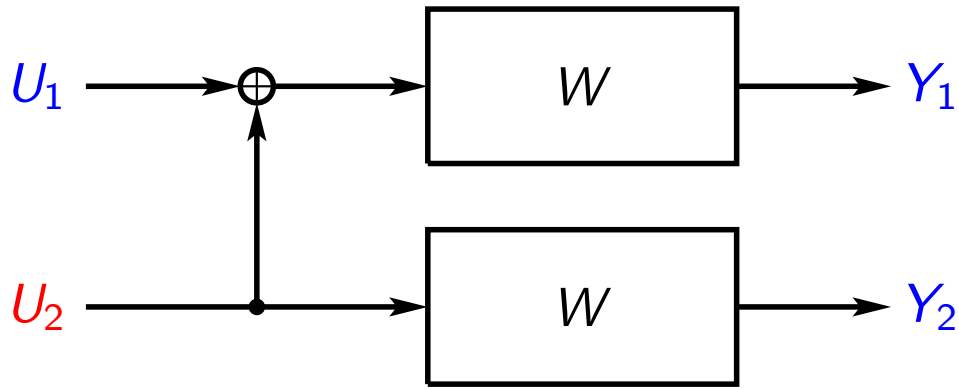
consider the transformation above to generate two channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ with

$$W^-(y_1 y_2|u_1) = \sum_{u_2} \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

$$W^+(y_1 y_2 u_1|u_2) = \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$



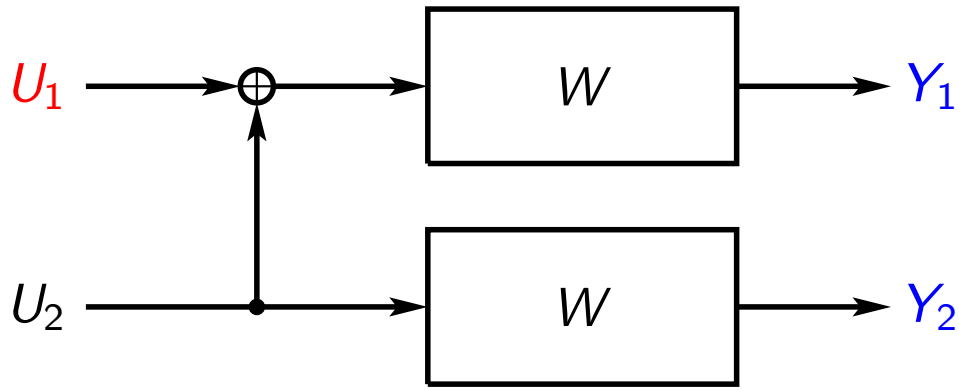
consider the transformation above to generate two channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ with

$$W^-(y_1 y_2|u_1) = \sum_{u_2} \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

$$W^+(y_1 y_2 u_1|u_2) = \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$



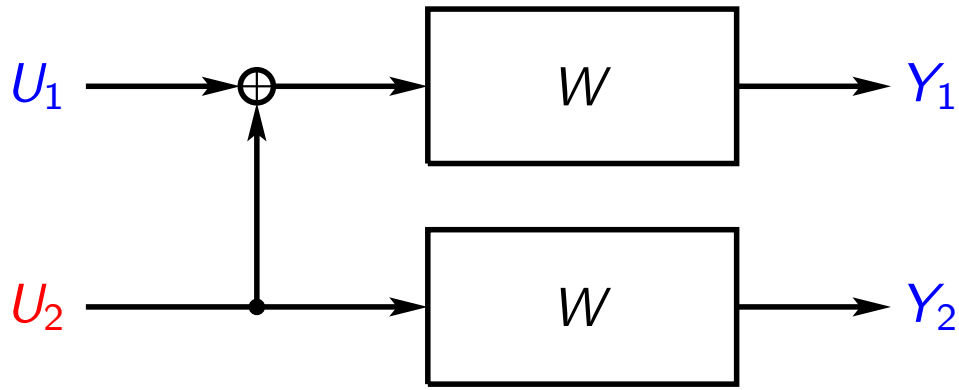
consider the transformation above to generate two channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ with

$$W^-(y_1 y_2|u_1) = \sum_{u_2} \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

$$W^+(y_1 y_2 u_1|u_2) = \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \stackrel{\Delta}{=} \{0,1\} \to \mathcal{Y}$



consider the transformation above to generate two channels $W^- : F_2 \to \mathcal{Y}^2$ and $W^+ : F_2 \to \mathcal{Y}^2 \times F_2$ with

$$W^-(y_1 y_2|u_1) = \sum_{u_2} \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

$$W^+(y_1 y_2 u_1|u_2) = \tfrac{1}{2} W(y_1|u_1+u_2) W(y_2|u_2)$$

The 2x2 transformation is information lossless

- With independent, uniform U_1, U_2,

$$I(W^-) = I(U_1; Y_1 Y_2),$$
$$I(W^+) = I(U_2; Y_1 Y_2 U_1).$$

- Thus,

$$\begin{aligned} I(W^-) + I(W^+) &= I(U_1 U_2; Y_1 Y_2) \\ &= 2I(W), \end{aligned}$$

- and $I(W^-) \leq I(W) \leq I(W^+)$.

The 2x2 transformation "creates" cutoff rate

With independent, uniform U_1, U_2,

$$R_0(W^-) = R_0(U_1; Y_1 Y_2),$$
$$R_0(W^+) = R_0(U_2; Y_1 Y_2 U_1).$$

Theorem (2005)

Correlation helps create cutoff rate:

$$R_0(W^-) + R_0(W^+) \geq 2R_0(W)$$

with equality iff W is a perfect channel, $I(W) = 1$, or a pure noise channel, $I(W) = 0$. Cutoff rates start polarizing:

$$R_0(W^-) \leq R_0(W) \leq R_0(W^+)$$

The 2x2 transformation "creates" cutoff rate

With independent, uniform U_1, U_2,

$$R_0(W^-) = R_0(U_1; Y_1 Y_2),$$
$$R_0(W^+) = R_0(U_2; Y_1 Y_2 U_1).$$

Theorem (2005)

Correlation helps create cutoff rate:

$$R_0(W^-) + R_0(W^+) \geq 2R_0(W)$$

with equality iff W is a perfect channel, $I(W) = 1$, or a pure noise channel, $I(W) = 0$. Cutoff rates start polarizing:

$$R_0(W^-) \leq R_0(W) \leq R_0(W^+)$$

The 2x2 transformation "creates" cutoff rate

With independent, uniform U_1, U_2,

$$R_0(W^-) = R_0(U_1; Y_1 Y_2),$$
$$R_0(W^+) = R_0(U_2; Y_1 Y_2 U_1).$$

Theorem (2005)

Correlation helps create cutoff rate:

$$R_0(W^-) + R_0(W^+) \geq 2R_0(W)$$

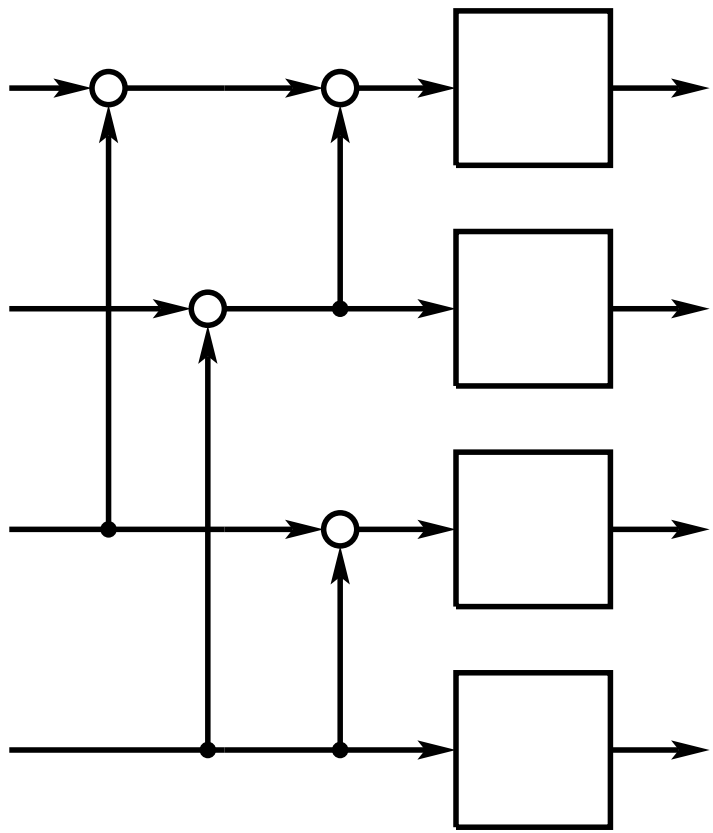
with equality iff W is a perfect channel, $I(W) = 1$, or a pure noise channel, $I(W) = 0$. Cutoff rates start polarizing:

$$R_0(W^-) \leq R_0(W) \leq R_0(W^+)$$

Recursive continuation

Do the same recursively: Given W,

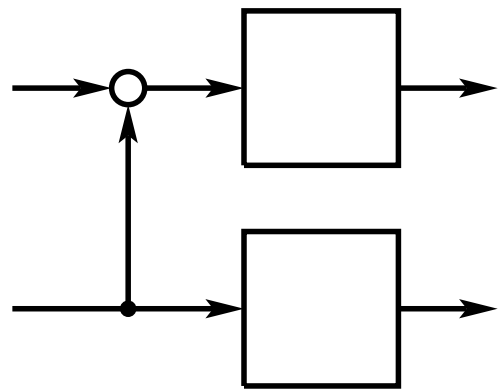
- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- ...



Recursive continuation

Do the same recursively: Given W,

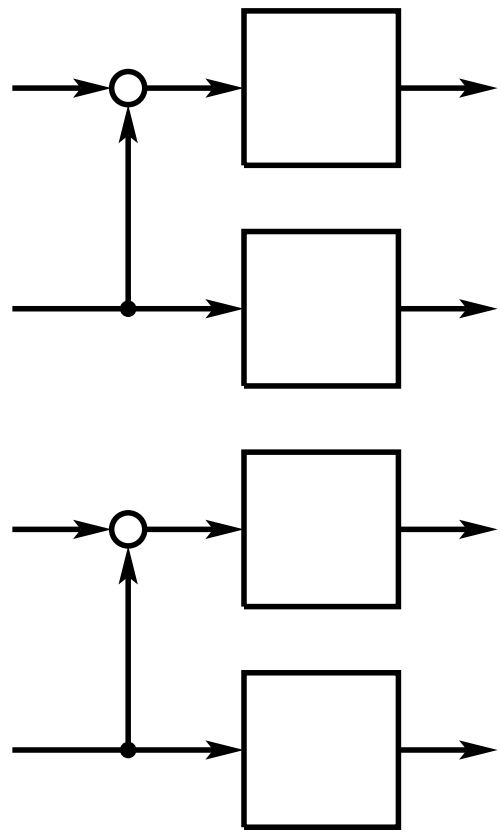
- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- ...



Recursive continuation

Do the same recursively: Given W,

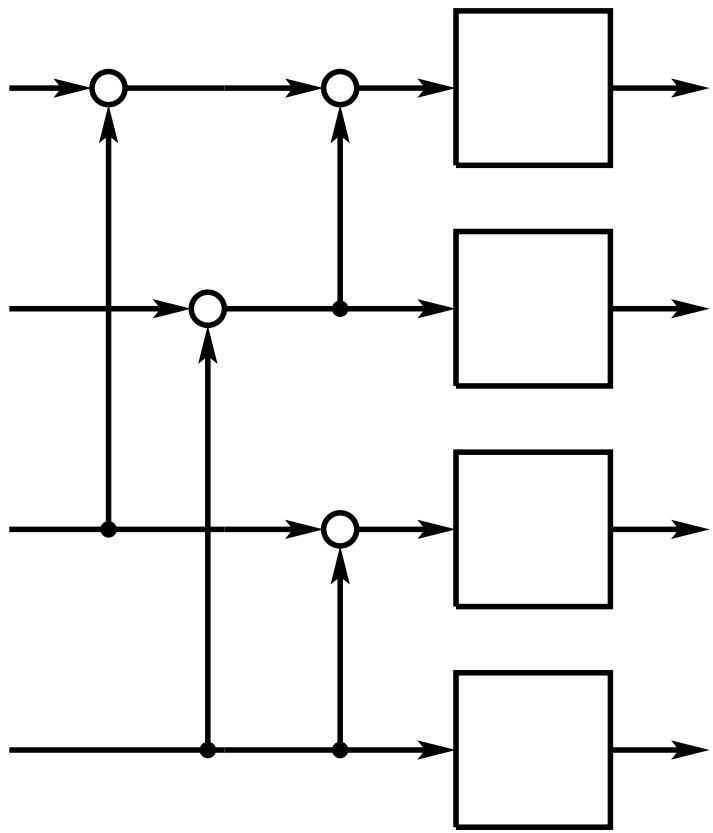
- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- ...



Recursive continuation

Do the same recursively: Given W,

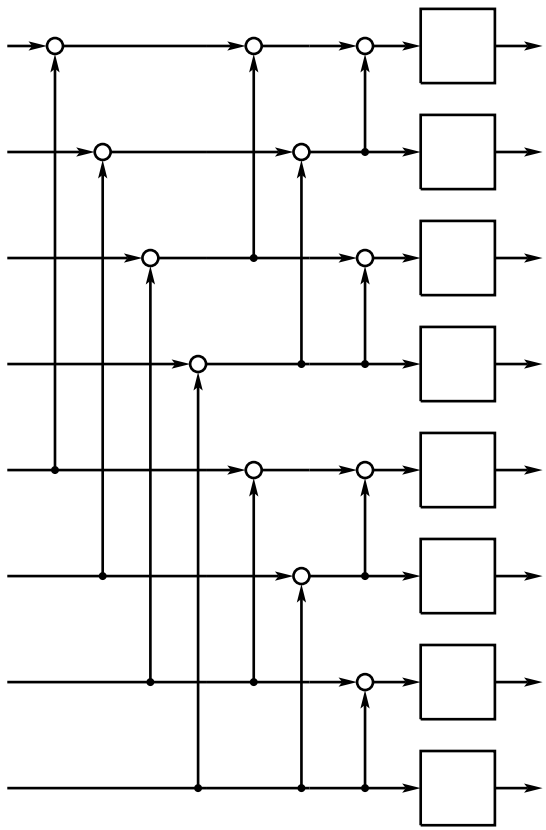
- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- ...



Recursive continuation

Do the same recursively: Given W,

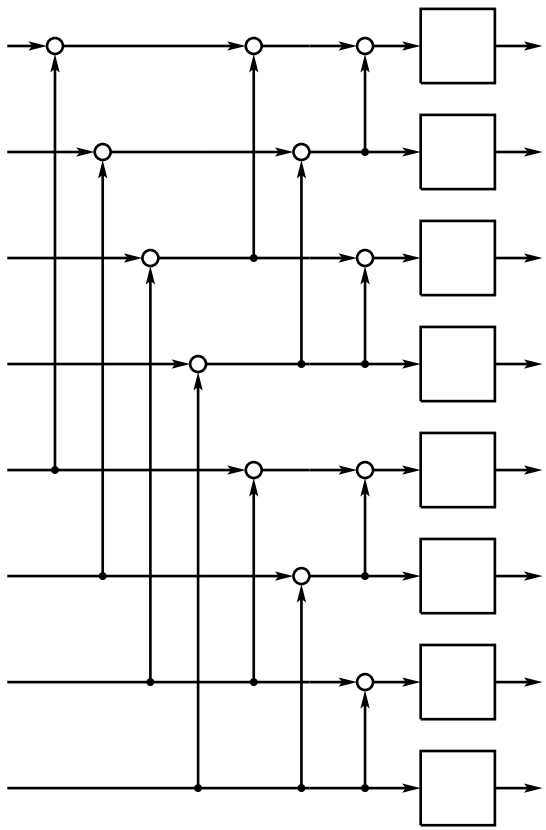
- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- ...



Recursive continuation

Do the same recursively: Given W,

- Duplicate W and obtain W^- and W^+.
- Duplicate W^- (W^+),
- and obtain W^{--} and W^{-+} (W^{+-} and W^{++}).
- Duplicate W^{--} (W^{-+}, W^{+-}, W^{++}) and obtain W^{---} and W^{--+} (W^{-+-}, W^{-++}, W^{+--}, W^{+-+}, W^{++-}, W^{+++}).
- ...



Polarization Process

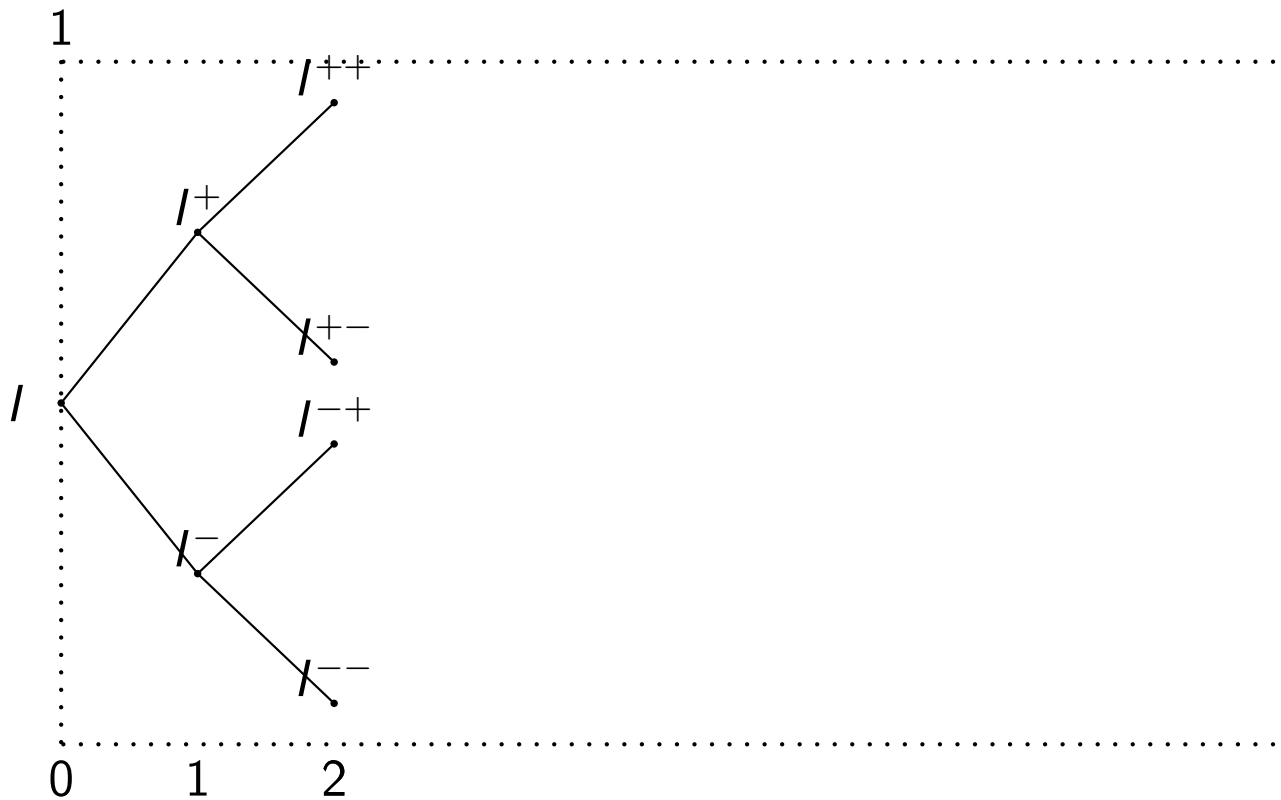
Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.

Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.

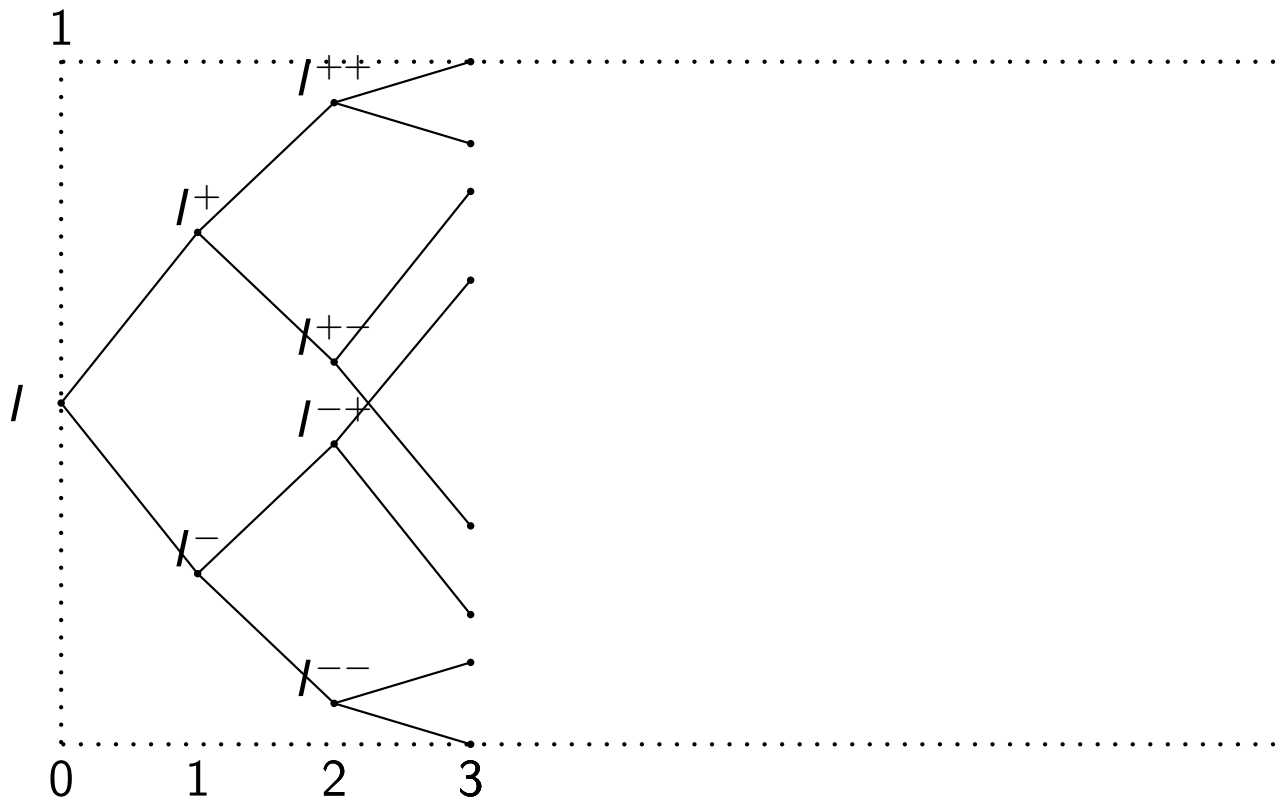
Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



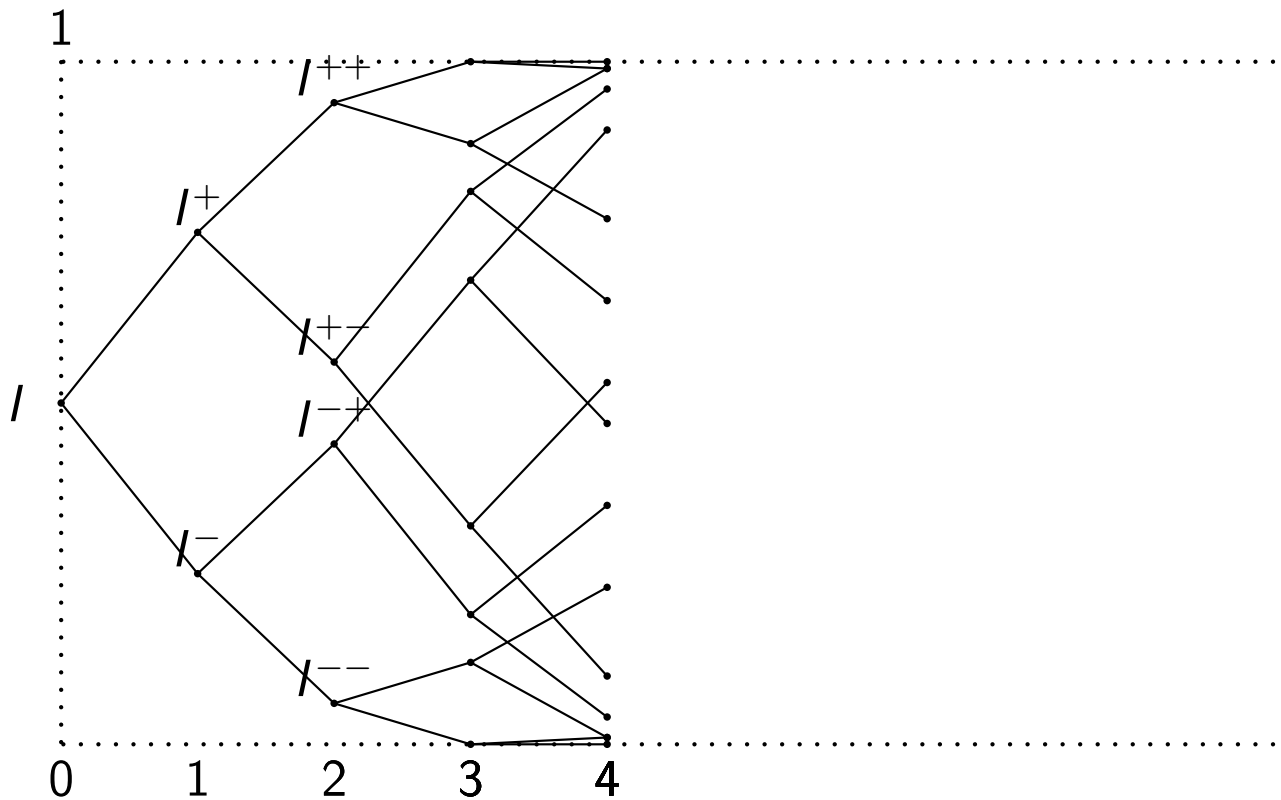
Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



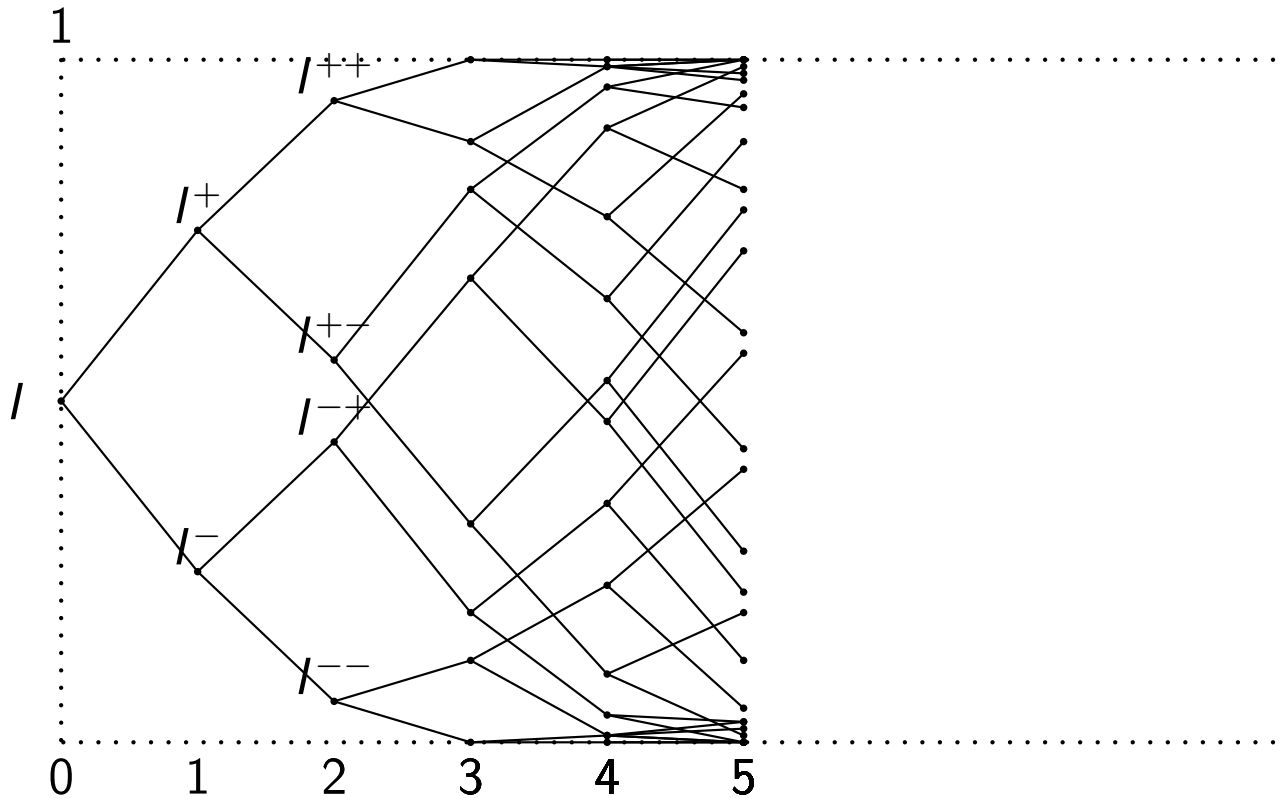
Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



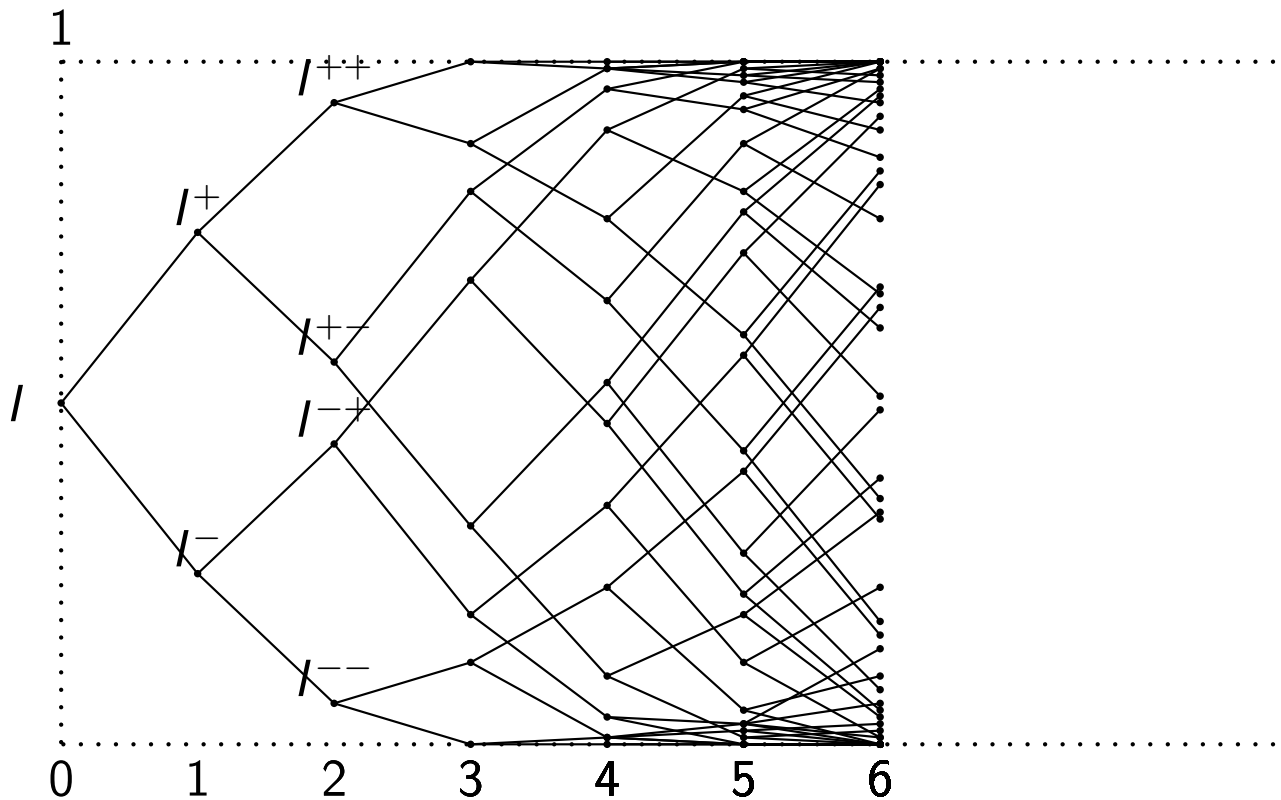
Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



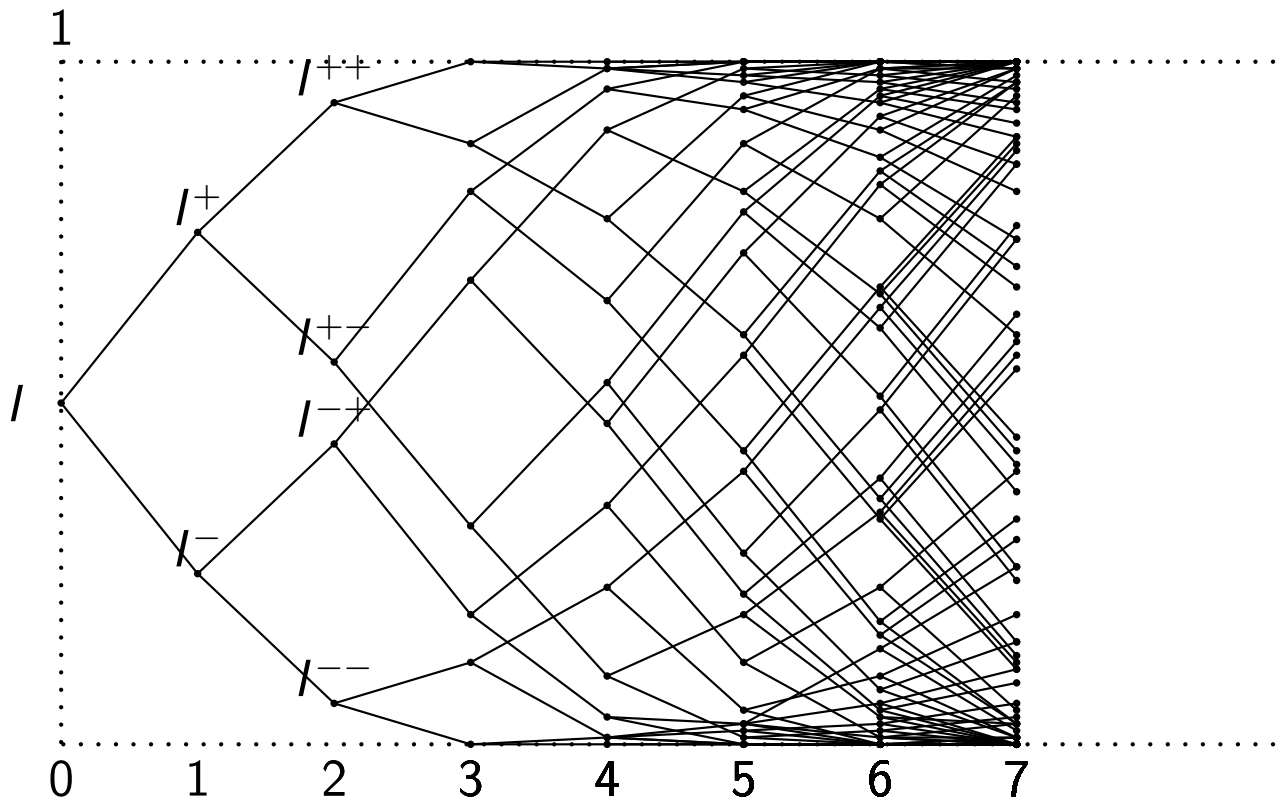
Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



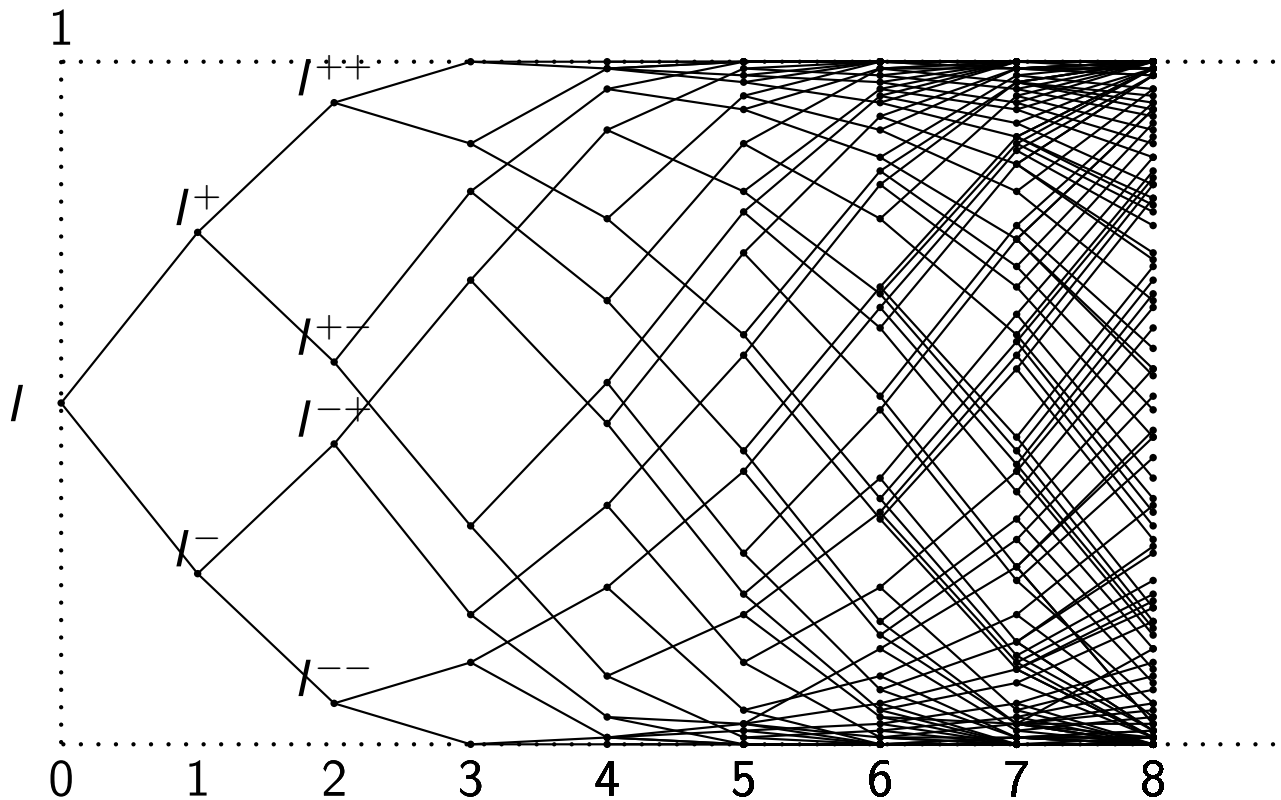
Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



Polarization Process

Evolution of $I = I(W)$, $I^+ = I(W^+)$, $I^- = I(W^-)$, etc.



Cutoff Rate Polarization

Theorem (2006)

The cutoff rates $\{R_0(U_i; Y^N U^{i-1})\}$ *of the channels created by the recursive transformation converge to their extremal values, i.e.,*

$$\frac{1}{N}\#\{i : R_0(U_i; Y^N U^{i-1}) \approx 1\} \to I(W)$$

and

$$\frac{1}{N}\#\{i : R_0(U_i; Y^N U^{i-1}) \approx 0\} \to 1 - I(W).$$

Remark: $\{I(U_i; Y^N U^{i-1})\}$ also polarize.

Cutoff Rate Polarization

Theorem (2006)

The cutoff rates $\{R_0(U_i; Y^N U^{i-1})\}$ *of the channels created by the recursive transformation converge to their extremal values, i.e.,*

$$\frac{1}{N}\#\left\{i : R_0(U_i; Y^N U^{i-1}) \approx 1\right\} \to I(W)$$

and

$$\frac{1}{N}\#\left\{i : R_0(U_i; Y^N U^{i-1}) \approx 0\right\} \to 1 - I(W).$$

Remark: $\{I(U_i; Y^N U^{i-1})\}$ also polarize.

Cutoff Rate Polarization

Theorem (2006)

The cutoff rates $\{R_0(U_i; Y^N U^{i-1})\}$ *of the channels created by the recursive transformation converge to their extremal values, i.e.,*

$$\frac{1}{N}\#\left\{i : R_0(U_i; Y^N U^{i-1}) \approx 1\right\} \to I(W)$$

and

$$\frac{1}{N}\#\left\{i : R_0(U_i; Y^N U^{i-1}) \approx 0\right\} \to 1 - I(W).$$

Remark: $\{I(U_i; Y^N U^{i-1})\}$ also polarize.

Cutoff Rate Polarization

Theorem (2006)

The cutoff rates $\{R_0(U_i; Y^N U^{i-1})\}$ *of the channels created by the recursive transformation converge to their extremal values, i.e.,*

$$\frac{1}{N}\#\left\{i : R_0(U_i; Y^N U^{i-1}) \approx 1\right\} \to I(W)$$

and

$$\frac{1}{N}\#\left\{i : R_0(U_i; Y^N U^{i-1}) \approx 0\right\} \to 1 - I(W).$$

Remark: $\{I(U_i; Y^N U^{i-1})\}$ also polarize.

Sequential decoding with successive cancellation

- Use the recursive construction to generate N bit-channels with cutoff rates $R_0(U_i; Y^N U^{i-1})$, $1 \leq i \leq N$.
- Encode the bit-channels independently using convolutional coding
- Decode the bit-channels one by one using sequential decoding and successive cancellation
- Achievable sum cutoff rate is

$$\sum_{i=1}^{N} R_0(U_i; Y^N U^{i-1})$$

which approaches $N\,I(W)$ as N increases.

Final step: Doing away with sequential decoding

- Due to polarization, rate loss is negligible if one does not use the "bad" bit-channels
- Rate of polarization is strong enough that a vanishing frame error rate can be achieved even if the "good" bit-channels are used uncoded
- The resulting system has no convolutional encoding and sequential decoding, only successive cancellation decoding

Polar coding

To communicate at rate $R < I(W)$:

- Pick N, and $K = NR$ good indices i such that $I(U_i; Y^N U^{i-1})$ is high,
- let the transmitter set U_i to be uncoded binary data for good indices, and set U_i to random but publicly known values for the rest,
- let the receiver decode the U_i successively: U_1 from Y^N; U_i from $Y^N \hat{U}^{i-1}$.

Polar coding

To communicate at rate $R < I(W)$:

- Pick N, and $K = NR$ good indices i such that $I(U_i; Y^N U^{i-1})$ is high,
- let the transmitter set U_i to be uncoded binary data for good indices, and set U_i to random but publicly known values for the rest,
- let the receiver decode the U_i successively: U_1 from Y^N; U_i from $Y^N \hat{U}^{i-1}$.

Polar coding

To communicate at rate $R < I(W)$:

- Pick N, and $K = NR$ good indices i such that $I(U_i; Y^N U^{i-1})$ is high,
- let the transmitter set U_i to be uncoded binary data for good indices, and set U_i to random but publicly known values for the rest,
- let the receiver decode the U_i successively: U_1 from Y^N; U_i from $Y^N \hat{U}^{i-1}$.

Polar coding

To communicate at rate $R < I(W)$:

- Pick N, and $K = NR$ good indices i such that $I(U_i; Y^N U^{i-1})$ is high,
- let the transmitter set U_i to be uncoded binary data for good indices, and set U_i to random but publicly known values for the rest,
- let the receiver decode the U_i successively: U_1 from Y^N; U_i from $Y^N \hat{U}^{i-1}$.

Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding, polar codes achieve the capacity $I(W)$ with

- *encoding complexity $N \log N$,*
- *decoding complexity $N \log N$,*
- *and probability of frame error better than $2^{-N^{0.49}}$*

Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding, polar codes achieve the capacity $I(W)$ with

- *encoding complexity $N \log N$,*
- *decoding complexity $N \log N$,*
- *and probability of frame error better than $2^{-N^{0.49}}$*

Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding, polar codes achieve the capacity $I(W)$ with

- *encoding complexity $N \log N$,*
- *decoding complexity $N \log N$,*
- *and probability of frame error better than $2^{-N^{0.49}}$*

Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding, polar codes achieve the capacity $I(W)$ with

- *encoding complexity $N \log N$,*
- *decoding complexity $N \log N$,*
- *and probability of frame error better than $2^{-N^{0.49}}$*

Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding, polar codes achieve the capacity $I(W)$ with

- *encoding complexity $N \log N$,*
- *decoding complexity $N \log N$,*
- *and probability of frame error better than $2^{-N^{0.49}}$*

Next lecture

- Details of the construction, encoding and decoding algorithms
- Survey of important results about polar codes
- Potential for applications

Next lecture

- Details of the construction, encoding and decoding algorithms
- Survey of important results about polar codes
- Potential for applications

Next lecture

- Details of the construction, encoding and decoding algorithms
- Survey of important results about polar codes
- Potential for applications

Thank you!