From Sequential Decoding to Polar Codes

Erdal Arıkan

Electrical-Electronics Engineering Department
Bilkent University
Ankara, Turkey

6th International Symposium on Turbo Codes & Iterative Information Processing
6-10 September 2010
Brest, France
Goals

- Show how polar coding originated from attempts to boost the cutoff rate of sequential decoding
- In particular, discuss the papers:
  - Pinsker (1965) “On the complexity of decoding”
  - Massey (1981) “Capacity, cutoff rate, and coding for a direct-detection optical channel”
Goals

- Show how polar coding originated from attempts to boost the cutoff rate of sequential decoding

- In particular, discuss the papers
  - Pinsker (1965) “On the complexity of decoding”
  - Massey (1981) “Capacity, cutoff rate, and coding for a direct-detection optical channel”
Show how polar coding originated from attempts to boost the cutoff rate of sequential decoding

In particular, discuss the papers

- Pinsker (1965) “On the complexity of decoding”
- Massey (1981) “Capacity, cutoff rate, and coding for a direct-detection optical channel”
Goals

- Show how polar coding originated from attempts to boost the cutoff rate of sequential decoding
- In particular, discuss the papers
  - Pinsker (1965) “On the complexity of decoding”
  - Massey (1981) “Capacity, cutoff rate, and coding for a direct-detection optical channel”
Outline

- Searching
  - Sequential decoding
  - Pinsker’s scheme
  - Massey’s scheme
  - Polarization
Outline

- Searching
  - Sequential decoding
    - Pinsker's scheme
    - Massey's scheme
    - Polarization
Outline

- Searching
- Sequential decoding
  - Pinsker’s scheme
- Massey’s scheme
- Polarization
### Outline

- Searching
- Sequential decoding
- Pinsker’s scheme
- **Massey’s scheme**
- Polarization
Outline

- Searching
- Sequential decoding
- Pinsker’s scheme
- Massey’s scheme
- Polarization
Find $X$ after observing $Y$

Two types of search:
- Binary: “Is $X$ in the set $S$?”
- Pointwise: “Is $X$ equal to $x$?”

Binary search leads to notions of entropy, mutual information, channel capacity.

Pointwise search leads to notions of Renyi entropy, cutoff rate.
Find $X$ after observing $Y$

Two types of search
- Binary: “Is $X$ in the set $S$?”
- Pointwise: “Is $X$ equal to $x$?”

Binary search leads to notions of entropy, mutual information, channel capacity

Pointwise search leads to notions of Renyi entropy, cutoff rate
Find $X$ after observing $Y$

Two types of search

- Binary: “Is $X$ in the set $S$?”
  - Pointwise: “Is $X$ equal to $x$?”

Binary search leads to notions of entropy, mutual information, channel capacity

Pointwise search leads to notions of Renyi entropy, cutoff rate
Search

Find $X$ after observing $Y$

Two types of search
- Binary: “Is $X$ in the set $S$?”
- Pointwise: “Is $X$ equal to $x$?”

Binary search leads to notions of entropy, mutual information, channel capacity

Pointwise search leads to notions of Renyi entropy, cutoff rate
Search

- Find $X$ after observing $Y$
- Two types of search
  - Binary: “Is $X$ in the set $S$?”
  - Pointwise: “Is $X$ equal to $x$?”

- Binary search leads to notions of entropy, mutual information, channel capacity
- Pointwise search leads to notions of Renyi entropy, cutoff rate
Find $X$ after observing $Y$

Two types of search
- Binary: “Is $X$ in the set $S$?”
- Pointwise: “Is $X$ equal to $x$?”

Binary search leads to notions of entropy, mutual information, channel capacity

Pointwise search leads to notions of Renyi entropy, cutoff rate
**Search complexity**

- **Binary search: Is \( X \in S? \)**
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- **Pointwise search: Is \( X = x? \)**
  - Each ‘NO’ answer reduces the size of the search space by \( \frac{1}{2} \)
  - No of queries may be very high — there is a “cutoff” phenomenon
Search complexity

- Binary search: Is $X \in S$?
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- Pointwise search: Is $X = x$?
  - Each ‘NO’ answer reduces the size of the search space by 1
  - No of queries may be very high — there is a “cutoff” phenomenon
Search complexity

- **Binary search: Is** $X \in S$?
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- **Pointwise search: Is** $X = x$?
  - Each ‘NO’ answer reduces the size of the search space by 1
  - No of queries may be very high — there is a “cutoff” phenomenon
Search complexity

- **Binary search**: Is $X \in S$?
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- **Pointwise search**: Is $X = x$?
  - Each ‘NO’ answer reduces the size of the search space by 1
  - No of queries may be very high — there is a “cutoff” phenomenon
Search complexity

- **Binary search:** Is $X \in S$?
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- **Pointwise search:** Is $X = x$?
  - Each ‘NO’ answer reduces the size of the search space by 1
  - No of queries may be very high — there is a “cutoff” phenomenon
Search complexity

- Binary search: Is $X \in S$?
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- Pointwise search: Is $X = x$?
  - Each ‘NO’ answer reduces the size of the search space by 1
  - No of queries may be very high — there is a “cutoff” phenomenon
Search complexity

- **Binary search**: Is $X \in S$?
  - Each ‘NO’ answer may halve the search space
  - No of queries is hardly an issue
  - Who answers the questions at what cost is a different question

- **Pointwise search**: Is $X = x$?
  - Each ‘NO’ answer reduces the size of the search space by 1
  - No of queries may be very high — there is a “cutoff” phenomenon
Pointwise search: Cutoff phenomenon

\((X, Y) \sim P_{X,Y}\) with \(X\) uniform on \(\{1, \ldots, M\}\) and

\[
Y = \begin{cases} 
X & \text{with probability } 1 - \epsilon \\
? & \text{with probability } \epsilon
\end{cases}
\]

Let \(G_{X|Y}\) be the number of questions asked until finding \(X\)

\[
E[G_{X|Y}] = (1 - \epsilon) \cdot 1 + \epsilon \cdot (M/2)
\]

Search complexity is \(\mathcal{O}(1)\) if \(M = o(1/\epsilon)\)

For higher-order \(M\), there is a “complexity cutoff”
(X, Y) \sim P_{X,Y} \text{ with } X \text{ uniform on } \{1, \ldots, M\} \text{ and }\nabla
\begin{align*}
Y &= \begin{cases} 
X & \text{with probability } 1 - \epsilon \\
? & \text{with probability } \epsilon
\end{cases}
\end{align*}

Let G_{X|Y} be the no of questions asked until finding X

\begin{align*}
E[G_{X|Y}] &= (1 - \epsilon) \cdot 1 + \epsilon \cdot (M/2) \\
\text{Search complexity is } &O(1) \text{ if } M = o(1/\epsilon) \\
\text{For higher-order } M, \text{ there is a "complexity cutoff"}
\end{align*}
(X, Y) \sim P_{X,Y} \text{ with } X \text{ uniform on } \{1, \ldots, M\} \text{ and }

Y = \begin{cases} 
X & \text{with probability } 1 - \epsilon \\
? & \text{with probability } \epsilon 
\end{cases}

Let \( G_{X|Y} \) be the no of questions asked until finding \( X \)

\[ E[G_{X|Y}] = (1 - \epsilon) \cdot 1 + \epsilon \cdot (M/2) \]

Search complexity is \( \mathcal{O}(1) \) if \( M = o(1/\epsilon) \)

For higher-order \( M \), there is a “complexity cutoff”
Pointwise search: Cutoff phenomenon

- \((X, Y) \sim P_{X,Y}\) with \(X\) uniform on \(\{1,\ldots,M\}\) and
  
  \[
  Y = \begin{cases}
    X & \text{with probability } 1 - \epsilon \\
    ? & \text{with probability } \epsilon
  \end{cases}
  \]

- Let \(G_{X|Y}\) be the no of questions asked until finding \(X\)

- \(E[G_{X|Y}] = (1 - \epsilon) \cdot 1 + \epsilon \cdot (M/2)\)

- Search complexity is \(O(1)\) if \(M = o(1/\epsilon)\)

- For higher-order \(M\), there is a “complexity cutoff”
**Pointwise search: Cutoff phenomenon**

- $(X, Y) \sim P_{X,Y}$ with $X$ uniform on $\{1, \ldots, M\}$ and
  
  $$
  Y = \begin{cases} 
  X & \text{with probability } 1 - \epsilon \\
  ? & \text{with probability } \epsilon 
  \end{cases}
  $$

- Let $G_{X|Y}$ be the no of questions asked until finding $X$
  
  $$
  E[G_{X|Y}] = (1 - \epsilon) \cdot 1 + \epsilon \cdot (M/2)
  $$

- Search complexity is $O(1)$ if $M = o(1/\epsilon)$

- For higher-order $M$, there is a “complexity cutoff”
Convolutional codes were invented by P. Elias (1955)

- Sequential decoding by J. M. Wozencraft (1957)
- SD enjoyed popularity in 1960s
- First coding system in space
- Viterbi algorithm (1967)
- SD lost ground to Viterbi algorithm in 1970s and never recovered
Convolutional codes were invented by P. Elias (1955)

Sequential decoding by J. M. Wozencraft (1957)


SD enjoyed popularity in 1960s

First coding system in space

Viterbi algorithm (1967)

SD lost ground to Viterbi algorithm in 1970s and never recovered
Convolutional codes, Sequential decoding, ...

- Convolutional codes were invented by P. Elias (1955)
- Sequential decoding by J. M. Wozencraft (1957)
- SD enjoyed popularity in 1960s
- First coding system in space
- Viterbi algorithm (1967)
- SD lost ground to Viterbi algorithm in 1970s and never recovered
Convolutional codes were invented by P. Elias (1955)
Sequential decoding by J. M. Wozencraft (1957)
SD enjoyed popularity in 1960s
First coding system in space
Viterbi algorithm (1967)
SD lost ground to Viterbi algorithm in 1970s and never recovered
Convolutional codes, Sequential decoding, ...

- Convolutional codes were invented by P. Elias (1955)
- Sequential decoding by J. M. Wozencraft (1957)
- SD enjoyed popularity in 1960s
- **First coding system in space**
  - Viterbi algorithm (1967)
  - SD lost ground to Viterbi algorithm in 1970s and never recovered
Convolutional codes were invented by P. Elias (1955)
Sequential decoding by J. M. Wozencraft (1957)
SD enjoyed popularity in 1960s
First coding system in space
Viterbi algorithm (1967)

SD lost ground to Viterbi algorithm in 1970s and never recovered
Convolutional codes, Sequential decoding, ...

- Convolutional codes were invented by P. Elias (1955)
- Sequential decoding by J. M. Wozencraft (1957)
- SD enjoyed popularity in 1960s
- First coding system in space
- Viterbi algorithm (1967)
- SD lost ground to Viterbi algorithm in 1970s and never recovered
Sequential decoding: the algorithm

SD is a search algorithm for the correct path in a tree code
Sequential decoding: the metric

Sequential decoding uses a “metric” to distinguish the correct path from the incorrect ones. Fano’s metric is given by:

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

- $\Gamma$: path length
- $y^n$: candidate path
- $x^n$: received sequence
- $R$: code rate
Sequential decoding: the 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 \( R \)
Sequential decoding: the 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: \( R \)
Sequential decoding: the cutoff rate

- SD achieves arbitrarily reliable communication at constant average complexity per bit at rates below a (computational) cutoff rate $R_{\text{comp}}$

- For a channel with transition probabilities $W(y|x)$, $R_{\text{comp}}$ equals

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

- Achievability: Wozencraft (1957), Reiffen (1962), Fano (1963), Stiglitz and Yudkin (1964)
- Converse: Jacobs and Berlekamp (1967)
Sequential decoding: the cutoff rate

- SD achieves \textit{arbitrarily} reliable communication at constant average complexity per bit at rates below a (computational) cutoff rate $R_{\text{comp}}$.
- For a channel with transition probabilities $W(y|x)$, $R_{\text{comp}}$ equals

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

- Converse: Jacobs and Berlekamp (1967).
Sequential decoding: the cutoff rate

- SD achieves **arbitrarily** reliable communication at **constant** average complexity per bit at rates below a (computational) cutoff rate $R_{\text{comp}}$.

- For a channel with transition probabilities $W(y|x)$, $R_{\text{comp}}$ equals

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

- **Achievability:** Wozencraft (1957), Reiffen (1962), Fano (1963), Stiglitz and Yudkin (1964)

- **Converse:** Jacobs and Berlekamp (1967)

Sequential decoding: the cutoff rate

- SD achieves **arbitrarily** reliable communication at constant average complexity per bit at rates below a (computational) cutoff rate $R_{\text{comp}}$.
- For a channel with transition probabilities $W(y|x)$, $R_{\text{comp}}$ equals

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

- Achievability: Wozencraft (1957), Reiffen (1962), Fano (1963), Stiglitz and Yudkin (1964)
- **Converse**: Jacobs and Berlekamp (1967)
Sequential decoding: the cutoff rate

- SD achieves **arbitrarily** reliable communication at constant average complexity per bit at rates below a (computational) cutoff rate $R_{\text{comp}}$
- For a channel with transition probabilities $W(y|x)$, $R_{\text{comp}}$ equals

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

- Achievability: Wozencraft (1957), Reiffen (1962), Fano (1963), Stiglitz and Yudkin (1964)
- Converse: Jacobs and Berlekamp (1967)
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order.
- Forgets what it saw beyond level $N$ upon backtracking.
- Let $G_N$ be the number of nodes searched (visited) at level $N$ until the correct node is found.
- Let $R$ be the code rate.
- There exist codes such that $E[G_N] \lesssim 1 + 2^{-N(R_0 - R)}$.
- For any code of rate $R$, $E[G_N] \leq 1 + 2^{-N(R_0 - R)}$. 

\[\frac{45}{143}\]
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order
- Let $G_N$ be the number of nodes searched (visited) at level $N$ until correct node is found.
- Let $R$ be the code rate.
- There exist codes s.t. $E[G_N] \leq 1 + 2^{-N(R_0 - R)}$.
- For any code of rate $R$, $E[G_N] \gtrsim 1 + 2^{-N(R_0 - R)}$.
- Forgets what it saw beyond level $N$ upon backtracking.
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order

  ![Code tree diagram]

- Forgets what it saw beyond level $N$ upon backtracking

  

Let $G_N$ be the number of nodes searched (visited) at level $N$ until the correct node is found.

- There exist codes such that

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

- For any code of rate $R$,
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order
- Forgets what it saw beyond level $N$ upon backtracking

Let $G_N$ be the number of nodes searched (visited) at level $N$ until the correct node is found.

There exist codes such that

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

For any code of rate $R$,

$$E[G_N] \geq 1 + 2^{-N(R_0 - R)}$$
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order
- Forgets what it saw beyond level $N$ upon backtracking
- Let $G_N$ be the number of nodes searched (visited) at level $N$ until correct node is found
- Let $R$ be the code rate
- There exist codes s.t.
  \[ E[G_N] \leq 1 + 2^{-N(R_0 - R)} \]
- For any code of rate $R$,
  \[ E[G_N] \gtrsim 1 + 2^{-N(R_0 - R)} \]
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order

- Let $G_N$ be the number of nodes searched (visited) at level $N$ until correct node is found

- Let $R$ be the code rate

- There exist codes s.t.

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

- For any code of rate $R$,

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

- Forgets what it saw beyond level $N$ upon backtracking
**Rules of the game: pointwise, no “look-ahead”**

- SD visits nodes at level \( N \) in a certain order
- Let \( G_N \) be the number of nodes searched (visited) at level \( N \) until correct node is found
- Let \( R \) be the code rate
  - There exist codes s.t.
    \[
    E[G_N] \leq 1 + 2^{-N(R_0-R)}
    \]
  - For any code of rate \( R \),
    \[
    E[G_N] \geq 1 + 2^{-N(R_0-R)}
    \]
- Forgets what it saw beyond level \( N \) upon backtracking
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order
- Forgets what it saw beyond level $N$ upon backtracking
- Let $G_N$ be the number of nodes searched (visited) at level $N$ until correct node is found
- Let $R$ be the code rate
- There exist codes s.t.
  \[
  E[G_N] \leq 1 + 2^{-N(R_0 - R)}
  \]
- For any code of rate $R$,
  \[
  E[G_N] \geq 1 + 2^{-N(R_0 - R)}
  \]
Rules of the game: pointwise, no “look-ahead”

- SD visits nodes at level $N$ in a certain order.
- Forgets what it saw beyond level $N$ upon backtracking.
- Let $G_N$ be the number of nodes searched (visited) at level $N$ until correct node is found.
- Let $R$ be the code rate.
- There exist codes s.t.
  \[ E[G_N] \leq 1 + 2^{-N(R_0 - R)} \]
- For any code of rate $R$,
  \[ E[G_N] \gtrsim 1 + 2^{-N(R_0 - R)} \]
Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$

- $R_0$ is a fundamental limit if one follows the rules of the game:
  - Single searcher
  - No look-ahead

- To boost the cutoff rate, change one or both of these rules
  - Use multiple sequential decoders
  - Provide look-ahead
Boosting the cutoff rate

- Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$
- $R_0$ is a fundamental limit if one follows the rules of the game:
  - Single searcher
  - No look-ahead
- To boost the cutoff rate, change one or both of these rules:
  - Use multiple sequential decoders
  - Provide look-ahead
Boosting the cutoff rate

- Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$
- $R_0$ is a fundamental limit if one follows the rules of the game:
  - Single searcher
  - No look-ahead

- To boost the cutoff rate, change one or both of these rules
  - Use multiple sequential decoders
  - Provide look-ahead
Boosting the cutoff rate

- Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$
- $R_0$ is a fundamental limit if one follows the rules of the game:
  - Single searcher
  - No look-ahead

- To boost the cutoff rate, change one or both of these rules:
  - Use multiple sequential decoders
  - Provide look-ahead
Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$

$R_0$ is a fundamental limit if one follows the rules of the game:
- Single searcher
- No look-ahead

To boost the cutoff rate, change one or both of these rules
- Use multiple sequential decoders
- Provide look-ahead
Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$.

$R_0$ is a fundamental limit if one follows the rules of the game:
- Single searcher
- No look-ahead

To boost the cutoff rate, change one or both of these rules:
- Use multiple sequential decoders
- Provide look-ahead
Goal: Finding SD schemes with $R_{\text{comp}}$ larger than $R_0$

$R_0$ is a fundamental limit if one follows the rules of the game:
- Single searcher
- No look-ahead

To boost the cutoff rate, change one or both of these rules
- Use multiple sequential decoders
- Provide look-ahead
Pinseker’s scheme (1965)

- Block coding just below capacity: $K/N \approx C(W)$
- $N$ large, block error rate small: $P_e \sim 2^{-O(N)}$
- Each SD sees a memoryless BSC with $R_0$ near 1
- Boosts the cutoff rate to capacity
Pinssker’s scheme (1965)

- Block coding just below capacity: $K/N \approx C(W)$
- $N$ large, block error rate small: $P_e \sim 2^{-O(N)}$
- Each SD sees a memoryless BSC with $R_0$ near 1
- Boosts the cutoff rate to capacity
Pinsker’s scheme (1965)

- Block coding just below capacity: $K/N \approx C(W)$
- $N$ large, block error rate small: $P_e \sim 2^{-O(N)}$
- Each SD sees a memoryless BSC with $R_0$ near 1
- Boosts the cutoff rate to capacity
Block coding just below capacity: $K/N \approx C(W)$

$N$ large, block error rate small: $P_e \sim 2^{-\mathcal{O}(N)}$

Each SD sees a memoryless BSC with $R_0$ near 1

Boosts the cutoff rate to capacity
A scheme that doesn’t work

No improvement in cutoff rate
Equivalent scheme

Derived (Vector) Channel

Cutoff rate = $R_0$(Derived vector channel)
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
Other attempts to boost the cutoff rate

- D. Falconer, 1966.
- P. R. Chevillat and D. J. Costello Jr., 1977.
- J. L. Massey, 1981.
- J. Belzile and D. Haccoun, 1993.
- ...
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
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

\[ C(QEC) = 2(1 - \epsilon) \]
\[ R_0(QEC) = \log \frac{4}{1 + 3\epsilon} \]

Independent coding of BECs

\[ C(BEC) = (1 - \epsilon) \]
\[ R_0(BEC) = \log \frac{2}{1 + \epsilon} \]
Introduction  Searching  Sequential decoding  Pinsker’s scheme  Massey’s scheme  Polarization

Capacity, cutoff rate for one QEC vs two BECs

**Ordinary coding of QEC**

\[ C(QEC) = 2(1 - \epsilon) \]
\[ R_0(QEC) = \log \frac{4}{1 + 3\epsilon} \]

- \( C(QEC) = 2 \times C(BEC) \)

**Independent coding of BECs**

\[ C(BEC) = (1 - \epsilon) \]
\[ R_0(BEC) = \log \frac{2}{1 + \epsilon} \]
Capacity, cutoff rate for one QEC vs two BECs

**Ordinary coding of QEC**

\[ C(QEC) = 2(1 - \epsilon) \]

\[ R_0(QEC) = \log \frac{4}{1 + 3\epsilon} \]

**Independent coding of BECs**

\[ C(BEC) = (1 - \epsilon) \]

\[ R_0(BEC) = \log \frac{2}{1 + \epsilon} \]

- \[ C(QEC) = 2 \times C(BEC) \]
- \[ R_0(QEC) \leq 2 \times R_0(BEC) \] with equality iff \( \epsilon = 0 \) or 1.
Cutoff rate improvement by splitting

- Cutoff rate of QEC
- Cutoff rate of BEC
- Sum cutoff rate after splitting
- Capacity of QEC

Erasure probability $\varepsilon$
Capacity, cutoff rate (bits)
Why does Massey’s scheme work?

- Why do we have \( 2R_0(\text{BEC}) \geq R_0(\text{QEC}) \)?
- Let \( G_N \) denote the number of guesses at level \( N \) until finding the correct node.
- Joint decoder has quadratic complexity.

\[
G_N(\text{QEC}) = G_N(\text{BEC}_1) G_N(\text{BEC}_2)
\]
\[= G_N(\text{BEC}_1)^2 \quad \text{correlated erasures}
\]

- Thus,

\[
E[G_N(\text{QEC})] = E[G_N(\text{BEC}_1)^2] \geq (E[G_N(\text{BEC}_1)])^2
\]

- Second moment of \( G_N(\text{BEC}) \) becomes exponentially large at a rate below \( R_0(\text{BEC}) \).
Why does Massey’s scheme work?

- Why do we have $2 R_0(\text{BEC}) \geq R_0(\text{QEC})$?
- Let $G_N$ denote the number of guesses at level $N$ until finding the correct node.

  Joint decoder has quadratic complexity

  $$G_N(\text{QEC}) = G_N(\text{BEC}_1) \cdot G_N(\text{BEC}_2)$$
  $$= G_N(\text{BEC}_1)^2 \quad \text{correlated erasures}$$

  Thus,

  $$E[G_N(\text{QEC})] = E[G_N(\text{BEC}_1)^2] \geq (E[G_N(\text{BEC}_1)])^2$$

  Second moment of $G_N(\text{BEC})$ becomes exponentially large at a rate below $R_0(\text{BEC})$. 
Why does Massey’s scheme work?

- Why do we have $2 R_0(\text{BEC}) \geq R_0(\text{QEC})$?
- Let $G_N$ denote the number of guesses at level $N$ until finding the correct node.
- Joint decoder has quadratic complexity.

$$G_N(\text{QEC}) = G_N(\text{BEC}_1) G_N(\text{BEC}_2)$$

$$= G_N(\text{BEC}_1)^2 \quad \text{correlated erasures}$$

- Thus,

$$E[G_N(\text{QEC})] = E[G_N(\text{BEC}_1)^2] \geq (E[G_N(\text{BEC}_1)])^2$$

- Second moment of $G_N(\text{BEC})$ becomes exponentially large at a rate below $R_0(\text{BEC})$. 
Why does Massey’s scheme work?

- Why do we have $2 R_0(\text{BEC}) \geq R_0(\text{QEC})$?
- Let $G_N$ denote the number of guesses at level $N$ until finding the correct node.
- Joint decoder has quadratic complexity

$$G_N(\text{QEC}) = G_N(\text{BEC}_1) \cdot G_N(\text{BEC}_2)$$

$$= G_N(\text{BEC}_1)^2 \quad \text{correlated erasures}$$

- Thus,

$$E[G_N(\text{QEC})] = E[G_N(\text{BEC}_1)^2] \geq (E[G_N(\text{BEC}_1)])^2$$

- Second moment of $G_N(\text{BEC})$ becomes exponentially large at a rate below $R_0(\text{BEC})$. 
Why does Massey’s scheme work?

- Why do we have $2 R_0(\text{BEC}) \geq R_0(\text{QEC})$?
- Let $G_N$ denote the number of guesses at level $N$ until finding the correct node.
- Joint decoder has quadratic complexity

$$G_N(\text{QEC}) = G_N(\text{BEC}_1) G_N(\text{BEC}_2)$$

$$= G_N(\text{BEC}_1)^2 \quad \text{correlated erasures}$$

- Thus,

$$E[G_N(\text{QEC})] = E[G_N(\text{BEC}_1)^2] \geq (E[G_N(\text{BEC}_1)])^2$$

- Second moment of $G_N(\text{BEC})$ becomes exponentially large at a rate below $R_0(\text{BEC})$. 
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
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
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
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**
Let $V : \mathbb{F}_2 \overset{\Delta}{=} \{0, 1\} \to \mathcal{Y}$ 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

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

The (symmetric) cutoff rate is defined as

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

Let $V : \mathbb{F}_2 \trianglerighteq \{0, 1\} \rightarrow \mathcal{Y}$ 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

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

The (symmetric) cutoff rate is defined as

$$R_0(V) \triangleq R_0(X; Y) \triangleq -\log \left[ \sum_{y \in \mathcal{Y}} \left( \sum_{x \in \mathbb{F}_2} \frac{1}{2} \sqrt{V(y|x)} \right)^2 \right]$$
Let \( V : \mathbb{F}_2 \triangleq \{0, 1\} \rightarrow \mathcal{Y} \) 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

\[
I(V) \triangleq I(X; Y) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathbb{F}_2} \frac{1}{2} V(y|x) \log \frac{V(y|x)}{\frac{1}{2} V(y|0) + \frac{1}{2} V(y|1)}
\]

The (symmetric) cutoff rate is defined as

\[
R_0(V) \triangleq R_0(X; Y) \triangleq - \log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathbb{F}_2} \frac{1}{2} \sqrt{V(y|x)} \right]^2
\]
Notation

- Let \( V : \mathbb{F}_2 \xrightarrow{\Delta} \{0, 1\} \rightarrow \mathcal{Y} \) 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

\[
I(V) \triangleq I(X; Y) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathbb{F}_2} \frac{1}{2} V(y|x) \log \frac{V(y|x)}{\frac{1}{2} V(y|0) + \frac{1}{2} V(y|1)}
\]

- The (symmetric) cutoff rate is defined as

\[
R_0(V) \triangleq R_0(X; Y) \triangleq - \log \sum_{y \in \mathcal{Y}} \left[ \sum_{x \in \mathbb{F}_2} \frac{1}{2} \sqrt{V(y|x)} \right]^2
\]
The basic construction

Given two copies of a binary input channel $W : \mathbb{F}_2 \triangleq \{0, 1\} \rightarrow \mathcal{Y}$

consider the transformation above to generate two channels $W^- : \mathbb{F}_2 \rightarrow \mathcal{Y}^2$ and $W^+ : \mathbb{F}_2 \rightarrow \mathcal{Y}^2 \times \mathbb{F}_2$ with

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

$$W^+(y_1 y_2 u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 + u_2) W(y_2 | u_2)$$
The basic construction

Given two copies of a binary input channel \( W : \mathbb{F}_2 \triangleq \{0, 1\} \rightarrow \mathcal{Y} \)

\[
W : \mathbb{F}_2 \rightarrow \mathcal{Y}
\]

\[
U_1 \quad W \quad Y_1
\]

\[
U_2 \quad W \quad Y_2
\]

consider the transformation above to generate two channels

\( W^- : F_2 \rightarrow \mathcal{Y}^2 \) and \( W^+ : F_2 \rightarrow \mathcal{Y}^2 \times F_2 \) with

\[
W^-(y_1 y_2 | u_1) = \sum_{u_2} \frac{1}{2} W(y_1 | u_1 + u_2) W(y_2 | u_2)
\]

\[
W^+(y_1 y_2 u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 + u_2) W(y_2 | u_2)
\]
The basic construction

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

Consider the transformation above to generate two channels

$W^- : \mathbb{F}_2 \rightarrow \mathcal{Y}^2$ and $W^+ : \mathbb{F}_2 \rightarrow \mathcal{Y}^2 \times \mathbb{F}_2$ with

$$W^-(y_1y_2|u_1) = \sum_{u_2} \frac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2)$$

$$W^+(y_1y_2u_1|u_2) = \frac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2)$$
The basic construction

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

\[
\begin{align*}
U_1 &\quad \oplus \quad W \quad \rightarrow \quad Y_1 \\
U_2 &\quad \rightarrow \quad W \quad \rightarrow \quad Y_2
\end{align*}
\]

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

\[
W^-(y_1, y_2 | u_1) = \sum_{u_2} \frac{1}{2} W(y_1 | u_1 + u_2) W(y_2 | u_2)
\]

\[
W^+(y_1, y_2, u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 + u_2) W(y_2 | u_2)
\]
The basic construction

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

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

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

$$W^+(y_1 y_2 u_1 | u_2) = \frac{1}{2} W(y_1 | u_1 + u_2) W(y_2 | u_2)$$
The basic construction

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

Consider the transformation above to generate two channels $W^- : \mathbb{F}_2 \to \mathcal{Y}^2$ and $W^+ : \mathbb{F}_2 \to \mathcal{Y}^2 \times \mathbb{F}_2$ with

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

$$W^+(y_1 y_2 u_1|u_2) = \frac{1}{2} W(y_1|u_1 + u_2) W(y_2|u_2)$$
Sequential decoding with successive cancellation

\[ (Y_1, Y_2) + U_2 \]

\[ m_1 \rightarrow E1 \rightarrow U_1 \rightarrow X_1 \rightarrow W \rightarrow Y_1 \]

\[ m_2 \rightarrow E2 \rightarrow U_2 \rightarrow X_2 \rightarrow W \rightarrow Y_2 \]

\[ \hat{m}_1 \rightarrow SD1 \]

\[ \hat{m}_2 \rightarrow SD2 \]
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,

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

- and $I(W^-) \leq I(W) \leq I(W^+)$.
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,
  \[
  I(W^-) + I(W^+) = I(U_1 U_2; Y_1 Y_2) \\
  = 2I(W),
  \]

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

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

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

[Diagram of 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 $W^{--}$ ($W^{+-}$, $W^{++}$) and obtain $W^{---}$ and $W^{---}$ ($W^{+++}$, $W^{+++}$, $W^{+++}$, $W^{+++}$).
- ...
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 $W^{---}$
- 
  ...
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 $W^{---}$ and $W^{---}$ ($W^{---}$, $W^{+++}$, $W^{+++}$, $W^{+++}$, $W^{+++}$).
- ...
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.

\[ I = I(W), \quad I^+ = I(W^+), \quad I^- = I(W^-), \quad \text{etc.} \]
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.

\[
\begin{align*}
&I, I^+, I^{++}, I^{+-}, I^{--}, \\
&I^-, I^{+-}, I^{--}, \ldots
\end{align*}
\]
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.
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.,

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

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.,

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

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.
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.
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.
Sequential decoding with successive cancellation

- Use the recursive construction to generate $N$ bit-channels with cutoff rates $R_0(U_i; Y^NU^{i-1})$, $1 \leq i \leq N$.
- **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
  \[
  \sum_{i=1}^{N} R_0(U_i; Y^NU^{i-1})
  \]

which approaches $NI(W)$ as $N$ increases.
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.
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
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
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.
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 $R < I(W)$:

1. Pick $N$, and $K = NR$ good indices $i$ such that $I(U_i; Y^N U^{i-1})$ is high,
2. 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,
3. let the receiver decode the $U_i$ successively: $U_1$ from $Y^N$; $U_i$ from $Y^N \hat{U}^{i-1}$.
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 $Y^N \hat{U}^{i-1}$.
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 $Y^N$; $U_j$ from $Y^N \hat{U}^{i-1}$.
Theorem (2007)

*With the particular one-to-one mapping described here and with the successive cancellation decoding*

- polarization codes are ‘$I(W)$ achieving’,
- encoding complexity is $N \log N$,
- decoding complexity is $N \log N$,
- probability of error decays like $2^{-\sqrt{N}}$ (with E. Telatar, 2008).
Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding

- polarization codes are ‘\(I(W)\) achieving’,
- encoding complexity is \(N \log N\),
- decoding complexity is \(N \log N\),
- probability of error decays like \(2^{-\sqrt{N}}\) (with E. Telatar, 2008).
Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding

- polarization codes are ‘$I(W)$ achieving’,
- encoding complexity is $N \log N$,
- decoding complexity is $N \log N$,
- probability of error decays like $2^{-\sqrt{N}}$ (with E. Telatar, 2008).
Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the **successive cancellation decoding**

- polarization codes are ‘$I(W)$ achieving’,
- encoding complexity is $N \log N$,
- decoding complexity is $N \log N$,
- probability of error decays like $2^{-\sqrt{N}}$ (with E. Telatar, 2008).
Polar coding complexity and performance

Theorem (2007)

With the particular one-to-one mapping described here and with the successive cancellation decoding

- polarization codes are ‘\( I(W) \) achieving’,
- encoding complexity is \( N \log N \),
- decoding complexity is \( N \log N \),
- probability of error decays like \( 2^{-\sqrt{N}} \) (with E. Telatar, 2008).
Acknowledgments
Acknowledgments
Acknowledgments