Decoding of Polar Codes with Reinforcement Learning

Nghia Doan¹, Seyyed Ali Hashemi², and Warren Gross¹

¹McGill University, Québec, Canada
²Stanford University, California, USA

IEEE GLOBECOM
Taipei, Taiwan
Dec 9, 2020
Polar codes: selected for the eMBB control channel in 5G

Cyclic redundancy check (CRC) is concatenated with polar codes in 5G for error detection

Belief Propagation (BP): reasonable error-correction performance, highly parallel
Background

- Polar codes: selected for the eMBB control channel in 5G
- Cyclic redundancy check (CRC) is concatenated with polar codes in 5G for error detection
- Belief Propagation (BP): reasonable error-correction performance, **highly parallel** → **high decoding throughput**
Polar codes

- Introduced by Arıkan in 2009
- $\mathcal{P}(N, K)$, $N$: code length, $K$: message length
- Code construction: based on polarization phenomenon
  - $K$ most reliable channels: information bits
  - $(N - K)$ least reliable channels: frozen bits

$\mathcal{P}(8, 5)$ with $u_0$, $u_1$, and $u_2$ are frozen bits

---

CRC-Aided BP (CABP) Decoder of Polar Codes

▶ Exploit the extrinsic information of the CRC-polar factor graphs
  ▶ Perform BP decoding on the polar factor graph $l_{th}$ iterations
  ▶ Unsatisfied CRC test: run BP decoding on the CRC graph

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph $I_{th}$ iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph $l_{th}$ iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

▶ Exploit the extrinsic information of the CRC-polar factor graphs
▶ Perform BP decoding on the polar factor graph $l_{th}$ iterations
▶ Unsatisfied CRC test: run BP decoding on the CRC graph

Exploit the extrinsic information of the CRC-polar factor graphs

- Perform BP decoding on the polar factor graph $l_{th}$ iterations
- Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph $l_{th}$ iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph $l_{th}$ iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph \( i_{th} \) iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph $l_{th}$ iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph \( I_{th} \) iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph $i_{th}$ iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

CRC-Aided BP (CABP) Decoder of Polar Codes

- Exploit the extrinsic information of the CRC-polar factor graphs
  - Perform BP decoding on the polar factor graph \( l_{th} \) iterations
  - Unsatisfied CRC test: run BP decoding on the CRC graph

---

Polar Decoding with Permuted Factor Graphs

- Permuting the PE layers of the polar code factor graph does not change the code
- The error probability of BP decoding can be improved by using different factor graph permutations

Random factor-graph permutations of polar codes
Polar Decoding with Permuted Factor Graphs

- Permuting the PE layers of the polar code factor graph does not change the code.
- The error probability of BP decoding can be improved by using different factor graph permutations.
- Open problem: Under a specific channel output, select a factor graph that results in a correct codeword.

Random factor-graph permutations of polar codes
Previous Work

- Cyclic factor-graph permutations [a]
- Random factor-graph permutations [b]
- Monte Carlo based methods [c-e]

---


Previous Work

- Cyclic factor-graph permutations [a]
- Random factor-graph permutations [b]
- Monte Carlo based methods [c-e] → require a pre-construction of the factor graphs

---


Given a channel output $y$, select a set of $T$ "good" factor-graph permutations for CABP during the decoding.
Given a channel output $y$, select a set of $T$ ”good” factor-graph permutations for CABP during the decoding

Formalize the factor-graph permutation selection as a multi-armed bandit problem
Decoding of Polar Codes with Reinforcement Learning

- Given a channel output $y$, select a set of $T$ "good" factor-graph permutations for CABP during the decoding.

- Formalize the factor-graph permutation selection as a multi-armed bandit problem → use state-of-the-art bandit algorithms to solve the problem.
Generate $k$ random actions (sets of permutations)

Each action contains the original permutation and a set of $T - 1$ random permutations

The decoder selects a set of permutations (an action) to perform CABP decoding
Decoding of Polar Codes with Reinforcement Learning

Definitions:

- $a_j$: an action index (permutation set index)
- $n_{a_j}$: the number of times that action $a_j$ has been selected
- $v_{a_j}$: the estimated value of selecting action $a_j$ (how good action $a_j$ is)
- $\alpha_j, \beta_j$: a pair of shape parameters for a Beta distribution in $[0, 1]$ associated with action $a_j$, initially $\alpha_j = \beta_j = 1$ for all $j$
Decoding of Polar Codes with Reinforcement Learning

- Definitions:
  - $a_j$: an action index (permutation set index)

- $(\alpha_j, \beta_j)$: a pair of shape parameters for a Beta distribution in $[0, 1]$ associated with action $a_j$, initially $\alpha_j = \beta_j = 1 \forall j$
Definitions:

- $a_j$: an action index (permutation set index)
- $n_{a_j}$: the number of times that action $a_j$ has been selected
Decoding of Polar Codes with Reinforcement Learning

- Definitions:
  - $a_j$: an action index (permutation set index)
  - $n_{a_j}$: the number of times that action $a_j$ has been selected
  - $\nu_{a_j}$: the estimated value of selecting action $a_j$ (how good action $a_j$ is)
  - $(\alpha_{a_j}, \beta_{a_j})$: a pair of shape parameters for a Beta distribution in $[0, 1]$ associated with action $a_j$, initially $\alpha_{a_j} = \beta_{a_j} = 1$ ∀ $j$
Definitions:

- $a_j$: an action index (permutation set index)
- $n_{a_j}$: the number of times that action $a_j$ has been selected
- $v_{a_j}$: the estimated value of selecting action $a_j$ (how good action $a_j$ is)
- $(\alpha_j, \beta_j)$: a pair of shape parameters for a Beta distribution in $[0, 1]$ associated with action $a_j$, initially $\alpha_j = \beta_j = 1 \forall j$
Decoding of Polar Codes with Reinforcement Learning

Definitions:

- \( a_j \): an action index (permutation set index)
- \( n_{a_j} \): the number of times that action \( a_j \) has been selected
- \( v_{a_j} \): the estimated value of selecting action \( a_j \) (how good action \( a_j \) is)
- \((\alpha_j, \beta_j)\): a pair of shape parameters for a Beta distribution in \([0, 1]\) associated with action \( a_j \), initially \( \alpha_j = \beta_j = 1 \ \forall j \)
Decoding of Polar Codes with Reinforcement Learning

▷ Action selection:
Decoding of Polar Codes with Reinforcement Learning

- Action selection:
  - $\epsilon - Greedy$

  $$a^* = \begin{cases} \
  \arg \max_{a_j} v_{a_j} & \text{with prob. } 1 - \epsilon \text{ (exploitation)} \\
  a_{\text{random}} & \text{with prob. } \epsilon \text{ (exploration)} 
  \end{cases}$$
Decoding of Polar Codes with Reinforcement Learning

- Action selection:
  - $\epsilon - Greedy$
    $$a^* = \begin{cases} 
    \arg\max_{a_j} v_{a_j} & \text{with prob. } 1 - \epsilon \text{ (exploitation)} \\
    a_{\text{random}} & \text{with prob. } \epsilon \text{ (exploration)}
    \end{cases}$$

- Upper Confidence Bound (UCB)
  $$a^* = \arg\max_{a_j} \left[ v_{a_j} + c \sqrt{\frac{\ln t}{n_{a_j}}} \right]$$
  - exploitation
  - exploration
Action selection:

\[ a^* = \arg \max_{\forall j} \lambda(a_j) \]

The action selection can be obtained prior to the actual decoding.
Action selection:

Thompson Sampling (TS)

\[ \nu_{a_j} = \text{BetaDist}(\alpha_j, \beta_j) \]

An action \( k \) is then selected as \( a^* = \arg \max_{\forall j} \nu_{a_j} \)
Action selection:

- Thompson Sampling (TS)

\[ v_{a_j} = \text{BetaDist}(\alpha_j, \beta_j) \]

An action \( k \) is then selected as \( a^* = \arg \max_{\forall j} v_{a_j} \)

- The action selection can be obtained prior to the actual decoding.
Parameter update: 

- The event that taking action $a^*$ results in a CRC-satisfied codeword.

- $E_{a^*} = 1$ if $E_{a^*}$ occurs and $0$ otherwise.

- $\epsilon$-Greedy and UCB:
  
  - $v_{a^*} = v_{a^*} + 1 - E_{a^*} - n_{a^*}$

- Thompson Sampling (TS):
  
  - $\alpha_k = \alpha_k + 1 - E_{a_k}$
  
  - $\beta_k = \beta_k + 1 - E_{a_k}$
Decoding of Polar Codes with Reinforcement Learning

Parameter update:

- $E_{a^*}$: the event that taking action $a^*$ results in a CRC-satisfied codeword.
Parameter update:

- $E_{a^*}$: the event that taking action $a^*$ results in a CRC-satisfied codeword.

- $\mathbb{1}_{E_{a^*}} = 1$ if $E_{a^*}$ occurs and $\mathbb{1}_{E_{a^*}} = 0$, otherwise.
Parameter update:

$E_{a^*}$: the event that taking action $a^*$ results in a CRC-satisfied codeword.

$\mathbb{1}_{E_{a^*}} = 1$ if $E_{a^*}$ occurs and $\mathbb{1}_{E_{a^*}} = 0$, otherwise.

$\epsilon$-Greedy and UCB:

$$v_{a^*} := v_{a^*} + \frac{\mathbb{1}_{E_{a^*}} - v_{a^*}}{n_{a^*}}$$
Parameter update:

- $E_{a^*}$: the event that taking action $a^*$ results in a CRC-satisfied codeword.

- $\mathbb{1}_{E_{a^*}} = 1$ if $E_{a^*}$ occurs and $\mathbb{1}_{E_{a^*}} = 0$, otherwise.

- $\epsilon$-Greedy and UCB:

  $$v_{a^*} := v_{a^*} + \frac{\mathbb{1}_{E_{a^*}} - v_{a^*}}{n_{a^*}}$$

- Thompson Sampling (TS):

  $$\alpha_k := \alpha_k + 1 - \mathbb{1}_{E_{a_k}}$$
  $$\beta_k := \beta_k + \mathbb{1}_{E_{a_k}}$$
Performance comparison of various multi-armed bandit algorithms used by RL-CABP decoding. The simulation is obtained at $E_b/N_0 = 3.0$ dB with $k = 500$, $\varepsilon = 2^{-4}$, and $c = 2^{-3}$. 
Error-correction performance of different factor-graph permutation selection schemes for $\mathcal{P}(128, 64)$ with a 24-bit CRC used in 5G (24C).
Error-correction performance of RL-CABP decoding and other decoding algorithms of polar codes.
Propose an algorithm that selects the good factor-graph permutations during the course of decoding

Significantly reduce the error-probability of BP decoding while still maintain the parallelism property of BP decoding

The factor-graph selection can be pipelined with the decoding process
Thank You!