Data-Driven Polar Codes for Unknown Channels With and Without Memory

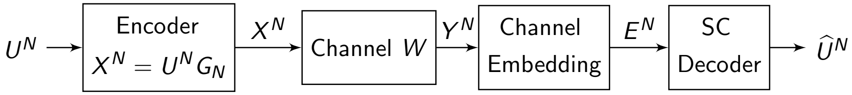
Ziv Aharoni[1], Bashar Huleihel[1], Henry D. Pfister[2], Haim H. Permuter[1]

[1]Ben-Gurion University of the Negev

[2]Duke University

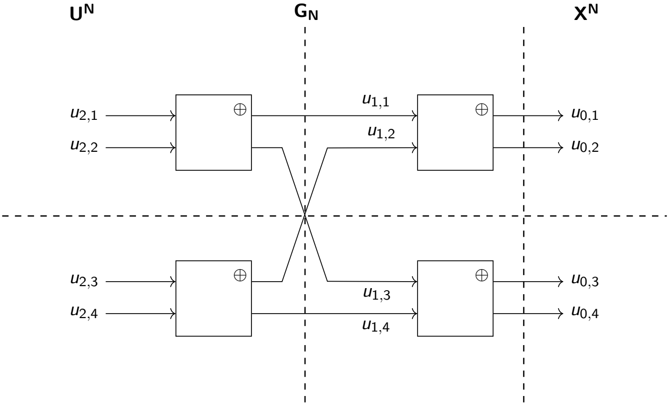
International Symposium on Information Theory
December 1, 2025

Polar Coding Scheme



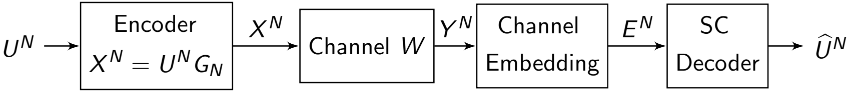
- Input alphabet $\mathcal{X} = \{0, 1\}$
- Block length $N = 2^n$, $n \in \mathbb{N}$
- Information set $\mathcal{A} \subset [N]$
- Arikan's transform G_N
- Output alphabet $\mathcal{Y}$
- Transition kernel W

Polar Encoder



- $u_{i,j}$ is the j-th bit in the i-th decoding stage, $i \in \{0\} \cup [\log_2(N)], j \in [N]$
- Stages $\log_2(N), 0$ corresponds to U^N, X^N, respectively

Successive Cancellation Decoder



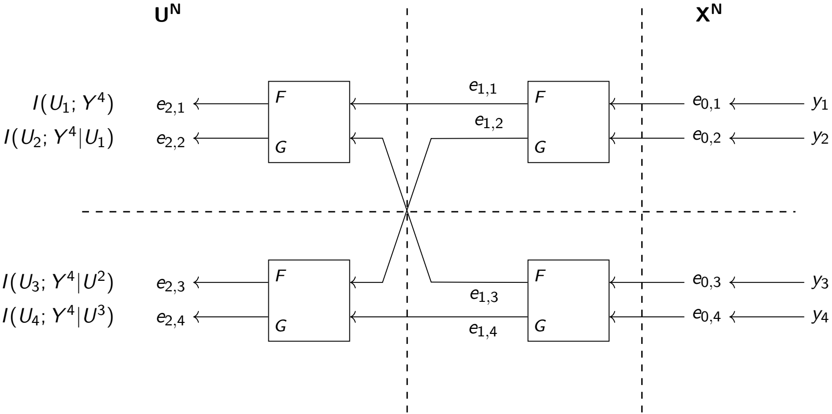
- For memoryless channels:

$$\begin{array}{ll} \text{Sufficient statistics:} & E(y) = \log \frac{W(1|y)}{W(0|y)} \\ \text{Check-node:} & F(e_1, e_2) = 2\tanh^{-1}\left(\tanh\frac{e_1}{2}\tanh\frac{e_2}{2}\right) \\ \text{Bit-node:} & G(e_1, e_2, u) = e_2 + (-1)^u e_1 \\ \text{Soft decision:} & H(e_1) = \sigma(e_1) \end{array}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the logistic function.

- Computational complexity $O(N \log N)$
- Finite state channels: SC trellis decoder (Wang et. al. 15')
Computational complexity $O(|\mathcal{S}|^3 N \log N)$

SC Decoder



Sufficient statistics:	$E(y) = \log \frac{W(1\|y)}{W(0\|y)}$
Check-node:	$F(e_1, e_2) = 2\tanh^{-1}\left(\tanh\frac{e_1}{2}\tanh\frac{e_2}{2}\right)$
Bit-node:	$G(e_1, e_2, u) = e_2 + (-1)^u e_1$
Soft decision:	$H(e_1) = \sigma(e_1)$

Data-Driven Polar Code for Memoryless Channels

Setting:

- Unknown memoryless channel W
- Uniform input distribution P_X
- Observations $\mathcal{D}_M = \{x_i, y_i\}_{i=1}^M \sim (P_X \otimes W)^{\otimes M}$

Problem:

- The channel LLRs E are unknown

Solution:

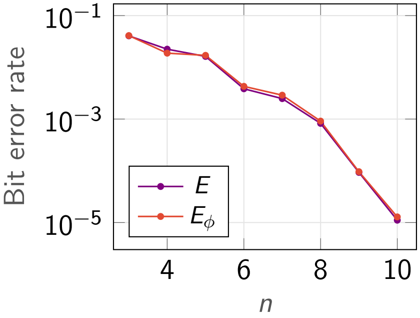
- Use the MINE Algorithm (Belghazi et. al. 18')
- The optimal solution of MINE:

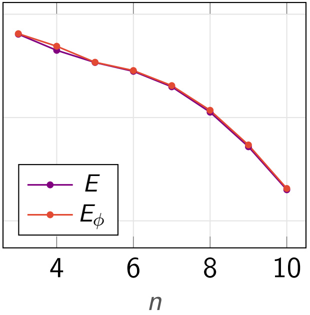
$$T^*(x,y) = \log \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)} + c$$

- Relation with the channel LLRs $E(y) = \log \frac{W(1|y)}{W(0|y)}$:

$$E_\phi(y) = T_\phi(1,y) - T_\phi(0,y)$$

Experiments - Memoryless Channels





BSC:

$$Y = \begin{cases} X & \text{w.p. } 0.9 \\ X \oplus 1 & \text{w.p. } 0.1 \end{cases}$$

AWGN:

$$Y = X + Z,$$
$$Z \sim \mathcal{N}(0, 0.5)$$

Data-Driven Polar Code for Channels with Memory

Setting:

- Unknown channel with memory $W_{Y^{MN}\|X^{MN}}$
- Uniform input distribution P_X
- Observations $\mathcal{D}_{MN} = \{x_i, y_i\}_{i=1}^{MN} \sim P_{X^{MN}} \otimes W_{Y^{MN}\|X^{MN}}$

Problem:

- The channel LLRs E are unknown
- SC decoder is unknown

Solution:

- Define channel embedding $E_\phi : \mathcal{Y} \to \mathbb{R}^d$
- Define a neural SC (NSC) decoder

Neural SC decoder

A NSC is defined by $F_{\theta_1}, G_{\theta_2}, H_{\theta_3}$ with parameters $\theta = \{\theta_1, \theta_2, \theta_3\}$. The functions satisfy:

- $F_\theta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the check-node NN.
- $G_\theta : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{X} \to \mathbb{R}^d$ is the bit-node NN.
- $H_\theta : \mathbb{R}^d \to [0, 1]$ is the soft decision NN.

- SC decoding with $E_\phi, F_\theta, G_\theta, H_\theta$ yields $P^{\phi,\theta}_{U_i|U^{i-1},Y^N}$, an estimate of $P_{U_i|U^{i-1},Y^N}, \quad i \in [N]$

Learning the NSC

- The mutual information of the effective bit channels

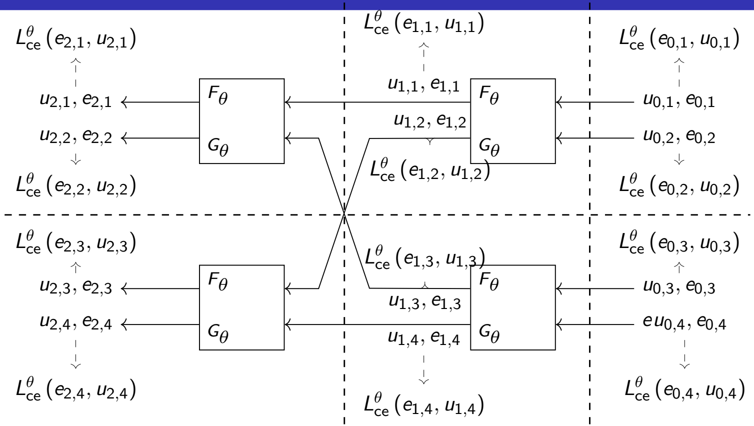
$$\begin{aligned} \mathrm{I}\left(U_i; Y^N | U^{i-1}\right) &= \mathrm{H}\left(U_i | U^{i-1}\right) - \mathrm{H}\left(U_i | U^{i-1}, Y^N\right) \\ &= 1 - \mathrm{H}\left(U_i | U^{i-1}, Y^N\right) \end{aligned}$$

  - $X_i \overset{iid}{\sim} \mathrm{Ber}(0.5) \Rightarrow U_i \overset{iid}{\sim} \mathrm{Ber}(0.5)$

- **The NSC objective:**
  - Let $\mathcal{D}_{M,N} \sim P_{X^{MN}} \otimes W_{Y^{MN} \| X^{MN}}$
  - $E_\phi, F_\theta, G_\theta, H_\theta$
  - Let $u_{i,j} = (x_{i,1}^N G_N)_j$. Then, for all $i \in [N]$

$$\mathrm{H}_{\Phi,\Theta}^{M}\left(U_i | U^{i-1}, Y^N\right) = \min_{\theta,\phi} \left\{ -\frac{1}{M} \sum_{j=1}^{M} \log P_{U_i | U^{i-1}, Y^N}^{\phi,\theta}\left(u_{j,i} | u_{j,1}^{i-1}, y_{j,1}^N\right) \right\}$$

Neural SC Decoder - Loss computation



- The complete loss is an average of all loss terms:

$$L^{\theta}_{ce}(e, u) = -u \log\left(\mathrm{H}_\theta(e)\right) - (1-u)\log\left(1 - \mathrm{H}_\theta(e)\right).$$

- We denote this computation by NSCTrain
- Computational complexity $O(dkN \log N)$

NSC Learning Algorithm

Algorithm Data-driven polar code for channels with memory

input: Dataset $\mathcal{D}_{M,N}$, block length n_{t}
output: Learnt NSC

Initiate the weights of $E_\phi, F_\theta, G_\theta, H_\theta$
$N = 2^{n_{\text{t}}}$
for $k = 1$ to $\mathrm{N}_{\text{iters}}$ **do**
  Sample x^N, y^N
  $u^N = x^N G_N$
  Compute e^N by $e_i = E_\phi\left(y_i\right)$
  Compute $\mathrm{L} = \mathrm{NSCTrain}\left(e^N, u^N, 0\right)$
  Minimize L w.r.t. $E_\phi, F_\theta, G_\theta, H_\theta$ via SGD
end for

Consistency

Theorem (Consistency for Time Invariant Channels)

Let $\mathbb{X}, \mathbb{Y}$ be jointly stationary and ergodic stochastic processes. Let $\mathcal{D}_{M,N} \sim P_{X^{MN}} \otimes W_{Y^{MN}\|X^{MN}}$, where $N = 2^n$, $M, n \in \mathbb{N}$. Let $u_{j,i} = (x_{j,1}^N G_N)_i$ and let $W_N^{(i)}, \forall i \in [N]$ be the effective bit channels. Then, for every $\varepsilon > 0$ there exist $E_\phi, F_\theta, G_\theta, H_\theta$ and $m \in \mathbb{N}$ such that for $M > m$ and $i \in [N]$

$$\left| \mathrm{H}_{\phi,\theta}^M \left(U_i | U^{i-1}, Y^N \right) - \mathrm{H} \left(U_i | U^{i-1}, Y^N \right) \right| < \varepsilon. \quad (1)$$

Consistency

Theorem (Consistency for Time Invariant Channels)

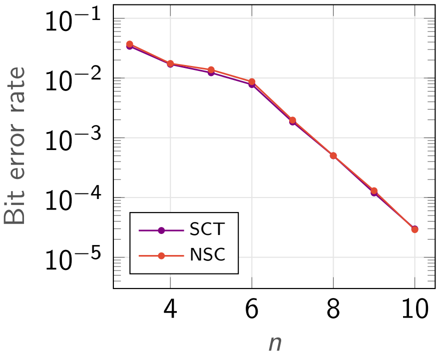
Let $\mathbb{X}, \mathbb{Y}$ be jointly stationary and ergodic stochastic processes. Let $\mathcal{D}_{M,N} \sim P_{X^{MN}} \otimes W_{Y^{MN} \| X^{MN}}$, where $N = 2^n$, $M, n \in \mathbb{N}$. Let $u_{j,i} = (x_{j,1}^N G_N)_i$ and let $W_N^{(i)}, \forall i \in [N]$ be the effective bit channels. Then, for every $\varepsilon > 0$ there exist $E_\phi, F_\theta, G_\theta, H_\theta$ and $m \in \mathbb{N}$ such that for $M > m$ and $i \in [N]$

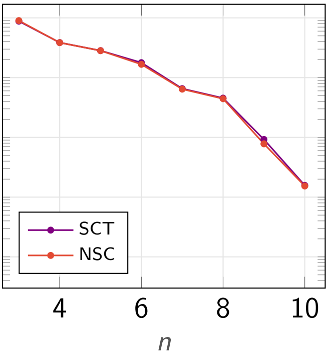
$$\left| \mathrm{H}_{\phi,\theta}^{M} \left(U_i | U^{i-1}, Y^N \right) - \mathrm{H} \left(U_i | U^{i-1}, Y^N \right) \right| < \varepsilon. \qquad (1)$$

Sketch of proof:

- Identifying the structure of $P_{U_i|U^{i-1},Y^N}$
- Universal approximation of $P_{U_i|U^{i-1},Y^N}$ by NNs
- Estimation of expectations with empirical means.

Experiments - Finite State Channels





Ising:

$$Y = \begin{cases} X & \text{w.p. } 0.5 \\ S & \text{w.p. } 0.5 \end{cases}$$

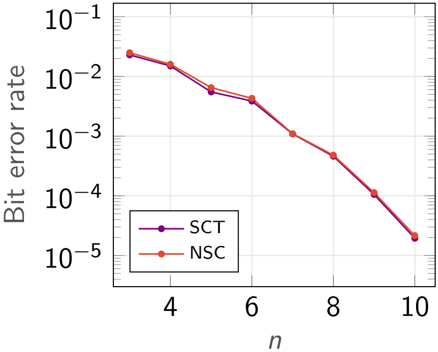
$$S' = X$$

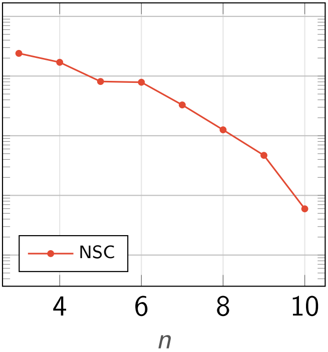
Trapdoor:

$$Y = \begin{cases} X & \text{w.p. } 0.5 \\ S & \text{w.p. } 0.5 \end{cases}$$

$$S' = X \oplus Y \oplus S$$

Experiments - Large Channel Memory





ISI:

$$Y_t = \sum_{i=0}^{m} h_i X_{t-i} + Z_i$$

$$Z_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

MA-AGN:

$$Y_t = X_t + \tilde{Z}_t$$

$$\tilde{Z}_t = Z_t + \alpha Z_{t-1}$$

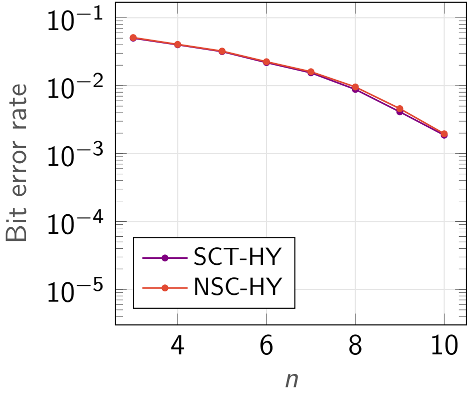
Extension to Stationary Input Distribution

- Let P_X be a stationary input distribution
- Now

$$\mathrm{I}\left(U_i; Y^N | U^{i-1}\right) = \mathrm{H}\left(U_i | U^{i-1}\right) - \mathrm{H}\left(U_i | U^{i-1}, Y^N\right)$$

- According to [HY13], apply NSC twice
  - First with $E_\phi^X(y) = e_X \in \mathbb{R}^d$ for all $y \in \mathcal{Y} \Rightarrow \mathrm{L}_X$
  - Second with E_ϕ as before $\Rightarrow \mathrm{L}_Y$
- Minimize $\mathrm{L}_X + \mathrm{L}_Y$ with respect to ϕ, θ

Experiments - Ising Channel with Stationary Input Distribution



Ising

Summary

- Data-driven polar code for channels with and without memory
  - For memoryless: estimate channel embedding via MINE
  - For memory: learnt embedding + NSC
  - Consistency
  - Computational Complexity $O(kdN \log N)$
- Extension to stationary inputs via Honda-Yamamoto scheme

Future work

- Extension to list decoding
- Applications to capacity estimation

Summary

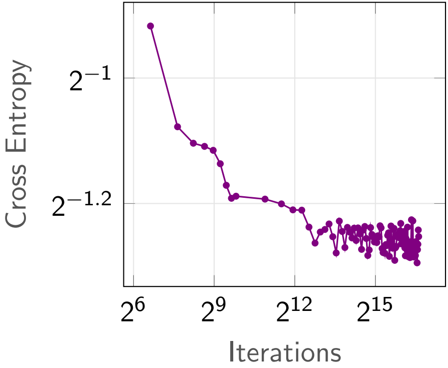
- Data-driven polar code for channels with and without memory
  - For memoryless: estimate channel embedding via MINE
  - For memory: learnt embedding + NSC
  - Consistency
  - Computational Complexity $O(kdN \log N)$
- Extension to stationary inputs via Honda-Yamamoto scheme

Future work

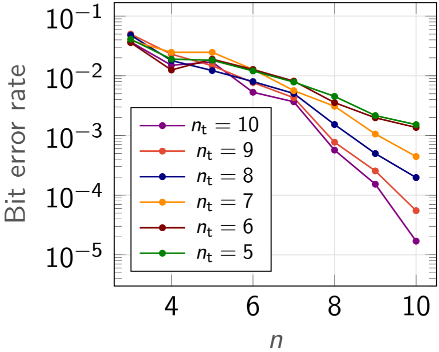
- Extension to list decoding
- Applications to capacity estimation

Thank You!

Experiments - NSC Learning



Convergence



Varying n_t