Polar Codes for the Deletion Channel: Weak and Strong Polarization

Ido Tal[1] Henry D. Pfister[2] Arman Fazeli[3] Alexander Vardy[3]

[1]Technion [2]Duke [3]UCSD

Big picture first

A polar coding scheme for the deletion channel where the:

- Deletion channel has *constant* deletion probability δ
- Fix a hidden-Markov input distribution[1].
- Code rate converges to information rate
- Error probability decays like $2^{-\Lambda^{\gamma}}$, where $\gamma < \frac{1}{3}$ and Λ is the codeword length
- Decoding complexity is at most $O(\Lambda^{1+3\gamma})$
- Achieves hidden-Markov capacity!

[1] i.e., a function of an aperiodic, irreducible, finite-state Markov chain

Big picture first

A polar coding scheme for the deletion channel where the:

- Deletion channel has *constant* deletion probability δ
- Fix a hidden-Markov input distribution[1].
- Code rate converges to information rate
- Error probability decays like $2^{-\Lambda^\gamma}$, where $\gamma < \frac{1}{3}$ and Λ is the codeword length
- Decoding complexity is at most $O(\Lambda^{1+3\gamma})$
- Achieves hidden-Markov capacity! Equals true capacity?

[1] i.e., a function of an aperiodic, irreducible, finite-state Markov chain

Big picture first

A polar coding scheme for the deletion channel where the:

- Deletion channel has *constant* deletion probability δ
- Fix a hidden-Markov input distribution[1].
- Code rate converges to information rate
- Error probability decays like $2^{-\Lambda^\gamma}$, where $\gamma < \frac{1}{3}$ and Λ is the codeword length
- Decoding complexity is at most $O(\Lambda^{1+3\gamma})$
- Achieves hidden-Markov capacity! Equals true capacity?
- Key ideas:
  - Polarization operations defined for trellises
  - Polar codes modified to have guard bands of '0' symbols

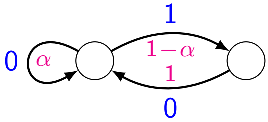
[1] i.e., a function of an aperiodic, irreducible, finite-state Markov chain

A brief history of the binary deletion channel

- Early Work: Levenshtein [Lev66] and Dobrushin [Dob67]
- LDPC Codes + Turbo Equalization: Davey-MacKay [DM01]
- Coding and Capacity Bounds by Mitzenmacher [Mit09] and many more: [FD10], [MTL12], [CK15], [RD15], [Che19]
- Polar codes: [TTVM17], [TFVL17], [TFV18]
- Our Contributions:
  - Proof of weak polarization for constant deletion rate
  - Strong polarization for constant deletion rate with guard bands
  - Our trellis perspective also establishes weak polarization for channels with insertions, deletions, and substitutions

Hidden-Markov input process

Example: $(1, \infty)$ Run-Length Constraint



- Input process is (X_j), $j \in \mathbb{Z}$
- Marginalization of (S_j, X_j), $j \in \mathbb{Z}$
- State (S_j), $j \in \mathbb{Z}$, is Markov, stationary, irreducible, aperiodic
- For all j, it holds that

$$P_{S_j, X_j | S_{-\infty}^{j-1}, X_{-\infty}^{j-1}} = P_{S_j, X_j | S_{j-1}}$$

Code rate

The code rate of our scheme approaches

$$\mathcal{I}(X;Y) = \lim_{N\to\infty} \frac{1}{N} H(\mathrm{X}) - \lim_{N\to\infty} \frac{1}{N} H(\mathrm{X}|\mathrm{Y}) ,$$

- $\mathrm{X} = (X_1, \ldots, X_N)$ is hidden-Markov input
- Y is the deletion channel output

Theorem (Strong polarization)

Fix a regular hidden-Markov input process. For any fixed $\gamma \in (0, 1/3)$, the rate of our coding scheme approaches the mutual-information rate between the input process and the deletion channel output. For large enough blocklength Λ, the probability of error is at most $2^{-\Lambda^\gamma}$.

Uniform input process

- It is known that a memoryless input distribution is suboptimal
- To keep this talk simple, we will however assume that the input process is uniform, and thus memoryless
- That is, the X_i are i.i.d. and $\mathrm{Ber}(1/2)$

The polar transform

- Let $\mathrm{x} = (x_1, \ldots, x_N) \in \{0,1\}^N$ be a vector of length $N = 2^n$
- Define
  - minus transform: $\mathrm{x}^{[0]} \triangleq (x_1 \oplus x_2, x_3 \oplus x_4, \ldots, x_{N-1} \oplus x_N)$
  - plus transform: $\mathrm{x}^{[1]} \triangleq (\quad x_2, \quad x_4, \ldots, \quad x_N)$
  - Both are vectors of length $N/2$
- Define $\mathrm{x}^{[b_1, b_2, \ldots, b_\lambda]}$ recursively:

$$\mathrm{z} = \mathrm{x}^{[b_1, b_2, \ldots, b_{\lambda-1}]}, \quad \mathrm{x}^{[b_1, b_2, \ldots, b_\lambda]} = \mathrm{z}^{[b_\lambda]}$$

- The polar transform of x is $\mathrm{u} = (u_1, u_2, \ldots, u_N)$, where for

$$i = 1 + \sum_{j=1}^{n} b_j 2^{n-j}$$

we have

$$u_i = \mathrm{x}^{[b_1, b_2, \ldots, b_n]}$$

Polarization of trellises

- The decoder sees the received sequence y
- Ultimately, we want an efficient method of calculating

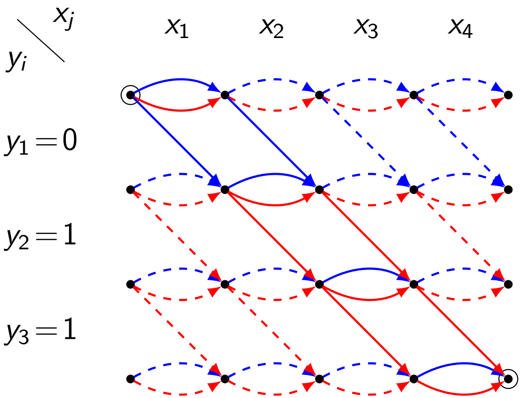
$$P(U_i = \hat{u}_i | \mathrm{U}^{i-1} = \hat{\mathrm{u}}^{i-1}, \mathrm{Y} = \mathrm{y})$$

- Towards this end, let us first show an efficient method of calculating the joint probability

$$P(\mathrm{X} = \mathrm{x}, \mathrm{Y} = \mathrm{y})$$

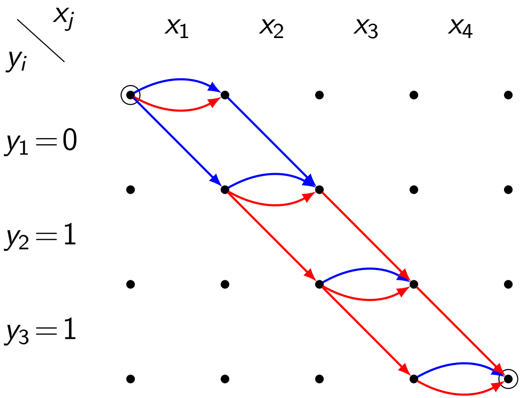
  - Generalizes the SC trellis decoder of Wang et. al. [WLH14], and the polar decoder for deletions by Tian et. al. [TFVL17]

Deletion channel trellis



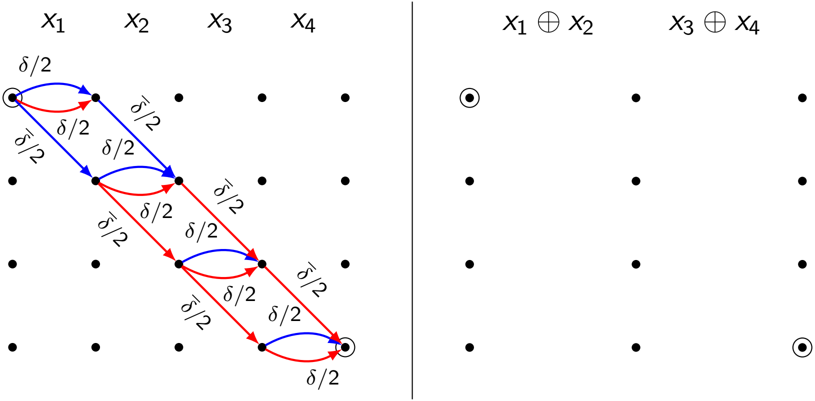
- Example: $N = 4$ inputs with length-3 output 011
- Edge labels: blue $x_j = 0$ and red $x_j = 1$
- Direction: diagonal = no deletion and horizontal = deletion

Deletion channel trellis



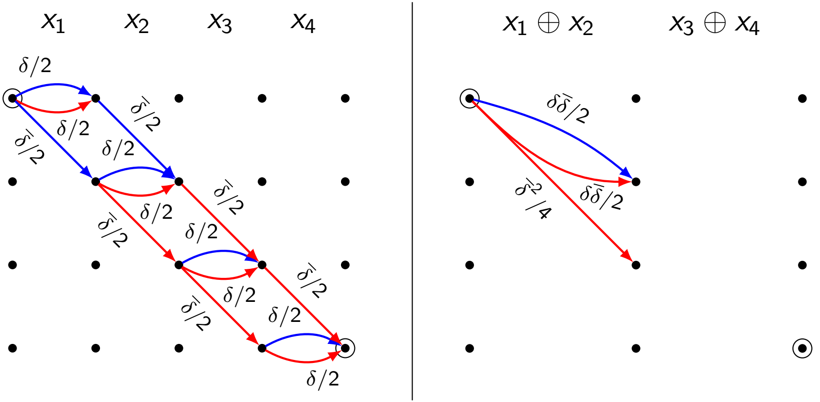
- Example: $N = 4$ inputs with length-3 output 011
- Edge labels: blue $x_j = 0$ and red $x_j = 1$
- Direction: diagonal = no deletion and horizontal = deletion

Deletion channel trellis and the minus operation



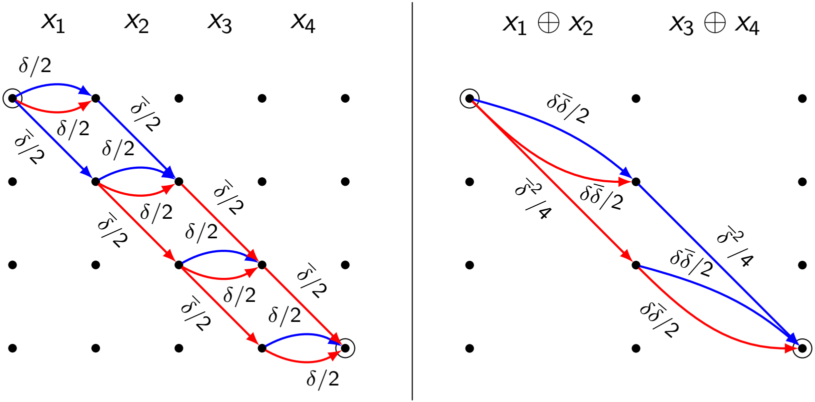
- Half as many sections representing twice the channel uses

Deletion channel trellis and the minus operation



- Half as many sections representing twice the channel uses
  - Edge weight is product of edge weights along length-2 paths
  - Edge label (i.e., color) is the xor of labels along length-2 paths

Deletion channel trellis and the minus operation



- Half as many sections representing twice the channel uses
  - Edge weight is product of edge weights along length-2 paths
  - Edge label (i.e., color) is the xor of labels along length-2 paths

Weak polarization

Theorem

For any $\epsilon > 0$,

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ i \in [N] \mid H(U_i | U_1^{i-1}, \mathrm{Y}) \in [\epsilon, 1-\epsilon] \right\} \right| = 0$$

The proof follows along similar lines as the seminal proof:

- Define a tree process
- Show that the process is a submartingale
- Show that the submartingale can only converge to 0 or 1

All the above follow easily, once we notice the following

- Let $\mathrm{X} \odot \mathrm{X}'$ be two concatenated inputs to the channel
- Denote the corresponding output $\mathrm{Y} \odot \mathrm{Y}'$
- Then,

$$H(A|B, \mathrm{Y} \odot \mathrm{Y}') \geq H(A|B, \mathrm{Y}, \mathrm{Y}')$$

Strong polarization

- Fix $N = 2^n$,

$$n_0 = \lfloor \gamma \cdot n \rfloor \quad \text{and} \quad n_1 = \lceil (1-\gamma) \cdot n \rceil$$

- Define

$$N_0 = 2^{n_0} \qquad \text{and} \qquad N_1 = 2^{n_1}$$

- Let $\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_{N_1}$ by i.i.d. blocks of length N_0
- Suppose the channel input is $\mathsf{X}_1 \odot \mathsf{X}_2 \odot \cdots \odot \mathsf{X}_{N_1}$
- Decoder sees $\mathsf{Y}_1 \odot \mathsf{Y}_2 \odot \cdots \odot \mathsf{Y}_{N_1}$
- If only we had a genie to "punctuate" the output to $\mathsf{Y}_1, \mathsf{Y}_2, \ldots, \mathsf{Y}_{N_1}$, proving strong polarization would be easy...

A “good enough” genie

- We would like this:

A “good enough” genie

- We would like this:

- We will settle for this:

A "good enough" genie

- We would like this:

- We will settle for this:

- No head. . .

A “good enough” genie

- We would like this:

- We will settle for this:

- No head. . .
- No tail. . .

A “good enough” genie

- Decoder sees

$$\mathsf{Y}_1 \odot \mathsf{Y}_2 \odot \cdots \odot \mathsf{Y}_{N_1}$$

- Decoder wants a genie to punctuate the above into

$$\mathsf{Y}_1, \mathsf{Y}_2, \ldots, \mathsf{Y}_{N_1}$$

- Our “good enough” genie will give the decoder

$$\mathsf{Y}^\star_1, \mathsf{Y}^\star_2, \ldots, \mathsf{Y}^\star_{N_1}$$

  where $\mathsf{Y}^\star_i$ is Y_i, with leading and trailing ‘0’ symbols removed
- Asymptotically, we have sacrificed nothing because

$$\mathcal{I}(\mathsf{X};\mathsf{Y}) = \mathcal{I}(\mathsf{X};\mathsf{Y}^\star)$$

Building our genie

- Guard bands added at the encoder
- Denote $\mathrm{x} = \mathrm{x}_{\mathrm{I}} \odot \mathrm{x}_{\mathrm{II}} \in \mathcal{X}^{2^n}$, where $\mathcal{X} = \{0,1\}$ and

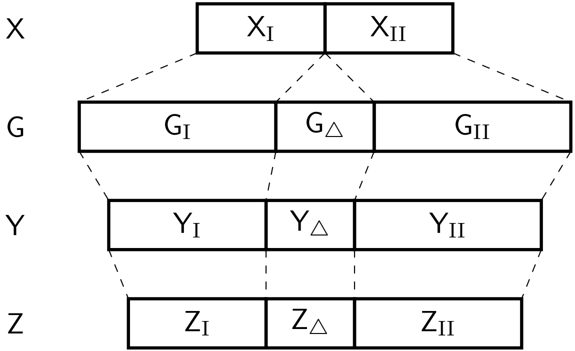
$$\mathrm{x}_{\mathrm{I}} = x_1^{2^{n-1}} \in \mathcal{X}^{2^{n-1}}, \quad \mathrm{x}_{\mathrm{II}} = x_{2^{n-1}+1}^{2^n} \in \mathcal{X}^{2^{n-1}}$$

- That is, instead of transmitting x, we transmit, $g(\mathrm{x})$, where

$$g(\mathrm{x}) \triangleq \begin{cases} \mathrm{x} & \text{if } n \le n_0 \\ g(\mathrm{x}_{\mathrm{I}}) \odot \overbrace{00\ldots0}^{\ell_n} \odot g(\mathrm{x}_{\mathrm{II}}) & \text{if } n > n_0, \end{cases}$$
$$\ell_n \triangleq 2^{\lfloor (1-\epsilon)(n-1) \rfloor}$$

- ϵ is a 'small' constant

The genie in action



- Z is Y with leading and trailing '0' symbols removed
- Guard band $Z_\triangle$ removed by splitting Z in half, and then removing leading and trailing 0 symbols from each half
- Genie successful if the middle of Z falls in the guard band

Conclusions

- Strong polarization for the deletion channel with constant deletion probability δ
- Error rate $2^{-\Lambda^{\gamma}}$ comes from balancing strong polarization and guard-band failure
- If capacity of deletion channel achievable by hidden-Markov inputs, then we can achieve capacity!

References I

[Che19] Mahdi Cheraghchi.
Capacity upper bounds for deletion-type channels.
Journal of the ACM (JACM), 66(2):9, 2019.

[CK15] Jason Castiglione and Aleksandar Kavcic.
Trellis based lower bounds on capacities of channels with synchronization errors.
In *Information Theory Workshop*, pages 24–28, Jeju, South Korea, 2015. IEEE.

[DM01] Matthew C Davey and David JC MacKay.
Reliable communication over channels with insertions, deletions, and substitutions.
IEEE Transactions on Information Theory, 47(2):687–698, 2001.

[Dob67] Roland L'vovich Dobrushin.
Shannon's theorems for channels with synchronization errors.
Problemy Peredachi Informatsii, 3(4):18–36, 1967.

References II

[FD10] Dario Fertonani and Tolga M Duman.
Novel bounds on the capacity of the binary deletion channel.
IEEE Transactions on Information Theory, 56(6):2753–2765, 2010.

[Lev66] V. I. Levenshtein.
Binary codes capable of correcting deletions, insertions and reversals.
Soviet Physics - Doklady, 10(8):707–710, February 1966.

[Mit09] Michael Mitzenmacher.
A survey of results for deletion channels and related synchronization channels.
Probability Surveys, 6:1–33, 2009.

[MTL12] Hugues Mercier, Vahid Tarokh, and Fabrice Labeau.
Bounds on the capacity of discrete memoryless channels corrupted by synchronization and substitution errors.
IEEE Transactions on Information Theory, 58(7):4306–4330, 2012.

References III

[RD15] Mojtaba Rahmati and Tolga M Duman.
Upper bounds on the capacity of deletion channels using channel fragmentation.
IEEE Transactions on Information Theory, 61(1):146–156, 2015.

[TFV18] Kuangda Tian, Arman Fazeli, and Alexander Vardy.
Polar coding for deletion channels: Theory and implementation.
In *IEEE International Symposium on Information Theory*, pages 1869–1873, 2018.

[TFVL17] Kuangda Tian, Arman Fazeli, Alexander Vardy, and Rongke Liu.
Polar codes for channels with deletions.
In *55th Annual Allerton Conference on Communication, Control, and Computing*, pages 572–579, 2017.

[TTVM17] E. K. Thomas, V. Y. F. Tan, A. Vardy, and M. Motani.
Polar coding for the binary erasure channel with deletions.
IEEE Communications Letters, 21(4):710–713, April 2017.

References IV

[WLH14] Runxin Wang, Rongke Liu, and Yi Hou.
Joint successive cancellation decoding of polar codes over intersymbol interference channels.
arXiv preprint arXiv:1404.3001, 2014.