An Information-Theoretic Perspective on Successive-Cancellation List Decoding

Mustafa Cemil Coşkun[*,0] and Henry D. Pfister[†]
[*]Nokia Bell Labs and [†]Duke University
mustafa.coskun@nokia-bell-labs.com

The Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology (MIT)

Cambridge, MA
November 2, 2022

[0]supported by German Research Foundation (DFG) under Grant KR 3517/9-1 when I was with the Technical University of Munich (TUM)

Motivation

- Recent interest in efficient transmission of short information blocks[1]
  - 5G New Radio (5G NR) and already 6G[2,3]
  - Internet of things (IoT) and wireless sensor networks (WSN), etc.[3]
- Successive-cancellation list (SCL) decoding combined with outer cyclic redundancy checks (CRCs) makes polar codes competitive for short blocks
  - Adopted for uplink and downlink control information for the enhanced mobile broadband (eMBB) communication service[4]
  - Candidate for ultra-reliable low-latency communications (URLLCs) and massive machine-type communications (mMTCs)[4]
  - Useful for communicating over the fading channels with no CSI[5,6]
  - Part of possible solutions for unsourced random access problem[7,8]

---

1 [Polyanskiy et al., 2010], Channel coding rate in the finite blocklength regime... (Trans. Inf. Theory)

2 [Durisi et al., 2016], Towards massive, ultra-reliable, and low-latency wireless communications with short packets... (IEEE Proc.)

3 [Mahmood et al., 2020], White paper on critical and massive machine type communication towards 6G... (CoRR)

4 [Bioglio et al., 2021], Design of polar codes in 5G new radio... (IEEE Commun. Surveys & Tutorials)

5 [Xhemrishi et al., 2019], List decoding of short codes for communication over unknown fading channels... (Asilomar)

6 [Yuan et. al., 2021], Polar-coded non-coherent communication... (Commun. Lett.)

7 [Fengler et. al., 2022], Pilot-based unsourced RA with a massive MIMO receiver, interference cancellation, and power control... (J-SAC)

8 [Gkagkos et al., 2022], FASURA: A scheme for quasi-static massive MIMO unsourced random access channels... (CoRR)

Outline

1. Background
2. An Information-Theoretic Perspective
   - Binary Erasure Channel
3. Numerical Results
4. Conclusions

Outline

1. Background
2. An Information-Theoretic Perspective
   - Binary Erasure Channel
3. Numerical Results
4. Conclusions

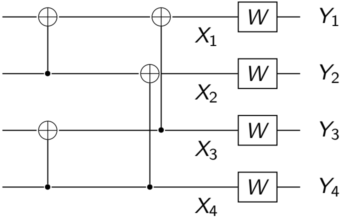
Linear Codes based on Polar Transform

$$X^N = U^N \boldsymbol{G}_N \quad \text{where} \quad \boldsymbol{G}_N \triangleq \boldsymbol{B}_N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n} \quad \text{and} \quad N = 2^n$$

frozen/info sets
$\mathcal{F} = \{3\}$
$\mathcal{A} = \{1, 2, 4\}$

frozen/info bits
$U_i = f_i\left(U^{i-1}\right),\ i \in \mathcal{F}$
$U_{\mathcal{A}} = V_1^K$

info $U_1 = V_1$
info $U_2 = V_2$
frozen $U_3 = U_1 \oplus U_2$
info $U_4 = V_3$

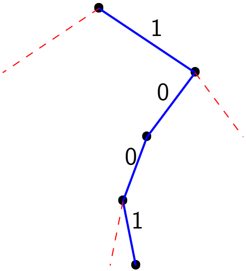


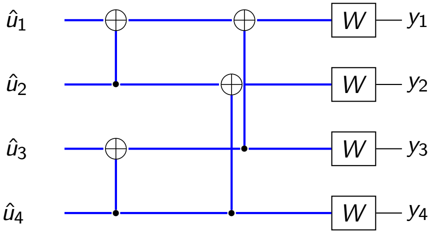
- Reed–Muller (RM) codes: optimize $\mathcal{F}$ to maximize minimum distance[9,10]
- Polar codes: optimize $\mathcal{F}$ for successive cancellation (SC) decoding[11,12]
  Requires estimation of $H(W_N^{(i)}) \triangleq H\left(U_i | Y^N, U^{i-1}\right)$, for $i \in \{1, \ldots, N\}$
- Any binary linear code is obtained with suitable $\mathcal{F}$ and f_i,[13] $i \in \mathcal{F}$

9 [Muller, 1954], Application of boolean algebra to switching circuit design and to error detection... (Trans. IRE Elect. Comp.)
10 [Reed, 1954], A class of multiple-error-correcting codes and the decoding scheme... (Trans. IRE Inf. Theory)
11 [Stolte, 2002], Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung... (PhD Thesis, TU Darmstadt)
12 [Arıkan, 2009], Channel polarization: A method for constructing capacity-achieving codes for BMSCs... (Trans. Inf. Theory)
13 [Trifonov and Miloslavskaya, 2016], Polar subcodes... (J-SAC)

Successive-Cancellation Decoding

Example: $u_3 = 0$ (frozen bit)





- Errors made by SC decoding cannot be corrected by later decisions
- Use SC list (SCL) decoding[11,14,15] to approach ML decoding performance

---

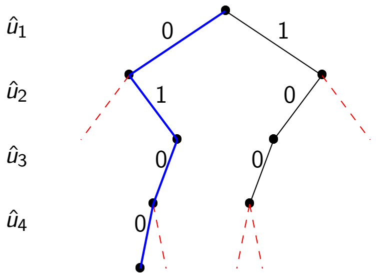
11 [Stolte, 2002], Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung... (PhD Thesis, TU Darmstadt)

14 [Dumer and Shabunov, 2006], Soft-decision decoding of Reed-Muller codes: recursive lists... (Trans. Inf. Theory)

15 [Tal and Vardy, 2015], List decoding of polar codes... (Trans. Inf. Theory)

Successive-Cancellation List Decoding

For each i, both options are stored for decision on $\hat{u}_i$ (i-th decoding stage,) which doubles the number of paths at each decoding stage



$L = 2$

rank based on $P_{U^i|Y^N}(\hat{u}^i|y^N)$

After N-th stage

$$\hat{u}^N = \arg\max_{u^N \in \mathcal{L}} P_{U^N|Y^N}\left(u^N|y^N\right)$$

- Keep a list $\mathcal{L}$ of past decisions for a set of L likely paths
  Use an optimal rule to choose between final codewords on the list $\mathcal{L}$
- Recall that any binary linear code can be represented as a polar code
  Can be used for any binary linear block code; may not be efficient

Motivating Question

- What list size is sufficient to approach ML decoding performance for a given code and channel?
  - Can be attacked via simulation but quite complex for long codes and lists
  - Simulation alone unlikely to provide insight into the question
  - A theoretical answer might enable better code designs for SCL decoding
  - A recent related work co-authored by Vardy[16] provides an upper bound on the sufficient list size by generalizing a previous result co-authored by Urbanke[17]

[16] [Fazeli et al., 2021], List decoding of polar codes: How large should the list be to achieve ML decoding?... (ISIT)

[17] [Hashemi et al., 2018], Decoder partitioning: Towards practical list decoding of polar codes... (Trans. Commun.)

Summary of Results

- New bounds related to the list size required for near-ML performance
  - To avoid losing true codeword, its rank must not be larger than list size
  - The expected log-rank of correct codeword is upper bounded by an entropy
  - Bounds on this entropy are derived and relatively easy to compute
  - The log-rank of correct codeword concentrates around this mean
  - As an application, the analysis is used to modify RM codes
    New codes outperform 5G codes under SCL decoding with practical list sizes
- For the binary erasure channel (BEC)
  - This entropy equals the dimension of an affine subspace
  - The random dimension sequence can be approximated by a Markov chain
  - For a fixed number of erasures, the approximation is reasonably accurate

Outline

1. Background
2. An Information-Theoretic Perspective
   - Binary Erasure Channel
3. Numerical Results
4. Conclusions

An Information Theoretic Perspective (1)

Basic Idea: After m steps, consider the conditional entropy $H\big(U^m | Y^N\big)$

The chain rule of entropy implies:

$$H\big(U^m | Y^N\big) = \sum_{i=1}^{m} H\big(U_i \big| U^{i-1}, Y_1^N\big)$$
$$= \sum_{i=1}^{m} H\big(W_N^{(i)}\big)$$

But what about the frozen bits?

An Information-Theoretic Perspective (2)

- For the first m input bits, the information/frozen sets are denoted

$$\mathcal{A}^{(m)} \triangleq \mathcal{A} \cap \{1, \ldots, m\} \quad \text{and} \quad \mathcal{F}^{(m)} \triangleq \mathcal{F} \cap \{1, \ldots, m\}$$

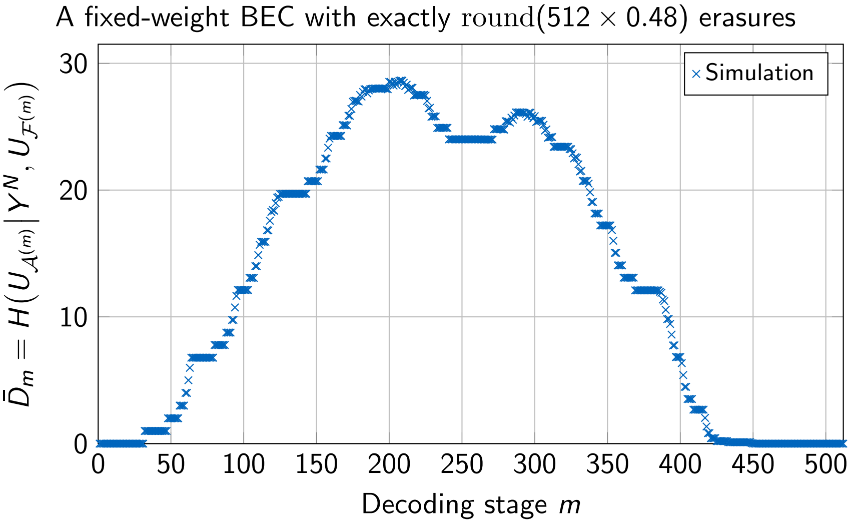
- **Key Idea:** information entropy given frozen bits

$$d_m(y^N) \triangleq H\big(U_{\mathcal{A}^{(m)}} \big| Y^N = y^N, U_{\mathcal{F}^{(m)}}\big) \quad \text{and} \quad \bar{D}_m = H\big(U_{\mathcal{A}^{(m)}} \big| Y^N, U_{\mathcal{F}^{(m)}}\big)$$

$$\sum_{i \in \mathcal{A}^{(m)}} H\big(W_N^{(i)}\big) - \sum_{i \in \mathcal{F}^{(m)}} \big(1 - H\big(W_N^{(i)}\big)\big) \leq \bar{D}_m \leq \sum_{i \in \mathcal{A}^{(m)}} H\big(W_N^{(i)}\big)$$

Q: Is the sequence $\{\bar{D}_m\}$ relevant for the list size for near-ML decoding on BMSCs?

An Exemplary $(512, 256)$ Code



Bounding the List Size

Theorem

Upon observing y^N when u^N is sent, we define the set (for $\alpha \in (0,1]$)
$\mathcal{S}_\alpha^{(m)}(u^m, y^N) \triangleq \{\tilde{u}^m : \mathbb{P}(\tilde{u}_{\mathcal{A}^{(m)}} | y^N, \tilde{u}_{\mathcal{F}^{(m)}}) \geq \alpha \mathbb{P}(u_{\mathcal{A}^{(m)}} | y_1^N, u_{\mathcal{F}^{(m)}})\}$. *Then,*

$$\mathrm{E}\left[\log_2 |\mathcal{S}_\alpha^{(m)}|\right] \leq \bar{D}_m + \log_2 \frac{1}{\alpha} = H\big(U_{\mathcal{A}^{(m)}} \big| Y^N, U_{\mathcal{F}^{(m)}}\big) + \log_2 \frac{1}{\alpha}$$

- Choosing $\alpha < 1$ (say 0.94) captures near misses and matches entropy better
- Consider SCL decoding with max list size L_m after the m-th decoding step
  - It needs to satisfy $L_m \geq |\mathcal{S}_1^{(m)}|$ for the true u_1^m to stay on the list

    A first-order code design criterion: $\log_2 L_m \geq \bar{D}_m$
    (reduce the peak $\bar{D}_{\max} \triangleq \max_m \bar{D}_m$)

A Few Remarks

- This approach currently has two weaknesses:
  - Entropy determined by typical events but coding cares about rare events
  - Sequence $\bar{D}_m$ averaged over Y^N but decoder sees one realization $d_m(y^N)$

Theorem

For a wide range of BMS channels, the random variable D_m concentrates around its mean $\bar{D}_m$, i.e., for any $\beta > 0$, we have

$$\mathbb{P}\left\{\frac{1}{N}|D_m - \bar{D}_m| > \beta\right\} \leq 2\exp\left(-\frac{\beta^2}{c^2}N\right)$$

where c is a positive constant defined by the channel

Outline

1. Background
2. An Information-Theoretic Perspective
   - Binary Erasure Channel
3. Numerical Results
4. Conclusions

Successive Cancellation Inactivation Decoding

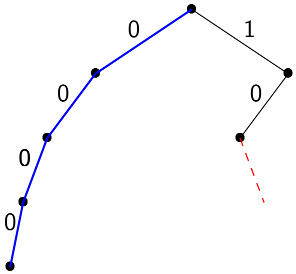
- Consider SCL decoding with unbounded list size on the BEC
  - Set of all valid paths after any decoding stage forms an affine subspace
  - SCL decoding tracks all valid paths defined by this subspace
- SC inactivation (SCI) decoding instead stores a basis for space[18]
  - If SC decoding step outputs erasure, inactivate the bit and add basis vector
  - Later messages in decoder are functions of inactivated bits (i.e., basis vectors)
  - If SC decoding of frozen bit is an unerased message, then resulting equation may allow one to consolidate the basis (i.e., remove a basis vector)

[18] [Coșkun, Neu, and Pfister, 2020], Successive cancellation inactivation decoding for modified Reed–Muller and eBCH codes... (ISIT)

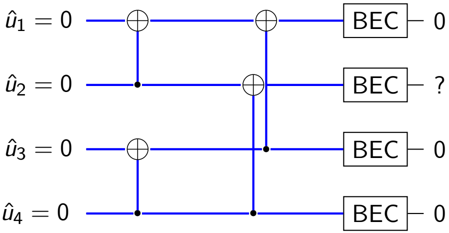
Successive Cancellation Inactivation Decoding: Algorithm

Example: $u_3 = 0$ (frozen bits)

$$d_m\left(y_1^N\right) = \text{subspace dimension after } m\text{-th decoding stage}$$



$d_4\left(y_1^4\right) = 0$



- Unique solution only if $d_N\left(y_1^N\right) = 0$; otherwise declare an error
- Equivalent to SCL decoding with unbounded list size aka ML decoding

Outline

1 Background

2 An Information-Theoretic Perspective
- Binary Erasure Channel

3 Numerical Results

4 Conclusions

PAC Codes and the Dynamic Reed-Muller Ensemble

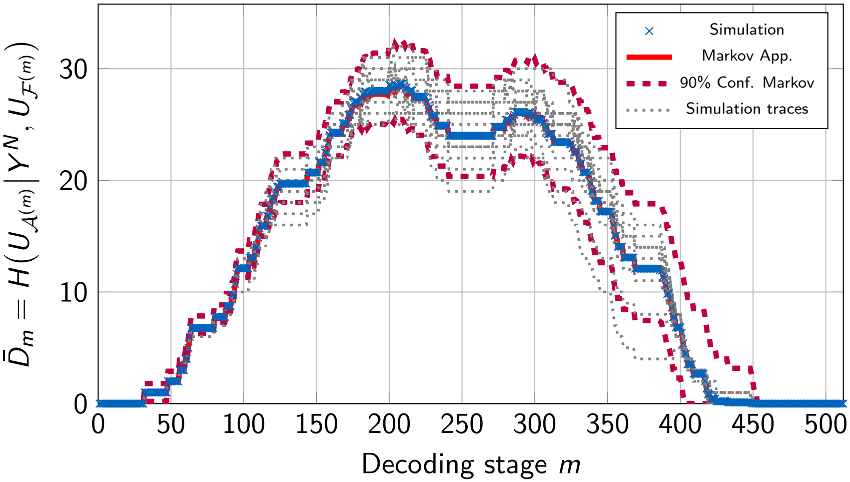
- Polarization-adjusted convolutional (PAC) codes[19]:
  - Given set $\mathcal{A}$ and a rate-1 convolutional code (CC) with memory ν
  - Encode information and frozen bit sequence with CC before applying polar transform
  - Decode using SCL or other methods, e.g., sequential decoding
  - For short lengths, RM frozen indices appear to be a good choice
- Dynamic RM (dRM) code ensemble[18]:
  - Let $\mathcal{A}$ be the information indices of an RM code
  - Modified RM code where frozen bits are random linear function of past bits
  - Closely related to PAC codes

19 [Arıkan, 2019], From sequential decoding to channel polarization and back again... (CoRR)

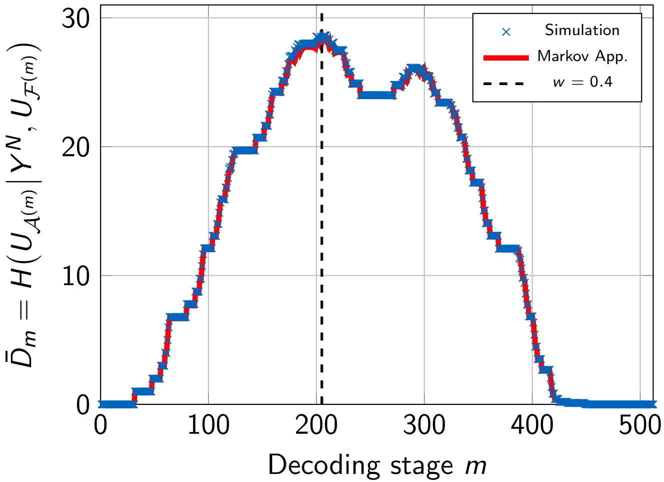
18 [Coşkun, Neu and Pfister, 2020] Successive cancellation inactivation decoding for modified Reed–Muller and eBCH codes... (ISIT)

$(512, 256)$ dRM Code

A fixed-weight BEC with exactly $\mathrm{round}(512 \times 0.48)$ erasures



$(512, 256)$ dRM Code

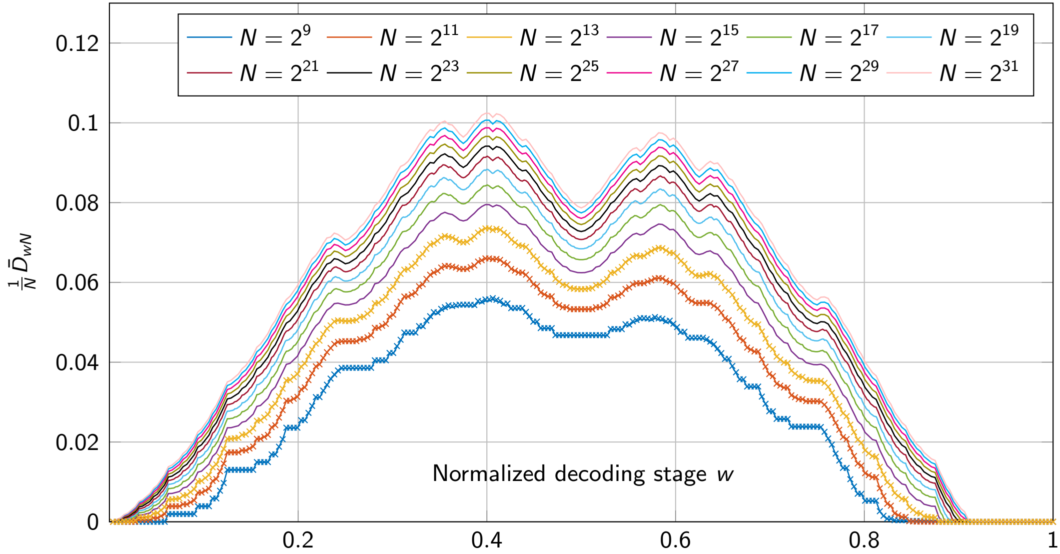


- How does $\bar{D}_m$ behave as block length increases?
- Let $w \triangleq \frac{m}{N}$, $m \in \{1, \ldots, N\}$, define the sequence $\frac{1}{N}\bar{D}_{wN}$

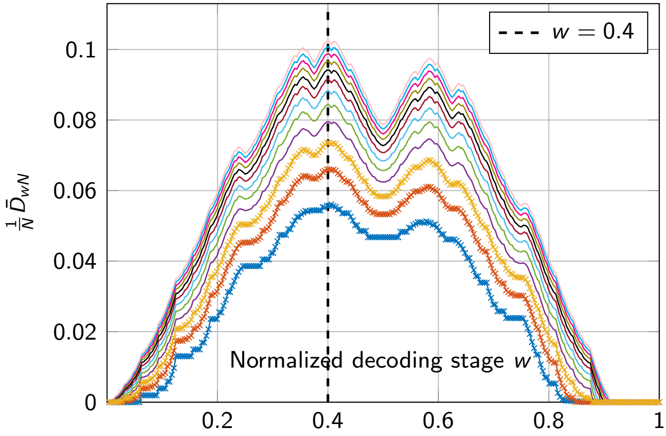
  E.g.: $w = 0.4 \rightarrow \frac{1}{512}\bar{D}_{0.4N}$

Growth Rate of Subspace Dimension for dRM Codes with $R = 0.5$

A fixed-weight BEC with exactly $\mathrm{round}(N \times 0.48)$ erasures



Growth Rate of Subspace Dimension for dRM Codes with $R = 0.5$

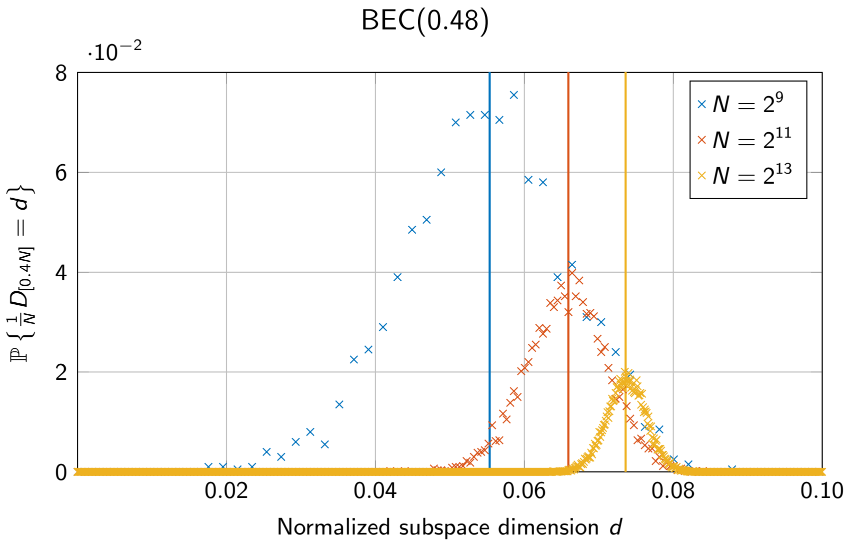


- How does the PMF of $\frac{1}{N}\bar{D}_{wN}$ behave as block length increases for a fixed w?
- Set $w = 0.4$ and plot

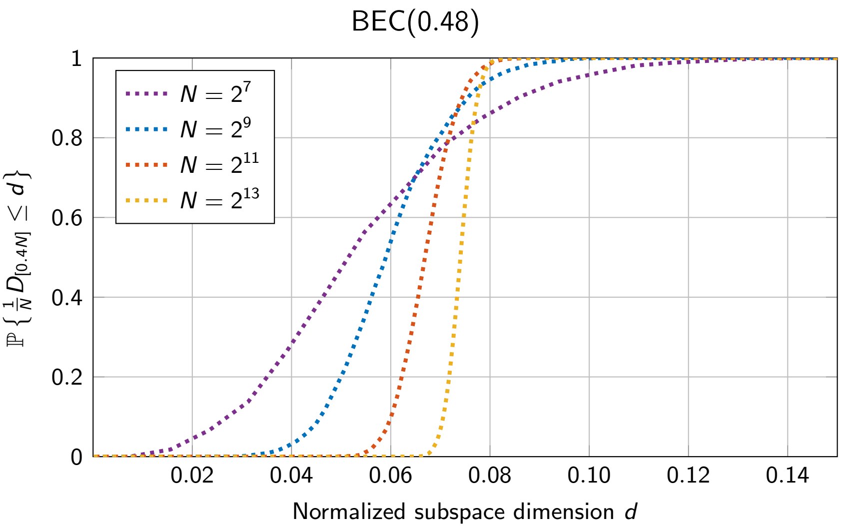
$$\mathbb{P}\left\{\frac{1}{N}\bar{D}_{[0.4N]} = d\right\}$$

where $d \in \mathcal{D}_N \subset [0, R]$

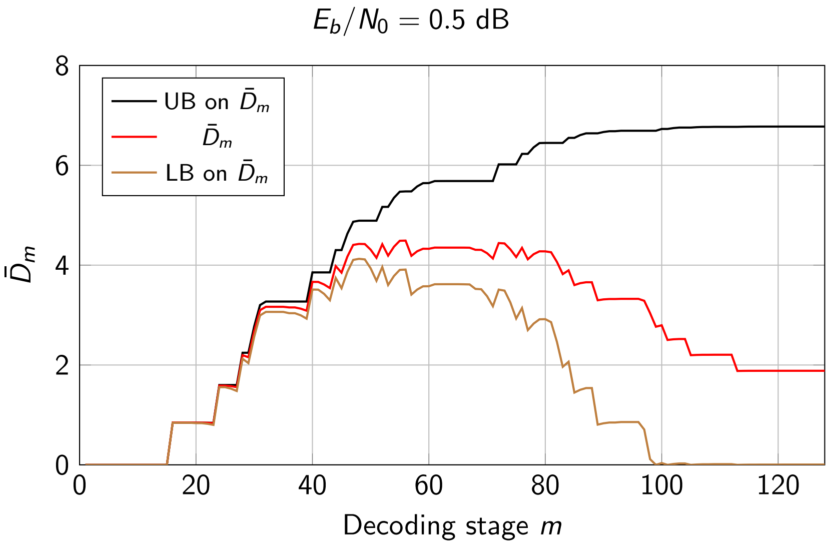
PMFs for $\frac{1}{N} D_{[0.4N]}$ for dRM Codes with $R = 0.5$



CDFs for $\frac{1}{N} D_{[0.4N]}$ for dRM Codes with $R = 0.5$



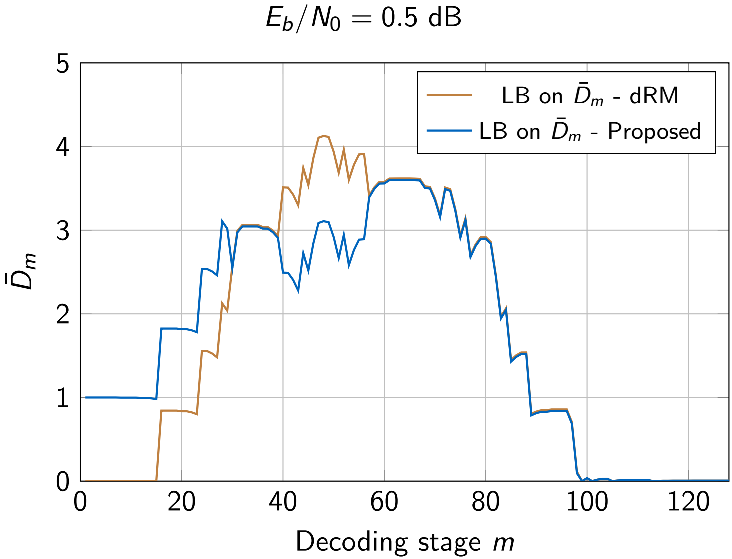
$(128, 64)$ dRM Code over the BAWGNC



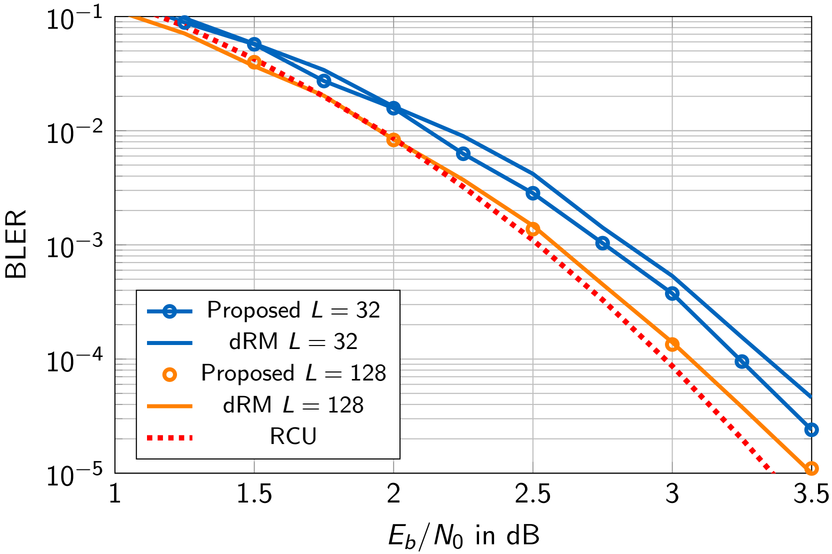
$(128, 64)$ Proposed vs dRM Code over the BAWGNC

Proposed Code

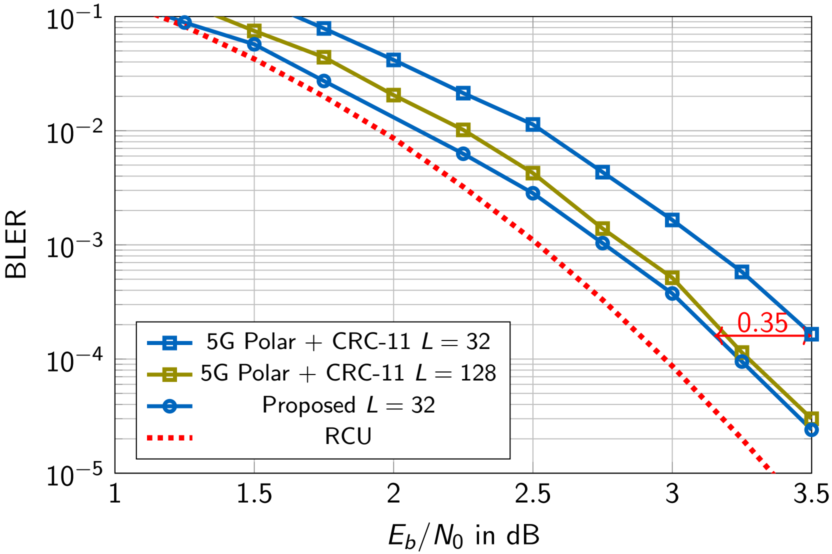
- $u_{\{30,40\}}$ dynamic frozen bits
- $u_{\{1,57\}}$ info. bits



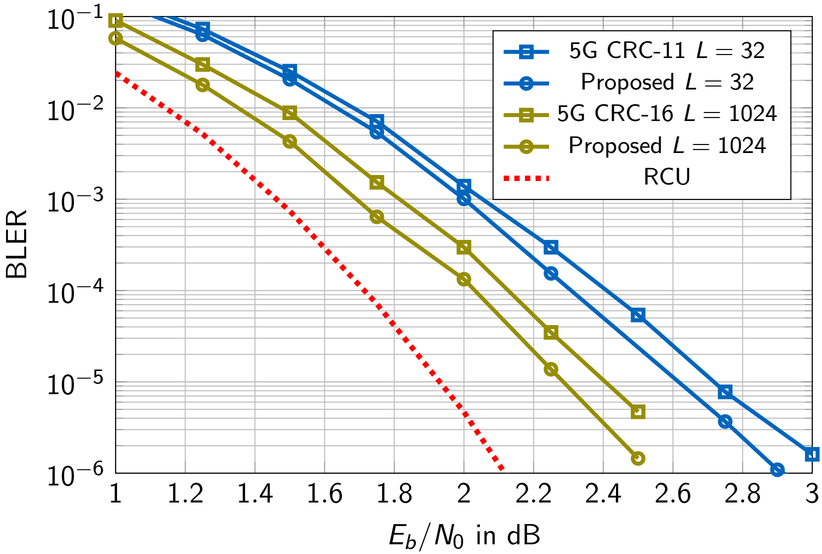
(128, 64) Codes over the BAWGNC



$(128, 64)$ Codes over the BAWGNC



$(512, 256)$ Codes over the BAWGNC



Outline

1. Background
2. An Information-Theoretic Perspective
   - Binary Erasure Channel
3. Numerical Results
4. Conclusions

Conclusions

- Recent advances in polar codes allow performance near random coding union bound for codes up to 512 bits of length with moderate complexity
  - Dynamic frozen bits act as outer code that improves minimum distance
  - SCL decoding can fully exploit outer code with large enough list size
- "What list size is sufficient to approach maximum-likelihood (ML) decoding performance under an SCL decoder?"
  - Information theory provides some estimates of required list size
  - For the BEC, the estimate is quite accurate and even relevant for optimum decoding
- Analysis leads to improved designs (in comparison with the PAC code and 5G polar codes) under SCL decoding with list sizes $L \in [8, 1024]$

Thanks for your attention