

# An Information-Theoretic Perspective on Successive-Cancellation List Decoding

**Mustafa Cemil Coşkun**<sup>\*,0</sup> and Henry D. Pfister<sup>†</sup>

<sup>\*</sup>Nokia Bell Labs and <sup>†</sup>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

---

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

- Recent interest in **efficient** transmission of **short** information blocks<sup>1</sup>
  - 5G New Radio (5G NR) and already 6G<sup>2,3</sup>
  - Internet of things (IoT) and wireless sensor networks (WSN), etc.<sup>3</sup>
- **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<sup>4</sup>
  - **Candidate** for ultra-reliable low-latency communications (URLLCs) and massive machine-type communications (mMTCs)<sup>4</sup>
  - **Useful** for communicating over the fading channels with no CSI<sup>5,6</sup>
  - **Part of possible solutions** for unsourced random access problem<sup>7,8</sup>

---

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

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

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

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

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

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

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

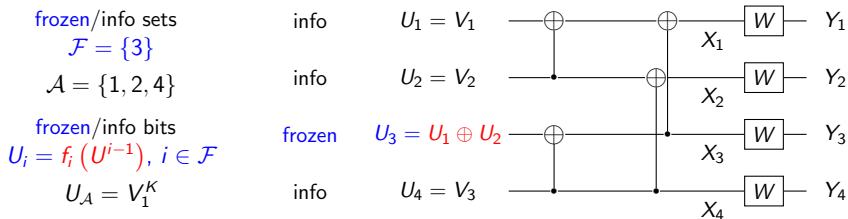
<sup>8</sup> [Gkagkos et al., 2022], FASURA: A scheme for quasi-static massive MIMO unsourced random access channels... (CoRR)

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

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

# Linear Codes based on Polar Transform

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



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

<sup>9</sup>[Muller, 1954], Application of boolean algebra to switching circuit design and to error detection... (Trans. IRE Elect. Comp.)

<sup>10</sup>[Reed, 1954], A class of multiple-error-correcting codes and the decoding scheme... (Trans. IRE Inf. Theory)

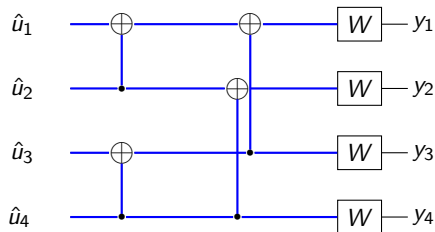
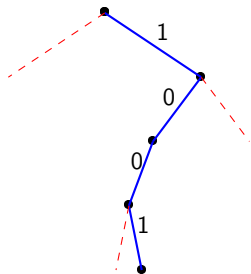
<sup>11</sup>[Stolte, 2002], Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung... (PhD Thesis, TU Darmstadt)

<sup>12</sup>[Arkan, 2009], Channel polarization: A method for constructing capacity-achieving codes for BMSCs... (Trans. Inf. Theory)

<sup>13</sup>[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<sup>11,14,15</sup> to **approach ML decoding performance**

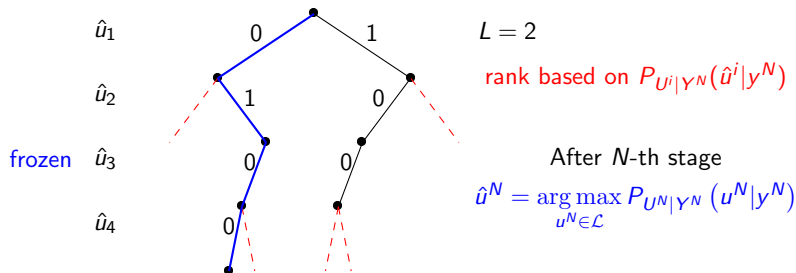
<sup>11</sup>[Stolte, 2002], Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung... (PhD Thesis, TU Darmstadt)

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

<sup>15</sup>[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



- 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<sup>16</sup> provides an **upper bound on the sufficient list size** by generalizing a previous result co-authored by Urbanke<sup>17</sup>

---

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

<sup>17</sup>[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**

- 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  $H(U^m|Y^N)$

The chain rule of entropy implies:

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

But what about the frozen bits?

## An Information-Theoretic Perspective (2)

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

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

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

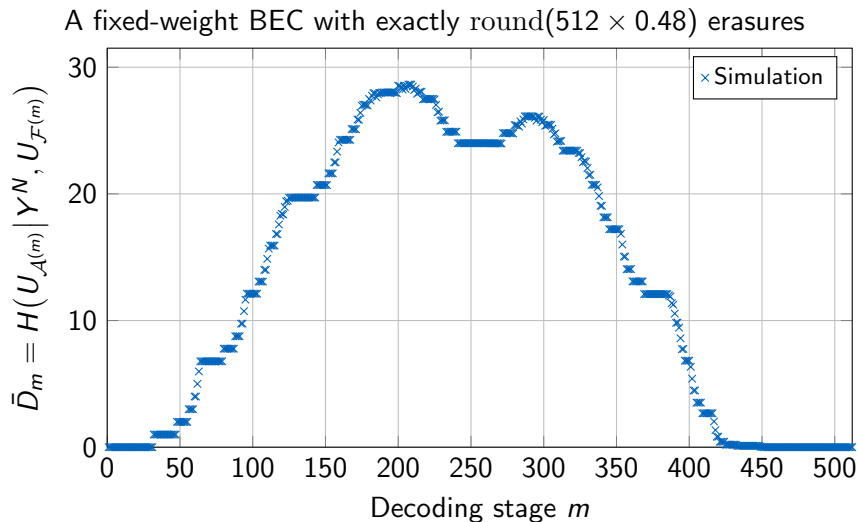
$$d_m(y^N) \triangleq H(U_{\mathcal{A}^{(m)}} | Y^N = y^N, U_{\mathcal{F}^{(m)}}) \quad \text{and} \quad \bar{D}_m = H(U_{\mathcal{A}^{(m)}} | Y^N, U_{\mathcal{F}^{(m)}})$$

---

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

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

# An Exemplary (512, 256) Code



## 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,

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

- 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*

- 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**<sup>18</sup>
  - 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)

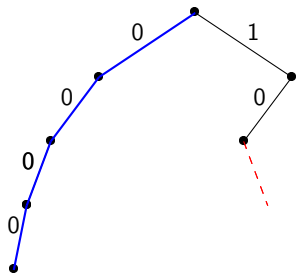
---

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

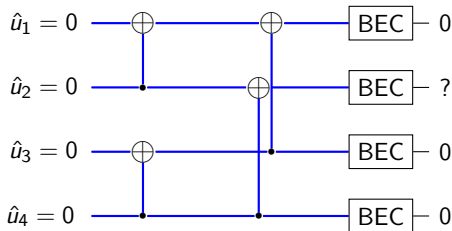
# Successive Cancellation Inactivation Decoding: Algorithm

Example:  $u_3 = 0$  (frozen bits)

$d_m(y_1^N)$  = subspace dimension after  $m$ -th decoding stage



$$d_4(y_1^4) = 0$$



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

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

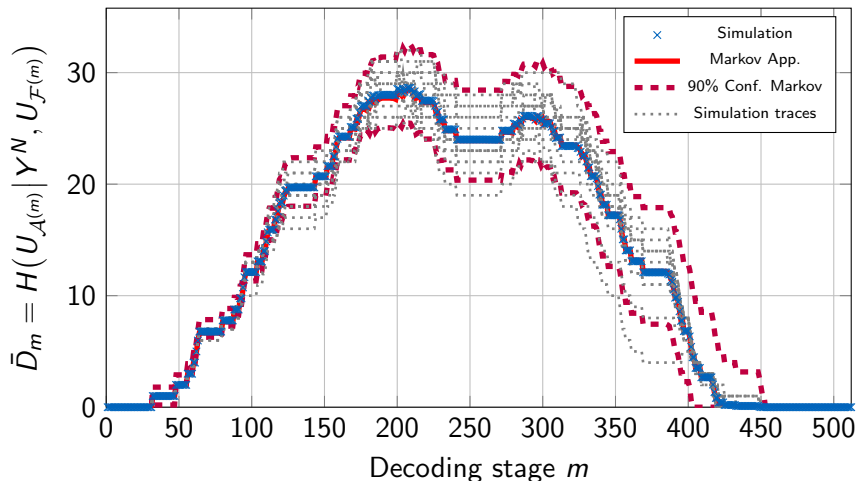
- Polarization-adjusted convolutional (PAC) codes<sup>19</sup>:
  - Given set  $\mathcal{A}$  and a rate-1 convolutional code (CC) with memory  $\nu$
  - Encode information and frozen bit sequence with CC before applying polar transform
  - Decode using SCL or other methods, e.g., sequential decoding
  - For short lengths, RM frozen indices appear to be a good choice
- Dynamic RM (dRM) code ensemble<sup>18</sup>:
  - 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

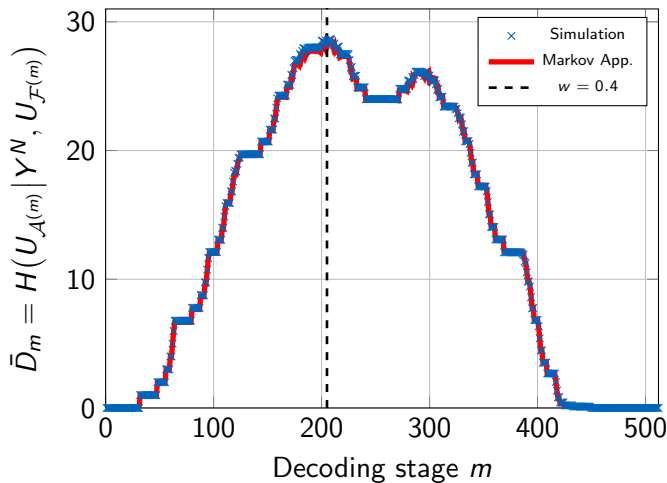
---

<sup>19</sup>[Arkan, 2019], From sequential decoding to channel polarization and back again... (CoRR)

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

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

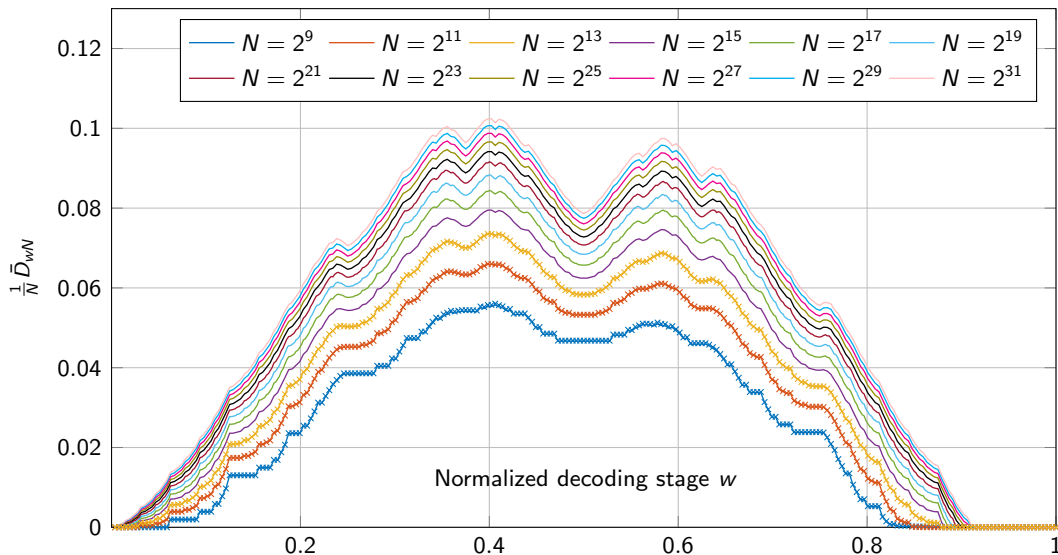




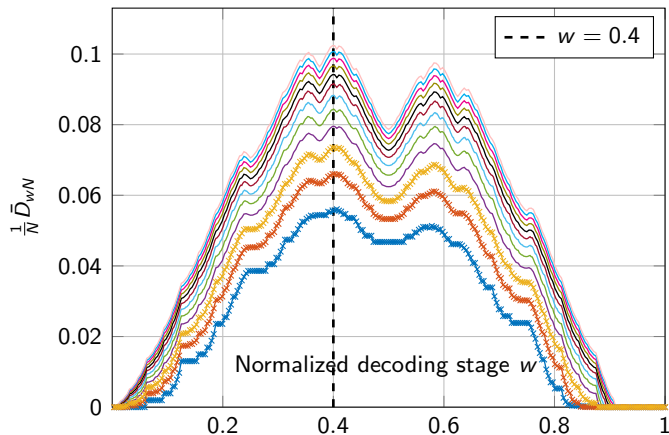
- How does  $\bar{D}_m$  behave as block length increases?
- Let  $w \triangleq \frac{m}{N}$ ,  $m \in \{1, \dots, N\}$ , define the sequence  $\frac{1}{N} \bar{D}_{wN}$   
E.g.:  $w = 0.4 \rightarrow \frac{1}{512} \bar{D}_{0.4N}$

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

A fixed-weight BEC with exactly  $\text{round}(N \times 0.48)$  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  $w$ ?

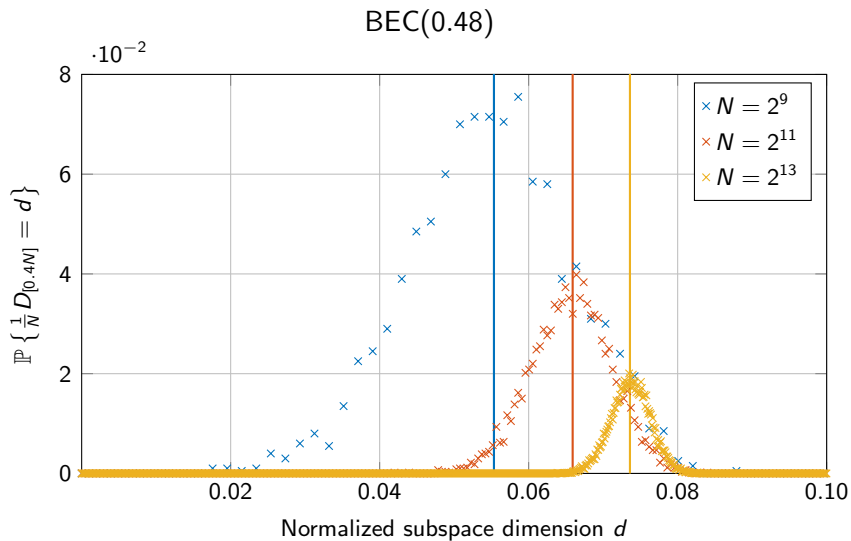
- Set  $w = 0.4$  and plot

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

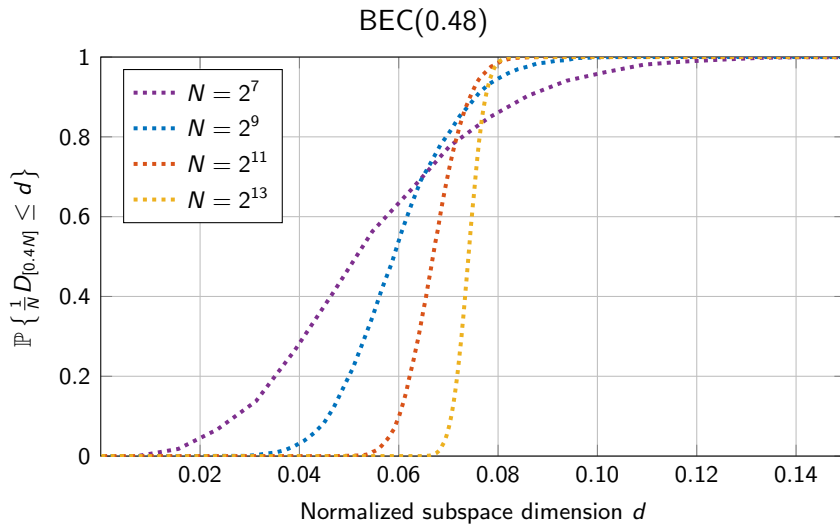
where  $d \in \mathcal{D}_N \subset [0, R]$



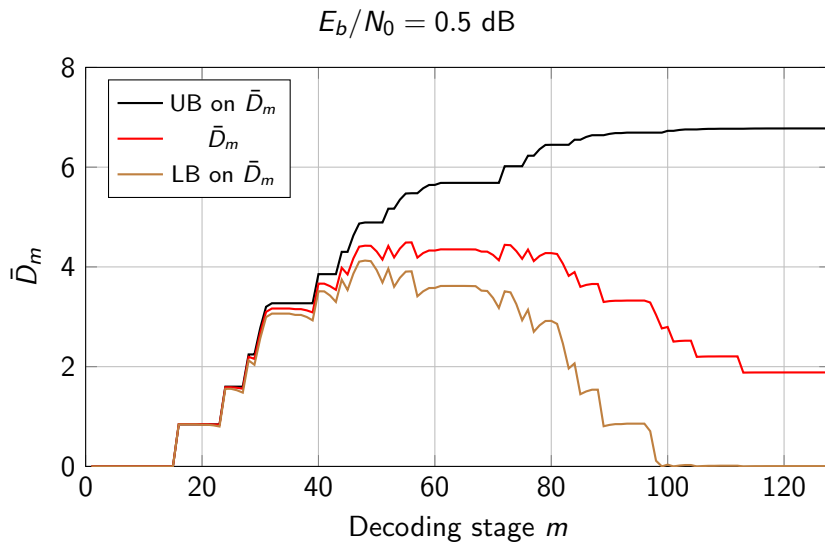
# PMFs for $\frac{1}{N}D_{[0.4N]}$ for dRM Codes with $R = 0.5$



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



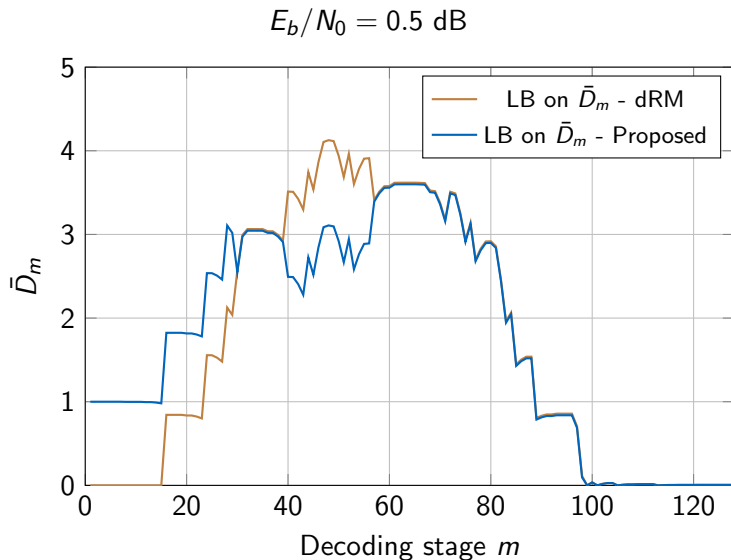
# (128, 64) dRM Code over the BAWGNC



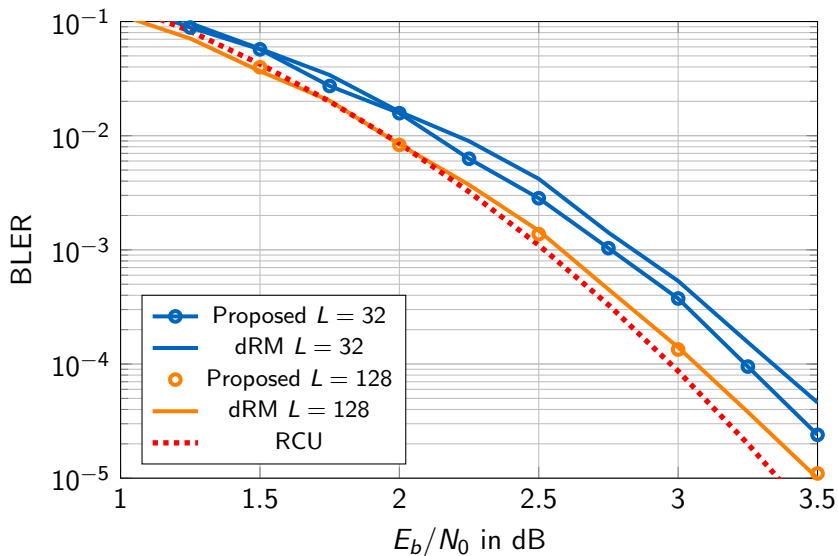
# (128, 64) Proposed vs dRM Code over the BAWGNC

Proposed Code

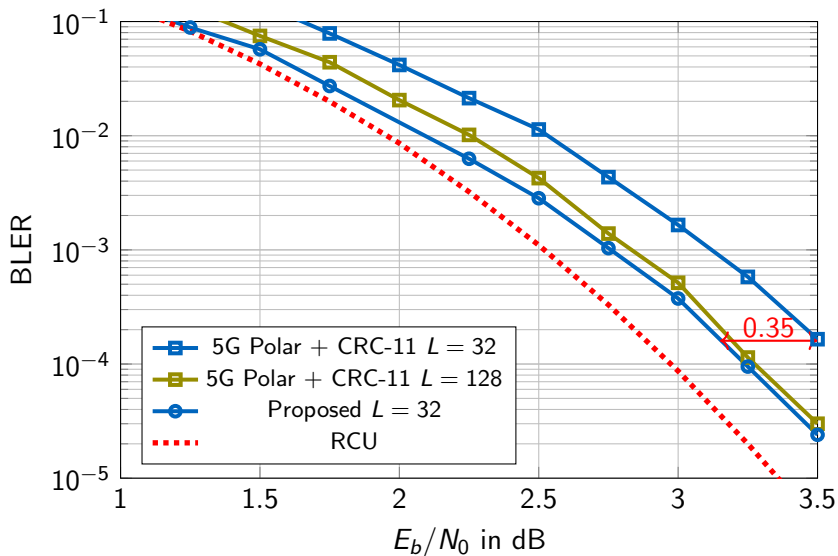
- $u_{\{30,40\}}$  dynamic frozen bits
- $u_{\{1,57\}}$  info. bits



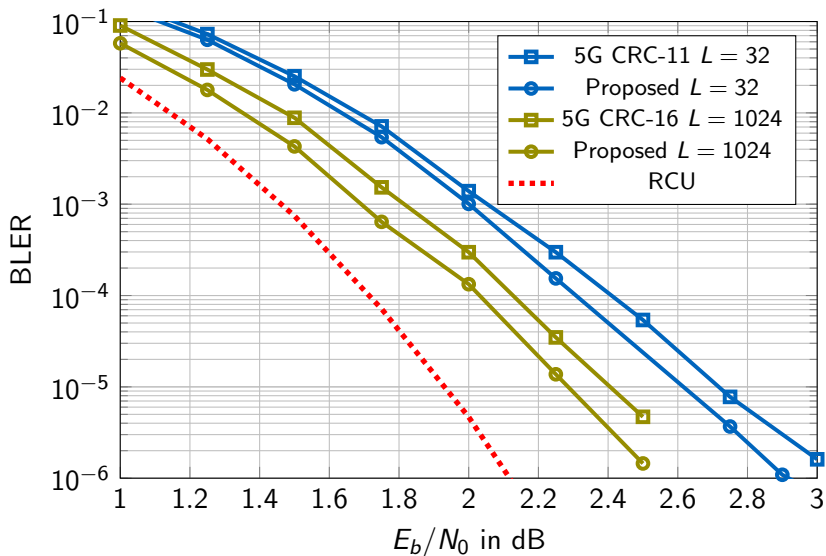
# (128, 64) Codes over the BAWGNC



# (128, 64) Codes over the BAWGNC



# (512, 256) Codes over the BAWGNC



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



A lot of recent work in this area. Here is partial list of some notable papers:

- Optimization of frozen bit positions for SCL decoding
  - E.g., recent works co-authored by Viterbo [RV19], ten Brink [EEtB19], Hashemi [LHCG22, LHYC19] and Ye [LYH21, LYH22]
- Optimization of polar codes with dynamic frozen bits for SCL decoding (another keyword: [pre-transformed polar codes](#))
  - E.g., co-authored by Miloslavskaya on short designs [MV20, MVL<sup>+</sup>21]
  - E.g., co-authored by Zhang on their concatenation with CRCs [LGZ21]
- Distance spectrum analysis of polar codes with dynamic frozen bits
  - E.g., a recent award-winning work co-authored by Vardy [YFV21] and another work co-authored by Zhang [LZL<sup>+</sup>21]
  - E.g., on PAC codes co-authored by Vardy [YFV20] and Viterbo [RBV20]

- 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**

- [EECtB19] A. Elkelesh, M. Ebada, S. Cammerer, and S. ten Brink.  
Decoder-tailored polar code design using the genetic algorithm.  
*IEEE Transactions on Communications*, 67(7):4521–4534, 2019.
- [LGZ21] B. Li, J. Gu, and H. Zhang.  
Performance of CRC concatenated pre-transformed RM-polar codes.  
*CoRR*, abs/2104.07486, 2021.
- [LHCG22] Y. Liao, S. A. Hashemi, J. M. Cioffi, and A. Goldsmith.  
Construction of polar codes with reinforcement learning.  
*IEEE Transactions on Communications*, 70(1):185–198, 2022.
- [LHYC19] Y. Liao, S. A. Hashemi, H. Yang, and J. M. Cioffi.  
Scalable polar code construction for successive cancellation list decoding: A graph neural network-based approach.  
*CoRR*, abs/2207.01105, 2019.

- [LYH21] G. Li, M. Ye, and S. Hu.  
A dynamic programming method to construct polar codes with improved performance.  
*CoRR*, abs/2111.02851, 2021.
- [LYH22] G. Li, M. Ye, and S. Hu.  
Adjacent-bits-swapped polar codes: A new code construction to speed up polarization.  
*CoRR*, abs/2202.04454, 2022.
- [LZL<sup>+</sup>21] Y. Li, H. Zhang, R. Li, J. Wang, G. Yan, and Z. Ma.  
On the weight spectrum of pre-transformed polar codes.  
*CoRR*, abs/2102.12625, 2021.
- [MV20] V. Miloslavskaya and B. Vucetic.  
Design of short polar codes for SCL decoding.  
*IEEE Transactions on Communications*, 68(11):6657–6668, 2020.
- [MVL<sup>+</sup>21] V. Miloslavskaya, B. Vucetic, Y. Li, G. Park, and O. Park.  
Recursive design of precoded polar codes for SCL decoding.  
*IEEE Transactions on Communications*, 69(12):7945–7959, 2021.

- [RBV20] M. Rowshan, A. Burg, and E. Viterbo.  
Polarization-adjusted convolutional (PAC) codes: Fano decoding vs list decoding.  
*CoRR*, abs/2002.06805, 2020.
- [RV19] M. Rowshan and E. Viterbo.  
How to modify polar codes for list decoding.  
In *IEEE International Symposium on Information Theory*, pages 1772–1776, 2019.
- [YFV20] H. Yao, A. Fazeli, and A. Vardy.  
List decoding of Arıkan’s PAC codes.  
*CoRR*, abs/2005.13711, 2020.
- [YFV21] H. Yao, A. Fazeli, and A. Vardy.  
A deterministic algorithm for computing the weight distribution of polar codes.  
In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 1218–1223, 2021.