Polar Coding
A perspective for Applications

Erdal Arıkan
İhsan Doğramacı Bilkent University, Ankara, Turkey

WCNC 2018 Workshop on Polar Coding
15 April 2018, Barcelona, Spain

Polar codes for applications: Topics

- Code construction algorithms
  - Density evolution
  - Gaussian approximation
  - Huawei method
- Improved decoding algorithms
  - List decoding with and without CRC
  - Sequential methods
  - ML methods
- Improved codes
  - Concatenation schemes
  - Nonbinary kernels
  - Universal codes
- Length adjustment
  - Shortening
  - Puncturing
- Rate-compatible constructions
- Incremental redundancy
  - Hybrid ARQ
  - Rateless coding
- Blind detection/Early termination
- Implementation issues
  - Throughput
  - Energy efficency
  - Power density
  - Latency
  - Gap to Shannon limit

Outline

- Polar code performance
  - List decoding
  - Effect of CRC
- Puncturing/Shortening/HARQ
- Some implementation Issues

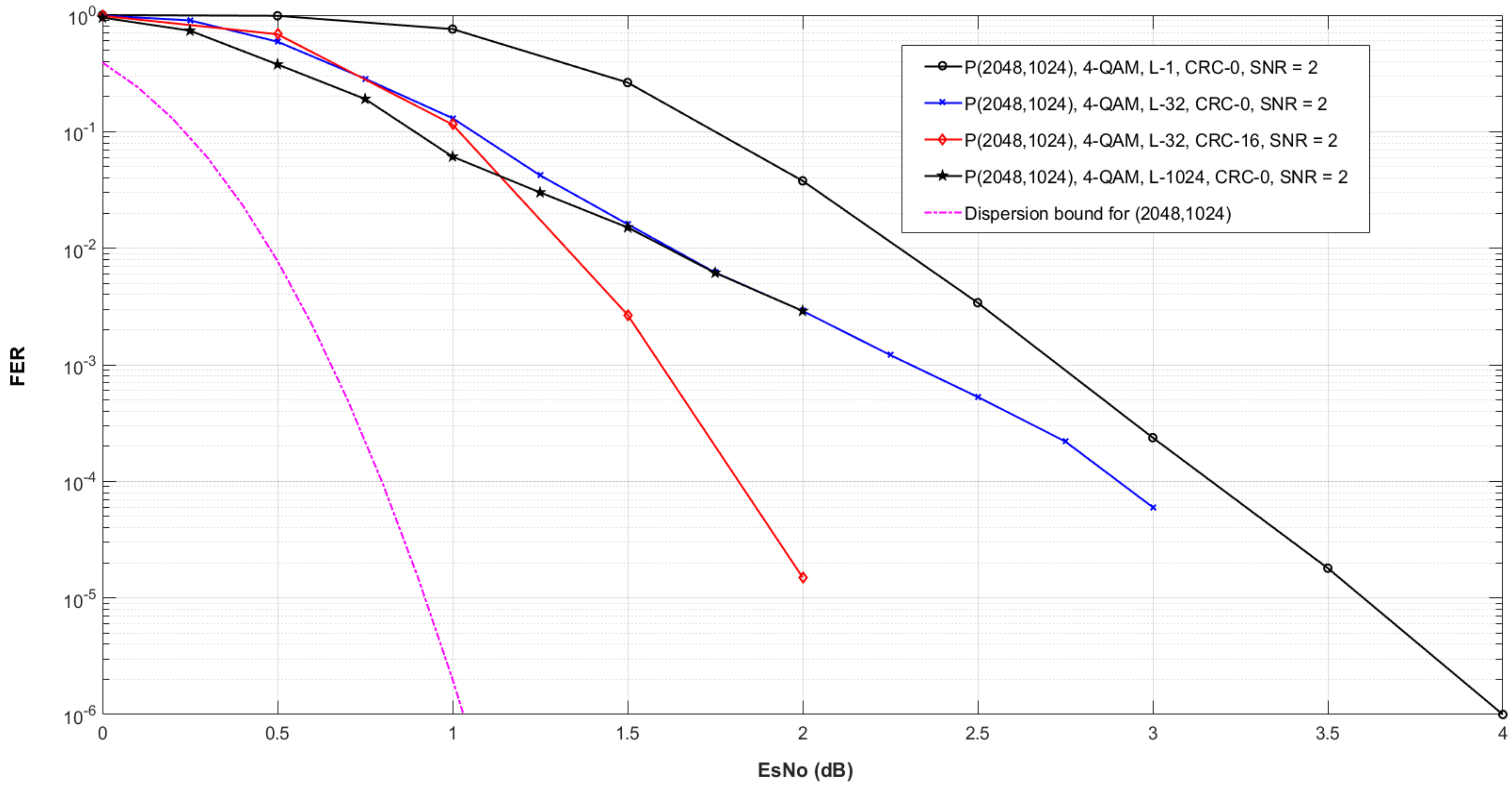
Decoding

- Tal, I. and Vardy, A. (2011) ‘List decoding of polar codes’, in *2011 IEEE International Symposium on Information Theory Proceedings*. *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 1–5. doi: 10.1109/ISIT.2011.6033904.
- Dumer, I. (2017b) ‘On decoding algorithms for polar codes’, *arXiv:1703.05307 [cs, math]*. Available at: http://arxiv.org/abs/1703.05307 (Accessed: 17 March 2017).
- Arikan, E. (2010) ‘A survey of reed-muller codes from polar coding perspective’, in *2010 IEEE Information Theory Workshop (ITW)*. *2010 IEEE Information Theory Workshop (ITW)*, IEEE, pp. 1–5. doi: 10.1109/ITWKSPS.2010.5503223.
- Li, B., Shen, H. and Chen, K. (2015) ‘A Decision-Aided Parallel SC-List Decoder for Polar Codes’, *arXiv:1506.02955 [cs, math]*. Available at: http://arxiv.org/abs/1506.02955 (Accessed: 10 June 2015).
- Li, B., Shen, H. and Tse, D. (2012) ‘An Adaptive Successive Cancellation List Decoder for Polar Codes with Cyclic Redundancy Check’, *arXiv:1208.3091*. Available at: http://arxiv.org/abs/1208.3091 (Accessed: 16 August 2012).
- Li, B., Shen, H. and Tse, D. (2013) *Parallel Decoders of Polar Codes*. arXiv e-print 1309.1026. Available at: http://arxiv.org/abs/1309.1026 (Accessed: 5 September 2013).
- Niu, K. and Chen, K. (2012) ‘Stack decoding of polar codes’, *Electronics Letters*, 48(12), pp. 695–697. doi: 10.1049/el.2012.1459.

Performance

- Polar codes achieve channel capacity at low complexity but their performance is not very good (even under ML decoding)
- When concatenated with a CRC, the performance improves immensely; furthermore, list decoding achieves near ML performance at reasonable complexity
- What is the explanation for this improved performance?
- How is it that list decoding achieves near ML performance?

List decoding with CRC: Tal and Vardy (2011)



Adaptive list decoder

- Bin, Shen and Tse (2012) experimented with an adaptive list decoder with CRC, in which the list size was progressively doubled until successful decoding or some limit being reached
  - Start with L=1
  - If failure, set L=2
  - If failure, set L=4
  - Until either success of L = Lmax
- The results show that there is a «cutoff rate» phenomenon in effect!

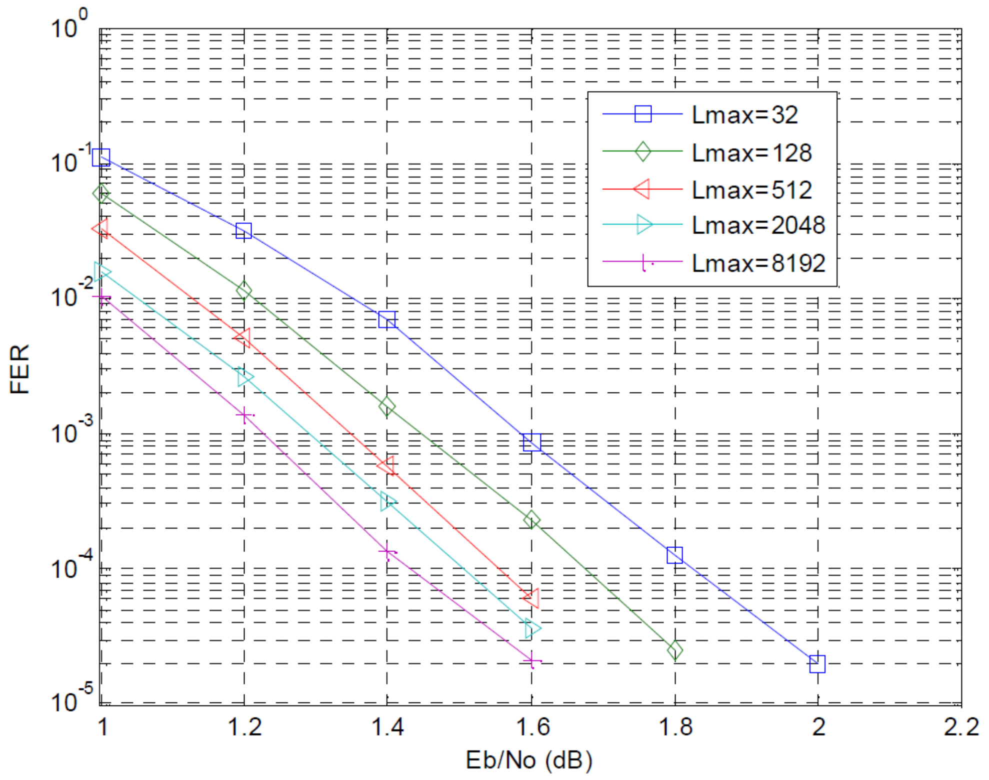


TABLE I. THE MEAN OF L OF THE ADAPTIVE SC-LIST DECODER

E_b/N_o(dB)	1.0	1.2	1.4	1.6	1.8	2.0
L_{max}=32	16.64	8.03	3.86	2.04	1.39	1.14
L_{max}=128	35.31	12.16	4.52	2.17	1.41	
L_{max}=512	70.41	19.14	5.45	2.27		
L_{max}=2048	133.40	30.80	6.64	2.36		
L_{max}=8192	271.07	52.59	7.88	2.47		

Gap to Shannon limit

- Li, Shen and Tse (2012) further report that a (2048,1024) polar code, under adaptive list decoding with Lmax = 262,144 and with a 24 bit CRC, performs within 0.2 dB of the information-theoretic limit at FER 10^{-3}
- This is a performance that theory did not predict, and we still do not have a full explanation!
- The near-optimal performance of polar codes under CRC-aided list decoding is the main reason why polar codes are making an impact in practice
- How can we explain the near-optimal performance of polar codes when they are concatenated with a relatively short CRC?

References for polar code performance analysis

- E. Arikan and E. Telatar, “On the Rate of Channel Polarization,” *0807.3806*, Jul. 2008.
- R. Mori and T. Tanaka, “Performance and Construction of Polar Codes on Symmetric Binary-Input Memoryless Channels,” *0901.2207*, Jan. 2009.
- S. H. Hassani, R. Mori, T. Tanaka, and R. L. Urbanke, “Rate-dependent analysis of the asymptotic behavior of channel polarization,” *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2267–2276, 2013.
- Goli, A., Hassani, S. H. and Urbanke, R. (2012) ‘Universal Bounds on the Scaling Behavior of Polar Codes’, *arXiv:1205.2876*. Available at: http://arxiv.org/abs/1205.2876 (Accessed: 15 May 2012)
- V. Guruswami and P. Xia, “Polar Codes: Speed of Polarization and Polynomial Gap to Capacity,” *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 3–16, Jan. 2015.
- Arıkan, E. (2015) ‘A Packing Lemma for Polar Codes’, *arXiv:1504.05793 [cs, math]*. Available at: http://arxiv.org/abs/1504.05793 (Accessed: 23 April 2015).

What does theory say about performance?

- Frame error for polar codes with length $N = 2^n$ and rate R satisfies

$$\log(-\log P_e) \approx \frac{n}{2} + \sqrt{n}\, Q^{-1}(R/I)$$

where I is symmetric capacity of channel, and Q^{-1} is inverse of the complementary CDF of unit normal.

- Optimal codes of the same length and rate have a frame error probability that goes to zero as

$$\log(-\log P_e) \approx n + \log E_r(R)$$

where $E_r(R)$ is the channel reliability exponent.

- The inferior performance of polar codes is explained by the fact that polar codes have a minimum distance of $O(\sqrt{N})$ whereas optimal codes have minimum distance of $O(N)$.

Looking for an explanation

- For an RM(r,m) code, the number of codewords with minimum Hamming weight are

$$A_{2^{m-r}} = 2^r \prod_{i=0}^{m-r-1} \left[\frac{2^{m-i}-1}{2^{m-r-i}-1} \right] \leq 2^{m^2} = 2^{(\log N)^2}$$

- This is a large number but not exponential in the block length N.
- The entire inner shell of codewords at minimum distance from a codeword can be eliminated by a CRC of length $\sim(\log N)^2$.

Case study for $N = 2048$

RM order r	Dimension K	Rate R	Minimum Distance d_{min}	Multip. at d_{min} A	CRC length ~$\log_2(A)$	Norm. CRC length ~$\log_2(A)/K$
0	1	0.0005	2048	1	0.0	0.0%
1	12	0.0059	1024	4.1 E+03	12.0	100.0%
2	67	0.0327	512	2.8 E+06	21.4	32.0%
3	232	0.1133	256	4.1 E+08	28.6	12.3%
4	562	0.2744	128	1.4 E+10	33.7	6.0%
5	1024	0.5000	64	1.1 E+11	36.7	3.6%
6	1486	0.7256	32	2.3 E+11	37.7	2.5%
7	1816	0.8867	16	1.1 E+11	36.7	2.0%
8	1981	0.9673	8	1.3 E+10	33.6	1.7%
9	2036	0.9941	4	3.6 E+08	28.4	1.4%
10	2047	0.9995	2	2.1 E+06	21.0	1.0%
11	2048	1.0000	1	2.0 E+03	11.0	0.5%

Expurgation, spectral thinning

- Imposing a CRC constraint of length L on the encoder input block leaves only a fraction 2^{-L} of the original codewords, but changes the code rate by a fraction L/K.
- The RM distance study shows that, at a negligible loss L/K in rate, the entire inner shell of nearest-neighbor codewords can be eliminated, improving the code minimum distance.
- The experimental results suggest that CRC eliminates much more than just the nearest neighbors.
- Conjecture: A CRC of length $\mathrm{L} \sim (\log N)^2$ improves the minimum distance of polar code to $O(N)$.
- Although motivated by different goals, a paper related to this subject is «Arıkan, E. (2015) ‘A Packing Lemma for Polar Codes’, arXiv:1504.05793 [cs, math]»

Outline

- Polar code performance
- Puncturing/Shortening/HARQ
- Some implementation Issues

Puncturing and shortening

- Shin, D.-M., Lim, S.-C. and Yang, K. (2013) ‘Design of Length-Compatible Polar Codes Based on the Reduction of Polarizing Matrices’, *IEEE Transactions on Communications*, 61(7), pp. 2593–2599. doi: 10.1109/TCOMM.2013.052013.120543.
- Niu, K., Chen, K. and Lin, J. R. (2013) ‘Beyond turbo codes: Rate-compatible punctured polar codes’, in *2013 IEEE International Conference on Communications (ICC)*. *2013 IEEE International Conference on Communications (ICC)*, pp. 3423–3427.
- Wang, R. and Liu, R. (2014) ‘A Novel Puncturing Scheme for Polar Codes’, *IEEE Communications Letters*, 18(12), pp. 2081–2084.
- Zhang, L. *et al.* (2014) ‘On the puncturing patterns for punctured polar codes’, in *2014 IEEE International Symposium on Information Theory*. *2014 IEEE International Symposium on Information Theory*, pp. 121–125. doi: 10.1109/ISIT.2014.6874807.
- Miloslavskaya, V. (2015) ‘Shortened Polar Codes’, *IEEE Transactions on Information Theory*, 61(9), pp. 4852–4865.
- Chandesris, L., Savin, V. and Declercq, D. (2017) ‘On Puncturing Strategies for Polar Codes’, *arXiv:1701.04746 [cs, math]*.
- Bioglio, V., Gabry, F. and Land, I. (2017) ‘Low-Complexity Puncturing and Shortening of Polar Codes’, in *Wireless Communications and Networking Conference Workshops (WCNCW).*

Puncturing and Shortening

Puncturing	Shortening
A fixed segment of the codeword is not sent; both are methods of adjusting the length of a code	
Unsent segment is not fixed	Unsent segment is fixed
Encode data as usual	Encoder has to be designed for a subcode; or a systematic encoder may be used
Decoder treats punctured segment as erasures	Decoder treats shortened segment as known
The original code design no longer optimal	
Punctured bits may be sent incrementally as a method of IR-HARQ	N/A

A Shortening method: Wang and Liu (2014)

- Polar transform is an operation of the form $x = uG$ where $G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}$
- Wang and Liu (2014) proposed that shortening set $S \subset \{1,2,\cdots,2^n\}$ be selected so that

$$(u_S = 0) \Longrightarrow (x_S = 0).$$

- This simplifies encoding since now an encoder for the unshortened code can be used
- The constraint can be satisified by a «greedy» algorithm that picks S by giving priority to rows of G with greater Hamming weight.
- After selecting S to adjust the length, the rate can be adjusted by picking a frozen set F distinct from S.
- The resulting shortened code has length $M = 2^n - |S|$, and rate $1 - \frac{|S|+|F|}{M}$.

Puncturing: A conservation law by Shin *et al* (2013)

- Punctured bits are treated as erasures by the decoder
- The erasures in the received codeword due to shortening lead to an equal number of erasures at the bit-channels of the source vector
  - A given subset P of punctured bits, the erasures caused by these bits percolate to a corresponding subset P' of erasued bits on the source side, with $|P'| = |P|$
  - This is not specific to polar codes or to any decoding method; it is an information-theoretic identity of general nature
- This leads to a basic design constraint for choosing a puncturing set for polar codes: Choose P so that P' does not overlap with information bits on the source side

Puncturing/Shortening Summary

- Optimal frozen bit pattern on the source side depends on the particular puncturing/shortening pattern on the codeword side
- Optimal puncturing/shortening pattern on the codeword side depends on the particular frozen bit pattern on the source side
- No efficient method is known for joint optimization of frozen/puncturing/shortening patterns
- Even if a solution is found, it may be too complex for applications where all these patterns have to be stored or computed on the fly
- Bioglio et al (2017) propose a pragmatic approach in which the frozen pattern is fixed by the mother polar code design, the puncturing pattern is chosen as the bit-reversed image of $\{1,2,\ldots,P\}$, and the shortening pattern is chosen as the bit-reversed image of $\{N-S+1,\ldots,N\}$, where N is the length of the mother code, P is the number of punctured bits, and S is the number of shortened bits

Rate-compatible codes

- A sequence of nested codes $C_1, C_2, \cdots, C_n$ of decreasing code rate such that the codeword bits of C_i are embedded in the codeword bits of C_{i+1}.
- RC codes are attractive because they can be encoded and decoded using the same hardware.
- All RC codes can be thought of as being obtained by puncturing or extending extending.
- In puncturing, one typically begins with an optimize mother code C_n, and obtains the other codes by puncturing.
- In extending, one begins with an optimized C_1 and obtaines the other codes by extending C_1 by adding further parities.
- Clearly, one can also start somewhere in the middle and apply both puncturing downstream and extending upstream.
- In any case, the given any RC code $C_1, C_2, \cdots, C_n$ one may view this as being constructed by extending or puncturing.
- In HARQ extending is the preferred approach since one desires to maximize chances of success in the first trial.
- A special case of RC coding is rateless coding where the number of codes in the RC code sequence can be arbitrarily large.

HARQ with Polar Codes

- Chen, K., Niu, K. and Lin, J. (2013) 'A Hybrid ARQ Scheme Based on Polar Codes', *IEEE Communications Letters*, 17(10), pp. 1996–1999. doi: 10.1109/LCOMM.2013.090213.131670.
- Chen, K. *et al.* (2014) 'Polar coded HARQ scheme with Chase combining', in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 474–479. doi: 10.1109/WCNC.2014.6952074.
- Saber, H. and Marsland, I. (2015) 'An Incremental Redundancy Hybrid ARQ Scheme via Puncturing and Extending of Polar Codes', *IEEE Transactions on Communications*, 63(11), pp. 3964–3973. doi: 10.1109/TCOMM.2015.2477082.
- El-Khamy, M. *et al.* (2015) 'HARQ Rate-Compatible Polar Codes for Wireless Channels', in *2015 IEEE Global Communications Conference (GLOBECOM)*. *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6. doi: 10.1109/GLOCOM.2015.7417429.
- Tavildar, S. R. (2016) 'A H-ARQ scheme for polar codes', *arXiv:1606.08545 [cs, math]*. Available at: http://arxiv.org/abs/1606.08545 (Accessed: 29 June 2016).
- Ruslan Morozov and Peter Trifonov (2017) *Chained polar codes for HARQ*. Available at: http://seim-conf.org/media/materials/2017/proceedings/SEIM-2017_Short_Papers.pdf#page=22 (Accessed: 4 November 2017).
- Ma, L., JieXiong and Wei, Y. (2017) 'An Incremental Redundancy HARQ Scheme for Polar Code', *arXiv:1708.09679 [cs, math]*. Available at: http://arxiv.org/abs/1708.09679 (Accessed: 25 October 2017).

Rateless/HARQ Polar Codes

- Li, B. et al. (2015) 'Capacity-Achieving Rateless Polar Codes', arXiv:1508.03112 [cs, math]. Available at: http://arxiv.org/abs/1508.03112 (Accessed: 7 September 2016).
- Li, B. et al. (2016) 'Capacity-achieving rateless polar codes', in 2016 IEEE International Symposium on Information Theory (ISIT). 2016 IEEE International Symposium on Information Theory (ISIT), pp. 46–50. doi: 10.1109/ISIT.2016.7541258.
- Hong, S. N., Hui, D. and Marić, I. (2017) 'Capacity-Achieving Rate-Compatible Polar Codes', IEEE Transactions on Information Theory, PP(99), pp. 1–1. doi: 10.1109/TIT.2017.2756668.
- Mondelli, M. et al. (2017) 'Capacity-Achieving Rate-Compatible Polar Codes for General Channels', in 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). 2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), pp. 1–6. doi: 10.1109/WCNCW.2017.7919107.
- Feng, B., Zhang, Q. and Jiao, J. (2017) 'An Efficient Rateless Scheme Based on the Extendibility of Systematic Polar Codes', IEEE Access, PP(99), pp. 1–1. doi: 10.1109/ACCESS.2017.2762363.

Capacity-achieving rateless polar codes

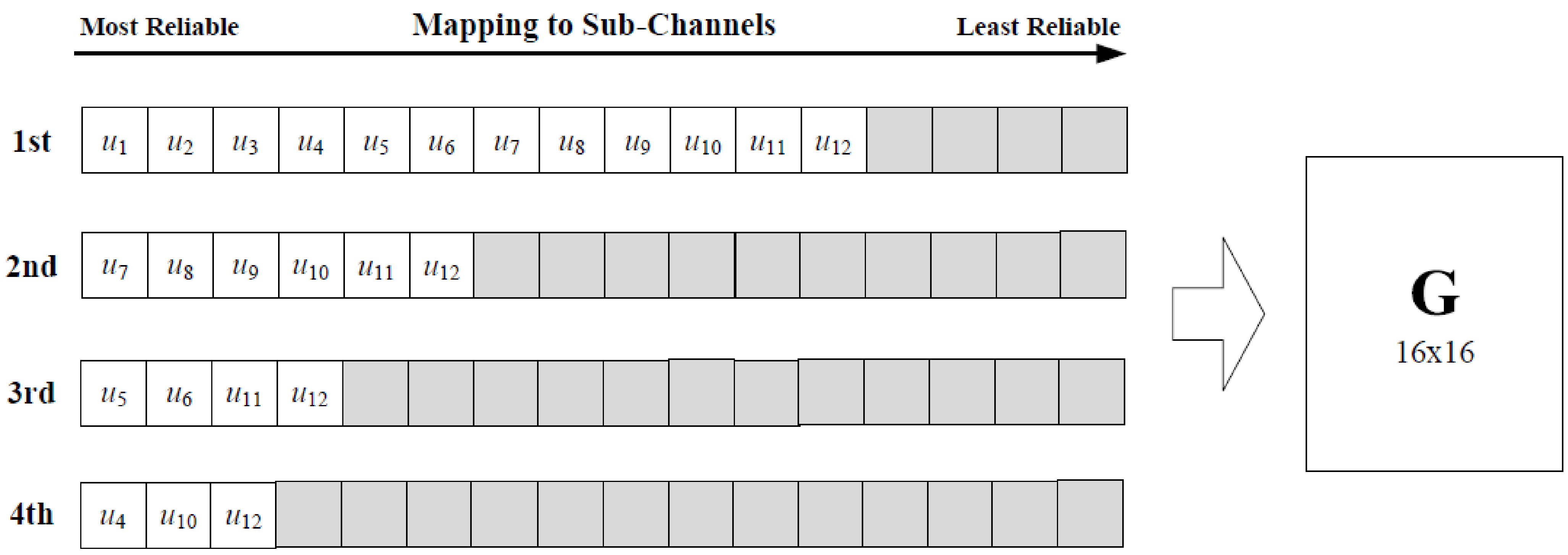


Figure from «Li, B. et al. (2016) 'Capacity-achieving rateless polar codes', in 2016 IEEE International Symposium on Information Theory (ISIT). 2016 IEEE International Symposium on Information Theory (ISIT), pp. 46–50. doi: 10.1109/ISIT.2016.7541258».

Outline

- Polar code performance
- Puncturing/Shortening/HARQ
- Some implementation Issues

Tb/s challenge: Back to the future?

- Before the age of VLSI, FEC theory was far ahead of technology
  - LCPC codes were too complex for 1960s
  - INTELSAT used Massey's Threshold Decoding
- Are we headed into a similar situation for high-throughput applications?
  - Silicon technology has stalled, while there is demand for ever increasing data rates
  - Will existing FEC schemes scale well for next generation systems that demand Tb/s throughput?
  - Are we going to have to sacrifice coding gain for high throughput?

FEC for Tb/s wireless

System Specifications

- Throughput : 1 Tb/s
- Coding gain : Best Effort
- Latency : Best Effort
- Flexibility : Best effort
- Rate : From 1/2 to 15/16
- Modulation : 4- to 64-QAM

Technology Constraints

- Technology : 7 nm
- Chip area : 10 mm^2
- Energy efficiency : 1 pJ/b
- Power density : 0.1 W/mm^2
- Clock speed : 1 GHz

Tb/s polar codes: Design considerations

- Since the clock speed is limited to 1 GHz, the decoder must produce on average 1000 bit decisions per clock tick to provide a 1 Tb/s throughput
- Successive Cancellation Decoding and its variants (including Belief Propagation) are inherently sequential: pipelined implementations needed
- Majority Logic Decoding provides parallelism but is suboptimal and more complex than SC decoding

References for high throughput decoders

- A. Balatsoukas-Stimming, A. J. Raymond, W. J. Gross, and A. Burg, “Hardware Architecture for List SC Decoding of Polar Codes,” *arXiv:1303.7127 [cs, math]*, Mar. 2013.
- P. Giard, G. Sarkis, C. Thibeault, and W. J. Gross, “A 237 Gbps Unrolled Hardware Polar Decoder,” *arXiv:1412.6043 [cs]*, Dec. 2014.
- Y. S. Park, Y. Tao, S. Sun, and Z. Zhang, “A 4.68Gb/s belief propagation polar decoder with bit-splitting register file,” in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, 2014, pp. 1–2.
- O. Dizdar and E. Arıkan, “A High-Throughput Energy-Efficient Implementation of Successive Cancellation Decoder for Polar Codes Using Combinational Logic,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 3, pp. 436–447, Mar. 2016.
- P. Giard *et al.*, “PolarBear: A 28 nm FD-SOI ASIC for Decoding of Polar Codes,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. PP, no. 99, pp. 1–1, 2017.

Pipelined Polar Decoder for Tb/s

- Designs are for length 1024 polar codes
- Decoding algorithms is a hybrid of SC and Majority Logic Decoding
- Work carrried out at Bilkent University and Polaran Ltd.

	Code Rate	**15/16**		**5/6**		**4/5**		**3/4**		**1/2**	
Component counts	Adders	37,558		51,539		52,772		53,556		59,871	
	XOR Gates	6,625		8,531		8,835		8,911		9,555	
	Multiplexers	98,317		132,459		137,324		150,380		172,377	
	Flip-Flops	720,000		1,067,000		1,117,000		1,098,000		1,074,000	
Gates	Number of gates	6,484,098		9,519,086		9,948,891		9,850,301		9,797,537	
	Logic / Storage	11 %	89 %	10%	90%	10%	90%	11%	89%	12%	88%
Area	mm^2	9.48		13.90		14.53		14.39		14.33	
	Logic / Storage	12%	88%	11%	89%	10%	90%	12%	88%	13%	87%
Power	Watt	1.31		1.92		2.01		1.99		1.98	
	Logic / Storage	11%	89%	11%	89%	10%	90%	11%	89%	13%	87%

Pipelined polar decoder scaled to 7 nm

Code Rate	15/16	5/6	4/5	3/4	1/2
Throughput (Gb/s)	1123	999	958	898	599
Area (mm^2)	0.2	0.26	0.27	0.3	0.3
Pow. Den. (W/mm^2)	4.7	4.72	4.72	4.7	4.7
Energy Eff. (pJ/bit)	0.7	1.21	1.32	1.4	2.1
Freq. (MHz)	1169	1169	1169	1169	1169

Pipelined polar decoder scaled to 7 nm with design adjustments

- Design adjustments made:
  - Increased area to reduce power density (use dark silicon)
  - Reduced clock speed to 1 GHz

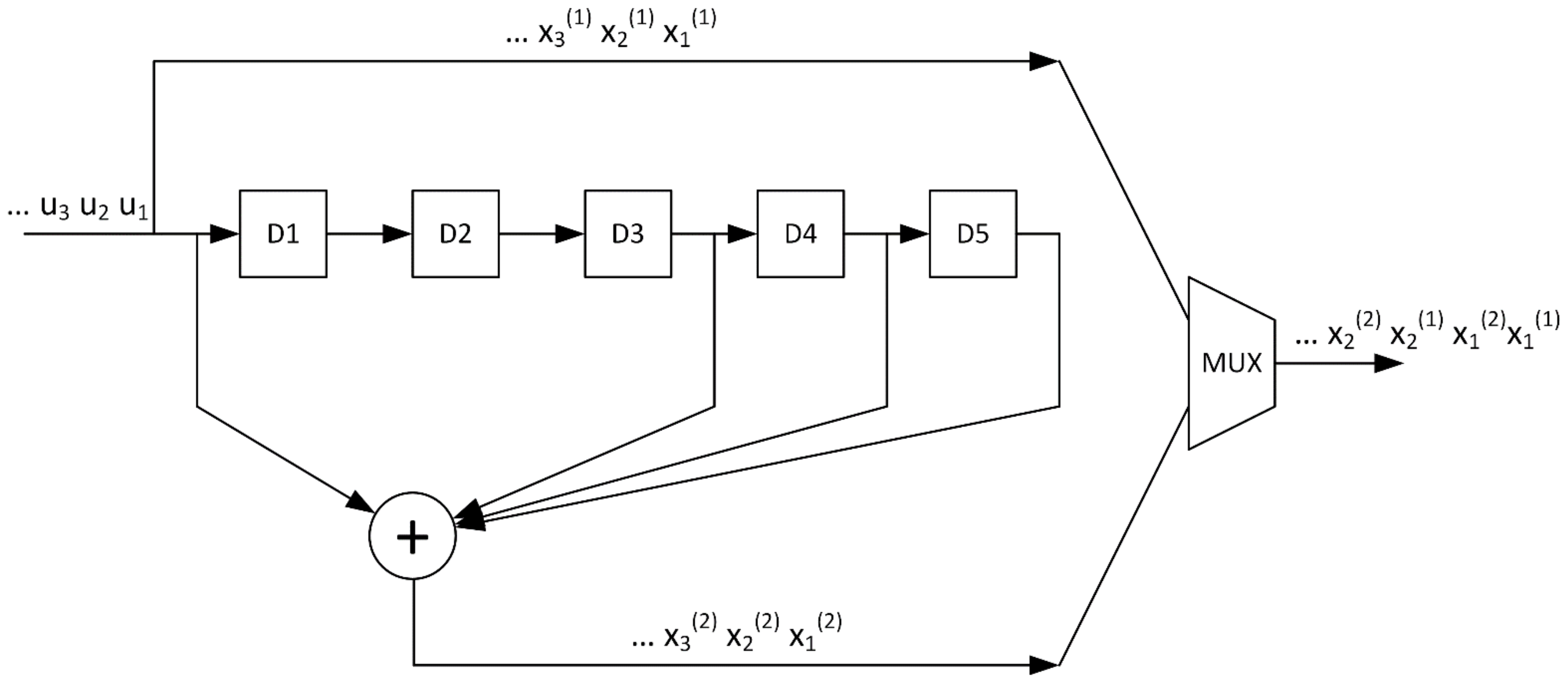
Code Rate	15/16	5/6	4/5	3/4	1/2
Throughput (Gb/s)	960.0	854.0	819.0	768.0	512.0
Area (mm^2)	10	10	10	10	10
Pow. Den. (W/mm^2)	0.1	0.1	0.1	0.1	0.1
Energy Eff. (pJ/bit)	0.7	1.6	1.6	1.4	2.1
Freq. (MHz)	1000	1000	1000	1000	1000

Lessons learned

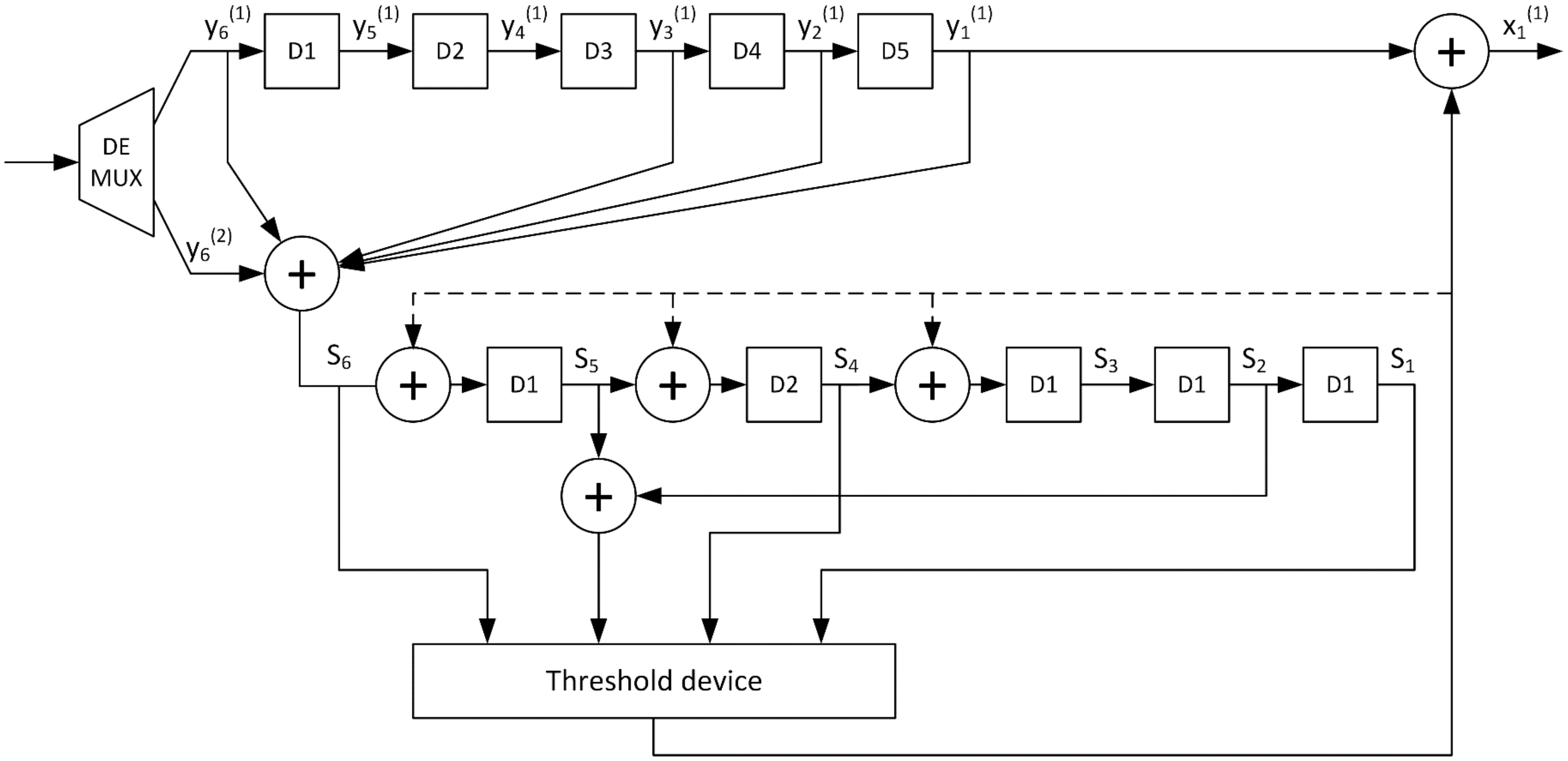
- As we scale polar codes to 7 nm for 1 Tb/s throughput
  - Power density emerges as the main design bottleneck
  - Power density can be brought under control by artificially inflating chip area
- Storage consumes 80% of area and power in a pipelined architecture
- Need a custom design FEC

Can old FEC techniques by a solution?

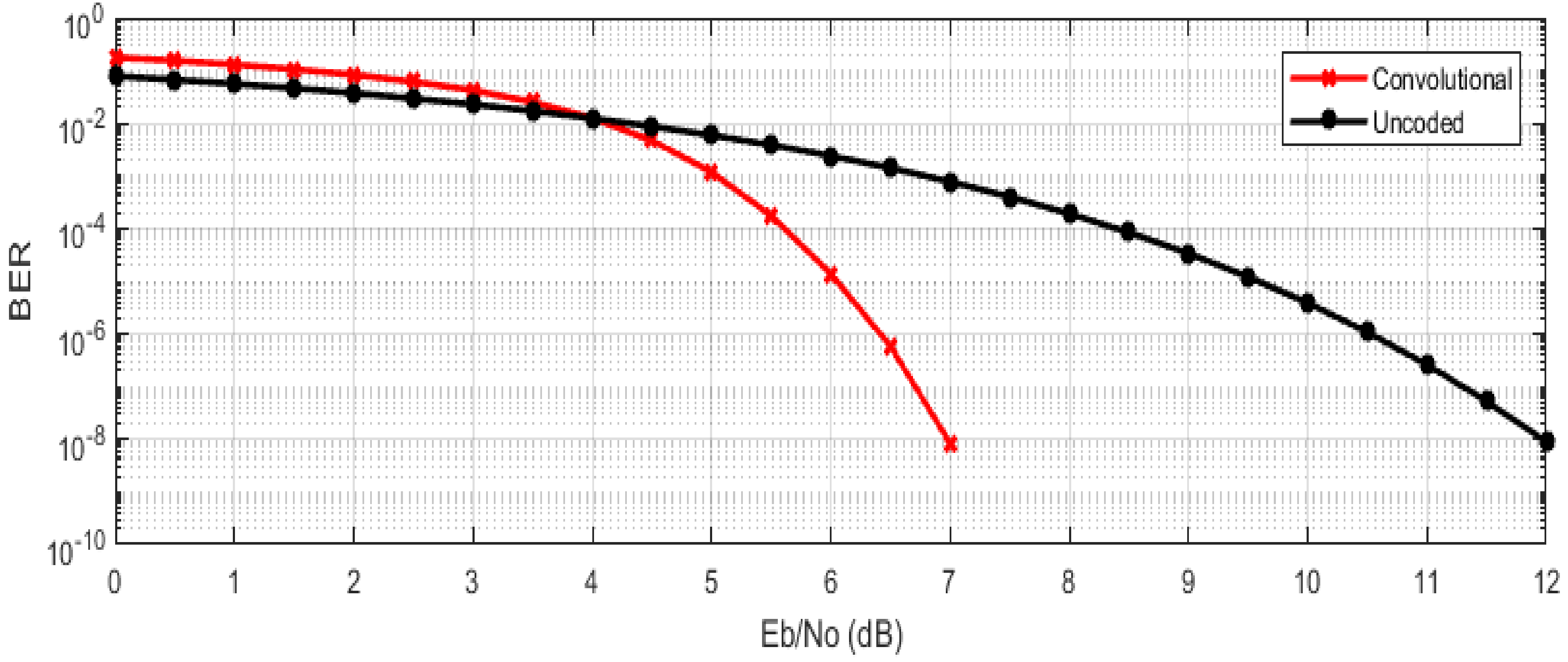
- In a recent study at Bilkent University, convolutional coding under threshold decoding was considered for Tb/s communications



Threshold decoder



BER Performance: 3.6 dB coding gain at 1E-5



Implementation of threshold decoder

Technology (nm)	350 nm (original design)	7 nm figures obtained by scaling	7 nm figures after area and clock adjustments
Througput (Gb/s)	0.13	3.0	1000
Area (mm2)	1.4E-02	1.2E-05	10
Power (W)	7.2E-03	3.2E-03	1.1
Area Eff. (Gb/s/mm2)	9.7	2.4E+05	100.0
Pow. Den. (W/mm2)	0.5	258.8	0.11
Energy Eff. (pJ/bit)	54.4	1.1	1.1
Latency (us)	2.3E-02	1.0E-03	6.0E-03
Freq. (MHz)	266.7	5920.7	1000

- Even this very simple decoder is challenged to meet the power density requirement

Thank you!

Acknowledgements

This work has been carried out by a support from the EPIC project with the funding from the European Union’s Horizon 2020 research and innovation programme under grant No. 760150.