Next-Generation Channel Coding Towards Terabit/s Wireless Communications

The European Union Horizon 2020 EPIC Project https://epic-h2020.eu/

Norbert Wehn
Microelectronic System Design Research Group
TU Kaiserslautern
67663 Kaiserslautern, Germany
Email: wehn@eit.uni-kl.de

Onur Sahin
InterDigital Europe
London, UK, EC2A 3QR
Email: Onur.Sahin@InterDigital.com

Abstract—The continuous demands on increased spectral efficiency, higher throughput, lower latency and lower energy in communication systems impose large challenges on the baseband processing in wireless communication. This applies in particular to channel coding (Forward Error Correction) that is a core technology component in any digital baseband. Future Beyond-5G use cases are expected to require wireless data rates in the Terabit/s range in a power envelope in the order of 1-10 Watts. In the past, progress in microelectronic silicon technology driven by Moore's law was an enabler of large leaps in throughput, lower latency, lower power etc. However, we have reached a point where microelectronics can no more keep pace with the increased requirements from communication systems. In addition, advanced technology nodes imply new challenges such as reliability, power density, cost etc. Thus, channel coding for Beyond-5G systems requires a real cross layer approach, covering information theory, algorithm development, parallel hardware architectures and semiconductor technology. The EPIC project addresses these challenges and aims to develop new Forward Error Correction (FEC) schemes for future Beyond-5G use cases targeting a throughput in the Tb/s range. Focus will be on the most advanced FEC schemes, i.e. Turbo codes, Low Density Parity Check (LDPC) codes and Polar codes [1].

Keywords—Forward Error Correction, Beyond-5G, Terabit/s throughput

I. INTRODUCTION

The available power budget for baseband processing in state-of-the-art transceivers is typically limited to some Watts. E.g., if we target a throughput of $1\,\text{Tb/s}$ in a $1\,\text{W}$ power envelope for the FEC Intellectual Property (IP), only $1\,\text{pJ}$ energy is available to process a single bit. In comparison, in $28\,\text{nm}$ technology, a single 64-bit double precision floating point operation consumes about $20\,\text{pJ}$ as well as transferring 256 bits over a $40\,\text{mm}$ wire on a chip costs about $10\,\text{nJ}$. In the foreseeable future, progress in microelectronics will yield some improvements in the three important metrics, i.e. area, energy, and frequency. Extrapolating down from $28\,\text{nm}$ to $7\,\text{nm}$ technology will bring approximately a factor of 12x reduction in area and a factor of 4x improvement in energy efficiency [2][3]. Moreover due to the fact that area density increases faster than the improvement in power, power density will be one of the largest challenges and is already a big issue in today's technologies, which is known as "dark silicon phenomenon" [4]. The maximum frequency is expected

to improve up to 3x. However, the maximum frequency in baseband IPs has an upper limit in the order of $1\,\text{GHz}$ due to power and design issues. Achieving a throughput of $1\,\text{Tb/s}$ with a maximum frequency of $1\,\text{GHz}$ results in about 1000 information bits that have to be decoded in each clock cycle which demands for extreme parallelism. Thus, for Beyond-5G FEC, improvement from new silicon technology has to be complemented with improvements on the code level, the decoding algorithms and the architectural level.

II. IMPLEMENTATION CHALLENGES

There is a fundamental discrepancy between information theory objectives and efficient implementation objectives [5]. Advanced channel codes like Turbo codes and LDPC codes combine randomness with very limited locality, i.e. interleaver for Turbo codes and Tanner graph for LDPC codes, and some structures with iterative/sequential decoding techniques to achieve near channel capacity. On the implementation side, however, large locality, large regularity and large parallelism are mandatory to obtain energy efficient, high-throughput architectures. To bridge the gap between information theory and efficient implementation EPIC pursues a holistic approach in which information theory, code design, algorithms/architecture co-design, front-end and back-end implementation optimization are considered together. This approach is shown in Figure 1.

III. EPIC APPROACH

Figure 2 exhibits the overall methodology of EPIC and the corresponding work flow. This work flow is partitioned into 5 basic steps. Step 3 is EPICs core step in which the aforementioned holistic approach is pursued for the three different advanced coding schemes, i.e. Turbo codes, LDPC codes and Polar codes.

The investigation starts with the definition of use cases and corresponding Key Performance Indicators (KPIs). EPIC identified a set of Beyond-5G use cases by analyzing state-of-the-art literature and current standardization efforts. In total eight Beyond-5G use cases were identified. We considered only use cases that are particularly interesting and challenging in the context of advanced FEC technologies. The selected use cases are data kiosk, virtual reality, intra-device communication,
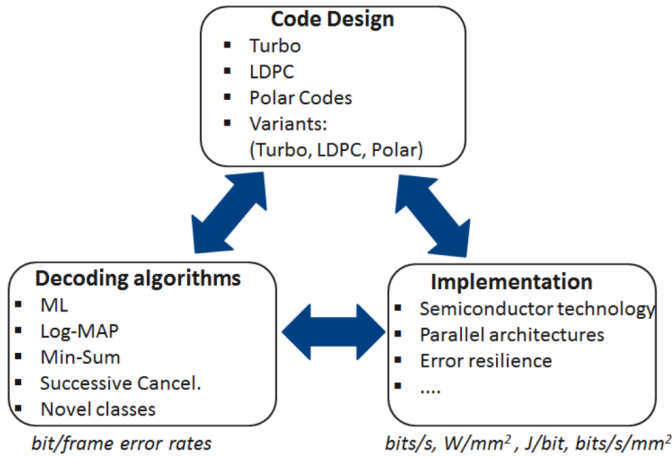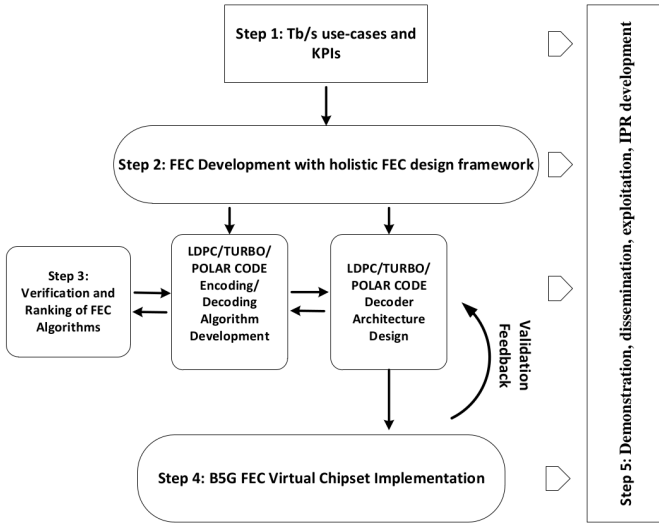
Fig. 1. EPIC FEC design framework

Fig. 2. EPIC methodology steps and work flow

technology nodes: $28\,\mathrm{nm}$ that is state-of-the-art technology for many published FEC IPs and a projected advanced $7\,\mathrm{nm}$ technology node. The corresponding implementation KPIs for the two technology nodes are shown in Table II.

	Bit Error Rate (BER)	Flexibility	Latency	Throughput [Gbit/s]
Data Kiosk	$< 10^{-12}$	Low	0.5 ms	1000
Virtual Reality	$< 10^{-6}$	high	0.5 ms	500
Intra-device	$< 10^{-12}$	low	100 ns	500
Fronthaul	$< 10^{-13}$	medium	25 ns	1000
Backhaul	10^{-8}	medium	100 ns	250
Data Center	$< 10^{-12}$	medium	100 ns	1000
Wireless Fiber	10^{-12}	medium	200 ns	1000
HT Satellite	10^{-10}	medium	10 ms	100-1000

TABLE I.   COMMUNICATION/SYSTEM KPIs FOR EPIC USE CASES.

	Area efficiency [Gbit/s/mm^2]		Power density [W/mm^2]		Energy efficiency [pJ/bit]	
	28nm	7nm	28nm	7nm	28nm	7nm
Data Kiosk	100	220	0.09	0.20	0.90	0.90
Virtual Reality	50	54	0.02	0.03	0.48	0.48
Intra-device	50	50	0.13	0.50	1.00	1.00
Fronthaul	100	100	0.17	0.06	0.60	0.60
Backhaul	25	25	0.09	0.09	3.60	3.60
Data Center	100	162	0.20	0.12	0.75	0.75
Wireless Fiber	100	120	0.23	0.14	1.13	1.13
HT Satellite	100	n/a	0.27	n/a	0.50	0.50

TABLE II.   IMPLEMENTATION KPIs FOR SEVEN EPIC USE CASES.

Based on this analysis and the resulting implementation KPIs, the following targets were fixed for the EPIC project: meet the communication performance requirements for Beyond-5G use cases and at the same time, an area efficiency of $100\,\mathrm{Gbit/s/mm^2}$, an energy efficiency of about $1\,\mathrm{pJ/bit}$ and a power density in the order of $0.1\,\mathrm{W/mm^2}$ for a $7\,\mathrm{nm}$ technology node.

IV. HIGH-THROUGHPUT DECODERS

To achieve a throughput of $1\,\mathrm{Tb/s}$ parallelism has to be exploited at all levels [6]. The three prominent channel codes, namely Turbo codes, LDPC codes and Polar codes largely differ in this respect: Turbo code decoding is inherently serial and is mainly performed on data-flow graphs; LDPC code decoding is inherently parallel and is mainly performed on a data-flow graph; Polar code decoding is performed on a traversal of a code tree structure. Thus, the challenges for high-throughput decoder implementations are fundamentally different between Turbo codes, LDPC codes and Polar codes resulting in a large difference in the achievable throughput and energy efficiency of today's state-of-the-art decoders. The two most prominent techniques to achieve high- throughput on architectural level are spatial parallelism and functional parallelism (pipelining). Pipelining has some efficiency advantages, i.e. large locality, compared to spatial parallelism, but is limited in its applicability if control-flow, e.g. iteration control, plays a major role. In addition, in pipelined architectures throughput is increased at the cost of additional storage since several blocks have to be kept simultaneously in the pipeline stages. Thus deeply pipelined architectures can suffer from large storage requirements that are a major source of power consumption. Channel decoding algorithms are mainly data-flow dominated.
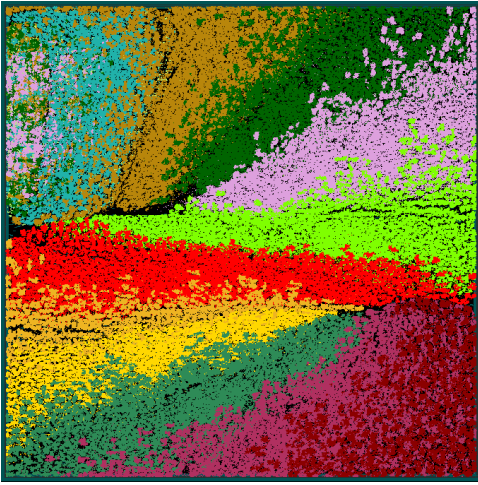
wireless fronthaul/backhaul, data center, hybrid fiber-wireless networks, and high throughput satellites. Collectively, these use cases present a diverse set of FEC design challenges. For all use cases key performance indicators were determined. These KPIs are divided into two sets: communication/system performance related KPIs, i.e. BER/FER, flexibility, throughput, latency, and implementation related KPIs, i.e. area, power, and resulting area efficiency, power density and energy efficiency. For each use case we analyzed the available cost budget, available power budget for the FEC IP and chip volume from a market driven perspective. The corresponding area implementation KPI results from the correlation of cost budget and volume with wafer and mask costs for specific technology nodes. If the cost budget permits larger areas than $10\,\mathrm{mm^2}$, we limit the area to $10\,\mathrm{mm^2}$, that is feasible from an IP perspective. The other implementation KPIs are derived in a similar way. Obviously this method is only an estimate and neglects some factors, but it gives some quantitative indications on the FEC challenges for future systems. Table I shows the resulting communication/system KPIs for the seven use cases. Corresponding implementations KPIs were calculated for two

Fig. 3. Layout LDPC code decoder in $28\,\mathrm{nm}$. The area is $2.8\,\mathrm{mm}^2$.

Fig. 4. Layout Polar code decoder in $28\,\mathrm{nm}$. The area is $4.3\,\mathrm{mm}^2$.

Thus, to achieve throughput far beyond $100\,\mathrm{Gbit/s}$, dataflow and tree structures must be flattened, or "unrolled", and pipelined respectively [7][8]. In the following we present two high-throughput decoders that are based on this principle.

The first decoder is an LDPC decoder supporting the 802.11ad standard having a block size of 672 bits and code rate 13/16. Decoding is performed with the min-sum algorithm and 9 iterations. These iterations are "unrolled" and pipelined. To shorten the critical path three additional pipeline stages are added in each iteration, resulting in overall 27 pipeline stages. The decoder was synthesized on a $28\,\mathrm{nm}$ Fully Depleted Silicon on Insulator (FD-SOI) technology under worst case Process, Voltage and Temperature (PVT) assumption. The corresponding layout is shown in Figure 3. The total area is $2.8\,\mathrm{mm}^2$. One recognizes the dataflow architecture in which each color represents an iteration pipeline stage. The decoder was characterized for different voltages. At $0.9\,\mathrm{V}$ and a frequency of $220\,\mathrm{MHz}$, a throughput of $160\,\mathrm{Gbit/s}$, power consumption of $960\,\mathrm{mW}$ and energy efficiency of 6pJ/bit can be achieved. Outcome at $0.6\,\mathrm{V}$ are $70\,\mathrm{Gbit/s}$, $210\,\mathrm{mW}$ and $3\,\mathrm{pJ/bit}$. Extrapolating these numbers to $7\,\mathrm{nm}$ gives a maximum throughout of $480\,\mathrm{Gbit/s}$ and $1.5\,\mathrm{pJ/bit}$.

The second decoder is a Polar code decoder with block size 1024 and code rate 1/2. Decoding is performed with the successive cancellation algorithm. This decoder was synthesized on the same technology and under the same conditions as the LDPC decoder. The corresponding layout is shown in Figure 4, area is $4.3\,\mathrm{mm}^2$. In total 385 nodes have to be visited during the polar factor tree traversal when running the successive cancellation algorithm. This results theoretically in 385 pipeline stages. Each color in the layout represents the corresponding logic block of such a pipeline stage. Since the bitwidth of the different pipeline stages largely vary dependent on the tree level, i.e., the bitwidth of the root node is n, the width of its child nodes $n/2$ etc., the blocks have different sizes. To minimize the storage requirements and the power consumption, the decoder is unlike the LDPC decoder not deeply pipelined but only partially pipelined resulting in only 99 pipeline stages. At $0.9\,\mathrm{V}$ and a frequency of $606\,\mathrm{MHz}$, a throughput of $620\,\mathrm{Gbit/s}$, power consumption of $2.8\,\mathrm{W}$ and energy efficiency of $4.6\,\mathrm{pJ/bit}$ can be achieved under worst case PVT conditions. Extrapolating these numbers to $7\,\mathrm{nm}$ with a maximum frequency of $1\,\mathrm{GHz}$ results in an energy efficiency of less than $2\,\mathrm{pJ/bit}$.
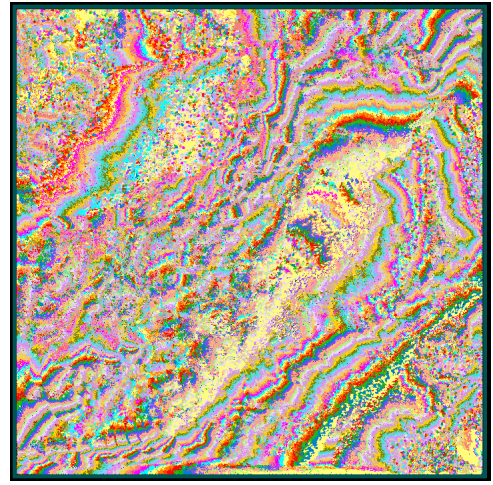
V. CONCLUSION

The presented decoder architectures demonstrate that unrolled dataflow architectures allow very high throughput approaching the 1Tb/s wall in $7\,\mathrm{nm}$ technology. However, there are trade-offs between communication performance and high throughput that depend on the particular use case. High communication performance typically requires complex decoding algorithms to achieve near maximum-likelihood performance and large block lengths to approach the Shannon bound. But unrolling under 1GHz frequency and area constraints is only feasible for smaller block lengths, limits the complexity of the decoding algorithms. Furthermore, unrolled architectures suffer on flexibility w.r.t. code rate, block length, iteration control etc. So the biggest implementation challenge is achieving the requirements on energy efficiency and power density while maintaining communication performance and the necessary flexibility required by the use cases. The EPIC project is dedicated to tackle these challenges by utilizing its holistic implementation-aware channel code design framework and develop one of the first Beyond-5G FEC technology solutions.

REFERENCES

[1] E. Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009.

[2] O. Villa, D. R. Johnson, M. Oconnor, E. Bolotin, D. Nellans, J. Luitjens, N. Sakharnykh, P. Wang, P. Micikevicius, A. Scudiero, S. W. Keckler, and W. J. Dally. Scaling the Power Wall: A Path to Exascale. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 830–841, Nov 2014.

[3] ITRS 2.0. International Technology Roadmap for Semiconductors, 2015 Edition, Section 5: More Moore.

[4] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger. Dark silicon and the end of multicore scaling. *IEEE Micro*, 32(3):122–134, May 2012.

[5] S. Scholl, S. Weithoffer, and N. Wehn. Advanced iterative channel coding schemes: When Shannon meets Moore. In *2016 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, pages 406–411, Sept 2016.

[6] S. Weithoffer, M. Herrmann, C. Kestel, and N. Wehn. Advanced wireless digital baseband signal processing beyond 100 gbit/s. In *2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, pages 1–6, Oct 2017.

[7] P. Schläfer, N. Wehn, M. Alles, and T. Lehnigk-Emden. A new dimension of parallelism in ultra high throughput LDPC decoding. In *SiPS 2013 Proceedings*, pages 153–158, Oct 2013.

[8] P. Giard, G. Sarkis, C. Thibeault, and W. J. Gross. Multi-mode unrolled architectures for polar decoders. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 63(9):1443–1453, Sept 2016.