IMT Atlantique / Département MEE
Laboratoire Lab-STICC / équipe 2AI
Technopole Brest Iroise – CS 83818 – 29238 Brest – France

Project: PhD Thesis

Domain: Hardware design, Channel coding, digital communications

Topic: Hardware implementation of fully pipelined Turbo decoders

Duration: 36 months

Start: from 01.09.2021

Supervisor: Stefan Weithoffer

Co-Supervisor(s): Charbel Abdel Nour, Catherine Douillard

Contact: firstname.lastname@imt-atlantique.fr

A. Context

The information society we live in today relies on forward-error-correction (FEC) techniques for point-to point communications. In the wireless domain, FEC based on Turbo codes enabled the mobile internet and ubiquitous video streaming with third and fourth generation (3G/4G) wireless communication systems. Indeed, for standards like UMTS, which only required a throughput of 1Mb/s in its 1999 release, to LTE, which required 100Mb/s in the 2008 release, Turbo codes with their excellent error correcting performance for the targeted error rates, built-in code rate flexibility and ease of encoding were a perfect match. Turbo codes are also used in current and future releases of LTE-A Pro (4G+), which requires a throughput of several Gb/s and support for almost 200 different frame sizes and a wide range of code rates. However, the widespread use of Turbo codes and the low level of targeted rates led to only a few decoding algorithms being considered for practical use cases. The BCJR/MAP algorithm and its variants in the logarithmic domain represent the standard algorithms used in software and hardware implementations alike [1, 2, 3]. This textbook knowledge has not been significantly challenged in almost 30 years [4]. The MAP algorithm suffers from significant limitations through the inherently serial recursive calculations along the code trellis. These impose a lower boundary on the computational complexity of the algorithms and also limit the achievable parallelism for high throughput and low latency implementations in both hardware and software.

On another note, the iterative decoding principle introduced with Turbo codes played a significant role in the re-discovery of Low Density Parity Check (LDPC) codes [5]. Their advent coincided with the need for higher throughput. It seemed then natural to resort to LDPC codes for high throughput applications taking into account the ease of achieving high degrees of parallelism with their corresponding decoders. As a consequence of the absence of major novelty for a long period of time and the availability of an alternative solution (LDPC codes) for high throughput applications a common belief was set that Turbo codes no longer hold any mystery for the experts of the field and that there is no room for overcoming their above-listed drawbacks. An example is the 5G New Radio (NR) [6], where LDPC codes were chosen over Turbo codes despite advantages still offered by Turbo codes in terms of error correcting performance, especially for short frame sizes and severe channel conditions. Moreover, Turbo codes need little to no overhead to support a wide range of frame size and code rates.

Indeed, one of the outcomes of the recently finished EU project EPIC [7] targeting ultra-high throughput coding, where we were actively involved, clearly states that there is no silver bullet for the challenging requirements posed by todays and upcoming communication systems Beyond 5G (B5G). Applications like large-scale virtual/augmented reality, holographic calls, etc., which are expected to be enabled by B5G communication systems, require throughputs in the order of Tb/s. Today, there is no code family that is able to achieve jointly ultra-high throughput, low latency, reasonable chip area and power consumption, excellent error correction capabilities while providing the level of support for frame size and code rate flexibilities required for practical applications. The higher area efficiency of the contending high-throughput LDPC and Polar decoders comes at the cost of a deficit in flexibility with respect to

code rate and frame size. Most of the Turbo decoders either are highly flexible with respect to frame sizes, or can be made flexible [8, 9]. However, when supported with the classical MAP decoding algorithm [1, 10], this flexibility represents one of the main factors that lead to the reduced efficiency of Turbo decoders. The consequences are two-fold: first, there is a high potential for lowering Turbo decoder complexity for applications with limited flexibility requirements and second, for applications requiring high flexibility, Turbo codes can be clearly superior to LDPC or Polar codes if the factors limiting the achievable throughput for hardware implementations can be overcome.

To set Turbo codes, invented in Brest 30 years ago by Claude Berrou [10], again on the path of being THE leading technology, their decoding algorithms and decoder hardware architectures need to be re-invented. The road to Tb/s Turbo decoders goes through the joint introduction of profound modifications to their decoding algorithm and to their code design while taking into account implementation and application constraints which should be possible helped by the advances in the field related to the other major code families. This disruptive task is also motivated by the scientific and the political aspects targeting jointly scientific leadership and French/European sovereignty over major technologies.

B. Positioning and recent works/results

In recent works, we started to tackle this challenging and disruptive task. In [4], we proposed the first fully pipelined iteration unrolled Turbo decoder and presented a hardware implementation achieving for the first time more than 100 Gb/s. The implementation which was placed and routed in 28 nm technology still required an area of almost 24 mm^2. Therefore, we proposed in recent works [11, 12] to combine different methods of parallelization which enabled us to reduce the area consumption to 15 mm^2 for the same throughput and frame size (K=128 bits), an improvement of up to two orders of magnitude in throughput for Turbo decoder hardware implementations over state-of-the-art. The implementation from [11] also shows that a significant degree of frame-size flexibility can be retained for fully pipelined Turbo decoders.

While the above works are still based on the MAP algorithm, we also introduced a new trellis-based decoding algorithm, the Local-SOVA (LSOVA) algorithm [13]. It employs a forward and backward recursion like the conventional max-Log-MAP (MLM) algorithm, but features a new way of computing the extrinsic information. This new way is based upon a new path-merging operation which leads to a hierarchical structure and allows for a reduction in the computational complexity for high radix orders. Decoding with a higher radix is an essential technique for increasing the decoder throughput, since several consecutive trellis steps are processed in parallel. This leads to a reduction in decoding latency and in state metric memory, which in turn, increases the decoder throughput and area efficiency for conventional Turbo decoders (PMAP, XMAP). Therefore, a high radix decoding is highly desirable. However, the increased complexity and long critical path for higher radix orders in the past made radix orders ≥8 inefficient and Turbo decoders typically use radix-2 or radix-4 computations. In [13], a computational complexity reduction of around 30% w.r.t. add- and compare-operations for LSOVA is shown for radix orders 4 and 8 on the example of the (8-state) LTE code. Motivated by these large complexity reductions, we recently proposed hardware architectures for the computational units of the LSOVA algorithm in [14]. They apply additional complexity reductions to the LSOVA algorithm leading to efficient hardware architectures for the Add-Compare-Select-Units (ACSU) and Soft-Output-Units (SOU). These works gave, for the first time, implementation results for the LSOVA computational units at a 22nm technology. The results not only confirm the projected complexity reductions from [15] which were based on a mere counting of arithmetic operations, but exceed the complexity reduction projections. Indeed, up to 75% area savings for the SOU and approximately 45% for the overall comparison with the respective MLM units were observed.

The abovementioned algorithmic simplifications and hardware architectures, which we proposed during the EPIC project, represent only a short subset of the ideas that were identified with high technical potential and as serious candidates for patent filings. These ideas are planned as future works within this PhD topic.

C. Topic, environment and outcomes

Building on initial results on the hardware implementation of new decoding algorithms for Turbo decoding, the ANR funded JCJC project "TurboLEAP", which is set to start 01.03.2021 and led by Stefan Weithoffer, will investigate the design of ultra-high throughput hardware architectures for the decoding of Turbo codes, focusing mainly on hardware design aspects. Within TurboLEAP, the PhD candidate is expected to play a significant role in the original approach applied traditionally by the members of the Mathematical and Electrical Engineering (MEE) department of IMT Atlantique consisting of a close interaction between Silicon and Algorithm design. Contrary to classical procedures, in this approach, algorithms are thought and derived from the start to consider at equal footing the hardware implementation and performance related constraints. Indeed, the candidate is expected to explore the full potential of fully pipelined iteration unrolled Turbo decoder hardware architectures. Throughputs of more than 1 Tb/s for large frame sizes with several thousands of information bits are targeted. Furthermore, decoder hardware architectures for Spatially coupled Turbo codes shall be investigated [16].

Under the supervision of well-established researchers in the field of Turbo coding/decoding and hardware implementations, mainly Prof. Catherine Douillard and associate professors Charbel Abdel Nour and Stefan Weithoffer, he will work in close collaboration with a PhD candidate planned to be recruited in the context algorithm design. Furthermore, a close collaboration with the Post-doctoral research fellow to be recruited for the TurboLEAP project and working both on the design of new Turbo codes and algorithms tailored for fully pipelined hardware architectures, is expected. Together, the assembled team shall achieve the required level of synergy between the simplified design of decoding algorithms and their impact on hardware implementations.

D. Financing

The thesis financing (36 months) is provided through the ANR funded JCJC TurboLEAP project.

E. Candidate Profile

The candidate should hold a Master degree or an engineering degree in digital communications. The following qualifications are beneficial for the completion of the project:

- Experience with one or more of the following languages: C, C++, Python
- Experience with one of the following hardware description languages: VHDL, Verilog, SystemC
- Experience with hardware development on FPGA and/or ASIC
- Advanced lectures on channel coding and communication systems

F. References

[1] L. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (Corresp.)," *IEEE Transactions on Information Theory,* vol. 20, no. 2, pp. 284-287, 1974.

[2] B. L. Gal and C. Jego, "Low-latency and high-throughput software turbo decoders on multi-core architectures," *Annales Des Télécommunications,* vol. 75, no. 1, pp. 27-42, 2020.

[3] S. Weithoffer, F. Pohl and N. Wehn, "On the applicability of trellis compression to Turbo-Code decoder hardware architectures," in *9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, 2016.

[4] S. Weithoffer, C. A. Nour, N. Wehn, C. Douillard and C. Berrou, "25 Years of Turbo Codes: From Mb/s to beyond 100 Gb/s," in *IEEE 10th International Symposium on Turbo Codes & Iterative Information Processing (ISTC)*, 2018.

[5] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronics Letters,* vol. 33, no. 6, pp. 457-458, 1996.

[6] A. G. W. Yong, "Draft LTI on 3GPP 5G initial description template," 2017.

[7] EPIC, "Enabling practical wireless tb/s communications with next generation channel coding (EPIC)," [Online]. Available: https://epic-h2020.eu/. [Accessed 27 01 2021].

[8] T. Ilnseher, F. Kienle, C. Weis and N. Wehn, "A 2.15GBit/s turbo code decoder for LTE advanced base station applications," in *7th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, 2012.

[9] C. Roth, S. Belfanti, C. Benkeser and Q. Huang, "Efficient Parallel Turbo-Decoding for High-Throughput Wireless Systems," *IEEE Transactions on Circuits and Systems,* vol. 61, no. 6, pp. 1824-1835, 2014.

[10] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proceedings of ICC '93 - IEEE International Conference on Communications*, 1993.

[11] S. Weithoffer, O. Griebel, R. Klaimi, C. Abdel Nour and N. Wehn, "Advanced Hardware Architectures for Turbo Code Decoding Beyond 100 Gb/s," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, 2020.

[12] S. Weithoffer, R. Klaimi, C. Abdel Nour, N. Wehn and C. Douillard, "Fully Pipelined Iteration Unrolled Decoders the Road to TB/S Turbo Decoding," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.

[13] V. H. S. Le, C. A. Nour, E. Boutillon and C. Douillard, "Revisiting the Max-Log-Map Algorithm With SOVA Update Rules: New Simplifications for High-Radix SISO Decoders," *IEEE Transactions on Communications,* vol. 68, no. 4, pp. 1991-2004, 2020.

[14] S. Weithoffer, R. Klaimi, C. A. Nour, N. Wehn and C. Douillard, "Low-complexity Computational Units for the Local-SOVA Decoding Algorithm," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020.

[15] V. H. S. Le, C. A. Nour, C. Douillard and E. Boutillon, "A Low-Complexity Dual Trellis Decoding Algorithm for High-Rate Convolutional Codes," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020.

[16] S. Moloudi, M. Lentmaier and A. Graell i Amat, "Spatially Coupled Turbo-Like Codes: A New Trade-Off Between Waterfall and Error Floor," *IEEE Transactions on Communications,* vol. 67, no. 5, pp. 3114-3123, 2019.