

# A 100 Gbps LDPC Decoder for the IEEE 802.11ay Standard

Meng Li, Veerle Derudder, Claude Desset,  
Andy Dewilde, André Bourdoux  
imec, Kapeldreef 15, B-3001 Heverlee, Belgium

Yanxiang Huang  
Nvidia, 2788 San Tomas Expy,  
Santa Clara, CA 95051, USA

**Abstract**—IEEE 802.11ay is the amendment to the 802.11 standard that enables Wi-Fi devices to achieve 100 Gbps using the unlicensed mm-Wave (60 GHz) band at comparable ranges to today’s commercial 60 GHz devices based on the 802.11ad standard. In this paper, we propose a full row-based layered LDPC decoder supporting all the coding rates for 802.11ay. Taking the property of the parity check matrix of 802.11ay, combining multiple layers into single layer improves the hardware utilization hence increases the throughput. The throughput is further increases by interleaving multiple frames to improve the utilization of each pipeline stage. The decoder is synthesized at both 28 nm and 16 nm CMOS technology and power estimated with stimuli at 7db and 3.5db. The 28 nm implementation running at 600 MHz and achieves a throughput of 67 Gbps for coding rate 13/16 at 4 iterations with area efficiency of 160 Gbps/sqmm and consumes an average power consumption of 408 mW and 141 mW, yielding energy efficiency of 6.05 pJ/bit and 2.1 pJ/bit at 3.5db and 7db. The 16 nm implementation running at 1 GHz and achieves a throughput of 112 Gbps at 4 iterations with area efficiency of 589 Gbps/sqmm and consumes an average power of 408mW and 163mW, yielding energy efficiency of 3.64 pJ/bit and 1.45 pJ/bit at 3.5db and 7db.

## I. INTRODUCTION

Low-density parity-check (LDPC) codes [1] [2] [3] play an important role in the wireless communication domain, since the first adoption in the DVB-S2 standard in 2003. This is due to the near Shannon limit error correction capability and quite mature hardware implementations.

The 60 GHz spectrum is known as the oxygen absorption band. That means radio waves at those frequencies are actually degraded by the presence of oxygen in the air. For this reason, 60 GHz was considered only appropriate for point-to-point, outdoor applications using highly-directional antennas (e.g., wireless links between two networks) until recently. IEEE 802.11ad for the Gigabit-wireless local area network (WLAN) and IEEE 802.15.3c for the wireless personal area network (WPAN) are multi-gigabit data rate standards working in the 60 GHz band.

While enabling multi-Gbps wireless local communications was a significant achievement, throughput and reliability requirements of new applications, such as augmented reality/virtual reality (AR/VR) and wireless fronthauling/backhauling, exceed what 802.11ad can offer. To meet the demanding requirements of such diverse applications, the IEEE 802.11 Task Group ay (802.11ay) was formed in 2015 to define PHY and MAC amendments to the 802.11 standard

that enable Wi-Fi devices to achieve 100 Gbps using the unlicensed mm-Wave (60 GHz) band at comparable ranges to today’s commercial 60 GHz devices based on the 802.11ad standard. Where 802.11ad uses a maximum of 2.16 GHz bandwidth, 802.11ay bonds four of those channels together for a maximum bandwidth of 8.64 GHz. MIMO is added with a maximum of 8 streams. The link rate per stream is 44 Gbps with 256-QAM, with four streams this goes up to 176 Gbps.

In recent years, there are many published designs of multi-Gbps LDPC decoders. However, designing an LDPC decoder with 100+ Gbps throughput, low power consumption and low area cost while maintaining good BER performance has been a big challenge.

The deep scaled CMOS technology brings significant implementation gains for digital circuits design [4] in terms of speed, area and power consumption. It serves as a major enabler of wireless communication at higher data rates with improved energy and area cost. The commercial 16 nm Fin-FET technology from TSMC [5] claims around 50% speed gain under same power consumption when compared to the 28 nm node. However, technology scaling alone is not enough to cover the increased throughput requirements. Code innovations and new architectures are the key driver to make ultra-high speed LDPC decoders.

There is a multi-dimensional design space for high speed LDPC decoders [6], such as decoding scheduling, optimization of check node calculation, parallelism extensions, pipeline stage optimization, frame level pipeline, etc. To design a high throughput decoder, the designer should consider these multi-dimensional optimizations. The optimization requires a trade-off in performance, power and area, depending on the application.

A fully parallel LDPC decoder [7] can reach the highest throughput. The state of art (SOA) CMOS technology is able to support this kind of architecture for short code length (<1000). However, as the code length increases, long interconnects create routing congestions that limit an efficient design and result in a large decoder with low area utilization, poor timing and power consumption.

For architecture-aware (AA) LDPC codes, a partial row-based parallel architecture [8] is a good alternative. It reduces the routing congestion problem while increasing the throughput when compared to a block serial decoding [9].

The SOA multi-gigabit LDPC decoders [10][11] apply a

row-based scheme, extending the bit node parallelism to the code length while keeping the check node parallelism same as the size of the sub-matrix. This achieves high throughput while reducing complexity compared to a fully parallel architecture.

The unrolling architecture [12] is based on a fully parallel architecture and further increases the decoding parallelism by unrolling the iteration loop. Dedicated hardware is added  $N$  times that unrolls the iterations in a two-phase decoding architecture, where  $N$  equals to the maximum iteration number. The unrolled architecture is pipelined, to allow processing one block per clock cycle. This architecture also reduces the routing congestion compared to classical fully parallel architectures. The unrolling architecture significantly increase the throughput and opens a door for 100+ Gbps FEC [13] [14]. However, this architecture lacks the flexibility in coding rates and early stop capability.

In this work, we propose a full row-based layer LDPC decoder with frame interleaved scheduling for the 802.11ay standard. Ultra high throughput is achieved using frame interleaved and highly compressed decoding scheduling making best use of hardware resources. For example, it only takes 3 cycles to finish message passing for 4 layers for the 1/2 rate. Different from the unrolling architectures, the decoding parallelism is increased by changing message update scheduling and make the best utilization of a row-based layer LDPC decoder, hence the early stop and flexibility in rates are inherited from row-based layer LDPC decoder. The proposed architecture is verified in 28 nm CMOS as well as in 16 nm FinFET technology. With these advanced technologies, the proposed decoder can reach 100 Gbps throughput at clock rate of 1 GHz and shows great implementation results in both area and energy efficiency.

The rest of the paper is organized as follows. In section II, the LDPC codes in the 802.11ay standard are introduced, followed by a brief introduction of layered decoding algorithm. Section III details the proposed frame level pipelined full row-based layer architecture. The implementation results are shown in section IV. Section V concludes the paper.

## II. BACKGROUND

### A. LDPC codes in IEEE 802.11ay

An LDPC code is a linear block code defined by an  $M \times N$  sparse parity check matrix  $H$ , where  $N$  is the code length and  $M$  is the number of parity checks. The LDPC codes in IEEE 802.11ay are a kind of architecture-aware (AA) LPDC codes. The parity check matrix is described as an  $M_b \times N_b$  based matrix  $H_b$ , with  $M_b = M/z$  and  $N_b = N/z$ , where  $z$  is the size of the sub-matrix. A sub-matrix is either a zero matrix or a cyclic permutation of rows of an identity matrix. In our case, the length of the codes is equal to 1344 or 672 and the size of the sub-matrix is 42. Several coding rates, 1/2, 5/8, 2/3, 3/4, 13/16, 5/6 and 7/8 are supported for each code length. The LDPC codes for 802.11ay are derived from the LDPC codes in 802.11ad by lifting [16]. The rate 2/3, 5/6 are created by shortening from rate 3/4, 7/8 and rate 7/8 is created by puncturing from rate 13/16. There are two different

Fig. 1. Parity check matrix for rate 1/2  $N=672$

Fig. 2. Parity check matrix for rate 1/2  $N=1344$  with column switch

modes of coding rate 7/8 for short code length. The second mode of rate 7/8 is called superimpose mode. The parity check matrix of the superimposed rate 7/8 is created based on the parity check matrix of rate 3/4 by applying modulo-2 addition of permutation matrices from two layers, causing two diagonal in most permutation matrices. In the following paper, the presented architecture does not supports this mode.

The base matrix of rate 1/2 for short code length and lifting code length are shown in Fig. 1 and in Fig. 2, respectively. The number written in each block denotes the cyclic-right-permutation value for the identity sub-matrix. An empty block denotes a null sub-matrix of size  $42 \times 42$ . The lifted parity check matrix has the same check node and bit node degree as the base parity check matrix.

### Algorithm 1 Layered decoding

**Initialization:**  
 $\vec{Q}_n = \vec{l}\vec{r}_n; \vec{r}_{mn} = \vec{0}$   
**Decoding:**  
**for**  $t = 1, \dots, t_{max}$  {iteration }  
**for**  $m = 1, \dots, M_b$  {sub-iteration }  
**Check node processing**  
 $\vec{q}_{mn} = |\vec{Q}|_n^{\epsilon_{mn}} - \vec{r}_{mn}$   
 $\vec{r}_{mn} = f(\vec{q}_{mk}), k \in N(m) \setminus n$   
**Bit node processing**  
 $\vec{Q}_n = \vec{q}_{mn} + \vec{r}_{mn}, n \in N(m)$   
 $|\vec{Q}|_n = \vec{Q}_n^{\tau_{mn}}$   
**Hard decision according to**  $|\vec{Q}|_n$

### B. Layered decoding schedule

This kind of architecture aware AA-LDPC codes can be decoded by partial parallel Sum-Product Algorithm (SPA)

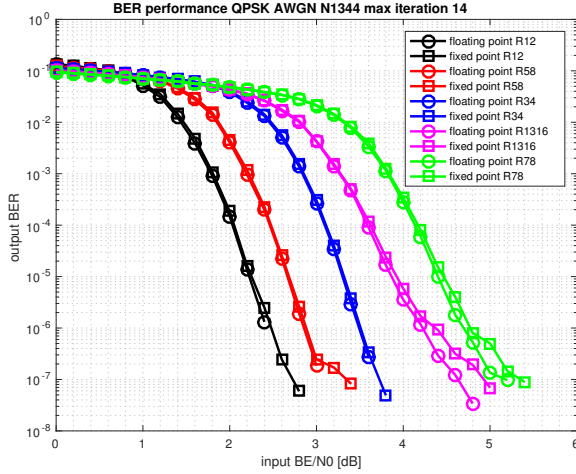


Fig. 3. BER for all coding rates over AWGN channel with maximum 14 iterations (5bits)

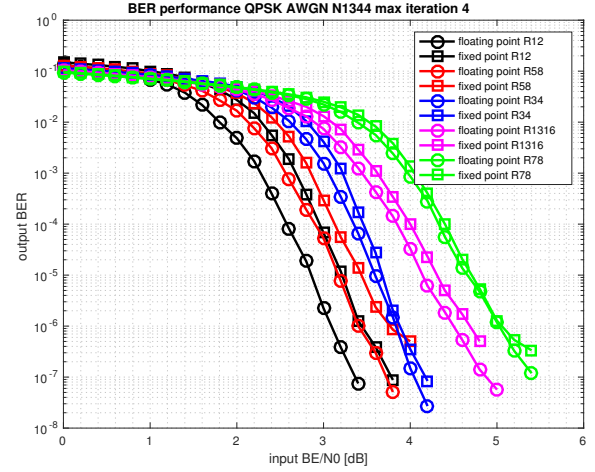


Fig. 4. BER for all coding rates over AWGN channel with maximum 4 iterations (4bits)

with layered scheduling. A group of information can be calculated and passed between bit nodes and check nodes by applying permutation at one clock cycle. For a block serial LDPC decoder, the *a posterior* information vector  $\vec{Q}_n$  with size of  $z$  is initialized with input log likelihood ratio  $llr_n$  and the *extrinsic* information vector  $\vec{r}_{mn}$  is initialized with zeros. Each iteration consist of  $M_b$  sub-iterations. During each sub-iteration, the *a posterior* information  $\vec{Q}_n$  passes through a permutation network with a shifting value  $\epsilon_{m,n}$  equal to  $P(m,n)$ , where  $P(m,n)$  is the shifting value indicated in the base matrix. The check nodes utilize the *intrinsic* information  $\vec{q}_{mn}$  which is achieved by subtracting the *extrinsic* information  $\vec{r}_{mn}$  from the *a posterior* information  $\vec{Q}_n$  to calculate the updated *extrinsic* information by using SPA or simplified min-sum [24] algorithm. Afterwards, the updated *extrinsic* information  $\vec{r}_{mn}$  is added to the *a priori* information. Finally, the *a posterior* information passes through the permutation network again with a shifting value  $\tau_{m,n}$  equal to  $z - P(m,n)$  and saved back to memory.

Fig. 3 and Fig. 4 show the performance comparison between floating and fixed point under different maximum iteration number for the five coding rates with code length 1344 using QPSK modulation over AWGN channel. The 5 bits quantization performance shows negligible performance when compared with floating point decoding with maximum 14 iterations, which gives optimal BER performance. The following decoder is configured at 4 bits quantization with 4 maximum iterations, when compared to optimal performance in Fig. 3, the performance degradation is around 0.3 to 1 dB for different coding rates.

### III. PROPOSED ARCHITECTURE

In this section, we detail a full row-based layered LDPC decoder with frame interleaved scheduling for the 802.11ay standard.

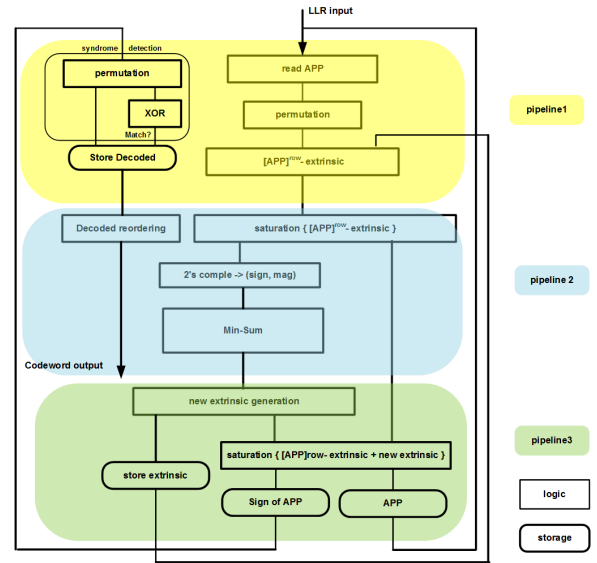


Fig. 5. Full row-based layered architecture

The architecture is shown in Fig. 5. There are three pipeline stages in the architecture and it takes 3 clock cycles to finish one layered decoding. In the first cycle, the *a posterior* information  $\vec{Q}_n$  is read from 32 storage elements, then passing through the barrel shifter and subtracted by *extrinsic* information to produce the *a priori* information  $\vec{q}_{mn}$ . The syndrome detection is also performed in the first pipeline stage. In the second cycle, the *a priori* information is saturated into smaller bit width and fed into the CFU (check node function unit). The CFU takes charge of finding the minimum absolute value and overall sign of all the fed in *a priori* information. In the last cycle, the output module generates the *extrinsic* information, added to the pre-stored *a priori* information and then directly written back to the storage elements.

In the parity check matrix of the code rate 1/2 and code

	R12	R58	R34	R1316	R78
672	4	4	4	3	3
1344	4	4	4	3	4

TABLE I

NUMBER OF EFFECTIVE LAYERS IN HARDWARE IMPLEMENTATION FOR EACH CODE RATES

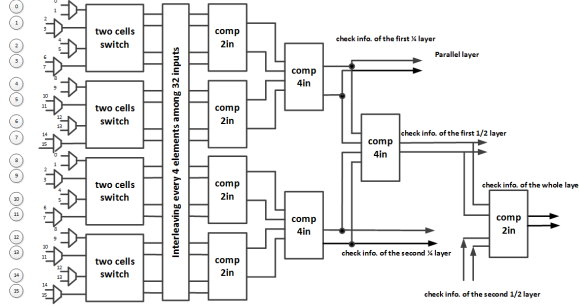


Fig. 6. Min-sum search tree with 16 inputs

rate 5/8, there are groups of two layers with no overlapped connection with any bit node and the degree of each combined layer is not larger than 8. Based on this property, it is feasible to perform message exchanges between two layers at one layered decoding. This is detailed in [10]. Every consecutive two layers in the lifted parity check matrix do not have overlapped connection with any bit node as well. Hence performing two consecutive layers in one layered decoding is feasible. We call this kind of layer as compressed layer. Taking coding rate 1/2 whose parity check matrix is shown in Fig. 2 as example, it is even feasible to perform 4 specific layers in one layer decoding in 3 clock cycles, which boosts the throughput by a factor of 2 compared to a full row-based layer LDPC decoder for the 802.11ad standard. We call this kind of layer as super compressed layer. Taking rate 5/8 with code length 1344 as example, it has 2 compressed layers and 2 super compressed layers. Each compressed and super compressed layer is one effective layer in hardware implementation as shown in I.

A min-sum search architecture with 16 inputs is shown in Fig. 6. In the proposed architecture, two of these are instantiated. The 32 inputs min-sum search tree is able to create check node information for normal layer, compressed layer and super compressed layer. How to map the *a priori* information from bit node of a combined layer to min-sum tree is not straightforward. In 802.11ad, column and layer re-ordering [10] guarantees every two-bit node vector has a one-to-one mapping to the first half and second half of the min-sum search tree. In 802.11ay, an additional two level routing network is required. The first level switches the input order of every two inputs, which is called a cell. The second level switches the input order of every two cells. One cell consists of 4 elements in the parity check matrix and has two possible orders as show in Fig. 2 top with a combined parity check matrix with column and layer switch for coding rate 1/2 is shown at the bottom.

In the architecture proposed in Fig. 5, a 3-stage pipeline

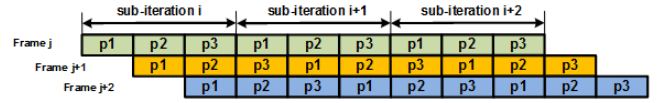


Fig. 7. Frame interleaved pipeline

to perform message exchange between bit nodes and check nodes, as defined in the compressed parity check matrix. To design for high throughput and optimal hardware utilization, a frame interleaved scheduling architecture is used. This makes sure that all pipeline stages are actives on each clock cycle, working on 3 frames at once as shown in Fig. 7.

#### IV. IMPLEMENTATION RESULTS

The LDPC decoder is implemented in 16 nm FinFET as well as 28 nm CMOS technology. Synthesis is performed with Cadence Encounter using Worst Case PVT settings. The design is synthesized at 1 GHz in 16 nm and 600 MHz in 28 nm resulting in an area of 0.19  $mm^2$  resp 0.42  $mm^2$ . For a fixed number of iterations, the minimum decoding throughput of the frame-interleaved decoder is calculated as

$$Throughput = \frac{N_{ldpc} * N_{frame} * f_{clk}}{N_{iteration} * N_{layer} * C_{layer}} \quad (1)$$

where  $N_{ldpc}$  is 1344,  $N_{frame}$  is 3,  $f_{clk}$  is the clock speed,  $C_{layer}$  is 3 and  $N_{layer}$  as listed in Table I. For 16 nm, running at 1 GHz, this gives a throughput of 112 Gbps for code rate 13/16 at 4 iterations and 67 Gbps throughput for the other rates.

Table. II gives a comparison between this work and the SOA high speed LDPC decoders for 802.11ad. Results reported for this work use a netlist from front-end logical synthesis. Power number is obtained through simulation using input LLRs stimuli created at  $E_b/N_0$  3.5 dB without early stop and 7.0 dB with early stop. The advanced 16nm technology node brings a factor of 2.2 gain in the eare and a factor of 1.6 in speed and energy efficiency when compared to 28nm.

#### V. CONCLUSION

A full row-based layered LDPC decoder with frame interleaved schedule is proposed for 802.11ay, implemented in 16 nm FinFET as well as 28 nm CMOS. The proposed architecture is flexible to support all coding rates and early stop as well. The 16 nm design has an area of 0.19  $mm^2$  with a power consumption of 408 mW and 163 mW at 3.5dB and 7dB running on an 0.8 V supply at 1 GHz clock. It achieves a throughput of 112 Gbps with maximum 4 iterations at the highest coding rate, resulting in area efficiency of 589 Gb/iter/sqmm and an energy efficiency of 3.64 pJ/bit and 1.45 pJ/bit. The implemented results show significant improvement in throughput, area and energy efficient when compared to SOA. These performance gains are related to the architecture improvements and going to a more advanced technology node.

	This work				ASSCC[20]	ISSCC[19]	ISTC[12]	VLSI[23]
Standard	802.11ay				802.11ad	802.11ad	802.11ad	802.11ad
Code length	1344/672				672	672	672	672
Algorithm	Layered				Layered	Two-phase	Two-phase	Two-phase
VFU/CFU parallelism	1344/42				336/42	672/42	672/546	672/42
Frame pipeline	YES				NO	YES	NO	YES
Unrolling	NO				NO	NO	YES	NO
No. cores	1				4	1	1	1
Support rates	10				4	4	1	4
Quantization	4				4	5	4	4
Implementation	logical synthesis				chip	chip	chip	chip
Technology	16nm		28nm		28nm	28nm	65nm	28nm
Work frequency	1000 MHz		600 MHz		470 MHz	150 MHz	257 MHz	202 MHz
Throughput (Gbps)	112@4it		67@4it		18@4it	12@3.75it	160@9it	6.78@10it
Area (sqmm)	0.19		0.42		0.78	0.63	12	1.99
VDD (V)	0.8		0.9		0.9	1.1	1.2	0.9
Power@ rate	R13/16		R13/16		R1/2	R1/2	R13/16	R13/16
Power@ SNR (db)	3.5	7	3.5	7	high	5	unknown	unknown
Early stop	NO	YES	NO	YES	YES	NO	NO	YES
Latency (ns)	36@4it	9@1it	60@4it	12@1it	138@4it	320@4it	35@9it	317@10it
Power (mW)	408	163	406	141	166	180	5360	104
Energy efficiency (pJ/iter)	3.64	1.45	6.05	2.10	18.4	15	33.5	15.34
Area efficiency (Gbps/sqmm)	589	589	160	160	23.58	19.05	13.6	3.41

TABLE II  
COMPARISONS WITH PUBLISHED MULTI-GIGABIT LDPC DECODERS.

#### ACKNOWLEDGMENT

The research leading to these results has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 760150 - project EPIC.

#### REFERENCES

- [1] R. G. Gallager, "Low-density parity-check codes", IRE Trans. Inf. Theory, vol. 8, no. 1, pp. 2128, Jan. 1962.
- [2] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices", IEEE Trans. Inf. Theory, vol. 45, no. 2, pp. 3994-31, Mar. 1999.
- [3] S.-Y. Chung, G. D. Forney, Jr., T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit", IEEE Commun. Lett., vol. 5, no. 2, pp. 5860, Feb. 2001.
- [4] ITRS 2.0, International Technology Roadmap for Semiconductors, 2015 Edition, Section 5: More Moore.
- [5] S. Y. Wu, et al., "A 16nm FinFET CMOS technology for mobile SoC and computing applications", 2013 IEEE International Electron Devices Meeting
- [6] P. Schlafer, C. Weis, N. Wehn, M. Alles, "Design space of flexible multigigabit LDPC decoders", VLSI Design, vol. 2012, February 2012.
- [7] A. Darabiha, A.C. Carusone, F.R. Kschischang, "A 3.3-gbps bit-serial block-interlaced min-sum ldpc decoder in 0.13um cmos", in Custom Integrated Circuits Conference, 2007. CICC 07. IEEE, sept. 2007, pp. 459-462.
- [8] M. Li, F. Naessens, P. Debacker, P. Raghavan, C. Desset, A. Dejonghe, L. Van der Perre, "An area and energy efficient half-row-paralleled layer LDPC decoder for the 802.11AD standard", in Proc. IEEE Workshop Signal Process. Syst., Oct. 2013, pp. 112-117
- [9] Mansour, M.M., "A Turbo-Decoding Message-Passing Algorithm for Sparse Parity-Check Matrix Codes", Signal Processing, IEEE Transactions on, 2006
- [10] M. Li, Y. Lee, Y. Huang, L. Van der Perre, "Area and energy efficient 802.11ad LDPC decoding processor", IEEE Transactions Electronics Letters, 2015
- [11] M. Weiner, B. Nikolic, Z. Zhengya, "LDPC decoder architecture for high-data rate personal-area networks", IEEE Int. Symp. on Circuits and Systems (ISCAS), pp. 1784-1787, May 2011.
- [12] P. Schlafer, N. Wehn, M. Alles, T. Lehnigk-Emden, "A new dimension of parallelism in ultra high throughput LDPC decoding", SiPS 2013 Proceedings, Taipei City, 2013, pp. 153-158.
- [13] A. Balatsoukas-Stimming and M. Meidlinger and R. Ghanaatian and G. Matz and A. Burg, "A fully-unrolled LDPC decoder based on quantized message passing", 2015 IEEE Workshop on Signal Processing Systems (SiPS), pp. 1-6.
- [14] P. Giard, G. Sarkis, C. Thibeault, W. J. Gross, "237 Gbit/s unrolled hardware polar decoder", Electronics Letters, 2015
- [15] IEEE Standard for Information Technology Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11ad-2012, Dec. 2012, pp. 1-628.
- [16] P802.11ay/D1.4 Draft Standard for Information Technology 10 Telecommunications and Information Exchange 11 Between Systems Local and Metropolitan Area 12 Networks Specific Requirements Part 11: Wireless 13 LAN Medium Access Control (MAC) and Physical Layer 14 (PHY) Specifications, July. 2018
- [17] M. Weiner and B. Nikolic and Z. Zhang, "LDPC decoder architecture for high-data rate personal-area networks", 2011 IEEE International Symposium of Circuits and Systems (ISCAS), pp. 1784-1787
- [18] Hocevar, D.E., "A reduced complexity decoder architecture via layered decoding of LDPC codes", Signal Processing Systems, 2004. SIPS 2004. IEEE Workshop on, 2004
- [19] M. Weiner, M. Blagojevic, S. Skotnikov, A. Burg, P. Flatresse, B. Nikolic, "A scalable 1.5-to-6 Gbps 6.2-to-38.1 mW LDPC decoder for 60 GHz wireless networks in 28 nm UTBB FDSOI", IEEE Int. Solid-State Circuits Conf. Dig. Tech. Pap., pp. 464-465, February 2014.
- [20] M. Li, et al., "An energy efficient 18Gbps LDPC decoding processor for 802.11ad in 28nm CMOS", 2015 IEEE Asian Solid-State Circuits Conference (A-SSCC)
- [21] H. Motozuka, N. Yosoku, T. Sakamoto, T. Tsukizawa, N. Shirakata, K. Takinami, "A 6.16Gb/s 4.7pJ/bit/iteration LDPC decoder for IEEE 802.11ad standard in 40 nm LP-CMOS", in Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP), Dec. 2015, pp. 1289-1292.
- [22] S. Scholl and S. Weithoffer and N. Wehn, "Advanced iterative channel coding schemes: When Shannon meets Moore", 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC), 2016, Brest, France
- [23] M. Milicevic, P. G. Gulak, "A Multi-Gb/s Frame-Interleaved LDPC Decoder With Path-Unrolled Message Passing in 28-nm CMOS", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2018
- [24] Jinghu Chen, A. Dholakia, E. Eleftheriou, M. P. C. Fossorier, Xiao-Yu Hu, "Reduced-Complexity Decoding of LDPC Codes", IEEE Transactions on Communications, 2005