

An Enhanced Check-Node Architecture for 5G New Radio LDPC Decoders

Fakhreddine Ghaffari*, Khoa Le*[†],

* ETIS UMR8051, CY Université, ENSEA, CNRS, F-95000, Cergy, France

[†] Faculty of Electrical and Electronics Engineering, Hochiminh City University of Technology, VNU-HCM, Vietnam
{fakhreddine.ghaffari, khoa.letrung}@ensea.fr;

Abstract—This paper presents an efficient hardware architecture of the Check Node (CN) units for the fifth generation (5G) new-radio Low-Density Parity-Check (LDPC) decoders. The proposed CN architecture is designed by splitting the high-degree CN operations into several phases and simplifying computing circuitry and connection wires. The critical path is shortened while the latency increment for one decoding iteration is negligible. Also, the proposed architecture allows to apply adaptively different offset factors when decoding different CN degree. This technique enhances the error rate and performance of our quantized LDPC decoder. The ASIC synthesis results confirm the advantages of the proposed architecture. This later helps reduce the decoder complexity by up to 30% while the operating frequency is enhanced by 10% compared to the conventional solution.

I. INTRODUCTION

Low-Density Parity-Check codes have been shown as a powerful error correction codes and they were used in many practical applications [1], [2]. More especially, they have been adopted in the enhanced mobile broadband as error correction code for data transmission in 5G [3]. The LDPC decoders take the Log-Likelihood Ratio (LLR) values of the received signal as their inputs and the decoding is an iterative process between two groups of computing units: the Variable Nodes (VNs) and the Check Nodes (CNs), until satisfying the stopping conditions. The predefined calculation methods in VNs and CNs form different decoding algorithms in which the Sum-Product (SP) algorithm is known as the most powerful error correction decoder, approaching the channel capacity with the cost of high computing complexity. A simplified version of the SP decoding algorithm, named Min-Sum (MS) algorithm, is then proposed to relax the complexity computing by using only the addition and comparison operations. However, a significant decoding performance loss is observed. Several variants of MS decoding algorithm are introduced, such as Offset Min-Sum (OMS) and Normalized Min-Sum (NMS), to compensate this performance loss [4].

In practical, the MS and its variant (OMS and NMS) are used and they are implemented in the finite precision, *i.e.*, the exchanged messages are presented with a limited number of bits. These quantized decoders can be implemented in *flooding* scheduling or in *layered* scheduling. The differences of these two scheduling can be found in several works in the literature such as in [4],[5] and the references therein. This paper discusses on the layered scheduling. The layered scheduling is favorable and practical where a decoding iteration is divided into several layers. Moreover, the converging speed of the decoder with layered scheduling is faster than that of the one with flooding scheduling [4]. Note that, the layered scheduling can be applied only for the Quasi-Cyclic LDPC (QC-LDPC) code and it is introduced in detail in the next section. In layered scheduling practical implementations, as in [5], [6], the CN processing unit (CNU) is the most complex module and it affects the maximal achievable operating frequency of the decoder, depending on its degree. For the irregular LDPC codes where the CN (and VN) degree is different from one CN to another, the CNU is usually implemented with the maximum degree and some inputs are disable when the CNU is used for lower degree CNs [7],[8], [9]. This maximum CN degree has a great impact on the decoder complexity since it decides the number of implemented processing units connecting to this CNU as well as the corresponding connection wires.

Recently, the telecommunications standard development in the 3rd Generation Partnership Project (3GPP) have decided to use LDPC codes for the enhanced mobile broadband in 5G [3]. Two QC-LDPC base matrices, named the BG1 and the BG2, have been introduced which support compatible rate and scalable data transmission. These LDPC codes are constructed by a concatenation of a high rate LDPC code and a low-density generating matrix code, which leads to the fact that these codes are dramatically irregular [10], [11]. For example in the BG1 base matrix, the CN degree varies from 3 to

maximum 19 in which only 4 in total of 46 rows are with degree 19. This high CN degree in the 5G LDPC codes would increase significantly the decoder complexity and result in a very long critical path when implementing the CNU (leading to low achievable decoder operating frequency). Furthermore, this significant difference in CN degrees and the minority of the high degree CNs would result in an inefficient hardware utilization when applying the CNU conventional realizations of [5], [6].

This work presents an efficient hardware architecture of the Check Node (CN) units for the 5G LDPC codes. The proposed CN architecture is designed by splitting the high-degree CN operations into several phases and simplifying computing circuitry and connection wires. The critical path is shortened while the latency increment for one decoding iteration is negligible. Also, the proposed architecture allows to apply adaptively different offset factors when decoding different CN degree. Also, by applying the offset scheduling only on the selective CNU, the decoding performance is impressively improved being much better than other floating-point counterparts and approaches to the performance of the Sum-Product decoder. Also, by applying the offset only on selected CNU, the decoding performance is impressively improved being much better than other floating-point counterparts and approaches the performance of the Sum-Product decoder. The paper is organized as follows. Section II presents the LDPC codes and decoding principle. This section also recalls the LDPC decoder hardware architecture and highlights the drawback when implementing the fully parallel CNU for the 5G LDPC decoders. Section III presents the proposed CNU design and analyses its effects in terms of hardware reduction and decoding throughput. The CNU modification on the offset scheduling is also presented in this section. The ASIC synthesis results and decoding performance are introduced and analyzed in Section IV. Section V concludes the paper.

II. 5G NEW RADIO (NR) LDPC CODES AND DECODERS

A. Preliminaries

An LDPC code is presented by a parity check matrix $H(M \times N)$ with N columns corresponding to N binary bits and M rows corresponding to M parity check equations on these column bits. The coded bit n ($1 \leq n \leq N$) is parity-checked by the equation m ($1 \leq m \leq M$) if the entry $H(m, n) = 1$ and they are called the neighbors. The neighbors set of the CN m is defined as $\mathcal{N}(m)$ and that of the VN n is as $\mathcal{N}(n)$.

The VN (CN) degree is defined by the number of its neighbors, denoted by d_v (d_c), *i.e.*, $d_{vn} = |\mathcal{N}(n)|$ and $d_{cm} = |\mathcal{N}(m)|$. In practical, a type of the LDPC codes, named as Quasi-Cyclic LDPC (QC-LDPC) code, is used in most of the LDPC practical applications such as communications, storage systems [1], [2], [12], [13] since it facilitates the hardware implementation of the decoder while providing a good decoding performance compared to the random-constructed codes [4], [14]. A QC-LDPC code is presented by a base matrix $H_B(R_B \times C_B)$, with the integer entries $-1 \leq H_B(i, j) < Z - 1$, ($1 \leq i \leq R_B, 1 \leq j \leq C_B$). The QC-LDPC parity check matrix is obtained by extending the base matrix such a way that each entry (i, j) is replaced by a $H_B(i, j)$ -times cyclic-shift of the $Z \times Z$ identity matrix in the case $H_B(i, j) \geq 0$, otherwise it is replaced by a $Z \times Z$ all-zero matrix (Z is called the expansion factor). The error correction principle of LDPC codes is as follows. The data information, in form of a vector with K bits, is encoded by adding M additional bits to form a vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ($|\mathbf{x}| = N$) such that $H\mathbf{x}^T = \mathbf{0}$ and \mathbf{x} is called a codeword. This codeword \mathbf{x} is assumed to be affected by noise. This noise appears when \mathbf{x} is transmitted through the communication channels (when LDPC is applied in communications) or when \mathbf{x} is being stored in the storage medium, being read from or being written to that medium. Before being transmitted through a communication channel, \mathbf{x} is modulated and in this paper, we assume that the Binary Phase Shift Keying (BPSK) modulation and the Additive White Gaussian Noise (AWGN) model are used. The received information at the receiver is denoted by $\mathbf{y} = (y_1, y_2, \dots, y_n)$ ($|\mathbf{y}| = N$): $y_n = (1 - 2 \times x_n) + z_n$ where z_n is the Gaussian noise with zero mean and variance σ^2 . The quantized LDPC decoders take the quantized version of \mathbf{y} as their inputs. We denote the decoder input as $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_N)$ such that $\gamma_n = [\mu \cdot y_n]$ where $[a]$ returns the closest integer of a in the integer set $\mathcal{A} = (-Q, \dots, -1, 0, +1, \dots, +Q)$, ($Q = 2^{q-1} - 1$). $\mu > 0$ is known as the channel scale factor to be optimized for a good error correction performance [5]. The LDPC decoding is a message-passing process where the messages are iteratively passed between the VNs and the CNs. In a decoding iteration, the VNs and CNs will compute and send their messages. This iterative process has continuously proceeded until the correct codeword is successfully found (decoding success) or the maximum number of iterations (denoted by It_{max}) is reached (decoding failure). The LDPC decoding implementation could be flooding where all the VNs will update their

messages at the same time then all CNs updates concurrently the messages. The QC-LDPC codes provide the possibility to be decoded on a “layered” way in which a decoding iteration comprises several decoding layers. In a decoding layer, only CNs belonging to 1 (or some) row(s) of the base matrix update their messages and the VNs connecting to that CNs update their a-posteriori-probability (APP). A decoding iteration is finished when all of the CNs of H update their messages. The layered decoding scheduling algorithm and its hardware architecture can be found in detailed in [5] and we follow the architecture in that paper for our work. It should be noted from the architecture that, there are d_c saturation (SAT) modules, d_c VN processing units (VNUs) and d_c barrel shifter (BS) modules implemented to produce d_c CNU inputs. Similarly, there are d_c APP modules and d_c inverted BS (nBS) implemented to process d_c CNU outputs. Therefore, the large value of d_c will result in a high complexity implementation. The CNU module is implemented in parallel to maximize the decoding throughput. Also, for the irregular LDPC codes, *i.e.*, the CN degree may be different from one to another, the CNU is implemented with $d_{c_{max}}$ inputs and $d_{c_{max}}$ outputs [5], [8], [9].

B. The 5G NR LDPC codes and remarks

Recently, the structure of the LDPC codes used for 5G was released by the 3GPP and the code details, including 2 base matrices BG1 and BG2, can be found in [3][10][15]. This 5G NR LDPC is the concatenation of a high rate LDPC core (the first 4 rows of the base matrices) with a low-density generator matrix [10]. This low-density generator matrix (the extension part) is for the Hybrid Automatic Repeat reQuest (HARQ) with Incremental Redundancy (IR-HARQ) and results in a significant number of degree-1 [16]. The structure of one of the base matrices, the BG1, is illustrated in Fig. 1 in which the sub-matrices **A** and **D** is called the core (or kernel) and the **E**, **O** and **I** are the extensions (**O** is all-zero matrix while **I** is the identity matrix).

A considerable remark on the 5G LDPC codes is that these codes are dramatically irregular, for both VN and CN sides [16]. The CN degree distribution is listed in the table I for the two matrices BG1 and BG2. It can be seen that in the base matrix BG1, the CN degree varies largely from 3 to 19. More specially, only 4 rows are with degree 19 in the total of 46 rows. With the conventional architecture for irregular codes mentioned above, the implemented CNUs with degree $d_{c_{max}} = 19$ are fully used in only 4 layers while they are operated

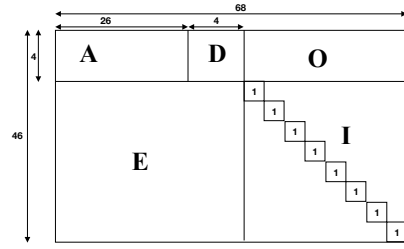


Figure 1. The structure of the base matrix BG1 proposed for 5G.

with $d_c < 11$ in the 42 other layers (26 layers are with degree 5 and 6). It is also the case for the BG2 where there are 2 rows (layers) with $d_{c_{max}} = 10$ while having 20 rows with degree 4. The considerable variation in CN degree would result in the high inefficiency in hardware utilization in conventional architecture.

Table I. CN DEGREE DISTRIBUTION OF THE BG1 AND BG2 5G LDPC CODES. THE BOLD NUMBERS ARE OF THE KERNEL PART OF THE CODES.

	CN degree									total
	3	4	5	6	7	8	9	10	19	
Nb of rows in BG1	1	5	18	8	5	2	2	1	4	46
Nb of rows in BG2	6	20	9	3	-	2	-	2	-	42

III. THE PROPOSED CNU ARCHITECTURE FOR 5G LDPC CODES

As mentioned in the previous sections, applying the conventional LDPC decoding architecture to the 5G LDPC decoder would result in inefficiency in hardware utilization, and this is due to the high irregularity in CN degree. The principle of our modification on the CNU is on the separation of its operations into several phases, and each phase can be performed in separate clock cycles. Indeed, the CN calculation is mainly on the finding of the first and second minimum of a set of inputs. This operation can be divided into the same operation on the sub-set of the inputs and the CN calculation is fully finished when all the inputs are covered in these subsets. For example, instead of implementing the CNU with 19 inputs (degree 19) for the BG1 5G LDPC decoding, we may implement the CNU with 10 inputs (degree 10). In this case, the CN calculations on the layers having degree smaller or equal to 10 is performed as normally while that of the layer with degree 19 will be performed in $\lceil 19/10 \rceil = 2$ phases (finding the minimum values in the first 10 inputs in the first phase and then continuing with the other 9 inputs in the second phase). The design of the proposed CNU is shown in Fig 4, on comparison with the conventional CNU design. It can

be seen in the figure that the conventional CNU of the MS/OMS decoder is designed with dc_{max} inputs and dc_{max} outputs to cover all the possibilities of CN degree. When operating on the degree being smaller than dc_{max} , the input 2-to-1 multiplexers will disable the unused inputs by multiplexing a predefined “max” values to its output. On the contrary, the proposed CNU is designed with $dc < dc_{max}$ inputs and $dc < dc_{max}$ outputs. In order to keep the precise result, the memory elements are implemented to store the tentative min_1 , min_2 and the index of min_1 from one phase to another. When all calculating phases are finished, the values in these additional memory elements are used to produce the CNU computing results. The advantage of our proposed solution comes at the decoder complexity reduction. As analyzed above, the number of CNU inputs and outputs decide the number of computing circuits connecting to the CNU, and it is true for all computing components. When excluding the memory blocks in the decoder design, we can formulate the normalized complexity of the proposed decoder to the conventional one as $\mathcal{C} = \frac{dc}{dc_{max}}$. The proposed method would help reduce the critical path and hence, enhance the operating frequency. It comes from the fact that the CNU design is for a smaller number of inputs. Furthermore, with the lower complexity design, the synthesizer may have more space in the routing optimization. This is confirmed in the synthesis results of the next section. The proposed method, however, requires more clock cycles to finish a decoding iteration compared to the conventional design, which may affect the decoding throughput. The decoding throughput of an LDPC decoder is computed using the Equ. 1 where f_{max} denotes the maximum decoding frequency and It_{max} is the maximum number of iterations to decode one codeword. $(2*L+\sigma)$ denotes the number of clock cycles required for each decoding iteration in which 2 clock cycles are needed to decode 1 layer, L is the number of layers and σ is the additional cycles when applying our method ($\sigma = 0$ in the conventional design).

$$\theta = \frac{N * f_{max}}{It_{max} * (2 * L + \sigma)} \quad (1)$$

When assuming the decoding frequencies are equivalent, the normalized throughput of the proposed design to the conventional one can be formulated as $\mathcal{T} = \frac{2*L}{2*L+\sigma}$. Our CNU design requires 2 clock cycles for 1 phase of the CNU computing (1 cycle for finding min_1 and $index-min_1$, and 1 cycle for the min_2). The $\tilde{\gamma}$ updating for each extra phase requires one cycle. In general, σ depends on the implemented CNU degree and the 5G

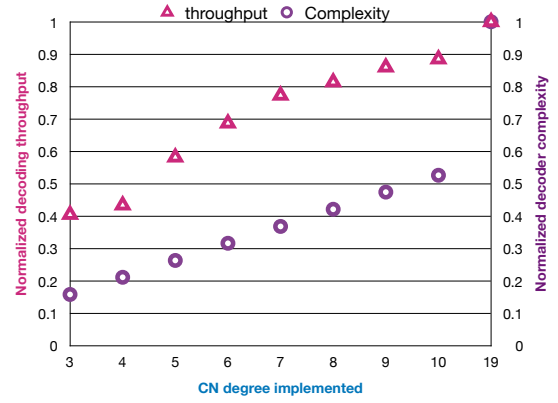


Figure 2. The complexity gain and throughput loss as a function of the implemented CNU degree for the BG1 5G LDPC code.

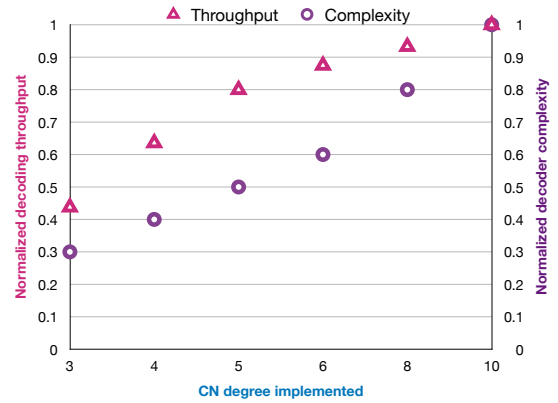


Figure 3. The complexity gain and throughput loss as a function of the implemented CNU degree for the BG2 5G LDPC code.

LDPC matrix. In order to visualize Complex gain and the Throughput loss of our solution, we plot \mathcal{C} and \mathcal{T} as the functions of the CNU implemented degree for the two 5G LDPC base matrices, BG1 and BG2, in Fig. 2 and Fig. 3, respectively. It can be seen that \mathcal{C} and \mathcal{T} vary dramatically on the implemented CNU degree and the gain in complexity is more significant than the loss of decoding throughput. This is true for the two matrices. However, the gain of our method is more interesting in the BG1 matrix. It is because the BG1 has a higher maximum CN degree (which is 19) than the one of the BG2 (which is 10). A quantitative gain example is when implementing the CNU with degree 10 instead of 19 (of the conventional design) on the BG1. The complexity reduces 50% while the throughput decreases 10%. In other words, in a complexity-equivalent comparison, applying our solution would increase 80% decoding throughput.

Another modification in our CNU design compared to the conventional one is on the selective application of

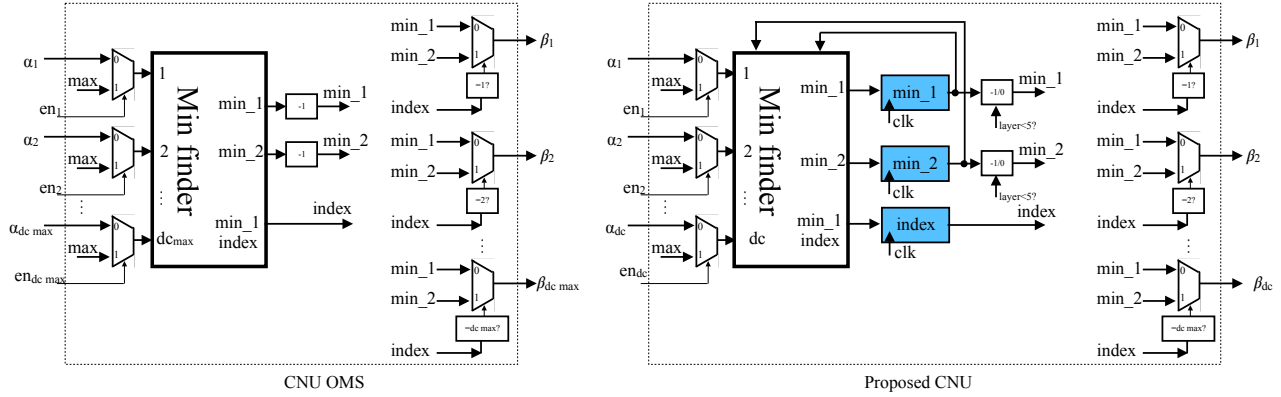


Figure 4. The architecture of the conventional CNU and the proposed CNU for the 5G LDPC OMS decoder.

the offset factor on the computed minimum values. This contribution helps improve significantly the decoding performance of the 5G LDPC decoder, by adding the extra control on the offset operations. As mentioned in the description of the 5G matrices, they are dramatically irregular on both VN and CN degrees. This irregularity results in the performance loss when applying an offset factor on the quantized OMS decoder. Note that, on the quantized decoder, the offset factor could be only the integers (typically, it is 1). The decoding performance loss on irregular codes were observed in many works on the literature such as in [17], [5] and it is even more significant on the 5G LDPC codes as seen in Fig 5. We also observed on our statistics that the VNs with low degree tend to be more prone to error. In particular, on the degree-1 VNs of the extension part of the 5G LDPC codes, the error probability is 10^3 times higher than other VN degrees. The proposed CNU design allows to apply selectively the offset factor as a function of the VN/CN degree. For example, for the 5G LDPC codes, only for the first four layers (where all the neighbor VNs have the minimum degree of 4 for BG1 and 5 for BG2) is applied the offset factor, *i.e.*, $\lambda = 1$ if $l < 5$ else $\lambda = 0$. From the hardware complexity point of view, the additional complexity for this control could be negligible since this single verification can be done globally and propagated to all the CNUs.

IV. THE SYNTHESIS RESULTS ANALYSIS AND DECODING PERFORMANCE

In order to confirm the efficiency of our solution, we have implemented the conventional OMS decoder (CNU with degree 10) and our proposed decoder (with CNU degree 6) on the $K = 520, Z = 52$, code rate $R = 0.25, N = 2080$ 5G BG2 LDPC code. The proposed

decoders are synthesized using the 90nm technology with the Digital Standard Cell Library SAED-EDK90-CORE, Process 1P9M 1.2v/2.5v. The hierarchical design flow is with the Synopsys Design Platform. The post-synthesis results are reported in table II. It is confirmed that the proposed solution helps reduce the decoder complexity (around 30% in this implementation) while the operating frequency is enhanced by 10% compared to the conventional solution (204MHz compared to 184MHz). The decoding throughput of our solution provides a slight enhancement (1.285 Gbps compared to 1.275 Gbps). This is because the proposed CNU requires some extra clock cycles to finish its operations.

Table II. THE SYNTHESIS RESULTS ON THE $K = 520, Z = 52, R = 0.25, N = 2080$ 5G BG2 LDPC CODE.

	The proposed decoder CNU degree = 6	The OMS decoder CNU degree = 10
Area (mm^2)	2.5	3.6
f_{max} (Mhz)	204	184
Throughput (Gbps)	1.285	1.275

For decoding performance, we have simulated the proposed decoder on the above 5G LDPC code ($K = 520, Z = 52, R = 0.25, N = 2080$) with the $It_{max} = 15$. We make comparison with performances of quantized MS and OMS decoders. The performance of the floating-point SP, OMS and MS decoder benchmarks are also added. It can be seen in the Fig 5 that the proposed solution has an impressive performance gain. Our quantized decoder is even better than the floating-point OMS and MS decoders and approaches to the performance of the SP decoder, especially on the high error rate region.

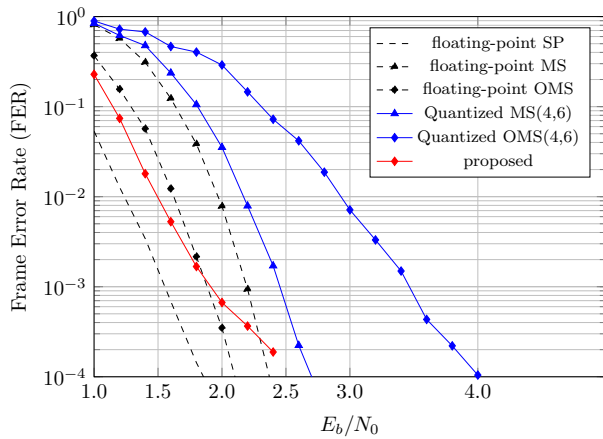


Figure 5. Decoding performance of the decoders on the $K = 520$, $Z = 52$, $R = 0.25$, $N = 2080$ 5G LDPC code.

V. CONCLUSIONS

We introduce in this paper an efficient hardware architecture for the 5G New Radio LDPC decoders. We focus on optimizing the architecture of the check node processing units (CNU) to adapt to the highly irregular 5G LDPC code. The proposed CNU separates the operations into several phases and performs them in different clock cycles. By avoiding implementing the high-degree CNUs, we show that the decoder complexity can be significantly reduced and the operating frequency is enhanced, compared to the architecture on the state-of-the-art. By applying the offset factor only on the CNs with the high degree neighbor Variable Nodes, the decoding performance of our implemented decoder is impressively improved, much better than other floating-point decoders and approaches to the performance of the Sum-Product decoder.

REFERENCES

- [1] "Wireless lan medium access control and physical layer specifications: Enhancements for higher throughput, p802.11n/d3.07," *IEEE-802.11n*, Mar. 2008.
- [2] "Local and metropolitan area networks - part 16, ieee std 802.16a-2003," *IEEE-802.16*.
- [3] "Wf on ldpc parity check matrices," *document R1-1711982, 3GPP TSG RAN WG1 NR AH 2*, [Online]. Available: <http://www.3gpp.org>, 2017.
- [4] D. Declercq, M. Fossorier, and E. Biglieri, "Channel coding: Theory, algorithms, and applications," *Academic Press Library in Mobile and Wireless Communications*, Elsevier, ISBN: 978-0-12-396499-1, 2014.
- [5] T. T. Nguyen-Ly, V. Savin, K. Le, D. Declercq, F. Ghaffari, and O. Boncalo, "Analysis and design of cost-effective, high-throughput LDPC decoders," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, no. 3, pp. 508–521, March 2018.
- [6] C. Zhang, S. Weng, X. You, and Z. Wang, "Area-efficient check node unit architecture for single block-row quasi-cyclic ldpc

codes," in *2014 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Nov 2014, pp. 431–434.

- [7] A. Katyushnyj, A. Krylov, A. Rashich, C. Zhang, and K. Peng, "Fpga implementation of ldpc decoder for 5g nr with parallel layered architecture and adaptive normalization," in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, 2020, pp. 34–37.
- [8] A. Verma and R. Shrestha, "A new partially-parallel vlsi-architecture of quasi-cyclic ldpc decoder for 5g new-radio," in *2020 33rd International Conference on VLSI Design and 2020 19th International Conference on Embedded Systems (VLSID)*, 2020, pp. 1–6.
- [9] H. Wu and H. Wang, "A high throughput implementation of qc-ldpc codes for 5g nr," *IEEE Access*, vol. 7, pp. 185 373–185 384, 2019.
- [10] T. Richardson and S. Kudekar, "Design of low-density parity check codes for 5g new radio," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 28–34, MARCH 2018.
- [11] H. Cui, F. Ghaffari, K. Le, D. Declercq, J. Lin, and Z. Wang, "Design of high-performance and area-efficient decoder for 5g ldpc codes," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 2, pp. 879–891, 2021.
- [12] K. Le, F. Ghaffari, D. Declercq, and B. Vasić, "Efficient hardware implementation of probabilistic gradient descent bit-flipping," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 4, pp. 906–917, April 2017.
- [13] N. Jiang, K. Peng, J. Song, C. Pan, and Z. Yang, "High-throughput qc-ldpc decoders," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 251–259, 2009.
- [14] V. L. Petrovi, M. M. Markovi, D. M. E. Mezeni, L. V. Saranovac, and A. Radoevi, "Flexible high throughput qc-ldpc decoder with perfect pipeline conflicts resolution and efficient hardware utilization," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 12, pp. 5454–5467, 2020.
- [15] T. Lin, S. Cao, S. Zhang, and S. Xu, "A unified reconfigurable datapath for 5g compatible ldpc decoding," in *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, 2018, pp. 215–218.
- [16] K. Le, F. Ghaffari, and D. Declercq, "An adaptation of min-sum decoder for 5g low-density parity-check codes," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, Sapporo, Japan, May 2019, pp. 1–5.
- [17] F. Cochachin, D. Declercq, E. Boutillon, and L. Kessal, "Density evolution thresholds for noise-against-noise min-sum decoders," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–7.