

A 588-Gb/s LDPC Decoder Based on Finite-Alphabet Message Passing

Reza Ghanaatian¹, Alexios Balatsoukas-Stimming, Thomas Christoph Müller, *Student Member, IEEE*, Michael Meidlinger, Gerald Matz, *Senior Member, IEEE*, Adam Teman², and Andreas Burg, *Member, IEEE*

Abstract—An ultrahigh throughput low-density parity-check (LDPC) decoder with an unrolled full-parallel architecture is proposed, which achieves the highest decoding throughput compared to previously reported LDPC decoders in the literature. The decoder benefits from a serial message-transfer approach between the decoding stages to alleviate the well-known routing congestion problem in parallel LDPC decoders. Furthermore, a finite-alphabet message passing algorithm is employed to replace the VN update rule of the standard min-sum (MS) decoder with lookup tables, which are designed in a way that maximizes the mutual information between decoding messages. The proposed algorithm results in an architecture with reduced bit-width messages, leading to a significantly higher decoding throughput and to a lower area compared to an MS decoder when serial message transfer is used. The architecture is placed and routed for the standard MS reference decoder and for the proposed finite-alphabet decoder using a custom pseudo-hierarchical backend design strategy to further alleviate routing congestions and to handle the large design. Postlayout results show that the finite-alphabet decoder with the serial message-transfer architecture achieves a throughput as large as 588 Gb/s with an area of 16.2 mm² and dissipates an average power of 22.7 pJ per decoded bit in a 28-nm fully depleted silicon on insulator library. Compared to the reference MS decoder, this corresponds to 3.1 times smaller area and 2 times better energy efficiency.

Index Terms—28-nm FD-SOI, finite-alphabet decoder, low-density parity-check (LDPC) code, min-sum (MS) decoding, unrolled architecture.

I. INTRODUCTION

LOW-DENSITY parity-check (LDPC) codes have become the coding scheme of choice in high data-rate communication systems after their rediscovery in the 1990s [1], due to their excellent error correcting performance along with the availability of efficient high-throughput hardware implementations in modern CMOS technologies.

Manuscript received March 10, 2017; revised June 10, 2017 and August 27, 2017; accepted October 4, 2017. Date of publication November 16, 2017; date of current version January 19, 2018. This work was supported by the Swiss National Science Foundation under Project 200021-153640. (*Corresponding author: Reza Ghanaatian.*)

R. Ghanaatian, A. Balatsoukas-Stimming, C. Müller, A. Teman, and A. Burg are with the Telecommunications Circuits Laboratory, EPFL Lausanne, 1015 Lausanne, Switzerland (e-mail: reza.ghanaatian@epfl.ch; alexios.balatsoukas@epfl.ch; christoph.mueller@epfl.ch; adam.teman@epfl.ch; andreas.burg@epfl.ch).

M. Meidlinger and G. Matz are with the Vienna University of Technology, 1040 Vienna, Austria (e-mail: mmeidlin@nt.tuwien.ac.at; gmatz@nt.tuwien.ac.at).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVLSI.2017.2766925

LDPC codes are commonly decoded using iterative message passing (MP) algorithms in which the initial estimations of the bits are improved by a continuous exchange of messages between decoder computation nodes. Among the various MP decoding algorithms, the min-sum (MS) decoding algorithm [2] and its variants [e.g., offset min-sum (OMS) and scaled MS] are the most common choices for hardware implementation. LDPC decoder hardware implementations traditionally start from one of these established algorithms (e.g., MS decoding), where the exchanged messages represent log likelihood ratios (LLRs). These LLRs are encoded as fixed point numbers in two's-complement or sign-magnitude representation, using a small number of uniform quantization levels, in order to realize the message update rules with low-complexity conventional arithmetic operations.

Recently, there has been significant interest in the design of finite-alphabet decoders for LDPC codes [3]–[7]. The main idea behind finite-alphabet LDPC decoders is to start from one or multiple arbitrary message alphabets, which can be encoded with a bit-width that is acceptable from an implementation complexity perspective. The message update rules are then crafted as generic mapping functions to operate on this alphabet. The main advantage of such finite-alphabet decoders is that the message bit width can be reduced significantly with respect to a conventional decoder, while maintaining the same error-correcting performance [5], [6]. The downside of this approach is that the message update rules of finite-alphabet decoders usually cannot be described using fast and area-efficient standard arithmetic circuits.

Different hardware architectures for LDPC decoders have been proposed in the literature in order to fulfill the power and throughput requirements of various standards. More specifically, various degrees of resource sharing result in flexible decoders with different area requirements. On the one hand, partial-parallel LDPC decoders [8], [9] and block-parallel LDPC decoders [10], [11] are designed for medium throughput, with modest silicon area. Full-parallel [12], [13] and unrolled LDPC decoders [5], [14], on the other hand, achieve very high throughput (in the order of several tens or hundreds of Gigabits per second) at the expense of large area requirements. Even though, in principle, LDPC decoders are massively parallelizable, the implementation of ultrahigh speed LDPC decoders still remains a challenge, especially for long LDPC codes with large node degrees [15]. While synthesis results for such long-length codes, as for

example reported in [16], show the potential for a very high throughput, the actual implementation requires several further considerations mainly due to severe routing problems and the impact of parasitic effects.

A. Contributions

In this paper, we propose an unrolled full-parallel architecture based on serial transfer of the decoding messages, which enables an ultrahigh throughput implementation of LDPC decoders for codes with large node degrees by reducing the required interconnect wires for such decoders. Moreover, we employ a finite-alphabet LDPC decoding algorithm in order to decrease the required quantization bit-width, and thus, to increase the throughput, which is limited by the serial message transfer in the proposed architecture. We also adopt a linear floorplan for the unrolled full-parallel architecture as well as an efficient pseudohierarchical flow that allows the high-speed physical implementation of the proposed decoder. To the best of the authors' knowledge, by combining the aforementioned techniques, we present the fastest fully placed and routed LDPC decoder in the literature.

B. Outline

The remainder of this paper is organized as follows. Section II gives an introduction to decoding of LDPC codes, as well as more details on existing high-throughput implementations of LDPC decoders. Section III describes our proposed ultrahigh throughput decoder architecture that employs a serial message-transfer technique. In Section IV, our algorithm to design a finite-alphabet decoder with nonuniform quantization is explained and applied to the serial message-transfer decoder of Section III. Section V describes our proposed approach for the physical implementation and the timing and area optimization of our serial message-transfer decoders. Finally, Section VI analyzes the implementation results, and Section VII concludes this paper.

II. BACKGROUND

In this section, we first briefly summarize the fundamentals of LDPC codes and the iterative MP algorithm for the decoding. We then review the state-of-the-art in high-speed LDPC decoder architectures to set the stage for the description of our implementation.

A. LDPC Codes and Decoding Algorithms

A binary LDPC code is a set of codewords that are defined through an $M \times N$ binary-valued sparse parity-check matrix as

$$\{\mathbf{c} \in \{0, 1\}^N \mid \mathbf{H}\mathbf{c} = \mathbf{0}\} \quad (1)$$

where all operations are performed modulo 2. If the parity-check matrix contains exactly d_v ones per column and exactly d_c ones per row, the code is called a (d_v, d_c) -regular LDPC code. Such codes are usually represented with a Tanner graph, which contains N variable nodes (VNs) and M check nodes (CNs), and VN n is connected to CN m if and only if $\mathbf{H}_{mn} = 1$.

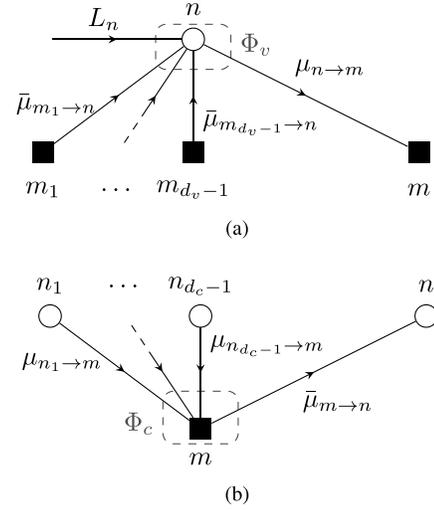


Fig. 1. (a) VN update and (b) CN update for $\mathcal{N}(n) = \{m, m_1, \dots, m_{d_v-1}\}$ and $\mathcal{N}(m) = \{n, n_1, \dots, n_{d_c-1}\}$.

LDPC codes are traditionally decoded using MP algorithms, where information is exchanged as messages between the VNs and the CNs over the course of several decoding iterations. At each iteration, the message from VN n to CN m is computed using a mapping $\Phi_v : \mathbb{R}^{d_v} \rightarrow \mathbb{R}$, which is defined as

$$\mu_{n \rightarrow m} = \Phi_v(L_n, \bar{\boldsymbol{\mu}}_{\mathcal{N}(n) \setminus m \rightarrow n}) \quad (2)$$

where $\mathcal{N}(n)$ denotes the neighbors of node n in the Tanner graph, $\bar{\boldsymbol{\mu}}_{\mathcal{N}(n) \setminus m \rightarrow n} \in \mathbb{R}^{d_v-1}$ is a vector that contains the incoming messages from all neighboring CNs except m , and $L_n \in \mathbb{R}$ denotes the channel LLR corresponding to VN n . Similarly, the CN-to-VN messages are computed using a mapping $\Phi_c : \mathbb{R}^{d_c-1} \rightarrow \mathbb{R}$, which is defined as

$$\bar{\mu}_{m \rightarrow n} = \Phi_c(\boldsymbol{\mu}_{\mathcal{N}(m) \setminus n \rightarrow m}). \quad (3)$$

Fig. 1 illustrates the message updates in the Tanner graph. In addition to Φ_v and Φ_c , a third mapping $\Phi_d : \mathbb{R}^{d_v+1} \rightarrow \{0, 1\}$ is needed to provide an estimate of the transmitted codeword bits in the last VN iteration based on the incoming CN messages and the channel LLR L_n according to

$$\hat{c}_n = \Phi_d(L_n, \bar{\boldsymbol{\mu}}_{\mathcal{N}(n) \rightarrow n}). \quad (4)$$

Messages are exchanged until a valid codeword has been decoded or until the maximum number of iterations I has been reached.

For the widely used MS algorithm, the mappings (2) and (3) are

$$\Phi_v^{\text{MS}}(L, \bar{\boldsymbol{\mu}}) = L + \sum_i \bar{\mu}_i \quad (5)$$

and

$$\Phi_c^{\text{MS}}(\boldsymbol{\mu}) = \text{sign } \boldsymbol{\mu} \min |\boldsymbol{\mu}| \quad (6)$$

where $\min |\boldsymbol{\mu}|$ denotes the minimum of the absolute values of the vector components and $\text{sign } \boldsymbol{\mu} = \prod_j \text{sign } \mu_j$. The decision

mapping Φ_d is defined as

$$\Phi_d^{\text{MS}}(L, \vec{\mu}) = \frac{1}{2} \left(1 - \text{sign} \left(L + \sum_i \bar{\mu}_i \right) \right). \quad (7)$$

B. High-Throughput LDPC Decoders

Several high-throughput LDPC decoders have been developed during the past decade in order to satisfy the high data-rate requirements of optical and high-speed Ethernet networks. These decoders usually rely on a full-parallel isomorphic [17] architecture and a flooding schedule, which directly maps the algorithm for one iteration to hardware. More specifically, the CN and VN update equations are directly mapped to M CN and N VN processing units, and a hard-wired routing network is responsible for passing the messages between them. From an implementation perspective, while such an architecture enables a very high throughput by fully exploiting the inherent parallelism of each iteration, the complexity of the highly unstructured routing network turns out to be a severe bottleneck. In addition to this routing problem, such full-parallel decoders usually require one or two clock cycles for each iteration and in the worst case as many cycles as the maximum number of iterations for each codeword, which is another throughput limitation factor.

Several solutions have been proposed to alleviate the routing problem in full-parallel decoders, on both architectural and algorithmic levels. Darabiha *et al.* [18] and [19] suggest using a bit-serial architecture, which only requires a single wire for each variable-to-check and check-to-variable-node connection. While this approach can reduce the routing congestion, it also leads to a significant reduction in the decoding throughput. The decoder in [19], for example, only achieves a throughput of 3.3 Gb/s when implemented using a 130-nm CMOS technology. Another architectural technique is reported in [13], where the long wires of the decoder are partitioned into several short wires with pipeline registers. As a result, the critical path is broken down into shorter paths, but the decoder throughput is also affected since more cycles are required to accomplish each iteration. Nevertheless, the decoder of [13] is still able to achieve 13.2 Gb/s in 90-nm CMOS with 16 iterations.

On an algorithmic level, Mohsenin *et al.* [20] propose a MP algorithm, called MS split-row threshold, which uses a column-wise division of the \mathbf{H} matrix into S_{pn} partitions. Each partition contains N/S_{pn} VNs and M CNs, and global interconnects are minimized by only sharing the minimum signs between the CNs of each partition. This algorithmic modification was used to implement a full-parallel decoder for the challenging (2048, 1723) LDPC code in 65-nm CMOS, which achieves a throughput of 36.2 Gb/s with 11 decoding iterations. Another decoder, reported in [21], uses a hybrid hard/soft decoding algorithm, called differential binary MP algorithm, which reduces the interconnect complexity at the cost of some error-correcting performance degradation. A full-parallel (2048, 1723) LDPC decoder using this algorithm was implemented in 65-nm CMOS technology, achieving a throughput of up to 126 Gb/s [22]. The work of [23] also proposes another algorithmic level modification, called the

probabilistic MS algorithm, where a probabilistic second minimum value is used instead of the true second minimum value to simplify the CN operation and to facilitate high-throughput implementation of full-parallel LDPC decoders. Further, a mix of tree and butterfly interconnect network is proposed in the CN unit to balance the interconnect complexity and the logic overhead and to reduce the routing complexity. The implementation of a decoder for the (2048, 1723) LDPC code with the proposed techniques in 90-nm CMOS technology achieves a throughput of 45.42 Gb/s.

Stochastic decoding of LDPC codes [24] was another important improvement based on both algorithm and gate-level implementation considerations to solve the routing problem of LDPC decoders, where probabilities are interpreted as binary Bernoulli sequences. This approach, on the one hand, reduces the complexity of CNs and the routing overhead, but, on the other hand, introduces difficulties in VN update rules due to correlated stochastic streams, which may deteriorate the error-correcting performance especially in longer-length codes. To solve this correlation problem by rerandomizing the VN output streams, the work of [25] proposes to use majority-based tracking forecast memories in each VN, which results in a decoder with full-parallel architecture for a (2048, 1723) LDPC code that achieves a throughput of 61.3 Gb/s in 90-nm CMOS. An alternative method to track the probability values, called delayed stochastic decoding, is reported in [26], and the full-parallel decoder for the same code can deliver a throughput as large as 172.4 Gb/s in 90-nm CMOS.

To solve the problem of throughput limitations in full-parallel decoders from potentially using multiple iterations for decoding, the work of [14] presents an unrolled full-parallel LDPC decoder. In the proposed architecture, each decoding iteration is mapped to distinct hardware resources, leading to a decoder with I iterations that can decode one codeword per clock cycle, at the cost of significantly increased area requirements with respect to nonunrolled full-parallel decoders. This unrolled architecture achieves a throughput of 161 Gb/s for a (672, 546) LDPC code with $d_v = 3$ and $d_c = 6$, when implemented in a 65-nm CMOS technology. It is noteworthy that an unrolled decoder has 50% reduced wires between adjacent stages compared to a nonunrolled decoder since one stage of VNs is connected to one stage of check nodes (CNs) with unidirectional data flow per decoding iteration. Even though this measure leads to a lower routing congestion, it is still challenging to fully place and route such a decoder. This routing issue becomes more and more severe when considering longer LDPC codes and especially with increasing CN and VN degrees to achieve better error-correcting performance, as required in wireline applications such as for the (2048, 1723) code with $d_v = 6$ and $d_c = 32$ used in the IEEE 802.3an standard [15].

III. SERIAL MESSAGE-TRANSFER LDPC DECODER

Unrolled full-parallel LDPC decoders provide the ultimate throughput with smaller routing congestion than conventional full-parallel decoders. However, they are still not trivial to implement for long LDPC codes with high CN and VN degrees, which suffer from severe routing congestion.

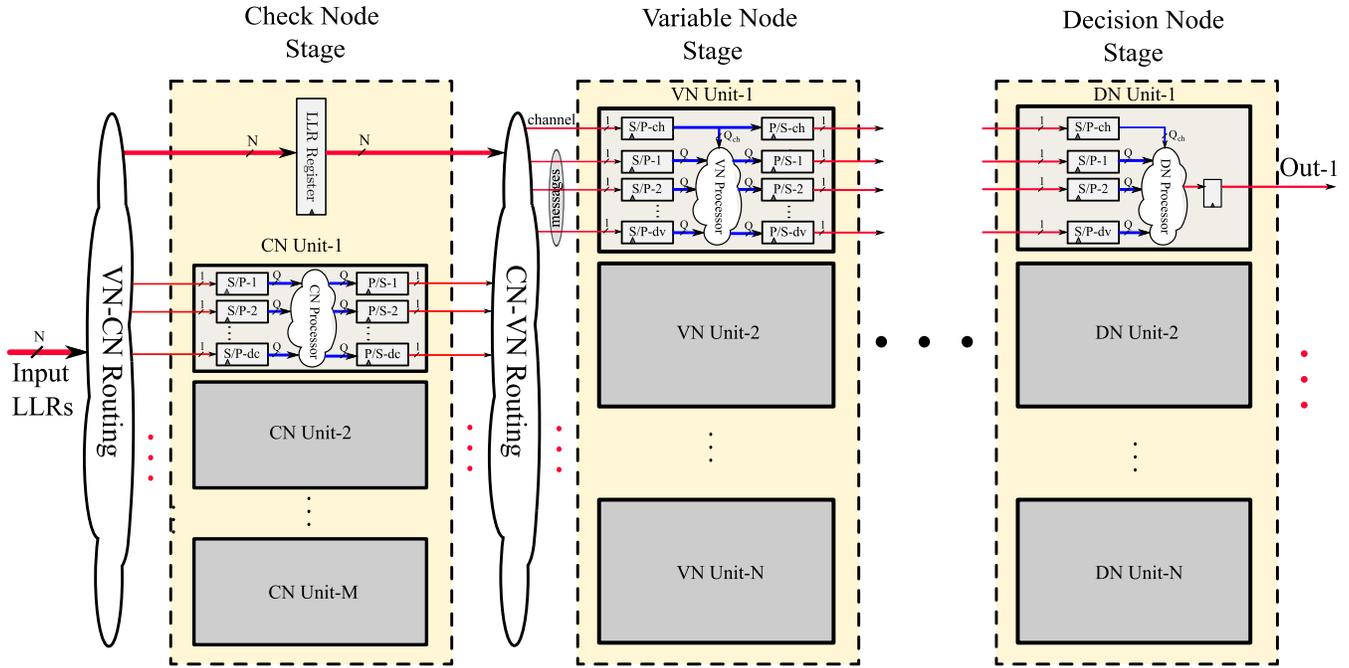


Fig. 2. Serial message-transfer decoder architecture.

Hence, in this section, we propose an unrolled full-parallel LDPC decoder architecture that employs a serial message-transfer technique between CNs and VNs.¹ This architecture is similar to the bit-serial implementations of [18] and [19] in the way the messages are transferred; however, as we shall see later, it differs in the fact that it is unrolled and in the way the messages are processed in the CNs and VNs.

A. Decoder Architecture Overview

An overview of the proposed unrolled serial message-transfer LDPC decoder architecture is shown in Fig. 2. As with all unrolled LDPC decoders, each decoding iteration is mapped to a distinct set of N VN and M CN units, which form a processing pipeline. In essence, the unrolled LDPC decoder is a systolic array, in which a new set of N channel LLRs is read in every clock cycle, and a decoded codeword is put out in every clock cycle.

Even though both the CNs and VNs can compute their outgoing messages in a single clock cycle, similar to the architecture in [5], in the proposed serial message-transfer architecture each message is transferred one bit at a time between the consecutive VN and CN stages over the course of Q_{msg} clock cycles, where Q_{msg} is the number of bits used for the messages. More specifically, each CN and VN unit contains a serial-to-parallel (S/P) and parallel-to-serial (P/S) conversion unit at the input and output, respectively, which are clocked Q_{msg} times faster than the processing clock to collect and transfer messages serially, while keeping the

¹We note that a first implementation of our decoder was based on parallel (word-level) message-transfer. The place and route tool for this implementation was hardly able to converge even when the area utilization was unacceptable. We, therefore, propose serial message-transfer architecture and we adopt special implementation methodologies for this architecture, which will be explained in Section V.

overall decoding throughput constant. More details on the architecture of the CN and VN units as well as the proposed serial message-transfer mechanism are provided in the sequel.

B. Decoder Stages

The unrolled LDPC decoder, illustrated in Fig. 2, consists of three types of processing stages, which are described in more detail below. We note that the CN/VN processors of this reference decoder are similar to those of a standard MS decoder, and our modifications for these parts (to realize a finite-alphabet decoder) are discussed in Section IV.

1) *Check Node Stage:* Each check node stage consists of M CN units, each of which contains three components: a CN processor, which implements (6) similar to [5], [14], d_c S/P units for the d_c input messages, and d_c P/S units for the d_c output messages. Moreover, the complete check node stage contains a register bank that is used to store the channel LLRs, which are not directly needed by the check node stage, but nevertheless must be forwarded to the following VN stage and thus need to be buffered. Hence, no S/P and P/S units are required for the channel LLR buffers in the check node stage as they are simply forwarded serially to the following variable node stage.

2) *Variable Node Stage:* Each variable node stage consists of N VN units, each of which contains a VN processor and S/P and P/S units at the inputs and outputs, respectively, similar to the CN unit structure. Each VN processor implements the update rule (5) similar to [5].

3) *Decision Node Stage:* The last variable node stage is called a decision node stage because it is responsible for taking the final hard decisions on the decoded codeword bits. The structure of this stage is similar to a variable node stage, but a decision node (DN) has a simpler version of the VN processor

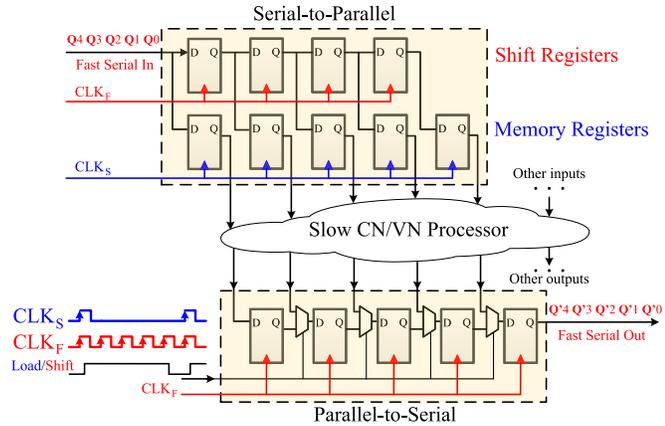


Fig. 3. Message receive and transfer mechanism by S/P and P/S shift registers, enabled by the fast clock (CLK_F), and the message process enabled by the slow clock (CLK_S) for $Q_{\text{msg}} = 5$.

that only computes sum of all inputs and put out its sign [5], and thus no P/S unit is required at its output.

C. Message Transfer Mechanism

One of the modifications, compared to [14] and [5], is the serial transfer of the channel and message LLRs between the stages of the decoder, which reduces the required routing resources by a factor of Q_{msg} . This modification is applied to make the placement and routing of the decoder feasible, especially for large values of d_b and d_c . To this end, as explained in the previous section, an S/P and a P/S shift register are added to each input and each output of the CN and VN units, as illustrated in Fig. 3. We see that the S/P unit consists of a $(Q_{\text{msg}} - 1)$ -bit shift register and Q_{msg} memory registers, while the P/S unit has Q_{msg} registers with multiplexed inputs. The serial messages are transferred with a fast clock, denoted by CLK_F , that is Q_{msg} times faster than the slow processing clock, denoted by CLK_S . More specifically, at each CN unit and VN unit input, data are loaded serially into the S/P shift register using the fast CLK_F , and after the Q_{msg} th cycle all message bits are stored in memory registers, clocked by the slow CLK_S . The CN/VN processing can then be performed in one CLK_S cycle, and the output messages are saved in the output P/S shift register and transferred serially to the next stage using again CLK_F . At the same time, a new set of messages is serially loaded into the input shift register. We note that for simpler clock tree generation, all registers in Fig. 3 are clocked by CLK_F , while CLK_S is actually implemented as a pulsed clock, which is generated using a clock-gating cell controlled by a finite state machine.

D. Decoder Hardware Complexity and Performance Analysis

In this section, we describe the required memory complexity of the proposed decoder as well as the decoding latency and the throughput.

1) *Memory Requirement*: The memory complexity can easily be characterized by counting the number of required registers and can be approximated by

$$R_{\text{tot}} \approx N(d_b + 1)Q(6I - 1) \quad (8)$$

where I is the number of decoding iterations (which in unrolled decoders strongly affects the memory requirements) and $Q = Q_{\text{msg}} = Q_{\text{ch}}$, which is often the case for MS LDPC decoders. From (8), one can easily see that the quantization bit-width linearly increases the memory requirement for the proposed architecture.

2) *Decoding Latency*: Since each stage has a delay of two CLK_S cycles and there are two stages for each decoding iteration, the decoder latency is $4I \text{ CLK}_S$ cycles or, equivalently, $4IQ_{\text{max}} \text{ CLK}_F$ cycles, where $Q_{\text{max}} = \max(Q_{\text{msg}}, Q_{\text{ch}})$.

3) *Decoding Throughput*: In the proposed unrolled architecture, one decoder codeword is output in each CLK_S cycle. Therefore, the coded throughput of the decoder is

$$T = Nf_{S_{\text{max}}} \quad (9)$$

where $f_{S_{\text{max}}}$ is the maximum frequency of CLK_S , while the maximum frequency of CLK_F or simply maximum frequency of the decoder is $f_{\text{max}} = f_{F_{\text{max}}} = Q_{\text{max}}f_{S_{\text{max}}}$. For the proposed architecture, we have

$$f_{S_{\text{max}}} = \left\{ \max \left((Q_{\text{max}}T_{\text{CP,route}}), (T_{\text{CP,VN}}), (T_{\text{CP,CN}}) \right) \right\}^{-1} \quad (10)$$

where $T_{\text{CP,VN}}$ and $T_{\text{CP,CN}}$ are the delay of the critical paths of the CN unit and the VN unit, respectively, and $T_{\text{CP,route}}$ is the critical path delay of the (serial) routing between the decoding stages. Thus, the decoder throughput will be limited by the routing, if the VN/CN delay is smaller than Q_{max} times the routing delay. Hence, on the one hand, the serial message-transfer decoder alleviates the routing problem by reducing the required number of wires, but on the other hand, the decoder throughput for large quantization bit-widths may be affected, as the serial message-transfer delay will become the limiting factor.

IV. FINITE-ALPHABET SERIAL MESSAGE-TRANSFER LDPC DECODER

Even though the serial message-transfer architecture alleviates the routing congestion of an unrolled full-parallel LDPC decoder, it has a negative impact on both throughput and hardware complexity, as discussed in the previous section. In our previous work [5], [6], we have shown that finite-alphabet decoders have the potential to reduce the required number of message bits, while maintaining the same error rate performance. In this section, we will review the basic idea and our design method for this new type of decoders and then show how the bit-width reduction technique of [5], [6] can be applied verbatim in order to increase the throughput and reduce the area of the serial message-transfer architecture.

A. Mutual Information-Based Finite-Alphabet Decoder

In our approach of [5], [6], the standard MP decoding algorithm update rules are replaced by custom update rules that can be implemented as simple lookup tables (LUTs). These LUTs take integer-valued input messages and produce a corresponding output message. Moreover, the input-output mapping that is represented by the LUTs is designed in a way that maximizes the mutual information between the LUT output messages and the codeword bit that these messages correspond to. We note that a similar approach was also used

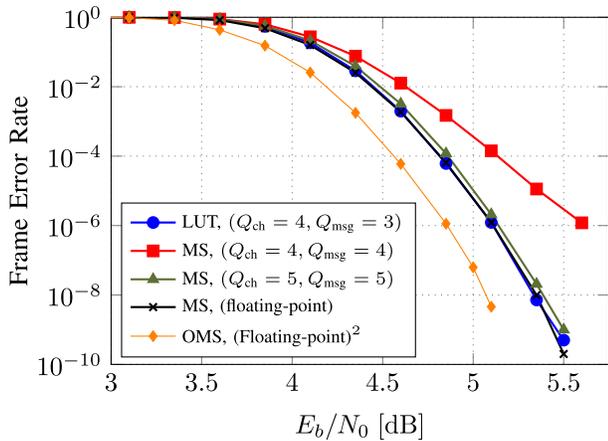


Fig. 4. Frame error rate of the IEEE 802.3an LDPC code under floating-point MS decoding, fixed-point MS decoding with different bit-widths, LUT-based decoding, and floating-point OMS decoding (offset = 0.5) as reference, all with $I = 5$ decoding iterations.

in [7], but the corresponding hardware implementation would have a much higher hardware complexity than the method of [5] and [6]. This happens because, contrary to [7], in [5] and [6], we used LUTs only for the VNs, while the CNs use the standard MS update rule.

B. Error-Correcting Performance and Bit-Width Reduction

In Fig. 4, we compare the performance of the IEEE 802.3an LDPC code under floating-point MS decoding, fixed-point MS decoding (with $Q_{ch} = Q_{msg} \in \{4, 5\}$), and LUT-based decoding (with $Q_{ch} = 4$ and $Q_{msg} = 3$) when performing $I = 5$ decoding iterations with a flooding schedule. We also show the performance of a floating-point OMS decoder as a reference.² We observe that the fixed-point decoder with $Q_{ch} = Q_{msg} = 5$ has almost the same performance as the floating-point decoder, while the fixed-point decoder with $Q_{ch} = Q_{msg} = 4$ shows a significant loss with respect to the floating-point implementation. Thus, a standard MS decoder would need to use at least $Q_{ch} = Q_{msg} = 5$ quantization bits. The LUT-based decoder, however, can match the performance of the floating-point decoder with only $Q_{ch} = 4$ channel quantization bits and $Q_{msg} = 3$ message quantization bits.³ Additionally, with the above quantization bit choices, no noticeable error floor has been observed for both the MS decoder and LUT-based decoder when 10^{10} frames have been transmitted (which corresponds to a bit error rate $\approx 10^{-12}$ with the current block length). We note that, for the LUT-based decoder, the performance in the error floor region can be traded with the performance in the waterfall region by an appropriate choice of the design signal-to-noise ratio (SNR) for the LUTs [6].

²The reference simulation was obtained and matched with our simulation by using the open-source simulator provided by: Adrien Cassagne; Romain Tajan; Mathieu Lonardon; Baptiste Petit; Guillaume Delbergue; Thibaud Tonnellier; Camille Leroux; Olivier Hartmann, AFF3CT: A Fast Forward Error Correction Tool, 2016. [Online]. Available: <https://doi.org/10.5281/zenodo.167837>

³We note that reducing Q_{ch} further results in a non-negligible loss with respect to the floating-point decoder.

C. LUT-Based Decoder Hardware Architecture

The LUT-based serial message-transfer decoder hardware architecture is very similar to the MS decoder architecture, described in Section III. However, the LUT-based decoder can take advantage of the significantly fewer message bits that need to be transferred from one decoding stage to the next. This reduction reduces the number of CLK_F cycles per iteration, which in turn increases the throughput of the decoder according to (10) provided that the CN/VN logic is sufficiently fast. Moreover, the size of the buffers needed for the S/P and P/S conversions is also reduced significantly, which directly reduces the memory complexity of the decoder [see (8)].

On the negative side, we remark that the VN units for each VN stage (decoder iteration) of the LUT-based decoder are different, which slightly complicates the hierarchical physical implementation as we will see later. Furthermore, since $Q_{ch} > Q_{msg}$, we now need to transfer the channel LLRs with multiple (two) bits per cycle to avoid the need to artificially limit the number of CLK_F cycles per iteration to Q_{ch} rather than to the smaller Q_{msg} . To reflect this modification, we redefine (10) as $Q_{max} = \max(Q_{msg}, \lceil Q_{ch}/2 \rceil)$. While this partially parallel transfer of the channel LLRs impacts routing congestion, we note that the overhead is negligible since the number of channel LLRs is small compared to the total number of messages.

V. IMPLEMENTATION

Despite the use of a serial message-transfer, the physical implementation of the decoders proposed in the previous sections requires special scrutiny since the number of global wires is still significant and the overall area is particularly large. Therefore, in this section, we propose and describe a pseudohierarchical design methodology to implement the serial message-transfer architecture.

A. Physical Design

Due to the large number of identical blocks in the decoder architecture, a bottom-up flow is expected to provide the best results. The CN, VN, and DN units are first placed and routed individually to build hard macros,⁴ and their timing and physical information are extracted. These macros are then instantiated as large cells in the decoder top level. We propose to treat the macros as custom standard cells with identical height to be able to perform the placement using the standard-cell placement, rather than the less capable macro placement of the backend tool, since in our case the number of hard macro instances is extremely large and the interconnect pattern is complex and highly irregular.

Fig. 5 illustrates the proposed physical floorplan for the decoders with the unrolled architecture. In this floorplan, the CN and VN macros within each stage are constrained to be placed into dedicated regions [placement regions in Fig. 5(a)]. This measure enforces the high-level structure of systolic array pipeline, but it also leaves freedom to the placement tool to choose the location for the macros in each stage to

⁴Note that for the LUT-based decoder there are different macros for each VN stage as apposed to the MS decoder.

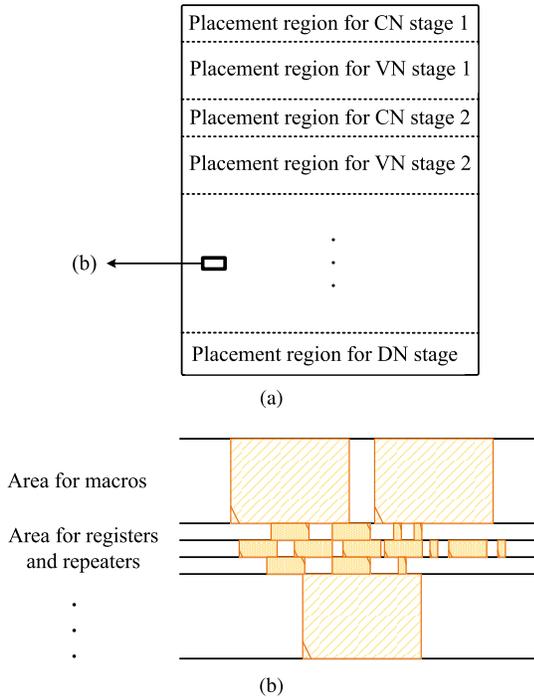


Fig. 5. Physical floorplan for serial message-transfer architecture. (a) High-level overview of the floorplan with dedicated placement regions for each decoder stage. (b) Zoomed-in view of rows structure for custom macros (large colored blocks) and conventional standard-cells (small colored blocks) placement.

minimize routing congestion between stages. Note that the linear floorplan has also the advantage of being scalable in the number of iterations since little interaction or interference exists between stages. Furthermore, the CN and VN macros are placed in dedicated rows while the area between these rows is left for repeaters and for the register standard-cells for the channel LLRs in the CN stages, as shown in Fig. 5(b). We note that the proposed floorplan and the encapsulation of the VN and CN macros as large standard-cells exploit the automated algorithm to optimize both custom and conventional standard cells placement in order to alleviate the significant routing congestion.

B. Timing and Area Optimization Flow

Although the synthesis results can give an approximate evaluation for timing and area of the physical implementation, several iterations with different constraints are required to reach an optimal layout. To this end, we propose the methodology illustrated in the flowchart of Fig. 6 to effectively implement the serial message-transfer architecture. The main idea behind this methodology is that three main factors directly contribute to the decoder throughput and also indirectly to the decoder area, as discussed in Section III and specifically summarized in (10). Our goal is to maximize the throughput at a minimum area.

We define the timing constraint applied to CLK_S as $T_{\text{CSTR},\text{CLK}_S}$, and the timing constraint applied to CLK_F as $T_{\text{CSTR},\text{CLK}_F}$. The first step is to place and route the CN/VN macros based on $T_{\text{CSTR},\text{CLK}_S}$. This step is followed by the implementation of the decoder using $T_{\text{CSTR},\text{CLK}_F}$. (The initial

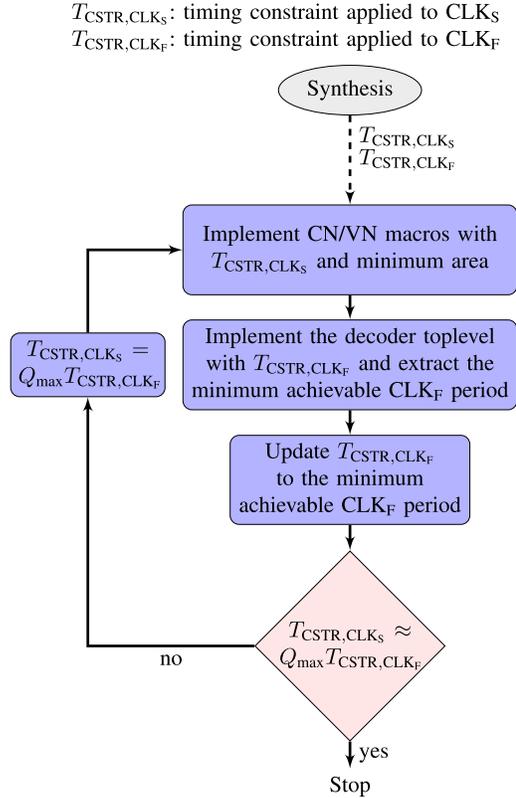


Fig. 6. Proposed flowchart to optimize timing and area for the serial message-transfer architecture.

constraints for the backend are thereby extracted from synthesis timing results.) The fully placed and routed design can give an accurate routing delay, which will be used to update $T_{\text{CSTR},\text{CLK}_F}$ and then $T_{\text{CSTR},\text{CLK}_S}$ according to (10). The updated $T_{\text{CSTR},\text{CLK}_S}$ will be used to reimplement the CN and VN macros within the minimum achievable area.

We note that for a long LDPC code with a large area and long routing delay (such as the one of the IEEE 802.3an standard), the first implementation starts with $T_{\text{CSTR},\text{CLK}_S} < Q_{\text{max}} T_{\text{CSTR},\text{CLK}_F}$. After obtaining a realistic value for the CLK_F period (and hence for $T_{\text{CSTR},\text{CLK}_F}$) at the end of the implementation, the $T_{\text{CSTR},\text{CLK}_S}$ will be updated to a larger value to approach $T_{\text{CSTR},\text{CLK}_S} \approx Q_{\text{max}} T_{\text{CSTR},\text{CLK}_F}$. Consequently, the CN and VN macro area and thus the decoder area will shrink in the second iteration, which result in larger achievable CLK_F frequency and hence smaller $T_{\text{CSTR},\text{CLK}_F}$ and $T_{\text{CSTR},\text{CLK}_S}$. The feedback loop will reach the optimum point after a few iterations.

VI. RESULTS AND DISCUSSIONS

To study the impact of the serial message-transfer architecture and the finite-alphabet decoding scheme, we have implemented the proposed architecture by employing the methodology explained in Section V, and we analyzed the results for both MS and LUT-based decoding. We used the parity-check matrix of the LDPC code defined in the IEEE 802.3an standard [15], i.e., a (2048, 1723) LDPC code of $R = 0.8413$ with $d_b = 6$ and $d_c = 32$. We used $I = 5$ for both decoders and $Q_{\text{msg}} = Q_{\text{ch}} = 5$ for the MS decoder

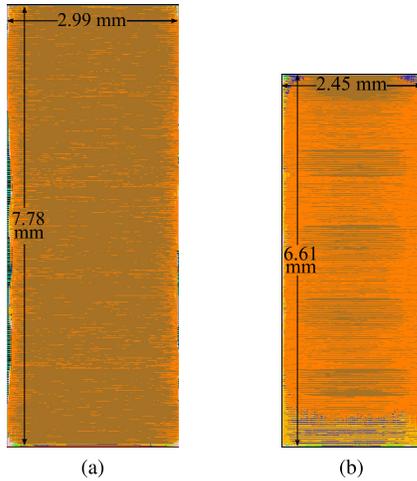


Fig. 7. Layouts for (a) MS decoder and (b) LUT-based decoder.

and $Q_{\text{msg}} = 3$ and $Q_{\text{ch}} = 4$ for the LUT-based decoder to achieve the same error-correction performance, as described in Section IV. The decoders were synthesized from a VHDL description using Synopsys Design Compiler and placed and routed using Cadence Encounter Digital Implementation. The layouts are shown in Fig. 7. The results are reported for a 28-nm FD-SOI library under typical operating conditions ($V_{\text{DD}} = 1$ V and $T = 25$ °C).

A. Delay Analysis

In the serial message-transfer architecture, the critical path and, hence, the maximum decoding frequency are defined by (10). To investigate the impact of serially transferring the messages on the decoder throughput, we consider the delay of the following register-to-register critical paths for both the MS and LUT-based decoder.

1) *CN Critical Path*: The CN critical path ($T_{\text{CP,CN}}$) is the path from the S/P memory registers to the P/S shift register within the CN unit. For both decoders, this path is essentially comprised of the logic cells for a sorter tree with a depth of four.

2) *VN Critical Path*: The VN critical path ($T_{\text{CP,VN}}$) is the path from S/P memory registers to P/S shift register within the VN unit. This path is dominated by an adder tree for the MS decoder and an LUT tree for the LUT-based decoder.

3) *Routing Critical Path*: The routing critical path ($T_{\text{CP,route}}$) comprises mainly the interconnect wires (and buffers) that connect the CN/VN unit S/P shift register to the VN/CN unit P/S shift register.

Table I summarizes the critical path delays of the CN/VN and the routing path. Together with (10), the values in the table dictate the maximum achievable frequency for CLK_S and CLK_F , respectively, for both of the decoders with the proposed serial message-transfer architecture. We note that the critical paths are reported after the timing and area constraints for the CN/VN macros and for the decoder toplevel are jointly optimized according to the flow shown in Fig. 6. According to (10), we observe that in both decoders the message transfer limits the slow clock CLK_S to a period of 5×1.51 ns = 7.55 ns and 3×1.16 ns = 3.48 ns

TABLE I
CRITICAL PATH DELAYS FOR MS AND LUT-BASED DECODER

Path	MS decoder	LUT-based decoder
CN [ns]	2.38	1.42
VN [ns]	0.96	1.24
Routing [ns]	1.51	1.16

TABLE II
DETAILED AREA FOR CN/VN UNIT*

Component	MS decoder	LUT-based decoder
CN unit logic [μm^2]	1578	485
CN unit register [μm^2]	1695	971
CN macro [μm^2]	3607	1510
VN unit logic [μm^2]	315	403 [†]
VN unit register [μm^2]	381	235 [†]
VN macro [μm^2]	755	646[†]

*Logic and register areas are obtained from synthesis, and macro areas are the final post-layout results.

[†]Although the VNs are different for each stage of the LUT-based decoder, their areas are similar and the result for one of them is reported here.

TABLE III
POWER AND ENERGY EFFICIENCY COMPARISON FOR MS AND LUT-BASED DECODER

	MS decoder	LUT-based decoder
Total power @ f_{max} [mW]	12248	13350
Total power @ 662 MHz [mW]	12248	10257
Leakage power [mW]	7.44	5.27
Energy efficiency [pJ/bit]	45.2	22.7

for the MS and the LUT-based decoder, respectively, where 1.51 and 1.16 ns are the corresponding minimum CLK_F periods. Consequently, in our flow, the VN and CN units end up as optimized for minimum area only with relaxed and easy to meet timing constraints.

B. Area Analysis

Fig. 8 illustrates the area distribution among the various components after the layout. The area utilization is approximately 67% for both decoders. While almost 62% of the layout is filled with CN/VN macros and registers, the clock tree and routing buffers occupy around 5%. Furthermore, we see a 44% difference in total area between the decoders due to the fact that the total area for CN and VN macros is 14.12 mm^2 in the MS decoder, as opposed to only 9.56 mm^2 in the LUT-based decoder.

To understand this fact, we list the area of each CN/VN macro in Table II. According to this table, the finite-alphabet MP algorithm leads to significantly smaller CN processors because of two important factors: first, the bit-width reduction of the messages directly affects the data-path area, and second, the quantized messages in the LUT-based decoder are processed directly in the sorter tree of the CNs without the need to compute their absolute values. However,

TABLE IV
IMPLEMENTATION RESULTS FOR MS AND LUT-BASED DECODER AND COMPARISON WITH OTHER WORKS

	MS decoder	LUT-based decoder	[14]	[8]	[20]	[23]	[27]	[26]
Process technology	28 nm FD-SOI	65 nm CMOS	65 nm CMOS	65 nm CMOS low-power	65 nm CMOS	90 nm CMOS	90 nm CMOS	90 nm CMOS
Supply voltage [V]	1.0	1.2	1.2	1.2	1.3	0.9	1.2	1.0
LDPC code	(2048, 1723)	(672, 546)	(2048, 1723)	(2048, 1723)	(2048, 1723)	(2048, 1723)	(2048, 1723)	(2048, 1723)
Node degree (d_v, d_c)	(6, 32)	(3, 6)	(6, 32)	(6, 32)	(6, 32)	(6, 32)	(6, 32)	(6, 32)
Algorithm	min-sum	finite-alphabet	min-sum	offset min-sum with post processor	split-row	normalized probabilistic min-sum	reduced-complexity min-sum	delayed stochastic
I_{\max}	5	5	9	8	11	9	30	–
Quantization bits	5	3	4	4	5	4	6	5
E_b/N_0 @ BER = 10^{-7} [dB]	4.97	4.95	–	4.25	4.55	4.4	4.32	4.7
Architecture	unrolled full-parallel	unrolled full-parallel	unrolled full-parallel	partial-parallel	full-parallel	full-parallel	full-parallel	full-parallel
Core area [mm ²]	23.3	16.2	12.9	5.05	4.84	9.6	3.84	3.93
Area utilization [%]	66.4	65.9	76	84.5	97	91	–	93
Max. frequency (f_{\max}) [MHz]	662	862	257	700	195	199.6	226	750
Latency [ns]	151	69.6	105	137	56.4	45.09	–	800
Throughput @ I_{\max} [Gb/s]	271	588	160.8	13.3	36.3	45.42	12.8	172.4*
Power @ f_{\max} [mW]	12248	13350	5360	2800	1359	1110	1040	–
Area eff. [Gb/s/mm ²]	11.6	36.3	12.5	2.63	7.5	4.73	3.34	43.86
Energy per bit @ I_{\max} [pJ/b]	45.2	22.7	33.3	210.5	37.4	24.44	81.2	–
Scaled area eff. [†] [Gb/s/mm ²]	11.6	36.3	156.4	32.9	93.8	157.1	110.9	1456.8
Scaled energy per bit [‡] [pJ/b]	45.2	22.7	10	63	9.5	9.4	17.5	–

*Maximum throughput @ $E_b/N_0 = 5.5$ dB (Note that throughput @ I_{\max} is not reported in the original paper)

[†]Scaling is done by S^3 where S is the relative dimension to 28 (Note that this is very rough and optimistic since it does not apply to the interconnects)

[‡]Scaling is done by $1/SU^2$ where U is the relative voltage to 1.0

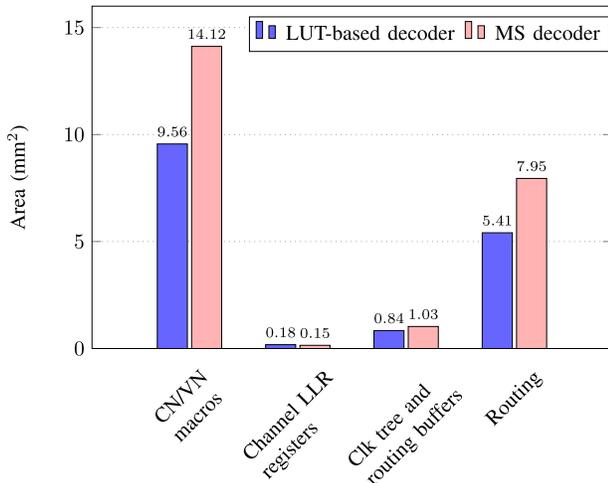


Fig. 8. Detailed area results for the LUT-based and MS decoder with a total area of 16.2 and 23.3 mm², respectively.

VN processors are less area-efficient in the LUT-based decoder in comparison with the ones of the MS decoder. This is caused by the fact that the LUT-based computations are, in general, less area-efficient than the conventional arithmetic-based update rules. Thus, the logic area of the VN in the LUT-based decoder is larger, even though their input/output bit-width is smaller. Another contributing factor in the Table II is the register area, which is defined by the number of S/P and P/S registers. For those, the 40% reduction of bit-width in the LUT-based decoder is directly noticeable in the register area for both CN and VN units. Altogether, the CN and VN macros in the LUT-based decoder are 58% and 14% smaller, respectively, compared to those of the MS decoder.

C. Power Analysis

The energy which is consumed by each decoder is proportional to the capacitance, which in turn is related to the

decoder area. Also, the number of required CLK_F cycles for the serial message-transfer to decode one codeword, which is inversely proportional to the decoding throughput at a constant frequency, directly contributes to the consumed energy for each decoded bit. Therefore, we analyze both the total power and the energy efficiency of the decoders using post-layout vector-based power analysis.⁵ The results are reported in Table III. We note that the total powers are calculated at f_{\max} for both decoders. Also, for comparison purpose, we have calculated the total powers at a constant CLK_F frequency, here $\min(f_{\max, \text{MS}}, f_{\max, \text{LUT}}) = 662$ MHz, for both decoders and note them in Table III. According to this table, the total power consumption of the LUT-based decoder is 16.2% smaller than that of the MS decoder. Furthermore, by considering the fact that the LUT-based decoder has 66.7% higher throughput than the MS decoder at a similar CLK_F frequency, the energy efficiency of the LUT-based decoder is almost two times better in comparison with the MS decoder.

D. Summary and Final Comparison

The final postlayout results for our MS and LUT-based decoders and also for some other recently implemented decoders are summarized in Table IV. Our LUT-based decoder runs at a maximum CLK_F frequency of $f_{\max, \text{LUT}} = 862$ MHz and delivers a sustained throughput of 588 Gb/s, while it occupies 16.2 mm² area and dissipates 22.7 pJ/bit. Compared to the MS decoder, the LUT-based decoder is $1.4\times$ smaller, $2.2\times$ faster, and thus $3.1\times$ more area efficient. It also has 16.2% lower power dissipation and $2\times$ better energy efficiency, when the decoding throughput is taken into account.

The work in [14] is the only other unrolled full-parallel decoder in literature, but it is designed for the

⁵We first extract the parasitic information of both the hard macros and the top level from the placement and routing tool and then read and link them using a power computation tool to generate the complete parasitic information.

IEEE 802.11ad [28] code, which has a shorter block length and smaller node degrees ($d_v = 3$ and $d_c = 6$ as opposed to $d_v = 6$ and $d_c = 32$ for the code used in the design reported in this paper). The work of [8], [20], [23], [26], and [27] are for the same IEEE 802.3an code considered in this paper, but with partial-parallel and full-parallel architectures. The proposed LUT-based decoder has more than an order of magnitude higher throughput compared to [20] and [23], and three times higher throughput compared to [26], while the maximum throughput of the proposed decoder is maintained for all SNR scenarios as it does not require early termination to achieve a high throughput. The area efficiency of the proposed unrolled full-parallel architecture, however, is inferior to the one of the decoders in [20], [23], and [27] with full-parallel architecture due to the repeated routing overhead between the decoder stages in our design.

VII. CONCLUSION

An ultrahigh throughput LDPC decoder with a serial message-transfer architecture and based on nonuniform quantization of messages was proposed to achieve the highest decoding throughput in literature. The proposed decoder architecture is an unrolled full-parallel architecture with serialized messages for CN/VN units, which was enabled by employing S/P and P/S shift registers at the inputs and outputs of each unit. The proposed quantized MP algorithm replaces conventional MS, resulting in 40% reduction in message bit-width without any performance penalty. This algorithm was implemented by using generic LUTs instead of adders for VNs while the CNs remained unchanged compared to MS decoding. Placement and routing results in 28-nm FD-SOI show that the LUT-based serial message-transfer decoder delivers 0.588-Tb/s throughput and is 3.1 times more area efficient and two times more energy efficient in comparison with the MS decoder with serial message-transfer architecture.

REFERENCES

- [1] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [2] M. P. C. Fossorier, M. Mihaljević, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 673–680, May 1999.
- [3] S. K. Planjery, D. Declercq, L. Danjean, and B. Vasic, "Finite alphabet iterative decoders—Part I: Decoding beyond belief propagation on the binary symmetric channel," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4033–4045, Oct. 2013.
- [4] B. M. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Noise thresholds for discrete LDPC decoding mappings," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2008, pp. 1–5.
- [5] A. Balatsoukas-Stimming, M. Meidlinger, R. Ghanaatian, G. Matz, and A. Burg, "A fully-unrolled LDPC decoder based on quantized message passing," in *Proc. IEEE Int. Workshop Signal Process. Syst.*, Oct. 2015, pp. 1–6.
- [6] M. Meidlinger, A. Balatsoukas-Stimming, A. Burg, and G. Matz, "Quantized message passing for LDPC codes," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2015, pp. 1606–1610.
- [7] F. J. C. Romero and B. M. Kurkoski, "Decoding LDPC codes with mutual information-maximizing lookup tables," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 426–430.
- [8] Z. Zhang, V. Anantharam, M. J. Wainwright, and B. Nikolić, "An efficient 10GBASE-T Ethernet LDPC decoder design with low error floors," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 843–855, Apr. 2010.
- [9] A. Cevrero, Y. Leblebici, P. Jenne, and A. Burg, "A 5.35 mm² 10GBASE-T Ethernet LDPC decoder chip in 90 nm CMOS," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2010, pp. 1–4.
- [10] C. Roth, P. Meinerzhagen, C. Studer, and A. Burg, "A 15.8 pJ/bit/liter quasi-cyclic LDPC decoder for IEEE 802.11n in 90 nm CMOS," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2010, pp. 1–4.
- [11] T.-C. Kuo and A. N. Willson, Jr., "A flexible decoder IC for WiMAX QC-LDPC codes," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2008, pp. 527–530.
- [12] A. J. Blanksby and C. J. Howland, "A 690-mW 1-Gb/s 1024-b, rate-1/2 low-density parity-check code decoder," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 404–412, Mar. 2002.
- [13] N. Onizawa, T. Hanyu, and V. C. Gaudet, "Design of high-throughput fully parallel LDPC decoders based on wire partitioning," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 3, pp. 482–489, Mar. 2010.
- [14] P. Schläfer, N. Wehn, M. Alles, and T. Lehnigk-Emden, "A new dimension of parallelism in ultra high throughput LDPC decoding," in *Proc. IEEE Int. Workshop Signal Process. Syst.*, Oct. 2013, pp. 153–158.
- [15] *IEEE Standard for Information Technology Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks Specific Requirements Part 3: Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment 4: Ethernet Operation Over Electrical Backplanes*, IEEE Standard 802.3an, Sep. 2006.
- [16] K. Cushon, P. Larsson-Edefors, and P. Andrekson, "Low-power 400-Gbps soft-decision LDPC FEC for optical transport networks," *J. Lightw. Technol.*, vol. 34, no. 18, pp. 4304–4311, Sep. 15, 2016.
- [17] H. Kaeslin, *Digital Integrated Circuit Design: From VLSI Architectures to CMOS Fabrication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [18] A. Darabiha, A. C. Carusone, and F. R. Kschischang, "A 3.3-Gbps bit-serial block-interlaced min-sum LDPC decoder in 0.13- μ m CMOS," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2007, pp. 459–462.
- [19] A. Darabiha, A. C. Carusone, and F. R. Kschischang, "Power reduction techniques for LDPC decoders," *IEEE J. Solid-State Circuits*, vol. 43, no. 8, pp. 1835–1845, Aug. 2008.
- [20] T. Mohsenin, D. N. Truong, and B. M. Baas, "A low-complexity message-passing algorithm for reduced routing congestion in LDPC decoders," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 5, pp. 1048–1061, May 2010.
- [21] N. Mobini, A. H. Banihashemi, and S. Hemati, "A differential binary message-passing LDPC decoder," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2518–2523, Sep. 2009.
- [22] K. Cushon, S. Hemati, C. Leroux, S. Mannor, and W. J. Gross, "High-throughput energy-efficient LDPC decoders using differential binary message passing," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 619–631, Feb. 2014.
- [23] C.-C. Cheng, J.-D. Yang, H.-C. Lee, C.-H. Yang, and Y.-L. Ueng, "A fully parallel LDPC decoder architecture using probabilistic min-sum algorithm for high-throughput applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 9, pp. 2738–2746, Sep. 2014.
- [24] S. S. Tehrani, W. J. Gross, and S. Mannor, "Stochastic decoding of LDPC codes," *IEEE Commun. Lett.*, vol. 10, no. 10, pp. 716–718, Oct. 2006.
- [25] S. S. Tehrani, A. Naderi, G.-A. Kamendje, S. Hemati, S. Mannor, and W. J. Gross, "Majority-based tracking forecast memories for stochastic LDPC decoding," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4883–4896, Sep. 2010.
- [26] A. Naderi, S. Mannor, M. Sawan, and W. J. Gross, "Delayed stochastic decoding of LDPC codes," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5617–5626, Nov. 2011.
- [27] F. Angarita, J. Valls, V. Almenar, and V. Torres, "Reduced-complexity min-sum algorithm for decoding LDPC codes with low error-floor," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 7, pp. 2150–2158, Jul. 2014.
- [28] *ISO/IEC/IEEE International Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band (Adoption of IEEE Std 802.11ad-2012)*, IEEE Standard 802-11:2012/Amd.3:2014(E), Mar. 2014, pp. 1–634.



Reza Ghanaatian was born in Jahrom, Iran, in 1988. He received the M.Sc. degree in digital systems from the Department of Electrical Engineering, Sharif University of Technology (SUT), Tehran, Iran, in 2012. He is currently working toward the Ph.D. degree at the Telecommunications Circuits Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

He was with the Advanced Integrated Circuit Design Laboratory, SUT, from 2011 to 2013, where he was involved in field-programmable gate array-based systems for wireless and optical communication applications. His current research interests include VLSI circuits for signal processing and communications, as well as approximate computing techniques for energy efficient system design.



Alexios Balatsoukas-Stimming received the Diploma and M.Sc. degrees in electronics and computer engineering from the Technical University of Crete, Chania, Greece, in 2010 and 2012, respectively, and the Ph.D. degree in computer and communications sciences from École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, where he performed his research at the Telecommunications Circuits Laboratory.

His current research interests include VLSI circuits for signal processing and communications, as well as error correction coding theory and practice.

Dr. Balatsoukas-Stimming received a scholarship from the Alexander S. Onassis foundation for his M.Sc. studies. He serves as a reviewer for several IEEE journals and conferences, and has been recognized as an Exemplary Reviewer by the IEEE WIRELESS COMMUNICATIONS LETTERS in 2013 and 2014.



Thomas Christoph Müller (S'16) received the bachelor's degree in information technology from the Schmalkalden University of Applied Sciences, Schmalkalden, Germany, in 2011, and the master's degree in system-on-chip from Lund University, Lund, Sweden, in 2013. He is currently pursuing the Ph.D. degree with the Telecommunication Circuits Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, with a research focus on digital implementation and, an emphasis on low power, variation mitigation, and design

methodology.

He has been a Project/Research Assistant with Lund University and with the Technical University of Denmark, Kongens Lyngby, Denmark.



Michael Meidlinger was born in 1989. He received the B.Sc. and M.Sc. degrees (with distinction) from Technische Universität Vienna, Vienna, Austria, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree with the Communication Theory Group.

From 2011 to 2013, he was involved in mobile communication research and in the development of the Vienna LTE-A simulators. His current research interests include quantizer design for telecommunication receivers as well as error correction coding and superposition modulation techniques.



Gerald Matz (SM'07) received the Dipl.-Ing. and Dr. Techn. degrees in electrical engineering and the Habilitation degree in communication systems from the Vienna University of Technology, Vienna, Austria, in 1994, 2000, and 2004, respectively.

He has held visiting positions at the Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, Toulouse, France, in 2004, the Communication Theory Laboratory, ETH Zürich, Zürich, Switzerland, in 2007, and the École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse, Toulouse, France, in 2011. He currently holds a tenured Associate Professor position with the Institute of Telecommunications, Vienna University of Technology. He has directed or actively participated in several research projects funded by the Austrian Science Fund, by the Viennese Science and Technology Fund, and by the European Union. He has published over 200 scientific articles in international journals, conference proceedings, and edited books. He is the coeditor of the book *Wireless Communications Over Rapidly Time-Varying Channels* (New York, NY, USA: Academic, 2011). His current research interests include wireless networks, statistical signal processing, information theory, and big data.

Prof. Matz was a member of the IEEE SPS Technical Committee on Signal Processing Theory and Methods and the IEEE SPS Technical Committee on Signal Processing for Communications and Networking. He is a member of the ÖVE. He has been a member of the Technical Program Committee of numerous international conferences. He was the Technical Program Chair of Asilomar 2016, the Technical Program Co-Chair of EUSIPCO 2004, the Technical Area Chair for MIMO Communications and Signal Processing at Asilomar 2012, and the Technical Area Chair for Array Processing at Asilomar 2015. He was a recipient of the Cardinal Innitzer Most Promising Young Investigator Award in 2006. He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY (2013–2015), the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006–2010), the *EURASIP Journal Signal Processing* (2007–2010), and the IEEE SIGNAL PROCESSING LETTERS (2004–2008).



Adam Teman received the Ph.D., M.Sc., and B.Sc. degrees in electrical engineering from Ben-Gurion University (BGU), Be'er Sheva, Israel, in 2006, 2011, and 2014, respectively.

He was a Design Engineer at Marvell Semiconductors, Yokne'Am Illit, Israel, from 2006 to 2007, with a focus on physical implementation. He was a Post-Doctoral Researcher at the Telecommunications Circuits Laboratory, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, under a Swiss Government Excellence Scholarship from 2014–2015. In 2015, he joined the Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel, where he is currently a Tenure Track Researcher with the Department of Electrical Engineering and a Leading Member of the Emerging Nanoscaled Integrated Circuits and Systems Laboratories. He has authored over 40 scientific papers and 4 patent applications. His current research interests include low-voltage digital design, energy efficient SRAM, NVM, and eDRAM memory arrays, low-power CMOS image sensors, low-power design techniques for digital and analog VLSI chips, significance-driven approximate computing, and process tolerant design techniques.

Dr. Teman was honored with the Electrical Engineering Department's Teaching Excellence recognition at BGU in 2010–2012. He was a recipient of the BGU's Outstanding Project Award in 2011, the Yizhak Ben-Yaakov HaCohen Prize in 2010, the BGU Rector's Prize for Outstanding Academic Achievement in 2012, the Wolf Foundation Scholarship for excellence of 2012, and the Intel Prize for Ph.D. students in 2013. His doctoral studies were conducted under a Kreitman Foundation Fellowship. He is an Associate Editor of the *Microelectronics Journal* and a Technical Committee Member of several IEEE conferences.



Andreas Burg (S'97–M'05) was born in Munich, Germany, in 1975. He received the Dipl.-Ing. degree from the Swiss Federal Institute of Technology Zürich (ETH Zürich), Zürich, Switzerland, in 2000, and the Dr.Sc.Techn. degree from the Integrated Systems Laboratory, ETH Zürich, in 2006.

In 1998, he was with Siemens Semiconductors, San Jose, CA, USA. He was with the Bell Labs Wireless Research, New Jersey, USA, for one year. From 2006 to 2007, he was a Post-Doctoral Researcher at the Integrated Systems Laboratory and at the Communication Theory Group, ETH Zürich. In 2007, he co-founded Celestrius, an ETH-spinoff in the field of MIMO wireless communication, where he was responsible for the ASIC development as the Director for VLSI. In 2009, he joined ETH Zürich as an SNF Assistant Professor and the Head of the Signal Processing Circuits and Systems Group, Integrated Systems Laboratory. Since 2011, he has been a Tenure Track Assistant Professor with the École Polytechnique

Federale de Lausanne, Lausanne, Switzerland, where he is leading the Telecommunications Circuits Laboratory.

Mr. Burg is also a member of the EURASIP SAT SPCN and the IEEE TC-DISPS. In 2000, he was a recipient of the Willi Studer Award and the ETH Medal for his diploma and his diploma thesis, respectively. He was also a recipient of ETH Medal for his Ph.D. dissertation in 2006. In 2008, he received a 4-year grant from the Swiss National Science Foundation (SNF) for an SNF Assistant Professorship. With his students he received the Best Paper Award from the *EURASIP Journal on Image and Video Processing* in 2013 and the Best Demo/Paper Awards at the ISCAS 2013, ICECS 2013, and ACSSC 2007. He has served on the TPC of various conferences on signal processing, communications, and VLSI. He was the TPC Co-Chair for VLSI–SoC 2012 and the TCP Co-Chair for the ESSCIRC 2016 and SIPS 2017. He served as an Editor for the IEEE TRANSACTION OF CIRCUITS AND SYSTEMS in 2013 and is on the Editorial board of the *Springer Microelectronics Journal* and the *MDPI Journal on Low-Power Electronics and its Applications*.