

# REDUCED-COMPLEXITY TRELLIS MIN-MAX DECODER FOR NON-BINARY LDPC CODES

*Huyen Pham Thi and Hanho Lee*

Dept. of Information and Communication Engineering,  
Inha University, Incheon, 22212, Korea  
E-mails: phamhuyenmta87@gmail.com, hhlee@inha.ac.kr

## ABSTRACT

In this paper, a novel algorithm and corresponding reduced-complexity decoder architecture are proposed for decoding the trellis min-max NB-LDPC code. This proposal reduces the number of messages exchanged between check node and variable node as well as the hardware complexity. Thus, the memory requirement and the wiring congestion is decreased, which increases the throughput of the decoder with a negligible error-correcting performance loss. A layered decoder architecture is implemented for the (2304, 2048) NB-LDPC code over GF(16) based on the proposed algorithm with a 90-nm CMOS technology. The results show an area reduction of 19.4% for the check node unit, 26.56% for the whole decoder and a throughput of 1396 Mbps with almost similar error-correcting performance, compared to previous works.

**Index Terms**— Nonbinary LDPC codes, check node processing, trellis min-max, layered decoding, VLSI design.

## 1. INTRODUCTION

Non-binary low-density parity-check (NB-LDPC) codes are defined over Galois-Field GF( $q$ ) ( $q > 2$ ) including  $q$  elements, which are investigated by Davey and MacKay [1]. NB-LDPC codes outperform their binary counterparts in terms of error-correcting performance when the code length is moderate. However, their decoder architectures have high complexity and large memory requirements, especially for the check node unit (CNU). This limits the maximum throughput and minimum area for their architectures.

The original NB-LDPC decoding algorithm is Q-ary sum-of-product algorithm (QSPA) [1], which is well-known for the optimal error-correcting performance at the cost of high complexity. To reduce the decoding complexity, several approximation algorithms such as extended min-sum (EMS) [2] and min-max [3] algorithms were proposed to decrease the CNU complexity as the bottleneck of the NB-LDPC decoder. The min-max algorithm [3, 4] using comparisons instead of

additions in [2] leads to a simplify in the CNU architecture and a prevention of the numerical growth. However, both algorithms have a throughput problem because of the forward-backward scheme applied in the CNU.

Recently, the trellis EMS (T-EMS) algorithm in [5] was proposed to increase the throughput of NB-LDPC decoders at the cost of large area, where the output messages in the CNU are simultaneously generated. In [6], simplified trellis min-max (S-TMM) algorithm was proposed based on the work in [5] to not only increase the decoder throughput but also reduce the CNU complexity. However, a large memory requirement is required to store the CNU output messages. Moreover, a large number of exchanged messages between check node and variable node causes the routing congestion, which limits the maximum throughput of the decoder. Several algorithms [7, 8] have been proposed to further reduce the CNU complexity and the exchanged messages.

In this paper, a novel algorithm and decoder architecture are proposed for NB-LDPC codes to reduce both the CNU complexity and the memory requirement when increasing the  $d_c$  value, compared to the conventional implementation [6]. First, the CNU output messages are generated from only the minimum values that achieves by 1-min finder instead of 2-min finder. Thus an area reduction is obtained. Second, a small set of the CNU output messages is kept to generate all the messages in the variable node unit (VNU). The proposed decoder architecture is implemented for the (2304, 2048) NB-LDPC code over GF(16). The results show an area reduction of 19.4% for the CNU, 26.56% for the whole decoder and a throughput of 1396 Mbps with a negligible error-correcting performance loss.

The rest of this paper is organized as follows: Section 2 presents the modified TEC-TMM decoding algorithm. In Section 3, the CNU and the overall decoder architecture are proposed, and the implementation results are discussed. Finally, conclusions are given in Section 4.

This work was supported by the Basic Science Research Program through the NRF funded by the Ministry of Science, ICT and Future Planning 2016R1A2B4015421.

---

**Algorithm 1** Modified TEC-TMM algorithm

---

**Input:**  $Q_{mn}, z_n = \arg \min_{a \in GF(q)} Q_{mn}(a); \forall n \in N(m)$

- 1:  $\Delta Q_{mj}(\eta = a \oplus z_j) = Q_{mj}(a); (0 \leq j < d_c)$
- 2:  $\beta = \sum_{j=0}^{d_c-1} z_j \in GF(q)$
- 3:  $\{m1(a), I(a)\} = \varphi\{\Delta Q_{mk}(a)|_{k=0}^{d_c-1}\}$
- 4:  $\{\Delta Q'(a), d1(a), d2(a)\} =$   
 $\min_{\eta'_k(a) \in \text{conf}(1,2)} \{\max_{k=1,2} (m1(\eta'_k(a)))\};$   
 $\Delta Q''(a) = \min_{\eta'_k(a) \in \text{conf}(1,2)} \{\max_{k=1,2} (m1(\eta'_k(a)))\}$
- 5:  $\{\Delta Q'_{m1}, \Delta Q'_{m2}, a_{m1}, a_{m2}\} = \varphi'\{\Delta Q'(a)|_{a=1}^{q-1}\}$

**Output:**  $\begin{cases} \Delta Q'_{m1}, \Delta Q'_{m2}, a_{m1}, a_{m2} \\ \Delta Q''(a) \\ d1(a) \\ d2(a) \\ z_n^* = z_n \oplus \beta \end{cases}$

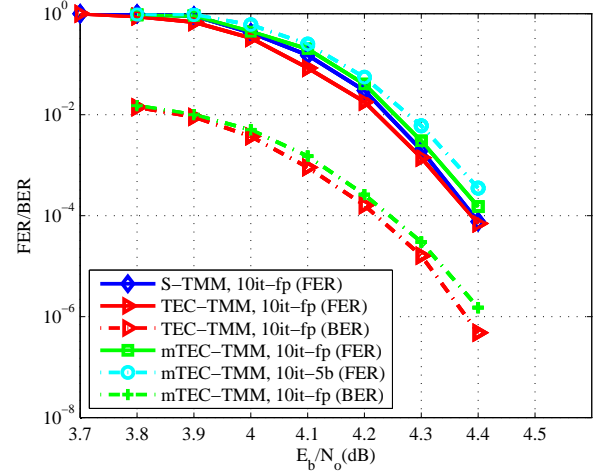
---

## 2. MODIFIED TEC-TMM DECODING ALGORITHM

### 2.1. Review of NB-LDPC codes

NB-LDPC codes are defined by a sparse parity-check matrix  $\mathbf{H}$  including  $M$  rows and  $N$  columns, where nonzero elements  $h_{mn}$  belong to the  $GF(q)$ . Let  $d_c$  ( $d_v$ ) be row weight (column weight) of matrix  $\mathbf{H}$ . Regular NB-LDPC codes are considered with the fixed values of  $d_c$  and  $d_v$ . Furthermore, NB-LDPC codes are presented by a Tanner graph corresponding to matrix  $\mathbf{H}$ , where variable nodes present  $N$  columns and check nodes present  $M$  rows.  $N(m)$  and  $M(n)$  denote the set of  $d_c$  variable nodes connected to check node  $m$  and  $d_v$  check nodes connected to variable node  $n$ , respectively.  $Q_{m,n}(a)$  and  $R_{n,m}(a)$  denote the exchanged messages from  $n$  variable node to  $m$  check node (V2C) and from  $m$  check node to  $n$  variable node (C2V) for each symbol  $a \in GF(q)$ .

In [6], the S-TMM algorithm generates the CNU output messages in a parallel way by constructing an extra column  $\Delta Q(a)$ . First,  $d_c$  log-likelihood ratios (LLRs) vectors from variable nodes  $Q_{m,n}(a)$  are converted to the delta domain as  $\Delta Q_{m,n}(a + z_n) = Q_{m,n}(a)$ , where  $z_n$  is the  $n$ -th hard-decision symbol with the highest reliability. Thus, the first row of the delta trellis includes the most reliable symbols that is the optimal path for zero field element (path 0). The optimal path of nonzero field elements is the deviation from path 0, which has at least one nonzero field element with nonzero LLR. The minimum values  $m1(a)$  in the remain trellis rows are used to construct the extra column  $\Delta Q(a)$ . The configuration set  $\text{conf}(1, 2)$  [2] is a set of possible paths, which generated from the most reliable messages  $m1(a)$ , and a maximum of two deviations. The LLR value and field element of each path is denoted as the maximum of LLRs and sum of the field elements of the nodes in the path, respectively. For each nonzero field element  $a$ , there are  $q/2$  possible paths from the  $\text{conf}(1, 2)$ . The path with the smallest LLR value among  $q/2$  paths is the optimal path, and its LLR value is  $\Delta Q(a)$  value for field element  $a$ .



**Fig. 1.** FER and BER performance of (2304, 2048) NB-LDPC code over  $GF(16)$  under the AWGN channel.

### 2.2. Modified TEC-TMM algorithm

In this work, a modified two-extra-column trellis min-max (mTEC-TMM) algorithm is proposed for decoding NB-LDPC codes, as shown in Algorithm 1. Two extra columns  $\Delta Q'(a)$ ,  $\Delta Q''(a)$  and column indexes of deviations such as  $d1(a)$  and  $d2(a)$  are constructed using only the most reliable messages  $m1(a)$  in a similar way with [8]. Then only first two minimum values from  $\Delta Q'(a)$  such as  $\{\Delta Q'_{m1}, \Delta Q'_{m2}\}$  and the extra column  $\Delta Q''(a)$  are used to update the C2V messages. Thus, the number of exchanged bits is reduced from  $2 \times (q-1) \times w + 2 \times (q-1) \times \lceil \log_2 d_c \rceil + d_c \times p$  bits [8, 9] to  $(q+1) \times w + 2 \times (q-1) \times \lceil \log_2 d_c \rceil + (d_c + 2) \times p$  bits that corresponding to  $\{\Delta Q'_{m1}, \Delta Q'_{m2}, a_{m1}, a_{m2}\}$ ,  $\Delta Q''(a)$ ,  $d1(a)$ ,  $d2(a)$ , and  $z_n^*$ .

The C2V messages  $\Delta R_{m,n}(a)$  are updated as follows. For two elements  $a_{m1}$  and  $a_{m2}$ , the C2V messages are  $\Delta Q'_{m1}$  and  $\Delta Q'_{m2}$  without deviation or the compensate values  $\Delta Q''(a_{m1})$  and  $\Delta Q''(a_{m2})$  at columns with deviations  $d1(a)$  and  $d2(a)$ , respectively. For the  $(q-3)$  remain elements,  $\Delta Q'(a)$  values are approximated using the second minimum value  $\Delta Q'_{m2}$  as depicted (1), where  $\gamma = 1.5$  is selected. Therefore, the C2V messages of the remain elements are updated using either  $\Delta Q'(a)$  or  $\Delta Q''(a)$ , depending on the deviations  $d1(a)$  and  $d2(a)$ . Finally, the C2V messages in the delta domain  $\Delta R_{m,n}(a)$  are converted to the normal domain  $R_{m,n}(a)$  using  $z_j^*$ , where  $z_j^* = z_j \oplus \beta$ .

$$\Delta Q'(a) = \gamma \Delta Q'_{m2} \quad (1)$$

Fig. 1 represents the frame error rate (FER) and bit error rate (BER) performance of (2304, 2048) NB-LDPC code over  $GF(16)$  using the proposed mTEC-TMM algorithm, which show a negligible performance loss, compared to S-TMM [6]

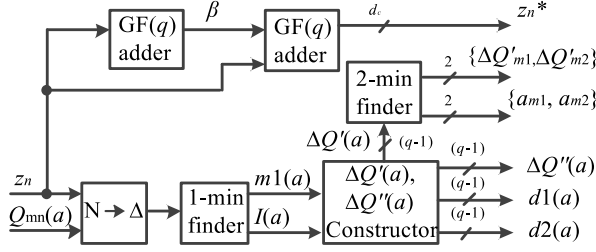


Fig. 2. Top-level CNU architecture.

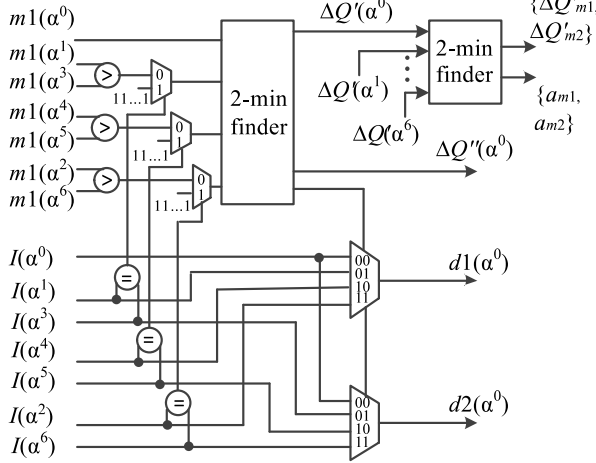


Fig. 3. Two extra column constructor for  $\alpha^0$  over GF(8).

and original two-extra-column trellis min-max (TEC-TMM) [8]. A fixed-point scheme with 5-bit quantization is applied, which introduces a negligible degradation, compared to the floating-point scheme.

### 3. PROPOSED DECODER ARCHITECTURE

#### 3.1. CNU architecture

Fig. 2 shows the proposed top-level CNU architecture. In this work, 1-min finders with  $d_c$  inputs are proposed to find the most reliable messages  $m1(a)$  and its indexes  $I(a)$ . Then, the two extra columns  $\Delta Q'(a)$ ,  $\Delta Q''(a)$  are constructed based on the  $m1(a)$  values using the 2-min finders with  $q/2$  inputs. It is noted that only one 2-min finder with  $(q-1)$  inputs is proposed to find the first two minimum values ( $\Delta Q'_{m1}$ ,  $\Delta Q'_{m2}$ ) and their indexes ( $a_{m1}$ ,  $a_{m2}$ ) from the first extra column  $\Delta Q'(a)$ . The proposed CNU architecture provides a lower complexity because of using the 1-min finders instead of 2-min finders to find the  $m1(a)$  values, compared to previous work [5, 6, 9]. The CNU area is greatly reduced when increasing the  $d_c$  value and  $d_c > q/2$ . In addition, the number

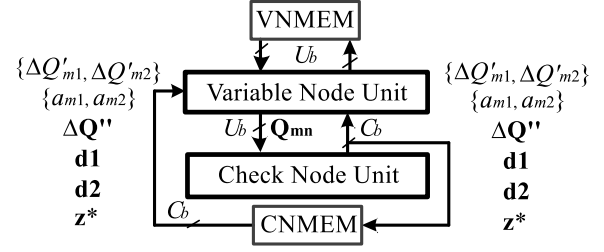


Fig. 4. Top-level decoder architecture.

Table 1. CNU complexity comparison

GF(16), $d_c = 36$	[6]	[9]	Proposed
Gate count (NAND)	80.2K	59.4K	47.86K
Memory (bits)	2880	474	417
Quantization	5	5	5

of exchanged bits is reduced because of only two elements in  $\Delta Q'(a)$  kept. Fig. 3 shows an example for the proposed two extra column constructor for  $\alpha^0$  over GF(8). Updating the C2V messages and conversion these messages from delta to normal domain are performed in the VNU.

Table 1 shows the CNU complexity of the (2304, 2048) NB-LDPC code over GF(16) with  $(d_c, d_v) = (36, 4)$  and the circulant submatrix size  $QC = 64$ , where the methods in [10, 11] are applied to generate the  $\mathbf{H}$  matrix. The results demonstrate that the proposed CNU reduces not only the gate count by 40.3% but also the exchanged bits between CNU and VNU by 85.5%, compared to [6]. Moreover, the proposed CNU reduces the area and the number of exchanged bits by 19.4% and 12%, respectively, compared to [9].

#### 3.2. Top-level decoder architecture

In this work, a layered decoder scheme is implemented, where one row per layer is decoded at one time. In the layered scheme, since the CNU output messages in one iteration are used in the next decoding iteration, a check node memory (CNMEM) with a size of  $M \times ((q+1) \times w + 2 \times (q-1) \times \lceil \log_2 d_c \rceil + (d_c + 2) \times p)$  bits is required to store these messages. The variable node memory (VNMEM) with a size of  $N \times q \times w$  bits is used to store the LLR values from channel in the first iteration, and the LLR values from decoding outputs in the next iterations.

Fig. 4 shows the top-level layered decoder architecture for the NB-LDPC codes, where the CNU architecture is presented in Fig. 2. The VNU module is responsible for 1) permutation and de-permutation of the variable messages according to the nonzero values  $h_{mn}$  of  $\mathbf{H}$  matrix 2) decompression network that generates the C2V messages in the delta domain and convert these messages from delta to normal domain 3) normalization to ensure that the smallest value in each V2C

**Table 2.** Comparison of the proposed (2304, 2048) NB-LDPC layered decoder over GF(16) with other works

GF(16), $d_c = 36$	[6]	[9]	Proposed
Report	Post-layout	Post-layout	Synthesis
Technology	90-nm	90-nm	90-nm
Quantization	5	5	5
Gate count (NAND)	1882K	975K	716K
$f_{clk}$	330.8	333.3	433.8
Throughput (Mbps)	957.5	964.7	1396
Efficiency (Mbps/M gates)	508.8	989.4	1949.7

message vector, which is input vector of the CNU, is equal to zero.

The synthesis results of the proposed (2304, 2048) NB-LDPC layered decoder architecture over GF(16) are shown in Table 2. A 90-nm CMOS standard cell library and synopsys design tools are used to synthesize the proposed decoder architectures. The throughput of the decoder is calculated, as shown in (2), where  $seg$  is the number of pipeline stages used to obtain a balance between throughput and area, and  $I_{max}$  is the maximum number of decoding iterations. In the proposed decoder architecture, a throughput of 1396 Mbps is achieved with  $seg = 6$  and  $I_{max} = 10$ . The area of memories is calculated based on considering that one bit memory is equal to 1.5 NAND gates. In addition, six pipeline stages used in the decoder require 12.4-K registers.

$$Throughput = \frac{f_{clk}[\text{MHz}] \times N \times p}{I_{max} \times (M + d_v \times seg) + QC} [\text{Mbps}] \quad (2)$$

As can be seen that the proposed decoder provides a great reduction of gate count by 62% and 26.56%, compared to [6] and [9], respectively. Reducing both the CNU complexity and the exchanged bits between CNU and VNU leads to a great improvement in the maximum frequency. Thus, the overall efficiency of the proposed decoder is almost twice as high as the one of the decoder in [9].

#### 4. CONCLUSION

In this paper, a novel mTEC-TMM algorithm and corresponding decoder architecture are proposed for decoding the NB-LDPC codes. The results demonstrate that the proposed mTEC-TMM layered decoder greatly reduces both the CNU complexity, the wiring congestion and the memory requirements with negligible performance loss, compared to conventional trellis min-max decoders.

#### 5. REFERENCES

- [1] H.C Davey and D.C MacKay, "Low-density parity check codes over GF(q)," in *Information Theory Workshop*, Jun. 1998, pp. 165–167.
- [2] D. Declercq and M. Fossorier, "Decoding algorithms for nonbinary LDPC codes over GF(q)," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 633–643, Apr. 2007.
- [3] V. Savin, "Min-max decoding for nonbinary LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 960–964.
- [4] H. P. Thi, A. Sabooh, and H. Lee, "High-throughput partial-parallel block-layered decoding architecture for nonbinary LDPC codes," *Integration, the VLSI Journal*, vol. 59, pp. 52–63, May. 2017.
- [5] E. Li, D. Declercq, and K. Gunnam, "Trellis-based extended min-sum algorithm for non-binary LDPC codes and its hardware structure," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2600–2611, Jul. 2013.
- [6] J. O. Lacruz, F. García-Herrero, D. Declercq, and J. Valls, "Simplified trellis min-max decoder architecture for nonbinary low-density parity-check codes," *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 23, no. 9, pp. 1783–1792, Sep. 2015.
- [7] J. O. Lacruz, F. García-Herrero, M. J. Canet, and J. Valls, "High-performance NB-LDPC decoder with reduction of message exchange," *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 24, no. 5, pp. 1950–1961, May. 2016.
- [8] H. P. Thi and H. Lee, "Two-extra-column trellis min-max decoder architecture for nonbinary LDPC codes," *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 25, no. 5, pp. 1787–1791, May. 2017.
- [9] J. O. Lacruz, F. Garcia-Herrero, M. Canet, J. Valls, and A. Pérez-Pascual, "A 630 Mbps non-binary LDPC decoder for FPGA," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May. 2015, pp. 1989–1992.
- [10] C.S. Choi, H. Lee, N. Kaneda, and Y. Chen, "Concatenated non-binary LDPC and HD-FEC codes for 100gb/s optical transport systems," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May. 2012, pp. 1783–1786.
- [11] B. Zhou, J. Kang, S. Song, S. Lin, K. Abdel-Ghaffar, and M. Xu, "Construction of non-binary quasi-cyclic LDPC codes by arrays and array dispersions," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1652–1662, Jun. 2009.