

A 1.31Gb/s, 96.6% Utilization Stochastic Nonbinary LDPC Decoder for Small Cell Applications

Xin-Ru Lee^{1,2}, Chih-Wen Yang^{1,2}, Chih-Lung Chen¹, Hsie-Chia Chang¹ and Chen-Yi Lee¹

¹Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan

²MediaTek Inc., Hsinchu, Taiwan

Email: {mark, hcchang, cylee}@si2lab.org

Abstract—In this paper, an over Gb/s stochastic nonbinary LDPC (NB-LDPC) decoder chip is first-reported. The operation of proposed decoder is transformed to logarithm domain, so that the decoding complexity is mitigated by the simpler summations and fewer bit-width. In addition, the storage requirements are dramatically reduced by truncated TFM architecture. After, benefited from architecture optimizations and symbol-serial property, the routing capability of proposed decoder is extraordinarily enhanced. According to the measurement results, this decoder can deliver 1.31Gb/s throughput under 368MHz clock frequency with the corresponding energy-efficiency of 0.45nJ/bit. Compared to other NB-LDPC decoders, our stochastic NB-LDPC decoder with 96.6% chip utilization improves 2x area-efficiency and 7x energy-efficiency.

I. INTRODUCTION

As the rapid growth of mobile communications, data offloading is considered as a more efficient technique to manage the radio spectrum [1]. A small cell appears as a cost-effective way to alleviate data traffic and enhance the indoor coverage. However, the challenge of small cells comes from the severer interference since the density of small cells can be very high in some area. The low-density parity-check (LDPC) codes are widely used to provide powerful error correcting capability under the lowest signal-to-noise ratio (SNR) [2]. Currently, nonbinary LDPC (NB-LDPC) codes receive more considerable attention due to its ability of resisting against burst errors and better error-rate performance; while the first NB-LDPC decoder chip has revealed a large power consumption of 3.034nJ/bit (3.866W) [3], indicating that the energy-hungry NB-LDPC decoder with complicated processing units and a large amount of storages cannot meet requirements of wireless applications. In this respect, stochastic computation is a promising decoding method for NB-LDPC codes, as shown in Fig. 1. It has higher routing efficiency and simpler computations due to symbol-serial representation of probability. The message vectors in the decoders are converted from a log-density-ratio (LDR) value to a single symbol over $GF(2^p)$. Therefore, not only the parity-check operation is simplified to a finite field addition, but the routing networks can be alleviated with much fewer wires. Assumed that the bit-width of LDR value is 6-bit and the finite field is $GF(2^4)$, wires per edge are reduced from 192 wires of LDR decoding to 8 wires of stochastic decoding. Although pure stochastic algorithms as tracking forecast memories (TFMs) [4] can efficiently reduce the complexity, the lower throughput is

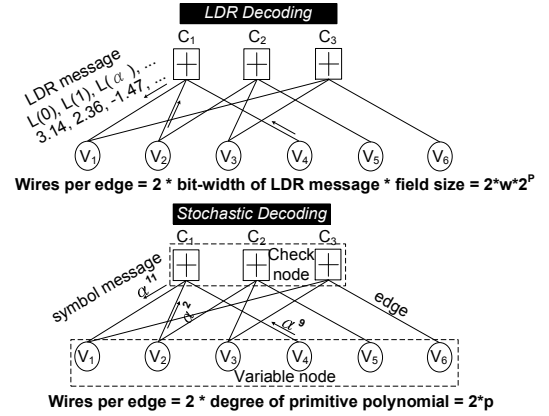


Fig. 1. Comparison of LDR and stochastic decoding.

a bottleneck for practical applications [5]. On the contrary, the relaxed half-stochastic (RHS) algorithm [4], performing symbol-based communication between check and variable nodes but probability-based computation at each variable node locally, provides higher throughput with larger area cost. Hence, it is still a design challenge for NB-LDPC decoders to enhance both area efficiency and routing capability.

In this work, we implement a 0.45nJ/bit, 1.31Gb/s with 96.6% chip utilization NB-LDPC decoder chip for small cell applications. This chip features: (1) logarithm domain operations are proposed to simplify the circuitry and reduce storages. (2) truncated TFM architecture is presented to achieve lower complexity of variable node units (VNUs). (3) a two-stage random number generator is adopted for stochastic process. (4) decoding scheduling is optimized to improve the critical path. The remainder of this paper is organized as follows. Section II introduces the proposed RHS decoding in the logarithm domain. In Section III, an area-efficient VNU architecture as well as its random number design is presented. Section IV reports the implementation results of proposed stochastic NB-LDPC decoder chip, including bit-error-rate (BER) performance and measurement results. Finally, the conclusions are made in Section V.

II. LOG-BASED RELAXED HALF-STOCHASTIC DECODING

In the original RHS algorithm, the check node messages are processed in the symbol domain; while the variable node

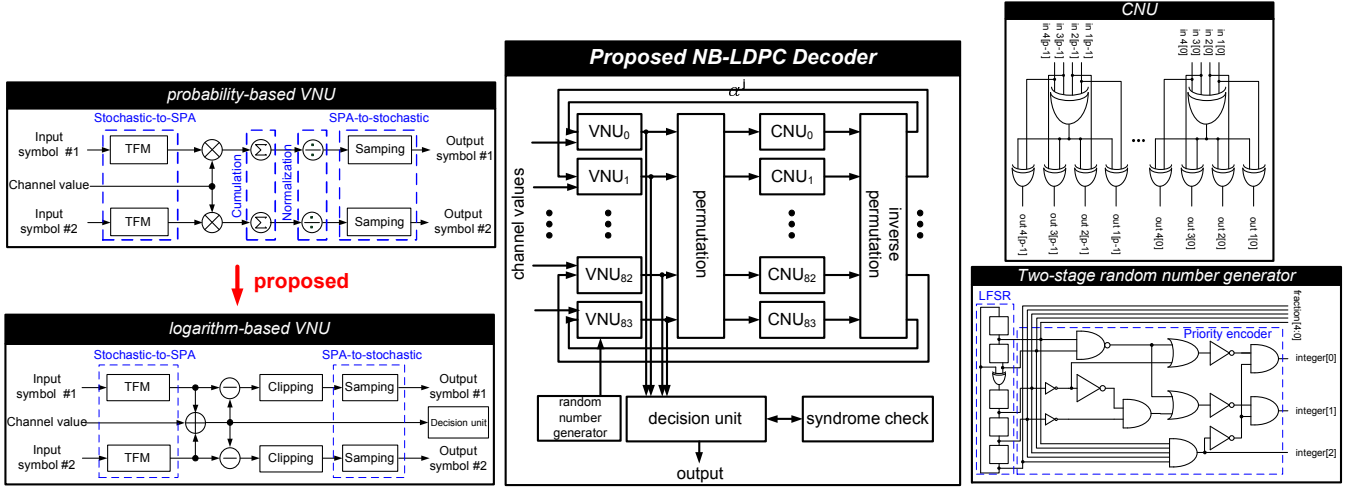


Fig. 2. Proposed stochastic NB-LDPC decoder using log-domain RHS algorithm.

messages are processed in the probability domain based on the sum-product algorithm (SPA) decoding. As shown in Fig. 2, the RHS variable node can be decomposed into two decoding stages, including to handle incoming stochastic symbols and convert them to probabilities, and vice versa. The probability of each stochastic symbol is converted by using a probability tracker called TFM. Then, term-by-term multiplications and message normalization are involved in the SPA decoding. Finally, the generating of stochastic symbols can be fulfilled by comparing SPA results with random numbers. In this respect, the problem of practical implementation is the complicated calculations of VNU.

A hardware-friendly decoding process in which variable nodes are operated in the logarithm domain is introduced. The main idea of proposed logarithm-based VNUs comes from the derivation of successive relaxation [4] in the RHS algorithm as follows:

$$\begin{aligned}
 PMF_t[\gamma] &= (1 - \beta)PMF_{t-1}[\gamma] + \beta \\
 \Rightarrow \log(PMF_t[\gamma]) &= \log(1 - \beta) + \log(PMF_{t-1}[\gamma]) + x \\
 \Rightarrow LMF_t[\gamma] &= LMF_{t-1}[\gamma] - \text{offset} \\
 \Rightarrow LMF_t[\gamma] &= LMF_{t-1}[\gamma] - (a * LMF_{t-1}[\gamma] + b) \\
 \xrightarrow{\text{approx.}} LMF_t[\gamma] &= LMF_{t-1}[\gamma] - (LMF_{t-1}[\gamma] \gg \text{shift-bit}).
 \end{aligned}$$

In other words, the decoding algorithm can be approximated by piecewise linear functions; moreover, the piecewise functions are further simplified by right shift operations to reduce the complexity.

III. PROPOSED AREA-EFFICIENT NB-LDPC DECODER

Our partial parallel (168, 84) over GF(2⁴) NB-LDPC decoder, composed of 84 4-input check node units (CNUs) and 84 2-input variable node units (VNUs), is demonstrated in Fig. 2. One of decoding round between VNUs and CNUs is finished within 2 clock cycles. Because the convolution operations are transformed into stochastic symbol operations, each CNU can be implemented by only exclusive-OR gates. The permutation and inverse permutation deal with finite field multiplication and division. Benefited from logarithm domain

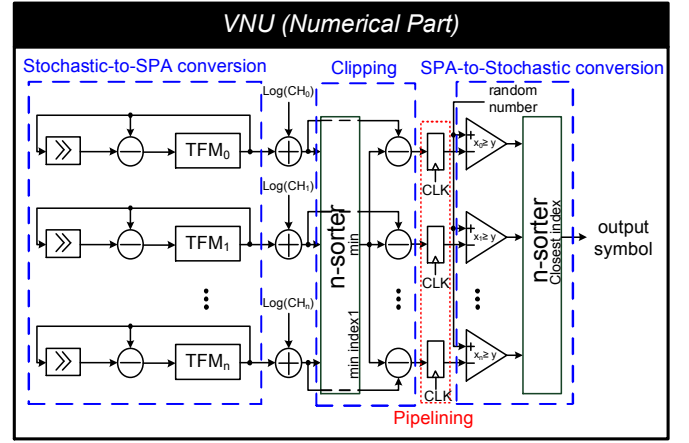


Fig. 3. Proposed VNU architecture in the logarithm domain.

operations, the multiplications of original VNU are reduced to simpler summations, and the bit-width of computation is also decreased. Based on above improvements and symbol-based communication between VNUs and CNUs, the proposed decoder chip achieves 96.6% chip utilization, which is the highest value among silicon-proven LDPC decoders.

The architecture of proposed VNU implemented in the log-domain is illustrated in Fig. 3. Each VNU consists of 2 steps: the first step is a stochastic-to-SPA conversion, in which tracking forecast memories (TFMs) storing the possibilities of each symbol are updated by the input symbol. And the second step is a SPA-to-stochastic conversion, in which a stochastic symbol is generated for neighboring CNUs. Since the updating equations are transformed to logarithm domain, the circuitry can be implemented by right shift operations and subtractors. To prevent numeric overflow, the sorter is used to search the minimal value then normalization is performed by subtractions. The SPA-to-stochastic conversion contains the compare-select method to generate an output symbol. On the other hand, the shuffled scheduling is proposed to trade-off between hardware cost and throughput. Table I shows

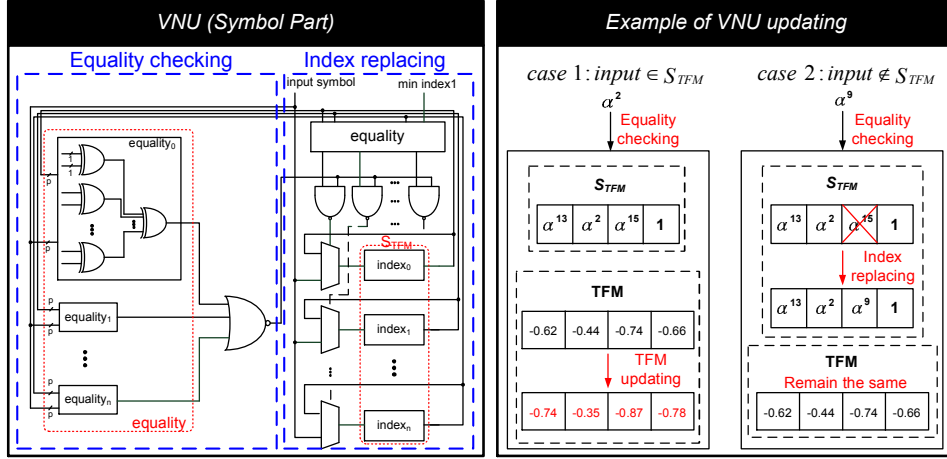


Fig. 4. Architecture of message truncation technique.

TABLE I
SYNTHESIS RESULTS OF DIFFERENT SHUFFLED GROUP

Shuffled group (G)	1	2	4
Critical path (ns)	4.30	2.65	1.90
Decoder gate count (K-gate)	1486	1004	734
Throughput (Mb/s)	1662	1349	941
Hardware efficiency (Mb/s/K-gate)	1.12	1.34	1.28

the synthesis results of different shuffled group numbers. According to the results, the case of 2 shuffled group ($G=2$) achieves the best hardware efficiency. Namely, a pipeline register is inserted within the VNU to improve the critical path. At the first cycle, the stochastic-to-SPA conversion is operated with the preceding 84 variable nodes, while the succeeding 84 nodes are processed at the next cycle. For the SPA-to-stochastic conversion, the generation of stochastic symbol streams depends on comparing a random number with TFM values and sorting the closest symbol.

The architecture of message truncation technique, including equality checking and index replacing, is illustrated in Fig. 4. To reduce the number of registers, there are only n symbols storing in the symbol set S_{TFM} of VNU as output candidates in SPA-to-stochastic conversion. First, the equality checking will check the symbol set S_{TFM} which contains the input symbol or not. It is accomplished by parallel detections composed of exclusive-OR gates. If the input symbol belongs to S_{TFM} , then the TFMs will be updated based on the log-based RHS decoding. That is, the value which corresponds to the input symbol is increased, and other values are decreased. Otherwise, if the input symbol is not involved in the symbol set S_{TFM} , the symbol with the smallest value in the symbol set is replaced by the input symbol; meanwhile, all of values of the TFM remain the same. This index replacing can be implemented through NAND gates and multiplexers. Simulation reveals the length-5 symbol set ($n=5$) maintains BER performance of the proposed decoder.

Due to the distribution of a uniform random number mapped to the logarithm domain is not uniform distributed, a two-stage random number generator consists of a priority encoder and a linear feedback shift register (LFSR) is proposed to

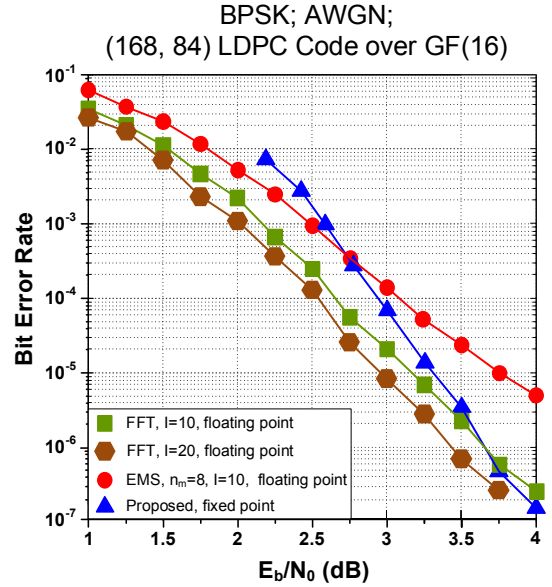


Fig. 5. BER performance of proposed decoder.

generate integer and fraction, respectively. Fig. 2 shows the architecture of random number generator with 5-bit fraction and 3-bit integer. The fraction bits are generated by the LFSR; while the integer bits are generated by the combination of LFSR and priority encoder.

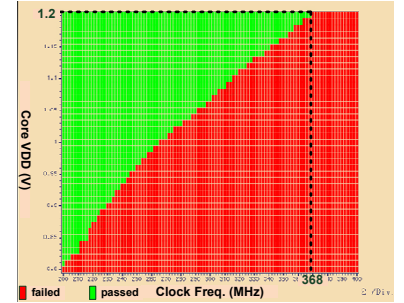
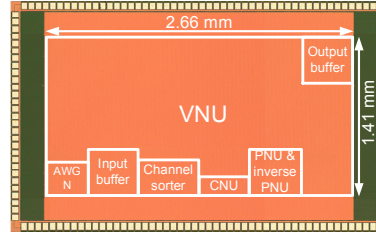
IV. IMPLEMENTATION AND MEASUREMENT RESULTS

The test chip was fabricated in UMC 90nm 1P9M CMOS process. The core area of this chip is $3.75mm^2$, including an AWGN generator and testing circuits. The bit-error-rate (BER) performance is graphed in Fig. 5. According to the performance curves, the proposed decoder outperforms the extended min-sum (EMS) algorithm, which is adopted in the most of NB-LDPC decoders. Fig. 6(a) shows the measurement results. At the standard performance condition which is room temperature and 1V supply, the chip can be operated at 264MHz, achieving a throughput of 943.7Mb/s with 188 computation cycles and consuming a power of 347.1mW. For

TABLE II
COMPARISON OF ENERGY EFFICIENCY AND HARDWARE EFFICIENCY AMONG NB-LDPC DECODERS.

	This work (168, 84) over GF(16)	JSSC' 15 [3] (160, 80) over GF(64)	TCAS-I' 12 [6] (837, 726) over GF(32)	TSP' 13 [7] (837, 726) over GF(32)	TVLSI' 13 [8] (837, 726) over GF(32)	TCAS-II' 15 [9] (168, 84) over GF(16)
Block length	672	960	4185	4185	4185	672
Process (nm)	90	65	180	90	180	90
Gate count (K-gate)	1103	2780	1293	8510	871	1253
Core Area (mm^2)	3.75	7.04	-	46.18	-	-
Frequency (MHz)	368	700	200	250	200	286
Throughput (Mb/s)	1315	1221	64	233.53	66	1131
Power (mW)	587.7	3704	-	893	-	1174
Area efficiency ($Mb/s/mm^2$)	350.67	173.44	-	5.06	-	-
Energy efficiency (nJ/bit)	0.45	3.034	-	3.824	-	1.04
Algorithm	Stochastic	EMS	Simplified Min-Sum	Max-Log-QSPA	Min-Max	Stochastic
Status	measurement	measurement	synthesis	post-layout	synthesis	post-layout

Code	(168, 84) GF(16) (2, 4)-regular		
Block length	672		
Process (nm)	90		
Gate count (K-gate)	1103		
Core Area (mm^2)	3.75		
Chip utilization (%)	96.6		
Supply (V)	1.2	1.0	0.8
Frequency (MHz)	368	264	194
Throughput (Mb/s)	1315	944	693
Power (mW)	587.7	347.1	203.5
Area efficiency ($Mb/s/mm^2$)	350.7	251.7	184.8
Energy efficiency (nJ/bit)	0.45	0.37	0.29



(a) measurement results.

(b) microphotograph.

(c) shmoo plot.

Fig. 6. Implementation results of proposed stochastic NB-LDPC decoder chip

a higher throughput scenario, we increase the supply voltage to 1.2V, where a clock rate of 368MHz is achieved and delivers a throughput of 1.31Gb/s and an area efficiency of $350.67 Mb/s/mm^2$. For low power applications, the supply voltage can be scaled down to 0.8V with a lower operating frequency of 194MHz, leading to the best energy efficiency of 0.29nJ/bit (associated with a power of 203.5mW). Compared to the latest NB-LDPC designs [3] as shown in Table II, the power consumption and core area of this chip are reduced by 85% and 47%, respectively. The microphotograph and shmoo plot of this test chip is shown in Fig. 6(b) and Fig. 6(c).

V. CONCLUSION

In this paper, a 1.31Gb/s partial parallel decoder of a (168, 84) regular-(2, 4) nonbinary LDPC code over $GF(2^4)$ in 90nm CMOS is presented. The logarithm domain transformation as well as a message truncation technique is proposed to effectively reduce the bit-width and storage requirement of message. Moreover, the simpler routing networks profited from stochastic computation with optimized computation units deliver 96.6% logic utilization, which is the highest value of silicon-proven LDPC decoders. Compared to the state-of-the-art works, our proposal achieves better energy efficiency by 7x and area efficiency by 2x as well, making it very suitable for energy-efficient wireless communications.

ACKNOWLEDGMENT

This work is funded by MOE 5Y50B program, MOST 103-2221-E-009-198-MY3, MOST 101-2628-E-009-013-MY3,

and MediaTek-NCTU research center.

REFERENCES

- [1] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile Data Offloading: How Much Can WiFi Deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, Apr. 2013.
- [2] M. Weiner, M. Blagojevic, S. Skotnikov, A. Burg, P. Flatresse, and B. Nikolic, "A Scalable 1.5-to-6Gb/s 6.2-to-38.1mW LDPC Decoder for 60GHz Wireless Networks in 28nm UTBB FDSOI," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2014, pp. 464–465.
- [3] Y. S. Park, Y. Tao, and Z. Zhang, "A Fully Parallel Nonbinary LDPC Decoder With Fine-Grained Dynamic Clock Gating," *IEEE Journal of Solid-State Circuits*, vol. 50, pp. 464–475, Feb. 2015.
- [4] G. Sarkis, S. Hemati, S. Mannor, and W. J. Gross, "Stochastic Decoding of LDPC Codes over $GF(q)$," *IEEE Transactions on Signal Processing*, vol. 61, pp. 939–950, Mar. 2013.
- [5] C. W. Yang, X. R. Lee, C. L. Chen, H. C. Chang, and C. Y. Lee, "Area-efficient TFM-based Stochastic Decoder Design for Non-binary LDPC Codes," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Jun. 2014, pp. 409–412.
- [6] X. Chen and C. L. Wang, "High-Throughput Efficient Non-Binary LDPC Decoder Based on the Simplified Min-Sum Algorithm," *IEEE Transactions on Circuits and Systems—Part I: Regular Papers*, vol. 59, pp. 2784–2794, Nov. 2012.
- [7] Y. L. Ueng, K. H. Liao, C. H. Chou, and C. J. Yang, "A High-Throughput Trellis-Based Layered Decoding Architecture for Non-Binary LDPC Codes Using Max-Log-QSPA," *IEEE Transactions on Signal Processing*, vol. 61, pp. 2940–2951, Nov. 2013.
- [8] F. Cai and X. Zhang, "Relaxed Min-Max Decoder Architectures for Nonbinary Low-Density Parity-Check Codes," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, pp. 2010–2023, Nov. 2013.
- [9] X. R. Lee, C. W. Yang, C. L. Chen, H. C. Chang, and C. Y. Lee, "An Area-Efficient Relaxed Half-Stochastic Decoding Architecture for Nonbinary LDPC Codes," *IEEE Transactions on Circuits and Systems—Part II: Express Briefs*, vol. 62, pp. 301–305, Mar. 2015.