24.2 A 1.15Gb/s Fully Parallel Nonbinary LDPC Decoder with Fine-Grained Dynamic Clock Gating

Youn Sung Park, Yaoyu Tao, Zhengya Zhang

University of Michigan, Ann Arbor, MI

The primary design goal of a communication or storage system is to allow the most reliable transmission or storage of more information at the lowest signal-to-noise ratio (SNR). State-of-the-art channel codes including turbo and binary LDPC have been extensively used in recent applications [1-2] to close the gap towards the lowest possible SNR, known as the Shannon limit. The recently developed nonbinary LDPC (NB-LDPC) code, defined over Galois field (GF), holds great promise for approaching the Shannon limit [3]. It offers better coding gain and a lower error floor than binary LDPC. However, the complex nonbinary decoding prevents any practical chip implementation to date. A handful of FPGA designs and chip synthesis results have demonstrated throughputs up to only 50Mb/s [4-6]. In this paper, we present a 1.15Gb/s fully parallel decoder of a (960, 480) regular-(2, 4) NB-LDPC code over GF(64) in 65nm CMOS. The natural bundling of global interconnects and an optimized placement permit 87% logic utilization that is significantly higher than a fully parallel binary LDPC decoder [7]. To achieve high energy efficiency, each processing node detects its own convergence and applies dynamic clock gating, and the decoder terminates when all nodes are clock gated. The dynamic clock gating and termination reduce the energy consumption by 62% for energy efficiency of 3.37nJ/b, or 277pJ/b/iteration, at a 1V supply.

An NB-LDPC code is formed by grouping bits to symbols using GF elements [3]. The factor graph of an NB-LDPC code has fewer edges compared to an equivalent binary LDPC code, as shown in Fig. 24.2.1, leading to simpler wiring in a decoder implementation. Each edge in an NB-LDPC factor graph carries a vector of messages, thus the global interconnects can be bundled for higher area efficiency. However, GF processing involves a complex check node (CN) and variable node (VN): CN runs a forward-backward algorithm to consider possible symbols and combinations, and both VN and CN use vector sorting to decide the likely candidates [8]. Compared to binary LDPC decoders, an NB-LDPC decoder is heavy on logic gates and light on wiring, making it well suited to a parallel implementation.

We demonstrate a fully parallel (960, 480) GF(64) NB-LDPC decoder architecture that comprises of 160 2-input VNs and 80 4-input CNs, illustrated in Fig. 24.2.2. The VNs and CNs are connected as in the factor graph, with permutation and inverse permutation between VN and CN for GF multiplication and division, and normalization to prevent numerical overflow. The decoder implements the extended min-sum algorithm [8] using an optimized 5b quantization to achieve a very low frame error rate (FER) of 10^{-8} at 3.6dB SNR. Each CN is divided into 6 elementary CN (ECN) steps: 1 for forward traversal of the trellis that describes a parity equation, 1 for backward traversal, and 4 for merging the partial likelihoods to 4 sets of CN-to-VN (C2V) messages. Each VN consists of 3 elementary VN (EVN) steps, 2 of which merge C2V messages with prior likelihoods to compute VN-to-CN (V2C) messages, and 1 computes the posterior likelihood for hard decision. Each ECN and EVN performs searching and sorting, and the processing pipelines are overlapped for a low latency of 48 clock cycles per decoding iteration.

The decoder is mapped to a 6 rows × 40 columns grid floorplan: the inner 2 rows for the 80 CNs and the outer 4 rows for the 160 VNs. Starting from an initial placement, we apply a heuristic node swapping algorithm to minimize the longest Manhattan distance between nodes. The optimization yields a 7.04mm^2 667MHz chip design in the worst-case process corner at 0.85V and 125°C. Compared to a fully parallel binary LDPC decoder that is dominated by interconnect complexity [7], this fully parallel NB-LDPC decoder is placed and routed with an initial density of 80% for a final density of 87%.

The coding gain of the NB-LDPC decoder depends on the limit on decoding iterations, L. Increasing L from 10 to 30 improves the SNR by 0.15dB at a FER of 10^{-6}, but the energy efficiency worsens. Interestingly, we observe that the vast majority of the VNs converge very quickly, long before reaching the iteration limit. In addition, VNs and CNs are dominated by sequential circuits that consume significant clock switching power. Therefore we design each VN to detect its convergence and gate its clock to save power. The convergence detection is based on two criteria: meeting the minimum number of iterations M, and the VN having maintained the same decision for the last T consecutive iterations. The convergence detection is done at each VN level, seen in Fig. 24.2.3, permitting a fine-grained dynamic clock gating. Once a VN is clock gated, parts of the CNs are also clock gated by the enable signals propagated by the VN. A CN is fully clock gated once all its connected VNs are clock gated. The technique is highly effective over a wide range of SNR levels. For instance, setting L = 30, M = 10 and T = 10 at 3.6dB SNR allows VNs and CNs to be clock gated for 19.97 iterations out of 30 iterations, on average. Simulation confirms that the clock gating criteria assure excellent coding gain comparable to L = 100 at moderate-to-high SNR.

The test chip was fabricated in a STMicroelectronics 65nm CMOS process. The NB-LDPC decoder core occupies 7.04mm^2. In its periphery, AWGN generators produce the inputs to the decoder, and error collector measures the decoder FER and BER performance. Input sorters are added to initialize the log-likelihood vector for each VN. The test chip is fully functional with measured FER and BER performance graphed in Fig. 24.2.4. At room temperature and a 1V supply voltage, the chip operates at a maximum frequency of 700MHz for a throughput of 1.4Gb/s (with 10 iterations) and energy efficiency of 2.94nJ/b. Increasing the iteration limit L from 10 to 30 improves the coding gain by 0.15dB at a FER of 10^{-6}, but degrades the throughput to 467Mb/s and energy to 8.93nJ/b.

To achieve better energy efficiency, we enable fine-grained dynamic clock gating with L = 30, M = 10 and T = 10 to reduce the power consumption by 46% and cut the energy to 4.84nJ/b. To boost the throughput, we allow the decoder to terminate and move on to next input after all VNs are clock gated. Termination detection is based on monitoring clock gating enable signals, shown in Fig. 24.2.3. With L = 30, M = 10 and T = 10, enabling termination increases the throughput to 1.15Gb/s at 3.6dB SNR, and improves the energy to 3.37nJ/b. Hence, by combining fine-grained dynamic clock gating and termination, we achieve an excellent coding gain, a high throughput of 1.15Gb/s and energy efficiency of 3.37nJ/b (277pJ/b/iteration). Scaling the supply voltage from 1V to 675mV reduces the frequency from 700MHz to 400MHz for much improved energy efficiency of 1.10nJ/b (90pJ/b/iteration). Steps of the energy optimization are illustrated in Fig. 24.2.5. The throughput, energy efficiency and area efficiency of this chip are 24×, 3× and 18× better, respectively, than the latest NB-LDPC decoder design as shown in Fig. 24.2.6. The results also compare favorably to the binary LDPC decoder that provides the best coding gain [2]. The microphotograph of the NB-LDPC chip is shown in Fig. 24.2.7.

Acknowledgements:
This work was funded by NSF CCF-1054270 and the Broadcom Foundation. We acknowledge STMicroelectronics for chip fabrication and Marvell for testing facilities. We thank Pascal Urard and Engling Yeo for advice.

References:
[1] C. Studer, *et al.*, "A 390Mb/s 3.57mm^2 3GPP-LTE turbo decoder ASIC in 0.13µm CMOS," *ISSCC Dig. Tech. Papers*, pp. 274-275, 2010.
[2] P. Urard, *et al.*, "A 360mW 105Mb/s DVB-S2 Compliant Codec Based on 64800b LDPC and BCH Codes Enabling Satellite-Transmission Portable Devices," *ISSCC Dig. Tech. Papers*, pp. 310-311, 2008.
[3] M. Davey, D. MacKay, "Low-Density Parity Check Codes Over GF(q)," *IEEE Commun. Lett.*, vol. 2, no. 6, pp.165-167, Jun. 1998.
[4] X. Zhang, F. Cai, "Efficient Partial-Parallel Decoder Architecture for Quasi-Cyclic Nonbinary LDPC Codes," *IEEE Trans. Circuits and Systems-I*, vol. 58, no. 2, pp. 402-414, Feb. 2011.
[5] Y. Ueng, *et al.*, "An Efficient Layered Decoding Architecture for Nonbinary QC-LDPC Codes," *IEEE Trans. Circuits and Systems-I*, vol. 59, no. 2, pp. 385-398, Feb. 2012.
[6] Y. Tao, *et al.*, "High-Throughput Architecture and Implementation of Regular (2, d_c) Nonbinary LDPC Decoders," *IEEE Int. Symp. Circuits Syst.*, pp. 2625-2628, 2012.
[7] A. Blanksby, C. 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, 2002.
[8] A. Voicila, *et al.*, "Low-Complexity Decoding for Non-Binary LDPC Codes in High Order Fields," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1365-1375, 2010.

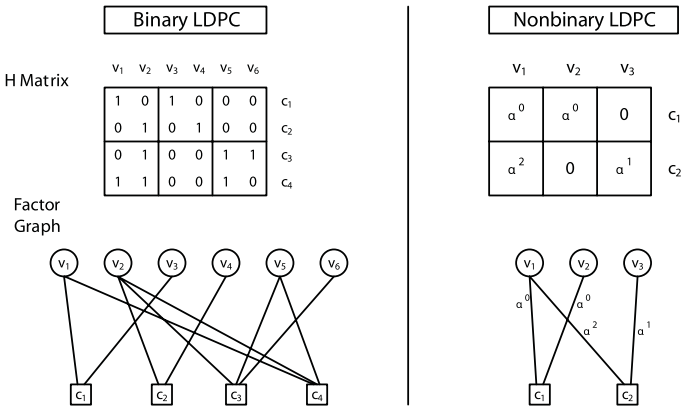

Figure 24.2.1: Comparison of binary and nonbinary LDPC (NB-LDPC) code.

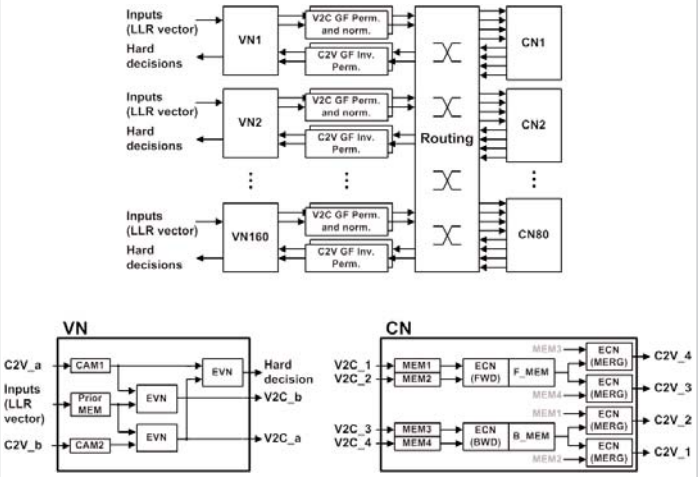

Figure 24.2.2: Architecture of the fully parallel NB-LDPC decoder.

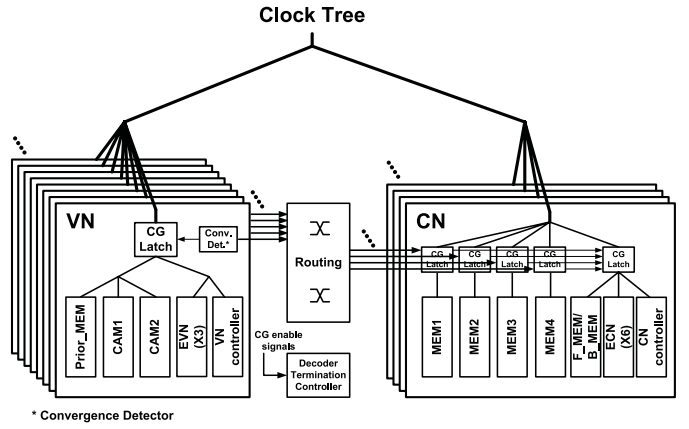

Figure 24.2.3: Illustration of clock tree and clock gating strategy.

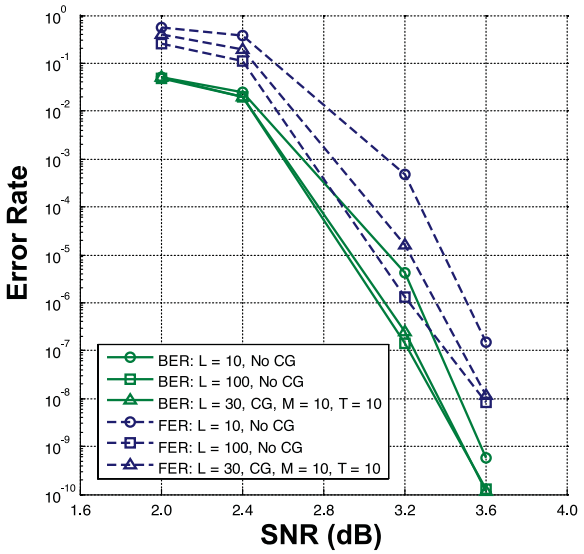

Figure 24.2.4: Measured BER and FER performance using 5b quantization.

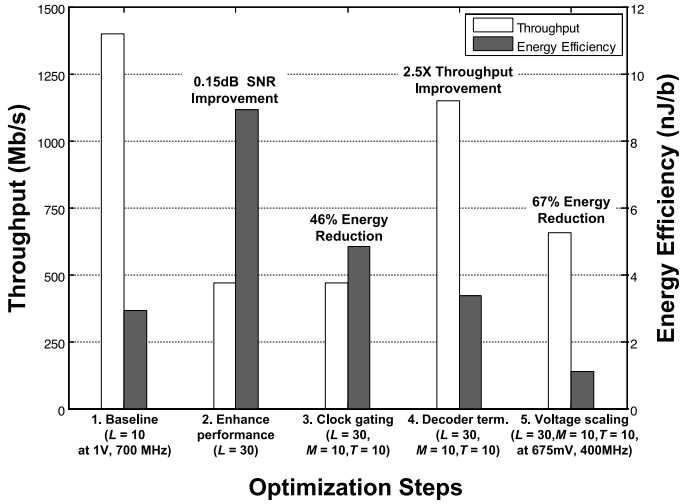

Figure 24.2.5: Throughput and energy optimization measured at 3.6dB SNR.

	Urard ISSCC09 [2]	Zhang TCAS11 [4]	Ueng TCAS12 [5]	This work	
Code	binary LDPC	NB-LDPC GF(32)	NB-LDPC GF(32)	NB-LDPC GF(64)	
Block Length	64800 variable rate	3720 rate 0.88	1240 rate 0.55	960 rate 0.5	
Design	silicon	FPGA	layout	silicon	
Process	65nm	-	90nm	65nm	
Core Area	6.07mm^2	-	10.33mm^2	7.04mm^2	
Utilization	-	-	-	87%	
Gate Count	-	1.13M (XOR)	1.92M (NAND)	2.78M (NAND)	
Supply	1.2V	-	-	1.0V	0.675V
Frequency	174MHz	106MHz	260MHz	700MHz	400MHz
Iterations	-	15	10	10 to 30**	10 to 30**
Throughput	135Mb/s	9.3Mb/s	47.69Mb/s	1.15Gb/s	656Mb/s
Power	360mW (max)	-	479.8mW	3.866W	726mW
Energy Efficiency	2.67nJ/b (max) -	- -	10.06nJ/b 1.006nJ/b/iter	3.37nJ/b 0.277nJ/b/iter	1.10nJ/b 0.090nJ/b/iter
Area Efficiency*	22.2Mb/s/mm^2	-	8.85Mb/s/mm^2	163Mb/s/mm^2	93Mb/s/mm^2

*Area normalized to 65nm for comparison
**Clock gating enabled with decode termination

Figure 24.2.6: Chip summary and comparison with state-of-the-art designs.

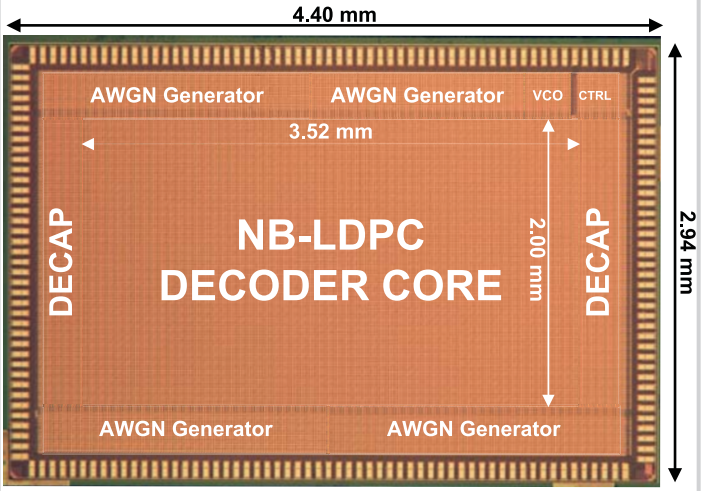

Figure 24.2.7: Micrograph of the NB-LDPC test chip 65nn CMOS.

978-1-4673-4516-3/13/$31.00