18.7 A 2.4mm² 130mW MMSE-Nonbinary-LDPC Iterative Detector-Decoder for 4×4 256-QAM MIMO in 65nm CMOS

Chia-Hsiang Chen, Wei Tang, Zhengya Zhang

University of Michigan, Ann Arbor, MI

The latest multiple-input multiple-output (MIMO) wireless systems have adopted iterative detection and decoding (IDD) to reduce the signal-to-noise ratio (SNR) required for a reliable transmission. An IDD system consists of a soft-in soft-out (SISO) detector to cancel interference, and a SISO forward-error correction (FEC) decoder to remove errors. The two blocks exchange soft information to improve the SNR iteratively. State-of-the-art IDDs based on sphere decoding (SD) and low-density parity-check (LDPC) FEC have been demonstrated in [1], [2] for up to 4×4 64-QAM MIMO, achieving 396Mb/s detection [1] and 586Mb/s decoding [2]. Compared to an SD detector [1]-[3], a minimum mean-square error (MMSE) detector [4] has a lower complexity and can be easily scaled to support a high-order modulation for a high data rate and spectral efficiency. Compared to LDPC FEC [1-3], nonbinary LDPC (NBLDPC) FEC offers better coding gain [5] and improves the detection-decoding performance [6]. Despite NBLDPC's higher complexity, efficient approximate decoding [5] is possible and it is well suited for a high-order modulation.

In this work, we demonstrate an MMSE-NBLDPC iterative detector-decoder for a 4×4 256-QAM MIMO system to achieve an excellent error rate that improves with iterations, as shown in Fig. 18.7.1. To minimize latency over the iterative loop and improve throughput, the MMSE detector is divided into 4 task-based coarse pipeline stages so that all stages can operate in parallel. Both the number of stages and the stage latency of the detector are minimized, and the long critical paths are interleaved and placed in a slow clock domain to support a high data rate in a cost-effective way. The resulting MMSE detector achieves an 82% higher throughput compared to [4], and almost 3.5× the throughput of the latest SD detector [1]. The NBLDPC decoder is implemented using 78 processing nodes to enable fully parallel message passing. Serial Galois field (GF) processing is pipelined using a data forwarding technique to cut the decoding latency by 30% over the latest design [5]. The detector and decoder exchange symbol log-likelihood ratios (LLR) that are efficiently computed based on the L_1 distance to the nearest neighbors in the QAM constellation. To lower the power consumption, automatic clock gating is applied to stage boundary and buffer registers to save 53% of the detector power and 61% of the decoder power. The results are demonstrated in a 65nm MMSE-NBLDPC iterative detector-decoder test chip that achieves 1.38Gb/s detection and 1.02Gb/s decoding (5 iterations), consuming 26.5mW and 103mW, respectively.

The MMSE detector design is comprised of 4 coarse pipeline stages as depicted in Fig. 18.7.2. Channel information and a-priori symbol LLRs from the decoder are pre-processed in the first stage to generate the MMSE matrix. The matrix is then inverted using LU decomposition (LUD) for MMSE filtering in the second and third stage, while interference cancellation is done in parallel. The final stage computes the SNR and symbol LLRs as the input to the NBLDPC decoder. The LUD in the second stage contains the critical paths and requires a long latency, causing unbalanced pipeline stages and a throughput bottleneck. As the Newton-Raphson reciprocal unit dictates the inner loop of the LUD, we reformulate the reciprocal computation in a parallel structure to shorten the second stage from 18 cycles [4] to 12 cycles. To loosen the timing constraint on the long critical paths in the second and third stage, we create a 2× slow clock domain for the two stages to allow the gates to be downsized, and recoup the throughput by interleaving between two copies of the datapaths without stalling the pipeline. After gate downsizing, the duplication costs only 24% additional area over the baseline, but the throughput is increased by 38%. In the final stage, we use an algorithmic property to simplify the SNR computation from 4 complex multiply and add [4] to 2 real multiply and 1 add of shorter bit widths, reducing the area of the final stage by 50% and power by 46%. The final MMSE detector in 65nm CMOS achieves a throughput of 1.38Gb/s.

An NBLDPC code offers superior coding gain even at moderate block length, and the coding performance improves with higher GF size. In this work, we use a (52, 26) regular-(2, 4) NBLDPC code over GF(256) with a binary block length of 416b. The NBLDPC decoder instantiates 52 variable nodes (VN) and 26 check nodes (CN), as shown in Fig. 18.7.3. VN and CN adopt the truncated extended min-sum (EMS) algorithm using the most reliable 12 GF elements in GF(256) processing to reduce complexity, while still maintaining a good coding performance. Due to the serial nature of GF processing, CN has to stall for VN to complete a GF vector and vice versa in a conventional design [5], resulting in a latency and throughput bottleneck. We apply a forwarding technique to provide GF element outputs directly from VN to CN to eliminate stalls, and reallocate the memory in VN to enable the forwarding from CN to VN in a similar manner. Forwarding shortens each decoding iteration from 36 cycles to 25, and it skips the CN-to-VN message storage. The final NBLDPC decoder in 65nm CMOS achieves a throughput of 1.02Gb/s (5 iterations) with a latency of 81.4ns per decoding iteration.

The MMSE detector and the NBLDPC decoder exchange symbol LLRs instead of bit LLRs. In an IDD system with a binary FEC, the detector generates the LLR by searching and computing the Euclidean distance to the nearest constellation points for each bit, as shown in Fig. 18.7.4. In comparison, nonbinary symbol LLR generation requires searching and computing the distance to the nearest constellation points once for each multibit symbol. As the LLR inputs to the NBLDPC decoder are normalized to the most likely constellation point, the LLR generation can be further reduced to calculating the L_1 distance to the nearest points along the real and imaginary axis. As Fig. 18.7.4 illustrates, symbol LLRs are computed using bit invert, shift and add in our design, and the nearest points are determined directly based on the soft detector output, eliminating costly search and multiply.

A total of 70.9kb registers are used for buffering data in and between stages of the detector and the decoder. Registers are used in place of memory arrays to support high access bandwidth and the flexibility of placing small memory blocks. Registers are power hungry, but we recognize a power reduction opportunity as most of the registers used in our design are infrequently updated due to the coarse pipelining, e.g., 1 update every 12 cycles for the 7.6kb stage boundary registers in the detector, and 1 update every 25 cycles for the 26.2kb CN buffer registers in the decoder. We exploit the access pattern to reduce power by enabling clock gating of the registers when they are idling, saving the detector power and the decoder power by 53% and 61%, respectively.

The fabricated test chip is fully functional. The MMSE detector core and the NBLDPC decoder core occupy 0.7mm² and 1.7mm², respectively. At room temperature and 1.0V supply, the MMSE detector runs at a maximum frequency of 517MHz for a throughput of 1.38Gb/s, the highest reported throughput of a SISO MMSE detector [4]. The MMSE detector consumes 19.2pJ/b, an order of magnitude lower than previous SISO detector designs [1], [2], [4], demonstrating the advantage of the optimized MMSE detection for IDD. At room temperature and 1.0V supply, the decoder runs at 307MHz for a throughput of 1.02Gb/s (5 iterations). The NBLDPC decoder consumes 20.1pJ/b/iteration, the lowest reported energy of an NBLDPC decoder [5], and it matches the efficiency of the binary LDPC decoder used in IDD [2]. Although our NBLDPC code is about half the size of [5], the order-of-magnitude improvement in energy [5] is significant. The energy efficiency can be further improved by voltage and frequency scaling, as shown in Fig. 18.7.5. At 500mV supply, the MMSE detector and the NBLDPC decoder consume 9.7pJ/b and 6.9pJ/b/iteration, respectively, for throughputs above 200Mb/s. Our work is compared with state-of-the-art MIMO detector and decoder designs in Fig. 18.7.6. The die photo is shown in Fig. 18.7.7.

Acknowledgements:
This work was funded in part by NSF CCF-1054270 and Intel. We thank Christoph Studer, Youn Sung Park, Yaoyu Tao, and Tai-Chuan Ou for advice.

References:
[1] B. Noethen, *et al.*, "A 105GOPS 36mm² Heterogeneous SDR MPSoC with Energy-Aware Dynamic Scheduling and Iterative Detection-Decoding for 4G in 65nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 188-189, Feb. 2014.
[2] F. Borlenghi, *et al.*, "A 2.78 mm² 65 nm CMOS Gigabit MIMO Iterative Detection and Decoding Receiver," *European Solid-State Circuits Conf.*, pp. 65-68, 2012.
[3] M. Winter, *et al.*, "A 335Mb/s 3.9mm² 65nm CMOS flexible MIMO Detection-Decoding Engine Achieving 4G Wireless Data Rates," *ISSCC Dig. Tech. Papers*, pp. 216-217, Feb. 2012.
[4] C. Studer, S. Fateh, D. Seethaler, "ASIC Implementation of Soft-Input Soft-Output MIMO Detection using MMSE Parallel Interference Cancellation," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1754-1765, 2011.
[5] Y. S. Park, Y. Tao, Z. Zhang, "A 1.15 Gb/s Fully Parallel Nonbinary LDPC Decoder with Fine-Grained Dynamic Clock Gating," *ISSCC Dig. Tech. Papers*, pp. 422-423, Feb. 2013.
[6] S. Pfletschinger, D. Declercq, "Getting Closer to MIMO Capacity with Non-Binary Codes and Spatial Multiplexing," *IEEE Globecom*, 2010.

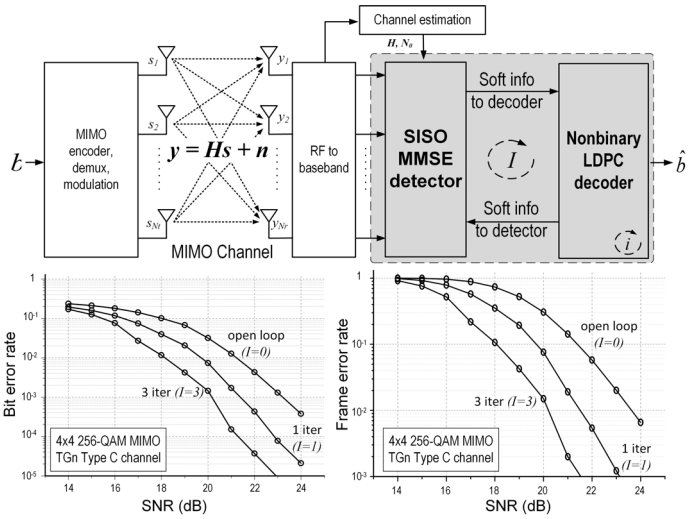


Figure 18.7.1: MMSE-NBLDPC IDD design for MIMO system.

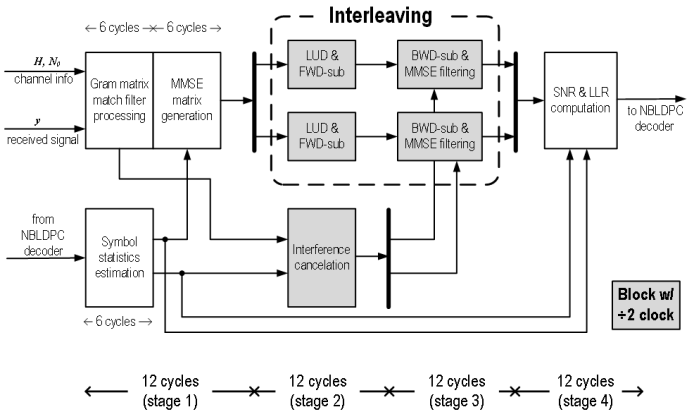


Figure 18.7.2: Block diagram of the MMSE detector.

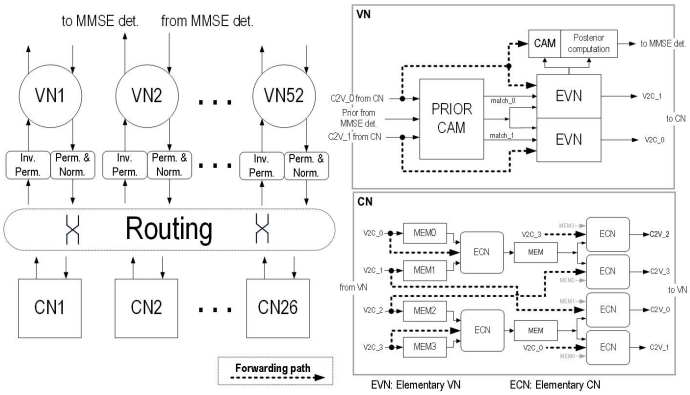
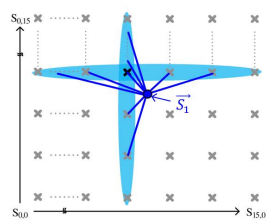
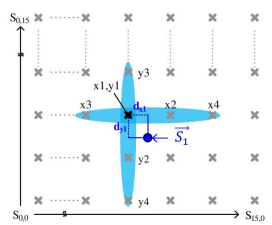


Figure 18.7.3: NBLDPC decoder with data forwarding.



bit	LLR
b0	$\lvert\overrightarrow{S_1}-S_{14,3}\rvert^2-\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2$
b1	$\lvert\overrightarrow{S_1}-S_{13,3}\rvert^2-\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2$
b2	$\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2-\lvert\overrightarrow{S_1}-S_{11,3}\rvert^2$
b3	$\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2-\lvert\overrightarrow{S_1}-S_{7,3}\rvert^2$
b4	$\lvert\overrightarrow{S_1}-S_{12,1}\rvert^2-\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2$
b5	$\lvert\overrightarrow{S_1}-S_{12,2}\rvert^2-\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2$
b6	$\lvert\overrightarrow{S_1}-S_{13,3}\rvert^2-\lvert\overrightarrow{S_1}-S_{12,4}\rvert^2$
b7	$\lvert\overrightarrow{S_1}-S_{12,8}\rvert^2-\lvert\overrightarrow{S_1}-S_{12,3}\rvert^2$

Bit LLR computation (before SNR scaling) for the soft detector output $\overrightarrow{S_1}$ in 256-QAM.



	LLR
x1	0
x2	$(1-d_{x1})\times 4$
x3	$(1+d_{x1})\times 4$
x4	$(2-d_{x1})\times 8$
y1	0
y2	$(1-d_{y1})\times 4$
y3	$(1+d_{y1})\times 4$
y4	$(2-d_{y1})\times 8$

Symbol LLR computation (before SNR scaling) for the soft detector output $\overrightarrow{S_1}$ in 256-QAM. (Note that the x and y entries are cross added to obtain the symbol LLRs.)

Figure 18.7.4: Bit LLR and symbol LLR computation.

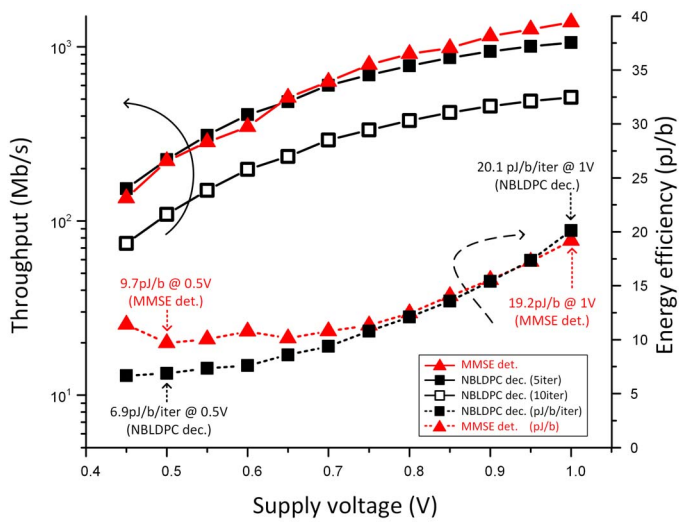


Figure 18.7.5: Chip measurement results.

Detector	Noethen [1]	Borlenghi [2]	Winter [3]	Studer [4]	This work
IDD design	yes	yes	no	yes	yes
Algorithm	SD SISO	SD SISO	SD SO	MMSE SISO	MMSE NB-SISO
MIMO system	≤ 4x4	≤ 4x4	≤ 4x4	4x4	4x4
Modulation	≤ 64	≤ 64	≤ 64	≤ 64	256
Technology [nm]	65	65	65	90	65
Core area [mm²]	-	2.78	0.31	1.5	0.7
Preprocessing area [kGE]	383[a]	[b]	[b]	410	347[c]
Detection area [kGE]		872	215		
Frequency [MHz]	445	135	333	568	517
Power [mW]	87	-	38	189	26.5
Throughput [Mb/s]	396	194	296-807	757	1379
Area efficiency [Mb/s/kGE]	1.03	0.22	1.37-3.75	1.85	3.68
Energy efficiency [pJ/b]	220	920	48	250	19.2
Decoder	**Noethen [1]**	**Borlenghi [2]**	**Winter [3]**	**Park [5]**	**This work**
IDD design	yes	yes	no	no	yes
Code	LDPC	LDPC	LDPC	NBLDPC GF(64)	NBLDPC GF(256)
Block length	768	1944	768	960	416
Technology [nm]	65	65	65	65	65
Core area [mm²]	-	0.78	3.6	7.04	1.7
Decoding area [kGE]	-	-	-	2780	935
Frequency [MHz]	500	299	267	700	307
Power [mW]	-	-	367	3866	103
Iterations	10	10	10	10 to 30[d]	5 10
Throughput [Mb/s]	155	586	235.2	1150	1024 512
Area efficiency [Mb/s/mm²]	100.92	751	65.33	163	602 301
Energy efficiency [pJ/b/iter]	232	21	170	277	20.1

[a]: memory for data exchange included
[b]: data pre-processing block (QRD) not included
[c]: total area is 264 kGE if no interleaving processing
[d]: with early termination.

Figure 18.7.6: Comparison with state-of-the-art designs.

18

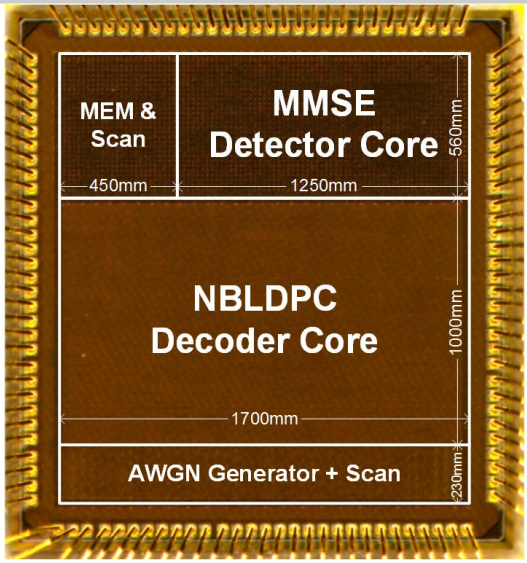


Figure 18.7.7: Chip microphotograph.

978-1-4799-6224-2/15/$31.00