

Refresh-Free Dynamic Standard-Cell Based Memories: Application to a QC-LDPC Decoder

Pascal Meinerzhagen^{*§}, Andrea Bonetti^{*}, Georgios Karakonstantis^{*}, Christoph Roth[†],
Frank Gürkaynak[†], and Andreas Burg^{*}

^{*}Telecommunications Circuits Laboratory, EPFL, Lausanne, VD, 1015 Switzerland; [§]Now with Intel Labs, Intel Corporation;
E-Mail: pascal.a.meinerzhagen@intel.com, [andrea.bonetti, georgios.karakonstantis, andreas.burg]@epfl.ch

[†]Integrated Systems Laboratory, ETHZ, Zurich, ZH, 8092 Switzerland; E-Mail: rothc@iis.ee.ethz.ch, kgf@ee.ethz.ch

Abstract—The area and power consumption of low-density parity check (LDPC) decoders are typically dominated by embedded memories. To alleviate such high memory costs, this paper exploits the fact that all internal memories of a LDPC decoder are frequently updated with new data. These unique memory access statistics are taken advantage of by replacing all static standard-cell based memories (SCMs) of a prior-art LDPC decoder implementation by dynamic SCMs (D-SCMs), which are designed to retain data just long enough to guarantee reliable operation. The use of D-SCMs leads to a 44% reduction in silicon area of the LDPC decoder compared to the use of static SCMs. The low-power LDPC decoder architecture with refresh-free D-SCMs was implemented in a 90 nm CMOS process, and silicon measurements show full functionality and an information bit throughput of up to 600 Mbps (as required by the IEEE 802.11n standard).

I. INTRODUCTION

Embedded memories are a crucial building block of most embedded digital signal processing (DSP) systems and often account for a dominant area and power share. In a large variety of such embedded DSP systems, the memories need to hold data only for a short time before they are written again with new data. In such application scenarios with periodic and frequent write updates, the conventionally used SRAM macrocells or standard-cell based memories (SCMs) [1] can be replaced with embedded dynamic memories to improve area efficiency. As a representative example of embedded DSP systems containing frequently updated memories, this paper focuses on low-density parity check (LDPC) decoders used extensively in wireless communications, and shows the benefits of integrating refresh-free, dynamic SCMs (D-SCMs).

A good example for the high memory share in LDPC decoders is [2], where the embedded memories consume 68% and 58% of the overall power and area, respectively. To improve the overall area and energy efficiency of LDPC decoders, many recent works have focused on alternative implementations of the embedded memories, in addition to the persisting efforts to limit the memory footprint and bandwidth requirements. In particular, replacing conventional SRAM macrocells with SCMs was shown to entail a considerable 40% power reduction due to the ability to merge the memories with logic, better data locality, less routing, and consequently lower active power consumption [1], [3]. This active energy advantage arising from SCMs, as opposed to SRAM hard-macros, is further demonstrated by an extensive, comparative analysis of various LDPC decoder architectures [4]. However, the energy savings provided by SCMs come at the cost of an increased silicon area, which can become 50% larger compared to SRAM hardmacros [1]. As a further alternative to SRAM macrocells, a recent work [5] proposes to use gain-cell based eDRAMs in a high-throughput LDPC decoder. These eDRAM macrocells can be operated without refresh operation

due to frequent write access and lead to an overall better area and energy efficiency.

In this paper, we propose to combine the aforementioned advantages of SCMs, especially the high data locality and the low switching power, with the high storage density of dynamic bitcells. This approach reduces the area penalty of previous LDPC decoders using static SCMs [6], and can be safely adopted, even without the need for explicit refresh cycles. Refresh-free operation is possible as the write and read access statistics of all internal memories of the considered LDPC decoder are known *a priori* to exhibit frequent write updates. Our contributions can be summarized as follows: 1) for the first time, we propose refresh-free D-SCMs, which unite the assets of both SCMs and dynamic bitcells as embedded storage arrays for dedicated and highly optimized DSP sub-systems containing frequently updated memories, with LDPC decoders serving as a representative example; 2) the D-SCM architecture and bitcell design contains many novelties, including a) selective insertion of high-threshold (high- V_T) gates for increased noise margins at negligible speed penalty; b) design of custom-designed, multifunctional standard-cell combining a 3-transistor (3T) dynamic latch and NAND functionality; c) circuit optimizations to avoid short-circuit currents during all operating modes; and d) capacitive coupling from the read word-line (RWL) onto the dynamic storage node (SN) for increased read speed and robustness; as well as 3) proof of superior area and energy efficiency enabled by D-SCMs through their integration into a LDPC decoder; this circuit is manufactured in 90 nm CMOS, and compared with a reference design in the same node.

II. QC-LDPC DECODER ARCHITECTURE

LDPC codes and in particular quasi-cyclic (QC)-LDPC codes [7] are among the most popular and capable error-correcting codes adopted in many modern standards including DVB-S2 [8] and IEEE 802.11n [9]. The decoding of a QC-LDPC code is in general performed by iterative message passing between variable nodes which represent the code bits and check nodes which represent the parity check equations of the code-specific parity-check matrix \mathbf{H} . The messages going from variable nodes to check nodes are denoted as Q-messages and the messages exchanged in the other direction as R-messages. In addition, an L-value is associated with each variable node representing the reliability information for the corresponding code bit in the form of an estimate of the a-posteriori log-likelihood ratio (LLR). In this work, we use an LDPC decoder based on the offset-min-sum (OMS) message-update rules combined with the layered decoding schedule in order to profit from a good balance between convergence speed and VLSI implementation complexity [7], [4].

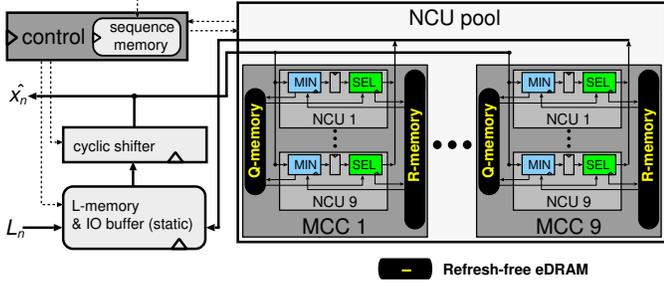


Fig. 1. Architecture of the quasi-cyclic LDPC decoder with refresh-free dynamic memories (highlighted in yellow).

A. Architecture Details

The considered LDPC decoder architecture [3] is shown in Fig. 1. The decoder starts by initializing the L-memory with the initial LLRs of the code bits obtained from the baseband receiver and continues with the sequential processing of the loaded parity-check matrix. To this end, Z node computation units (NCUs) sequentially execute the OMS algorithm for each layer of \mathbf{H} , where Z denotes the number of parity-check equations per layer. Each NCU follows a two-step procedure. In the first step, the MIN unit iteratively computes all Q-messages and other intermediate data of the current layer using the corresponding R-messages and L-values from the previous iteration. During this process, the cyclic shifter shifts the Z successive L-values fetched from the L-memory according to the quasi-cyclic property of \mathbf{H} in order to feed all MIN units with the proper values. In the second step, the SEL unit iteratively updates the R-messages and L-values based on the old R-messages and on the buffered Q-messages and intermediate data provided by the MIN unit. This process is repeated for all layers of \mathbf{H} and until a predefined number of iterations has been reached or an online stopping-criterion has been triggered. As shown in Fig. 1, several NCUs are grouped together with the corresponding Q-memory and R-memory sub-blocks in order to maximize data locality [3].

B. Memory Requirements and Characteristics

Interestingly, we observe that the Q- and R-messages as well as the L-values need to be stored only for a short time, before the corresponding memories are updated again with new data, which allows us to use refresh-free dynamic storage elements. Three types of memories are required in the considered decoder architecture: the R, Q, and L memories store the R-messages, Q-messages, and L-values, respectively. The total size requirement of each memory type for the highest-rate parity-check matrix with $Z = 81$ specified for the IEEE 802.11n standard is shown in Tbl. I. In order to enable refresh-free operation the memories are characterized according to the following definitions: 1) The *retention time* t_{ret} denotes the time interval between the first write access and the last corresponding read access to a memory block. 2) Moreover, the *update rate* t_{up} is defined as the time interval between a write access to a word and the next write access to the same word. Note that at a time t_{ret} after writing, all addresses must still read out correctly while up to a time t_{up} after writing, the data levels in the dynamic storage cell should still be strong enough to avoid short-circuit currents (unless they can be avoided by circuit techniques). Tbl. I shows the retention time (t_{ret}) requirements as well as the effective,

TABLE I
MEMORY SIZES, RETENTION TIMES, AND UPDATE RATES.

	R Memory	L Memory	Q Memory
Size [bits]	35640	9720	9720
t_{ret} [ns, clock cycles]	287.8, 88	287.8, 88	78.5, 24
t_{up} [ns, clock cycles]	287.8, 88	287.8, 88	287.8, 88

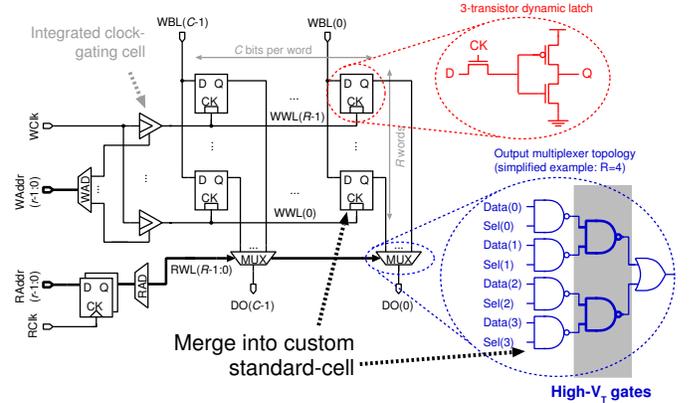


Fig. 2. Architecture of dynamic, standard-cell based memory.

guaranteed update rates (t_{up}) of all memory types contained in the QC-LDPC decoder architecture. The table assumes an operating frequency of 305.8 MHz, necessary for an information bit throughput of 600 Mbps (at 10 iterations), required by the highest-rate mode of the IEEE 802.11n standard.

III. DYNAMIC STANDARD-CELL BASED MEMORY DESIGN

Among many different SCM architectures presented and compared in [1], an architecture (shown in Fig. 2) based on latches as basic storage cells, integrated clock-gating cells for the generation of write select pulses, and static CMOS multiplexers for the readout of the selected word is most suitable in terms of area efficiency, power consumption, and speed. Due to the frequent write updates, a custom-designed dynamic latch is proposed as basic storage cell rather than a commercially available static latch. In order to aggressively push for minimum area, a 3-transistor (3T) dynamic latch topology is adopted as starting point, as shown in Fig. 2. To further improve the area efficiency, the 3T latch is merged with the first stage of the read multiplexer, namely a NAND gate, into a single, custom standard-cell. The conceptual schematic of this standard-cell is shown in Fig. 3(a).

As a protection mechanism against excessive leakage in case of potentially weak output levels of the dynamic storage cell, the second stage of the readout multiplexer (i.e., all logic gates directly following the basic storage cell with NAND functionality) is implemented with high threshold-voltage (high- V_T) gates, as shown in Fig. 2. Since only one cell in a long combinatorial path is replaced with a high- V_T cell the impact on the speed is negligible.

A. Bitcell Optimization to Avoid Short-Circuit Currents

The initial cell shown in Fig. 3(a) uses a single NMOS transistor to transfer a logic level from the write bitline (WBL) to the storage node (SN) as soon as a write operation is initiated by rising the write wordline (WWL). While logic “0” levels are properly transferred to the SN, logic “1” levels are degraded by the threshold voltage drop across the NMOS write transistor (MW), as we do not use a WWL overdrive voltage for straightforward integration of this cell into a design

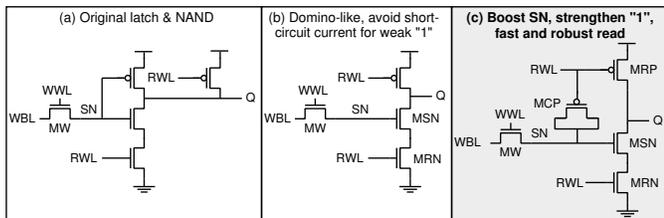


Fig. 3. Design exploration of custom standard-cells combining dynamic latch and NAND functionality.

with a single core supply voltage. Charge injection and clock feedthrough further deteriorate the logic “1” level during de-assertion of the WWL. These deteriorated logic “1” levels bare the risk for short-circuit currents during readout of the cell, i.e., as soon as the read wordline (RWL) is asserted and goes high. To avoid such excessive short-circuit currents which would last for an entire clock cycle, the PMOS transistor connected to SN is removed, as shown in Fig. 3(b). This results in a cell that operates similarly to domino logic: prior to a read access, the output node Q is precharged to V_{DD} , as at that time the RWL is still de-asserted and low. During the read access, the RWL is high, and the output node Q is safely discharged even with a deteriorated, weak logic “1” level on the storage node, while the node Q remains in its pre-charged state if SN holds a logic “0”. In addition to avoiding short-circuit current during read, there is no risk for short-circuit currents during non-read cycles (including potential standby times of the LDPC decoder) either. In fact, the output node Q of the domino-like dynamic bitcell is always properly charged to V_{DD} during non-read cycles, which circumvents short-circuit currents in its output stage and in subsequent logic gates. This property distinguishes the presented cell from conventional, dynamic memory and logic cells.

B. Increasing Read Robustness

Transistor MSN in Fig. 3(b) suffers from the body effect: its positive source-to-body voltage V_{SB} increases its threshold voltage V_T , which aggravates the readout process of an already deteriorated logic “1”. Similarly to a common practice in gain-cell eDRAM design [10], adding a coupling capacitor in form of a MOS capacitor (MCP) between the SN and the RWL, as shown in Fig. 3(c), was found to considerably improve the read “1” robustness of our bitcell, as well. The positive RWL transition during the onset of a read operation couples onto the SN and temporarily rises the SN level, thereby strengthening the logic “1” and leading to a faster read operation.

The layout of the basic storage cell with NAND functionality is shown on the left-hand side of Fig. 4. Compared with a minimum-drive, minimum-size, static latch and NAND gate from a commercial standard cell library, the silicon area of the proposed, custom-designed, multifunctional standard-cell is reduced by 70%.

IV. SILICON MEASUREMENT RESULTS

The above-described QC-LDPC decoder was manufactured in a 90nm CMOS technology. A chip microphotograph and a complete layout picture of the decoder core, surrounded by a pad-frame are shown in the middle of Fig. 4. A total of 8 packaged dies were verified on a HP93000 digital tester; all measured dies were fully functional within the expected voltage and frequency range.

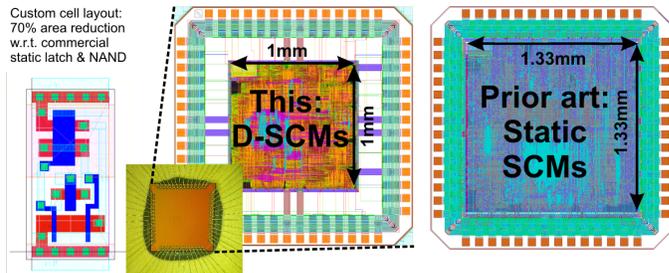


Fig. 4. (Left) Layout of custom standard-cell; (Middle) Chip microphotograph and layout picture of the proposed LDPC decoder; and (Right) Layout picture of the same LDPC decoder architecture using static SCMs.

A. Frequency and Voltage Characterization

Fig. 5 shows the percentage of failing chips as a function of the frequency and the supply voltage V_{DD} . As expected, there is a maximum and a minimum operating frequency, together defining a frequency range for valid circuit operation. The maximum frequency is determined by the critical path delay and decreases with the supply voltage. There is a sharp transition from 0 to 100% failing chips, which means that die-to-die variations between the 8 measured dies (all from the same wafer) do not significantly affect the critical path delay. The need for a minimum operating frequency (below which the chips fails) arises from the dynamic memories, which are designed to retain data only for the minimum required time of 287.8 ns. We observe a rather slow transition from 0 to 100% failing chips when gradually slowing down the clock frequency for a given V_{DD} . Compared to a few critical timing paths whose delays are determined by the transistor’s on-current (I_{on}) that varies only slightly from die to die, the minimum retention time of the dynamic storage cells is determined by several leakage mechanisms and is much more sensitive to parametric variations. This behavior is well aligned with previous reports on gain-cell based eDRAMs whose retention time is very sensitive even to within-die parametric variations [11]. Supply voltage scaling has two complementary effects on the retention time of the considered dynamic bitcell: 1) weakened leakage currents (e.g., the subthreshold conduction of MW decreases with V_{DS} , which in turn decreases with V_{DD}); and 2) lower noise margins (i.e., less headroom for deterioration of logic storage levels due to leakage). According to the measurements shown in Fig. 5, the weaker leakage currents at lower V_{DD} are the dominant effect, allowing longer retention times and lower frequencies at lower V_{DD} . The same behavior, i.e., improved retention times at scaled voltages, has also previously been observed in logic-compatible, gain-cell based eDRAMs [12]. For all voltages between 0.8 and 1.2 V, there is a large range of frequencies where all measured LDPC decoder chips function correctly. Within these admissible voltage and frequency ranges, the decoder supports different throughput modes, as exemplified by the markers in Fig. 5.

B. Comparison with Prior-Art Implementations

The 70% area reduction of the multifunctional, dynamic standard-cell results in a considerable 44.4% reduction in the area cost of the LDPC decoder, compared to its implementation with static SCMs. In fact, the core size of the proposed decoder is only 1.00 mm^2 , while it is 1.77 mm^2 with static SCMs, as shown on the right-hand side of Fig. 4.

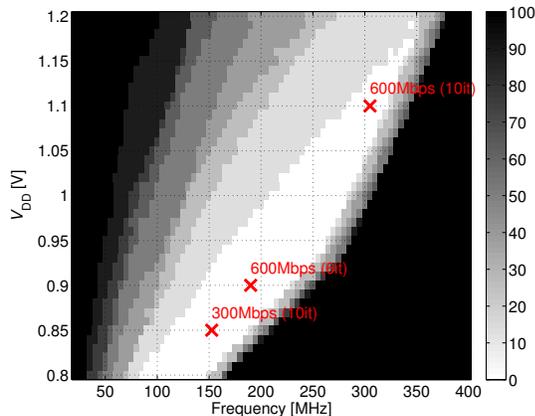


Fig. 5. Percentage of failing chips as a function of the frequency and V_{DD} .

The leakage current of the presented decoder architecture is dominated by the leakage current of the embedded memories. Replacing the static SCMs with the proposed D-SCMs (which have a built-in mechanism to avoid short-circuit currents) results in an average decoder's leakage current reduction of 55% compared to the decoder using static SCMs. However, as the leakage current is small compared to the switching current, and as all parts other than the basic storage cell remain unchanged, the proposed decoder implementation exhibits only a small total power reduction of 5.5% on average (among all measured dies) compared to the same decoder architecture using static SCMs. The corresponding, average decoding energy is 14.7 pJ/bit/iteration (measured at 1.0 V, 305.8 MHz, for 10 iterations, computed over the coded throughput, 600 Mbps information bit throughput, averaged over 8 dies) in case of D-SCMs and 15.5 pJ/bit/iteration in case of static SCMs. Finally, as shown in Tbl. II, the proposed LDPC decoder is compared with a selection of the best—in terms of hardware efficiency A [mm²/Gbps] and energy efficiency E [pJ/bit/iter]—, recent, silicon-proven LDPC decoders for the IEEE 802.11n or the WiMAX standards. All metrics are scaled to the 90 nm CMOS node and a voltage of 1.1 V and reported in parenthesis, in addition to the original values. The proposed decoder compares favorably with prior art by achieving both good hardware and energy efficiency. Only one work [13] (using SRAM macrocells) has slightly better hardware efficiency, at the cost of worse energy efficiency, while only one work [14] has slightly better energy efficiency, at the cost of worse area efficiency. Compared to [15], which is also based on SRAM macrocells, the proposed D-SCM based decoder implementation has both better hardware and better energy efficiency.

According to [4], LDPC decoders using SCMs achieve higher decoding throughput than decoders using SRAM macrocells. Moreover, the use of dynamic logic further promotes high operating frequencies compared to CMOS logic. Therefore, the proposed LDPC decoder based on D-SCMs is a favorable choice to achieve the high throughput requirements of future wireless communications systems.

V. CONCLUSIONS

Due to frequent and periodic write updates, all embedded memories of a low-power QC-LDPC decoder can be implemented using area-efficient dynamic storage cells. The newly proposed and seamlessly integrated dynamic, standard-

TABLE II
COMPARISON WITH PRIOR-ART LDPC DECODER IMPLEMENTATIONS

Publications	[3]	[15]	[13]	[14]	This
Technology [nm]	90	180	130	65	90
V_{DD} [V]	1.0	1.8	1.2	1.2	1.0
Core area [mm ²]	1.77	14.3	3.03	3.36	1.0
Maximum throughput T [Mbps]	679	640 (1280)	728 (1052)	880 (636)	600
Hardware efficiency A [mm ² /Gbps]	2.6	22.3 (2.8)	4.16 (1.4)	3.8 (10.1)	1.7
Energy efficiency E [pJ/bit/iter]	15.8 (19.1)	123 (23.0)	39 (22.7)	10.9 (12.7)	14.7 (17.8)

Scaling to 90 nm, 1.1 V: $T \sim s$, $A \sim 1/s^2$, $E \sim 1/s \cdot (1.1 V/V_{DD})^2$

cell based memories lead to a silicon area and leakage current reduction of 44.4% and 55.0%, respectively. The proposed multifunctional, dynamic storage cell avoids short-circuit currents by changing the read logic from CMOS to domino style and is optimized for robust read by inserting a coupling capacitor between the storage node and the read wordline. A potential drawback of the proposed decoder is the need of a minimum operating frequency, below which the refresh-free dynamic storage elements start to lose their data. However, all measured dies have a large range of safe operating frequencies compatible with various throughput modes. The silicon-proven LDPC decoder exhibits a core area of 1.0 mm² in a 90 nm CMOS node, dissipates 14.7 pJ/bit/iteration, and runs at all frequencies from 85 to 345 MHz for a voltage range from 0.8 to 1.2 V.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (project # PP002-119057), a Marie Curie grant (# 304186), and an Intel Ph.D fellowship (P. Meinerzhagen).

REFERENCES

- [1] P. Meinerzhagen *et al.*, "Towards generic low-power area-efficient standard cell based memory architectures," in *IEEE MWSCAS*, 2010.
- [2] C. Studer *et al.*, "Configurable high-throughput decoder architecture for quasi-cyclic LDPC codes," in *IEEE Asilomar Conference*, 2008.
- [3] C. Roth *et al.*, "A 15.8 pJ/bit/iter quasi-cyclic LDPC decoder for IEEE 802.11n in 90nm CMOS," in *IEEE A-SSCC*, 2010.
- [4] —, "Area, throughput, and energy-efficiency trade-offs in the VLSI implementation of LDPC decoders," in *IEEE ISCAS*, 2011.
- [5] Y. S. Park *et al.*, "Low-power high-throughput LDPC decoder using non-refresh embedded DRAM," *IEEE JSSC*, 2014.
- [6] Q. Xie *et al.*, "A high parallel macro block level layered LDPC decoding architecture based on dedicated matrix reordering," in *IEEE SiPS*, 2011.
- [7] K. K. Gunnam *et al.*, "Multi-rate layered decoder architecture for block LDPC codes of the IEEE 802.11n wireless standard," in *IEEE ISCAS*, 2007.
- [8] *Digital Video Broadcasting (DVB) User guidelines for the second generation system for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)*, Feb. 2005.
- [9] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: Enhancements for higher throughput," *IEEE P802.11n/D5.02, Part 11*, July 2008.
- [10] P. Meinerzhagen *et al.*, "Design and failure analysis of logic-compatible multilevel gain-cell-based DRAM for fault-tolerant VLSI systems," in *IEEE GLSVLSI*, 2011.
- [11] K. C. Chun *et al.*, "A 2T1C embedded DRAM macro with no boosted supplies featuring a 7T SRAM based repair and a cell storage monitor," *IEEE JSSC*, 2012.
- [12] R. Iqbal *et al.*, "Two-port low-power gain-cell storage array: Voltage scaling and retention time," in *IEEE ISCAS*, 2012.
- [13] B. Xiang *et al.*, "An 847-955 Mb/s 342-397 mw dual-path fully-overlapped QC-LDPC decoder for WiMAX system in 0.13um CMOS," *IEEE JSSC*, 2011.
- [14] X. Peng *et al.*, "A 115mW 1Gbps QC-LDPC decoder ASIC for WiMAX in 65nm CMOS," in *IEEE A-SSCC*, 2011.
- [15] M. M. Mansour and N. R. Shanbhag, "A 640-Mb/s 2048-bit programmable LDPC decoder chip," *IEEE JSSC*, 2006.