Flash Memory Summit

Codelucida

Ultra-Low Resource FPGA implementation of Finite Alphabet Iterative Decoders

D. Declercq

B. Reynwar, V. Yella, B. Vasic and S. Planjery

Contact: declercq@codelucida.com

Website: www.codelucida.com

Low Resource vs. High Throughput

- Storage using 3D TLC / QLC flash requires **LDPC ECC** for the improved performance against BCH
- LDPC is especially important to improve the **endurance** and the robustness to **retention**
- Throughput requirements are only increasing: PCIe-gen3 **(4GB/s)** and PCIe-gen4 **(8GB/s)**

However

- Strong Error Correction requires **more hardware resource**
- Larger throughputs requires **more hardware resource**

LDPC solutions are usually
too big to fit in FPGA based controllers

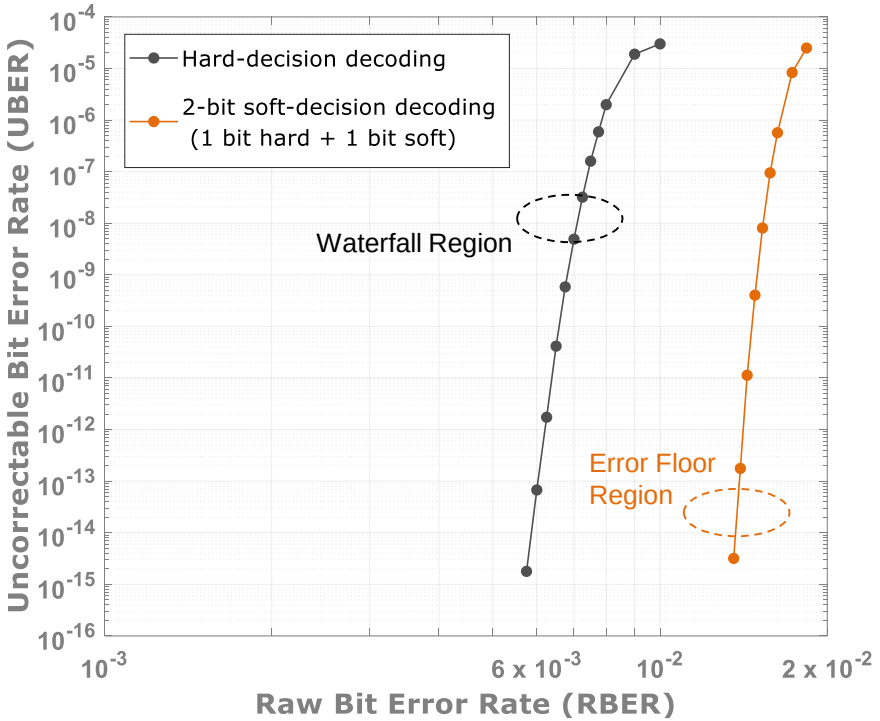
Vertically layered FAID decoding

FAID™ : Finite Alphabet Iterative Decoding

- Iterative decoders with 3 bits messages belonging to { -3 , -2 , -1 , 0 , +1 , +2 , +3 }
  - ✓ Low computational complexity
  - ✓ Low memory requirements
- Vertical column-wise scheduling for low memory and fast processing
- **Goal for this talk**: Propose specific FPGA implementation of FAID with extremely low-resource usage and high throughput

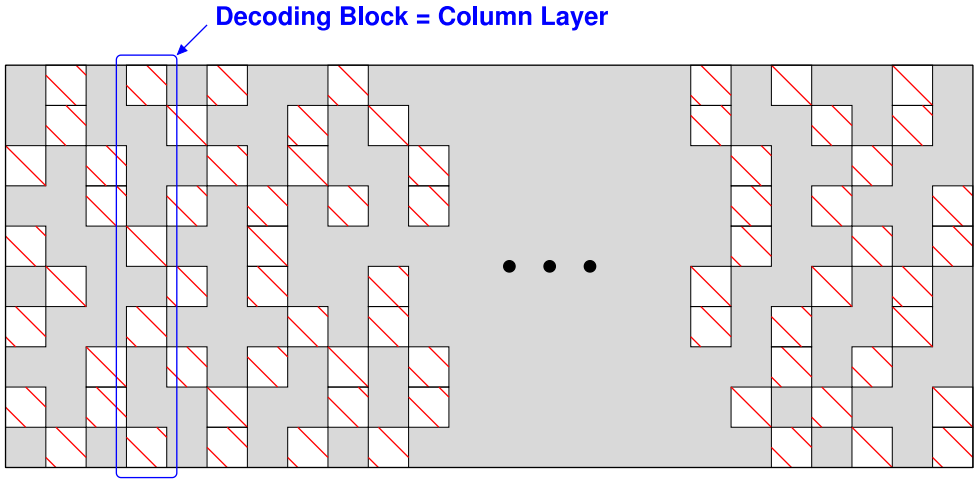
Why FAID have good ECC performance ?

- Low precision iterative decoders tend to **limit** the ECC performance, both in the waterfall and Error floor regions
- **Classical solution**: increase message precision, but with extra HW resource
- **FAID approach:** low precision, but **optimize** the non-linear VN update rules to recover the performance loss

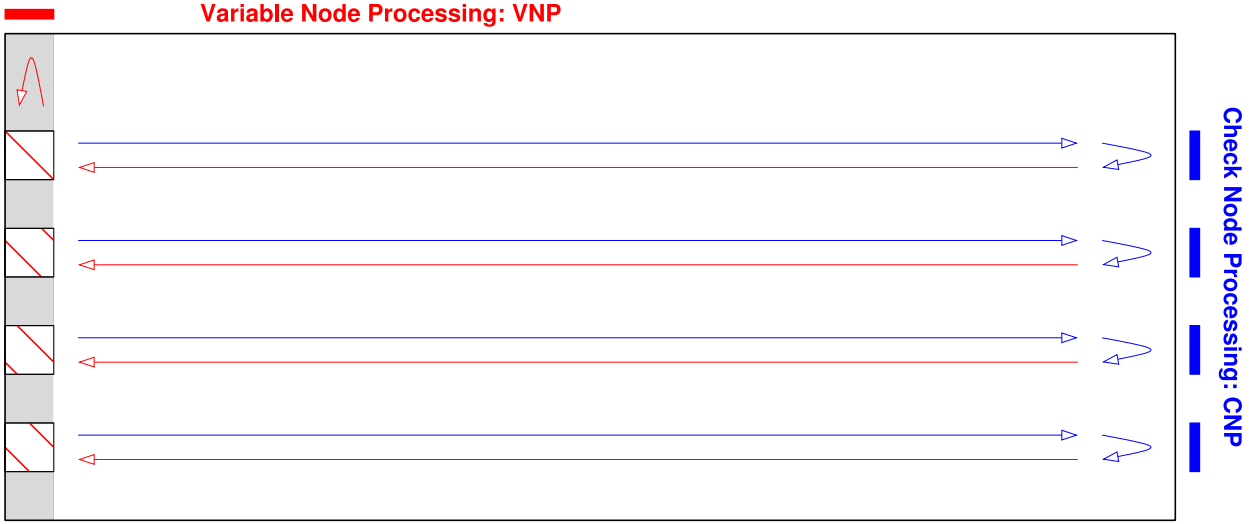


Vertical Processing

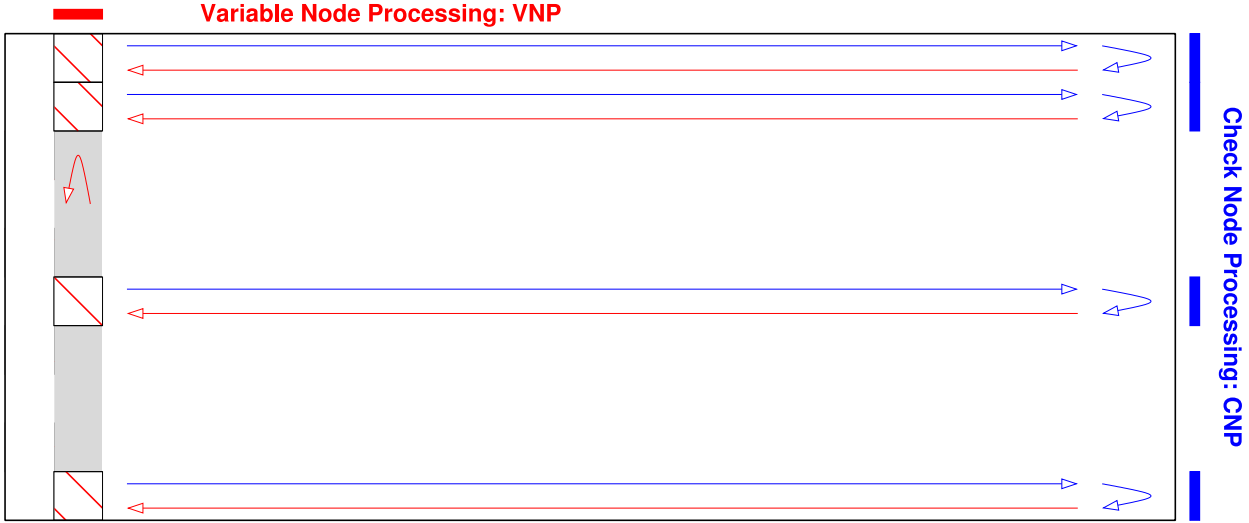
- **Quasi-cyclic LDPC codes** - parity-check matrix defined by circulant blocks
- Process the dv=4 circulants blocks of each column **in parallel**
- Add **constraints in the LDPC code** such that no 2 blocks overlap in consecutive columns



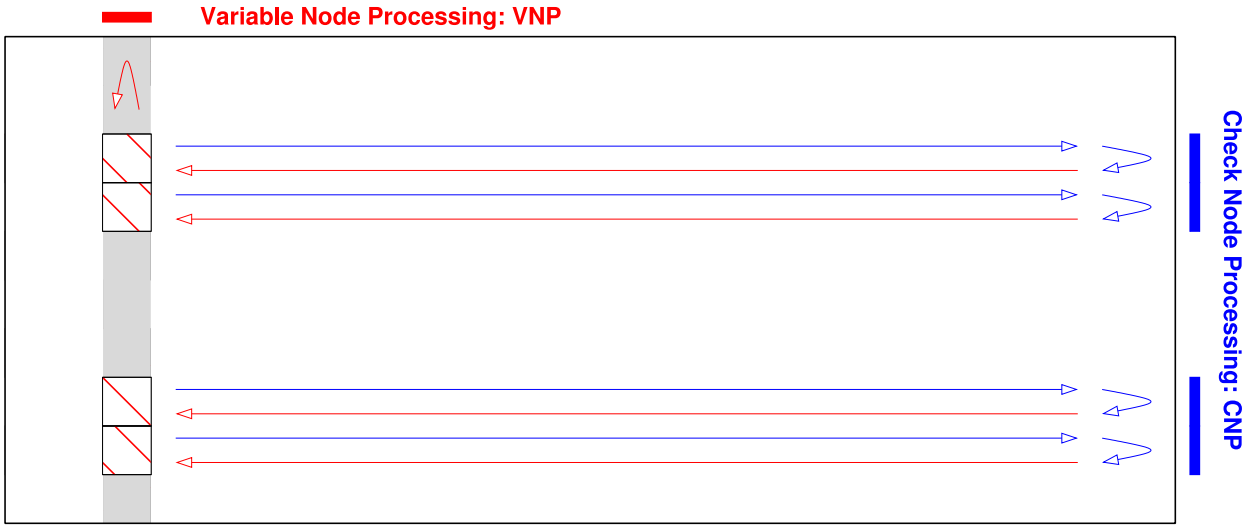
Vertical Processing: column #1



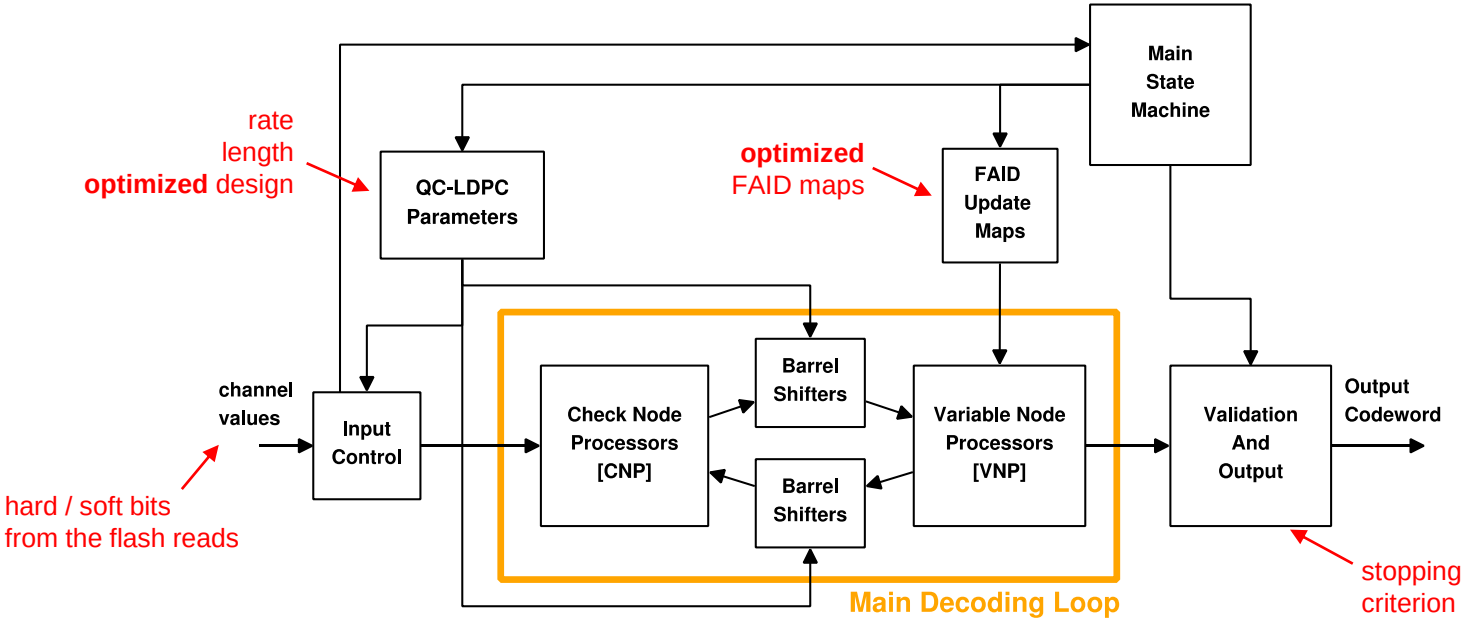
Vertical Processing: column #2



Vertical Processing: column #3



Top Level Architecture



Specific FPGA Optimization

- Implementation of the **FAID VNP** as Look-up-Tables rather than Boolean circuit

  ⟹ reduce the **Memory usage** and the **LUT usage**

- Optimization of the **pipelined architecture** to maximize chip frequency

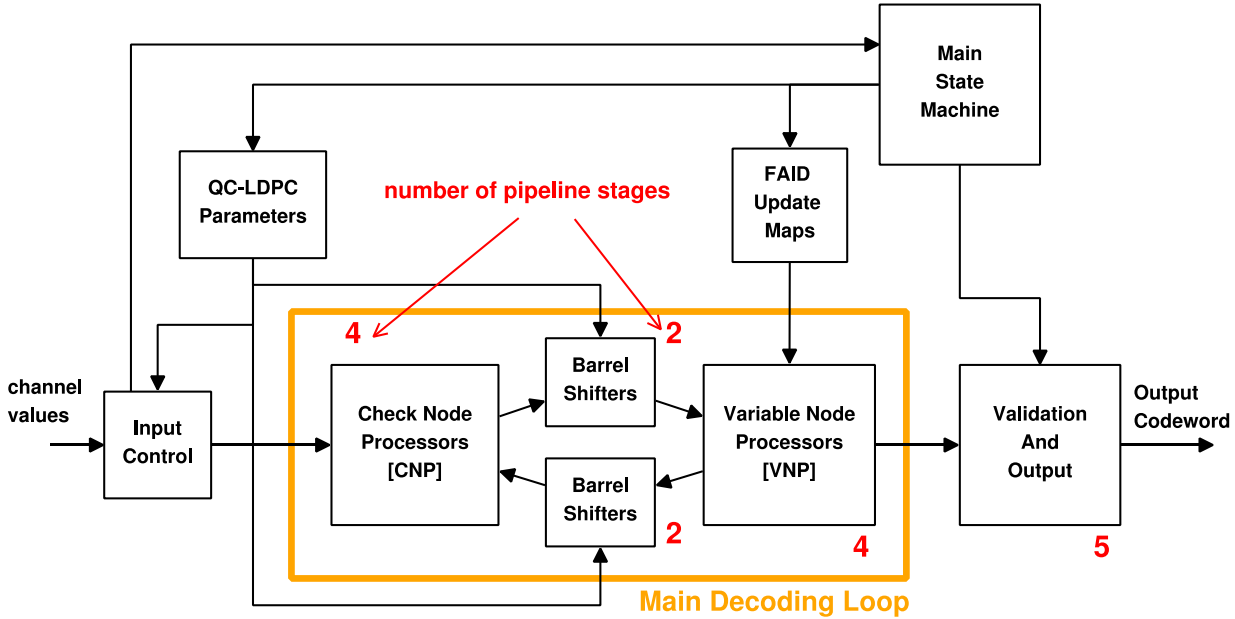
  ⟹ increase the **Frequency**

- Vertical layered decoding uses both **deep and shallow** memories. Optimization of the BRAM or LUTRAM implementation of the memories is crucial.

  ⟹ reduce the **Memory Usage** and increase **Frequency**

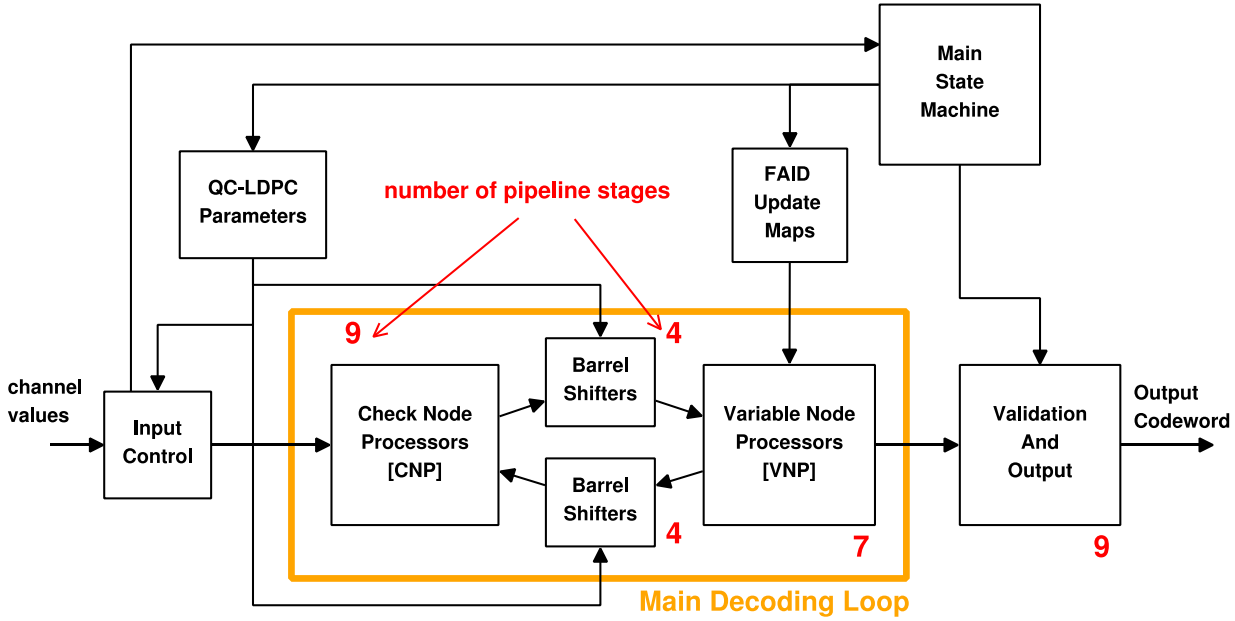
Pipeline Optimization @ 250Mz

Synthesis passes **@250Mhz** using **17 pipeline stages** at specific location



Pipeline Optimization @ 500Mz

Synthesis passes **@500Mhz** using **33 pipeline stages** at specific location



FPGA usage and frequency

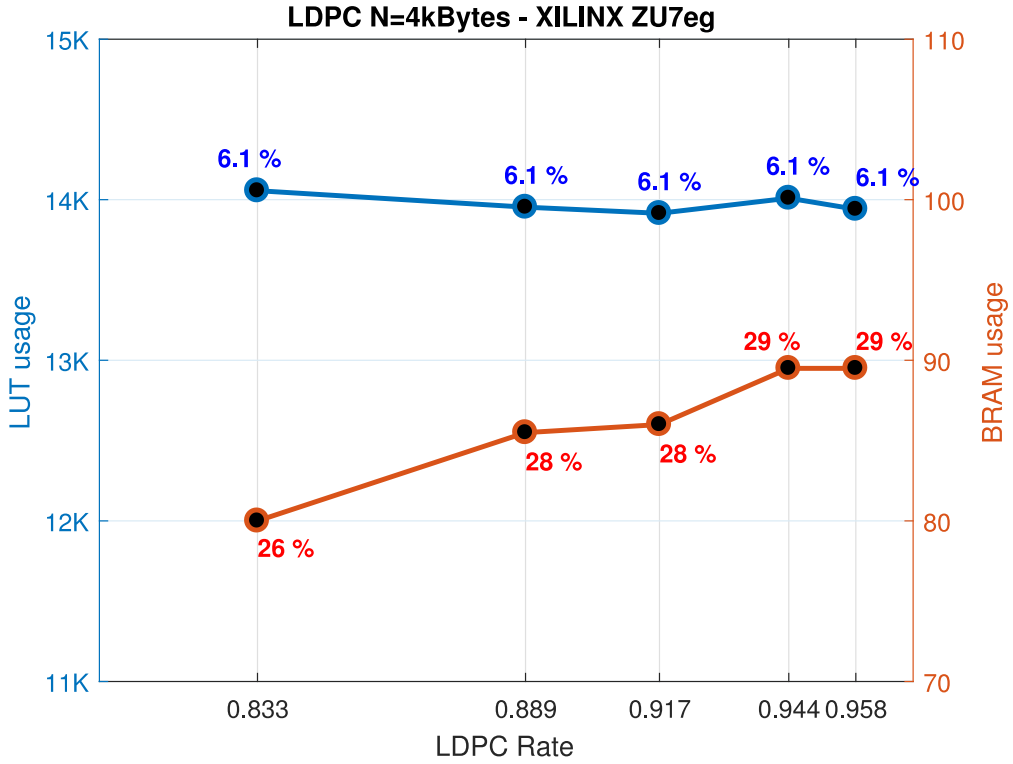
- Synthesis on **Zync Ultrascale ZU7eg** (xczu7eg-ffvf1517-2-e):
  - ✓ specs of the chip: 230K LUT - 460K Flipflops - 312 BRAM
  - ✓ maximum Frequency for Bram access 637Mhz
- 4kB QC-LDPC / Rate=0.89
  - ✓ synthesis passes **up to 600Mhz** (max. chip frequency)
  - ✓ very low resource usage: **< 7% LUT** and **< 11% BRAM**
  - ✓ **Tradeoff** between BRAM and LUT usage

@ 350Mhz T = 1.23 GB/s

	High BRAM	Low BRAM
LUT	14K	19K
FF	13K	19K
BRAM	85	33

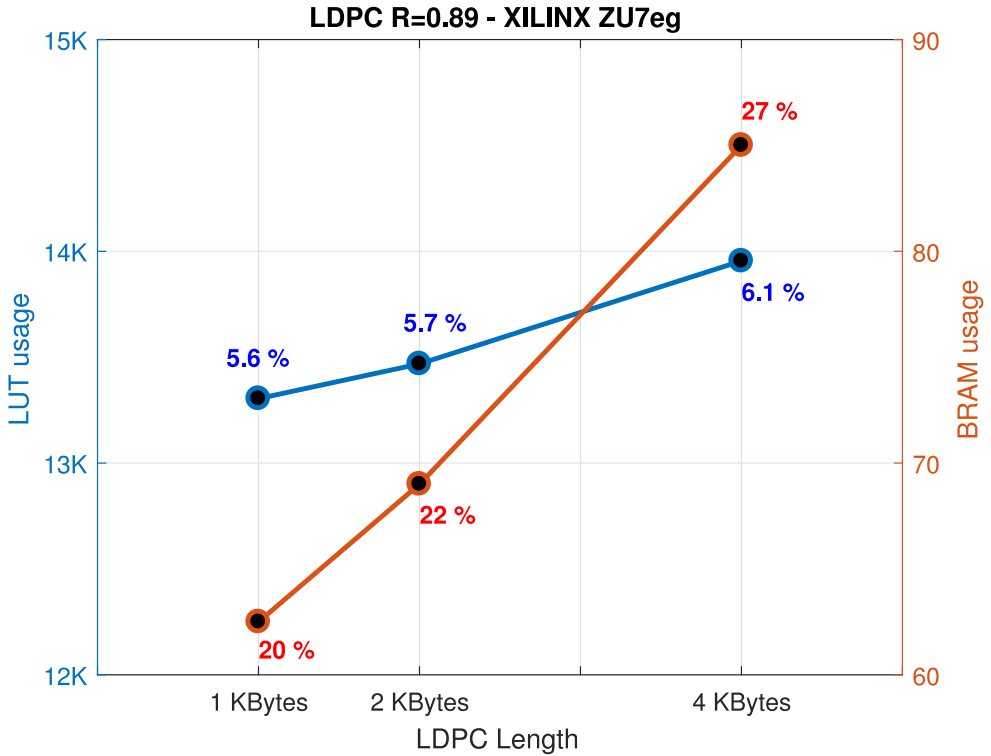
Usage vs. Coding Rate

- Results @350Mhz
  for T = 1.23 GB/sec
- LUT usage is **constant** whatever the rate
- Only slight increase in BRAM utilization
- Allows **multi-rate implementation** at the cost of a few extra BRAM
- **3% more BRAM** for each additional rate



Usage vs. Block Length

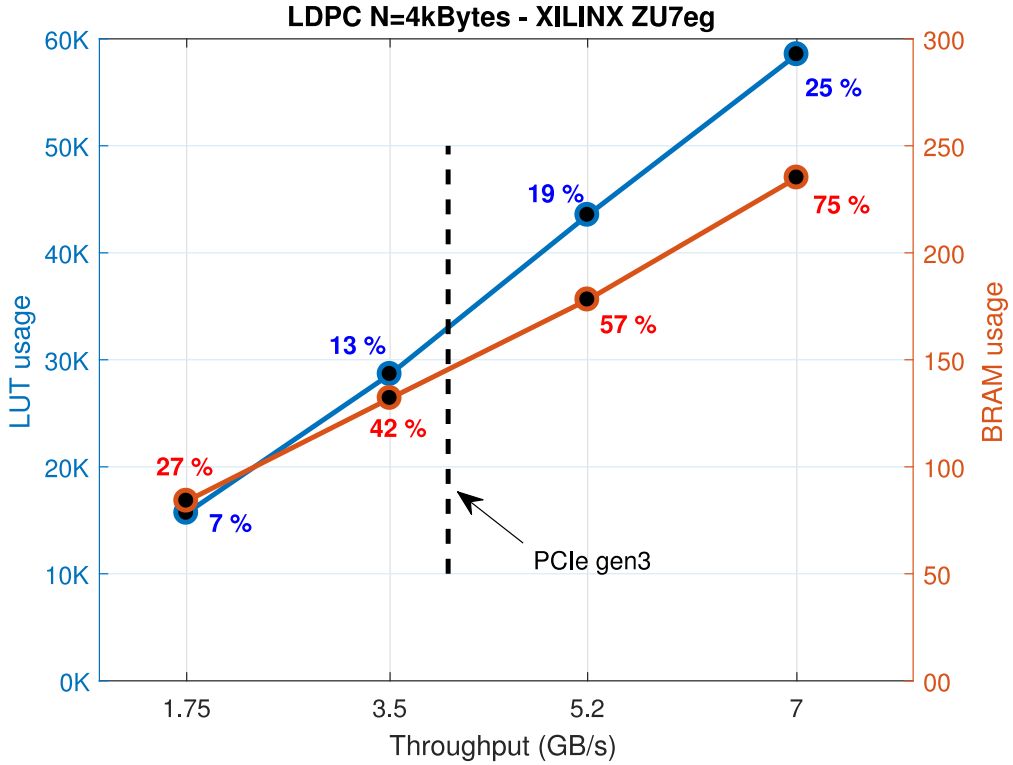
- Results @350Mhz
  for T = 1.23 GB/sec
- LUT usage is **almost constant** for all lengths
- Using small sizes requires **less BRAM**



Usage vs. Throughput

- Results @500Mhz
- **4GB/s** throughput with
  15% of LUT
  45% of the BRAM
  on XILINX ZU7eg
- LUT usage grows linearly
- At a given throughput, single core instance is **always beneficial** compared to multiple cores

@5.2 GB/s: 8% LUT saving
41% BRAM saving



Conclusion

- Proposed FAID LDPC decoder **optimized for FPGA**
  - ✓ Very low resource usage
  - ✓ Very high frequency
- Enable **PCIe throughputs** with low cost FPGA controllers
- Performance have been validated on actual **3D TLC NAND** Flash samples (in-house study),

Demonstration at the Booth #856
multi-core solution @500Mz on a Xilinx ZU7ev chip at T=6GB/s.