Introduction: Non-binary LDPC Decoder on FPGAs

- We explore a complex error-correction signal processing algorithm:
  - non-binary LDPC decoding (FFT-SPA)

- We utilize a high-level synthesis tool to design an LDPC decoder FPGA accelerator
- Vivado HLS allows:
  - fast design space exploration via directive optimizations
  - C/C++ as input for generating an FPGA accelerator

Proposed LDPC Decoder Accelerator

- LDPC decoder characteristics
  - 3-dimensions of computation:
    - \( N \times d \times M \times d_c \) probability mass functions (pmfs)
    - \( 2^m \) probabilities per pmf
    - \( d \times d_c \) pmf/node
    - \( 2^m \) is the Galois field dimension
  - each dimension is defined over a computation loop

- Applied LDPC computation:
  - Fast Walsh-Hadamard transform (WHT)
  - Hadamard products (\( \text{vn/cn} \))
  - Cyclic permutations ((de)permute)

Under the hood transformations:
- 3 different nested loop structures:
  - \( \text{cn/proc/vn/proc} \): 3 loops triple-nested
  - \( \text{depermute/permute} \): 2 loops double-nested
  - \( \text{vn/proc/nloop} \): 5 loops triple-nested
- no computation performed directly on DRAM data
- high bandwidth available but high latency of access
- data is moved to BRAM memory for computation at prologue and to DRAM memory at epilogue

High-Level Architecture

- Vivado HLS exports an accelerator design as an IP-XACT without external I/O, clock interface or AXI4 data mowers
- 1 DRAM and AXI-M controllers per SODIMM (2)
- 1 port on AXI-M controllers per accelerator instantiated (K)

Experimental Results: Latency vs. LUTs utilization

Comparison with RTL-based Decoders

Conclusions

- We show that combining the correct optimizations we are able to reach within 50% of RTL-based LDPC decoders

- Programming language is the same but programming model is different
- Code refactoring is still required
- Exploited parallelism dimensions are exposed in proper loop structures

- By instantiating the accelerators in a suitable high-level architecture we are able to fit multiple accelerators further elevating the parallelism level