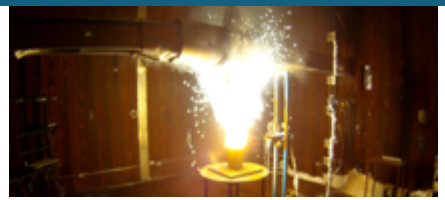
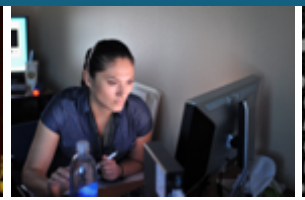
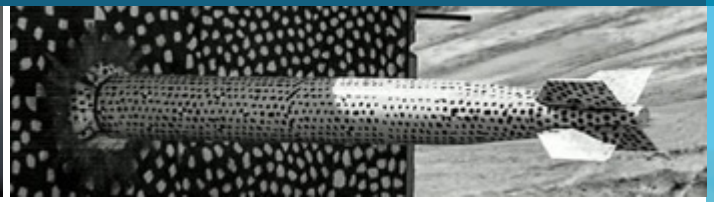
Sandia National Laboratories

SAND2020-12887PE

miniAMR port using oneAPI







Intern: Nicholas Miller

Mentor: Clayton Hughes

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Motivation

- Goal was to evaluate oneAPI for programmability and performance on FPGA's
- Ported the miniAMR proxy app to DPC++ and measured its performance on an Arria 10 development board
- MiniAMR was chosen due to the simplicity of the implementation and size of code base

miniAMR

- Part of the Mantevo proxy app suite
- Proxy app that simulates adaptive mesh refinement and distribution of work for multi-node systems
- Calculations performed on the grid can be easily changed to suit the programmer's needs
  - For the port we are calculating an average on the 7 point stencil of every point
- Calculation was the only part of the proxy app that was converted to oneAPI

Optimizations of Design

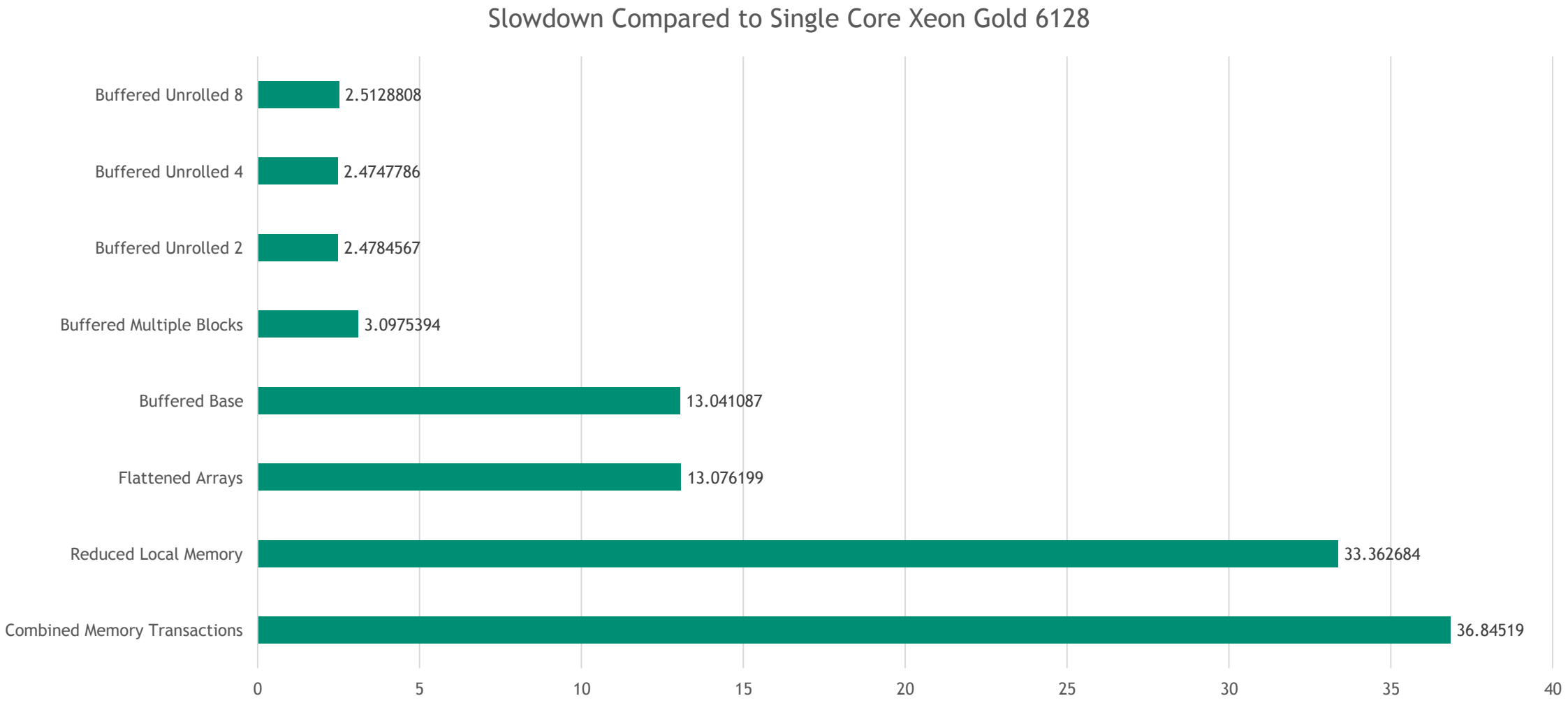
- Combined Memory Transactions
  - The optimization that provided the largest performance boost was to combine all the variable computations in a block into a single communication and computation step
  - This reduced the number of calls to the SYCL runtime by 40x
- Reduced Local Memory
  - Only maintains 1000 elements at a time in BRAM
- Flattened Arrays
  - Flattened all arrays to 1D accessors to make buffer creation and destruction faster
- Buffering
  - Buffers the SYCL runtimes to create overlap between kernel execution and command communication

Experimental Setup

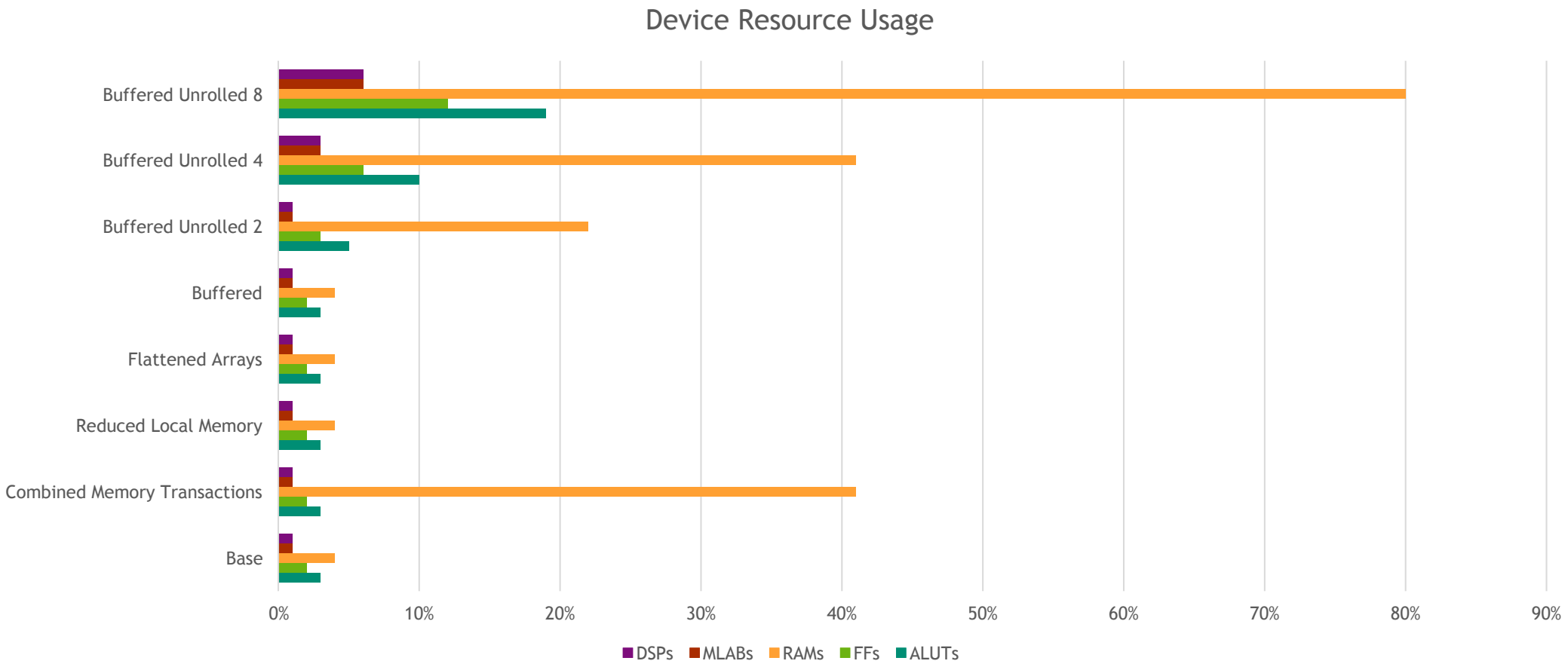
- Run on the Intel Devcloud system
- Increased number of blocks in a run to compare for the buffering tests

CPU	2 x Intel(R) Xeon(R) Gold 6128 CPU @ 3.40GHz
FPGA Family	Arria 10
FPGA Device	10AX115S2F45I2SGES
System Memory	196 GB
Base Parameters	No Parameters
Increased Blocks Parameters	--num_refine 4 --max_blocks 9000 --num_objects 1 --object 2 0 -1.71 -1.71 -1.71 0.04 0.04 0.04 1.7 1.7 1.7 0.0 0.0 0.0 --num_tsteps 25

Performance



Utilization



Conclusions

- Greatly reduced development time and digital logic knowledge needed compared to HDL
  - First iteration was complete within 2 days
- Best case design showed 2.4x slowdown compared to single core reference with ~22% BRAM usage
- Low transparency in device interactions
- Unable to fully customize hardware
- Full Paper can be found at: http://www.cs.sandia.gov/summerproceedings/CCR2020.html

Questions?