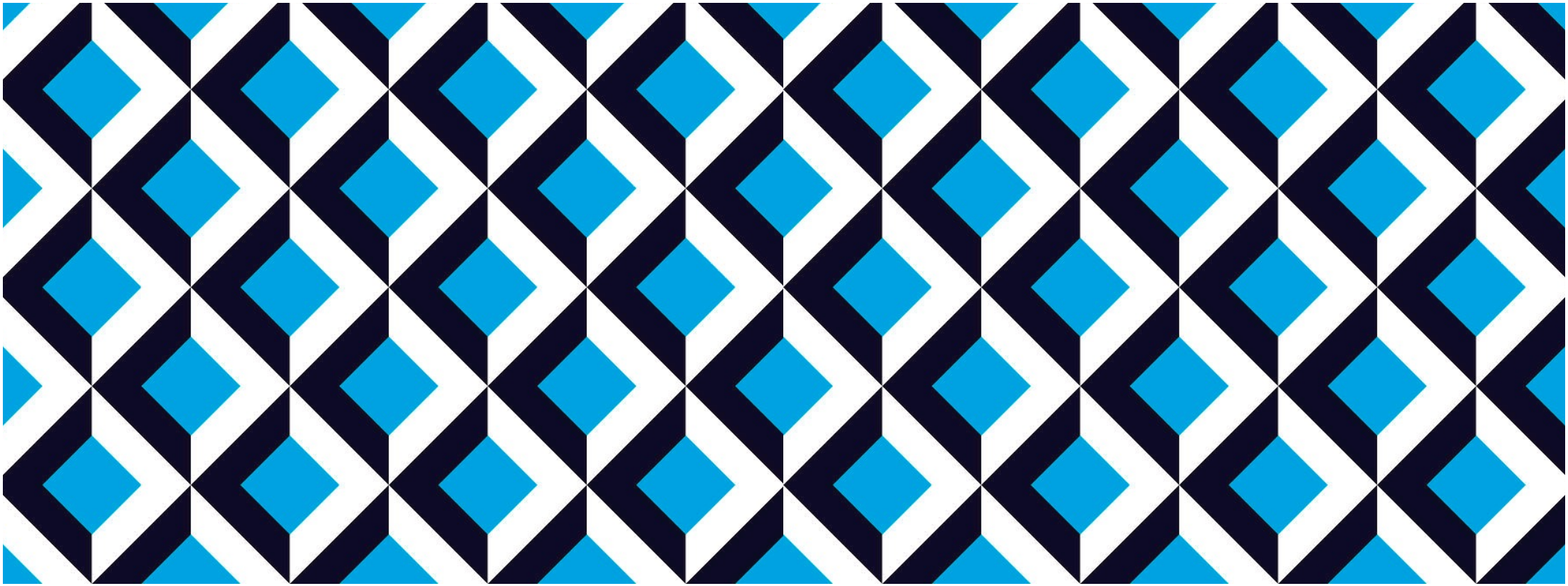


DEVELOPMENT AND PERFORMANCE OF HOUGH TRANSFORM ALGORITHM FOR EF TRACKING ON FPGA USING HLS

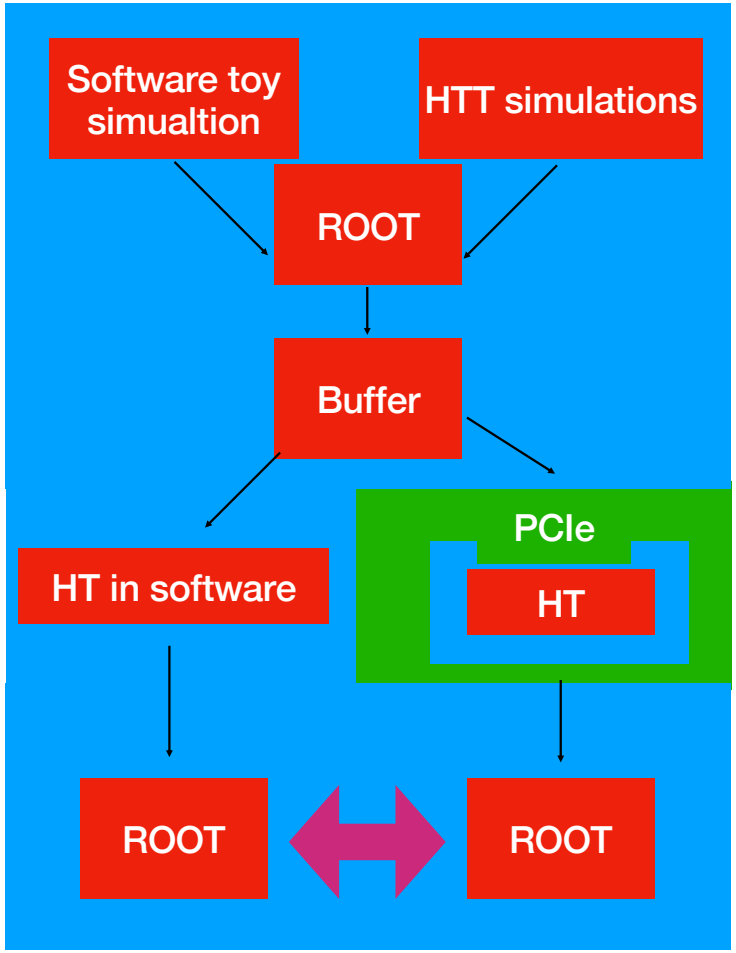
LUCY LEWITT
ALESSANDRO CERRI
MARIO GRANDI

UNIVERSITY OF SUSSEX

- We developed and tested a pattern recognition algorithm on FPGA using High Level Synthesis
  - Following the work performed for ATLAS on the development of a Hough Transform algorithm on FPGA with traditional tools (HDL)
  - Aim: assess viability of HLS
  - Performance
  - Resources
  - Portability
- What we have so far:
  - Software emulation → validation check and compare with existing implementations
  - HLS [“c++"] implementation to synthesize on FPGA
  - FPGA implementation optimized for performance and resource use
  - Performance and resource/power consumption studies

DEVELOPMENT SETUP

- Input:
  - Simplified SW simulation of detector hits
  - ATLAS simulation ”reference” ntuples from Phase II studies
- Intermediate format as ROOT files to runn HT either in:
  - software emulation
  - FW implemented on Xilinx accelerator card via standard OpenCL calls
- Outputs compared for debugging, sanity check, performance cross-check etc.



HARDWARE AT SUSSEX

Accelerators:

- 3x **U280 Alveo Cards**
- 2x U250 Alveo Cards
- 5x VCK5000
- ***3x U50***

Servers:

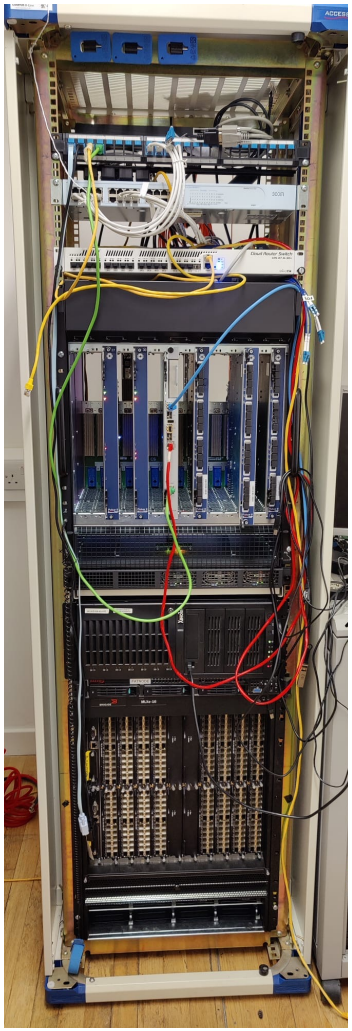
- PowerEdge T640 with one 2xU280, 1xU50
- PowerEdge R7920 with free accelerator slots
- PowerEdge R740 (128 GB) for firmware synthesis
- *PowerEdge R750 x2 (coming soon) able to host:*
  - *2 large and 4 small Accelerator*
  - *Gen 4 PCIe slots*

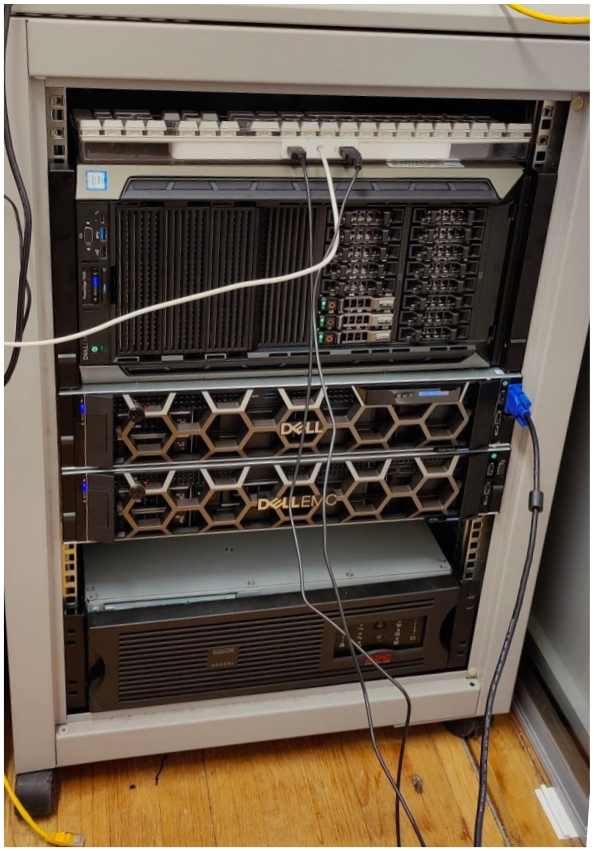
Other FPGA availabilities:

VC707, VC709 Virtex 7 Evaluation boards

VU37P Ultra scale Evaluation boards

4x VX690T on custom ATCA board





Contact us if you are interested in taking advantage of these resources

ACCELERATORS: QUICK FACTS

Card	Ext. RAM	HBM	SRAM	MLUT	PCIe	I/O	Slots	List Price $	
U280	32 GB 38 GBps	8 GB 460 GBps	41 MB	1.08	Gen4x8	2x100 GbE	2	8900	
U250	64 GB 77 GBps	-	54 MB	1.3	Gen3x16	2x100 GbE	2	9600	
U50	-	8 GB 316 GBps	28 MB	0.87	Gen3x16 2x Gen4x8	100 GbE	1	3400	
VCK5000	16 GB 70 GBps	-	24 MB	0.9	Gen3x16 Gen4x8	2x100 GbE	2	3400* 16400	VERSAL ACAP

SOFTWARE SIMULATION FOR DEVELOPMENT

Toy-pattern generator framework was developed at Sussex to quickly produce toy datasets with simplified:

- Particle gun generator
- Custom detector geometry emulator
- Cluster identification and parameterization

Simulated detector hits can be processed with software and Hardware implementation

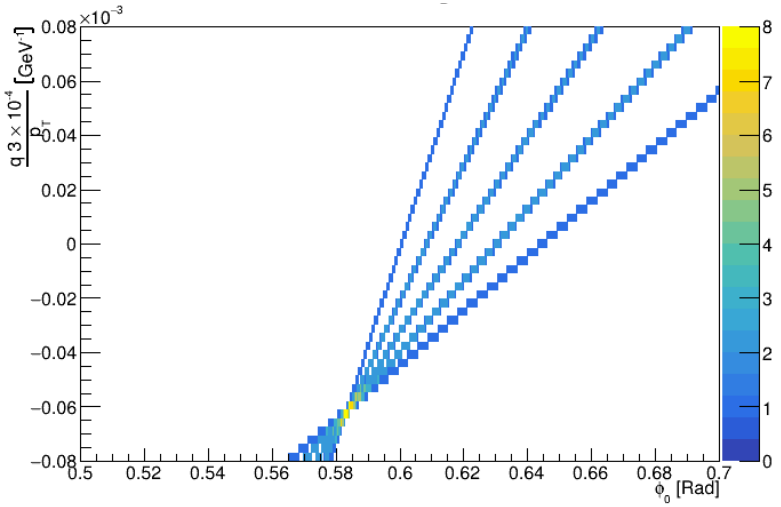
- Validation of firmware behavior

Fast tool enabling quick rough study of implementation

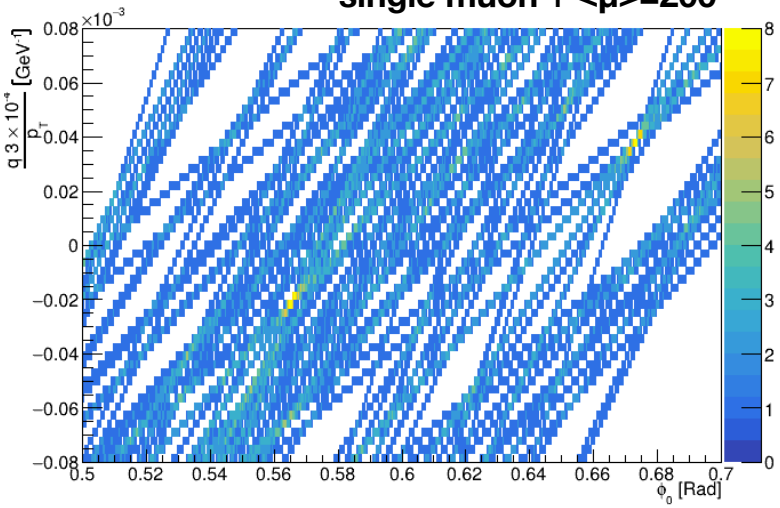
- High statistics
- Reduced simulation time (compared e.g. to full ATLAS simulation and firmware emulation)
- Validation of algorithm and comparison with HW

Useful as general purpos tool e.g. ML algorithms developed with same framework as part of undergraduate/master students projects at Sussex

single muon



single muon + <μ>=200



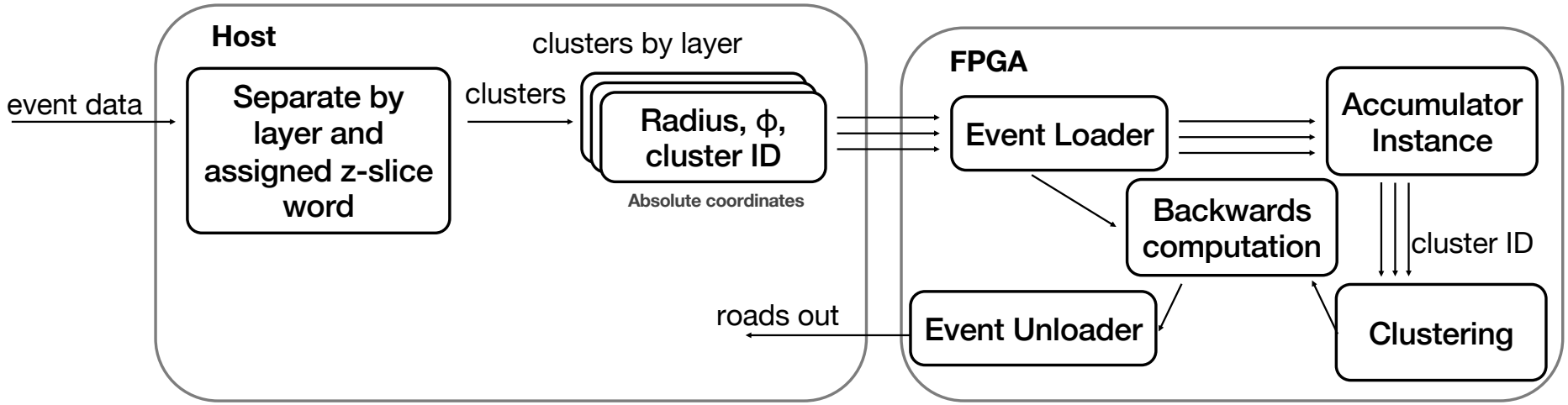
ALGORITHM IN HW

Hough Transform implemented on Xilinx U280/U50 Alveo FPGA PCIe accelerator, with the aim of:

- Extracting high performance parameters (latency, throughput, power,…) for tuning and benchmarks
- Investigation of data transfer between accelerator and host
- Educate ourselves on Xilinx's vendor-specific application tools (Vitis HLS)
- "By-produts":
  - Experience shared with community to help establish procedures, etc.
  - Flexible and re-configurable HT implementation adaptable to ATLAS use and beyond

Single implementation code can be shared between simulation and Vitis HLS

- Simulation and FPGA accelerator code consistent by design
- FPGA implementation and performance tuning driven by specific directives transparent to simulation
- …potential to evolve towards LHCb-style cross-architecture "Allen" framework



SW-HW COMPARISONS

Software

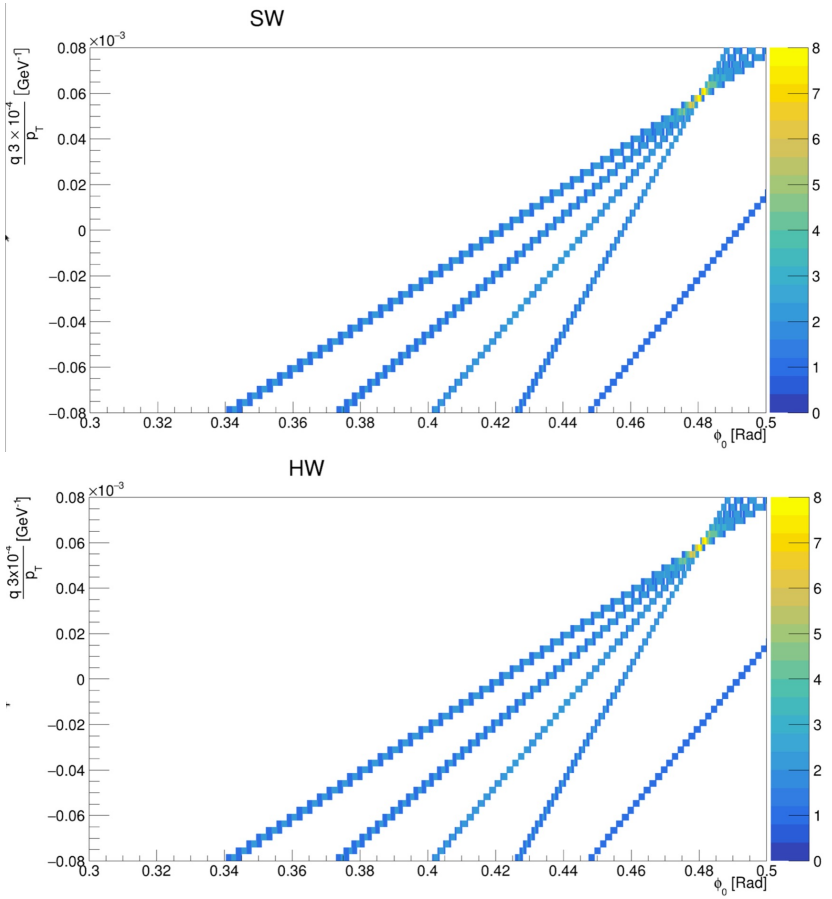
Row	EventNumber	number of roads
0	36810034	15
1	36810054	42
2	36810127	46
3	36810135	38
4	36810188	42
5	36810196	32
6	36810276	34
7	36810303	14
8	36810357	38
9	36810391	28

Row	Instance	ClusterGr	InvPtInde	Phi0Index	HitIndex
0	0	0	45	852	20
0	1	0	45	852	21
0	2	0	45	852	24
0	3	0	45	852	25
0	4	0	45	852	26
0	5	0	45	852	27
0	6	1	45	853	20
0	7	1	45	853	21
0	8	1	45	853	24
0	9	1	45	853	25
0	10	1	45	853	26
0	11	1	45	853	27
0	12	2	45	854	20
0	13	2	45	854	21
0	14	2	45	854	24
0	15	2	45	854	25
0	16	2	45	854	26
0	17	2	45	854	27
0	18	3	46	860	20
0	19	3	46	860	21
0	20	3	46	860	24
0	21	3	46	860	25
0	22	3	46	860	26
0	23	3	46	860	27

Hardware

Row	EventNumber	number of roads
0	36810034	15
1	36810054	42
2	36810127	46
3	36810135	38
4	36810188	42
5	36810196	32
6	36810276	34
7	36810303	14
8	36810357	38
9	36810391	28

Row	Instance	ClusterGr	InvPtInde	Phi0Index	HitIndex
0	0	0	45	852	20
0	1	0	45	852	21
0	2	0	45	852	24
0	3	0	45	852	25
0	4	0	45	852	26
0	5	0	45	852	27
0	6	1	45	853	20
0	7	1	45	853	21
0	8	1	45	853	24
0	9	1	45	853	25
0	10	1	45	853	26
0	11	1	45	853	27
0	12	2	45	854	20
0	13	2	45	854	21
0	14	2	45	854	24
0	15	2	45	854	25
0	16	2	45	854	26
0	17	2	45	854	27
0	18	3	46	860	20
0	19	3	46	860	21
0	20	3	46	860	24
0	21	3	46	860	25
0	22	3	46	860	26
0	23	3	46	860	27



- Cross-validation of HT implementation in HW and emulation
  - Number of “roads” found, position in HT space, comparison against emulation as well as full-precision SW implementation etc.
  - allows continuous cross-check throughout development, optimization, performance tuning
  - **Consistency is a requirement** → should become part of continuous integration

CONFIGURATIONS

Configuration	Accumulator (ϕ0 x qA/pT)	Hough space pattern recognition	z-slices	Accelerator
1	960x54	Sliding window (4 5 6 5 4)	4 (‖)	Alveo U50
2	64x216	Fixed (7)	19	Alveo U280/U50
3	64x216	Sliding window (4 5 6 5 4)	19	Alveo U50
4	960x216	Sliding window (4 5 6 5 4)	4 (‖)	Alveo U50
5	64x216	Fixed (7) with track xtrp	19	U280 (HBM and DDR)

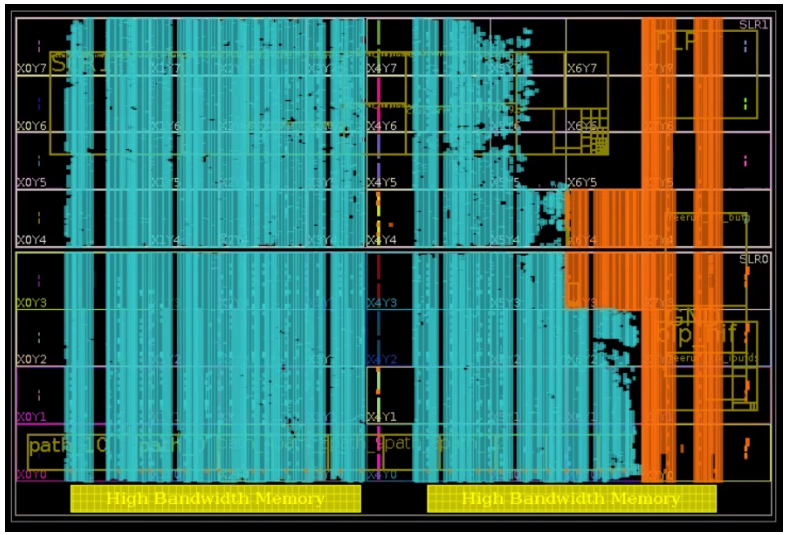
1. Inspired by original ATLAS studies
   - performance check
   - Early test of full synthesis, resource consumption and performance in actual accelerator
2. Reduced accumulator size (coarser)
   - inspired by subsequent ATLAS studies
3. “Complex” HT clustering (more resources than 2)
4. Extend 1 to pT>1
5. Include “track extension”: improve efficiency on smaller accumulator
   closest comparison to VHDL implementation

CONFIGURATION 1 - RESOURCE UTILISATION AND LATENCY

1 960x54 SlW 4||z 4 GeV

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	12.28%	2.83%	7.92%	13.24%	0.00%	0.07%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	42.14%	24.19%	11.25%	26.16%	0.00%	21.14%
Unused Resources	57.86%	75.81%	88.75%	73.84%	100.00%	78.86%
accumulator_instance (4)	23.67%	18.71%	4.03%	0.00%	0.00%	1.01%
backwards_computation (4)	9.81%	3.47%	3.52%	21.96%	0.00%	19.41%
cluster_unloader (1)	0.62%	0.17%	0.35%	0.09%	0.00%	0.00%
clustering (4)	6.92%	1.77%	2.81%	0.00%	0.00%	0.00%
event_loader (1)	1.12%	0.08%	0.53%	4.12%	0.00%	0.72%

Easily fits target accelerator platform



Synthesis:
Clock ~250 MHz w target 300 MHz
Synthesis time 9 h

CONFIGURATION 2 VS 3 - RESOURCE UTILIZATION ON SAME ACCELERATOR (U50)

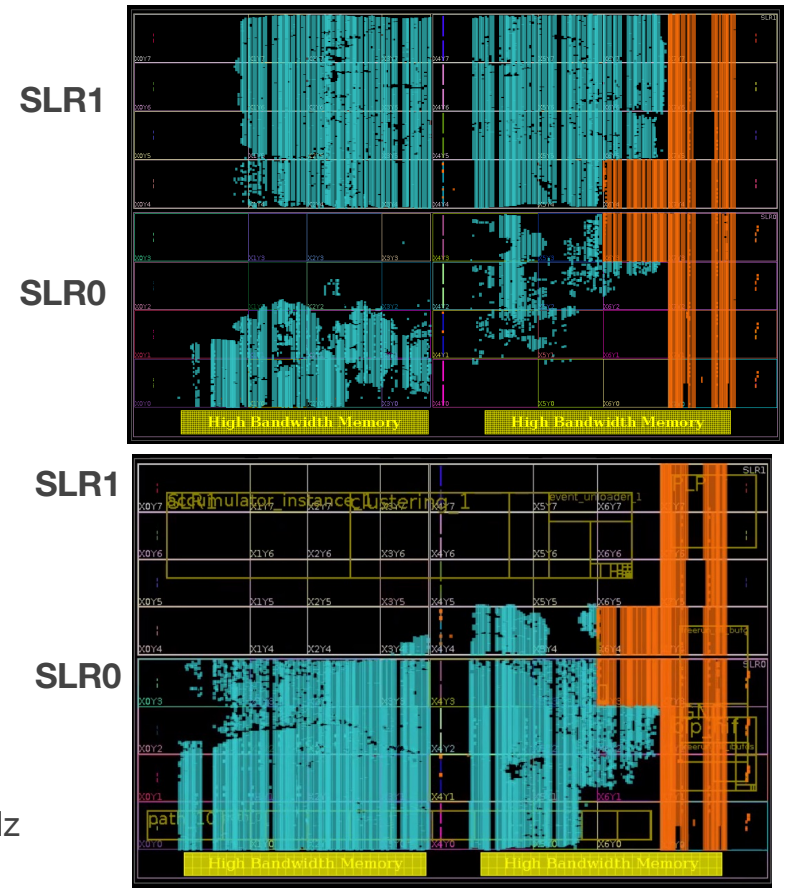
Alveo U50 – fixed thr.

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	12.15%	2.66%	7.78%	13.24%	0.00%	0.07%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	22.28%	6.51%	8.17%	9.69%	0.00%	5.83%
Unused Resources	77.72%	93.49%	91.83%	90.31%	100.00%	94.17%
accumulator_instance (1)	7.13%	3.63%	2.34%	0.00%	0.00%	0.25%
backwards_computation (1)	2.44%	0.87%	0.88%	5.49%	0.00%	4.85%
serial_cluster_unloader (1)	0.38%	0.16%	0.23%	0.09%	0.00%	0.00%
serial_event_loader (1)	1.05%	0.09%	0.46%	4.12%	0.00%	0.72%
tower_finder (1)	11.27%	1.77%	4.25%	0.00%	0.00%	0.00%

Alveo U50 w clustering

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	12.15%	2.66%	7.78%	13.24%	0.00%	0.07%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	22.73%	6.51%	8.74%	9.69%	0.00%	5.83%
Unused Resources	77.27%	93.49%	91.26%	90.31%	100.00%	94.17%
accumulator_instance (1)	6.89%	3.63%	2.34%	0.00%	0.00%	0.25%
backwards_computation (1)	2.44%	0.87%	0.88%	5.49%	0.00%	4.85%
clustering (1)	11.96%	1.77%	4.82%	0.00%	0.00%	0.00%
serial_cluster_unloader (1)	0.38%	0.16%	0.23%	0.09%	0.00%	0.00%
serial_event_loader (1)	1.05%	0.09%	0.46%	4.12%	0.00%	0.72%

3 64x216 SIW 19z 1 GeV



- **Processing z slices is serial rather than parallel**
- **Synthesized configuration 3 successfully full details TBD, resource consumption not much higher**

Synthesis:
Clock ~215 MHz/215 MHz w target 300 MHz
Synthesis time 6/6 h

CONFIGURATION 4- RESOURCE UTILIZATION ON U50

LARGE ACCUMULATOR (960X216) WITH SLIDING WINDOW CLUSTERING AND 4 Z SLICES

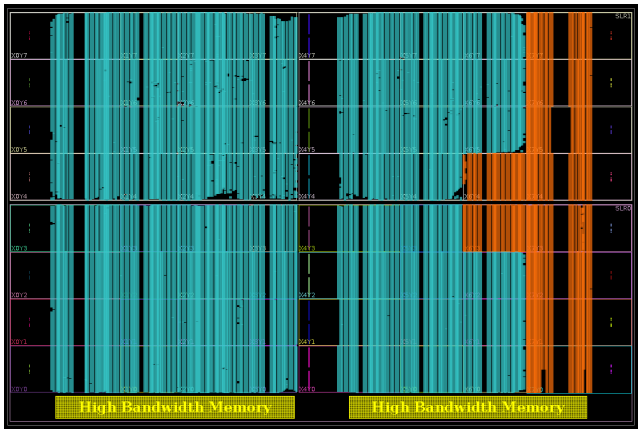
4 960x216 SlW 4||z 1 GeV

Alveo U50 w clustering Alveo U50 – fixed thr.

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	12.62%	3.57%	8.52%	13.24%	0.00%	0.07%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	64.48%	10.87%	25.69%	63.21%	0.00%	21.14%
Unused Resources	35.52%	89.13%	74.31%	36.79%	100.00%	78.86%
accumulator_instance (4)	4.11%	0.00%	1.21%	37.05%	0.00%	1.01%
backwards_computation (4)	9.79%	3.50%	3.56%	21.96%	0.00%	19.41%
cluster_unloader (1)	0.62%	0.17%	0.36%	0.09%	0.00%	0.00%
clustering (4)	48.86%	7.13%	20.03%	0.00%	0.00%	0.00%
event_loader (1)	1.10%	0.08%	0.53%	4.12%	0.00%	0.72%

SLR1

SLR0



- **Processing z slices is series rather than parallel**
- **Pt processing down to 1 GeV**

Synthesis:
Clock ~220 MHz MHz w target 300 MHz
Synthesis time 9h

CONFIGURATION 5 - RESOURCE UTILIZATION
HBM VS DDR

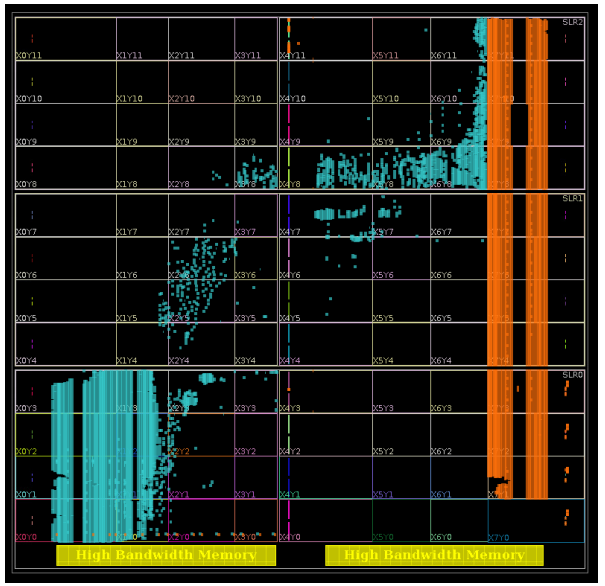
5 64x216 Fix7TX 19z 1 GeV

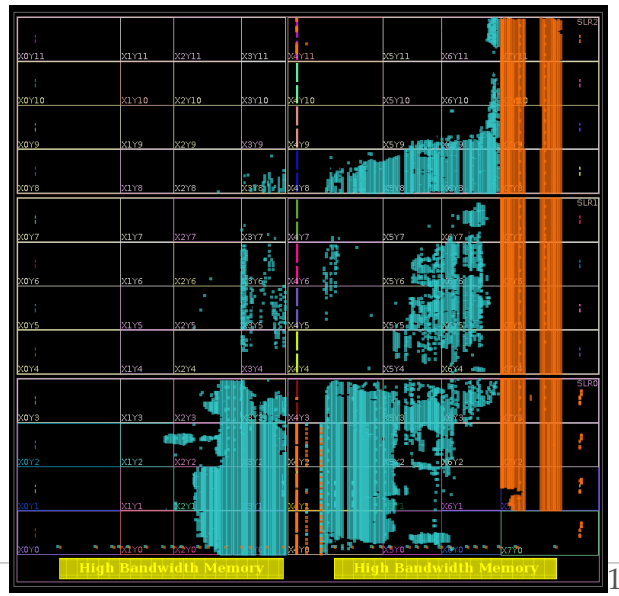
Alveo U280 using HBM

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	10.94%	2.92%	8.31%	14.14%	0.00%	0.08%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	4.60%	2.60%	2.54%	6.35%	0.00%	1.04%
Unused Resources	95.40%	97.40%	97.46%	93.65%	100.00%	98.96%
hough_transform_nau (1)	4.60%	2.60%	2.54%	6.35%	0.00%	1.04%
hough_transform_nau_1	4.60%	2.60%	2.54%	6.35%	0.00%	1.04%

Alveo U280 using DDR

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	7.92%	1.68%	5.78%	10.02%	0.00%	0.04%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	4.46%	2.56%	2.47%	6.06%	0.00%	1.04%
Unused Resources	95.54%	97.44%	97.53%	93.94%	100.00%	98.96%
hough_transform_nau (1)	4.46%	2.56%	2.47%	6.06%	0.00%	1.04%
hough_transform_nau_1	4.46%	2.56%	2.47%	6.06%	0.00%	1.04%





- **Processing z slices is serial rather than parallel**
- **Similar latency and resource usage with different types of memory**
  - **Resources and latency dominated by "processing" rather than "I/O"**

Synthesis:
Clock ~350 MHz/360 MHz w target 400 MHz
Synthesis time 5/7 h

CONFIGURATION 5 - RESOURCE UTILIZATION
U280 VS U50

5 64x216 Fix7TX 19z 1 GeV

Alveo U280

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	10.94%	2.92%	8.31%	14.14%	0.00%	0.08%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	4.60%	2.60%	2.54%	6.35%	0.00%	1.04%
Unused Resources	95.40%	97.40%	97.46%	93.65%	100.00%	98.96%
hough_transform_nau (1)	4.60%	2.60%	2.54%	6.35%	0.00%	1.04%
hough_transform_nau_1	4.60%	2.60%	2.54%	6.35%	0.00%	1.04%

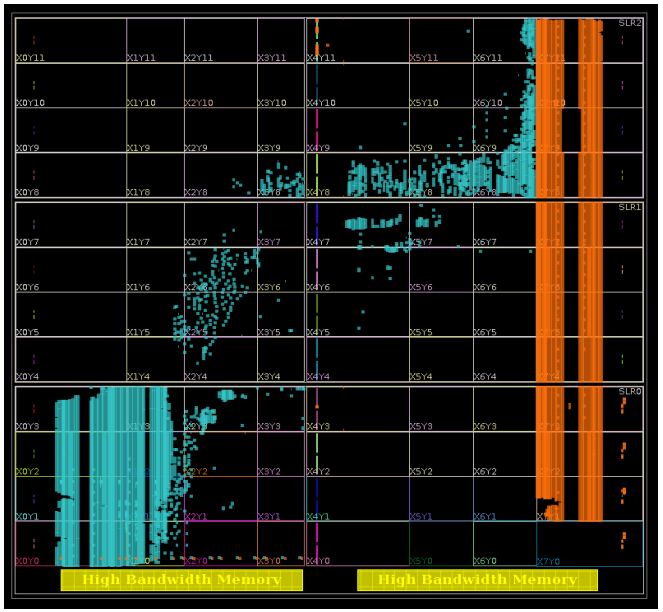
Alveo U50

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	11.65%	2.21%	7.21%	13.24%	0.00%	0.07%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	6.94%	3.85%	3.76%	9.43%	0.00%	1.58%
Unused Resources	93.06%	96.15%	96.24%	90.57%	100.00%	98.42%
hough_transform_nau (1)	6.94%	3.85%	3.76%	9.43%	0.00%	1.58%
hough_transform_nau_1	6.94%	3.85%	3.76%	9.43%	0.00%	1.58%

SLR2

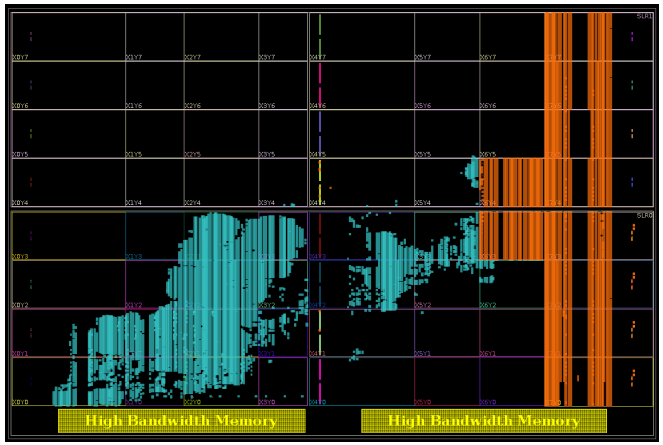
SLR1

SLR0



SLR1

SLR0

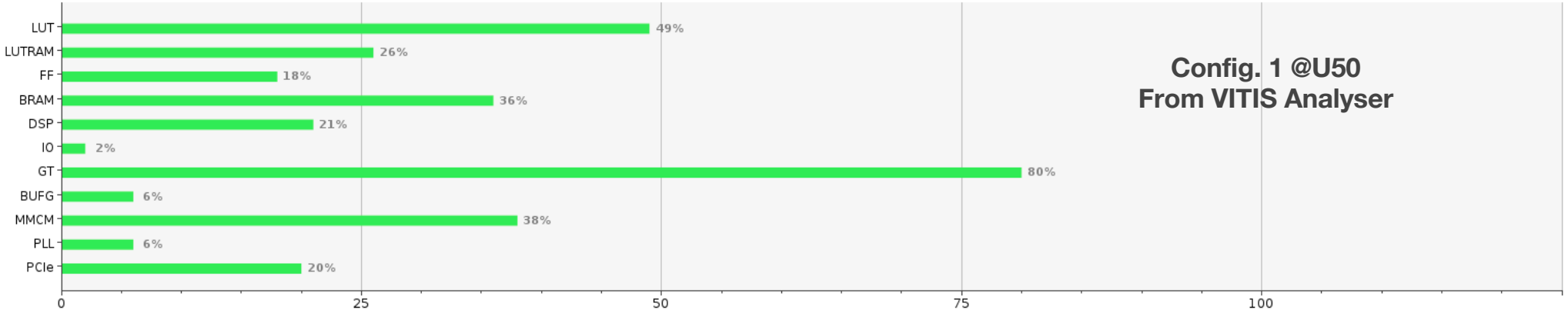


- **Comparable resource utilization (U50 *~66% of U280)***
- **No significant difference in terms of power/synthesis**

Synthesis:
Clock ~350 MHz/380 MHz w target 400 MHz
Synthesis time 5/5 h

CONFIGURATION 1 & 2 – RESOURCE UTILIZATION DETAILS

Pre vs post-synthesis resource estimation can differ significantly: is what the VITIS tools assess reliable?



Config. 1 @U50
From VITIS Analyser

Config. 1 @U50 From post-synthesis & implementation

Name	LUT	LUTAsMem	REG	BRAM	URAM	DSP
Platform	12.28%	2.83%	7.92%	13.24%	0.00%	0.07%
User Budget	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Used Resources	42.14%	24.19%	11.25%	26.16%	0.00%	21.14%
Unused Resources	57.86%	75.81%	88.75%	73.84%	100.00%	78.86%
accumulator_instance (4)	23.67%	18.71%	4.03%	0.00%	0.00%	1.01%
backwards_computation (4)	9.81%	3.47%	3.52%	21.96%	0.00%	19.41%
cluster_unloader (1)	0.62%	0.17%	0.35%	0.09%	0.00%	0.00%
clustering (4)	6.92%	1.77%	2.81%	0.00%	0.00%	0.00%
event_loader (1)	1.12%	0.08%	0.53%	4.12%	0.00%	0.72%

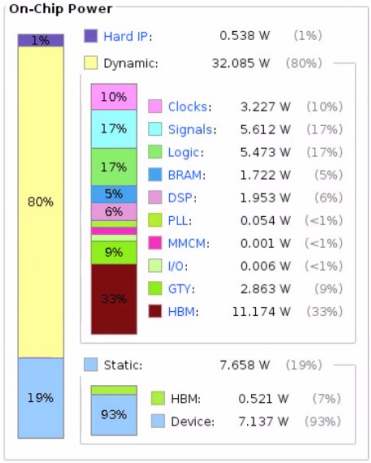
VITIS Analyser, very close to post-synthesis →corroborates numbers shown for all configurations (here config. 1)

POWER CONSUMPTION

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**40.294 W**
FPGA Power:	33.866 W
HBM Power:	6.428 W
Design Power Budget:	**63 W**
Power Budget Margin:	**22.706 W**
Junction Temperature:	**85.2°C**
Thermal Margin:	14.8°C (17.2 W)
Effective θJA:	0.8°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Medium

Launch Power Constraint Advisor to find and fix invalid switching activity

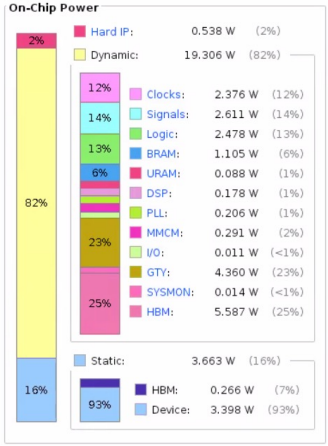


Config. 1 @U50
960x254 sl.w 4||z
250 MHz

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**23.535 W**
FPGA Power:	20.312 W
HBM Power:	3.223 W
Design Power Budget:	**Not Specified**
Power Budget Margin:	**N/A**
Junction Temperature:	**35.3°C**
Thermal Margin:	64.7°C (136.8 W)
Effective θJA:	0.4°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Low

Launch Power Constraint Advisor to find and fix invalid switching activity



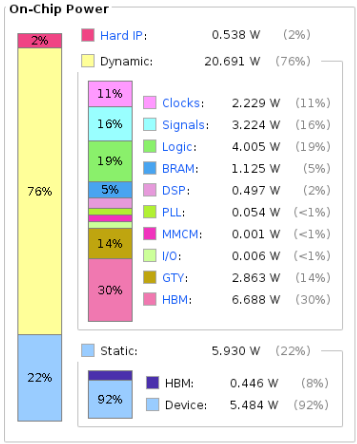
Config. 2 @U280
64x216 fix7 19z
200 MHz

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**27.175 W**
FPGA Power:	23.187 W
HBM Power:	3.988 W
Design Power Budget:	**63 W**
Power Budget Margin:	**35.825 W**
Junction Temperature:	**75.4°C**
Thermal Margin:	24.6°C (28.6 W)
Effective θJA:	0.8°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Medium

Launch Power Constraint Advisor to find and fix invalid switching activity

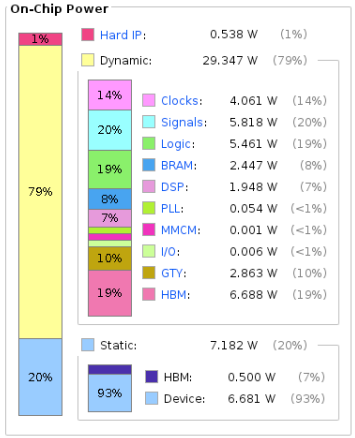
Config. 2 @U50
64x216 fix7 19z
220 MHz



Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**37.08 W**
FPGA Power:	33.039 W
HBM Power:	4.04 W
Design Power Budget:	**63 W**
Power Budget Margin:	**25.92 W**
Junction Temperature:	**82.8°C**
Thermal Margin:	17.2°C (19.9 W)
Effective θJA:	0.8°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Medium

Launch Power Constraint Advisor to find and fix invalid switching activity



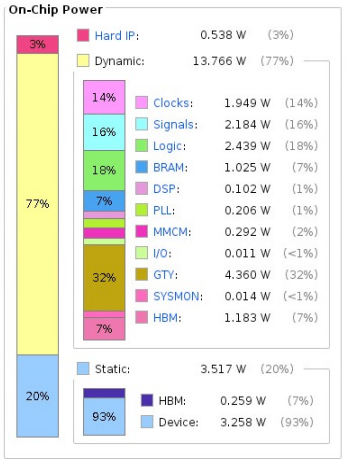
Config. 4 @U50
960x216 sl.w 4||z
220 MHz

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**17.828 W**
FPGA Power:	16.981 W
HBM Power:	0.848 W
Design Power Budget:	**Not Specified**
Power Budget Margin:	**N/A**
Junction Temperature:	**32.8°C**
Thermal Margin:	67.2°C (142.4 W)
Effective θJA:	0.4°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Low

Launch Power Constraint Advisor to find and fix invalid switching activity

Config. 5 @U280
64x216 fix7 19z TX
350 MHz

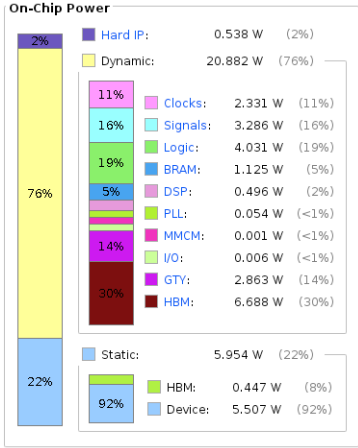


Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

otal On-Chip Power:	**27.387 W**
FPGA Power:	23.398 W
HBM Power:	3.989 W
esign Power Budget:	**63 W**
ower Budget Margin:	**35.613 W**
unction Temperature:	**75.5°C**
hermal Margin:	24.5°C (28.4 W)
ffective θJA:	0.8°C/W
ower supplied to off-chip devices:	0 W
onfidence level:	Medium

aunch Power Constraint Advisor to find and fix ivalid switching activity

Config. 3 @U50
64x216 sl.w 19z
360 MHz



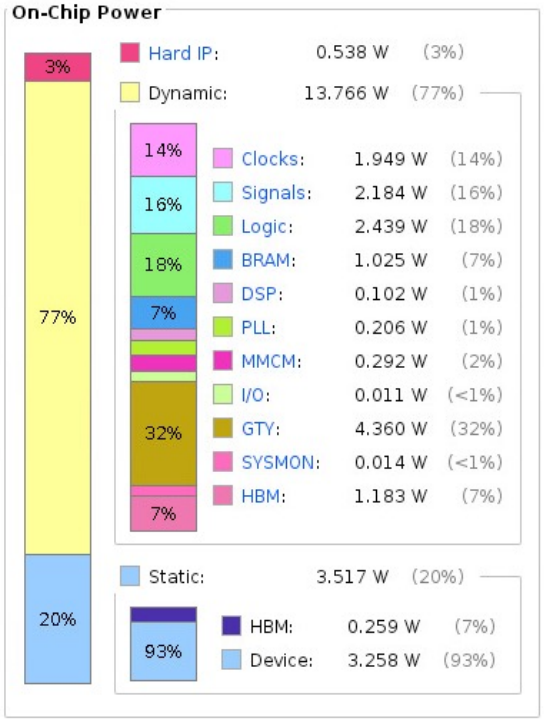
- Similar power & clock and about ¼-½ resources compared to what presented 5/7/22

CONFIGURATION 5 – POWER ANALYSIS - 64x216 fix7 19z TX

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**17.828 W**
FPGA Power:	16.981 W
HBM Power:	0.848 W
Design Power Budget:	**Not Specified**
Power Budget Margin:	**N/A**
Junction Temperature:	**32.8°C**
Thermal Margin:	67.2°C (142.4 W)
Effective θJA:	0.4°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Low

Launch Power Constraint Advisor to find and fix invalid switching activity

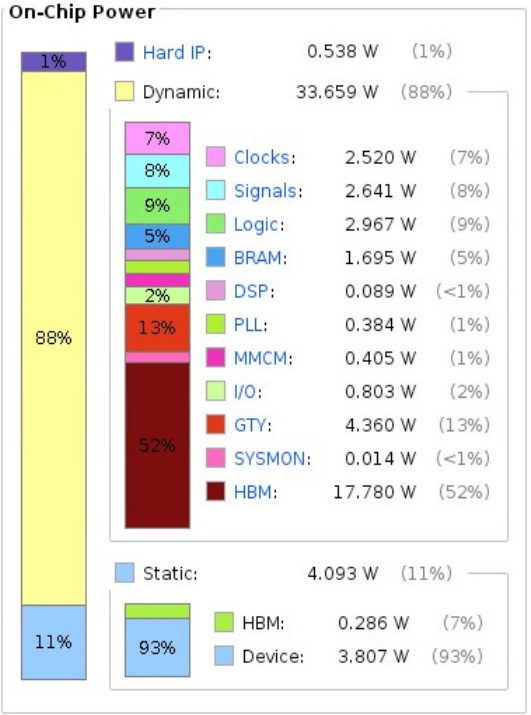


Config. 5 @ U280-HBM

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**38.293 W**
FPGA Power:	28.537 W
HBM Power:	9.756 W
Design Power Budget:	**Not Specified**
Power Budget Margin:	**N/A**
Junction Temperature:	**41.7°C**
Thermal Margin:	58.3°C (122.5 W)
Effective θJA:	0.4°C/W
Power supplied to off-chip devices:	0 W
Confidence level:	Low

Launch Power Constraint Advisor to find and fix invalid switching activity

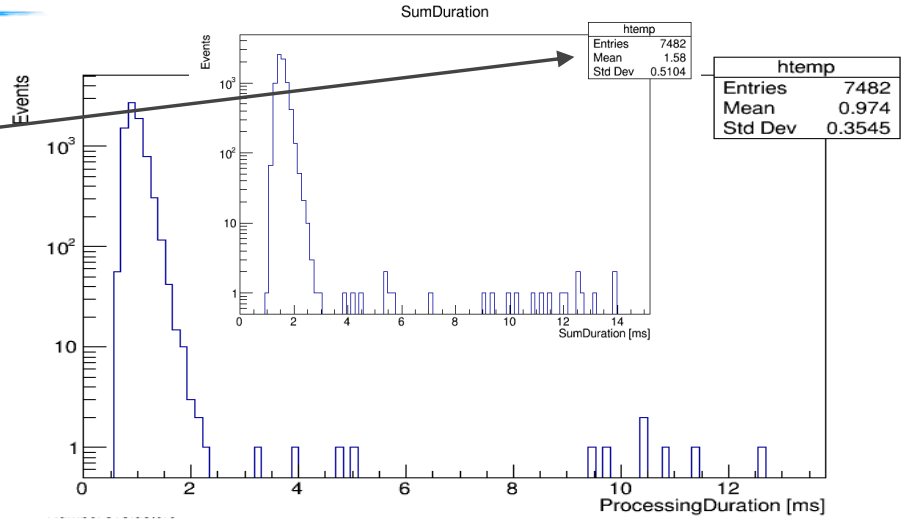


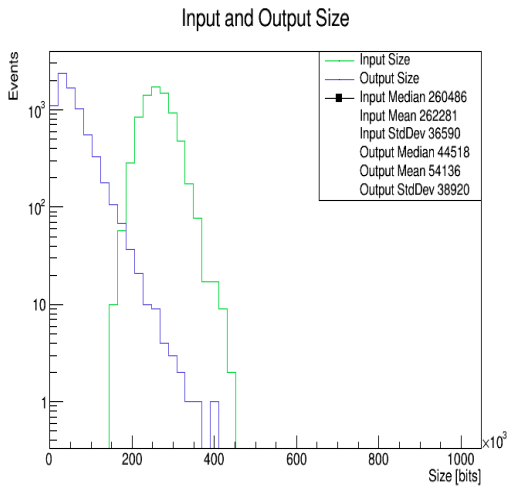
Config. 5 @ U280-DDR

…uninitialized HBM?

CONFIG. 1 LATENCY HTT SAMPLE 200 PU, SINGLE MUON @ **4 GEV** (01-0.3 ETA), 7.5K EVENTS

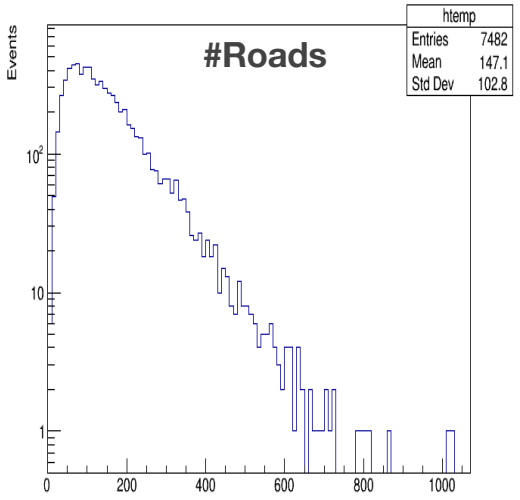
- <Latency>
  - With PCIe transfer ~1.6 ms (stdev 0.5 ms)
  - Without PCIe tranfer ~1 ms (stdev 0.35 ms)

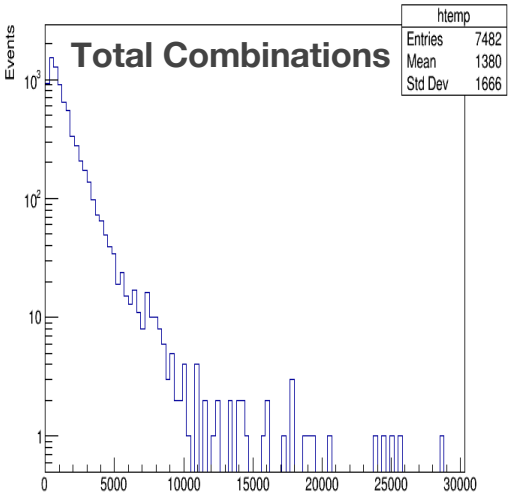


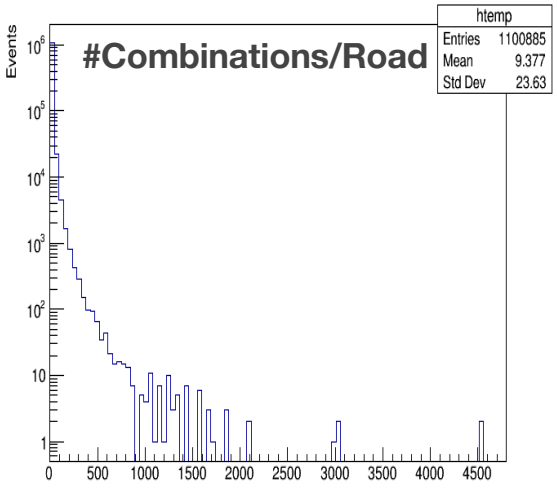


From ATLAS studies

scenario	configuration	$\langle n_{\text{roads}} \rangle$	$\langle n_{\text{combos}} \rangle$
s1	nominal	300	2,800
s1	space points	670	2,700
s1	stubs	860	5,700
s1	space points and stubs	1,200	3,800
s2	nominal	370	3,800
s2	space points	1,100	5,500
s2	stubs	1,100	7,700
s2	space points and stubs	2,000	8,300
s4	nominal	780	3,800







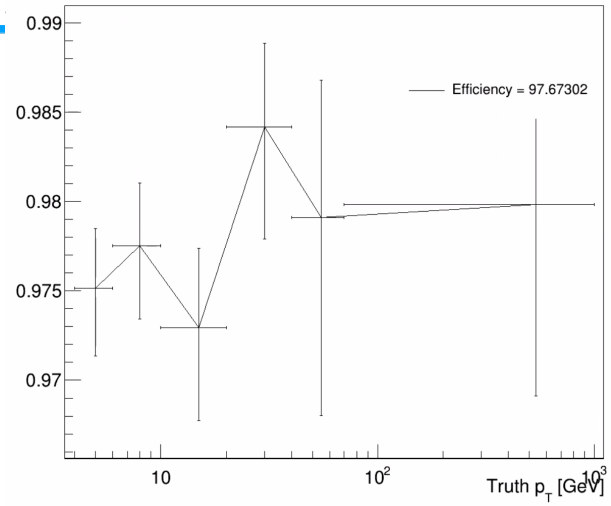
RESULTS SUMMARY TABLE

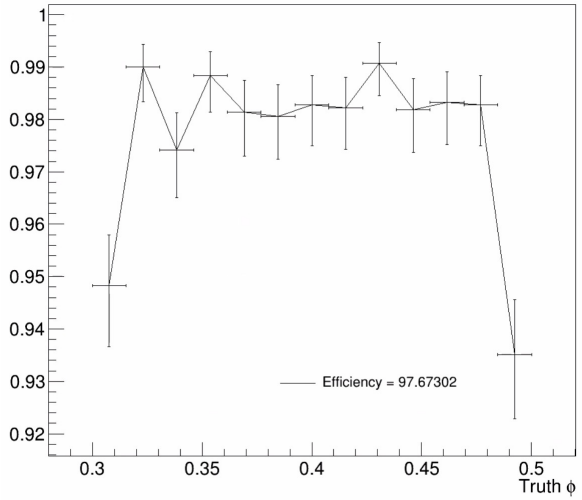
	Configuration	Latency ms		Resources [U50]	Power W	Energy mJ/ev	Clock MHz	Synth. Time h	ε	Input size kB mean±RMS Median	Output size kB mean±RMS Median	#output roads mean±RMS Median	#output comb. mean±RMS Median	# comb./road mean±RMS Median
		Excl. PCIe	Incl. PCIe											
1	960x54 SlW 4||z 4 GeV	1±0.3 0.9	1.6±0.5 1.5	50%	40	60	250	9	96%	33±5 32	7±5 5.5	150±100 122	1400±1700 910	9±24 25
2	64x216 Fix7 19z 1 GeV	3.3±0.2 3.3	3.9±0.6 3.8	23%	27	100	215	6	85%	73±9 73	2.2±1.4 1.8	38±25 33	750±1170 470	20±75 54
3	64x216 SlW 19z 1 GeV	3.3±0.4 3.2	3.8±0.5 3.8	23%	27	100	215	6	-	73±9 73	0.1±0.2 1.2	1.3±3.5 0.7	66±540 136	48±210 42
4	960x216 SlW 4||z 1 GeV	7.3±1.5 7.2	7.9±1.6 7.7	65%	37	285	220	9	95%	73±9 73	25±16 22	550±350 480	5200±4700 4000	9±22 25
5	64x216 Fix7TX 19z 1 GeV	0.8±0.4 0.75	1.6±0.7 1.5	5% U280	18	27	350	5	88%	73±9 73	6.7±4.5 5.5	100±65 87	3500±5300 2200	35±140 200

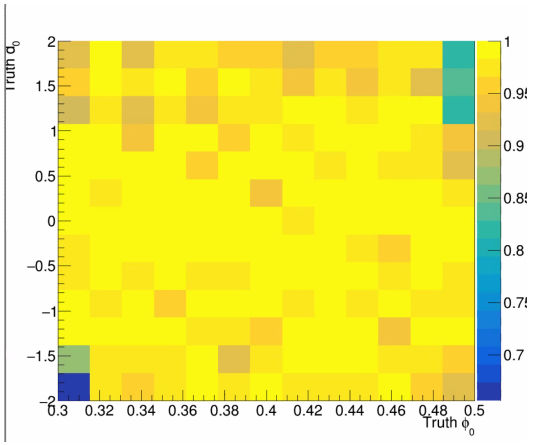
Single muon events wit PU=200

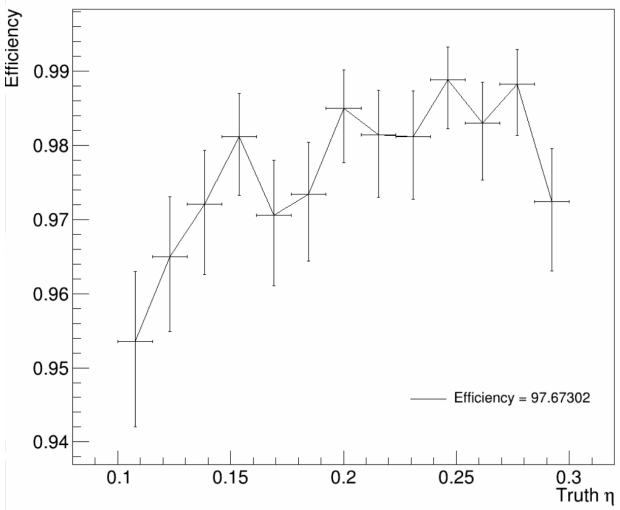
- Latency well within online farm needs: could e.g. parallelise to save FPGA resources
- Impact of data transfer/PCIe not critical
- Resources optimisation can massively improve requirements (e.g. 5 vs 2)
- Power: how does this compare to CPU use?

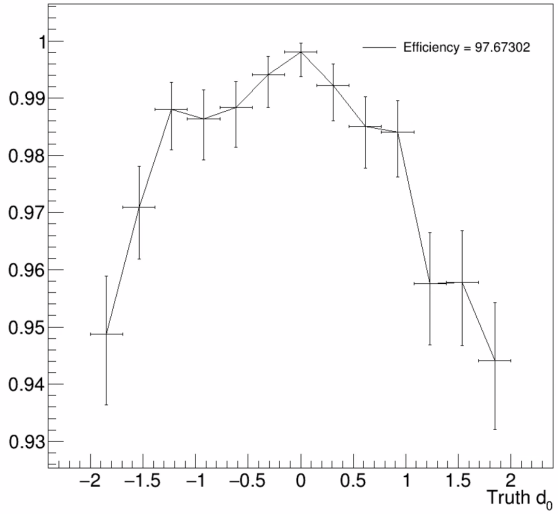
CONFIG. 1 PATTERN RECOGNITION PERFORMANCE SIMULATION @ 200 PU, SINGLE MUON @ 4 GEV

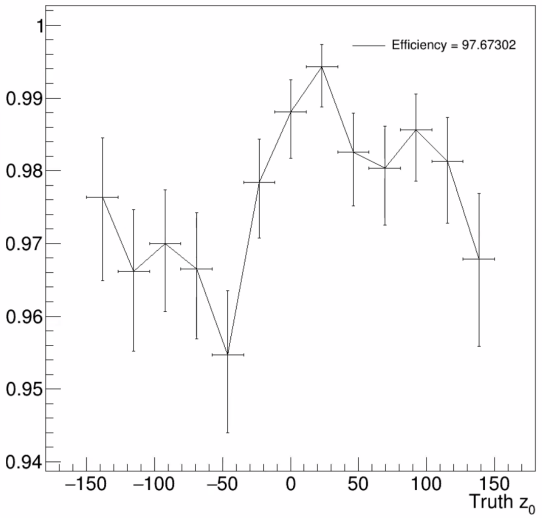






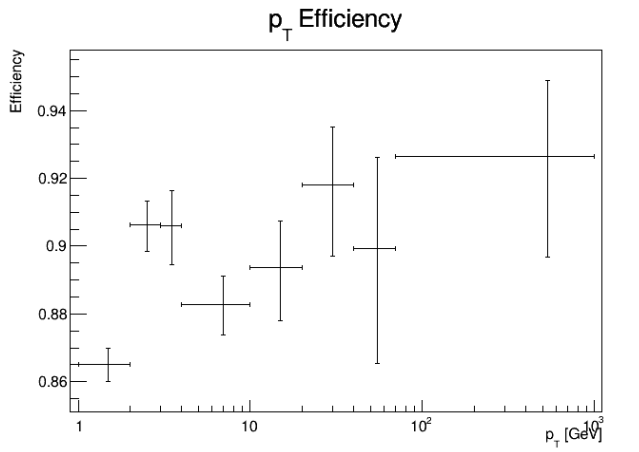


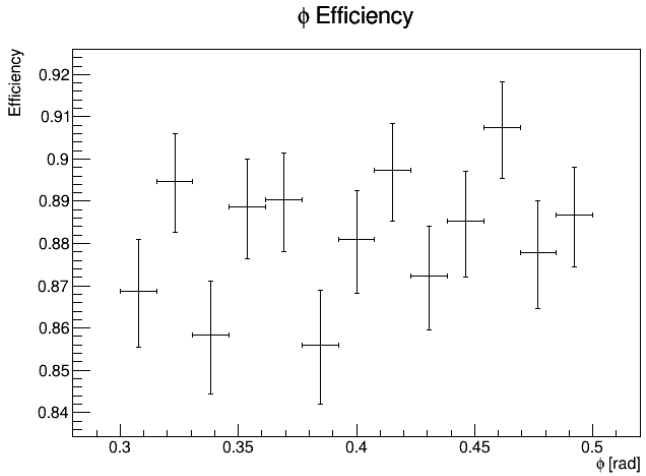


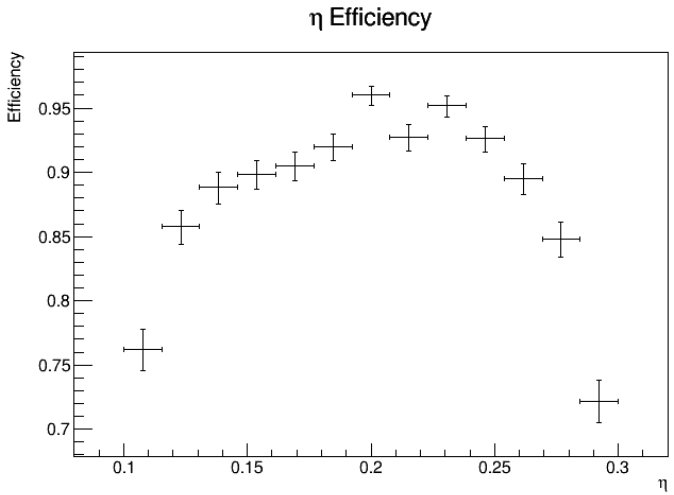


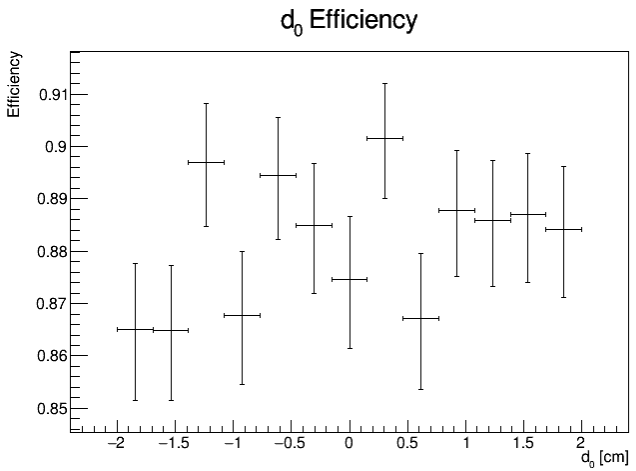
Efficiencies look fairly reasonable and follow what seen in ATLAS studies

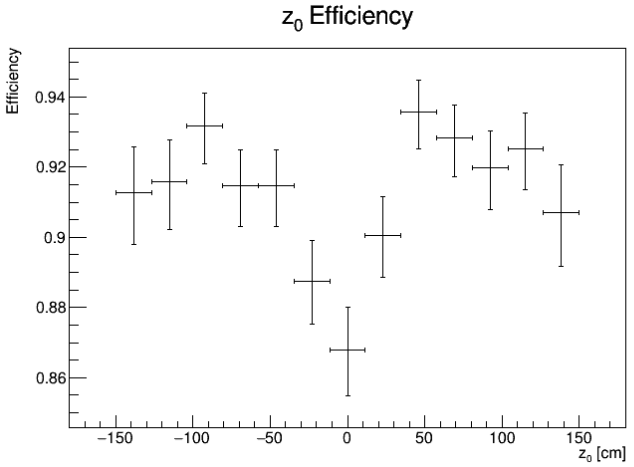
CONFIG. 5 PATTERN RECOGNITION PERFORMANCE SIMULATION @ 200 PU, SINGLE MUON @ 1 GEV











Efficiencies look fairly reasonable and follow what seen in ATLAS studies

CONCLUSION AND OUTLOOK

HLS Prototype of Hough Transform dimensioned on ATLAS studies has been developed and tested at Sussex joining SWIFT-HEP and ATLAS phase II HLT efforts

- Efficiency, combinatorics etc. consistent with expectations/other studies
- Encouraging benchmarks (resources, power, performance)
  - Will itereate further on the basis of first results

Implemented on commercial FPGA accelerators with **HLS**:

- Xilinx vendor-specific development tools
- Requirements and performance are in line with (if not better than!) what seen with VHDL so far
- Highly re-configurable:
  - HT parameters (binning, ranges, clustering, etc.)
  - Portability to other FPGA accelerators (e.g. U50/U250/U280/VCK5000 etc.)
- Open to cross-platform seamless operation “a` la LHCb”

- REMINDER: Hardware available at Sussex for developing/testing

TWiki page documenting all relevant information is being maintained as part of “SWIFT-HEP” UK effort

- Useful tool to centralise resources and results → Please use, contribute, and suggest!

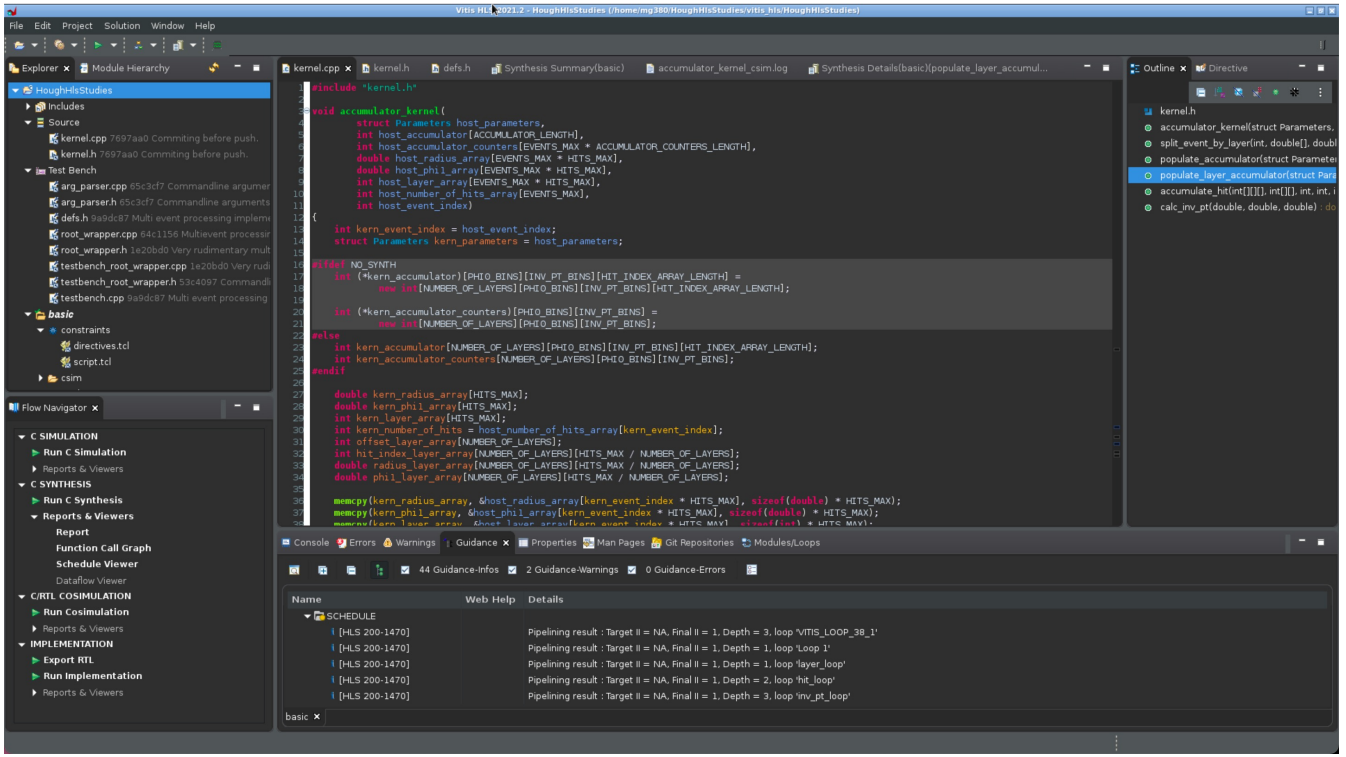
Integration of HT on FPGA accelerators with ACTS framework would allow for further future studies and more thorough physics performance evaluation

BACK-UP

OPTIMISATION USING VITIS HLS GUI

GUI allows for easy editing and debugging

- “C simulation”
  - → compiles the code in C++
  - → Debugging tools available
- “C synthesis”
  - → Simulates compilation of code on FPGA
  - → Summary and reports generated on expected resource consumption and processing time
- “C/RTL cosimulation”
  - → allows user to verify that code is functionally identical to C++ source code



Once all steps are completed the code can be exported to RTL and run on FPGA

OPTIMISATION WITH VITIS HLS GUI - SUMMARY AND REPORTS

C synthesis report allows for the resource and timing information to be evaluated for the accelerated algorithm

- Values are estimated and are not fully accurate
- Useful in giving an idea of which step of the process can be further optimised

Detailed breakdown of resources consumption, timing etc… can be viewed for each module/function of the accelerated algorithm

- Resource usage estimates can be exported and stored for further study