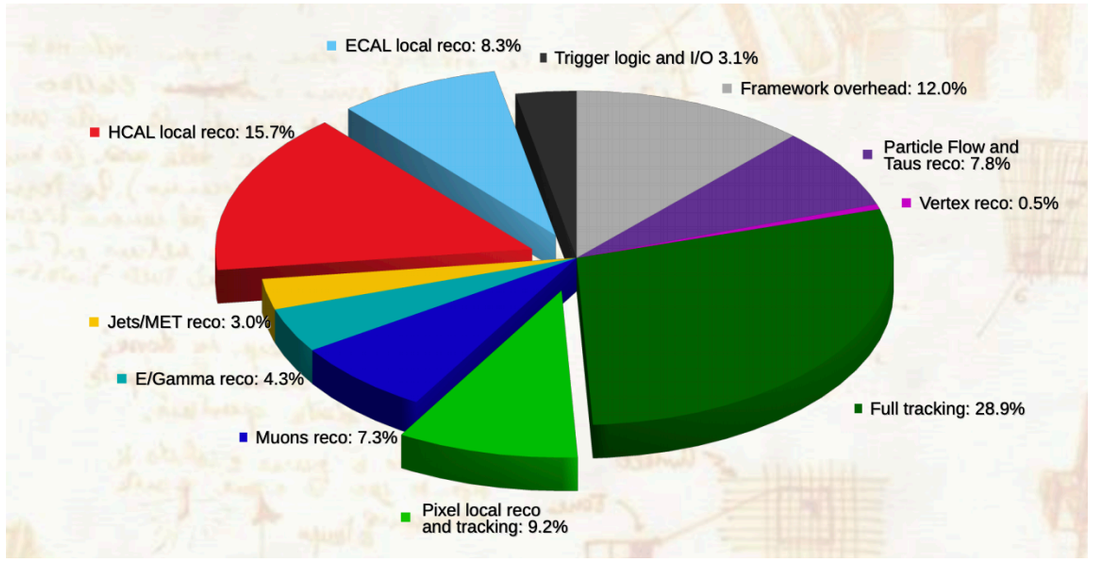
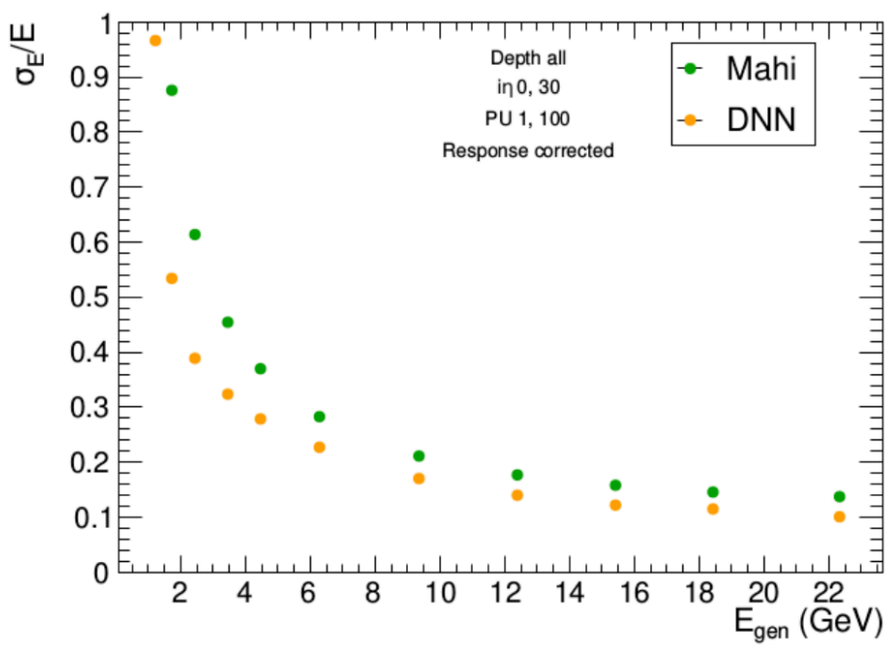
IRIS-HEP Review

P. Harris, D.Rankin
J. Krupa(not here) & FML team

Core Project

- Our main goal : get FPGAs setup to run in LHC workflows
  - Project has a large scope + other funding (HDR/CSSI)
  - Within IRIS-HEP focus has been on Hcal DNN algorithm
    - Despite title not a GNN (not sure how/why the label)
- Work has been done to implement Hcal DNN on FPGAs



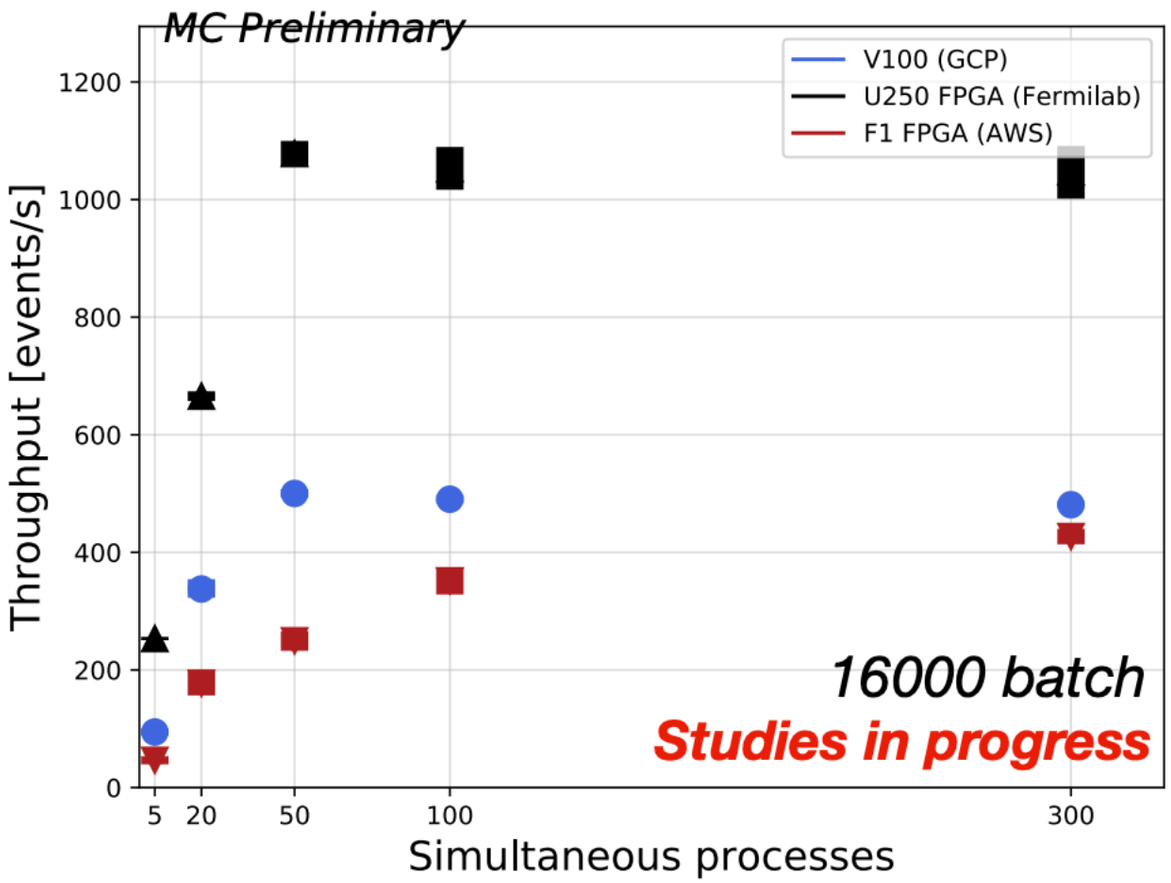


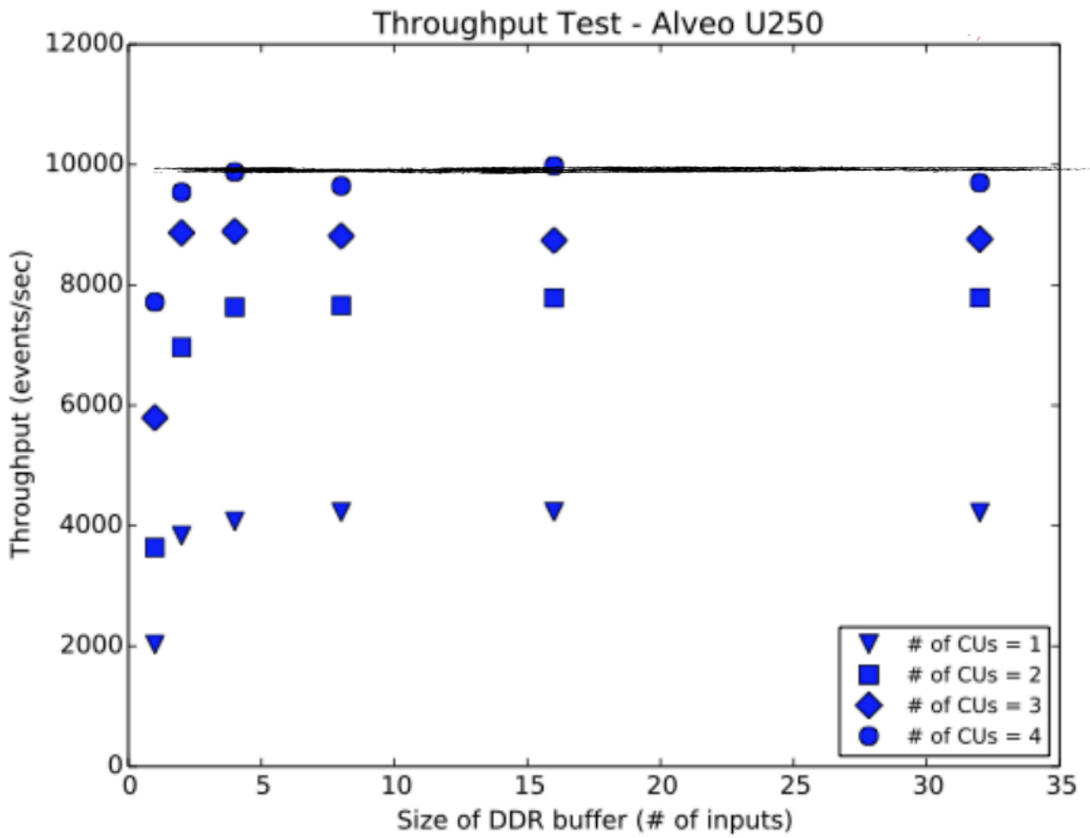
Related Projects

- Co-founder of Fast Machine Learning project
  - HLS4ML and Co-processor studies : 40+ members across domains
- Other funding:
  - NSF Exploring clouds to accelerate science: grant ended
  - NSF CSSI (with LIGO) : developing ML tools for ligo
  - NSF Harnessing Data Revolution (with Mark+LIGO+UW)
    - Work is synergistic with IRIS-HEP but not overlapping
  - USCMS computing operations budget (New!)

Highlights

- Dylan should present soon at an IRIS-HEP topical
  - We have it working on an FPGA in HLT

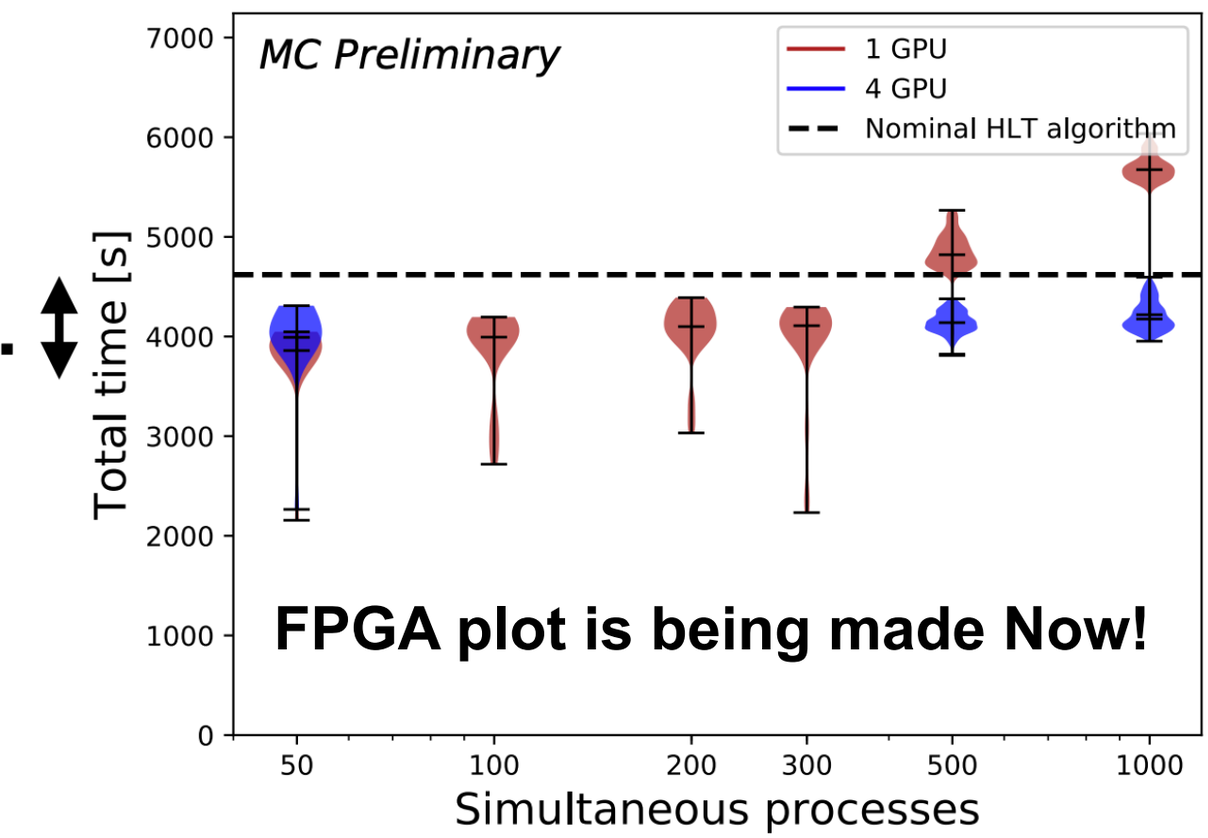


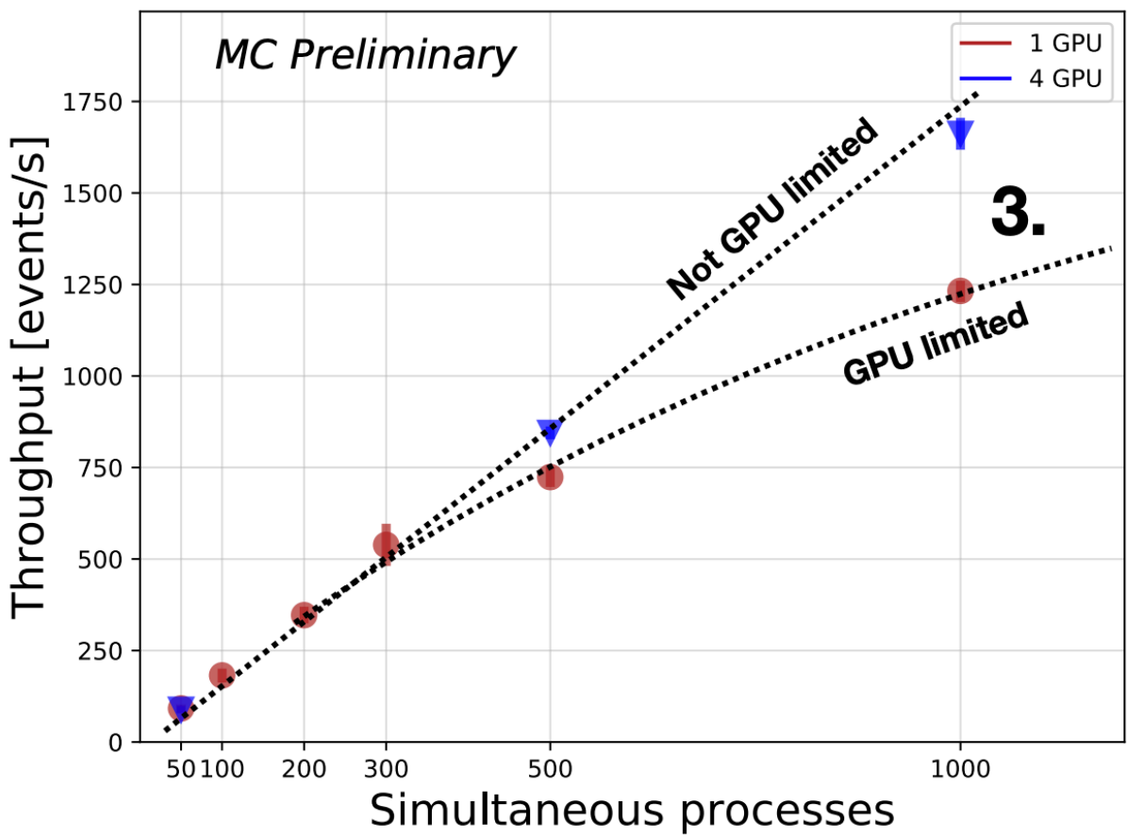


10 kHz on a single FPGA

What it is doing?

- One FPGA can run 10k events/s
- 10 FPGAs can run the full HLT
- This reduces the HLT size by 10-15% (5k cores)
- GPU can run 500 events/s





Time

Current

Input

Algo 1

Output

Thread

Algo 2

Event

Input

Algo 1

Output

Thread

Algo 2

Event

Processor as-a-Service

Input

Algo 1

Output

Thread

Input

Event

Algo 1

Output

Event

Reduced Latency

GPU or FPGA

Algo 2

Algo 2

.....

Other machines

Goals for 2 years

- Done with Mark Neubauer
- Identify specific use cases for FPGA based use cases in HLT or otherwise
  - For MIT, we found the Hcal algo (a simple example)
- Asses performance relative to GPUs/CPUs
  - See previous slide
- Release software to the HEP community w/HLS4ML
  - See previous slide also our SONIC infrastructure is in a PR for CMSSW
- Hold community workshop :
  - IRIS-HEP Blueprint https://indico.cern.ch/event/822126/

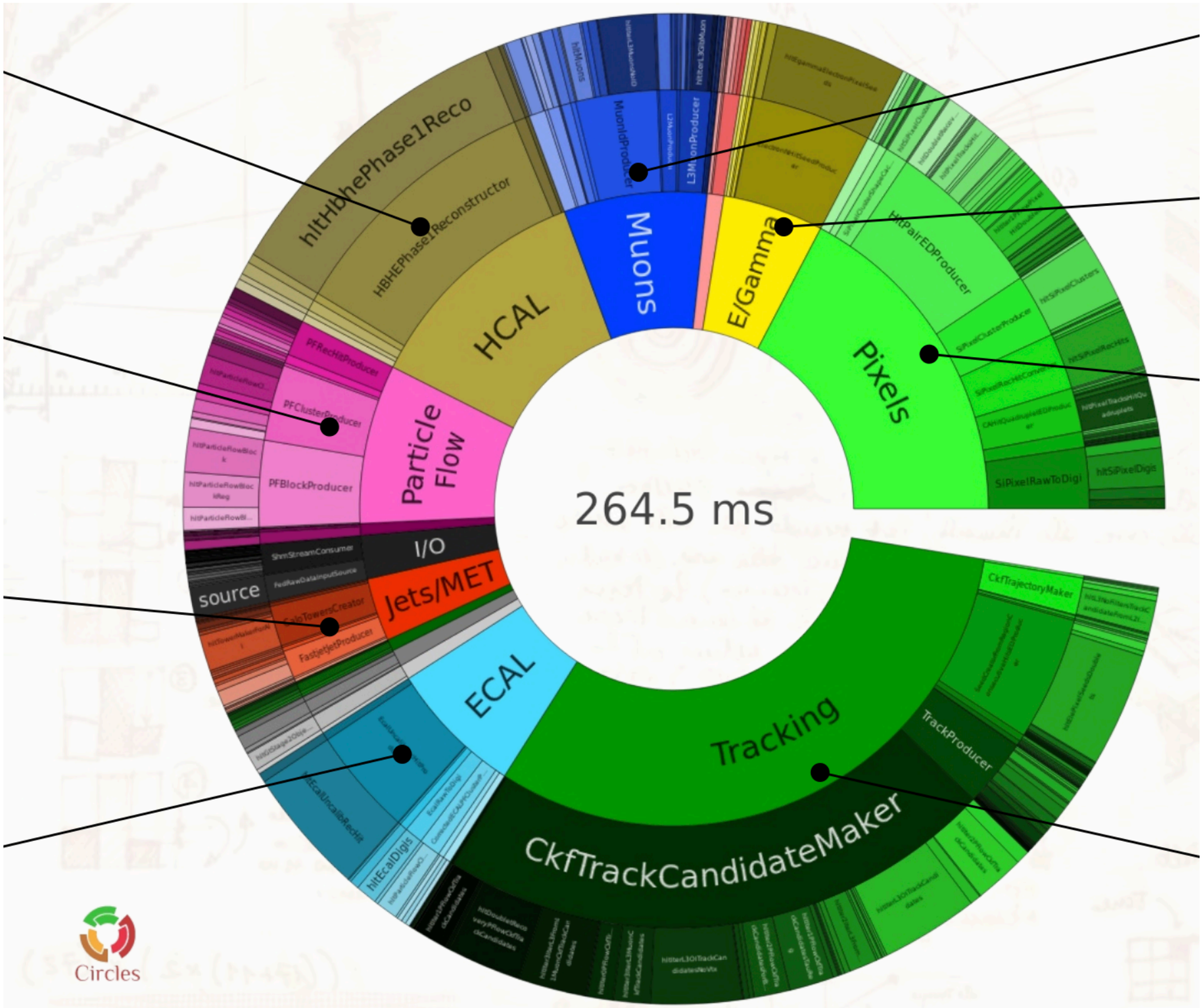
Year 3 Goals

- Aim to demo completely within the HLT (GPU/ FPGA/CPU)
  - Running at CMS with all or fraction of HLT (HW dependt)
  - Integration with a number of parallel projects (not in IRIS-HEP)
- Working to expand this work to include Hcal clustering+depth
  - Investigating Graphs/PVCNN/…models
  - Additionally aiming to expand this to include ecal
  - Same infrastructure can be used, but needs development
- Focus in IRIS-HEP is developing model and running on an FPGA

Challenges?

- Building the infrastructure to run FPGAs in HLT/Offline
- Parallel effort in GPUs also being performed
- A good challenge : full reco or HLT with X% on GPU/FPGA
  - We could also do this in the context of analysis
  - Strategy involves scheduling etc… with CPUs + others
- Additionally this could be tied to 10 PB challenge
- Obtain concrete numbers for level of improvement & scaling needs

Thanks!


hltHbheP
HBHEPhase1Reconstructor
HCAL
hltMuons
MuonIdProducer
L3MuonProducer
Muons
E/Gamma
Pixels
HltPairEDProducer
hltSiPixelClusters
SiPixelClusterProduc
hltSiPixelRecHits
SiPixelRecHitConver
SiPixelRawToDigi
hltSiPixelDigis
PFRecHitProducer
PFClusterProducer
PFBlockProducer
Particle Flow
ShmStreamConsumer
FedRawDataInputSource
CaloTowersCreator
FastjetJetProducer
I/O
Jets/MET
source
hltEcalUncalibRecHit
hltEcalDigis
ECAL
CkfTrajectoryMaker
TrackProducer
Tracking
CkfTrackCandidateMaker
264.5 ms
Circles