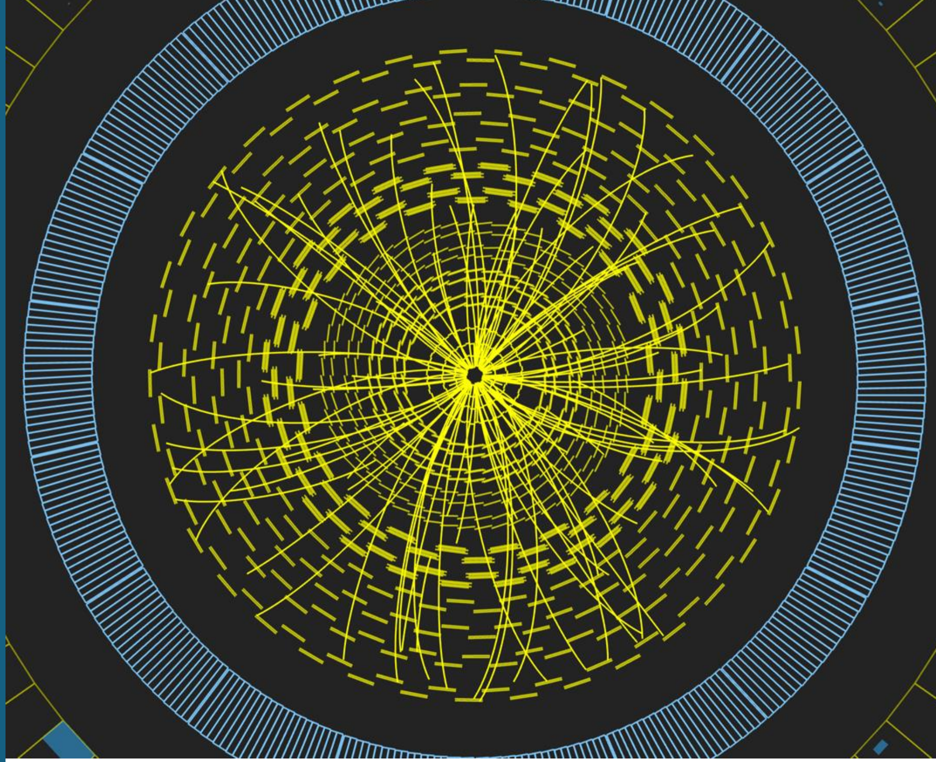
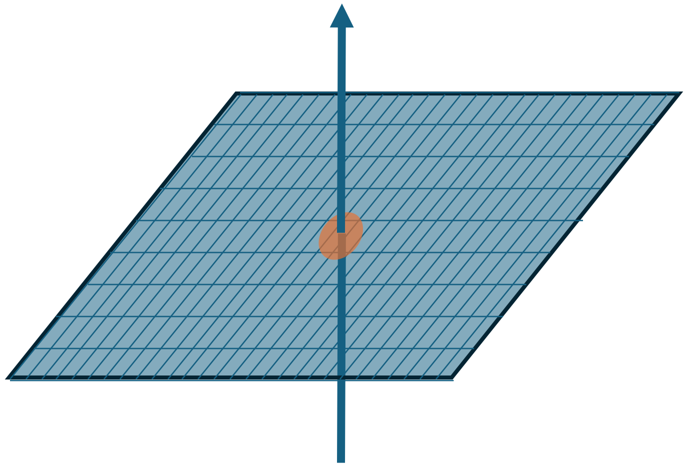
Smartpixels with data reduction at the source

Jannicke Pearkes, University of Colorado
On behalf of the **smartpixels collaboration**

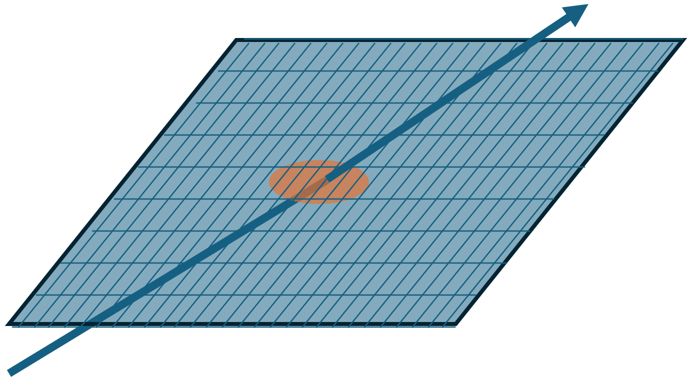
NGT workshop: hls4ml HEP Community Forum
September 27th, 2024



Concept behind smartpixels



Charged particle path perpendicular to sensor:
Regular charge cluster shape

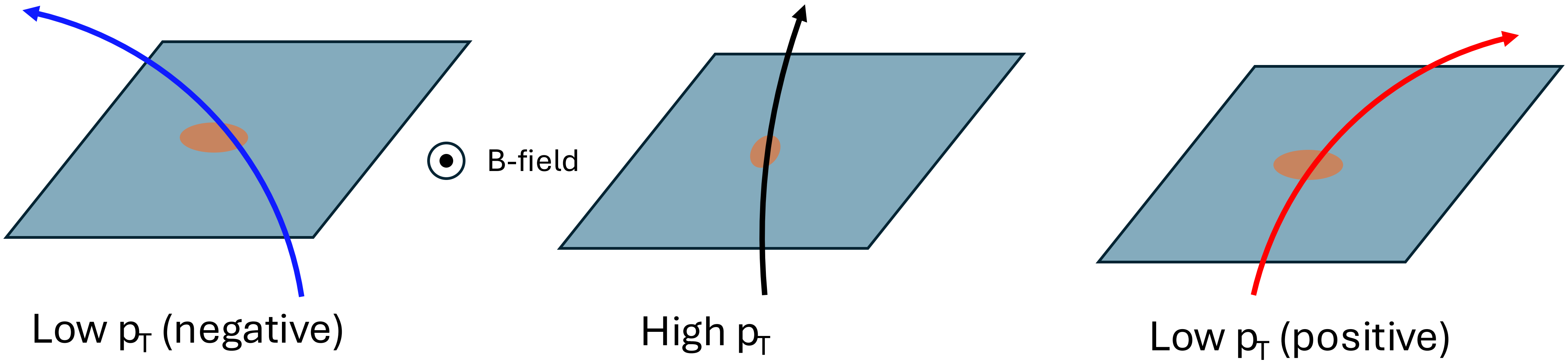


Charged particle path at angle to sensor:
Smeared charge cluster shape

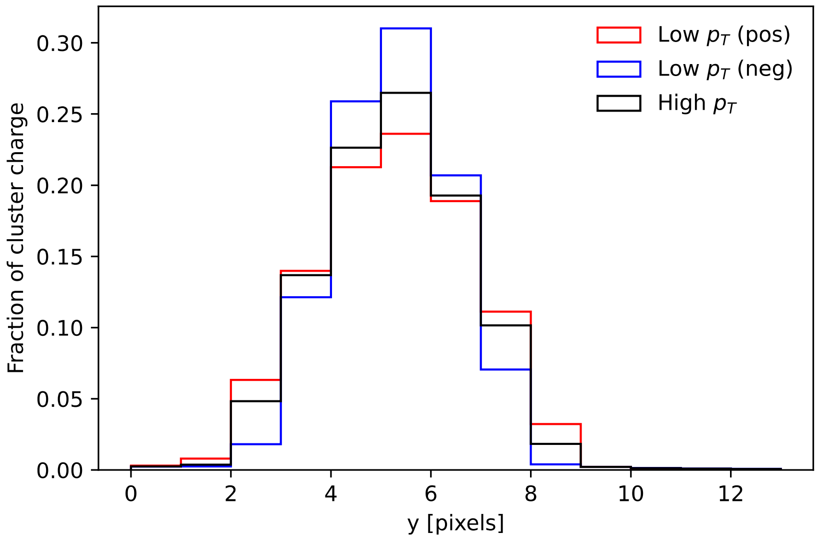
Use cluster shape to extract incident angle of particle traversing pixel sensors

Concept behind smartpixels

Yoo et al 2024 Mach. Learn.: Sci. Technol. 5 035047

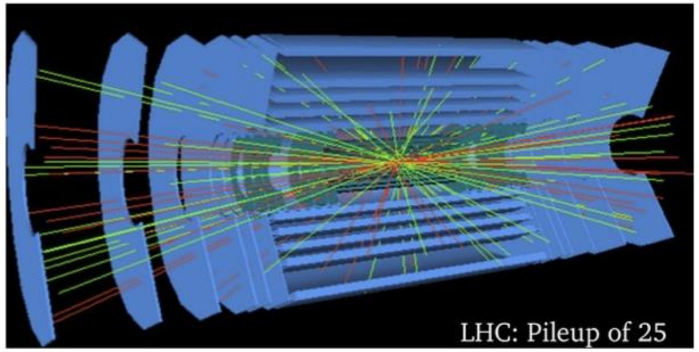


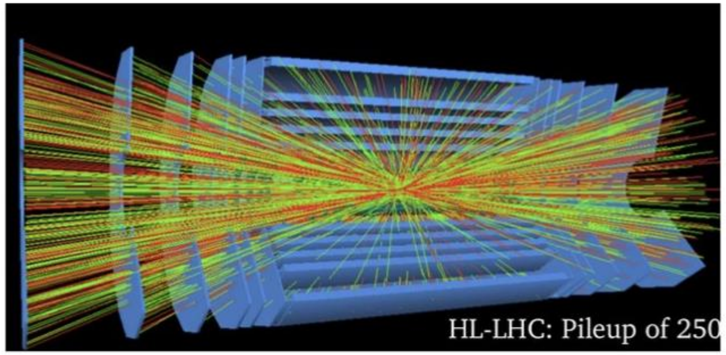
- Can use a locally customized neural network in sensor readout to distinguish low p_T from high p_T charged particles
- Lorentz drift shifts cluster charge distribution



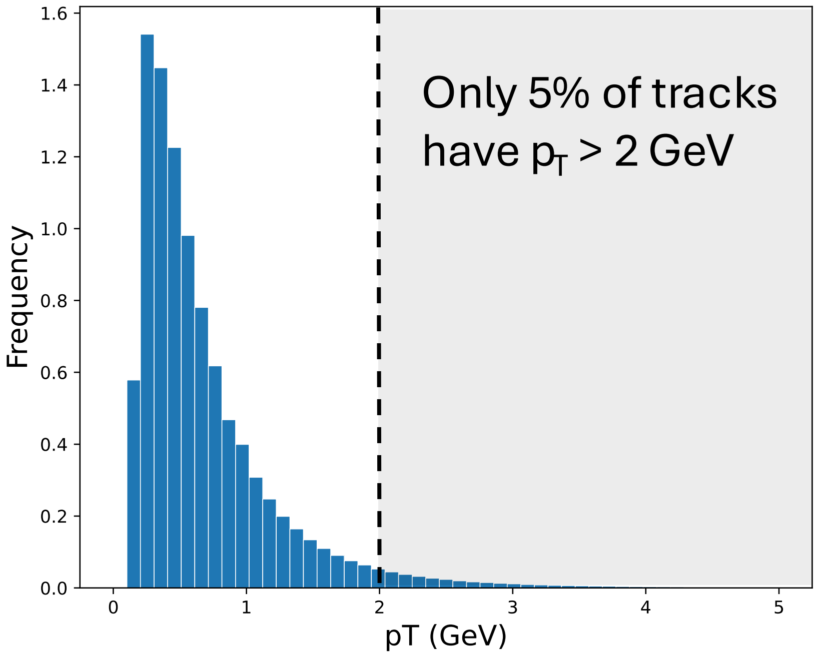
Why smartpixels?

- High-luminosity LHC is going to result in unprecedented data rates especially in the tracking detectors





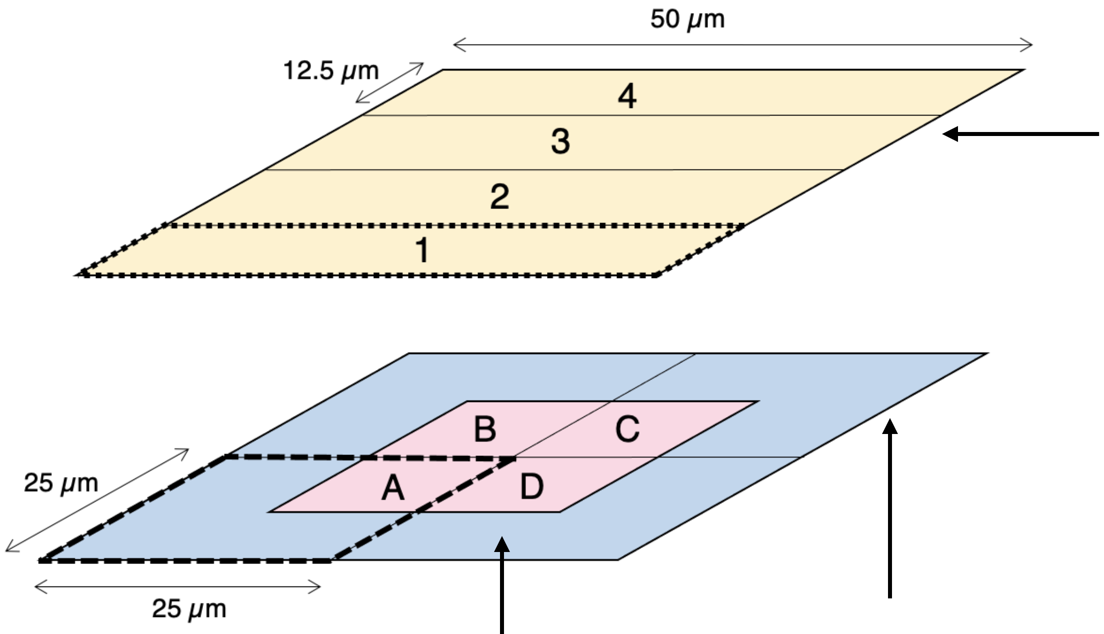
- **By filtering on track momentum with smartpixels we could reduce the data volume at the source,** lowering both rates and power consumption



Momentum distribution of tracks reconstructed by the CMS detector during Run 2 data taking 2310.02474

Smartpixels design

Yoo et al 2024 Mach. Learn.: Sci. Technol. 5 035047



sensor pixels
- 100 μm thick single layer of silicon
- 12.5 x 50 μm^2 pixel size

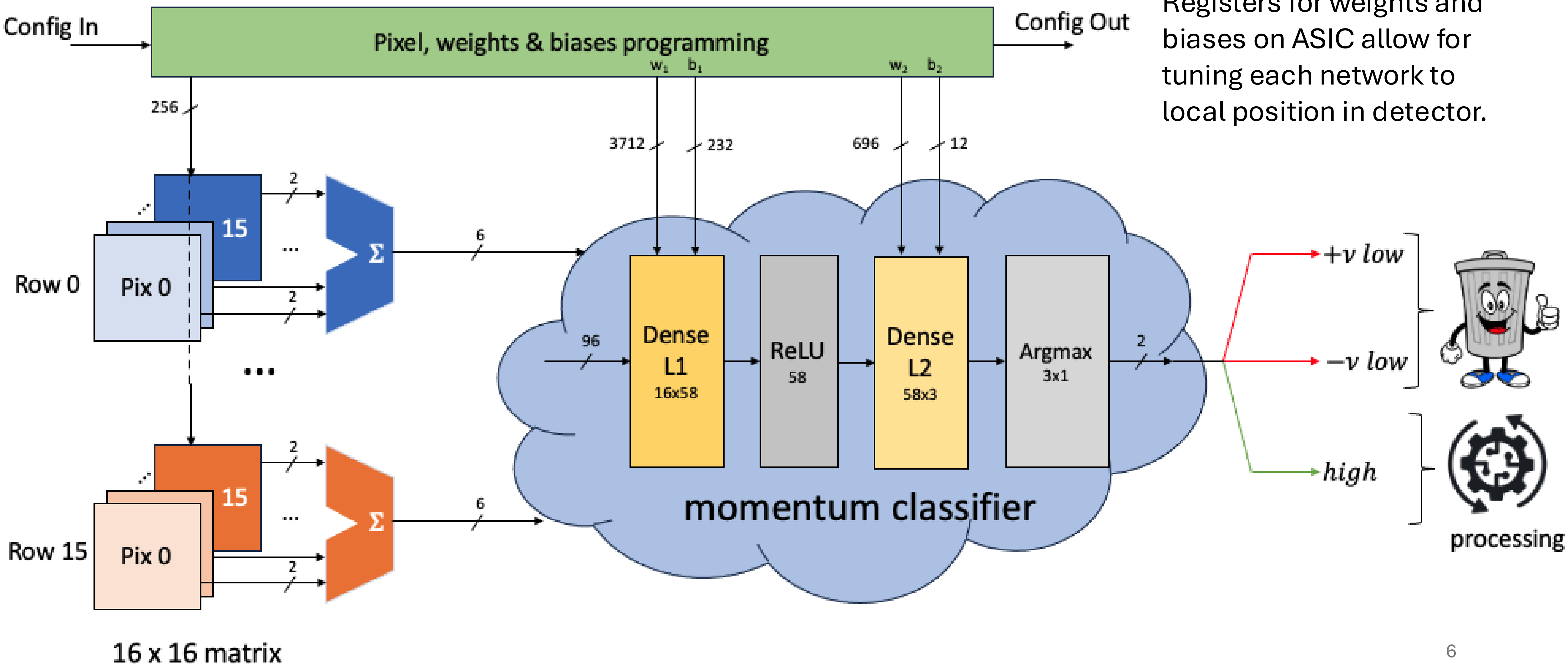
digital logic with ADC inputs and locally customized NN for filtering based on p_T

analog charge amplification and autozero comparators for ADC

Neural network for p_T filter

Image: Benjamin Parpillon

Registers for weights and biases on ASIC allow for tuning each network to local position in detector.



How well does it work?

Yoo et al 2024 Mach. Learn.: Sci. Technol. 5 035047

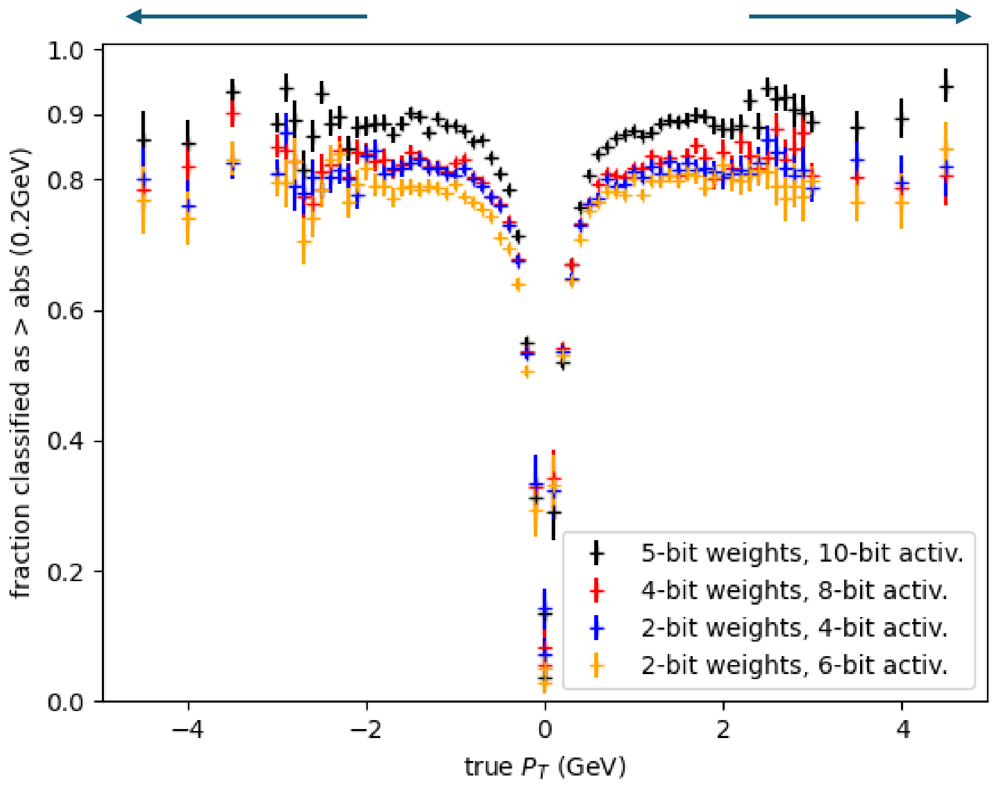
- Tested multiple network architectures and quantization options
- **Can reduce data rate by 54.4% - 75.4%**
- Expected power consumption 300 μW/cm^2
- Expected latency 3.9ns

Simulation details:

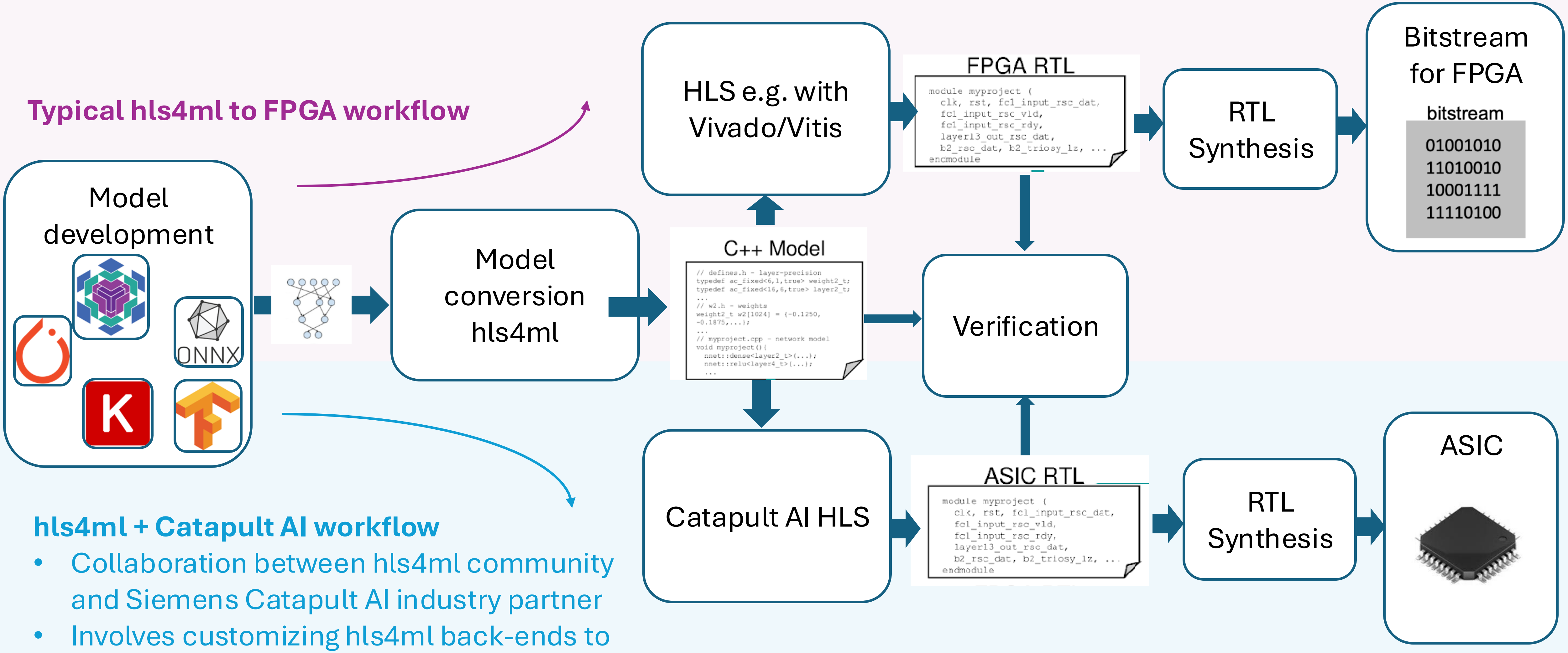
- Tracked data taken from CMS with p_T up to ~5 GeV
- Untracked data not included and includes CMS acceptances
- PixelAV simulation of silicon sensor
- Sensor placed at r=30mm in 3.8T magnetic field
- Single 100μm thick layer of silicon with 12.5x50 μm^2 pixels
- Overall sensor area 16x16 mm^2
- Bias voltage of -100 V

Model	Sig. efficiency	Bkg. rejection
Full precision	93.3 %	25.1 %
Quantized inputs	88.8 %	25.8 %
Quantized weights & inputs	87.3 %	28.2 %

flat efficiency for track p_T > 2 GeV



Integration of hls4ml and Catapult AI



hls4ml + Catapult AI workflow

- Collaboration between hls4ml community and Siemens Catapult AI industry partner
- Involves customizing hls4ml back-ends to meet Catapult AI specifications

Chip tape-out and testing:

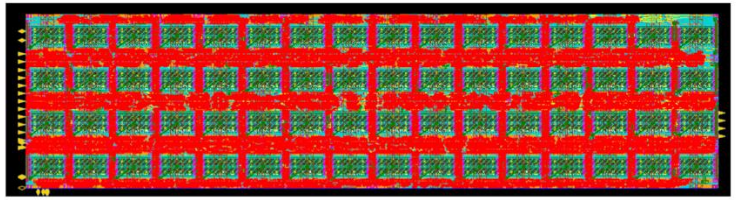
- Prototype 1.5mm^2 ASIC with momentum filtering NN in 28nm CMOS has been fabricated*
- Tests of bare chip currently in progress

Next steps after testing protype:

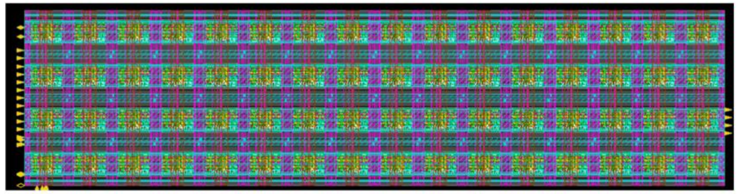
- Build a bigger chip to bump bond to sensor & test in a testbeam

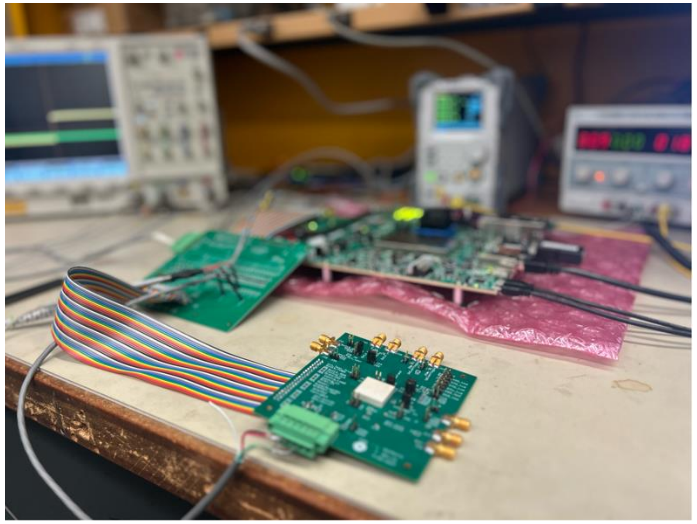
https://arxiv.org/abs/2406.14860

Red = classifier algorithm



Floorplan with analog pixels with power and bias grid

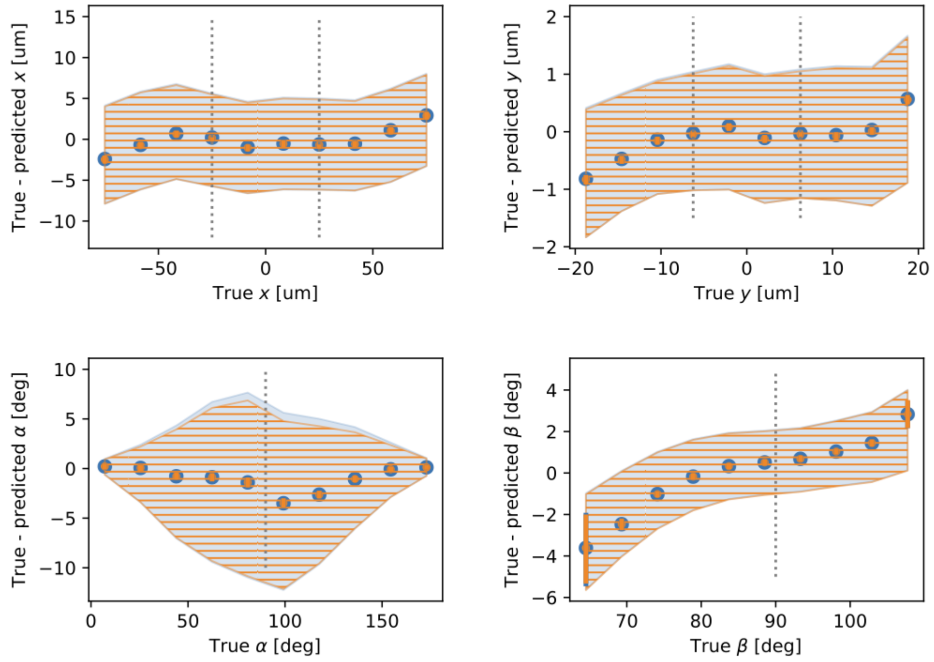




*for characterizations of radiation hardness of 28nm, see G. Borghello, TWEPP 2023

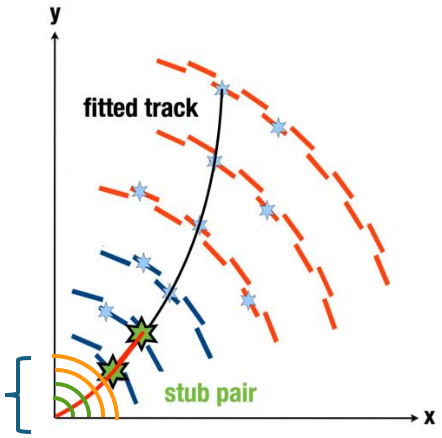
Further developments

https://arxiv.org/abs/2312.11676



- Regression of position, angles and associated uncertainties of charged particle track with mixture density networks
- Regression could further reduce data volume by compressing pixel hits into salient physics quantities
- Combination of regression + momentum filtering could be used to include inner tracker in CMS L1 track trigger, with standard pixel hits being read out in parallel given an L1 accept
- Studies examining technical feasibility and physics outcomes of integrating within CMS L1 track trigger for Phase III (~2035)
- Applications in future colliders, e.g. rejecting beam-induced-backgrounds at muon colliders

Pixel layers unused in CMS L1 track trigger upgrade



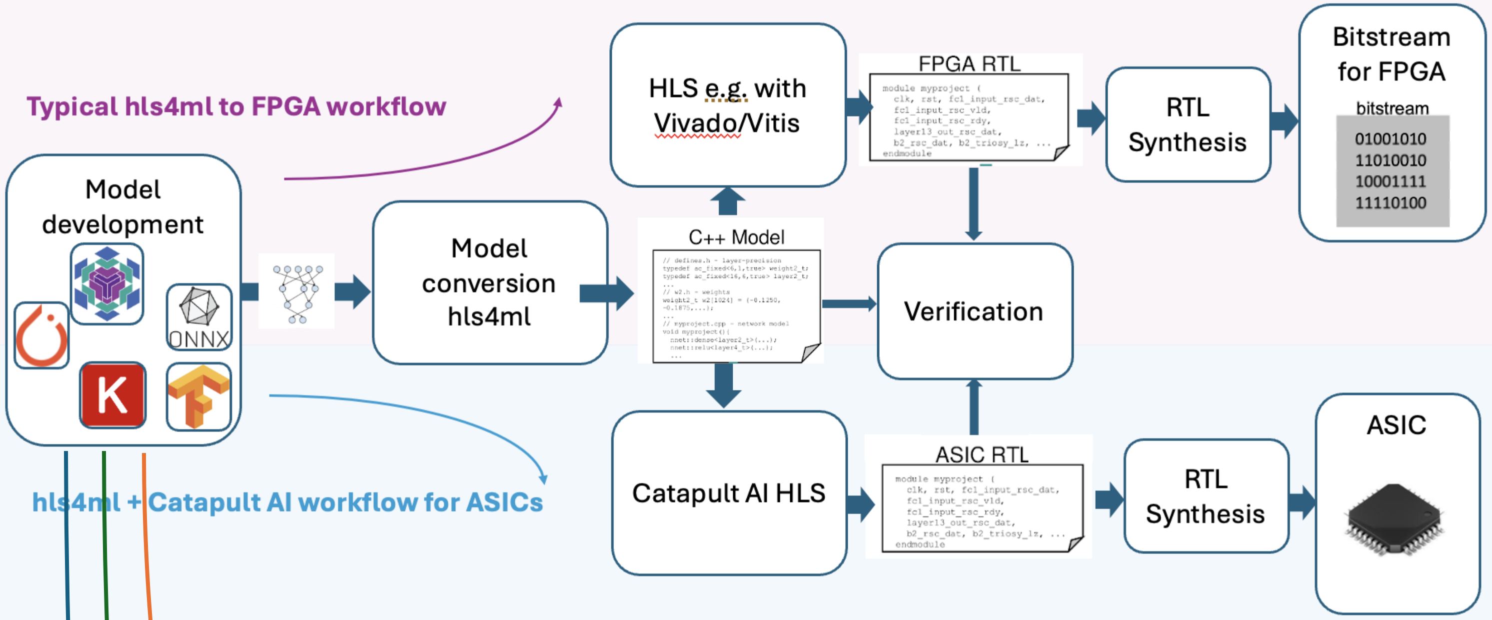
What we want to see from the hls4ml community:

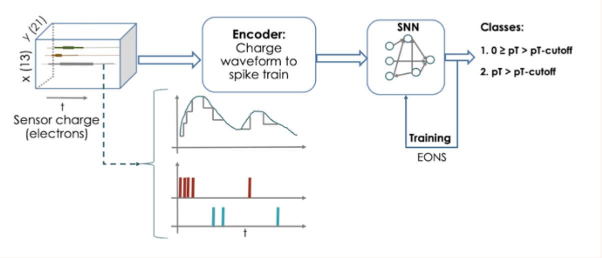
- Continued collaboration with Siemens + Catapult AI to keep architectures supported by hls4ml also supported in Catapult AI backend
- hls4ml on ASICs opens up space for larger range of "Smart detectors" applications
  e.g ML compressed readouts for high granularity or dual readout calorimeters

Bold indicates tested in Catapult AI Workflow so far, typically on a per-project basis

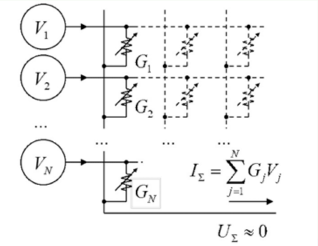
NN Layers	Activation Functions	Pooling/Padding/Reshaping
Conv1D, Conv2D	ELU	AveragePooling1D, AveragePooling2D
SeparableConv1D, SeparableConv2D	**LeakyReLU**	**MaxPooling1D, MaxPooling2D**
BatchNormalization	PReLU	UpSampling1D, UpSampling2D
Dense	**ReLU**	**ZeroPadding1D, ZeroPadding2D**
DepthwiseConv1D, DepthwiseConv2D	**Softmax**	**Resize**
PointwiseConv1D, PointwiseConv2D	TernaryTanh	Transpose
LSTM	ThresholdedReLU	Merge
SimpleRNN		Dot
TernaryDense		Concatenate
		Clone

“Can the hls4ml paradigm be applied across novel beyond CMOS microelectronics back-ends?”

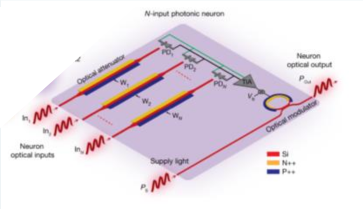




hls4ml for Spiking Neural Networks?
SNNs for smartpixels: https://arxiv.org/abs/2307.11242



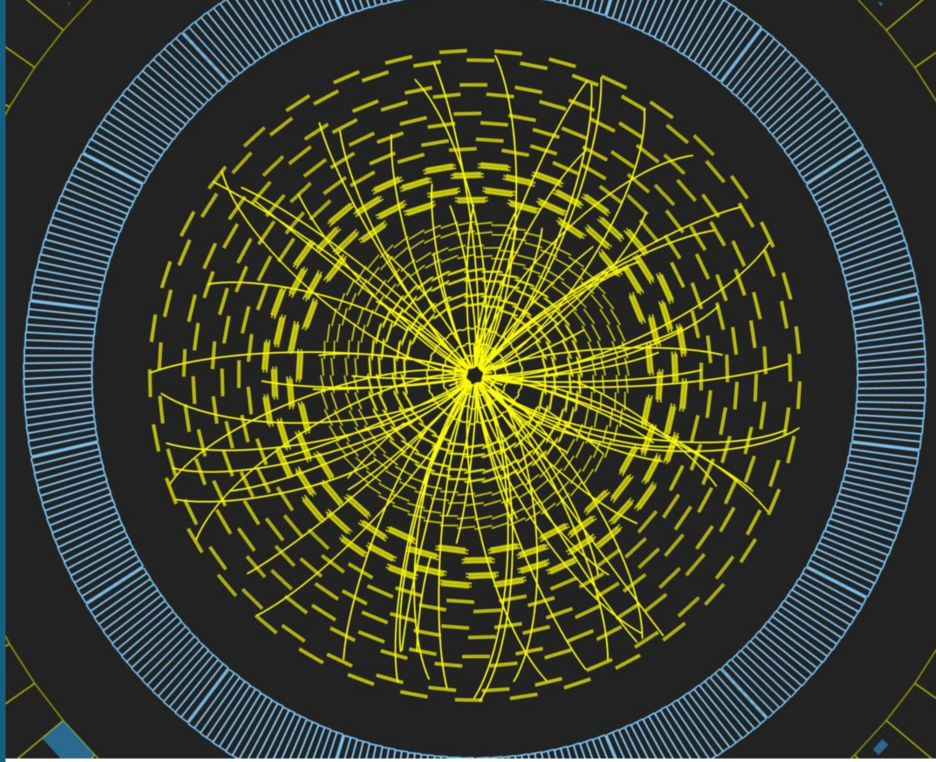
hls4ml for analog circuit-based neural networks?



hls4ml for optical neural networks?
... and many, many more.

See review in [Applications and Techniques for Fast Machine Learning in Science]

Thank you!

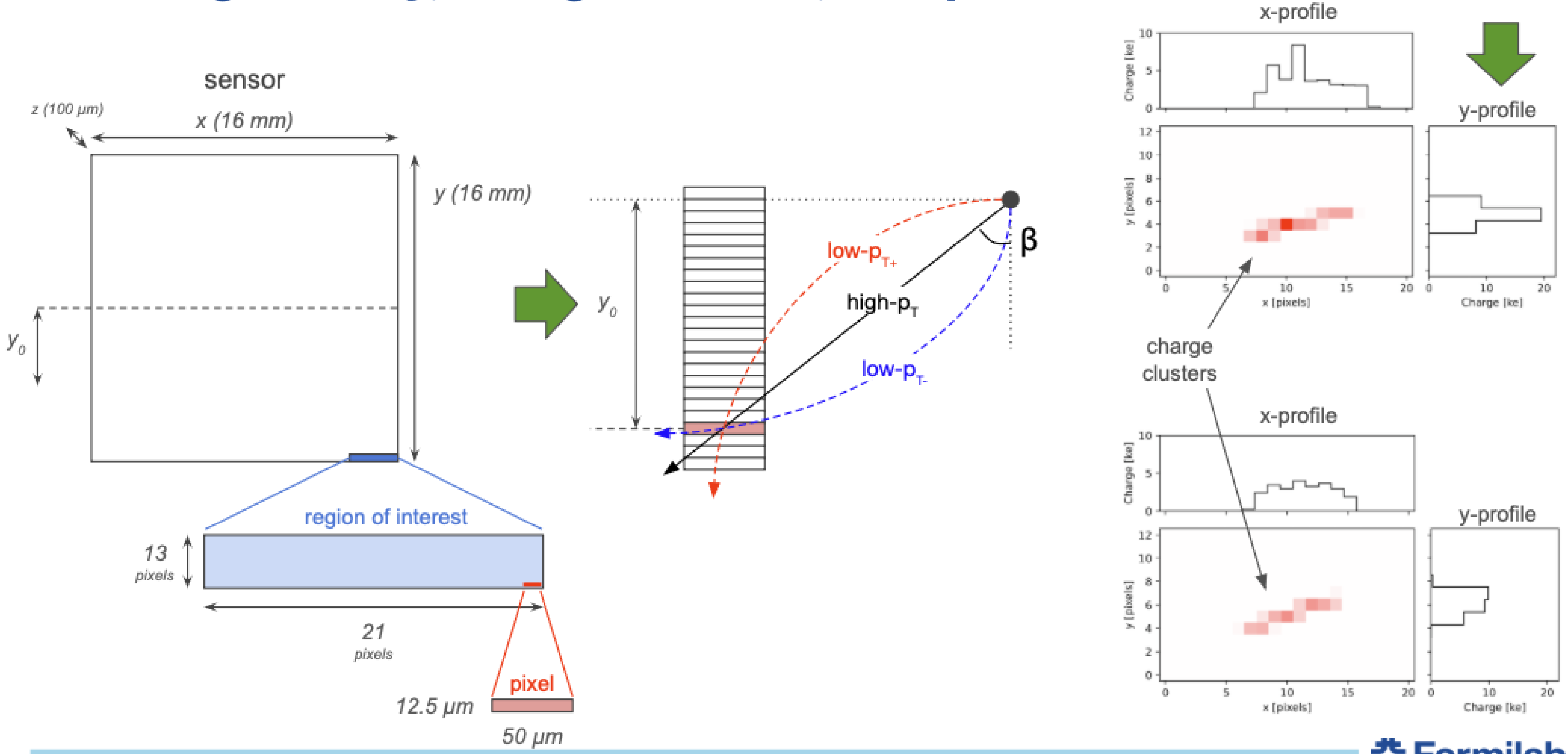


smartpixels

References:

- On-Sensor Data Filtering using Neuromorphic Computing for High Energy Physics Experiments (Jul 2023) https://arxiv.org/abs/2307.11242
- Smart pixel sensors: towards on-sensor filtering of pixel clusters with deep learning (Oct 2023) https://iopscience.iop.org/article/10.1088/2632-2153/ad6a00
- Smartpixels: Towards on-sensor inference of charged particle track parameters and uncertainties (Dec 2023) https://arxiv.org/abs/2312.11676
- Smart Pixels: In-pixel AI for on-sensor data filtering (June 2024) https://arxiv.org/abs/2406.14860
- Siemens Catapult AI + hls4ml press release (May 2024) https://newsroom.sw.siemens.com/en-US/siemens-catapult-ai-nn/
- Jennet Dickinson, LCWS 2023 (May 2023) https://indico.slac.stanford.edu/event/7467/contributions/5966/
- Gisueppe Di Guglielmo, FastML @ ICCAD (Nov 2023) https://fastmachinelearning.org/iccad2023/file/fastml4science_iccad_20231102.pdf
- Gisueppe Di Guglielmo, Siemens User2User (April 2024) https://docs.google.com/presentation/d/1O4VMajkHL81xM0IaN0ZJ3sUANdr0aeQfKN0AmvlnbkE/edit#slide=id.g2c55b12b4ca_0_96
- Anthony Badea ICHEP 2024 https://indico.cern.ch/event/1291157/contributions/5888438/

Sensor geometry, charge clusters, and profiles



https://fastmachinelearning.org/iccad2023/file/fastml4science_iccad_20231102.pdf

RD53 Collaboration, IEEE 10182033

RD53 Benchmarks

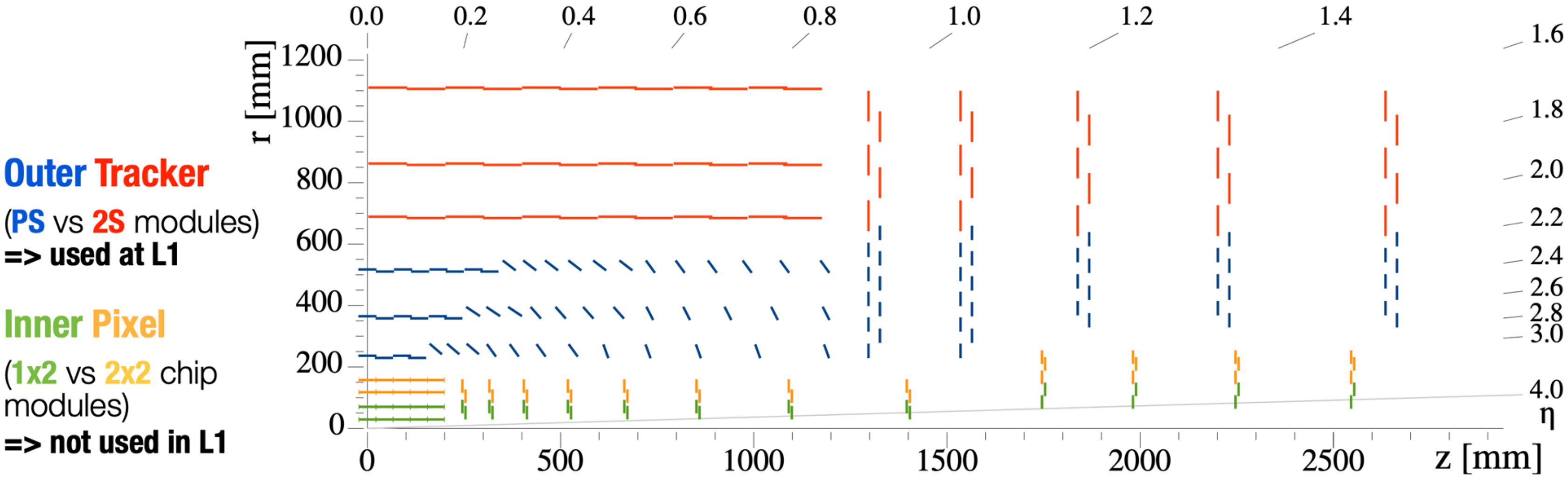
Talk by Flavio Loddo

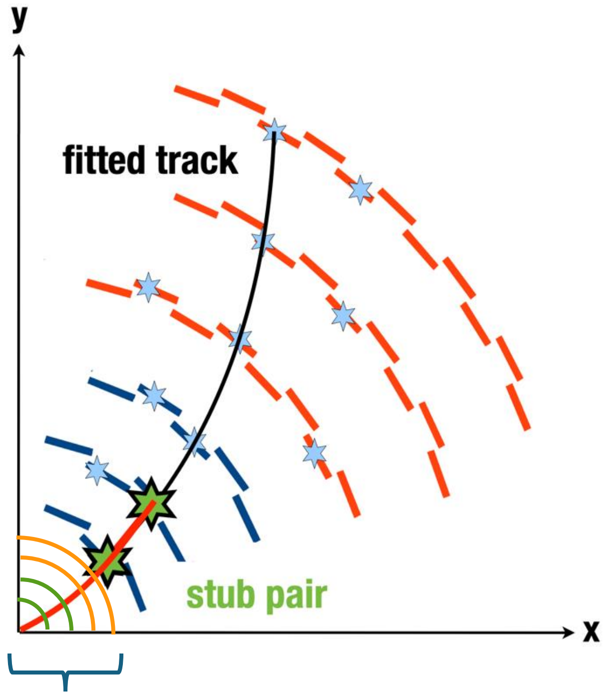
	ATLAS/CMS
Chip size	20x21mm^2/21.6x18.6mm^2
Pixel size	50x50 µm^2
Hit rate	3 GHz/cm^2
Trigger rate	1 MHz/750kHz
Trigger latency	12.5 us
Min. threshold	600 e-
Radiation tolerance	500 Mrad @-15C
Power	< 1W/cm^2

Why smartpixels?

Images adapted from L. Skinnari

HL-LHC L1 track finding does not include hits from the inner tracker





pixel layers unused in CMS L1 trigger upgrade

- **Smartpixels could make L1 triggering with the inner tracker feasible.**
- Better L1 tracking resolutions
- Better L1 vertexing and pile-up rejection
- Better L1 b-tagging (e.g. for targeted HH->4b triggers)