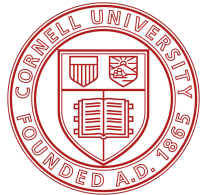


Smartpixels: Intelligent pixel detectors

Towards a radiation hard ASIC with on-chip machine learning in 28nm CMOS

By: **Ben Weiss** on behalf of the Smartpixels group

FastML 2025 – 4 September 2025



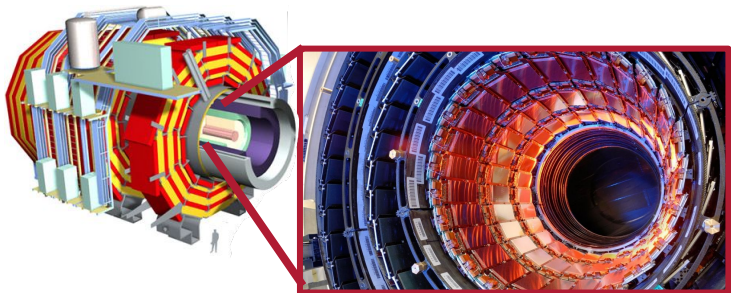
UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN



JOHNS HOPKINS
UNIVERSITY



Data rates at the LHC



Compact Muon Solenoid
(CMS) Pixel detector
~Raw O(TB/s)

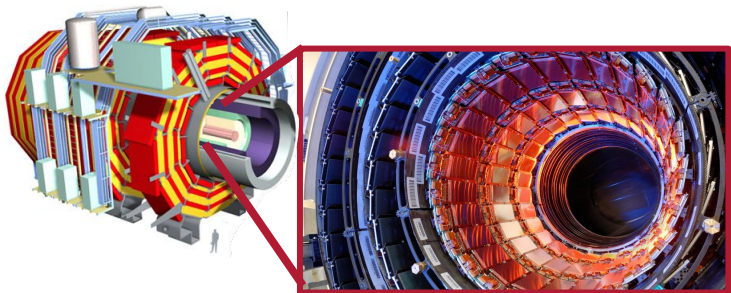


*Very rough estimate, YouTube streams ~60B hrs/year



YouTube's total average
streaming data rate*
~O(TB/s)

Data rates at the LHC



Compact Muon Solenoid
(CMS) Pixel detector
~Raw O(TB/s)



*Very rough estimate, YouTube streams ~60B hrs/year

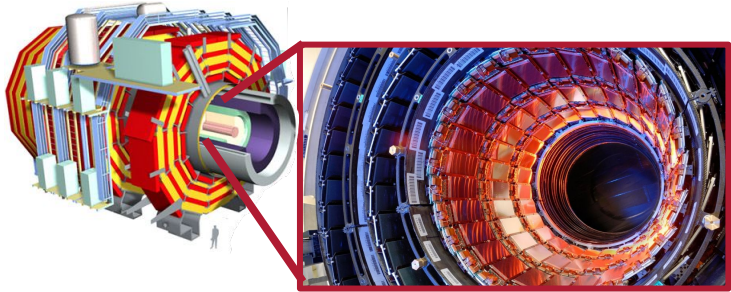


YouTube's total average
streaming data rate*
~O(TB/s)

Triggers



Data rates at the LHC



Compact Muon Solenoid
(CMS) Pixel detector
~Raw O(TB/s)



*Very rough estimate, YouTube streams ~60B hrs/year



YouTube's total average
streaming data rate*
~O(TB/s)

Triggers

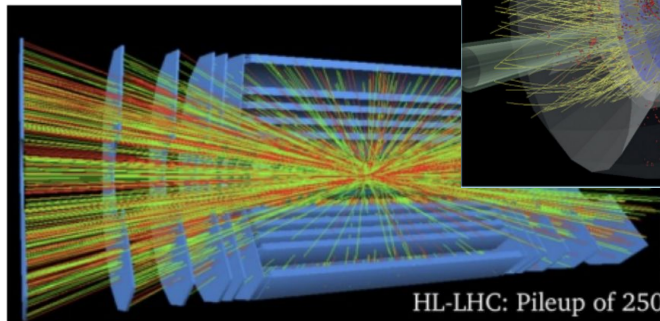


HL LHC: 5x data rate

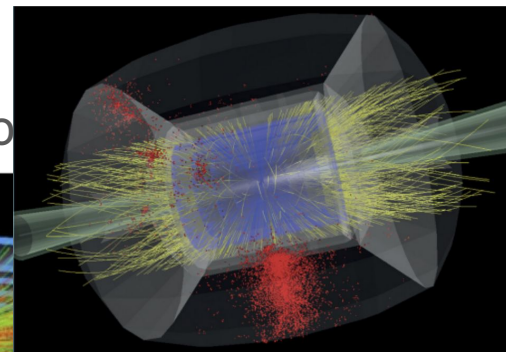
CMS phase 3 & beyond

- Even worse at a future colliders like FCC-hh or a Muon collider (induced background)
- **100 μm x 25 μm \rightarrow 50 μm x 12.5 μm**
Current pitch Future pitch

FCC-hh PU= 4x



Muon Collider:



FCC-hh: The Hadron Collider

Future Circular Collider Conceptual Design Report Volume 3

Clearly, more sophisticated trigger algorithms, like isolation for muons or longitudinal segmentation for the calorimetry are needed to keep the trigger rates at acceptable levels. This essentially means that today's offline algorithms have to be migrated to the trigger.

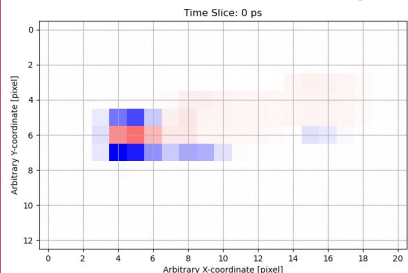
At 4x smaller pitch, how much can we refine readout on-pixel with ML?

Enter Smartpixels!

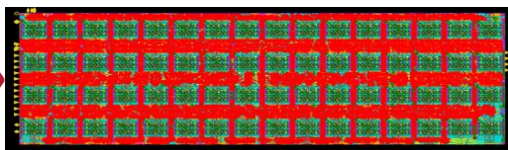
Smartpixels is an on-pixel, ML based, **pre-trigger** data refinement implementation for pixel detector ROCs.

- Filtering and regression of track parameters from a **single** silicon layer
- **Extreme-edge** implementation targeting future collider experiments

Pixel array charge



SmartPix ROC bonded to pixel



Digital NN
Analog readout

Predictions:

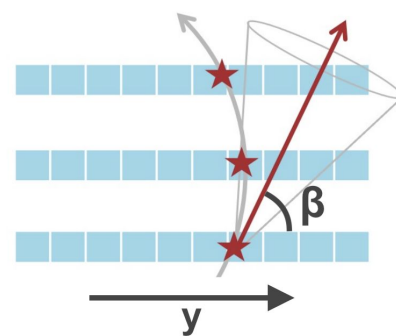
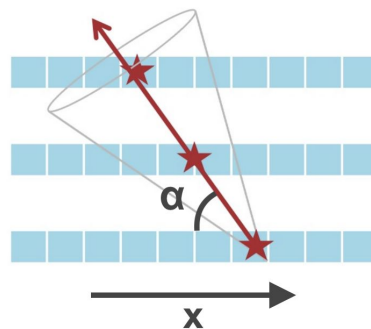
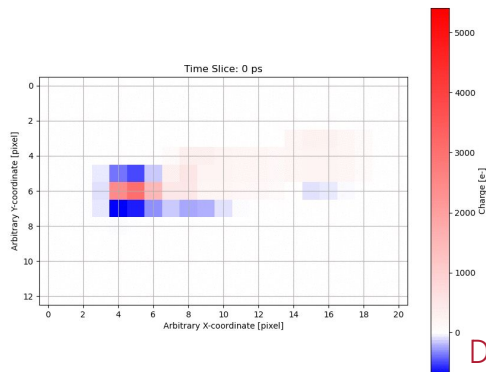
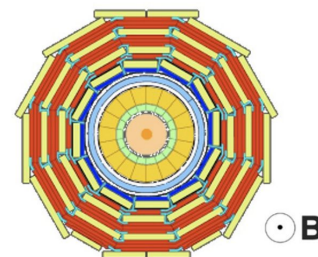
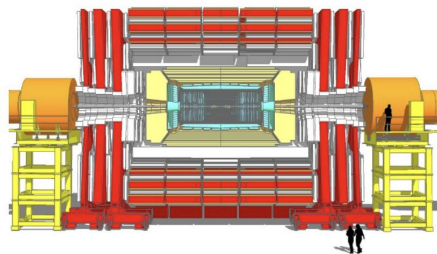
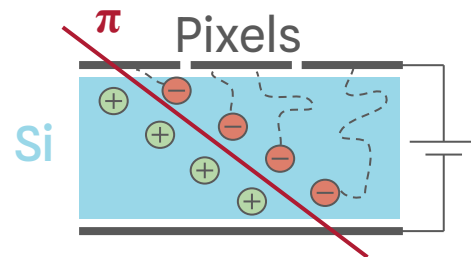
- High vs. low pt?
- Hit position
- Track angle

L1 Trigger
@ 40 MHz

Smartpixels Datasets

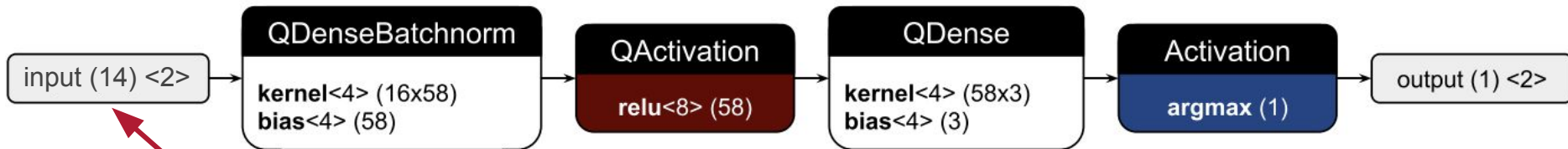
Incident pion kinematics taken from real CMS data.
Sensor charge collection simulated with **PixelAV**.

- 16x16 pixel array
- **50 μm x 12.5 μm** pixel pitch
- 20 timesteps (200ps apart)
- Subset available on [Zenodo](#)
- 500k-4M events/dataset



Dataset production work by Danush Shekar, Morris Swartz

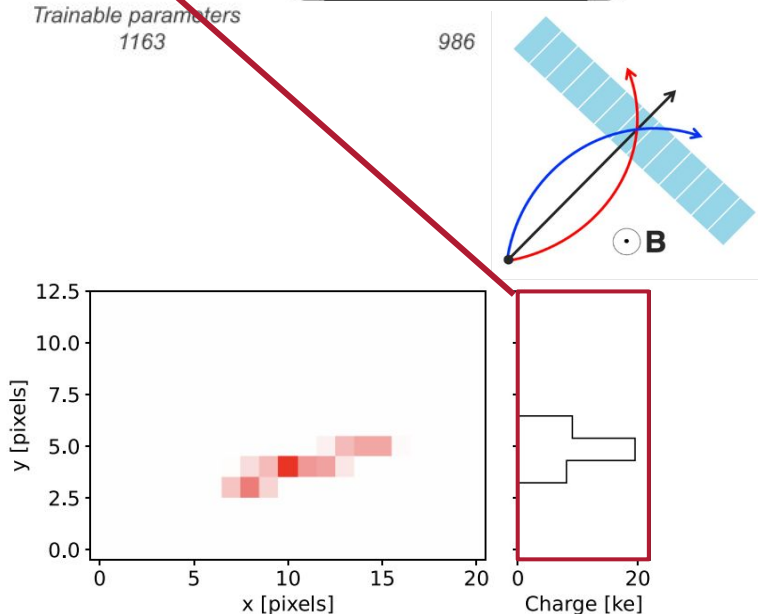
Filtering algorithm



Trainable parameters
1163

986

177



Classifies high vs low momentum particles
with ~ 0.2 GeV threshold

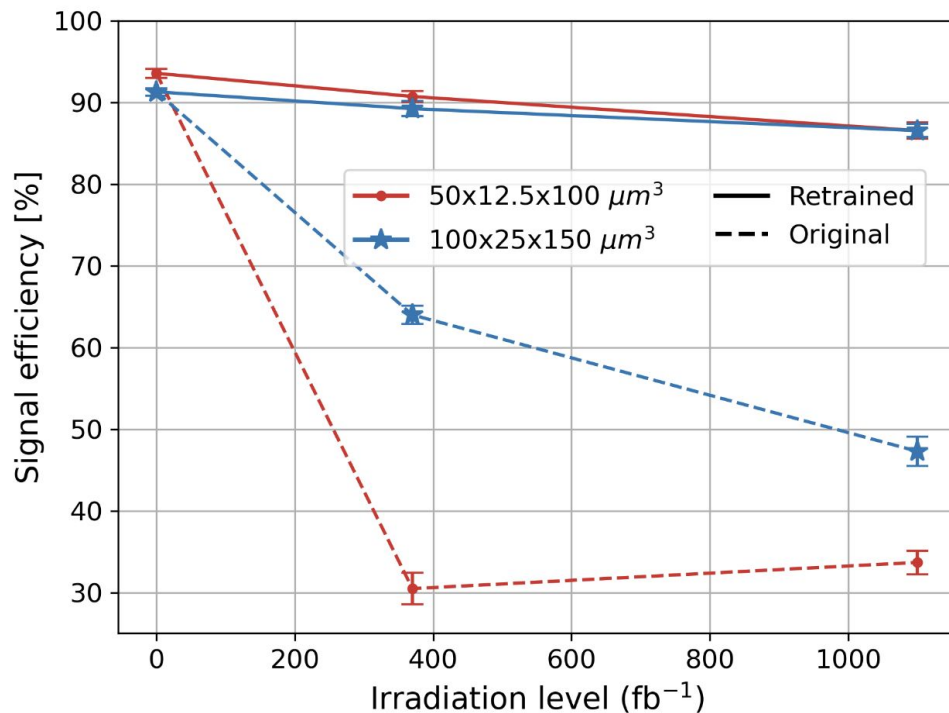
- **13x21** pixel array with **2-bit** quantization on front end charge
- Dense network trained on y-projection of charge cluster images
- 4 Bit quantization of weights & biases

Filtering algorithm performance

Multiple geometries (0.2 GeV thresh.):

Sensor geometry [μm^3]	Bias voltage [V]	Signal efficiency %	Data reduction %
50 X 10 X 100	100	95.4 \pm 0.5	33.1 \pm 1.0
50 X 12.5 X 100	100	93.9 \pm 0.5	33.1 \pm 0.9
50 X 15 X 100	100	93.3 \pm 0.5	30.7 \pm 0.9
50 X 20 X 100	100	91.2 \pm 0.9	28.4 \pm 0.9
50 X 25 X 100	100	88.3 \pm 0.7	27.3 \pm 0.8
100 X 25 X 100	100	88.6 \pm 0.9	26.9 \pm 1.0
100 X 25 X 150	175	91.9 \pm 0.7	29.7 \pm 1.0

Retrainable to endure irradiation:

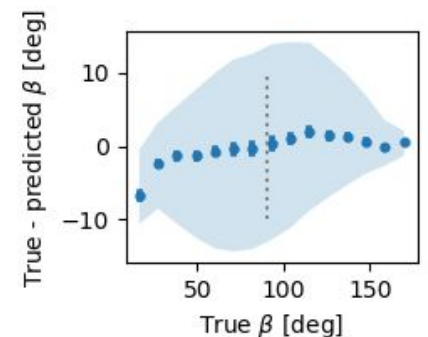
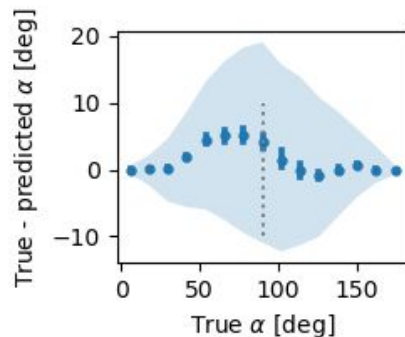
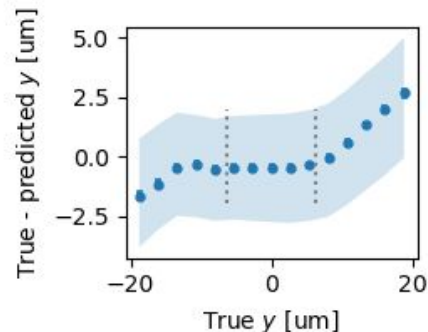
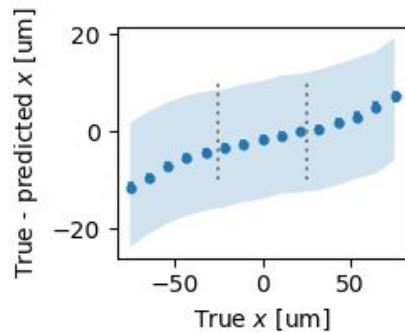


Regression algorithm

Infers particle track parameters (**position & angle**) from a single silicon layer

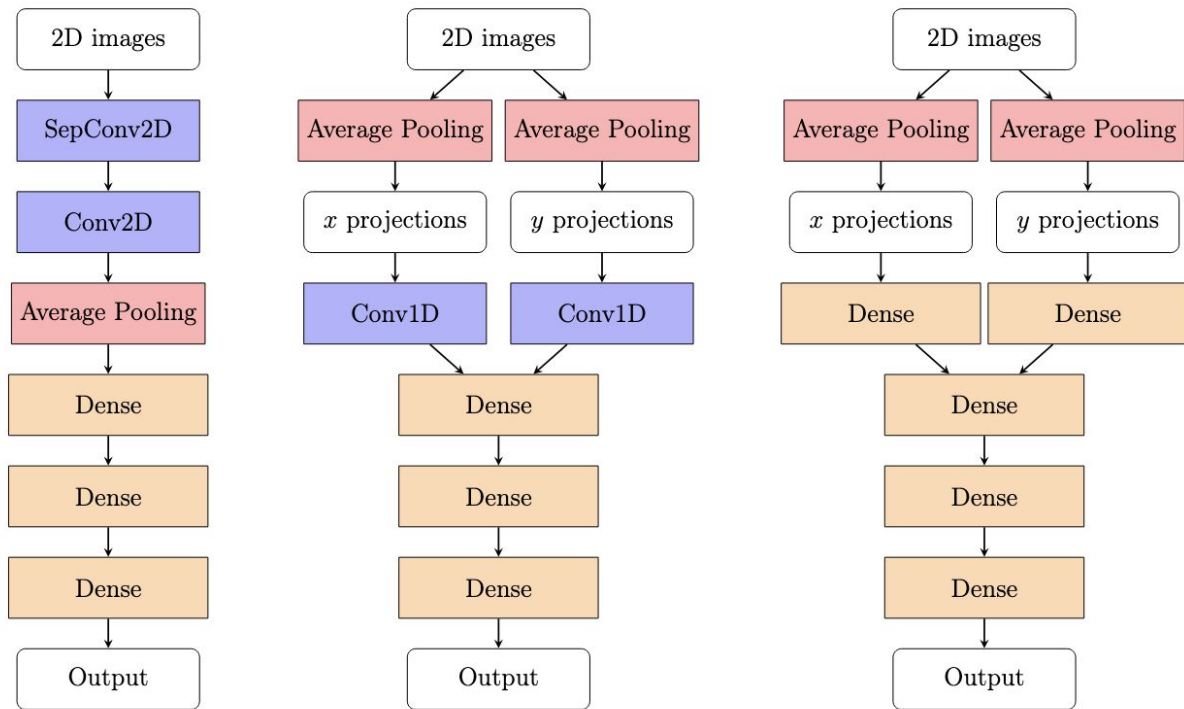
- 2 time slices of 16x16 charge cluster images [2-bit input quantization]
- 2 Convolutional layers with pooling fed into dense network [8 bit network quantization]
- Predicting **variable and uncertainty**

$$\text{Loss} = \frac{1}{4} \sum_{v_i} \frac{(v_i - \mu_i)^2}{\sigma_i} \quad \text{where, } v_i \in \{x, y, \cot(\alpha), \cot(\beta)\}$$



Regression Alg. design space exploration

Reducing # of operations 



Current work to reduce model size for hardware implementation ($<0.2 \text{ mm}^2$)

- On ASIC, generally: **Conv2D > Conv1D > Dense**
- Performance is similar for compressed models!
- Synthesis estimates coming soon...

Regression on a realistic FE

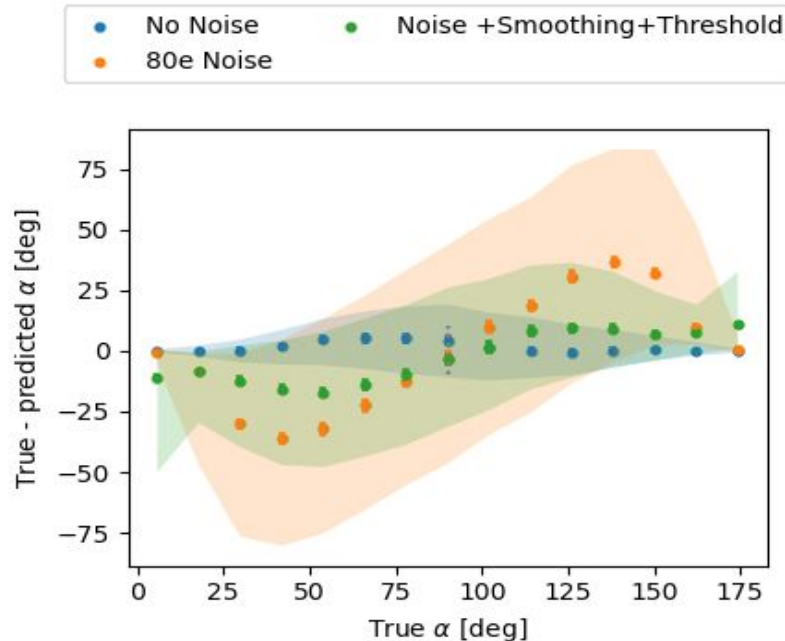
Regression performance degrades significantly with nominal **front end noise**.

Potential solutions:

- 5σ threshold (anticipated)
- Smoothing with average pooling
- *Denoising Autoencoder*.

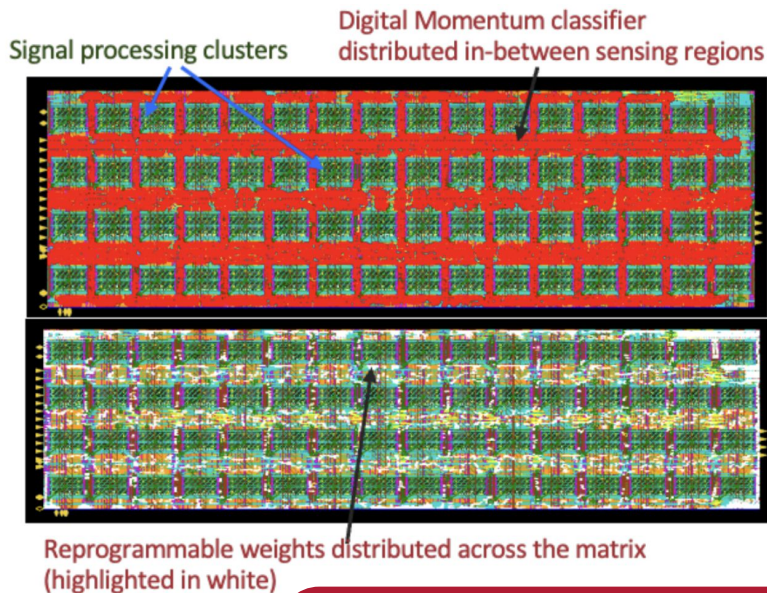
Dovetails with **quantization** optimization:

- Allow keras to optimize charge intervals *during training*



ADC output	Charge interval [e^-]
00	< 400
01	400 – 1600
10	1600 – 2400
11	> 2400

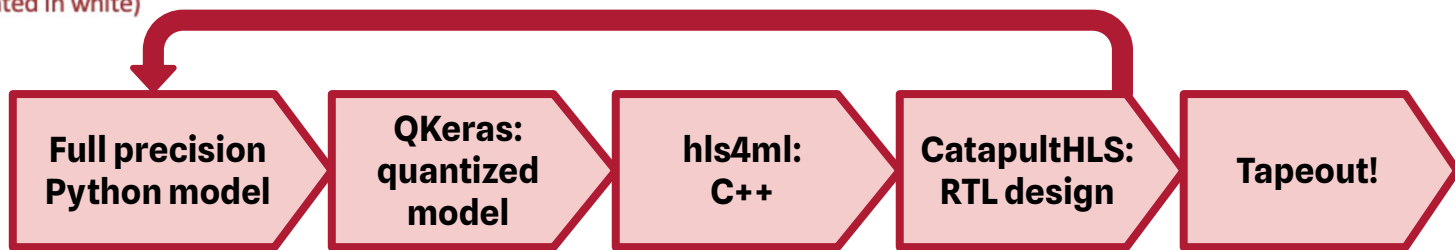
ASIC synthesis



Established pipeline to translate quantized python level model to RTL design

- hls4ml converts QKeras model to C++
- Catapult HLS converts C++ to RTL
- Estimates model footprint (mm^2), power consumption
- Feedback for python optimization

Balancing performance, area, power

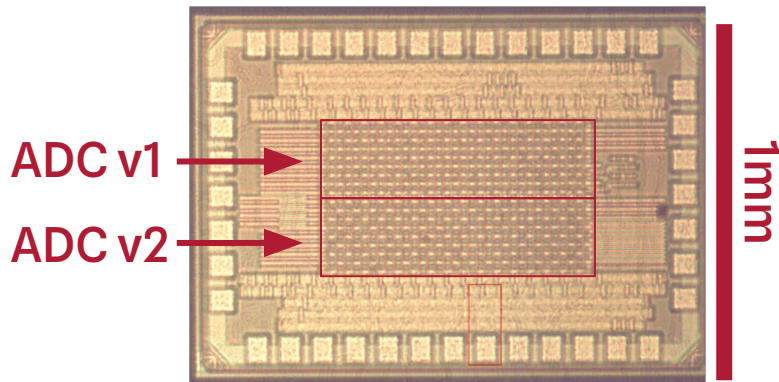
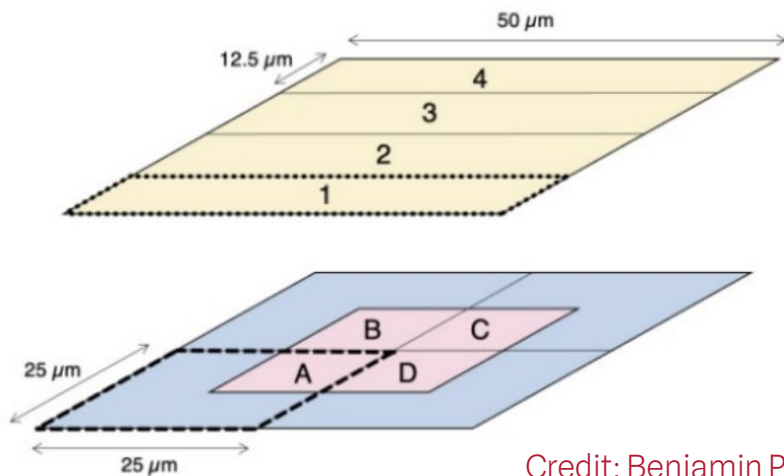


New Prototype SmartPixels ROIC

Newest Smartpixels **TSMC 28 nm** chip hosts per-pixel charge injection and programmable network for **classifier**.

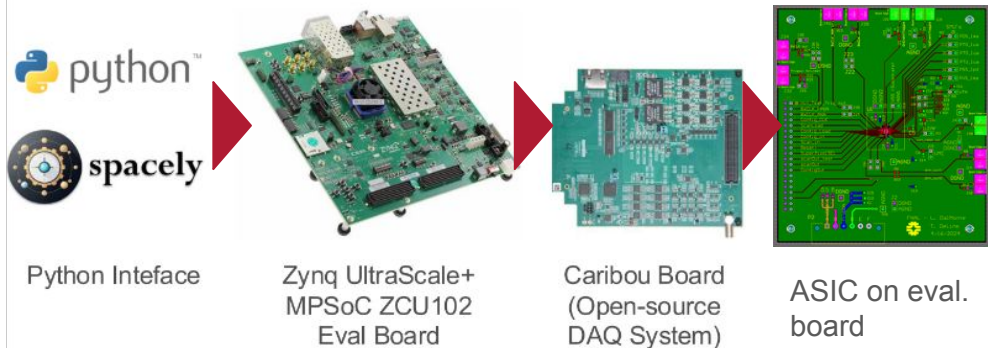
- 2 sets of 8x32 pixel arrays (2 ADC vers.)
- ADC+DNN consumes $\sim 6 \mu\text{W}/\text{pixel}$ @ 40 MS/s

Spec.	Value	Units
Parameters	4652	w+b
DNN+ADC area	~ 0.2	mm^2
Total area	~ 1.6	mm^2
Clock frequency	40	MHz
Latency	50	ns



Credit: Benjamin Parpillon, Anthony Badea

Chip testing efforts



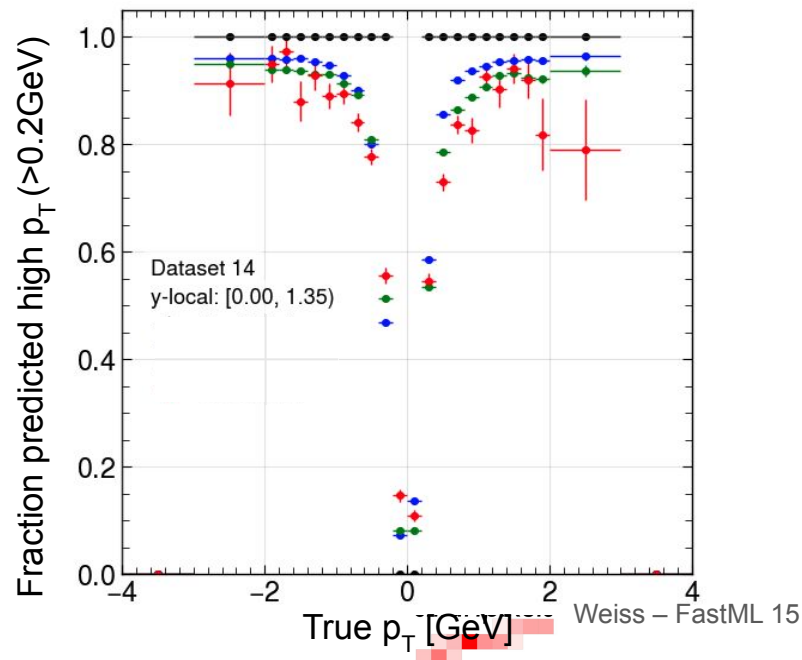
Test routines developed to measure:

- Power consumption
- Pixel S-curves
- DNN programming/
predictions on chip



FPGA based setup to communicate with the smartpixels ROIC

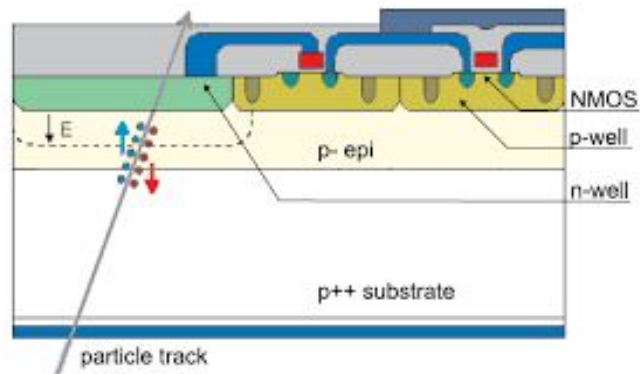
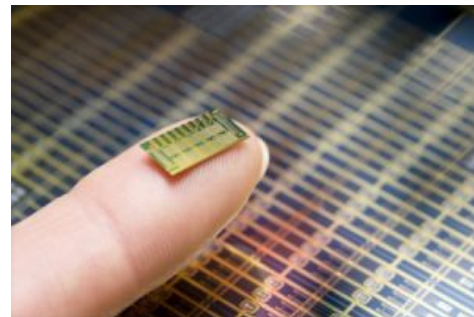
- Spacely+Caribou+Peary to convert python test routines into bitstreams



Future trajectory for Smartpixels:

This was a whirlwind tour of Smartpixels, **much** more in the works!

- Physics impact studies
- ROI expert algorithm
- **Tapeout of regression algorithm chip!**
- Training on real readout charge clusters
- Test beam of Smartpixels chips
- Real irradiation retainability
- Analog input for new architectures like MAPS



Conclusions

Smartpixels is an on-ASIC, ML based, particle-track data **compression** implementation for pixel detector ROCs.

- Data rates will render current triggering schemes unviable for future collider experiments
- Algorithms to **filter** and **regress** track parameters from a single silicon layer
 - Current compressing these algorithms and adding realism
- Synthesizing ASIC logic design for tapeout
- First results from filtering algorithm **on chip!**

The Smartpixels group!

Fermi National Accelerator Laboratory: Abhijith Gandrakota, Benjamin Parpillon, Chinar Syal, Douglas Berry, Farah Fahim, Gauri Pradhan, Giuseppe Di Guglielmo, James Hirschauer, Jennet Dickinson, Lindsey Gray, Nhan Tran, Ron Lipton

Cornell University: Jennet Dickinson, Ben Weiss

Johns Hopkins University: Dahai Wen, Morris Swartz, Petar Maksimovic

Northeastern University: Nick Manganelli

Northwestern University: Manuel Blanco Valentin

Oak Ridge National Laboratory: Aaron Young, Shruti R. Kulkarni

Purdue University: Mia Liu, Arghya Das

University of Chicago: Karri DiPetrillo, Anthony Badea, Carissa Kumar, Emily Pan, Rachel Kovach-Fuentes, Aidan Nicholas, Eliza Howard, Eric You

University of Colorado Boulder: Jannicke Pearkes, Ricardo Silvestre

University of Illinois Chicago: Corrinne Mills, Danush Shekar, Jieun Yoo, Mohammad Abrar Wadud

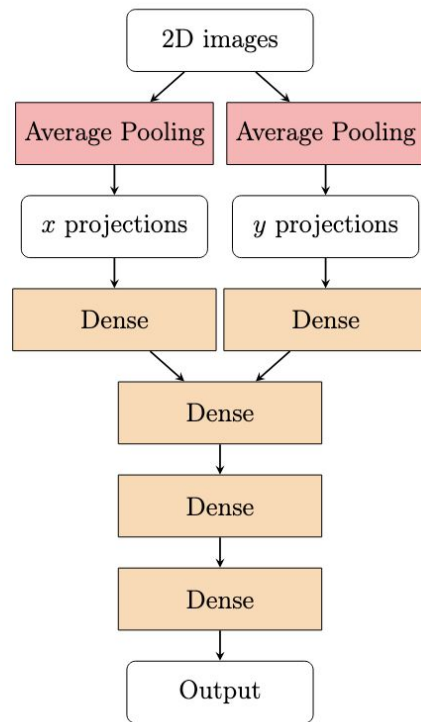
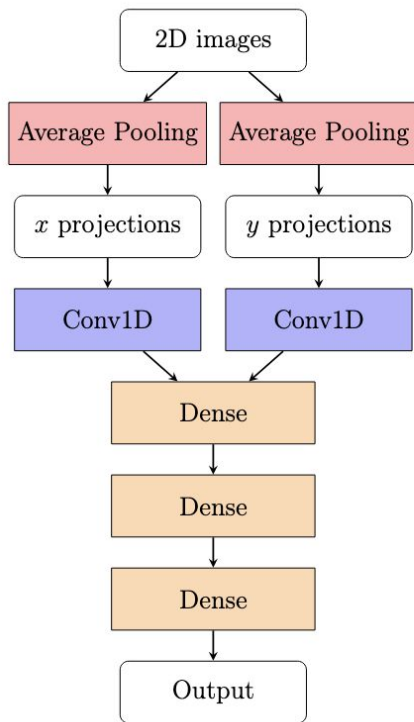
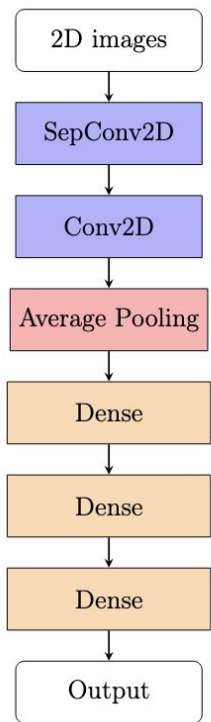
University of Illinois Urbana-Champaign: Mark S. Neubauer, David Jiang

University of Kansas: Alice Bean



Regression Alg. design space exploration

Reducing # of operations 

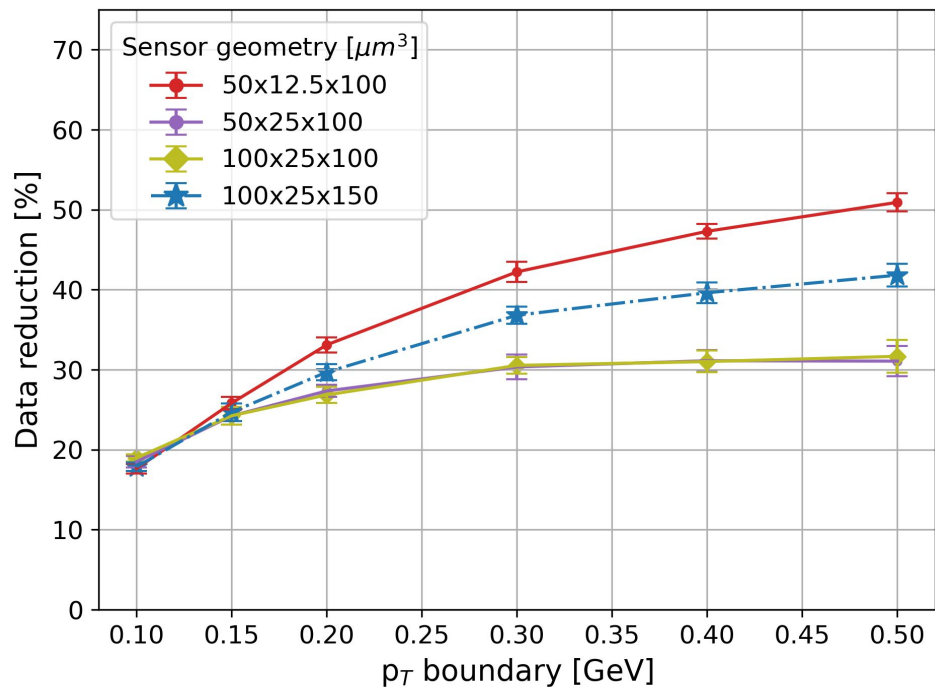
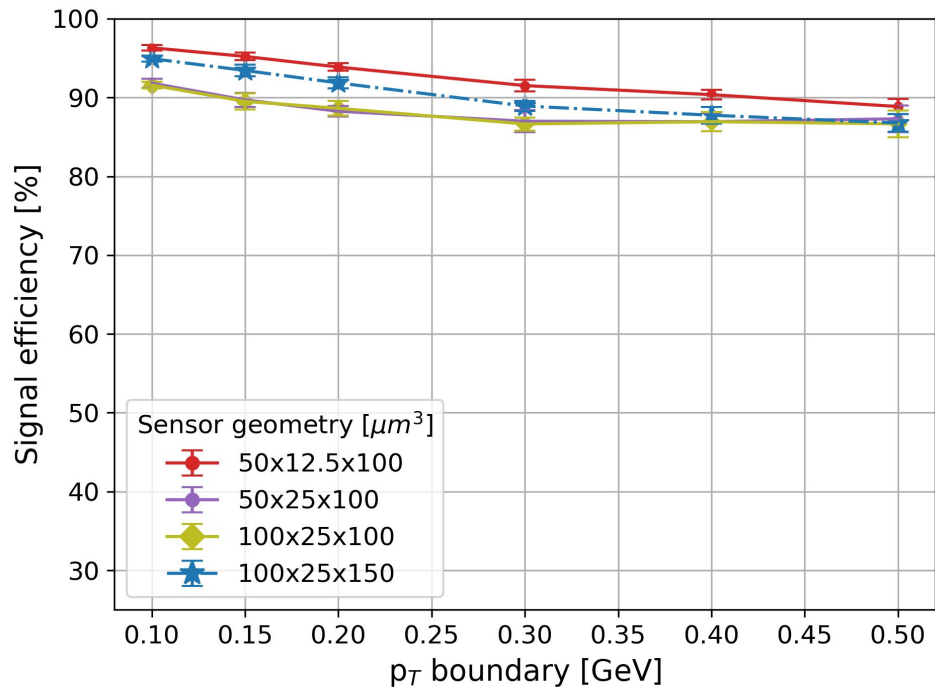


Current development to reduce model size for hardware implementation

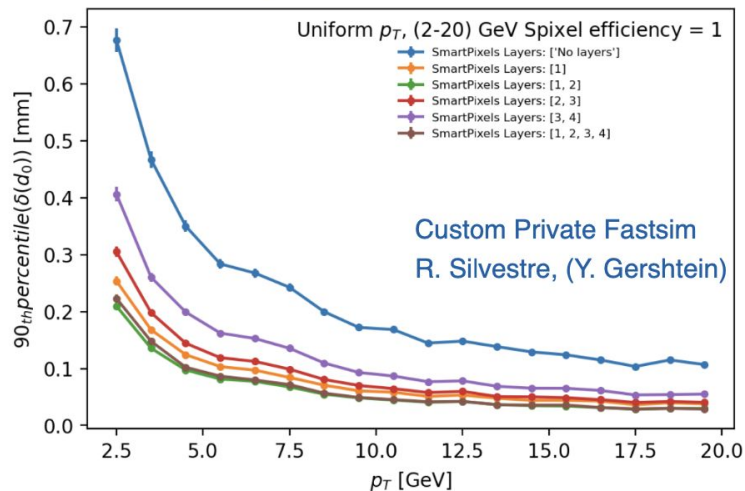
- On ASIC, generally:
Conv2D > Conv1D > Dense

ASIC Synthesis (45nm)	
Clock Period	5 ns
Area Estimate	1.4 mm ²
Buffer Area	0.0017 mm ²
Inverter Area	0.055 mm ²
Logic Area	0.80 mm ²
Sequential Area	0.52 mm ²
Latency	27 μ s

Varying pT Threshold

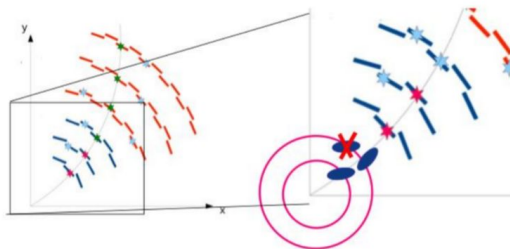
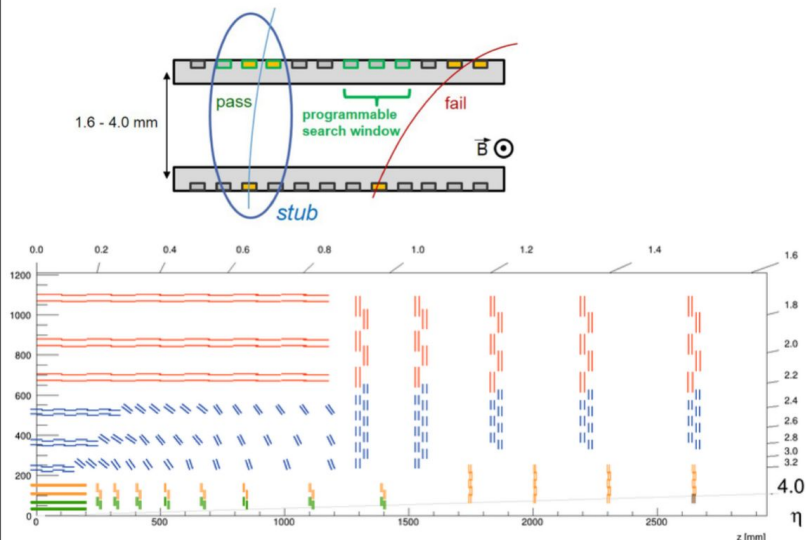


An Aside on Transverse Impact Parameters



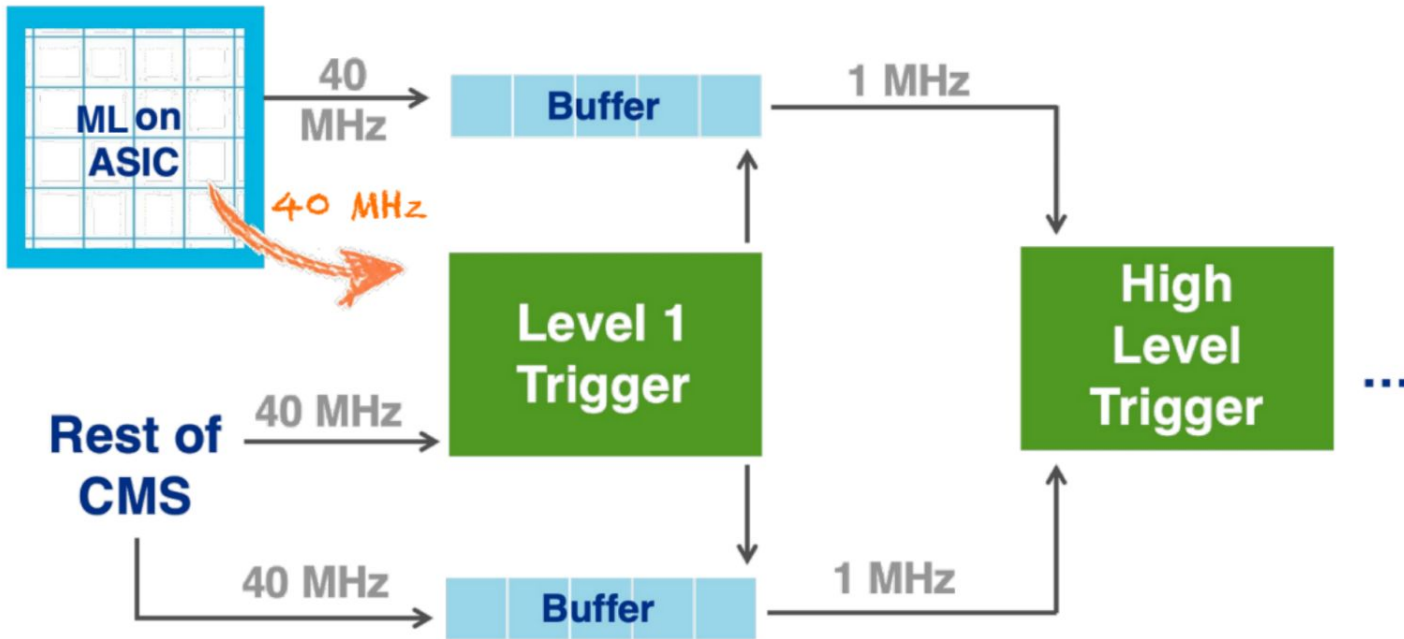
- Transverse impact parameter resolution heavily influenced by pixel cluster precision and multiple scattering
 - LHC detector trackers are *heavy*
- Without innermost measurements:
 - extrapolated track fit has enormous errors at low p_T
 - Physics improvements only at large lifetimes / high transverse boost
 - *Poor precision* for physically interesting phase space

Using smartpixels data in a detector system








- We get roughly 2-3 degree resolution in the azimuthal direction
 - “Inside-out” tracking will not work very well, combinatorics too high
 - However we can use the outer tracker tracks to identify regions of interest and then the angular reconstruction from smartpixels to match the extrapolated track
- With more processing on detector we could try to find “pixel seeds” instead
 - More clean and pure, but this would require on detector track finding to deal with combinatorics

Readout chain: a futuristic smart pixels detector



Use ML to perform physics-motivated data reduction on-ASIC

Regression ROIC s'

- Algo #1: “conv2d model”
 - I/O
 - 13x21x2 inputs, bit-width 4 (fixed<4,1>)
 - 14 outputs fixed<8,2>
 - HLS area **0.81 mm²** (Syn **1.23 mm²**, PnR **1.46 mm²** )
- Algo #2: “conv1d model”
 - I/O
 - 13x21x2 inputs fixed<4,1>, bit-width 4 (fixed<4,1>)
 - 14 outputs fixed<8,2>
 - HLS area **0.39 mm²** (Syn **0.6 mm²** , PnR **0.72 mm²** )
- Algo #3: “MLP model”
 - I/O
 - 16x16x2 inputs, fixed<16,6>
 - 3 outputs fixed<16,6> (x and y + angle)
 - HLS area **0.52 mm²** (Syn **0.78 mm²** , PnR **0.94 mm²** )