Using an Optical Processing Unit for tracking and calorimetry at the LHC

- Optical Processing Units
- High-energy physics colliders in 3 slides
- OPU for Tracking
- OPU for Calorimetry

The team:

Biswajit Biswas, Aishik Ghosh, David Rousseau (~~LAL~~-Orsay)

Laurent Basara (LRI-Orsay) IJCLab

Many thanks to the LightOn team in particular Laurent Daudet, Iacopo Poli for access to LightOn OPU

And to Steve Farrell, Wahid Bhimji for access to the dataset and useful discussions

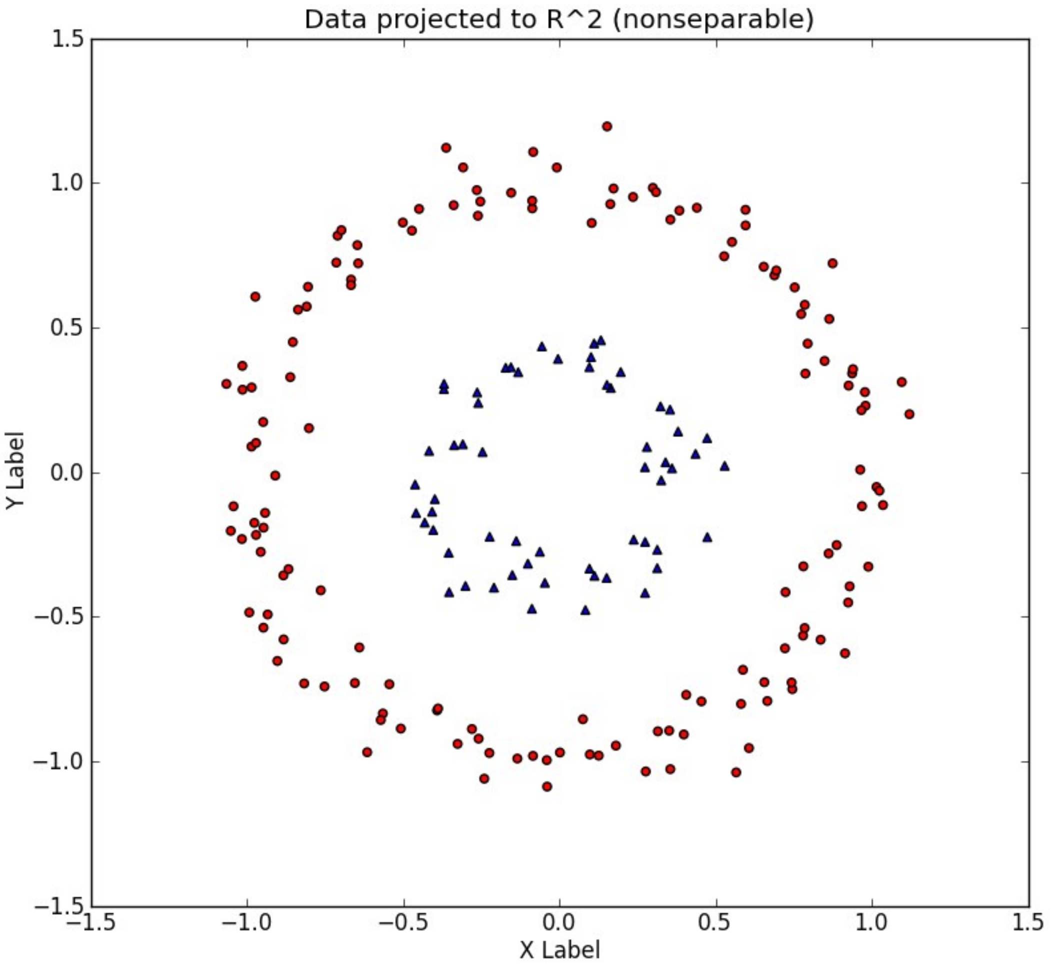
Using an Optical Processing Unit for tracking and calorimetry at the LHC

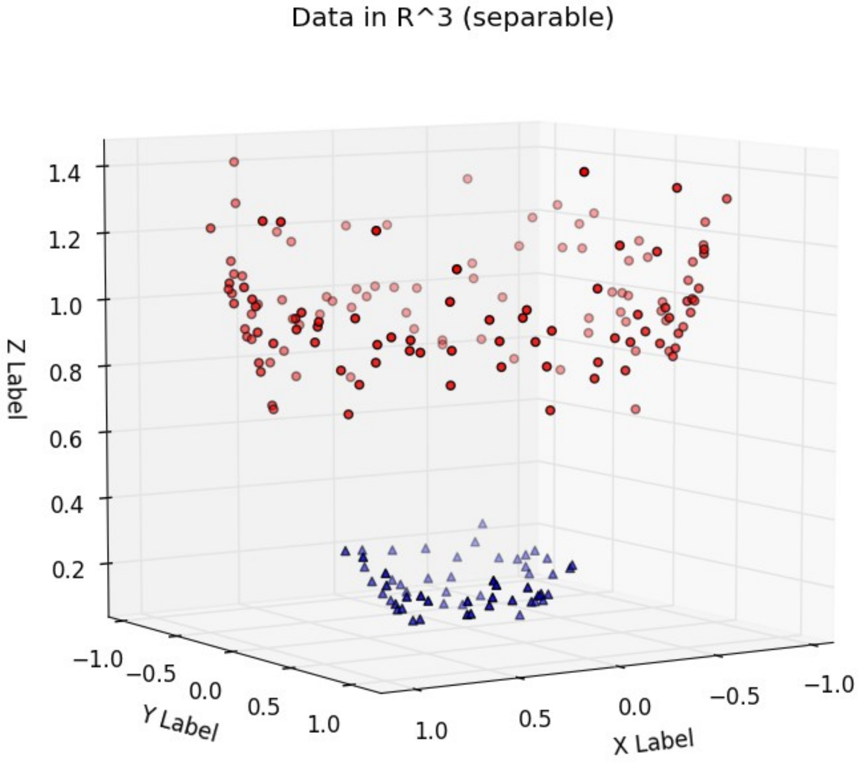
- Optical Processing Units
- High-energy physics colliders in 3 slides
- OPU for Tracking
- OPU for Calorimetry

« Using a term like
nonlinear science
is like referring
to the bulk of zoology
as the study
of non-elephant animals »

— Stanislaw Ulam

Non-linear problems, SVM and the kernel trick





Kitchen Sinks

Random Features for Large-Scale Kernel Machines

Ali Rahimi and Ben Recht

Abstract

To accelerate the training of kernel machines, we propose to map the input data to a randomized low-dimensional feature space and then apply existing fast linear methods. Our randomized features are designed so that the inner products of the transformed data are approximately equal to those in the feature space of a user specified shift-invariant kernel. We explore two sets of random features, provide convergence bounds on their ability to approximate various radial basis kernels, and show that in large-scale classification and regression tasks linear machine learning algorithms that use these features outperform state-of-the-art large-scale kernel machines.

Everything about the kitchen sink

To fit a kernel SVM, you normally fit a weighted sum of Radial Basis Functions to data:

$$f(x;\alpha) = \sum_{i=1}^{N} \alpha_i k(x, x_i)$$

We showed how to approximate each of these basis functions in turn as a sum of some random functions that did not depend on the data:

$$k(x, x') \approx \sum_{j=1}^{D} z(x;\omega_j) z(x';\omega_j)$$

A linear combination of a linear combination is another linear combination, but with this new linear combination has many fewer (D) parameters:

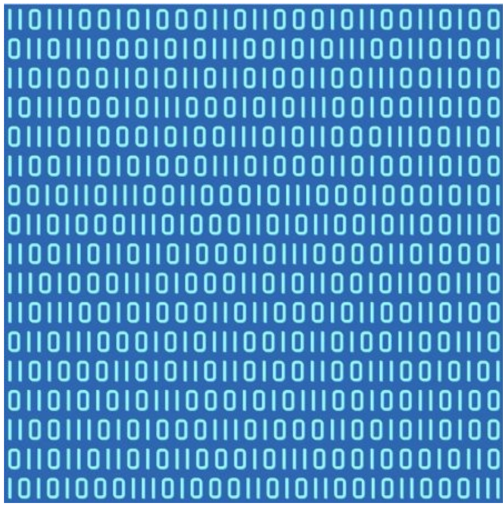
$$f(x;\alpha) \approx \sum_{j=1}^{D} \beta_i z(x;\omega_j)$$

We showed how to approximate a variety of radial basis functions and gave bounds for how many random functions you need to approximate them each of them well.

Original paper: https://people.eecs.berkeley.edu/~brecht/papers/07.rah.rec.nips.pdf

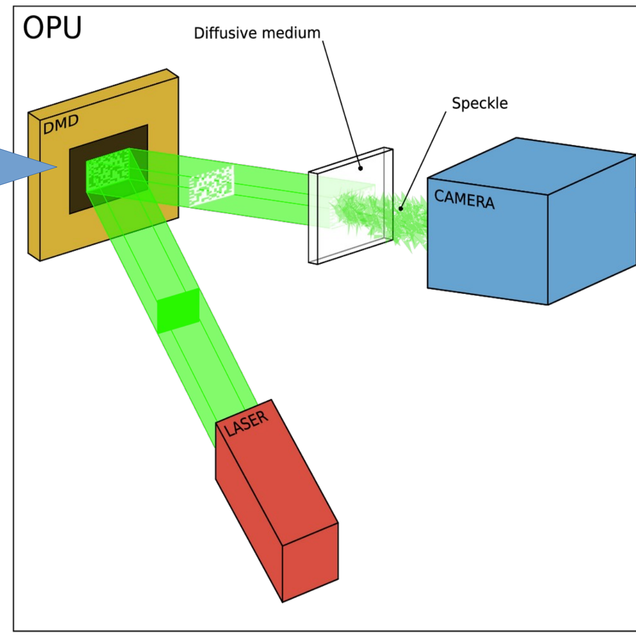
Popularization: http://www.argmin.net/2017/12/05/kitchen-sinks/

Optical Processing Unit

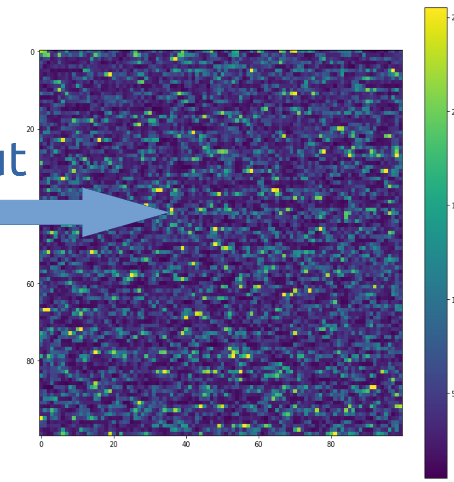


https://docs.lighton.ai/notes/opu.html

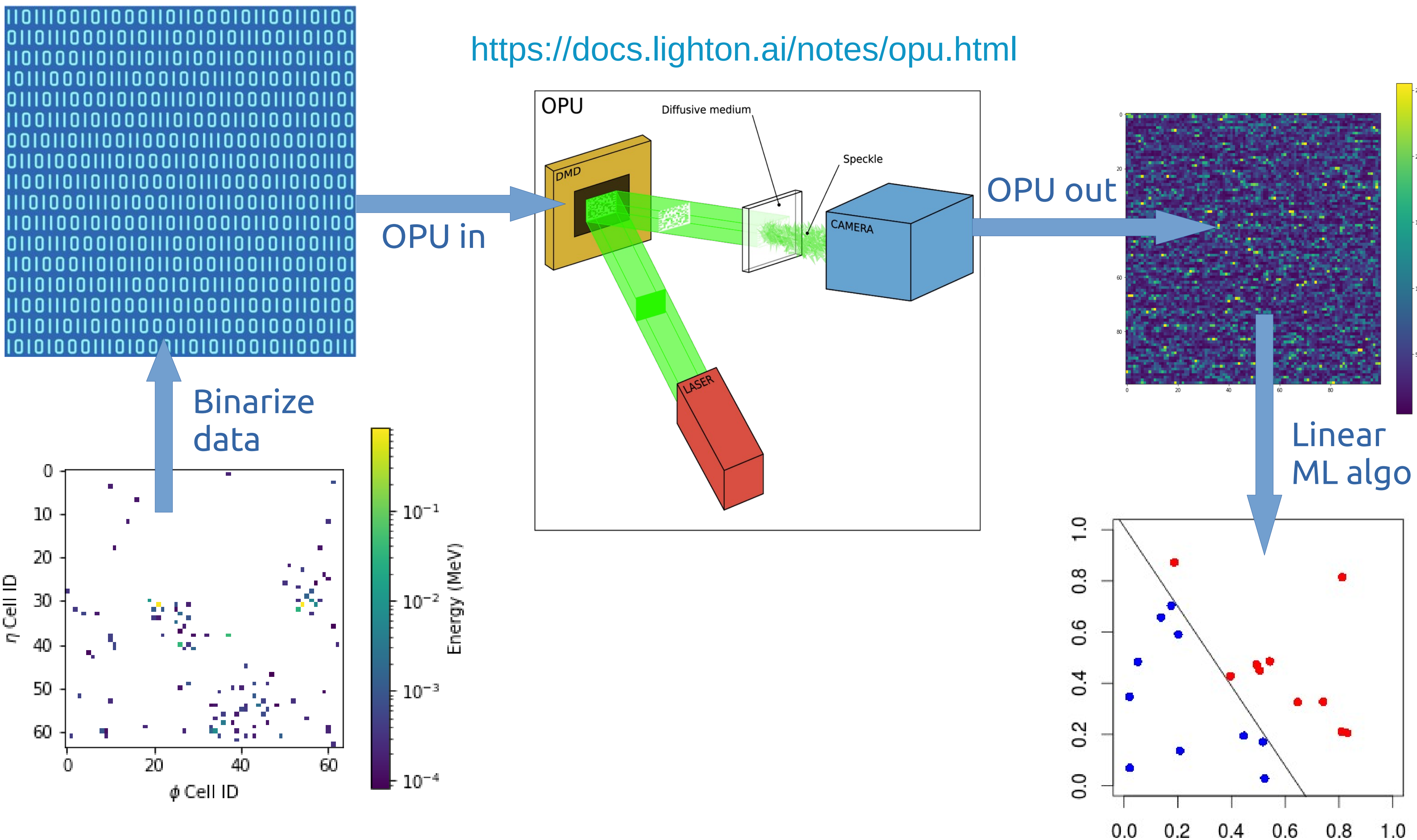
OPU in



OPU out



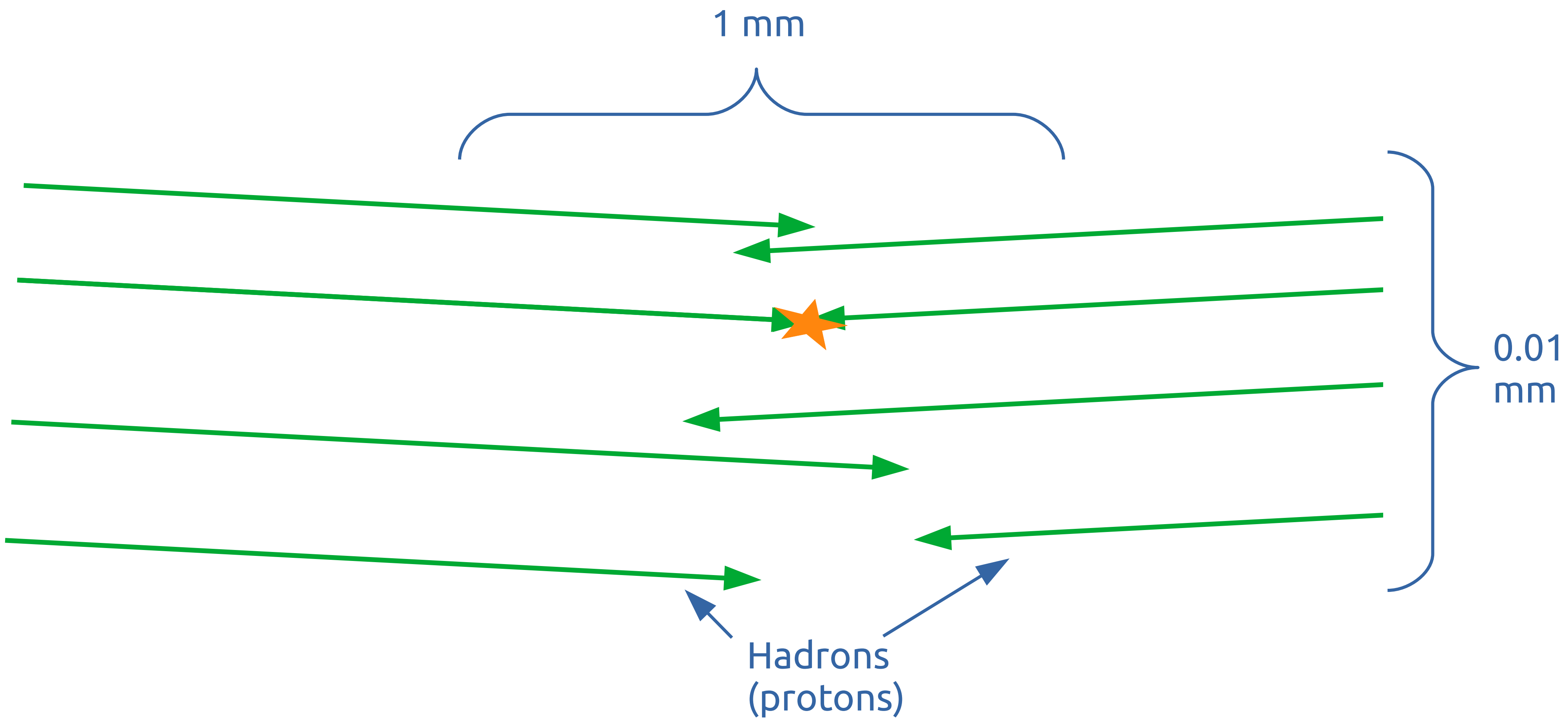
Optical Processing Unit ML workflow



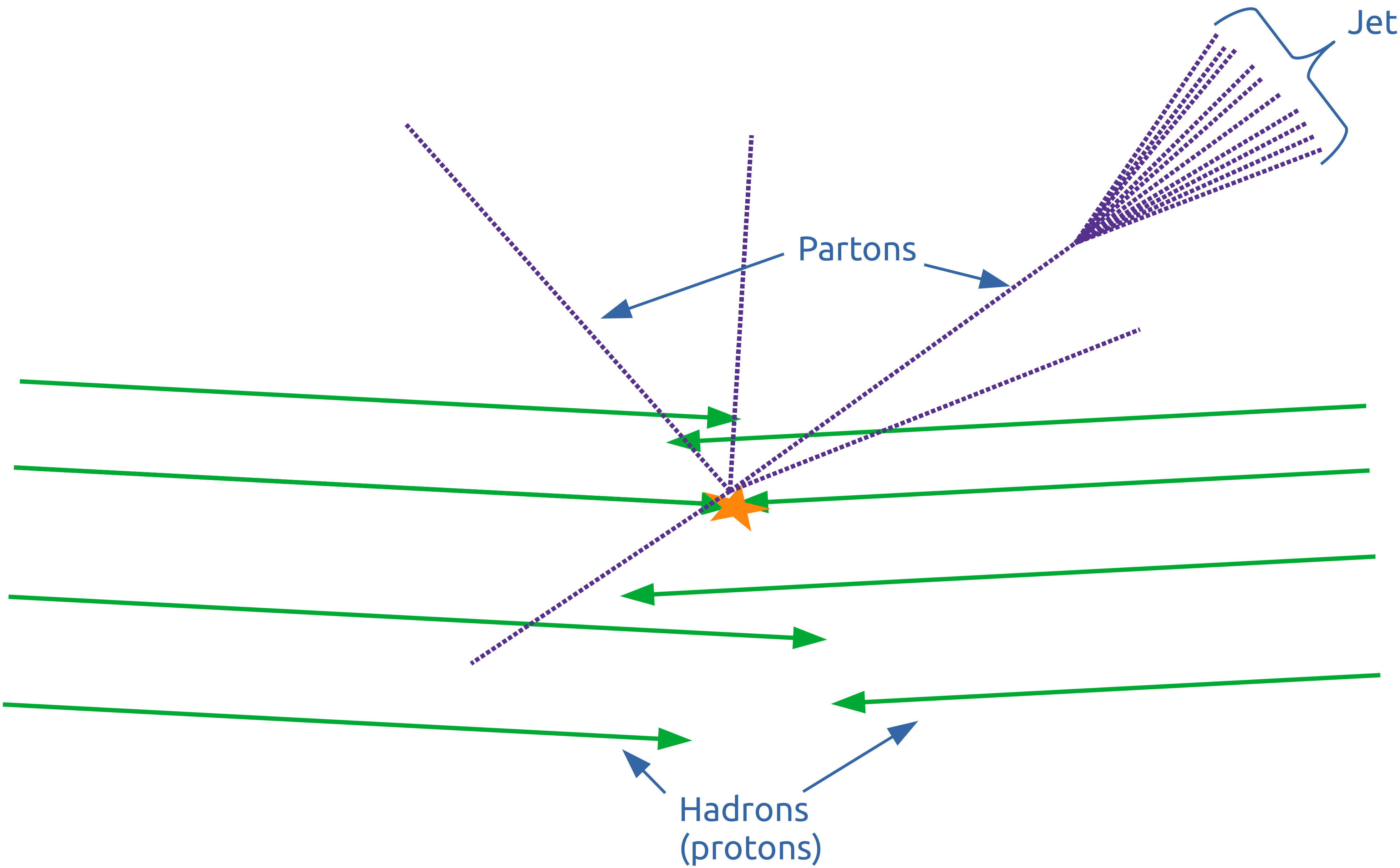
Using an Optical Processing Unit for tracking and calorimetry at the LHC

- Optical Processing Units
- High-energy physics colliders in 3 slides
- OPU for Tracking
- OPU for Calorimetry

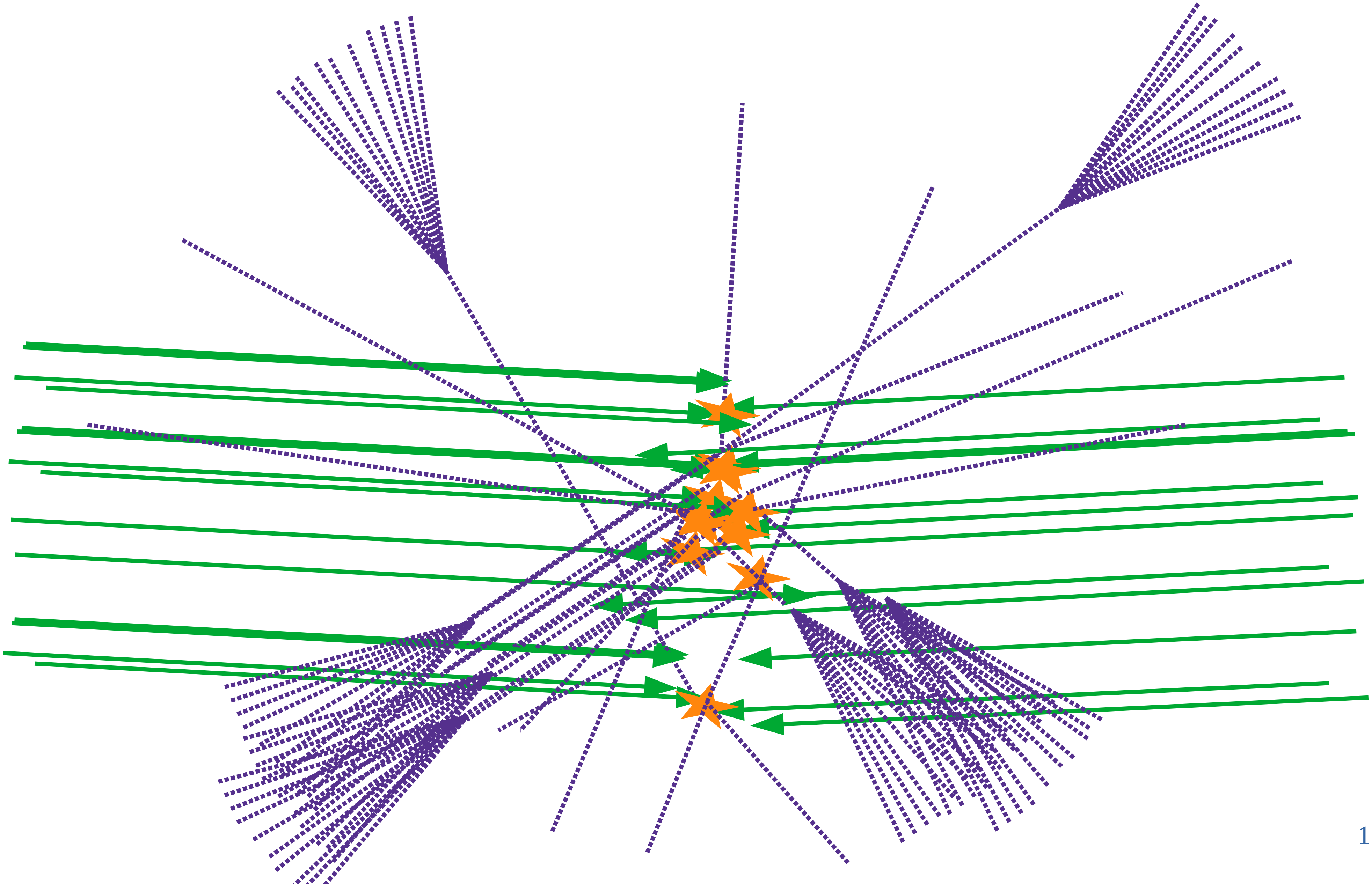
Large Hadron Collider

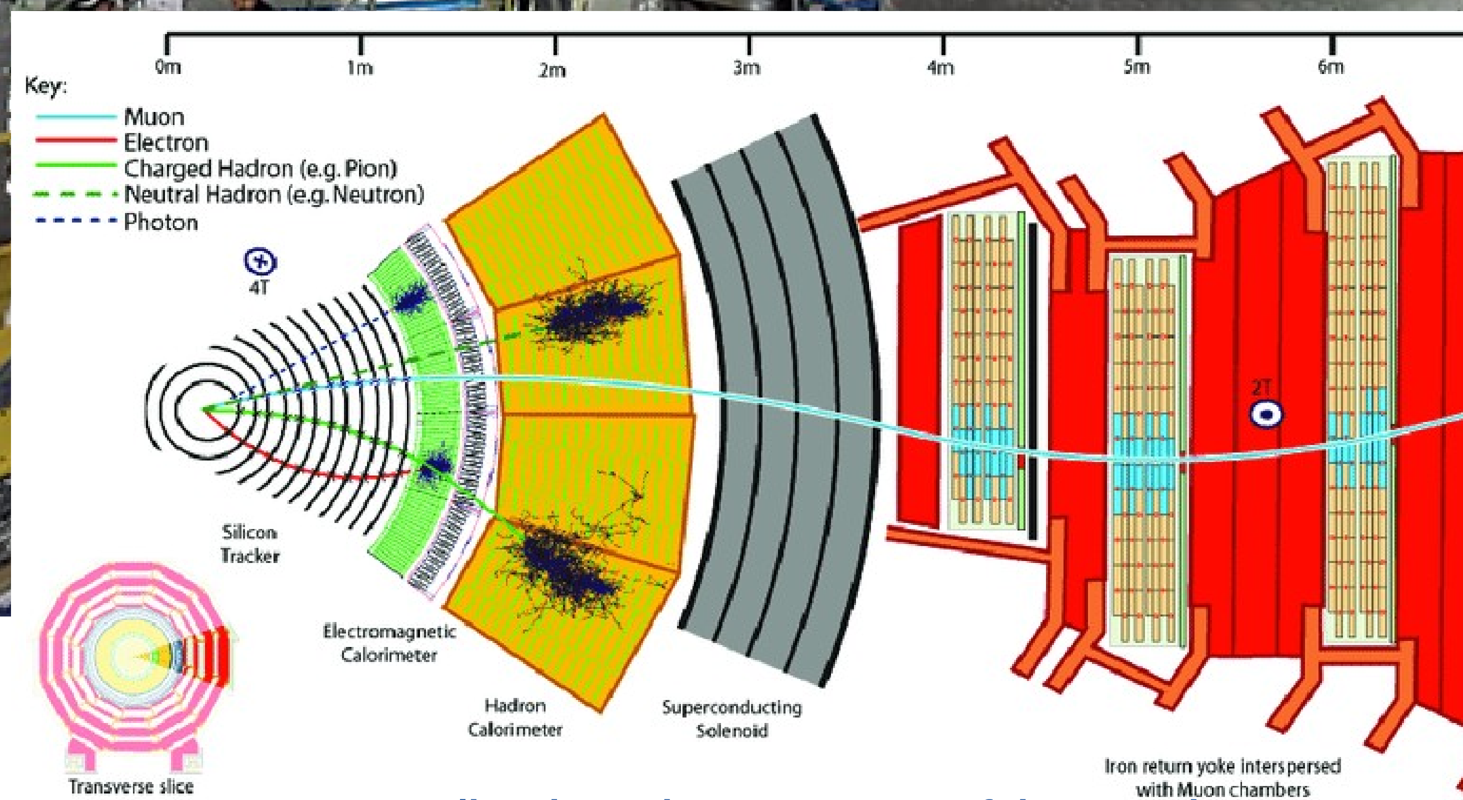


Particle creation



Pile-up





Detection

Transverse slice through one segment of the CMS detector

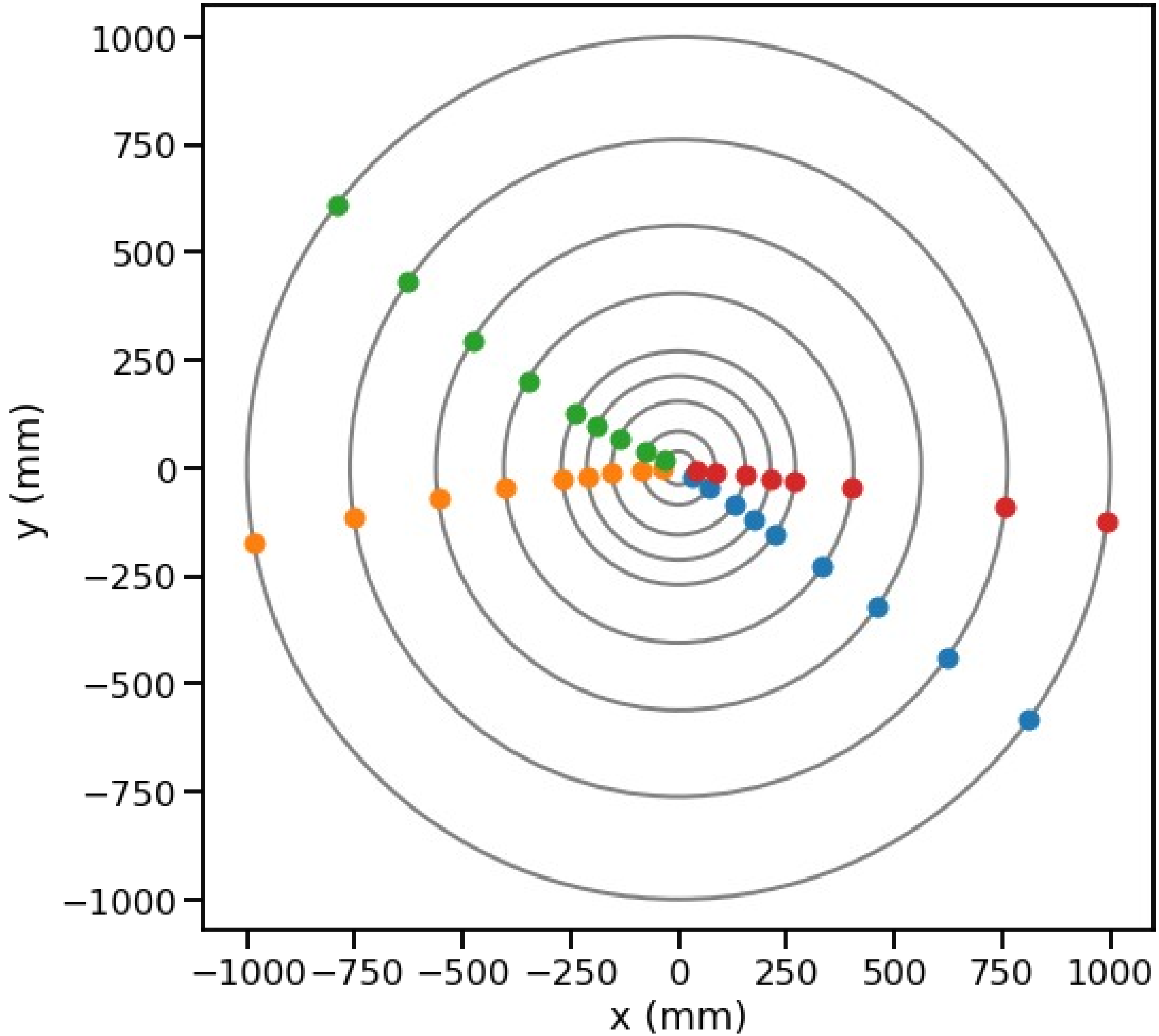
(HL-)LHC: a few numbers

- ~10 PB of data per year
- Pile-up: 50 → 200
- 10 K particles / collision
- 100 K 3D points / collision
- 3-20 hits per particle
- Looking for innovative data analysis on LHC next generation

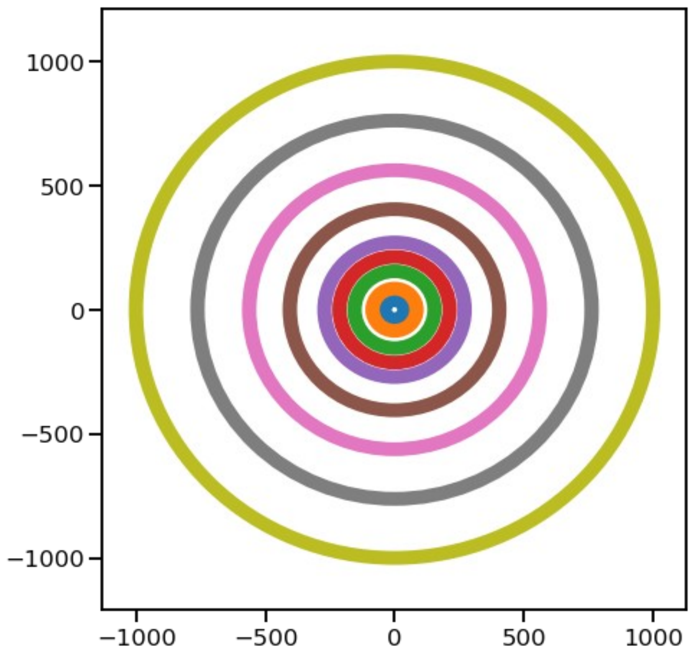
Using an Optical Processing Unit for tracking and calorimetry at the LHC

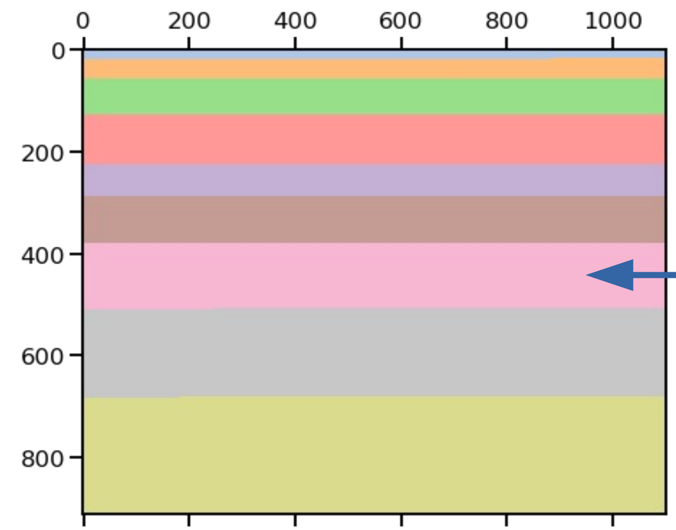
- Optical Processing Units
- High-energy physics colliders in 3 slides
- OPU for Tracking
- OPU for Calorimetry

Reduced 2D dataset

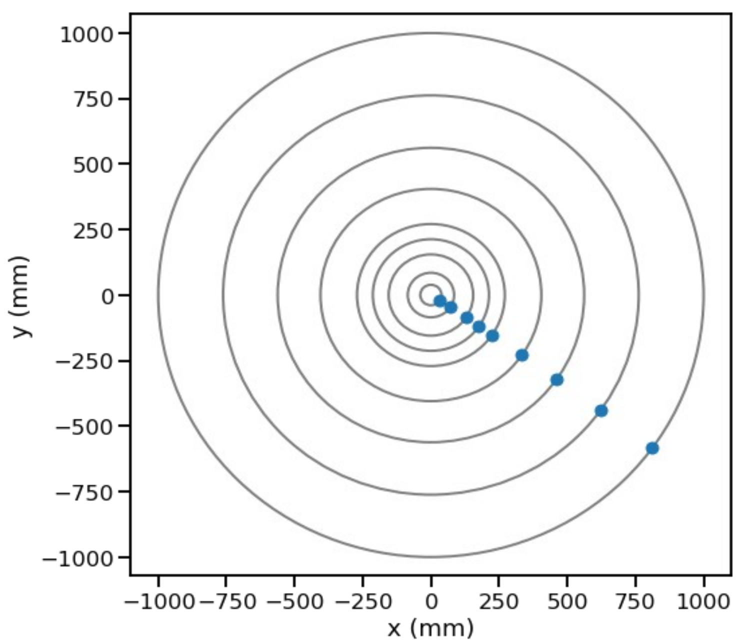


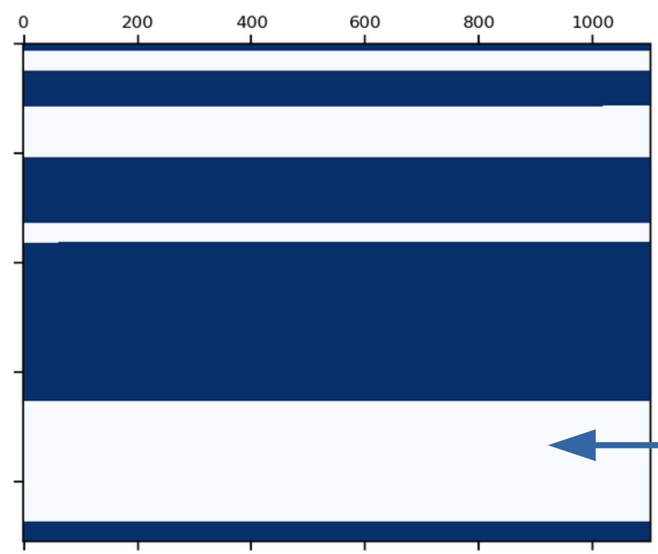
Binary encoding





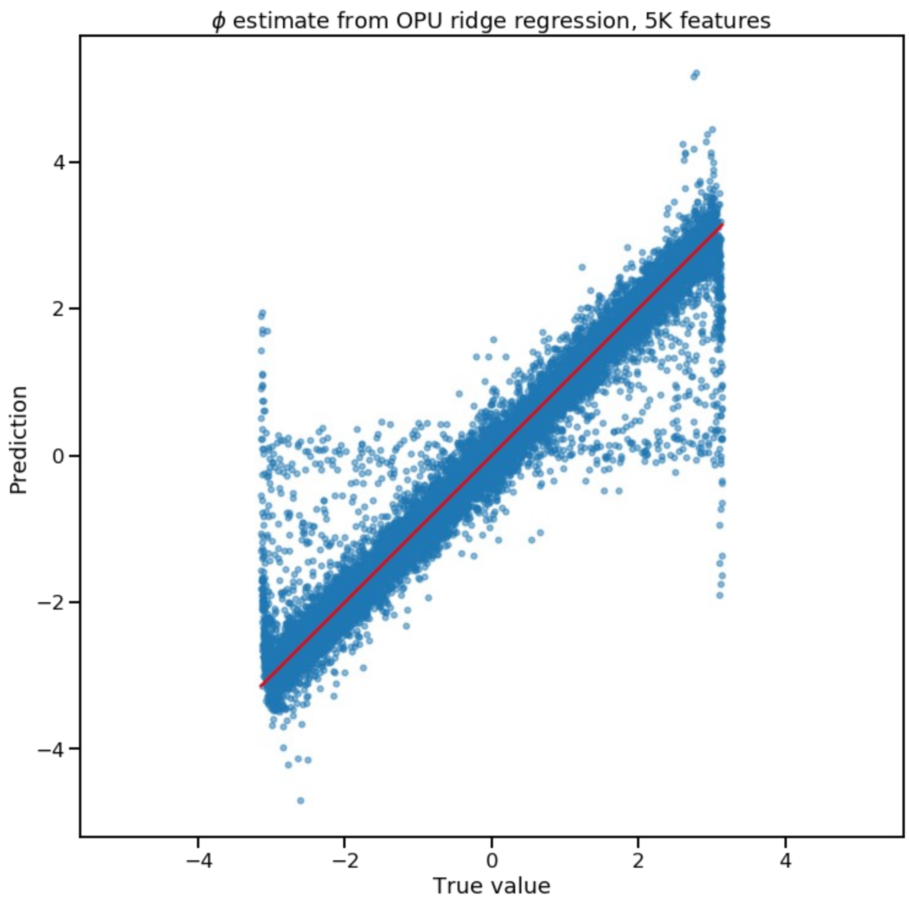
~ 2 times less tracker pixels than DMD pixels
→ all pixels represented on DMD at once

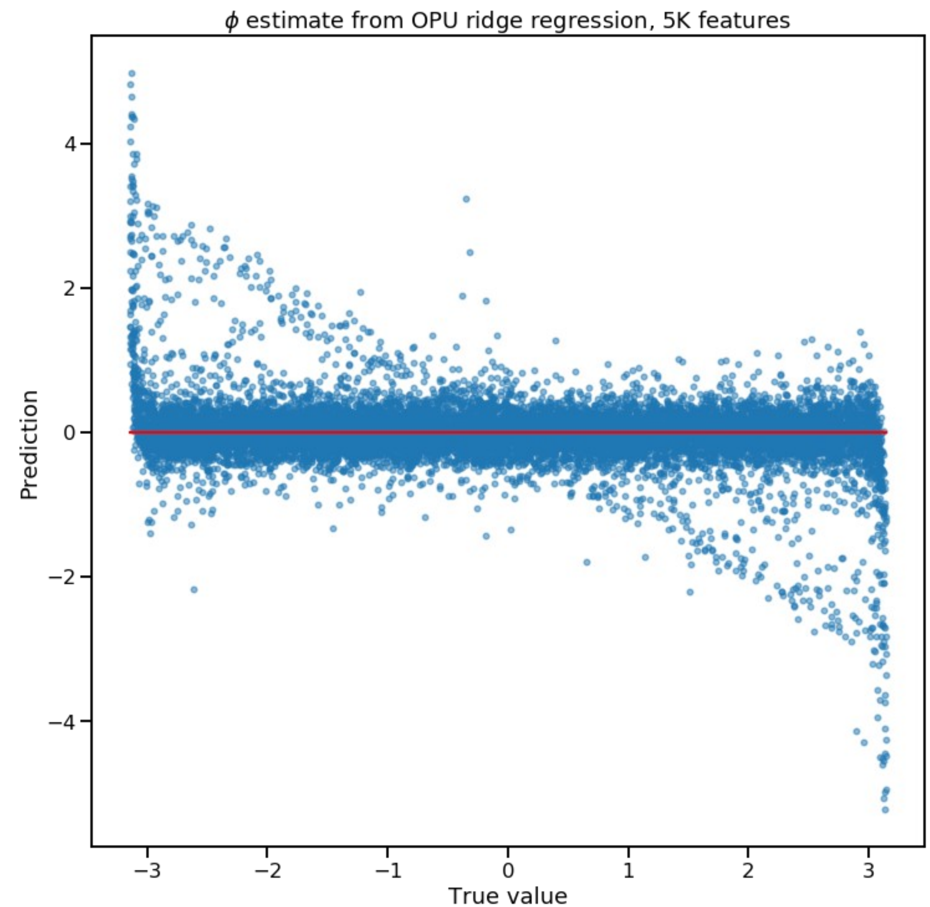




Each layer at most one hit
we change bool value
at each hit seen
→ ~ 50 % of DMD lit

Estimation of initial angle

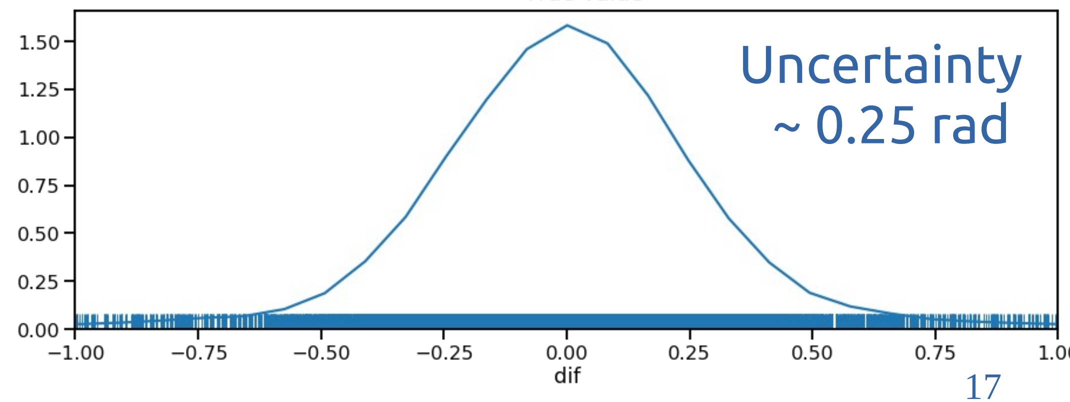




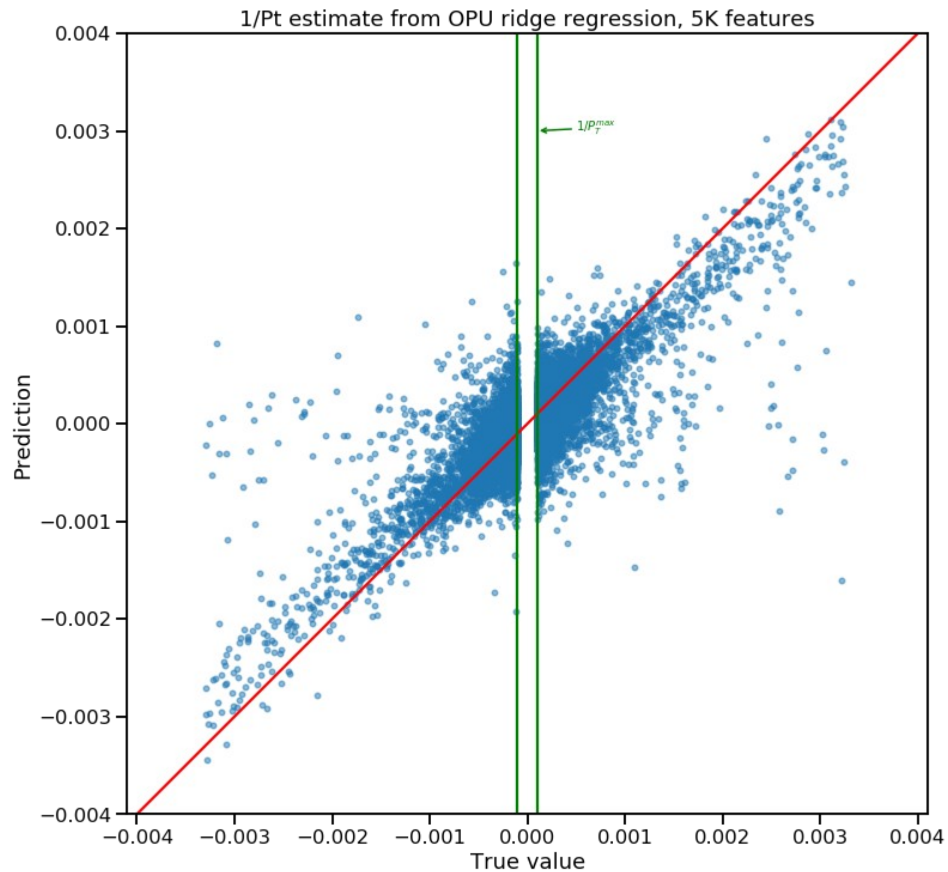
$$\boxed{\min_{\beta \in R^{m \times n}} \; ||X\beta - y|| + ||\gamma\beta||}$$

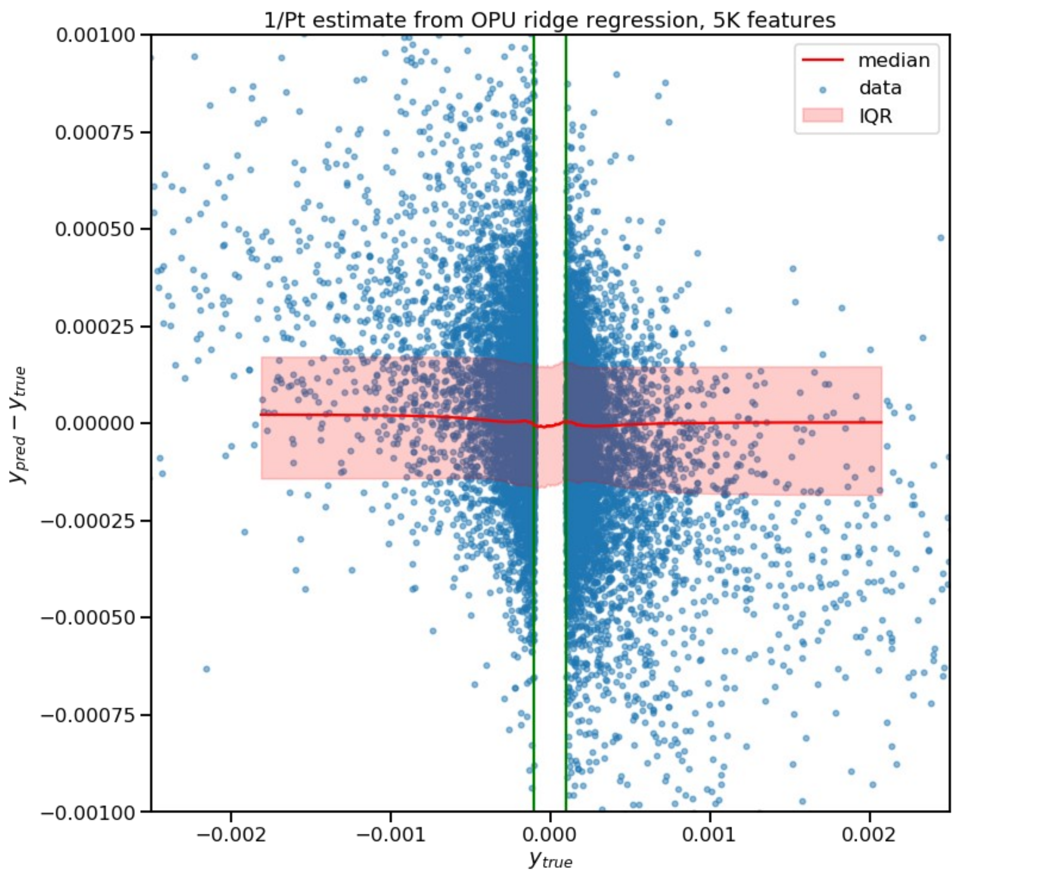
(5K) random features

Ground truth angle



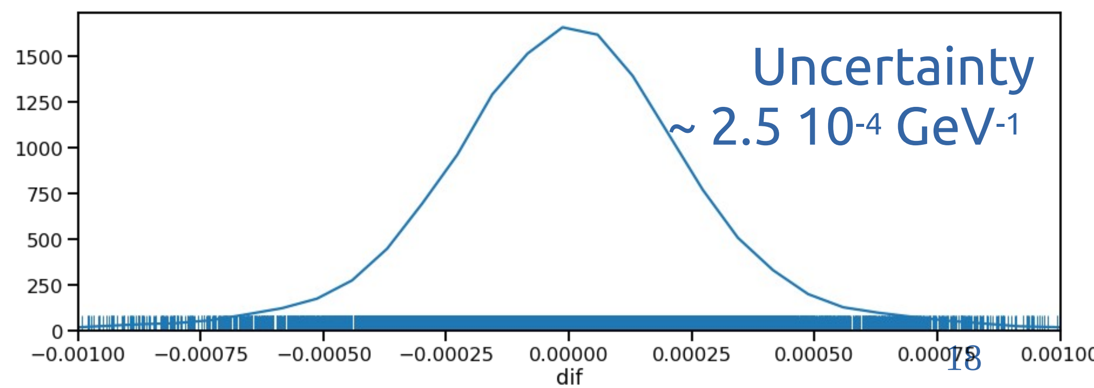
Estimation of (inverse) momentum





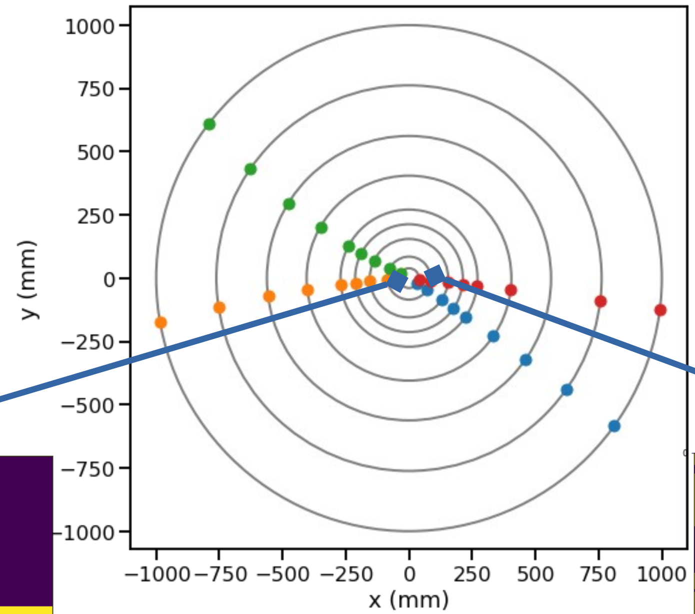
$$\boxed{\min_{\beta \in R^{m \times n}} ||X\beta - y|| + ||\gamma\beta||}$$

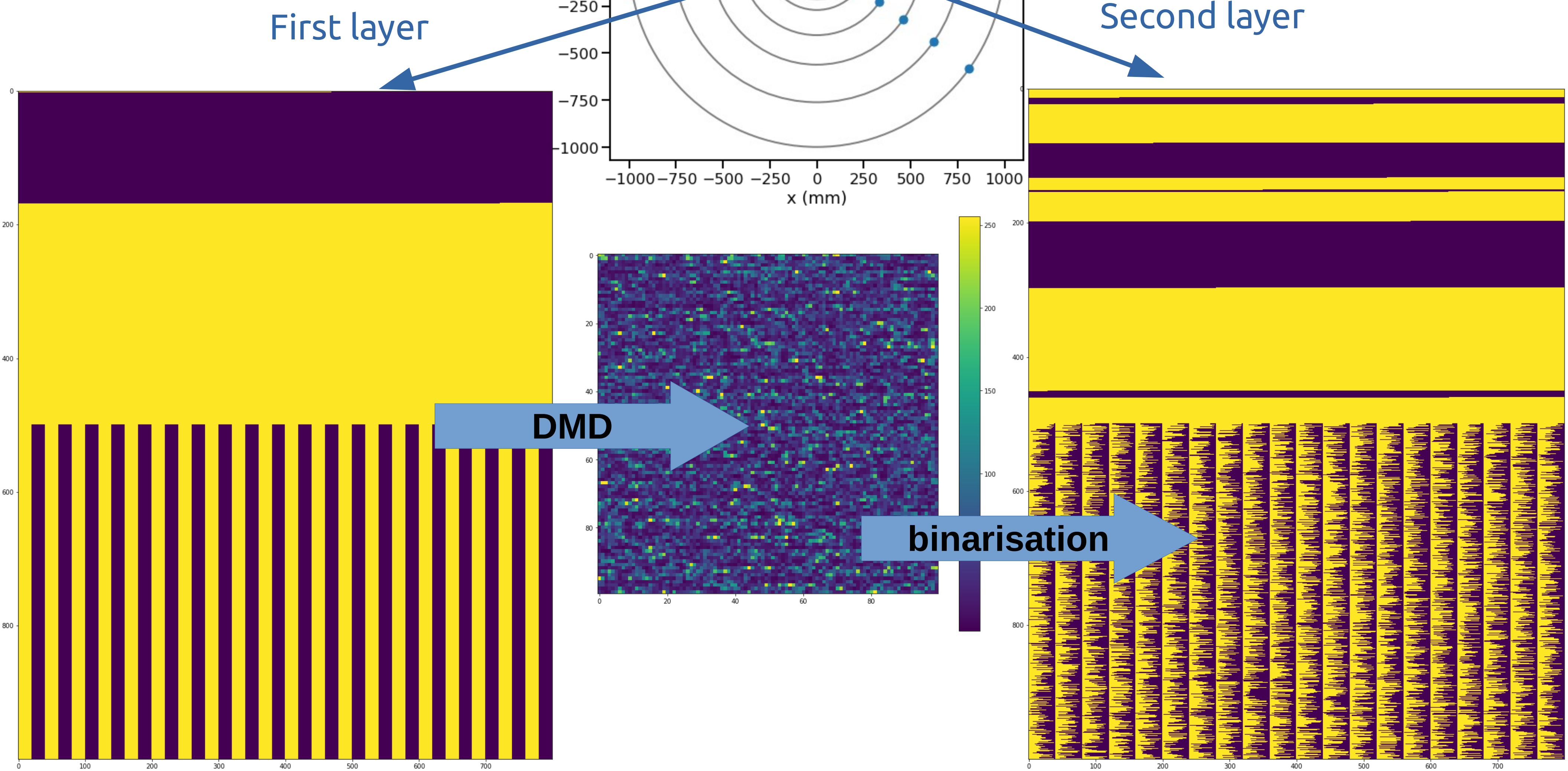
Inverse momentum (~curvature)



Uncertainty
~ 2.5 10^{-4} GeV^{-1}

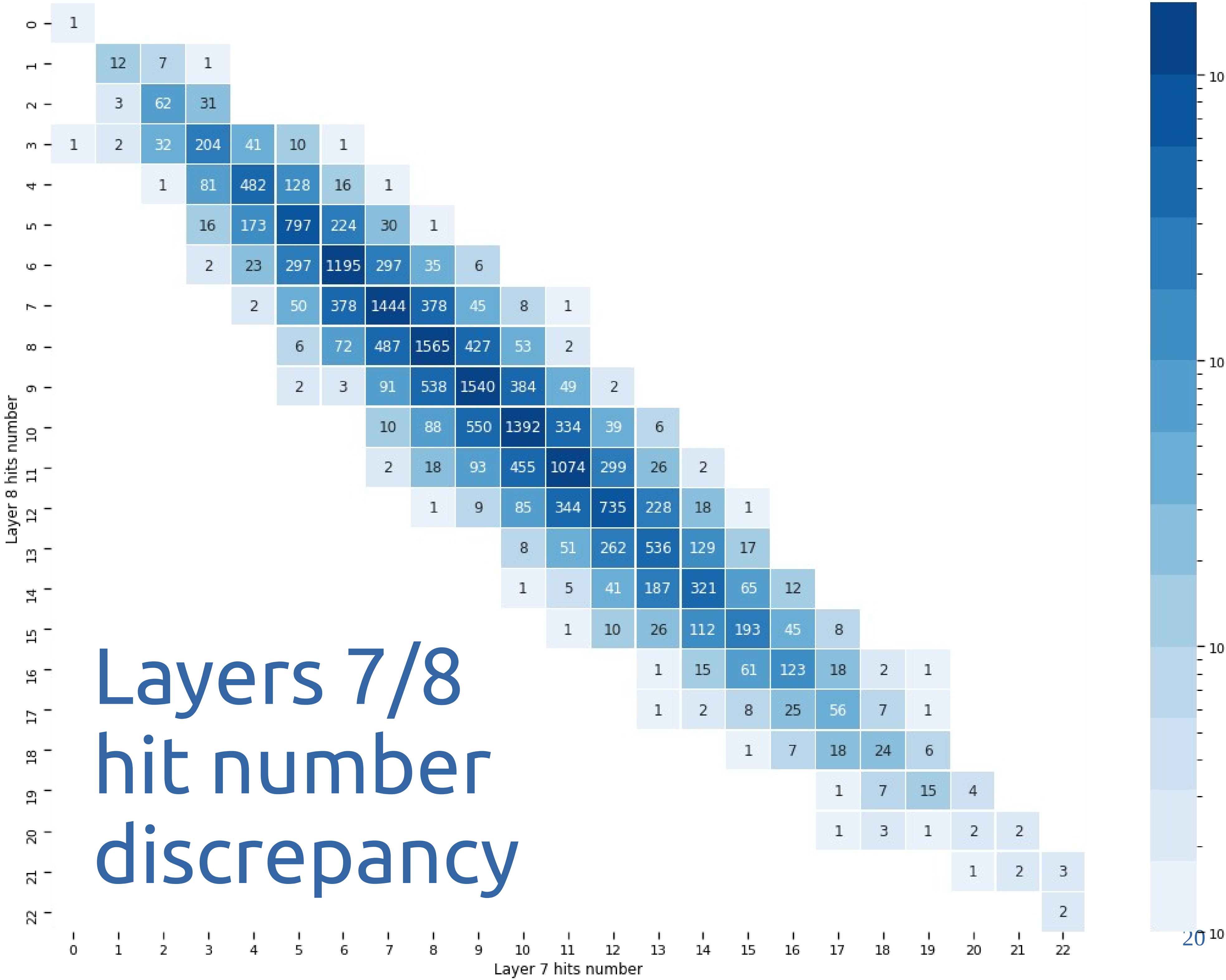
Binary encoding



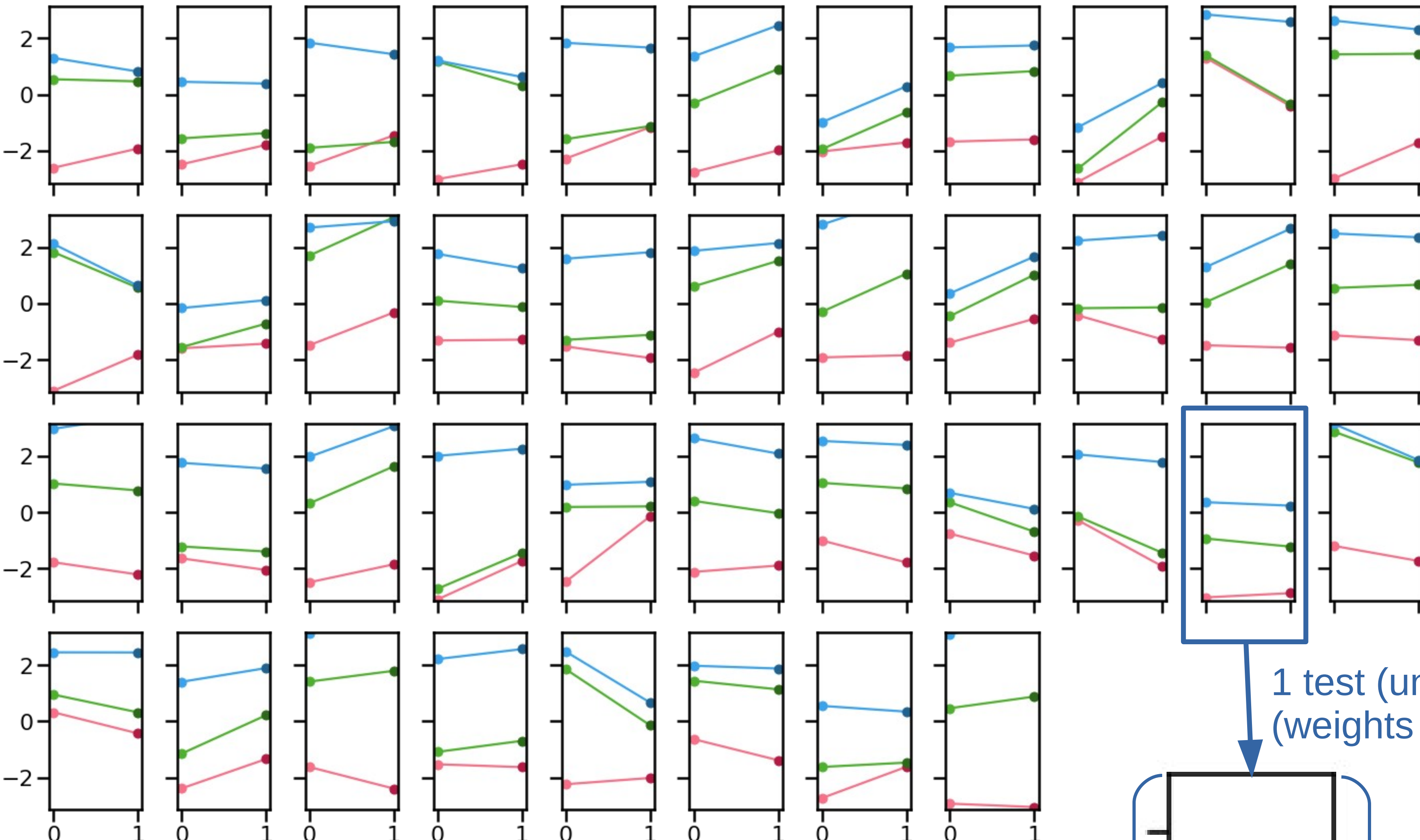


See echo state networks in Davide Faranda talk yesterday
https://indico.in2p3.fr/event/20187/contributions/78673/

Layers 7/8 hit number discrepancy



Oracle on number of hits (3)

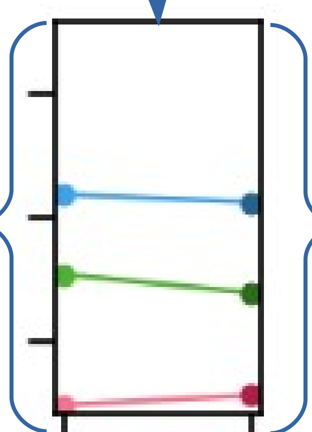


$$\min_{\beta \in R^{m\times n}} ||X\beta - y|| + ||\gamma\beta||$$

(3) last layer hit position
X (20K) events

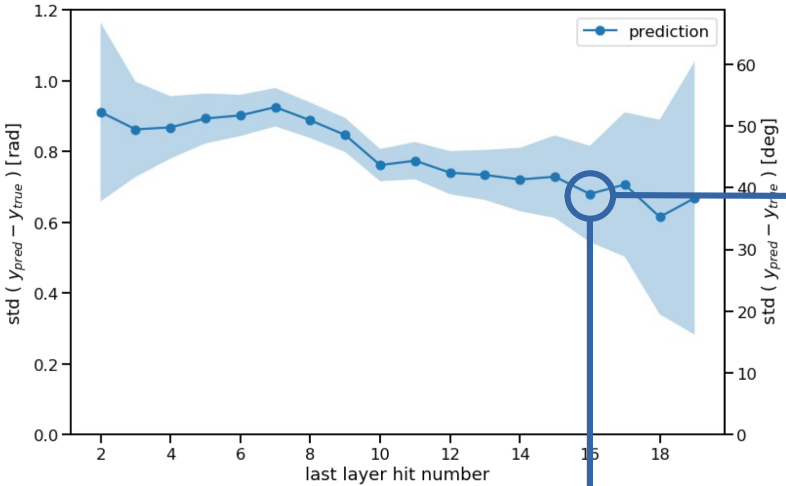
1 test (unseen) event
(weights from training)

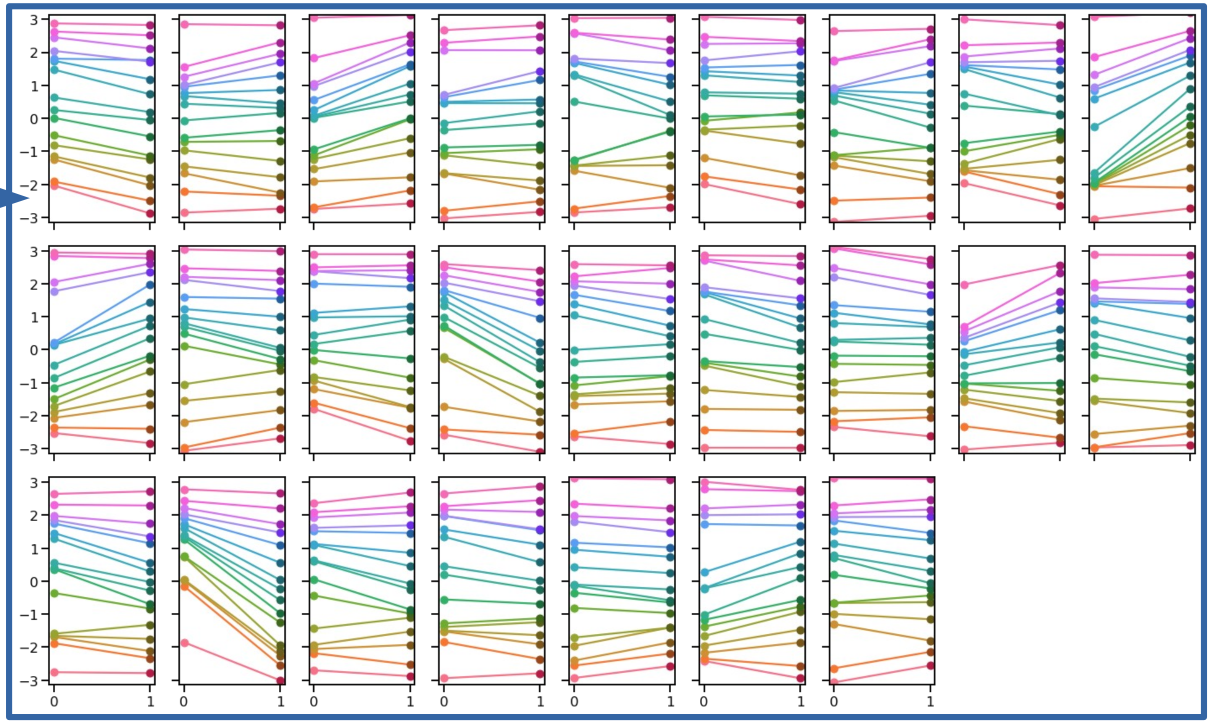
True L8
angle

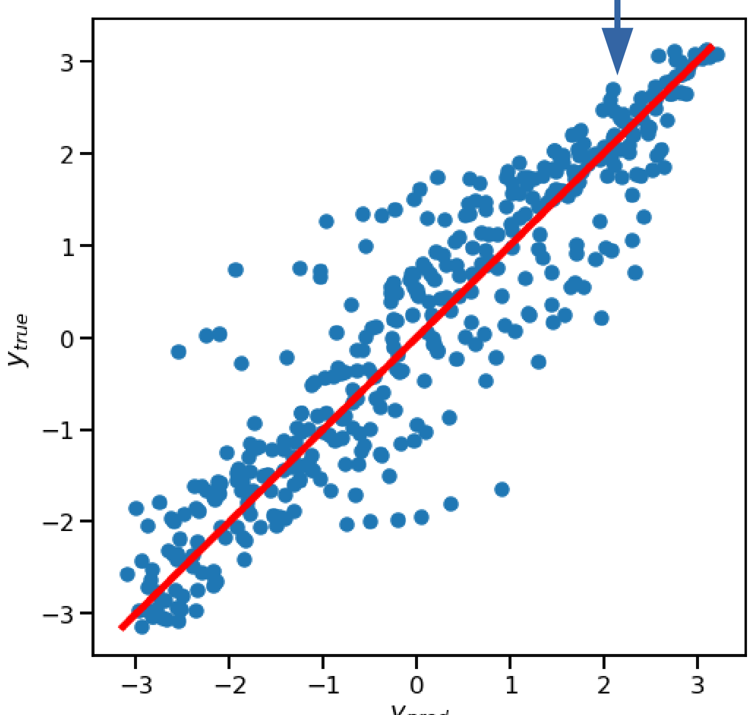


Predicted
L8 angle

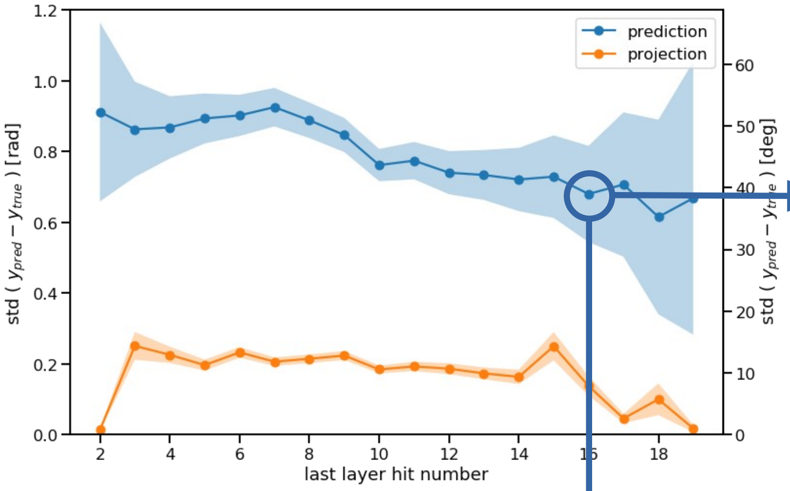
Standard deviation wrt hit number

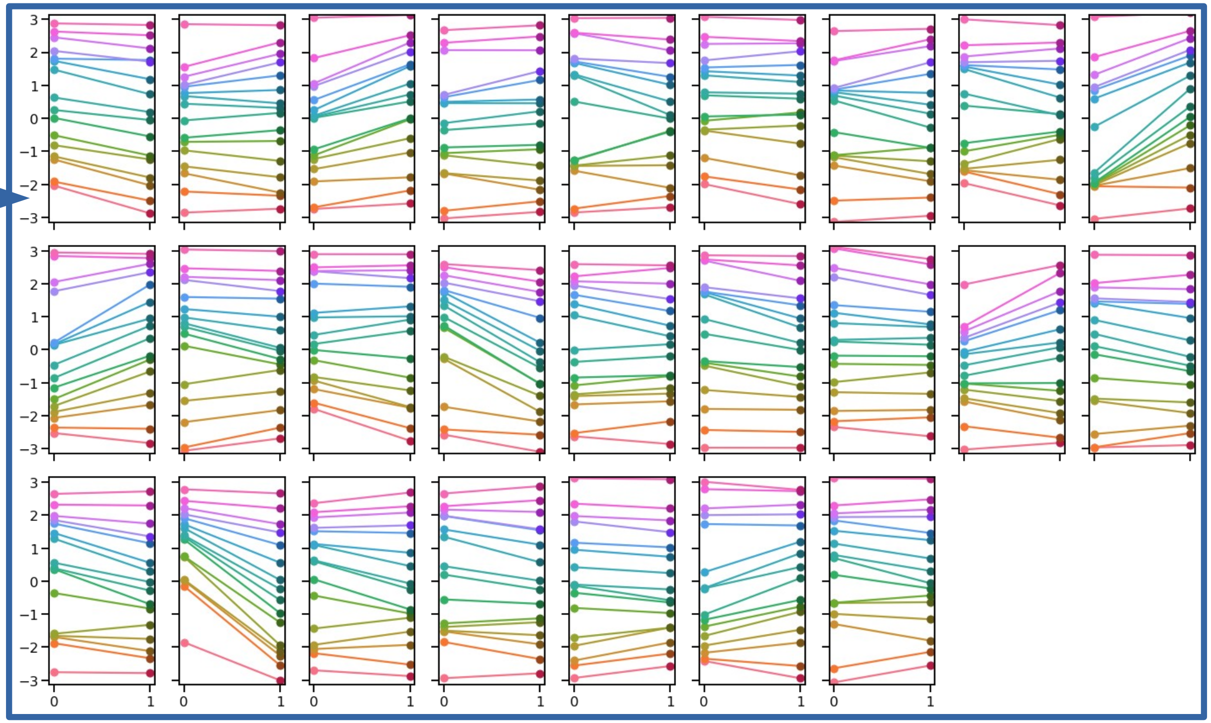


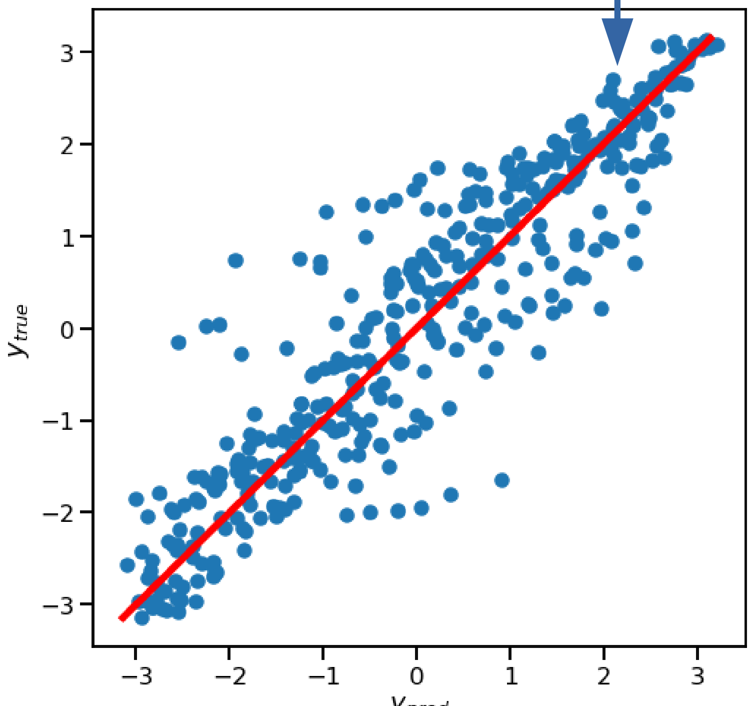




Standard deviation wrt hit number



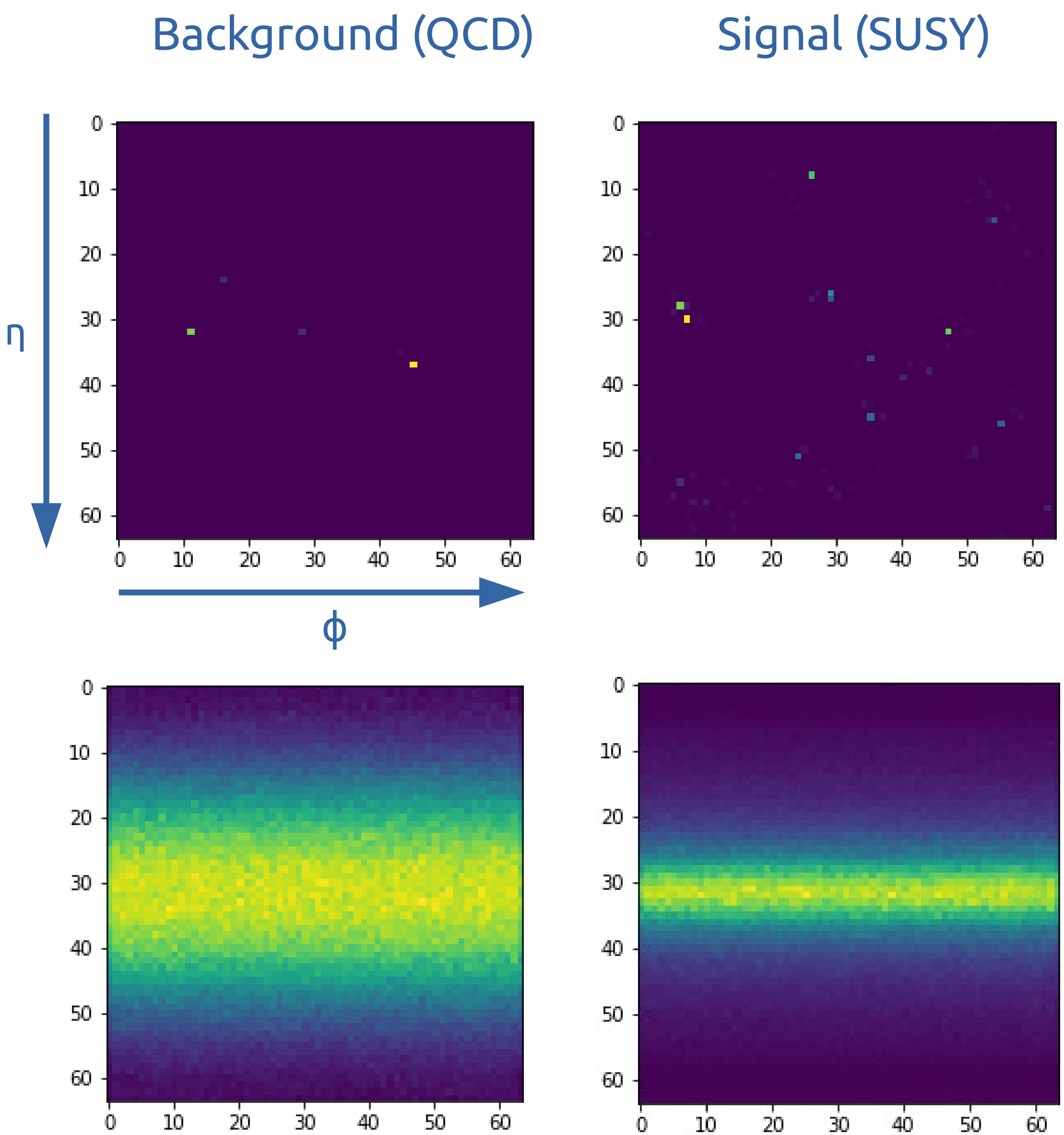




Using an Optical Processing Unit for tracking and calorimetry at the LHC

- Optical Processing Units
- High-energy physics colliders in 3 slides
- OPU for Tracking
- OPU for Calorimetry

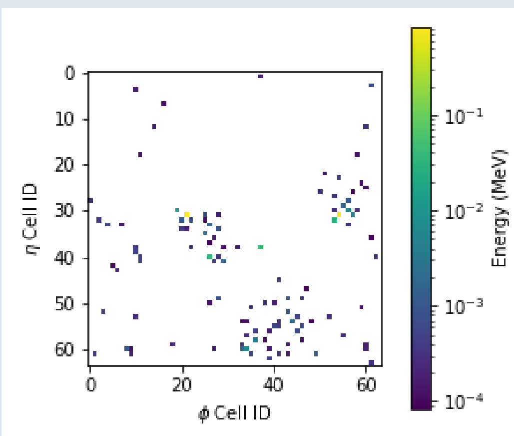
Event classification case study



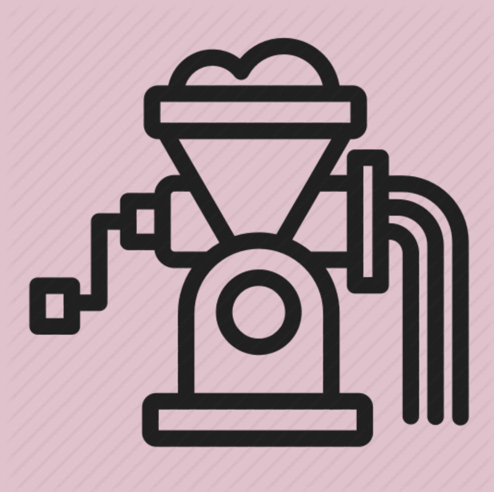
- Raw calorimeter readings binned into 64X64 images
- Pixel value = energy at corresponding location
- Each image = data of whole calorimeter (no cropping)
- Bottom row= normalized distribution of whole dataset

Supervised ML on Calo data

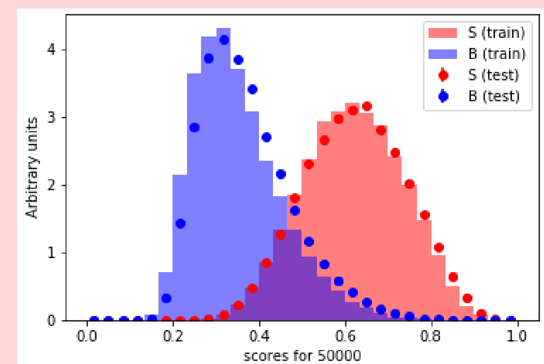
Calorimeter image
+ ground truth



Supervised
ML algorithm

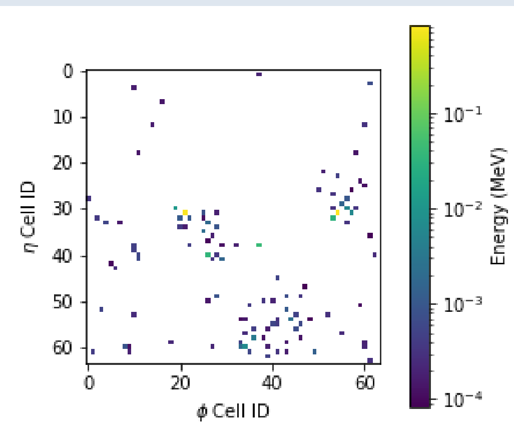


Signal / background
separation



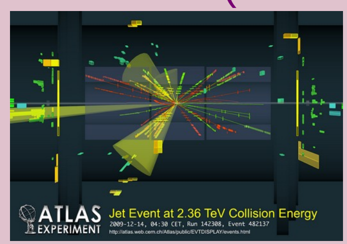
Beyond feature engineering

Calorimeter image
+ ground truth

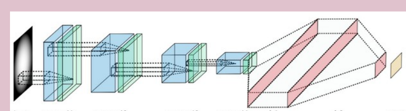


Supervised
ML algorithm

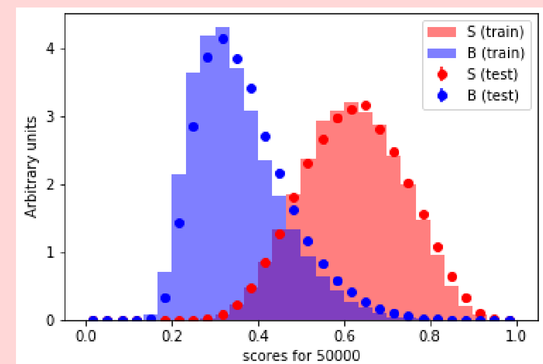
Feature engineering
Classical ML (BDT…)



Raw features
Modern ML (CNN...)

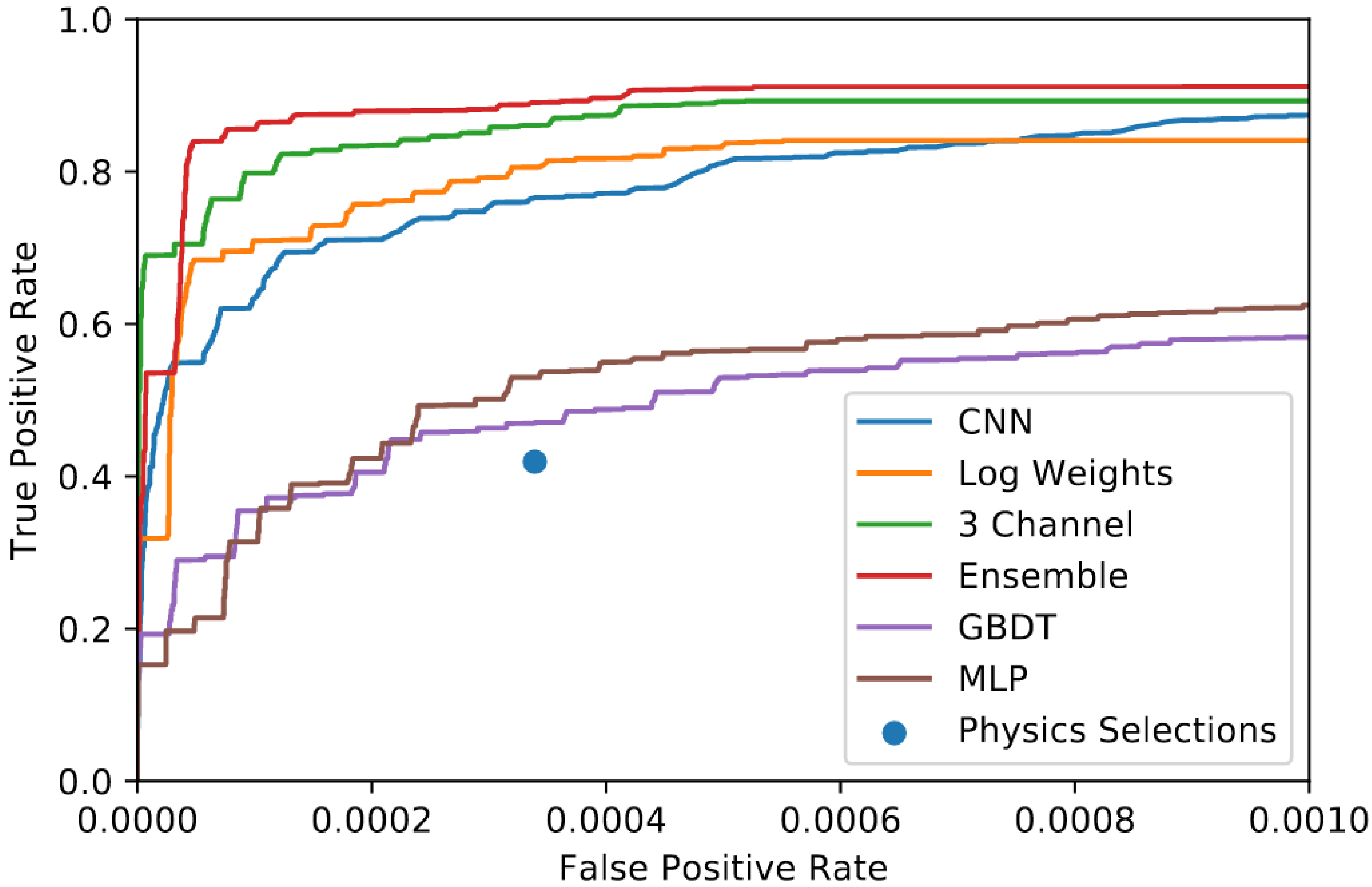


Signal / background
separation



See Joao Coelho talk from yesterday :
Calorimeter reconstruction with computer vision at LHCb
https://indico.in2p3.fr/event/20187/contributions/78787/

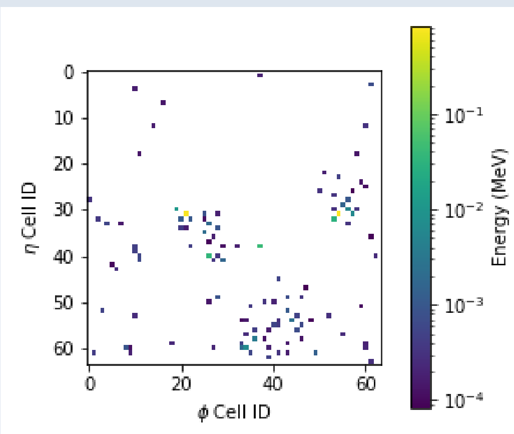
CNN results



ROC curve comparison of different CNN implementations
with physics selections and shallow classifiers (arXiv:1711.03573)

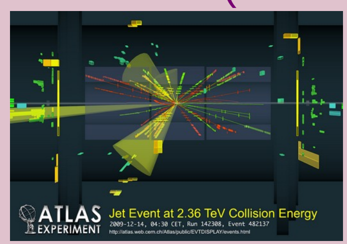
OPU competitive with CNN ?

Calorimeter image
+ ground truth



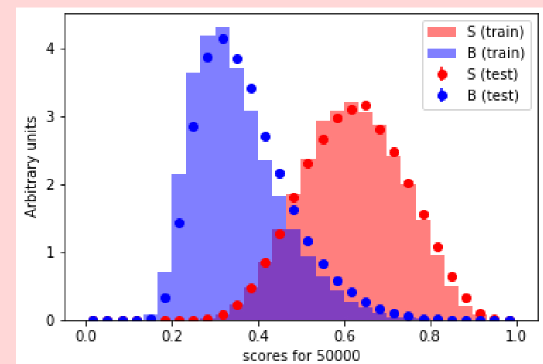
Supervised
ML algorithm

Feature engineering
Classical ML (BDT…)



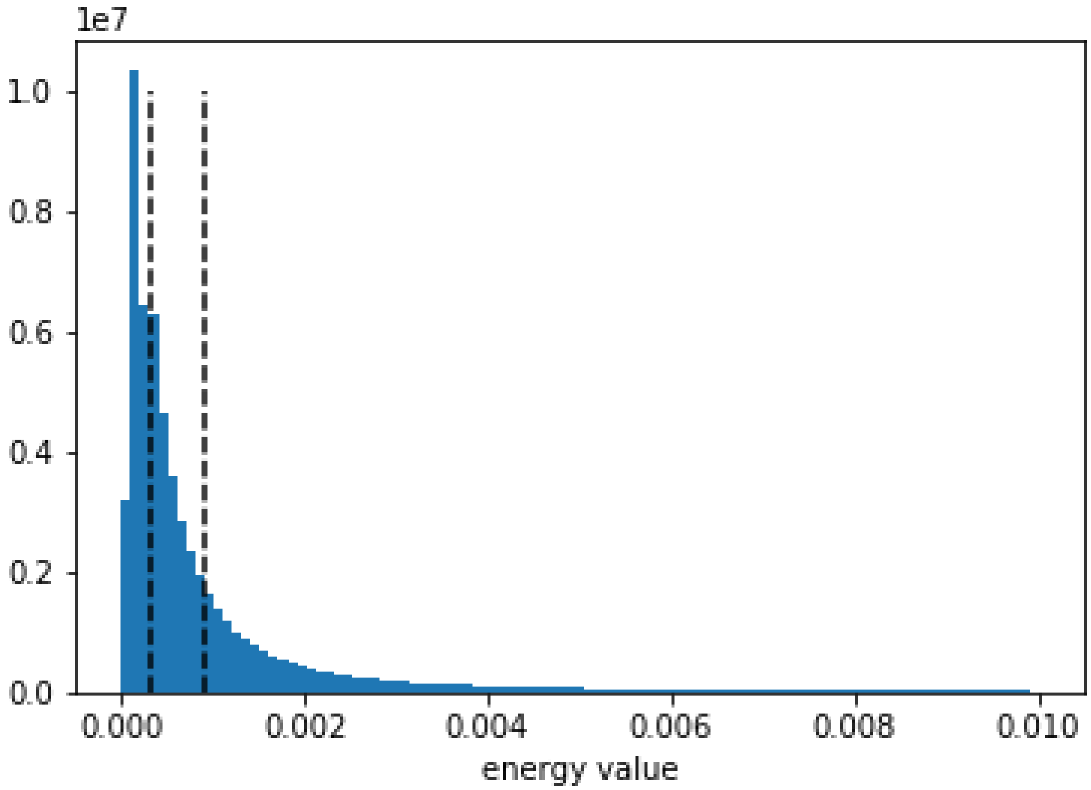
Raw features
Modern ML :
- CNN
- OPU ?

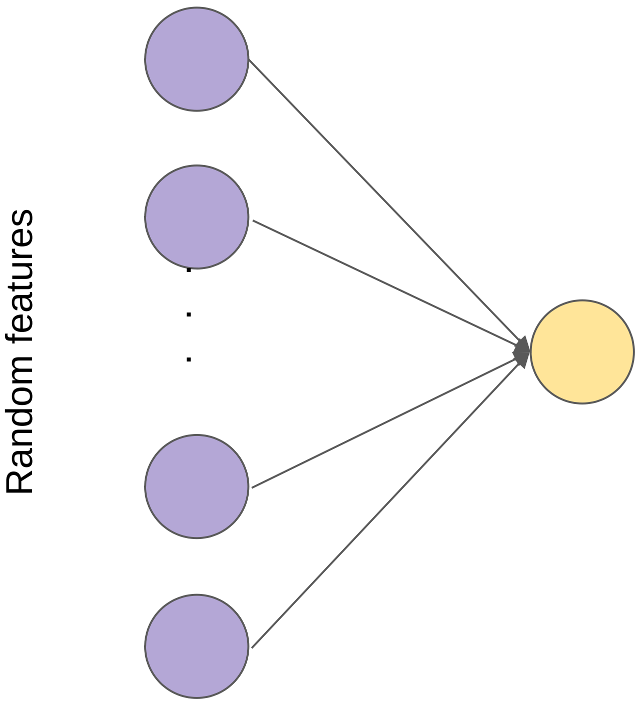
Signal / background
separation



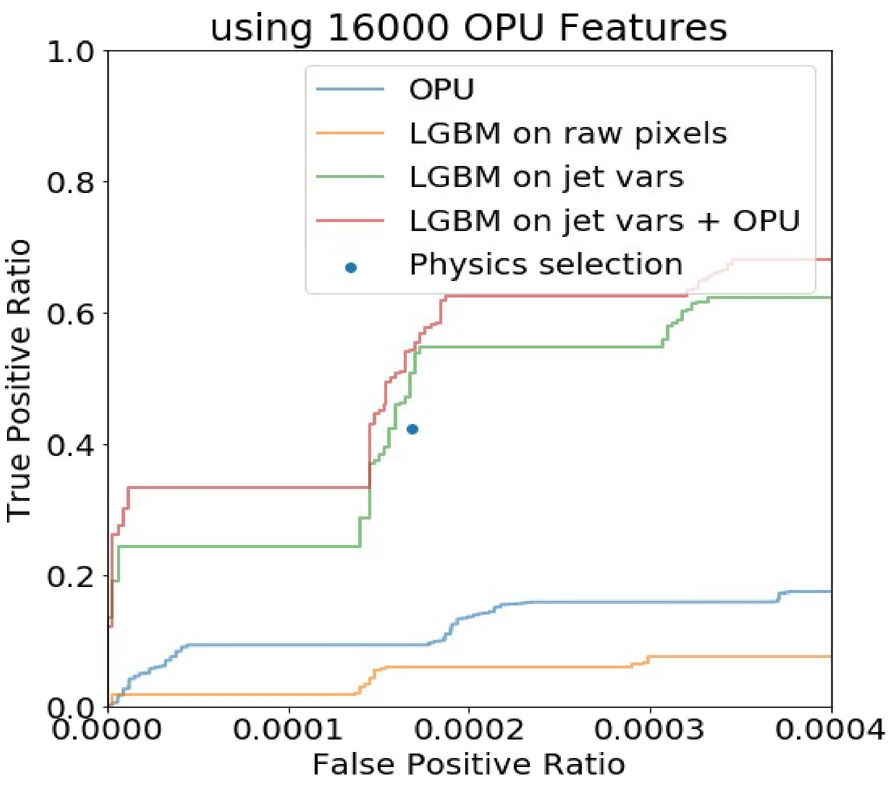
Modelization

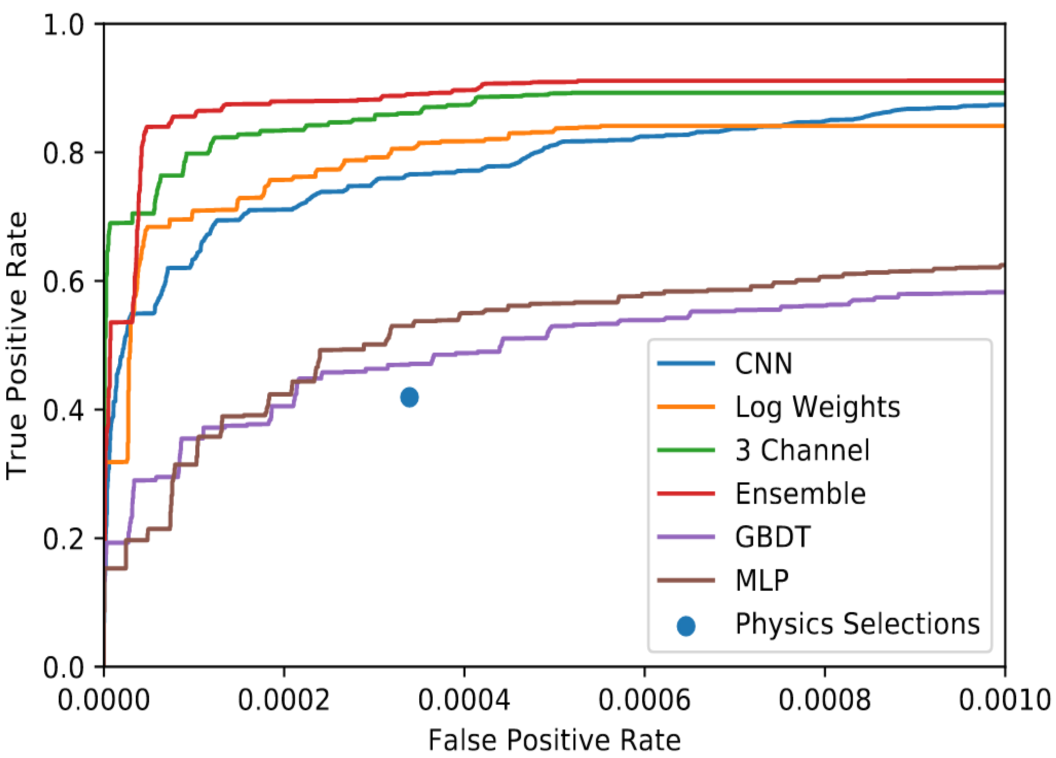
- Intensity-based binning (3 bits per pixel)
- Linear regression: single output node neural network.



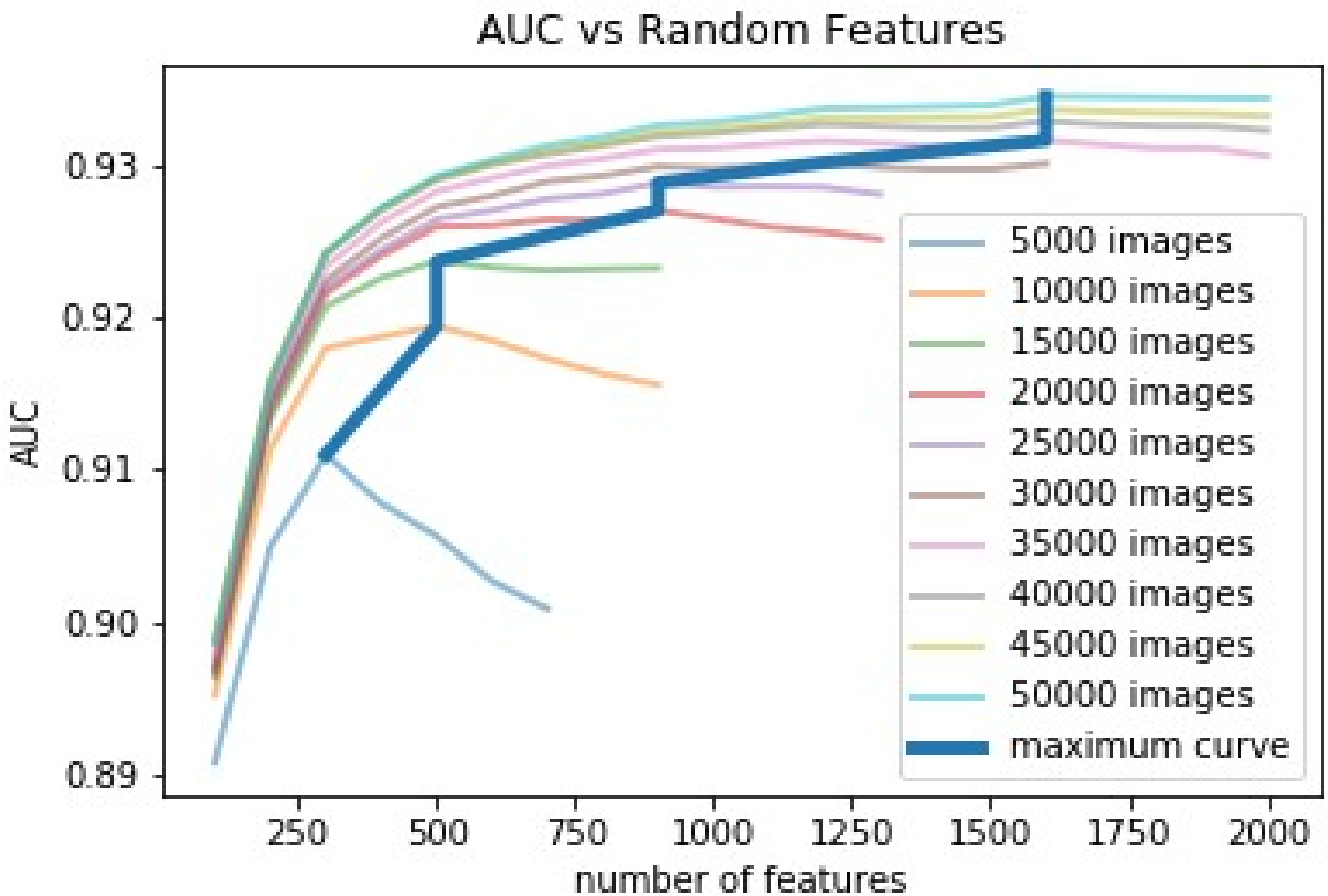


Comparison with CNN





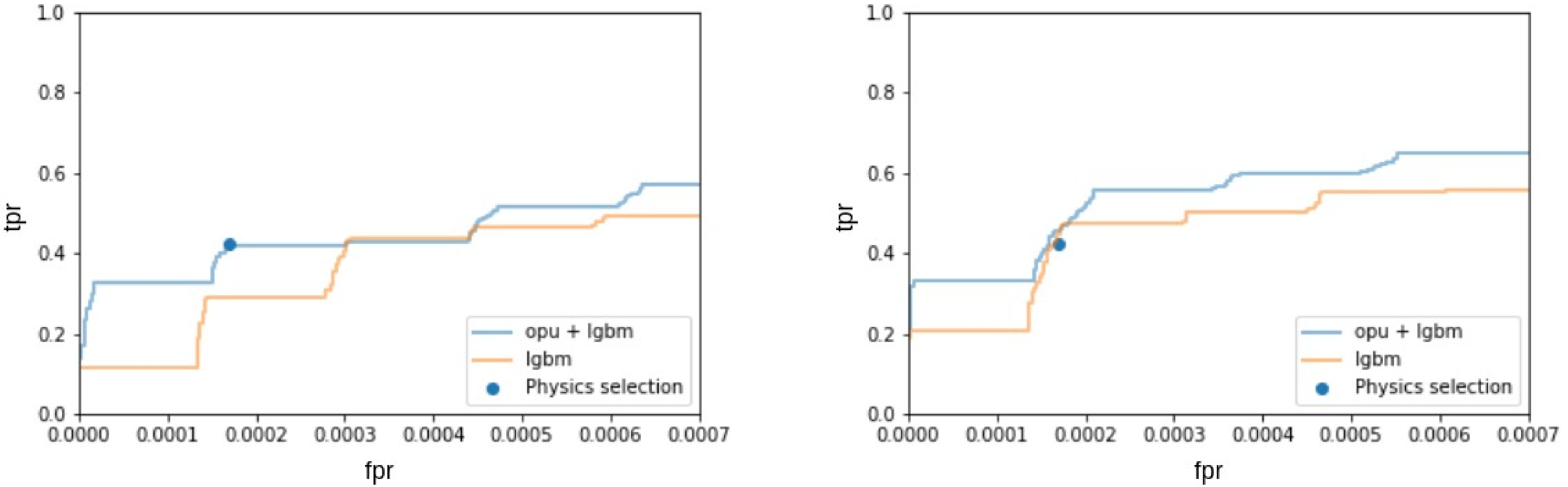
Optimal performance



- Optimal number of features increases with number of training images
- Even low number of images allows high accuracy

OPU utility

- NN require a large amount of training data
- OPU + BDTs scalable even when $N_{events} \simeq N_{pixels}$



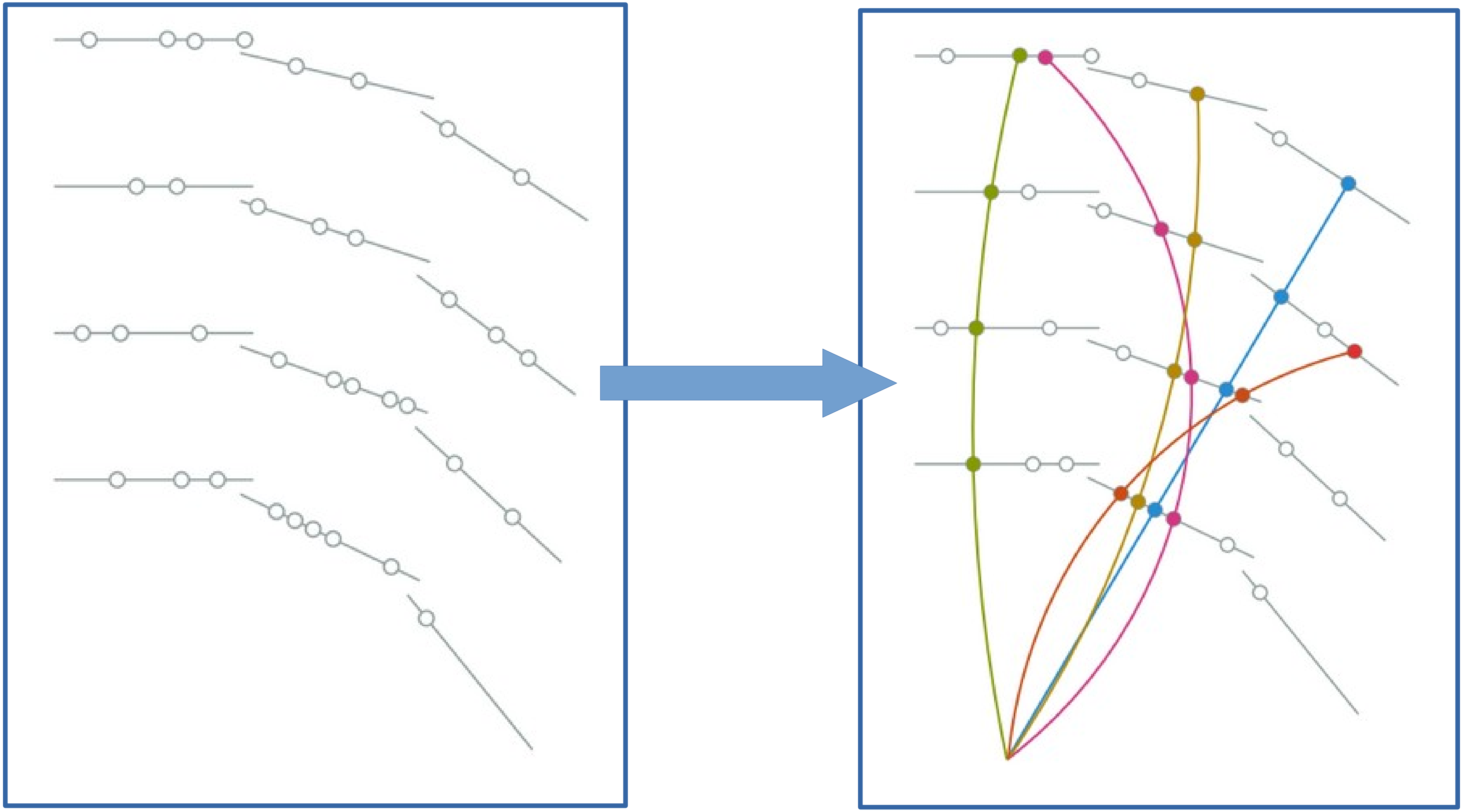
(a)4096 Training images (b)8192 Training images

Conclusions

- OPU provides physical device to reduce dimensionality / training time
- Use for detector tracking / calorimetry?
- Casting a Tracking problem for OPU is hard ; nonetheless estimations of
  - Single particle parameters (angle, inverse momentum)
  - Number of particles, position projected on next layer
  - OPU « makes sense » without matching traditional methods
- Calorimetry
  - Faster training than CNNs, far less training data, more robust
  - Performance not comparable to CNNs but fairly good even when $N_{features} \simeq N_{pixels}$
  - BDT can combine handcrafted variables with regression output from OPU random features
  - Outlook: extend to similar problems with finer granularity (arXiv:1807.00083)

Backup

The TrackML challenge: connect the dots

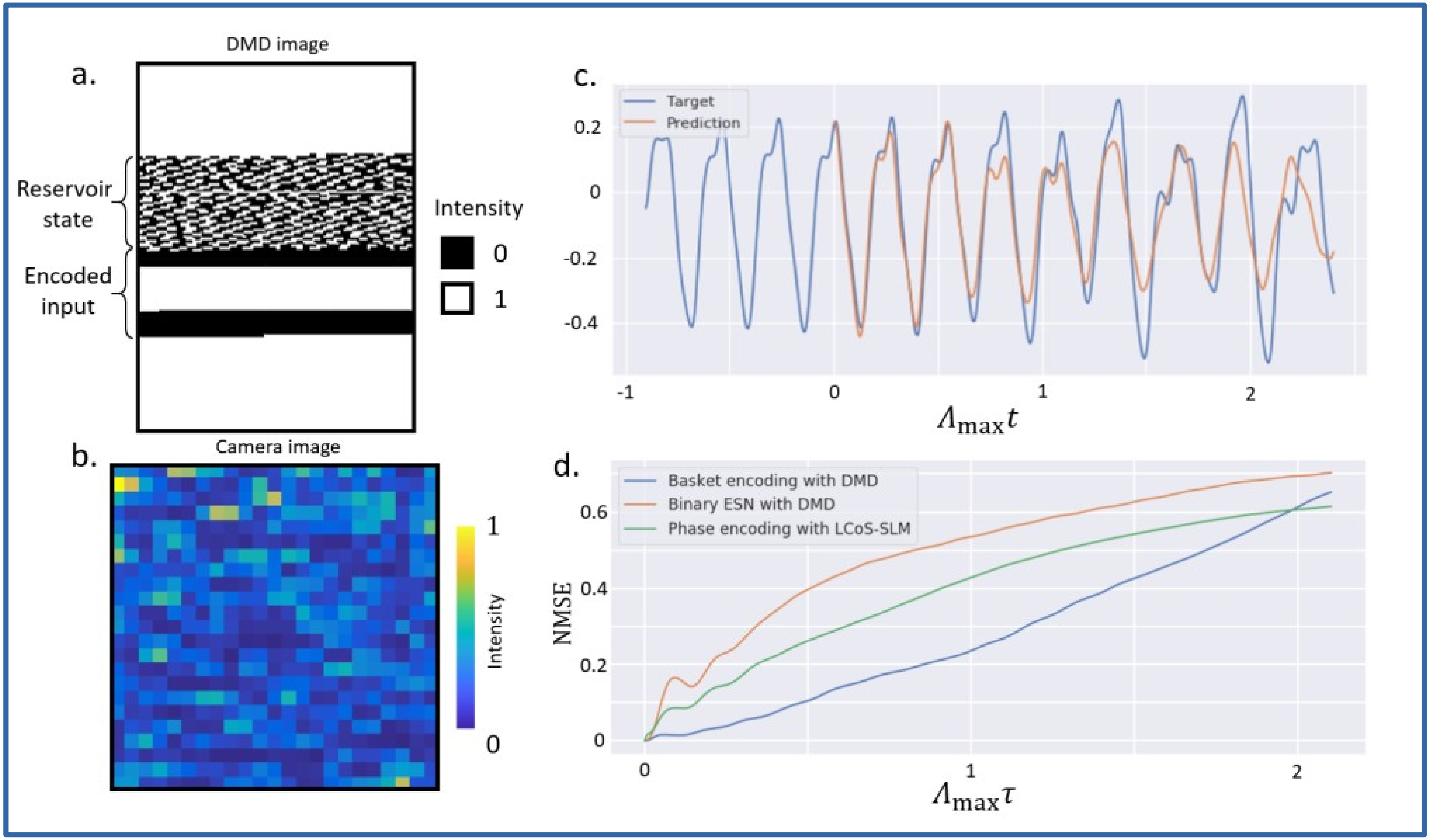


How to proceed?

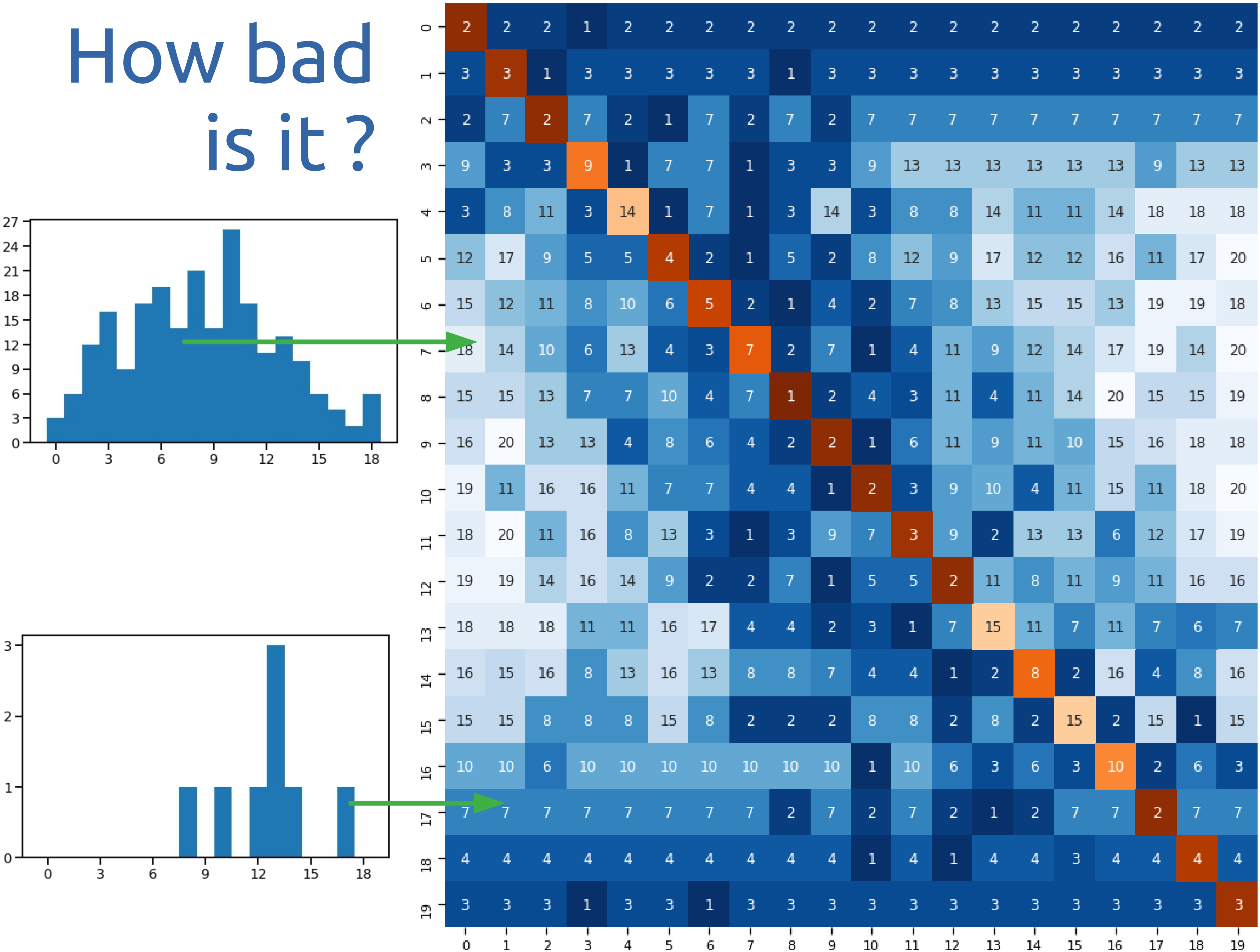
- Track following? No simple geometry of successive layers
- Compress the hits seen in electronics?
  - 2B electronic channels (!) → 1M OPU bits
  - Test with layered tSVD, autoencoders… didn't give anything interesting
- Use a more manageable dataset
  - Simplified dataset from RAMP challenge

Reservoir computing

Inspired by arxiv:1907.00657, J. Dong and al. : « Optical Reservoir Computing using multiple light scattering for chaotic systems prediction »

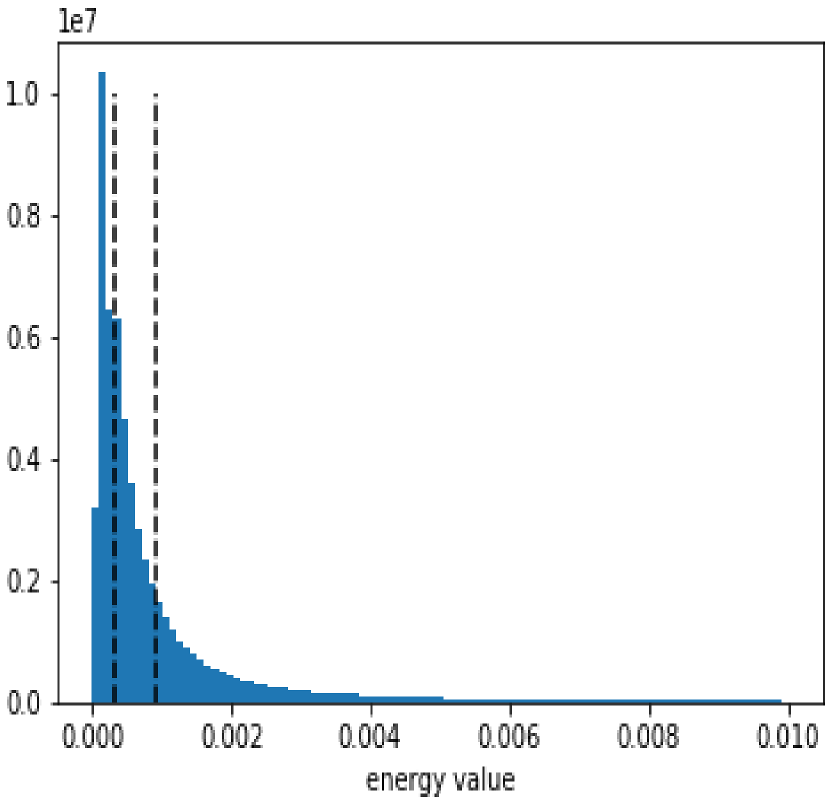


How bad is it ?



27 24 21 18 15 12 9 6 3 0
0 3 6 9 12 15 18
3 2 1 0
0 3 6 9 12 15 18
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 1 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 3
2 7 2 7 2 1 7 2 7 2 7 7 7 7 7 7 7 7 7 7
9 3 3 9 1 7 7 1 3 3 9 13 13 13 13 13 13 9 13 13
3 8 11 3 14 1 7 1 3 14 3 8 8 14 11 11 14 18 18 18
12 17 9 5 5 4 2 1 5 2 8 12 9 17 12 12 16 11 17 20
15 12 11 8 10 6 5 2 1 4 2 7 8 13 15 15 13 19 19 18
18 14 10 6 13 4 3 7 2 7 1 4 11 9 12 14 17 19 14 20
15 15 13 7 7 10 4 7 1 2 4 3 11 4 11 14 20 15 15 19
16 20 13 13 4 8 6 4 2 2 1 6 11 9 11 10 15 16 18 18
19 11 16 16 11 7 7 4 4 1 2 3 9 10 4 11 15 11 18 20
18 20 11 16 8 13 3 1 3 9 7 3 9 2 13 13 6 12 17 19
19 19 14 16 14 9 2 2 7 1 5 5 2 11 8 11 9 11 16 16
18 18 18 11 11 16 17 4 4 2 3 1 7 15 11 7 11 7 6 7
16 15 16 8 13 16 13 8 8 7 4 4 1 2 8 2 16 4 8 16
15 15 8 8 8 15 8 2 2 2 8 8 2 8 2 15 2 15 1 15
10 10 6 10 10 10 10 10 10 10 1 10 6 3 6 3 10 2 6 3
7 7 7 7 7 7 7 7 2 7 2 7 2 1 2 7 7 2 7 7
4 4 4 4 4 4 4 4 4 4 1 4 1 4 4 3 4 4 4 4
3 3 3 1 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Encoding Scheme

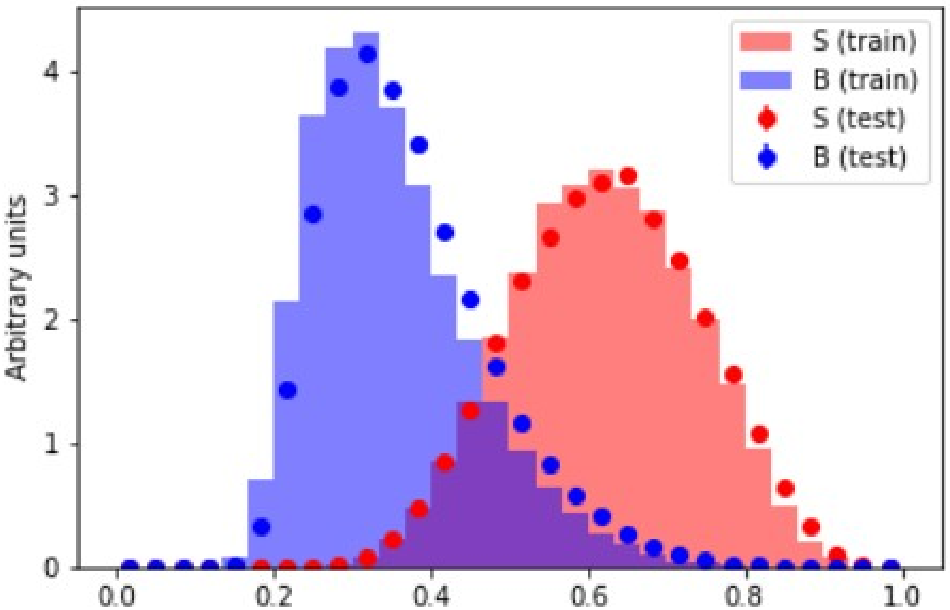


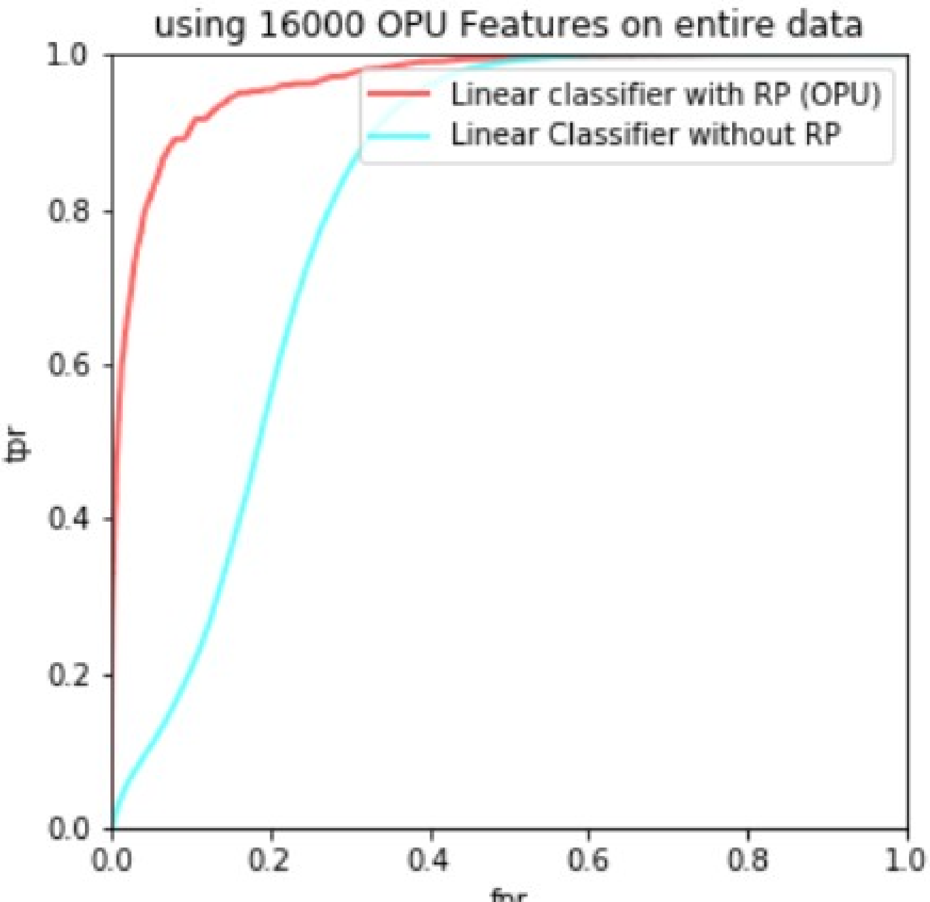
Distribution of energy (Excluding zeros)

Pixel (energy) value	encoding
x = 0	000
x>0 and x<=.00031528	001
$x>3.1528*10^{-4}$ and $x<=9.1565*10^{-4}$	011
$x>9.1565*10^{-4}$	111

The intensity based binning performed much better then auto-encoders

Predictions using OPU





Estimate next layer hits number

$$\min_{\beta \in R^{m\times n}} ||X\beta - y|| + ||\gamma \beta||$$

(10K) random features
X (10K) events

(1) last layer hit number
X (10K) events