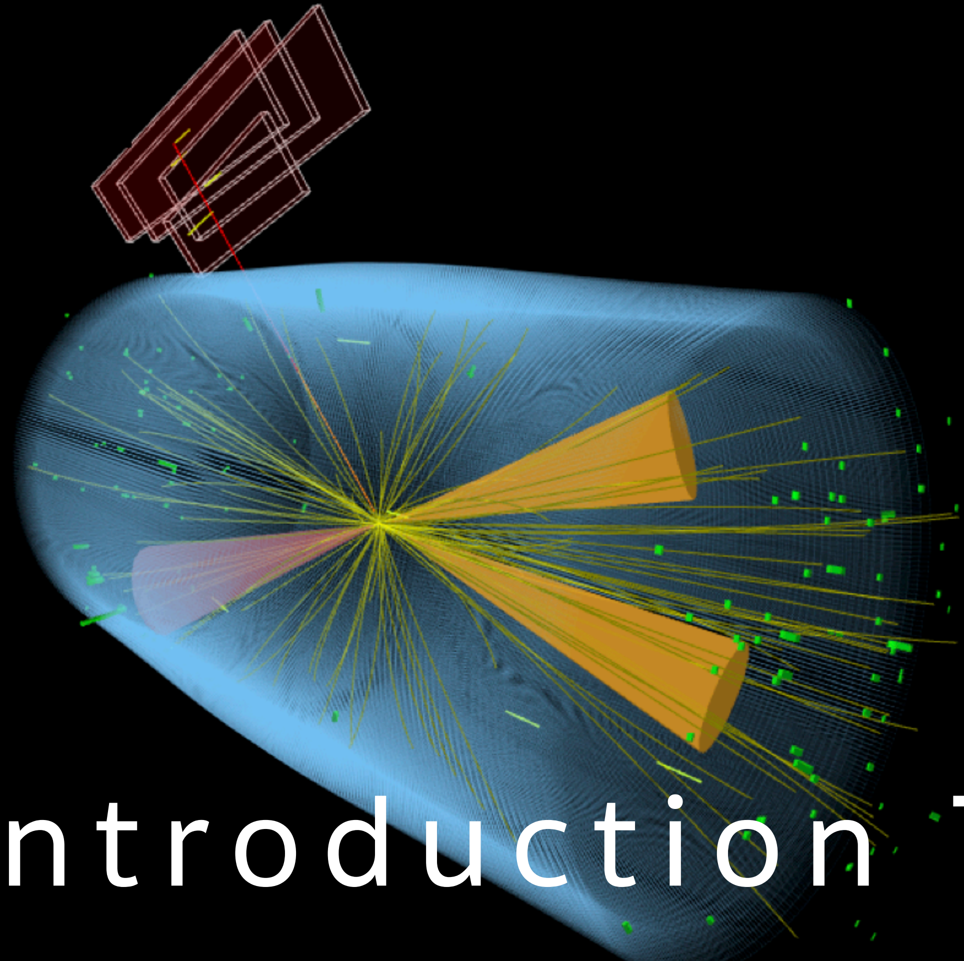
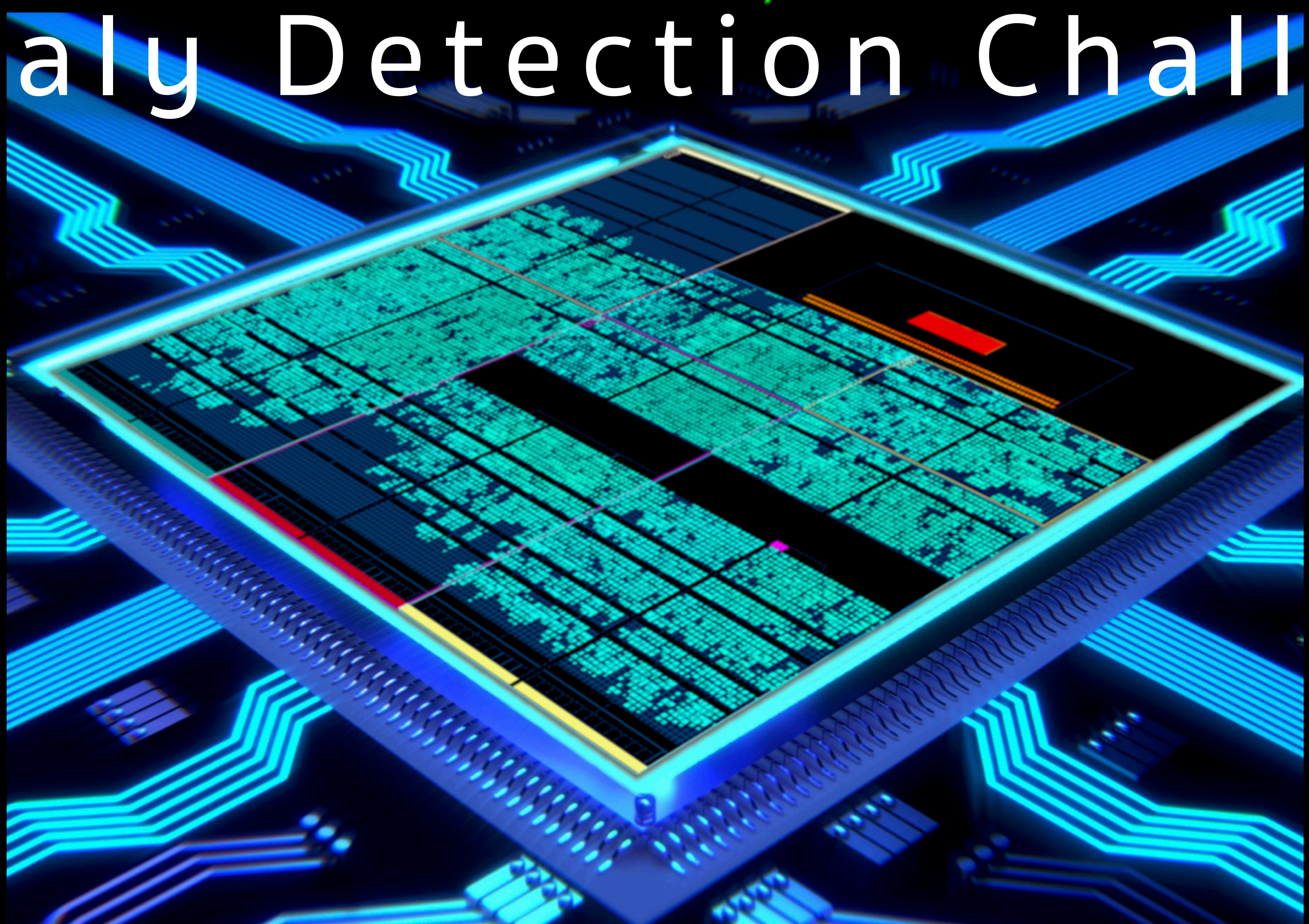


Introduction To Anomaly Detection Challenge



Thea Arrestad[1]

Katya Govorkova[1]

Jennifer Ngadiuba[2]

Maurizio Pierini[1]

Ema Puljak[1]

Kinga Wozniak[1]

[1]CERN [2]FNAL

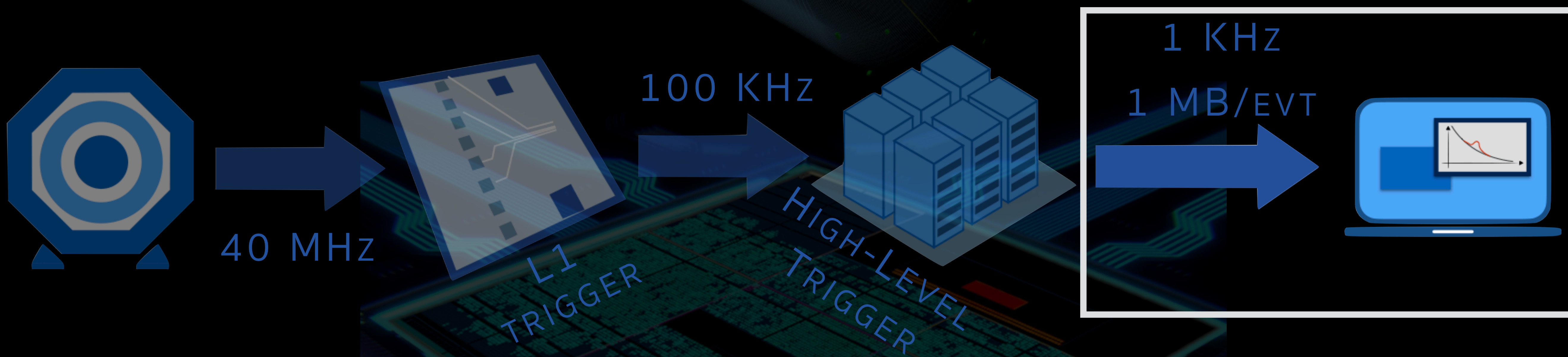
ML④Jets2021

7 July 2021

Unsupervised New Physics detection at 40 MHz
Challenge

Usual Search for **New Physics** is performed by designing a **dedicated trigger** for a specific signal hypothesis and performing complicated statistical analysis offline

But what if you **do not** have a **signal hypothesis**?...

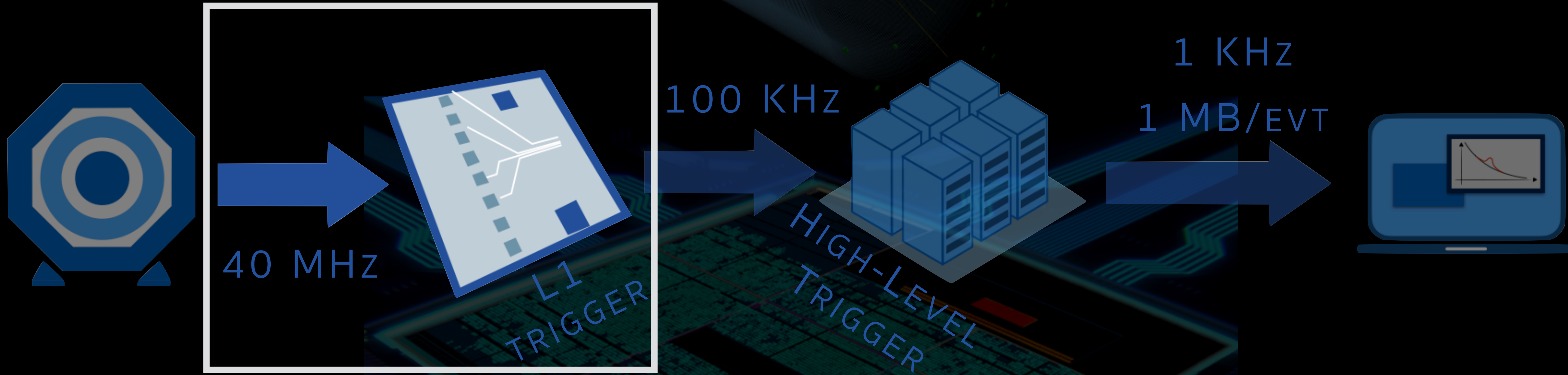


A lot of **Unsupervised** algorithms for **offline** processing — LHC Olympics, The Dark Machines, etc

Unsupervised New Physics detection at 40 MHz Challenge

Idea is to **look for** something **very rare and unusual** directly in the **Level-1 Trigger** without any signal hypothesis in mind

The challenge is to find a-priori **unknown** and **rare New Physics** hidden in a data sample dominated by ordinary Standard Model processes



The deliverable is a developed **algorithm** that can be deployed and run **in L1** with strict **latency** requirement of **< 1 microsecond**

The task is therefore to design an architecture that maximises the **sensitivity for New Physics** but at the **lowest** possible **resource** and **latency** budget

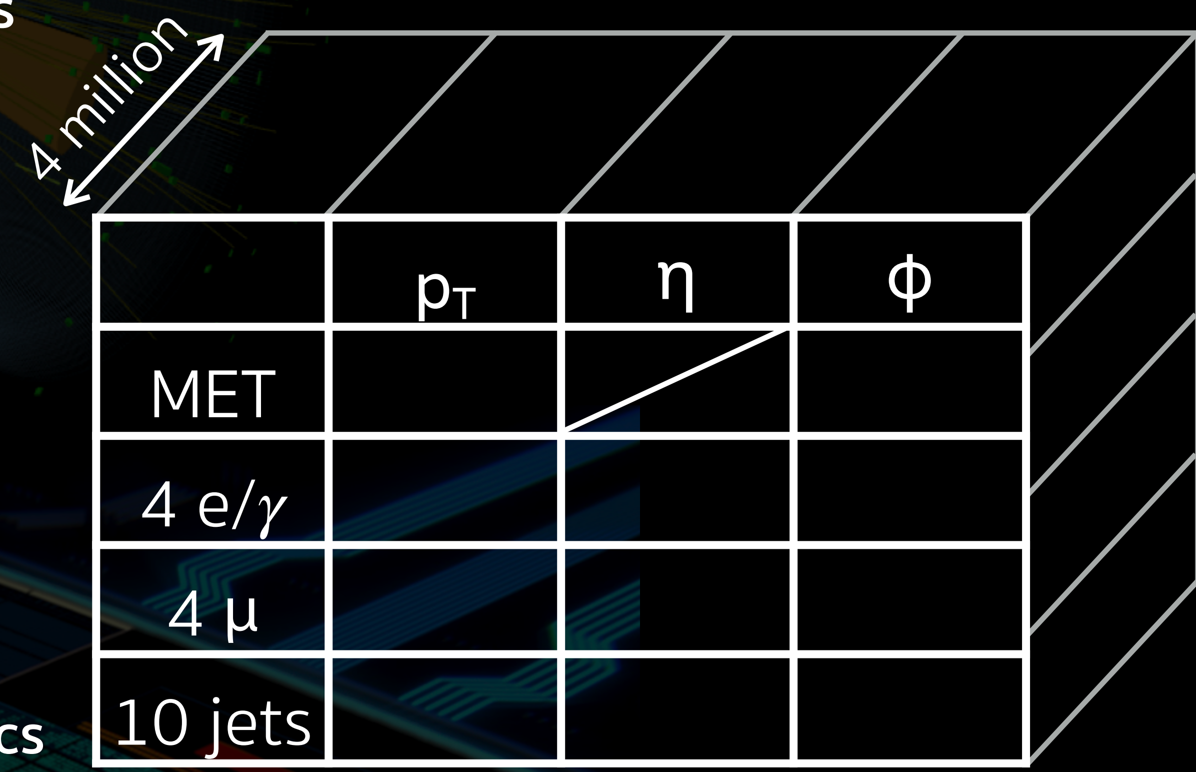
Data Samples

The data is represented as an array of MET, up to 4 e/γ, 4 μ and 10 jets each described by p_T, η and ϕ to **mimic L1 data format**

Train using provided **4 million background-like events** simulated with Delphes 💾

Events are **pre-filtered** to have **at least one lepton**

- Inclusive W production, with W → lν (59.2%)
- Inclusive Z production, with Z → ll (6.7%)
- tt production (0.3%)
- QCD multijet production (33.8%)



Evaluate performance on several different **New Physics simulated samples**

- Neutral scalar boson A, 50 GeV → 4 l 💾
- Leptoquark, 80 GeV → b τ 💾
- Scalar boson, 60 GeV → τ τ 💾
- Charged scalar boson, 60 GeV → τ ν 💾
- Black Box 🎁

Unsupervised New Physics detection at 40 MHz

Step-by-Step

- **Gather** in a team/by yourself
- **Get a cool name** for your team, for example team " **DeepAnomaly** " 👨‍👨‍👦
- Get yourself familiar with the **details on the challenge webpage**
- **Investigate** available **datasets** and **example codes**

SM Dataset: Training	BSM Dataset: A to 4 leptons	BSM Dataset: h^+ to τν
Blackbox Dataset	BSM Dataset: LQ to bτ	BSM Dataset: h^0 to ττ

- **Design** your AD **model**
- **Evaluate performance** and submit **results**
- **Best models** will be **published** in a White Paper (and perhaps deployed in L1 trigger of CMS!!)

Unsupervised New Physics detection at 40 MHz

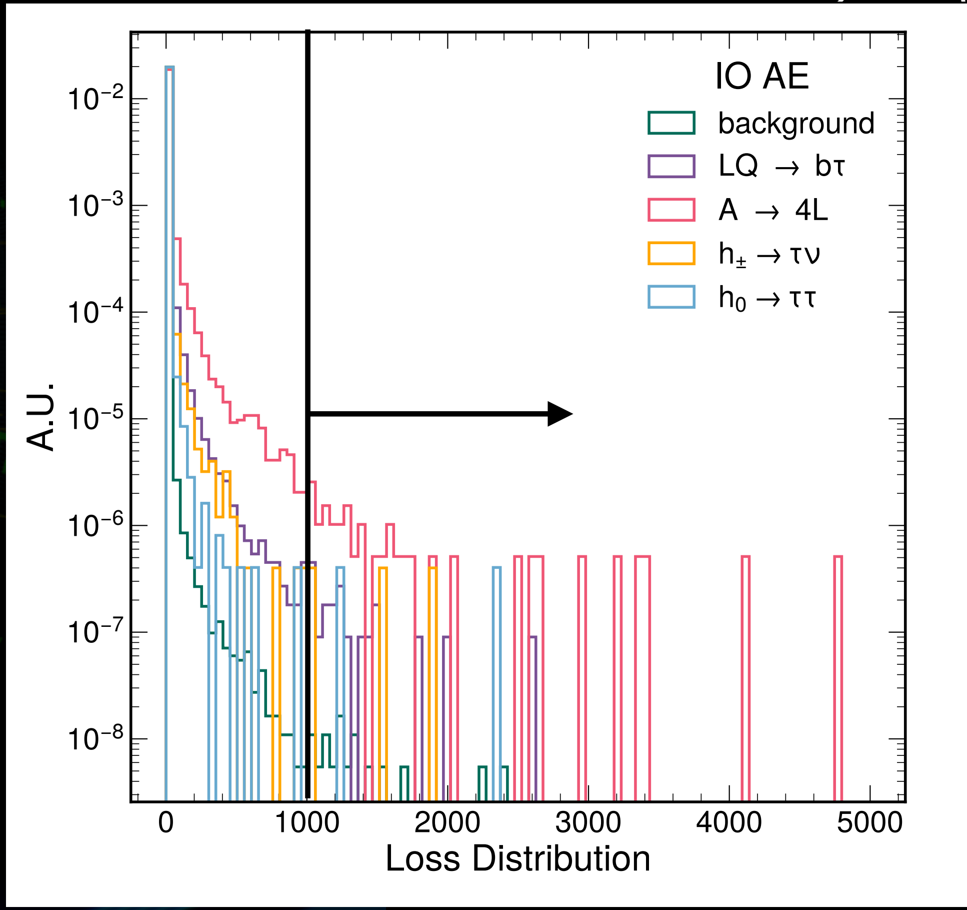
Example Team " DeepAnomaly " 👨‍👨‍👦

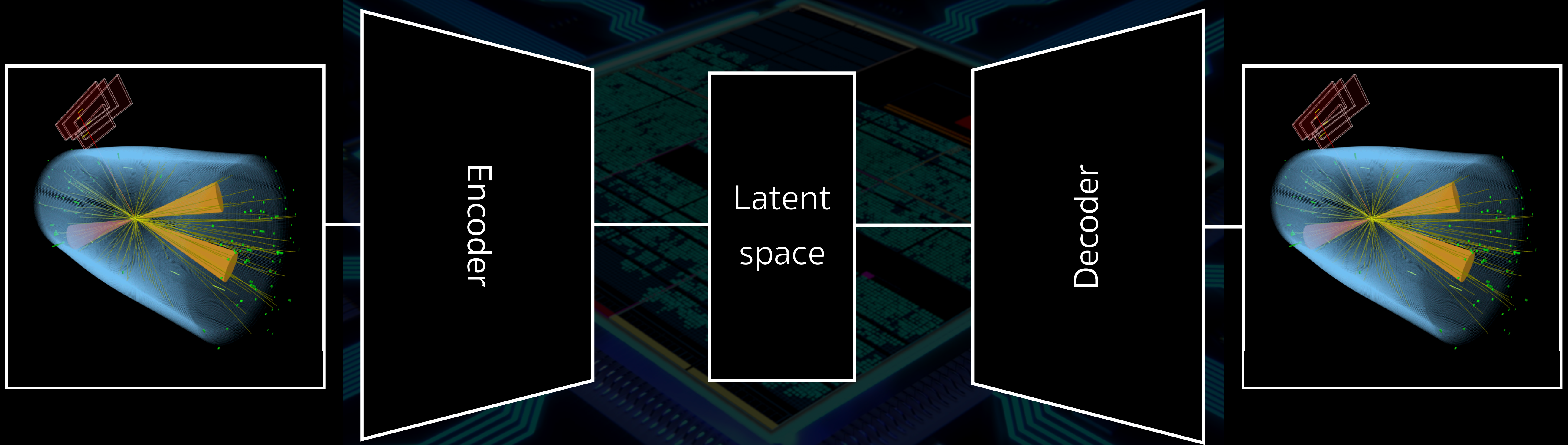
Designs algorithm

Decide on the **algorithm** that you want to explore

The Example Team " **DeepAnomaly** " has chosen **Autoencoders**

- **Encode** input in **smaller** dimensional **space**
- **Train** on typical LHC **background**
- **Anomalous** data will have **higher loss**
- **Calculating** the loss requires to **store the input** until the output is computed

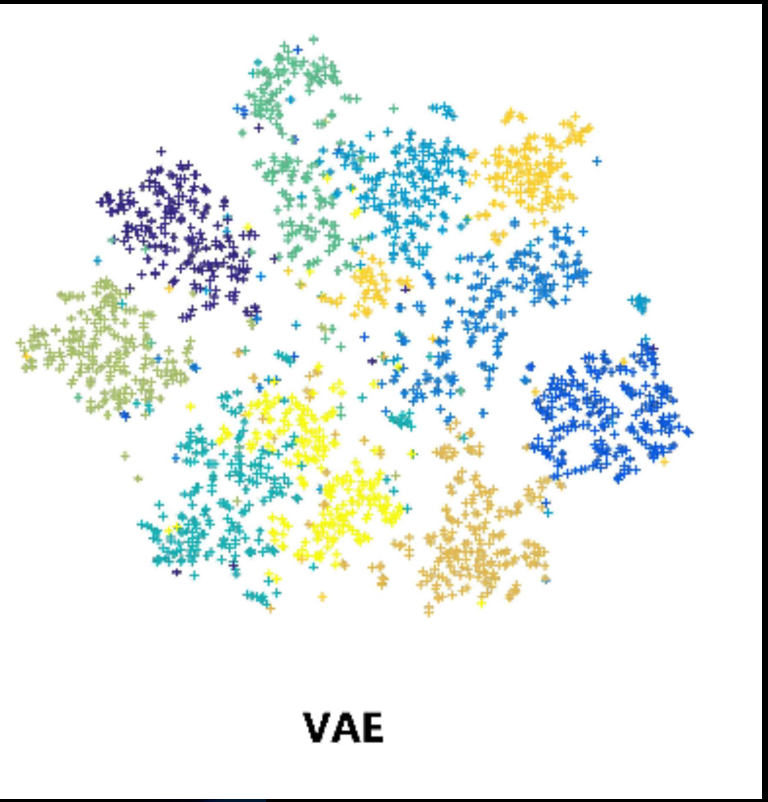




Unsupervised New Physics detection at 40 MHz

Example Team " DeepAnomaly " 👨‍👨‍👦

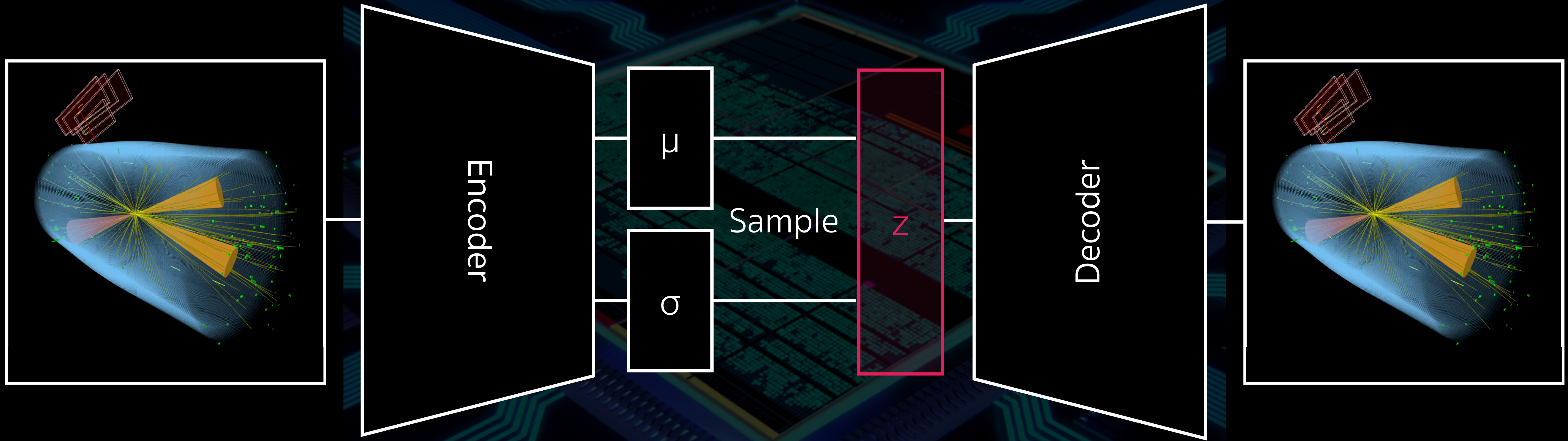
Designs algorithm



The " **DeepAnomaly** " team has also considers **Variational Autoencoders**

- **The latent space** is sampled from Encoder output
- Can be used **to generate new samples**
- **Inference** can be done only **on the latent space**
- **No** need to store **input** and deployment of **Encoder is enough**

  (e.g. saves resources and latency in comparison to AE)



Unsupervised New Physics detection at 40 MHz

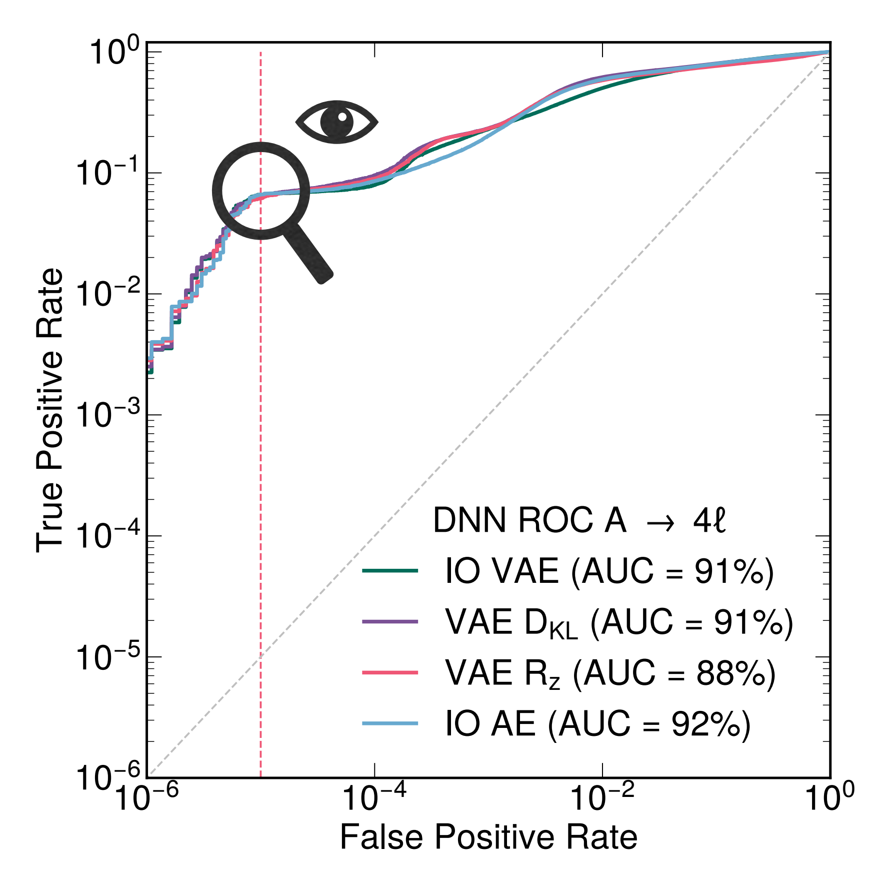
Example Team " DeepAnomaly " 👨‍👨‍👦

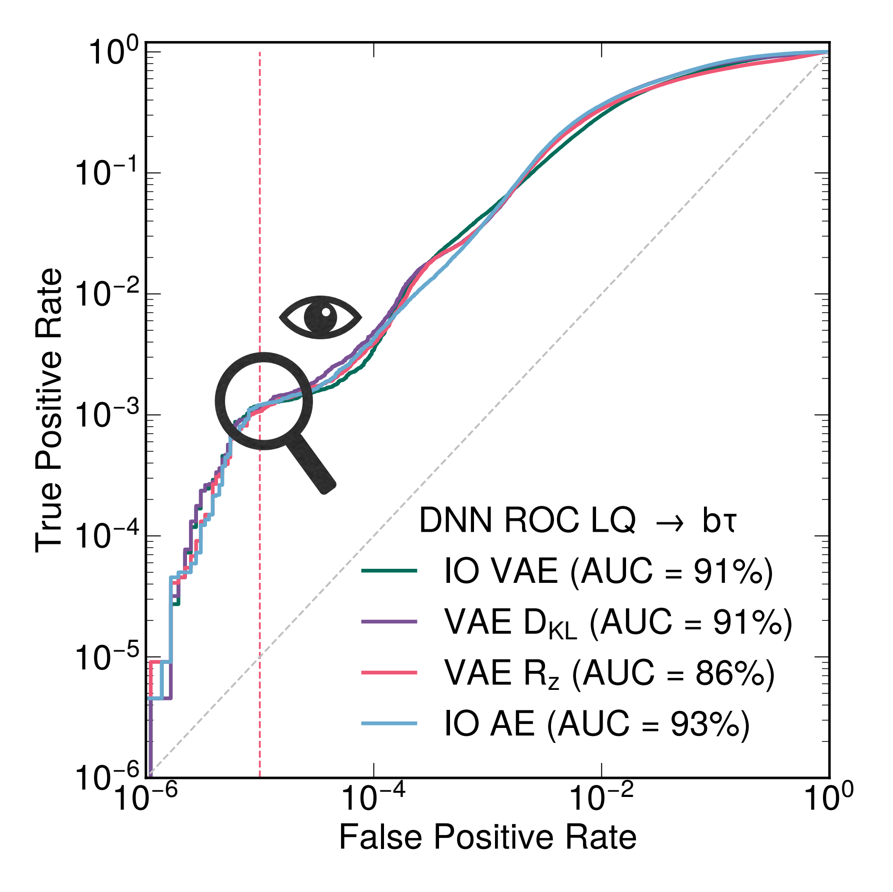
Evaluates Performance on Signal samples

Goal is to **maximise** TPR at FPR 10^{-5} (roughly corresponding to the available output data rate budget for a trigger algorithm) for each of the provided anomaly

The Team " **DeepAnomaly** " checks AE vs VAE

- The **Inference** can be done only **on the latent space**
- **No** need to store **input** and deployment of **Encoder is enough**





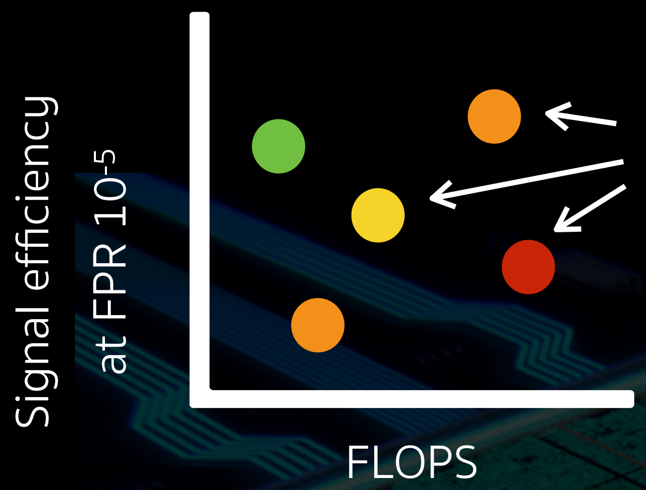
Unsupervised New Physics detection at 40 MHz

Example Team " DeepAnomaly " 👨‍👨‍👦

Submits results

- For the **Blackbox** dataset get the event number of the **1000 most anomalous** events based on the algorithm metrics

- An estimate of the **algorithm efficiency** can be obtained by calculating **the floating-point operations per second (FLOPS)** 🔗



Different algorithm performance

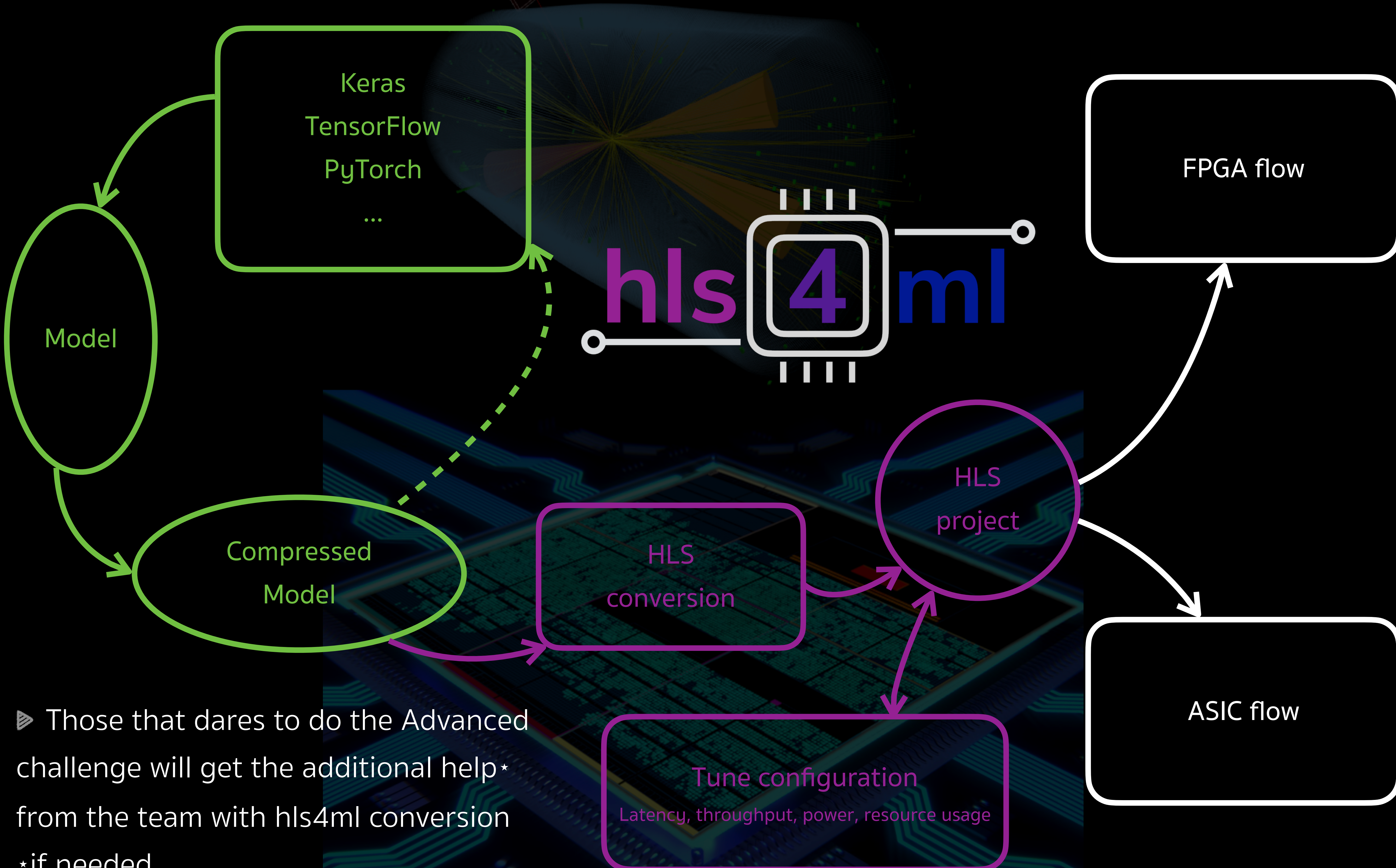
The best one is chosen by min(FLOPS) & max(sig eff)

- **The submission** should be in a form of a HDF5 file, DeepAnomaly.h5, containing a **numpy array** with the **identification numbers** of each selected event, plus a **dictionary** with the algorithm deployment **performance** 🔗

Upload contribution!

Unsupervised New Physics detection at 40 MHz

✪ Advanced Challenge



- Those that dares to do the Advanced challenge will get the additional help⋆ from the team with hls4ml conversion

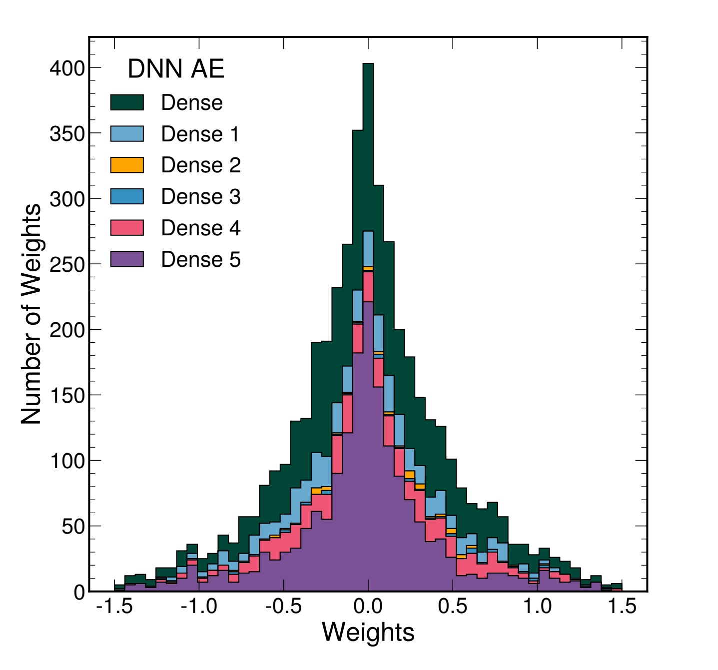
⋆if needed

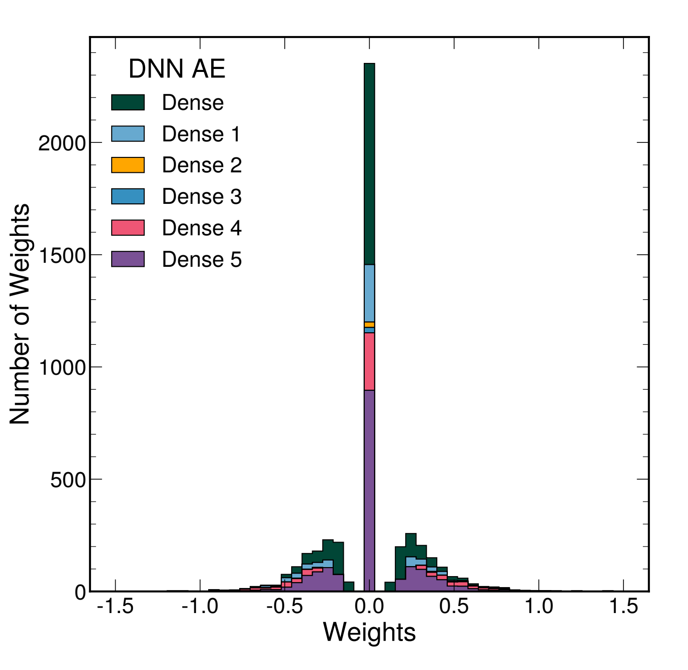
Unsupervised New Physics detection at 40 MHz

Example Team " DeepAnomaly " 👨‍👨‍👦

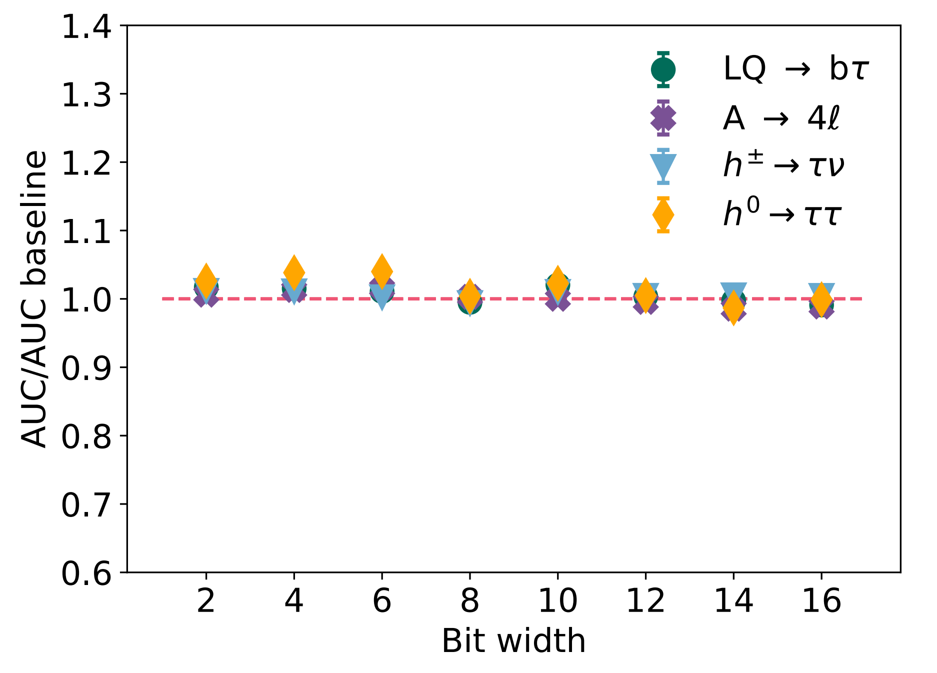
✪ Participates in Advanced Challenge

- Prune the model with the target sparsity of 50%





- Use Quantization Aware Training (QAT) with QKeras to quantize the model



- Use hls4ml to convert model to firmware and evaluate latency and performance

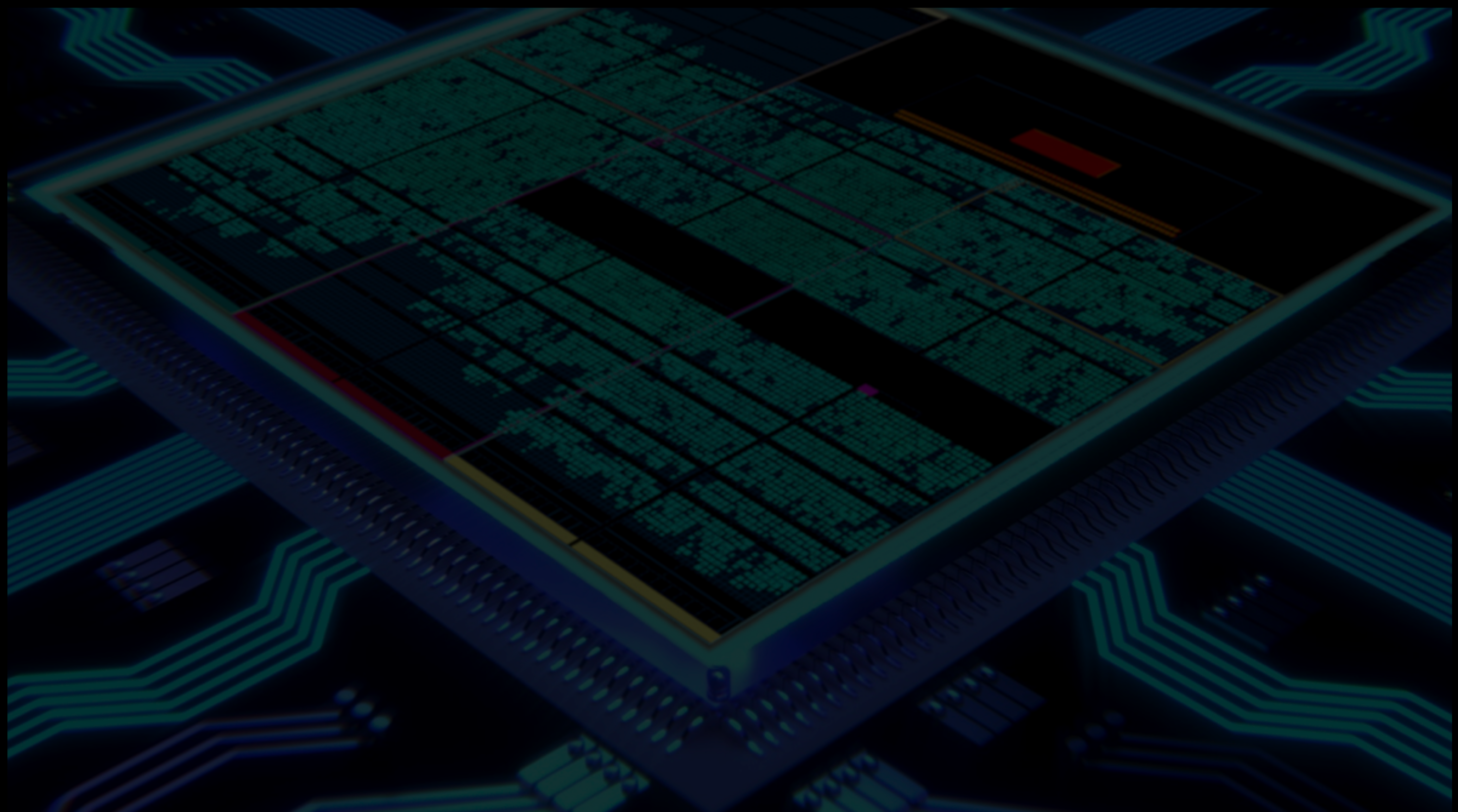
Model	Hardware	DSP[%]	LUT[%]	FF[%]	BRAM[%]	Latency[ns]
DNN AE	Xilinx VU9P	7	9	0.5	0.00	70
DNN VAE	Xilinx VU9P	2	3	0.2	0.05	45

- Results shown for " DeepAnomaly " are soon to be published

All the information is available at

Good luck!! 🚀

And remember New Physics might be just behind the corner…



And if you still have questions

📩

ekaterina.govorkova@cern.ch
ema.puljak@cern.ch
jennifer.ngadiuba@cern.ch
maurizio.pierini@cern.ch
thea.aarrestad@cern.ch