How machine learning changes particle physics

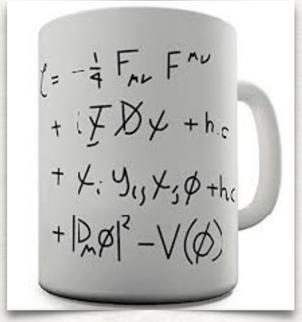
MIHOKO M. NOJIRI (KEK)
ISCO2023 MARCH 1, 2023

Particle physics

- Particle Physics has been answering big questions, and it keeps challenging remaining big questions
- Example
  - Higgs boson (explaining origin of mass)
  - Dark matter in the Universe
  - Origin of the matter

  and they are Science!
- Big Science connects people! ex CERN LHC ATLAS experiment: 3000 researcher 1200 PhD 42 countries, 182 Institutions

HOW PARTICLE PHYSICS RESEARCHWORKS (DATA AND THEORY)



SM lagrangian

+

Prediction of yet unseen particles

" New Physics"

A few very precise data

Field Theory
Quantum Mechanics
+ general relatively

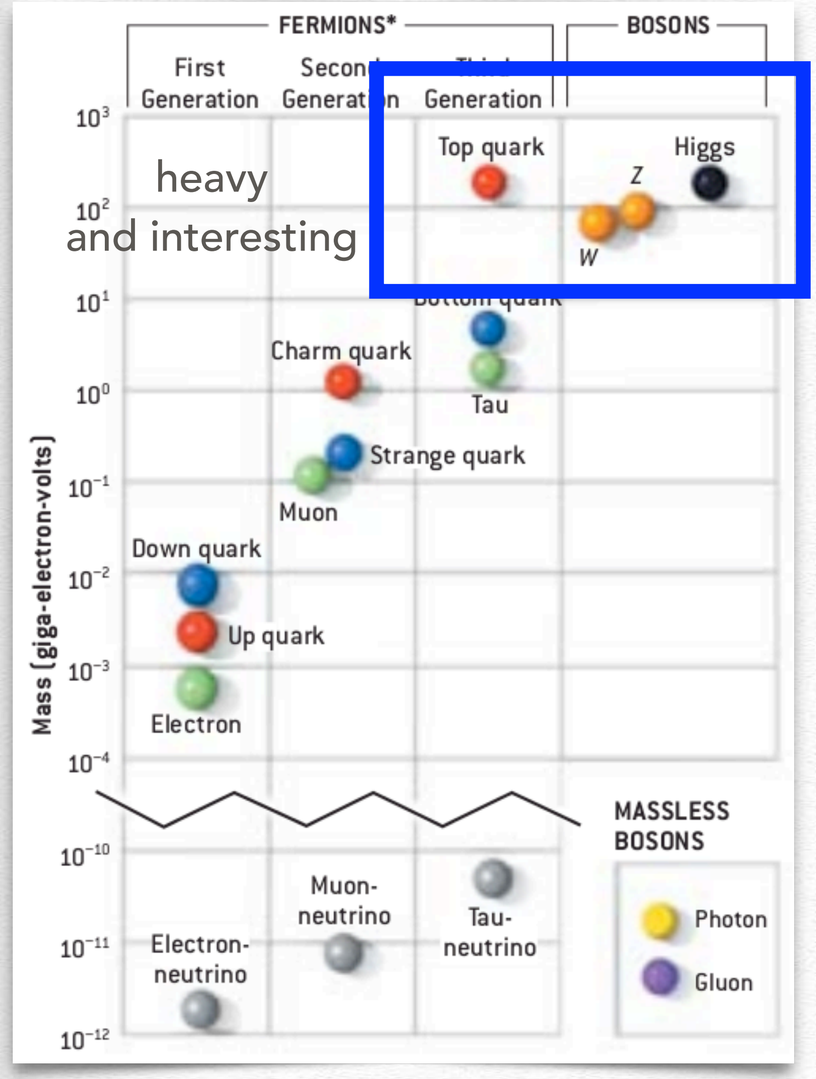
Other data

Phenomena at Early Universe
.....

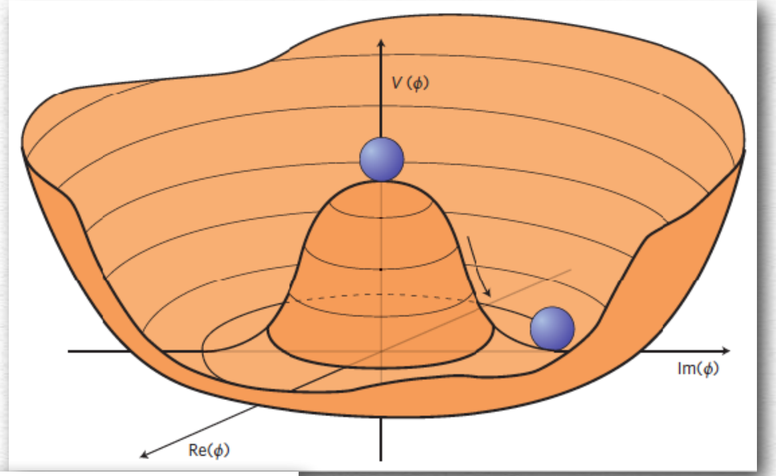
Predictions

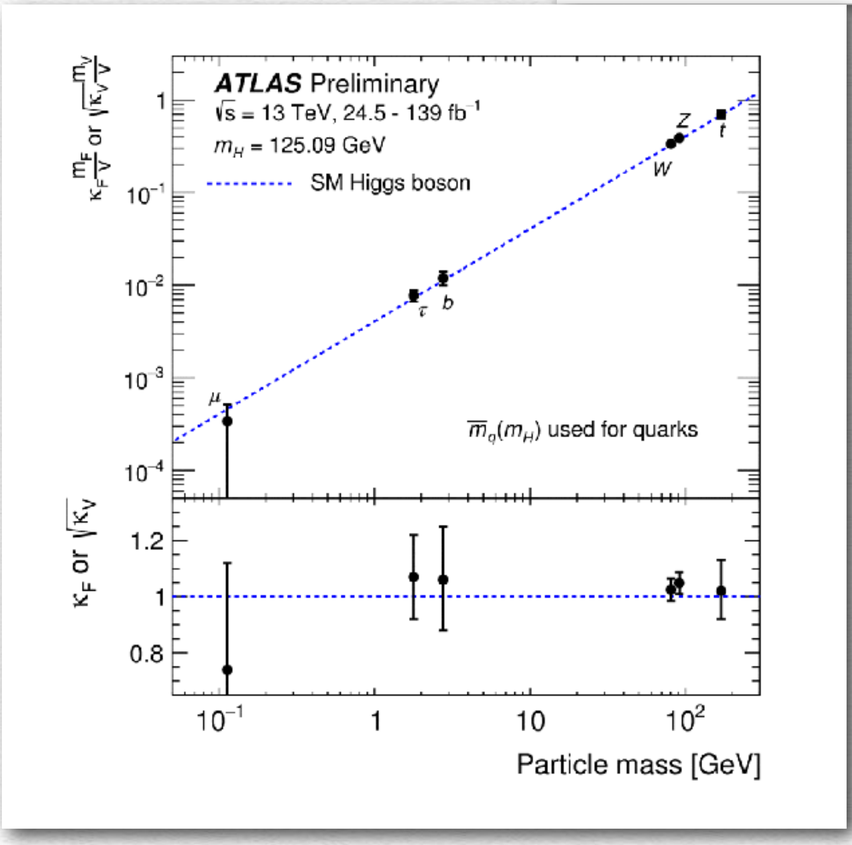
Interaction of particles at any scale

LHC Higgs discovery and its nature



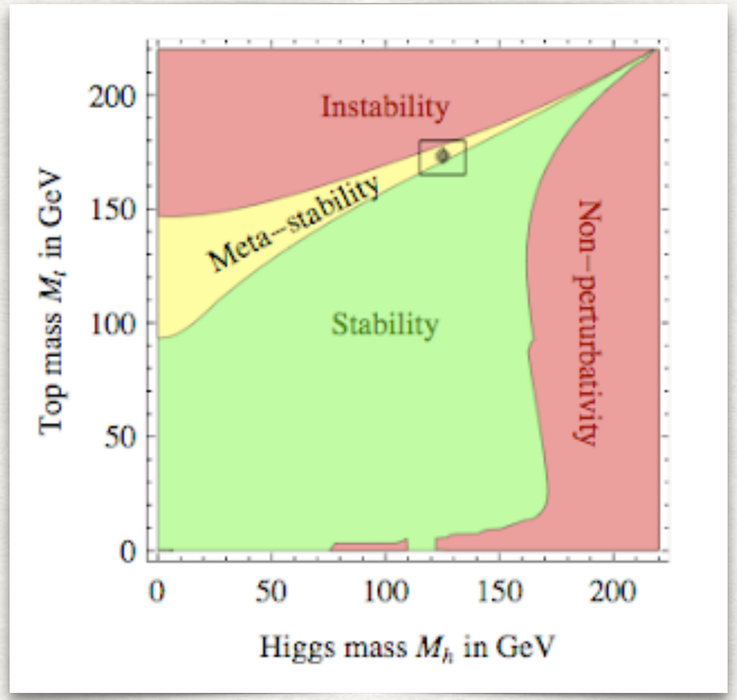
Higgs order param. v
"yukawa coupling " y
"particle mass" ~ yv

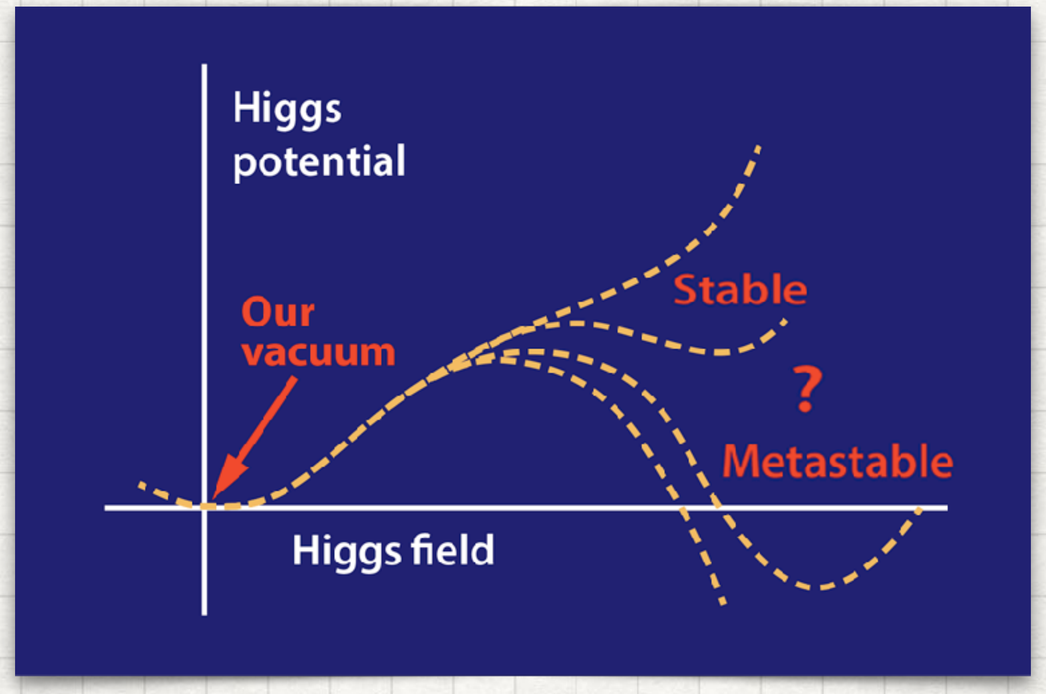


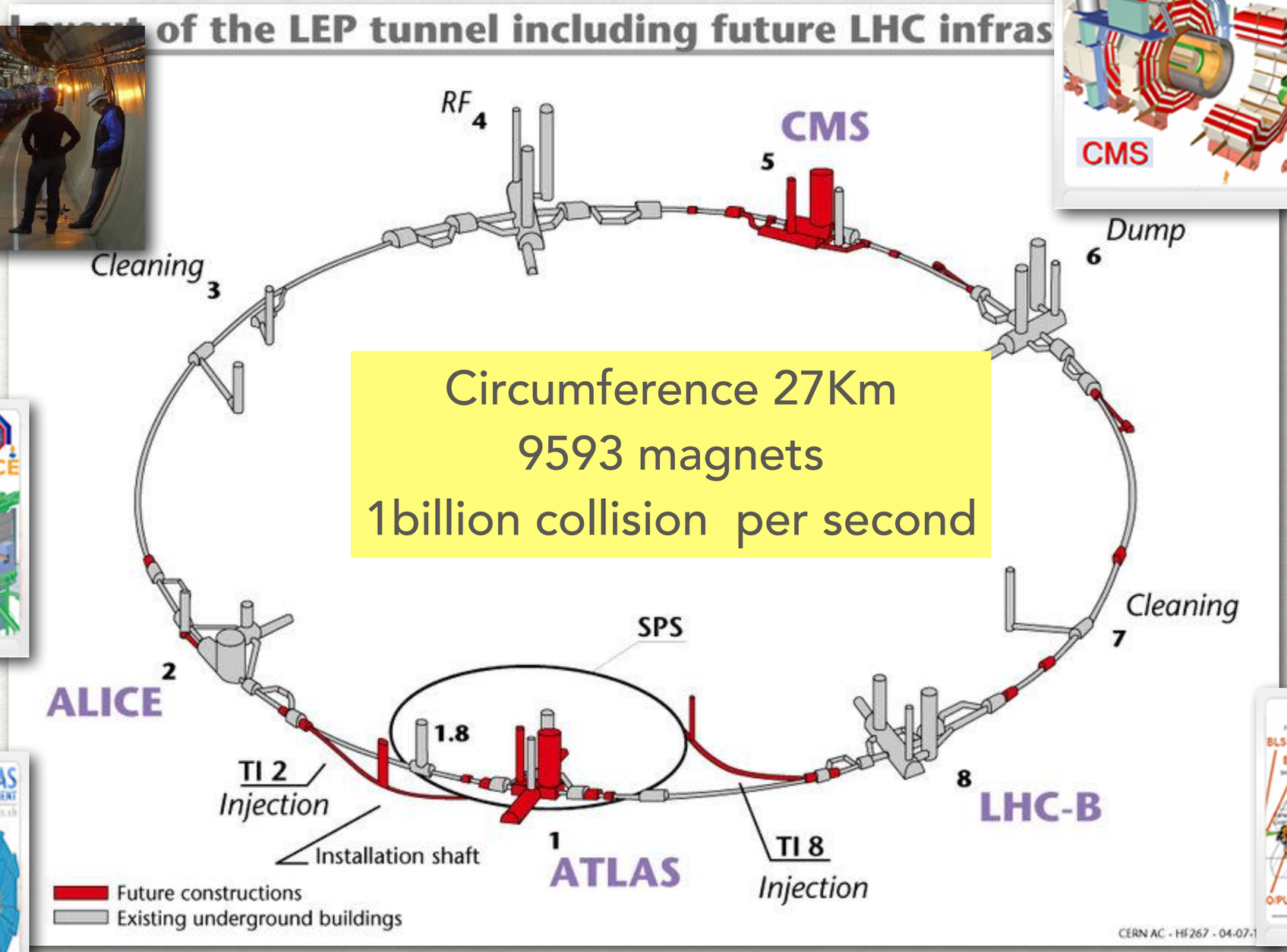


HIGGS BOSON AND OUR UNIVERSE

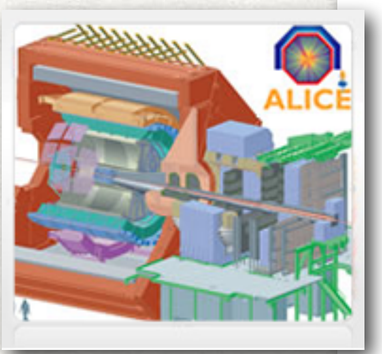
- Our vacuum is not stable.
- This is problem! quantum mechanism says that the vacuum tunnel to the true vacuum sometime.
- We do not know how we settle in the current vacuum
  - It is worth to study Higgs boson in detail!

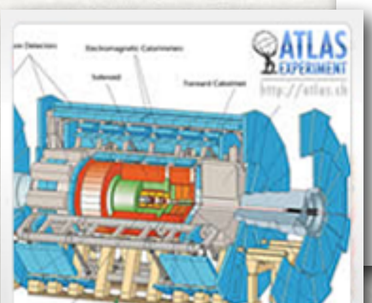


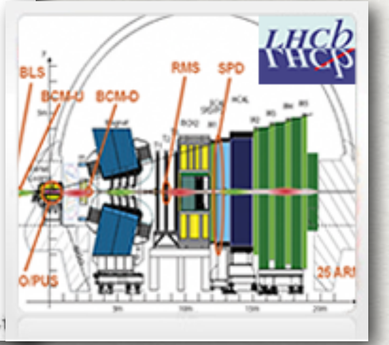




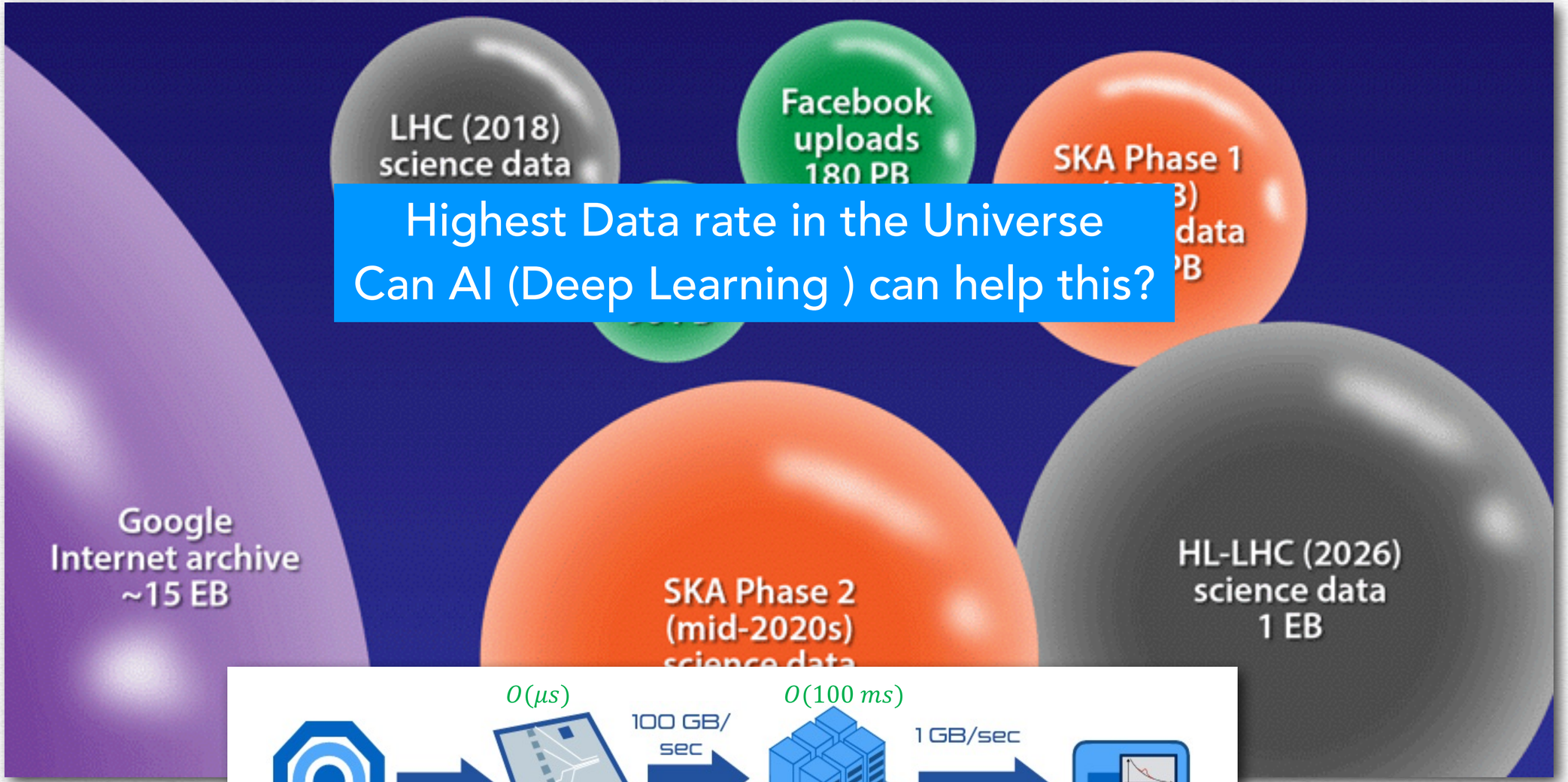
Circumference 27Km
9593 magnets
1billion collision per second



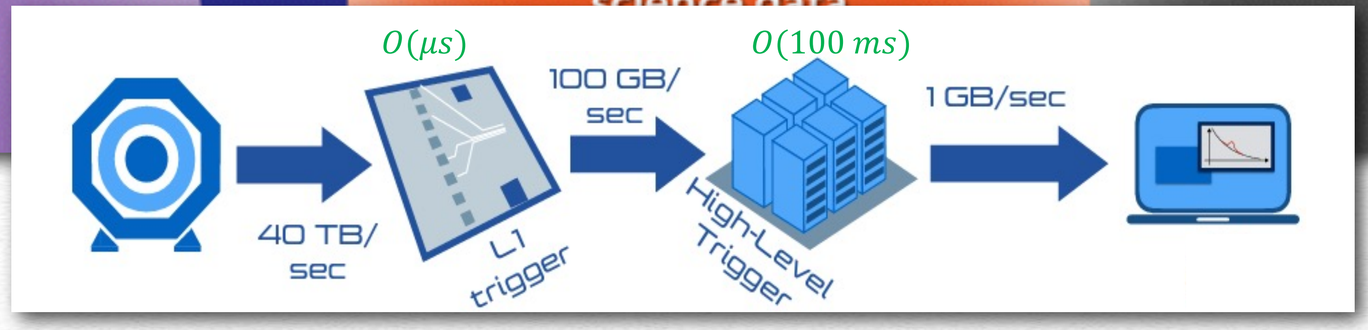




HL-LHC Challenge



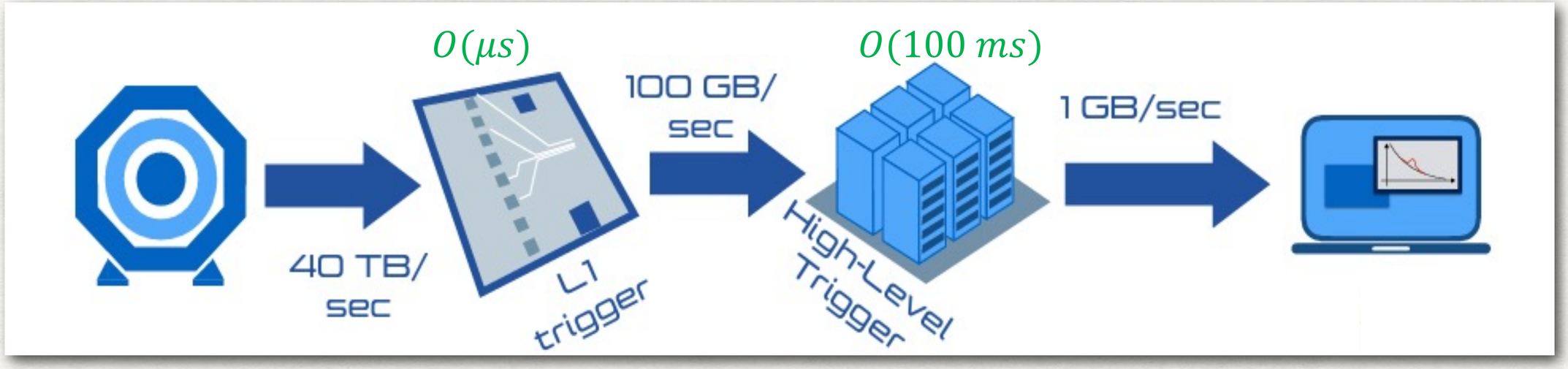
Highest Data rate in the Universe
Can AI (Deep Learning) can help this?



How ML can uncover "new physics" at LHC

- new particle production and its decay to known particles(such as top quark , Higgs boson) or unexpected distribution (deviation from SM)
  - New gauge bosons, supersymmetric particles
  - high energy top quark and Higgs boson can be signal
- Identify unexpected events (anomaly search)
  - Select "strange events" among huge event samples
  - dark matter, hidden sector particles (dark photons)

HOW ML HELP EXPERIMENT ITSELF



- High rate + many data (factor 100 increase)
  - We have to process event 10 times faster (reconstruct , recognize important events , record it)
  - Whole system should not exceed computing budget, and <u>naively it will exceed</u> (because the roadmap takes into account improvements by "R&D")

"ML at HEP workshop" in Japan

23-24 February 2023
KEK (High Energy Accelerator Research Organization)
Asia/Tokyo timezone

Beyond AI
連携事業

量子場計測システム国際拠点
International Center for Quantum-field Measurement Systems for Studies of the Universe and Particles
WPI research center at KEK

MLPhys Foundation of "Machine Learning Physics"
Grant-in-Aid for Transformative Research Areas (A)

- Overview
- Timetable
- Registration
- Participant List
- Poster session
- Venue and access to KEK
- Covid-19 measures at KEK
- Accomodations
- Information about entries to Japan
- Restaurants

We are delighted to announce that the "ML
from February 23rd to 24th, 2023. After a s
2022, this is our first time to organize such
leading researchers on Machine-Learning t
together researchers in high energy physics

The scope of the workshop is to overview t
and beyond, and to provide lively discussions about the directions of ML & AI technical developments in the field.

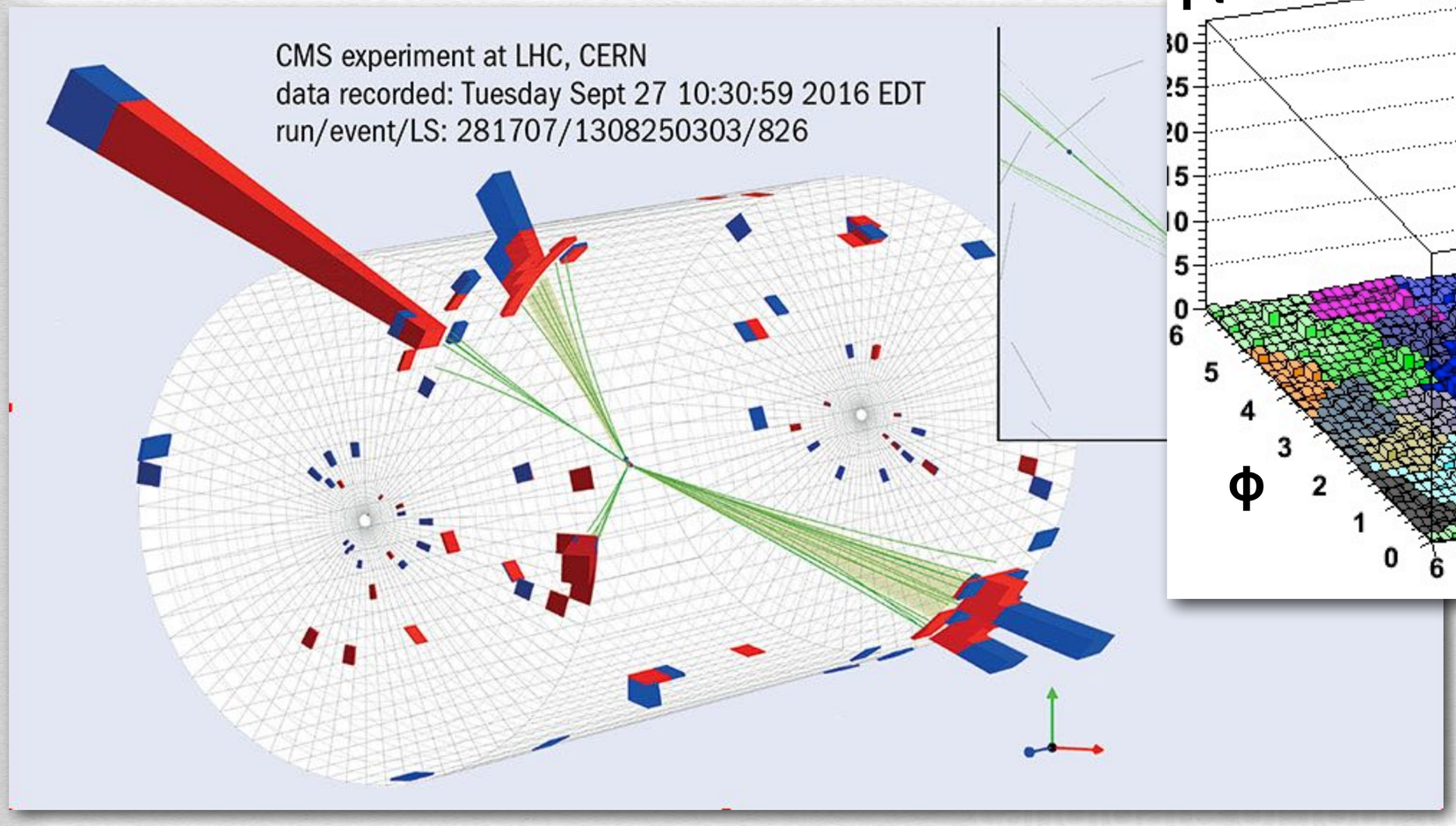
The topics of the workshop program include:

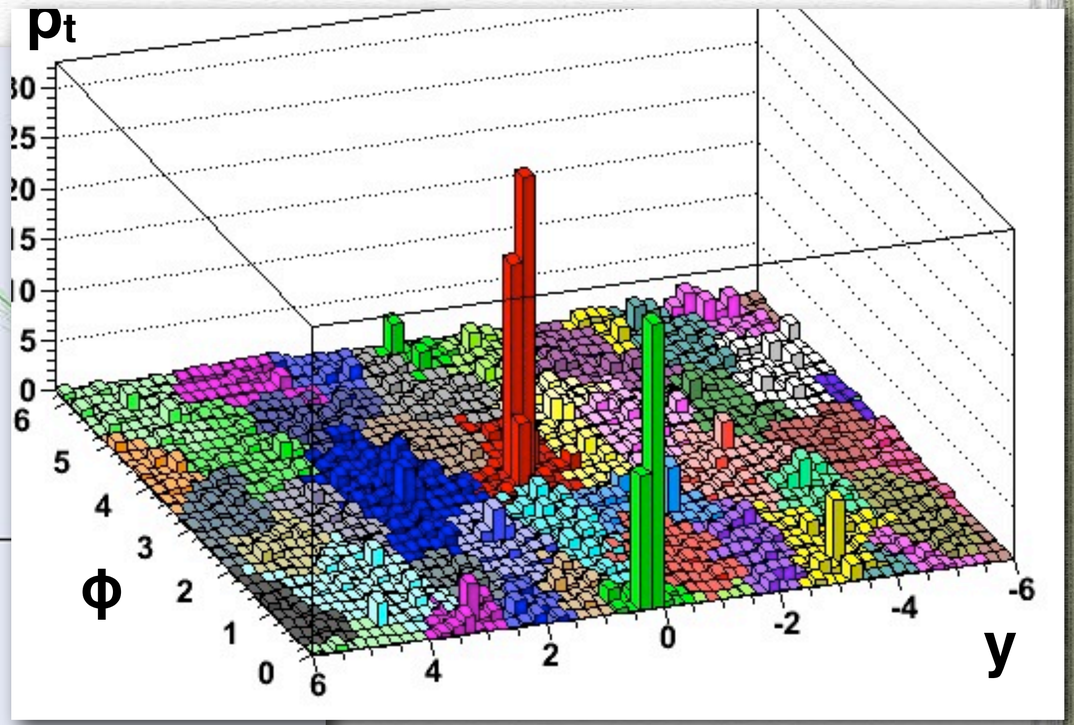
- Advanced ML techniques useful for HEP,
- Advanced architectures for ML,
- ML applications in HEP and beyond,
- Advanced applications at accelerators and systems, and
- Future directions in ML and AI x HEP and beyond.

The workshop format is hybrid (in-person at KEK and online via zoom). Presentations are given by

Most of the subject of my talks
can be found here
https://kds.kek.jp/event/44830/

How machine Leaning help Analysis

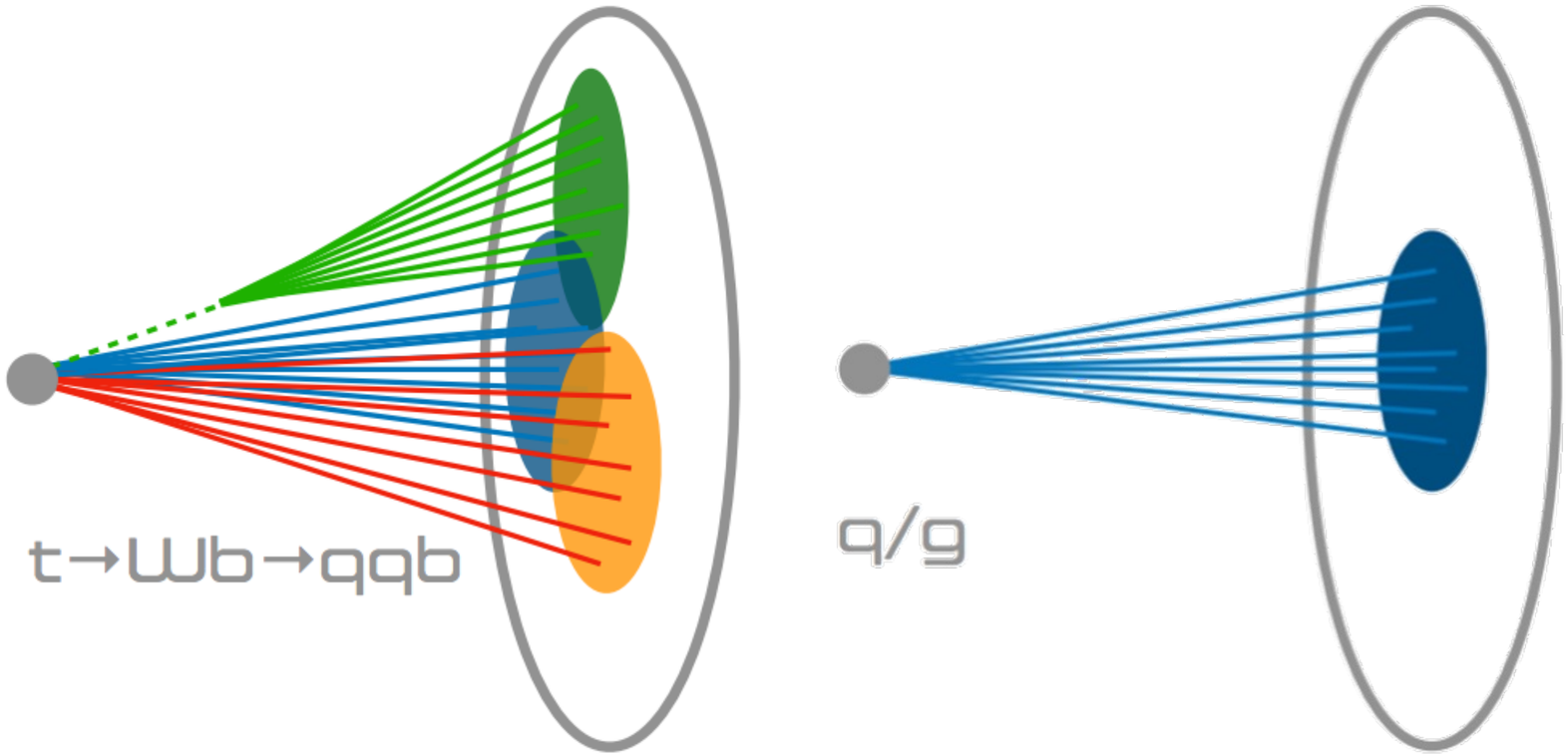




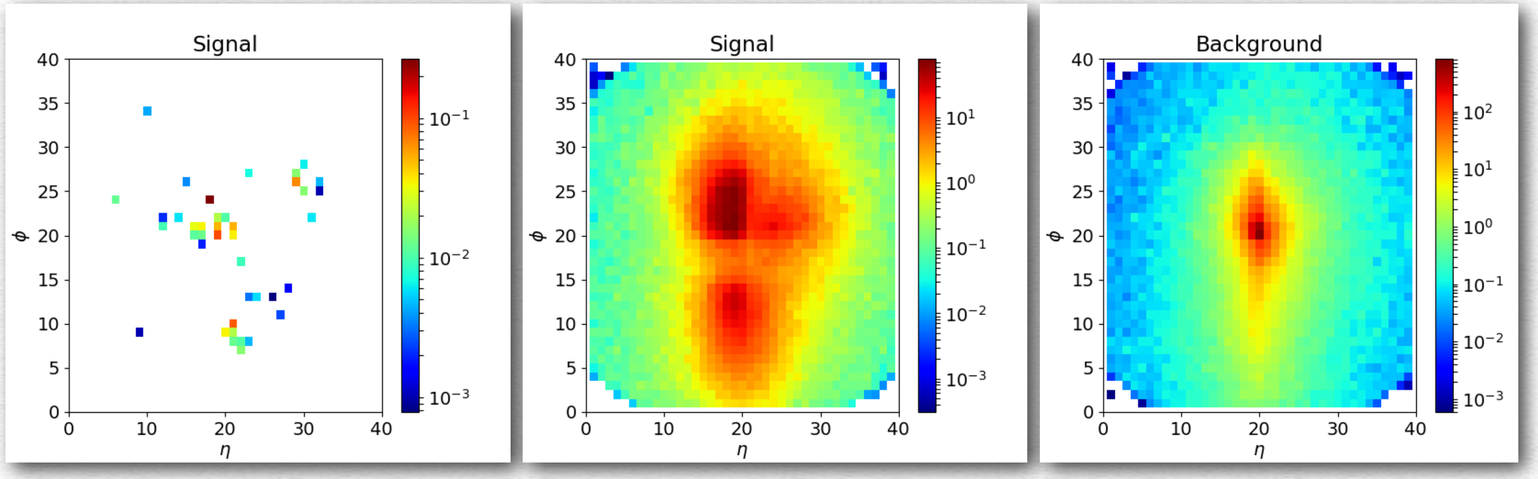
Jet clustering

Example boosted top identification

Task: Top-quark jets vs single quarks/gluons

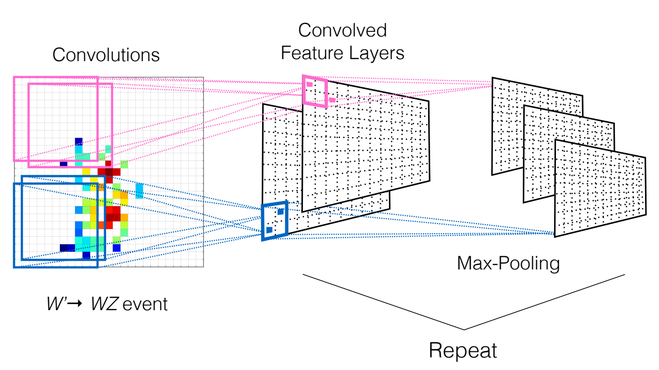


New Physics search with ML



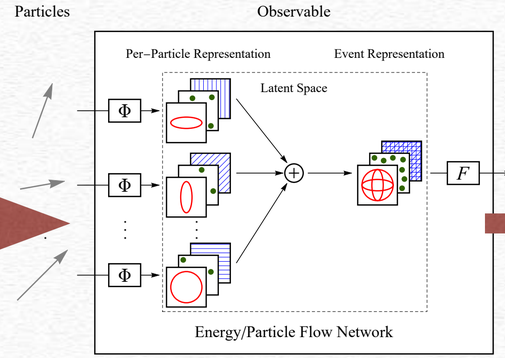
high energy top quark
is similar to light quark and gluons
but there is some difference

Jet as Image



CNN(2014)

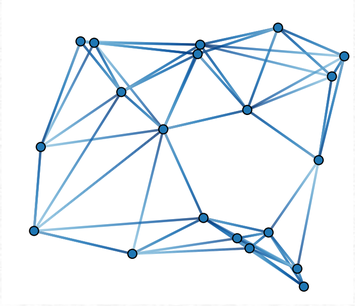
as sets



permutation invariance
(stable)

Easy to implement
QCD "values"

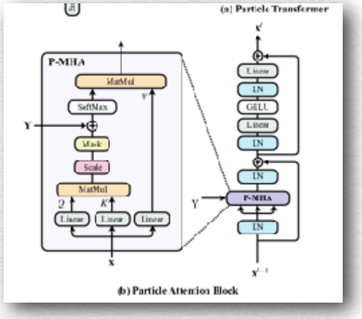
as graphs



sparse data
1902.08570
Particle Net

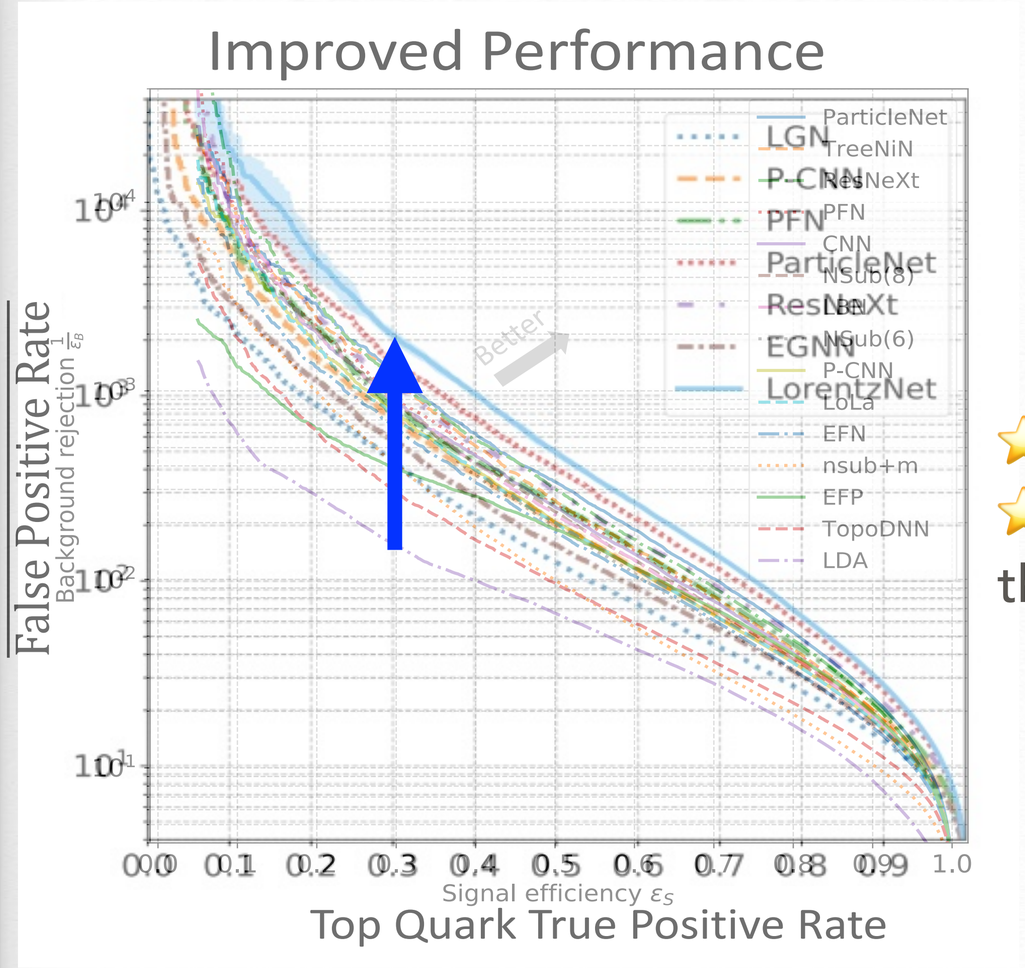
Lorentz net
(Graph respect special relativity)

transformer



building key
and query
2202.03772

2018 to 2021



2201.08187

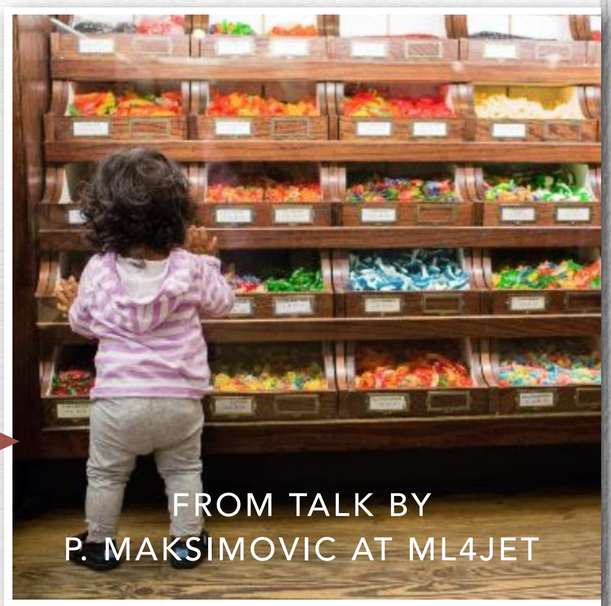
More than factor 10 reduction of background
→ current cross section limit improve by factor 10

Two questions

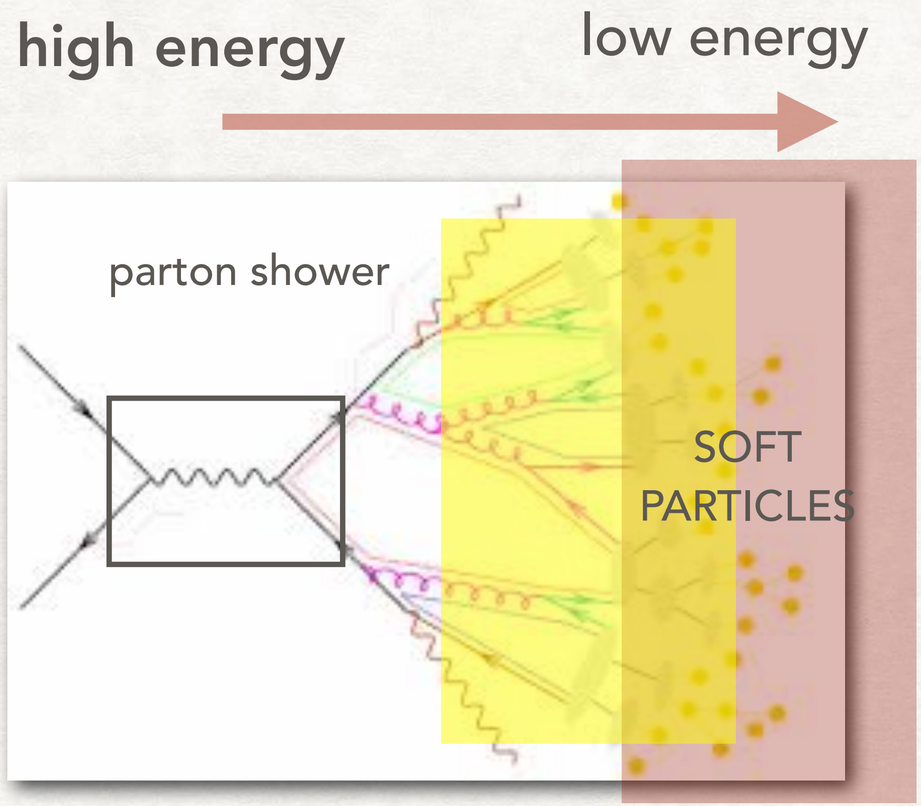
⭐if it improve drastically, "why theory cannot explain it?

⭐Theorists start to get complaint from experimentalists that performance isn't that impressive.

"an experimentalist wondering why they cannot touch the top of the shelf "



MACHINE DOESN'T CARE PHYSICS VALUES



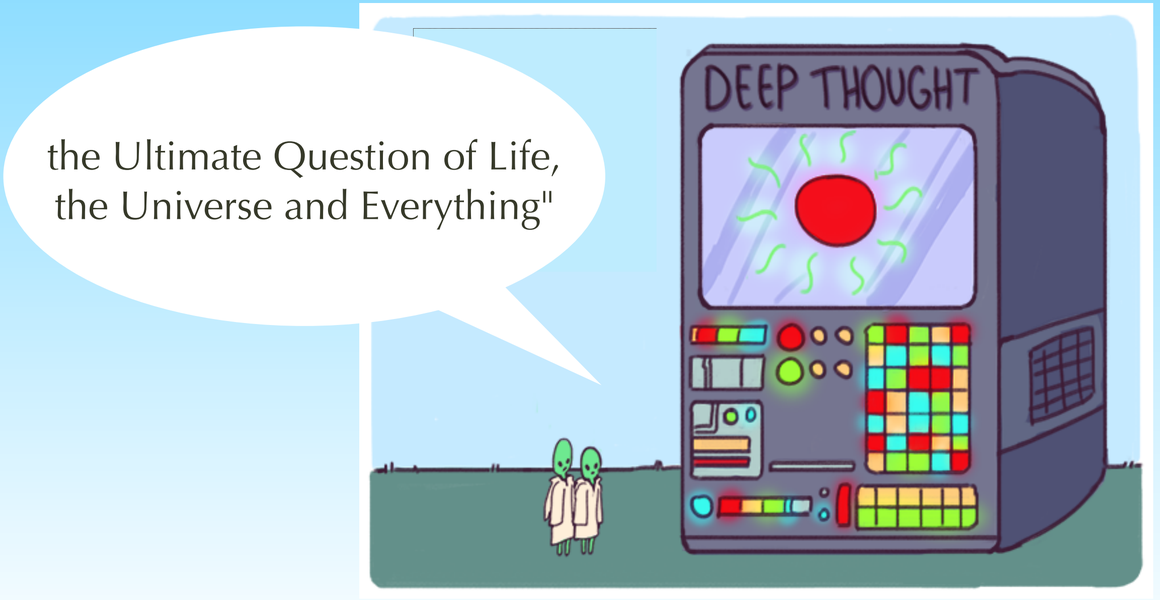
- IRC (infrared collinear) safe quantities : theoretically predictable
- **IRC "sensitive" quantity: (soft physics)** (ex low energy particles, number of particles) Difficult to predict theoretically because of the divergence of the theory. We need modelings but they often do not agree with data.

MESSAGE

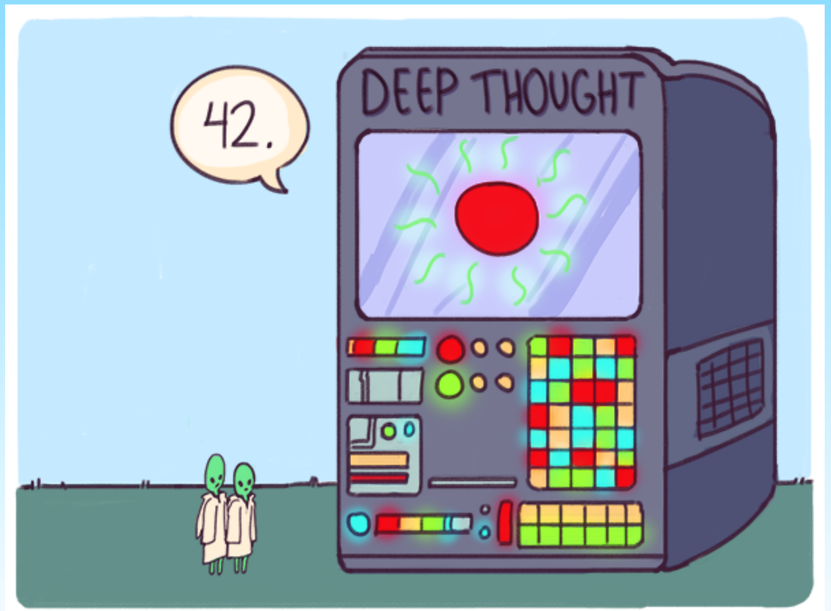
Modern Deep learning models take into account ALL correlation.
HEP analysis need simulation and we need updated our simulation to be consistent at all energy.

- **Soft phyiscs might be useful because low energy particle distribution has information of branching history strong interaction. just we need careful calibration**

About asking "something" to computer



750M years



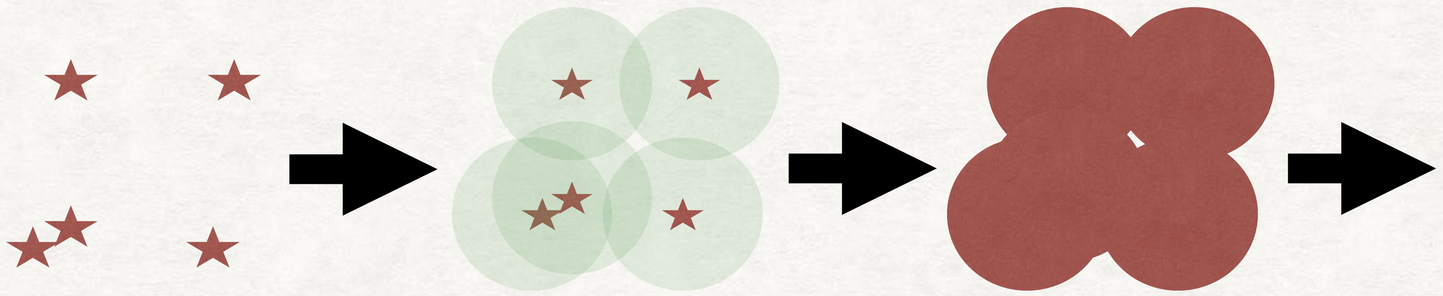
*" I think the problem, to be quite honest with you, is that **you've never actually know what the question is**"*
*"So **once you do know what the question actually is, you will know what the answer means**"*

from "The Hitchhiker's Guide to the Galaxy"
by Douglas Adams

Physics ≠Getting good results by using ML and satisfy

SOFT PARTICLE GEOMETRY

Lim, Nojiri *Phys.Rev.D* 105 (2022) 1, 014004
Furuichi, Lim, Nojiri (in preparation)



jet image

circle (radias R) at every hit

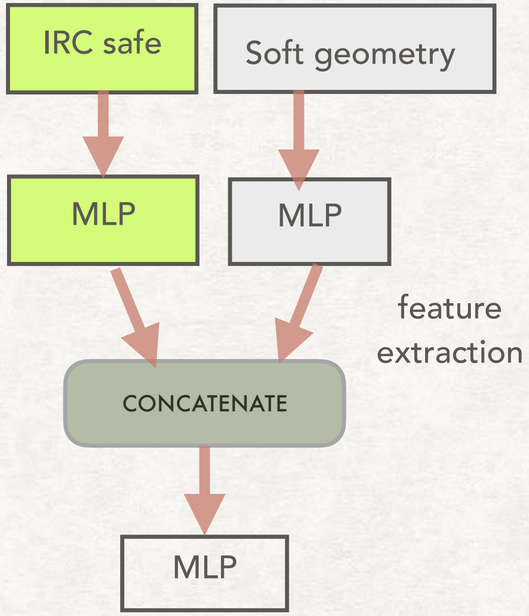
resulting image

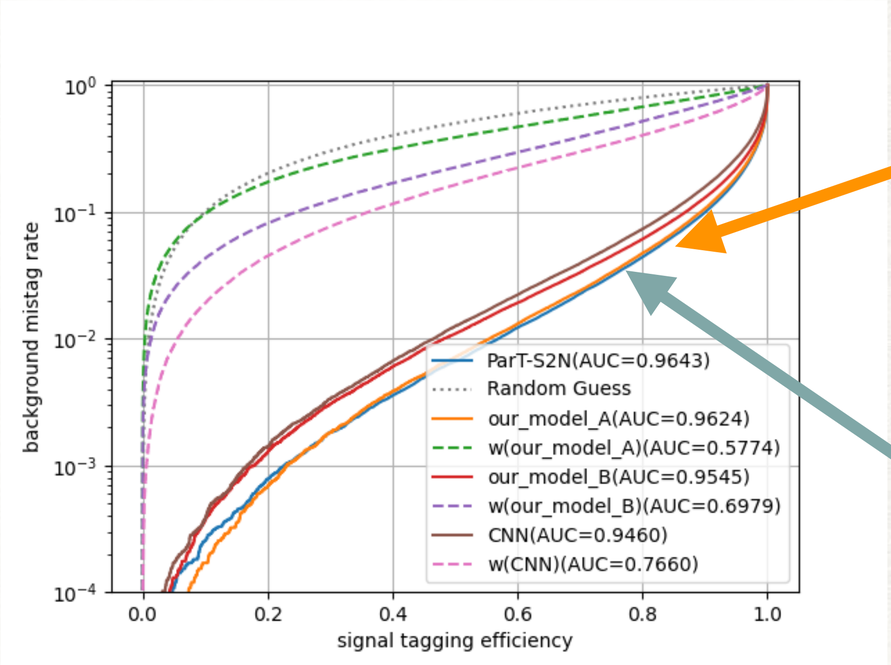
S(R) R=R1, R2, R3,,,

Area (A),boundary length(L)
Euler character (E) of S (R)

1. compress spase data into real numbers
2. resilient to fluctuation of the inputs
3. permutation invariant

soft geometry model





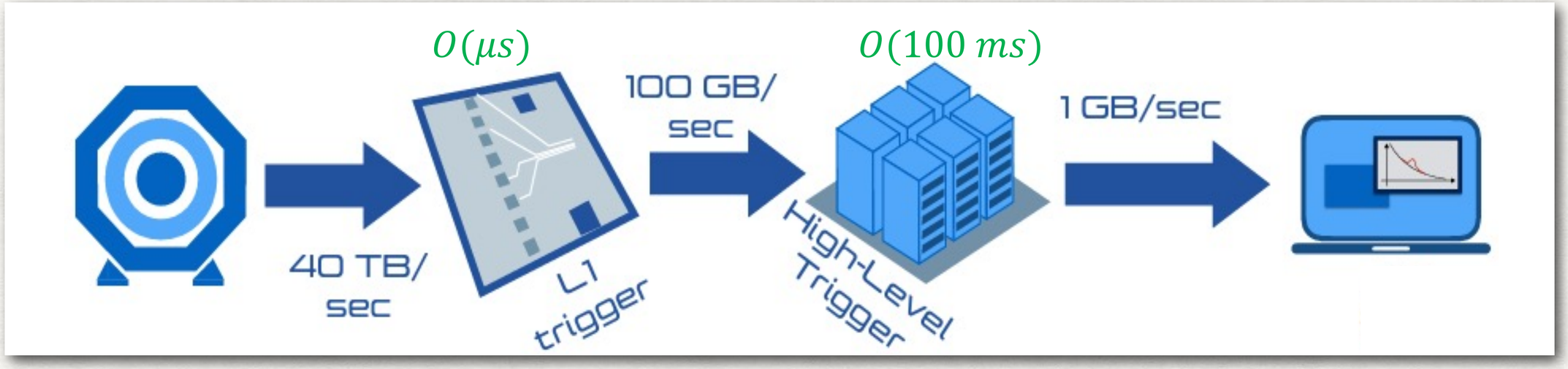
soft geometry model

inputs has clear interpretation on distance scale
→ easy to infer the reasoning

Maybe we can create better simulation by DL

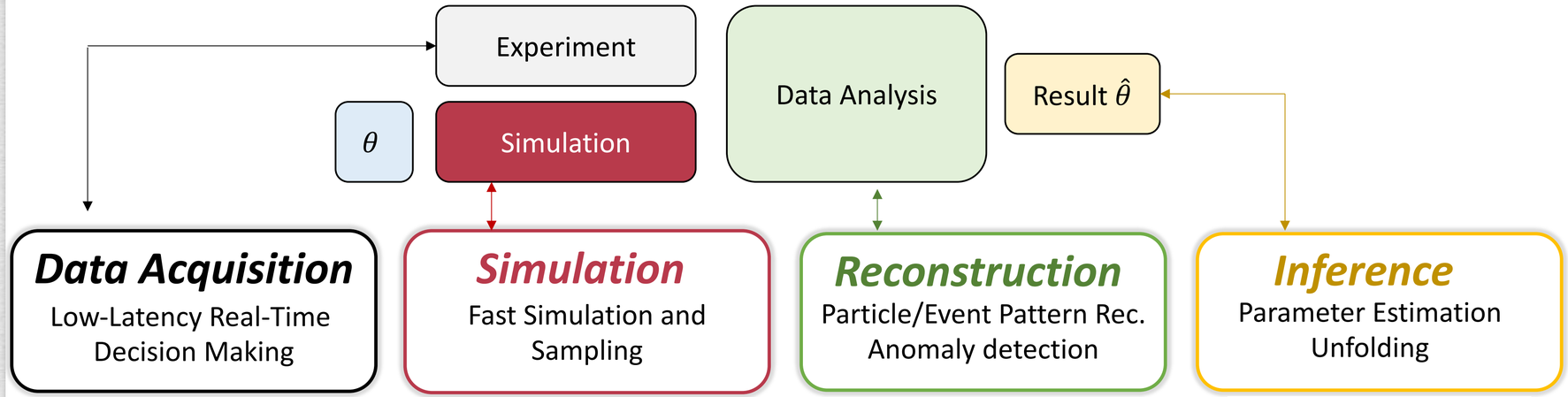
Particle Transformer
(current world best set up)

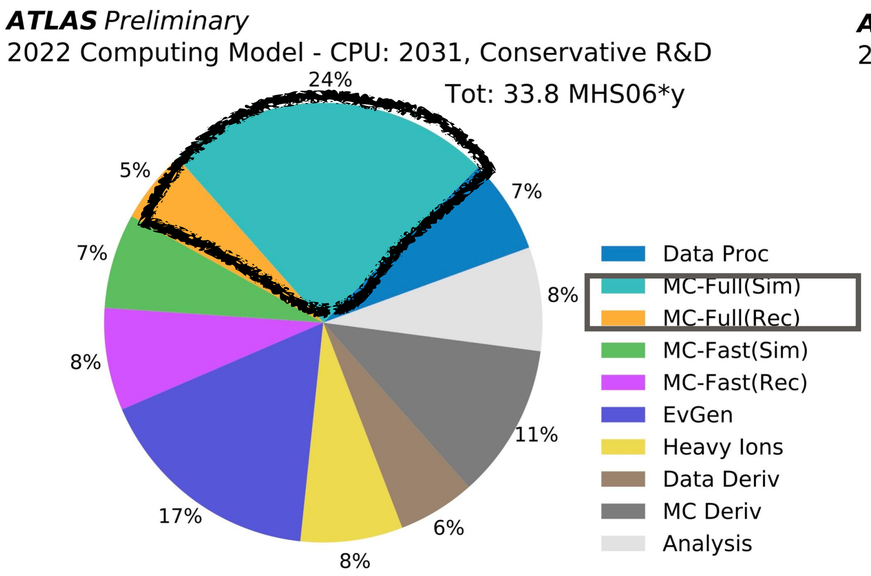
HOW ML HELP EXPERIMENT ITSELF

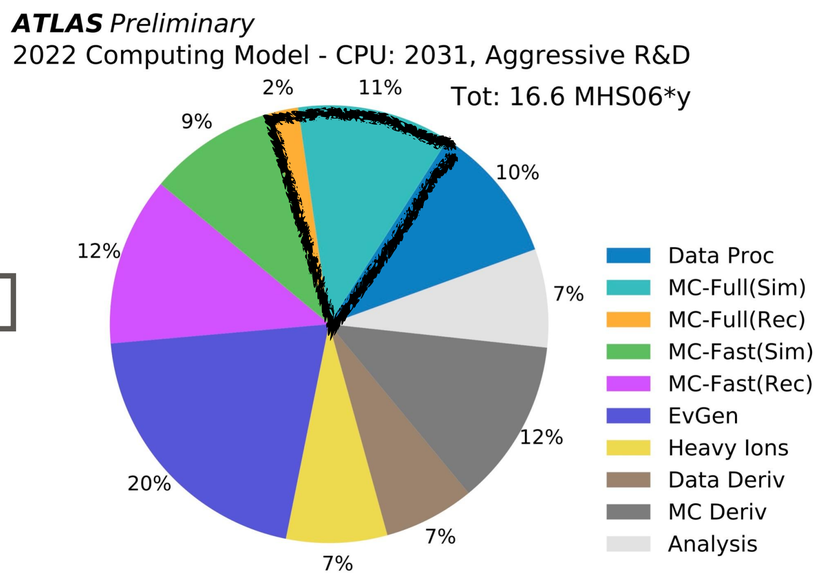


- High rate + many data (factor 100 increase)
  - Experiment should process events 10 times faster (reconstruct , recognize important events , record it)
  - **Whole analysis chain should not exceed computing budget,** and naively it exceeds (because the roadmap takes into account improvement by "R&D")

Machine Learning Across Data Analysis







Agressive R& D assume improvement by ML

Kagan "ML at HEP"

event generation by ML

Calorimater

generating one event takes 3min

Bang!

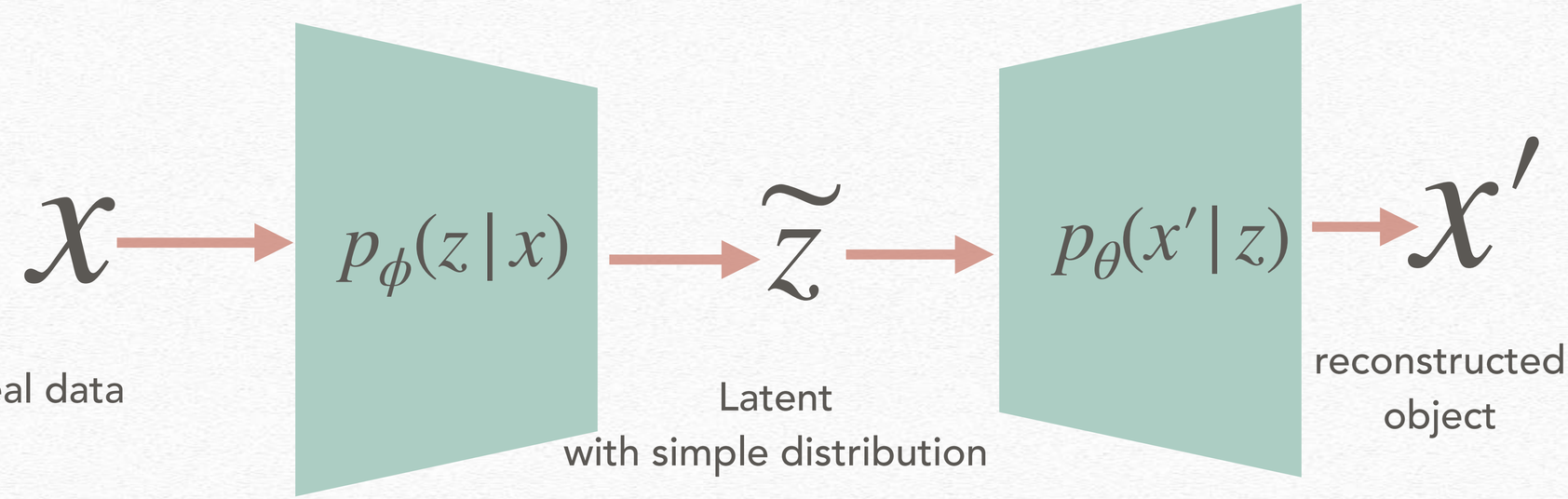
input particle {x}

ML model

E-Deposit to
each component

P({E} | {x})

Variational AutoEncoder(VAE): bottle
neck structure x→ z → x'



real data

Latent
with simple distribution

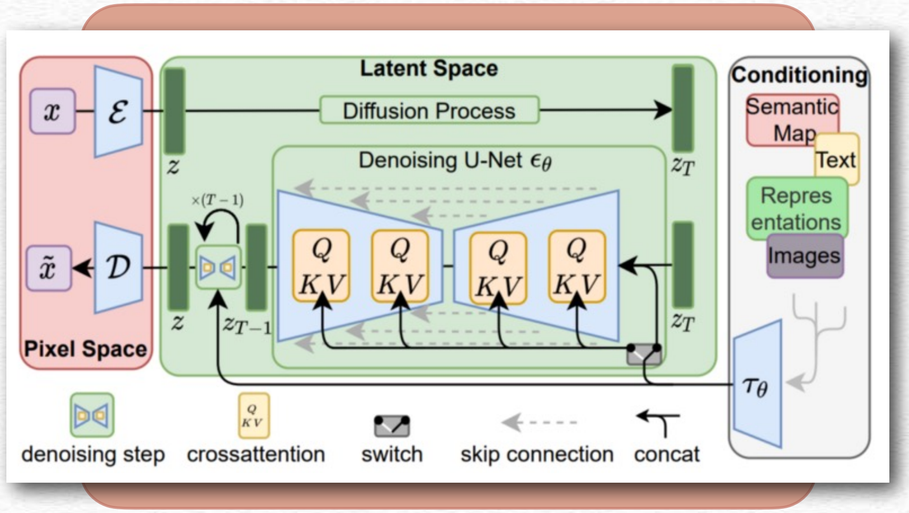
reconstructed
object

event generation

Calorimater

Bang!

input particle {x}



E-Deposit to each component

P({E} | {x})

ML toolbox

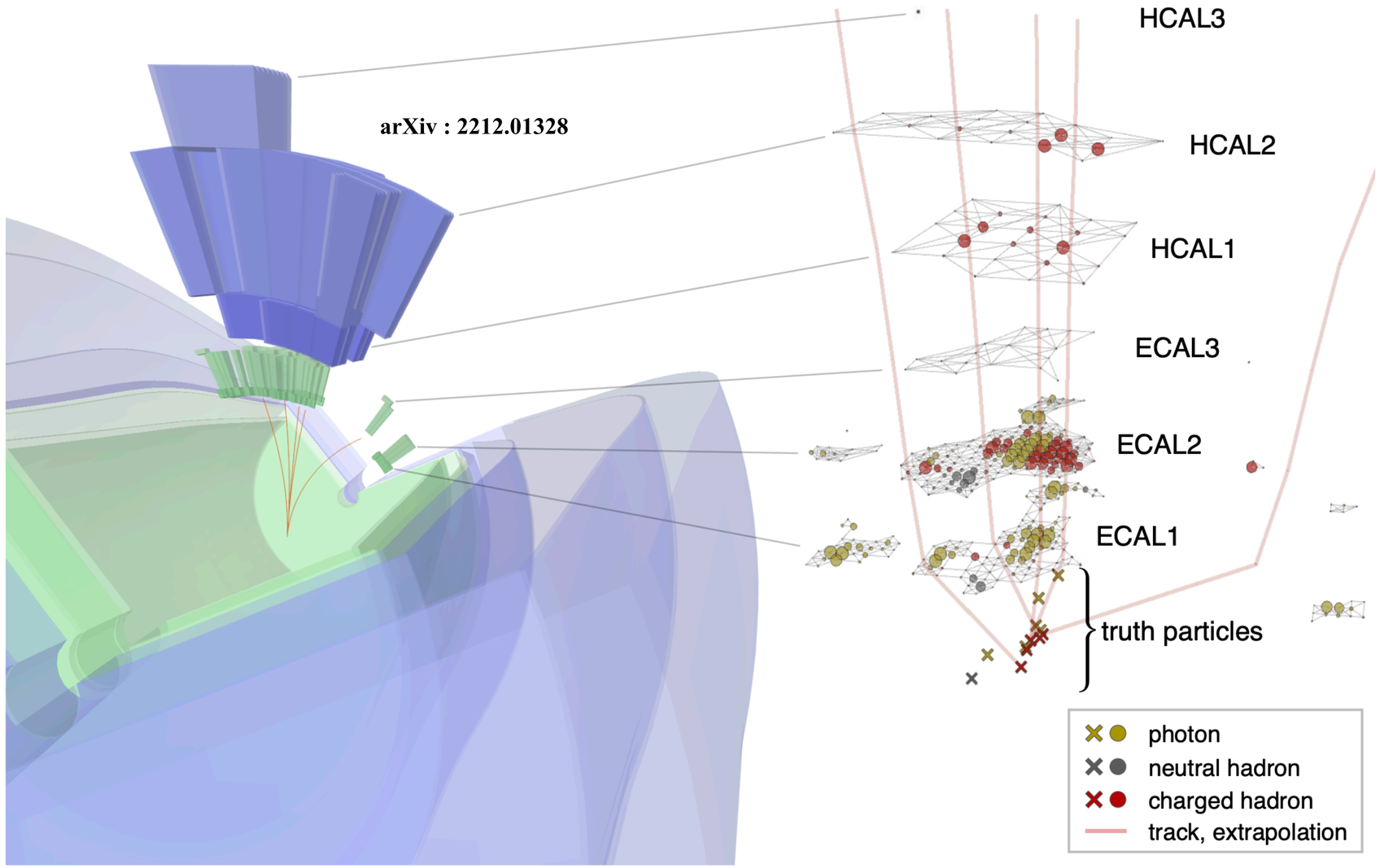
GAN:
competition between generator and discriminator

Normalizing Flows:
map simple distribution to nontrivial function

Diffusion : Add noise step by step

Once we complete this map, we have **fast and more precise model** for everybody (even theorists) . DL model is differentiable so that you can do parameter search using ML

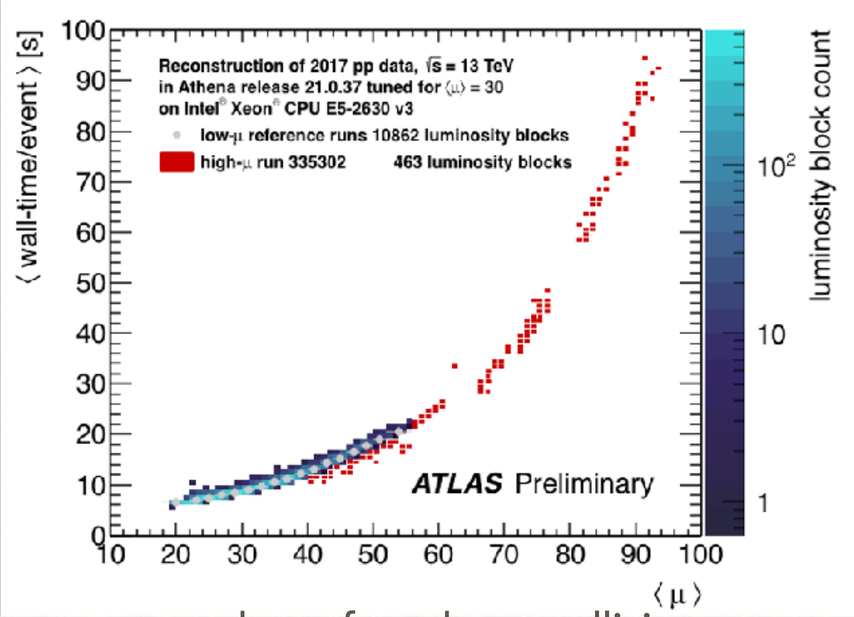
The PF reconstruction algorithm



PF lepton, hadron, photon = F_{PF} (track hits + calo cells)

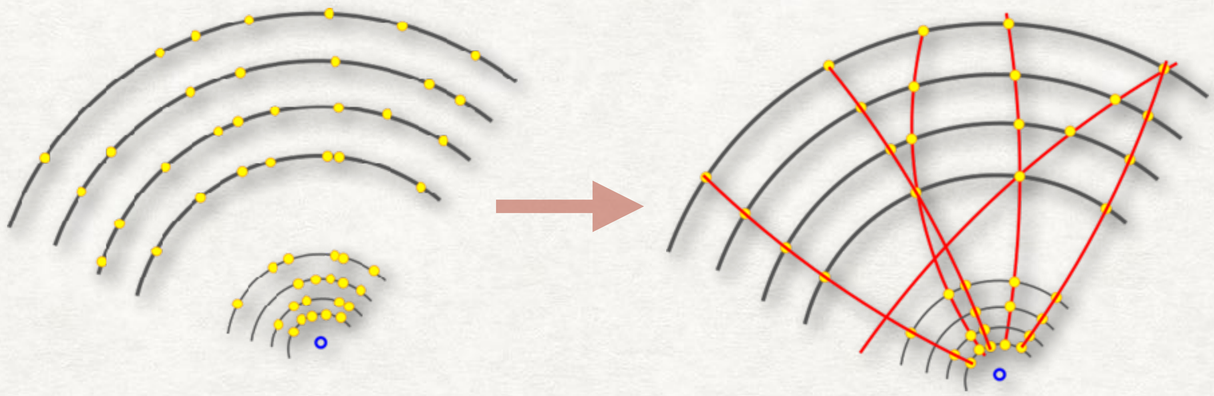
BREAKING THE WALL

events in pipeline →**select interesting one**→save (physical constraints)

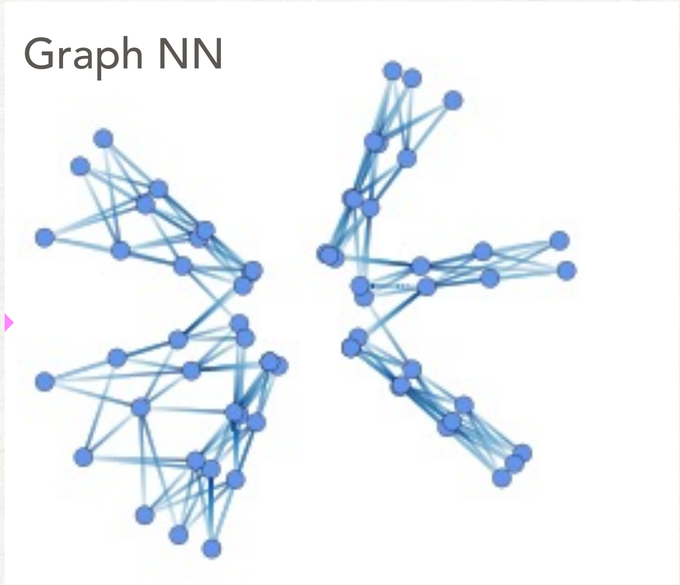


number of tracks at collision

tracking



Graph NN



EXA.TrkX (Eur. Phys J. C(2021) 81:87

detemine near by node → GNN to handle edge classification

training time ~36 hour(by A100)、

prediction 2.2 s/event (GPU), 220 s/event (CPU)

ML on hardware

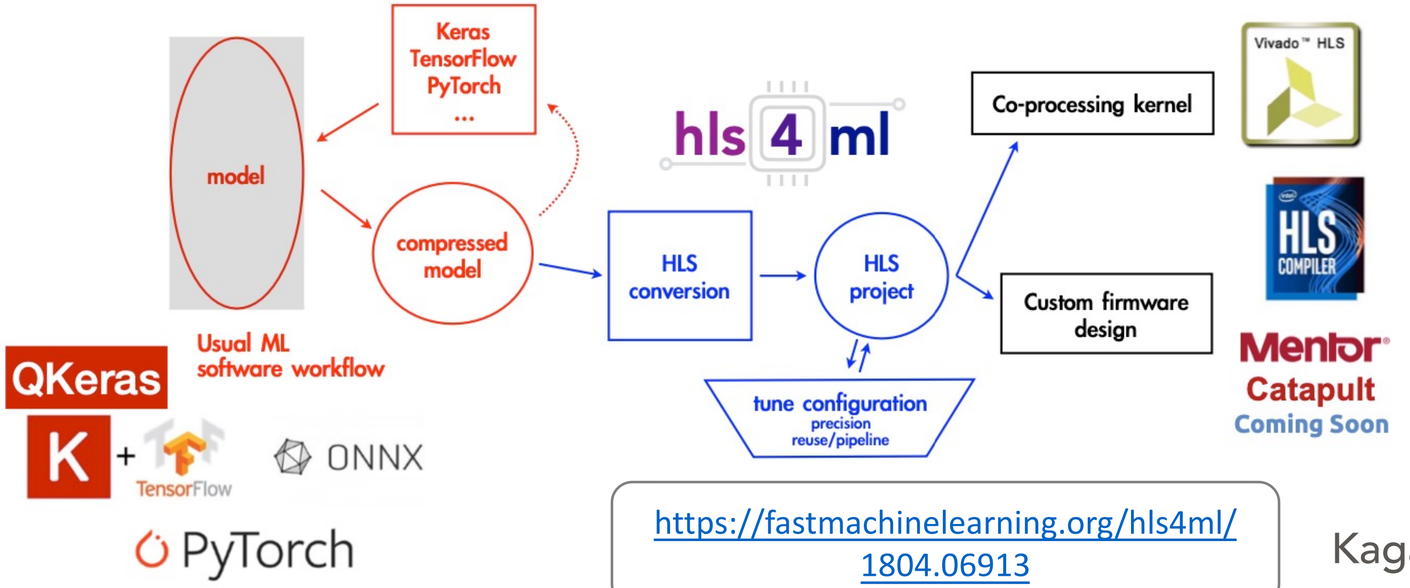
Real-Time Low Latency ML

Goal: Move improvements from ML for reconstruction to on

FPGA= Field Programmable Gate Array
HLS =its compiler

hls4ml:

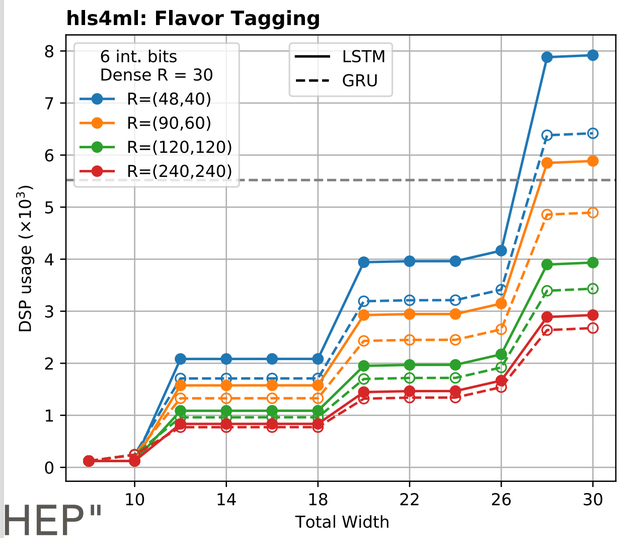
- HEP Community developed tool do deploy NN as electronic circuit on hardware
- FPGA-as-accelerator library → Whole model on chip
- Model Compression essential



https://fastmachinelearning.org/hls4ml/
1804.06913

Kagan "ML at HEP"

Example: Resource Usage in Jet Classification with RNNs



Khoda, ..., **MK**, et. al, 2207.00559

summary

- HEP people are working to merge DL to most precise, large data science under extreme environment.
- The outcome will certainly shared to outside community, so stay tuned.