Advancing Particle Physics With Deep Learning

LPC Topic of the Week
23 Feb, 2021

Jean-Roch Vlimant (California Institute of Technology)
jvlimant@caltech.edu @vlimant

Outline

I. Overview of Machine Learning
II. The case for Deep Learning in HEP
III. Deep Learning Applications in HEP

An Introduction to Machine Learning

a bird's view ...

A Definition

"Giving computers the ability to learn without explicitly programming them" A. Samuel (1959).

Is fitting a straight line machine learning ?
Models that have enough capacity to define its own internal representation of the data to accomplish a task : **learning from data.**

In practice : a statistical method that can extract information from the data, not obviously apparent to an observer.

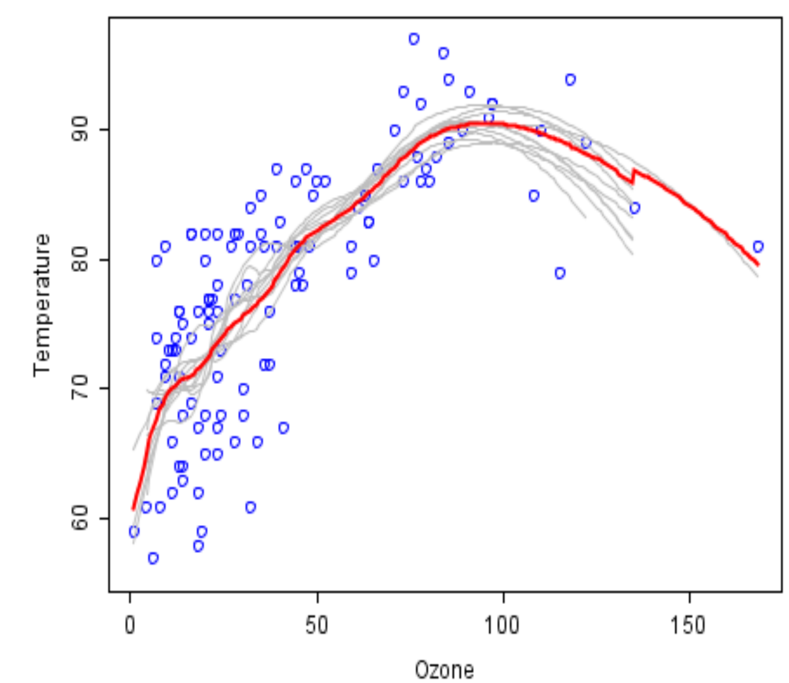
→ Most approach will involve a **mathematical model** and a cost/reward function that needs to be **optimized.**

→ The more **domain knowledge** is incorporated, the better.

Supervised Learning

- Given a dataset of samples, a subset of features is qualified as **target**, and the rest as **input**
- Find a **mapping from input to target**
- The mapping should **generalize to any extension** of the given dataset, provided it is generated from the same mechanism

$$dataset \equiv \{(x_i, y_i)\}_i$$
$$find\ \ function\ \ f\ \ s.t.\ \ f(x_i) = y_i$$



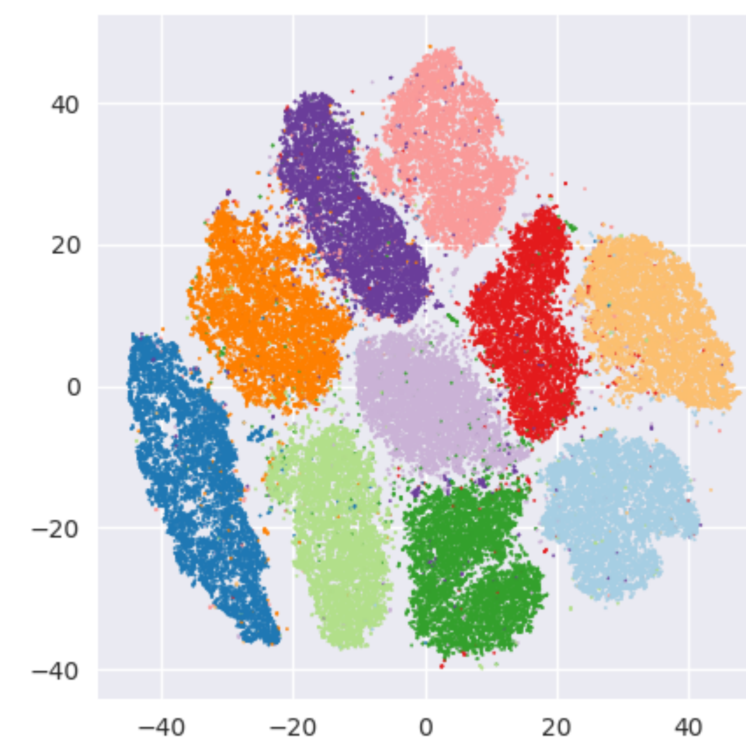
- Finite set of target values :
  → **Classification**
- Target is a continuous variable :
  → **Regression**

Unsupervised Learning

- Given a dataset of samples, but there is no subset of feature that one would like to predict
- Find mapping of the samples to a lower dimension manifold
- The mapping should generalize to any extension of the given dataset, provided it is generated from the same mechanism

$$dataset \equiv \{(x_i)\}_i$$
$$find \;\; f \;\; s.t. \;\; f(x_i) = p_i$$

- Manifold is a finite set
  → **Clusterization**
- Manifold is a lower dimension manifold :
  → **Dimensionality reduction, density estimator**



Reinforcement Learning

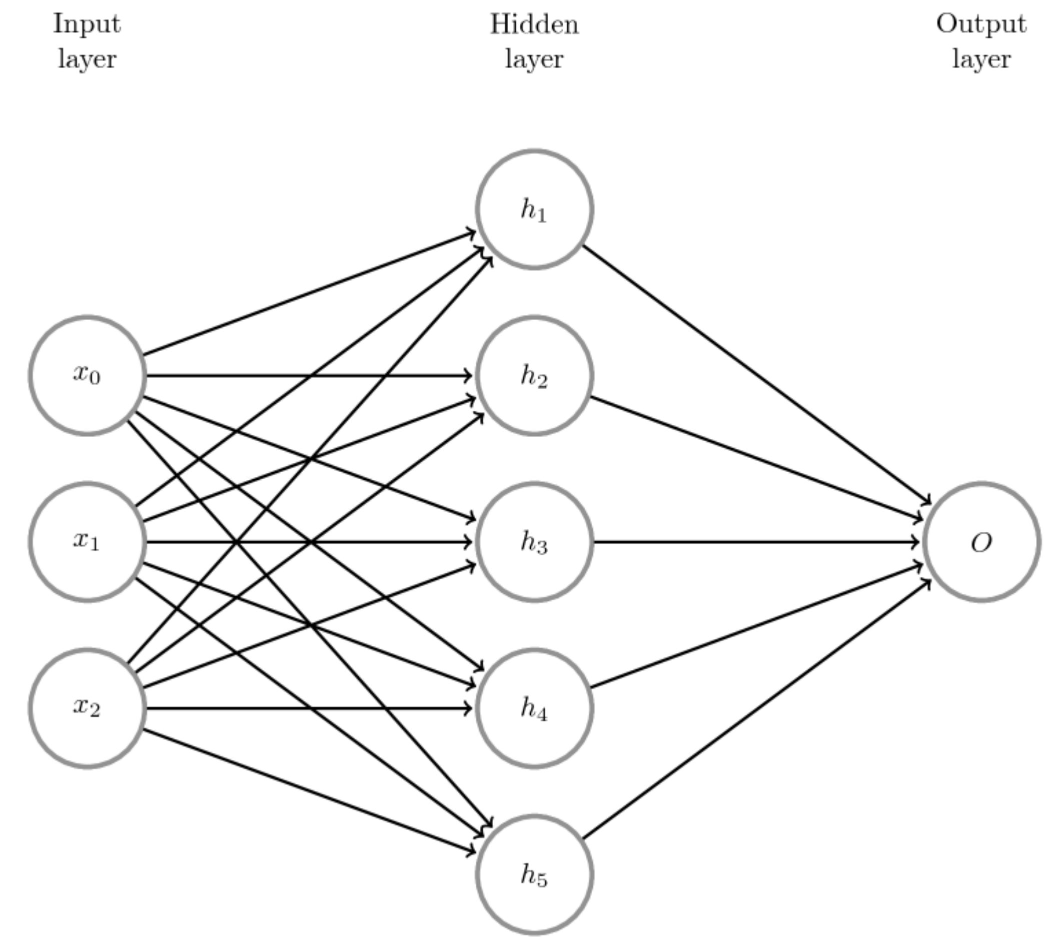
- Given an **environment** with multiple states, given a reward upon action being taken over a state
- Find an **action policy to drive** the environment toward maximum cumulative reward

$$s_{t+1} = Env(s_t, a_t)$$
$$r_t = Rew(s_t, a_t)$$
$$\pi(a|s) = P(A_t = a | S_t = s)$$
$$find\ \pi\ s.t.\ \sum_t r_t\ is\ maximum$$

Artificial Neural Network

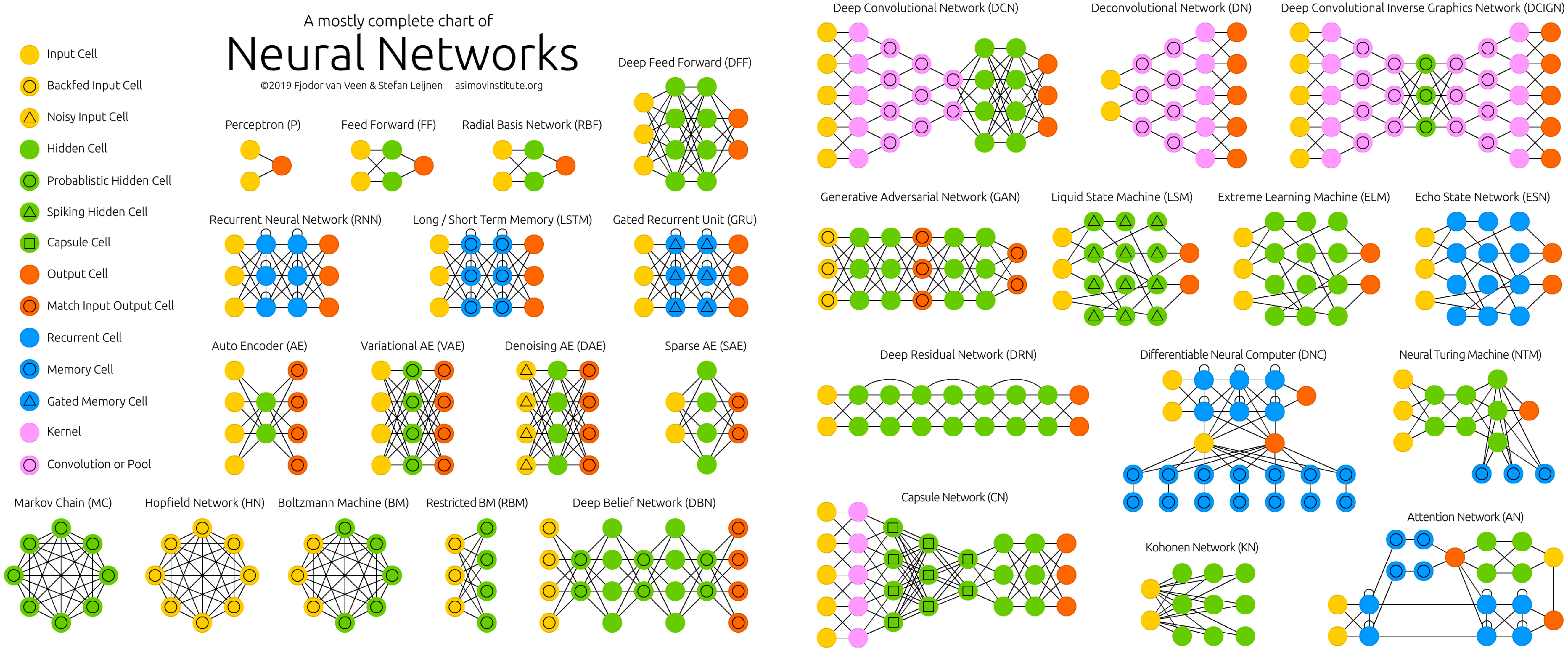
- **Biology inspired** analytical model, but **not bio-mimetic**
- Booming in recent decade thanks to large dataset, increased computational power and theoretical novelties
- Origin tied to logistic regression with change of data representation
- Part of any "deep learning" model nowadays
- Usually large number of parameters trained with stochastic gradient descent

$$h = \phi(Ux + v)$$
$$o(x) = \omega^T h + b$$
$$p_i \equiv p(y = 1 | x) \equiv \sigma(o(x)) = \frac{1}{1 + e^{-o(x)}}$$
$$loss_{XE} = -\sum_i y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)$$

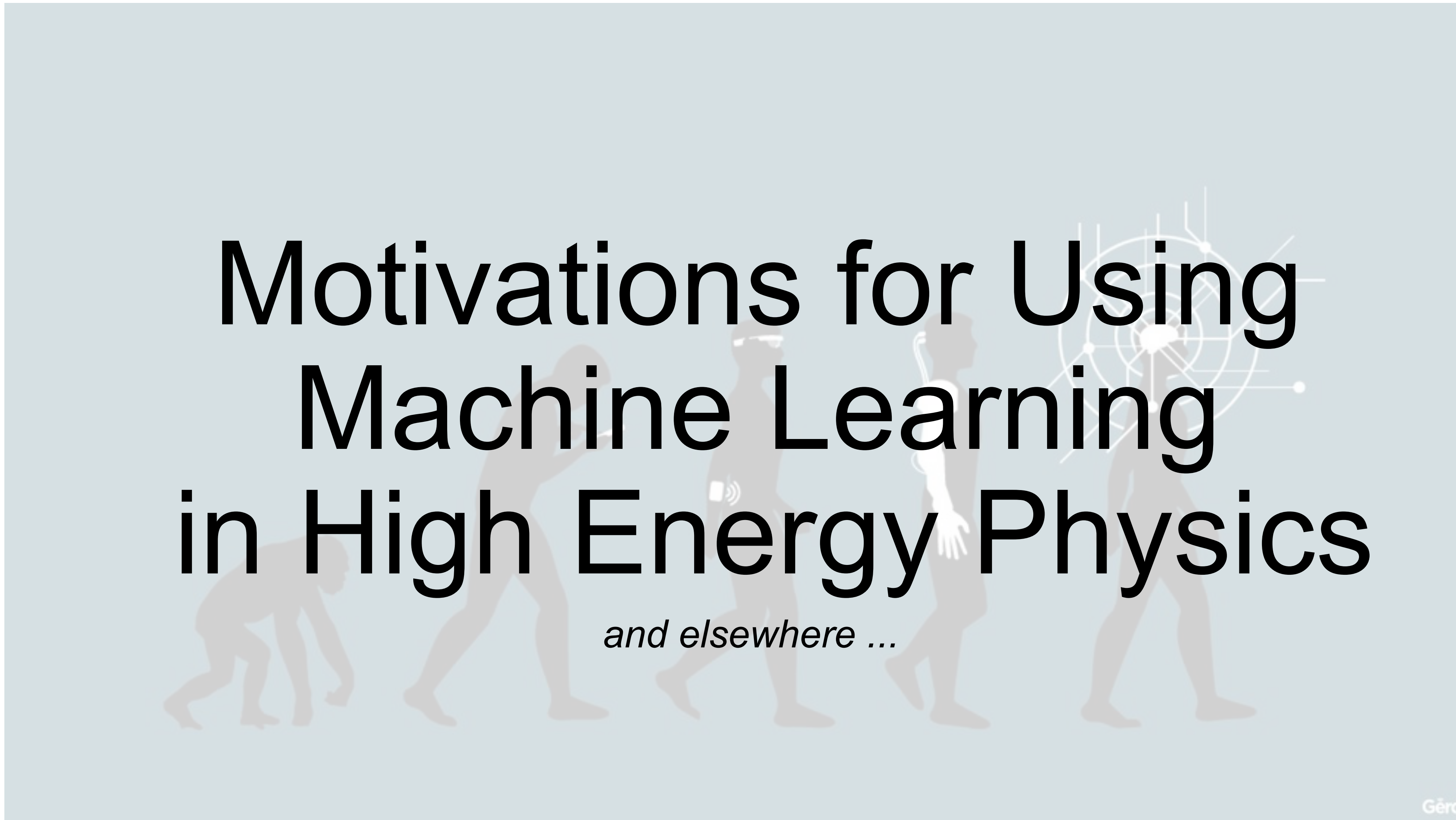


Neural Net Architectures

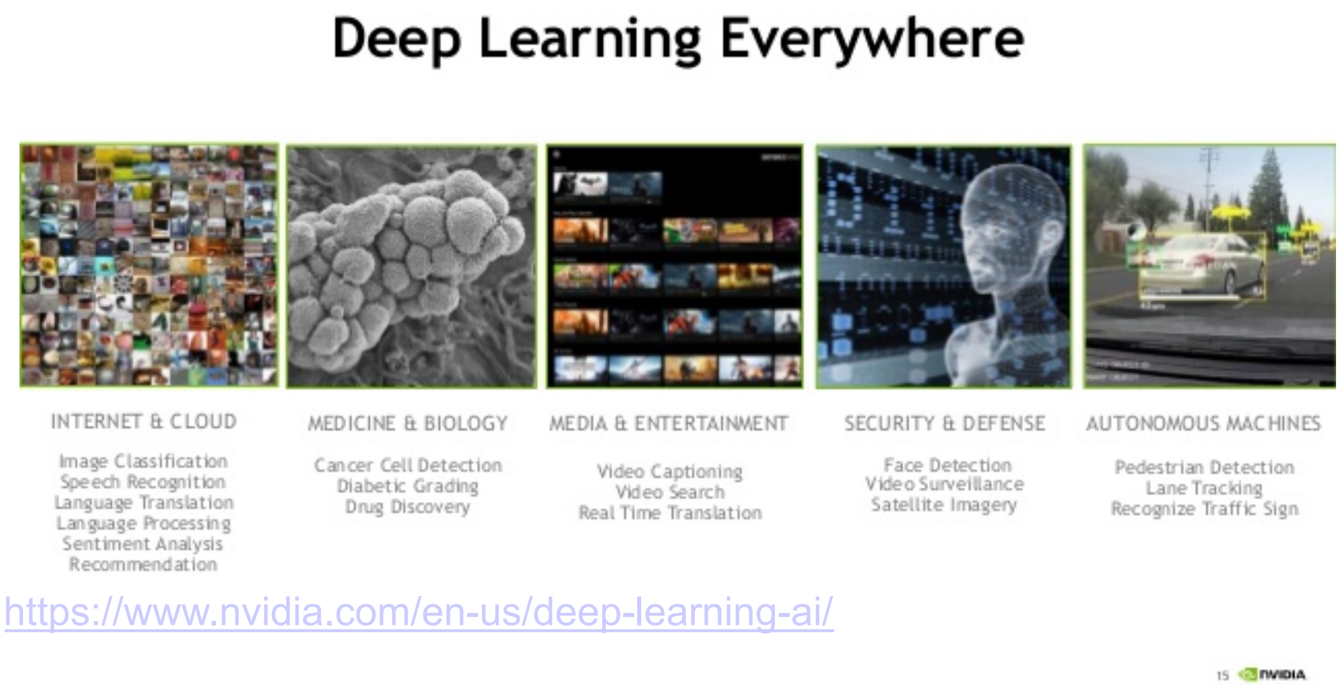
http://www.asimovinstitute.org/neural-network-zoo



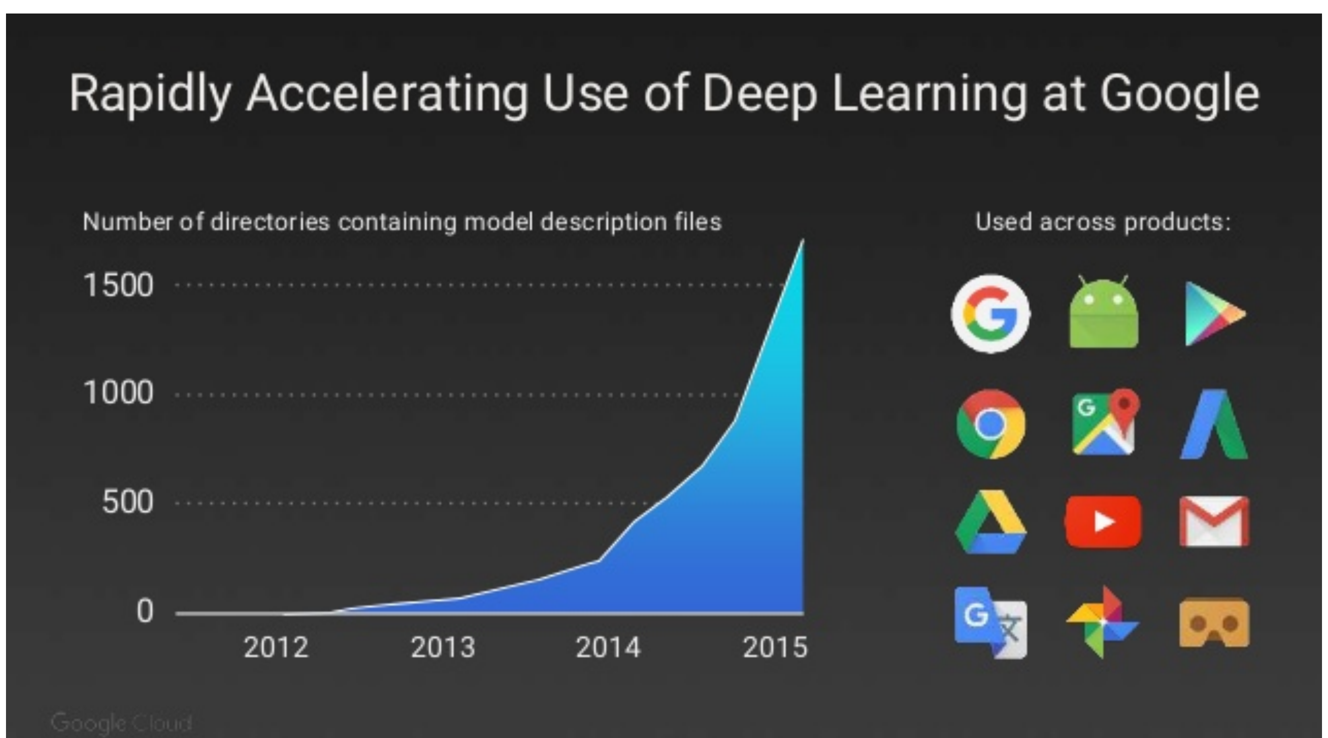
- Does not cover it all : densenet, graph network, ...

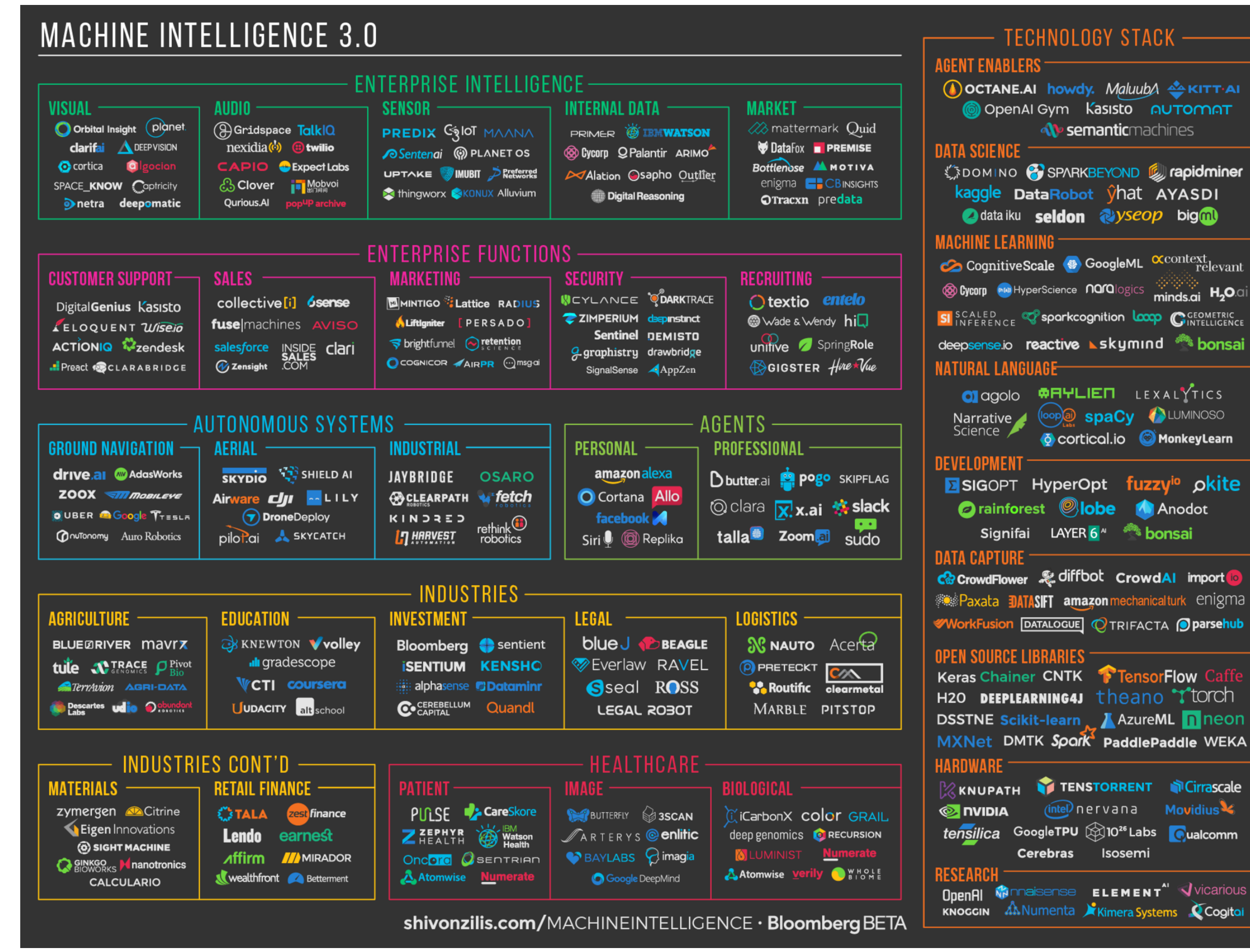

Motivations for Using
Machine Learning
in High Energy Physics
and elsewhere ...
Gërd

Machine Learning in Industry



https://www.nvidia.com/en-us/deep-learning-ai/

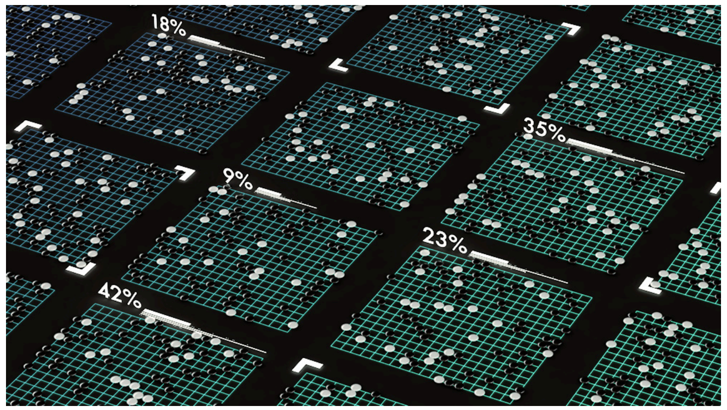


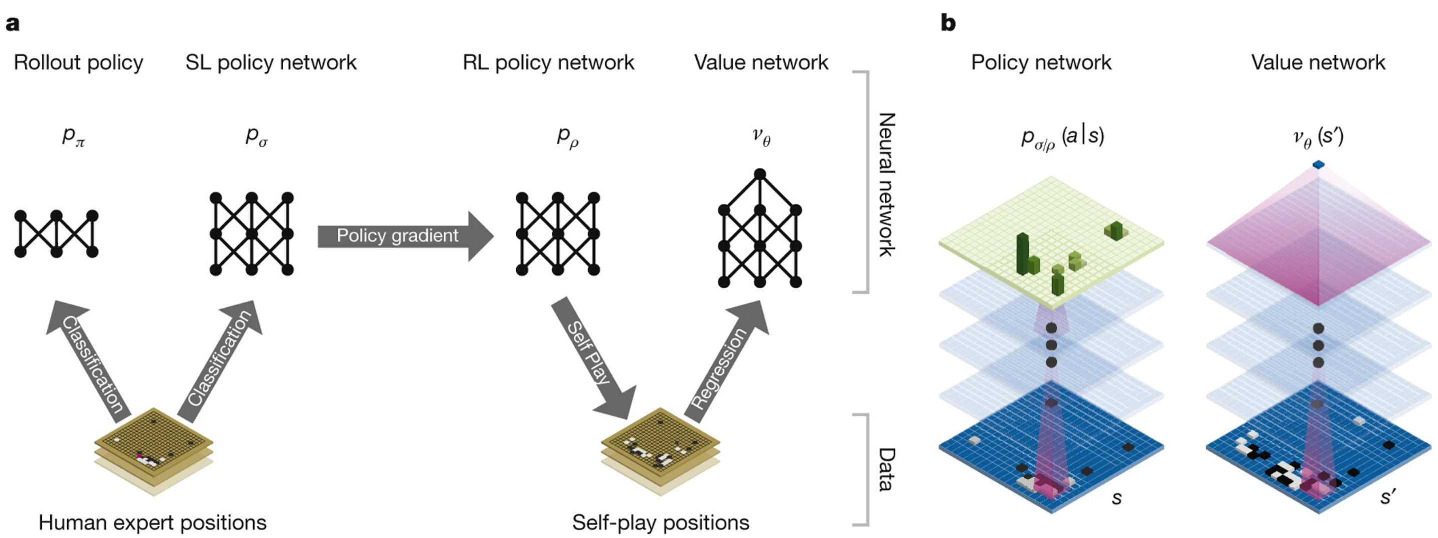


http://www.shivonzilis.com/machineintelligence

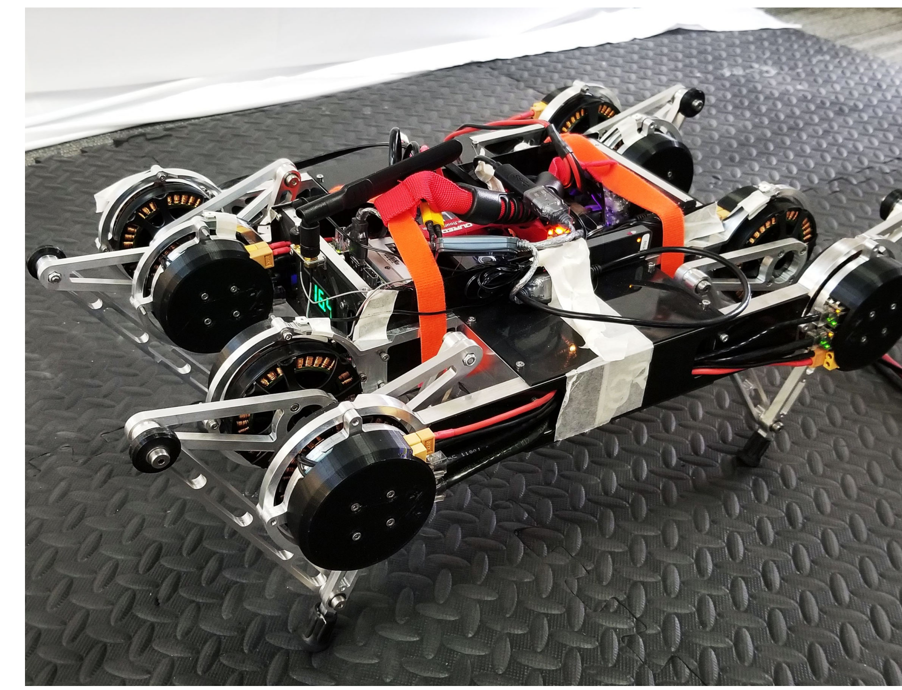
Prominent skill in industry nowadays. Lots of data, lots of applications, lots of potential use cases, lots of money. Knowing machine learning can open significantly **career horizons.**

Learning to Control





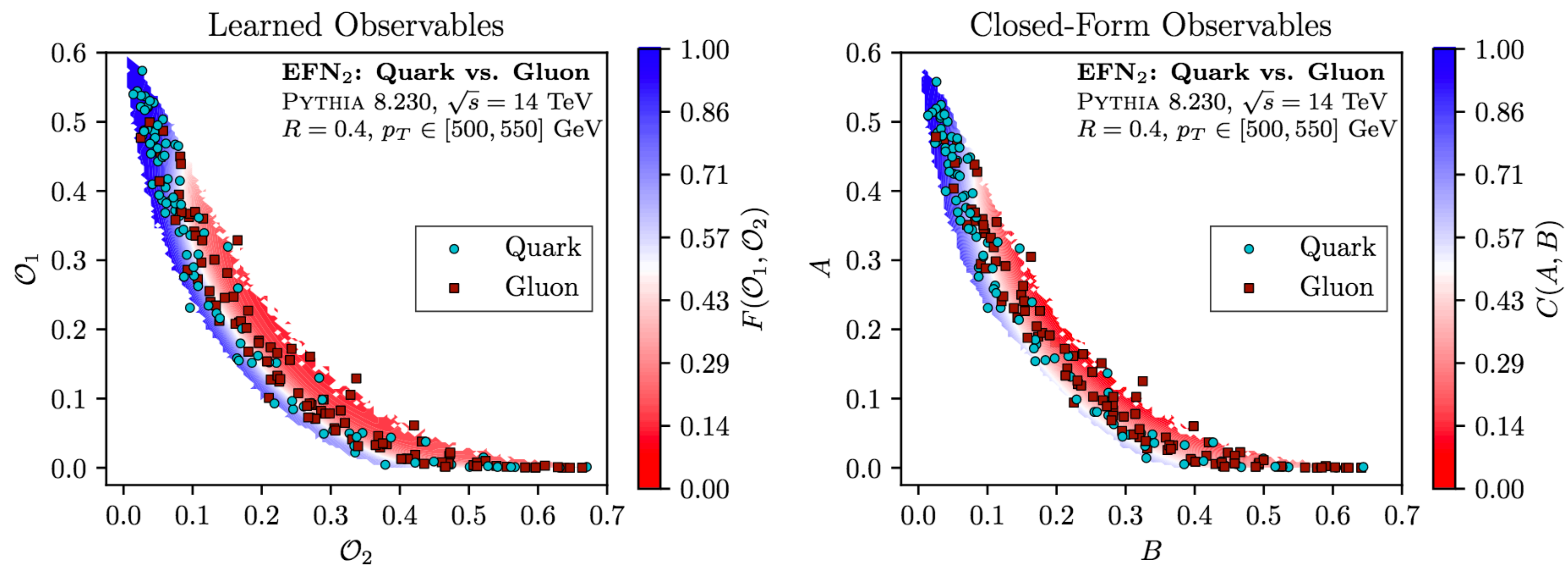
Mastering the game of Go with deep neural networks and tree search,
https://doi.org/10.1038/nature16961



Learning to Walk via Deep Reinforcement Learning
https://arxiv.org/abs/1812.11103

Modern machine learning **boosts control technologies**.
AI, gaming, robotic, self-driving vehicle, etc.

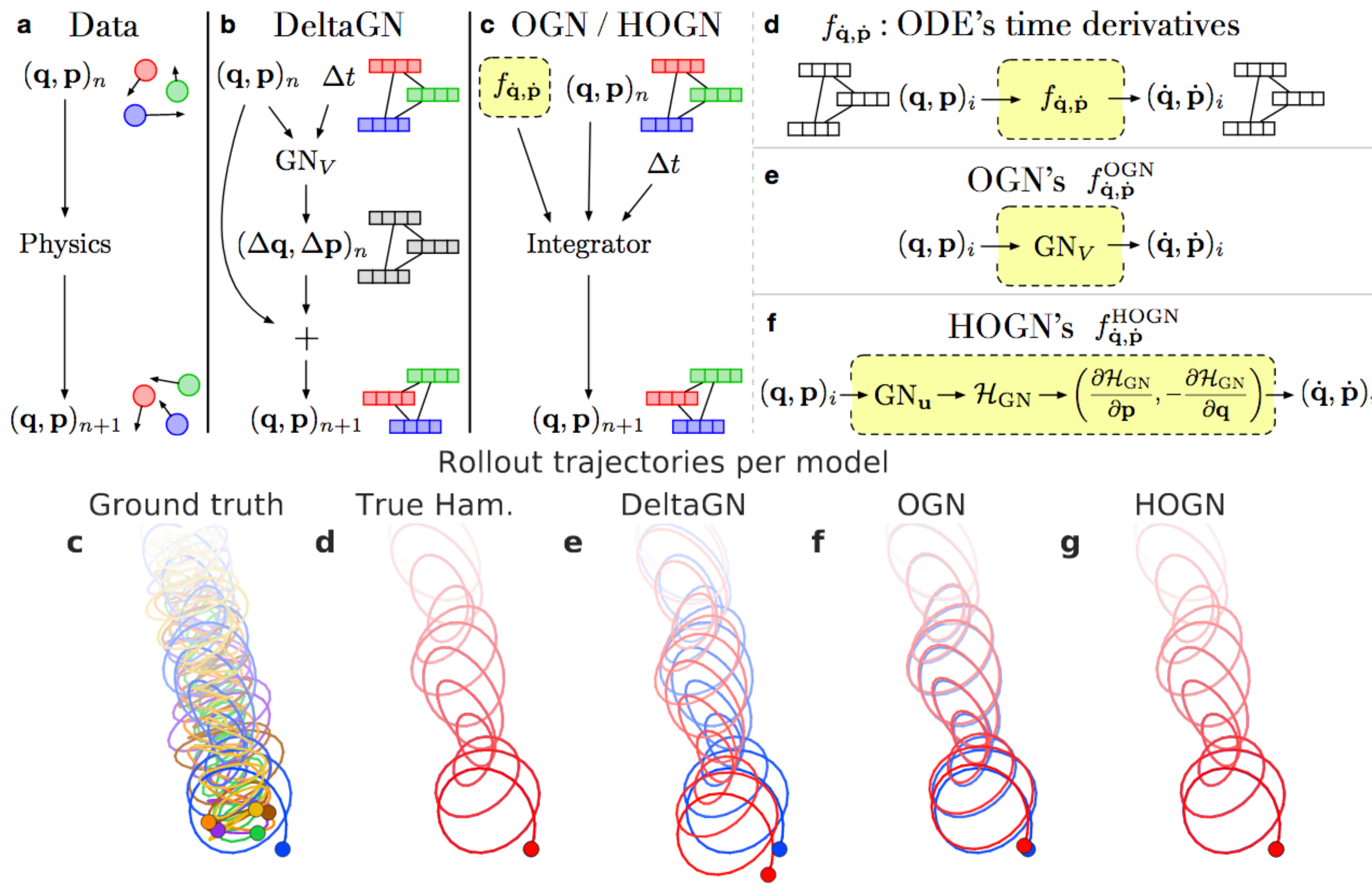
Physics Knowledge



P. Komiske, E. Metodiev, J. Thaler, [1810.05165]

Machine Learning can **help understand Physics**.

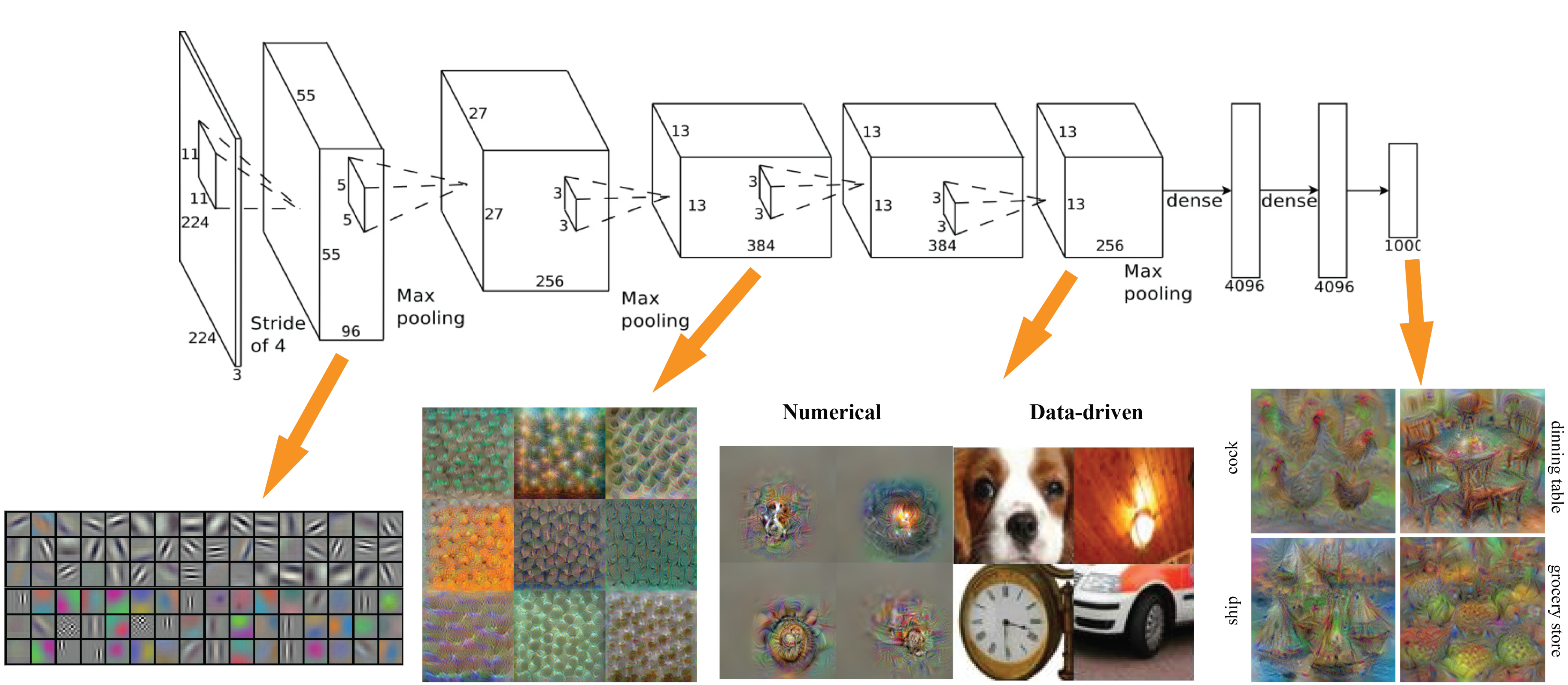
Use Physics



A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, P. Battaglia [1909.12790]

Let the model **include Physics principles** to master convergence

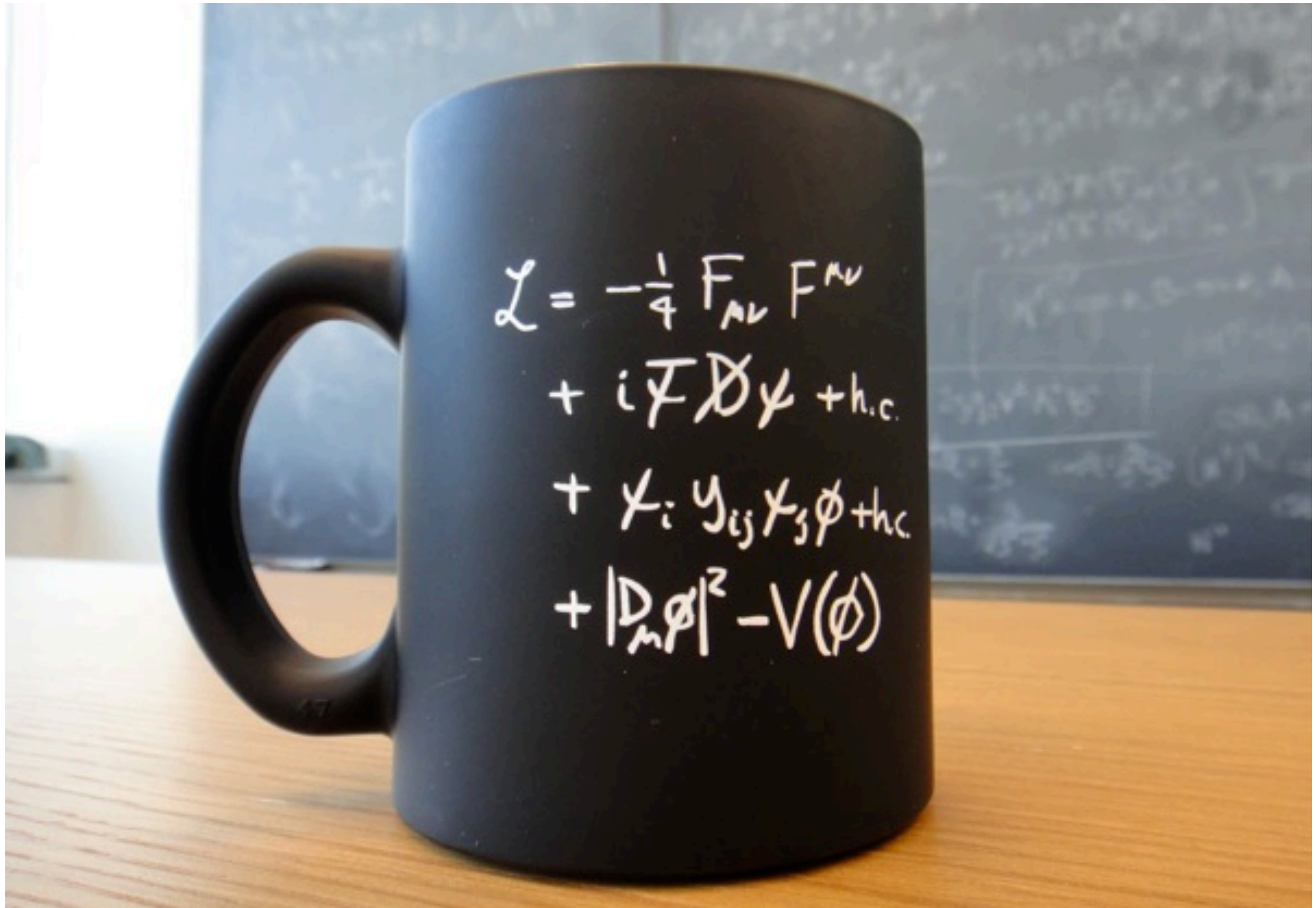
Learning from Complexity



Conv 1: Edge+Blob | Conv 3: Texture | Conv 5: Object Parts | Fc8: Object Classes

Machine learning model can **extract information from complex dataset.**
More classical algorithm counter part may
take **years of development.**

The Standard Model



Well demonstrated effective model.
Good amount of detailed, **“labelled” simulation available**.

The Sea Beyond Standard Model

Slide: A. Wulzner [H&N]

HEP yesterday

"Almost" **Simple H_1**

Focus on **few sharply-defined** alternative models (e.g., the Higgs)

Case-by-case design of **optimal test**

HEP today

"Very" **Composite H_1**

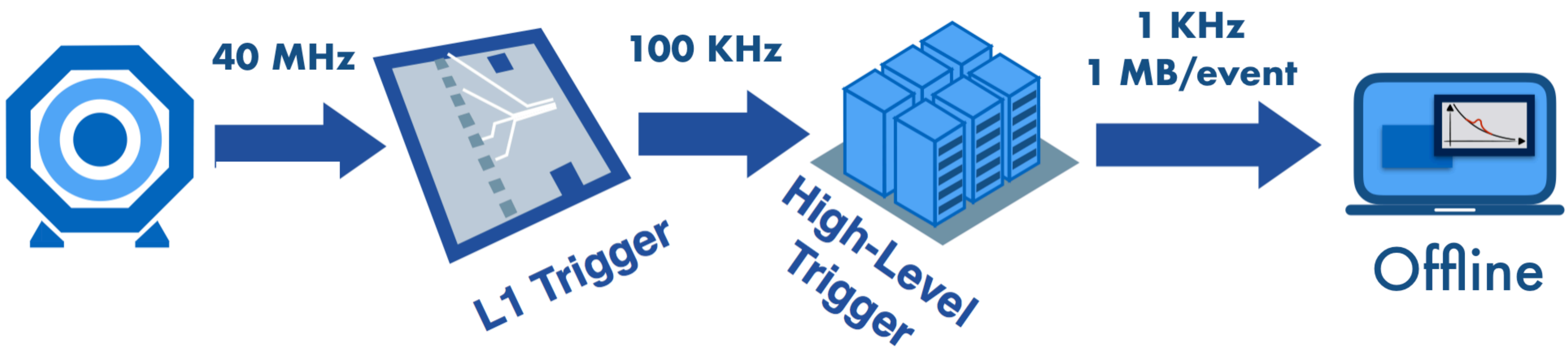
Huge set of alternatives

Case-by-case optimisation **unfeasible**

The **right H_1** likely **not yet formulated**

Event Triggering

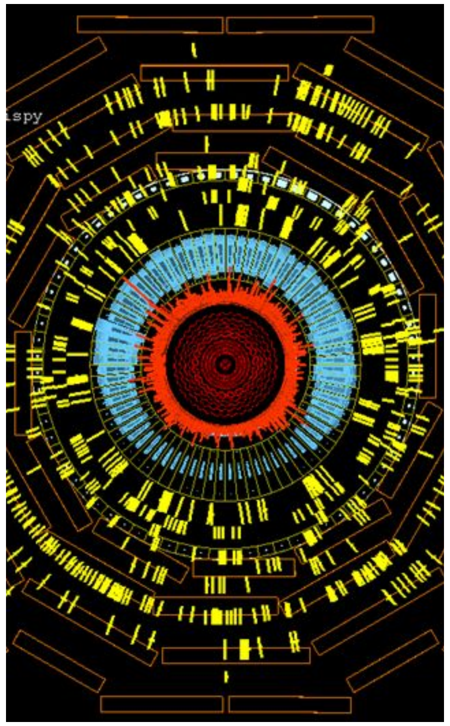
Select what is important to keep for analysis.
Ultra fast decision in hardware and software.

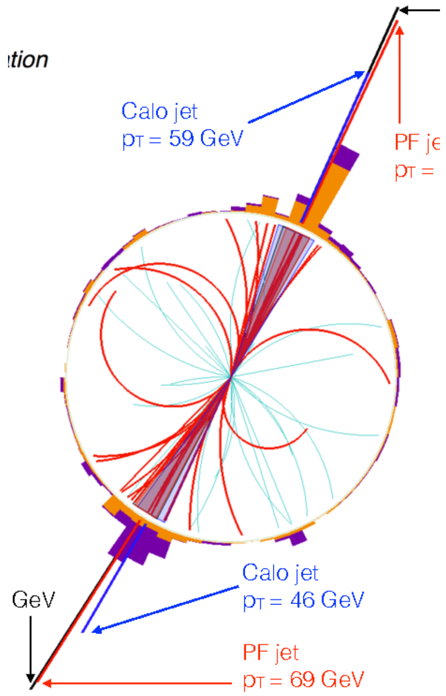


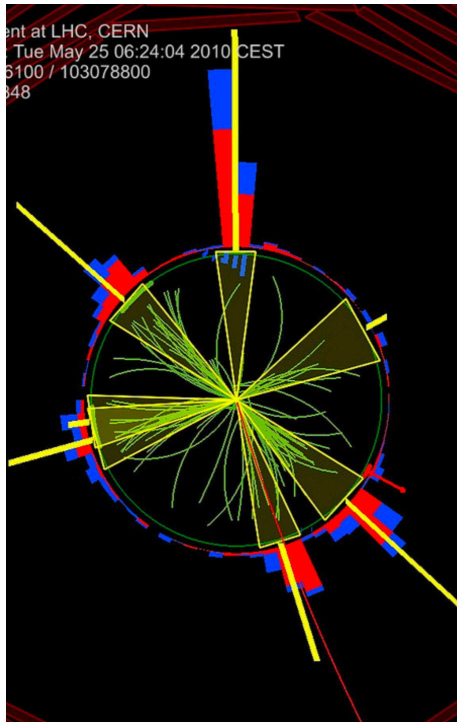
Reconstruction of the event under limited latency / bandwidth.
Better resolution help lowering background trigger rates,
Faster algorithms helps making more refined decisions.

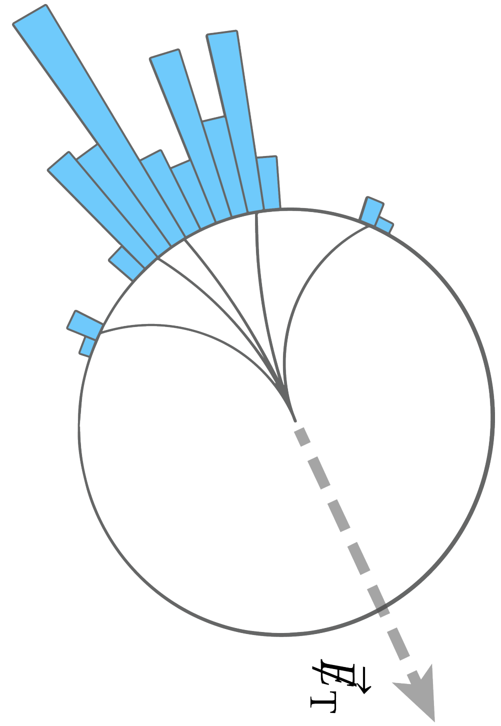
Reconstructing Collisions

Detector Data	Local reconstruction	Particle representation	Jet Clustering	High level features









Event Processing

Dimensionality reduction

Globalization of information

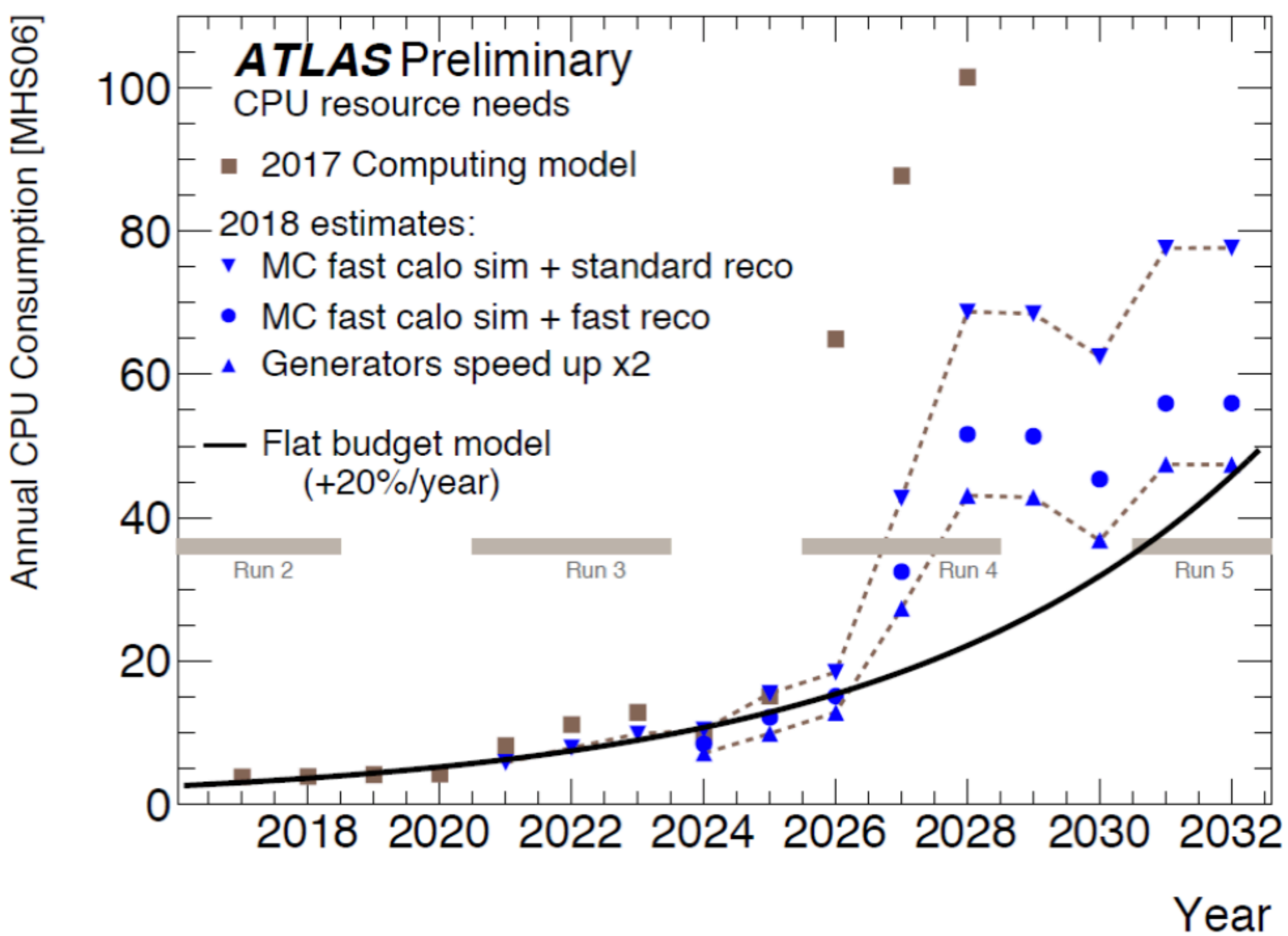
From detector signal to high-level features using **mostly pattern recognition**.
Complex and **computing intensive** series of tasks.

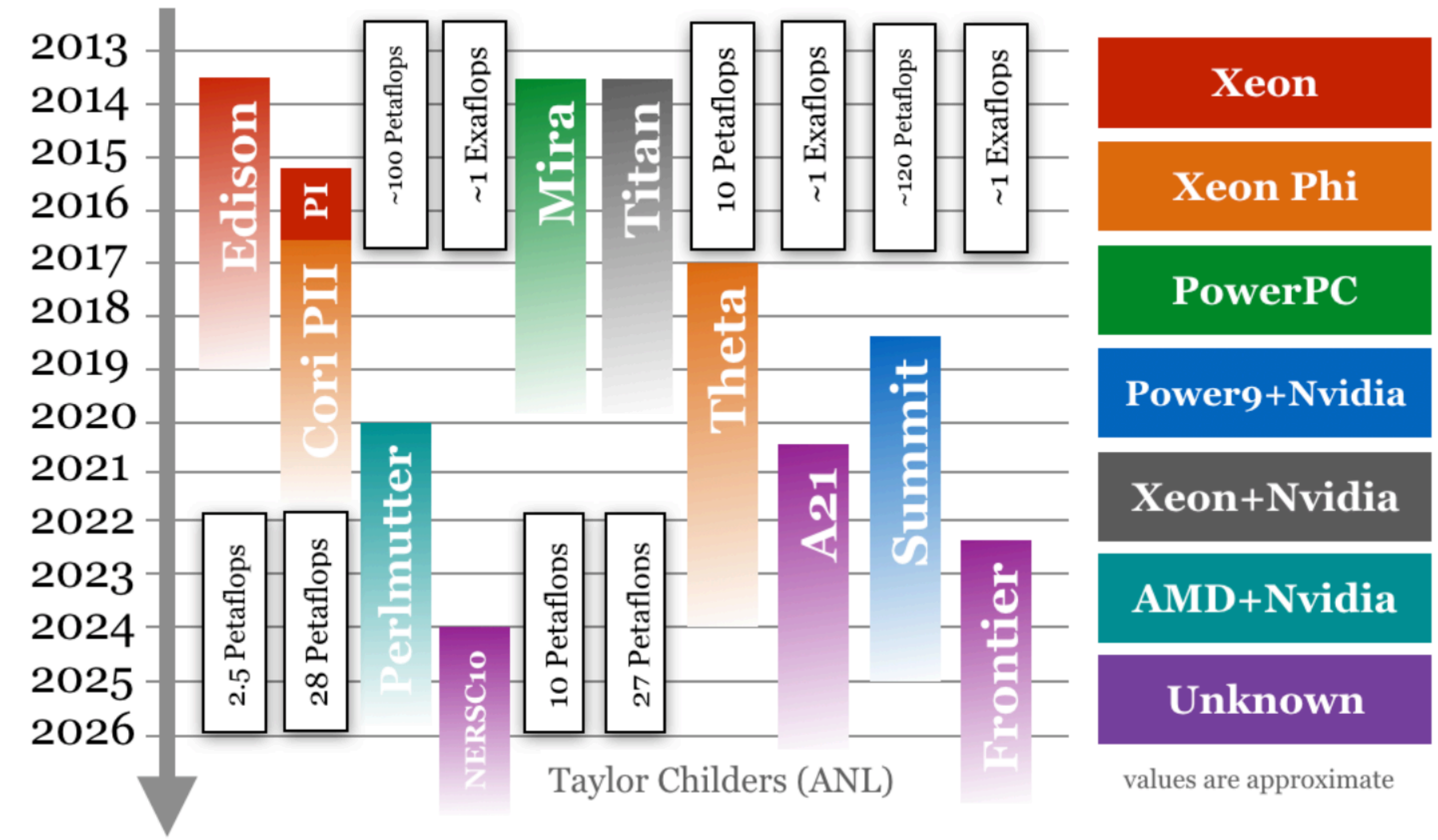
Simulating Collisions

Madgraph, Pythia, Sherpa, ...	**Event Generator**: compute predictions of the standard models to several orders of expansion in coupling constants (LO, NLO, NNLO, ...) using proton density functions.
Pythia, ...	**Hadronization**: phenomenological model of the evolution of hadrons under the effect of QCD.
GEANT 4, GEANT V	**Material simulator**: transports all particles throughout meters of detector, using high resolution geometrical description of the materials.
Homegrown software	**Electronic emulator:** converts simulated energy deposits in sensitive material, into the expected electronic signal, including noise from the detector.

Non-differentiable, **computing intensive** sequence of **complex simulators** of the signal expected from the detectors.

The Computing Cost of Science

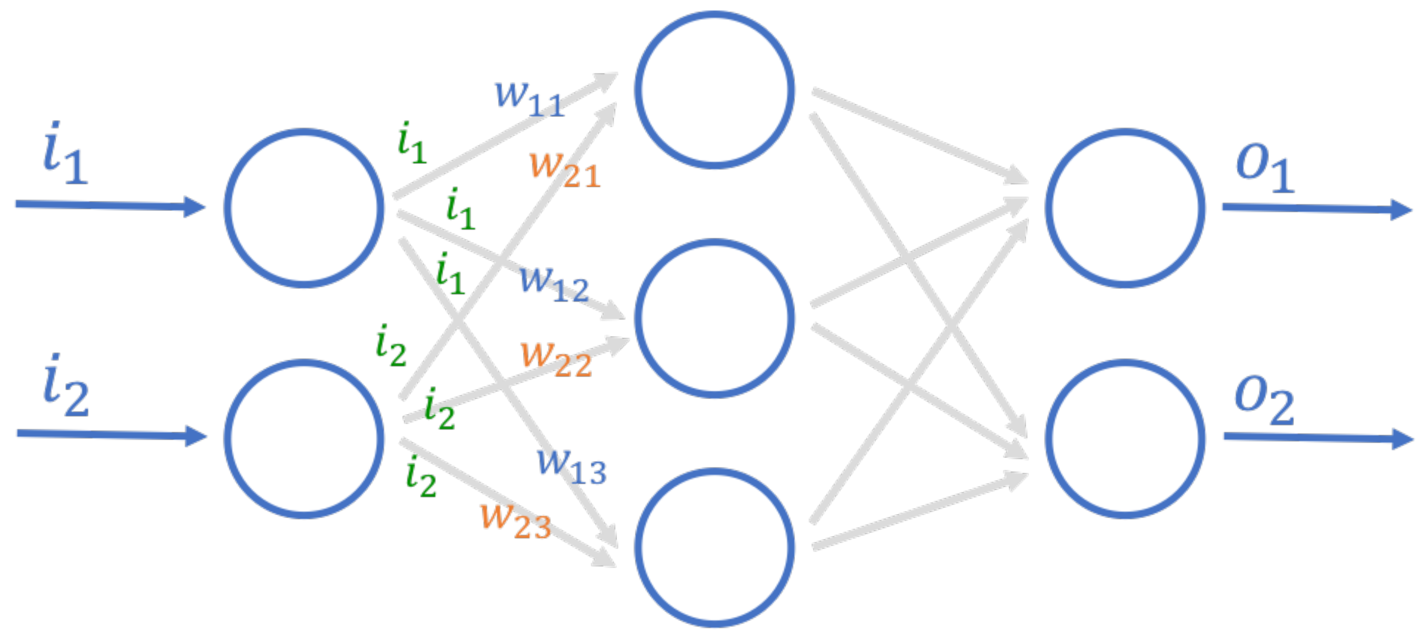




https://indico.cern.ch/event/822126/contributions/3500169/

Ever growing needs for computing resource.
Slowdown of classical architecture, over growth of GPU architecture.

Operation Vectorization

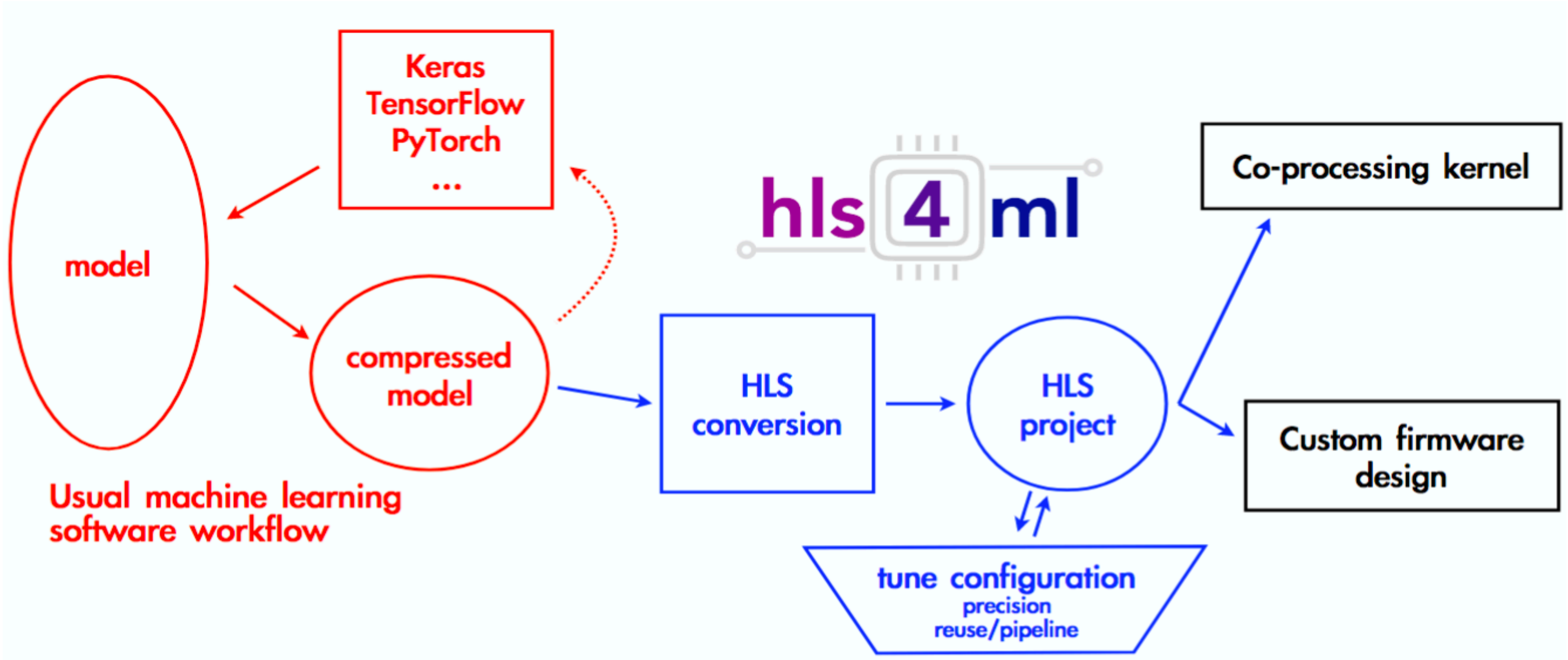


ANN ≡ matrix operations ≡ parallelizable

$$\begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ w_{13} & w_{23} \end{bmatrix} \cdot \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} (w_{11} \times i_1) + (w_{21} \times i_2) \\ (w_{12} \times i_1) + (w_{22} \times i_2) \\ (w_{13} \times i_1) + (w_{23} \times i_2) \end{bmatrix}$$

Computation of prediction from artificial neural network model can be **vectorized to a large extend.**

Hyper-Fast Prediction



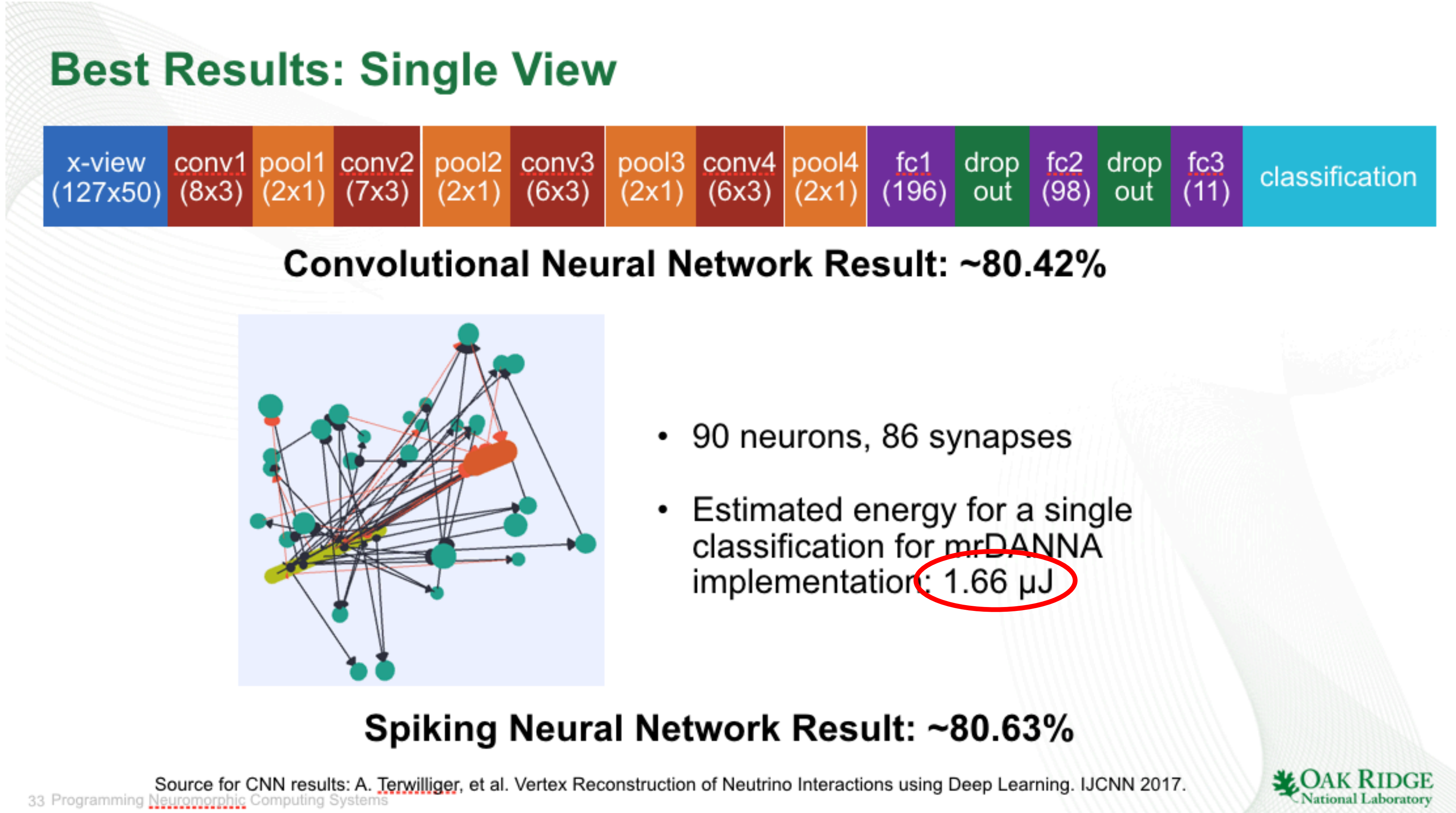
Synthesizing FPGA firmware from trained ANN

https://fastmachinelearning.org/hls4ml/

J. Duarte et al.[1804.06913]

Artificial neural network model can be **executed efficiently on FPGA**, GPU, TPU, ...

Low Power Prediction



https://indico.fnal.gov/event/13497/contribution/0 *Slide C. Schuman*

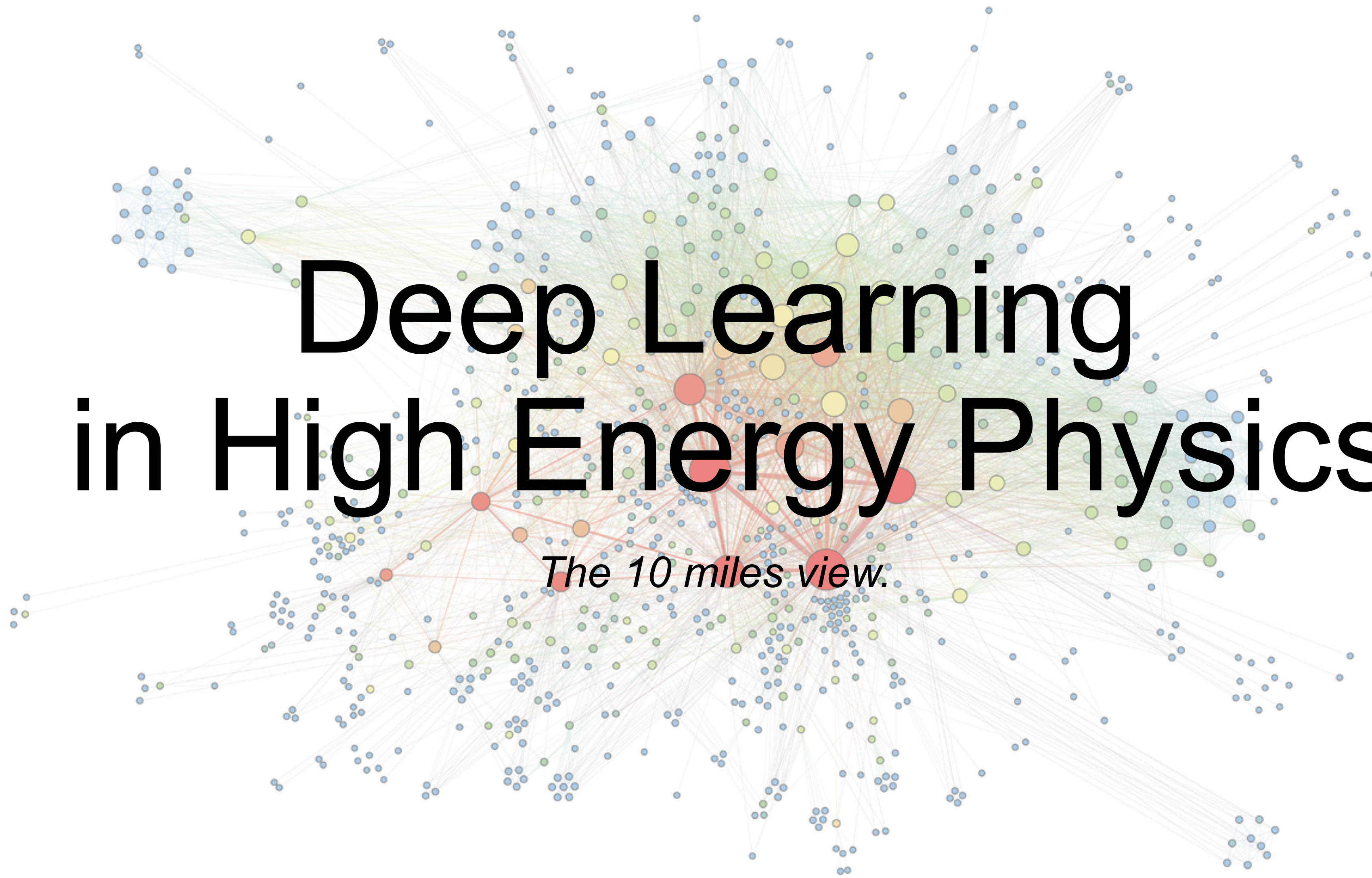
Neuromorphic hardware dedicated to **spiking neural networks**
Low power consumption by design

Take home message :

Machine Learning is a widely recognized and used technology in industry

Deep Learning has the potential of helping Science to make progress

Neural Networks could help with the computing requirements of Science

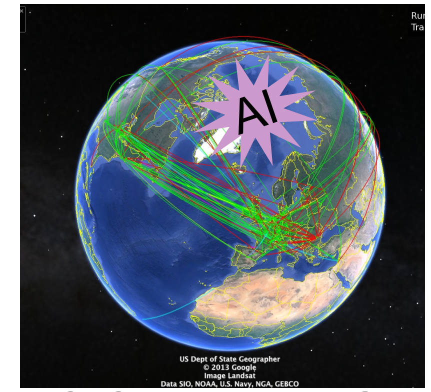


Deep Learning in High Energy Physics

The 10 miles view.

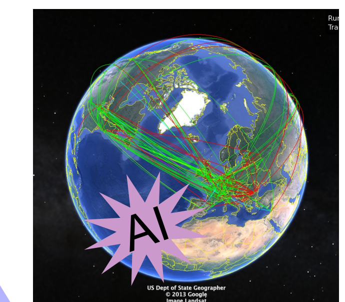
AI in HEP

Role of AI: accelerator control, data acquisition, event triggering, anomaly detection, new physics scouting, event reconstruction, event generation, detector simulation, LHC grid control, analytics, signal extraction, likelihood free inference, background rejection, new physics searches, ...

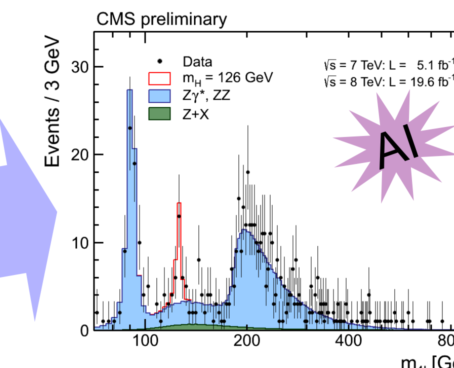


LHC Computing Grid
200k cores pledge to
CMS over ~100 sites

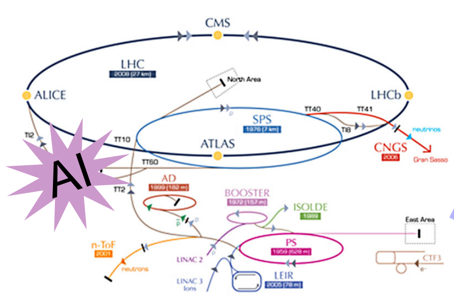
CERN Tier-0/Tier-1
Tape Storage
200PB total



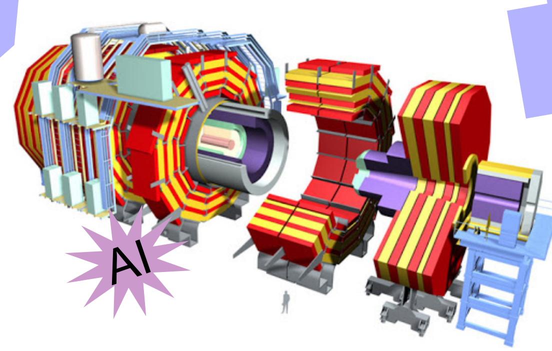
LHC Grid
Remote Access
to 100PB of data



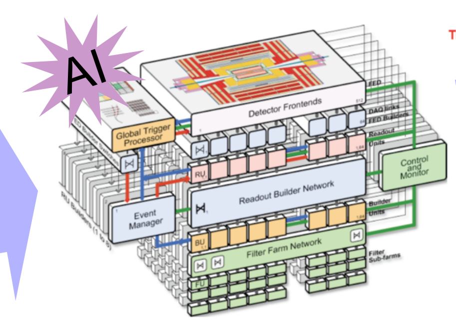
Rare Signal
Measurement
~1 out of 10^6



Large Hadron Collider
40 MHz of collision



CMS Detector
1PB/s

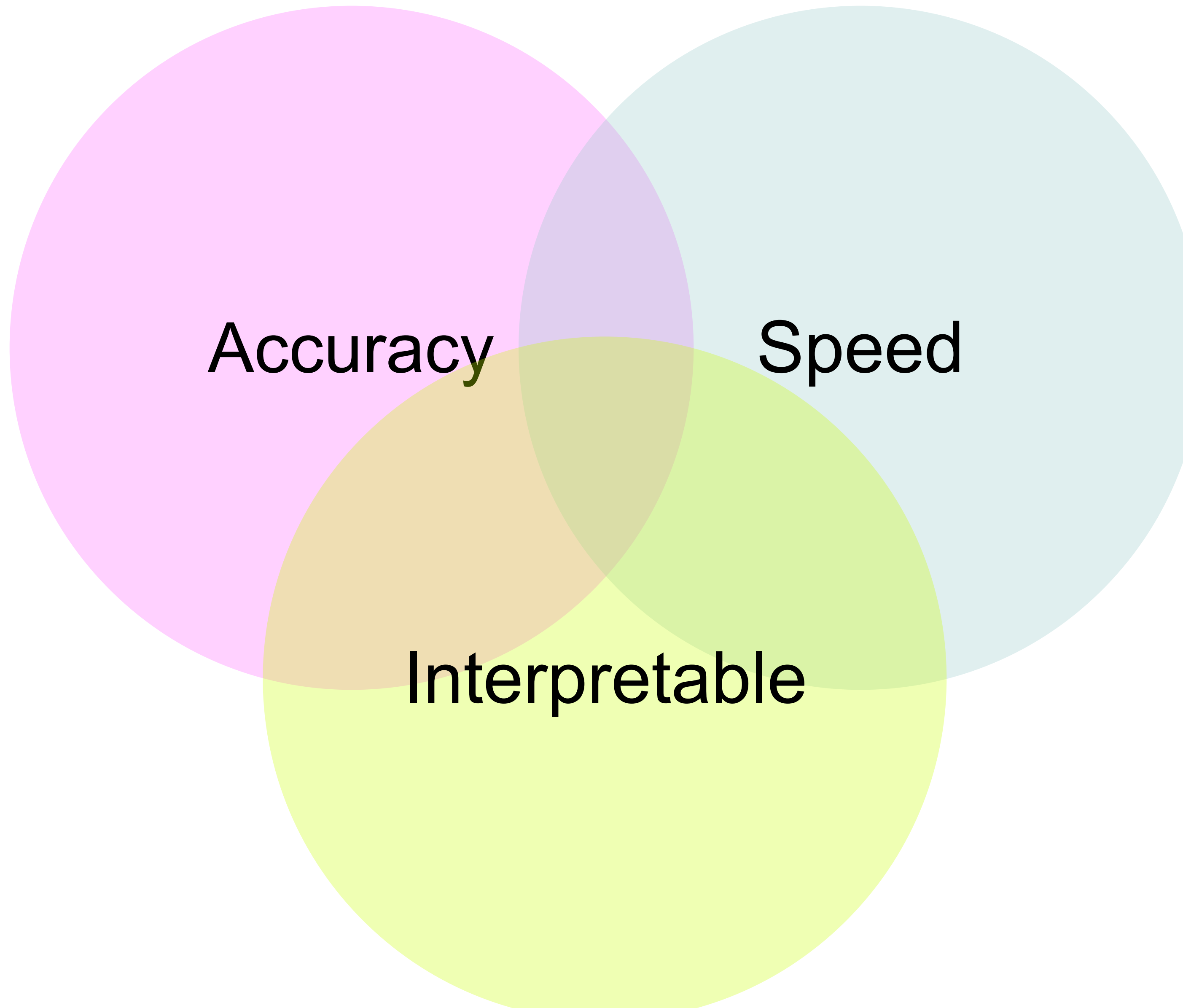


CMS L1 & High-Level Triggers
50k cores, 1kHz

CERN Tier-0
Computing Center
20k cores

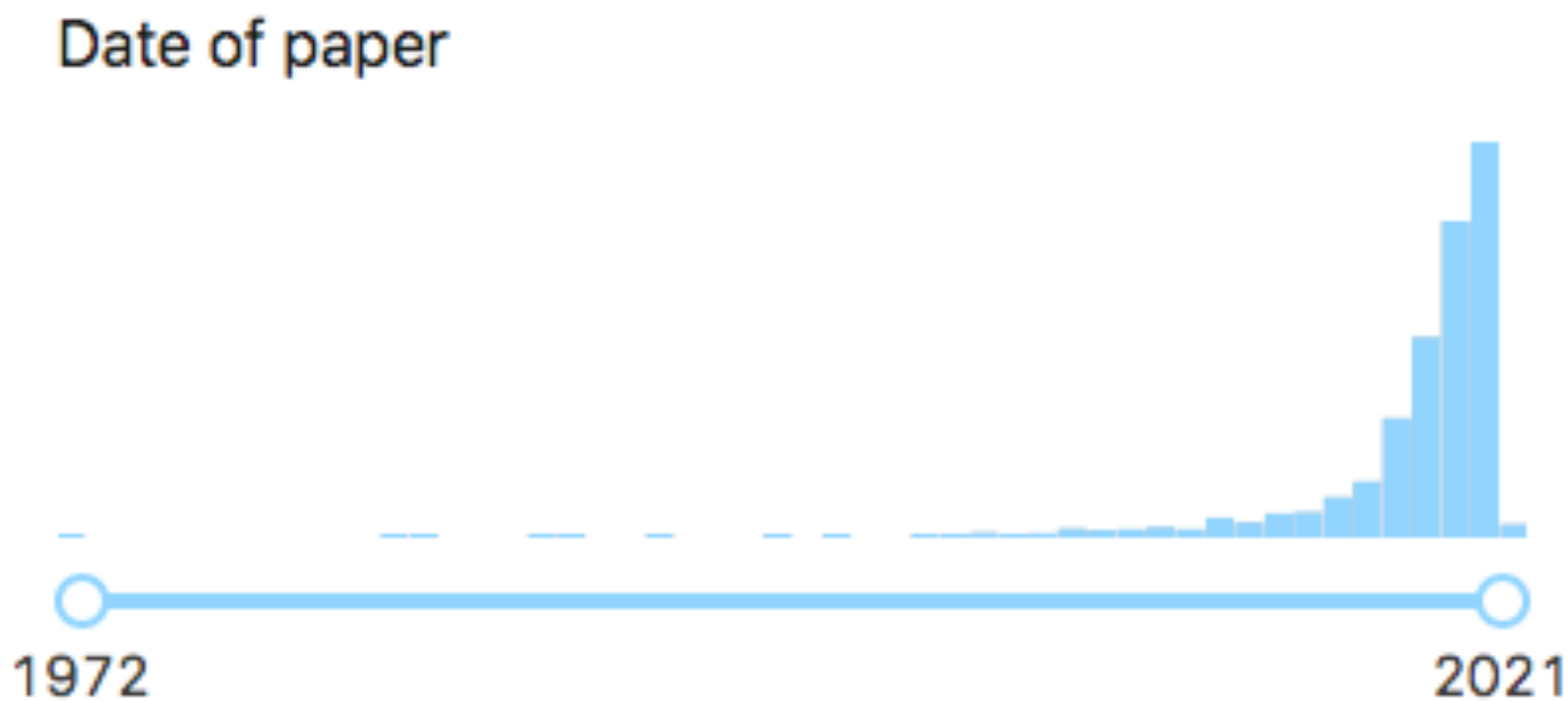
→ Up to date listing of references:
https://github.com/iml-wg/HEPML-LivingReview

Possible Utilizations



- **Fast surrogate** models (trigger, simulation, etc) ; even better if more accurate.
- **More accurate** than existing algorithms (tagging, regression, etc) ; even better if faster.
- Model performing **otherwise impossible tasks** (operations, etc)

Growing Literature



https://inspirehep.net/literature?q=machine learning or deep learning

Community-based up to date listing of references
https://iml-wg.github.io/HEPML-LivingReview/

Comments on Literature

- Most work and publications on fast simulation (Delphes, etc) :
  - ✓proof of concepts
- Numerous open datasets available on various tasks :
  - ✓simplifies greatly benchmarking
- Trend of sharing software with publication :
  - ✓improves community-wise effort.
- Growing number of publications by the collaborations :
  - ✓the “real deal”

Specialized journals appeared in the recent years:

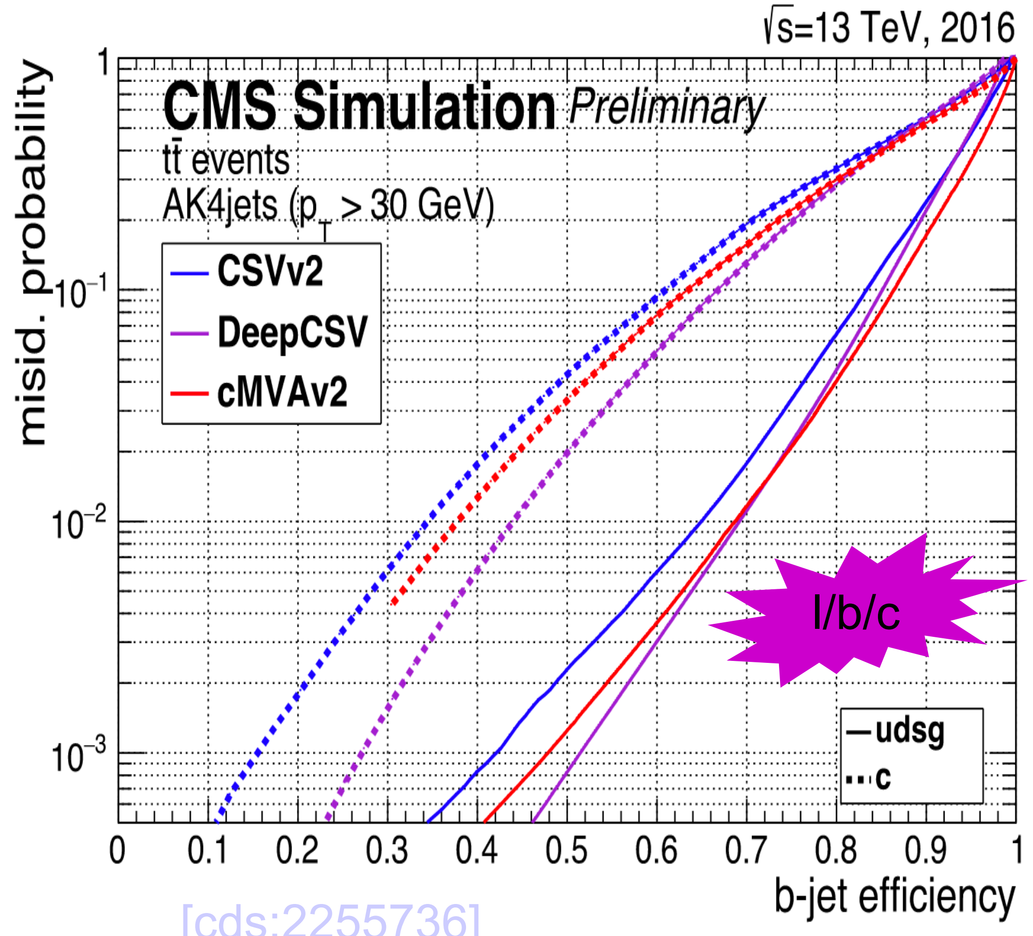
- Computing and Software for Big Science (CSBS)
  https://www.springer.com/journal/41781
- Machine Learning: Science and Technology (MLST)
  https://iopscience.iop.org/journal/2632-2153
- Big Data and AI in HEP
  https://www.frontiersin.org/big-data-and-ai-in-high-energy-physics

Experiment adoption takes time.
Adaptive publication rules might incentivize integration.

Deep Learning in CMS

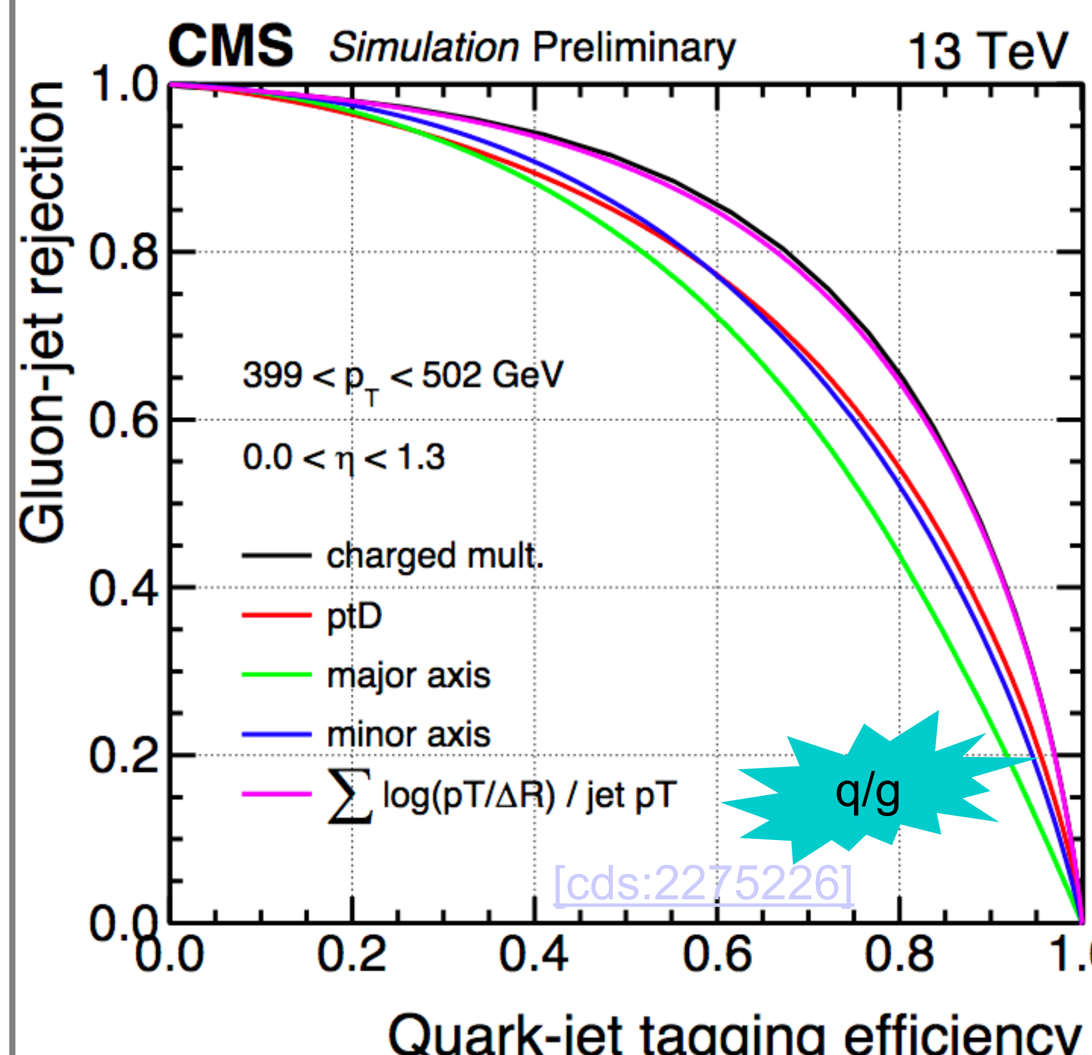
- Machine learning in particle identification, energy regression, and S/B classification since long ; yielding improved sensitivity
- Deep learning entered in jet tagging, monitoring, object identification, regression ; all state of the art performance
- Being carefully calibrated and deployed in analysis
- Deep learning R&D at upstream levels ; operation, monitoring, L1, anomaly search, track reconstruction, calorimeter reconstruction, pileup mitigation, etc (more in backup slides)

Jet Tagging

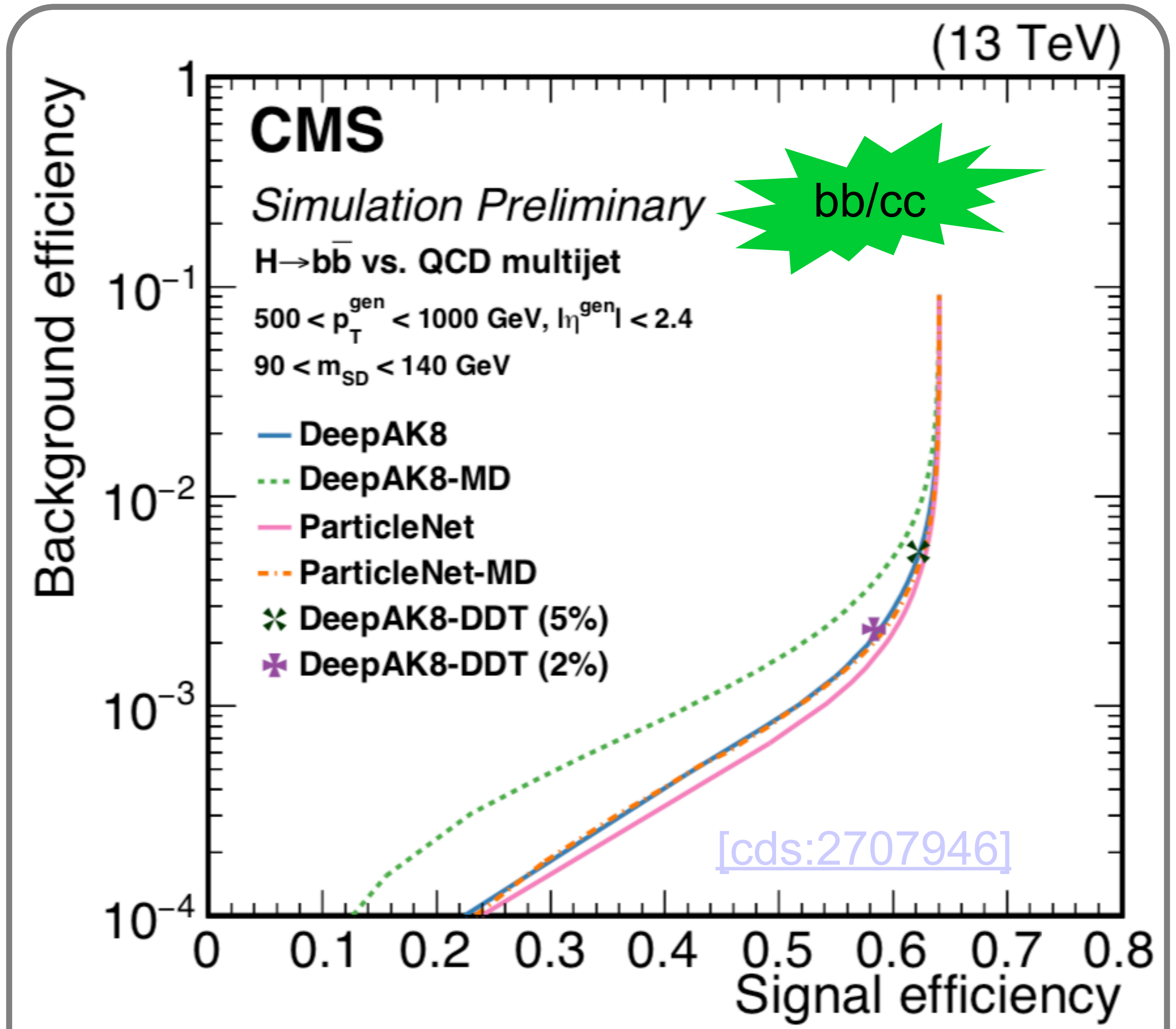


Combining jet-level, vertex-level and track-level features in FC neural net.

Jet-level features in a boosted decision tree for Quark/Gluon discrimination.

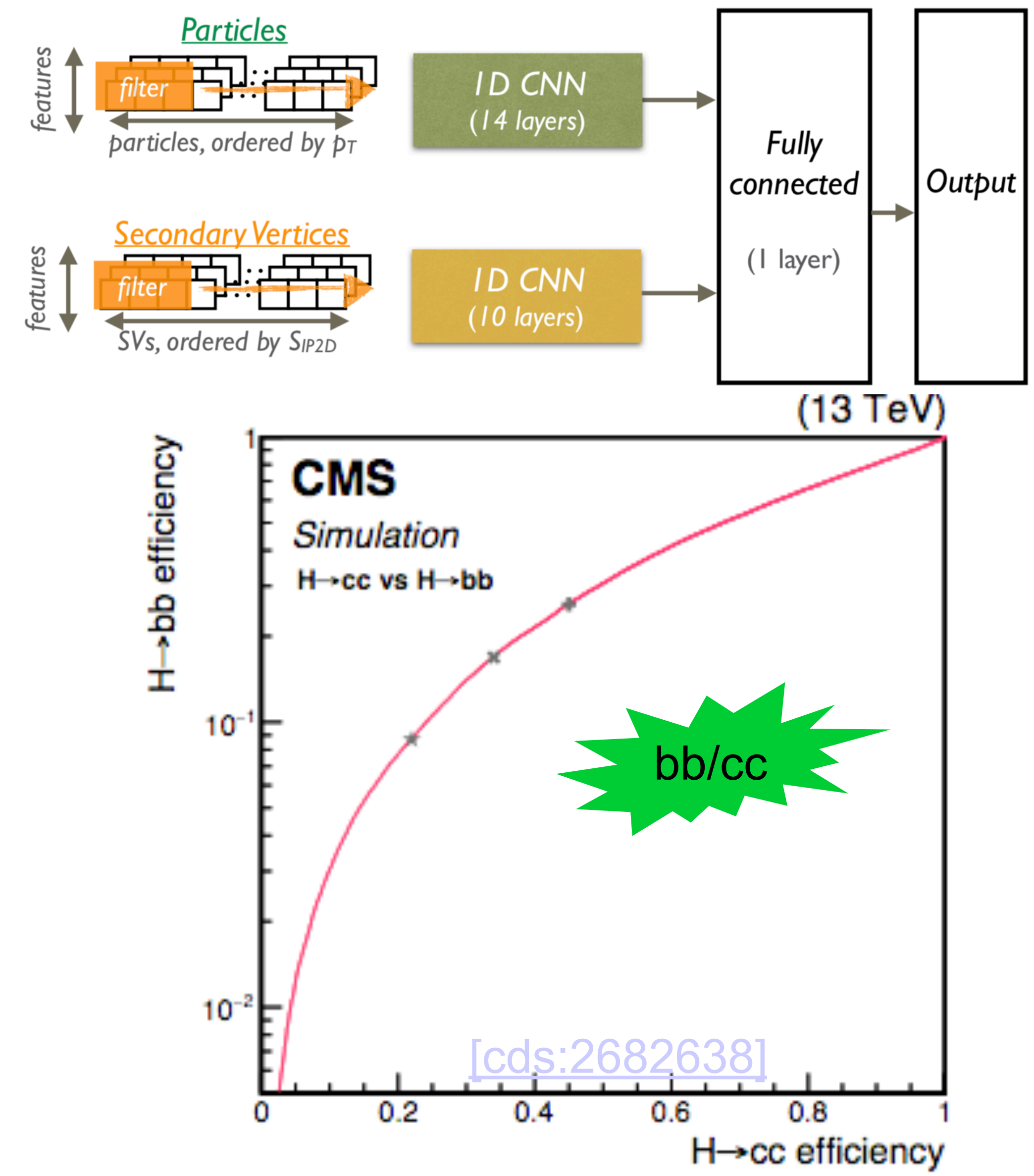


Higgs Tagging



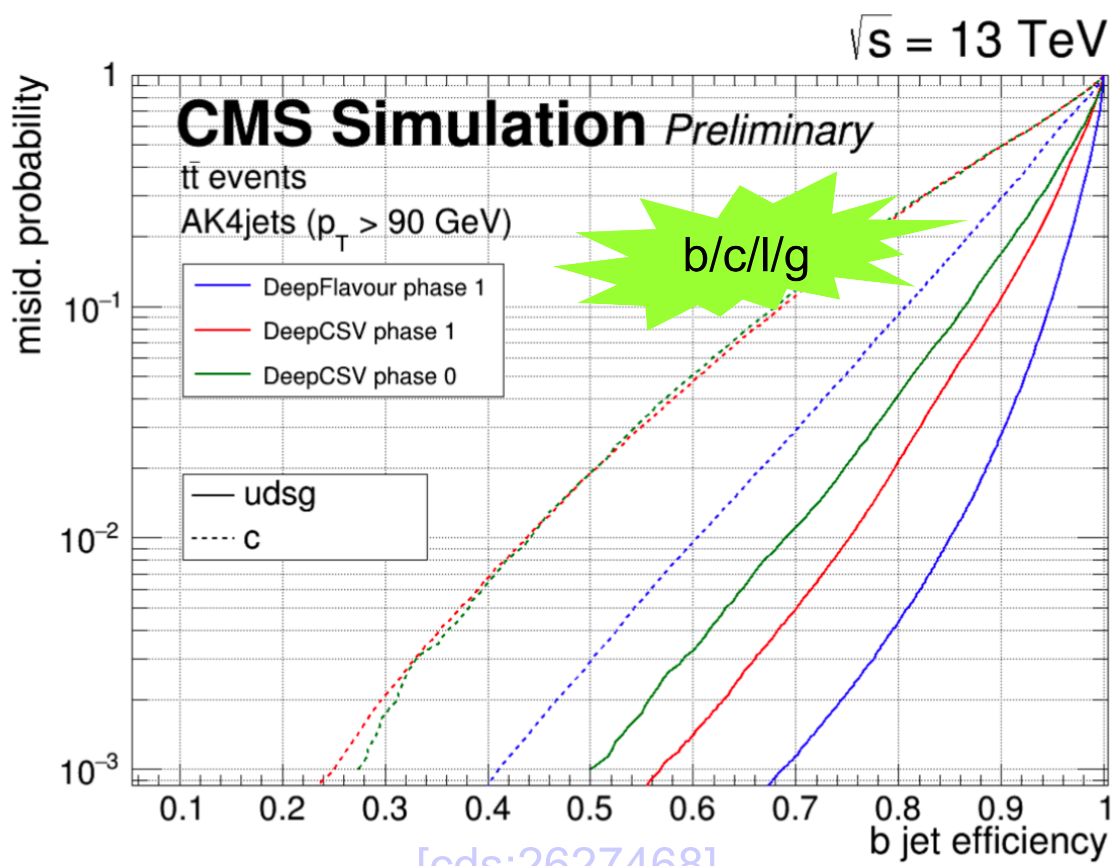
Using particle flow candidates as input to the Particle-Net architecture.

Combines particle and secondary vertex features in Conv1D/FC

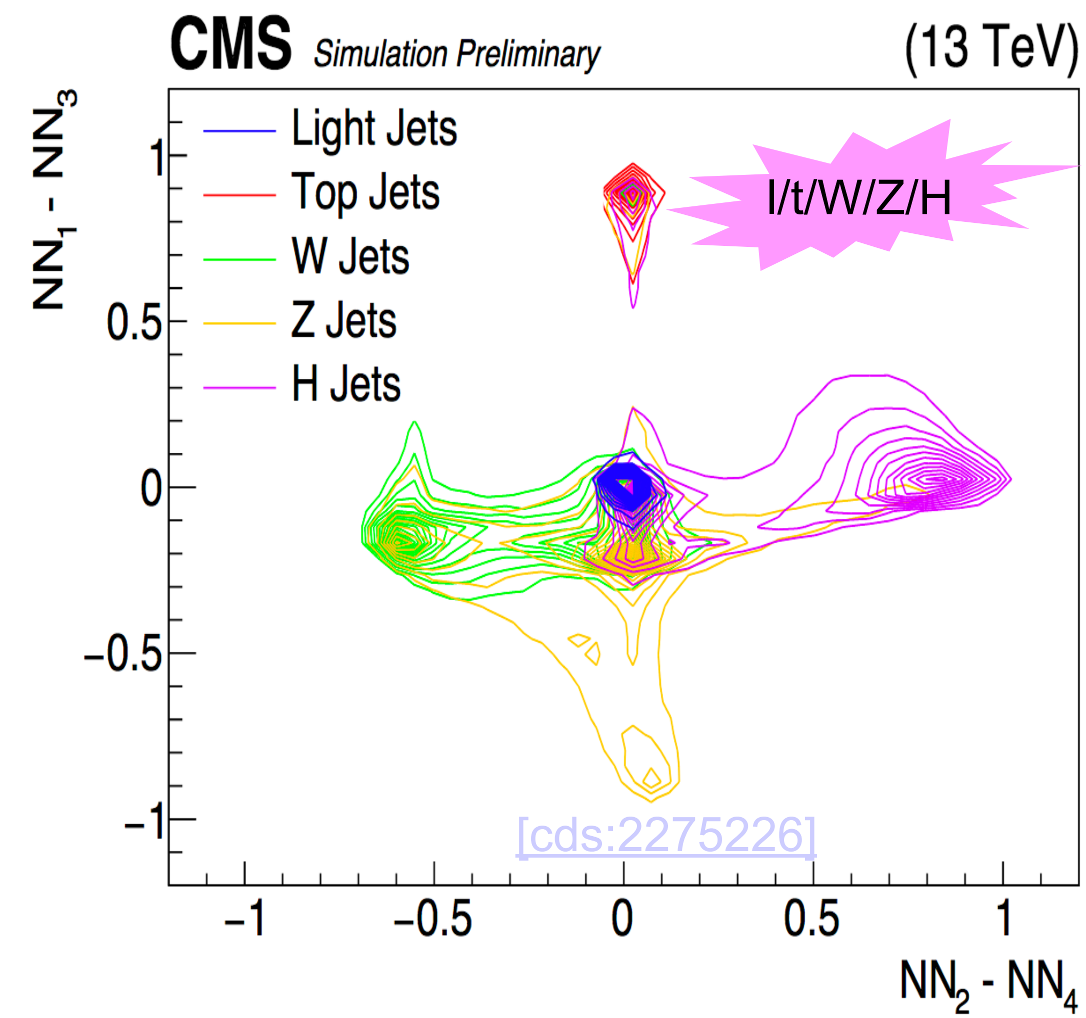


Jet x-Tagging

Jet-level, vertex-level and particle-level features in Conv1D/LSTM/FC multi-class neural net.



[cds:2627468]

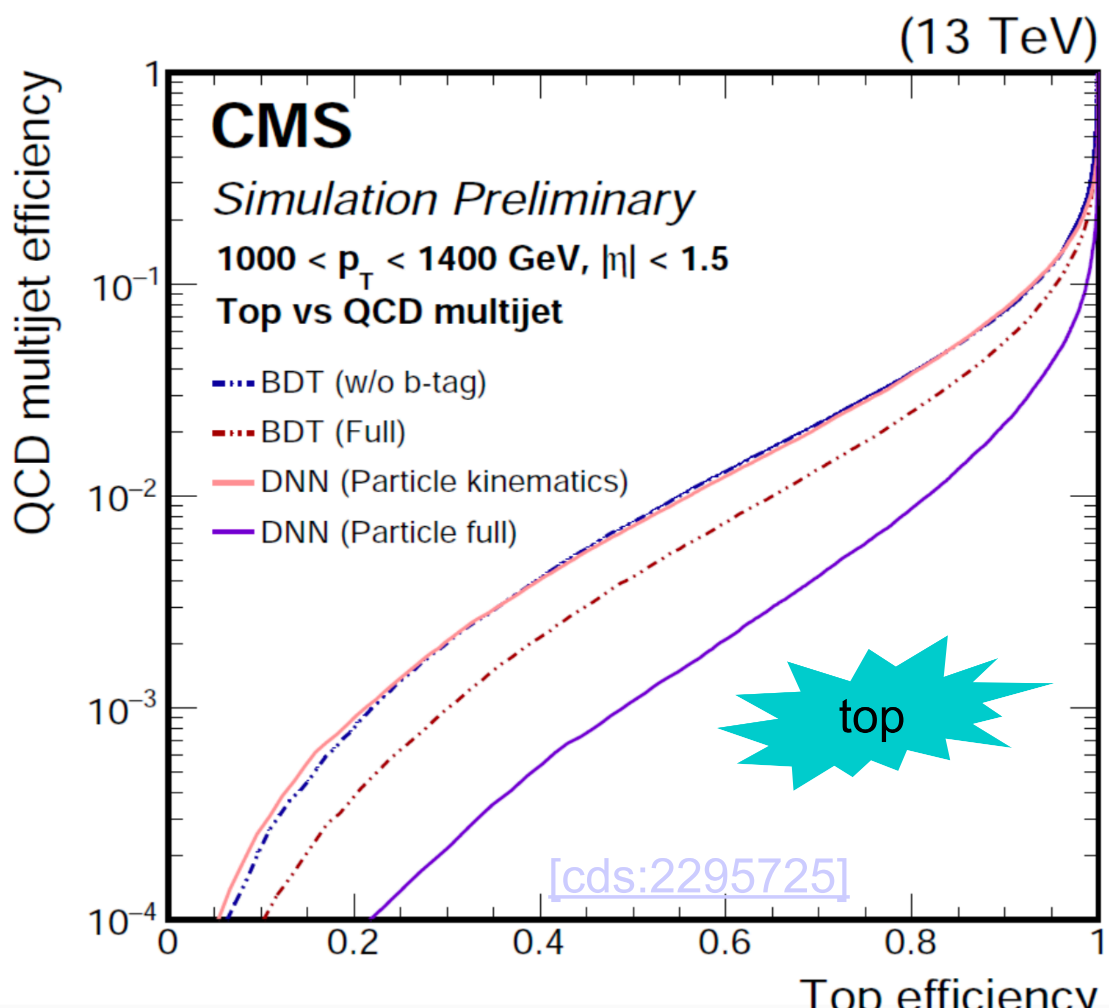


[cds:2275226]

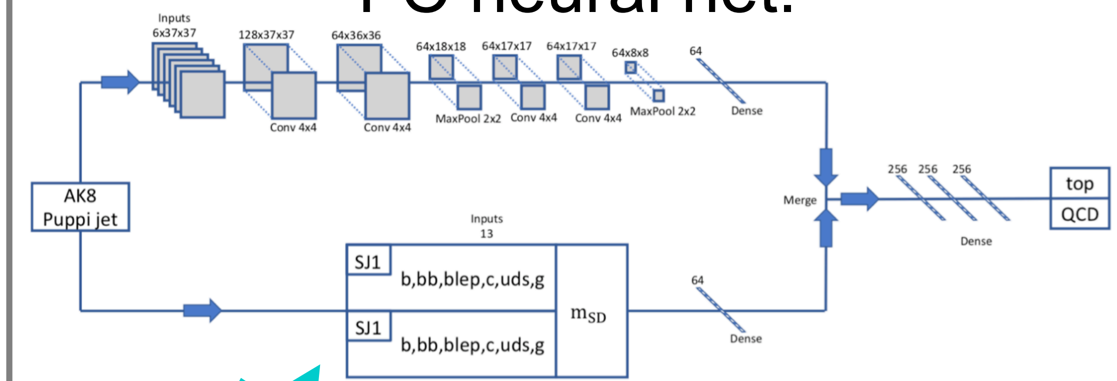
Jet-level features computed in hypothesized rest frame in dense multi-class neural net.

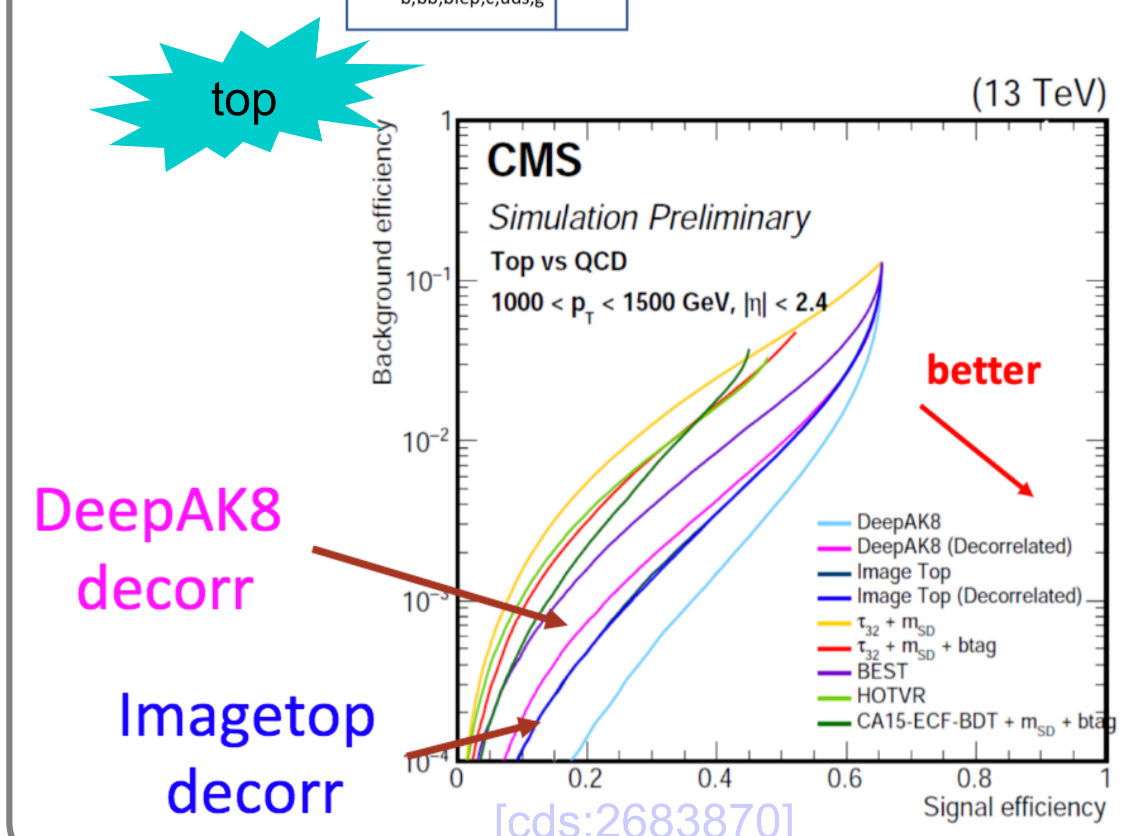
Top Tagging

Combining particle-level, track-level and vertex-level features in Conv1D/ FC neural net.

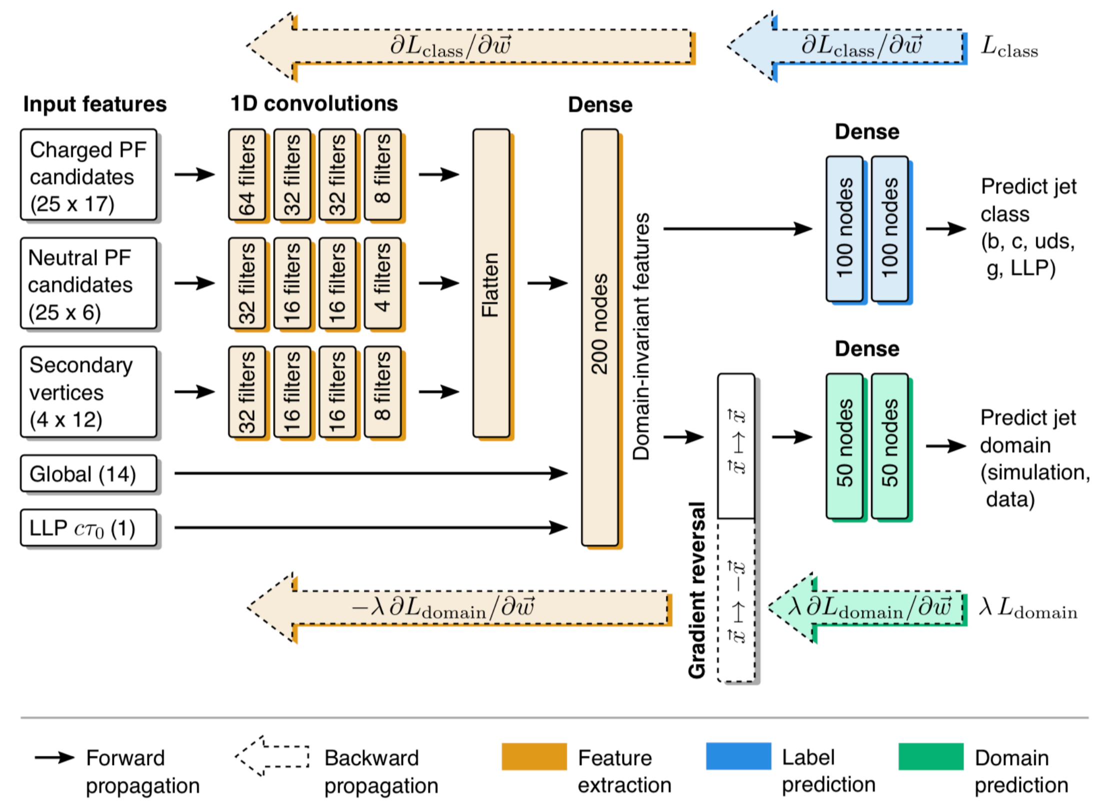


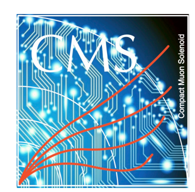
Combining particle-image and subjet tagging features with CNN/ FC neural net.

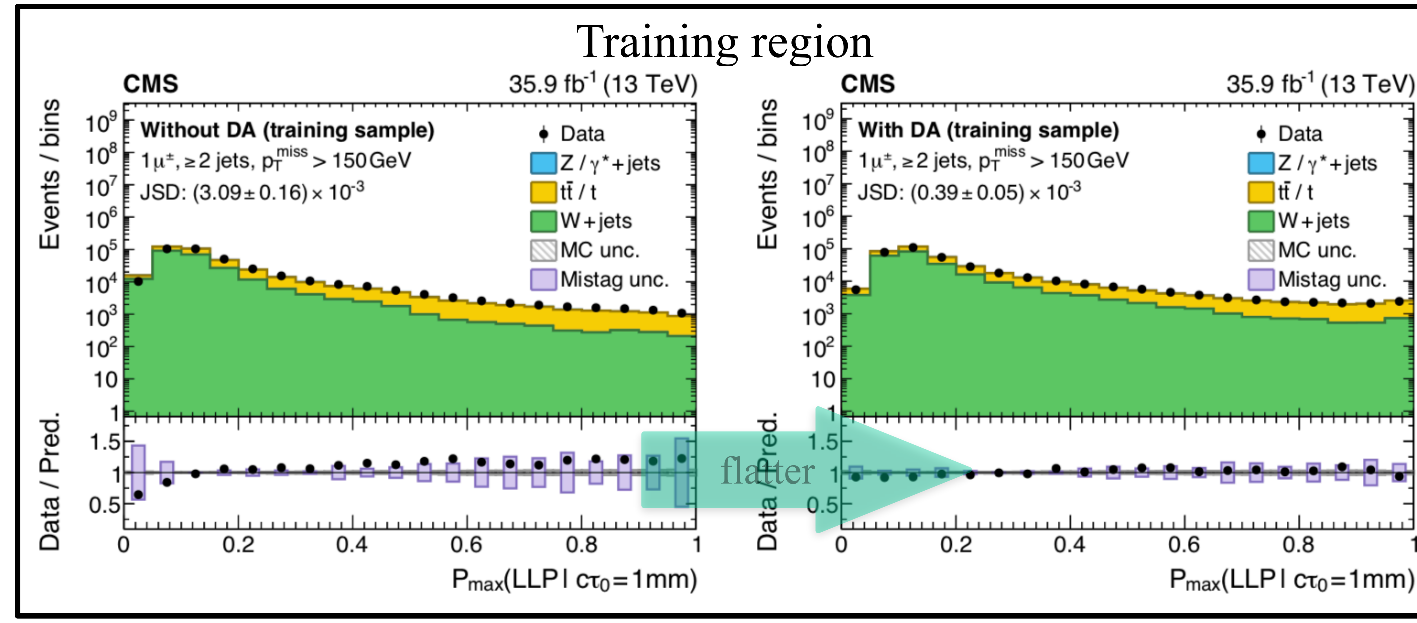


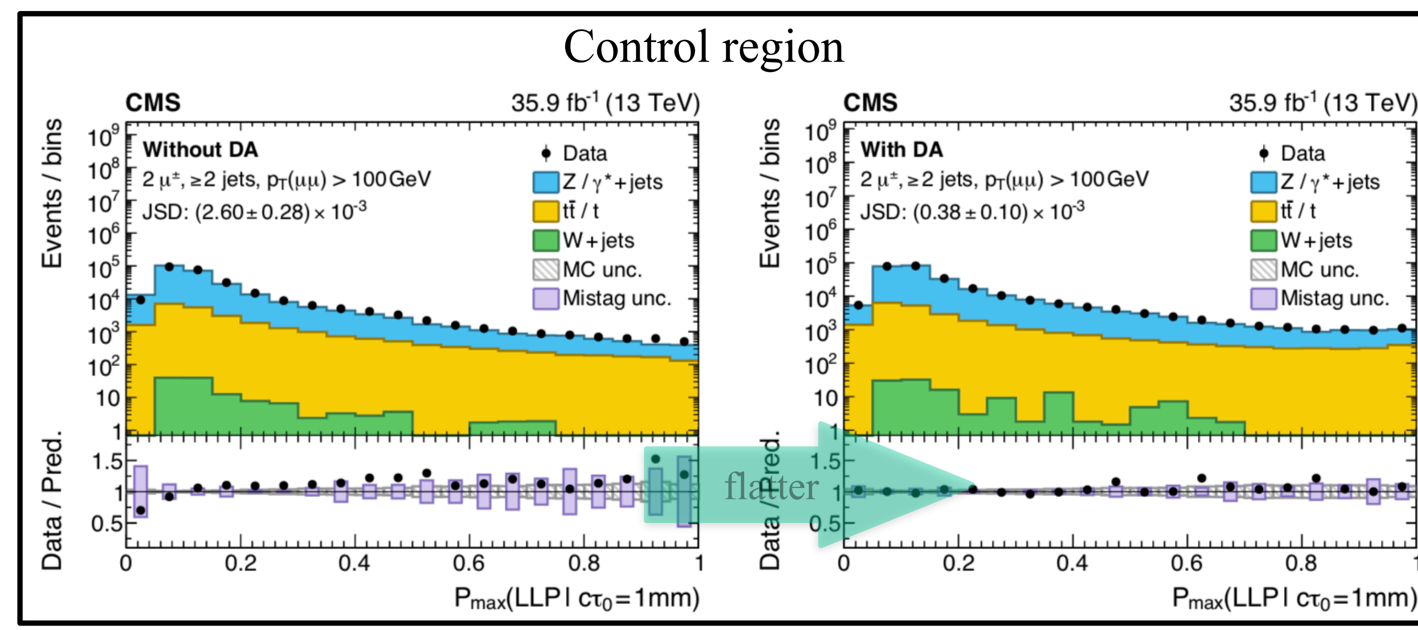


Domain in-Dependence







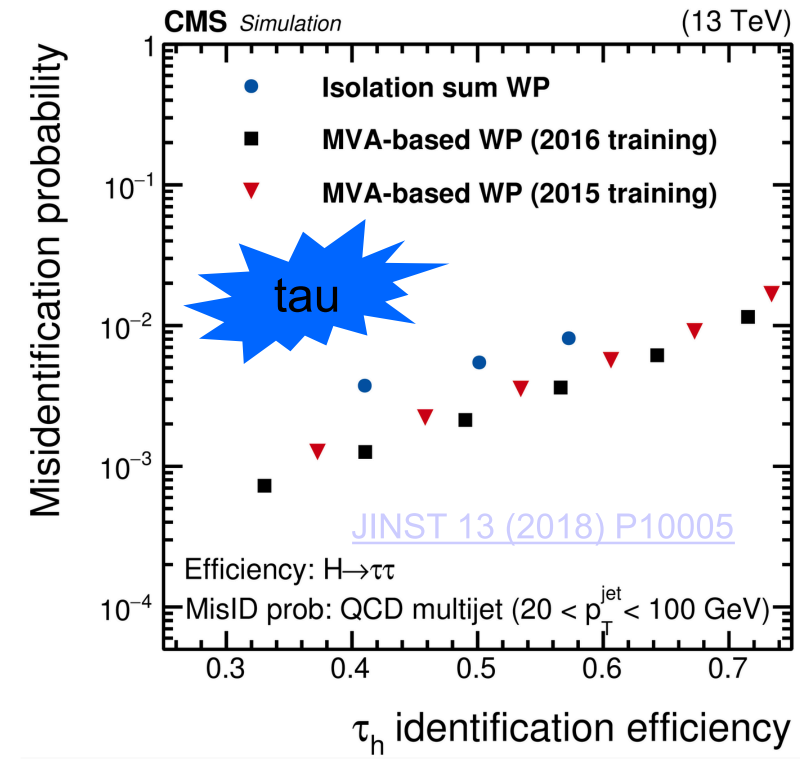


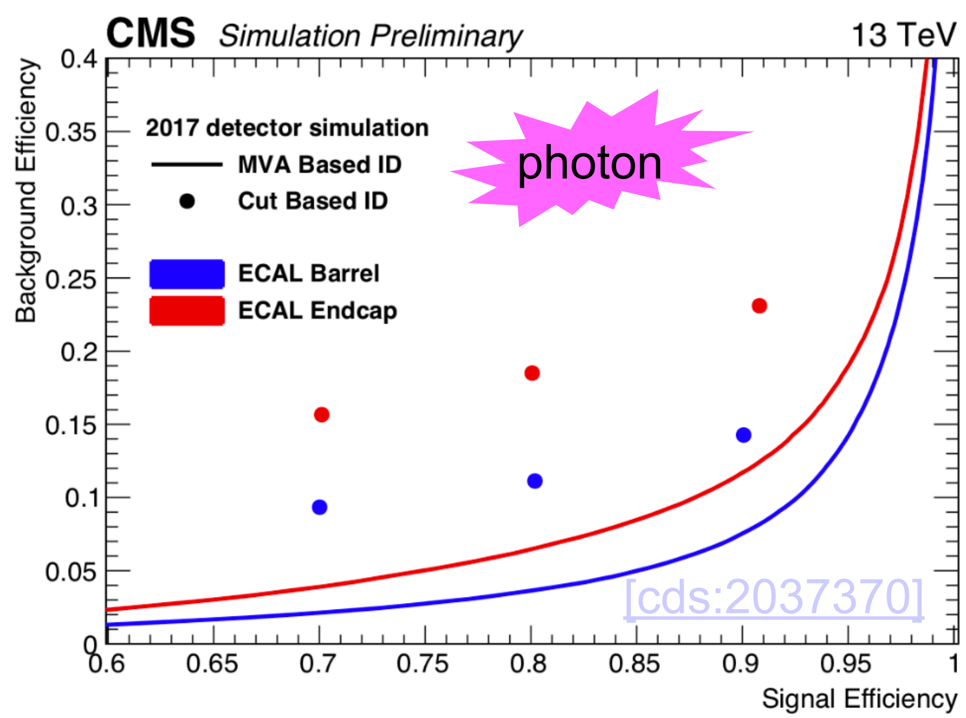
A deep neural network to search for new long-lived particles decaying to jets [doi:10.1088/2632-2153/ab9023]

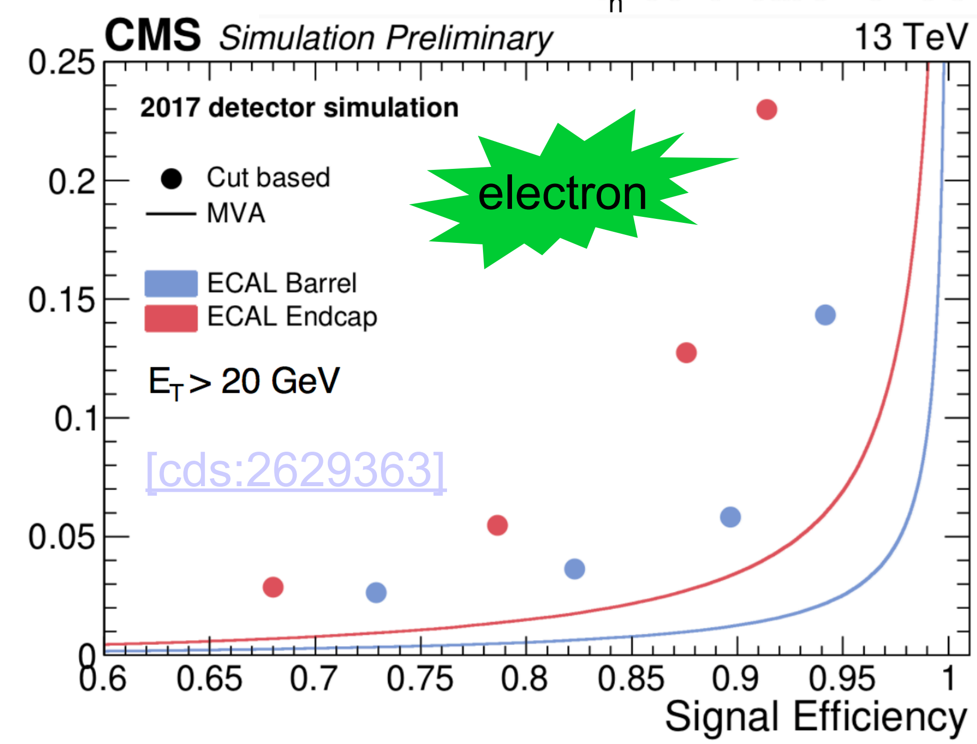
Gradient reversal on a domain-classifier to mitigate the discrepancies of classifier output between data and simulation.

Particle Id

Object-level features boosted decision tree classification. Analysis specific Muon identification with MVA.

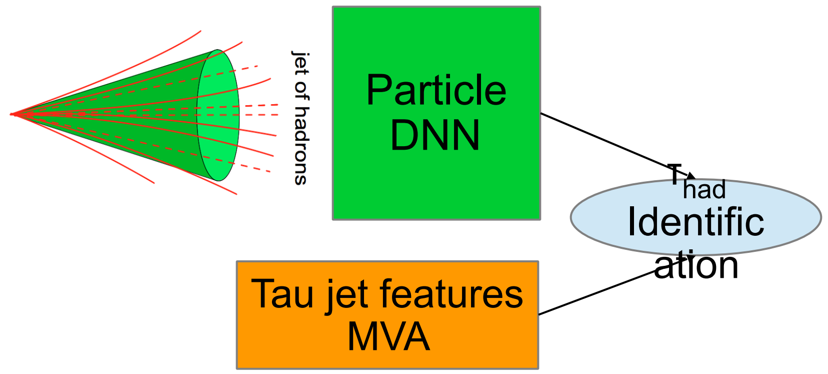


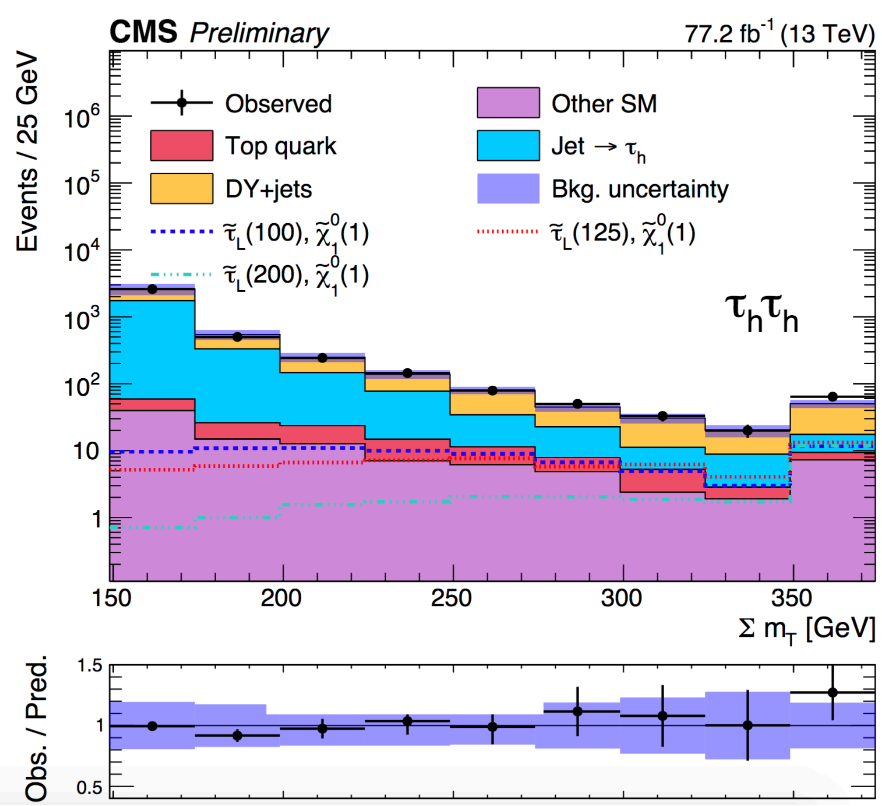




Tau-id with DNN

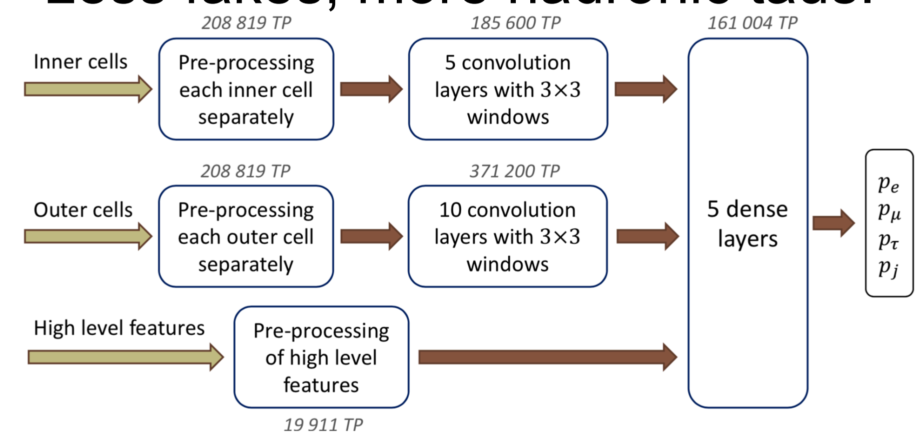
Combine the jet and particle features.
Reduction of fake hadronic taus.

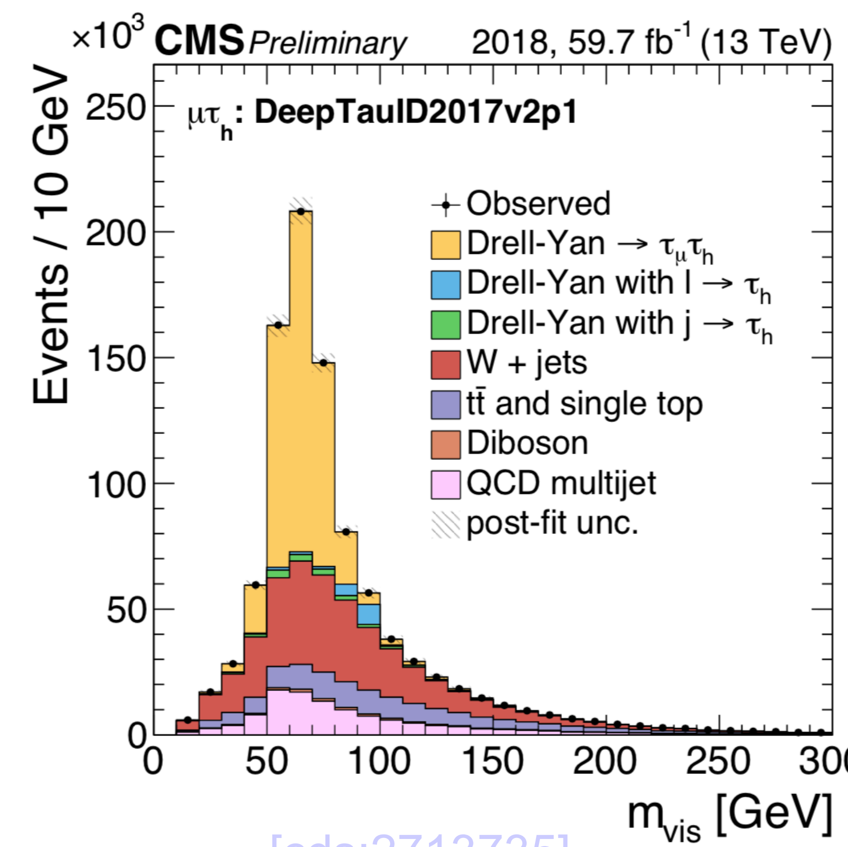




[cds:2669190]

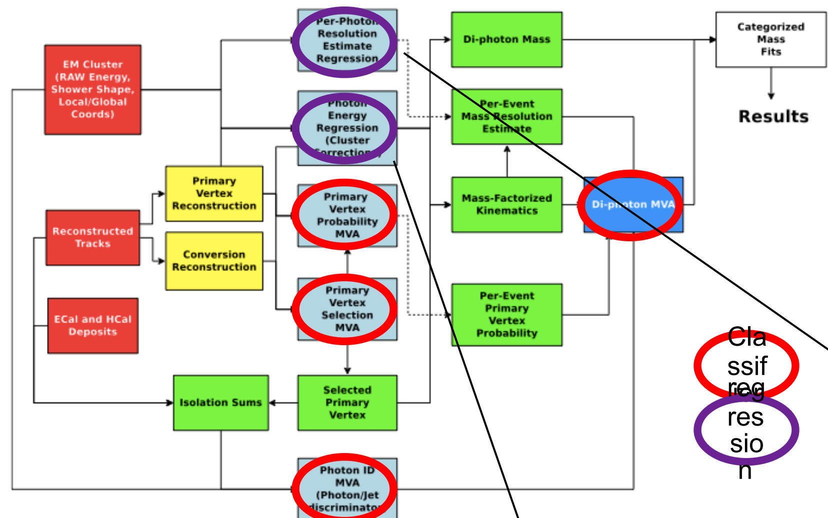
Combines jet and particle-image features.
Less fakes, more hadronic taus.



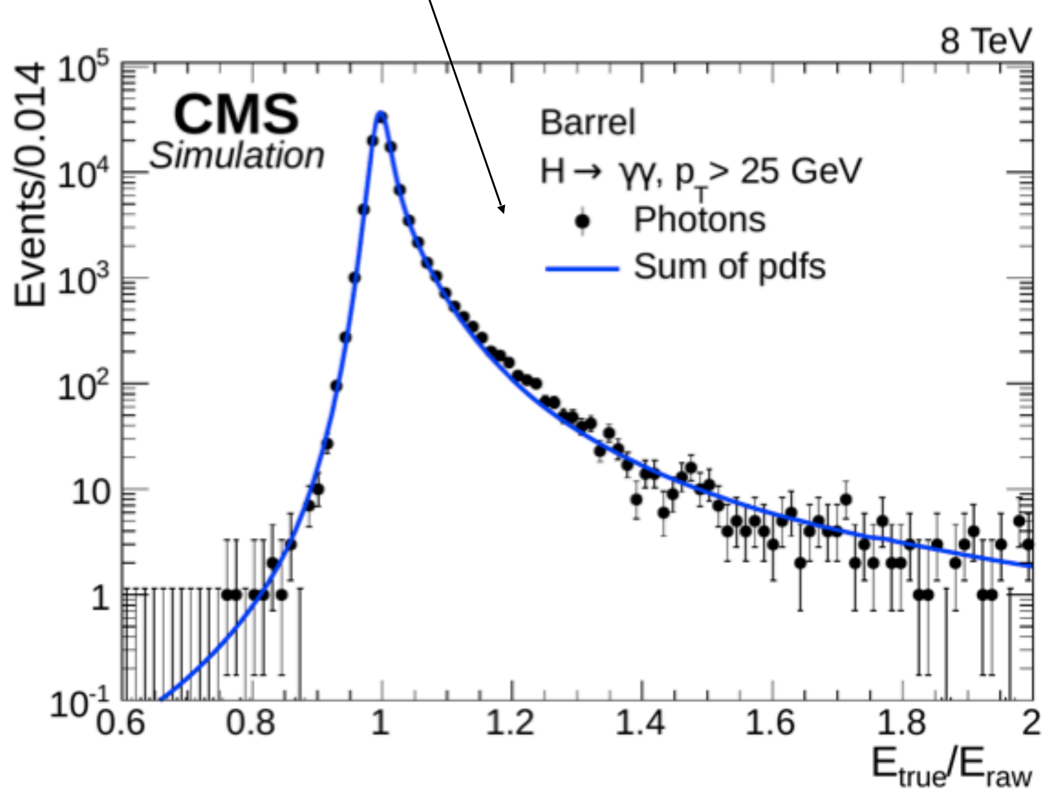


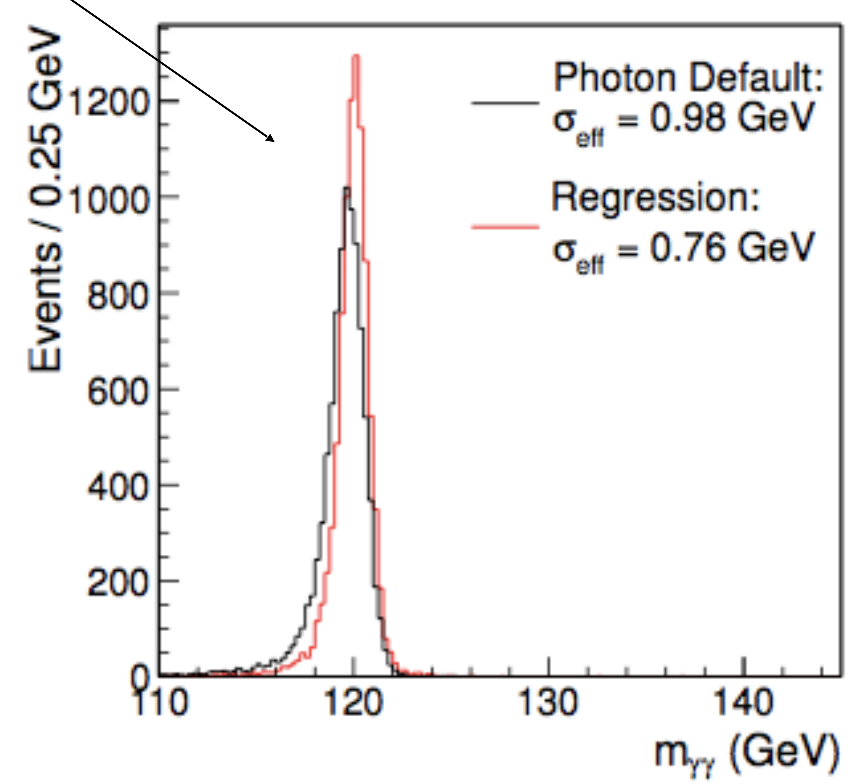
[cds:2713735]

Higgs to gamma^2



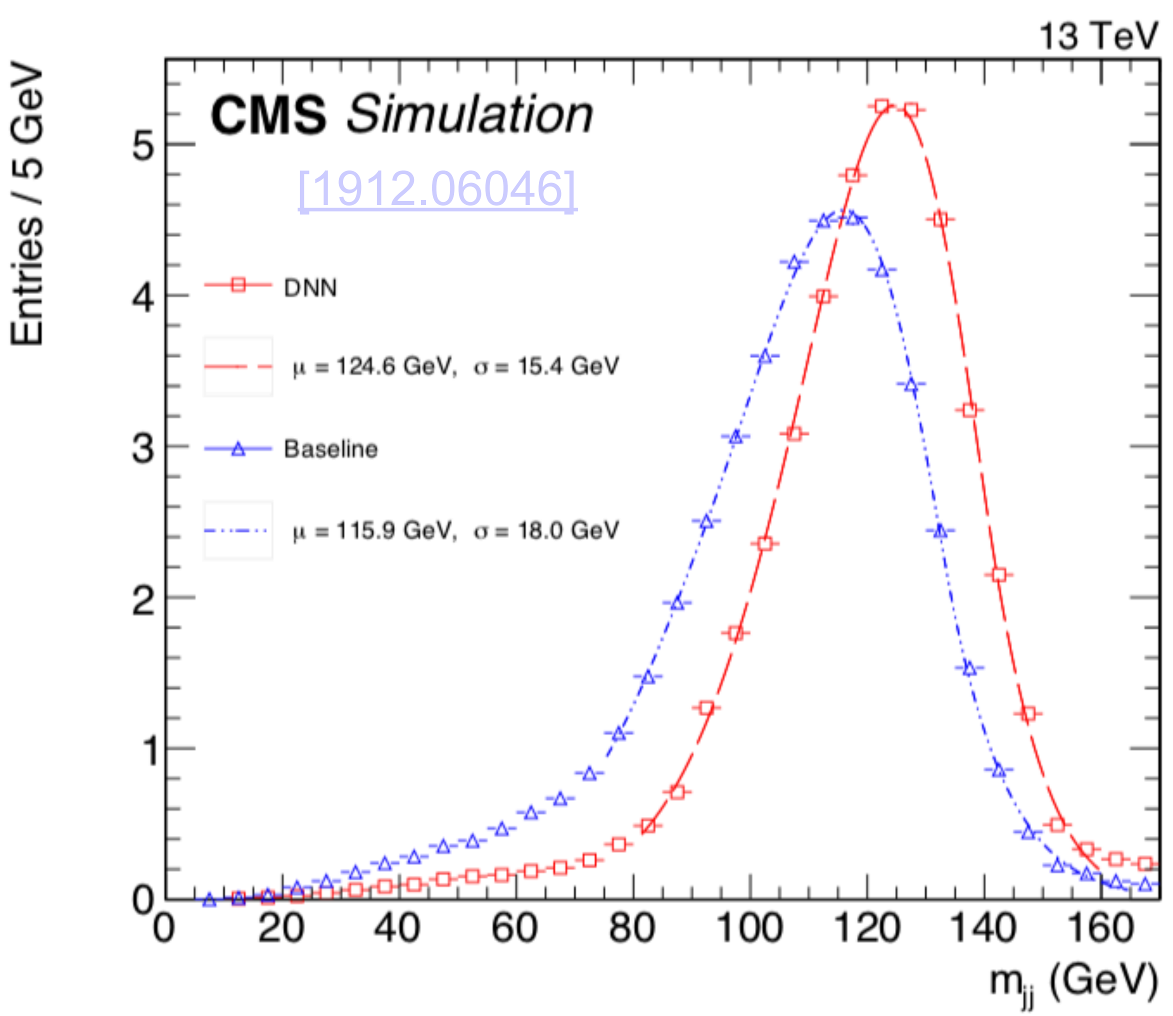
Per photon boosted decision tree based prediction of energy and resolution with semi-parametrized loss function.

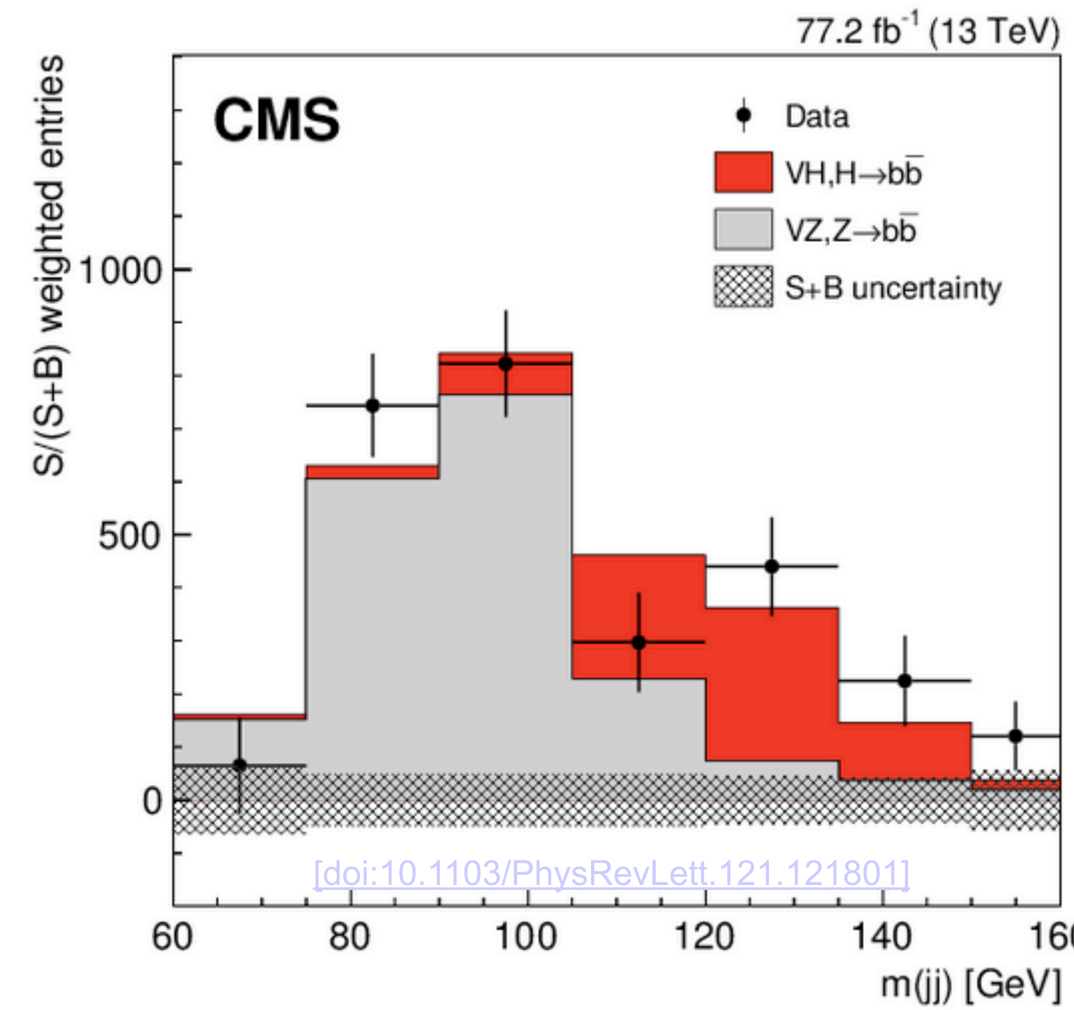




b-jet Energy & Resolution

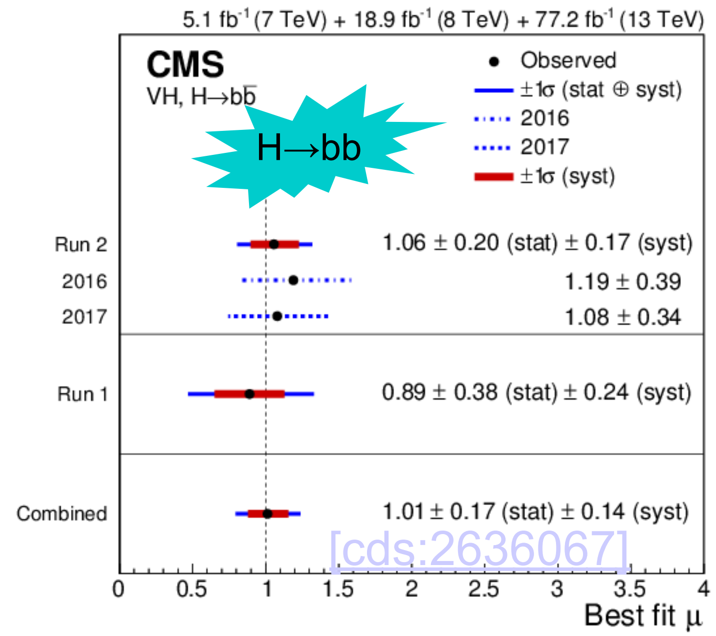
Fully connected jet-level features neural network predicts the jet energy correction and resolution using quantile regression .

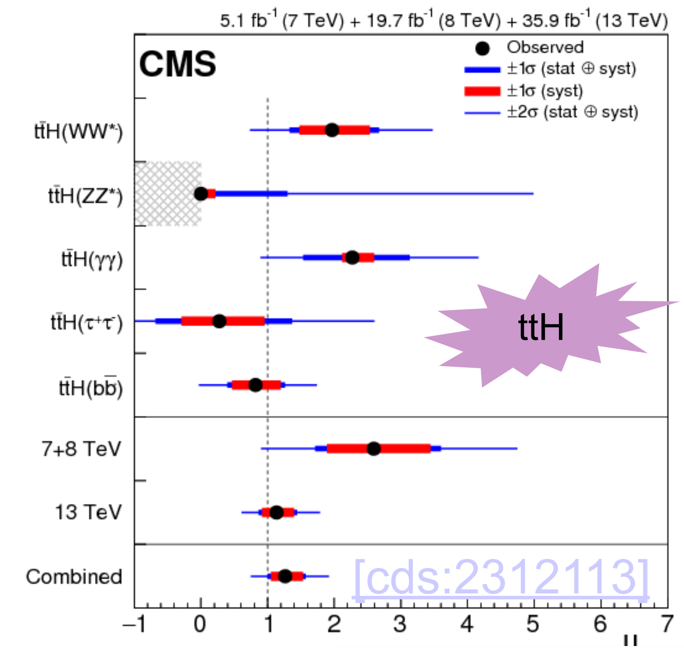


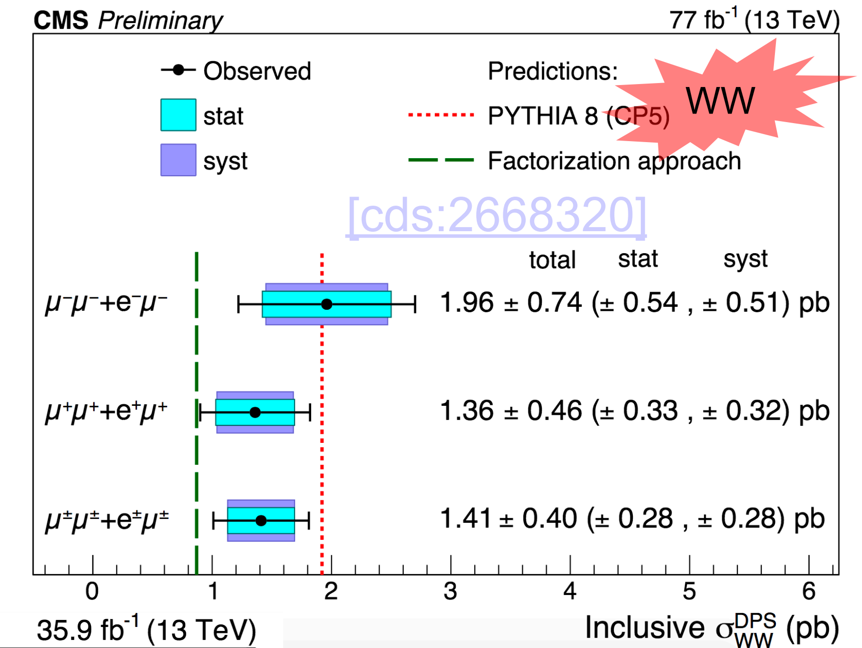


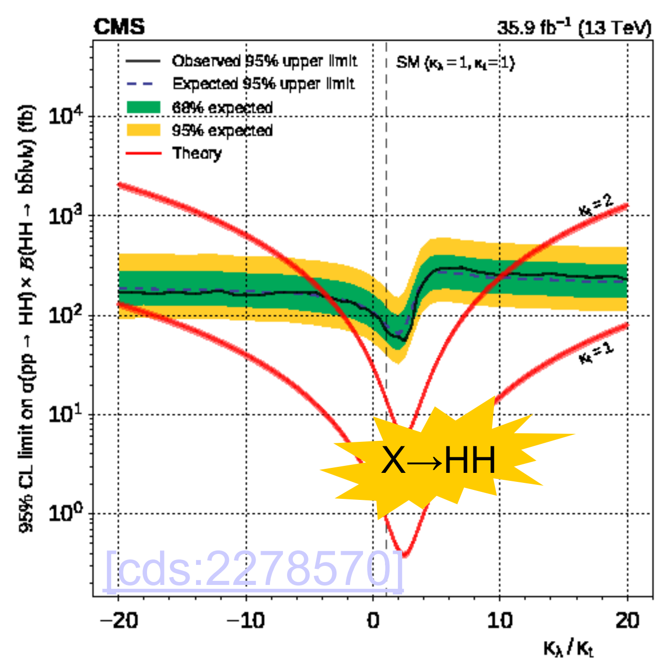
~20% improvement on Higgs mass resolution.

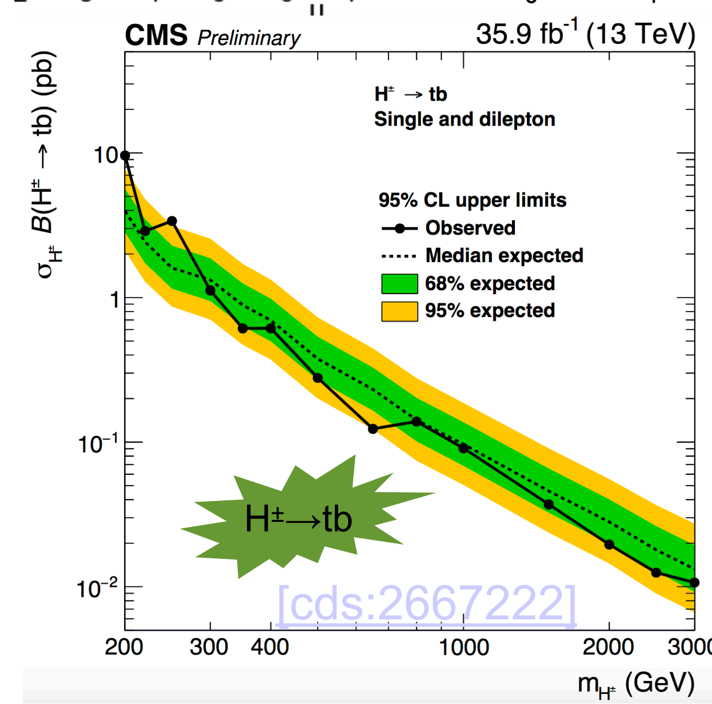
Increased Sensitivity





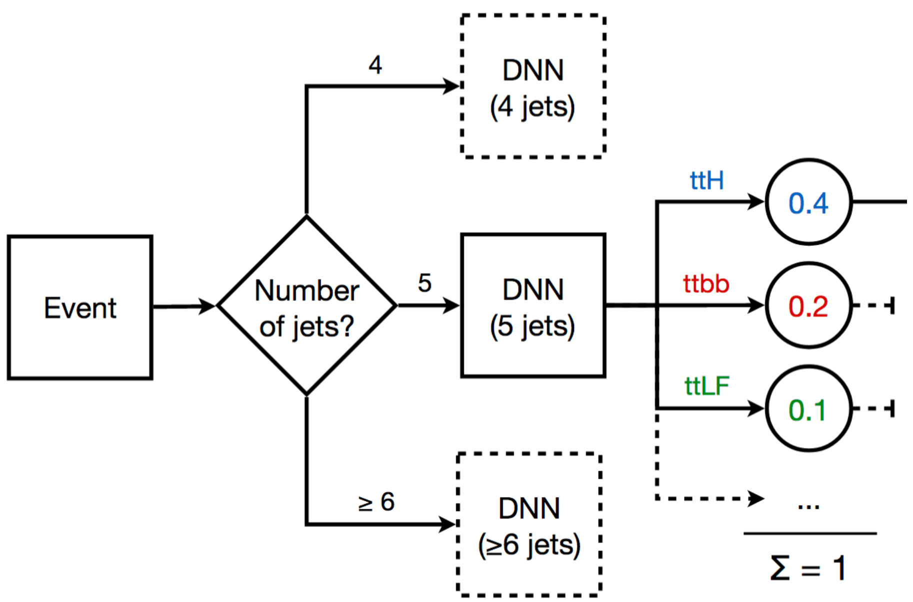




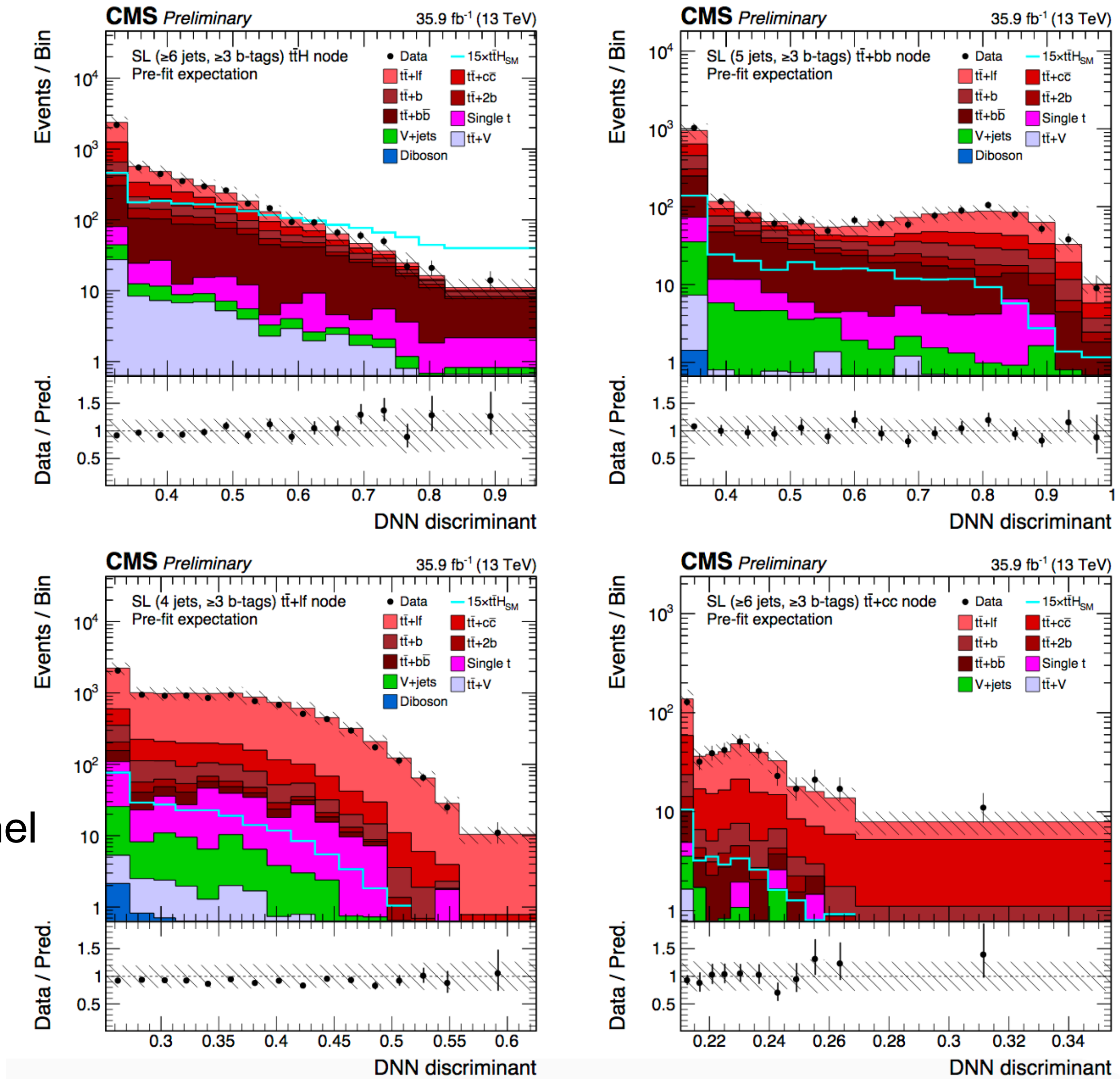


Increased sensitivity of analysis with BDT/NN signal extraction.
Would require more data otherwise.

Multi-category Classification



Slide M. Rieger

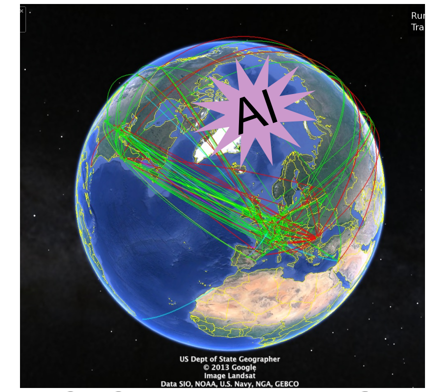


Search for ttH production in the H-to-bb decay channel with leptonic tt decays [cds:2308267]

Regular analysis fit categories sub-divided using DNN output nodes for added sensitivity.

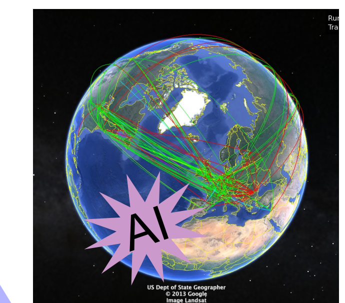
Deep Learning on the Edge of CMS

Role of AI: accelerator control, data acquisition, event triggering, anomaly detection, new physics scouting, event reconstruction, event generation, detector simulation, LHC grid control, analytics, signal extraction, likelihood free inference, background rejection, new physics searches, ...

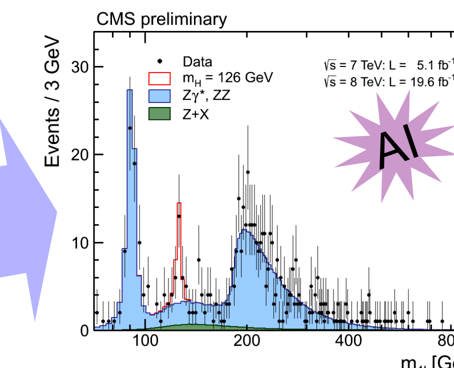


LHC Computing Grid
200k cores pledge to CMS over ~100 sites

CERN Tier-0/Tier-1 Tape Storage
200PB total

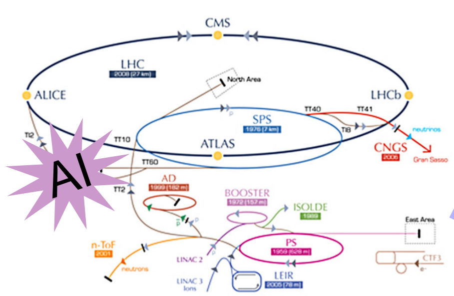


LHC Grid Remote Access to 100PB of data

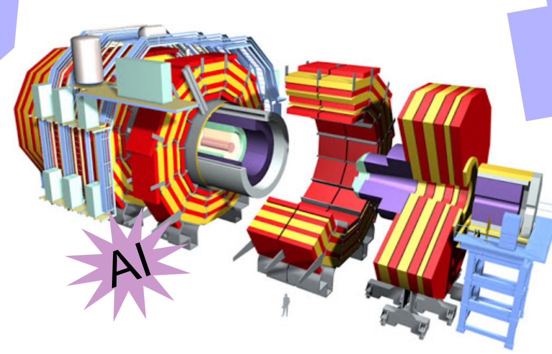


Rare Signal Measurement
~1 out of 10^6

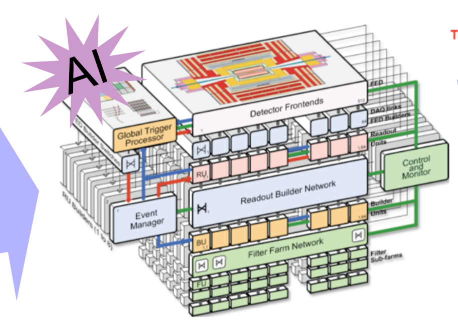
CERN Tier-0 Computing Center
20k cores



Large Hadron Collider
40 MHz of collision



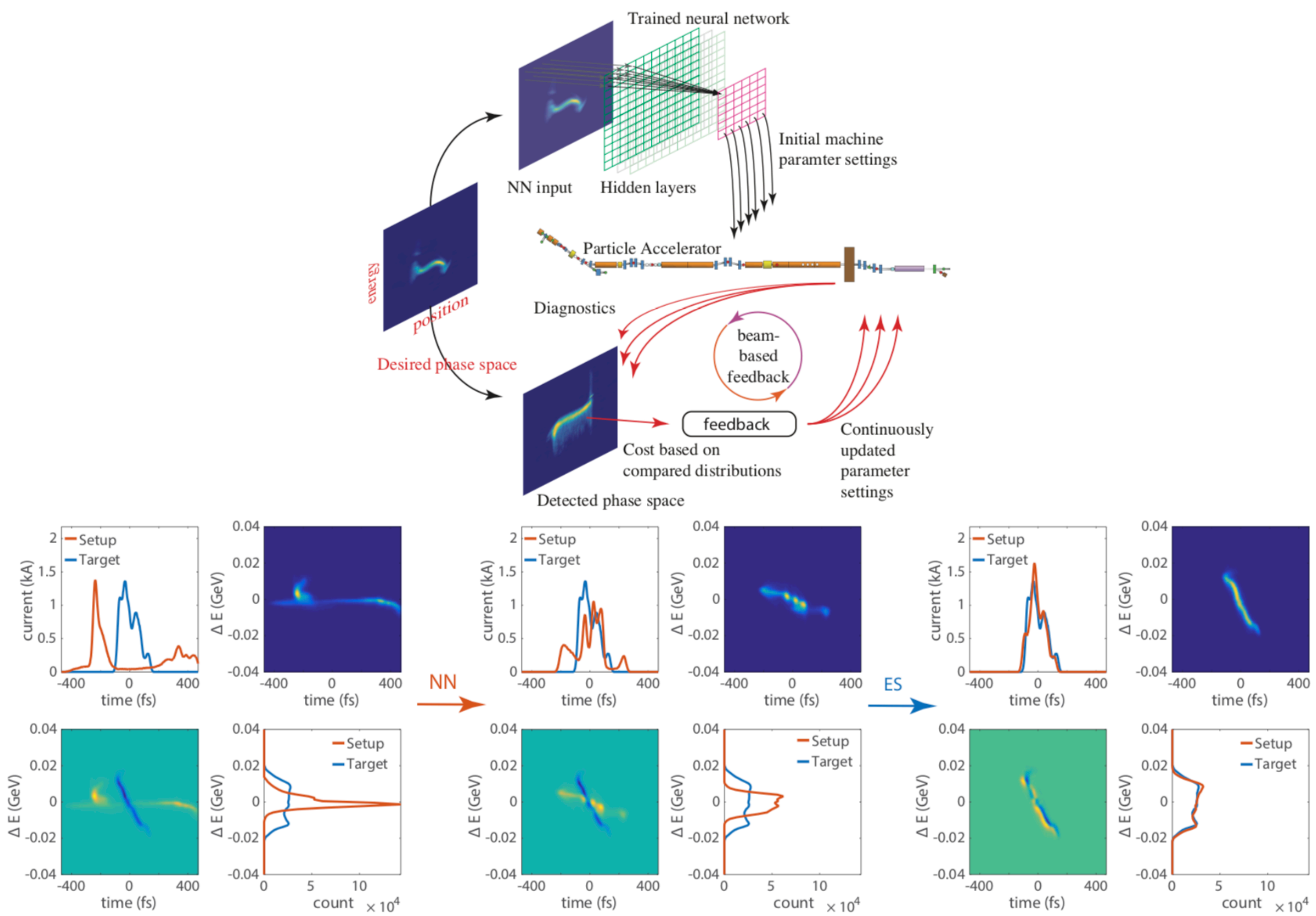
CMS Detector
1PB/s



CMS L1 & High-Level Triggers
50k cores, 1kHz

➡*Highlighting particular items next.*
➡*More in backup slides.*

Producing the Data



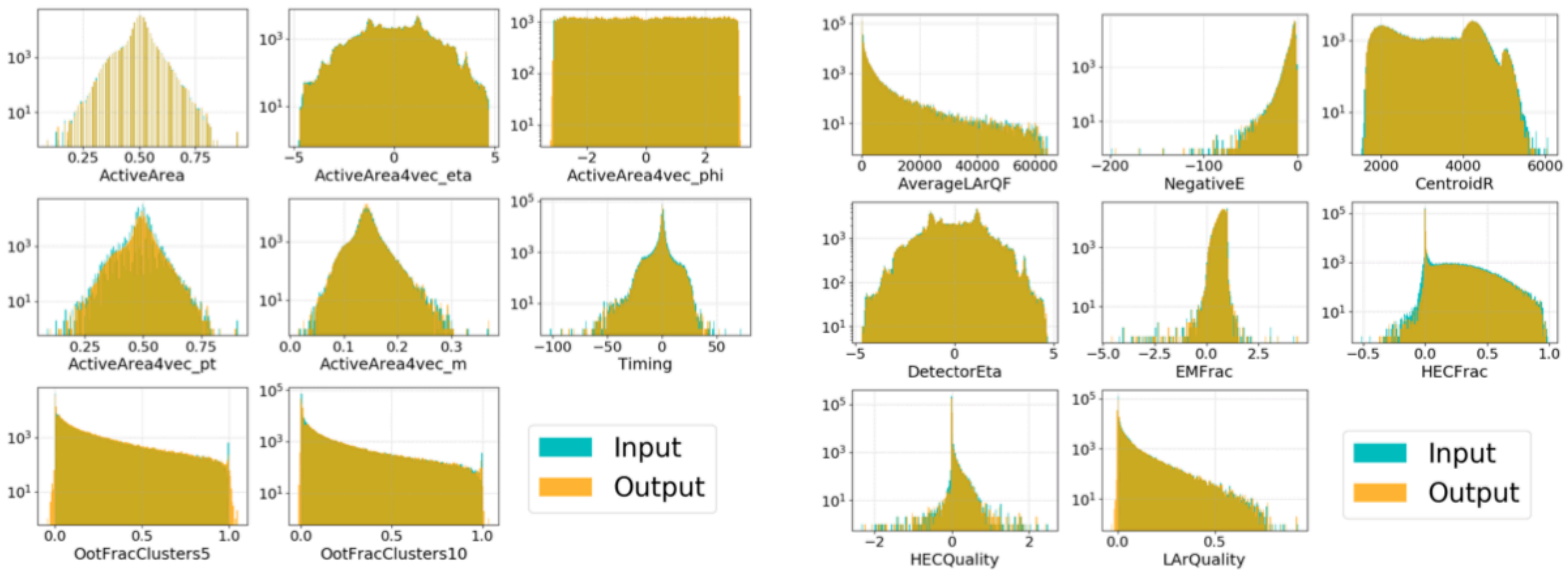
A. Scheinker, C. Emma, A.L. Edelen, S. Gessner
[2001.05461]

- Machine learning can be used to tune devices, control beams, perform analysis on accelerator parameters, etc.
- Already successfully deployed on accelerator facilities.
- More promising R&D to increase beam time.
- Potential for detector control ?

Opportunities in Machine Learning for Particle Accelerators [1811.03172]
Machine learning for design optimization of storage ring nonlinear dynamics [1910.14220]
Advanced Control Methods for Particle Accelerators (ACM4PA) 2019 Workshop Report [2001.05461]
Machine learning for beam dynamics studies at the CERN Large Hadron Collider [2009.08109]
…

Advertising: FrontiersIn Research Topics on Operational Intelligence

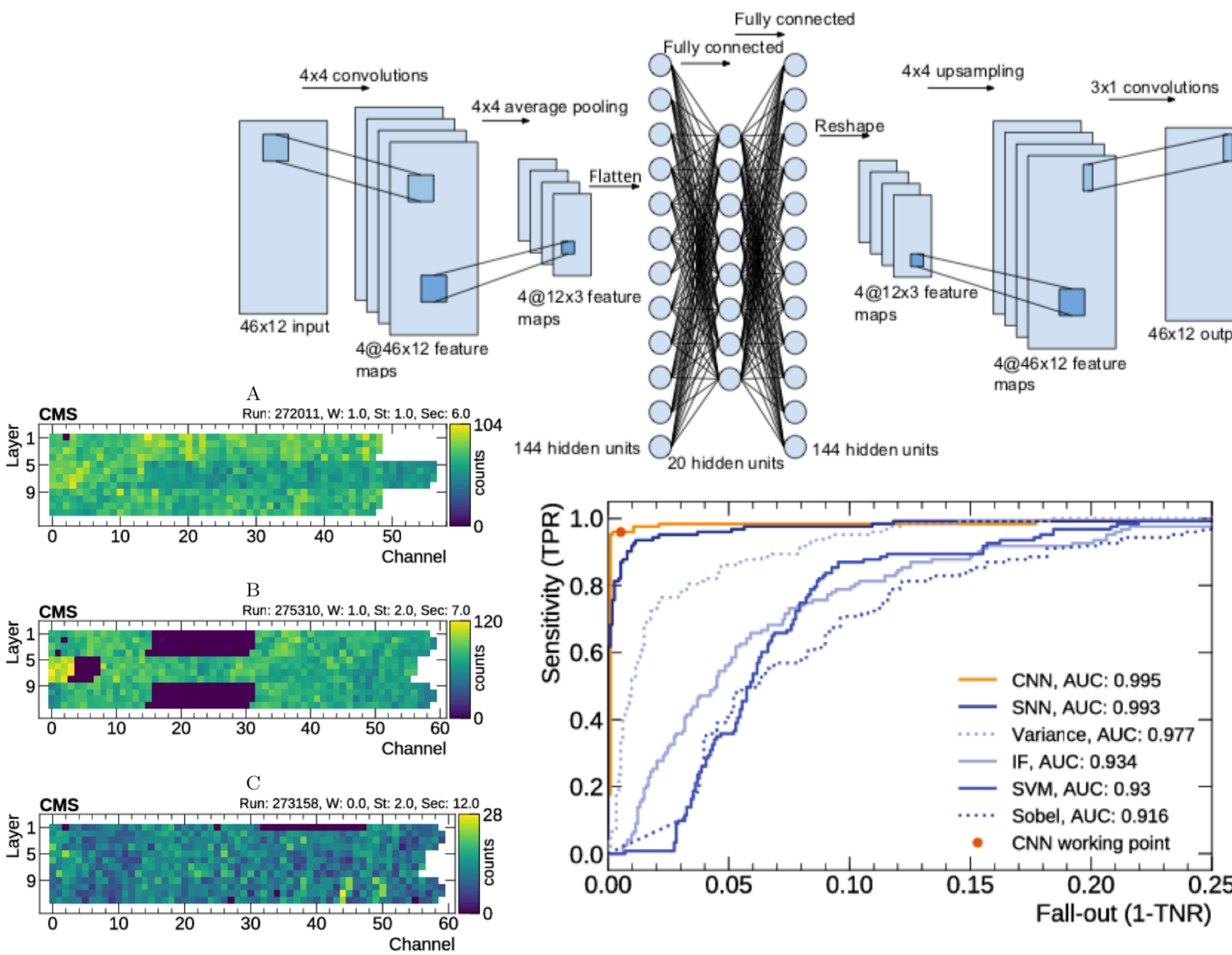
Compressing Data

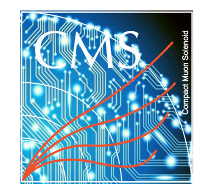


Deep Auto-Encoders for compression in HEP
http://lup.lub.lu.se/student-papers/record/9004751

- Rich literature on data compression of image with neural network.
- Make use of abstract semantic space for image compression.
- Image compression can suffer some loss of resolution.
- Saving on disk/tape cost. Potential in scouting strategies.
- R&D needed to reach the necessary level of fidelity.

Cleaning Data



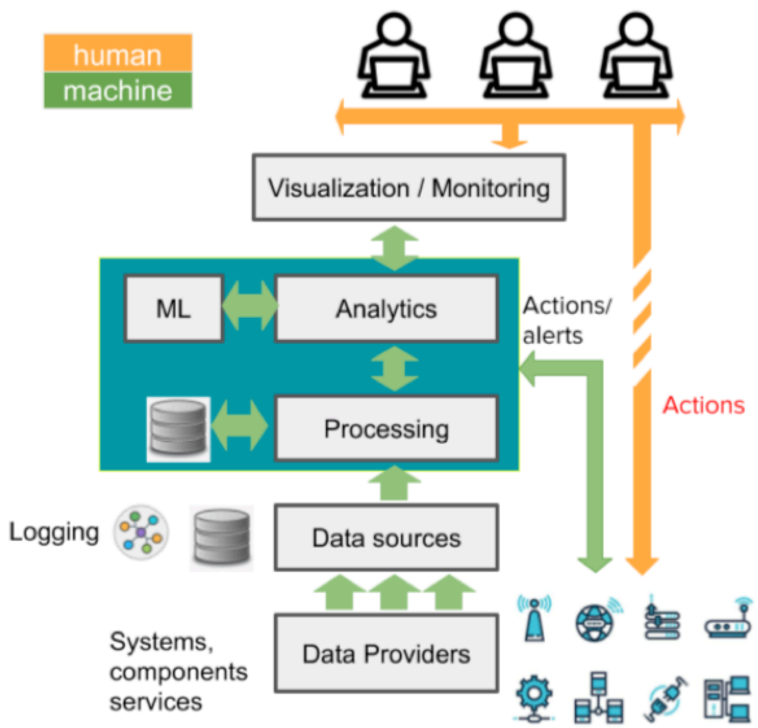


- Data quality is a person power intensive task, and crucial for swift delivery of Physics
- Machine learning can help with automation.
- Learning from operators, reducing workload.
- Continued R&D and experiment adoption.

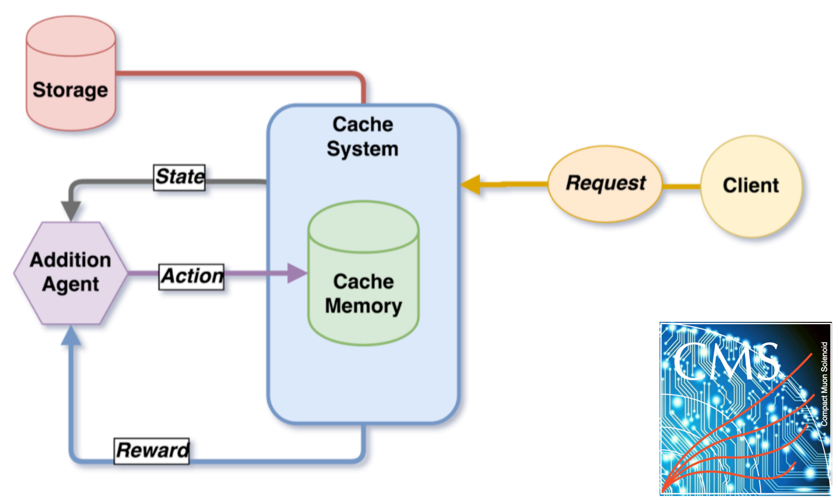
A.A. Pol, G. Cerminara, C. Germain, M. Pierini, A. Seth
[doi:10.1007/s41781-018-0020-1]

Towards automation of data quality system for CERN CMS experiment [doi:10.1088/1742-6596/898/9/092041]
LHCb data quality monitoring [doi:10.1088/1742-6596/898/9/092027]
Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider [1808.00911]
Anomaly detection using Deep Autoencoders for the assessment of the quality of the data acquired by the CMS experiment [doi:10.1051/epjconf/201921406008]
…

Managing Data



Operational Intelligence
[cds:2709338]



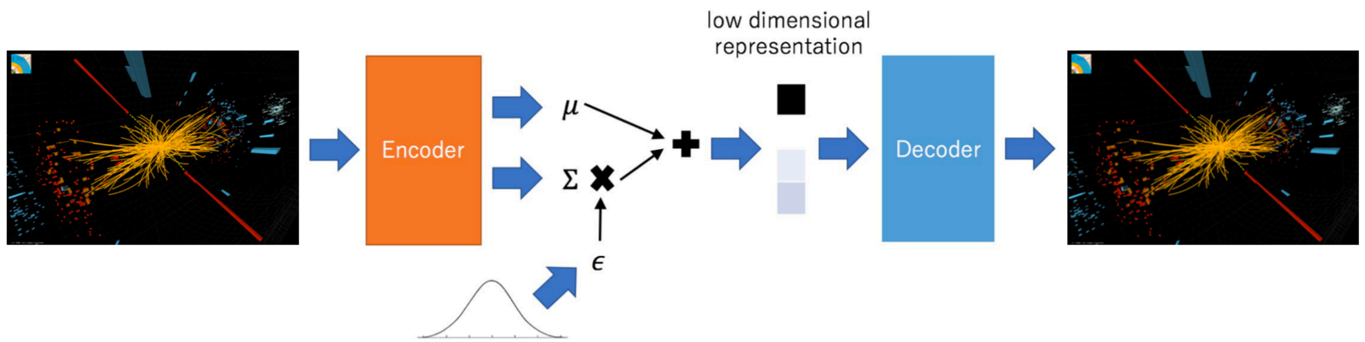
Cache Type	Throughput	Cost	Read on hit ratio	Band sat.	CPU Eff.
SCDL	**79.43%**	**50.68%**	21.22%	58.94%	58.75%
LFU	65.01%	104.73%	**33.29%**	**51.00%**	**60.92%**
Size Big	49.02%	111.73%	28.55%	54.40%	60.41%
LRU	47.15%	112.84%	27.64%	54.93%	59.90%
Size Small	46.71%	113.01%	27.39%	55.01%	59.73%

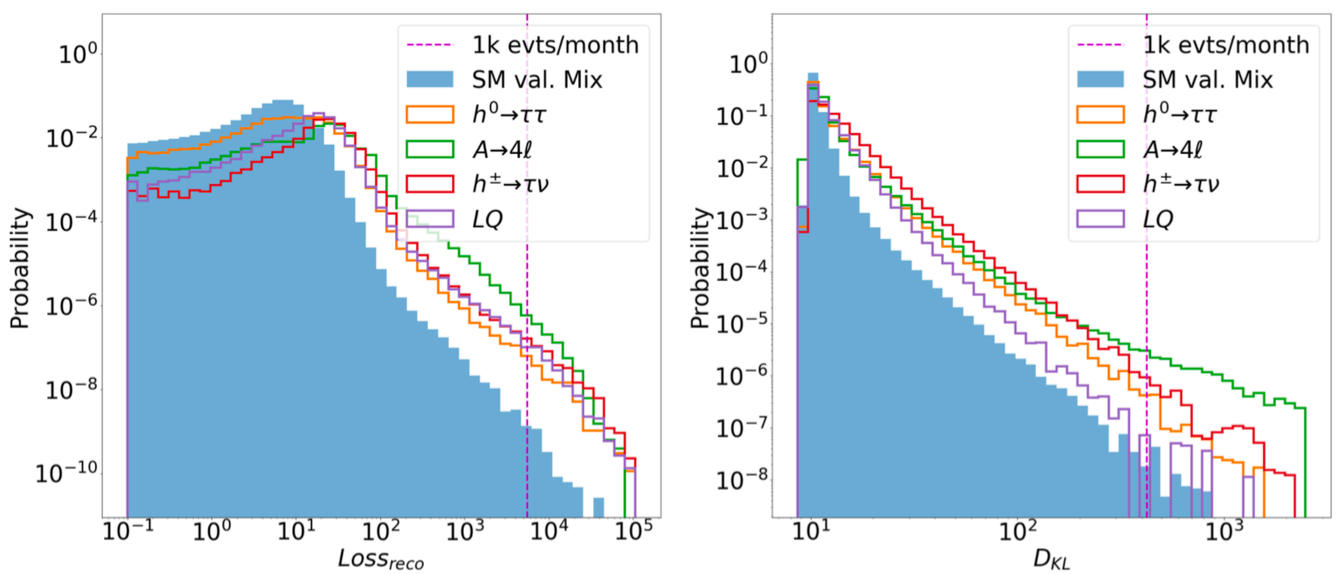
Caching suggestions using Reinforcement Learning
LOD 2020, in proceedings

- The LHC-grid is key to success of the LHC experiments.
- Complex ecosystem with dedicated operation teams.
- Person power demanding, and inefficient in some corner of the phase space.
- Potential for AI-aided operation.
- Lots of modeling and control challenges.
- R&D to increase operation efficiency.

Advertising: FrontiersIn Research Topics on Operational Intelligence

Detecting New Data



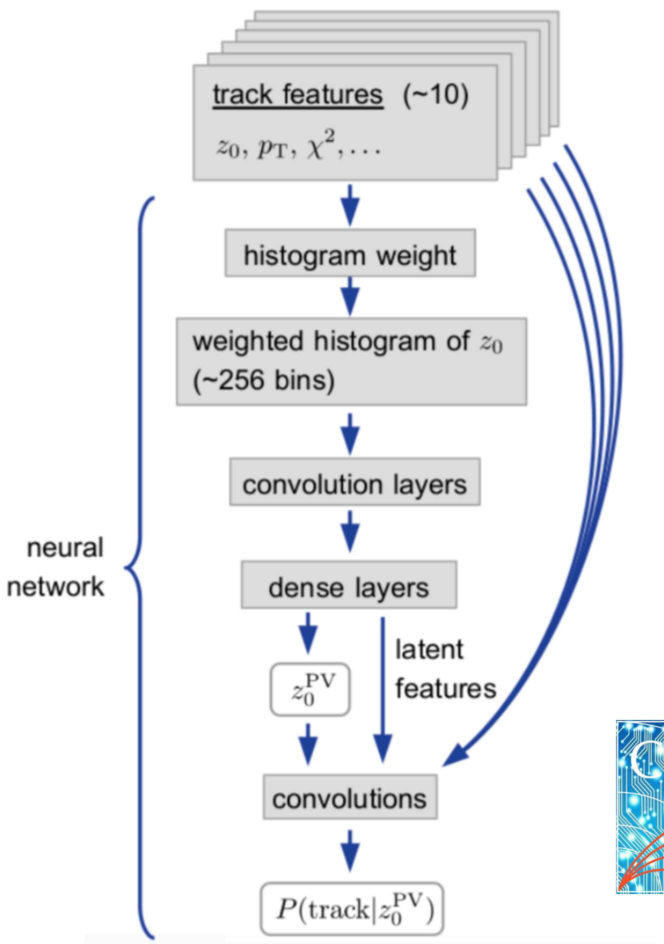


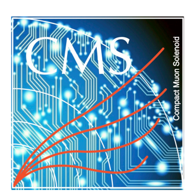
Use of variational auto-encoders directly on data to marginalize outlier events, for anomalous event hotline operation.
[doi:0.1007/JHEP05(2019)036]

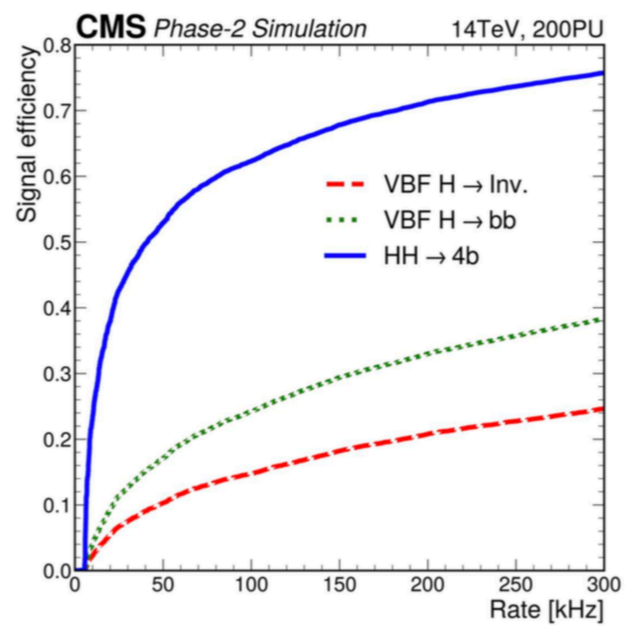
- Machine learning since long deployed in the trigger for selected signatures.
- Further potential for background trigger rate reduction.
- Emerging opportunity for triggering on unknown signatures : “a la Hotline”.
- More promising R&D and experiment adoption.

Triggering and Scouting

Vertex reconstruction at L1





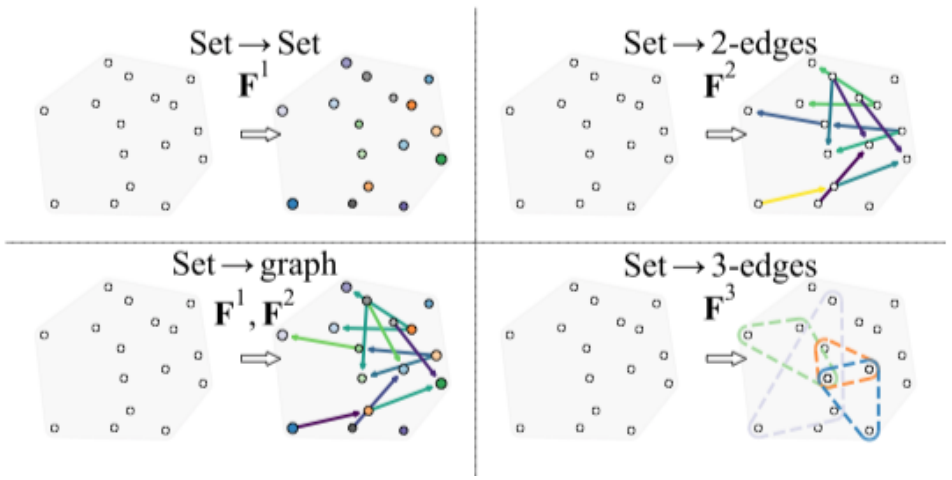


Anomaly detection at L1

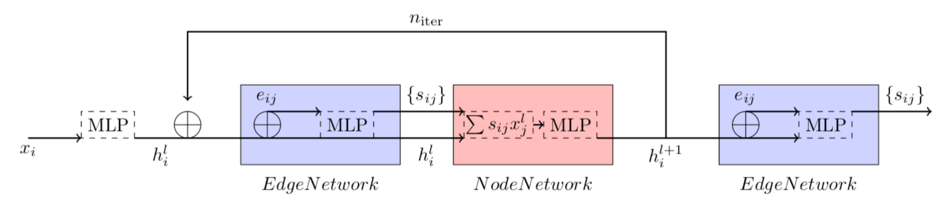
Phase-2 upgrade of the CMS L1-Trigger
[cds:2714892]

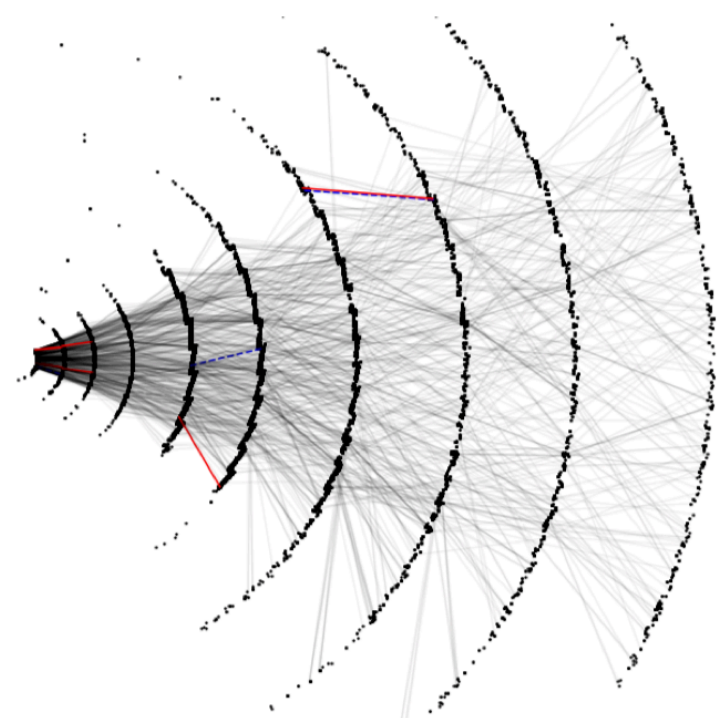
- Trigger benefit from fast reconstruction algorithms
- L1 needs FPGA implementation. hls4ml-enabled algorithms.
- Quality of selection increases with refinement of object reconstruction
- Having the best reconstruction is particularly important in scouting
- Balance between speed and accuracy

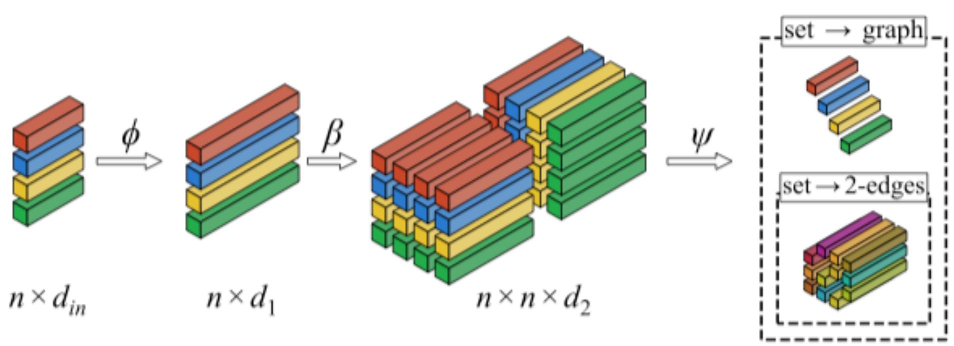
Reconstructing Data



GNN applied to charged particle tracking
[2007.00149]







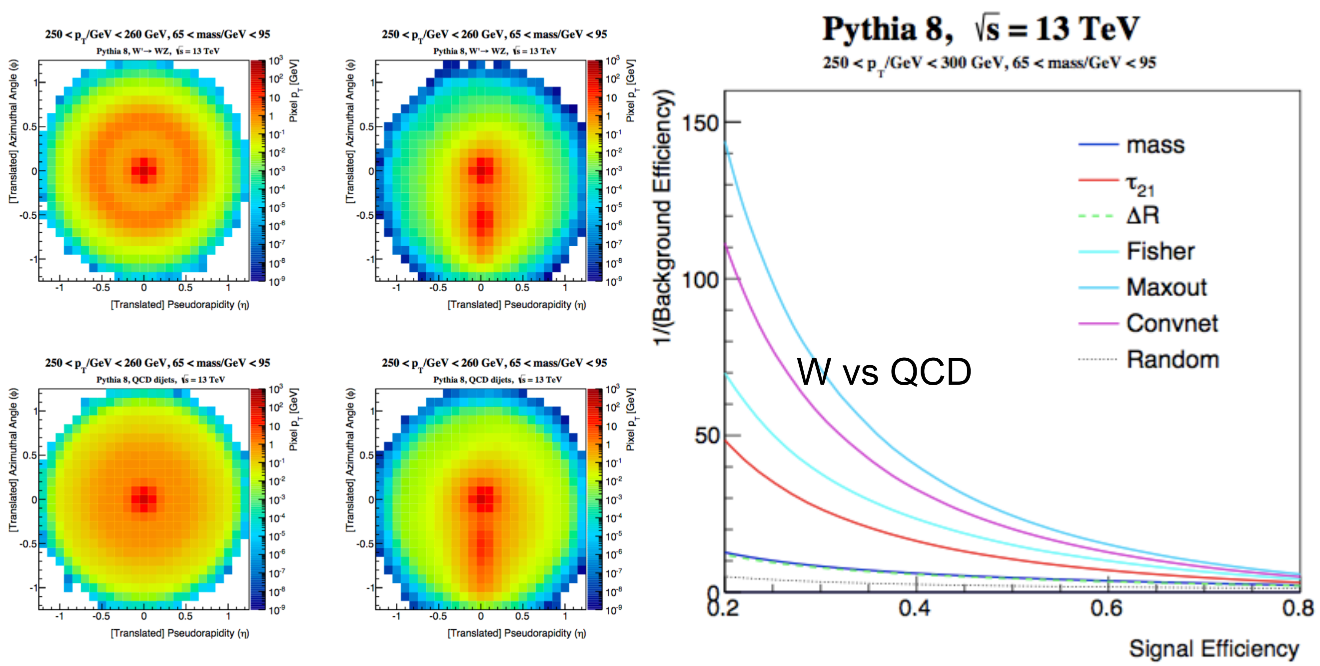
Learning graphs from sets, applied to vertexing
[2002.08772]

- Event reconstruction is pattern recognition to a large extend. Advanced machine learning techniques can help.
- Learn from the simulation, and/ or data.
- Learn from existing “slow reconstruction” or simulation ground truth.
- Automatically adapt algorithm to new detector design.
- Image base methods evolving towards graph-based methods.
- Accelerating R&D to exploit full potential.

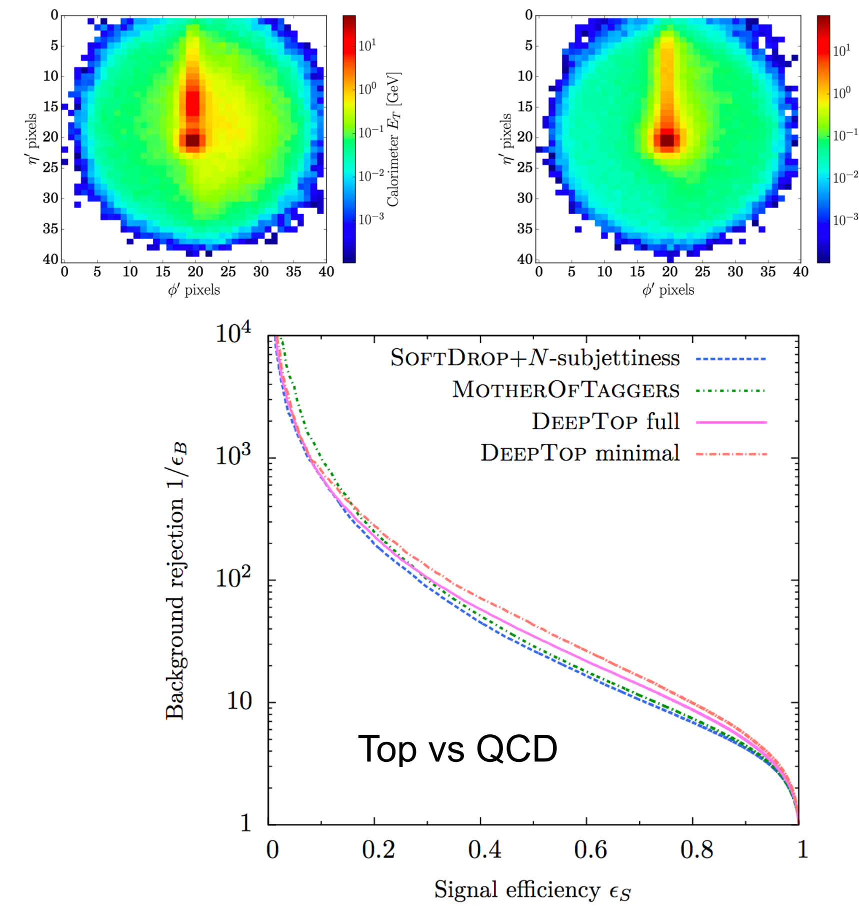
More of the relevant works at:
https://iml-wg.github.io/HEPML-LivingReview/

Image Representation

Jet-Images – Deep learning edition
[1511.05190]

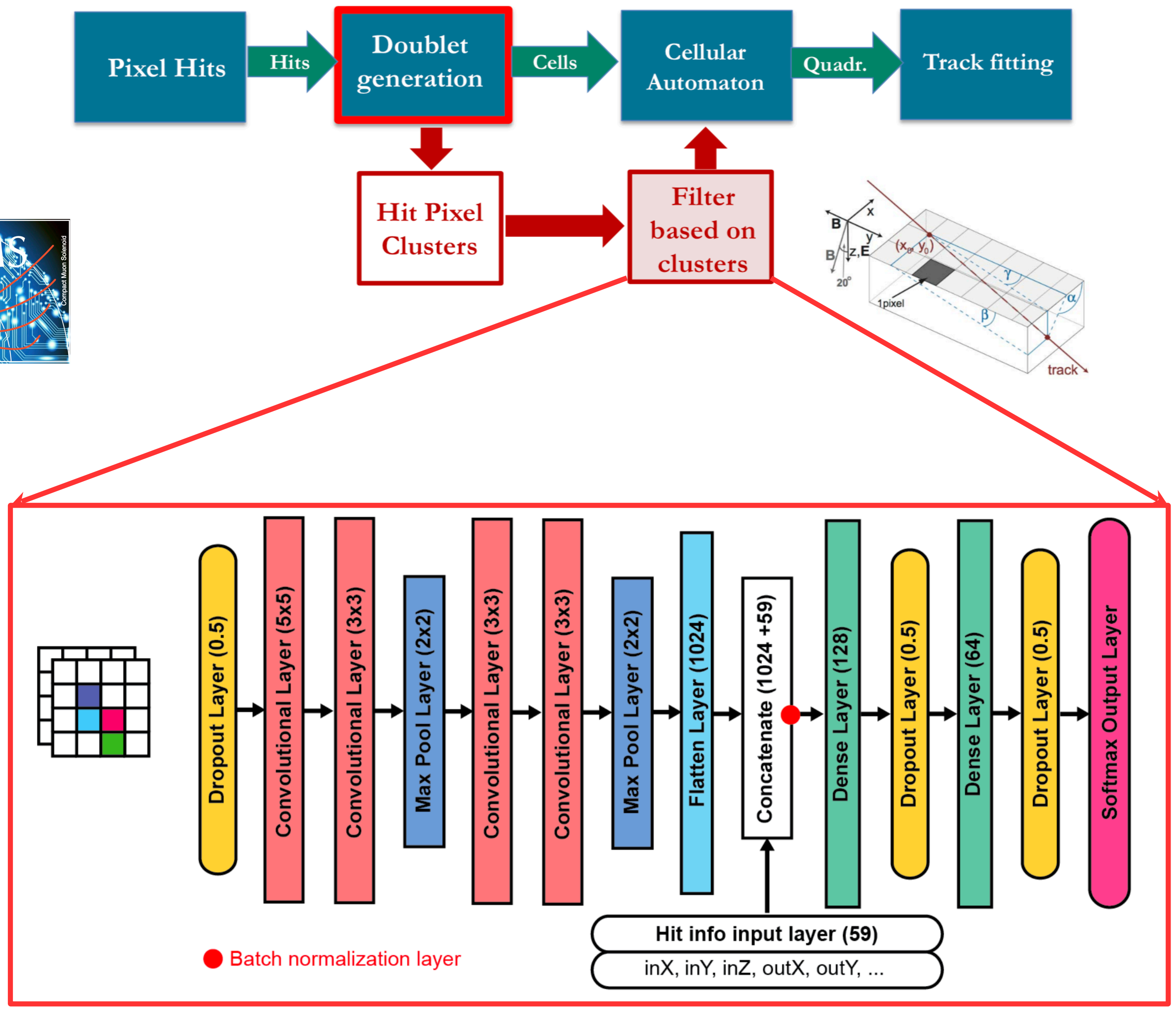


Deep-learning top taggers or the end of QCD?
[1701.08784]



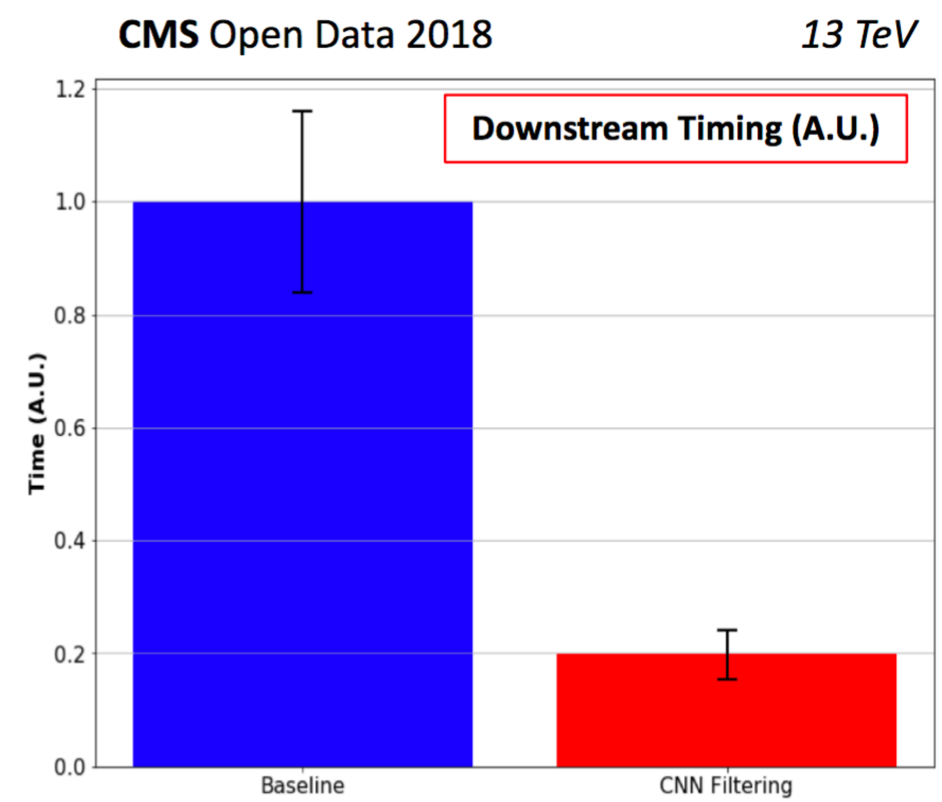
Calorimeter signal are image-like.
Projection of reconstructed particle properties onto images possible.
Potential loss of information during projection.

Seed Cleaning

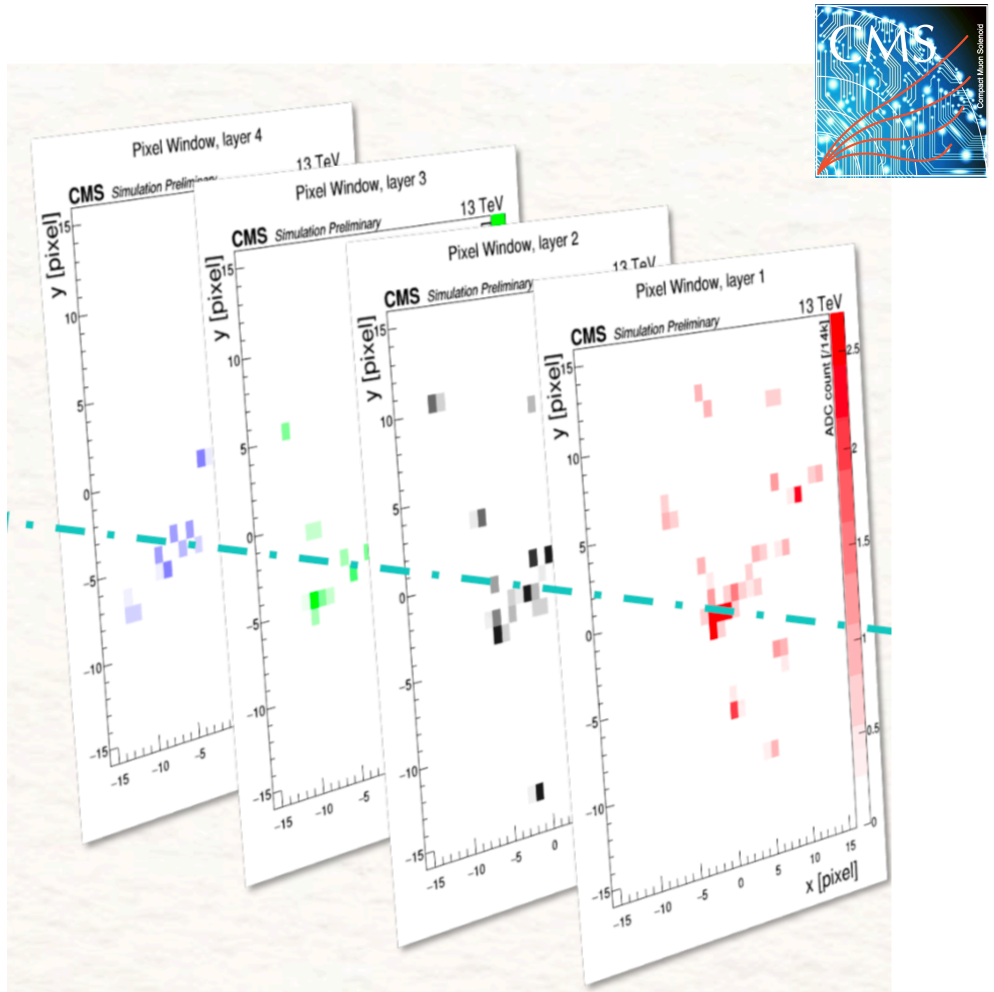


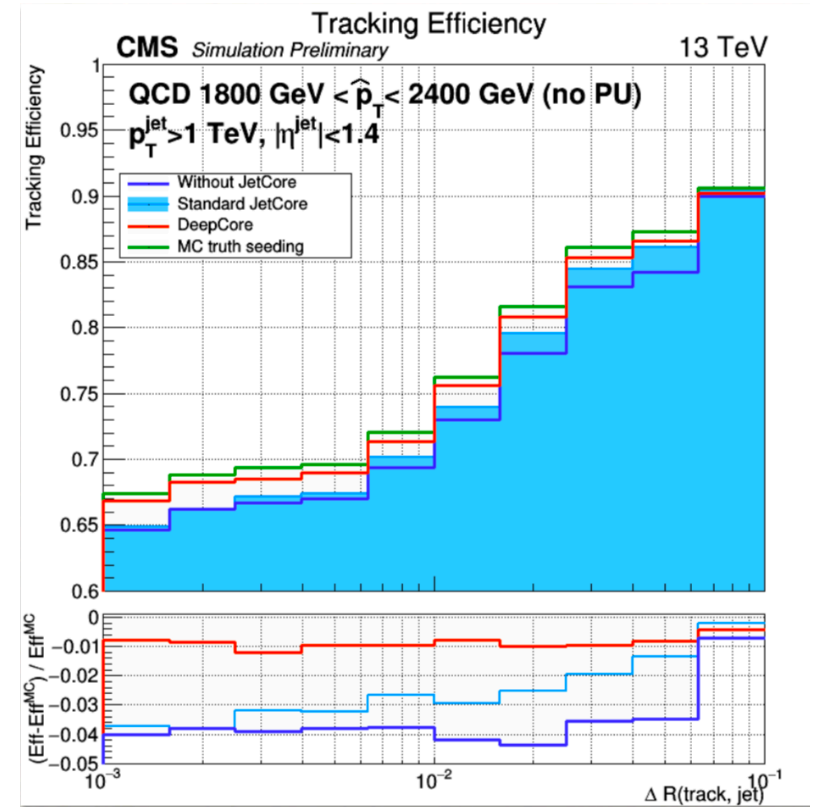
https://indico.cern.ch/event/742793/contributions/3298727

- Categorization of hits doublet using the pixel cluster shapes as input
- Significantly reduce timing in pattern recognition



Seed Finding in Jets

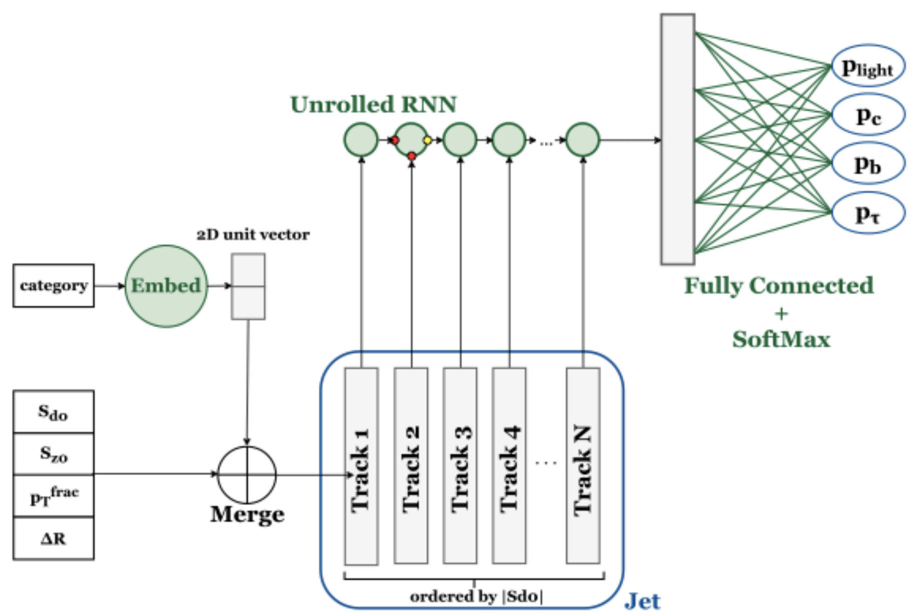




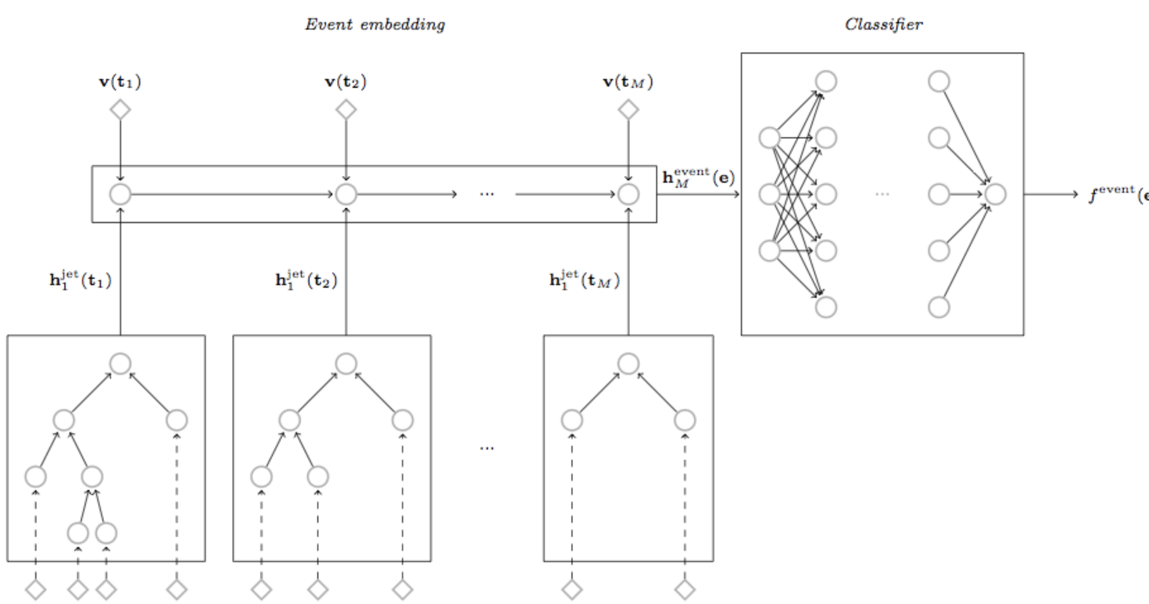
https://indico.cern.ch/event/742793/contributions/3274301/

- Predict tracklets parameters from raw pixels using CNN
- Approaching the maximum reachable performance

Sequence Representation



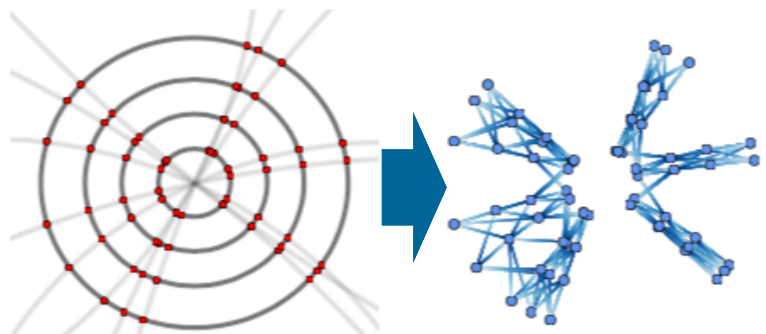
B-Jet with Recurrent Neural Networks
[cds:2255226]



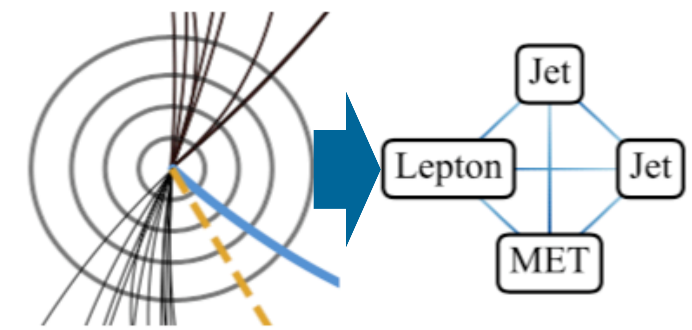
QCD-Aware Recursive Neural Networks for Jet Physics.
[1702.00748]

Somehow arbitrary choice on ordering with sequence representation.
Physics-inspired ordering as inductive bias.
Ordering can be learned too somehow.

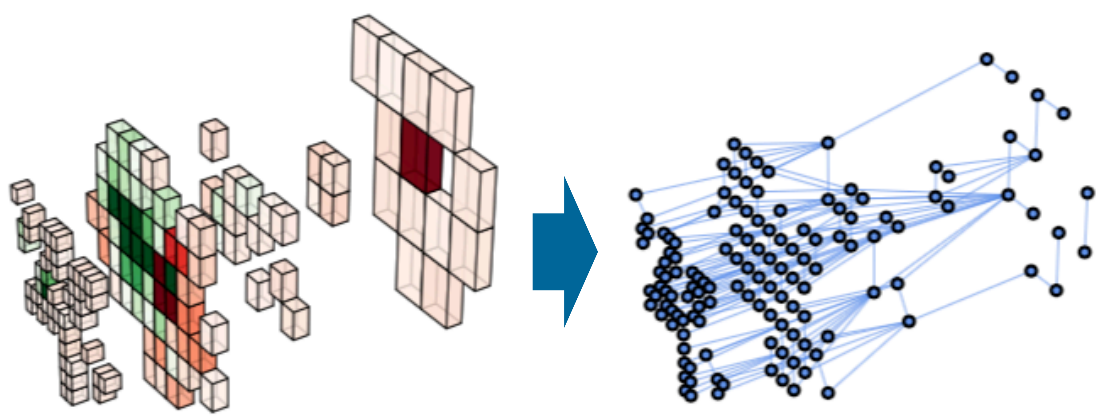
Graph Representation



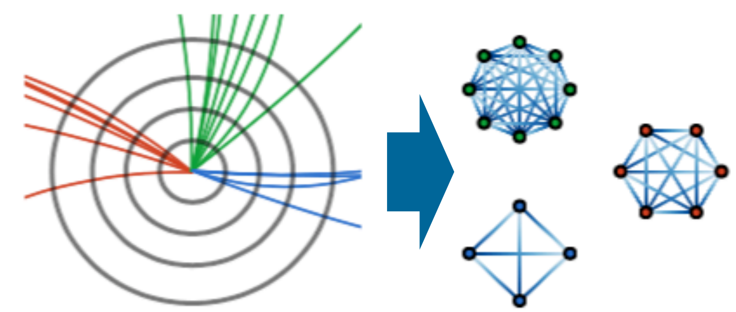
Hits in tracking detector



Objects in an event



Hits in calorimeter detector

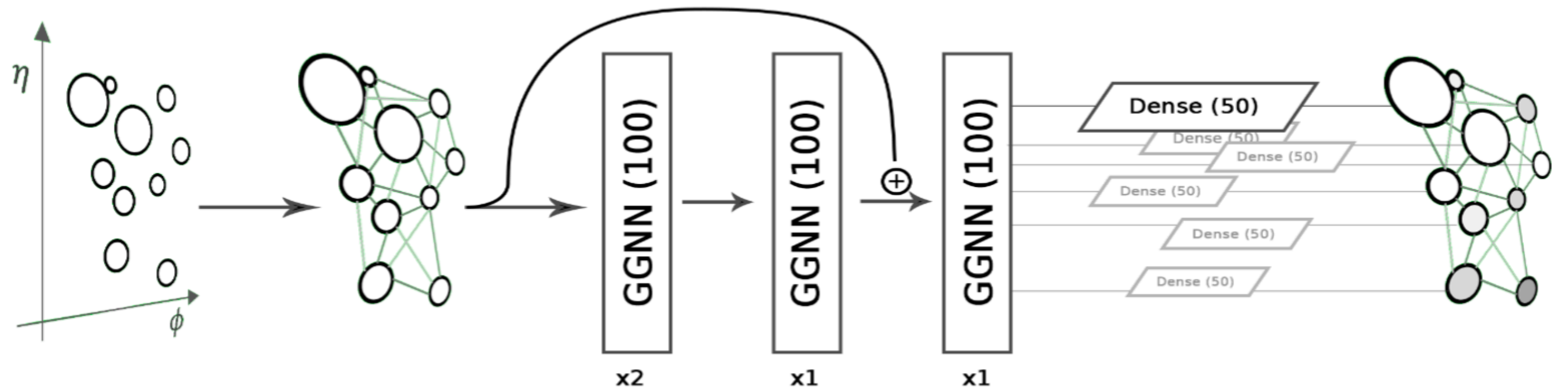


Object sub-structure in an event

Graph Neural Networks in Particle Physics
[2007.13681]

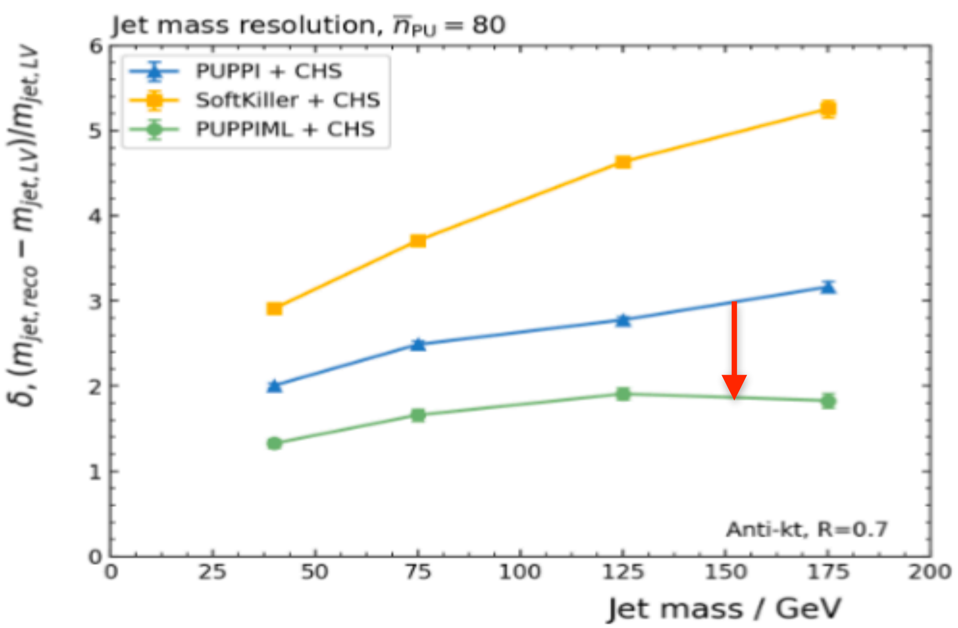
Heterogenous data fits well in graph/set representation.

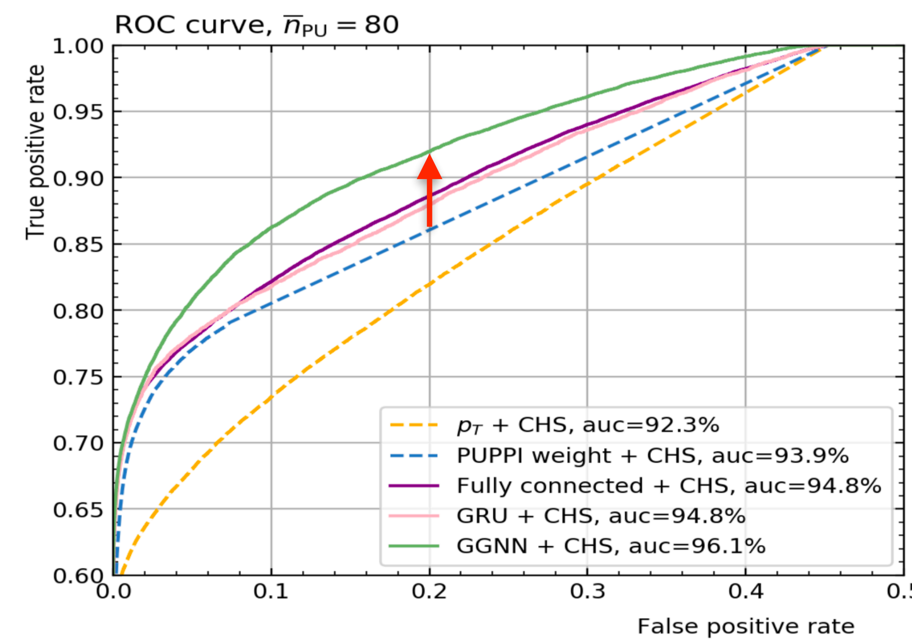
Pile-Up Mitigation



Pileup mitigation at the Large Hadron Collider with Graph Neural Networks [1810.07988]

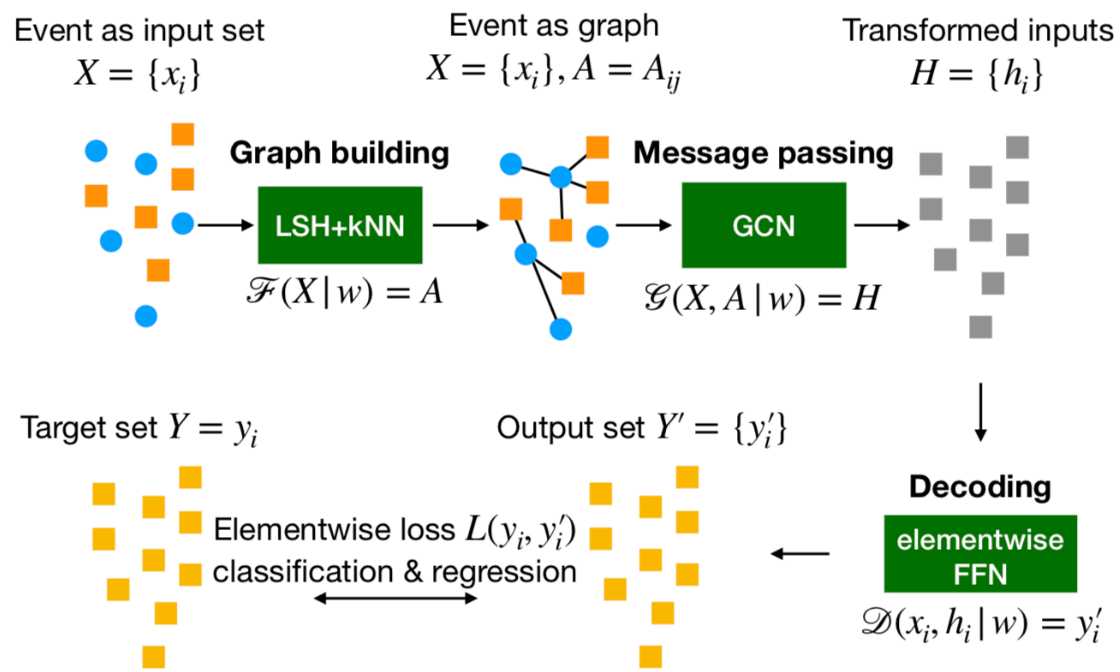
- Locally connected graph of reconstructed particle flow candidates
- Gated graph neural network (GGNN) to evolve node representations
- per-particle pile-up classification extract for neutrals



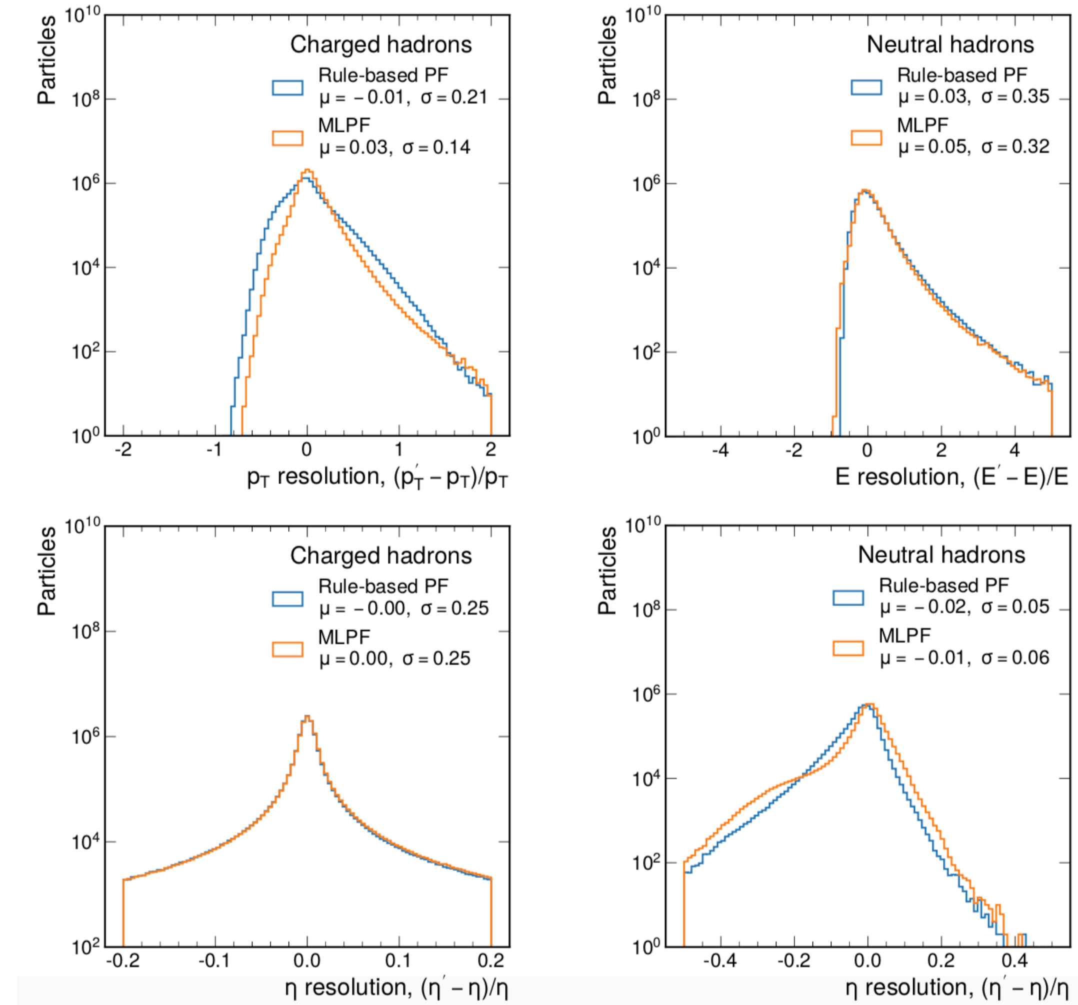


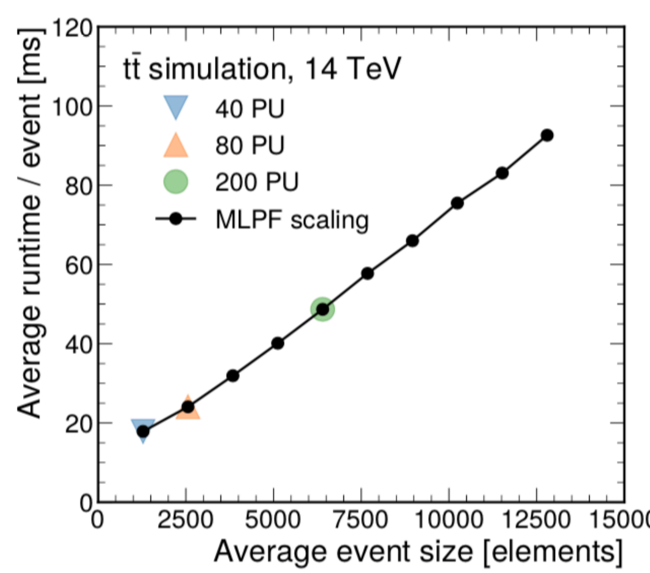
Particle Flow Reconstruction

MLPF: Efficient Machine-Learned Particle-Flow Reconstruction Using Graph Neural Networks [2101.08578]



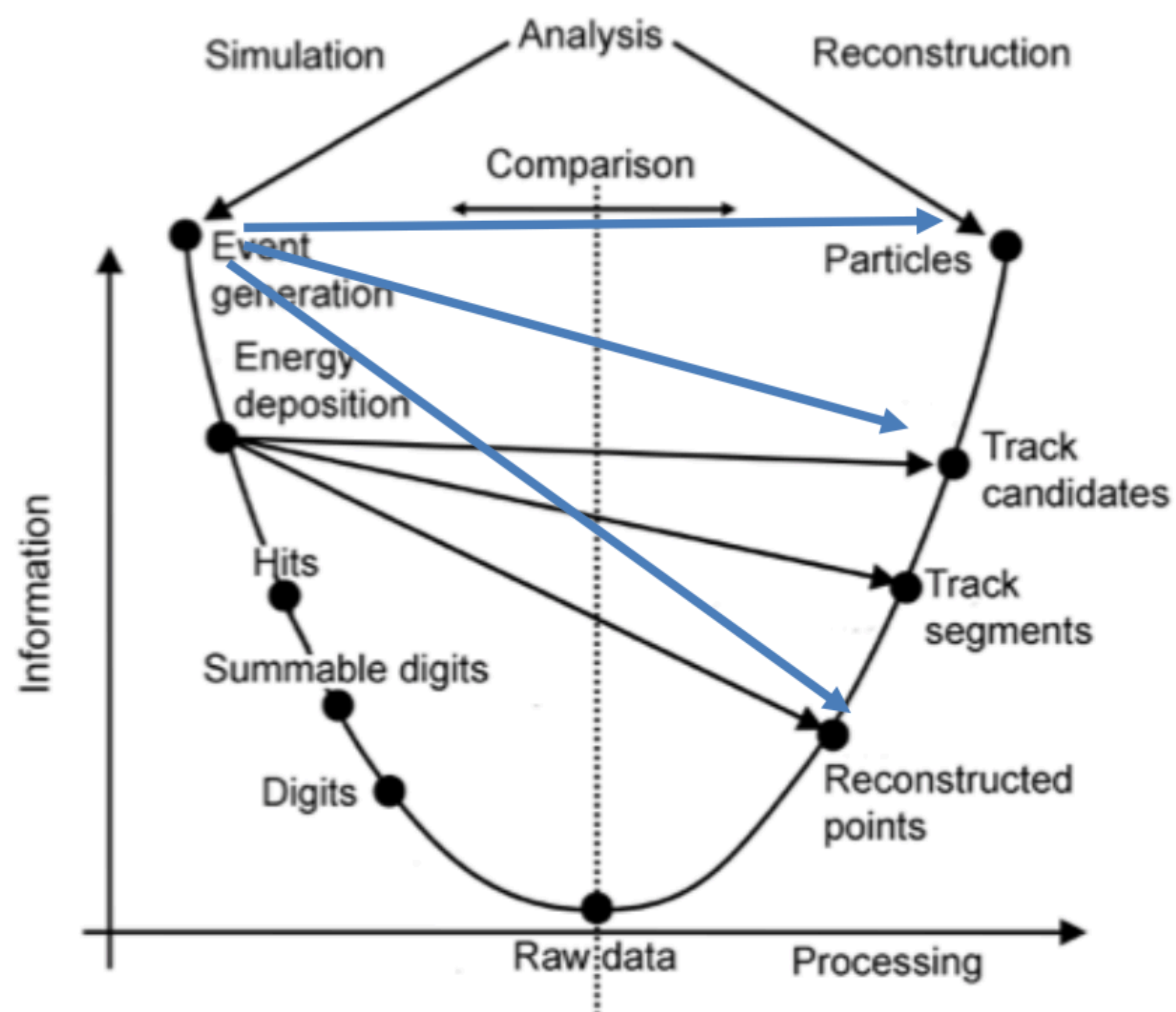
$$x_i = [\text{type}, p_T, E_{ECAL}, E_{HCAL}, \eta, \phi, \eta_{outer}, \phi_{outer}, q, \ldots], \text{type} \in \{\text{track}, \text{cluster}\}$$
$$y_i = [\text{PID}, p_T, E, \eta, \phi, q, \ldots], \text{PID} \in \{\text{none}, \text{charged hadron}, \text{neutral hadron}, \gamma, e^{\pm}, \mu^{\pm}\}$$
$$h_i \in \mathbb{R}^N, N = 256$$
Trainable neural networks: $\mathcal{F}, \mathcal{G}, \mathcal{D}$





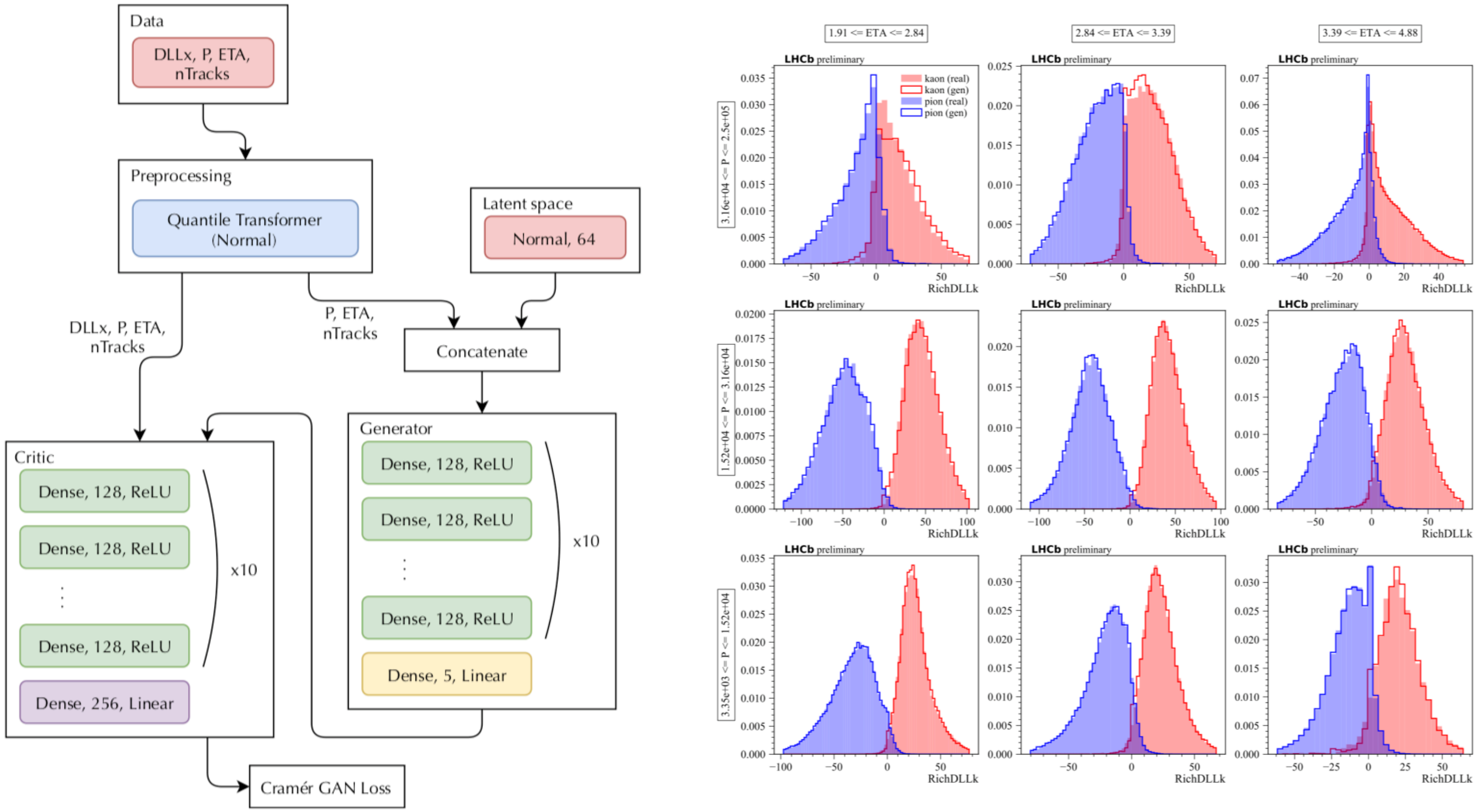
- Set of tracks & clusters in input
- Classify sub-set of graph nodes
- Regress parton kinematics
- Execution time linear with PU

Reconstruction ∘ Simulation ~ Identity



Simulation aims at predicting the outcome of collisions.
Reconstruction aims at inverting it.
Multiple ways to connect intermediate steps with deep learning.

Simulating Data

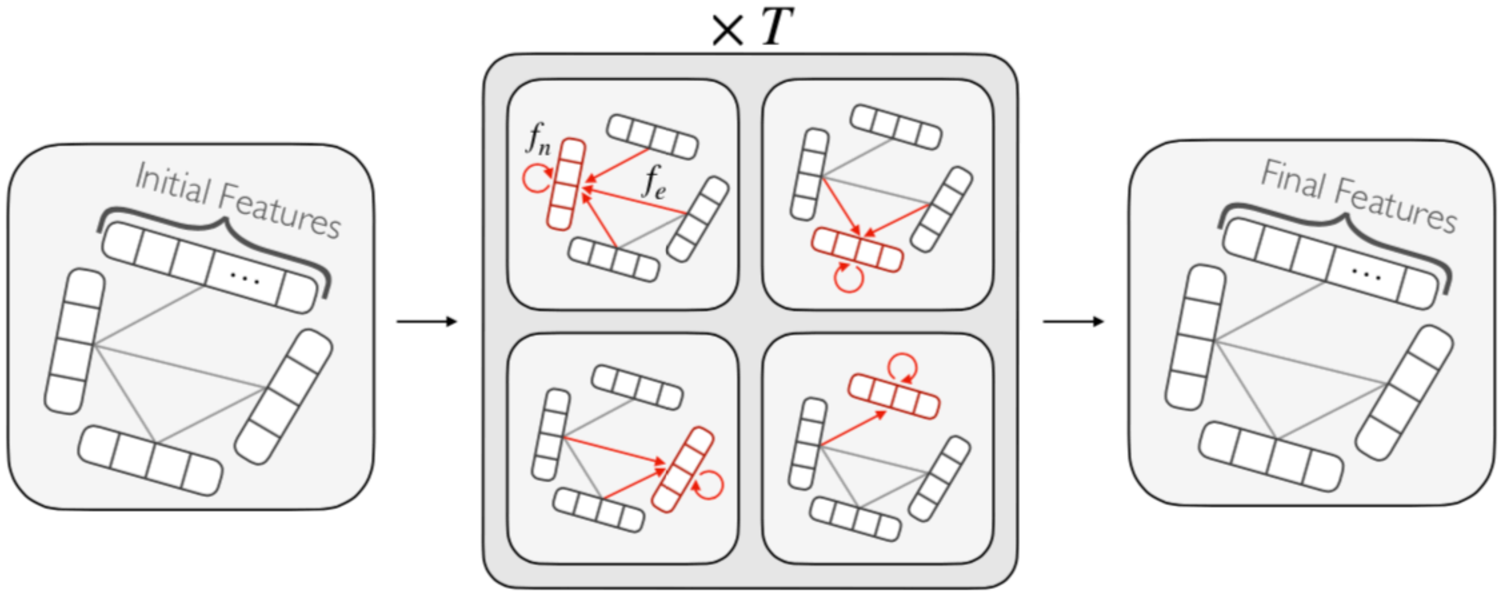


Generative Adversarial Networks for LHCb Fast Simulation [2003.09762]

- Fully detailed simulation is computing intensive.
- Fast and approximate simulators already in operation.
- Applicable at many levels : sampling, generator, detector model, analysis variable, etc
- Generative models can provide multiple 1000x speed-up.
- Careful study of statistical power of learned models over training samples.
- Many R&D, experiment adoption starting.

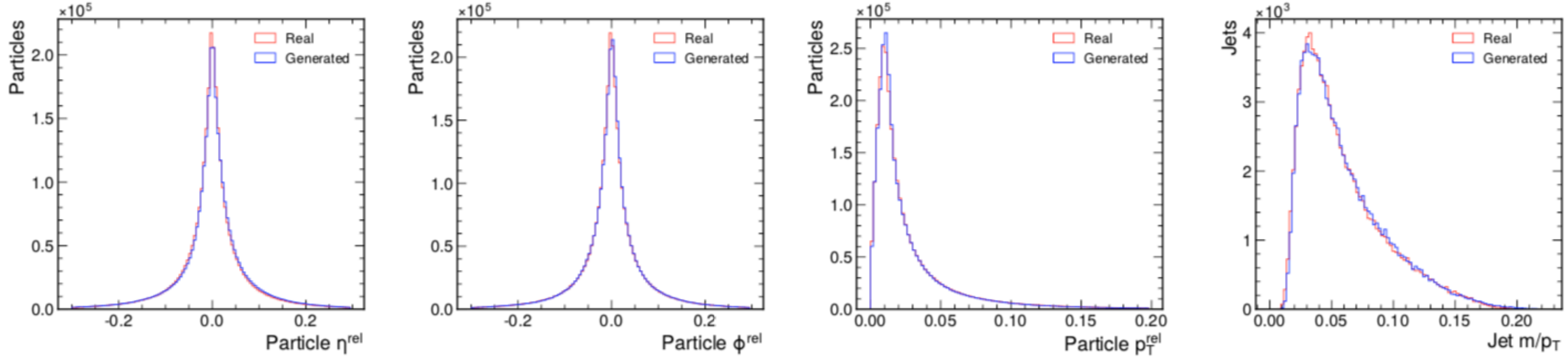
More of the relevant works at:
https://iml-wg.github.io/HEPML-LivingReview/

Graph Generative Models



- Events represented as a graph of particles
- Generator and discriminator networks as message passing graph neural networks
- Predicting particle kinematics

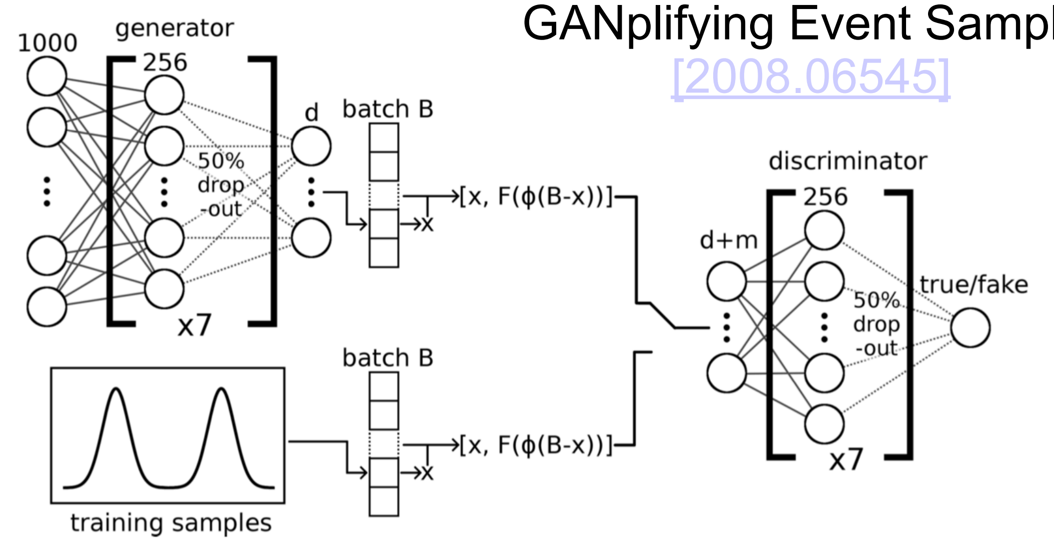
Graph Generative Adversarial Networks for Sparse Data Generation in High Energy Physics [2012.00173]

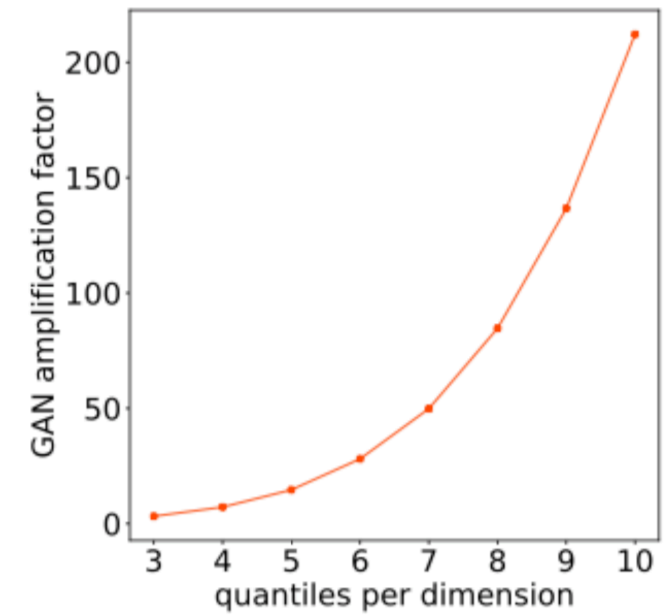


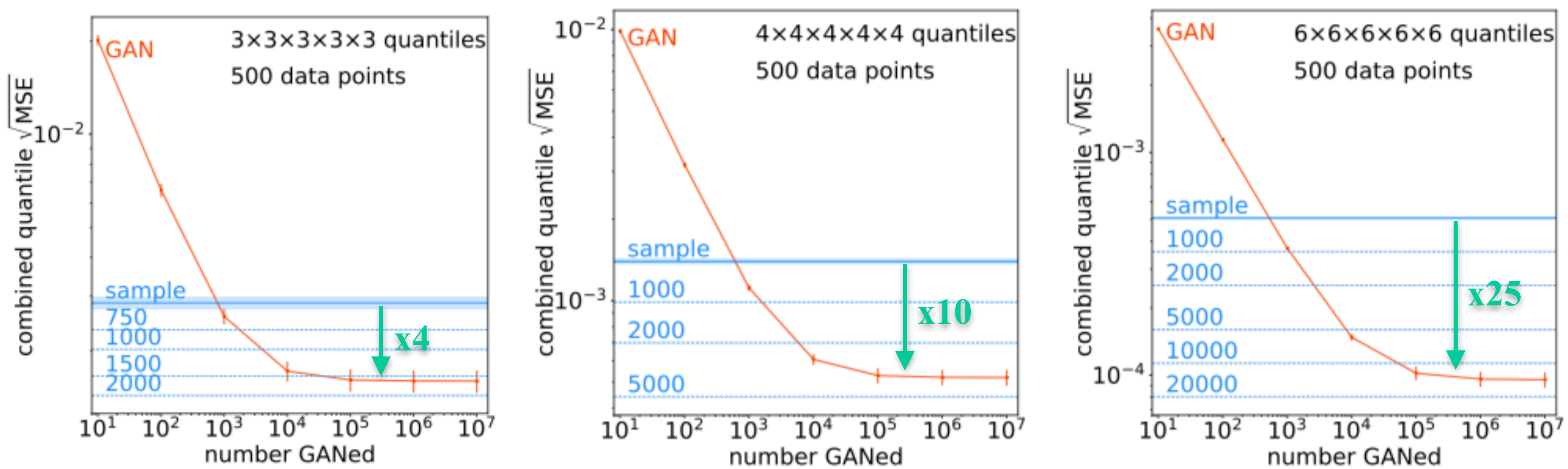
AI in HEP, LPC Topic of the Week, J-R Vlimant

Statistical Power

GANplifying Event Sample
[2008.06545]



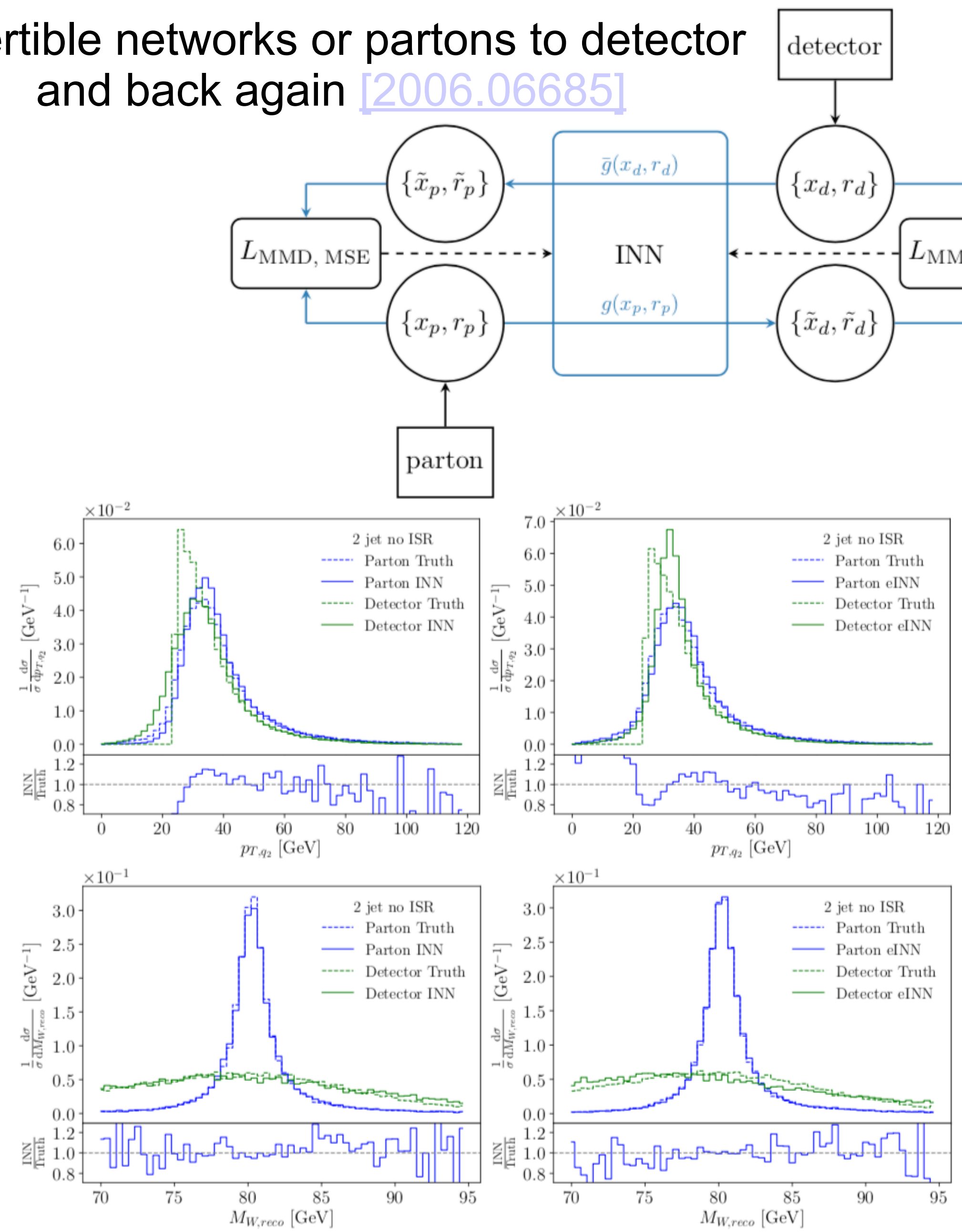


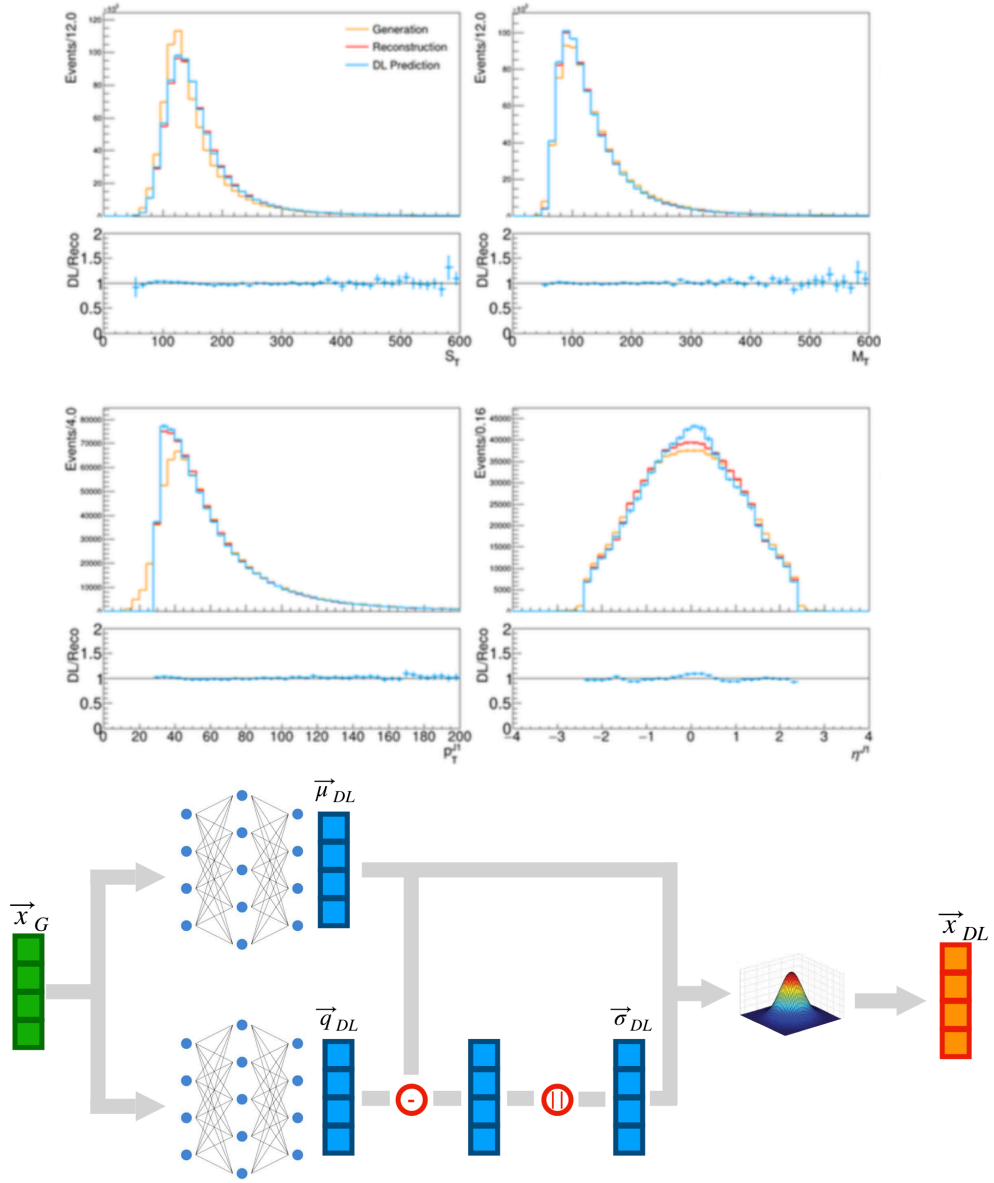


Generative adversarial network may help producing samples with **higher statistical power** than the one used for training.

Suiting Models

Invertible networks or partons to detector and back again [2006.06685]

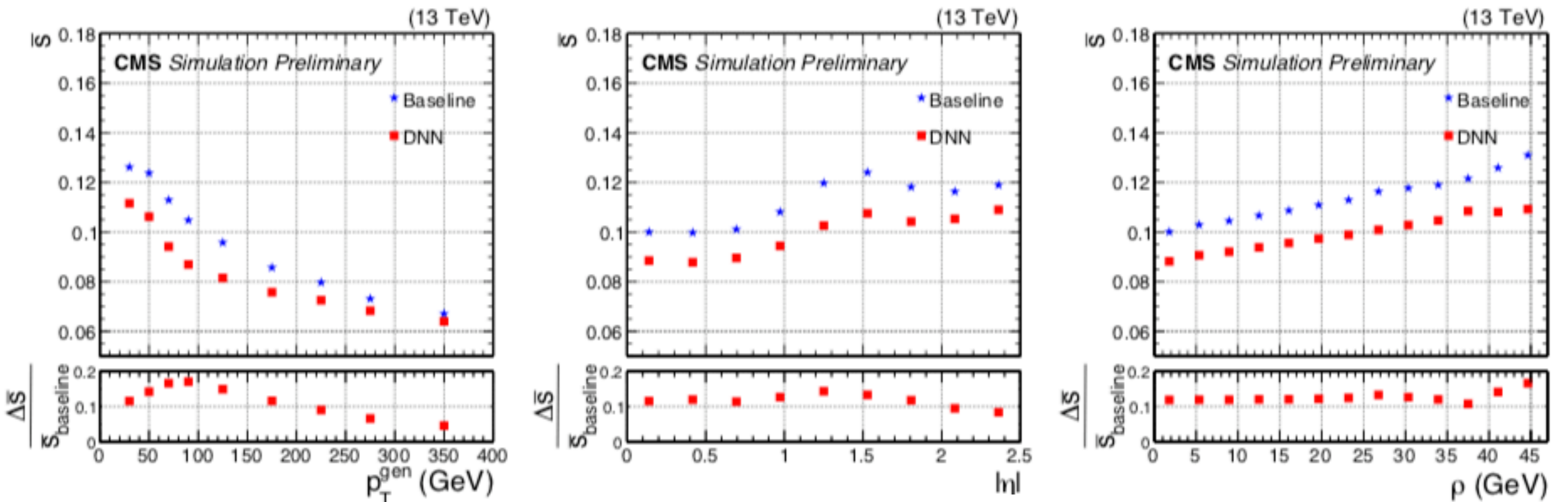


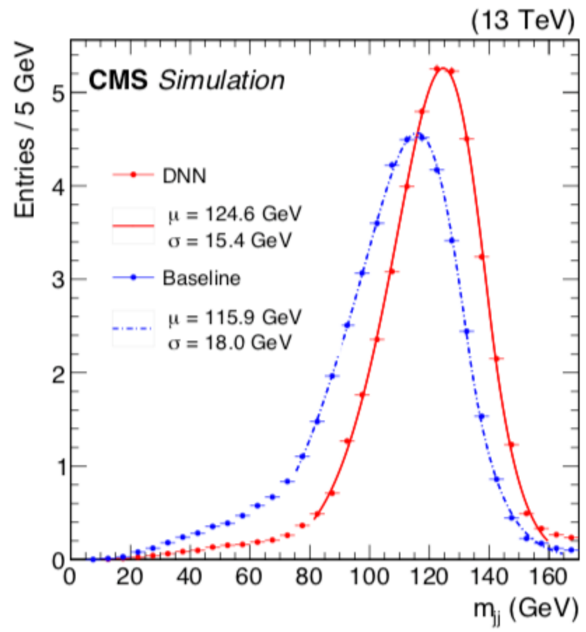


Data augmentation at the LHC through analysis-specific fast simulation with deep learning [2010.01835]

Learn the parton⇛detector function instead of generating samples from vacuum.

Calibrating Data

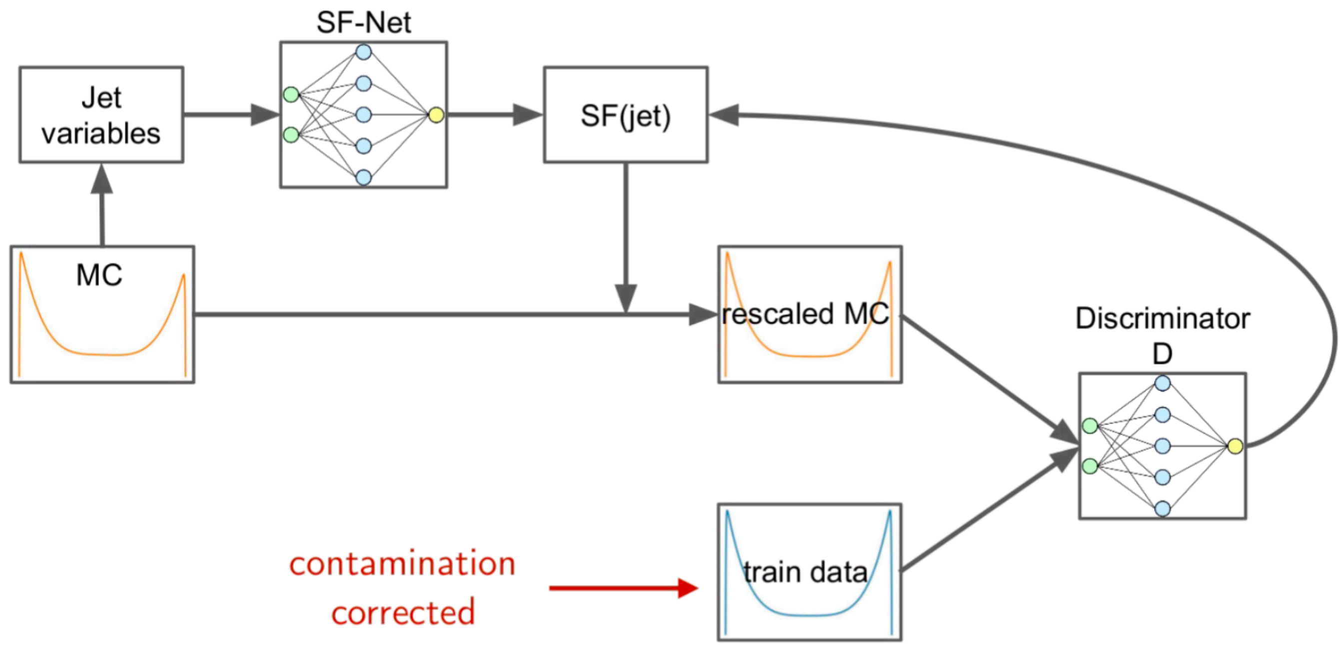




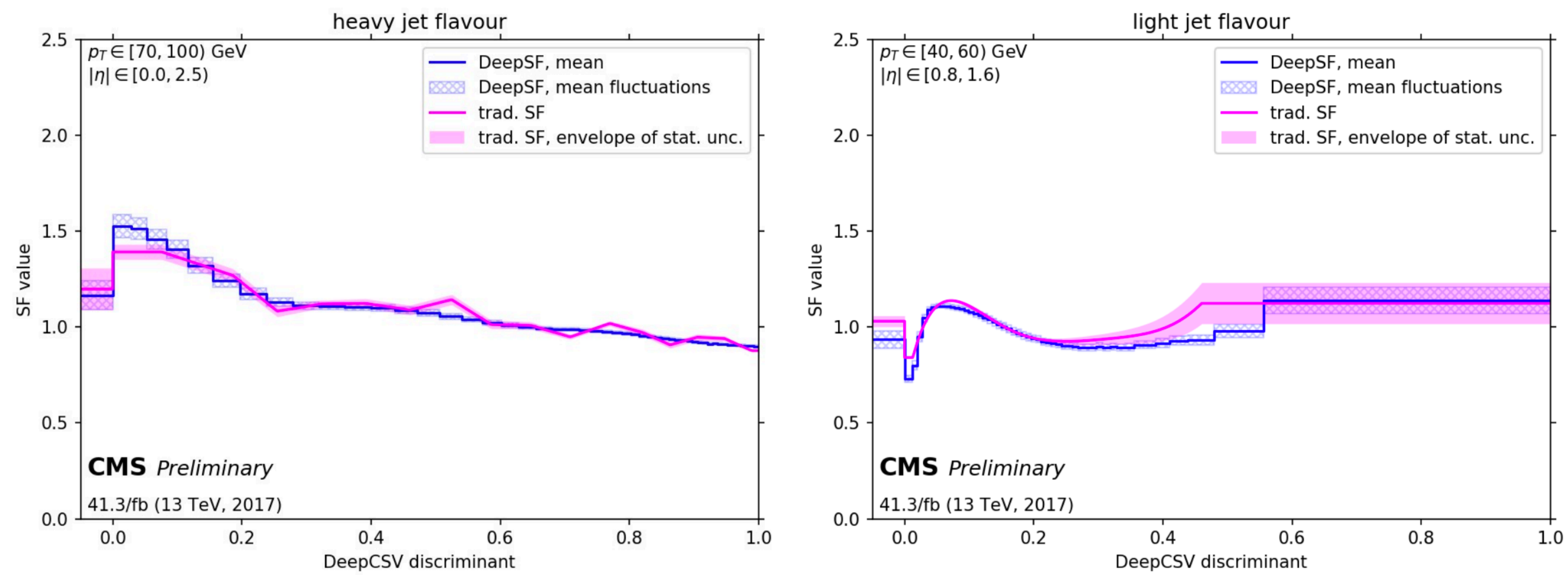
A deep neural network for simultaneous estimation of b jet energy and resolution [1912.06046]

- Energy regression is the most obvious use case.
- Learning calibrating models from simulation and data.
- Parametrization of scale factors using neural networks.
- Reducing data/simulation dependency using domain adaptation.
- Continued R&D

Tagging Scale Factor

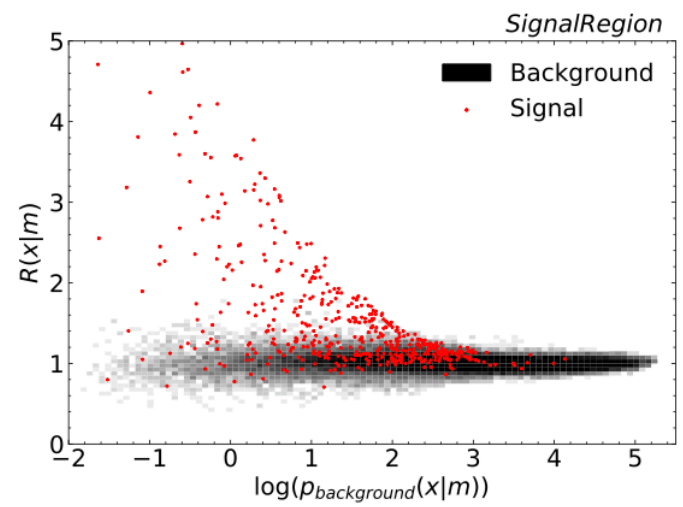


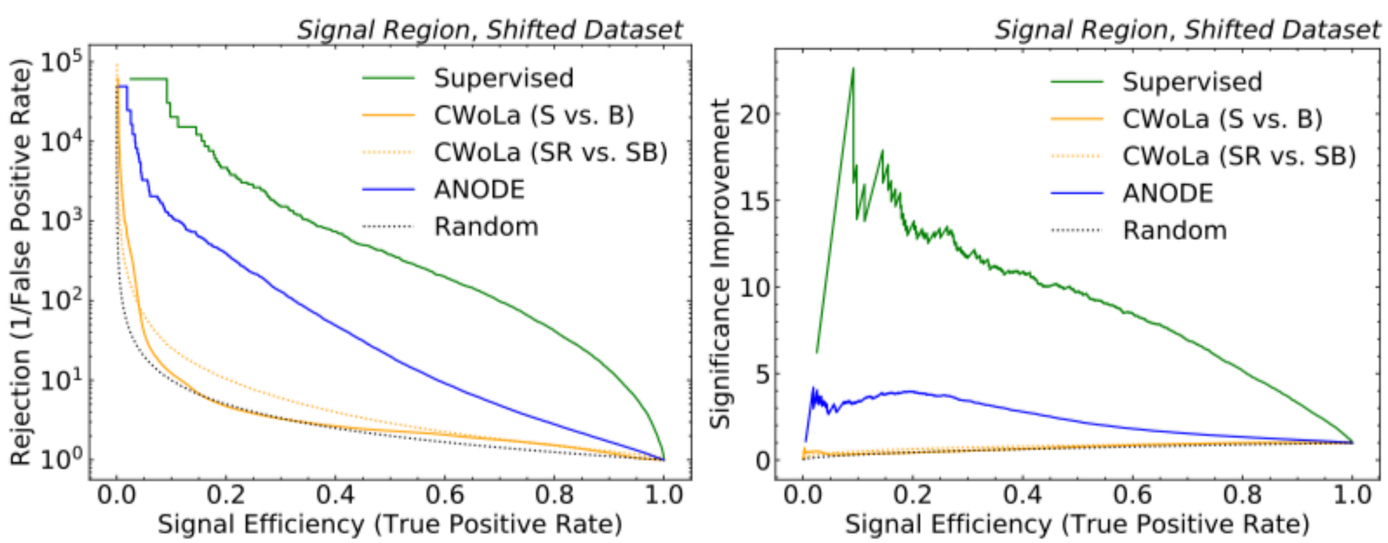
Learn per-jet data/MC scale factor using adversarial technique.



Adversarial Neural Network-based data-simulation corrections for heavy-flavor jet-tagging [cds:2666647]

Analyzing Data





Use of masked autoregressive density estimator with normalizing flow as model-agnostic signal enhancement mechanism.
[doi:10.1103/PhysRevD.101.075042]

More of the relevant works at:
https://iml-wg.github.io/HEPML-LivingReview/

- Machine learning has long infiltrated analysis for signal/bkg classification.
- Increasing number of analysis with more complex DNN.
- Application to signal categorization, bkg modeling, kinematics reconstruction, decay product assignment, object identification, …
- Breadth of new model agnostic methods for NP searches.
- Continued R&D and experiment adoption initiated.

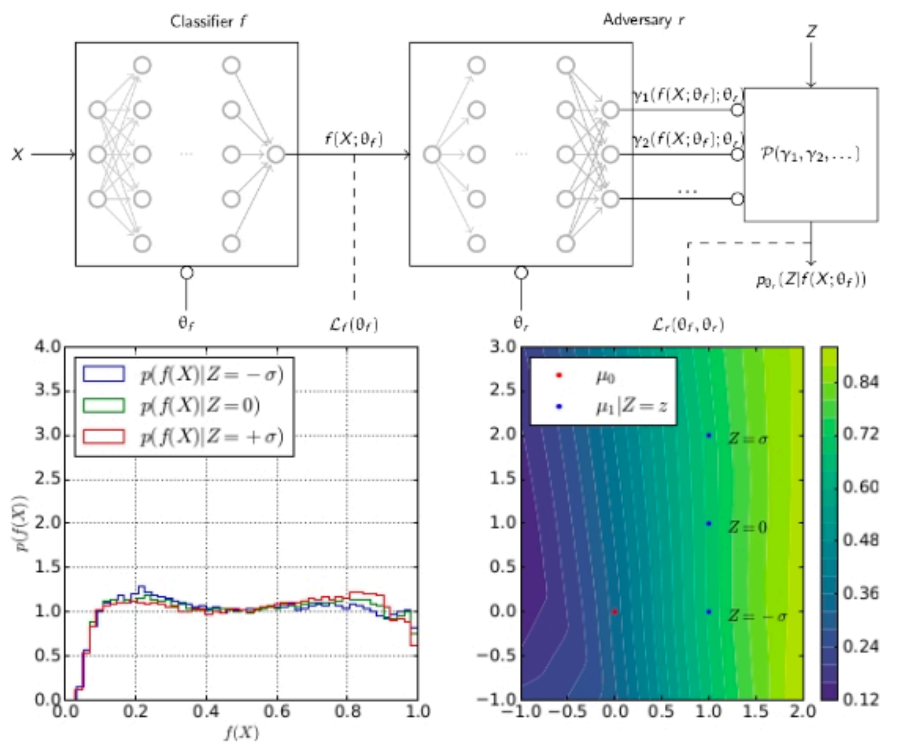
“One-Sided” Hypothesis Testing

- Rigor in calibrating the rate of anomaly is HEP specific (Anomaly detection is not).
- Some methods can serve as a hotline: notification of odd signals.
- Some methods can serve in analysis: calibrated rate of novelty.
- Also of great importance in data quality monitoring/certification.

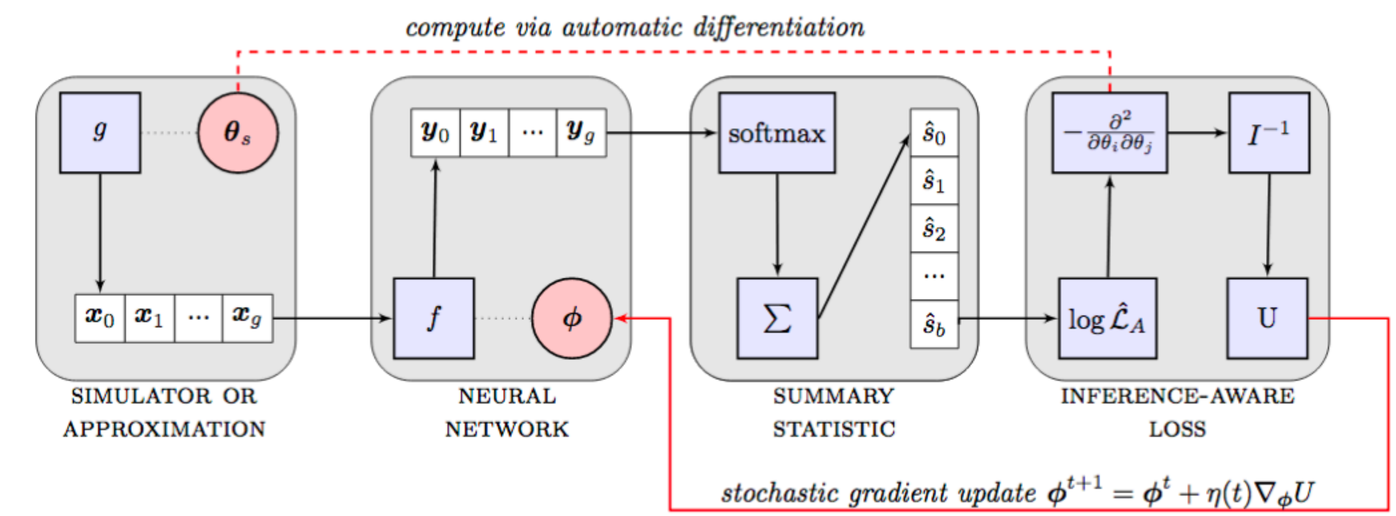
LHC Olympics 2020 [2101.08320]

Individual Approaches

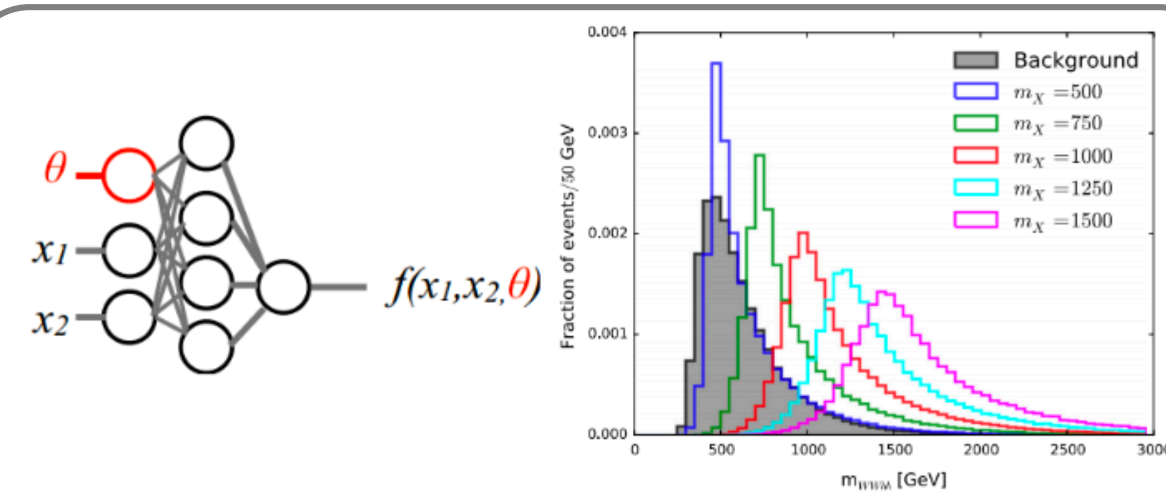
Syst. Estimation and Mitigation



Learn to pivot [1611.01046]



INFERNO: Inference-Aware Optimisation [1806.04743]

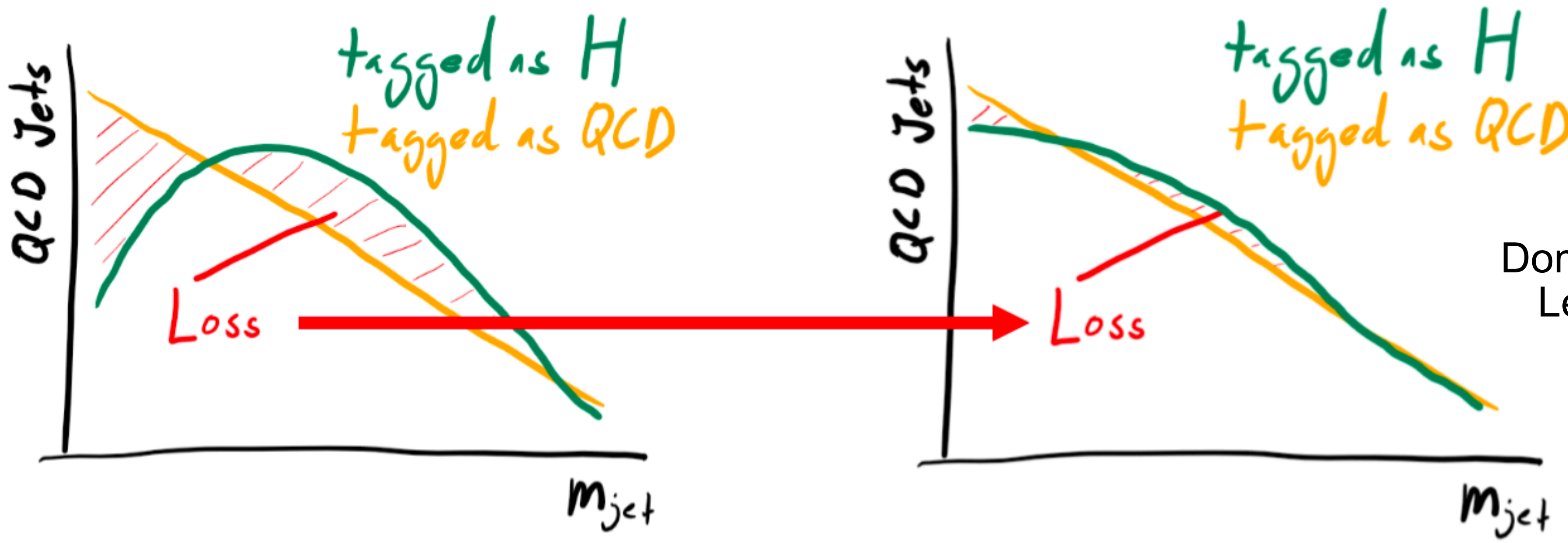


Parametrized Learning [1601.07913]

Systematic uncertainties can be propagated the usual ways.
No additional systematic from the model itself.
Methods to mitigate, propagate and optimize against systematic uncertainties.

De-correlation

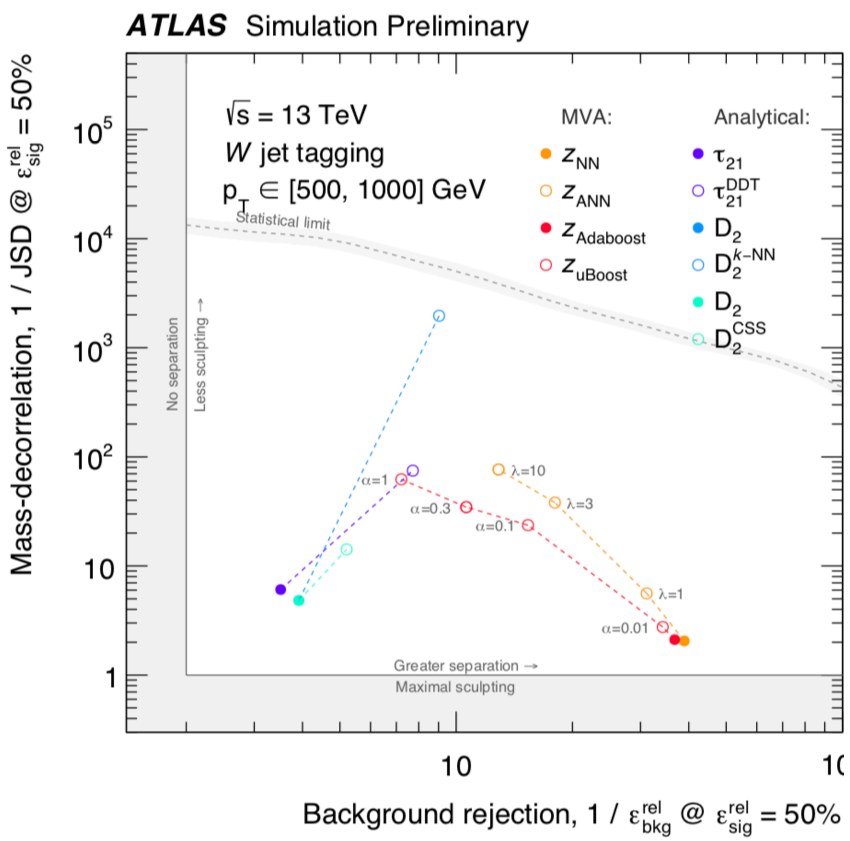
Most background estimation methods (side-bands, ABCD, parametrized fit, …) will require background shape to somehow be independent of analysis selections/processing (not only when using machine learning BTW).



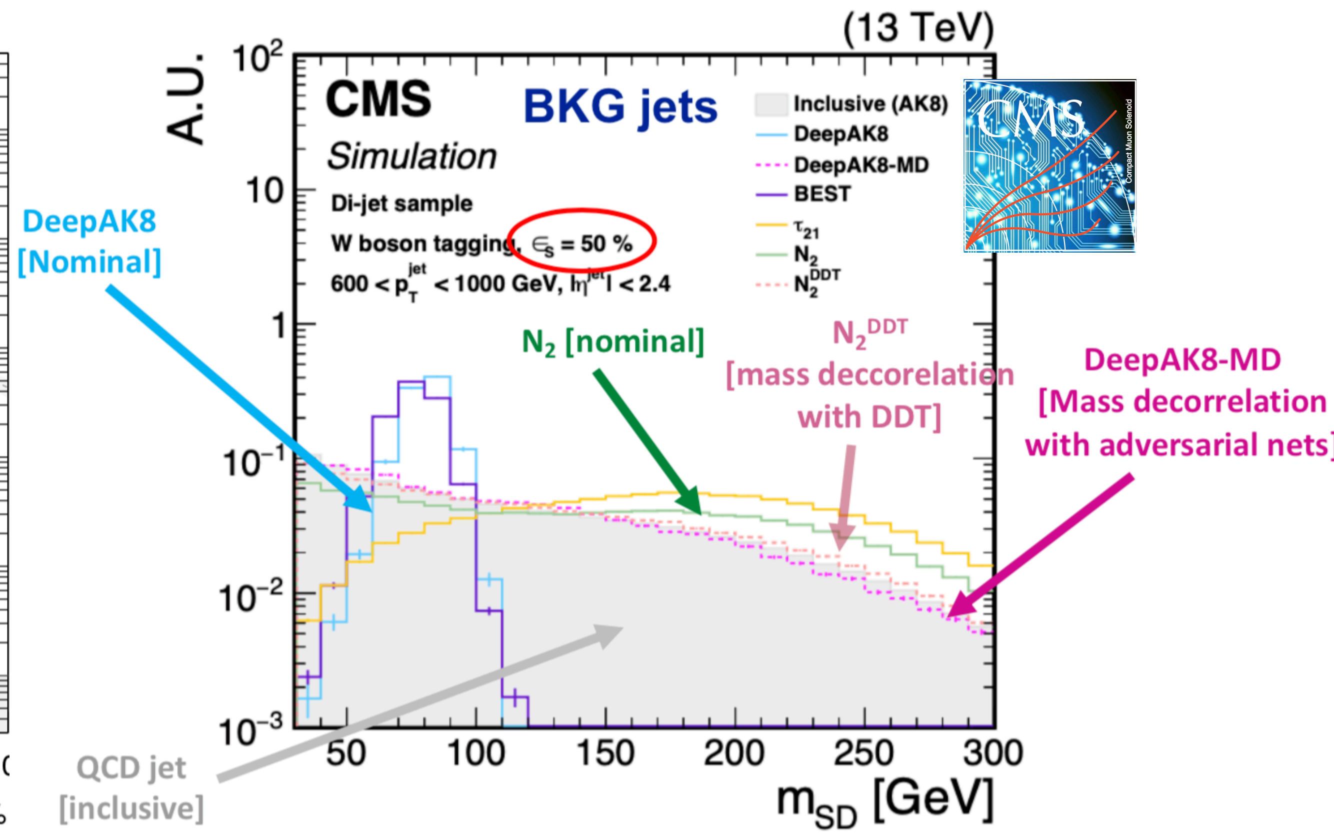
Domain adaptation [1409.7495]
Learn to Pivot [1611.01046]

Numerous methods proposed to de-correlate model predictions and quantities of interest (p_T, mass, …).
Usually adding a term in the loss to constrain de-correlation.

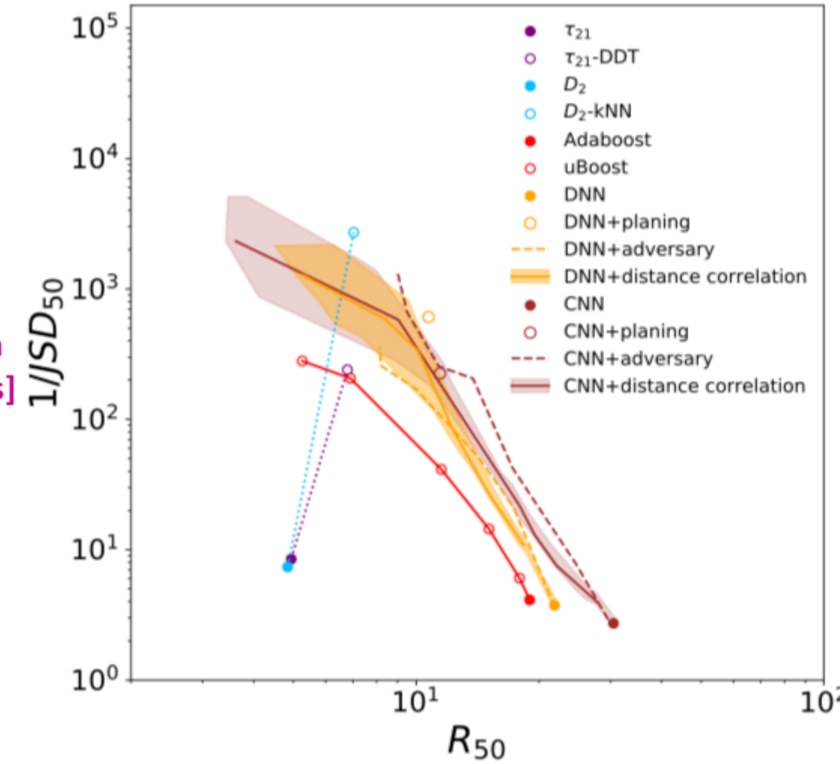
De-correlation Performance



ATLAS Collab. [cds:2630973]



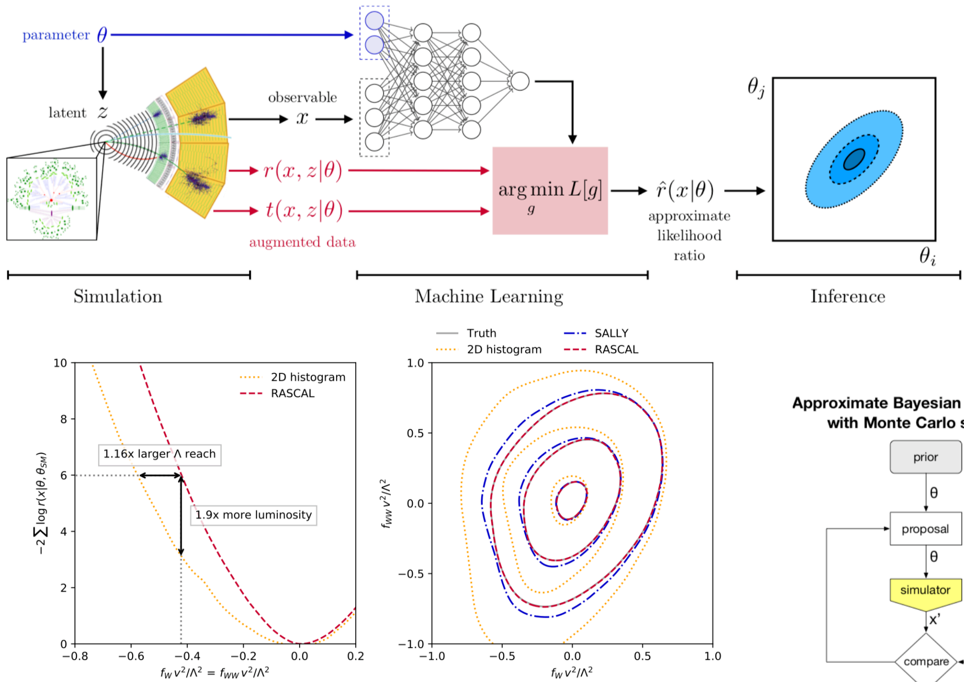
CMS Collab.
[doi:10.1088/1748-0221/15/06/P06005]



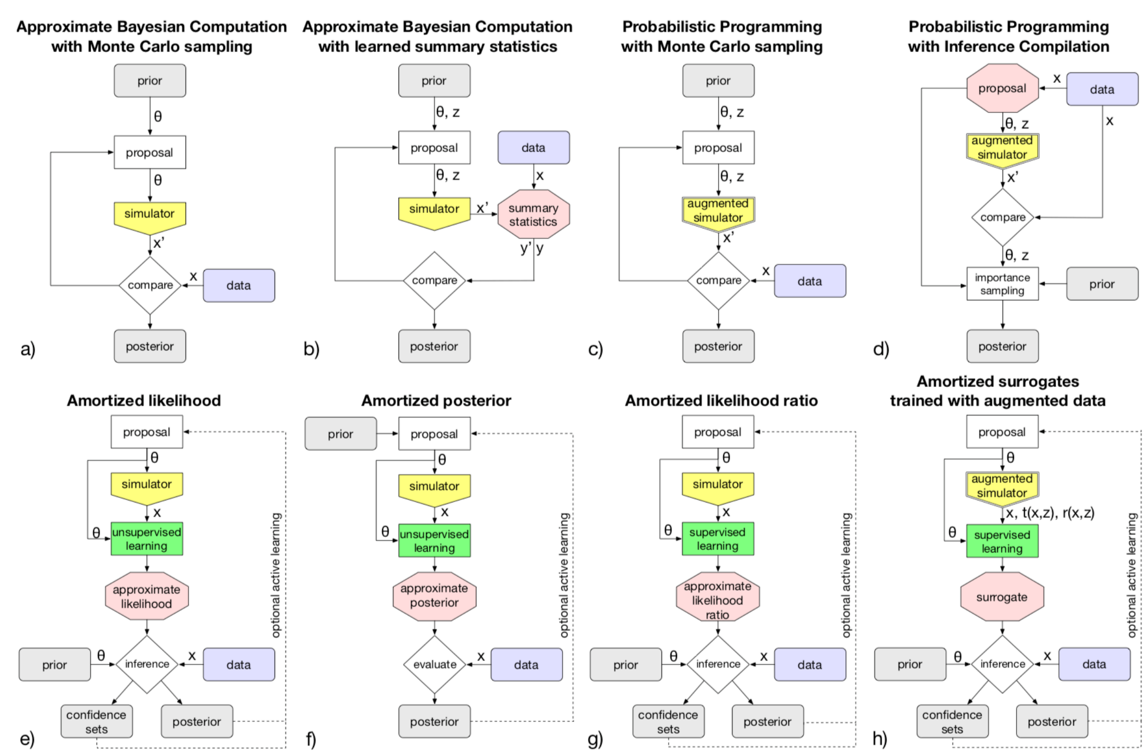
DISCO: Distance Correlation
[2001.05310]

Jenson-Shannon Divergence (JSD) as the comparison metric for shaping.
Residual shaping needs to enter systematics uncertainty estimation.

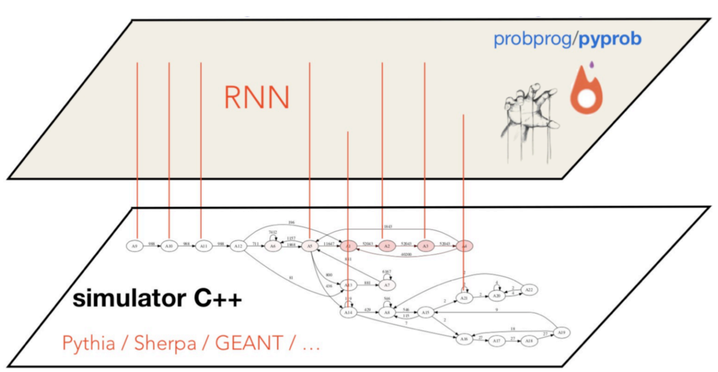
Theory Behind the Data



Constraining EFT with ML
[1805.00013]



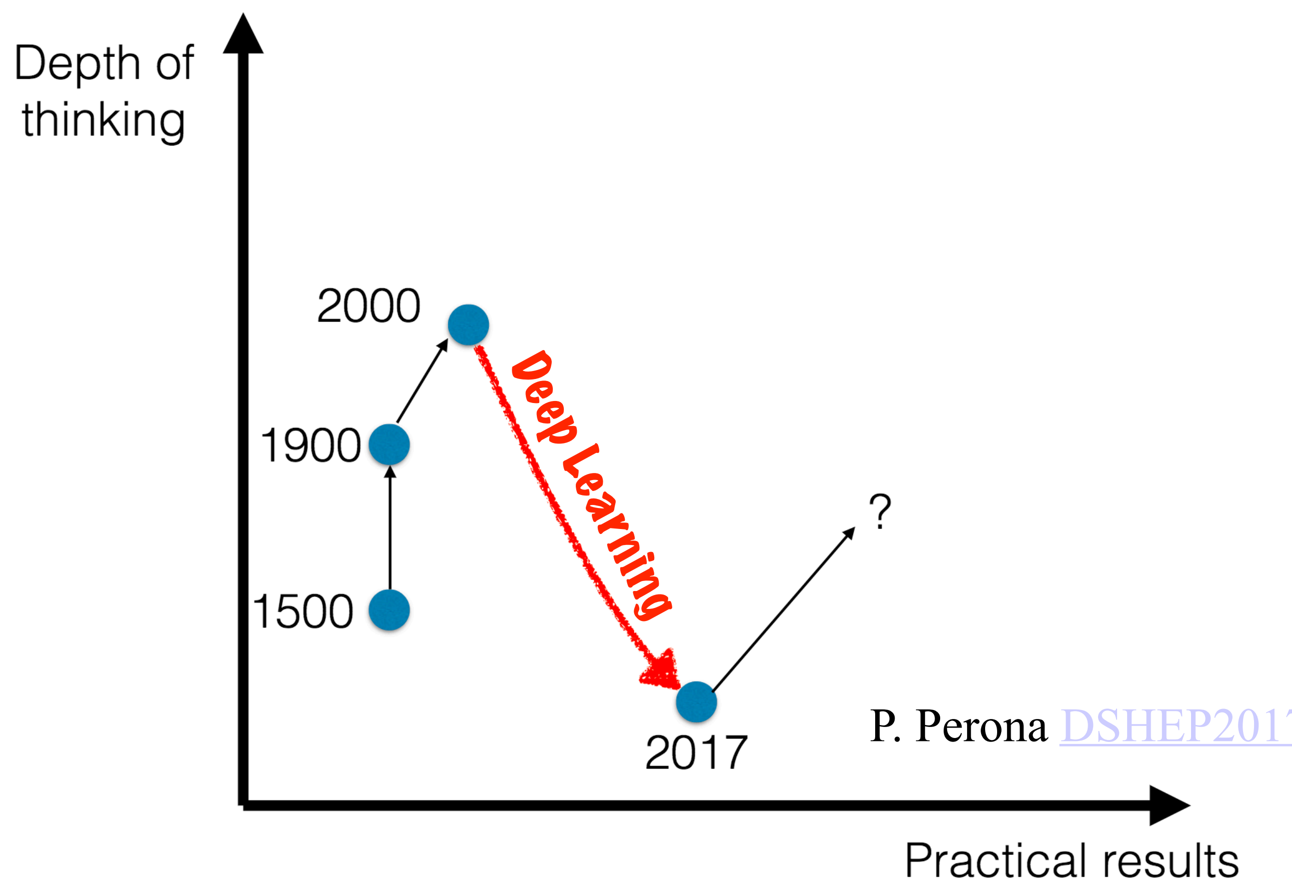
The frontiers of simulation-based inference
[1911.01429]



https://github.com/probprog/pyprob

- Hypothesis testing is the core of HEP analysis.
- Intractable likelihood hinders solving the inverse problem.
- Going beyond the standard approach using machine learning and additional information from the simulator.
- More precise evaluation of the priors on theory's parameters.
- May involve probabilistic programming instrumentation of HEP simulator.
- R&D to bring this in the experiment.

The Black-box Dilemma



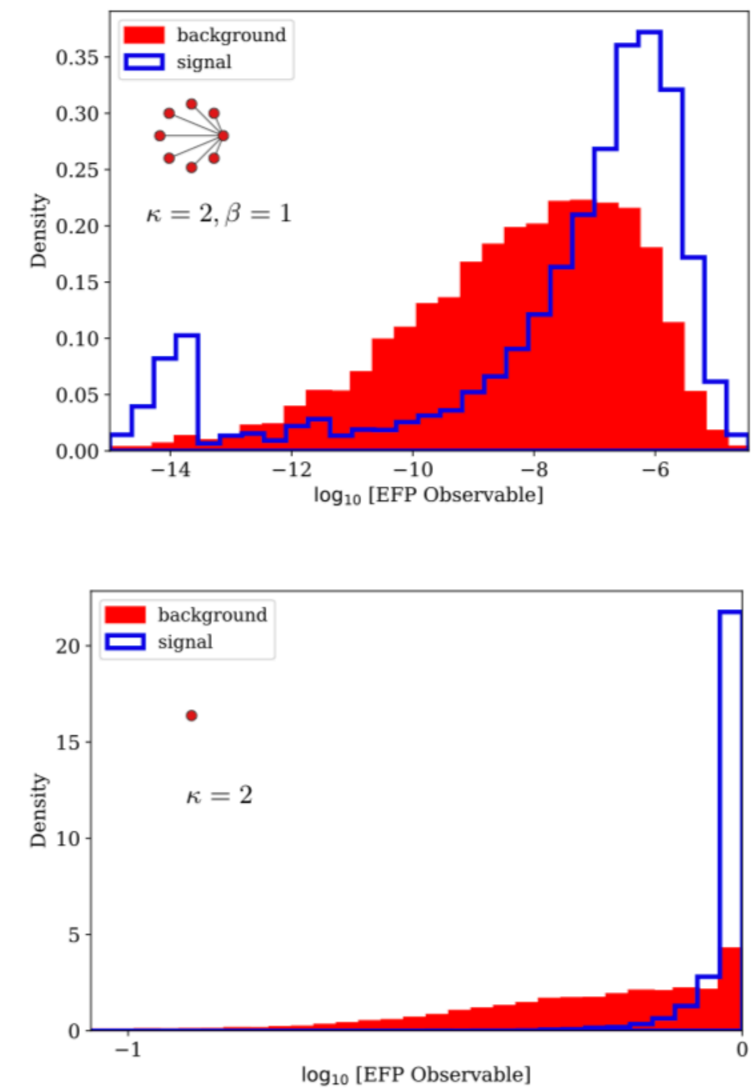
Deep learning may yield great improvements.
Having the “best classification performance” is not always sufficient.
Forming an understand of the processes at play is often crucial.

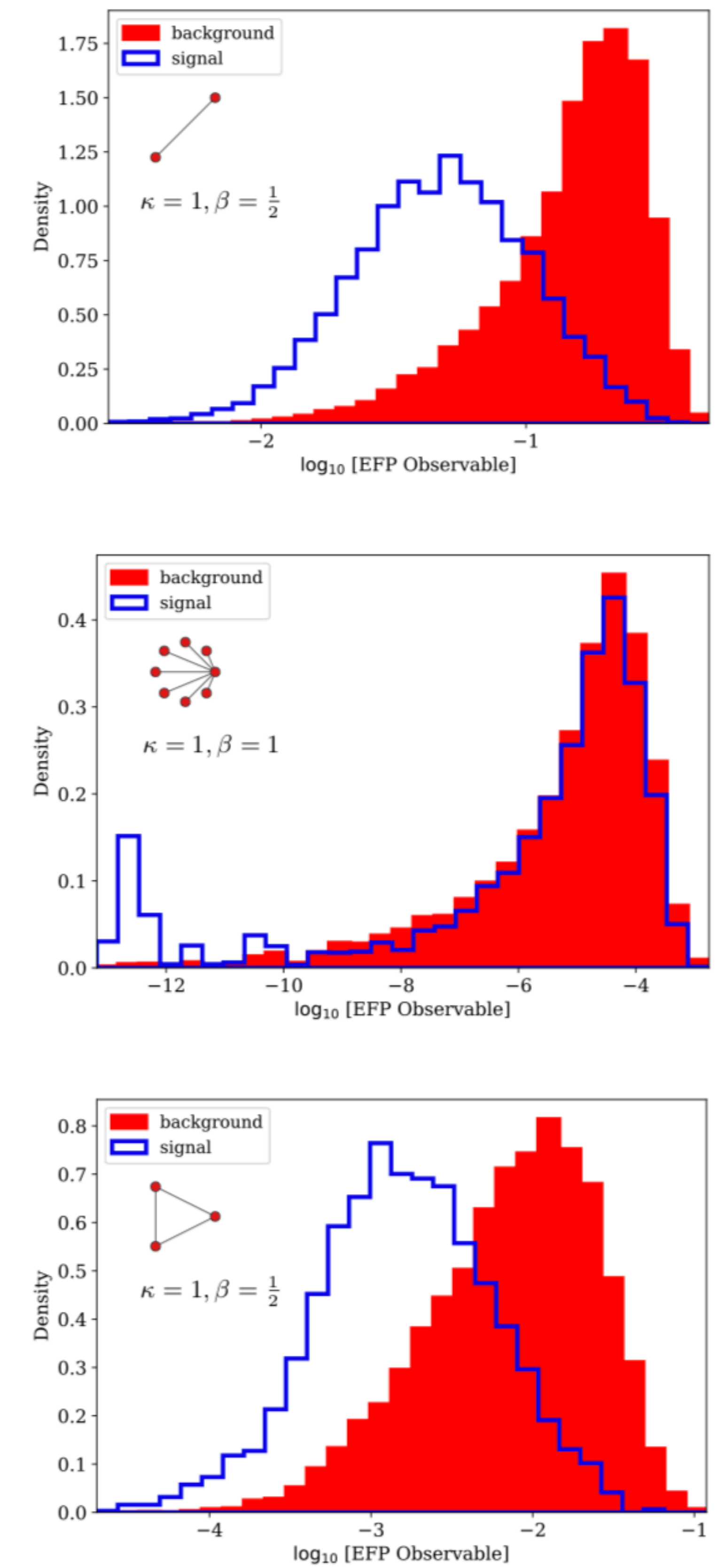
Learning Observables

Electron classification performance

Base		Additions (κ, β)		(AUC)
7HL				0.945
7HL	$+M_{\text{jet}}$			0.956
7HL		$(1, \frac{1}{2})$		0.970
7HL	$+M_{\text{jet}}$	$(1, 1)$	$(1, \frac{1}{2})$	0.971
7HL		$(2, -)$		0.970
7HL	$+M_{\text{jet}}$	$(2, 1)$	$(2, -)$	0.971
CNN				0.972

https://arxiv.org/abs/2010.11998
https://arxiv.org/abs/2011.01984





Search in the space of functions using decision ordering.
Simplified to the energy flow polynomial subspace.
Extract set of EFP that matches DNN performance.

Take home message :

Rapid growth of machine learning applications in HEP

(too) Slowly turning proofs of concept into production

Exciting time ahead exploiting further the potential of AI

Inference Engines

"On-Board accelerator"

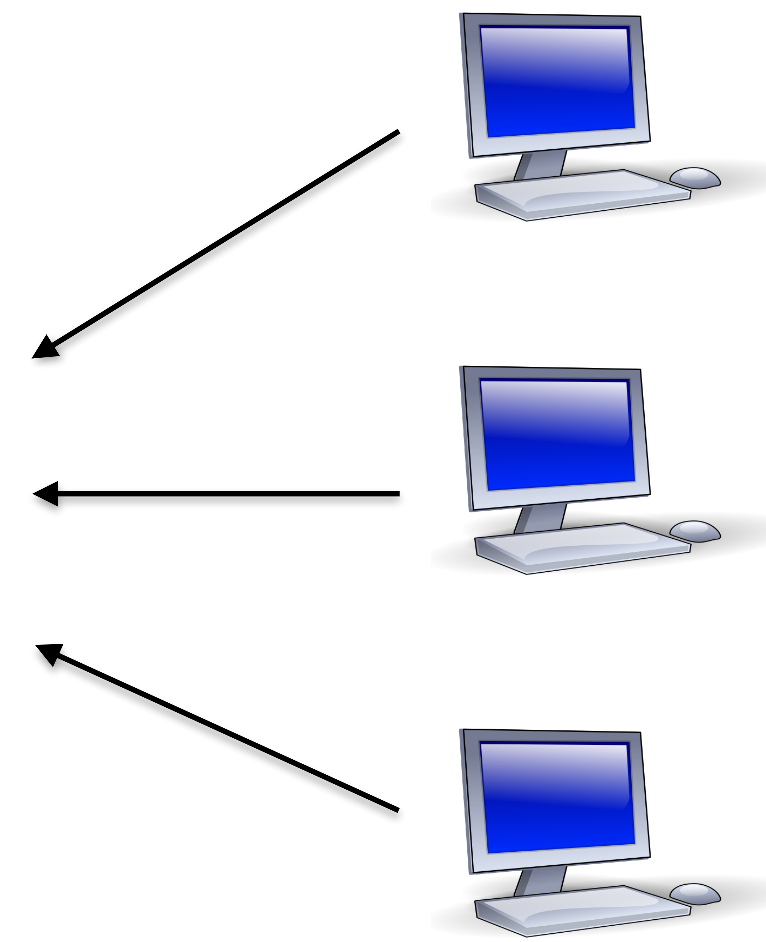
CPU
- Small models
- Small datasets
- Useful for design space exploration

GPU
- Medium-to-large models, datasets
- Image, video processing
- Application on CUDA or OpenCL

TPU
- Matrix computations
- Dense vector processing
- No custom TensorFlow operations

FPGA
- Large datasets, models
- Compute intensive applications
- High performance, high perf./cost ratio

"Remote accelerator"

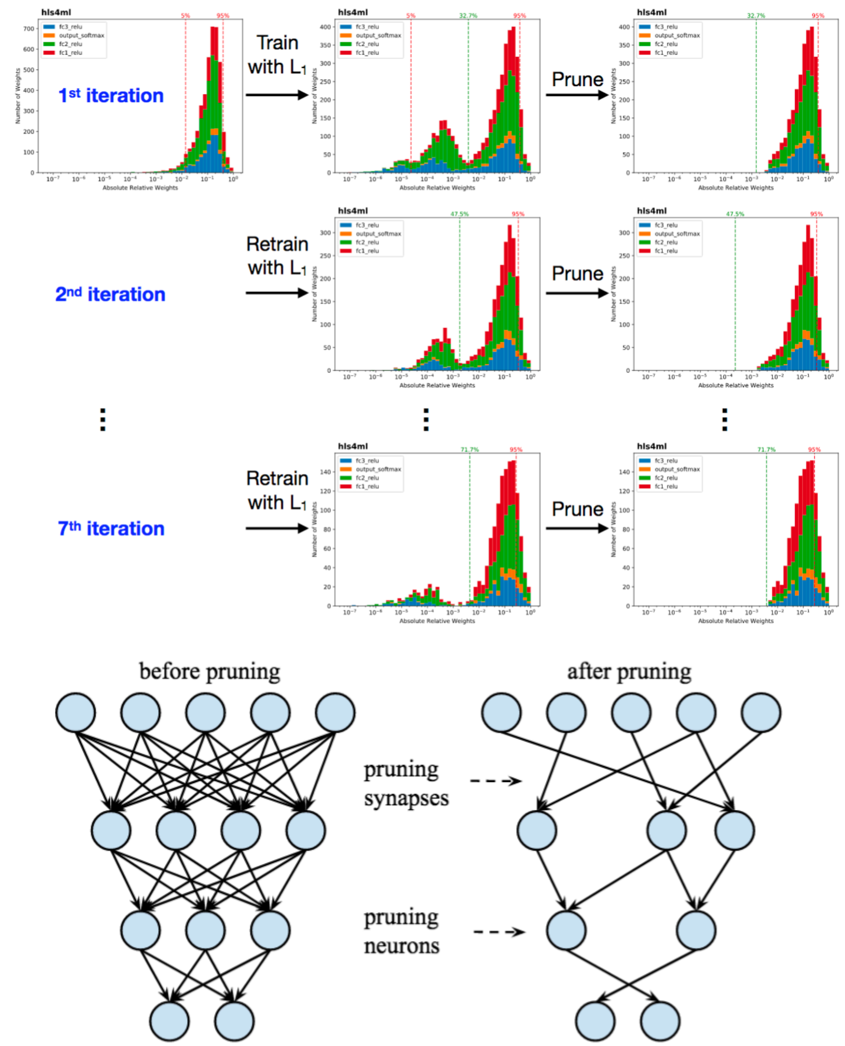


[1811.04492], [2007.10359], [2007.14781]

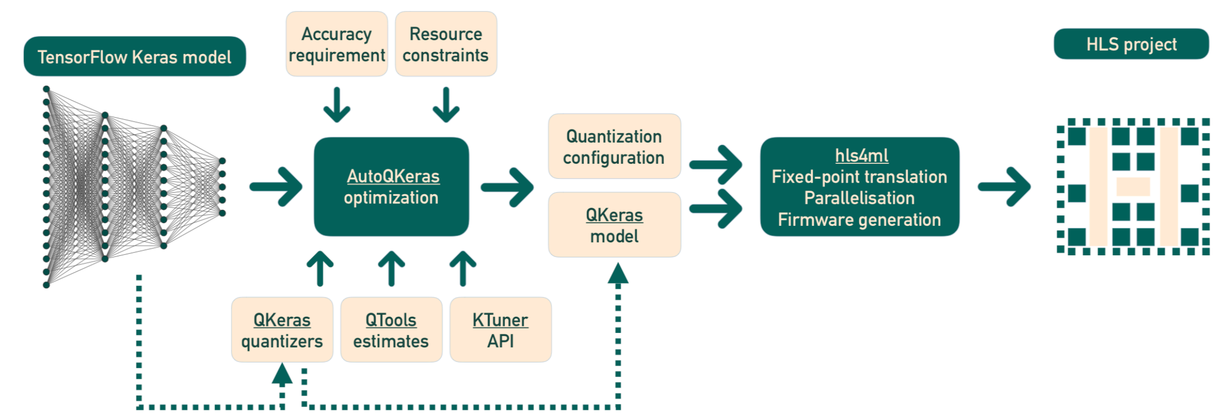
Growing list of deep learning accelerators.
Location of the device is driven by the environment (HLT, Grid, …).

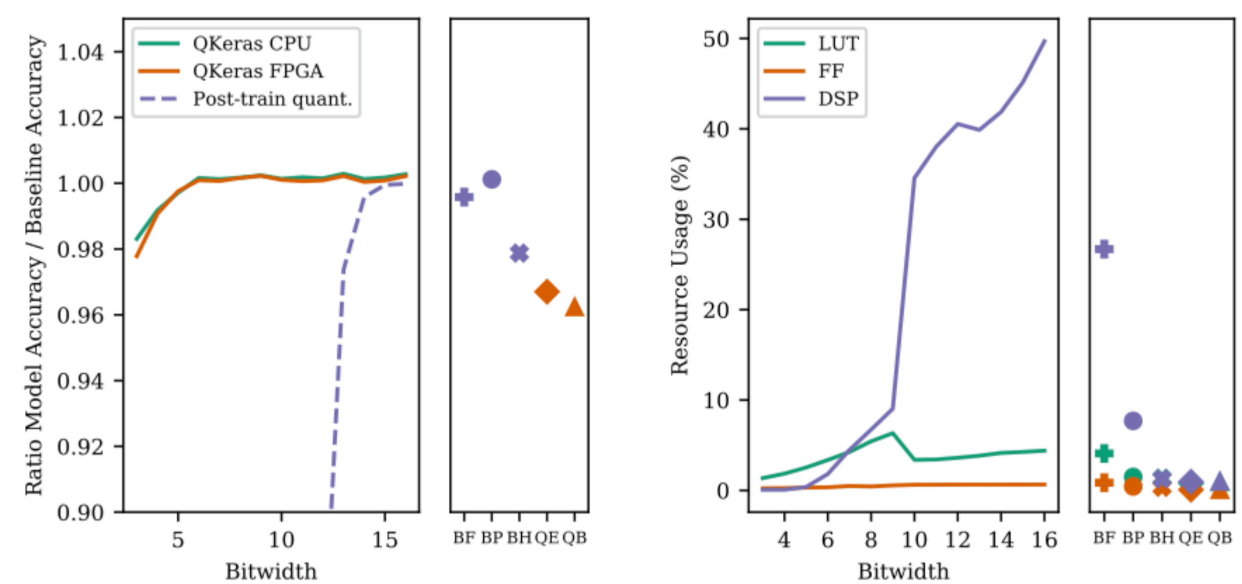
Model Compression

Fast inference of deep neural networks in FPGAs for particle physics [1804.06913]



Automatic deep heterogeneous quantization of Deep Neural Networks for ultra low-area, low-latency inference on the edge at particle colliders [2006.10159]





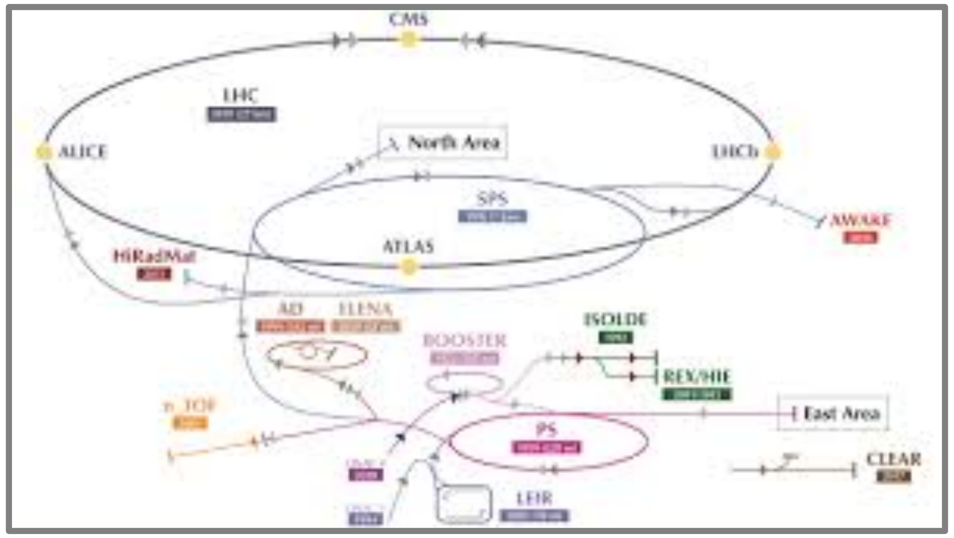
Model inference can be accelerated by reducing the number and size of operations.

Summary

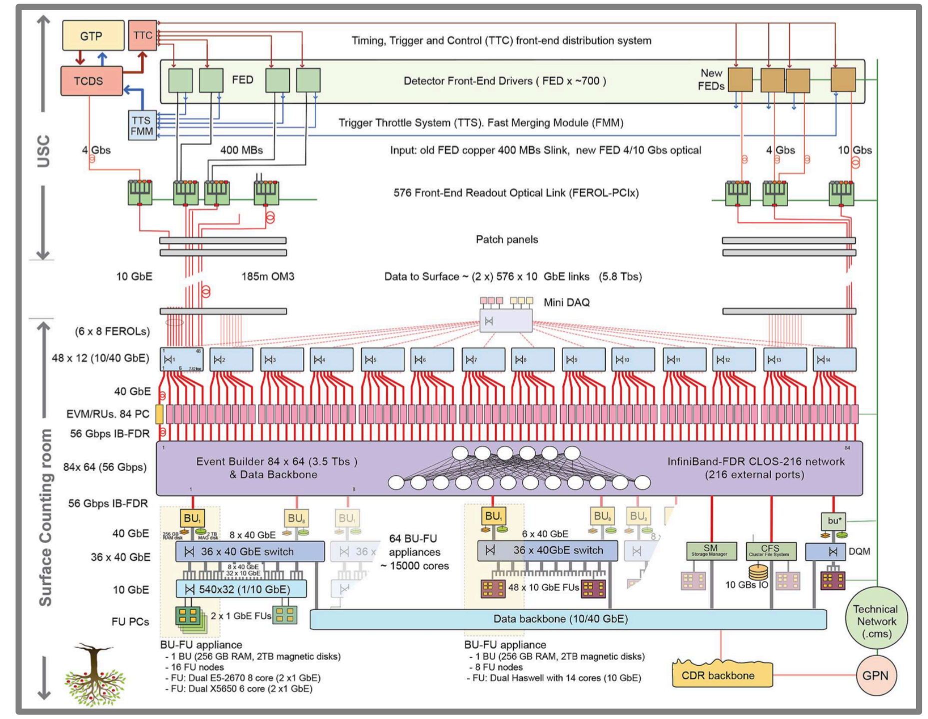
- ➡Physics at collider is a computing intensive endeavor. Extracting, simulating, reconstructing rare signal from large amount of data.
- ➡Deep learning offers great prospects for Science and Physicists. Fast and efficient data processing.
- ➡Deep learning is entering High Energy Physics data processing at all levels.
- ➡Advancing particle physics: smarter operation, faster data processing, light-speed simulation, more refined information extracted, ...
- ➡A lot done since DS@LHC15 , a long way to go for more integration to experimental workflows.

DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA
mPP
4pc

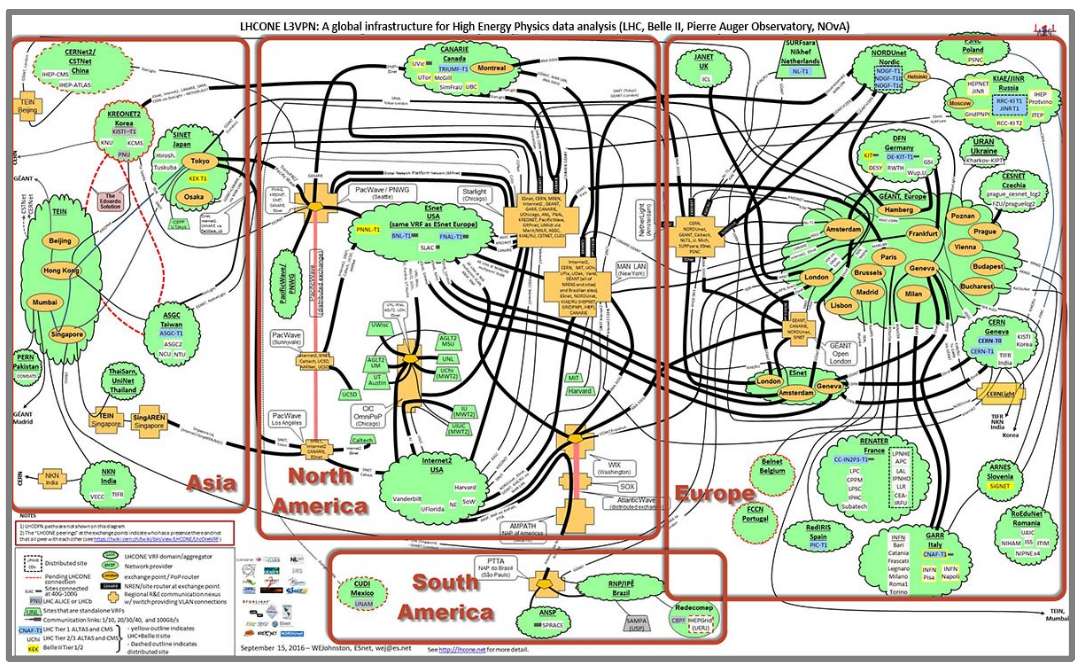
HEP Instruments



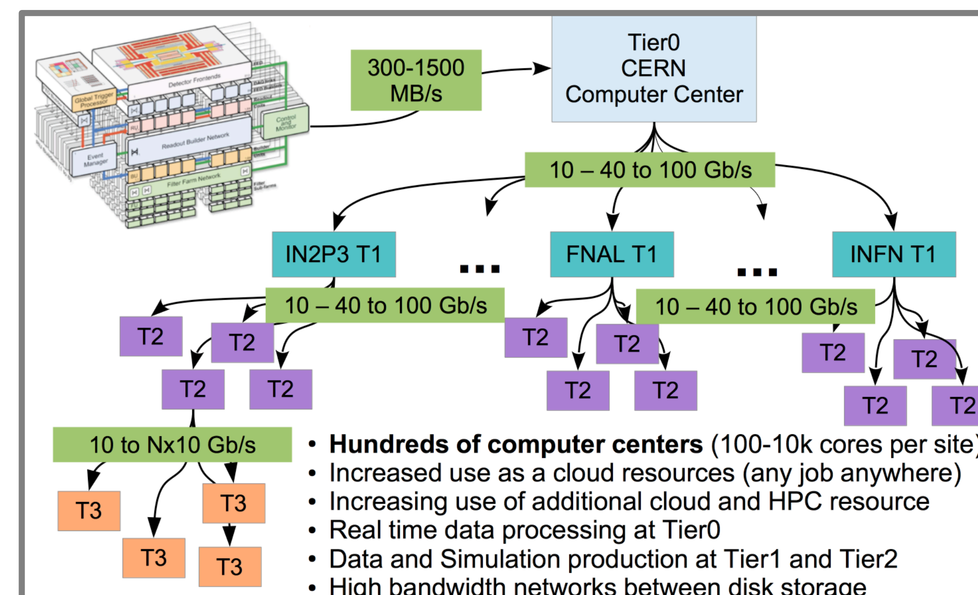
https://home.cern/science/accelerators/



DAQ [IEEE:7111380]



https://home.cern/science/computing/grid



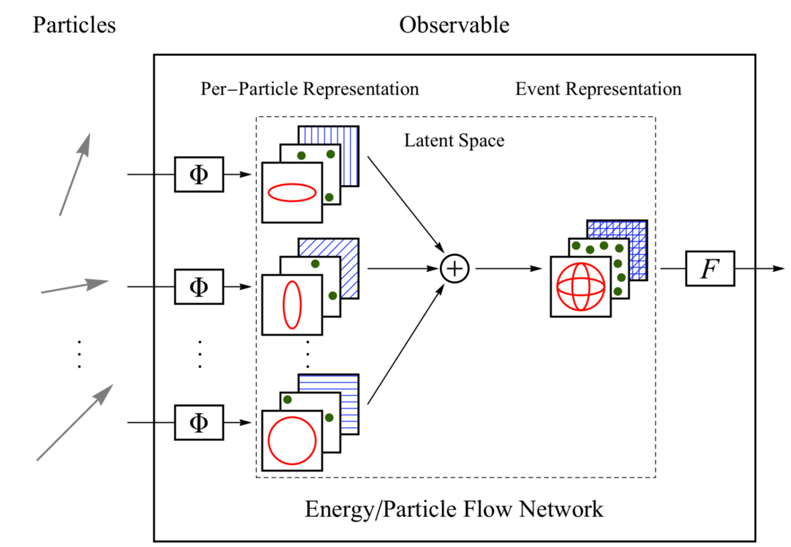
Unique set of complex apparatus for doing Science.

Inductive Bias

$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i} \begin{pmatrix} 1 & 0 & \cdots & 0 & C_{1,N+2} & \cdots & C_{1,M} \\ 0 & 1 & & \vdots & C_{2,N+2} & \cdots & C_{2,M} \\ \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & C_{N,N+2} & \cdots & C_{N,M} \end{pmatrix}$$

$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ w_{jm}^{(E)} \, E(\tilde{k}_m) \\ w_{jm}^{(d)} \, d_{jm}^2 \end{pmatrix}$$

Lorentz Learning Layer
[1707.08966]



Deep set
[1810.05165]

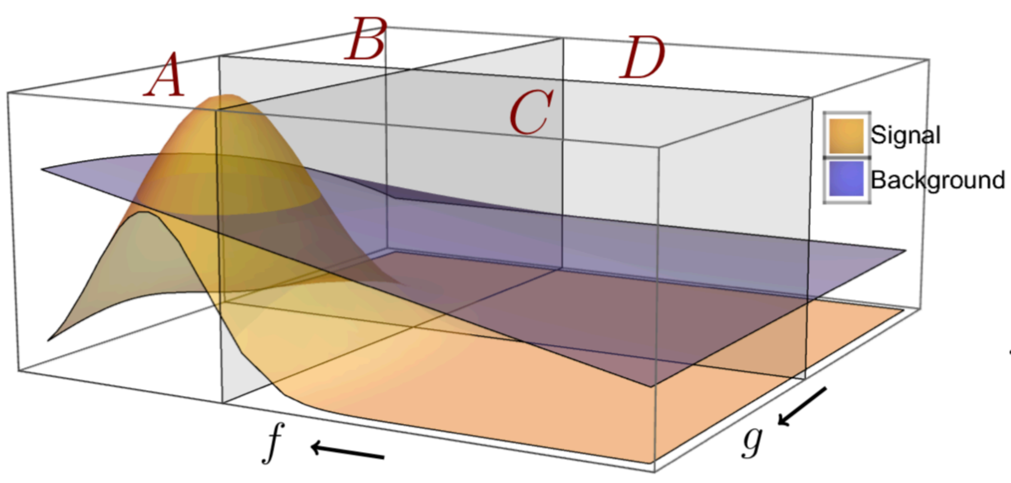
$$\text{In} - W_{\text{in}} - \mathcal{L}_{\text{CG}} - \text{MLP}_{\text{inv}} - \cdots - \mathcal{L}_{\text{CG}} - \text{MLP}_{\text{inv}} - \text{P}_{\text{inv}} - \sum_{\text{particles}} - W_{\text{out}} - \text{Out}$$

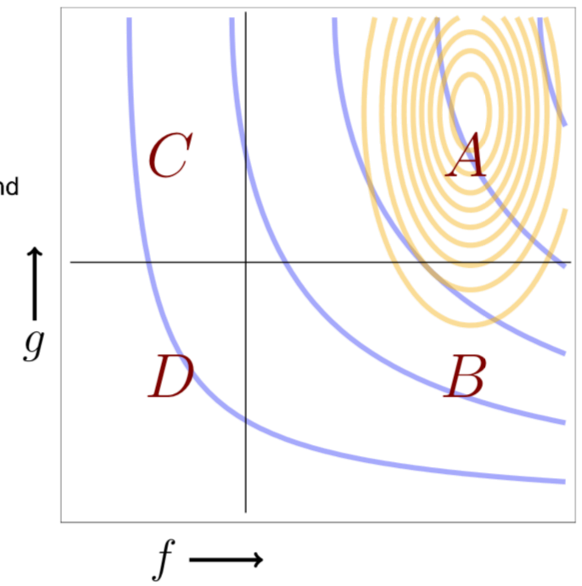
$$\mathscr{F}_i \mapsto W \cdot \left(\mathscr{F}_i \oplus \mathscr{F}_i^{\otimes 2} \oplus \sum_j f\left(p_{ij}^2\right) \cdot p_{ij} \otimes \mathscr{F}_j \right)$$

Lorentz group quivariant networks
[2006.04780]

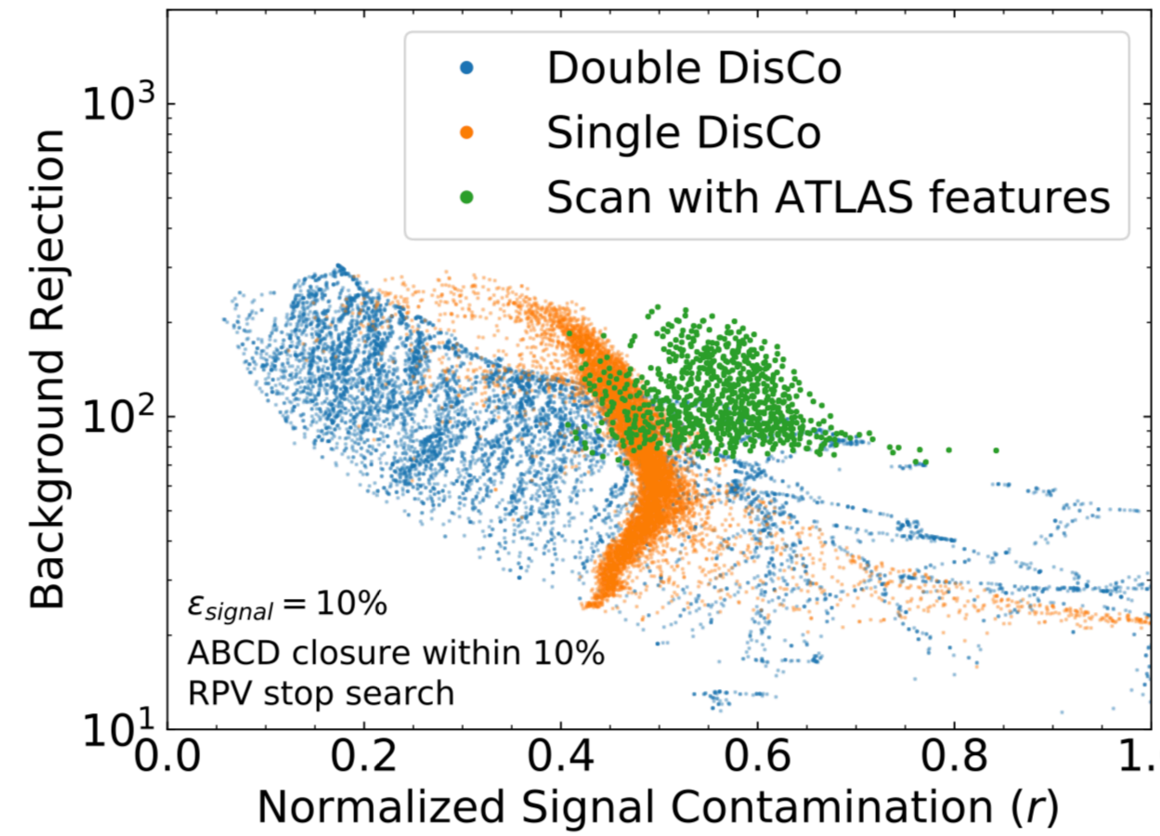
Embed the symmetry and invariance in the model.
Economy of model parameters.

Background Estimation





ABCD + Disco
[2007.14400]



Most popular background estimation method (ABCD), can be optimized for de-correlation, yielding increased significance.

Overview

Reinforcement Learning (cherry)

- The machine predicts a scalar reward given once in a while.
- **A few bits for some samples**

Supervised Learning (icing)

- The machine predicts a category or a few numbers for each input
- **10→10,000 bits per sample**

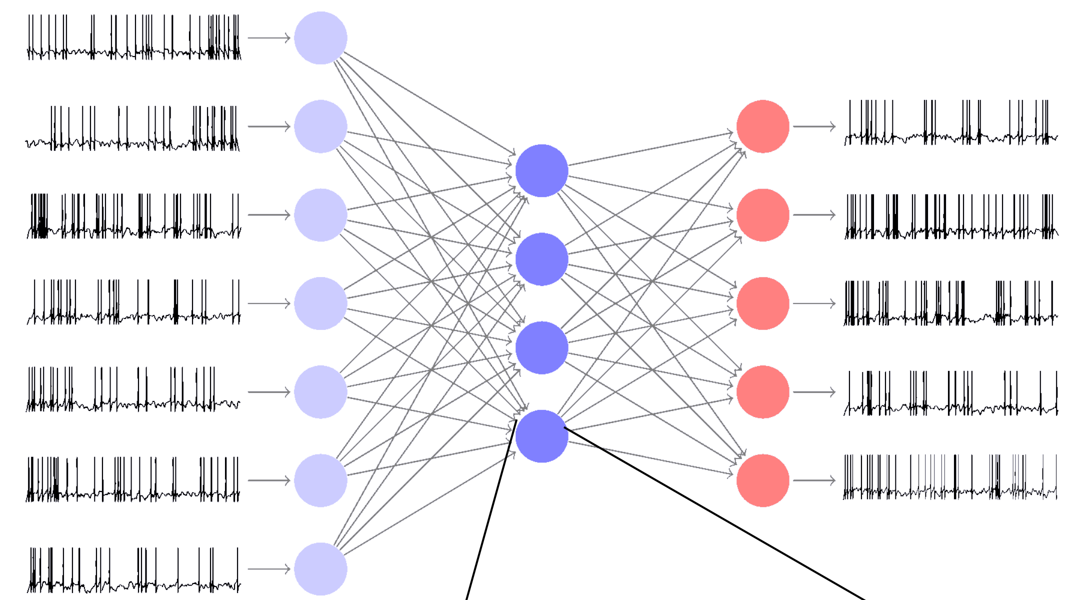
Unsupervised Learning (cake)

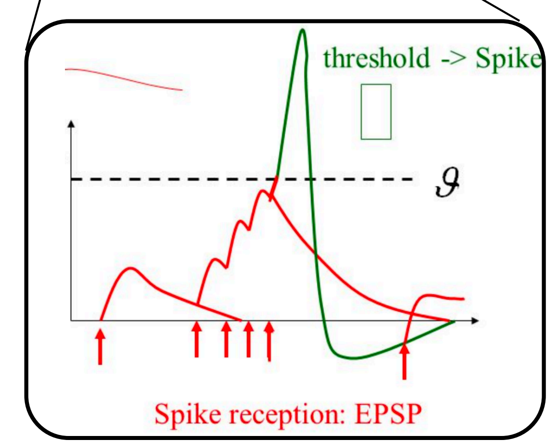
- The machine predicts any part of its input for any observed part.
- Predicts future frames in videos
- **Millions of bits per sample**

Yann Le cun, CERN, 2016

Spiking Neural Network

- Closer to the actual biological brain
- Adapted to temporal data
- Hardware implementation with low power consumption
- Trained using evolutionary algorithms
- Economical models

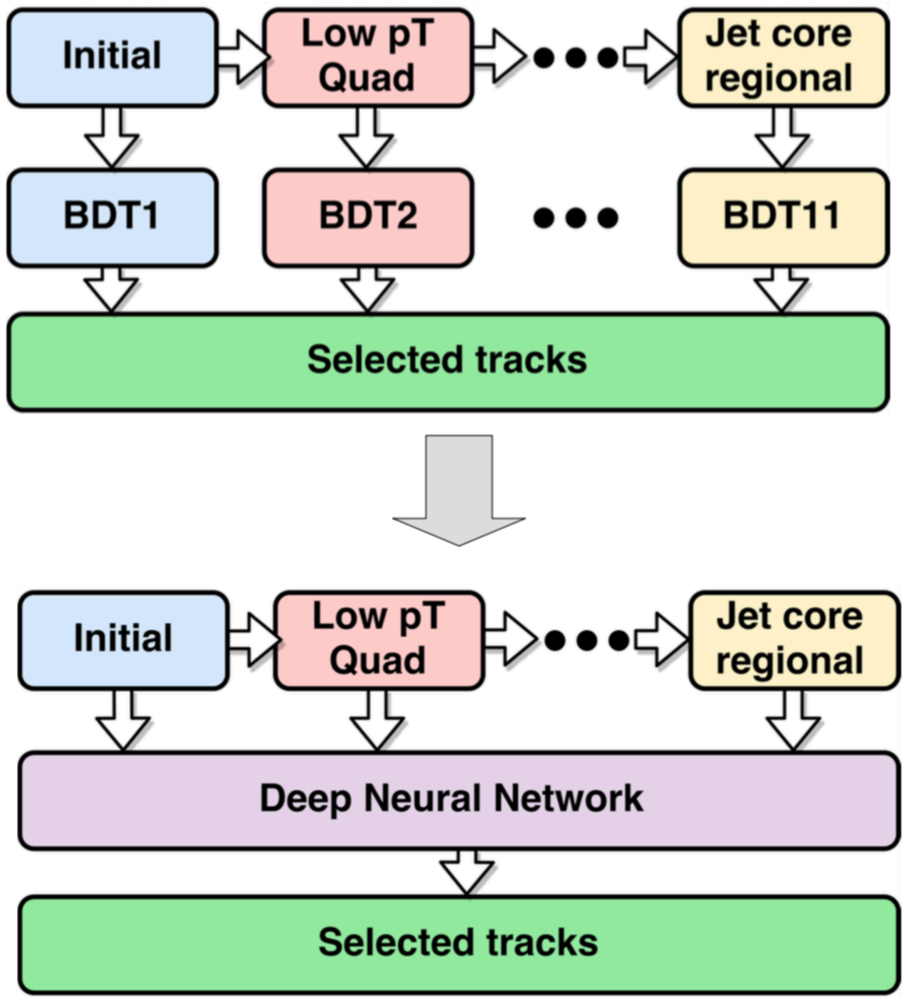


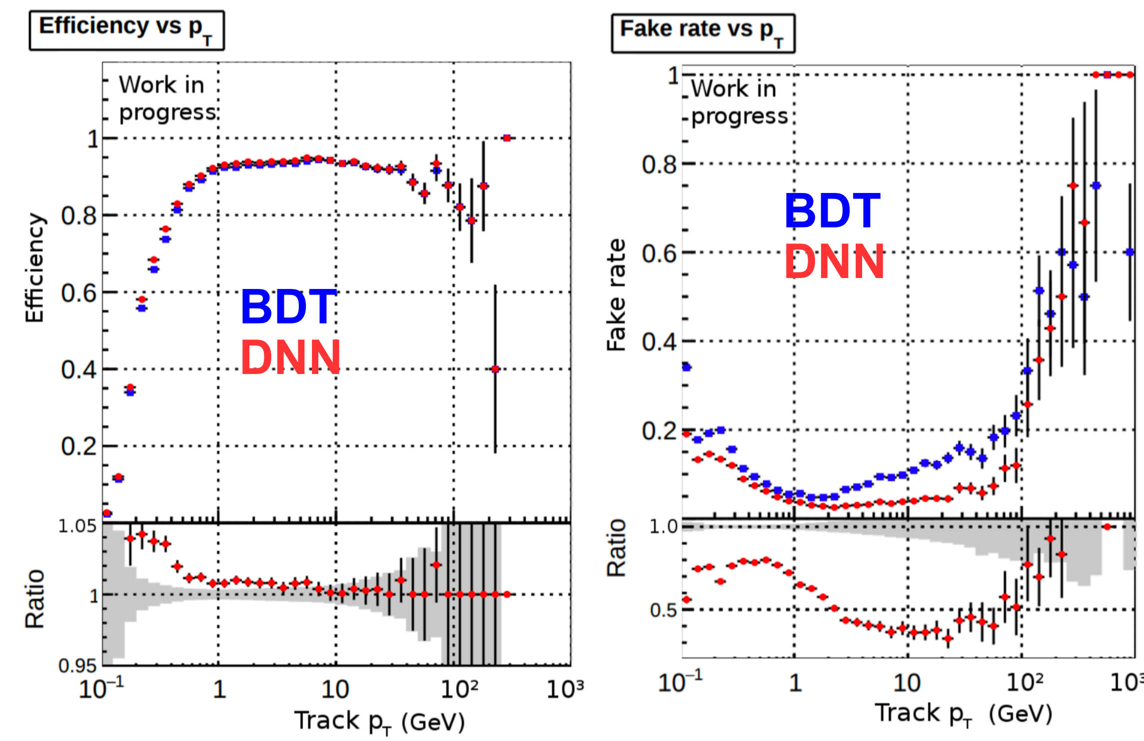


	Deep Learning	Spiking
Training Method	Back-propagation	Not well established (here, genetic algorithms)
Native Input Types	Images/Arrays of values	Spikes
Network Size	Large (many layers, many neurons and synapses per layer)	Relatively small (fewer neurons and sparser synaptic connections)
Processing Abilities	Good for spatial	Good for temporal
Performance	Well understood and state-of-the-art	Not well understood

Charged Particle Tracking R&D

Track Quality with DNN



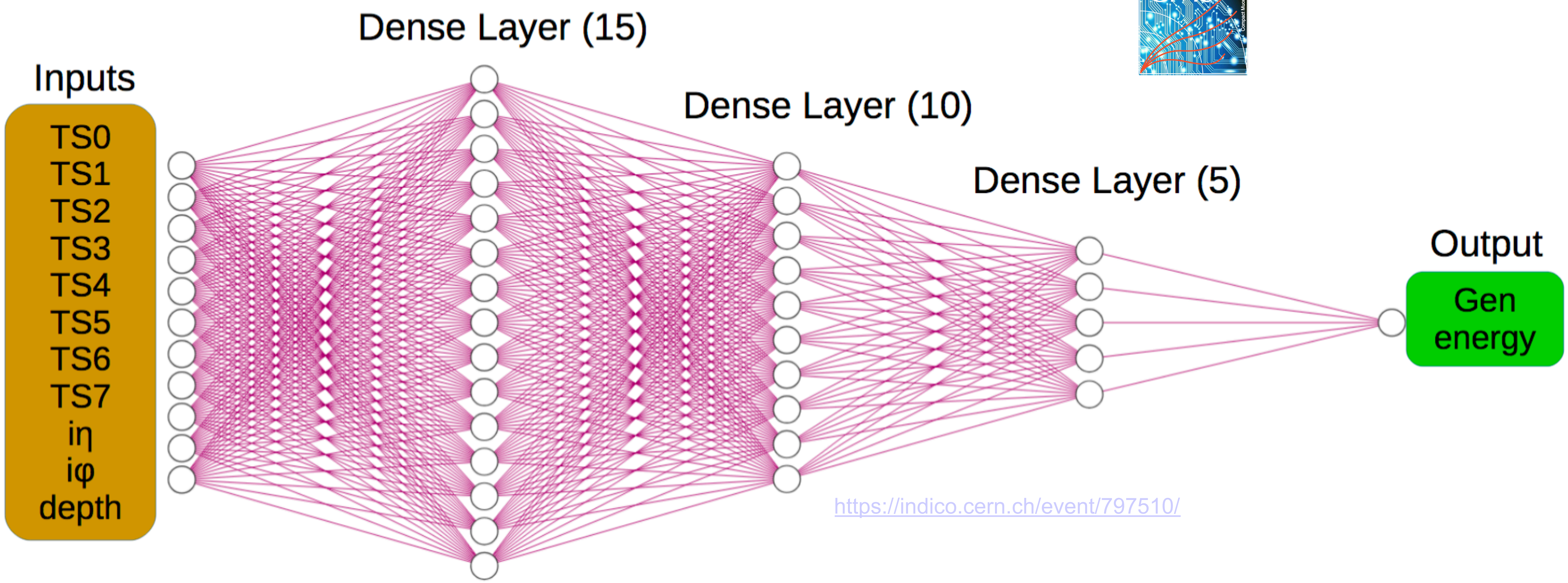


Simplifies and improves track selection within the scope of CMS iterative tracking

https://indico.cern.ch/event/658267/contributions/2813693/

Calorimeter – Jet R&D

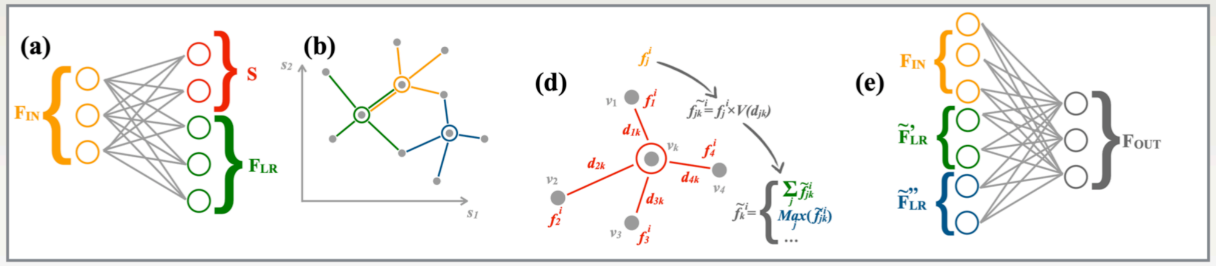
HCAL Energy



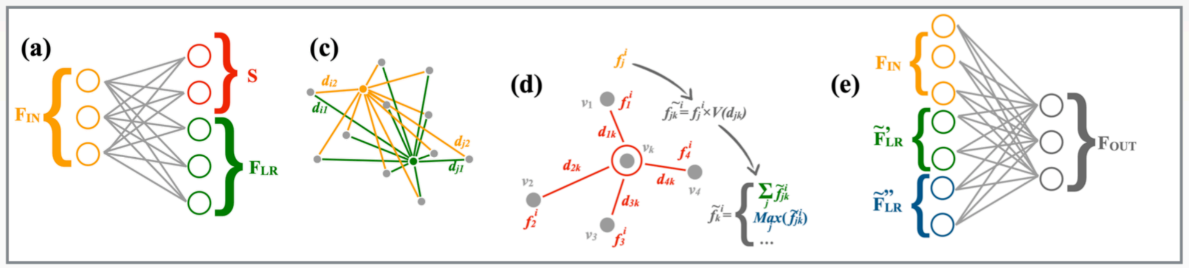
Learn the pre-pileup energy deposition in a regression from the sampled pulse shape.

HGCal Reconstruction

- GravNet

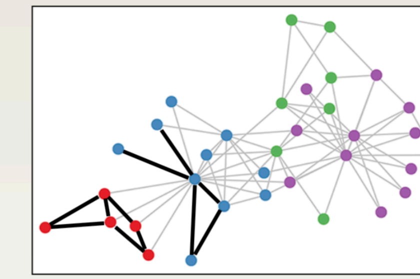


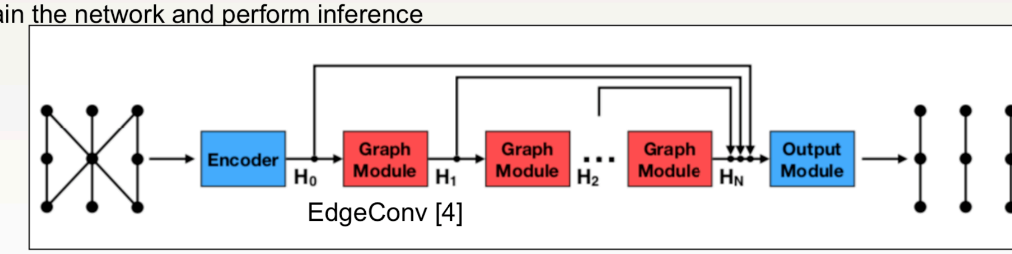
- GarNet



S.R. Qasim, J.K, Y. Iiyama, M Pierini arXiv:1902.07987, EPJC

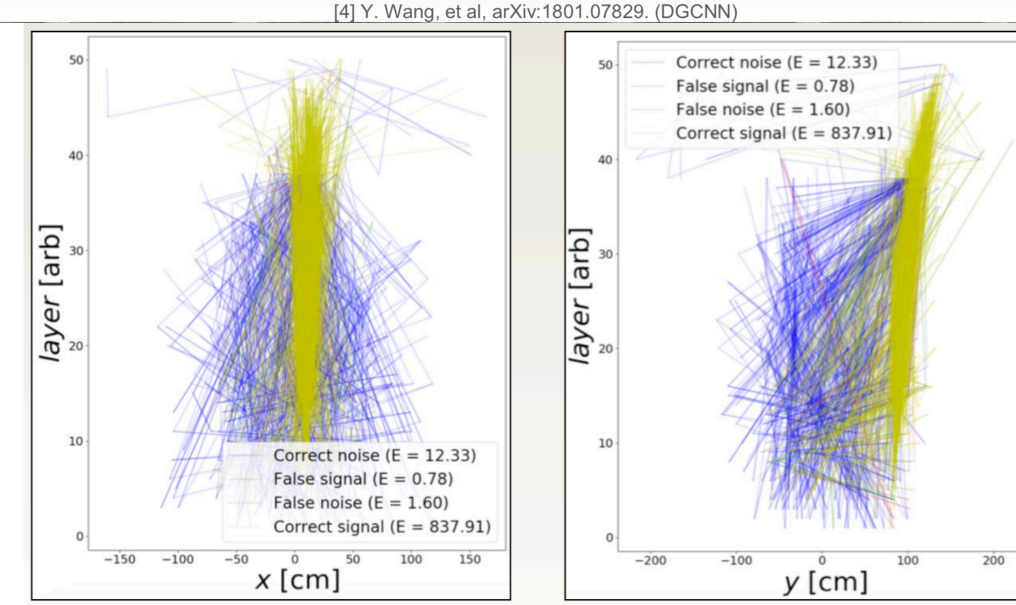
- Inspired by HEP.TrkX [1,2], edge classifiers can overcome the problem
- Objects appear as vertices that are connected to each other, but not connected to others
- Edges can carry additional information like particle ID
- Recipe [3]:
  - Pre-define a graph containing all possibly true edges (e.g. neighbours within a sphere)
  - Train the network and perform inference





  - Select edges with a predicted probability of more than 0.5 to be true as connections

[1] S. Farrel et al, arxiv:1810.06111,
[2] 10.1051/epjconf/201715000003
[3] X. Ju et al, https://ml4physicalsciences.github.io/files/NeurIPS_ML4PS_2019_83.pdf
[4] Y. Wang, et al, arXiv:1801.07829. (DGCNN)

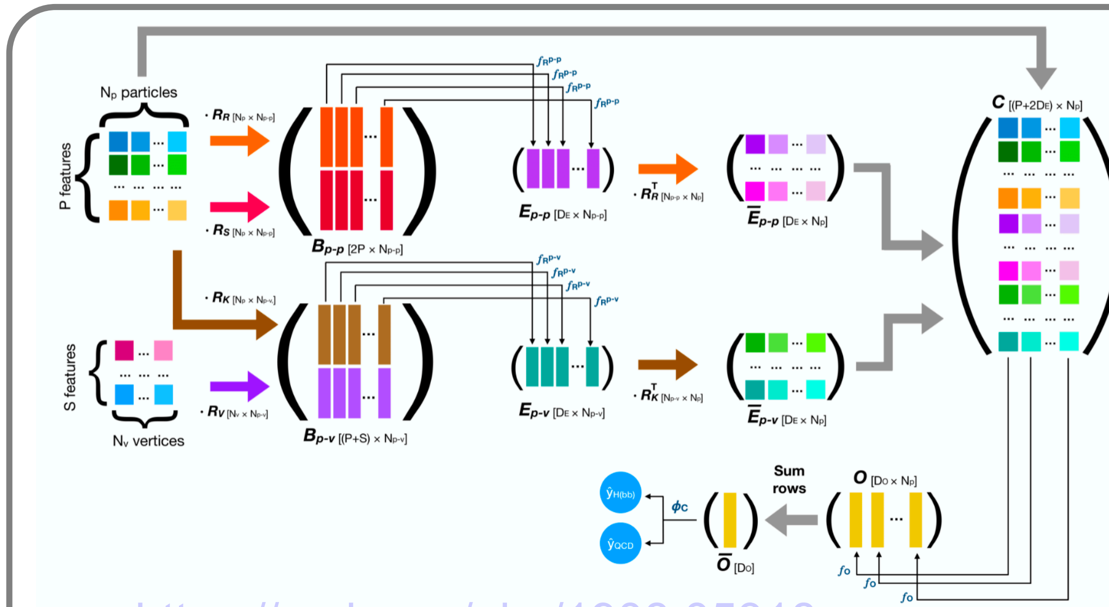


Use of graph models to perform reconstruction in the high granularity calorimeter.
Node clustering, Edge classification, node segmentation, ...

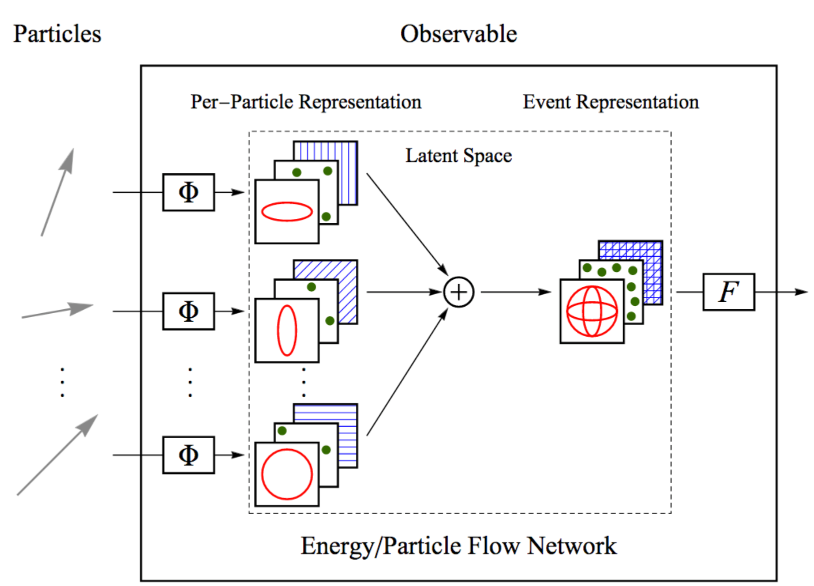
https://indico.cern.ch/event/847990/ *Slide J. Kieseler*

Particle-Cloud Jets

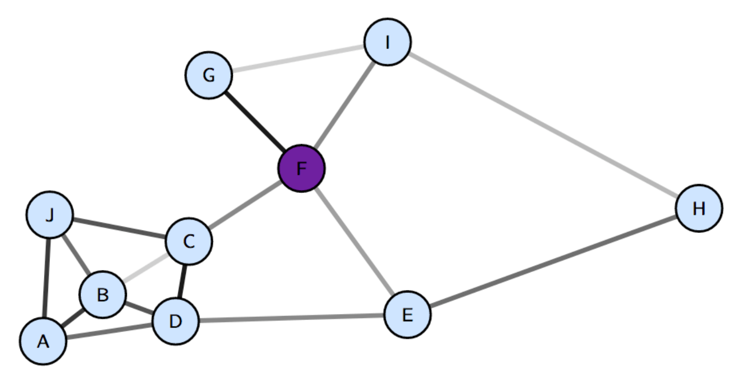
- Particle-flow jets are collection of reconstructed particles
- Graph / point-cloud representation is rather natural
- Connectivity of the graph depends on the model



https://arxiv.org/abs/1908.05318
https://arxiv.org/abs/1909.12285

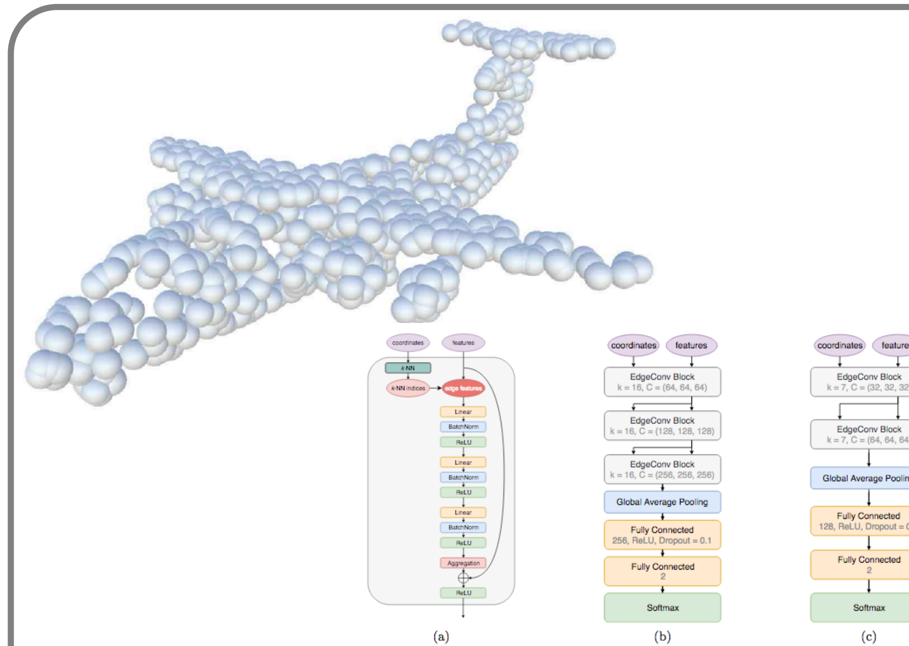


https://arxiv.org/abs/1810.05165



$$\tilde{m}_j^t = f(h_j^{t-1})$$
$$m_{j\to i}^t = \sigma(A_{ij}\tilde{m}_j^t)$$
$$h_i^t = \mathrm{GRU}(h_i^{t-1}, \Sigma_j m_{j\to i}^t)$$

https://tinyurl.com/y44ggt5b



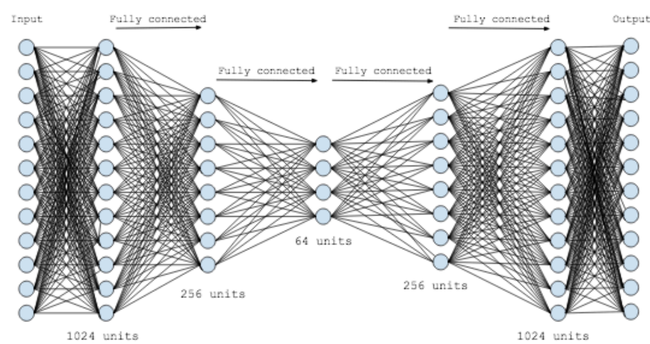
https://arxiv.org/abs/1902.08570

Monitoring R&D

Data Quality Monitoring

Chosen Autoencoder Architecture

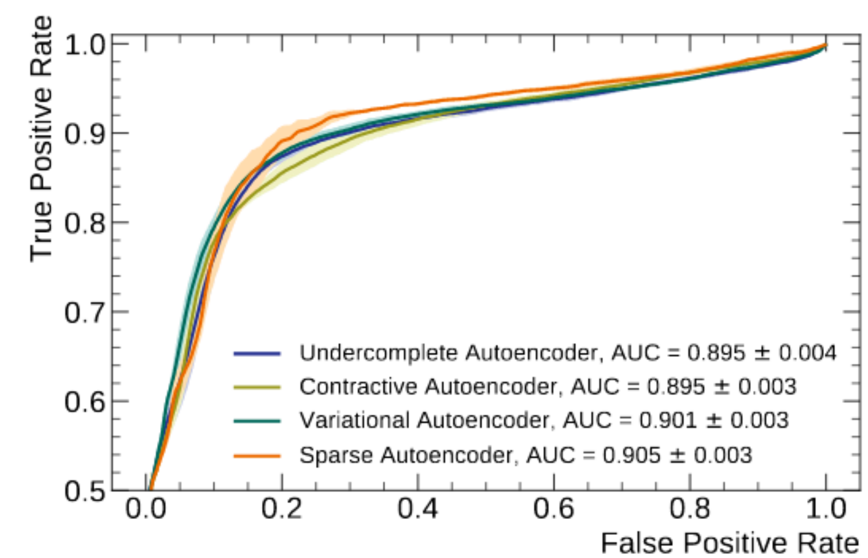
- Trained with *Keras/TensorFlow*.
- *Adam optimizer* (lr= 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.7$) and *early stopping* (patience = 32 epochs).
- Trained to minimize *mean squared error* between input vector and the output one: $\frac{1}{n}\sum_{i=0}^{n}(X_i - \hat{X}_i)^2$.
- Activations: *parametric rectified linear units*.



Proposed autoencoder architecture

Semi-supervised AD: Results

- Test set chosen gives representative values for ROC AUC.
- Anomaly score is the average reconstruction error squared over 100 worst reconstructed features $TOP100 = \frac{1}{100}\sum_{i=1}^{100} sorted(X_i - \hat{X}_i)^2$.
  - Contributions from well behaving features are irrelevant.



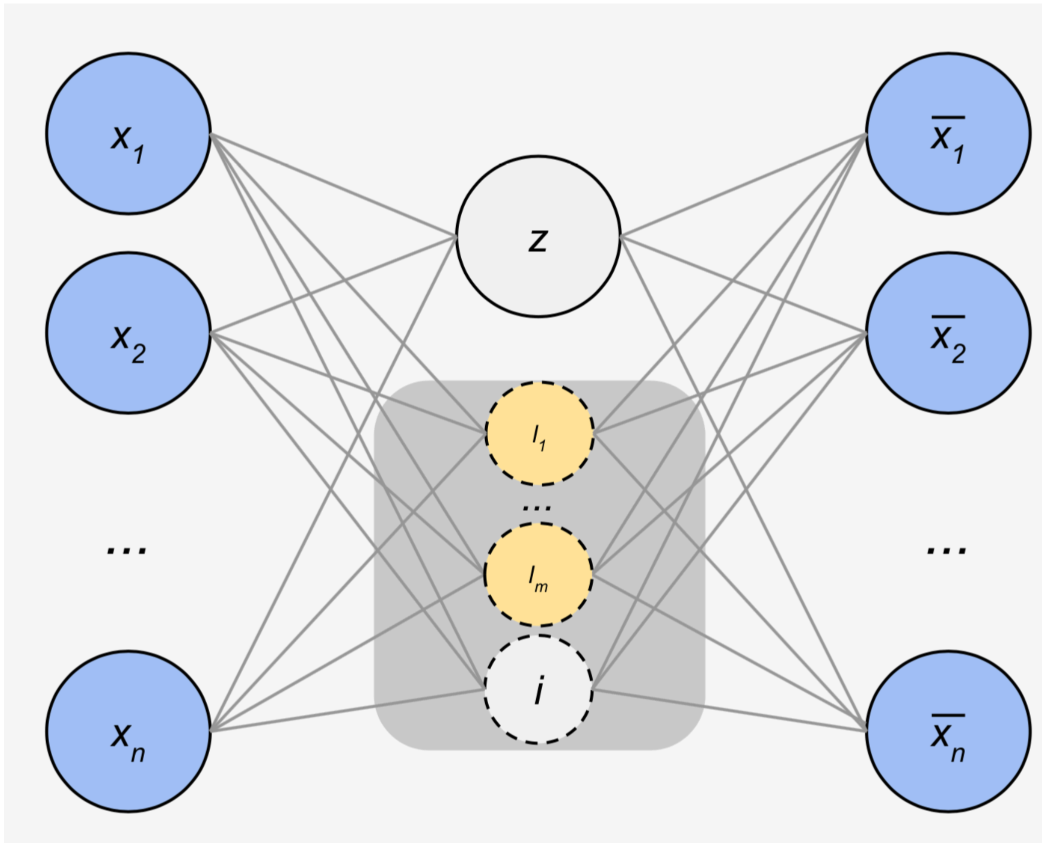
Slide A. Pol

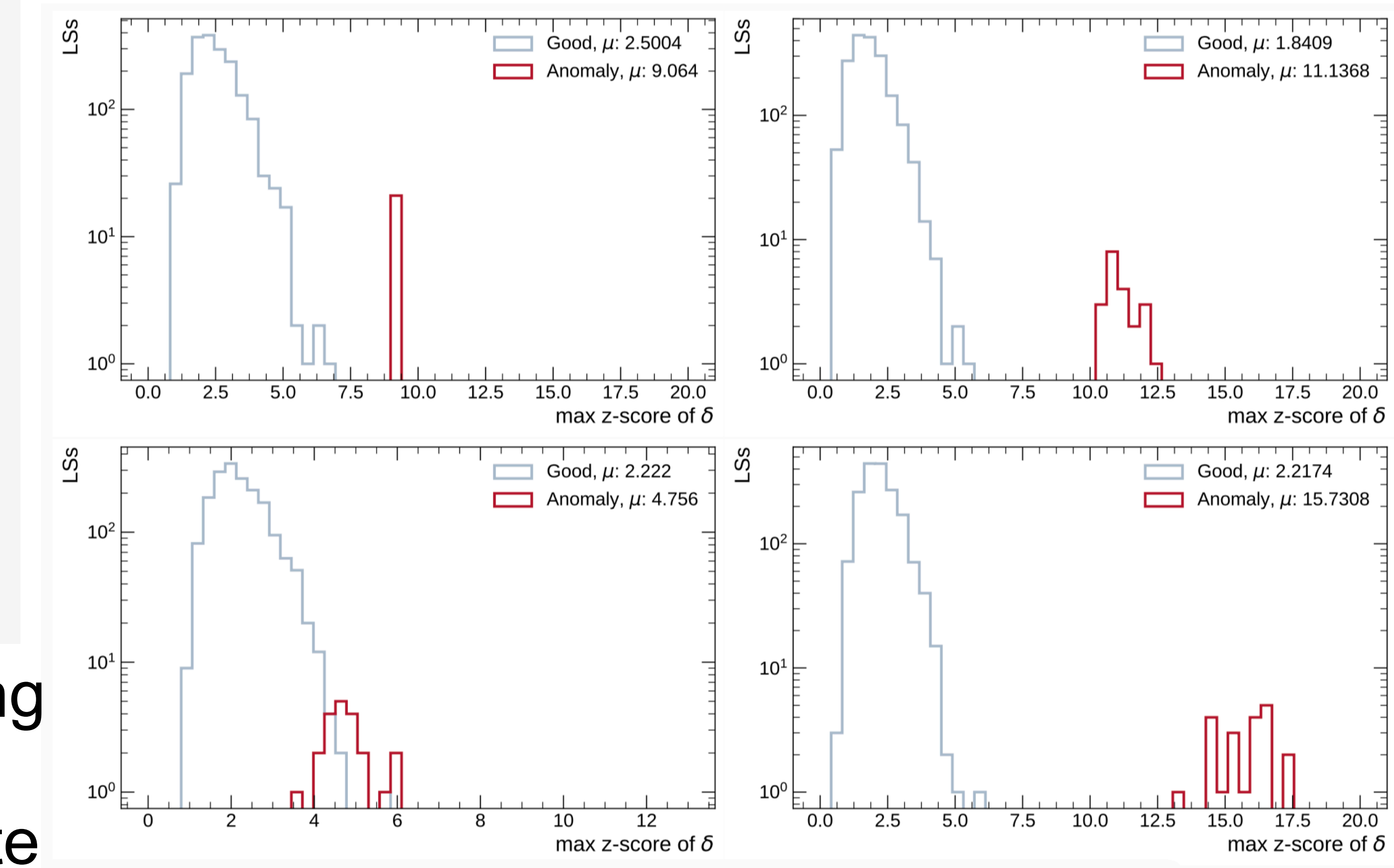
Performance of different AEs

Catch anomalies in data taking using auto-encoder of hundreds of features

https://indico.cern.ch/event/708041/contributions/3276189/

Trigger Rate Prediction



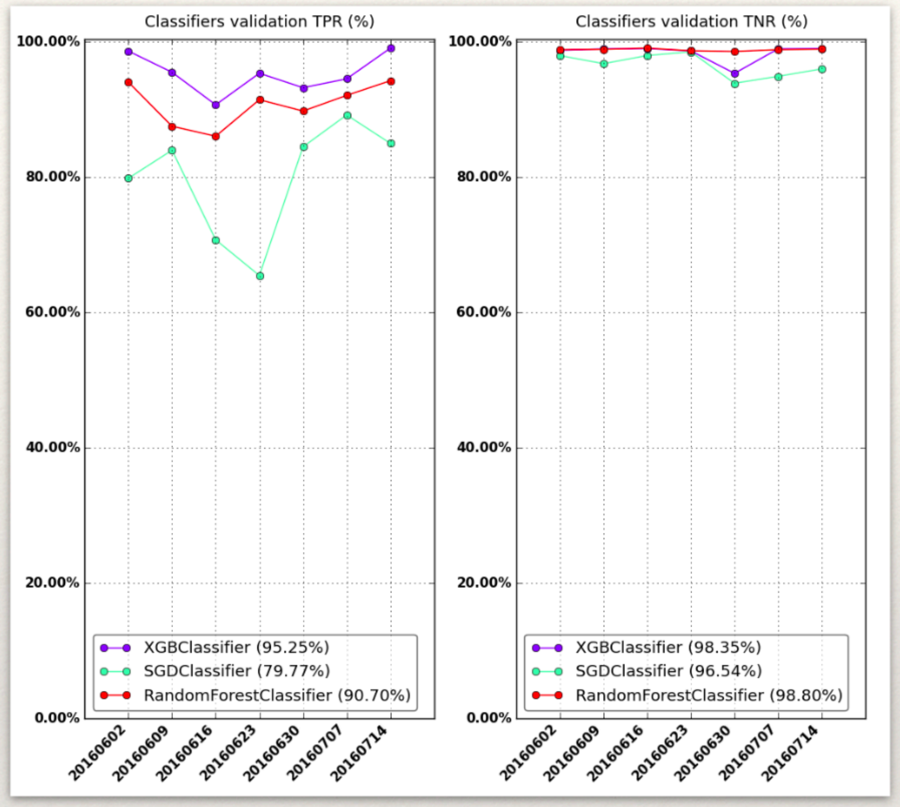


Detect deviation of trigger rate using variational auto-encoder on high level trigger rate, and L1 trigger rate in latent space

https://indico.cern.ch/event/708041/contributions/3276197/

Operation
R&D

Data Popularity



Performed studies with various classifiers:

RandomForest, SGD, XGboost

Found similar results with

SparkML

TPR=TP/(TP+FN)
TNR=TN/(TN+FP)

MINIAOD were introduced

in mid 2014

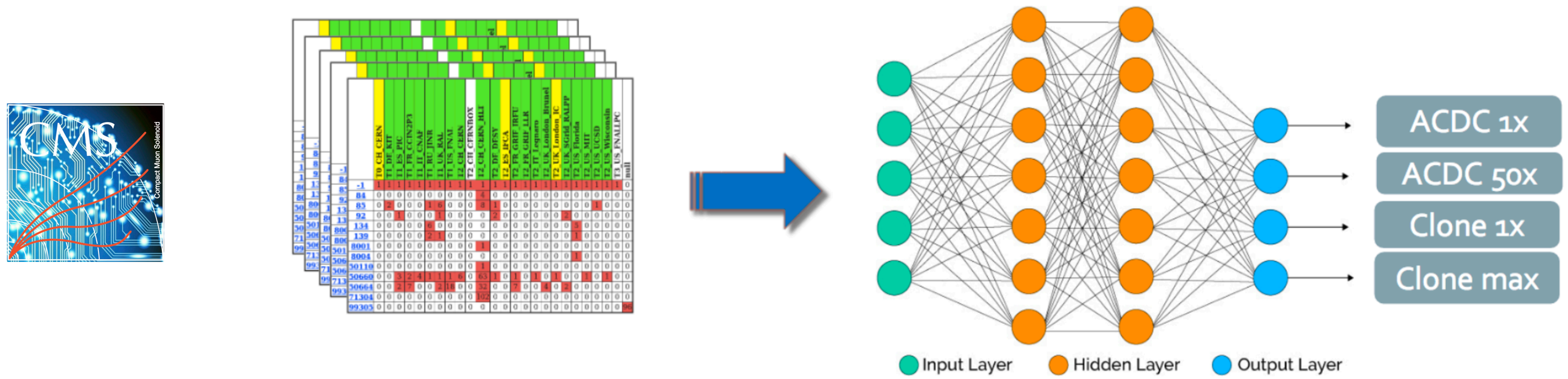
https://arxiv.org/abs/1602.07226

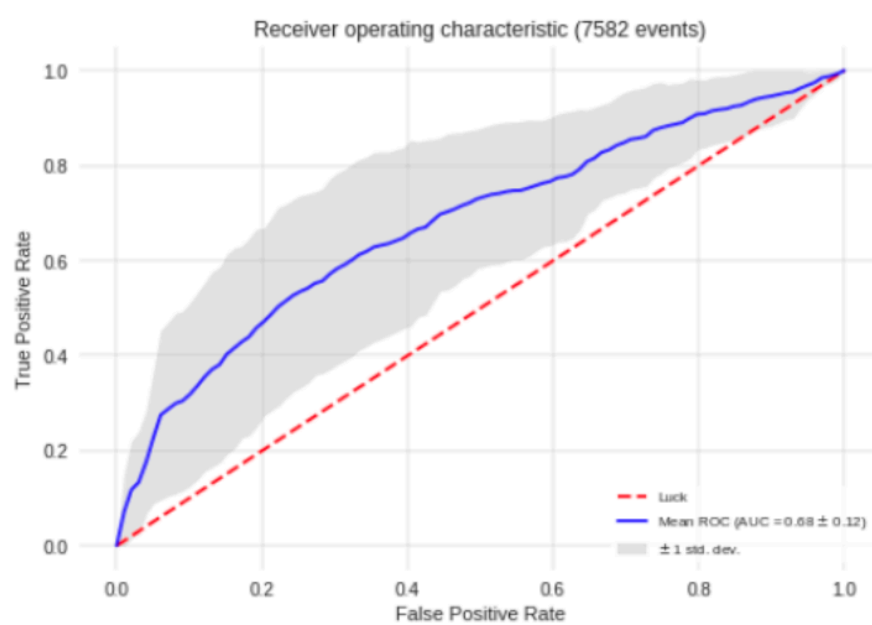
Data tier	TPR	TNR	FP	FN
AOD	0.97 ± 0.05	0.99 ± 0.02	0.005 ± 0.011	0.015 ± 0.029
AODSIM	0.93 ± 0.13	0.99 ± 0.02	0.008 ± 0.016	0.021 ± 0.045
MINIAOD	0.11 ± 0.32	0.99 ± 0.02	0.014 ± 0.026	0.001 ± 0.007
MINIAODSIM	0.49 ± 0.48	0.99 ± 0.02	0.009 ± 0.016	0.007 ± 0.031
USER	0.93 ± 0.15	0.98 ± 0.02	0.014 ± 0.021	0.011 ± 0.023

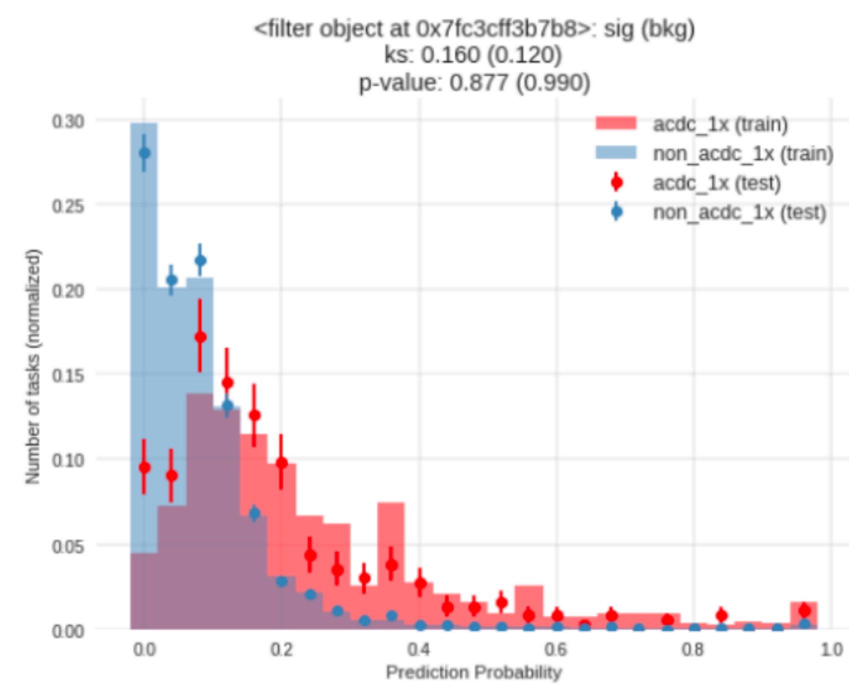
Slide V. Kuznetsov

R&D on predicting popularity of analysis datasets, in a view to a more efficient data placement.

Predicting Operator's Action



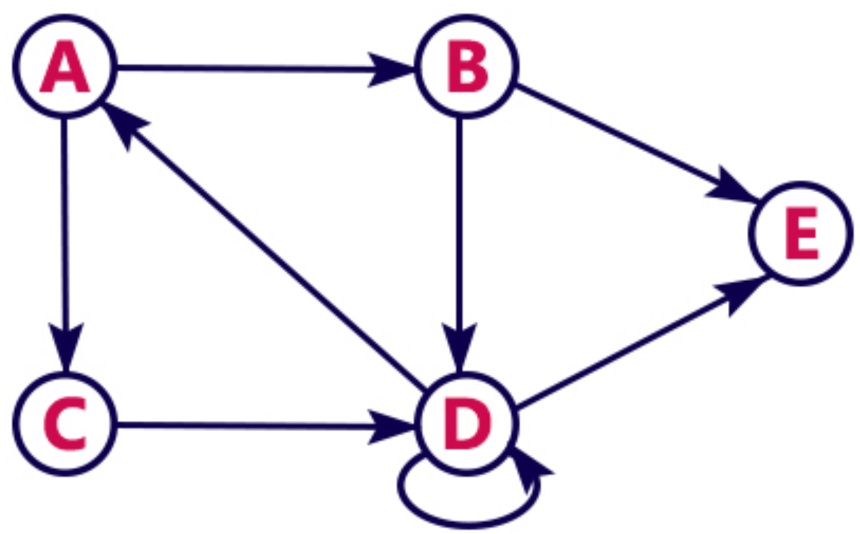




Challenging task of predicting the operator's action from the information they are provided with.

https://indico.cern.ch/event/587955/contributions/2937424/

Forewords on Graph



$$
\begin{array}{c|c}
 & \begin{matrix} A & B & C & D & E \end{matrix} \\
\begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} &
\begin{bmatrix}
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
$$

http://btechsmartclass.com/data_structures/graph-representations.html

A graph is composed of

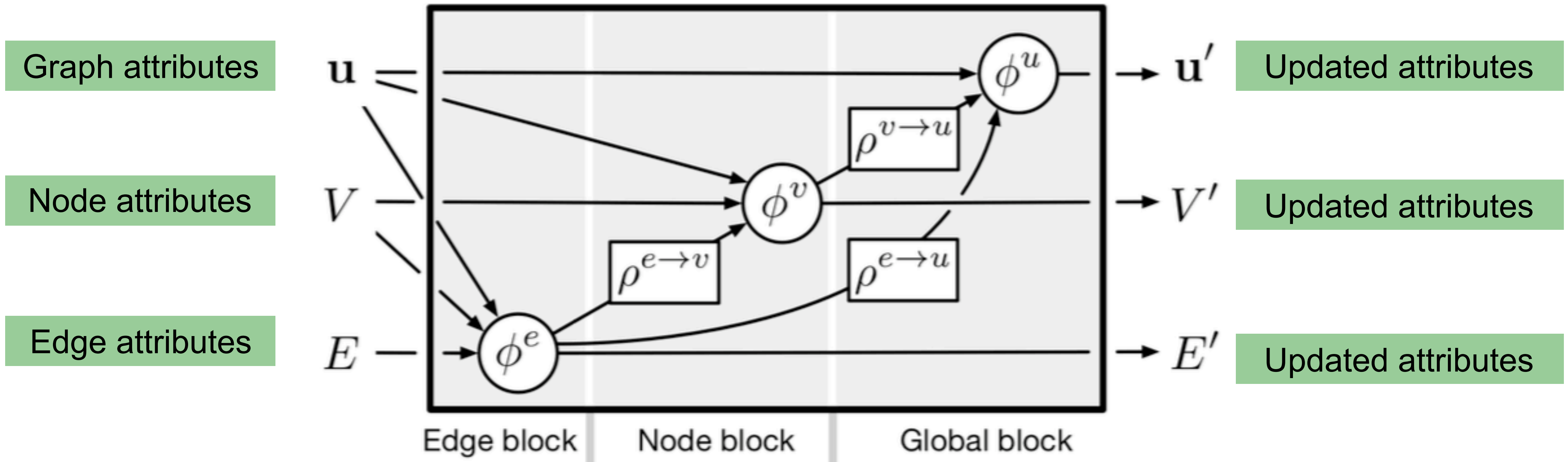
- **Nodes** that can be represented as a vector.
- **Edges** that can be represented with the adjacency matrix.

→ Flowing of information using matrix operations.

→ With machine learning on graphs, edges and nodes might acquire internal representations.

Graph Neural Networks Formalism

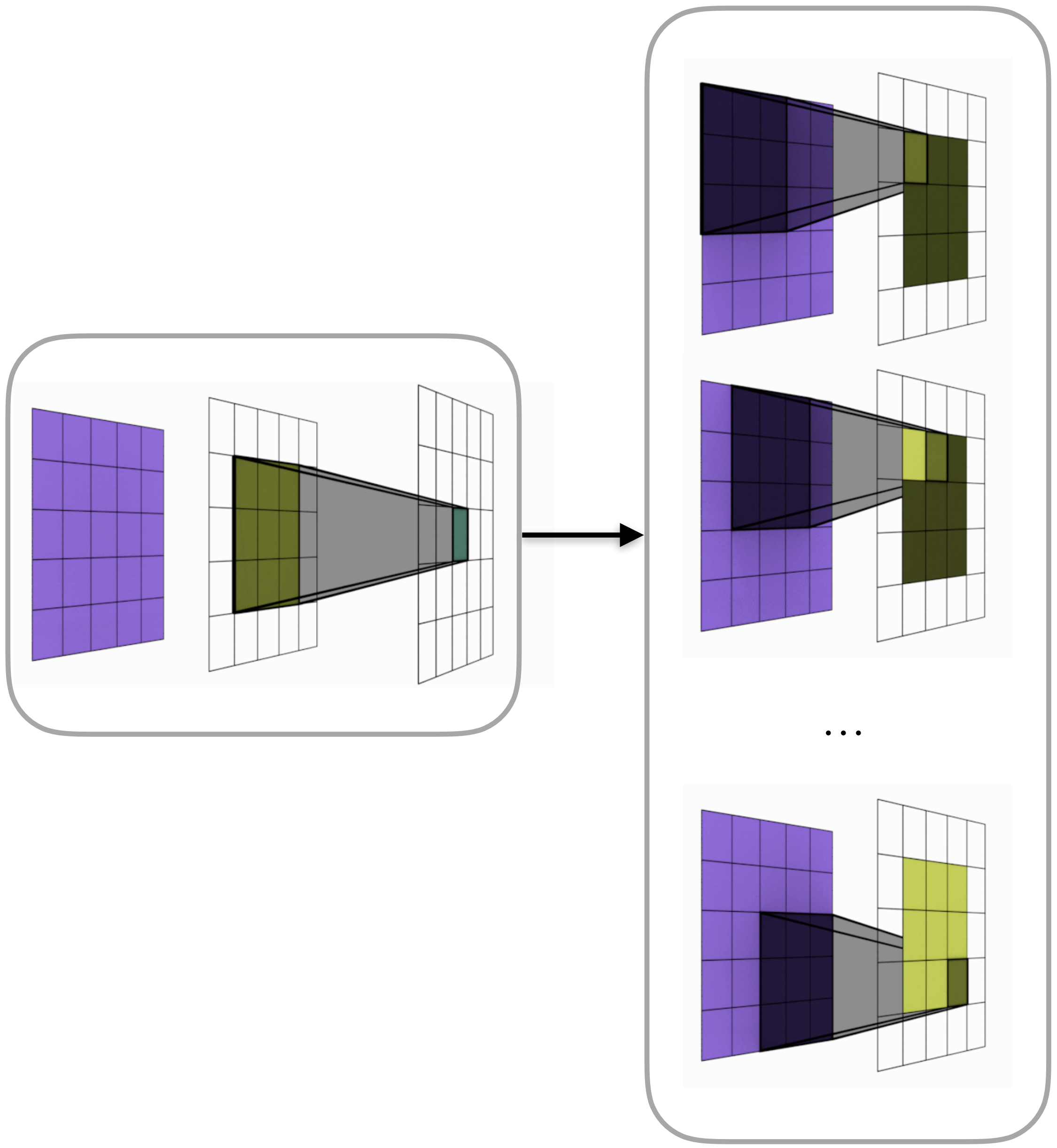
https://arxiv.org/abs/1806.01261



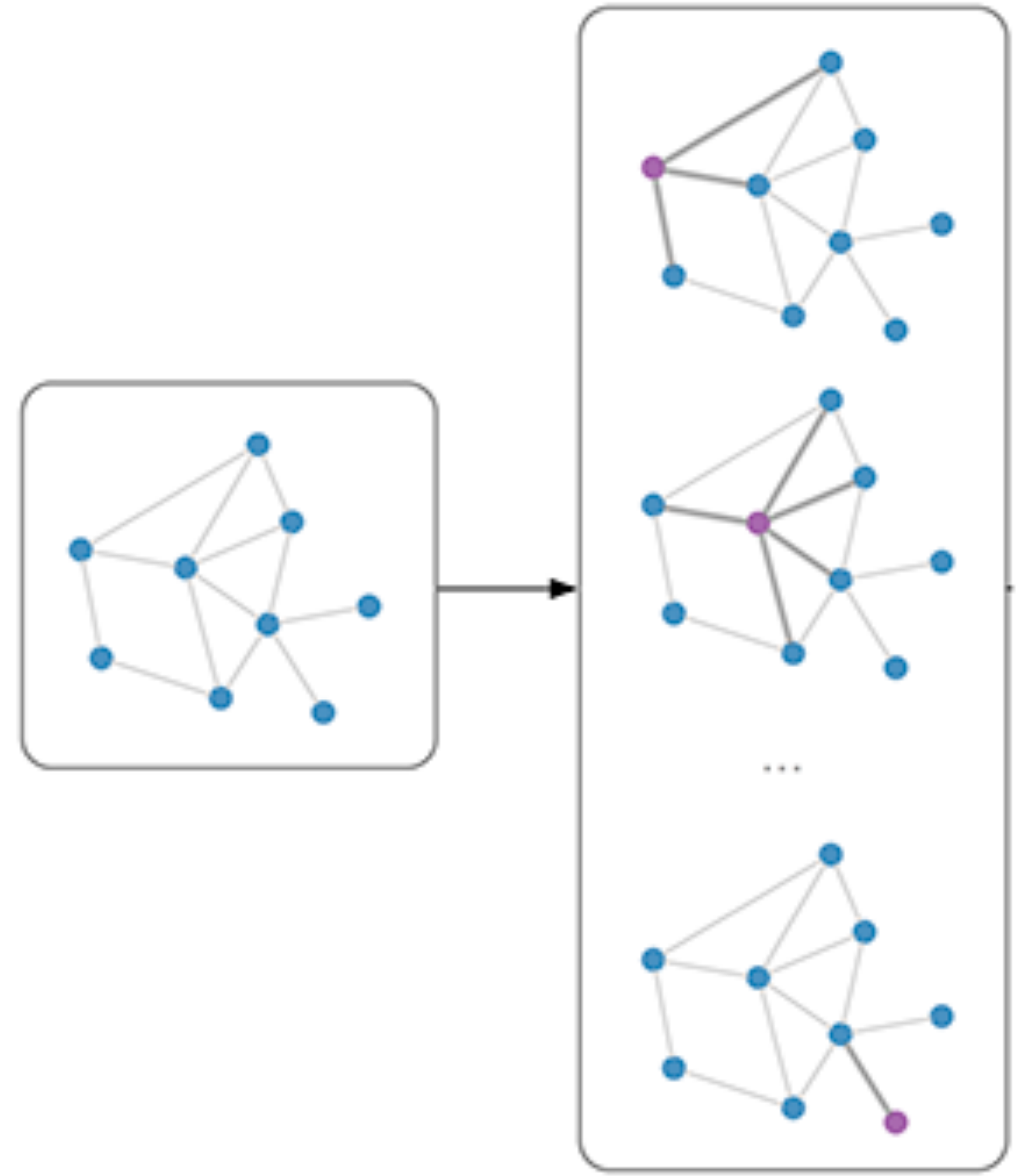
Lots of possibilities to operate on a graph.
Most available architectures can be expressed with Φ and ρ.

Readily software:
https://github.com/deepmind/graph_nets
https://github.com/rusty1s/pytorch_geometric

Graph Convolution



https://imgur.com/gallery/AIFHqe9



https://tkipf.github.io/graph-convolutional-networks/