Deep Learning at Colliders

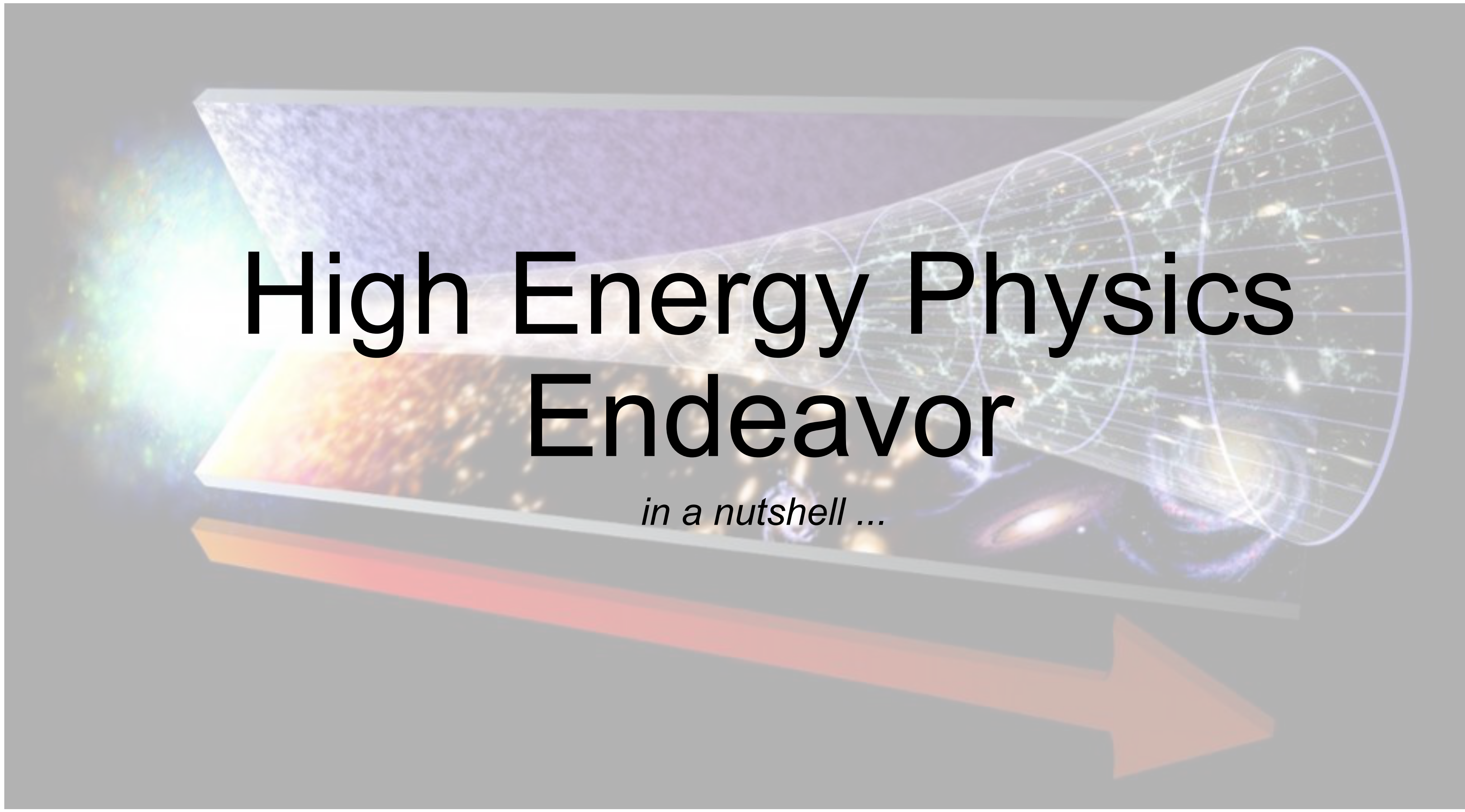
IN2P3 School Of Statistics 2021
18-29 Jan, 2021

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

School of Statistics

Outline

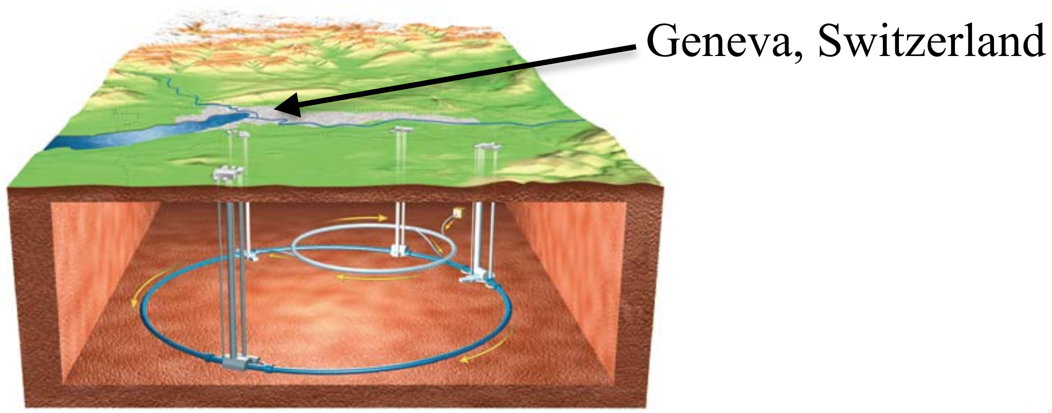
I. Physics at the Large Hadron Collider
II. The case for Deep Learning
III. Deep Learning Applications in HEP
IV. Collider-Specific AI

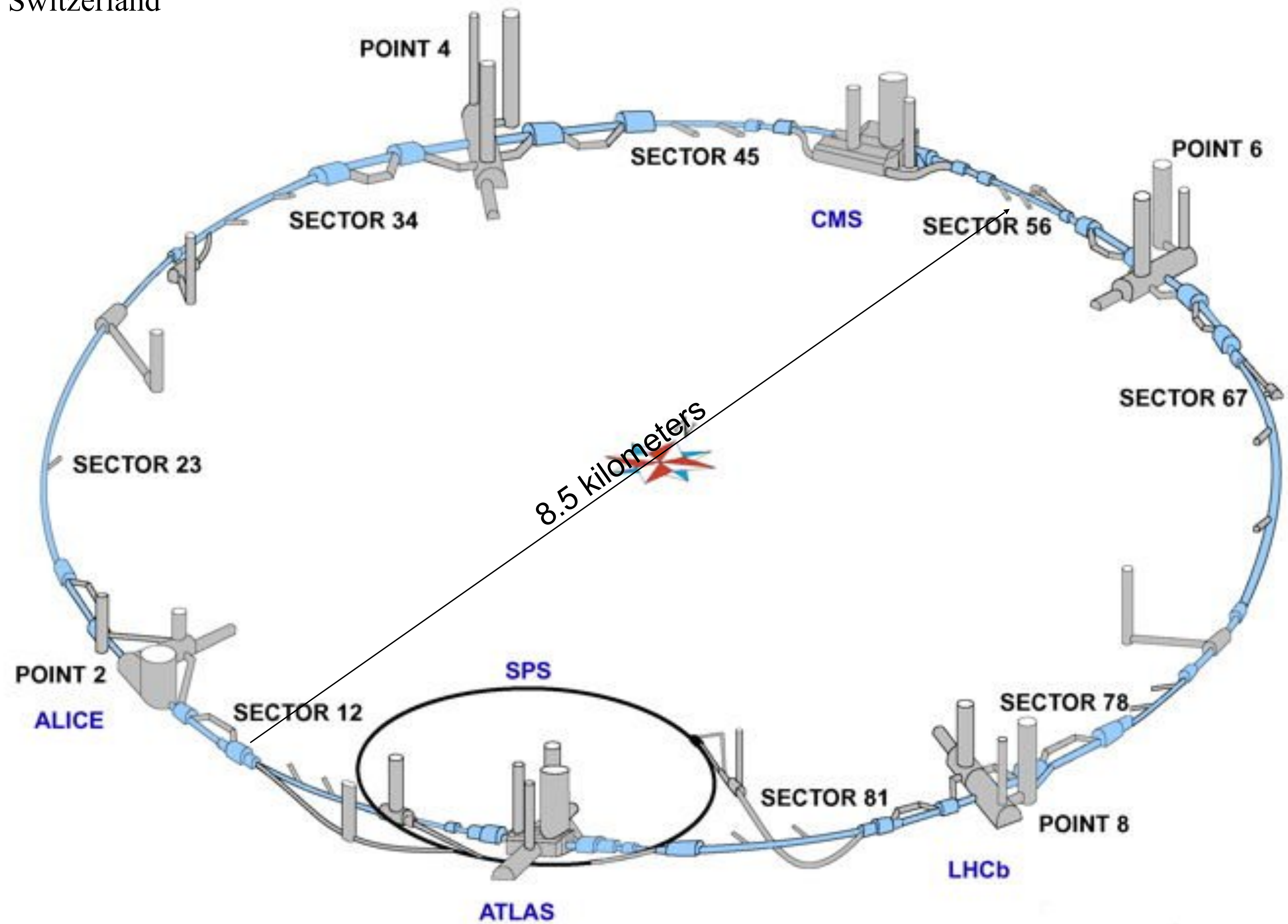


High Energy Physics Endeavor

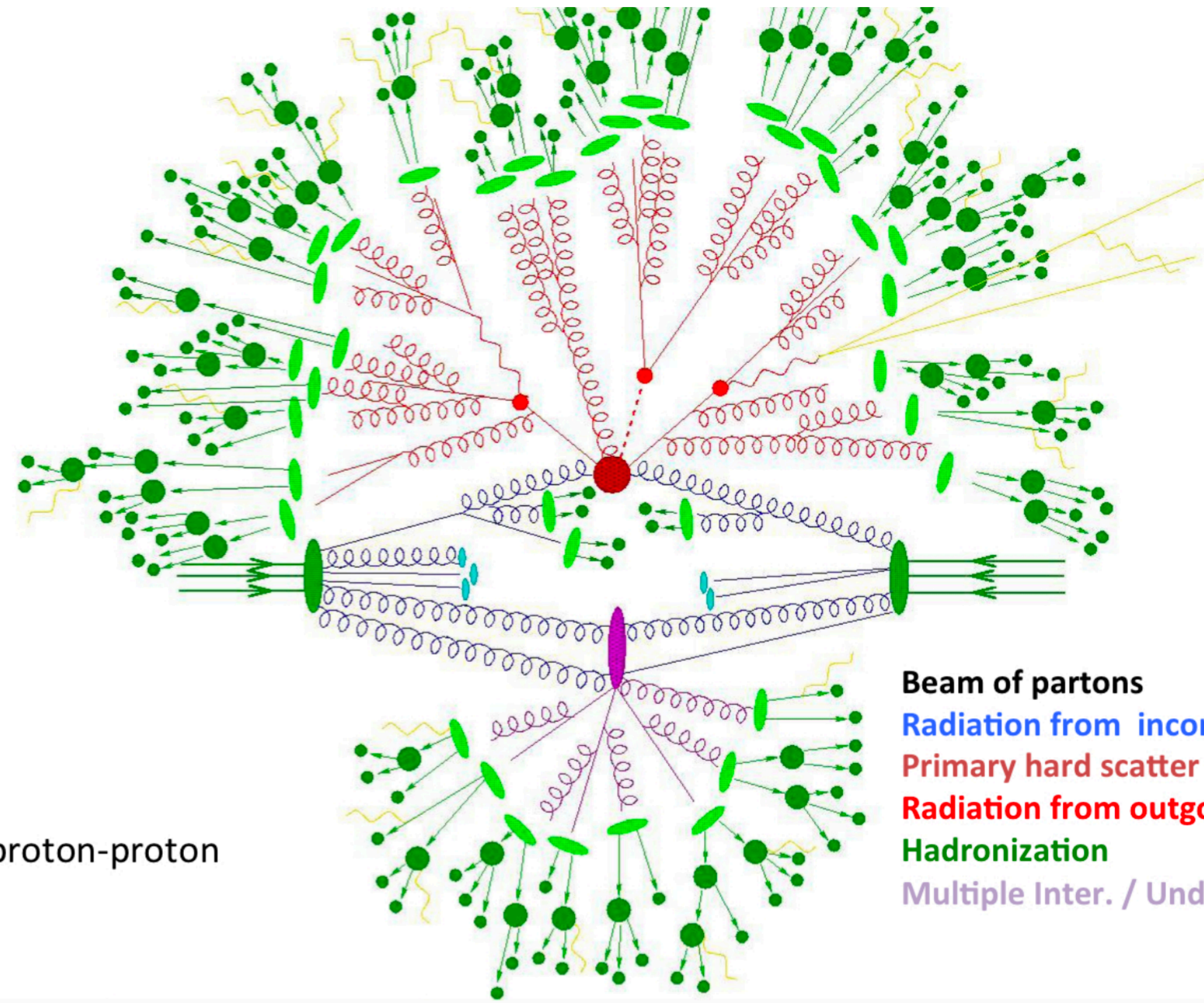
in a nutshell ...

The Large Hadron Collider





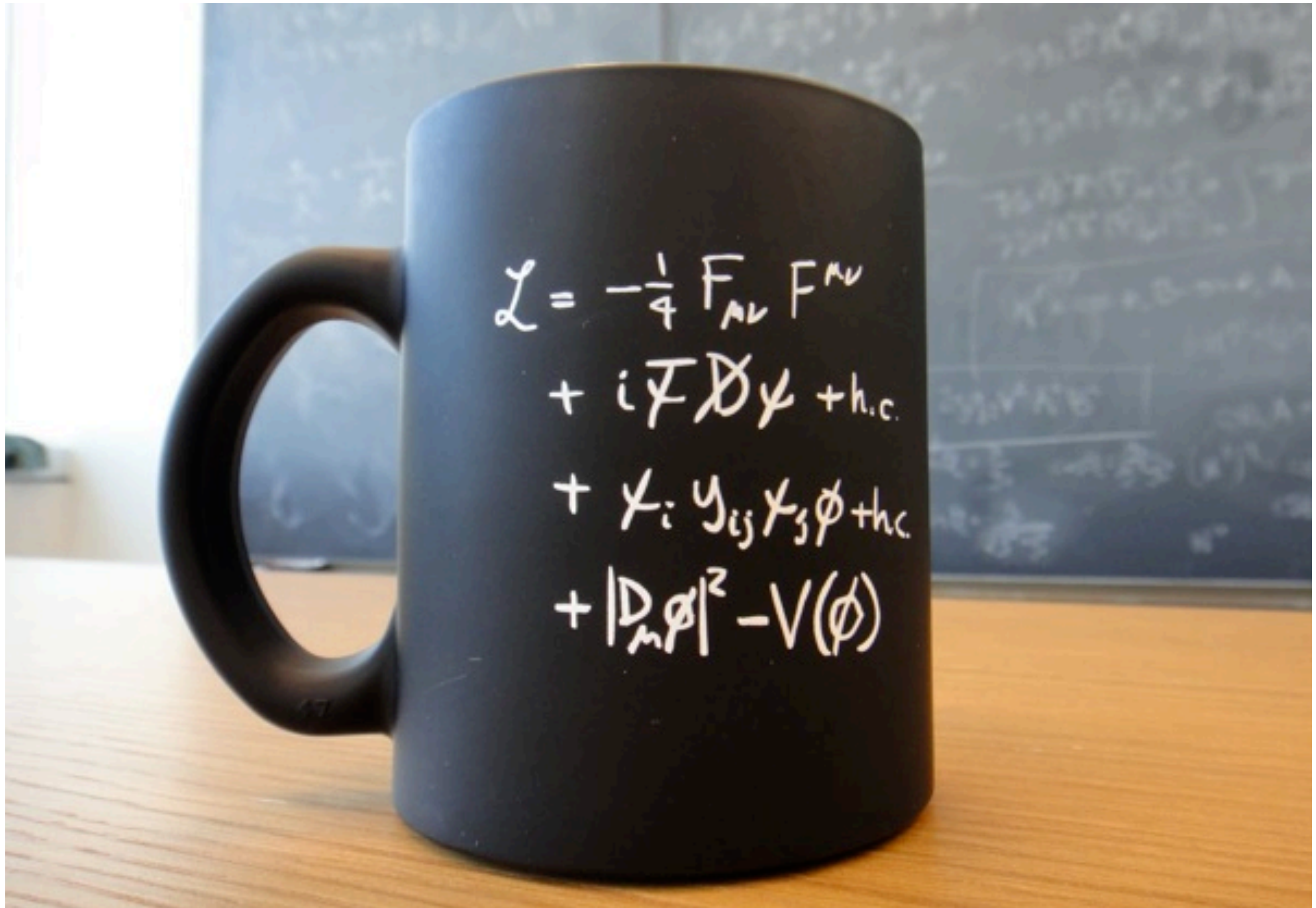
Colliding Hadrons



Typical proton-proton collision

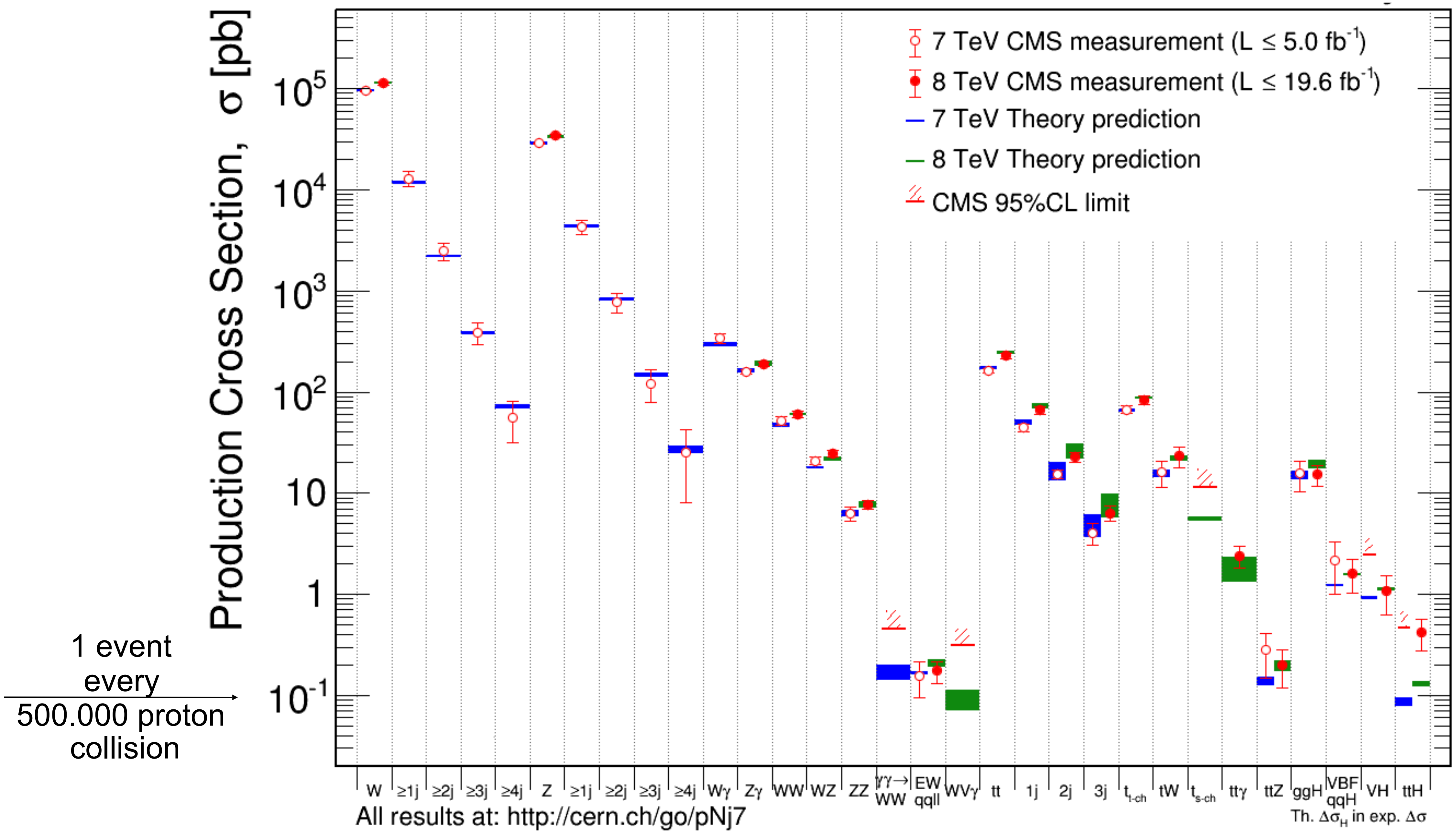
Probing fundamental laws of physics as large spectrum of particles (known and unknown) can be produced

The Standard Model



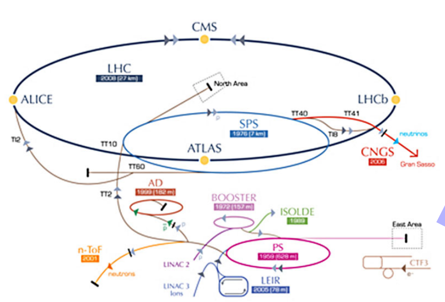
Well demonstrated effective model
We can predict most of the observations
We can use a large amount of simulation

Size Of The Challenge

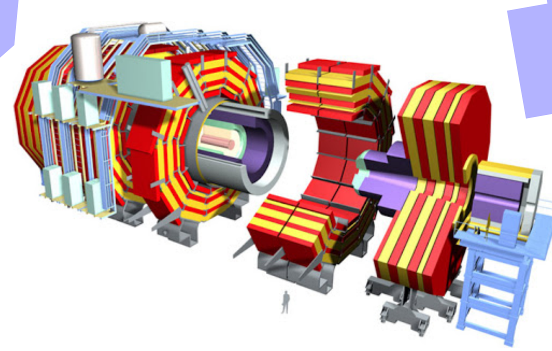


Low probability of producing exotic and interesting signals.
Observe rare events from a large amount of data.

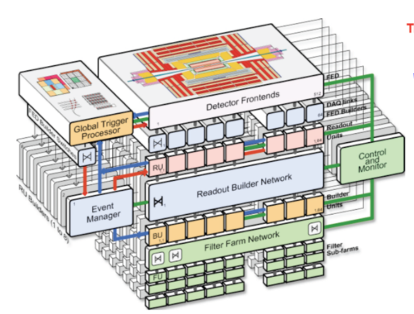
HEP Data Pipeline



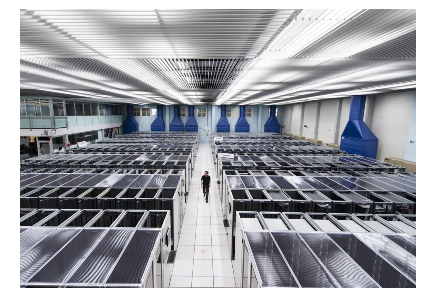
Large Hadron Collider
40 MHz of collision



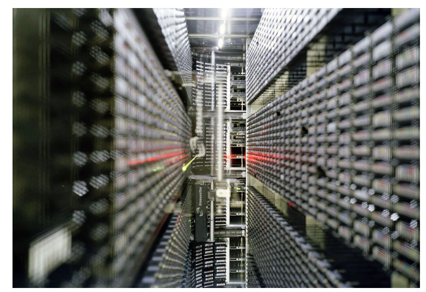
CMS Detector
1PB/s



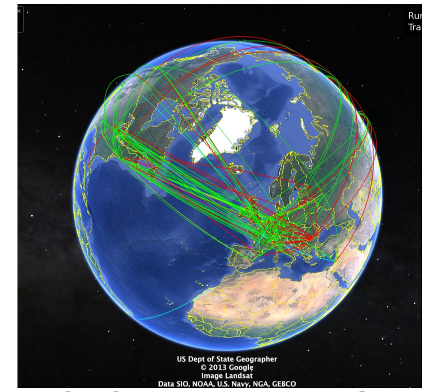
CMS L1 & High-Level Triggers
50k cores, 1kHz



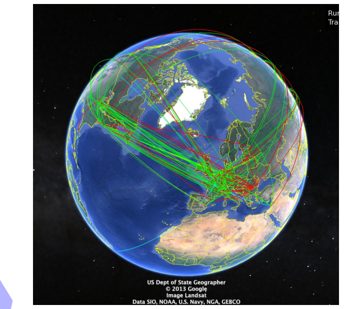
CERN Tier-0
Computing Center
20k cores



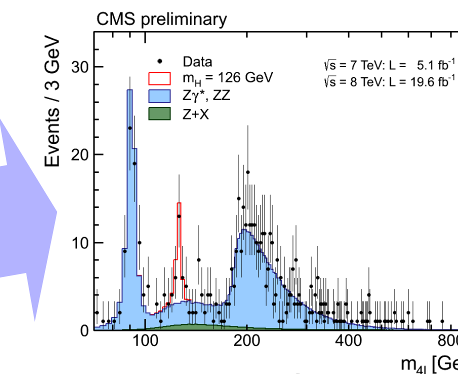
CERN Tier-0/Tier-1
Tape Storage
200PB total



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



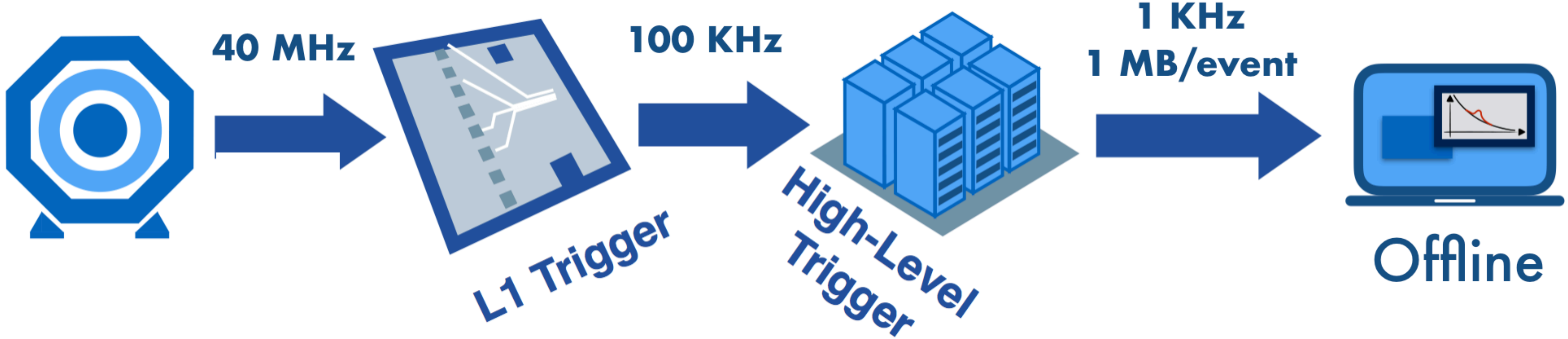
LHC Grid
Remote Access
to 100PB of data



Rare Signal
Measurement
~1 out of 10^6

Event Triggering

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

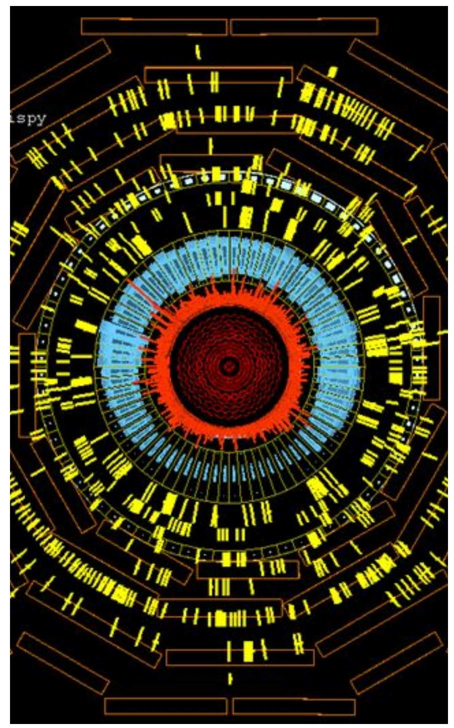


Reconstruction(s) of the event under limited latency.
Better resolution help lowering background trigger rates.
Approximate deep learning surrogates can help.

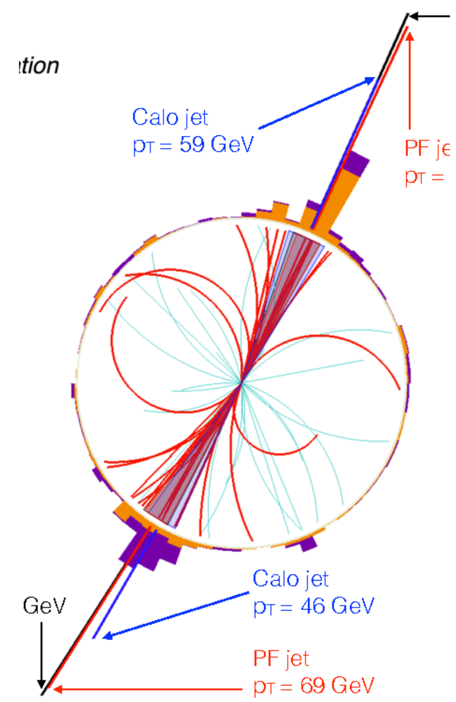
Reconstructing Collisions

Detector Data

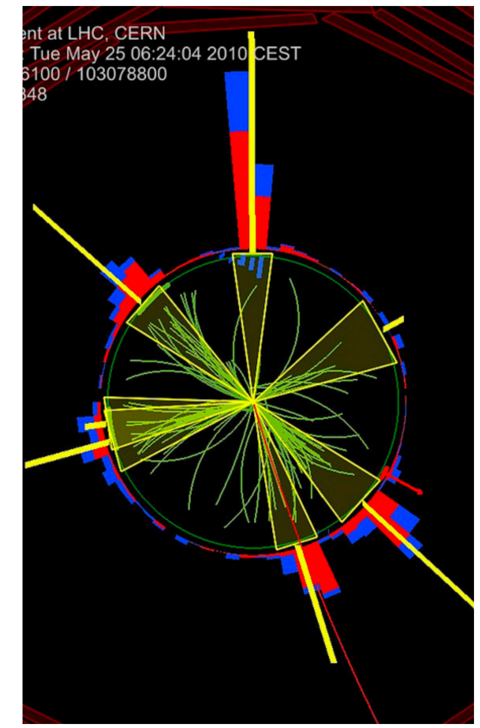
Local reconstruction



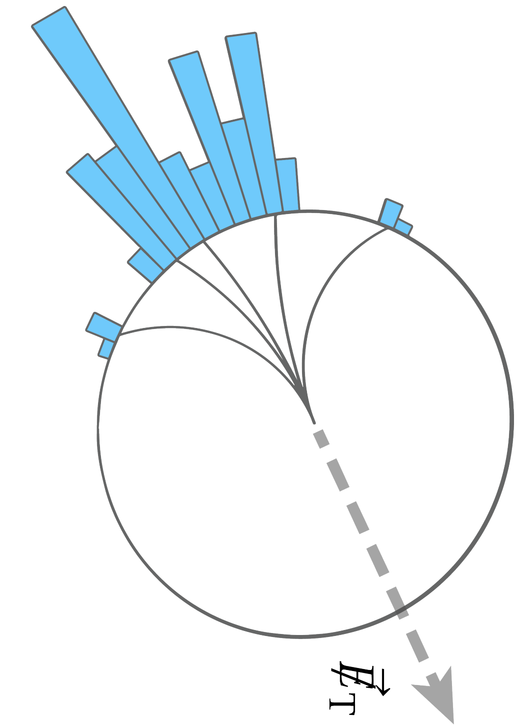
Particle representation



Jet Clustering



High level features



Event Processing

Dimensionality reduction

Globalization of information

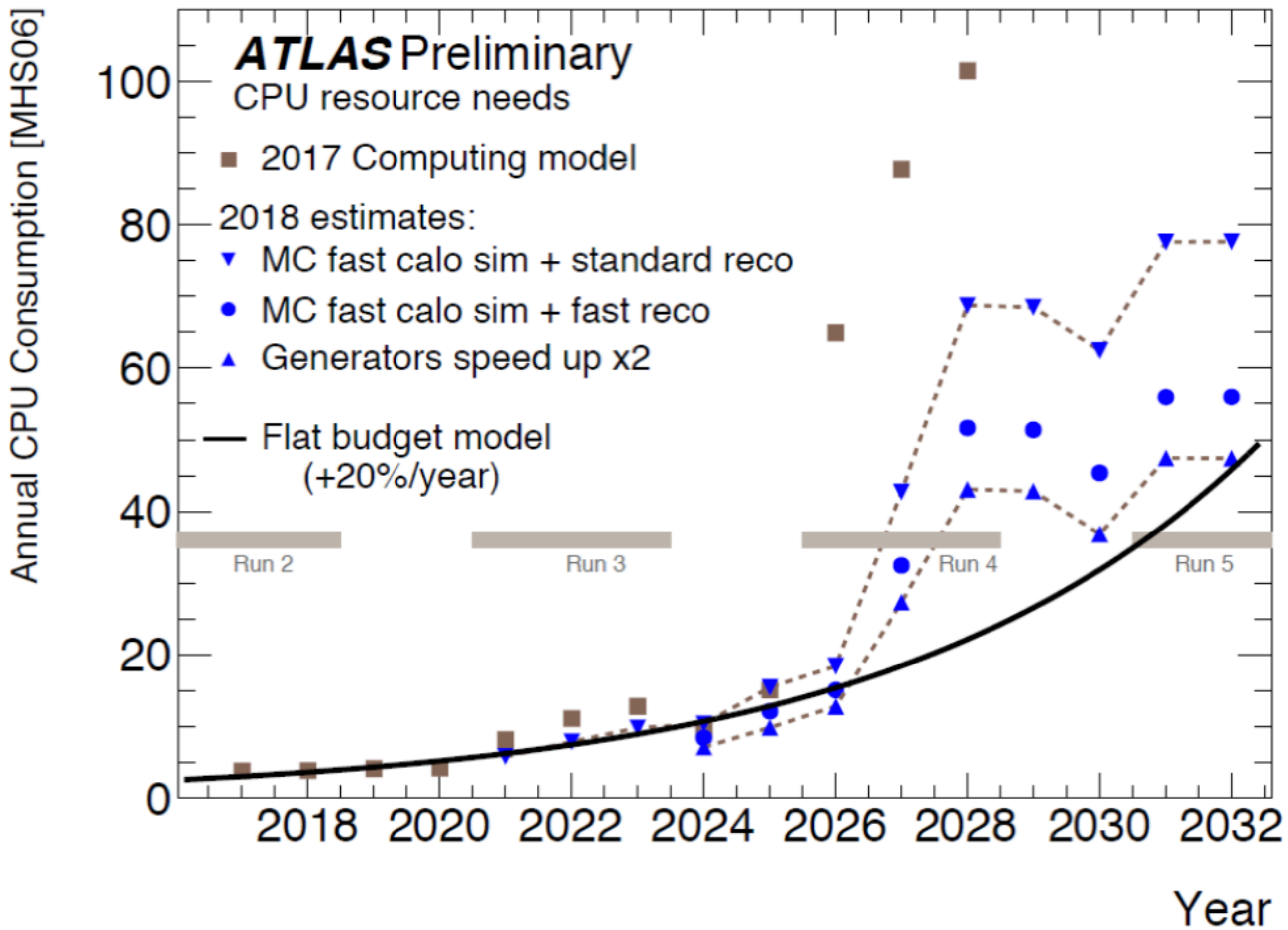
From digital signal, to local hits, to a sequence of objects, and high-level features.
Complex and computing intensive tasks.

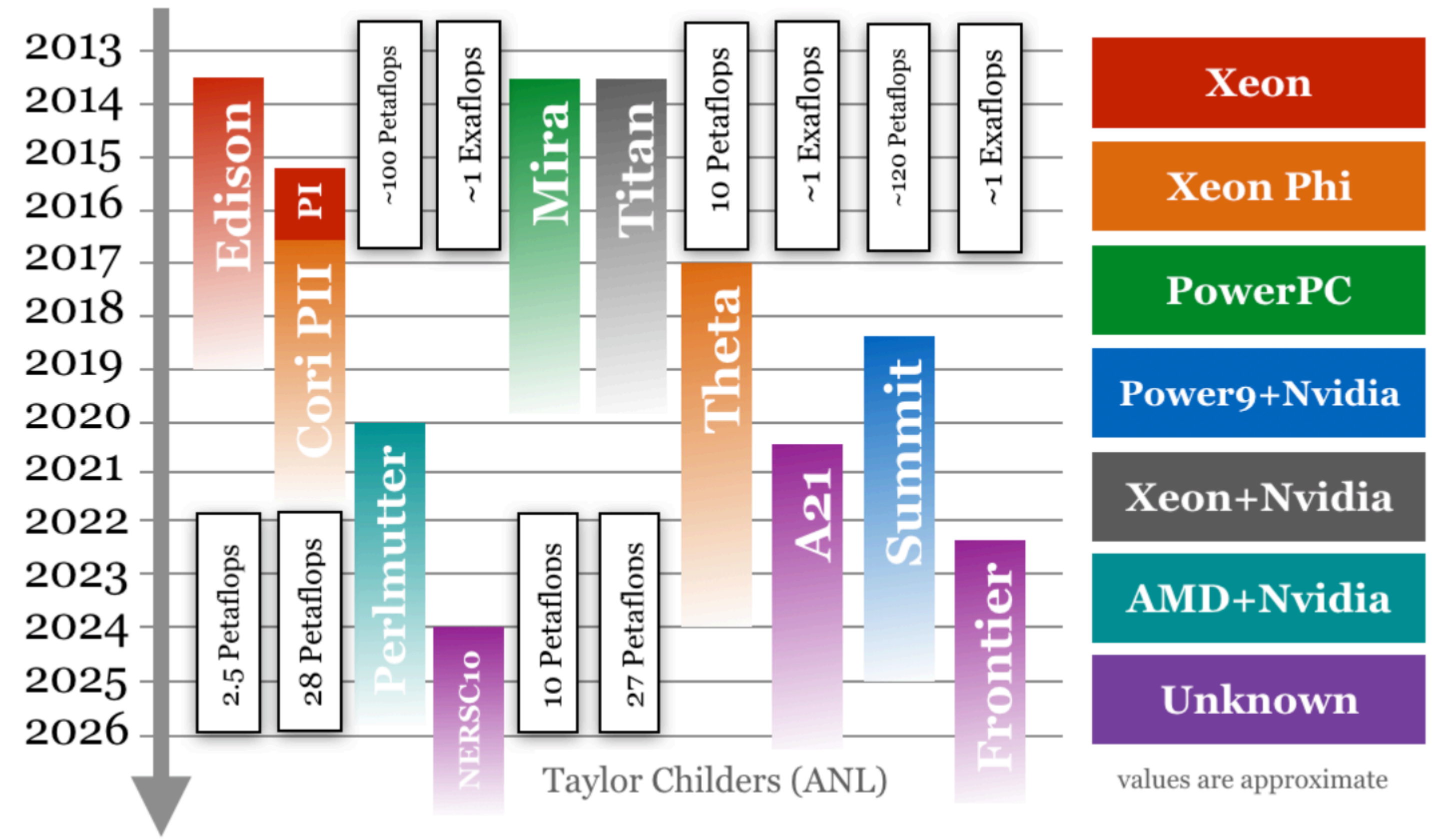
Simulating Collisions

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/

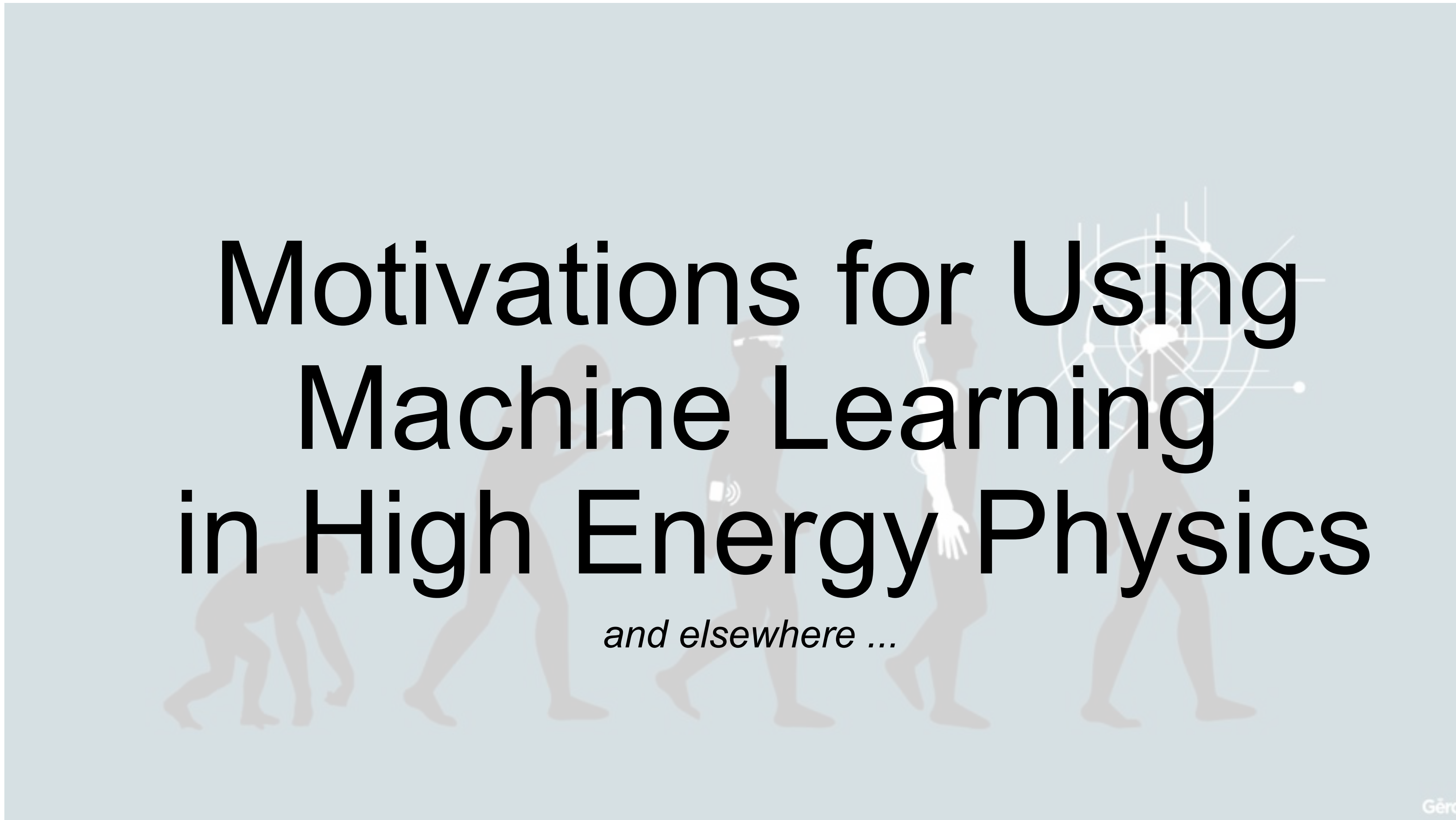
Ever growing needs for computing resource
Slowdown of classical architecture
Growth of GPU architecture

Take home message :

Measure rare and exotic processes from orders of magnitude larger backgrounds.

The Standard Model predicts with precision what to expect from many processes.

Reconstruct, identify and reject large amount of event within resource constraints.



Motivations for Using Machine Learning in High Energy Physics

and elsewhere ...

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**

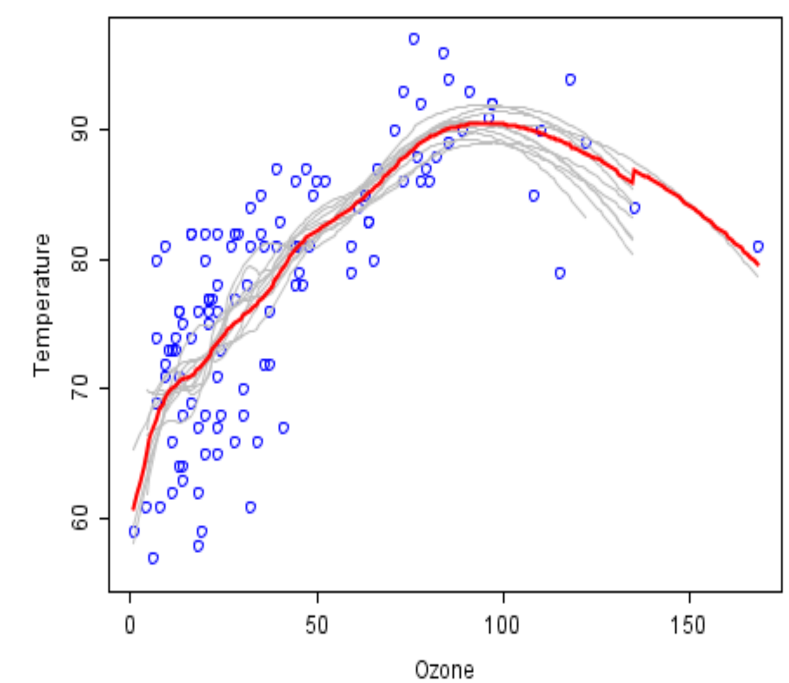
Yann Le cun, CERN, 2016

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

$$dataset \equiv \{(x_i, y_i)\}_i$$

$$find \;\; function \;\; f \;\; s.t. \;\; f(x_i) = y_i$$



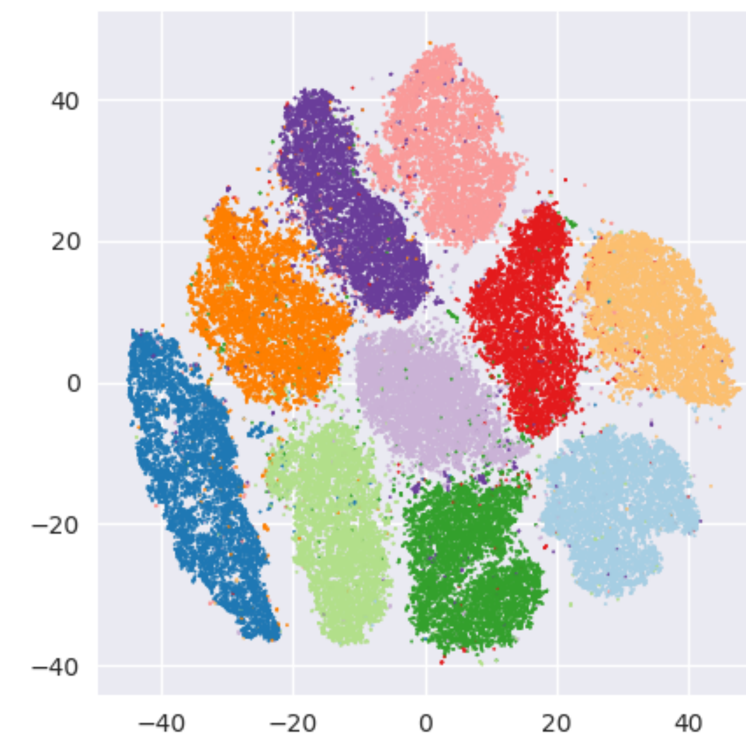
- Finite set of target values :
  - → **Classification**
- Target is a continuous variable :
  - → **Regression**

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**



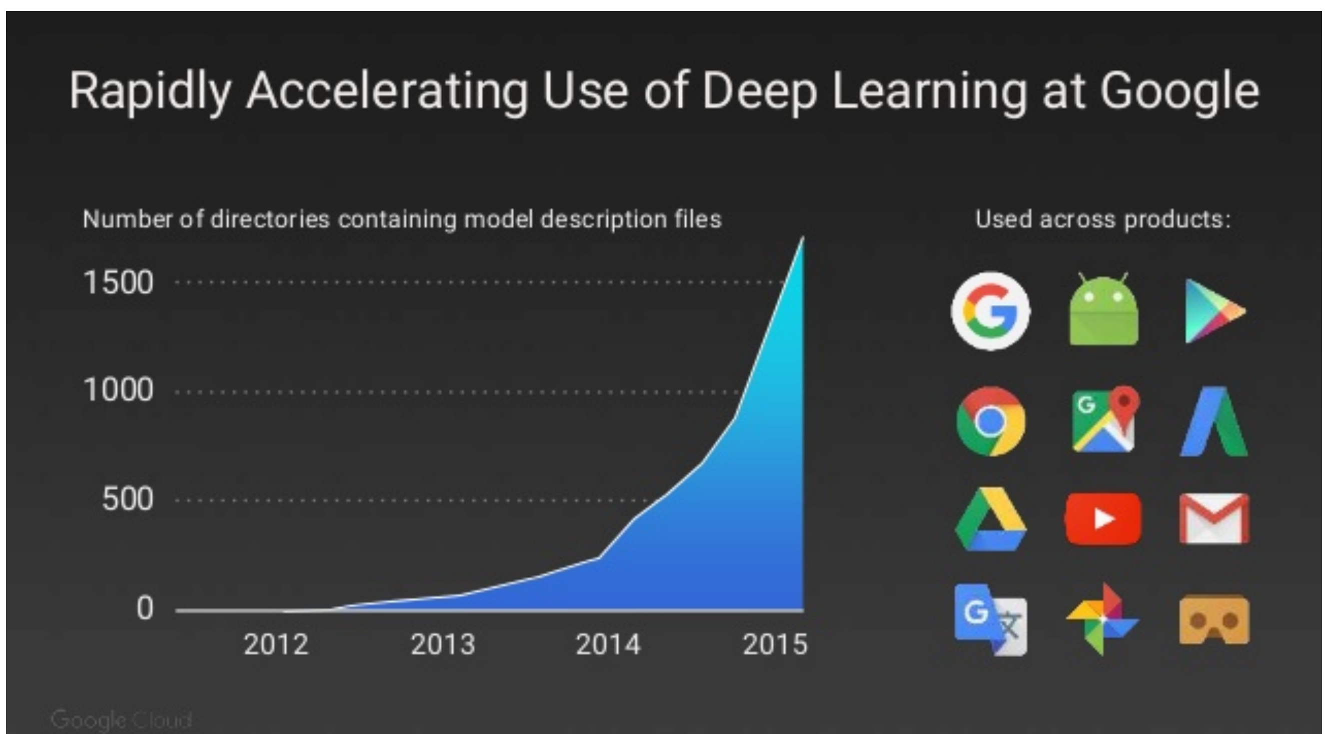
Reinforcement Learning

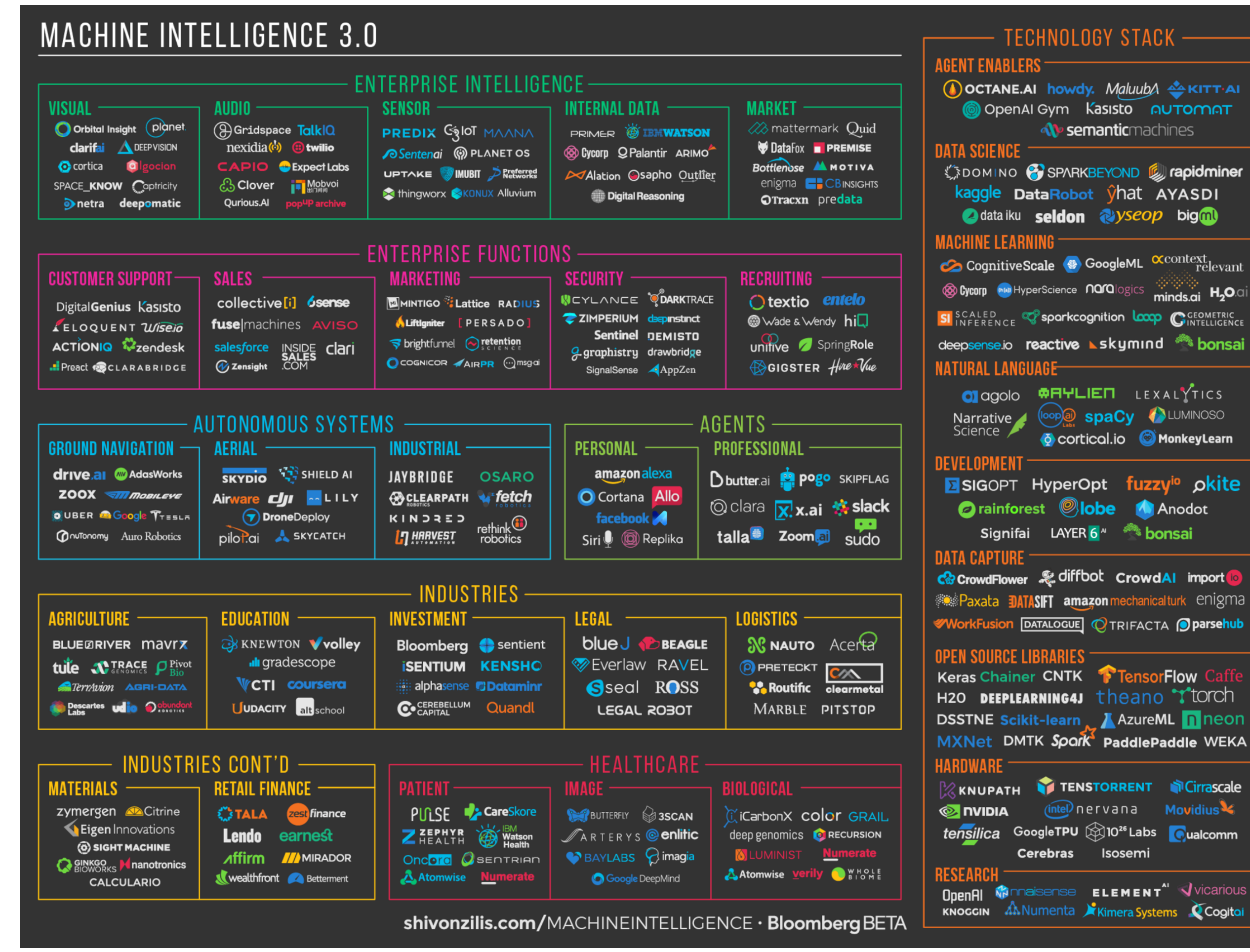
- Given an **environment** with multiple states, given a reward upon action being taken over a state
- Find an **action policy to drive** the environment toward maximum cumulative reward

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

Machine Learning in Industry

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

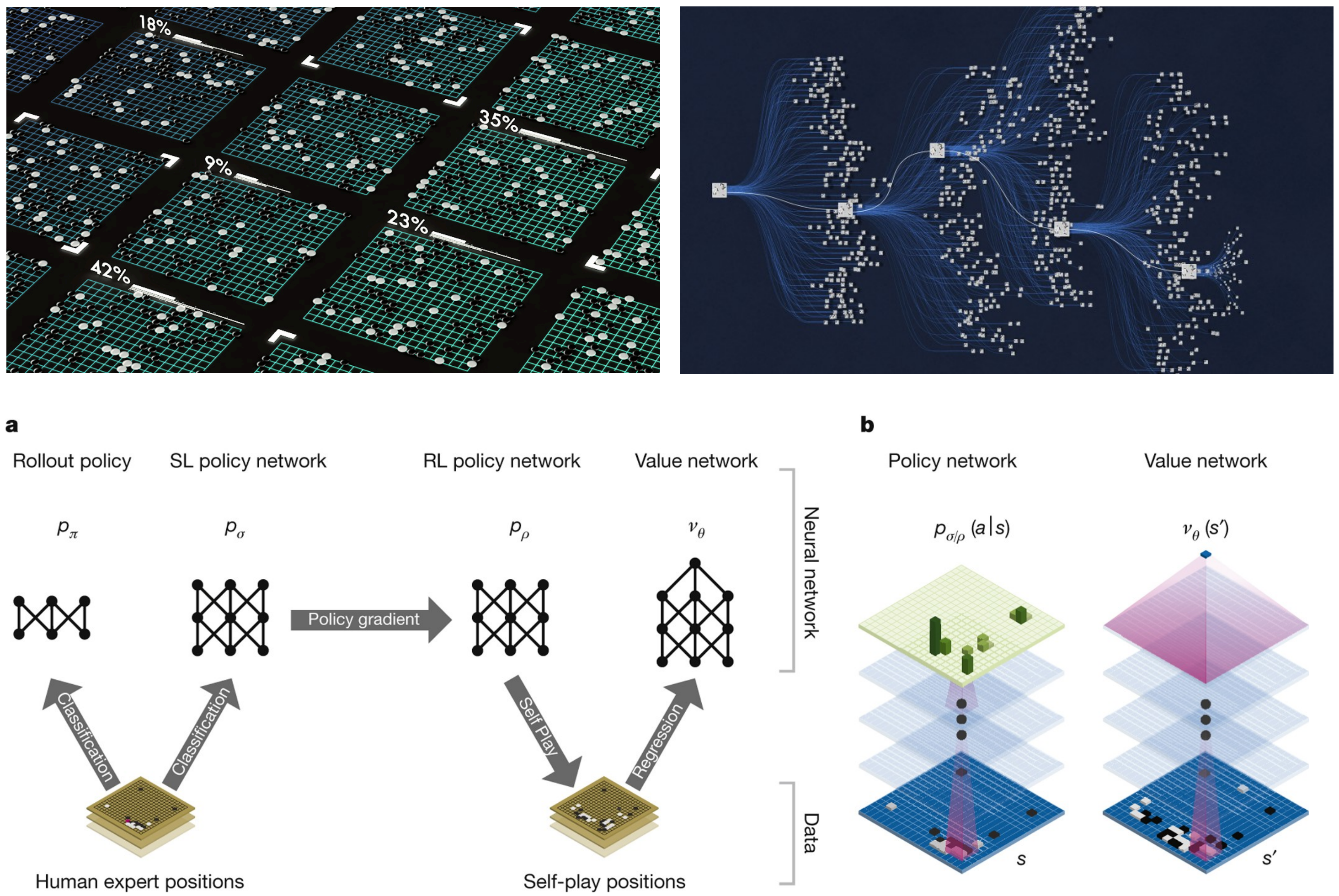




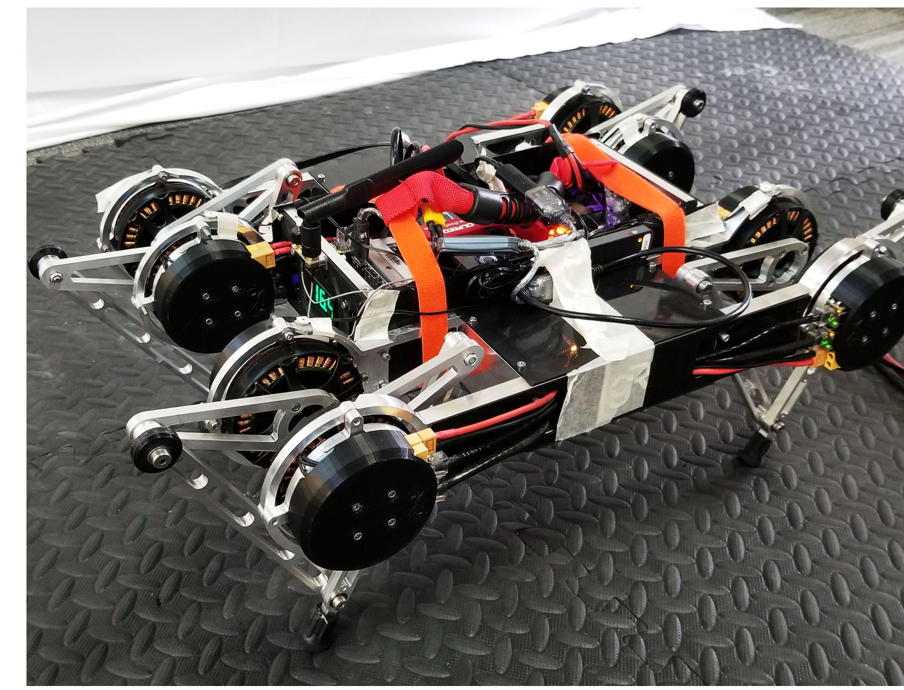
http://www.shivonzilis.com/machineintelligence

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



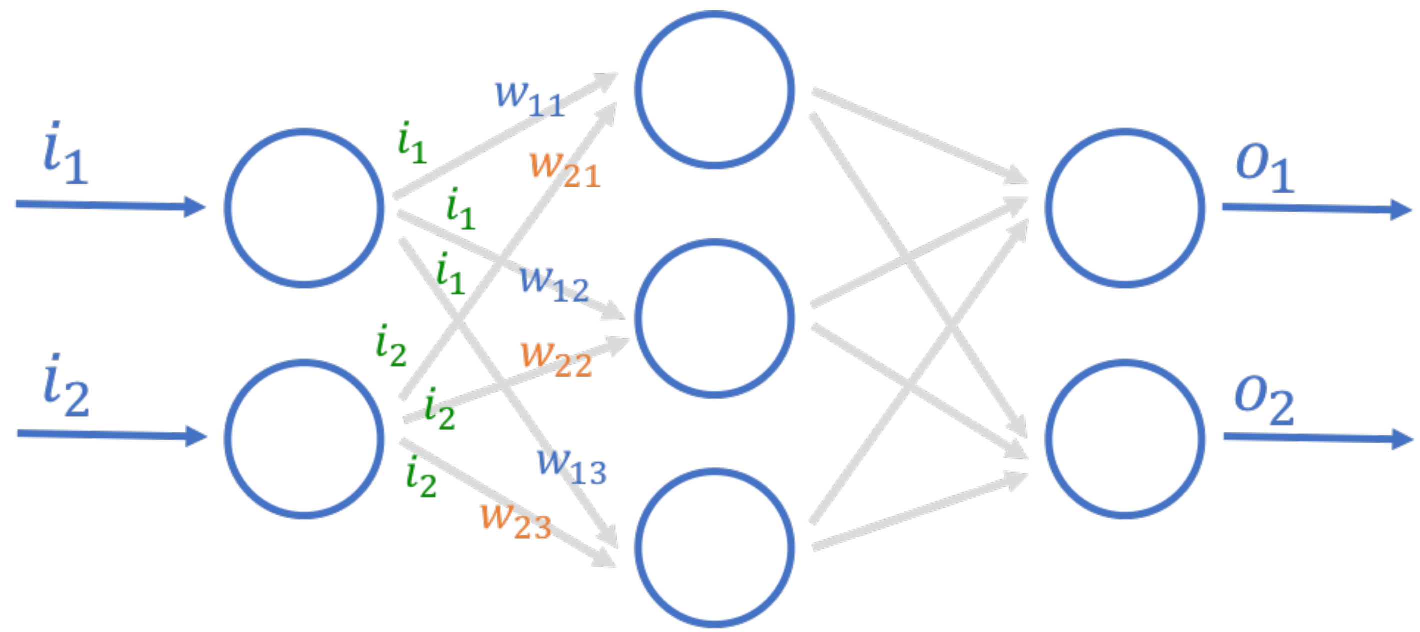
Mastering the game of Go with deep neural networks and tree search,
https://doi.org/10.1038/nature16961



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.

Operation Vectorization

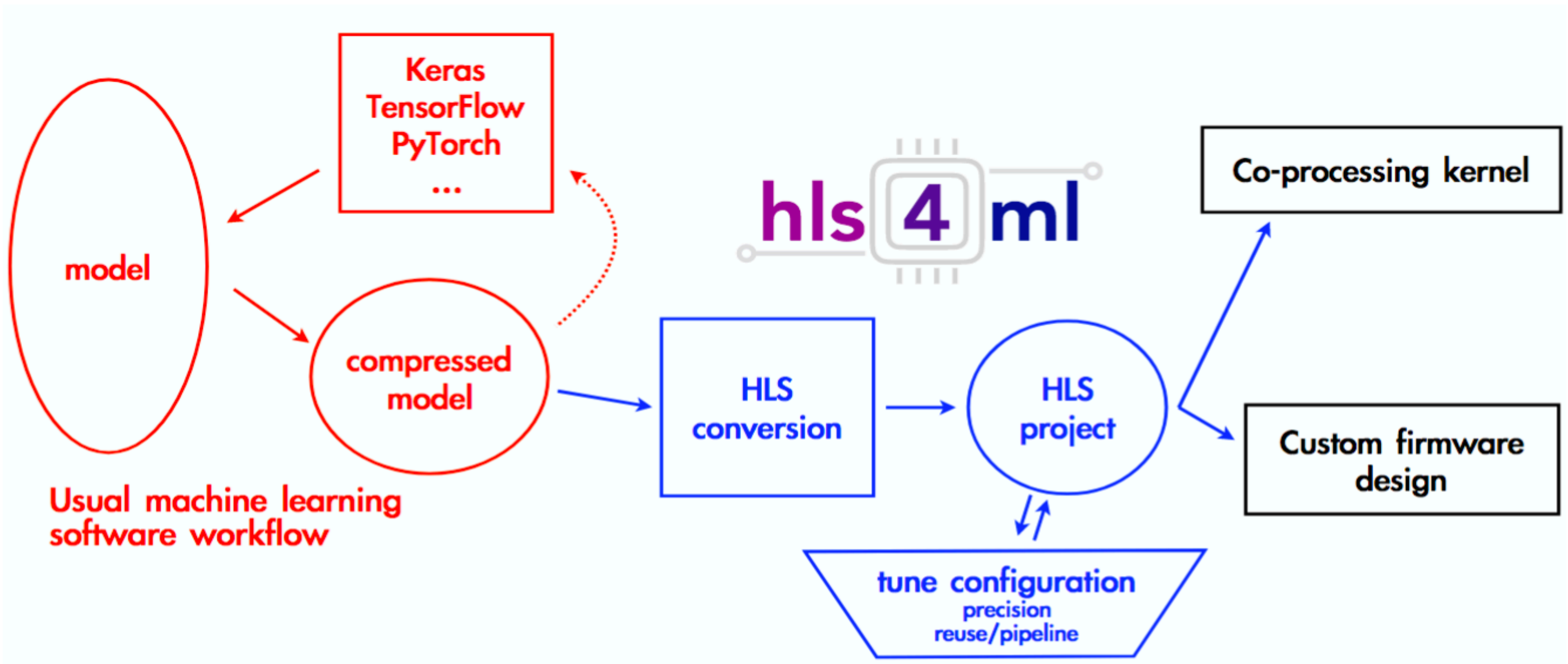


ANN ≡ matrix operations ≡ parallelizable

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

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]

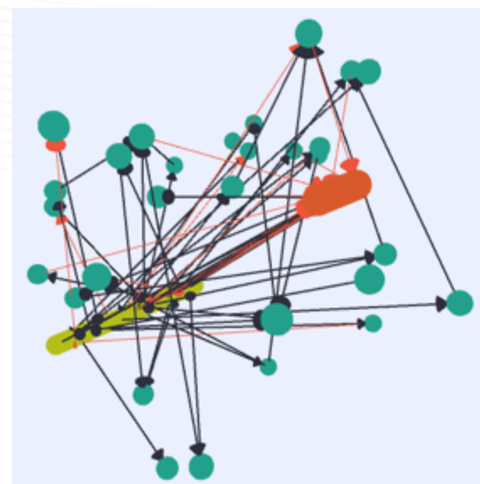
Artificial neural network model can be **executed efficiently on FPGA**, GPU, TPU, ...

Low Power Prediction

Best Results: Single View

x-view (127x50)	conv1 (8x3)	pool1 (2x1)	conv2 (7x3)	pool2 (2x1)	conv3 (6x3)	pool3 (2x1)	conv4 (6x3)	pool4 (2x1)	fc1 (196)	drop out	fc2 (98)	drop out	fc3 (11)	classification

Convolutional Neural Network Result: ~80.42%



- 90 neurons, 86 synapses
- Estimated energy for a single classification for mrDANNA implementation: 1.66 μJ

Spiking Neural Network Result: ~80.63%

Source for CNN results: A. Terwilliger, et al. Vertex Reconstruction of Neutrino Interactions using Deep Learning. IJCNN 2017.

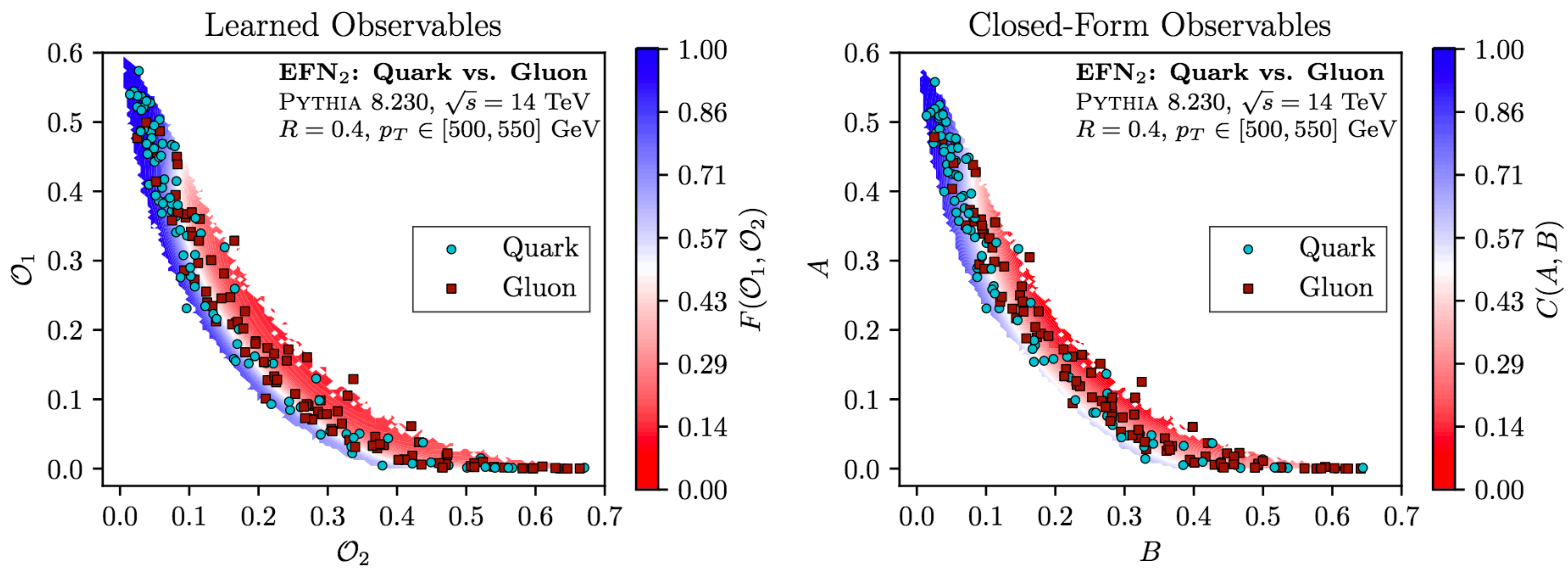
33 Programming Neuromorphic Computing Systems

OAK RIDGE National Laboratory

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

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

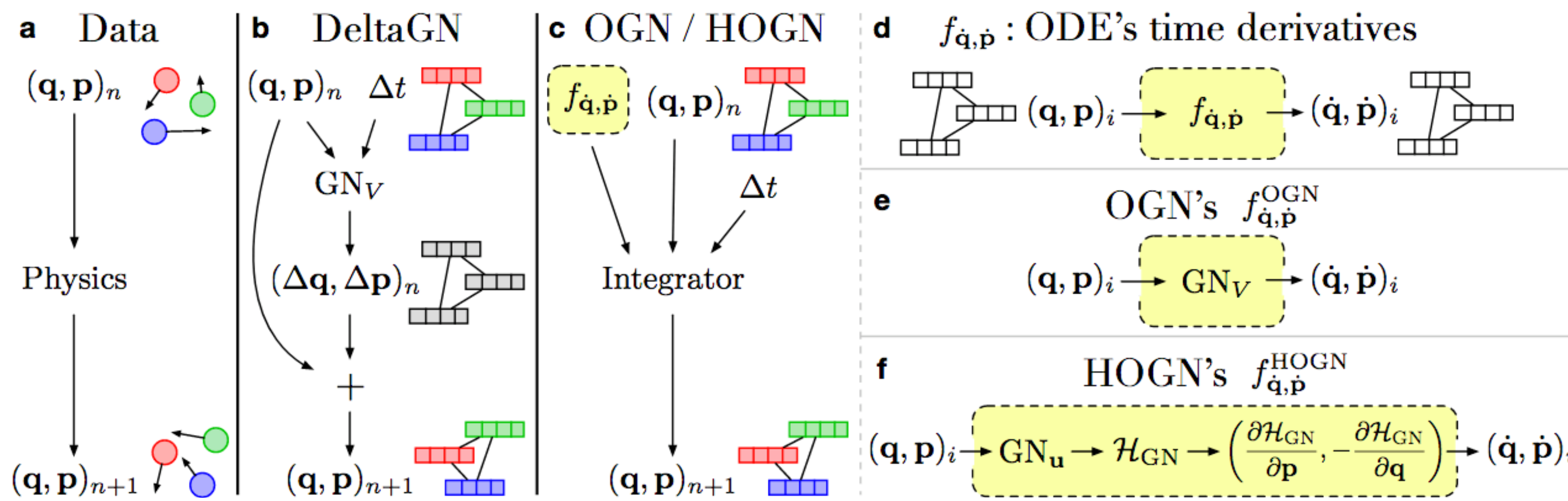
Physics Knowledge

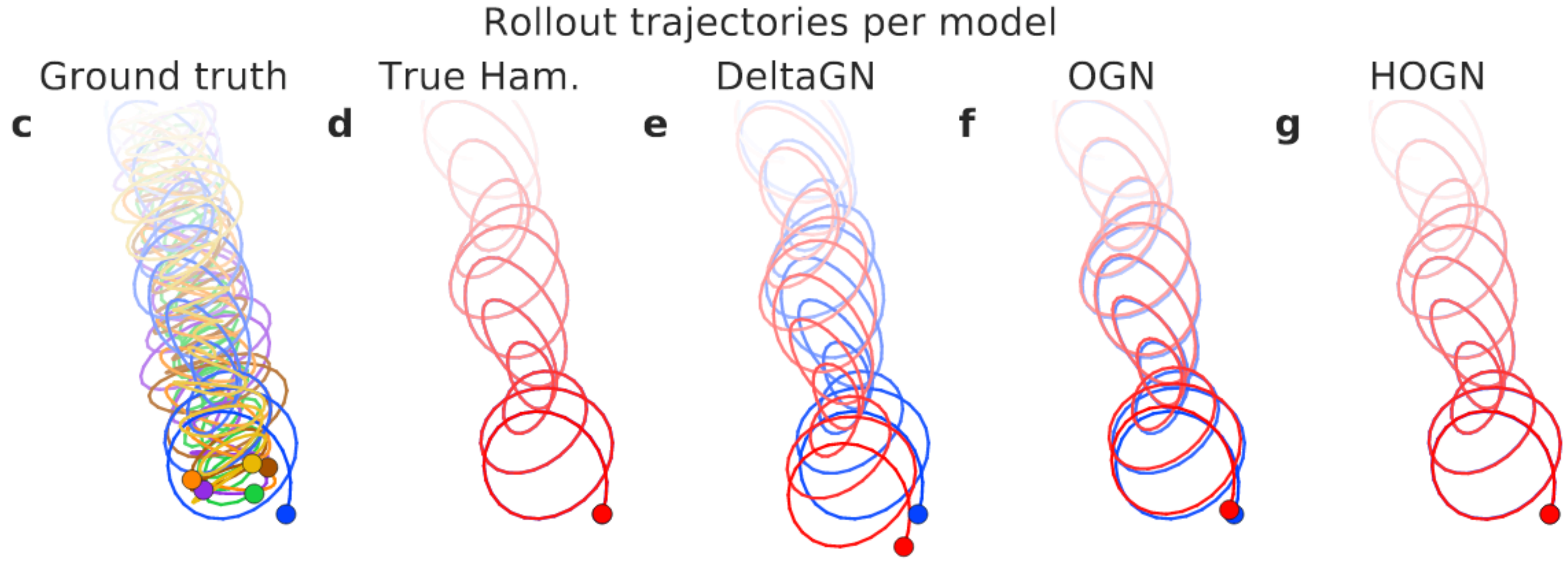


P. Komiske, E. Metodiev, J. Thaler, [1810.05165]

Machine Learning can **help understand Physics**.

Use Physics

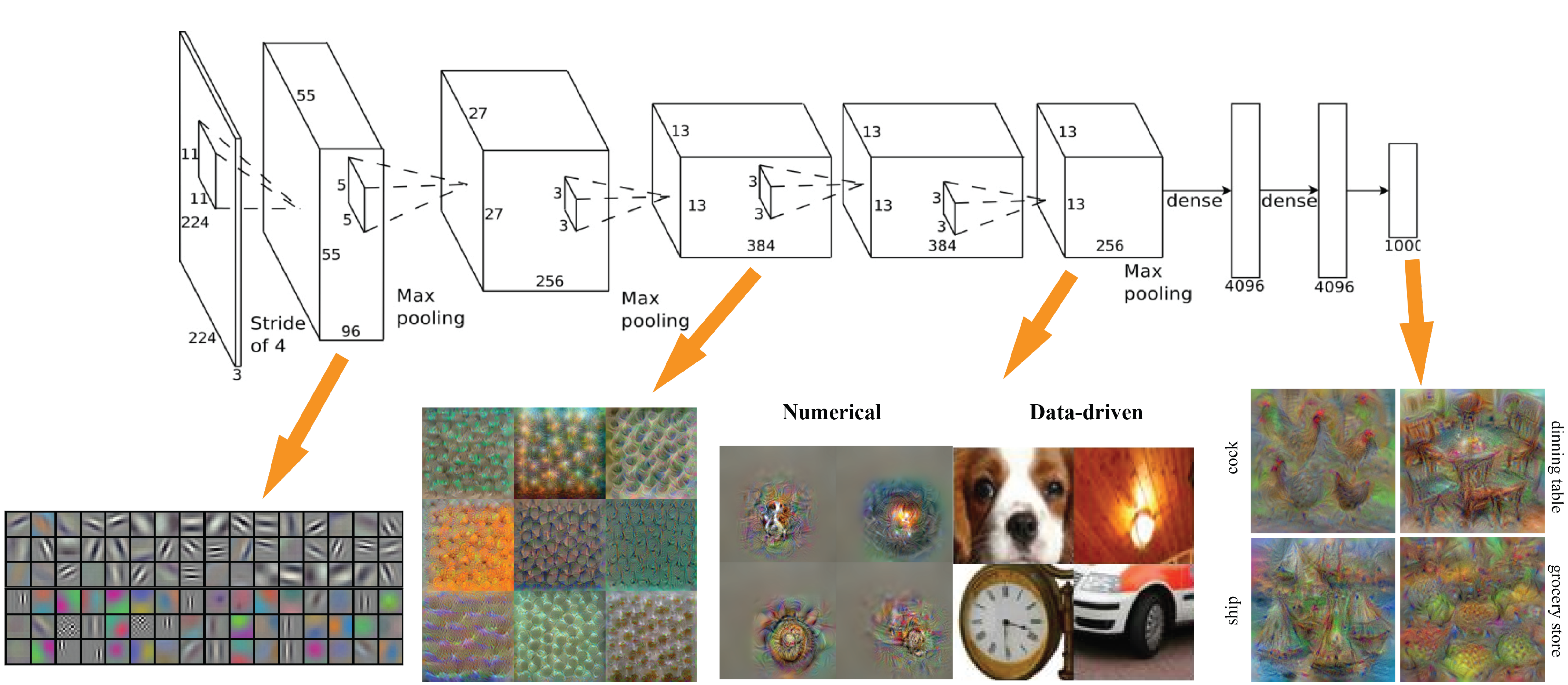




A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, P. Battaglia [1909.12790]

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

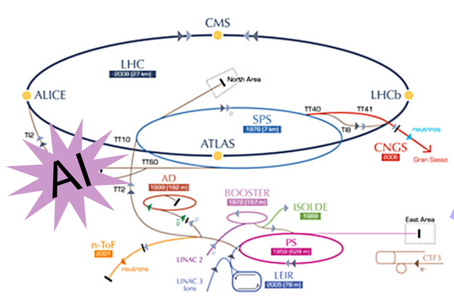
Machine learning model can **extract information from complex dataset.**

More classical algorithm counter part may

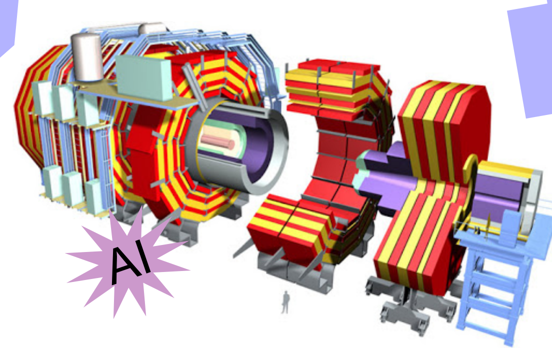
take **years of development.**

AI in HEP

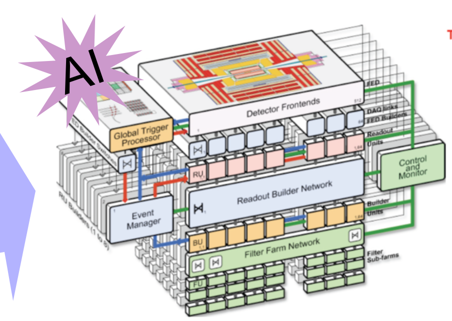
Role of AI: accelerator control, data acquisition, event triggering, anomaly detection, new physics scouting, event reconstruction, event generation, detector simulation, LHC grid control, analytics, signal extraction, likelihood free inference, background rejection, new physics searches, ...



Large Hadron Collider
40 MHz of collision



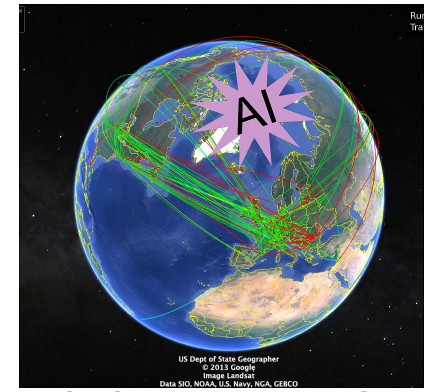
CMS Detector
1PB/s



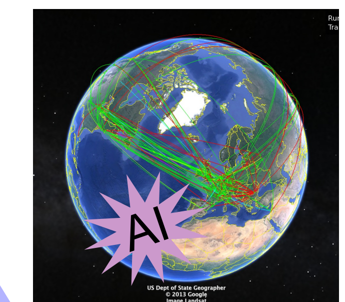
CMS L1 & High-Level Triggers
50k cores, 1kHz

CERN Tier-0
Computing Center
20k cores

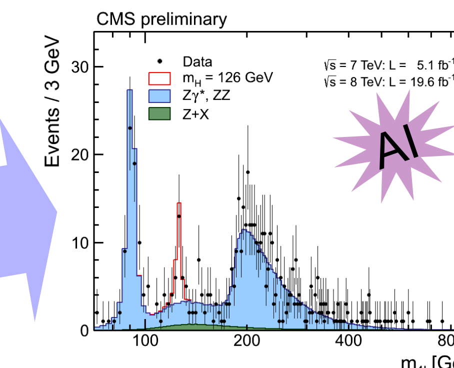
CERN Tier-0/Tier-1
Tape Storage
200PB total



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



LHC Grid
Remote Access
to 100PB of data



Rare Signal
Measurement
~1 out of 10^6

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

Reconstruction

Mostly pattern recognition tasks with regressions and classifications
Development on multiple tasks:

- Local energy reconstruction
- Jet reconstruction
- Particle (flow) reconstruction
- Tracking
- Vertexing
- …

Composite reconstruction and end-2-end approaches.

Graph neural networks are emerging as overarching solutions for reconstruction tasks

Generative Models in HEP

Enormous potential gain in computing performance (x thousands)
Extensive R&D effort on-going

- Better PDF
- Faster phase space integration
- Matrix element surrogates
- Particle shower energy deposit simulation
- Analysis-level sample production
- Learning the detector transfer function
- Reconstructed-particle-level fast simulation
- …

Extension of such models

- Unfolding mechanism
- Anomaly detection
- Background subtraction
- …

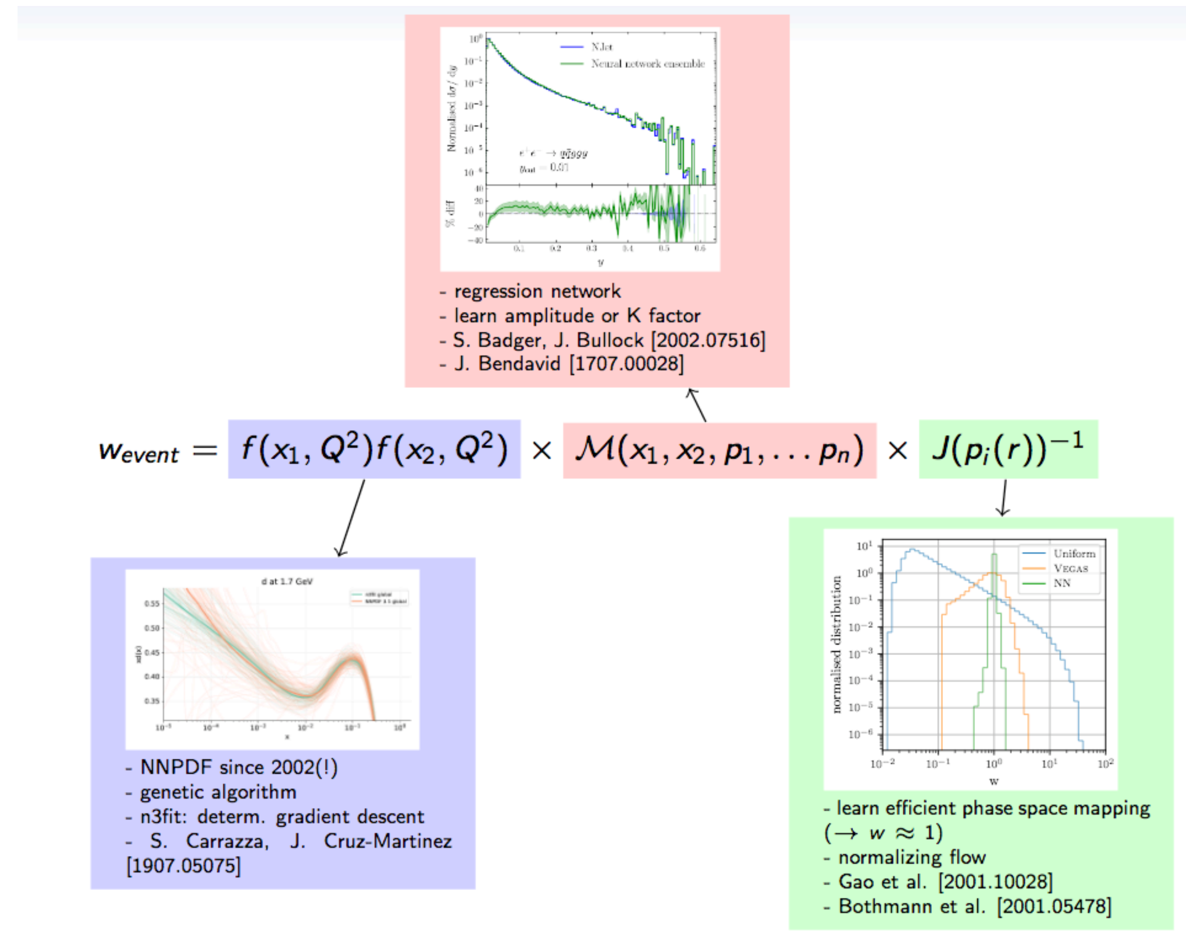


Diagram A. Butter

Generative adversarial network (GAN),Variational Auto-Encoder (VAE) and hybrid solutions taking the best of both methods

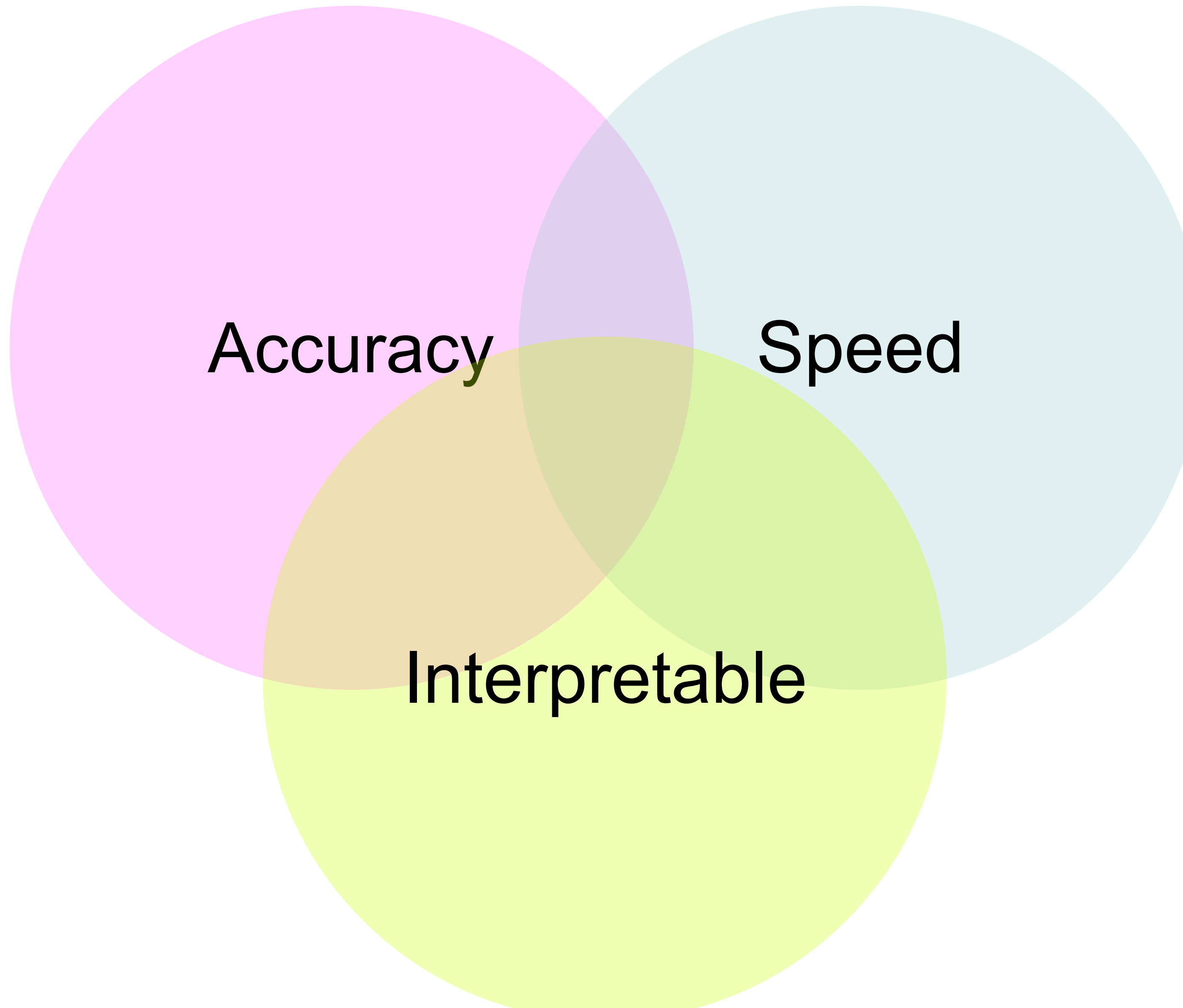
Taking Control of Apparatus

Multiple components of HEP data pipe-line can be operated with AI
- Accelerators
- Detectors
- DAQ/trigger
- Internet networking
- Data management
- Computing facilities
- …

Reinforcement learning for learning policies is challenging to put in practice (environment simulation / lots of data required, …). Alternative approaches using supervised learning.

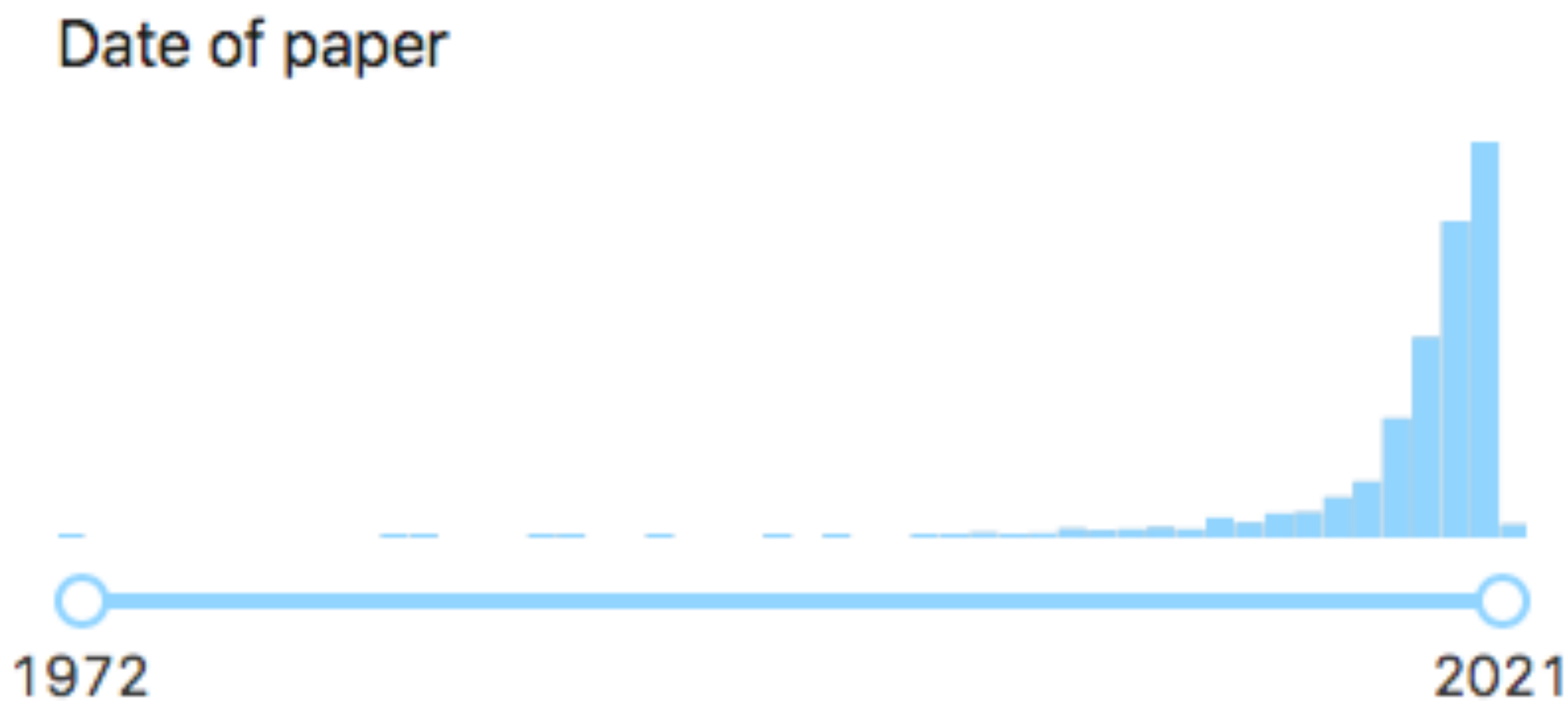
Potential gain in operation cost, and utilization efficiency.

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/

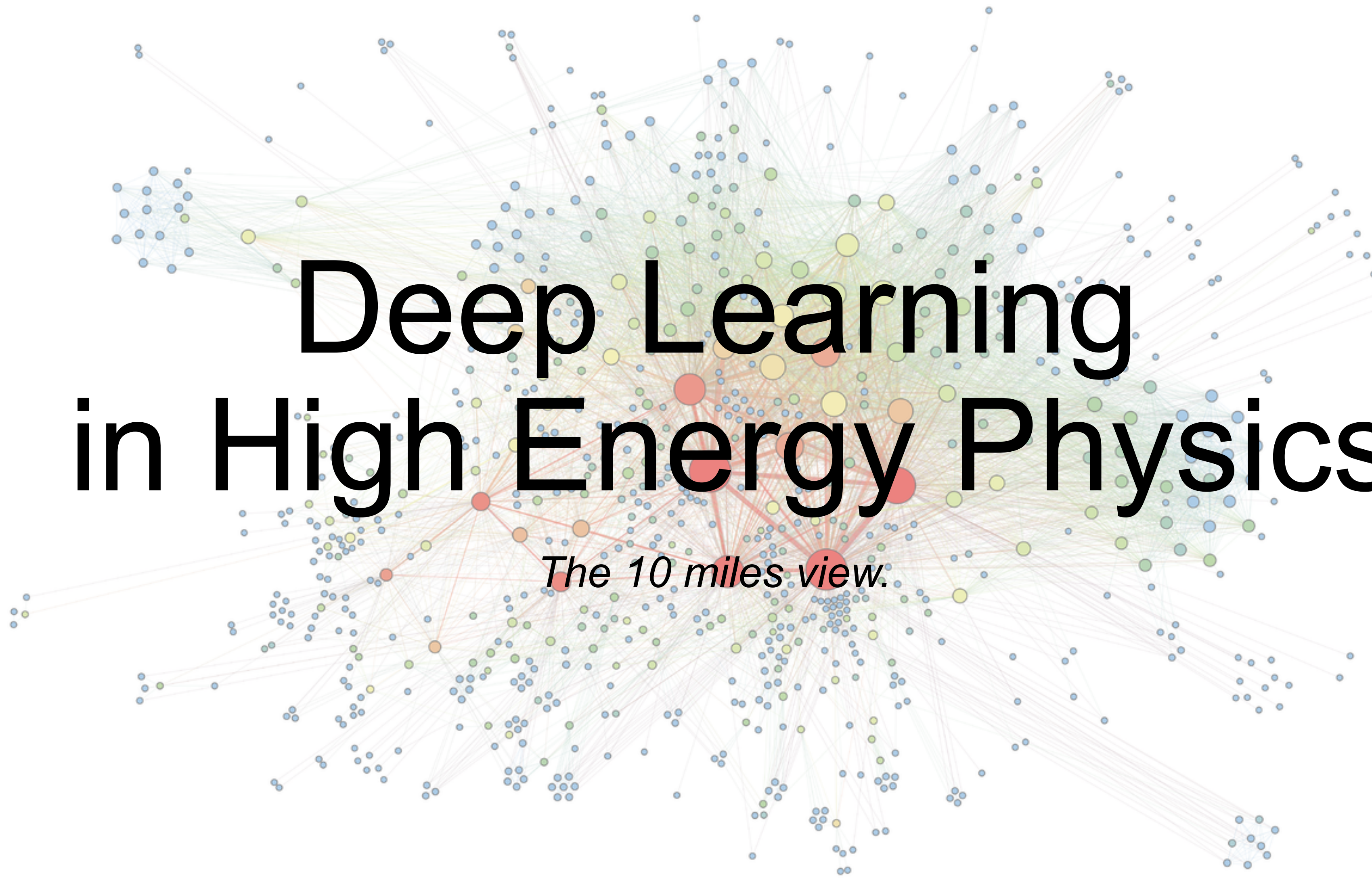
Take home message :

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

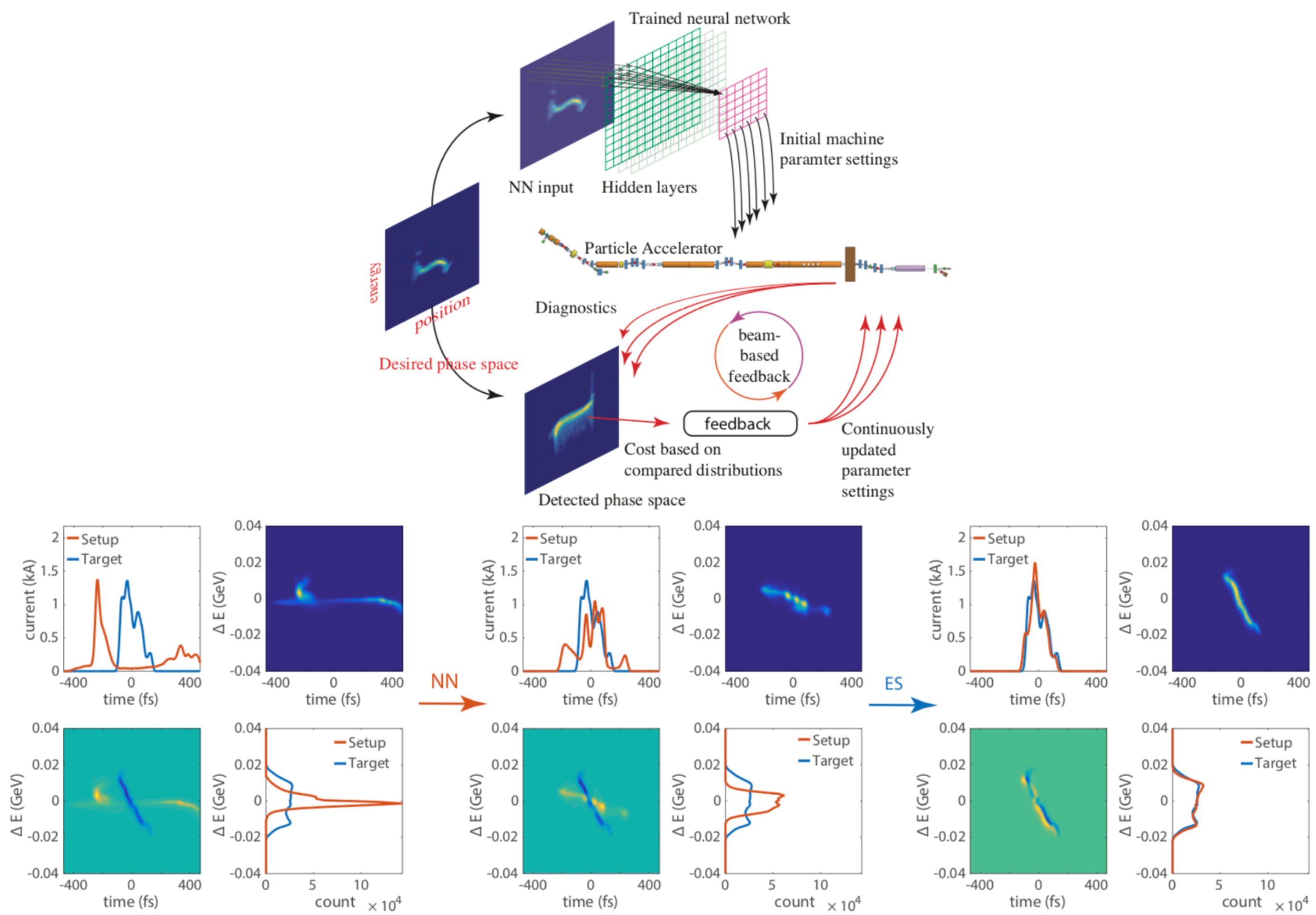
Wide range of potential applications



Deep Learning in High Energy Physics

The 10 miles view.

Producing the Data

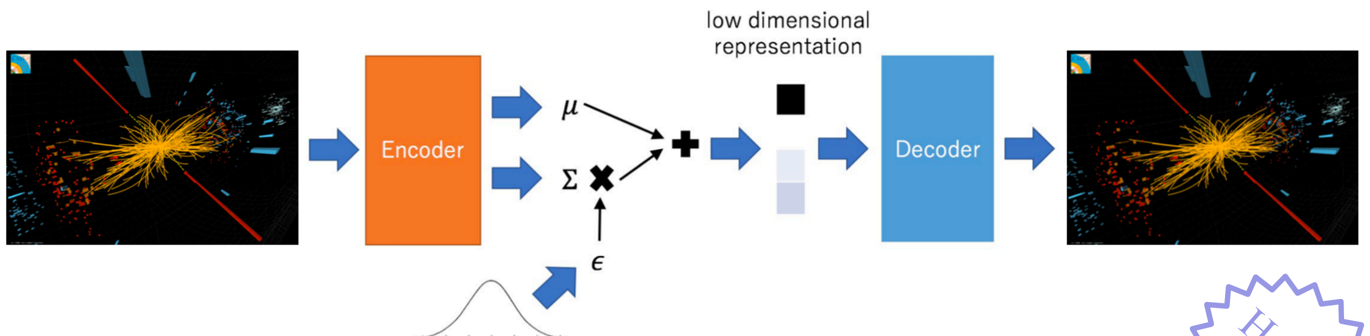


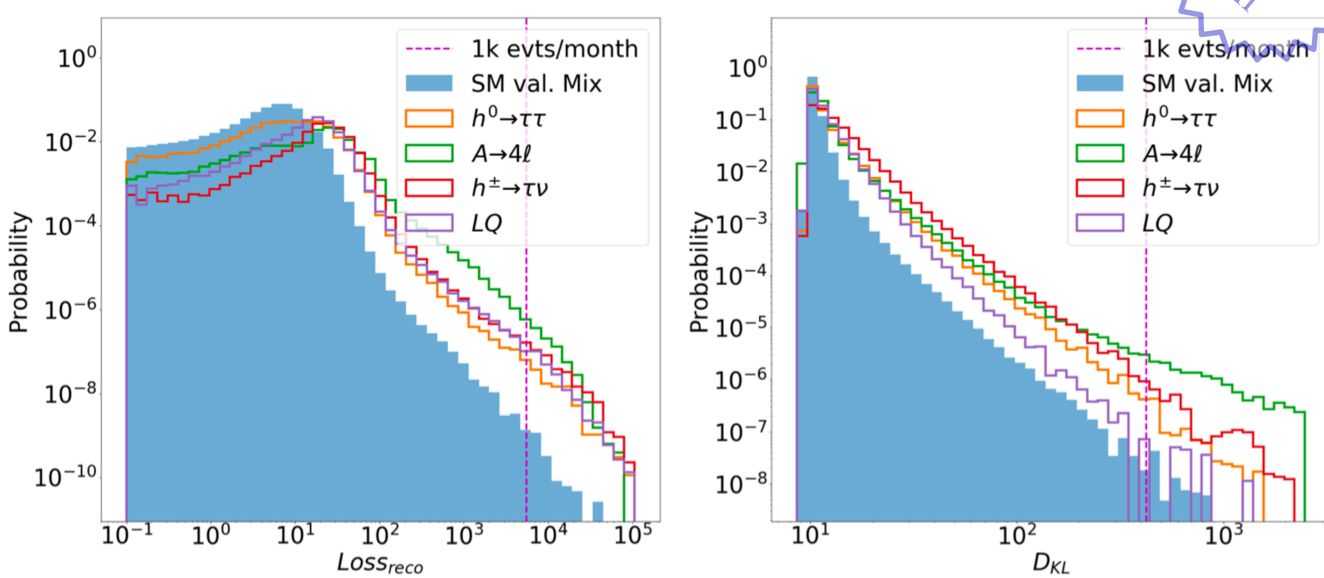
A. Scheinker, C. Emma, A.L. Edelen, S. Gessner [2001.05461]

- 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.

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]
…

Acquiring Data



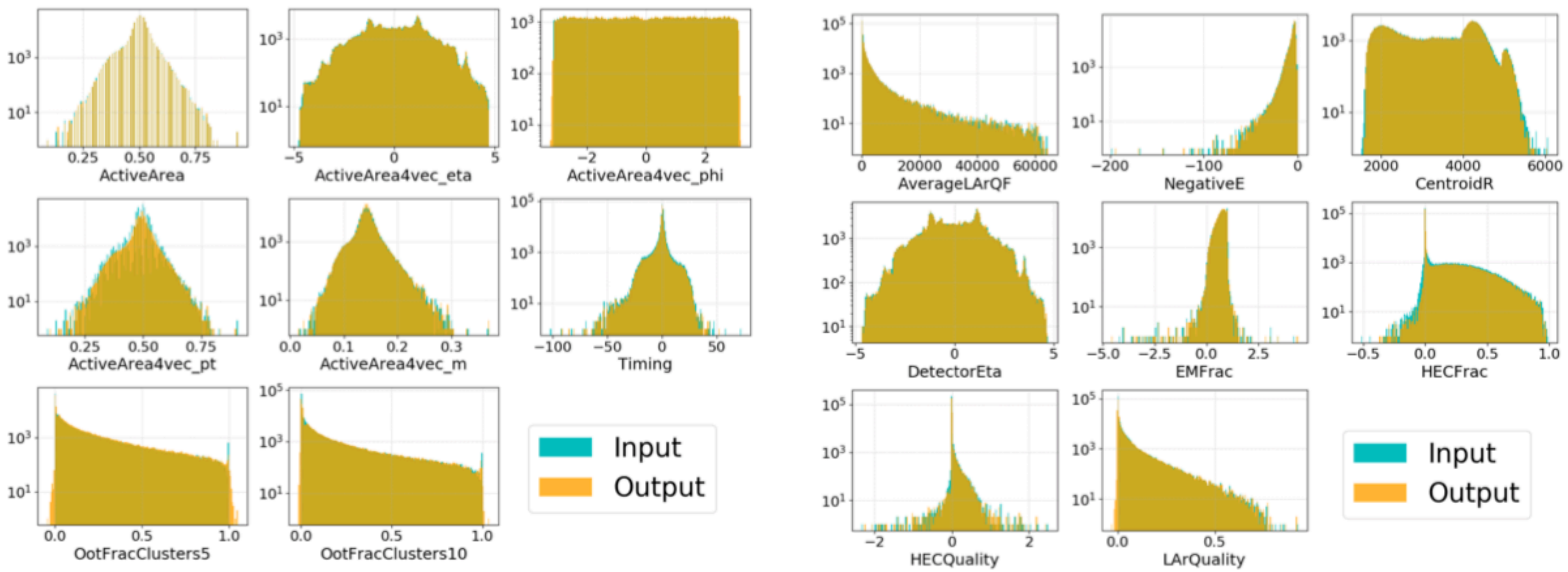


Hands-on

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.
- More promising R&D and experiment adoption.

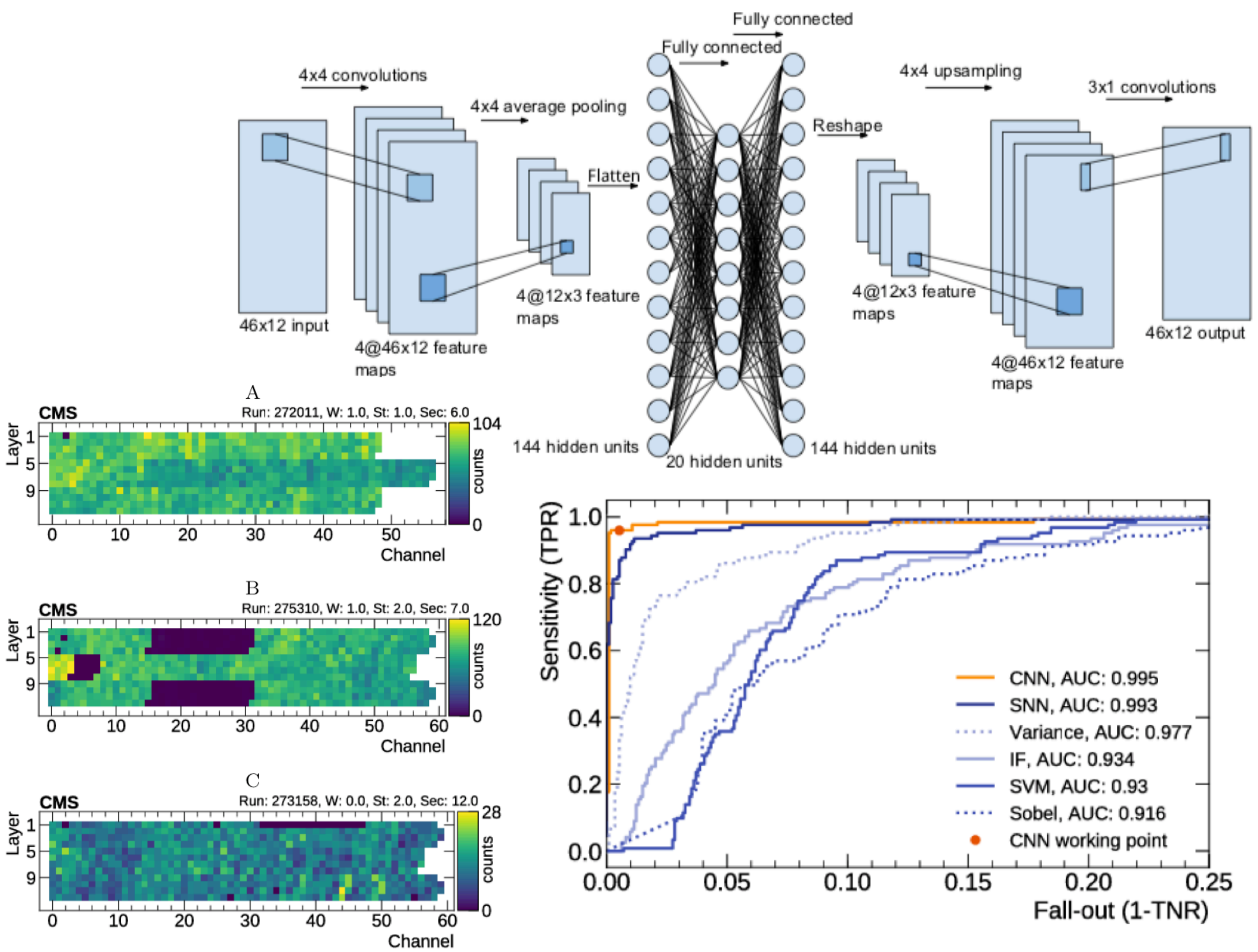
Compressing Data



Use of auto-encoder model
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 data analysis.
- R&D needed to reach the necessary level of fidelity.

Cleaning Data

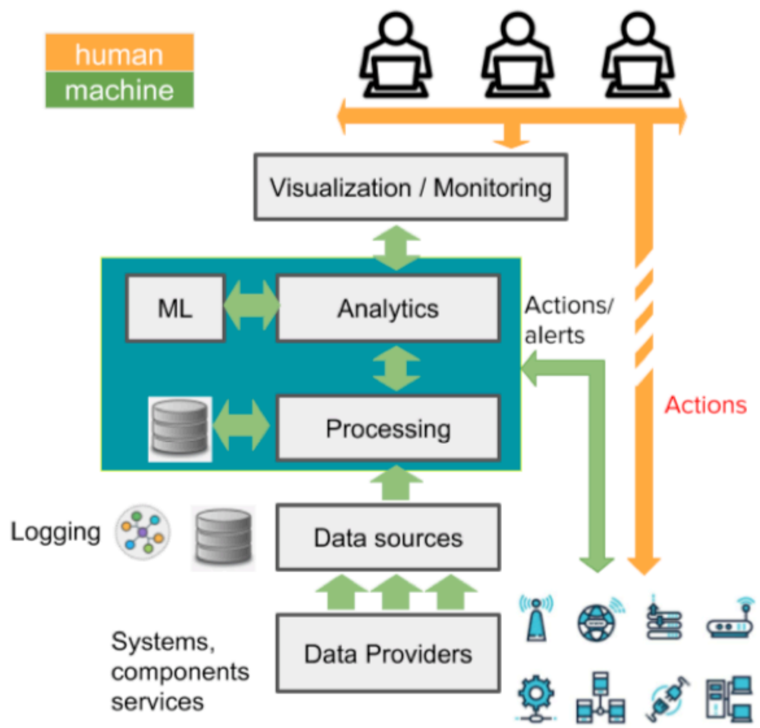


A.A. Pol, G. Cerminara, C. Germain, M. Pierini, A. Seth
[doi:10.1007/s41781-018-0020-1]

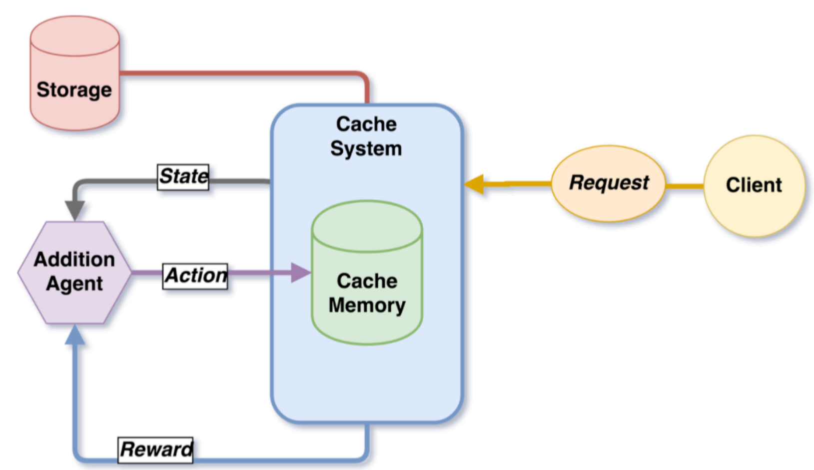
- Data quality is a person power intensive task, and crucial for swift delivery of Physics
- Machine learning can help with automation.
- Learning from operators, reducing workload.
- Continued R&D and experiment adoption.

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



[cds:2709338]
https://operational-intelligence.web.cern.ch

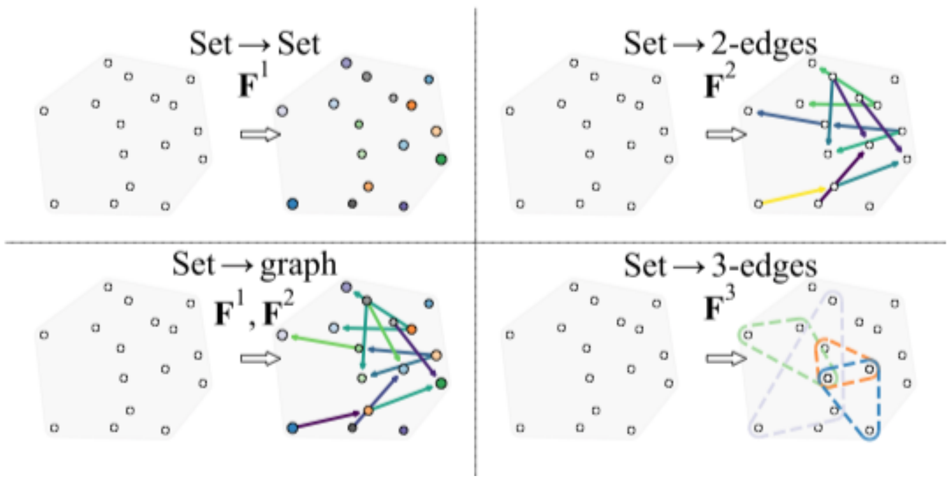


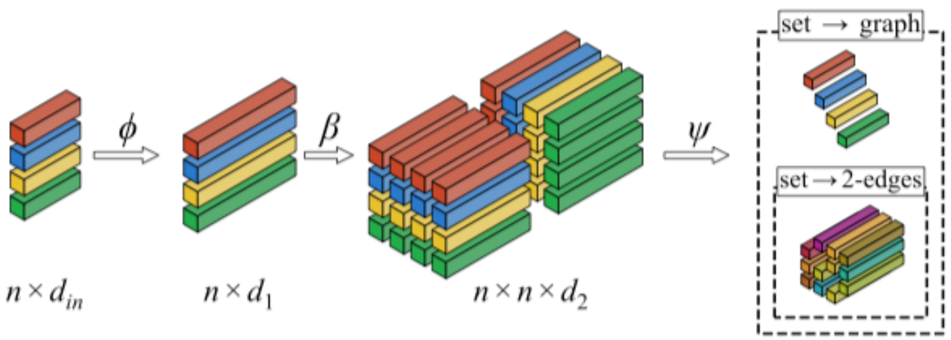
Cache Type	Throughput	Cost	Read on hit ratio	Band sat.	CPU Eff.
SCDL	**79.43%**	**50.68%**	21.22%	58.94%	58.75%
LFU	65.01%	104.73%	**33.29%**	**51.00%**	**60.92%**
Size Big	49.02%	111.73%	28.55%	54.40%	60.41%
LRU	47.15%	112.84%	27.64%	54.93%	59.90%
Size Small	46.71%	113.01%	27.39%	55.01%	59.73%

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.

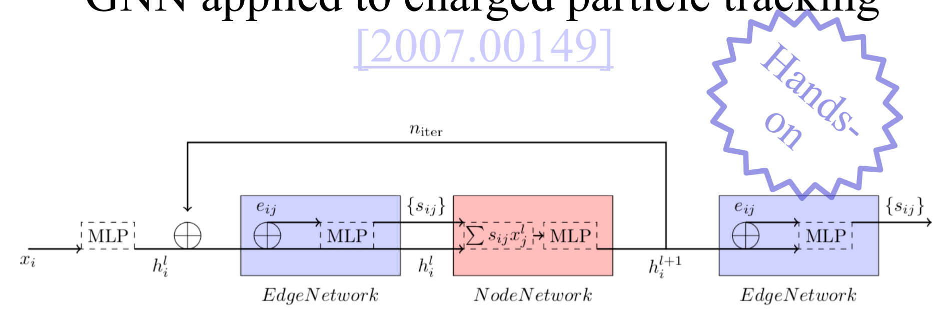
Reconstructing Data

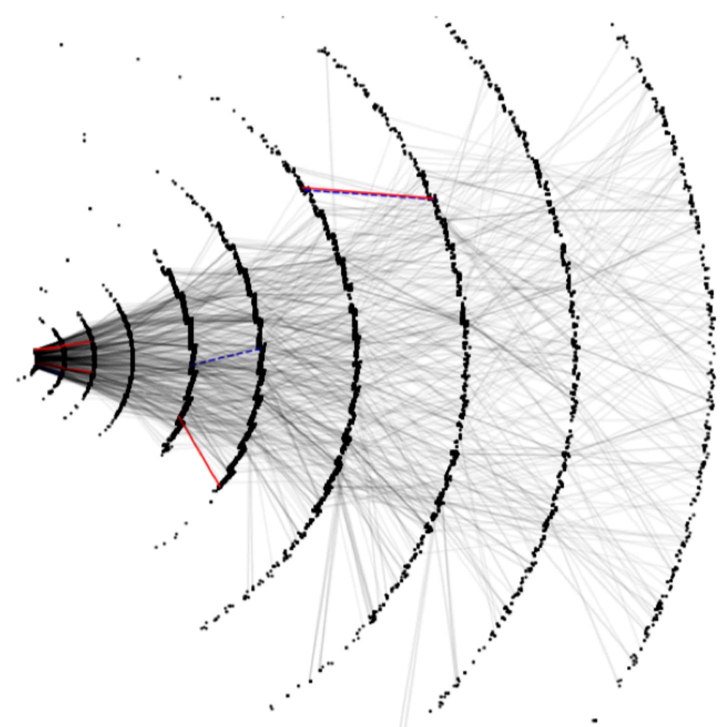




Learning graphs from sets, applied to vertexing
[2002.08772]

GNN applied to charged particle tracking
[2007.00149]

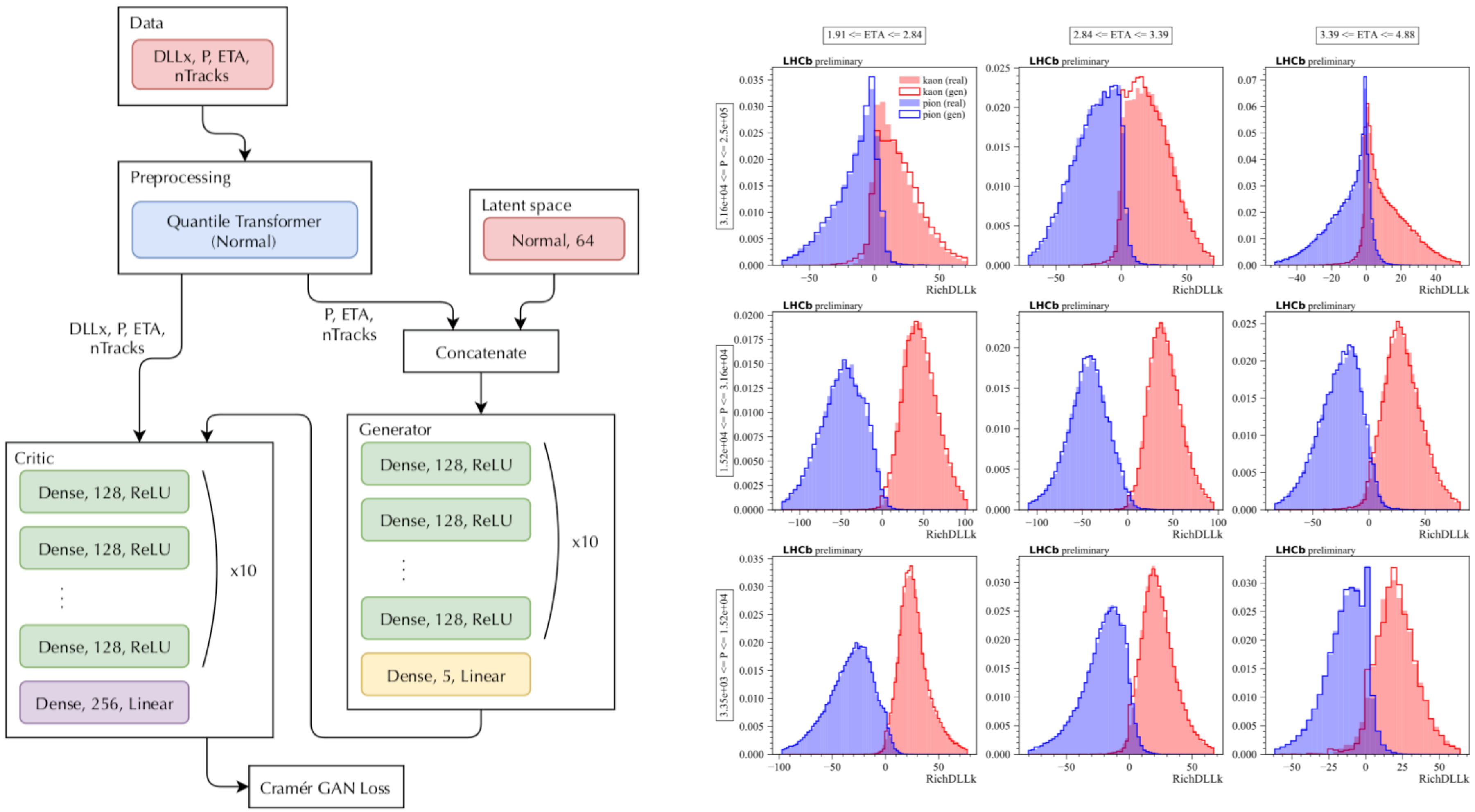




- Event reconstruction is pattern recognition to a large extend. Advanced machine learning techniques can help.
- Learn from the simulation, and/or data.
- Learn from existing "slow reconstruction" or simulation ground truth.
- Automatically adapt algorithm to new detector design.
- Image base methods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

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

Simulating Data

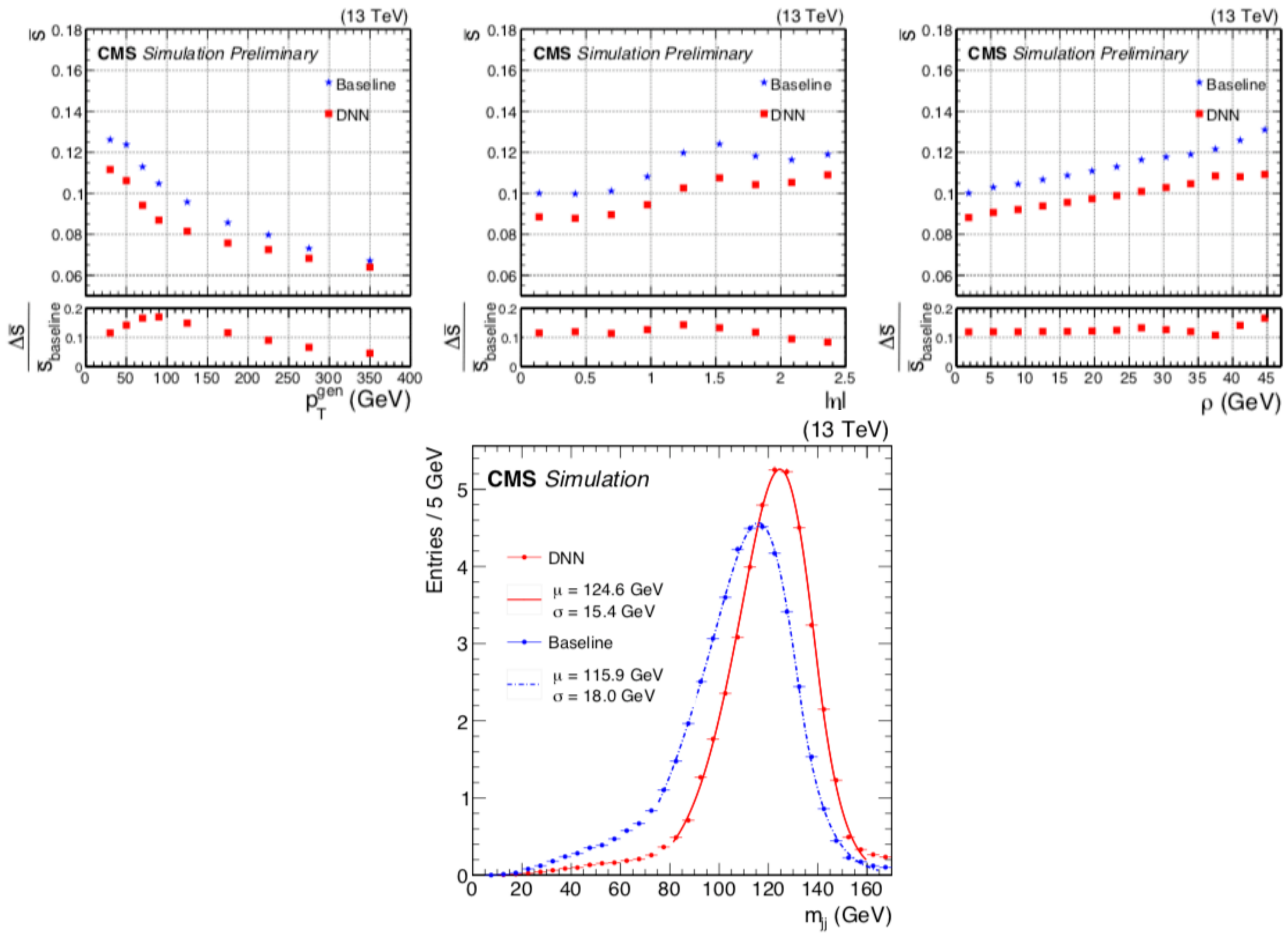


Generative Adversarial Networks for LHCb Fast Simulation
[2003.09762]

- 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.

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

Calibrating Data

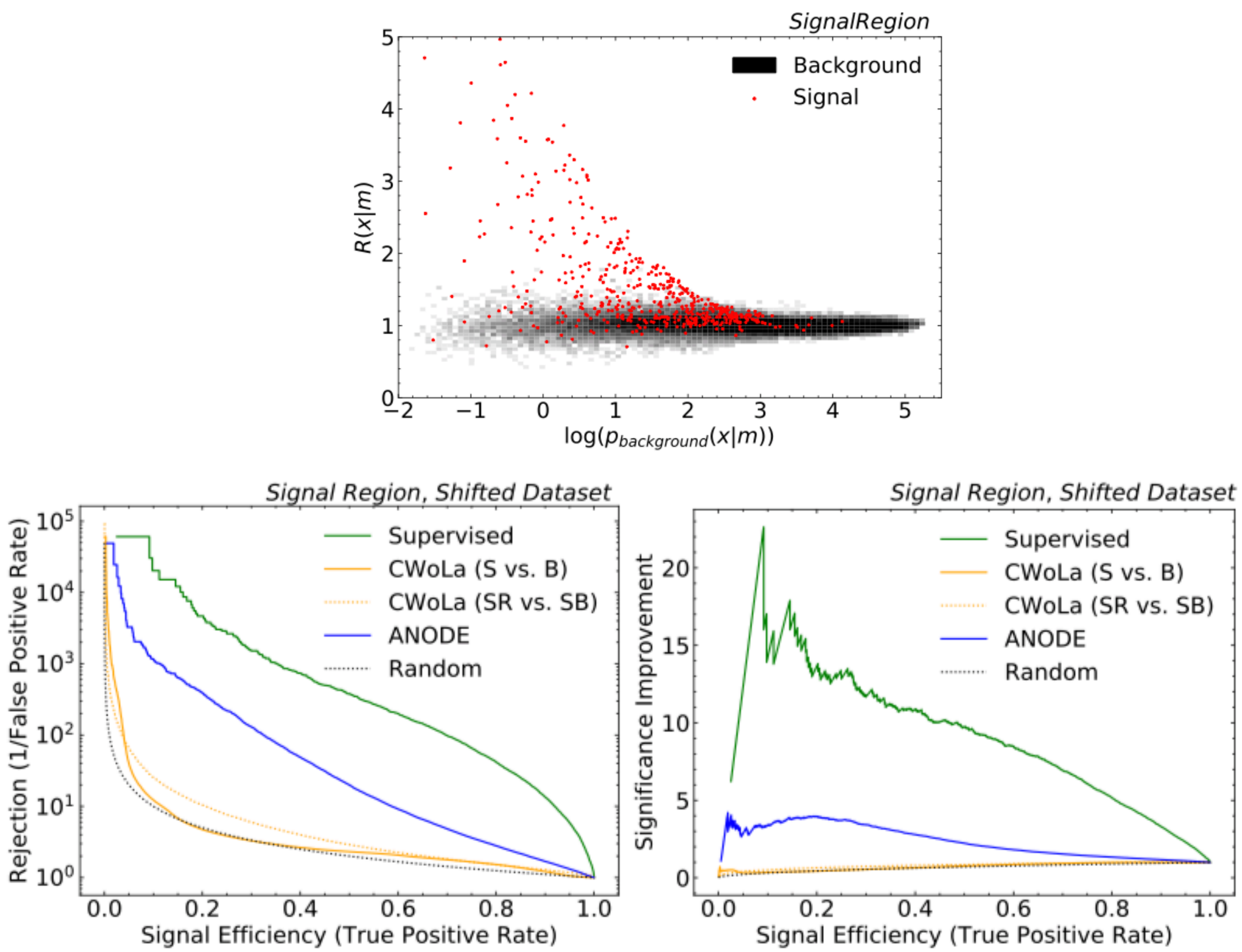


A deep neural network for simultaneous estimation of b jet energy and resolution [1912.06046]

- 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

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

Analyzing Data

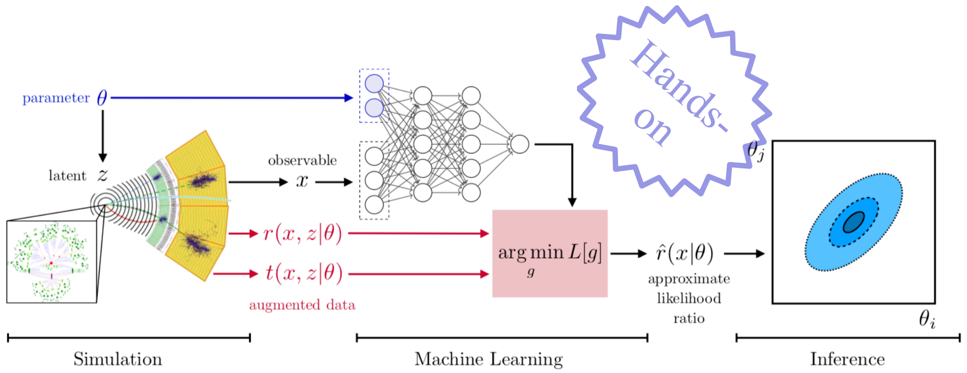


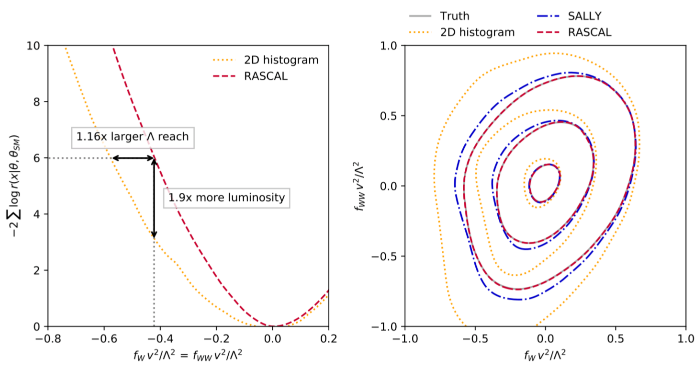
Use of masked autoregressive density estimator with normalizing flow as model-agnostic signal enhancement mechanism.
[doi:10.1103/PhysRevD.101.075042]

Much more relevant work going on.
https://iml-wg.github.io/HEPML-LivingReview/

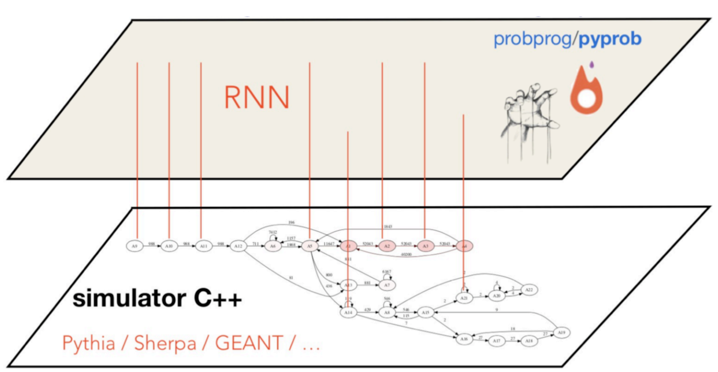
- Machine learning has long infiltrated analysis for signal/bkg classification.
- Increasing number of analysis with more complex DNN.
- Application to signal categorization, bkg modelling, kinematics reconstruction, decay product assignment, object identification, …
- Breadth of new model agnostic methods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

Theory Behind the Data

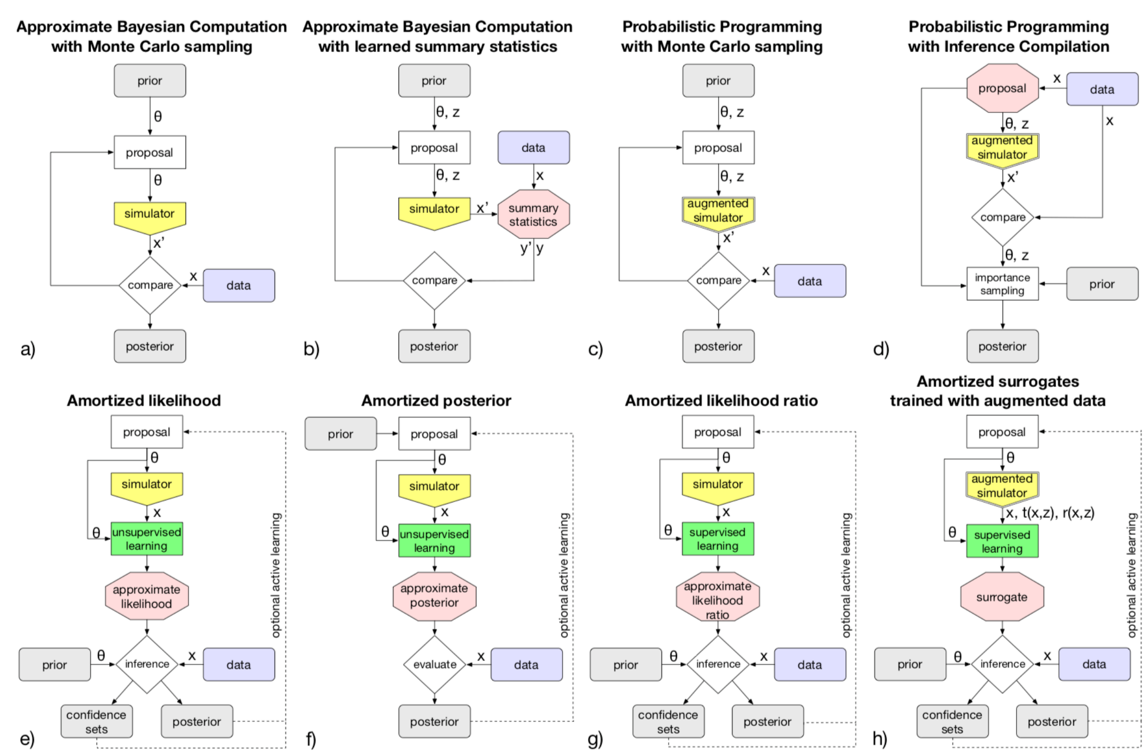




Constraining EFT with ML
[1805.00013]



https://github.com/probprog/pyprob



The frontiers of simulation-based inference
[1911.01429]

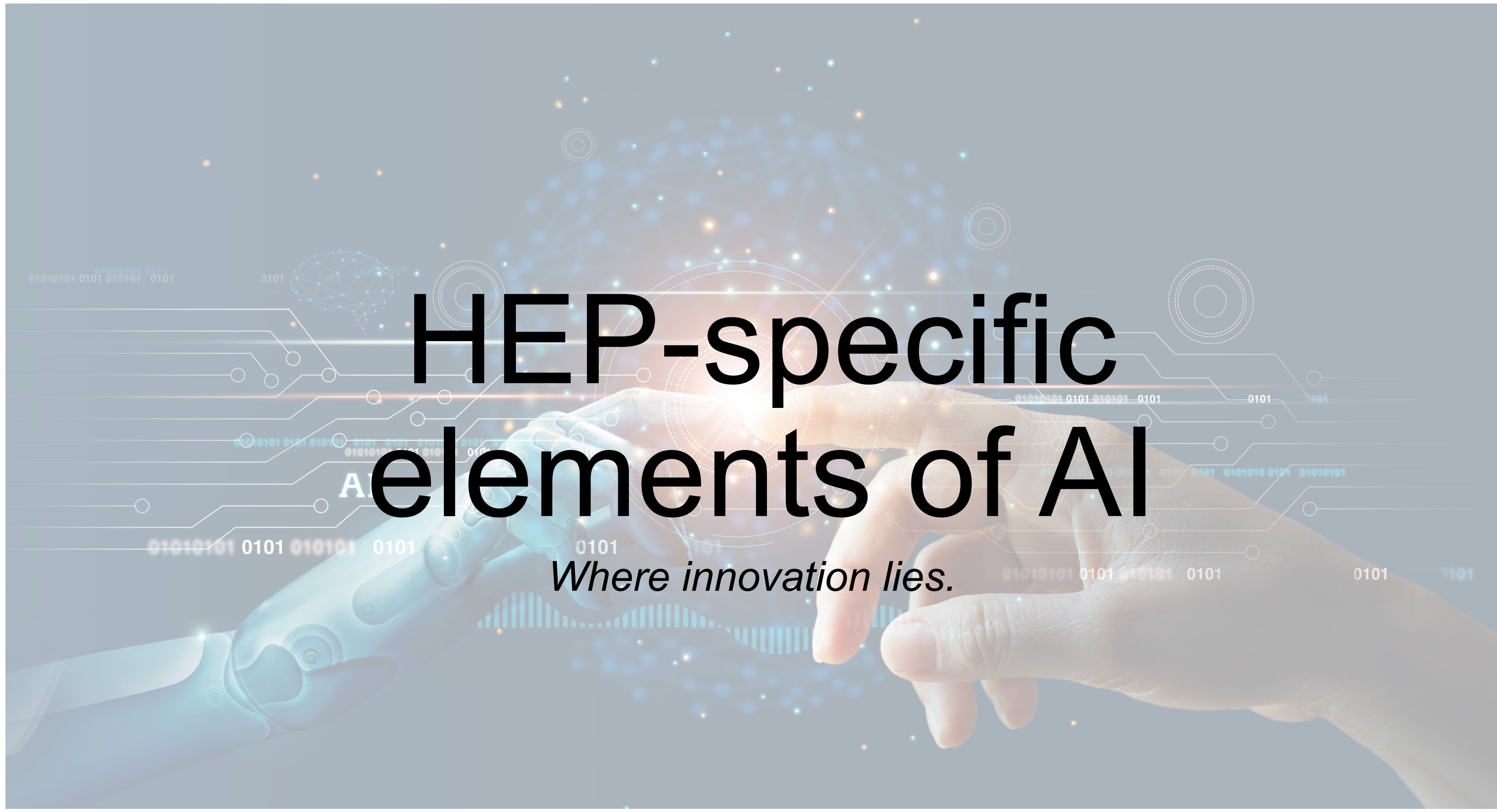
- Hypothesis testing is the core of HEP analysis.
- Intractable likelihood hinders solving the inverse problem.
- Going beyond the standard approach using machine learning and additional information from the simulator.
- More precise evaluation of the priors on theory's parameters.
- May involve probabilistic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

<u>Take home message :</u>

Rapid growth of machine learning applications in HEP

(too) Slowly turning proofs of concept into production

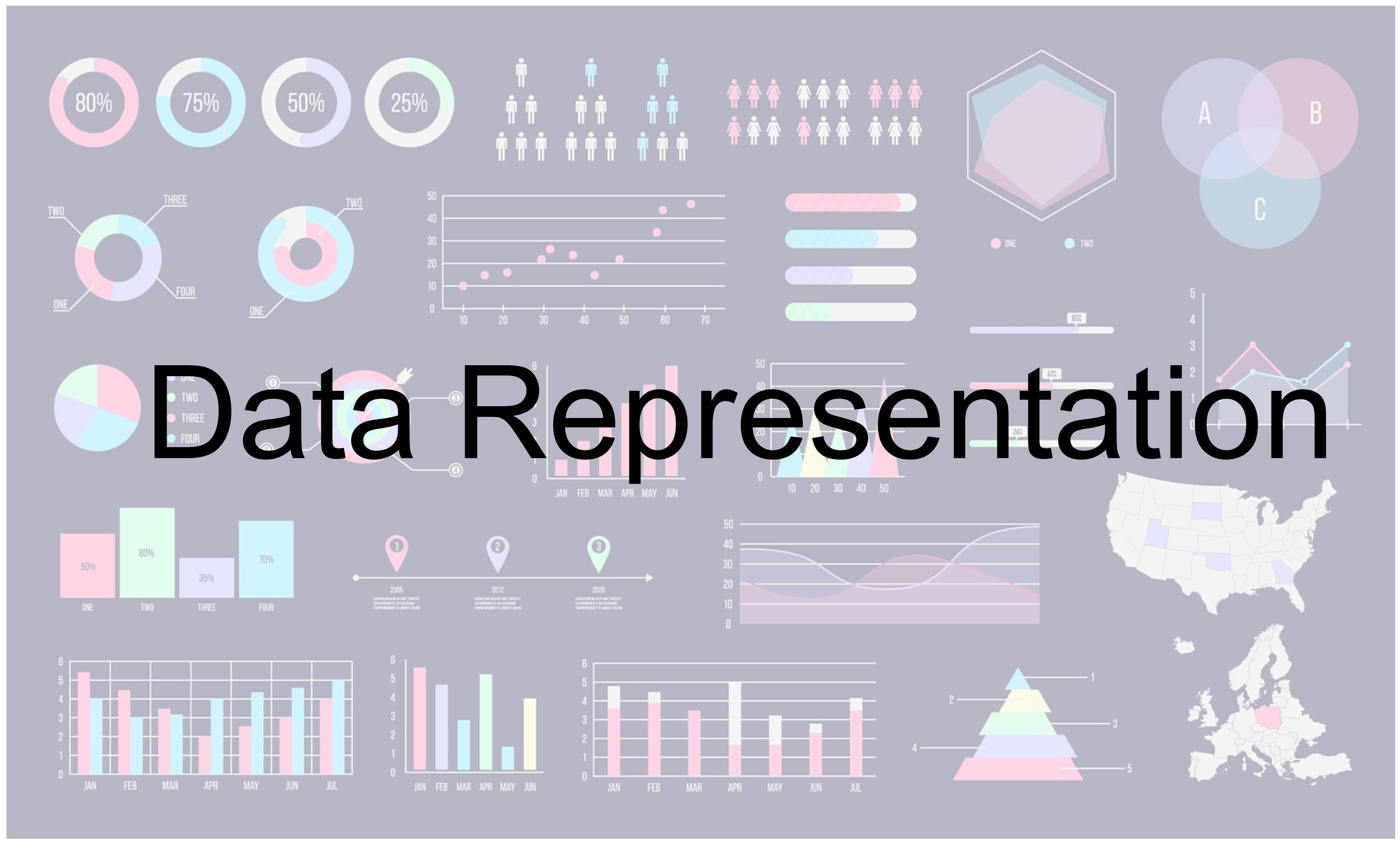
Exciting time ahead exploiting further the potential of AI



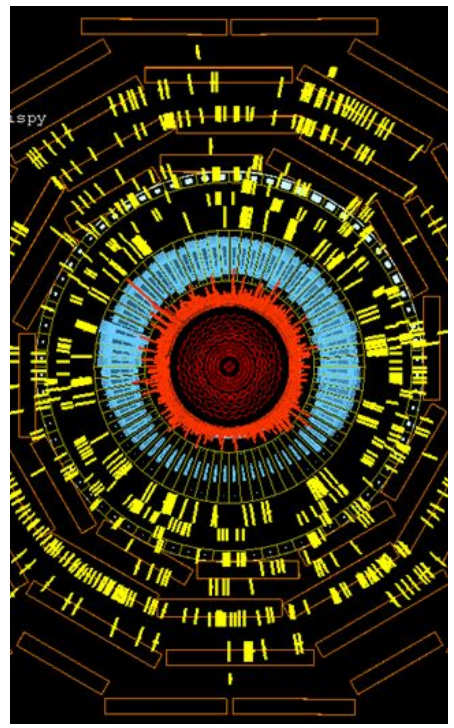
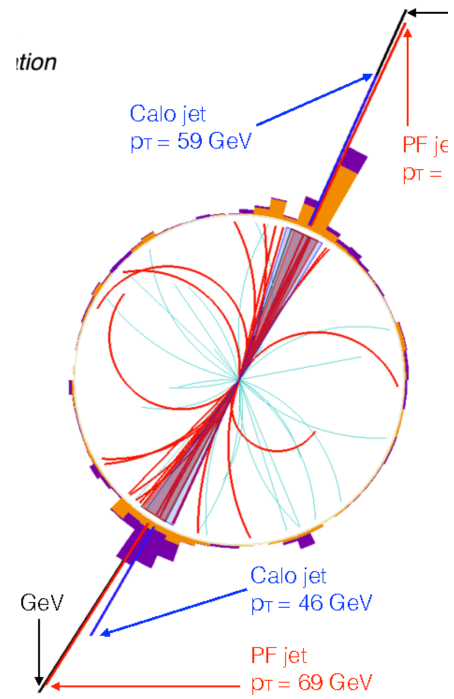
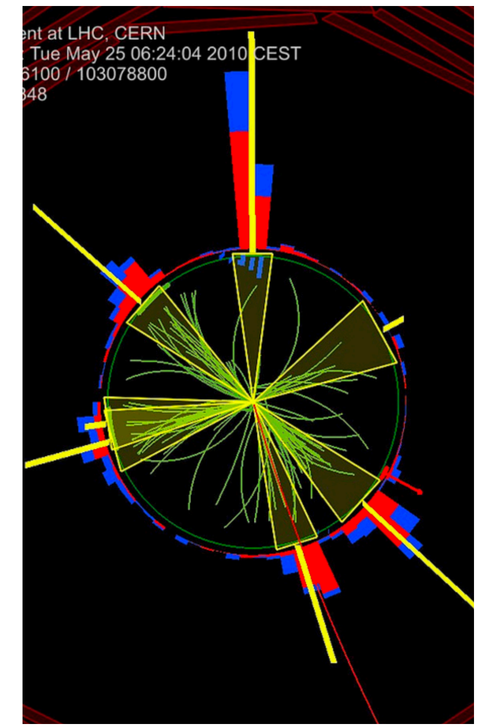
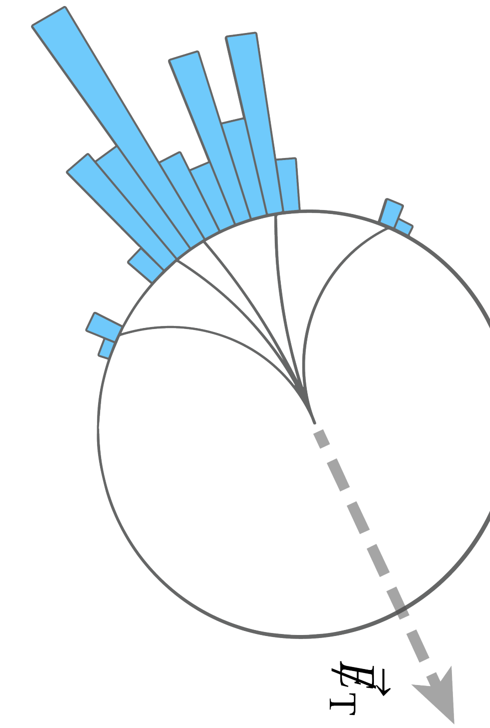
HEP-specific elements of AI

Where innovation lies.

Data Representation



From RAW to High Level Features

Detector Data	Local reconstruction	Particle representation	Jet Clustering	High level features
				

Event Processing

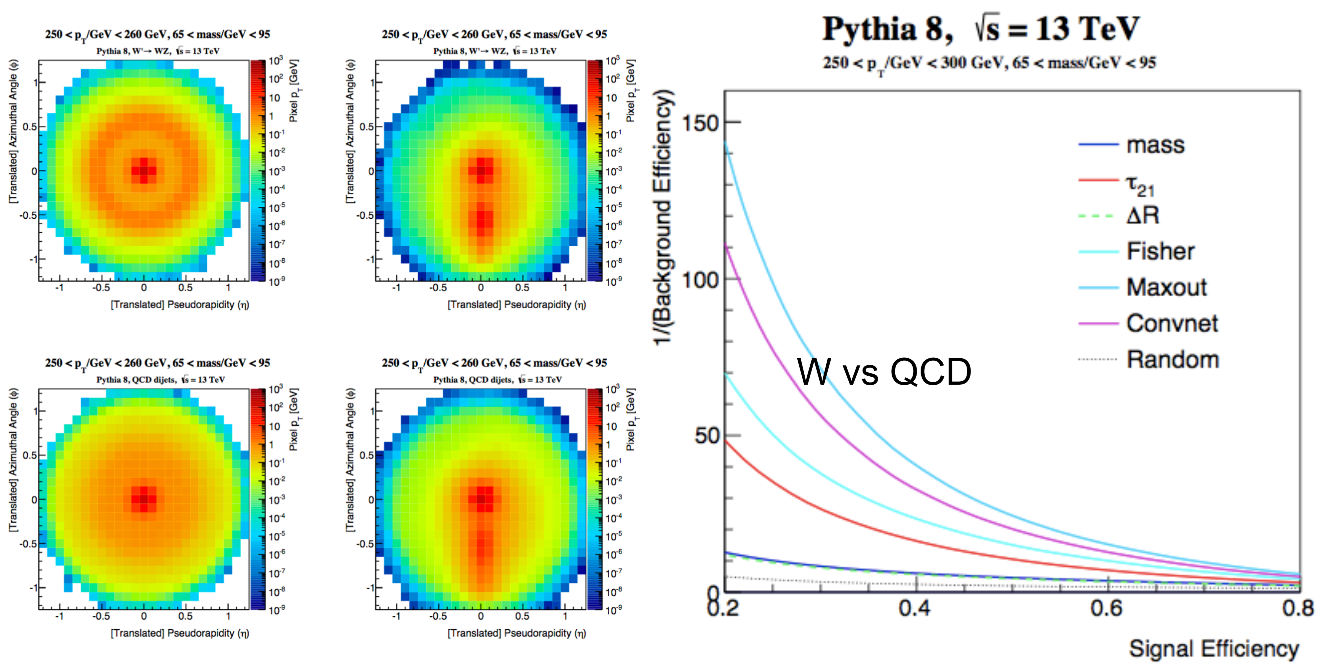
Dimensionality reduction

Globalization of information

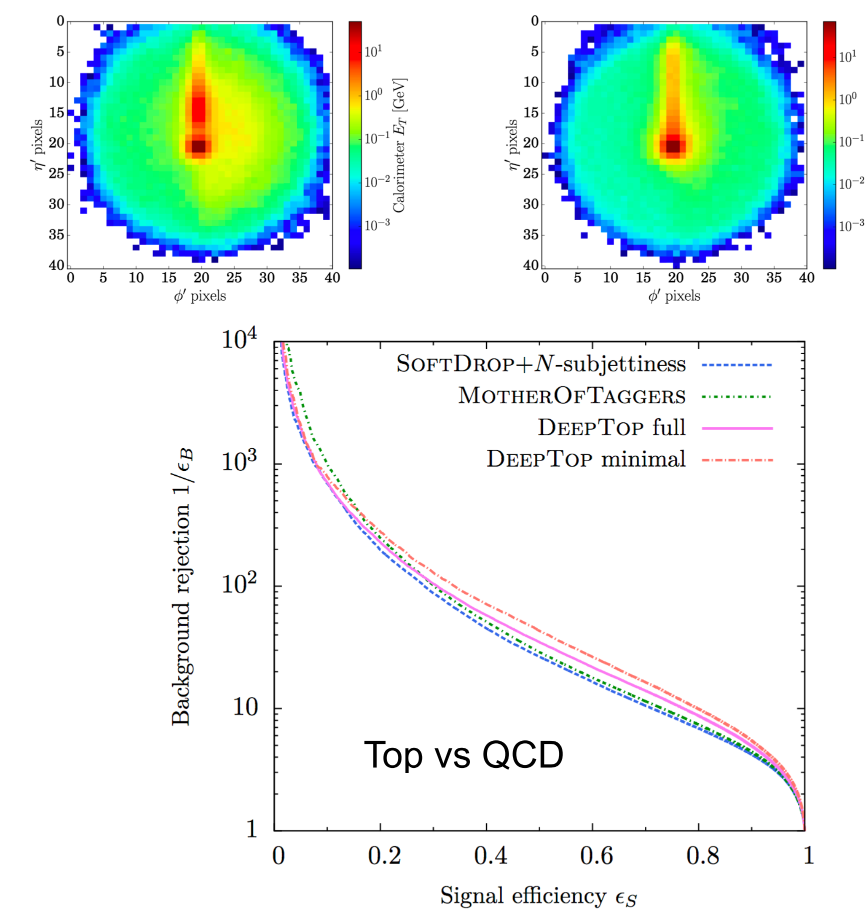
From digital signal, to local hits, to a sequence of objects, and high-level features.
Complex and computing intensive task that could find a match in ML application.

Image Representation

Jet-Images – Deep learning edition
[1511.05190]

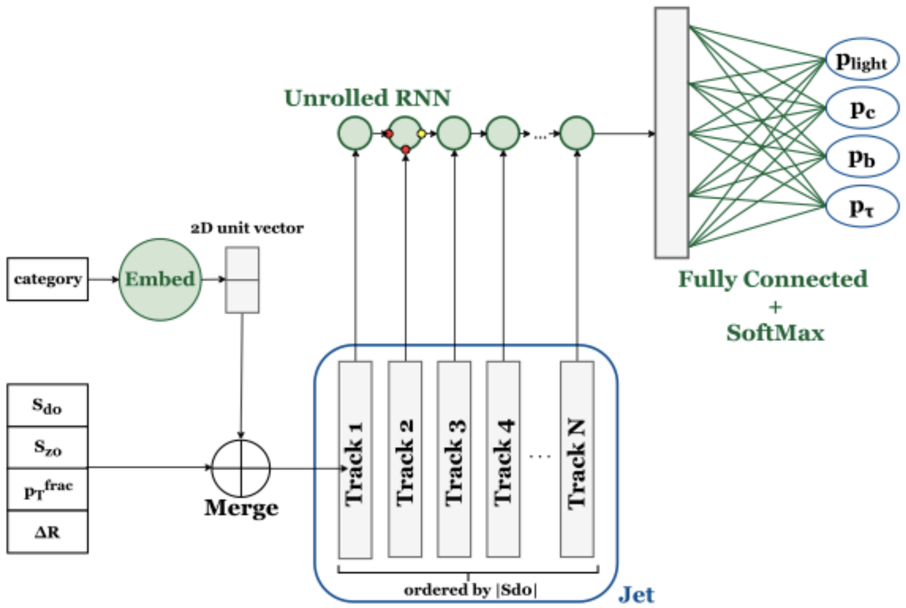


Deep-learning top taggers or the end of QCD?
[1701.08784]

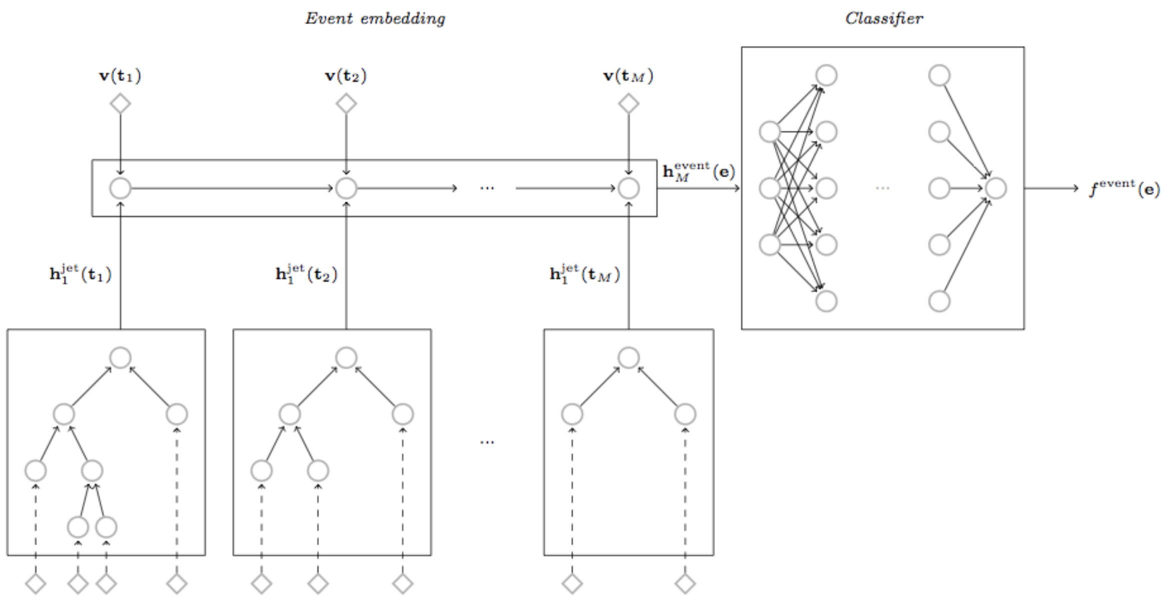


Calorimeter signal are image-like.
Projection of reconstructed particle properties onto images possible.
Potential loss of information during projection.

Sequence Representation



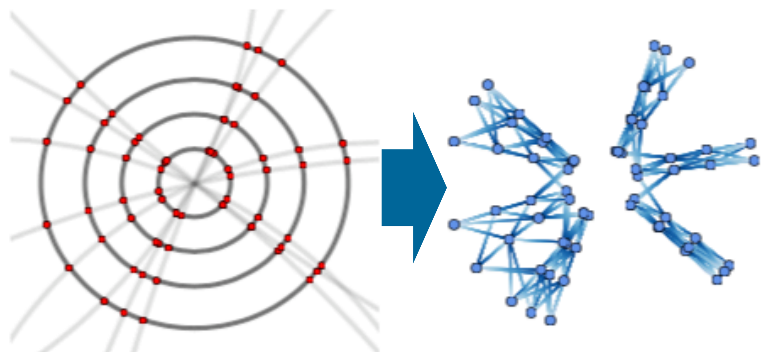
B-Jet with Recurrent Neural Networks
[cds:2255226]



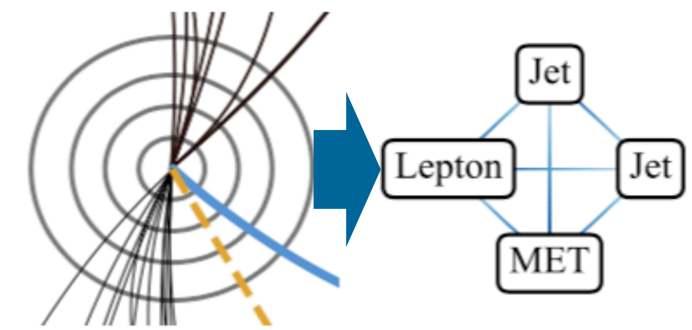
QCD-Aware Recursive Neural Networks for Jet Physics.
[1702.00748]

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

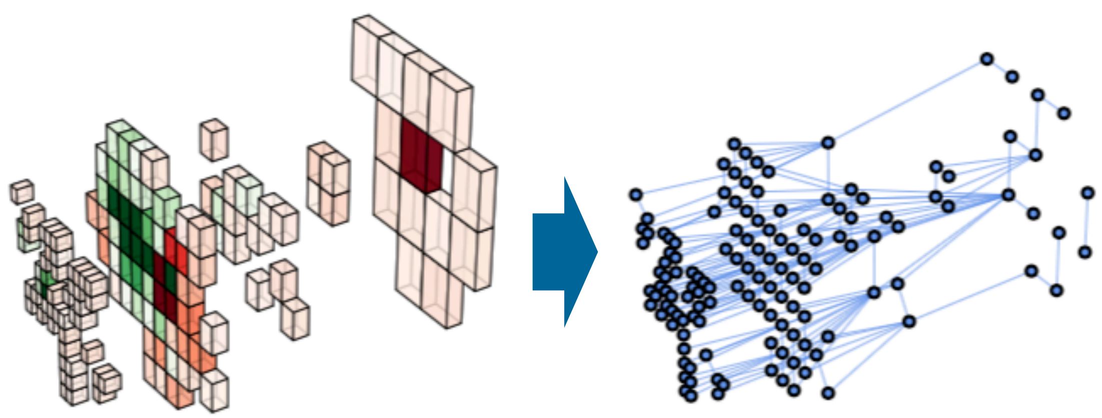
Graph Representation



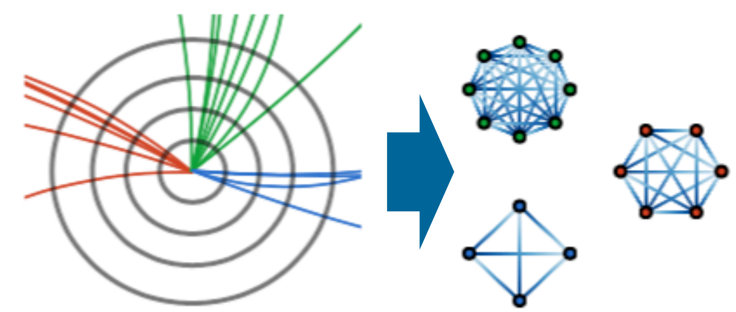
Hits in tracking detector



Objects in an event



Hits in calorimeter detector



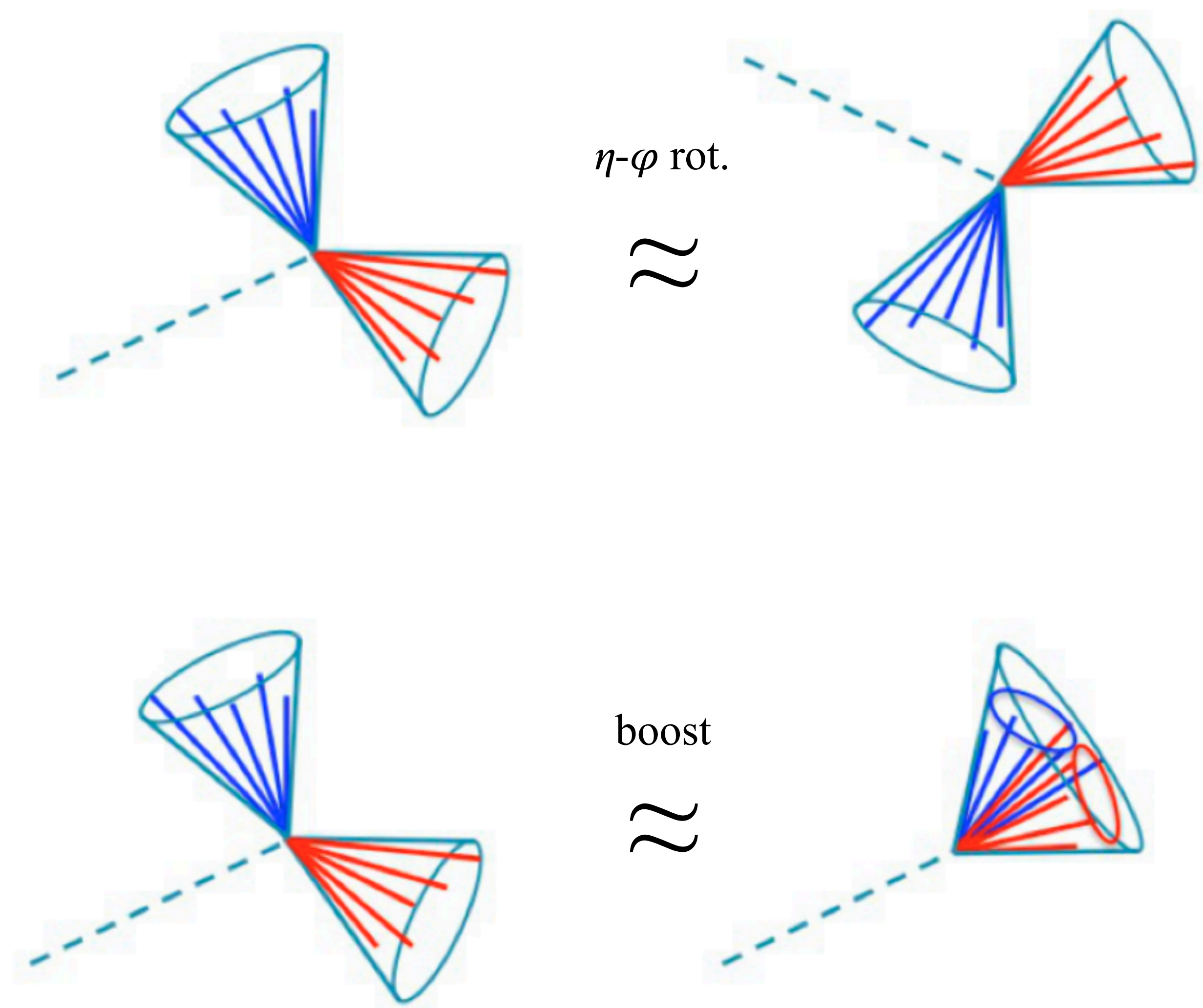
Object sub-structure in an event

Graph Neural Networks in Particle Physics
[2007.13681]

Heterogenous data fits well in graph/set representation.

Invariance and Symmetries

Dataset Degeneracy



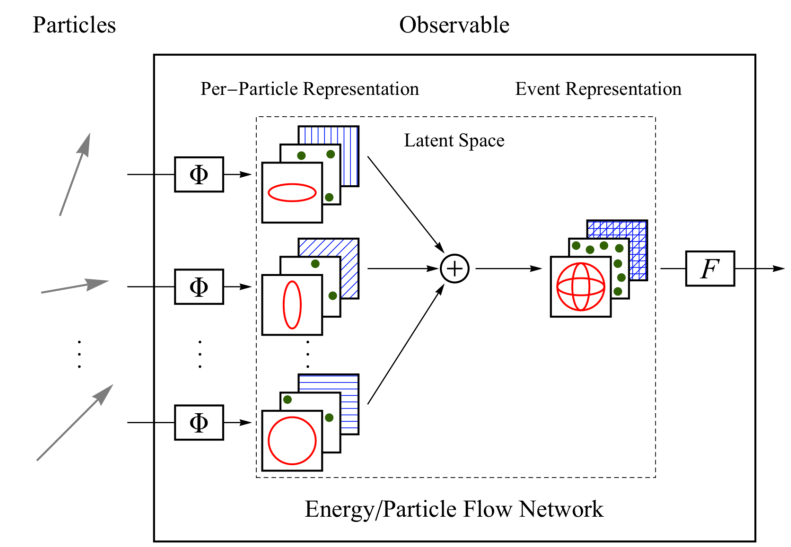
Pre-process the dataset to reduce degeneracy.
Model training improves as the invariance does not have to be learned.

Inductive Bias

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

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

Lorentz Learning Layer
[1707.08966]

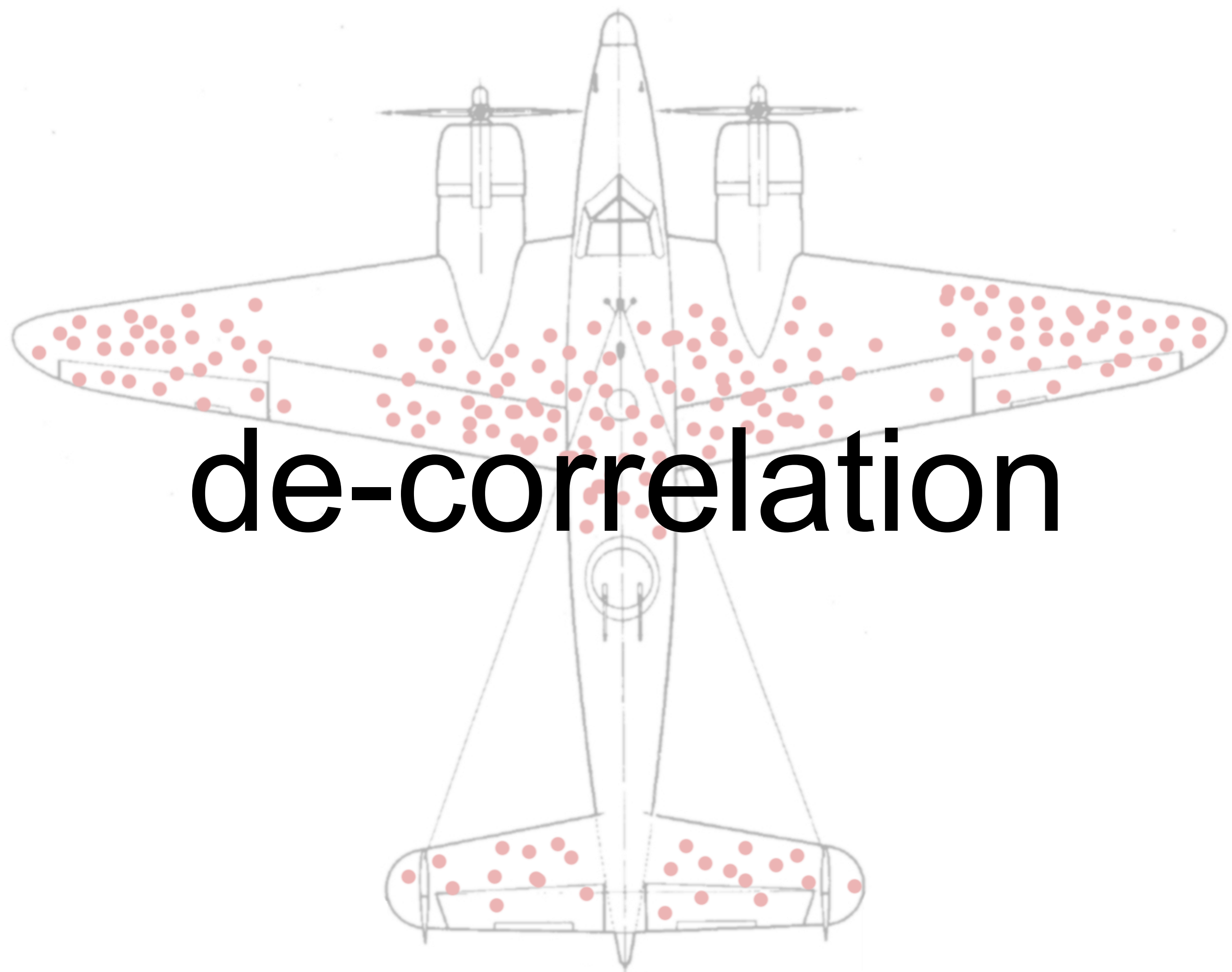


Deep set
[1810.05165]

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

Lorentz group quivariant networks
[2006.04780]

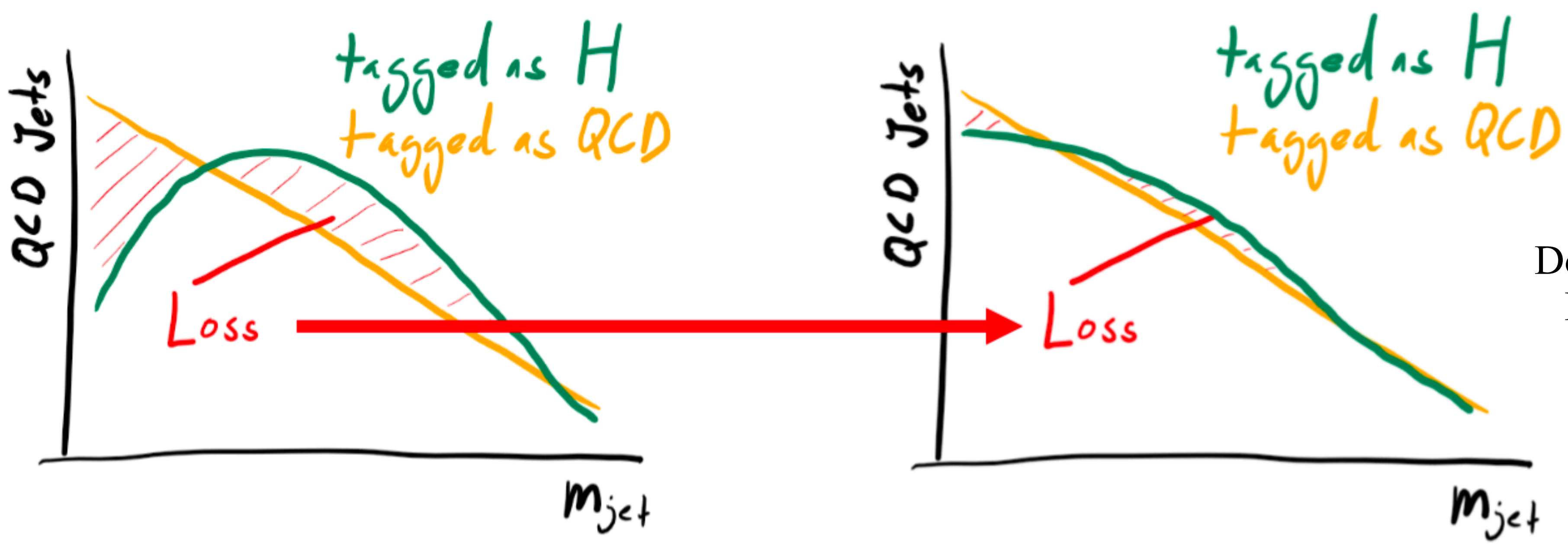
Embed the symmetry and invariance in the model.
Economy of model parameters.



de-correlation

De-correlation

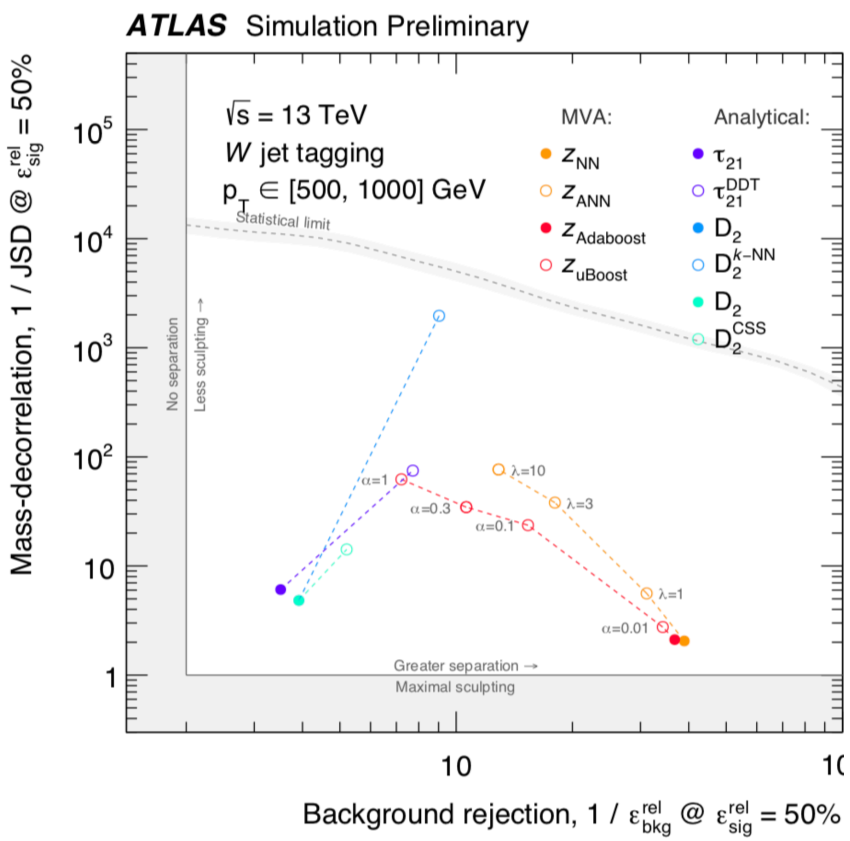
Most background estimation methods (side-bands, ABCD, parametrized fit, …) will require background shape to somehow be independent of analysis selections/processing (not only when using machine learning BTW).



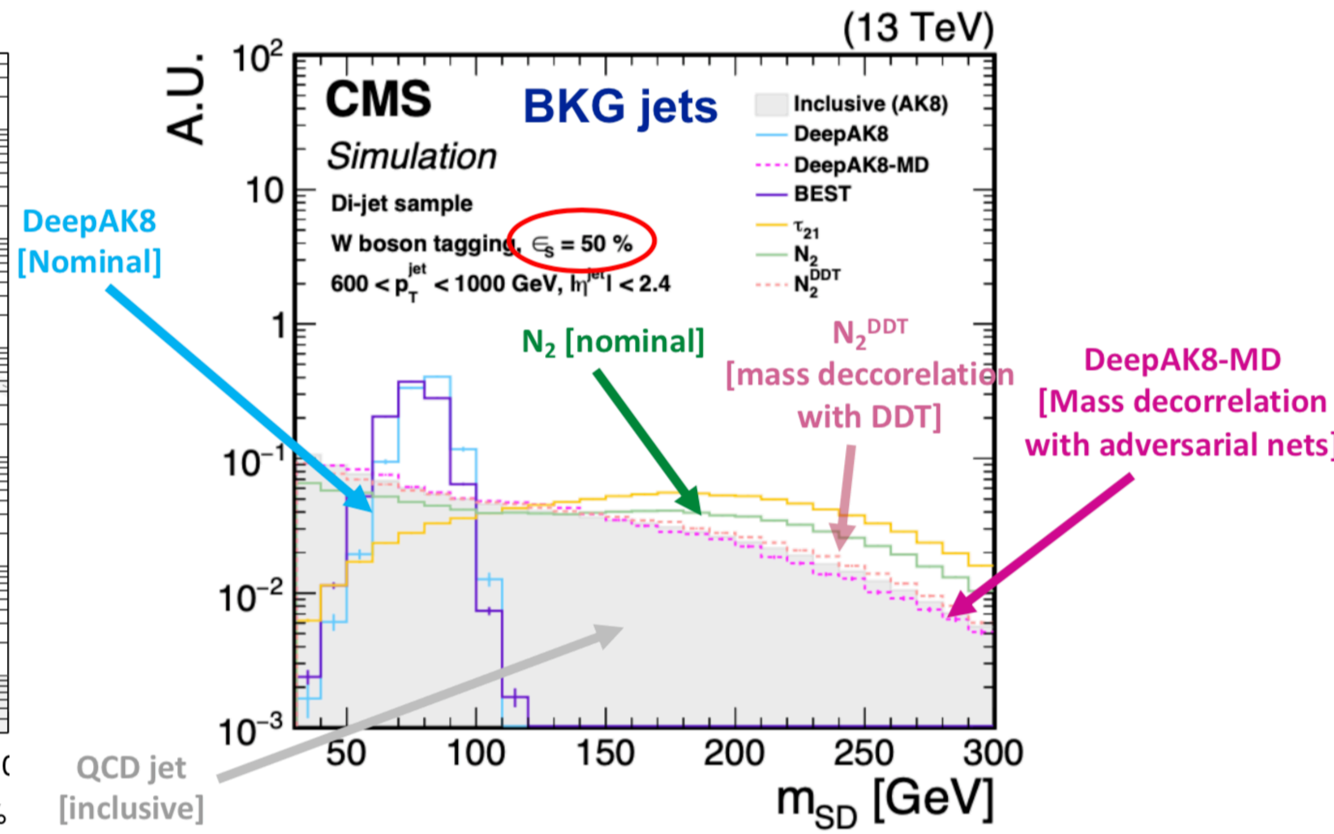
Domain adaptation [1409.7495]
Learn to Pivot [1611.01046]

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.

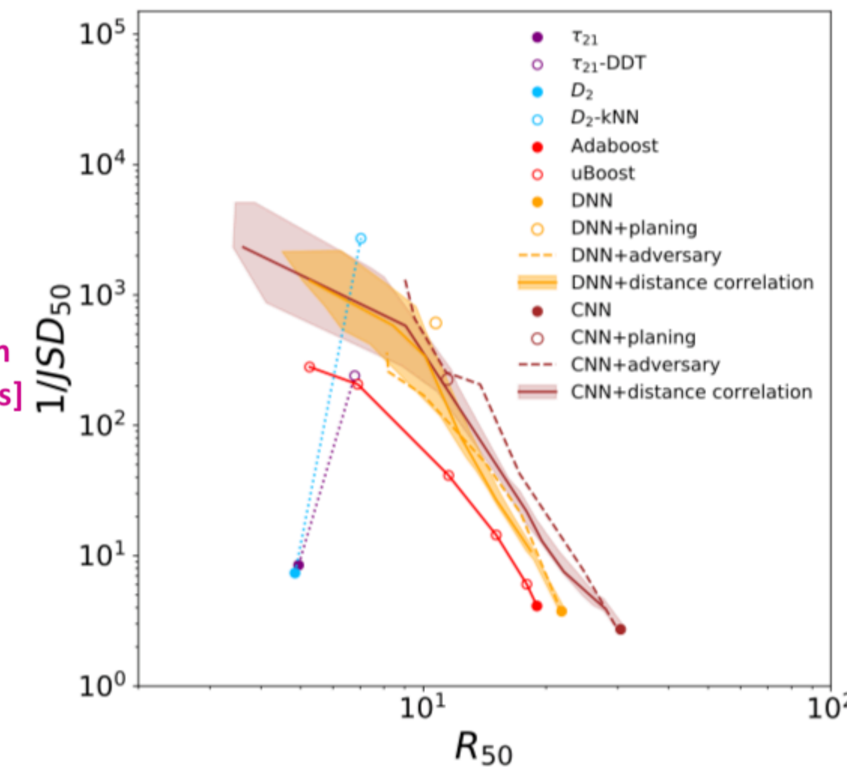
Performance



ATLAS Collab. [cds:2630973]



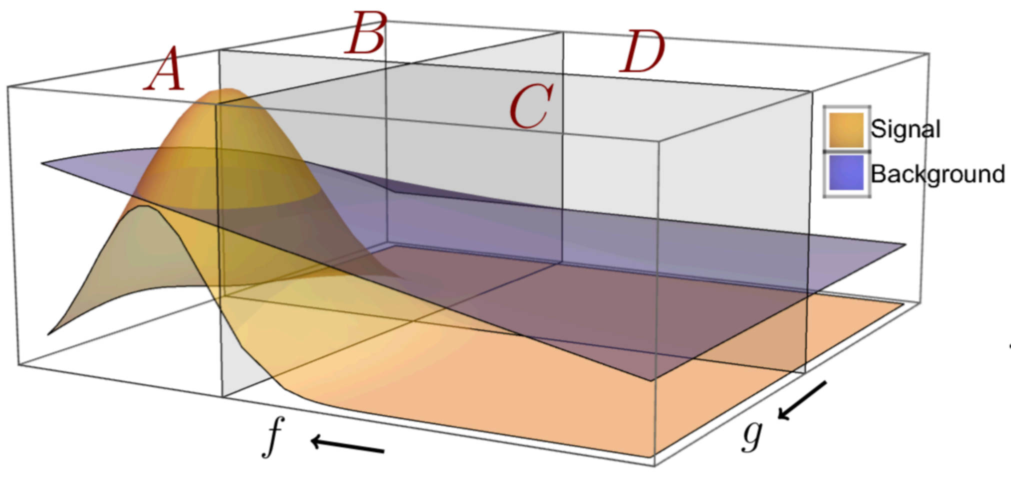
CMS Collab. [doi:10.1088/1748-0221/15/06/P06005]

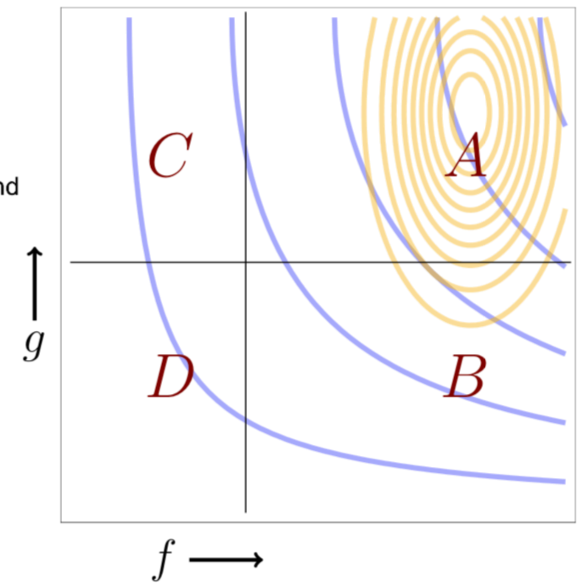


DISCO: Distance Correlation [2001.05310]

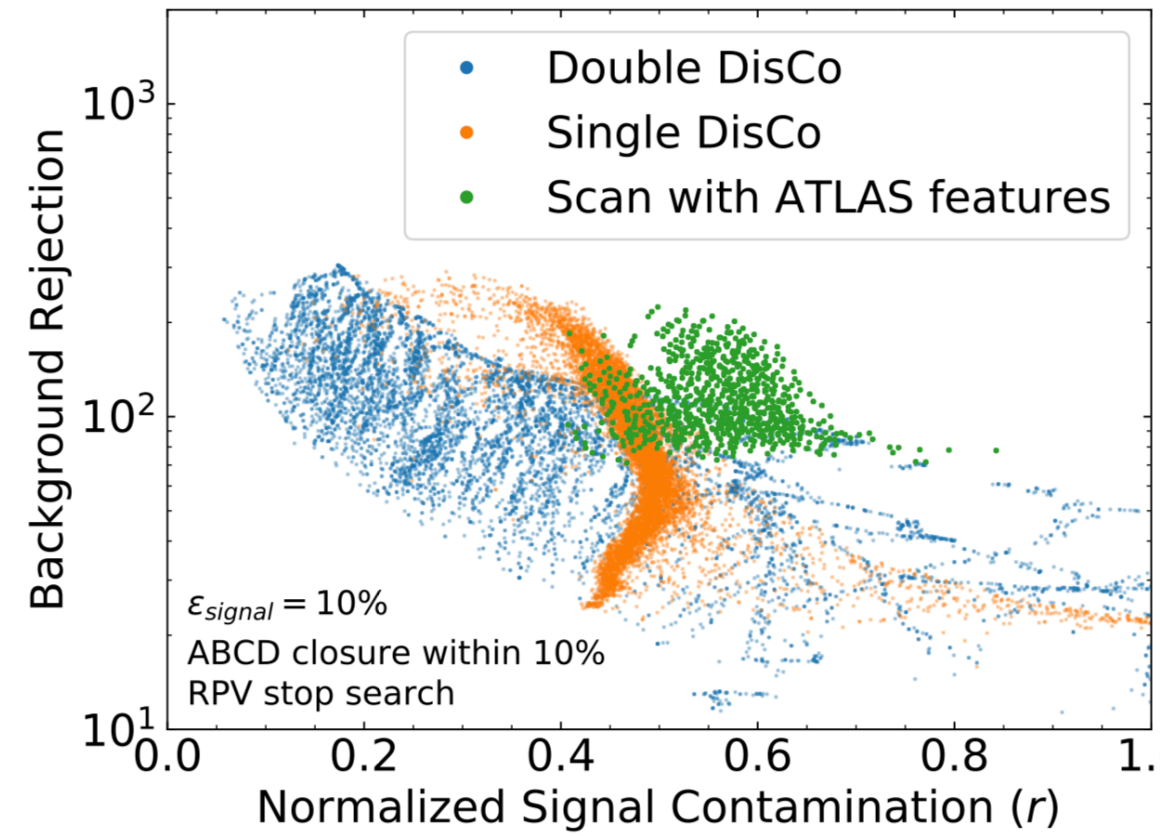
Jenson-Shannon Divergence (JSD) as the comparison metric for shaping.
Residual shaping needs to enter systematics uncertainty estimation.

Background Estimation



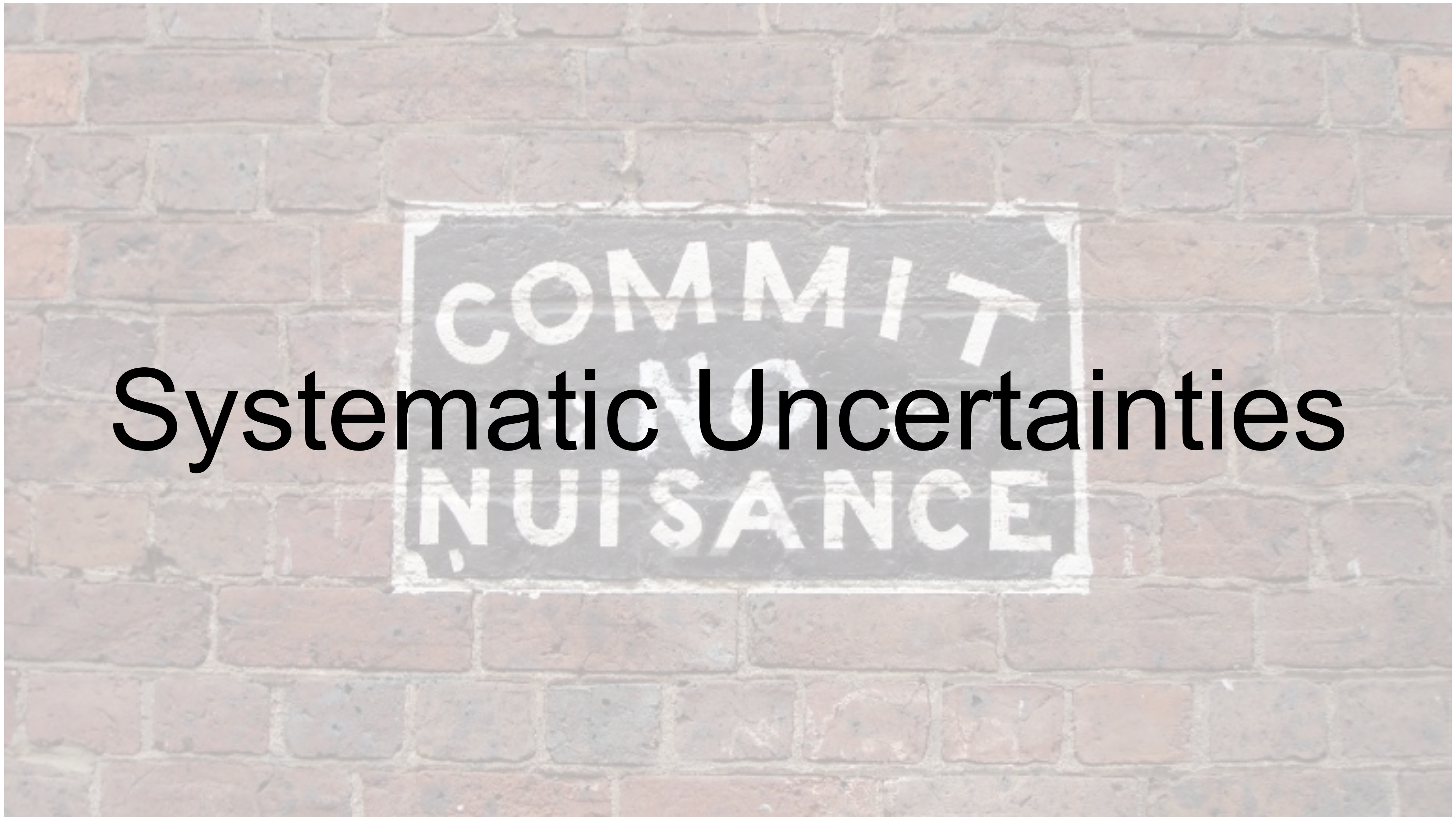


ABCD + Disco
[2007.14400]

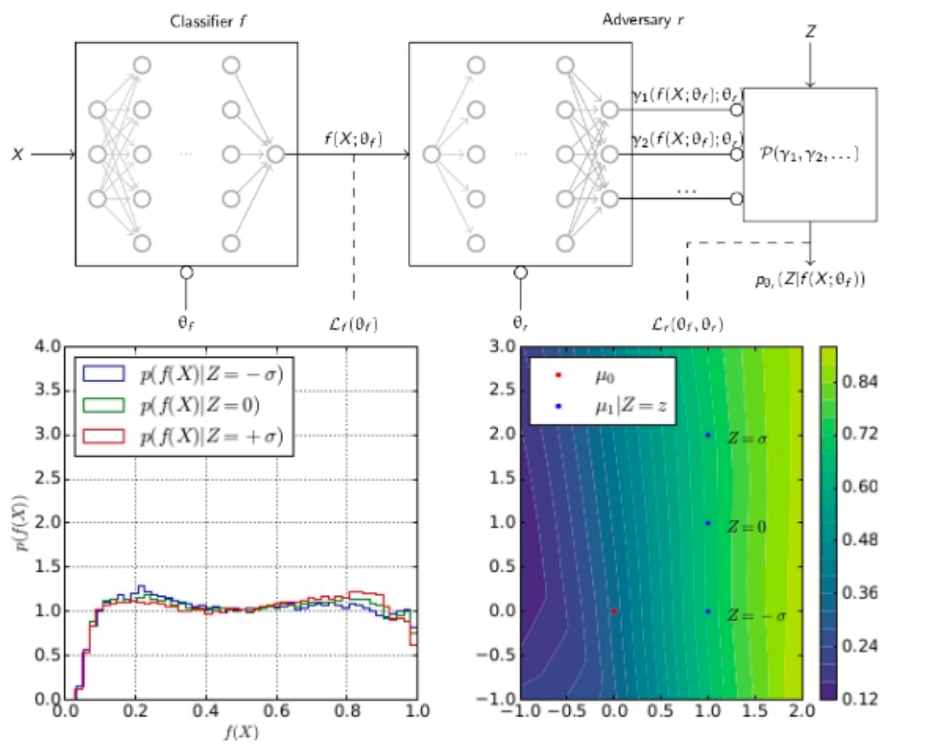


Most popular background estimation method (ABCD), can be optimized for de-correlation, yielding increased significance.

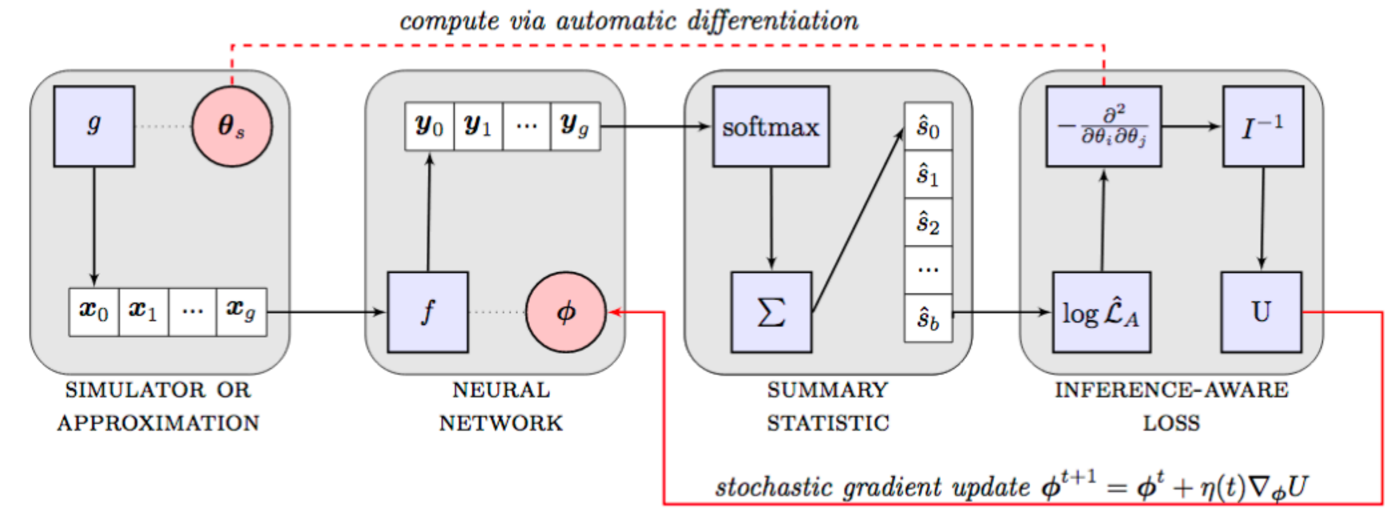
Systematic Uncertainties



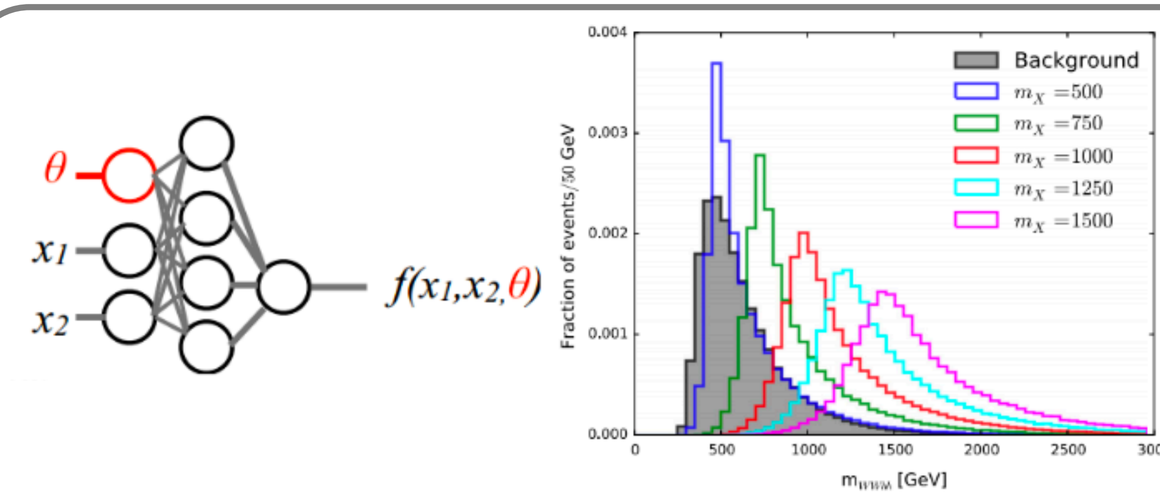
Syst. Estimation and Mitigation



Learn to pivot [1611.01046]



INFERNO: Inference-Aware Optimisation [1806.04743]

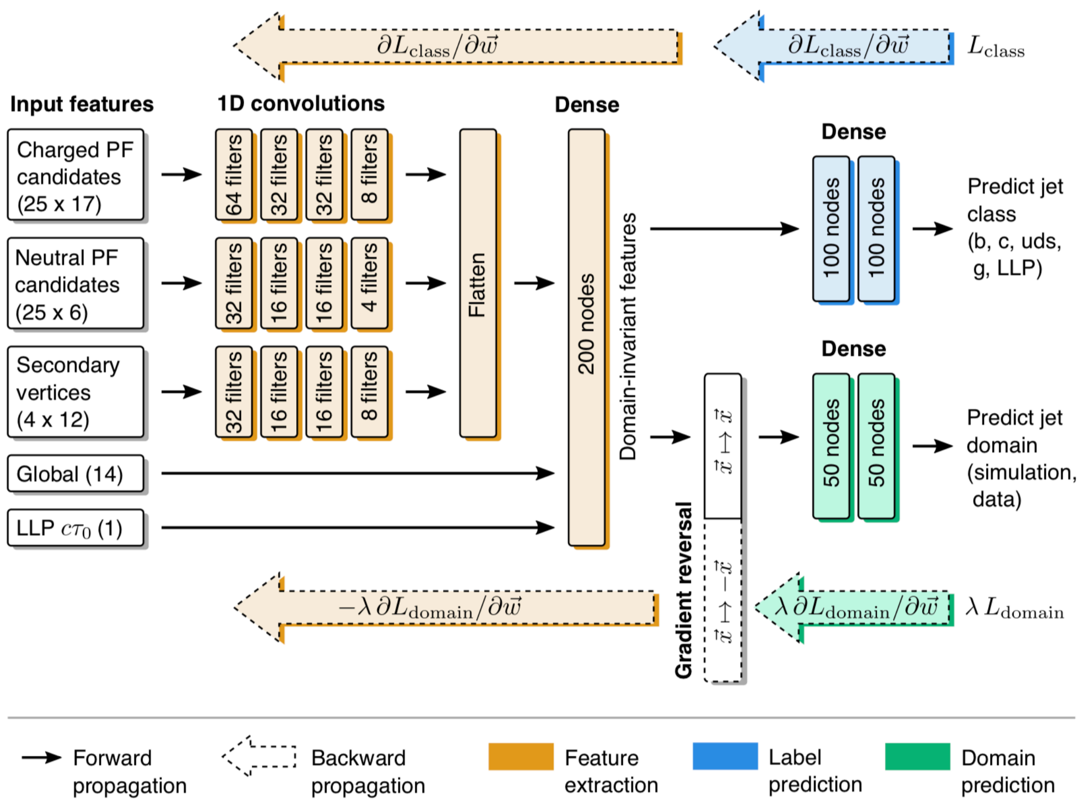


Parametrized Learning [1601.07913]

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

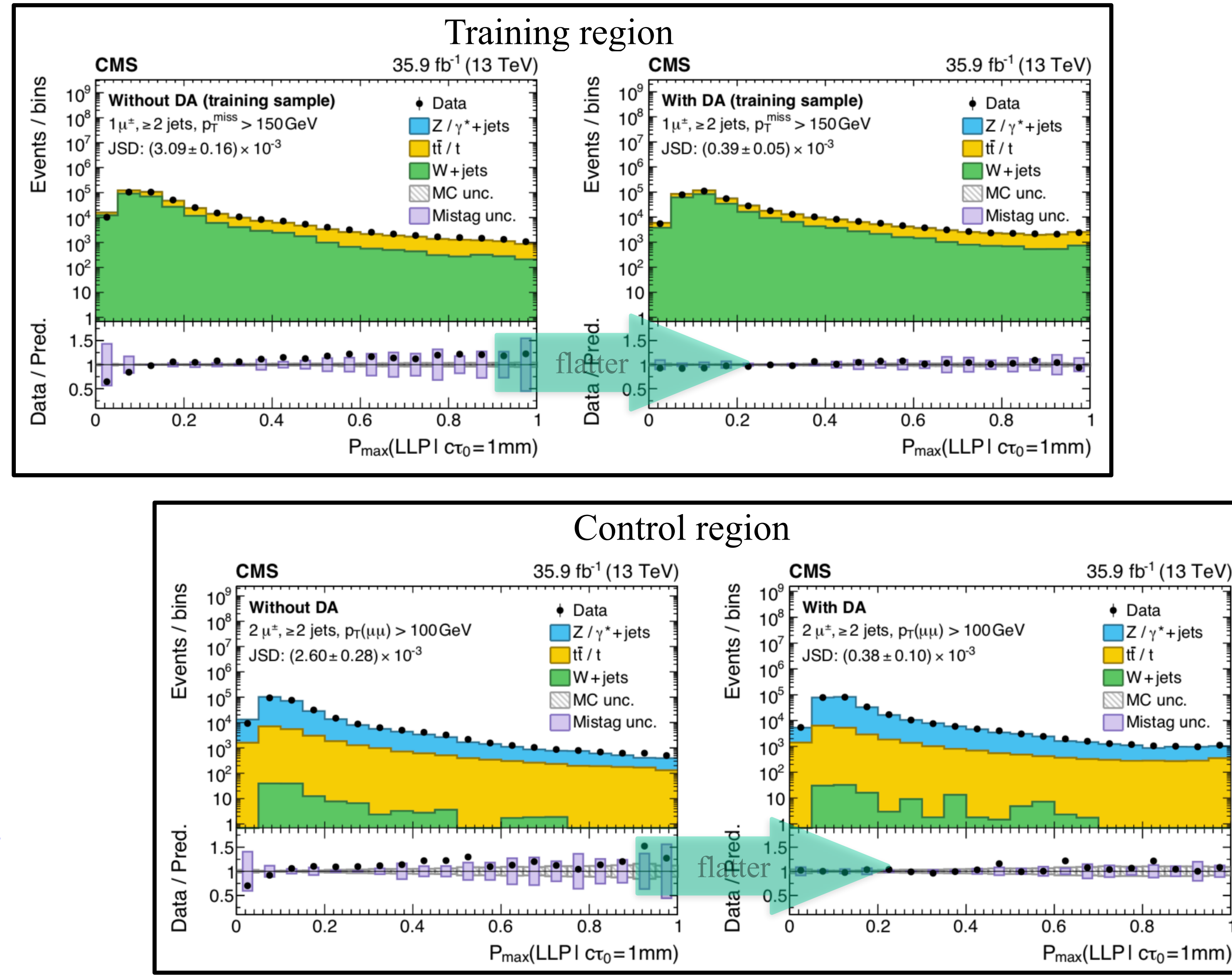
Domain Dependence

Domain in-Dependence

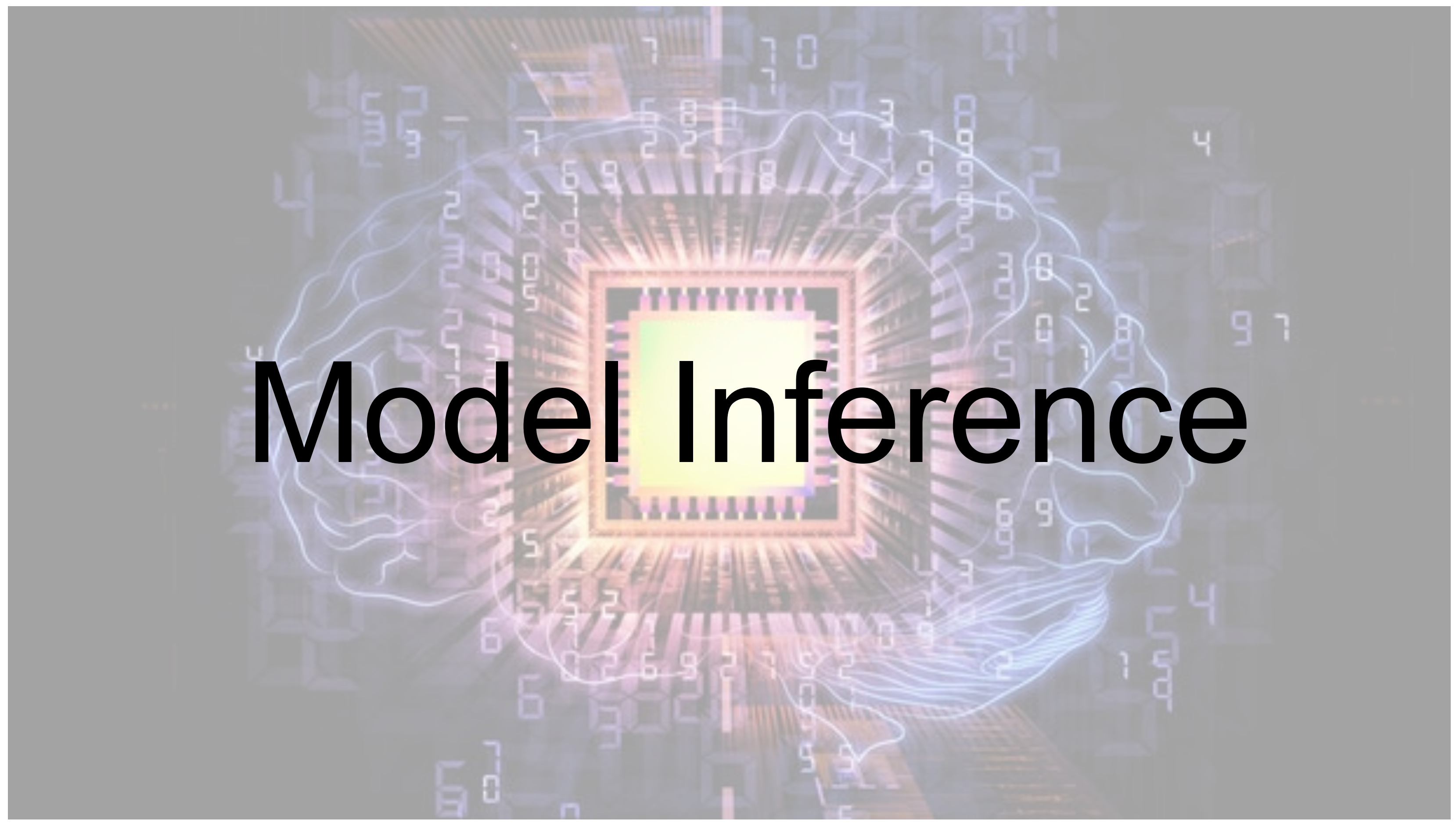


LLP jet tagger

[doi:10.1088/2632-2153/ab9023]



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



Model Inference

Inference Engines

"On-Board accelerator"

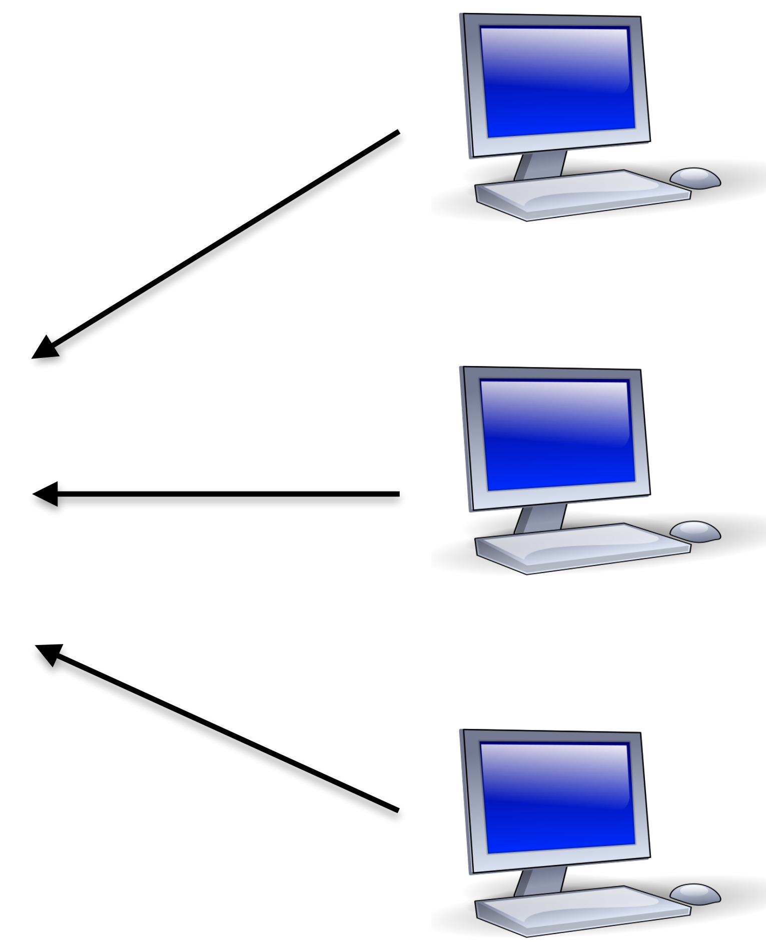
CPU
- Small models
- Small datasets
- Useful for design space exploration

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"

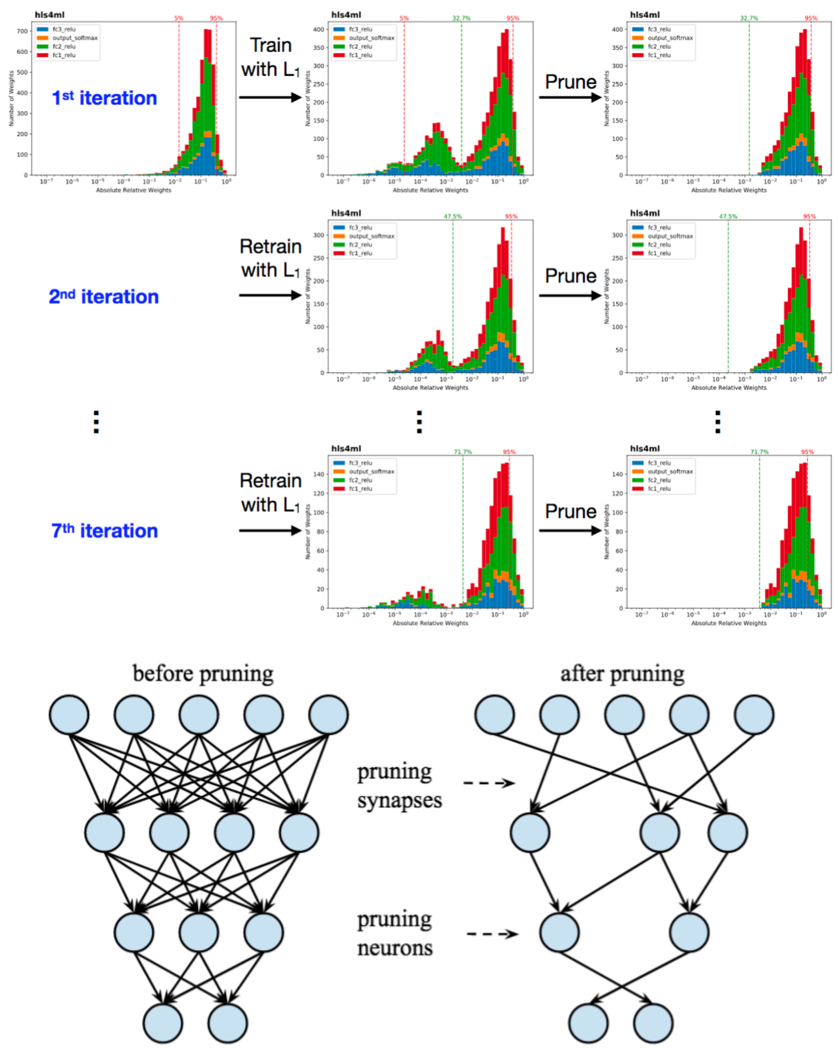


https://arxiv.org/abs/1811.04492
https://arxiv.org/abs/2007.10359
https://arxiv.org/abs/2007.14781

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

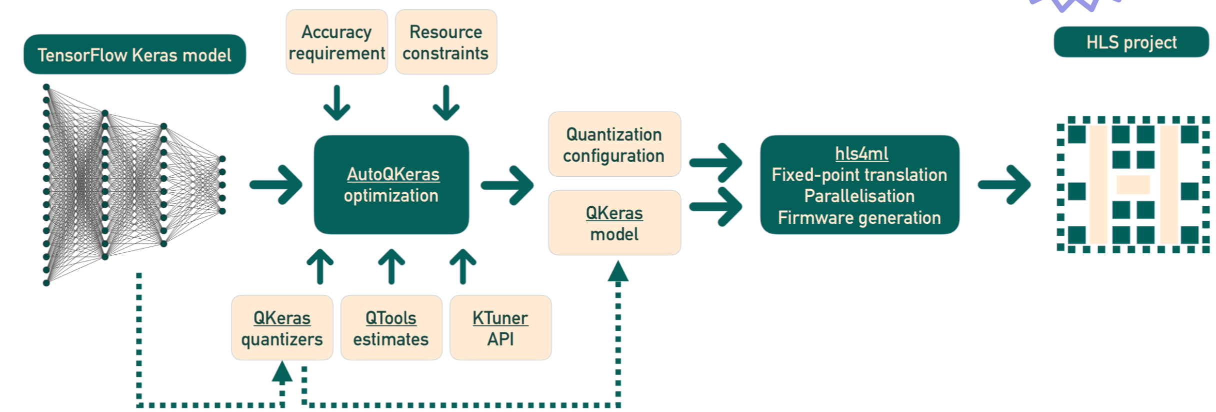
Model Compression

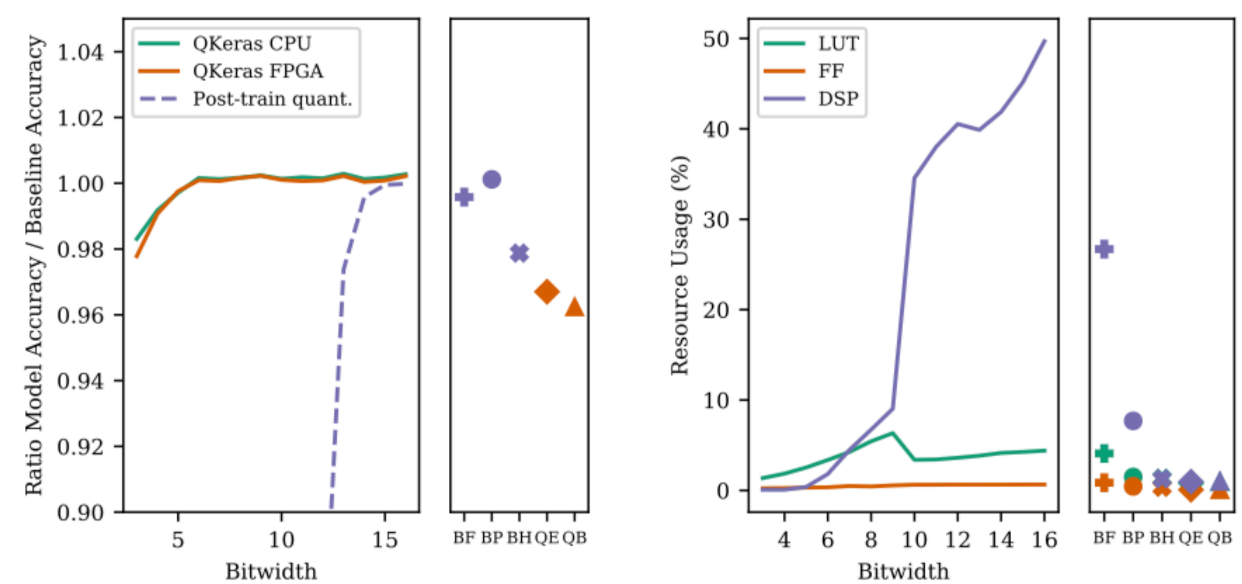
Pruning weights [1804.06913]



Quantization [2006.10159]

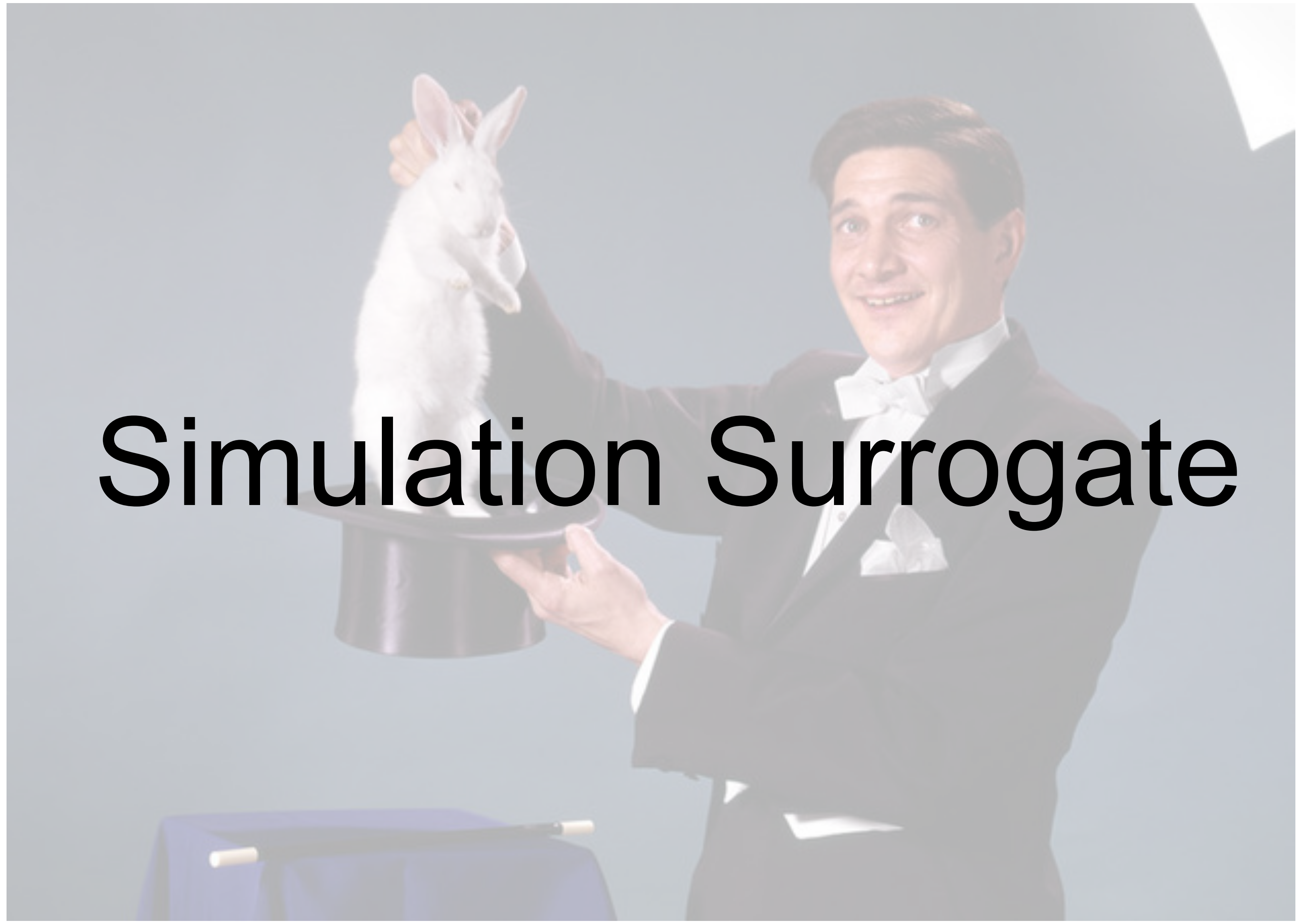
Hands-on



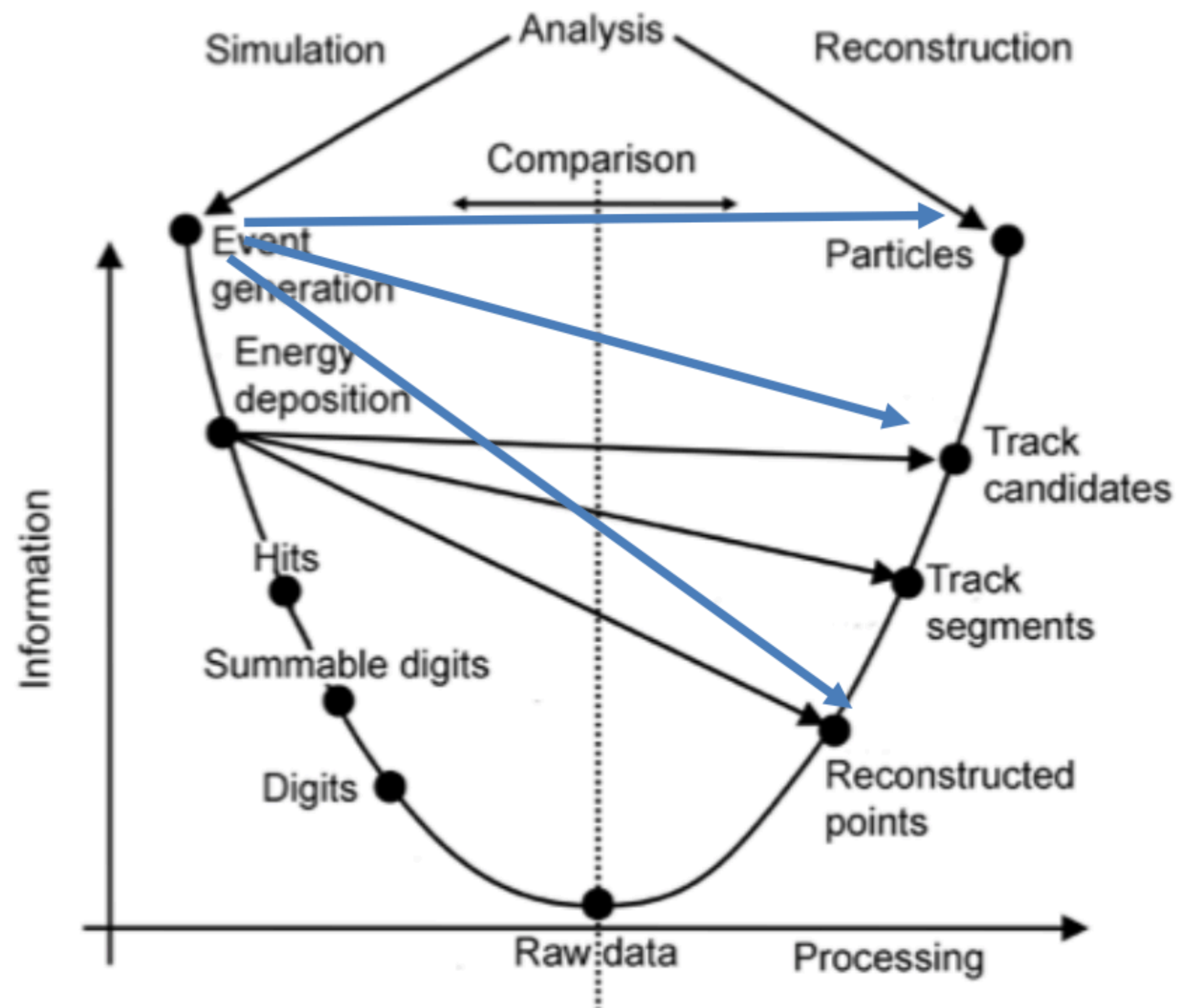


Model inference can be accelerated by reducing the number and size of operations.

Simulation Surrogate

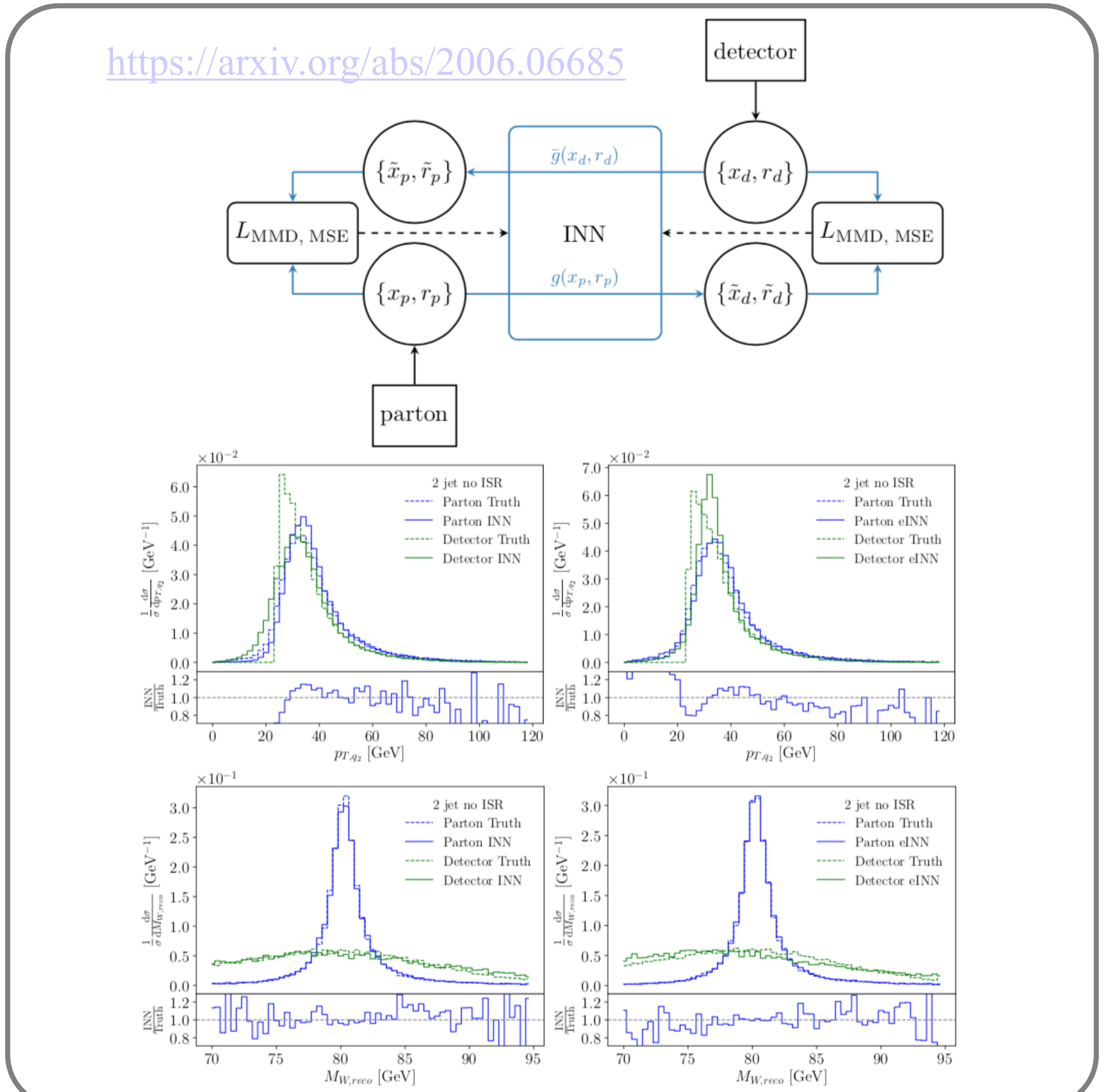


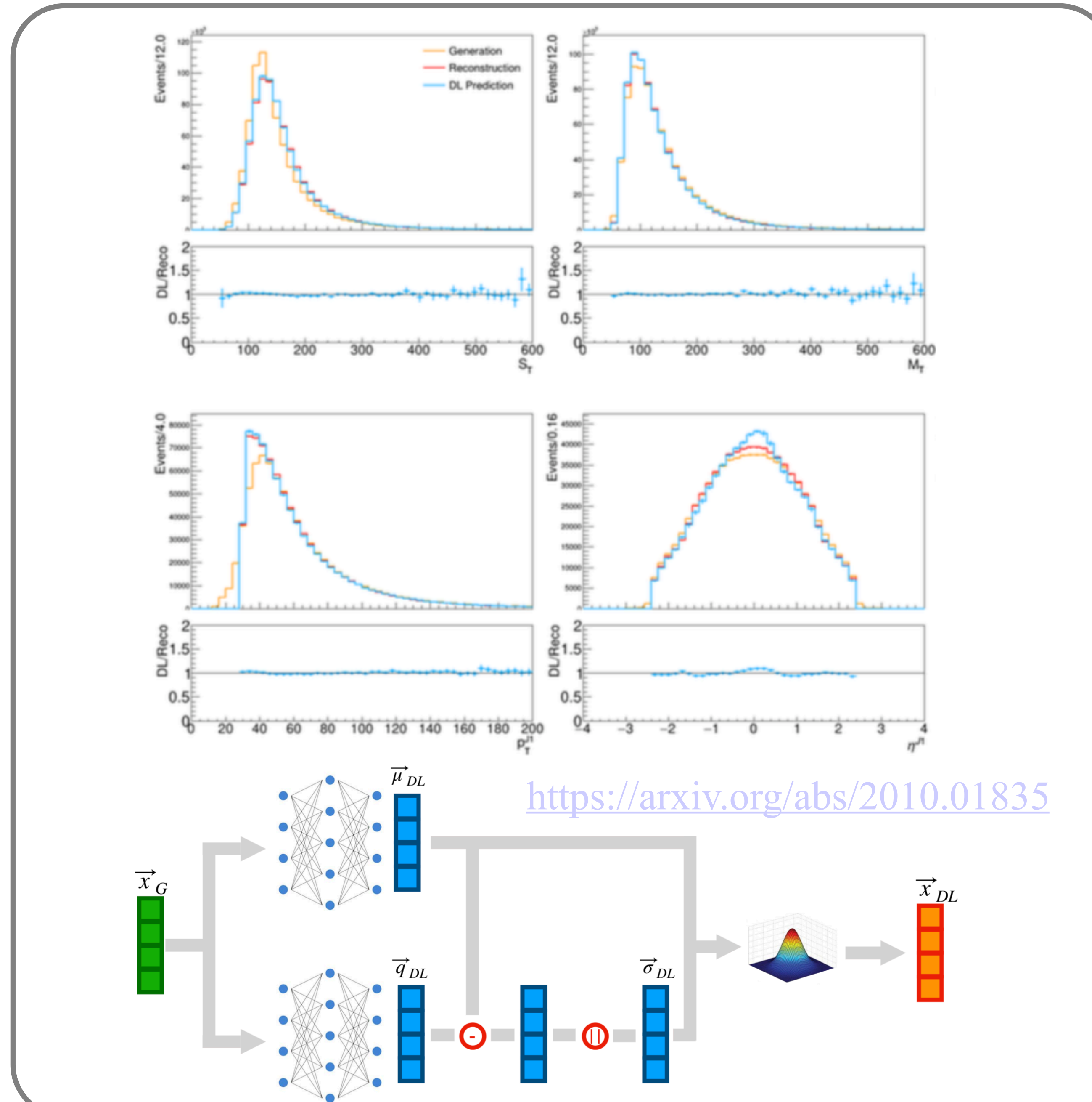
Reconstruction ∘ Simulation ~ Identity



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

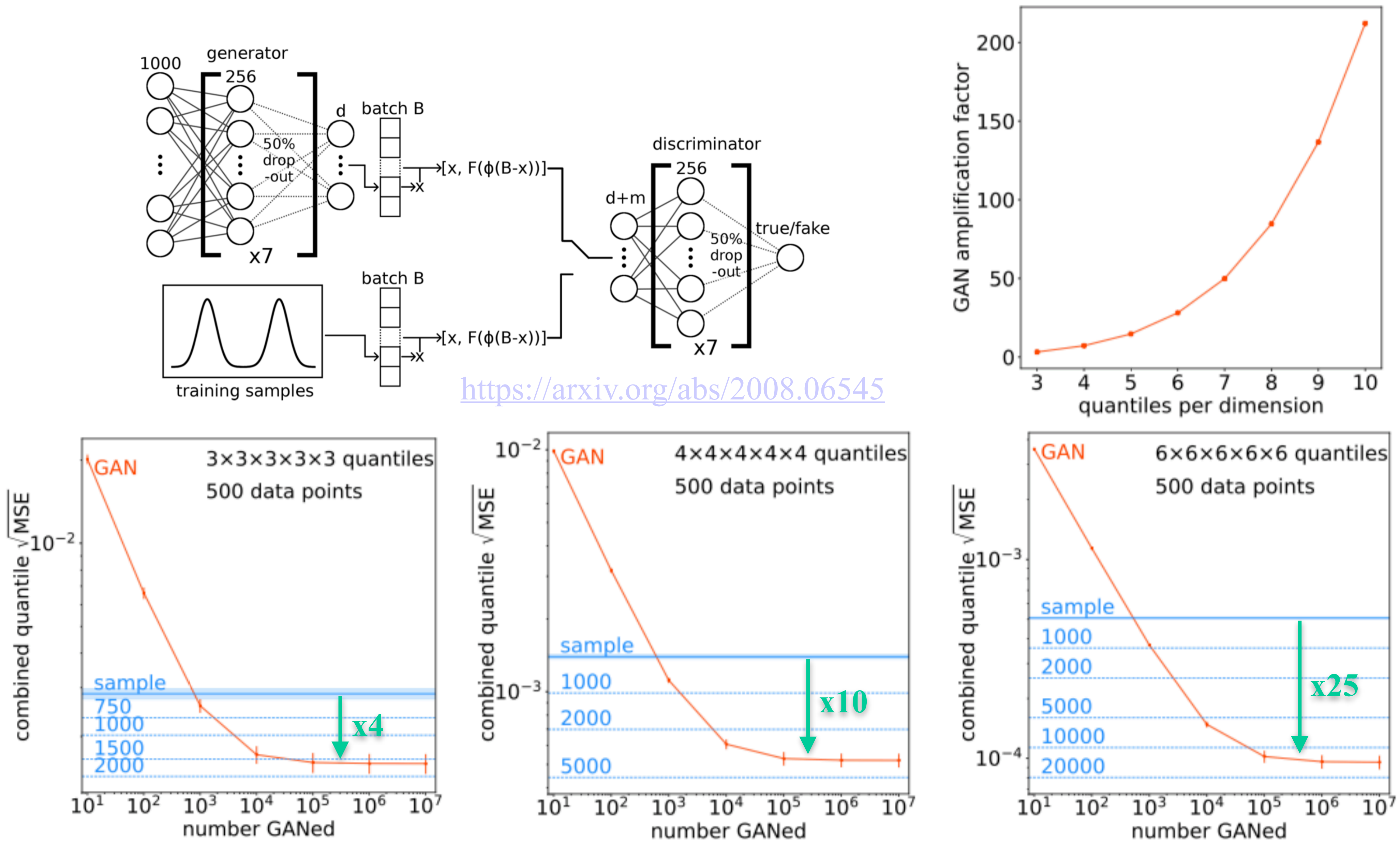
Suiting Models





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

Statistical Power



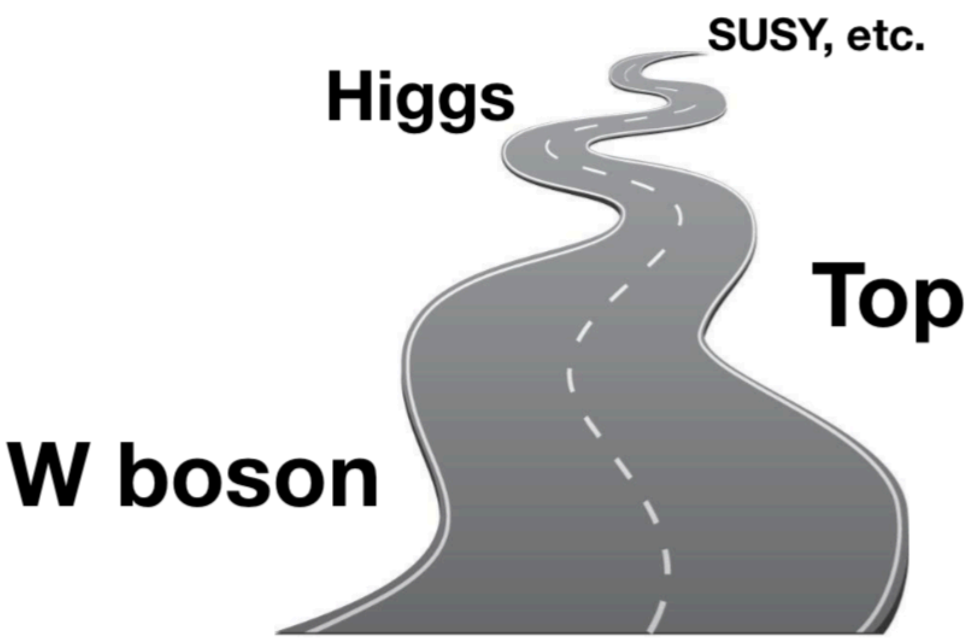
Generative adversarial network may help producing samples with higher statistical power than the one used for training.

Anomaly Search

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**

"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

3 Unsupervised

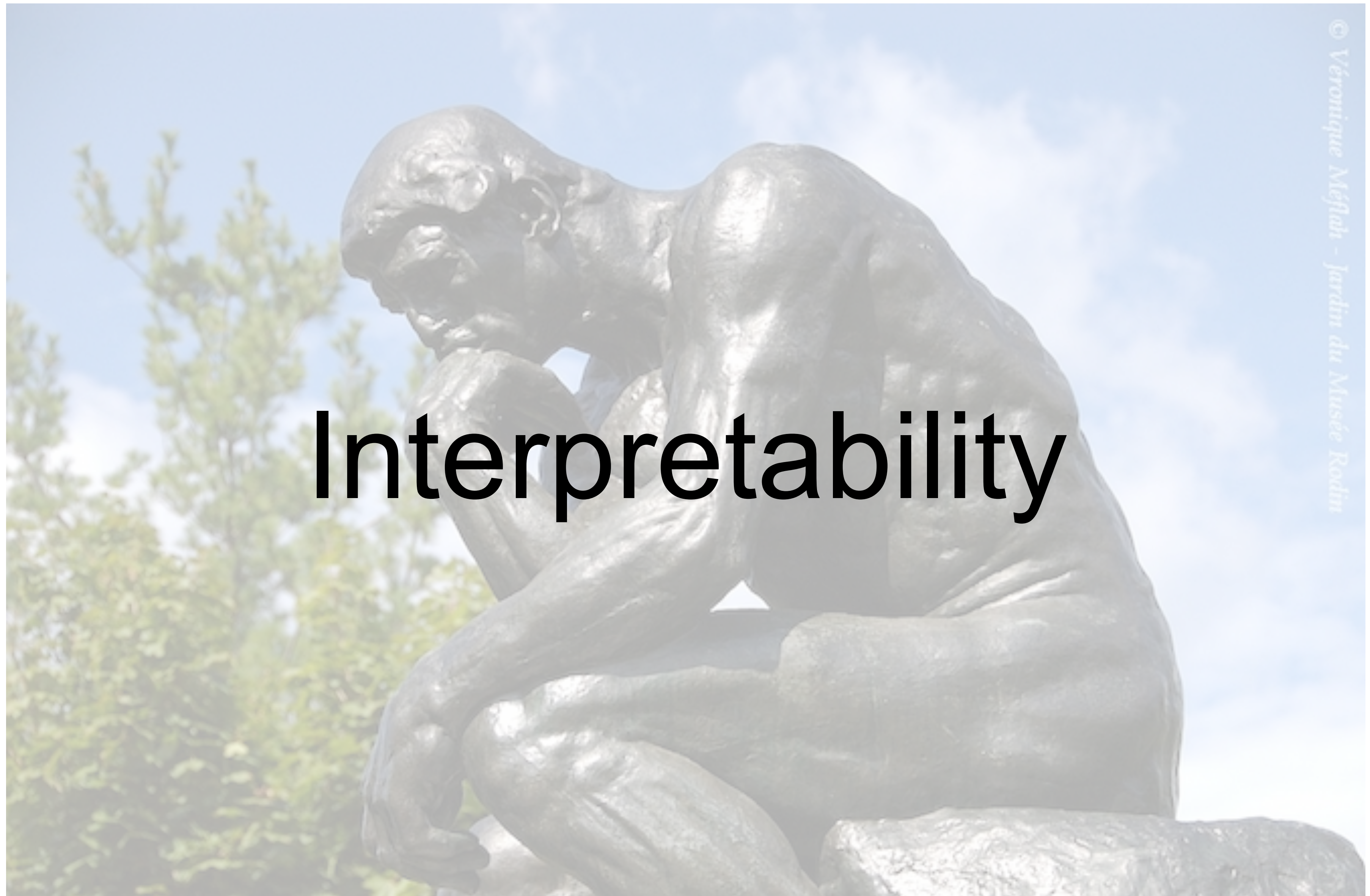
3.1 Anomalous Jet Identification via Variational Recurrent Neural Network
3.2 Anomaly Detection with Density Estimation
3.3 BuHuLaSpa: Bump Hunting in Latent Space
3.4 GAN-AE and BumpHunter
3.5 Gaussianizing Iterative Slicing (GIS): Unsupervised In-distribution Anomaly Detection through Conditional Density Estimation
3.6 Latent Dirichlet Allocation
3.7 Particle Graph Autoencoders
3.8 Regularized Likelihoods
3.9 UCluster: Unsupervised Clustering

4 Weakly Supervised

4.1 CWoLa Hunting
4.2 CWoLa and Autoencoders: Comparing Weak- and Unsupervised methods for Resonant Anomaly Detection
4.3 Tag N' Train
4.4 Simulation Assisted Likelihood-free Anomaly Detection
4.5 Simulation-Assisted Decorrelation for Resonant Anomaly Detection

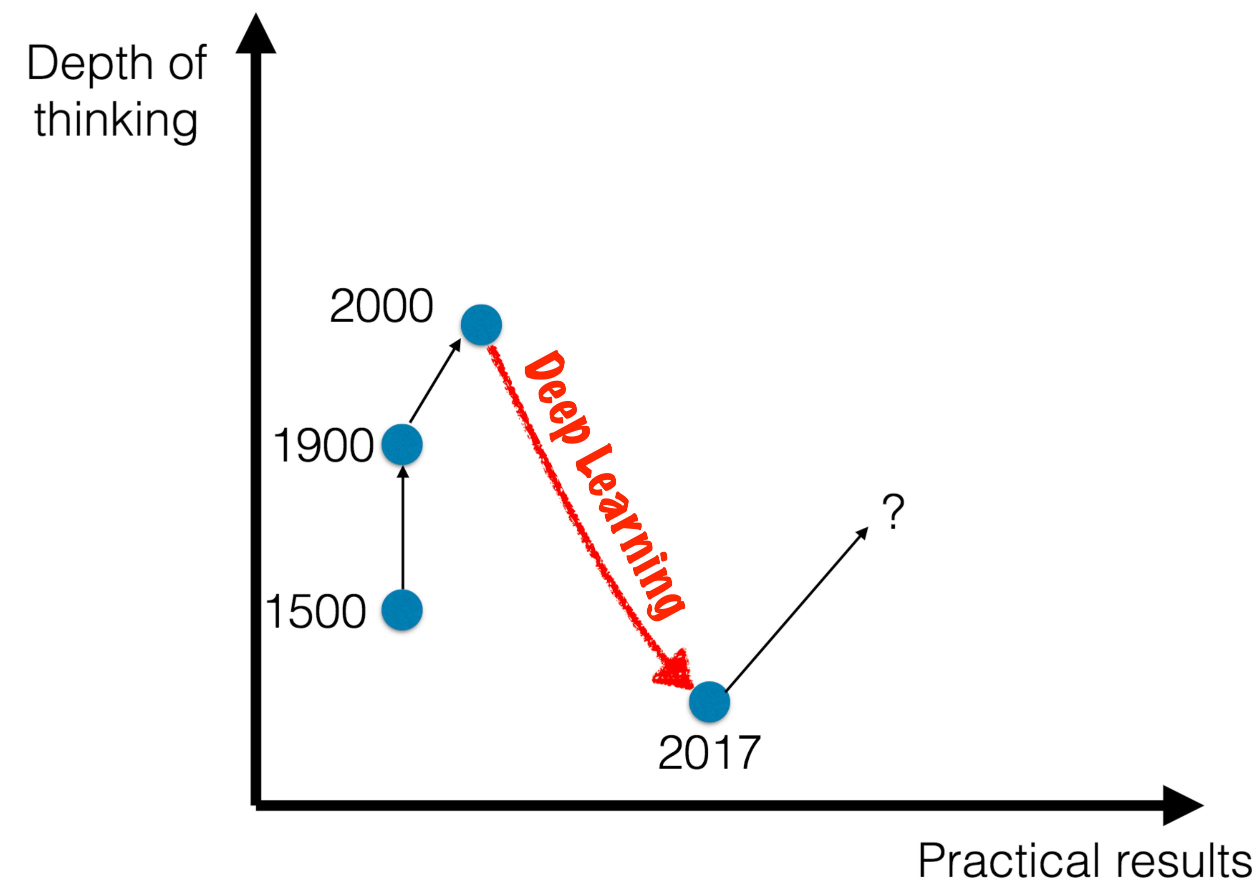
5 (Semi)-Supervised

5.1 Deep Ensemble Anomaly Detection
5.2 Factorized Topic Modeling
5.3 QUAK: Quasi-Anomalous Knowledge for Anomaly Detection
5.4 Simple Supervised learning with LSTM layers



Interpretability
© Véronique Méfflah - Jardin du Musée Rodin

The Black-box Dilemma



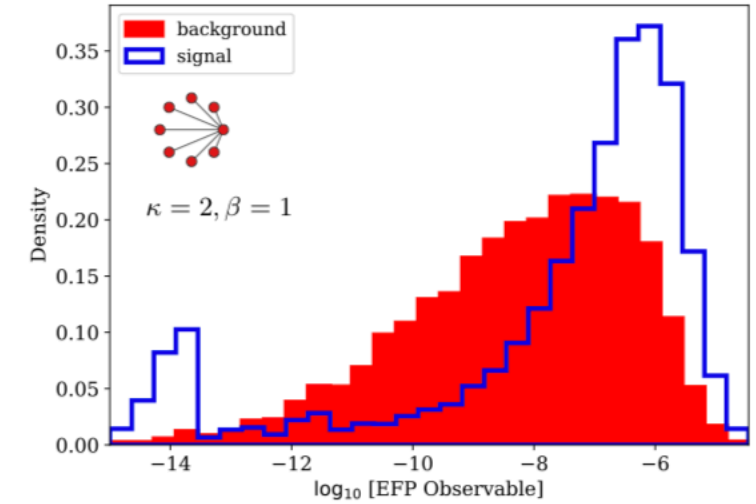
Deep learning may yield great improvements.
Having the “best classification performance” is not always sufficient.
Forming an understand of the processes at play is often crucial.

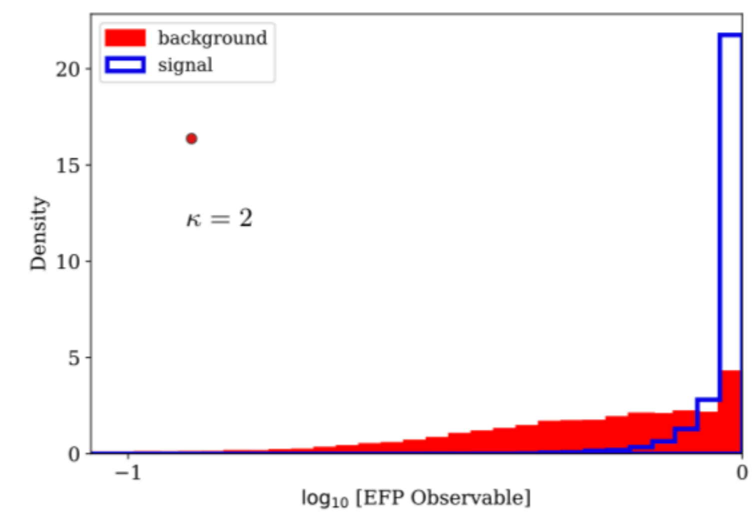
Learning Observables

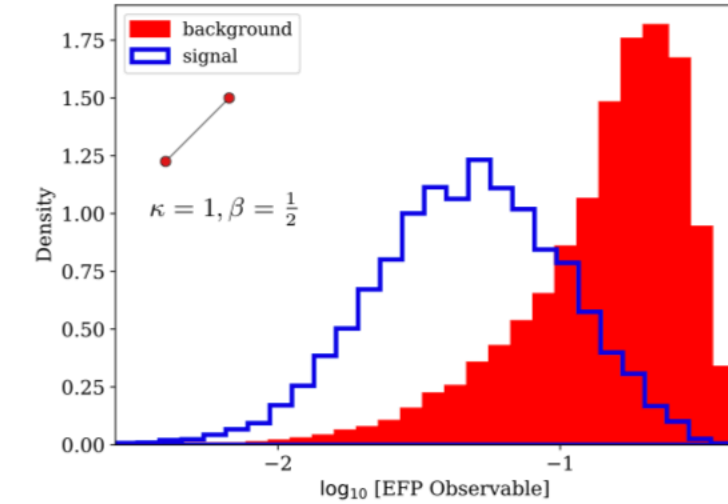
Electron classification performance

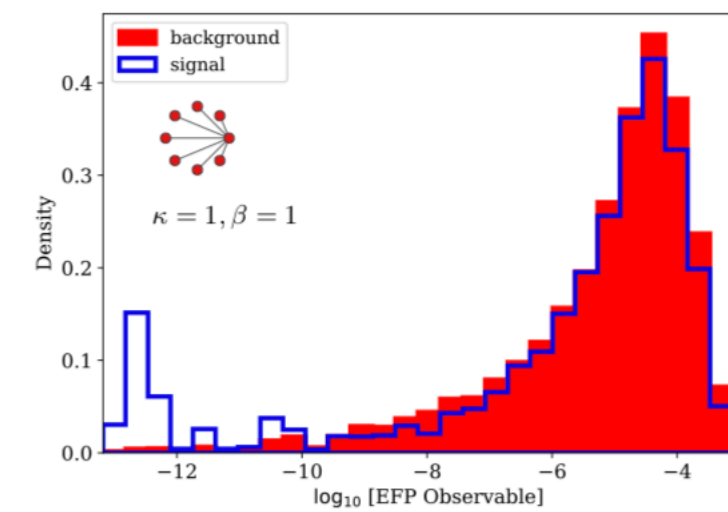
Base	Additions (κ, β)		(AUC)
7HL			0.945
7HL $+M_{\text{jet}}$			0.956
7HL	$(1, \frac{1}{2})$		0.970
7HL $+M_{\text{jet}}$	$(1, 1)$	$(1, \frac{1}{2})$	0.971
7HL	$(2, -)$		0.970
7HL $+M_{\text{jet}}$	$(2, 1)$	$(2, -)$	0.971
CNN			0.972

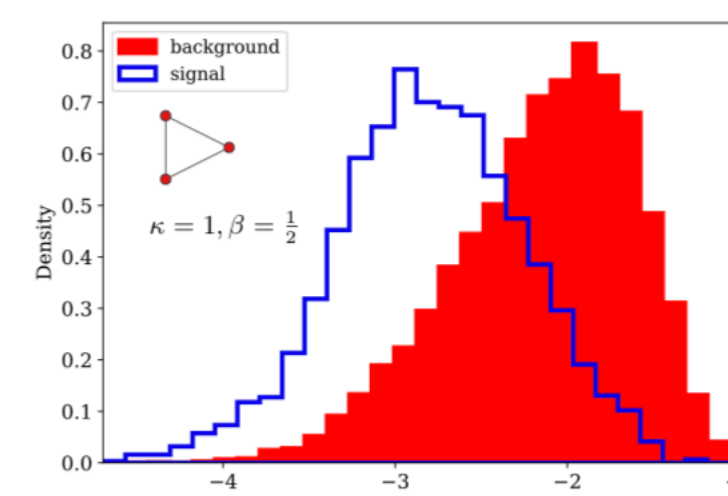
https://arxiv.org/abs/2010.11998
https://arxiv.org/abs/2011.01984







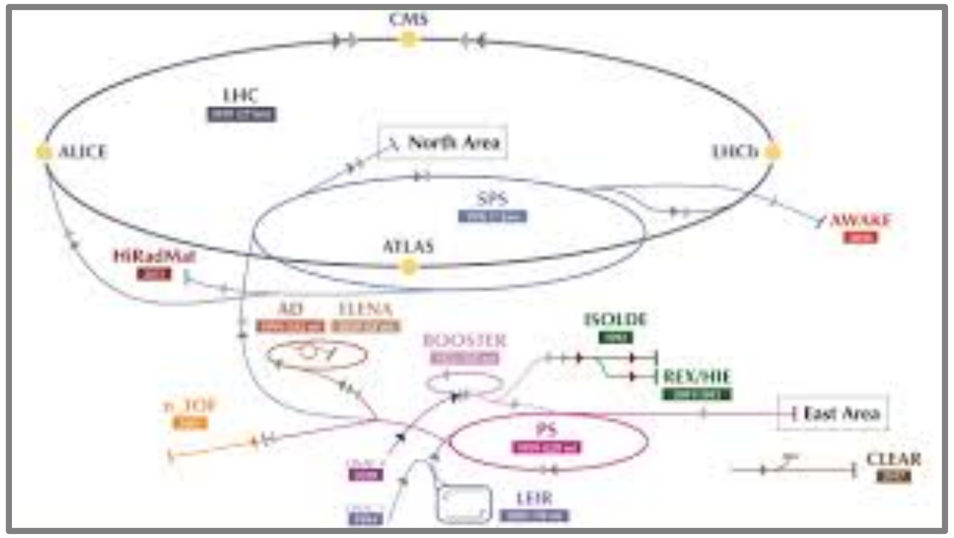




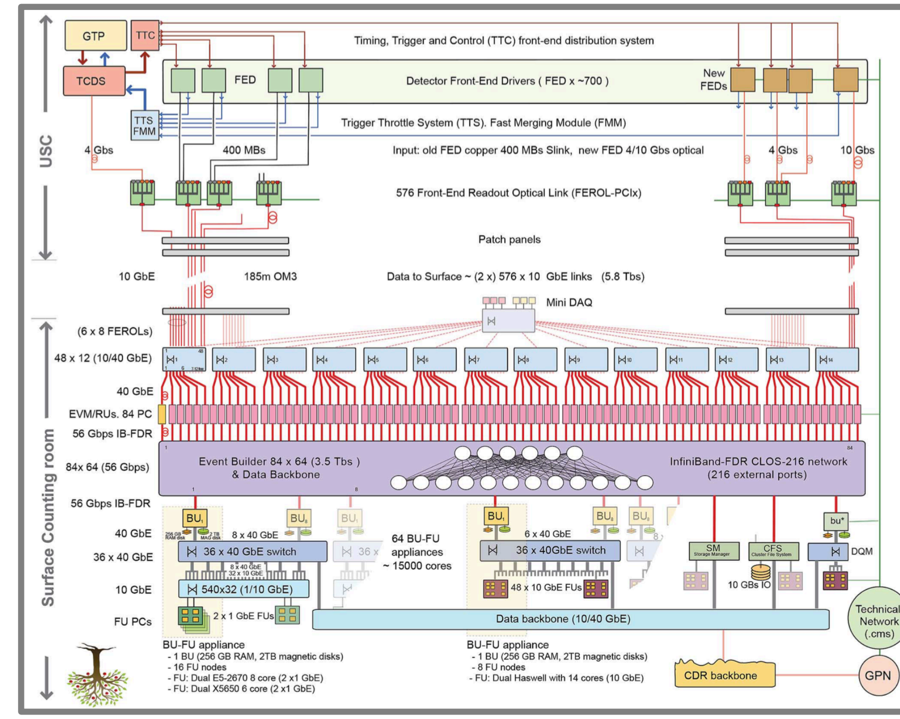
Search in the space of functions using decision ordering.
Simplified to the energy flow polynomial subspace.
Extract set of EFP that matches DNN performance.

Taking Control

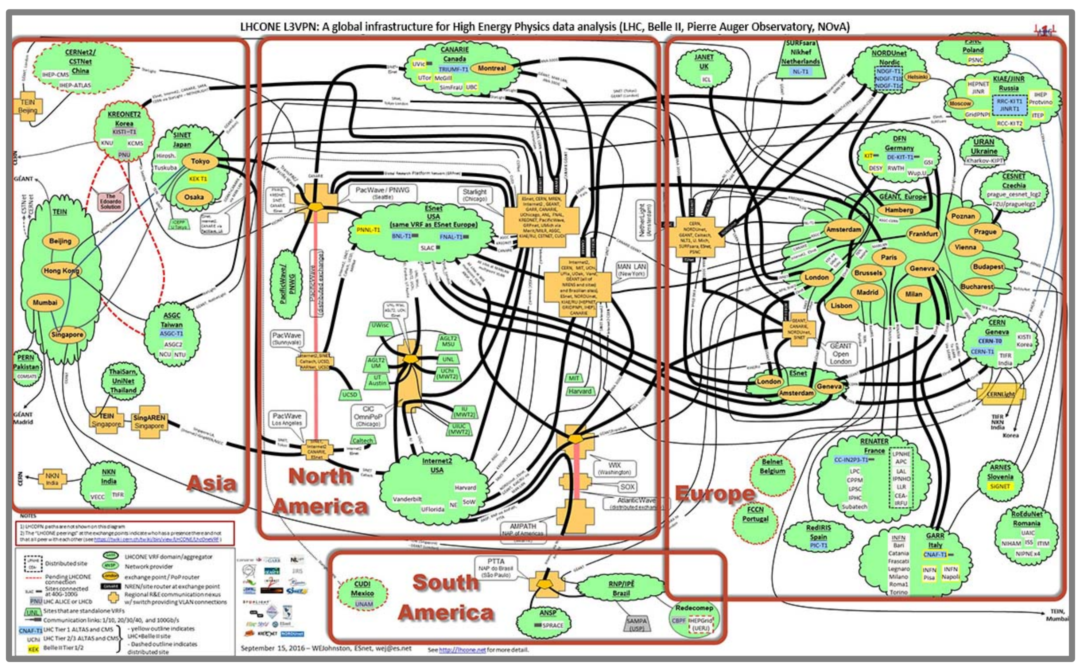
HEP Instruments



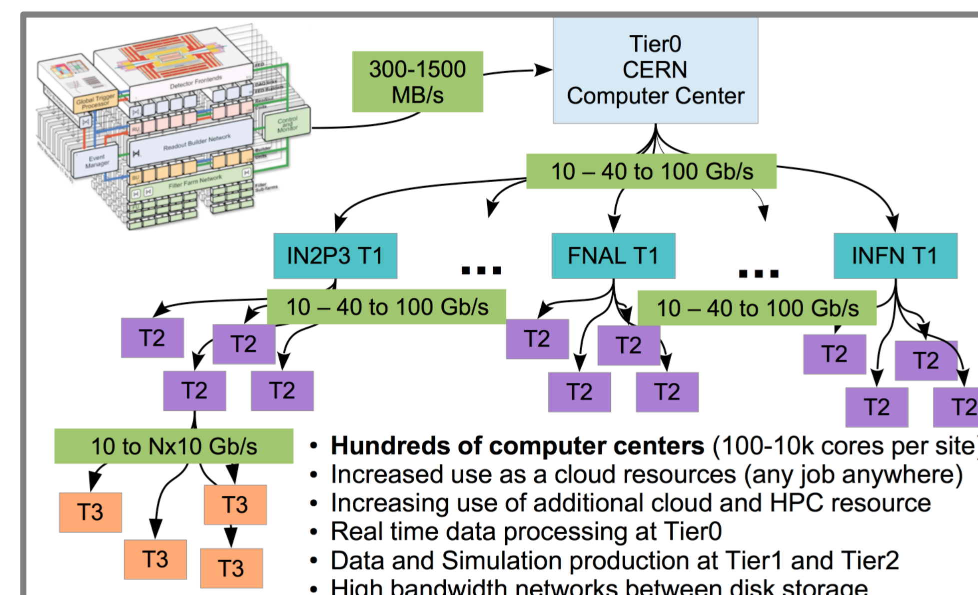
https://home.cern/science/accelerators/



DAQ [IEEE:7111380]



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



Unique set of complex apparatus for doing Science.

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. A lot done, a long way to go.

➡Doing AI at colliders requires to keep an eye on particular topics. Also relevant to other fields of Science.