

# Artificial Intelligence in the Heavens and on Earth

20<sup>th</sup> rencontres du Vietnam  
The Axion Quest  
August 8, 2024



Jean-Roch Vlimant (California Institute of Technology)

[jvlimant@caltech.edu](mailto:jvlimant@caltech.edu)



# Disclaimer

*This talk is very much collider physics oriented though most principles and arguments do transpose to other fields and axion quest.*

*When things are simple, there's always a way to make it more complicated ... so take what you need from this talk and leave the rest.*

*This is not a talk on ChatGPT, and I have not used it to write this talk.*



# Outline

- I. Machine Learning : an introduction
- II. Deep Learning in Particle Physics
- III. Prospects of Artificial Intelligence





# 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.



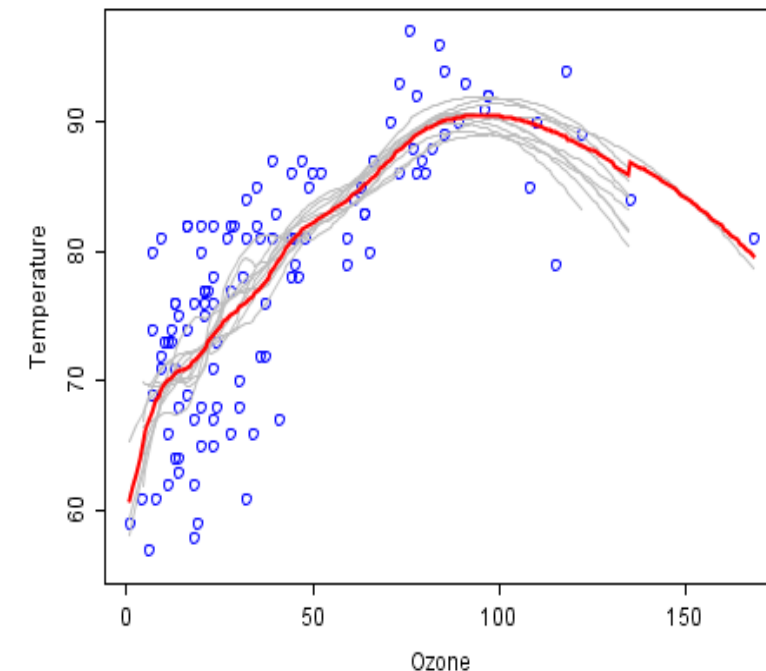
# Supervised Learning

- 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**



N.B. Weakly supervised, Classification without labels : [CWOLA](#)

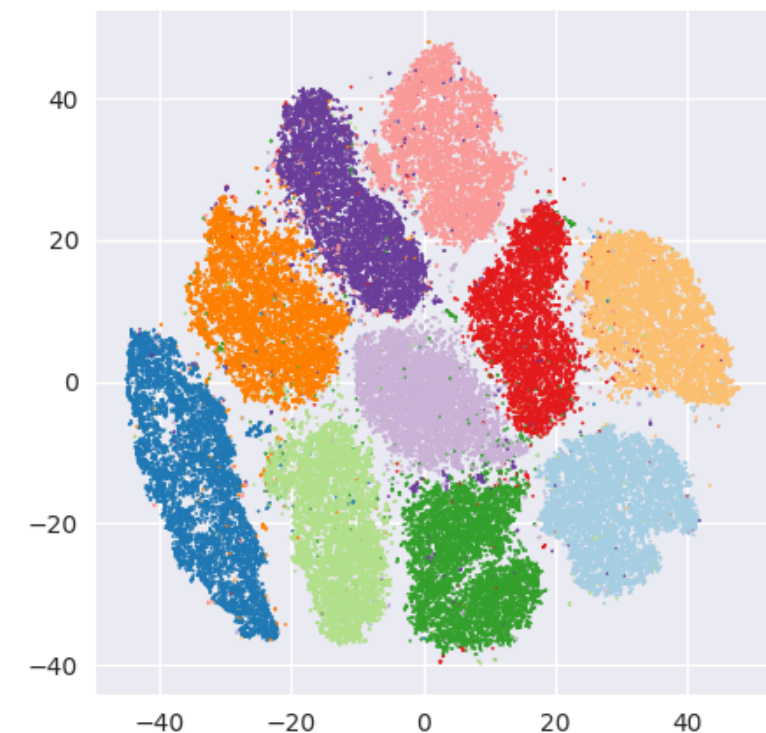


# Unsupervised Learning

- 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

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

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



# Reinforcement Learning

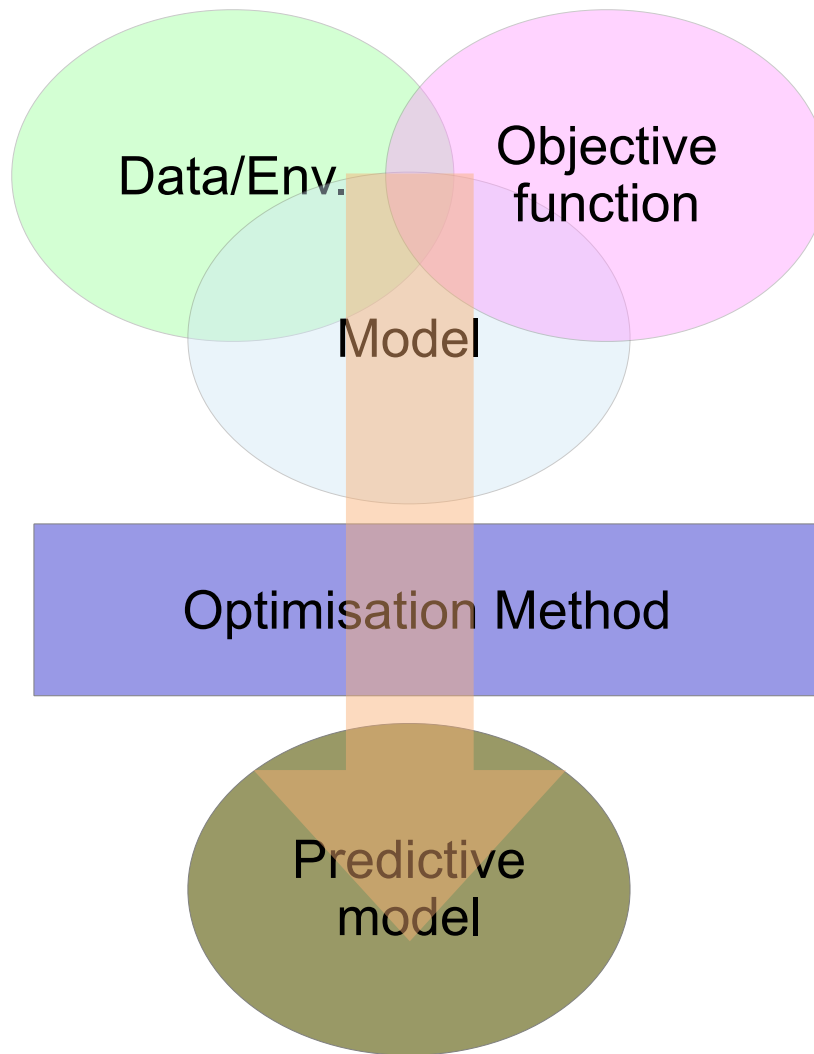
- Given an **environment** with multiple states, given a reward upon action being taken over a state
- Find an **action policy to drive** the environment toward maximum cumulative reward

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





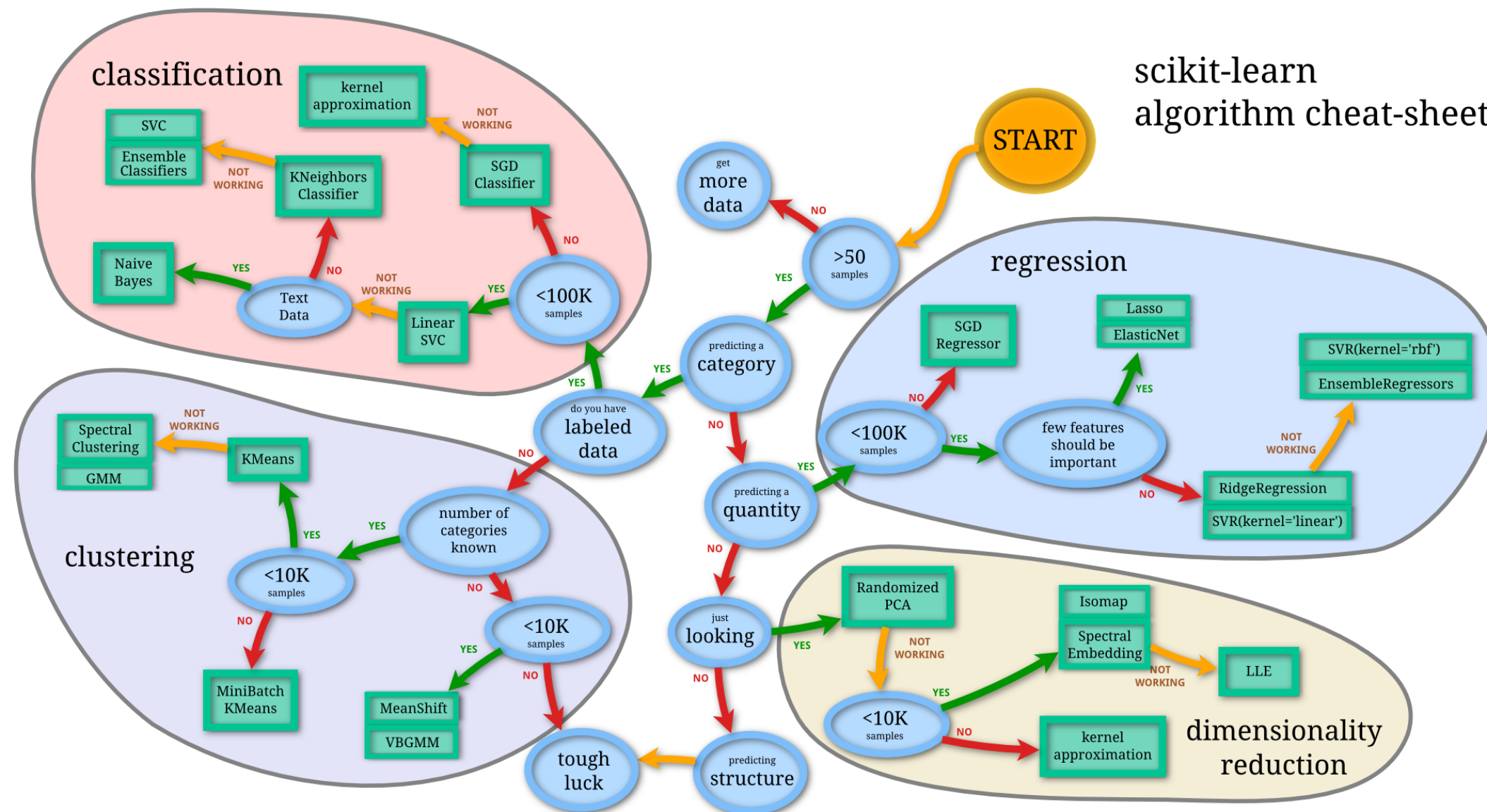
# Overview



Many optimisation methods adapted to the various type of the dataset, model, objective.  
Gradient descent, evolutionary algorithms, ...



# (Some) Machine Learning Methods



<http://scikit-learn.org/stable/tutorial/index.html>

Lot of ML algorithms out there.

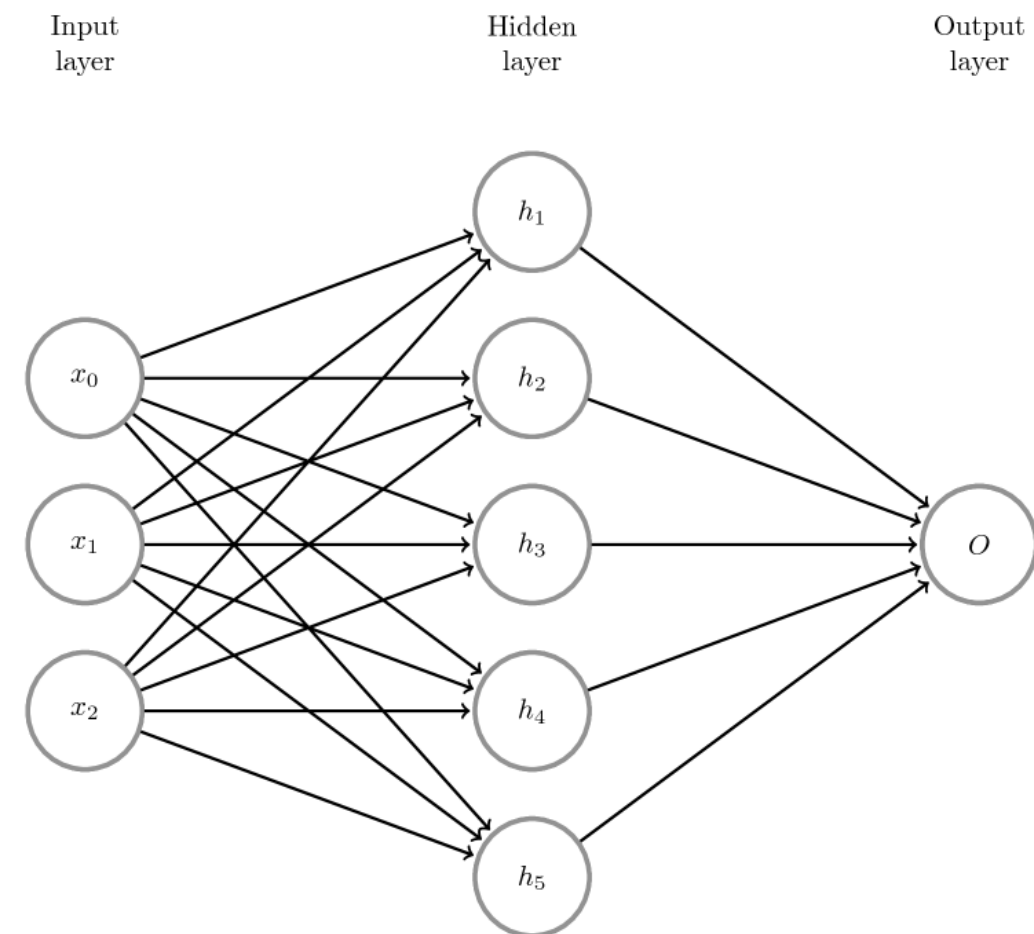


# Artificial Neural Network

$$h = \phi(Ux + v)$$
$$o(x) = \omega^T h + b$$

$$p_i \equiv p(y = 1 | x) \equiv \sigma(o(x)) = \frac{1}{1 + e^{-o(x)}}$$

$$\text{loss}_{XE} = - \sum_i y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)$$



**Biology inspired** analytical model, but **not bio-mimetic**

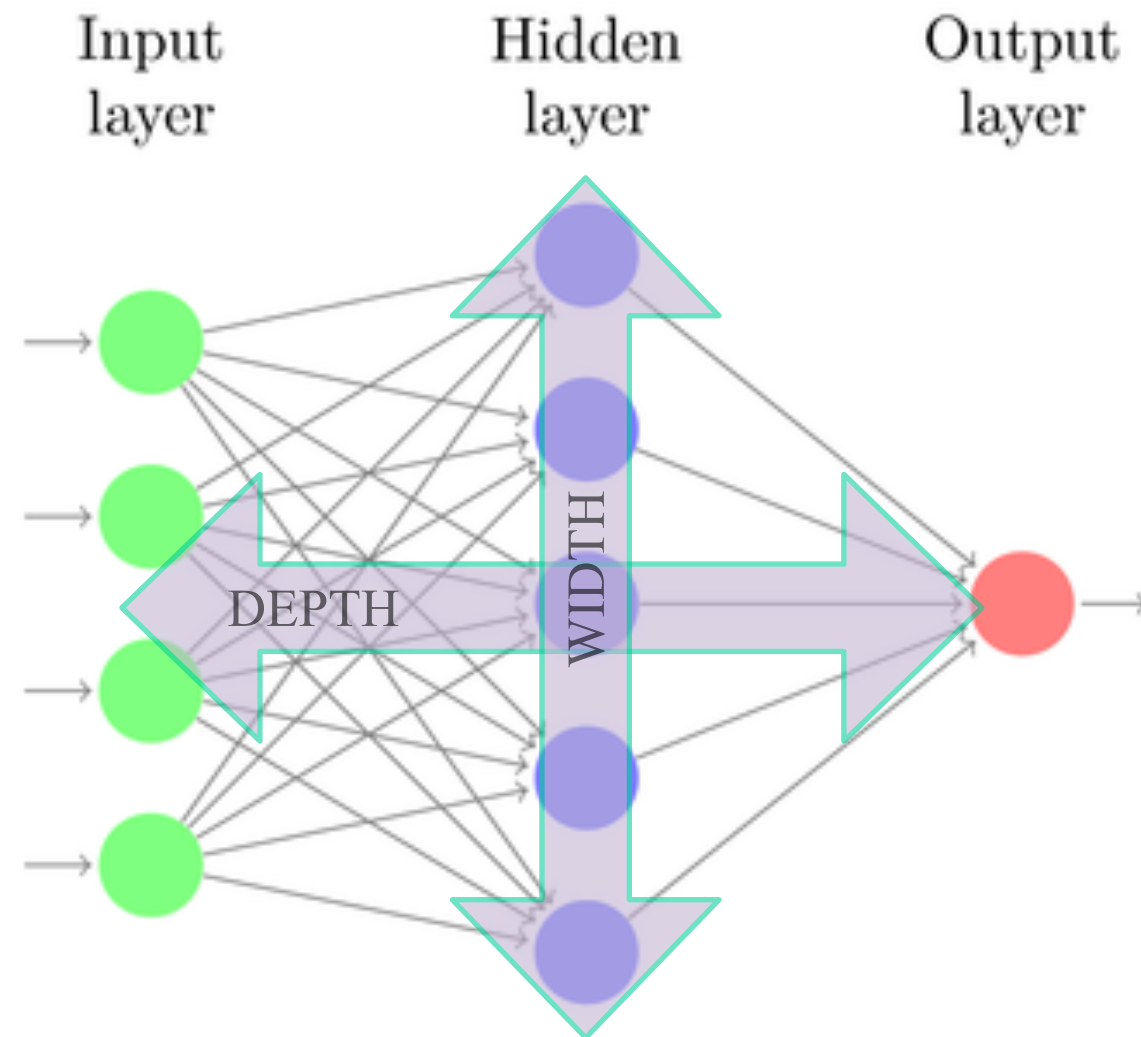
Booming in recent decade thanks to large dataset, increased computational power and theoretical novelties

Origin tied to logistic regression with change of data representation

Parameters usually trained with **stochastic gradient descent**



# Going Deep Learning



Kolmogorov's, and universal approximate theorems push towards **wide** and **deep** densely connected networks.

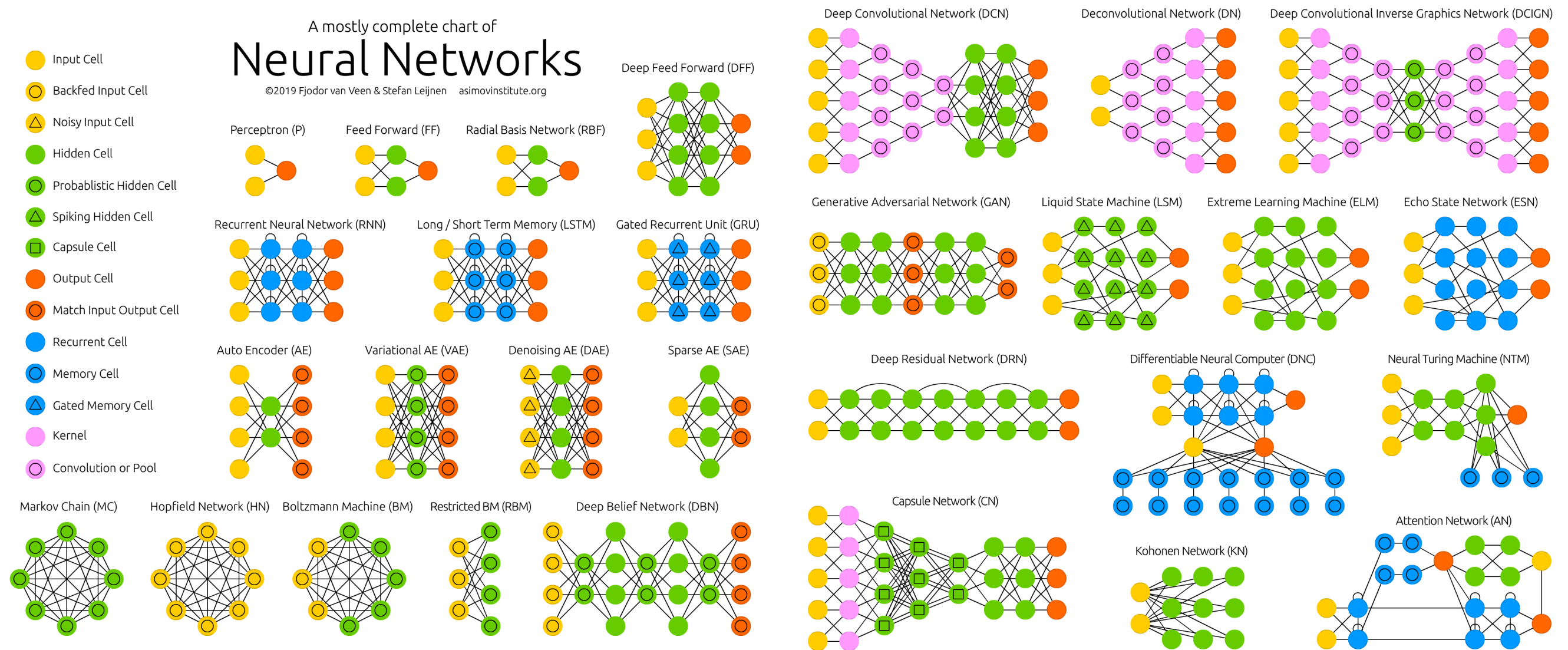
**Depth** helps with **decomposition**. **Width** helps with **approximation**.

**Still quite empirical field**, although theoretical progress are made



# Neural Net Architectures

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



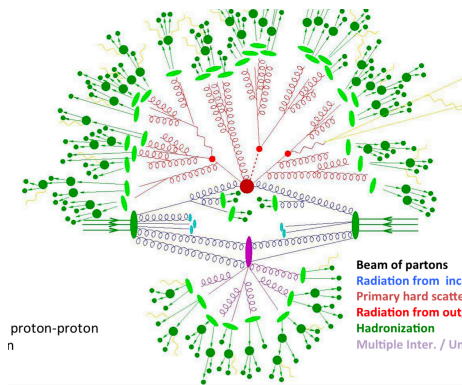
Many architectures and many more : graph network, transformer, ...  
**Plasticity and induction bias** with new generation of architecture.



# Generative Models

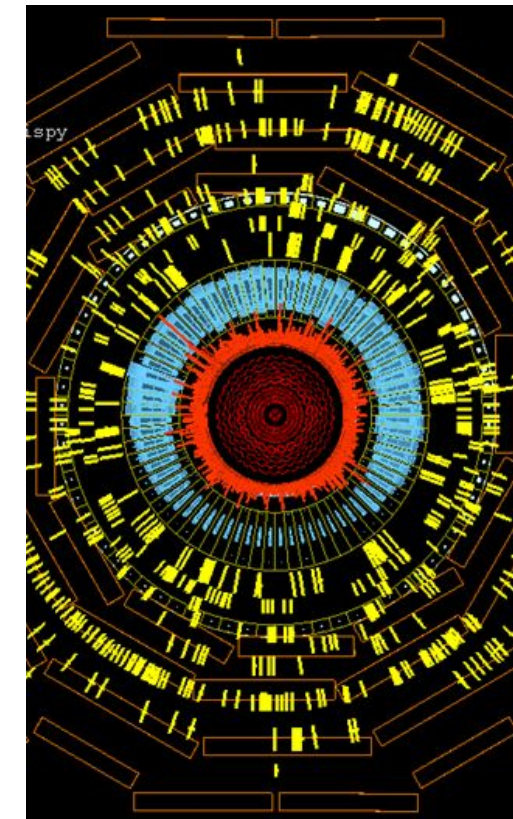
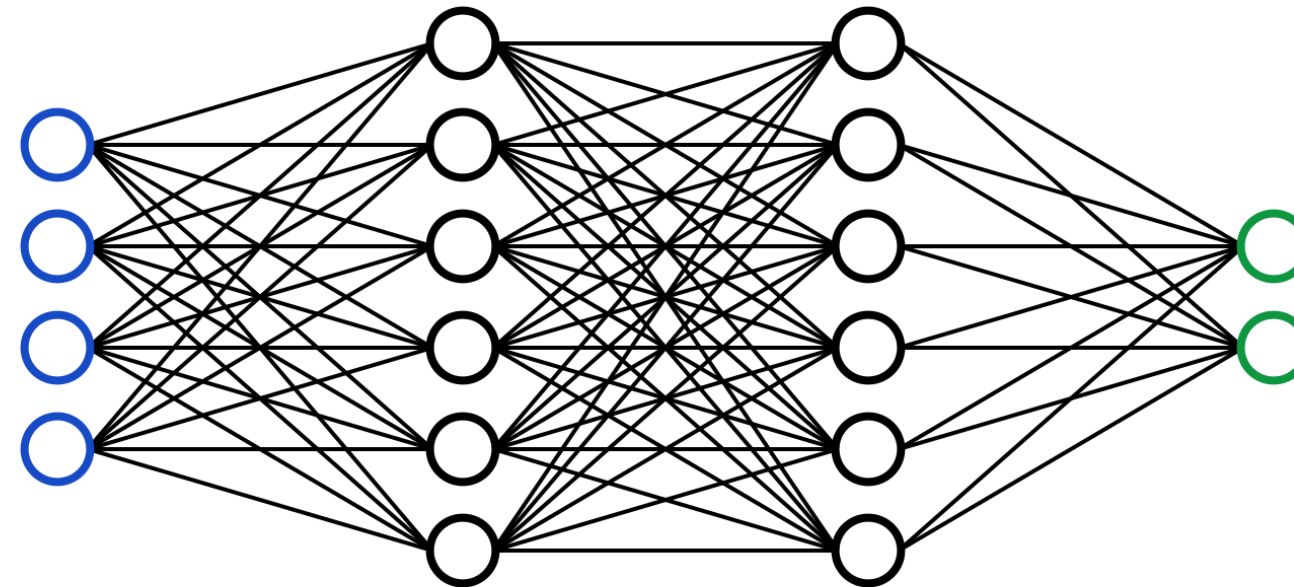


Random Numbers



proton-proton  
n

Conditioning



Machine Learning model able to learn the underlying data probability distribution function from the training data.

And produce stochastic output, as if they were new samples from the original dataset.

Extendable to data completion. Tune-able stochasticity.



# NLP — LLM

Natural Language Processing (NLP) has been one of the main driver of deep learning since the beginning.

Several leaps were made in the field over the last decades (RNN, LSTM, Attention, ...) and more recently thanks to Transformer-based models and generative pre-trained transformers (GPT).

Entering the era of large language Models LLM that have super-human performance.

e.g. ChatGPT : a model is able to converse in many languages, including **software language i.e writing code.**

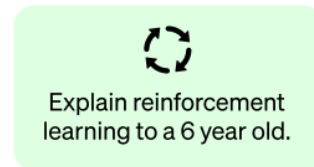


# Chat-GPT

Step 1

Collect demonstration data and train a supervised policy.

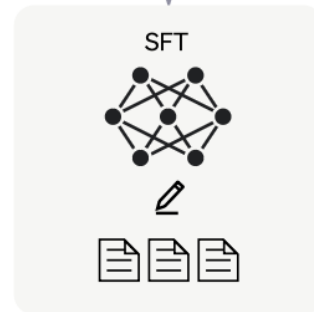
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3.5 with supervised learning.

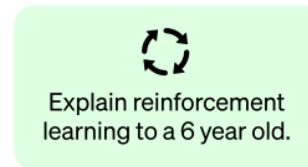


<https://openai.com/index/chatgpt/>

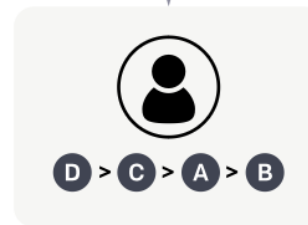
Step 2

Collect comparison data and train a reward model.

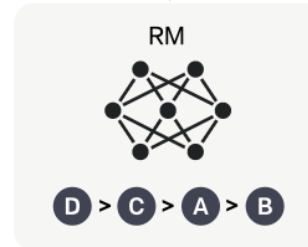
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

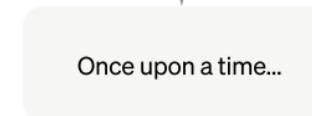
A new prompt is sampled from the dataset.



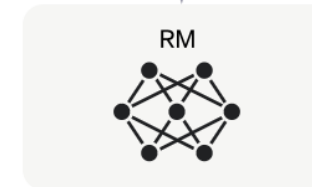
The PPO model is initialized from the supervised policy.



The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



Pre-trained to complete sentences over the IOT.

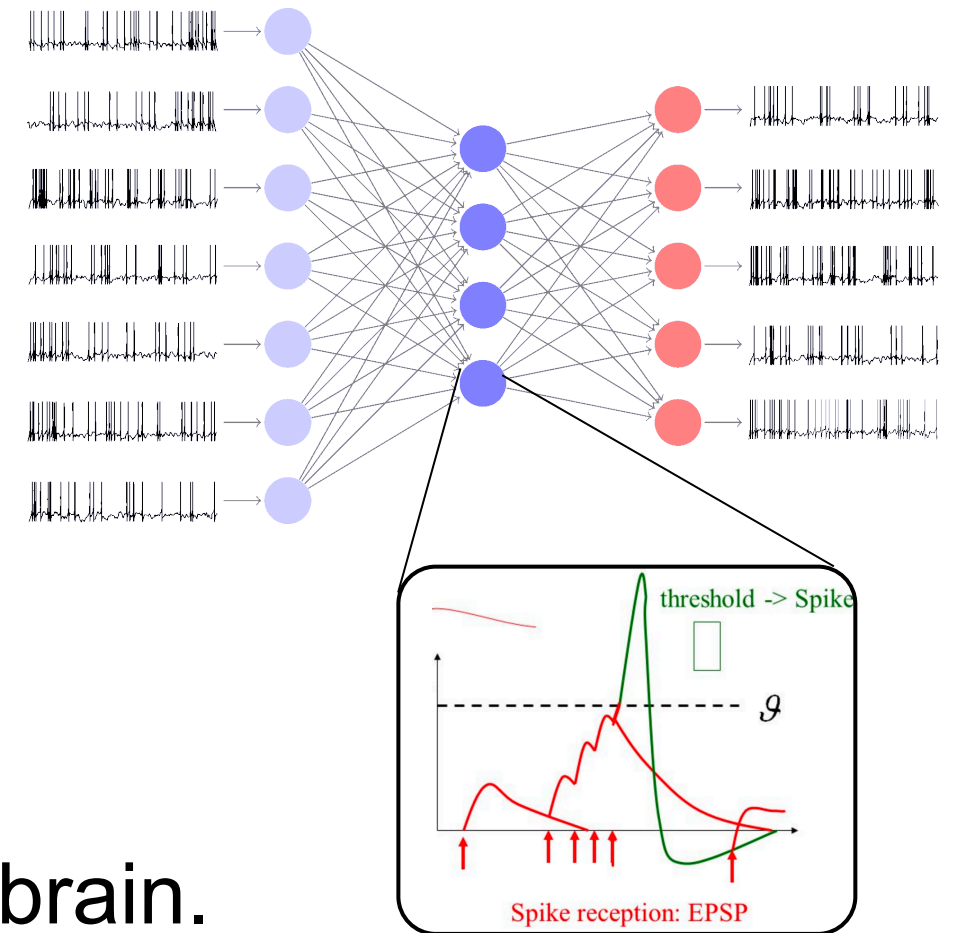
N.B. the need for human AI trainers





# Spiking Neural Network

|                      | Deep Learning  | Spiking   |
|----------------------|--|---|
| Training Method      | Back-propagation   | Not well established (here, genetic algorithms) <b>Partially solved</b> |
| 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   |



Closer to the actual biological brain.

Good for **temporal data**.

Hardware implementation with **low power consumption**.

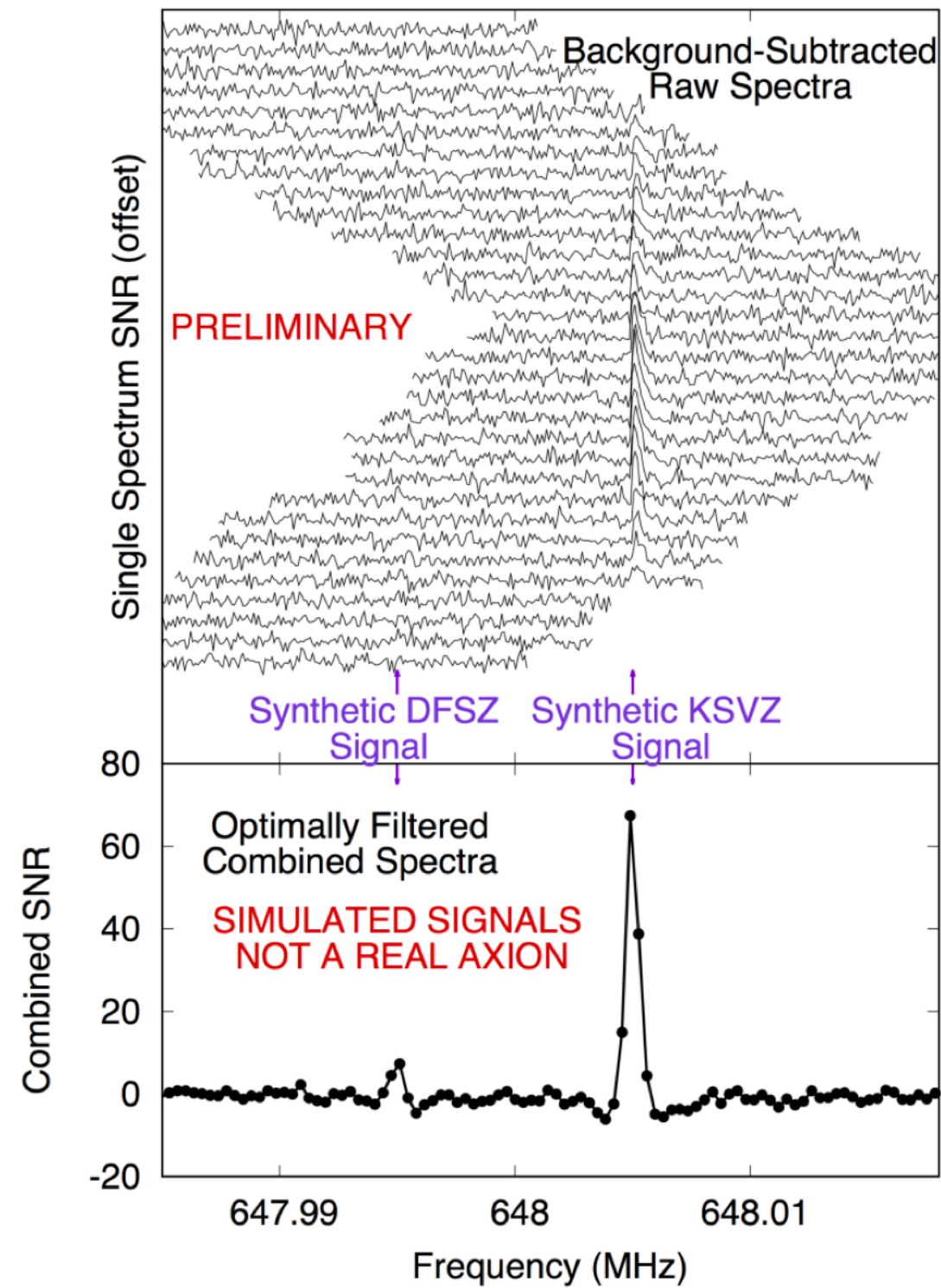
Trained using evolutionary algorithms, recent work on gradient-based.

Python libraries available : [slayer](#), [snntorch](#), [spikingjelly](#), [norse](#), ...

For ref [International Conference on Neuromorphic Systems](#).



# For example



Search for the QCD Axion with ADMX [[Nick Du](#)]

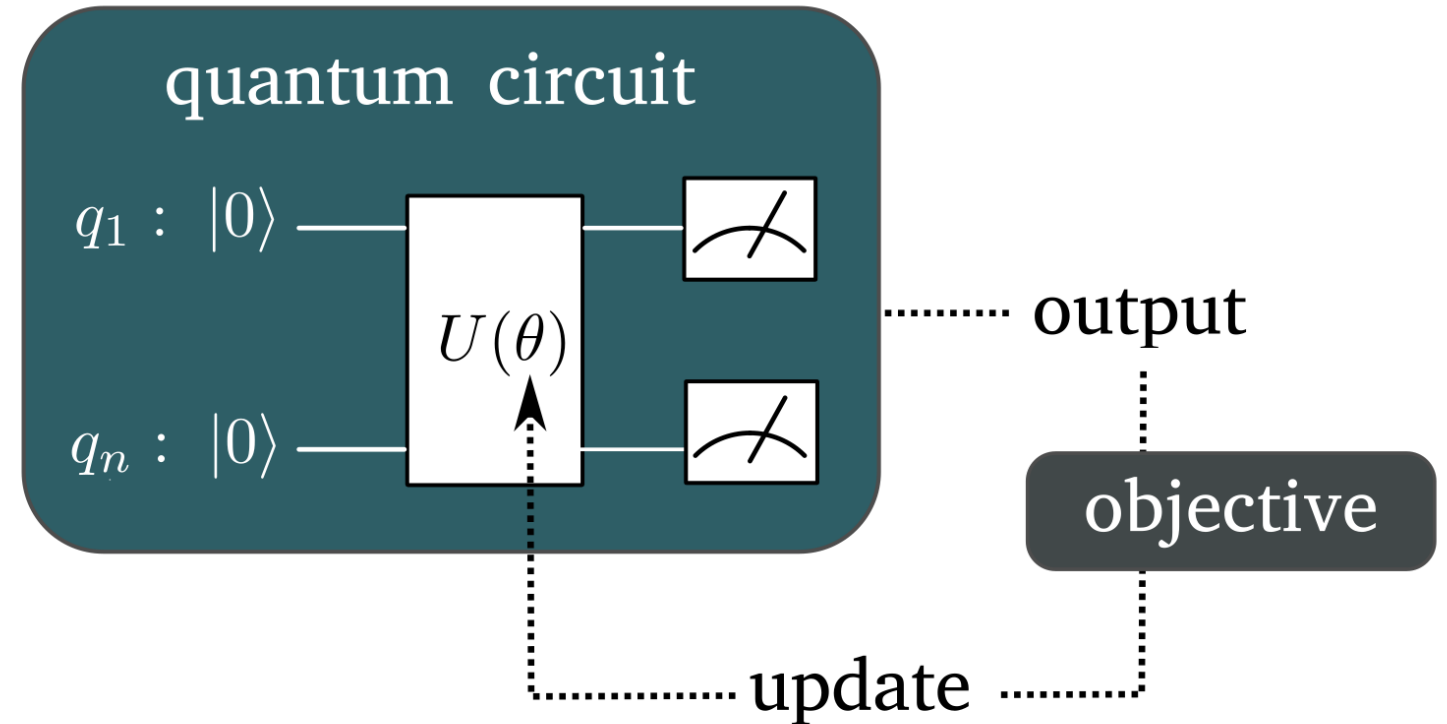
AI for the Axion Quest, J-R Vlimant



# Quantum Machine Learning



<https://www.ibm.com/quantum-computing/>



Objective based on quantum measurement.  
Parameters of a quantum circuits as weights.  
Trainable circuits for quantum machine learning.

Quantum Machine Learning [\[1611.09347\]](#)

Quantum Machine Learning in High Energy Physics [\[2005.08582\]](#)

Quantum Machine Learning Models are Kernel Methods [\[2101.11020\]](#)

Good for **generative model** or **learning from quantum data**.



# Artificial Intelligence

Science that aims at exhibiting “intelligence”, i.e : **capacity to perceive information, transform in knowledge, apply within an environment.**

*Narrow AI* : AI that matches human capability for a specific task.

*Artificial General Intelligence (AGI)* : AI that matches human capabilities across a wide range of tasks.

*Artificial Super Intelligence (ASI)* : AI with an intellectual scope beyond human intelligence.

Involves creativity, making analogies, extrapolating off domain, ...

An introspective study on how “human intelligence” is acquired.

Still very much used as hype ...



*Take home message :*

*Deep learning models have lots of capacity  
and demonstrate great performance.*

*Deep learning is still very much empirical.*

*AGI has still to be achieved.*



# Motivations for Using Machine Learning in Particle/Astro Physics

*and elsewhere ...*



# Motivations in a Nutshell

[\*More details in backups\*](#)

Experience with machine learning can significantly **open career horizons**.

Machine learning can provide ways to learn to **control environments**.

Physics **knowledge can be extracted** from data.

**Physics principles can be used** to boost model performances.

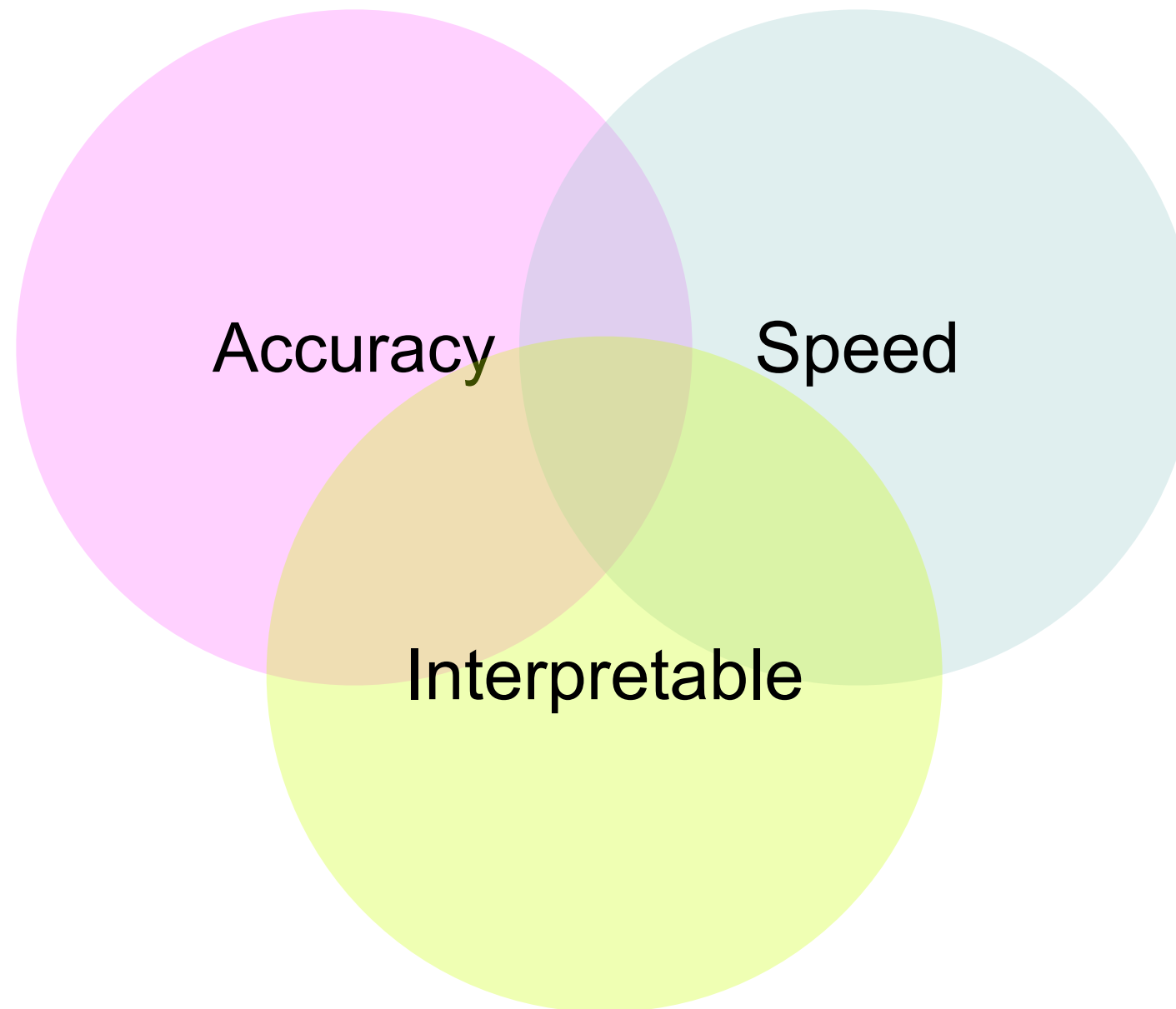
Deep learning can **learn from very complex data**.

Model can be **evaluated quite efficiently**.

Spiking models **use low power**.



# Possible Utilisation



- **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 cumbersome tasks** (operations, etc)





*Take home message :*

*Machine Learning is a widely recognised and used technology in industry*

*Deep Learning has the potential of helping Science to make progress*

*Neural Networks could help with the computing efficiency of Science*



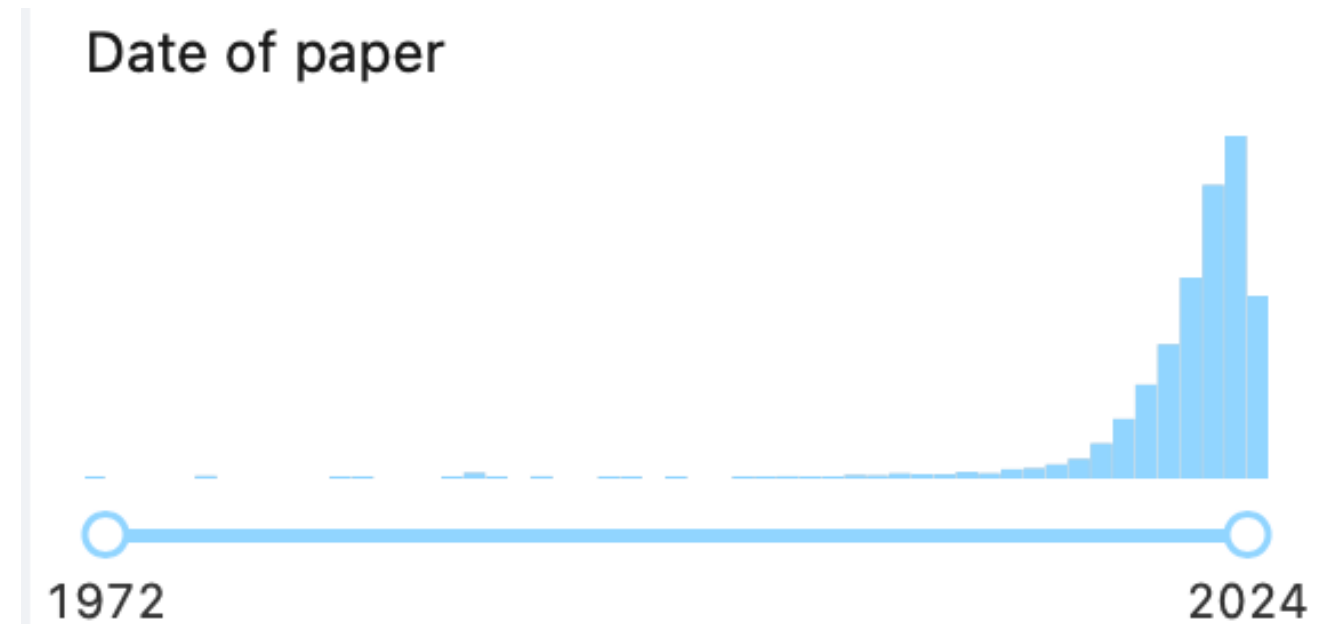
# Deep Learning in Particle Physics

*how far have we gone now ...*

Gerd



# Growing Literature



[https://inspirehep.net/literature?q=machine learning or deep learning](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/>



# Historical View

The AIHENP/ACAT series pioneered in showcasing AI in High Energy and Nuclear physics.

[DS@LHC15](#) and [DS@HEP](#) series lead us out of an AI winter.

Many workshop and conference series now exist for AI/Deep Learning in science and HEP : [ACAT](#), [Machine Learning and the Physical Sciences](#), [Fast Machine Learning for Science](#), [Machine Learning for Jets](#), [Hammers and Nails](#), ...

Omnipresent in main conferences and workshops.

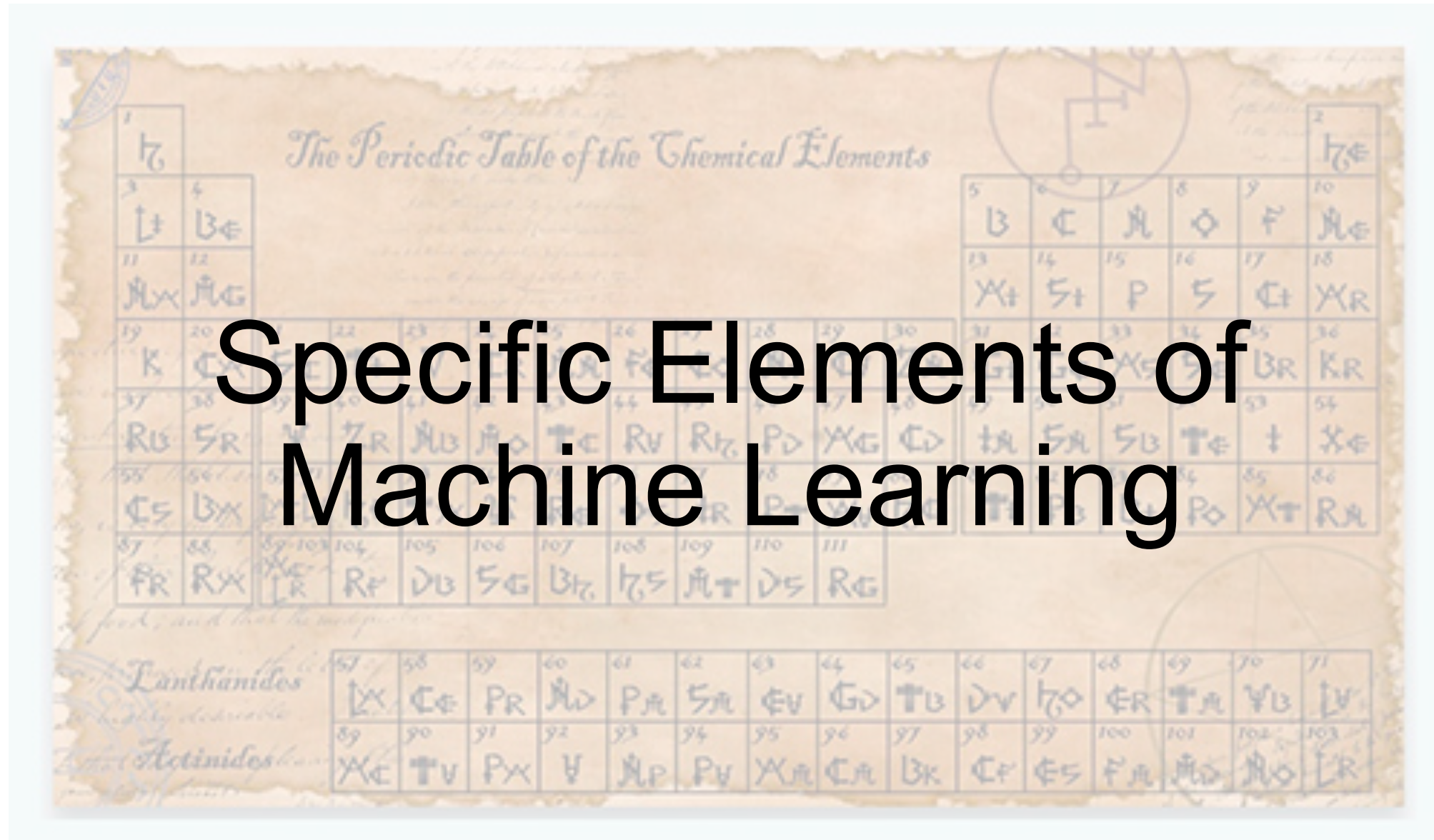
Many AI/DL topical workshops within the community.

## Past Workshops

|    | Workshop    | place                                    | Date                  |
|----|-------------|--|-----------------------|
| 1  | AIHENP 1990 | Lyon (France)                            | March 19-24 1990      |
| 2  | AIHENP 1992 | La Londe Les Maures (France)             | Jan. 13-18 1992       |
| 3  | AIHENP 1993 | Oberammergau (Germany)                   | Oct. 4-8 1993         |
| 4  | AIHENP 1995 | Pisa (Italy)                             | April 3-8 1995        |
| 5  | AIHENP 1996 | Lausanne (EPFL-UNIL) (Switzerland)       | Sept. 2-6 1996        |
| 6  | AIHENP 1999 | Heraklion (Crete, Greece)                | April 12-16, 1999     |
| 7  | ACAT 2000   | <a href="#">Chicago (FERMILAB) (USA)</a> | Oct. 16-20, 2000      |
| 8  | ACAT 2002   | <a href="#">Moscow (MSU)(Russia)</a>     | June 24-28, 2002      |
| 9  | ACAT 2003   | <a href="#">Tsukuba (KEK) (Japan)</a>    | Dec. 1-5, 2003        |
| 10 | ACAT 2005   | <a href="#">Zeuthen (DESY) (Germany)</a> | May 22-27, 2005       |
| 11 | ACAT 2007   | Amsterdam (NIKHEF) (The Netherlands)     | April 23-27, 2007     |
| 12 | ACAT 2008   | <a href="#">Erice (Italy)</a>            | Nov. 3-7, 2008        |
| 13 | ACAT 2010   | <a href="#">Jaipur (India)</a>           | Feb. 22-27, 2010      |
| 14 | ACAT 2011   | <a href="#">Uxbridge (UK)</a>            | Sept. 5-9, 2011       |
| 15 | ACAT 2013   | <a href="#">Beijing (China)</a>          | May 16-21 2013        |
| 16 | ACAT 2014   | <a href="#">Prague (Czech Republic)</a>  | Sept. 1-5, 2014       |
| 17 | ACAT 2016   | <a href="#">Valparaiso (Chile)</a>       | Jan. 18-22, 2016      |
| 18 | ACAT 2017   | <a href="#">Seattle (USA)</a>            | Aug. 21-25, 2017      |
| 19 | ACAT 2019   | <a href="#">Saas Fee (Switzerland)</a>   | Mar. 11-15, 2019      |
| 20 | ACAT 2021   | <a href="#">Daejeon (South Korea)</a>    | Nov. 29 - Dec 3, 2021 |
| 21 | ACAT 2022   | <a href="#">Bari (Italy)</a>             | Oct. 24-28, 2022      |

<https://twiki.cern.ch/twiki/bin/view/ACAT/PastWorkshop>





# Specific Elements of Machine Learning



# Specific Elements of ML in a Nutshell

[More details in backups](#)

Lots of labelled simulated data can be obtained from simulator.

Various level of data representation  $\Rightarrow$  image, sequence, graphs.

Various symmetries at play in the data  $\Rightarrow$  equivariant models.

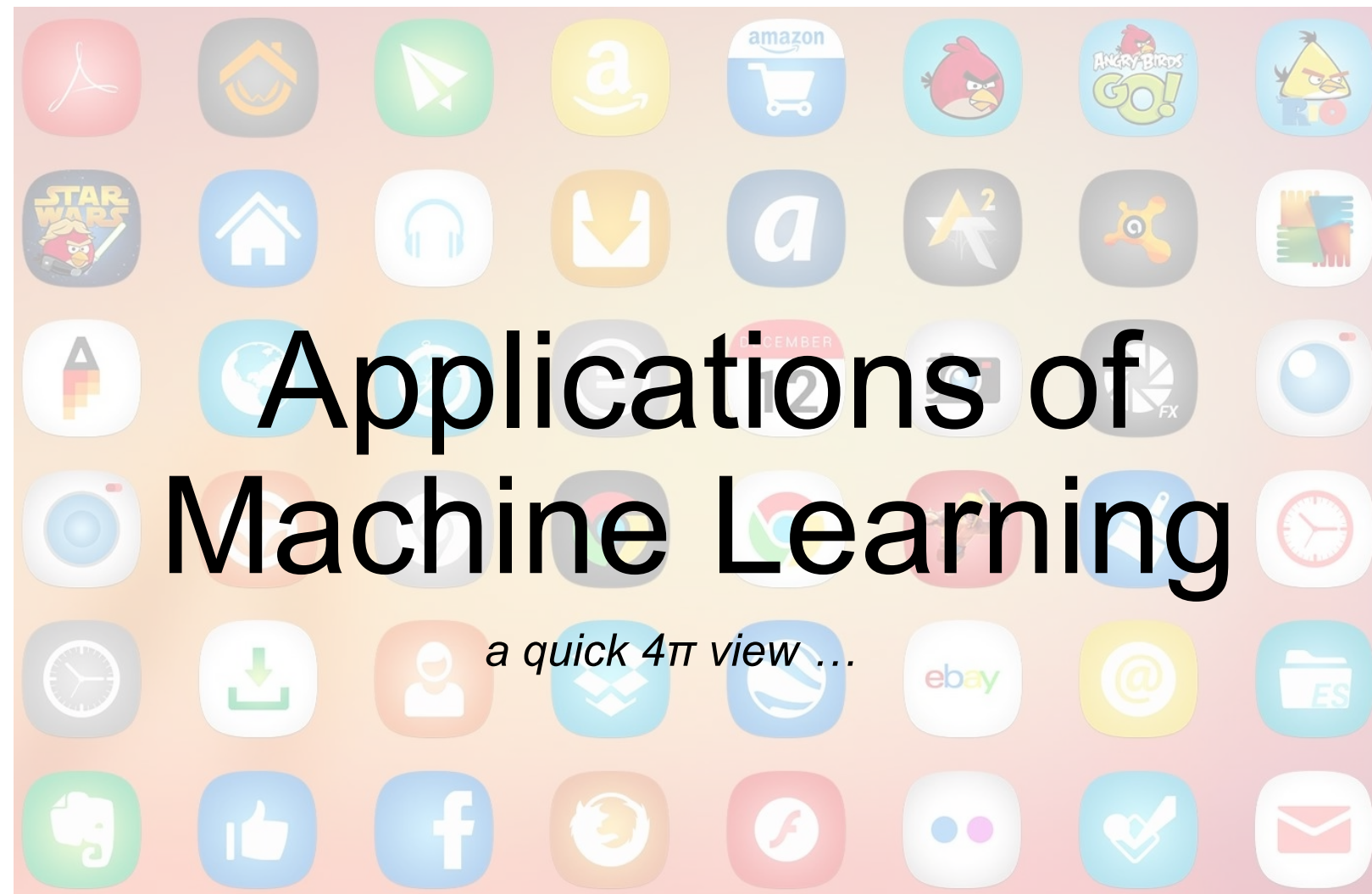
Need as least as possible biased models  $\Rightarrow$  decorrelation methods.

Uncertainties need to be taken into account  $\Rightarrow$  uncertainty aware models.

Data and simulation do not necessarily agree  $\Rightarrow$  domain adaptation.

Physicist need to understand the models  $\Rightarrow$  model interpretation.





*a quick 4π view ...*



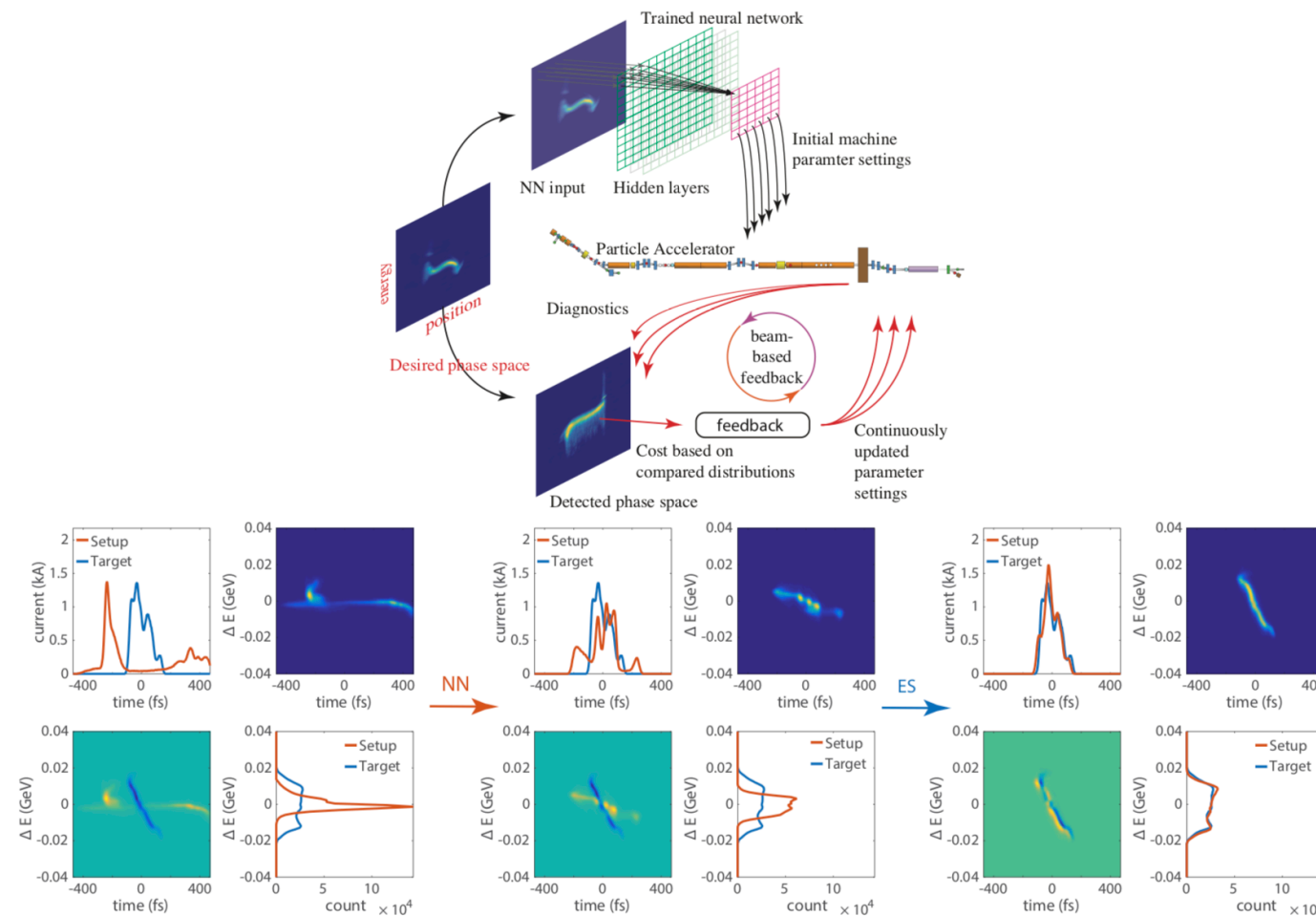


# Data Management





# Producing the Data



A. Scheinker, C. Emma, A.L. Edelen, S. Gessner  
[\[2001.05461\]](#)

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 [\[2009.08109\]](#)

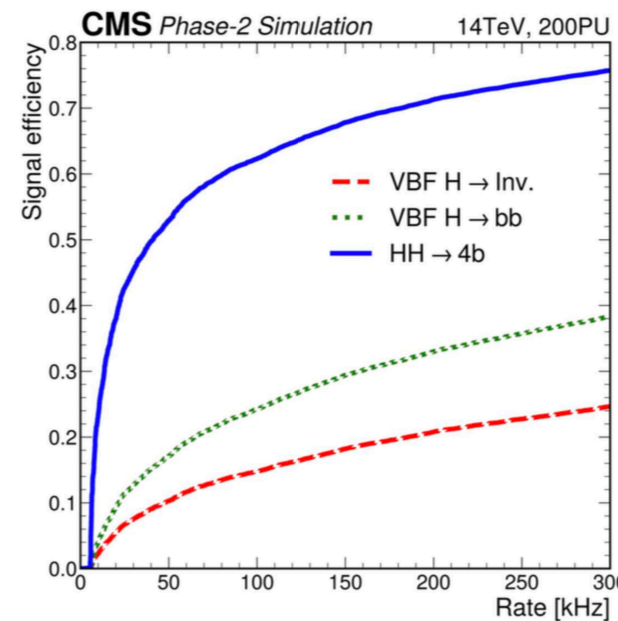
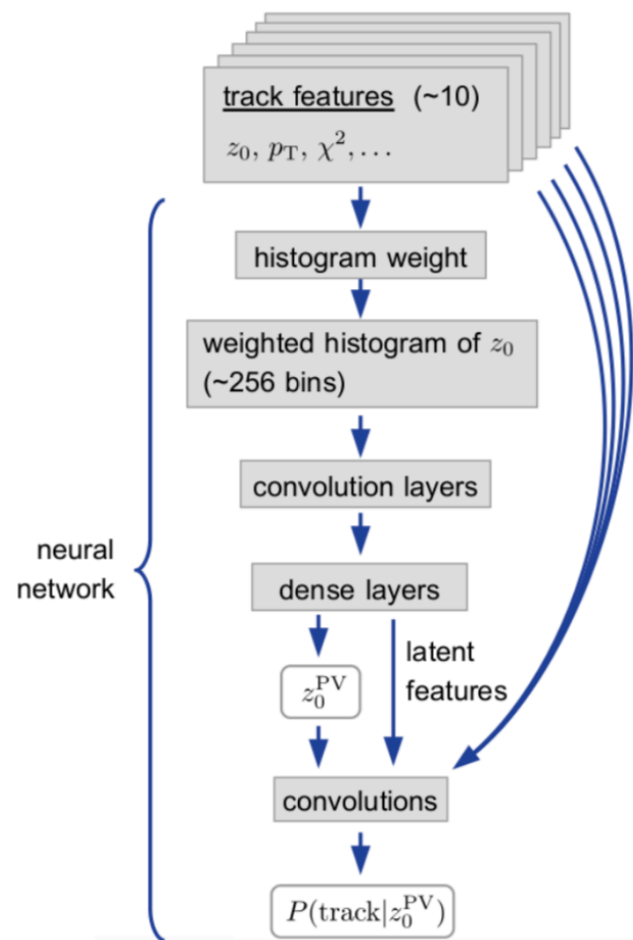
...

- 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.



# Triggering and Scouting

Vertex reconstruction at L1



Anomaly detection at L1

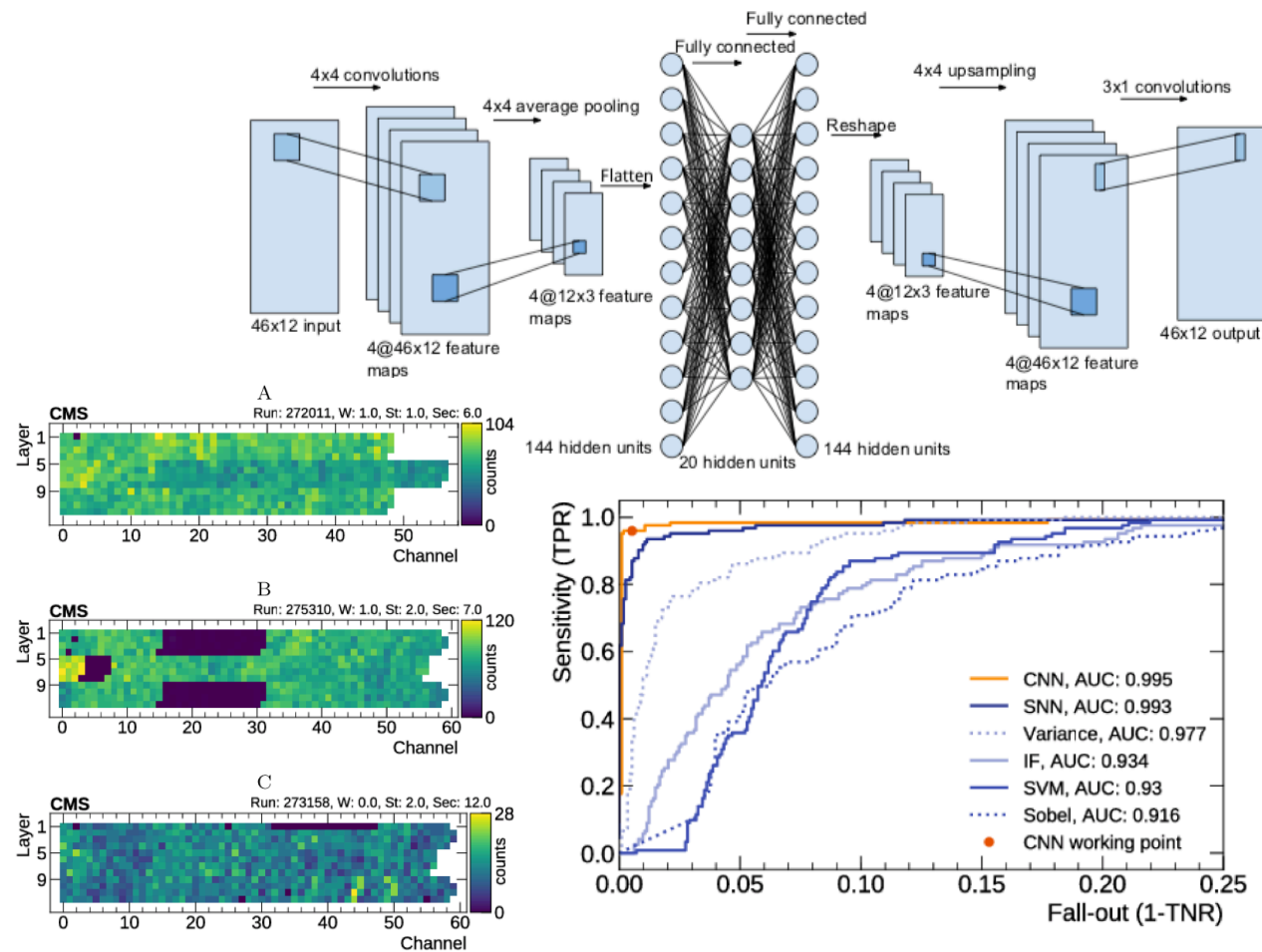
Phase-2 upgrade of the CMS L1-Trigger

[\[cds:2714892\]](https://cds.cern.ch/record/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



# 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

[\[doi:10.1007/s41781-018-0020-1\]](https://doi.org/10.1007/s41781-018-0020-1)

Towards automation of data quality system for CERN CMS experiment [\[doi:10.1088/1742-6596/898/9/092041\]](https://doi.org/10.1088/1742-6596/898/9/092041)

LHCb data quality monitoring [\[doi:10.1088/1742-6596/898/9/092027\]](https://doi.org/10.1088/1742-6596/898/9/092027)

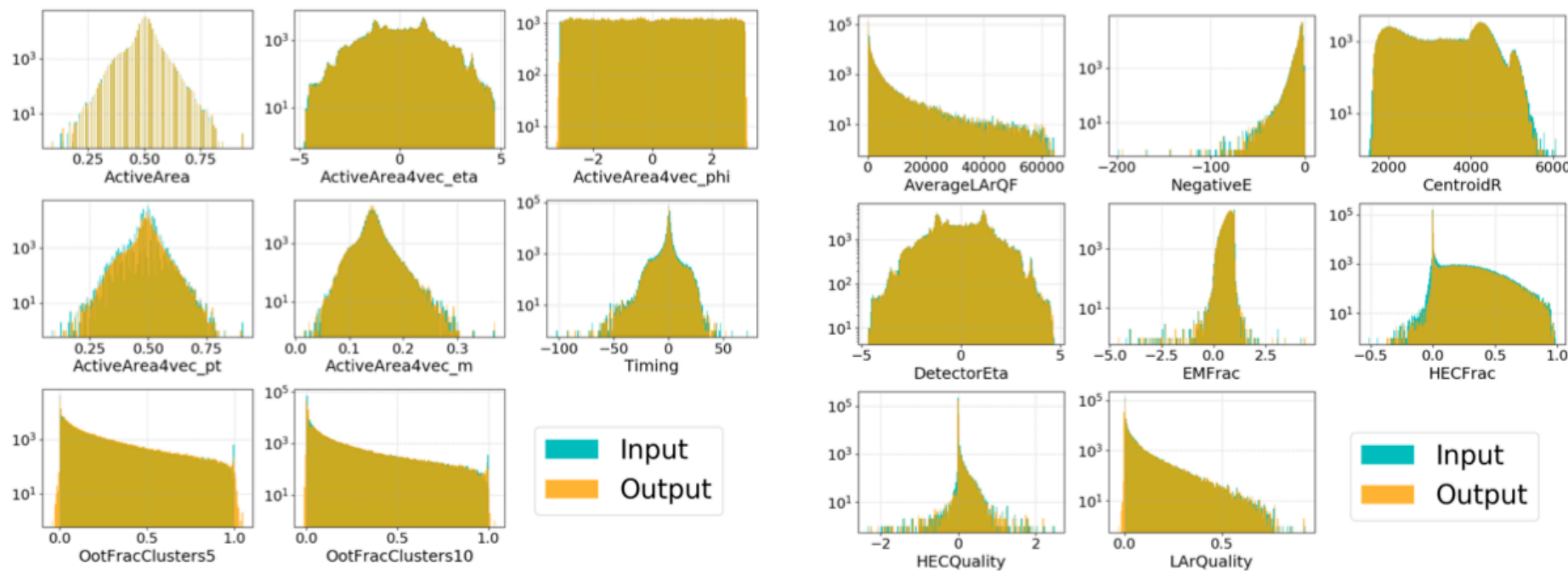
Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider [\[1808.00911\]](https://doi.org/10.1088/1742-6596/898/9/092041)

Anomaly detection using Deep Autoencoders for the assessment of the quality of the data acquired by the CMS experiment [\[doi:10.1051/epjconf/201921406008\]](https://doi.org/10.1051/epjconf/201921406008)

...



# Compressing Data

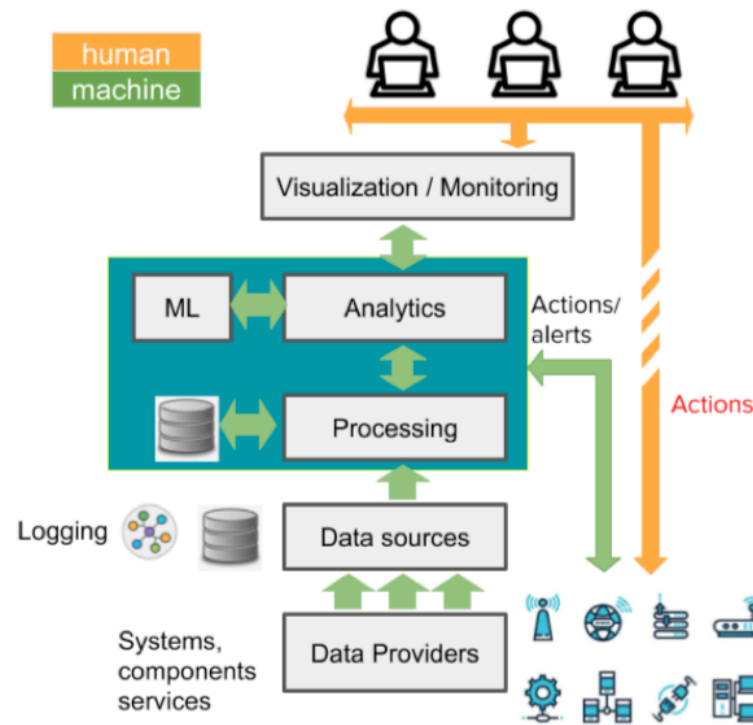


Deep Auto-Encoders for compression in HEP  
<http://lup.lub.lu.se/student-papers/record/9004751>

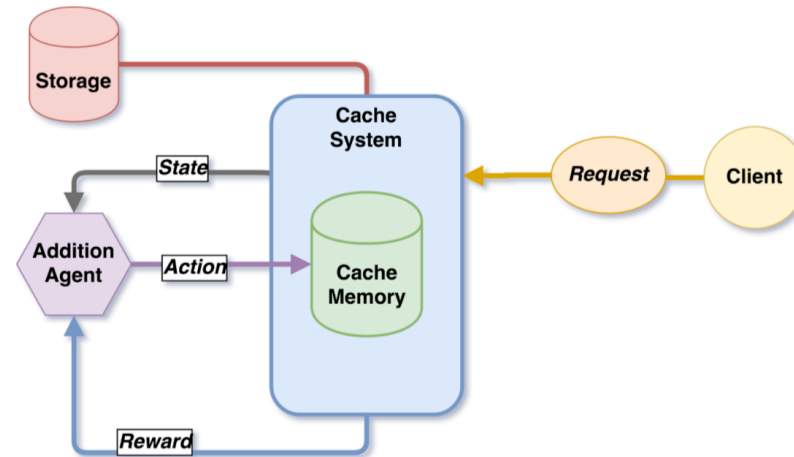
- Rich literature on data compression of image with neural network.
- Make use of abstract semantic space for image compression.
- Image compression can suffer some loss of resolution.
- Saving on disk/tape cost. Potential in scouting strategies.
- R&D needed to reach the necessary level of fidelity.



# Managing Data



Operational Intelligence  
[\[cds:2709338\]](https://cds.cern.ch/record/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](#)

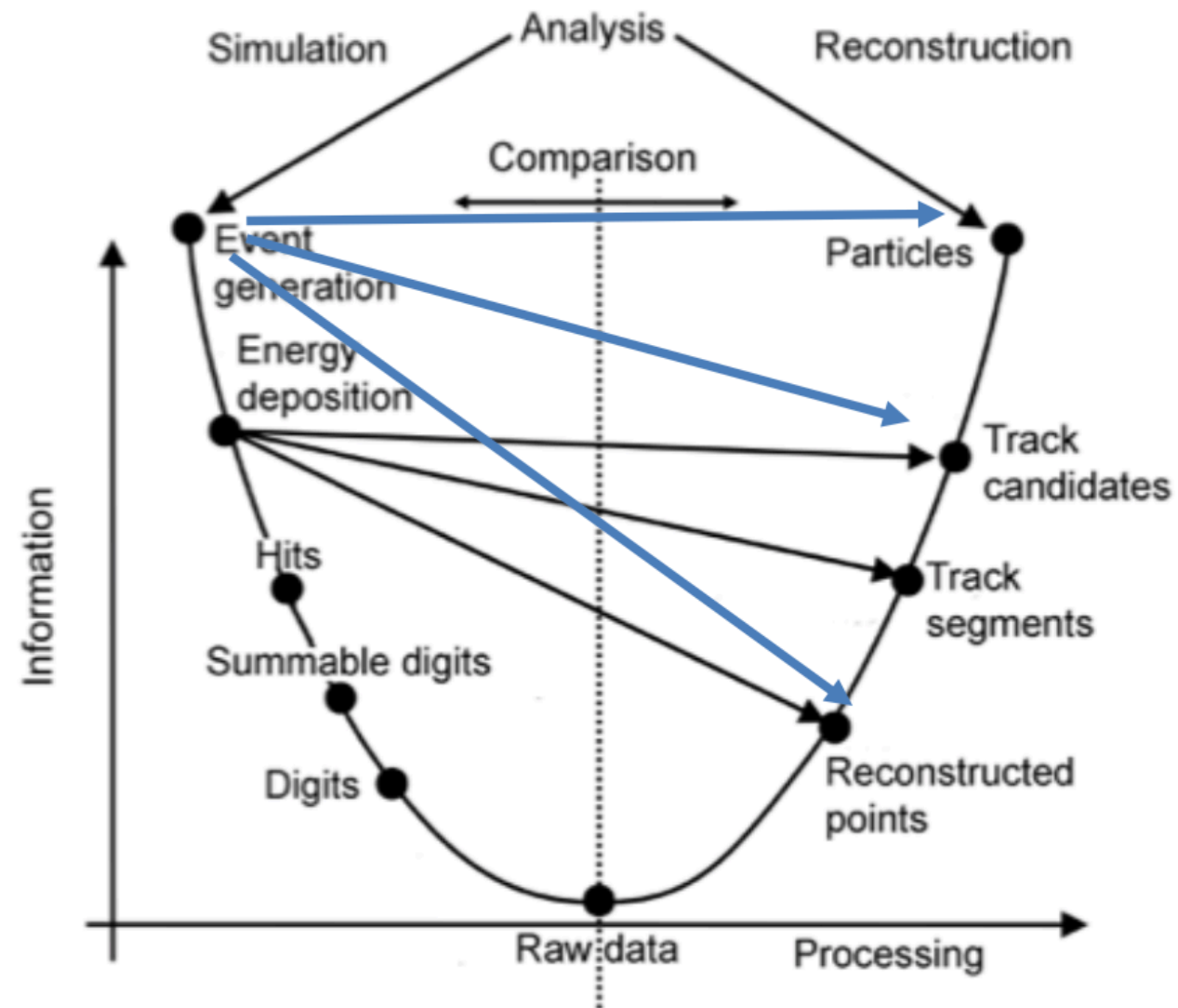




# Simulation/Reconstruction Surrogates



# Reconstruction ◦ Simulation ~ Identity



Simulation aims at predicting the outcome of collisions.

Reconstruction aims at inverting it.

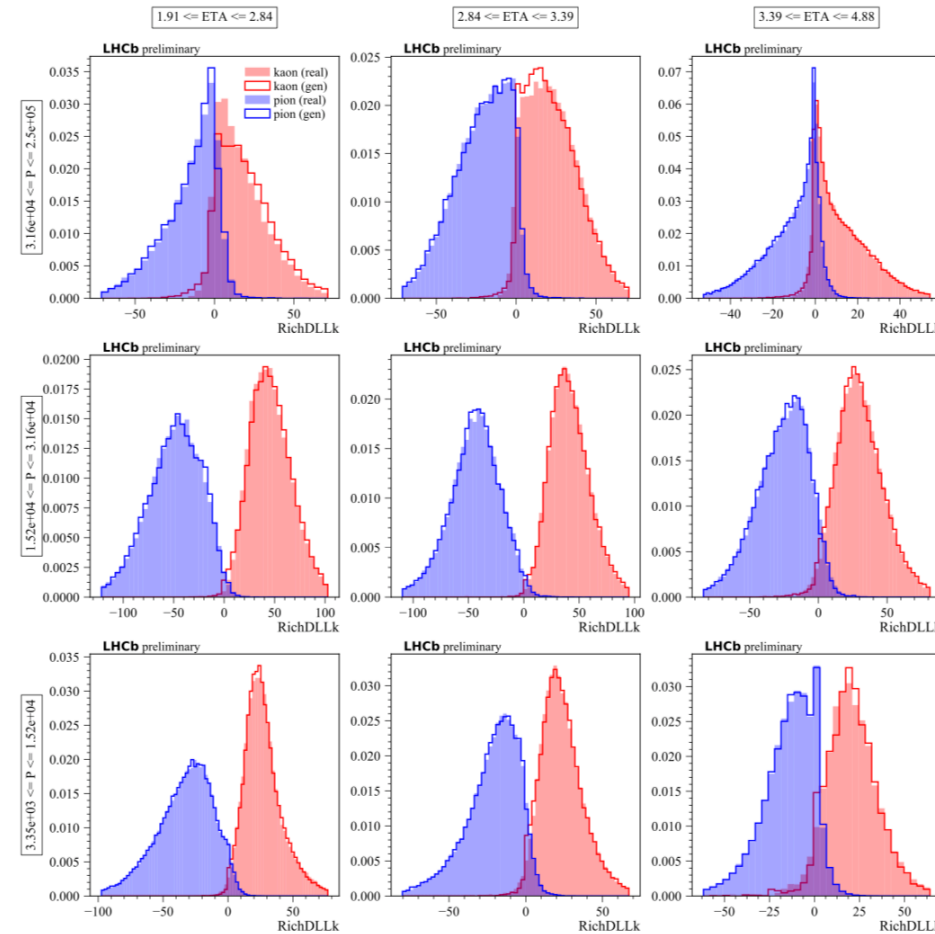
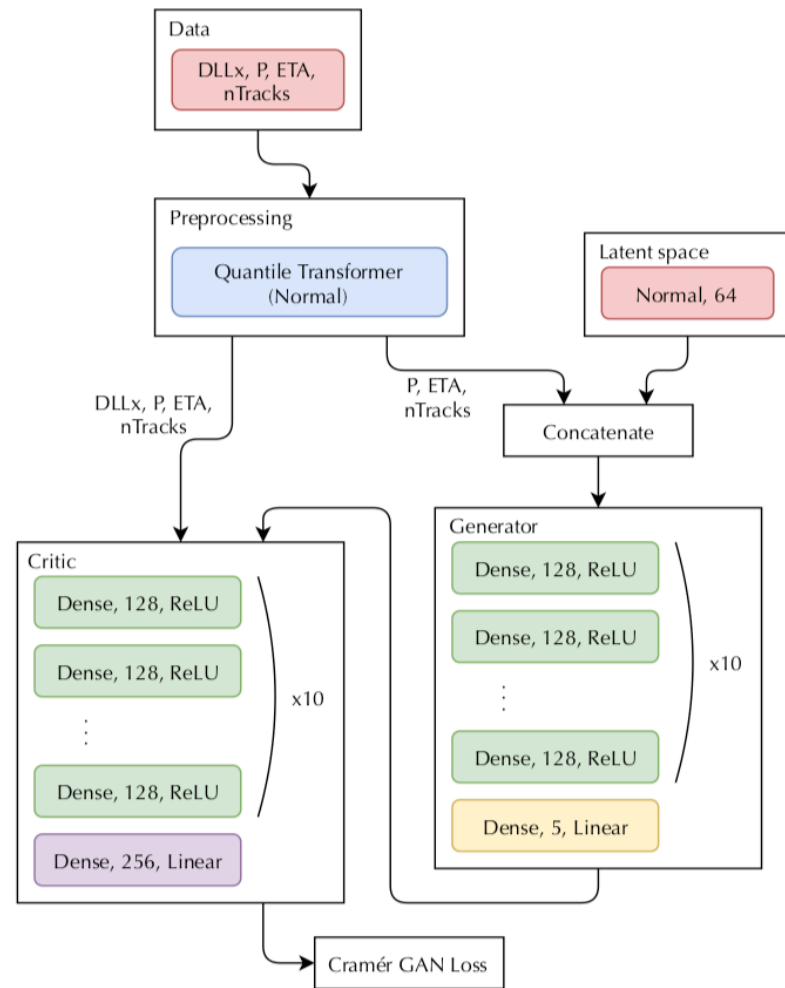
Multiple ways to connect intermediate steps with deep learning.







# Simulating Data



- 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.

Generative Adversarial Networks for LHCb Fast Simulation [\[2003.09762\]](#)

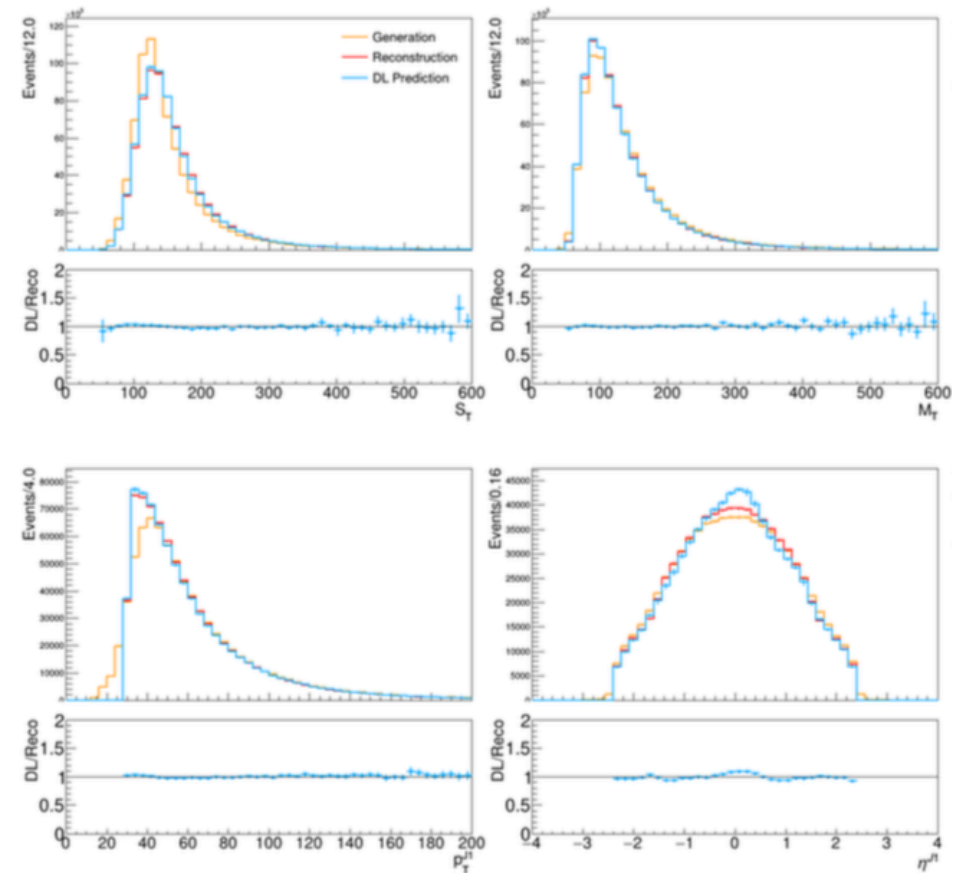
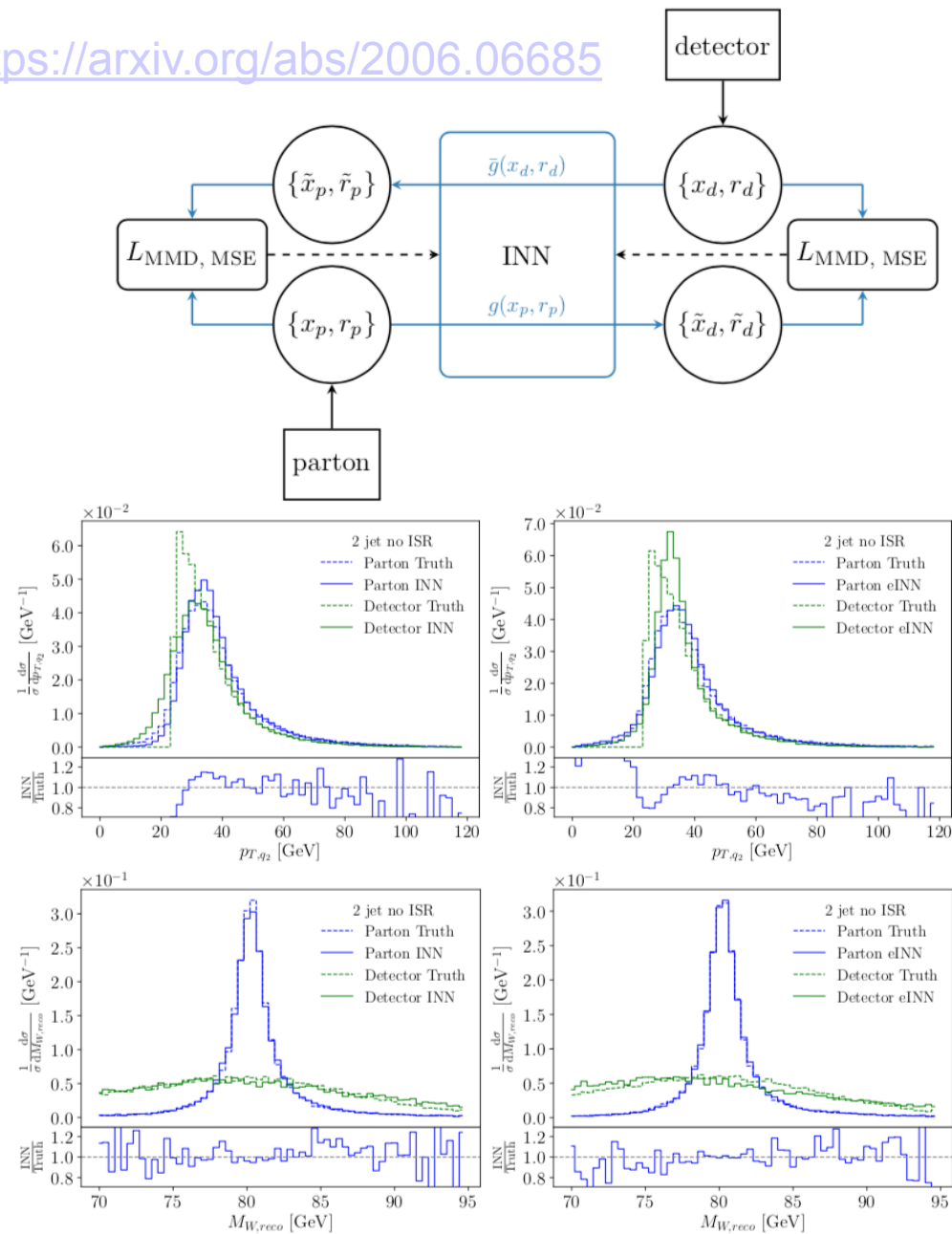
More of the relevant works at:

<https://iml-wg.github.io/HEPML-LivingReview/>

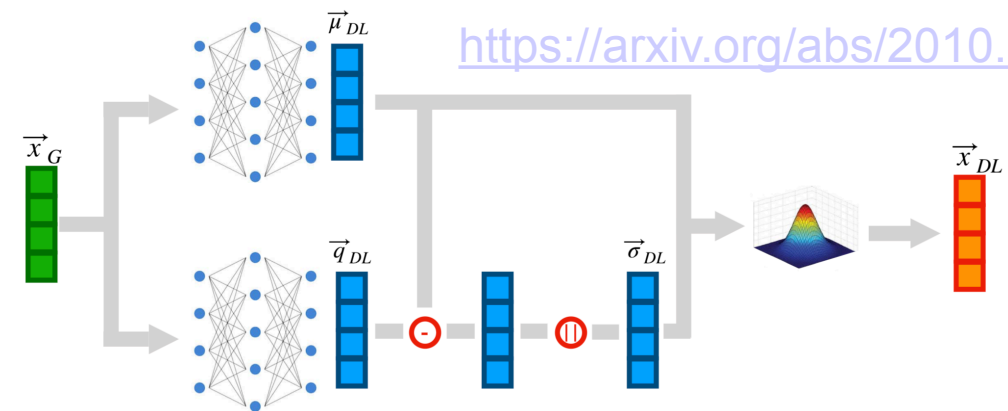


# Suiting Models

<https://arxiv.org/abs/2006.06685>



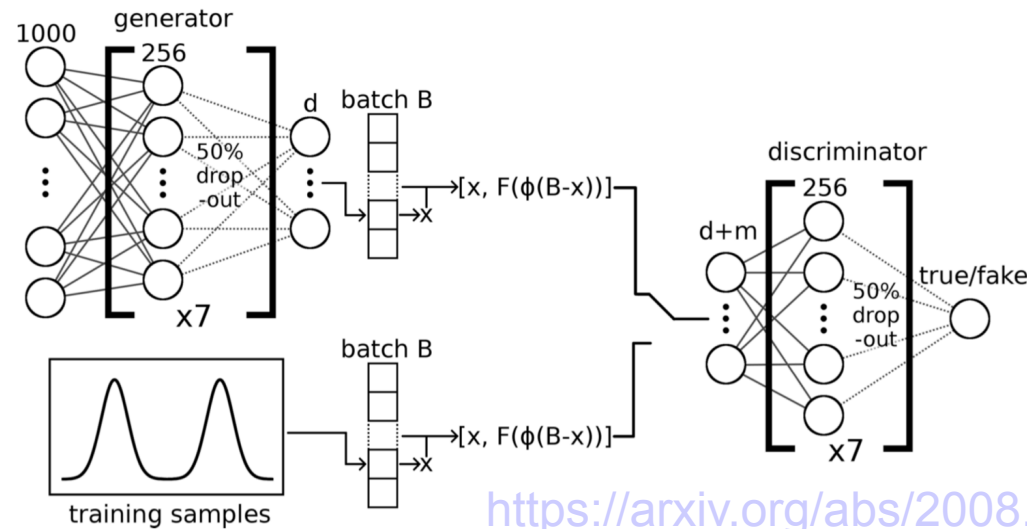
<https://arxiv.org/abs/2010.01835>



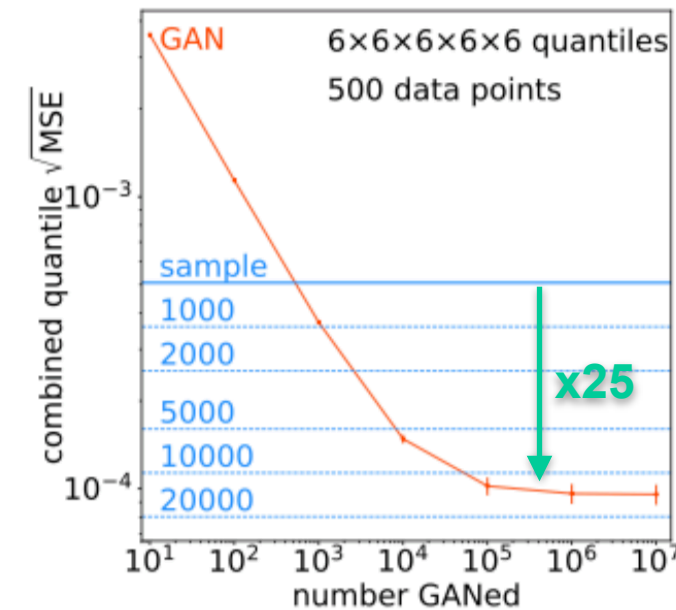
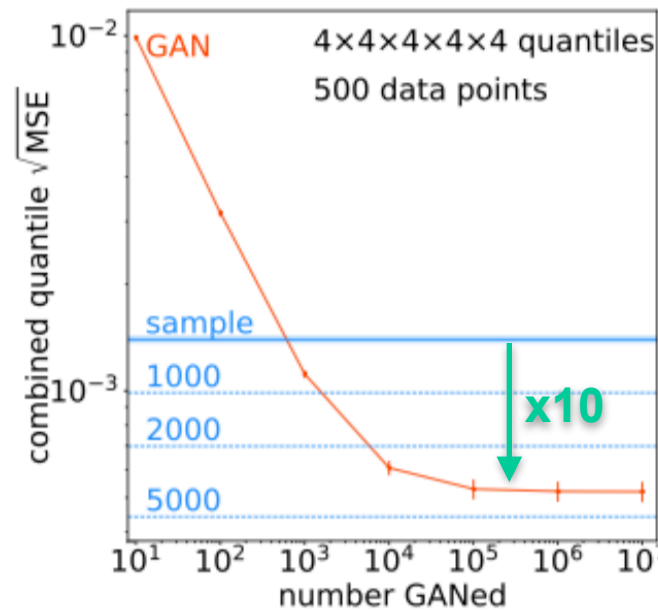
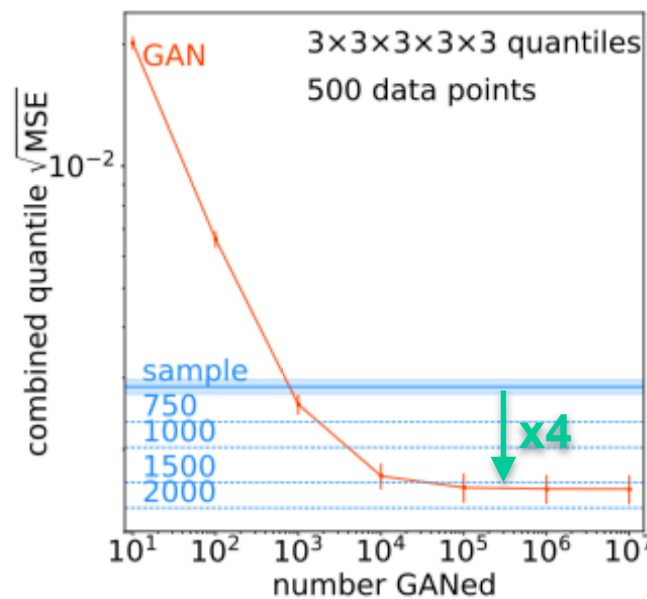
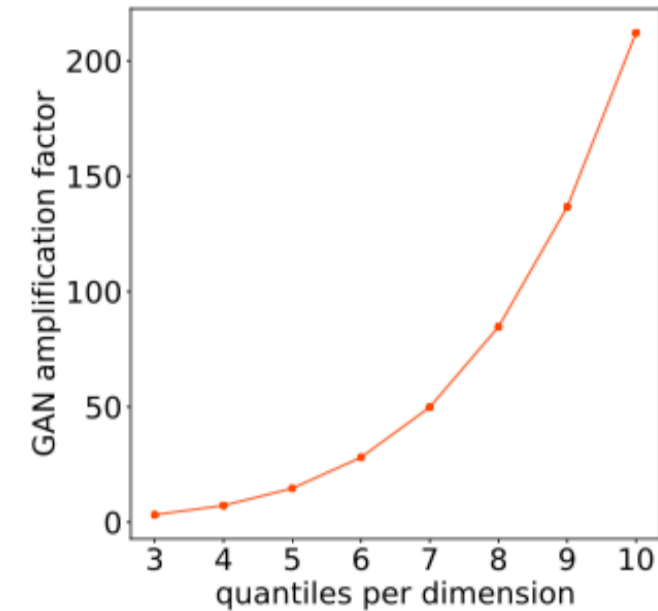
Learn the parton  $\Rightarrow$  detector function instead of generating samples from vacuum.



# Statistical Power

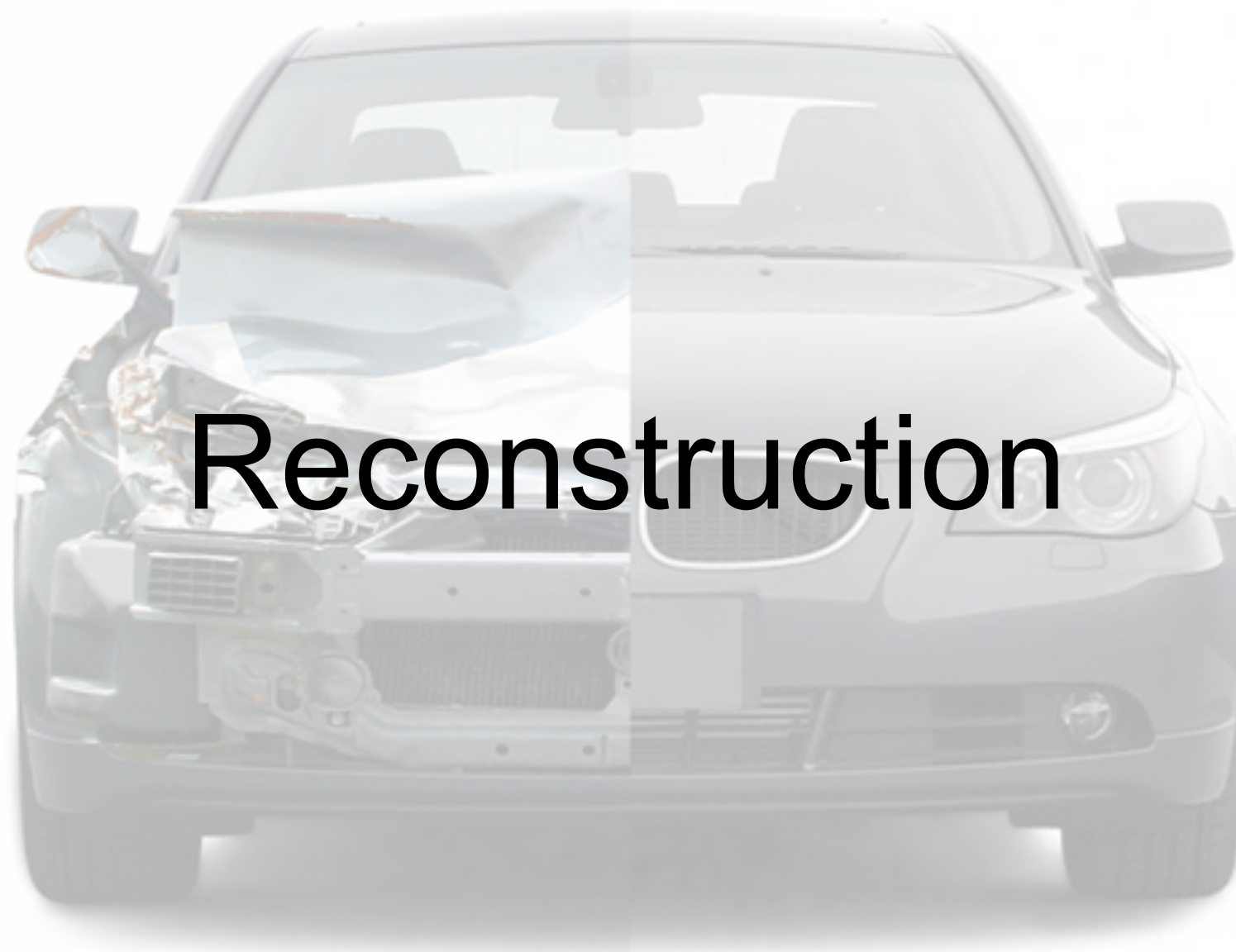


<https://arxiv.org/abs/2008.06545>



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

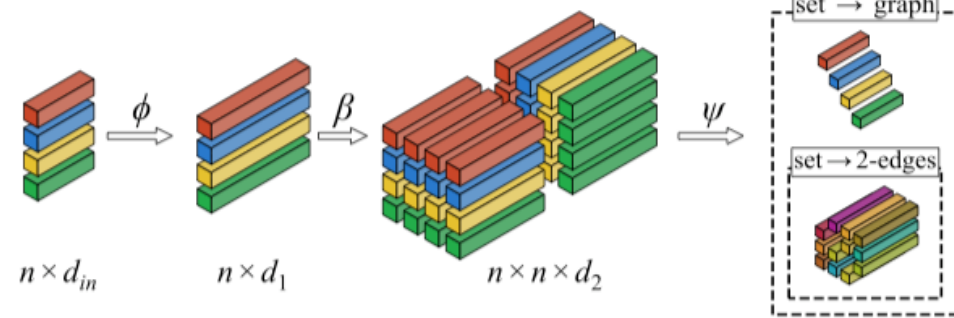
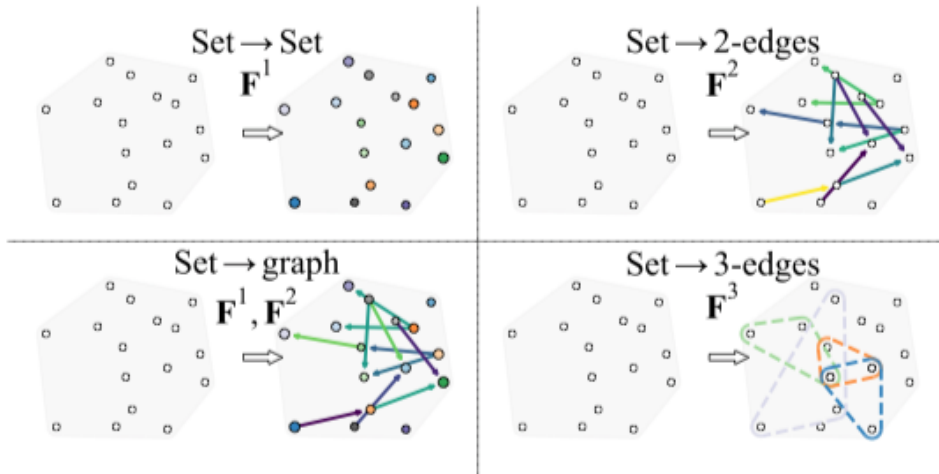




# Reconstruction

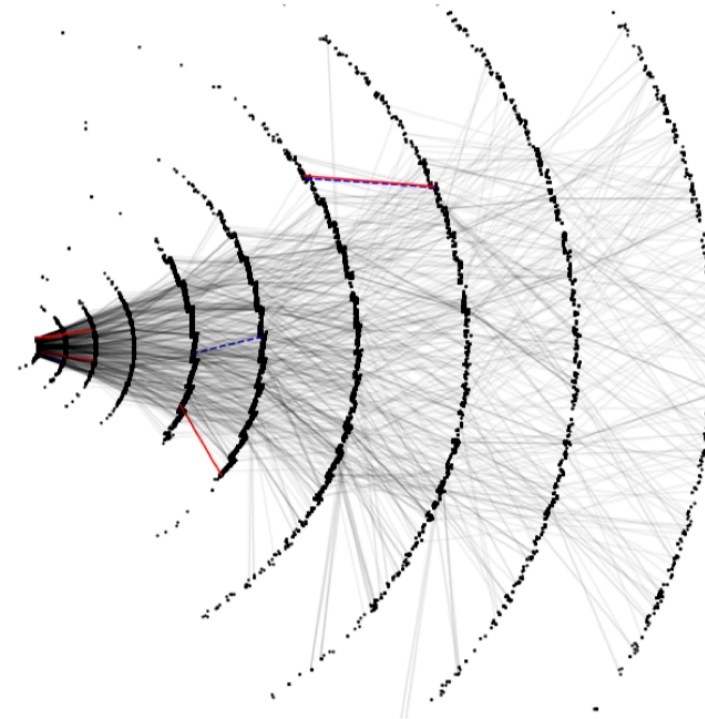
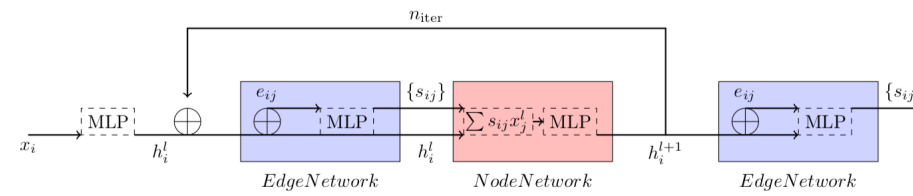


# Reconstructing Data



Learning graphs from sets, applied to vertexing  
[\[2002.08772\]](https://arxiv.org/abs/2002.08772)

GNN applied to charged particle tracking  
[\[2007.00149\]](https://arxiv.org/abs/2007.00149)

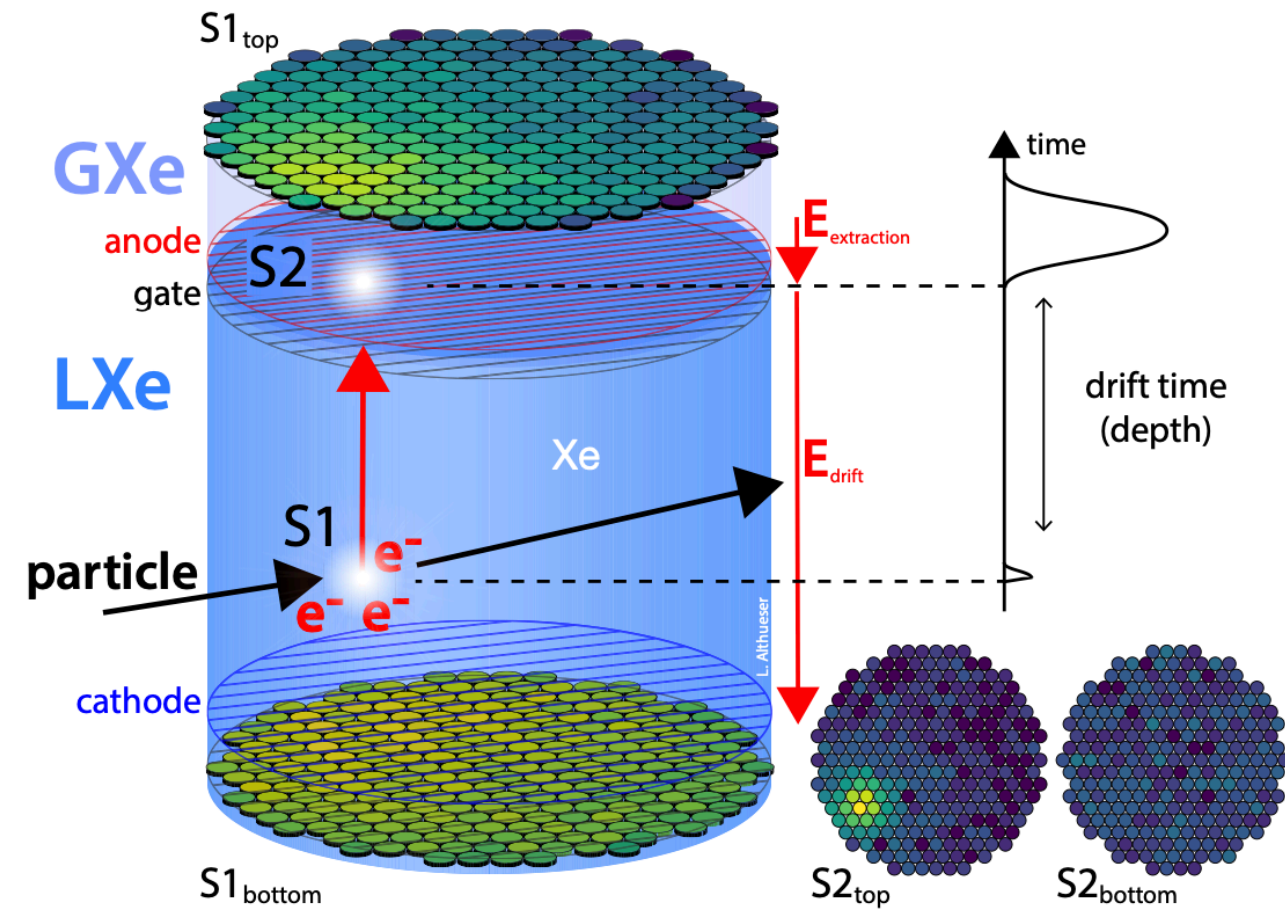


- Event reconstruction/tagging is pattern recognition to a large extent. Advanced machine learning techniques can help.
- Learn from the simulation, and/or data.
- Learn from existing “slow reconstruction” or simulation ground truth.
- Automatically adapt algorithm to new detector design.
- Image base 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/>



# For example

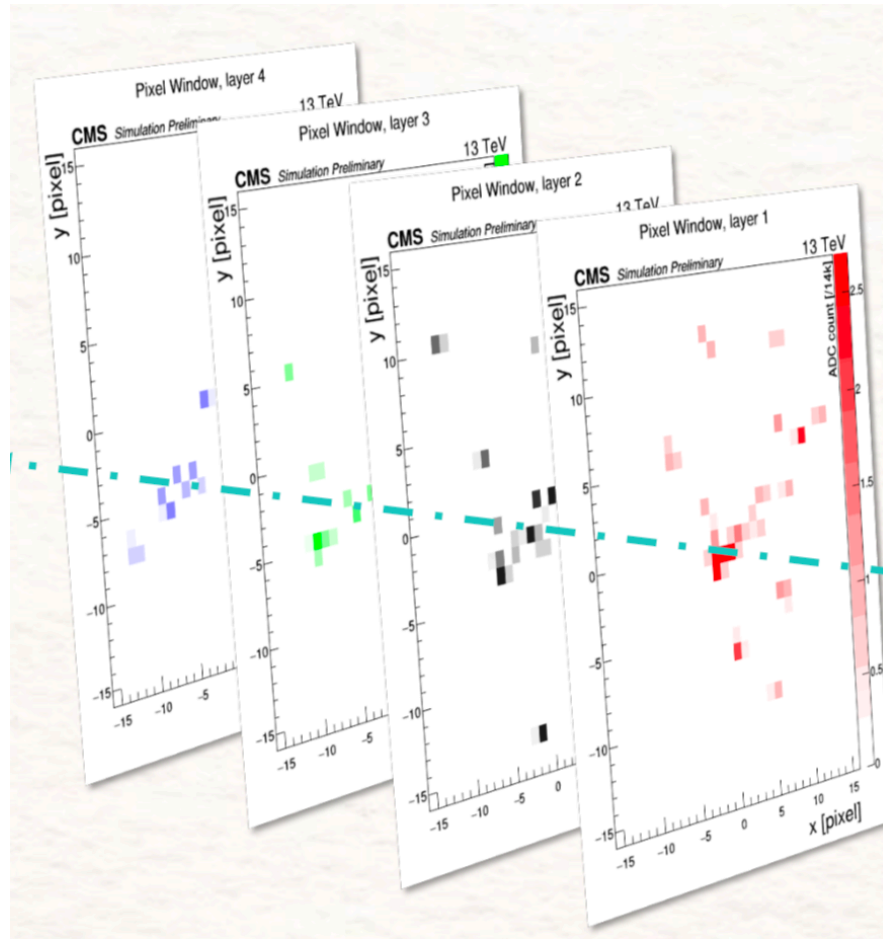


- Signal detection
  - Light signal (S1)
  - Charge signal (S2)
- Energy reconstruction
- 3D position reconstruction

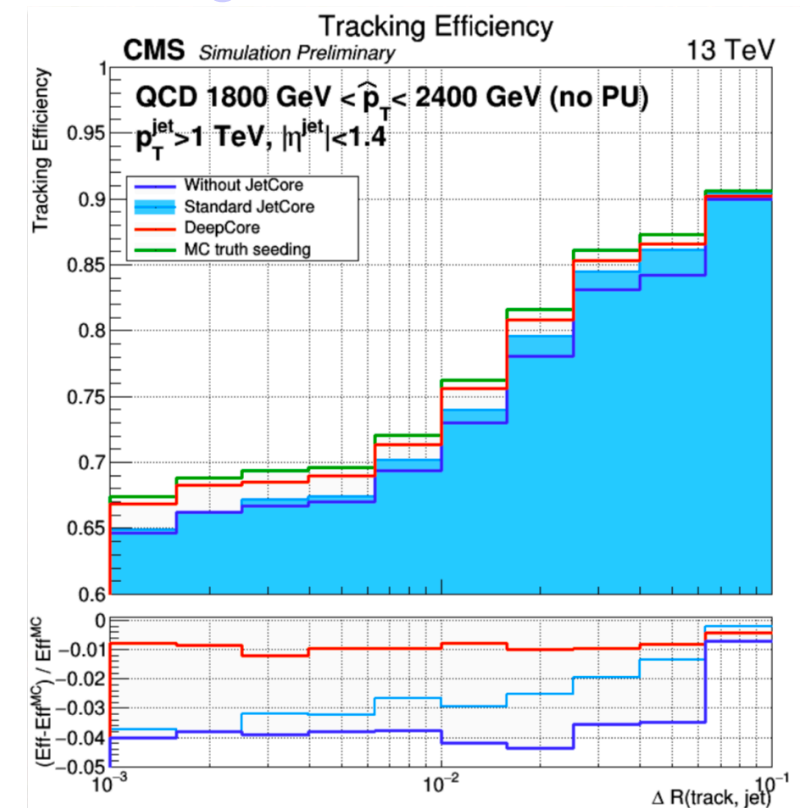
Search for Solar Axions and ALP Dark Matter with XENONnT [[Jingqiang Ye](#)]



# Track Seed Finding in Jets



## Deep Core Algorithm



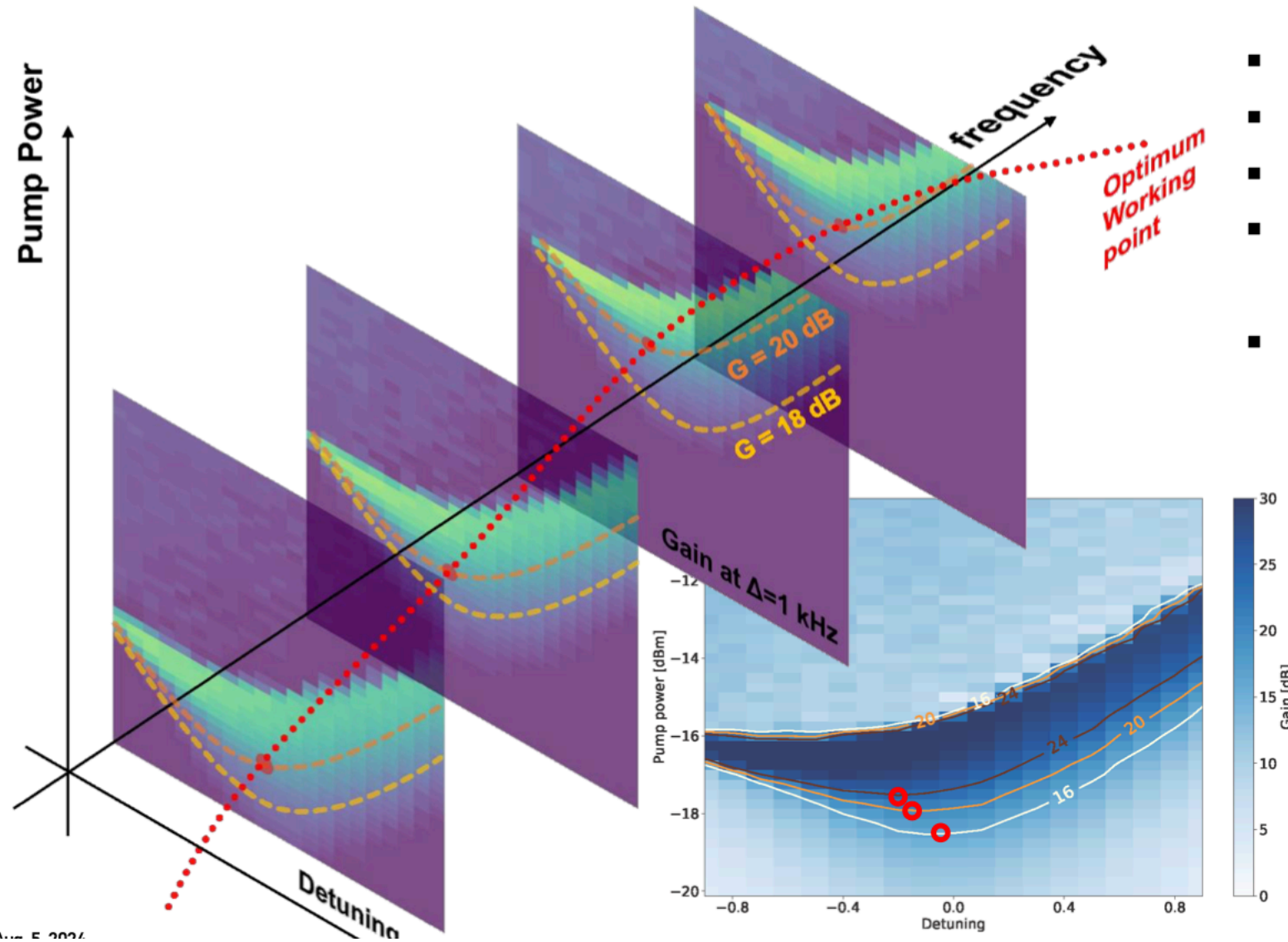
Predict tracklets parameters from raw pixels using CNN  
Approaching the maximum reachable performance



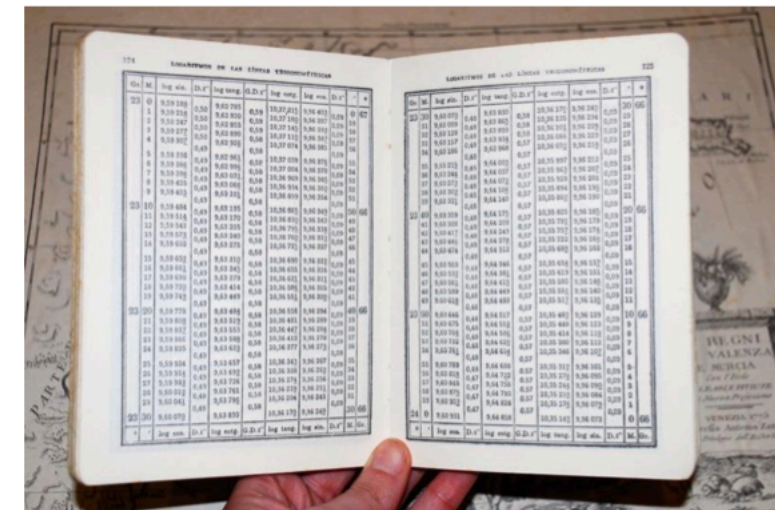
# For example

Caglar Kutlu et al  
arXiv:2305.088

## The Josephson Parametric Amplifier



- 2D JPA Paramap ( $f_{\text{passive}}, P_{\text{pump}}$ )
- Gain contour
- Lower pump power  $\rightarrow$  Lower added noise
- Get the lowest  $P_{\text{pump}}$  in the contour of given target gain (typically 20 dB)
- Every 0.2 ~ 0.5 MHz, interpolated for each tuning step

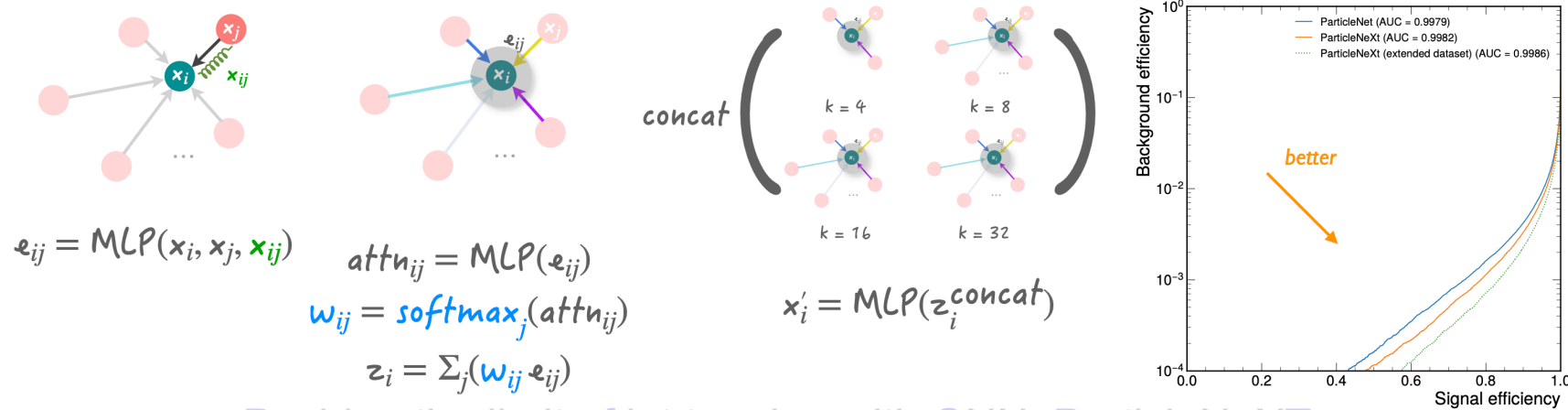


The CAPP-MAX axion haloscope: status and prospects [[Saebyeok Ahn](#)]



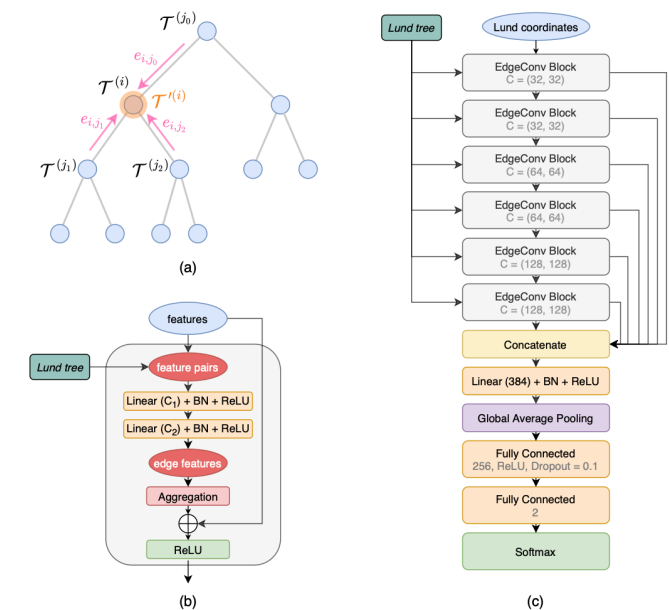
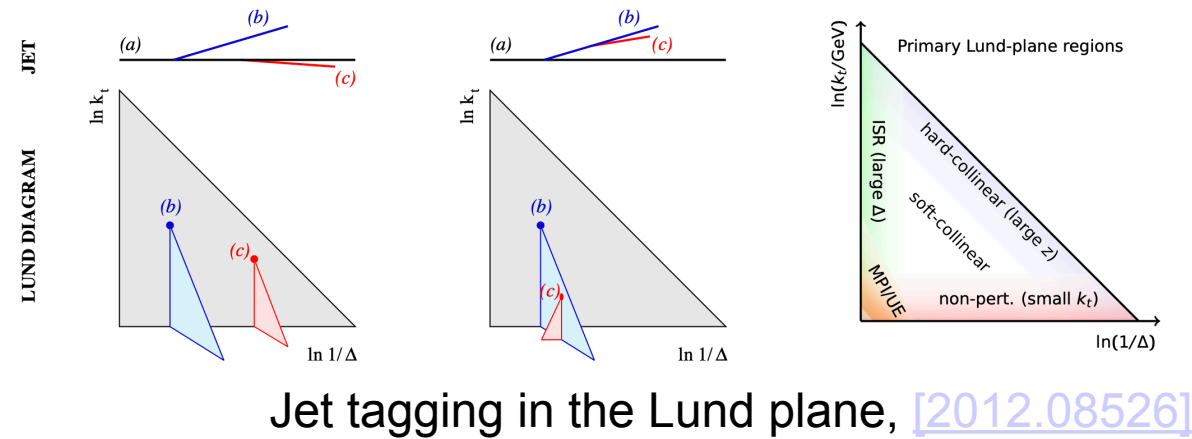


# Jet Tagging



Pushing the limit of jet tagging with GNN, ParticleNeXT

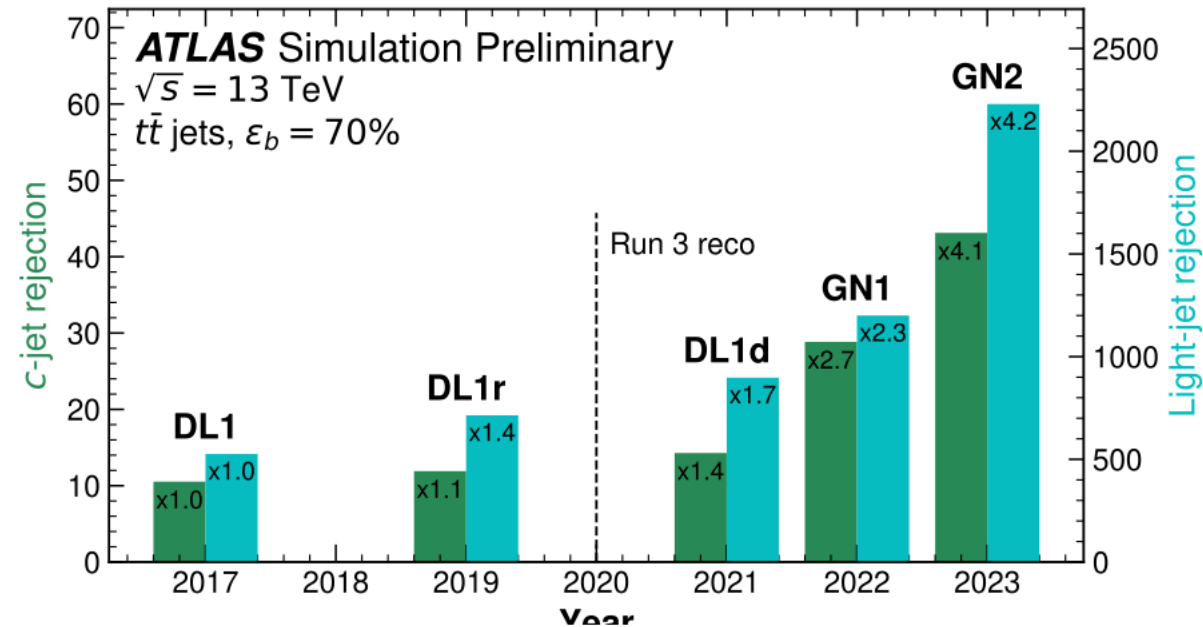
The Lund jet plane provides an efficient description of the radiation patterns within a jet



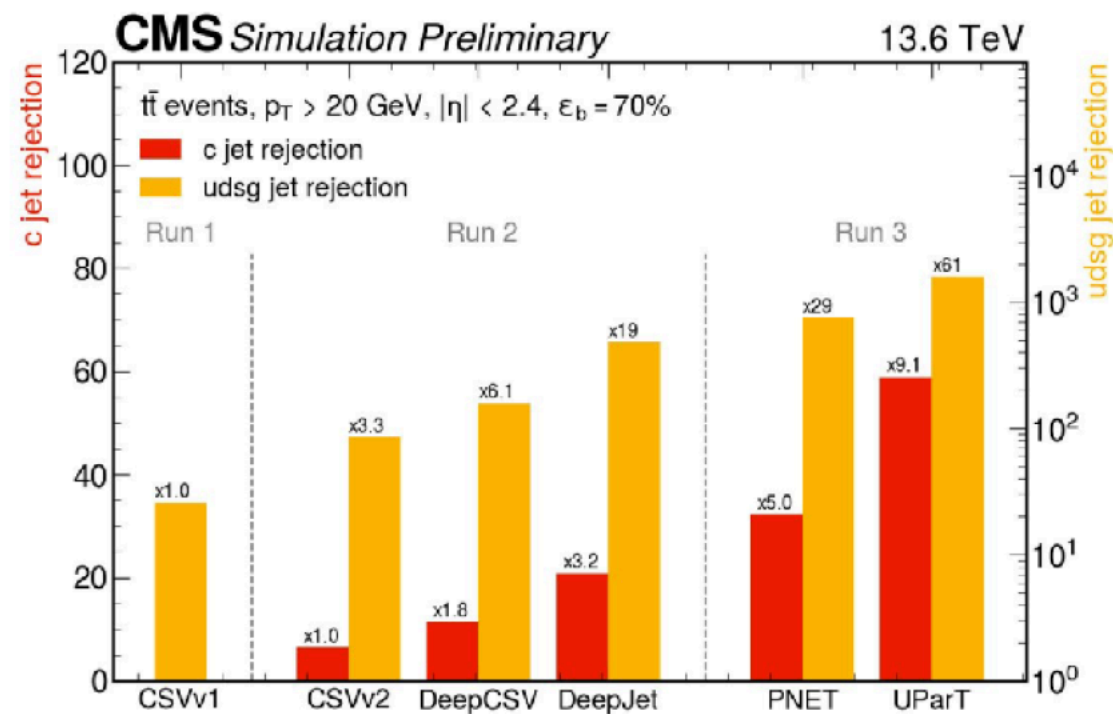
State of art graph based models are now supplanted by transformer based models.  
 Still a very rich field, in particular in developing inductive bias in the model  
 (symmetry, invariance, ... ).  
 Kinematic regression, substructure assignment, ... also possible.



# Jet Tagging



Flavor-tagging and flavour-rejection has improved significantly over the last years thanks to deep learning.

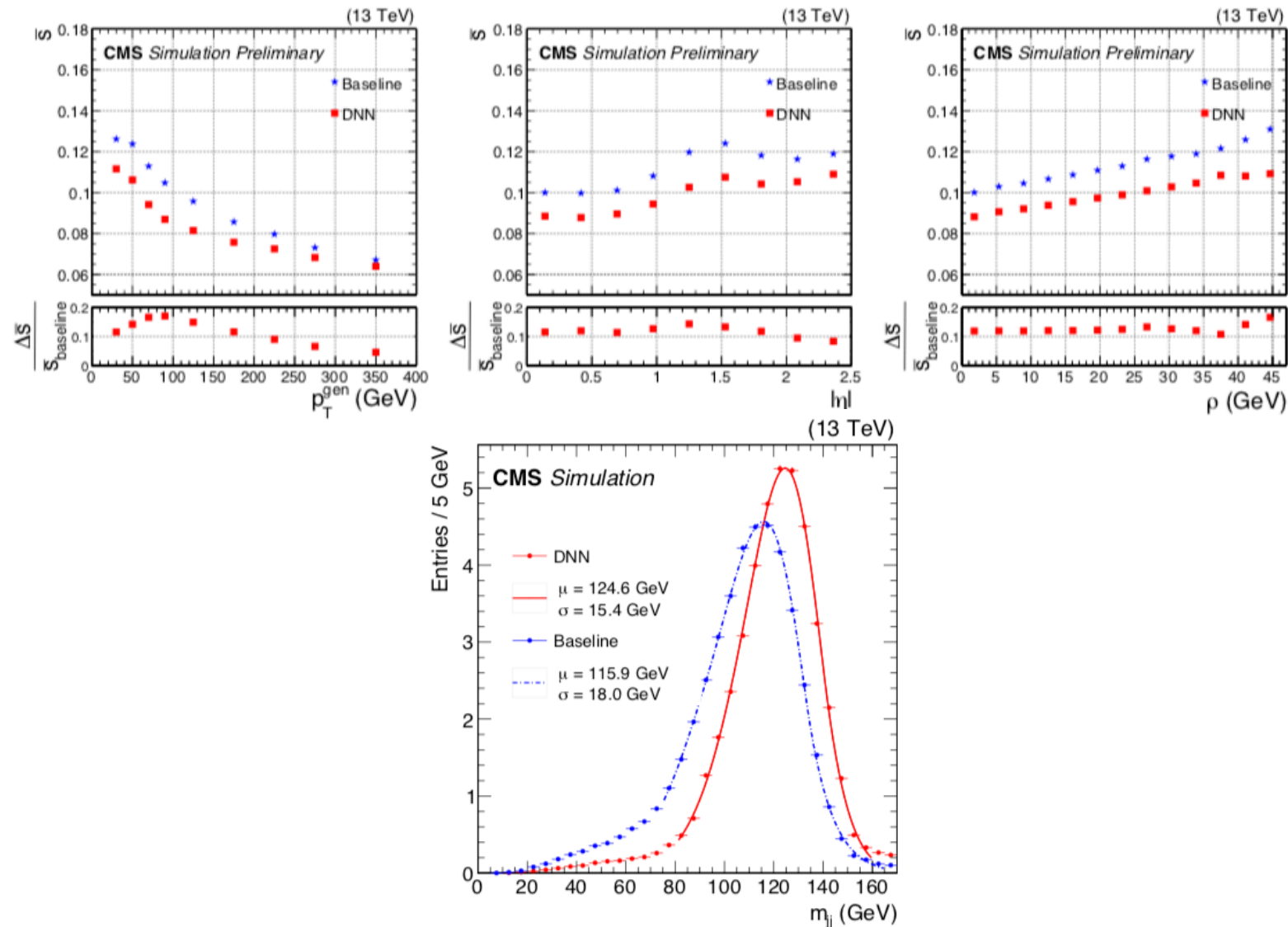


Boosting analysis efficiency and significance.

ICHEP 24 Summary [[Marumi Kado](#)]



# Calibrating Data

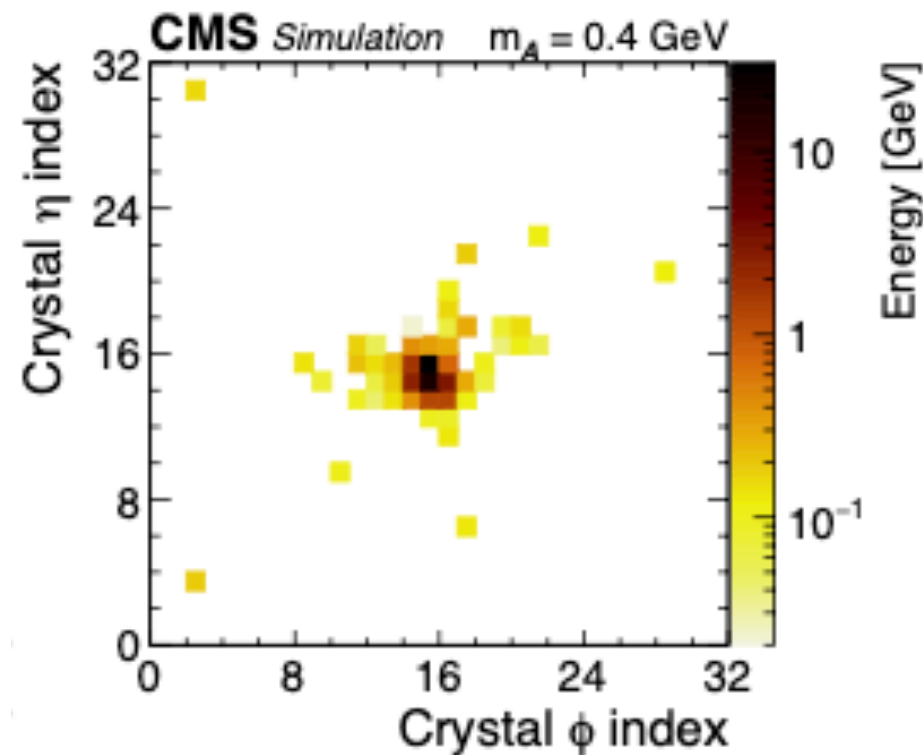


- Energy regression is the most obvious use case.
- Learning calibrating models from simulation and data.
- Parametrisation of scale factors using neural networks.
- Reducing data/simulation dependency using domain adaptation.
- Continued R&D

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

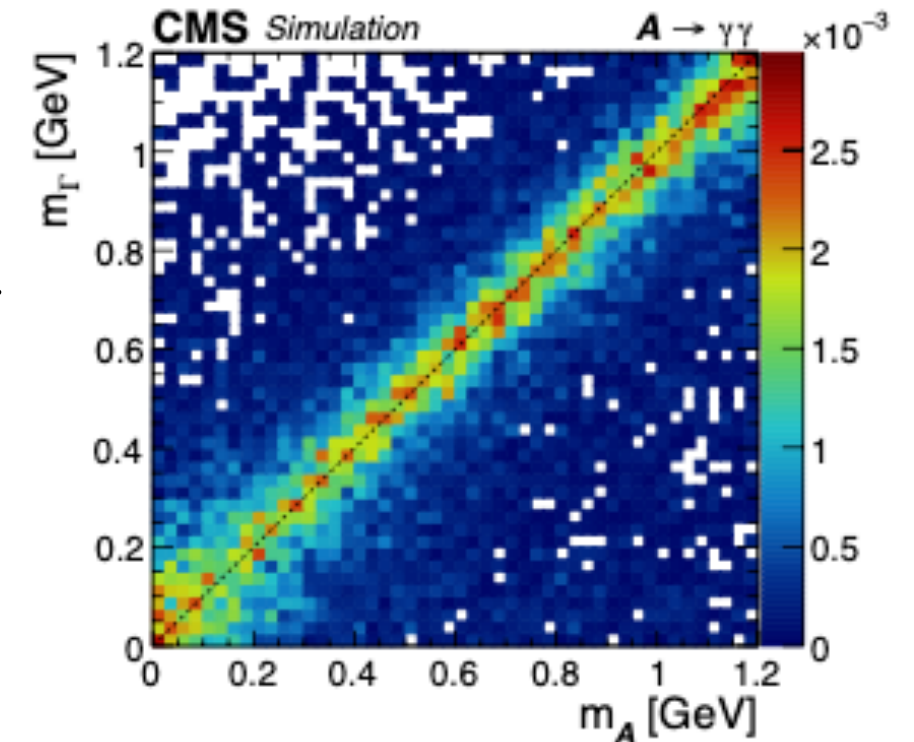


# di-photon Mass Regression



RESNET for mass regression  
+  
domain continuation a low mass

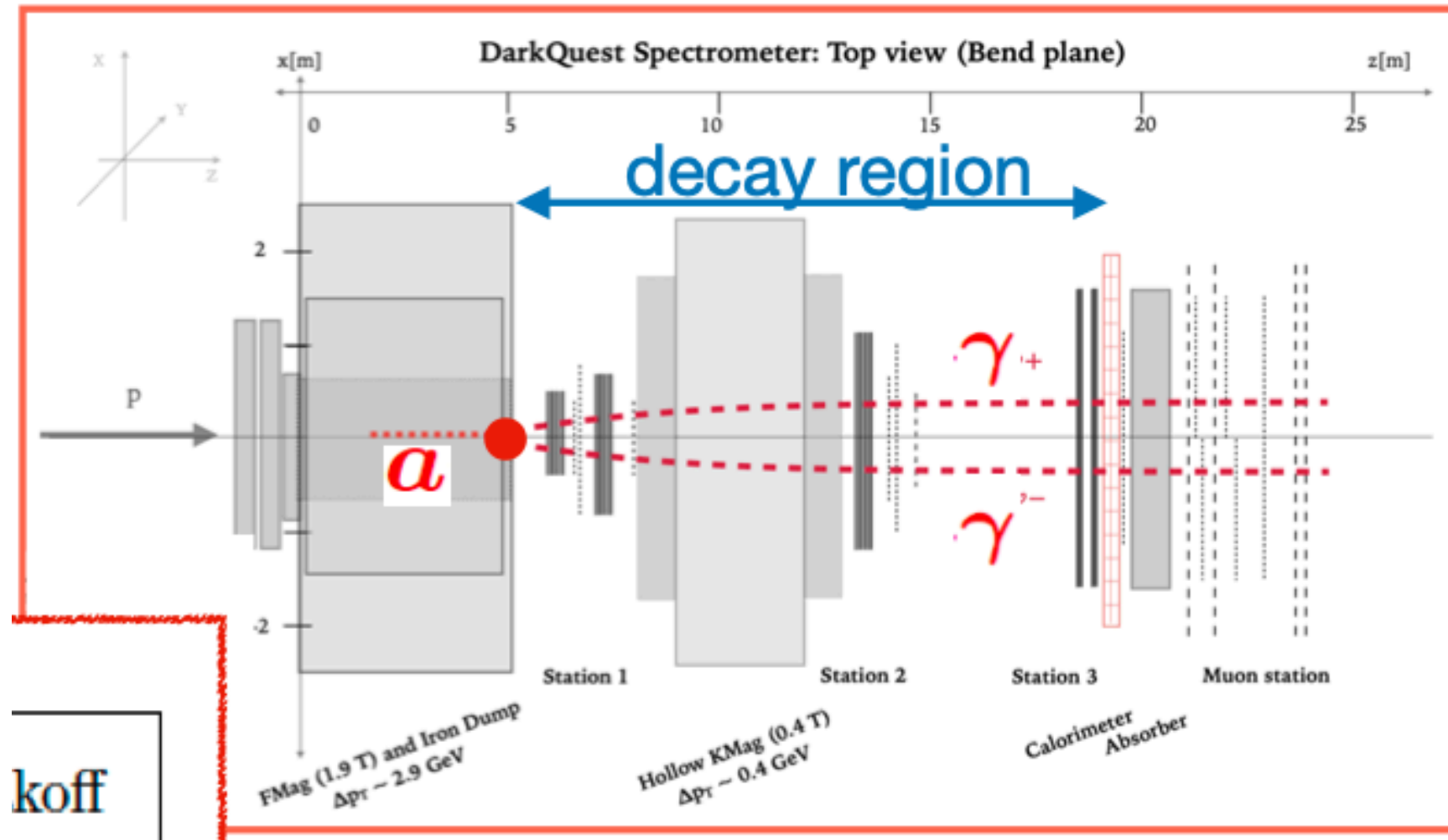
[10.1103/PhysRevD.108.052002](https://arxiv.org/abs/10.1103/PhysRevD.108.052002)



Learn the  $a$ /di-photon mass from the energy deposition in ECAL  
RESNET-based [\[1512.03385\]](https://arxiv.org/abs/1512.03385) model to process ECAL image  
Unprecedented reach at low mass

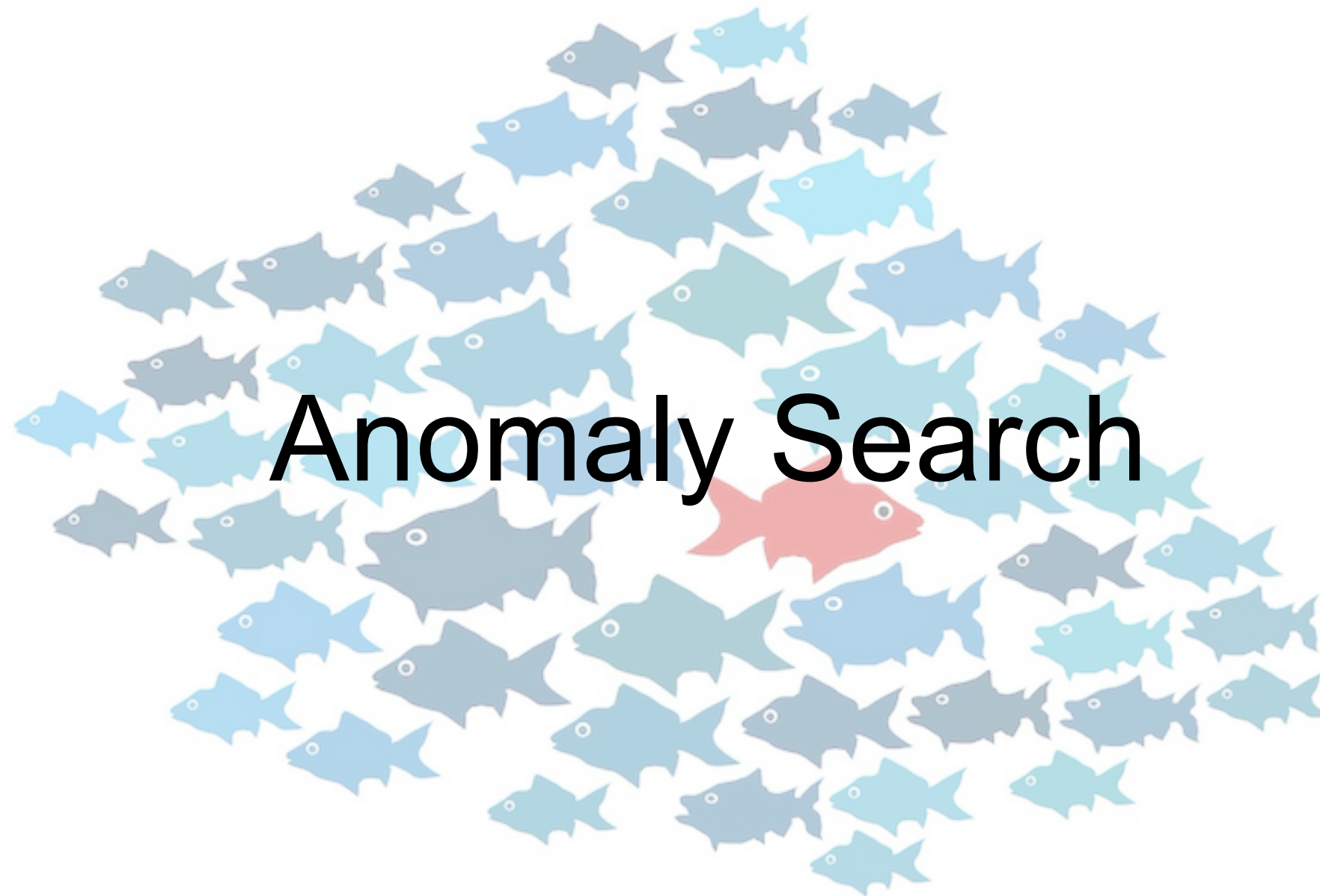


# For example



ALP searches with fixed target experiments [[Stefania Gori](#)]





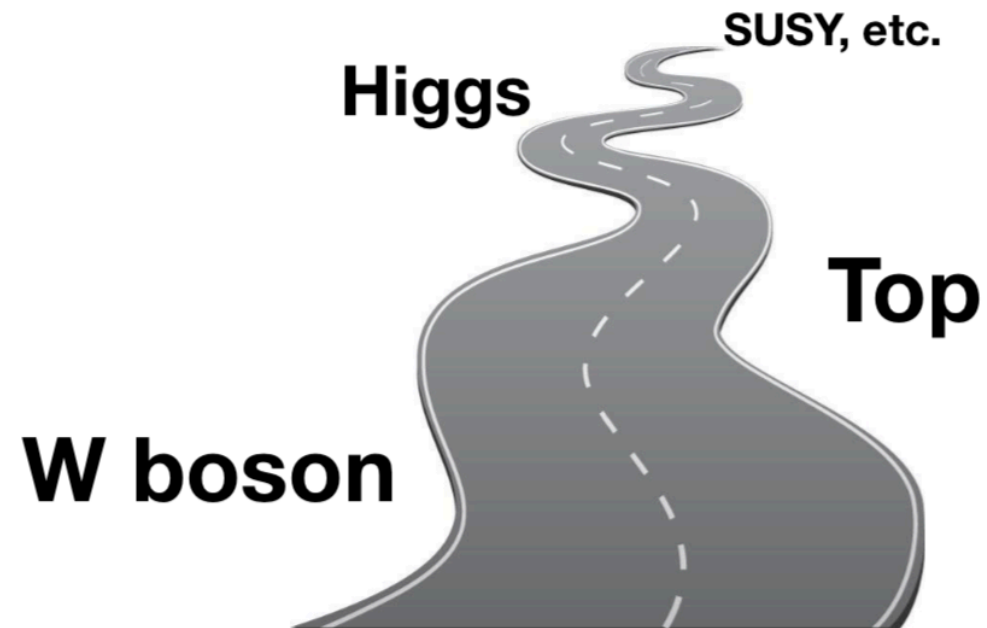
# Anomaly Search



# 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 $H_1$**

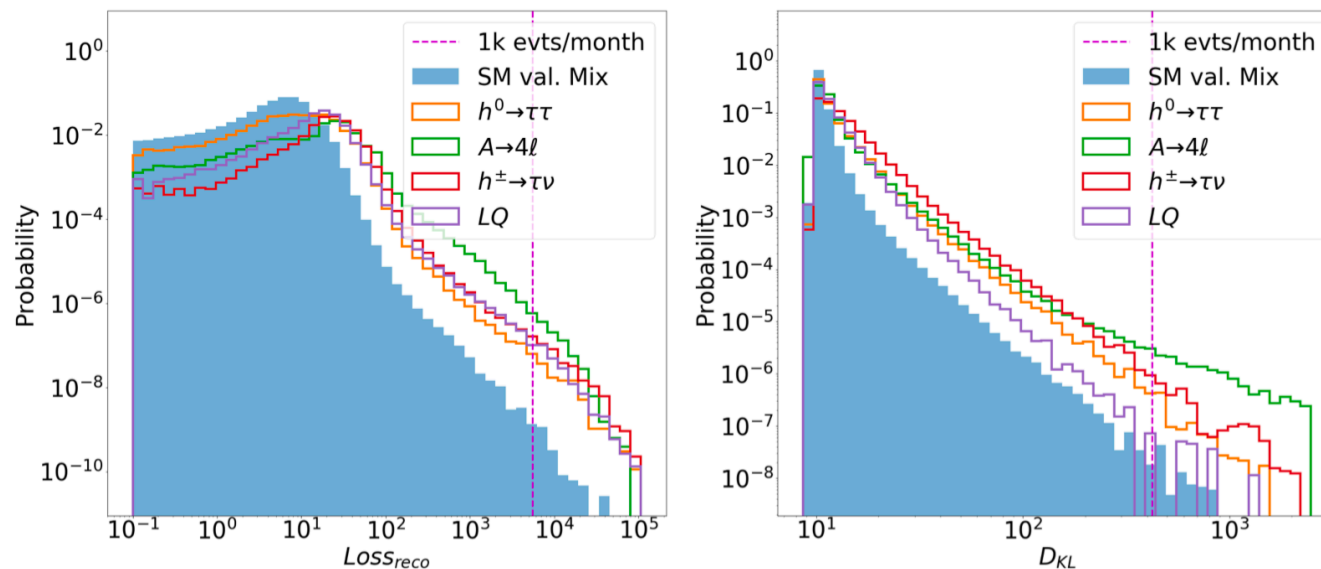
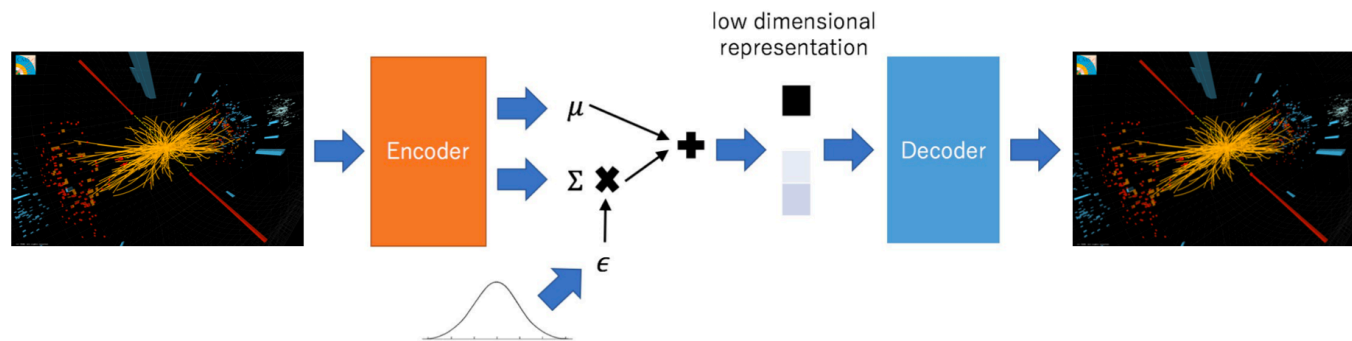
**Huge set** of alternatives

Case-by-case optimisation **unfeasible**

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



# 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\]](https://doi.org/10.1007/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.





# “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.

## Individual Approaches

LHC Olympics 2020 [\[2101.08320\]](#)

### 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

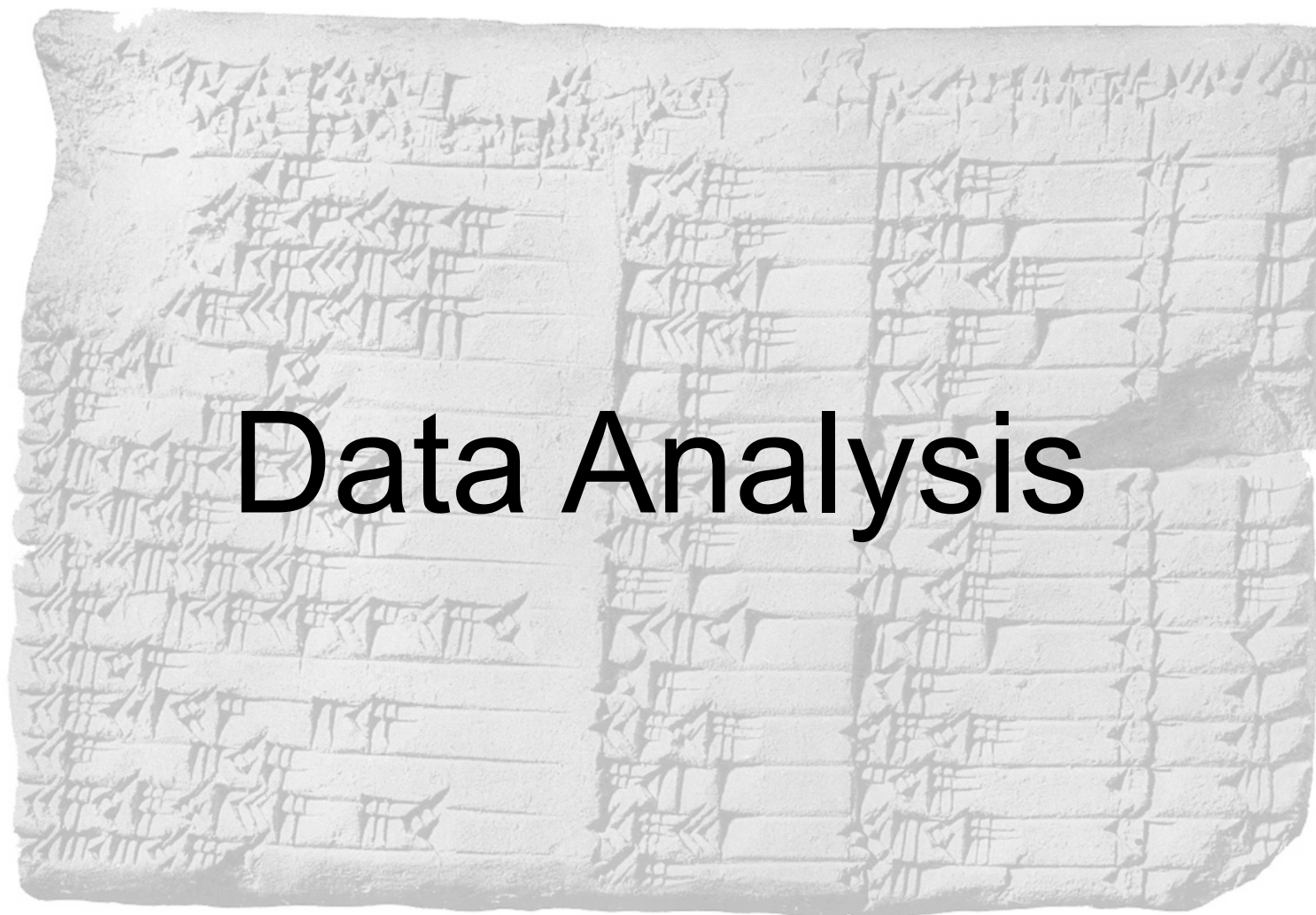
### 4 Weakly Supervised

- 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

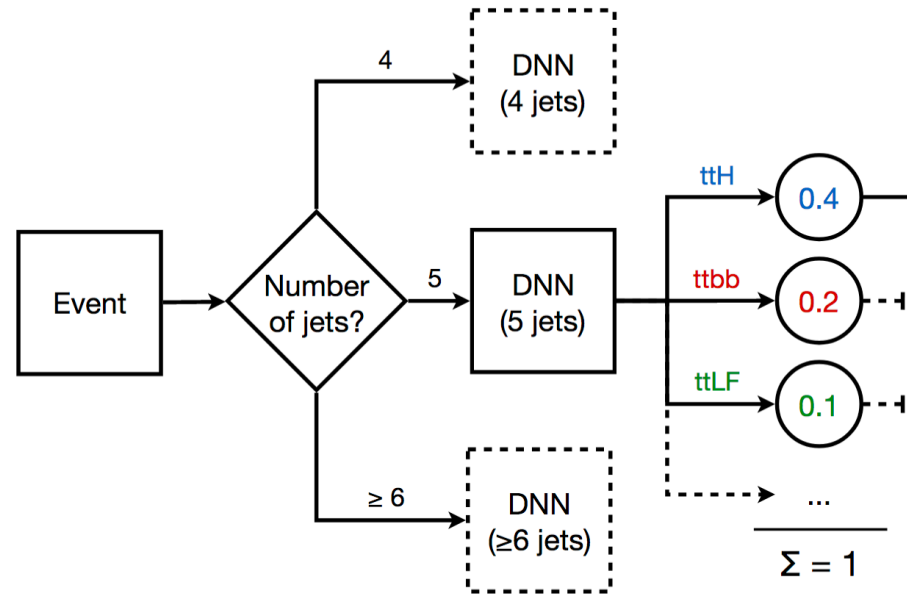




# Data Analysis

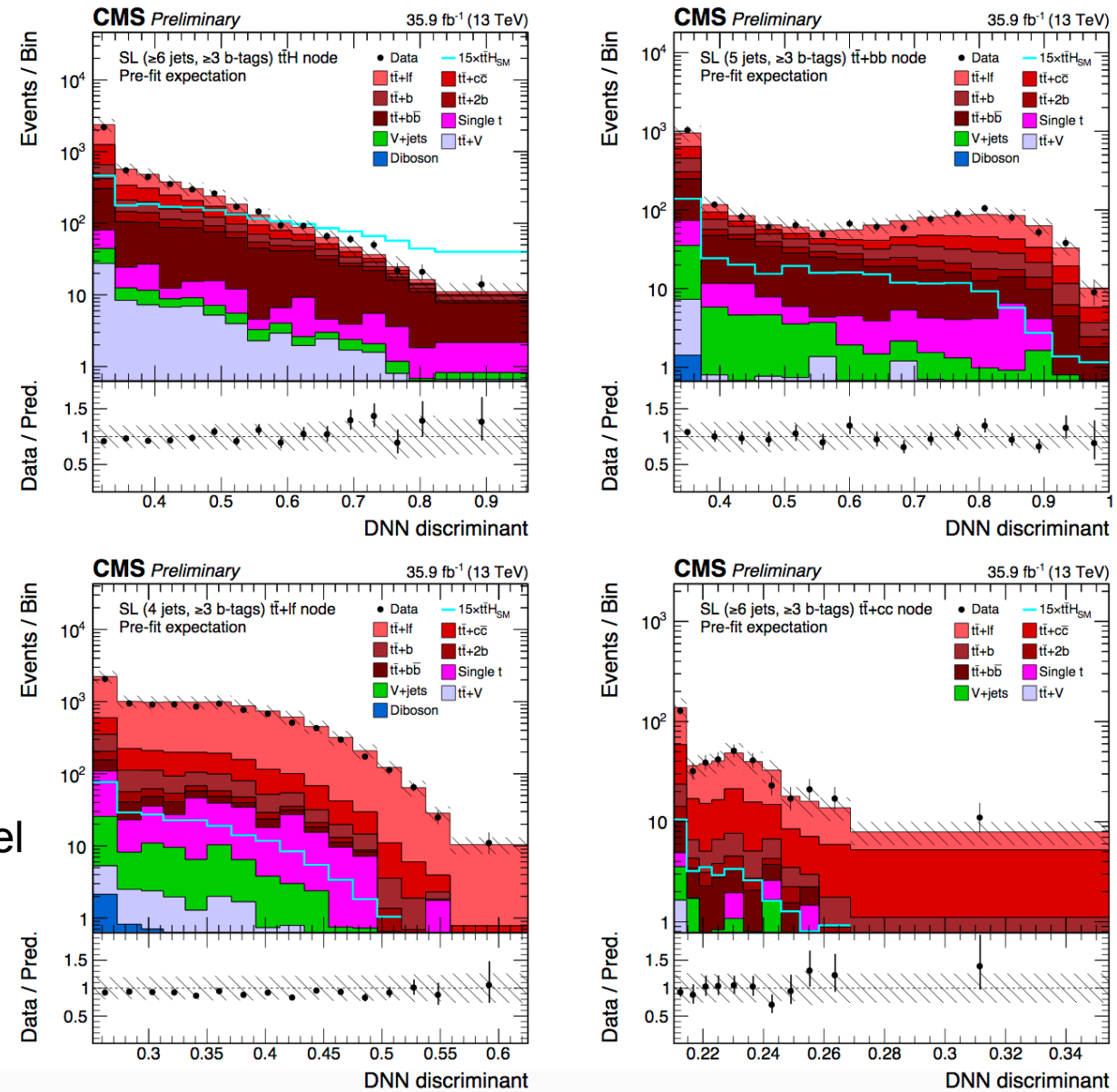


# Multi-category Classification



Slide M. Rieger

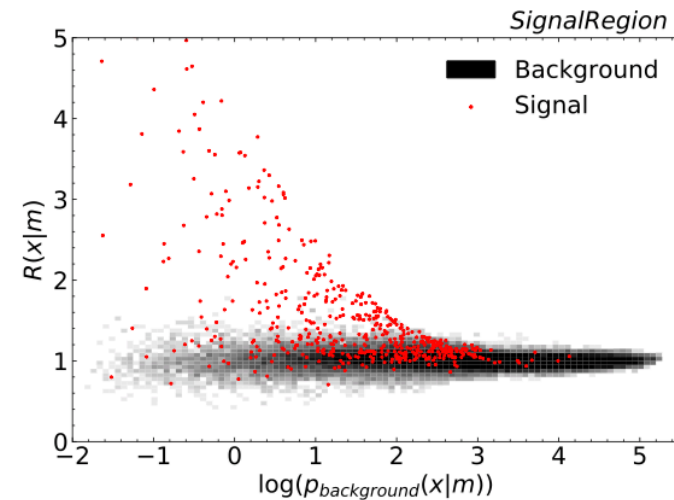
Search for ttH production in the H-to-bb decay channel with leptonic tt decays [\[cds:2308267\]](https://cds.cern.ch/record/2308267)



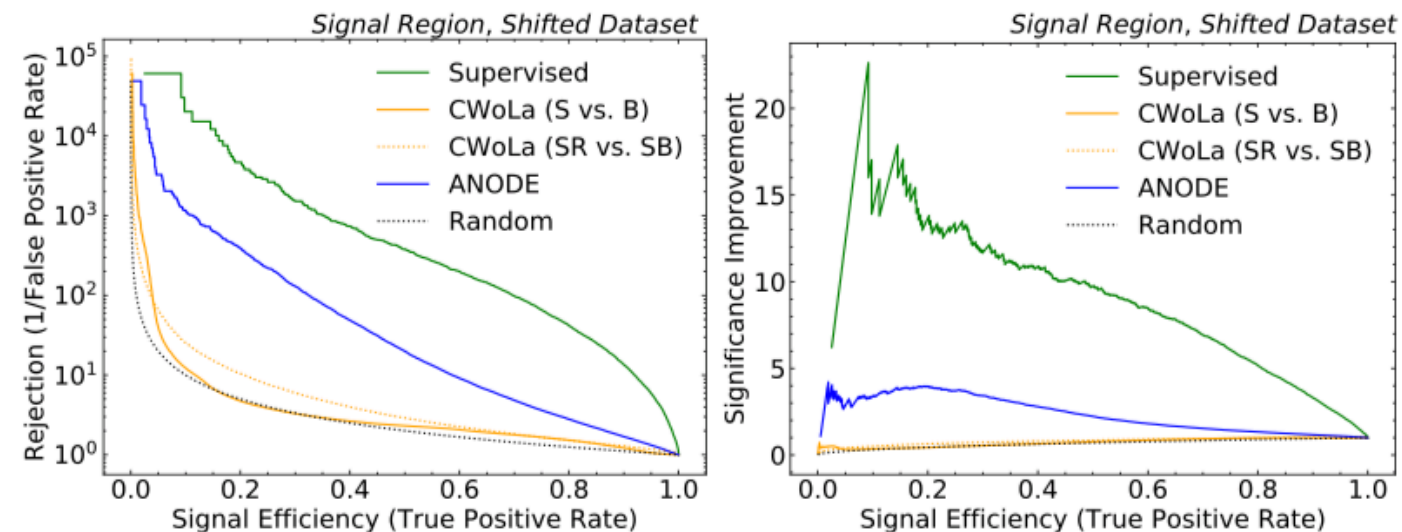
Regular analysis fit categories sub-divided using DNN output nodes for added sensitivity.



# Analysing Data



- 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.



Use of masked autoregressive density estimator with normalizing flow as model-agnostic signal enhancement mechanism.

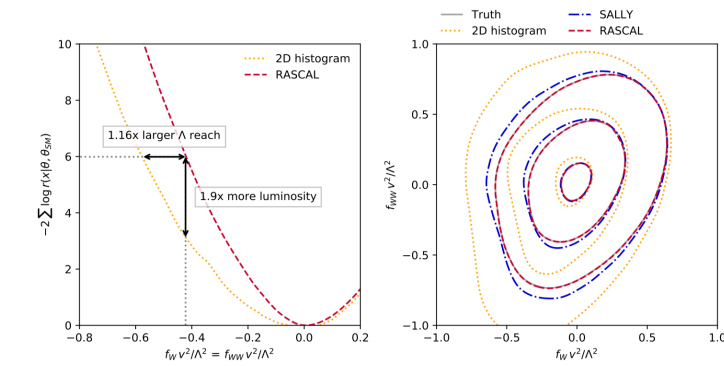
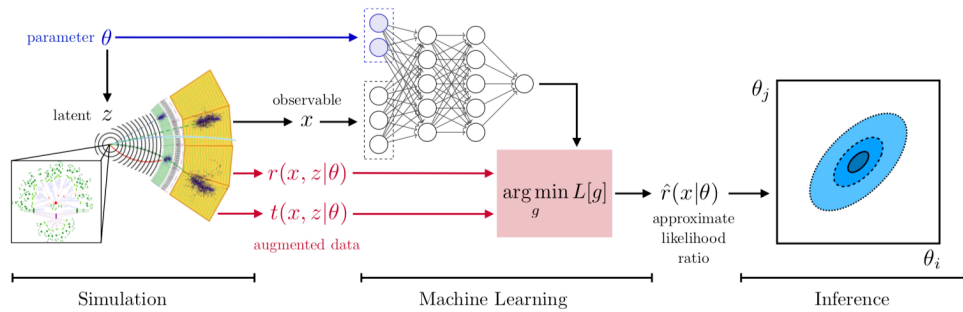
[\[doi:10.1103/PhysRevD.101.075042\]](https://doi.org/10.1103/PhysRevD.101.075042)

More of the relevant works at:

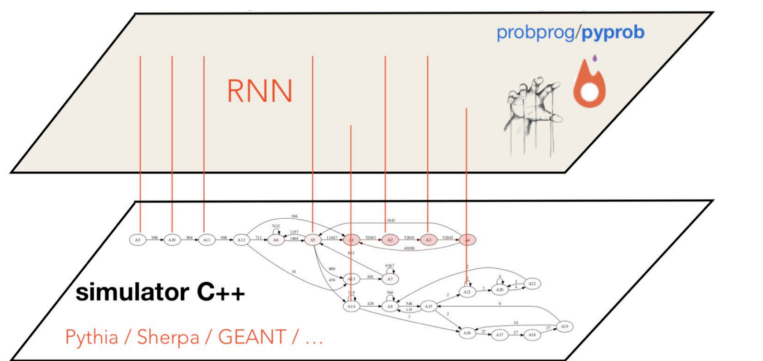
<https://iml-wg.github.io/HEPML-LivingReview/>



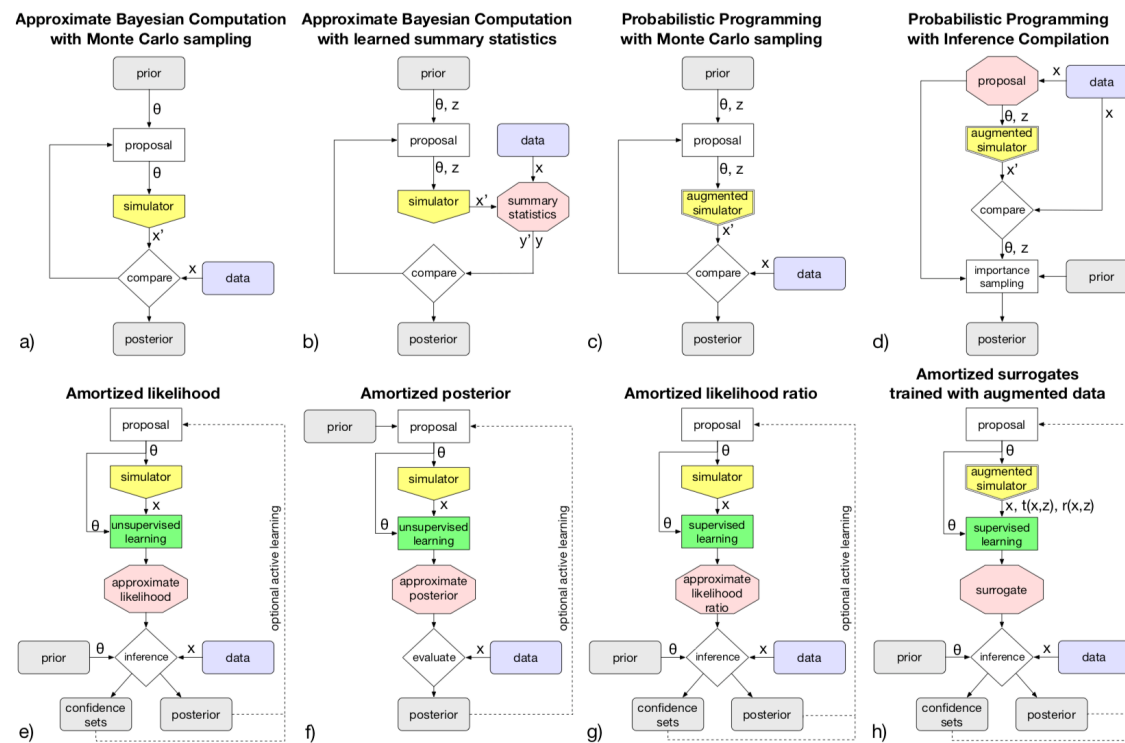
# Theory Behind the Data



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.

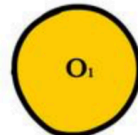
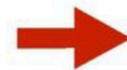
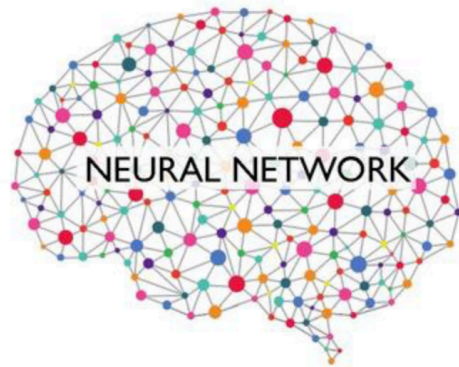
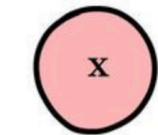


# Expected Parameter Values

## Regression by neural network

Lattice QCD gauge field

$\sim 10^7-10^9$  real numbers

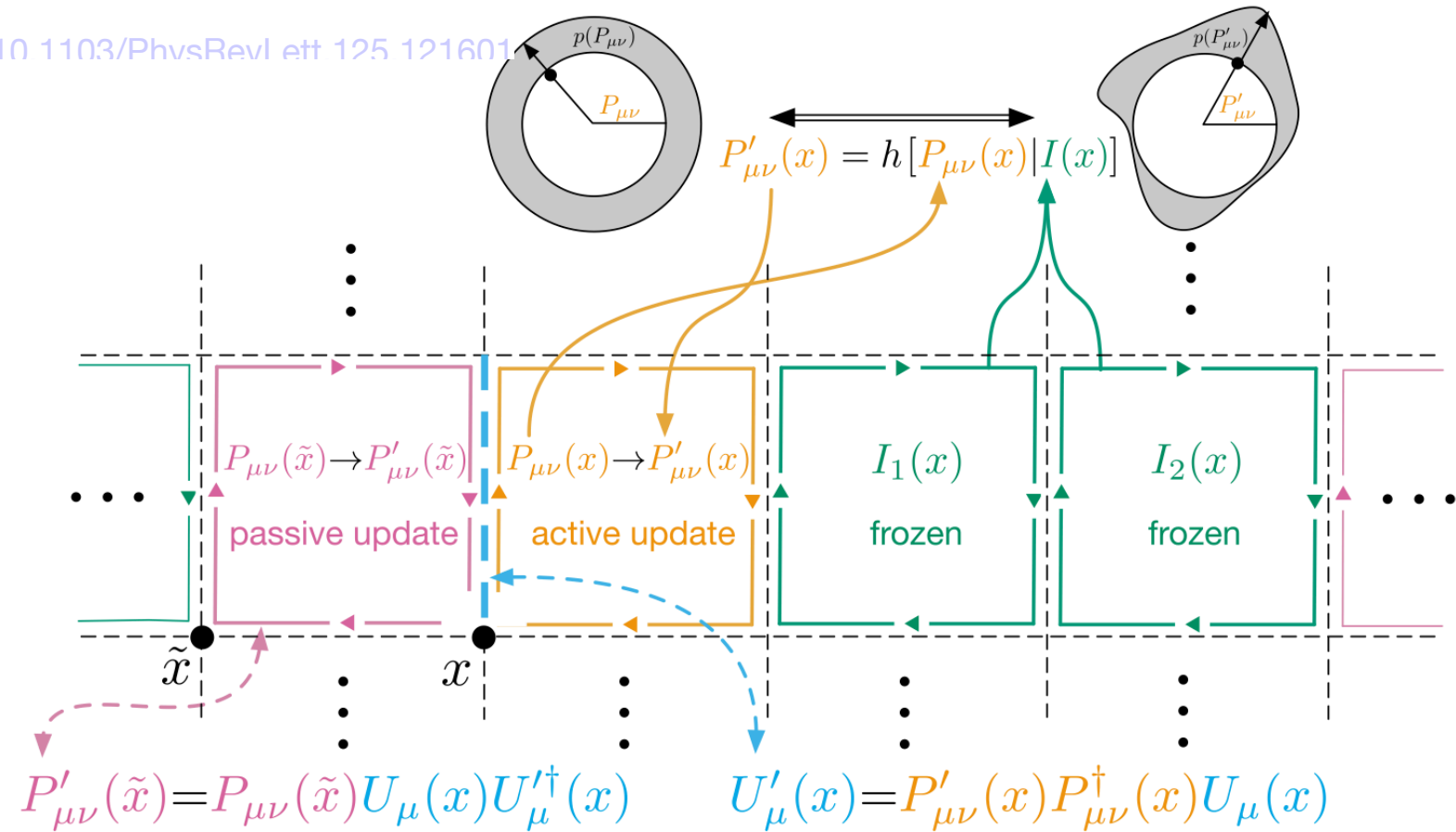


Parameters of lattice action

Few real numbers

Slide Phiala Shanahan

10.1103/PhysRevLett.125.121601



Lattice QCD is computationally expensive.



# For example

## Conclusions

- Significant progress.
- Issues of signal (statistics),  $N\pi$  contribution, and renormalization remain
- Gradient flow scheme is, so far, best for renormalization
- **quark-EDM**: Lattice QCD has provided results with  $\lesssim 5\%$  uncertainty
- **$\Theta$ -term**: Significant Progress. No reliable estimates yet
  - Statistics
  - Does  $N\pi$  provide leading excited-state contamination?
- **quark chromo-EDM**: Signal in the 3 methods being used
  - Renormalization and mixing (Working on gradient flow scheme)
  - Does  $N\pi$  provide leading excited-state contamination?
- **Weinberg  $G\tilde{G}G$  Operator** has signal
  - Address the mixing with  $\Theta$ -term in gradient flow scheme
- **Four-quark operators**: No calculations yet

Need 10-100 X Larger Computational Resources

Contributions of CP violating interactions to the neutron EDM from lattice QCD [[Rajan Gupta](#)]



*Take home message :*

*Particle Physics offers a unique playground  
for machine learning.*

*Rapid growth of deep learning applications  
in HEP.*

*Proofs of concept increasingly turning into  
production.*







# Outlooks

*what we could expect of future developments ...*



# Step Back on Artificial Intelligence

Developing AI requires an introspection on how intelligence is acquired, how new knowledge is created, is creation/innovation an extrapolation from the world of knowledge, is individual intelligence only based on experience and observation, ...  $\Rightarrow$  philosophical considerations.

Can this process be automatised, how much bias is there in AI-acquired knowledge feeding back the world of knowledge, would an AI exhibit the same bad traits of humanity, ...  $\Rightarrow$  ethical considerations.

How would an AI that develop a theory, makes a thought experiment, an experiment, an observation and confirm the theory look like. Would an AI develop new technologies and send AI to the moon. Would an AI devise a way to share resources, ...  $\Rightarrow$  behavioural considerations.

Not delving into these considerations here, but they are key elements to the future of AI.



# AI Scientist

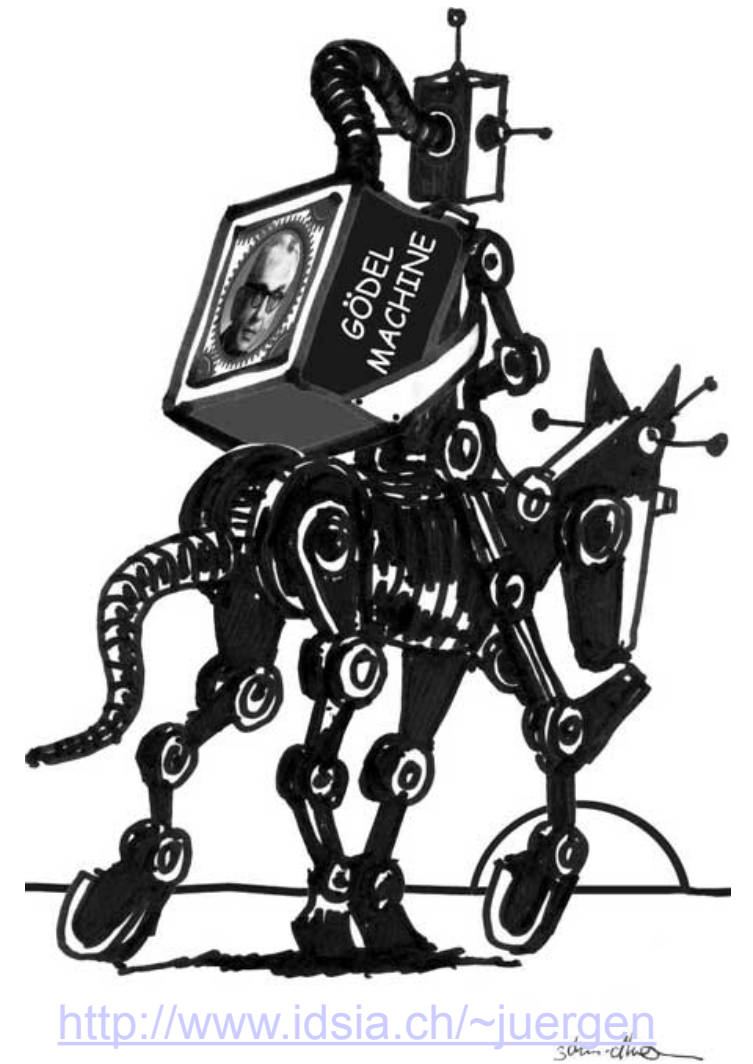
“Use DL” is not a good objective for applying deep learning. One needs to have a concrete objective function and data/simulator that contains enough information to realise the objective.

Trivial for simple objectives, might not be so trivial in more complex situations.

Deep learning is relevant when dealing with large sample datasets, complex simulation or data extraction algorithms, ...

Some data analysis method might not be considered because too computational expensive ; however deep learning might make them tractable.

A scientist should be efficient in all the tasks towards advancing knowledge ; “automate all the things”.

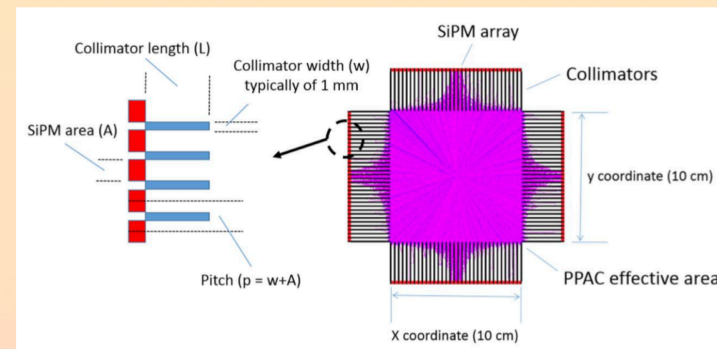
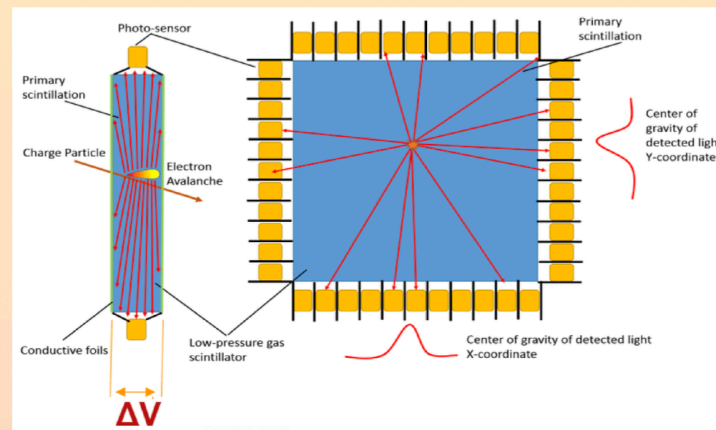


# Experimental Design

P.Vischia at ICHEP24

## Neutron Tomography

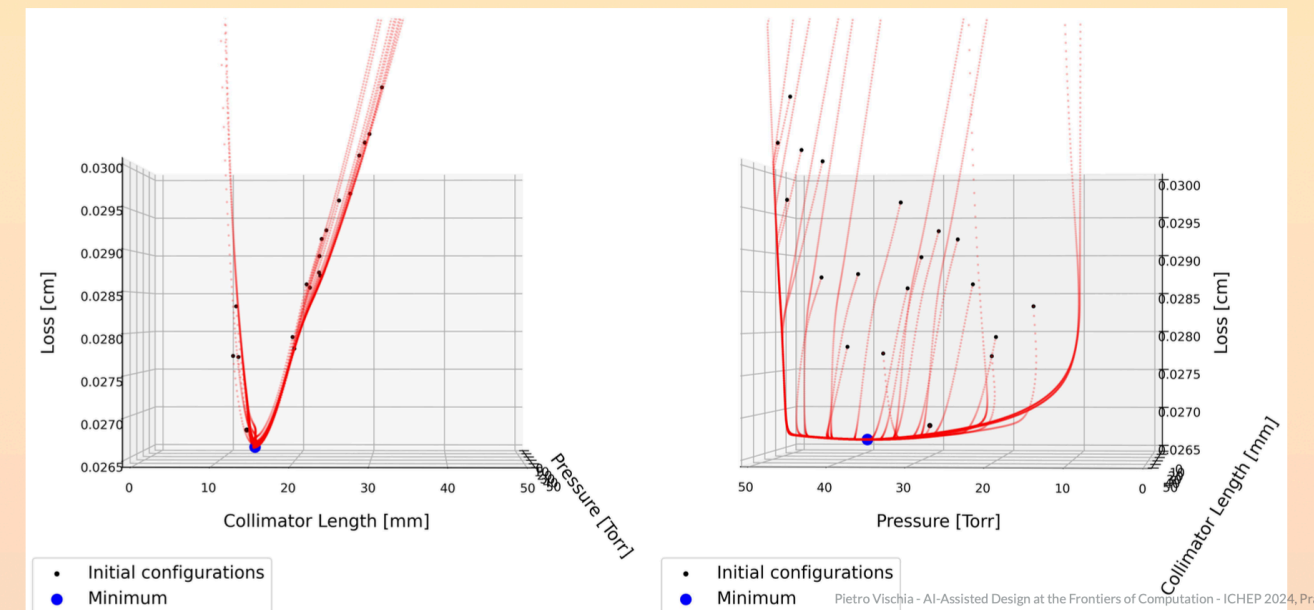
- GEANT4 model of a  $10 \times 10\text{cm}^2$  O-PPAC (Parallel-Plate Avalanche Counter with Optical Readout) from *Neutron Insights*
  - Parallel electrodes with  $3\text{mm}$  gap filled with low-pressure scintillating gas mixture (CF<sub>4</sub>) with high electroluminescent light yield
  - Readout via array of small silicon photomultipliers SiPMs
- Parametric neural network surrogate of the GEANT4 simulation
  - $p$ : higher pressure  $\rightarrow$  higher electroluminescence yield (up to a threshold), but larger voltage (energy expenditure)
  - $L$ : collimator length: tradeoff between accurate light localization (high  $L$ ) and higher photon statistics (low  $L$ )



<https://mode-collaboration.github.io/>

## Neutron Tomography: optimization

- Results for  $L$  give the same result as traditional studies from [10.1088/1748-0221/13/10/P10006](https://doi.org/10.1088/1748-0221/13/10/P10006)
- Remarkably stable **regardless of initial configuration**
- **MSc thesis of María Pereira Martínez**
  - Paper in preparation (w/ M.P.M., Xabier Cid Vidal)



Differentiable programming implementation of detector  $\Rightarrow$  measurement pipeline.

Optimisation of detector layout with gradient descent with respect to measurements.

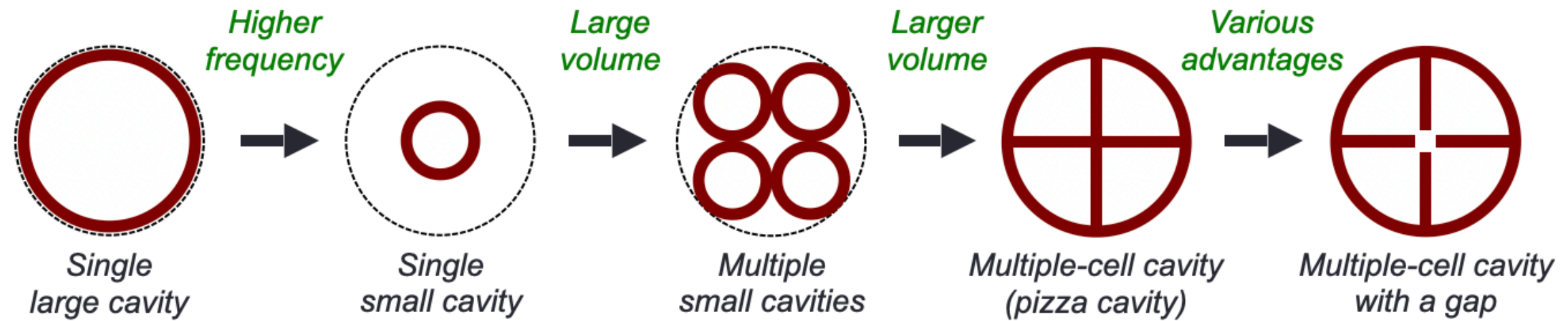
**More efficient experiment R&D.**



# For example



## Multi-cell (pizza)

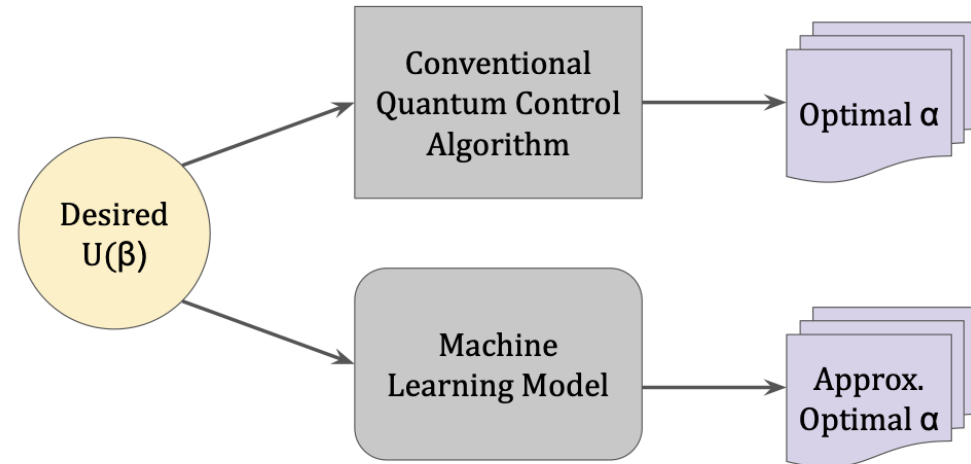


Cavity haloscopes for high-frequency axion search at IBS-CAPP [[Sungwoo Youn](#)]

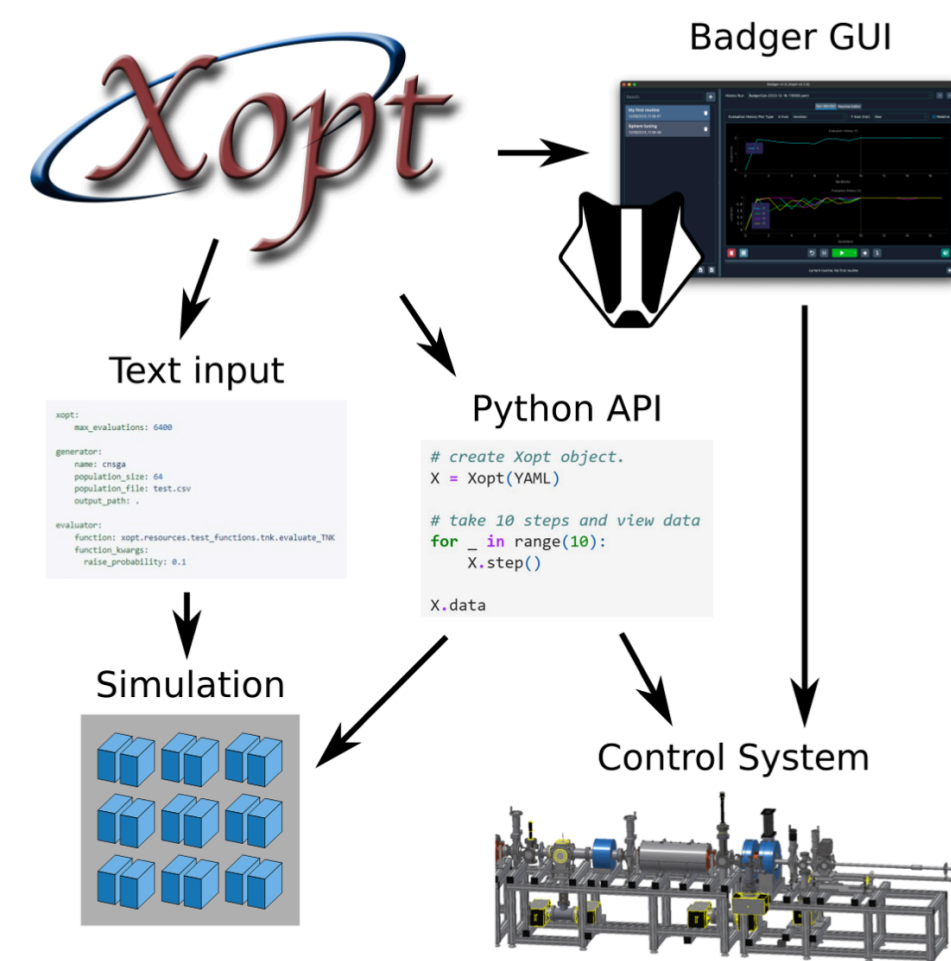


# Apparatus Control

... This algorithm is lightweight enough to fit on a low-resource FPGA and perform inference with a latency of 175ns and pipeline interval of 5 ns with > 0.99 gate fidelity. In the long term, such an accelerator could be used near quantum computing hardware where traditional computers cannot operate, enabling quantum control at a reasonable cost **at low latencies without incurring large data bandwidths outside of the cryogenic environment.**



Neural network accelerator for quantum control  
[10.1109/QCS56647.2022.00010](https://arxiv.org/abs/10.1109/QCS56647.2022.00010)



[10.18429/JACoW-IPAC2024-THPG85](https://arxiv.org/abs/10.18429/JACoW-IPAC2024-THPG85)

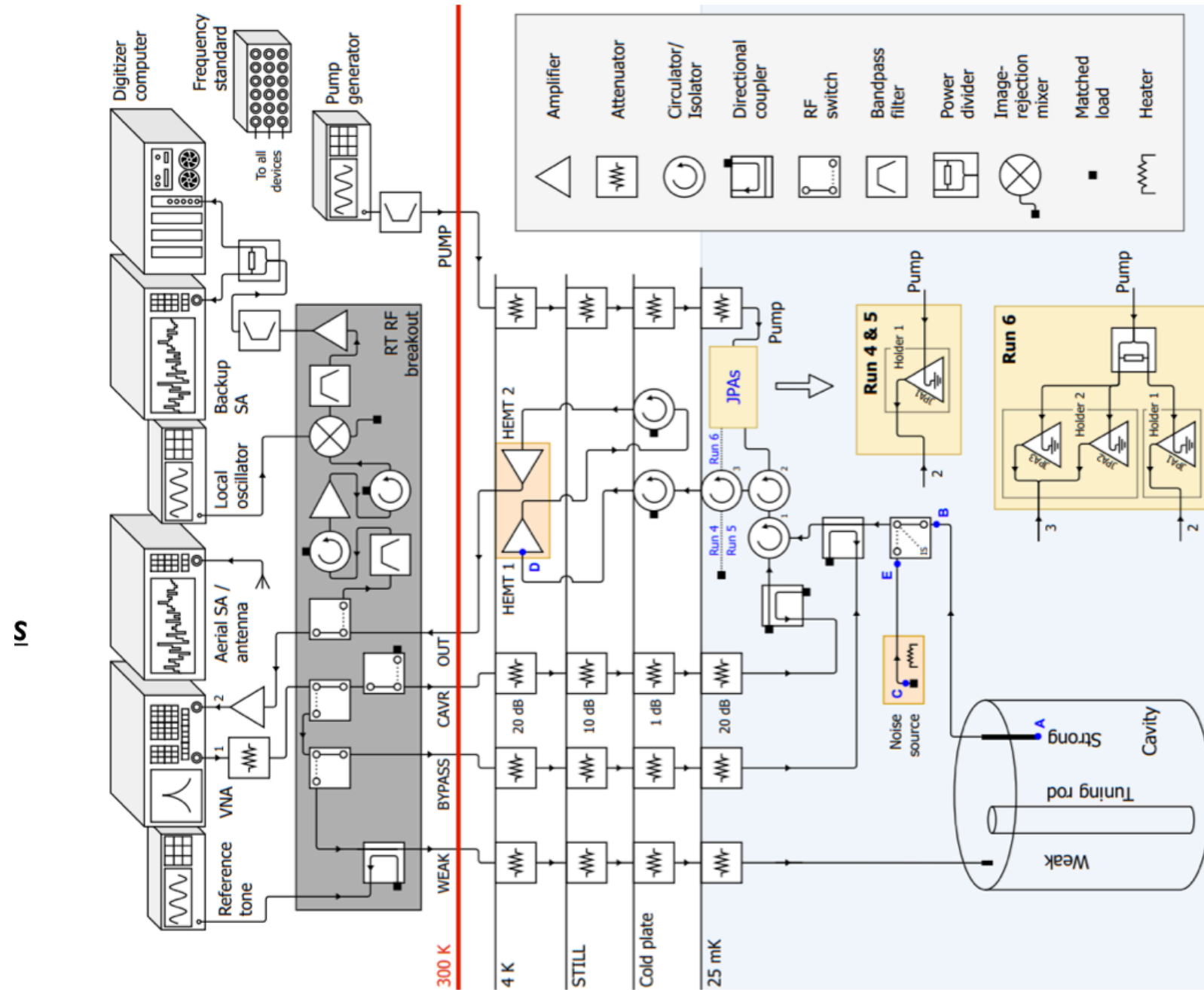
Progressing in achieving apparatus control : extend to accelerator, detector, DAQ, computing, networking, storage, ...

Potential for on-chip sensor control.

**Person, time and power efficient operation.**



# For example



The CAPP-MAX axion haloscope: status and prospects [[Saebyeok Ahn](#)]



# Analysis Description Languages



ChatGPT, write me a paper...

Hi ChatGPT. Below is an LHC physics analysis written in the Analysis Description Language syntax. Can you please use this to write the object selection and event selection sections of a scientific publication? Please keep in mind the following:

- Do not use lists, itemizations! Rather use paragraphs.
- Do not refer at all to "good" objects. Do NOT use the word "good" at all in the text. (..... give the ADL file from previous slide)

ChatGPT 4 is pretty good in writing ADL files from a given text.

Immediate use for education purposes.

[The Analysis Description Language Ecosystem](#)

ChatGPT 3.5

ChatGPT

Object selection in this analysis begins with the identification of muons satisfying a transverse momentum ( $p_T$ ) greater than 20 GeV and an absolute pseudorapidity ( $\eta$ ) less than 2.4. Similarly, electrons are chosen based on a  $p_T$  threshold greater than 20 GeV and  $|\eta|$  less than 2.5. A unified set of leptons is then formed by combining the selected electrons and muons.

Jets are selected with a  $p_T$  threshold of 30 GeV and  $|\eta|$  less than 2.4. To ensure separation from leptons, any jet within a distance  $dR < 0.4$  from a lepton is excluded.

Two key event variables are defined: the scalar sum of jet  $p_T$  (HT) and the transverse mass of the leading lepton and missing transverse energy (MET) (MTI). HT is calculated as the sum of the transverse momenta ( $p_T$ ) for all selected jets. MTI is computed as the square root of  $(2 \cdot p_T(\text{lepton}) \cdot MET \cdot (1 - \cos(\phi(MET) - \phi(\text{lepton}))))$ , where  $\phi$  represents the azimuthal angle.

Event selection is performed across three regions. The baseline region requires a minimum of 2 selected jets, a threshold on HT ( $HT > 200$ ), and a condition on the ratio of MET to HT ( $MET/HT \leq 1$ ). The signal region builds upon the baseline criteria but excludes events with any selected leptons and introduces an additional requirement on the azimuthal separation between MET and the leading jet ( $d\phi(MET, \text{jet}) > 0.5$ ). The control region, similar to the baseline, includes events with exactly one selected lepton and imposes a constraint on the transverse mass ( $MTI < 120$ ).

These selection criteria aim to isolate events with specific object characteristics and event topologies, enabling a detailed exploration of the physics processes under investigation.

Community-wide effort to define universal analysis description language.  
Still means writing a language and interpreter ...

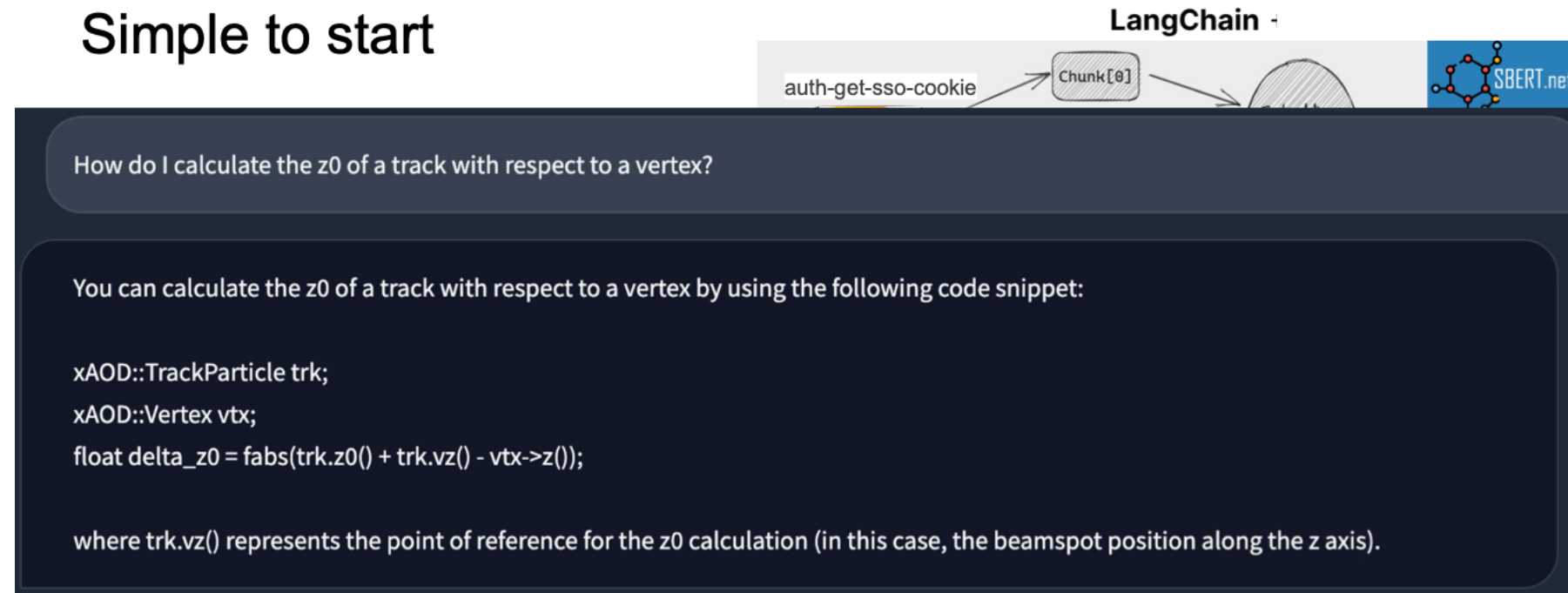
**Should GPT-like models be the Human  $\Rightarrow$  Code interpreter?**





# Code Generation

Simple to start



The screenshot shows a chat window with a question: "How do I calculate the z0 of a track with respect to a vertex?". The response includes a code snippet in C++:

```
xAOD::TrackParticle trk;  
xAOD::Vertex vtx;  
float delta_z0 = fabs(trk.z0() + trk.vz() - vtx->z());
```

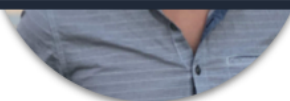
Below the code, it says: "where trk.vz() represents the point of reference for the z0 calculation (in this case, the beamspot position along the z axis)."

At the top of the chat window, there are labels: "LangChain", "auth-get-sso-cookie", "Chunk[0]", and "SBERT.net".



DM

chATLAS An AI Assistant for the ATLAS Collaboration



GF

\*Name credit: Sam Van Stroud  
Thanks to M. Hart, N. Pond, J. Barr, A. Vaitkus for the gentle ribbing  
IML Meeting, 9th April 2024

[ChATLAS: developing an AI assistant for the ATLAS collaboration](#)

In our field (and likely in others), writing code is a skill getting scarce. Learning a language is mostly done through “speaking and listening”

AI-generated code anyways needs reviewing and testing.

Conversing with an “AI writing code” is **actually a good way to teach people good coding skills and rules.**



# Documentation-bot

## AccGPT: Examples 3 – Coding Example

Could you give an example implementation of an optimization problem in COI?

What is the COI?

COI stands for "Common Optimization Interfaces". It is a project that aims to bring numerical optimization, machine learning, and reinforcement learning to the operation of the CERN accelerator complex. The COI provides common interfaces that make it possible to use numerical optimization and reinforcement learning on the same optimization problems. The `cernml-coi-utils` package provides utility functions and classes that make it easier to work with the COI, encapsulating common use cases to save time and prevent bugs.

Further readings:  
<https://acc-py.web.cern.ch/gitlab/geoff/cernml-coi-utils/docs/stable/index.html>

The example implementation of an optimization problem in COI is as follows:

```
python
import numpy as np
from scipy.optimize import minimize
from cernml import coi

class AwakeElectronBeamSteering(coi.SingleOptimizable):
    def __init__(self):
        self.corrector_scale = 1

    def compute_single_objective(self, params):
        rms = np.sqrt(np.mean(np.abs(self.corrector_scale * params)))
        return rms

    def compute_single_objective_function(self, params):
        return -np.mean(np.abs(self.corrector_scale * params))

    def get_initial_params(self, cycle_time):
        return np.zeros(self.corrector_scale.shape)

    def get_objective_function_name(self):
        return "ElectronBeamSteering"

coi.register("awake_electron_beam_steering", AwakeElectronBeamSteering)

# Run the optimization problem
coi.make("awake_electron_beam_steering")
coi.run()
```

For such tasks we aim to implement a dedicated coding LLM.

This implementation defines an optimization problem for electron beam steering in a particle accelerator. The `AwakeElectronBeamSteering` class inherits from `coi.SingleOptimizable` and overrides the `compute_single_objective`, `compute_single_objective_function`, `get_initial_params`, and `get_objective_function_name` methods. The `compute_single_objective` method calculates the root mean square (RMS) of the objective function. The `compute_single_objective_function` method calculates the objective function value. The `get_initial_params` method returns the initial parameters for the optimization. The `get_objective_function_name` method returns the name of the objective function.

Further reading:  
<https://acc-py.web.cern.ch/gitlab/geoff/cernml-coi/docs/stable/tutorials/implement-singleoptimizable.html>  
(Confidence: 69.0%)

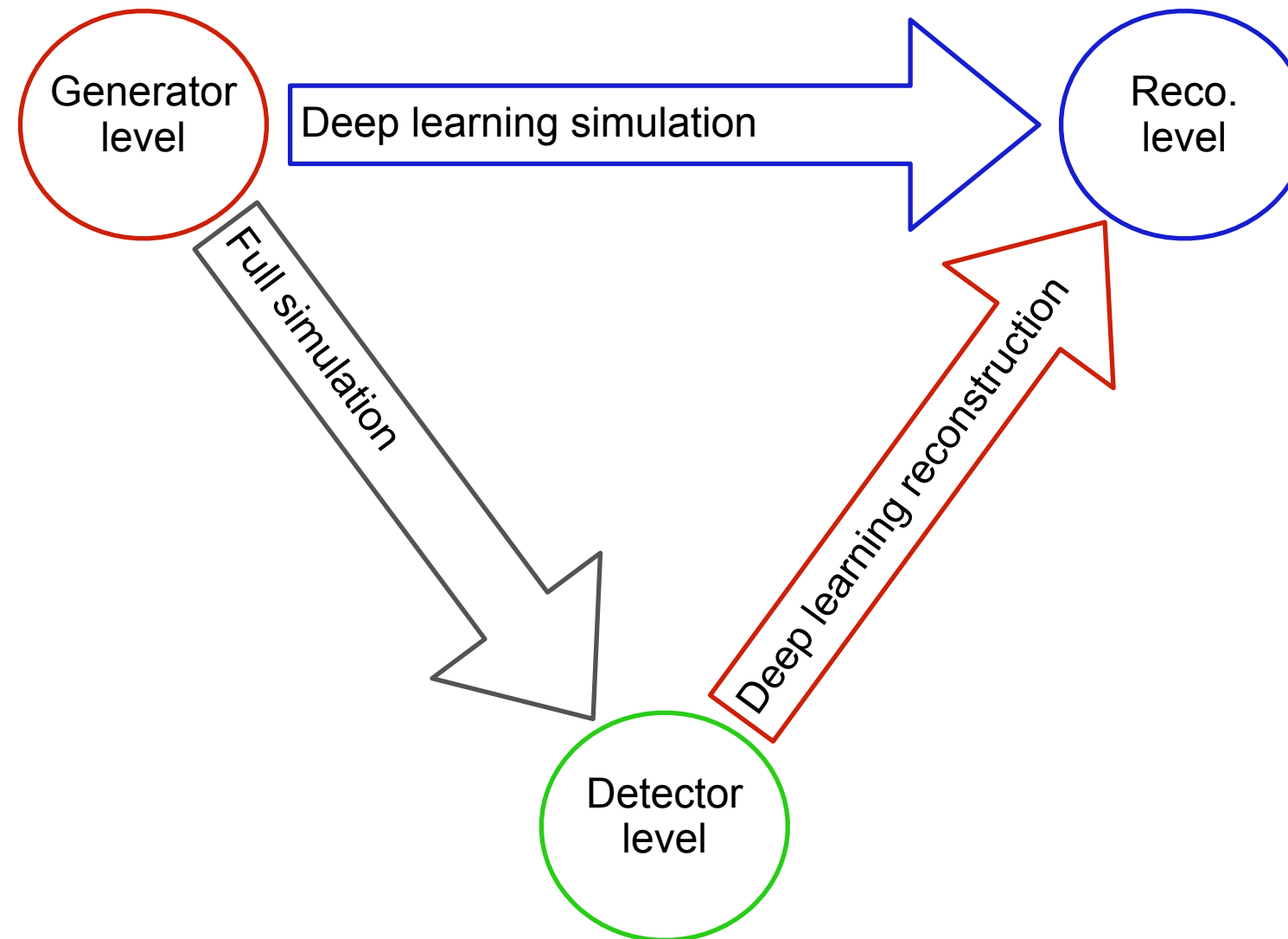


Natural to think about training one such model on experimental documentation. Multiple R&D projects underway. A community-wide effort seems to be forming on common issues.

**Can it be used to curate/create documentation ?**



# Closure of simulation/reconstruction



“All you need” is accurate detector simulation.

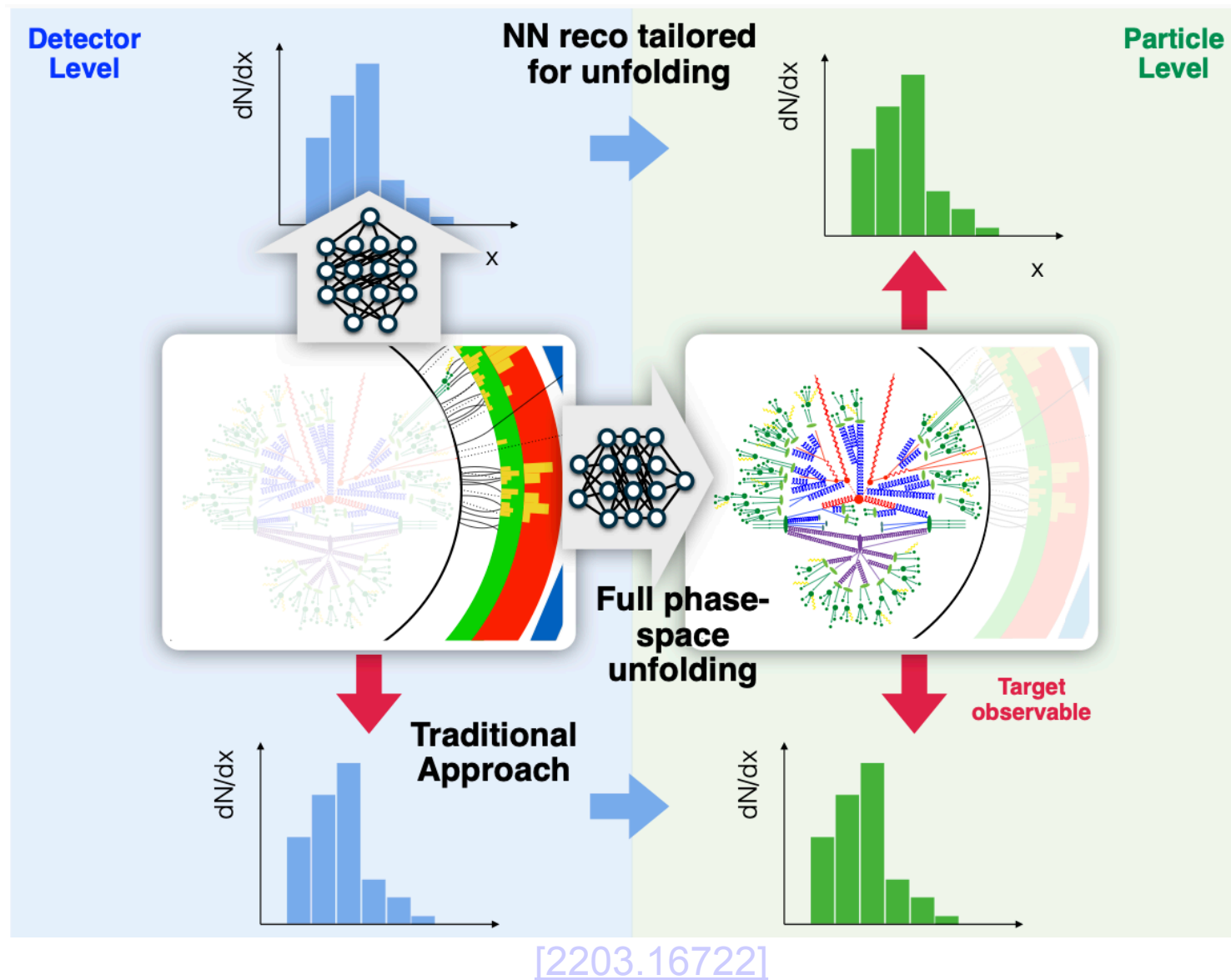
Deep learning reconstruction (generator as target) applied to data.

Deep learning simulation (reco. as target) for simulated samples.

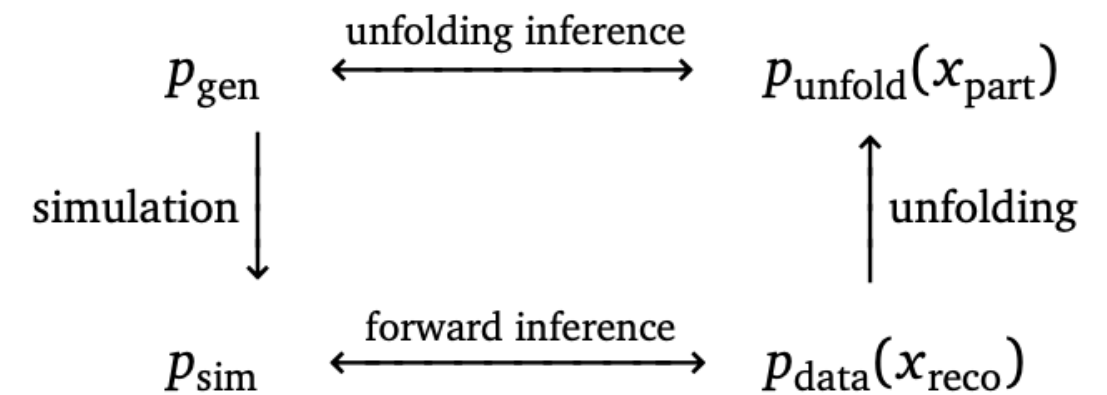
Invertible deep learning simulation can be used as unfolding.



# Simulation / Unfolding



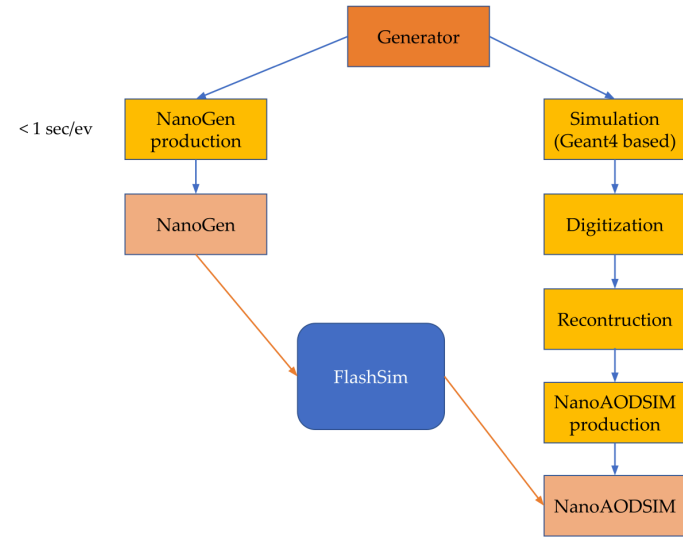
Invertible simulation models can perform unfolding.  
 Novel unfolding methods available.  
 Not a black-box design : **requires physicist's attention.**  
 Still a lot of work on R&D to be done.



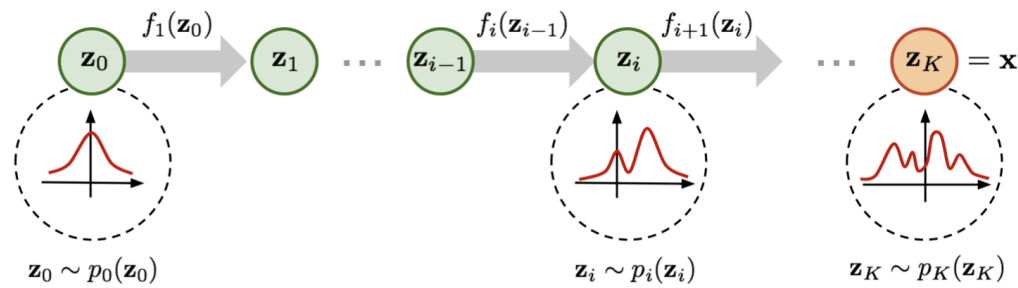
Overview of ML Unfolding [2404.18807]



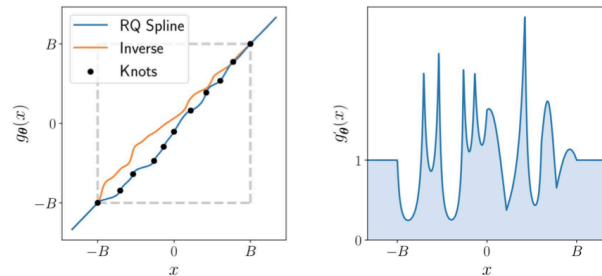
# Flash-Sim



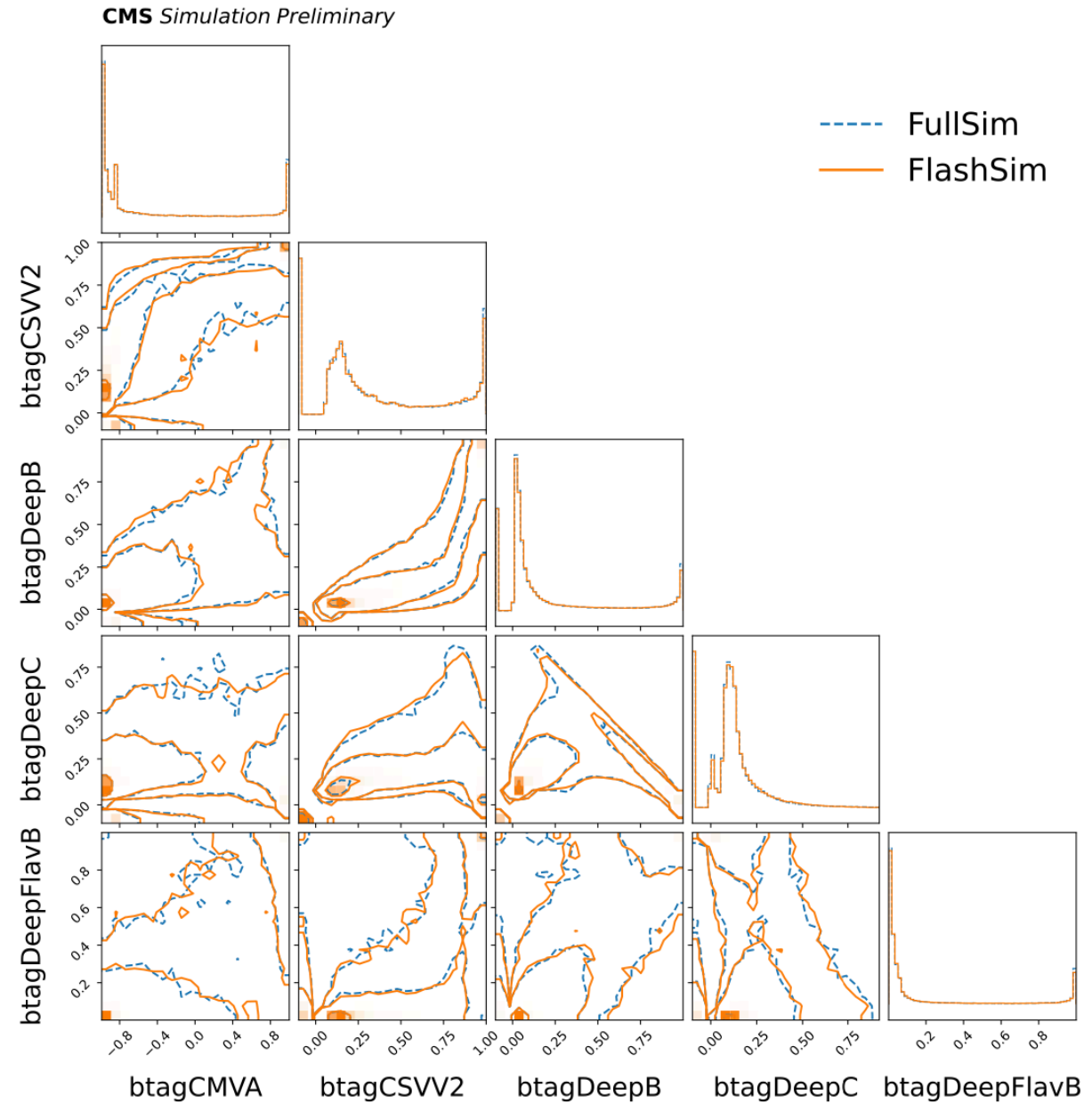
## NEURAL SPLINE FLOWS



$$p_X(x) = p_Z(z) \frac{dz}{dx} \rightarrow \log(p_X(x)) = \log(p_Z(f^{-1}(x))) - \log\left(\det\left|\frac{dx}{dz}\right|\right)$$



[2858890]



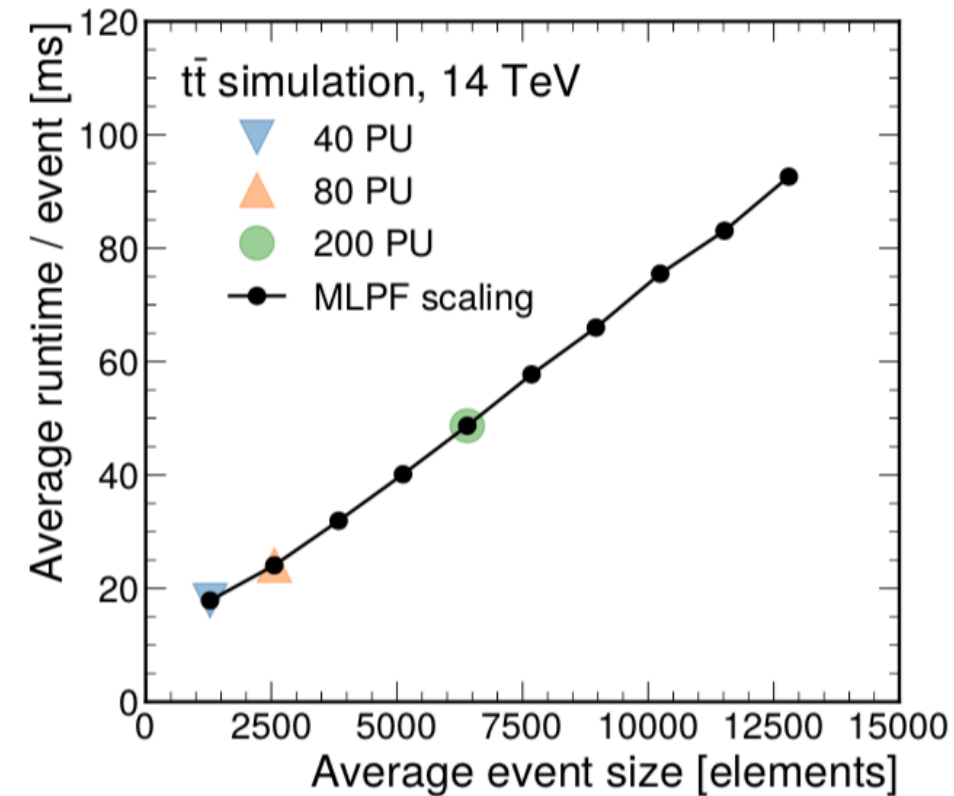
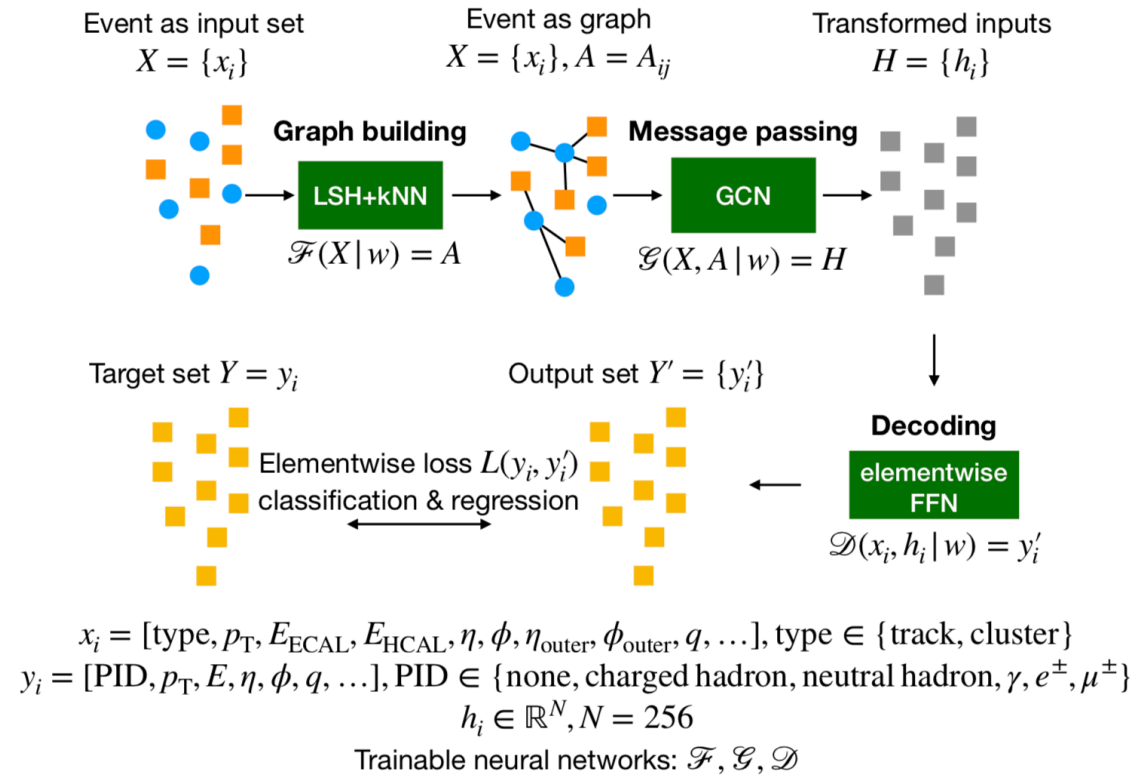
Predicting reconstructed NANO variables from generator-level quantities.  
Normalizing flow model, with neural spline yields astonishing performance.



# Reconstruction

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

[\[2101.08578\]](#)



Building reconstruction software is a lengthy and costly project.

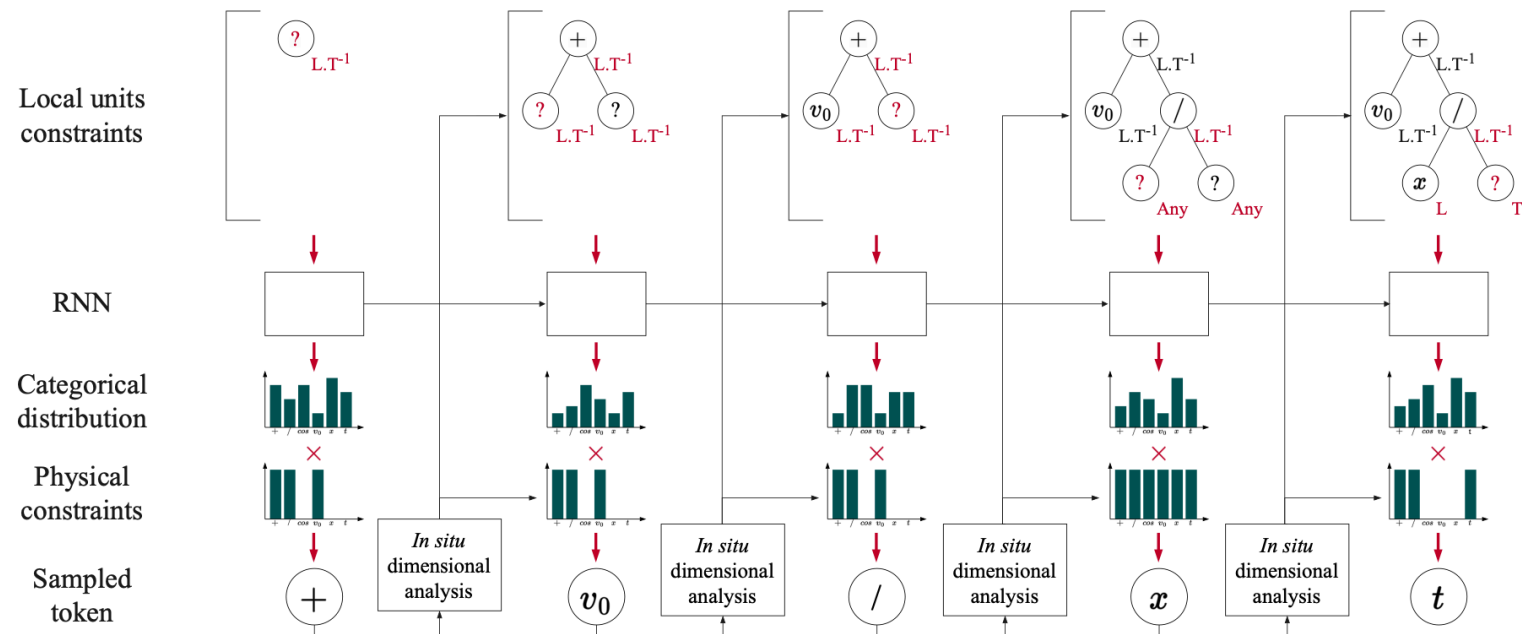
Surrogates learnable from existing software. End-to-end ML-reconstruction learnable from full simulation.

May yield better algorithmic complexity.

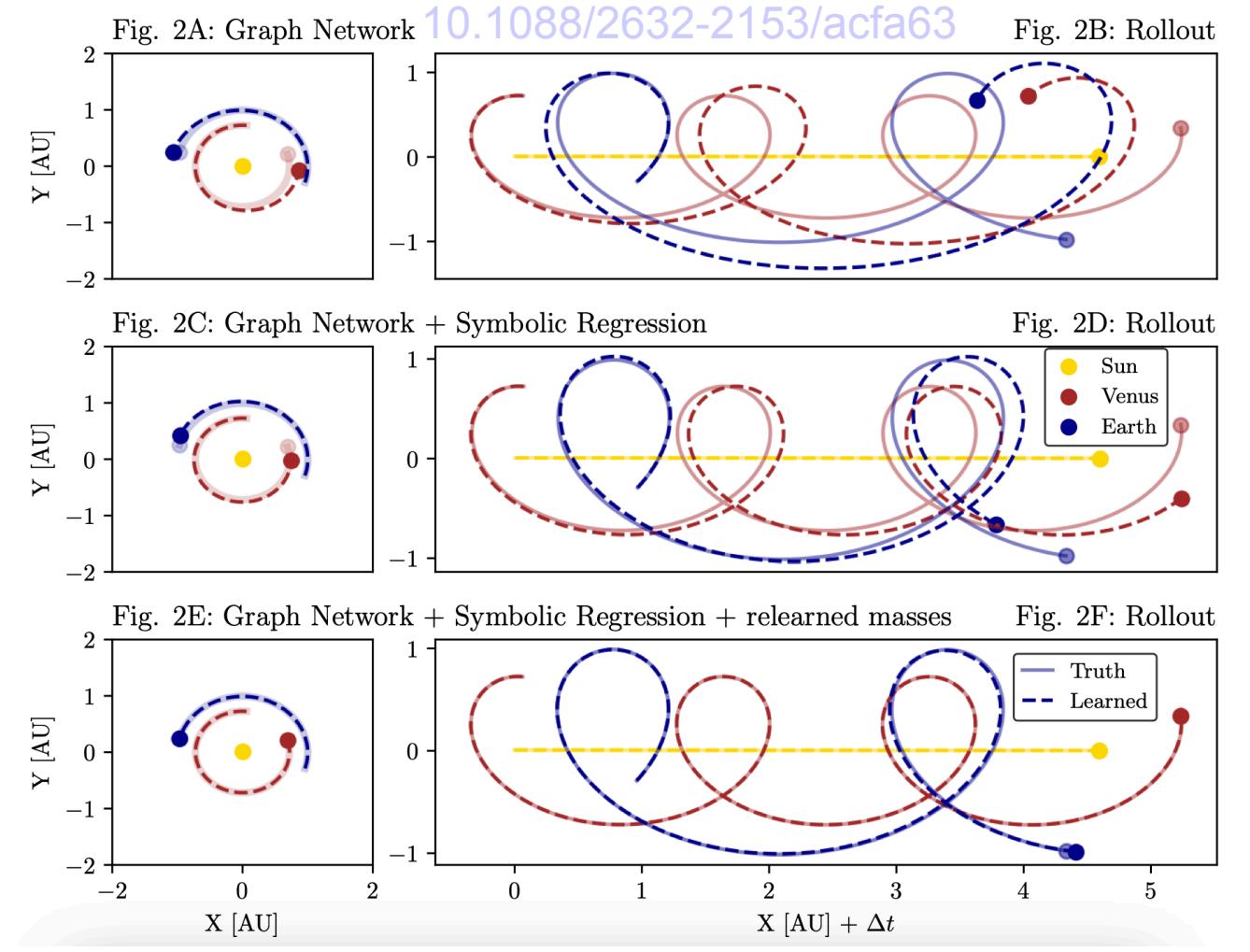
Still a lot of work on R&D to be done.



# Symbolic Learning



[NeurIPS ML4PS 2023 89.pdf](#)



Symbolic learning is still evolving.

Can be used to learn formulas, or as a mean of interpretability.

Complementary to physics-inspired models to learn effective Lagrangian, forces, interactions, ...



# Sustainability

Deep learning models are notoriously computational intensive to train : days, weeks, months ... on small to large computing facilities.

Computational cost of training is however largely offset by the gain in inference over classical algorithms, year-long over world-wide-resources.

Through AI-assisted operation, resource cost reduction could extend to accelerators, storage, networking, computing facilities, ...





*Take home message :*

*AI-assisted experiment design, data processing and apparatus control shall be more efficient.*

*Carefully crafted physics inspired model shall “learn all the things”.*



# Summary

- ➔ Machine learning has become extremely performant.
- ➔ Deep learning offers great prospects for Science.
- ➔ Several ways to improve on resource usage.
- ➔ A lot of R&D done since [DS@LHC15](#).
- ➔ Novel techniques entering experimental workflows.
- ➔ The future of science will be deep learning, always with a “Physicists in the loop”.

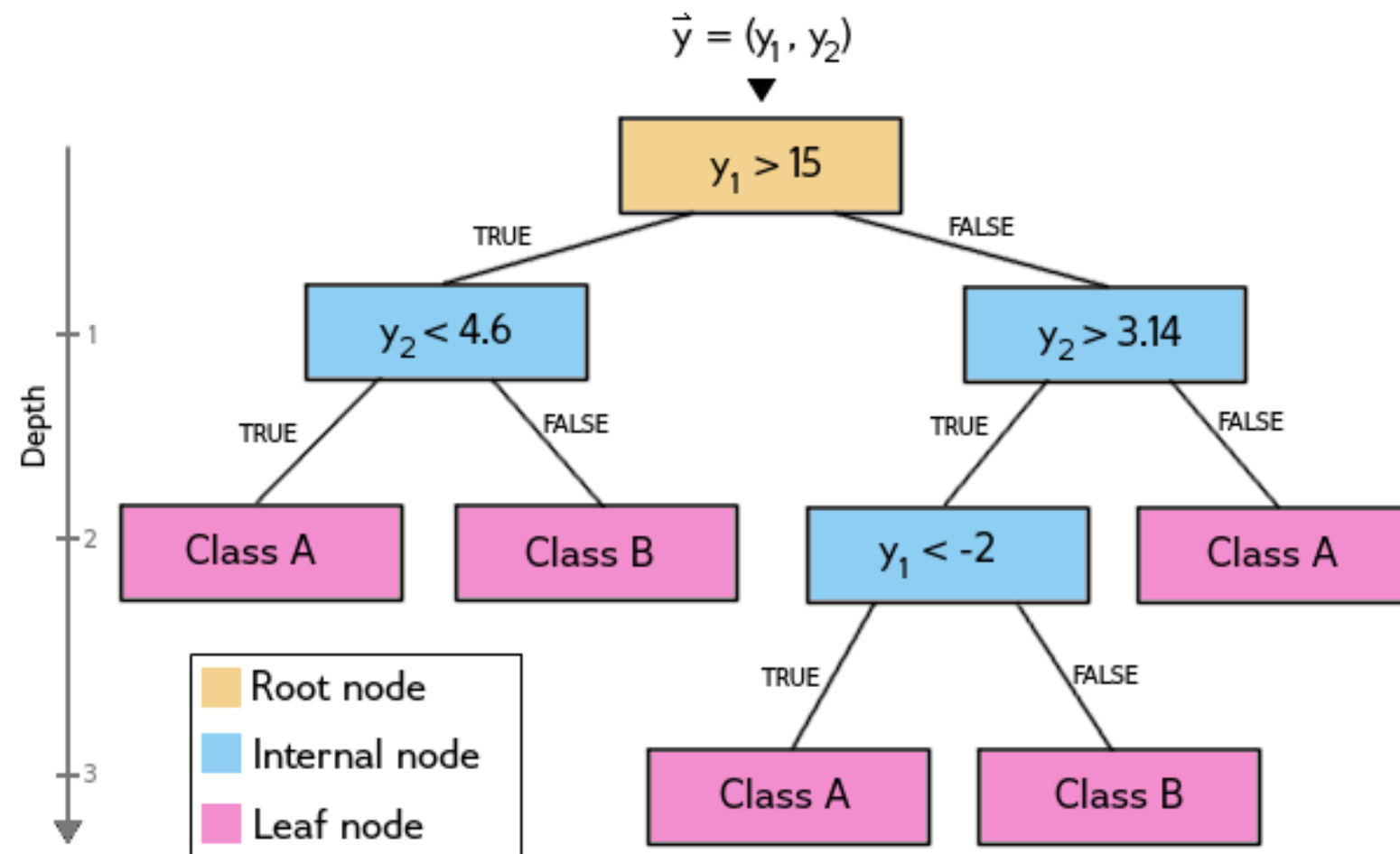
*This work is partially supported by the U.S. DOE, Office of Science, Office of High Energy Physics under Award No. DE-SC0011925, DE-SC0019219, and DE-AC02-07CH11359.*



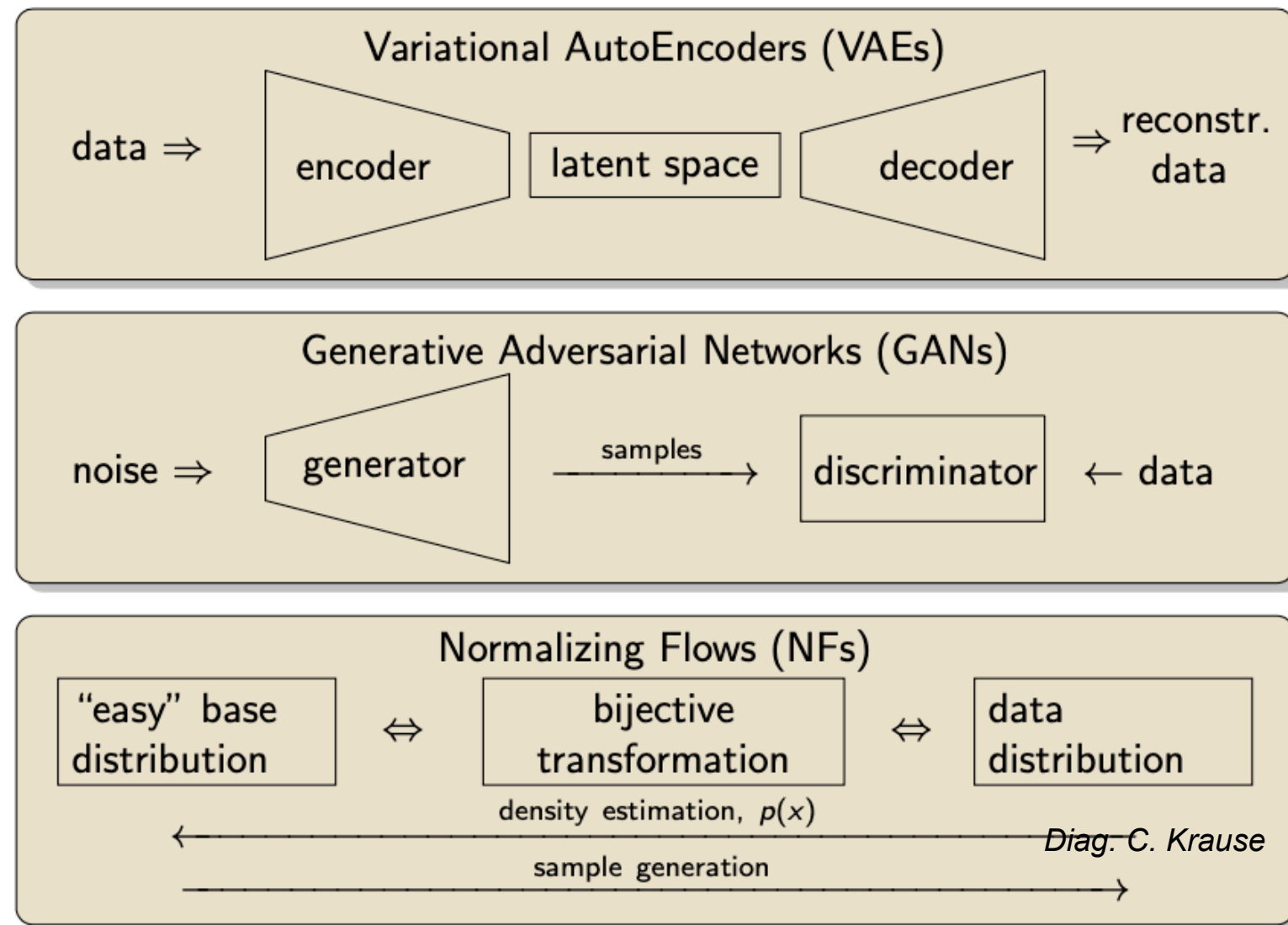


# Decision Tree

Decision trees is a well known tool in supervised learning.  
It has the advantage of being easily interpretable  
Can be used for classification or regression



# Diversity of Generative Models



Three main classes of generative models.



# The Mysteries of Deep Learning

Highly non-convex degenerate objective function optimisation with gradient descent is not guaranteed, and yet ...

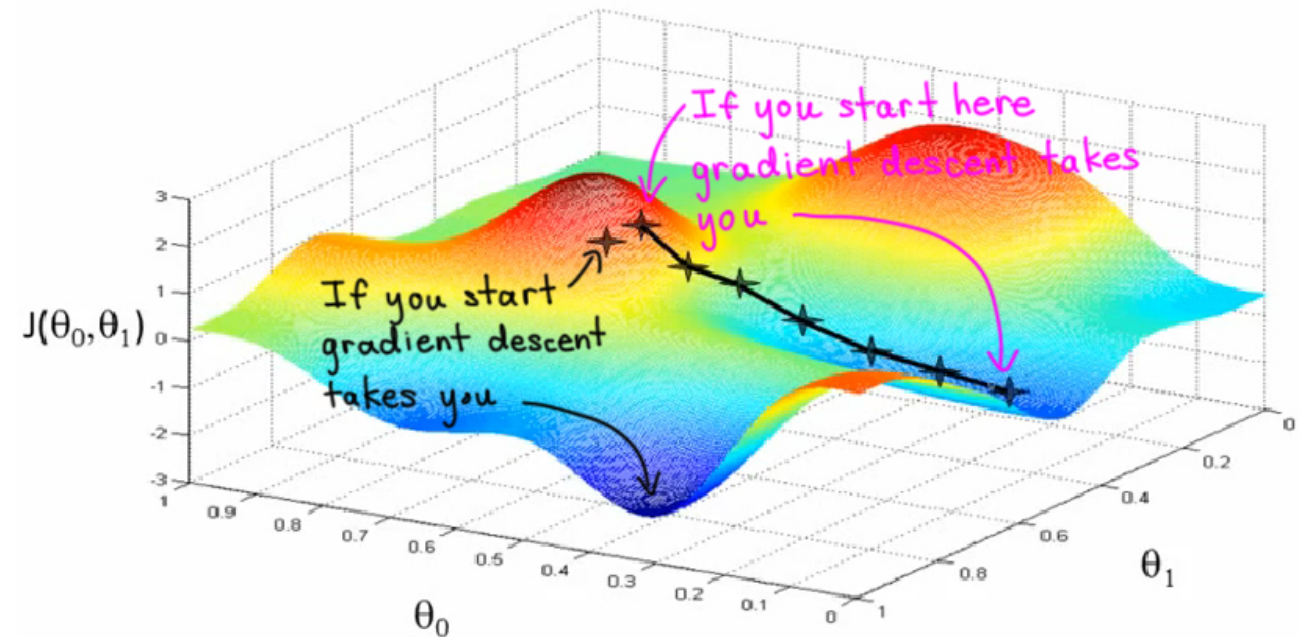
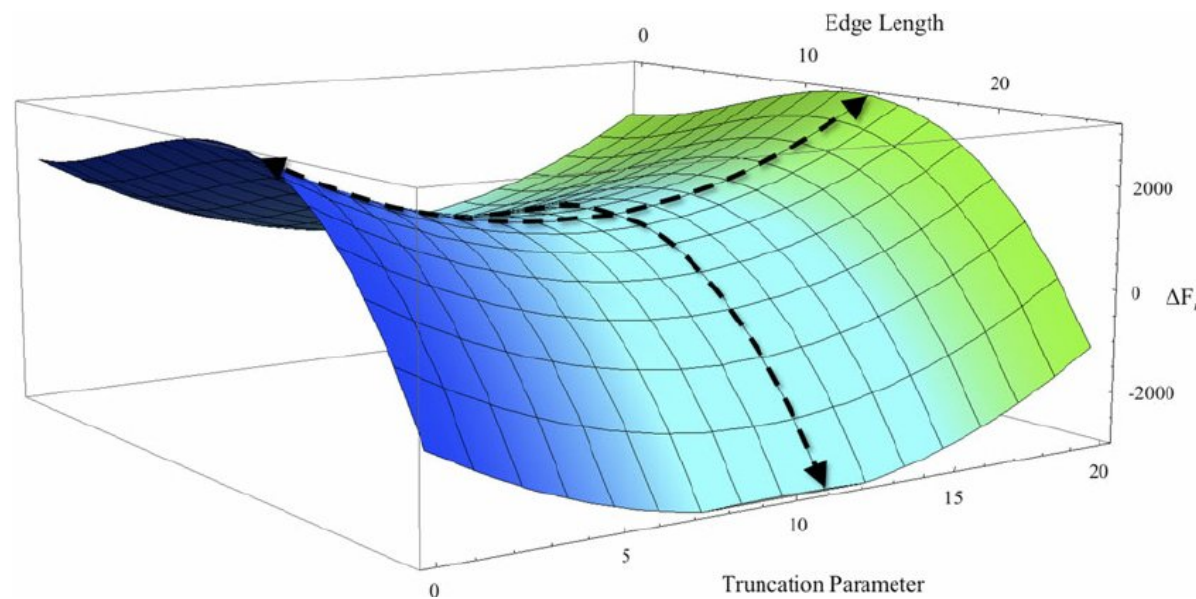
A complex “teacher” can train a simpler “student” that cannot learn from the data directly ...

Models are over-parametrised and yet manage to generalise ...

**Still quite empirical field**, although theoretical progress are made



# Non-Convex Optimisation



The objective functions optimised in machine learning are usually non-convex

Non guaranteed convergence of gradient descent  
Gradients may vanish near local optimum and saddle point



# Quantum Machine Learning

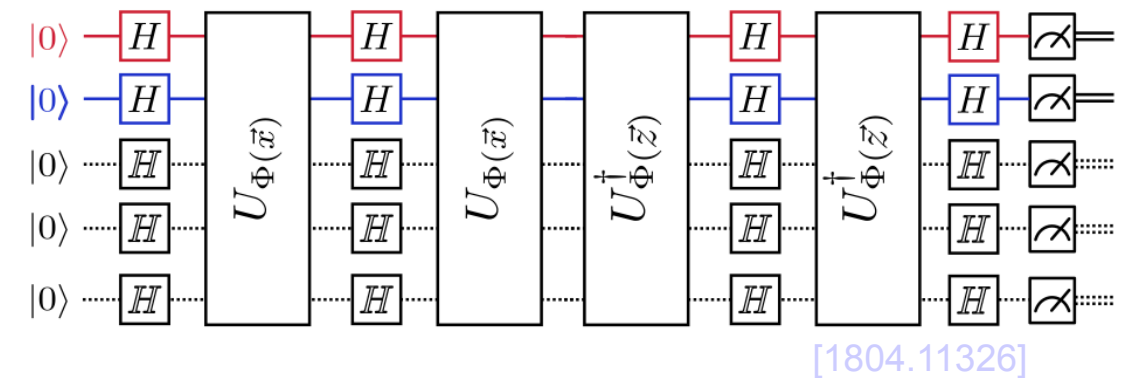
Deep learning is computing intensive, and de-facto enabled by use of GPU. People are looking for ways to leverage possible quantum advantage to accelerate machine learning techniques.

Main algorithms used in recent studies

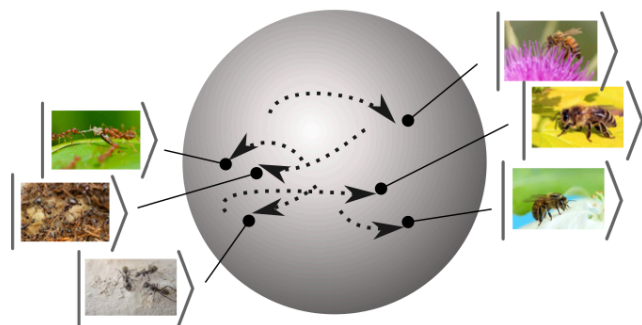
- ➔ Variational Quantum Circuits (VQC)
- ➔ Quantum Support Vector Machine (QSVM)
- ➔ Quantum Restricted Boltzman Machine (QRBM)
- ➔ Quantum Adiabatic Machine Learning (QAML)
- ➔ Quantum Generative Adversarial Network (QGAN)
- ➔ ...

Field in constant evolution. Embedding is crucial.  
Deep implications of kernel methods.

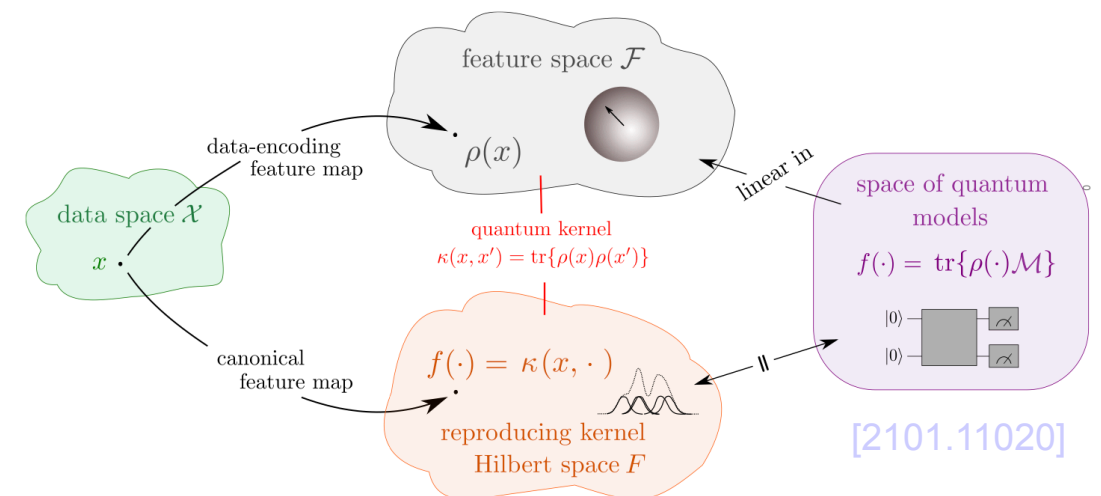
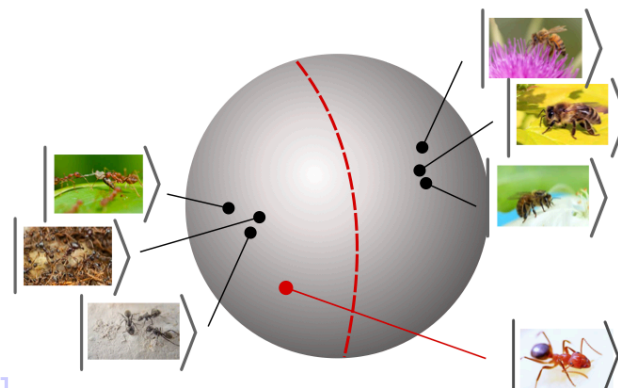
Software and toolkit available [pennylane](#) , [tf-quantum](#)



a. Training the embedding



b. Classification





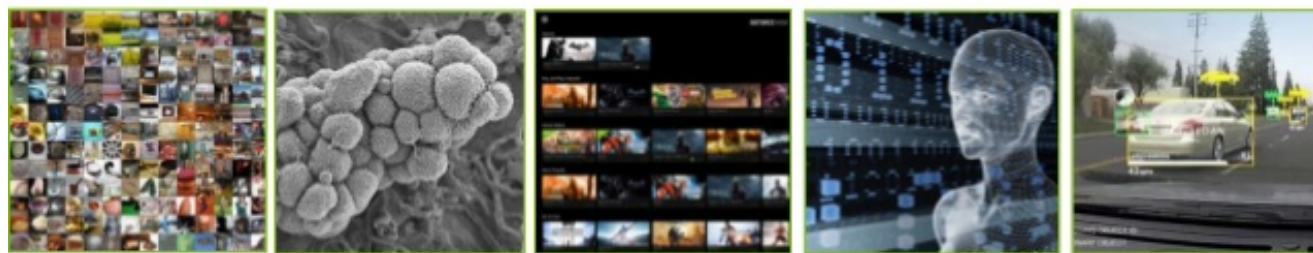
# Motivations for Using Machine Learning in Particle/Astro Physics

*and elsewhere ...*



# Machine Learning in Industry

## Deep Learning Everywhere

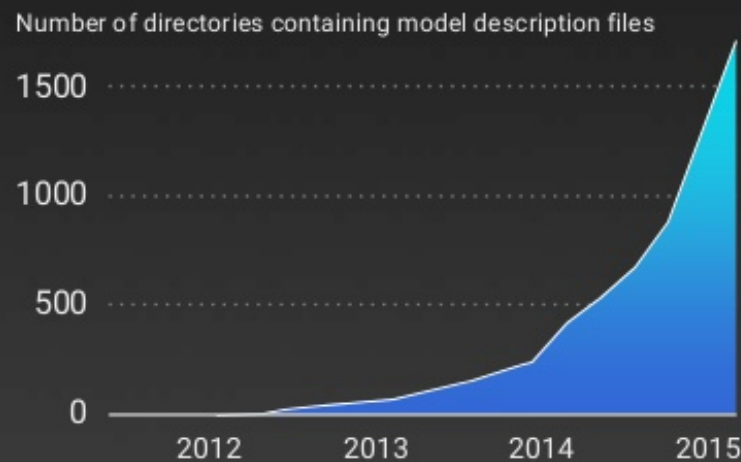


- | INTERNET & CLOUD  | MEDICINE & BIOLOGY  | MEDIA & ENTERTAINMENT   | SECURITY & DEFENSE  | AUTONOMOUS MACHINES   |
|---|---|---|---|---|
| <ul style="list-style-type: none"> <li>Image Classification</li> <li>Speech Recognition</li> <li>Language Translation</li> <li>Language Processing</li> <li>Sentiment Analysis</li> <li>Recommendation</li> </ul> | <ul style="list-style-type: none"> <li>Cancer Cell Detection</li> <li>Diabetic Grading</li> <li>Drug Discovery</li> </ul> | <ul style="list-style-type: none"> <li>Video Captioning</li> <li>Video Search</li> <li>Real Time Translation</li> </ul> | <ul style="list-style-type: none"> <li>Face Detection</li> <li>Video Surveillance</li> <li>Satellite Imagery</li> </ul> | <ul style="list-style-type: none"> <li>Pedestrian Detection</li> <li>Lane Tracking</li> <li>Recognize Traffic Sign</li> </ul> |

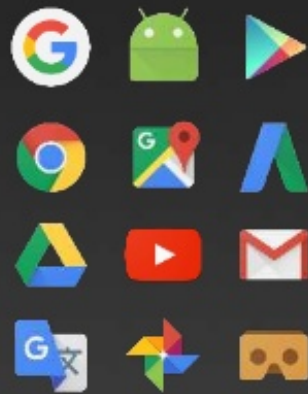
<https://www.nvidia.com/en-us/deep-learning-ai/>

15 NVIDIA

## Rapidly Accelerating Use of Deep Learning at Google



Used across products:



Google Cloud

**MACHINE INTELLIGENCE 3.0**

**ENTERPRISE INTELLIGENCE**

- VISUAL:** Orbital Insight, Clarifai, Cortica, SPACE KNOW, Netra, Planet, Deepvision, Goolson, Deepmatic
- AUDIO:** Gridspace, Clarifai, Nexidia, CAPIQ, Clover, CuriousAI, TalkIQ, Twilio, Expect Labs, Mobvoi, PopPP archive
- SENSOR:** PREDIX, Sentient, UPTAKE, thingworx, GIGOT, MAANA, PLANET OS, IMUBIT, KONUX Alluvium
- INTERNAL DATA:** PRIMER, Dyrpar, Alation, Digital Reasoning, IBM WATSON, Palantir ARIMO, Sapho, Outlier
- MARKET:** mattermark, Quid, Datafox, Botikbase, enigma, Tractix, PREMISE, MOTIVA, CB Insights, predata

**ENTERPRISE FUNCTIONS**

- CUSTOMER SUPPORT:** DigitalGenius, Eloquent, ACTIONIQ, Kasisto, Wiseio, Zendesk, Precis, CLARABRIDGE
- SALES:** Collective, Fuse, salesforce, Zensight, Inside Sales, Clari
- MARKETING:** MINTIGO, LiftIgniter, brightfunnel, COGNICOR, Lattice, RADIUS, PERSADO, retention, AIRPR, magid
- SECURITY:** CYCLANCE, ZIMPERIUM, graphistry, SignalSense, DARKTRACE, depinstinct, drawbridge, AppZen
- RECRUITING:** textio, Wade & Wendy, univie, SpringRole, GIGSTER, HiresVue

**AUTONOMOUS SYSTEMS**

- GROUND NAVIGATION:** drive.ai, ZOOX, UBER, Autonomy, AdastWorks, Google, TESLA, Auto Robotics
- AERIAL:** SKYDIO, Airware, DroneDeploy, SHIELD AI, LILY, SKYCATCH
- INDUSTRIAL:** JAYBRIDGE, CLEARPATH, HARVEST, OSARO, fetch, KINRED, rethink robotics

**AGENTS**

- PERSONAL:** amazon alexa, Cortana, facebook, Siri, Replika
- PROFESSIONAL:** butter.ai, @ clara, talla, pogo, x.ai, Zoom, SKIPFLAG, slack, sudo

**INDUSTRIES**

- AGRICULTURE:** BLUE RIVER, tule, Terrestrial, Bioscience Labs, mavrx, TRACE, Agrifly, udie, Pivot, Agri-Data, Durodot
- EDUCATION:** KNEWTON, CTI, UDACITY, gradescope, courseera, all school
- INVESTMENT:** Bloomberg, ISENTIUM, alpha sense, CEREBELLUM CAPITAL, Sentient, KENSHIC, Dataminr, Quandl
- LEGAL:** blue J, Everlaw, Seal, LEGAL ROBOT, BEAGLE, RAVEL, ROSS
- LOGISTICS:** NAUTO, PRETECKT, Routific, MARBLE, Acerta, clearmetal, PITSTOP

**INDUSTRIES CONT'D**

- MATERIALS:** zymergen, Eigen Innovations, SIGHT MACHINE, BINKO BIOWORKS, Citrine, nanotronics, CALCULARIO
- RETAIL FINANCE:** TALA, Lendo, Affirm, wealthfront, TALA, earnest, MIRADOR, Bettermint
- PATIENT:** PULSE, ZEPHYRUS HEALTH, Onco, Atomwise, CareScore, Watson Health, BENTRIAN, Numerate
- HEALTHCARE:** BUTTERFLY, ARTERYS, BAYLABS, Google DeepMind, 3SCAN, enlitic, imagga
- BIOLOGICAL:** CarbonX, deep genomics, LUMINIST, Atomwise, color, GRAIL, RECURSION, verily, GRAPE

**TECHNOLOGY STACK**

- AGENT ENABLERS:** OCTANE.AI, OpenAI Gym, semantic machines, howdy, MalubA, KITT.AI, Kasisto, AUTOMAT
- DATA SCIENCE:** DOMINO, kaggle, dataiku, seldn, SPARKBEYOND, DataRobot, yhat, AYASDI, rapidminer, enigma, CB Insights, bigml
- MACHINE LEARNING:** CognitiveScale, Dyrpar, HyperScience, SCALED INFERENCE, deepsense.io, reactive, deeplogic, GoogleML, context, relevant, minds.ai, H2O.ai, sparkcognition, loop, GEOMETRIC INTELLIGENCE, skymind, bonsai
- NATURAL LANGUAGE:** agolo, Narrative Science, Scaled Intelligence, cortical.io, MONKEY LEARN, LEXALYTICS, spaCy, LUMINOSO
- DEVELOPMENT:** SIGOPT, HyperOpt, rainforest, Signifai, LAYER 6, fuzzyio, pkite, Anodot, bonsai
- DATA CAPTURE:** CrowdFlower, Paxata, WorkFusion, diffbot, amazon, mechanical turk, enigma, DATASIFT, TRIFACTA, parsehub
- OPEN SOURCE LIBRARIES:** Keras, H2O, DSSTNE, MXNet, Chainer, CNTK, DEEPLARNING4J, Scikit-learn, DMTK, TensorFlow, theano, Scikit-learn, torch, Spark, PaddlePaddle, WEKA, AzureML, neon
- HARDWARE:** KNUPATH, NVIDIA, tersilica, GoogleTPU, Cerebras, TENSTORRENT, intel, nvidia, Cirrascale, nervana, Movidius, 10<sup>th</sup> Labs, qualcomm, Isosemi
- RESEARCH:** OpenAI, numasense, Numenta, Kimera Systems, Cogital, ELEMENT, vicarious

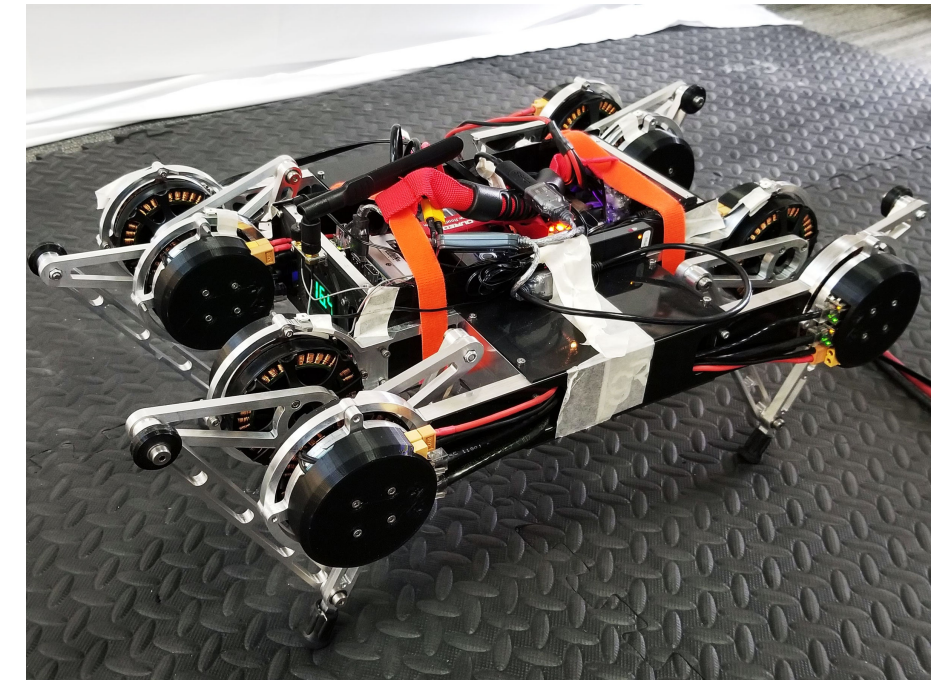
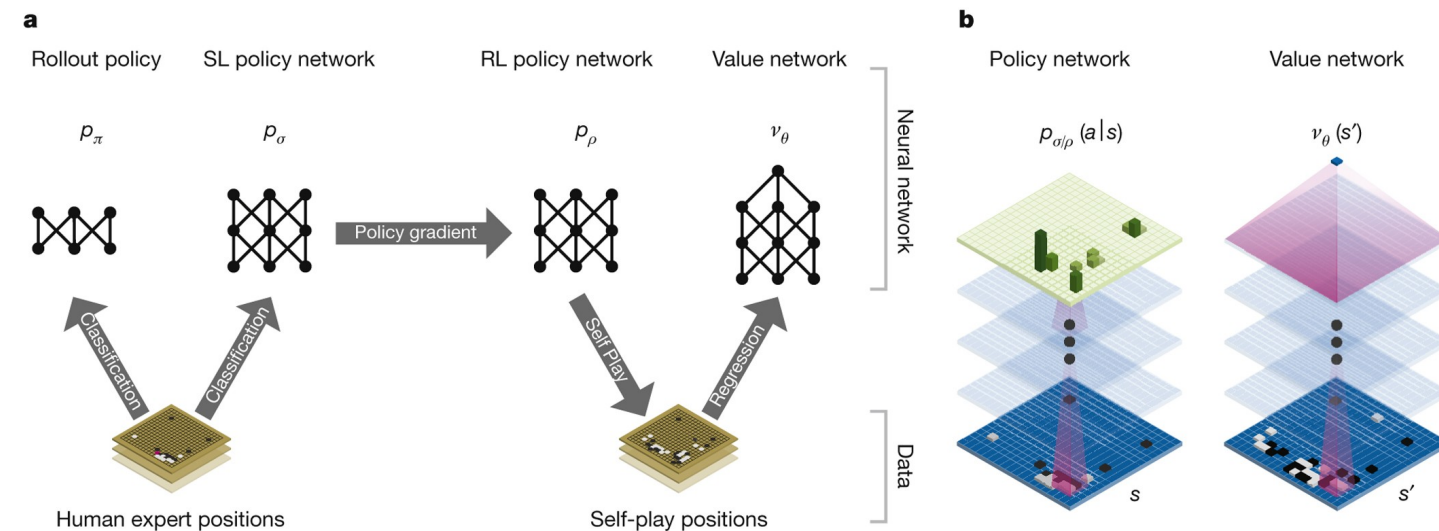
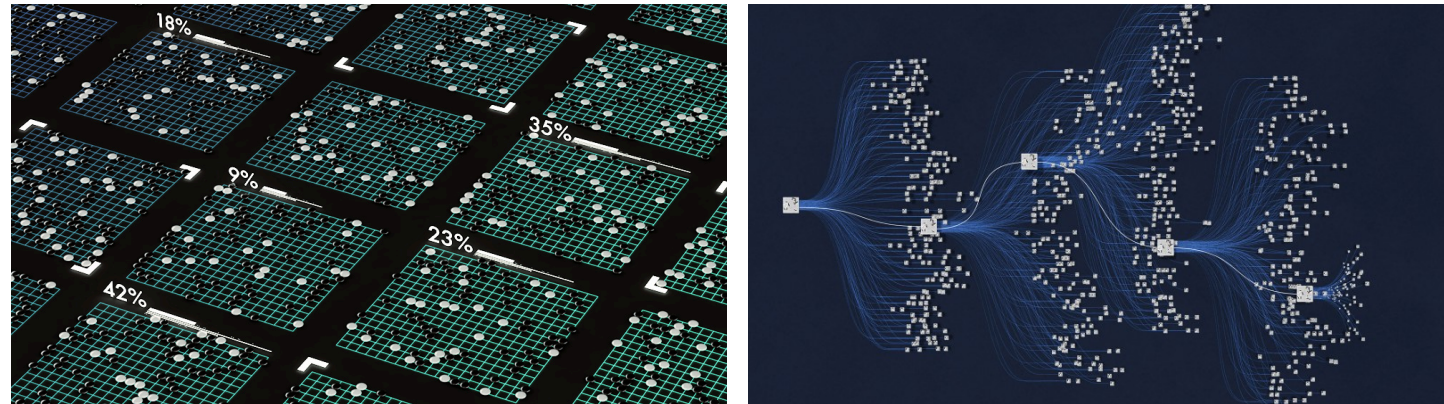
shivonzilis.com/MACHINEINTELLIGENCE · Bloomberg BETA

<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



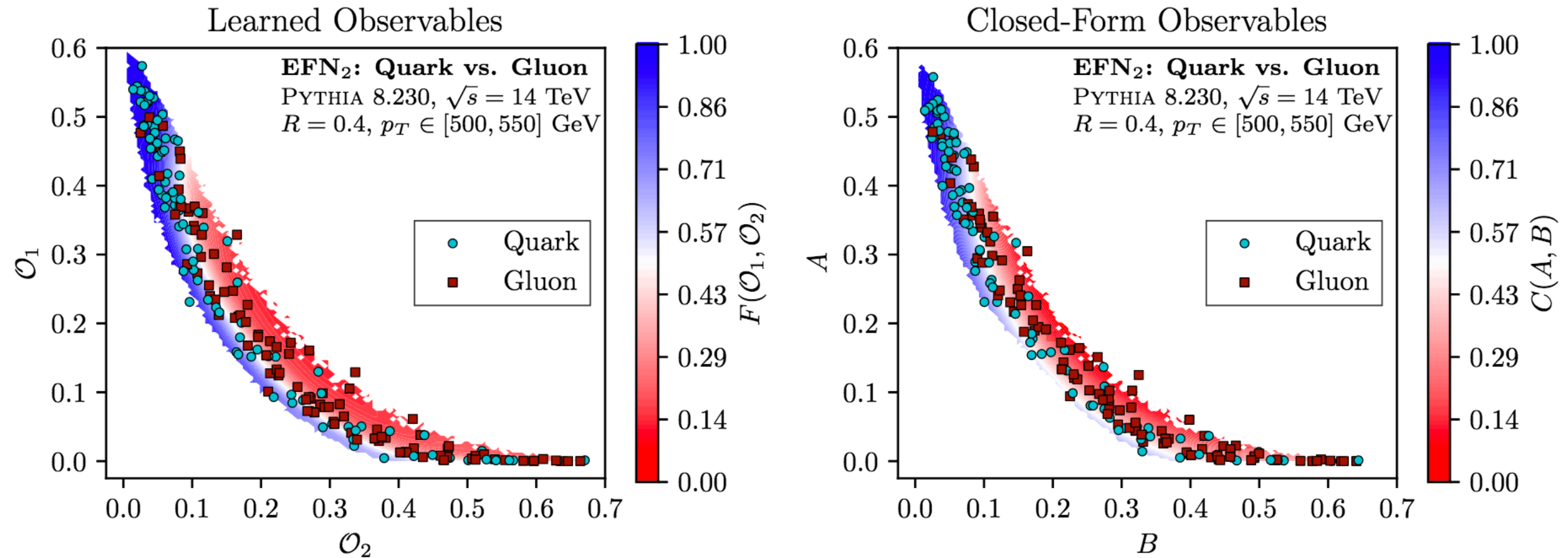
Learning to Walk via Deep Reinforcement Learning  
<https://arxiv.org/abs/1812.11103>

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

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



# Physics Knowledge

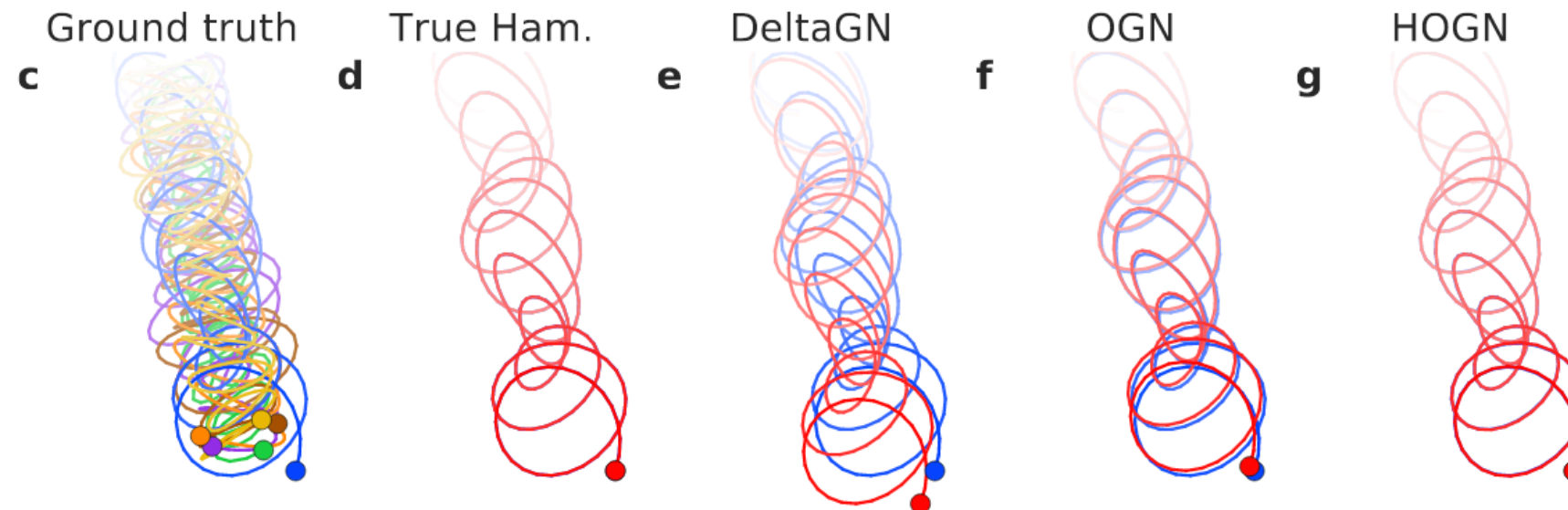
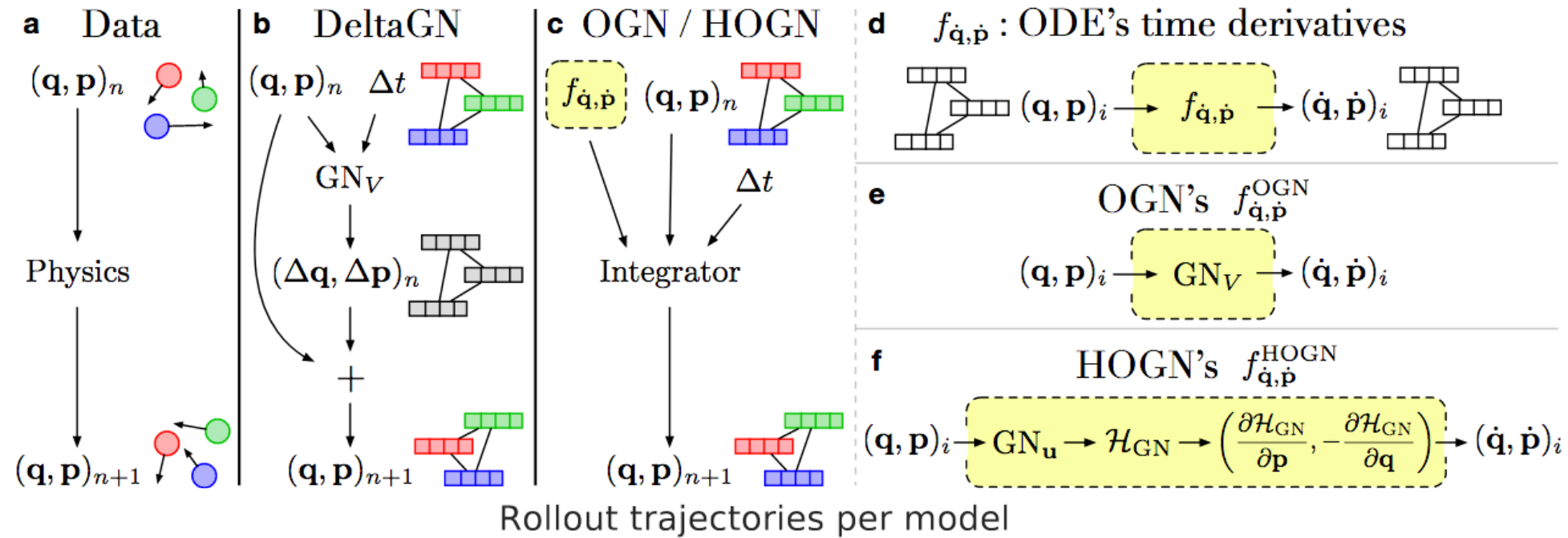


P. Komiske, E. Metodiev, J. Thaler, [\[1810.05165\]](#)

Machine Learning can **help understand Physics.**



# Use Physics

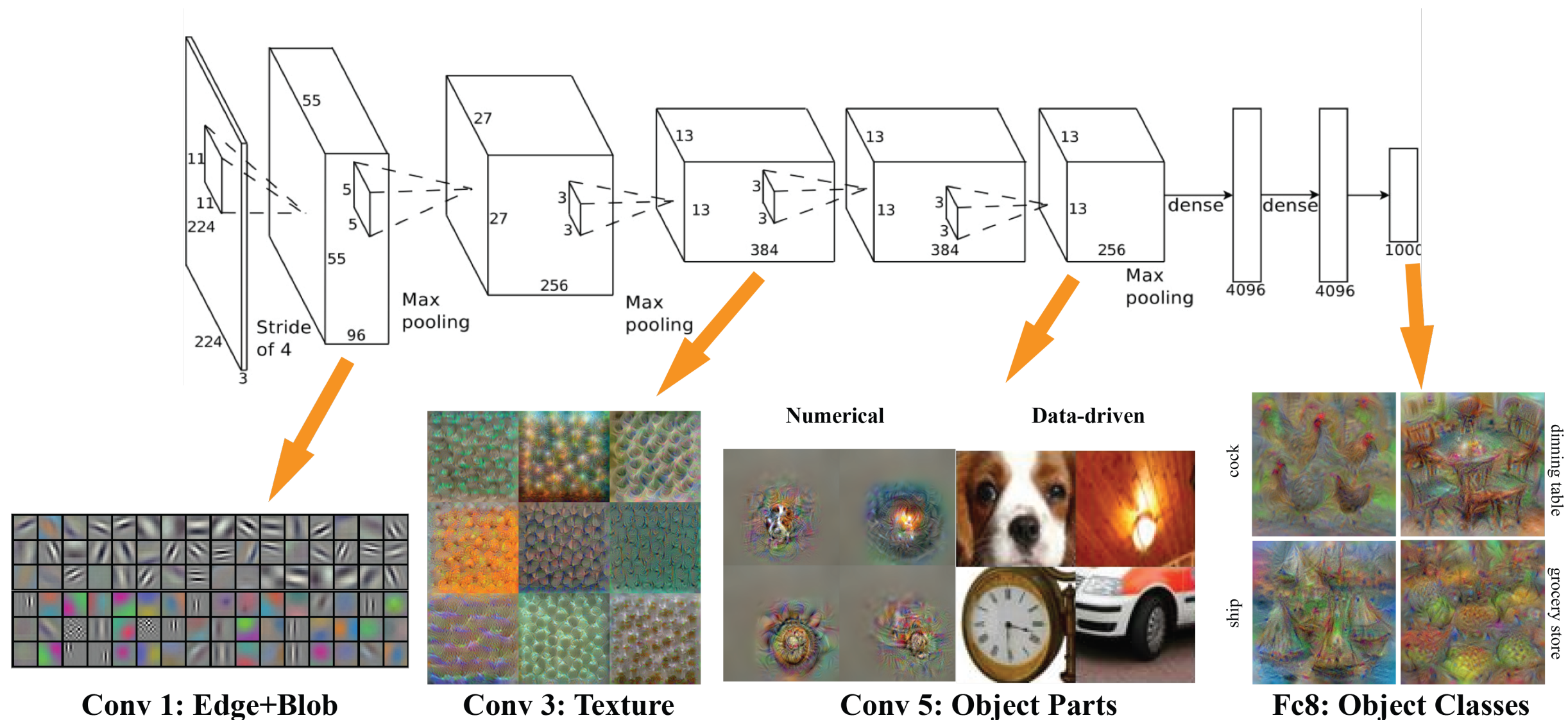


A. Sanchez-Gonzalez, V. Bapst, K. Cranmer, P. Battaglia [\[1909.12790\]](#)

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



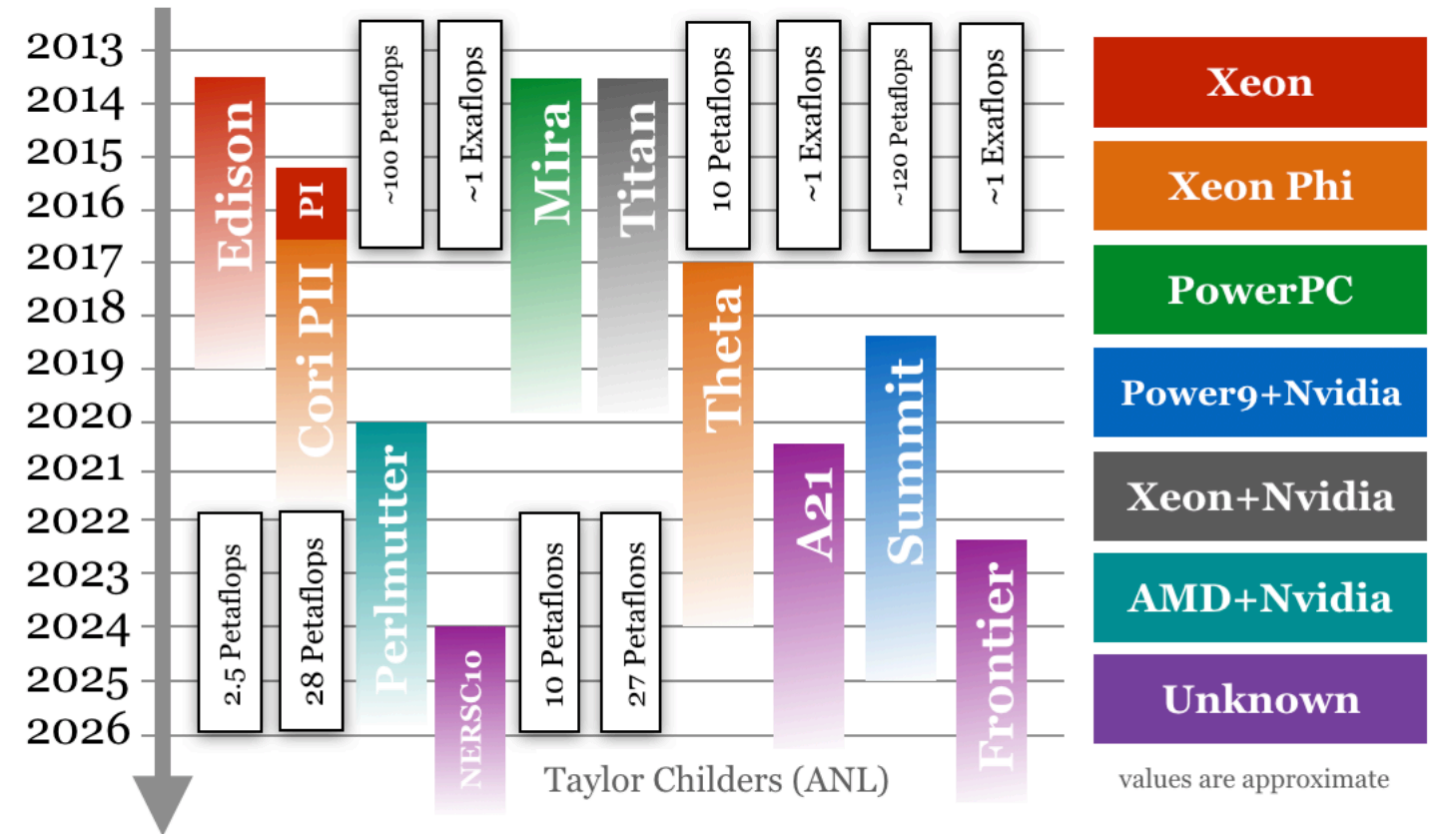
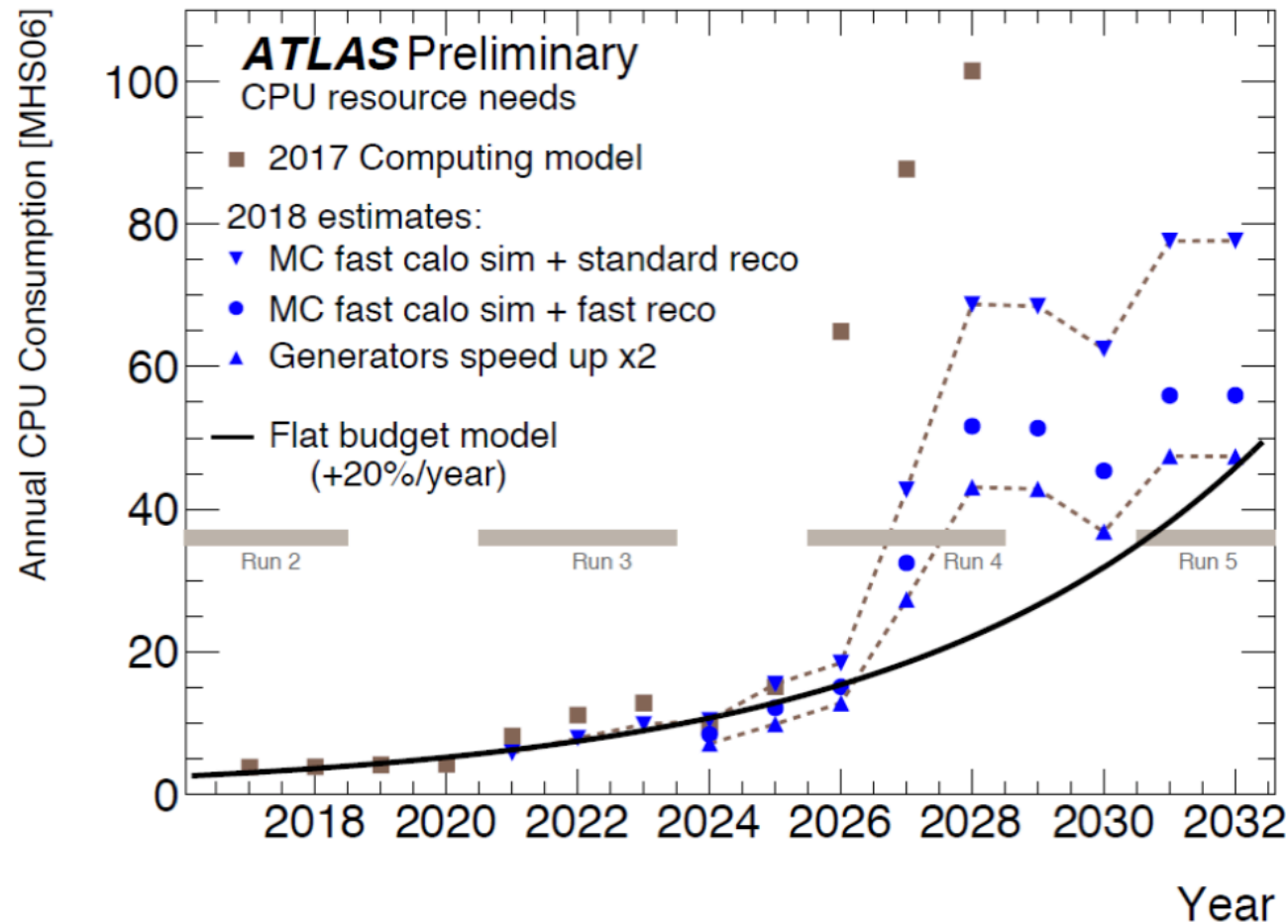
# Learning from Complexity



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



# The Computing Cost of Science

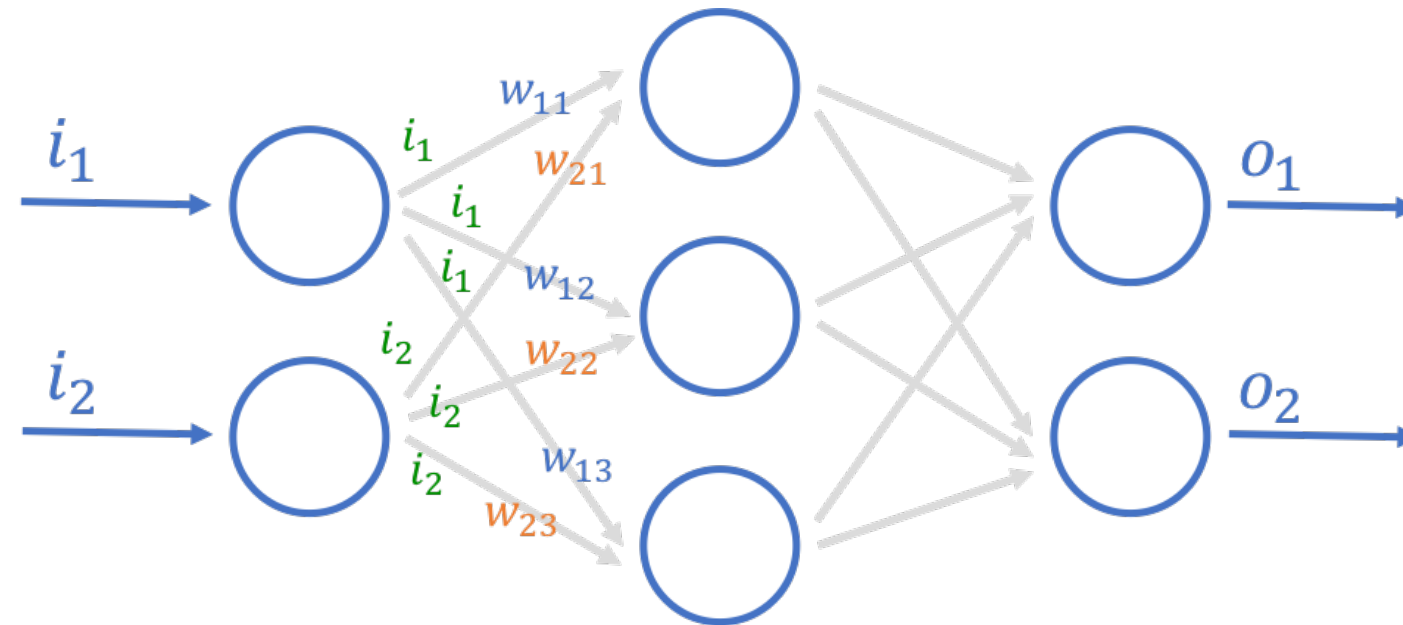


<https://indico.cern.ch/event/822126/contributions/3500169/>

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



# Operation Vectorisation



ANN  $\equiv$  matrix operations  $\equiv$  parallelizable

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

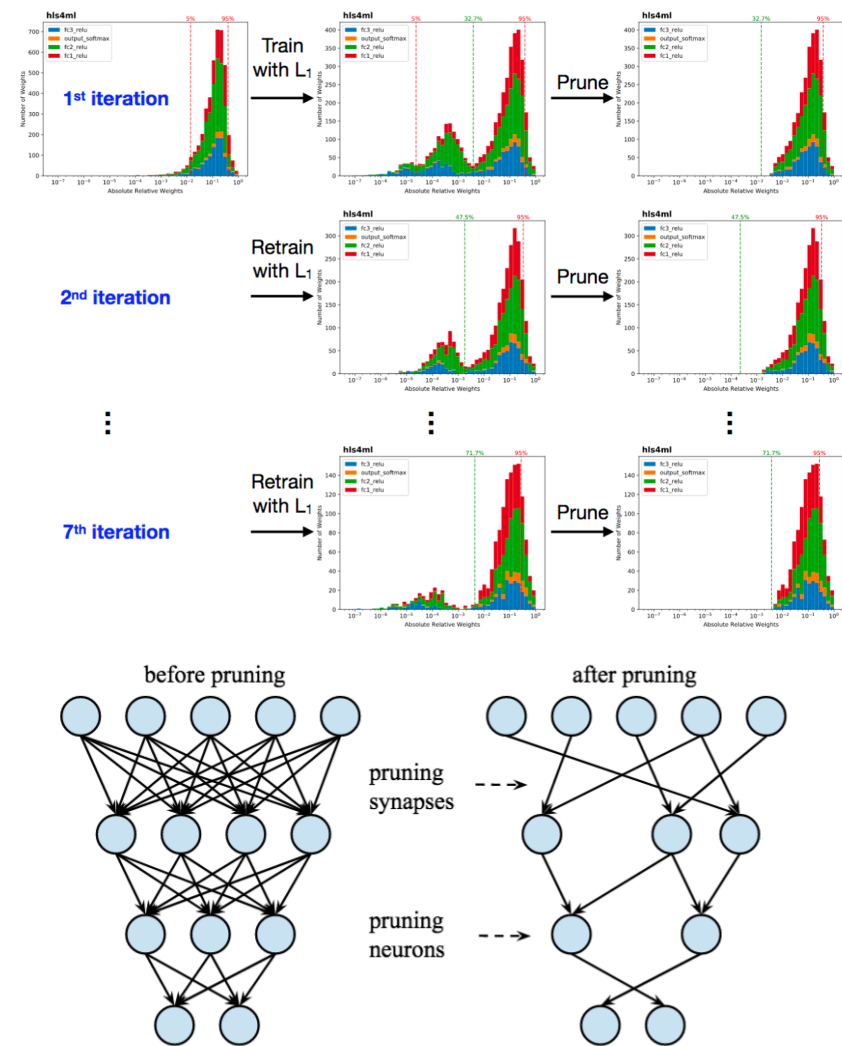
Computation of prediction from artificial neural network model can be **vectorised to a large extent.**



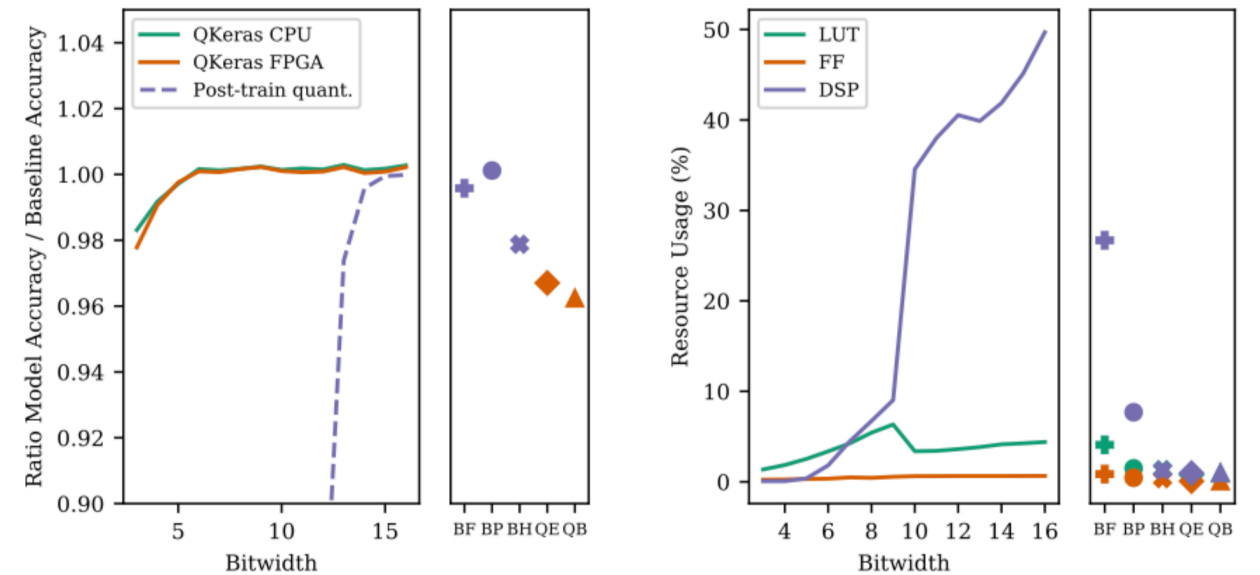
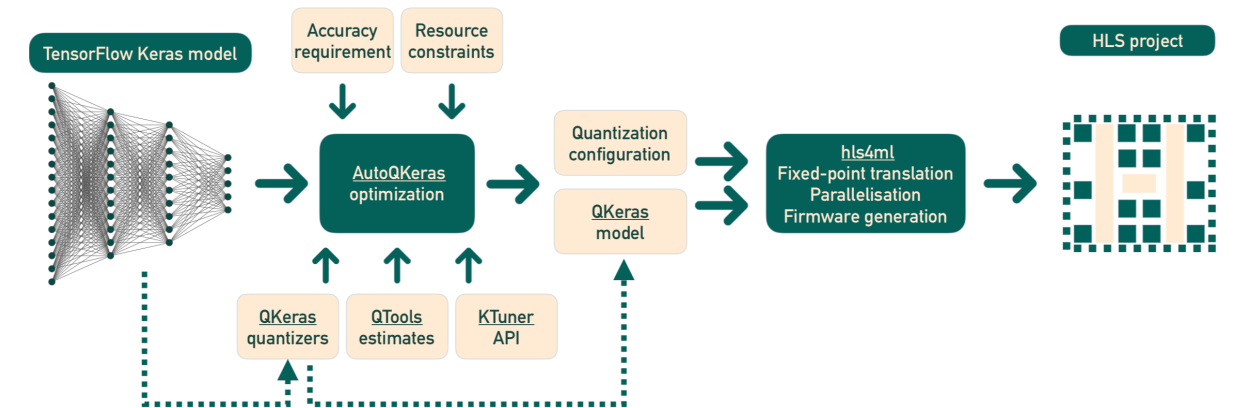


# Model Compression

Fast inference of deep neural networks in FPGAs for particle physics [\[1804.06913\]](#)



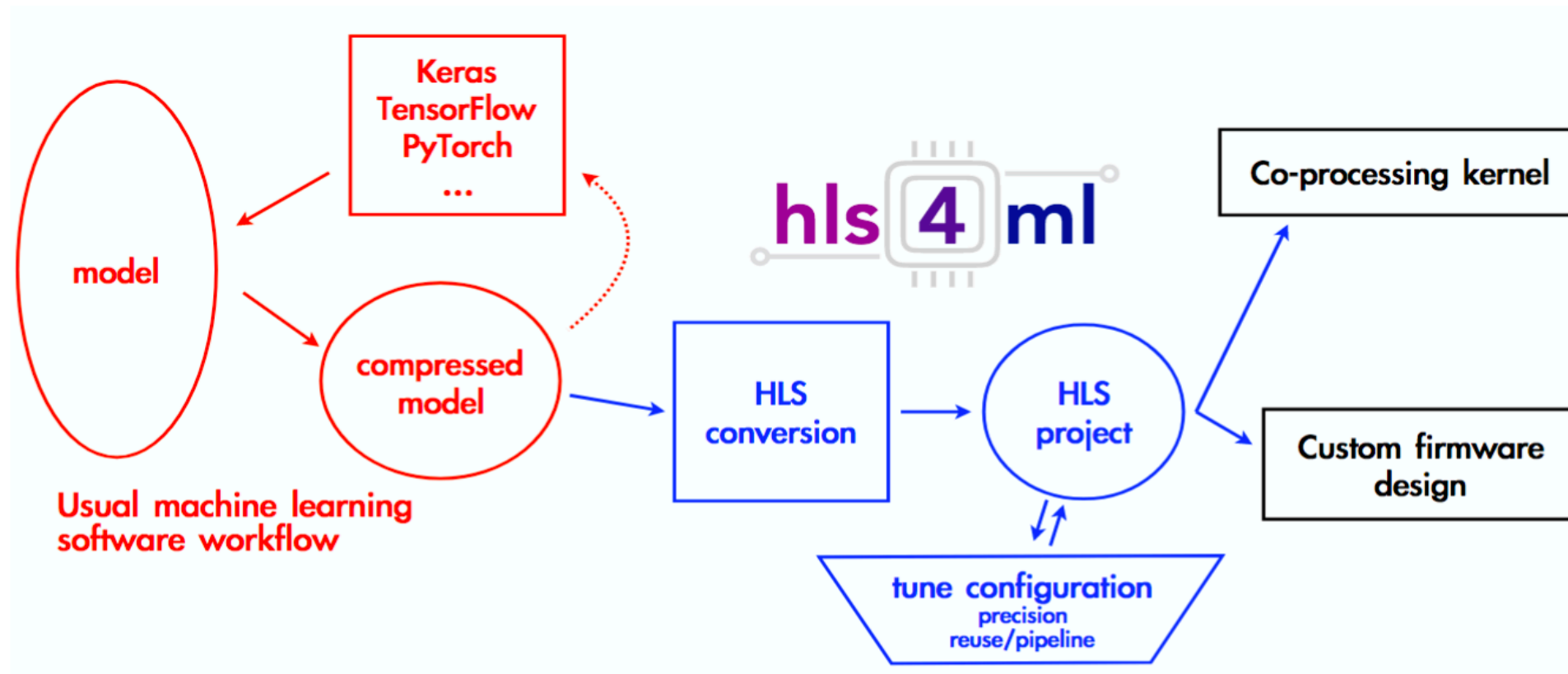
Automatic deep heterogeneous quantization of Deep Neural Networks for ultra low-area, low-latency inference on the edge at particle colliders [\[2006.10159\]](#)



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



# 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, ...**

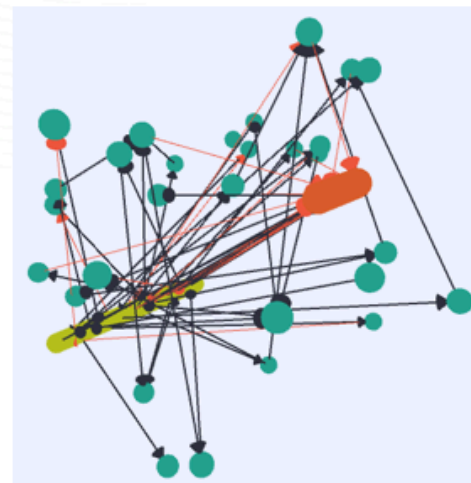


# Low Power Prediction

## Best Results: Single View



Convolutional Neural Network Result: ~80.42%



- 90 neurons, 86 synapses
- Estimated energy for a single classification for mrDANNA implementation: 1.66  $\mu$ J

Spiking Neural Network Result: ~80.63%

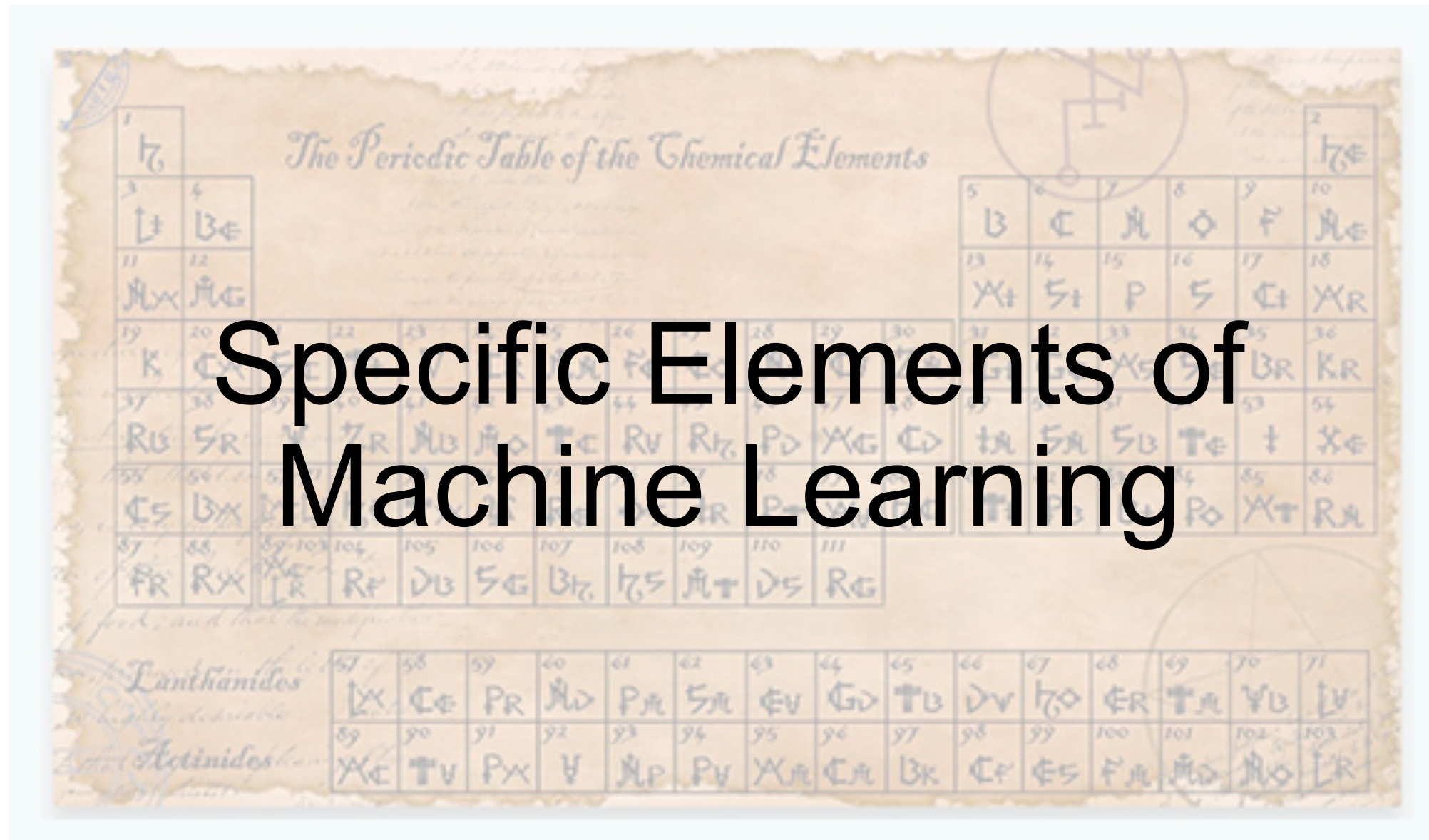
Source for CNN results: A. Terwilliger, et al. Vertex Reconstruction of Neutrino Interactions using Deep Learning. IJCNN 2017.  
33 Programming Neuromorphic Computing Systems



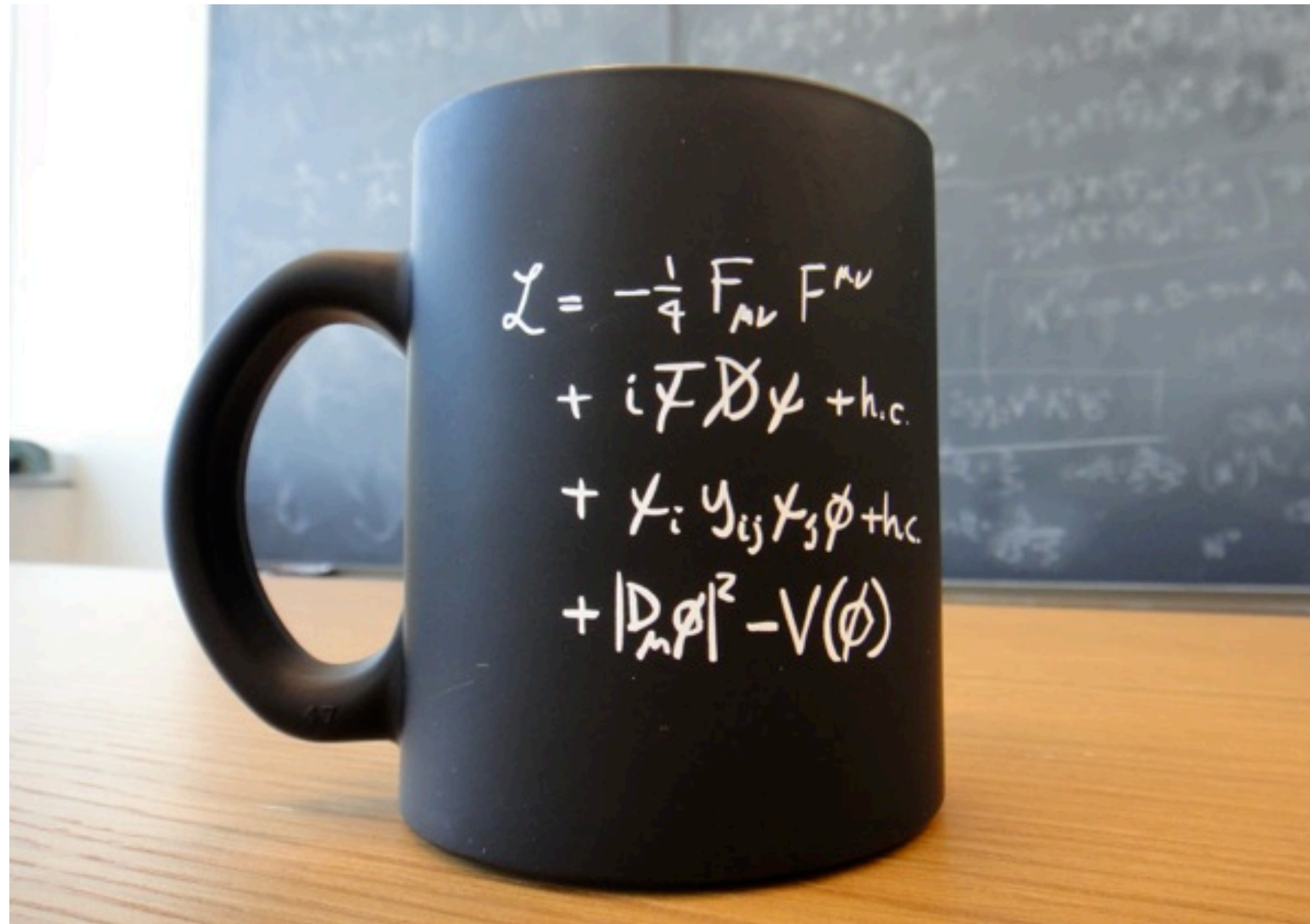
<https://indico.fnal.gov/event/13497/contribution/0> Slide C. Schuman

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





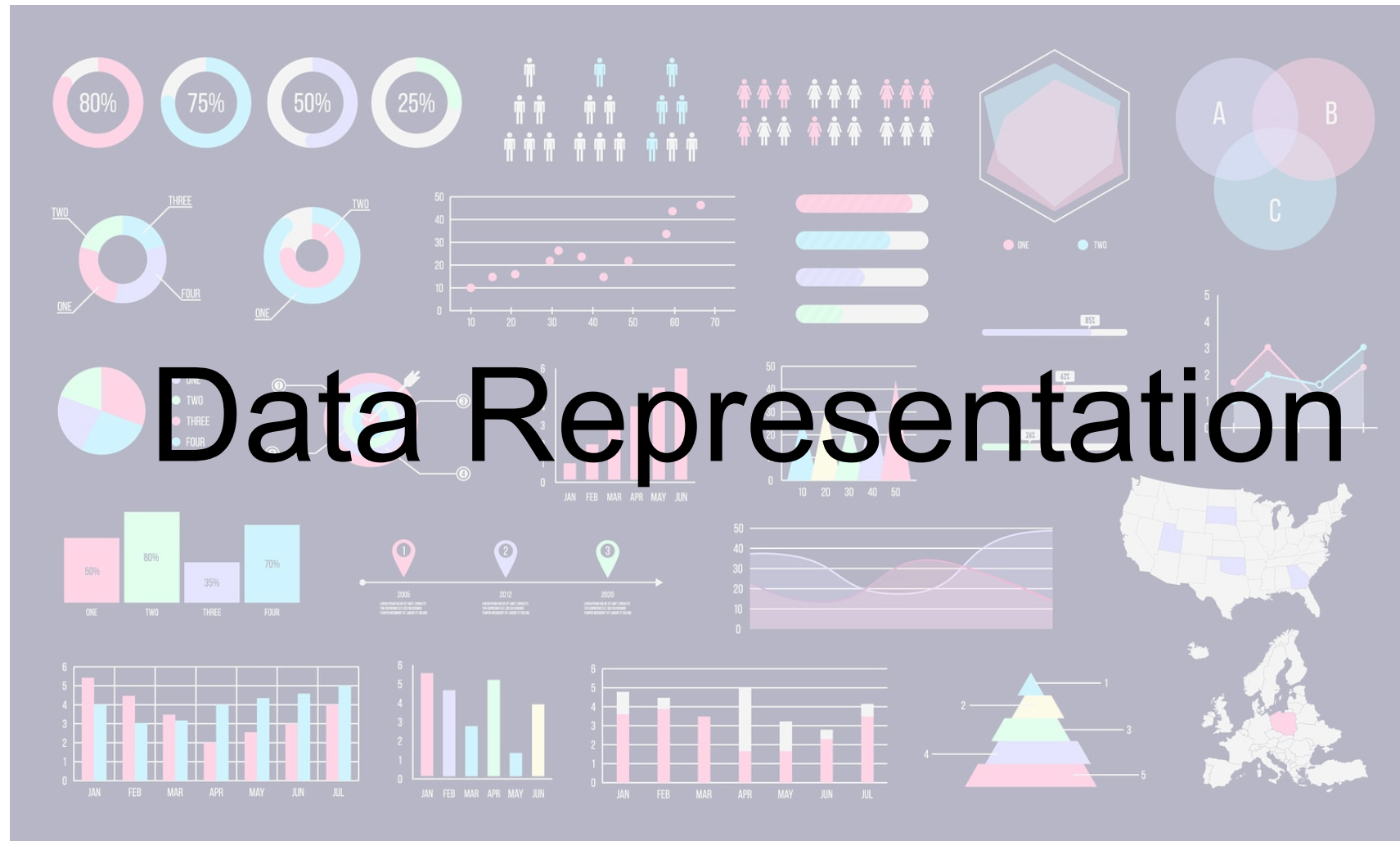
# The Standard Model



Well demonstrated effective model.

~infinite amount of detailed, **“labelled” simulation available.**

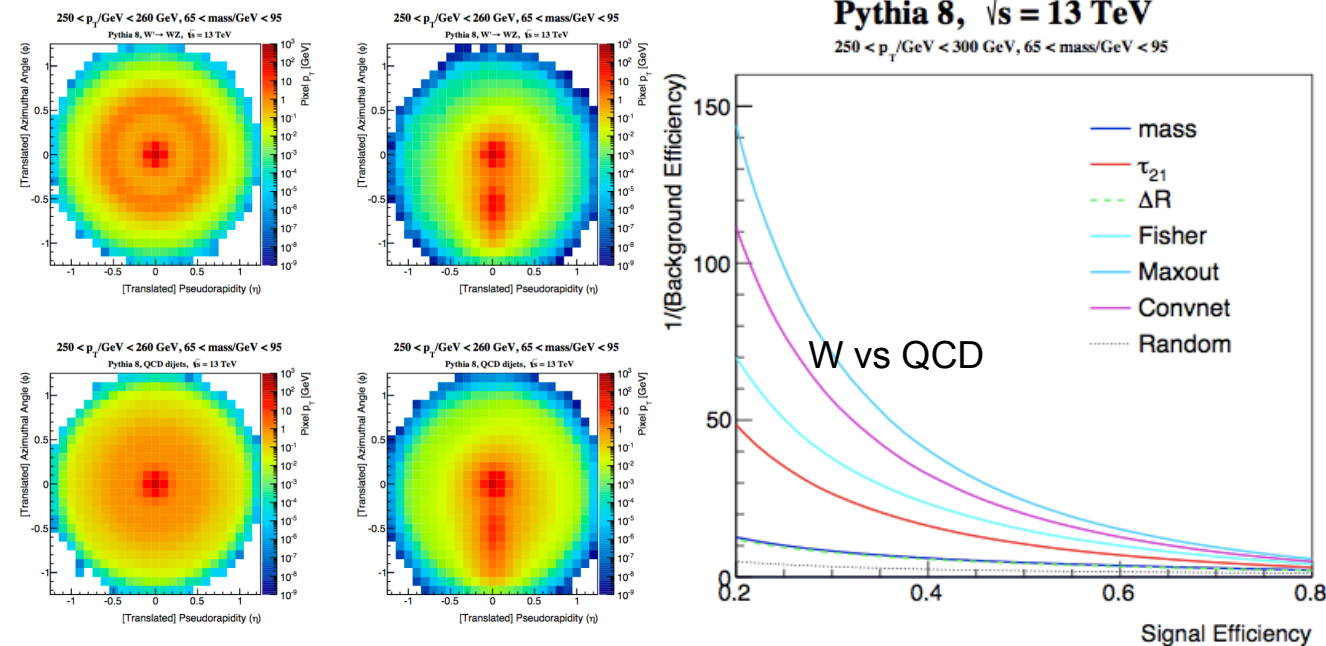




# Image Representation

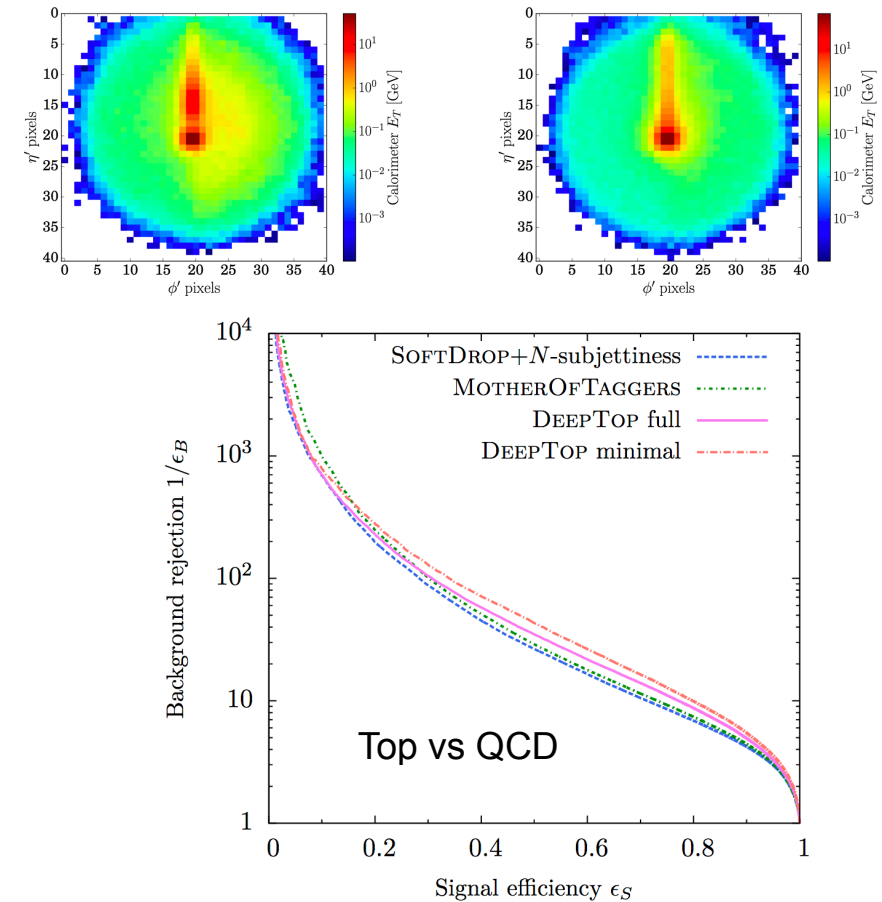
## Jet-Images – Deep learning edition

[1511.05190]



## Deep-learning top taggers or the end of QCD?

[1701.08784]



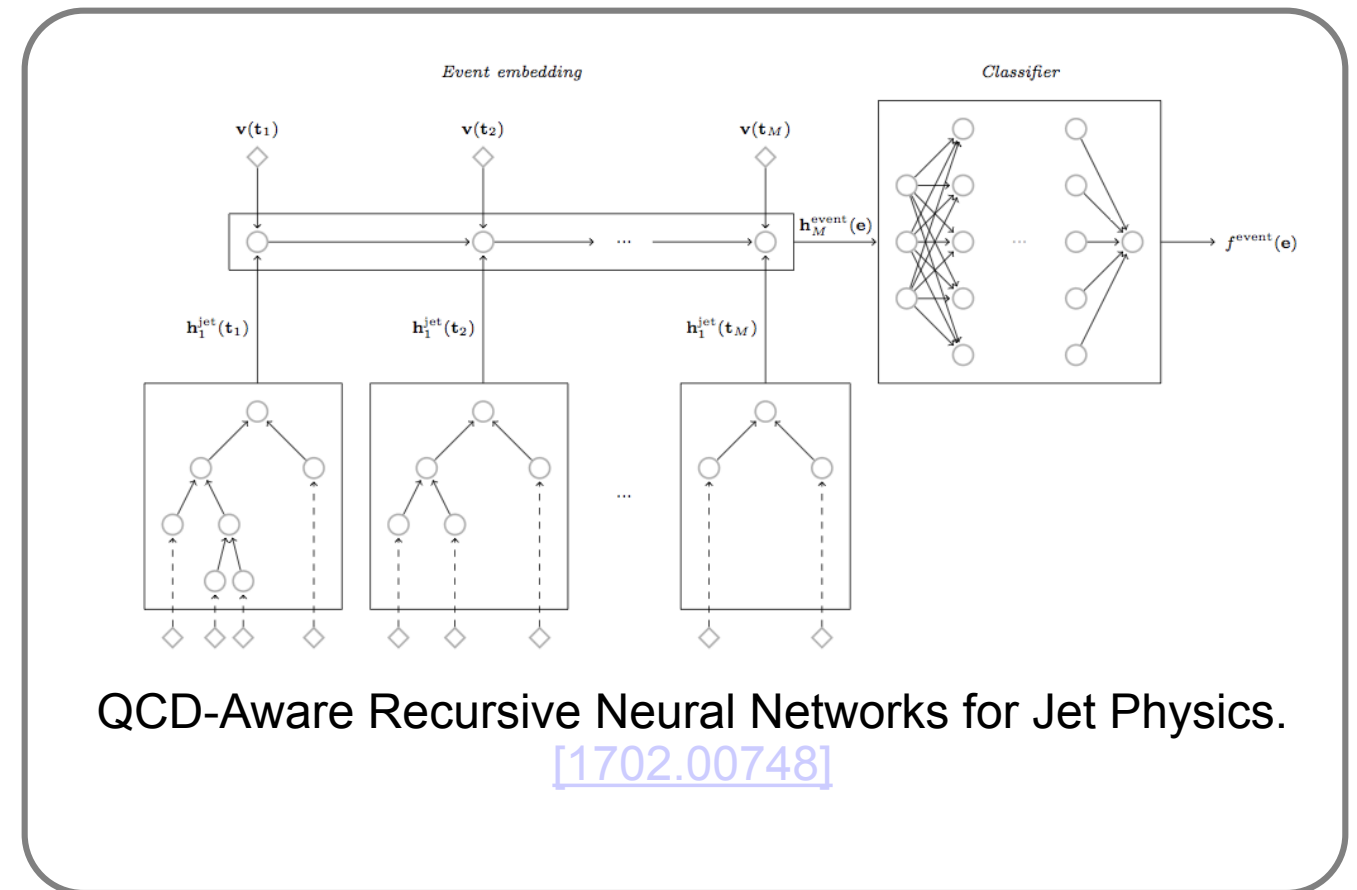
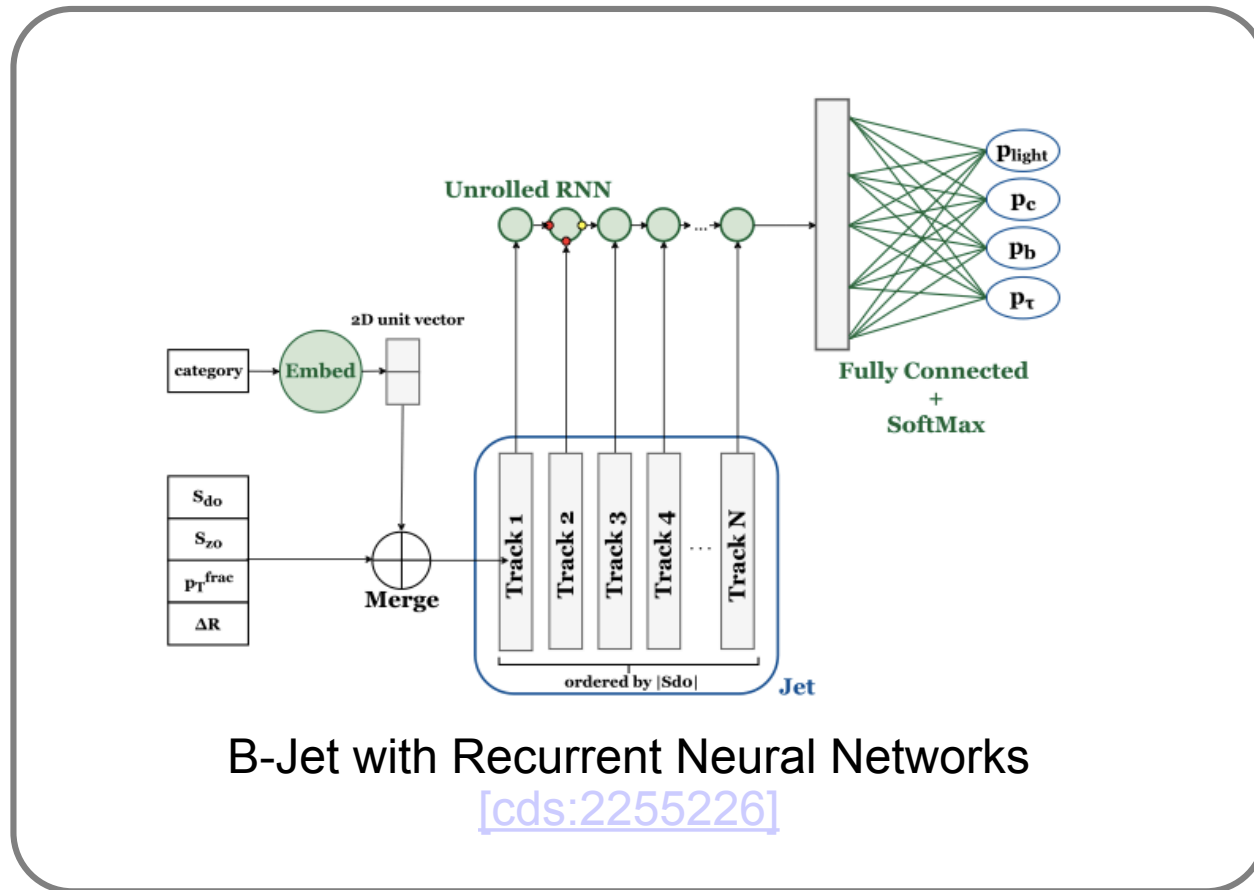
Calorimeter signal are image-like.

Projection of reconstructed particle properties onto images possible.

Potential loss of information during projection.



# Sequence Representation

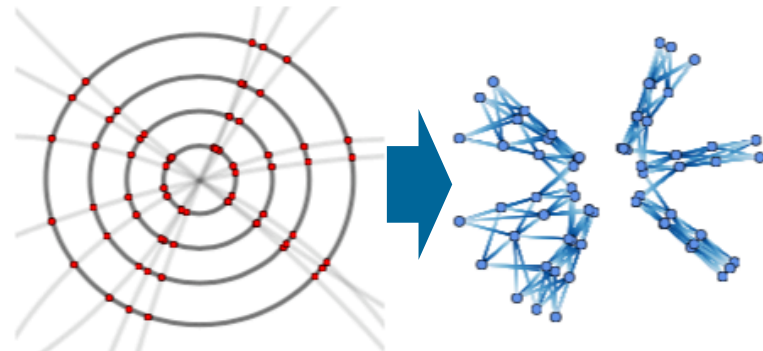


Somehow arbitrary choice on ordering with sequence representation.  
Physics-inspired ordering as inductive bias.  
Ordering can be learned too somehow.

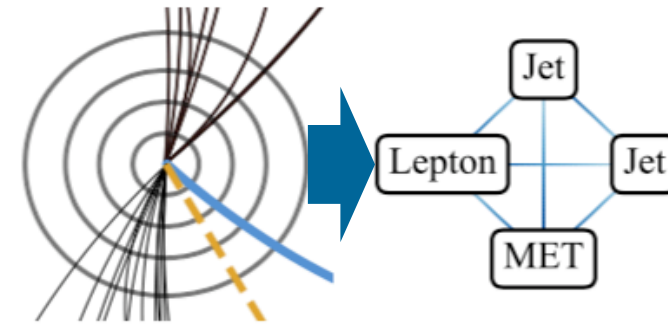




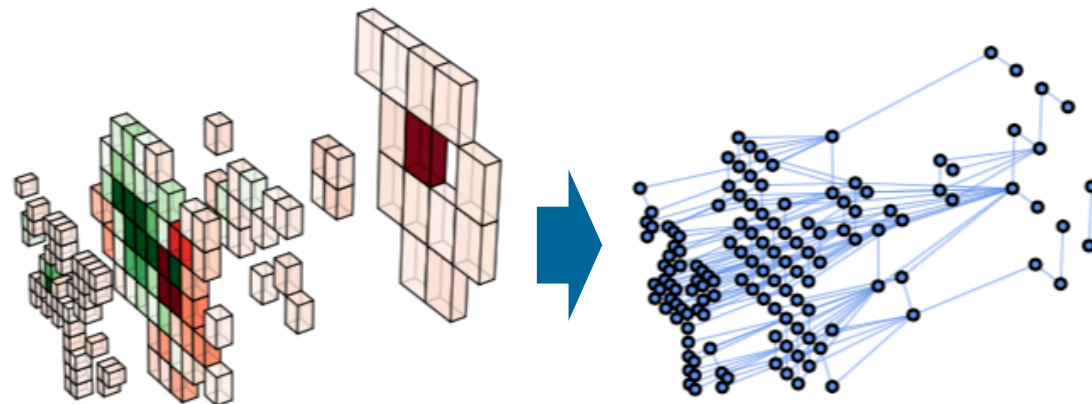
# Graph Representation



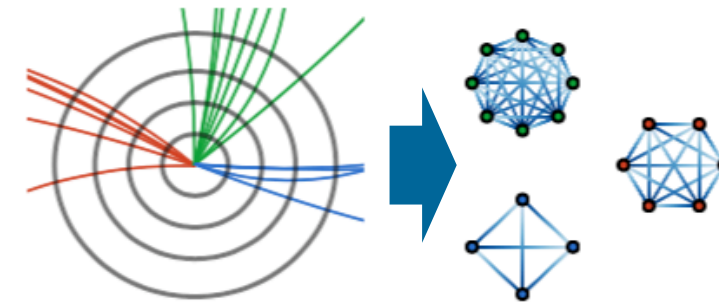
Hits in tracking detector



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.

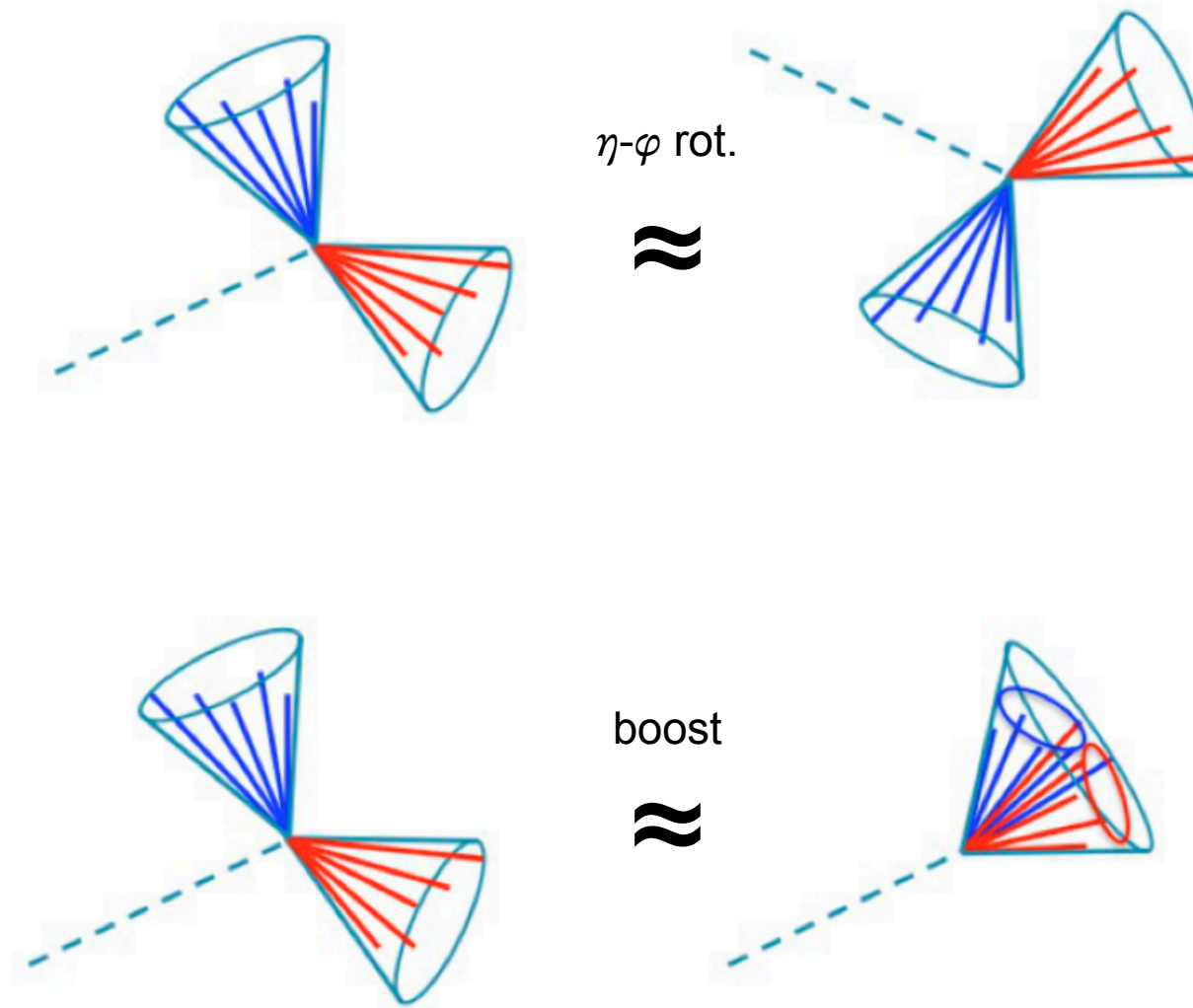




# Invariance and Symmetries



# Dataset Degeneracy



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



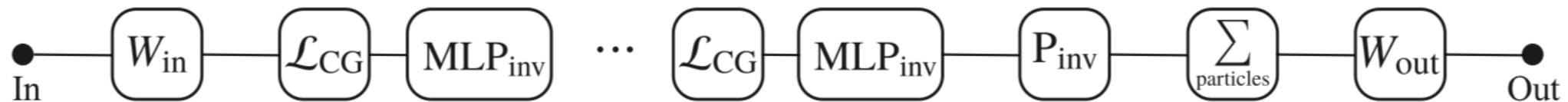
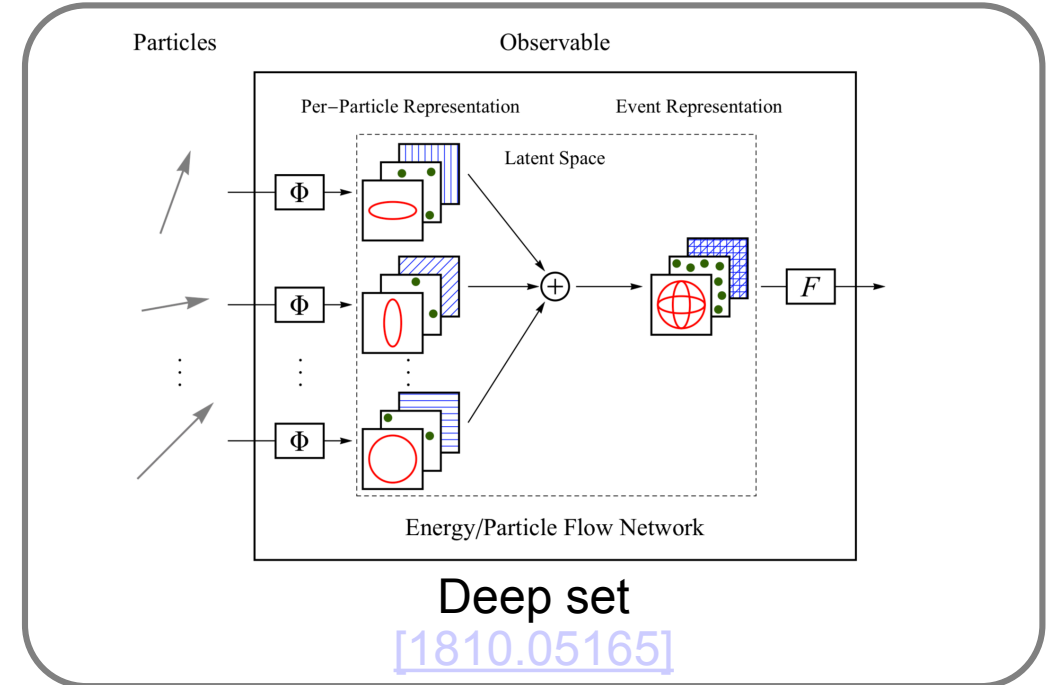
# Inductive Bias

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

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

Lorentz Learning Layer

[\[1707.08966\]](#)



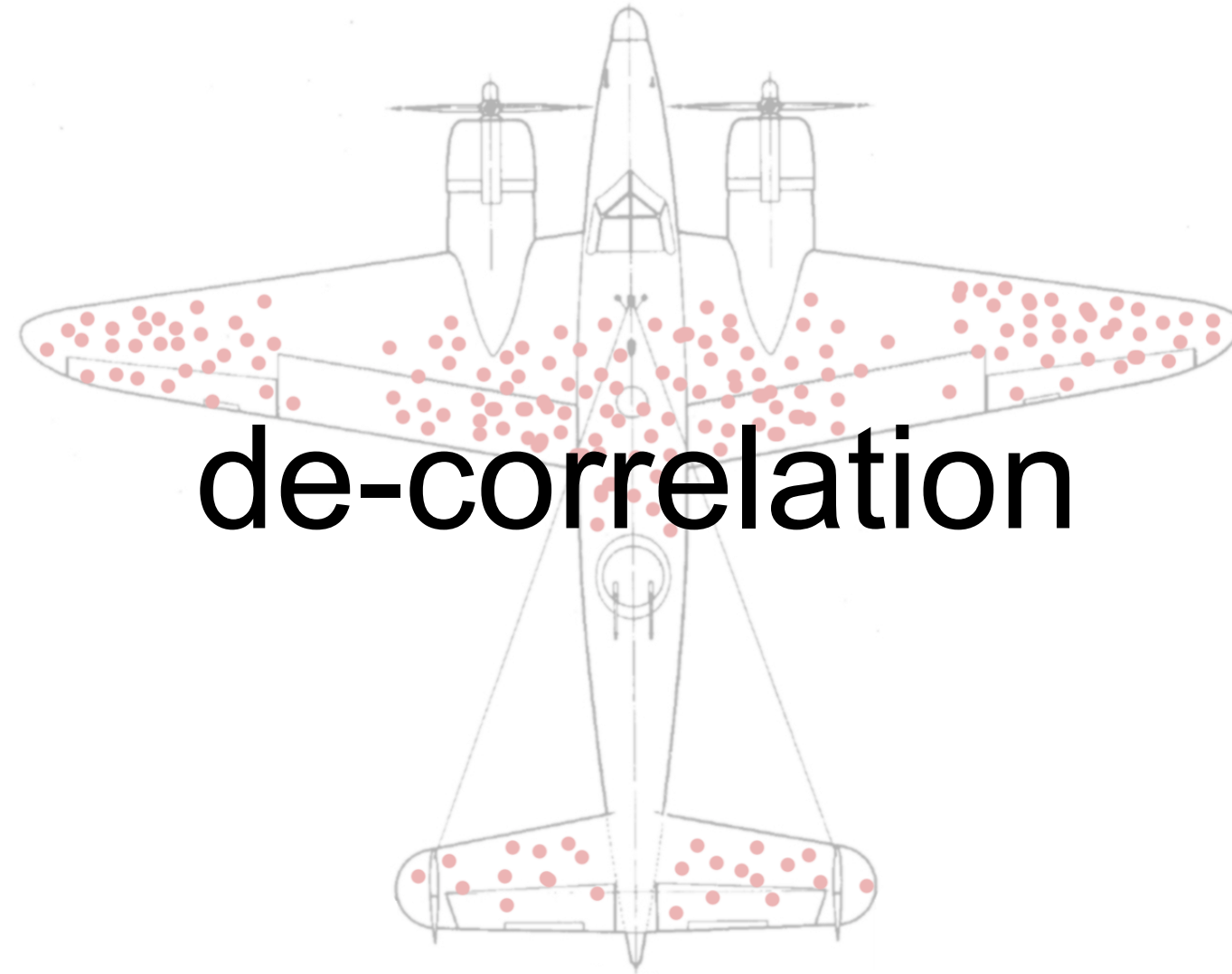
$$\mathcal{F}_i \mapsto W \cdot \left( \mathcal{F}_i \oplus \mathcal{F}_i^{\otimes 2} \oplus \sum_j f(p_{ij}^2) \cdot p_{ij} \otimes \mathcal{F}_j \right)$$

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).

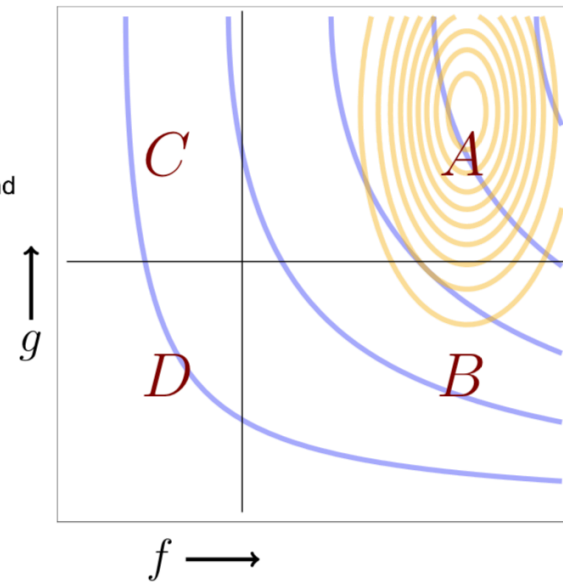
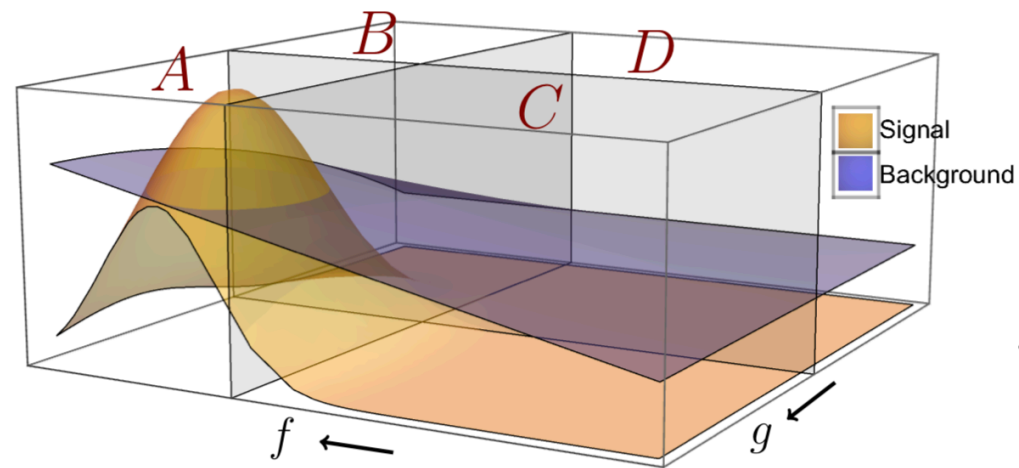


Numerous methods proposed to de-correlate model predictions and quantities of interest ( $p_T$ , mass, ...).

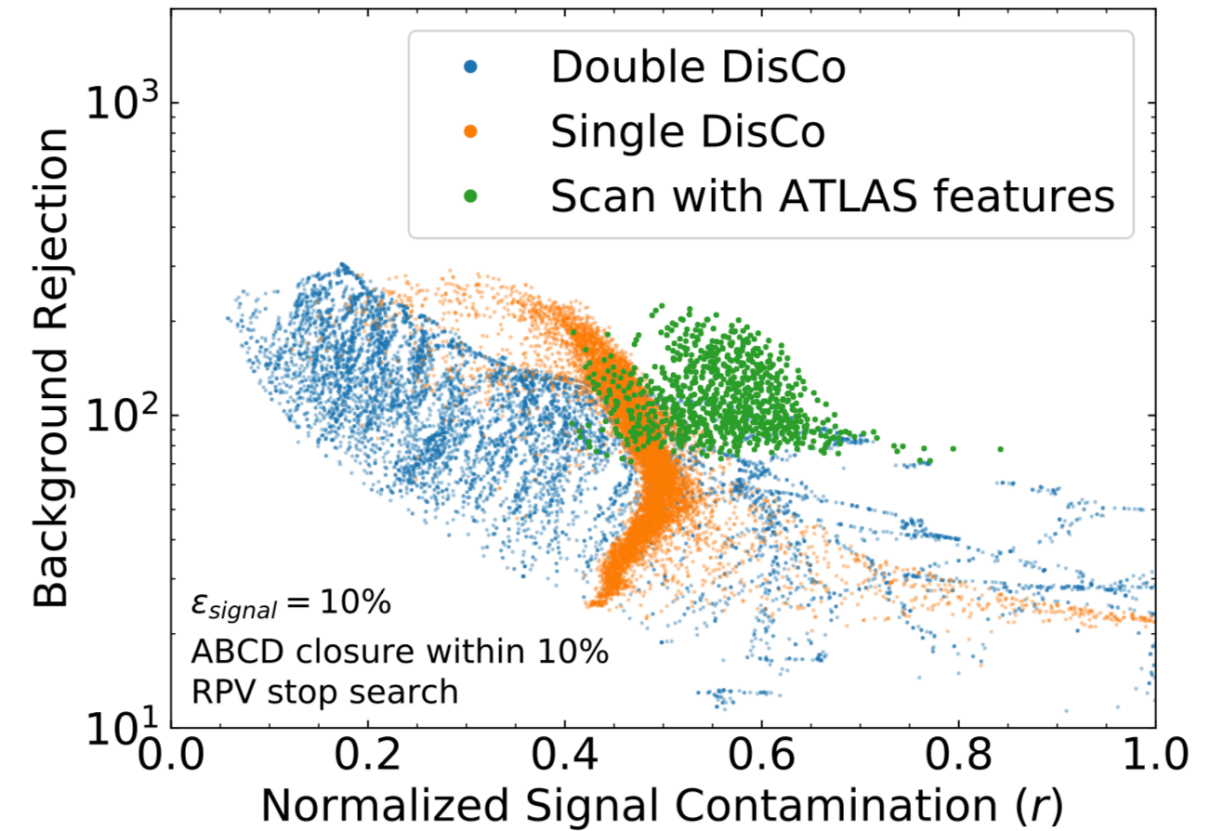
Usually adding a term in the loss to constrain de-correlation.



# Background Estimation

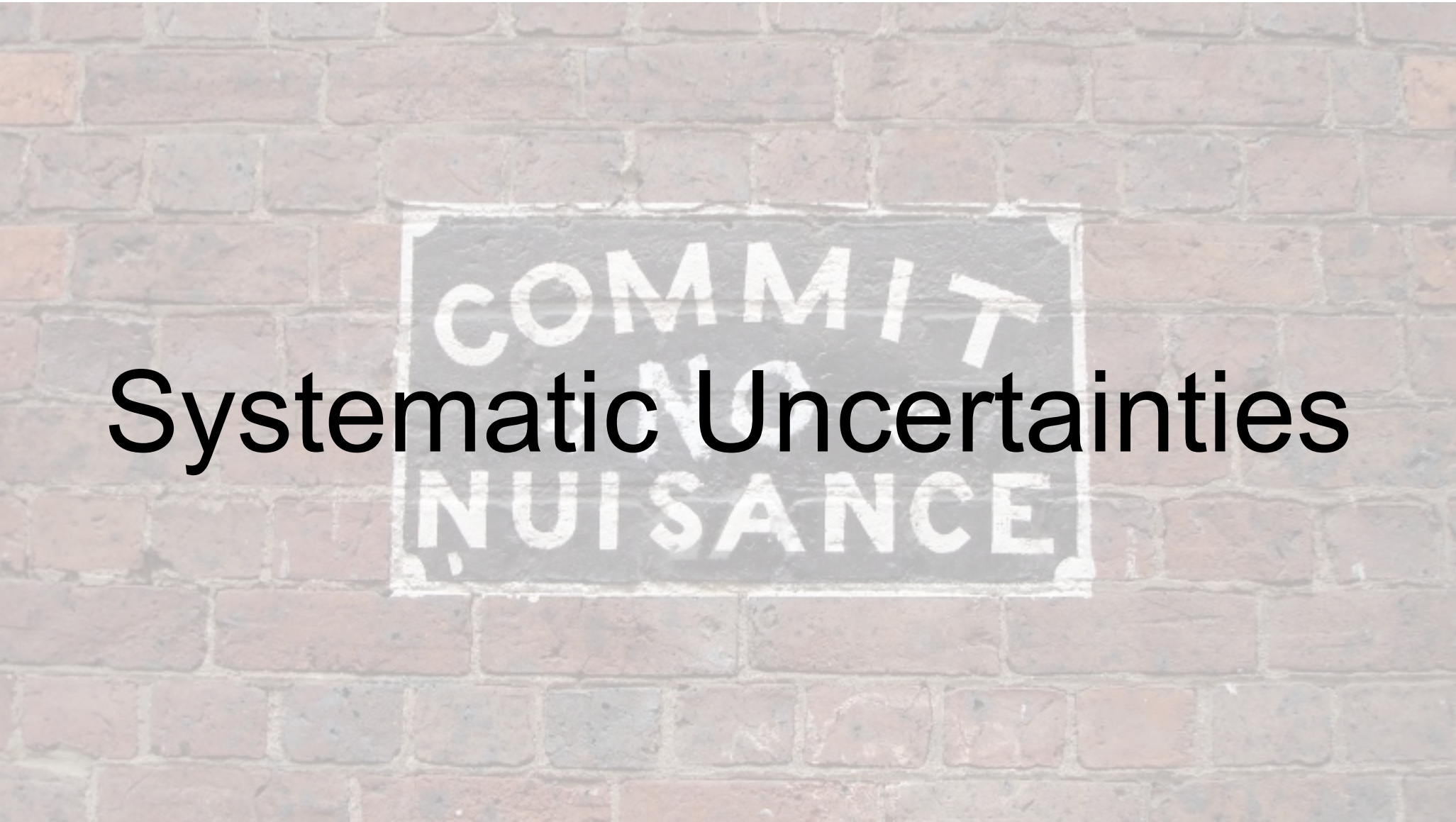


ABCD + Disco  
[\[2007.14400\]](#)



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



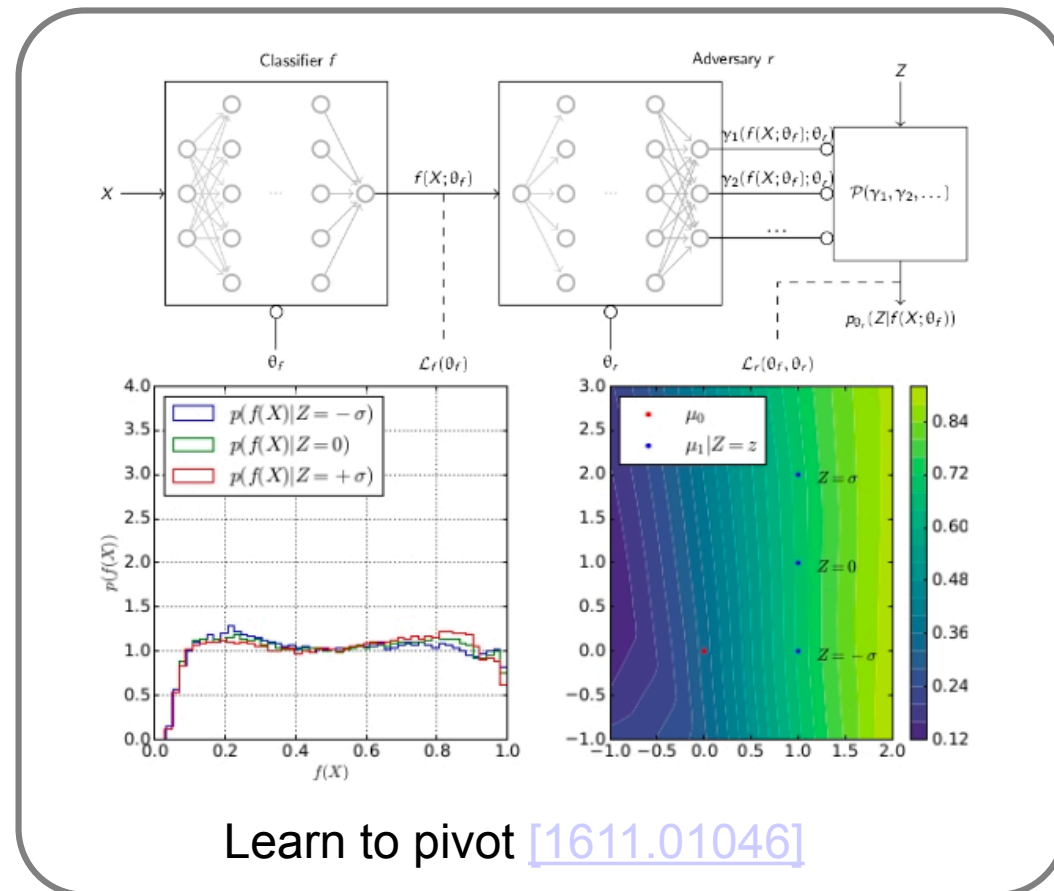


# Systematic Uncertainties

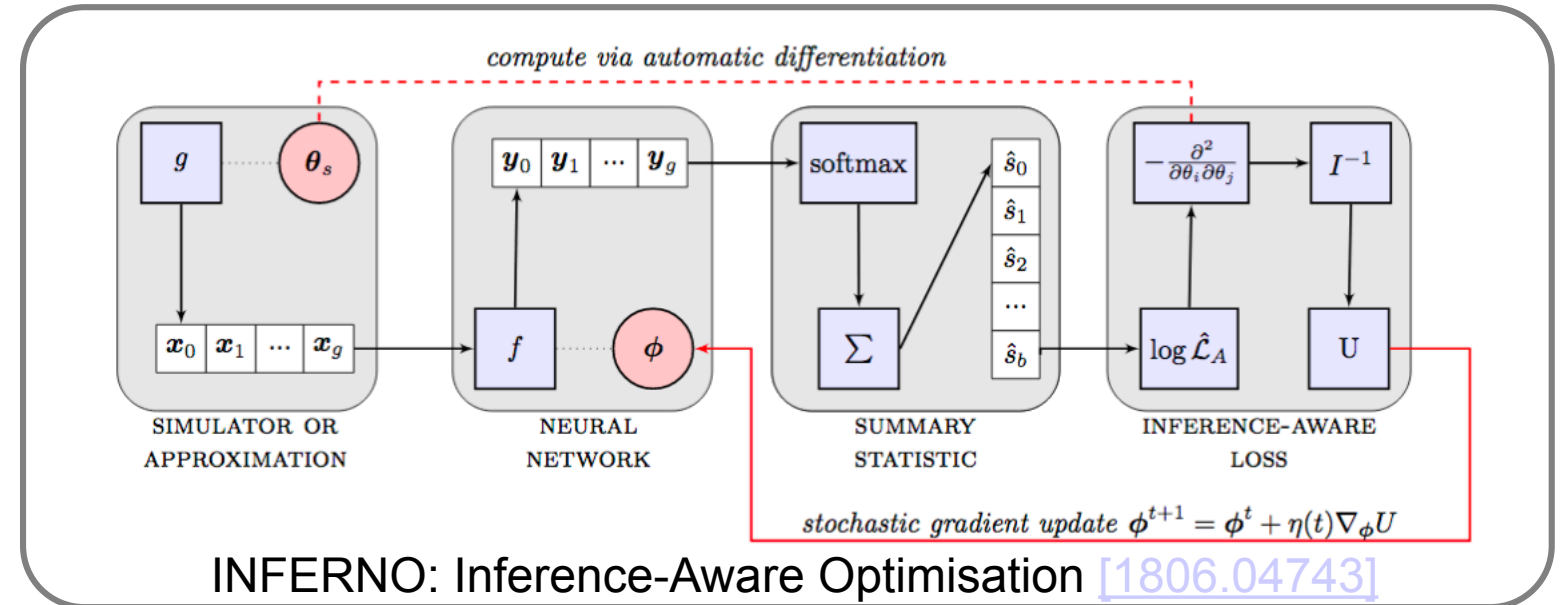




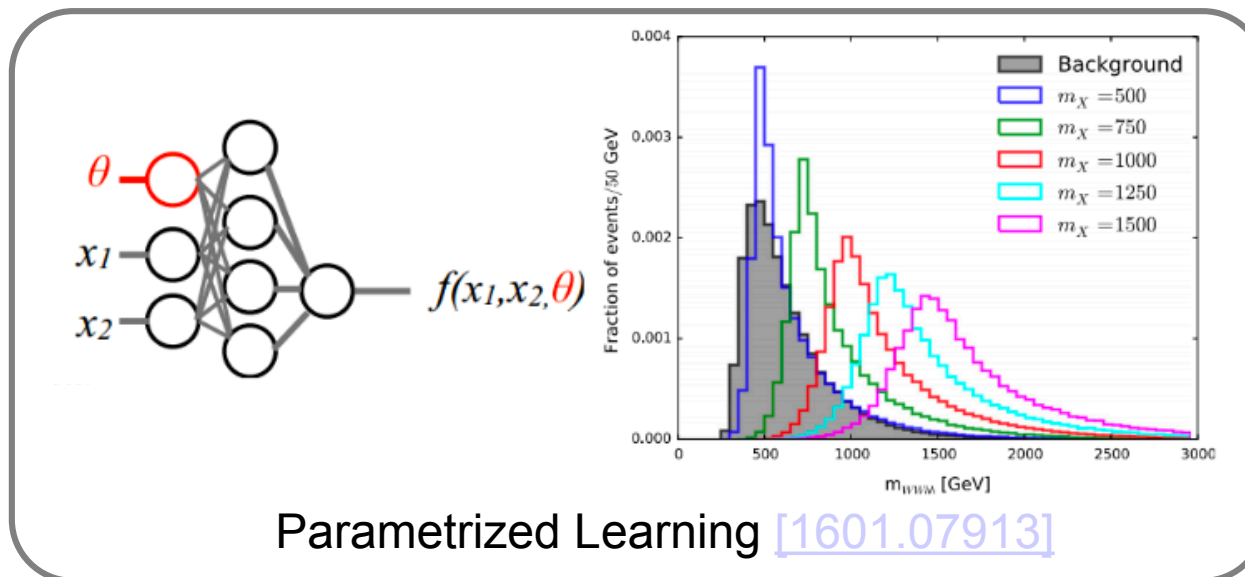
# System. Estimation and Mitigation



Learn to pivot [\[1611.01046\]](#)



INFERNNO: 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.

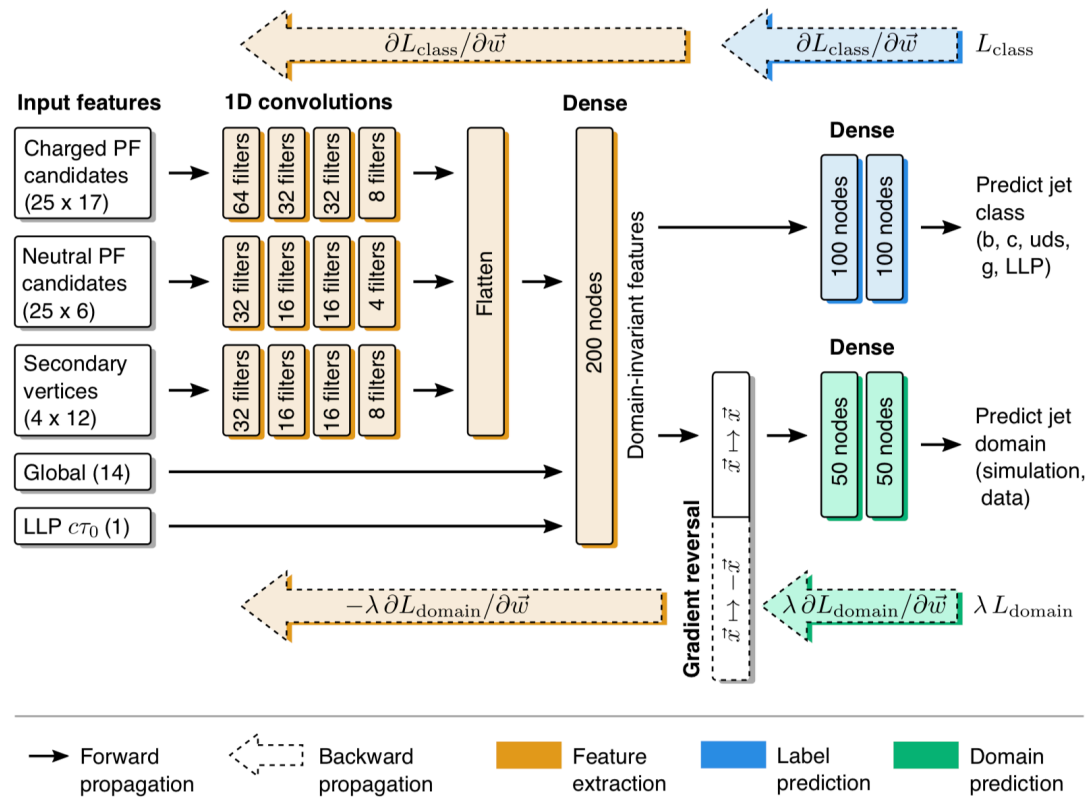




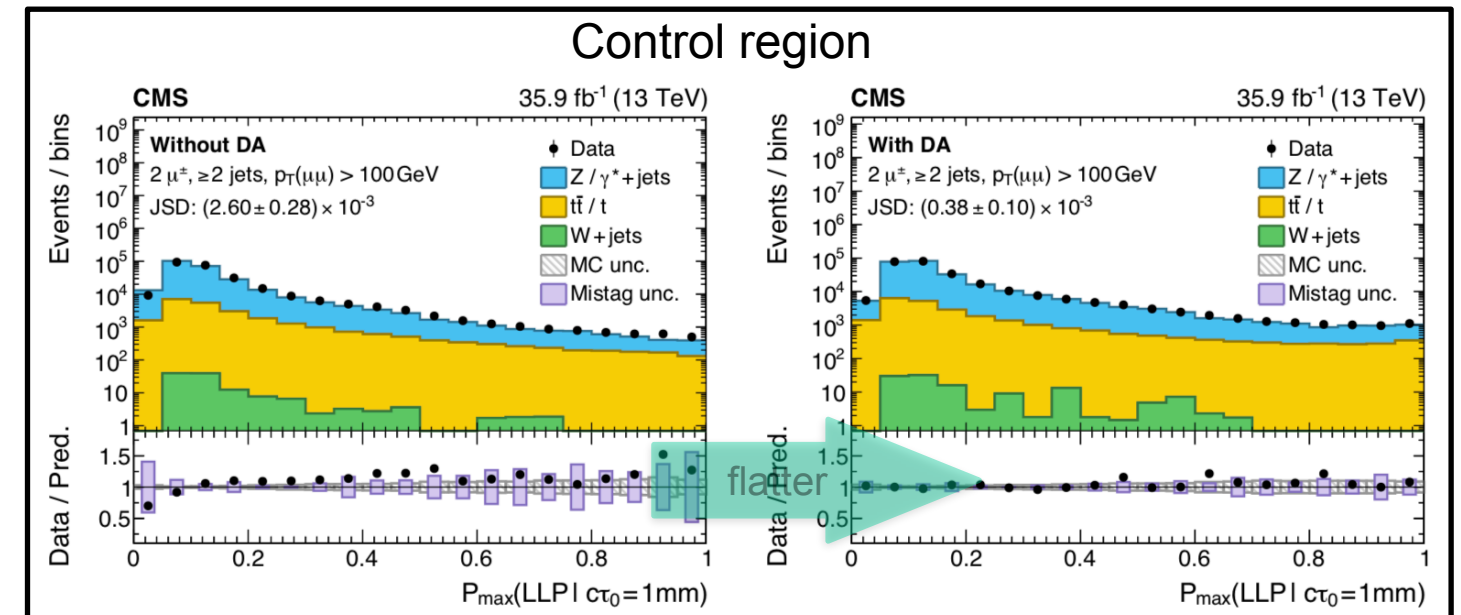
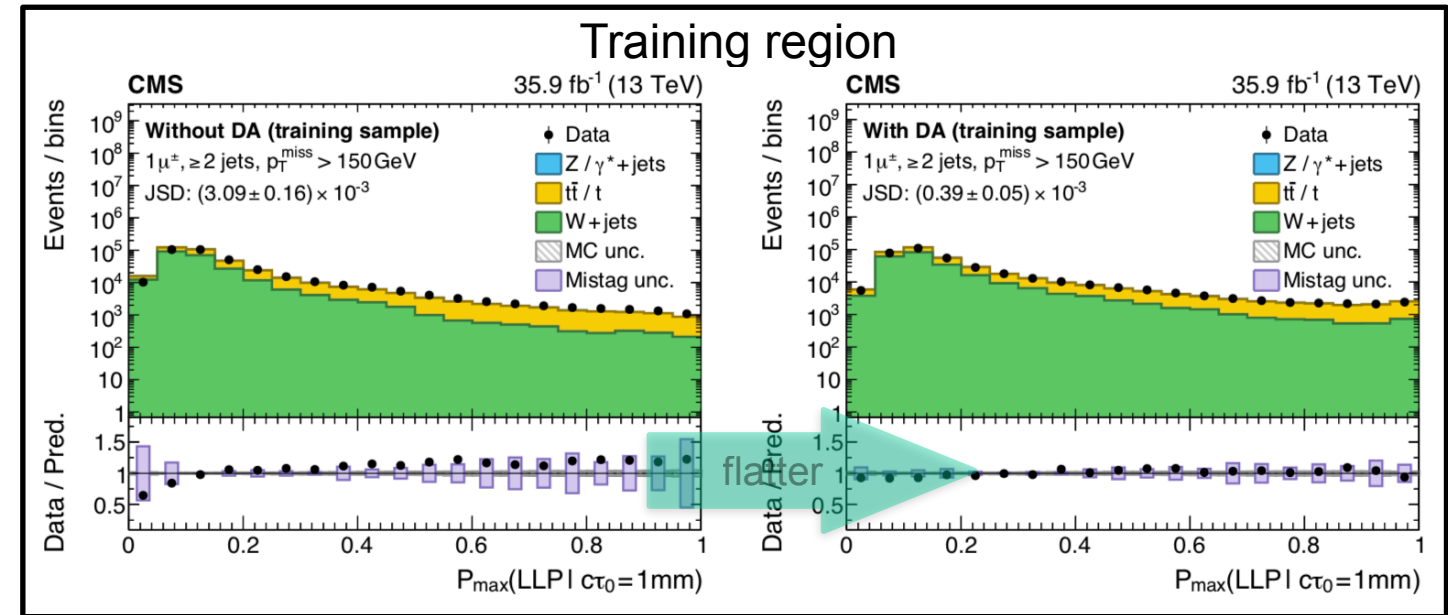
# Domain Dependence



# Domain in-Dependence

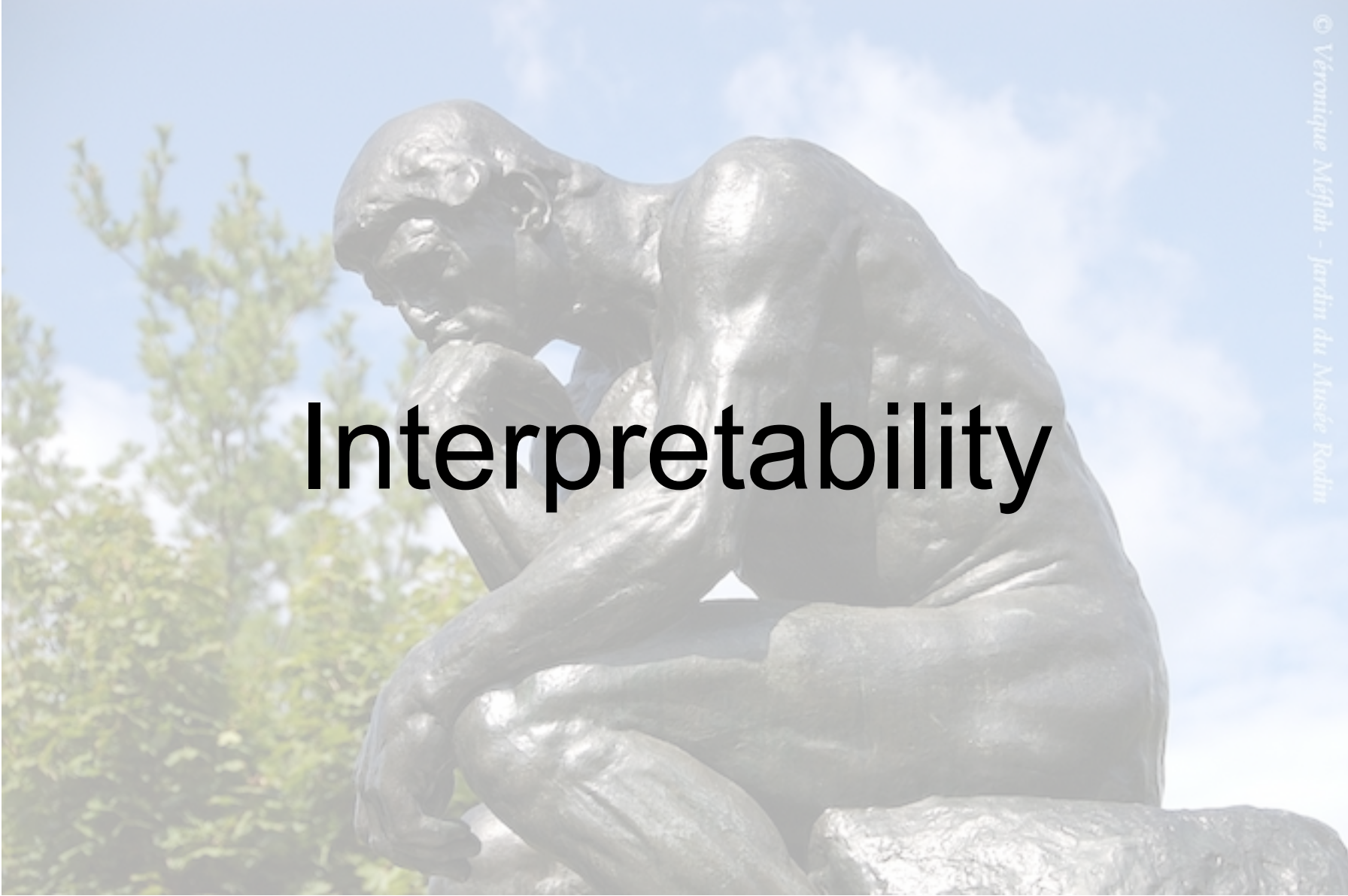


LLP jet tagger  
[\[doi:10.1088/2632-2153/ab9023\]](https://doi.org/10.1088/2632-2153/ab9023)



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



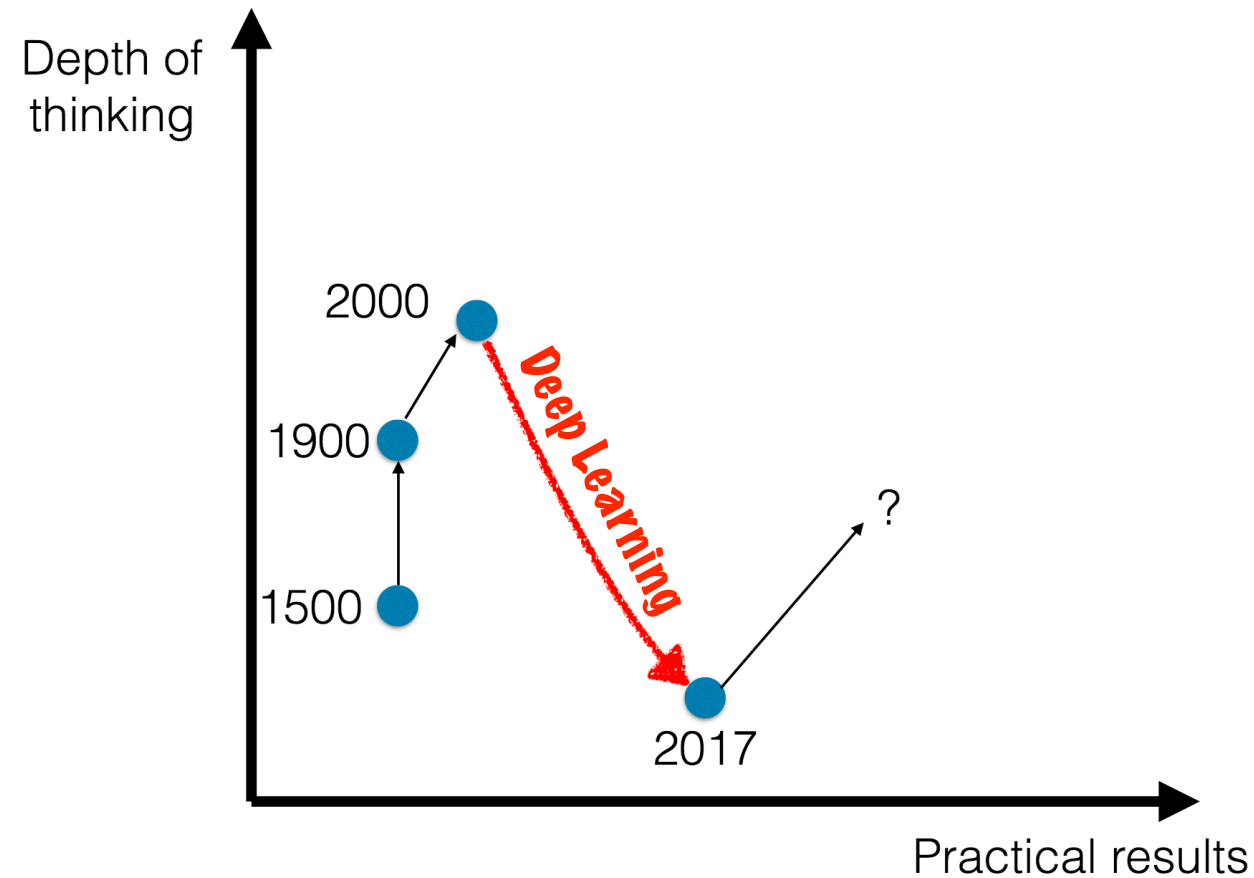


© Véronique Méflah - Jardin du Musée Rodin

# Interpretability



# 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.

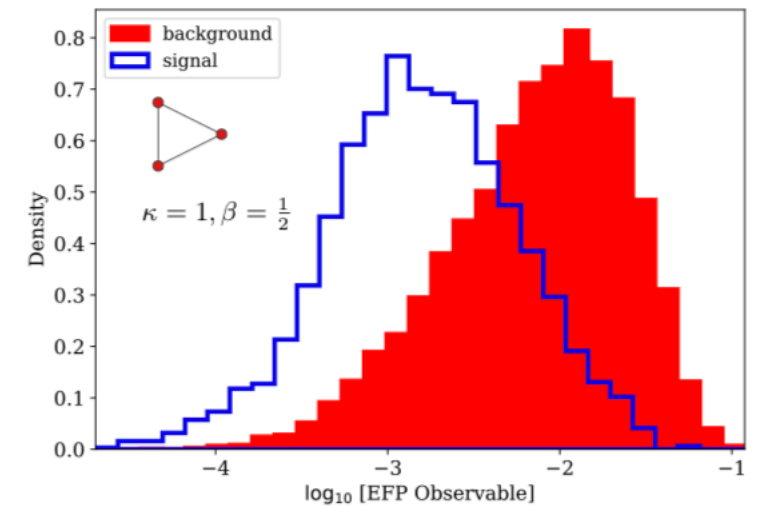
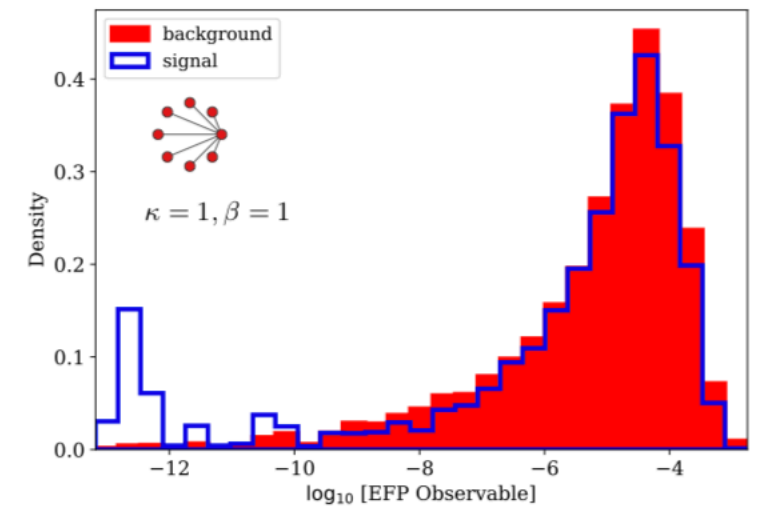
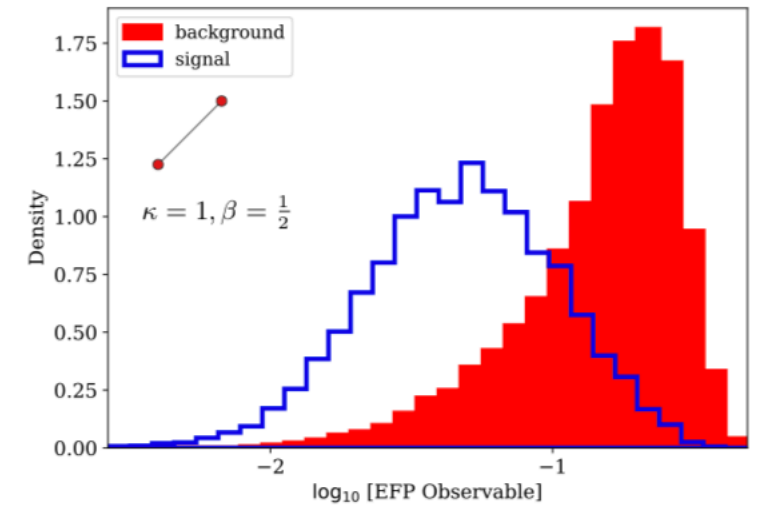
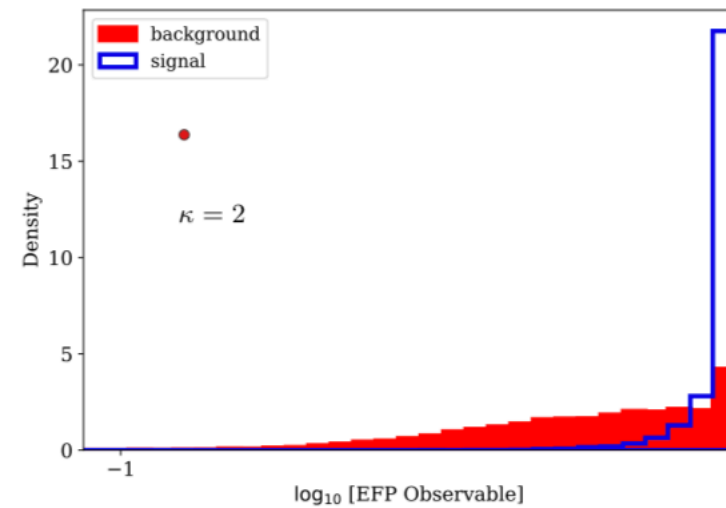
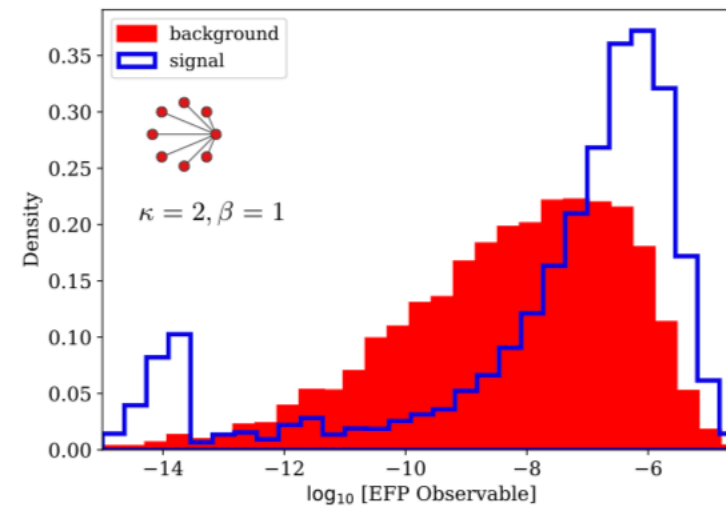


# Learning Observables

Electron classification performance

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

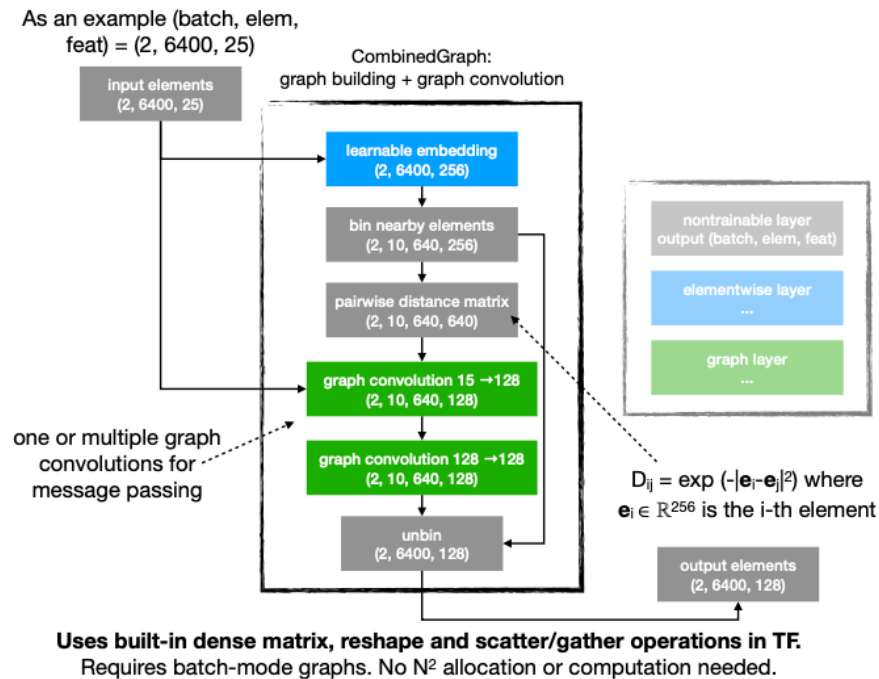
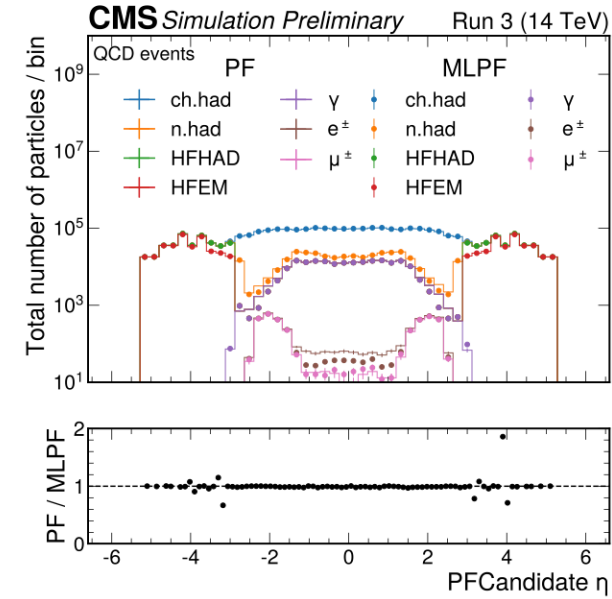
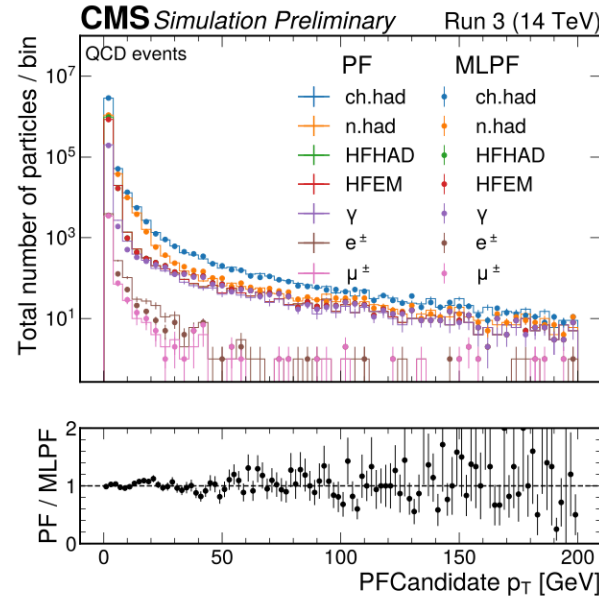
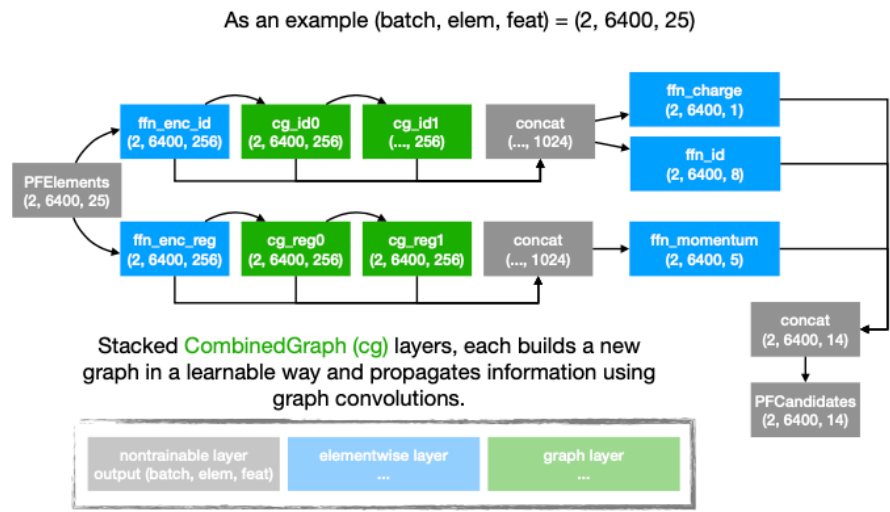
<https://arxiv.org/abs/2010.11998>  
<https://arxiv.org/abs/2011.01984>



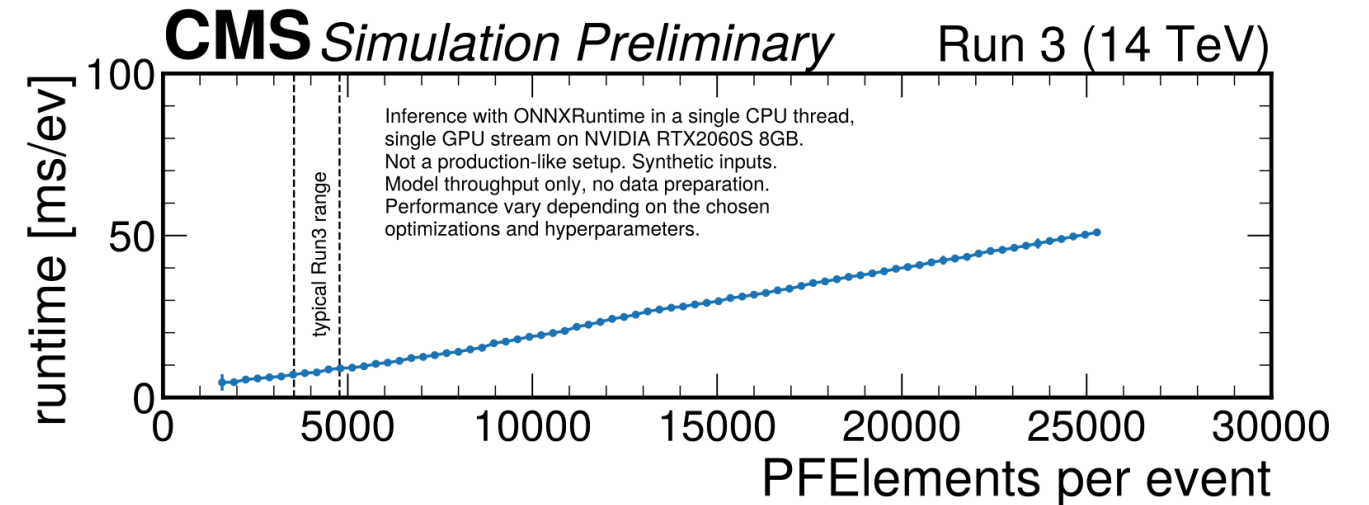
Search in the space of functions using decision ordering.  
 Simplified to the energy flow polynomial subspace.  
 Extract set of EFP that matches DNN performance.



# Particle-Flow Reconstruction



[2802826]



Set of tracks & clusters in input to graph-based model  
Classify sub-set of graph nodes as particle candidates

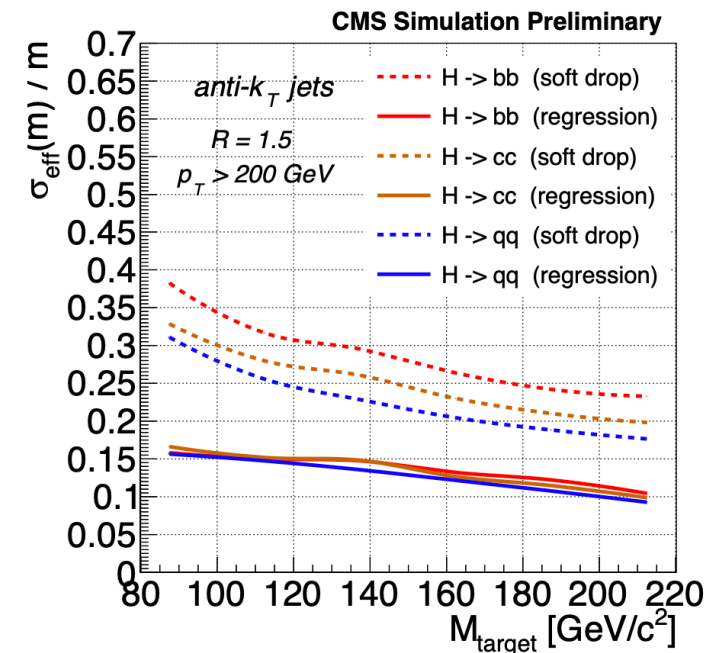
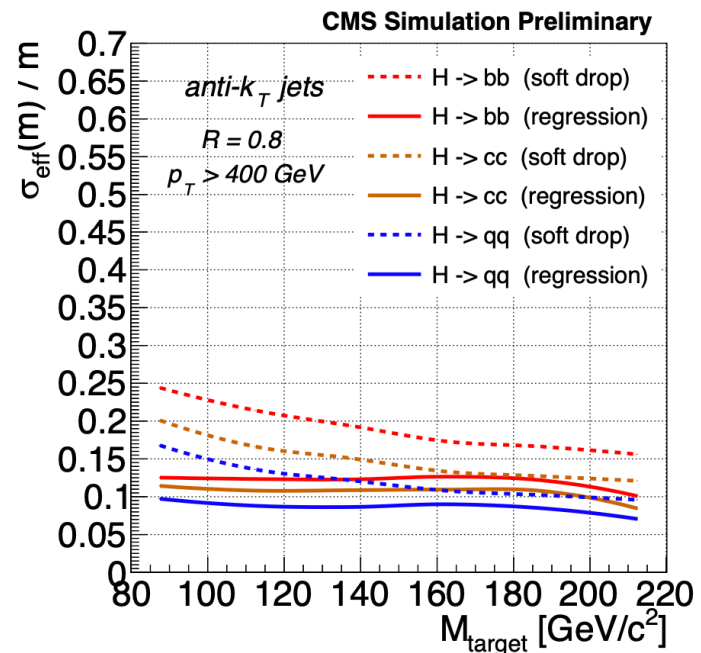
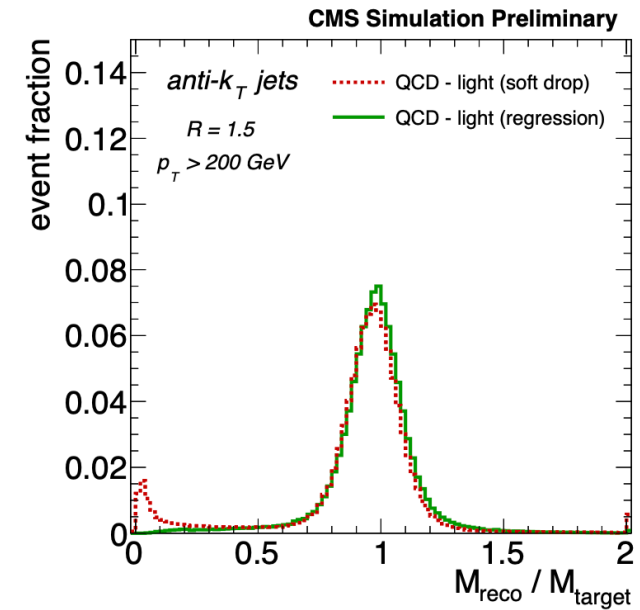
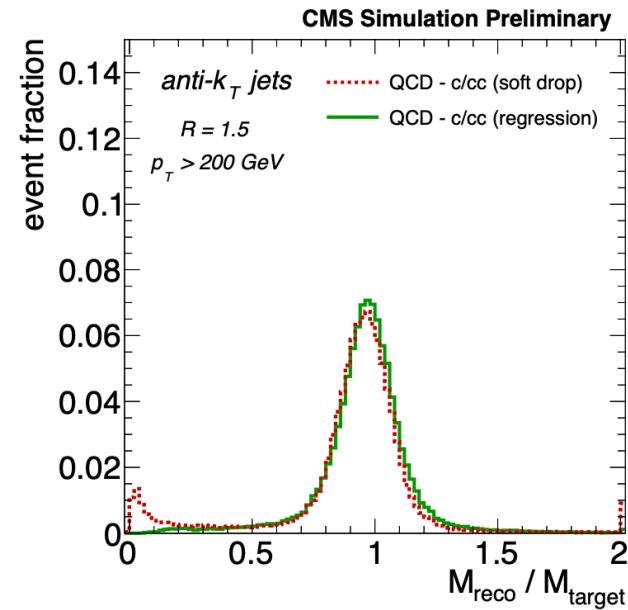
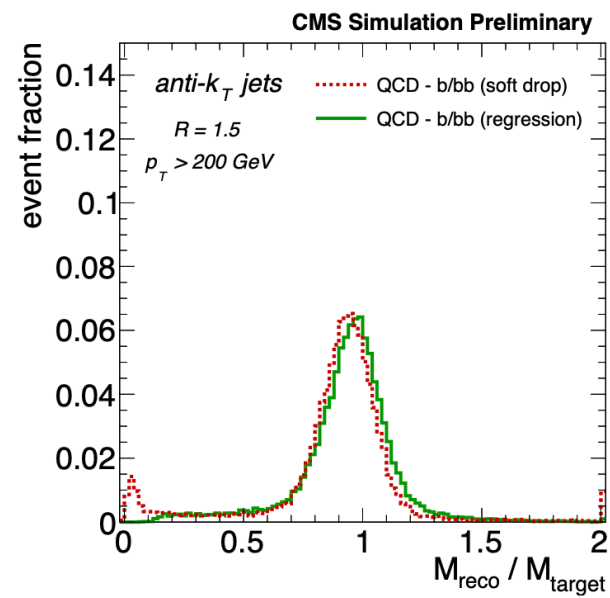
Regress parton kinematics from candidate

Model almost matching classical algorithm

Execution time quasi-linear with pile-up



# Jet Mass Regression

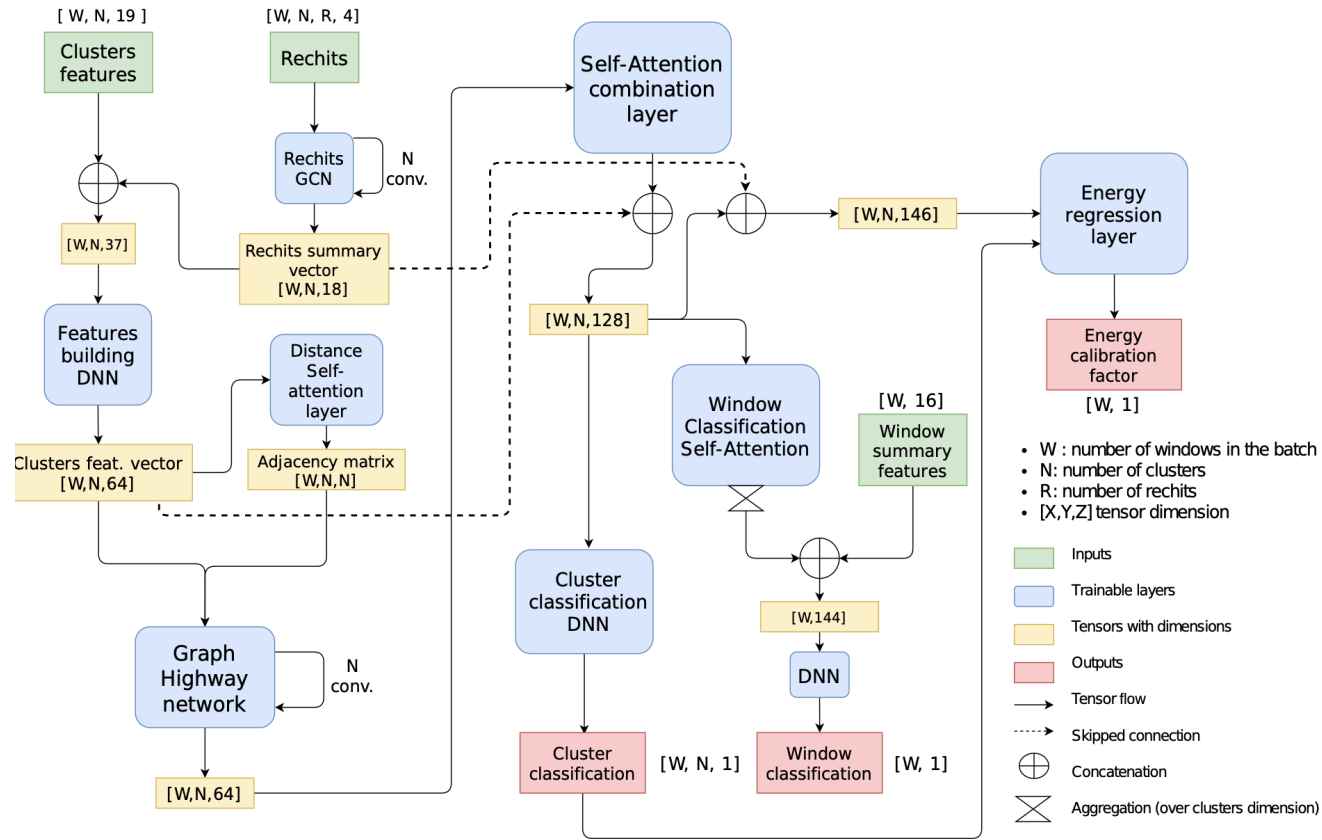


Jet mass regression using ParticleNet model, [\[2777006\]](#)

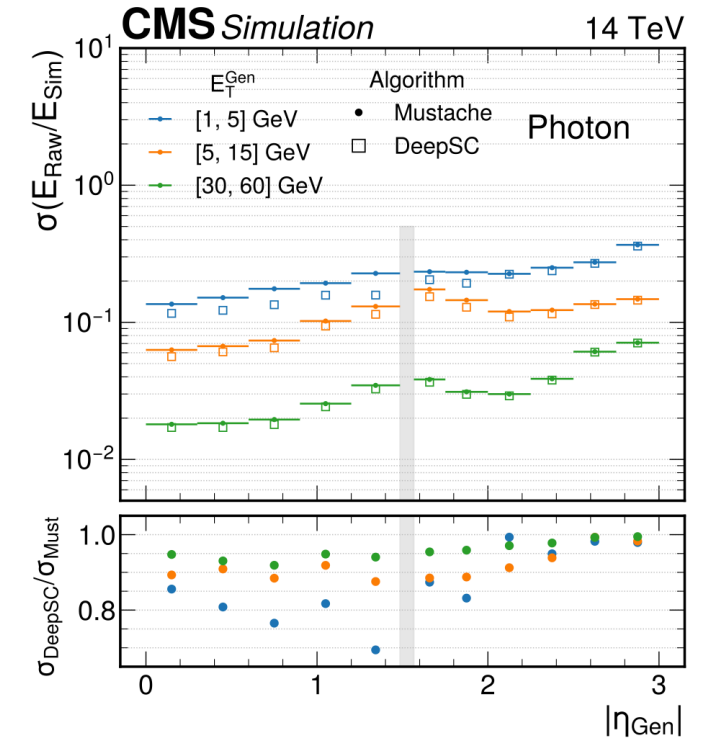
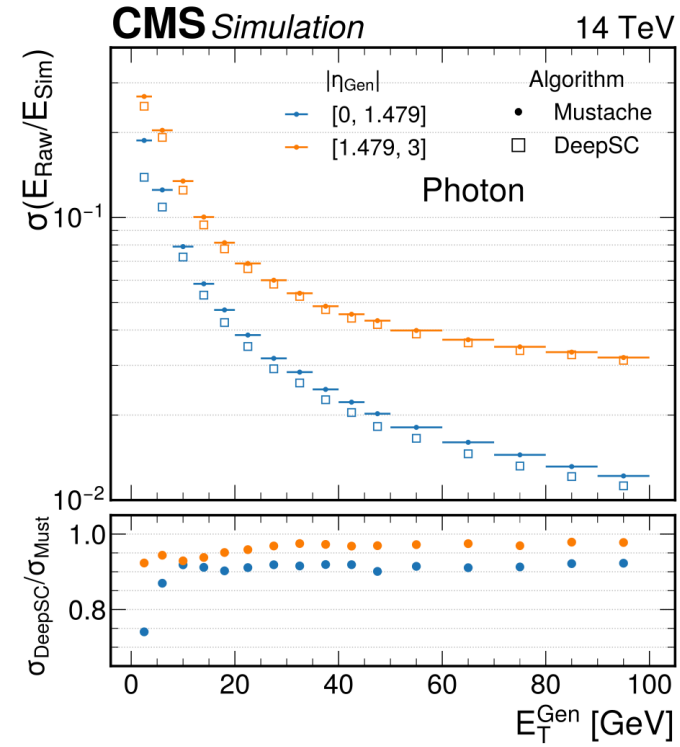




# Ecal Clustering/Regression



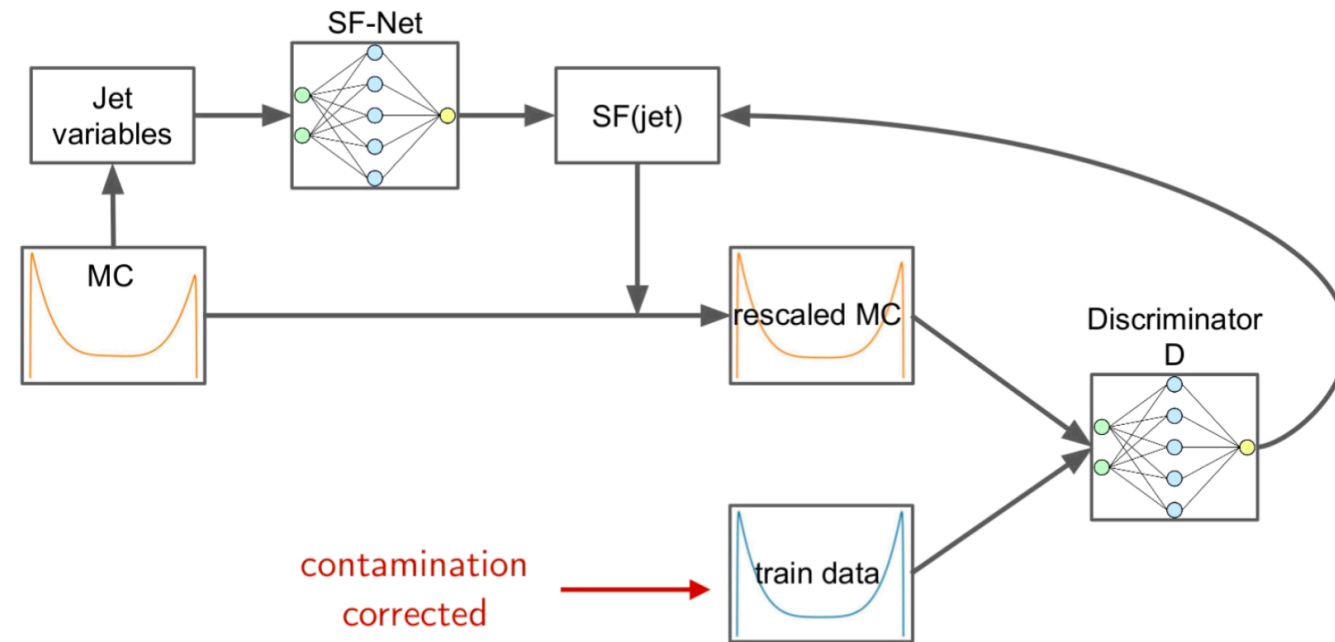
[2803235]



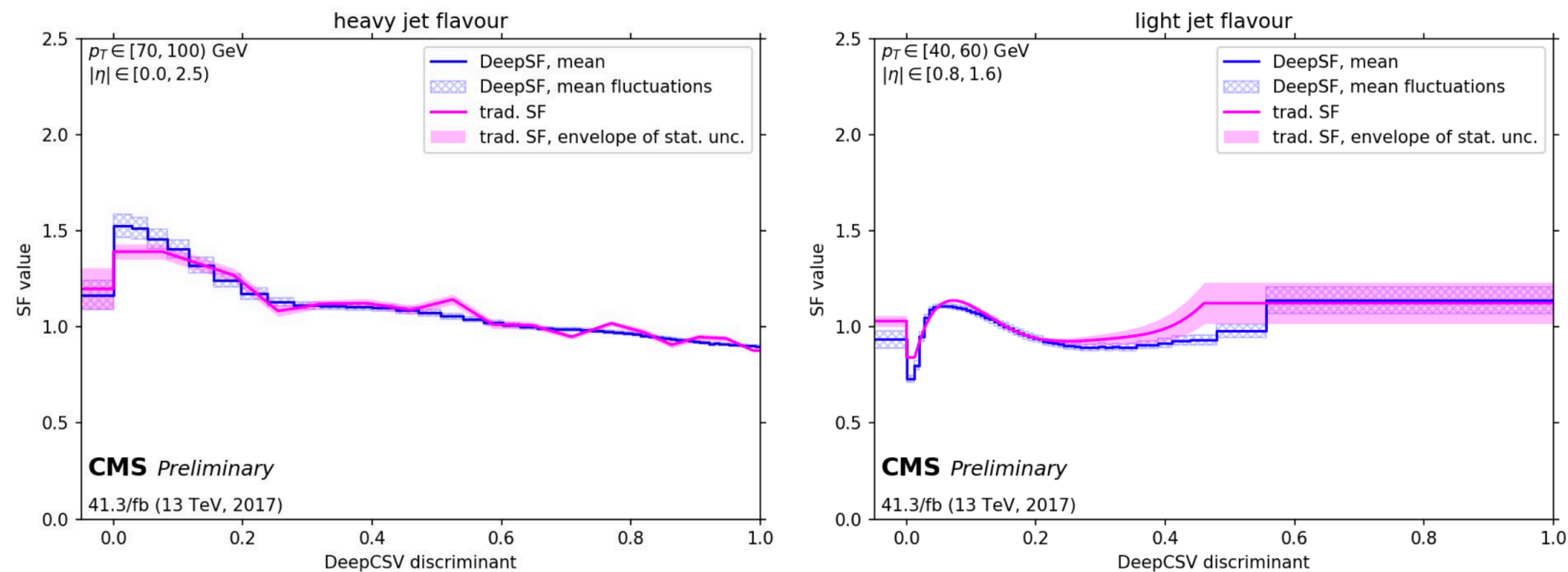
- Graph-based model with self-attention trained to :
  - ✓ seed-cluster classification
  - ✓ super-cluster classification
  - ✓ super-cluster energy regression
- Promising work in progress for calorimeter reconstruction



# Tagging Scale Factor



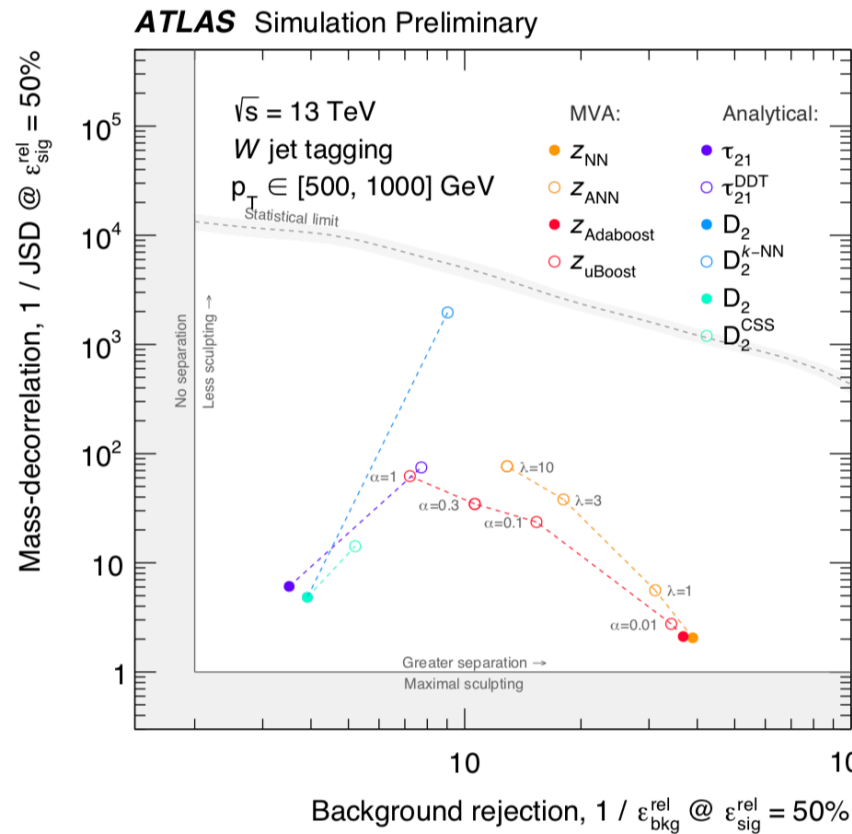
Learn per-jet data/MC scale factor using adversarial technique.



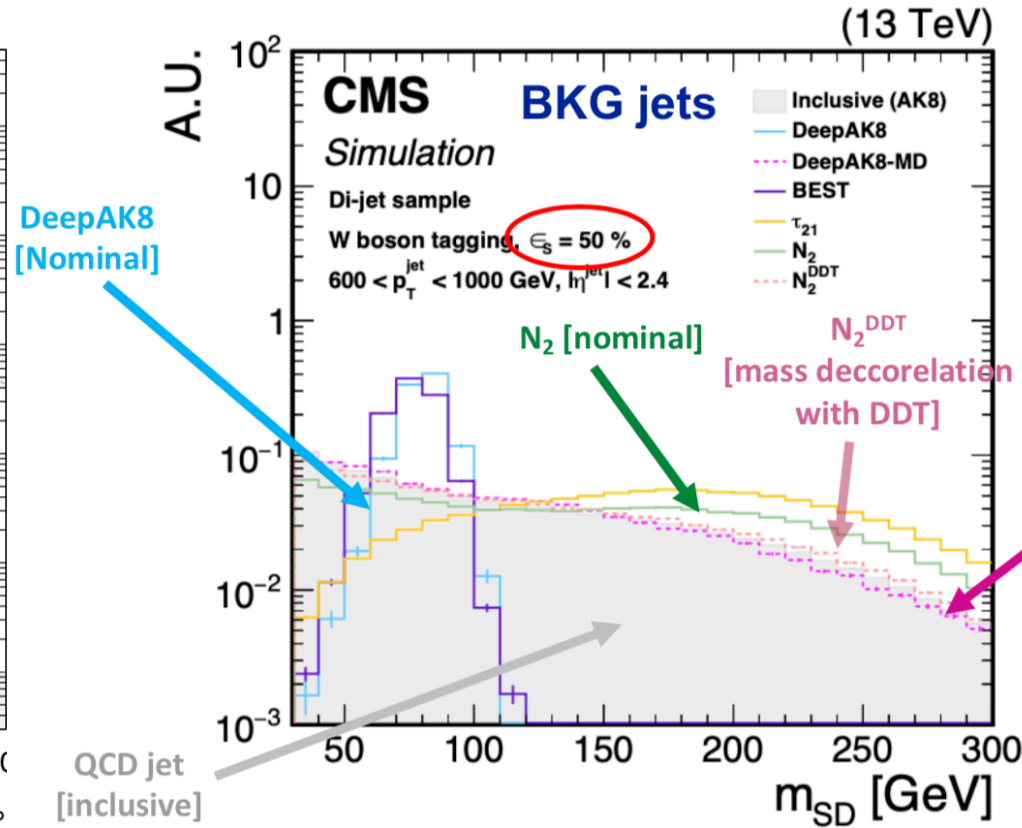
Adversarial Neural Network-based data-simulation corrections for heavy-flavor jet-tagging [\[cds:2666647\]](https://arxiv.org/abs/1708.02501)



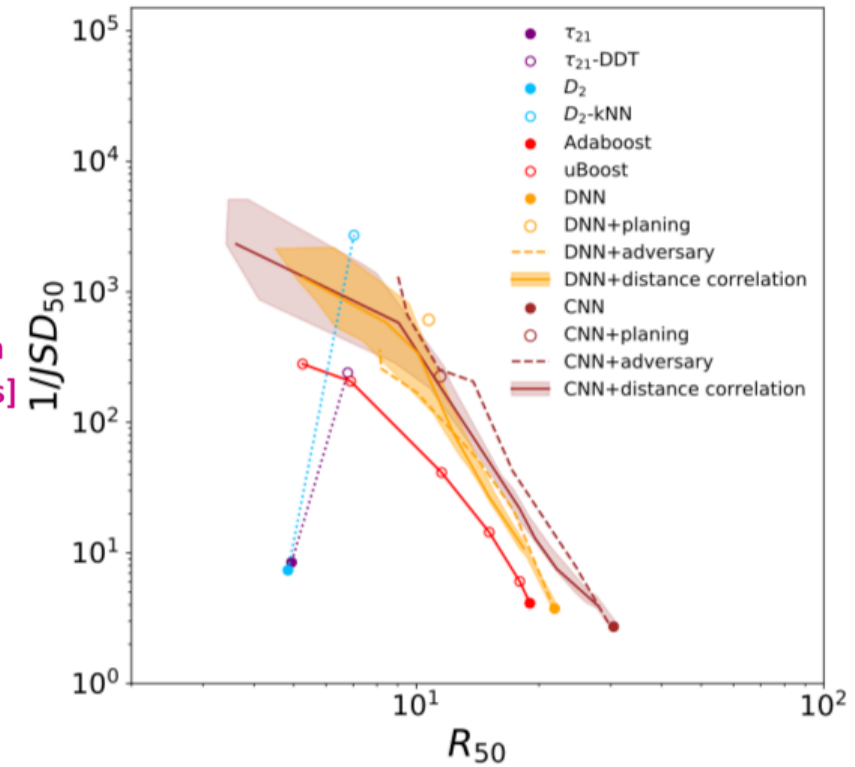
# Performance



ATLAS Collab. [\[cds:2630973\]](https://cds.cern.ch/record/2630973)



CMS Collab. [\[doi:10.1088/1748-0221/15/06/P06005\]](https://doi.org/10.1088/1748-0221/15/06/P06005)



DISCO: Distance Correlation [\[2001.05310\]](https://arxiv.org/abs/2001.05310)

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

