Deep Learning Particle Physics

Gregor Kasieczka

Email: gregor.kasieczka@uni-hamburg.de

Twitter: @GregorKasieczka

Mastodon: @gregor_k@sciencemastodon.com

ÖAW AI Winter School 2023

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

PUNCH 4NFDI

DASHH

Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CDCS
CENTER FOR DATA AND COMPUTING
IN NATURAL SCIENCES

PIER
Partnership of
Universität Hamburg and DESY

Bundesministerium
für Bildung
und Forschung

Overview

- Part I: Very brief introduction to particle physics
- Part II: Symmetries & how to treat them
- Part III: Generative models
- Part IV: Anomaly detection

Experimental particle physicist by training, do not expect formal proofs

Happy to take questions anytime, just raise your hands

Resources

REVIEWS

Check for updates

Machine learning in the search for new fundamental physics

Georgia Karagiorgi[1], Gregor Kasieczka[2], Scott Kravitz[3], Benjamin Nachman[3,4] and David Shih[5]

Abstract | Compelling experimental evidence suggests the existence of new physics beyond the well-established and tested standard model of particle physics. Various current and upcoming experiments are searching for signatures of new physics. Despite the variety of approaches and theoretical models tested in these experiments, what they all have in common is the very large volume of complex data that they produce. This data challenge calls for powerful statistical methods. Machine learning has been in use in high-energy particle physics for well over a decade, but the rise of deep learning in the early 2010s has yielded a qualitative shift in terms of the scope and ambition of research. These modern machine learning developments are the focus of the present Review, which discusses methods and applications for new physics searches in the context of terrestrial high-energy physics experiments, including the Large Hadron Collider, rare event searches and neutrino experiments.

For several decades, the standard model (SM) of particle physics has provided a clear theoretical guide to experiments, resulting in an extensive search programme that culminated with the discovery of the Higgs boson[1,2]. Although the SM is now complete, there are key experimental observations that compel the community to expand the search efforts for new particles and forces of nature beyond the SM (BSM). For example, the existence of dark matter (DM) and dark energy is well established[3], as are the mass of neutrinos[4,5] and the baryon–antibaryon asymmetry in the Universe[6] — yet none of these observations are explained by the SM. Additionally, 'aesthetic' problems plague the SM, including the unexplained weak-scale mass of the Higgs boson, the existence of three generations of fermions, and the minuteness of the neutron dipole moment[7]. Current and near-future high-energy physics (HEP) experiments have the potential to shed light on all of these fundamental challenges by creating new particles in the laboratory, or by observing interactions of new particles with normal matter or with other new particles.

This great potential for discovery comes with considerable data challenges. New particle interactions are expected to be rare, and their signature could be only subtly different from the SM. This means that researchers must collect and sift through an immense amount of complex data to isolate potential BSM physics. Machine learning (ML) offers a powerful solution to this challenge. Deep learning techniques (used here to mean modern ML, with deep neural networks (NNs) and other advanced tools that contain (much) more than tens of thousands of tunable parameters) are well suited for analysing large amounts of data in many dimensions to find subtle patterns. Multivariate analysis has been commonplace in HEP for decades (for example, the TMVA 'toolkit')[8], but the latest tools will qualitatively extend the sensitivity to 'hypervariate analysis' whereby the entire phase space of available experimental information can be analysed holistically. These new tools also allow for new analysis strategies independent of the dimensionality (density estimation, variable-length inputs and so on).

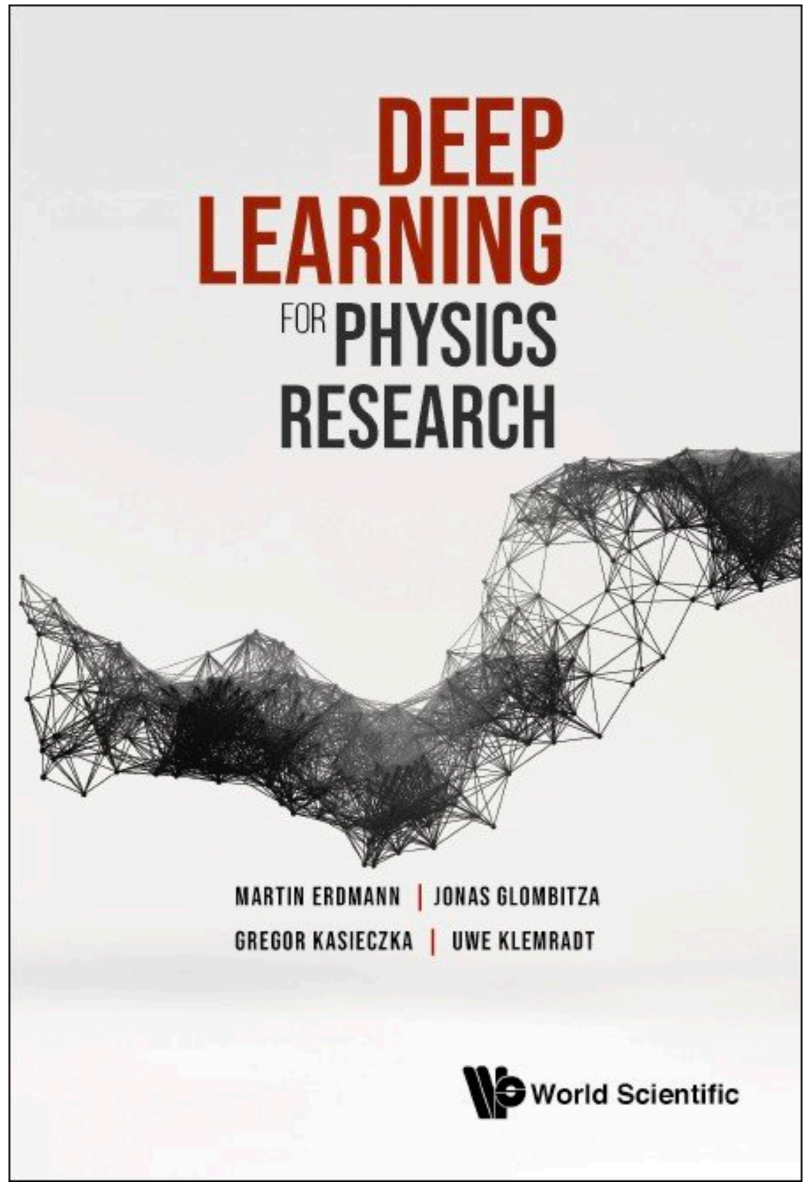
In tandem with the growing data volume, a related challenge is the increasing need for efficient (in terms of computational time, power and resource utilization) and accurate data processing for high-throughput applications. Efforts to that end include the development and acceleration of deep learning-based processing algorithms on power-efficient hardware platforms[9].

In addition to the growing data challenge, there is also the compounding challenge of simulating expectations for what experiments may observe. HEP experiments rely heavily on simulations for all aspects of research, from experimental design all the way to data analysis. Built on a thorough understanding of the SM and the fundamental laws of nature, these simulations are extremely comprehensive and sophisticated, but they are still only an approximation to nature. It is therefore often necessary to combine simulations with information directly from data to improve simulation accuracy. The corresponding ML models must be robust against inaccuracies and be able to integrate uncertainties.

[1]Department of Physics, Columbia University, New York, NY, USA.
[2]Institut für Experimentalphysik, Universität Hamburg, Hamburg, Germany.
[3]Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[4]Berkeley Institute for Data Science, University of California, Berkeley, CA, USA.
[5]NHETC, Department of Physics and Astronomy, Rutgers University, Piscataway, NJ, USA.
e-mail: georgia@nevis.columbia.edu; gregor.kasieczka@uni-hamburg.de; swkravitz@lbl.gov; bpnachman@lbl.gov; shih@physics.rutgers.edu

https://doi.org/10.1038/s42254-022-00455-1

NATURE REVIEWS | PHYSICS   VOLUME 4 | JUNE 2022 | 399



HEPML-LivingReview

A Living Review of Machine Learning for Particle Physics

Modern machine learning techniques, including deep learning, is rapidly being applied, adapted, and developed for high energy physics. The goal of this document is to provide a nearly comprehensive list of citations for those developing and applying these approaches to experimental, phenomenological, or theoretical analyses. As a living document, it will be updated as often as possible to incorporate the latest developments. A list of proper (unchanging) reviews can be found within. Papers are grouped into a small set of topics to be as useful as possible. Suggestions are most welcome.

download review

The purpose of this note is to collect references for modern machine learning as applied to particle physics. A minimal number of categories is chosen in order to be as useful as possible. Note that papers may be referenced in more than one category. The fact that a paper is listed in this document does not endorse or validate its content - that is for the community (and for peer-review) to decide. Furthermore, the classification here is a best attempt and may have flaws - please let us know if (a) we have missed a paper you think should be included, (b) a paper has been misclassified, or (c) a citation for a paper is not correct or if the journal information is now available. In order to be as useful as possible, this document will continue to evolve so please check back before you write your next paper. If you find this review helpful, please consider citing it using \cite{hepmllivingreview} in HEPML.bib.

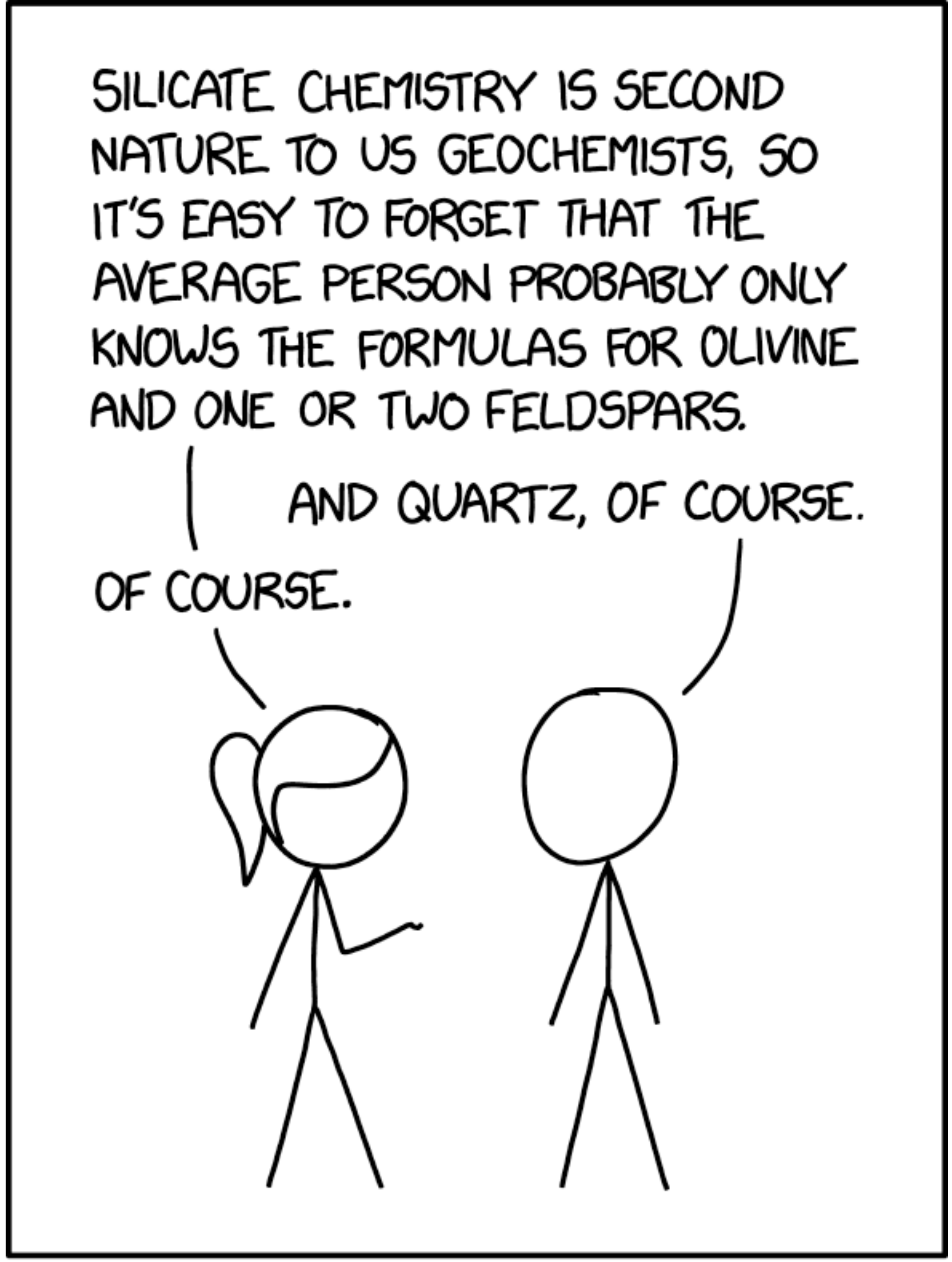
- Reviews
  - Modern reviews
    - Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning [DOI]
    - Deep Learning and its Application to LHC Physics [DOI]
    - Machine Learning in High Energy Physics Community White Paper [DOI]
    - Machine learning at the energy and intensity frontiers of particle physics

https://arxiv.org/abs/2112.03769

https://www.worldscientific.com/worldscibooks/10.1142/12294#t=aboutBook

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

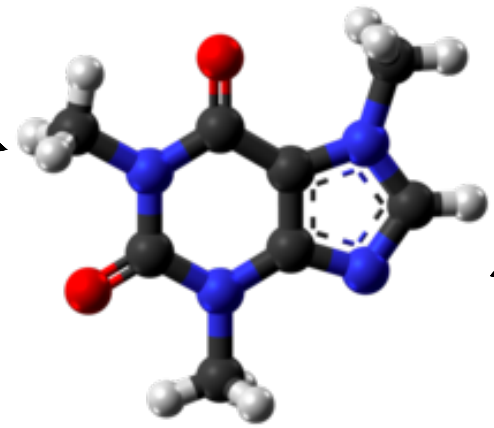
Part I: Very brief introduction to particle physics



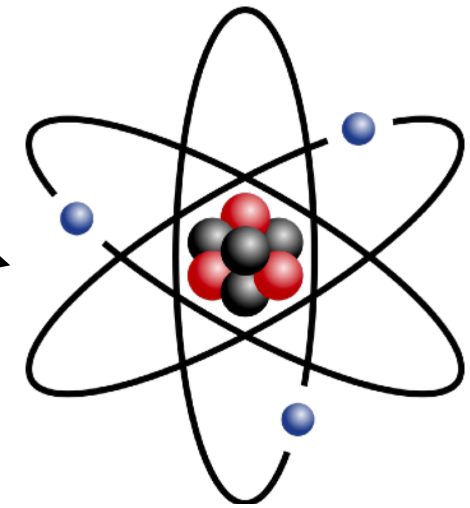
EVEN WHEN THEY'RE TRYING TO COMPENSATE FOR IT, EXPERTS IN ANYTHING WILDLY OVERESTIMATE THE AVERAGE PERSON'S FAMILIARITY WITH THEIR FIELD.

Setting the stage

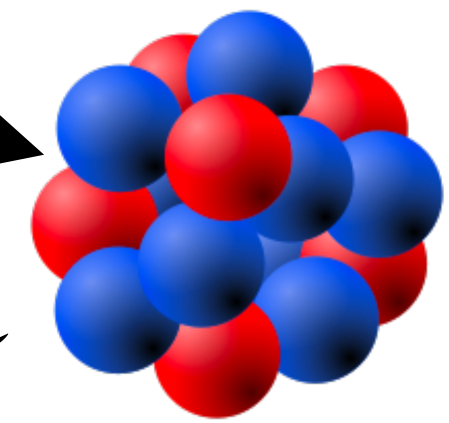
Daily Life: 10^{-3} - 10^{3} m



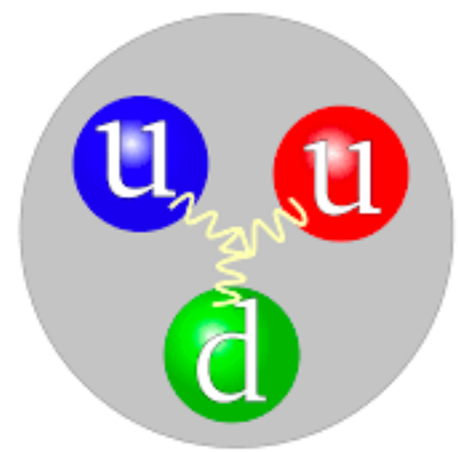
Molecule: 10^{-9} - 10^{-10} m



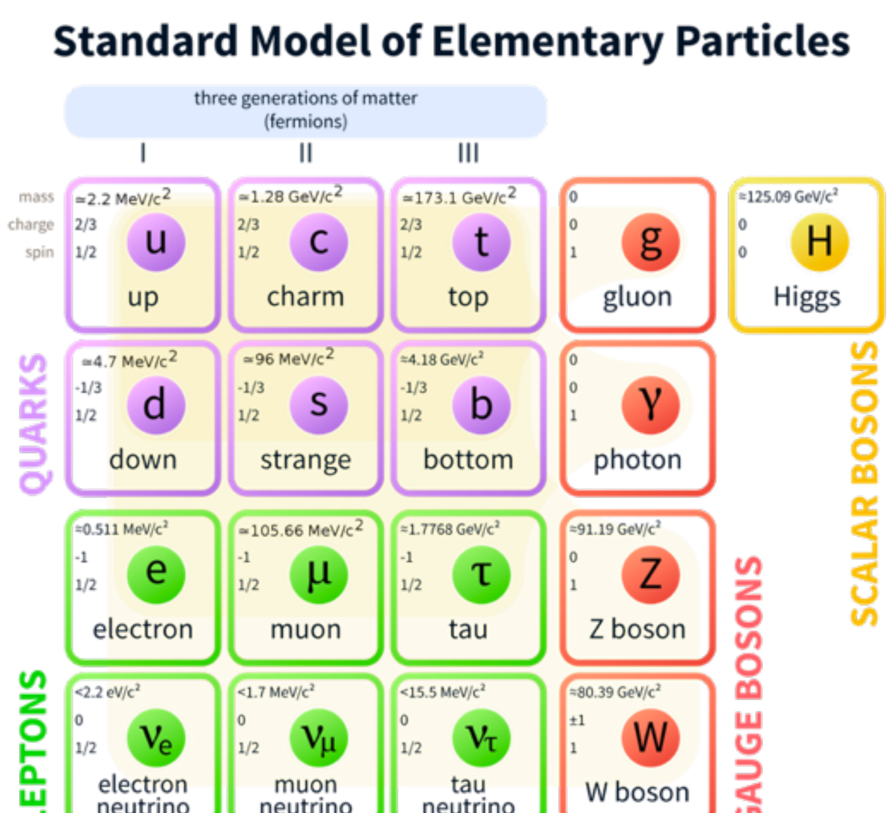
Atom: 10^{-10} m



Nucleus: 10^{-14} m

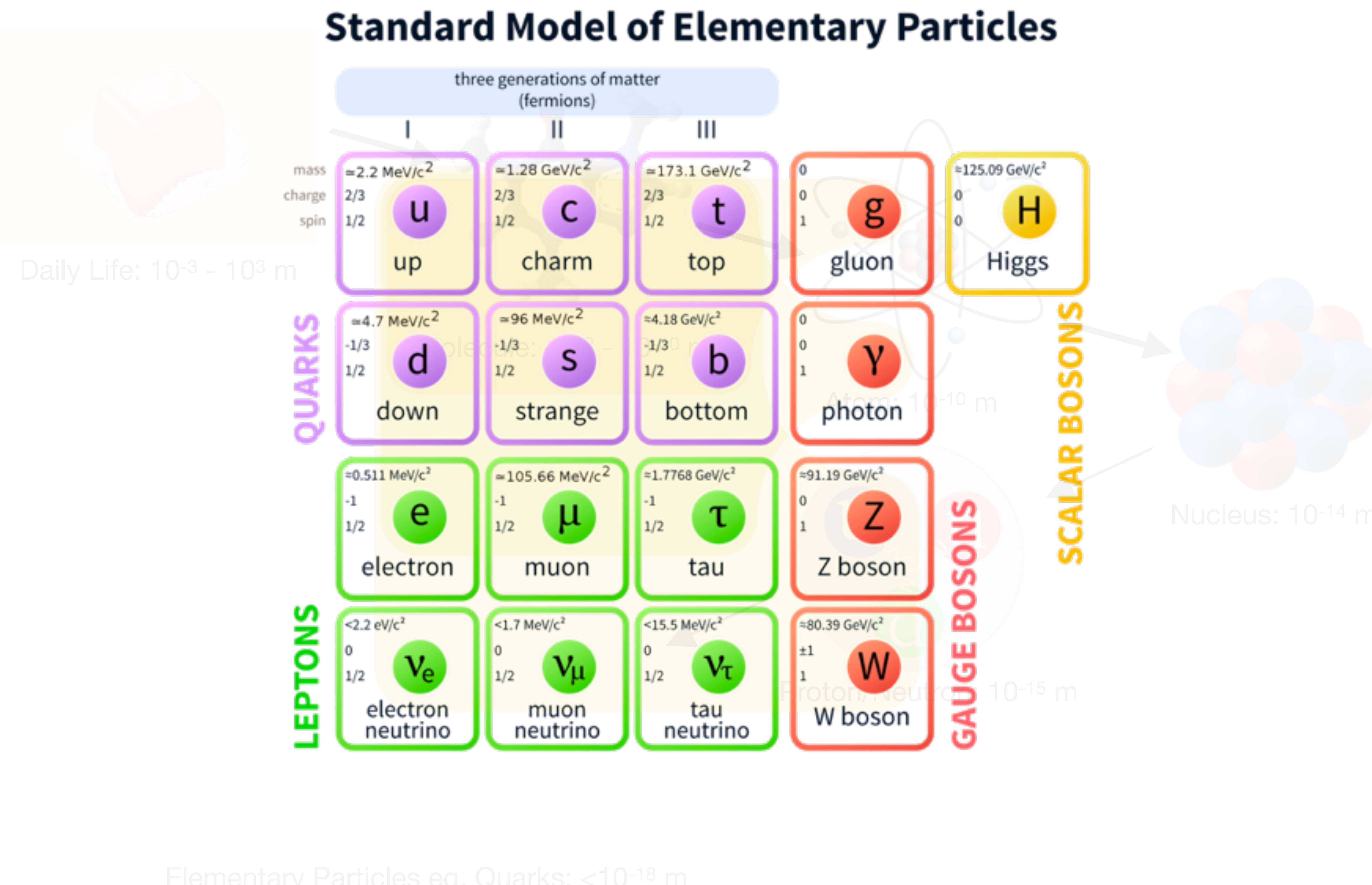


Proton/Neutron: 10^{-15} m



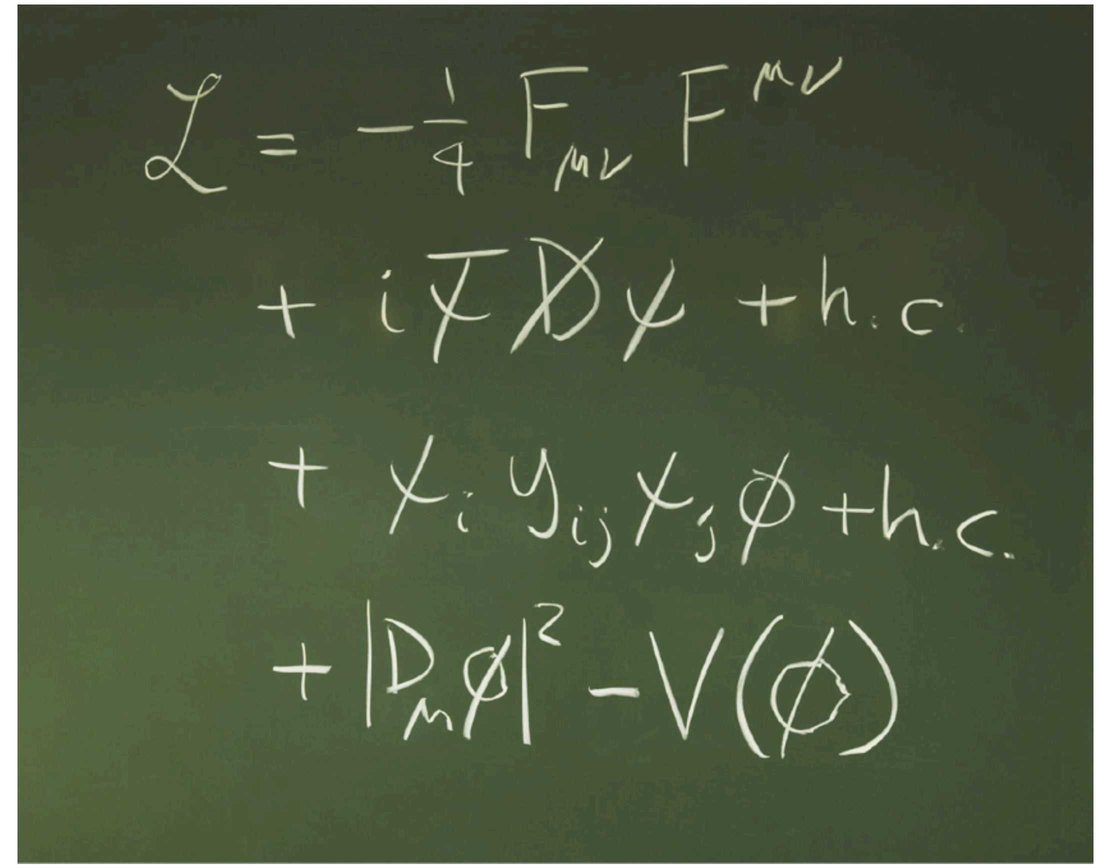
Elementary Particles eg. Quarks: <10^{-18} m

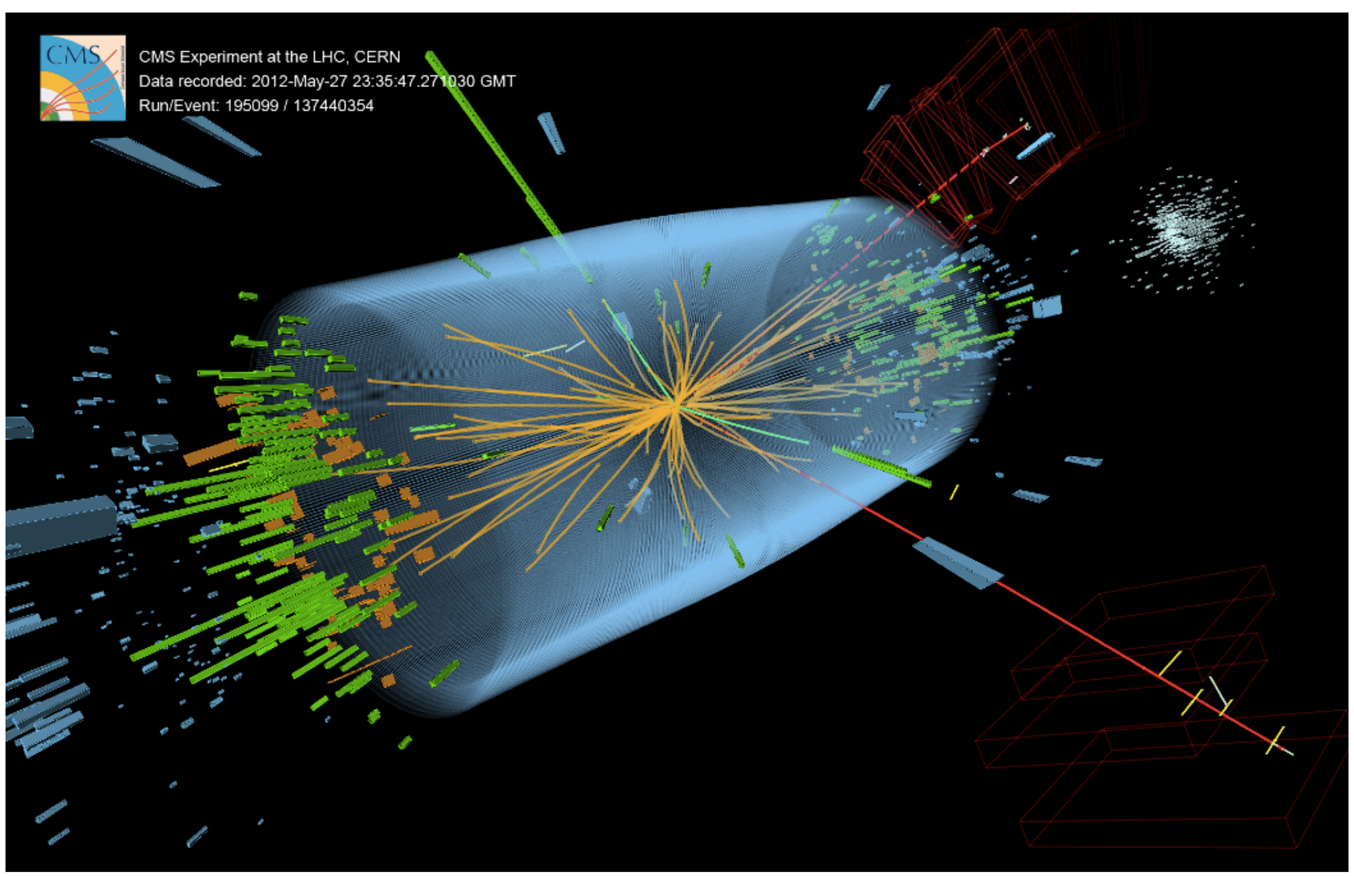
Setting the stage



Particle Physics

- Particle physicists study these smallest constituents of matter
- The Standard Model is an incredible scientific achievement and describes three of four fundamental forces
- Mathematical, quantum theoretical understanding of matter at the smallest scales



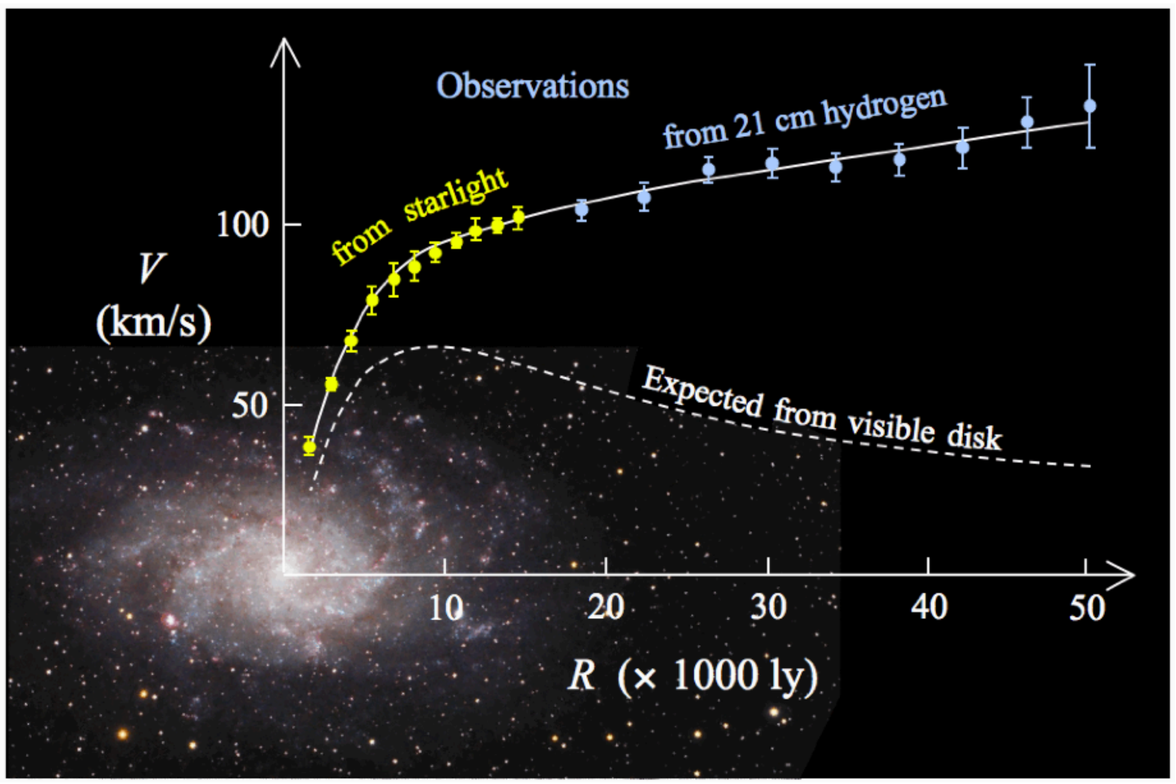


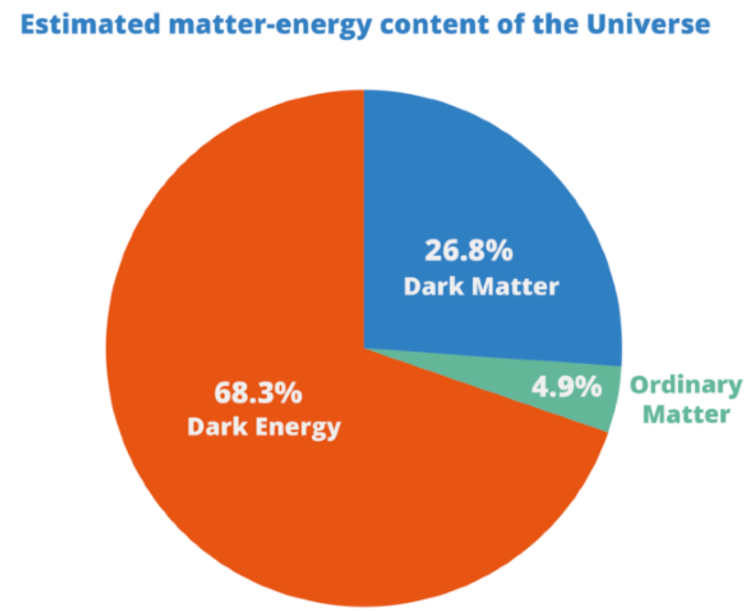
2012: Higgs Boson Discovery

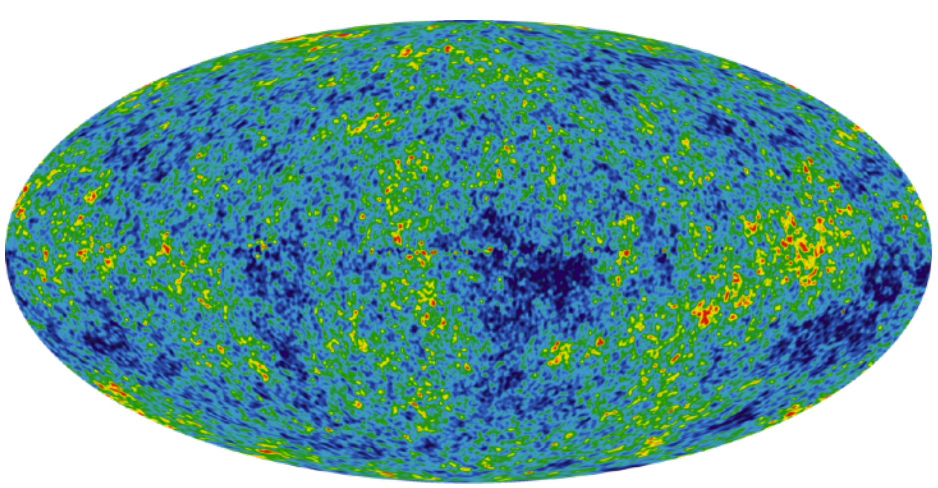
2013 Nobel Prize in Physics to Peter Higgs and François Englert "*for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles*"

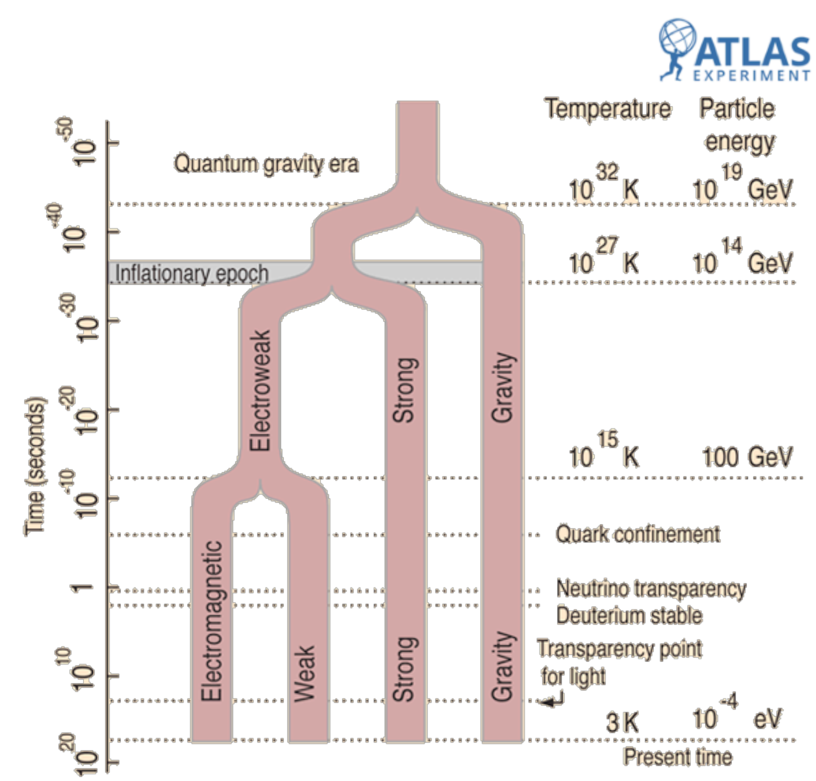
Open Questions

- The Standard Model cannot be the ultimate theory of Nature
- Both experimental and theoretical evidence
- Example: We know there must be a type of particle called dark matter; but we don't know what it is



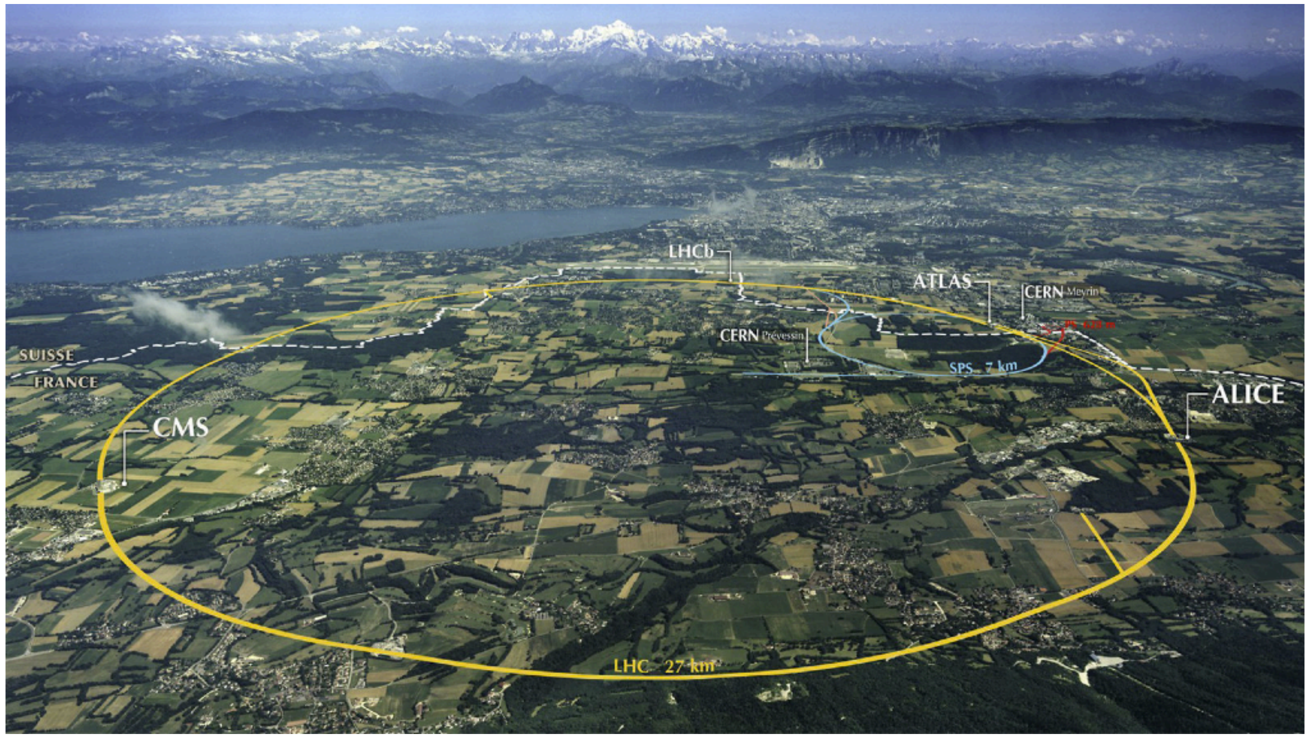




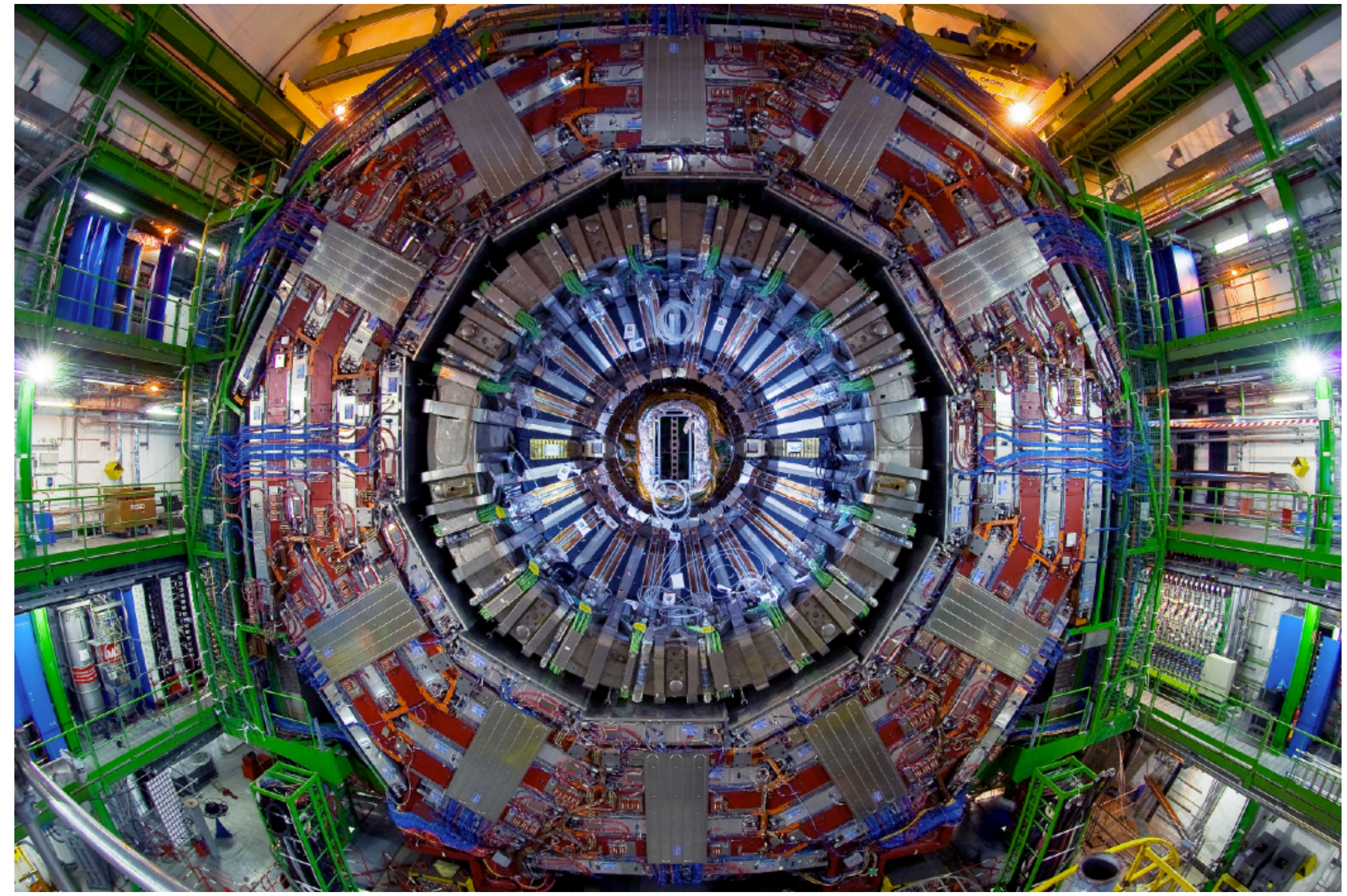


Current Experiments

Current Experiments



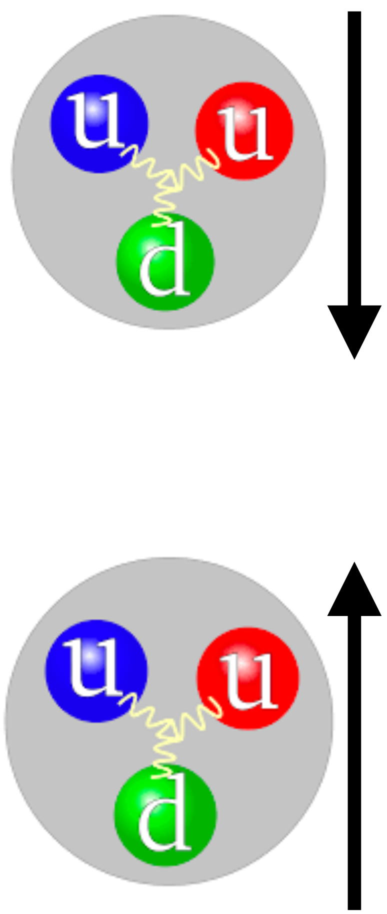
Aerial view of LHC (Geneva area)



CMS Experiment

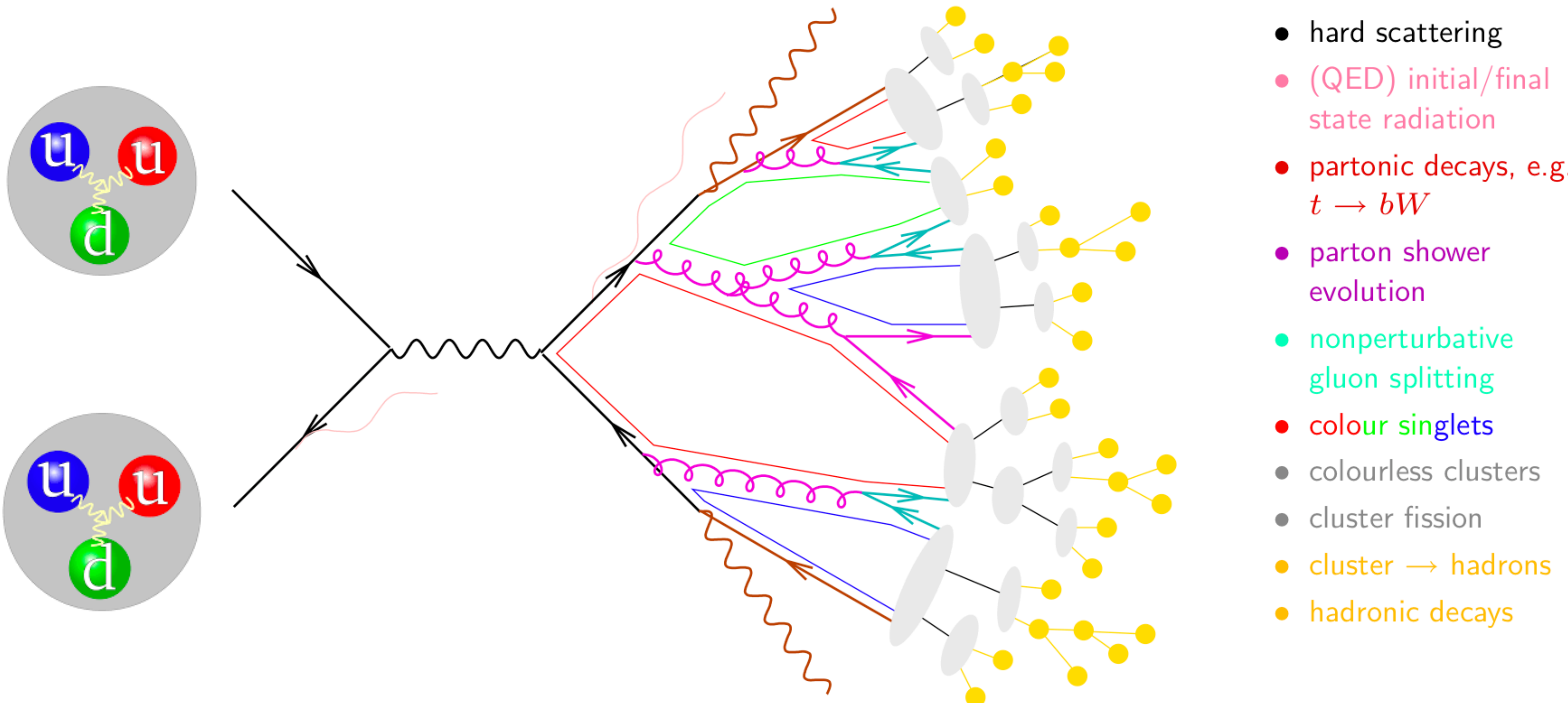
- Large Hadron Collider (LHC) at CERN laboratory 27 km circumference
- Collide pairs of protons with a centre-of-mass energy of 13 TeV (99.999999% of speed of light)
- 4 large experiments (ATLAS, CMS, LHCb, ALICE)
- 40 Million collisions/second / experiment
- ~25 Petabyte collision data/year / experiment

Particle Collisions



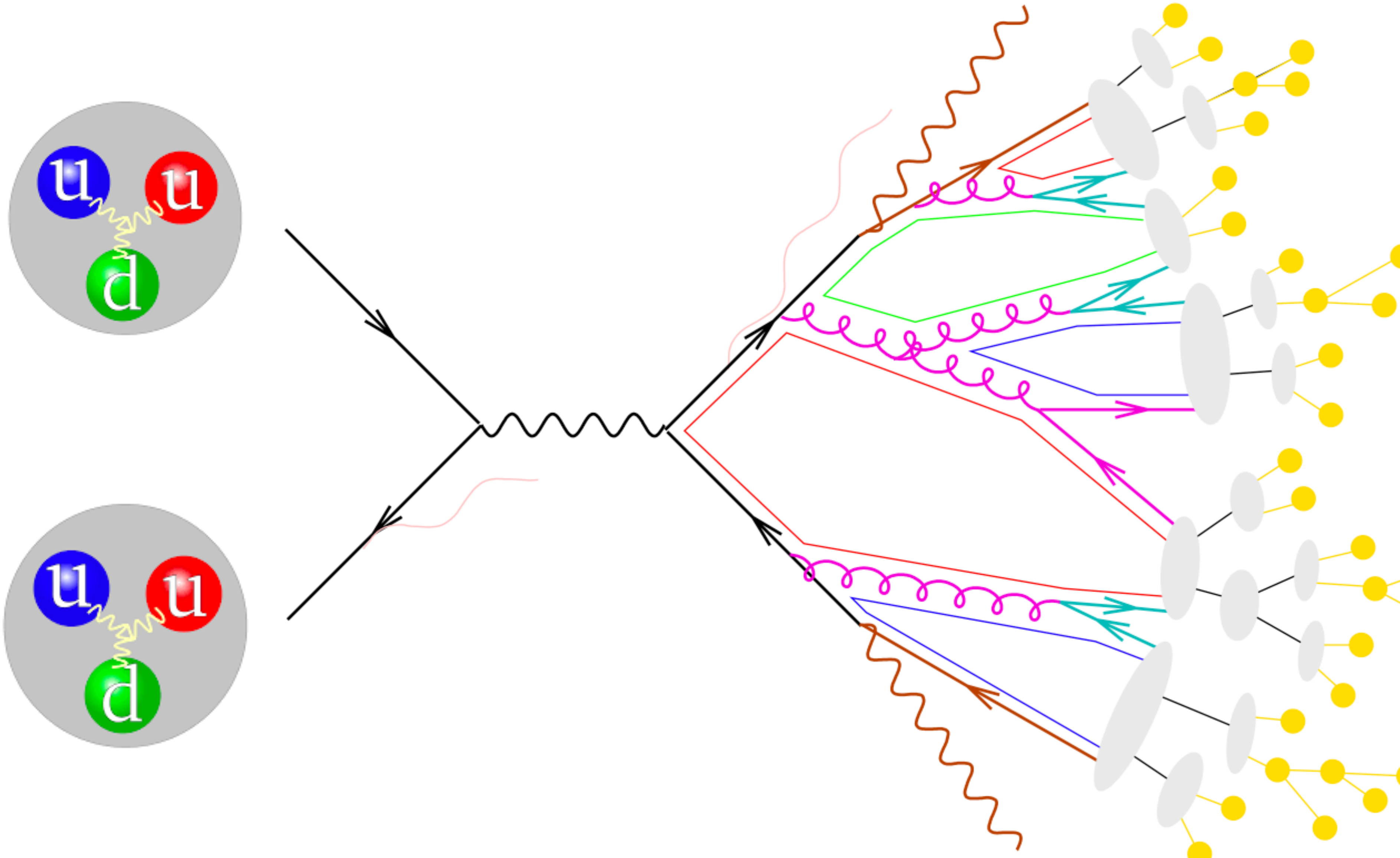
Particle Collisions

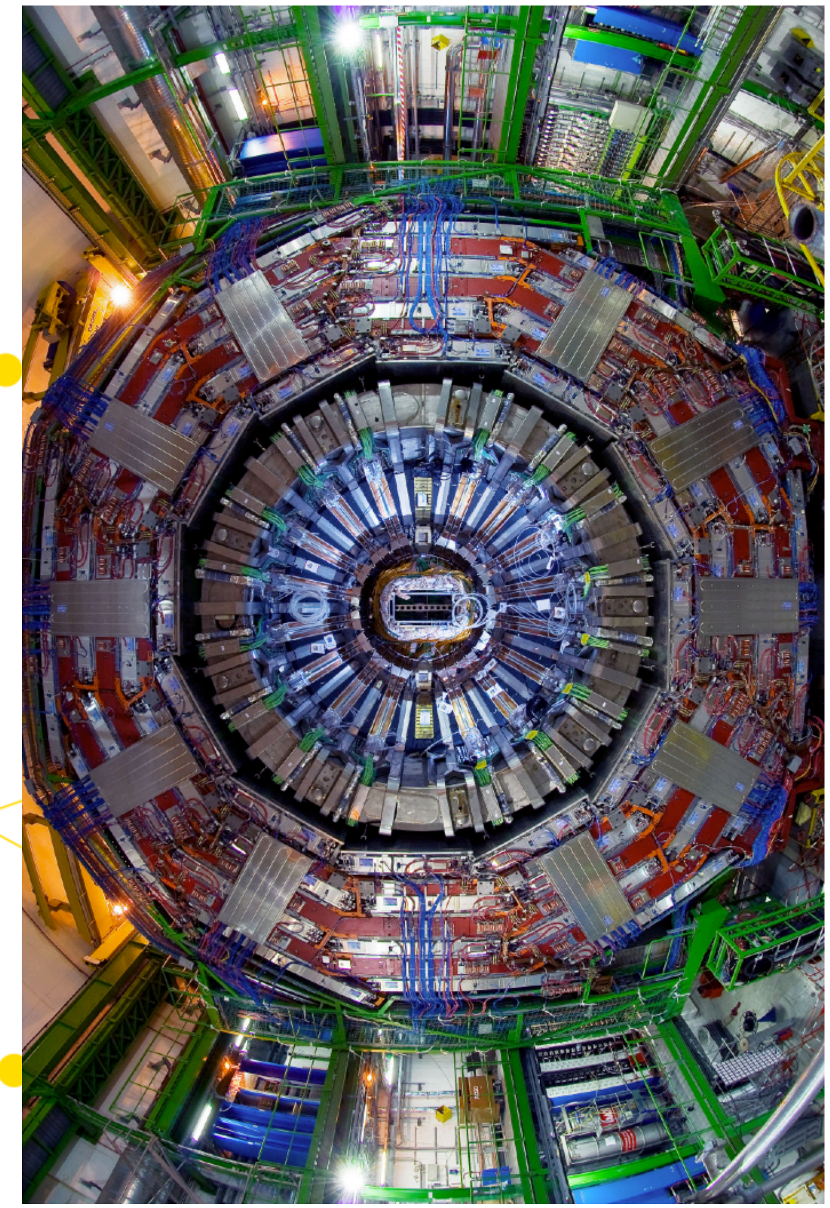
Particle Collisions



- Collisions turns kinetic energy into new particles ($E=mc^2$)
- Stochastic process, no control over *which* particles get produced
- Very short lived (e.g. 10^{-25} s for the top quark)
- Chain of particle decays

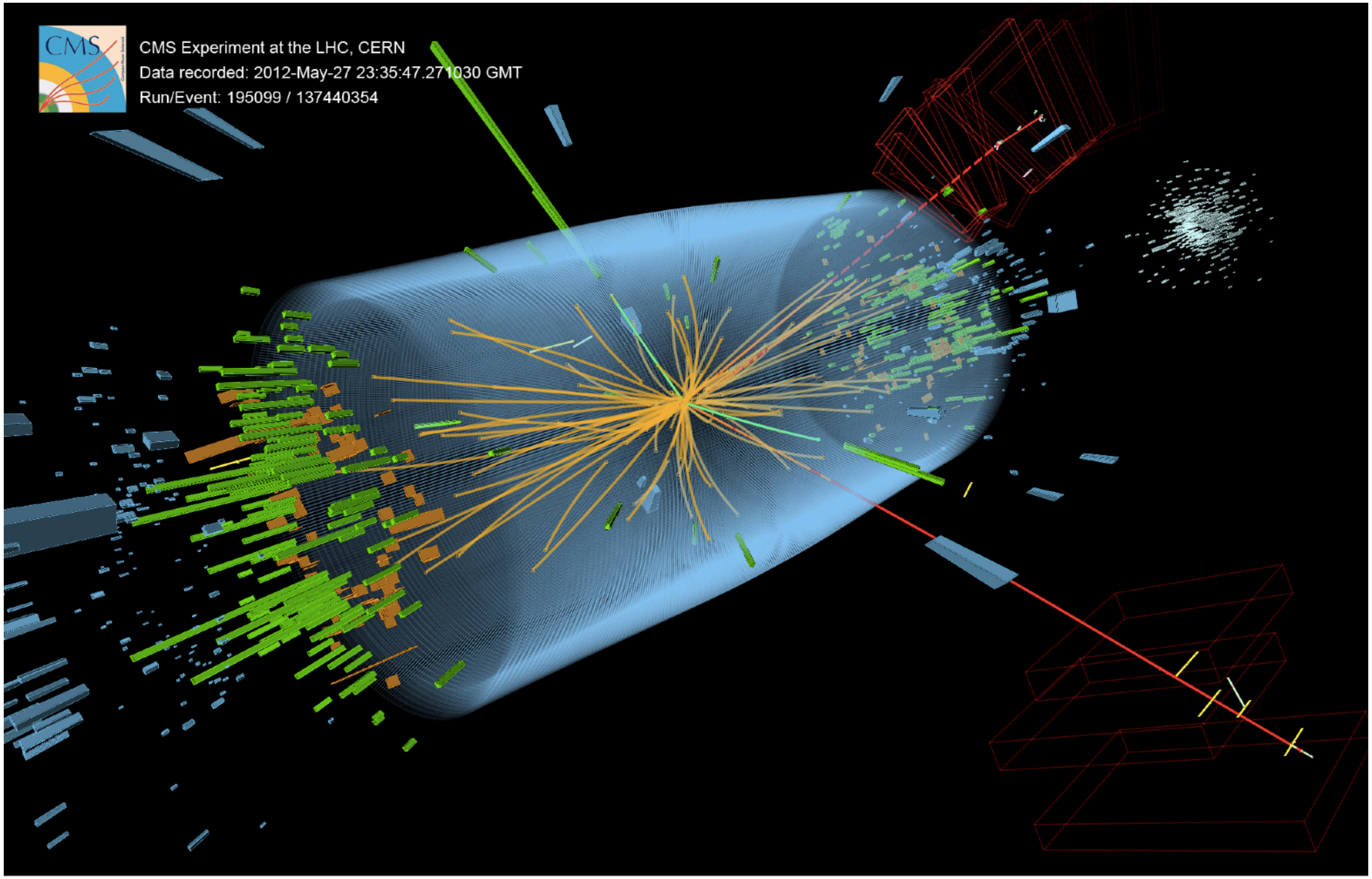
Particle Collisions





- Collisions turns kinetic energy into new particles ($E=mc^2$)
- Stochastic process, no control over *which* particles get produced
- Very short lived (e.g. 10^{-25} s for the top quark)
- Chain of particle decays, final states measures by detectors

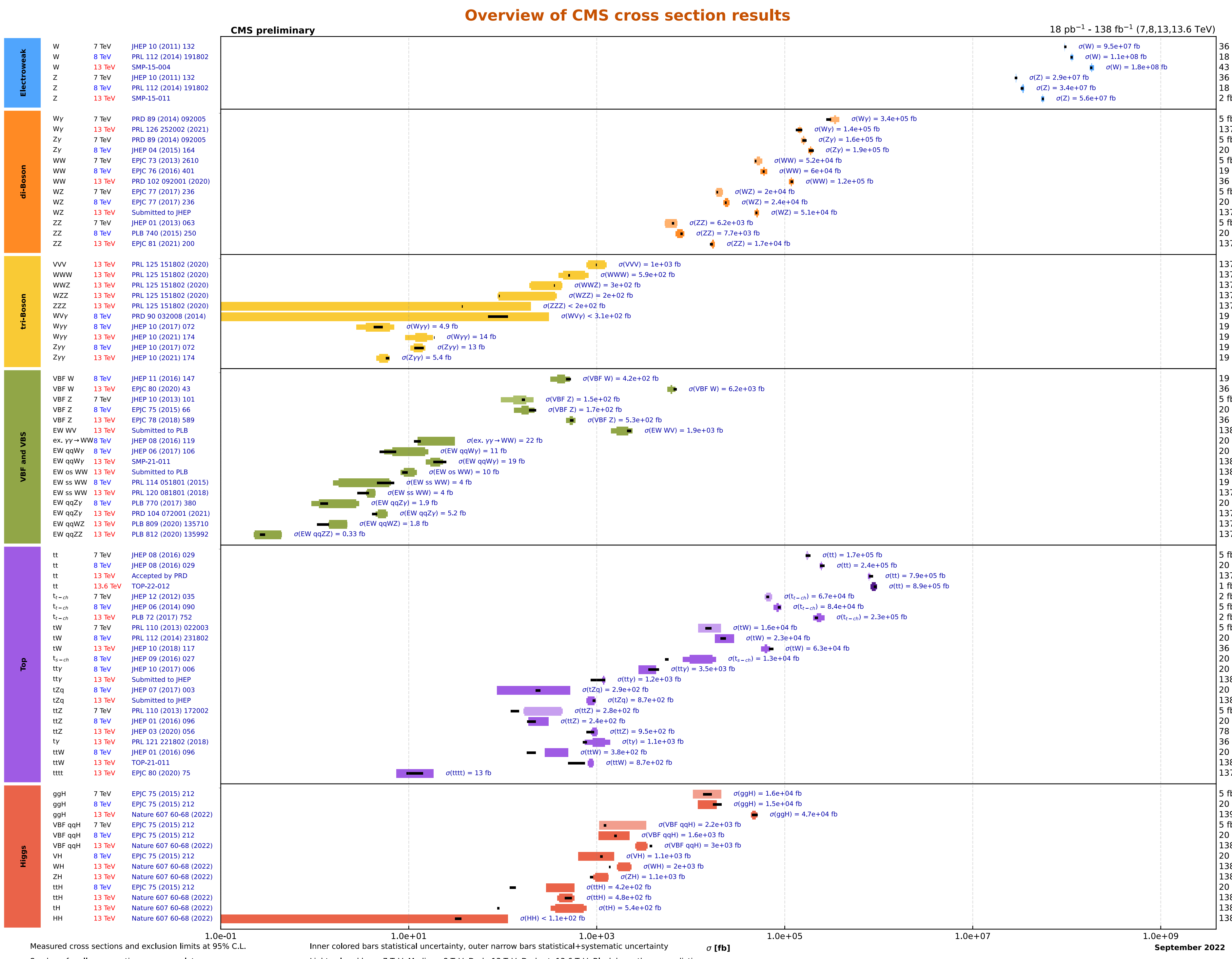
Particle Collisions



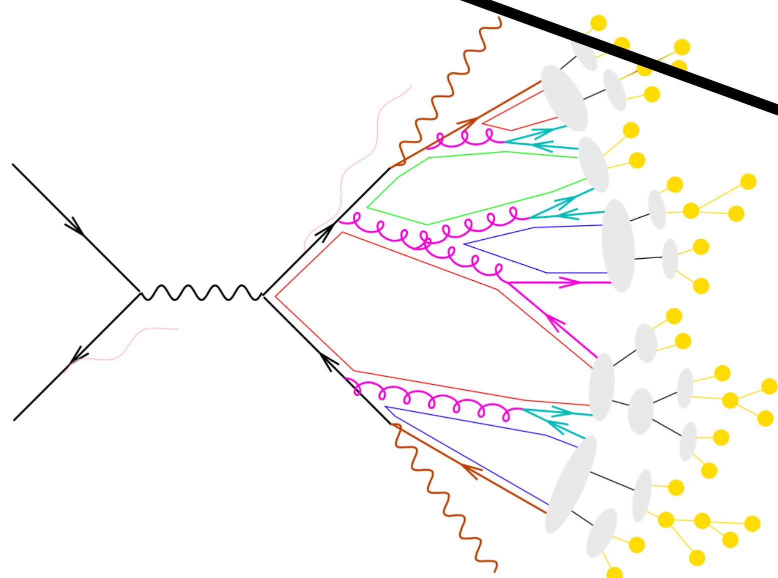
- Collisions turns kinetic energy into new particles ($E=mc^2$)
- Stochastic process, no control over *which* particles get produced
- Very short lived (e.g. 10^{-25} s for the top quark)
- Chain of particle decays, final states measures by detectors

‘Analysis’

- As we can’t control which particles get made, we just keep trying and store the data that look interesting, analyse afterwards
- One dataset allows many types of analyses:
    - Consider each analysis as one ‘experiment’

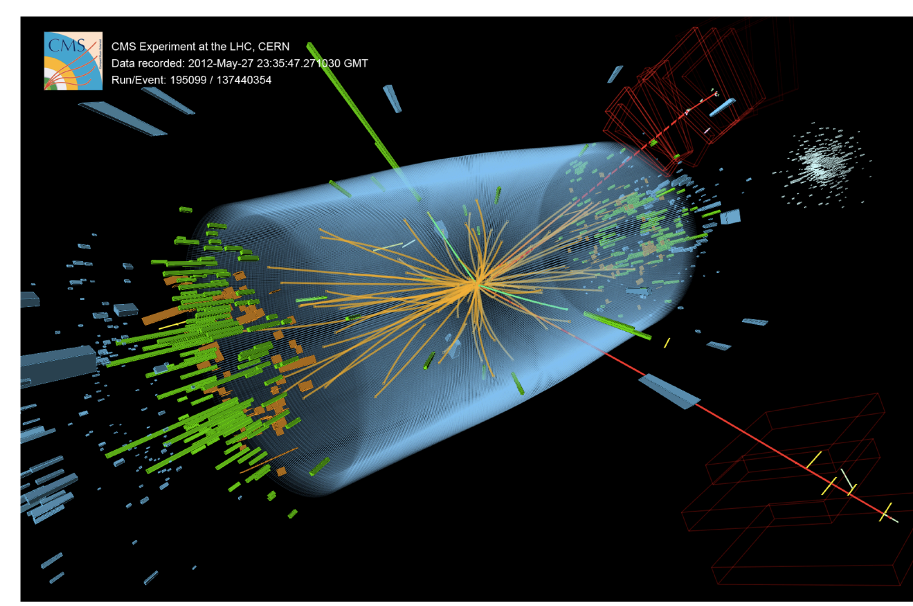


Experimental particle physics workflow



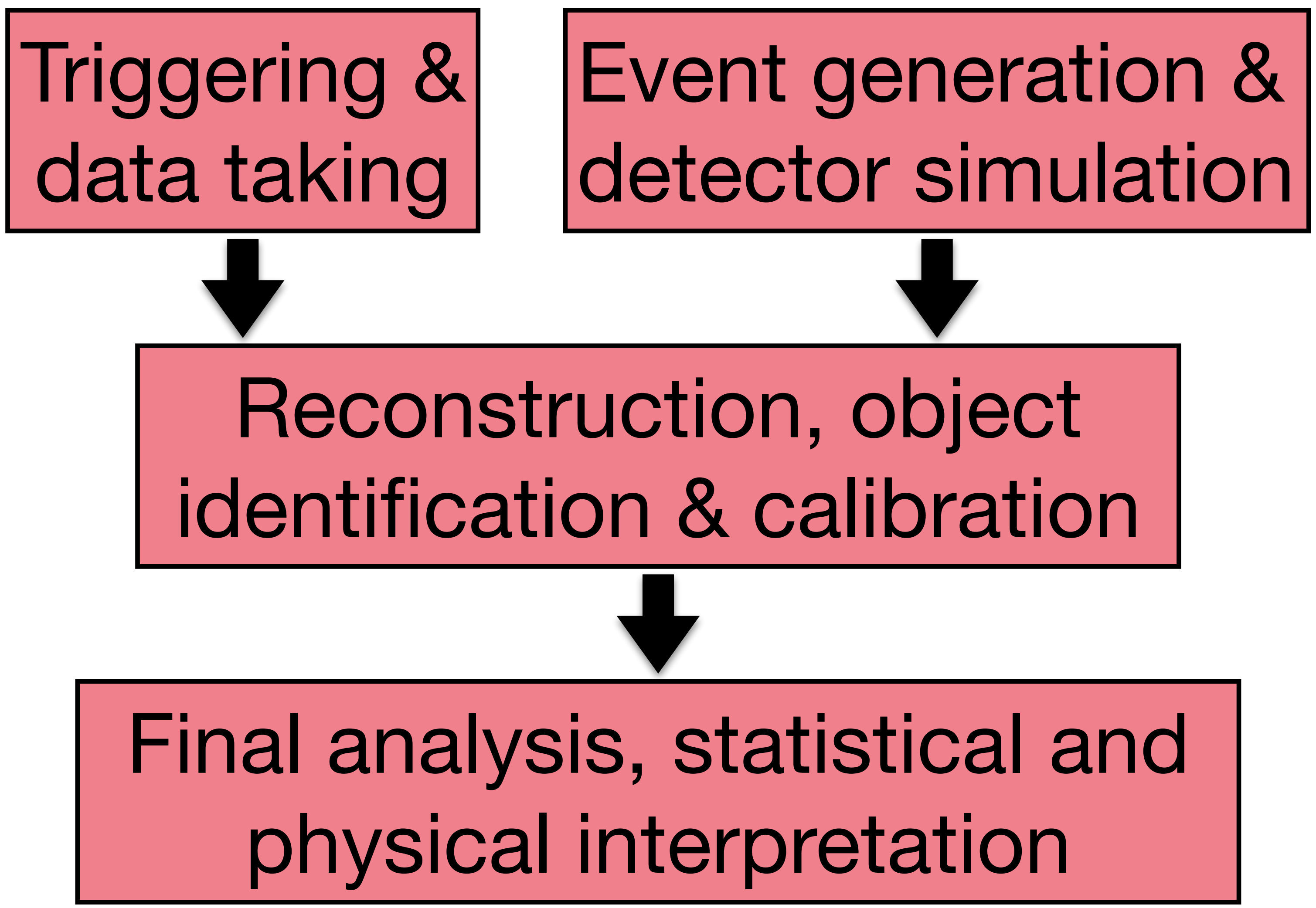
This is what happens in the experiment

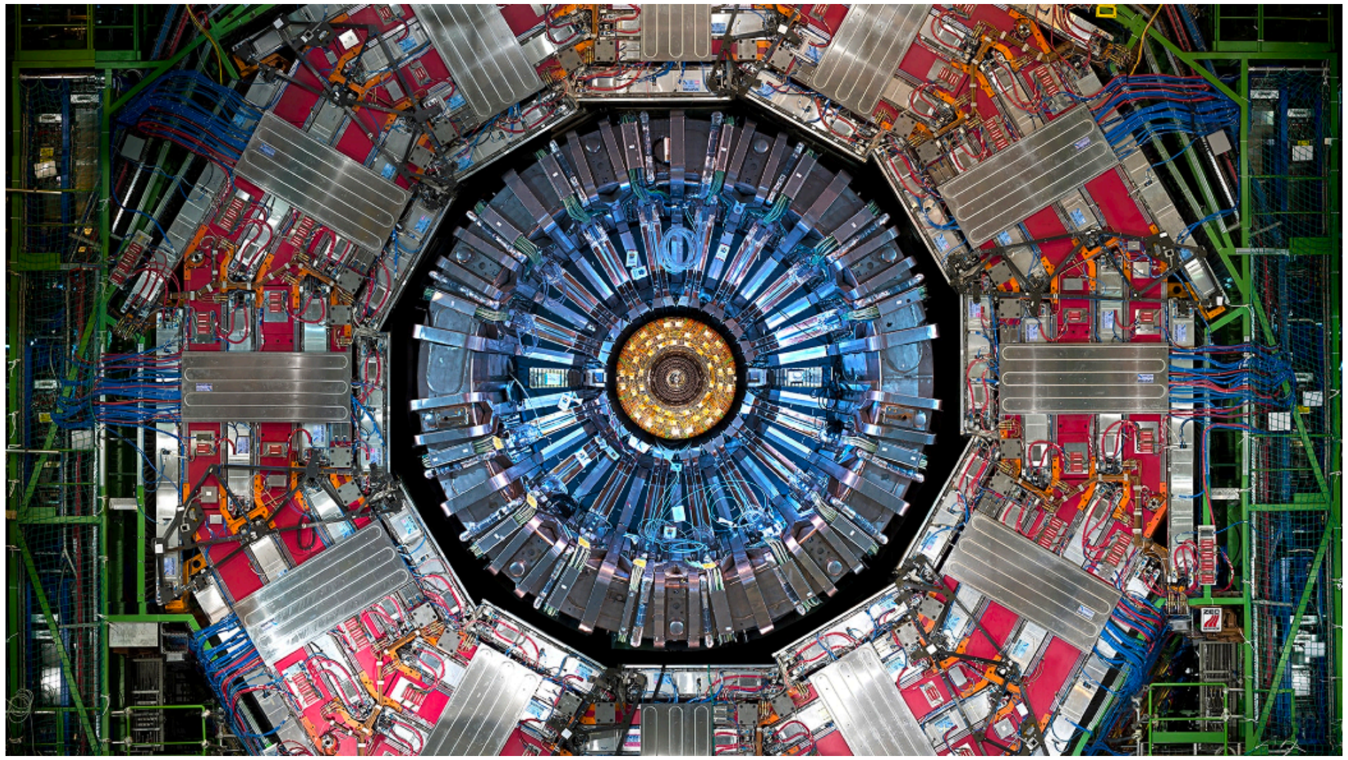
This is what we want to know



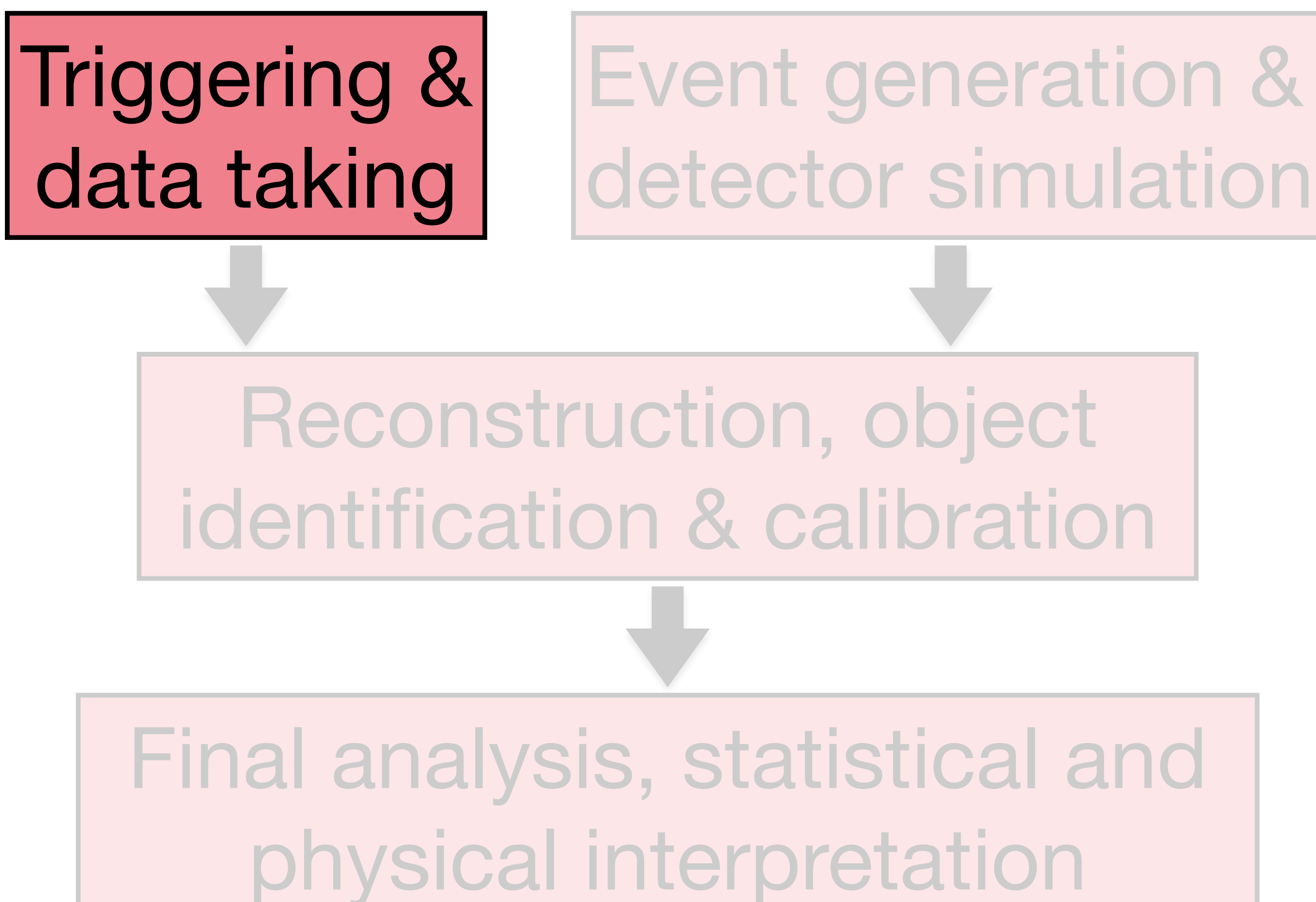
Connect observational data
with underlying theory:
Statistics & simulation

Experimental particle physics workflow





Triggering and data taking

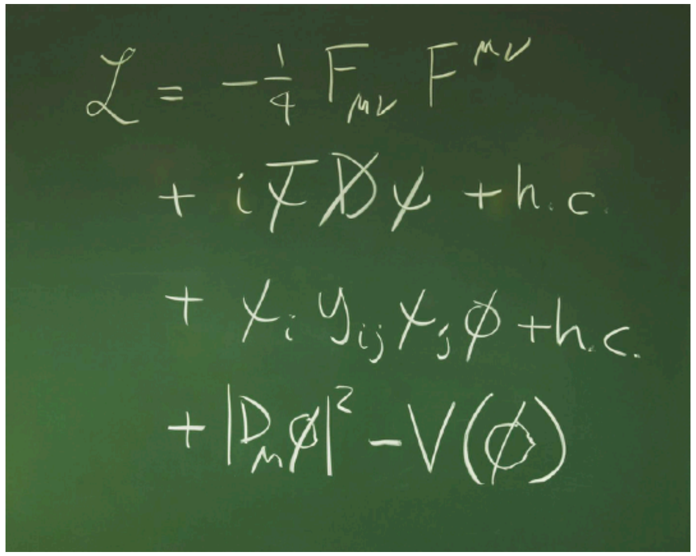


Particle collisions happen at a rate of 40 MHz with size ~1 MB/event.

Need to distill to ~1 kHz via lossy, irreversible filtering algorithms (Trigger).

Data is very heterogenous: low-level readouts in ~100M channels; can condense to O(10) high-level features

One collision = “one image”; sample i.i.d. from underlying physics distribution



&

Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Simulation

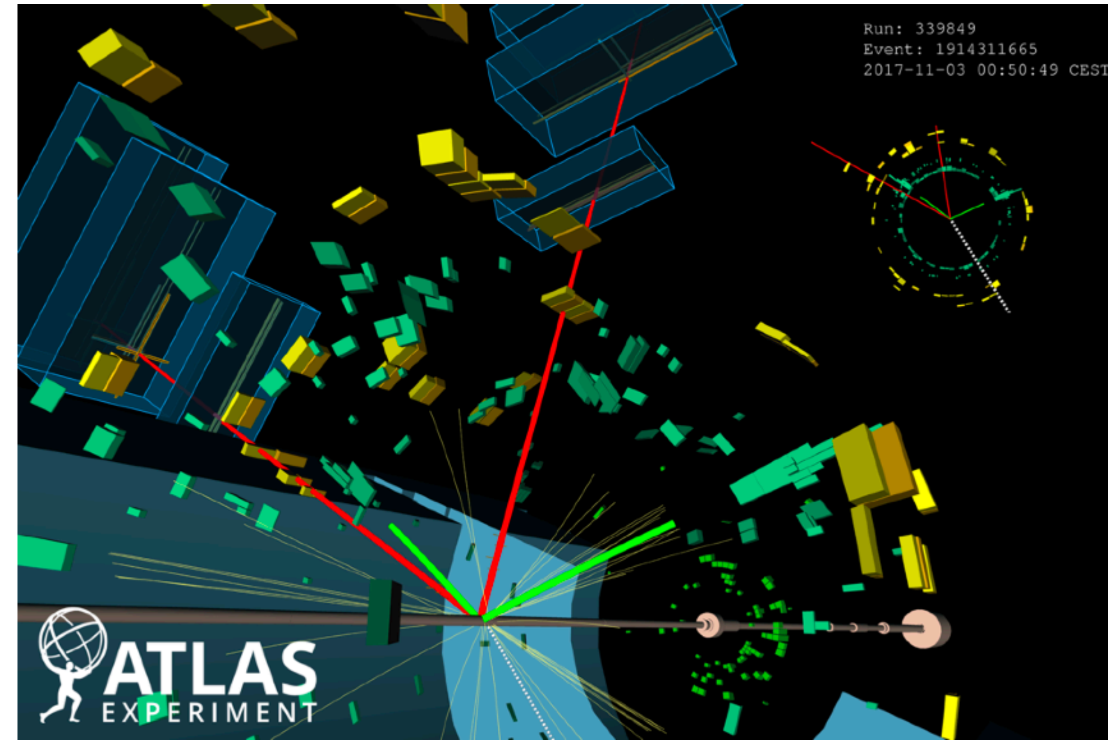
Theoretically well motivated Monte Carlo based simulations of known and hypothetical processes as well as detector responses.

As ~similar amount of simulated and real data is needed, significant compute goes here.

Reconstruction

Build high level objects (particles, leptons, jets, ..) from raw measurements in detectors and identify different particle decays.

Same processing chain for simulation and real data.



Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

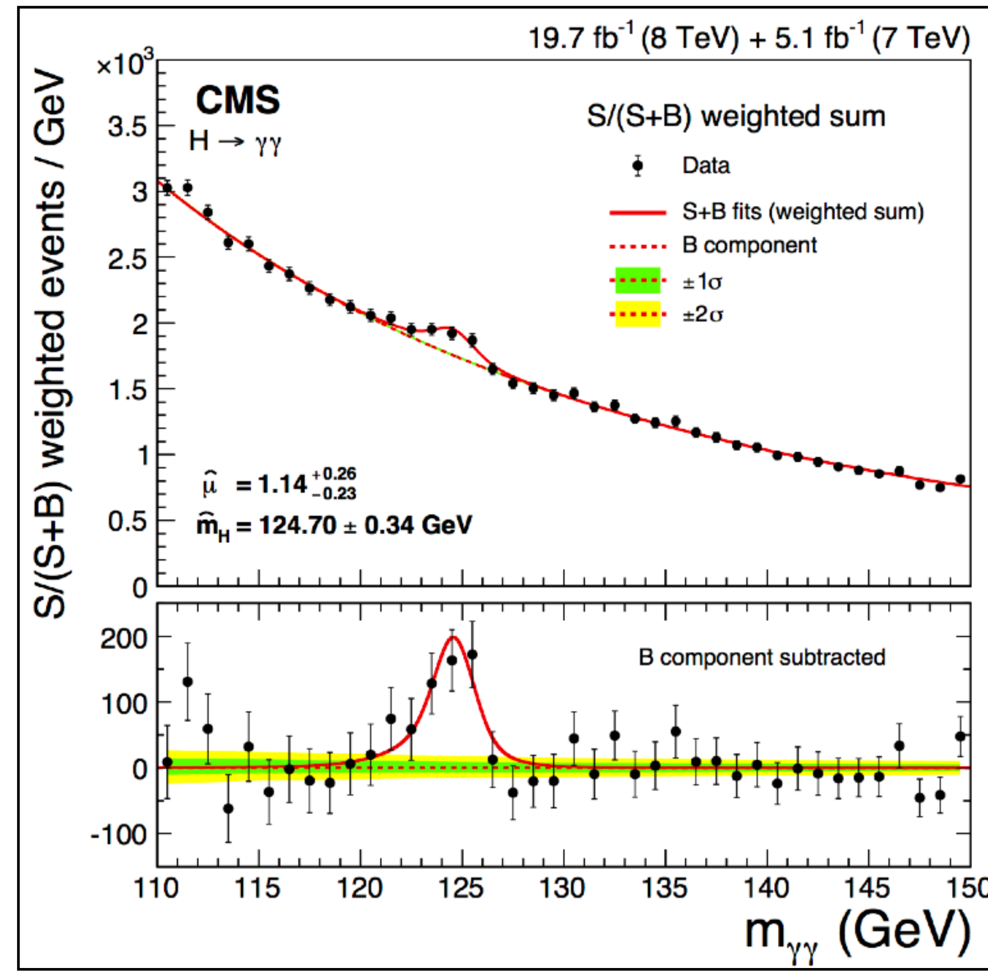
Final analysis, statistical and physical interpretation

Analysis

Previous steps dominated by central running; from here on increasingly local-compute dominated.

Select region of phase space that isolates a physical phenomen of interest and perform detailed statistical analysis.

Compares simulation and data, quantifies uncertainties.



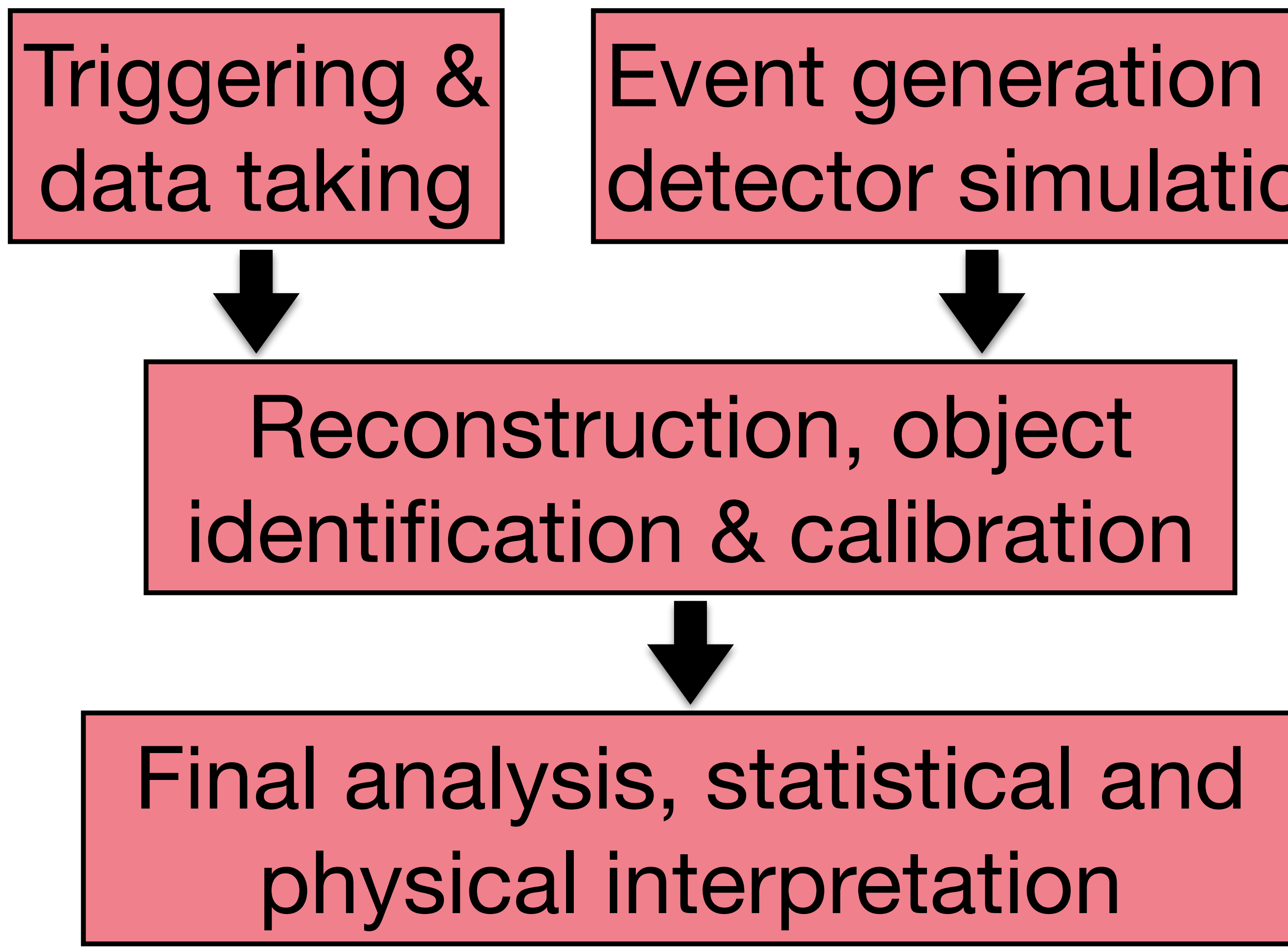
Triggering & data taking

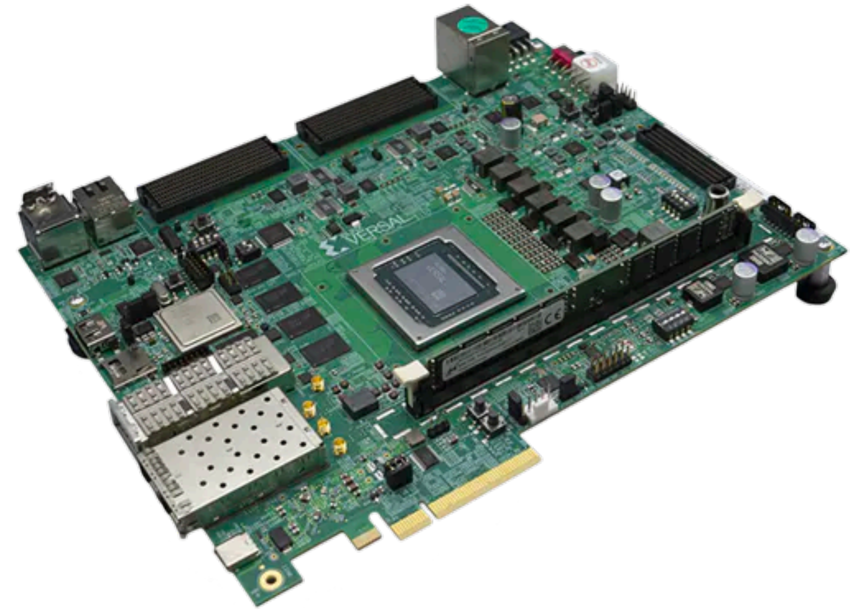
Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Machine learning plays an increasing role in all of these steps





Microsecond decisions needed for deciding whether to store events.

Triggering & data taking

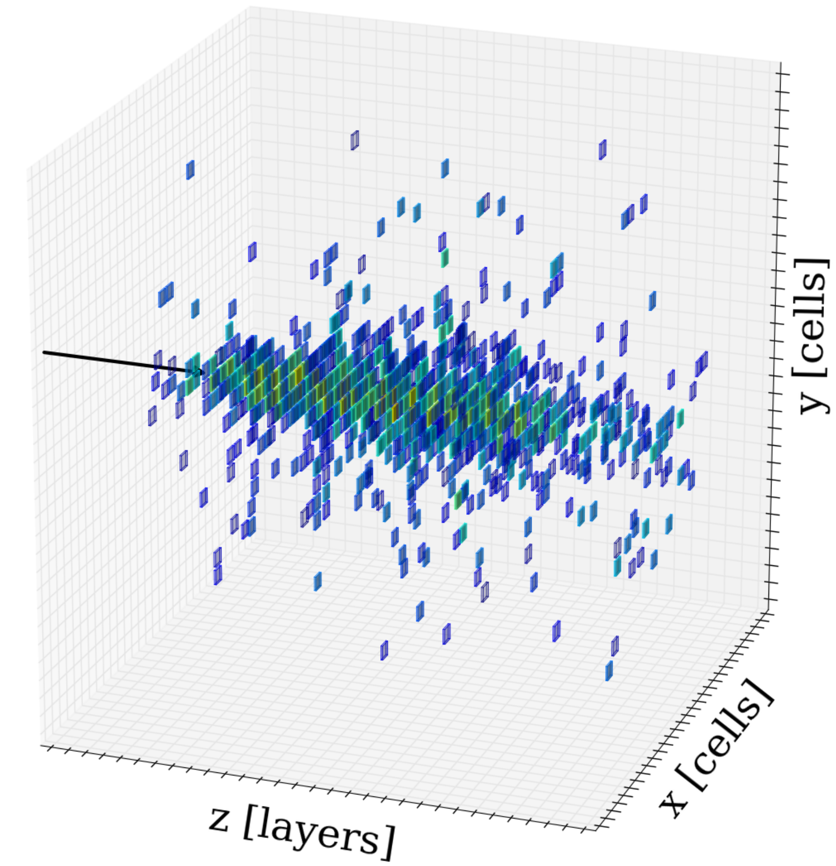
Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Triggering & data taking

Event generation & detector simulation



Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

Efficient simulation needed due to compute constraints.

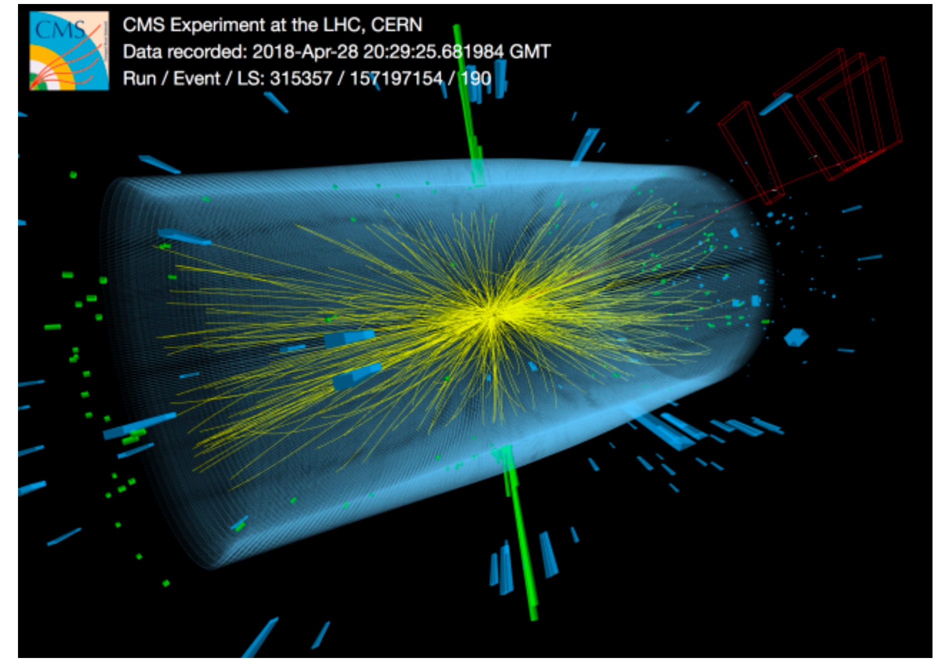
Considering various data representations and generative models (GANs, VAEs, flows, diffusion models)

Triggering & data taking

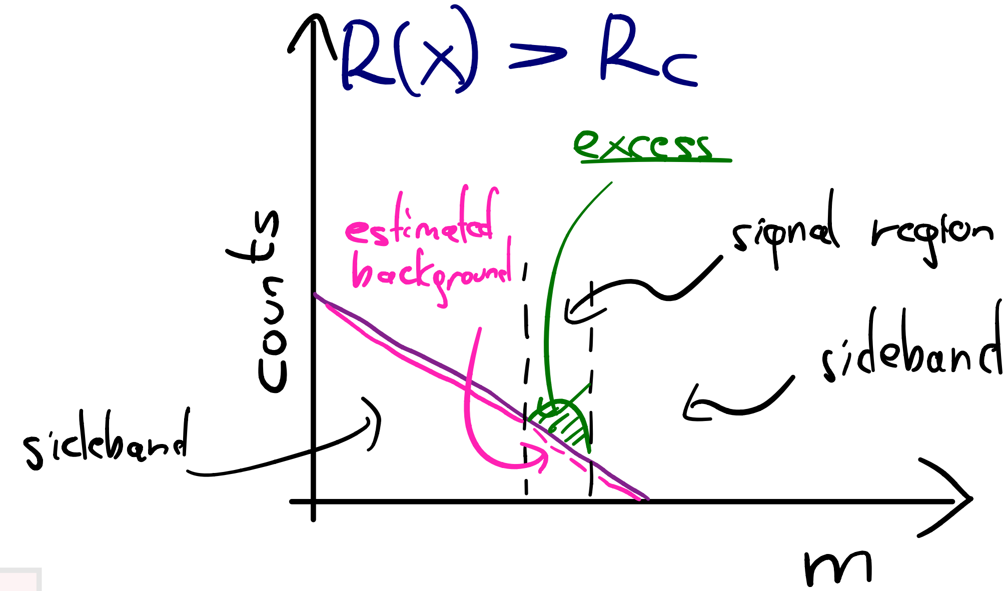
Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation



Difficult task of inferring 'true' physical process from energies measured in the detector.



Triggering & data taking

Event generation & detector simulation

Reconstruction, object identification & calibration

Final analysis, statistical and physical interpretation

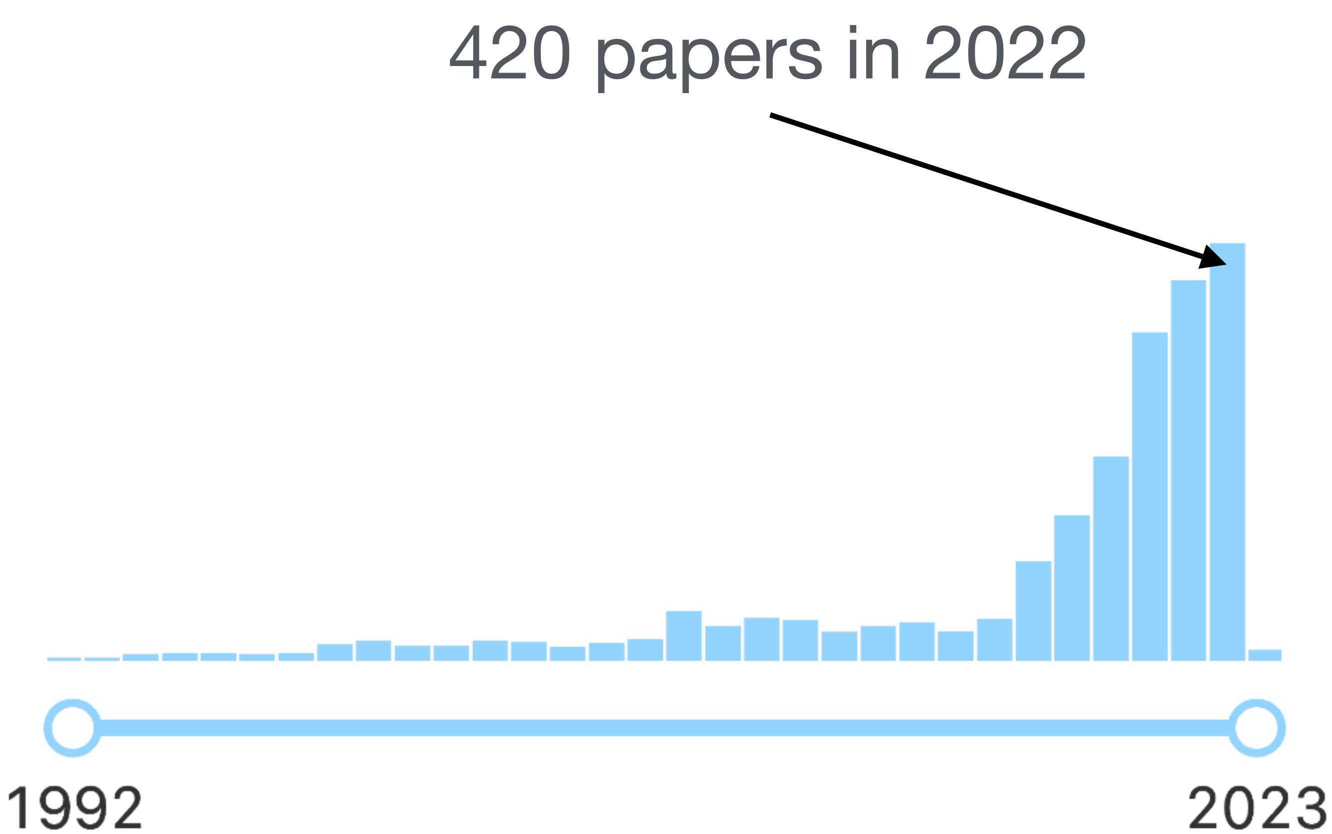
Enhanced bump-hunt: Then fit background from sidebands, compare to data in signal region

Use AI to for robust measurements or discovery of new particles, e.g. via anomaly detection

Machine Learning Particle Physics

Immense progress of machine learning in HEP over the last year

And corresponding increase of applications.

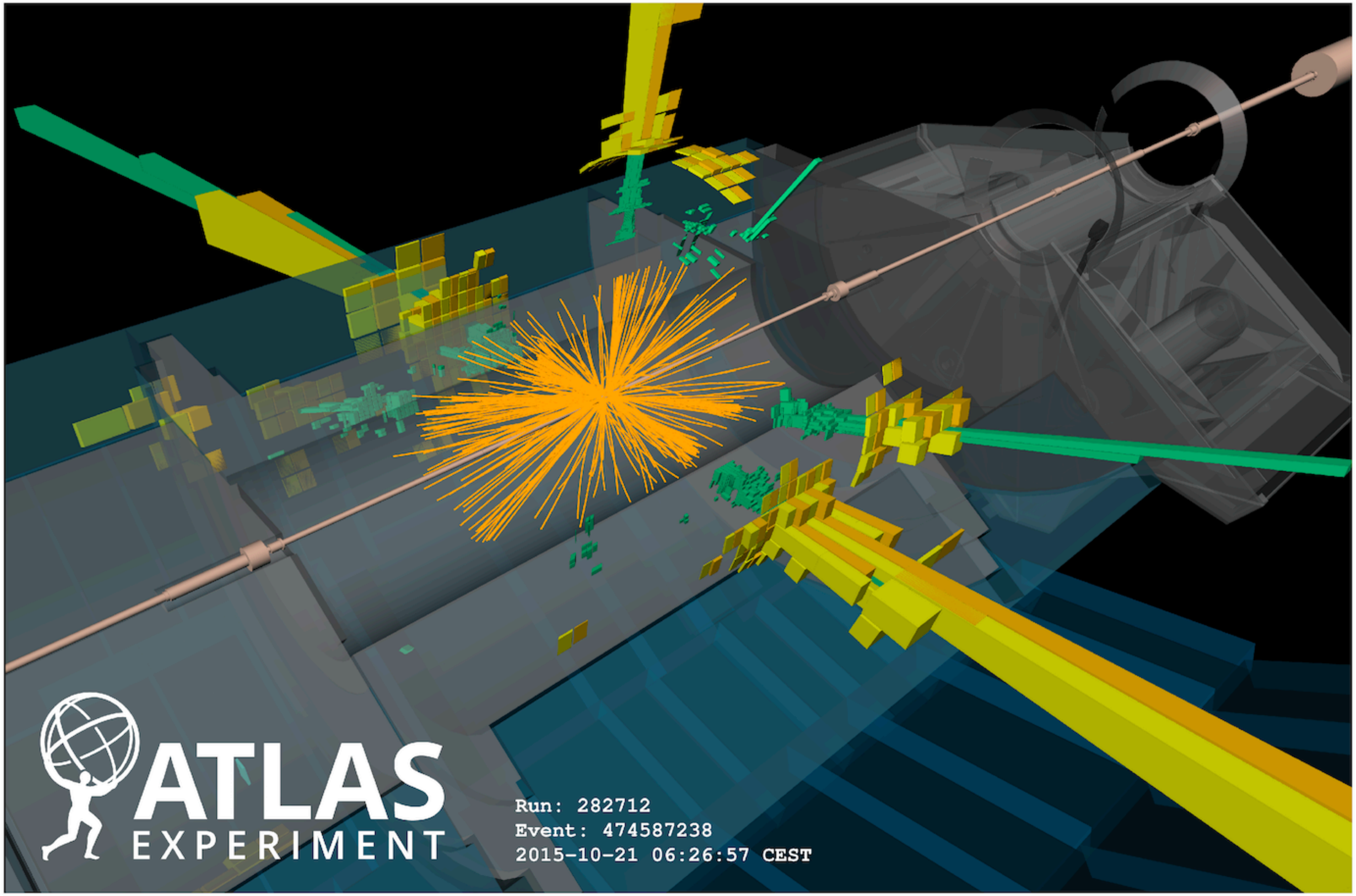


Inspire: ("machine learning" or "deep learning" or neural) and (hep-ex or hep-ph or hep-th)

- **Special role of HEP:**
- “Infinite” amounts of high quality labelled training data from realistic simulation accompanied by huge experimental datasets
- Interestingly structured data at multiple scales
- Detailed understanding of systematic uncertainties
- Asks fundamental questions about Nature

Part II: Symmetries & how to treat them

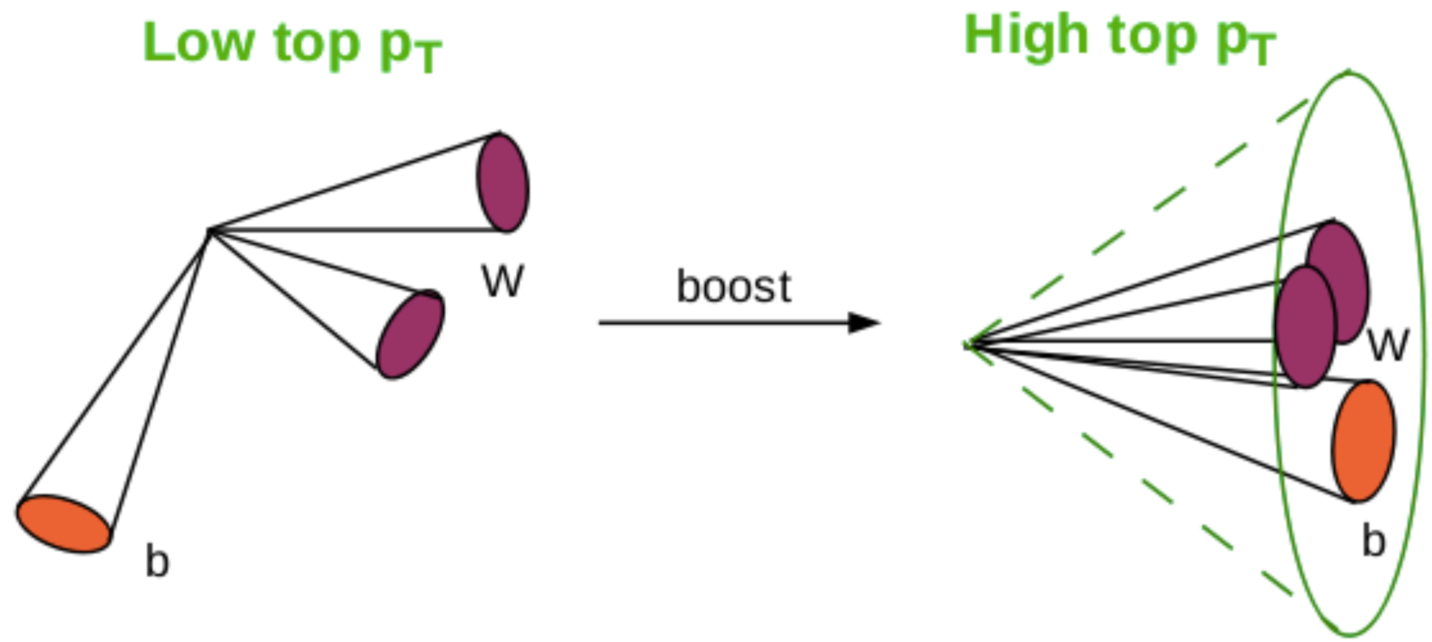
Jet tagging



- Intuitively a jet is:

  Collimated shower of particles in the detector

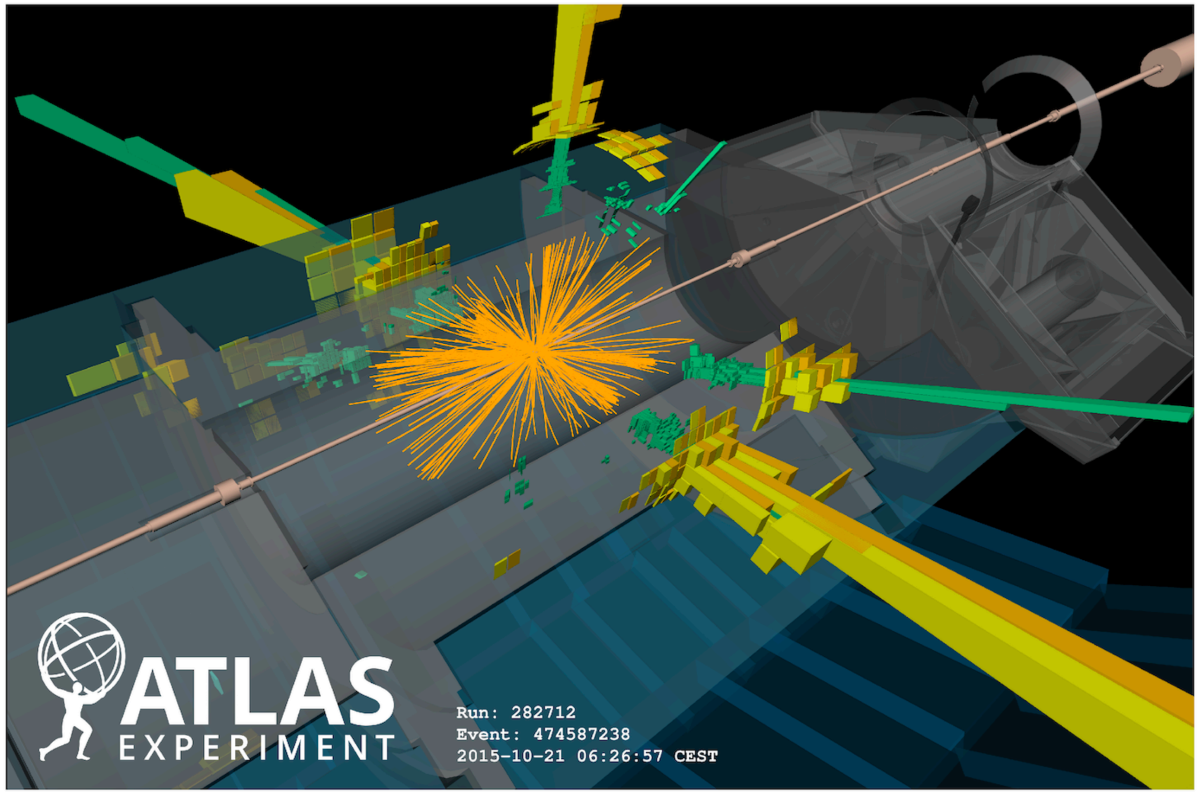
Top Quark Identification

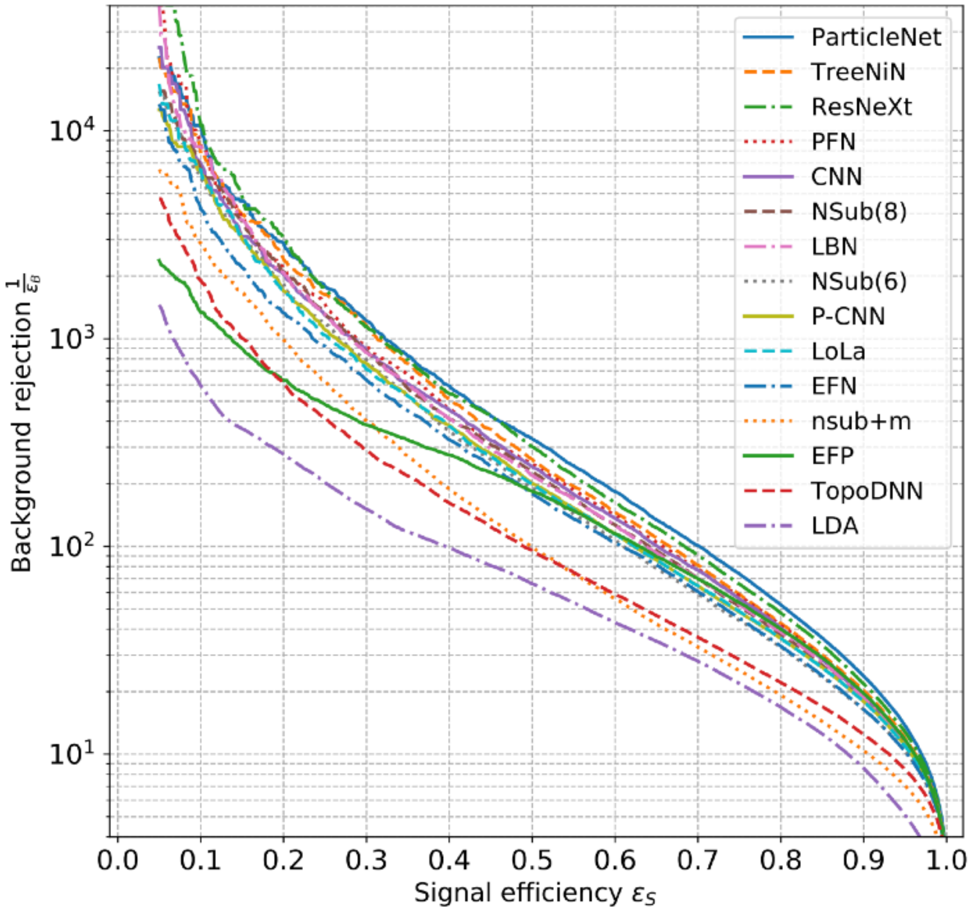


- Top quark:
  - Heaviest known elementary particle
  - Relevant for measurement and searches for new theories
- Hadronically decaying top/Higgs/W/Z
- Contained in one (large-R) jet
  - m/pT >= ~1
- How to distinguish from light quark/gluon jets (and from each other)
- Used for new physics searches (and SM studies)

Concrete task

- Distinguish jets initiated by a top quarks from jets from other particles
  - Binary classification task
- Use simulation as synthetic training data: perfect class labels available
  - (Leads to domain shift when applied to collider data)



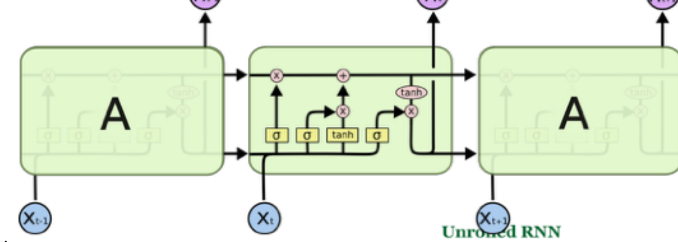


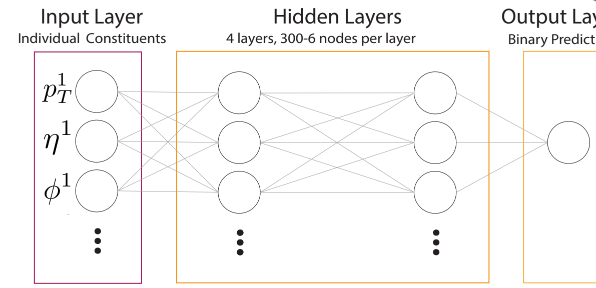
- 1.2M training examples (*jets*), 400k each for testing and validation
- Each example: Up to 200 particles with 3 features/particle (2D position on detector surface+ energy)
- Metrics: AUC: area under curve and R_{30}:1/FPR @ TPR=0.3 ()

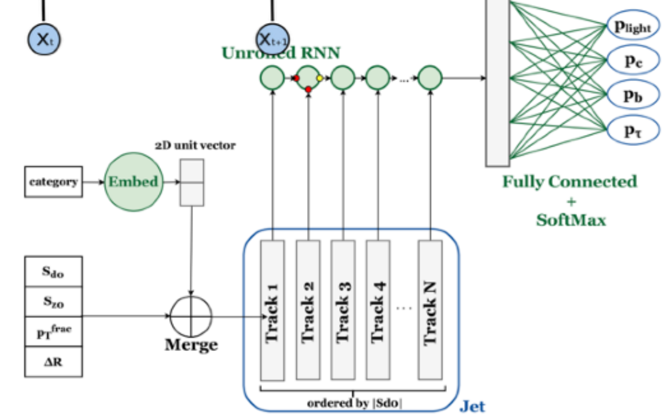
1902.09914

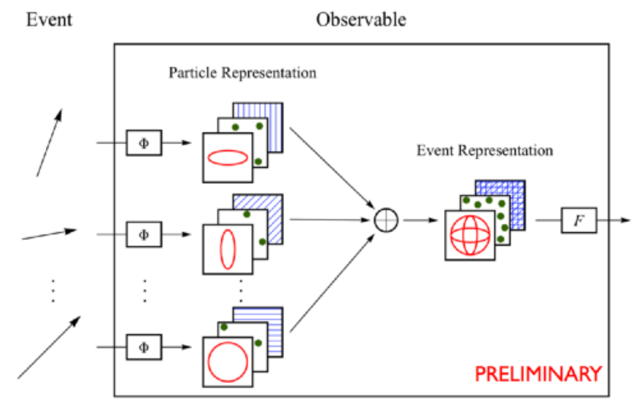
Enter deep learning

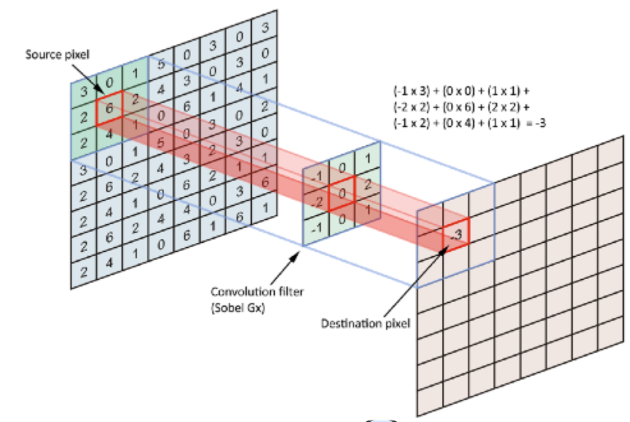
- Particles form a point cloud in space
  - Permutation symmetry
  - Symmetry of points in space:
    - Naively SO(3), actually Lorentz group
- How to solve with deep learning?
- Immense number of results, showcase some (useful) examples

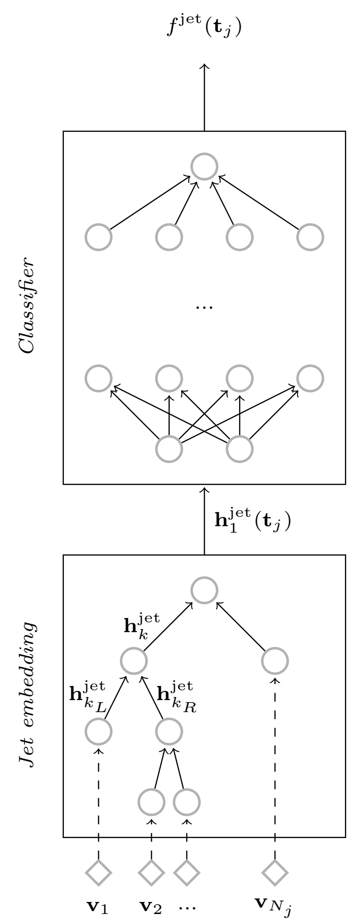


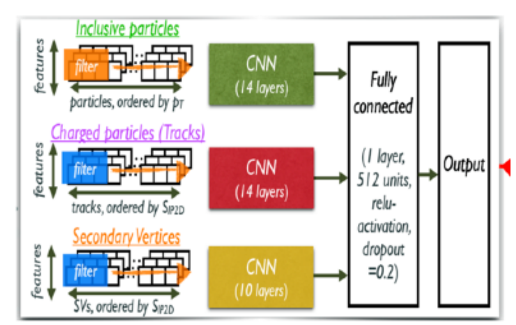


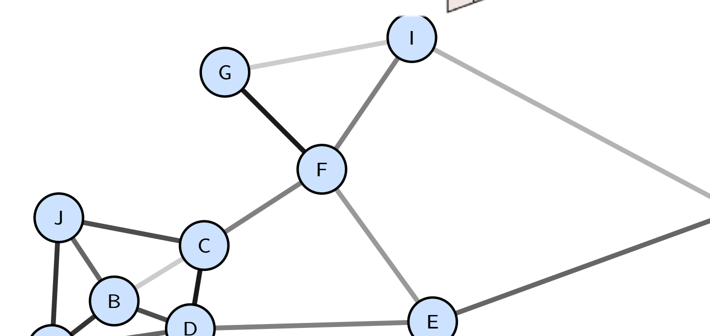












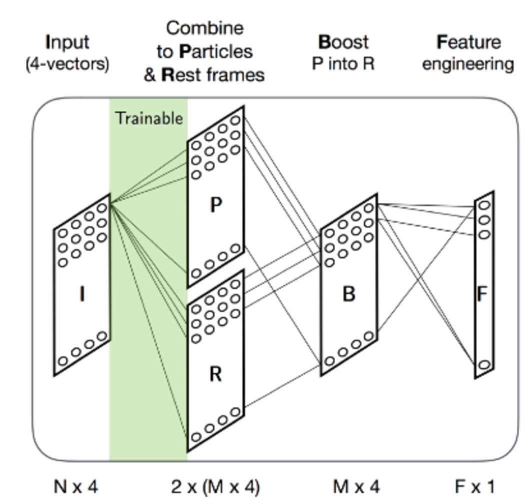
$$k_{\mu,i} = \begin{pmatrix} E_0 & E_1 & \cdots & E_N \\ p_{x,0} & p_{x,1} & \cdots & p_{x,N} \\ p_{y,0} & p_{y,1} & \cdots & p_{y,N} \\ p_{z,0} & p_{z,1} & \cdots & p_{z,N} \end{pmatrix}$$

Combination Layer (**CoLa**): create linear combinations:

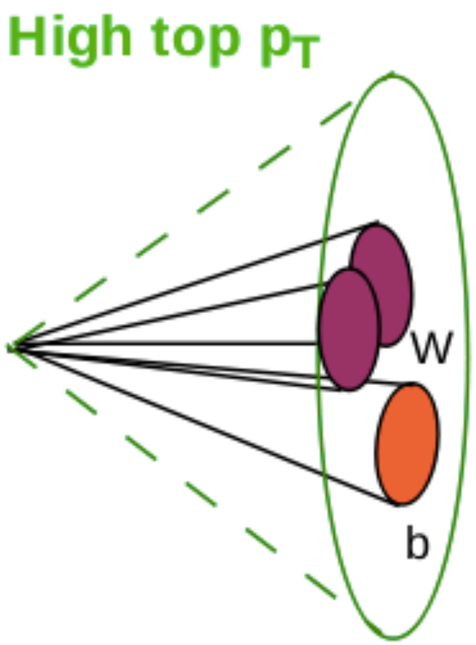
$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\, C_{ij}$$

Lorentz Layer (**LoLa**): Use resulting matrix to extract physics features.
Main assumption is the Minkowski metric

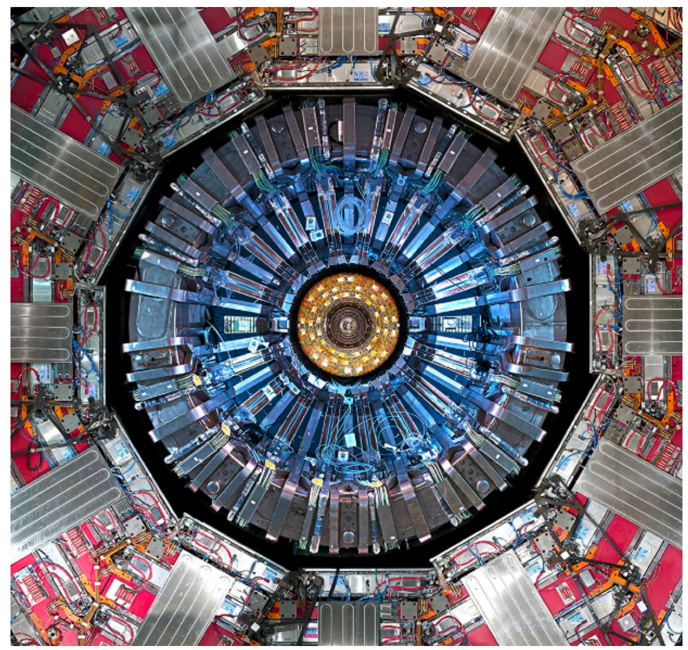
$$\widetilde{k}_{\mu,i} \to \sum_j (\widetilde{k}_i - \widetilde{k}_j)_\mu (\widetilde{k}_i - \widetilde{k}_j)_\nu \eta^{\mu\nu} B_{ij}$$



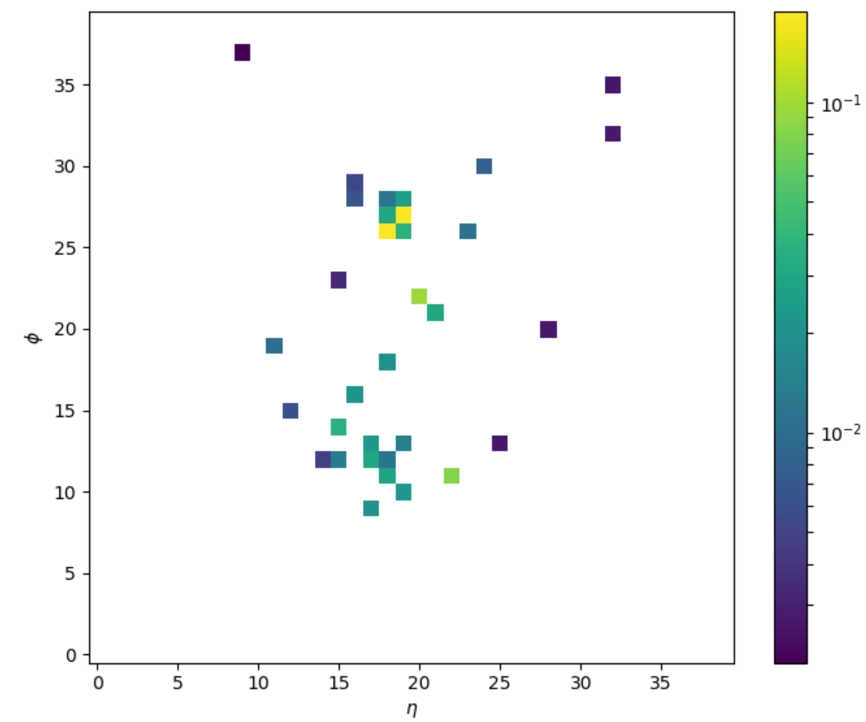
Jet Images



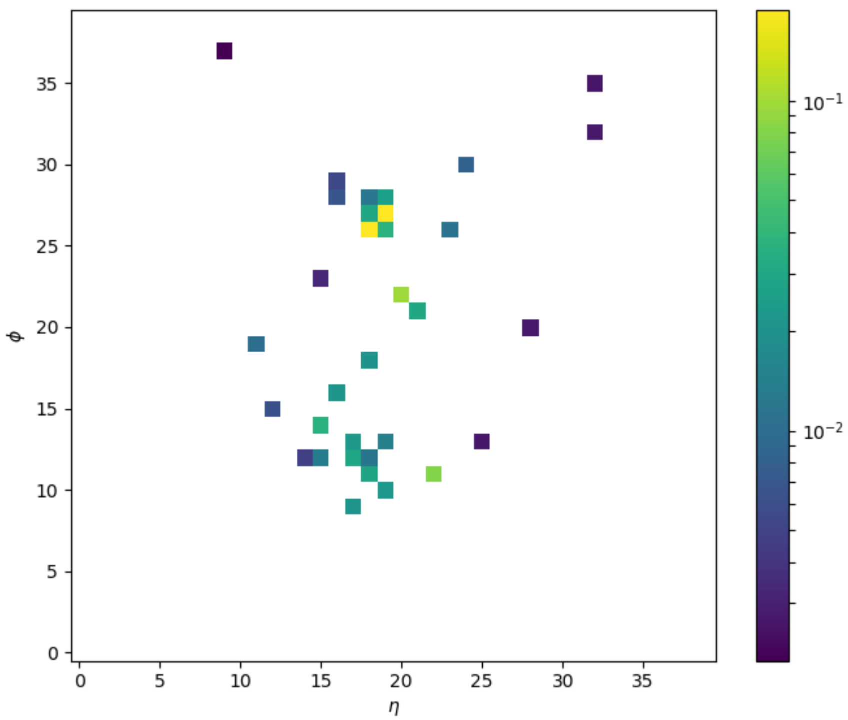
+



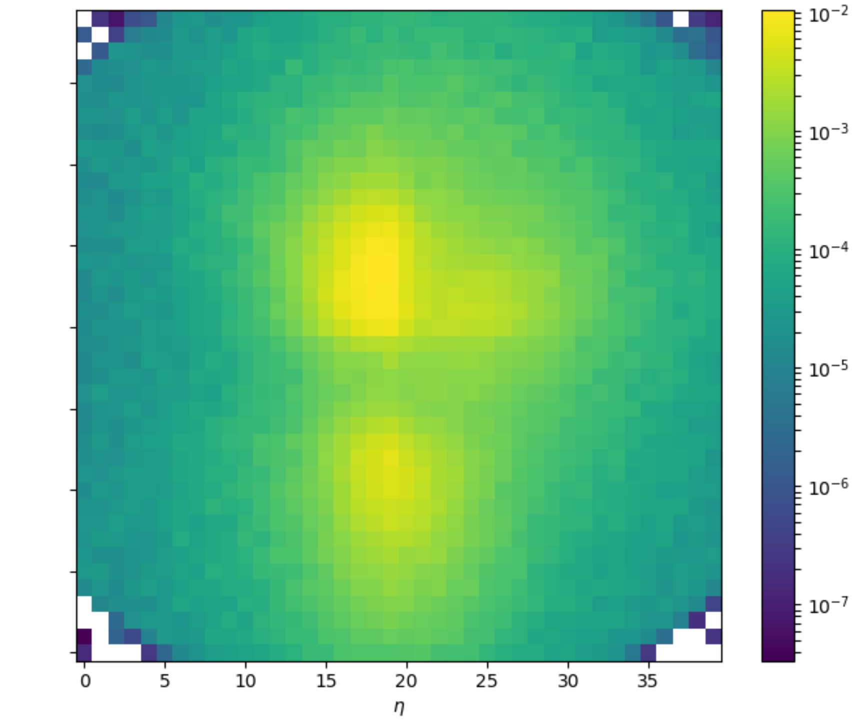
=



- Treat jets as images: Popular and done before deep learning (1407.5675, 1501.05968, 1511.05190, 1612.01551, 1701.08784, 1803.00107,….)
- Measure particle energies in calorimeter
- Image preprocessing
  - center, rotate, mirror, pixelate, trim, normalise

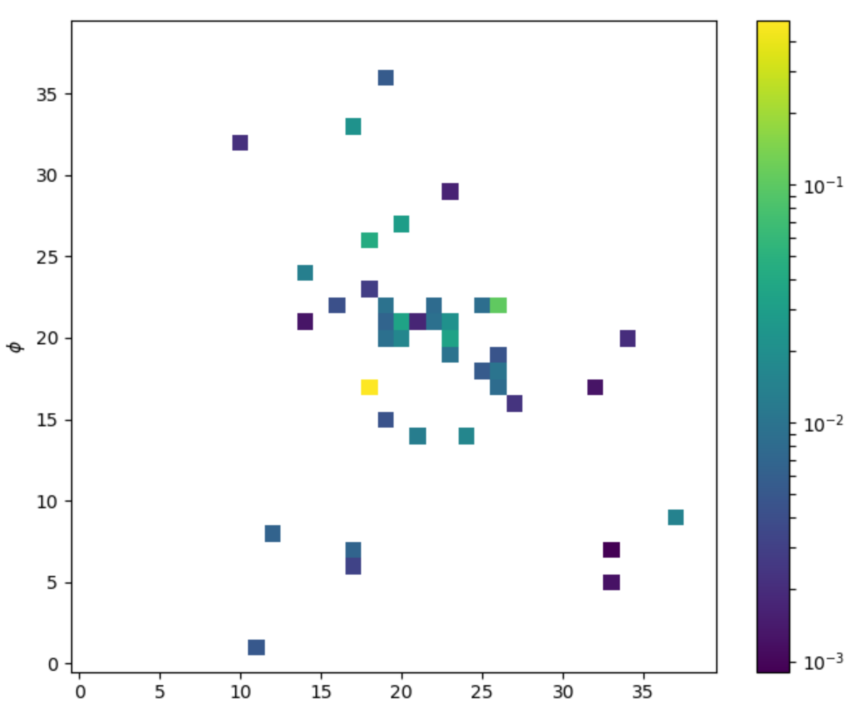


Single top jet

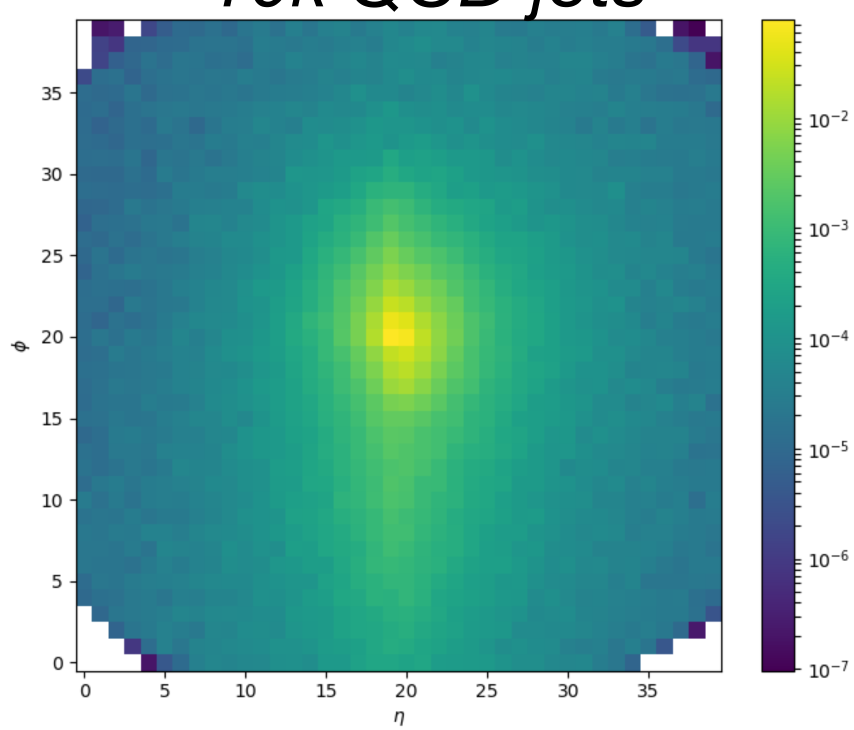


10k top jets

Single QCD jet



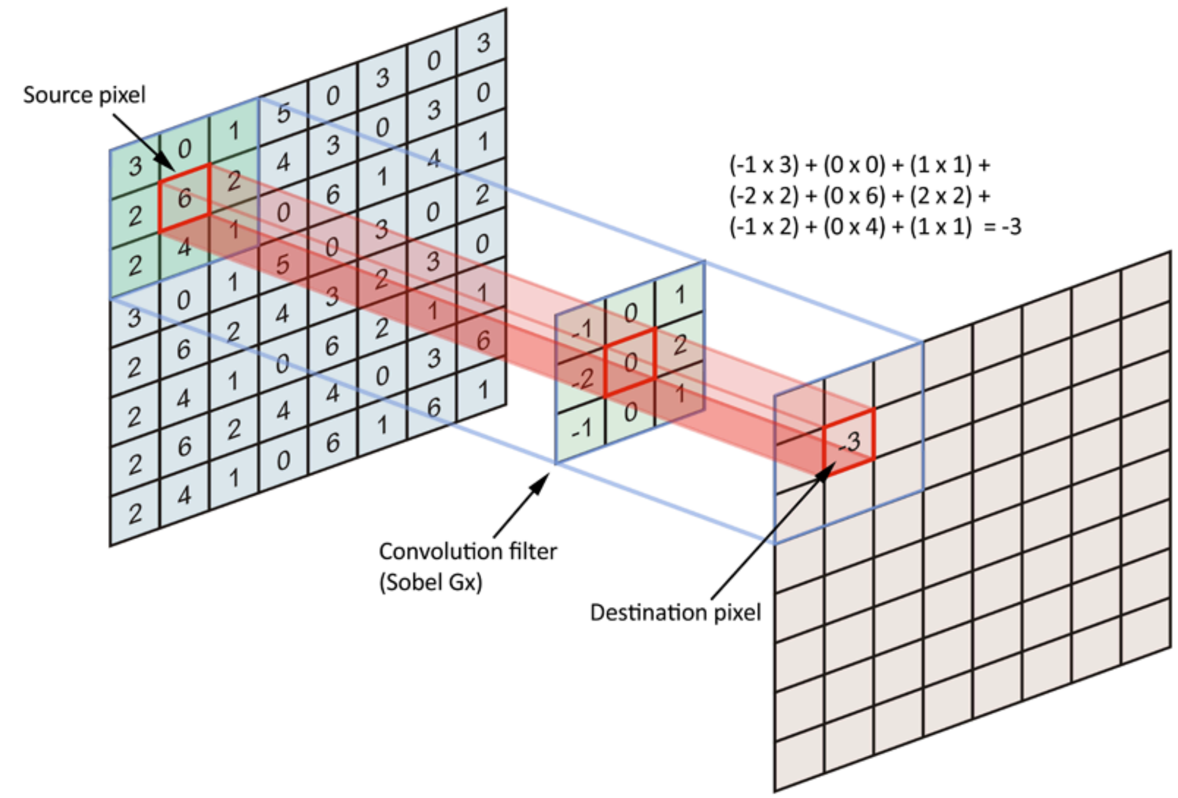
10k QCD jets

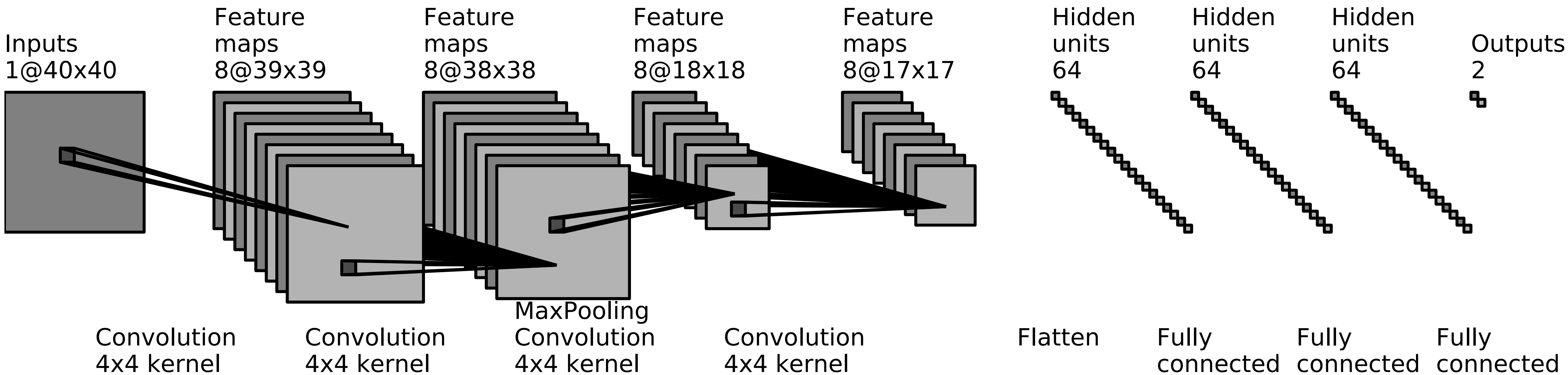


Different initial particles lead to different distributions of recorded energies

Convolutional network

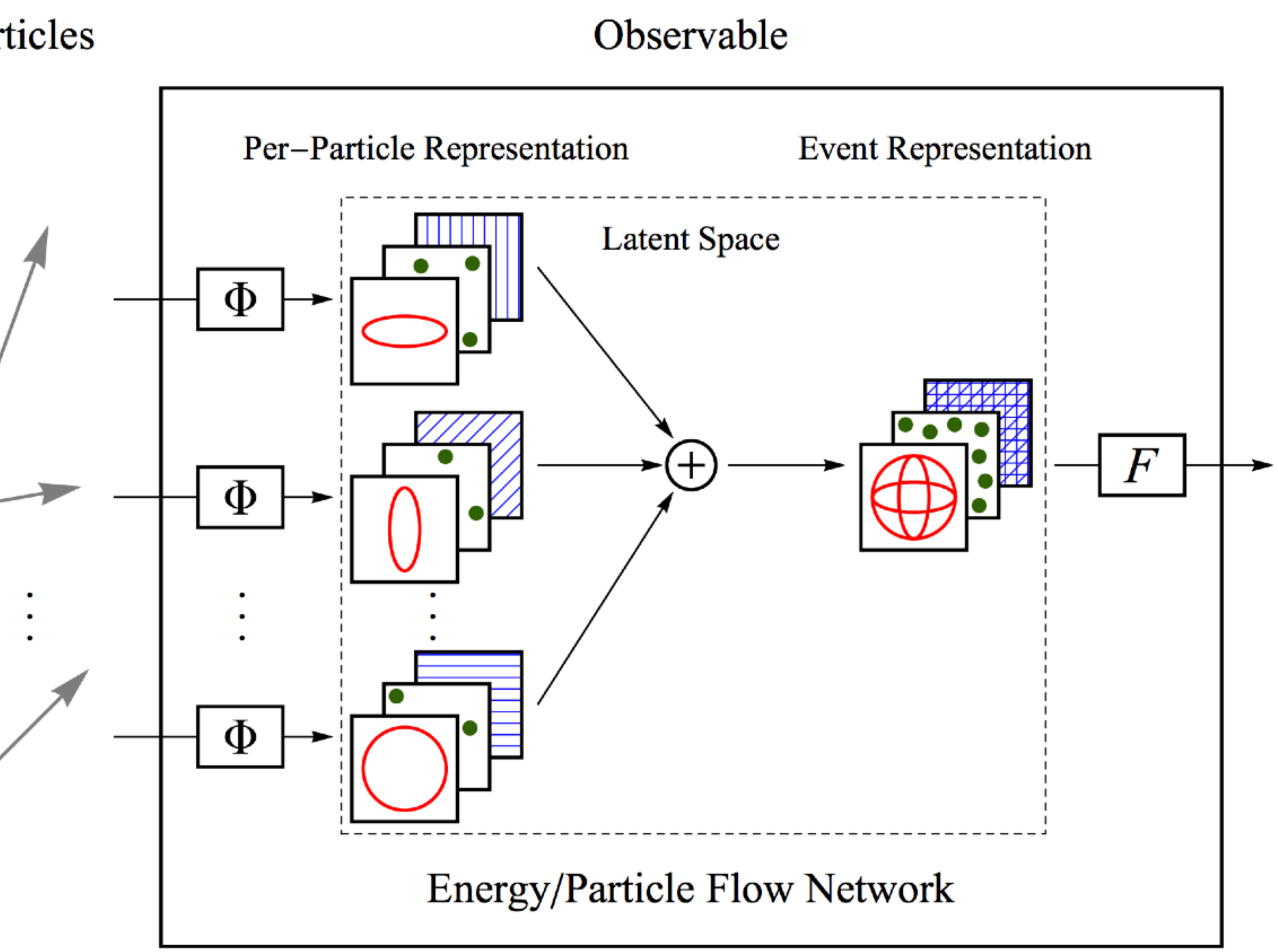
- Analyse grid-like data with convolutional networks
  - Same architectures as for computer vision
- Accounts for locality (correlation of nearby pixels) and *translation invariance*
  - *In fact not a symmetry of the images!*
- Potential limitation due to sparsity/pixelisation for high resolution data
  - No strong effect observed in this study
  - Careful how to pre-process (1803.00107)





Architecture from 1701.08784

Deep Sets



General :

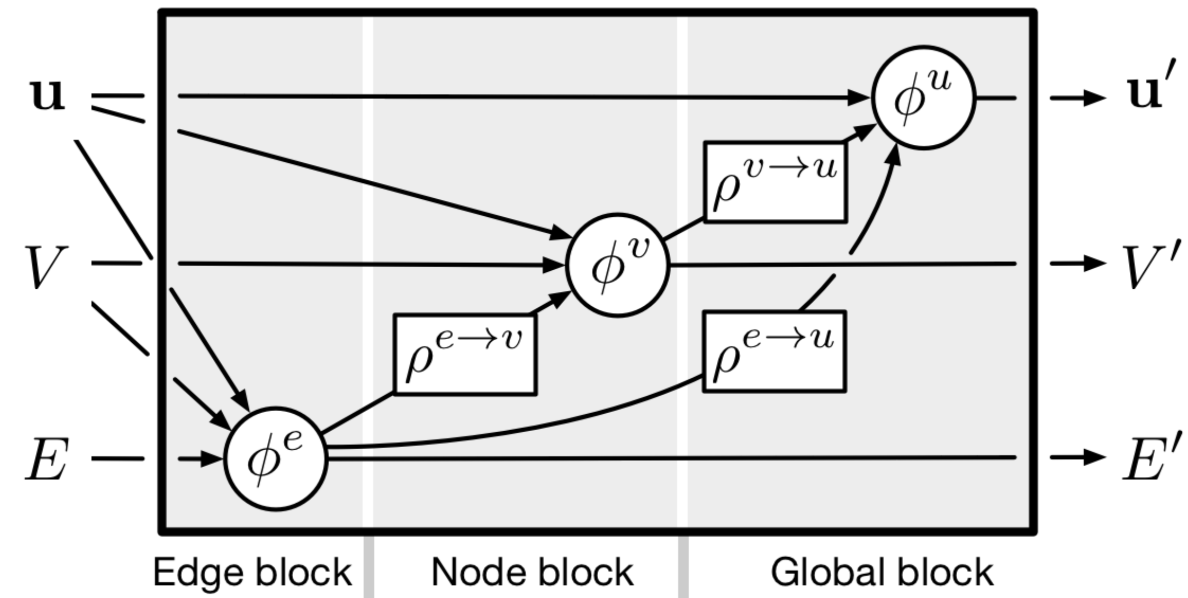
$$\text{PFN:} \quad F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

- Data is a permutation invariant point cloud: treat with set-based architecture
- Invariance/equivariance under symmetries
- How to make independent from ordering of four vectors?
  - Use permutation invariance of sum
  - →Deep set architecture (1703.06114)
  - Apply to jets: energy flow network (EFN) / particle flow network (PFN) (1810.05165)
- Simple and straightforward to implement but limited use of neighbourhood information

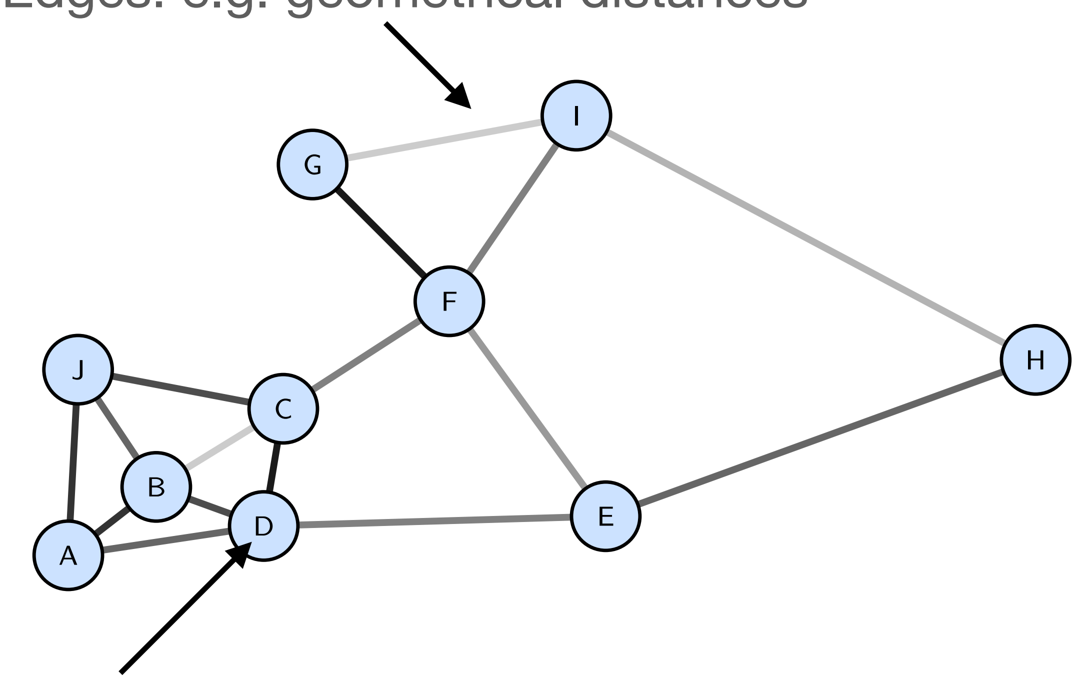
1810.05165

Graphs

- Basic motivation: Use physicists' knowledge about data as an implicit (or explicit bias) to help networks train faster / achieve better performance



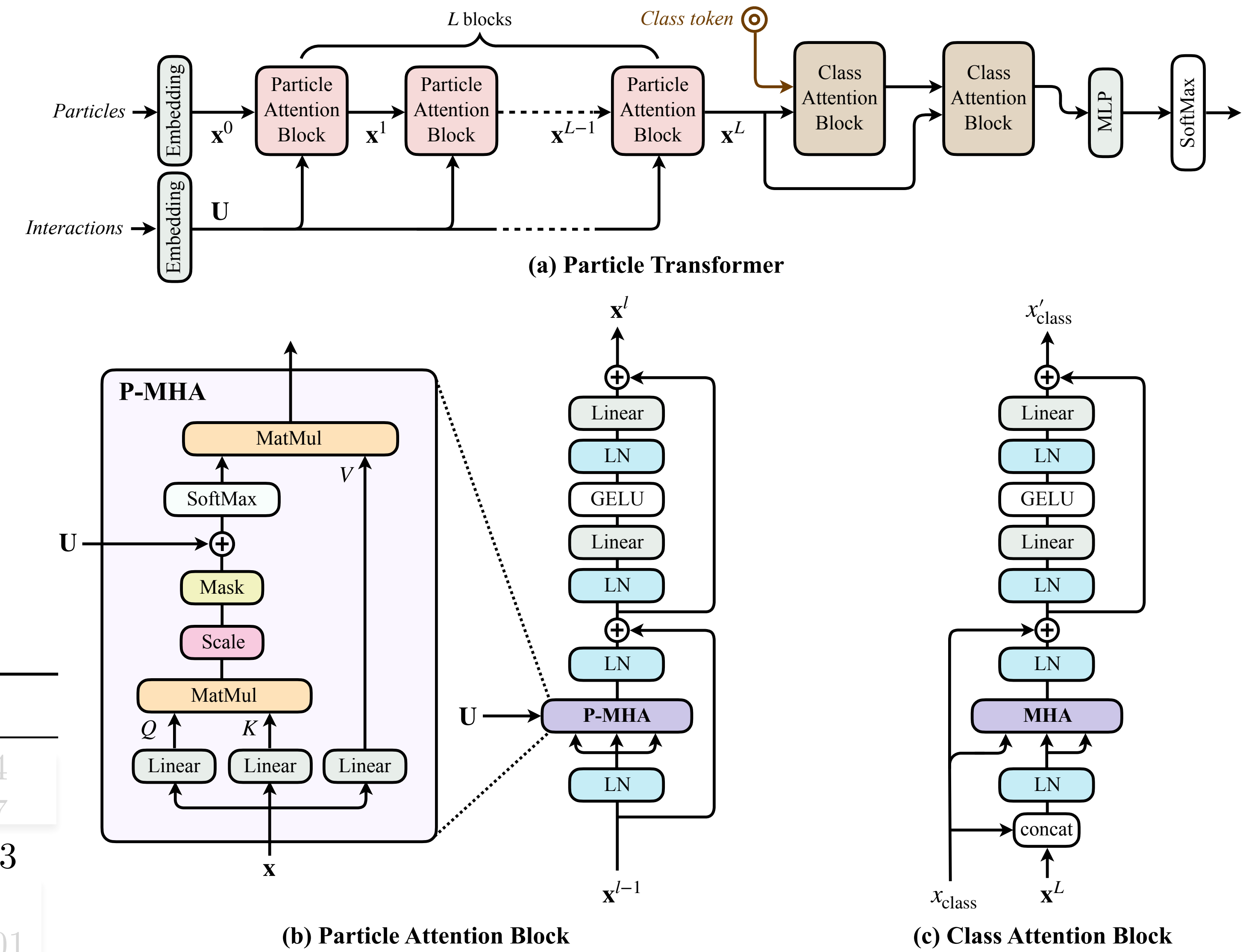
- **Graphs** are a general + powerful framework that captures relevant properties for particle tagging
  - e.g. best performance of ParticleNet (message passing graph) in top tagging comparison
  - versatile and well suited
- Can impose graph on set-like data e.g. by kNN clustering



Henrion et al ML4PS 2017; Qu, Gouskos 1902.08570; Shalom, Battaglia, Valiant 2007.13681 (review)

Transformers; Attention is all you need

- In ParticleNet, data-space geometry defines neighbourhood in graph; aggregation over all neighbours
- Attention allows the network to learn which parts of the input are truly relevant
- Attention is data-hungry, transfer-learning helps!



ParT architecture diagram

	Accuracy	AUC	$\text{Rej}_{50\%}$	$\text{Rej}_{30\%}$
P-CNN	0.930	0.9803	201 ± 4	759 ± 24
PFN	—	0.9819	247 ± 3	888 ± 17
ParticleNet	0.940	0.9858	397 ± 7	1615 ± 93
JEDI-net (w/ $\sum O$)	0.930	0.9807	—	774.6
PCT	0.940	0.9855	392 ± 7	1533 ± 101
LGN	0.929	0.964	—	435 ± 95
rPCN	—	0.9845	364 ± 9	1642 ± 93
LorentzNet	0.942	0.9868	498 ± 18	2195 ± 173
ParT	0.940	0.9858	413 ± 16	1602 ± 81
ParticleNet-f.t.	0.942	0.9866	487 ± 9	1771 ± 80
ParT-f.t.	**0.944**	**0.9877**	$\mathbf{691 \pm 15}$	$\mathbf{2766 \pm 130}$

Performance comparison on landscape dataset

Vaswani et al 1706.03762; Qu, Li, Qian 2202.03772; Mikuni, Canelli 2001.05311; ..

Attention is all you need

- In ParticleNet, data-space geometry defines neighbourhood in graph; aggregation over all neighbours
- Attention allows the network to learn which parts of the input are truly relevant
- Attention is data-hungry, transfer-learning helps! (Motivation for foundation models?)
- So far, observed trend:
  Higher physics performance comes at the cost of higher algorithm complexity & compute cost
- **Is this the only way?**

	Accuracy	# params	FLOPs
PFN	0.772	86.1 k	4.62 M
P-CNN	0.809	354 k	15.5 M
ParticleNet	0.844	370 k	540 M
ParT	**0.861**	2.14 M	340 M
ParT (plain)	0.849	2.13 M	260 M

(plain: standard multi-head-attention vs particle-multi-head-attention)

Vaswani et al 1706.03762; Qu, Li, Qian 2202.03772; Mikuni, Canelli 2001.05311; Gong et al 2201.08187 for a combination of transformers and attention

Aside:
Alternative to complex architecture

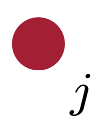
- Advantage of few high-level features:
  -easy to understand and calibrate
  -cheap to evaluate

- Advantage of complex architecture and low-level features: performance

- Can we combine both?

We need a basis

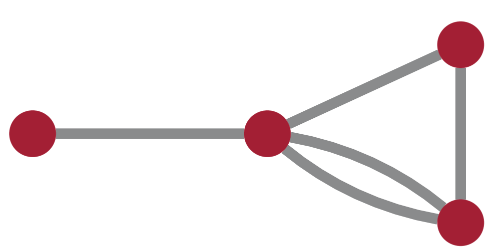
- Energy Flow Polynomials (EFPs) form a basis of jet substructure
  - Nodes: energy fractions
  - Edges: angular distances
- Depending on order considered, too many (e.g 7k) to efficiently train NN (many features work if there is structure, not so much for EFPs)



$$\Longleftrightarrow \sum_{i_j=1}^{M} z_{i_j},$$

$$\Longleftrightarrow \theta_{i_k i_\ell}$$

e.g.

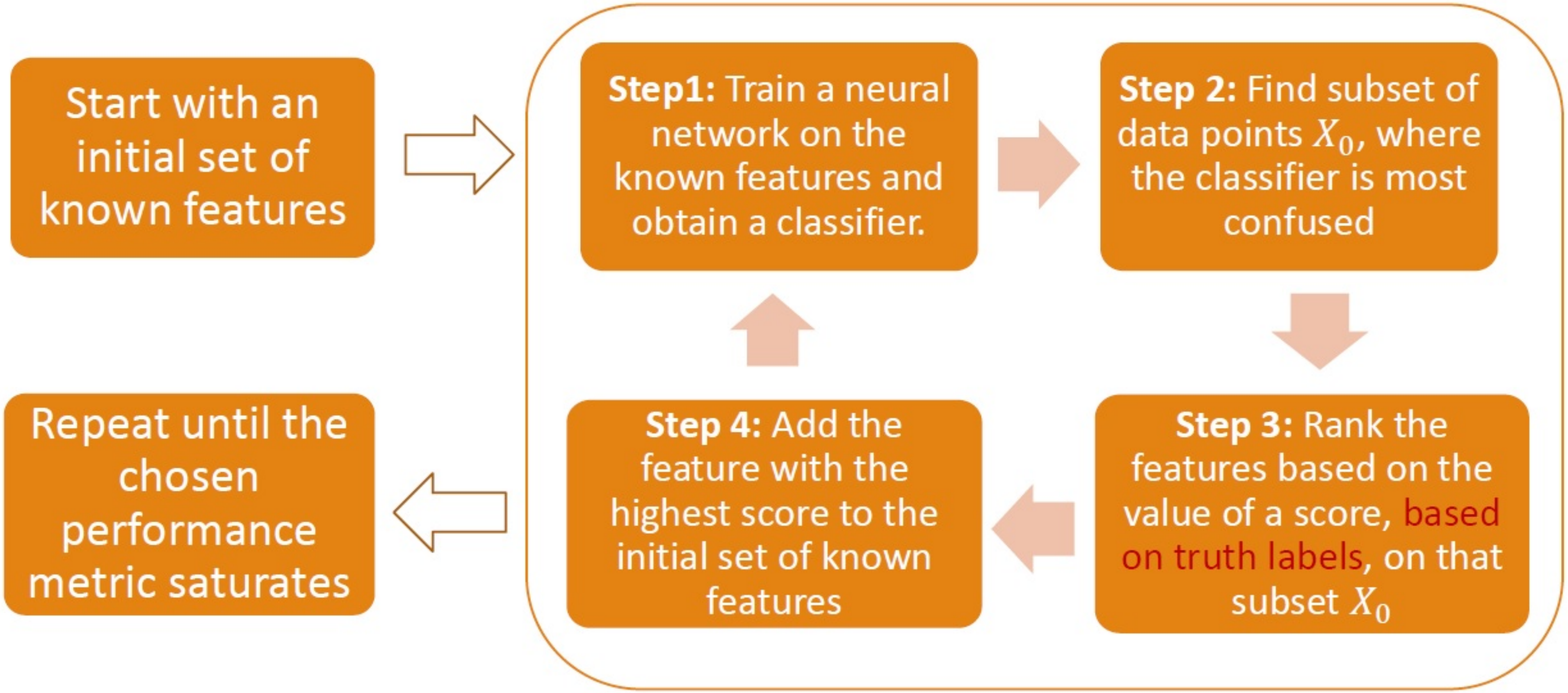


$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} \sum_{i_4=1}^{M} z_{i_1} z_{i_2} z_{i_3} z_{i_4} \theta_{i_1 i_2} \theta_{i_2 i_3} \theta_{i_2 i_4}^2 \theta_{i_3 i_4}.$$

Komiske, Metodiev, Thaler 1712.07124

Looking for optimal feature set

- Solution: Iterative feature selection, again based on DisCo



Das, GK, Shih 2212.00046; Faucett, Thaler, Whiteson, 2010.11998

Already selected features

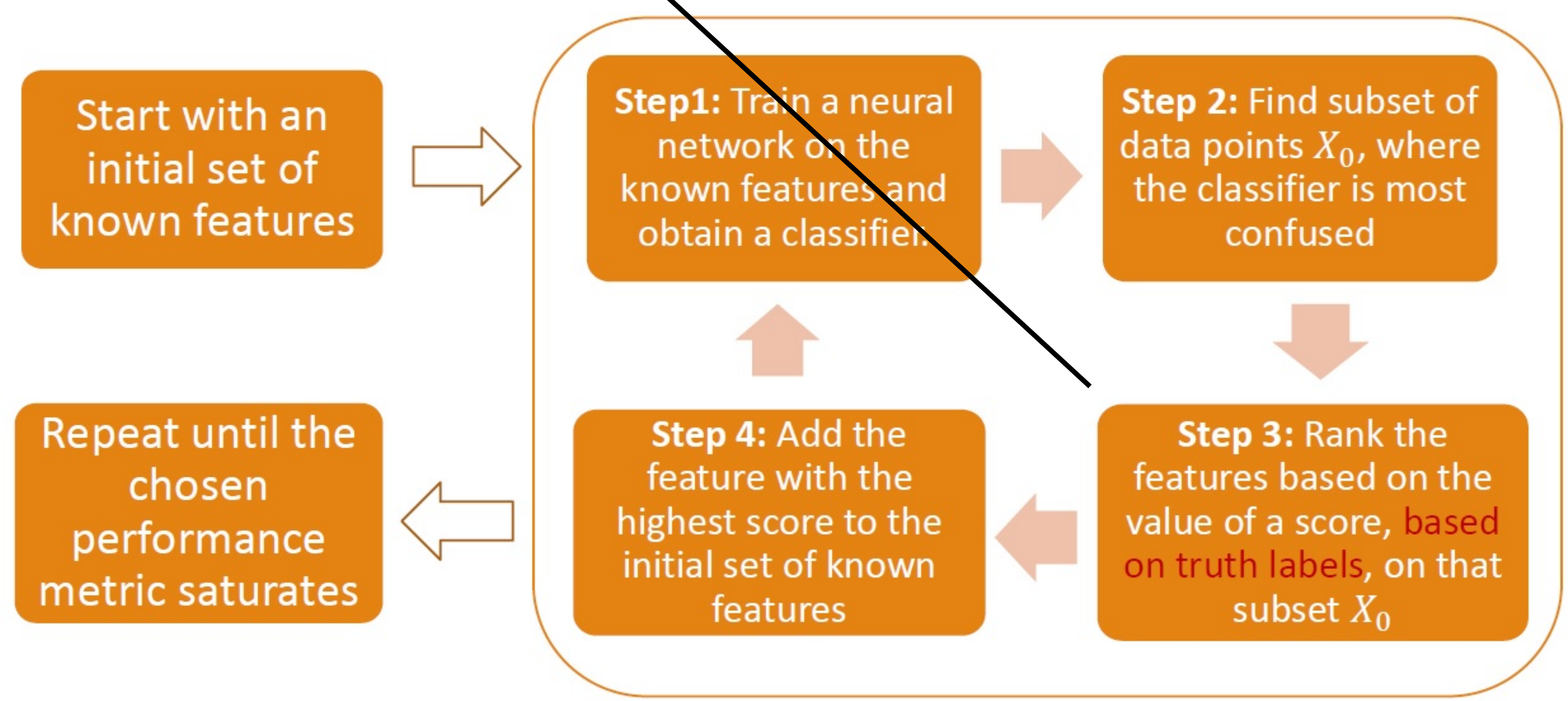
New feature under test

$$\mathcal{X} = \left\{ \left(f_{i_1}(\vec{x}), \ldots, f_{i_n}(\vec{x}), f_i(\vec{x}) \right) \middle| \vec{x} \in X_0 \right\}$$

$$\mathcal{Y} = \{ y_{\text{ref}}(\vec{x}) | \vec{x} \in X_0 \}$$

Truth label or reference classifier

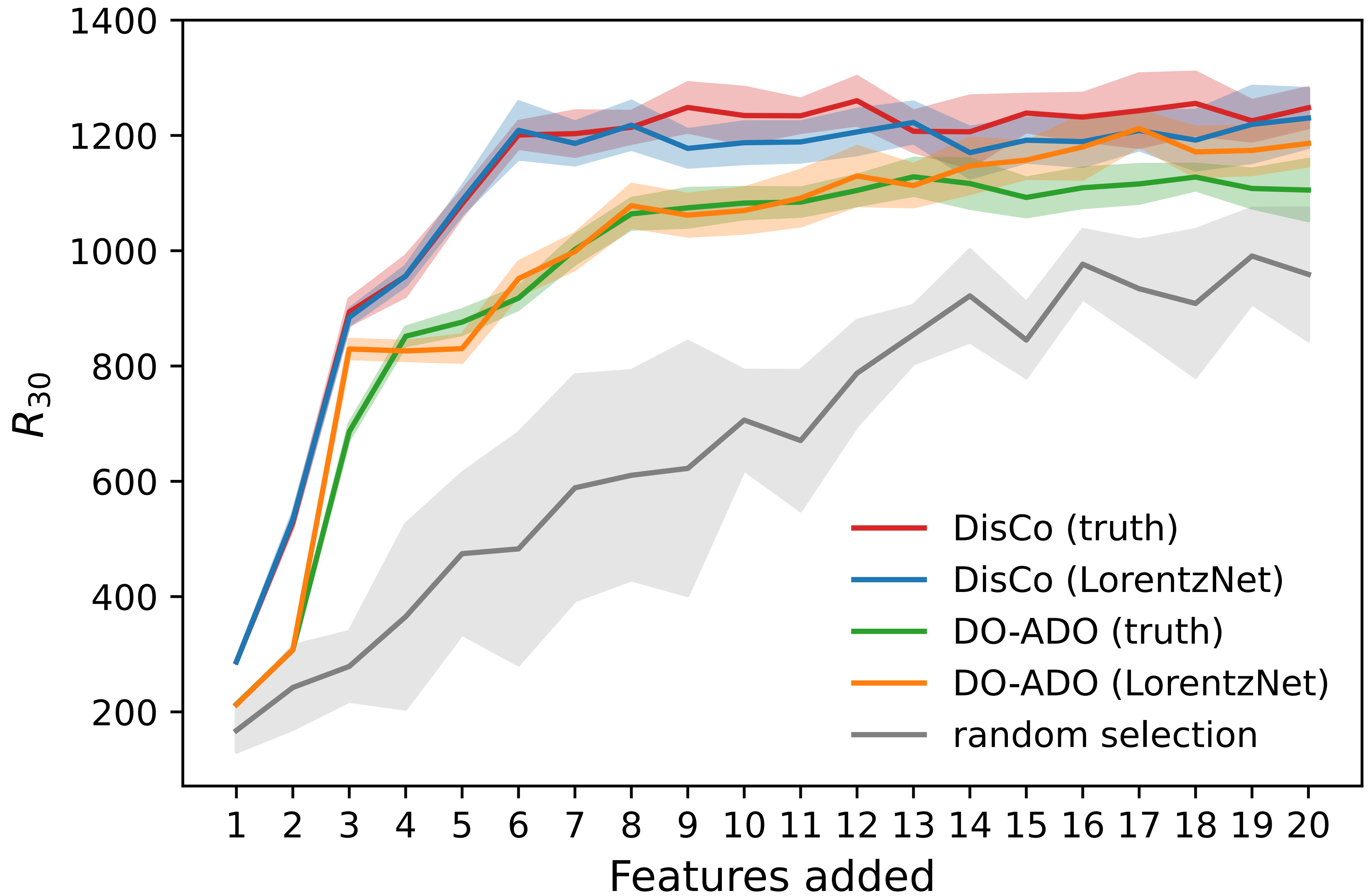
$$s_{f_i} = \text{Affine-DisCo}(\mathcal{X}, \mathcal{Y}).$$



Das, GK, Shih 2212.00046;

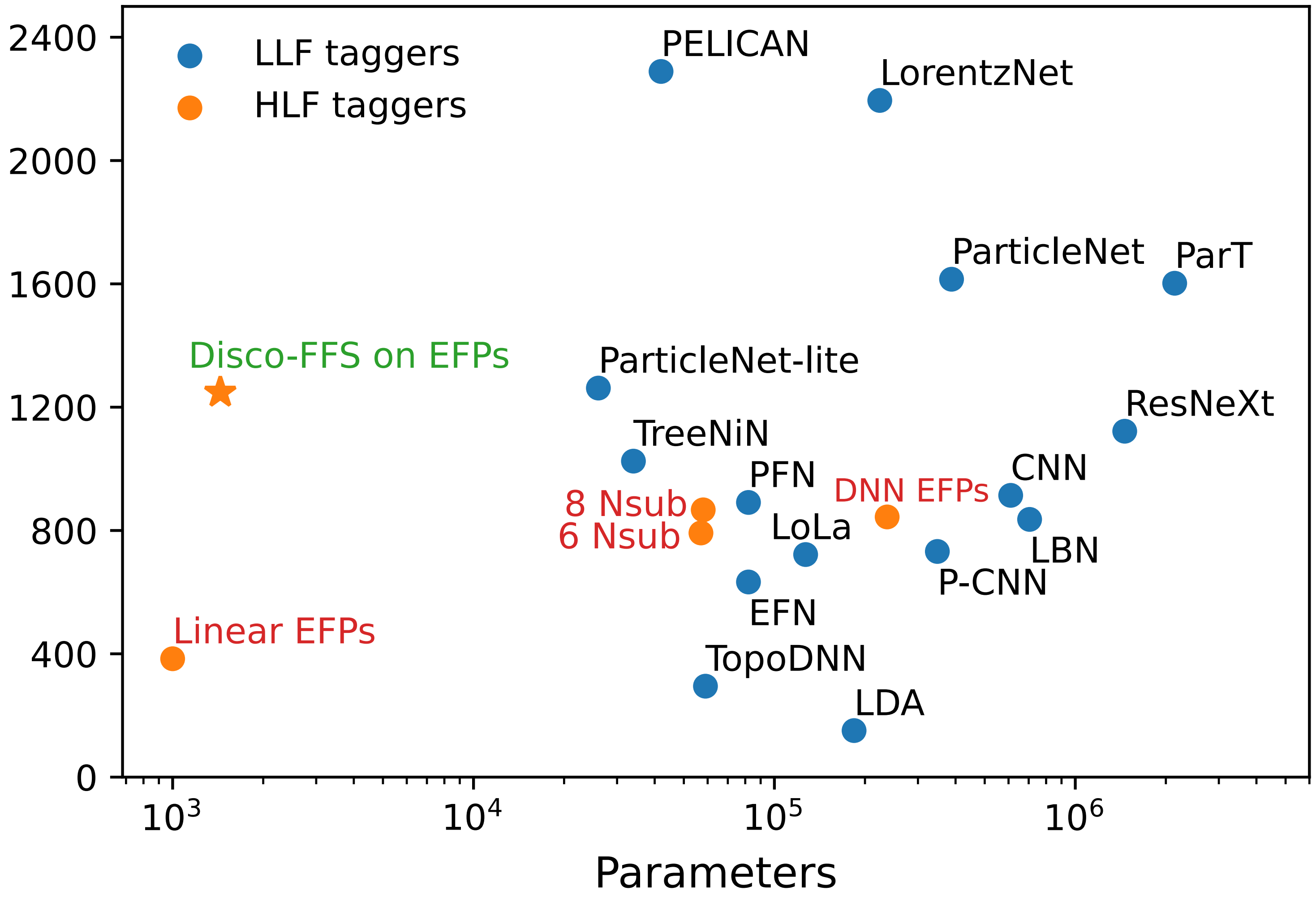
Results

- DiscoFFS find relevant features quicker than alternative feature selection methods



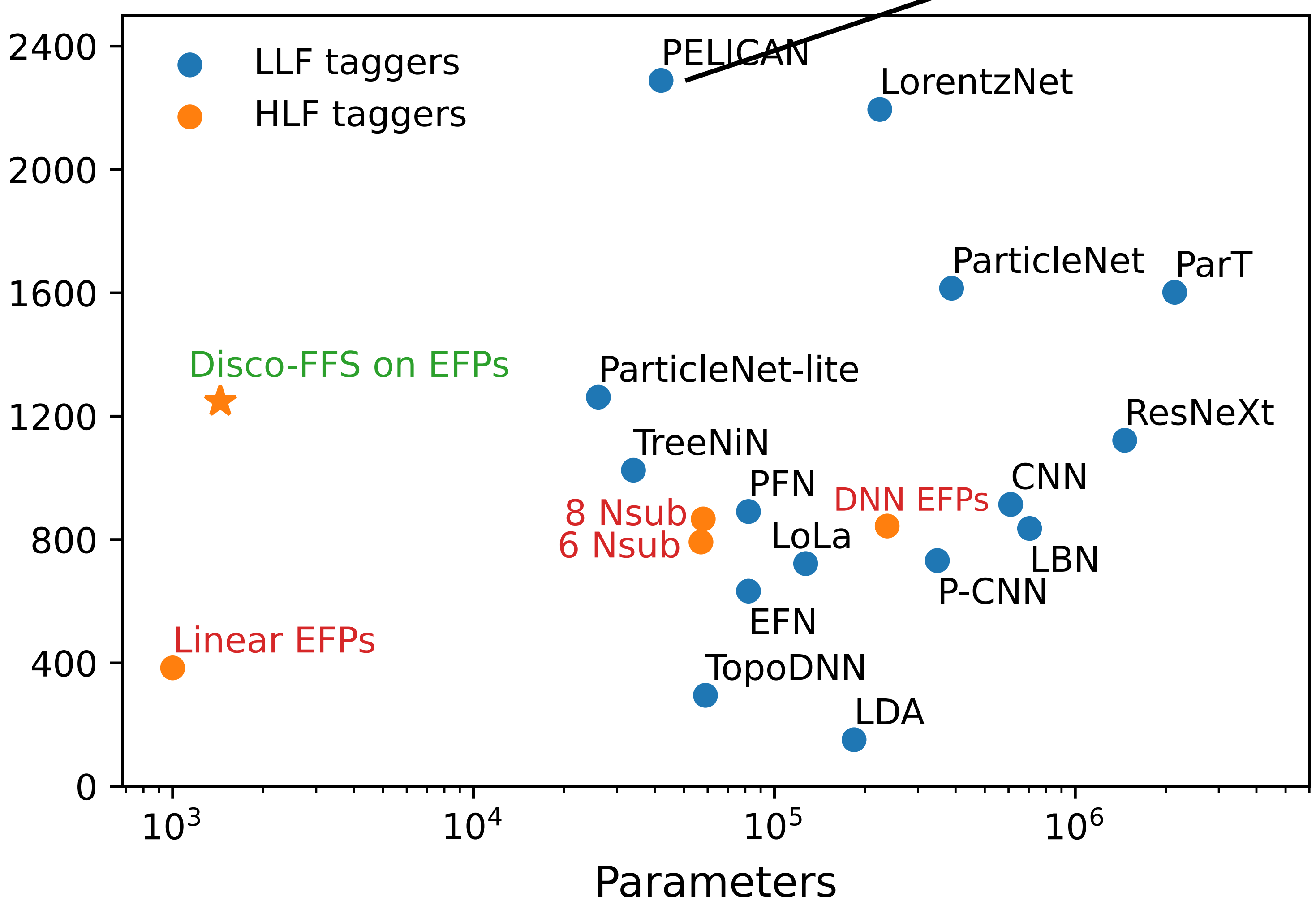
Das, GK, Shih 2212.00046;

Closing



- Many machine learning problems in particle physics
- Large amount of data and symmetries allow broad range of different approaches: from standard ML techniques to methods tailored to HEP data

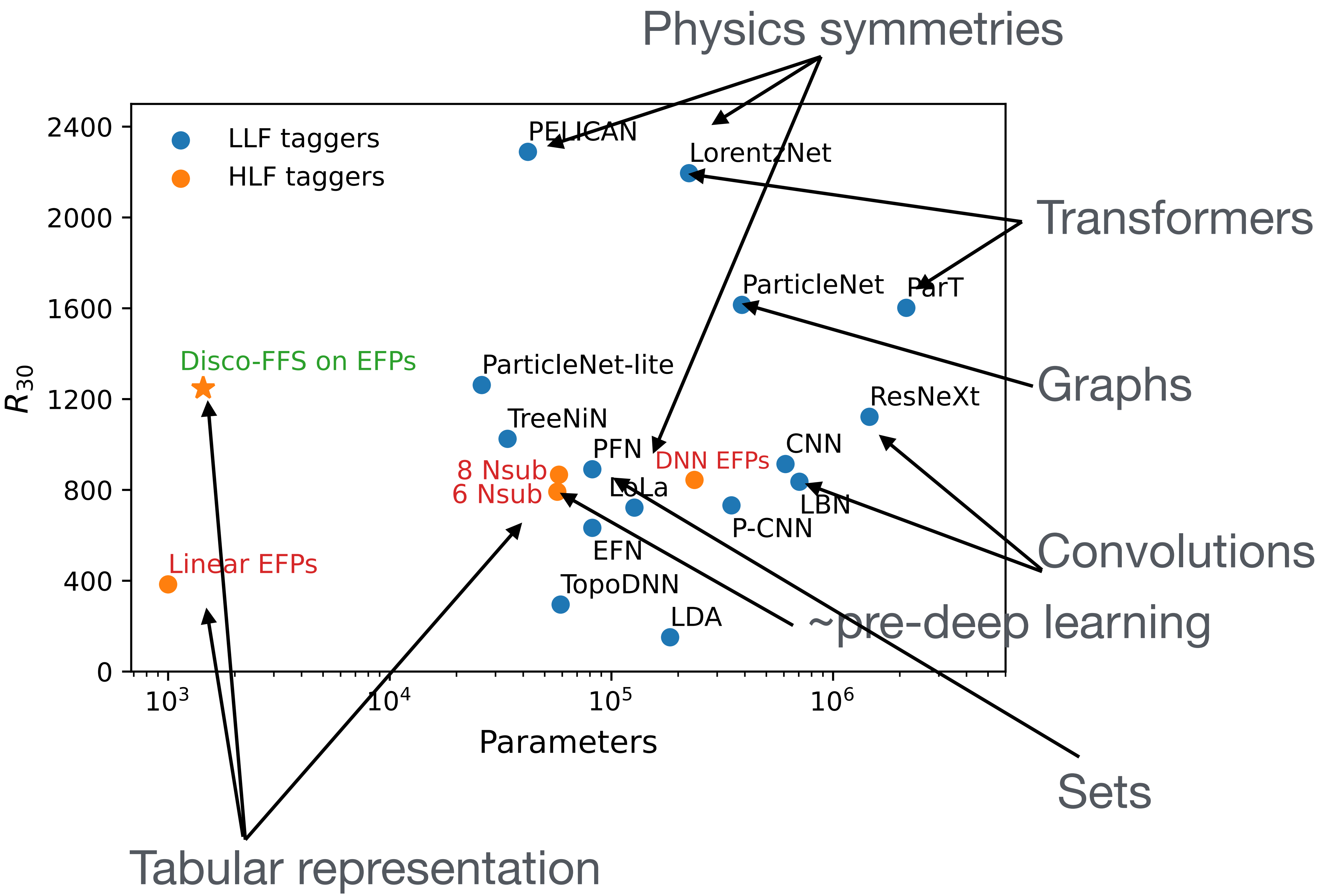
Closing



- Recent result that makes maximal use of symmetries of the problem: restrict learning to permutation invariant mappings between Lorentz tensors

  See 2211.00454

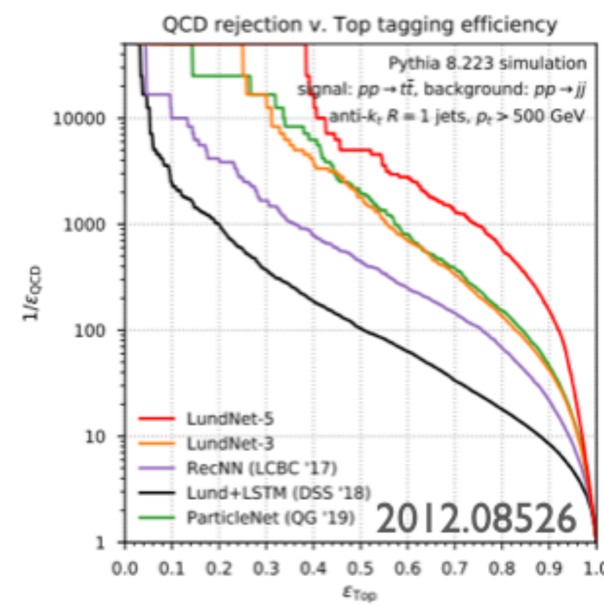
Closing

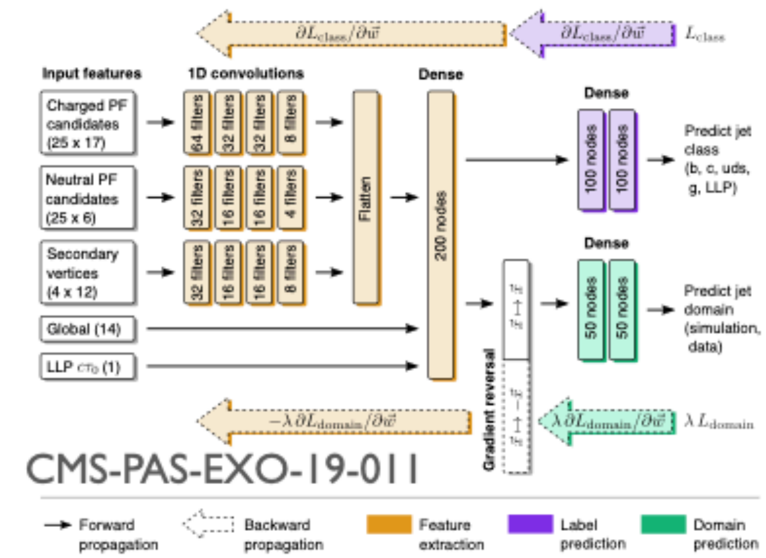


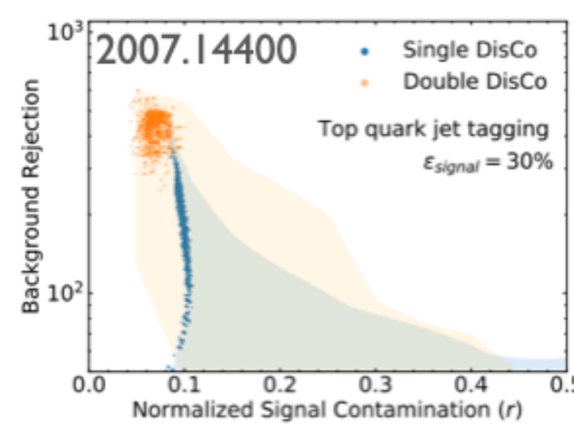
Physics symmetries
Transformers
Graphs
Convolutions
~pre-deep learning
Sets
Tabular representation
LLF taggers
HLF taggers
PELICAN
LorentzNet
ParticleNet
ParT
Disco-FFS on EFPs
ParticleNet-lite
ResNeXt
TreeNiN
PFN
DNN EFPs
CNN
8 Nsub
6 Nsub
LoLa
LBN
P-CNN
EFN
Linear EFPs
TopoDNN
LDA
R_{30}
2400
2000
1600
1200
800
400
0
10^3
10^4
10^5
10^6
Parameters

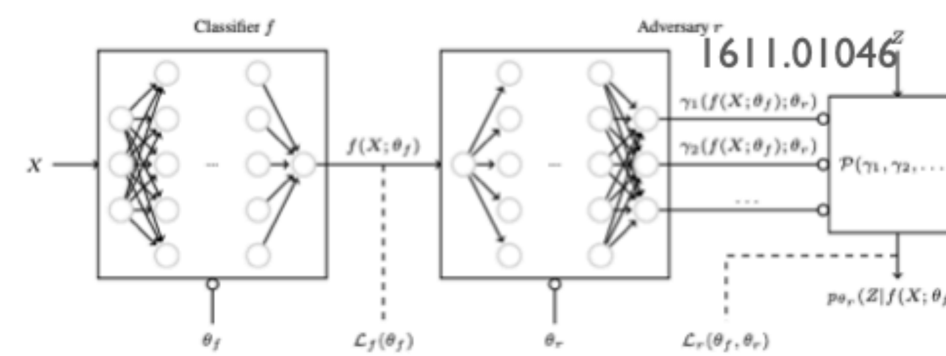
Are we done?

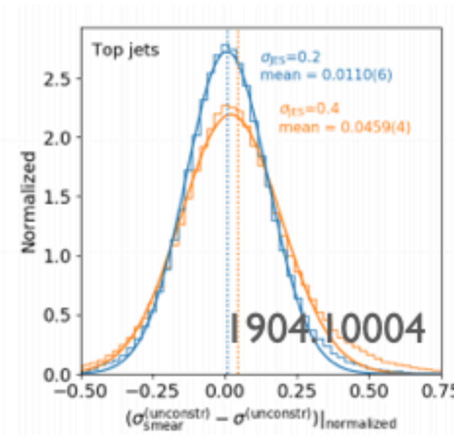
- **No (!)**
- Need:
  - Higger accuracy
    (easy to measure, many results)
  - Better stability
    (domain adaptation issue)
  - More control over uncertainties
  - Resource efficient implementations
  - Experimental integration
  - Theoretical understanding / explainability
  - More holistic learning
  - Problems beyond supervised learning
  - ….

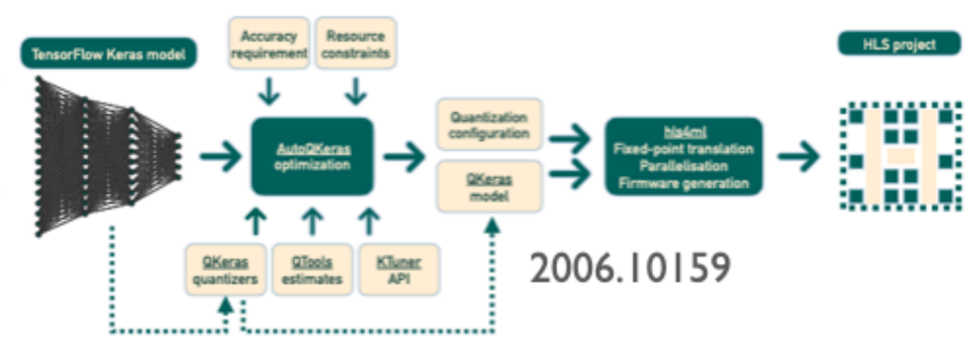












Are we done?

- **Yes**
- Need:
  - Higger accuracy
    (easy to measure, many results)
  - Better stability
    (domain adaptation issue)
  - More control over uncertainties
  - Resource efficient implementations
  - Experimental integration
  - Theoretical understanding / explainability
  - More holistic learning
  - Problems beyond supervised learning
  - ….