Machine Learning in High-Energy Physics: Successes and Future Directions

GRAEME A STEWART

CERN EP-SFT

HSF

HEP Software Foundation

CERN

Overview and Acknowledgements

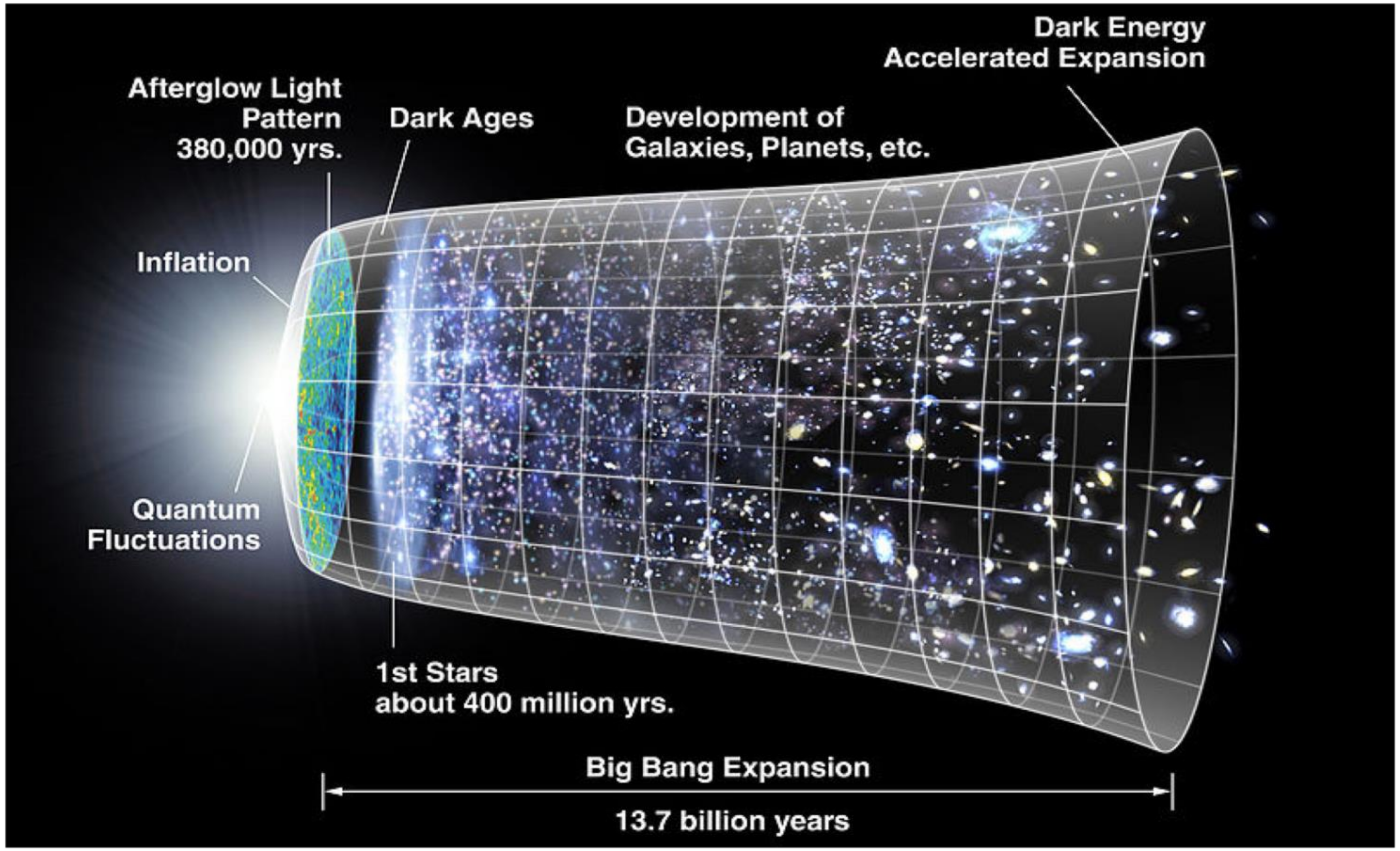
- Particle Physics Overview
  - Future Challenges for HEP
  - Including Computing Hardware
- Classification Problems and Physics Analysis
  - HiggsML Challenge
- HEP Data as Images
- Machine Learning for Generation
- Future Tracking Challenges and GraphNN
- Practical ML
  - Or How Can This All Work…?
- The Future and Conclusions

- Machine learning and artificial intelligence are extremely active topics in High-Energy Physics
- Summarising the work in the field in 45 minutes is essentially impossible
  - So you need to live with my particular selection
- I have drawn on the work or many, many colleagues, to whom I am extremely grateful
  - Of course, all mistakes I own
- In particular I'd like to thank James Catmore, Kyle Cranmer, Sergei Gleyzer, Lorenzo Moneta, David Rousseau, Andi Salzburger, Ariel Schwartzman, Jean-Roch Vilment

Introduction to Particle Physics

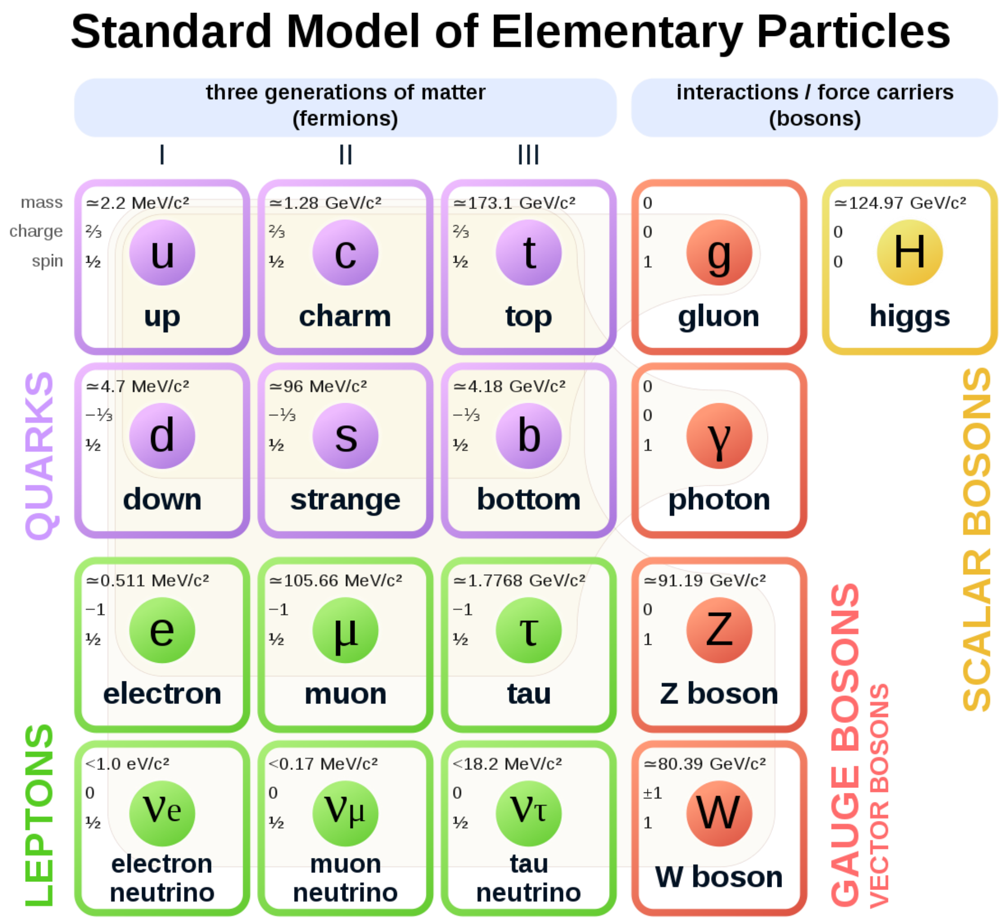
Particle Physics and Big (Data) Science

- Particle physics is trying to help answer questions
  - From the earliest moments of the Big Bang
  - Through to the universe as we see it today



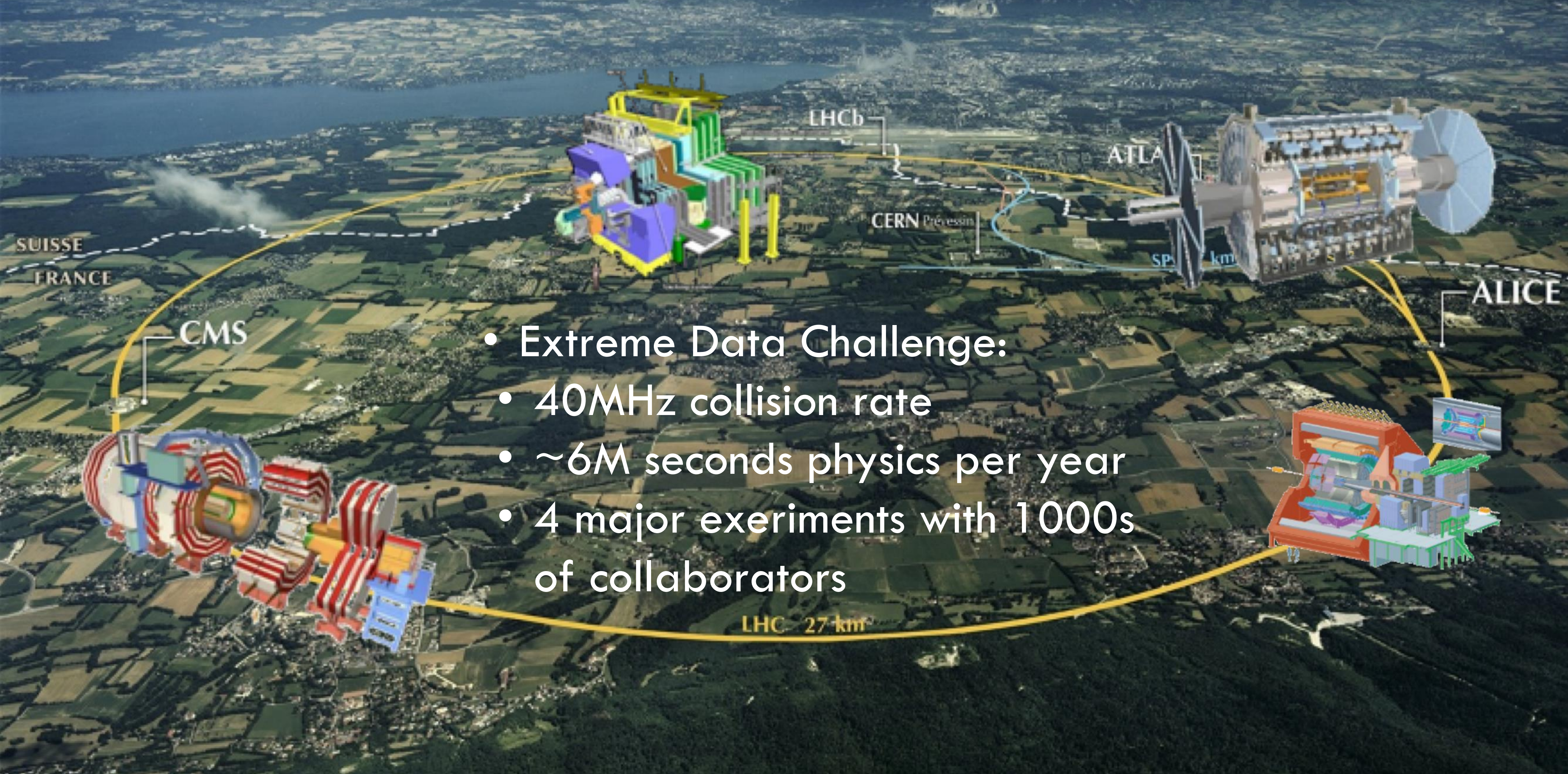
Key Questions

- How good is the Standard Model
  - Properties of the Higgs boson
  - Why is it so light?
- Dark Matter
  - If not SUSY, what is it?
- Dark Energy
- Anti-matter
  - Understand the asymmetry of the universe
- Quark gluon plasma at the Big Bang
  - Starting conditions for the universe we see today



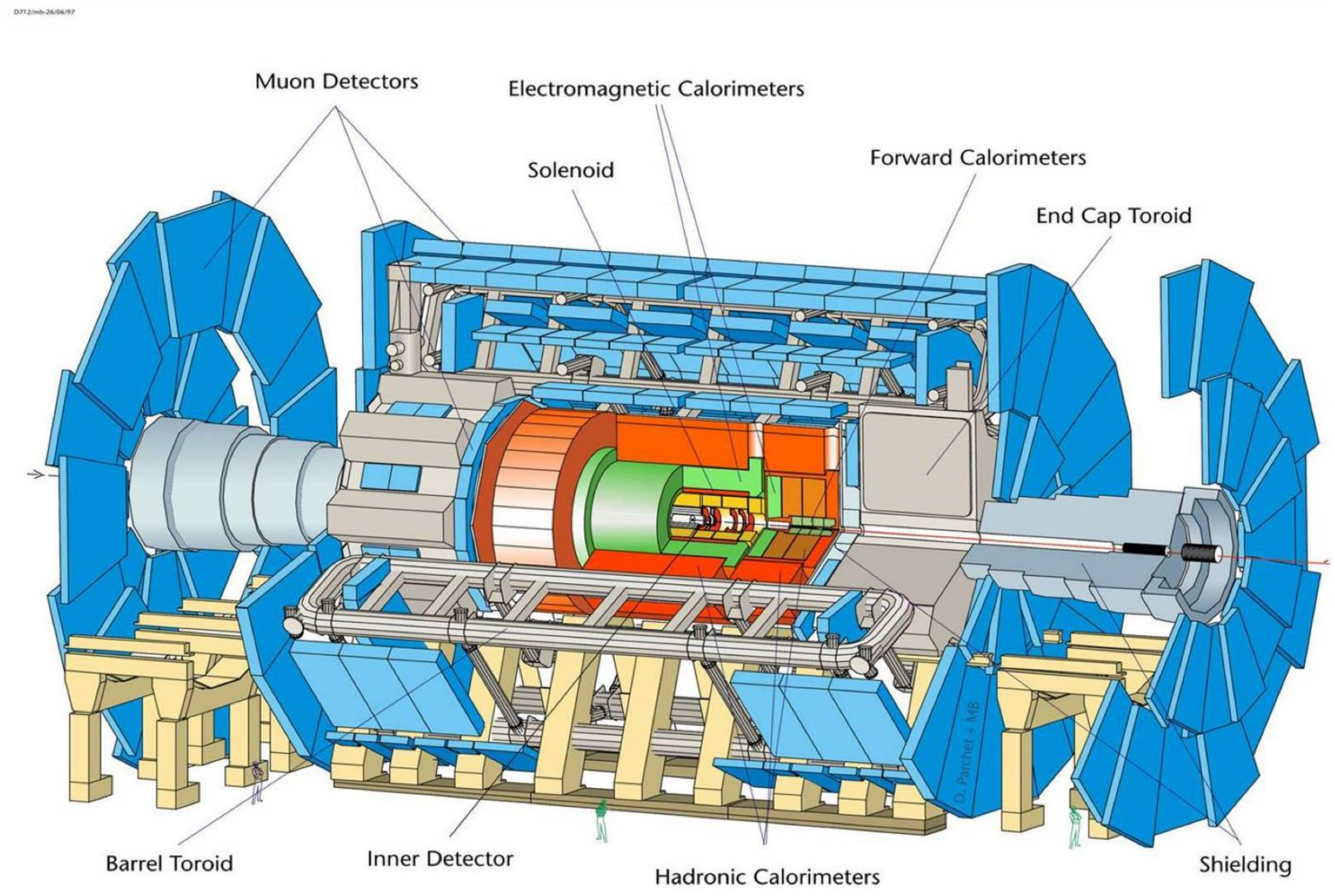
By MissMJ - Own work by uploader, PBS NOVA [1], Fermilab, Office of Science, United States Department of Energy, Particle Data Group, Public Domain, https://commons.wikimedia.org/w/index.php?curid=4286964

The Large Hadron Collider



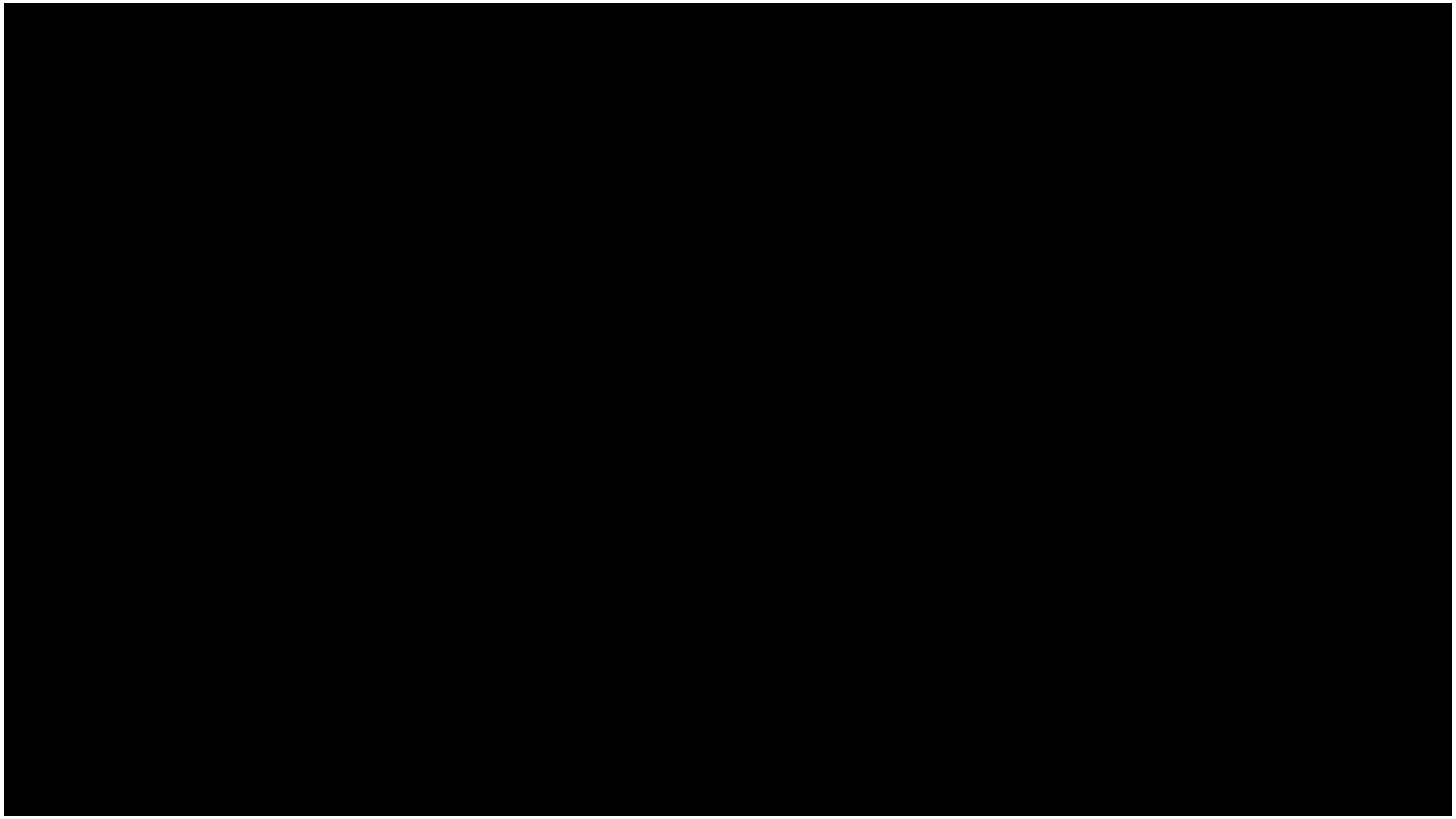
- Extreme Data Challenge:
  - 40MHz collision rate
  - ~6M seconds physics per year
  - 4 major exeriments with 1000s of collaborators

The ATLAS Experiment



Diameter	26m
Length	50m
Weight	7000tn
Field Strength	2T

An LHC Collision Event



Data Rates – now and the future…

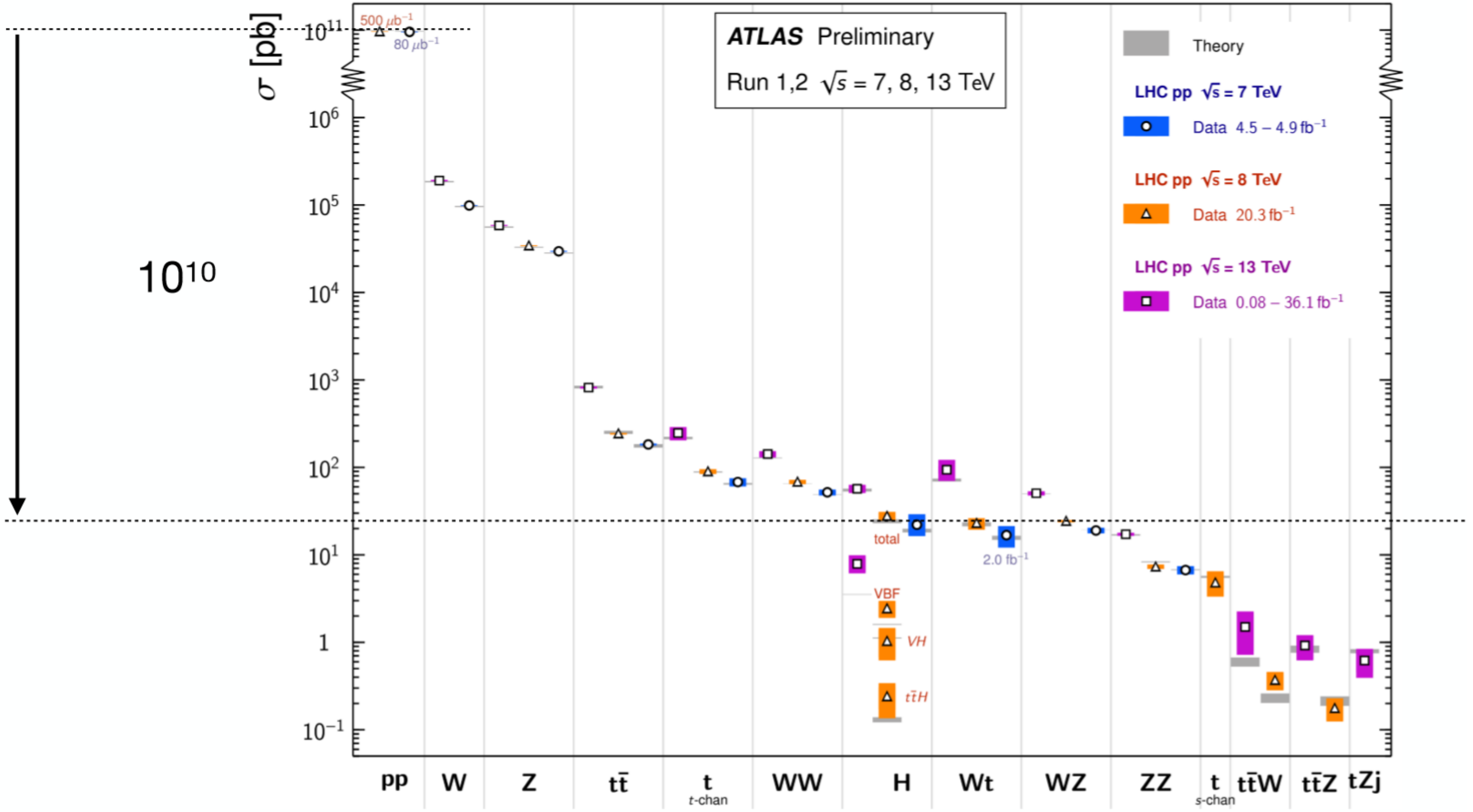
Parameter	ATLAS LHC Run 2 (2015-2018)	ATLAS HL-LHC Run 4 (2026 and beyond)
Collision Rate	40MHz	40MHz
Collision Energy	13TeV	14TeV
Pile-up	40	200
Level 1 Trigger	100kHz	1MHz
Data Accept Rate	1kHz	10kHz
Data Size	1MB	2MB
RAW Data Size (1 Year)	1PB	20PB

Event complexity

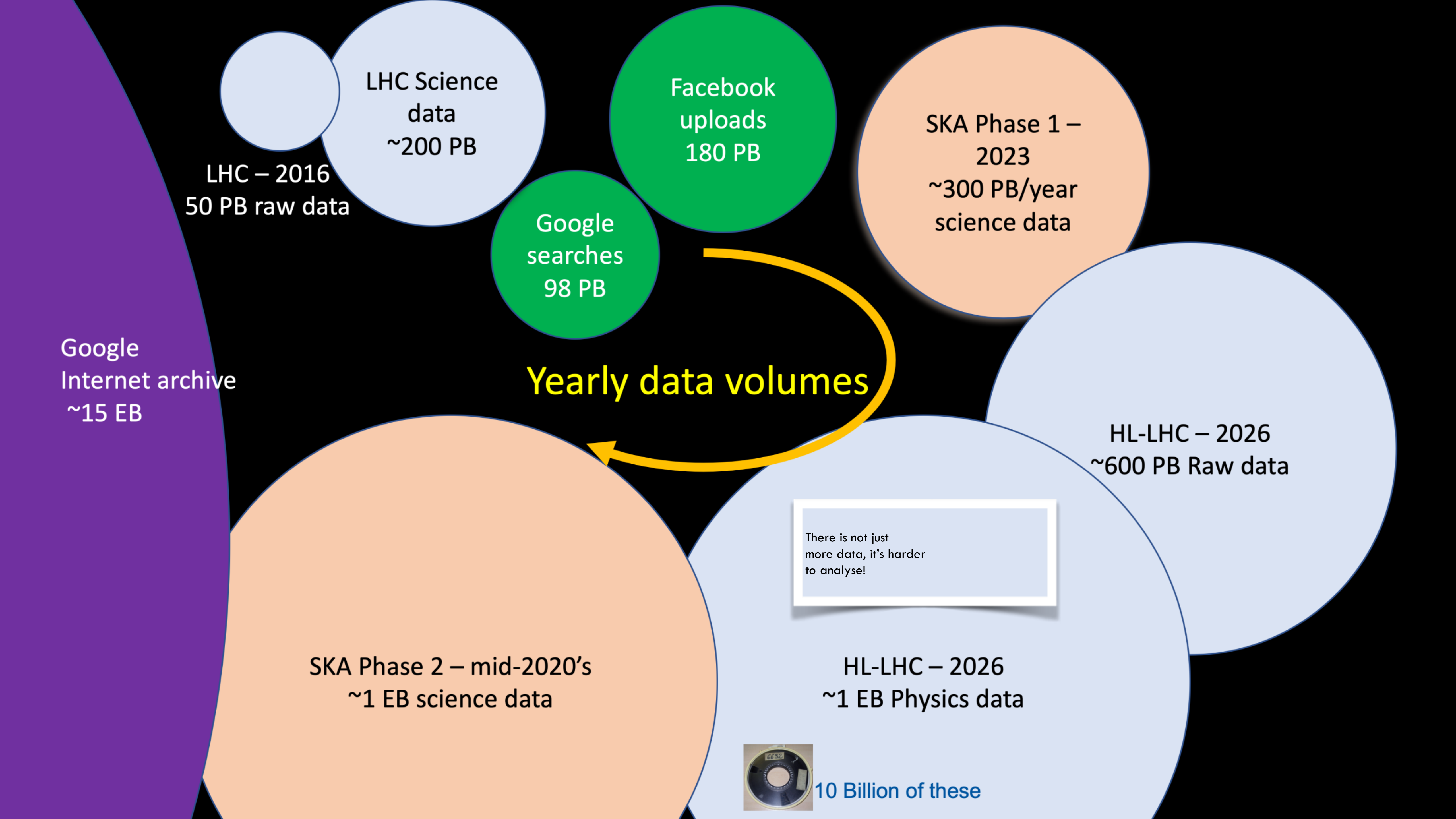
Event rate

- Considering the lifetime of the LHC we are only at the start of the journey
- Roughly x20 more data will be collected by the end of the project

What Are We Looking For?



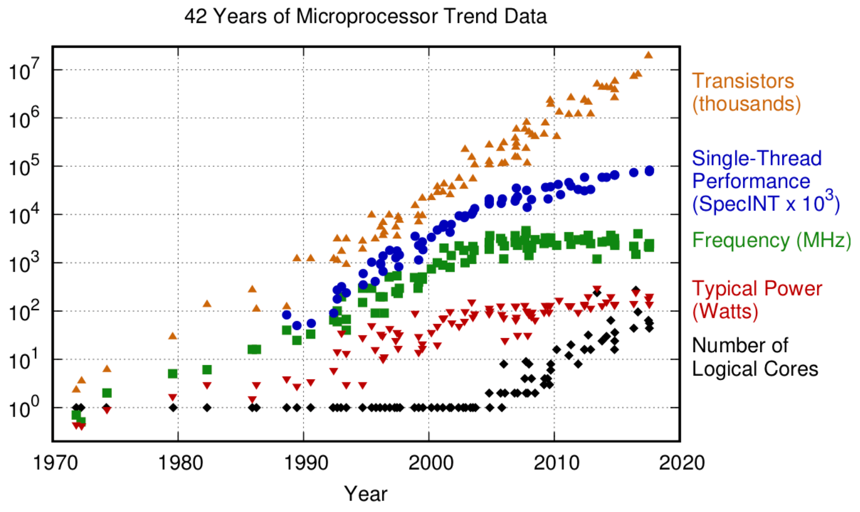
- Interesting physics events for the ATLAS experiment are relatively rare
- Very high backgrounds of well understood physics
- Background rejection has to be high
  - And also be fast
  - This is the experiment trigger system

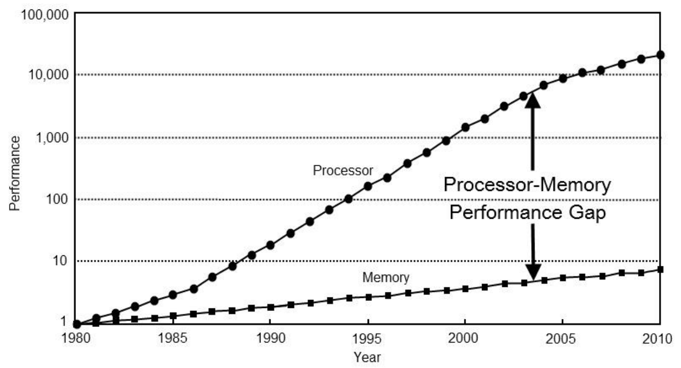


LHC – 2016
50 PB raw data
LHC Science data ~200 PB
Facebook uploads 180 PB
SKA Phase 1 – 2023 ~300 PB/year science data
Google searches 98 PB
Google Internet archive ~15 EB
Yearly data volumes
HL-LHC – 2026 ~600 PB Raw data
There is not just more data, it's harder to analyse!
SKA Phase 2 – mid-2020's ~1 EB science data
HL-LHC – 2026 ~1 EB Physics data
10 Billion of these

Technology Evolution and HEP Computing

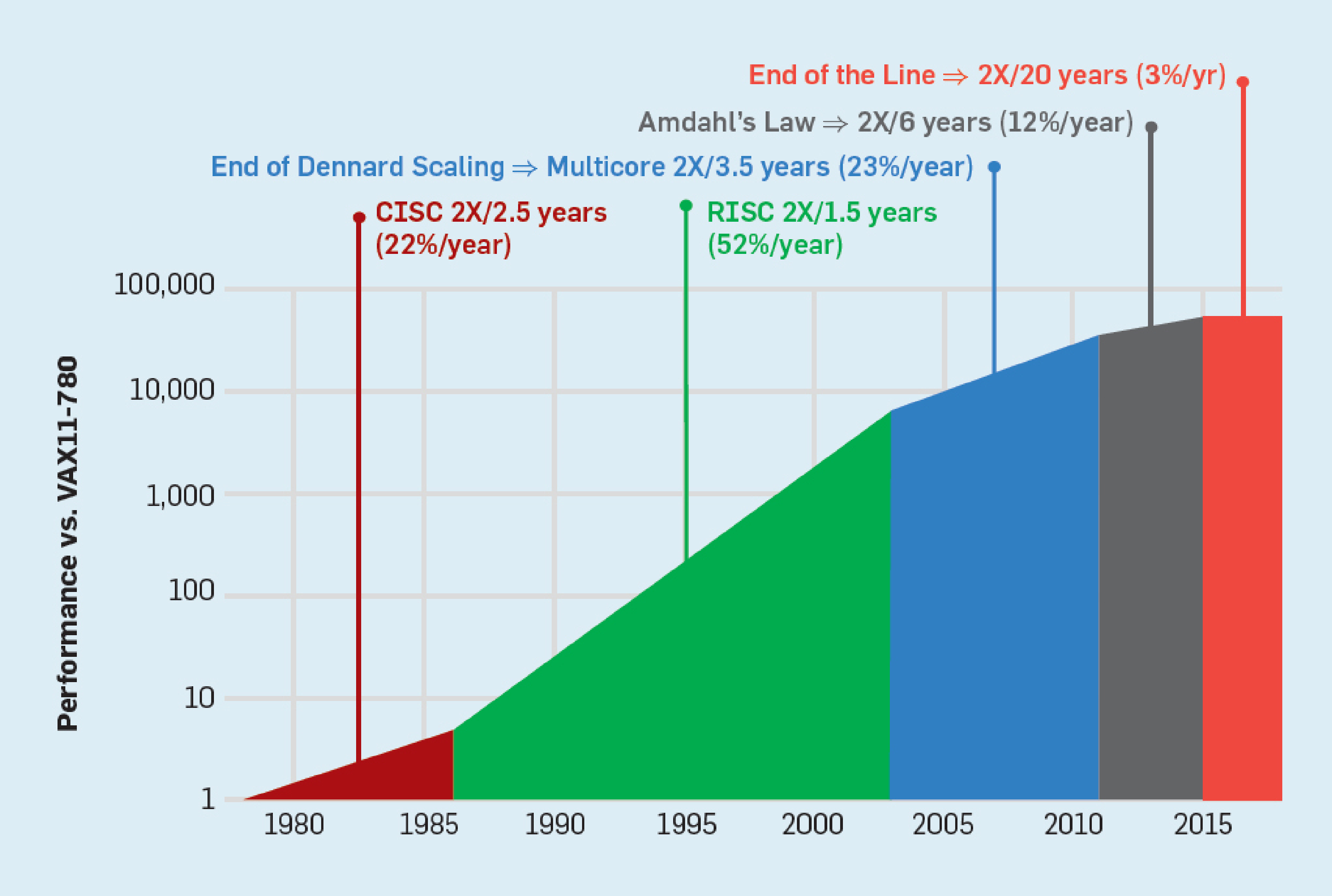
- Moore's Law continues to deliver increases in transistor density
  - But, doubling time is lengthening, and probably the end is in sight…
- Clock speed scaling failed around 2006
  - No longer possible to ramp the clock speed as process size shrinks
  - Leak currents become important source of power consumption
- So we are basically stuck at ~3GHz clocks from the underlying Wm-2 limit
  - This is the Power Wall
  - Limits the capabilities of serial processing
- Memory access times are now ~100s of clock cycles
- This is a serious headache for our software stack



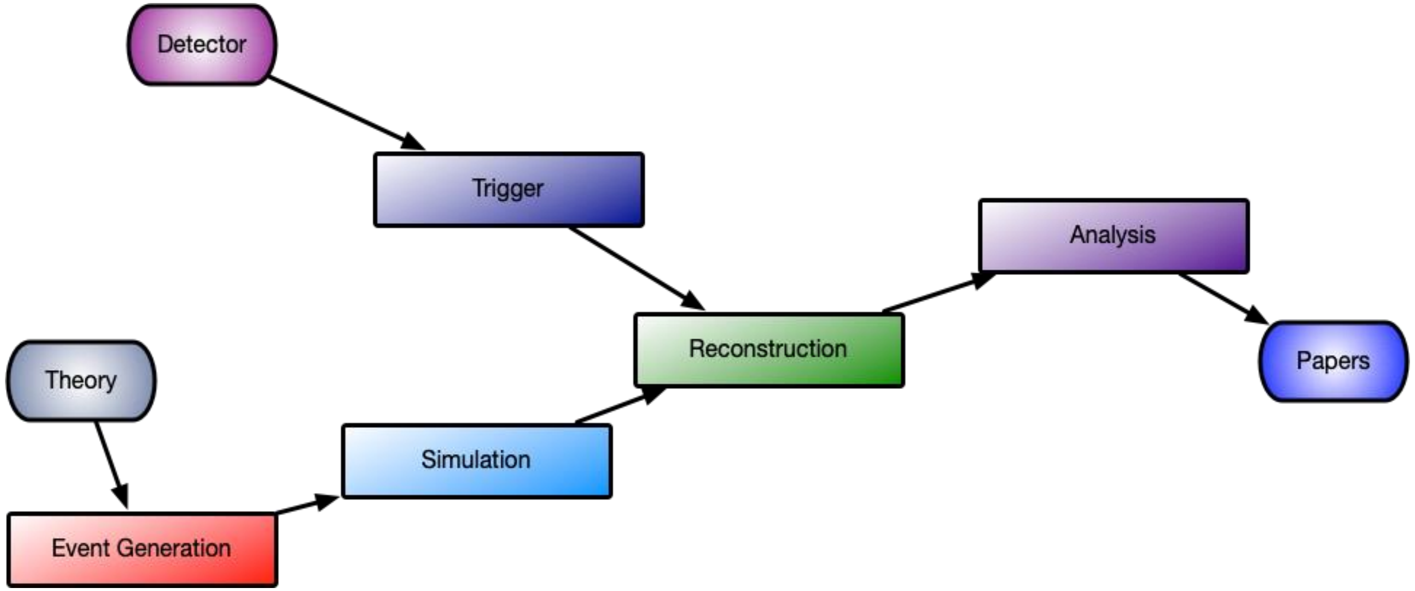


A New Landscape for Data Analysis

- Transition to concurrent execution for HEP software has been hard
  - Even just on CPUs
- Now technology pushes us towards different architectures
  - GPUs have become "standard"
  - FPGAs, TPUs also exist
- Prediction is for ever more diverse architectures
- One of the major drivers of new architectures is modern machine learning
- How can HEP best take advantage of that transition?



HEP Computing Workflow



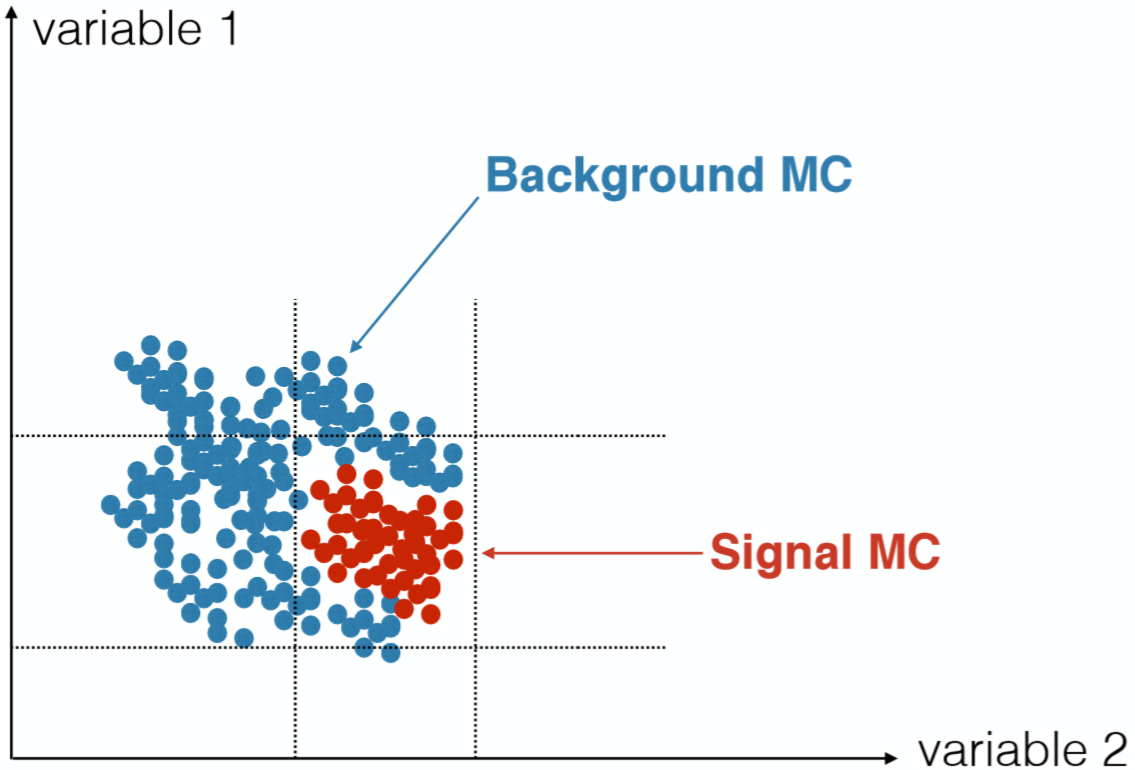
LHC experiment computing scale:

- Use 1M CPU cores every hour of every day
- Store 1000PB of data (600/400PB tape/disk split)
  - We are in the exabyte era already
- Run at 100s of computing centres worldwide
- Make 100PB of data transfers per year (10-100Gb links)

- Far from a homogeneous problem
- Many different phases to HEP computing today
- Mature code, with decades of physics experience built into it
  - About 50M lines of C++ and Python
- Machine learning finds many niches in the current workflow to be useful
  - Doing things **better,** improved outcomes
  - Doing things **faster,** quicker outcomes
  - Doing things **cheaper,** less resources
- Possibility that **advanced machine learning may be a disruptive technology**
  - Change the entire workflow in the future

Classification and Analysis

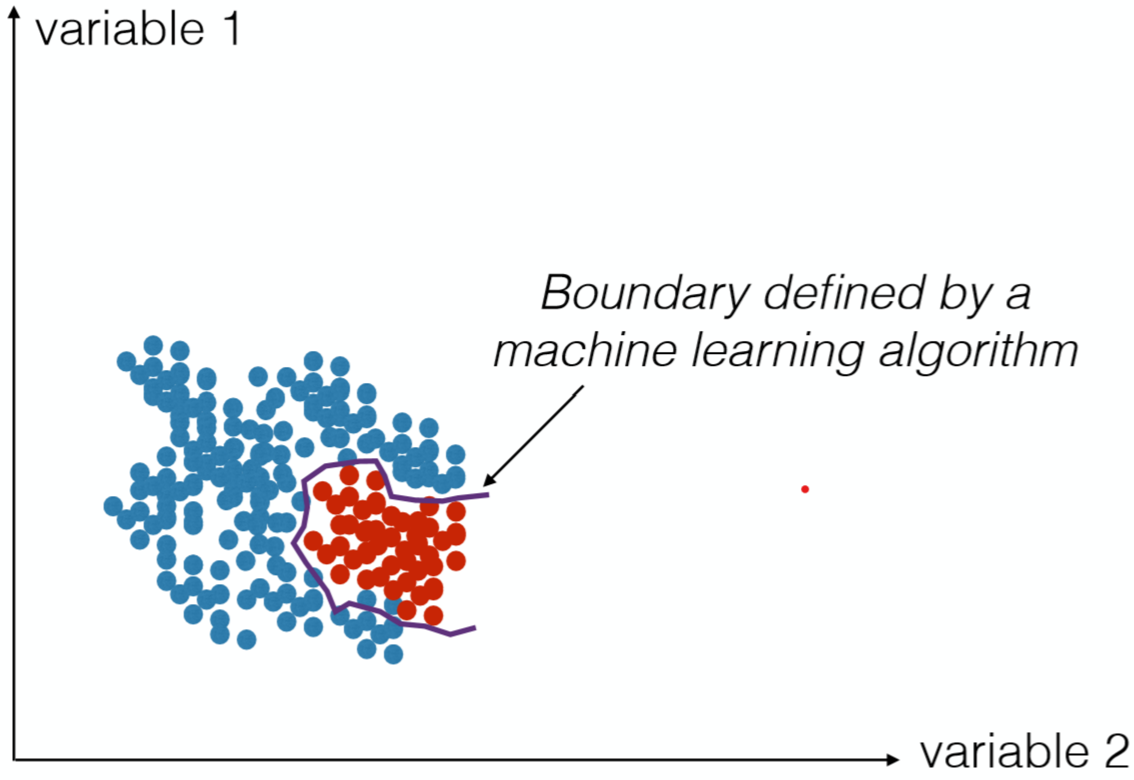
Classification and Analysis



Thanks to James Catmore, Oslo, for these graphics

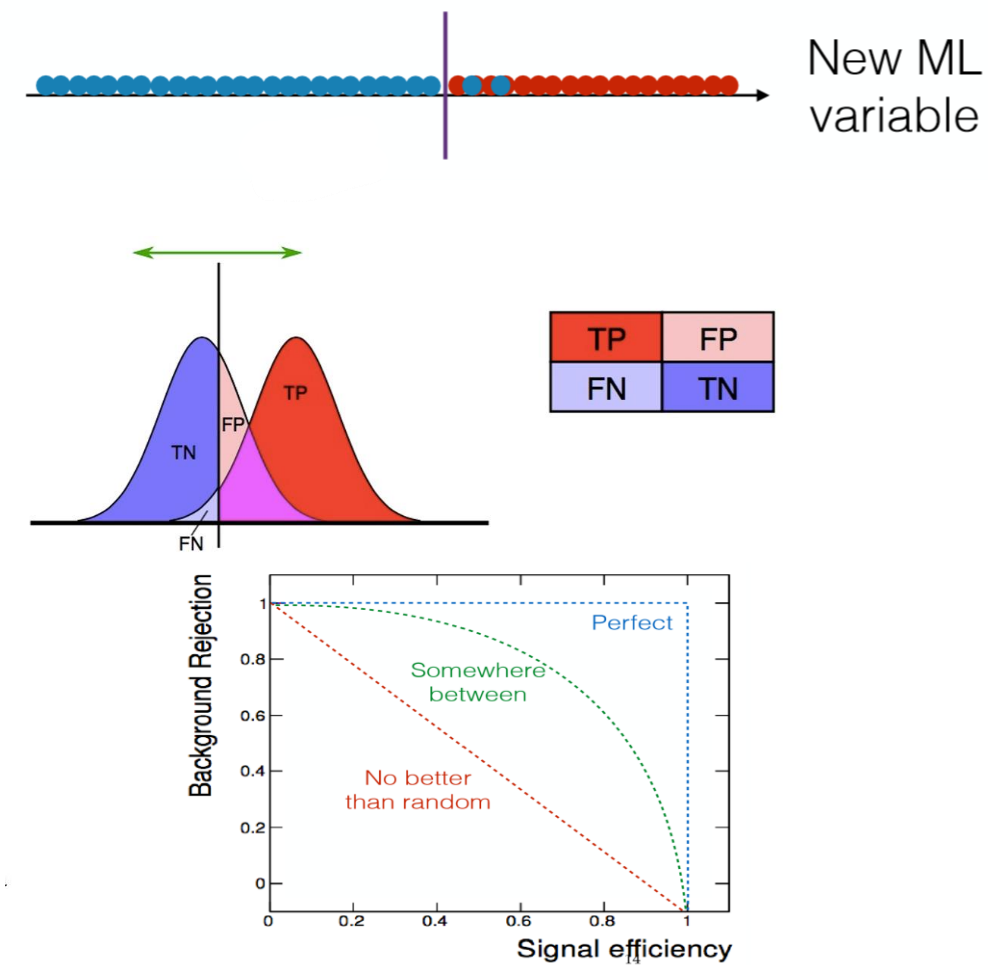
- A large part of physics analysis is signal/background discrimination
  - Signal can be at the 10^{-10} level
  - Reducible and irreducible backgrounds
- For many years we used cut based analysis techniques
  - $p(signal)$ higher if $v_{lo} < v < v_{hi}$
- A lot of effort went into defining higher level variables from lower level reconstruction outputs

Classification and Analysis



- Multi-variate analysis was a way to improve this
- Define signal and background from considering multiple variables at the same time
- Shallow neural networks often used
- Boosted decision trees became extremely popular in the field
  - Conceptually easy to understand the algorithm
  - Performance as good as shallow NNs
  - Efficient to store and quick to evaluate
  - We have a standard toolkit for this
    - TMVA toolkit in ROOT

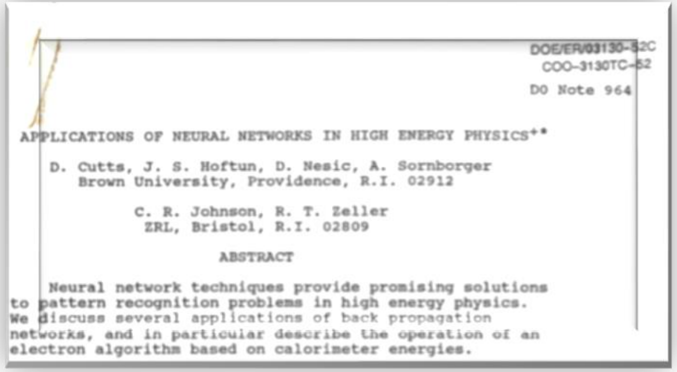
Classification and Analysis



- Effectively ML defines a new discrimination variable
  - $v_{ml} = f(v_1, v_2, v_3, v_4, v_5, \ldots)$
- Improvement in the discrimination power measured by the usual Receiver-Operator Curve (ROC)
  - Although nothing will beat the Baysiean Limit
  - But ML very useful when underlying PDFs are not known
- These techniques have were a significant success at both the Tevatron and at the LHC

Multi-Variate Success for HEP

Top quark mass measurement @Tevatron	Shallow NNs, BDTs
Single top quark discovery @ Tevatron	Shallow NNs, BDTs
Higgs discovery ($H \rightarrow \gamma\gamma$) @ CMS	BDT
Observation of $H \rightarrow bb$ @ ATLAS, CMS	BDT
Observation of $Bs \rightarrow \mu\mu$ @ ATLAS, CMS, LHCb	BDT
Observation of associated Higgs and top quark pair production ("ttH") @ ATLAS, CMS	BDT (XGBoost @ ATLAS)
Jet flavour tagging	Shallow NNs, BDT, Recurrent NNs


DOE/ER/03130-52C
COO-3130TC-52
D0 Note 964

APPLICATIONS OF NEURAL NETWORKS IN HIGH ENERGY PHYSICS**

D. Cutts, J. S. Hoftun, D. Nesic, A. Sornborger
Brown University, Providence, R.I. 02912

C. R. Johnson, R. T. Zeller
ZRL, Bristol, R.I. 02809

ABSTRACT

Neural network techniques provide promising solutions to pattern recognition problems in high energy physics. We discuss several applications of back propagation networks, and in particular describe the operation of an electron algorithm based on calorimeter energies.

- Papers from at least 30 years ago
  - As old as the world wide web!
- A huge success for the field

Getting Better: Machine Learning Challenge

Higgs Discovery

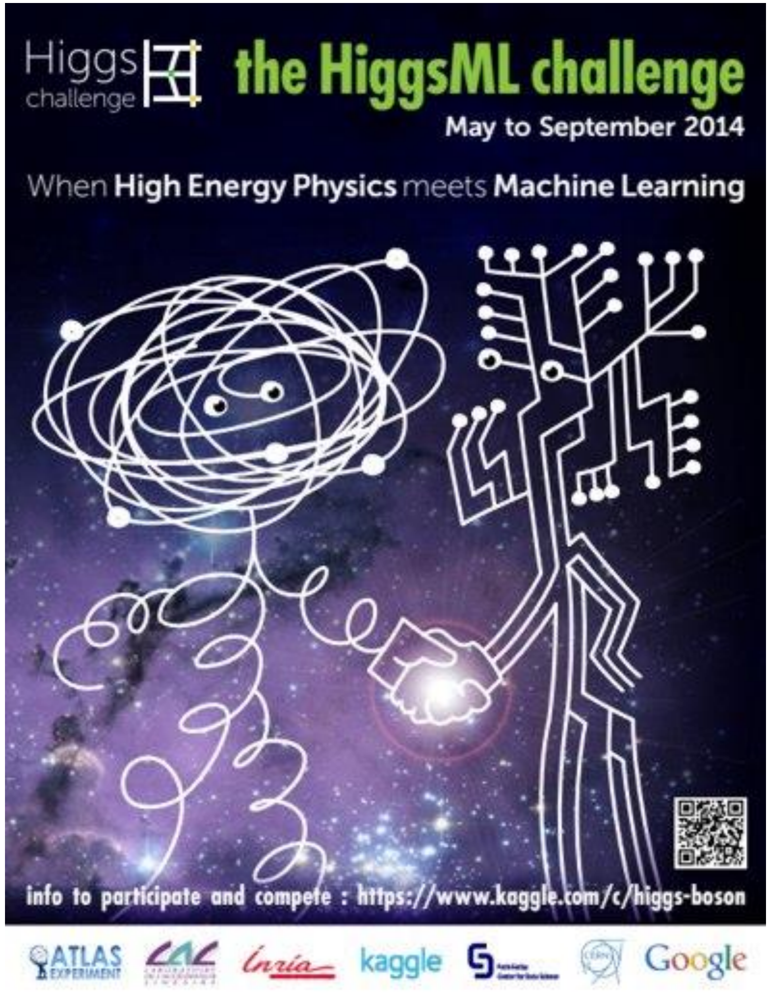
- Discovery of the Higgs boson in 2012 has been the greatest achievement of the LHC programme so far
- Machine learning provided significant increase in sensitivity
- If we had not used such techniques, we would have needed to collect significantly more data
  - LHC data is expensive!
- However, one can ask the question could we have done even better…?

Table 1 | Effect of machine learning on the discovery and study of the Higgs boson

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

Machine learning at the energy and intensity frontiers of particle physics, https://doi.org/10.1038/s41586-018-0361-2

Higgs Machine Learning Challenge

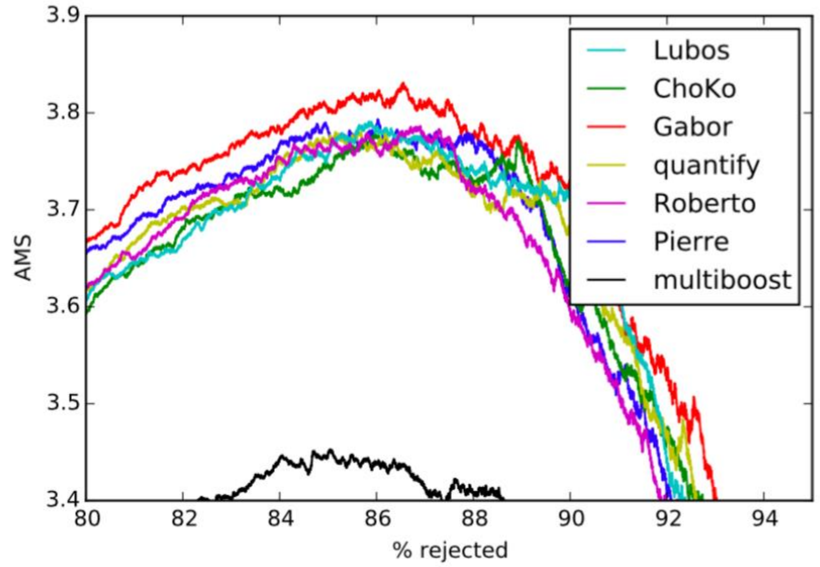


- Hosted on Kaggle platform
- Used 800k ATLAS simulated events
  - $H \rightarrow \tau\tau$ plus background events
  - 30 features per-event
    - Mix of low level reconstruction and derived features
- Huge excitement on the platform
  - 1785 teams took part
  - Physicists and data scientists
  - 35772 solutions were submitted
  - This was the largest challenge that they had hosted at the time
- Simple solution (untuned TMVA) beaten on day 1
- ‘Reasonable’ solution (multi-boost benchmark) beaten by day3

HiggsML Outcomes

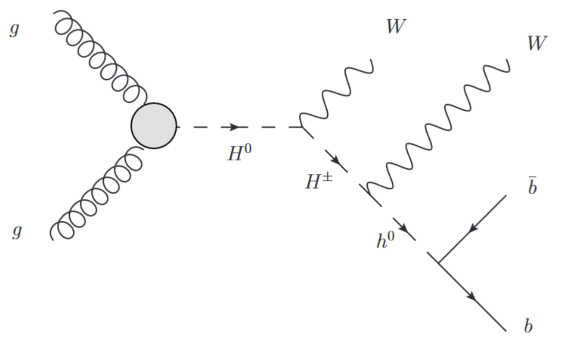
- Winners were from the computer science domain, not physicists
  - Deep Neural Network solution was the winner
  - XGBoost implementation proved to be highly successful
    - Winner of the HEP meets ML prize
    - Has gone on to be used in subsequent HEP analyses
- We learned a lot from this
- Nice boost for public interest in HEP!

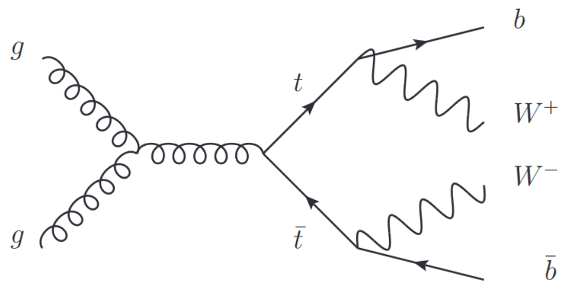
Rank	Team	Score	Entries	Method
1	Gábor Melis	3.80581	100	DNN
2	Tim Salimans	3.78913	57	RGF and meta ensemble
3	nhlx5haze	3.78682	254	Ensemble of neural networks
8	Luboš Motl's team	3.76050	589	XGboost and Intensive feature engineering
31	Mymo	3.72594	73	Ensemble of cascades and non-cascaded models
45	Crowwork	3.71885	94	XGBoost Tuned
782	Eckhard	3.49945	29	TMVA Tuned
902	Benchmark	3.40488	NA	MultiBoost
991	Benchmark	3.19956	NA	TMVA



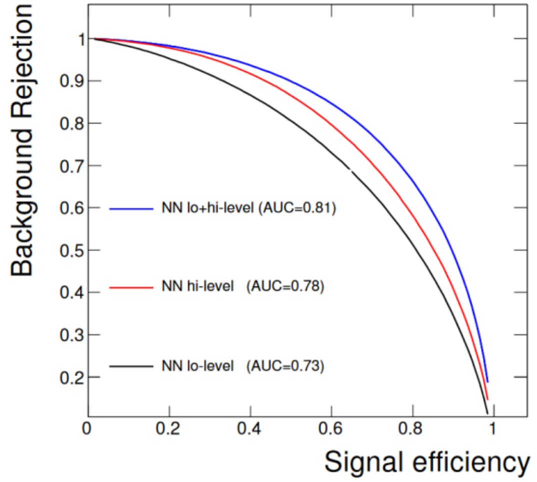
Lessons for Analysis

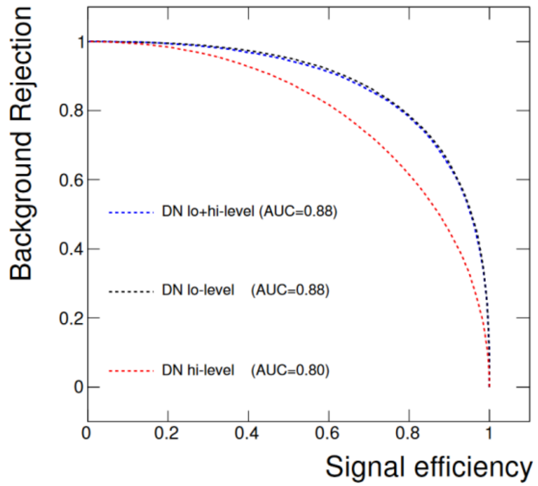
Baldi, Sadowski, Whiteson
https://arxiv.org/abs/1402.4735





- Minimal super-symmetric model at the LHC
  - Signal $H_0 \rightarrow WWbb$ vs. background $tt \rightarrow WWbb$
- Low level variables
  - 4-momentum vector
- High level variables
  - Pair-wise invariant masses
- Deep NN outperforms NN, and does not need high level variables
- **DNN learns the physics…?**

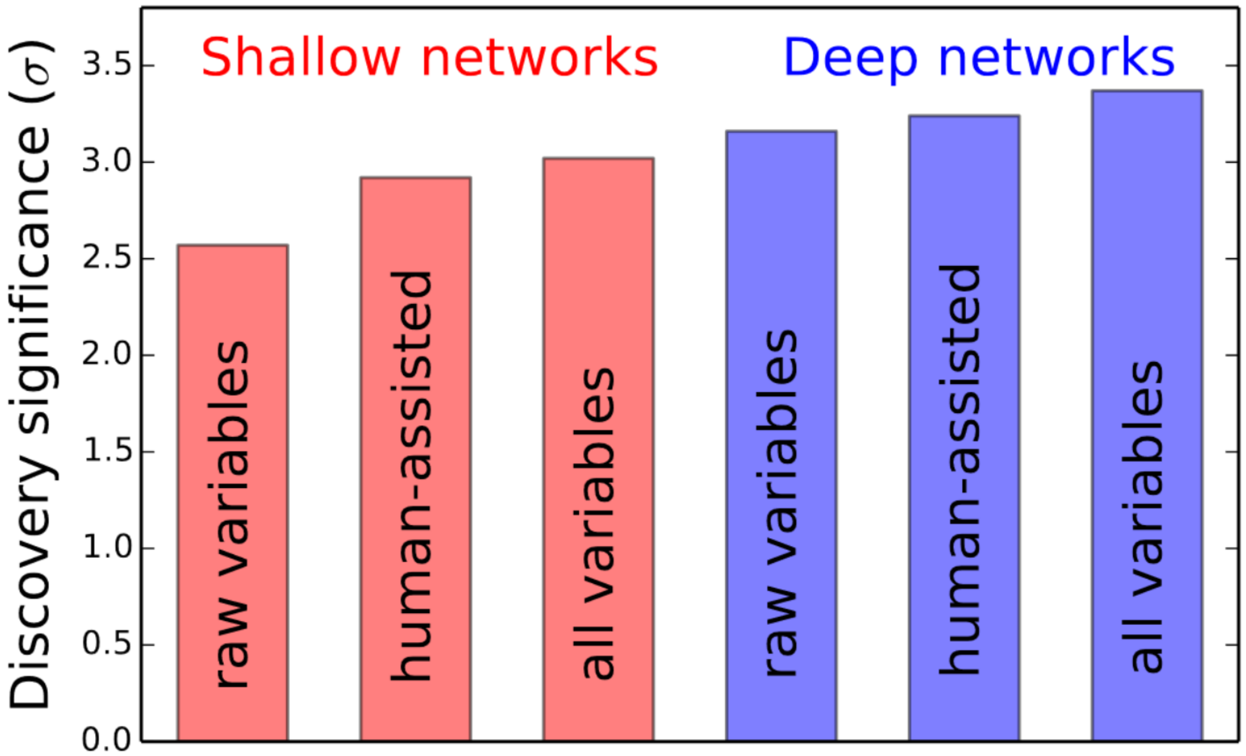




Thanks to David Rousseau, LAL for this example

$H \rightarrow \tau\tau$ Do-over

Baldi, Sadowski, Whiteson
https://arxiv.org/abs/1410.3469

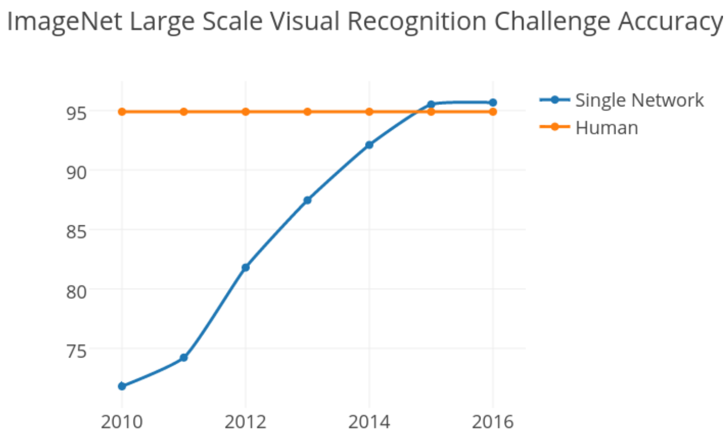


- LHC $H \rightarrow \tau\tau$ analysis with $Z \rightarrow \tau\tau$ background
  - 100M fast simulation training events from Delphes (this is a lot!)
- Low level variables
  - 4 momentum
- High level variables
  - transverse mass, delta R, centrality, jet variables, etc…
- Here the Deep Neural Network consistently outperforms the shallow NN
- **But high level variables always improve the significance**

Thanks to David Rousseau, LAL for this example

Image Recognition Problems

Image Recognition in HEP



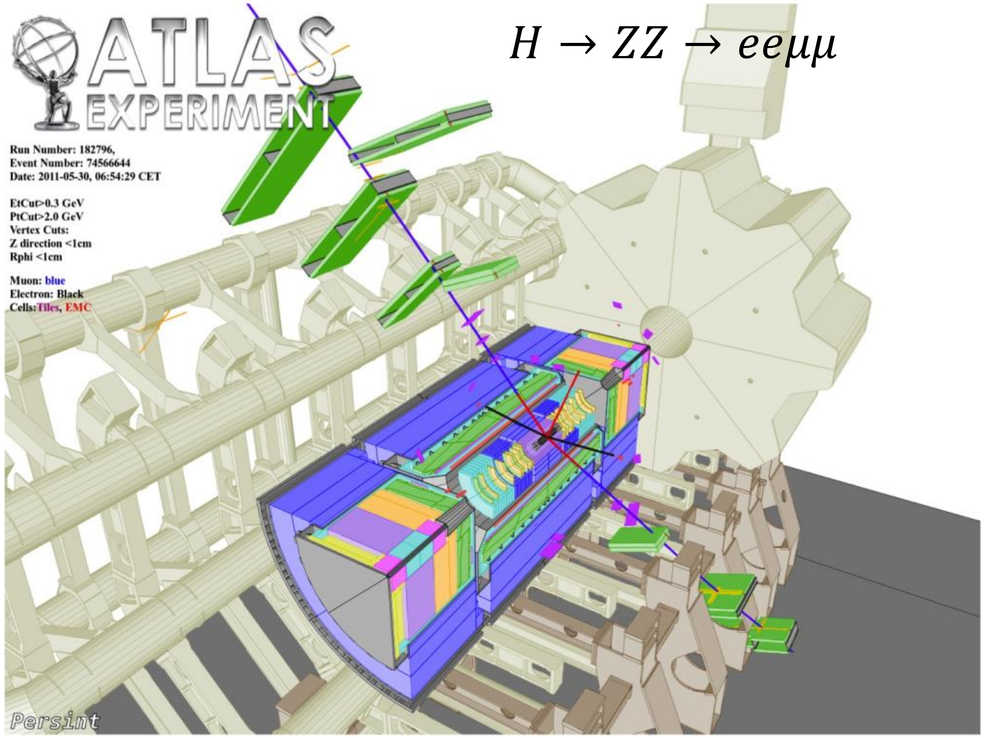
- Many advances in machine learning have been driven by the image recognition problem
- Development of new convolutional neural networks and powerful training methods
  - Convolutional networks scale better than fully connected layers
- Now exceed human accuracy on ImageNet database
- Can image recognition techniques be useful in HEP?

Chihuahua or muffin?

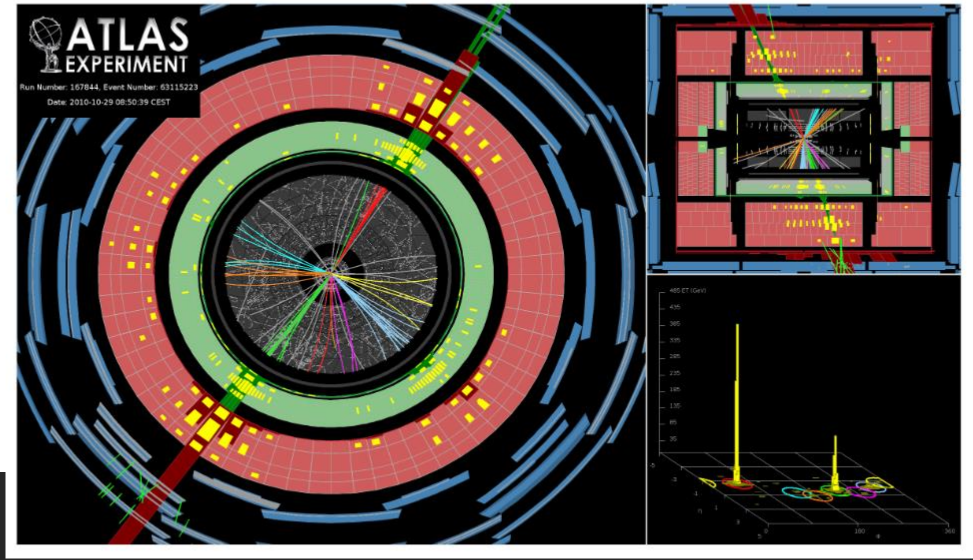
HEP events as Images

- Generally HEP events are not much like images at all
- Very different detector elements
- Complex geometry
- Occupancy low
    - i.e. most detector elements are not activated in each event
- However, unfolded calorimeter deposits do look rather regular and 'image-like'

$H \rightarrow ZZ \rightarrow ee\mu\mu$

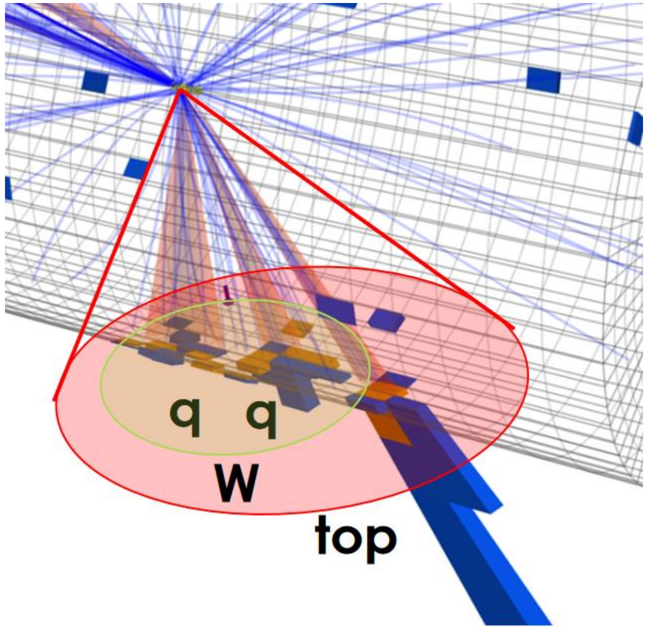


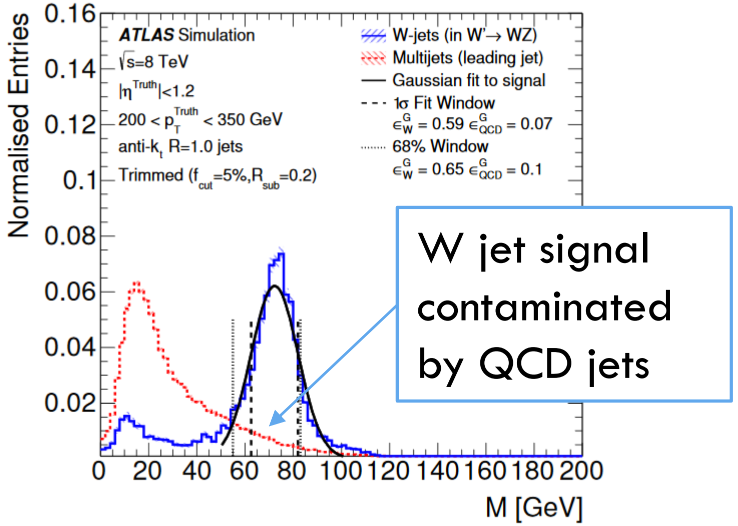
Not very much like a normal 2D image…



Jet Tagging at the LHC

- Important problem at the LHC is to distinguish jets, i.e. showers of multiple particles in the tracker and calorimeter
  - Those produced from the decay of b-quarks
  - Those produced by soft QCD processes
- Jets are a key observable
  - Can identity quarks and gluons in high-energy interactions
  - In combination can identify unstable heavy particles
    - Top quarks, Z, W or Higgs bosons
- At high LHC energies jets become more boosted
  - i.e. significant momentum parallel to the LHC beam
  - This causes jet elements to point close together and being to merge
  - Makes jet identification more challenging

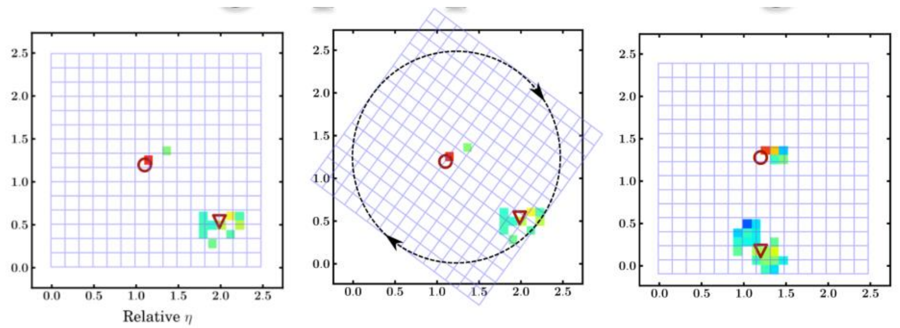


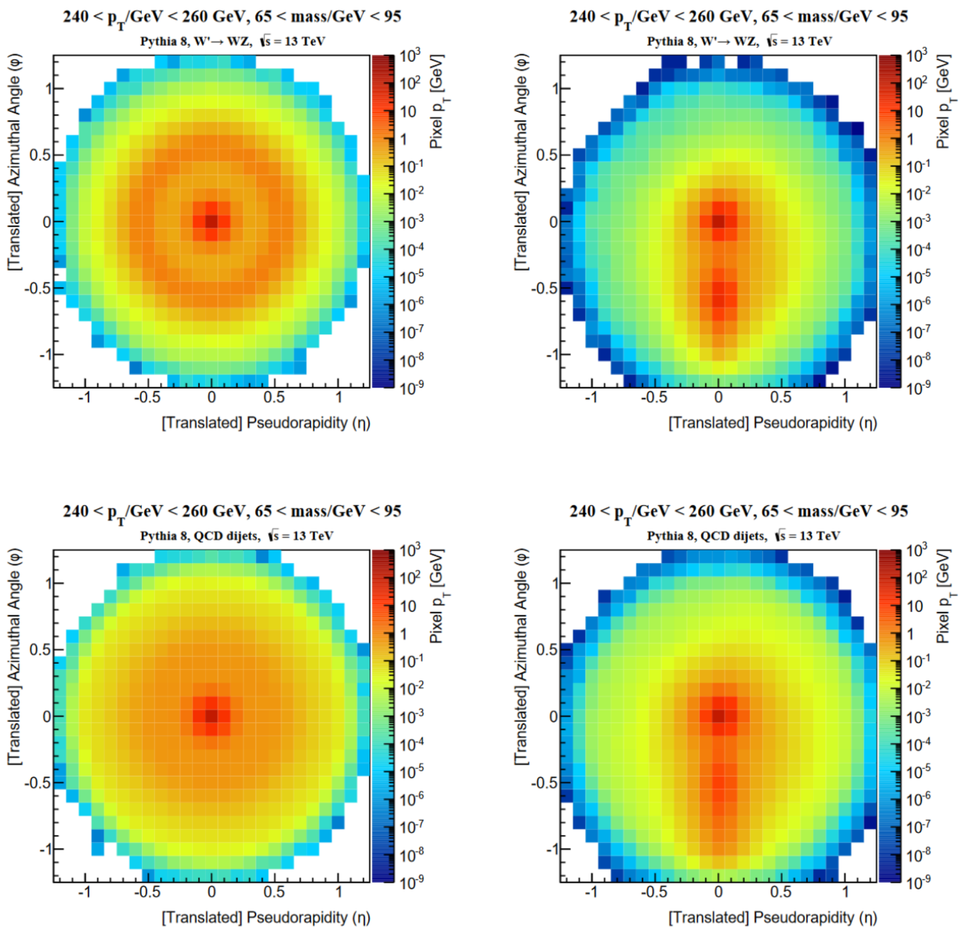


https://arxiv.org/abs/1510.05821

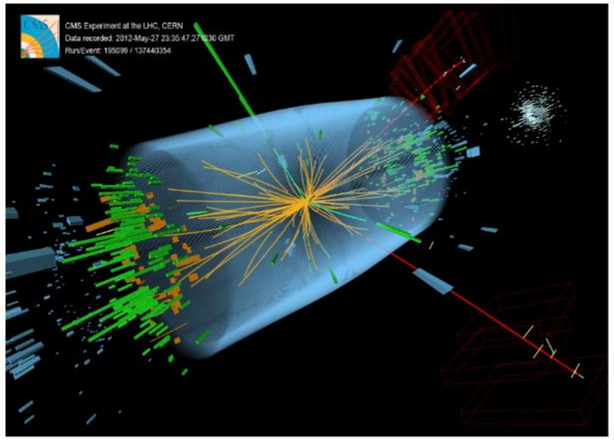
Jets as Images

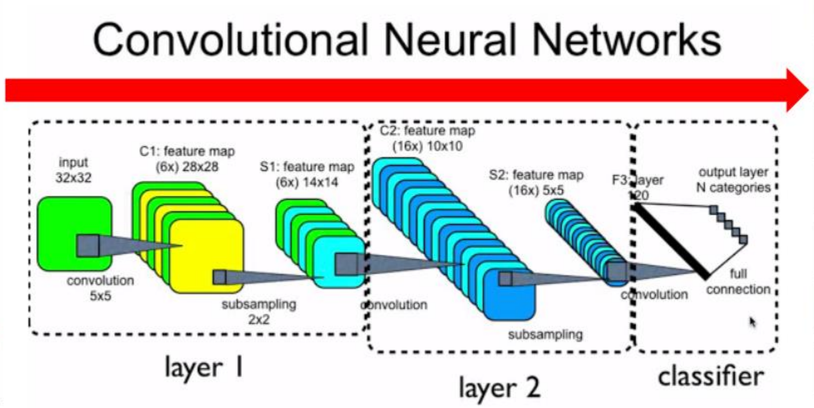
- Jets impact at random places in the calorimeter
- This variation in coordinates is not interesting
- Data preparation consists of
  - Centring
  - Rotating
  - Translate
- Need to be careful to make a physics preserving transformation
  - Calorimeter deposits measure energy, but the invariant quantity is transverse momentum
  - Caveat emptor: Naive application of image recognition techniques may get things wrong

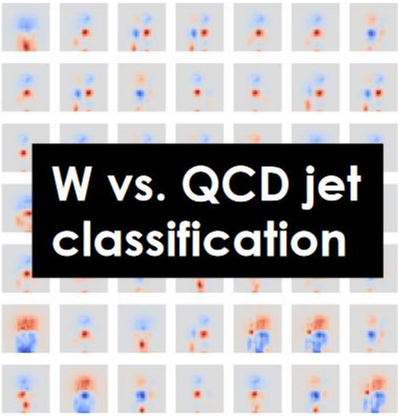




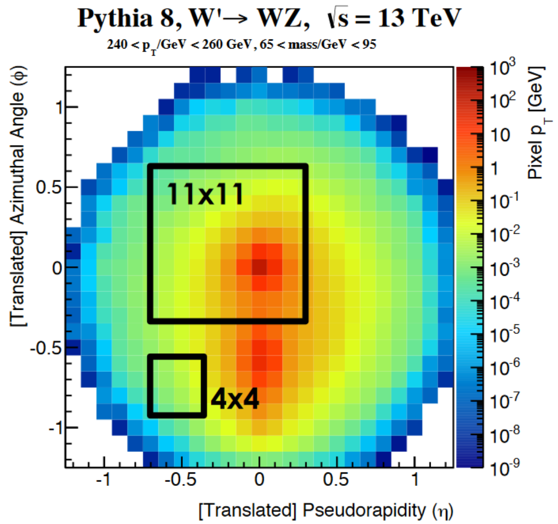
Learning and Tuning







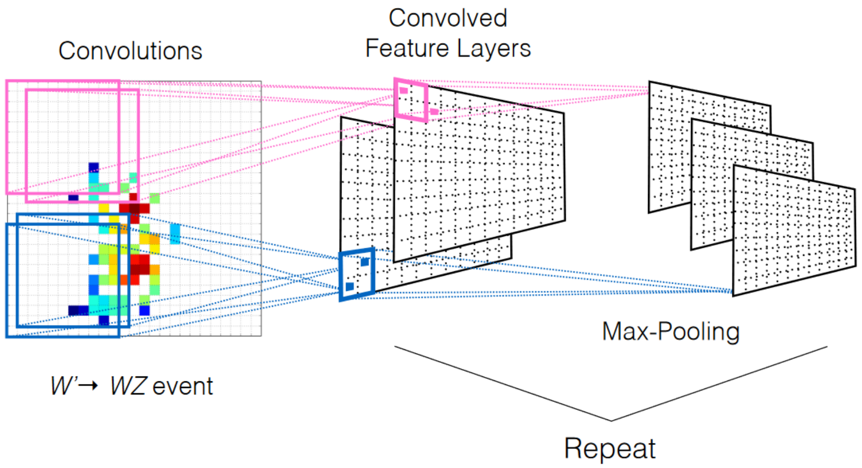
Filter size (1st layer)	(3x3)	(4x4)	(5x5)	(7x7)	(9x9)	(11x11)	(15x15)
AUC	14.8	12.5	11.1	13.3	17.3	**20.3**	18.1

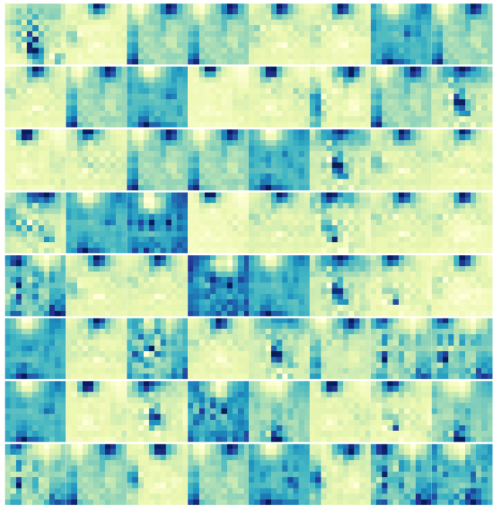


Use of larger filters than for normal image classification gives better results

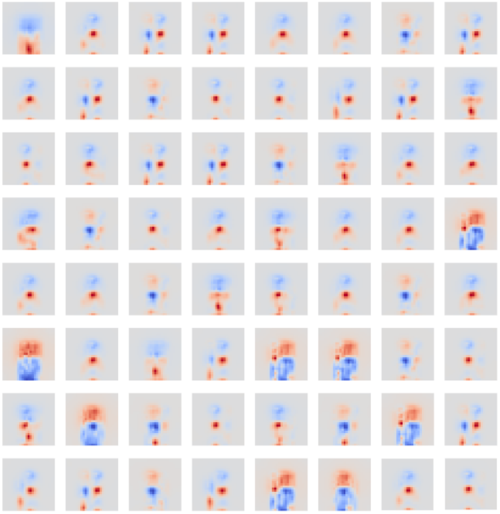
Results

- Results are significantly better than classical physics approaches (n-subjettiness) and classical image recognition ones (Fisher faces)

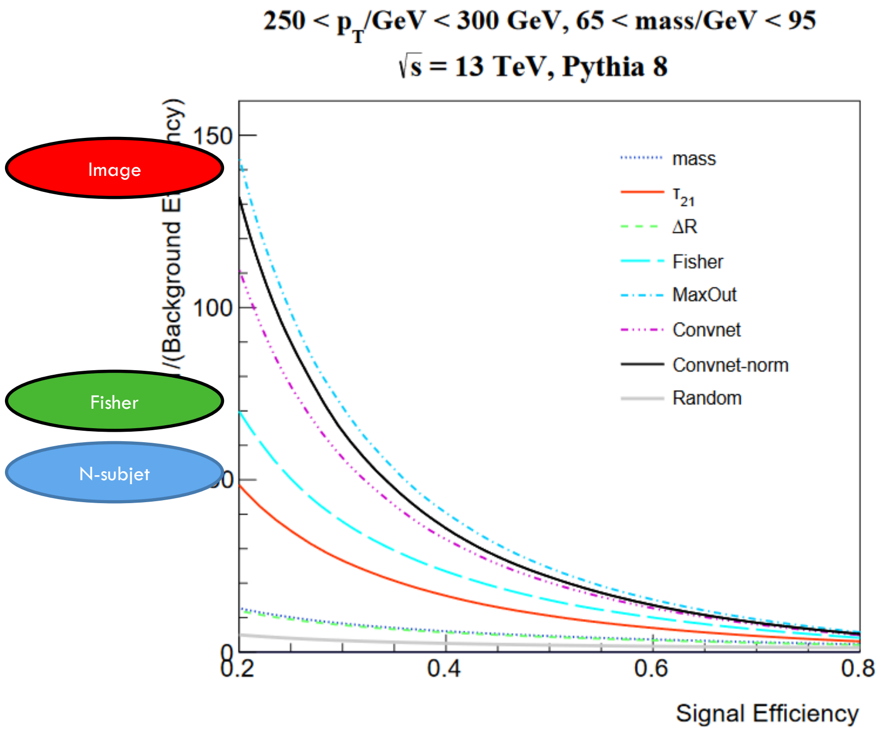




(a) (11 × 11) convolutional kernels from first layer

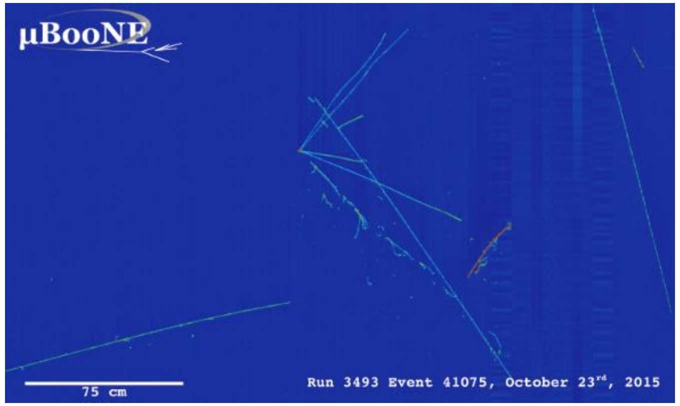


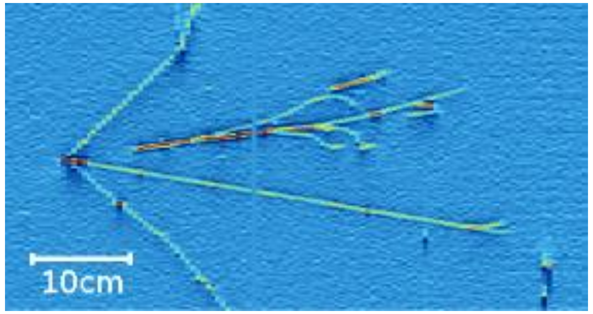
(b) Convolved Jet Image differences

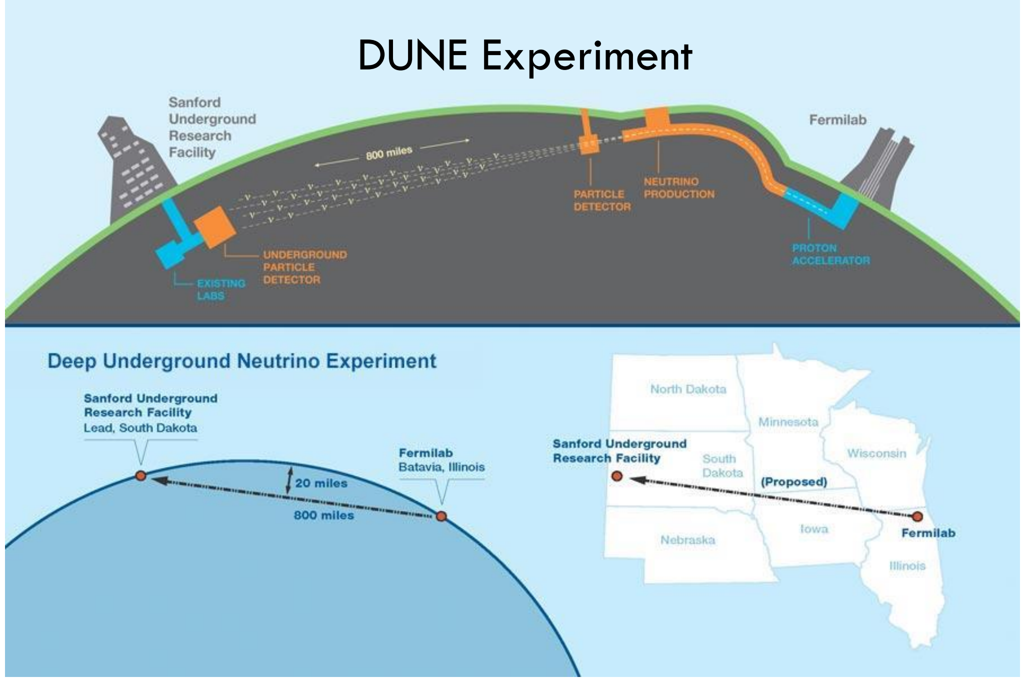


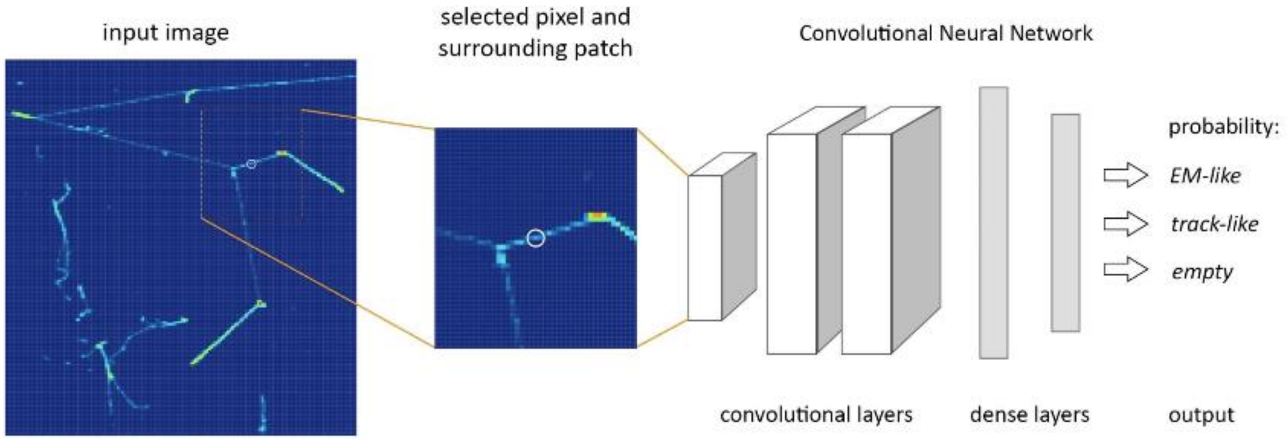
- Kernels indicate that many different features are being identified in the image-jets

CNNs for Neutrino Physics





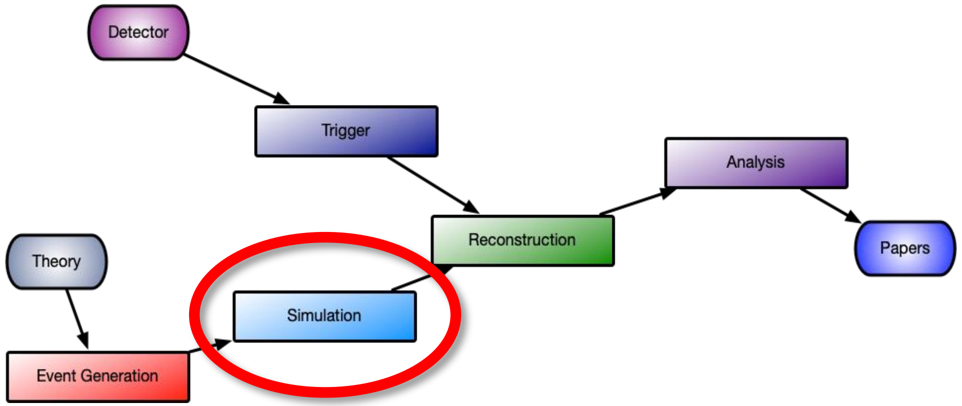




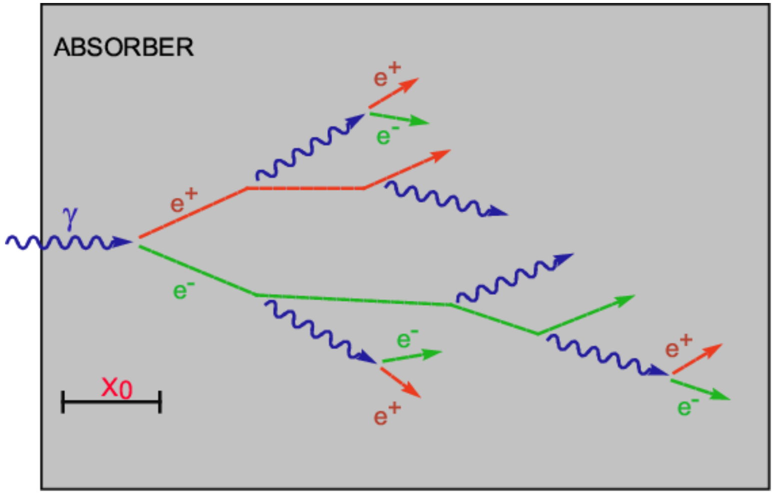
Generative Models

From Classification to Generation

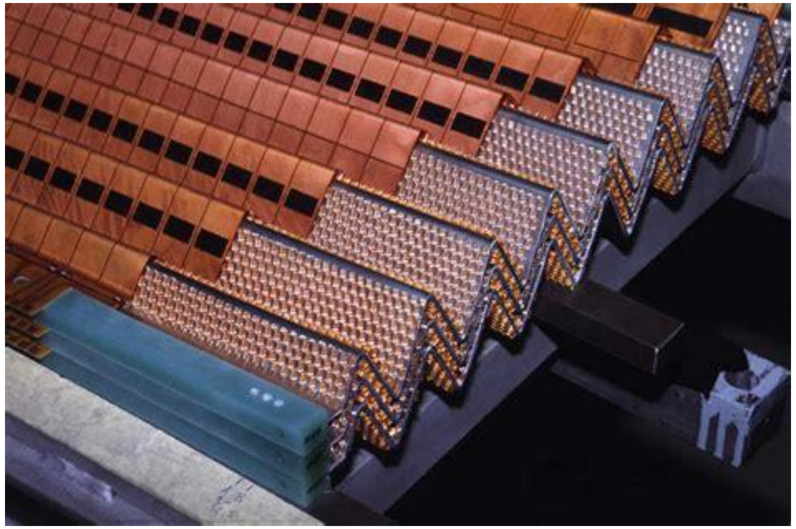
- Looking at LHC computing today the lion's share of resources go into detector simulation
- LHC detectors are highly complex devices and must be simulated in great accuracy for physics
  - From sensitive detectors to dead material for electronics, cables, cooling
  - Many different types of primary and daughter particles
  - Huge range of energies and different physics processes
  - Embedded in a complex magnetic field
- LHC produces a real particle collisions in just 0.000 000 025 seconds
- It may take 100 seconds to simulate this collision

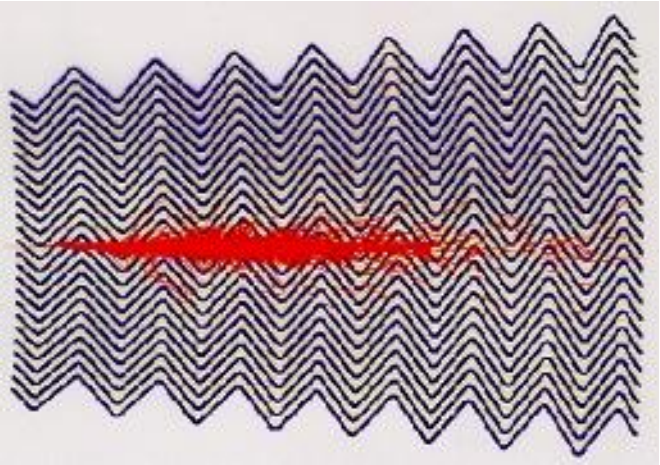


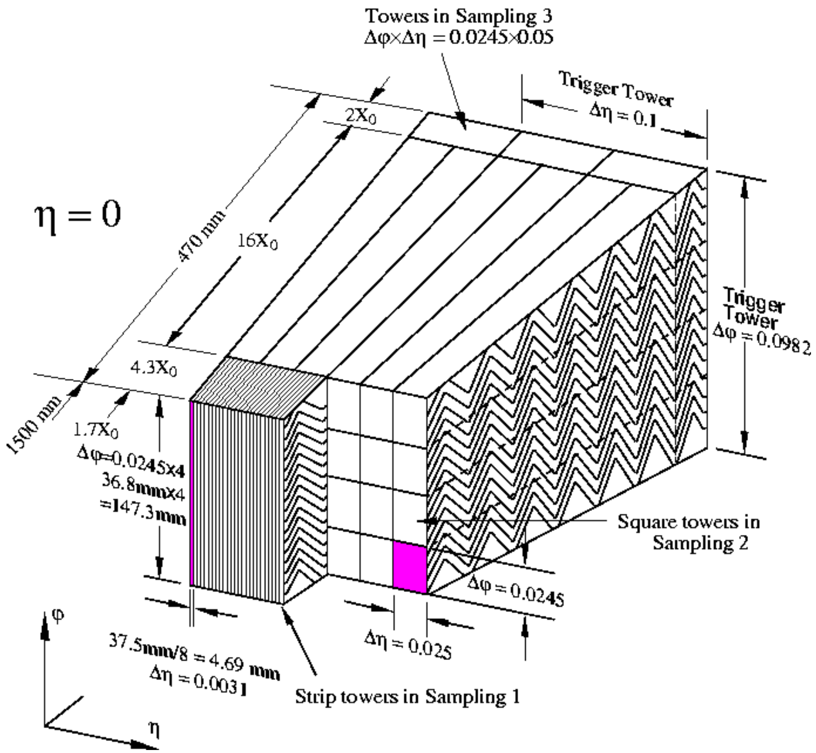
Electromagnetic Showers



- When a high energy photon or electron hits the calorimeter in a HEP experiment it produces an electromagnetic cascade
- The ATLAS calorimeter is a very complicated arrangement of liquid argon and accordion plates
  - Very time consuming to simulate with particle transport, ~75% of Geant4 simulation time for ATLAS

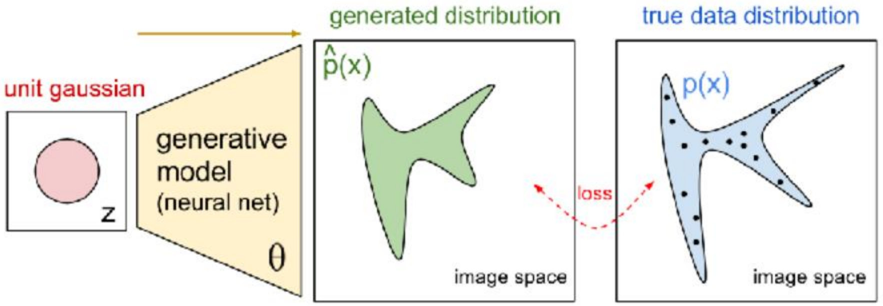


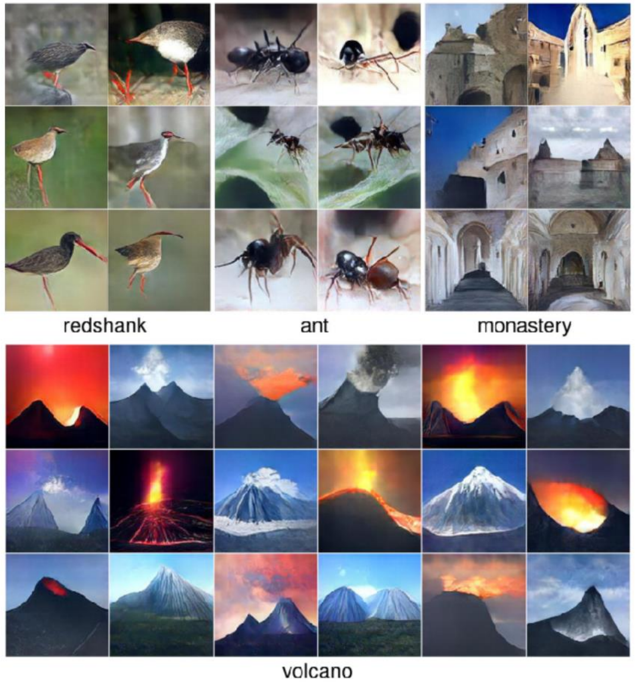




Machine Learning Inspiration: GAN

- Generative Adversarial Networks
  - Two neural networks play a game
    - The generator tries to generate events that look ‘real’
    - The discriminator tries to tell the difference between the real events and the generated ones
    - Can now generate extremely realistic looking faces

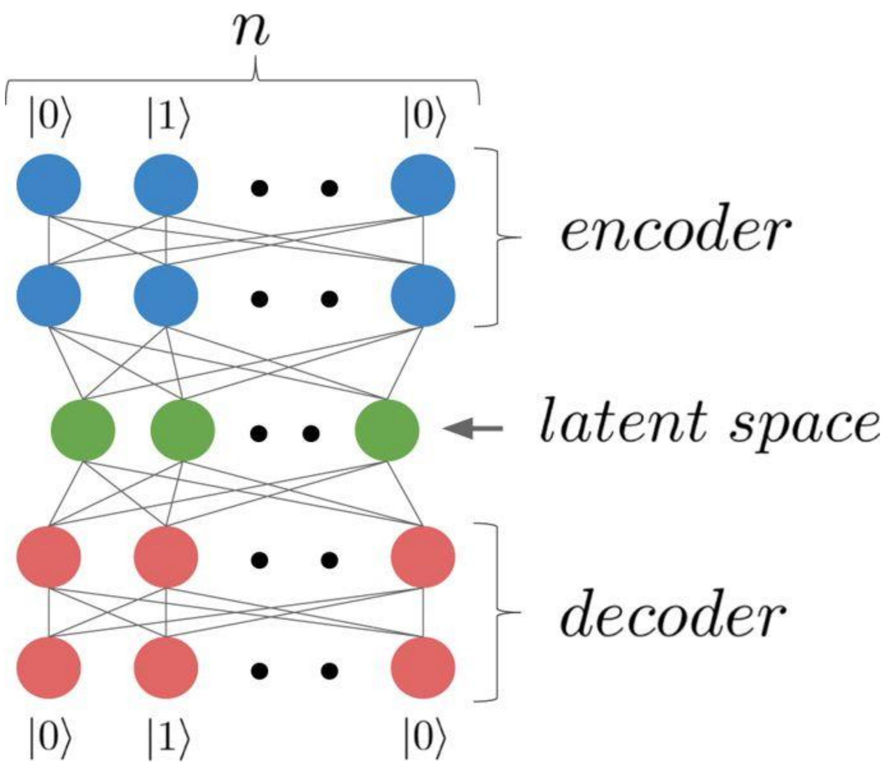




Thanks to Kyle Crammer, NYU

Machine Learning Inspiration: VAE

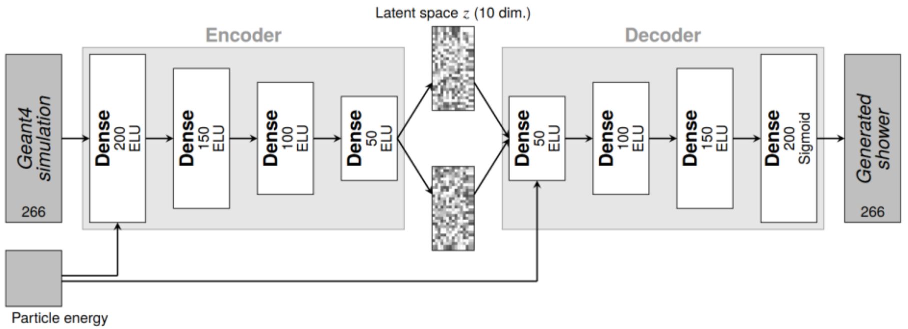
- Variational Auto Encoders are networks that map input data into a lower dimensional latent space
- Then remap that latent space back out to the original event
  - Relatively easy to train
- New events can be generated by sampling from the latent space to get new, unseen events
  - These should model the statistical properties of the original set of events
- VAE applications
  - Dimensionality reduction
  - Anomaly detection
    - This technique has also been used in HEP



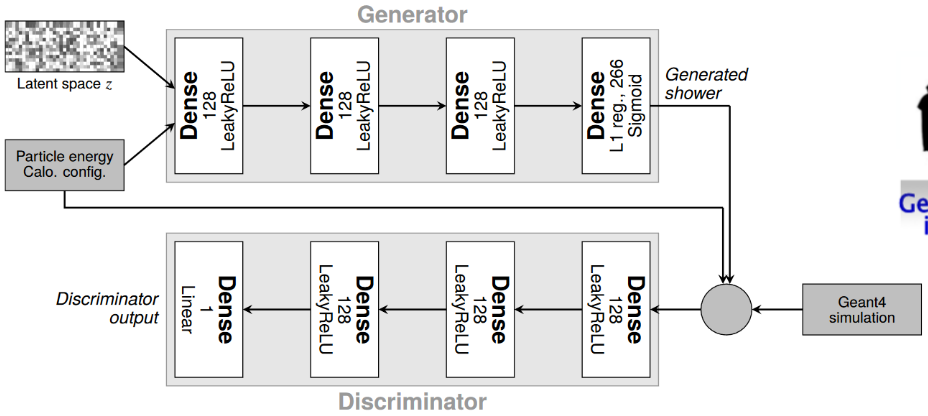
N.B. VAE's 'smoother' than GANs, but may be good for physics?

ATLAS ML Fast Simulation

VAE



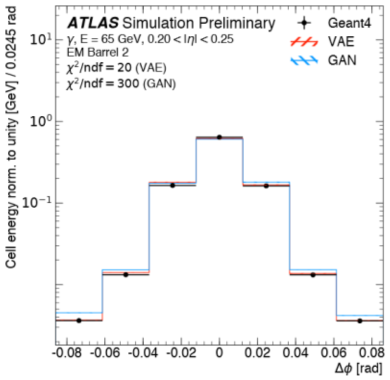
GAN

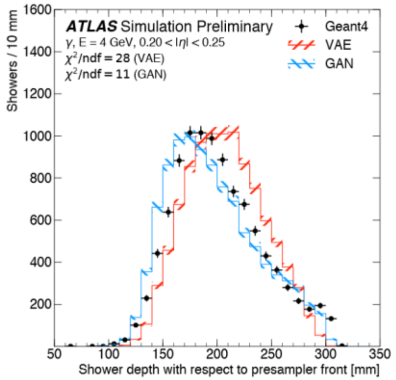


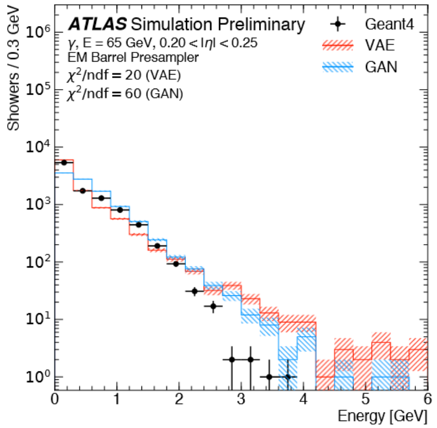
- Training times
  - VAE: 100 epochs, 2 minutes on CPU
  - GAN: 50 epochs, 7 hours on GPU
- VAE is very fast to train
  - Fairly broad hyperparameter scans feasible
- GAN adversarial step is long and increases the training time a lot
  - GAN also suffers from unstable training, may not converge to good results at all
  - Hyperparameter scans very expensive

ATLAS ML CaloSim Results

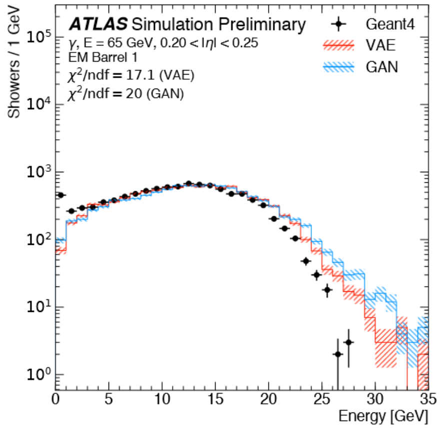
- Modeling of energy response of calorimeter is fairly good
- Note log scale for y-axis
- Difficulty in matching the energy in the tails
  - Energy deposition is small, therefore small in the loss function
- Shower width is good
- Shower depth deviates from expectation



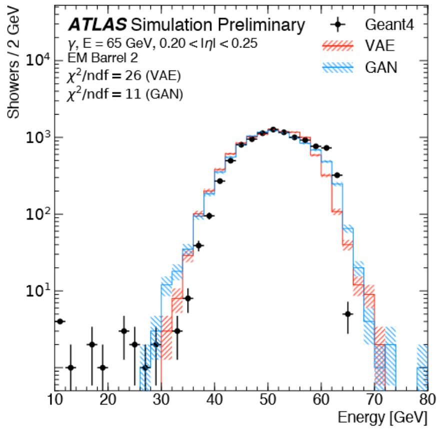




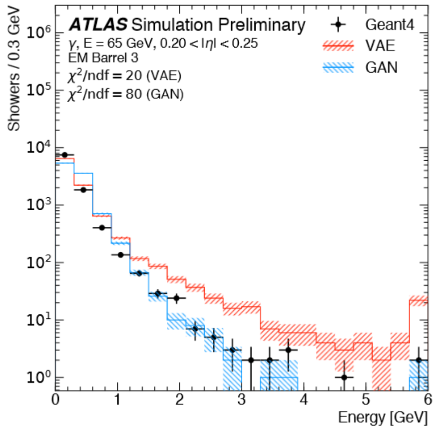
(a) Presampler



(b) Front layer



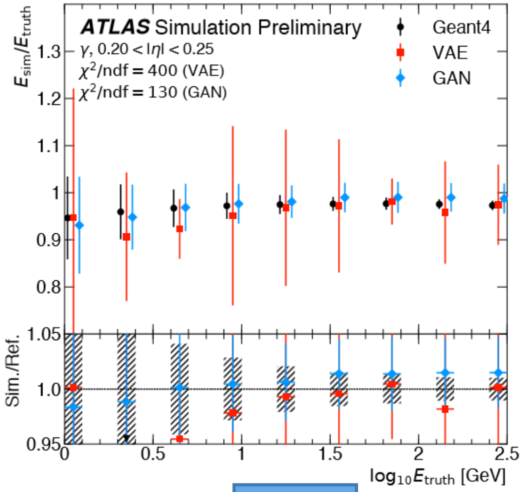
(c) Middle layer

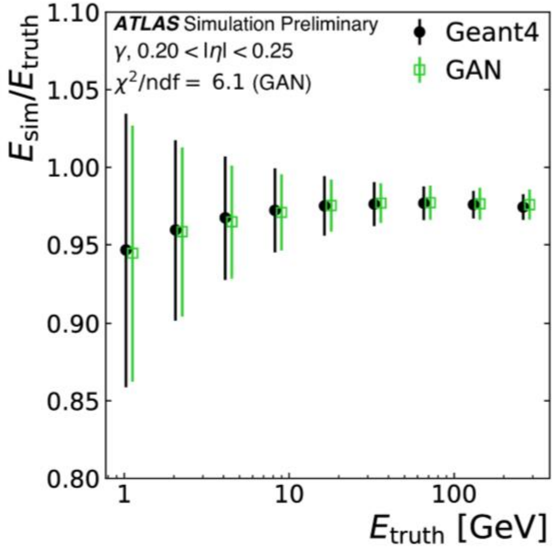


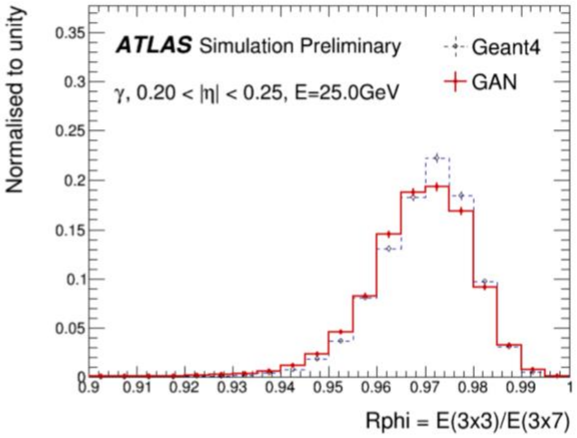
(d) Back layer

Improving the GAN

- Shower energy spread in the GAN did not match that from Geant4
- Add a second critic to the training process
  - Targets total energy of the showers
- Condition on the particle position in the target cell
  - Important to get energy deposition between central cells correct
- Re-optimise generator architecture
- Improved mean and width of energy distribution
- Good performance for interpolated energies

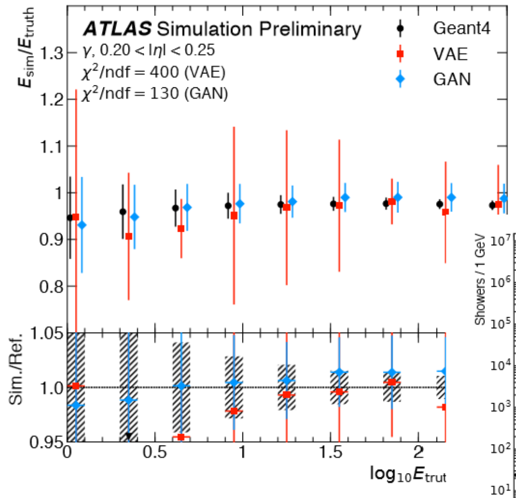


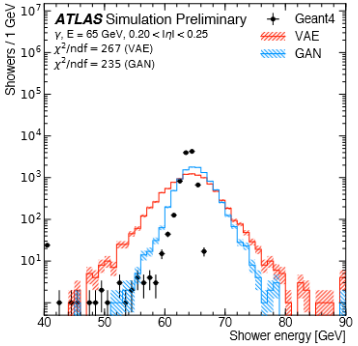


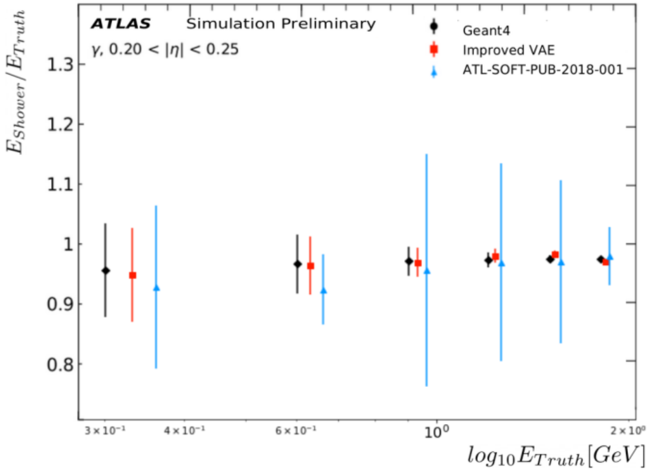


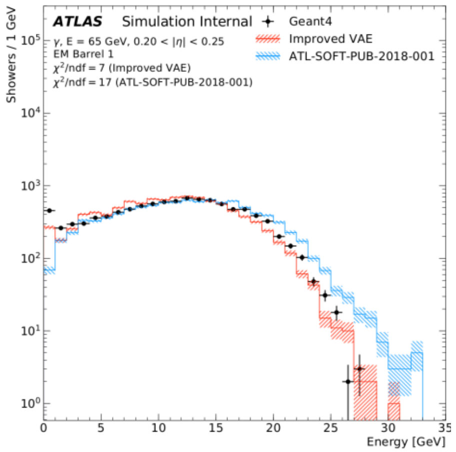
Improving the VAE

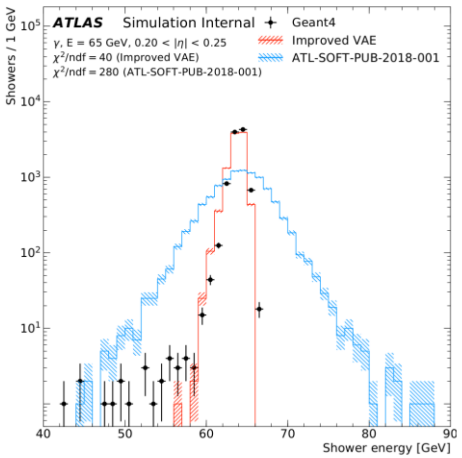
- Train on relative energies
  - E_{cell}/E_{layer}, E_{layer}/E_{total}, E_{total}/E_{truth}
- This re-weighting allows the network to balance energies more easily between layers
- Optimise loss function and architecture
- Much better energy matching
- Next steps
  - Use voxelization for modelling of internal cell structure
  - Different particle types
  - Expand eta range
  - Reapply conditioning to reduce network size







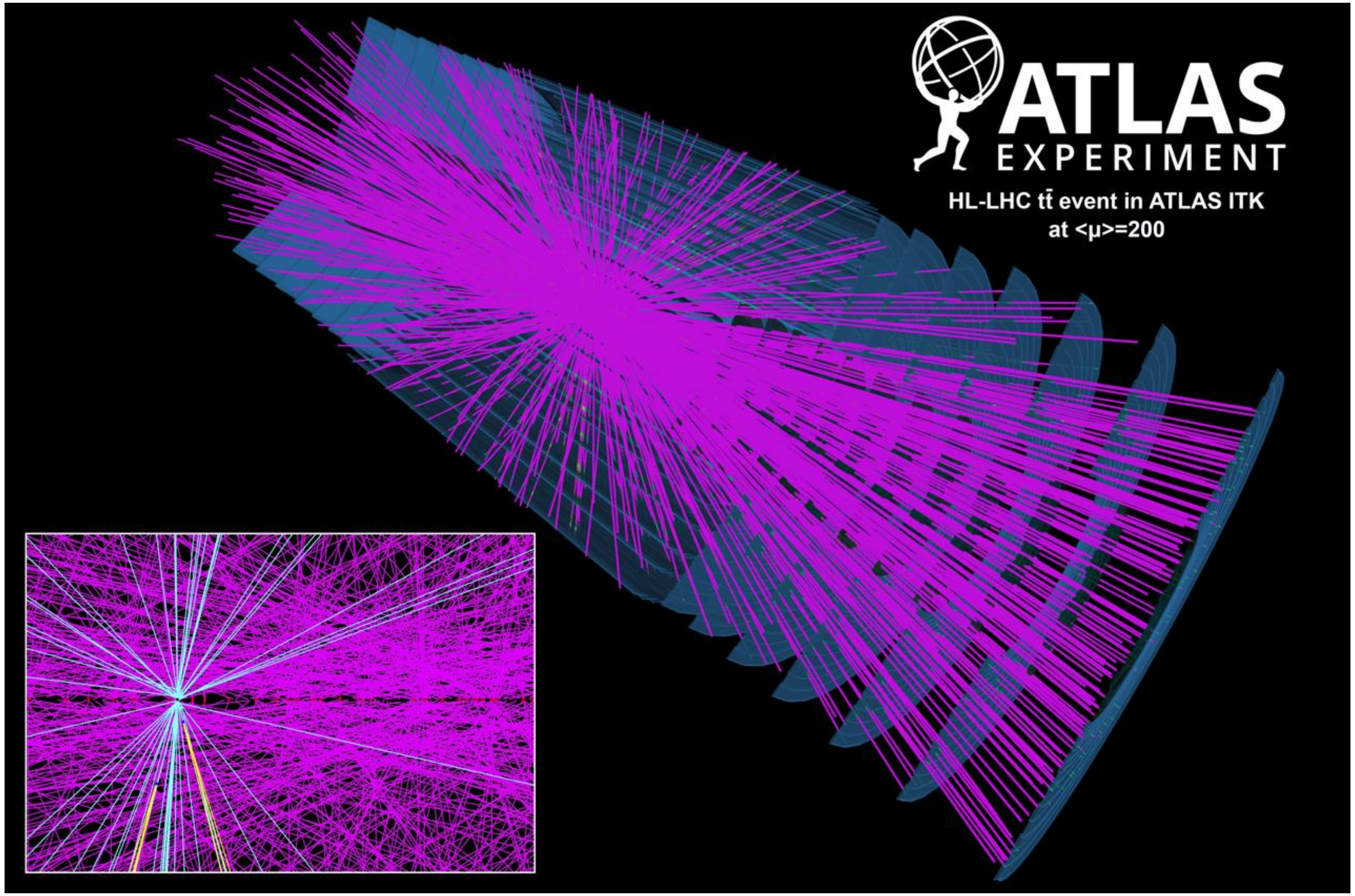




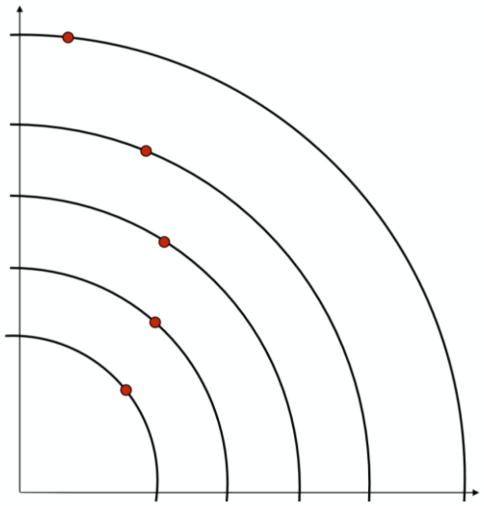
Tracking Challenge

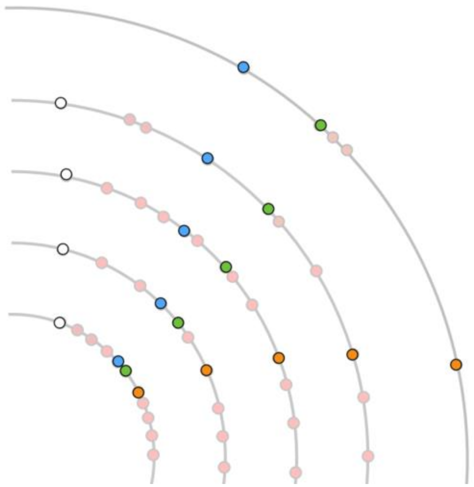
A Future Challenge

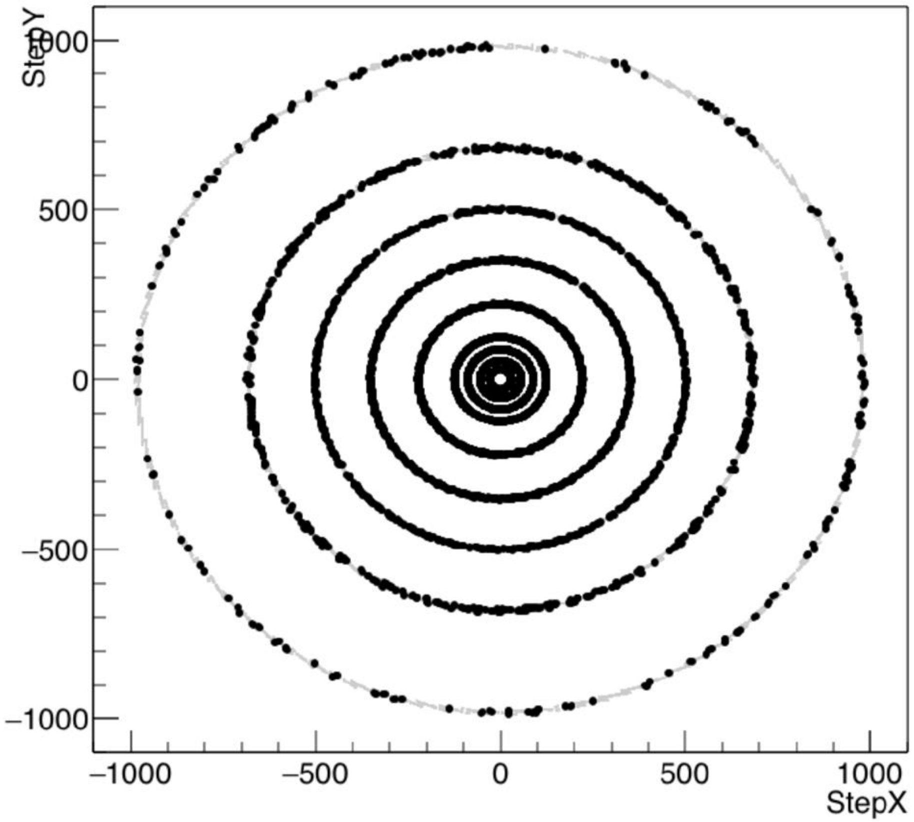
- High-Luminosity LHC will squeeze proton bunches together more tightly
- Instead of 40-60 pp collisions at a time we get 200
  - x4 chance of interesting physics
  - x4 background events, a.k.a pile-up
  - But combinatorics mean much worse than x4 increase in complexity



Tracking In A Few Pictures





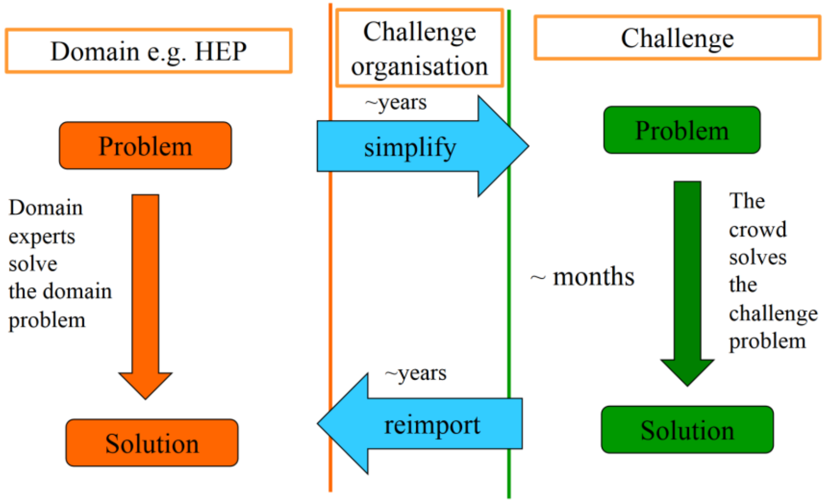


- Naïve tracking can almost be done 'by eye'
- Human visual system quickly overwhelmed in this case
  - Static points, no movement
- Challenge for pile-up 200 is considerable
- Many important physics effects that make this hard
  - Multiple scattering
  - Bremsstrahlung
  - Real detector geometry with inclined sensors, gaps, cables, …

Pictures thanks to Andi Salzburger, CERN

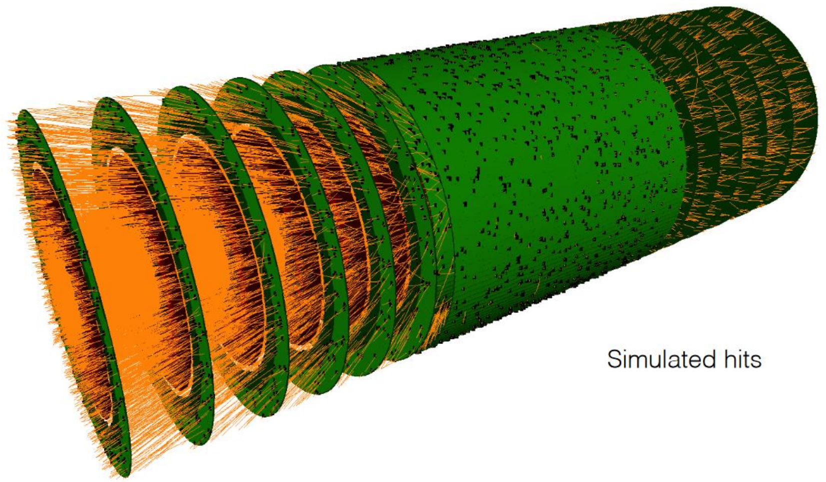
TrackML Challenge

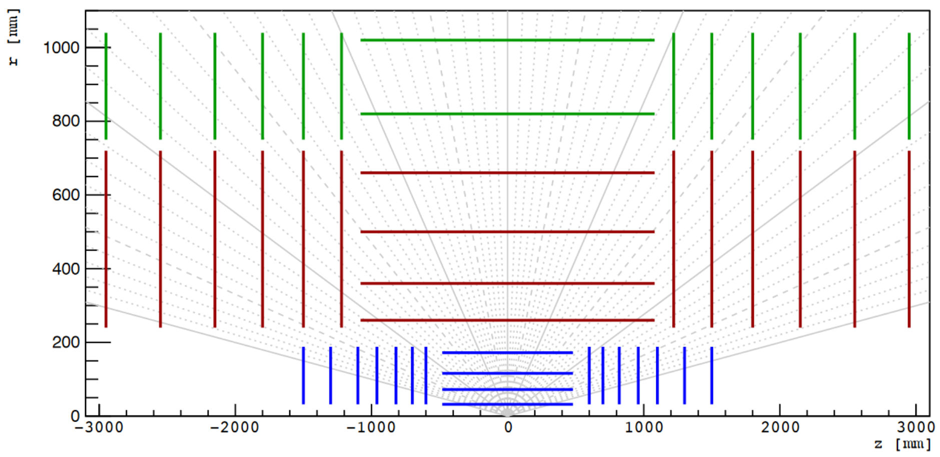
- Launched 30 April 2018
- Attempt to engage the ML community with this HEP problem
- One phase based on accuracy
  - Hosted on Kaggle
- One phase based on performance
  - Accuracy plus time
  - Hosted on Codalab



Thanks to David Rousseau, Andi Salzburger, Jean-Roch Vilmant in particular

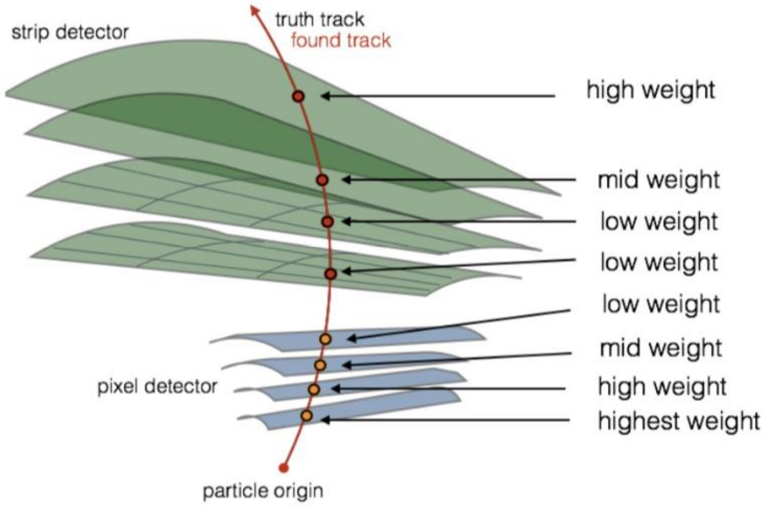
TrackML Dataset

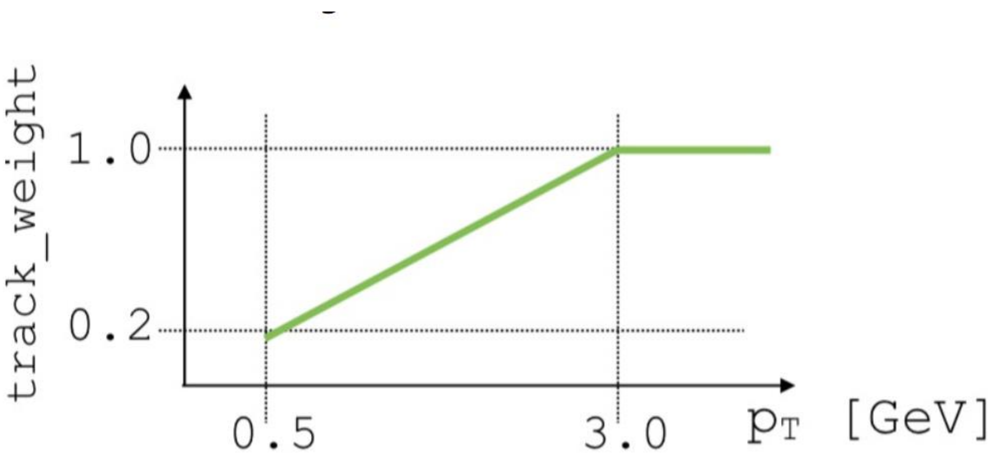




- Realistic model of a HEP tracking detector
  - Pixels, short strips, long strips
- Realistic simulation including multiple scattering, energy loss and hadronic interactions
- Detector and track data provided as CSV files
  - Actually useful for HEP as this has become a reference dataset for testing algorithmic improvements

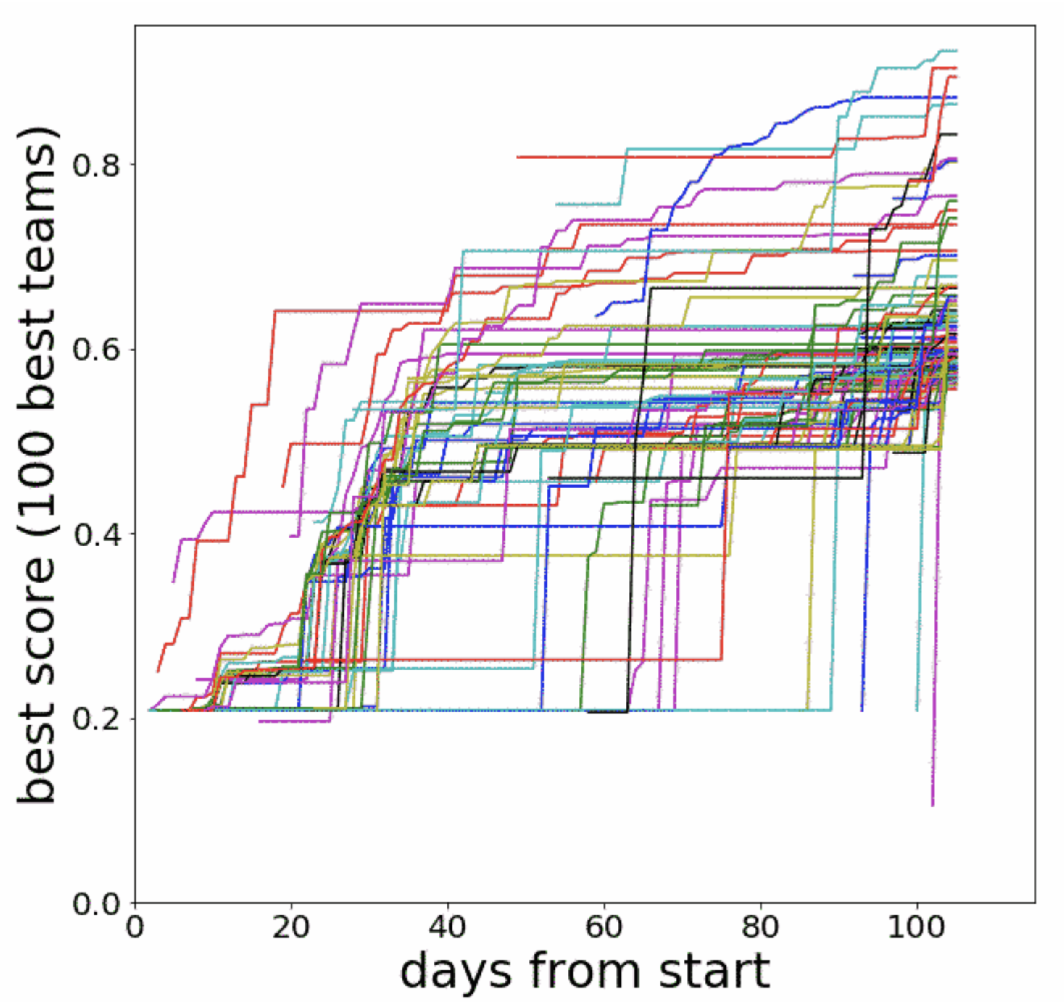
Accuracy





- Accuracy scored in a very physics relevant way
  - Favour tracks which pass right through the detector
  - High-momentum tracks are worth more
- At least 50% hits from the same ground truth particle
- At least 50% hits of the ground truth particle in the track
- Score normalised to sum over all possible tracks
  - Perfect score is 1

Accuracy Winners



#	△pub	Team Name	Notebook	Team Members	Score	Entries	Last
1	—	Top Quarks	Prize		0.92182	10	1y
2	—	outrunner	Prize		0.90302	9	1y
3	—	Sergey Gorbunov	Prize		0.89353	6	1y
4	—	demelian			0.87079	35	1y
5	—	Edwin Steiner			0.86395	5	1y
6	—	Komaki			0.83127	22	1y
7	—	Yuval & Trian	Jury pick		0.80414	56	1y
8	—	bestfitting			0.80341	6	1y
9	—	DBSCAN forever	Jury pick		0.80114	23	1y
10	—	Zidmie & KhaVo			0.76320	26	1y
11	—	Andrea Lonza			0.75845	15	1y
12	—	Finnies	Jury pick		0.74827	56	1y
13	—	Rei Matsuzaki			0.74035	12	1y
14	—	Mickey			0.73217	10	1y

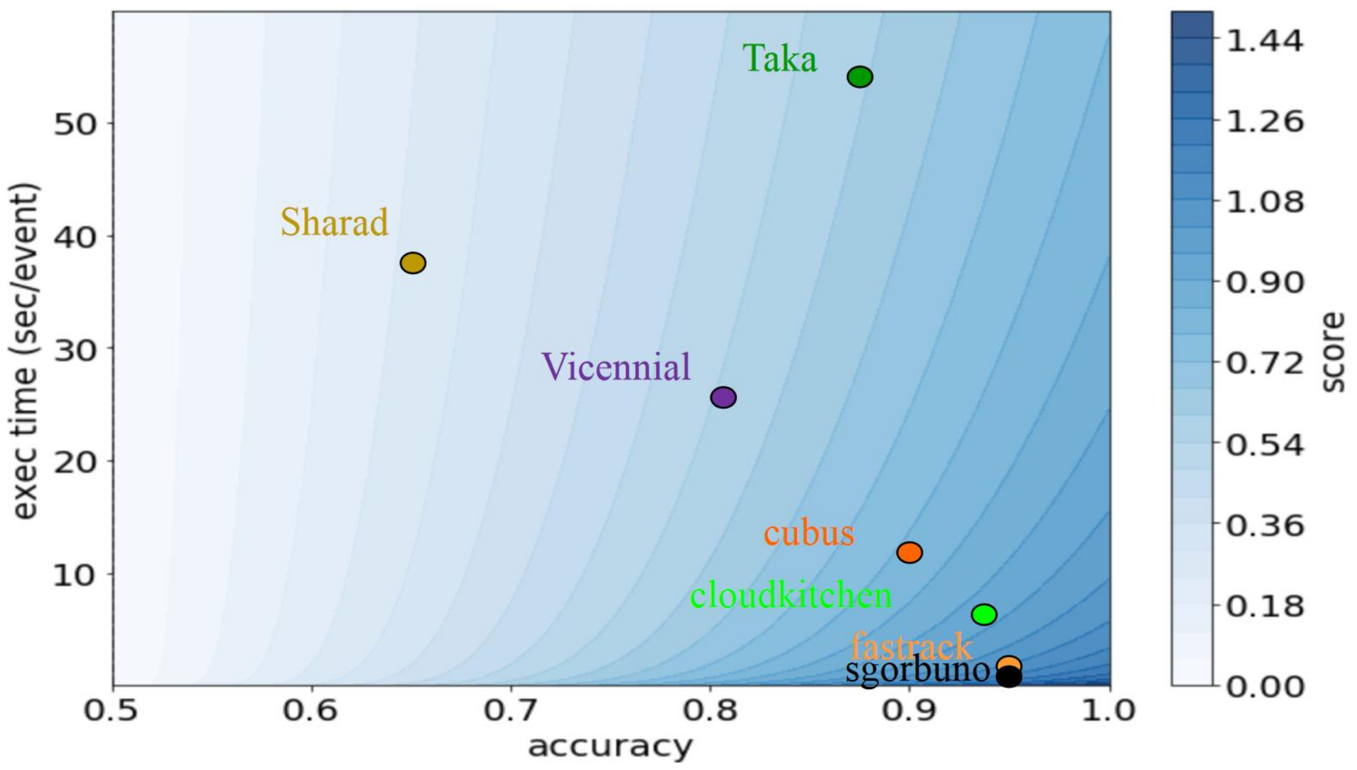
- Steady improvements in the scores over time
- A number of very innovative approaches

Accuracy Insights

- First: Top Quarks
  - Johan Sokrates is an industrial Mathematics master student
  - Pair seeding, triplet extension, trajectory following, track cleaning, all with machine learning for quality selection
- Second: Outrunner
  - Pei-Lien Chou is a software engineer in image-based deep learning in Taïwan
  - Machine learning to predict the adjacency matrix
- Third: Sergey Gorbunov
  - Sergey Gorbunov is a physicist, expert in tracking
  - Iterative steps, triplet seeding, trajectory following
- Jury Picks
  - Density Based Spatial Clustering (DBSCAN) appeared twice
  - LSTM and ML classifiers used to improve results

Throughput Phase

- Score based on accuracy and time
- $\sqrt{\log(1 + {}^{time}/_{600} \times (accuracy - 0.5)^2}$



Throughput Winners

HEP people

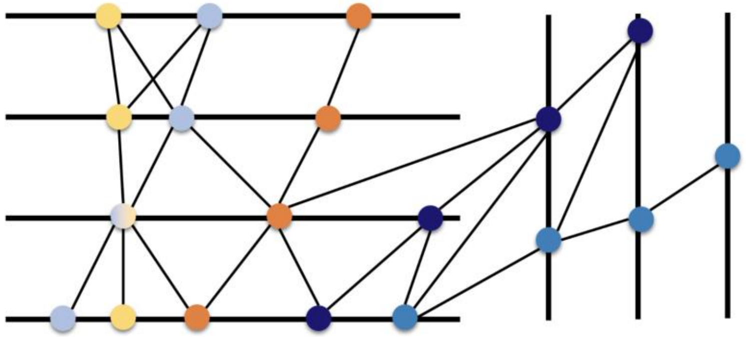
PH+CS

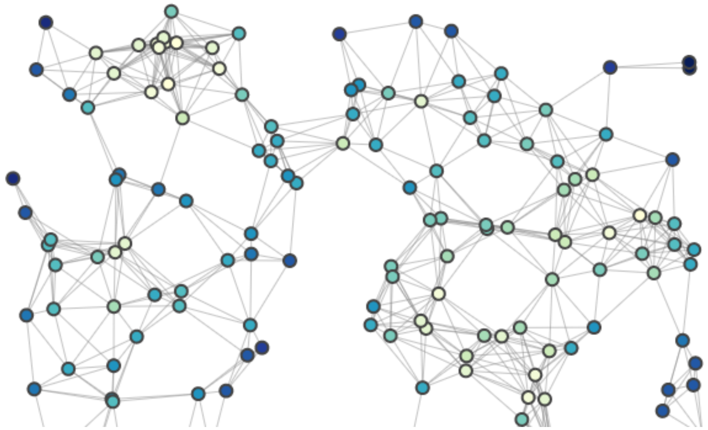
RESULTS									
#	User	Entries	Date of Last Entry	score ▲	accuracy_mean ▲	accuracy_std ▲	computation time (sec) ▲	computation speed (sec/event) ▲	Duration ▲
1	sgorbuno	9	03/12/19	1.1727 (1)	0.944 (2)	0.00 (14)	28.06 (1)	0.56 (1)	64.00 (1)
2	fastrack	53	03/12/19	1.1145 (2)	0.944 (1)	0.00 (15)	55.51 (16)	1.11 (16)	91.00 (6)
3	cloudkitchen	73	03/12/19	0.9007 (3)	0.928 (3)	0.00 (13)	364.00 (18)	7.28 (18)	407.00 (8)
4	cubus	8	09/13/18	0.7719 (4)	0.895 (4)	0.01 (9)	675.35 (19)	13.51 (19)	724.00 (9)
5	Taka	11	01/13/19	0.5930 (5)	0.875 (5)	0.01 (12)	2668.50 (23)	53.37 (23)	2758.00 (13)
6	Vicennial	27	02/24/19	0.5634 (6)	0.815 (6)	0.01 (10)	1270.73 (20)	25.41 (20)	1339.00 (10)
7	Sharad	57	03/10/19	0.2918 (7)	0.674 (7)	0.02 (4)	1902.20 (22)	38.04 (22)	1986.00 (12)

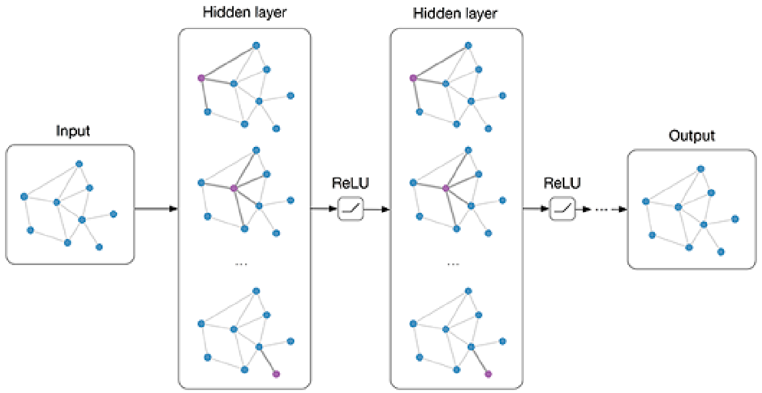
- Best accuracy and best speed went hand in hand in the solution
- It was much harder for non-HEP people to compete on speed
  - Many innovative ideas are interesting, but currently slow
- Classical HEP methods worked well, but ML used to boost accuracy, e.g. track cleaning

Graph Neural Networks

- Approach that was developed during the process of the challenge
  - Albeit independently by the HEP.TrkX project
- Recognise that current HEP approaches have excellent physics performance
  - But scale badly with resources when event complexity rises
- Represent tracker output as a graph
  - One track hit per node
  - Directed edges, inside → outside

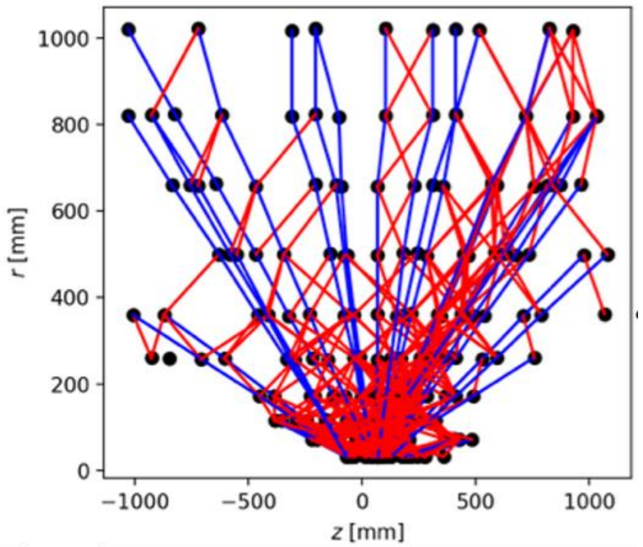






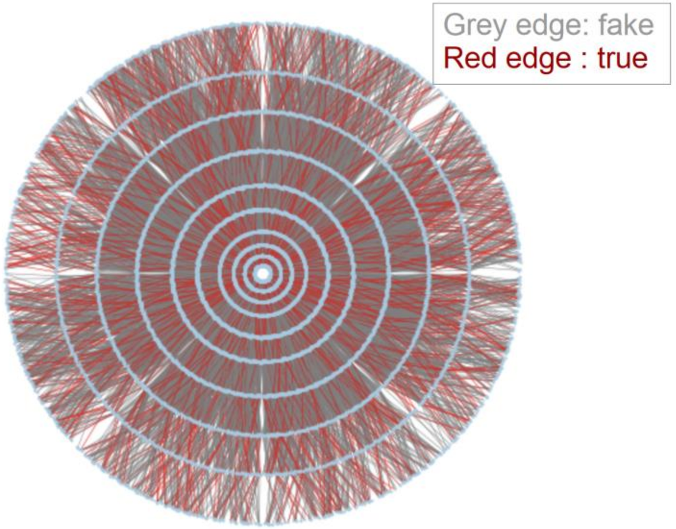
Deep Neural Network that reasons directly on a graph

Tracking and GraphNN



Blue true edge, red false

- Outstanding performance
  - 97% of GNN reconstructible tracks are found
    - N.B. this is a subset of all possible tracks!
  - A very natural representation for HEP tracker data
    - cf. image representation
- Graph is huge
  - 120k nodes
  - 4.4B edges
- Serious investigation underway in a number of areas
  - High-granularity calorimeter reconstruction, where the calorimeter is so fine if becomes like a tracker



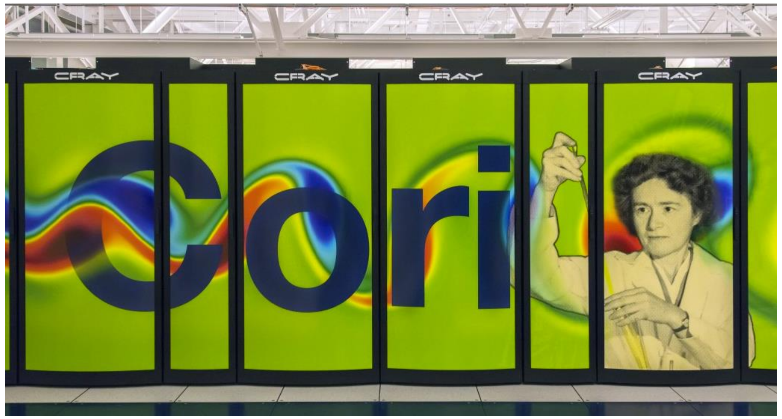
Practical ML for HEP

Practical Machine Learning for HEP

- Today HEP (really LHC) consumes huge computing resources, very costly
  - 1M CPU cores, 24x7x365 usage
  - 1EB of total storage
- These resources are mostly dedicated to our needs, fairly homogeneous and structured around traditional computing solutions
  - CPU cores
  - Mostly 2GB memory per core, few 100GB of local scratch space
  - Evolving from single core slots to multi-core (usually 8)
  - Essentially no inter-slot communication
    - HEP computing is traditionally embarrassingly parallel, because events are independent
- How is Machine Learning done?
  - Some organised production for development of BDTs and NNs in reconstruction
  - A lot of 'private' resources used for development and for training
  - Integration of inference into larger frameworks is still a moving target (lwtnn, TensorFlow, frugally-deep), but ongoing

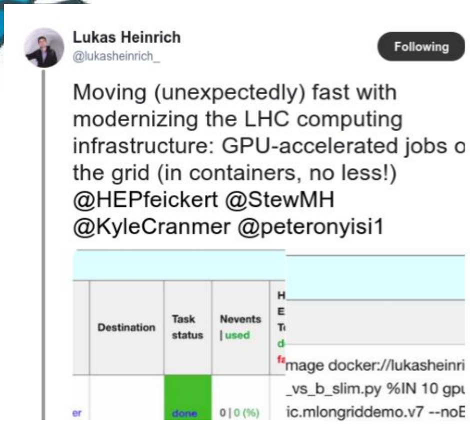
Future HEP Computing

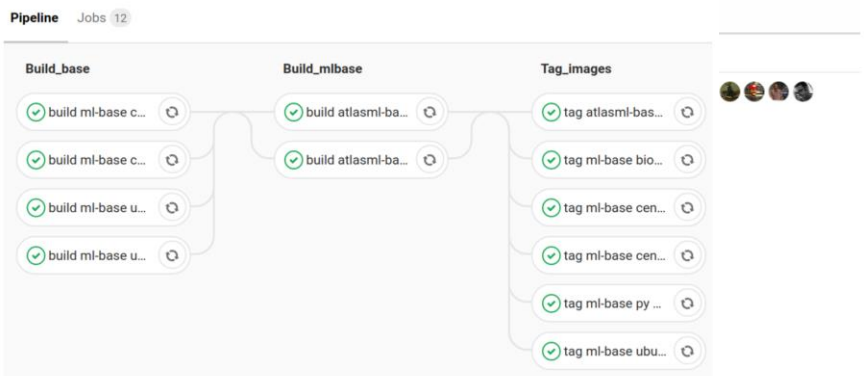
- LHC will not be the only exascale science in the future
  - Other HEP experiments: DUNE, FAIR
    - Can have important differences in their requirements
  - Other big science experiments: SKA, LSST, CTA
    - Different computing approaches, pipeline driven
- Future resources may look very different from those we have today
  - New generations of supercomputers are coming
  - Pressure from funding agencies to use these for HEP computing
  - Anticipated that 90-95% of potential throughput will be in GPUs here
    - Better have some idea what to do with them…
- Traditional computing approaches are not generally easy to adapt to these architectures
  - One of the big attractions of ML from a practical point of view is that it is optimised for these homogeneous high-FLOP architectures
  - And this work is being done by other people!
- Q. Do we have scaleable ML problems for supercomputers?



Portable ML and Data Preservation

- How to combine the latest advances in ML toolkits with running on managed grid resources?
  - Sites will not be installing TensorFlow on demand!
- Containers have started to revolutionise the portability of software stacks
  - Can build central containers for users
  - Who can tweak them if needed
- Now running and available on ATLAS grid resources
- Bonuses for
  - Reproducibility (still a significant issue at the training stage)
  - Continuous integration (testing and versioning)
  - Data preservation
- Still, we do continue to worry about the relatively short lifetime of Machine Learning software cf. a HEP experiment
  - What if software products just get abandoned?





Future Directions

The Future

- Machine learning: better, faster, cheaper… or different?

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

p(x, z | θ, ν)

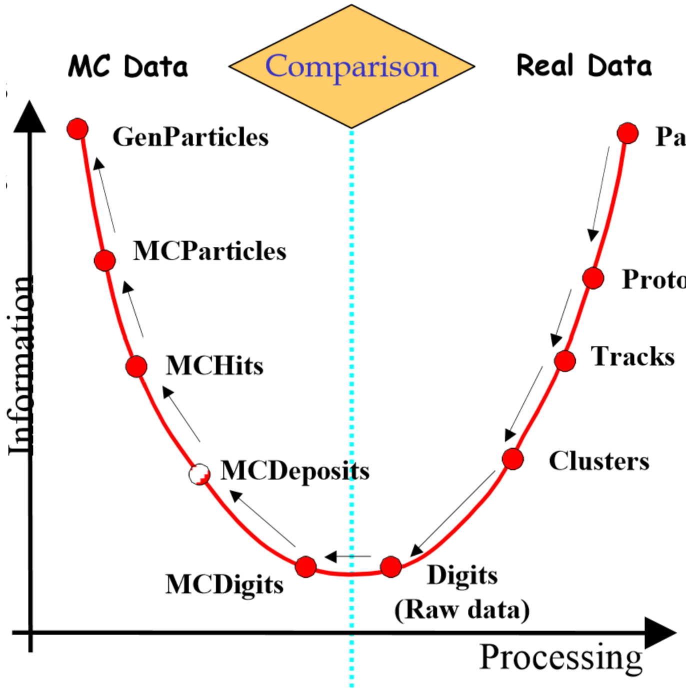
z
latent variables
Monte Carlo truth

x
observed data
simulated data

ν
nuisance parameters

INFERENCE

inverse problem
measurement
parameter estimation



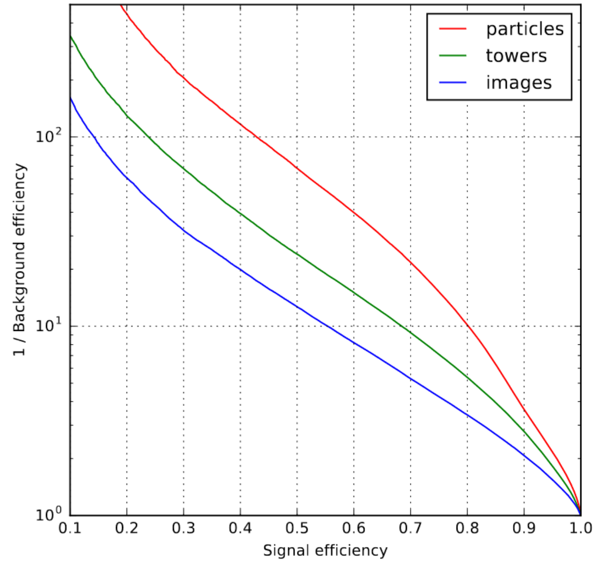
Always looking for shortcuts!

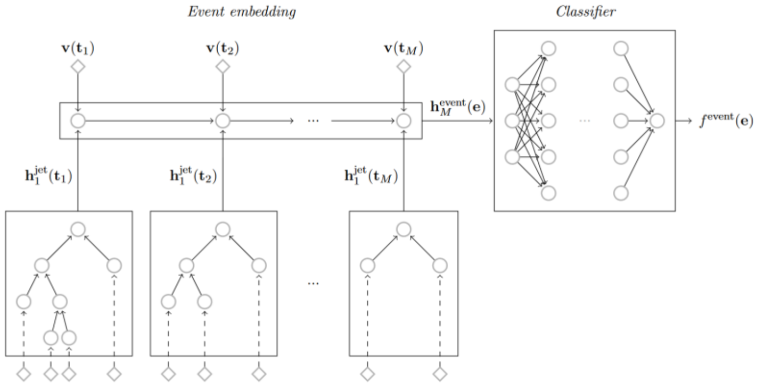
Sketch by Vincenzo Innocente, CERN

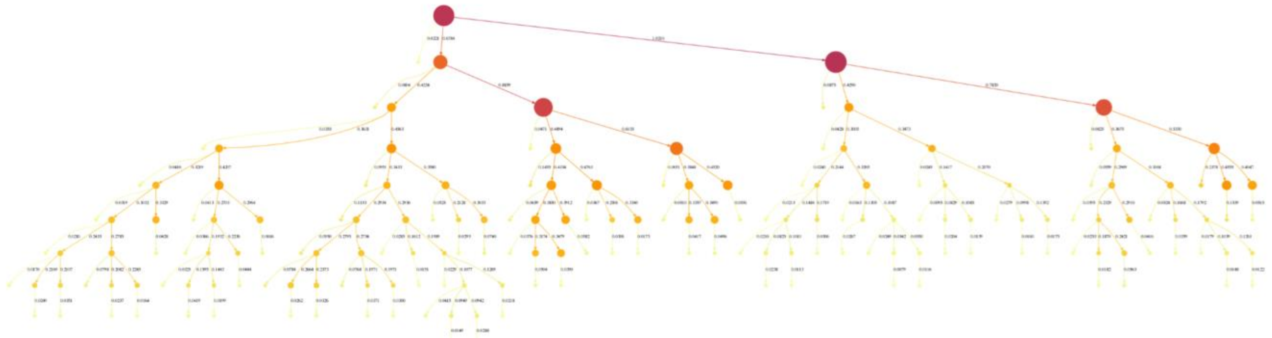
Mostly inspired by Kyle Crnmaer, NYU

Physics Aware Machine Learning

- Physics motivates the kernel used for the ML process
  - Vocabulary of kernels + grammar for composition
- Example is using a recursive neural network (RNN) for QCD jets
  - Encodes the topology of the event in a way that respects physical process
- Given better results than image based approaches



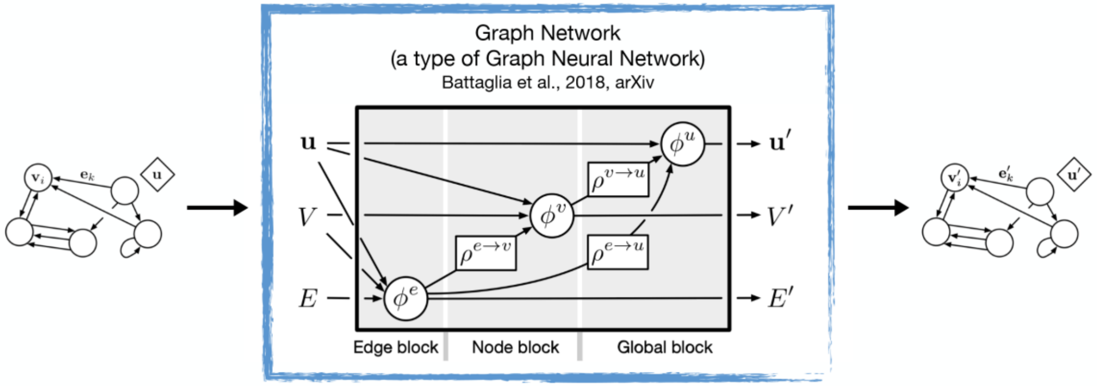




https://arxiv.org/abs/1702.00748

Machine Learning for Structured Data

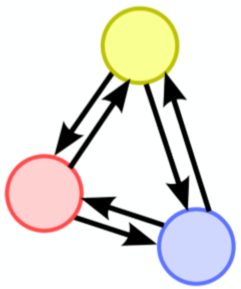
- Standard deep learning toolkits are not well suited to reasoning over structured representations
  - Multi-layer perceptron, convolutional neural network, recurrent neural network – how do I make my data fit the model?
- However, **graph neural networks** are suited to this topic



- Encode fundamental physics (or physical process) in the structure of the graph
- Output of the graph network respects topology, changes weights

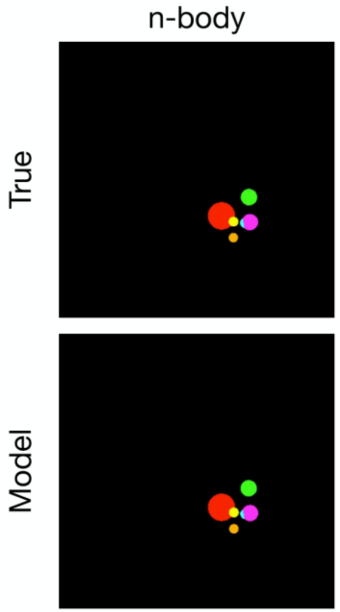
Peter Battaglia, Google Deep Mind

GraphNNs for Physics

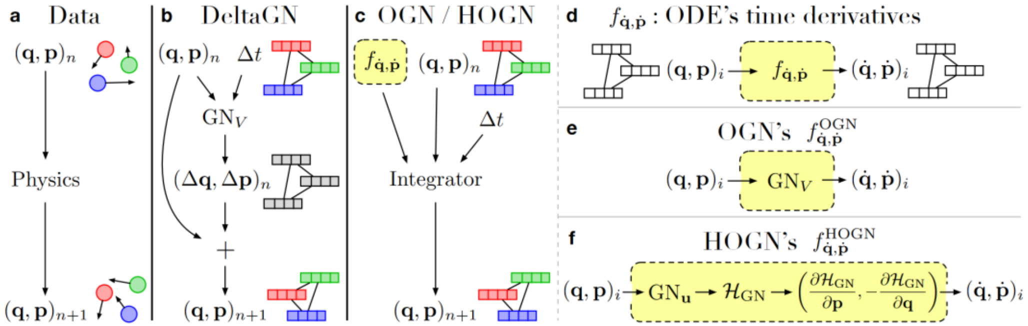


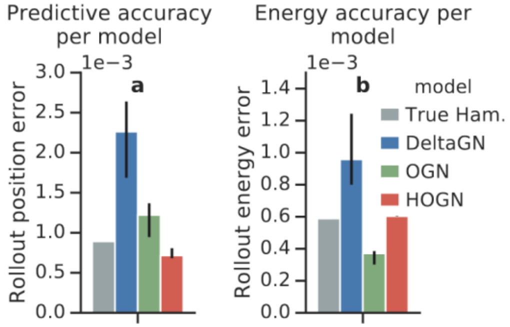
Nodes: bodies

Edges: gravitational forces



Results after 1000 steps





- GraphNN encoding a Hamiltonian
- This is extremely general!

- Modeling that naturally represents structured data
  - Available in TensorFlow
- Powerful toolkit for simulation and interpretation of behaviour

Peter Battaglia, Google Deep Mind

Conclusions

- Machine learning has been part of the toolkit for High-Energy Physics for a long time
- Advances in the field, driven by industry and academic research are opening up new opportunities for applications
- Improved discriminators are a natural application for HEP
  - Already we got important boosts to the statistical power of our analyses
- Applications are moving out to other parts of the HEP computing domain
  - Reconstruction applications, such as low level tracking – emphasis on speed of inference (even using FPGAs)
  - Generative networks for speeding up simulation
- Applying the techniques developed for other fields takes quite some care
  - Fruitful engagement with data science community through Physics ML challenges
- Machine Learning fits well into the evolution of computing hardware
  - That alone would be an important driver and brings practical **better, faster** and **cheaper** improvements
- Encoding domain knowledge of physics naturally into machine learning takes time and effort
  - This is the most radical option to really revolutionise HEP computing – will take years to be fully proven