

Machine learning and anomaly detection

at the ATLAS experiment

Baptiste Ravina

baptiste.ravina@cern.ch



University
of Glasgow



Overview

A (very brief!) look at ATLAS machine learning activities for non-collider physicists:

- A quick tour of the [experiment](#) (using Higgs-related examples)
- [Machine learning](#) can improve our measurements of the Higgs boson!
- Unsupervised searches and [anomaly detection](#) (aka is there something beyond the Higgs?)

CERN and the LHC

CERN: the European Organisation for Nuclear Research

The acronym doesn't make sense.

It's now also an international organisation beyond Europe.

And we mostly do high energy particle physics, rather than nuclear.

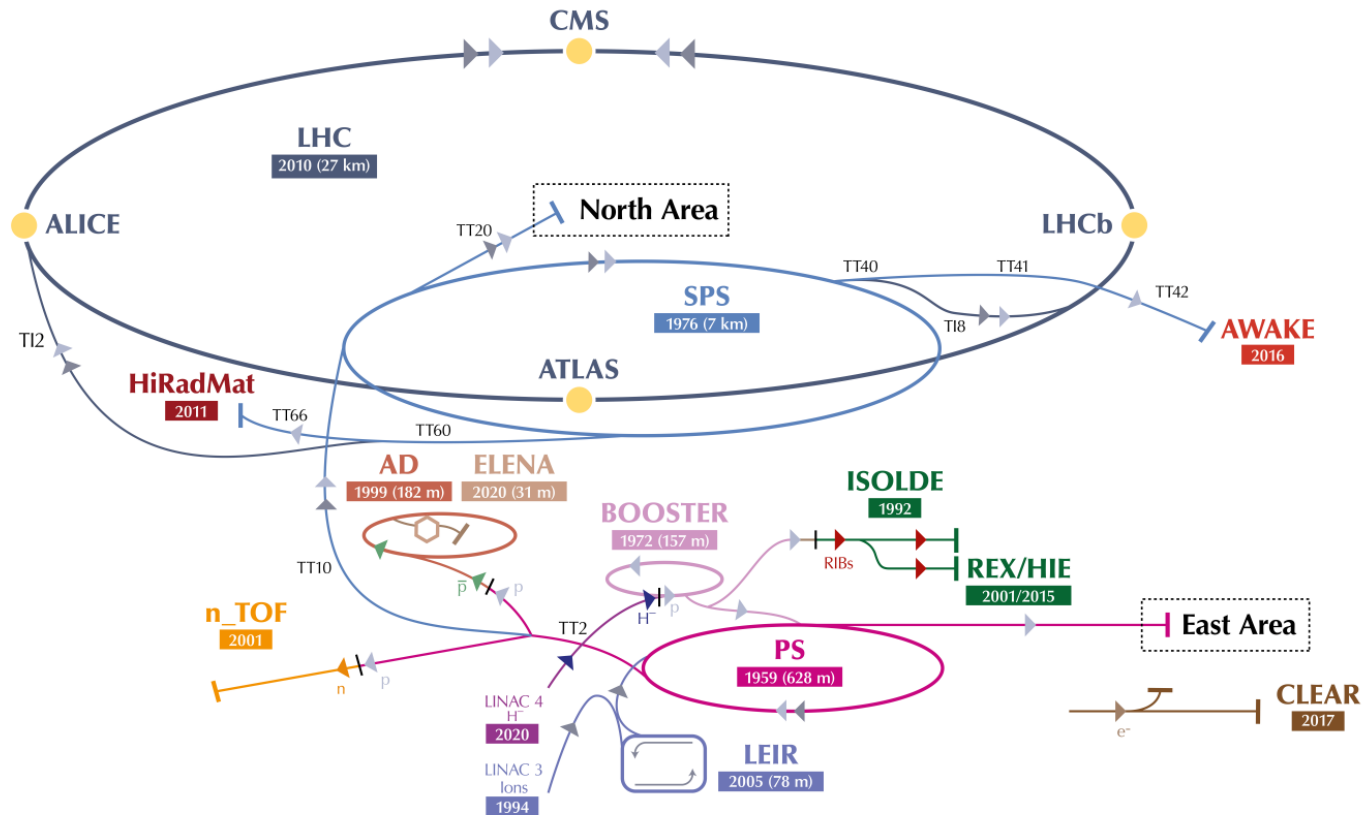


A brief history of CERN

- 1949: a concrete idea for a renewal of nuclear research and scientific excellence in Europe
- 1952: picking the right location (Geneva, Switzerland)
- 1954: start of construction and official birth of CERN
- 1957: the 600 MeV Synchro-Cyclotron starts up
- 1959: the 28 GeV Proton Synchrotron starts up
- 1971: first proton-proton collisions!
- 1976: the 400-450 GeV Super Proton Synchrotron starts up
- 1983: discovery of the W^{\pm} and Z^0 bosons!
- 1989: the 100-200 GeV Large Electron Positron collider starts up
- 1990: the first website is up at CERN
- 2008: the 7-8-13 TeV Large Hadron Collider starts up
- 2012: discovery of the Higgs boson!

The CERN accelerator complex

Complexe des accélérateurs du CERN



▶ H^- (hydrogen anions)
 ▶ p (protons)
 ▶ ions
 ▶ RIBs (Radioactive Ion Beams)
 ▶ n (neutrons)
 ▶ \bar{p} (antiprotons)
 ▶ e^- (electrons)

LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKEfield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE - Radioactive EXperiment/High Intensity and Energy ISOLDE // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n_TOF - Neutrons Time Of Flight //

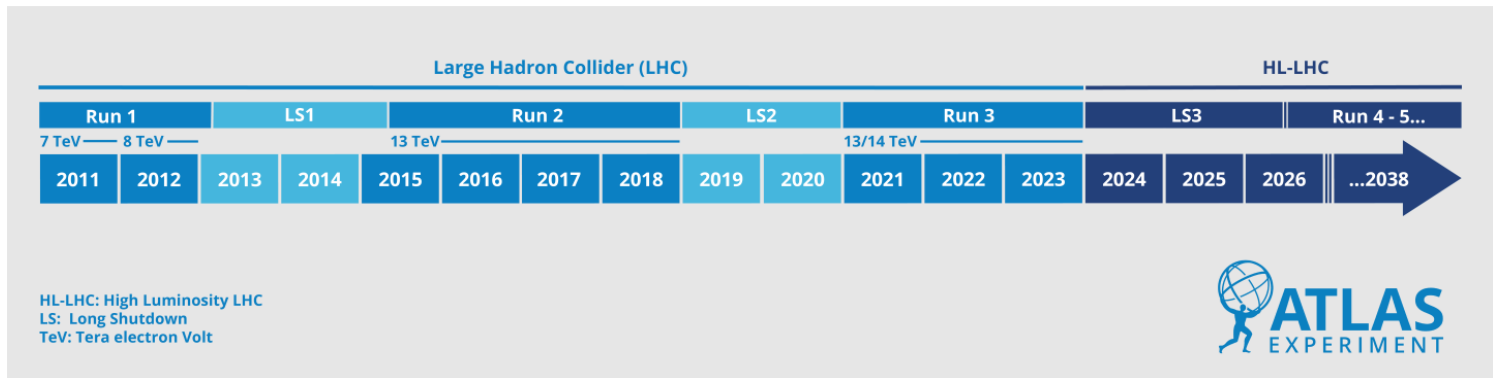
HiRadMat - High-Radiation to Materials

The LHC in numbers

Property	Value
Circumference	27 km
Depth	100 m
Magnet operating temperature	1.9 K (-271.3°C)
Number of magnets	9,593
Beam pressure	1.013×10^{-10} mbar
Nominal energy (protons)	6.5 TeV
Number of bunches per proton beam	2,808
Number of protons per bunch	1.2×10^{11}
Number of turns per second	11,245
Number of collisions per second	1 billion
Cost	CHF 4.3 billions
Energy consumption (CERN)	1.3 TWh/year
Energy production (Geneva)	3 TWh/year

Both colder and emptier than interstellar space!

The timeline



Of course, Covid-19 complicates things... Now expecting Run 3 to start in Q1 of 2022.

But there's still a lot to do with Run 2 data!

The biggest news so far

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC [☆]

ATLAS Collaboration [★]

This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.

ARTICLE INFO

Article history:

Received 31 July 2012

Received in revised form 8 August 2012

Accepted 11 August 2012

Available online 14 August 2012

Editor: W.-D. Schlatter

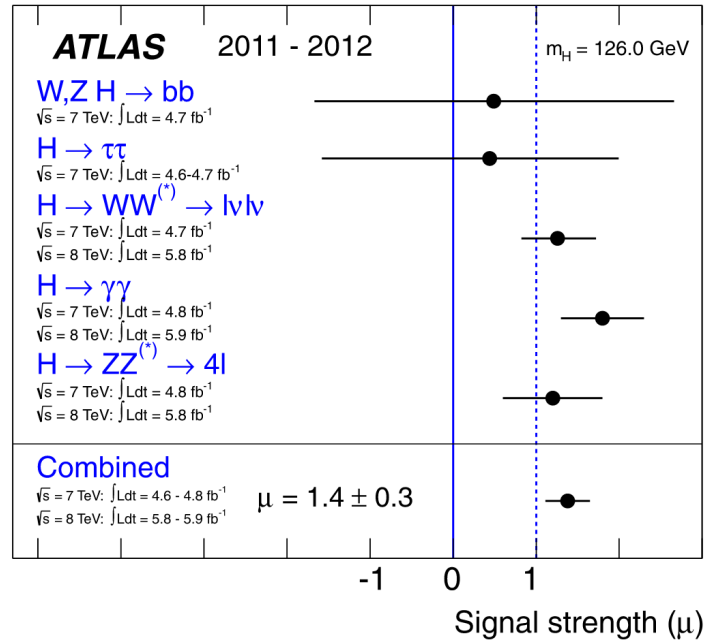
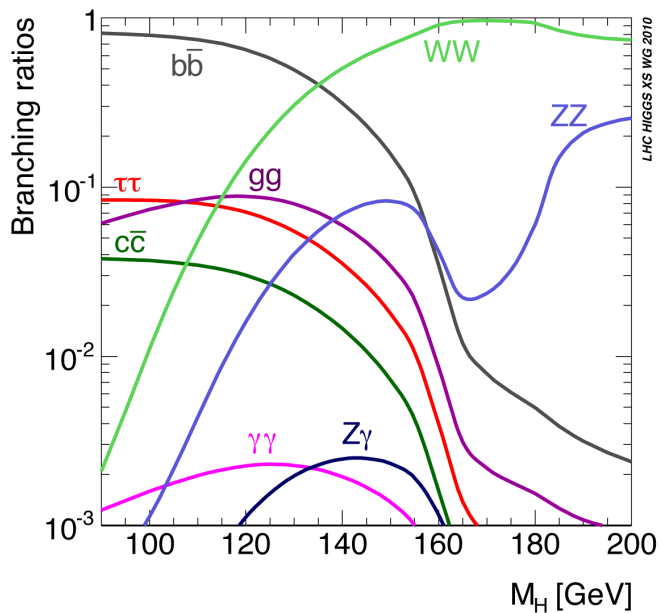
ABSTRACT

A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^{-1} collected at $\sqrt{s} = 7 \text{ TeV}$ in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ in 2012. Individual searches in the channels $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow \gamma\gamma$ and $H \rightarrow WW^{(*)} \rightarrow e\nu\mu\nu$ in the 8 TeV data are combined with previously published results of searches for $H \rightarrow ZZ^{(*)}$, $WW^{(*)}$, $b\bar{b}$ and $\tau^+\tau^-$ in the 7 TeV data and results from improved analyses of the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of $126.0 \pm 0.4 \text{ (stat)} \pm 0.4 \text{ (sys)} \text{ GeV}$ is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7×10^{-9} , is compatible with the production and decay of the Standard Model Higgs boson.

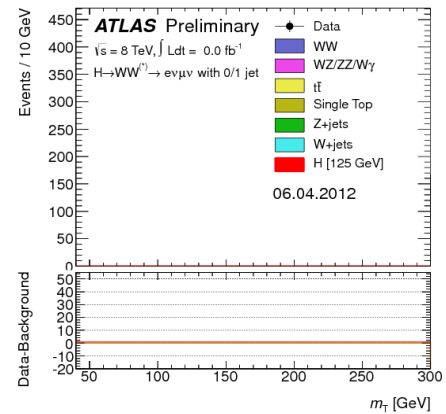
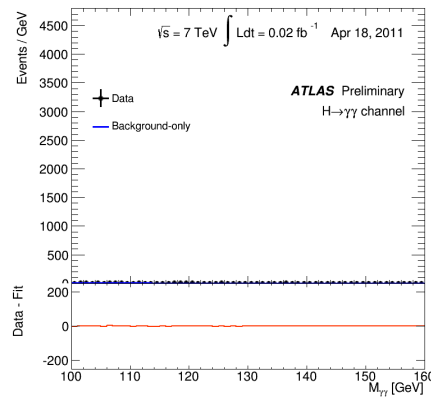
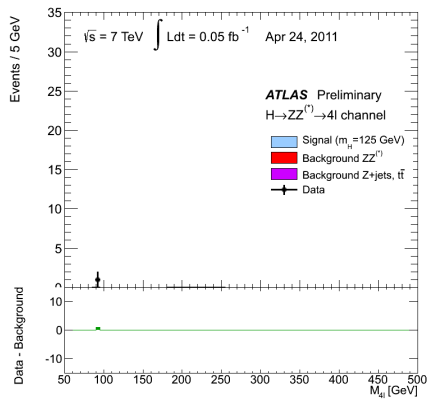
© 2012 CERN. Published by Elsevier B.V. Open access under [CC-BY-NC-ND license](#).

*Clear evidence for the production of a neutral boson
with a measured mass of 126 GeV [...] significance of 5.9 σ*

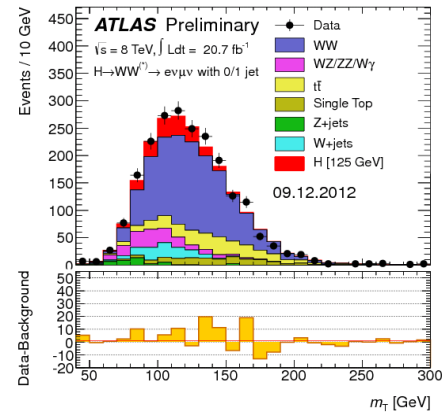
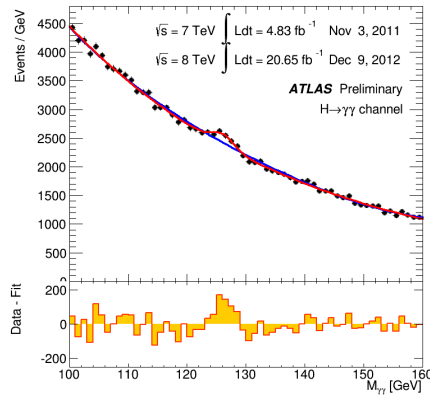
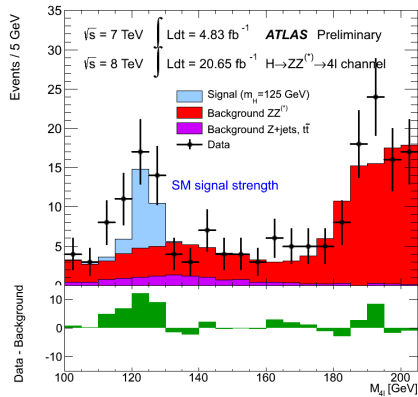
Branching ratios of the Higgs boson



A closer look at these discoveries



A closer look at these discoveries



The ATLAS experiment

The collaboration



- | | |
|----------------|--------------|
| Argentina | Morocco |
| Armenia | Netherlands |
| Australia | Norway |
| Austria | Poland |
| Azerbaijan | Portugal |
| Belarus | Romania |
| Brazil | Russia |
| Canada | Serbia |
| Chile | Slovakia |
| China | Slovenia |
| Colombia | South Africa |
| Czech Republic | Spain |
| Denmark | Sweden |
| France | Switzerland |
| Georgia | Taiwan |
| Germany | Turkey |
| Greece | UK |
| Israel | USA |
| Italy | CERN |
| Japan | JINR |

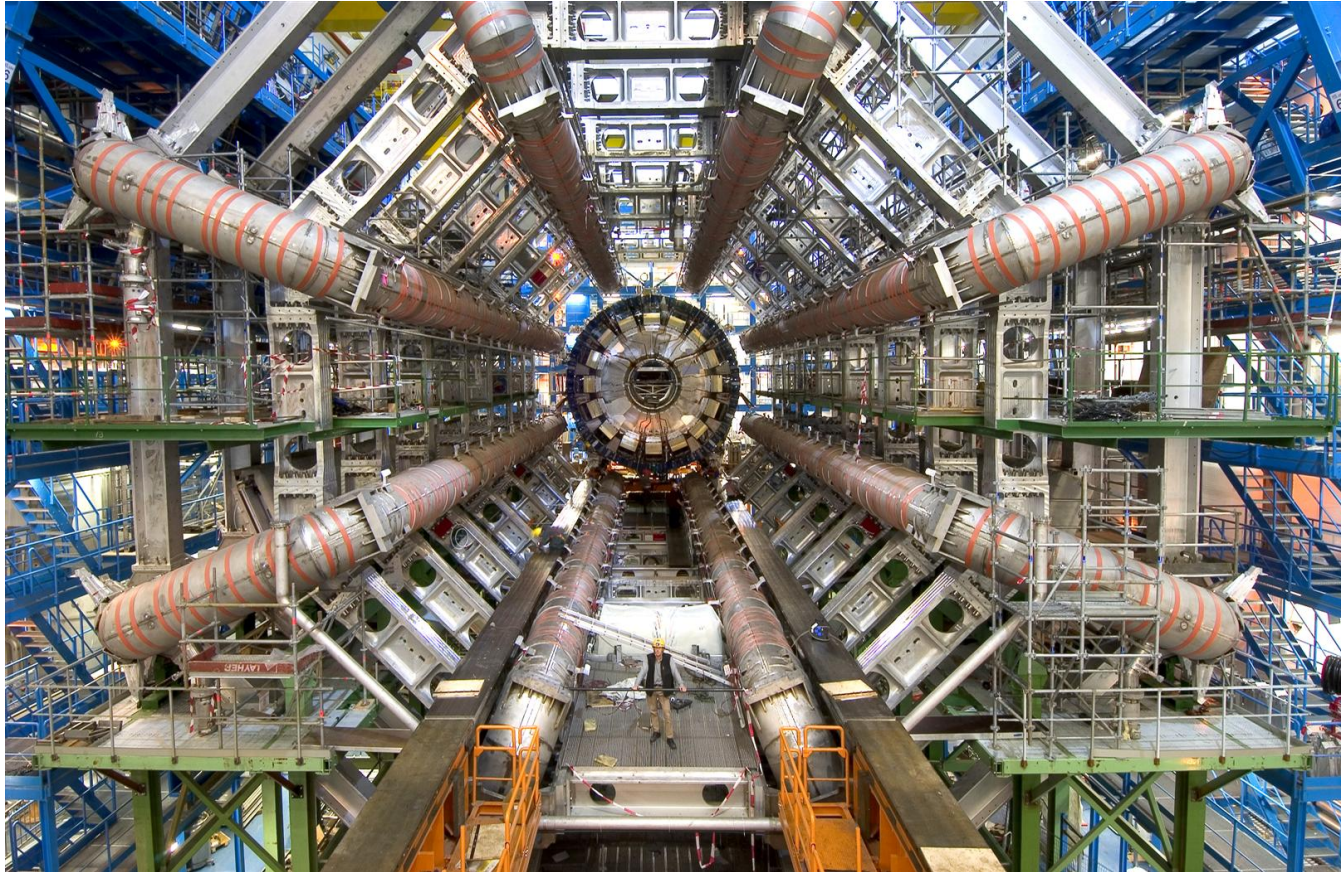
ATLAS Collaboration

180 institutions (235 institutes) from 38 countries



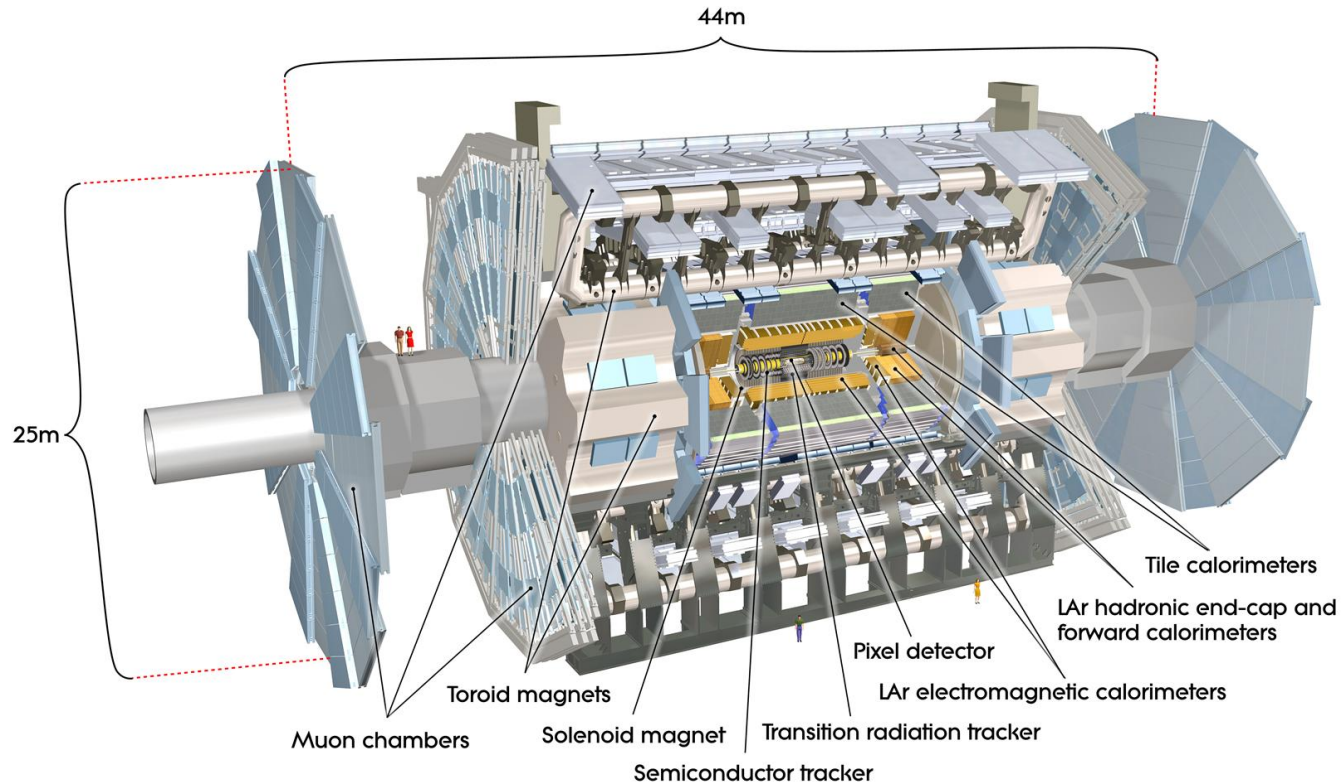
3,000 scientists!

The detector



The eight toroid magnets and the calorimeter.

The detector



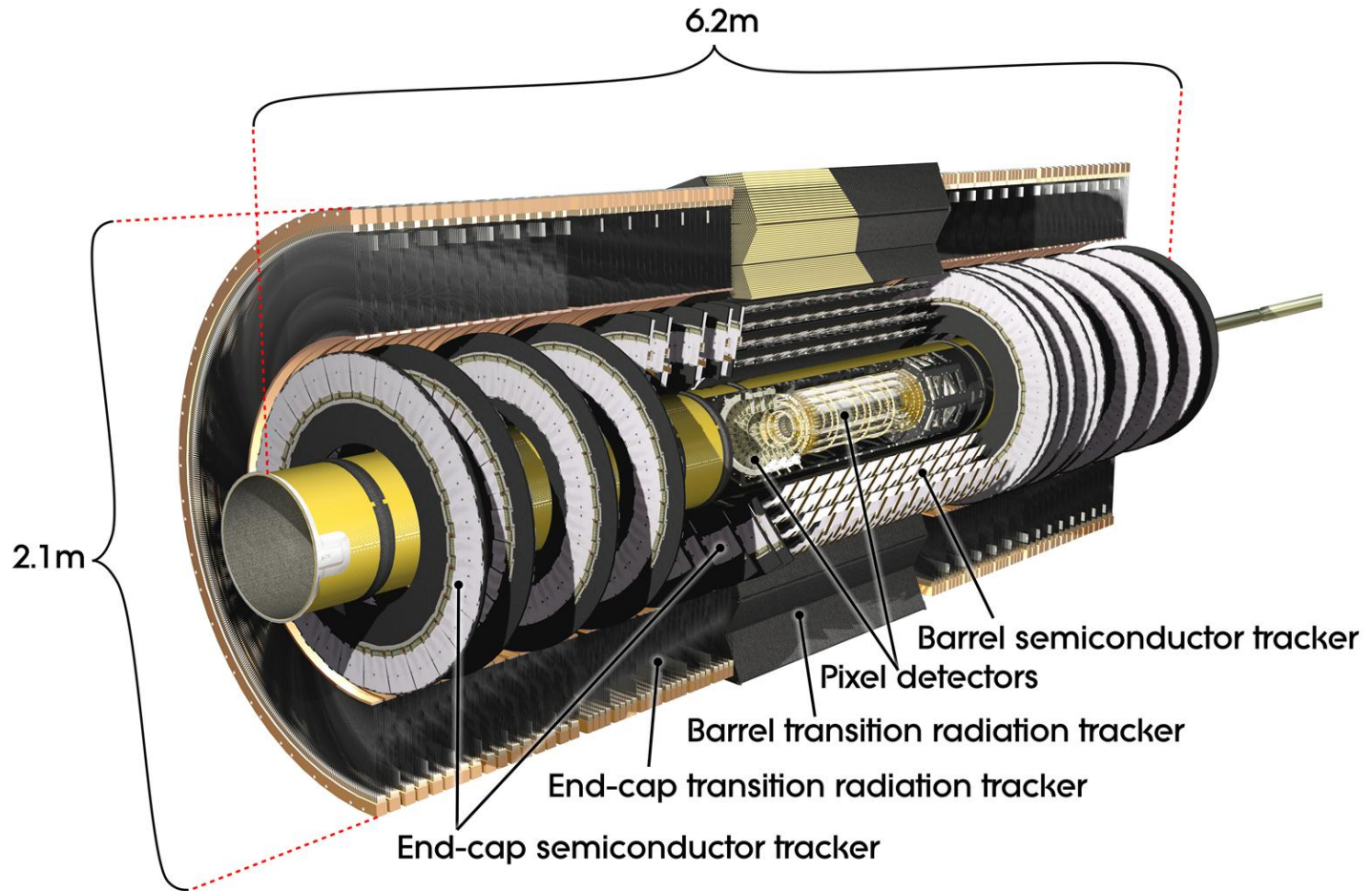
100 m underground / 7,000 tonnes / 100 million electronic channels / 3,000 km of cables

Detector technologies at a glance

Trackers measure the momentum of charged particles:

- gaseous detectors rely on the ionisation of gas (xenon, CO₂) and the ensuing **transition radiation**
- solid-state detectors enable the creation of electron-hole pairs in the dense material (Silicon), arranged in **strips** or **pixels**

Inner detector (tracker)

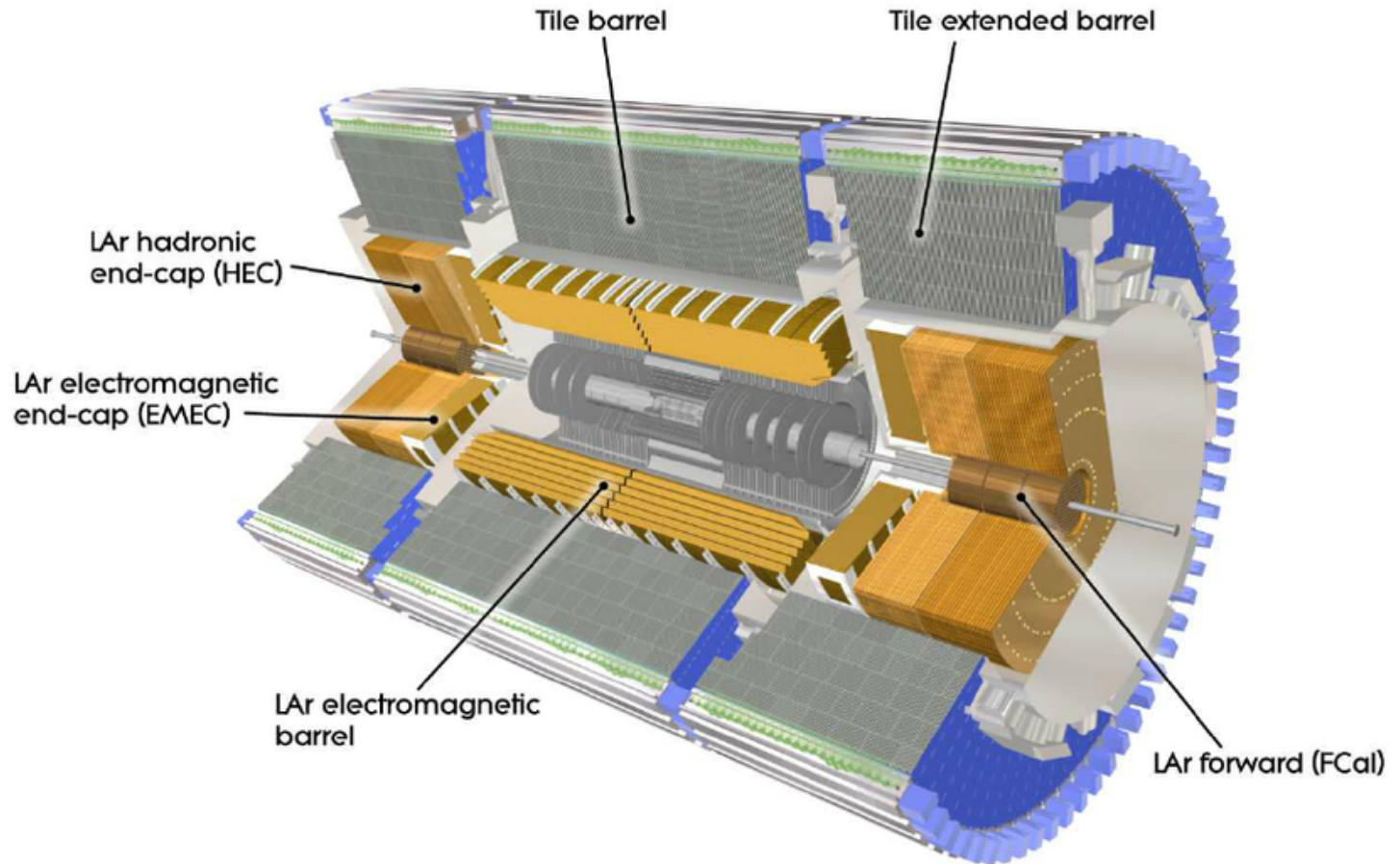


Detector technologies at a glance

Calorimeters measure the energy deposited by incoming particles as they travel through it:

- alternate layers of dense, absorbing material (iron or lead) and active medium (liquid argon, LAr)
- the master equation is $E = E_0 e^{-x/X_0}$ where X_0 is the **radiation length** \propto material
- electromagnetic cascade decays give precise measurements of electrons and photons, but hadronic cascade are much more complex
- in general, the energy resolution increases with energy

Calorimeters

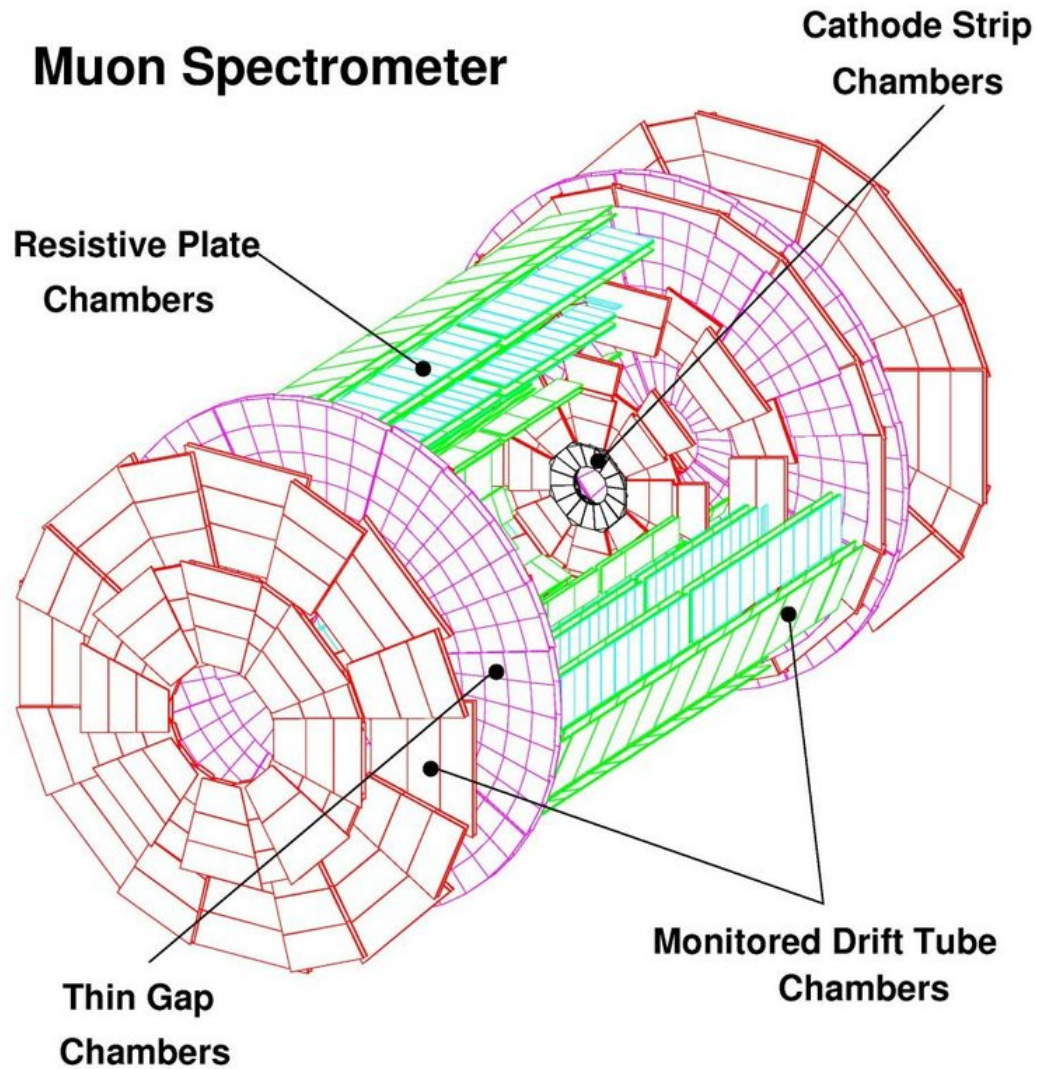


Detector technologies at a glance

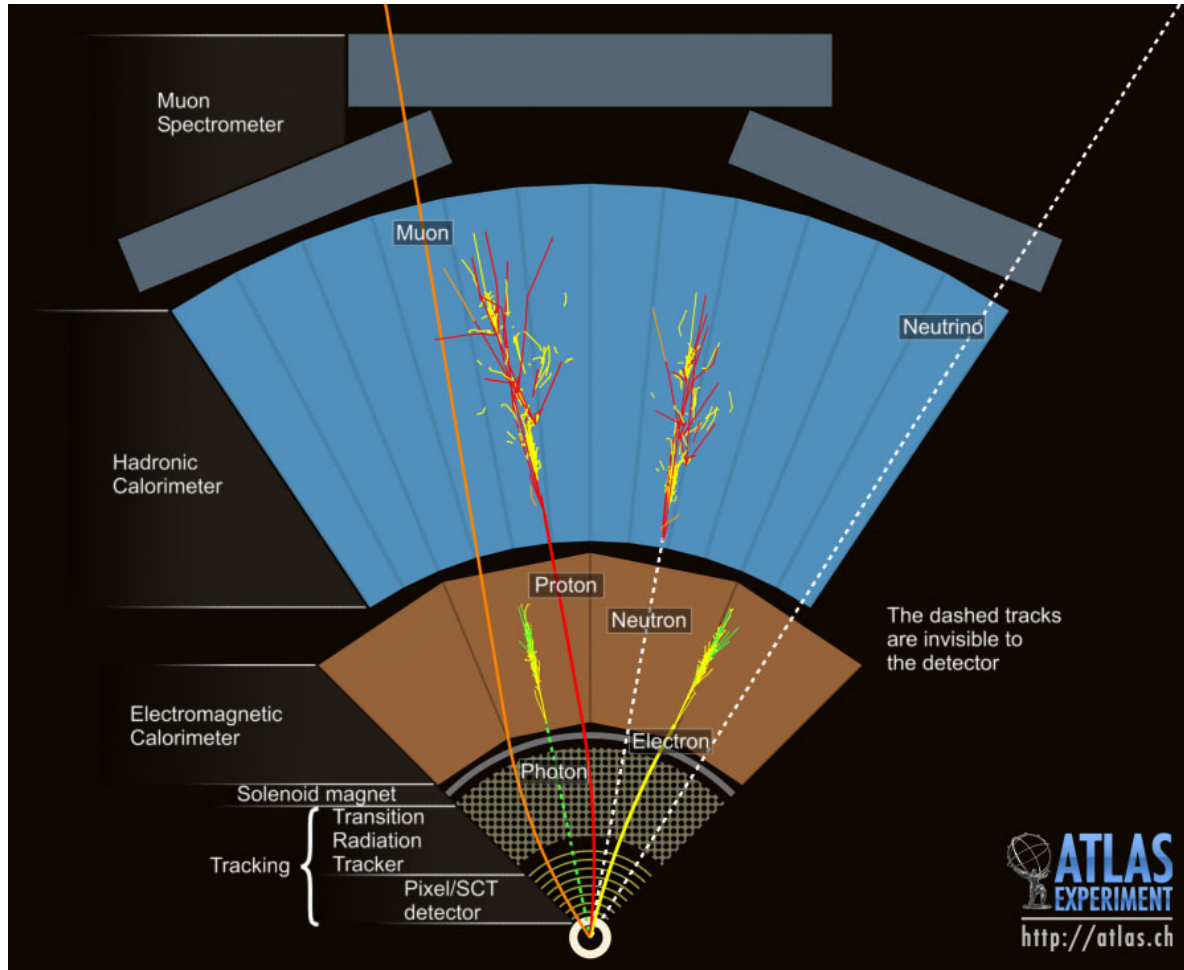
Muon spectrometers are essentially trackers too:

- use gas-based instrumentation (drift tubes, thin-gap and resistive-plate chambers) to measure the momentum of muons
- assuming that whatever made it through the ECAL+HCAL is a muon!

Muon spectrometer



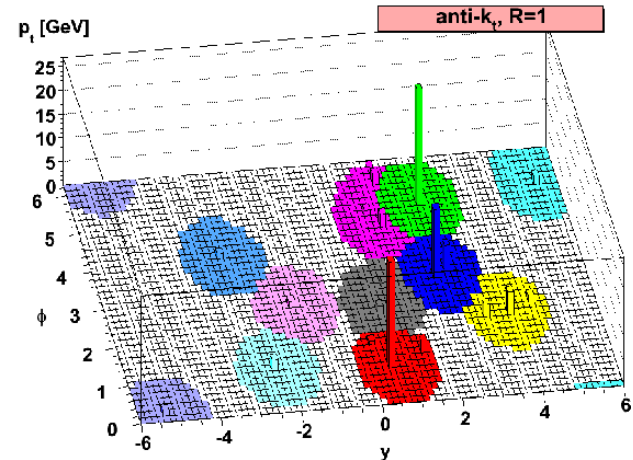
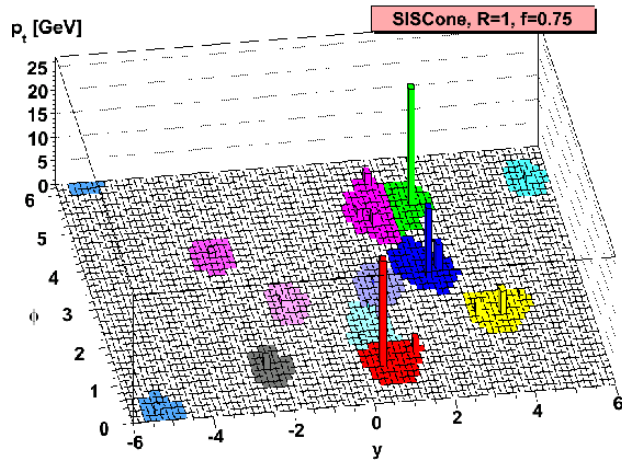
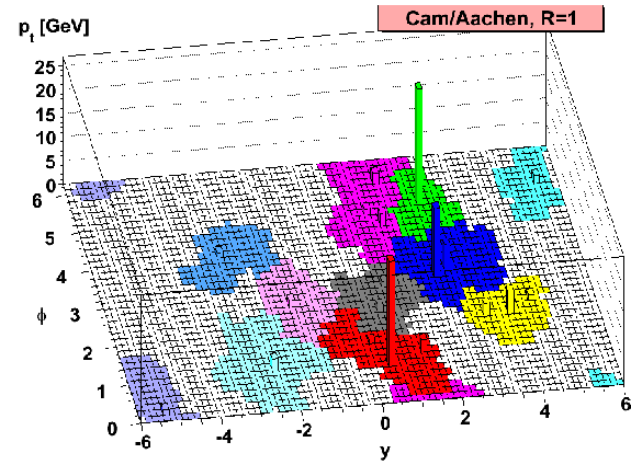
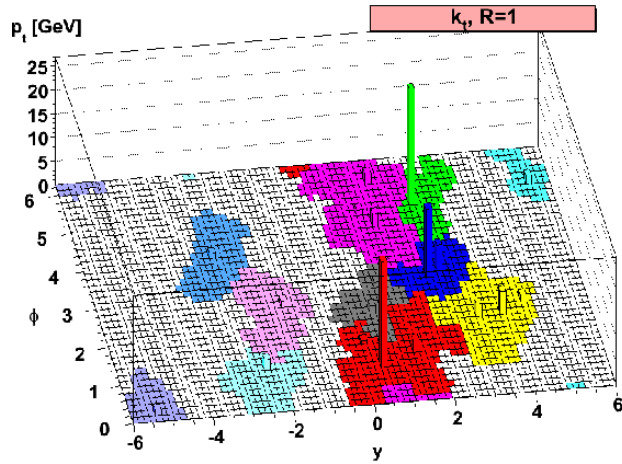
Particle identification in ATLAS



Quarks **hadronise** and form **jets**: messy! → many **jet constituents**

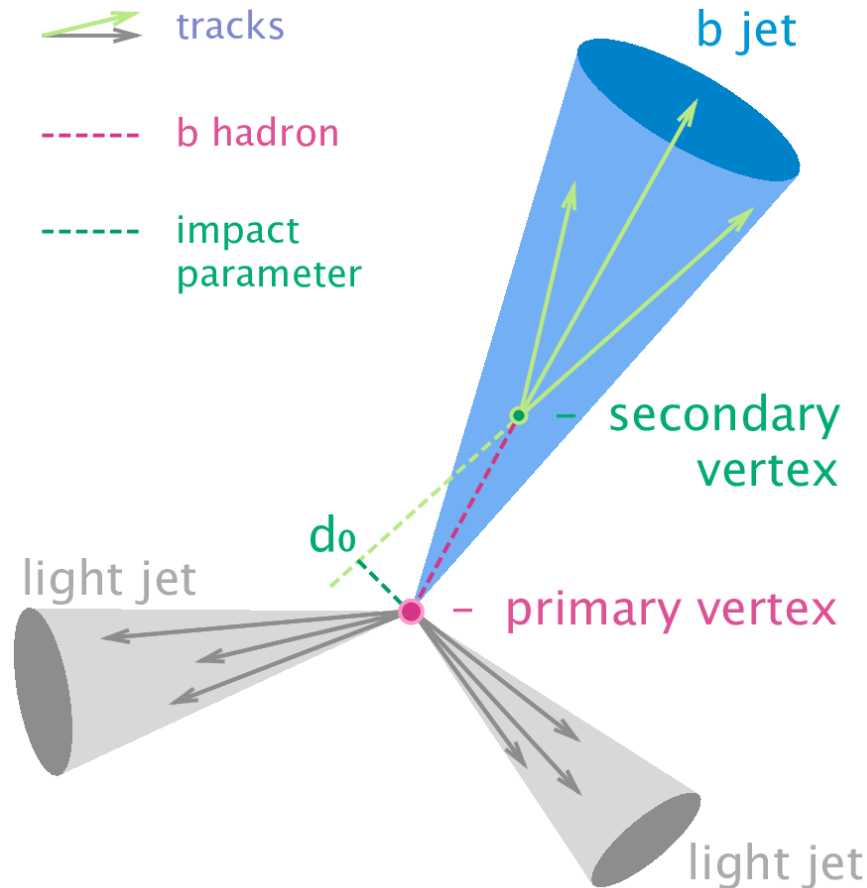
Particle identification in ATLAS

Jet clustering



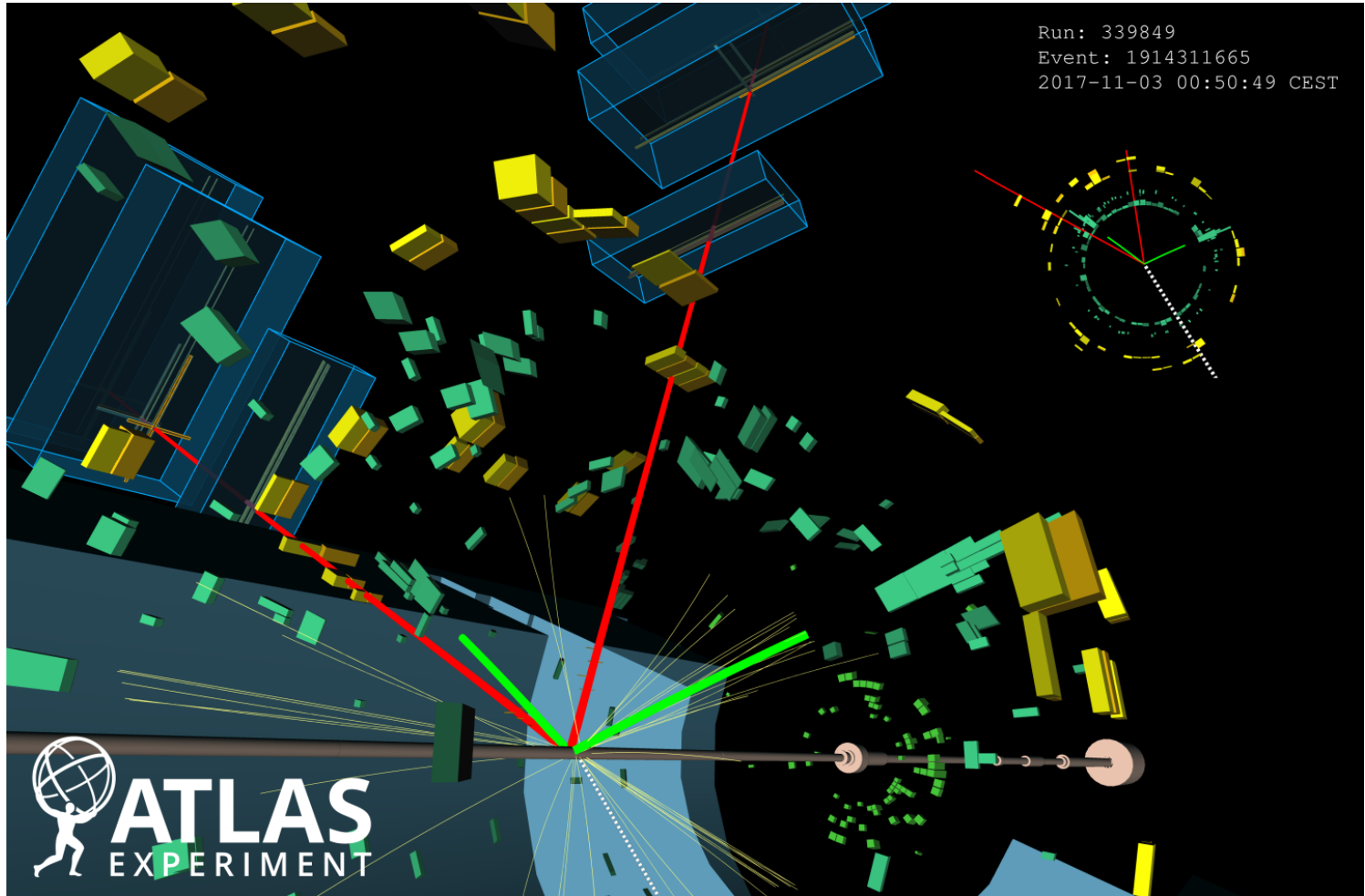
Particle identification in ATLAS

Jet tagging

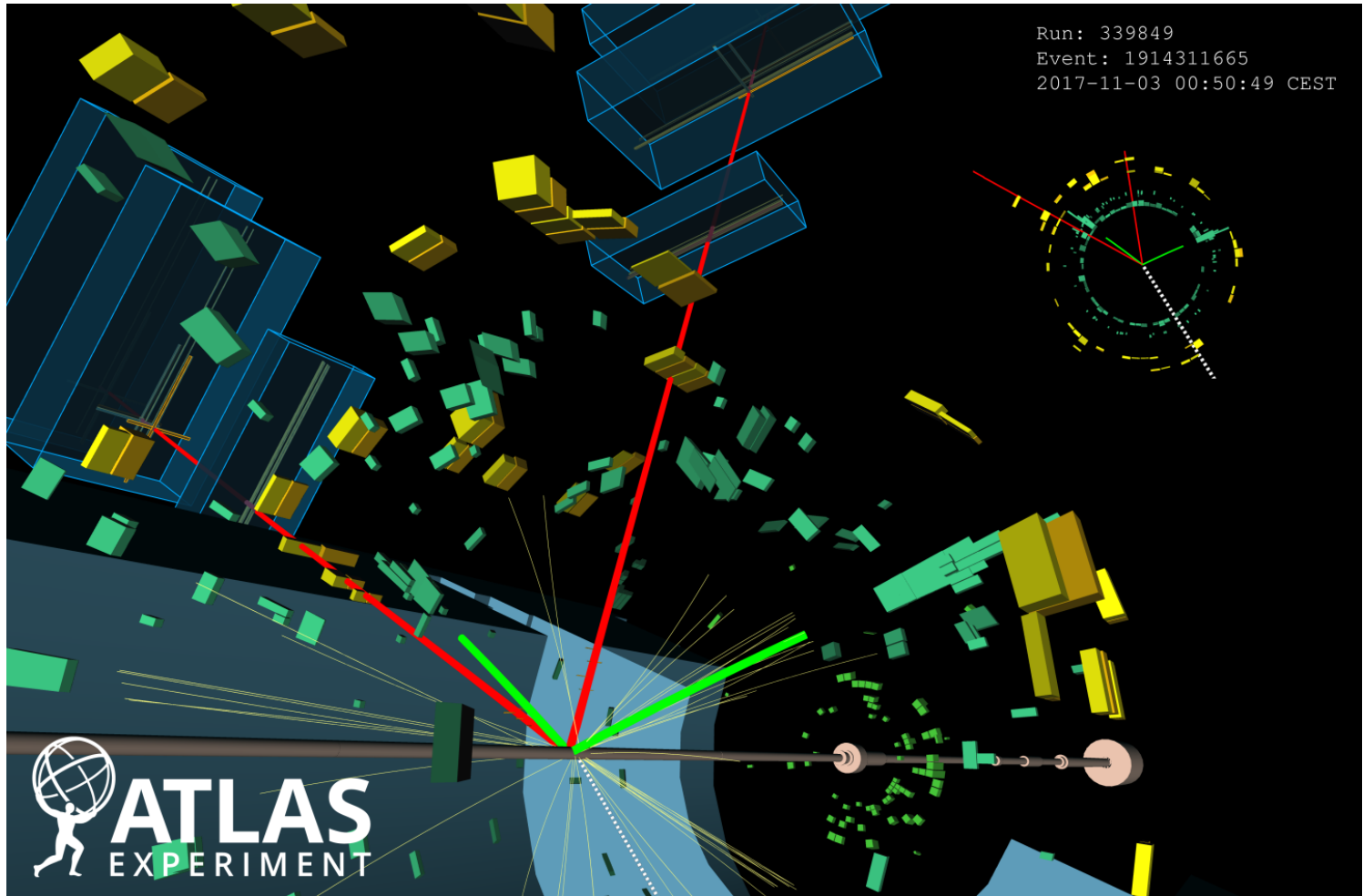


Quiz time: what's that Higgs?

Candidate event 1

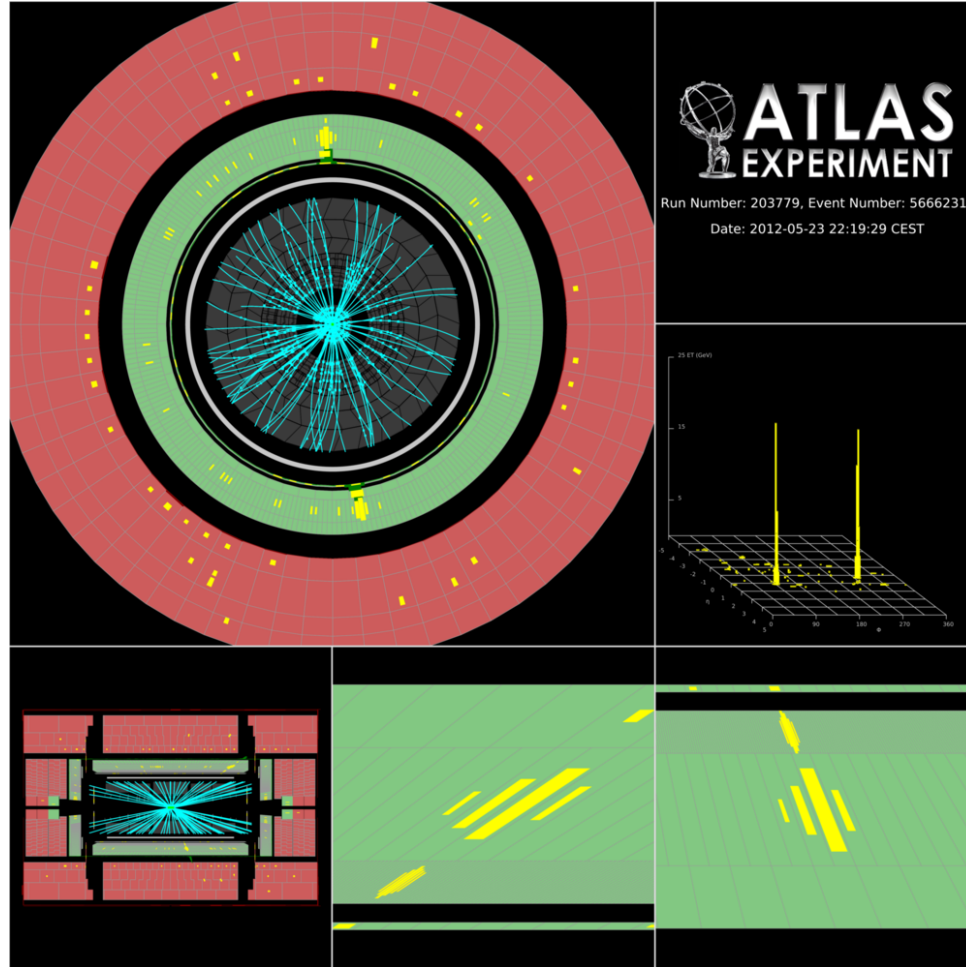


Candidate event 1

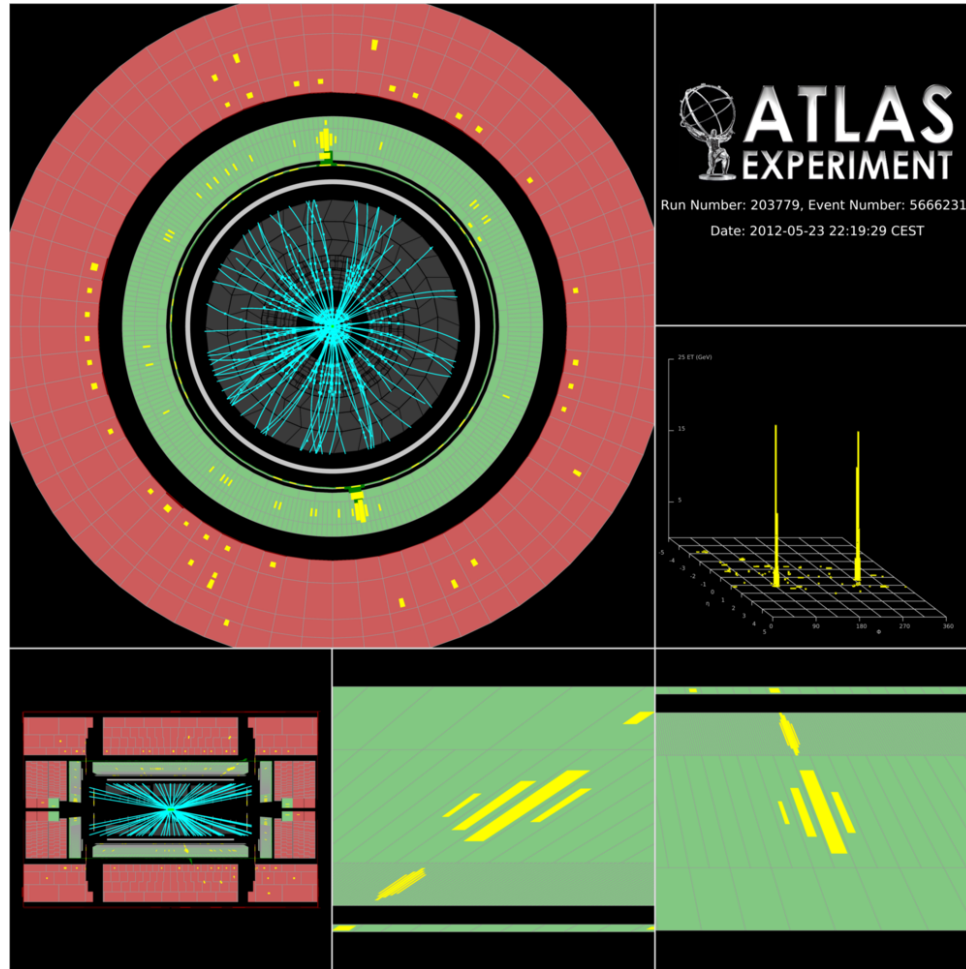


$$H \rightarrow ZZ^* \rightarrow e^+e^-\mu^+\mu^-$$

Candidate event 2



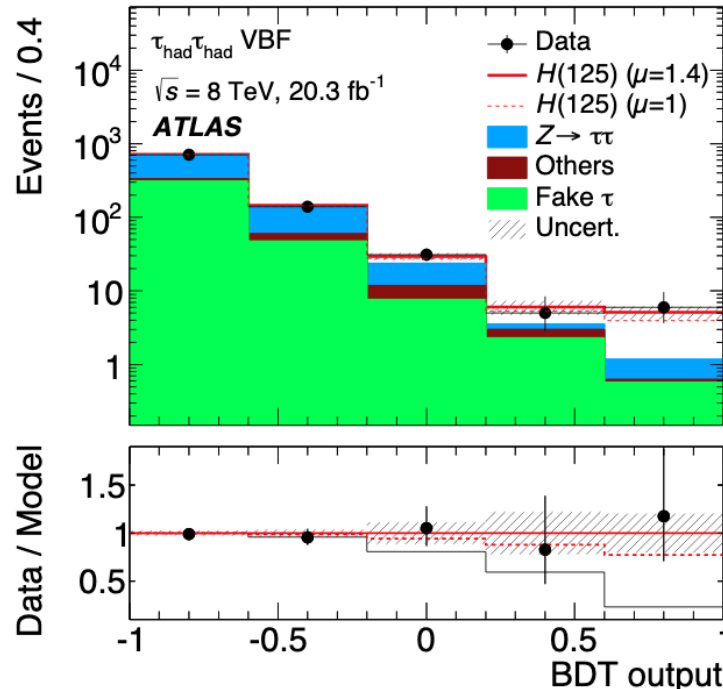
Candidate event 2



$$H \rightarrow \gamma\gamma$$

Machine learning at ATLAS

Early $H \rightarrow \tau^+ \tau^-$ at ATLAS



BDT (with ROOT TMVA) trained on ~ 10 variables, crucial for background discrimination and eventual observation of the Higgs coupling to taus! (4.5σ)

With and without Machine Learning...

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

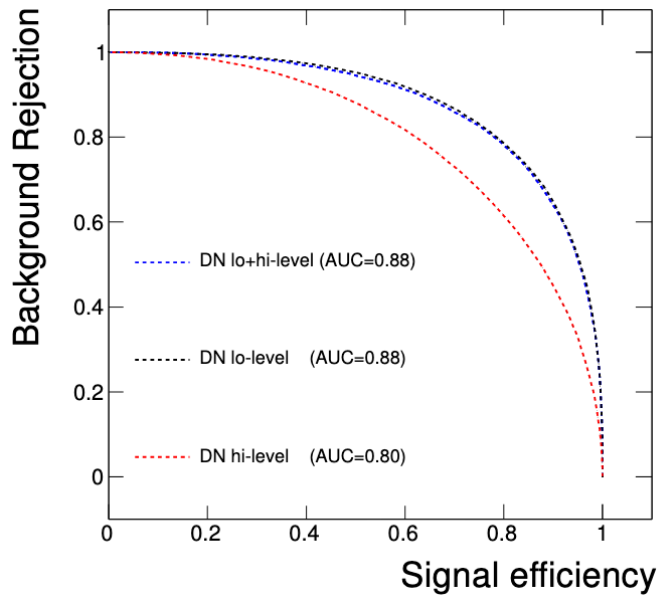
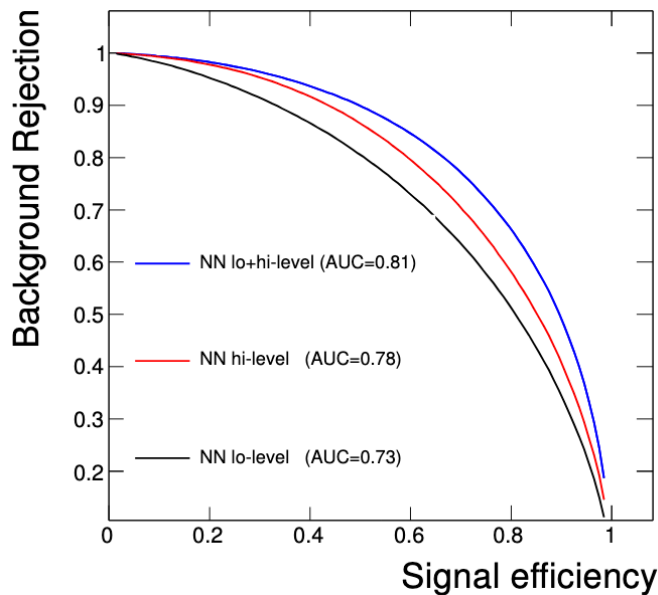
Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS ²⁴ $H \rightarrow \gamma\gamma$	2011–2012	2.2σ , $P = 0.014$	2.7σ , $P = 0.0035$	4.0	51%
ATLAS ⁴³ $H \rightarrow \tau^+\tau^-$	2011–2012	2.5σ , $P =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

Do **more** with **less**!

Machine learning instead of or beyond physics?

The $H \rightarrow \tau^+ \tau^-$ paper we saw previously used **high-level variables**: complex observables meant to represent **physical quantities of interest** (invariant masses, opening angles, sphericity, centrality...). These are very close to our understanding as (human!) physicists – it's how we usually approach the problem.

But is it the best way to go for a machine?



A booming field

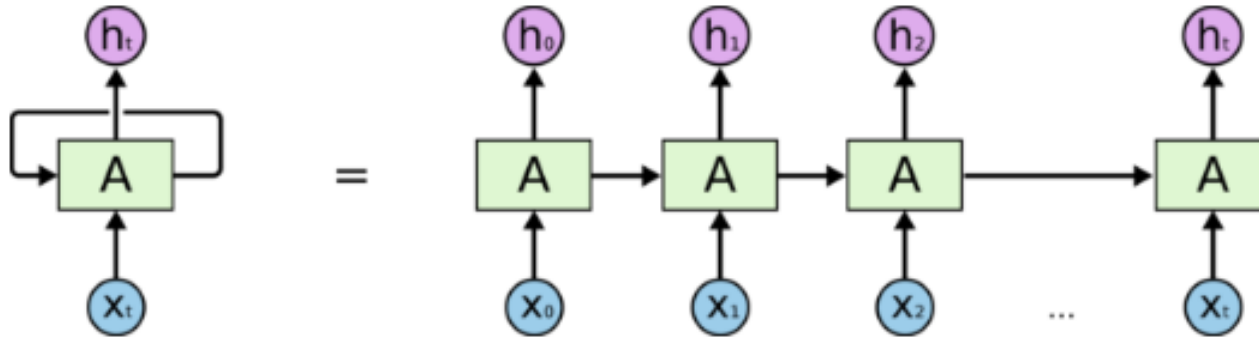
The number of applications of machine learning to high energy physics has **exploded** in recent years:

- (supervised) separation tasks: backgrounds vs signal
- sampling and optimisation of calibration algorithms
- reconstruction of particles (tracking, clustering)
- fast detector simulation
- adversarial networks to remove dependency on limiting systematic uncertainties
- efficient data compression and AI for triggers
- transfer learning
- graphs and sets to deal with many-particle systems
- likelihood-free inference
- **anomaly detection**

Tackling variability of inputs with **Recurrent Neural Networks**

Recurrent units

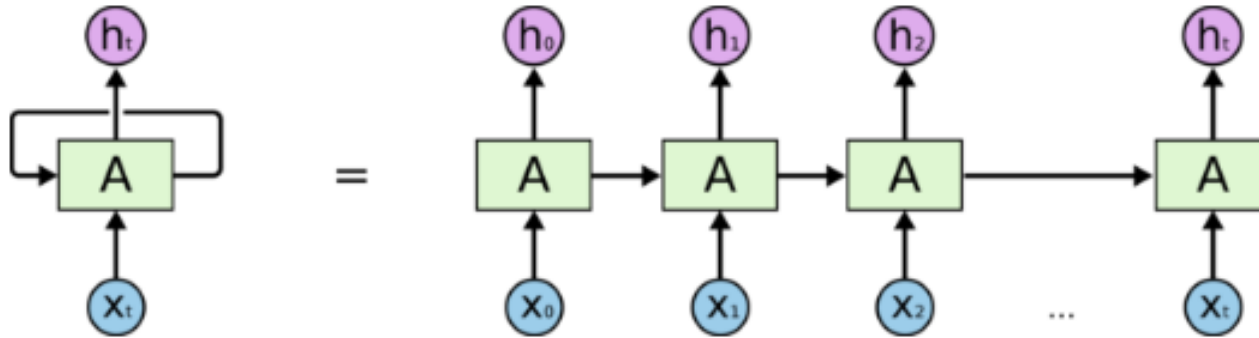
Introduce the concept of **states** to deal with **sequences**: time series, chains of molecules, syntactic elements, etc.



The state h_t is a function of the previous state h_{t-1} and the input feature x_t . In a sense, the network "replicates" itself with each pass (unfolding/unrolling) – it has a notion of "memory".

Recurrent units

Introduce the concept of **states** to deal with **sequences**: time series, chains of molecules, syntactic elements, etc.

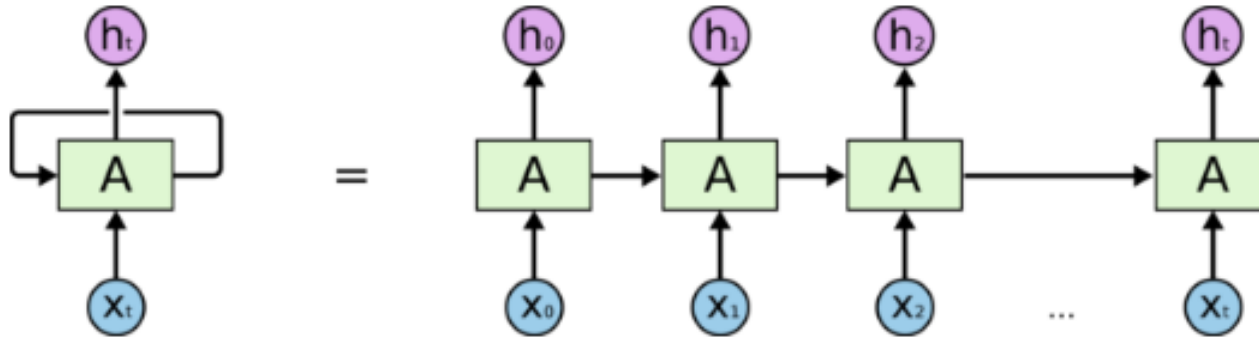


The state h_t is a function of the previous state h_{t-1} and the input feature x_t . In a sense, the network "replicates" itself with each pass (unfolding/unrolling) – it has a notion of "memory".

One common problem with long chains of RNN cells: **vanishing** or **exploding** gradients. The latter can be fixed with e.g. **gradient clipping**.

Recurrent units

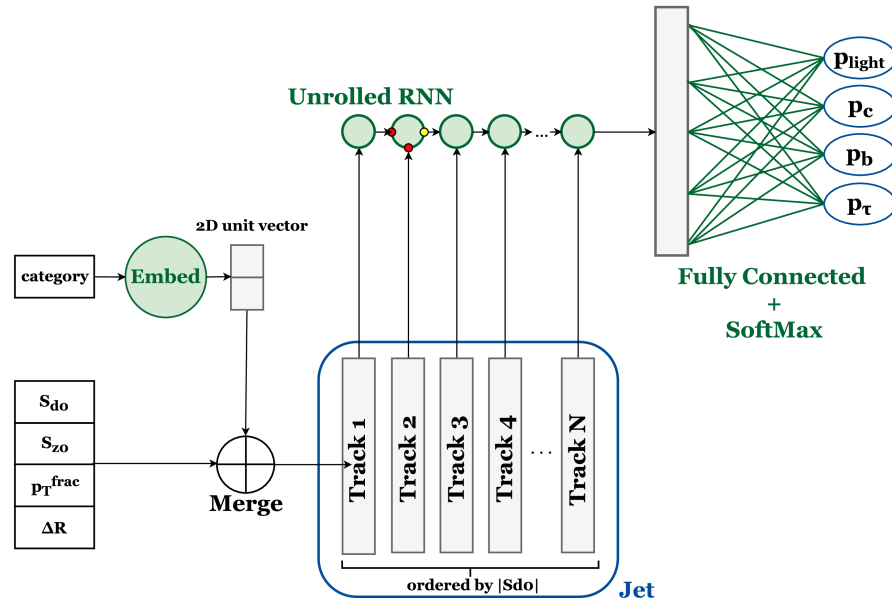
Introduce the concept of **states** to deal with **sequences**: time series, chains of molecules, syntactic elements, etc.



The state h_t is a function of the previous state h_{t-1} and the input feature x_t . In a sense, the network "replicates" itself with each pass (unfolding/unrolling) – it has a notion of "memory".

One common problem with long chains of RNN cells: **vanishing** or **exploding** gradients. The latter can be fixed with e.g. **gradient clipping**. To tackle the former, more complex cell structures can be used: **gated recurrent units** (GRUs) and **long short-term memory** networks (LSTM) are popular examples. They allow for "forget-gates" to regularise the learning (a bit like how dropout layers are used in DNNs)

Recurrent NNs for b-tagging

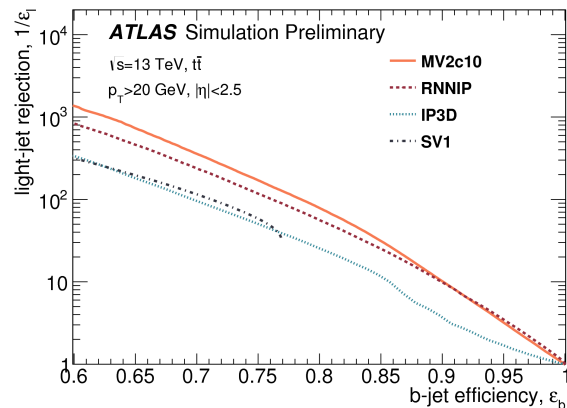
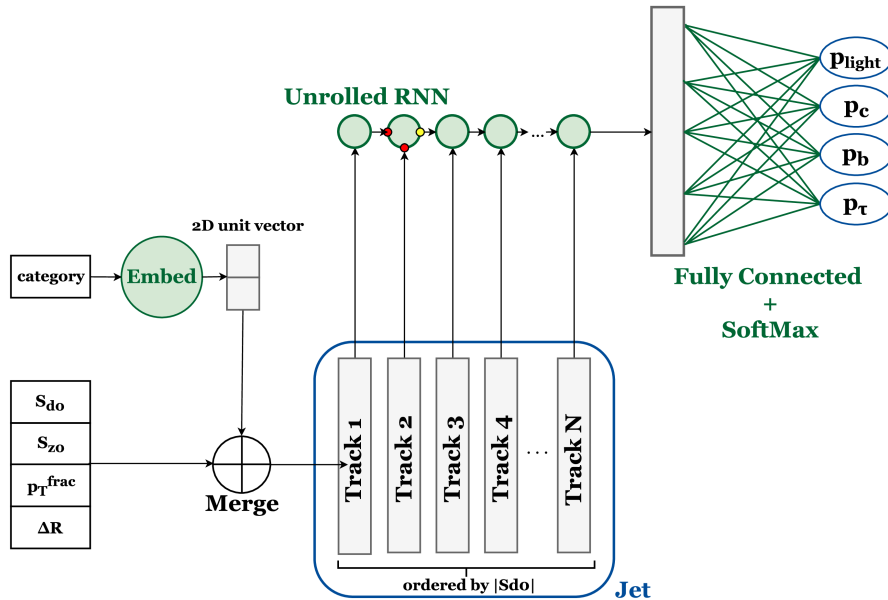


Standard neural networks and BDTs are **ill-suited** to the problem of dealing with a **variable number of jet constituents and tracks**.

Instead, take a list of tracks inside a jet and feed it to an RNN. The ordering is physics-inspired: significance of the impact parameter.

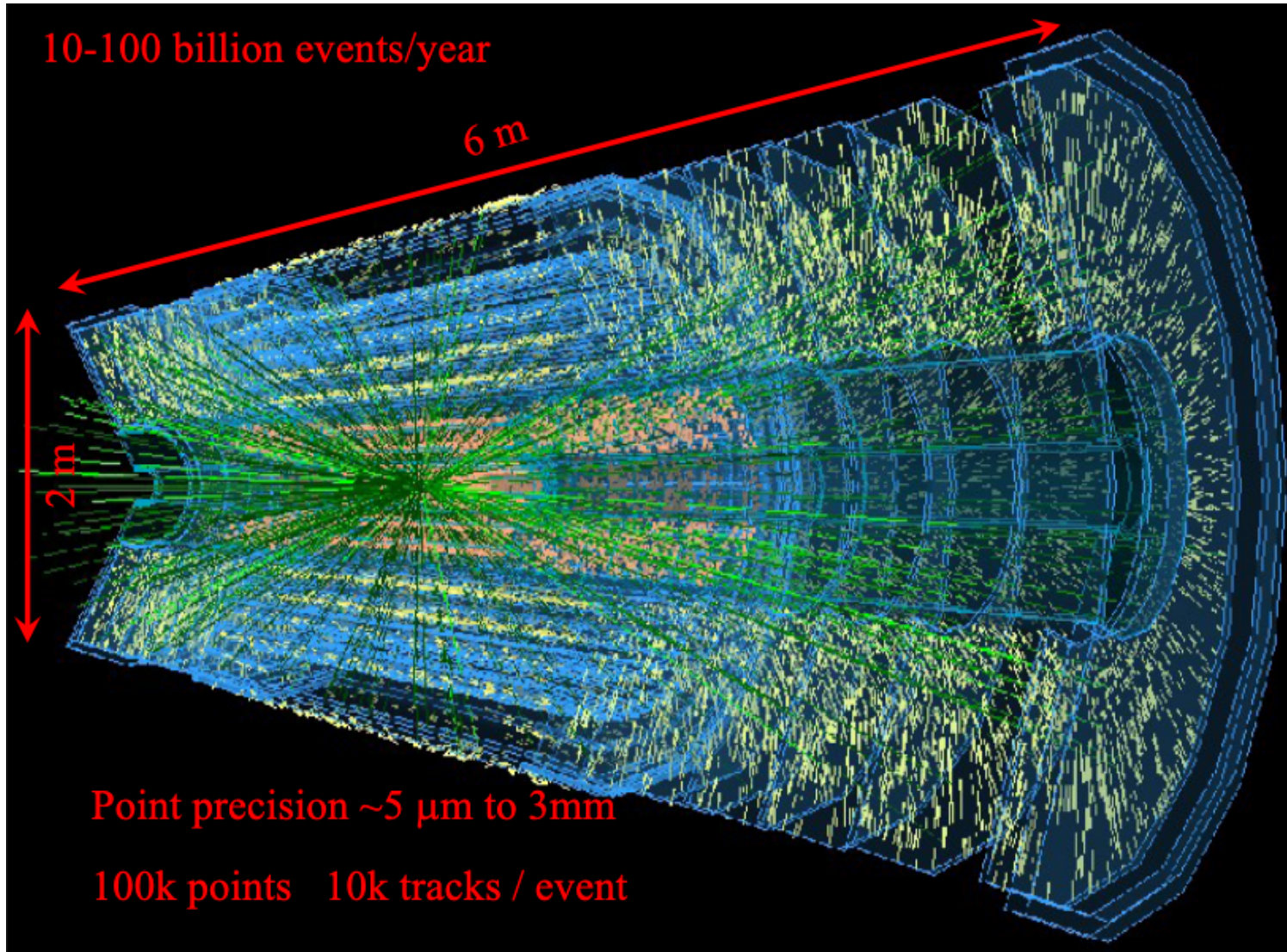
Without expliciting vertexing, excellent performance is still achieved!

Recurrent NNs for b-tagging



Staying **on track**(s): a quick interlude

The tracking crisis



Crowd-source it!

A banner for a Kaggle competition. The background is dark blue with a complex pattern of yellow and red lines and dots, resembling particle tracks. In the top left, there is a small icon of a person and the text "Featured Prediction Competition". The main title "TrackML Particle Tracking Challenge" is in large white font, followed by the subtitle "High Energy Physics particle tracking in CERN detectors". On the right side, "\$25,000" is written in large white font, with "Prize Money" below it. In the bottom left, there is a CERN logo and the text "CERN · 651 teams · a year ago".

Featured Prediction Competition

TrackML Particle Tracking Challenge
High Energy Physics particle tracking in CERN detectors

\$25,000
Prize Money

CERN · 651 teams · a year ago

Competition open to public on [Kaggle](#) → can amateurs do better than the pros?

Crowd-source it!

A banner for the TrackML Particle Tracking Challenge. The background is a dark blue field filled with a complex network of yellow and red lines and dots, representing particle tracks in a detector. In the top left corner, there is a small icon of a person and the text "Featured Prediction Competition". The main title "TrackML Particle Tracking Challenge" is in large white font, with the subtitle "High Energy Physics particle tracking in CERN detectors" below it. On the right side, "\$25,000 Prize Money" is displayed in white. In the bottom left corner, the CERN logo is followed by the text "CERN · 651 teams · a year ago".

Featured Prediction Competition

TrackML Particle Tracking Challenge
High Energy Physics particle tracking in CERN detectors

\$25,000
Prize Money

CERN · 651 teams · a year ago

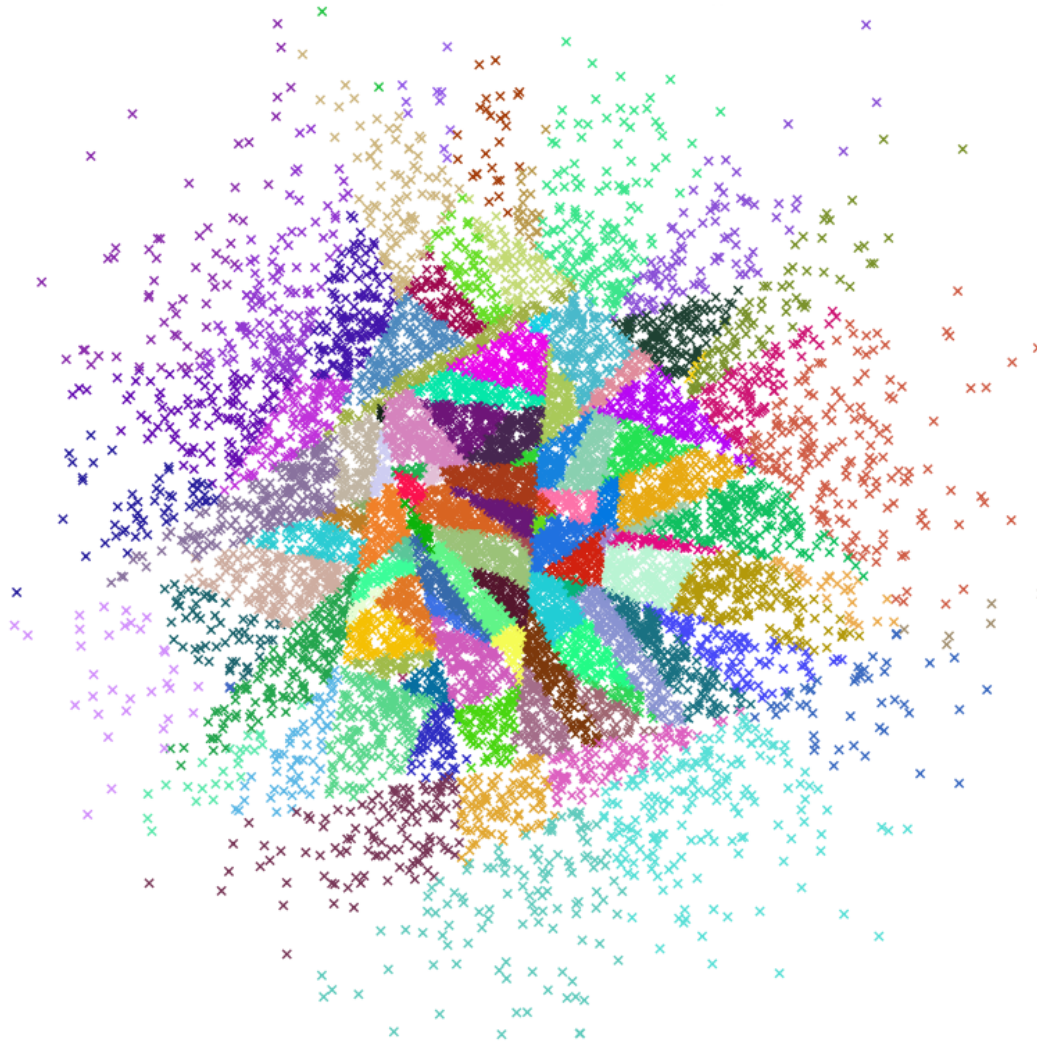
Competition open to public on [Kaggle](#) → can amateurs do better than the pros?

Sometimes, yes! New approaches (including deep learning) being folded in new generation of tracking algorithms at ATLAS :)

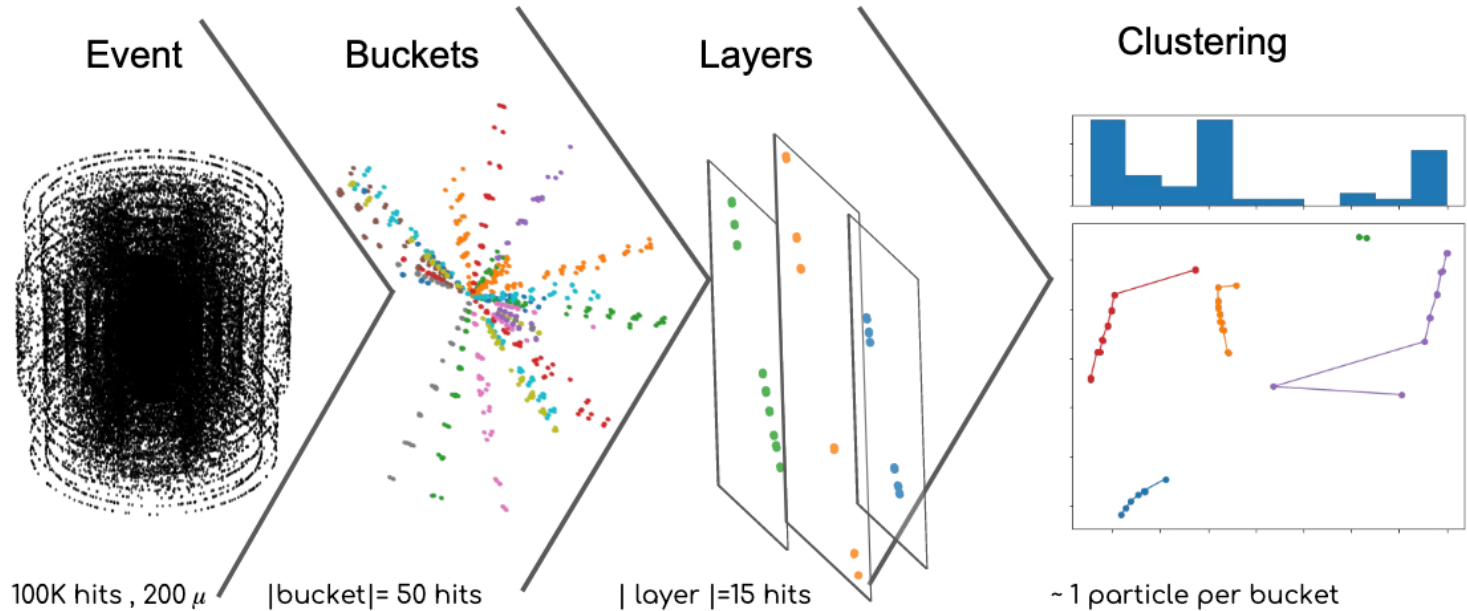
Spotify...

The screenshot displays the Spotify mobile application interface. At the top left, the Spotify logo is visible. Below it, a navigation menu includes 'Accueil', 'Rechercher', and 'Bibliothèque'. Under 'ÉCOUTÉ RÉCEMMENT', there are links for 'Valencià PLAYLIST', 'Valencià ALBUM', and 'Calm Vibes'. The main content area features a large album cover for 'Si al valencià' with the text 'Si al valencià' in speech bubbles. To the right, a list of songs is shown, including 'El Meu Poble', 'Governant', 'Corazón Viajero', 'Camp de Batalla', and 'Com Dues Gotes d'Aigua', all by 'Txarango · Som Riu'. The bottom section is titled 'Autoplay' and contains the text 'Autoplay similar songs when your music ends.' with a green toggle switch turned on. Below this are buttons for 'SHOW ADVANCED SETTINGS', 'LOG OUT', and 'About Spotify'. At the very bottom, the user's profile 'greysab' is shown, along with the current song 'El Meu Poble' by 'Txarango' and a progress bar indicating 1:56 out of 4:24.

Spotify-annoy!



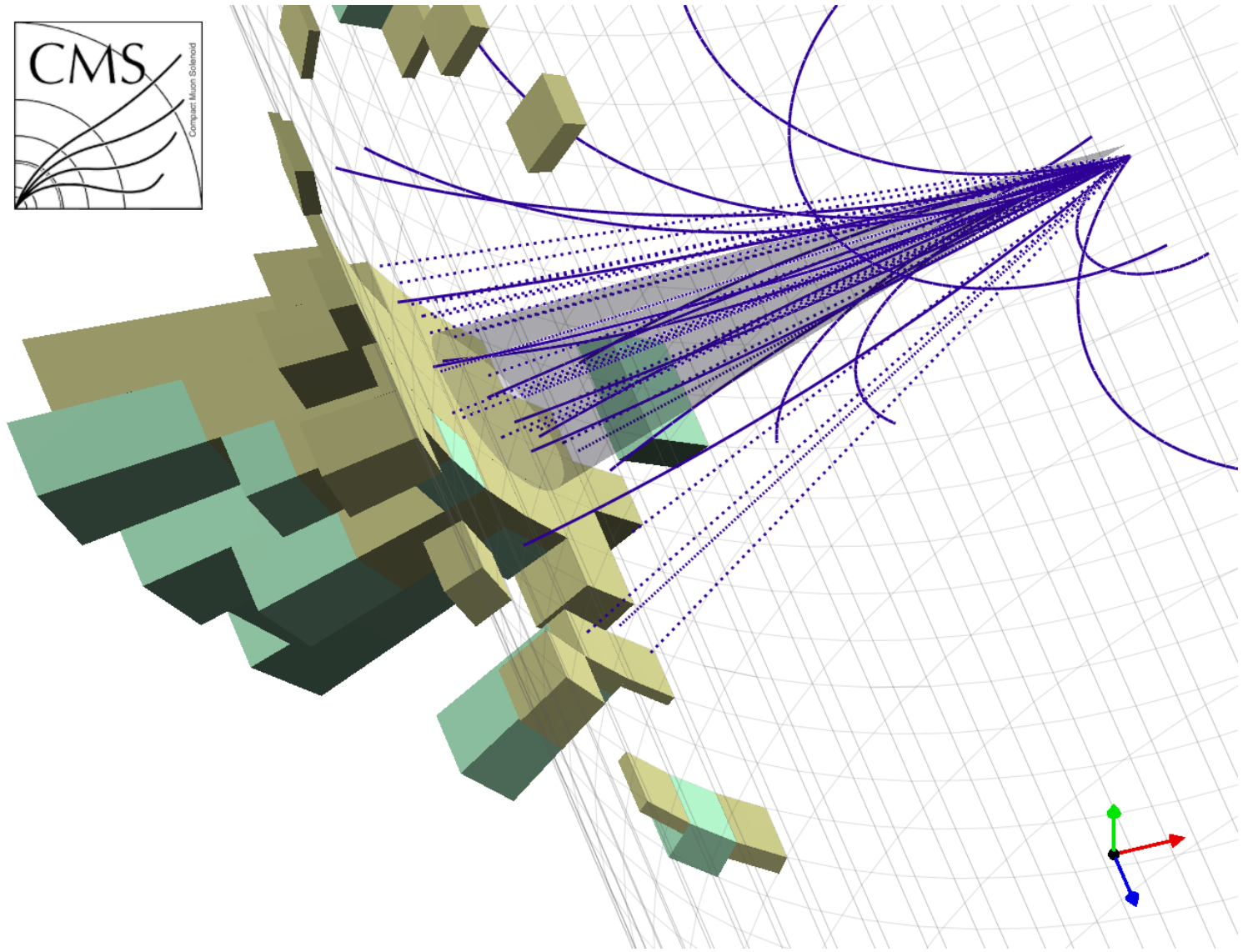
Introducing bucketing



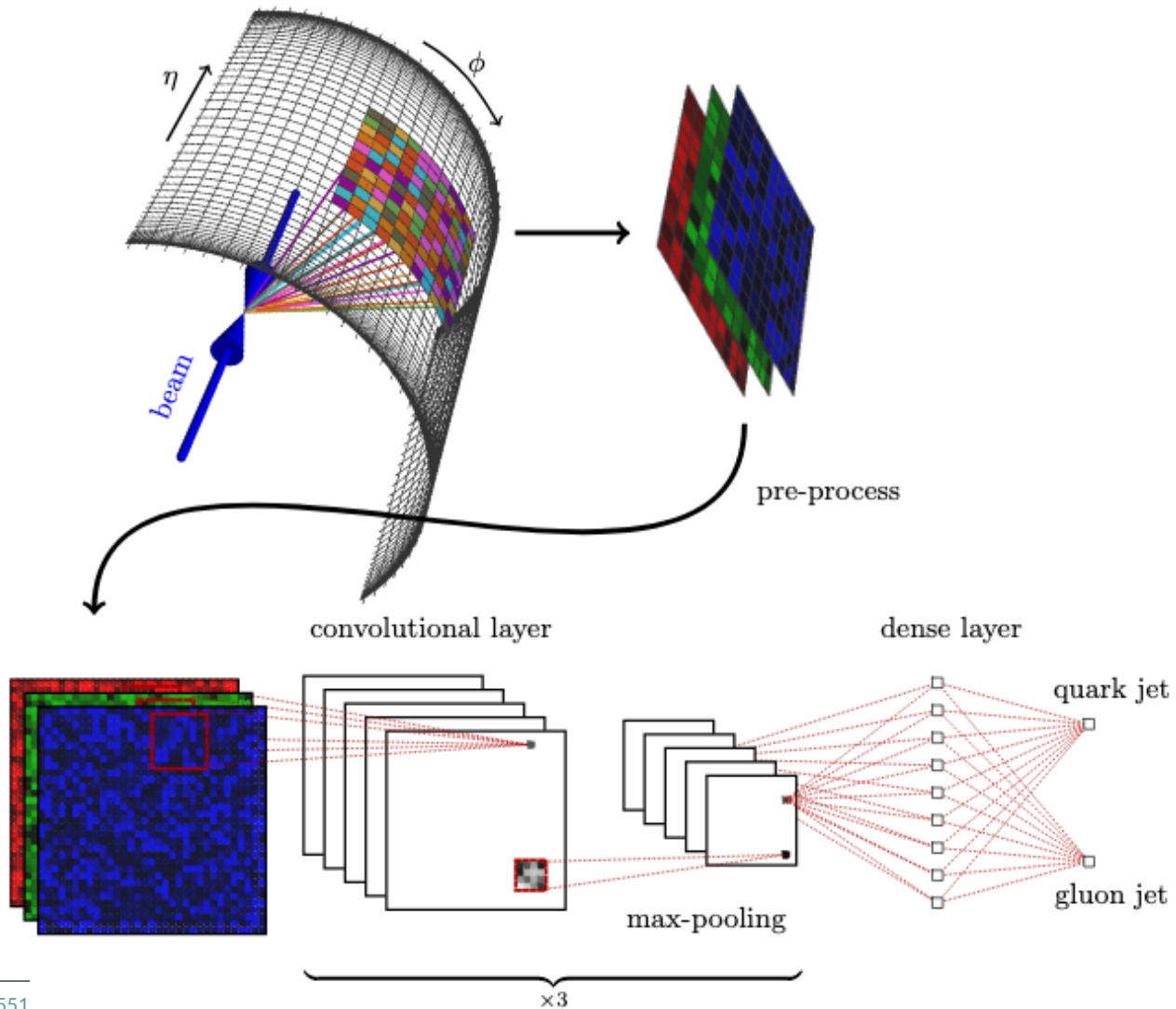
Drastic reduction in complexity! Allows for supervised learning on "buckets", or [unsupervised clustering approaches](#) (we'll come back to those in a moment)...

Jet images and **Convolutional Neural Networks**

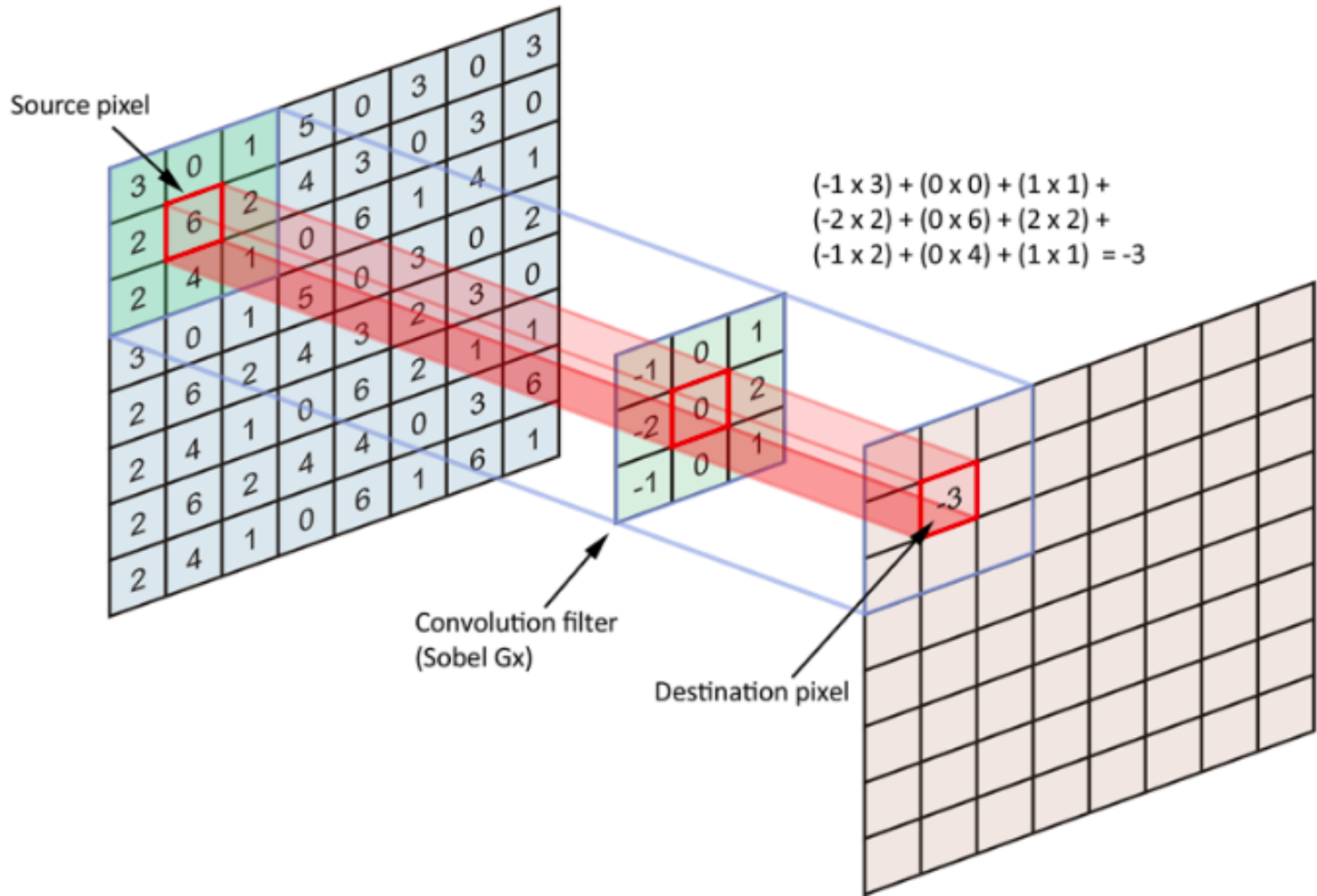
Jet images



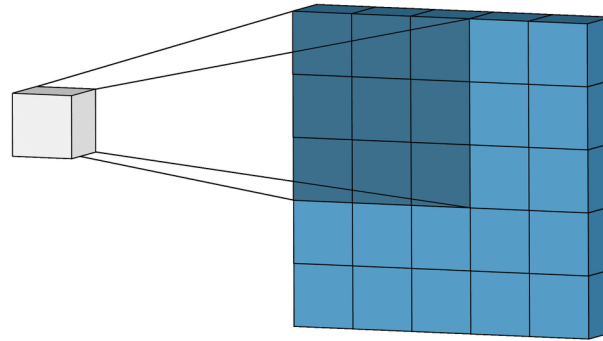
Jet images



Convolutional layers



Convolutional layers in action



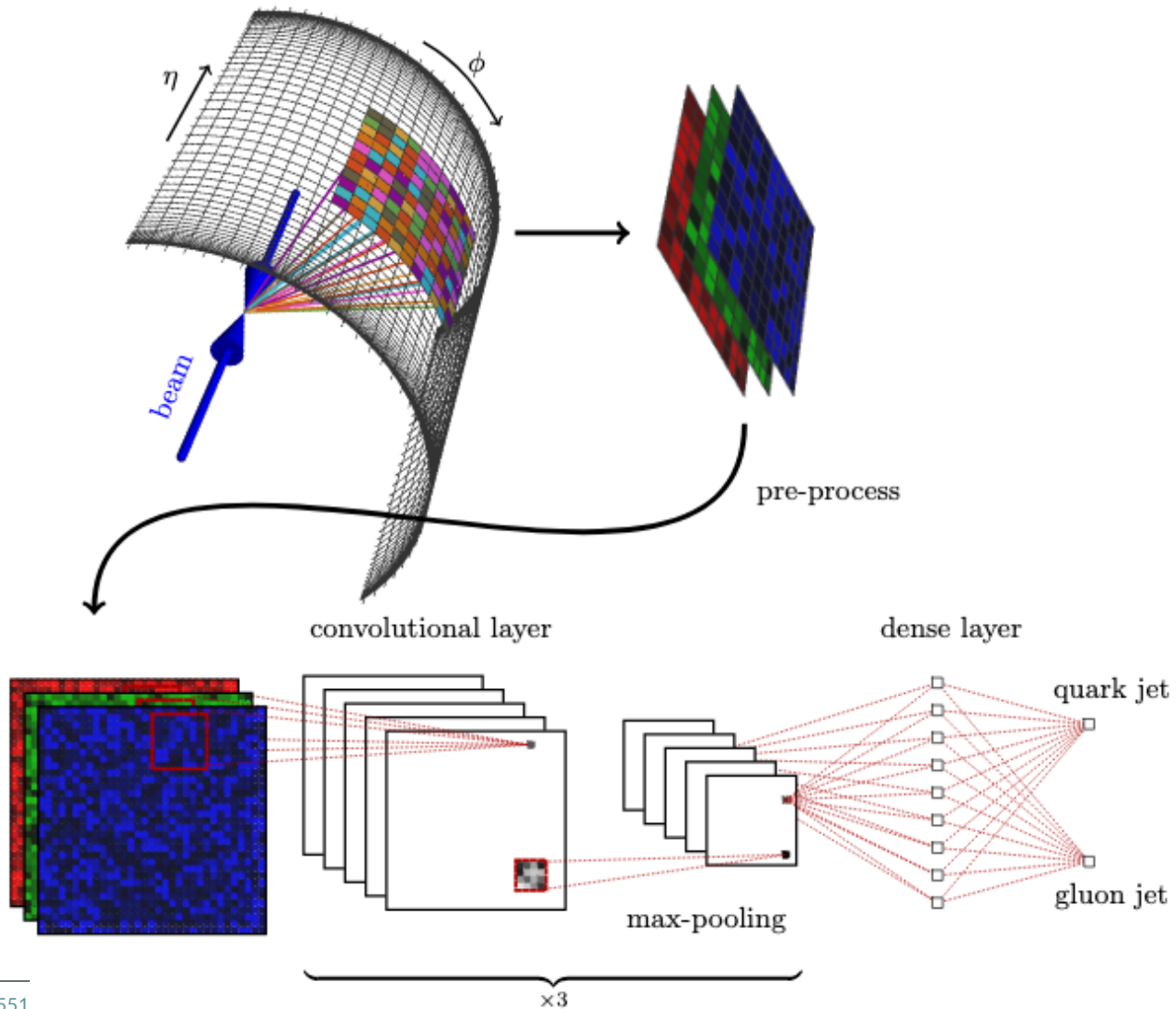
1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

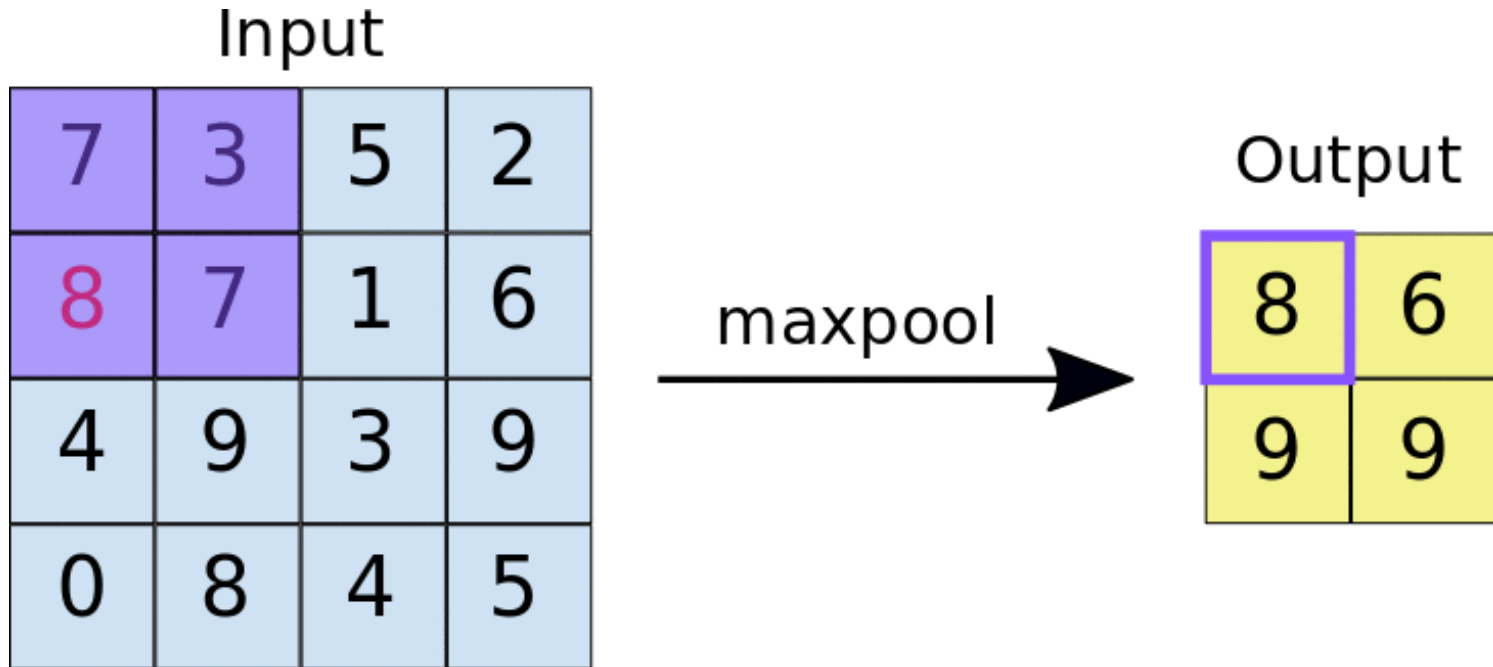
4		

Convolved
Feature

Jet images



Max-pooling



Max-pooling is a [sample-based discretisation process](#): reduce the dimensionality of the current layer by [down-sampling](#). This allows to focus on specific features.

Note: like the convolution, this operation can have overlaps!

An example: convolving cats

Imagine you have built a CNN to [identify animals](#). You may start with a picture of a cat:



An example: convolving cats

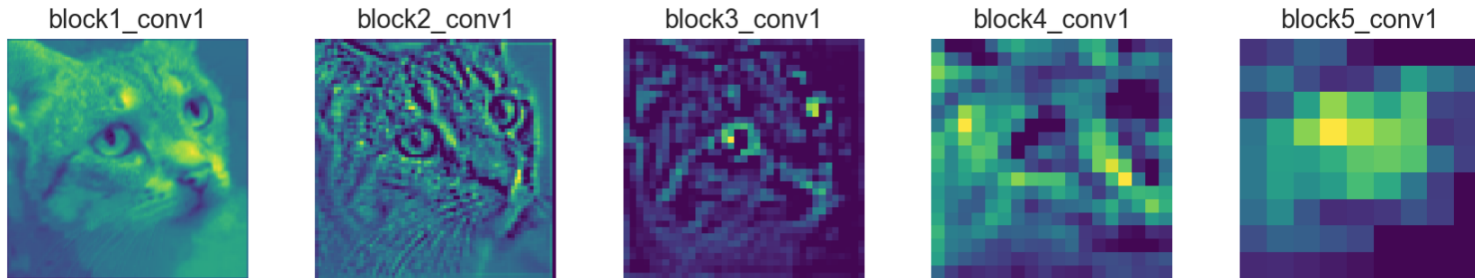
A neat trick to make your dataset richer and **less sensitive to variations** is **data augmentation**:



By **translating**, **rotating** and **smearing** this (poor) cat, we generate "new" data and tell the CNN "not to cheat" by assuming all cats are perfectly centered in the frame, look to the right, are the right way up etc.

An example: convolving cats

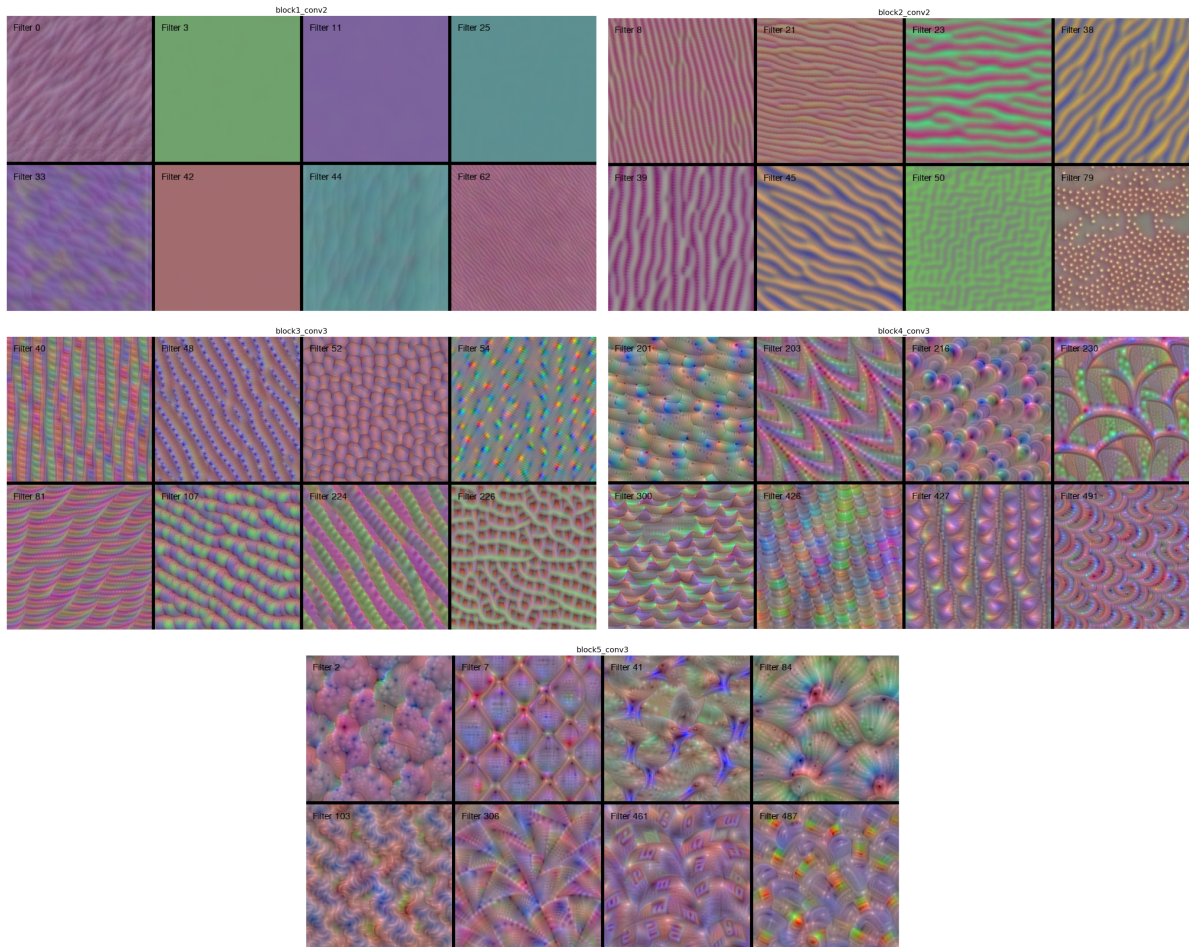
By passing our original image through each **convolutional filter**, it's easy to visualise what information the CNN is actually extracting:



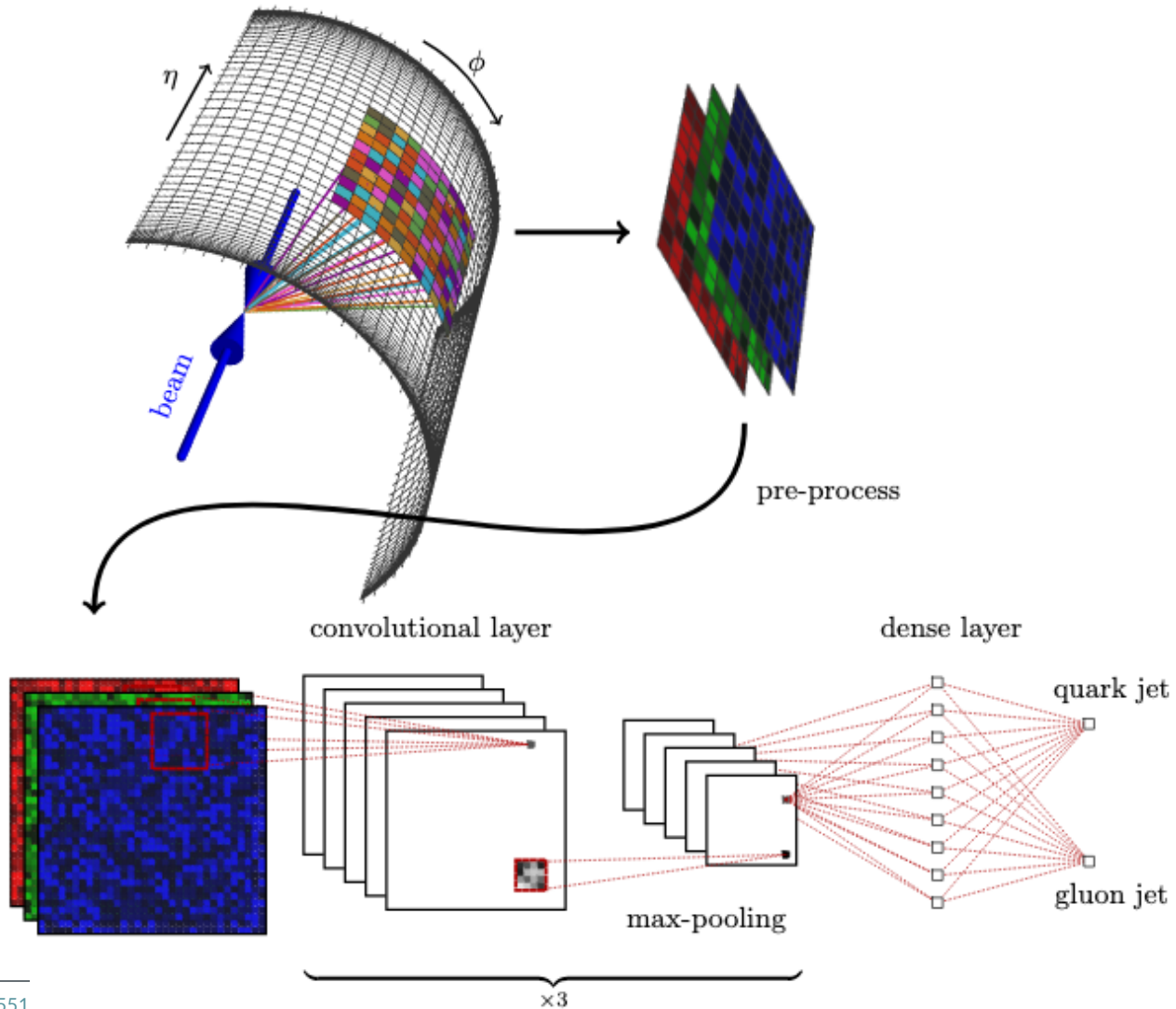
- The first layer retains most of the information, and starts **detecting edges** (through lighting perhaps?)
- The second layer makes this **edge detection** more explicit.
- The third layer has identified the **relative position** of the eyes and the nose
- After that, the CNN starts encoding features deeper and deeper, in a **low-level representation** that will be useful for classification!

An example: convolving cats

We can picture the filters themselves, by looking at what image they **respond most maximally to** (e.g. gradient ascent from a blank picture):

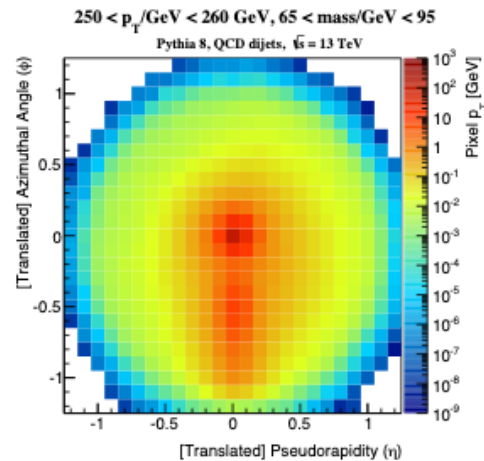
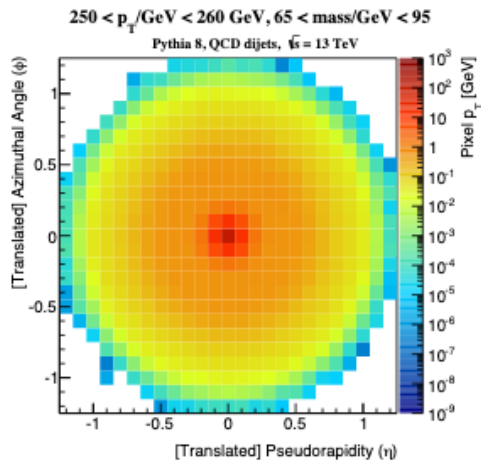
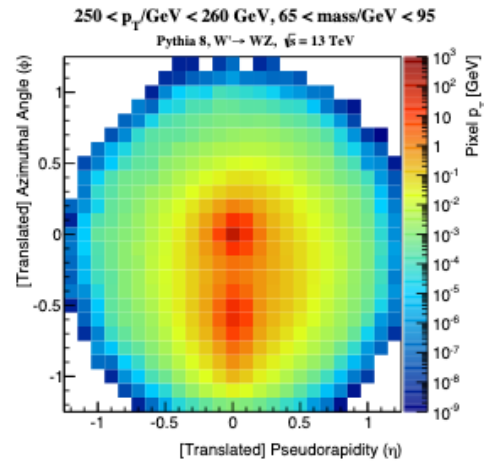
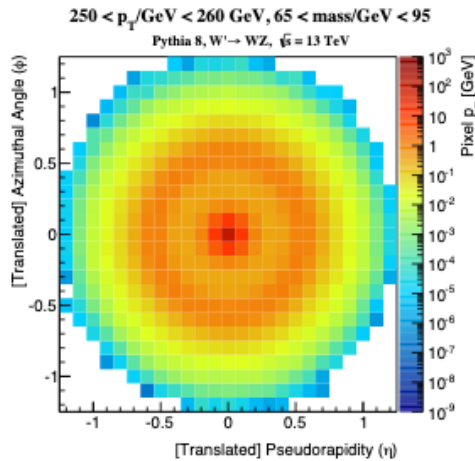


(back to) Jet images



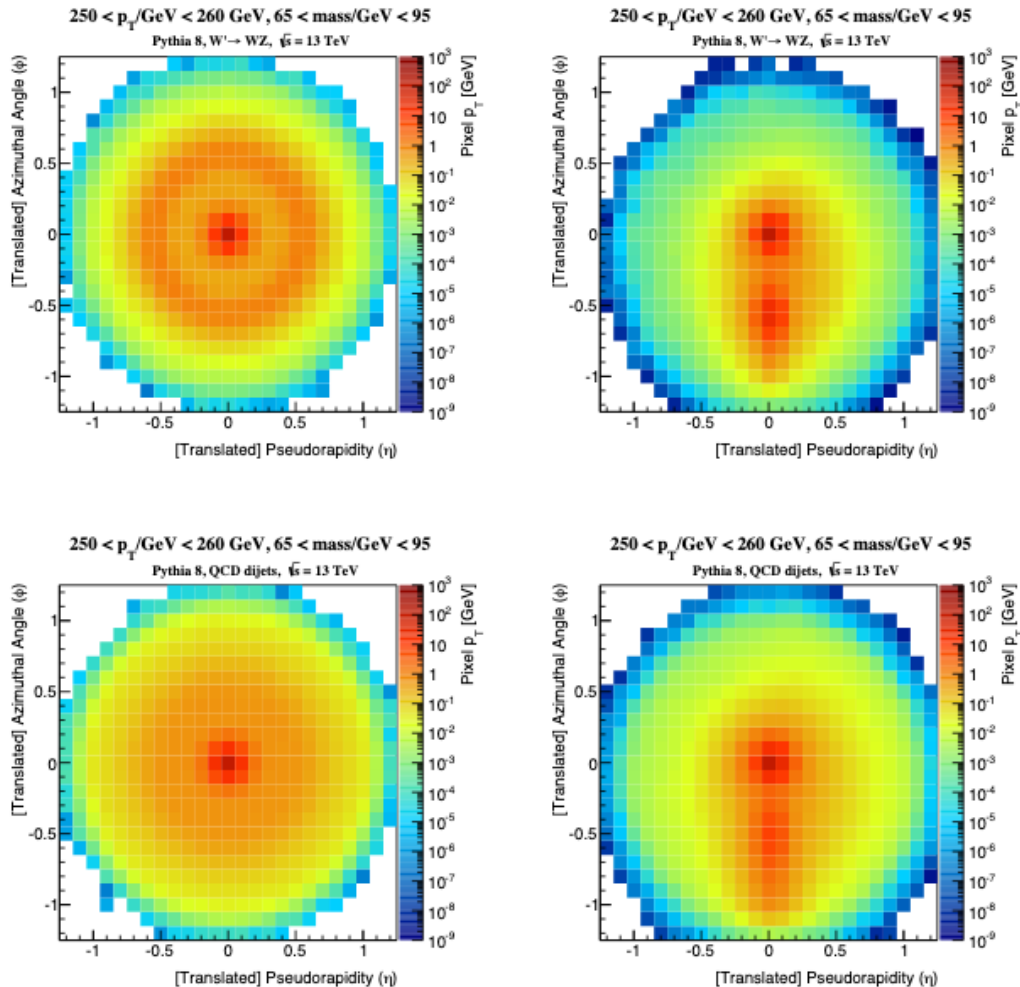
Average quarks and gluons

Pre-process the data: center, rotate and normalise → alternative to generating new data!

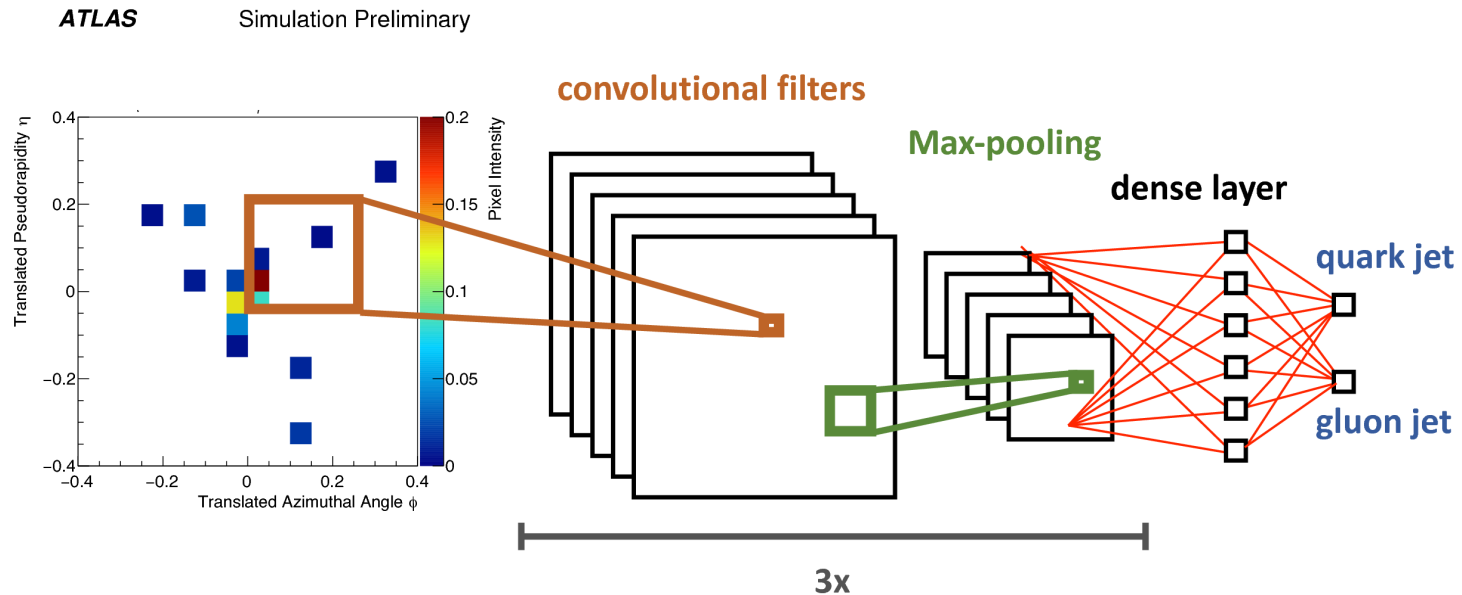


Average quarks and gluons

Can already observe some physics: **colour flow** and **octet radiation**! (more separation in W decays, more diffuse radiation in gluons)

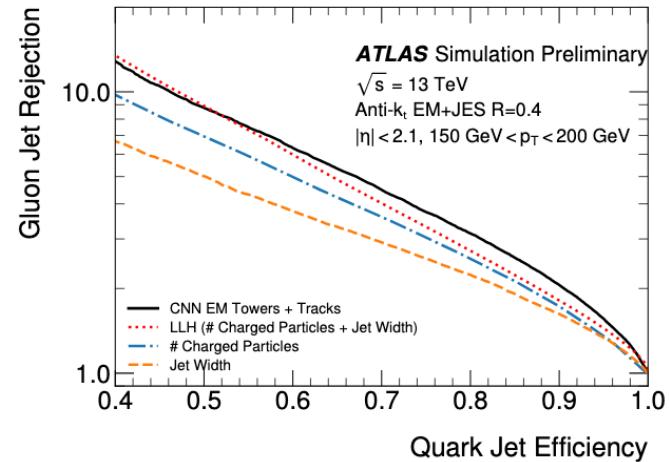
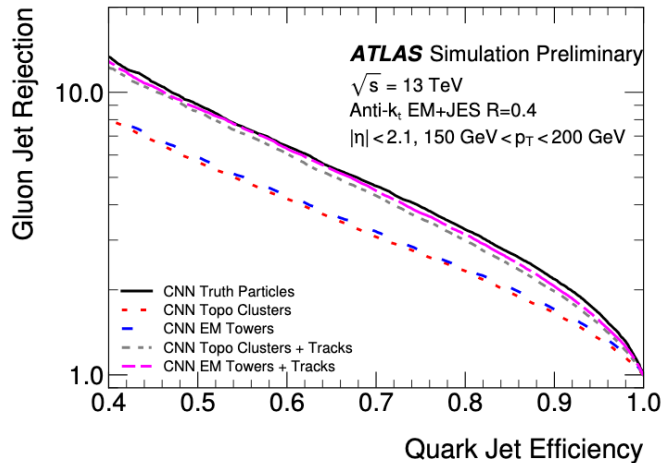


A CNN quark/gluon tagger



- Input is a **16x16 pixel image** of different types of constituents: truth particles, charged tracks, calorimeter clusters/towers
- The CNN learns **non-linear representations** of the image with the goal of **discriminating** between quark and gluon jet images

Performance



- Using calo+track information, the CNN **performs maximally** (as well as with truth information).
- And it also beats the log-likelihood model using a combination of substructure observables!
- However there are still some Monte Carlo generator biases...

Back to physics: **is there something beyond the Higgs?**

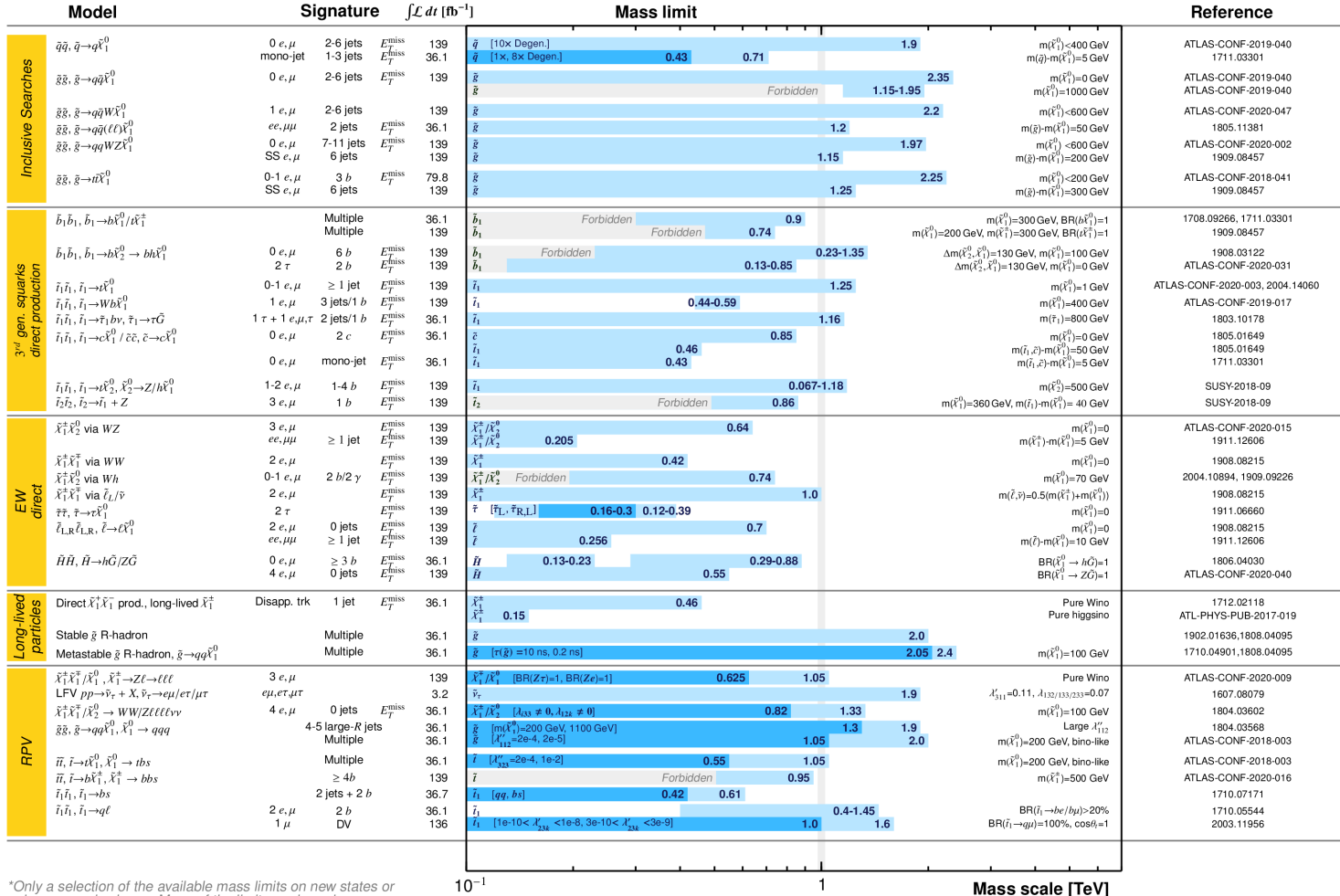
Supersymmetry?

ATLAS SUSY Searches* - 95% CL Lower Limits

July 2020

ATLAS Preliminary

$\sqrt{s} = 13$ TeV



*Only a selection of the available mass limits on new states or phenomena is shown. Many of the limits are based on simplified models, c.f. refs. for the assumptions made.

10⁻¹ 1 Mass scale [TeV]

Exotic particles?

ATLAS Exotics Searches* - 95% CL Upper Exclusion Limits

Status: May 2020

ATLAS Preliminary

$$\int \mathcal{L} dt = (3.2 - 139) \text{ fb}^{-1}$$

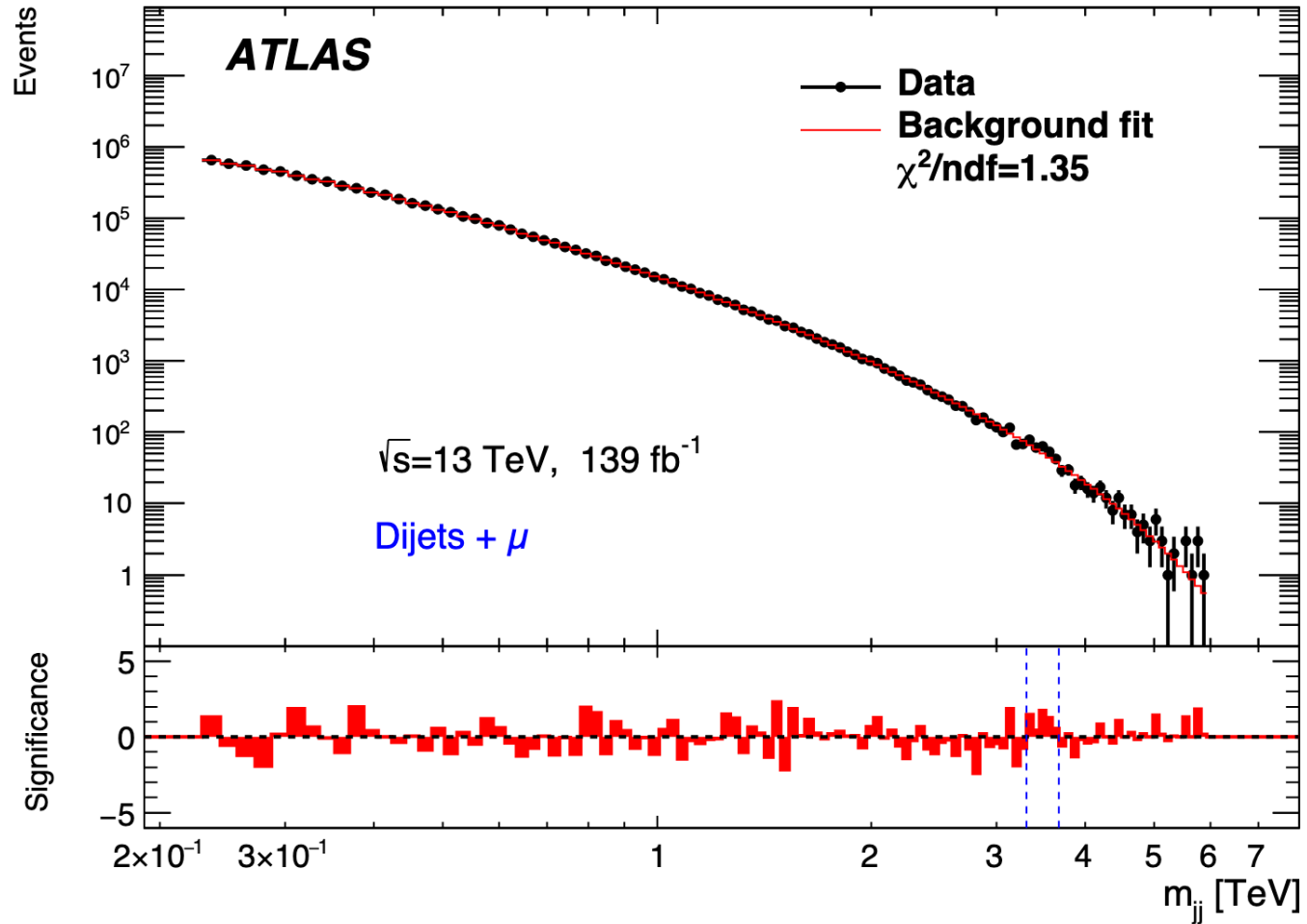
$$\sqrt{s} = 8, 13 \text{ TeV}$$

Model	ℓ, γ	Jets [†]	E_T^{miss}	$\int \mathcal{L} dt [\text{fb}^{-1}]$	Limit	Reference	
Extra dimensions	ADD $G_{KK} + g/q$	$0 e, \mu$	$1-4 j$	Yes	36.1	M_D 7.7 TeV	$n=2$
	ADD non-resonant $\gamma\gamma$	2γ	-	-	36.7	M_S 8.6 TeV	$n=3$ HLZ NLO
	ADD QBH	-	$2 j$	-	37.0	M_{th} 8.9 TeV	$n=6$
	ADD BH high Σp_T	$\geq 1 e, \mu$	$\geq 2 j$	-	3.2	M_{th} 8.2 TeV	$n=6, M_D = 3 \text{ TeV, rot BH}$
	ADD BH multijet	-	$\geq 3 j$	-	3.6	M_{th} 9.55 TeV	$n=6, M_D = 3 \text{ TeV, rot BH}$
	RS1 $G_{KK} \rightarrow \gamma\gamma$	2γ	-	-	36.7	G_{KK} mass 4.1 TeV	$k/\overline{M}_{pl} = 0.1$
	Bulk RS $G_{KK} \rightarrow WW/ZZ$	multi-channel	-	-	36.1	G_{KK} mass 2.3 TeV	$k/\overline{M}_{pl} = 1.0$
	Bulk RS $G_{KK} \rightarrow WV \rightarrow \ell\nu qq$	$1 e, \mu$	$2j/1 J$	Yes	139	G_{KK} mass 2.0 TeV	$k/\overline{M}_{pl} = 1.0$
	Bulk RS $G_{KK} \rightarrow tt$	$1 e, \mu$	$\geq 1 b, \geq 1J/2$	Yes	36.1	g_{KK} mass 3.8 TeV	$\Gamma/m = 15\%$
	ZUED / RPP	$1 e, \mu$	$\geq 2 b, \geq 3 j$	Yes	36.1	KK mass 1.8 TeV	Tier (1,1), $\mathcal{B}(A^{(1,1)} \rightarrow tt) = 1$
Gauge bosons	SSM $Z' \rightarrow \ell\ell$	$2 e, \mu$	-	-	139	Z' mass 5.1 TeV	
	SSM $Z' \rightarrow \tau\tau$	2τ	-	-	36.1	Z' mass 2.42 TeV	
	Leptophobic $Z' \rightarrow bb$	-	$2 b$	-	36.1	Z' mass 2.1 TeV	
	Leptophobic $Z' \rightarrow tt$	$0 e, \mu$	$\geq 1 b, \geq 2 J$	Yes	139	Z' mass 4.1 TeV	$\Gamma/m = 1.2\%$
	SSM $W' \rightarrow \ell\nu$	$1 e, \mu$	-	Yes	139	W' mass 6.0 TeV	
	SSM $W' \rightarrow \tau\nu$	1τ	-	Yes	36.1	W' mass 3.7 TeV	
	HVT $W' \rightarrow WZ \rightarrow \ell\nu qq$ model B	$1 e, \mu$	$2j/1 J$	Yes	139	W' mass 4.3 TeV	$g_V = 3$
	HVT $V' \rightarrow WV \rightarrow qq qq$ model B	$0 e, \mu$	$2 J$	Yes	139	V' mass 3.8 TeV	$g_V = 3$
	HVT $V' \rightarrow WH/ZH$ model B	multi-channel	-	-	36.1	V' mass 2.93 TeV	$g_V = 3$
	HVT $W' \rightarrow WH$ model B	$0 e, \mu$	$\geq 1 b, \geq 2 J$	Yes	139	W' mass 3.2 TeV	$g_V = 3$
LRSM $W_R \rightarrow tb$	multi-channel	-	-	36.1	W_R mass 3.25 TeV		
LRSM $W_R \rightarrow \mu N_R$	2μ	$1 J$	-	80	W_R mass 5.0 TeV	$m(N_R) = 0.5 \text{ TeV, } g_L = g_R$	
CI	CI $qqqq$	-	$2 j$	-	37.0	Λ 21.8 TeV	η_{LL}
	CI $\ell\ell qq$	$2 e, \mu$	-	-	139	Λ 35.8 TeV	η_{LL}
	CI $tttt$	$\geq 1 e, \mu$	$\geq 1 b, \geq 1 j$	Yes	36.1	Λ 2.57 TeV	$ C_{4\ell} = 4\pi$
DM	Axial-vector mediator (Dirac DM)	$0 e, \mu$	$1-4 j$	Yes	36.1	m_{med} 1.55 TeV	$g_a = 0.25, g_s = 1.0, m(\chi) = 1 \text{ GeV}$
	Colored scalar mediator (Dirac DM)	$0 e, \mu$	$1-4 j$	Yes	36.1	m_{med} 1.67 TeV	$g = 1.0, m(\chi) = 1 \text{ GeV}$
	$VV_{\chi\chi}$ EFT (Dirac DM)	$0 e, \mu$	$1 J, \leq 1 j$	Yes	3.2	M_{χ} 700 GeV	$m(\chi) < 150 \text{ GeV}$
	Scalar reson. $\phi \rightarrow \ell\chi$ (Dirac DM)	$0-1 e, \mu$	$1 b, 0-1 J$	Yes	36.1	m_{ϕ} 3.4 TeV	$y = 0.4, \lambda = 0.2, m(\chi) = 10 \text{ GeV}$
LQ	Scalar LQ 1 st gen	$1, 2 e$	$\geq 2 j$	Yes	36.1	LQ mass 1.4 TeV	$\beta = 1$
	Scalar LQ 2 nd gen	$1, 2 \mu$	$\geq 2 j$	Yes	36.1	LQ mass 1.56 TeV	$\beta = 1$
	Scalar LQ 3 rd gen	2τ	$2 b$	-	36.1	LQ_2^+ mass 1.03 TeV	$\mathcal{B}(LQ_2^+ \rightarrow b\tau) = 1$
	Scalar LQ 3 rd gen	$0-1 e, \mu$	$2 b$	Yes	36.1	LQ_3^+ mass 970 GeV	$\mathcal{B}(LQ_3^+ \rightarrow t\tau) = 0$
Heavy quarks	VLQ $TT \rightarrow Ht/Zt/Wb + X$	multi-channel	-	-	36.1	T mass 1.37 TeV	SU(2) doublet
	VLQ $BB \rightarrow Wt/Zb + X$	multi-channel	-	-	36.1	B mass 1.34 TeV	SU(2) doublet
	VLQ $T_{5/3} T_{5/3} T_{5/3} \rightarrow Wt + X$	$2(SS) \geq 3 e, \mu$	$\geq 1 b, \geq 1 j$	Yes	36.1	$T_{5/3}$ mass 1.64 TeV	$\mathcal{B}(T_{5/3} \rightarrow Wt) = 1, c(T_{5/3} Wt) = 1$
	VLQ $Y \rightarrow Wb + X$	$1 e, \mu$	$\geq 1 b, \geq 1 j$	Yes	36.1	Y mass 1.85 TeV	$\mathcal{B}(Y \rightarrow Wb) = 1, c_R(Wb) = 1$
	VLQ $B \rightarrow Hb + X$	$0 e, \mu, 2 \gamma$	$\geq 1 b, \geq 1 j$	Yes	79.8	B mass 1.21 TeV	$\kappa_B = 0.5$
	VLQ $QQ \rightarrow WqWq$	$1 e, \mu$	$\geq 4 j$	Yes	20.3	Q mass 690 GeV	
Excited fermions	Excited quark $q^* \rightarrow qg$	-	$2 j$	-	139	q^* mass 6.7 TeV	only u^* and $d^*, \Lambda = m(q^*)$
	Excited quark $q^* \rightarrow q\gamma$	1γ	$1 j$	-	36.7	q^* mass 5.3 TeV	only u^* and $d^*, \Lambda = m(q^*)$
	Excited quark $b^* \rightarrow bg$	-	$1 b, 1 j$	-	36.1	b^* mass 2.6 TeV	
	Excited lepton ℓ^*	$3 e, \mu$	-	-	20.3	ℓ^* mass 3.0 TeV	$\Lambda = 3.0 \text{ TeV}$
	Excited lepton ν^*	$3 e, \mu, \tau$	-	-	20.3	ν^* mass 1.6 TeV	$\Lambda = 1.6 \text{ TeV}$
Other	Type III Seesaw	$1 e, \mu$	$\geq 2 j$	Yes	79.8	N^0 mass 560 GeV	$m(W_R) = 4.1 \text{ TeV, } g_L = g_R$
	LRSM Majorana ν	2μ	$2 j$	-	36.1	N_R mass 3.2 TeV	1809.1105
	Higgs triplet $H^{\pm\pm} \rightarrow \ell\ell$	$2, 3, 4 e, \mu$ (SS)	-	-	36.1	$H^{\pm\pm}$ mass 870 GeV	DY production
	Higgs triplet $H^{\pm\pm} \rightarrow \ell\tau$	$3 e, \mu, \tau$	-	-	20.3	$H^{\pm\pm}$ mass 400 GeV	DY production, $\mathcal{B}(H^{\pm\pm} \rightarrow \ell\tau) = 1$
	Multi-charged particles	-	-	-	36.1	multi-charged particle mass 1.22 TeV	DY production, $ q = 5e$
	Magnetic monopoles	-	-	-	34.4	monopole mass 2.37 TeV	DY production, $ g = 1g_D, \text{ spin } 1/2$

*Only a selection of the available mass limits on new states or phenomena is shown.

† Small-radius (large-radius) jets are denoted by the letter j (J).

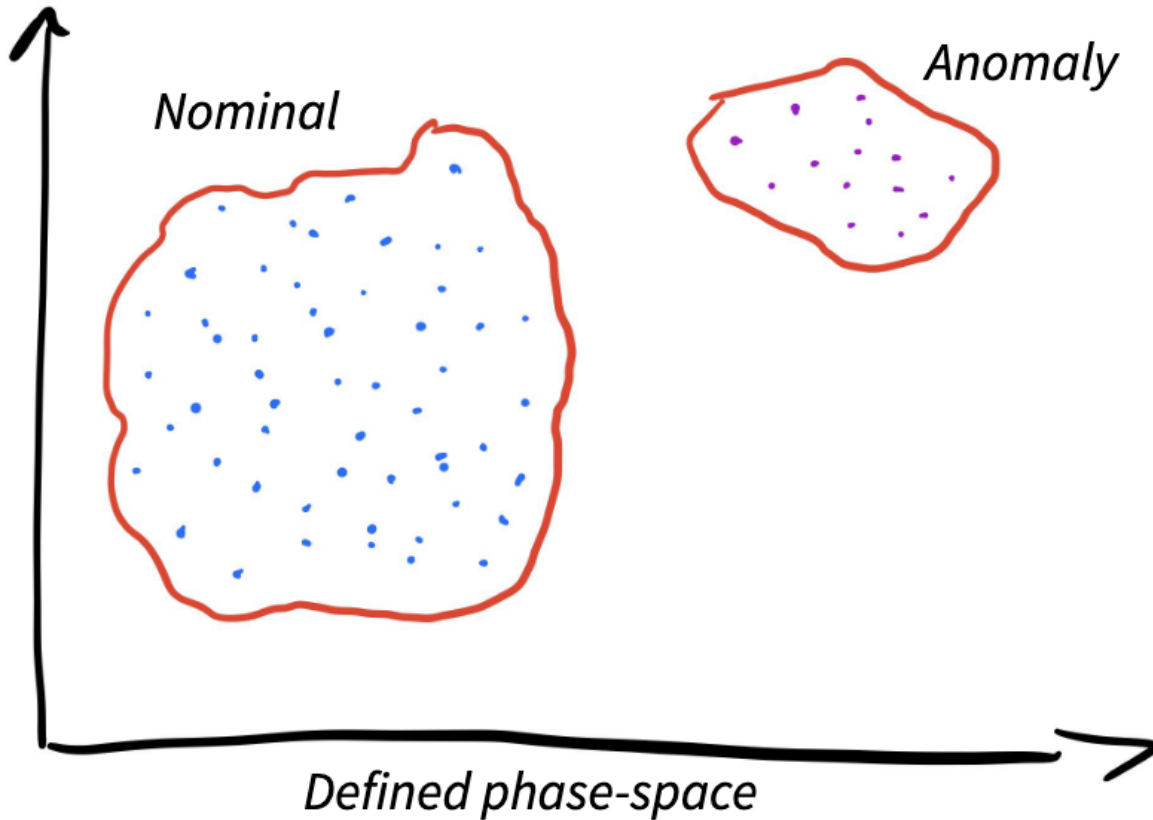
Resonances in bump hunts?



What is **anomaly detection**?

The basic idea

"Finding **patterns** that do not conform to **expected behaviour**." (C. Nellist, 2020)



The basic idea

"Finding **patterns** that do not conform to **expected behaviour**." (C. Nellist, 2020)

- Most commonly used in time series: e.g. fraud detection
- Assume ignorance of the type of anomaly / "catch-all" strategy: unsupervised learning!
- In particle physics: **model-independent BSM searches**

The basic idea

"Finding **patterns** that do not conform to **expected behaviour**." (C. Nellist, 2020)

- Most commonly used in time series: e.g. fraud detection
- Assume ignorance of the type of anomaly / "catch-all" strategy: unsupervised learning!
- In particle physics: **model-independent BSM searches**

Train your algorithm to form an **internal representation** of your training data (actual data or MC), apply to unseen events and use some **quality criterion**: how "new" is this unseen data?

Note: as in precision measurements, an anomaly (or excess/deviation) *doesn't necessarily translate* to BSM physics! It might be BSM, it might be poor detector performance, high-order QCD correction, extreme region of phase-space etc.

Bringing in **neural networks**: **CWoLa** hunting

Classifying WithOut Labels

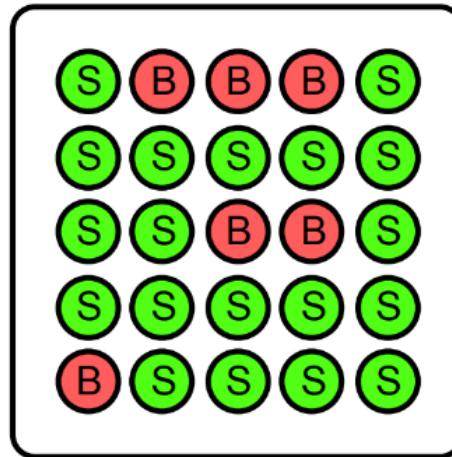
Beautiful (proven) theorem:

A classifier trained to optimally discriminate mixed sample 1 from mixed sample 2 is also optimal for discriminating S from B

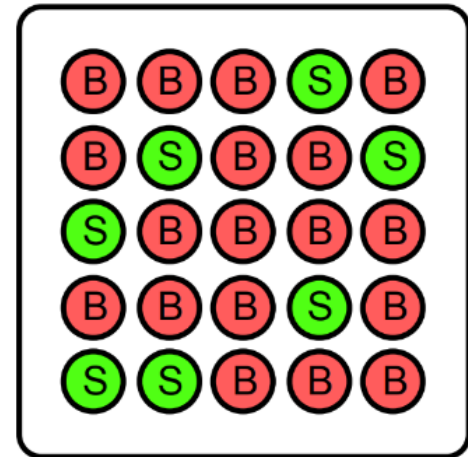


CWoLa

Mixed Sample 1



Mixed Sample 2



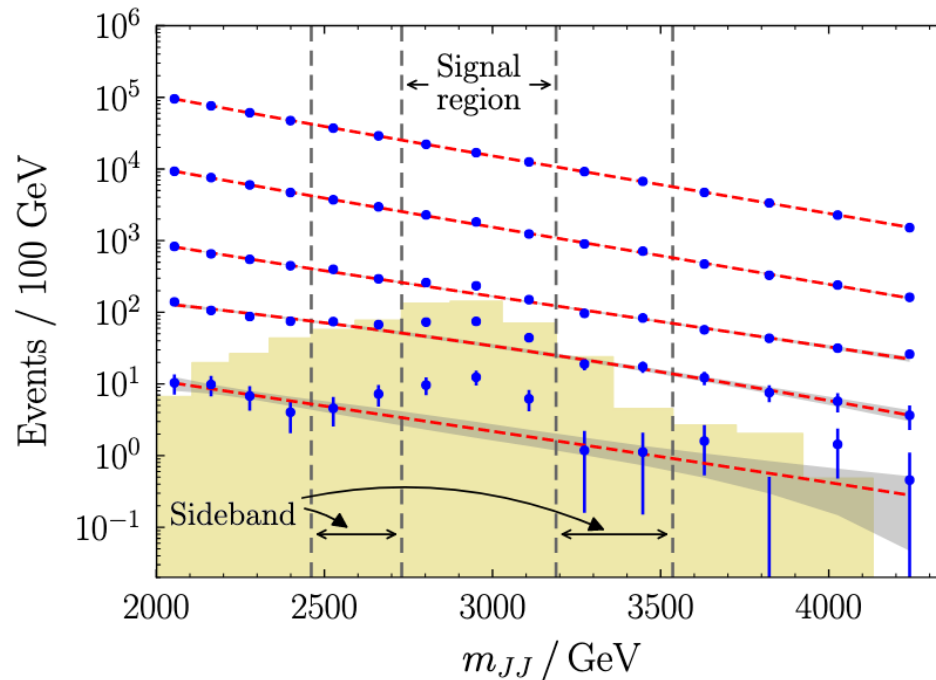
... so long as B and S are drawn from the same distribution in samples 1 & 2, and stats are large enough! (fair)

CWoLa hunting



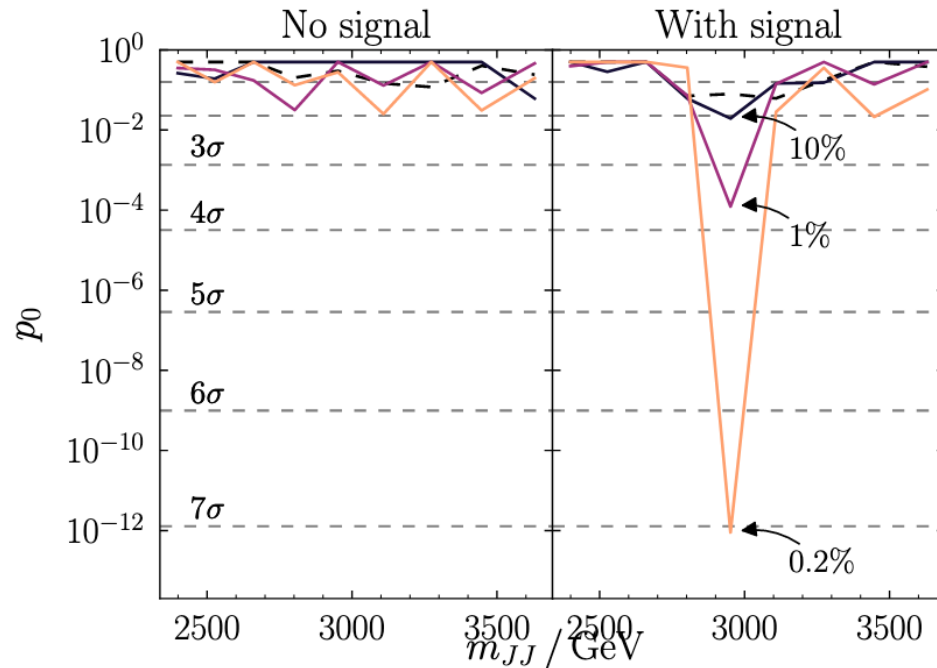
1. Scan range of interest, defining **sidebands** and **signal regions**
2. **Train** networks on background and signal, as defined in 1
3. Compute **p-value** and signal significance in signal region

CWoLa hunting



1. Scan range of interest, defining **sidebands** and **signal regions**
2. **Train** networks on background and signal, as defined in 1
3. Compute **p-value** and signal significance in signal region
4. Keep **rolling!**

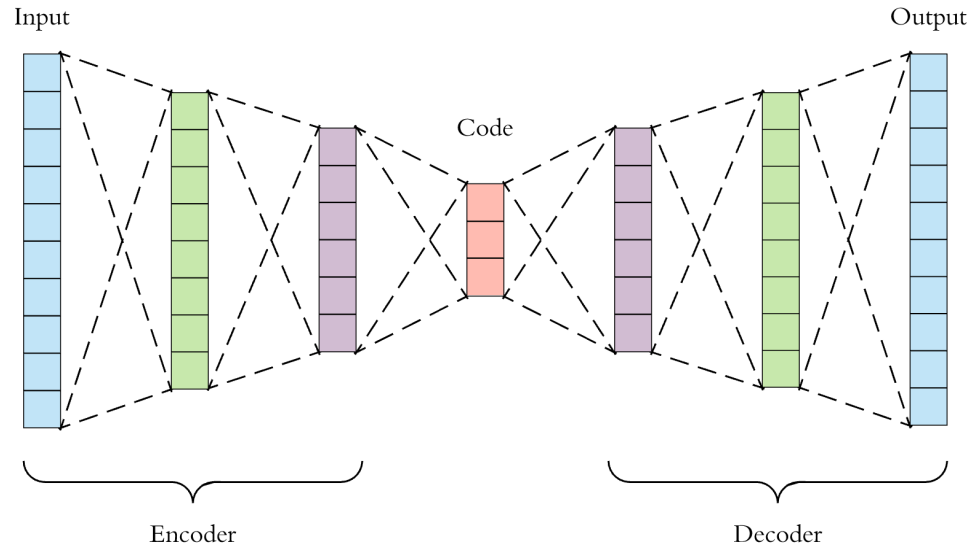
CWoLa hunting



1. Scan range of interest, defining **sidebands** and **signal regions**
2. **Train** networks on background and signal, as defined in 1
3. Compute **p-value** and signal significance in signal region
4. Keep **rolling!**

Compression and reconstruction: **AEs** and **VAEs**

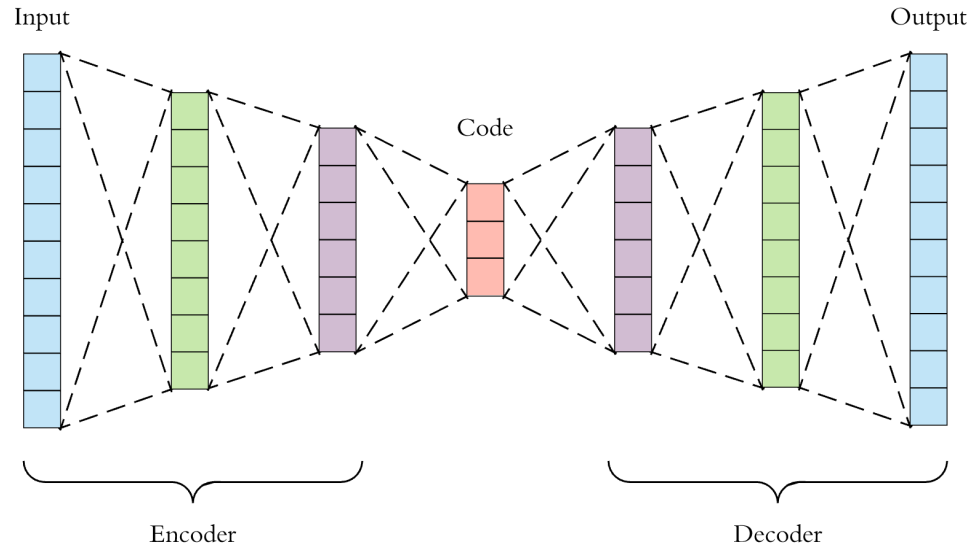
A neural network with a squeeze



An **autoencoder** is a DNN with 3 defining features:

1. the size of the output is the same as the size of the input
2. the loss is measured with respect to the input (and not some target!)
3. there is a **bottleneck**

A neural network with a squeeze

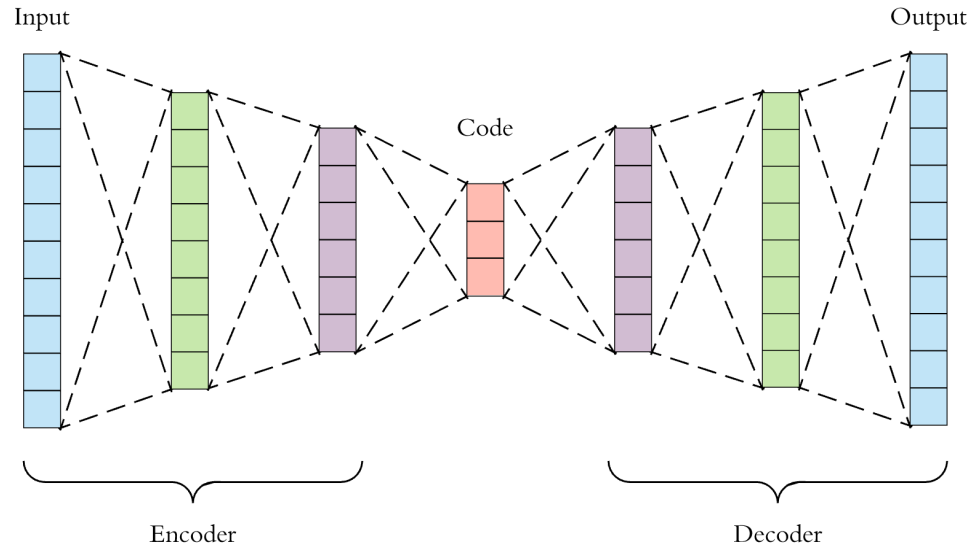


An **autoencoder** is a DNN with 3 defining features:

1. the size of the output is the same as the size of the input
2. the loss is measured with respect to the input (and not some target!)
3. there is a **bottleneck**

These conditions necessarily lead to a **latent space** (an internal representation of the input data), with the means to translate to/from it.

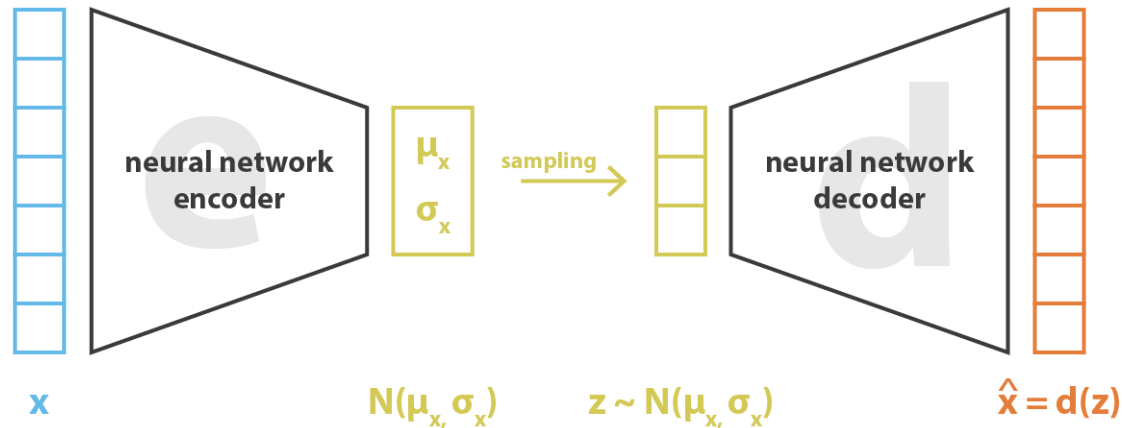
A neural network with a squeeze



An [autoencoder](#) therefore offers 3 main functionalities:

1. it can be used to de-noise data (noise not essential to latent representation!)
2. the latent space might offer new opportunities for discrimination
3. a large reconstruction error signals an anomaly

The re-parameterisation trick

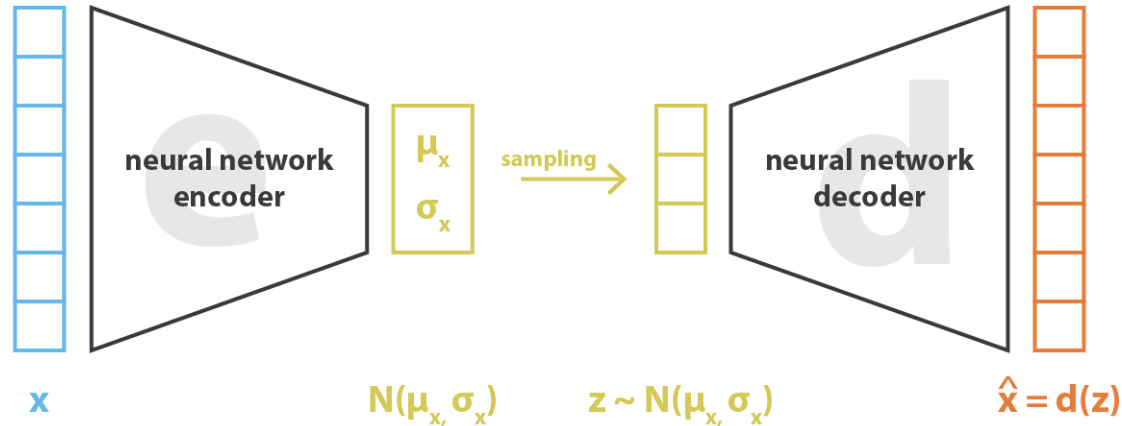


$$\text{loss} = \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = \|x - d(z)\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)]$$

A **variational autoencoder** connects the decoder to the encoder via a **sampling layer**: the Kullback-Leibler divergence (KLd) term in the loss enforces structure in the latent space.

The network will eventually learn the most efficient balance between the **reconstruction loss** and the **sampling loss**.

The re-parameterisation trick

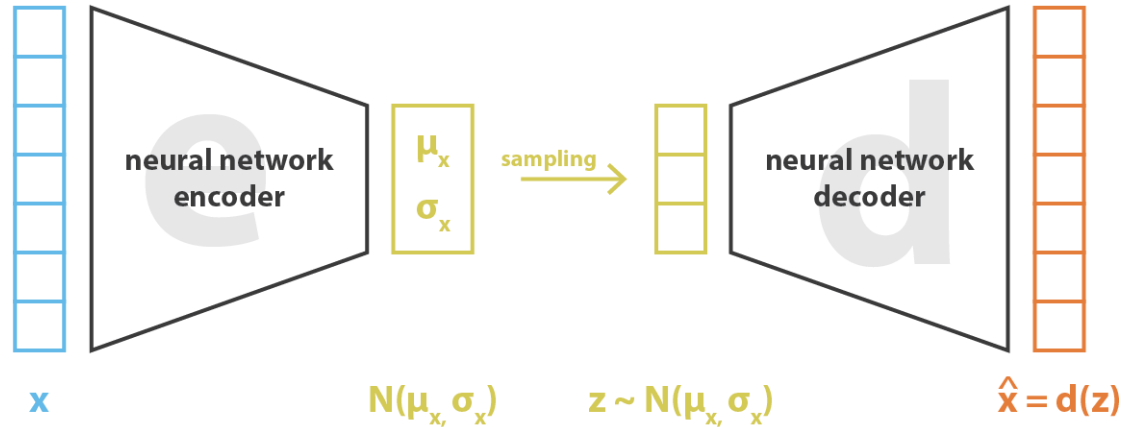


$$\text{loss} = \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = \|x - d(z)\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)]$$

If we consider only the MSE loss, we lose the constraint on $\mathcal{N}(\mu, \sigma) \sim \mathcal{N}(0, 1)$ and the VAE is allowed to "cheat" by clustering events arbitrarily far apart.

If we consider only the KLd loss, we force the structure of the latent space to be $\mathcal{N}(0, 1)$: we have generation without modelling!

The re-parameterisation trick

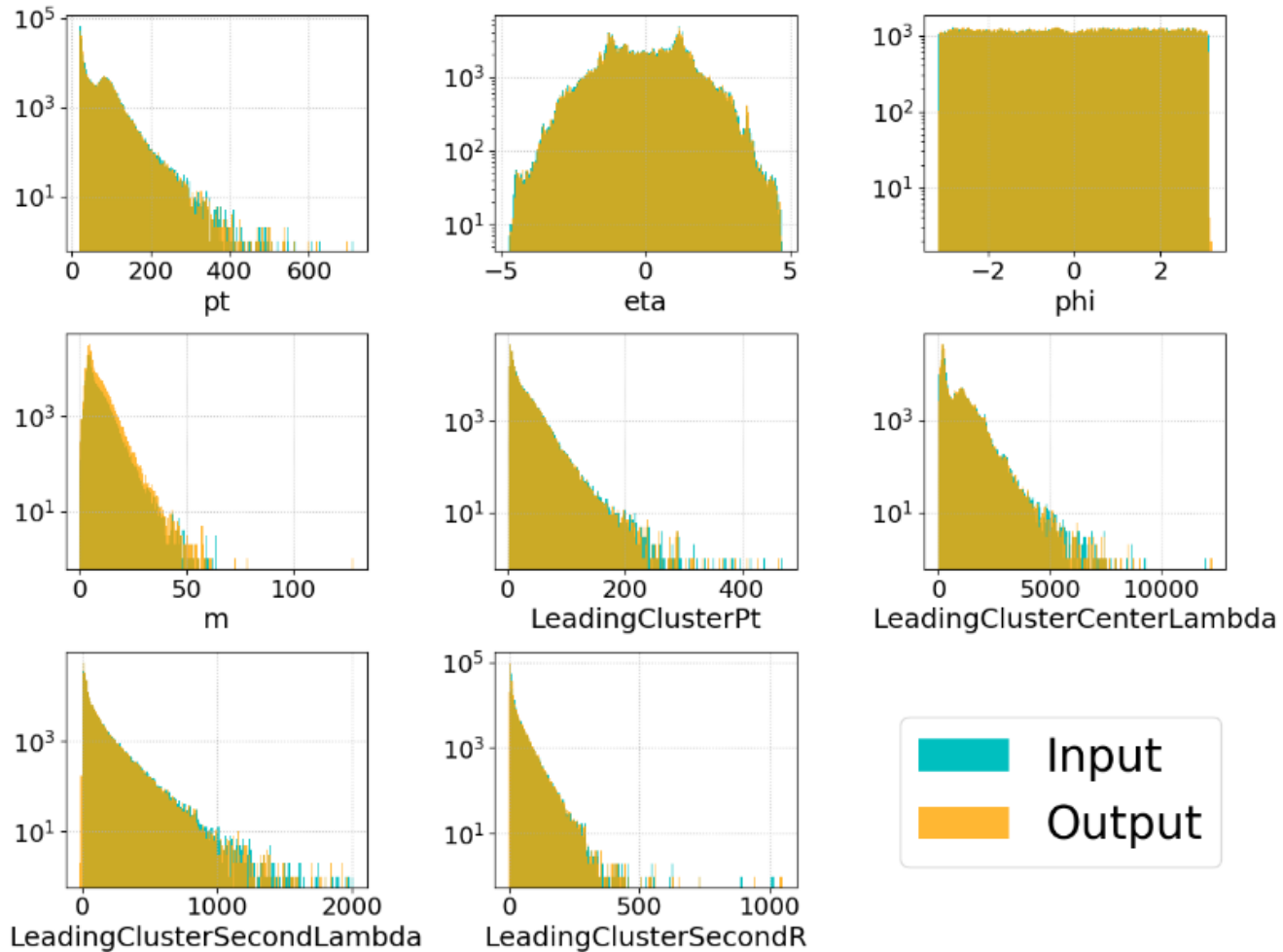


$$\text{loss} = \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = \|x - d(z)\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)]$$

If we've found a balance, we have a **VAE** that:

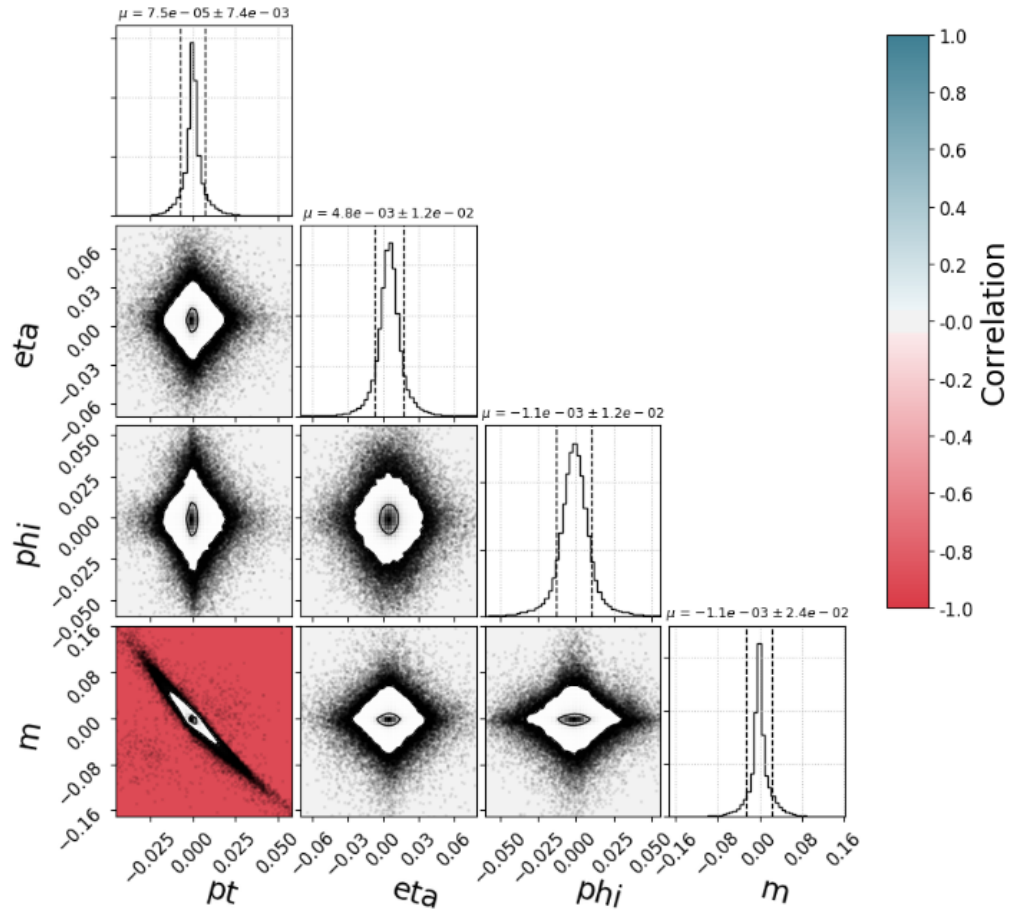
1. has a continuous latent space, allowing **generation of new events**
2. has a structured latent space, allowing **clustering and discrimination**
3. has a large reconstruction error on exotic data, allowing **anomaly detection**

Spot the differences



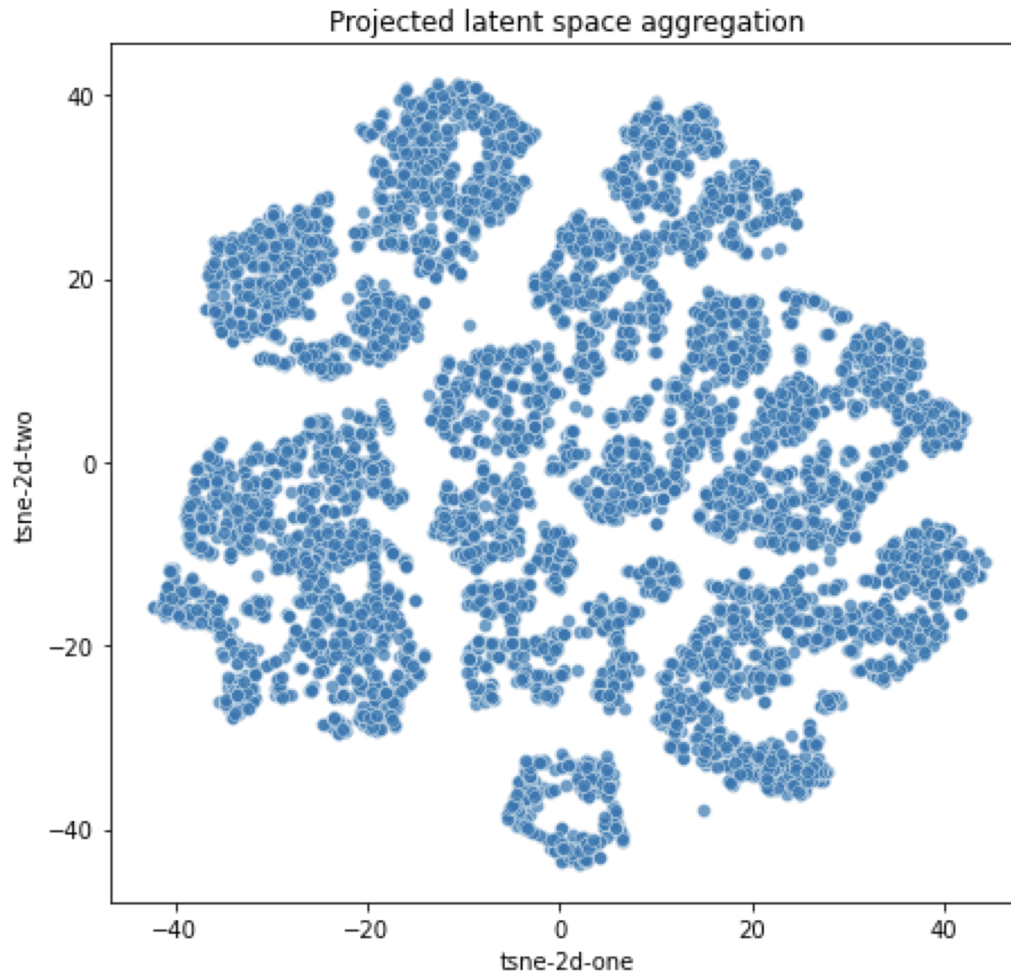
VAE setup: compress 27 input jet variables down to 14 in the latent space.

Understanding the latent space



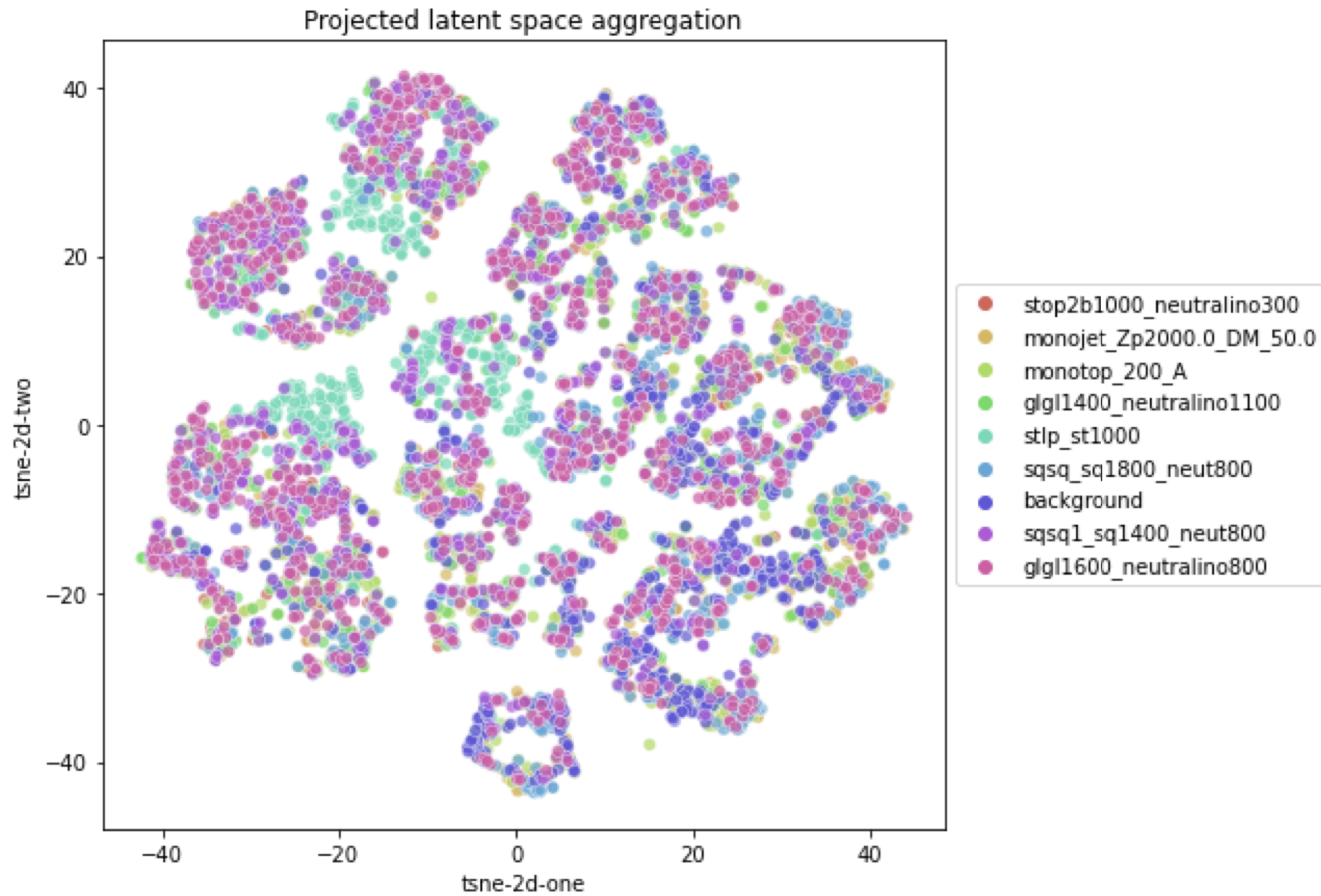
Work ongoing to **make sense of the latent space**: here see **strong correlation** between the residuals on m and p_T .

Understanding the latent space



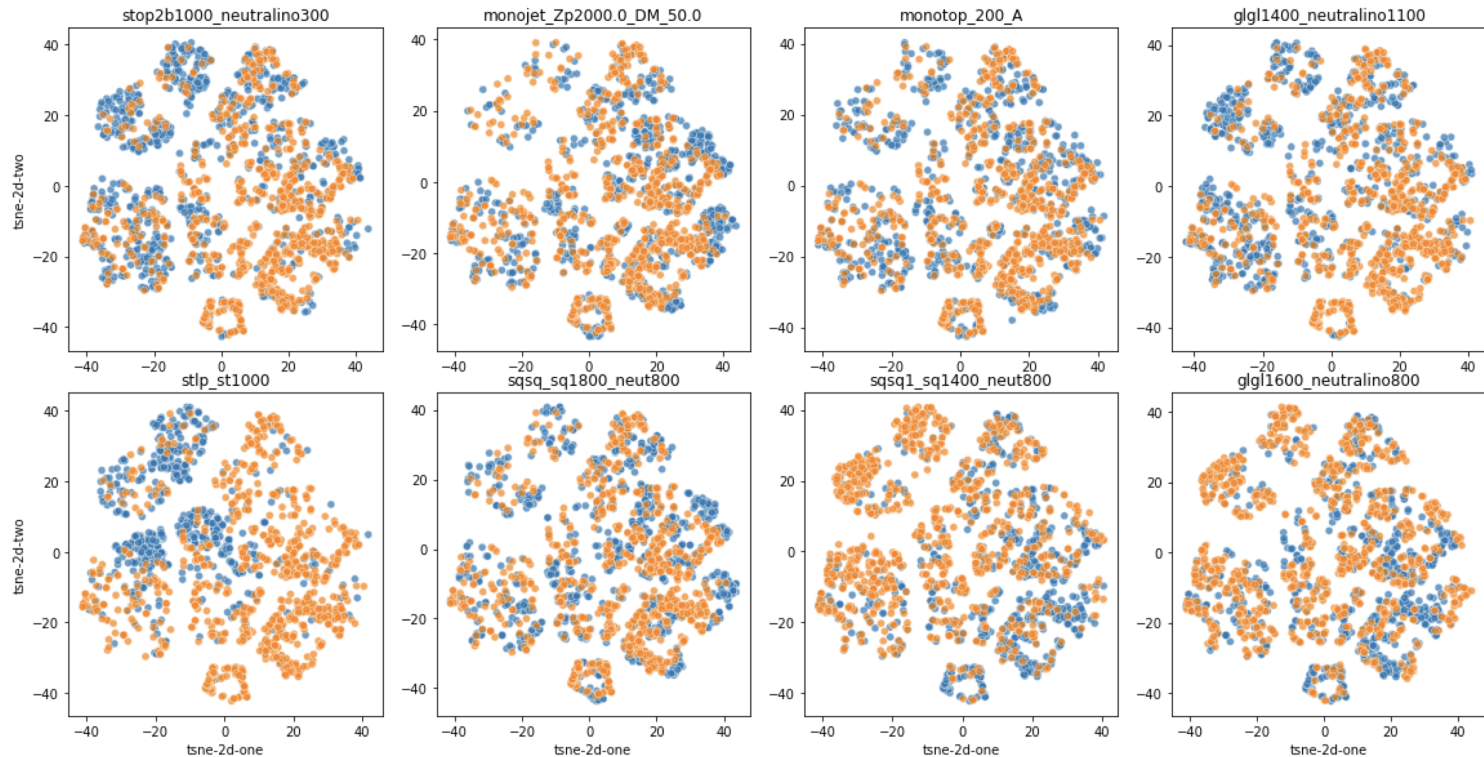
Un-labeled representation of the latent space.

Understanding the latent space



Labeled representation of the latent space. [How meaningful are these clusters?](#)

Understanding the latent space



Background is in orange, signal in blue — there is discrimination potential in the latent space!

Dark Machines

An international group of researchers *(not just LHC experimentalists!) using machine learning to go after Dark Matter in various forms:

- unsupervised collider searches (anomaly detection)
- high-dimensional sampling
- generative models
- strong lensing
- ... [your project here]

Expect the [first community white paper](#) on [event-level anomaly detection](#) very soon! (including many more interesting models and approaches I didn't have time to cover here)

darkmachines.org

Thank you.

Questions?

References

- A. De Simone & T. Jacques, *Guiding New Physics Searches with Unsupervised Learning*, [arXiv:1807.06038](#)
- M. Sugiyama & *al.*, *Least-squares two-sample test*, [Neural Networks 24 \(2011\) 735 – 751](#)
- E. Metodiev & *al.*, *Classification without labels: Learning from mixed samples in high energy physics*, [arXiv:1708.02949](#)
- J. Collins & *al.*, *Anomaly Detection for Resonant New Physics with Machine Learning*, [arXiv:1805.02664](#)
- J. Collins & *al.*, *Extending the Bump Hunt with Machine Learning*, [arXiv:1902.02634](#)
- C. Doersch, *Tutorial on Variational Autoencoders*, [arXiv:1606.05908](#)
- E. Wulff, *Deep Autoencoders for Compression in High Energy Physics*, [Lund Master thesis](#)