



UNIVERSITÉ
DE GENÈVE



Machine Learning @ ATLAS

*Tobias Golling, University of Geneva,
On behalf of the ATLAS Collaboration
Game of Flavours, Dubrovnik, May 3 2019*

Overview

- *Representative* selection of ML@ATLAS
- Classification & related topics
 - Will not have time to discuss regression tasks
- Clustering & data structures
- Generative models

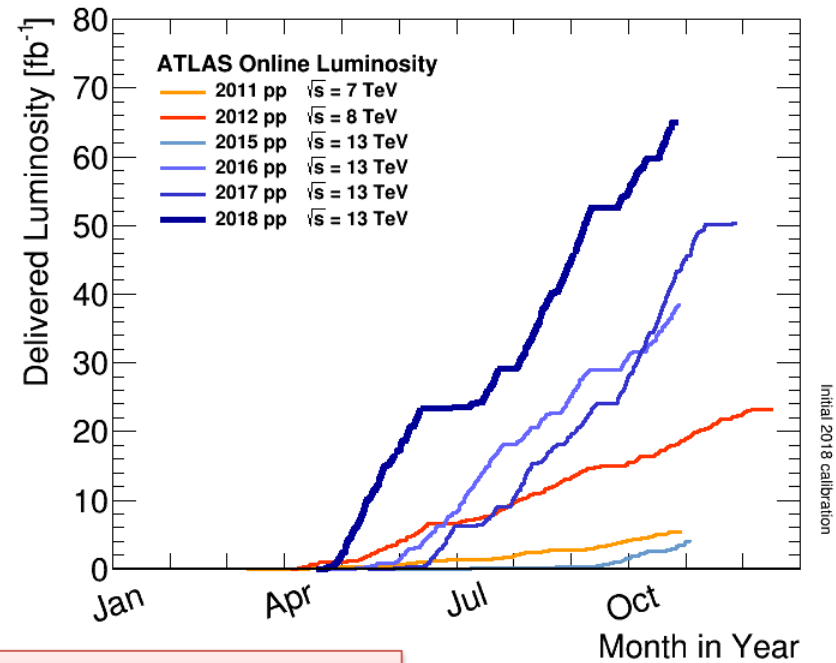
LHC interim evaluation

- Physics beyond the SM is **not** around the corner
- **Slow-growth era** of LHC has started: energy & luminosity
- **How to make rapid progress now?**

ATLAS Exotics Searches* - 95% CL Upper Exclusion Limits
 Status: July 2018
 ATLAS Preliminary
 $\int \mathcal{L} dt = (3.2 - 79.8) \text{ fb}^{-1}$ $\sqrt{s} = 8, 13 \text{ TeV}$

Model	τ, γ	Jets†	$E_{T,miss}^{\min}$	$\int \mathcal{L} dt [\text{fb}^{-1}]$	Limit	Reference	
Extra dimensions	ADD $G_{KK} + g/g$	$0, \mu, \mu$	1-4	Yes	36.1	MA, $\sqrt{s} = 7.7 \text{ TeV}$, $n=2$	1711.02001
	ADD non-resonant $\gamma\gamma$	$2, \gamma$	-	-	36.1	MA, $\sqrt{s} = 8.6 \text{ TeV}$, $n=3$	1707.04147
	ADD GBH	-	2	-	37.9	MA, $\sqrt{s} = 9.0 \text{ TeV}$, $n=6$	1703.06117
	ADD BH high Σ_{pp}	$\geq 1, \mu, \mu$	≥ 2	-	3.2	MA, $\sqrt{s} = 8.2 \text{ TeV}$, $n=6, M_{D,0} = 3 \text{ TeV}$, rot BH	1606.02085
	ADD BH multijet	$\geq 1, \mu, \mu$	≥ 3	-	3.6	MA, $\sqrt{s} = 9.55 \text{ TeV}$, $n=6, M_{D,0} = 3 \text{ TeV}$, rot BH	1512.02986
	RS1 $G_{KK} \rightarrow \gamma\gamma$	$2, \gamma$	-	-	36.1	MA, $\sqrt{s} = 4.1 \text{ TeV}$, $k_{IR} = 0.1$	1707.04147
	Bulk RS $G_{KK} \rightarrow WW/ZZ$	multi-channel	-	-	36.1	MA, $\sqrt{s} = 2.3 \text{ TeV}$, $k/M_{Pl} = 1.0$	CERN-EP-2018-179
	Bulk RS $G_{KK} \rightarrow \tau\tau$	multi-channel	$\geq 1b, \geq 1b, \geq 1b, \geq 1b$	Yes	36.1	MA, $\sqrt{s} = 3.8 \text{ TeV}$, $f/m = 15\%$	1804.10522
	gUED1 RSP	$1, \mu, \mu$	$\geq 2b, \geq 3$	Yes	36.1	MA, $\sqrt{s} = 1.8 \text{ TeV}$, $\text{Top}(1), \text{Top}(2), \text{Top}(3) \rightarrow \tau\tau = 1$	1803.06916
Charge bosons	SSM $Z' \rightarrow \ell\ell$	$2, \mu, \mu$	-	-	36.1	MA, $\sqrt{s} = 2.42 \text{ TeV}$, 4.5 TeV	1707.04044
	SSM $Z' \rightarrow \tau\tau$	$2, \tau$	-	-	36.1	MA, $\sqrt{s} = 2.42 \text{ TeV}$, 4.5 TeV	1709.07242
	Leptoquark $Z' \rightarrow bb$	-	2b	-	36.1	MA, $\sqrt{s} = 2.1 \text{ TeV}$	1803.06299
	Leptoquark $Z' \rightarrow \ell\ell$	$1, \mu, \mu$	$\geq 1b, \geq 1b, \geq 1b, \geq 1b$	Yes	36.1	MA, $\sqrt{s} = 3.0 \text{ TeV}$	1804.09023
	SSM $W' \rightarrow \ell\nu$	$1, \mu, \nu$	-	-	79.8	MA, $\sqrt{s} = 5.6 \text{ TeV}$	ATLAS-CONF-2018-017
	SSM $W' \rightarrow \tau\nu$	$1, \tau, \nu$	-	-	36.1	MA, $\sqrt{s} = 3.7 \text{ TeV}$	1801.06192
	HVT $V' \rightarrow WW \rightarrow \text{qqqq}$ model B	$0, \mu, \mu$	2j	-	79.8	MA, $\sqrt{s} = 4.15 \text{ TeV}$	ATLAS-CONF-2018-016
	HVT $V' \rightarrow WH/ZH$ model B	multi-channel	-	-	36.1	MA, $\sqrt{s} = 2.93 \text{ TeV}$	1713.04518
	LSSM $W'_2 \rightarrow tb$	multi-channel	-	-	36.1	MA, $\sqrt{s} = 3.26 \text{ TeV}$	CERN-EP-2018-142
CI	CI f/fqq	$2, \mu, \mu$	2j	-	37.0	MA, $\sqrt{s} = 21.8 \text{ TeV}$, η_{CI}	1703.06117
	CI f/fqq	$2, \mu, \mu$	-	-	36.1	MA, $\sqrt{s} = 40.0 \text{ TeV}$, η_{CI}	1707.04044
	CI f/fqq	$\geq 1, \mu, \mu$	$\geq 1b, \geq 1j$	Yes	36.1	MA, $\sqrt{s} = 2.57 \text{ TeV}$, $ \alpha_{CI} = 4\%$	CERN-EP-2018-174
DM	Axial-vector mediator (Dirac DM)	$0, \mu, \mu$	1-4j	Yes	36.1	MA, $\sqrt{s} = 1.55 \text{ TeV}$	1711.02001
	Colorlet scalar mediator (Dirac DM)	$0, \mu, \mu$	1-4j	Yes	36.1	MA, $\sqrt{s} = 1.67 \text{ TeV}$	1711.02001
	$W_{1,2}$ EFT (Dirac DM)	$0, \mu, \mu$	1, 1, 5, 1j	Yes	32.2	MA, $\sqrt{s} = 700 \text{ GeV}$	1711.02001
LQ	Scalar LQ 1 st gen	$2, e$	$\geq 2j$	-	32	MA, $\sqrt{s} = 1.1 \text{ TeV}$	1605.06035
	Scalar LQ 2 nd gen	$2, \mu$	$\geq 2j$	-	32	MA, $\sqrt{s} = 1.05 \text{ TeV}$	1605.06035
	Scalar LQ 3 rd gen	$1, \mu, \mu$	$\geq 1b, \geq 3j$	Yes	32.3	MA, $\sqrt{s} = 4.0 \text{ TeV}$	1504.04186
Flavor anomalies	VLO $TT \rightarrow HcZ/Wb + X$	multi-channel	-	-	36.1	MA, $\sqrt{s} = 1.37 \text{ TeV}$	ATLAS-CONF-2018-XXX
	VLO $BB \rightarrow WcZb + X$	multi-channel	-	-	36.1	MA, $\sqrt{s} = 1.34 \text{ TeV}$	ATLAS-CONF-2018-XXX
	VLO $T_{3,1} T_{3,3} / T_{3,3} \rightarrow Wc + X$	$2S(S)/3, 3, \text{ or } 21b, \geq 1j$	-	-	36.1	MA, $\sqrt{s} = 1.64 \text{ TeV}$	ATLAS-CONF-2018-071
	VLO $V \rightarrow Wb + X$	$1, \mu, \mu$	$\geq 1b, \geq 3j$	Yes	3.2	MA, $\sqrt{s} = 1.64 \text{ TeV}$	ATLAS-CONF-2018-072
	VLO $B \rightarrow Hb + X$	$0, \mu, \mu$	$\geq 1b, \geq 3j$	Yes	79.8	MA, $\sqrt{s} = 1.21 \text{ TeV}$	ATLAS-CONF-2018-XXX
	VLO $QQ \rightarrow WqWq$	$1, \mu, \mu$	$\geq 4j$	Yes	20.3	MA, $\sqrt{s} = 690 \text{ GeV}$	1509.04261
Excited fermions	Excited quark $q^* \rightarrow qq$	-	2j	-	37.0	MA, $\sqrt{s} = 6.0 \text{ TeV}$	1703.06117
	Excited quark $q^* \rightarrow q\gamma$	$1, \gamma$	-	-	36.1	MA, $\sqrt{s} = 6.3 \text{ TeV}$	1709.04440
	Excited quark $q^* \rightarrow \ell q$	$3, \mu, \mu$	-	-	36.1	MA, $\sqrt{s} = 2.6 \text{ TeV}$	1805.09299
	Excited lepton $\ell^* \rightarrow \ell\ell$	$3, \mu, \mu$	-	-	20.3	MA, $\sqrt{s} = 1.9 \text{ TeV}$	1411.2921
	Excited lepton $\ell^* \rightarrow \ell\gamma$	$3, \mu, \mu$	-	-	20.3	MA, $\sqrt{s} = 1.7 \text{ TeV}$	1411.2921
Other	Type III Seesaw	$1, \mu, \mu$	$\geq 2j$	Yes	79.8	MA, $\sqrt{s} = 580 \text{ GeV}$	ATLAS-CONF-2018-020
	LRSM Majorana ν	$2, \mu, \mu$	2j	-	20.3	MA, $\sqrt{s} = 2.9 \text{ TeV}$	1506.06200
	Higgs triplet $H^{\pm,0} \rightarrow \ell\ell$	$2, 2, 4, \mu, \mu$ (SS)	-	-	36.1	MA, $\sqrt{s} = 970 \text{ GeV}$	1710.04748
	Higgs triplet $H^{\pm,0} \rightarrow \tau\tau$	$3, \mu, \mu$	-	-	20.3	MA, $\sqrt{s} = 400 \text{ GeV}$	1411.2921
	Monopole (non-res prod)	$1, \mu, \mu$	1b	Yes	20.3	MA, $\sqrt{s} = 607 \text{ GeV}$	1410.0404
	Multi-charged particles	-	-	-	20.3	MA, $\sqrt{s} = 700 \text{ GeV}$	1504.04186
	Magnetic monopoles	-	-	-	7.0	MA, $\sqrt{s} = 1.34 \text{ TeV}$	1509.05059

*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 (L).



Opportunity !
Turning crank → innovation

In a nutshell



Our brain is not **trained** to analyze LHC data

⇒ Train artificial brain to do task for us

Machine Learning (ML)

What is ML?

- Inspired by how the brain works
- Learning from examples
- Condensing information to “knowledge”

How can ML help?

- Low hanging fruit
 - Better
 - Faster
 - Easier / automated
- More profound changes to how we approach physics ?!

Uniqueness of HEP data for ML

- Simulation can produce **highly valuable labeled training data** for supervised learning
- We have a **theory model (SM)**
 - How to inject our domain knowledge into ML
- **Systematic uncertainties**
- HEP not only customer but also *driver* of ML!

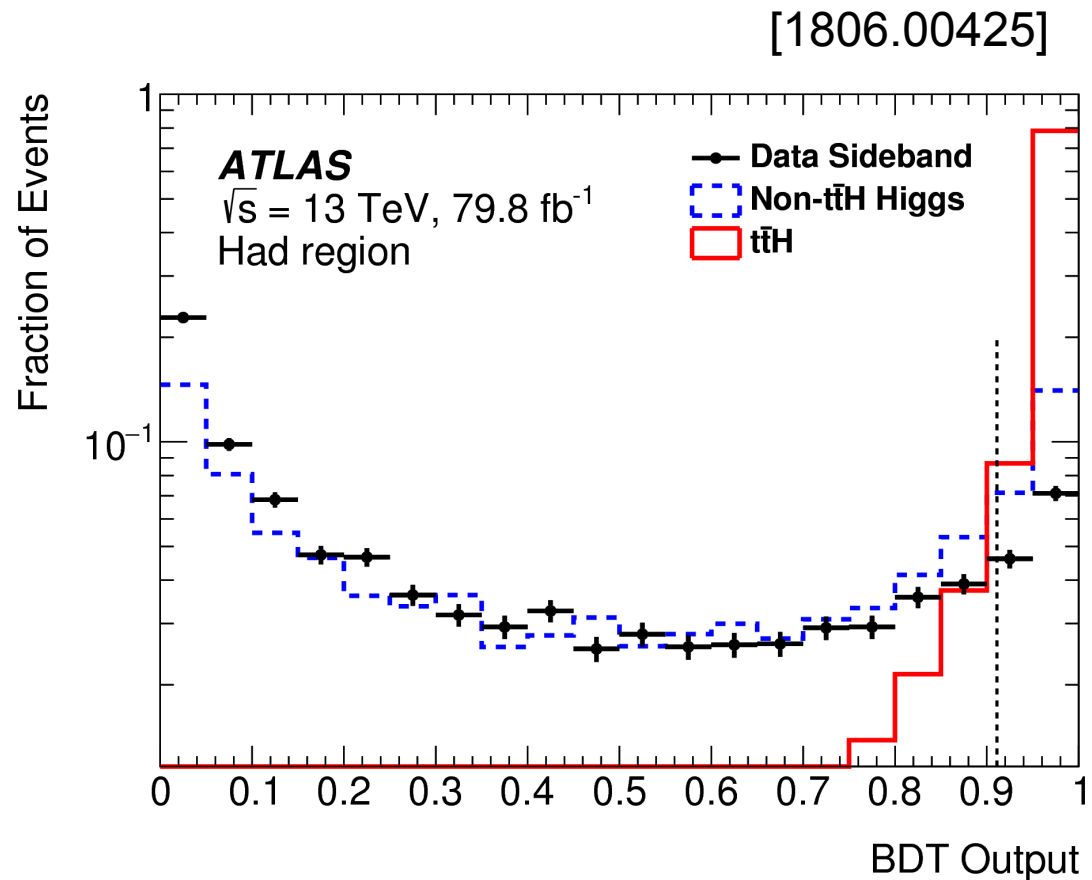
Classification with Machine Learning (better)



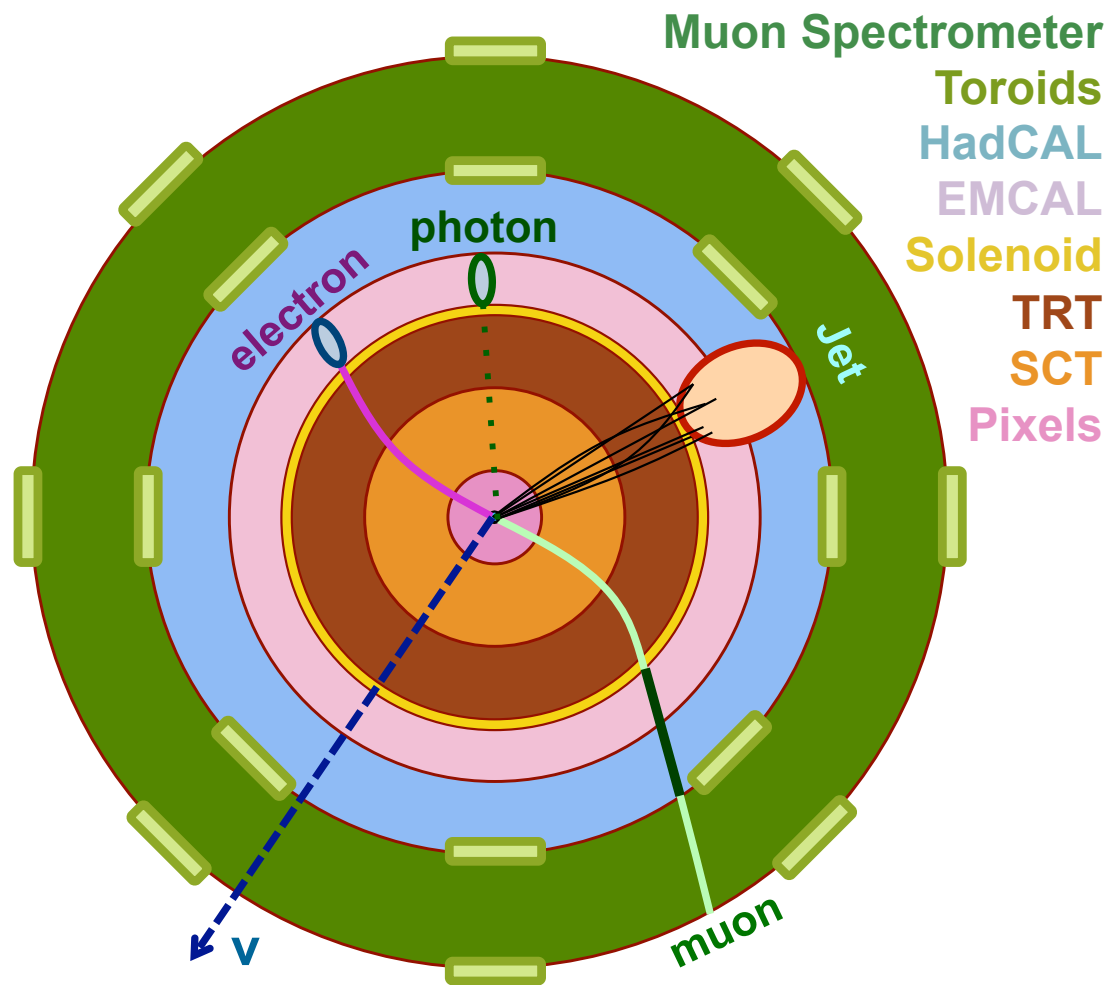
Classification task: edible or not?

Event-level discrimination

- We've used ML for decades
- Recent example: ttH discovery
- Human-engineered features (here 38 input variables)
- Many more examples exist

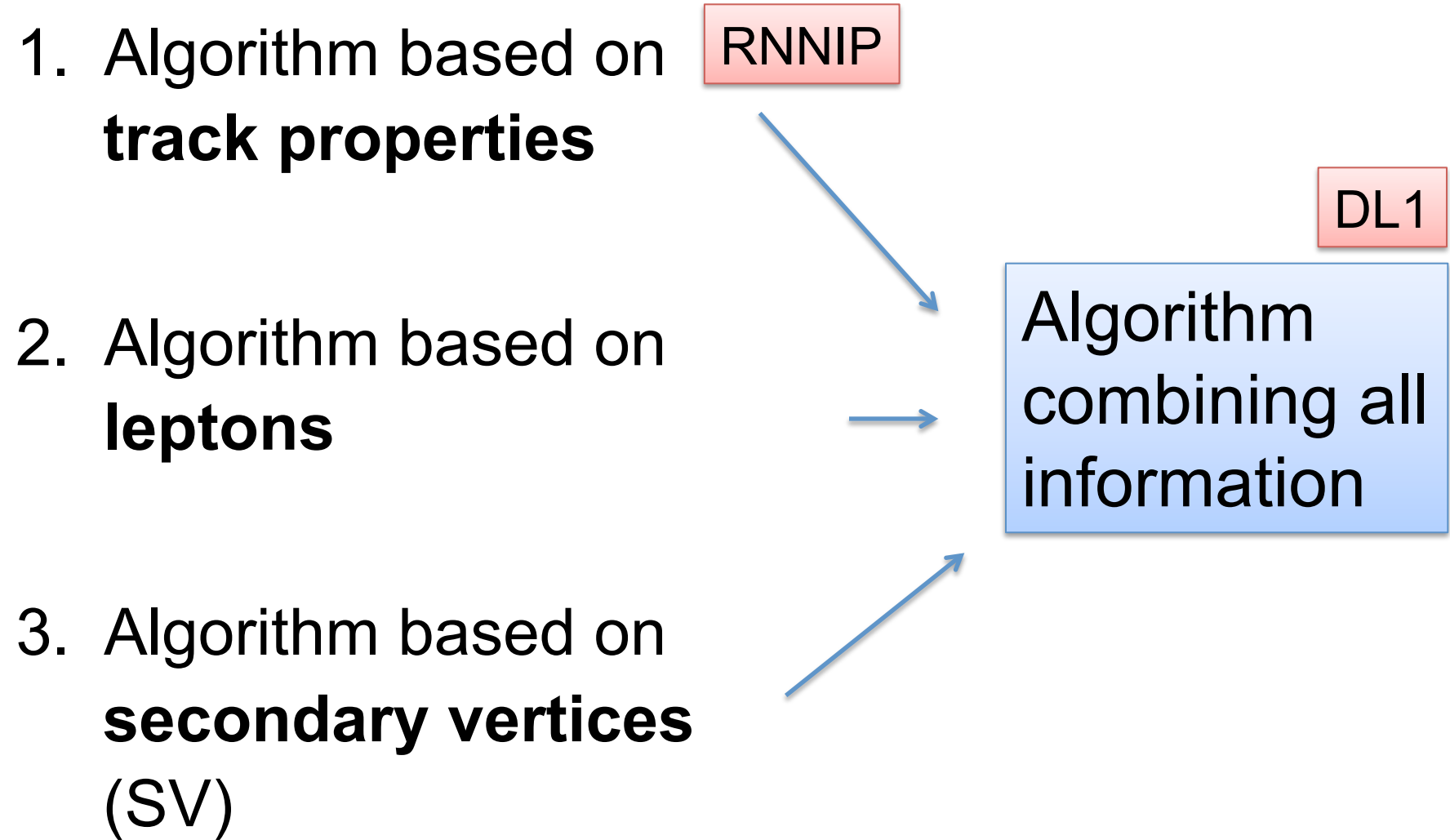


Ideal test ground: physics object classification



- Large statistics
- Excellent modeling
- Good return/effort
- Validate in Control Region

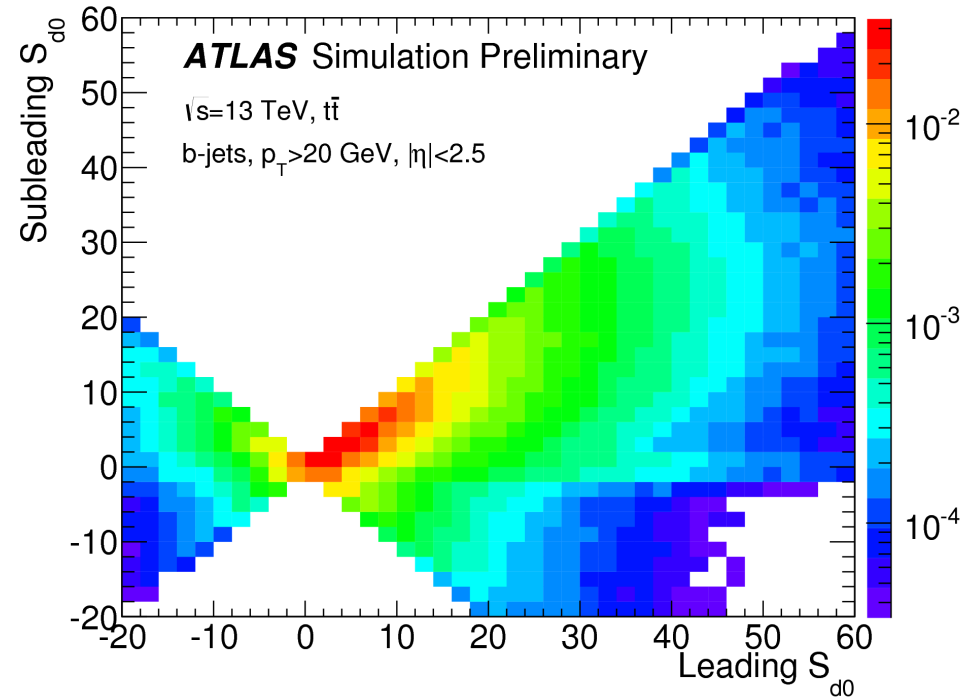
Flavor tagging algorithms



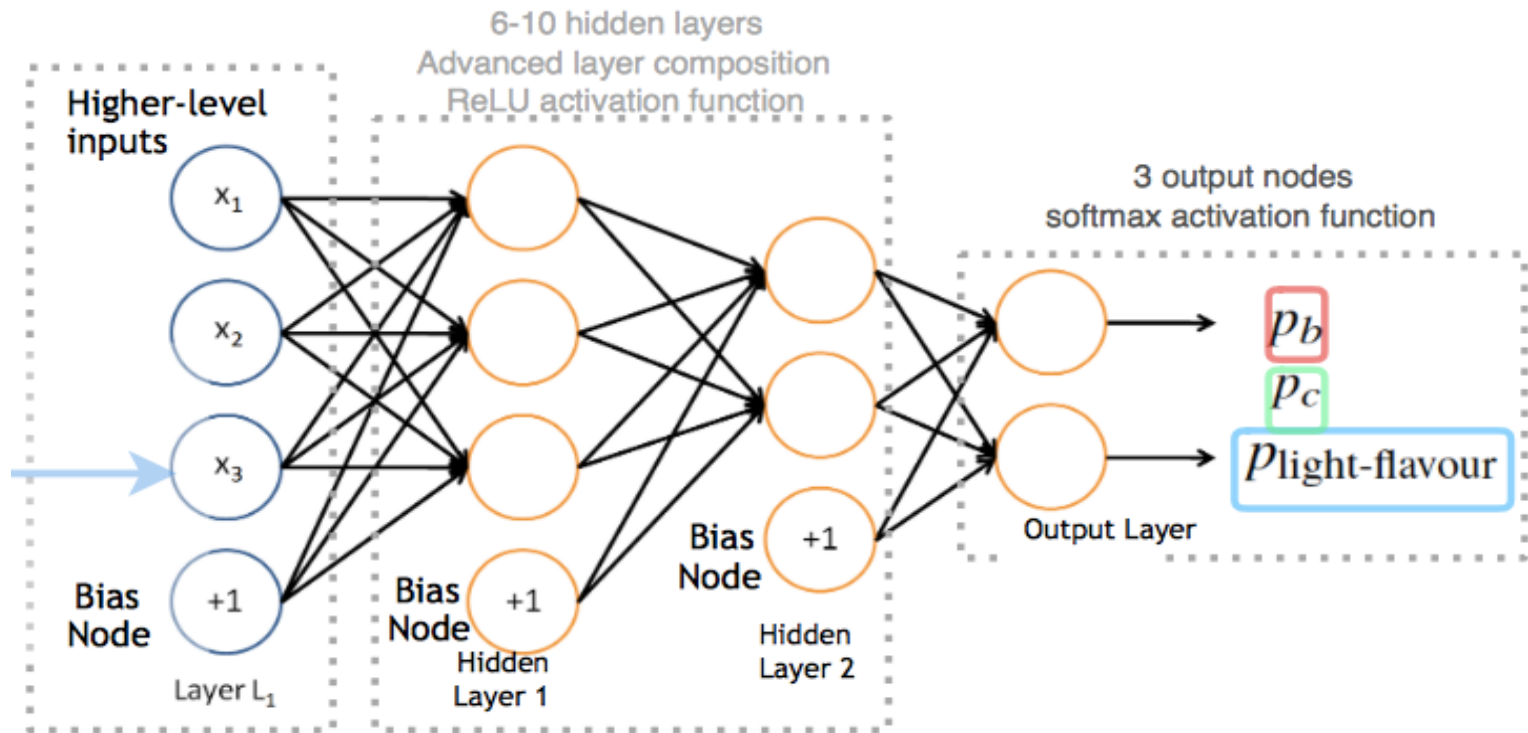
Recurrent NN: RNNIP

- Inputs: track properties of **arbitrary length**
- **Tracks as series**, ordered e.g. by d_0 significance
- Exploits **track correlations**
 - Long Short-Term Memory (LSTM) used to preserve memory and combat vanishing gradient problems

[ATL-PHYS-PUB-2017-003]



Deep Learning: DL1



[ATL-PHYS-PUB-2017-013]

- Trained using MC truth labels
- **Multi-class output** (easily extendable to more classes)
- **Flexibility:** one training for all OP for b- & c-tagging

[Parenthesis: training challenge]

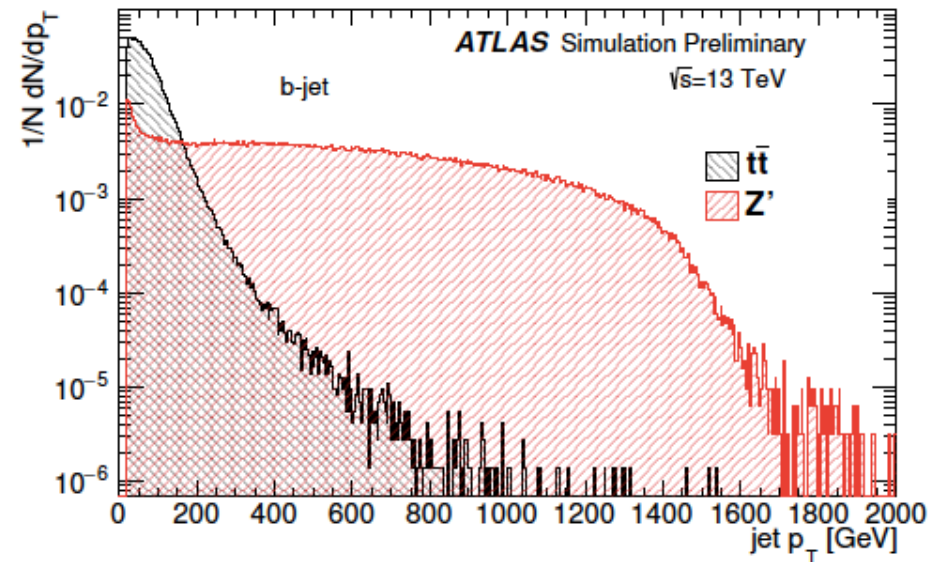
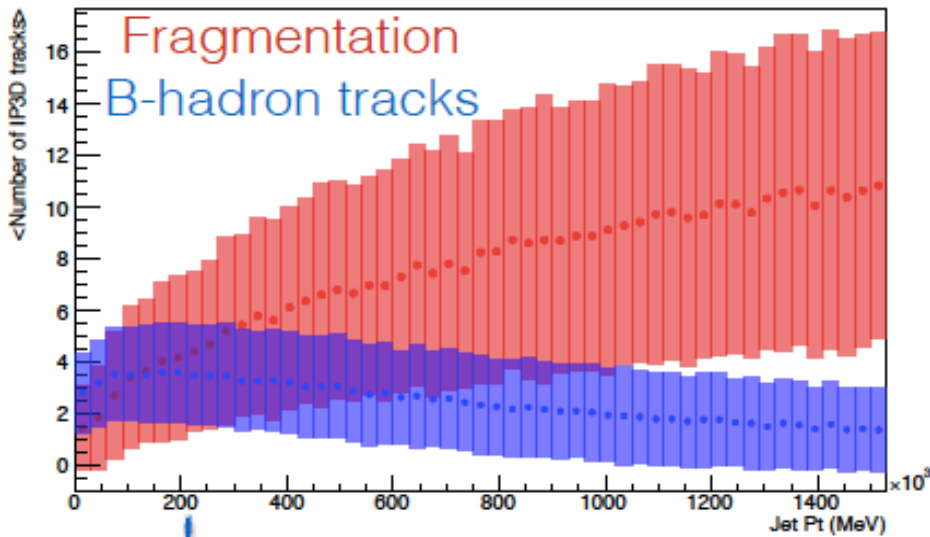
- How to find optimal hyperparameters
 - Brute force: grid search
- No off-the-shelf solution
- Toolsets exist, but no *instructions/theory*

Improve Flavor tagging at highest p_T

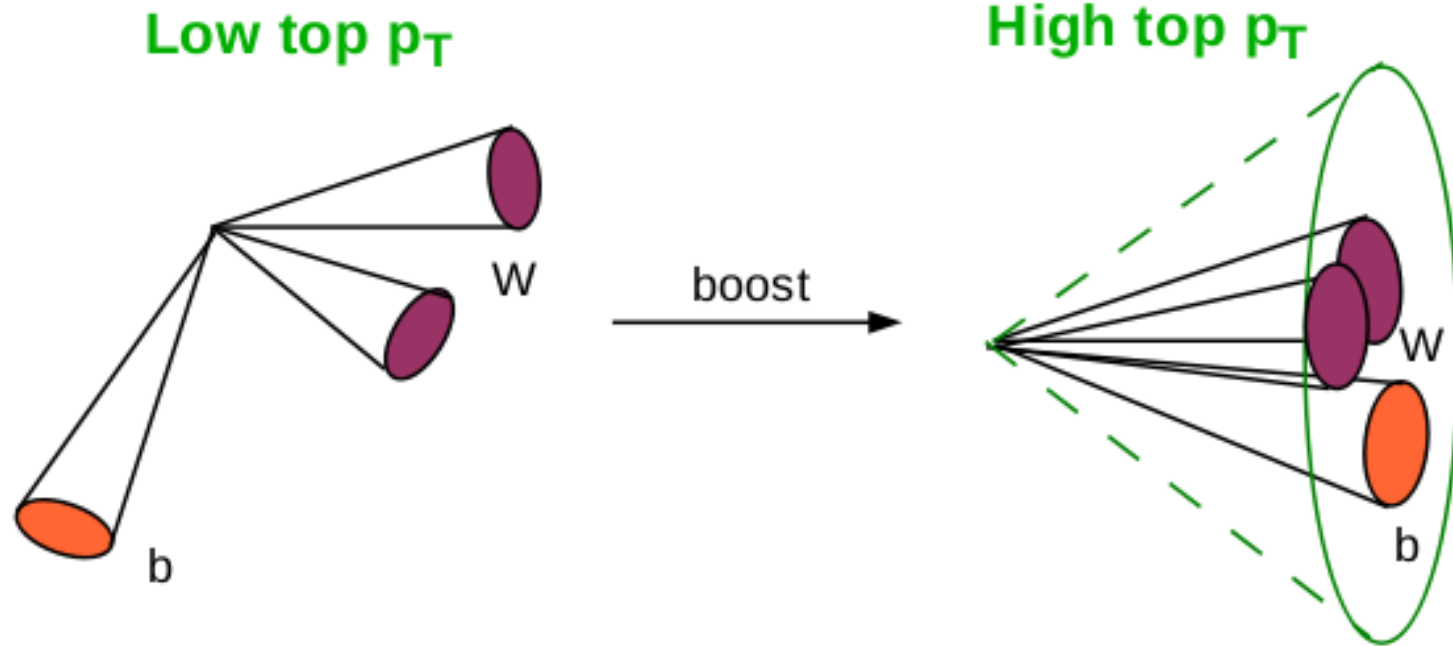
Provide algorithm with adequate training statistics at high p_T :
Use $Z' \rightarrow b\bar{b}/c\bar{c}/q\bar{q}$ (made \sim flat in p_T) instead of $t\bar{t}$

Flavor tagging very challenging at high p_T

[ATL-PHYS-PUB-2017-013]



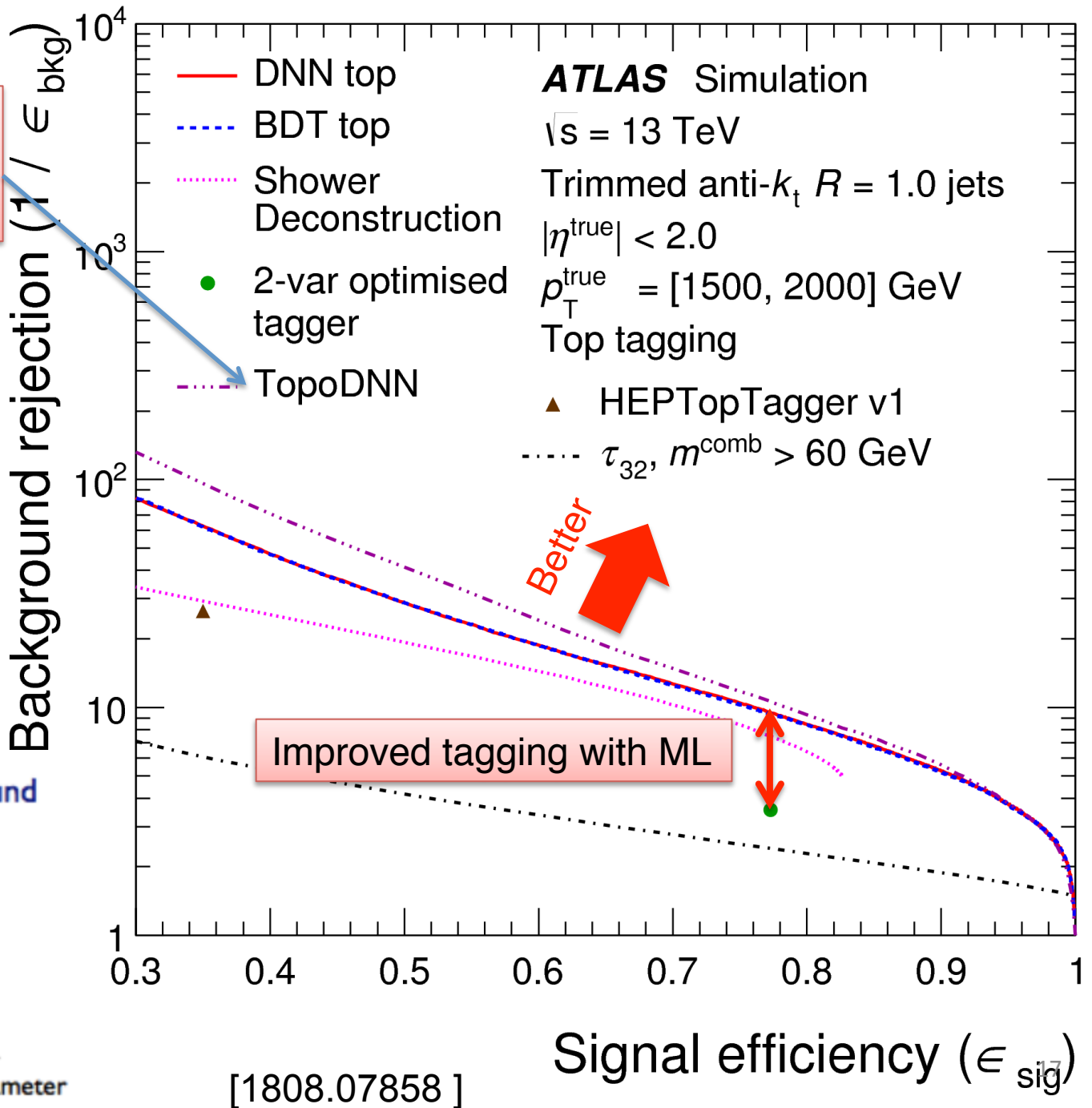
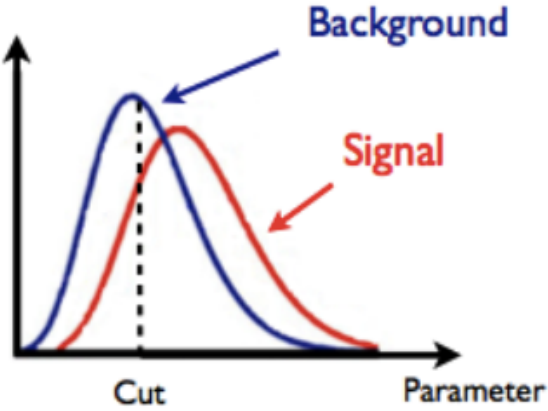
Boosted object tagging



Best results with raw data as input [1704.02124]

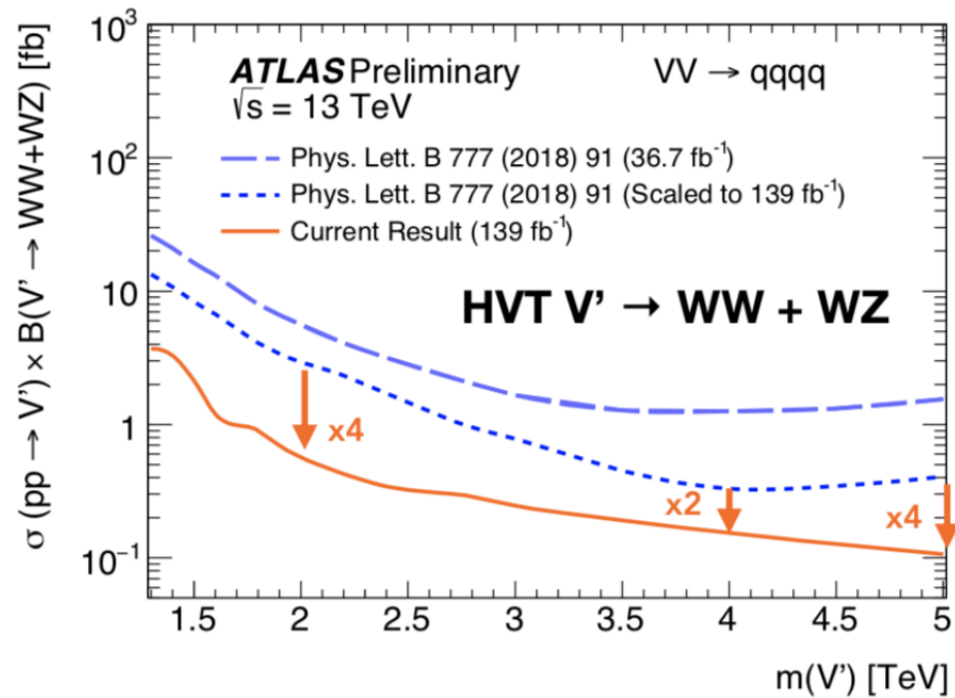
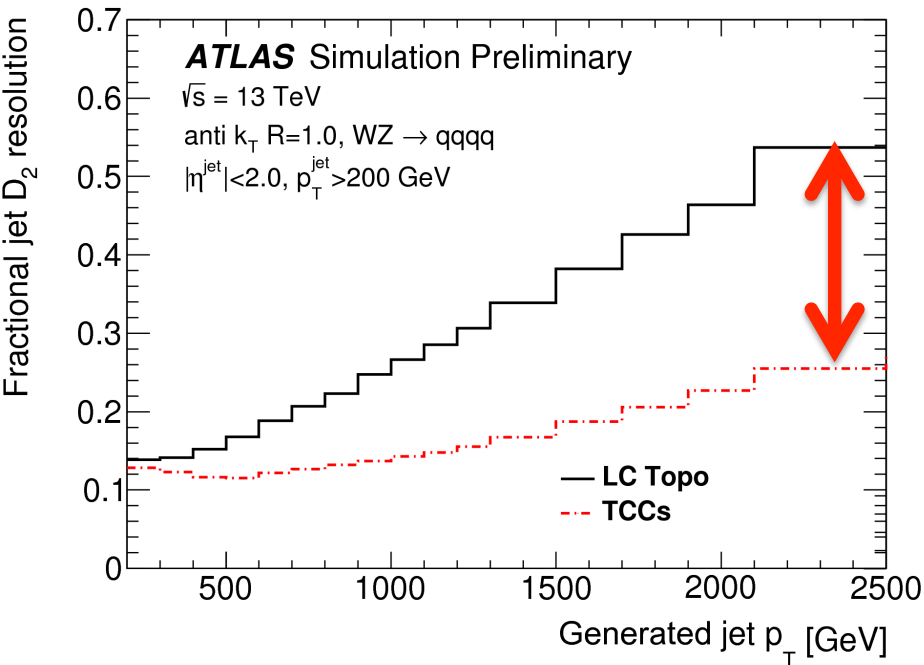
Is there more information to extract?

- Next frontier:
- Robustness
 - Versatility
 - Transparency

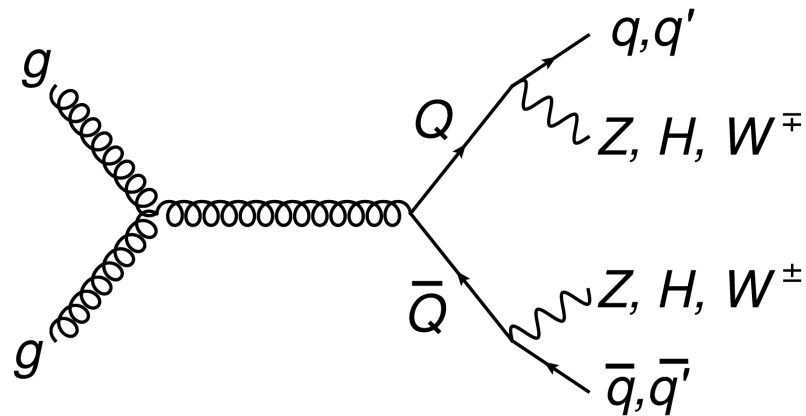


Boosted object tagging at highest p_T

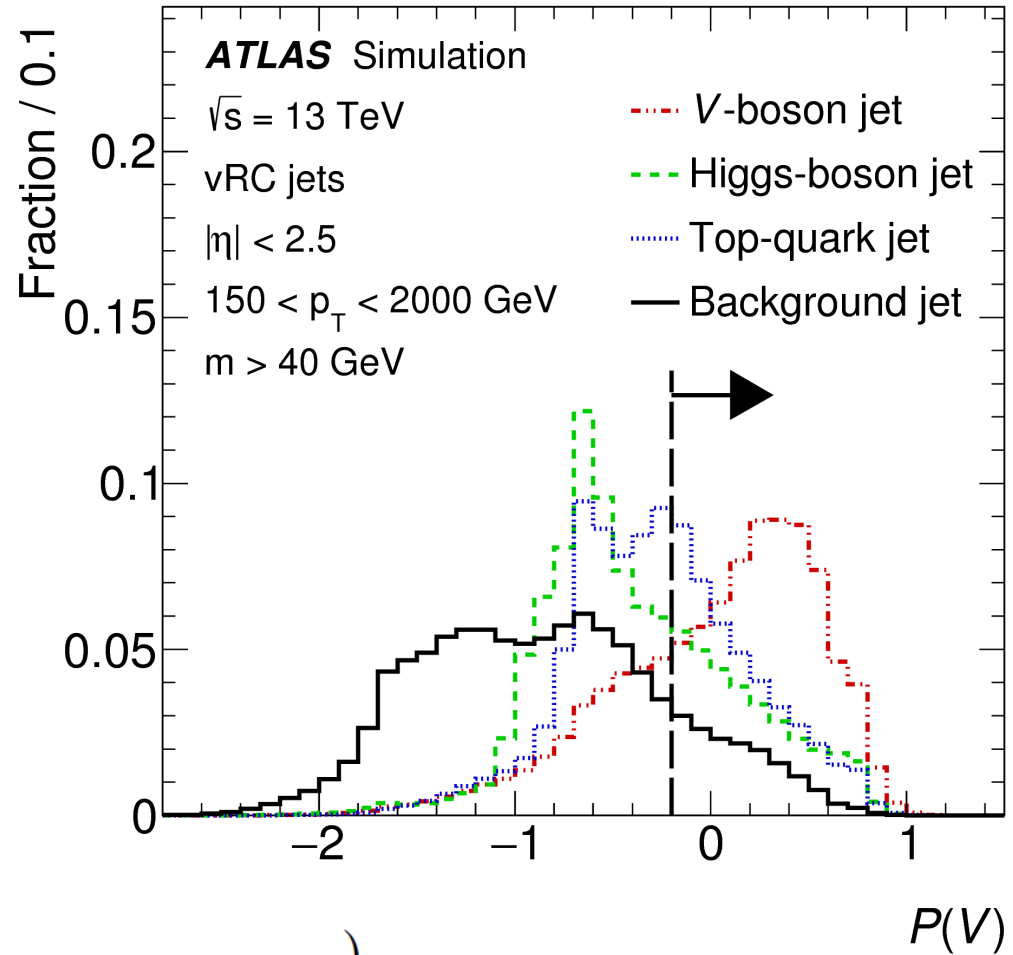
- Adequate training statistics at high p_T
- New features at high p_T : combine tracking & calorimeter info (Track-CaloCluster matching)



Multi-class tagging with deep NN



- Discriminate (in context of VLQ search)
 - W/Z
 - Higgs
 - Top
 - QCD-jet



$$P(V) = \log_{10} \left(\frac{D_{\text{DNN}}^V}{0.9 \cdot D_{\text{DNN}}^{\text{background}} + 0.05 \cdot D_{\text{DNN}}^t + 0.05 \cdot D_{\text{DNN}}^H} \right)$$

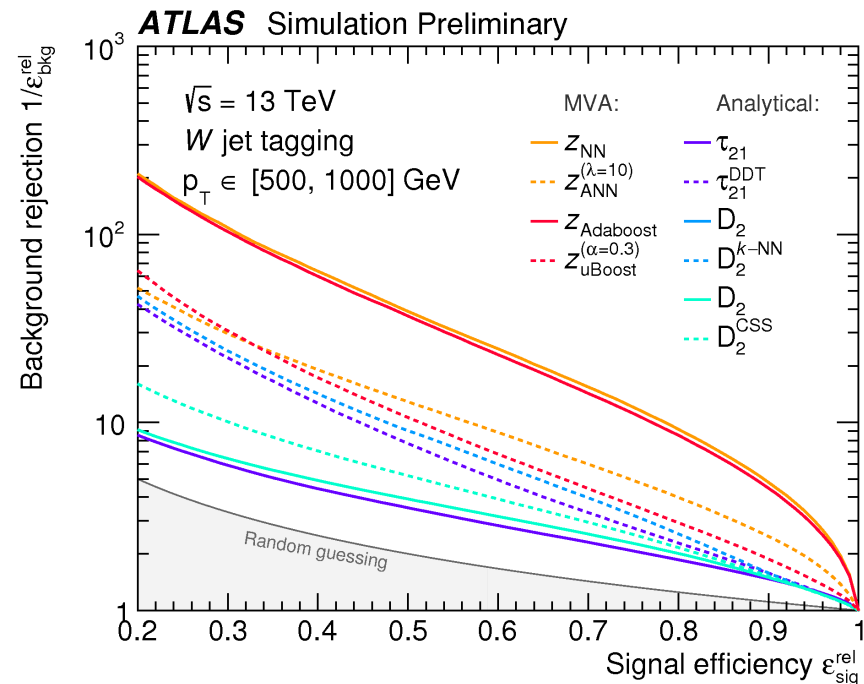
$P(V)$

Mass decorrelation

- Minimize sculpting of background jet mass distributions
- Enable more robust background estimation
- **Adversarial NN** penalizes classifier if mass is learned:

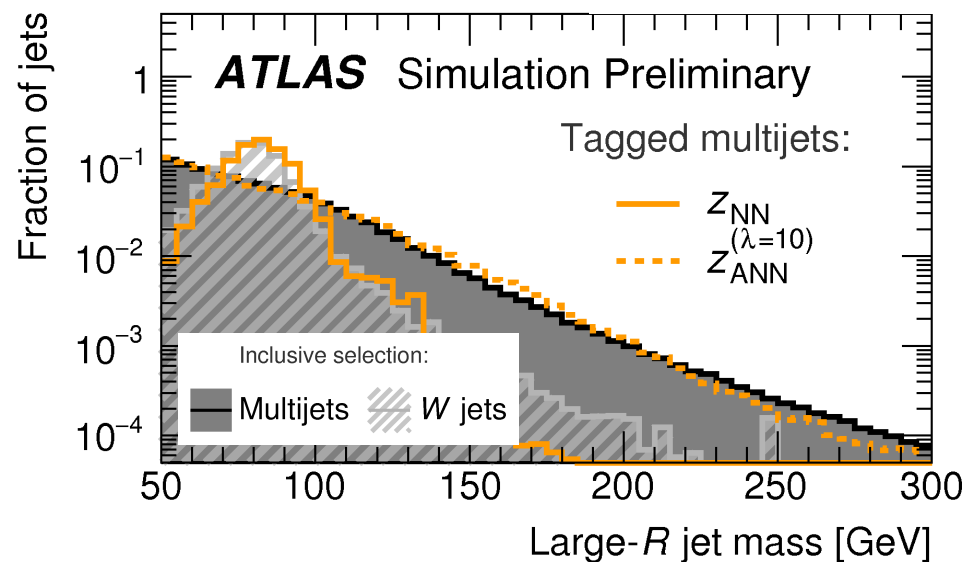
$$L = L_{\text{classifier}} + \lambda L_{\text{KL}}$$

- Sacrifice performance



+ Unsculpted BG

$\sqrt{s} = 13$ TeV, W jet tagging
Cuts at $\epsilon_{\text{sig}}^{\text{rel}} = 50\%$



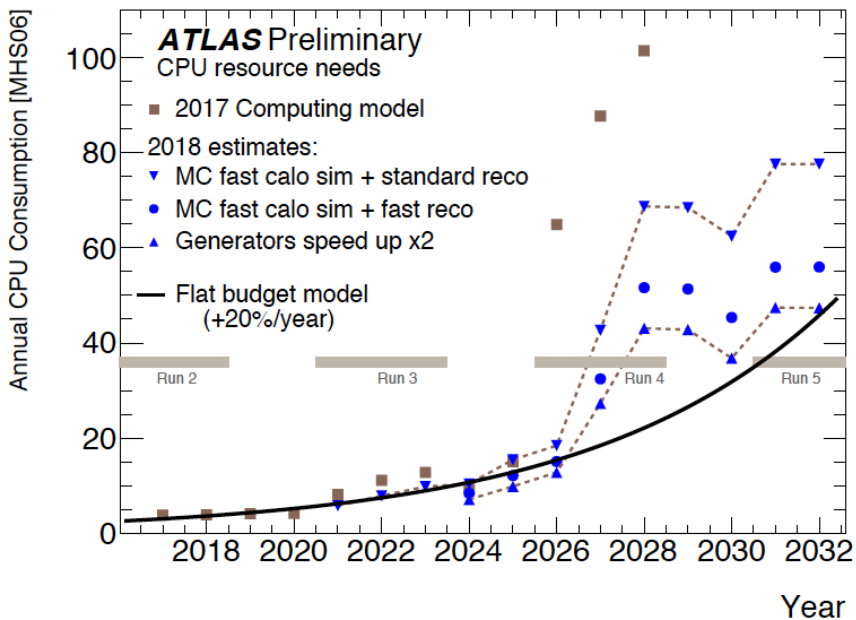
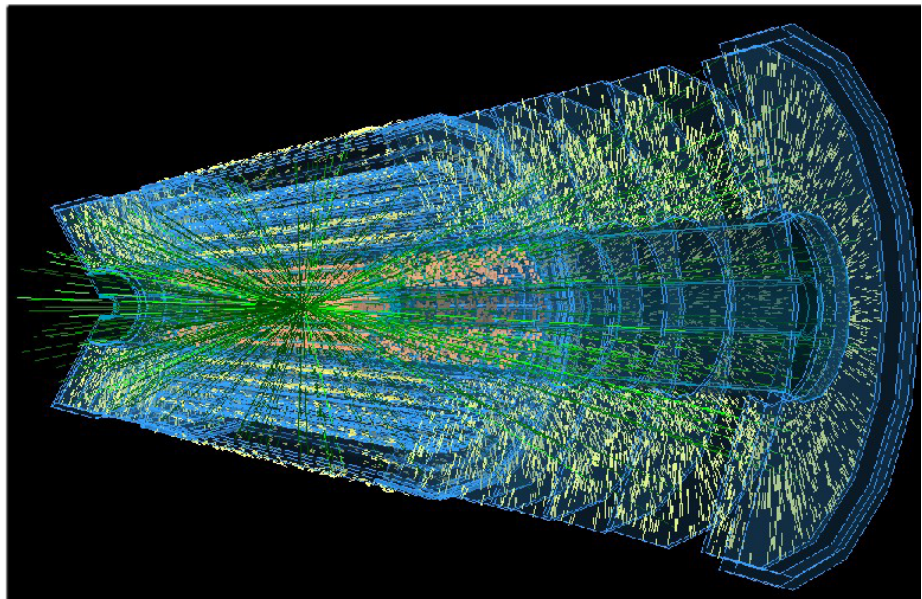
Mitigating impact of MC mismodeling (or pile-up dependence) is further application of ANN
Or train on data [1702.00414, 1708.02949]:
correct labeling → correct proportions

Beyond classification: clustering & generative models (faster)



2026: High-Lumi-LHC tracking crisis

100'000 space-points
10'000 tracks



Current algorithm:
combinatorial approach = slow!

Reconstruction limited by tracking

TrackML challenge

- Can ML help?
- *HL-LHC* data set with ACTS
 - $t\bar{t}$, 200 pile-up

ACTS: public high-fidelity simulation
[<http://acts.web.cern.ch/ACTS/>]

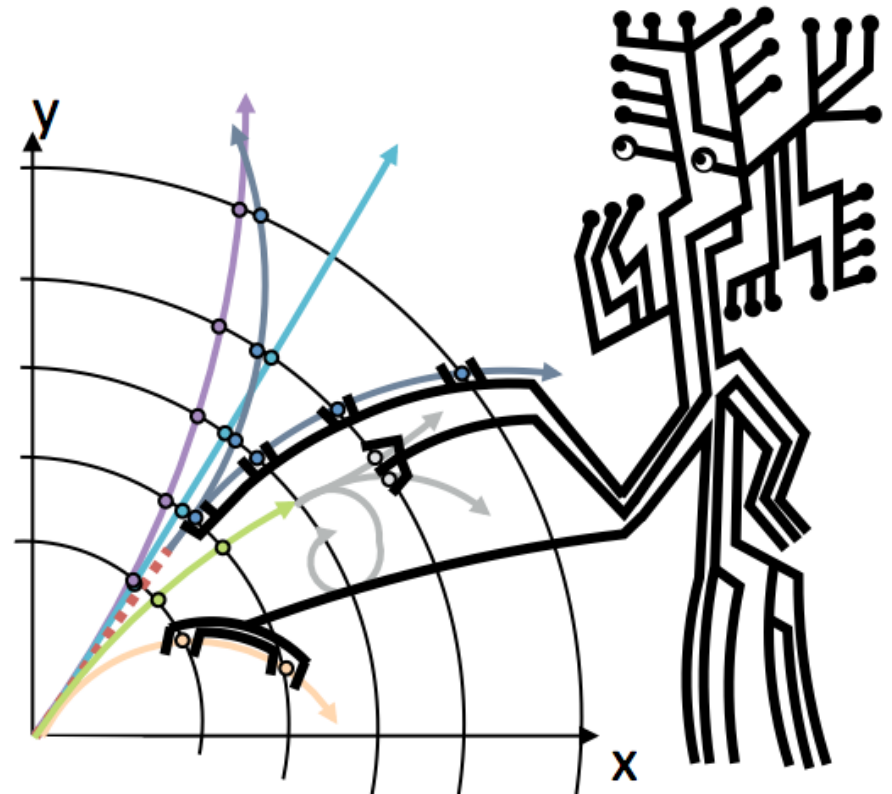
2 phases:

kaggle <https://www.kaggle.com/c/trackml-particle-identification>

Accuracy

Codalab <https://competitions.codalab.org/competitions/20112>

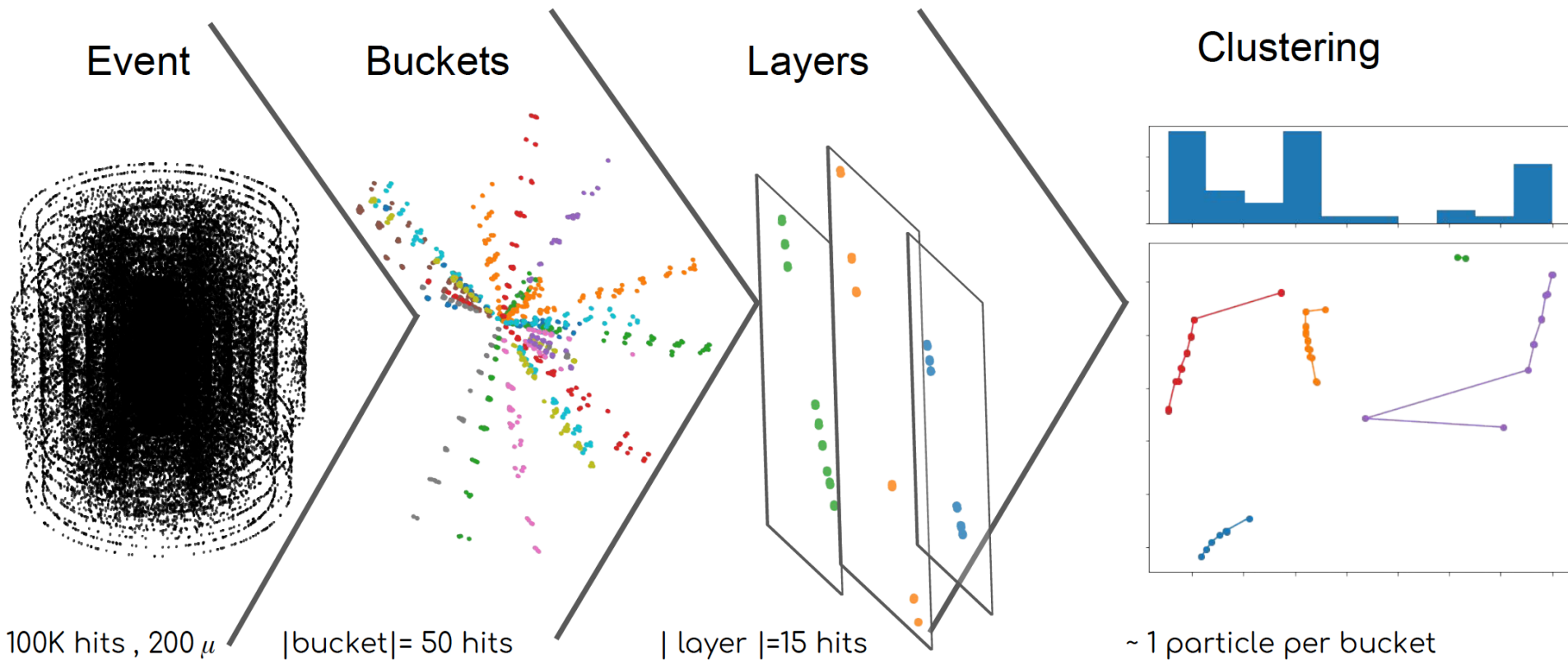
Throughput



Promising 2-Step Approach

1) Reduce combinatorial complexity

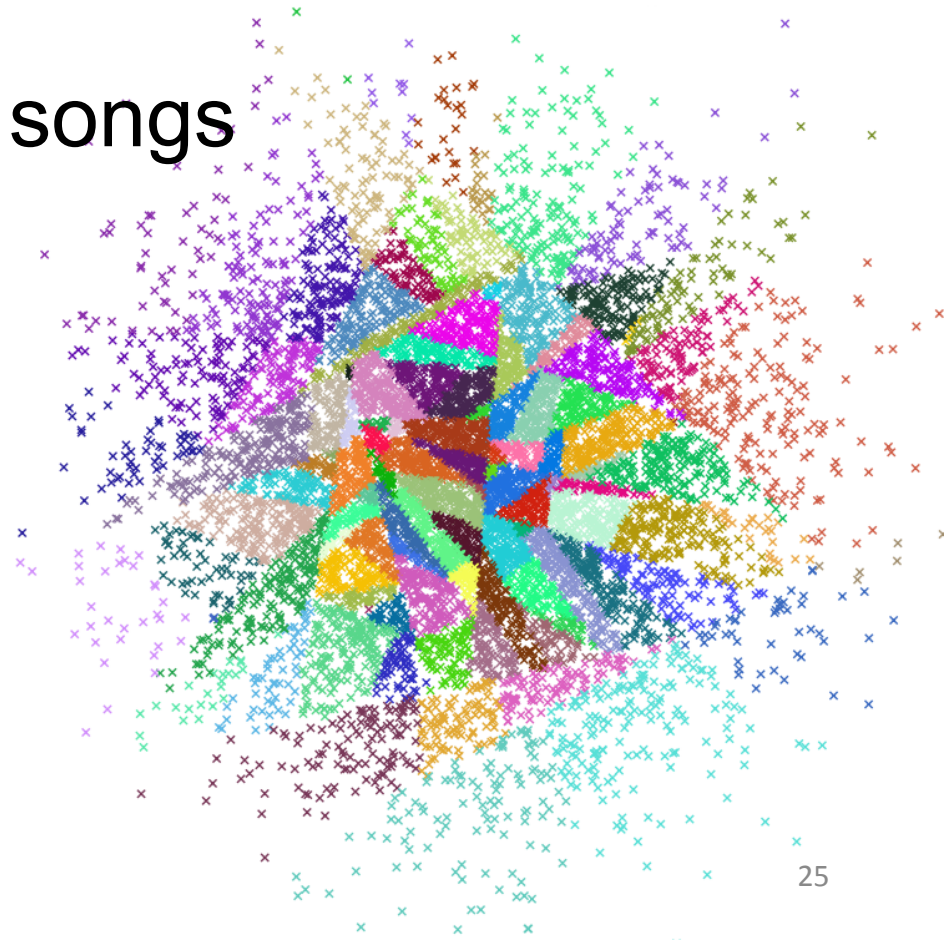
2) Track finding inside regions



Inspiration from spotify

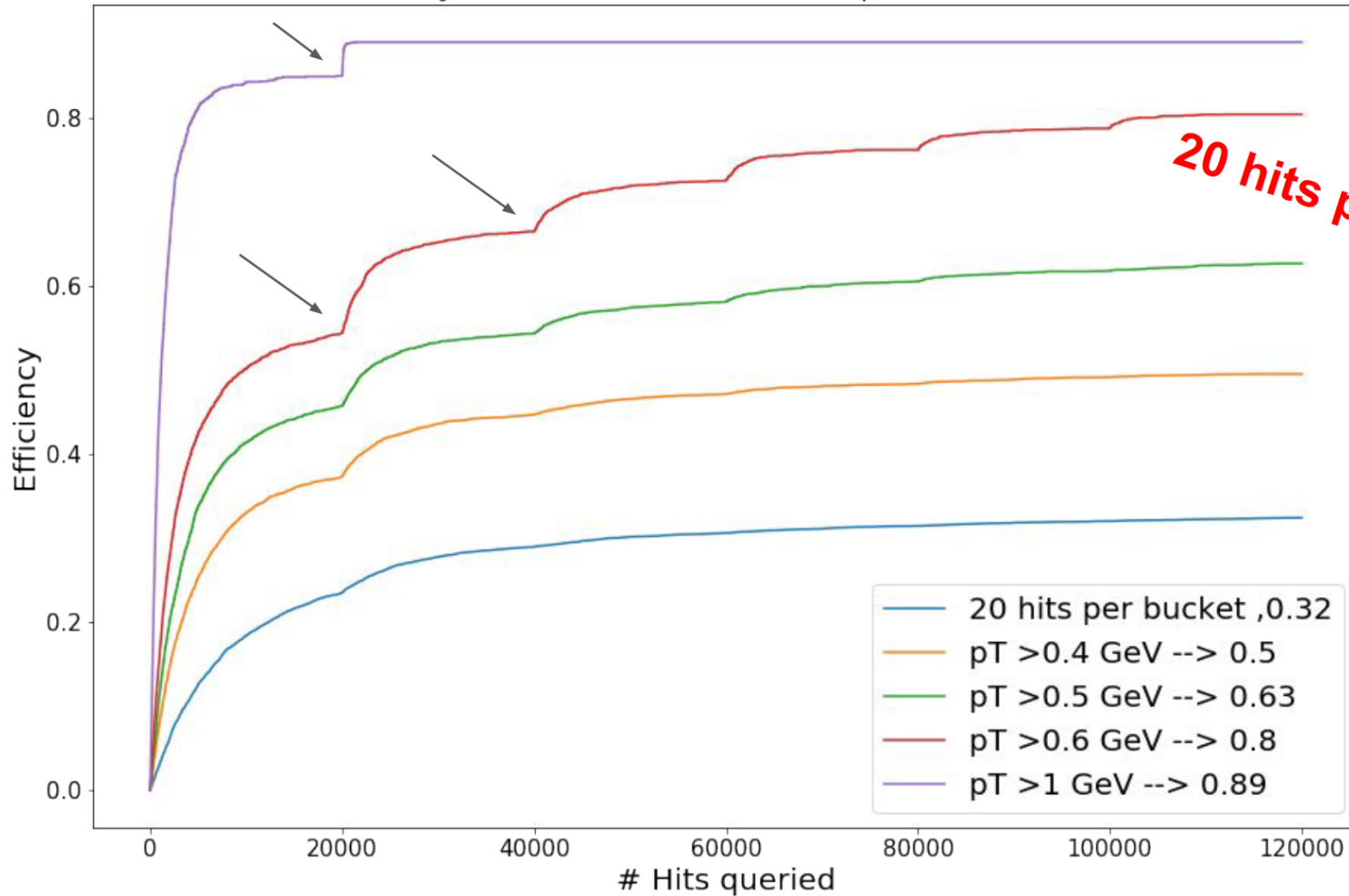


- Approximate Nearest Neighbors:
<https://github.com/spotify/annoy>
- $< 0.1\text{ms}$ to get n similar songs
– high-dimensional space
- Unsupervised
- Bucket definition based on angular distance



Bucketing provides fast & efficient seeding

Efficiency evolution, 6 trees of 20K queries , 8 hits min

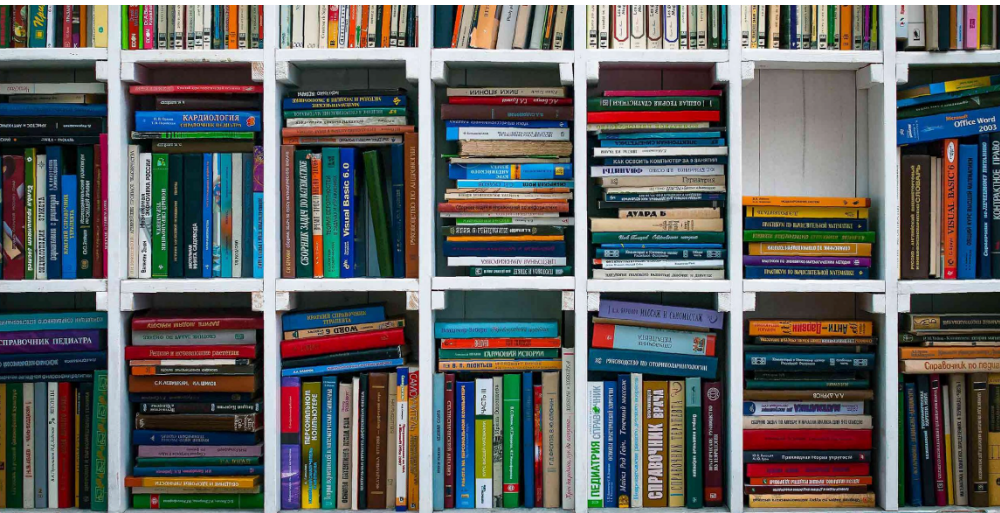


20 hits per bucket

**0.07 ms
per query**

Data Preparation

Data structure = Layout for memory

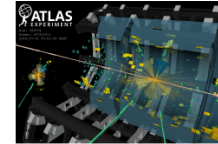
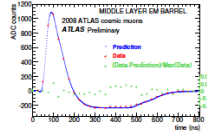
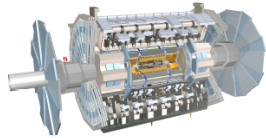
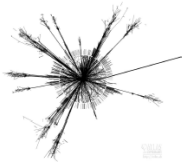


VS



- Most tracks are largely contained within a bucket
- Reduce problem to track finding inside buckets

Large-scale and high-fidelity simulation



Event generation

Detector simulation

Digitization

Reco.

Physics analysis

Bottleneck!

Group Production MC Reconstruction

Data Processing

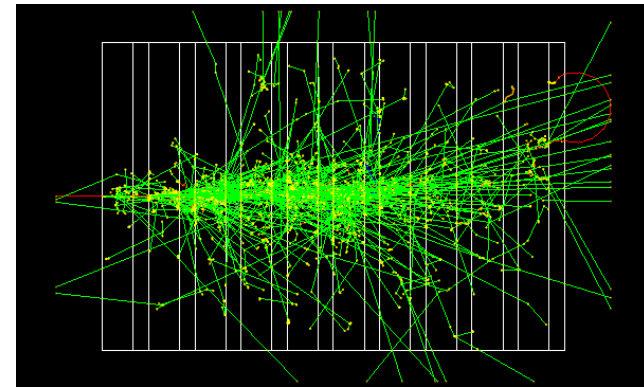
Analysis

T0 Processing

Others

MC Simulation

Dominated by Geant shower simulation!



Fast Calorimeter Simulation

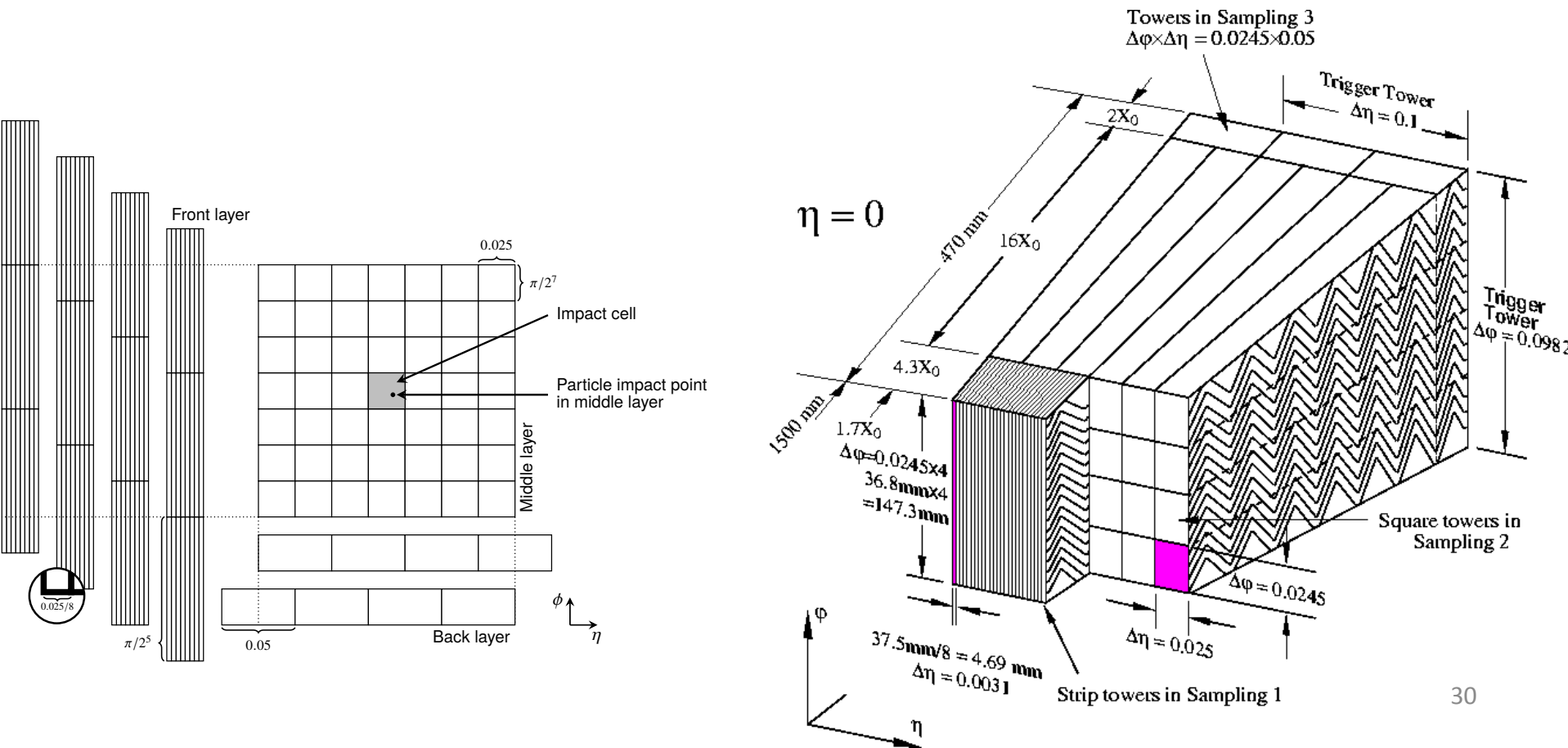
- Geant 4 too slow
- FastCaloSim V1 [ATL-PHYS-PUB-2010-013] used for years
- Improved FastCaloSim V2 [ATL-SOFT-PUB-2018-002] using PCA
- FastCaloSim fast enough but still not accurate enough for all simulation needs

- Objective: generative models to simulate calorimeter showers [ATL-SOFT-PUB-2018-001]

- Challenges:
 - Non-uniform geometry
 - Sparse data
 - Large dynamic range: tails

The ATLAS EM calorimeter

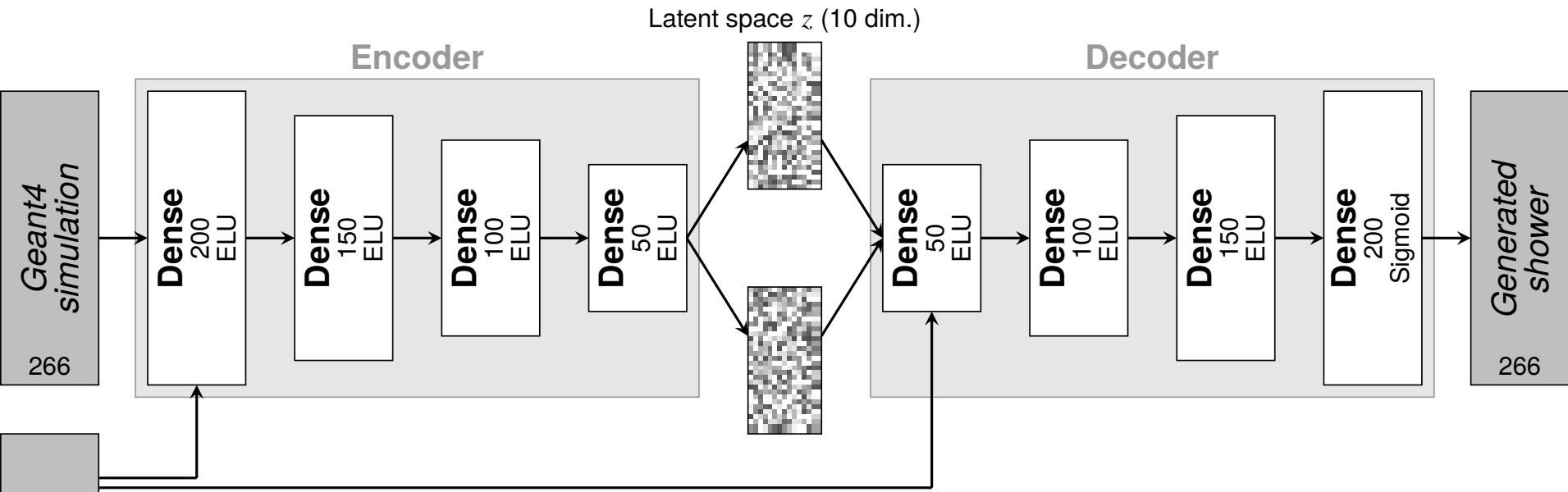
- Train and validate using G4 simulation of photons for the ATLAS geometry
 - Discrete particle energies logarithmically spaced between 1 and 260 GeV
 - Uniformly distributed in $0.20 < |\eta| < 0.25$



Deep Generative Models (VAE, GAN)

Variational Auto-Encoder (VAE) architecture:

[ATL-SOFT-PUB-2018-001]



Particle energy

Reconstruction loss

Constrain total energy

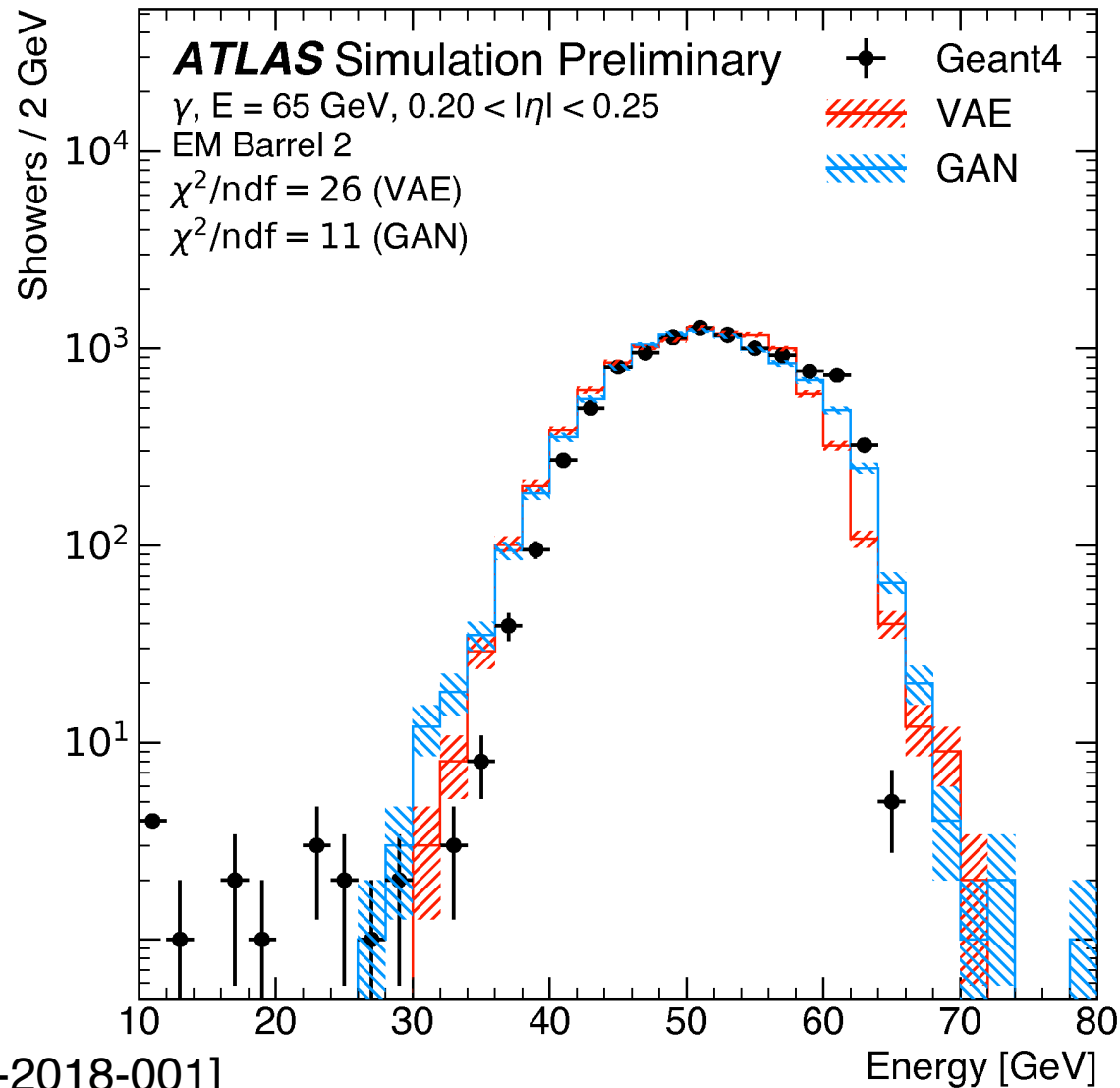
$$L_{\text{VAE}}(x, \tilde{x}) = w_{\text{reco}} \mathbb{E}_{z \sim q_{\theta}(z|x)} [\log p_{\phi}(x|z)] - w_{\text{KL}} \text{KL}(q_{\theta}(z|x) || p(z)) + w_{E_{\text{tot}}} L_{E_{\text{tot}}}(x, \tilde{x}) + \sum_i^M w_i L_{E_i}(x, \tilde{x}).$$

Kullback-Leibler divergence

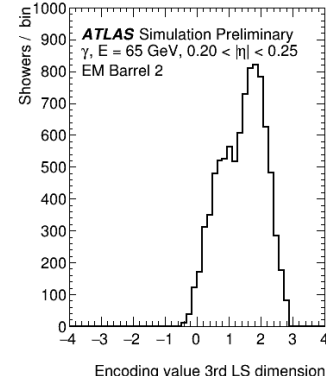
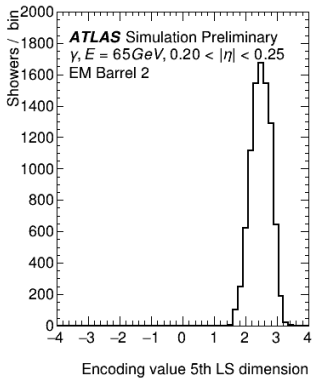
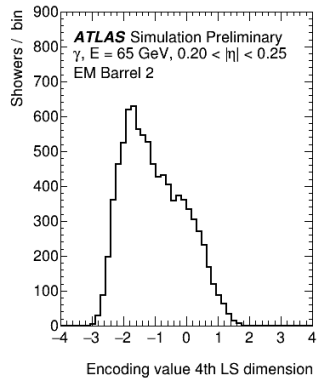
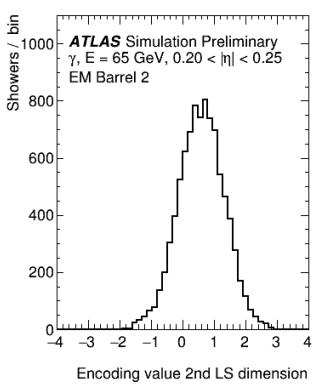
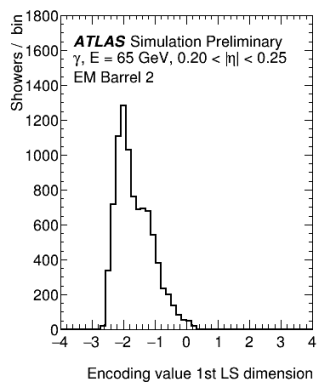
Constrain energy fractions in layers

Deep Generative Models (VAE, GAN)

Validation: promising



VAE Latent Space

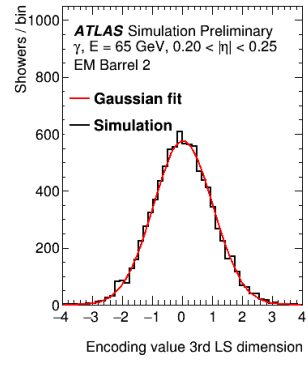
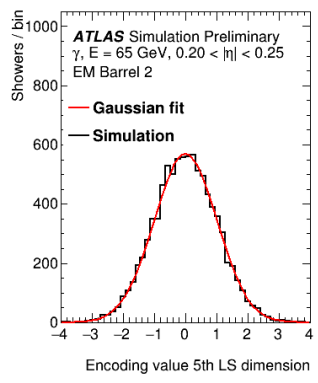
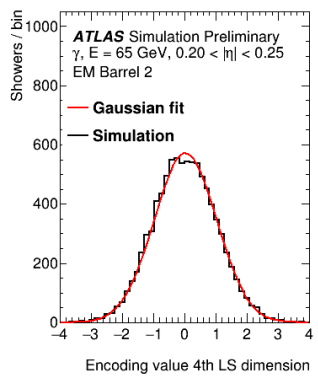
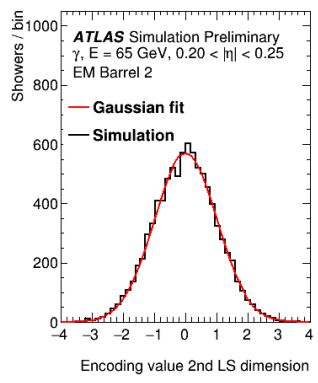
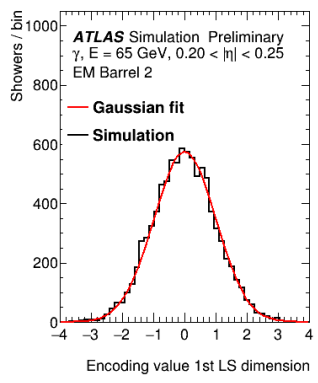


Non-Gaussian
5D latent space

Inverse Autoregressive
Flow [1606.04934]:
latent space more
Gaussian



Allows one to sample
from Gaussian
distribution



Integration into ATLAS (C++) Software

- Implemented with Lightweight Trained Neural Network (LWTNN) [<https://github.com/lwtnn/lwtnn>]
- Flag to switch trained model
- DNNCaloGAN same speed as FastCaloSim V2
 - 65 GeV single photon: G4 → FCS : 10 seconds → 70 ms
- LWTNN takes <1 ms per shower

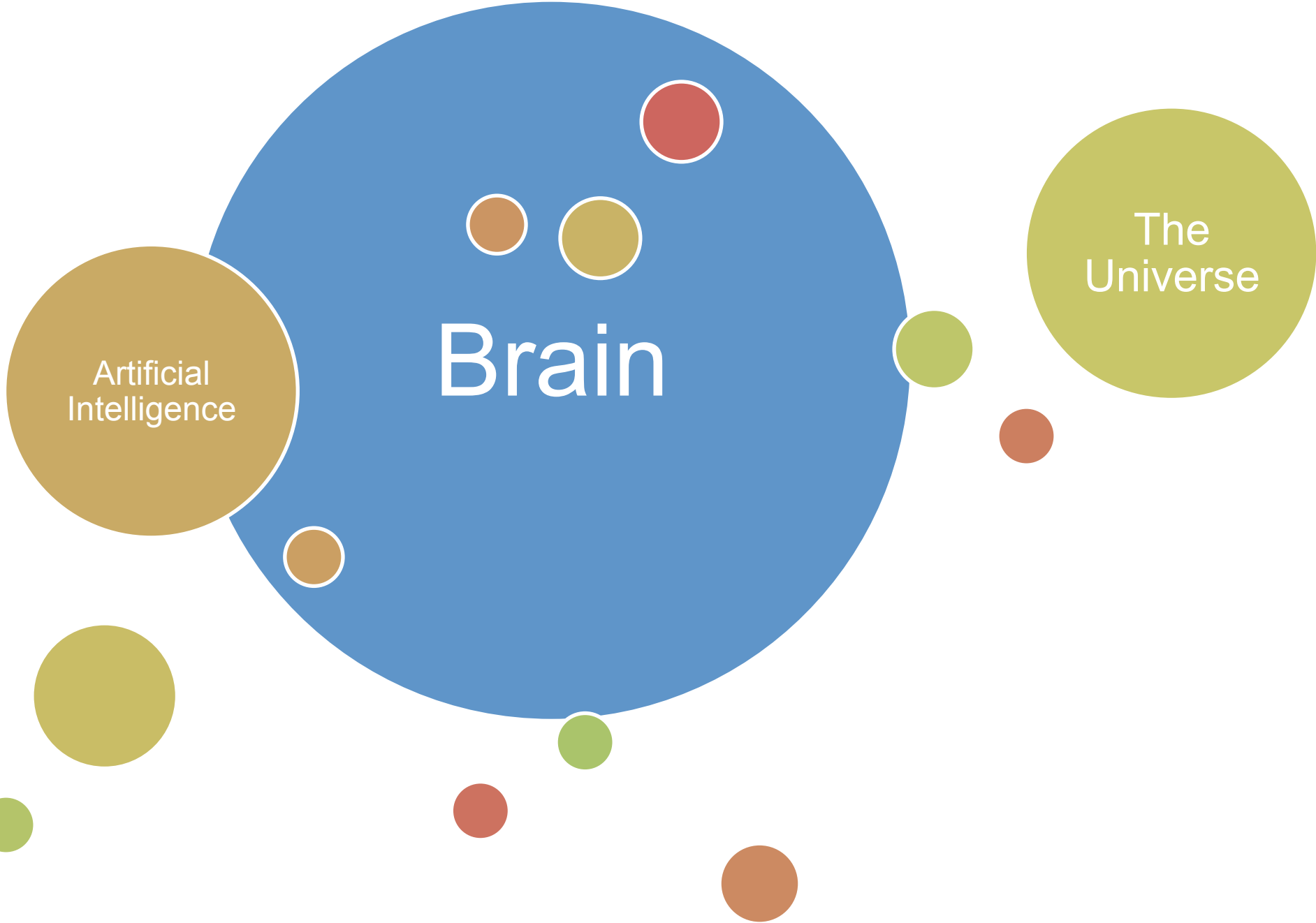
- Model trained on particles with fixed energy, but interpolates well to other energies

- Only central eta for now
- Plan: use higher uniform granularity for full eta range

The ML revolution in HEP has started

- Exciting time – room for **creativity**
 - Opportunity for young researchers to think outside the box
- Fruitful interdisciplinary work
 - ...but also hard work from proof of principle to realization in ATLAS
- Better, faster, more automated, ...
 - ...but also maintainability & memory footprint
- Trend from “ML in analysis” to “ML for simulation, reconstruction, trigger”
- “Raw data” vs. “human-engineered features”
- Domain knowledge vs. what machine learns
- New frontiers
 - Latency: trigger
 - Specialized hardware: FPGA, GPU, Custom DL chips, ...
 - Anomaly detection
 - Interpretability
 - ...

*Relevance for
flavour tagging*



Backup