



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR MATHEMATIK, INFORMATIK
UND NATURWISSENSCHAFTEN



FSP CMS
Erforschung von
Universum und Materie

EXZELLENZCLUSTER
QUANTUM UNIVERSE

Gefördert durch:

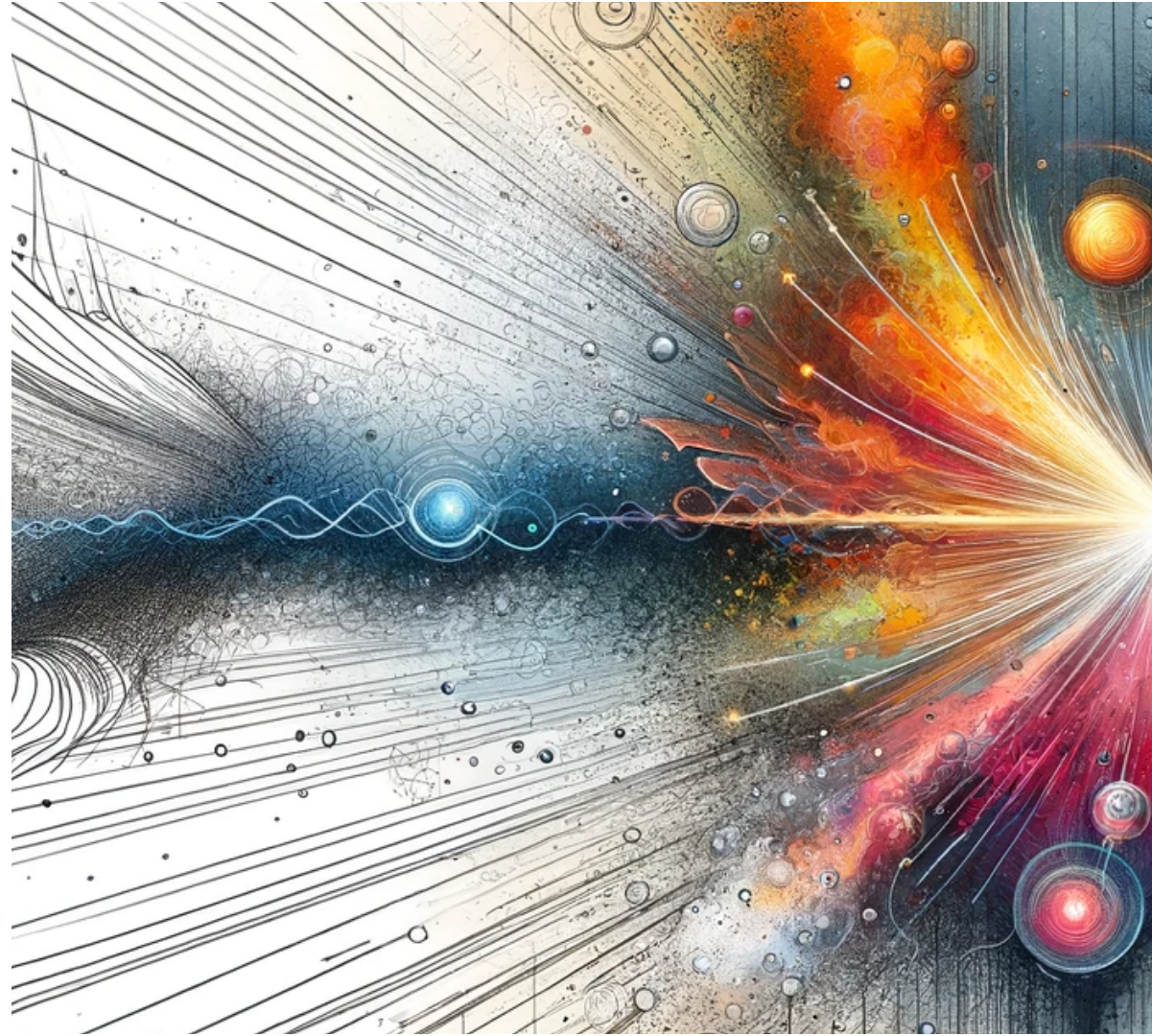


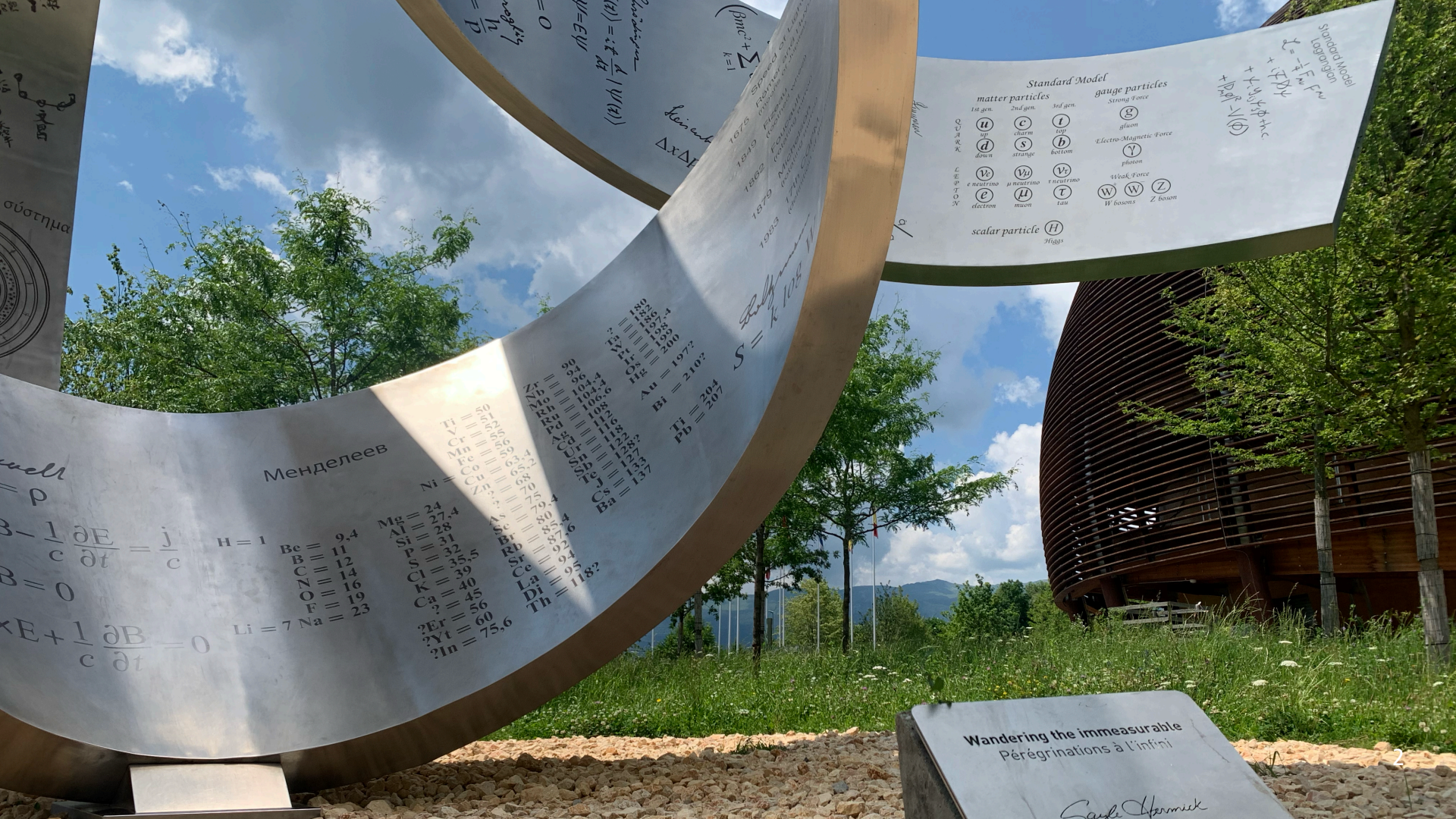
Bundesministerium
für Forschung, Technologie
und Raumfahrt

Transforming jet flavour tagging at the **ATLAS** and **CMS** experiments

Dr. Philipp Gadow

30.06.2026 | CPPS seminar
Universität Siegen





ψ
 ψ^*
 ψ^2
 ψ^3
 ψ^4
 ψ^5
 ψ^6
 ψ^7
 ψ^8
 ψ^9
 ψ^{10}
 ψ^{11}
 ψ^{12}
 ψ^{13}
 ψ^{14}
 ψ^{15}
 ψ^{16}
 ψ^{17}
 ψ^{18}
 ψ^{19}
 ψ^{20}
 ψ^{21}
 ψ^{22}
 ψ^{23}
 ψ^{24}
 ψ^{25}
 ψ^{26}
 ψ^{27}
 ψ^{28}
 ψ^{29}
 ψ^{30}
 ψ^{31}
 ψ^{32}
 ψ^{33}
 ψ^{34}
 ψ^{35}
 ψ^{36}
 ψ^{37}
 ψ^{38}
 ψ^{39}
 ψ^{40}
 ψ^{41}
 ψ^{42}
 ψ^{43}
 ψ^{44}
 ψ^{45}
 ψ^{46}
 ψ^{47}
 ψ^{48}
 ψ^{49}
 ψ^{50}
 ψ^{51}
 ψ^{52}
 ψ^{53}
 ψ^{54}
 ψ^{55}
 ψ^{56}
 ψ^{57}
 ψ^{58}
 ψ^{59}
 ψ^{60}
 ψ^{61}
 ψ^{62}
 ψ^{63}
 ψ^{64}
 ψ^{65}
 ψ^{66}
 ψ^{67}
 ψ^{68}
 ψ^{69}
 ψ^{70}
 ψ^{71}
 ψ^{72}
 ψ^{73}
 ψ^{74}
 ψ^{75}
 ψ^{76}
 ψ^{77}
 ψ^{78}
 ψ^{79}
 ψ^{80}
 ψ^{81}
 ψ^{82}
 ψ^{83}
 ψ^{84}
 ψ^{85}
 ψ^{86}
 ψ^{87}
 ψ^{88}
 ψ^{89}
 ψ^{90}
 ψ^{91}
 ψ^{92}
 ψ^{93}
 ψ^{94}
 ψ^{95}
 ψ^{96}
 ψ^{97}
 ψ^{98}
 ψ^{99}
 ψ^{100}



Maxwell

$$\nabla \cdot \mathbf{E} = \rho$$
$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{B} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j}$$

Менделеев

Periodic table of elements with atomic numbers and names in Cyrillic.

H=1	He=2																																																	
Li=7	Be=9.4	B=11	C=12	N=14	O=16	F=19	Ne=23																																											
								Na=23	Mg=24	Al=27.4	Si=28	P=31	S=32	Cl=35.5	K=40	Ca=45	Sc?	Ti=50	V=56	Cr=60	Mn=75.6	Fe=78.6	Ni=79	Cu=80	Zn=81	Ga=83	Ge=84	As=85	Se=87	Br=90	Kr=94	Rb=95	Sr=97	Y=98	Zr=101	Nb=102	Mo=106	Tc?	Ru=108	Rh=110	Pd=112	Ag=113	Cd=114	In=115	Sn=118	Pb=207	Bi=208	Po?	At?	Rn?

Handwritten mathematical formulas and notes.

$$\frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V \psi$$
$$\Delta x \Delta p \geq \frac{\hbar}{2}$$
$$\psi(x) = \sum_{k=1}^{\infty} \sin(kx)$$

1872 Standard Model
1840 Bohr
1869 Mendeleev
1879 Thomson
1903 Rutherford
1926 Schrödinger
1927 Heisenberg

Standard Model of particle physics chart.

matter particles		gauge particles	
1st gen.	2nd gen.	1st gen.	Strong Force
$\begin{pmatrix} u \\ d \end{pmatrix}$	$\begin{pmatrix} c \\ s \end{pmatrix}$	$\begin{pmatrix} g \\ g \\ g \end{pmatrix}$	gluon
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu \\ \nu_\tau \\ \tau \end{pmatrix}$	$\begin{pmatrix} \gamma \\ W \\ Z \end{pmatrix}$	Electro-Magnetic Force
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}$	$\begin{pmatrix} \nu_\mu \\ \mu \\ \nu_\tau \\ \tau \end{pmatrix}$	$\begin{pmatrix} W \\ W \\ Z \end{pmatrix}$	Weak Force
scalar particle $\begin{pmatrix} H \\ Higgs \end{pmatrix}$			

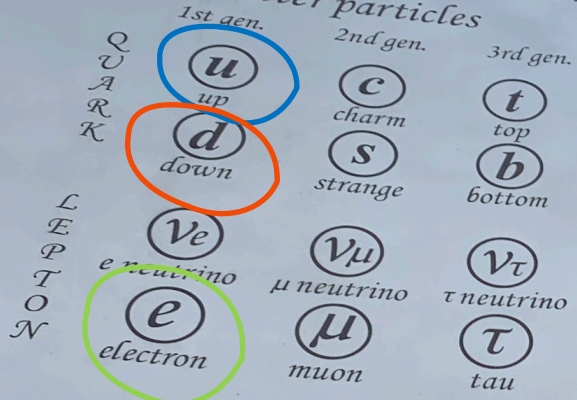
Wandering the immeasurable
Pérégrinations à l'infini

Sarah Hermick

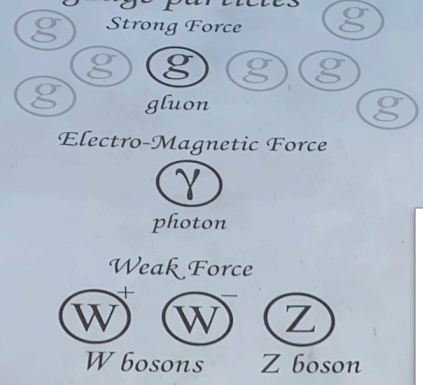
Handwritten signature

Standard Model

matter particles



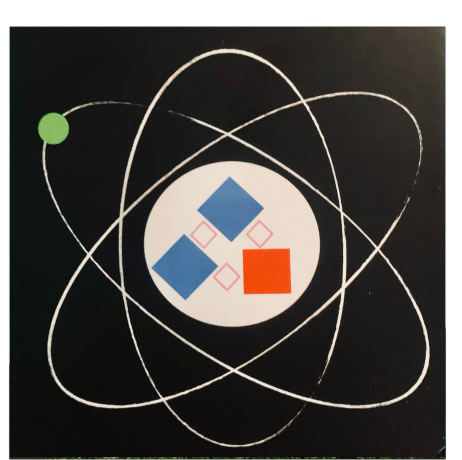
gauge particles



Standard Model Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{D} \psi + \text{h.c.}$$

Handwritten notes and symbols



Hydrogen atom

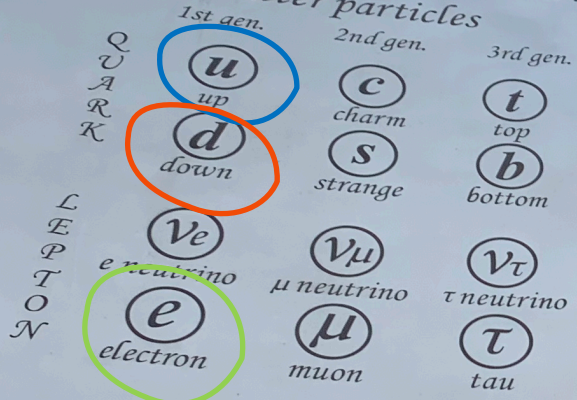
Wandering the Immeasurable
Pérégrinations à l'infini

Handwritten signature

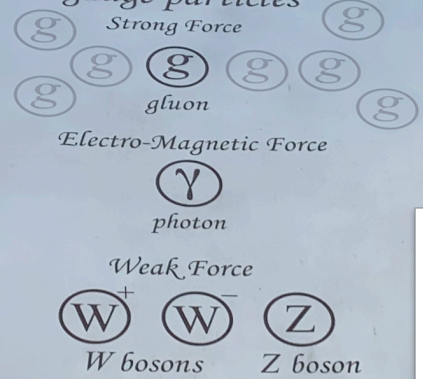
Handwritten signature

Standard Model

matter particles

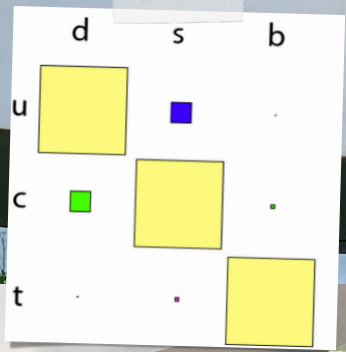


gauge particles

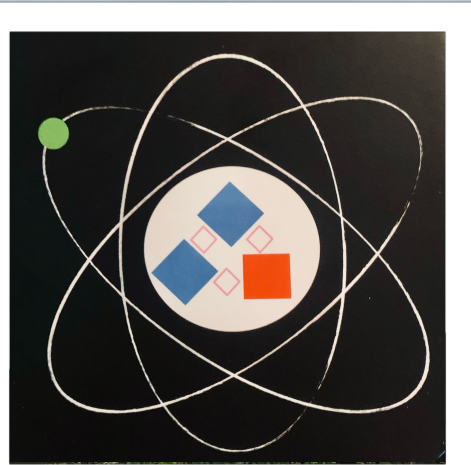


Standard Model Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{D} \psi + \text{h.c.} + \sum_{\text{bosons}} \mathcal{L}_{\text{boson}}$$



CKM matrix for quarks



Hydrogen atom

Wandering the Immeasurable
Pérégrinations à l'infini

Handwritten signature

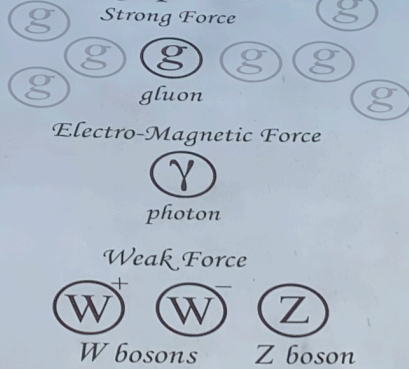
Handwritten signature

Standard Model

matter particles

	1st gen.	2nd gen.	3rd gen.
QUARK	u up	c charm	t top
	d down	s strange	b bottom
	ν_e e neutrino	ν_μ μ neutrino	ν_τ τ neutrino
LEPTON	e electron	μ muon	τ tau

gauge particles



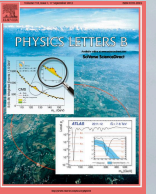
scalar particle H
Higgs

Higgs Boson Discovery
04. Juli 2012



Nobel Price
in Physics 2013

Standard Model
Lagrangian
 $\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \gamma^\mu \partial_\mu \psi + \dots$



Does the Brout-Englert-Higgs mechanism occur in Nature?

Discovery of a Higgs boson of mass 125 GeV at the Large Hadron Collider (LHC) completed the Standard Model (SM).

We will further explore the Higgs sector in the next decade.

$$\mathcal{L}_{SM} \supset -\frac{1}{4} F_{\mu\nu}^a F^{a,\mu\nu} + \bar{\psi}_i \gamma_\mu D_\mu^i \psi_j$$

constrained by gauge symmetry, tested to high precision

$$+ |D_\mu \phi|^2 + y_{ij} \bar{\psi}_i \phi \psi_j - V(\phi)$$

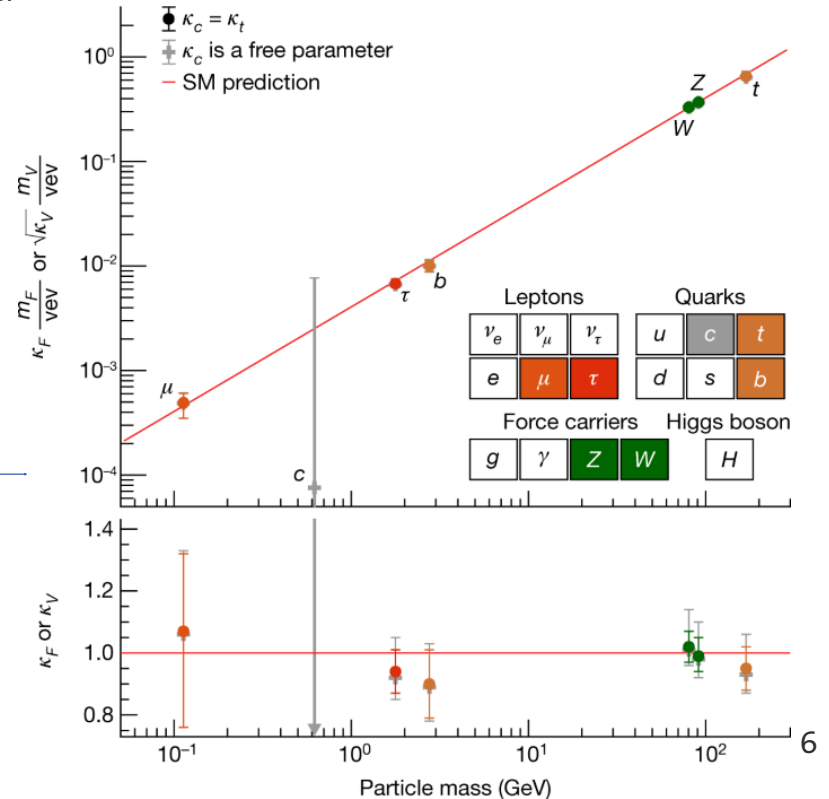
Interaction with weak gauge bosons

Yukawa couplings
Hierarchy of fermion masses and flavour

Higgs potential
Investigation at (HL)-LHC



ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". Nature 607, 52–59 (2022).



Does the Brout-Englert-Higgs mechanism occur in Nature?

Discovery of a Higgs boson of mass 125 GeV at the Large Hadron Collider (LHC) completed the Standard Model (SM).

We will further explore the Higgs sector in the next decade.

$$\mathcal{L}_{SM} \supset -\frac{1}{4}F_{\mu\nu}^a F^{a,\mu\nu} + \bar{\psi}_i \gamma_\mu D^\mu \psi_j$$

constrained by gauge symmetry, tested to high precision

$$+ |D_\mu \phi|^2 + y_{ij} \bar{\psi}_i \phi \psi_j - V(\phi)$$

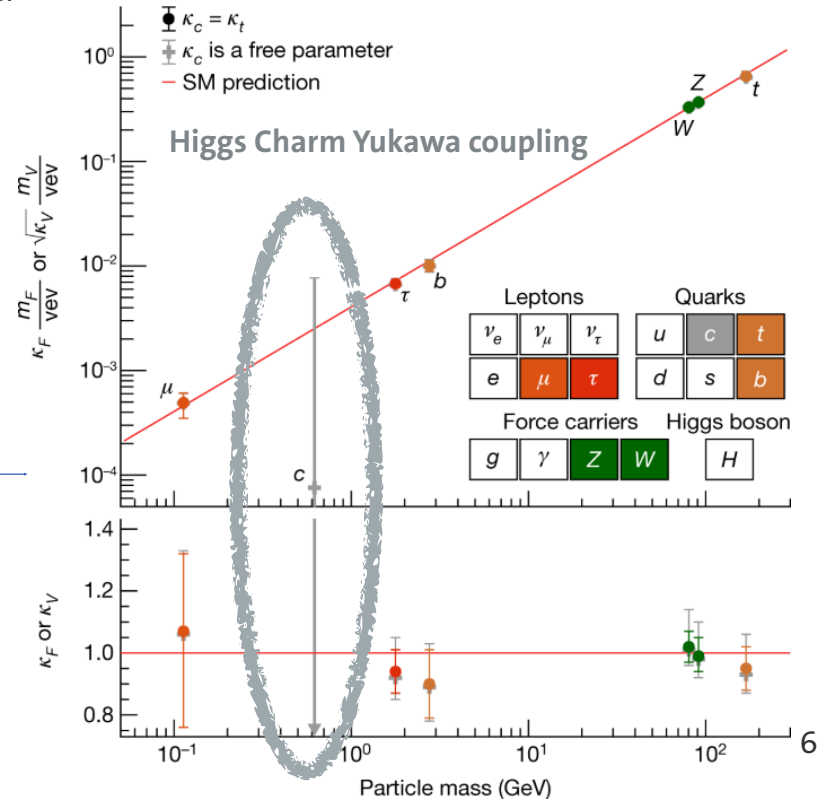
Interaction with weak gauge bosons

Yukawa couplings
Hierarchy of fermion masses and flavour

Higgs potential
Investigation at (HL)-LHC



ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". Nature 607, 52–59 (2022).



Does the Higgs potential look like a Mexican hat?

Source: "Advanced Course on Higgs Physics", J. Braathen 2024

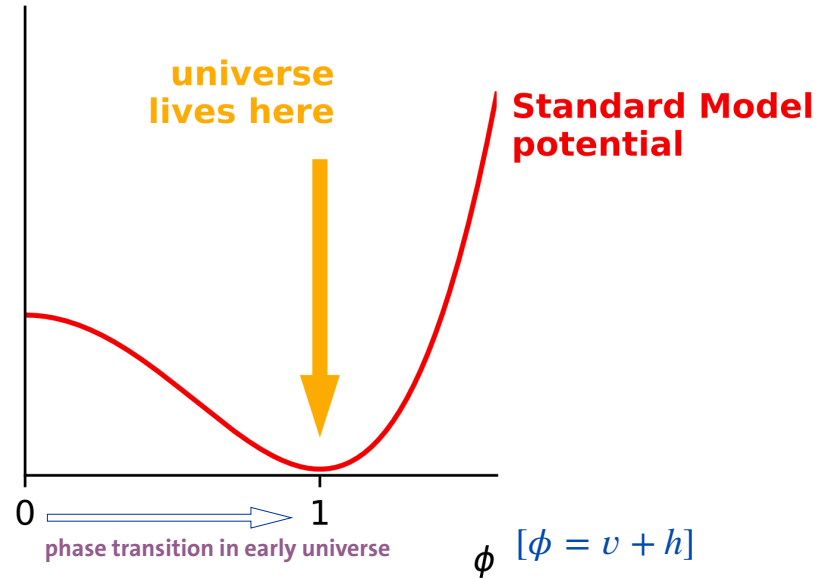
$$V_{SM}(\phi) = \mu^2 |\phi|^2 - \lambda |\phi|^4 - V_0$$

Higgs mass term

Quartic Higgs coupling
UV behaviour, vacuum stability

$$V(\phi) \supset \frac{1}{2} m_h^2 h^2 + \frac{1}{3!} \kappa_\lambda \left(\frac{3m_h^2}{v} \right) h^3 + \frac{1}{4!} \kappa_4 \left(\frac{3m_h^2}{v^2} \right) h^4 + \dots$$

$V(\phi)$, SM



Does the Higgs potential look like a Mexican hat?

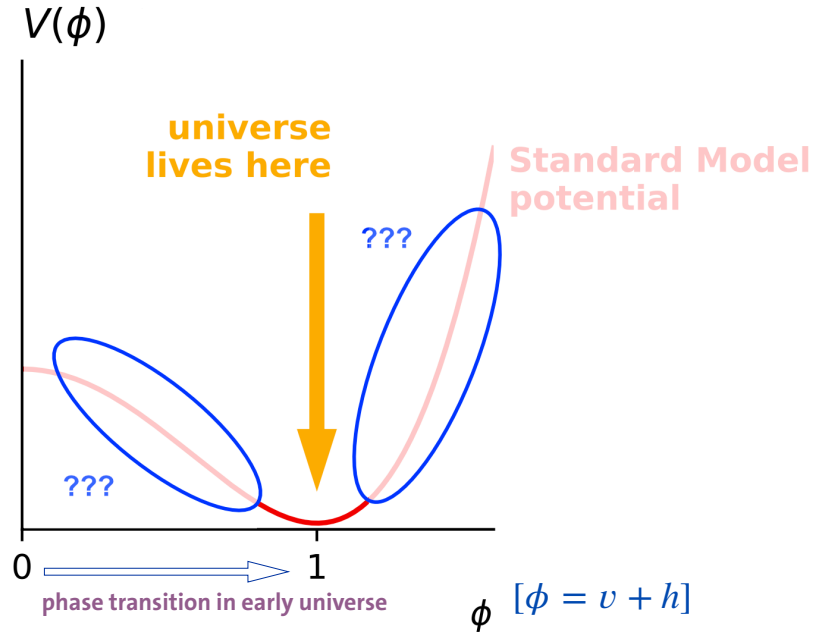
Source: "Advanced Course on Higgs Physics", J. Braathen 2024

$$V_{SM}(\phi) = \mu^2 |\phi|^2 - \lambda |\phi|^4 - V_0$$

Higgs mass term

Quartic Higgs coupling
UV behaviour, vacuum stability

$$V(\phi) \supset \frac{1}{2} m_h^2 h^2 + \frac{1}{3!} \kappa_\lambda \left(\frac{3m_h^2}{v} \right) h^3 + \frac{1}{4!} \kappa_4 \left(\frac{3m_h^2}{v^2} \right) h^4 + \dots$$



Does the Higgs potential look like a Mexican hat?

Source: "Advanced Course on Higgs Physics", J. Braathen 2024

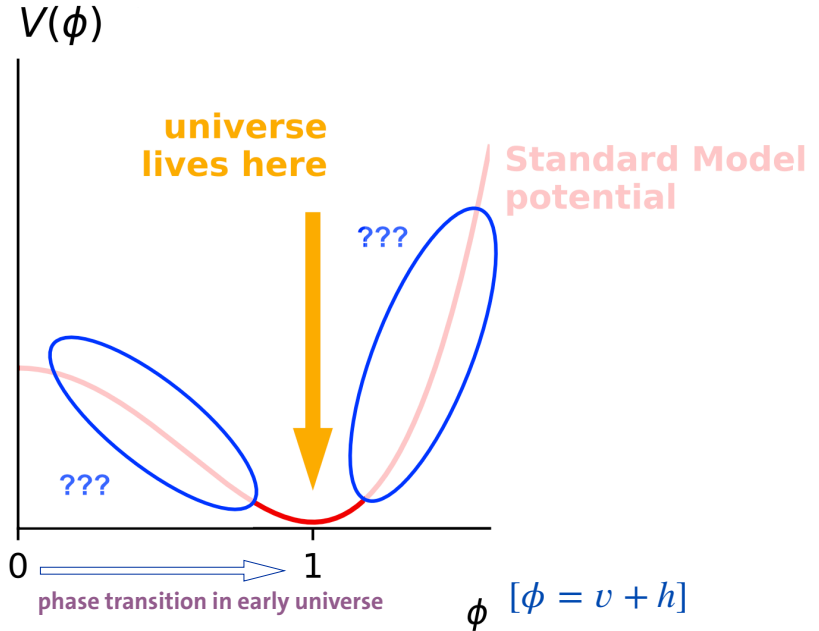
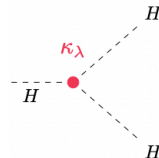
$$V_{SM}(\phi) = \mu^2 |\phi|^2 - \lambda |\phi|^4 - V_0$$

Higgs mass term

Quartic Higgs coupling
UV behaviour, vacuum stability

$$V(\phi) \supset \frac{1}{2} m_h^2 h^2 + \frac{1}{3!} \kappa_\lambda \left(\frac{3m_h^2}{v} \right) h^3 + \frac{1}{4!} \kappa_4 \left(\frac{3m_h^2}{v^2} \right) h^4 + \dots$$

λ_{HHH} predicts strength of HHH interaction and determines shape of potential



Does the Higgs potential look like a Mexican hat?

Source: "Advanced Course on Higgs Physics", J. Braathen 2024

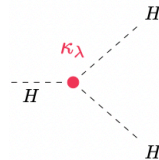
$$V_{SM}(\phi) = \mu^2 |\phi|^2 - \lambda |\phi|^4 - V_0$$

Higgs mass term

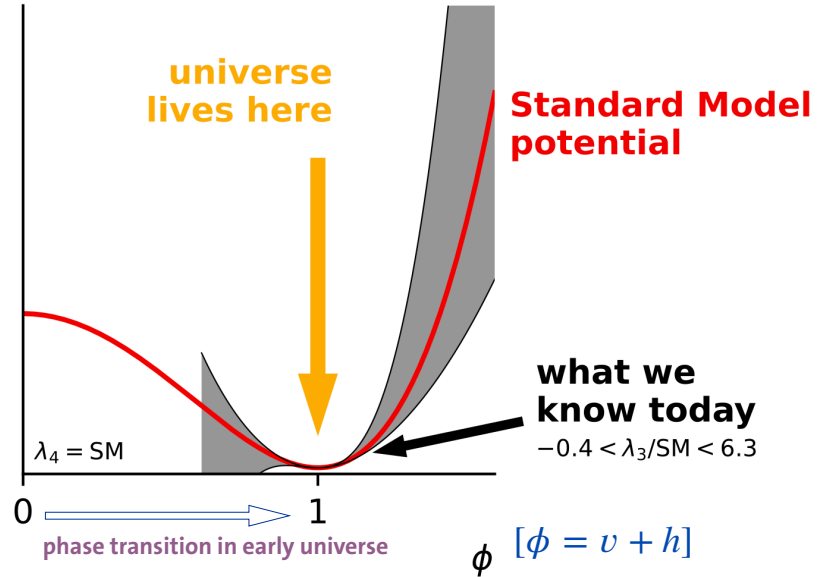
Quartic Higgs coupling
UV behaviour, vacuum stability

$$V(\phi) \supset \frac{1}{2} m_h^2 h^2 + \frac{1}{3!} \kappa_\lambda \left(\frac{3m_h^2}{v} \right) h^3 + \frac{1}{4!} \kappa_4 \left(\frac{3m_h^2}{v^2} \right) h^4 + \dots$$

λ_{HHH} predicts strength of HHH interaction and determines shape of potential



$V(\phi)$, today



Does the Higgs potential look like a Mexican hat?

Source: "Advanced Course on Higgs Physics", J. Braathen 2024

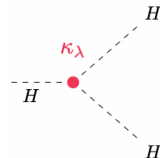
$$V_{SM}(\phi) = \mu^2 |\phi|^2 - \lambda |\phi|^4 - V_0$$

Higgs mass term

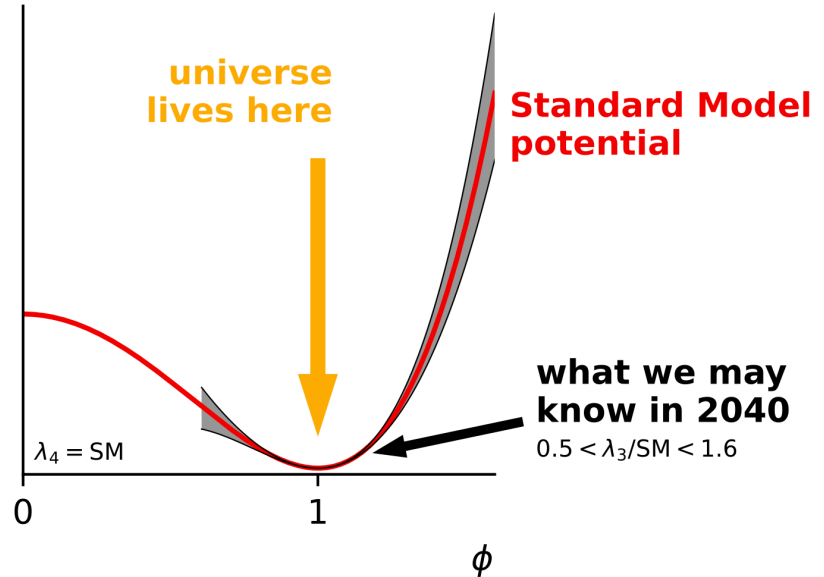
Quartic Higgs coupling
UV behaviour, vacuum stability

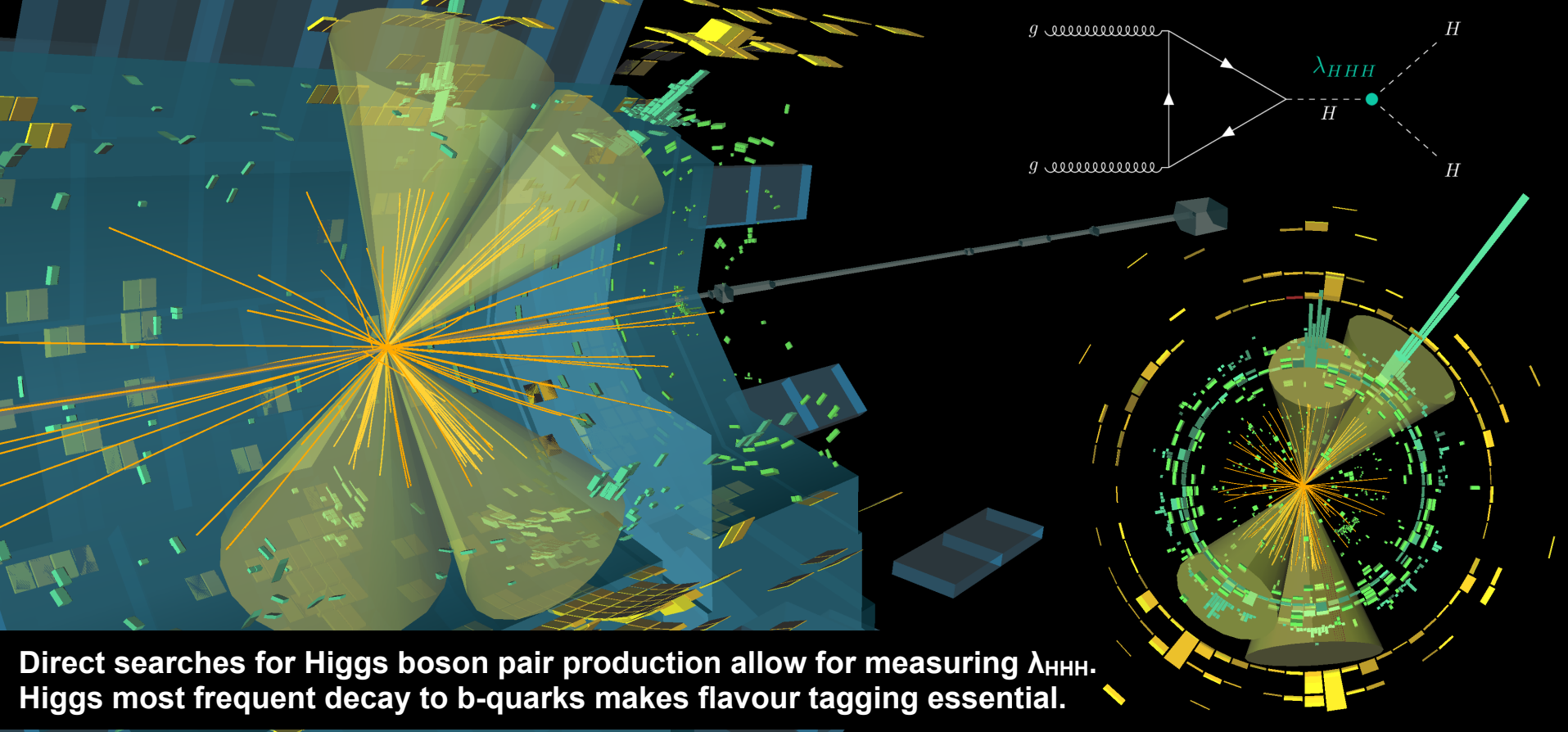
$$V(\phi) \supset \frac{1}{2} m_h^2 h^2 + \frac{1}{3!} \kappa_\lambda \left(\frac{3m_h^2}{v} \right) h^3 + \frac{1}{4!} \kappa_4 \left(\frac{3m_h^2}{v^2} \right) h^4 + \dots$$

λ_{HHH} predicts strength of HHH interaction and determines shape of potential



$V(\phi)$, 2040 (HL-LHC)

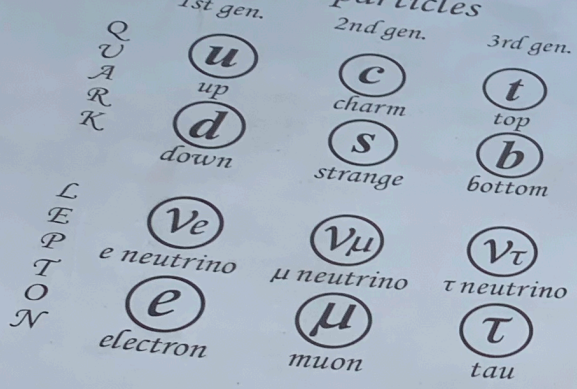




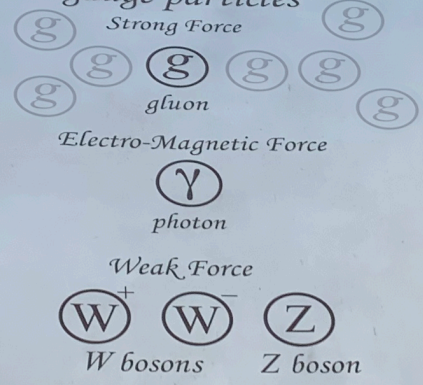
Handwritten signature

Standard Model

matter particles



gauge particles



scalar particle H
Higgs

Standard Model
Lagrangian
$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{D} \psi + \psi^\dagger \not{D} \psi + \text{h.c.}$$

Are there additional Higgs bosons?

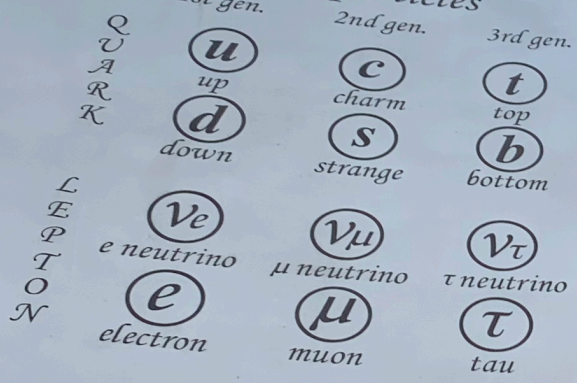
Wandering the Immeasurable
Pérégrinations à l'infini

Handwritten signature

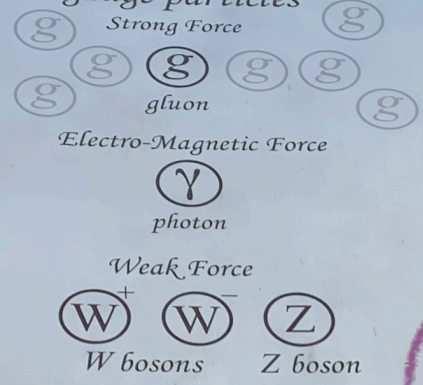
Handwritten signature

Standard Model

matter particles



gauge particles



scalar particle **H**
Higgs

Standard Model
Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{D} \psi + \text{h.c.}$$

Beyond the
Standard Model
(BSM)?

Are there additional Higgs bosons?

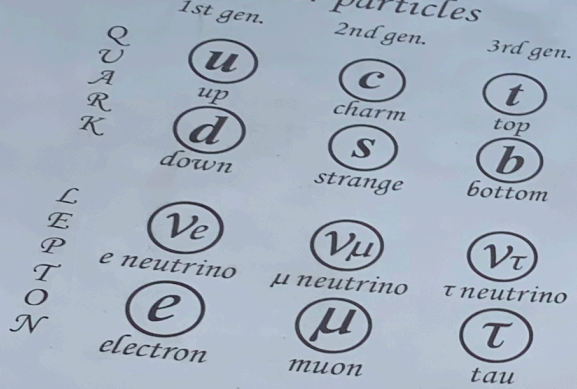
Wandering the Immeasurable
Pérégrinations à l'infini

Handwritten signature

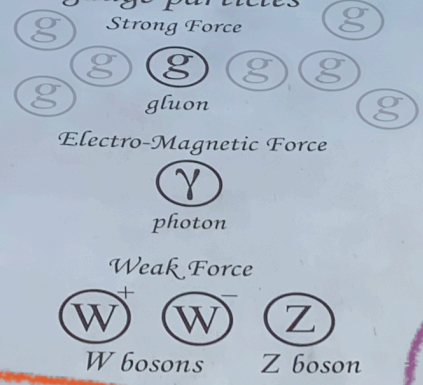
Handwritten signature

Standard Model

matter particles



gauge particles



Standard Model Lagrangian

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{D} \psi + \text{h.c.}$$


Beyond the Standard Model (BSM)?

Are there additional Higgs bosons?

Wandering the Immeasurable
Pérégrinations à l'infini

Handwritten signature

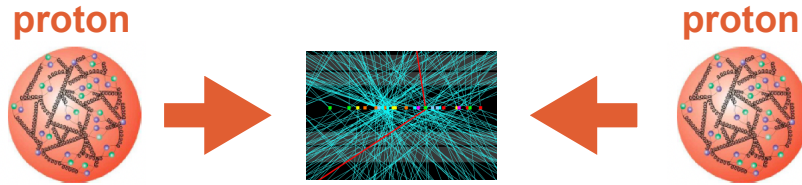
Some open questions

- How does the Higgs potential look like?
- What is the origin of the vast range of quark and lepton masses?
 - Are there modified interactions to the Higgs boson and known particles?
- What stabilises the Higgs mass versus high-scale new physics?
 - Are there new particles close to the mass of the Higgs boson?
- Research problems:
 1. Measuring the Higgs Charm Yukawa Coupling
 2. Direct searches for Higgs boson pair production
 3. Searches for additional Higgs bosons

The Large Hadron Collider – a discovery machine

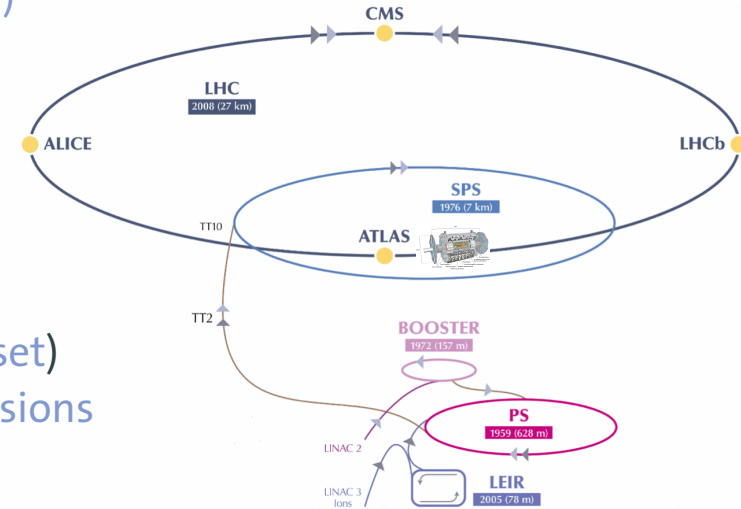
High-energy (E) particle collisions enable

- Probing the structure of matter (resolution $\sim 1/E$)
- Discovery of new and massive particles ($E = mc^2$)
- Investigating the conditions of the early universe ($E = kT$)



Large Hadron Collider

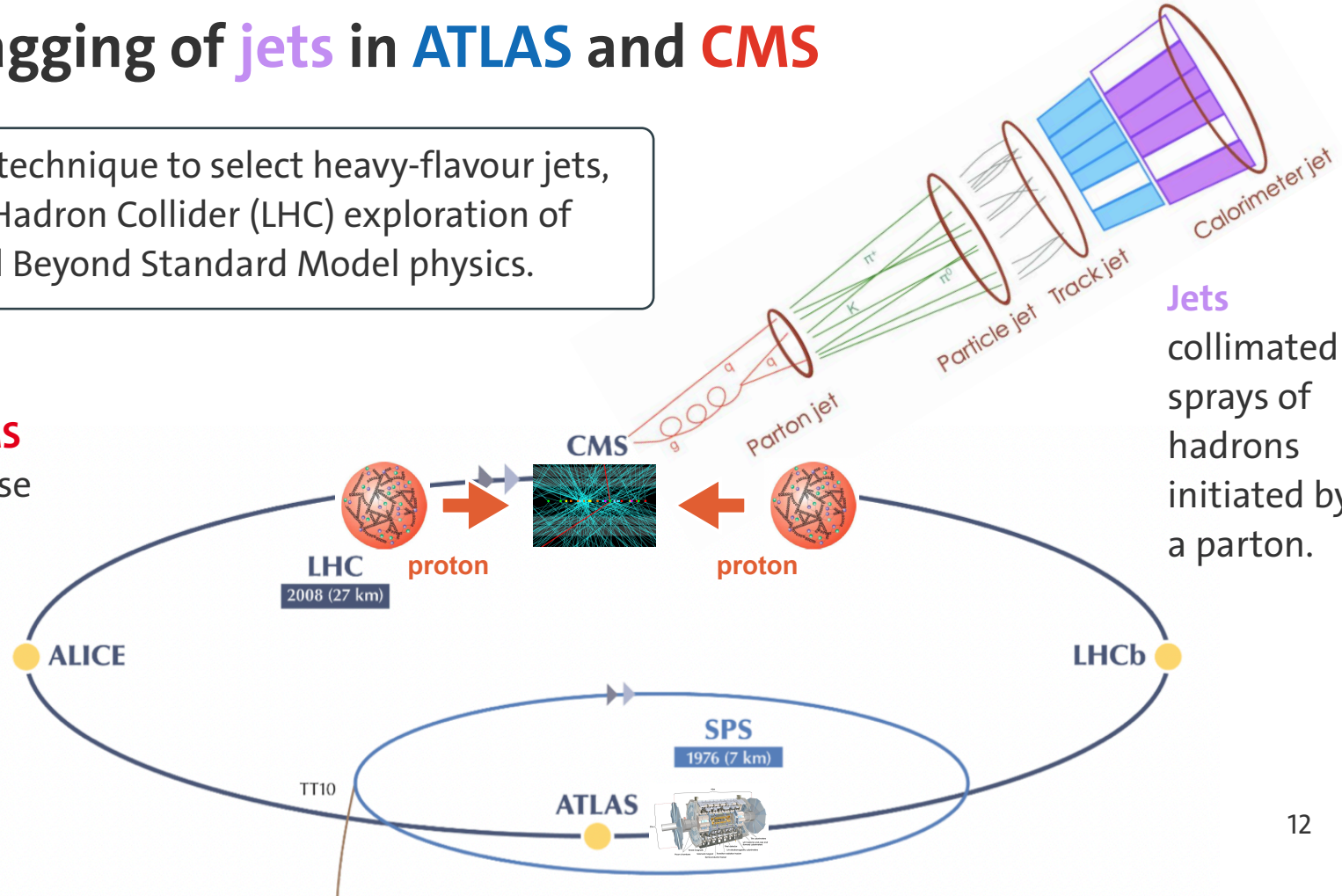
- Proton-proton (pp) collisions at 13 TeV (Run 2, large dataset) and 13.6 TeV (Run 3, larger dataset), also pA and AA collisions
- Four large experiments, today focus on ATLAS and CMS



Flavour tagging of jets in ATLAS and CMS

experimental technique to select heavy-flavour jets, driving Large Hadron Collider (LHC) exploration of Higgs, Top and Beyond Standard Model physics.

ATLAS and **CMS**
general purpose
experiments
at the LHC.

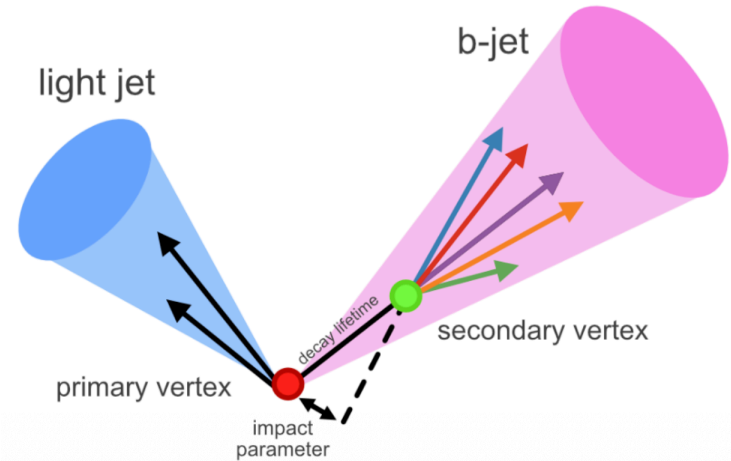
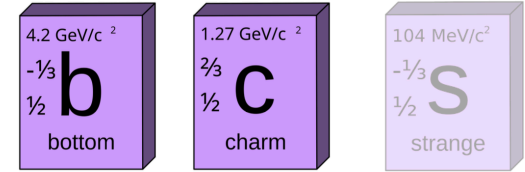


Jets
collimated
sprays of
hadrons
initiated by
a parton.

Flavour tagging

We can distinguish jets produced by heavy quarks (b-jets / c-jets) from other jets (light jets) by exploiting the distinct properties of b and c hadrons.

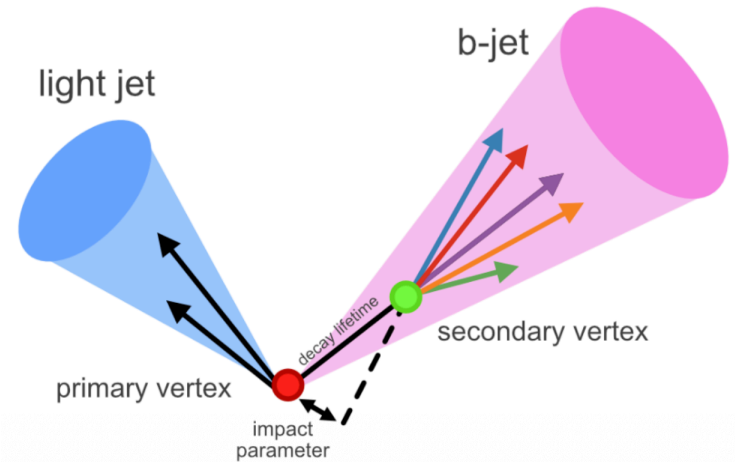
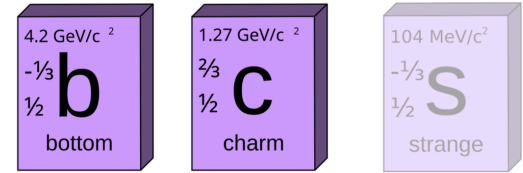
- Long **lifetime** (b: $\sim 1.5\text{ps}$ / c: $\sim 0.6\text{ps}$) results in displaced decays with **secondary vertex**
- High **mass** (compared to other hadrons)
- High **decay multiplicity** and **hard fragmentation**
- “Soft” **charged leptons** in displaced decay



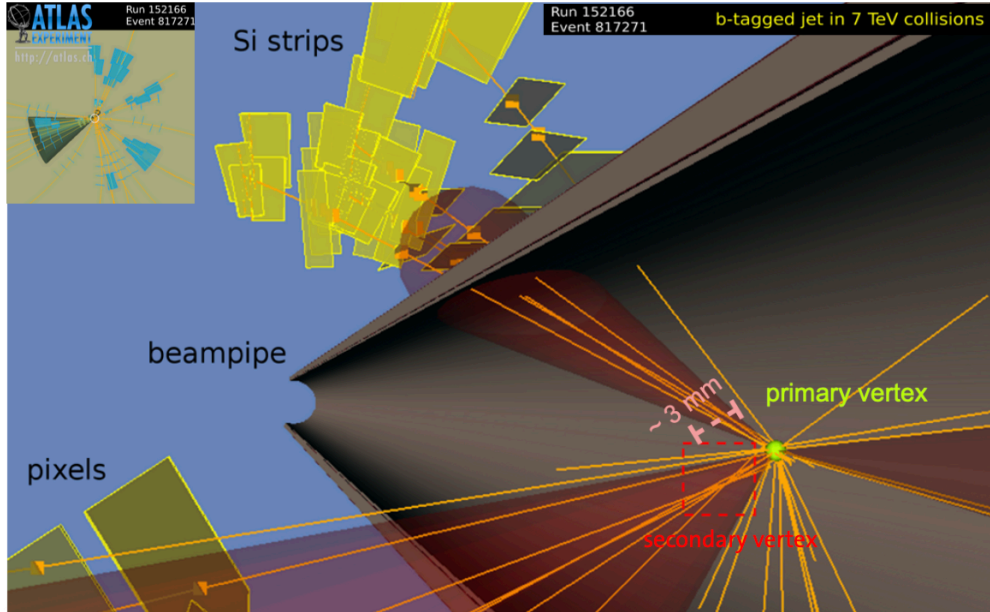
Flavour tagging

We can distinguish jets produced by heavy quarks (b-jets / c-jets) from other jets (light jets) by exploiting the distinct properties of b and c hadrons.

- Long **lifetime** (b: $\sim 1.5\text{ps}$ / c: $\sim 0.6\text{ps}$) results in displaced decays with **secondary vertex**
- High **mass** (compared to other hadrons)
- High **decay multiplicity** and **hard fragmentation**
- “Soft” charged leptons in displaced decay

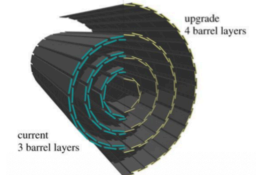
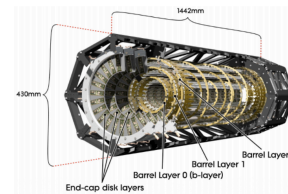


Tracking detectors enable flavour tagging



Displaced decay usually happens in beam pipe and results in **secondary vertex** with **tracks** that have **large impact parameters**.

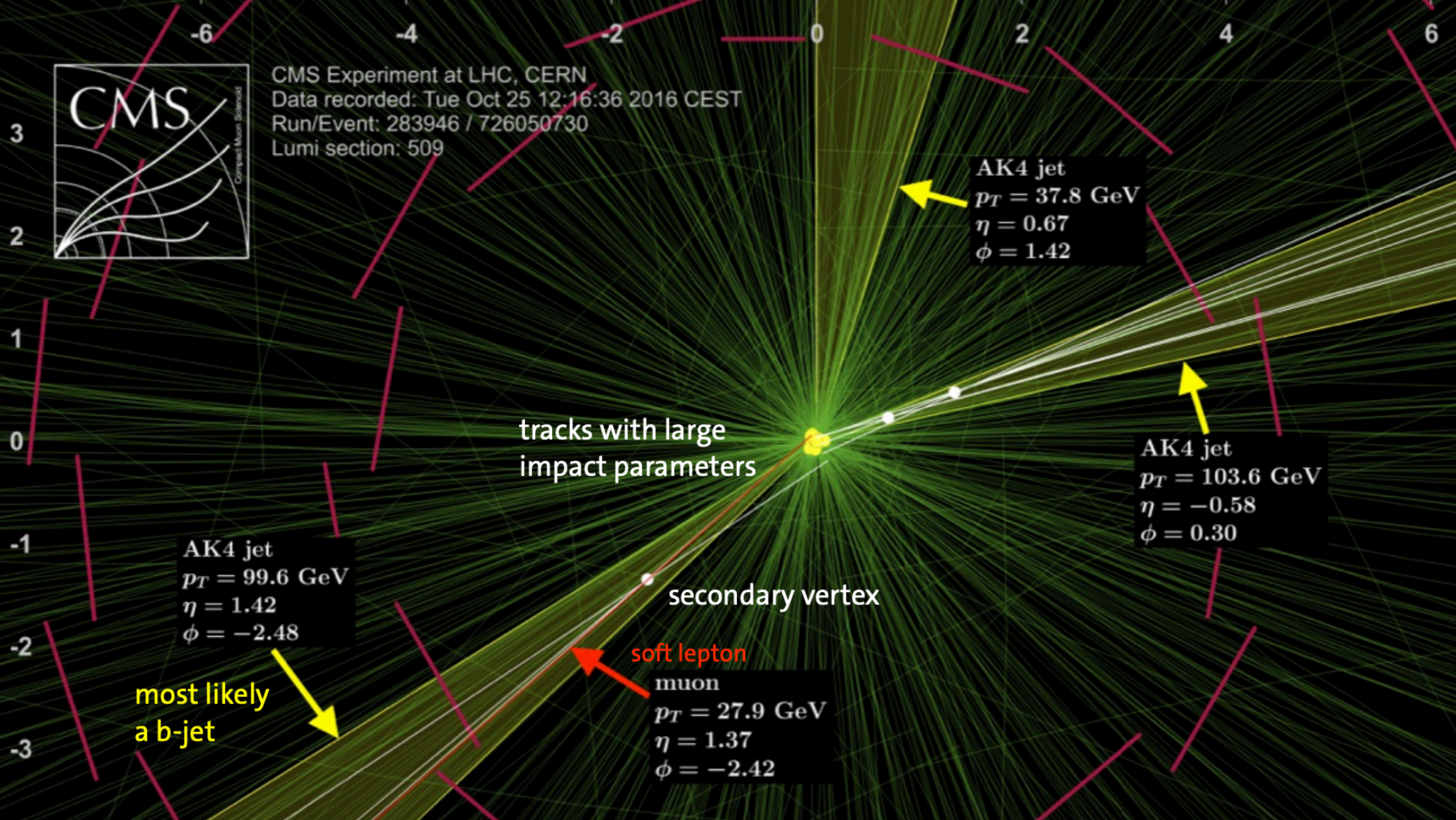
Measurement of track impact parameters and reconstruction of secondary vertices requires **excellent tracking detectors** located close to the beam pipe.



Challenge at end of pixel detector lifetime:
tracking radiation damage can degrade
flavour tagging performance



CMS Experiment at LHC, CERN
Data recorded: Tue Oct 25 12:16:36 2016 CEST
Run/Event: 283946 / 726050730
Lumi section: 509



AK4 jet
 $p_T = 37.8 \text{ GeV}$
 $\eta = 0.67$
 $\phi = 1.42$

AK4 jet
 $p_T = 103.6 \text{ GeV}$
 $\eta = -0.58$
 $\phi = 0.30$

AK4 jet
 $p_T = 99.6 \text{ GeV}$
 $\eta = 1.42$
 $\phi = -2.48$

most likely
a b-jet

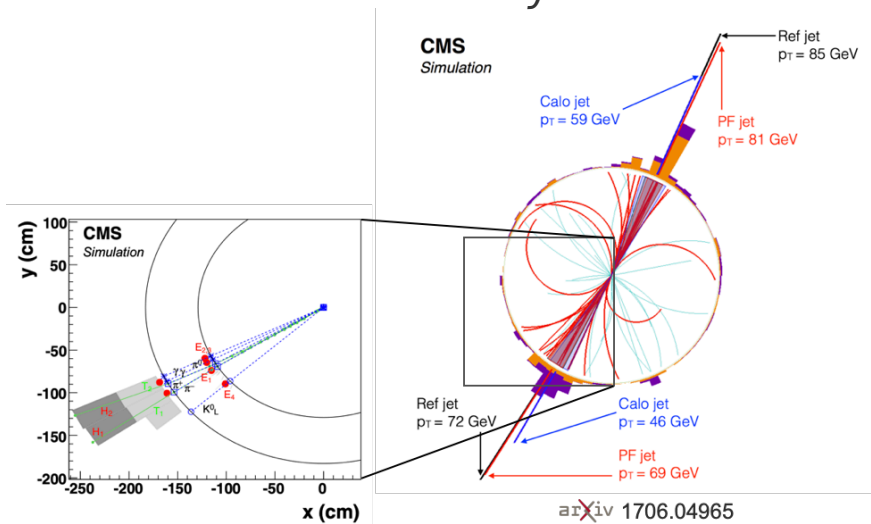
tracks with large
impact parameters

secondary vertex

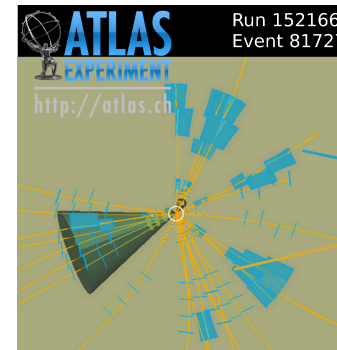
soft lepton
muon
 $p_T = 27.9 \text{ GeV}$
 $\eta = 1.37$
 $\phi = -2.42$

Ingredients: jets and **particle flow candidates** / **tracks**

- **CMS** particle flow event reconstruction: holistic approach by correlating basic elements from all detector layers.



ATLAS particle flow jets: reconstruction of jets based on combination of inner tracking detector tracks and calorimeter clusters.



Calorimeter clusters

Tracks

Particle Flow algorithm

Charged PFlow objects

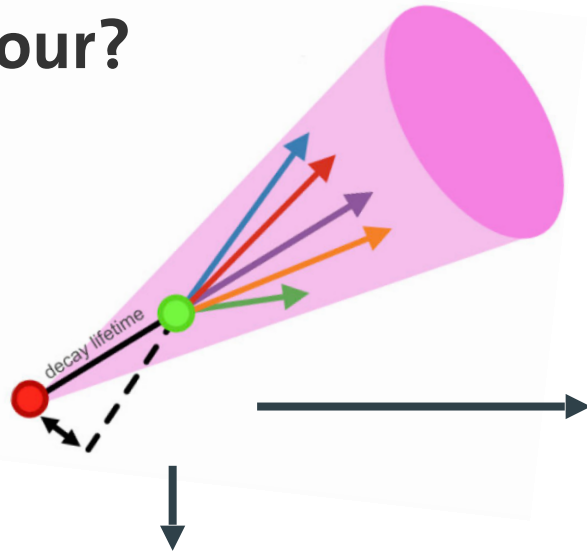
Neutral PFlow objects

Jet finding algorithm

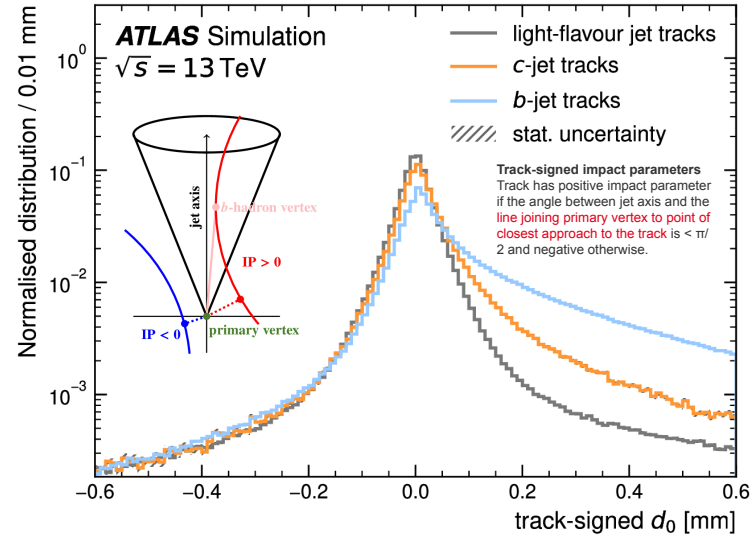
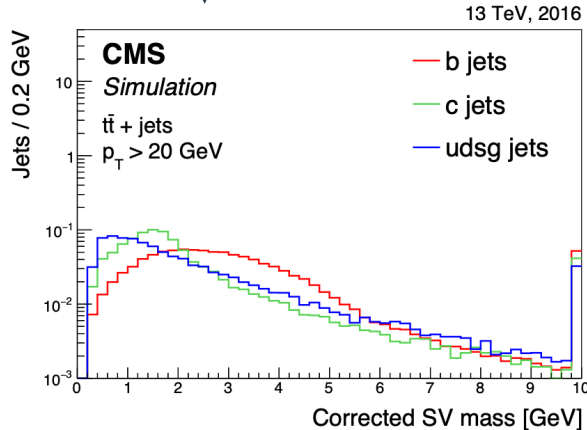
PFlow jets

How to tag flavour?

Charged particle tracks associated with the jet provide features with discrimination power to "tag" the jet flavour.



Secondary Vertex (SV)
The point from where the b or c hadron decays.



Transverse impact parameter d_0

Distance of closest approach of track to the primary vertex in transverse plane.

Signed impact parameters of **b-jets** (**c-jets**) are mostly **large and positive**, while light-flavour jets have values consistent with zero within resolution.

Machine learning

- No single property of heavy-flavour jets can provide optimal performance:
 - Track impact parameters
 - Reconstructed secondary vertices (flight length, mass, energy fraction, track multiplicity)
 - Soft charged leptons produced in decay of b/c hadrons
- Therefore, **combine all available information** in a **multivariate discriminant**.
 - Use machine learning tools: boosted decision trees, neural networks, transformers.

History of flavour tagging at LEP ([first report 1991](#)), HERA and Tevatron, already using neural networks ([JETNET 2.0](#) in F77).

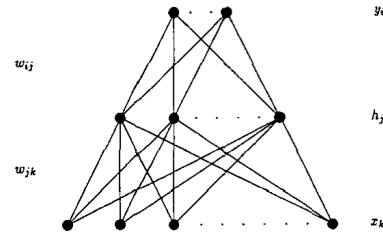


Figure 9. A feed-forward neural network with one layer of hidden units.

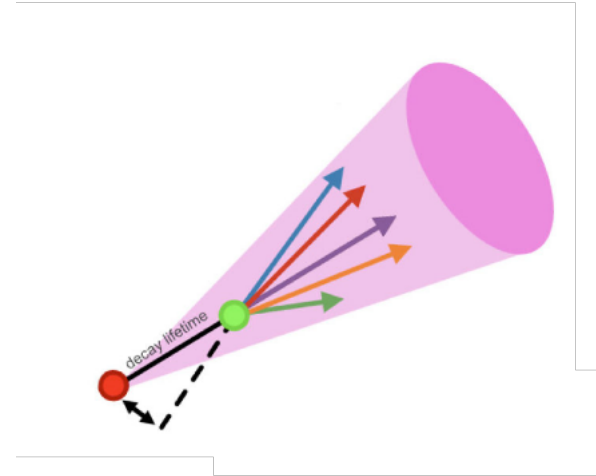
Problem statement

What we have:

- Jet constructed by anti-kt algorithm ($R=0.4$ / large radius) with kinematics (pt, eta, phi, mass)
- Collection of tracks and neutral particle flow candidates associated with jet: X_i with $i = \{1, \dots, n\}$
 - Each track has features $X_i \in \mathbb{R}^m$
- Jet has labels from simulation:
 - $Y: \{b, c, \text{light}, (\text{strange})\}$ or $\{bb, cc, \text{top}, \text{QCD}\}$

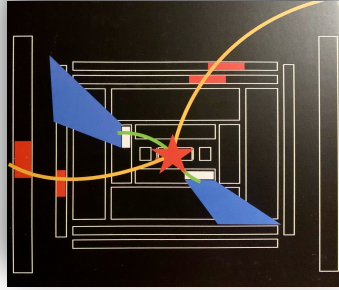
What we want: $p(Y | X_1, \dots, X_n)$

Note: definition of label (“jet flavour”) provides room for discussion. “What is a quark jet?”
ATLAS / CMS define jet flavour by ghost hadron / parton association

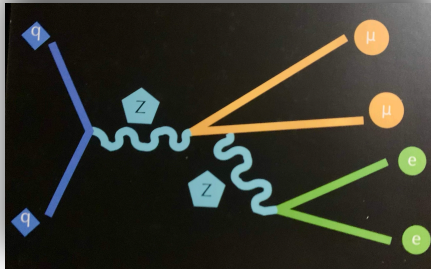


High dimensional problem:
 $n \cdot m \sim O(10^3)$

Data and analysis chain



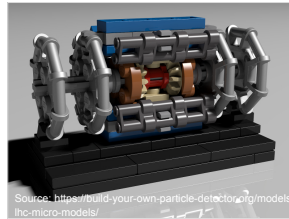
Detector



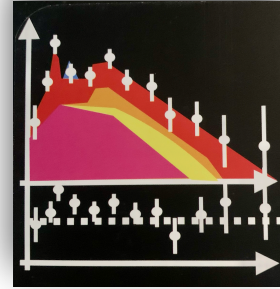
Theory

selection of collision events

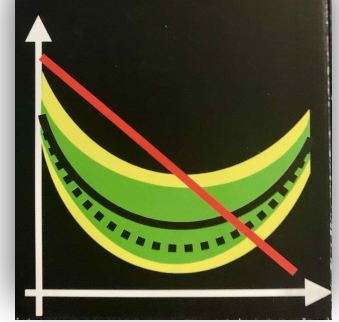
Calibration & alignment



"Digital twin"
(Simulation)



Reconstruction &
Data reduction



Statistical analysis

Simulation provides high-fidelity labelled data for training machine learning algorithms.

Source: "Particle Physics for Babies", L. Corpe

History of algorithms

ATLAS

JetProb, IPxD, MV1

MV2c20, DL1, DL1r/d

GN series

Run 1

Run 2

Run 3

Specialised algorithms

BDTs, deep neural networks

Graph networks, transformers



CMS

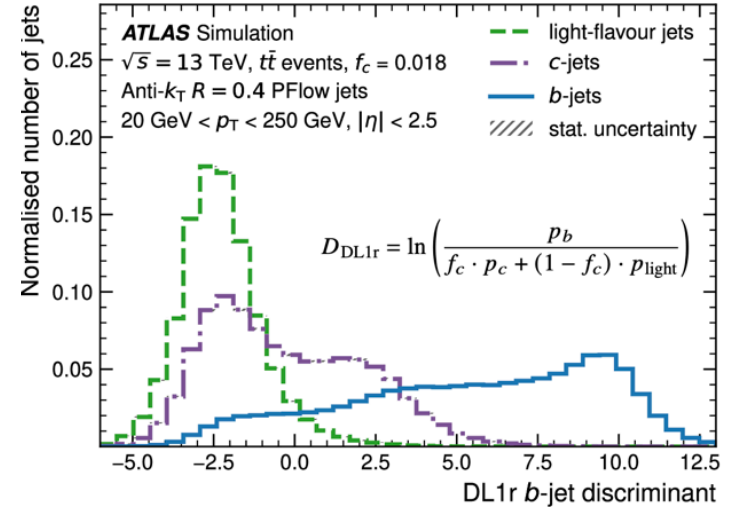
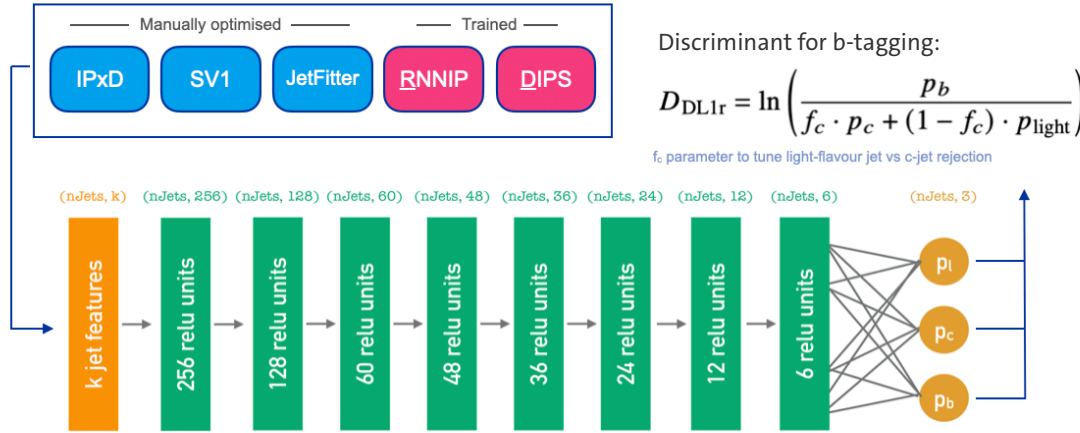
CSV

DeepCSV, DeepJet

ParticleNet, RobustParT, UParT

ATLAS DL1 algorithm

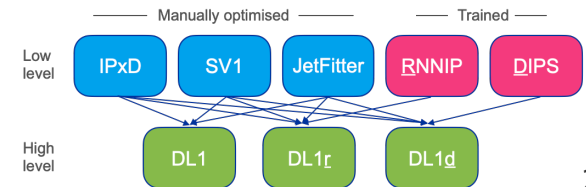
Low-level algorithms with specialised tasks compute inputs for neural network.



Two stage approach

Low-level algorithms use properties (e.g. impact parameters) of individual tracks or reconstruct secondary vertices.

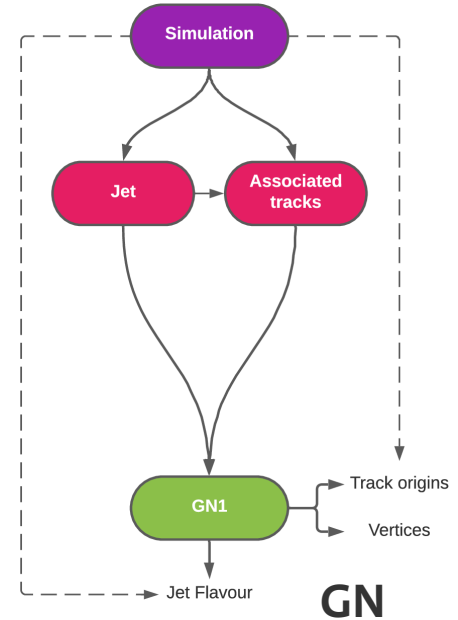
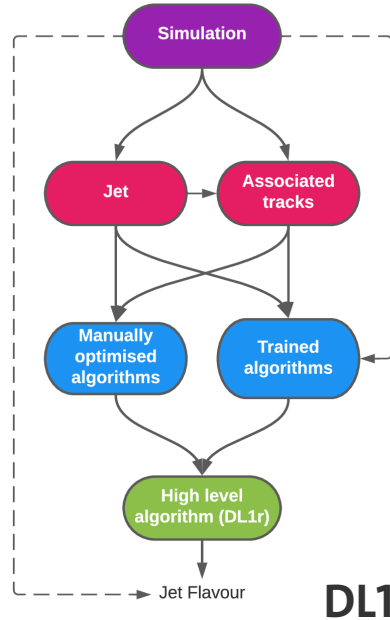
Their outputs are fed into the high-level taggers (deep neural networks) to provide a discriminant.



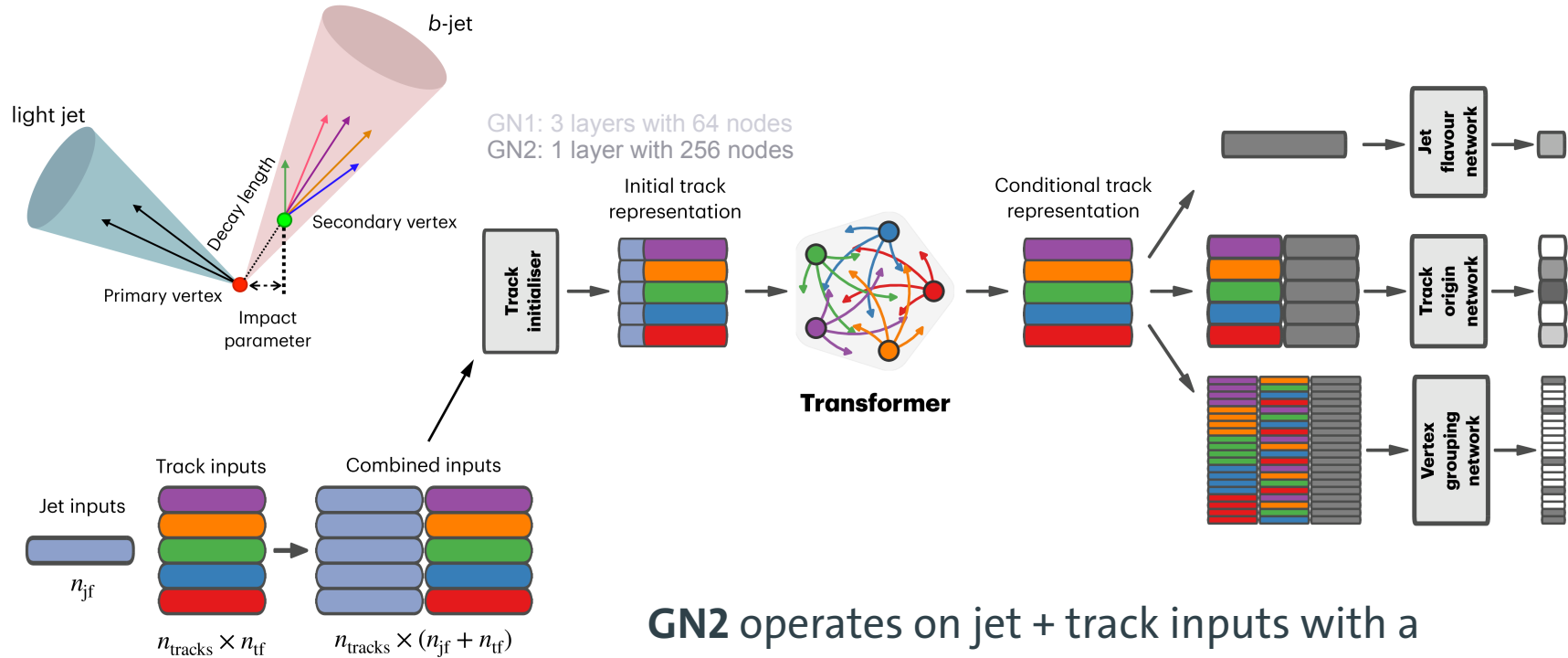
ATLAS GN algorithms

New approach: predict jet flavour with **no separation between low-level and high-level algorithms**, directly taking as inputs

- **jet** p_T and η
- **track** parameters, uncertainties, impact parameters
- **hit** information of tracks

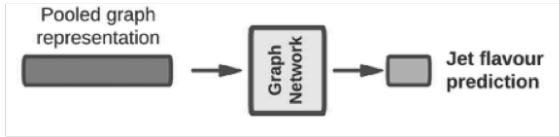
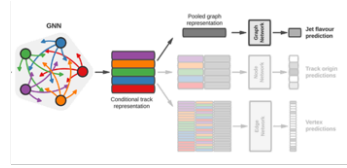


ATLAS GN2



GN2 operates on jet + track inputs with a transformer model using auxiliary tasks.

ATLAS GN2 b/c-tagging discriminant



p_b, p_c, p_u, p_{τ}

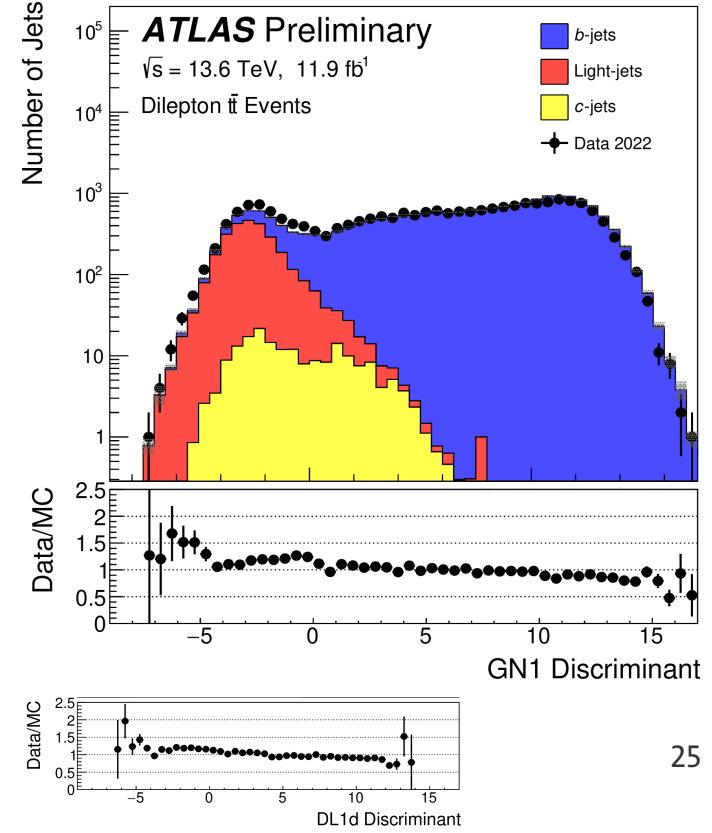
ATLAS approach (different to CMS): combine output probabilities of GN2 in one-dimensional discriminant defined as likelihood ratio

$$D_b = \log \left(\frac{p_b}{f_c p_c + f_{\tau} p_{\tau} + (1 - f_c - f_{\tau}) p_u} \right)$$

Straightforward re-definition of **discriminant for c-tagging** D_c by swapping p_b and p_c .

Similar quality of data/MC agreement (before calibration) as DL1 series

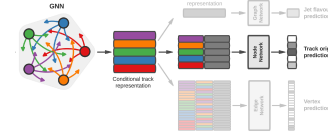
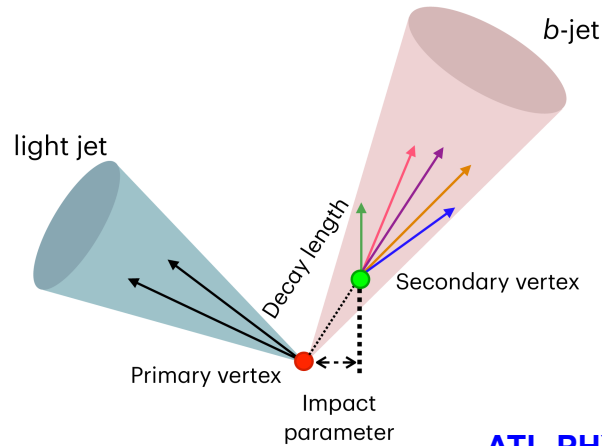
FTAG-2023-01



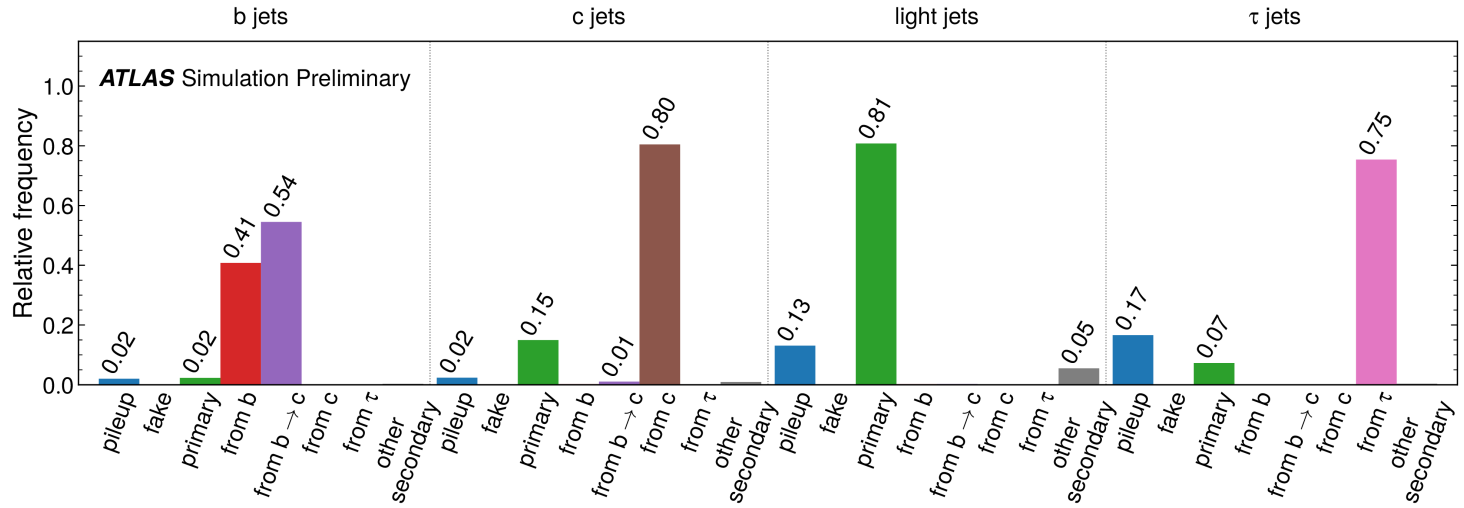
ATLAS GN2 auxiliary tasks

The GN series simultaneously trains jet classification with **auxiliary tasks** for

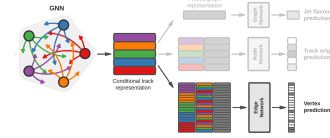
- prediction of the underlying physics process of each **track's origin**



[ATL-PHYS-PUB-2025-029](#)



ATLAS GN2 auxiliary tasks



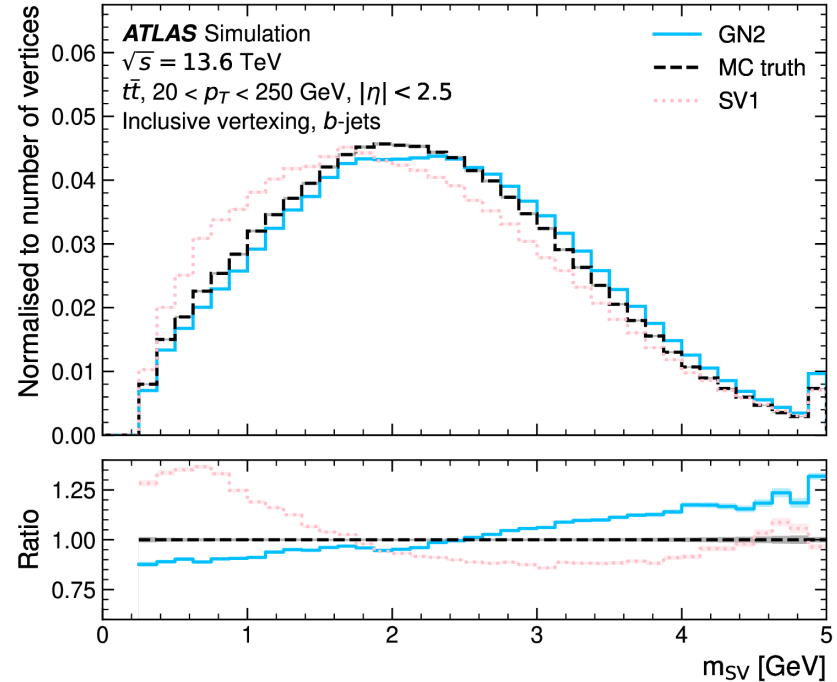
The GN series simultaneously trains jet classification with **auxiliary tasks** for

- prediction of the underlying physics process of each **track's origin**,
- grouping of tracks originating from a **common vertex**.

Total loss function is sum of loss functions for jet flavour categorisation and auxiliary task loss functions.

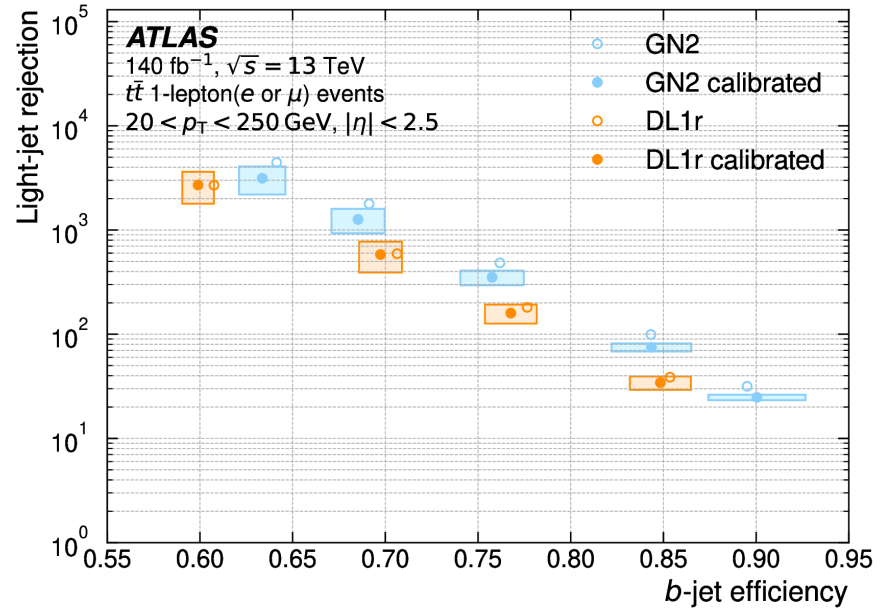
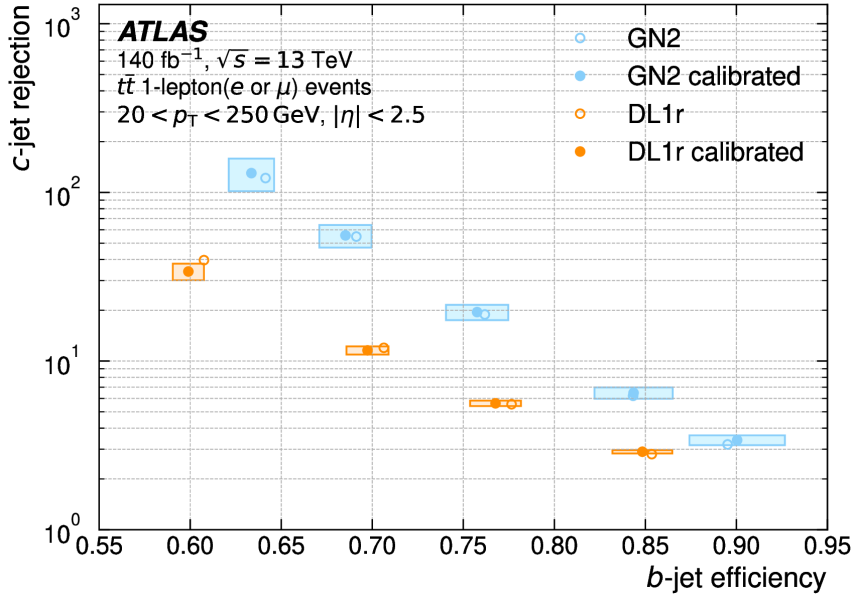
$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}}$$

[Nature Commun. 17 \(2026\) 541](#)



GN2 vertex reconstruction at least as good as dedicated vertex finding algorithm (SV1)

ATLAS GN2 algorithm performance



GN2 at 70% b-efficiency

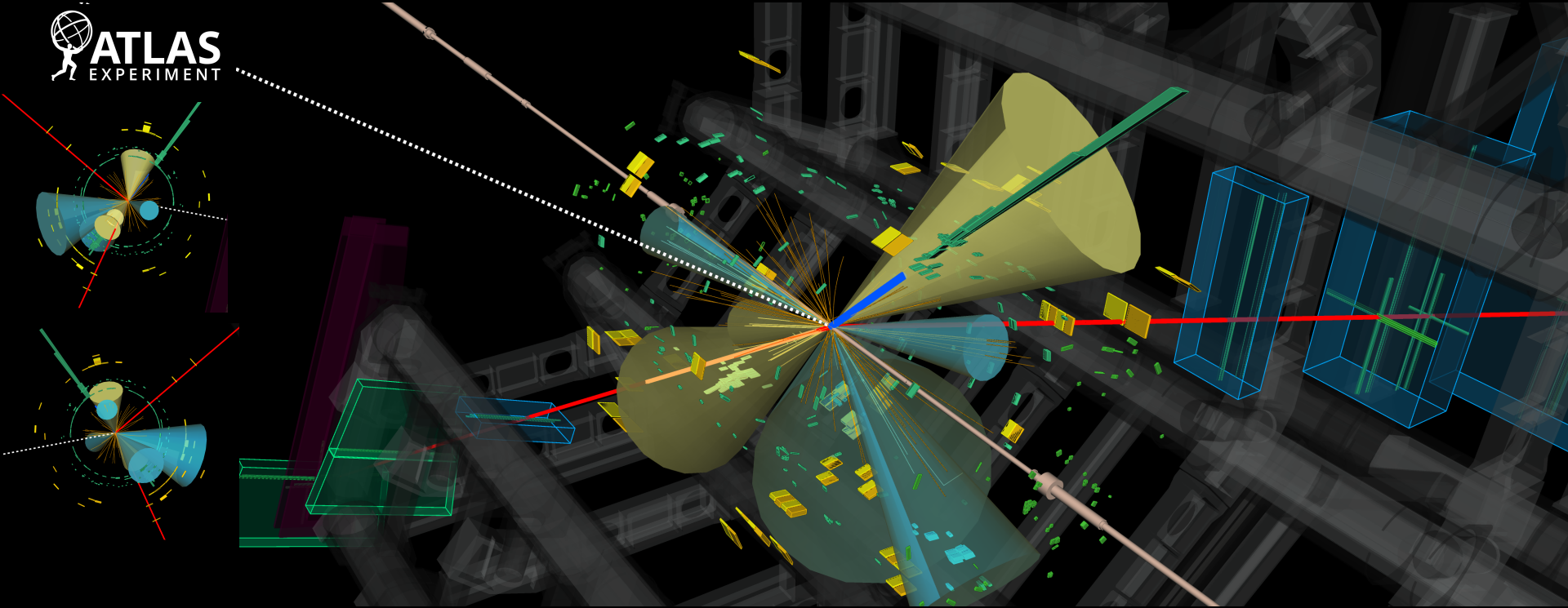
charm jet rejection

50

light-flavour jet rejection

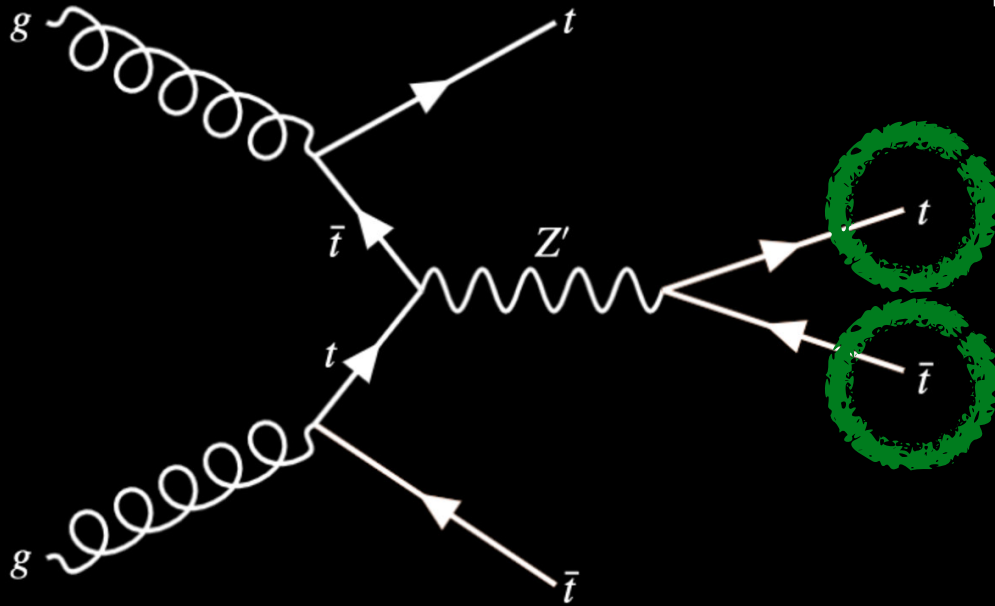
1500

Possible applications for GN2



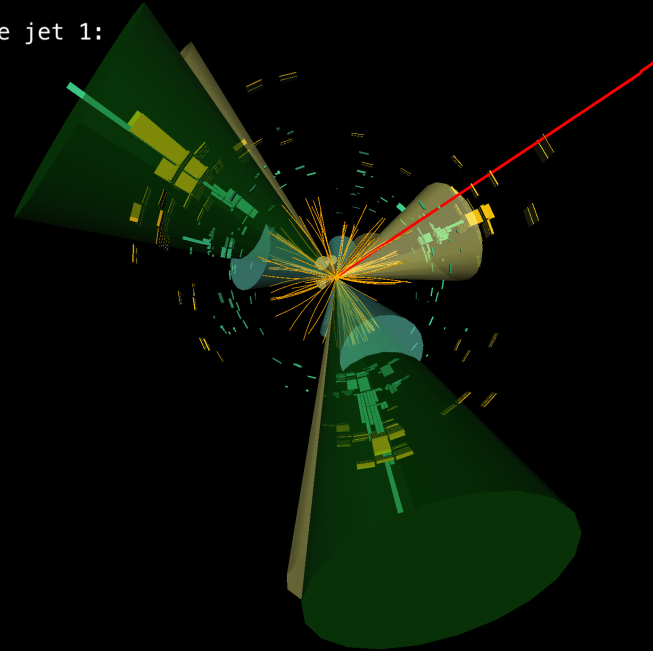
Very rare processes (SM $t\bar{t}t\bar{t}$: ~ 1800 expected in Run 2 dataset) with challenging complex final states.

Possible applications for GN2



Searches for new particles in events with four or three top quarks.

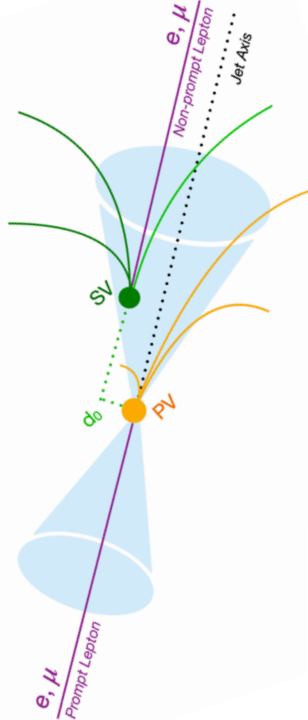
top candidate jet 1:
 $p_T = 463$ GeV
 $E^* = 538$ GeV



top candidate jet 2:
 $p_T = 469$ GeV
 $E = 646$ GeV



ATLAS GN2 for lepton isolation

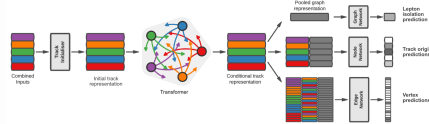


Background in e.g. $t\bar{t}H$ and $t\bar{t}t\bar{t}$ measurements from **non-prompt leptons** originating e.g. in semi-leptonic hadron decays.
 Suppression possible with **isolation** (measure activity around lepton).

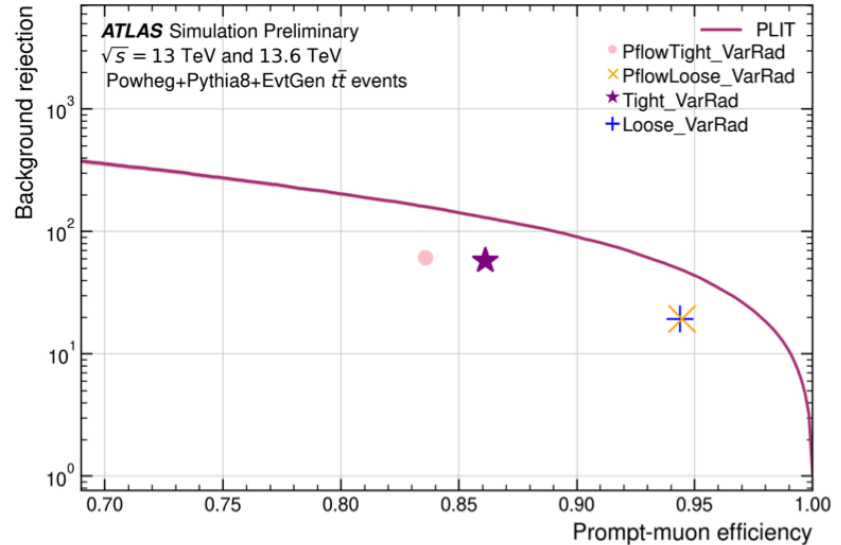
GN2



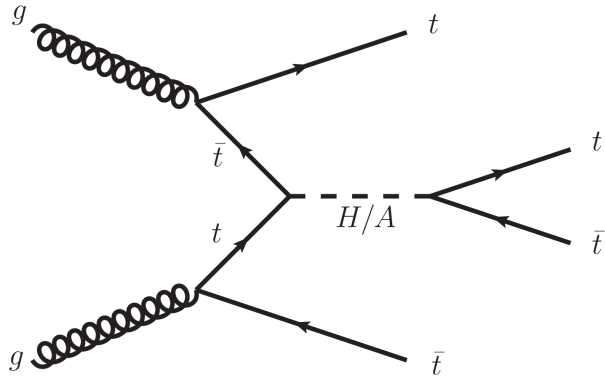
PLIT



Prompt lepton
isolation tagger

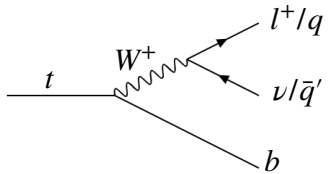


ATLAS GN2 architecture in four-top heavy Higgs searches



Four top quark final state

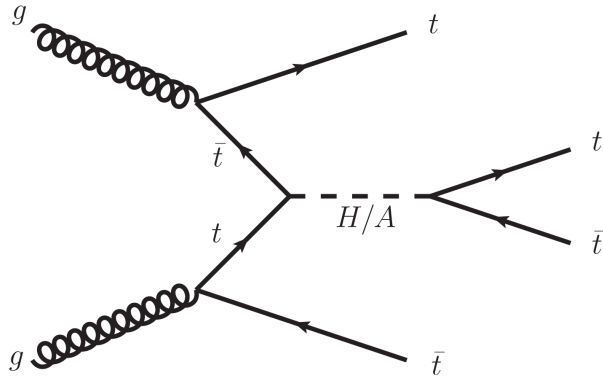
Each top quark decays to **Wb**,



Threefold improvements from multi-modal and multi-task GN2 model:

1. Improved b-tagging with GN2 results in larger b-efficiency at same background rates

ATLAS GN2 architecture in four-top heavy Higgs searches

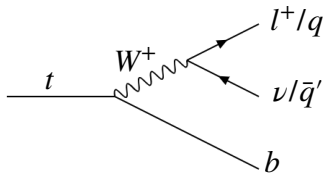


Four top quark final state

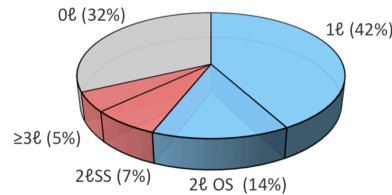
Threefold improvements from multi-modal and multi-task GN2 model:

1. Improved b-tagging with GN2 results in larger b-efficiency at same background rates
2. Once PLIT is commissioned, strong reduction of non-prompt lepton background

Each top quark decays to **Wb**,

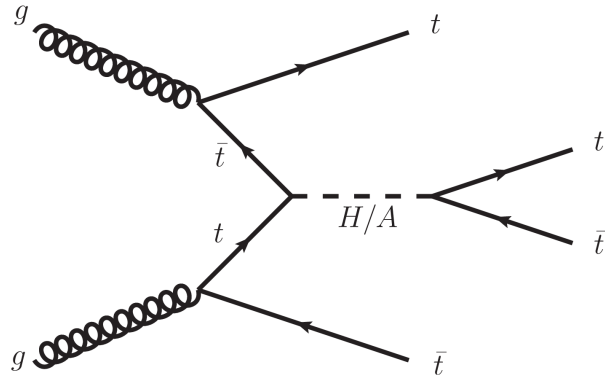


resulting in final state with



- only hadrons (0L)
 - 1 charged lepton (1L)
 - 2 opposite-sign charged leptons
 - 2 same-sign charged leptons
 - 3 or more charged leptons
- largest BR but large background
- small BR but very clean (**most sensitive channels**)

ATLAS GN2 architecture in four-top heavy Higgs searches

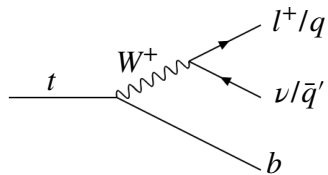


Four top quark final state

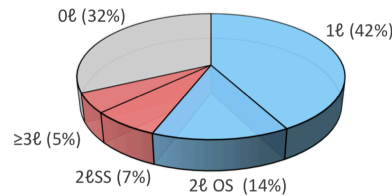
Threefold improvements from multi-modal and multi-task GN2 model:

1. Improved b-tagging with GN2 results in larger b-efficiency at same background rates
2. Once PLIT is commissioned, strong reduction of non-prompt lepton background
3. Use GN2 architecture for event classification

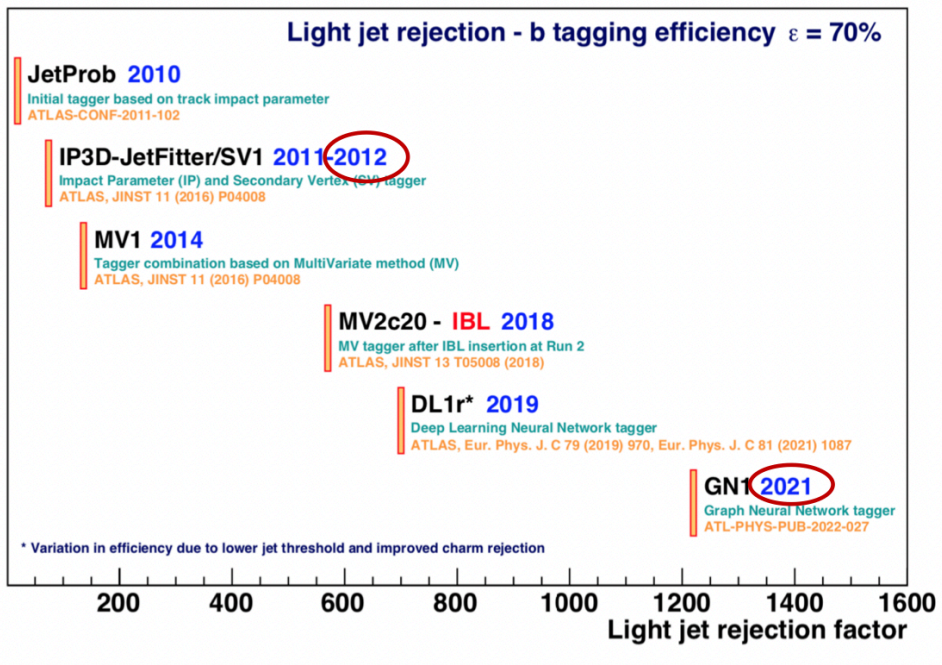
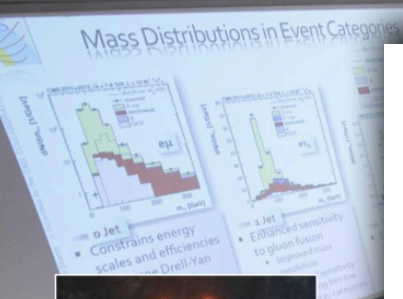
Each top quark decays to **Wb**,



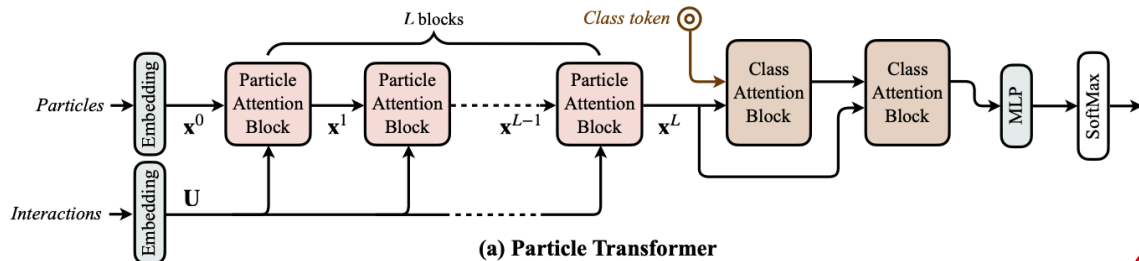
resulting in **final state with**



- only hadrons (0L)
 - 1 charged lepton (1L)
 - 2 opposite-sign charged leptons
 - 2 same-sign charged leptons
 - 3 or more charged leptons
- largest BR but large background
- small BR but very clean (**most sensitive channels**)



CMS Particle Transformer



Most recent **Unified Particle Transformer (UParT)** is based on Particle Transformer ([ParT](#)) architecture.

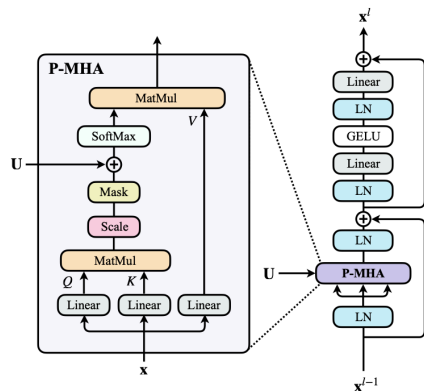
$$\Delta = \sqrt{(y_i - y_j)^2 - (\phi_i - \phi_j)^2}$$

$$k_T = \min(p_{T,i}, p_{T,j}) \Delta$$

$$z = \frac{\min(p_{T,i}, p_{T,j})}{p_{T,i} + p_{T,j}}$$

$$m^2 = (E_i + E_j)^2 - |\vec{p}_i + \vec{p}_j|^2$$

particle interactions

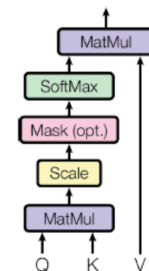


$$P\text{-MHA}(Q, K, V, U) = \text{SoftMax} \left(\frac{QK^T}{\sqrt{d_k}} + U \right) V$$

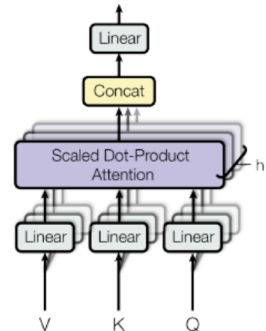
additional term U from particle interactions

[arXiv:1706.03762](https://arxiv.org/abs/1706.03762) [cs.CL]
"Attention Is All You Need"

Scaled Dot-Product Attention



Multi-Head Attention

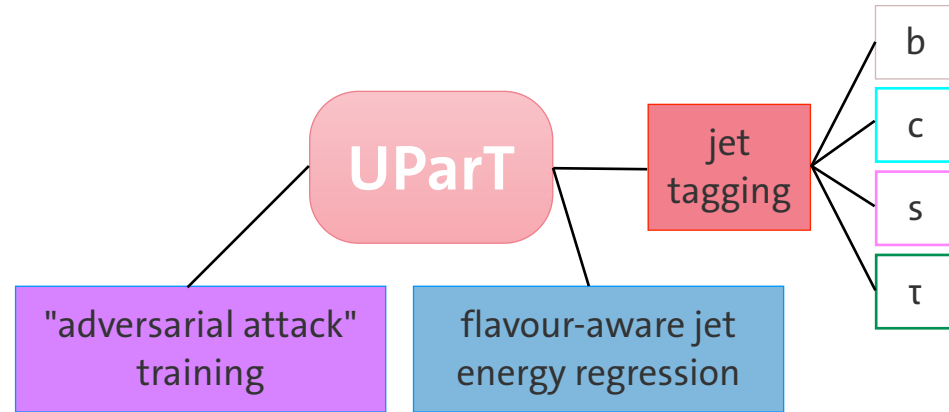


$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

CMS today: Unified Particle Transformer

Unified approach of having a **single model** for classification tasks

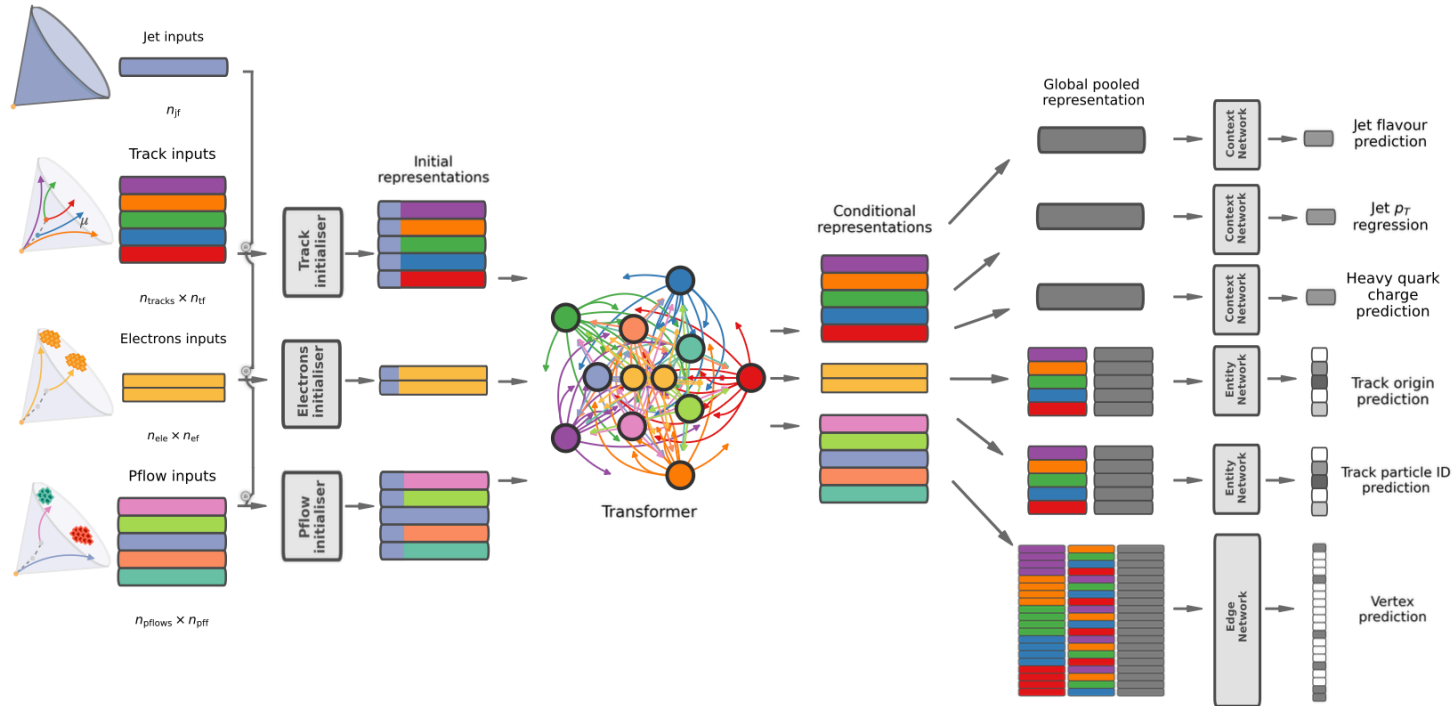
- **b/c tagging, quark- gluon tagging, strange tagging, tau tagging**
- and regression tasks:
- **flavour-aware jet energy correction and resolution**



UParT v2 is a modern version with architecture and training improvements and a bigger training dataset w.r.t. the first version.

UParTv2 at 70% b-efficiency	charm jet rejection	120
	light-flavour jet rejection	2800

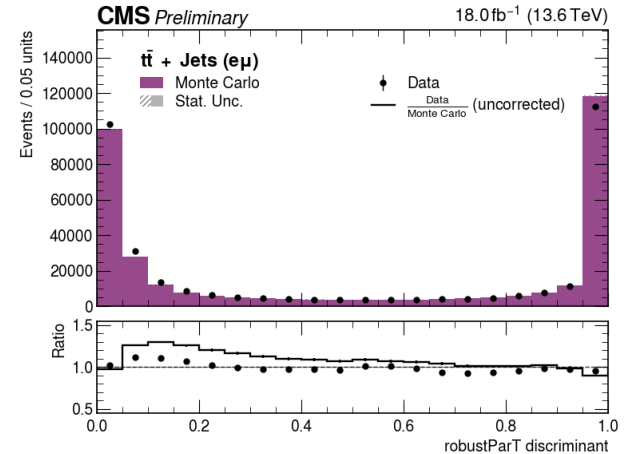
ATLAS GN3 state-of-the-art algorithm



Calibration of b/c-efficiency and mis-tag rates

[DP-2026-022](#)

- Performance of flavour-tagging algorithms is optimized and evaluated using **simulations** (mostly inclusive $t\bar{t}$ events).
- We cannot expect simulations to describe all effects impacting the performance of flavour-tagging algorithms accurately



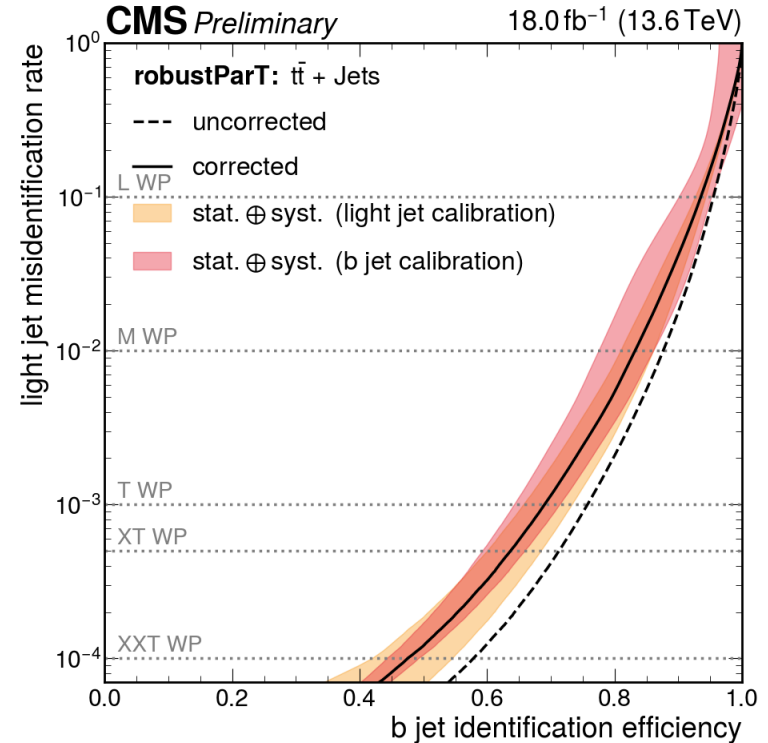
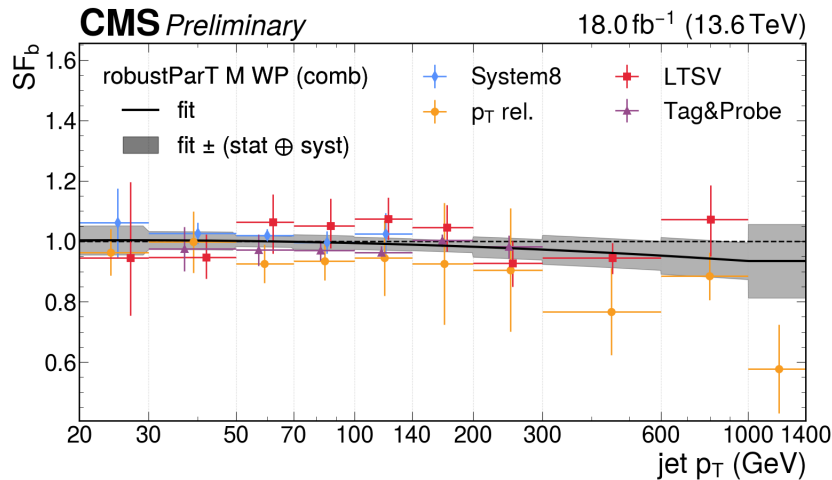
- **We need measurements of the b-/c- and light-jet tagging efficiencies.**
 - Need to extract samples of jets dominated by a single jet flavour
 - Provide results as data-to-simulation scale factors (SF: functions of p_t and η)
 - Assumption: SFs are process independent but not simulation independent

CMS Calibration results

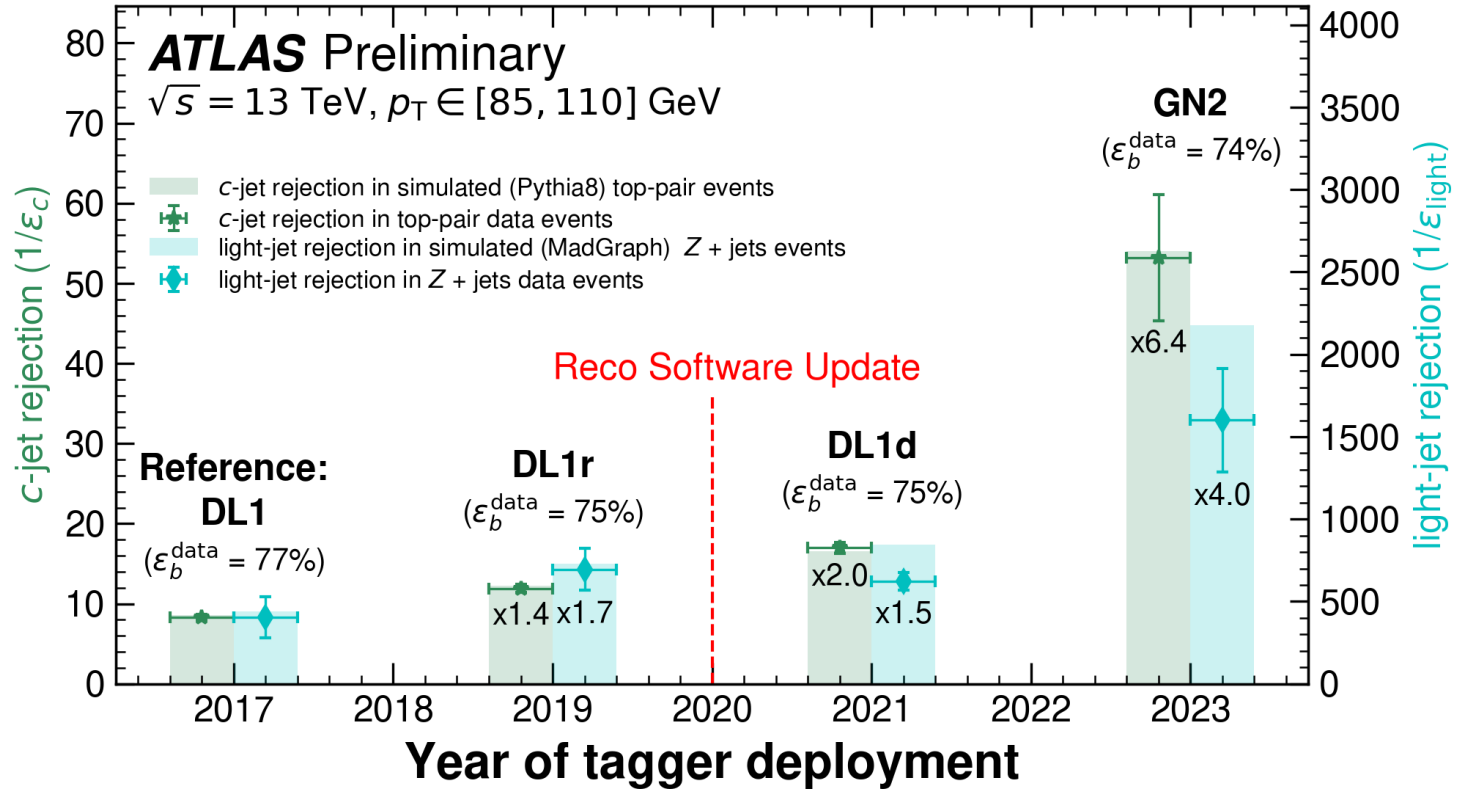
Scale factors
from calibration
measurements

$$SF_f(p_T) = \frac{\epsilon_f^{\text{data}}(p_T)}{\epsilon_f^{\text{MC}}(p_T)}$$

↑
b/c/light



Performance of algorithm
with applied corrections



Current research frontiers

1. Scaling laws
2. Foundation models
3. Charge / fragmentation tagging
4. Continuous calibration

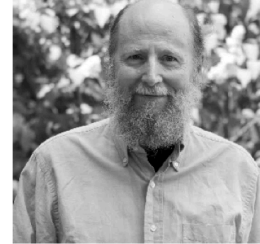
Biased selection! Not talking about: strange tagging, tagging at the trigger-level, tagging of large-radius jets, tagging in datasets with sparse information (TLA/scouting), better robustness (adversarial training with MiniFool), robust jet labelling (flavoured jet algorithms), ...

How to further improve algorithms?

The Bitter Lesson

Rich Sutton

March 13, 2019



The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking

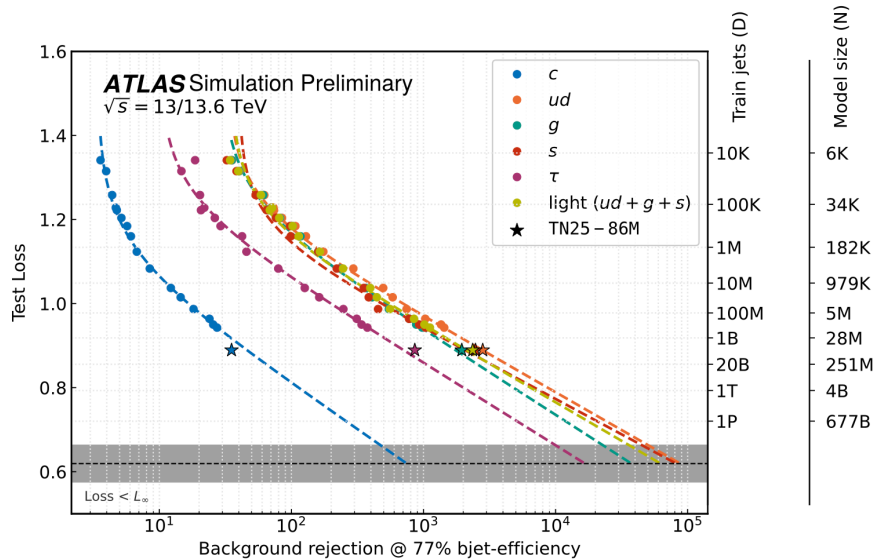
Researchers seek to leverage their human knowledge [...], but the only thing that matters in the long run is the leveraging of computation

There were many examples of AI researchers' belated learning of this bitter lesson.

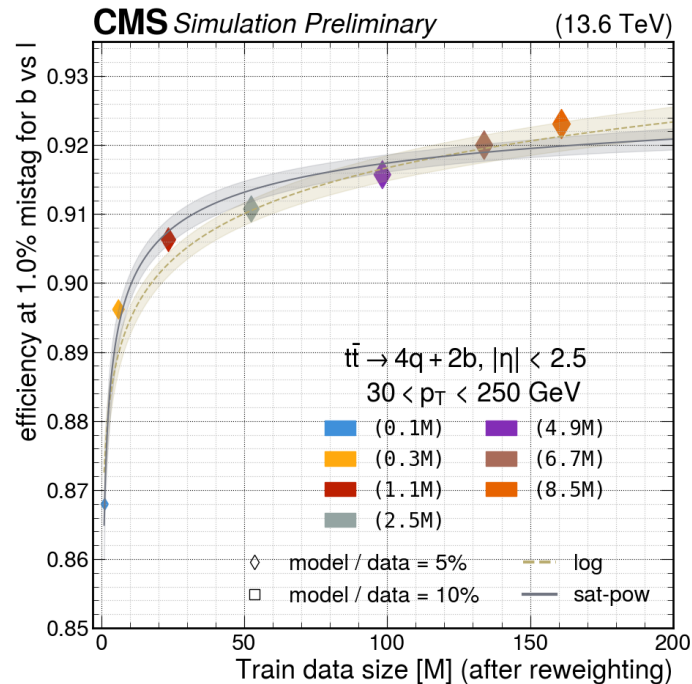
champion, Kasparov, in 1997, were based on massive, deep search. At the of computer-...
When a simple...
based chess res...
search may have won this time, but it was not a general strategy, and a...
wanted methods based on human input to win and were disappointed

... many examples of AI researchers' belated learning of this bitter lesson

Scaling laws



ATLAS scaling study training dataset with up to 7.7 billion jets and 86M parameter model

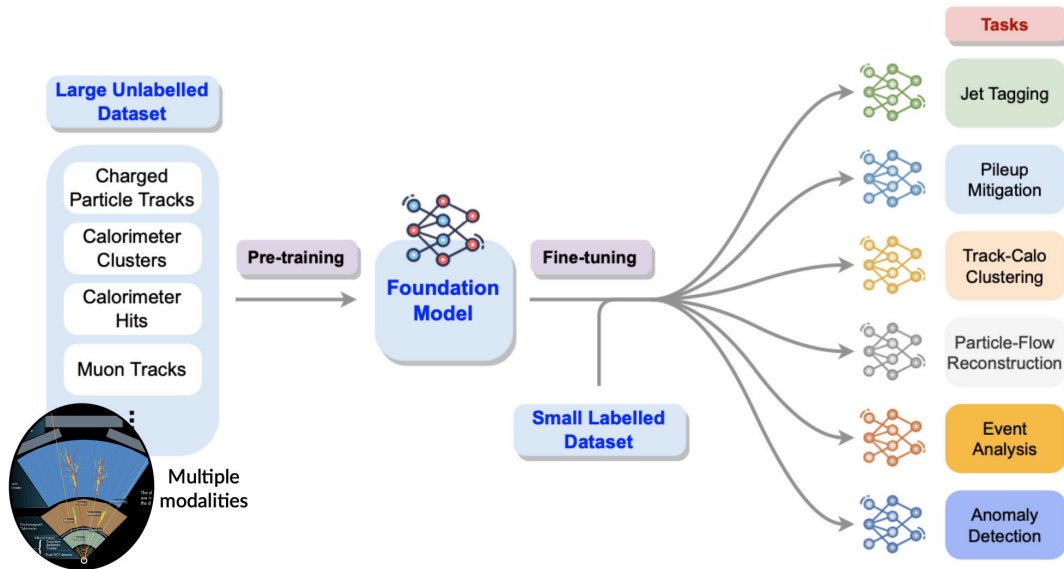


Newest taggers on both experiments improve mainly by training larger models on far more jets. (When) will scale become the dominant lever?

Foundation models

Nice overview in [A. Hallin, Foundation models for high-energy physics](#)

A foundation model is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of down-stream tasks.



[Submitted on 8 Mar 2024 (v1), last revised 7 Sep 2024 (this version, v2)]

Omnijet- α : The first cross-task foundation model for particle physics

Joschka Birk, Anna Hallin, Gregor Kasieczka

PAPER - OPEN ACCESS

Masked particle modeling on sets: towards self-supervised high energy physics foundation models

Tobias Golling, Lukas Heinrich, Michael Kagan, Samuel Klein*, Matthew Leigh, Margarita Osadchy and John Andrew Raine

Published 17 September 2024 · © 2024 The Author(s). Published by IOP Publishing Ltd
[Machine Learning, Science and Technology, Volume 5, Number 3](#)

[Submitted on 14 Dec 2024 (v1), last revised 6 May 2026 (this version, v2)]

Pretrained Event Classification Model for High Energy Physics Analysis

Joshua Ho, Benjamin Ryan Roberts, Shuo Han, Haichen Wang

Bumblebee: Foundation Model for Particle Physics Discovery

Andrew J. Wildridge, Jack P. Rodgers, Ethan M. Colbert, Yao yao, Andreas W. Jung, Miaoyuan Liu

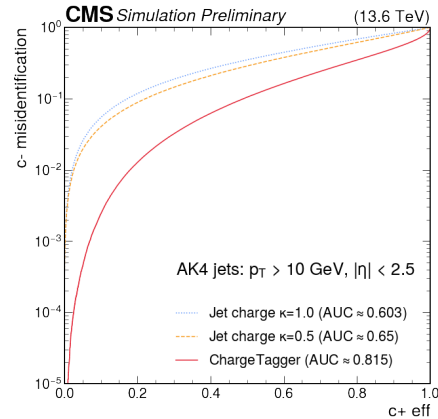
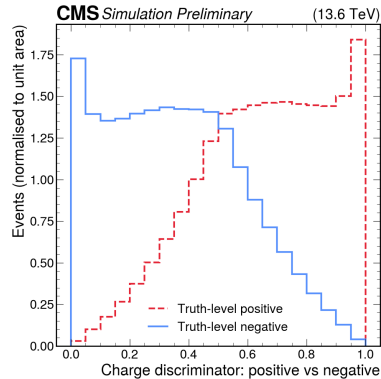
HEP-JEPA: A foundation model for collider physics using joint embedding predictive architecture

Jai Bardhan, Radhikesh Agrawal, Abhiram Tilak, Cyrin Neeraj, Subhadip Mitra

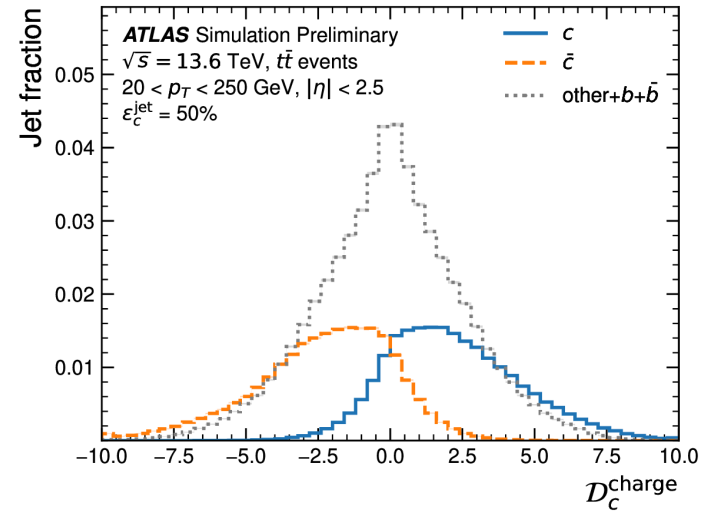
Charge / fragmentation tagging

New **CMS** Charge Tagger tags flavour + charge of jets in one single network

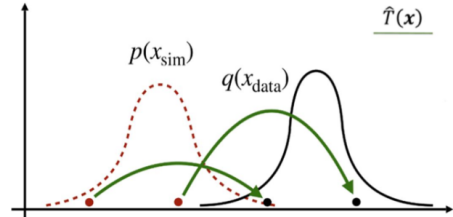
- Based on Particle Transformer with extended input features and two/three classification tasks
- Potentially opens charge-consistent pairing



Similar development at **ATLAS**:

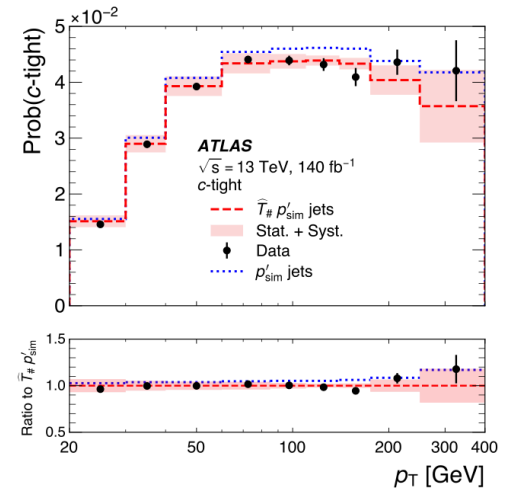
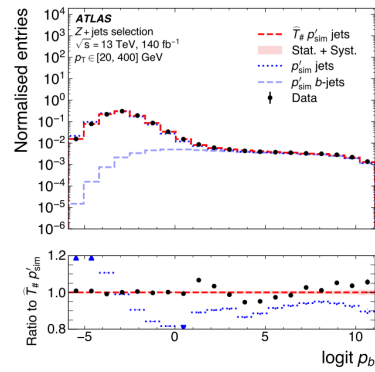
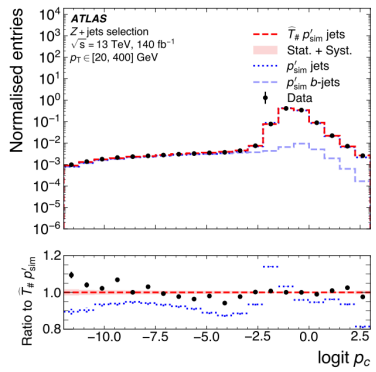
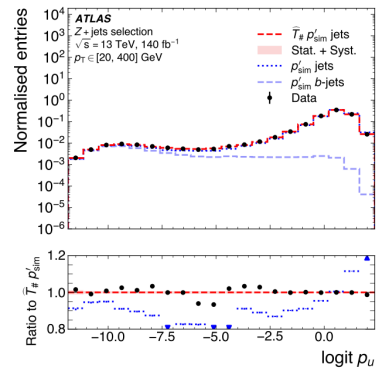


Continuous / optimal-transport calibration



Idea: obtain a mapping from $p_{b/c/u}^{\text{MC}}$ to $p_{b/c/u}^{\text{data}}$

- Architecture: Normalizing flow with a constraint to ensure the transport map is minimal.
- Closure with conventional calibration methods

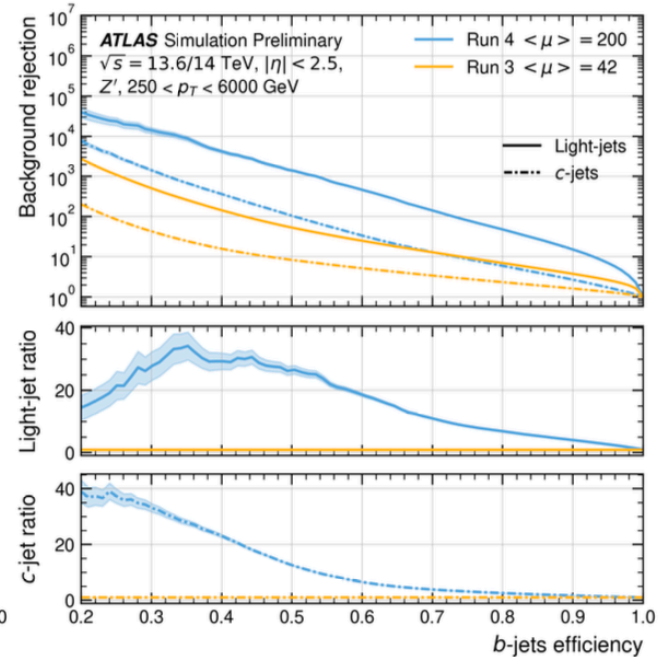
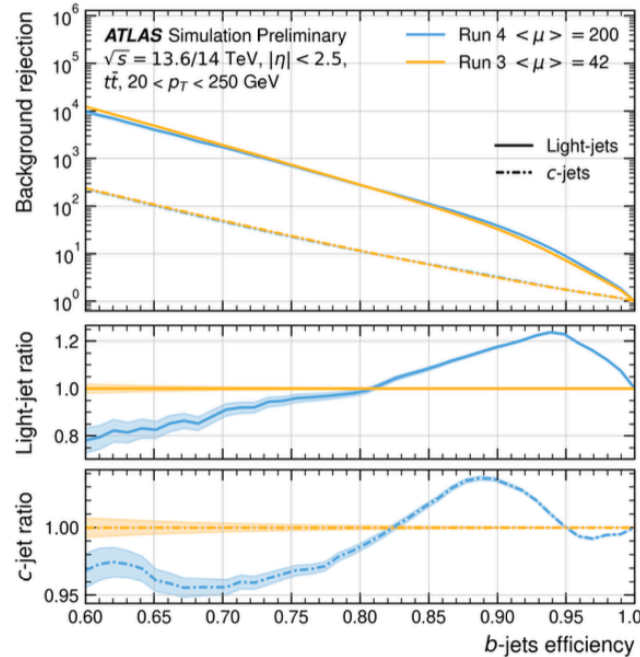
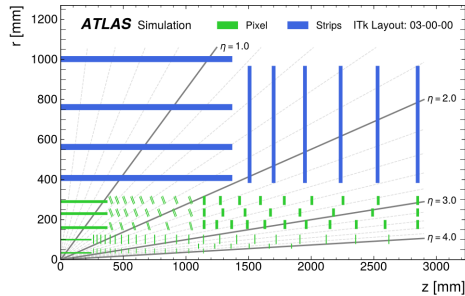


Outlook



Looking ahead: Flavour tagging in Run 4

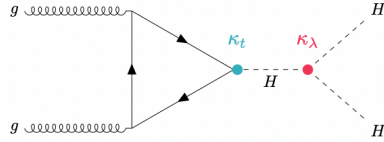
Phase-II upgrades for Run 4: **new inner detectors** with improved granularity and extended eta coverage



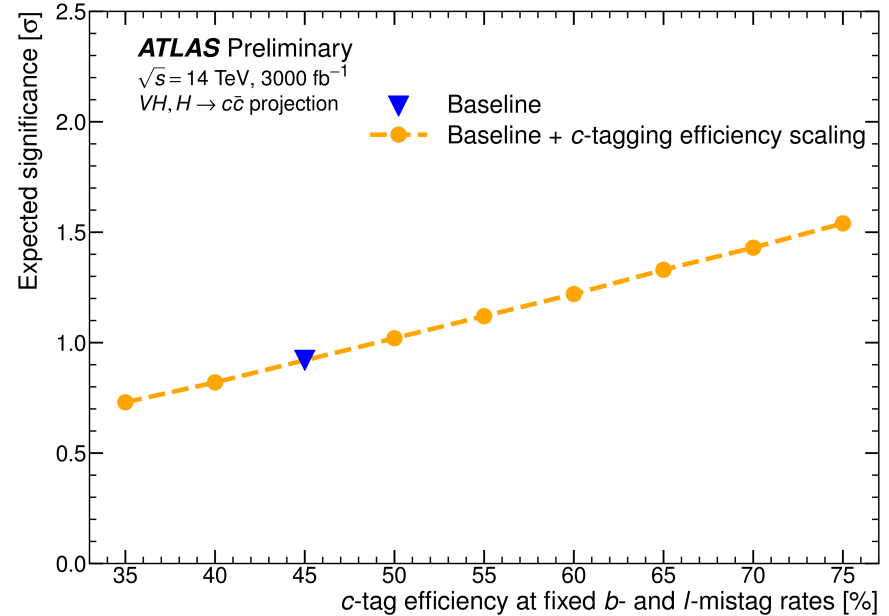
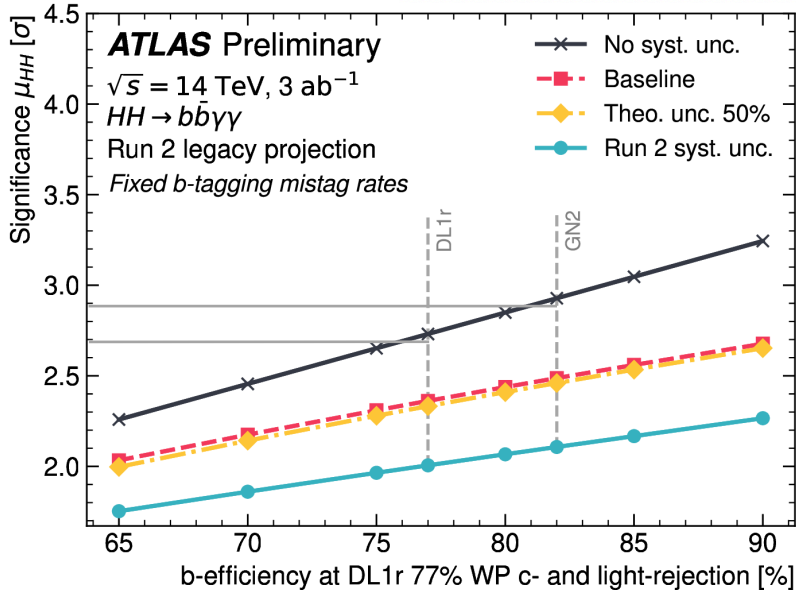
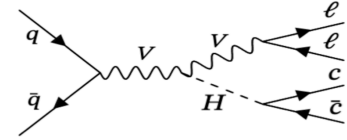
Showing here **ATLAS** b -tagging ROC curves to give idea of improvement w.r.t. Run 3 **despite massive increase in pile-up!**

Flavour tagging impact on Higgs searches

Higgs potential
e.g. $HH \rightarrow b\bar{b}\gamma\gamma$

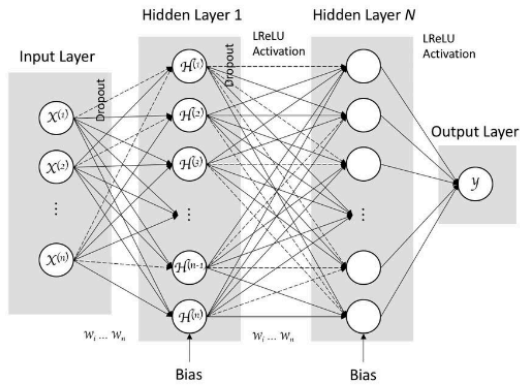


Yukawa couplings
 $VH, H(cc)$ searches



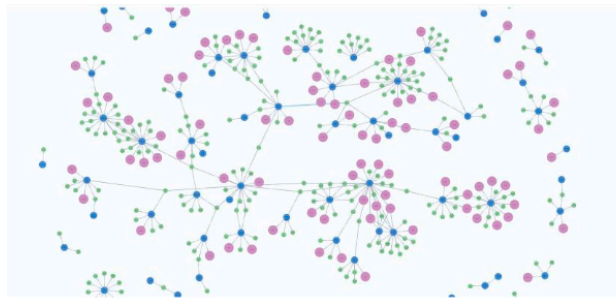
Machine Learning algorithms: Where next?

Vector-to-vector



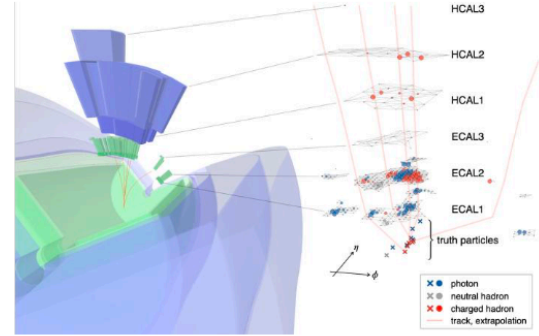
Known and used since many years

Set-to-Vector



Cardinality of the inputs is not fixed but output is

Set-to-Set



Cardinality of both the input and the output are not fixed

We are now mostly in the regime of characterising set-to-vector architectures

Summary

1. Flavour tagging is the experimental technique to select heavy-flavour jets, driving the LHC exploration of Higgs, Top and Beyond Standard Model physics.
2. ATLAS (GN2) and CMS (UParT) converged on multitask transformer taggers built on constituent/vertex interaction features which can be calibrated at fixed operating points
3. Frontiers:
 - Scaling laws (larger models + more data = better performance) and foundation models (self-supervised pretraining + transfer)
 - Continuous, multi-dimensional calibration (beyond 1D operating points)
 - Jet charge tagging, strange tagging, trigger-level tagging
 - HL-LHC upgrades: extended tracking acceptance, better impact-parameter resolution, and timing / 4D tracking for $\langle\mu\rangle\approx 200$



Kontakt



Dr. Philipp Gadow
Research Associate

Universität Hamburg
Institut für Experimentalphysik
Arbeitsgruppe Nikolopoulos
Luruper Chaussee 149
22607 Hamburg

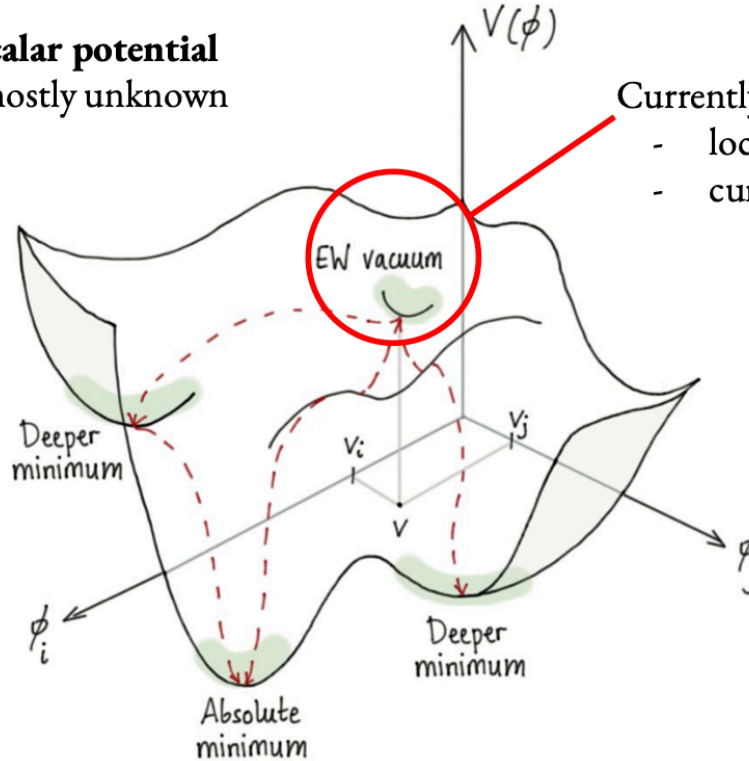
philipp.gadow@uni-hamburg.de

Backup



The shape of the **scalar potential** in our universe is mostly unknown

A more complex structure given by **additional scalar fields** could be realised in nature and hidden in experimental data



Currently known:

- location of the EW minimum: \mathbf{vev}
- curvature around the minimum: $\mathbf{m_H}$

Additional scalar states are appealing because of their relations to the **open questions of the SM:**

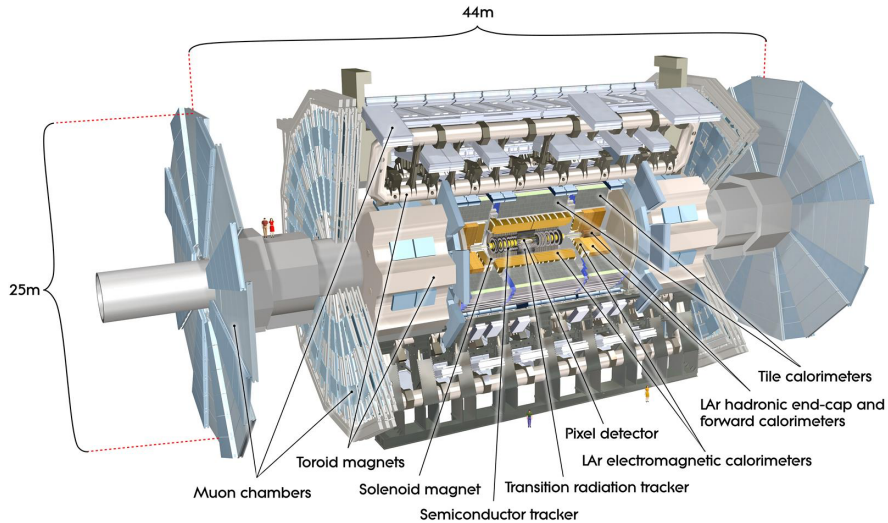
- origin of the EWSB
- stability
- baryogenesis
- portal to dark sector

The ATLAS and CMS experiments

Multi-purpose detectors with 4π coverage

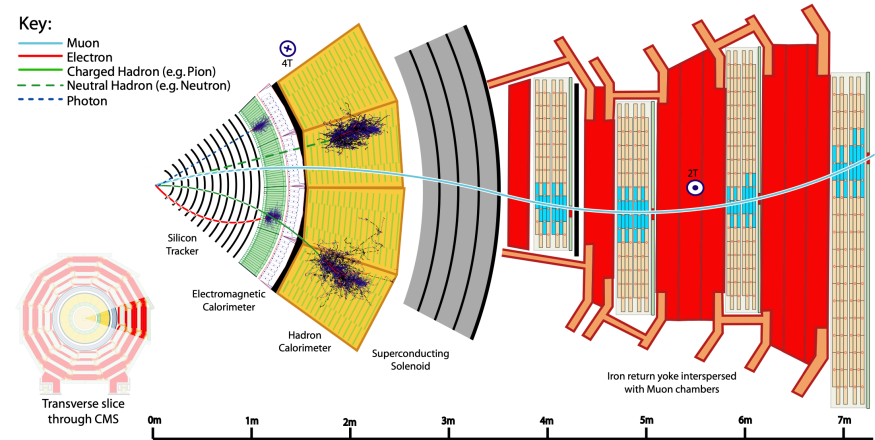
ATLAS Detector

ghj



CMS Detector

fjgh



Source: "Particle Detectors", Kolanoski / Wermes

Track impact parameters

Transverse impact parameter d_0

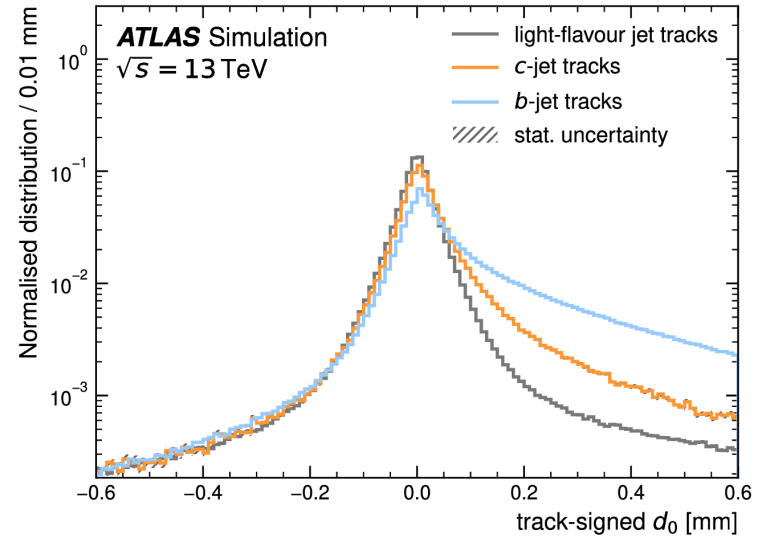
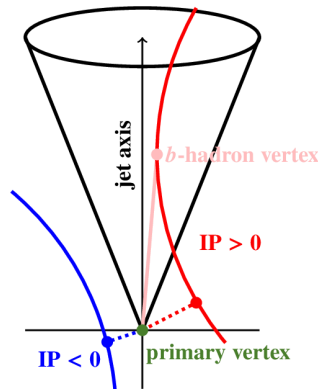
Distance of closest approach of track to the primary vertex in transverse plane.

Longitudinal track impact parameter z_0

Distance in z-direction between primary vertex and track at point of closest approach in x-y plane.

Track-signed impact parameters

Track has positive impact parameter if the angle between jet axis and the line joining primary vertex to point of closest approach to the track is $< \pi/2$ and negative otherwise.

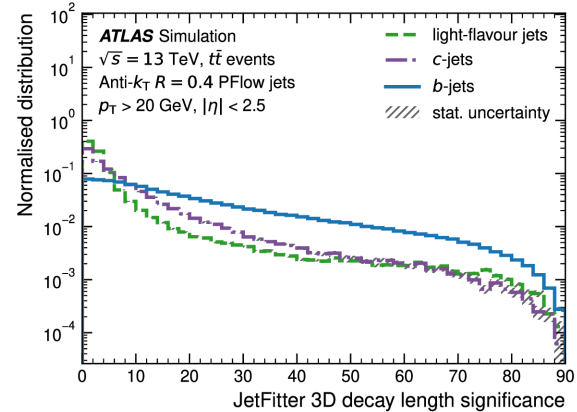
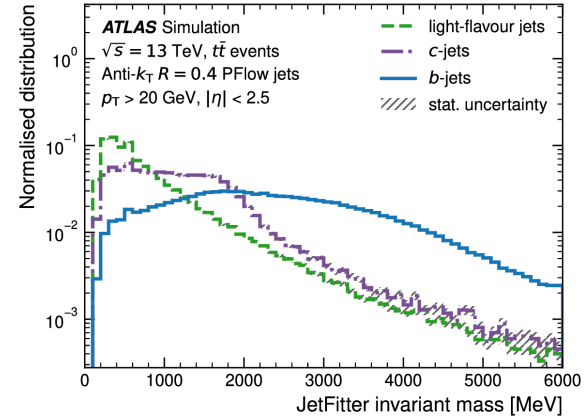
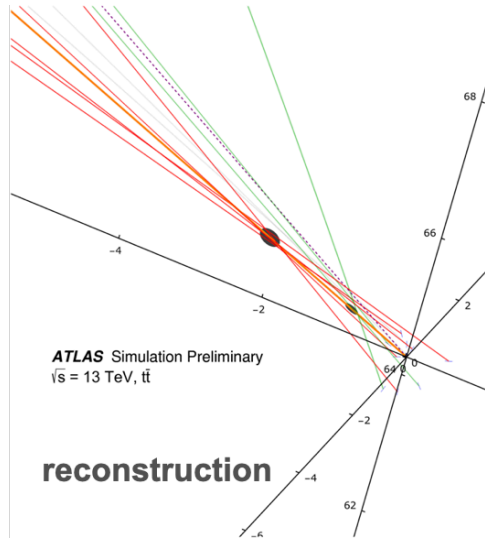
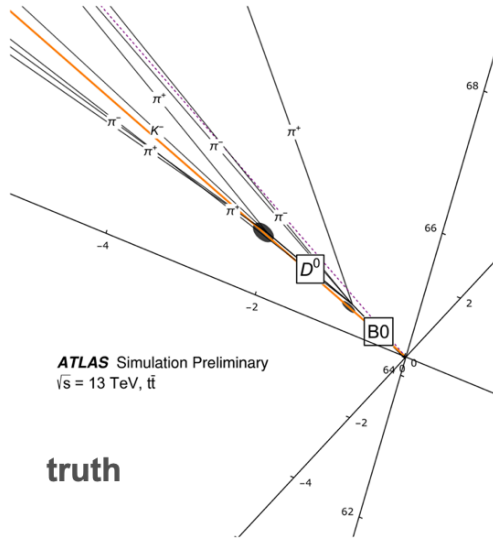


Signed impact parameters of **b-jets** (**c-jets**) are mostly **large and positive**, while light-flavour jets have values consistent with zero within resolution.

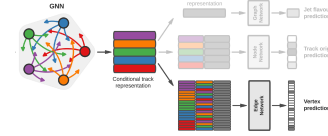
Secondary vertices

Algorithms aim to reconstruct secondary vertices from displaced decays.

Properties of fitted secondary vertices (mass, flight length) provide discrimination for heavy-flavour jet identification.



ATLAS GN2 vertex finding auxiliary task

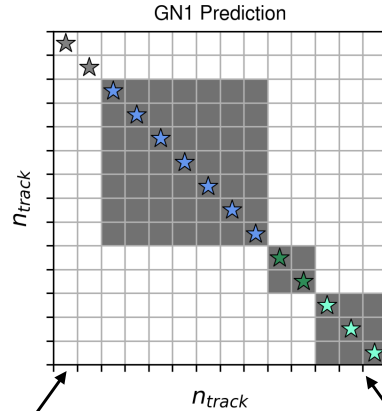
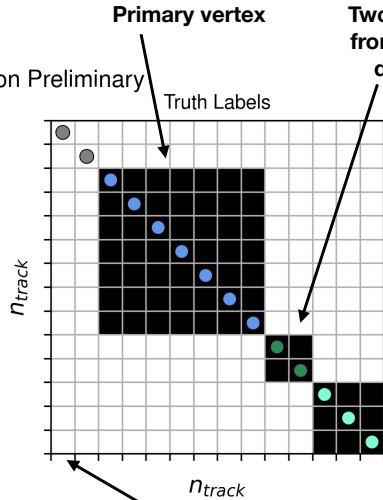


Edge classification as an auxiliary task grouping tracks from common vertices helps the jet classification.

GN1 successfully predicts jet flavour

ATLAS Simulation Preliminary
 $\sqrt{s} = 13$ TeV
 $t\bar{t}$ jets

Truth b -jet
 $p_T = 134.1$ GeV
 $p_b = 0.995$
 $p_c = 0.005$
 $p_u = 0.000$



GN1 vertex and origin prediction is perfect

- Truth
- ★ Predicted
- Pileup
- Fake
- Primary
- FromB
- FromBC
- FromC
- FromTau
- OtherSecondary

These metrics quantify the GN1 algorithm. Performance for GN2 is better.

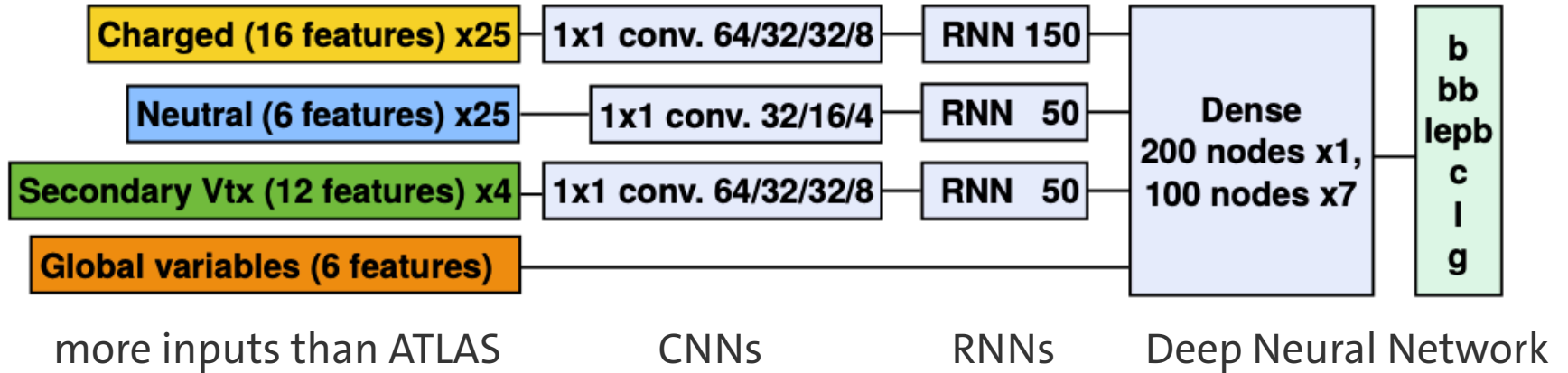
Efficient vertexing:

- recall of > 65% of tracks in truth vertex
- purity of > 50% of tracks in reco vertex

The algorithm correctly identifies 80% of truth vertices inside b-jets.

cherry-picked example

CMS DeepJet algorithm



UParT robustness: adversarial attacks

How to improve robustness against Monte-Carlo (MC) simulation mismodeling?

Learn from input feature distortions while keeping nominal performance.

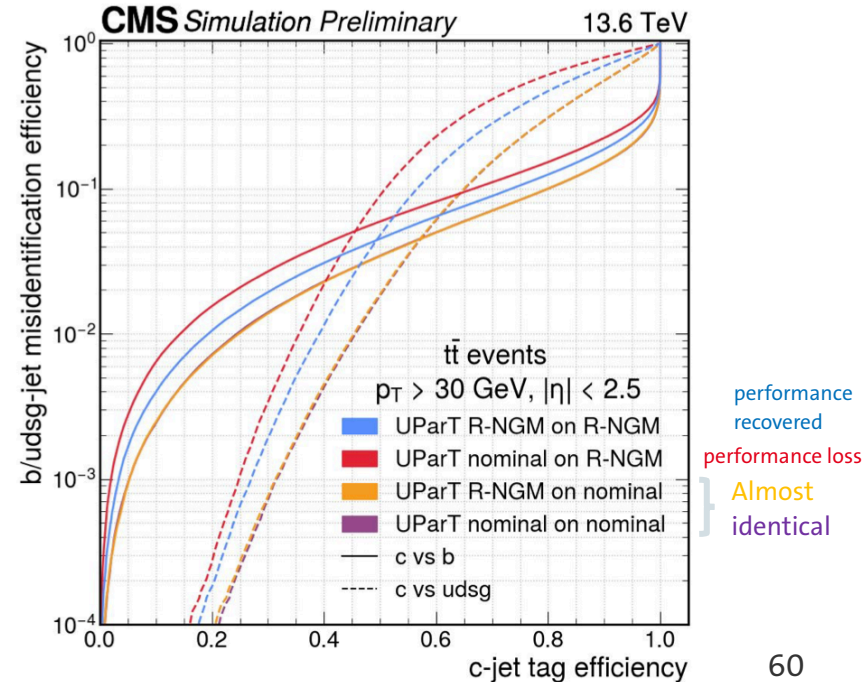
R-NGM (Rectified Normed Gradient Method):

double forward-backward pass, one to compute the adversarial attack,

$$x_{i,adv} = x_i + \epsilon \cdot |\nabla CE(x_i, \theta)|$$

one to train on nominal + adversarial samples using TRADES („trade adversarial robustness off against accuracy“) loss.

$$L = CE(x_{i,adv}, \theta) + \lambda \cdot KL(x_i, x_{i,adv})$$



Training with adversarial attacks

Nominal train

- 1) Data loader loads batch
- 2) Feed batch into model
- 3) Compute loss
- 4) Update weights

$$W_{new} = W - \ell \nabla_W L$$

Adversarial prediction

- 1) Data loader loads batch
- 2) Feed batch into model
- 3) Compute loss
- 4) Perform attack to obtain adversaries
 - 1) update inputs $I_{new} = I + \epsilon \nabla_I L$
 - 2) feed attacked batch into model
 - 3) Compute loss
- 5) Update weights

$$W_{new} = W - \ell \nabla_W L$$

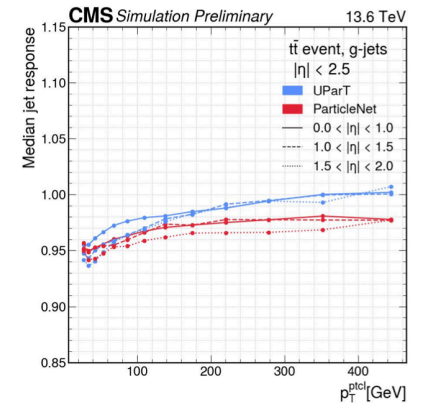
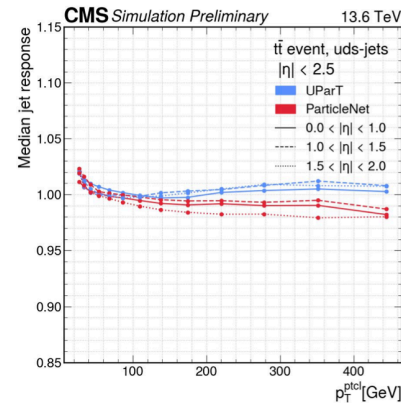
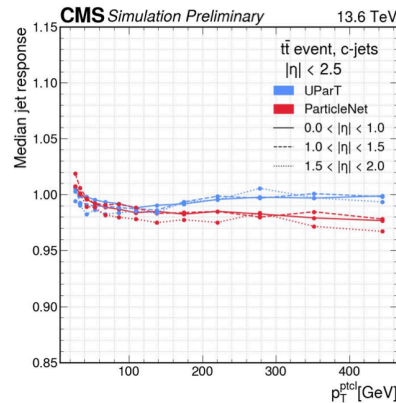
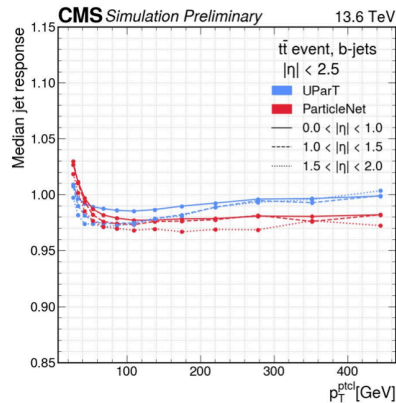
Jet energy regression in UParT

Classification

Regression

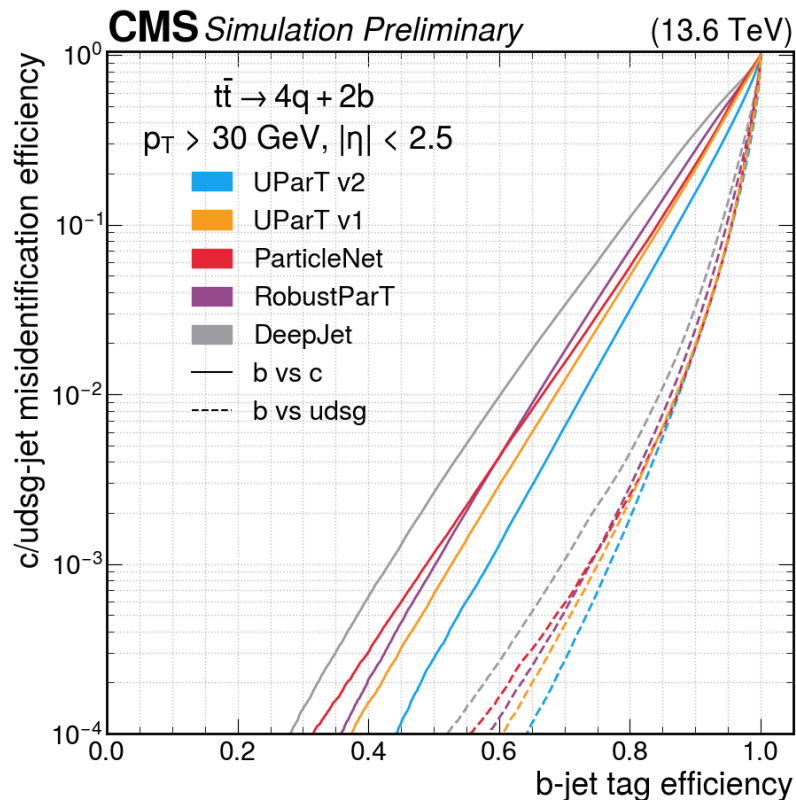
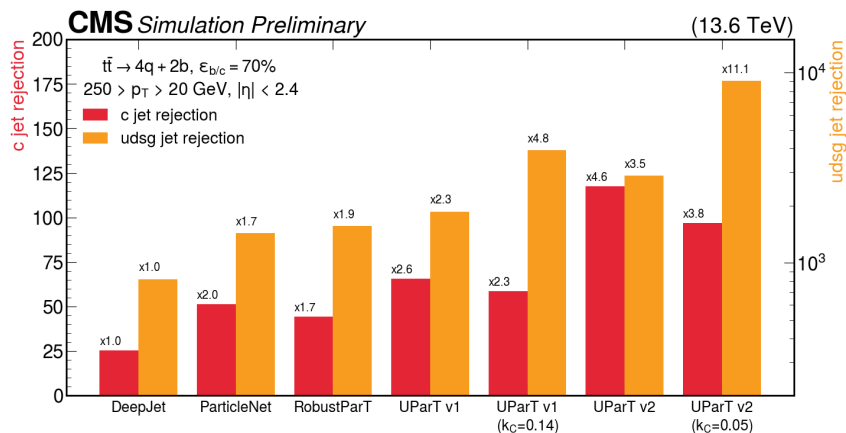
Quantile regression
(resolution estimation)

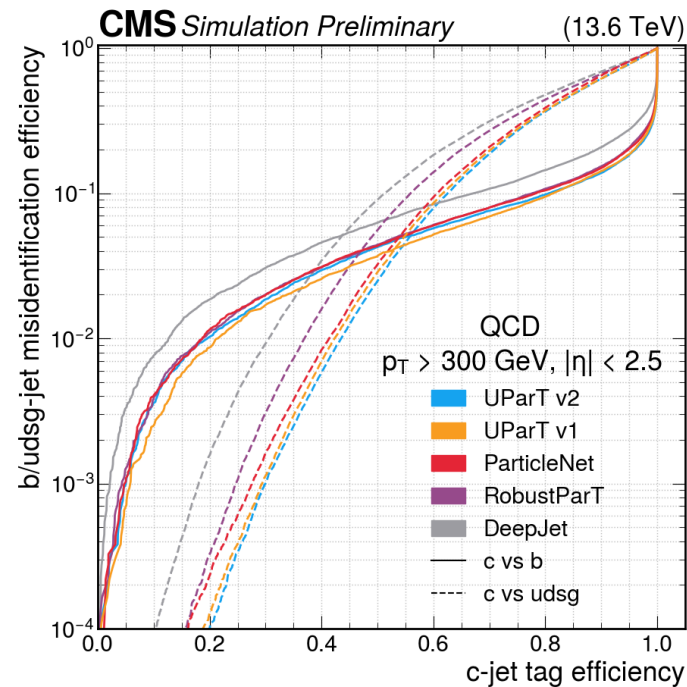
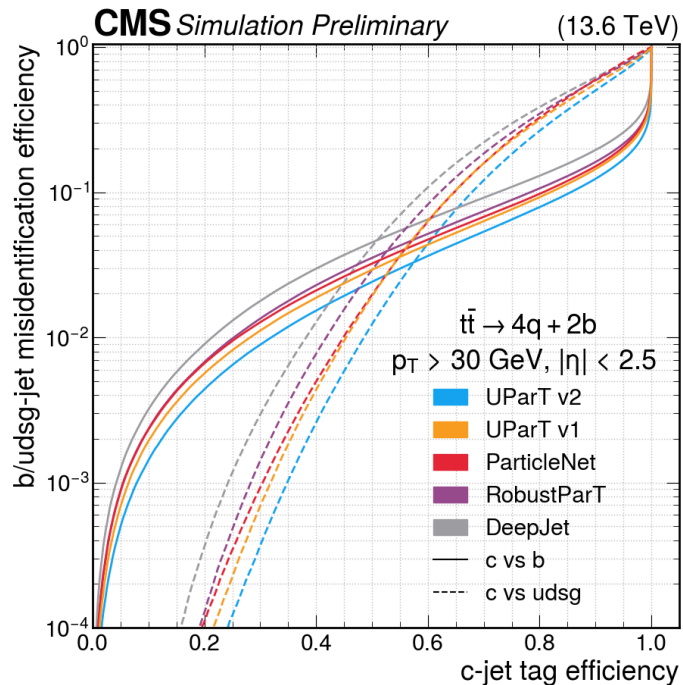
$$L = \text{CatEntropy}(x, x_{truth}) + \gamma_{\text{regr}} \cdot \log(\cosh(y - y_{truth})) + \gamma_{\text{quantile}} \cdot [p_{0.16}(z - z_{truth}) + p_{0.84}(z - z_{truth})]$$

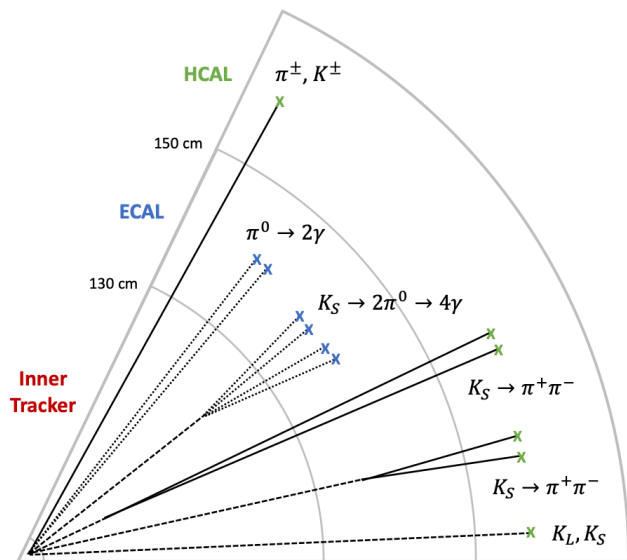


Differences between UParTv1 and UParTv2

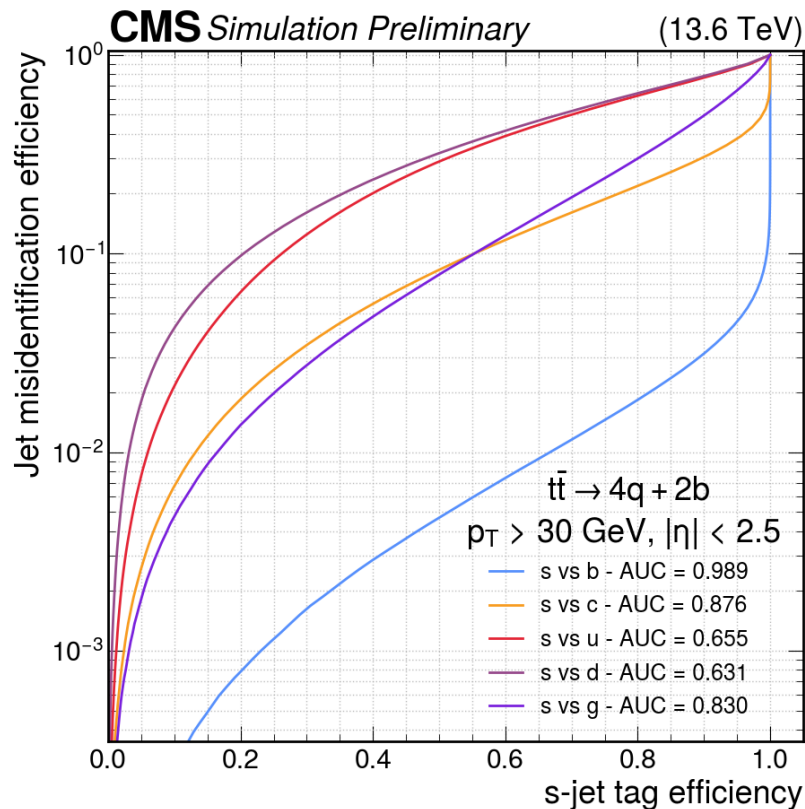
Hyperparameter	UParT v1	UParT v2
Number of layers	6	6
Embedding (hidden) size	128	192
Total number of parameters	2M	5.7M
Activation function	GELU	SILU/SwiGLU
Normalization layer	LayerNorm	RMSNorm
Size of training dataset (after reweighting)	35M	70M
Optimizer	Ranger	AdamW
Learning rate	2e-03	1e-03







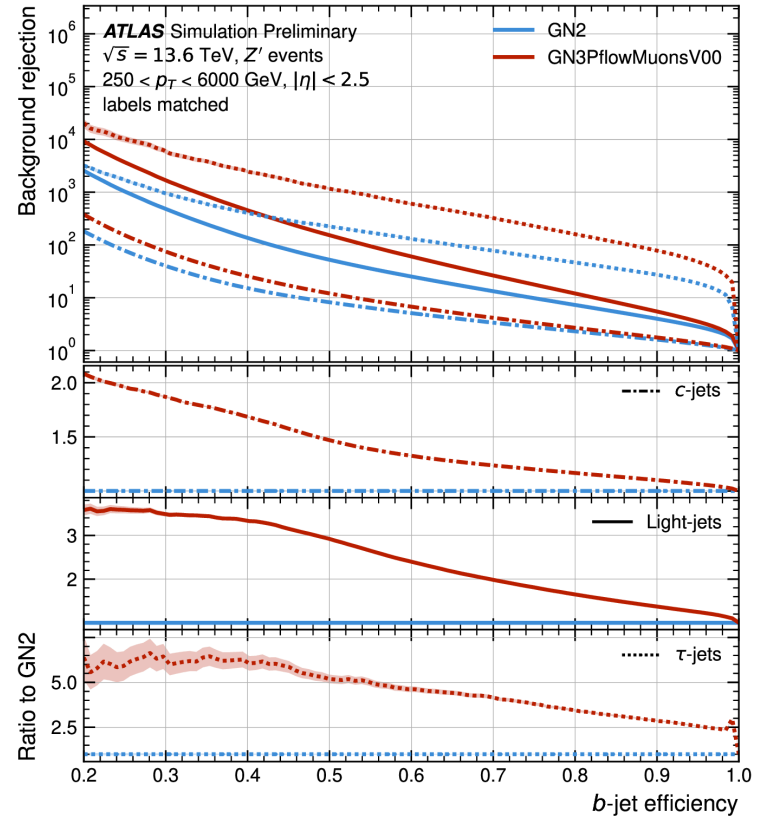
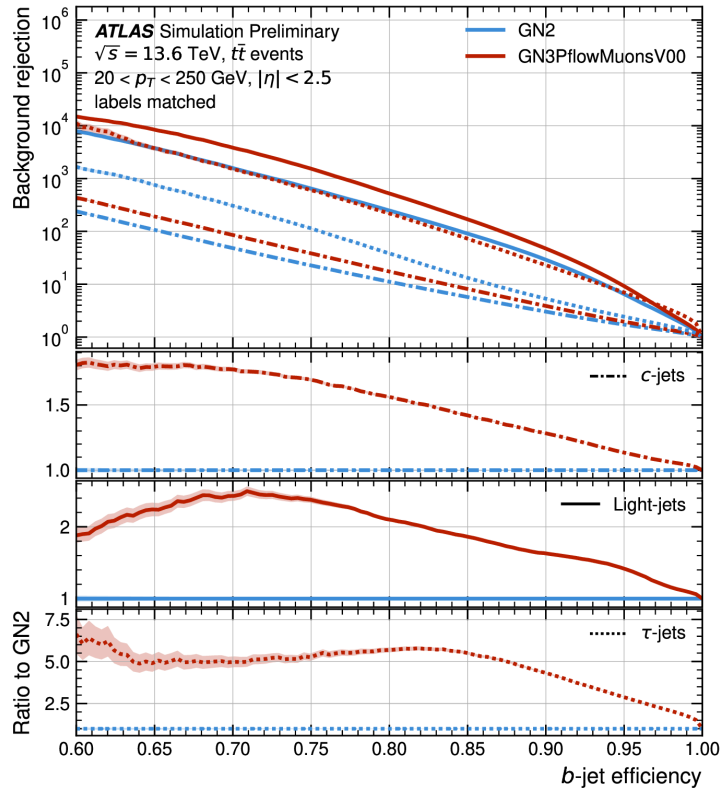
[arXiv:2003.09517](https://arxiv.org/abs/2003.09517) [hep-ph]



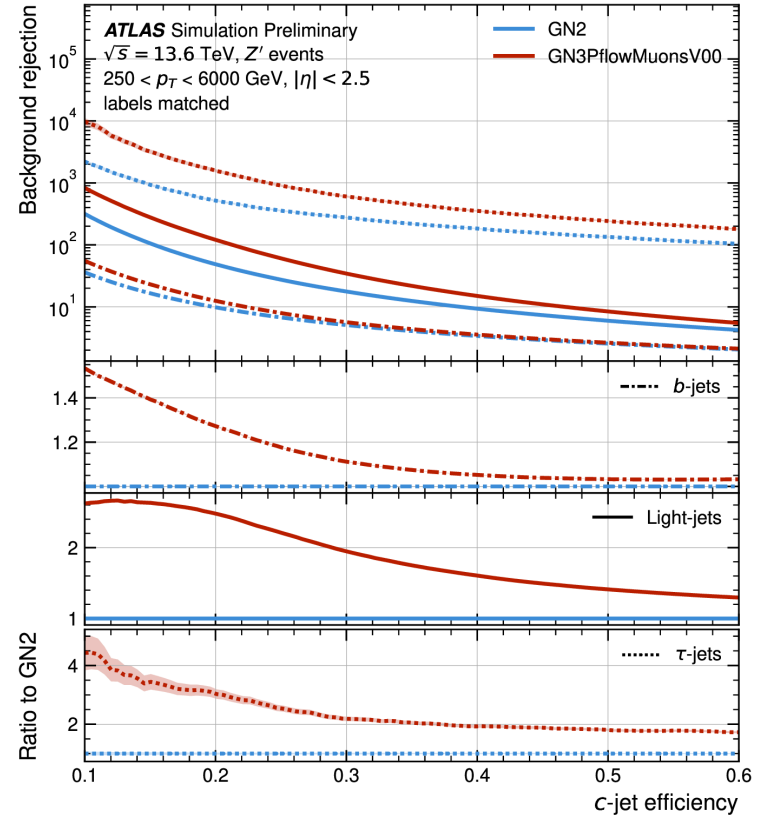
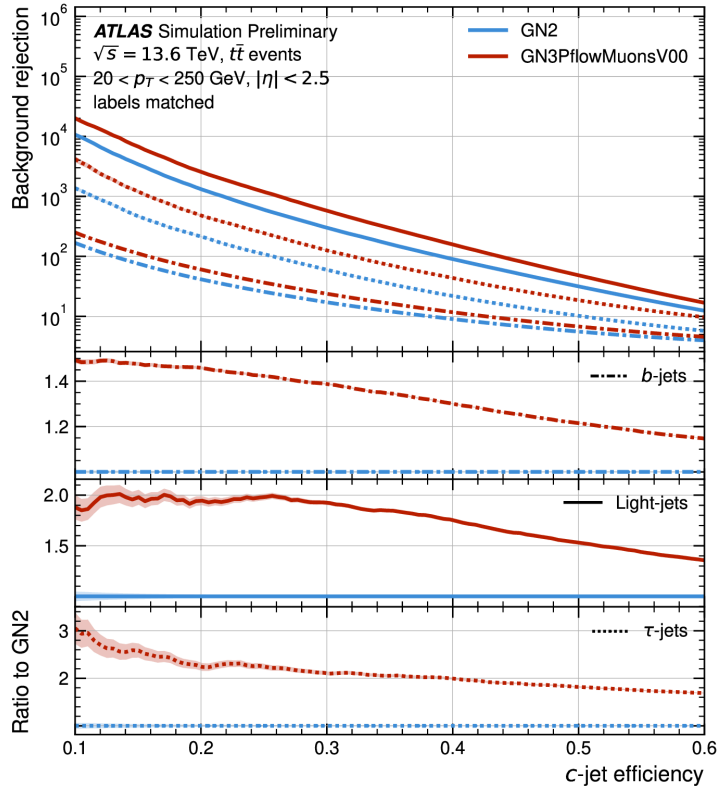
Differences between GN2 and GN3

Component	GN2	GN3
Track–jet association	ΔR –based association	Ghost-association
Track selection	$d_0 < 3.5$ mm, 40 tracks	$d_0 < 5$ mm, 50 tracks
Inputs	Jets and tracks	Jets, ghost-associated tracks, soft muons, PFlow objects
Activation function	ReLU	SiLU
Initialisation layers	256	512
Transformer encoder	4 scaled dot-product attention	4 Flash Attention, Gated Linear Units, 8 register tokens
Embedding dimension	256	512
Jet classification	4 classes: b, c, ℓ, τ	6 classes: b, c, ud, s, g, τ
Loss balancing	Fixed weights per task	Geometric mean of losses (GLS)
Optimiser	AdamW	Lion

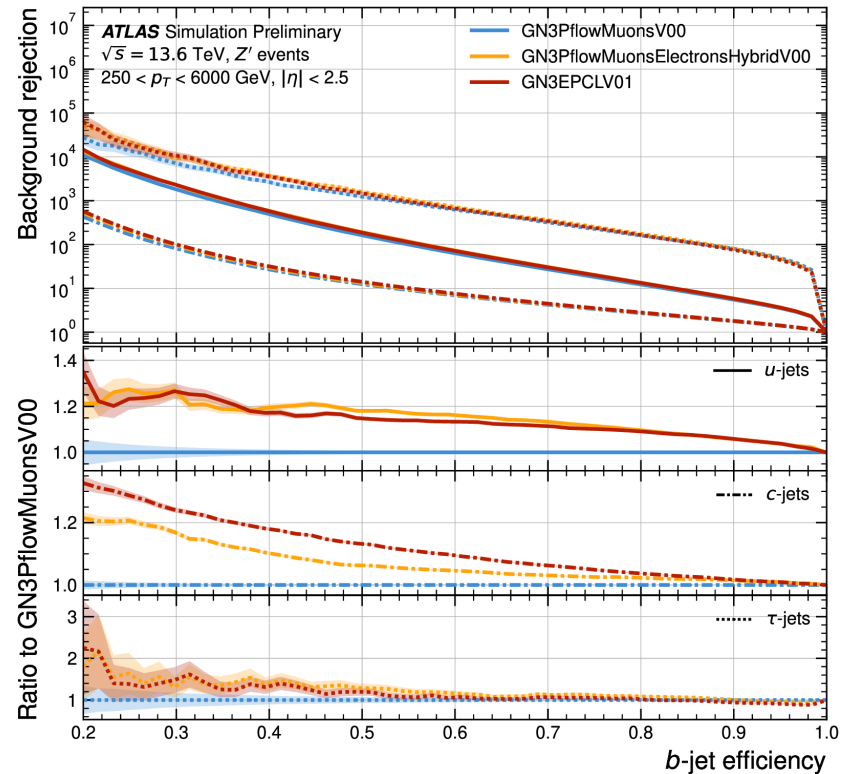
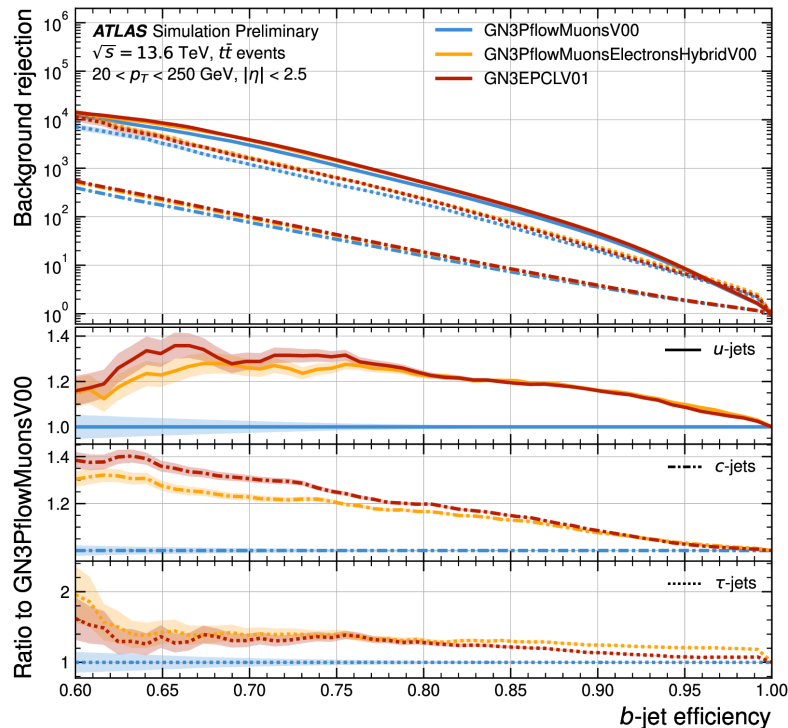
ATLAS GN3 b-tagging performance



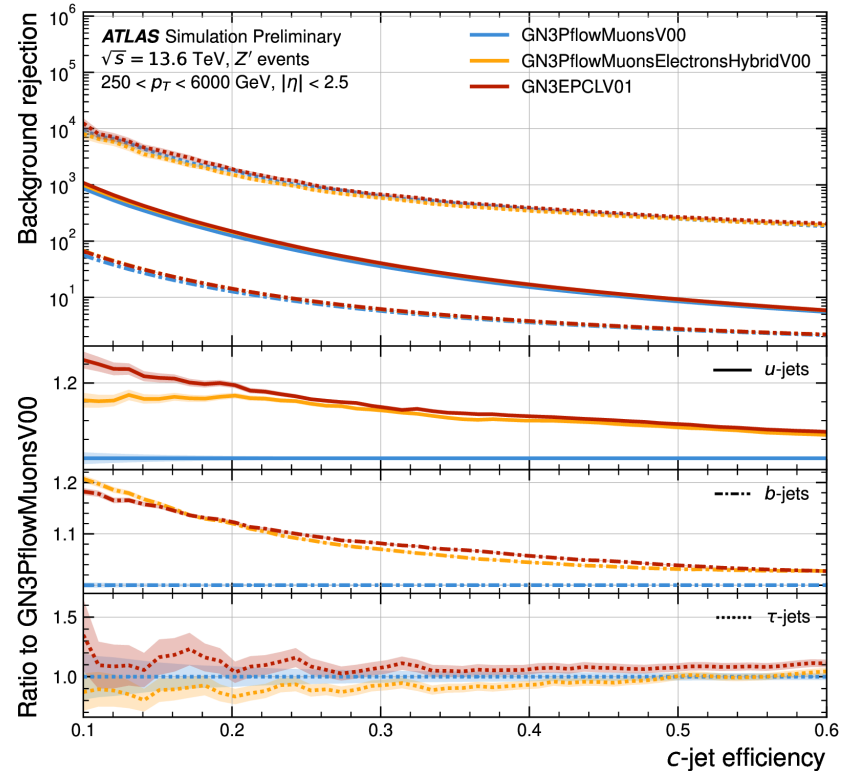
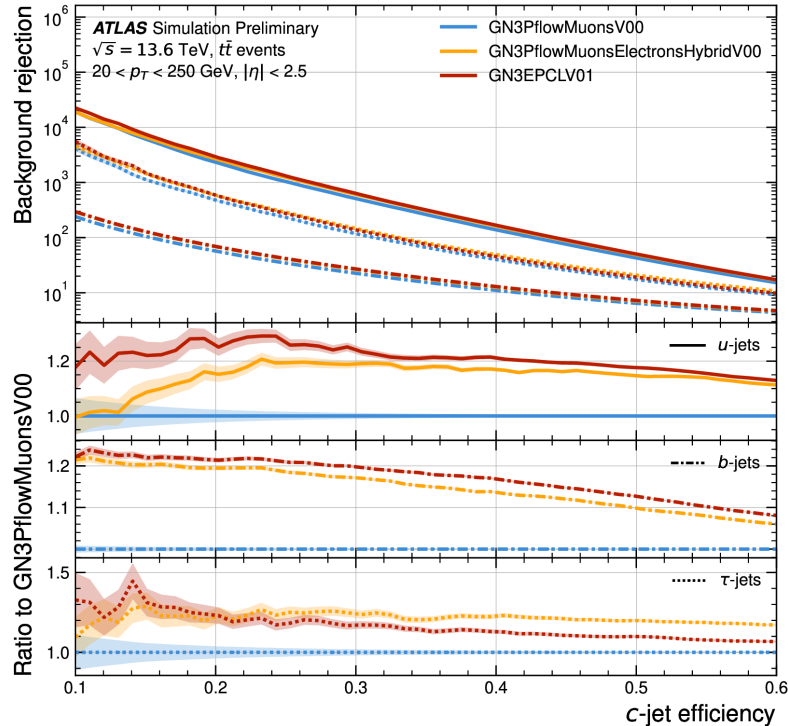
ATLAS GN3 c-tagging performance



ATLAS GN3EPCLV01 b-tagging performance



ATLAS GN3EPCLV01 c-tagging performance



Charge / fragmentation tagging generator dependence

