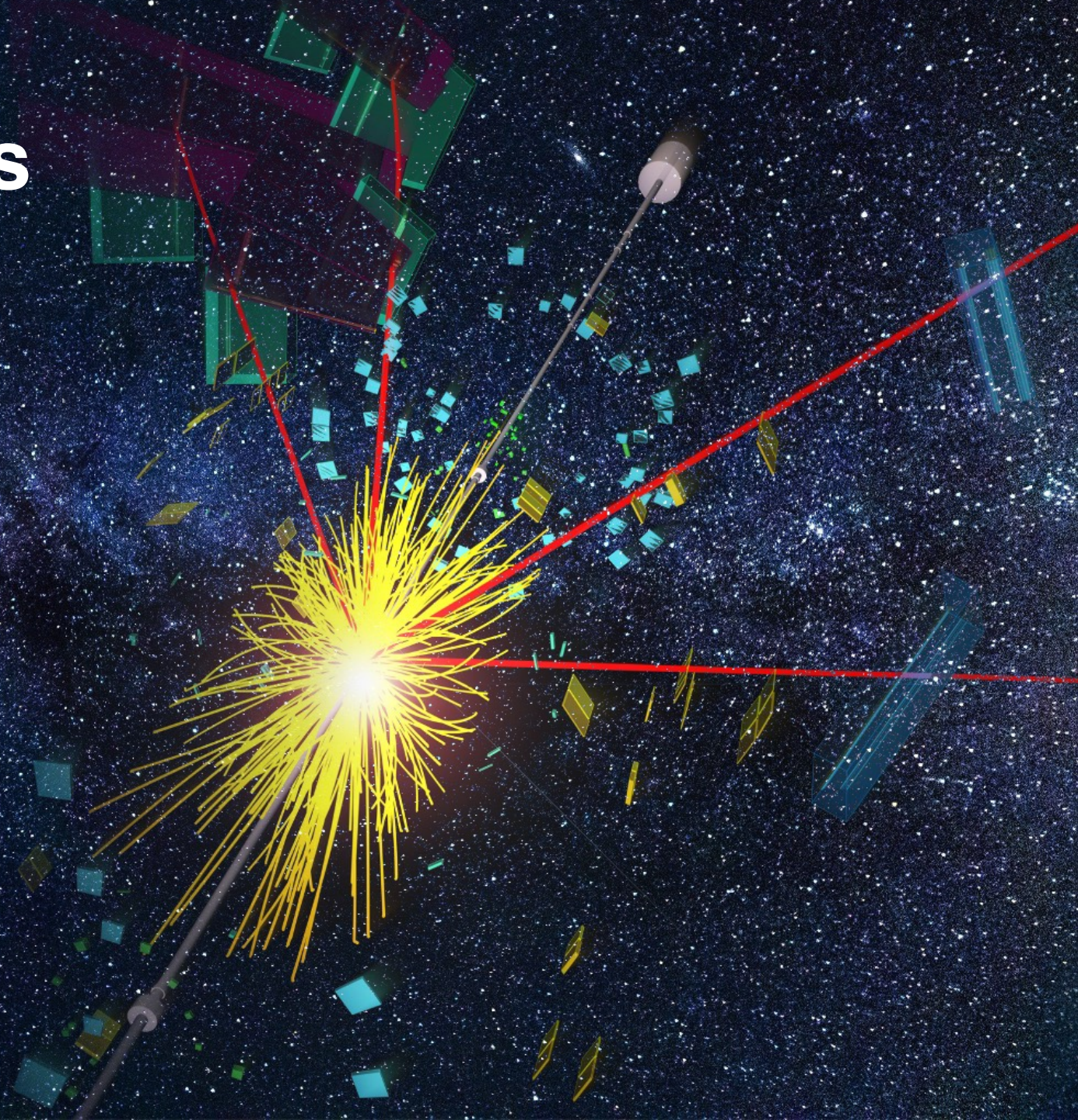


Experimental Physics at the LHC (2)

Andreas Hoecker (CERN)

57. Herbstschule HEP

1–11 September 2026, Bad Honnef, Germany

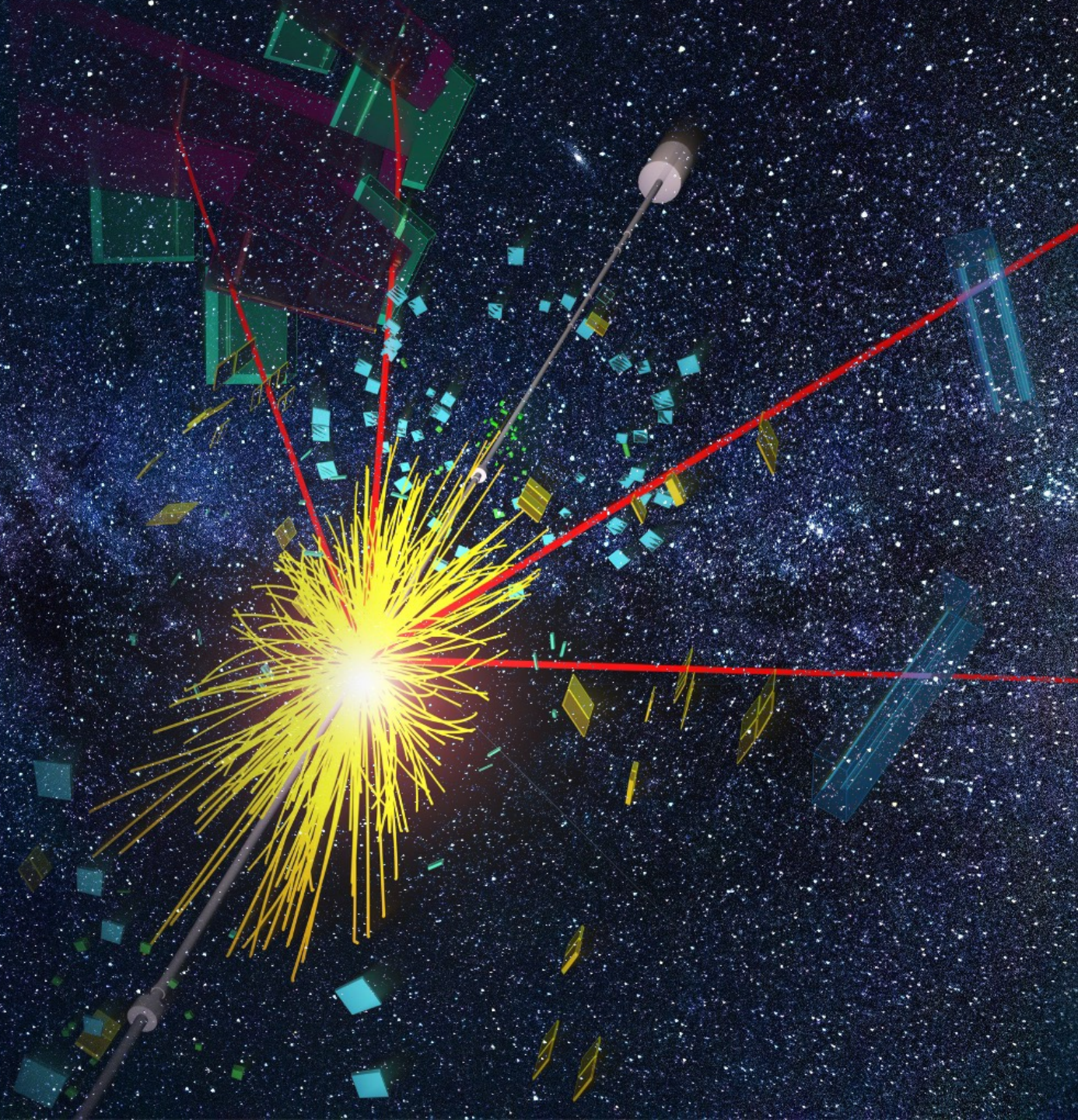


Today's lecture:

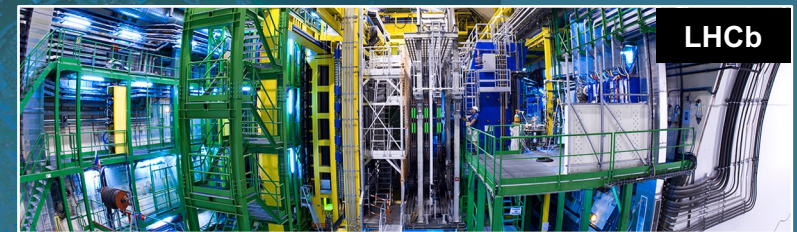
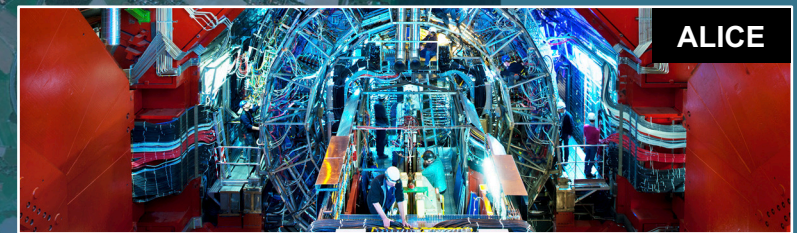
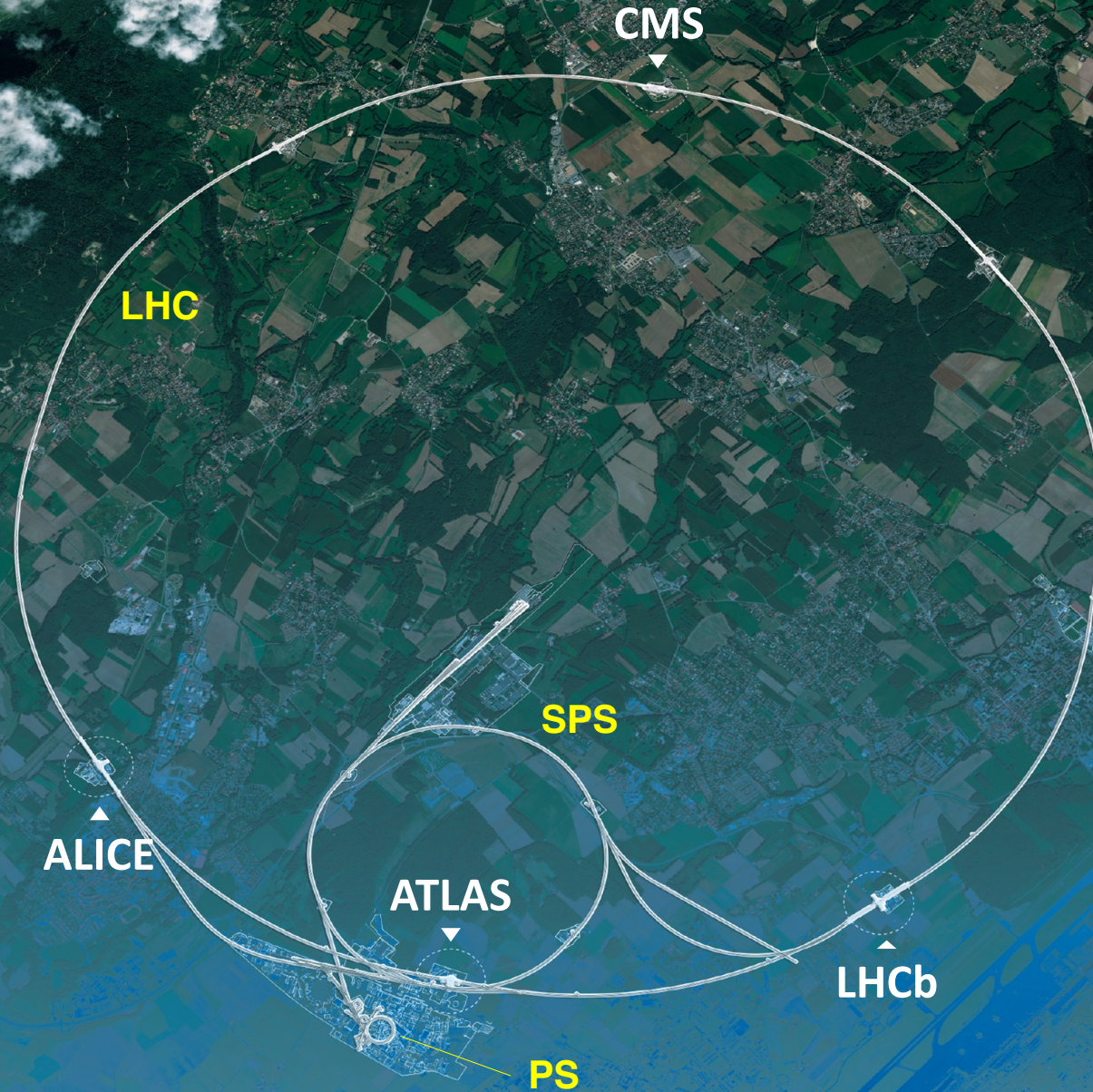
- Particle detection at the LHC
- Data reconstruction and analysis

Next lectures:

- Selected physics topics
- Upgrade to the High-Luminosity LHC
- New physics opportunities

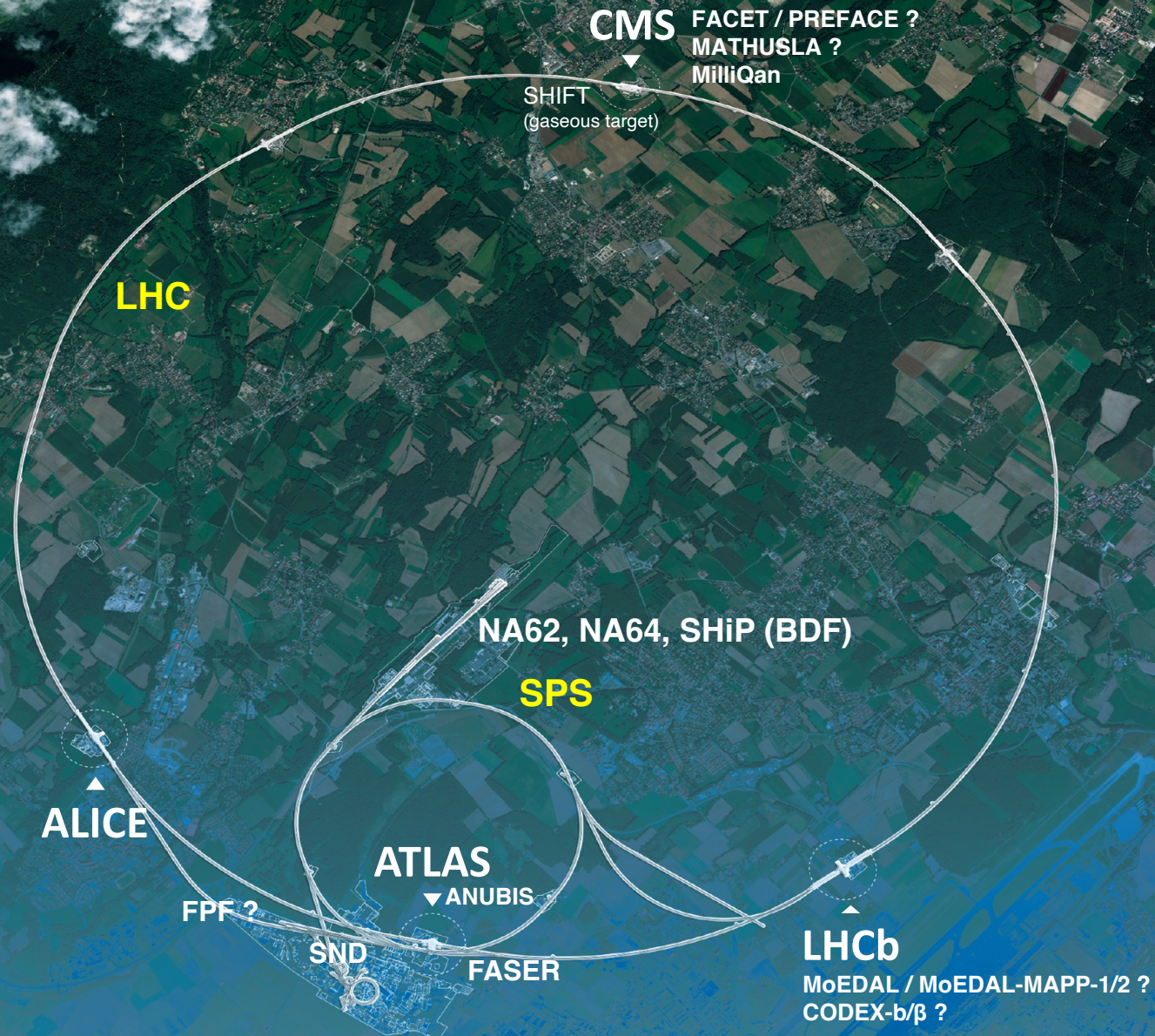


LHC Experiments



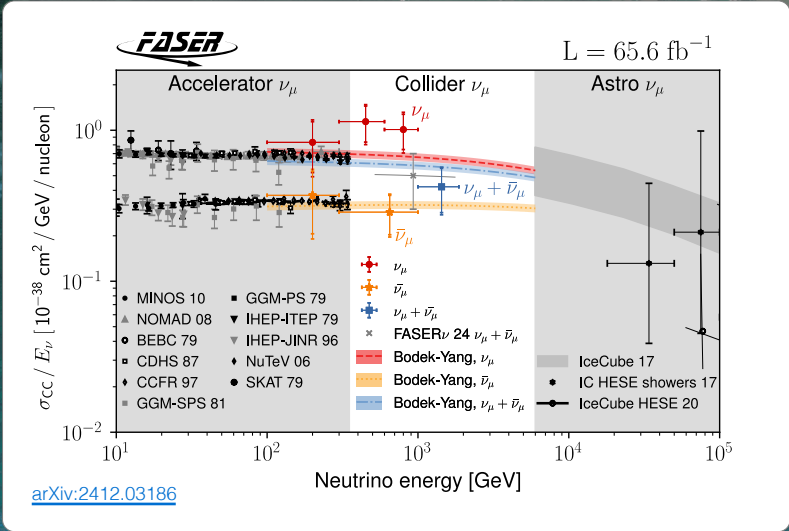
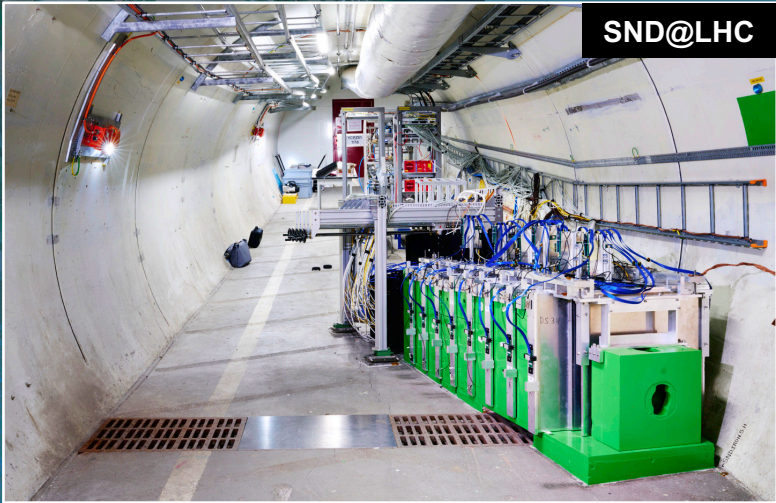
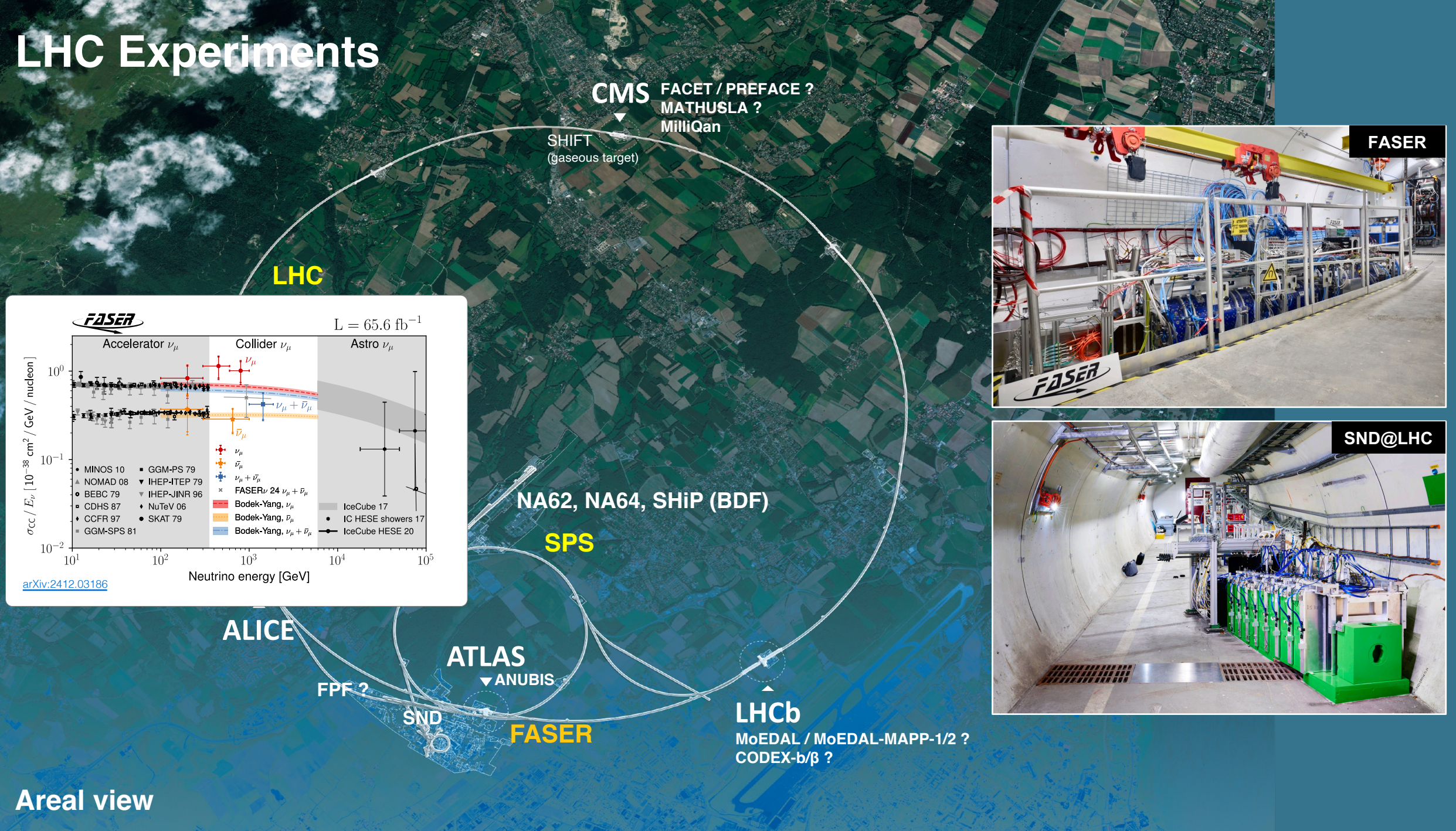
Areal view

LHC Experiments



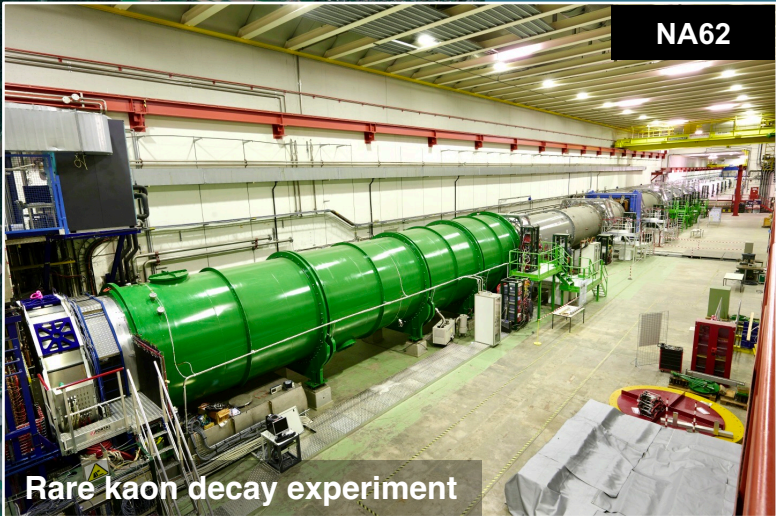
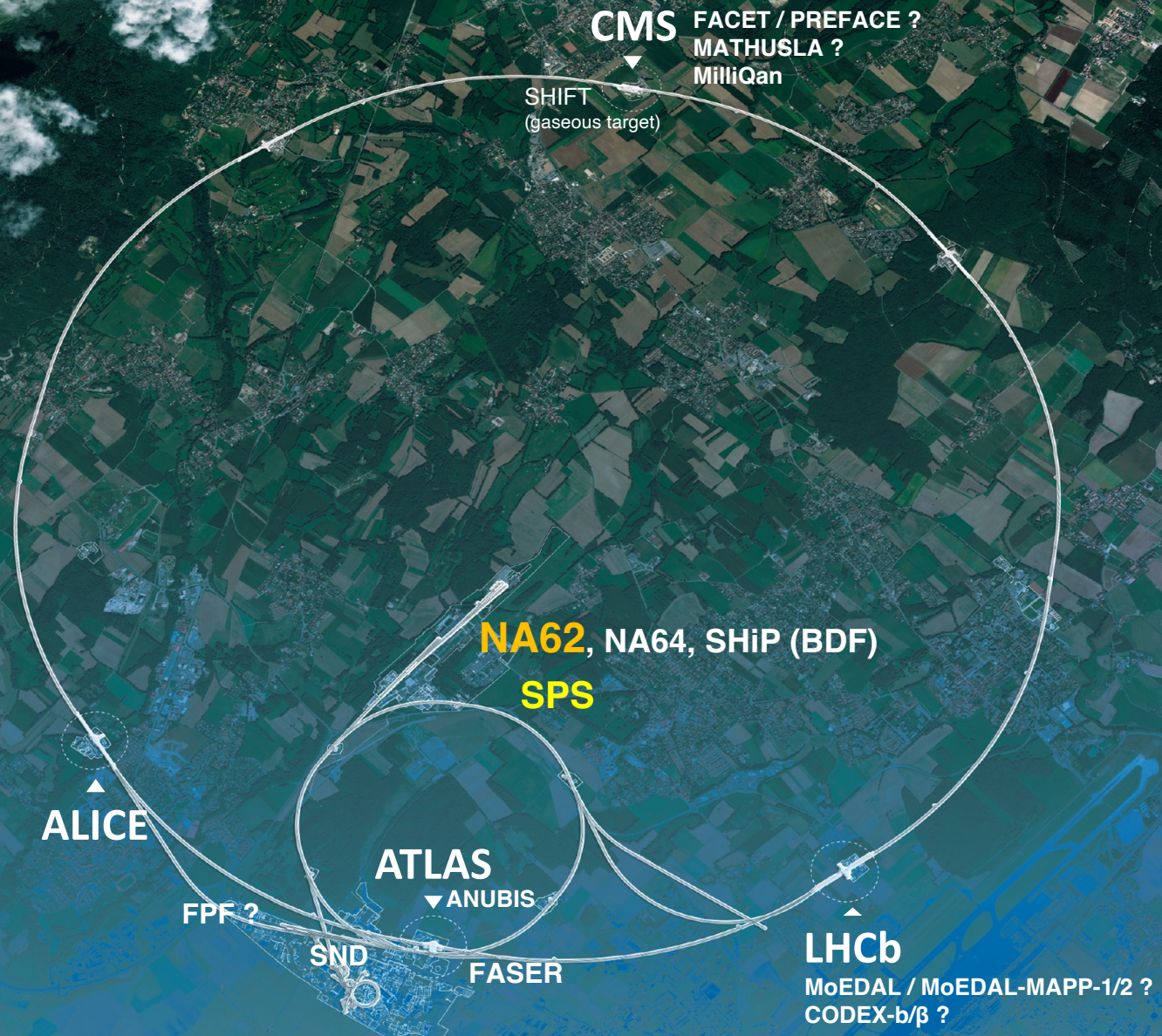
Areal view

LHC Experiments



Areal view

LHC Experiments

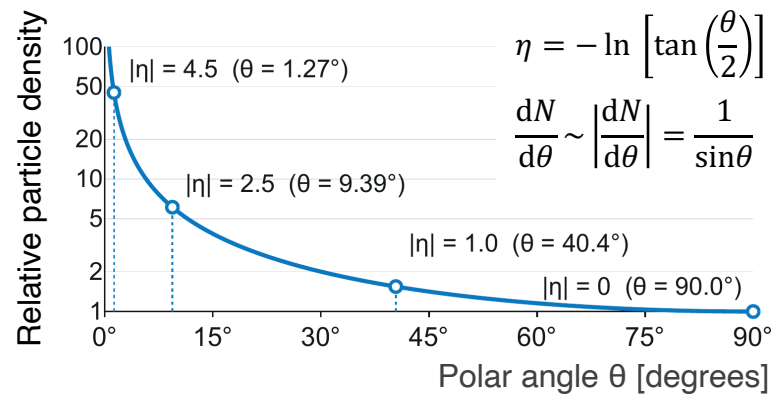


Areal view

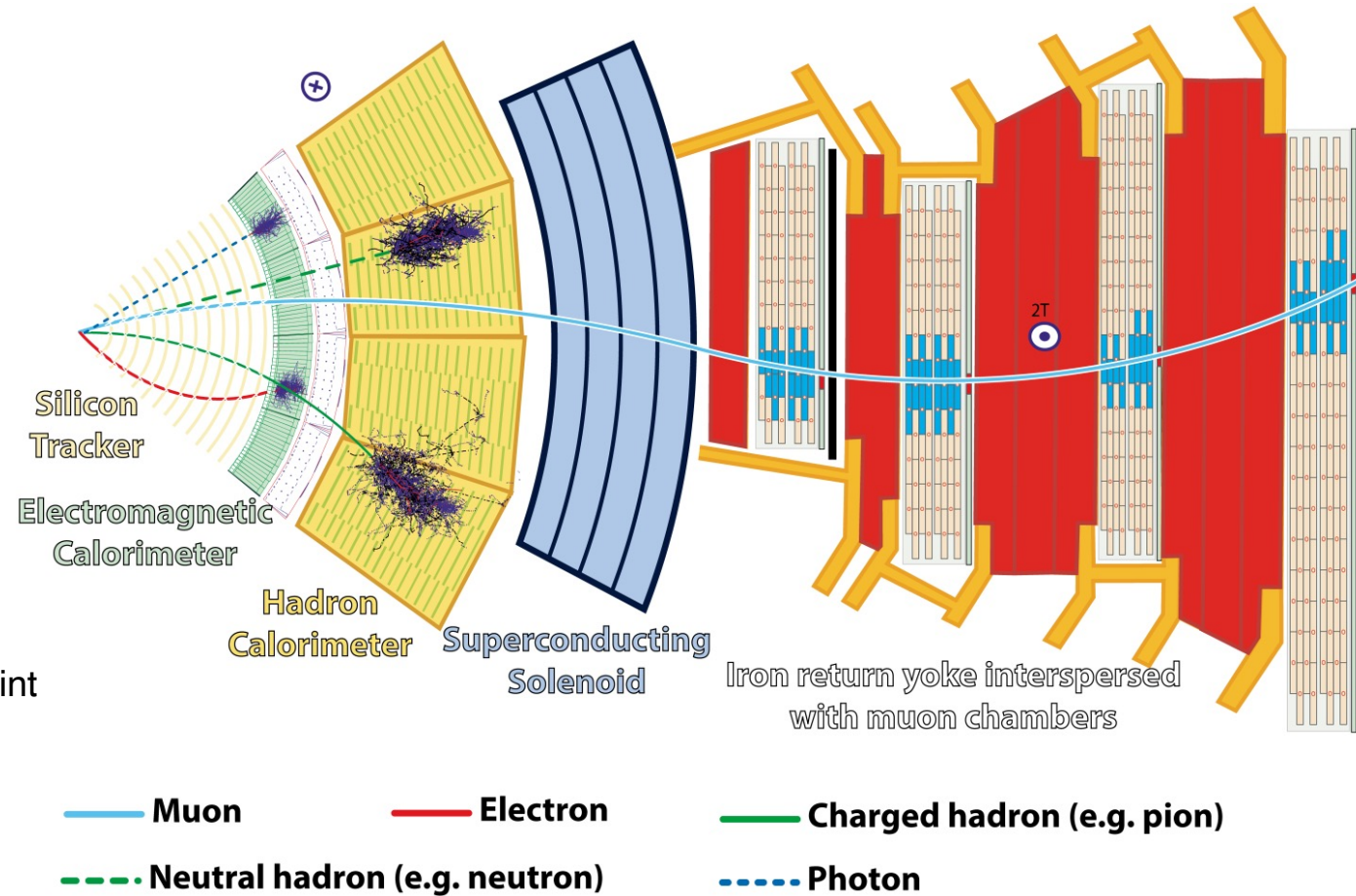
Particle detector basics

Requirements (up to 40 MHz bunch crossing rate, 64 pileup collisions — $2.5 \times$ original design!)

- Radiation hard inner tracker (silicon with large S/B) and forward calorimeters



- High-granular, low-material ($\sigma_{MS} \propto \sqrt{X/X_0}$) tracker*
- High-resolution tracking and vertexing close to collision point
(b/c-hadrons, τ -lepton have lifetimes: O(1.5), O(0.5), 0.3 ps $\rightarrow$ decay lengths ($\beta\gamma c\tau = (p/m)c\tau$) of 1.7, 1.5, 1.0 mm at $p = 20$ GeV)
- High-granularity, linear, uniform, low-noise, dense calorimeters
- High-acceptance muon tagging



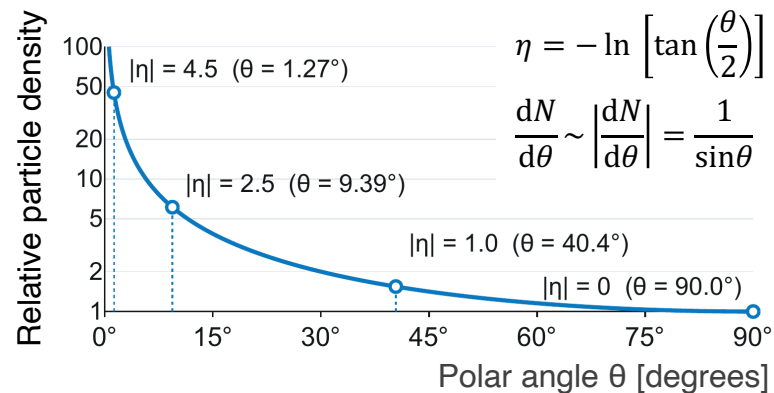
Slice showing CMS sub-detectors and how particles interact with them

*Radiation length X_0 : mean distance a high-E electron travels before losing 1/e of its energy by bremsstrahlung

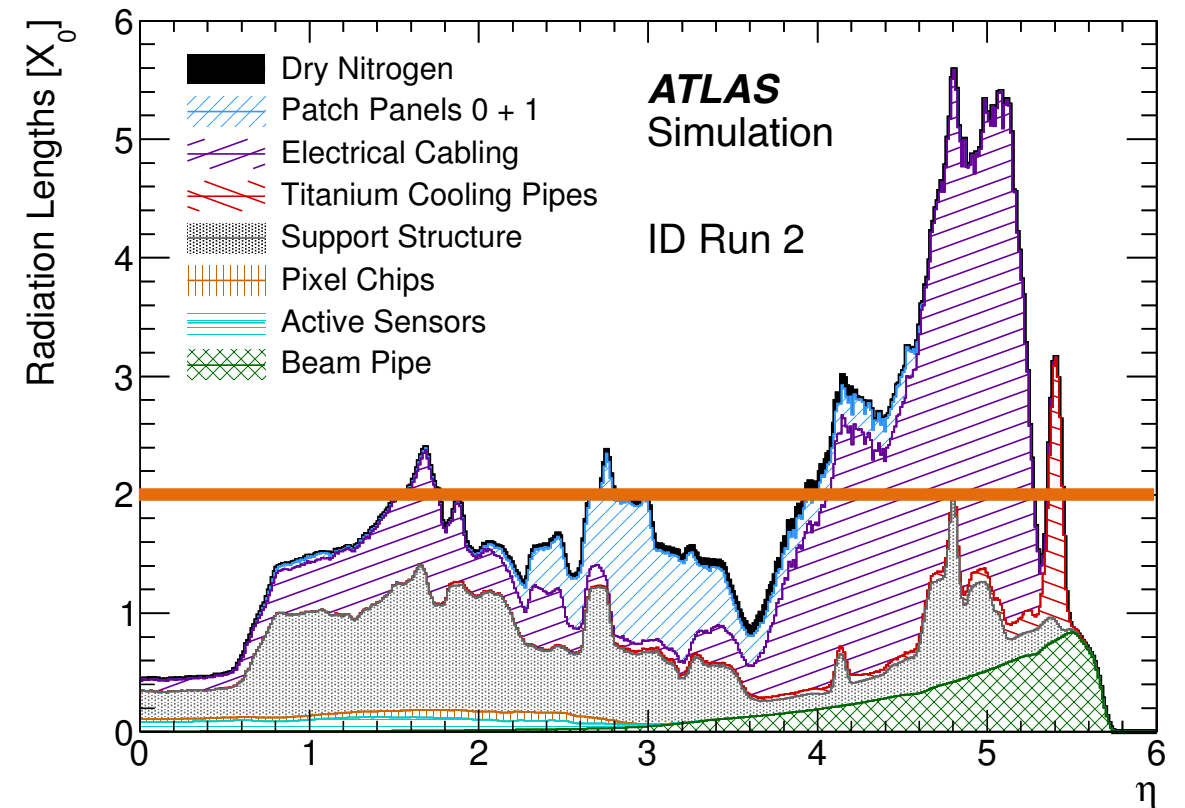
Particle detector basics

Requirements (up to 40 MHz bunch crossing rate, 64 pileup collisions — $2.5 \times$ original design!)

- Radiation hard inner tracker (silicon with large S/B) and forward calorimeters



- High-granular, low-material ($\sigma_{MS} \propto \sqrt{X/X_0}$) tracker*
- High-resolution tracking and vertexing close to collision point
(b/c-hadrons, τ -lepton have lifetimes: $O(1.5)$, $O(0.5)$, 0.3 ps $\rightarrow$ decay lengths ($\beta\gamma c\tau = (p/m)c\tau$) of 1.7, 1.5, 1.0 mm at $p = 20$ GeV)
- High-granularity, linear, uniform, low-noise, dense calorimeters
- High-acceptance muon tagging



Material leads to multiple scattering, nuclear interactions, bremsstrahlung, photon conversion

$$P(\gamma \rightarrow e^+e^-) \approx 1 - e^{-7X/9X_0}$$

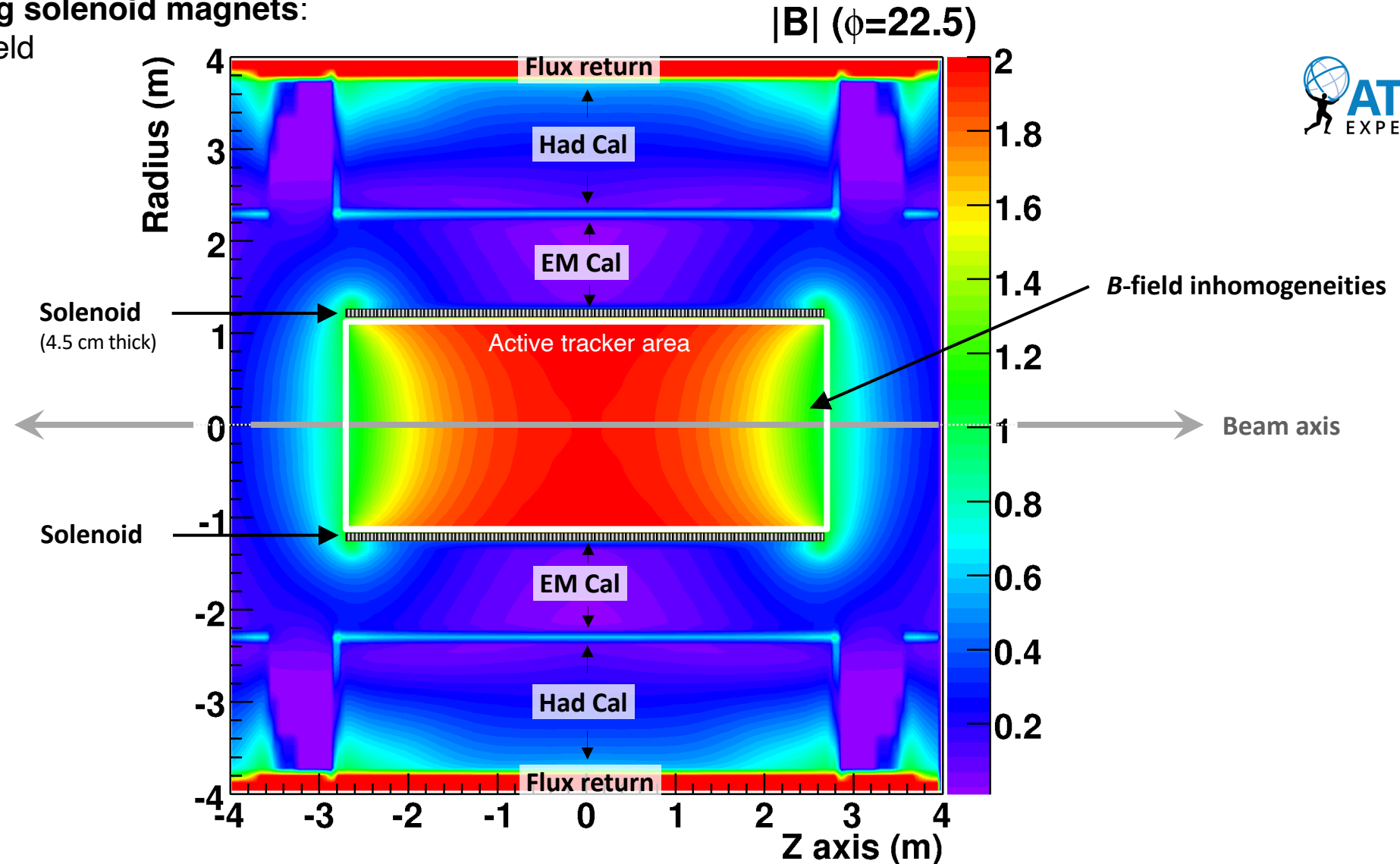
$$\approx 32/69\% \text{ at } \eta = 0/1$$

*Radiation length X_0 : mean distance a high-E electron travels before losing 1/e of its energy by bremsstrahlung

Particle detector basics — Solenoid magnets

Superconducting solenoid magnets:

homogenous B field
simplifies
tracking



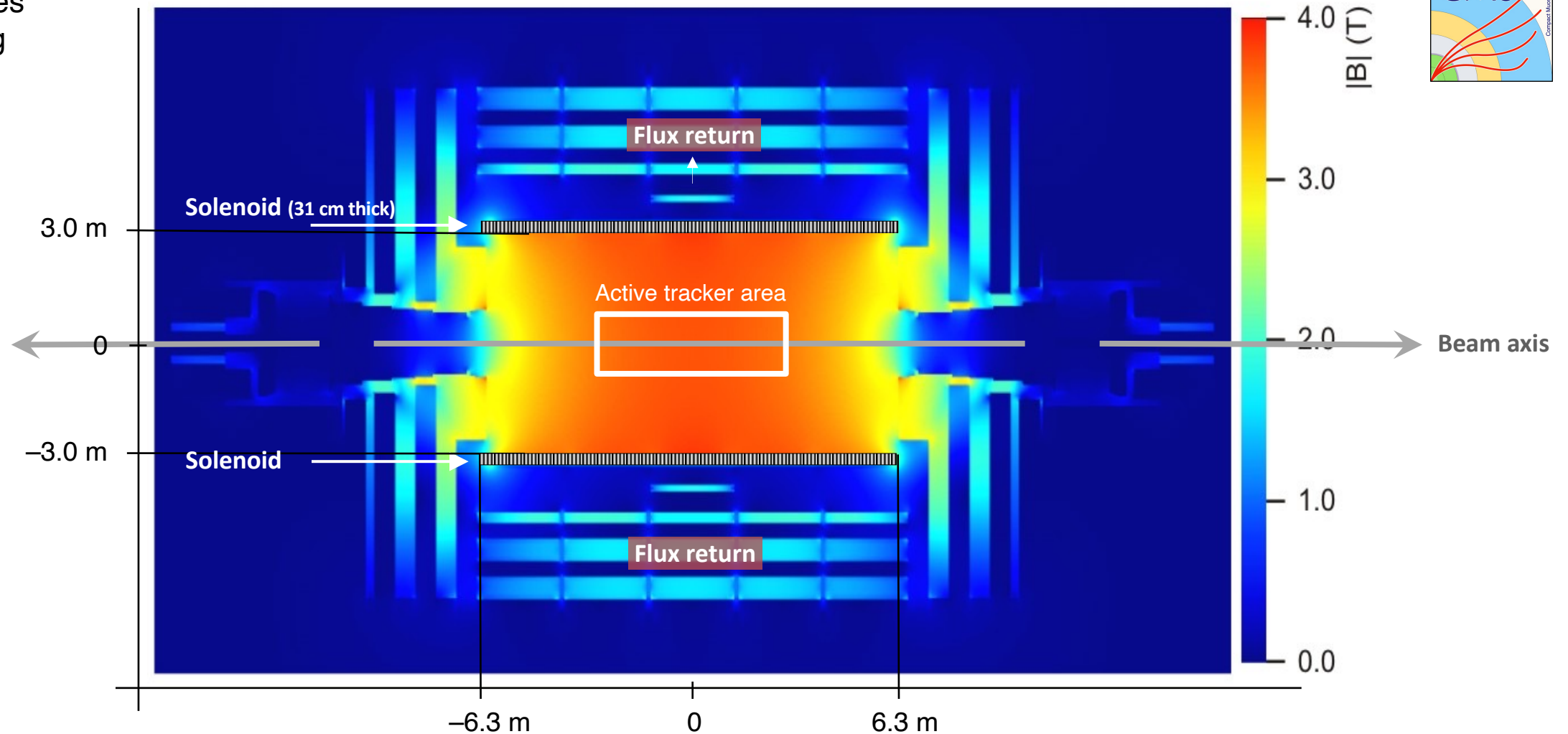
Particle detector basics — Solenoid magnets

Superconducting solenoid magnets:

homogenous B field

simplifies

tracking



Particle detector basics — Solenoid magnets

Superconducting solenoid magnets — stored energies

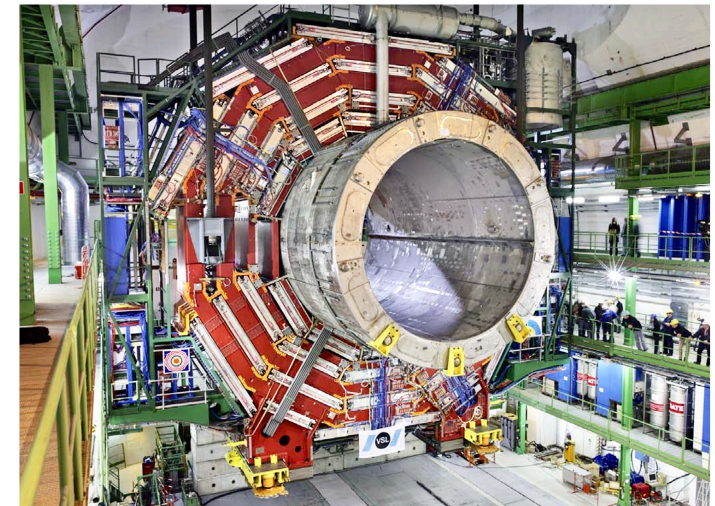
For homogeneous cylindrical solenoid:

$$E = \int \underbrace{\frac{B^2}{2\mu_0}}_{\text{Energy density}} dV \approx \frac{B^2}{2\mu_0} \pi L R^2 = \left\{ \begin{array}{l} \textbf{38 MJ for ATLAS} \text{ (} B = 2 \text{ T, } I = 7.7 \text{ kA, } V \sim 24 \text{ m}^3 \text{)} \\ \text{Extra-thin magnet (before calorimeter, } 0.66 X_0 \text{),} \\ \text{single winding, weight : 5.5 tons} \\ \\ \textbf{2'400 MJ for CMS} \text{ (} B = 3.8 \text{ T, } I = 18.2 \text{ kA, } V \sim 353 \text{ m}^3 \text{)} \\ \text{Thicker, stabilised magnet, four windings (} 3.9 X_0 \text{),} \\ \text{weight: 220 tons (} 40 \times \text{ ATLAS solenoid)} \end{array} \right.$$

$\left[\frac{\text{T}^2 \text{A}^2}{\text{N}} \right] = \left[\frac{\text{J}}{\text{m}^3} \right] = \left[\frac{\text{N} \cdot \text{m}}{\text{m}^3} \right] = [\text{Pa}]$

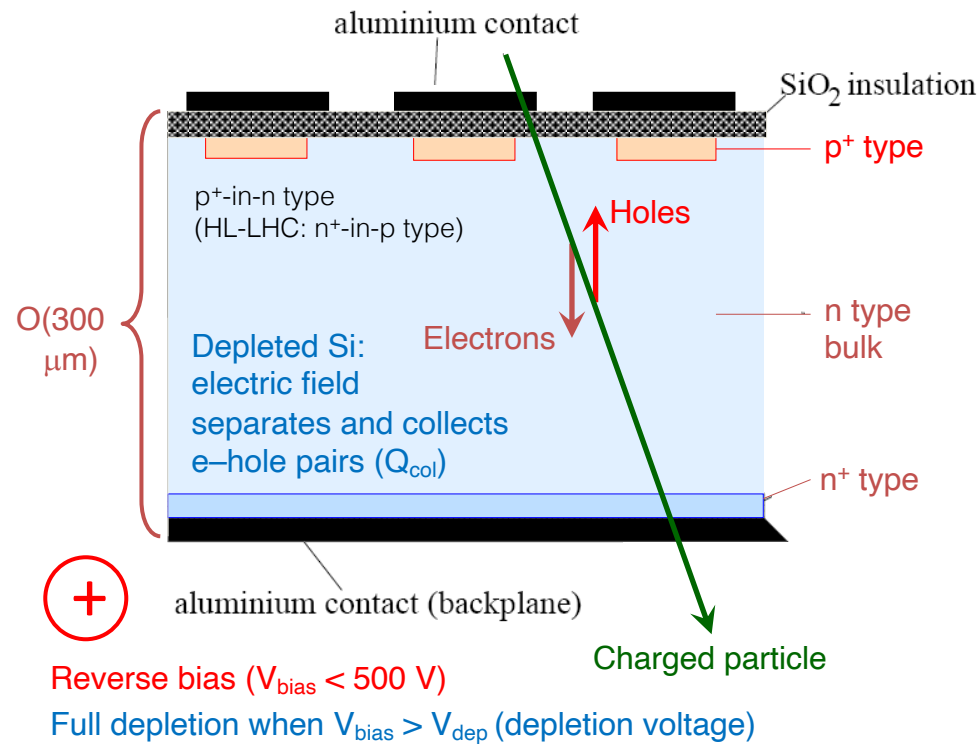
The 3.5 times larger magnetic pressure $B^2/(2\mu_0)$ per unit volume in CMS requires substantial mechanical reinforcement: the magnetic field is effectively trying to expand the coil

Insertion of ATLAS and CMS solenoids



Particle detector basics — Charged particle measurement

Tracking (inner tracker & muon system, exploiting ionization signal)



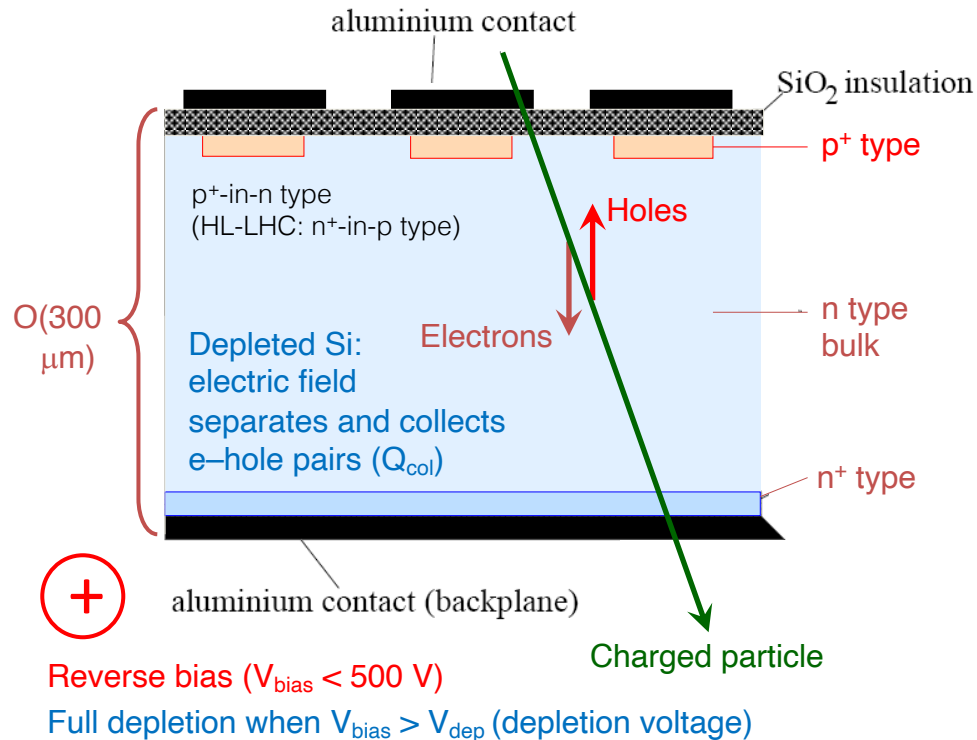
High Si density and low e-hole creation energy (3.6 eV vs ~36 eV in gas)
enable thin detectors with sufficient signal

High mobility of charge carriers in Si allows fast charge collection (~5 ns for e⁻)

Leakage current (I_{leak}) from thermally generated carriers suppressed by cooling

Particle detector basics — Charged particle measurement

Tracking (inner tracker & muon system, exploiting ionization signal)



After irradiation

$V_{dep} \uparrow$
(effective doping changes)

$Q_{col} \downarrow$
(carrier trapping)

$I_{leak} \uparrow$
(thermal carrier generation at radiation defects)



Bias recovery is limited by leakage current power ($P = I_{leak} \cdot V_{bias}$), which can cause thermal runaway*

High Si density and low e-hole creation energy (3.6 eV vs ~36 eV in gas) enable thin detectors with sufficient signal

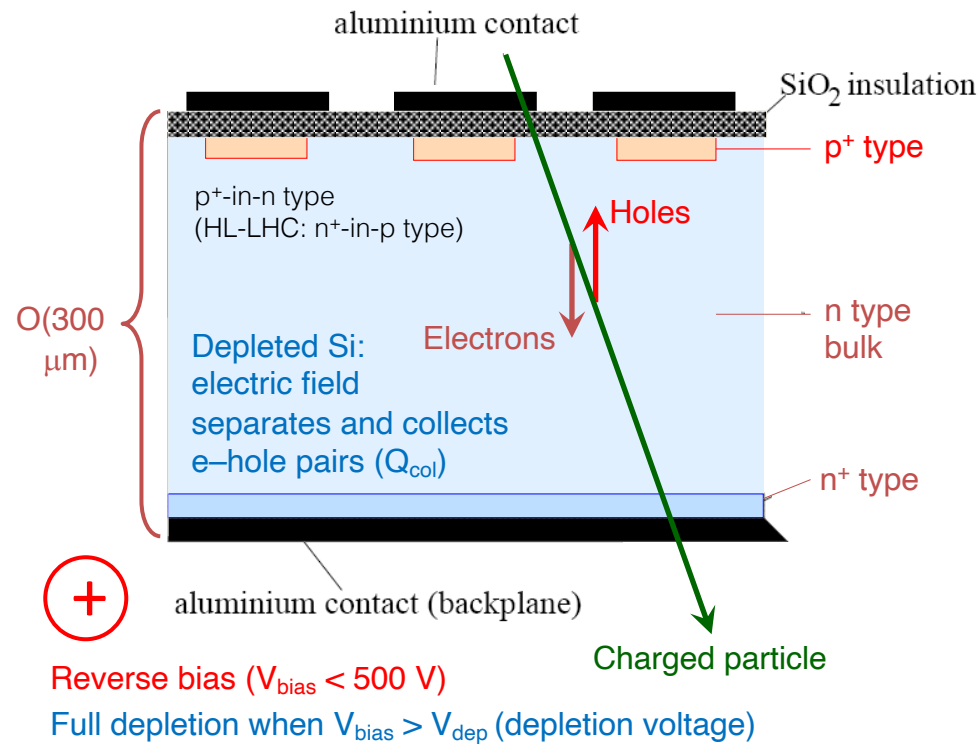
High mobility of charge carriers in Si allows fast charge collection (~5 ns for e⁻)

Leakage current (I_{leak}) from thermally generated carriers suppressed by cooling

*Thermal runaway: sensor heating increases leakage current and power dissipation, causing still more heating unless sufficient cooling stabilises the system

Particle detector basics — Charged particle measurement

Tracking (inner tracker & muon system, exploiting ionization signal)

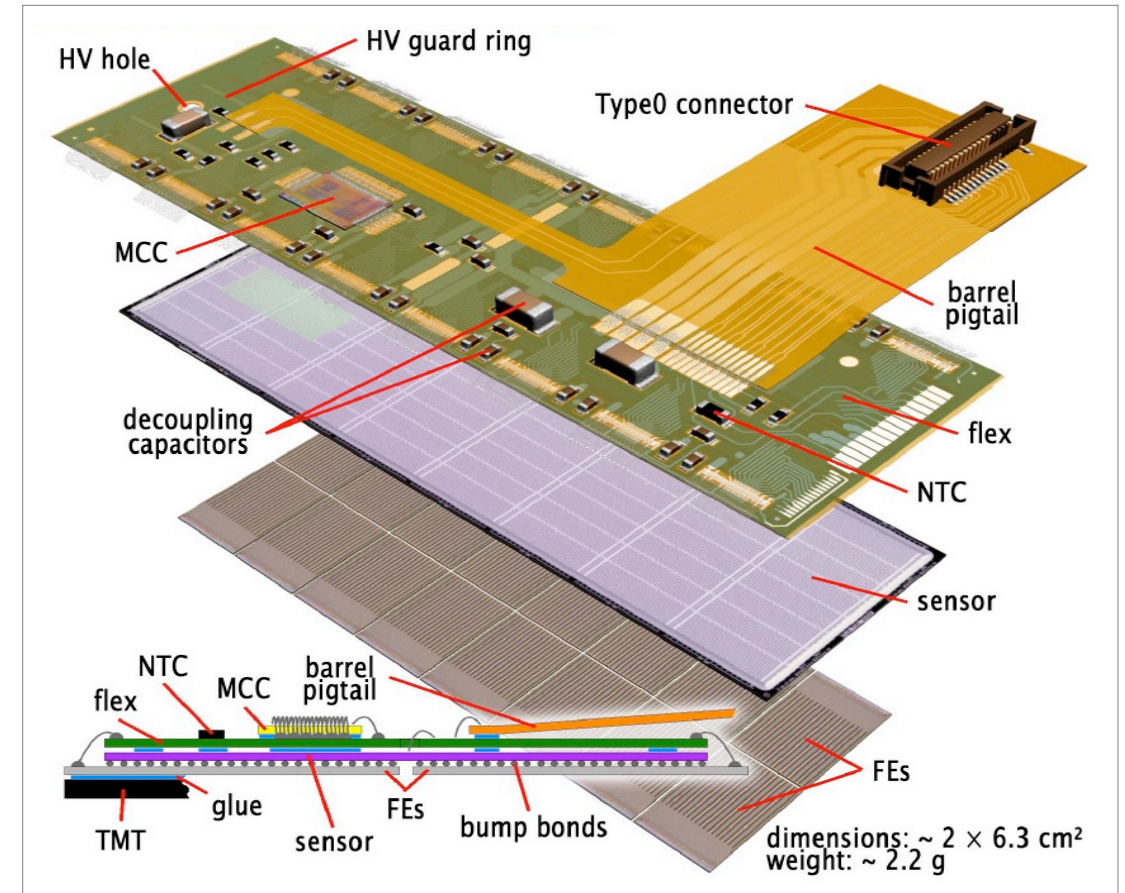


High Si density and low e-hole creation energy (3.6 eV vs $\sim 36 \text{ eV}$ in gas) enable thin detectors with sufficient signal

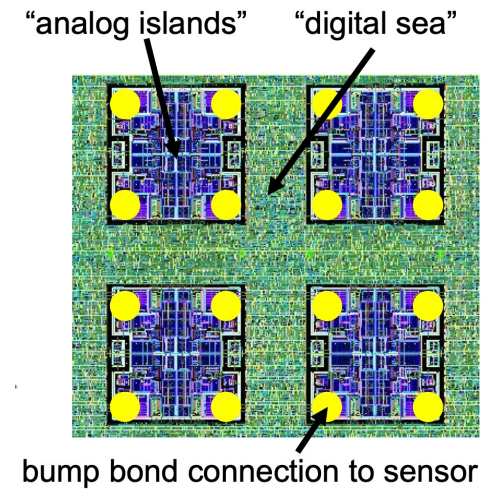
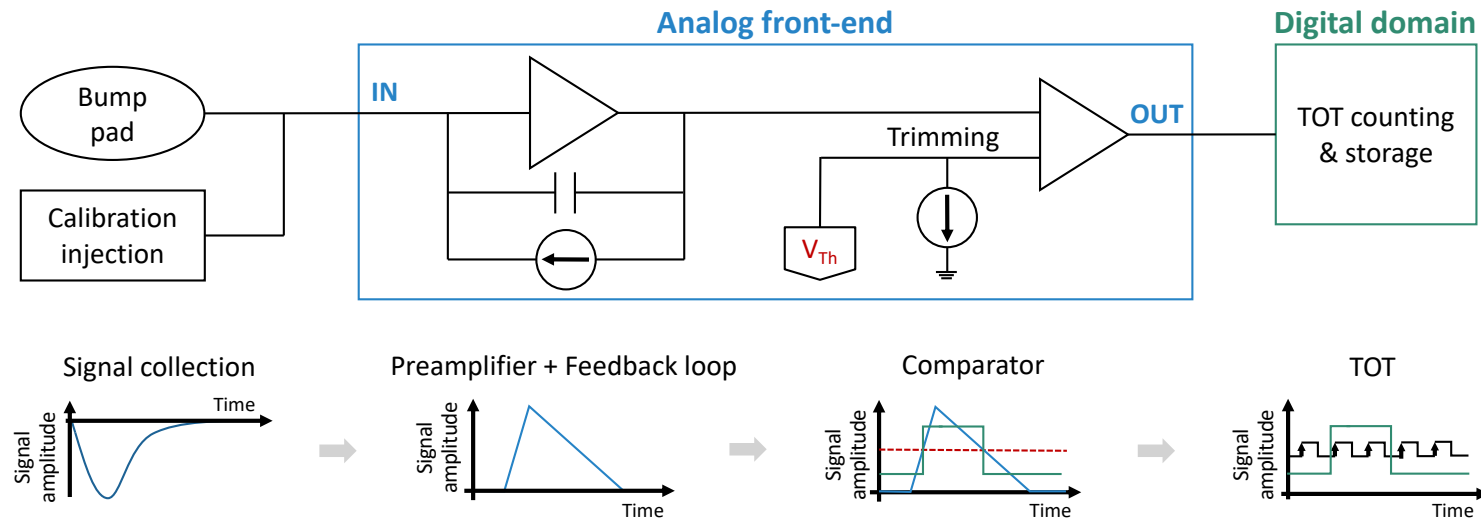
High mobility of charge carriers in Si allows fast charge collection ($\sim 5 \text{ ns}$ for e^-)

Leakage current (I_{leak}) from thermally generated carriers suppressed by cooling

Example: ATLAS silicon pixel module (16 FE chips, 2880 channels / chip)
("hybrid" technology: separate FE chips and sensor bump-bonded together)



Particle detector basics — Charged particle measurement



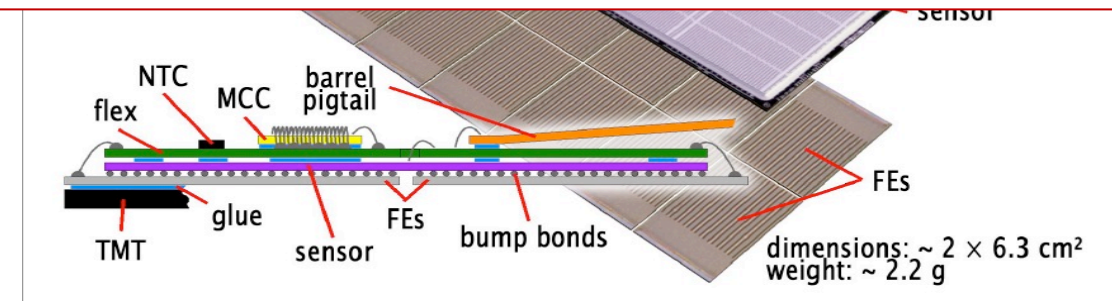
Readout ASIC combines an analog front-end (charge-sensitive preamplifier, feedback, discriminator, threshold tuning and calibration injection) that converts the tiny sensor charge into a digital hit signal, followed by digital logic for TOT digitization, time stamping, buffering, triggering and data readout [\[link to figure, RD53 chip\]](#)

+ aluminium contact (backplane)
Reverse bias ($V_{\text{bias}} < 500 \text{ V}$)
Full depletion when $V_{\text{bias}} > V_{\text{dep}}$ (depletion voltage)
Charged particle

High Si density and low e–hole creation energy (3.6 eV vs ~36 eV in gas) enable thin detectors with sufficient signal

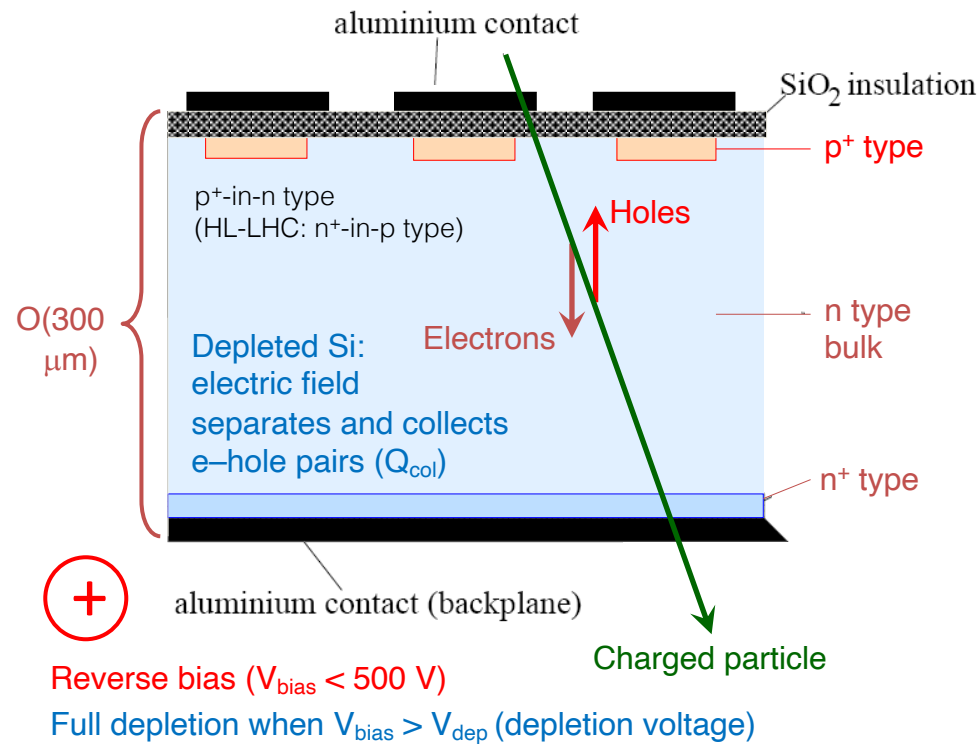
High mobility of charge carriers in Si allows fast charge collection (~5 ns for e[–])

Leakage current (I_{leak}) from thermally generated carriers suppressed by cooling



Particle detector basics — Charged particle measurement

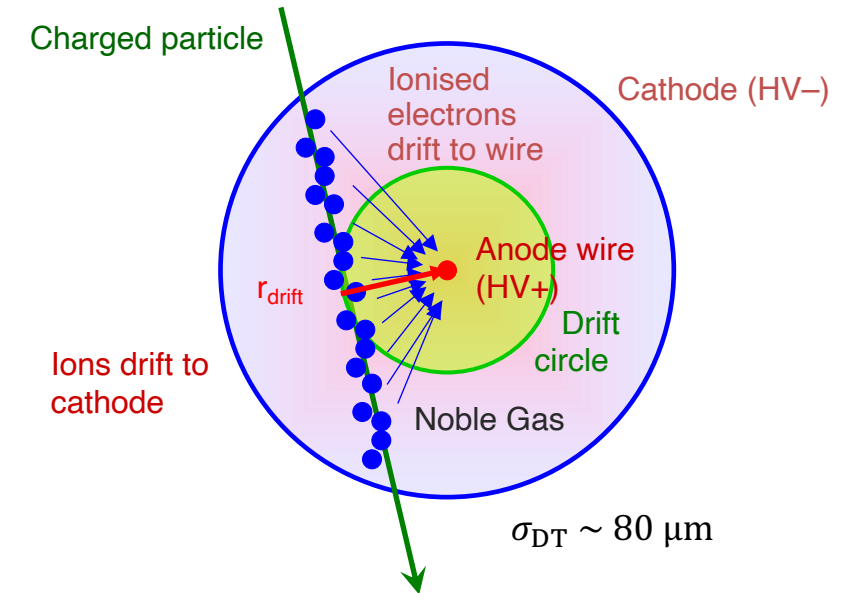
Tracking (inner tracker & muon system, exploiting ionization signal)



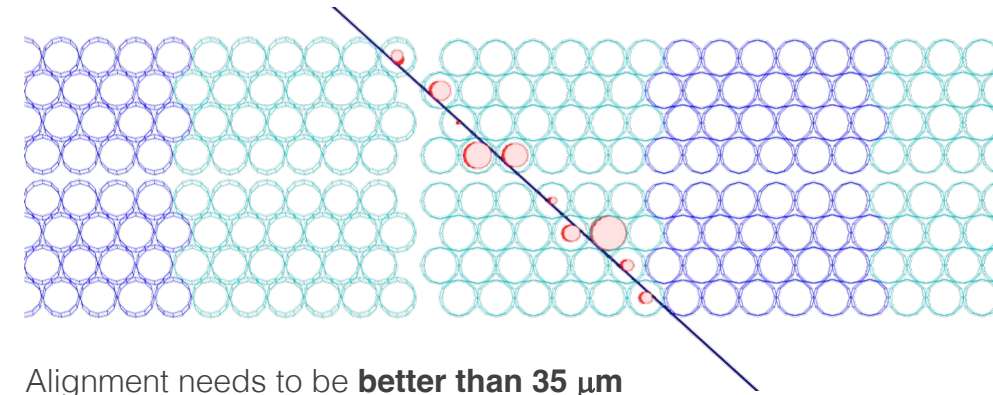
High Si density and low e-hole creation energy (3.6 eV vs $\sim 36 \text{ eV}$ in gas) enable thin detectors with sufficient signal

High mobility of charge carriers in Si allows fast charge collection ($\sim 5 \text{ ns}$ for e^-)

Leakage current (I_{leak}) from thermally generated carriers suppressed by cooling



Example: simulated muon in MDTs



Particle detector basics — Charged particle measurement

Momentum reconstruction of a charged track in a magnetic field (B)

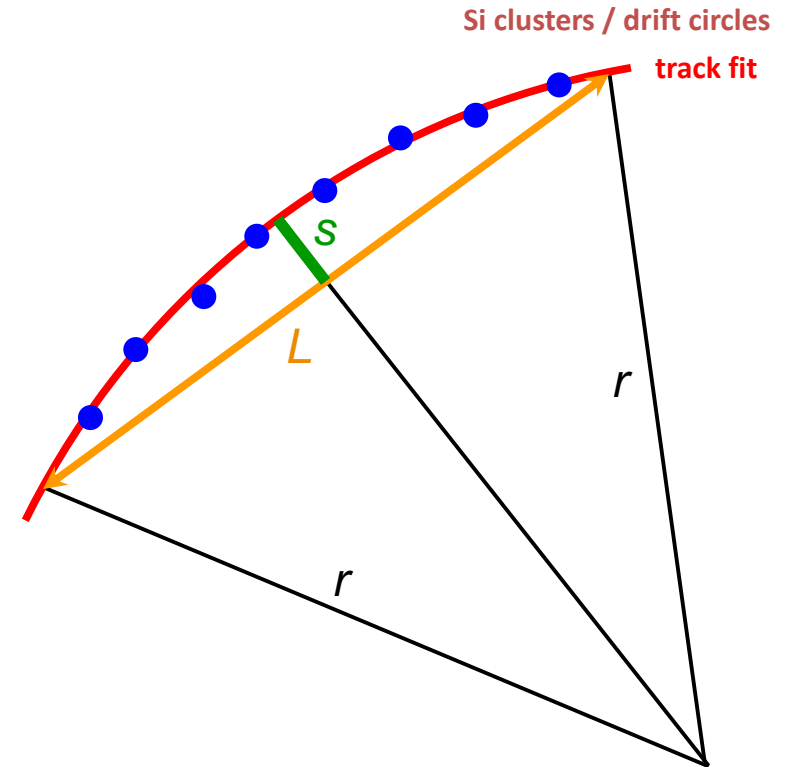
Obtain r and p_T from measurement of sagitta:

$$r \approx \frac{L^2}{8s} \quad (\text{if } s \ll L) \quad \Rightarrow \quad p_T [\text{GeV}] \approx 0.3 \cdot r [\text{m}] \cdot B [\text{T}]$$

Momentum resolution:

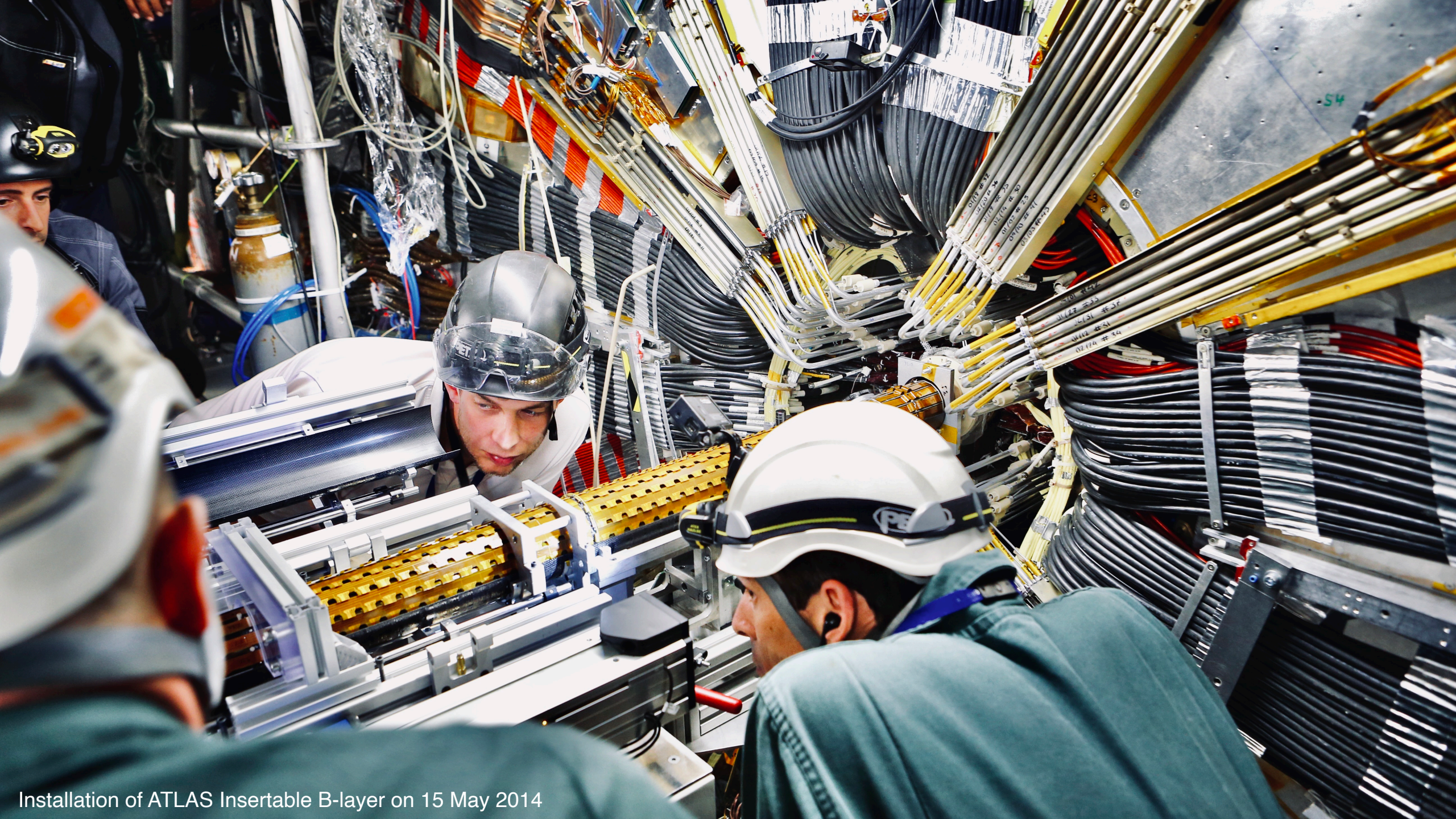
$$\frac{\sigma(p_T)}{p_T} \approx \frac{p_T}{0.3 \cdot B \cdot L^2} \cdot \frac{\sigma_{\text{hit}}}{\sqrt{N_{\text{eff}}}} \oplus \frac{0.045}{B \cdot L} \sqrt{\frac{X}{X_0}}$$

Multiple scattering term
(central tracks, $\beta = 1$)

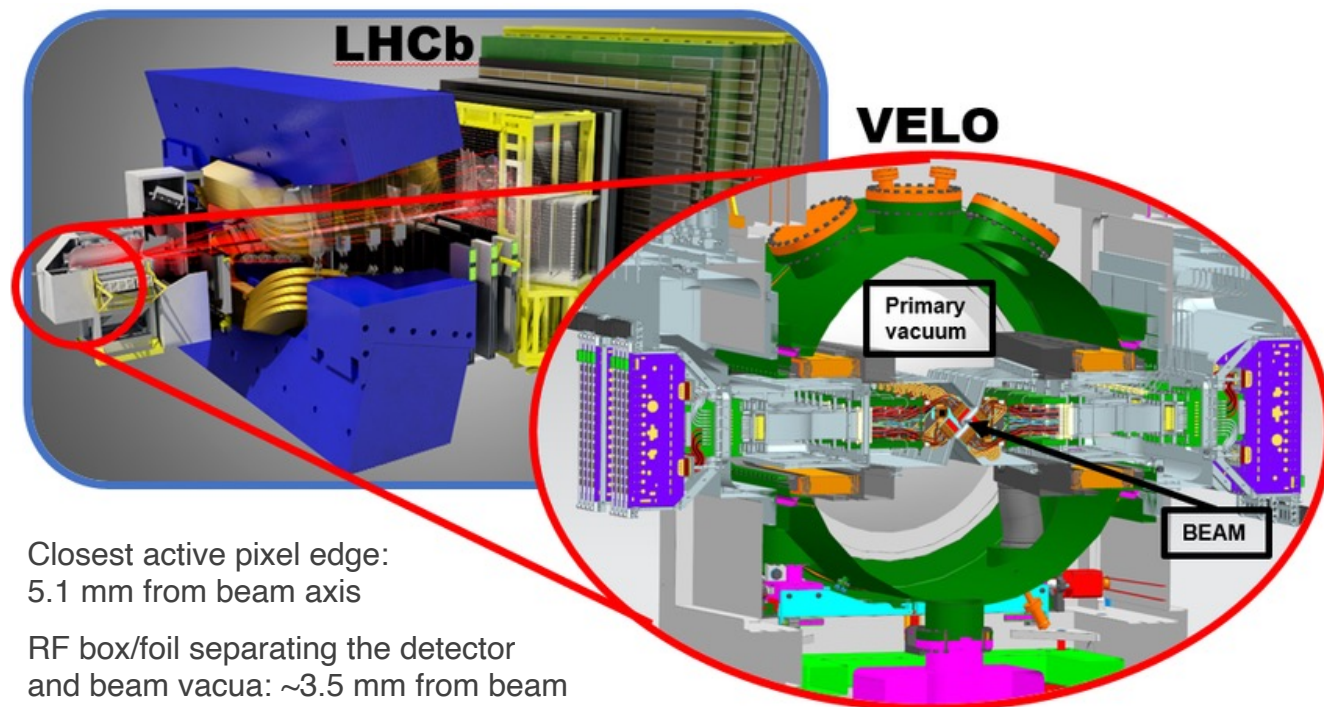
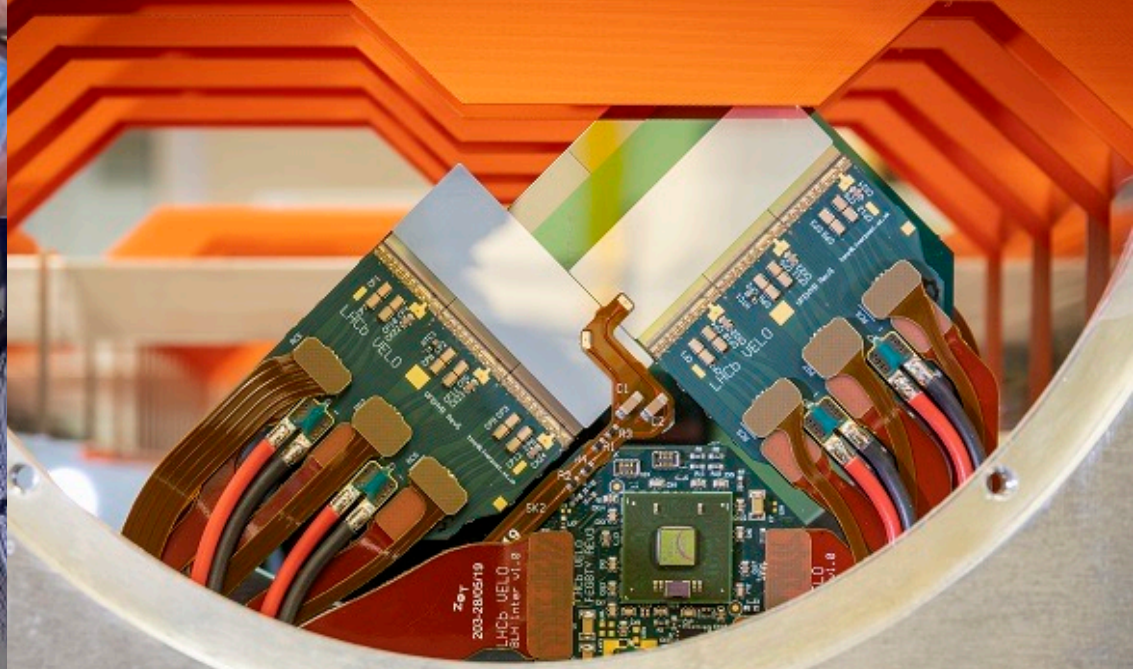
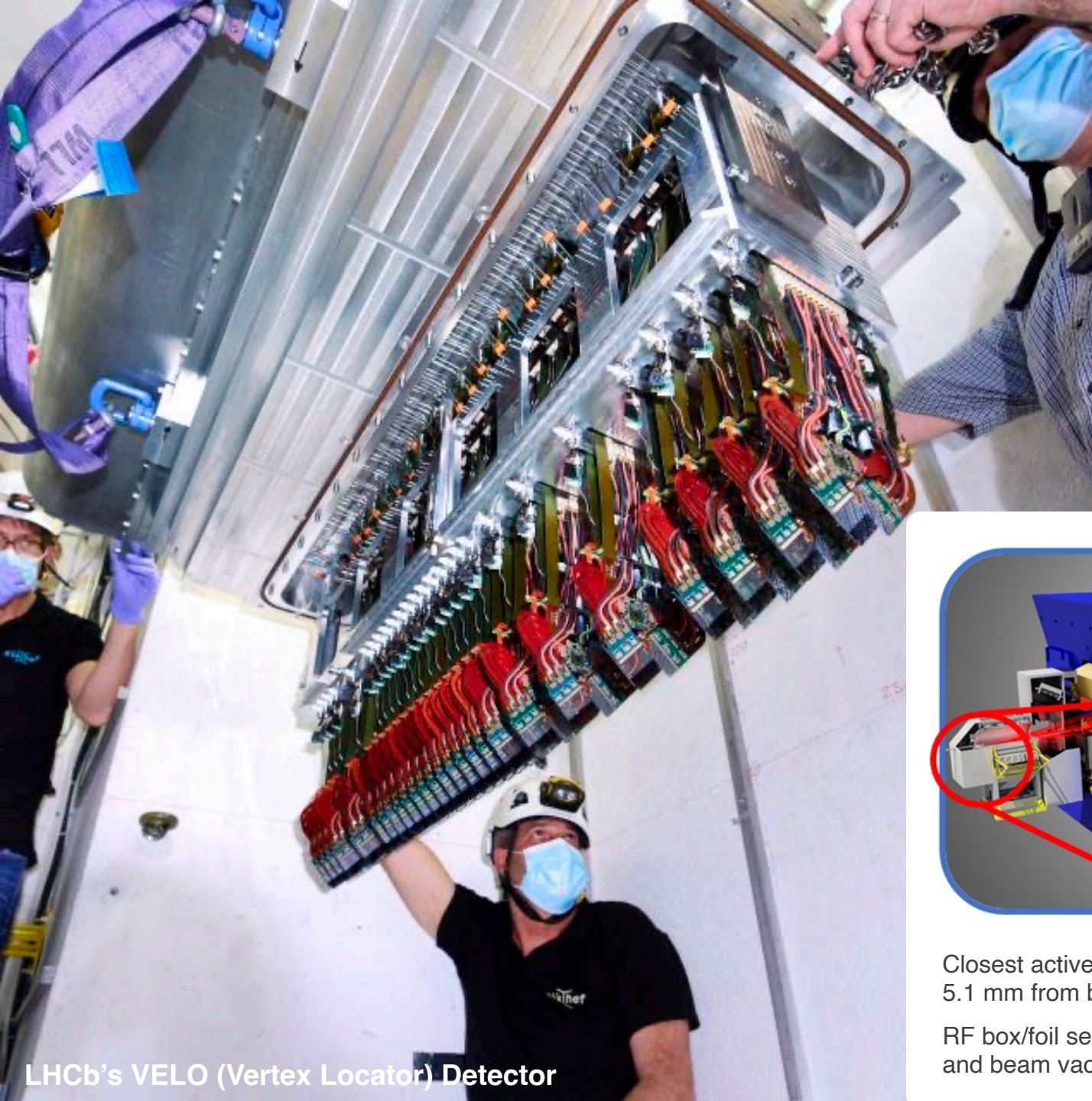


Track fitting in LHC environment is very challenging

- Several thousand tracks per proton-bunch collision, up to 20k tracks for a central PbPb collision
- Must handle ambiguities, hit overlaps, multiple scattering, bremsstrahlung, multiple vertices, ...
- Track fitters take Gaussian noise (e.g., Kalman filter) and non-Gaussian noise (e.g., Gaussian sum filter) into account
- Fitter must be fast, used in high-level trigger



Installation of ATLAS Insertable B-layer on 15 May 2014



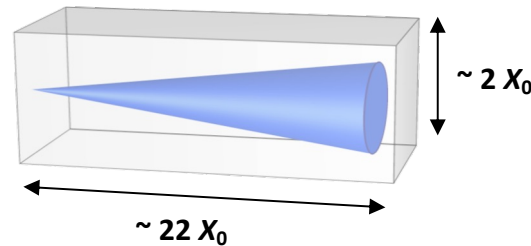
LHCb's VELO (Vertex Locator) Detector

Particle detector basics — EM Calorimeters

Calorimeters (high-energy particle incident on absorber initiates EM cascades and, if hadron, nuclear interactions)

EM: bremsstrahlung and pair production generate lower energy electrons and photons, transverse size given by Molière radius:

$$R_M = \frac{21 \text{ MeV}}{E_c} X_0, \quad E_c \approx \frac{600}{Z + 1.2} \approx 7_{\text{Pb}} \text{ MeV}$$



Calorimeters aim at large $X/X_0 \sim (20 - 30)$

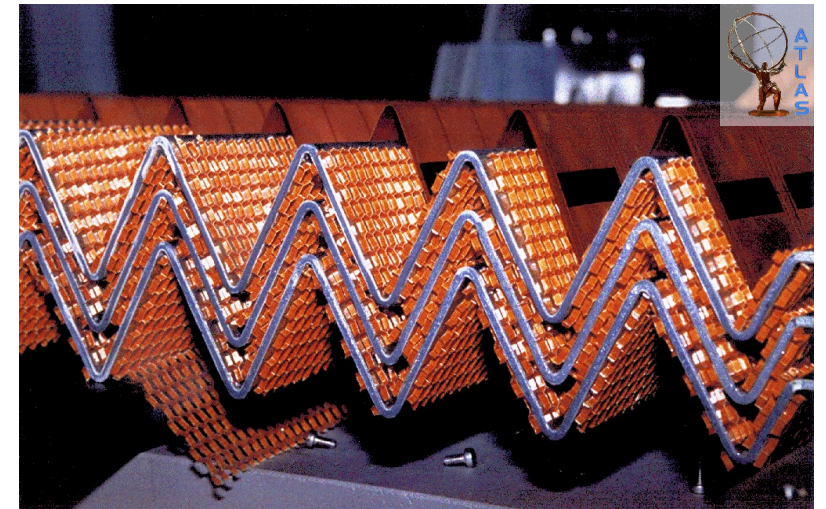
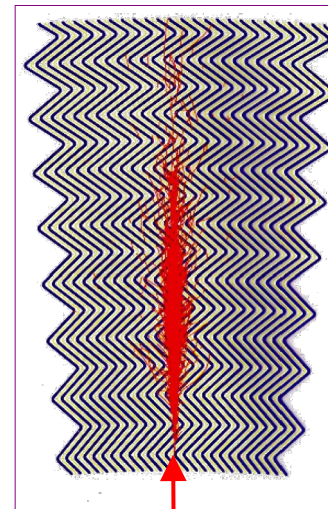
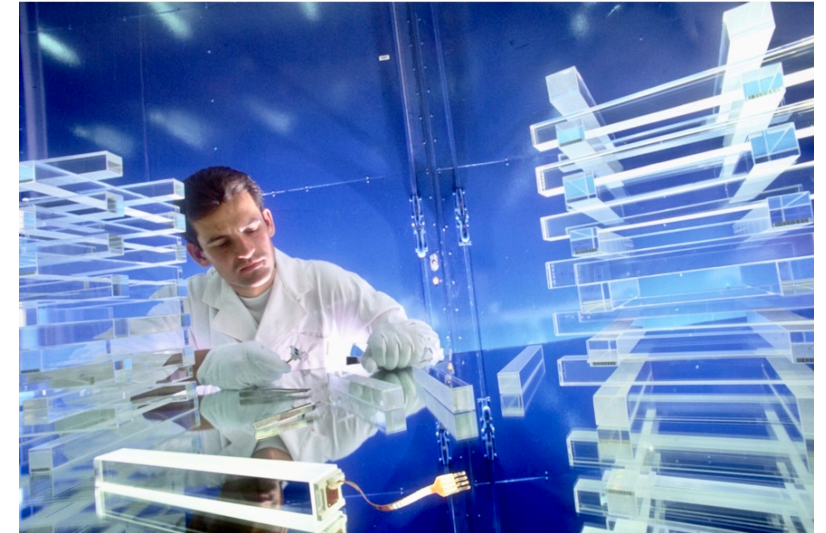
$X_0(\text{Pb, PbWO}_4) \approx 5.6, 8.9 \text{ mm}$

CMS: PbWO_4 scintillation crystal calorimeter

- Entire shower in active detector material, no particles lost in passive absorber, high granularity and resolution, light yield sensitive to radiation
- No longitudinal segmentation

ATLAS: LAr sampling calorimeter (cryogenic T at 87 K)

- Passive Pb absorber (1.1–1.5 mm thick [barrel]) inter-leaved with active detector material (liquid argon), high granularity, linearity, uniformity, radiation hard, medium resolution
- Longitudinal segmentation (3 layers)



Particle detector basics — EM Calorimeters

A calorimeter converts a particle shower into a signal proportional to energy. Since the number of shower particles $N \propto E$, counting fluctuations give $\sigma(E)/E \propto 1/\sqrt{E}$.

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E [\text{GeV}]}} \oplus b \oplus \frac{c}{E [\text{GeV}]}$$

Stochastic term

$$a/\sqrt{E}$$

Shower and sampling fluctuations

Dominates at intermediate energy

Constant term

$$b$$

Calibration, non-uniformity and leakage

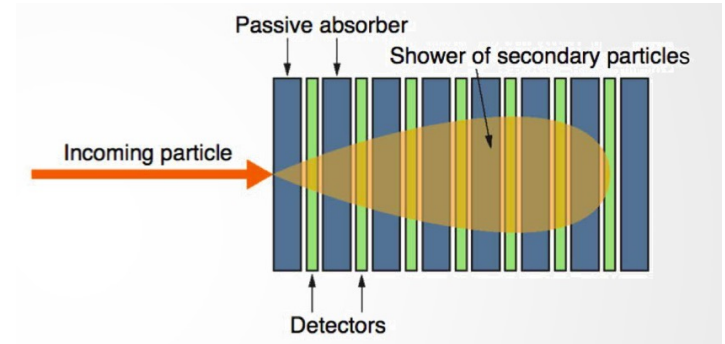
Dominates at high energy

Noise term

$$c/E$$

Electronics and pileup noise

Dominates at low energy



Sampling calorimeter

Only the active layers record energy; thinner absorber layers and a larger active fraction reduce the stochastic term

Representative EM performance: ATLAS Pb/LAr sampling: $10\sim12\% / \sqrt{E} \oplus 0.2\text{--}0.35\%$
CMS PbWO₄ crystals: $3\sim5.5\% / \sqrt{E} \oplus 0.5\%$

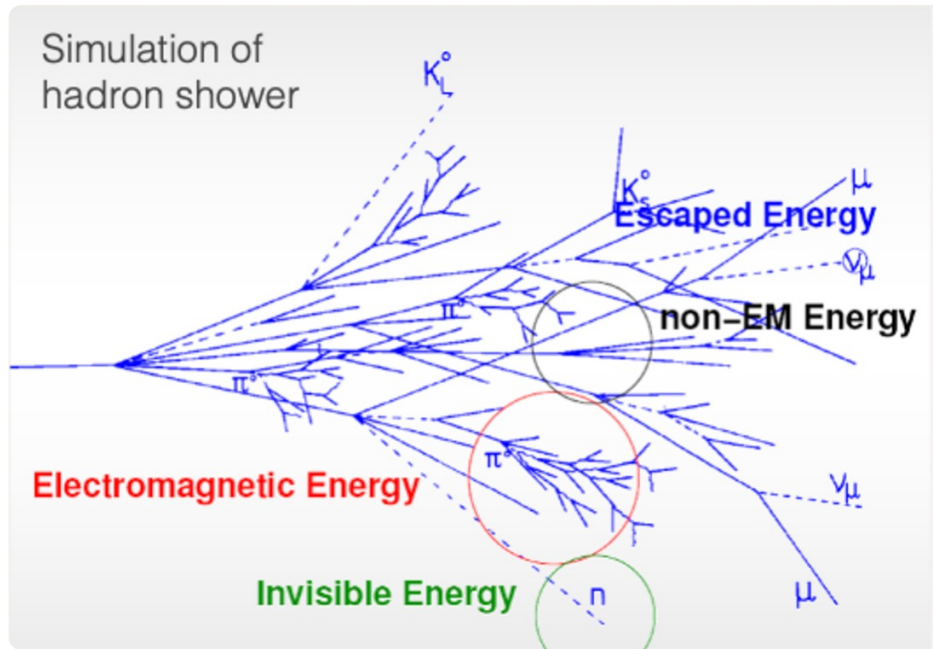
Takeaway: precision improves as $1/\sqrt{E}$ until calibration, uniformity and leakage set a constant-term floor

Particle detector basics — Hadron Calorimeters

HAD: Hadronic calorimeters need to fully absorb hadronic showers (initiated by interaction with nuclei)

Nuclear interaction length λ : mean free path of hadrons between strong collisions

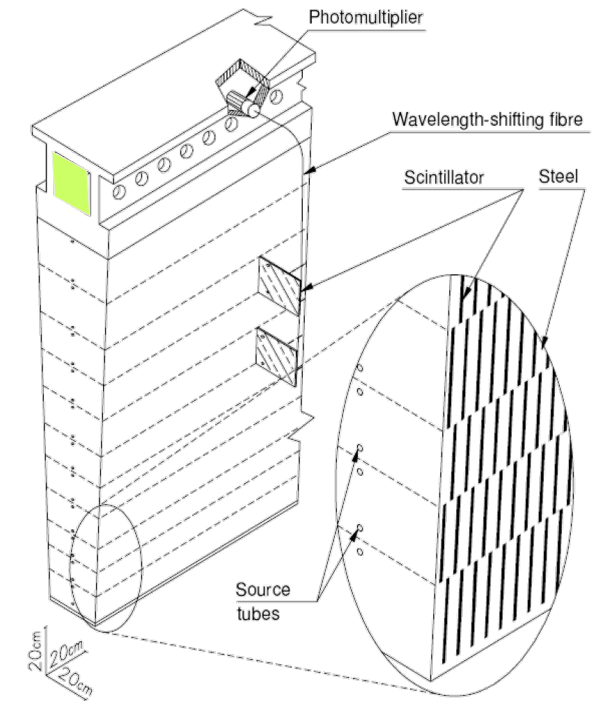
($\lambda > X[X_0] \rightarrow$ HAD showers are later than EM ones, eg, $\lambda_{\text{Fe/brass}} \sim 17 \text{ cm}$, need $\lambda > 10$)



Where the incident energy goes

- **EM component (~50%)**: $\pi^0, \eta \rightarrow \gamma\gamma$; average fraction grows with energy
- **Visible non-EM (~25%)**: ionisation and excitation by charged hadrons
- **Invisible (~25%)**: nuclear breakup and binding energy; slow neutrons
- **Escaped (~2%)**: neutrinos and particles leaving the instrumented volume

Larger fluctuations and invisible energy causes worse energy resolution for HAD showers



ATLAS / CMS

- $\lambda = 9.7\text{--}13.0 / 7.2\text{--}11.0$
- 3 / 1 sampling layers (barrel)
- Pions: stochastic: $\sim 55 / 70\%$, constant: $2.3 / 8.0\%$, noise: small
- Significant improvement in CMS through particle flow (less so for ATLAS)

Sketch of ATLAS Tile Hadronic Calorimeter (steel / scintillator, three radial layers — CMS 1 layer)

A complete ATLAS tile calorimeter module covering 1/64 of the full azimuth

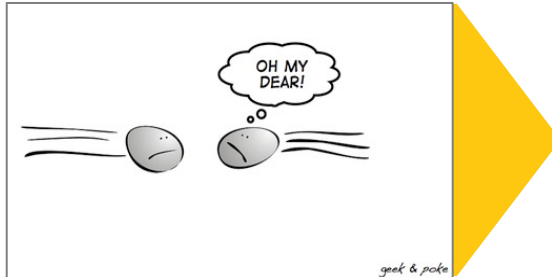




Transverse slice of the CMS detector

The data path in a nut-shell

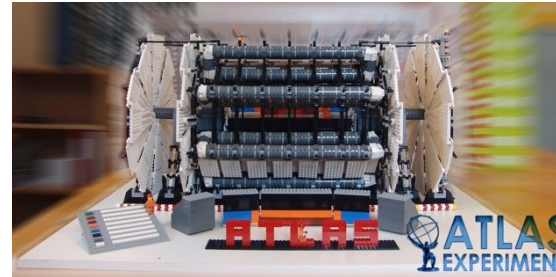
Large Hadron Collider



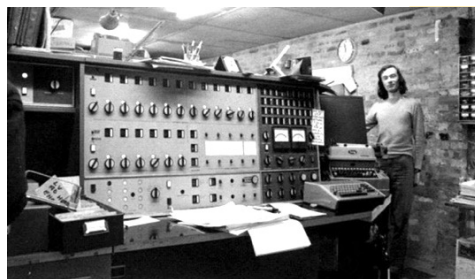
25 ns bunch distance

$$L_{\text{max}} \sim 2.2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$$

LHC Detector



Calibration & Reconstruction



48h calibration & data quality processing, then prompt reconstruction of data in Tier-0

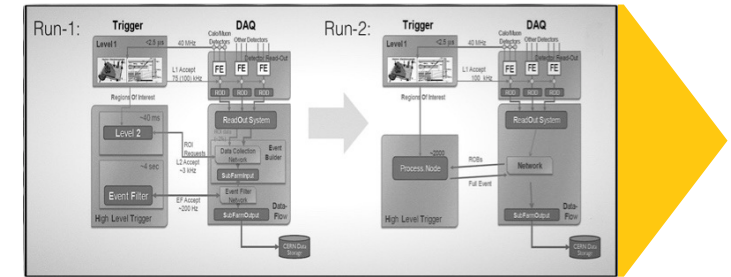
Distributed computing



Production of standardised derived datasets for physics and performance analysis

Also: MC production — O(100 billion) 13 TeV events produced per experiment

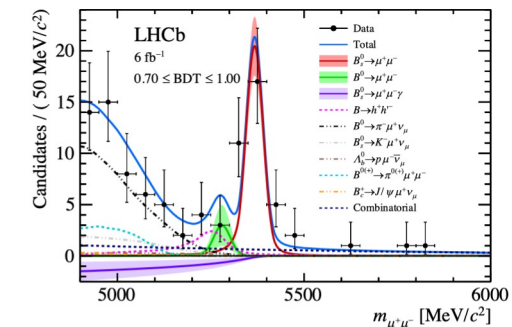
Trigger & Online monitoring



L1 (HW, up to 100 kHz) + HLT (SW, ~3 kHz)

Challenge: low-threshold triggers to keep physics efficiency high, use combined / topological signature information

Physics analysis



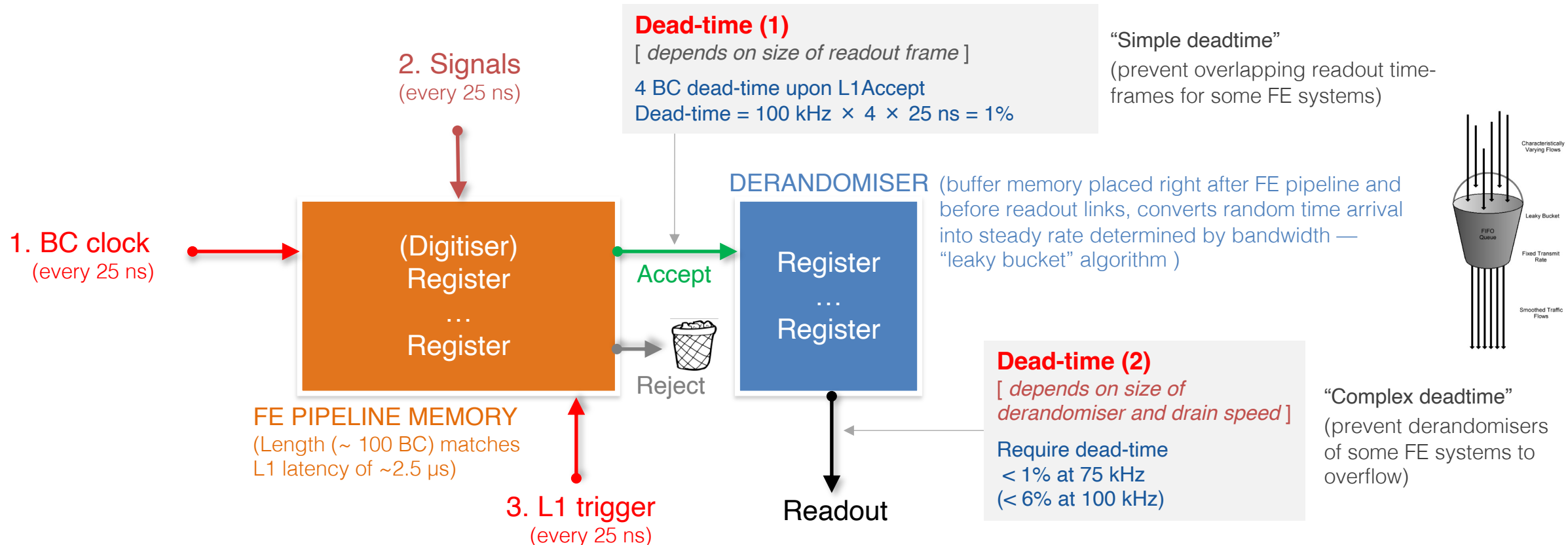
Performance analysis provides standard physics objects with calibrations and uncertainties, unified in analysis release

Physics analyses build upon this groundwork

Particle detector basics — Online event filter

Trigger: $\sim 1 / 10'000$ events kept, necessarily cutting into physics, **mostly driven by Level-1** (~ 100 kHz, driven by readout time)

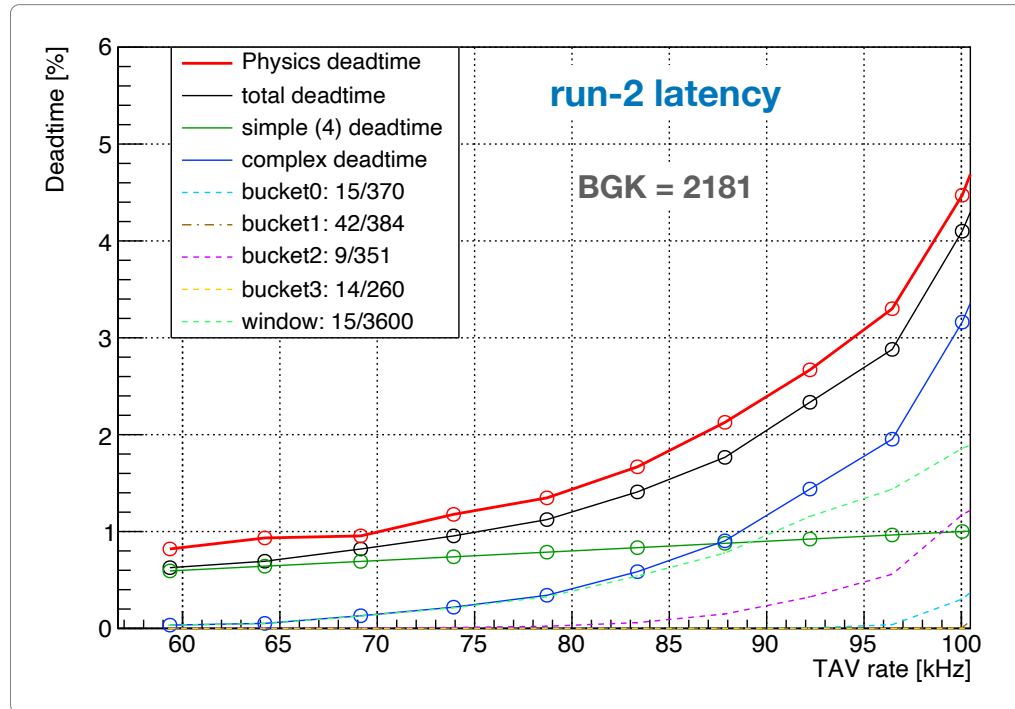
L1 trigger must deliver a decision every BC, despite a latency much longer than the BC period. Achieved by pipelining: the processing is split into steps, each done within one BC, with parallel logic handling different operations. The fixed latency is set by the number of steps plus data transfer times. Input data are fanned out to several processing elements, with data volume decreasing down the pipeline.



Particle detector basics — Online event filter

Trigger: $\sim 1 / 10'000$ events kept, necessarily cutting into physics, **mostly driven by Level-1** (~ 100 kHz, driven by readout time)

L1 trigger must deliver a decision every BC, despite a latency much longer than the BC period. Achieved by pipelining: the processing is split into steps, each done within one BC, with parallel logic handling different operations. The fixed latency is set by the number of steps plus data transfer times. Input data are fanned out to several processing elements, with data volume decreasing down the pipeline.



Move towards more and more powerful global FPGA-based triggers and deeper pipelines, need to watch the latency of the algorithms!

Dead-time (1)

[depends on size of readout frame]

4 BC dead-time upon L1Accept

Dead-time = $100 \text{ kHz} \times 4 \times 25 \text{ ns} = 1\%$

“Simple deadtime”

(prevent overlapping readout time-frames for some FE systems)

DERANDOMISER (buffer memory placed right after FE pipeline and before readout links, converts random time arrival into steady rate determined by bandwidth — “leaky bucket” algorithm)

Accept

Reject

Register
...
Register

Readout

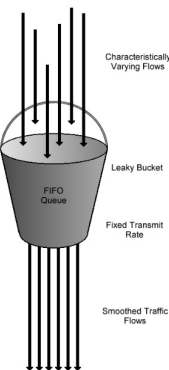
Dead-time (2)

[depends on size of derandomiser and drain speed]

Require dead-time
< 1% at 75 kHz
< 6% at 100 kHz

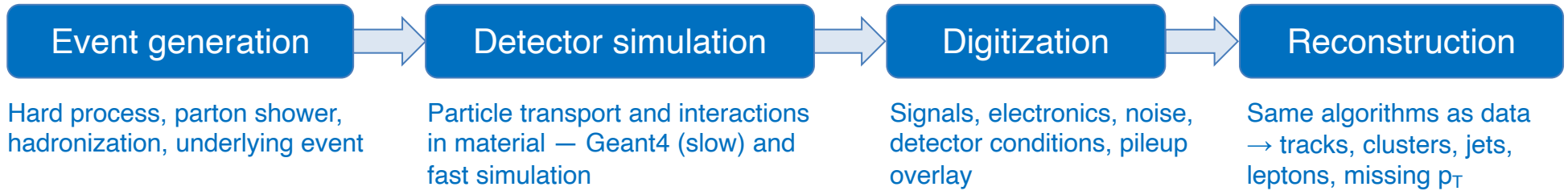
“Complex deadtime”

(prevent derandomisers of some FE systems to overflow)



Monte Carlo simulation is key for analysis

Simulation chain:

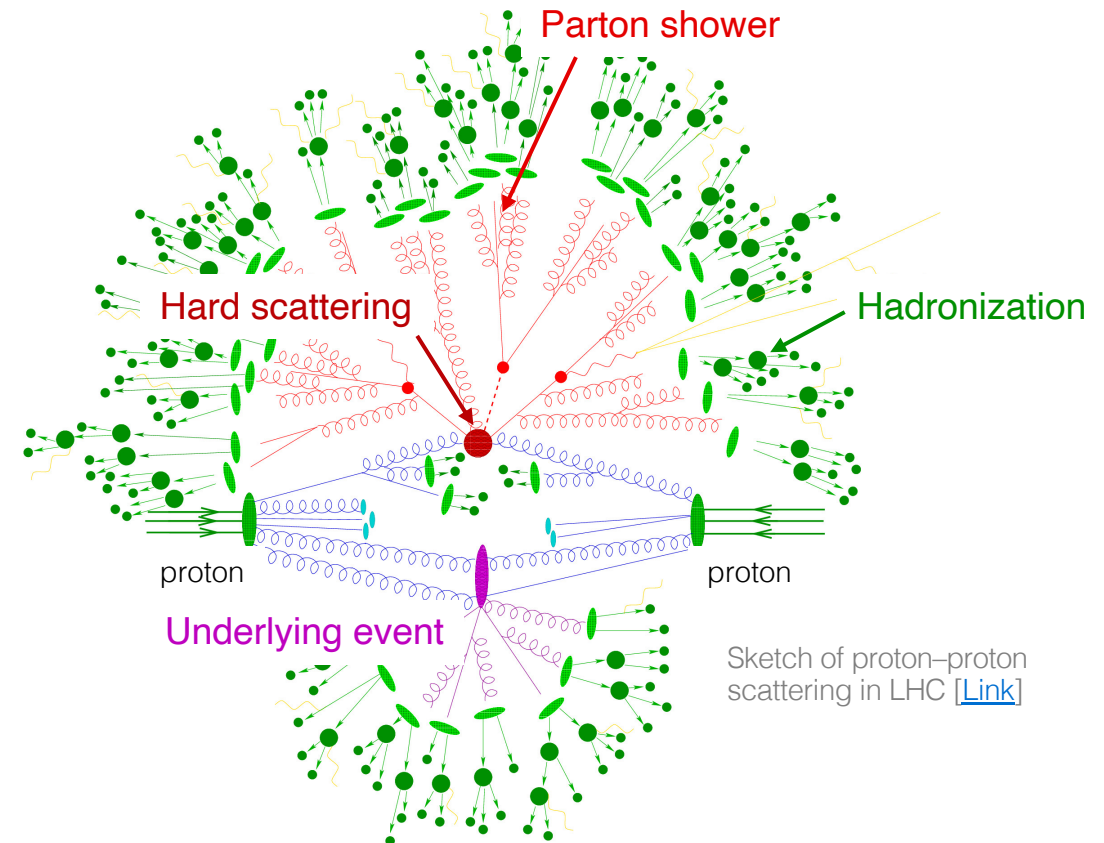


Event generation (defining the *MC truth* record)

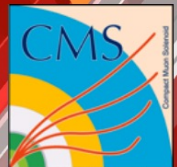
- **Hard process:** PDFs $\otimes$ matrix elements describe the short-distance scattering
- **QCD radiation:** parton shower plus matching / merging with matrix-element jets
- **Soft activity:** multiple-parton interactions and underlying event
- **Hadronization:** formation of hadrons followed by unstable-particle decays

MC is not a single prediction: alternative generators, scales, PDFs, and tunes used to quantify modelling uncertainties

Generator modelling is often a leading systematic uncertainty



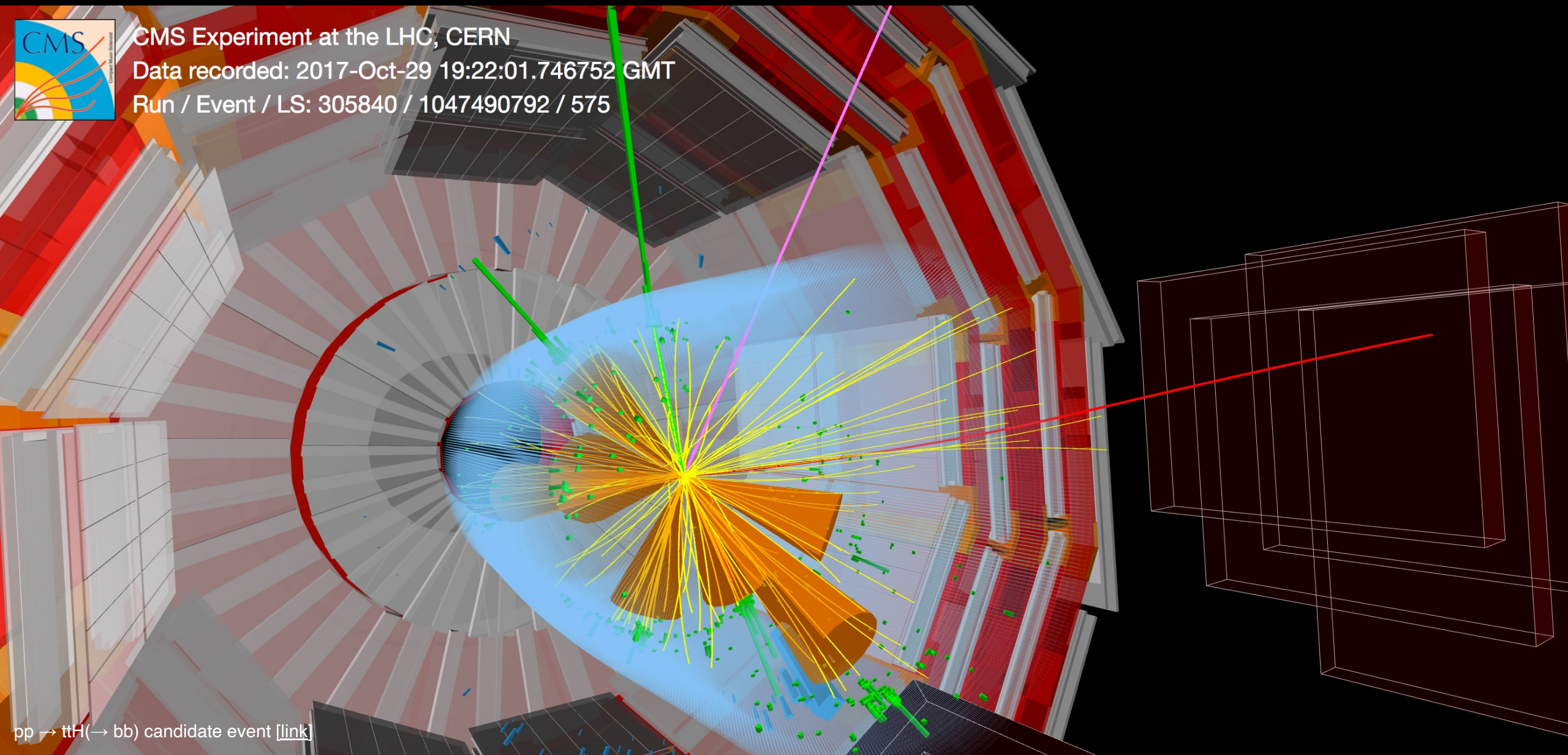
Particle (object) reconstruction, identification, and calibration



CMS Experiment at the LHC, CERN

Data recorded: 2017-Oct-29 19:22:01.746752 GMT

Run / Event / LS: 305840 / 1047490792 / 575

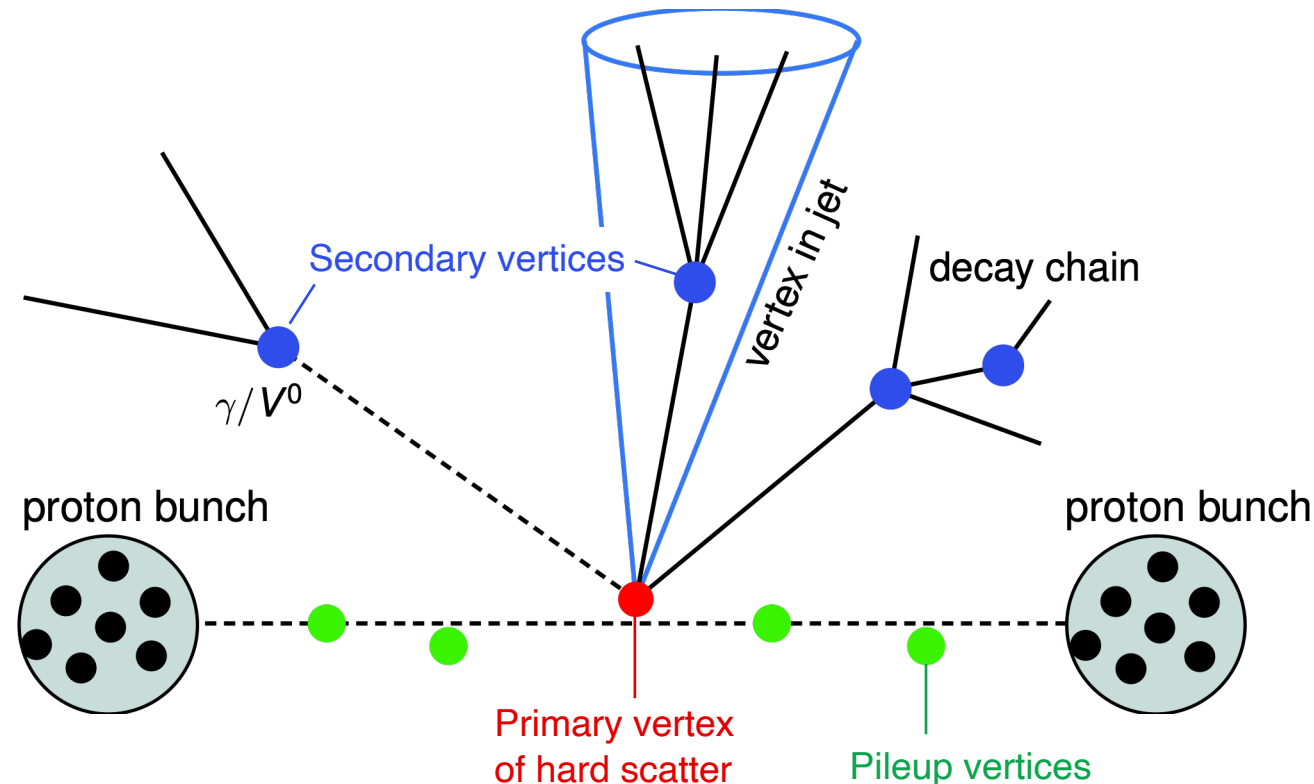


$pp \rightarrow t\bar{t}H(\rightarrow b\bar{b})$ candidate event [\[link\]](#)

Precision objects enable precision measurements — Vertexing

Tracks can be fitted to common vertices (constrained fit taking into account measurement uncertainties)

- The **hard-scatter primary vertex** identifies the pp collision of interest among pileup interactions ($\max \sum_{\text{tracks}} p_T^2$)
- Displaced **secondary and tertiary vertices** reveal decays of long-lived b- and c-hadrons or tau-leptons
- Photon conversions and hadronic interactions create material vertices, identified from their topology and location in the detector



LHC beamspot width in Run 3 (pp) at start of run (Gaussian profile):

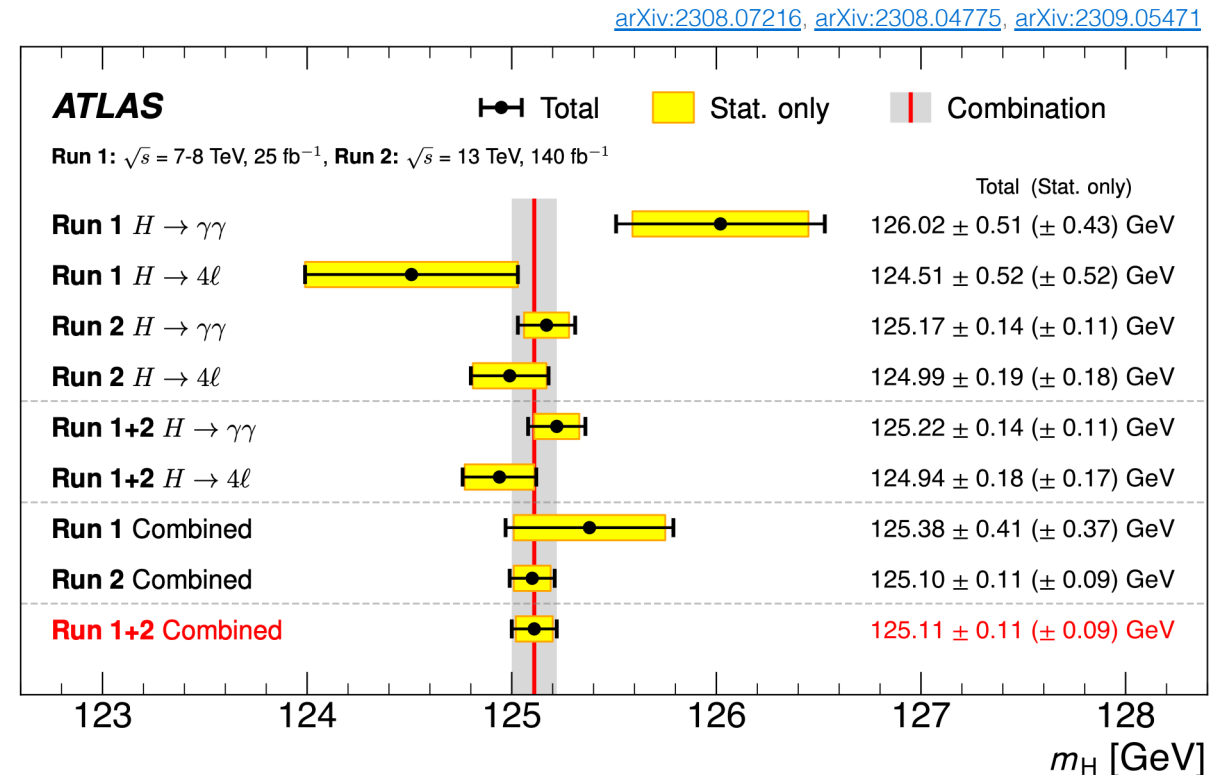
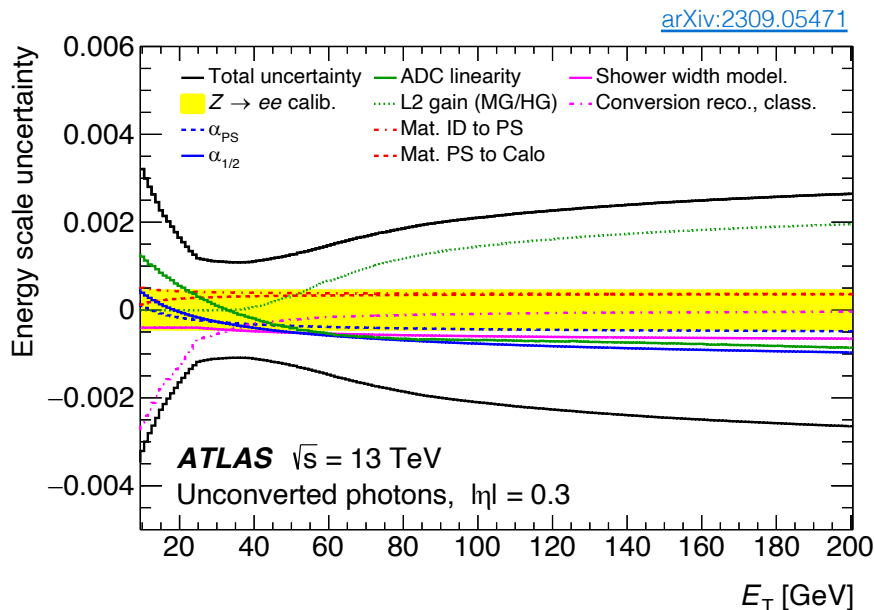
- $\sigma_{x/y} \approx 9 \mu\text{m}$
- $\sigma_z \approx 4 \text{ cm}$

Precision objects enable precision measurements — Electron / γ

Electron: isolated track matched to EM shower, energy measurement in EM calorimeter more accurate, calibrated *in-situ* with $Z, \Upsilon, J/\psi \rightarrow ee$ (“tag & probe”), and E/p . Note: large bremsstrahlung losses $P \propto 1/m^2$, $\Delta E/E \approx 1 - e^{-X/X_0} \approx 40/70\%$ in inner detector at $\eta \sim 0/1$, corrected in tracking

Photon: no track, isolated EM shower, calibrated with $Z \rightarrow ee(\gamma)$, $\pi^0 / \eta \rightarrow \gamma\gamma$ (low energy). Note: large conversion rate: mean free path: $\lambda_{\gamma \rightarrow e^+e^-} \approx 9/7 \cdot X_0 \rightarrow P(\gamma \rightarrow e^+e^-) \approx 1 - \exp(-X/\lambda_{\gamma \rightarrow e^+e^-}) = 1 - \exp(-7X/9X_0) \approx 32/69\%$ at $\eta \sim 0/1$

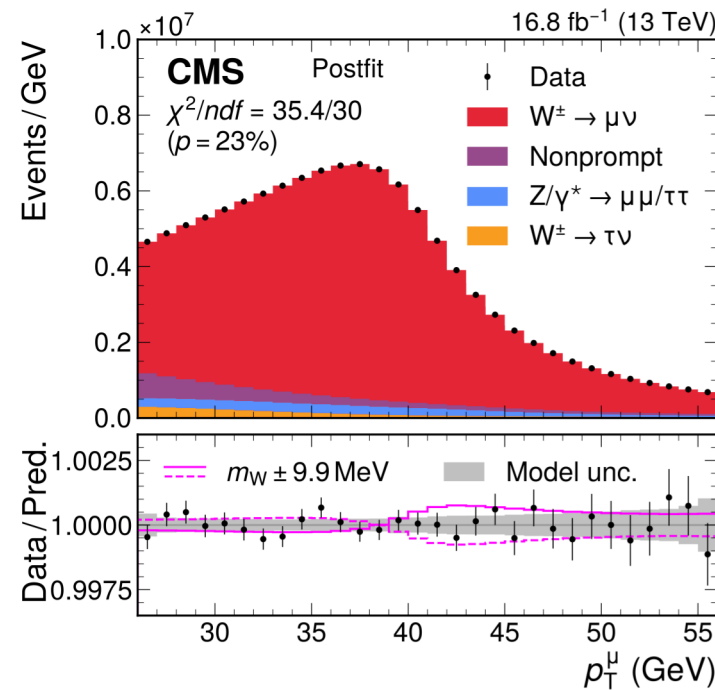
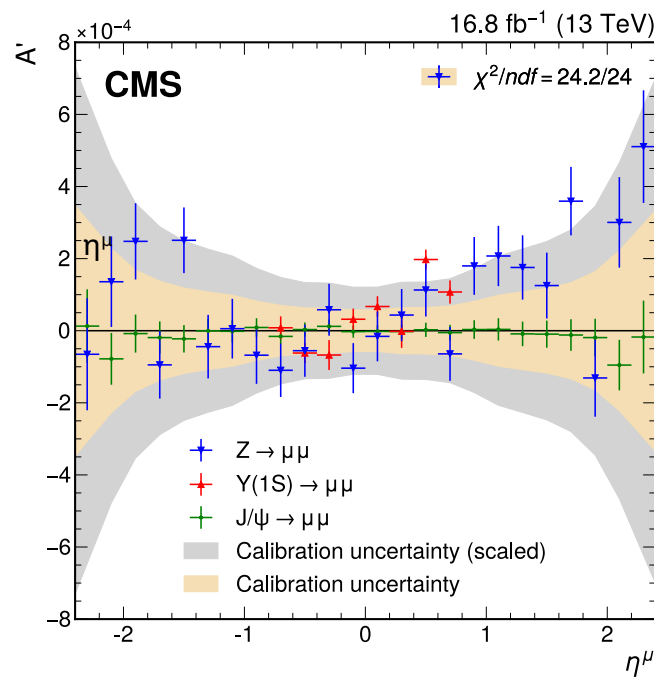
Detailed photon energy calibration enables precise Higgs mass measurement with $H \rightarrow \gamma\gamma$ (0.09%)



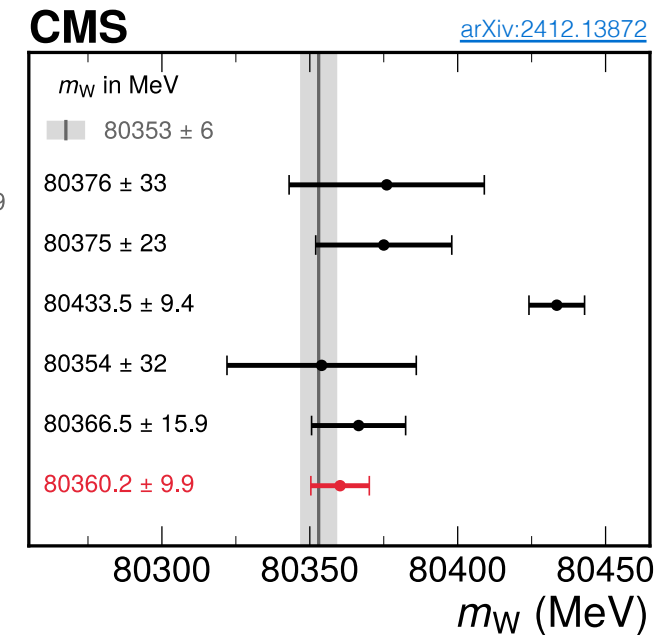
Precision objects enable precision measurements — Muon

Muon: isolated track in inner detector and muon system, combined momentum reconstruction (despite MIP, loses $\sim 3\text{--}5$ GeV in calorimeter through dE/dx), calibrated with $Z, \Upsilon, J/\psi \rightarrow \mu\mu$, requires accurate detector alignment

Detailed muon momentum calibration (and much more!) enables precise W mass measurement (10 MeV (0.012%) precision, also find agreement with LEP Z mass within 4.8 MeV precision)



Electroweak fit
PRD 110 (2024) 030001
LEP combination
Phys. Rep. 532 (2013) 119
D0
PRL 108 (2012) 151804
CDF
Science 376 (2022) 6589
LHCb
JHEP 01 (2022) 036
ATLAS
arXiv:2403.15085
CMS
This work

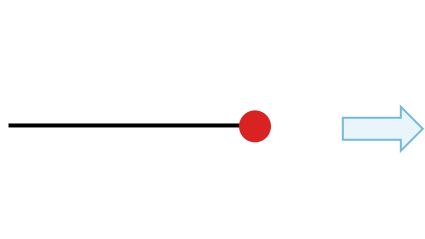


Precision objects enable precision measurements — Jet

A Jet is an experimentally defined, collimated spray of hadrons initiated by an energetic quark or gluon
QCD radiation builds the shower; colour confinement turns it into hadrons; a jet algorithm makes it measurable

1. Hard process

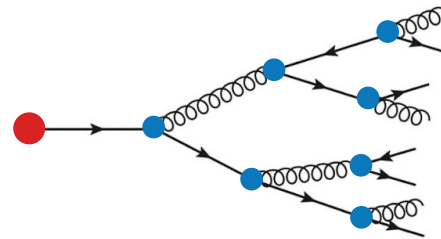
q/g



A coloured parton is produced at a hard scale Q

2. Parton shower

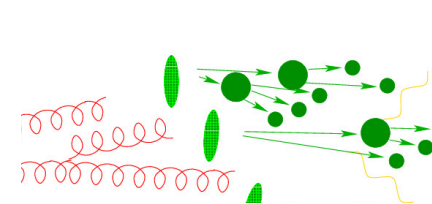
$q \rightarrow qg, g \rightarrow gg, g \rightarrow q\bar{q}$



Repeated emissions evolve from Q down to $Q_0 \approx 1 \text{ GeV}$

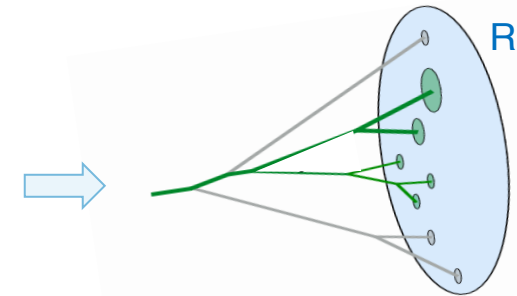
3. Hadronisation

$Q_0 \sim \Lambda_{\text{QCD}}$



Confinement converts partons into colour-singlet hadrons

4. Jet clustering



Anti- k_t clusters particles in $\eta - \phi$:
 $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ (typically $R = 0.4$)

Soft-collinear enhancement of gluon emission

$$dP \approx C_i \frac{\alpha_s}{\pi} \cdot \frac{dE}{E} \cdot \frac{d\theta^2}{\theta^2}$$

soft gluon: $E \rightarrow 0$
collinear gluon: $\theta \rightarrow 0$

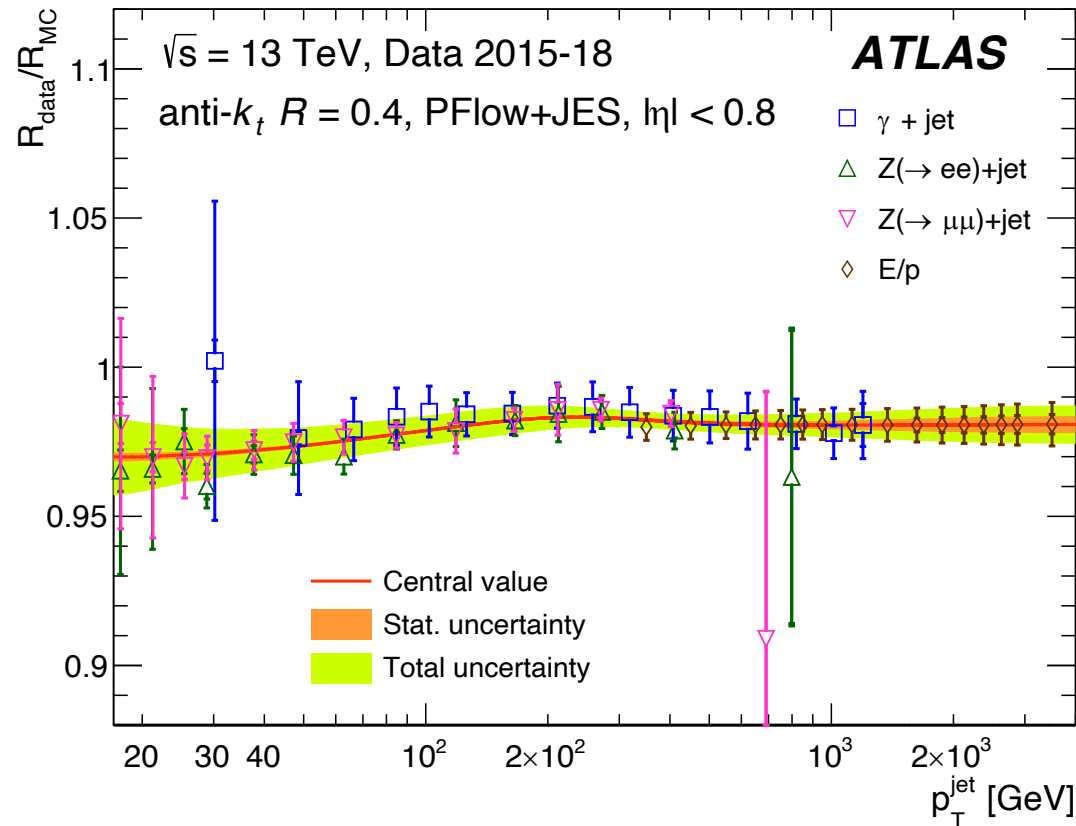
Why the measured cross section remains finite

For **infrared- and collinear-safe jet algorithms**, unresolved real emissions do not change the jet; their divergences cancel against virtual corrections. Repeated emissions generate Sudakov suppression

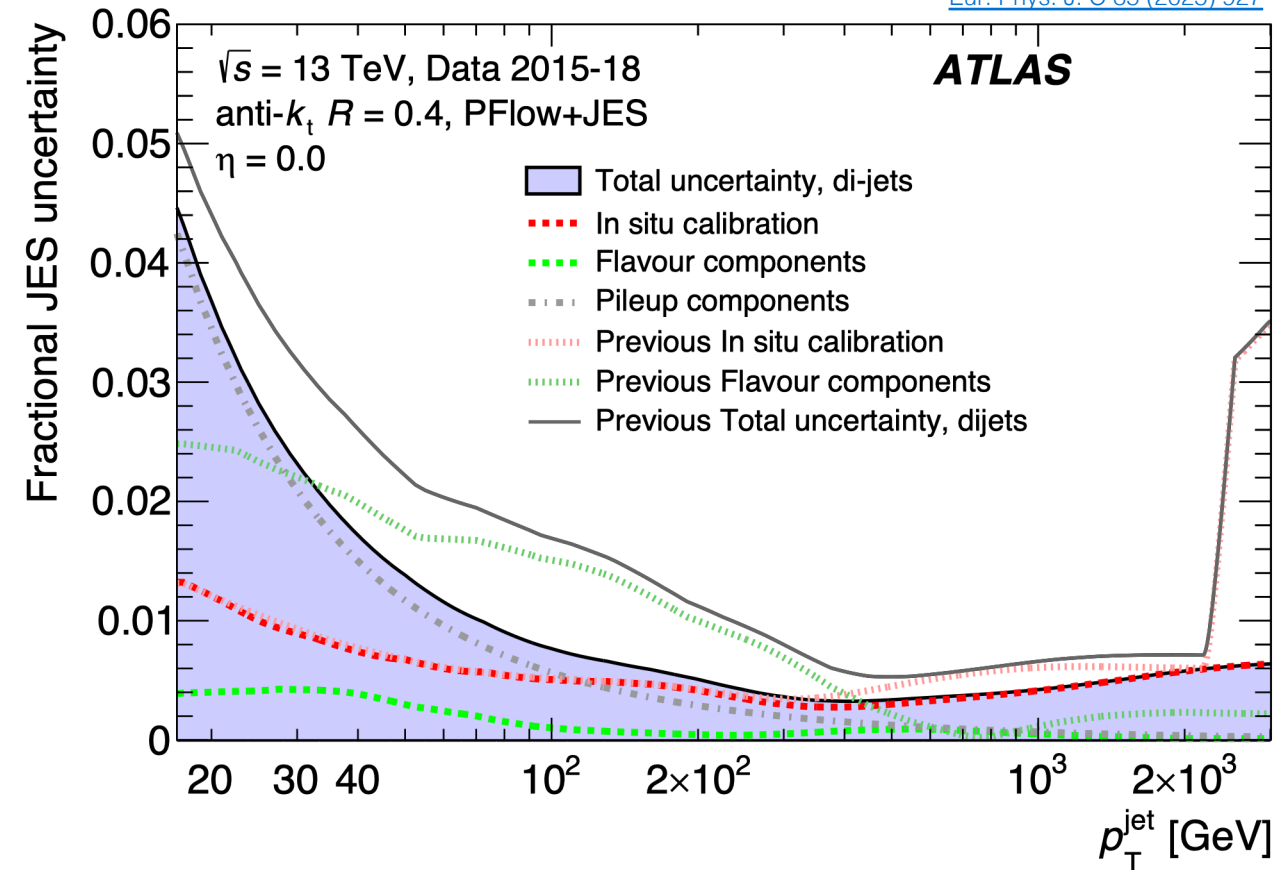
Jets are not fundamental particles: they are IRC-safe, algorithm-dependent representations of a hadronic final state

Precision objects enable precision measurements — Jet

Jet energy scale (JES) is calibrated to “particle level” using di-/multi-jet, γ + jet, Z + jet, $\tau \rightarrow \pi\nu$, $t \rightarrow W \rightarrow qq$ events



New JES calibration method using calorimeter response to single particles (test-beam, E/p from $\tau \rightarrow \pi\nu$), combined with p_T balance techniques for the most precise JES to date

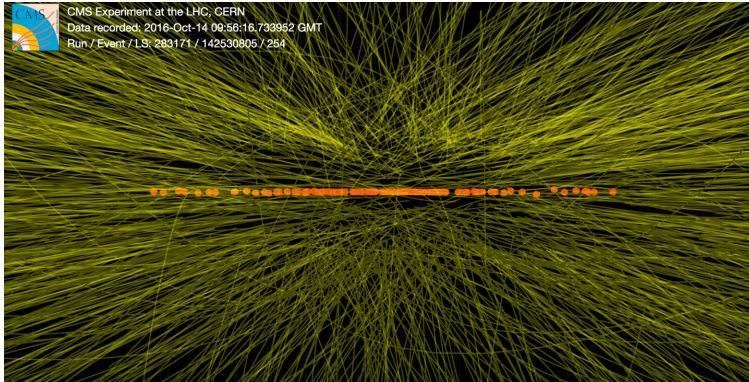


→ Total JES uncertainty of 3.8 / 0.3 / 0.6% at 20 / 300 / 4000 GeV
(uncertainty at low p_T dominated by pileup)

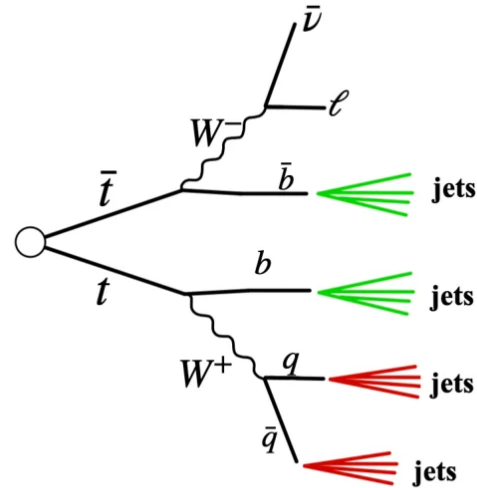
Precision objects enable precision measurements — Jet

Jet energy scale (JES) is calibrated to “particle level” using di-/multi-jet, γ + jet, Z + jet, $\tau \rightarrow \pi\nu$, $t \rightarrow W \rightarrow qq$ events

Precise jet energy calibration using in-situ processes enables, among others, high-precision top mass measurement (0.2%)

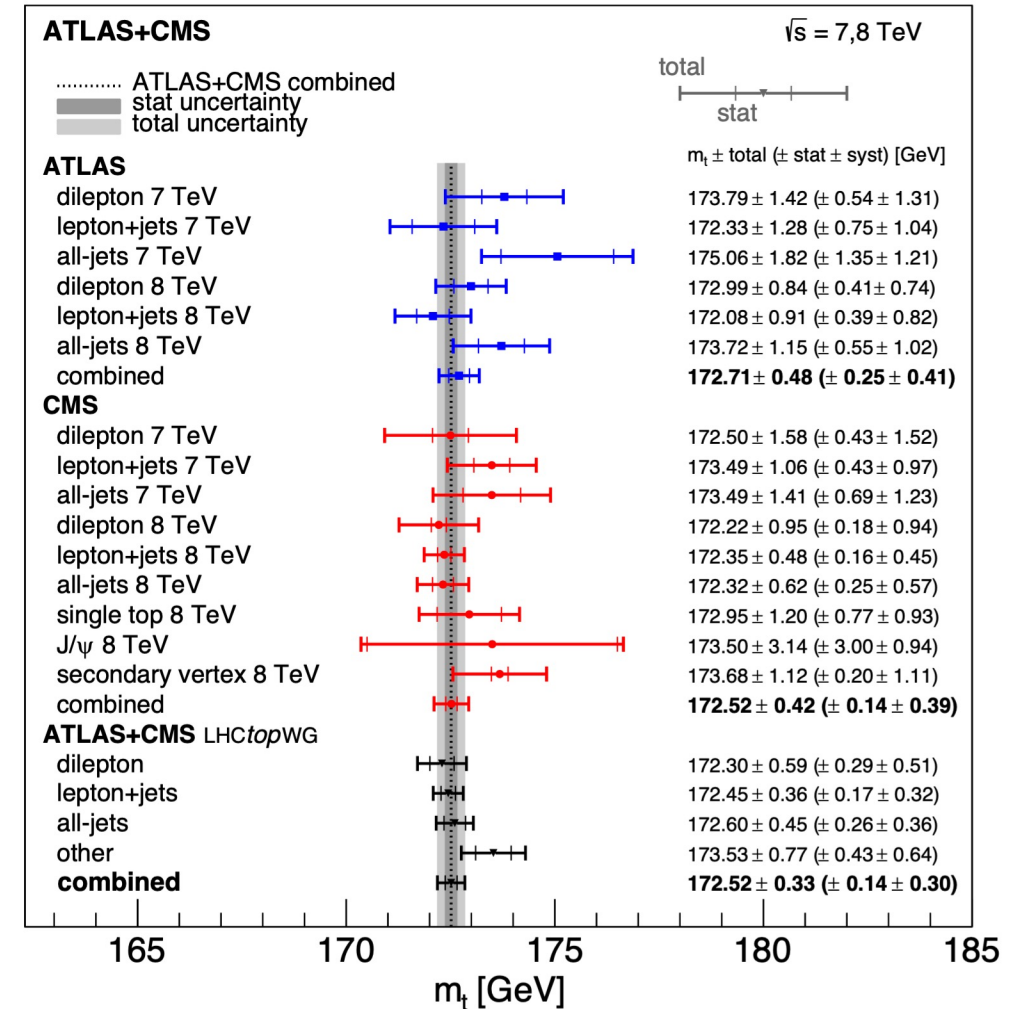


~100 pileup pp interactions in CMS detector



Pileup: mitigated / corrected through vertex association of tracks, area-based subtractions using soft event activity, particle flow candidate weight based on local activity and vertex (PUPPI, CMS)

[arXiv:2402.08713](https://arxiv.org/abs/2402.08713)



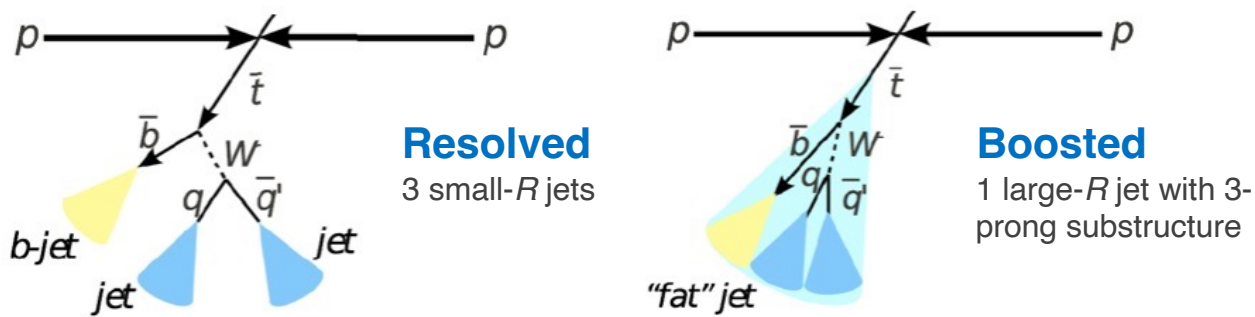
Lorentz boost and jet substructure

Opening angle $\theta \sim 2\gamma^{-1} = 2m/p_T$, $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \approx \theta$

At $p_T \gg m$, the hadronic decay products of W, Z, H and top become collimated and may merge into one reconstructed jet

$$\Delta R \approx 2m/p_T$$

Two-body, symmetric decay; $p_T \gg m$



When do standard $R = 0.4$ jets merge? $p_T \gtrsim 2m/R$

W/Z	$\approx 400\text{--}450\text{ GeV}$	2 prongs
H	$\approx 630\text{ GeV}$	2 prongs + flavour
top	$\approx 870\text{ GeV}$	3 prongs + b

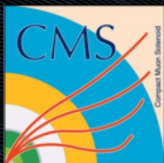
Rule of thumb: exact containment depends on decay kinematics, radiation and the jet algorithm

From a ‘fat’ jet to a physics object

- Capture** Cluster all decay products with a large radius, typically $R = 1.0$
- Groom** Remove pileup and soft wide-angle radiation, stabilise the jet mass
- Resolve prongs** Use jet mass, 2-/3-prong structure and b-tagged subjects

Modern low-level taggers learn these patterns directly from tracks and calorimeter/particle-flow constituents

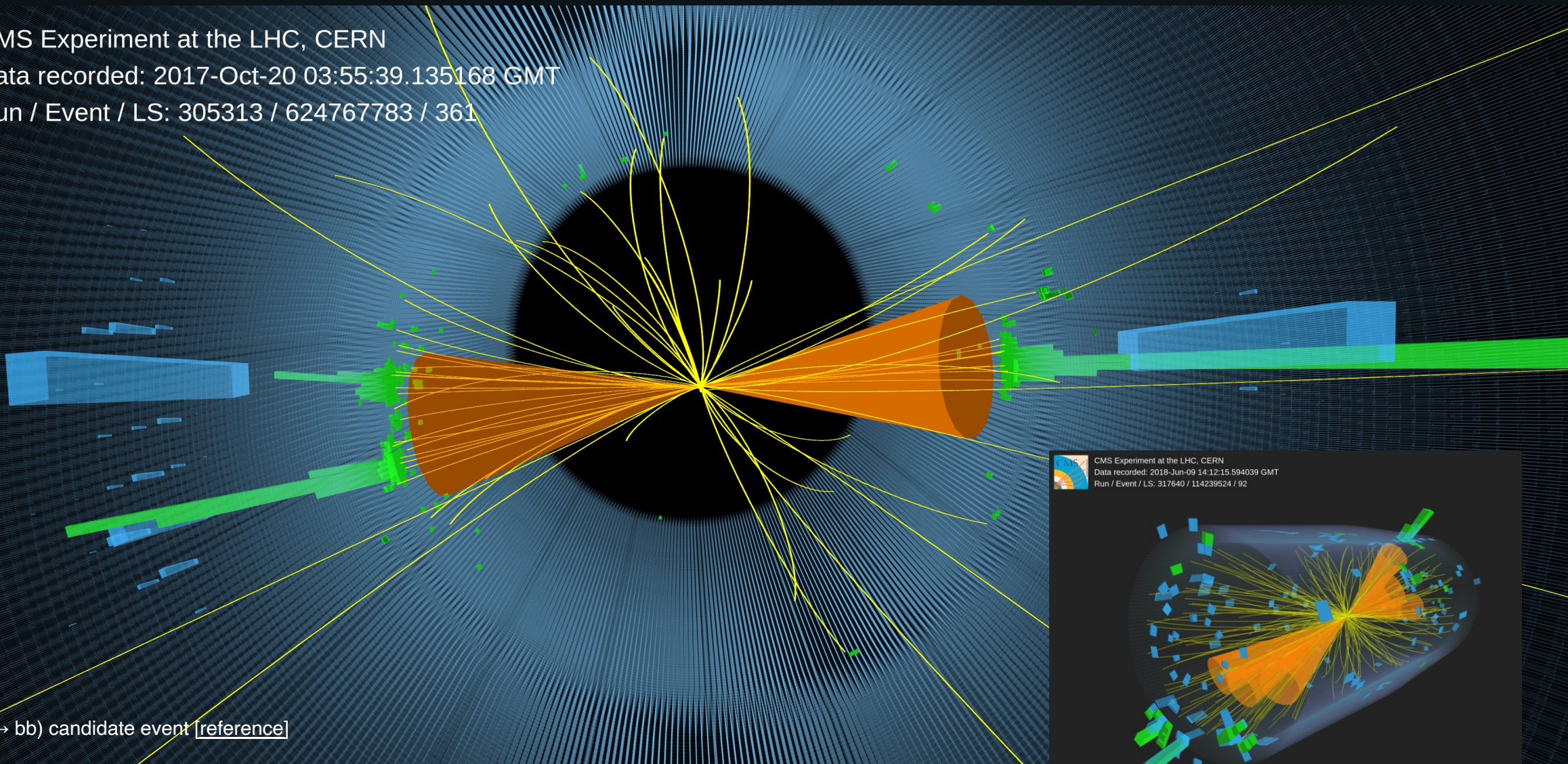
Boosted-object reconstruction converts lost angular separation into measurable internal jet structure



CMS Experiment at the LHC, CERN

Data recorded: 2017-Oct-20 03:55:39.135168 GMT

Run / Event / LS: 305313 / 624767783 / 361



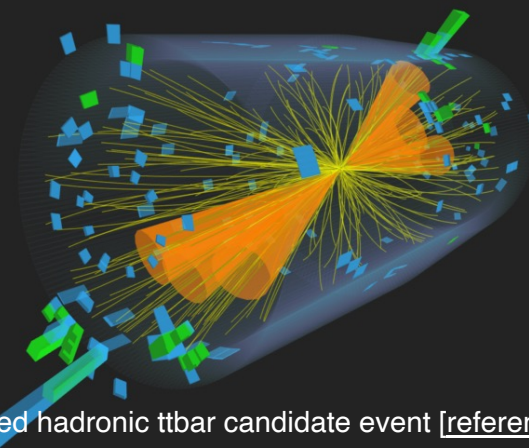
Jet + Higgs ($\rightarrow$ bb) candidate event [\[reference\]](#)



CMS Experiment at the LHC, CERN

Data recorded: 2018-Jun-09 14:12:15.594039 GMT

Run / Event / LS: 317640 / 114239524 / 92

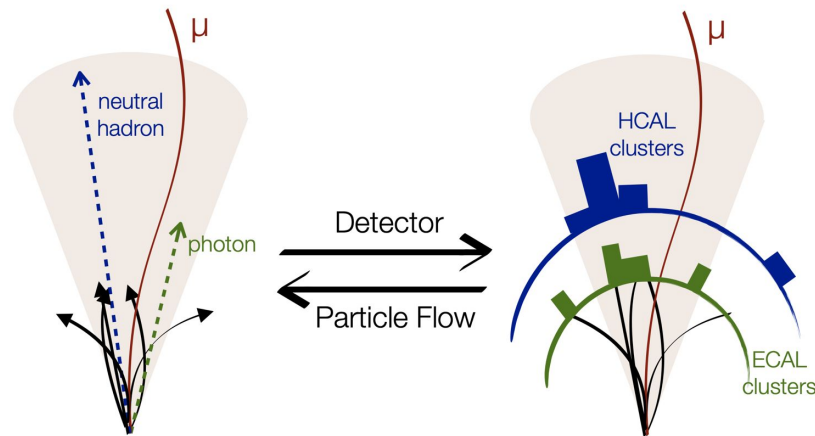


Boosted hadronic ttbar candidate event [\[reference\]](#)

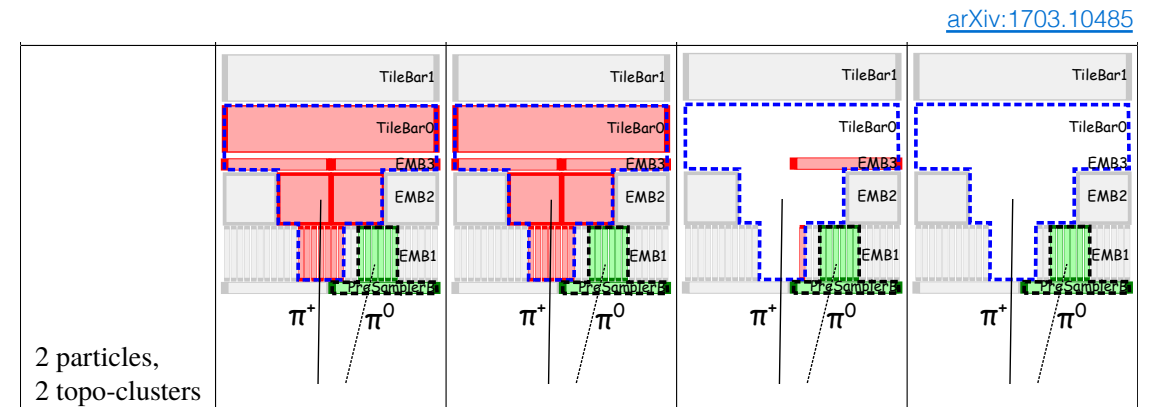
Particle flow: measure each particle with the best detector

Particle flow: a jet typically deposits ~65% of its visible energy via charged hadrons, 25% photons (mainly π^0), 10% neutral hadrons (n, K_L , ...)

- Measured about 1/3 in ECAL and 2/3 in the HCAL (poorer resolution, limiting jet energy resolution)
- Idea: replace energy of individual charged hadrons by track momenta (energy flow), so that ideally only 10% of energy stems from HCAL, leading to improved jet energy resolution for $E_{\text{jet}} \lesssim 150 \text{ GeV}$
- Generalisation to “particle flow” with aim to reconstruct individual particles not just energy deposits (ALEPH, ILC, CMS, ATLAS, FCC, ...)
- Many ideas for end-to-end particle flow involving machine learning

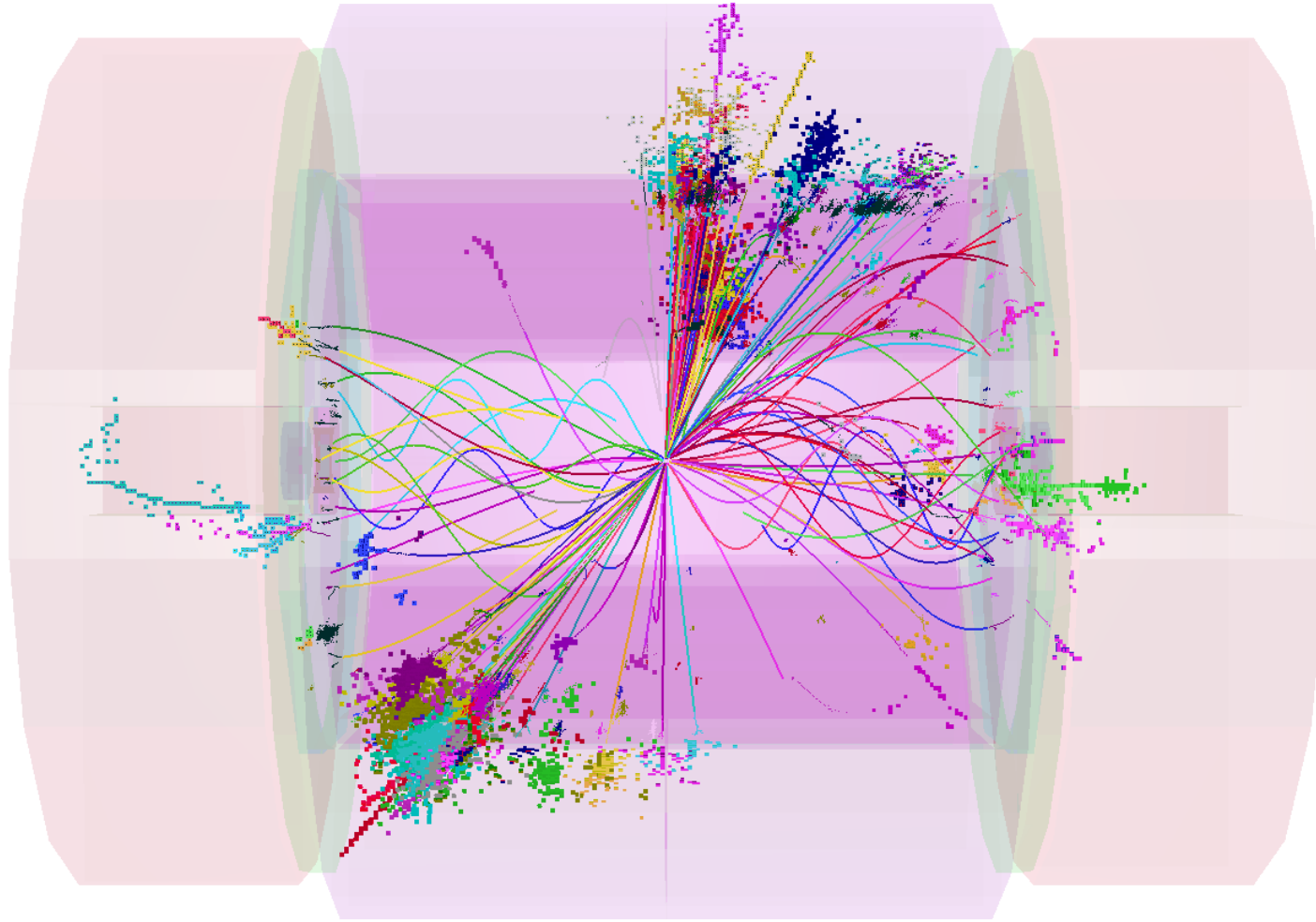


Particle flow solves the inverse problem of detector simulation → particles
[\[link to figure\]](#)



Idealised examples of particle flow algorithm handling different cases

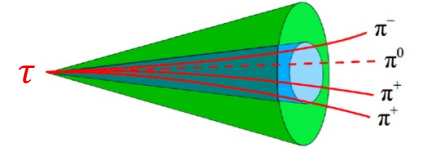
Particle flow: measure each particle with the best detector



Next-generation energy-frontier detectors are designed around particle flow
(here: CLIC detector design (CLICdet), silicon-tungsten ECAL & steel-scintillator HCAL)

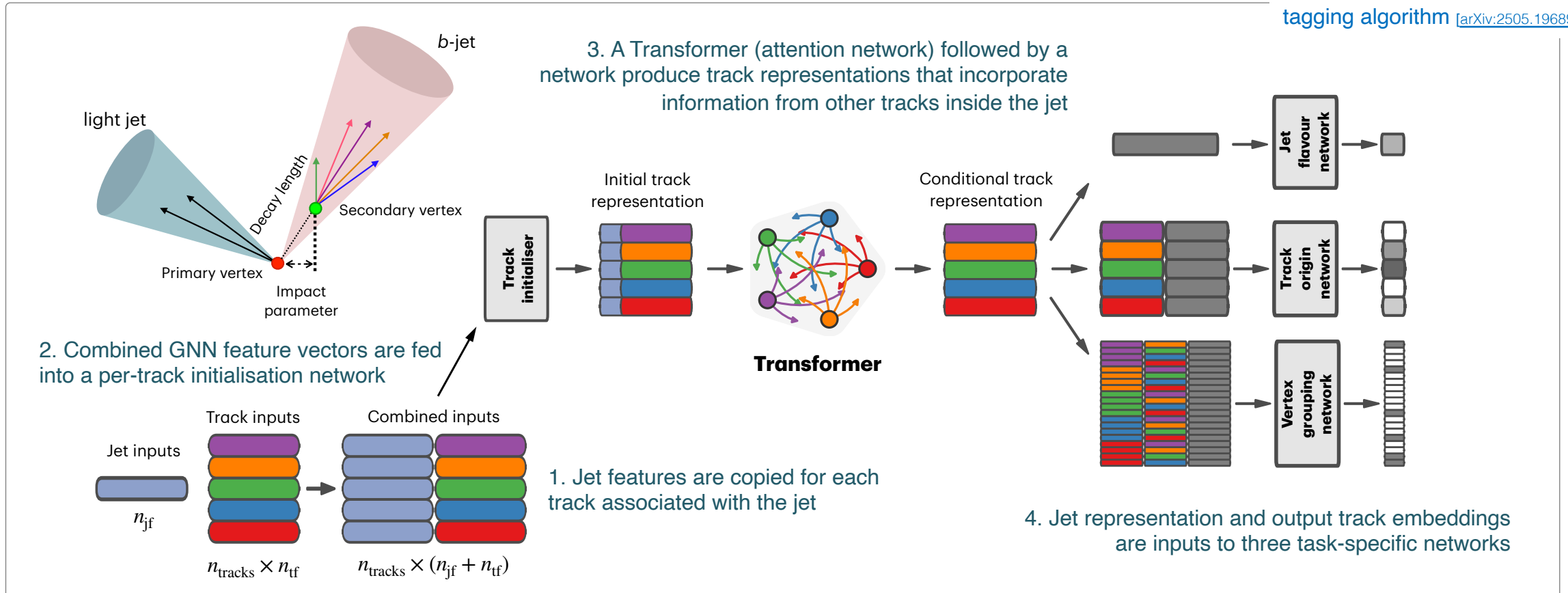
Flavour reconstruction / tagging — eldorado for modern ML

Tau: $\rightarrow e/\mu + \nu\nu$ (35%), $\rightarrow \pi/K + \nu$ (11.5%), $\pi/K+\pi^0 + \nu$ (26%), ..., isolated hadrons reconstructed in tracker and calorimeter (narrow cones) using particle flow and ML, calibrated with $Z \rightarrow \tau\tau$



b/c quarks: “tagged” through decay into b/c hadron, jet with massive secondary (b $\rightarrow$ c tertiary) vertex, non-isolated soft lepton(s), combined through ML, calibrated with jets containing muons, ttbar events

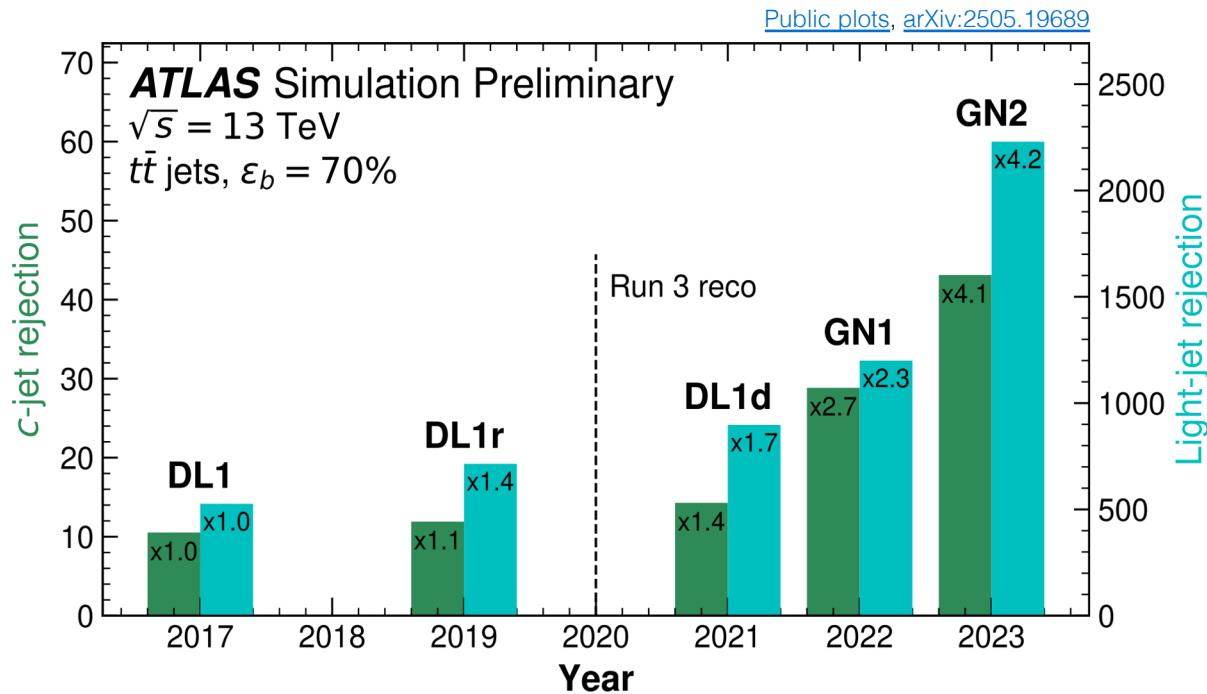
Illustration of ATLAS GN2 flavour tagging algorithm [\[arXiv:2505.19689\]](https://arxiv.org/abs/2505.19689)



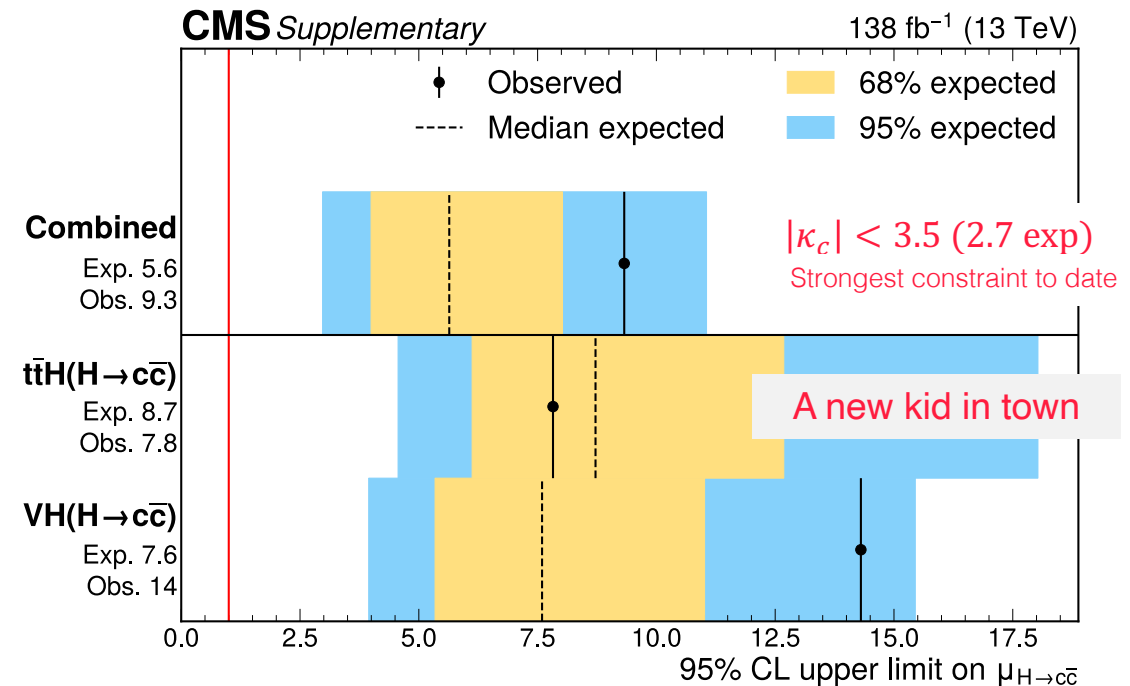
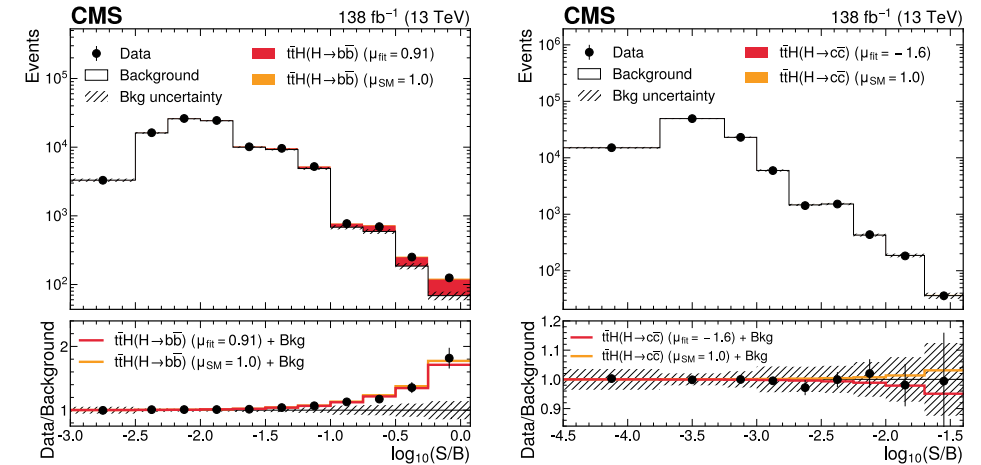
Jet/track features $\rightarrow$ Per-track embeddings $\rightarrow$ Transformer builds context across tracks $\rightarrow$ 3 heads deliver (i) jet flavour, (ii) track origins, and (iii) explicit displaced-vertex structure

Flavour reconstruction / tagging — eldorado for modern ML

Impressive flavour tagging and further ML based performance and analysis improvements enhance physics reach



CMS $t\bar{t}H(cc/bb)$ analysis used [ParticleNet](https://arxiv.org/abs/2509.22535) for jet flavour identification and [ParticleTransformer](https://arxiv.org/abs/2509.22535) for event classification

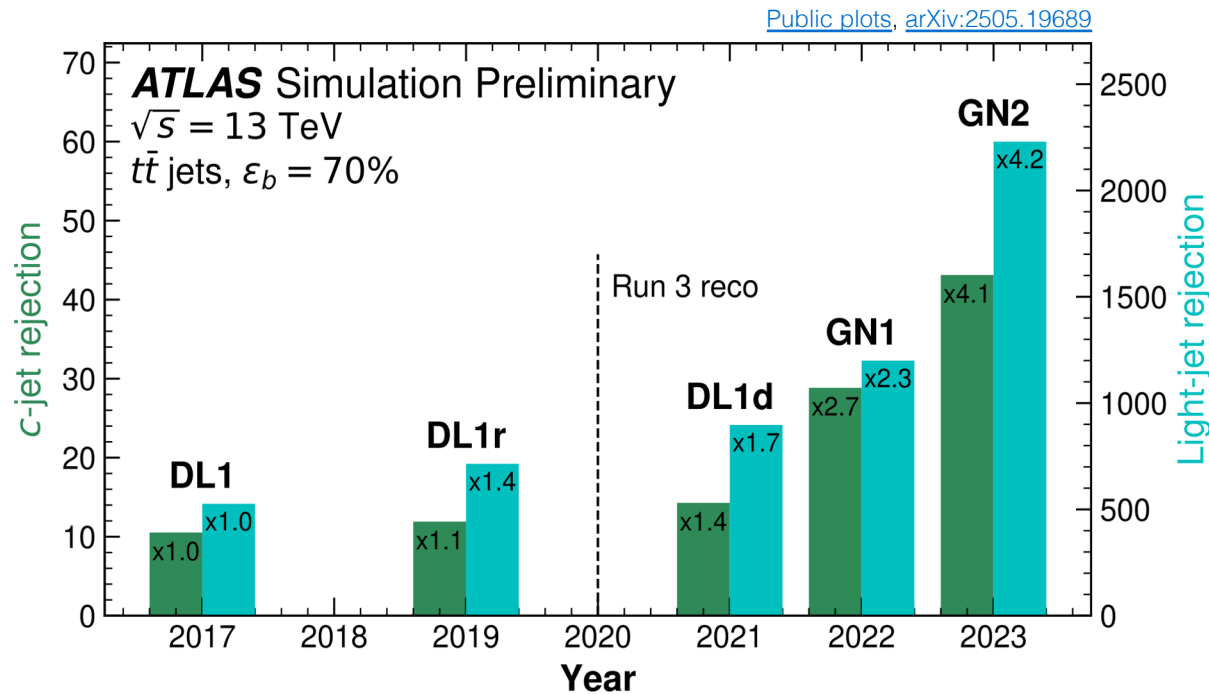


[arXiv:2509.22535](https://arxiv.org/abs/2509.22535)

Flavour reconstruction / tagging — eldorado for modern ML

Impressive flavour tagging and further ML based performance and analysis improvements enhance physics reach

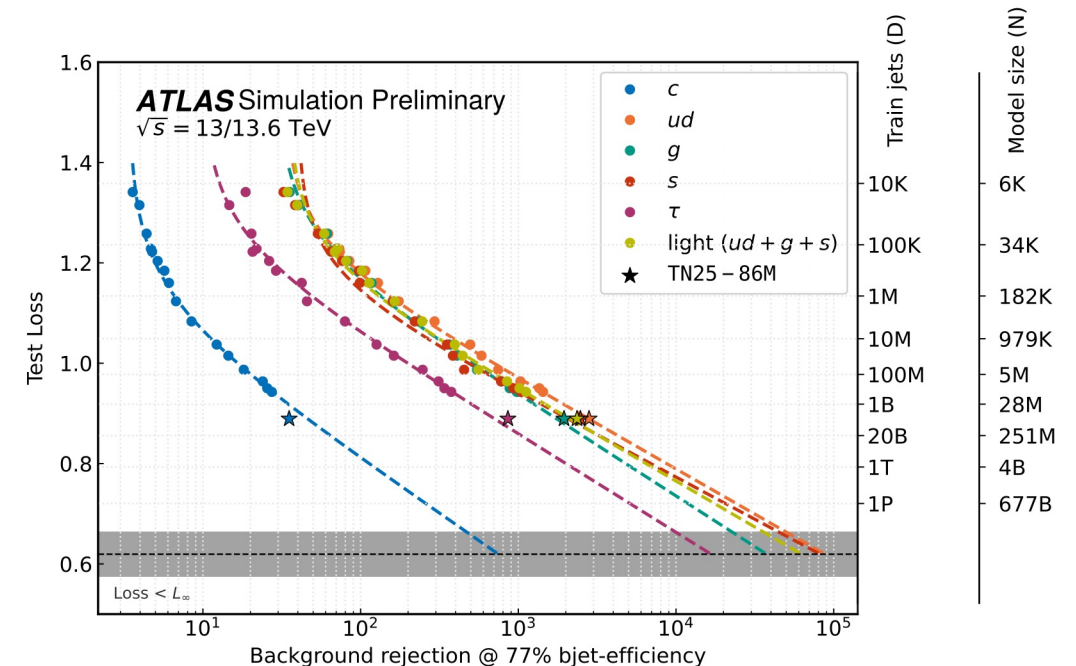
[ATL-SOFT-PUB-2026-002](#)



Large language models (LLMs) have demonstrated that scaling, driven by compute, can yield dramatic improvements in performance, with current state-of-the-art systems reaching up to one trillion parameters. In contrast, state-of-the-art jet flavour classifiers in HEP remain in the tens of millions of parameters. Scaling laws offer a framework for predicting how much HEP models benefit from systematic increases in capacity and training data.

Scaling behaviour of transformers — **how far can it go?**

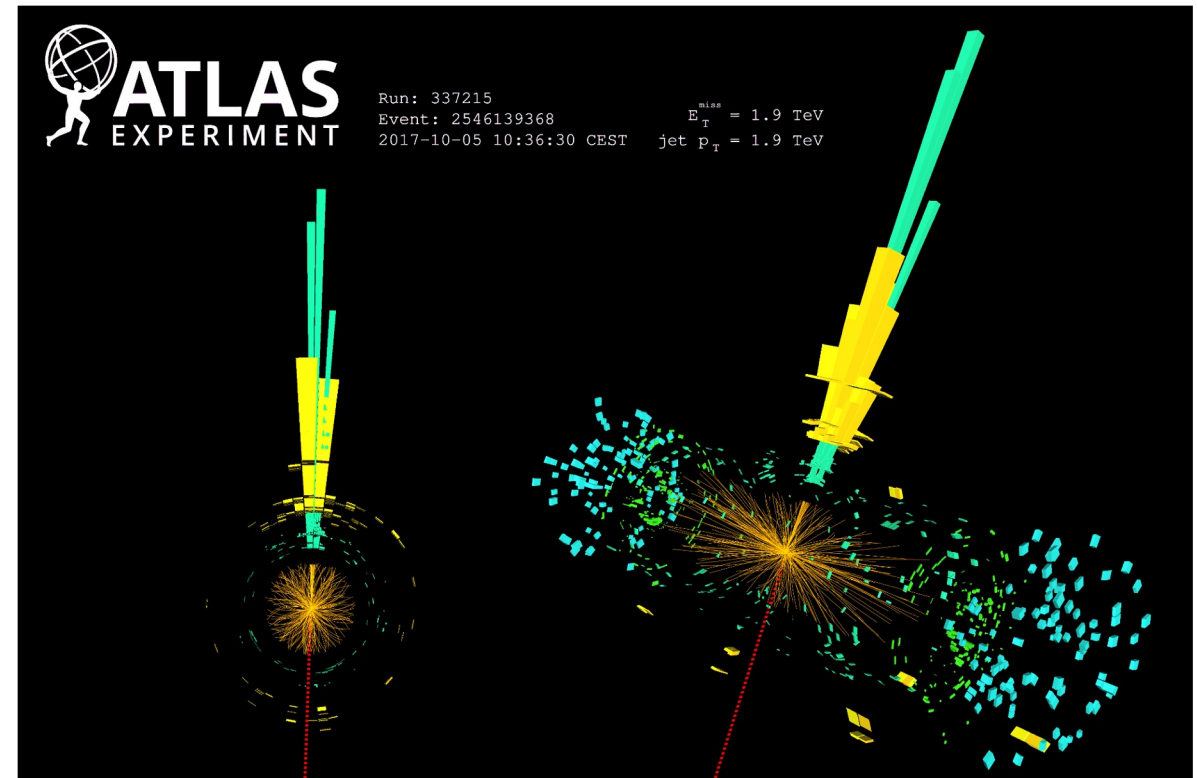
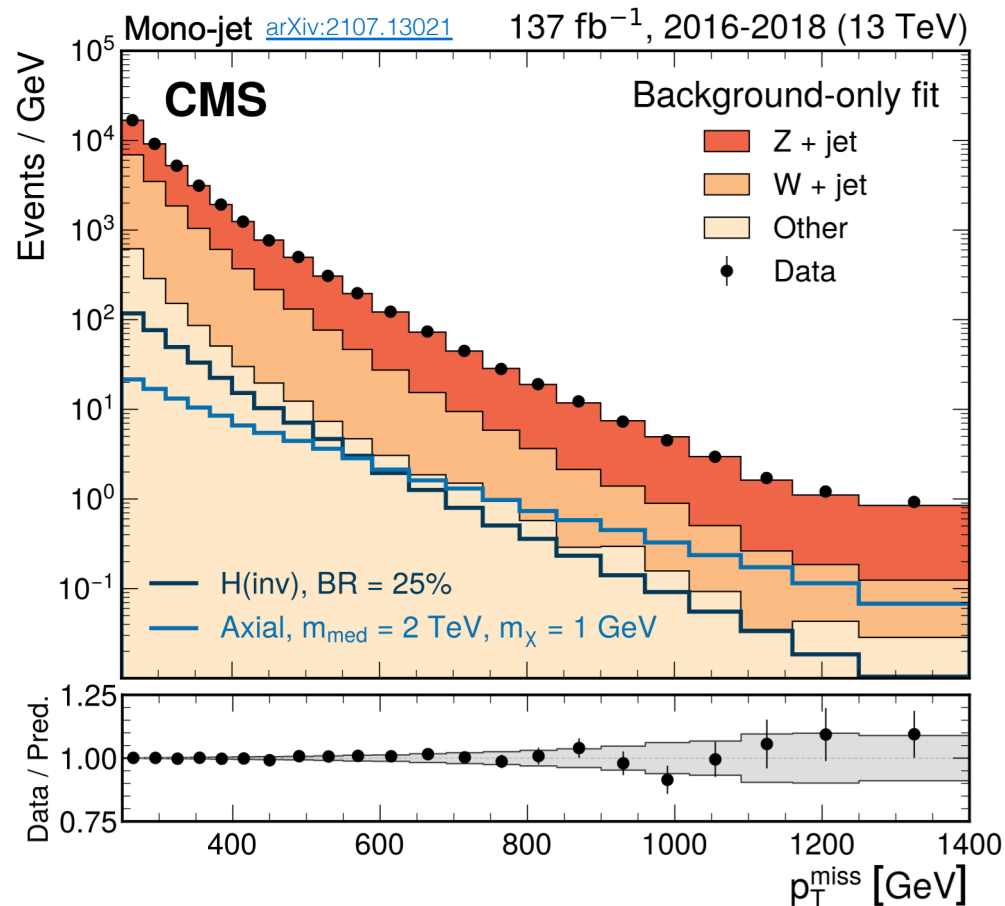
Expect further gains as training data and model sizes increase



Precision objects enable precision measurements — Missing p_T

Missing p_T : negative vector sum of momenta of all reconstructed objects and a “soft term” from isolated tracks / clusters

Precise missing p_T reconstruction allows to detect invisible particles

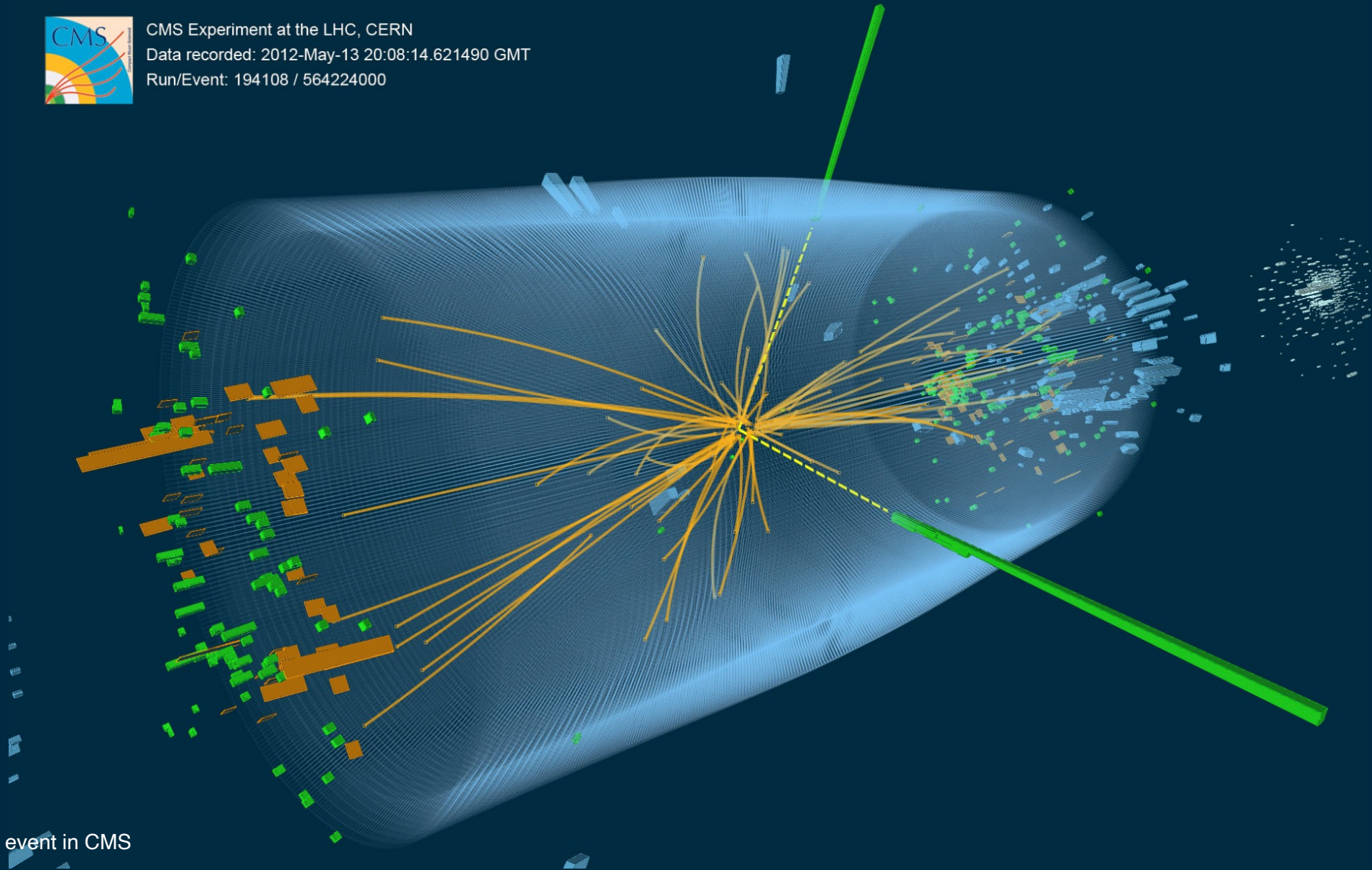


Monojet candidate event

Cross section, backgrounds, likelihood fits, systematic effects



CMS Experiment at the LHC, CERN
Data recorded: 2012-May-13 20:08:14.621490 GMT
Run/Event: 194108 / 564224000



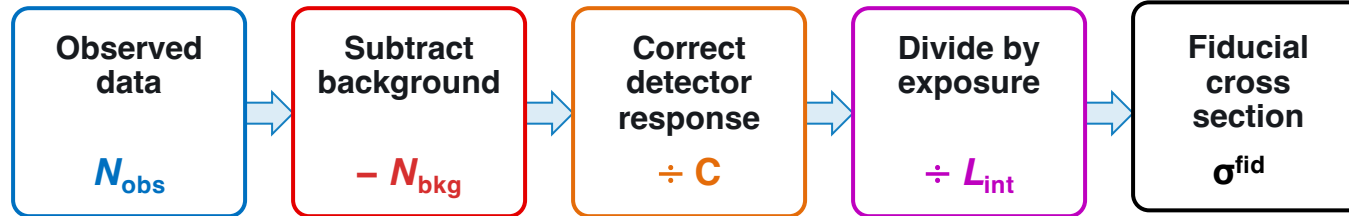
From event counts to a physical cross section

Cross section is obtained by subtracting backgrounds, correcting detector effects, and normalising to integrated luminosity

$$\sigma_{pp \rightarrow X \rightarrow \text{decay}}^{\text{fid}} = \frac{N_{\text{obs}} - N_{\text{bkg}}}{L_{\text{int}} \cdot C_X}$$

$$\sigma_{pp \rightarrow X}^{\text{tot}} = \frac{\sigma_{pp \rightarrow X \rightarrow \text{decay}}^{\text{fid}}}{A_X \cdot B_{X \rightarrow \text{decay}}}$$

Measurement recipe



$C_X = N_{\text{reco,selected}} / N_{\text{gen,fid}}$ — efficiency & resolution correction (from MC)

$A_X = N_{\text{gen,fid}} / N_{\text{gen,tot}}$ — acceptance factor (from MC / theory)

$B_{X \rightarrow \text{decay}}$ — branching fraction (from theory)

Uncertainty budget

- Statistics: finite event counts
- Background: normalisation and shape
- Detector: calibration, efficiency, migrations
- Luminosity: exposure scale
- Signal model: acceptance and unfolding dependence

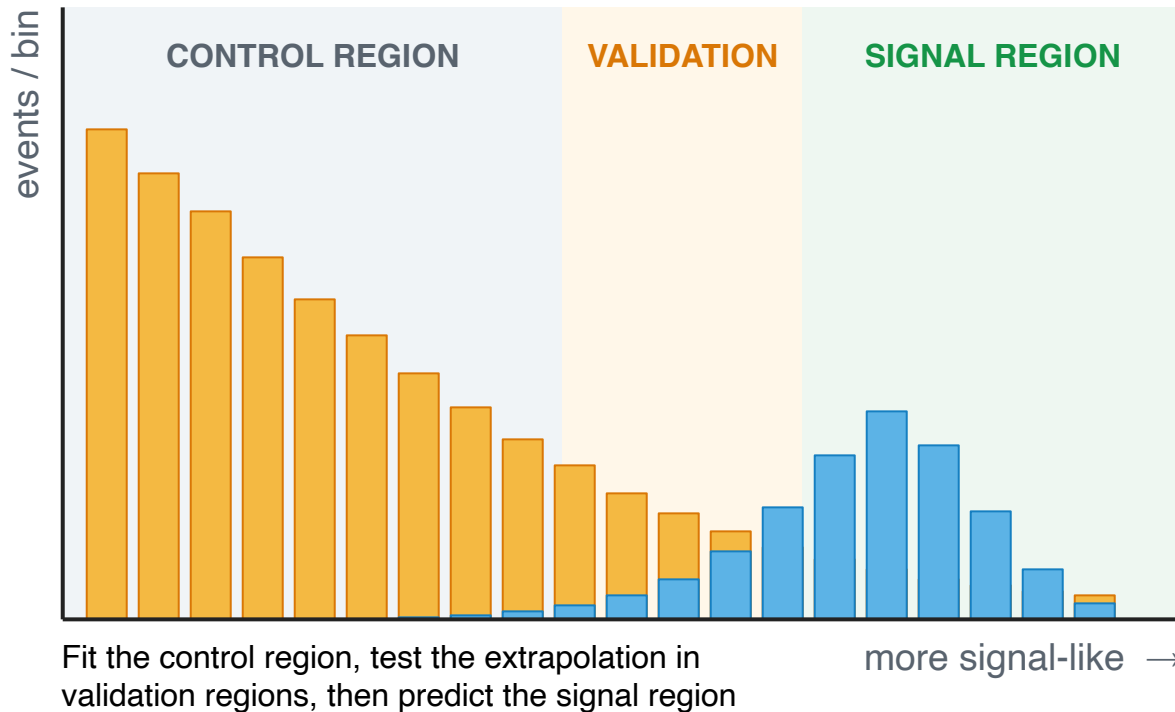
Differential results: invert a response matrix; regularisation trades variance against bias

Fiducial cross sections minimise extrapolation and are usually the most reusable result

Backgrounds are constrained with data whenever possible

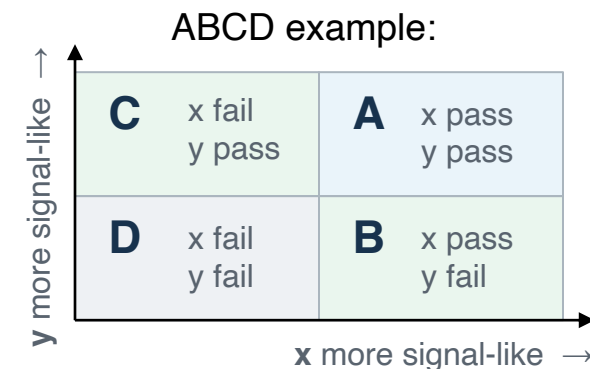
All LHC analyses must estimate backgrounds and their uncertainties: sensitivity of a BSM search often governed by the uncertainty on the background prediction — not only by its expected yield

Example of sensitive regions in high-tail BSM search:



3 complementary strategies:

- **Simulation / theory:** predicts normalisation and shape; limited by modelling and MC statistics in extreme tails
- **Control-region fit:** data constrain background normalisations; simulation transfers them to signal region
- **Data-driven extrapolation:** ABCD, sidebands or fake-factor methods reduce dependence on absolute simulation



$$N_A^{\text{pred}} = \kappa \cdot \frac{N_B N_C}{N_D}$$

κ corrects residual correlations; subtract non-target backgrounds and test signal leakage

Background yields and uncertainties are fitted, correlated quantities — not a fixed subtraction

Profile likelihood fit

Estimate the **parameter of interest** μ while the data constrain **backgrounds** and systematic uncertainties through **nuisance parameters** θ

1 Build the likelihood

Simple counting experiment

$$L(n \mid \mu, \theta) = \text{Pois}(n \mid \mu S + \theta B) \times G(\theta; 1, \sigma\theta)$$

μ signal strength
parameter of interest

θ background scale
nuisance parameter

n observed events | S expected signal | B nominal background
 $G(\theta; 1, \sigma\theta)$ is an auxiliary constraint from calibration or control data

2 Profile the nuisance parameters

For each tested μ , choose the θ values that maximise L



$$L_p(\mu) = L(\mu, \hat{\theta}(\mu))$$

$$\lambda(\mu) = L(\mu, \hat{\theta}(\mu)) / L(\hat{\mu}, \hat{\theta}); \quad q(\mu) = -2 \ln \lambda(\mu)$$

$\hat{\theta}(\mu)$: conditional best fit | $(\hat{\mu}, \hat{\theta})$: global best fit | $q(\mu)$: test statistic

In a realistic LHC analysis

Combined model

Product over bins, regions, channels, categories and processes

Control regions

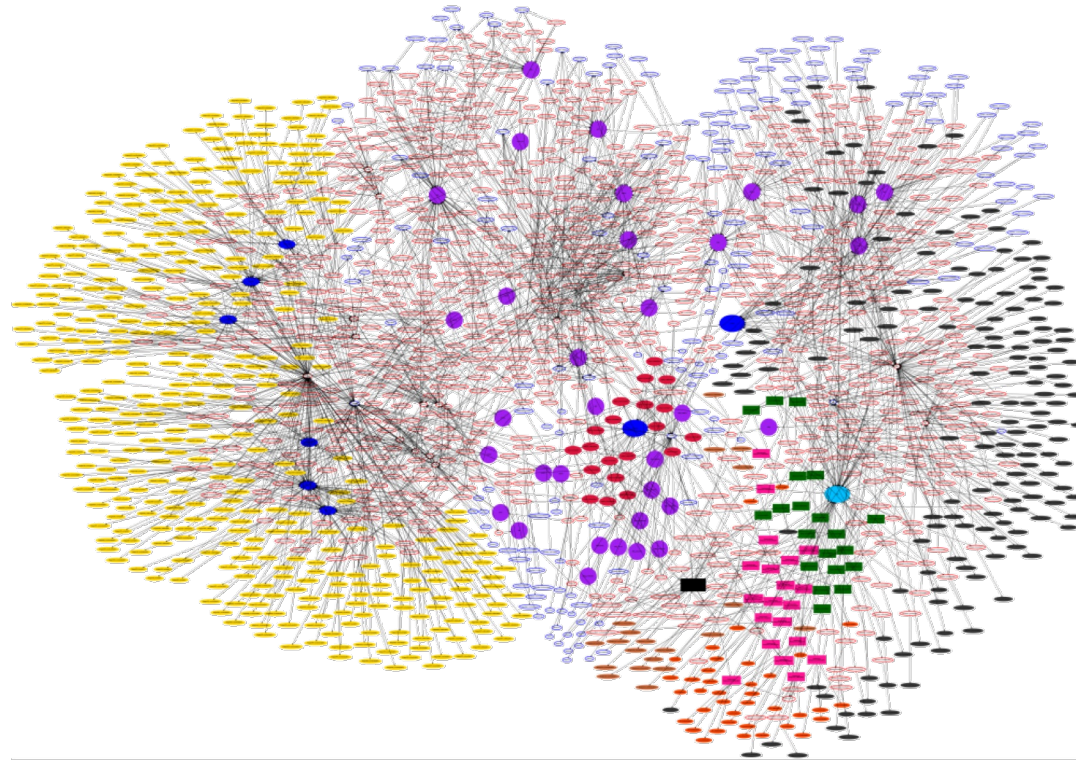
Constrain background yields and nuisance parameters directly from data

Correlations

Shared nuisance parameters propagate systematic effects across the fit

The profile-likelihood ratio provides confidence intervals, significances and upper limits

Profile likelihood fit



ATLAS $H \rightarrow \gamma\gamma$ likelihood
model used in fit

In a realistic LHC analysis

Combined model

Product over bins, regions, channels,
categories and processes

Control regions

Constrain background yields and nuisance
parameters directly from data

Correlations

Shared nuisance parameters propagate
systematic effects across the fit

The profile-likelihood ratio provides confidence intervals, significances and upper limits

Systematic uncertainties

They quantify how **imperfect knowledge** can impact the inferred physics result

Statistical: finite event counts; usually decreases as $1/\sqrt{N}$

Systematic: imperfect knowledge; may persist, but can be constrained by data (introducing a $1/\sqrt{N}$ component)

Common sources

Detector response and calibration

E and p_T scales, efficiencies, resolutions, alignment

Backgrounds and event modelling

Normalisations, shapes, data-driven transfer factors

Theory and simulation

Missing higher orders, PDFs, parton shower, hadronization

Data-taking conditions

Luminosity, pileup, time dependent conditions, detector stability

Treatment in a robust analysis

- 1 Identify and parameterise**
Represent each relevant effect by a nuisance parameter θ_i
- 2 Constrain with data**
Use calibrations, auxiliary measurements and control regions
- 3 Propagate shifts and correlations**
Carry rate and shape variations consistently through the likelihood
- 4 Validate and report**
Perform closure and stress tests; rank impacts and document assumptions

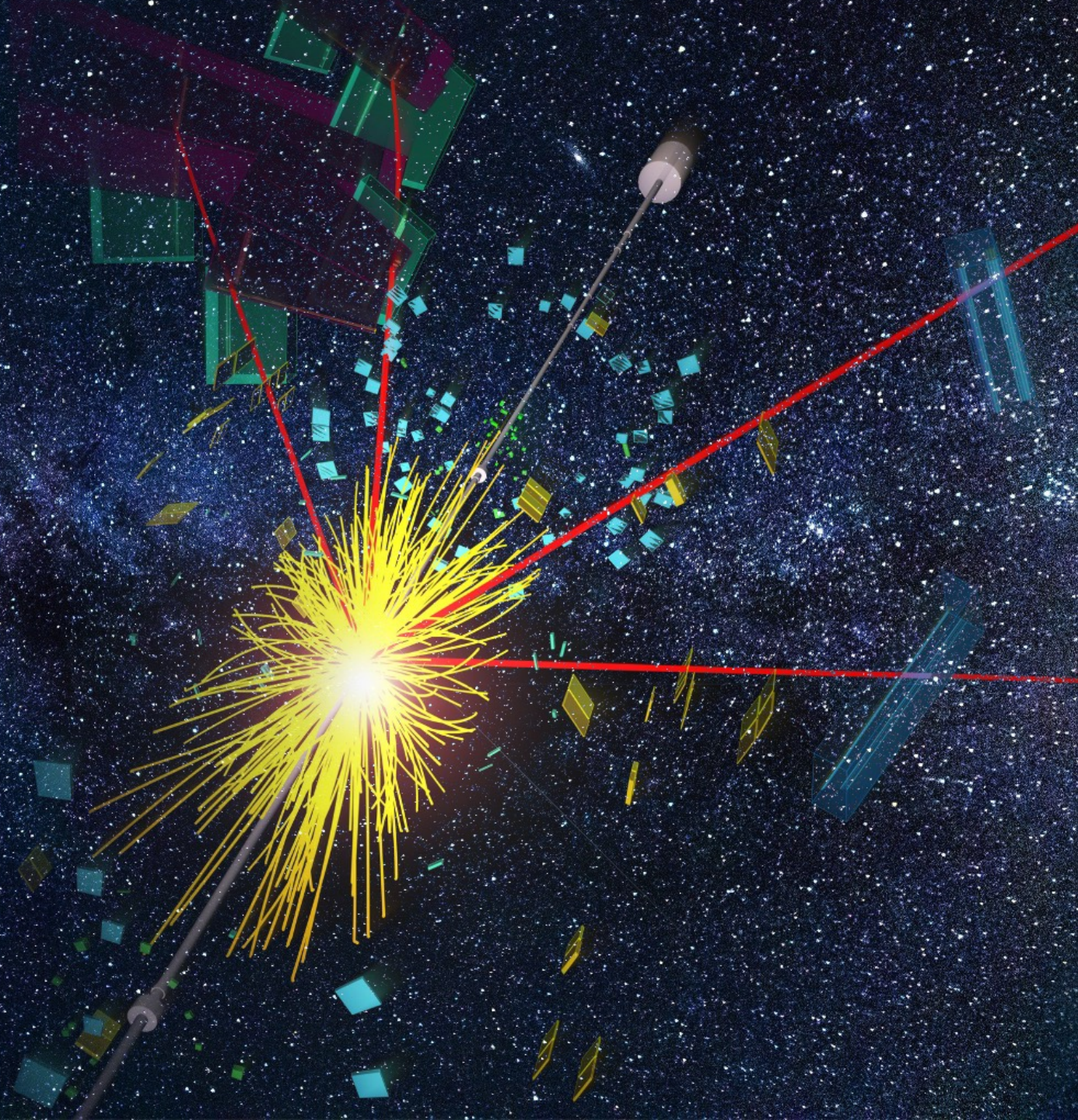
Goal: calibrated coverage, transparent assumptions, correct correlations — *right*, not over or under-conservative

Theory uncertainties: state the variation prescription explicitly; a Gaussian “ 1σ ” interpretation is not always justified

Summary of today's lecture

- A large facility, such as the LHC and its injectors, offers a diverse research programme, flagship experiments and much more
- The experimental environment of a hadron collider is especially demanding for detectors, event reconstruction, and their simulation. It is also a driver for innovation
- The groundwork invested into these areas is key for the successful physics exploitation
- Physics analysis relies on sophisticated statistical methods correlating as much as possible all contributions and their uncertainties
- There seems to be no limit to further improvement!

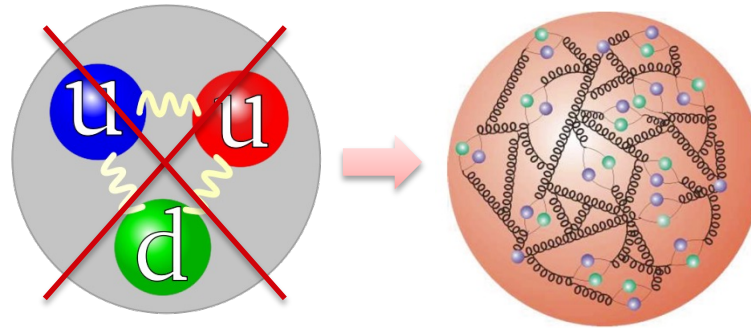
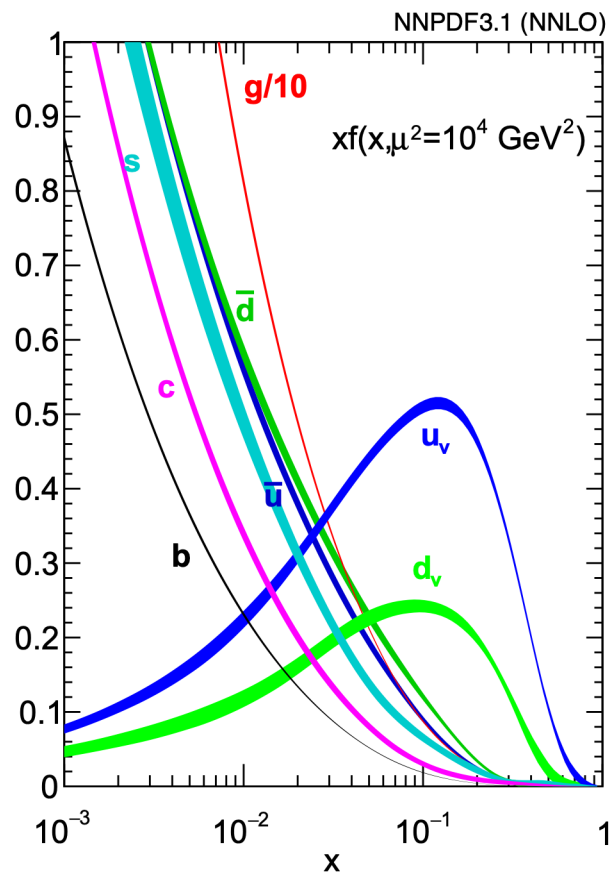
Extra slides



Proton–proton collisions

Proton–proton collisions correspond to a convolution of Parton Density Functions (PDF) with parton–parton scattering matrix element

Parton distribution functions

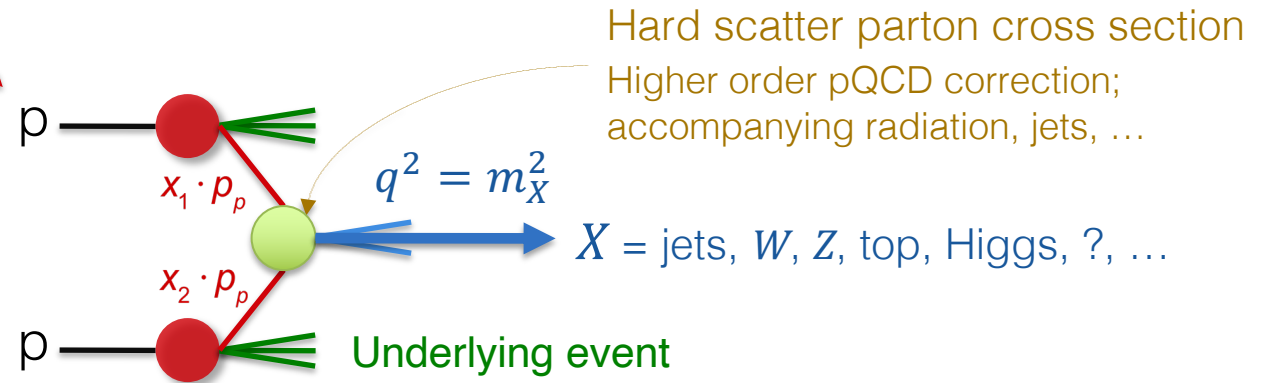


Represent structure of proton, extracted using experimental data and QCD properties

Proton–proton collisions

Proton–proton collisions correspond to a convolution of Parton Density Functions (PDF) with parton–parton scattering matrix element

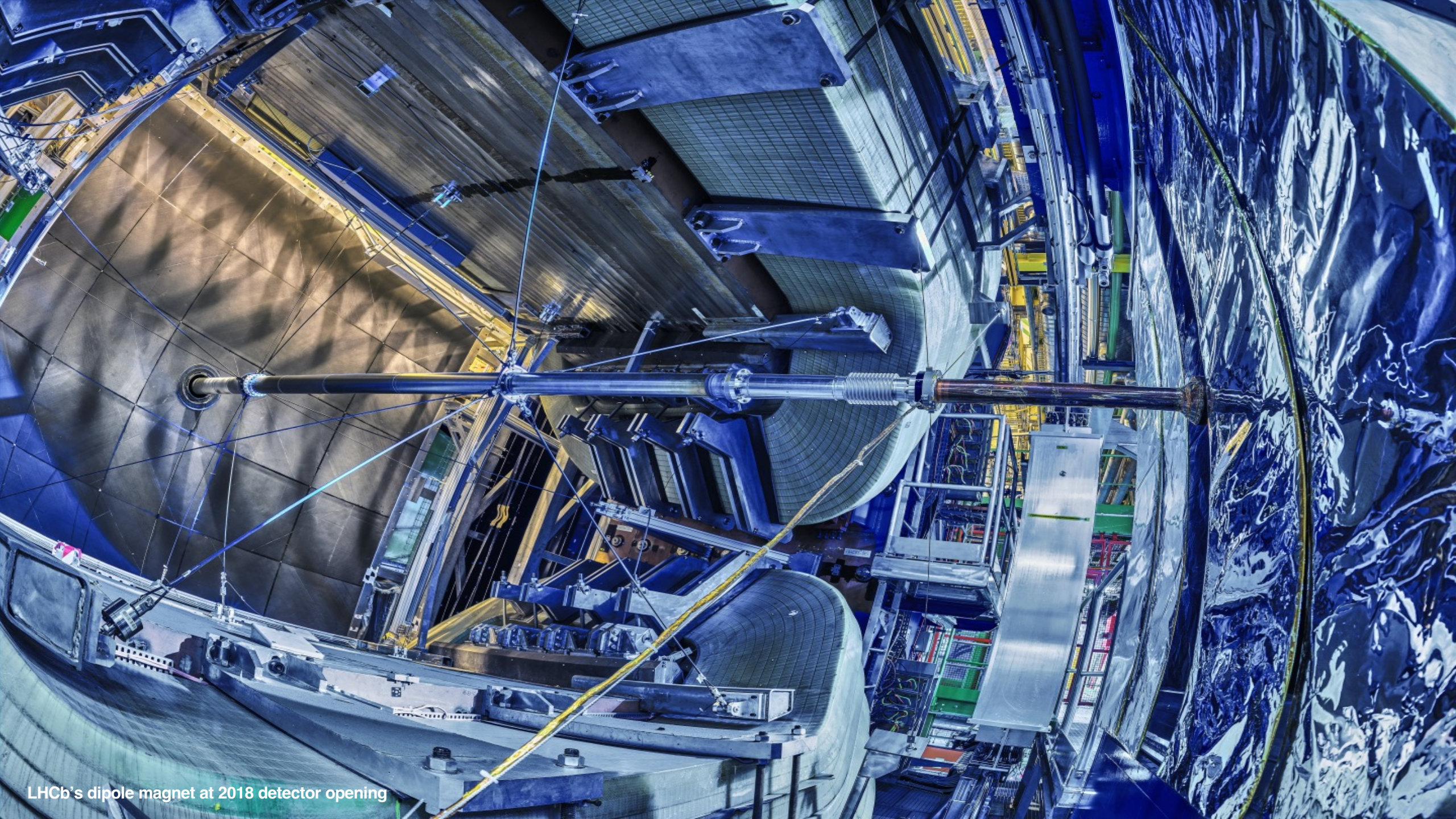
Parton distribution functions



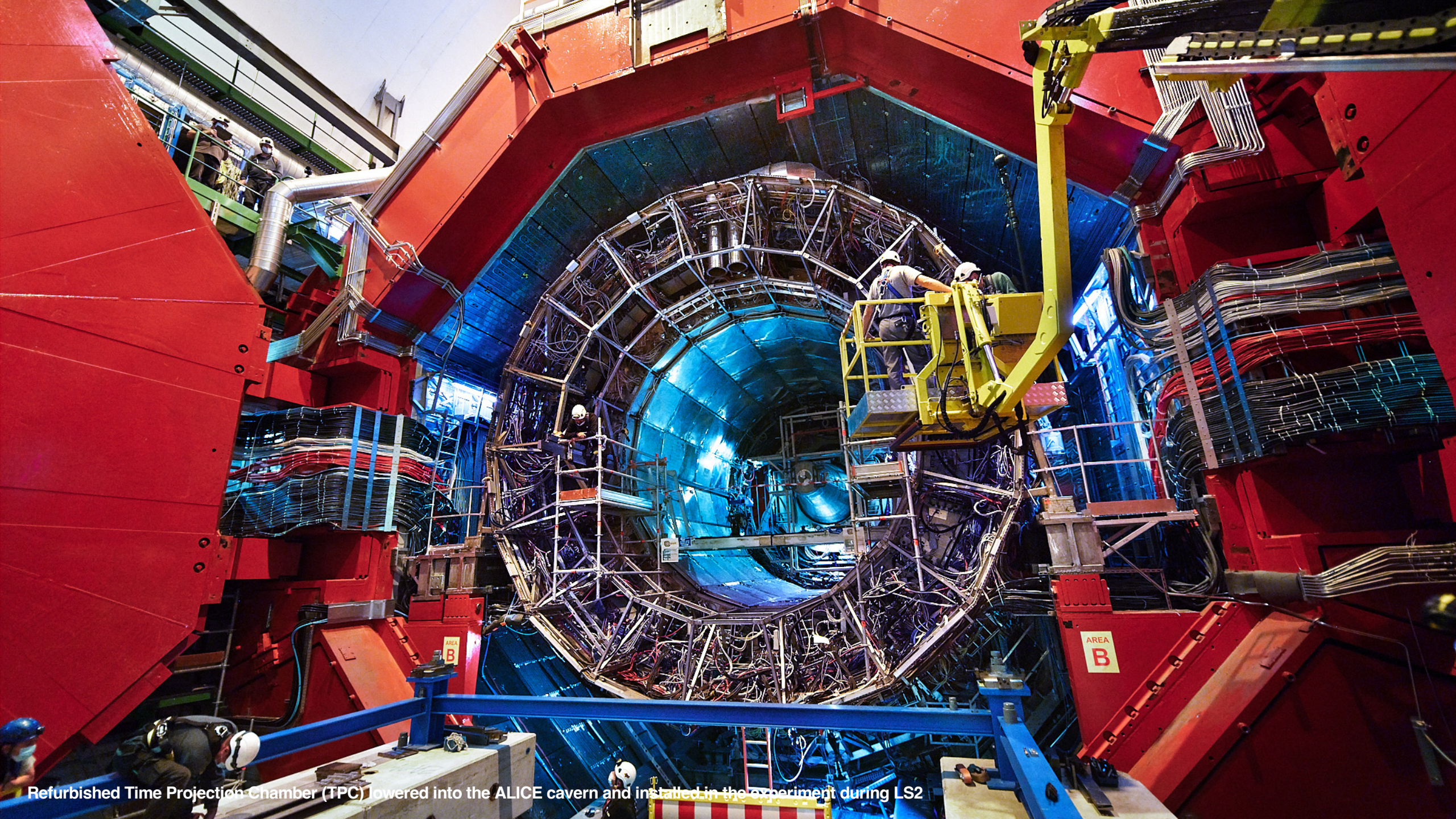
CM energy-squared of parton collision: $\hat{s} = M_X^2 = x_1 \cdot x_2 \cdot s_{\text{LHC}}$

Gluon and sea-quark PDFs rise dramatically towards low x :

- Higher cross section at higher proton–proton collision energy
- More luminosity also leads to higher achievable energy
- Low- x regime (eg, Higgs production) dominated by gluon–gluon collisions: “gluon collider”



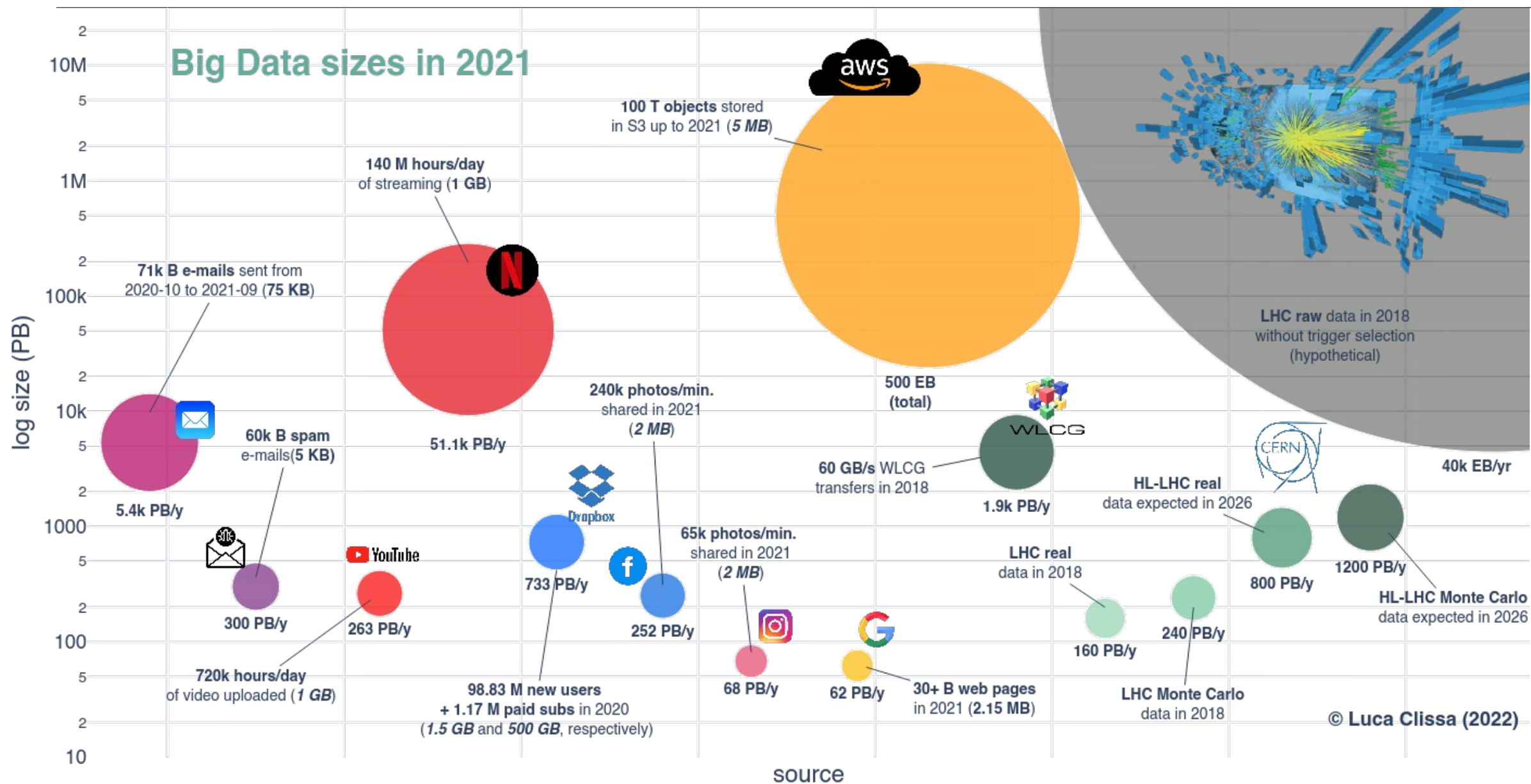
LHCb's dipole magnet at 2018 detector opening



Refurbished Time Projection Chamber (TPC) lowered into the ALICE cavern and installed in the experiment during LS2

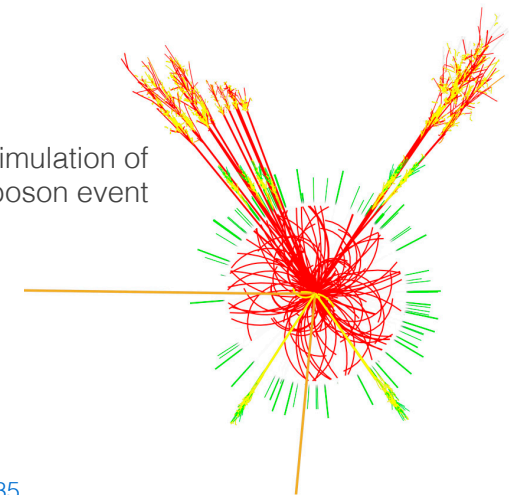
Particle detector basics — Data sizes

[Reference](#)



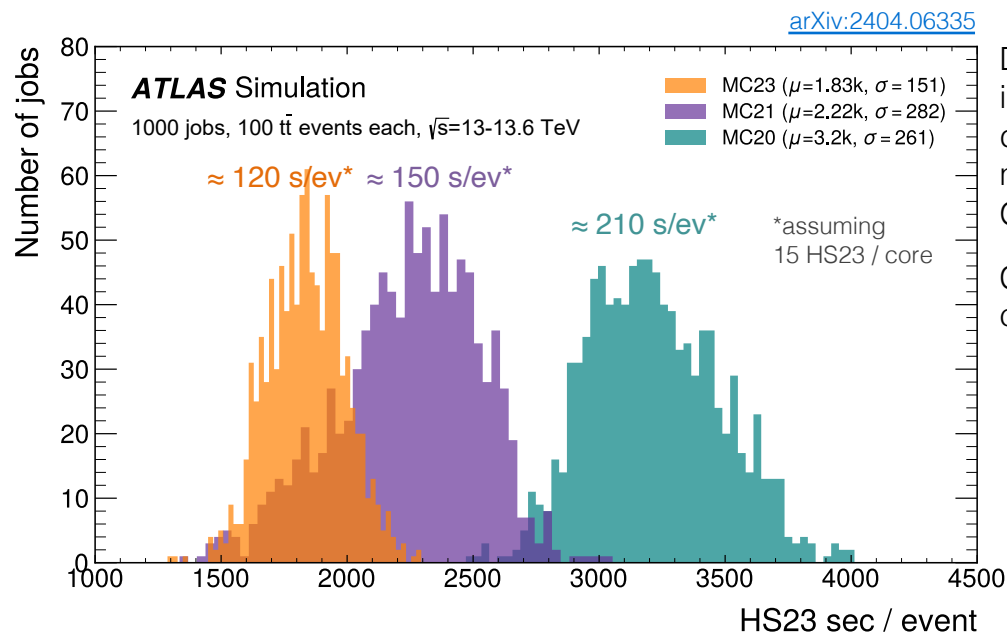
Detector simulation and generative ML

ATLAS simulation of
a Higgs boson event



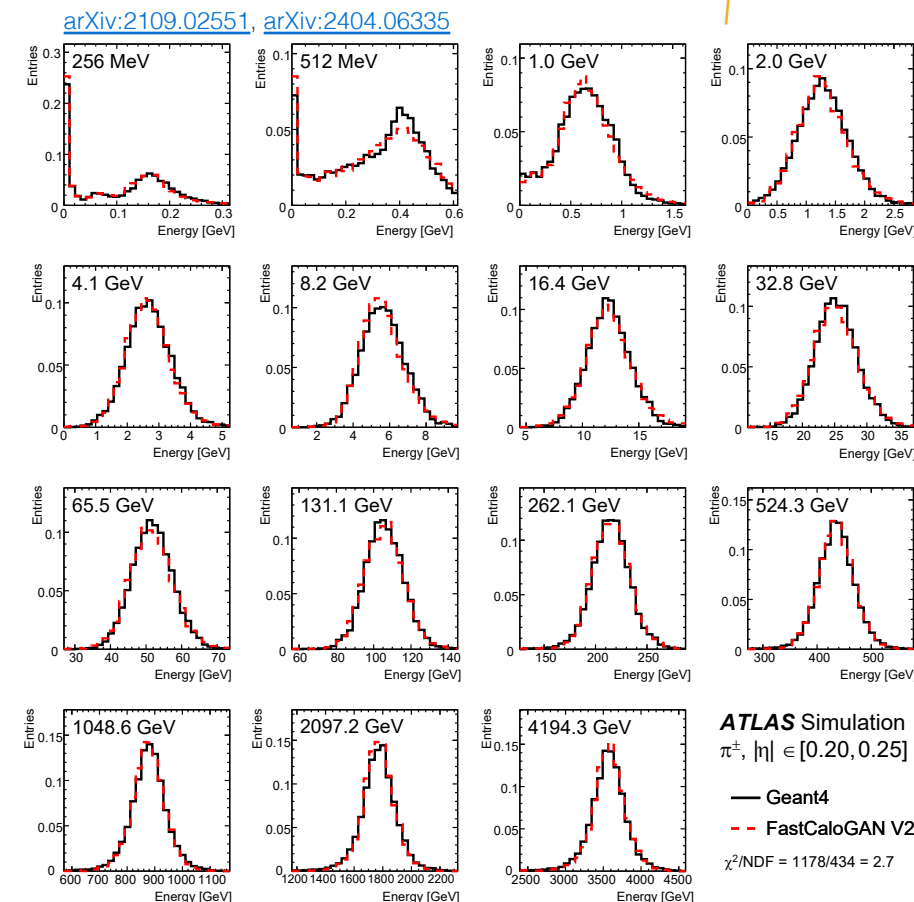
Detector simulation is traditionally performed by [Geant4](#) (G4)

- Geant4 simulates particle transport by stepping particles through the detector geometry, applying stochastic physics processes at each step, and recording energy deposits in sensitive volumes. These “hits” are then digitized into detector-like signals, so experiments can run reconstruction on them as if they were real data



Despite significant improvements, the detailed simulation necessarily makes G4 slow

CPU dominated by calorimeters



Fast simulation using generative ML algorithm trained with G4

- ATLAS developed a generative NN (FastCaloGAN) to simulate EM and HAD showers leading to an overall speed-up by 5 compared to G4

Precision objects enable precision measurements — Tracking

Charged particles deflection in magnetic field

- Lorentz force $\perp$ to B-field and to particle direction
- Particle trajectory projected onto plane $\perp$ to B-field is circle with radius: $r[\text{m}] \approx \frac{p_T[\text{GeV}]}{0.3 \cdot B[\text{T}]}$
 - ⇒ For $p_T = 10 \dots 1000 \text{ GeV} \rightarrow r = 17 \dots 1700 \text{ m}$ (cf, $r_{\text{ID}} \sim 1 \text{ m}$)

Obtain r and p_T from measurement of sagitta

$$r \approx \frac{L^2}{8s} \quad (\text{if } s \ll L) \quad \Rightarrow \quad p_T \propto \frac{1}{s} \quad \text{and} \quad \frac{\sigma(p_T)}{p_T} = \frac{\sigma(s)}{s} \propto p_T$$

Track fitting in LHC environment challenging

- Several thousand tracks per proton-bunch collision, up to 20k tracks for a central PbPb collision
- Must handle ambiguities, hit overlaps, **multiple scattering**, bremsstrahlung, multiple vertices, ...
- Track fitters take Gaussian noise (e.g., Kalman filter) and non-Gaussian noise (e.g., Gaussian sum filter) into account
- Fitter must be fast, used in high-level trigger



Multiple scattering (diffusion) of light passing through a wetter and wetter windscreen (left to right)

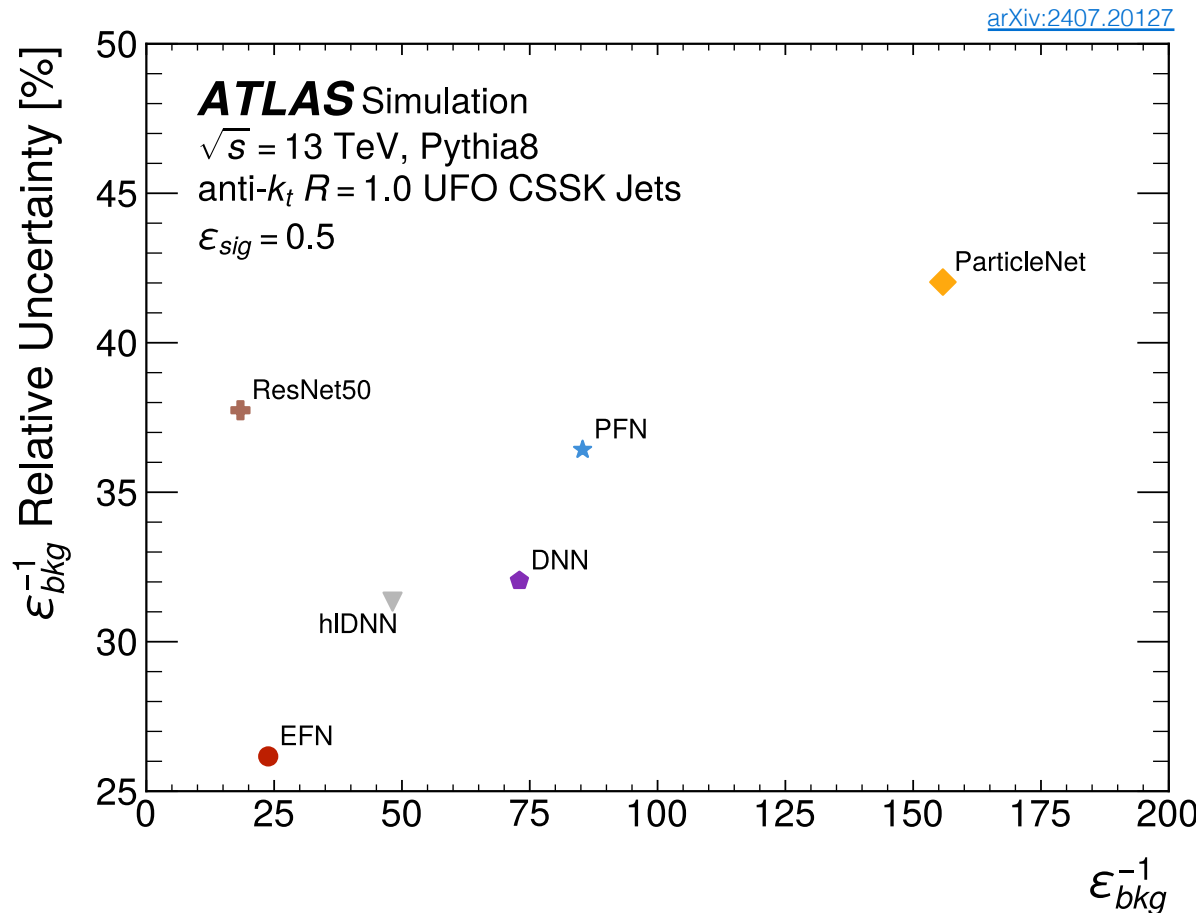
Lorentz boost and jet substructure

$$\text{Opening angle } \theta \sim 2\gamma^{-1} = 2m/p_T, \Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \approx \theta$$

At $p_T \gg m$, the hadronic decay products of W, Z, H and top become collimated and may merge into one reconstructed jet

$$\Delta R \approx 2m/p_T$$

Two-body, symmetric decay; $p_T \gg m$



Area of creative and powerful ML applications

Modelling uncertainty of constituent-based (except hIDNN) top taggers versus background rejection. Another prominent field of GNN applications in HEP

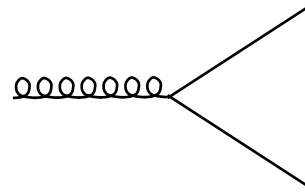
Detailed detector simulation is required to evaluate these taggers

ParticleNet: constituent based GNN, EFN: Energy-Flow Network, DeepSet structure, lower weights for low- p_T constituents, PFN: Particle Flow Network, as EFN, but without p_T weights, ResNet50: imaging CNN, DNN: densely connected NN with zero padding for variable structure, hIDNN: as DNN but with 15 high-level variables (subjettiness, splitting scales, thrust, etc)

Precision objects enable precision measurements — Jet

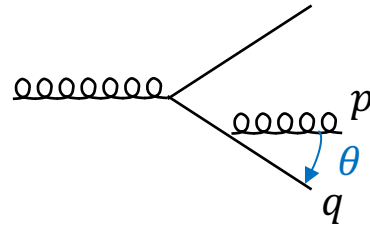
Jet: soft and collinear QCD radiation generates a parton shower, which hadronizes through colour confinement and is reconstructed as one or more jets

- Leading order:



No parton emission

- One real emission (NLO):



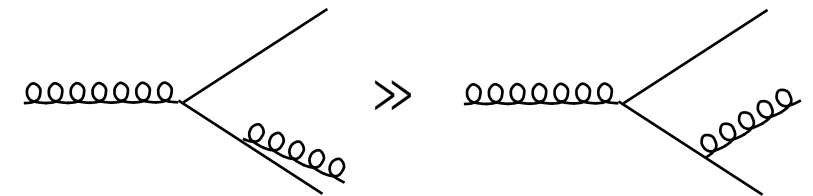
Propagator: $\frac{1}{(p+q)^2} \approx \frac{1}{2p \cdot q} = \frac{1}{2E_p E_q (1 - \cos\theta)} = \frac{1}{4E_p E_q \sin^2 \frac{\theta}{2}}$

Explodes if $E_p \rightarrow 0$ (soft divergence)
or $\theta \rightarrow 0$ (collinear divergence)

- Differential cross section:

$$d\sigma \propto \alpha_S \int_{\theta_{\min}}^R \int_{E_{\min}}^{E_{\max}} \frac{d\theta}{\theta} \frac{dE}{E} \sim \alpha_S \ln R \ln E \rightarrow$$

Double-log enhanced real-emission probability



Jets are experimentally defined using an infrared- and collinear-safe R -dependent algorithm, usually anti- k_t at the LHC