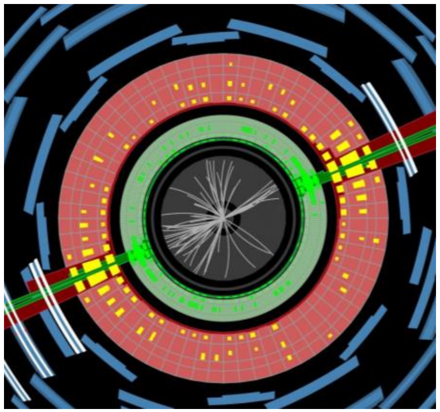
Data analysis in HEP

…with an emphasis on what we actually measure, and the role of software tools, especially simulation



Chacal, Witwatersrand University, RSA

16/1/2024

Jon Butterworth, University College London

Physics & Astronomy

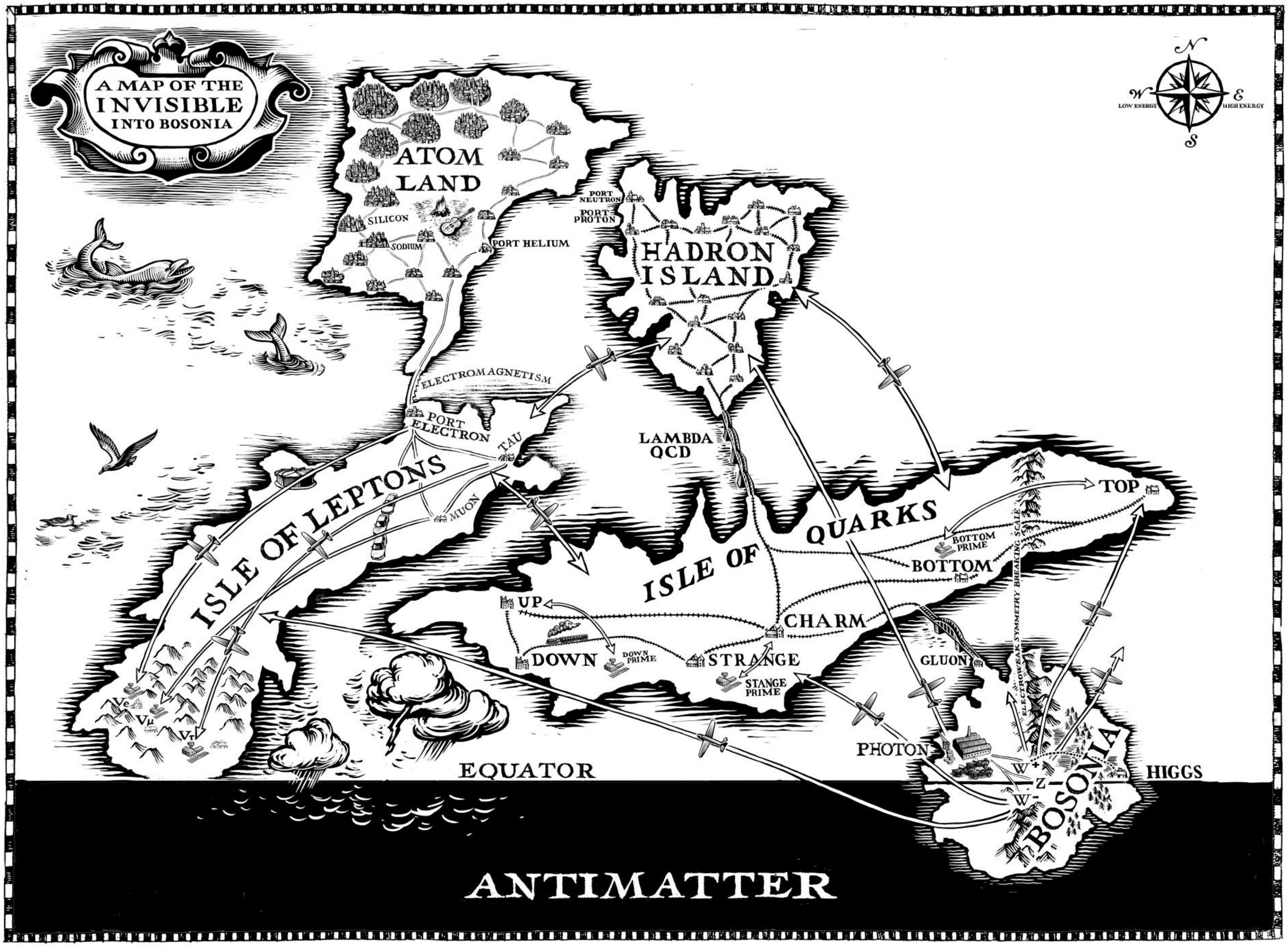
Who am I?

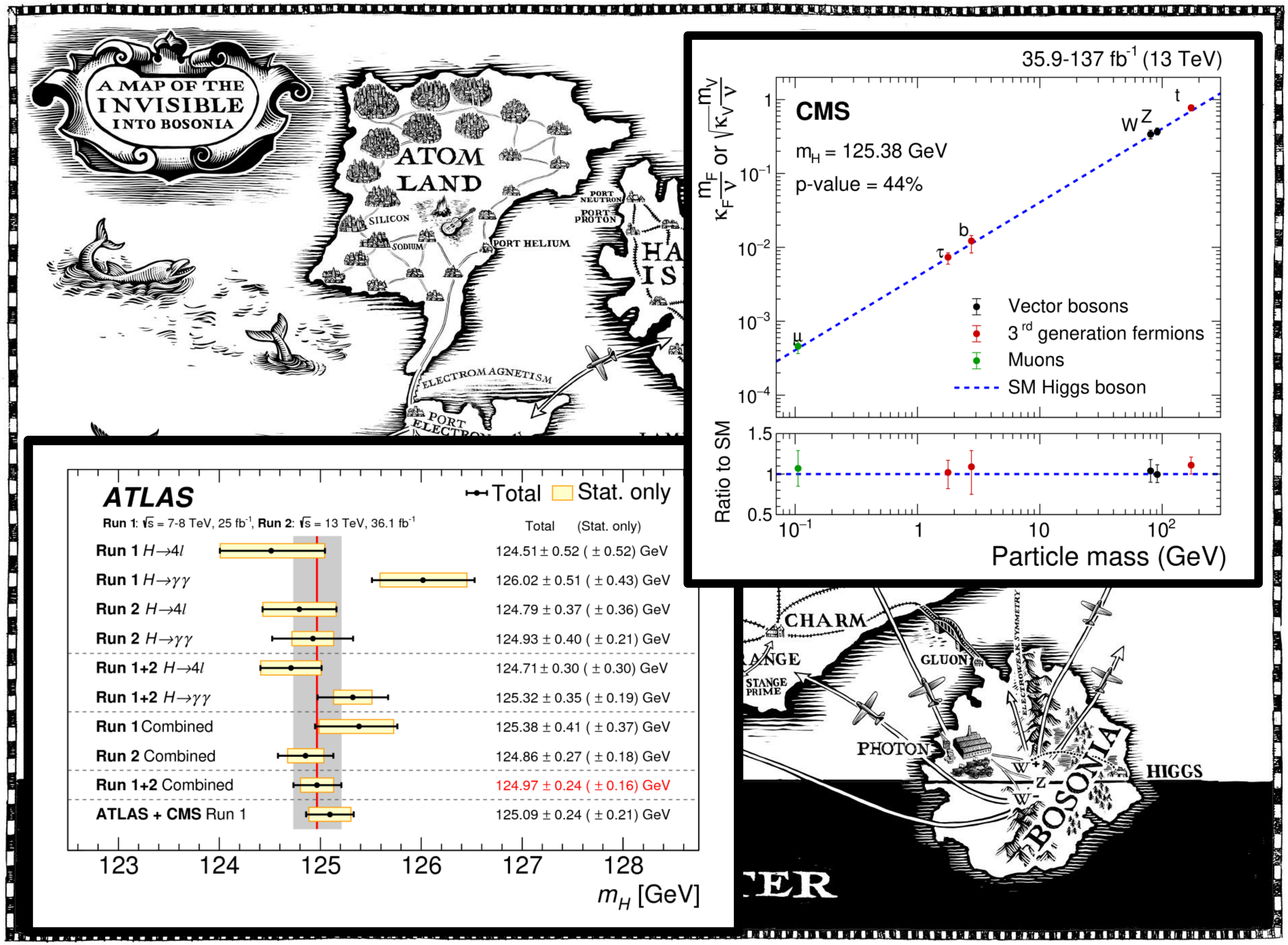
- Born 1967 (same age as the Standard Model).
- Grew up in Manchester, UK
- BA in Oxford 1986-1989
- DPhil also Oxford, on the ZEUS experiment in DESY, Hamburg, Germany 1989-1992
- Post doc Penn State, US (in Hamburg!) 1992-1995
- Lectureship at UCL 1995 and been there ever since (with a lot of time seconded to Hamburg and Geneva)
- Mainly ATLAS (CERN, LHC) since 2005
- Monte Carlo convener 2007-8, Standard Model convener 2010-2012, …
- …

Jon Butterworth

Jets, jet substructure.
Monte Carlo (MCnet)
Model independent measurements and their (re)interpretation

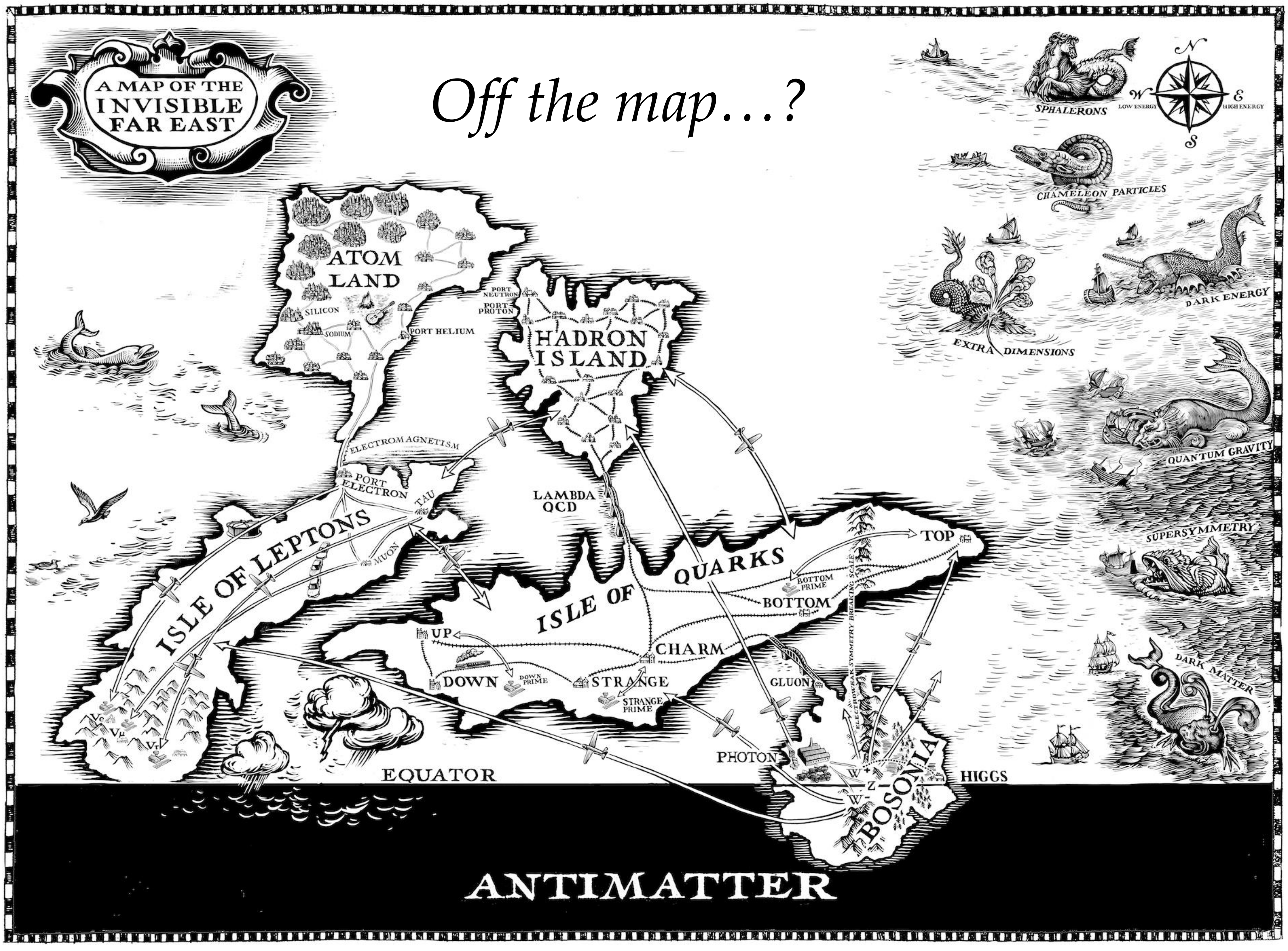
Two "popular" books (maps from one of them in this talk)
...


A MAP OF THE
INVISIBLE
INTO BOSONIA
N
W
E
S
LOW ENERGY
HIGH ENERGY
ATOM
LAND
SILICON
SODIUM
PORT HELIUM
PORT NEUTRON
PORT PROTON
HADRON
ISLAND
ELECTROMAGNETISM
PORT ELECTRON
TAU
MUON
LAMBDA QCD
ISLE OF LEPTONS
ISLE OF QUARKS
TOP
BOTTOM PRIME
BOTTOM
CHARM
UP
DOWN
DOWN PRIME
STRANGE
STANGE PRIME
GLUON
ELECTROWEAK SYMMETRY BREAKING SCALE
PHOTON
W+
Z
W-
BOSONIA
HIGGS
EQUATOR
ANTIMATTER



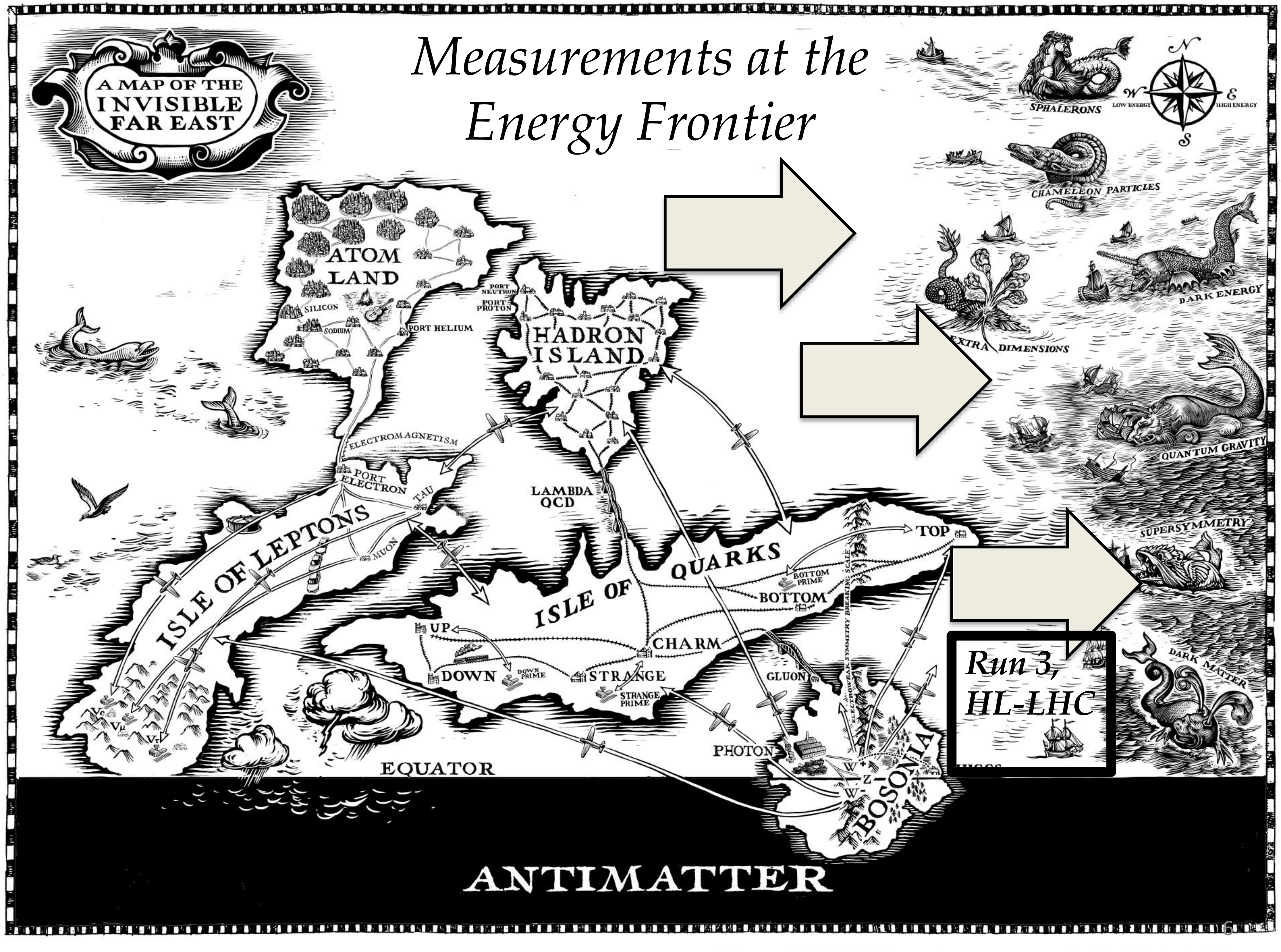
A MAP OF THE INVISIBLE INTO BOSONIA
ATOM LAND
SILICON
SODIUM
PORT HELIUM
PORT NEUTRON
PORT PROTON
ELECTROMAGNETISM
PORT ELECTRON
CHARM
GLUON
STANGE PRIME
ELECTROWEAK SYMMETRY
PHOTON
W+
Z
W-
HIGGS
BOSONIA
ATLAS
Total
Stat. only
Run 1: √s = 7-8 TeV, 25 fb-1, Run 2: √s = 13 TeV, 36.1 fb-1
Total (Stat. only)
Run 1 H→4l 124.51 ± 0.52 (± 0.52) GeV
Run 1 H→γγ 126.02 ± 0.51 (± 0.43) GeV
Run 2 H→4l 124.79 ± 0.37 (± 0.36) GeV
Run 2 H→γγ 124.93 ± 0.40 (± 0.21) GeV
Run 1+2 H→4l 124.71 ± 0.30 (± 0.30) GeV
Run 1+2 H→γγ 125.32 ± 0.35 (± 0.19) GeV
Run 1 Combined 125.38 ± 0.41 (± 0.37) GeV
Run 2 Combined 124.86 ± 0.27 (± 0.18) GeV
Run 1+2 Combined 124.97 ± 0.24 (± 0.16) GeV
ATLAS + CMS Run 1 125.09 ± 0.24 (± 0.21) GeV
123 124 125 126 127 128
mH [GeV]
35.9-137 fb-1 (13 TeV)
CMS
mH = 125.38 GeV
p-value = 44%
κF mF/v or √κV mV/v
μ τ b W Z t
Vector bosons
3rd generation fermions
Muons
SM Higgs boson
Ratio to SM
Particle mass (GeV)

Off the map…?



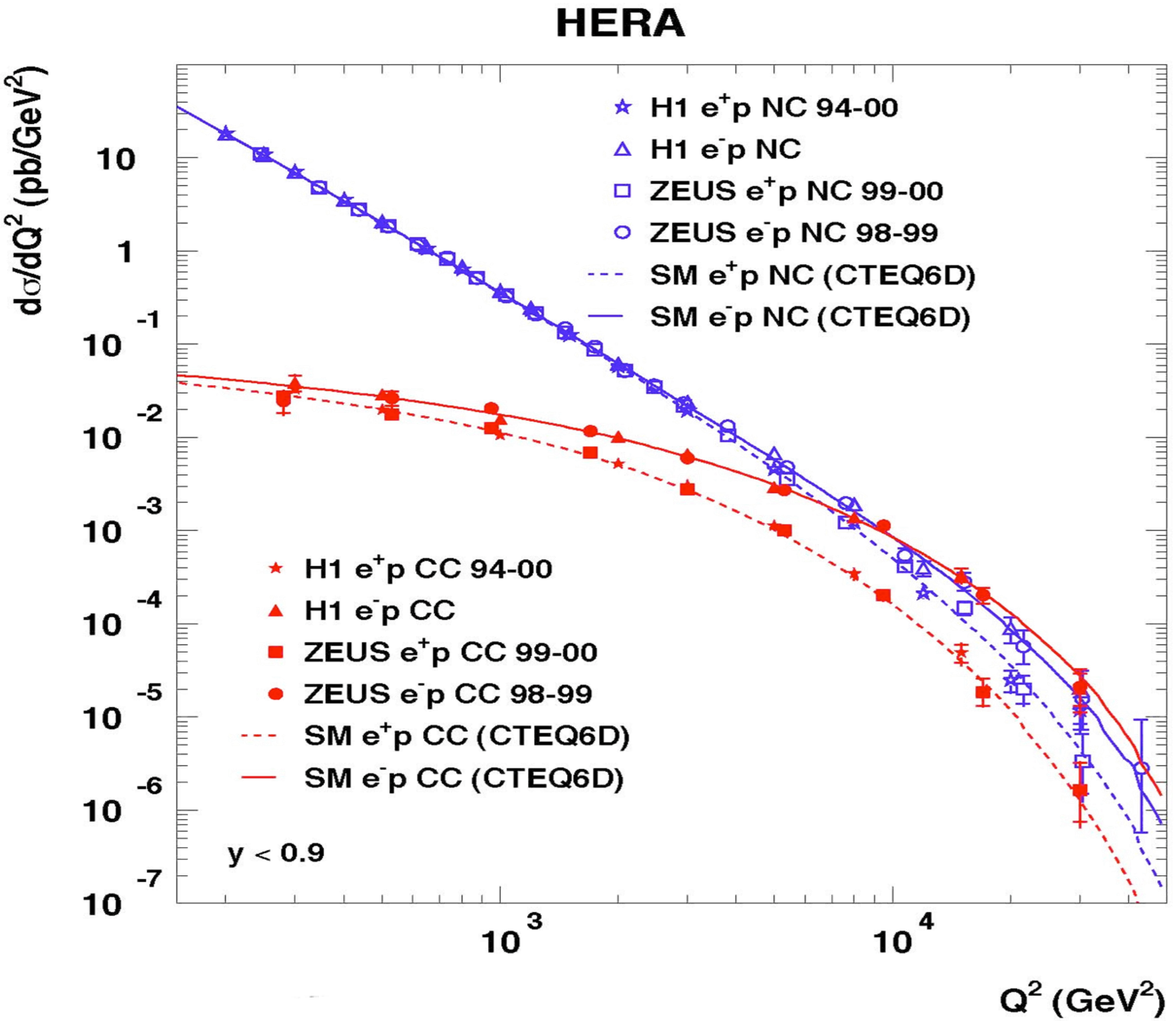
A MAP OF THE INVISIBLE FAR EAST
ATOM LAND
SILICON
SODIUM
PORT HELIUM
PORT NEUTRON
PORT PROTON
HADRON ISLAND
ELECTROMAGNETISM
PORT ELECTRON
TAU
MUON
ISLE OF LEPTONS
LAMBDA QCD
ISLE OF QUARKS
TOP
BOTTOM PRIME
BOTTOM
CHARM
UP
DOWN
DOWN PRIME
STRANGE
STRANGE PRIME
GLUON
PHOTON
ELECTROWEAK SYMMETRY BREAKING SCALE
W+
Z
W-
BOSONIA
HIGGS
EQUATOR
ANTIMATTER
N
W
E
S
LOW ENERGY
HIGH ENERGY
SPHALERONS
CHAMELEON PARTICLES
DARK ENERGY
EXTRA DIMENSIONS
QUANTUM GRAVITY
SUPERSYMMETRY
DARK MATTER

Measurements at the Energy Frontier



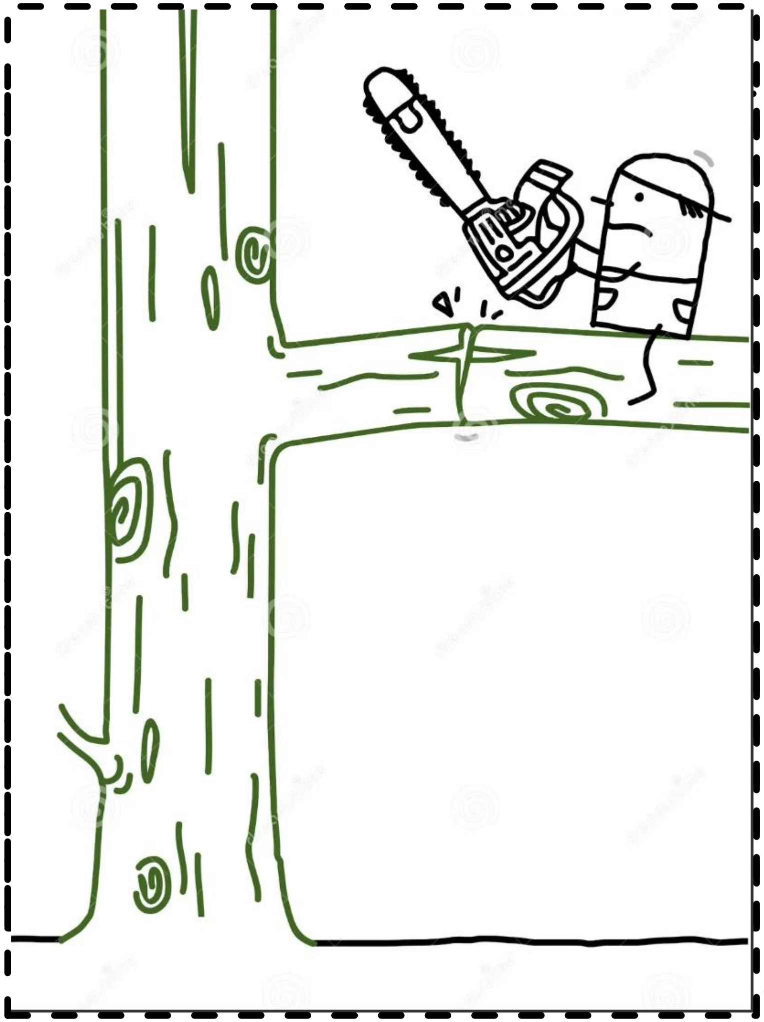
Run 3, HL-LHC

We are exploring the unknown



The power of computing and simulation

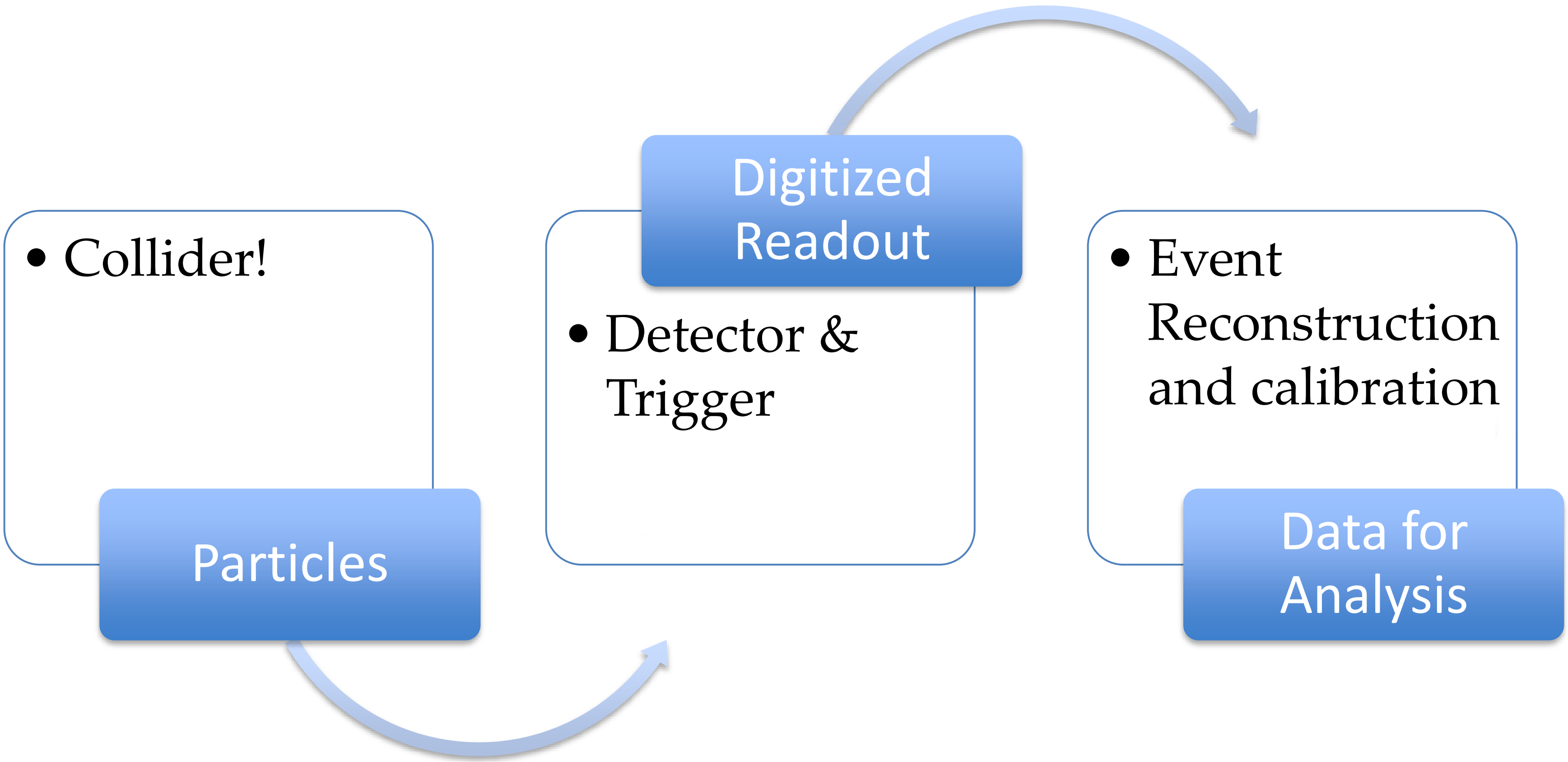
- Simulation, of our detectors and of the physics we probe with them, is an essential and increasingly powerful part of the calibration, analysis, measurement and interpretation process at colliders.



The Power: 1

- The role of Monte Carlo event generators and dectector simulation in a measurement

Simulation and Experiment



Simulation and Experiment

- MC Event Generator

Particle Four-Vectors

Digitized Readout

- Detector & Trigger Simulation

- Event Reconstruction, calibration, scale factors

Data for Analysis

Simulation and Experiment

- MC Event Generator

Particle Four-Vectors

Digitized Readout

- Detector & Trigger Simulation

- Event Reconstruction, calibration, scale factors

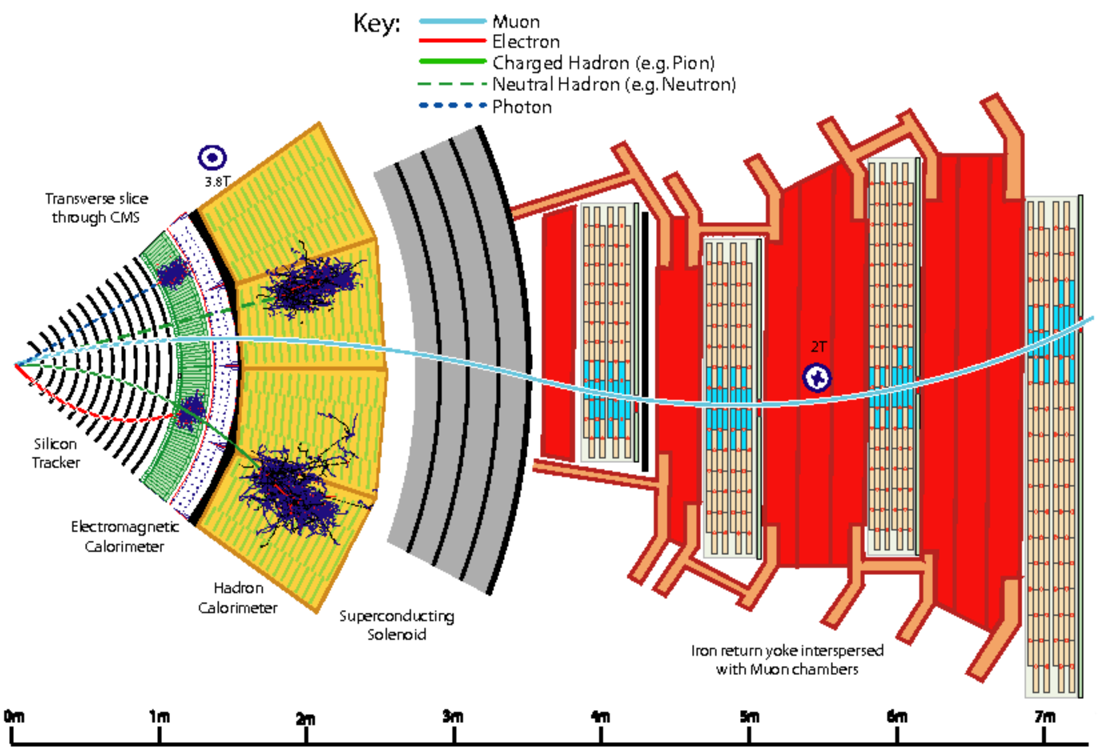
Data for Analysis

Unfolding & Data Correction:
Test and evaluate

"Detector effects"

- Efficiencies: there is a non-zero probability that a particle passing through a detector will not be reconstructed
- Fake backgrounds: there is a non-zero probability that a particle will be reconstructed even though it wasn't really there
- Smearing: the measured energies, momenta, angles of the particles and jets will be smeared due to the intrinsic resolution of the detectors

We need to know what our detector is doing so we can account for it and in some cases reverse it **(the red arrow)**



Simulation and Experiment

- Collider!

Particles

Digitized Readout

- Detector & Trigger

- Event Reconstruction and calibration

Data for Analysis

Unfolding & Data Correction:
Make the measurement!

What is a final state particle?

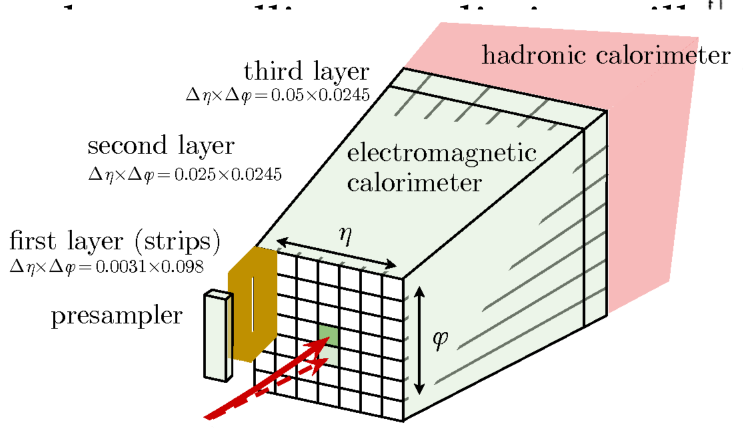
- Colour triplets (or octets) are not final state particles
  - Neither nature, nor our event generators, guarantee the physicality (or even the presence) of a "final-state top", for example
- Electroweak-scale particles (W, Z, H) are not final state particles
  - Decay lifetime is so short that coherence/interference effects cannot be neglected
  - As above. Focus on the leptons, hadrons, photons
- Operational definitions usually involve a lifetime cut (10ps), and/or distinguish between pre/post hadronization
  - Choices to make about t, B-hadrons etc
- Algorithmic combinations of final state objects
  - Hadronic Jets, Dressed leptons, Photon Isolation…

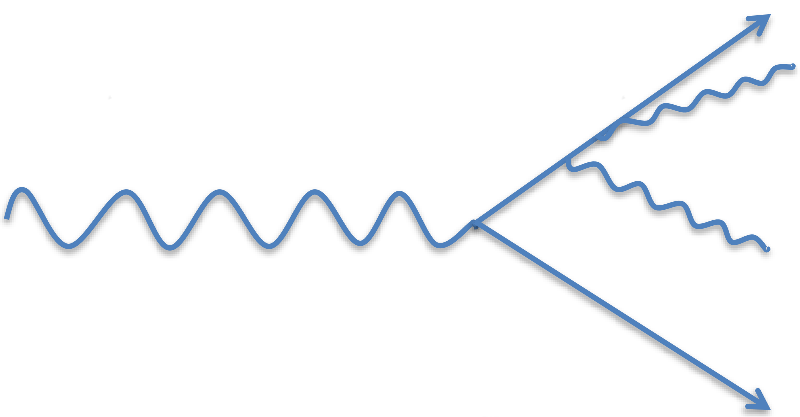
What is a final-state electron/muon?

- Electrons/muons from hadron decays are typically removed in the data analysis by isolation cuts / fake removal
  - ✓ Can define "prompt leptons" to be "not-from-hadron decays" and only consider these : this is more robust and model-independent than asking that the lepton comes from a certain propagator in the hard process
  - ✓ Well defined in Rivet (see tutorial), but you may need to also implement it in your experiment's software
- We don't usually define particle-level isolation, but rather correct for inefficiencies of these requirements
  - ➢ It might be worth reconsidering this in specific analyses where proximity to jets has a large effect on results

What is a final-state electron/muon?

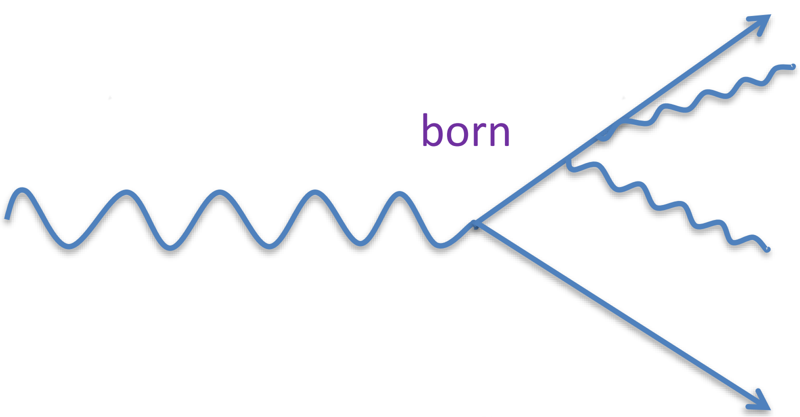
- Electrons and muons emit FSR photon radiation (and lots of it, especially in the collinear limit, especially for electrons).
  - ➢ For muons we measure the charged particle track, photon energy is not included
  - ➢ For electrons we cluster calorimeter cells be included in the energy measurement





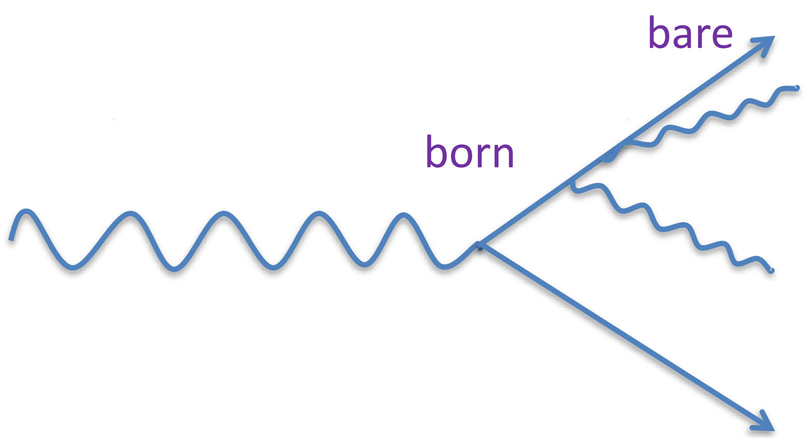
What is a final-state electron/muon?

- Electrons and muons emit FSR photon radiation (and lots of it, especially in the collinear limit, especially for electrons).
  - ➢ For muons we measure the charged particle track, photon energy is not included
  - ➢ For electrons we cluster calorimeter cells and most collinear radiation will be included in the energy measurement
- We can define lepton momenta as:
  1. <u>Born leptons</u> – as if FSR never happened (not what we measure)



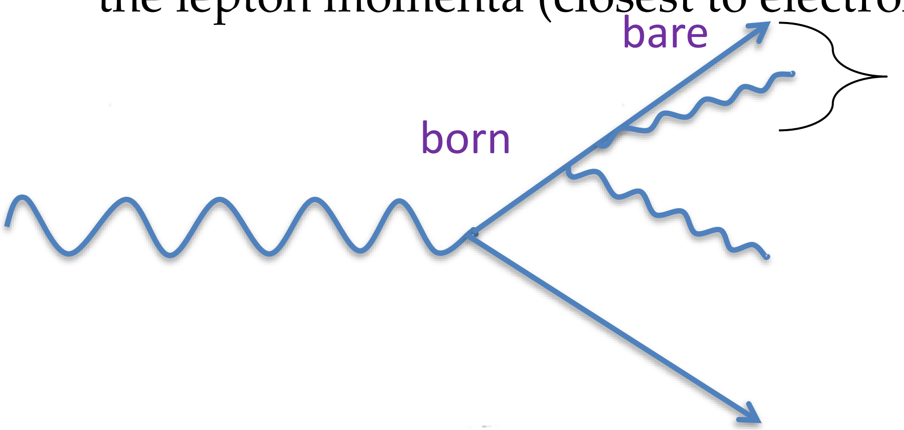
What is a final-state electron/muon?

- Electrons and muons emit FSR photon radiation (and lots of it, especially in the collinear limit, especially for electrons).
    - For muons we measure the charged particle track, photon energy is not included
    - For electrons we cluster calorimeter cells and most collinear radiation will be included in the energy measurement
- We can define lepton momenta as:
    1. <u>Born leptons</u> – as if FSR never happened (not what we measure)
    2. <u>Bare leptons</u> – after all FSR (closest to muon measurement)



What is a final-state electron/muon?

- Electrons and muons emit FSR photon radiation (and lots of it, especially in the collinear limit, especially for electrons).
  - ➢ For muons we measure the charged particle track, photon energy is not included
  - ➢ For electrons we cluster calorimeter cells and most collinear radiation will be included in the energy measurement
- We can define lepton momenta as:
  1. Born leptons – as if FSR never happened (not what we measure)
  2. Bare leptons – after all FSR (closest to muon measurement)
  3. Dressed leptons – with the momenta of close-by photons "clustered" into the lepton momenta (closest to electron measurement)



dressed : typically a $\Delta R < 0.1$ cone is used, but a jet algorithm may be better

What is a final-state electron/muon?

- Electron and muon final states can be very different for bare leptons, but much closer for born and dressed leptons: see Rivet tutorial

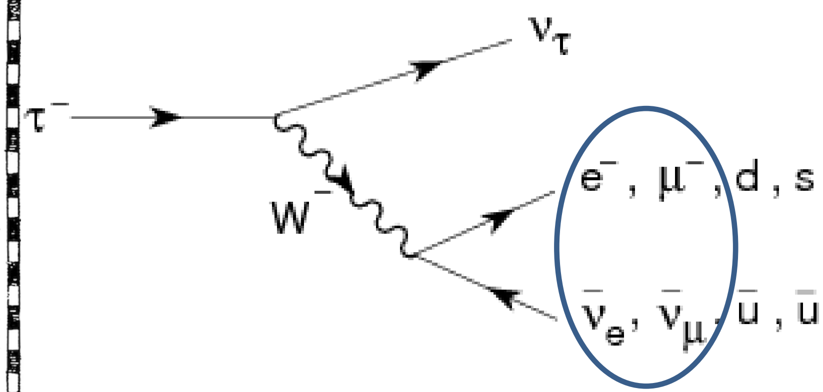
- It is often argued that dressed should be used for both to allow for *easy combination of final states.* Also bare versus dressed is much closer for muons than bare versus dressed for electrons

- Similarly, fiducial phase space cuts often harmonized for the two, requiring a small extrapolation in phase space for one

  - But electrons != muons
  - We may want to retain sensitivity to differences (cf LHCb…)
  - Perhaps it is better to measure both and publish correlations between uncertainties, and make choices that are best for each individual channel

What is a final-state tau?

Recall: unstable

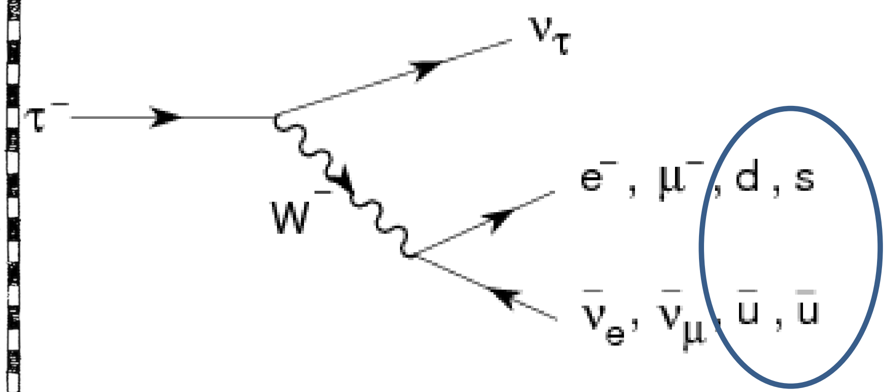


<u>Leptonic decays</u>

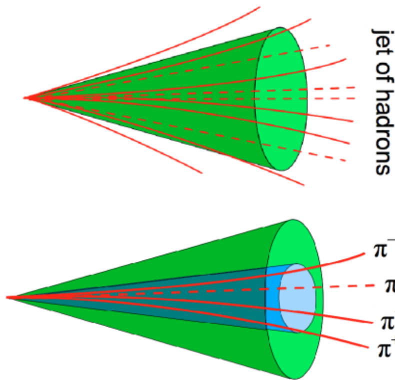
- The final state particles are electrons/muons and neutrinos
- Define fiducial phase-space with those (but we careful to check lepton efficiencies as e.g. impact parameter cuts can be less efficient for leptons from taus)

What is a final-state tau?

Recall: unstable (=0.1 mm



$\tau^\pm \rightarrow \pi^\pm \pi^0 \nu$
$\tau^\pm \rightarrow \pi^\pm \nu$
$\tau^\pm \rightarrow \pi^\pm \pi^0 \pi^0 \nu$
$\tau^\pm \rightarrow \pi^\pm \pi^\pm \pi \ \ \nu$
$\tau^\pm \rightarrow \pi^\pm \pi^\pm \pi \ \ \pi^0 \nu$

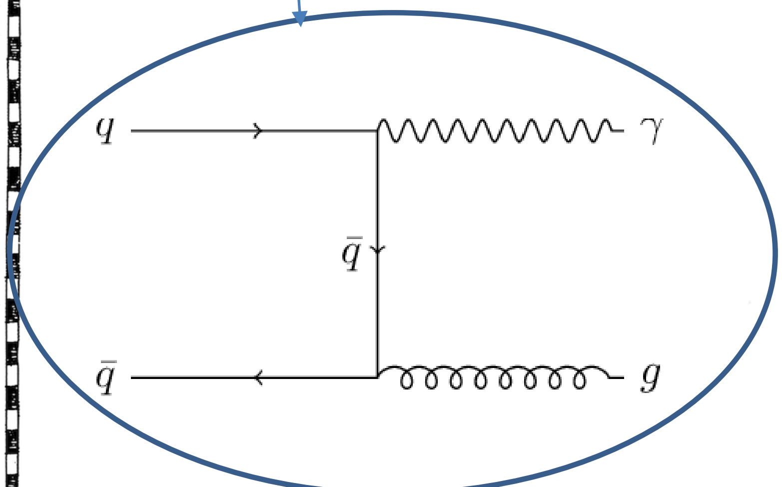


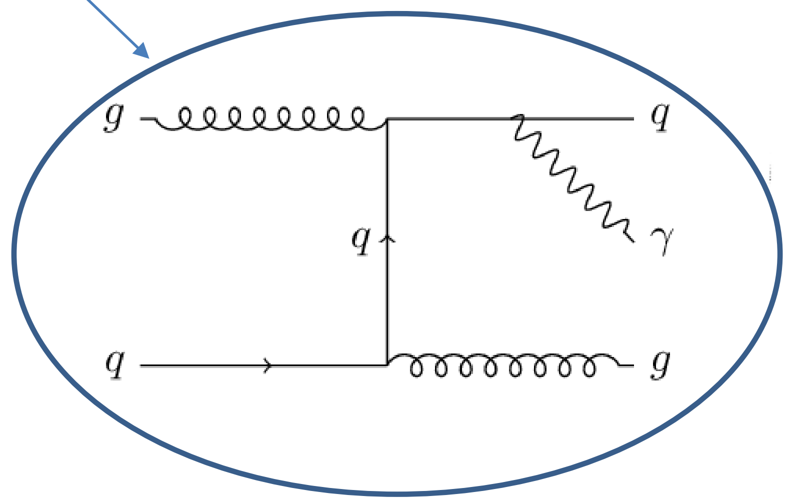
<u>Hadronic decays</u>

- Final state particles are hadrons (→ jets) and neutrinos
- Such a definition alone is complicated due to the large number of hadrons not from taus
- Experimental cuts reject backgrounds based on features of the jets, which are hard to replicate at the particle-level
- In this case a compromise might be best: require a hadron in the jet to have come from a prompt tau (this is not quite "final-state based")
- There is not much experience here, more detailed studies would be interesting

What is a final-state photon?

- Analyses usually measure prompt, isolated photons
- Recall: Prompt means not-from-a-hadron-decay
- In the calculation, prompt photons can be further divided into those from the "hard scatter" and those from parton fragmentation





A particle-level isolation criteria is necessary to replicate the isolation applied at reconstruction-level

Note in principle this could also be done for prompt leptons, but it is much less important

What is a final-state neutrino?

Invisible in the detector and existence inferred by p_T^{miss}

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = \underbrace{\underbrace{-\sum_{\substack{\text{selected}\\ \text{electrons}}} \mathbf{p}_{\mathrm{T}}^{e}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}} \underbrace{-\sum_{\substack{\text{accepted}\\ \text{photons}}} \mathbf{p}_{\mathrm{T}}^{\gamma}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}} \underbrace{-\sum_{\substack{\text{accepted}\\ \tau\text{-leptons}}} \mathbf{p}_{\mathrm{T}}^{\tau_{\mathrm{had}}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\tau_{\mathrm{had}}}} \underbrace{-\sum_{\substack{\text{selected}\\ \text{muons}}} \mathbf{p}_{\mathrm{T}}^{\mu}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}} \underbrace{-\sum_{\substack{\text{accepted}\\ \text{jets}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mathrm{jet}}}}_{\text{hard term}} \underbrace{\underbrace{-\sum_{\substack{\text{unused}\\ \text{tracks}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mathrm{soft}}}}_{\text{soft term}}$$

- Sometimes the momenta of (prompt?) invisible* particles are summed

- An alternative is to take – the sum of all the visible particles within detector acceptance, which is closer to what we measure but can be a bit complicated. E.g. what p_T of hadrons are we actually sensitive to?

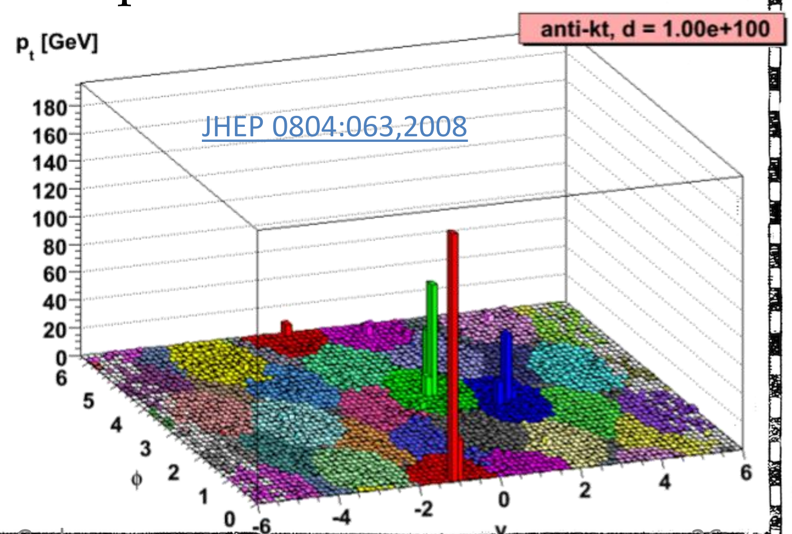
(More on this later)

*neutrinos are indistinguishable from BSM invisible particles

What is a final-state parton?

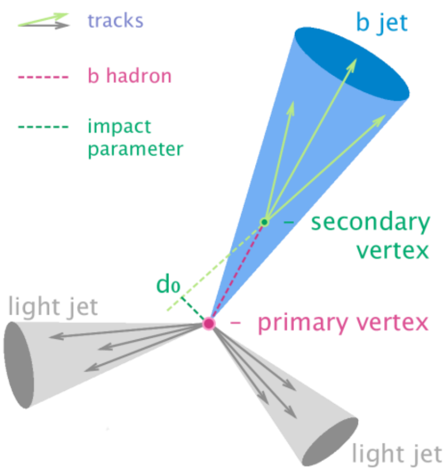
- Partons radiate more partons which hadronize.
- Run a jet algorithm on the final-state particles
  - Form a list of particles (this would be clusters / tracks at reco-level)
  - Merge the smallest pair according to a “distance” parameter
  - Iterate
- Algorithms assign each hadron to a jet. The energy/momentum of the jet represents the energy/momentum of the parton from the hard scatter
- Think carefully about what is included as inputs: Muons? Neutrinos?

Note: Depending on the reconstruction code, an electron will often form a jet initially. We remove these jets using overlap removal at both reco- and truth-level (e.g. remove any jets with $\Delta R < 0.4$ from a prompt electron)

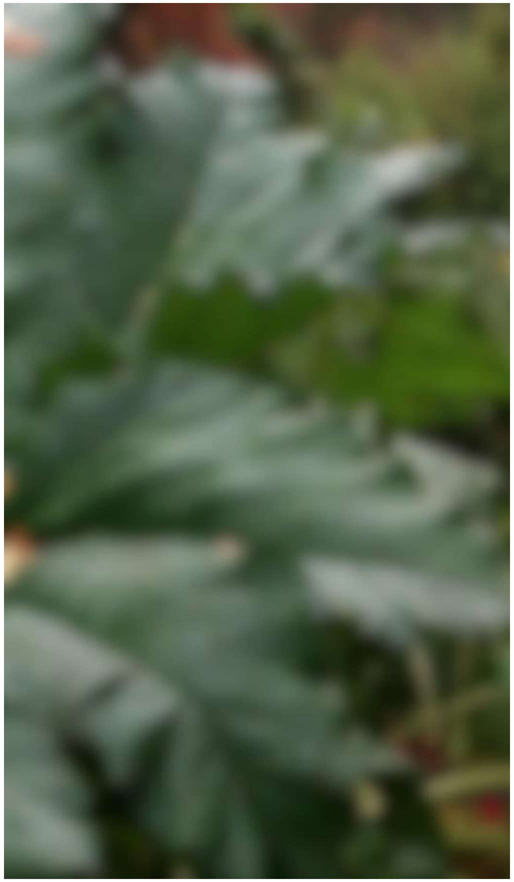


What is a final-state b-jet?

- Decay length for a 20 GeV b-hadron ~2 mm, they are therefore unstable and not included as final state particle
- However we select them experimentally by making displaced vertex selection cuts

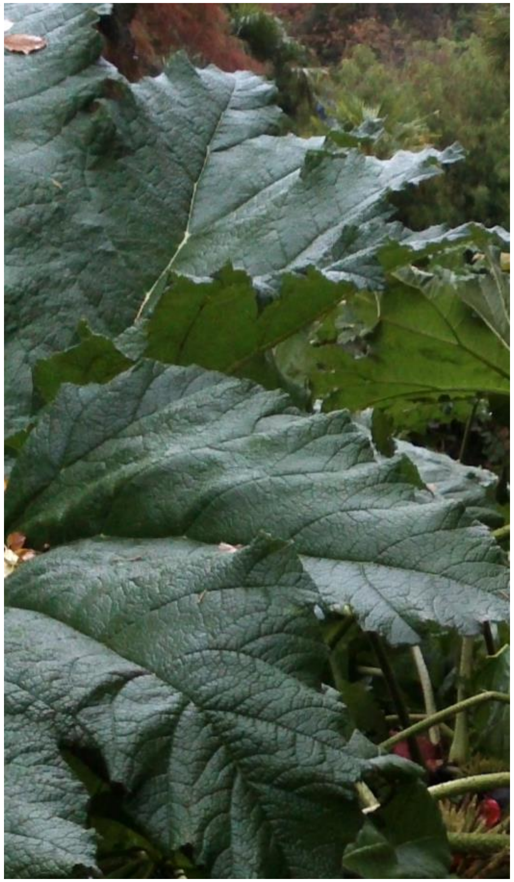


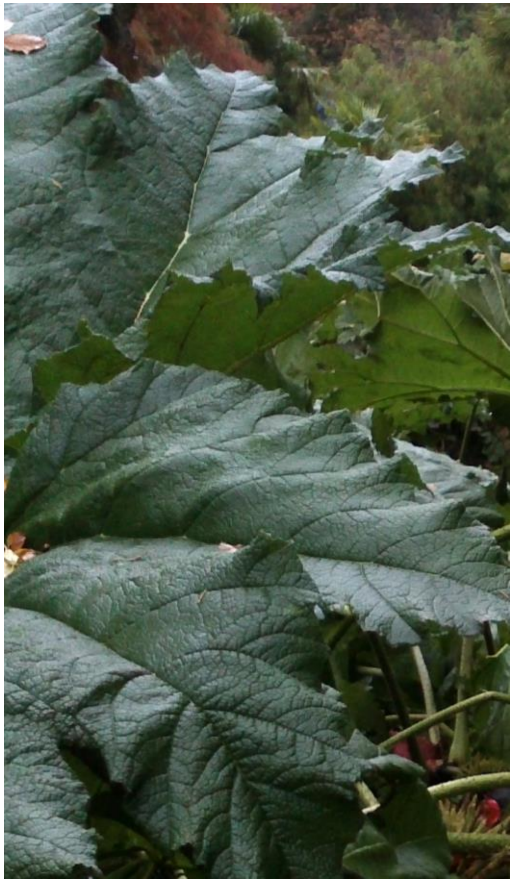
- Common "compromise" is to associate the *non-final state b*-hadrons to jets.
- If a jet contains a b-hadron it is considered a particle-level b-jet



What is a fiducial cross section?

Unfold





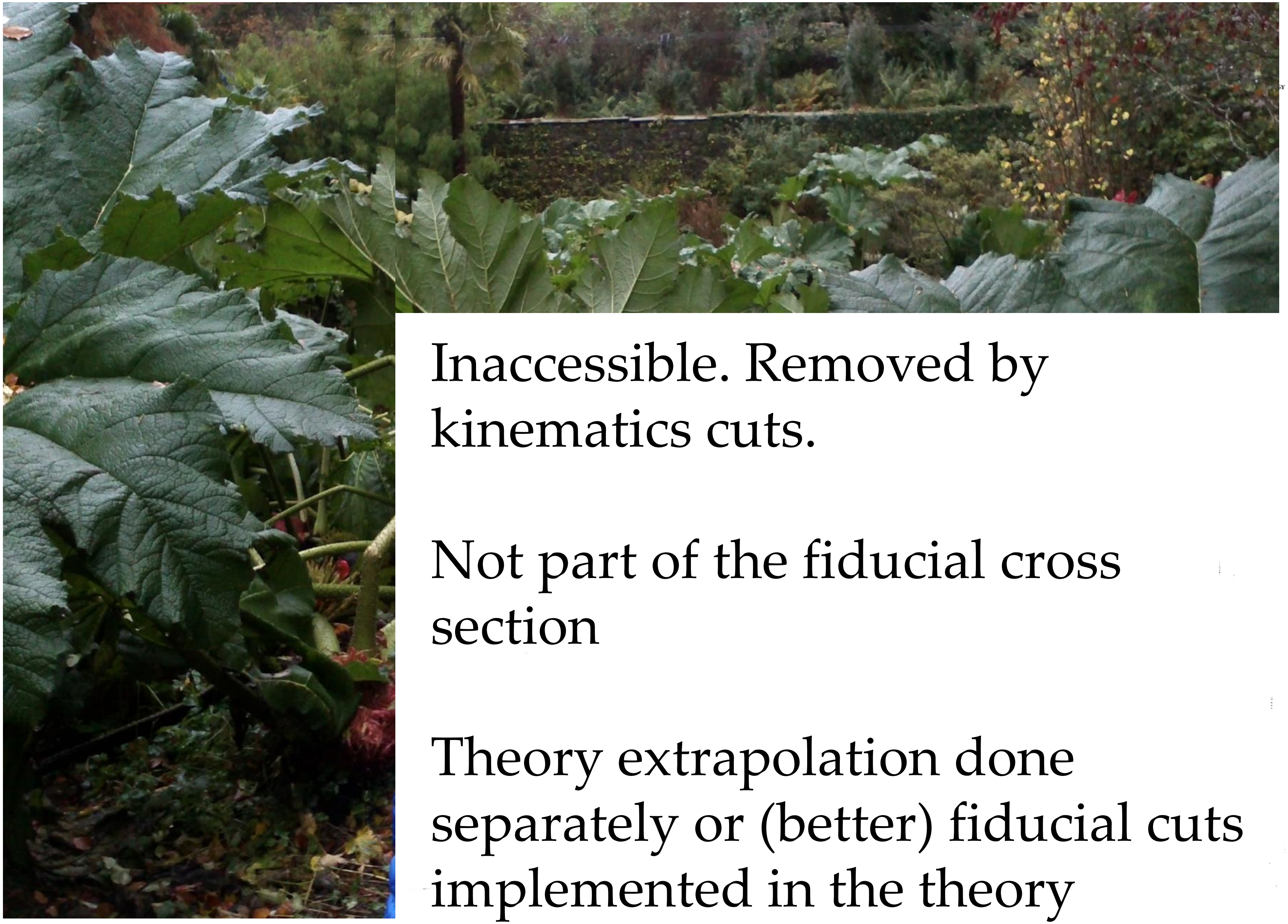
Increase acceptance

Increase acceptance

Extrapolate


But how reliably?

Fiducial phase-space



Inaccessible. Removed by kinematics cuts.

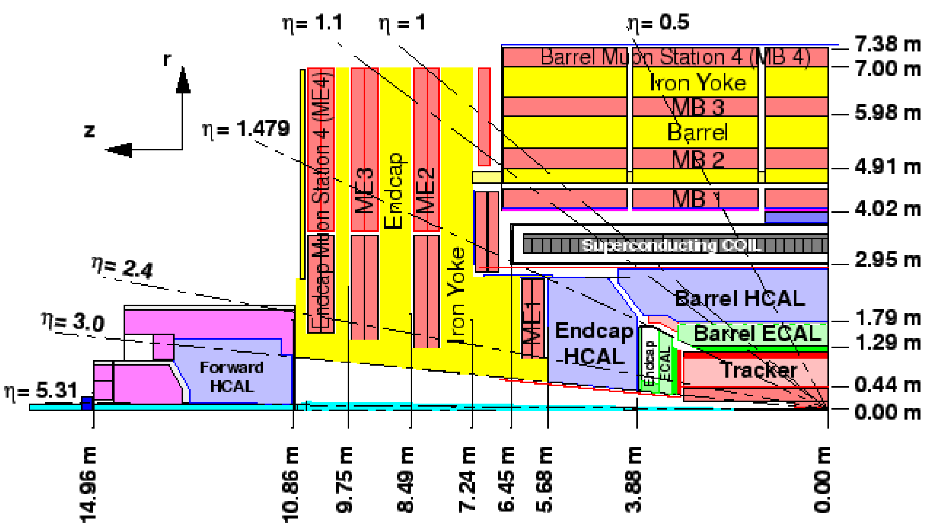
Not part of the fiducial cross section

Theory extrapolation done separately or (better) fiducial cuts implemented in the theory

Fiducial phase-space

Irrespective of detector efficiencies and resolution, there are particular kinematic regions that *we just don't measure at all.*

We do not have 4π detectors and we can't go down to zero p_T!

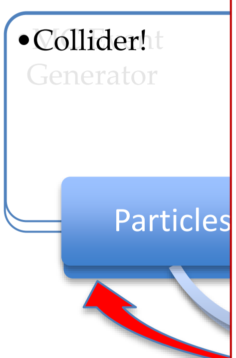


A fiducial phase-space is a set of selection criteria that can be applied to **final-state "truth" particles**

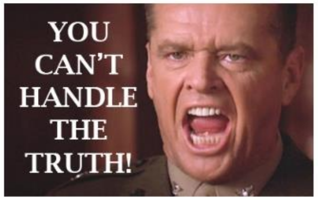
e.g.: Select events with one (and only one) muon with $p_T > 25$ GeV, $|\eta|<2.4$ and $p_T^{miss} > 30$ GeV.

Recap

- Outline of mak



- The SHG generator (or LHE) event record is often called "truth" – a dangerous phrase.
- We're doing quantum mechanics: there is no unambiguous truth!
  ⇒ event records are half-physics, half-debug-info. . . and zero indication of amplitude interference
- It gets worse: kinematic frames aren't defined (until the final-state) & momentum isn't necessarily conserved at vertices!
- **BEWARE!!**



Andy B, yesterday

- Defining what we measure
  - (Choice of) final state particles
  - Definition of fiducial phase space

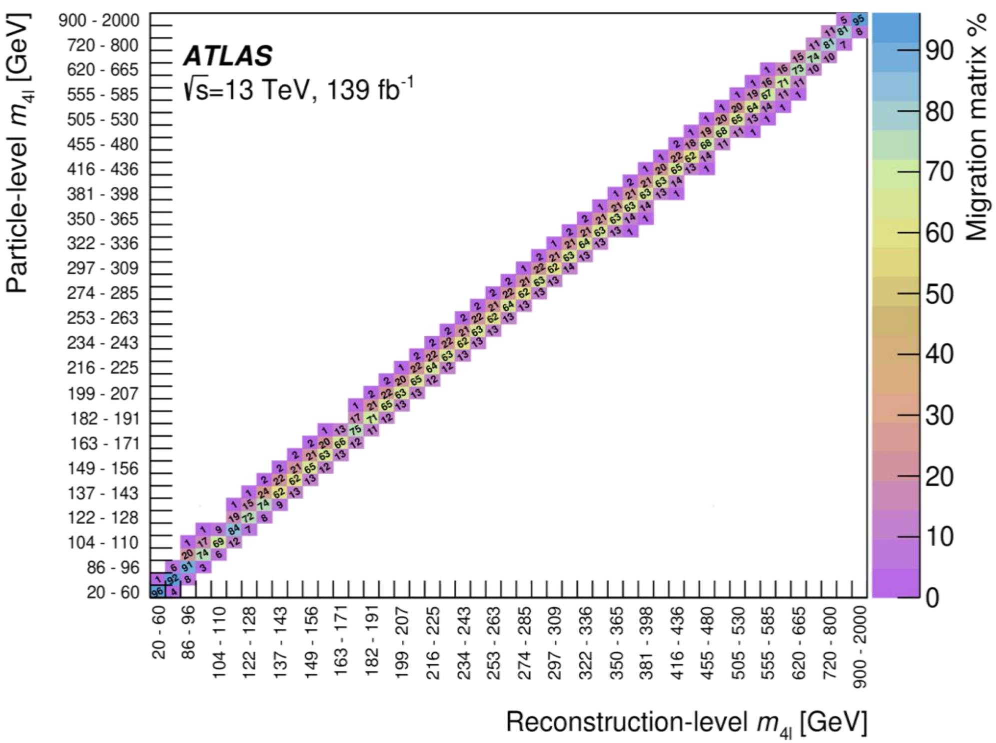
Today

- A bit more on unfolding
- Background (subtraction?)
- Jets and substructure

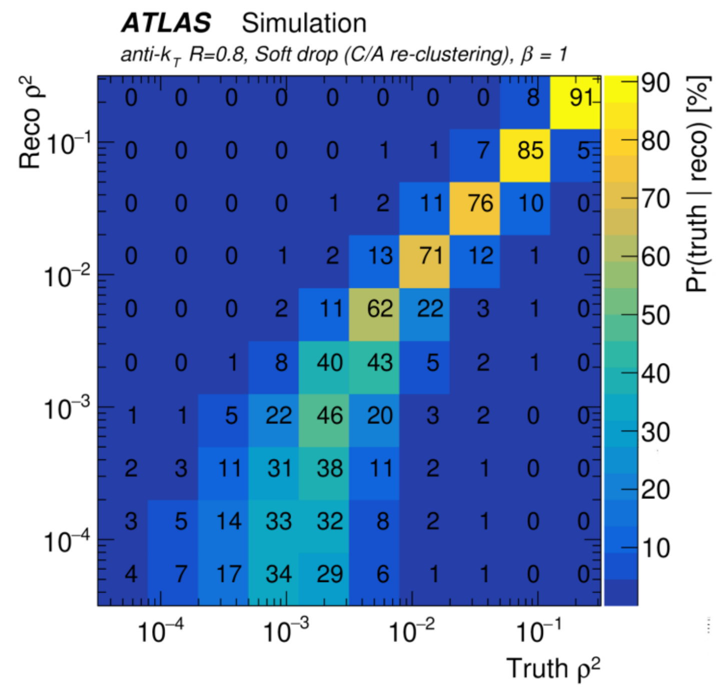
Unfolding to Particle Level

- If you
  - Have already calibrated the detector/reconstruction
  - Defined the **final state** carefully
  - Used this to define a fiducial phase space
  - Used a simulated prior that describes all relevant distributions
- … then "unfolding" is not a big final step
- Several standard techniques and implementations available

Unfolding Matrices (Examples)



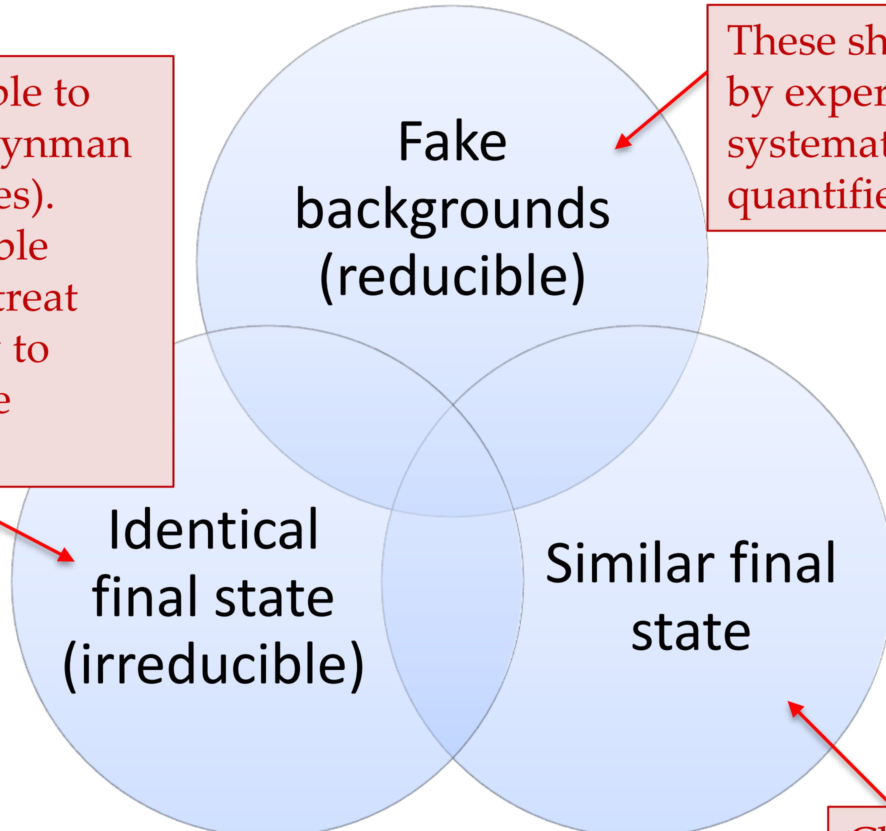
arXiv:2103.01918



arXiv:1711.08341

Background subtraction (or not?)

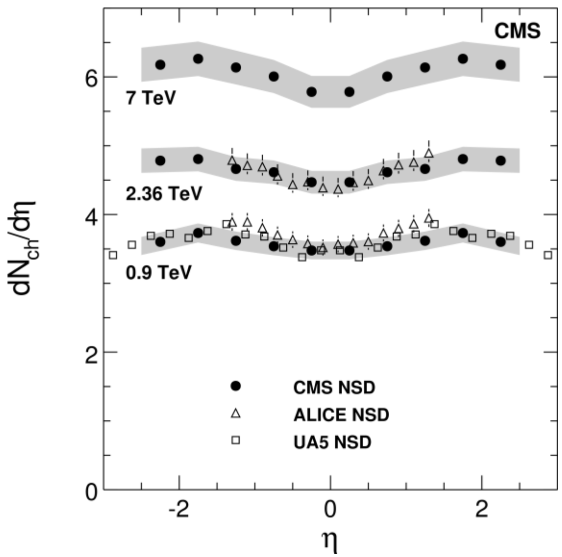
In general not possible to isolate individual Feynman diagrams (amplitudes). Sometime a reasonable approximation, but treat with caution and try to measure an inclusive observable too!



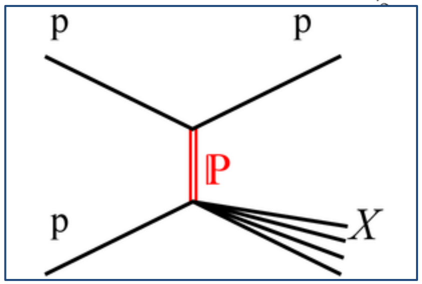
These should be subtracted by experimentalists and systematic uncertainties quantified

Choices, choices…

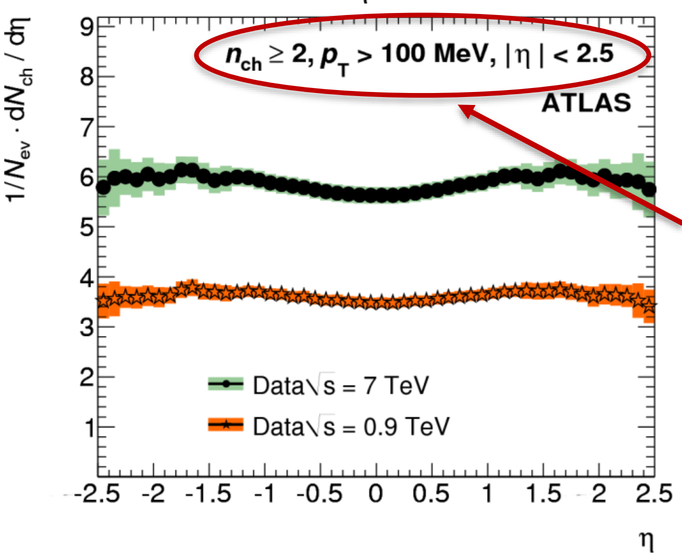
Similar final-states example: "Minimum Bias"



"Non-Single-Diffractive". That means that "single diffractive" events are subtracted using a MC simulation.



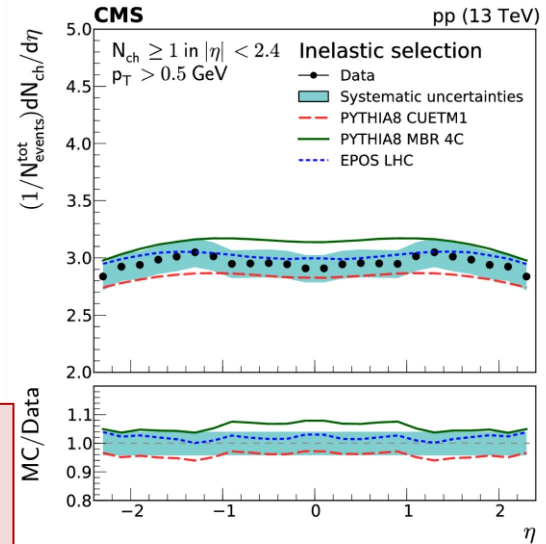
[Phys. Rev. Lett. 105 (2010) 022002](#)
Rivet: CMS_2010_S8656010



Fiducial final-state particle definition

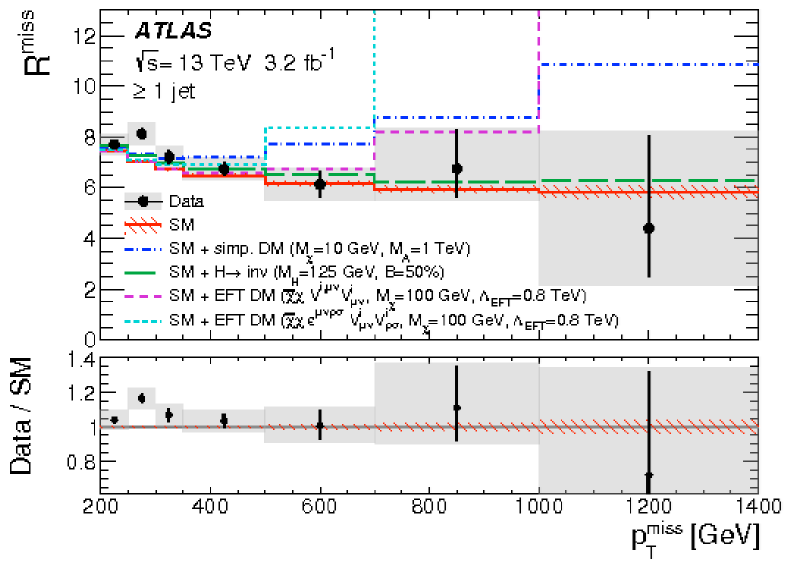
[New J. Phys. 13 (2011) 053033](#)
Rivet: ATLAS_2010_S8918562

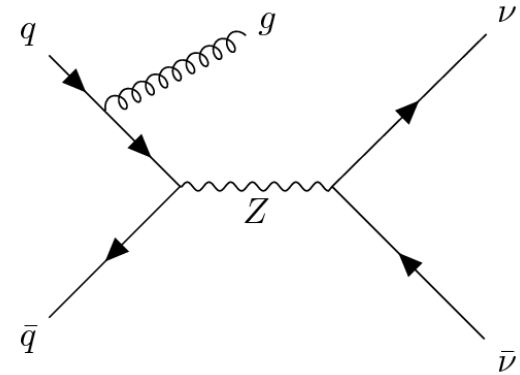
*All events with at least one particle satisfying this are **signal**.*



[Eur. Phys. J. C 78 (2018) 697](#)
Rivet: CMS_2018_I1680318

Similar final-states example: p_T^{miss} + jets





Eur. Phys. J. C 77 (2017) 765
Rivet: ATLAS_2017_I1609448

<u>Fiducial phase-space</u>: p_T^{miss} + jet(s)
p_T^{miss} > 200 GeV
No charged leptons with $|\eta|$< 2.5, p_T > 7 GeV

W→lν with "out of acceptance" leptons contribute, and look the same as Z→νν

- Background was determined using control regions+MC, and subtracted
- Perhaps these W's should be included as part of the "signal" definition?
  - ➢ This leaves the data uncontaminated and as close to "what we see" as possible.
  - ➢ Removes dependence on control regions and MC extrapolation between regions
- But be careful of fiducial phase-space definitions: e.g. how are out-of-acceptance leptons treated in the missing p_T definition?

What is a final-state neutrino?*

Invisible in the detector and existence inferred by p_T^{miss}

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = \underbrace{\underbrace{-\sum_{\substack{\text{selected}\\\text{electrons}}} \mathbf{p}_{\mathrm{T}}^{e}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}} \underbrace{-\sum_{\substack{\text{accepted}\\\text{photons}}} \mathbf{p}_{\mathrm{T}}^{\gamma}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}} \underbrace{-\sum_{\substack{\text{accepted}\\\tau\text{-leptons}}} \mathbf{p}_{\mathrm{T}}^{\tau_{\mathrm{had}}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\tau_{\mathrm{had}}}} \underbrace{-\sum_{\substack{\text{selected}\\\text{muons}}} \mathbf{p}_{\mathrm{T}}^{\mu}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}} \underbrace{-\sum_{\substack{\text{accepted}\\\text{jets}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,jet}}}}_{\text{hard term}} \underbrace{\underbrace{-\sum_{\substack{\text{unused}\\\text{tracks}}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}}}_{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,soft}}}}_{\text{soft term}}$$

Sum of all the visible particles within detector acceptance, which is closer to what we measure but can be a bit complicated. e.g.

- What p_T of hadrons are we actually sensitive to?
- What angle and p_T of leptons can we identify?

*Neutrinos are indistinguishable from BSM invisible particles, which is one of the points of this measurement…

Similar final-states example: Four-lepton production

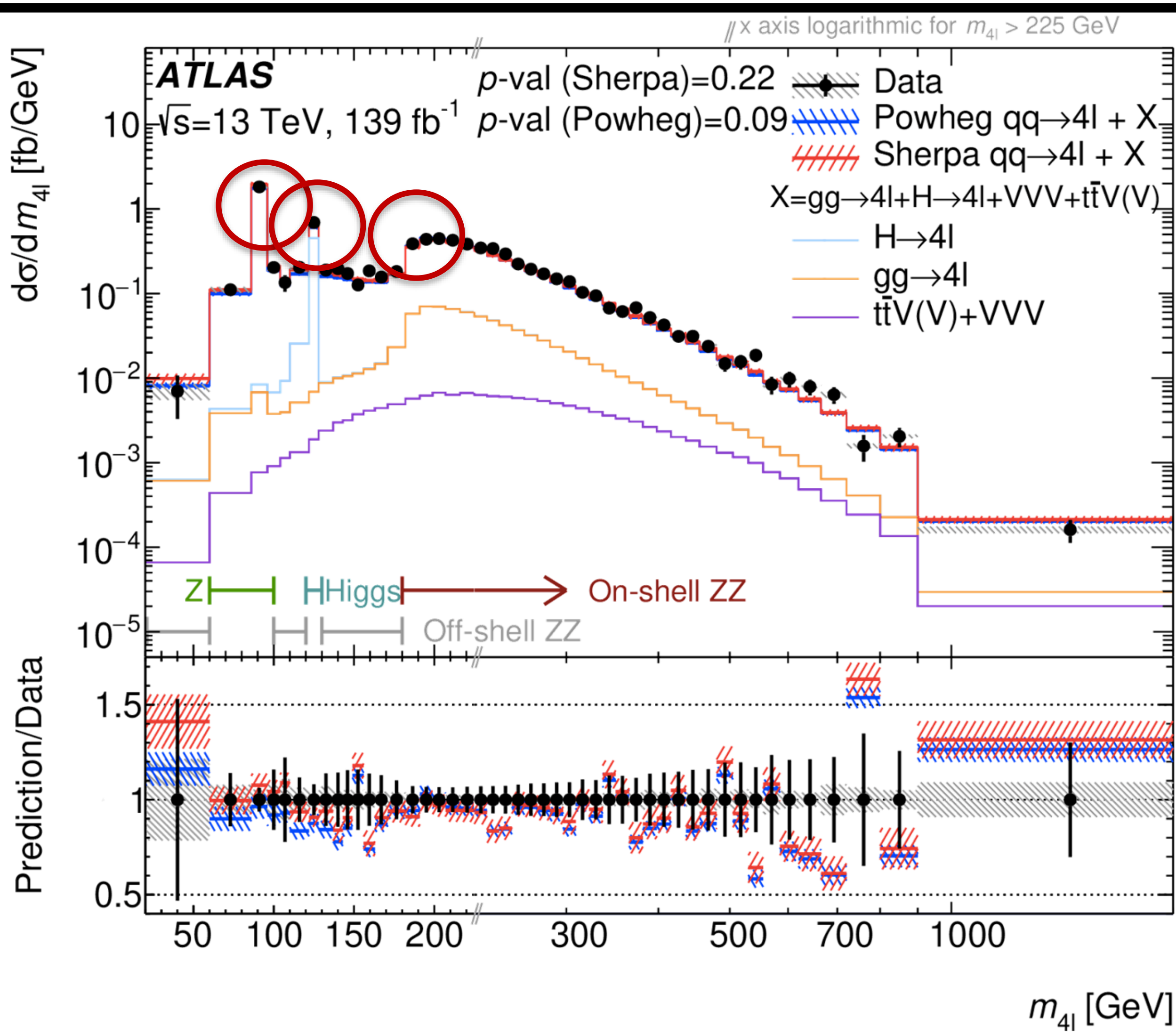
All events with:

e^+e^-, $\mu^+\mu^-$

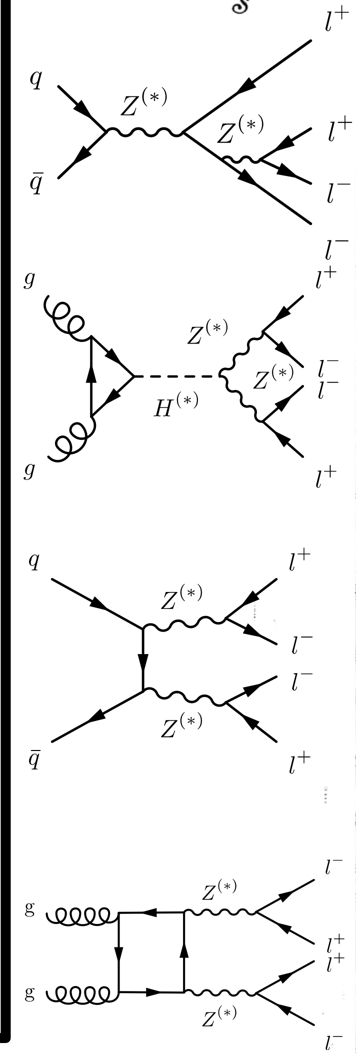
e^+e^-, e^+e^-

or

$\mu^+\mu^-$, $\mu^+\mu^-$



arXiv:2103.01918



Jet Substructure and Jet Grooming

"Boost"

Some new and different things about the LHC

- Unprecedented amounts of pile-up

Jets will be contaminated not just by underlying event but by particles from other pp interactions

- Unprecedented collision energy!

Electroweak-scale particles can be produced with high Lorentz boost, and lots of QCD radiation

Many Electroweak-scale particles/jets can be produced in the same event, and lots of QCD radiation

- Jets are everywhere

Ok that's not so new… but it's true

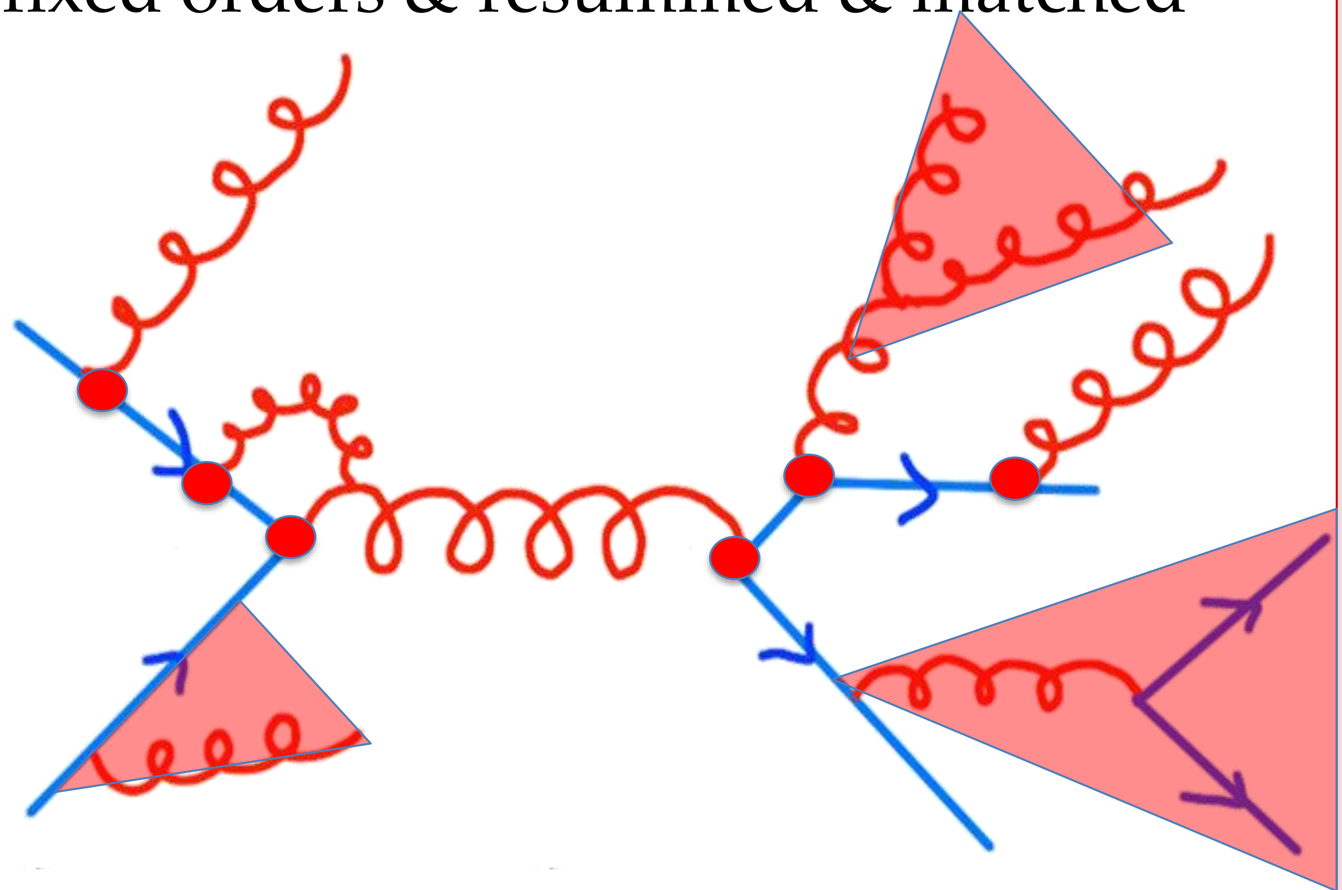
And may contain boosted Z, W, H, t etc decays

To look at it another way…

- Final stage of jet structure is "soft" non-perturbative QCD.
  - Formation of hadrons from gluons, 100 MeV energy scales (Λ_{QCD})
- Vast phase space between quark-gluon scatter (100s of GeV, few TeV) and Λ_{QCD}
- Most of jet substructure can potentially be analysed and understood perturbatively
  - Parton showers, ME matching →

To the TeV scale and beyond…

Precision theory, exclusive calculations, fixed orders & resummed & matched



Vertices → α_s, $\mathcal{O}(0.1)$

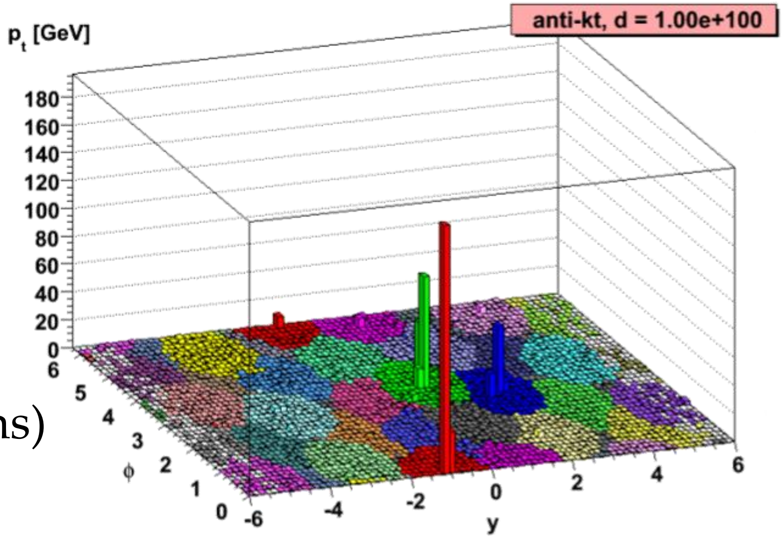
Soft or collinear kinematics → *log*(scales) $\mathcal{O}(10)$

α_s*log* terms $\mathcal{O}(1)$, must be resummed (exponentiated)

Still need fixed-order α_s outside these enhanced regions

- Jets are one key area in the data/theory comparison business
- Jets are not just less-well-measured leptons or "smeared" partons.
  - Hard radiation interference at amplitude level
  - Matching at high scales with Matrix element
  - Matching at low scales with parton densities and hadronisation model
  - potentially useful information in the internal jet structure, and in particle/energy flow between the jets
- Jets have **no existence independent of the algorithm**
  - even if the "algorithm" = event display + physicist

Jet finder(s) → Jet definers



$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2},$$

$$d_{iB} = k_{ti}^{2p},$$

- Form a list
  - (of particles, tracks, clusters or partons)
- Merge the smallest pair
- Iterate
- Anti-k_t : p = -1 (default. Use R=0.4, 0.6)
  - Stable, "circular" jets built around highest p_T regions
- Also used:
  - Cambridge/Aachen: p = 0. Based only on angular information
  - k_t : p = 1. Merging scales are physically meaningful

Cacciari, Salam, Soyez JHEP 0804:063,2008

Dokshitzer, Leder, Morretti, Webber (JHEP 08 (1997) 01; Wobisch and Wengler hep-ph/9907280

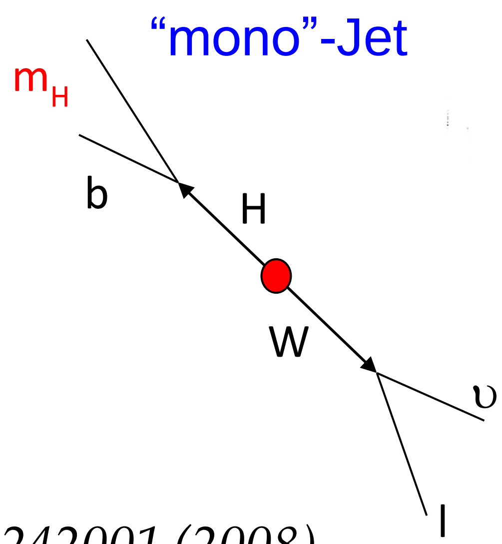
Catani et al Phys Lett B269 (1991); Nucl. Phys. B406 (1993); Ellis and Soper Phys Rev D48 (1993).

With well-defined algorithms, can look at, and use, the jet substructure

- An example from just before Run 1
  - Higgs production associated with a W or Z, Higgs decay to bb
  - Thought to be very difficult/impossible (TDRs)
  - Still, best chance for H→ bb
  - What can we do…?

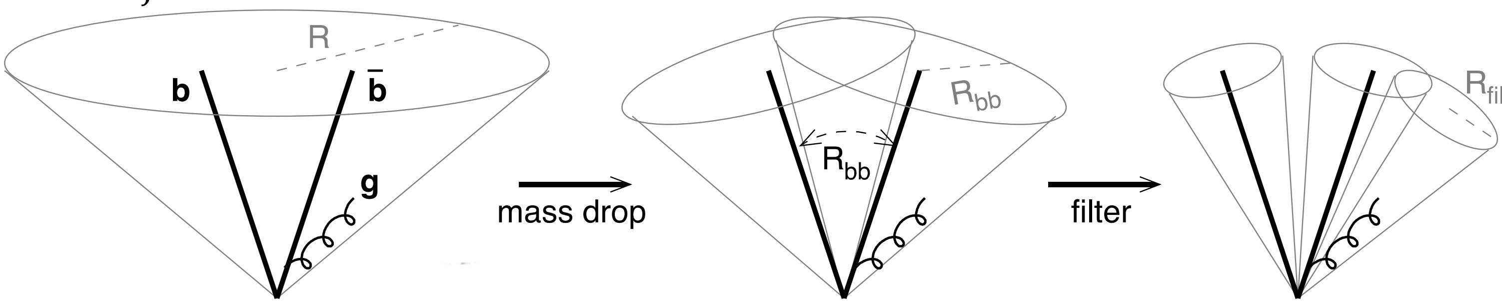
High p_T Higgs and Vector Boson

- By requiring that the Higgs and Vector Boson have a high transverse momentum, we lose a factor of ~20 in cross section
  - However, much of this would have failed other analysis cuts anyway
  - Background cross sections fall by a bigger factor (typically t-channel not s-channel)
- W/Z and H are all central
  - Better b-tagging, better jet resolution
- W/Z and H decay products collimated
  - Simpler topology, fewer combinatorials
  - Difficult for tops to fake this
- Z ➔ neutrinos becomes visible
  - High missing E_T
- *JMB, Davison, Rubin, Salam, Phys. Rev. Lett. 100, 242001 (2008)*



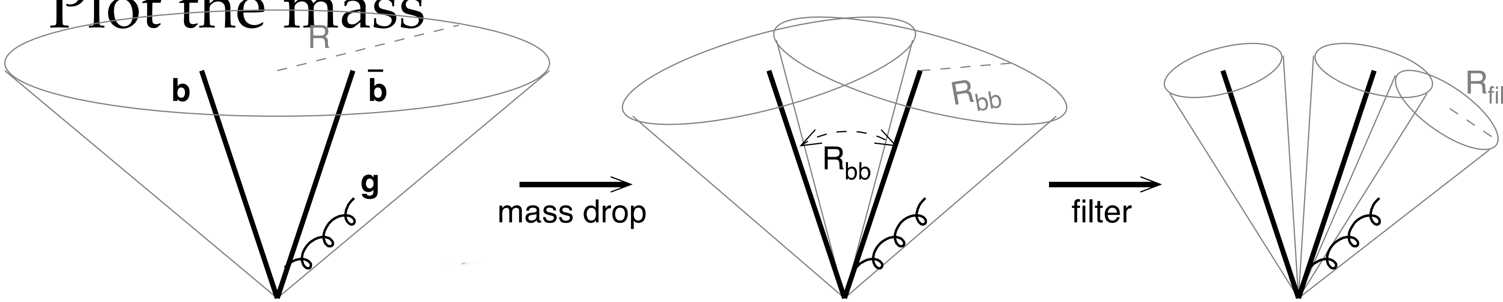
Sub-jet analysis

1. Start with Higgs candidate jet (highest p_T jet in acceptance) with mass m)
2. Undo last stage of clustering (reduce radius to R_{12})
   $J \rightarrow J_1, J_2$
3. If $\max(m_1, m_2) < 2m/3$
   Call this a "mass drop". This fixes the optimal radius for reconstructing the Higgs decay. Keep the jet J and call it the Higgs candidate.
   Else, go back to 2
4. Require $Y_{12} > 0.09$
   Dimensionless rejection of asymmetric QCD splitting
   Else reject the event
5. Require J_1, J_2 to each contain a b-tag
   Else reject the event

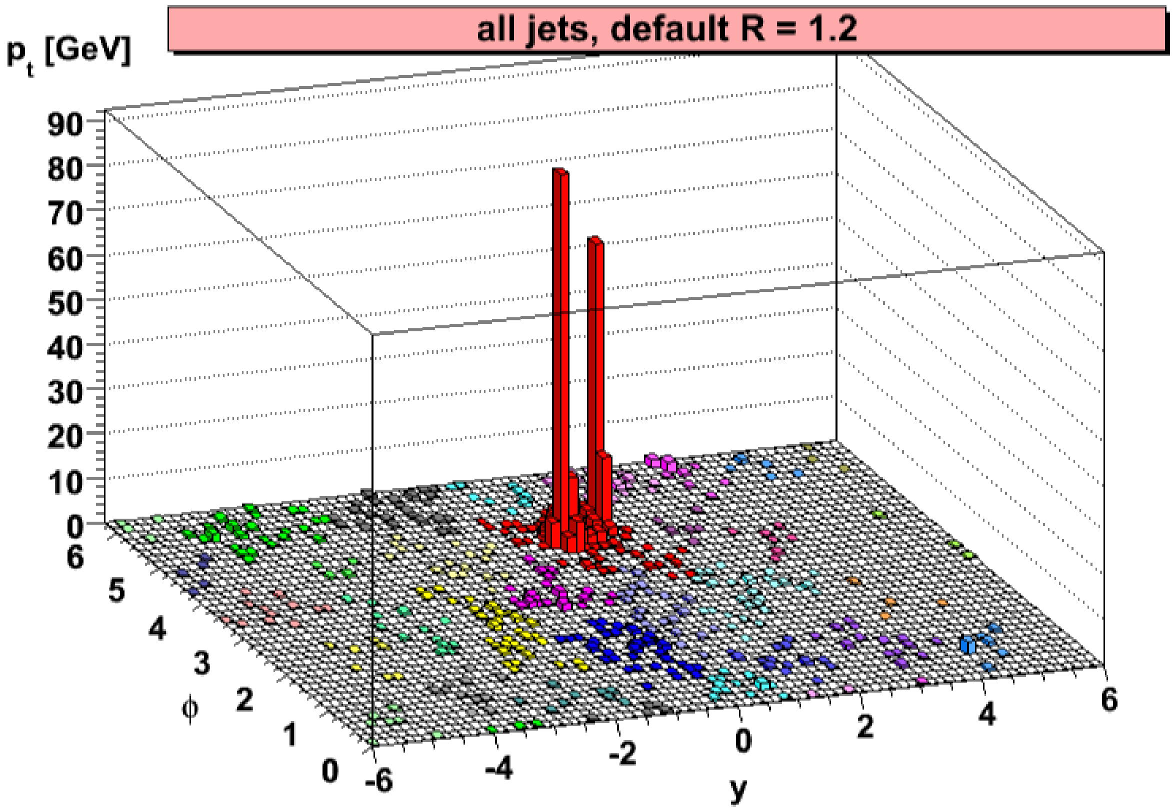


Sub-jet analysis

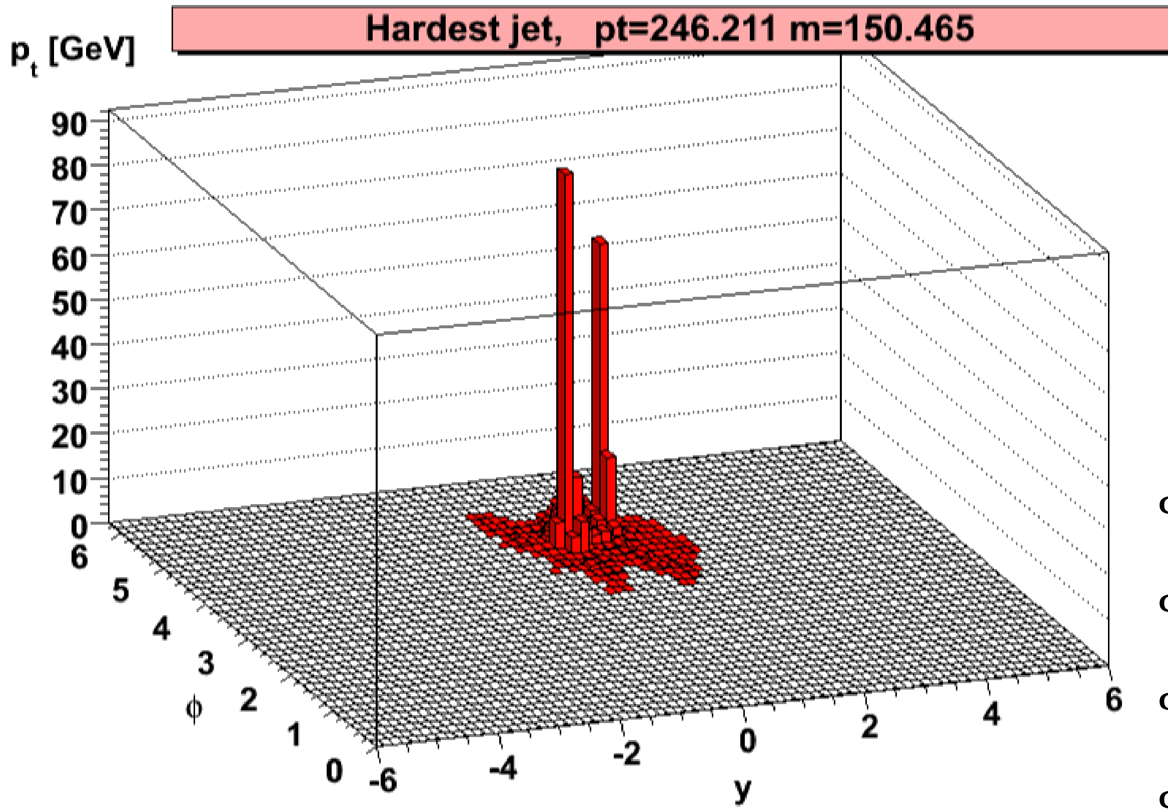
6. Define R_{filt} = min(0.3, R_{bb}/2)
   - Make use event-by-event of the known Higgs decay radius
   - Angular ordering means this is the characteristic radius of QCD radiation from Higgs products
   - Stuff outside of this is likely to be underlying event and pileup.
7. Recluster, with Cambridge/Aachen, R = R_{filt}
8. Take the 3 hardest subjets and combine to be the Higgs
   - b, anti-b and leading order final state gluon radiation
9. Plot the mass

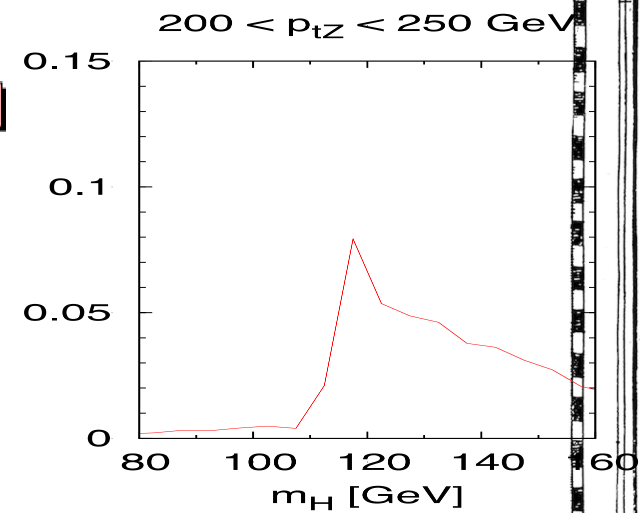


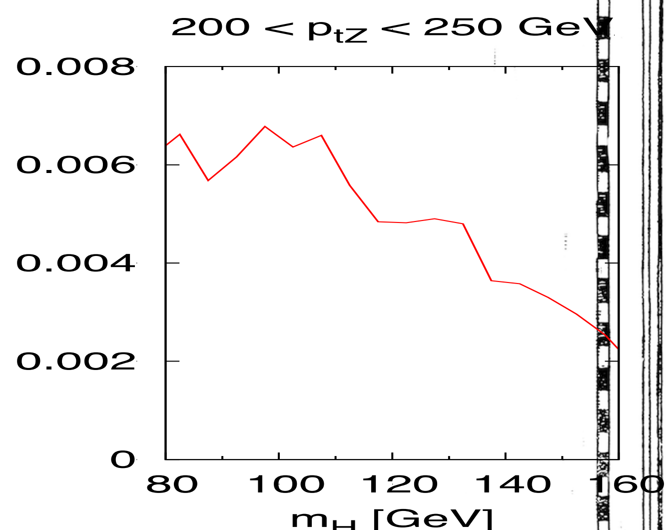
Improved subjet analysis



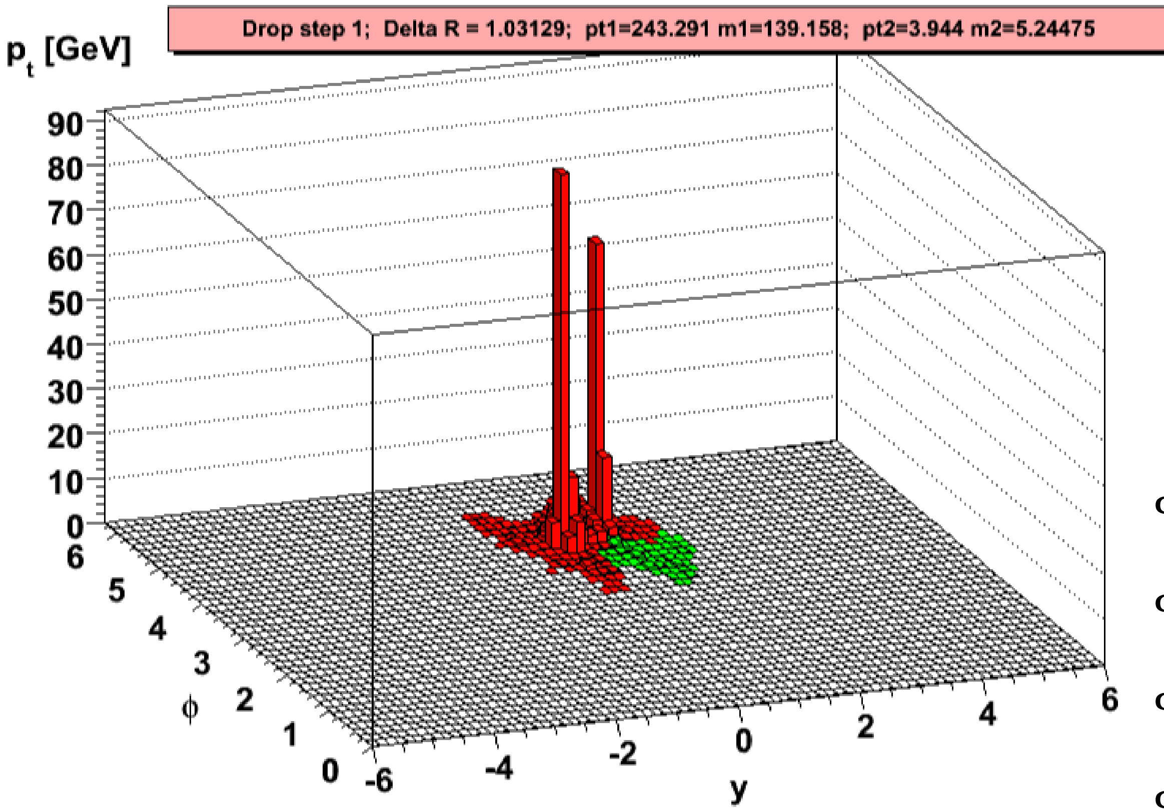
Improved subjet analysis

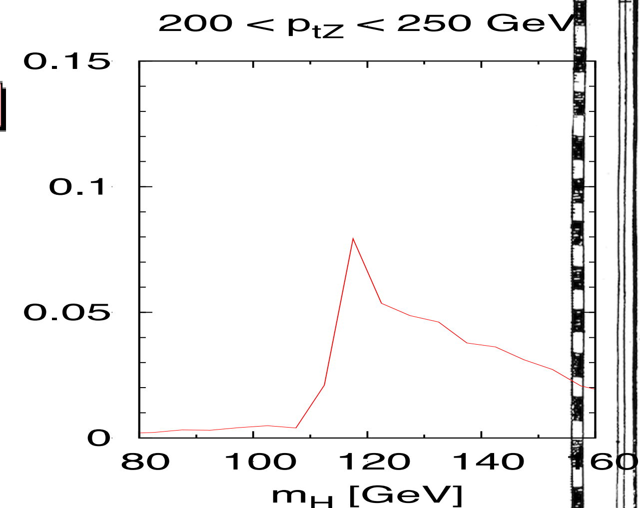


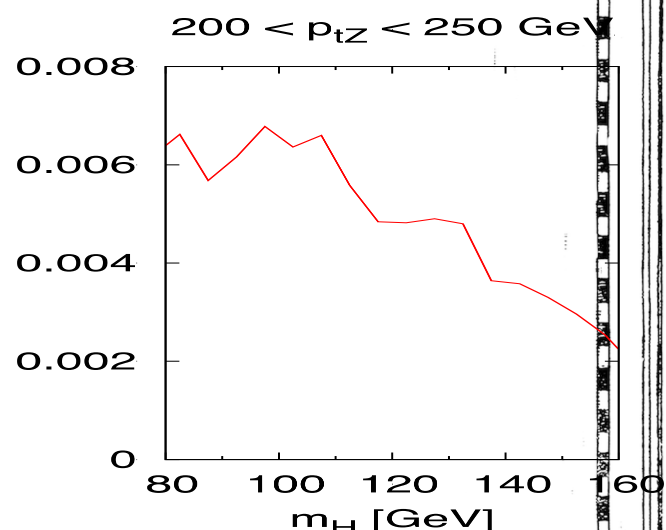




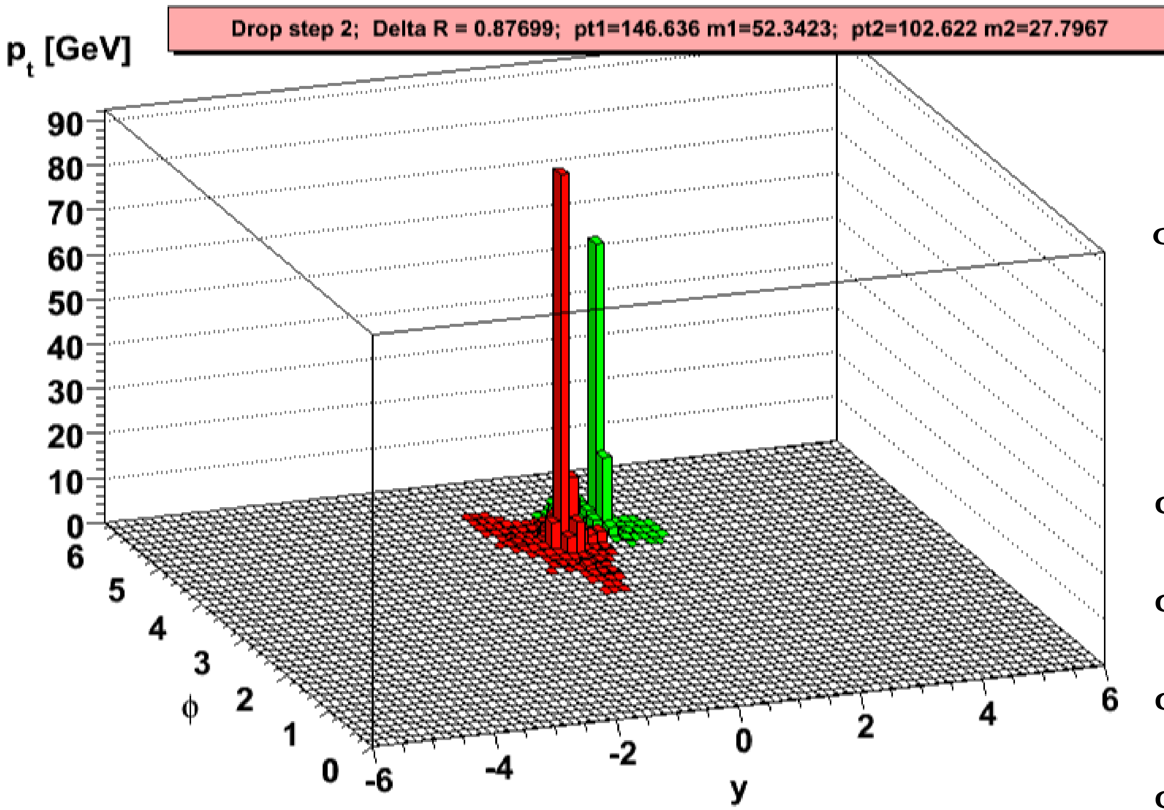
Improved subjet analysis

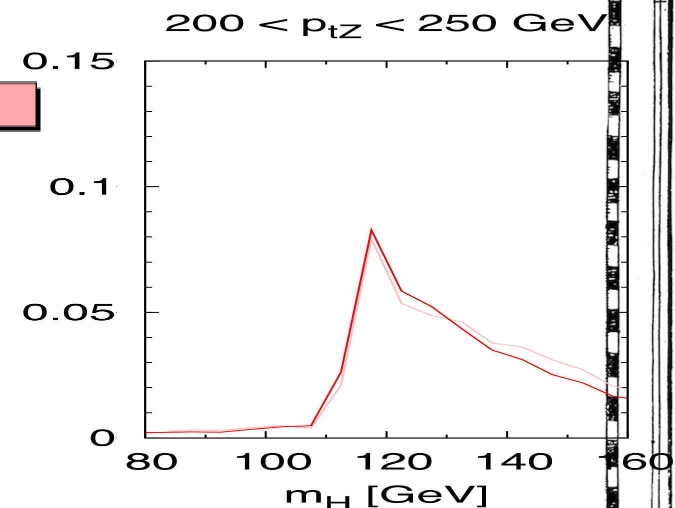


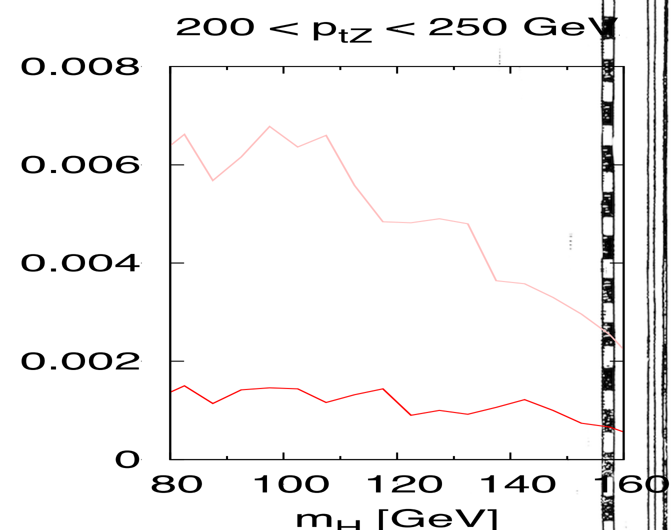




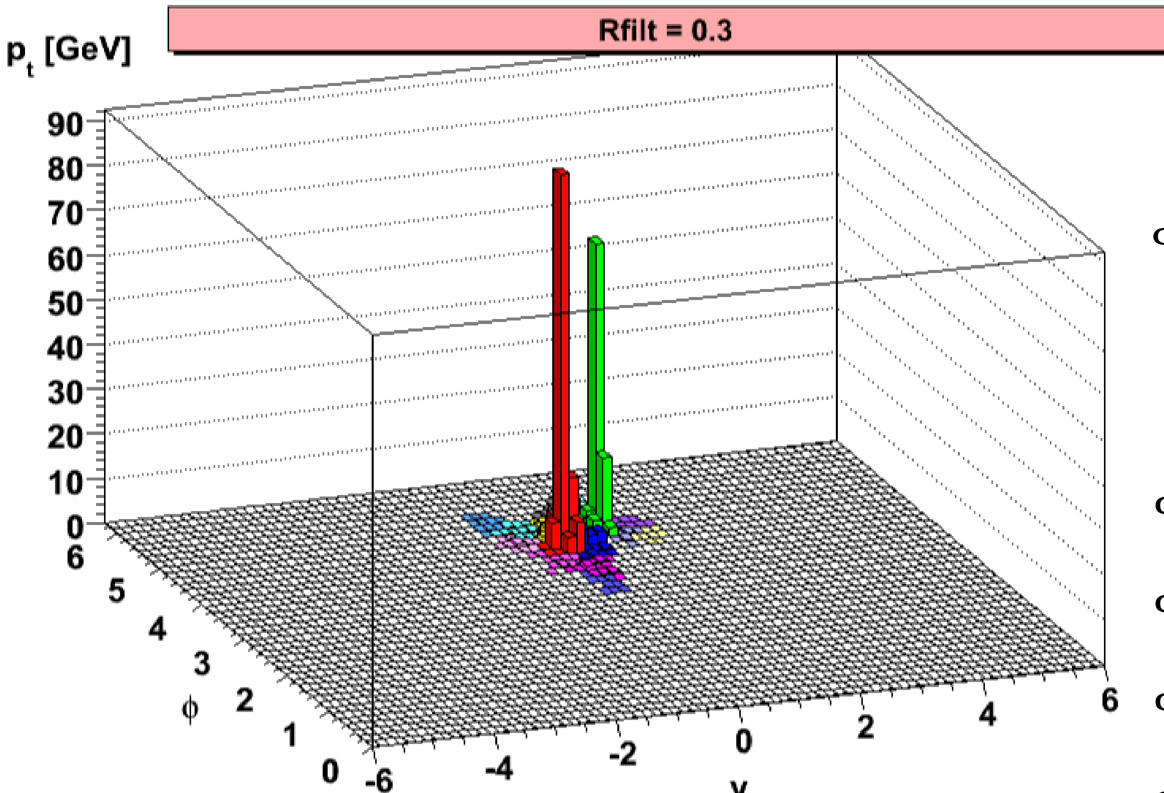
Improved subjet analysis

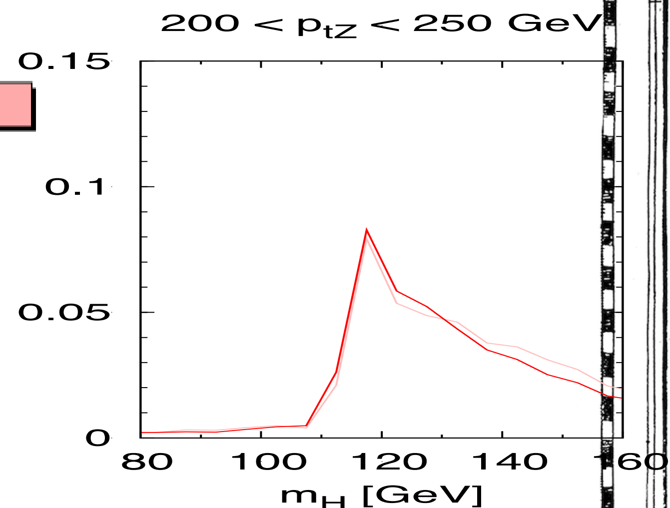


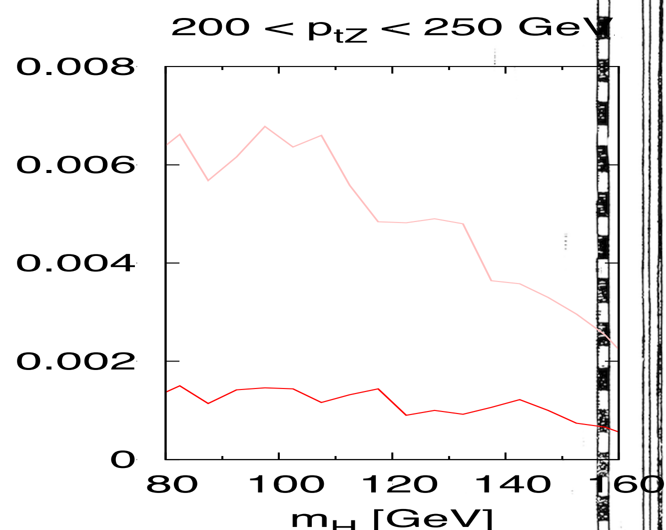




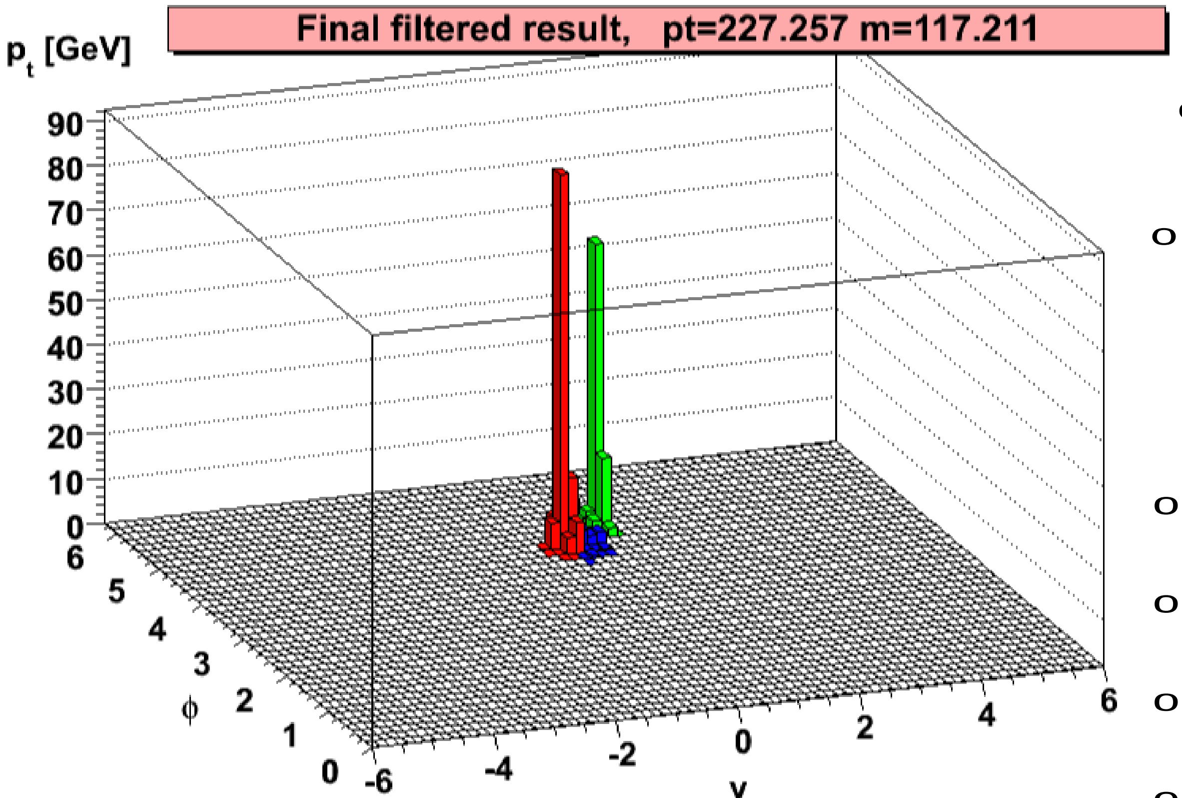
Improved subjet analysis

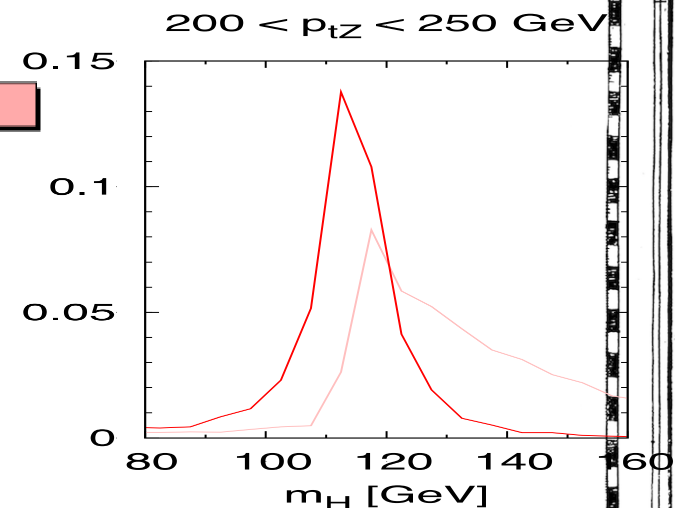


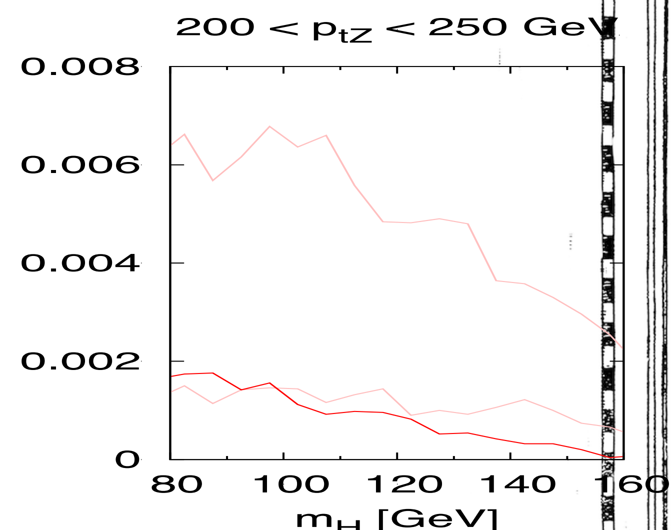




Improved subjet analysis







Roughly, two goals in jet substructure

- Grooming: Remove underlying event and pile-up
  - improved mass resolution, other properties reflect hard physics more closely
  - In a way which is perturbatively calculable
- Tagging: is this a q/g initiated jet or a heavy particle decay?
  - Background reduction, efficiency
  - NB interaction with B-tagging
  - Again, must be perturbatively calculable

Where are we now?

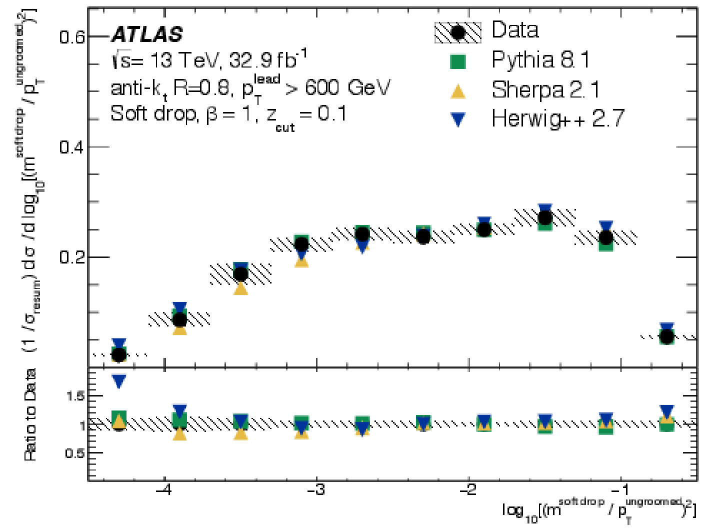
- "Soft drop" is the state-of-the art
  - Mass Drop Tagger (BDRS) → modified Mass Drop Tagger (follow p_T not mass). Dasgupta et al, https://arxiv.org/abs/1307.0007
  - mMDT tagging → jet grooming Larkoski et al, https://arxiv.org/abs/1402.2657

"We introduce a new jet substructure technique called "soft drop declustering", which recursively removes soft wide-angle radiation from a jet. The soft drop algorithm depends on two parameters--a soft threshold **zcut** *and an angular exponent* **β***--with the* **β=0** *limit corresponding roughly to the (modified) mass drop procedure."*

$$\frac{\min(p_{T,j_1}, p_{T,j_2})}{p_{T,j_1} + p_{T,j_2}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R}\right)^{\beta}$$

In use…

- ATLAS https://arxiv.org/abs/1711.08341



References

[1] There are nearly 100 public search results from ATLAS and CMS as well as an even larger number of phenomenological proposals to use jet substructure to enhance various searches. See e.g. Ref. [2, 3] for the first phenomenological proposals and Ref [4–7] for representative ATLAS and CMS performance studies.

[2] J. M. Butterworth, B. E. Cox and J. R. Forshaw, *WW scattering at the CERN LHC*, Phys. Rev. D **D65** (2002) 096014, arXiv: hep-ph/0201098.

[3] J. M. Butterworth et al., *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. **100** (2008) 242001, arXiv: 0802.2470 [hep-ph].

[4] ATLAS Collaboration, *Identification of high transverse momentum top quarks in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, JHEP **06** (2016) 093, arXiv: 1603.03127 [hep-ex].

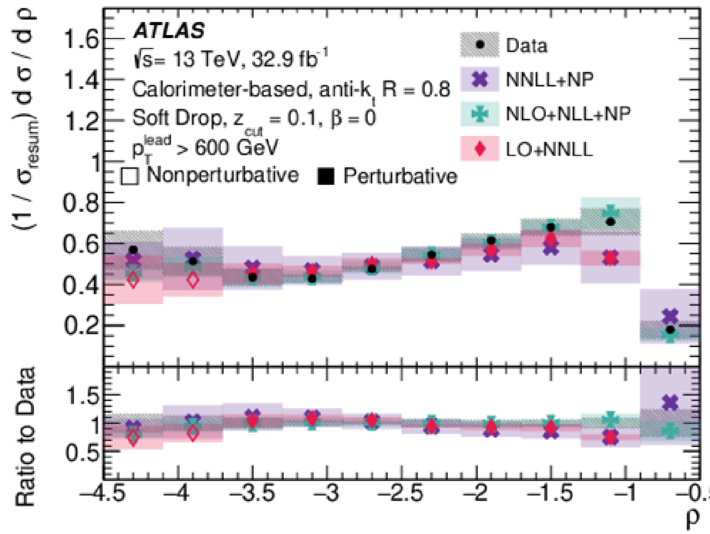
[5] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s}$ = 8 TeV*, Eur. Phys. J. C **76** (2016) 154, arXiv: 1510.05821 [hep-ex].

[6] CMS Collaboration, *Identification techniques for highly boosted W bosons that decay into hadrons*, JHEP **12** (2014) 017, arXiv: 1410.4227 [hep-ex].

[7] CMS Collaboration, *Boosted Top Jet Tagging at CMS*, CMS-PAS-JME-13-007 (2014), URL: https://cds.cern.ch/record/1647419.

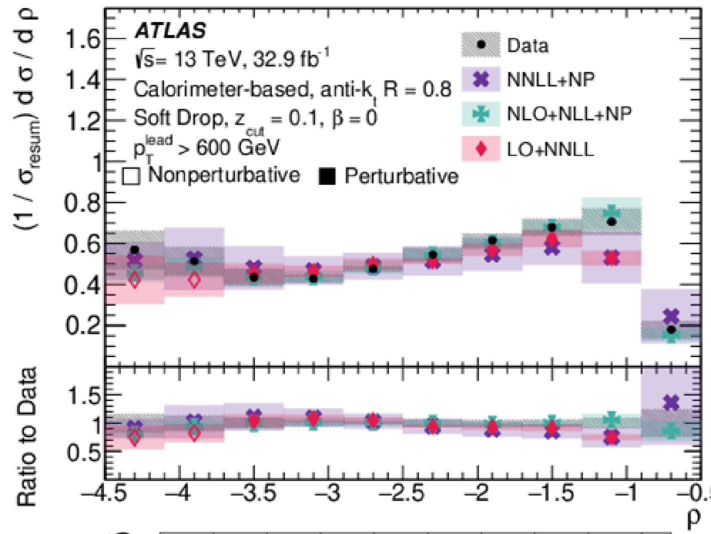
In use…

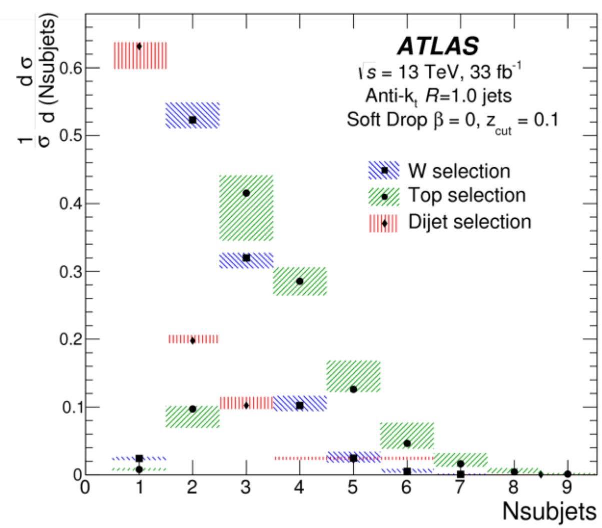
- ATLAS https://arxiv.org/abs/1711.08341
- https://arxiv.org/abs/1912.09837



In use…

- ATLAS https://arxiv.org/abs/1711.08341
- https://arxiv.org/abs/1912.09837
- https://arxiv.org/abs/1903.02942





In summary…

- Jet substructure and jet grooming (see also n-subjettiness etc) are ways of *defining new observables* which aim to be:
  - Calculable to controlled accuracy in the SM (QCD & EW)
  - More precise and more sensitive to the interesting hard physics (H, W, Z, t, BSM, novel QCD effects)
  - Less sensitive to soft/non-perturbative physics (hadronization, soft and collinear radiation)
  - Less sensitive to experimental noise (pile-up, mainly)
- Increasing role of machine learning (ML4JETS)

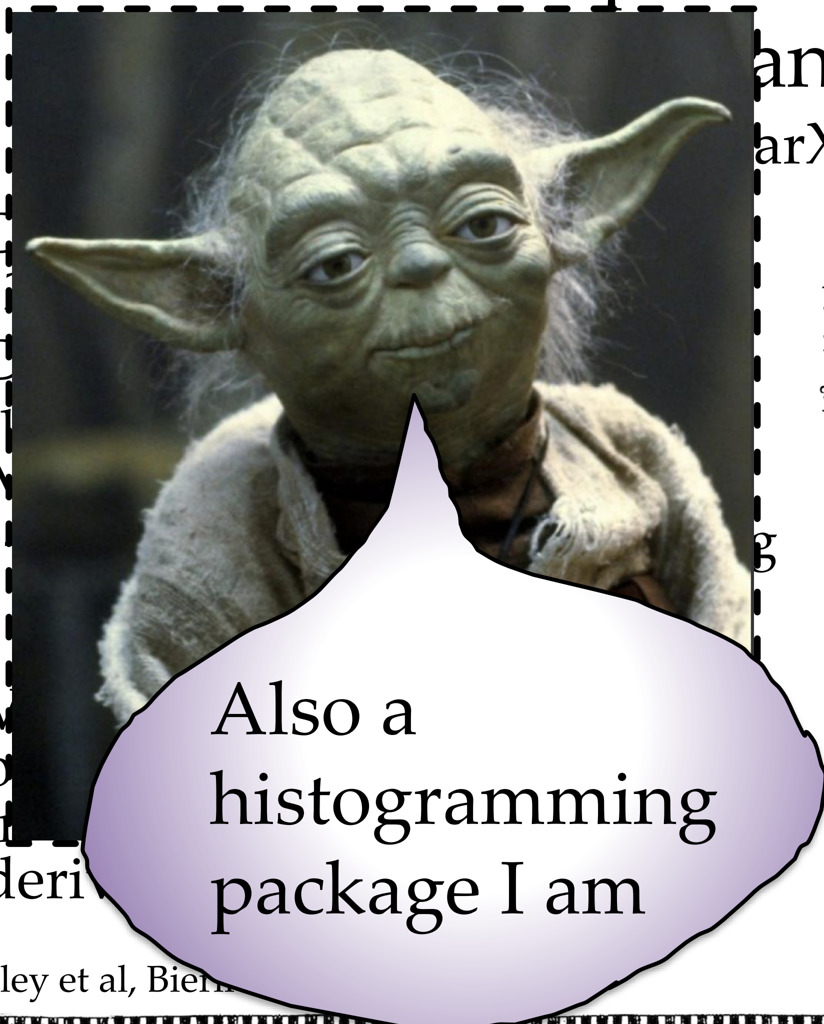
Finally… the Power: 2

- The role of Monte Carlo event generators as the best predictions we have for the Standard Model and Beyond
  - More details on how this is done this later in the school
  - Here I will just mention a couple of auxiliary tools which help

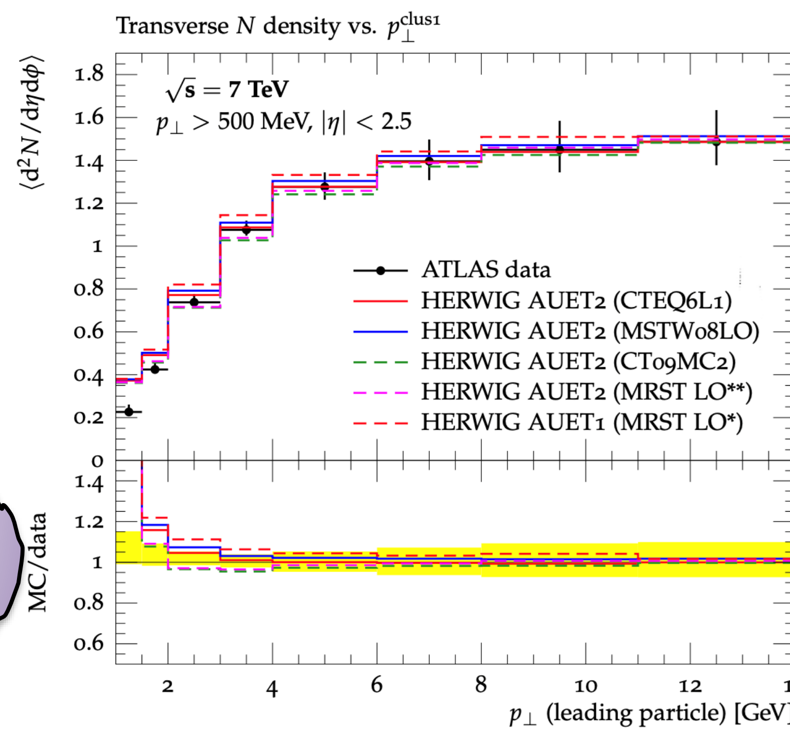
Introducing *Rivet* *and YODA*

“Robust Independent Validation of ... and Theory”

arXiv:1912.05451



- D
  (
- D
  t
  M
  g
- V
  o
  i
  deri



From ATL-PHYS-PUB-2011-008

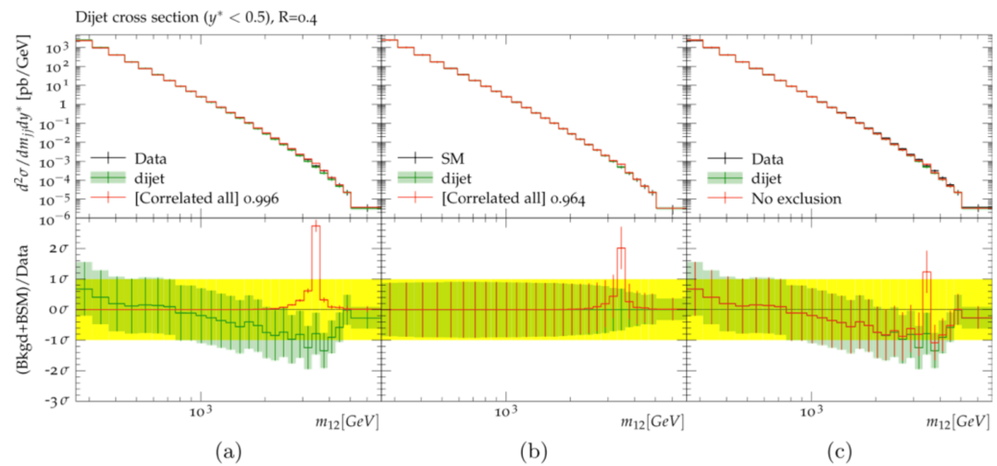
Buckley et al, Bier

Introducing *Contur*

"Constraints On New Theories Using Rivet"

arXiv:1605.05296, arXiv:2102.04377

- Extend the power of Rivet beyond the Standard Model
- Signal-injection of final-state particles from Beyond-the-SM physics events on to the measured cross sections in Rivet

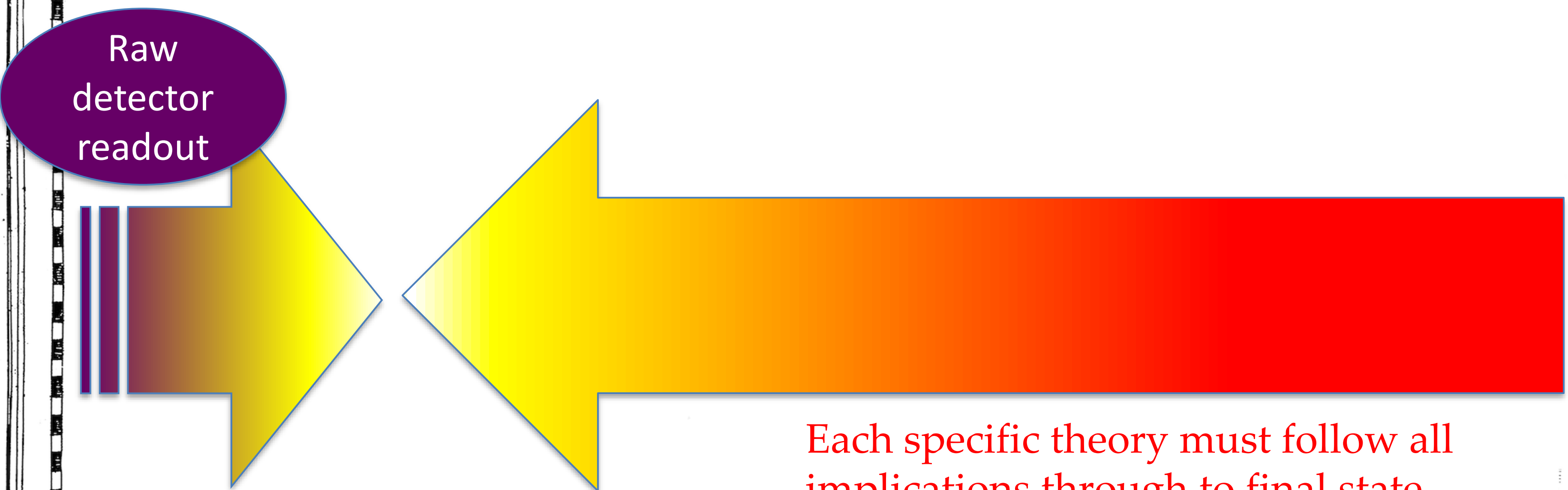


From Altakach, JMB, Ježo, Klasen, Schienbein arXiv:2111.15406

- Increasingly precise measurements and calculations together extend the reach

JMB, Grellscheid, Krämer, Sarrazin, Yallup; Buckley et al. **Tutorials**

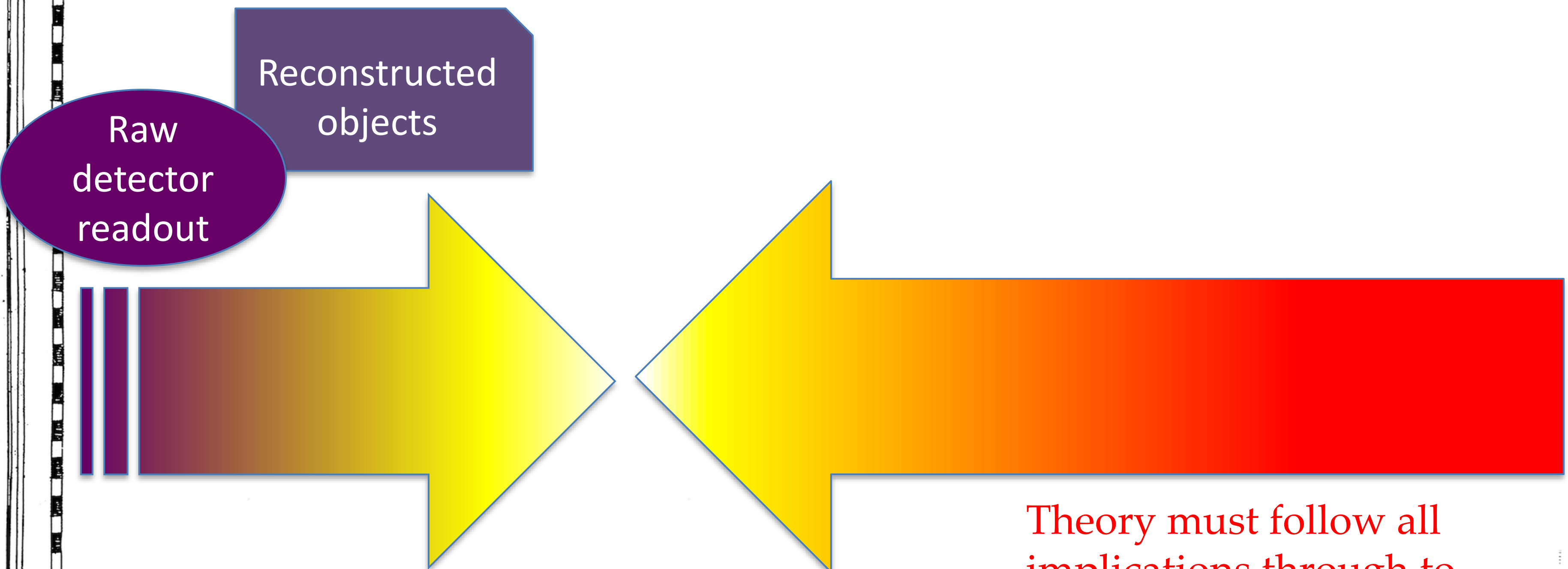
Where to compare nature to our ideas?



Zero model dependence.

Each specific theory must follow all implications through to final state particles and full detector simulation, including specific run conditions and time-dependent calibrations.

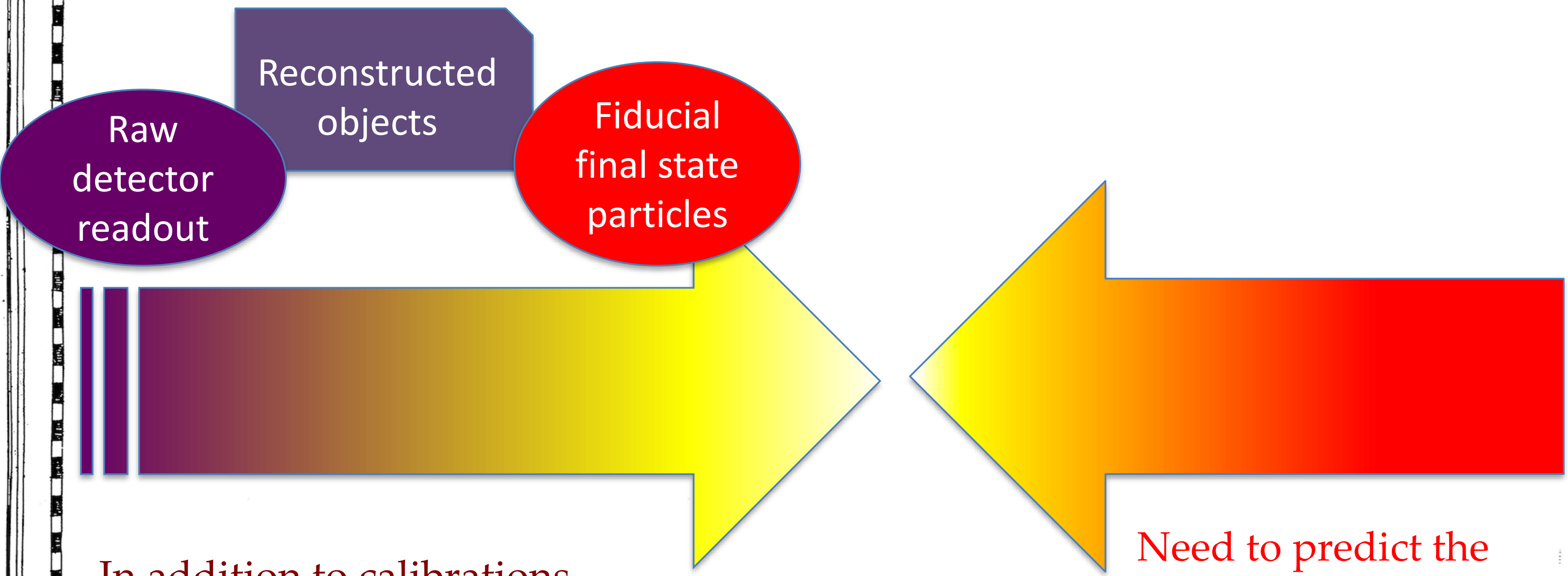
Where to compare nature to our ideas?



Calibrations applied which may have some dependence on models, but minimal and dependence can be controlled.

Theory must follow all implications through to final state and at least some parameterised approximation of detector resolution and efficiency.

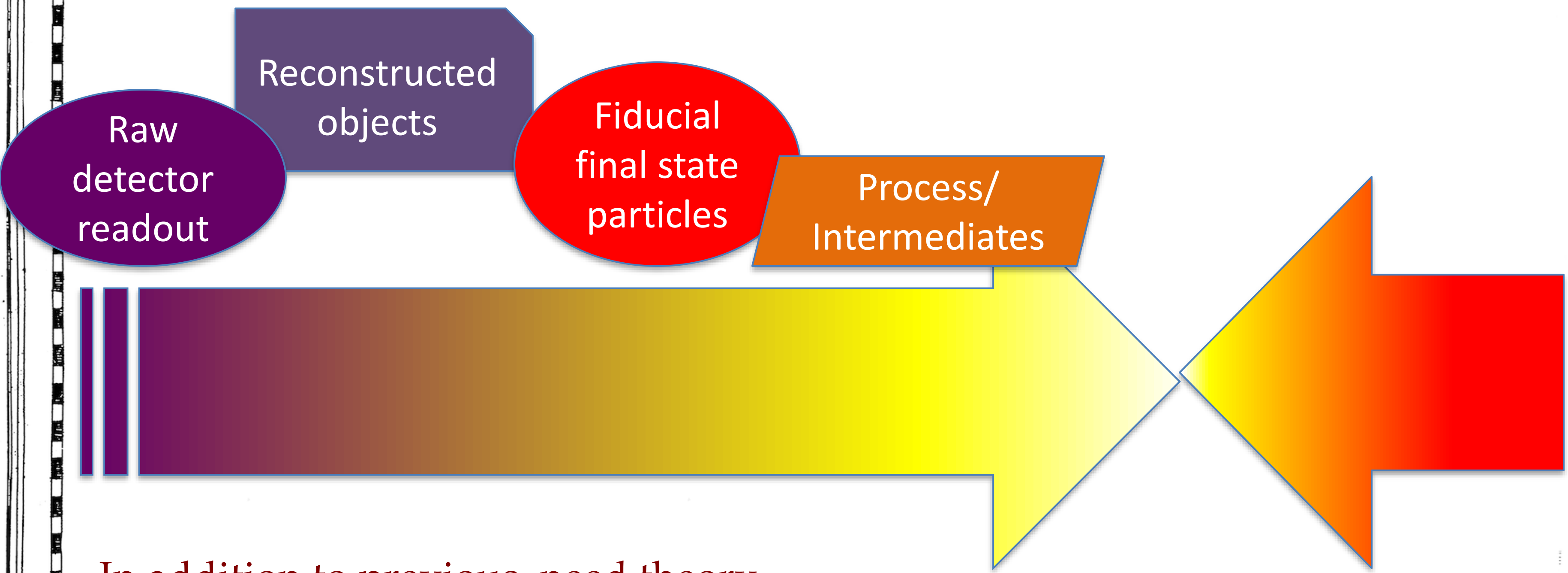
Where to compare nature to our ideas?



In addition to calibrations, need unfolding for resolution and efficiency, though uncertainties can generally be controlled.

Need to predict the exclusive final state.

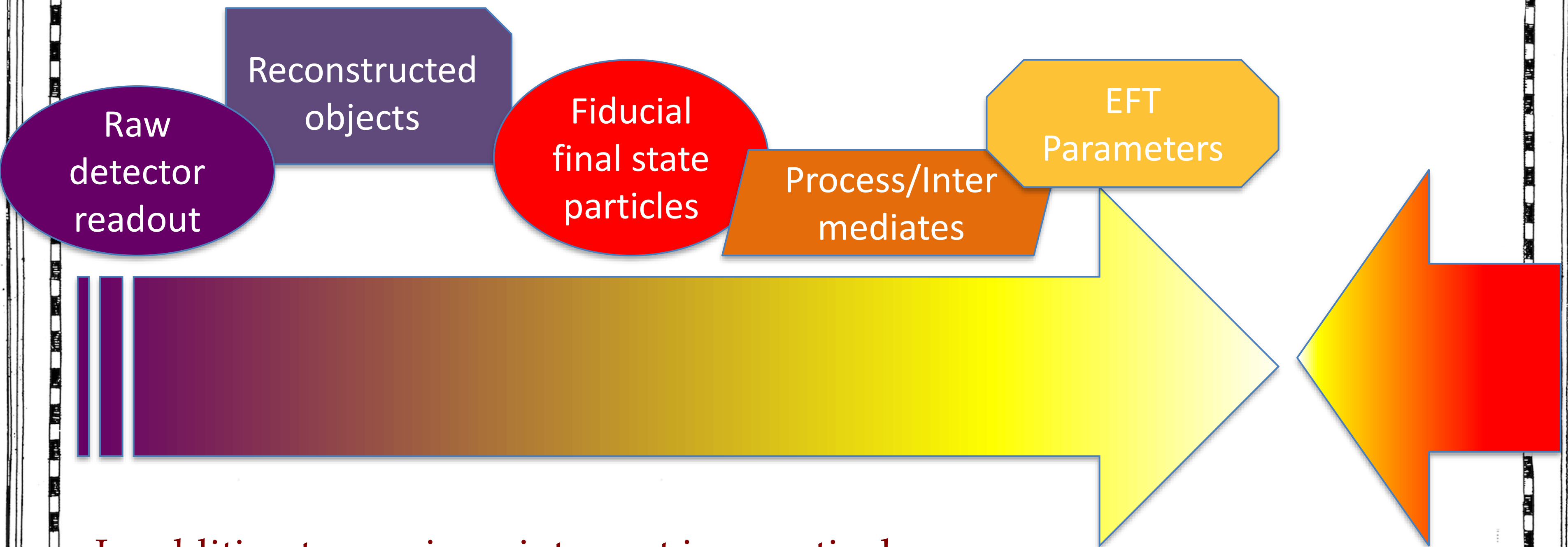
Where to compare nature to our ideas?



In addition to previous, need theory extrapolations into unobserved regions, theory background subtractions, and corrections for soft/long distance physics.

Can integrate over inclusive phase spaces and ignore soft/long distance physics.

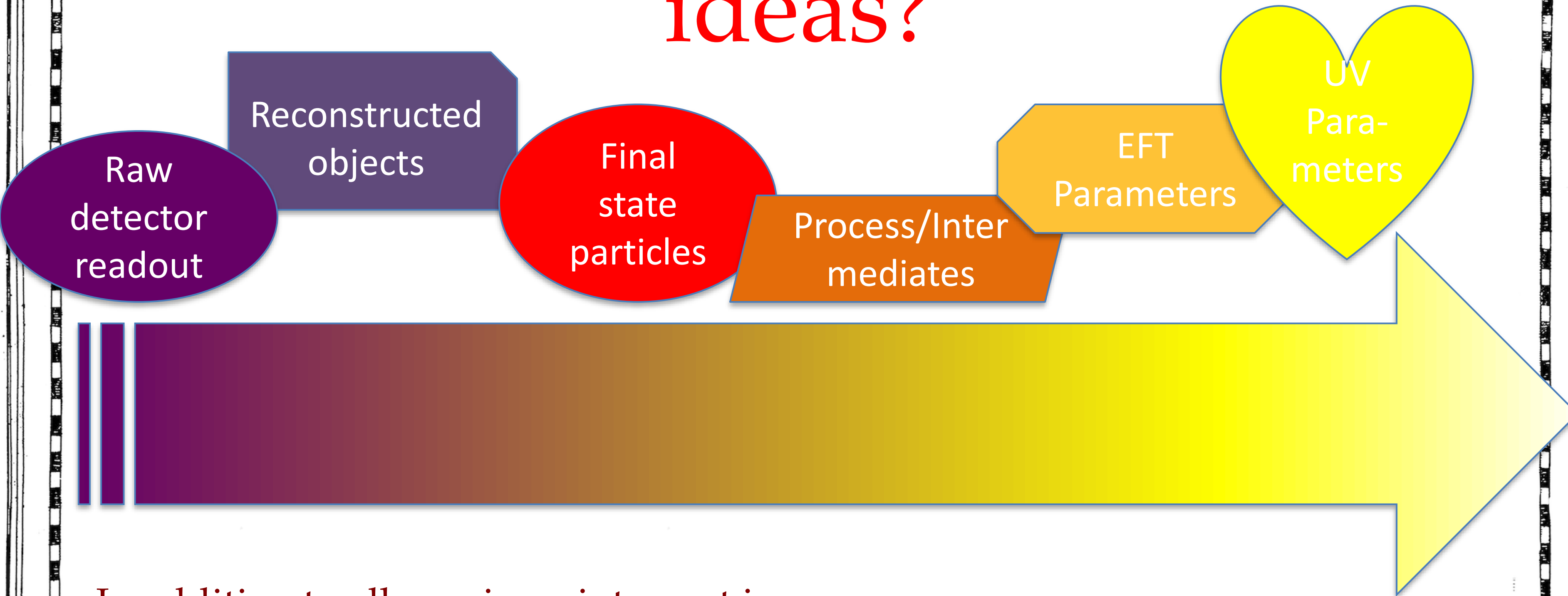
Where to compare nature to our ideas?



In addition to previous, interpret in a particular (simplified?) model.

Need to think about running from high energies, but not much else…

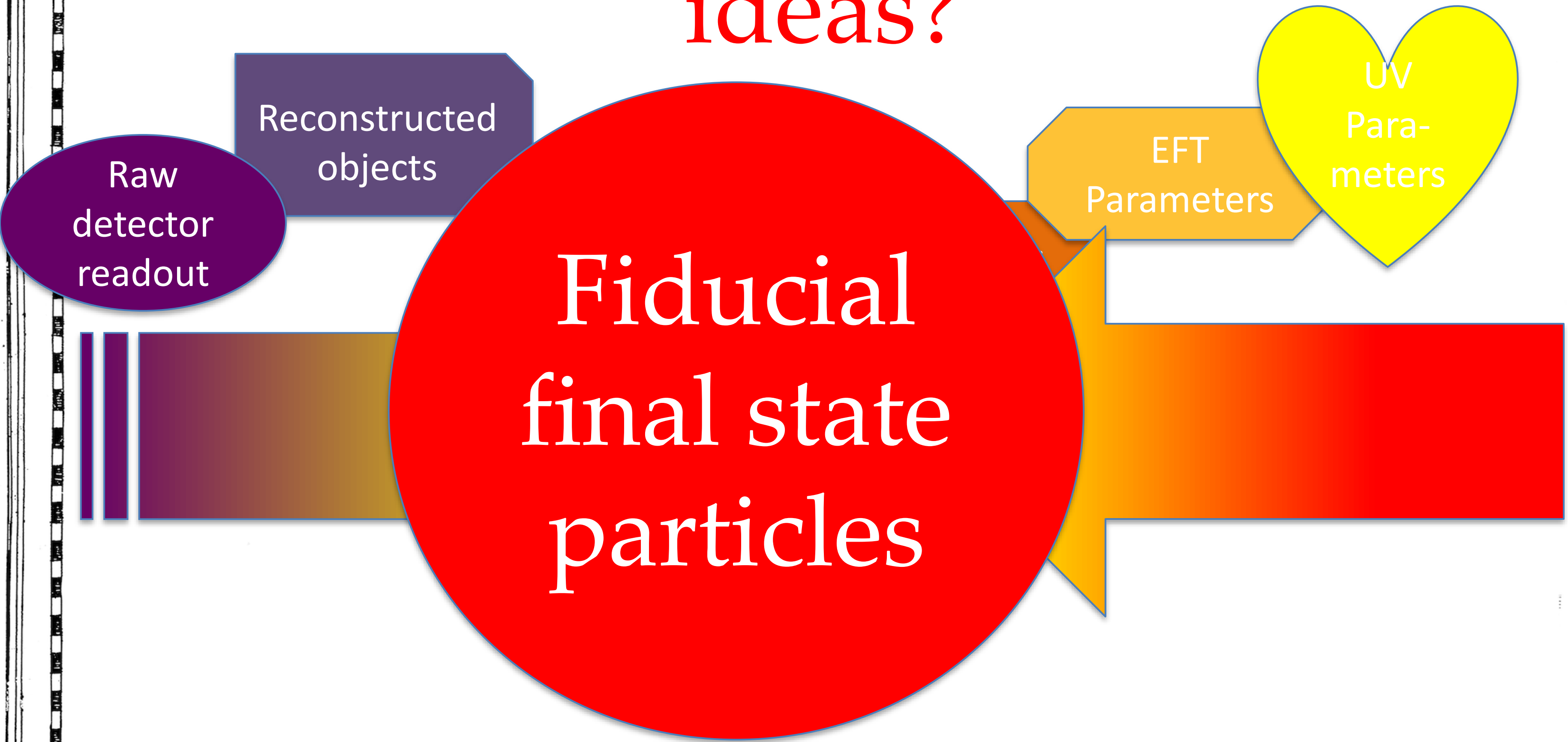
Where to compare nature to our ideas?



In addition to all previous, interpret in a particular UV complete model.

Have a good idea, then play golf.

Where to compare nature to our ideas?



The High Lumi LHC Era

- *Major experimental challenges*
- *High data rates*
    - *excellent trigger selection**
    - *fast, efficient analysis software**
    - *efficient/compact data storage*
- *High pile-up*
    - *tracking**
    - *vertexing*
    - *high granularity*
    - *jet "grooming" **

** Including Machine Learning and related techniques*

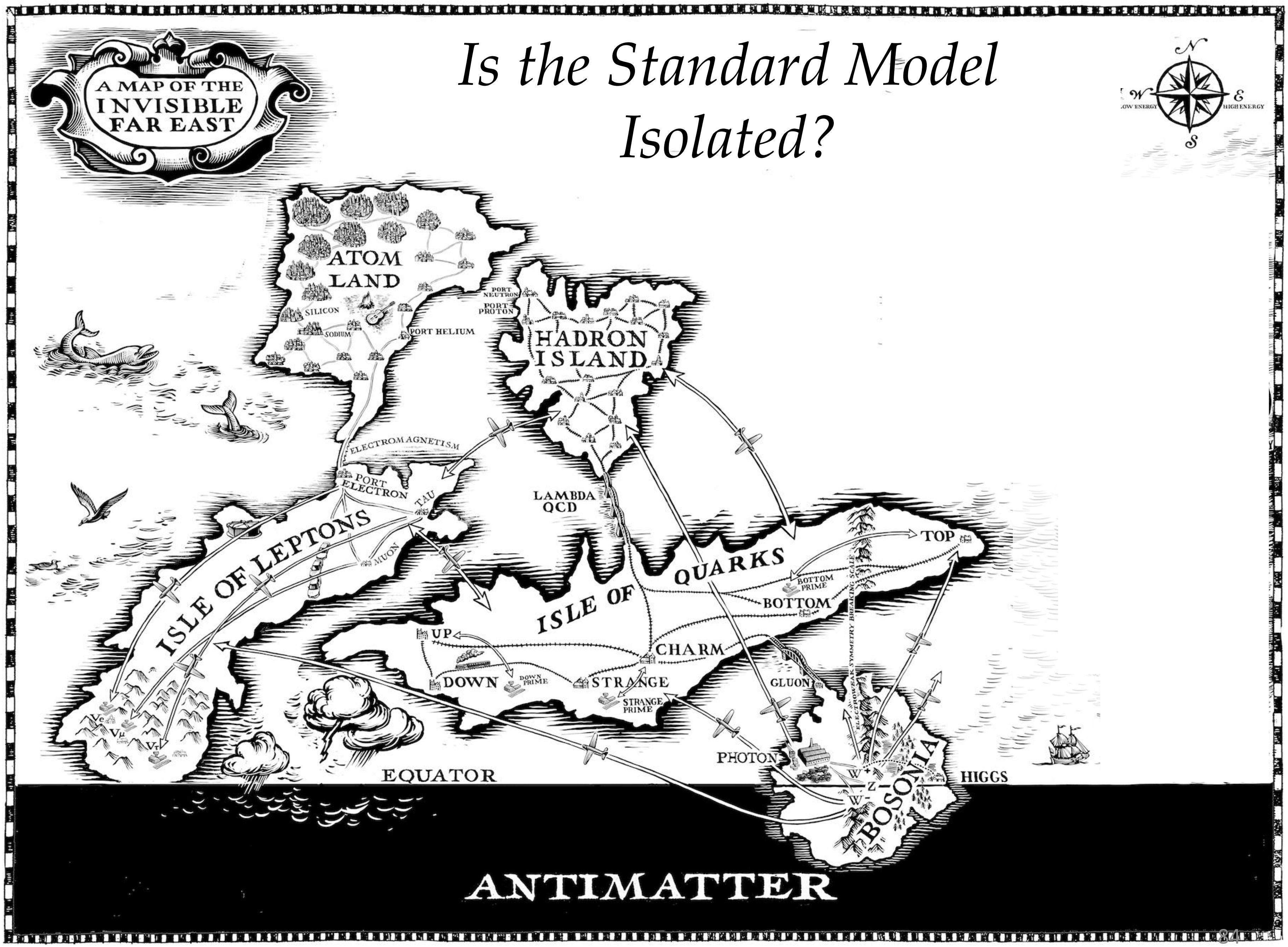
The High Lumi LHC Era

- *No agreed "favourite" extensions to the Standard Model*
  - *many ideas, connecting various anomalous phenomena*
  - *And we know the Standard Model is not a "Theory of Everything"*
- *Change of approach required*
  - *This is about exploration of new physics territory*
  - *No guarantee that Dark Matter, Supersymmetry, or indeed anything else beyond the Standard Model will be within reach*
- *Need precise, theory-independent measurements, and comparable calculations, in Standard Model & beyond. (As well as looking for "outliers".)*

The High Lumi LHC Era

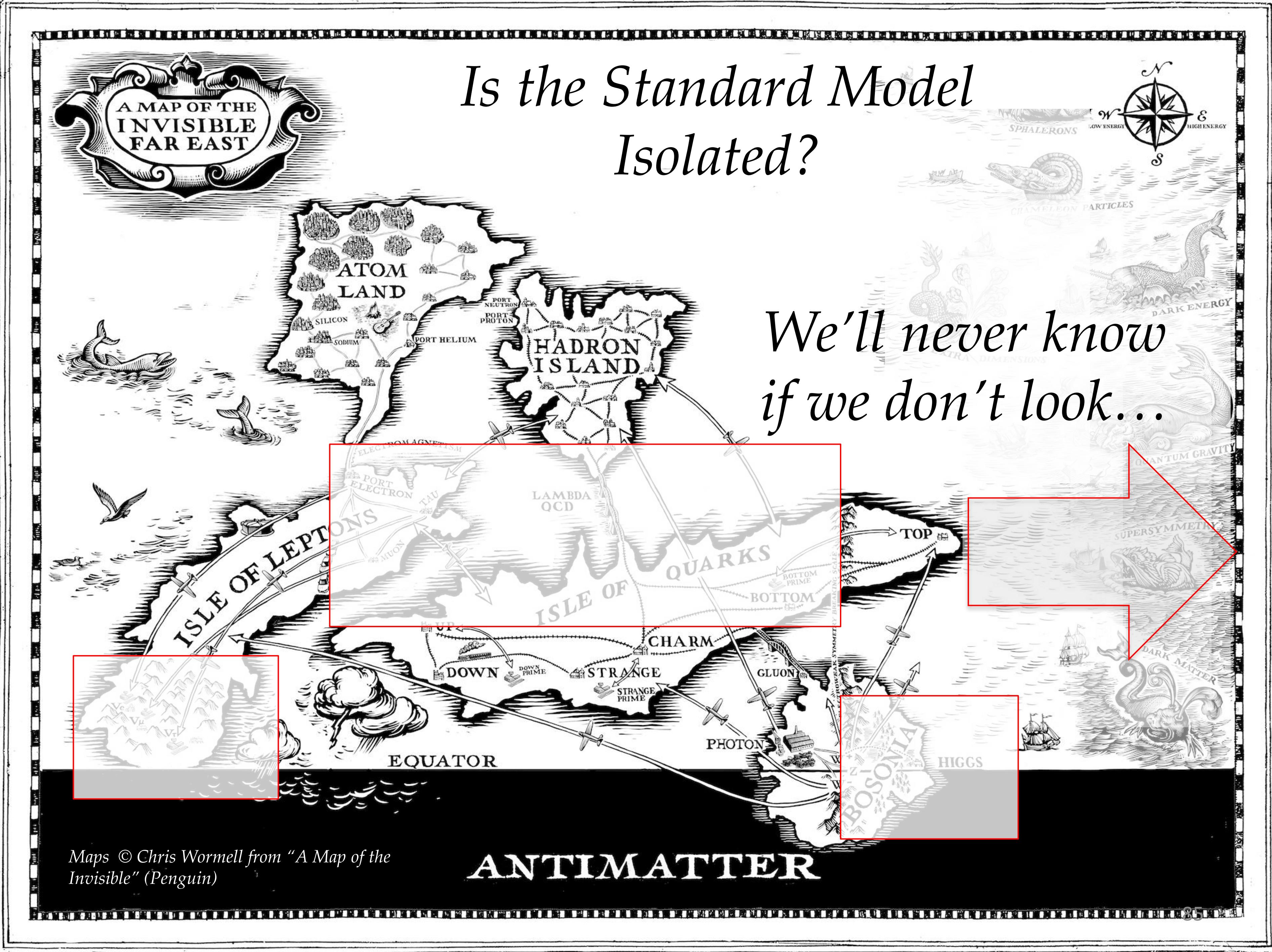
- *The guarantee is we will find out*
  - *Whether the Higgs-self coupling is as the Standard Model predicts*
  - *Whether or not the Standard Model continues to apply, well beyond the region in which it was developed, and to what precision*
  - *We will also push some amazing technologies, with likely benefits elsewhere, including computational and software techniques* *
- *And there may be surprises…*

Is the Standard Model Isolated?



Is the Standard Model Isolated?

We'll never know if we don't look…



Maps © Chris Wormell from "A Map of the Invisible" (Penguin)