Nikhef

Data Analysis in Particle Physics

Flavia de Almeida Dias (she/her)

@fladias_phys

CERN International Teacher Weeks Programme

14 August 2023

University of Amsterdam
Institute of Physics

ATLAS EXPERIMENT

About Me

CALTECH

Univ. Edinburgh

Niels Bohr Inst.

UvA/Nikhef

CERN

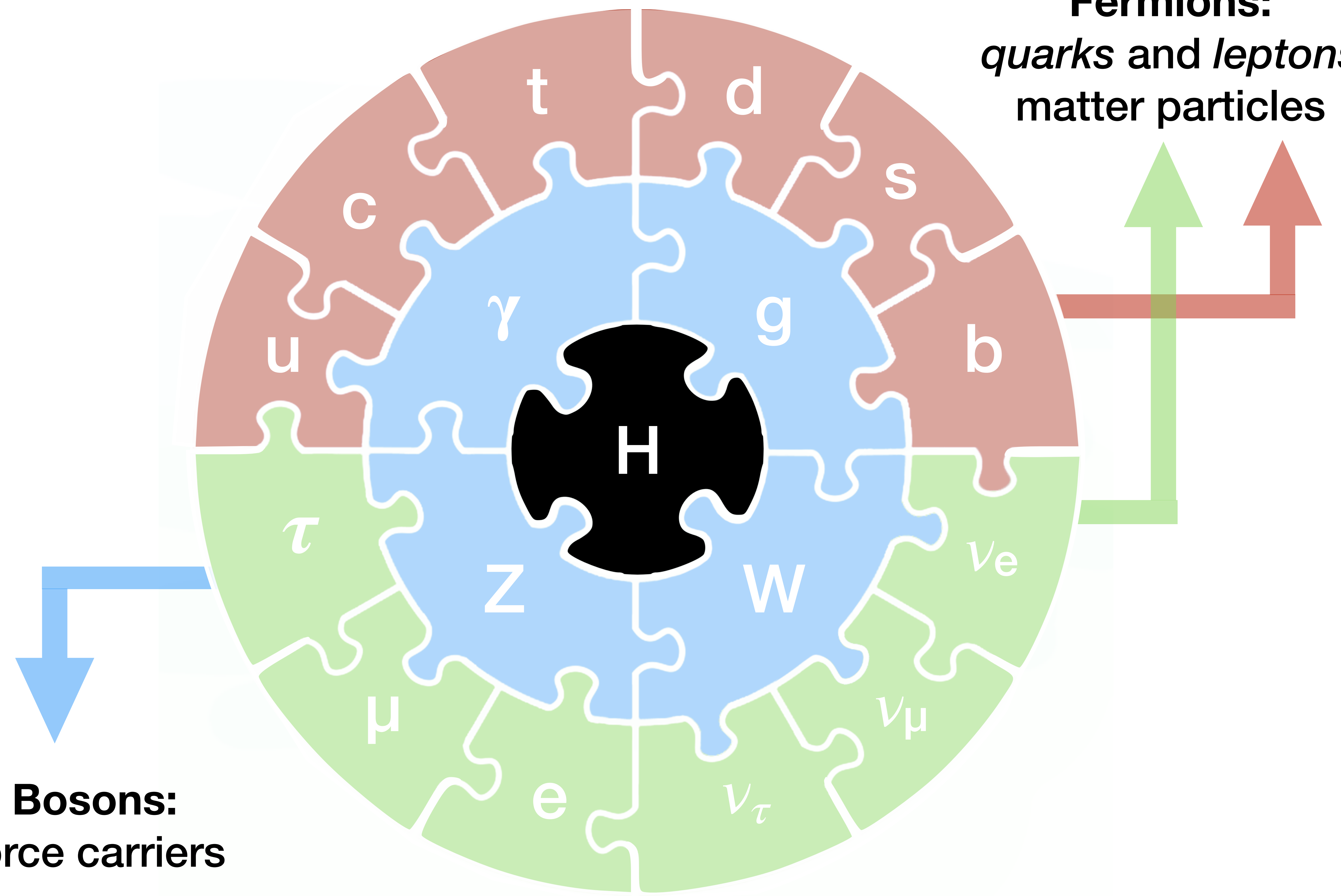
São Paulo Univ.

Recap: Particle Physics

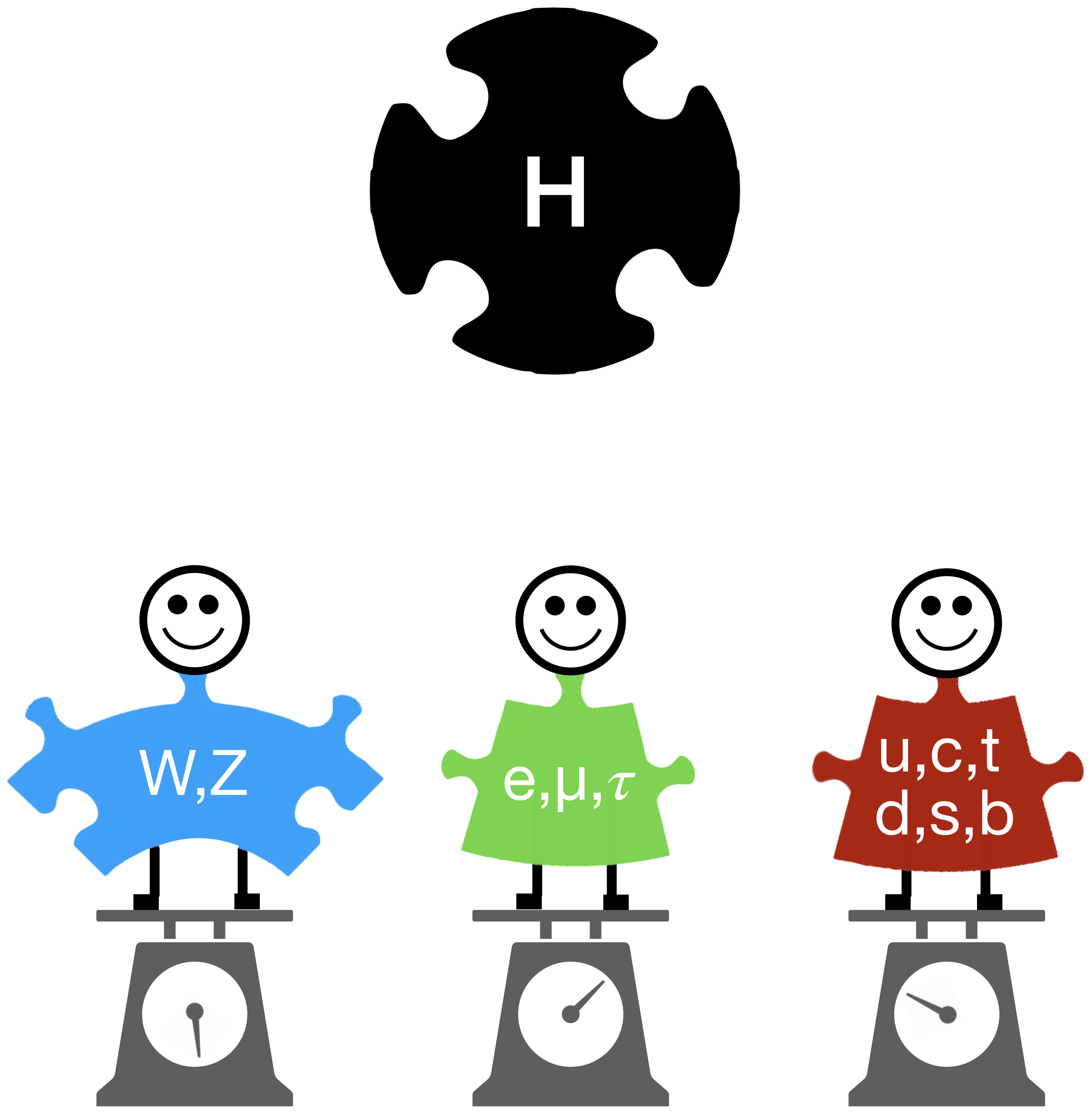
University of Amsterdam
Institute of Physics

ATLAS EXPERIMENT

Standard Model of Particle Physics

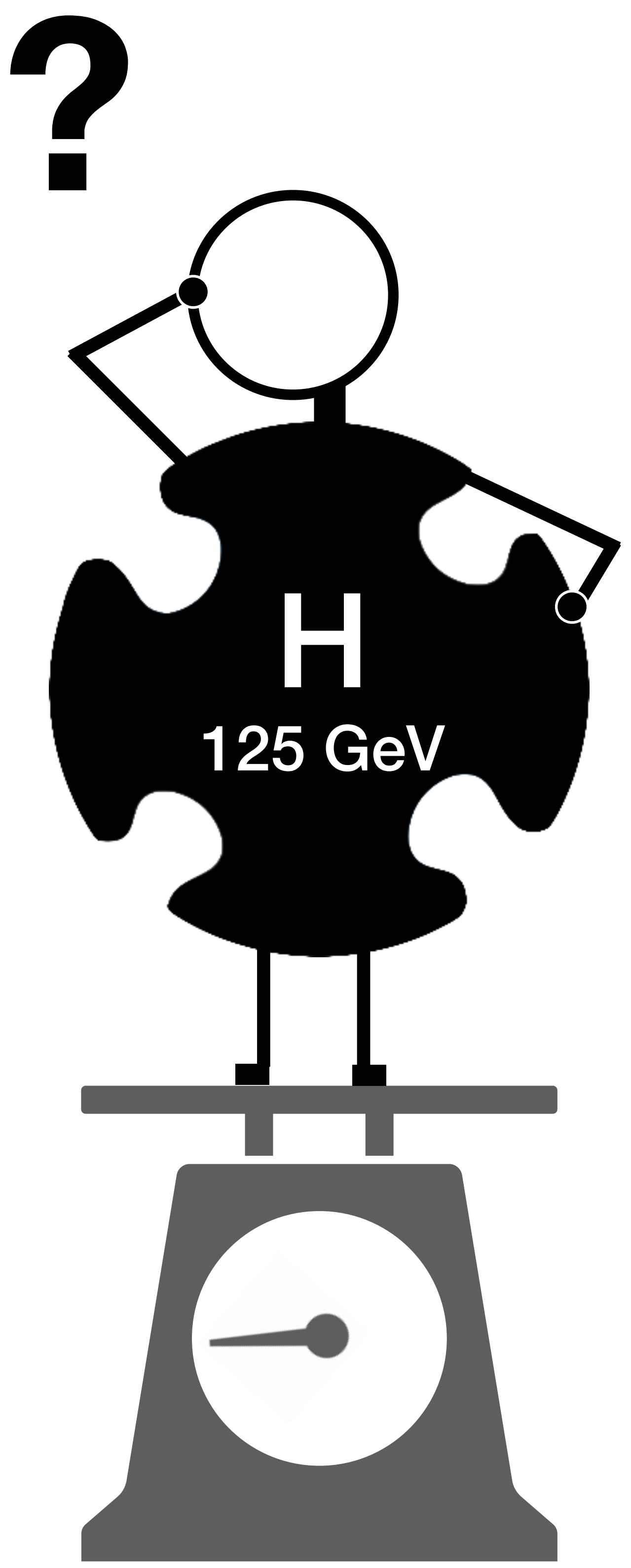


Higgs mechanism



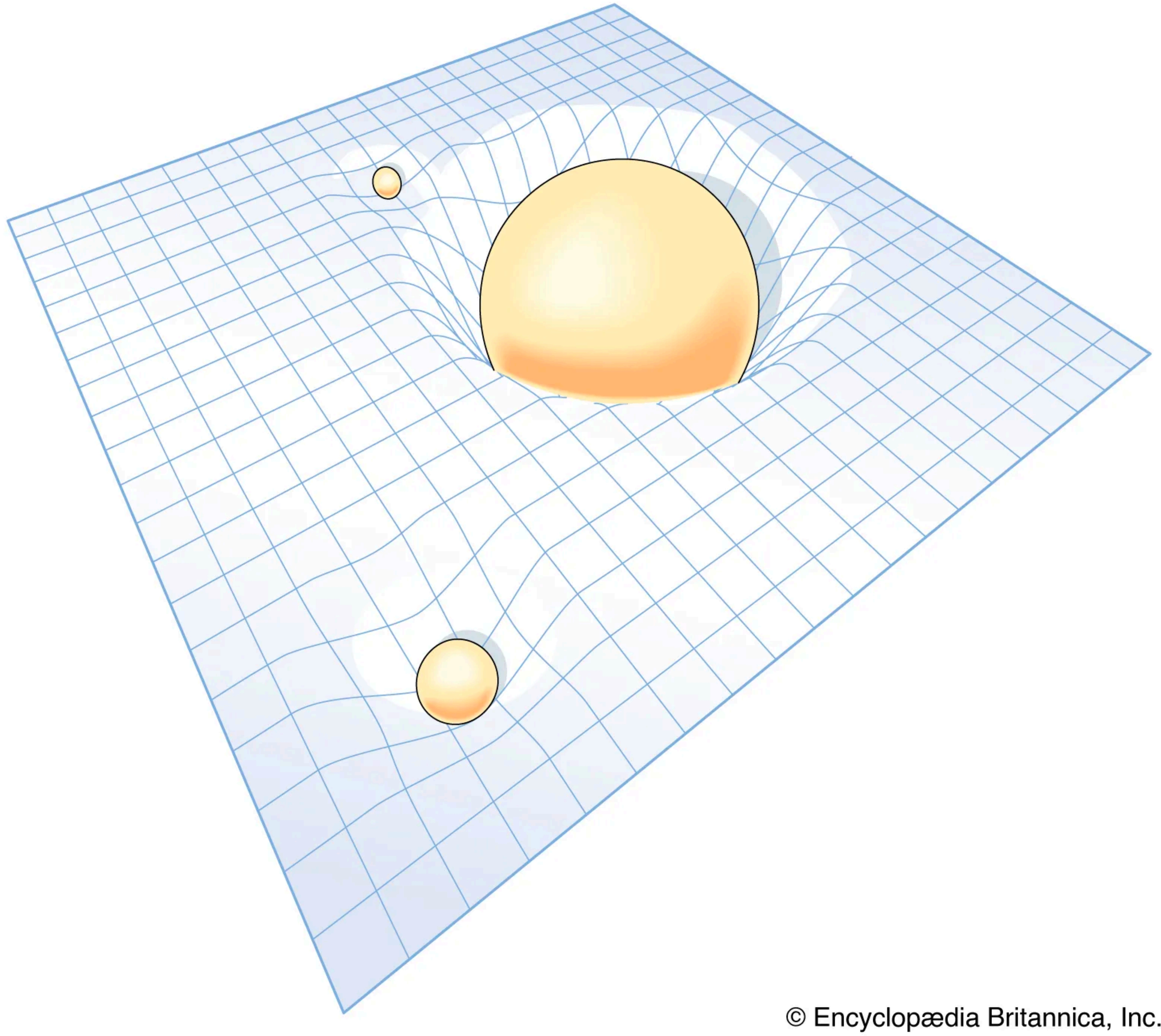
Open questions

Higgs Boson Mass



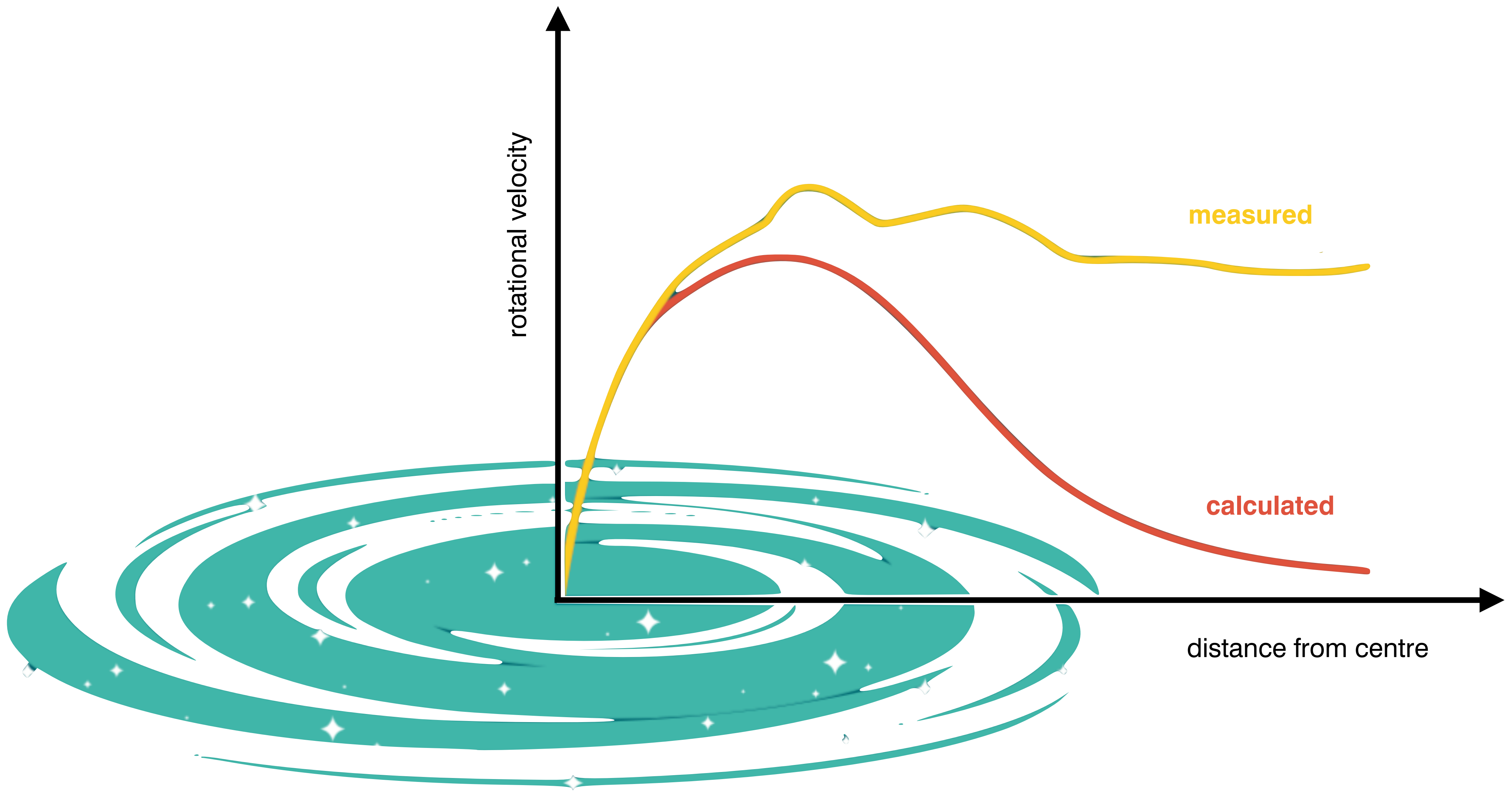
Due to *new particles* or *new interactions*?

Gravity



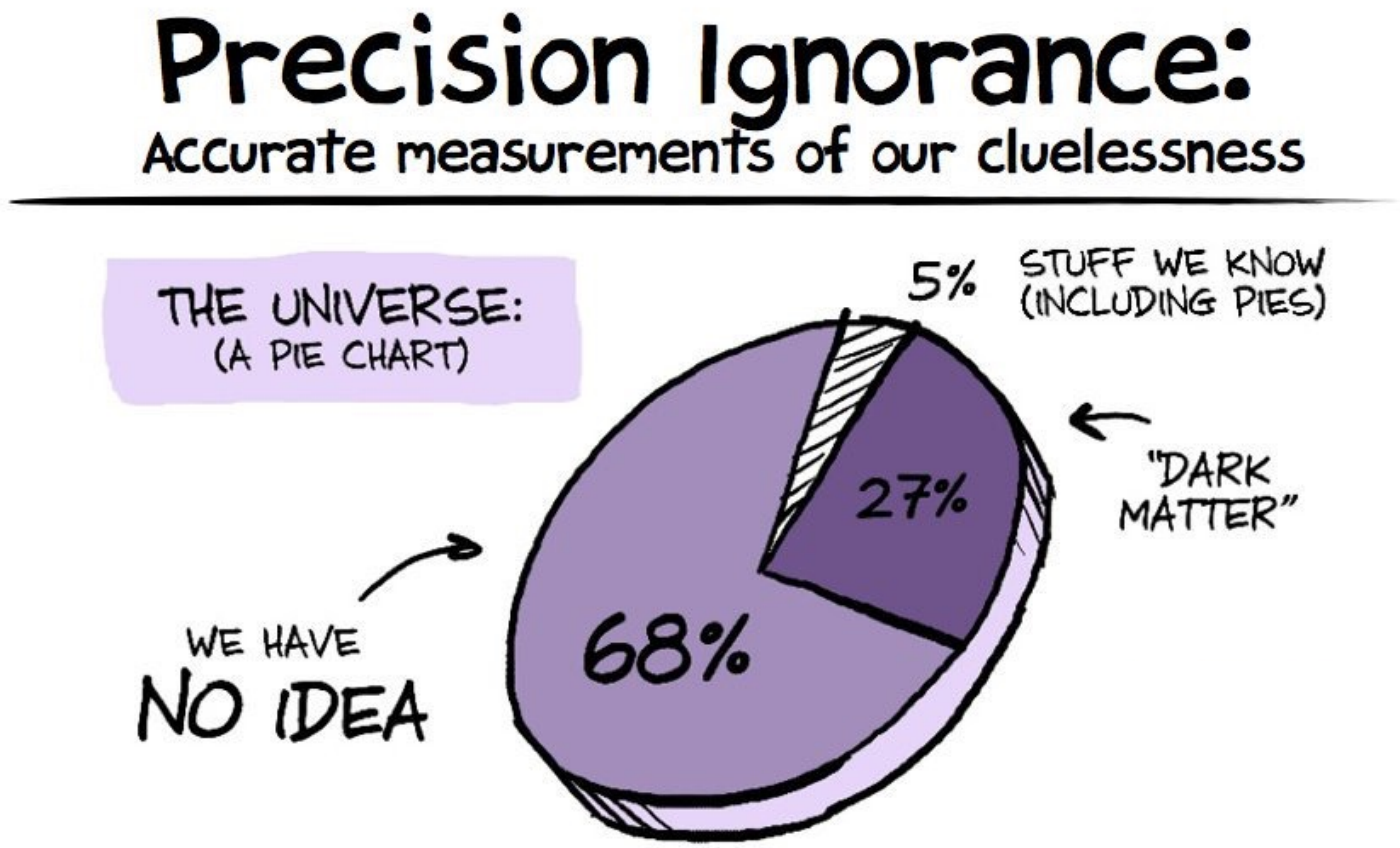
Open questions

Dark Matter



Credit: Higgs Boson & Beyond

Dark Energy

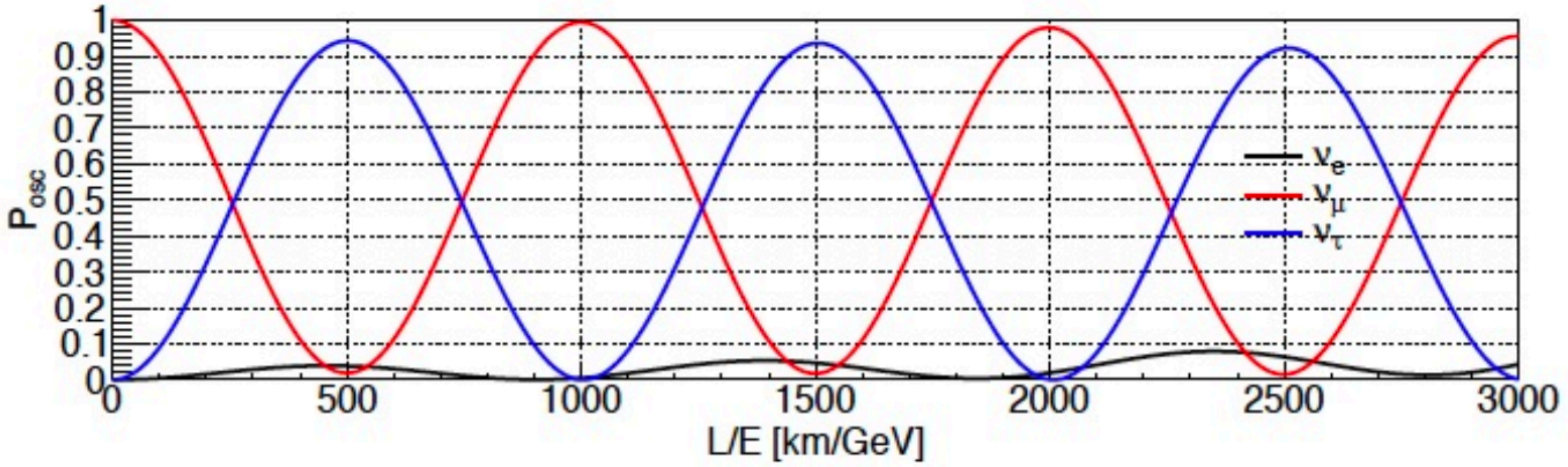


Credit: Cham & Whiteson

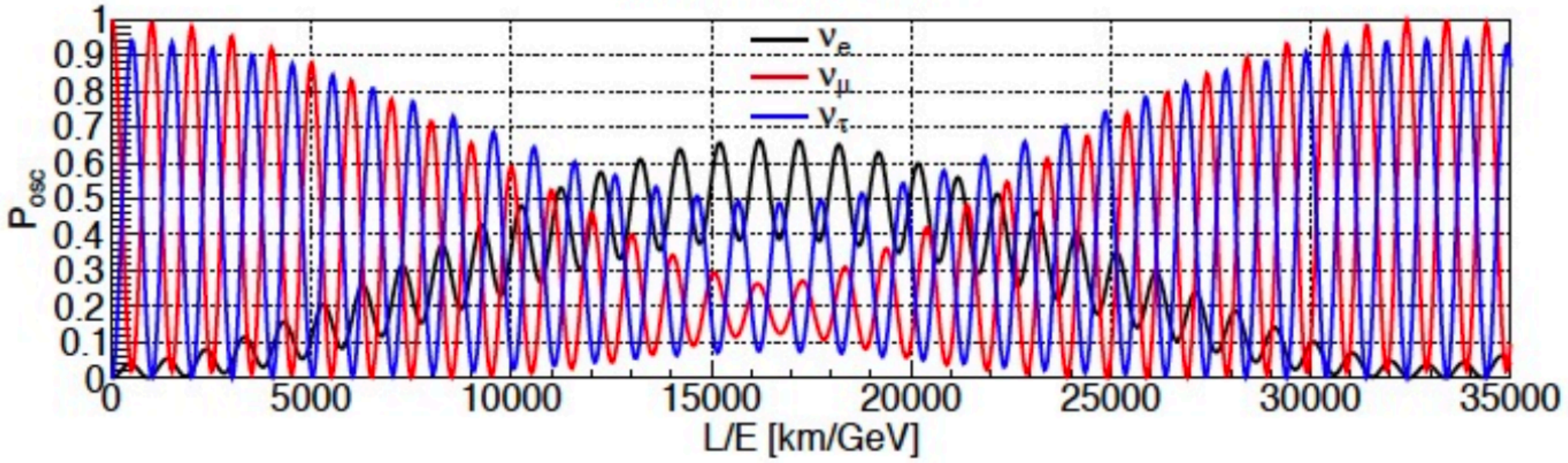
Open questions

Neutrino Masses

Oscillations in vacuum, starting with muon neutrino

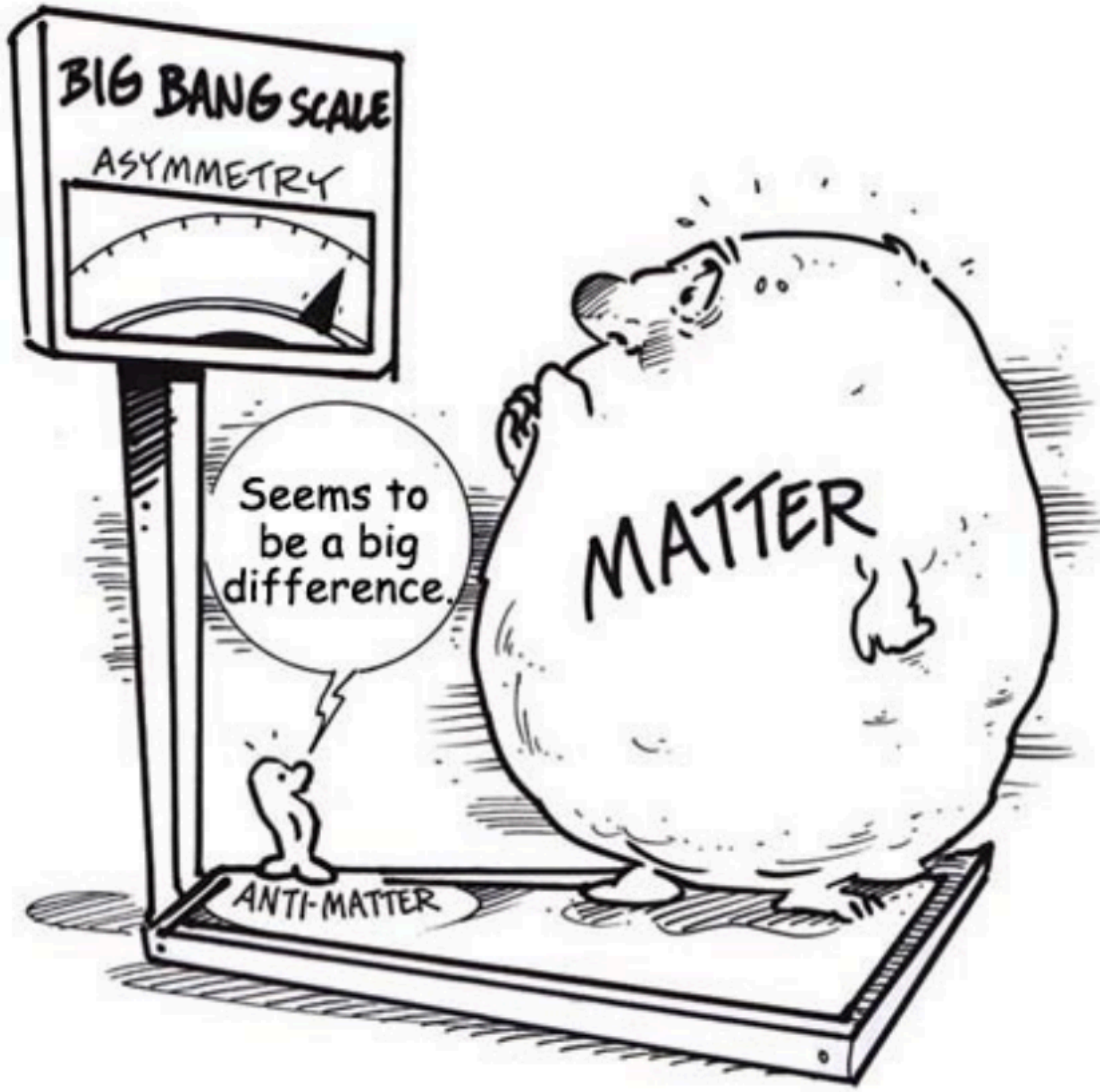


(c) ν_μ, short range



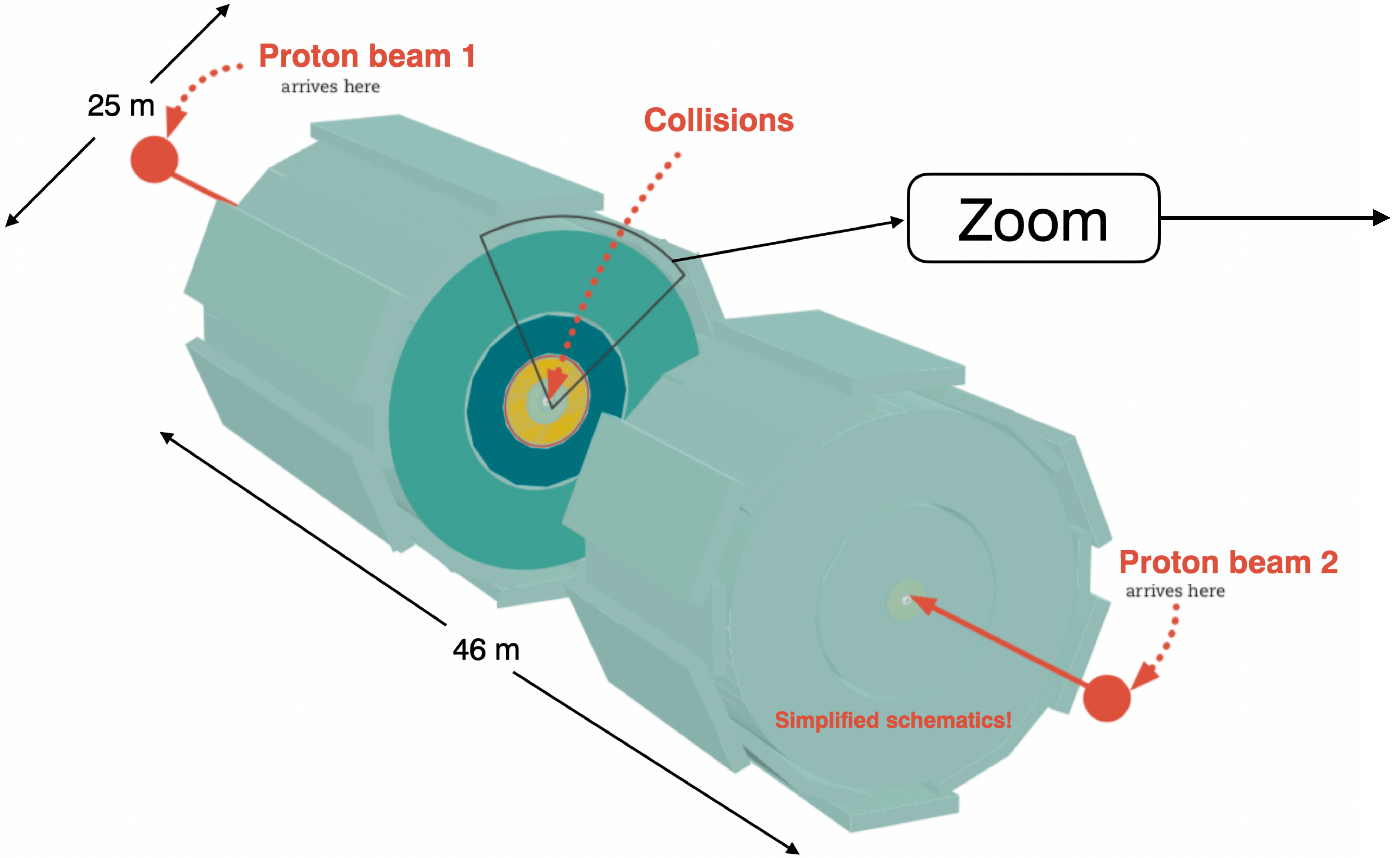
(d) ν_μ, long range

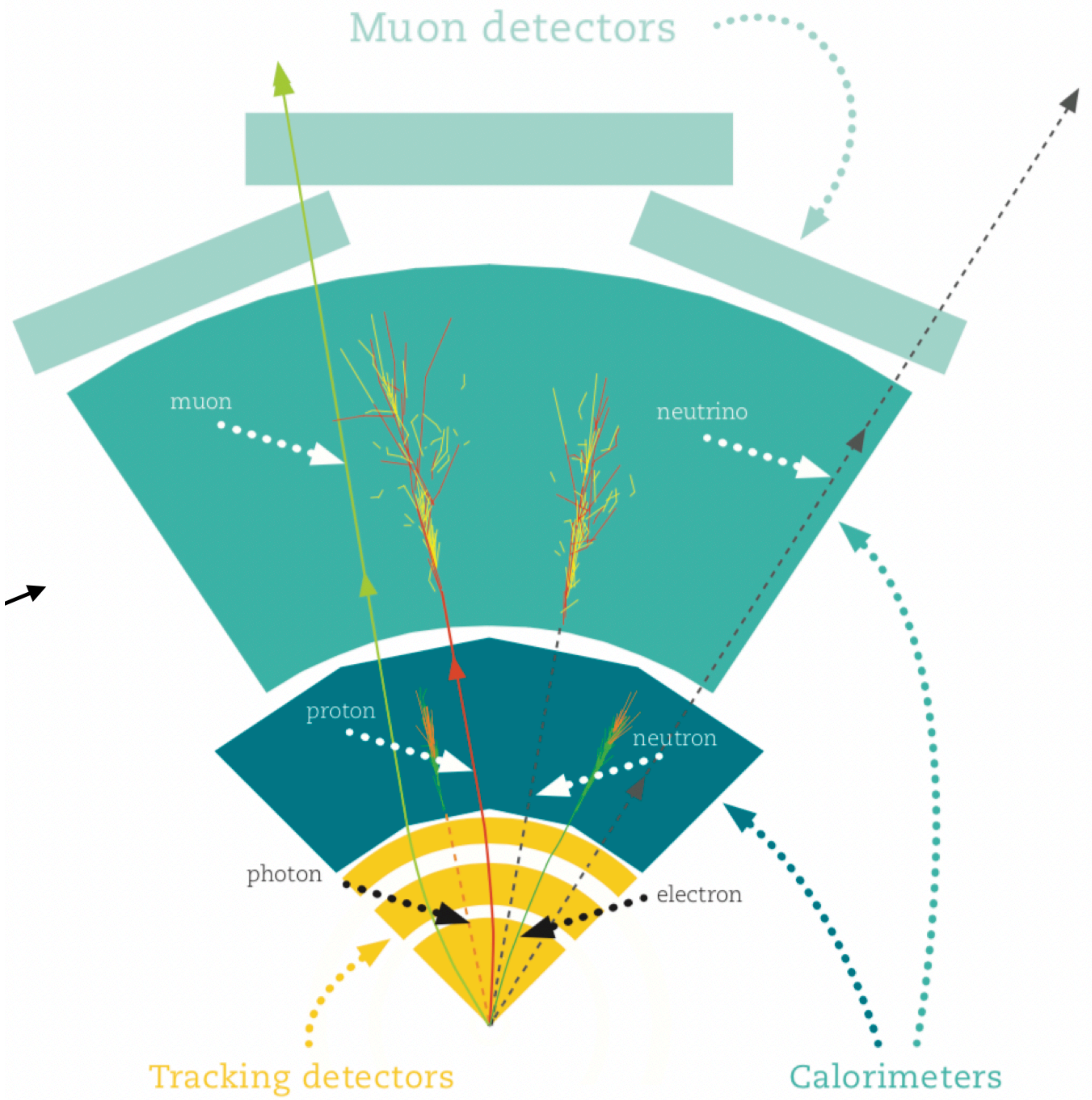
Matter-Antimatter Asymmetry



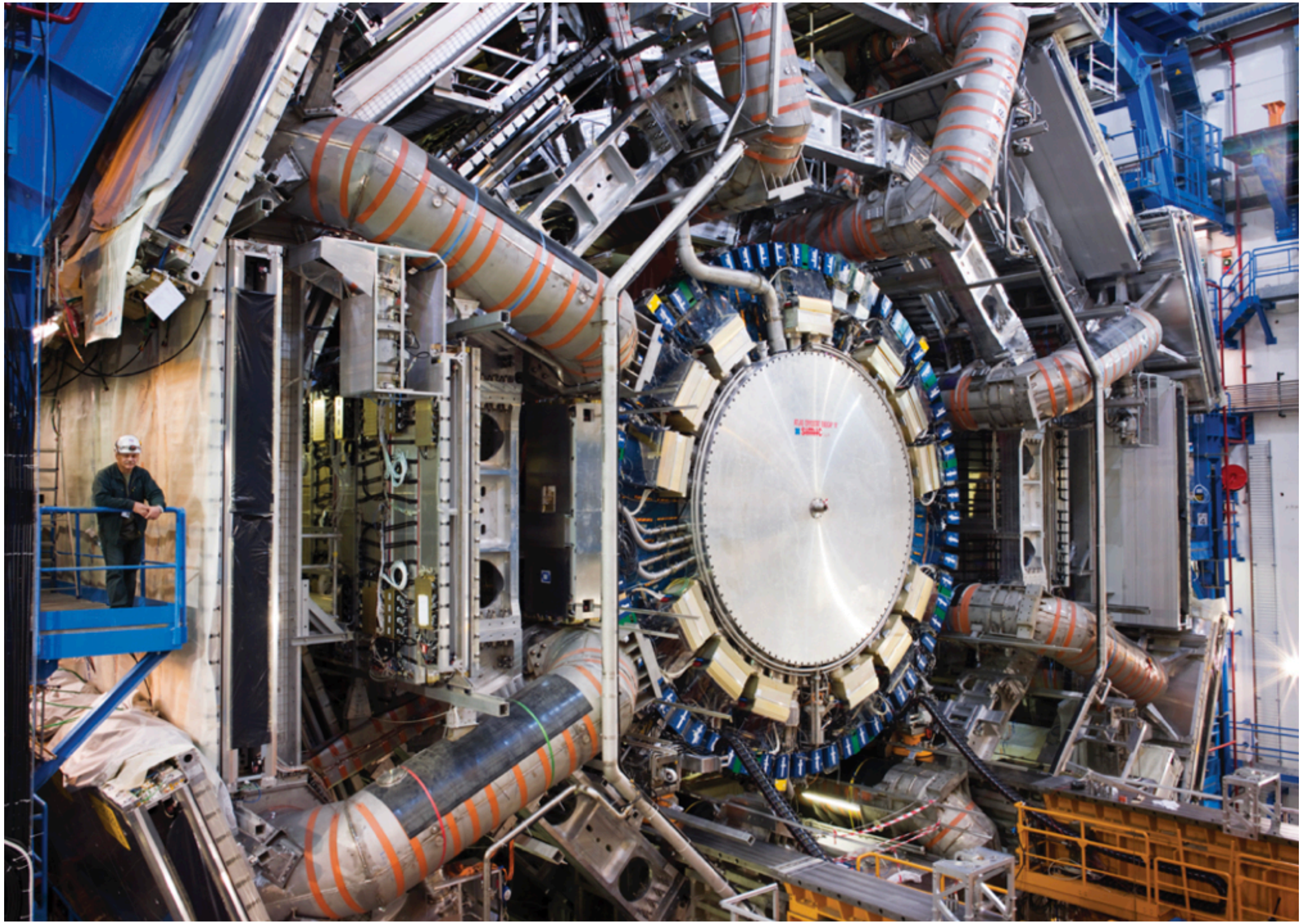
Large Hadron Collider

Particle Detectors





ATLAS Experiment



Data Analysis

What is an analysis?

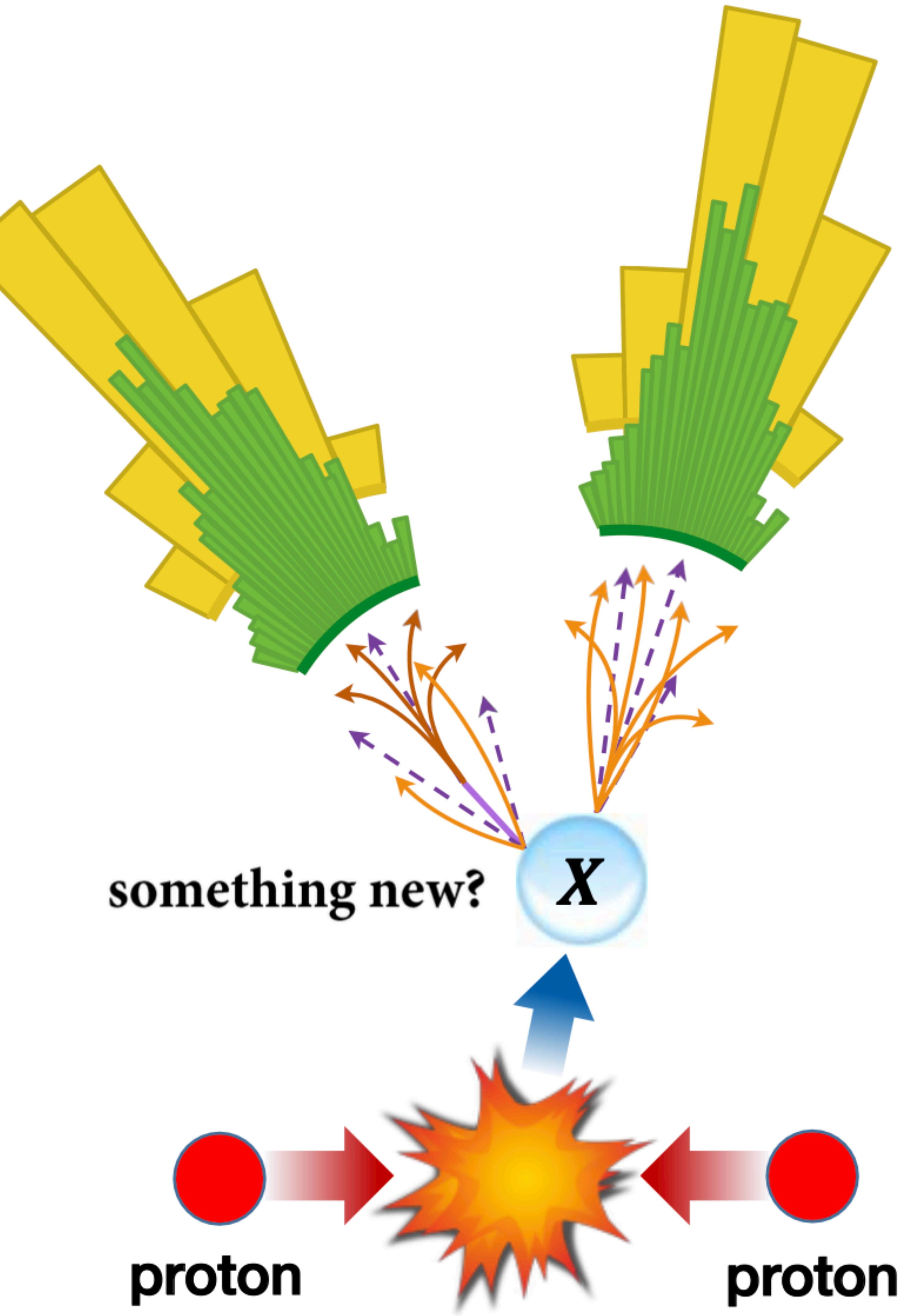
- A **scientific statement** from experimentation
- Result: Published numbers with uncertainties
- Types:
  - ➡ **Measurement**: This known process looks like this
  - ➡ **Search**: This new process exists or not
  - ➡ **Performance, R&D**: This algorithm/detector component works this well, improvements could be…

Analysis Ingredients

1. Define process of interest
2. Simulate how it would look like in the detector
3. Select events of interest
4. Estimate number of background events
5. Estimate uncertainties
6. Plot observables of interest
7. Perform statistical analysis to extract final parameter of interest
8. Pass peer-review (within and outside ATLAS)

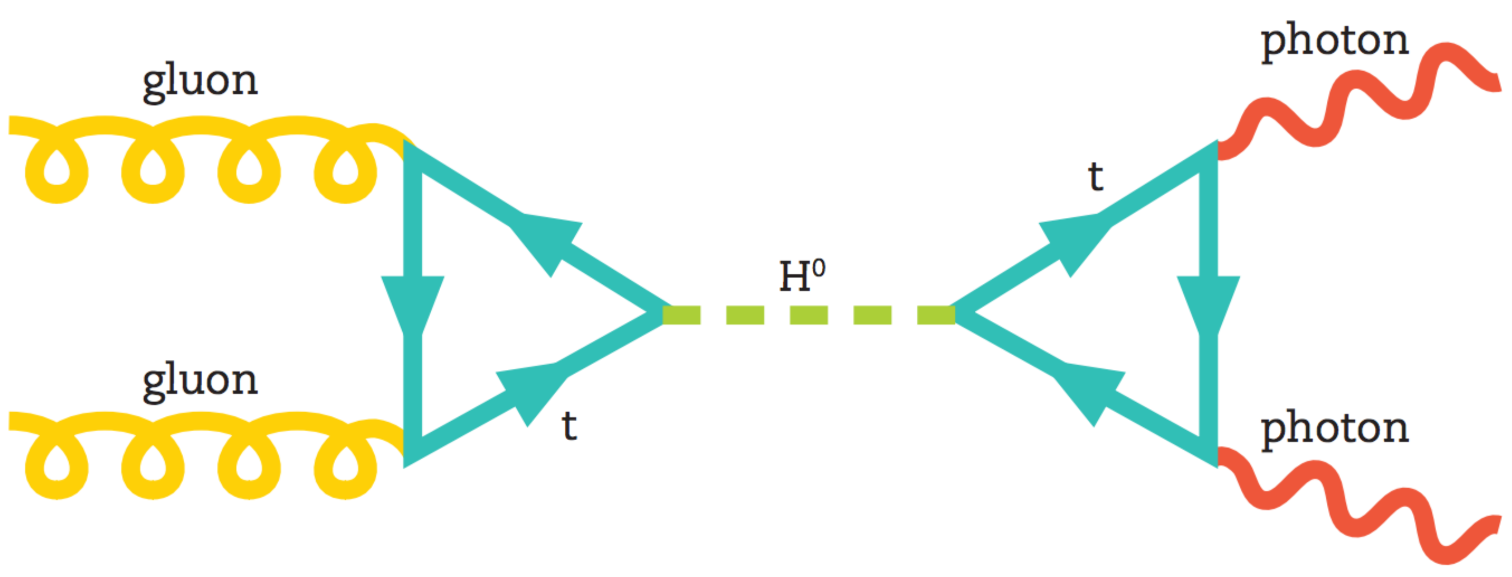
Step by step with example ATLAS analysis!

1. Define Process of Interest



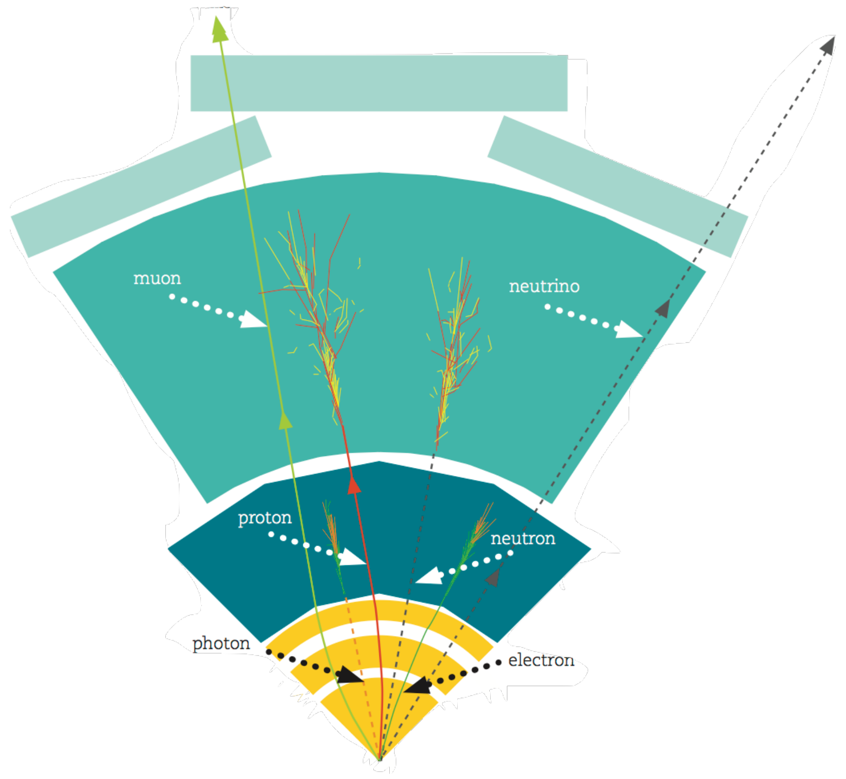
- **Chosen process:** pp → X → WW → JJ (signal)
- **Why?**
  - ➡ Related to Higgs mechanism
  - ➡ Probe for extra dimensions, new forces
  - ➡ Final state with jets probe highest collision energies
- **Current state-of-the-art**
  - ➡ Run-1 analysis had an excess
  - ➡ Use most up-to-date methods to identify jets

2. Simulate in ATLAS Detector



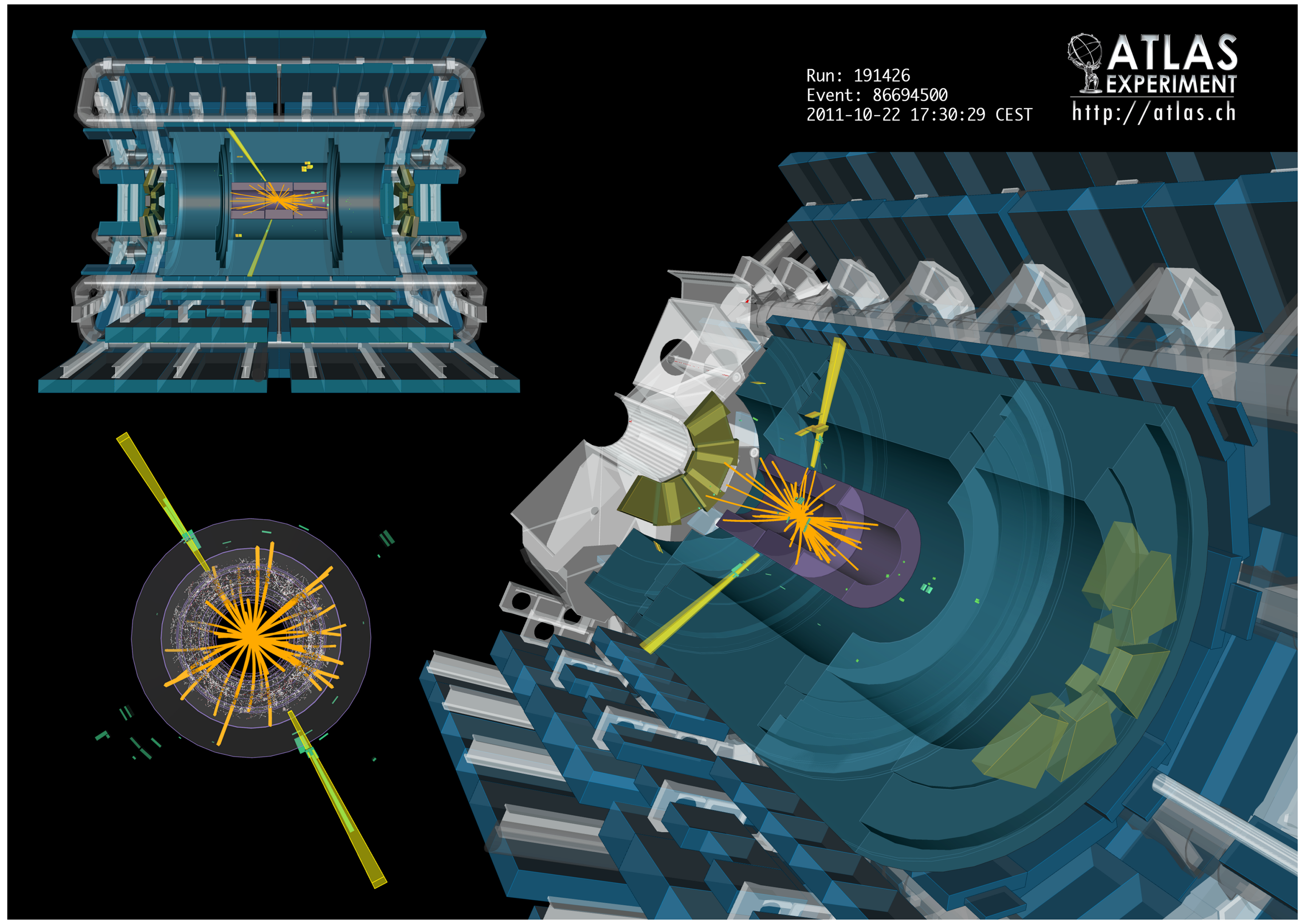
Feynman diagram

+



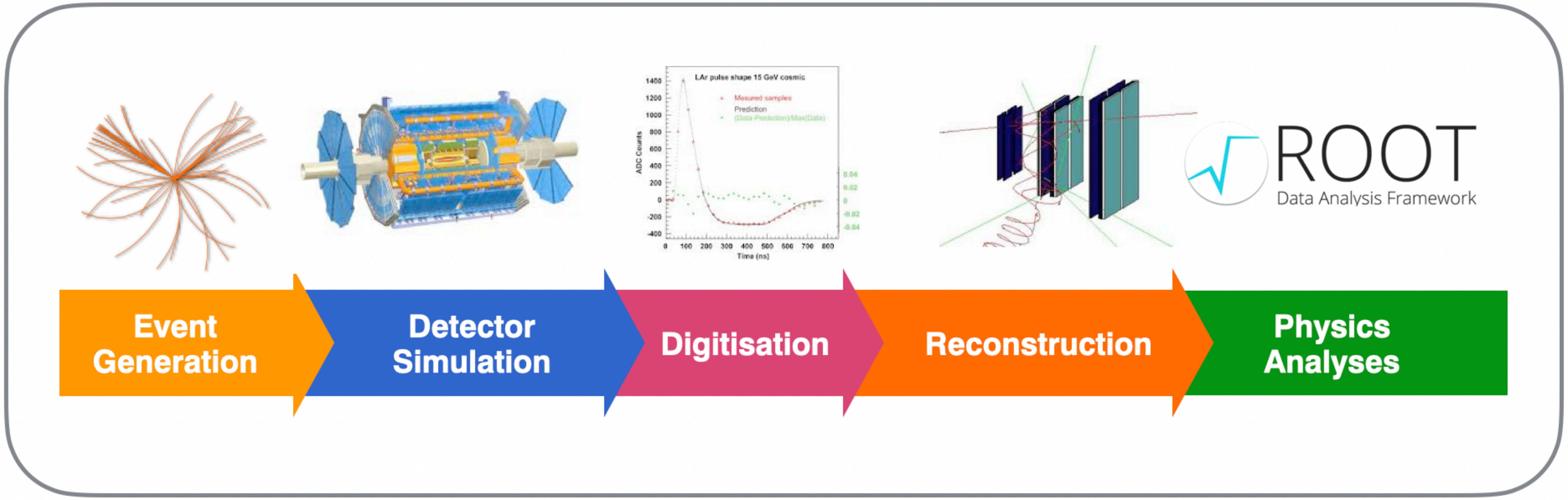
+

=



Nikhef

ATLAS Simulation

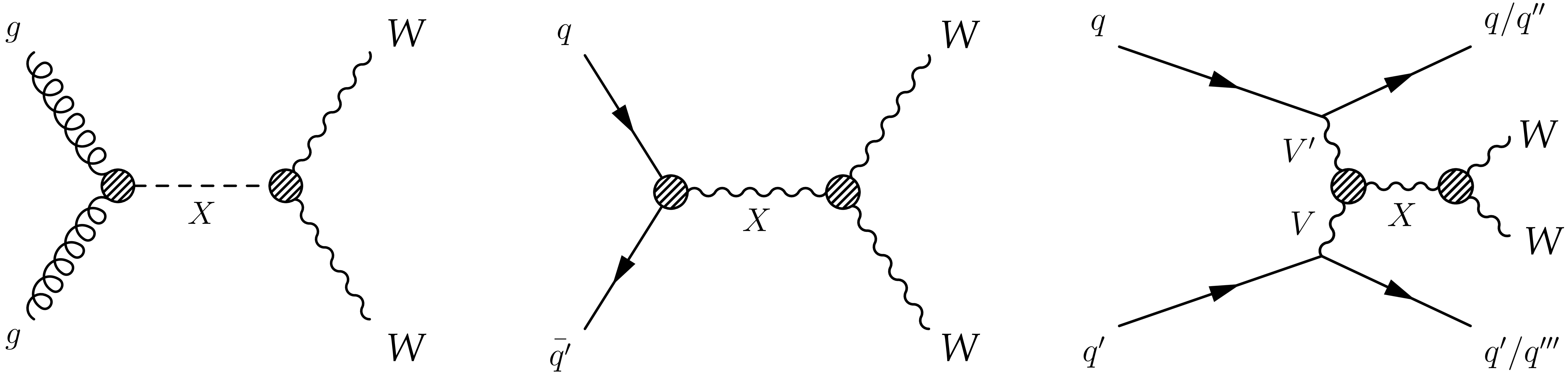


Collaborative C++/python inside ATLAS Athena framework

- Multi-step and computationally intensive procedure
  - ➡ Crucial to understand what we observe in the detector

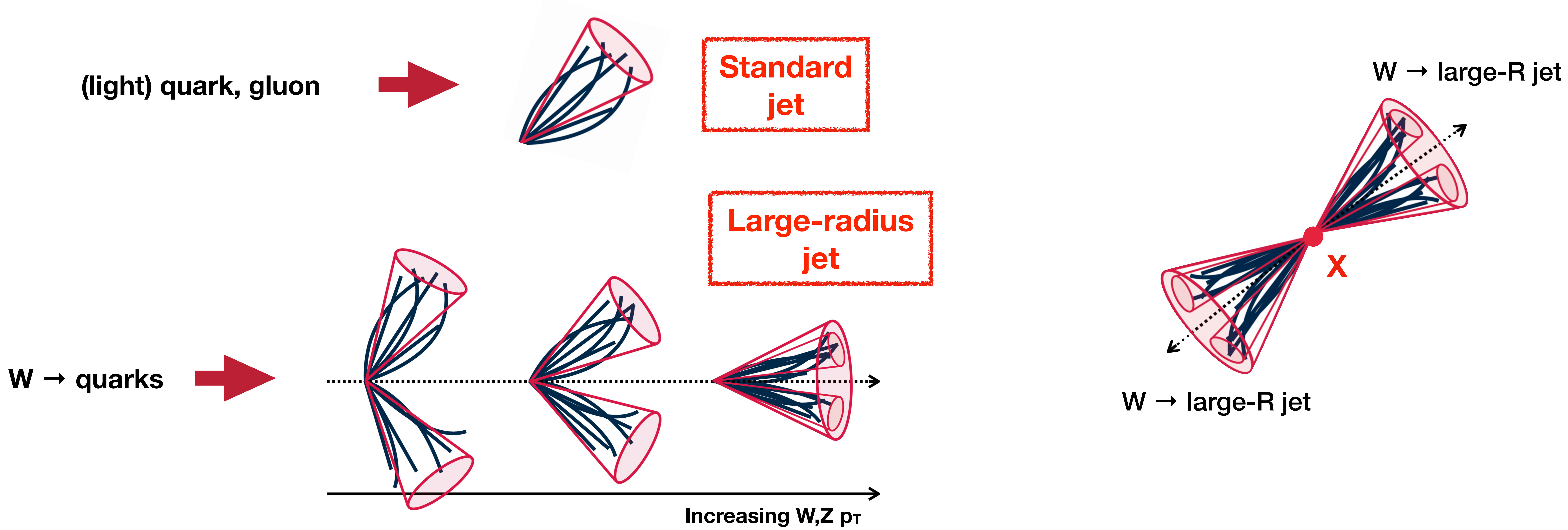
Our Signal: pp→X→WW→JJ

- Feynman diagrams:



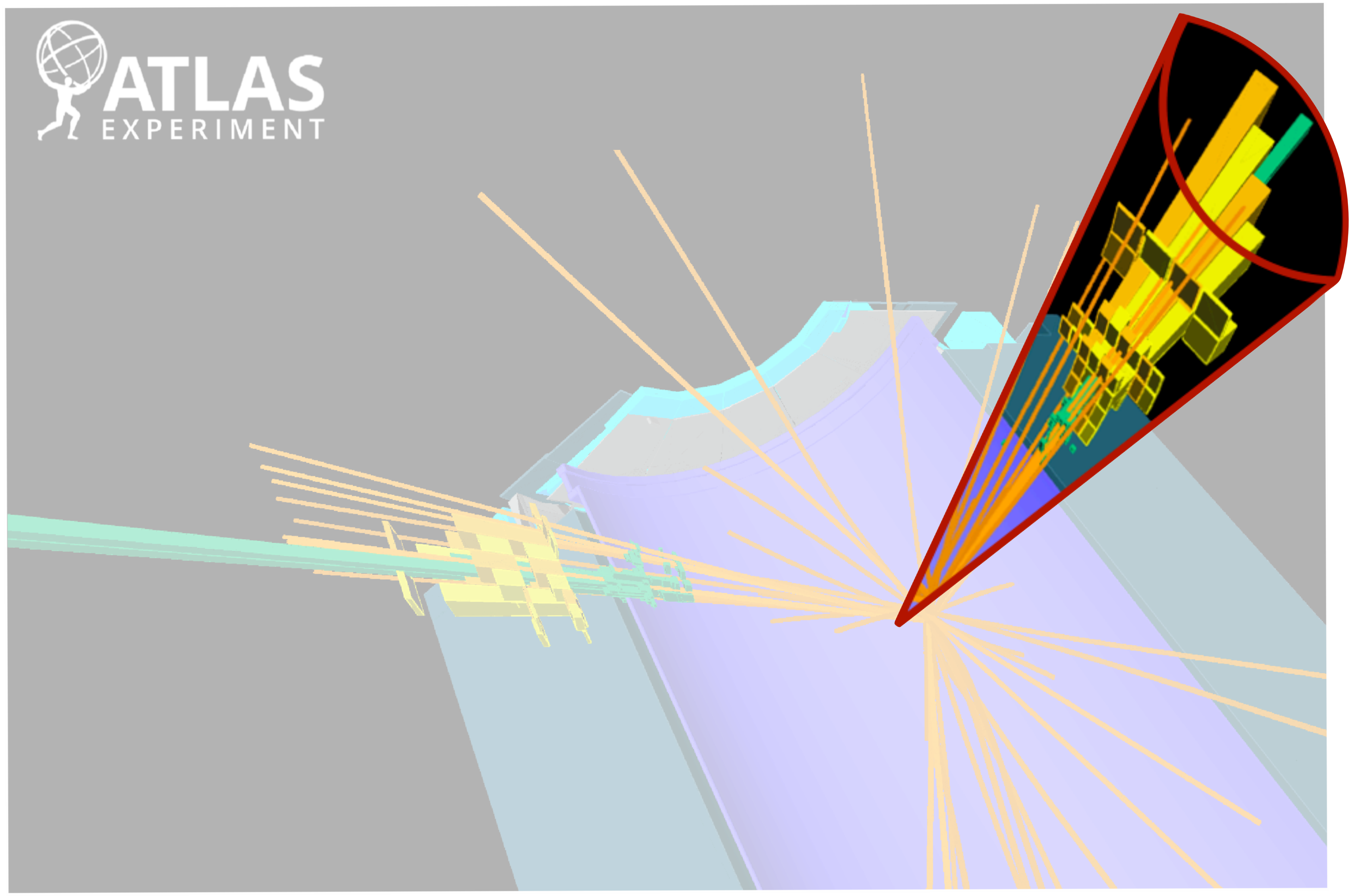
Nikhef

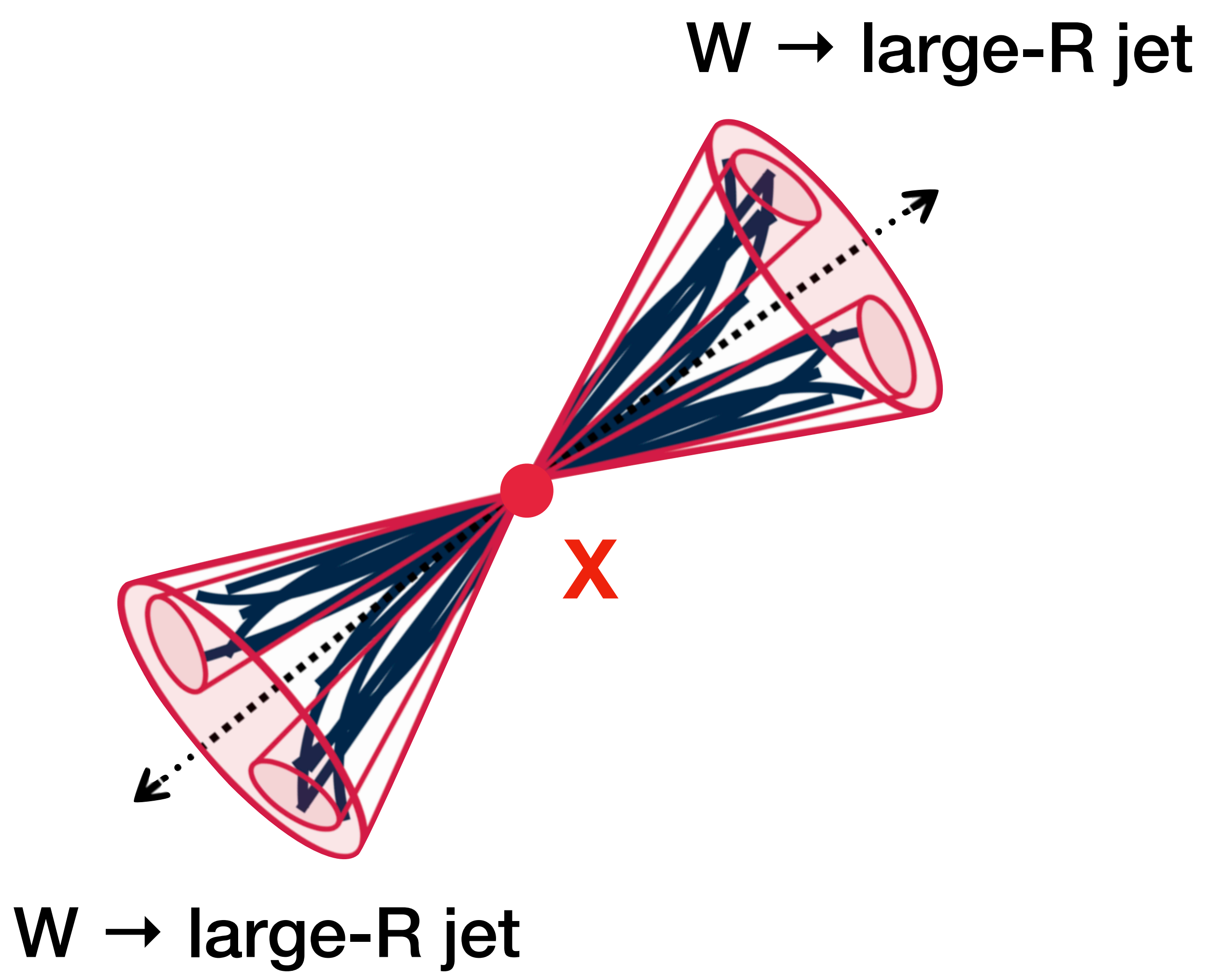
Our Signal: pp→X→WW→JJ



- X very heavy:
  - ➡ Each W boson will form a (large-radius) jet

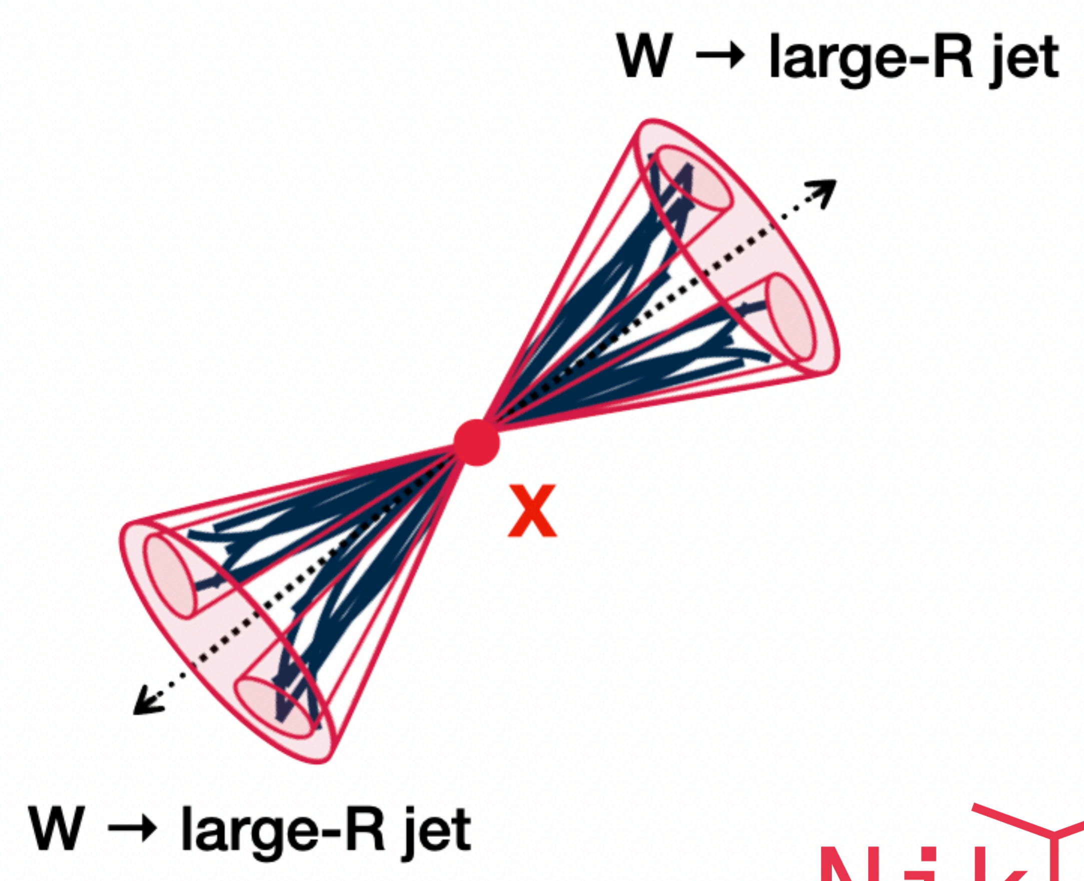
Our Signal: pp→X→WW→JJ





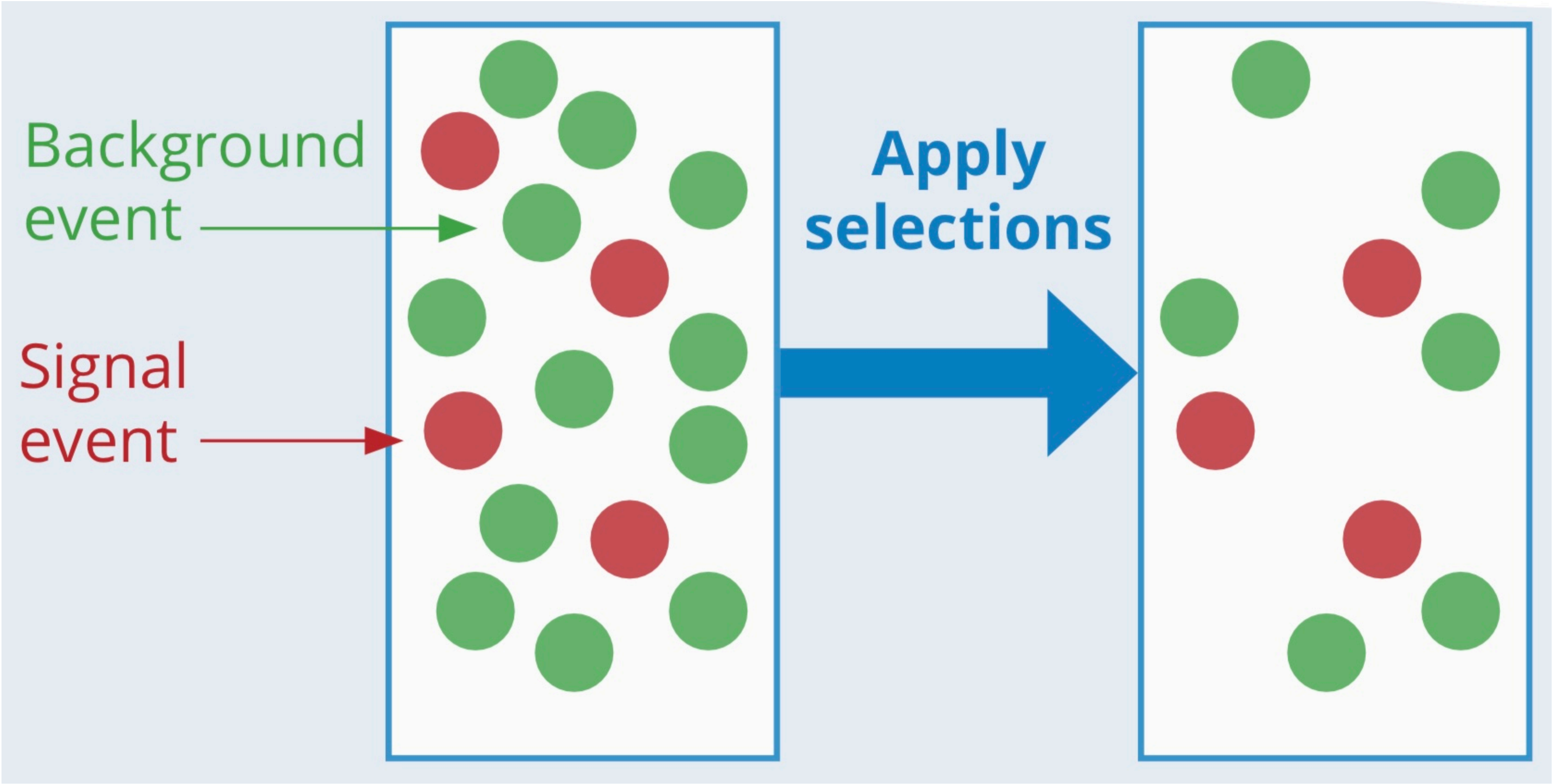
3. Select Events of Interest

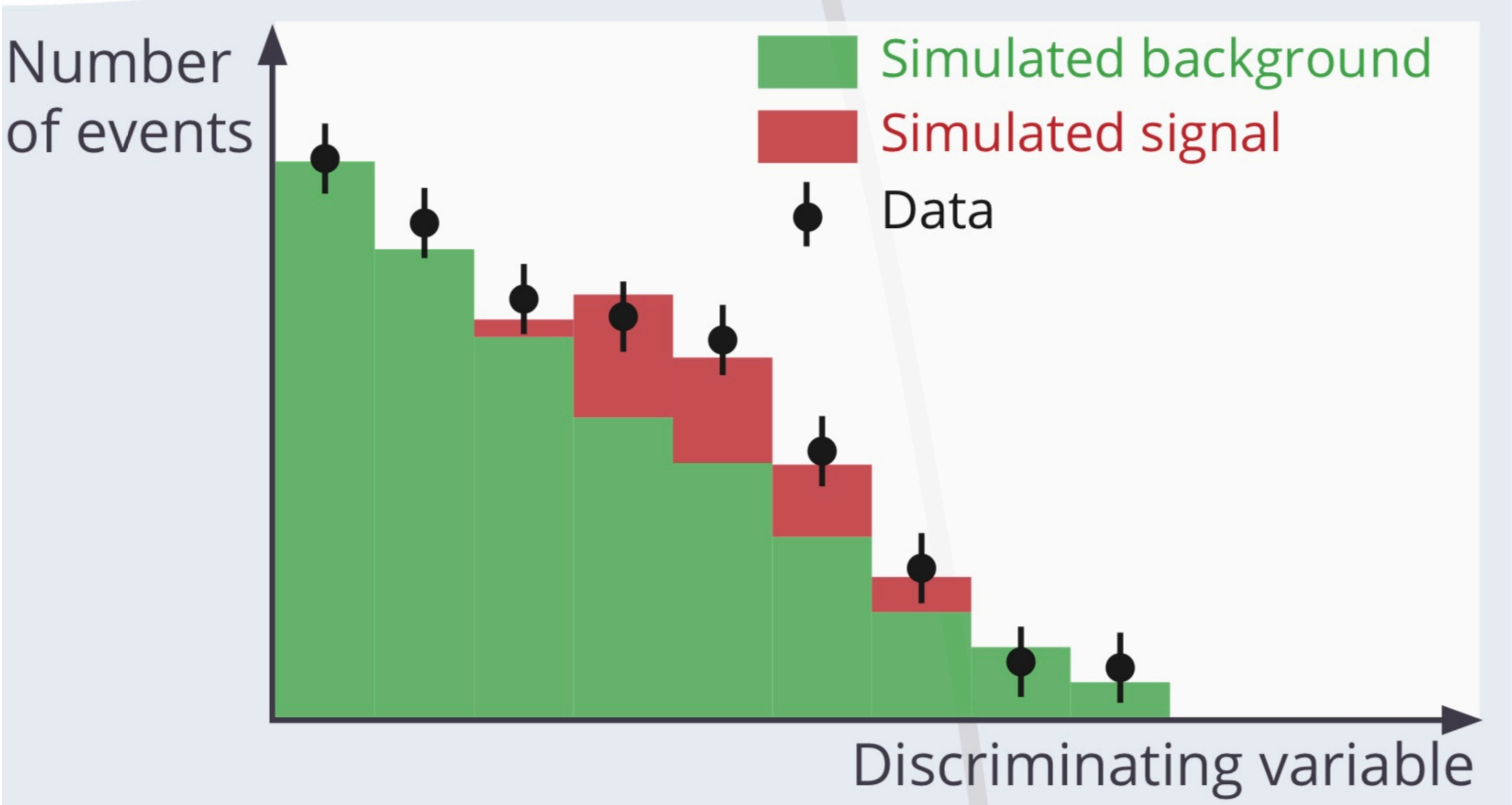
- **Online selection**: Trigger
  - ➡ Can't save all collision events!
    - 1.7 billion pp collisions per second (60M Mbps ⇒ 5400 simultaneous streams of 4K videos)
  - ➡ Select events with distinguishing characteristics that make them interesting!
    - Two stages: Level 1 hardware trigger (down to 100.000 events/s) and Level 2 software trigger (1.000 events/s)
- **Our analysis**: Trigger on very energetic jets
  - ➡ Special algorithms to select events with large-radius jets



Select Events of Interest

- **Offline selection:** Event selection
  - ➡ **Signal**: process of interest;
    Our analysis: X → WW → JJ
  - ➡ **Background**: any other process (in the Standard Model) which mimics the signal, with a similar signature in the detector
    Our analysis: QCD dijets, SM WW production
  - ➡ **Event selection:** increase signal-to-background ratio by favouring signal events
    Nowadays a lot of machine learning used!

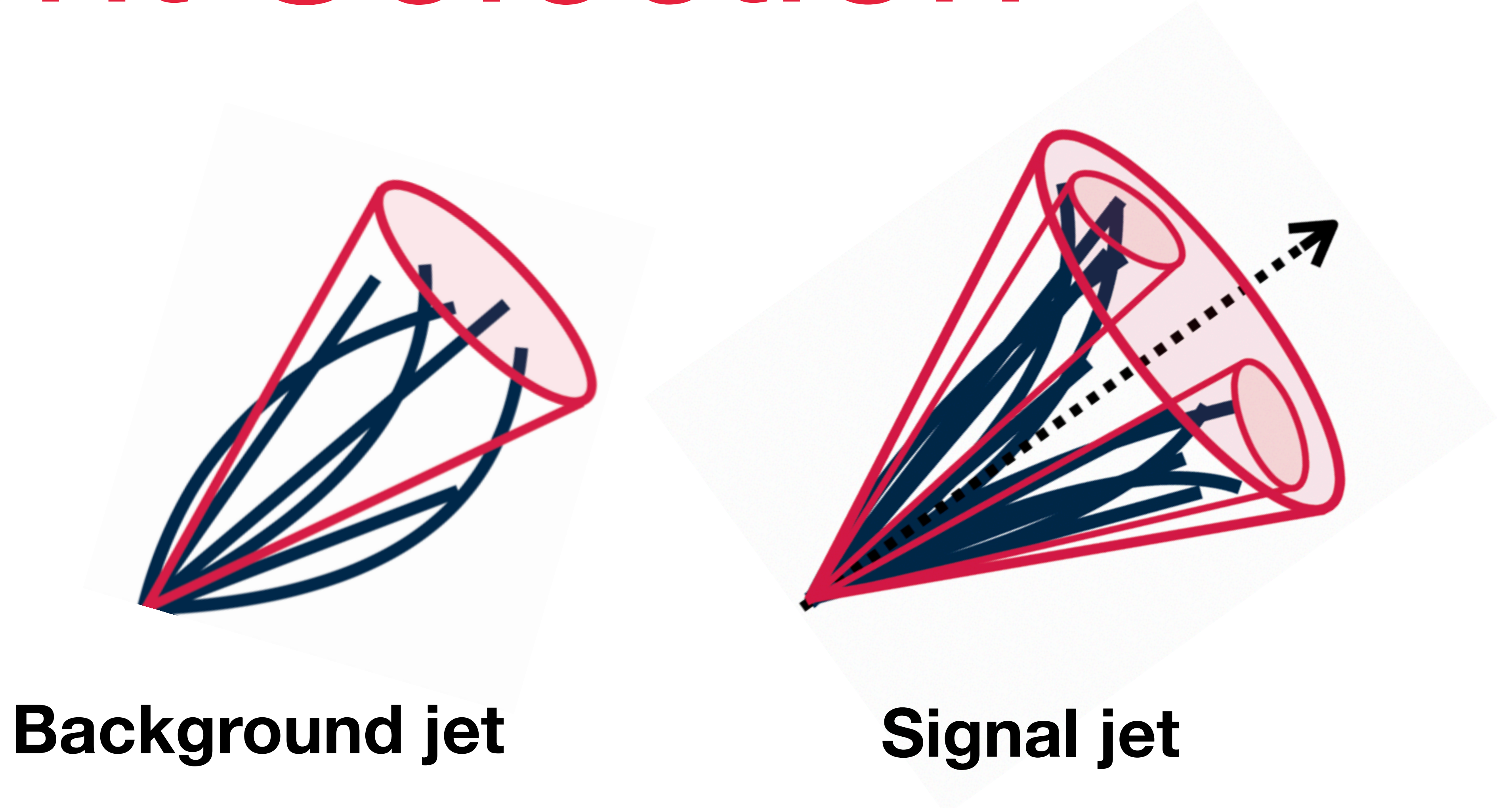


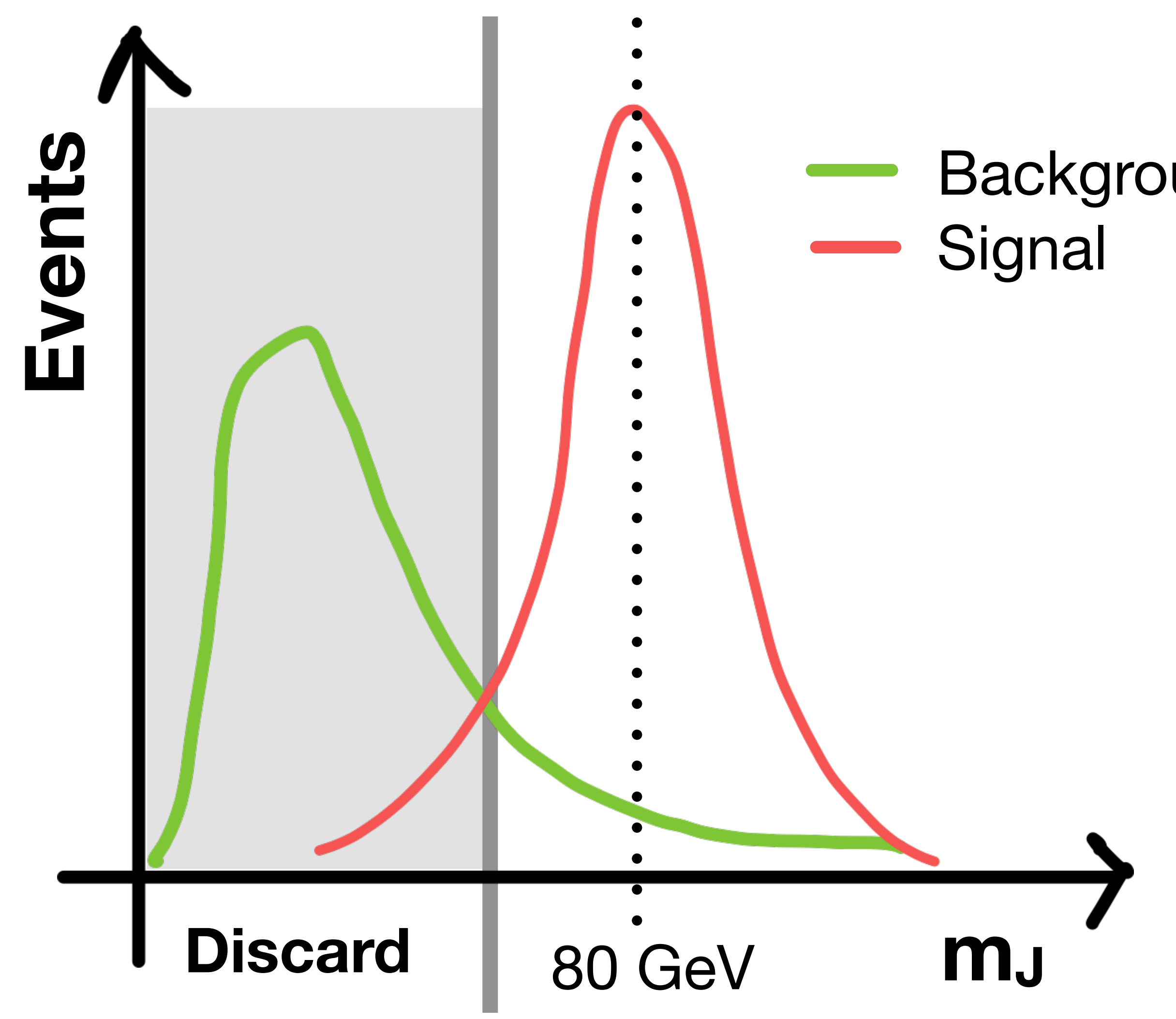


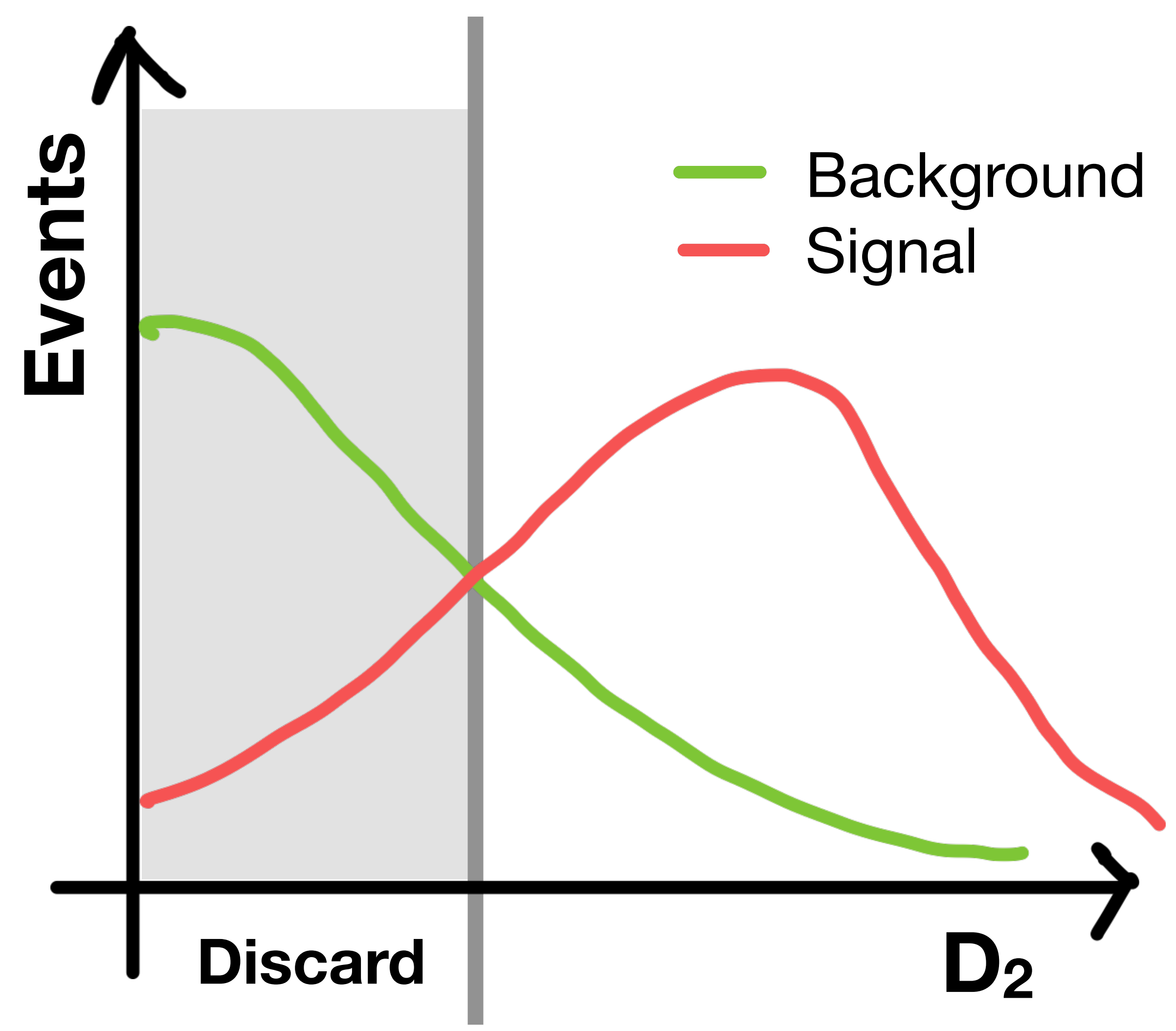
UvA

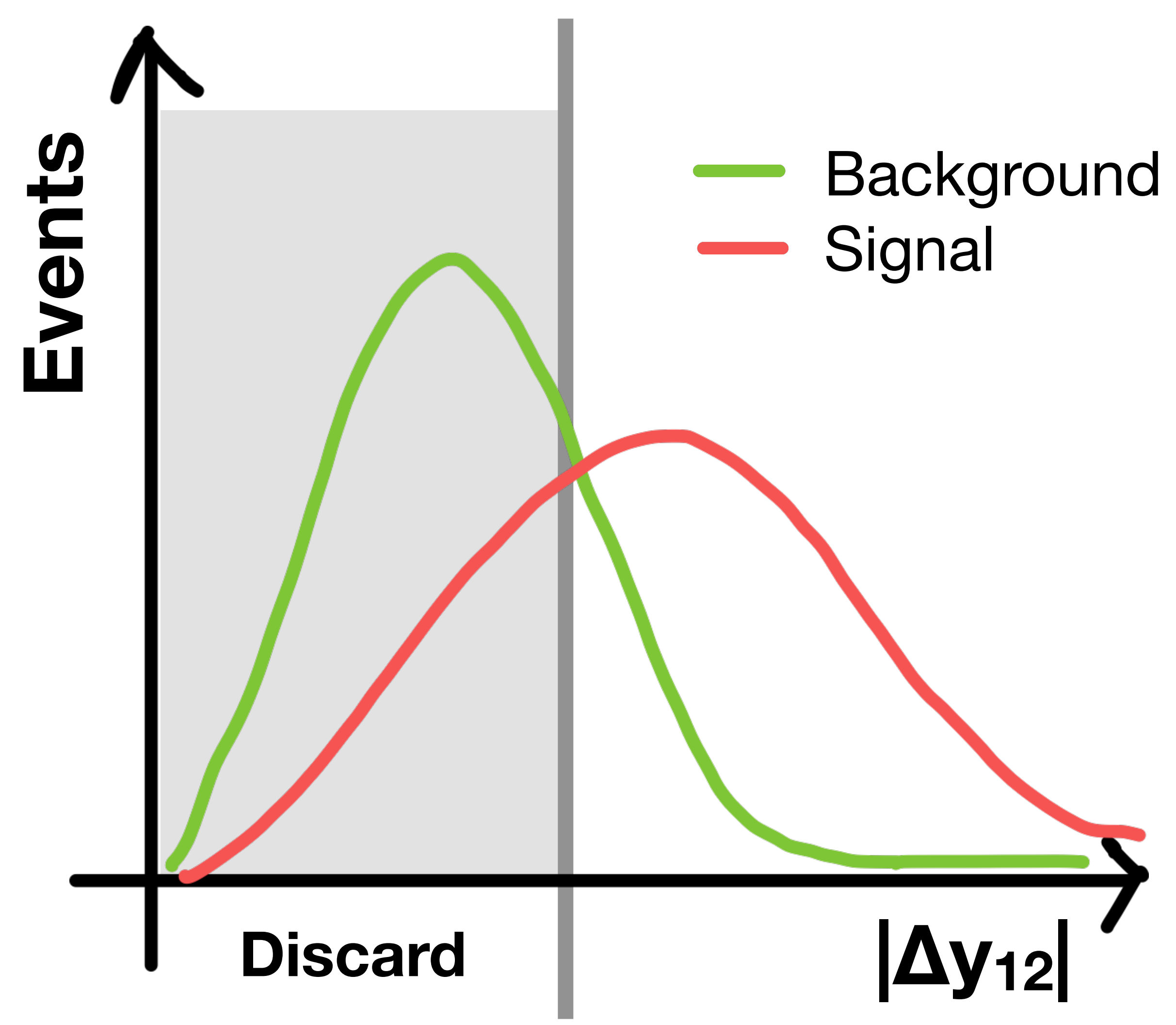
X→WW→JJ Event Selection

- Discriminant variables
  - ➡ Large-R jet mass (m_J)
  - ➡ Large-R jet energy correlation (D_2)
  - ➡ Spatial separation of jets ($|\Delta y_{12}|$)



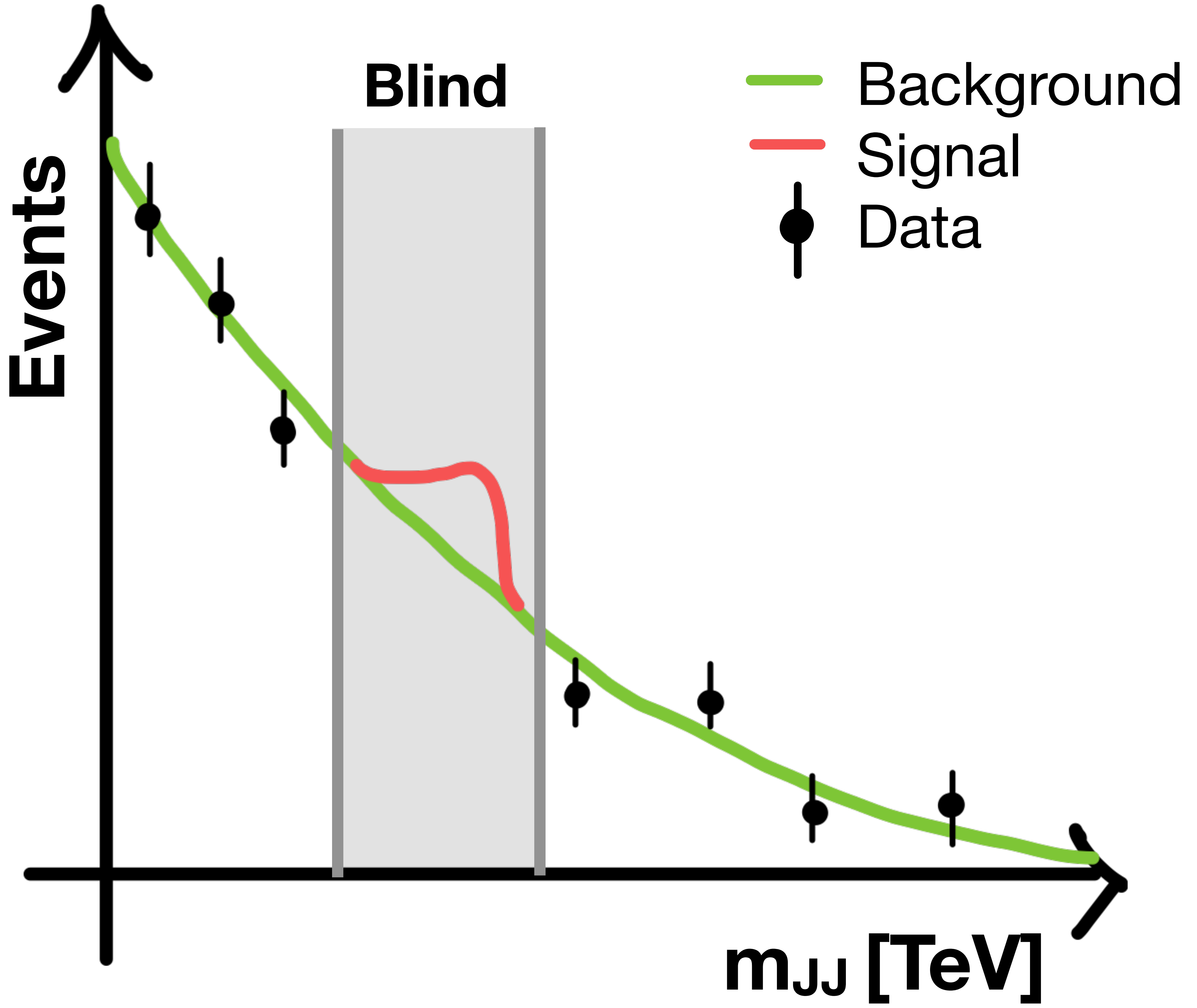




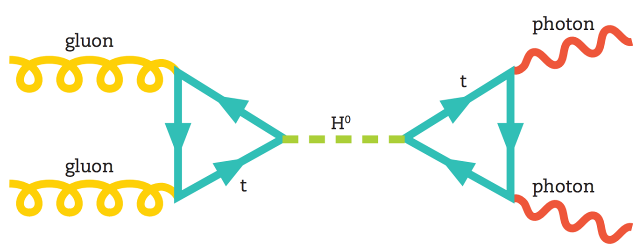


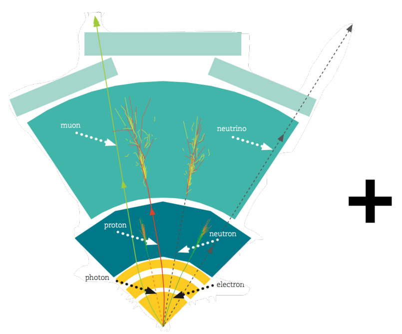
X→WW→JJ Event Selection

- Region of phase space with the most signal: **signal region**
  - ➡ Choose variable of the final discriminant
  - ➡ Check how background, signal and data behaves
  - ➡ Before all other steps are done: **blinding!**
    - Avoid bias when looking at the data
- X→WW→JJ signal region
  - ➡ Invariant mass of JJ: $m_{JJ} = \sqrt{(\sum \mathbf{E})^2 - |\sum \vec{\mathbf{p}}|^2}$
  - ➡ Look at m_{JJ} after all event selection from previous slide



4. Estimate Background Events

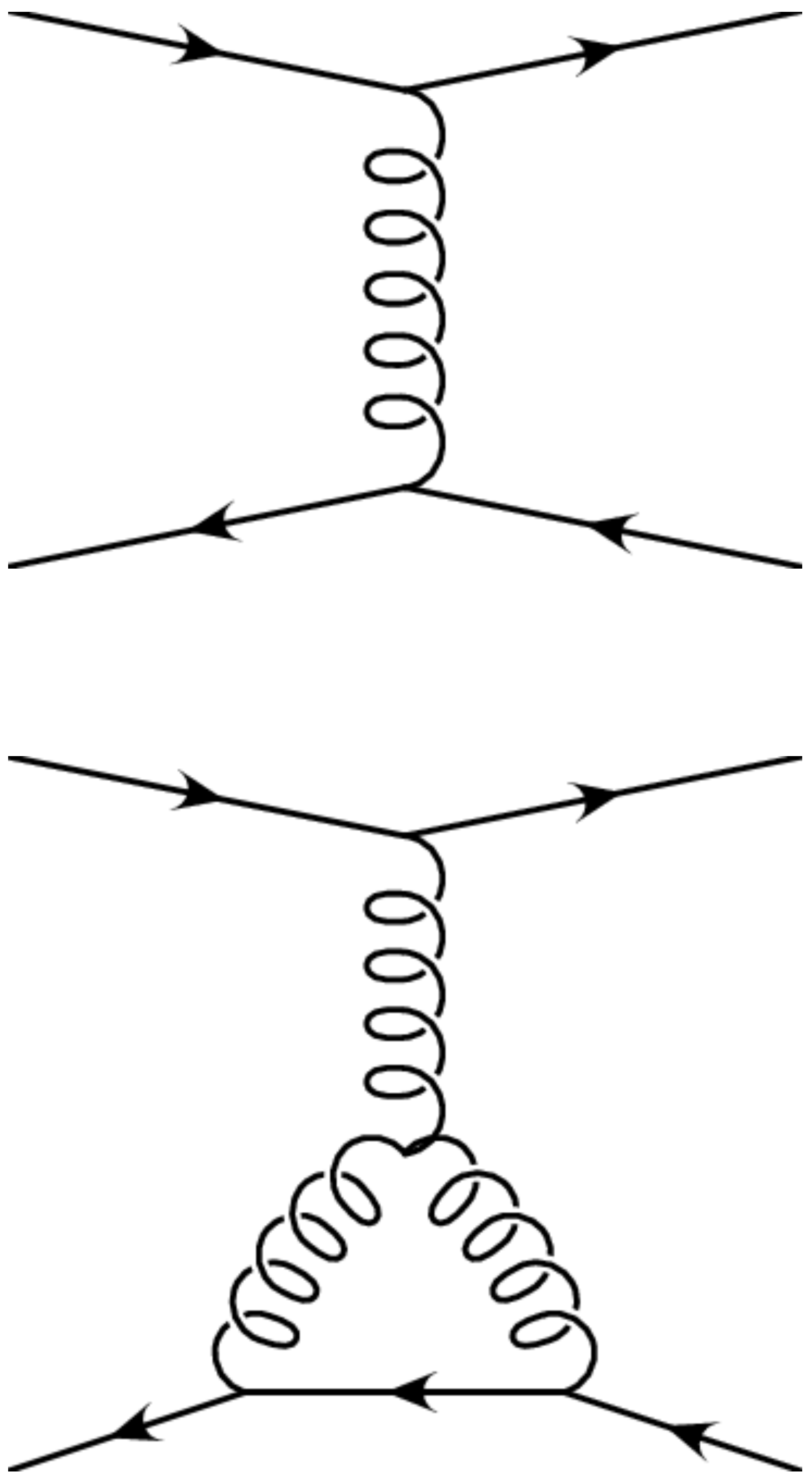
- You can't discover new physics without a good background estimate
- Background estimation techniques:
  - ➡ **Simulation/Monte-Carlo based**: 

+  + 
  - ➡ **Data-driven**: when backgrounds are too rare/hard to simulate

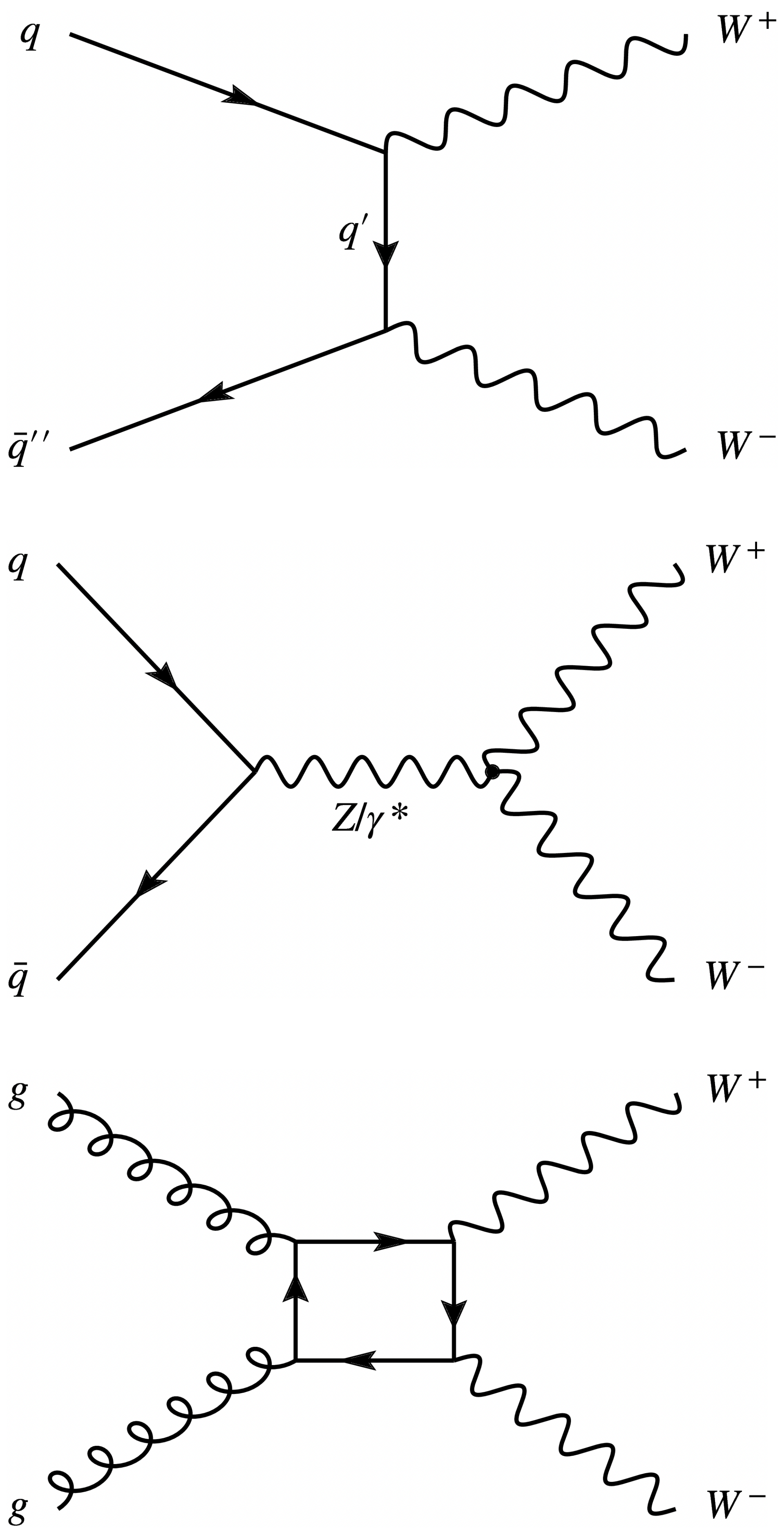
- Validation strategies:
  - ➡ **Control regions**: phase space depleted in signal but with similar kinematics to signal region
  - ➡ **Validation regions**: phase space less depleted in signal with closer kinematics to signal region

X→WW→JJ Backgrounds

SM QCD $q\bar{q}$ scattering



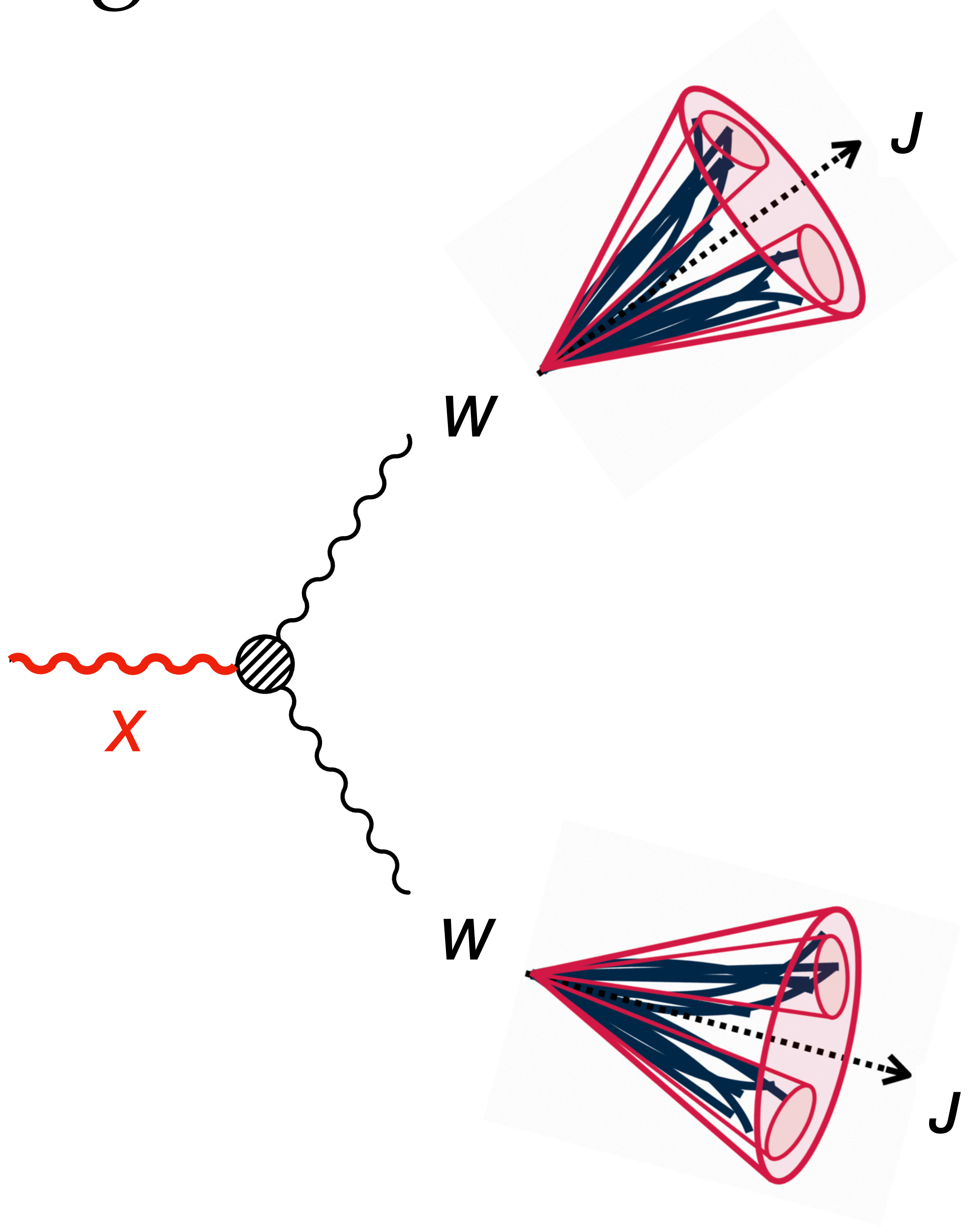
SM WW production

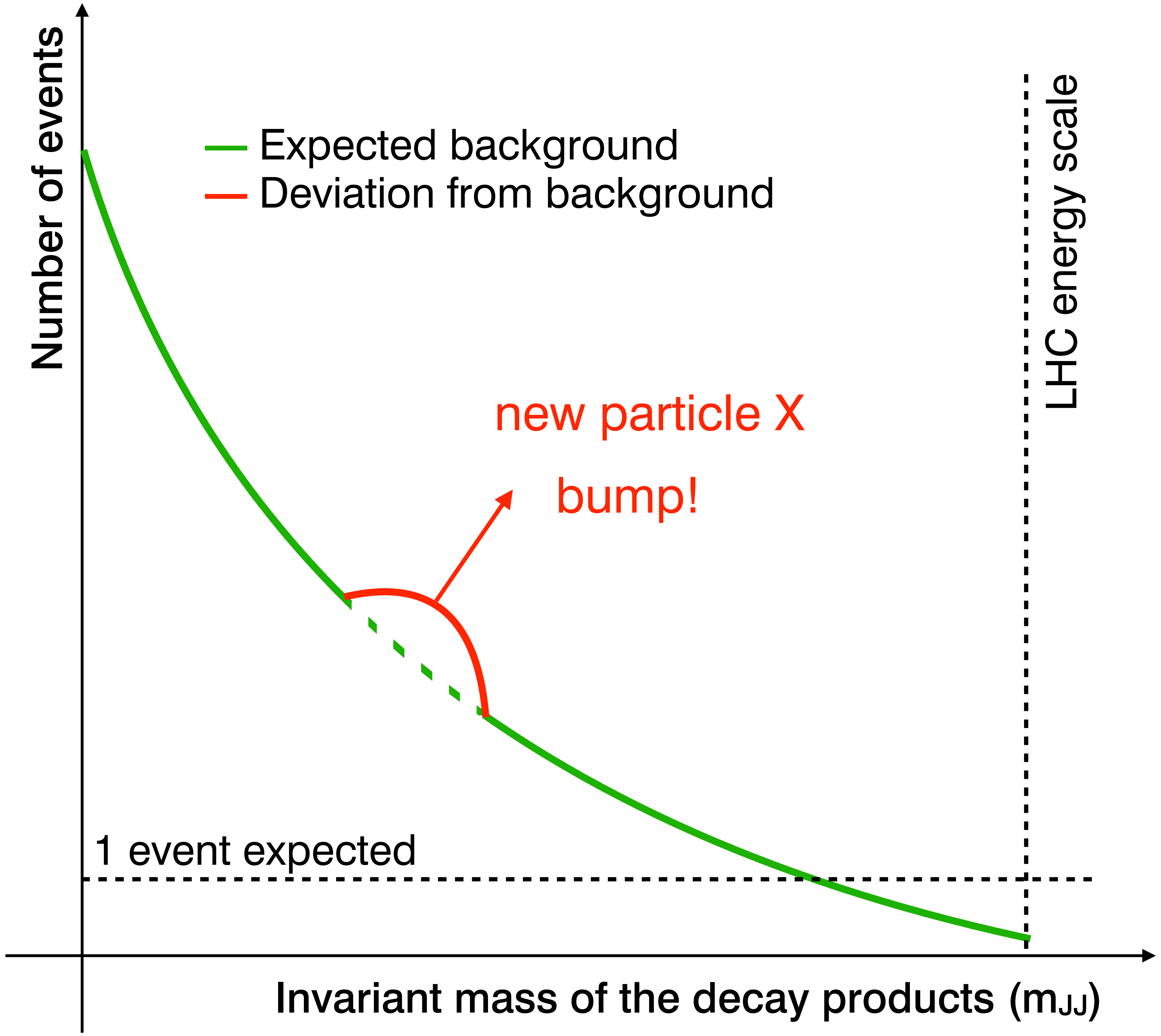


- Backgrounds from Standard Model are **non-resonant**
  - ➡ They don't make bumps
  - ➡ Signal is resonant and make bumps
- Backgrounds very hard to model using simulation
  - ➡ Data-driven approach

X→WW→JJ Background Strategy

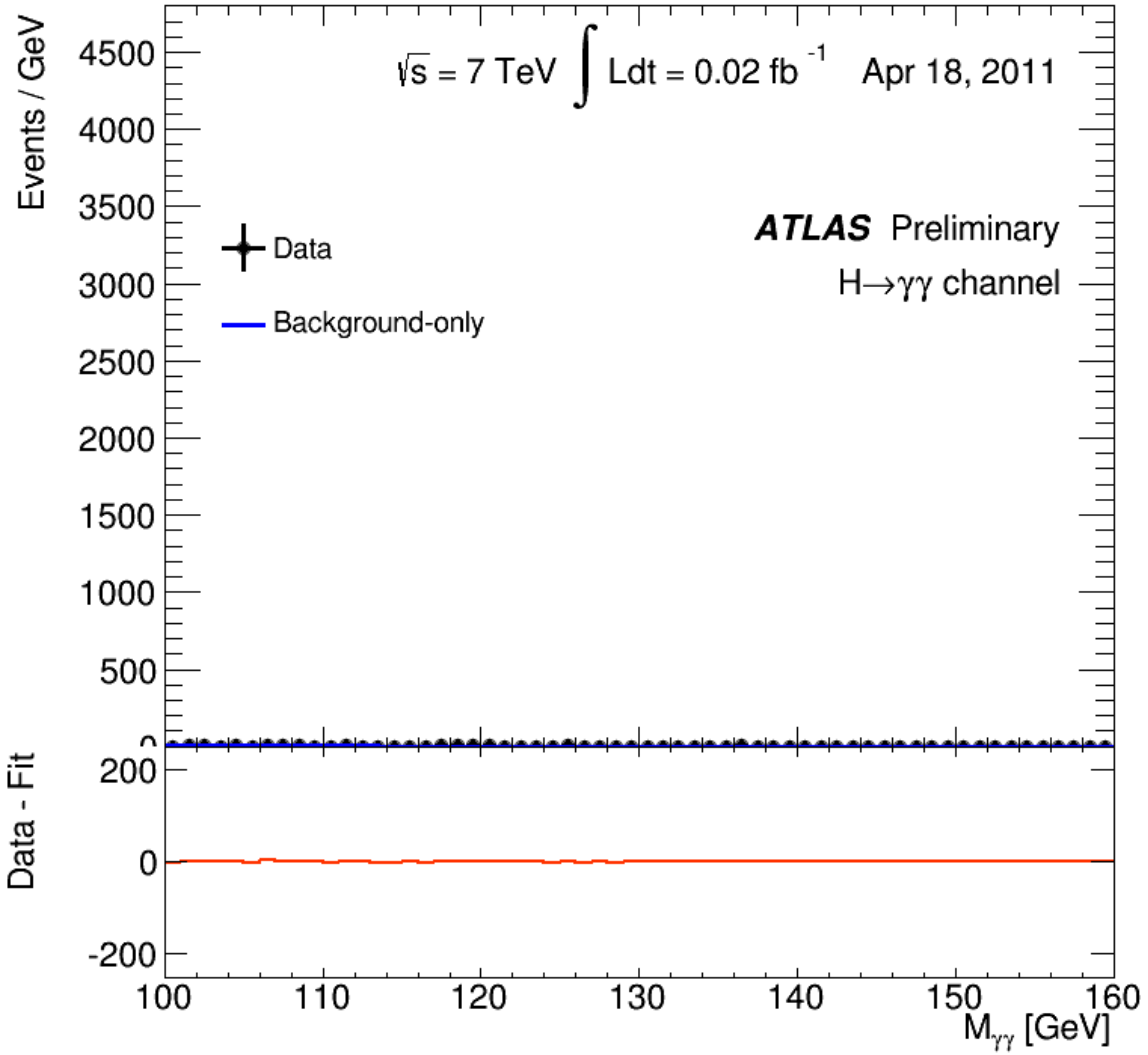
- Bump-hunt over a smoothly falling background





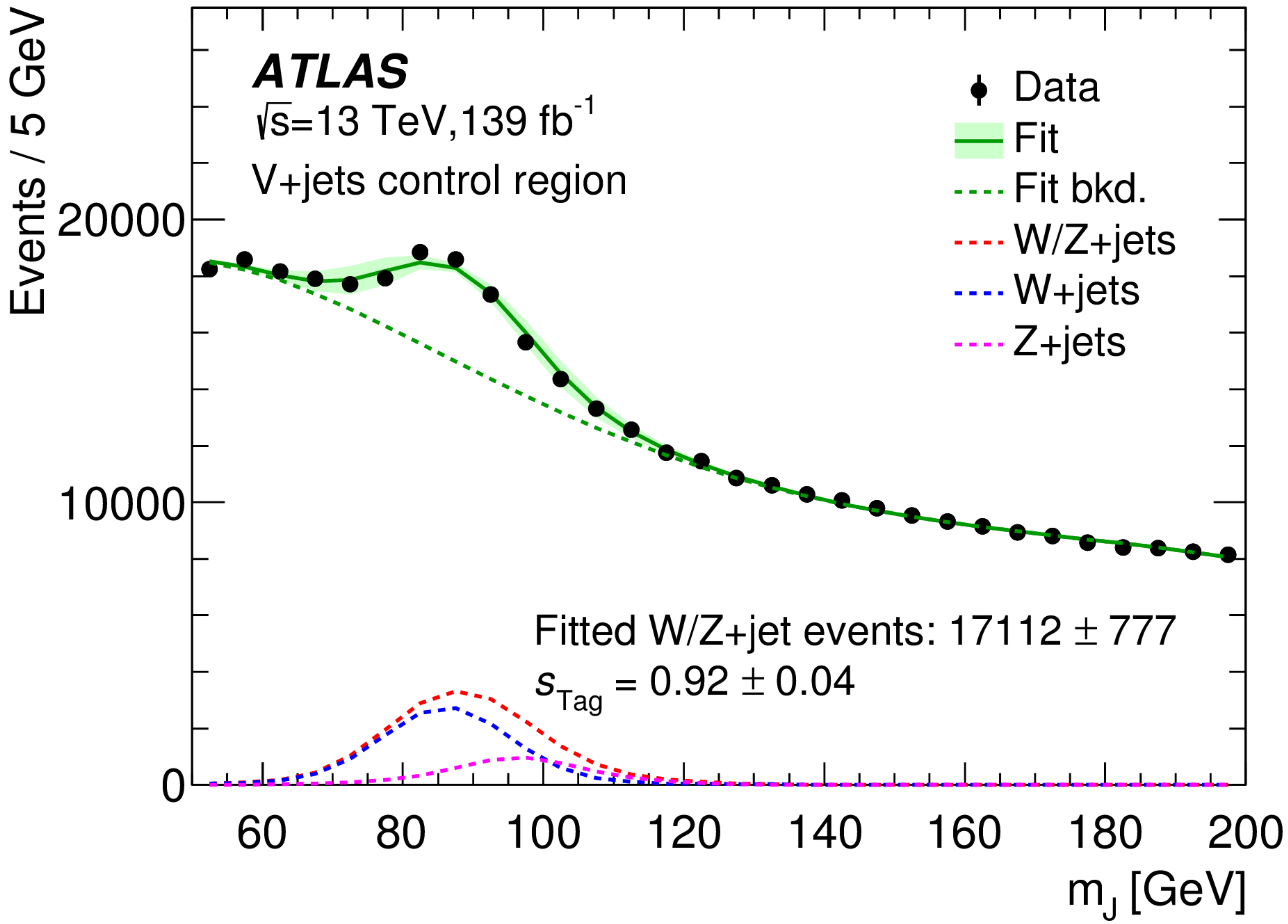
X→WW→JJ Background Strategy

- Bump-hunt over a smoothly falling background
- Famous example: **Higgs boson** decays to photons
- Plot from Run-1 with the data used for discovery



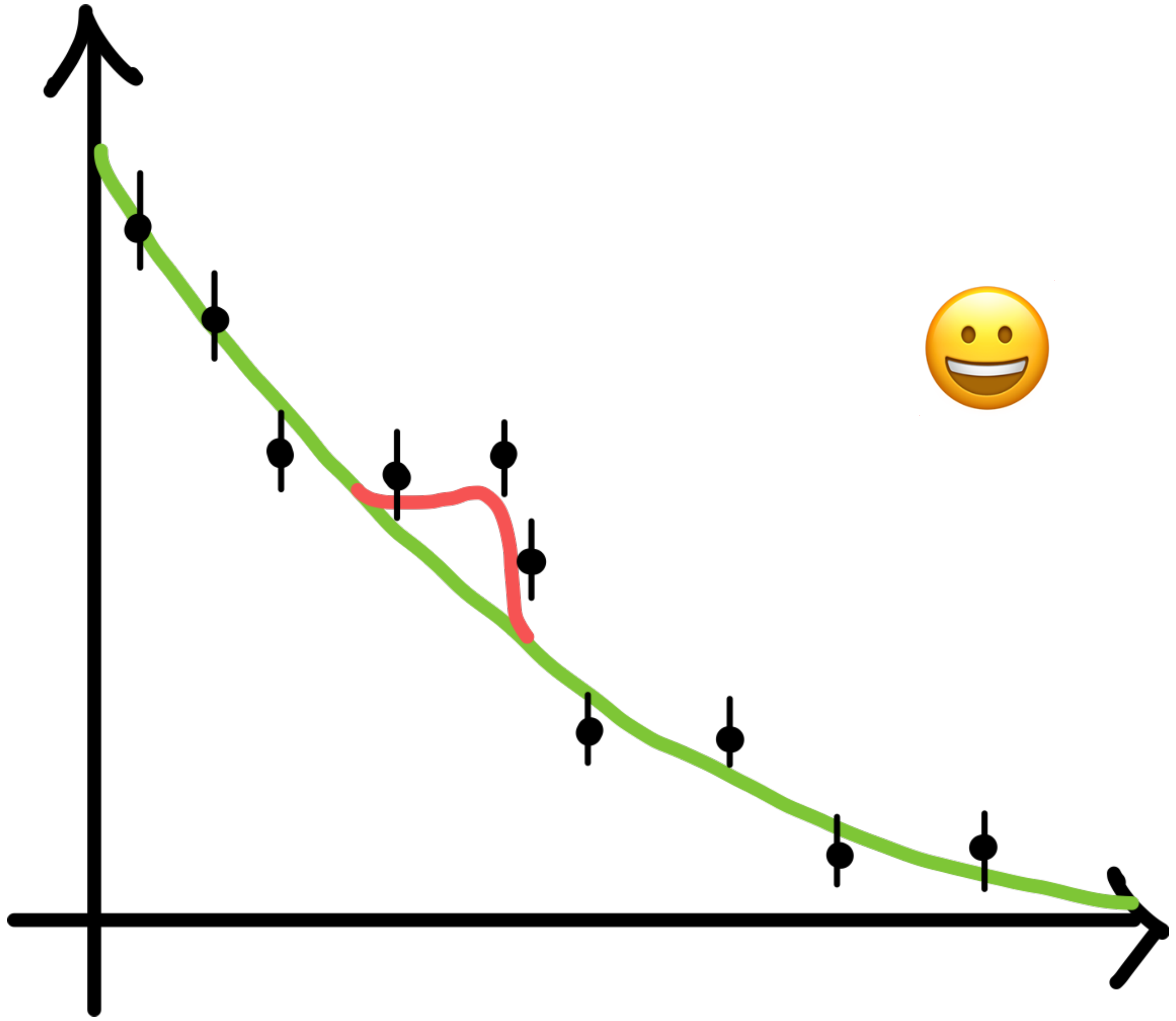
X→WW→JJ Background Strategy

- Validate our methods: measure known W/Z bosons in the same final state
- Use signal depleted **control region**



5. Estimate Uncertainties

- Arguably the hardest and most important part of an analysis!
- A number without an error is meaningless



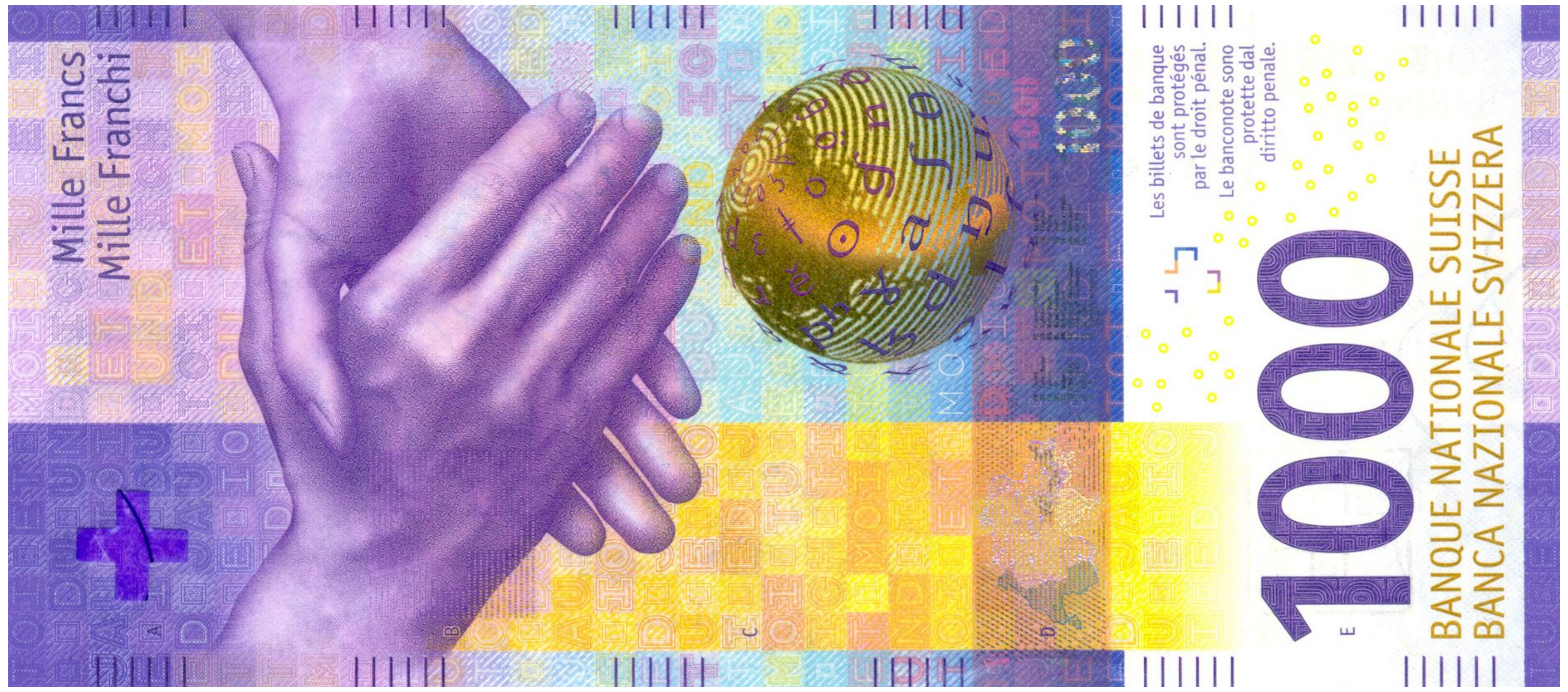
Statistical Uncertainties

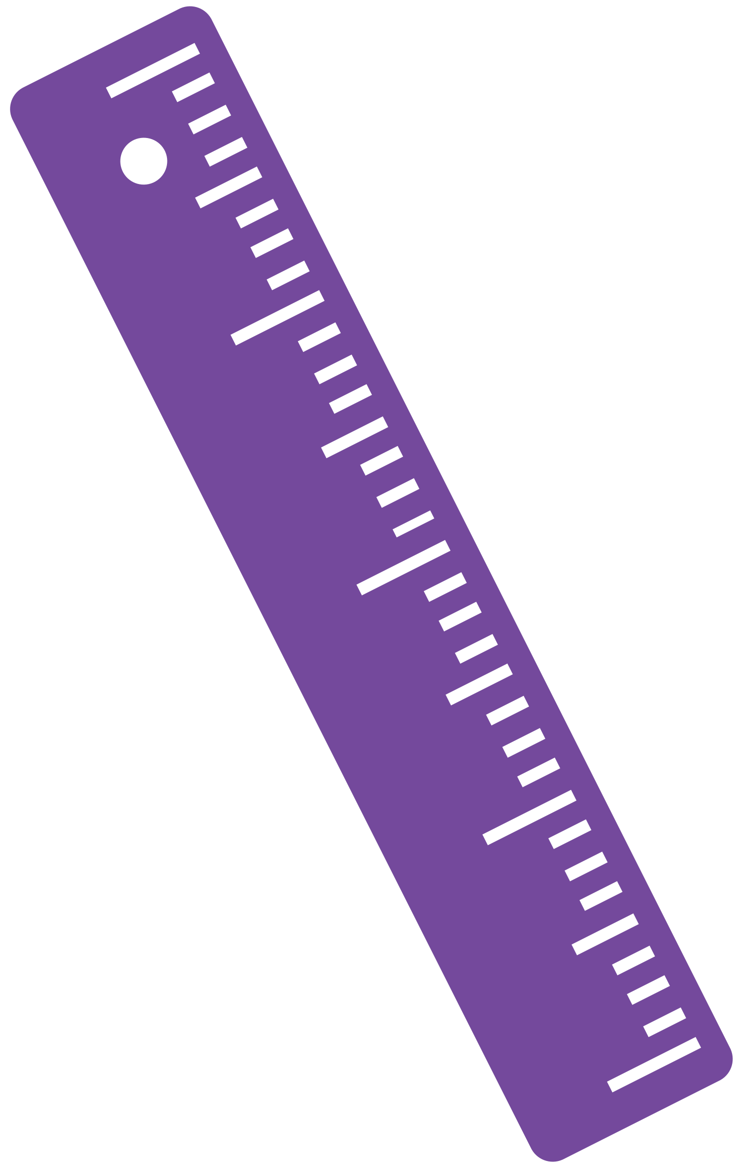
- From **stochastic fluctuations** arising from the fact that a measurement is based on a **finite set of observations**
  - ➡ Example: toss a coin; Is it heads or tails?

- Repeated measurements will give a set of observations different from each other
  - ➡ The statistical uncertainties are a measure of this variation
  - ➡ Calculated as Poisson fluctuations associated with random variations on the system one is examining

Systematic Uncertainties

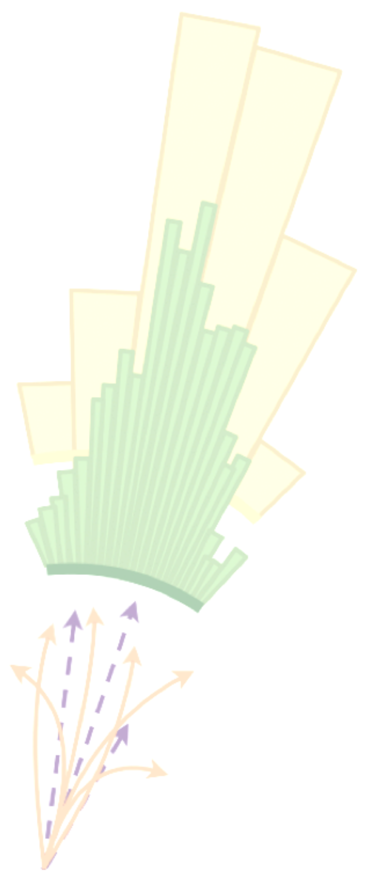
- Uncertainties associated with the measuring apparatus
  - ➡ Measuring the size of a 1000 CHF note:



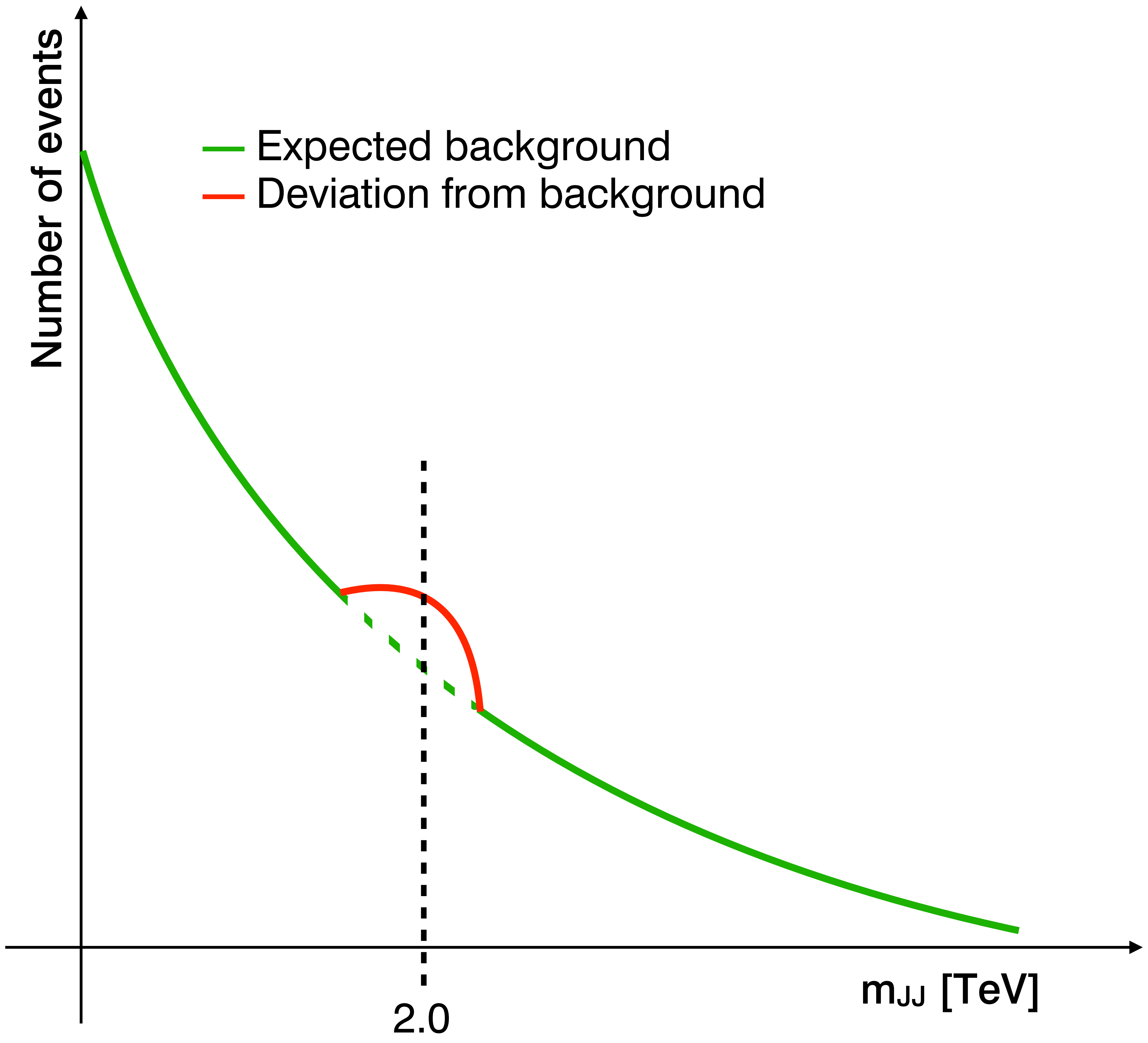


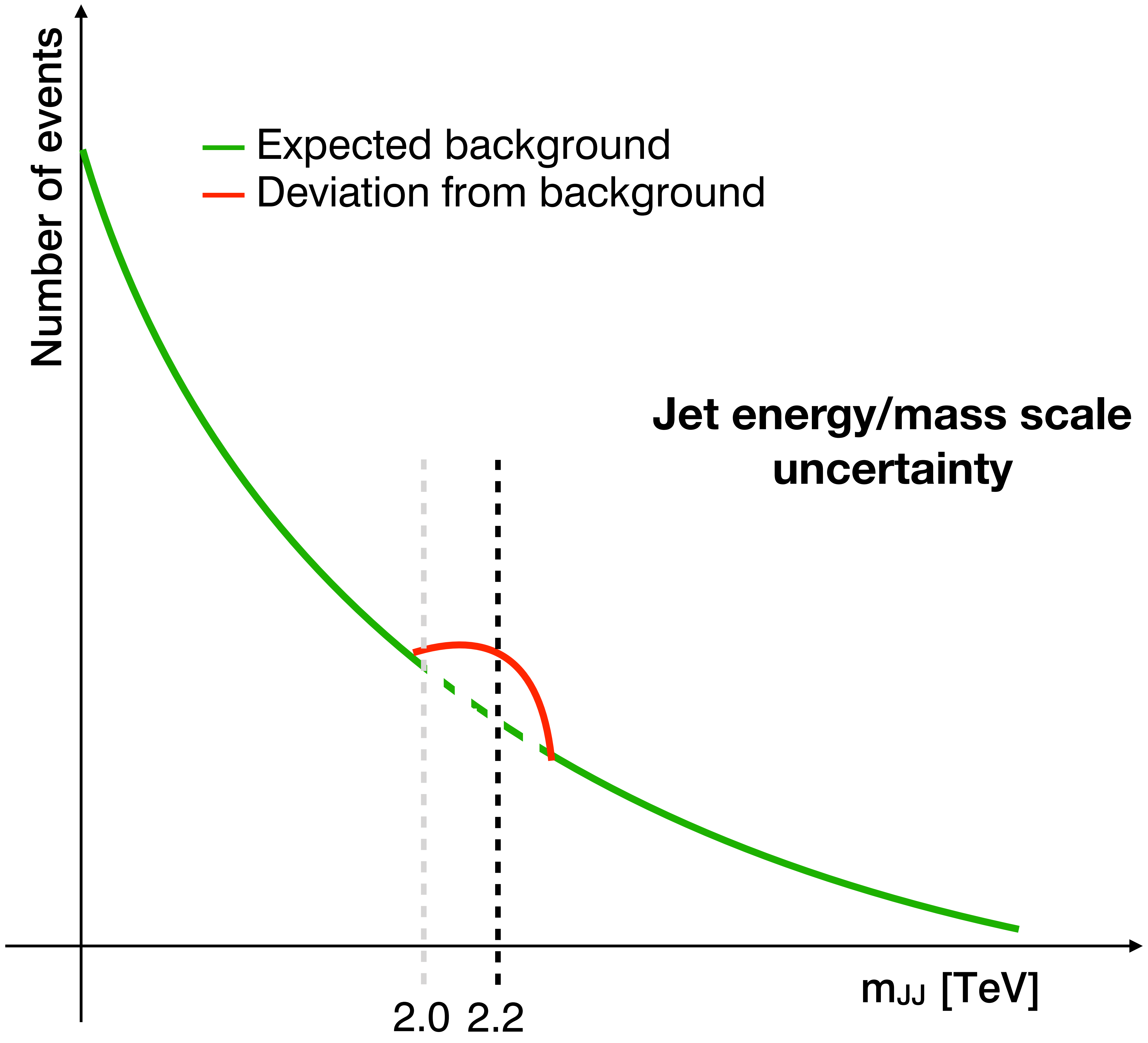
Systematic Uncertainties

- What are the assumptions underlying the measurement?
  - ➡ How accurate is your Monte Carlo simulation of your theory (Feynman diagrams)?
  - ➡ How precise are the models for your signal and background?
  - ➡ How well do you model how often your jets go outside the detector acceptance?
  - ➡ How well do you measure the jets themselves?

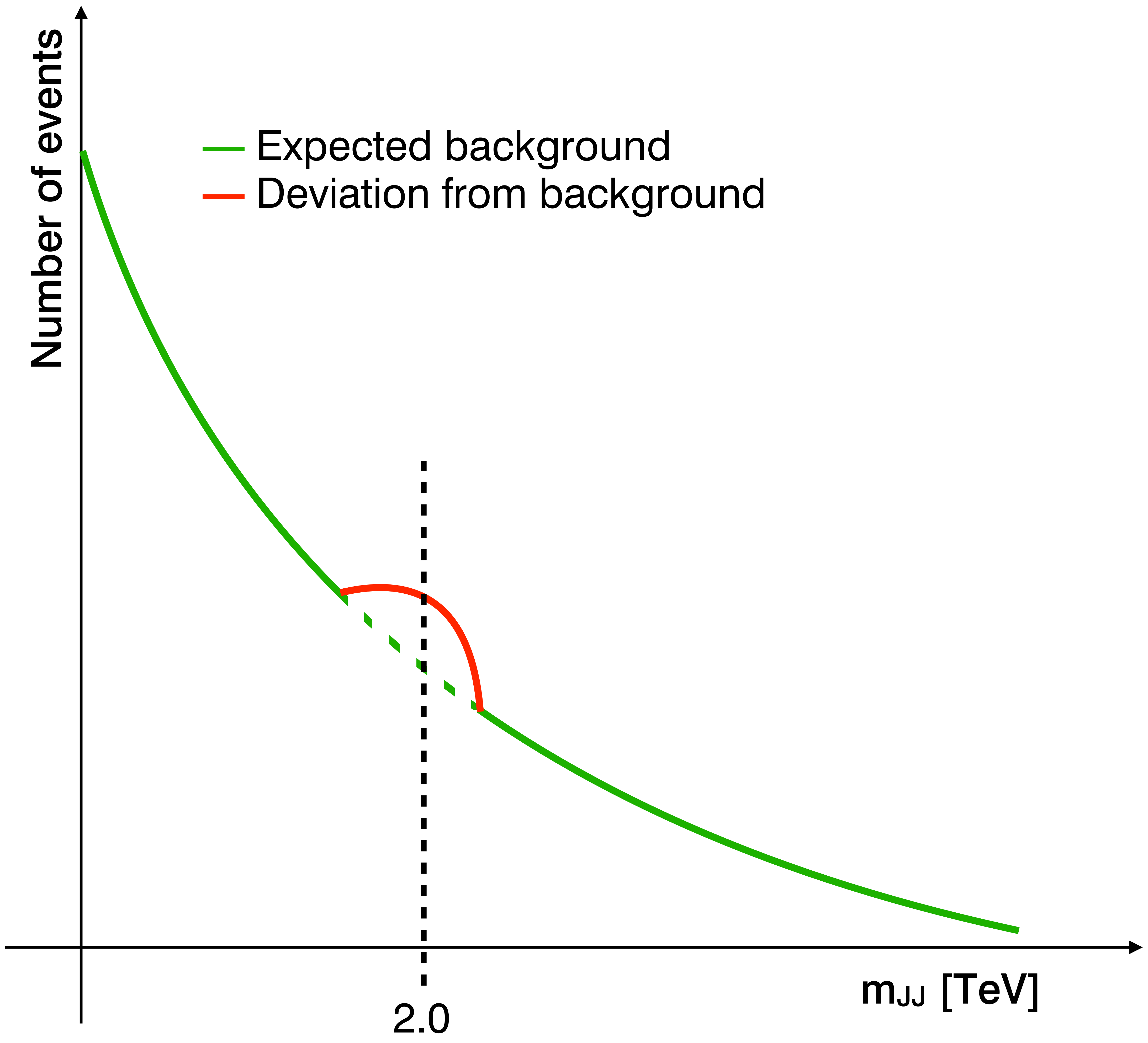


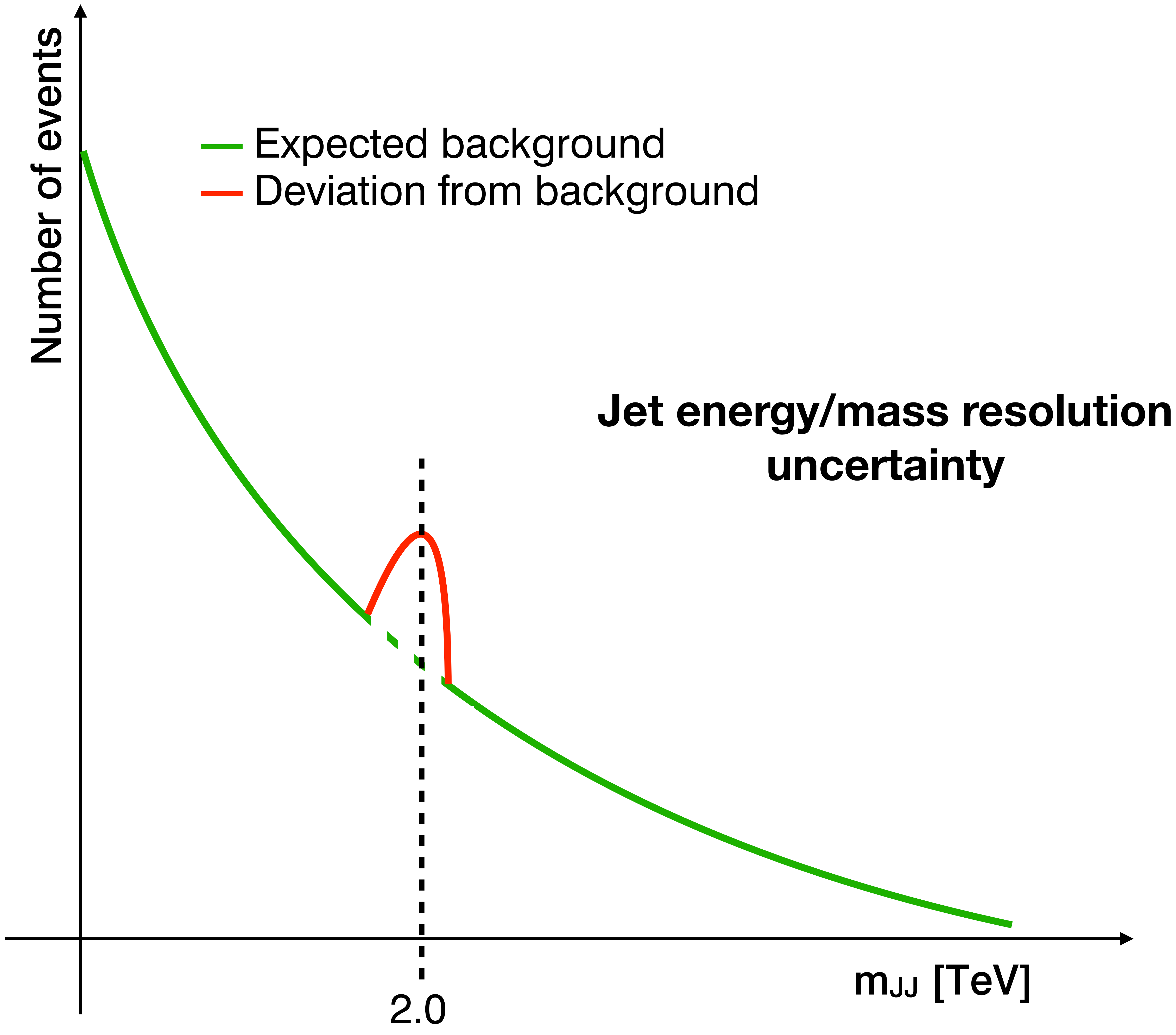
X→WW→JJ Uncertainties





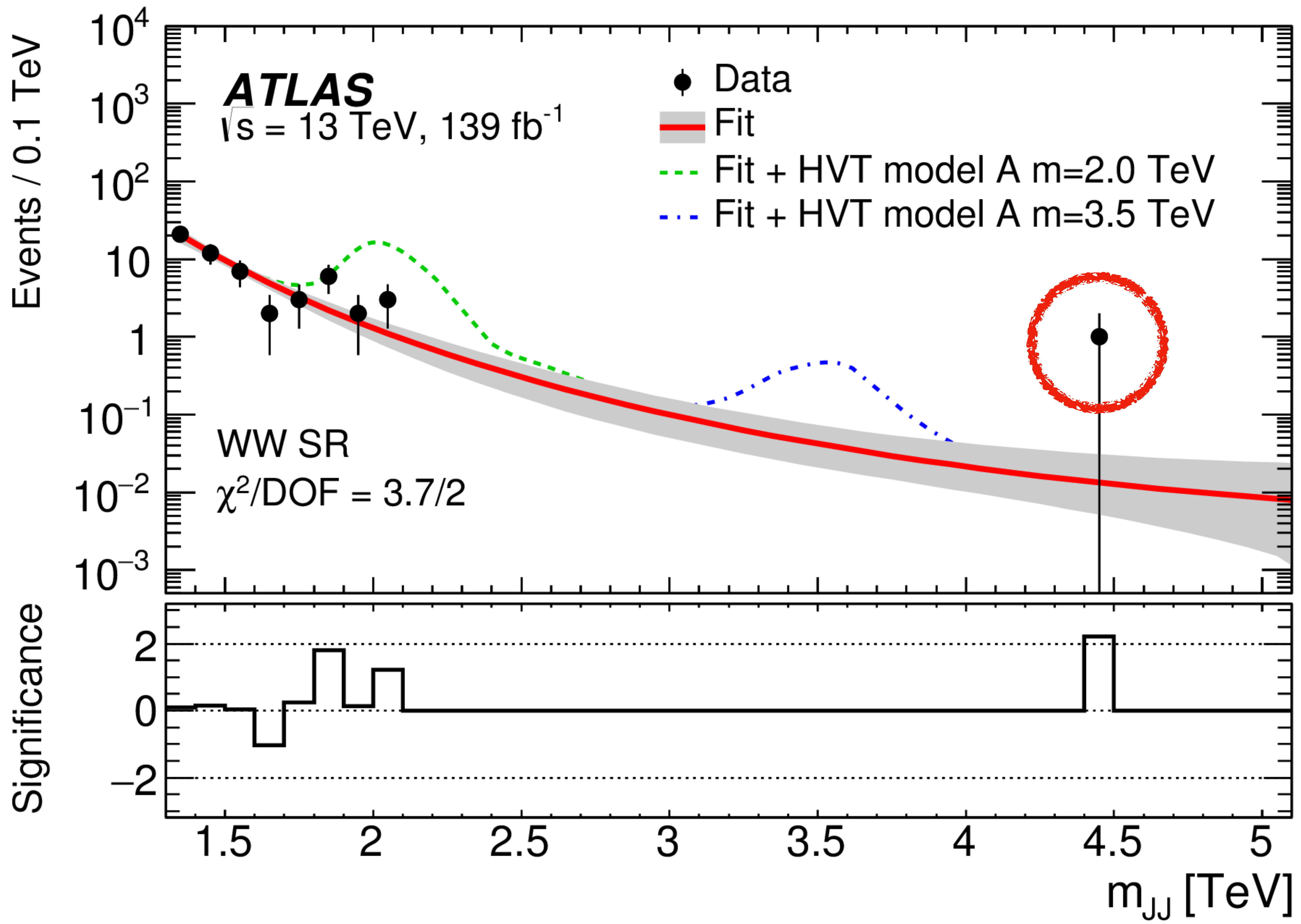
X→WW→JJ Uncertainties



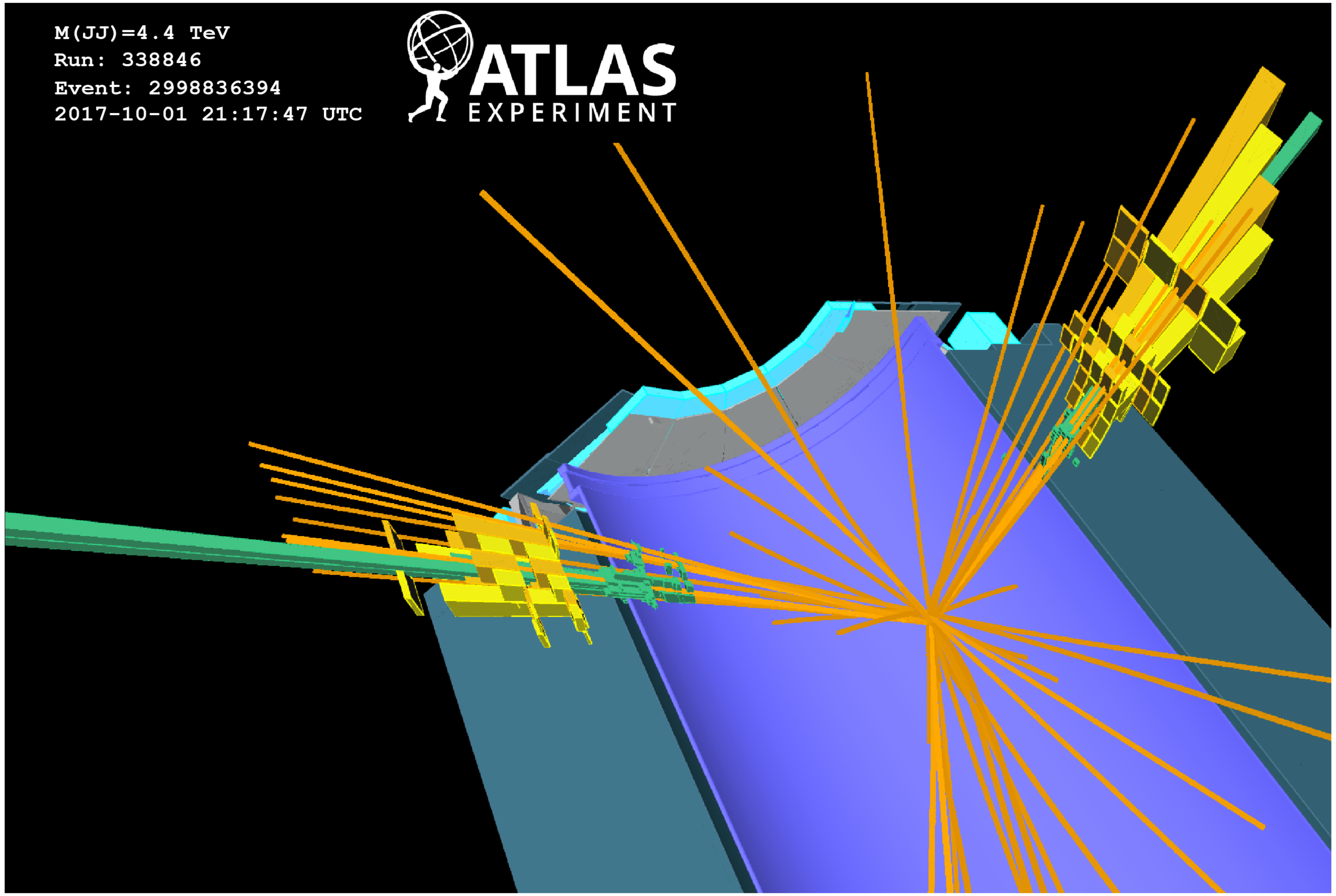


6. Plot Observables of Interest

- Time to look into the **unblinded** data!
  - ➡ Fit background
  - ➡ Check if the background is consistent with the observed data



X→WW→JJ Results

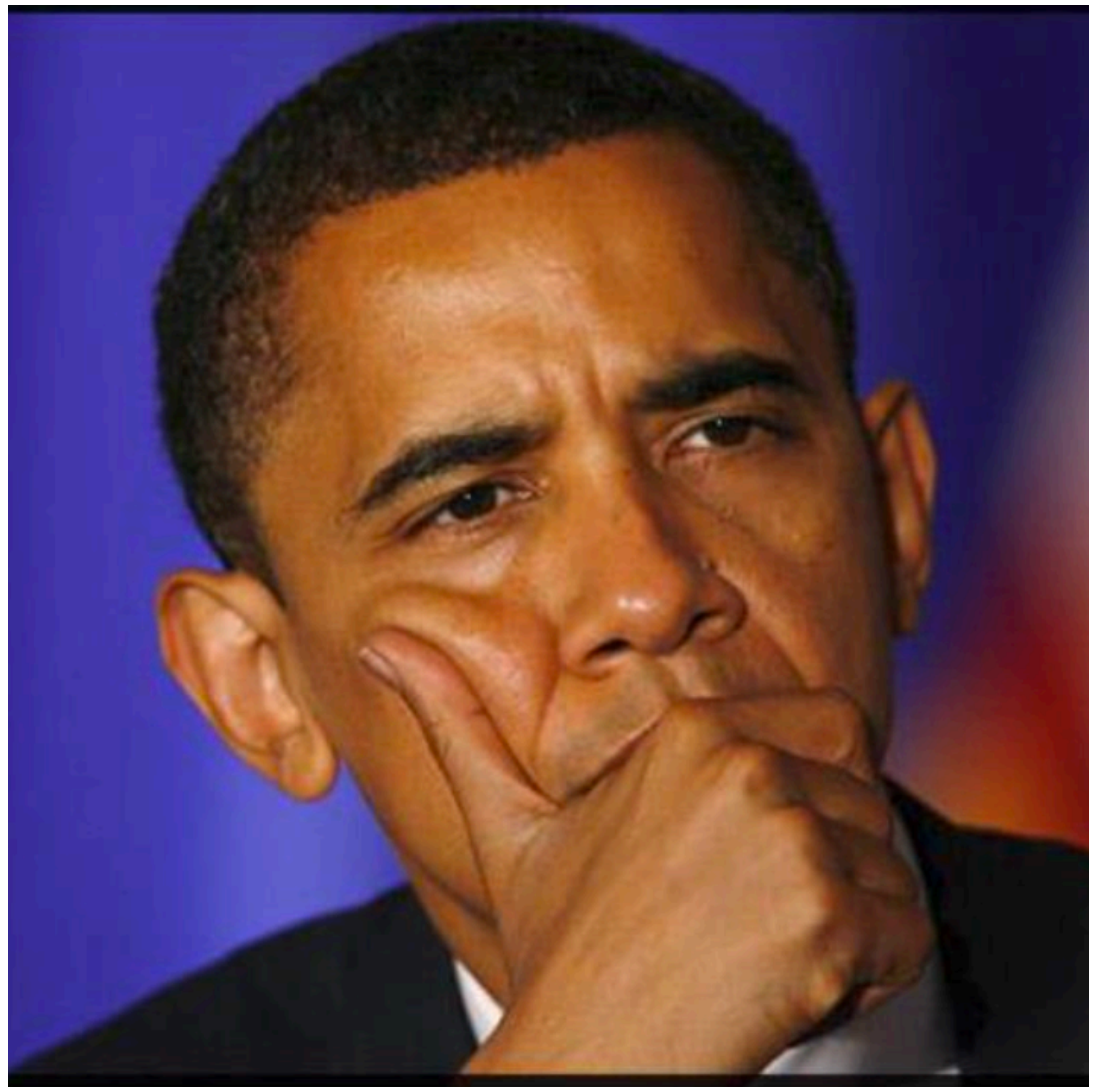


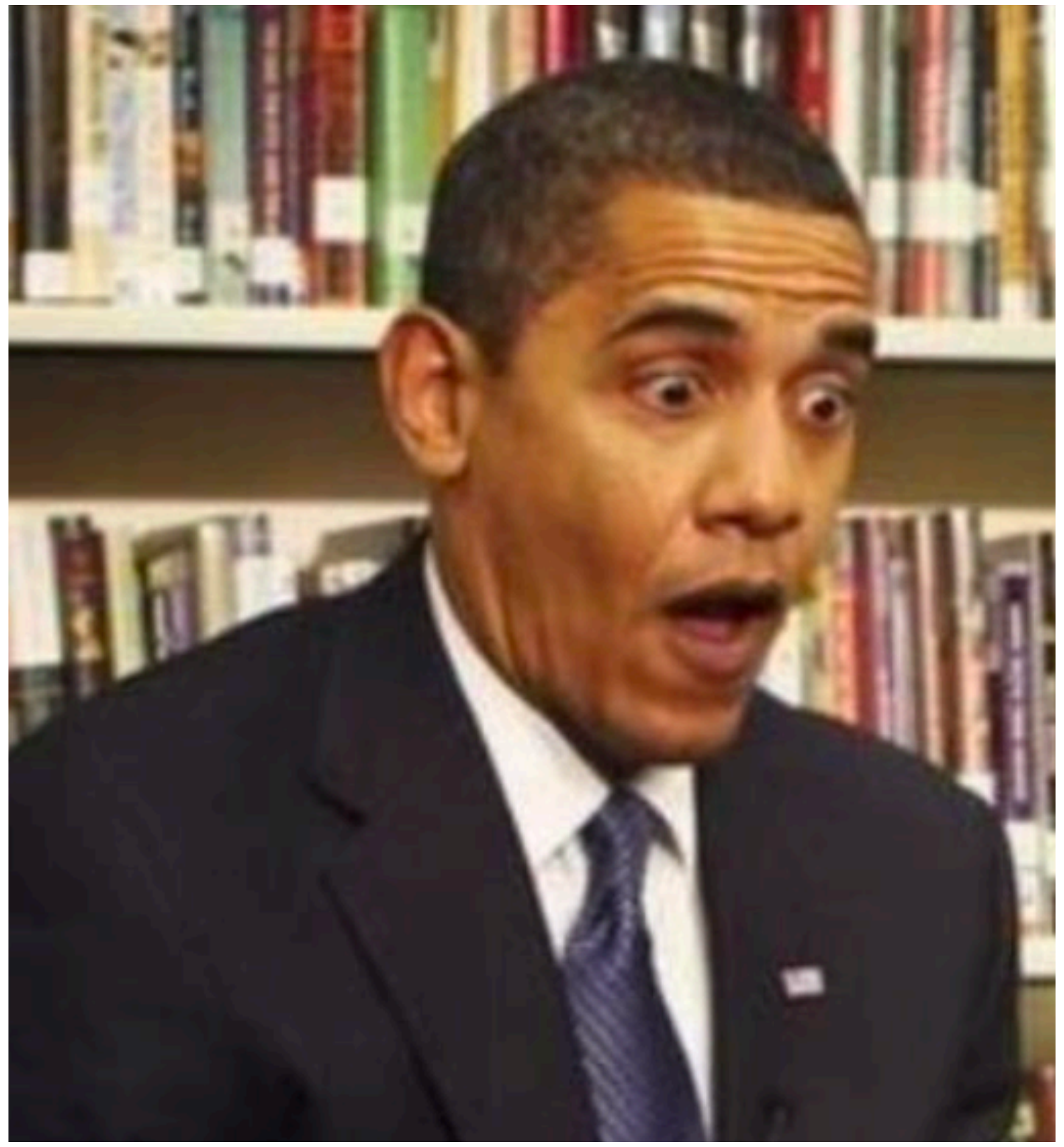
- m_{JJ} = 4.4 TeV
  - ➡ J_1: p_T=2.1 TeV, m_J=89 GeV
  - ➡ J_2: p_T=2.2 TeV, m_J=62.6 GeV
- Is one event enough to claim a discovery?
  - ➡ **NO!**

7. Statistical Analysis

- Probabilistic nature of particle physics: need to **accumulate data**
  - ➡ You can never tell from one even what was the process that caused it (even if it looks **a lot** like your signal)
- Estimate p-value / significance of observed events
  - ➡ **p-value**: compatibility with **background-only** hypothesis
    How likely the null-hypothesis is to explain my data?
    - High p-value (~1): nothing new in data
    - Very low p-value (0.00000035): discovery!
  - ➡ **Significance**: statistical measure of the strength of evidence for a particular observation
    Number of standard deviations σ that data differs from background

How many sigmas?



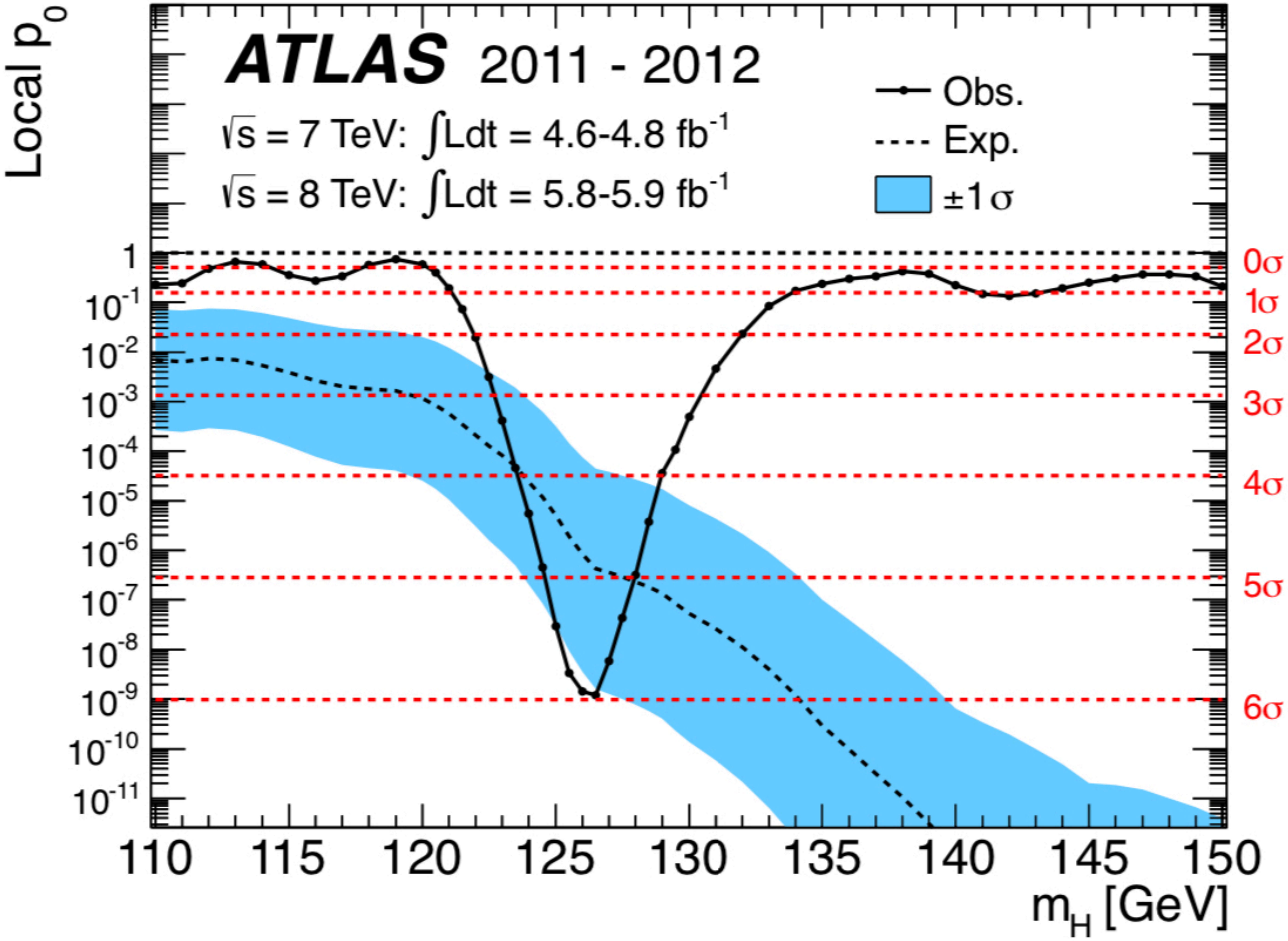


1-2 σ
1 in 3 times
1 in 22 times

3 σ
1 in 370 times
Hint/evidence

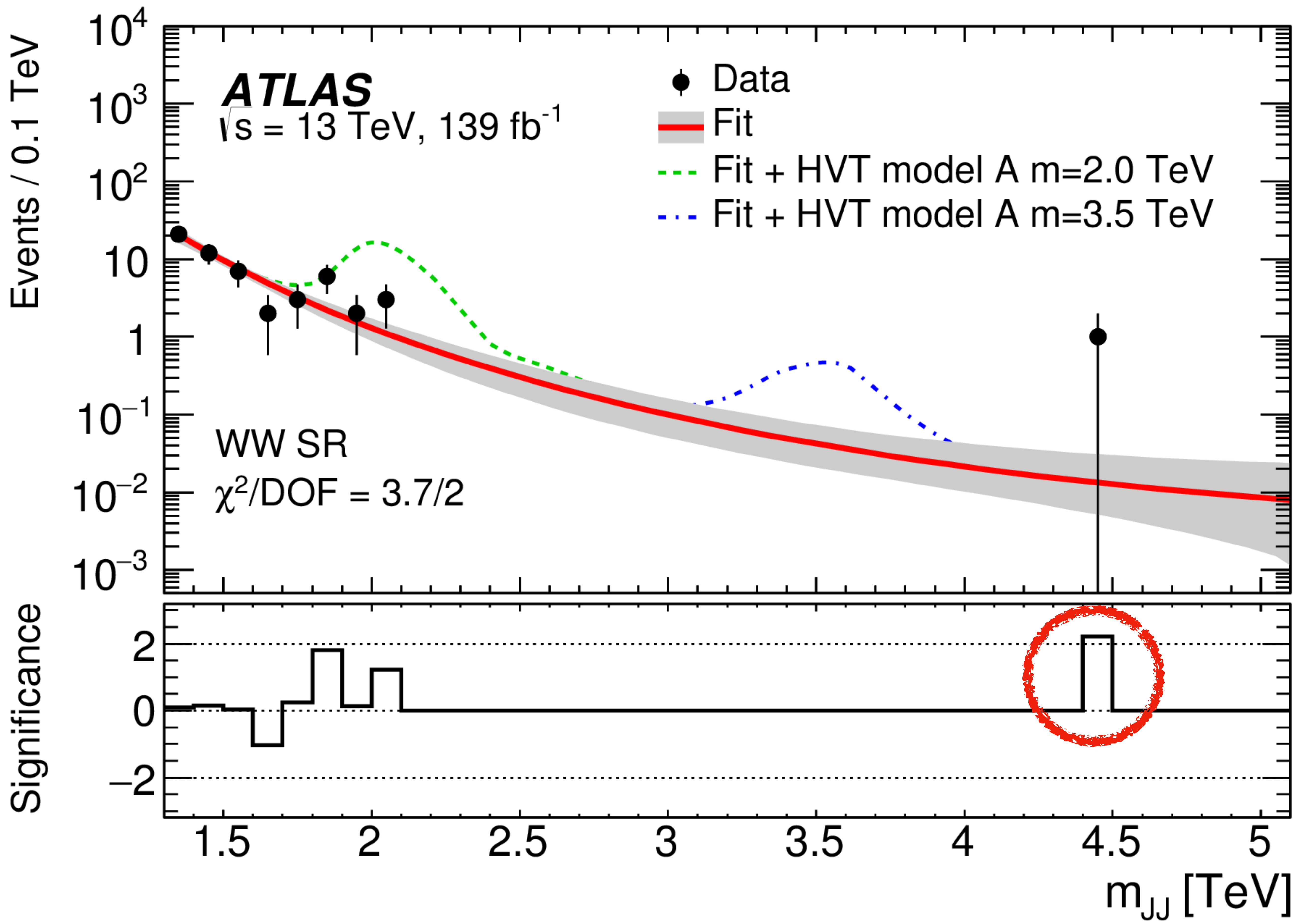
5 σ
1 in 3.5 million times
Discovery

How many sigmas? Higgs discovery

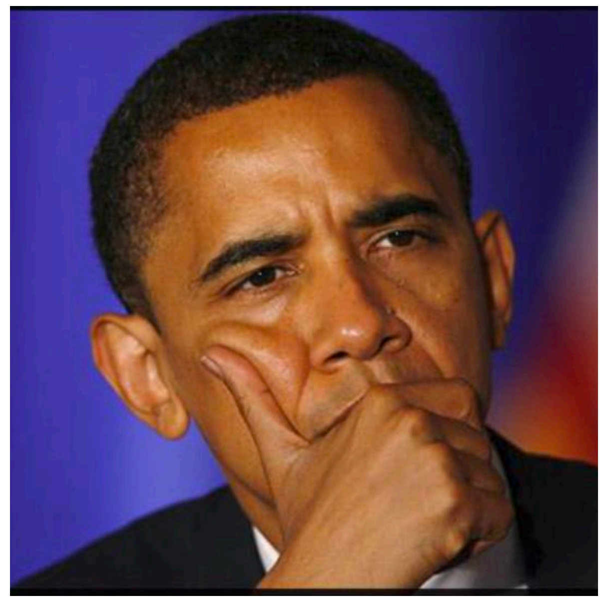


~6σ

How many sigmas: X→WW→JJ at 4.4 TeV



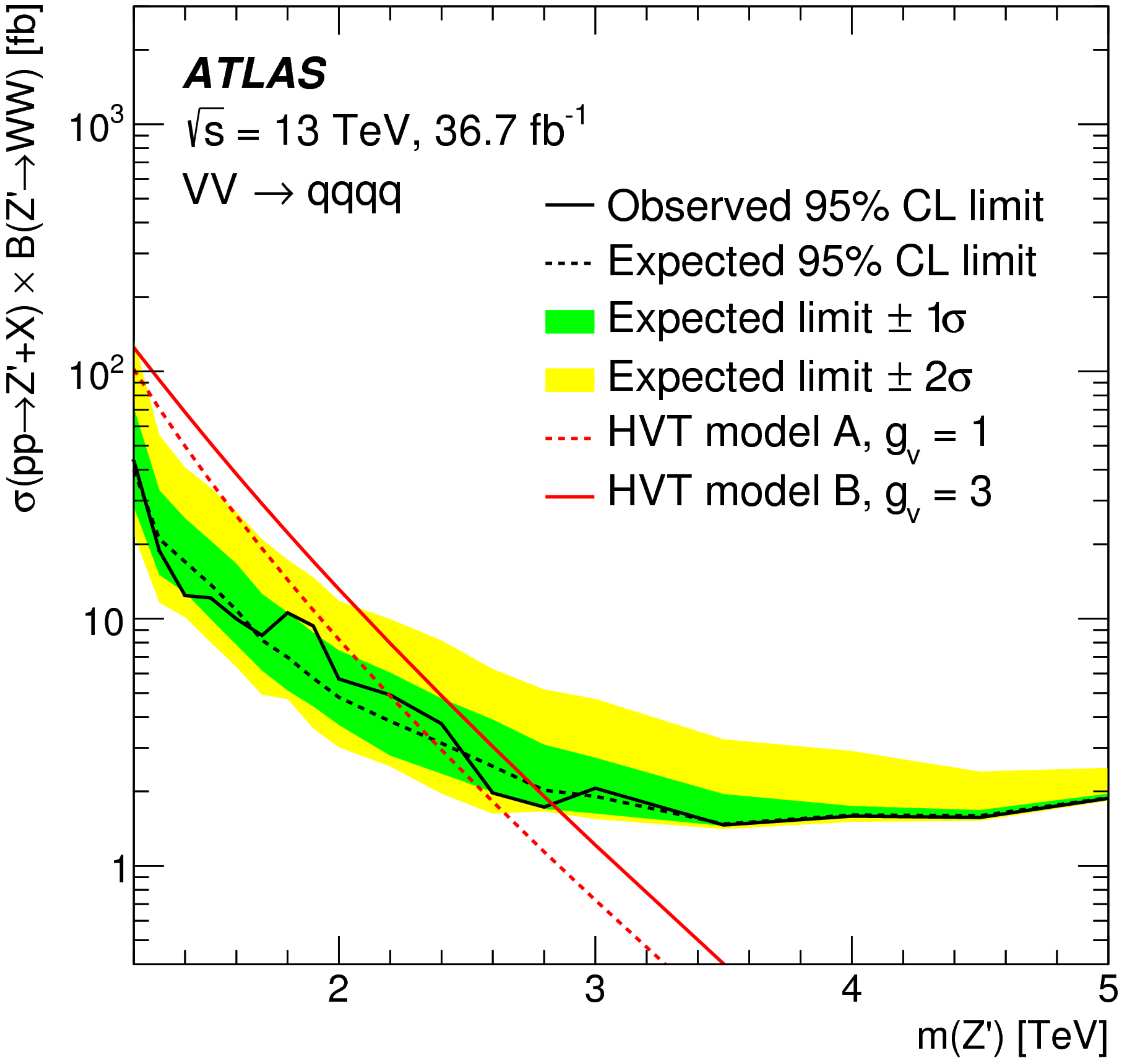
Just about 2σ



Statistical Analysis: Final Result

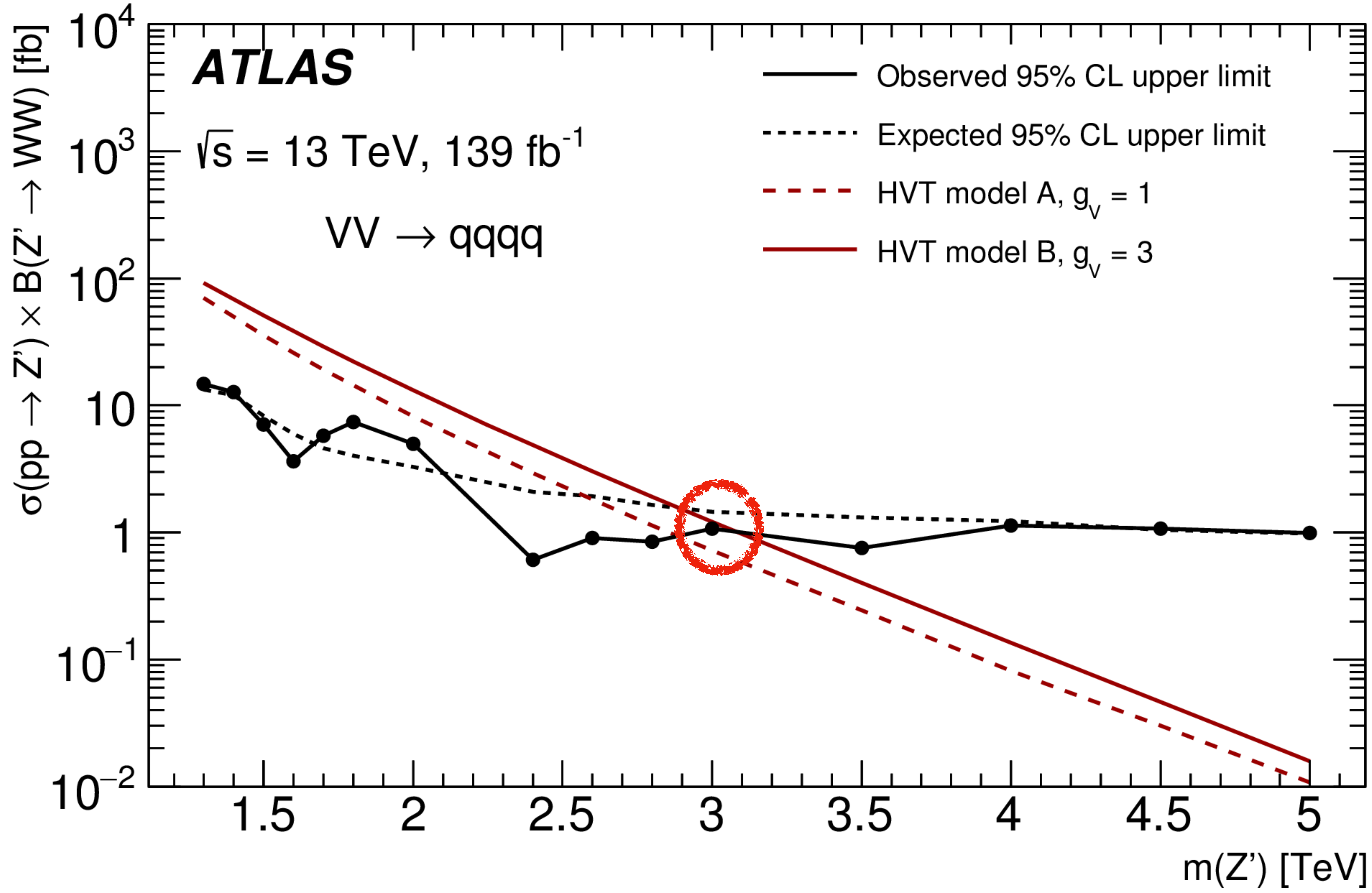
- In a search without a new particle: 95% C.L. upper limits on cross section x branching ratio

🇧🇷 **Brazil band plot**



Statistical Analysis: Final Result

- In a search: 95% C.L. upper limits on cross section x branching ratio



Exclusion mass in given model: ~3 TeV

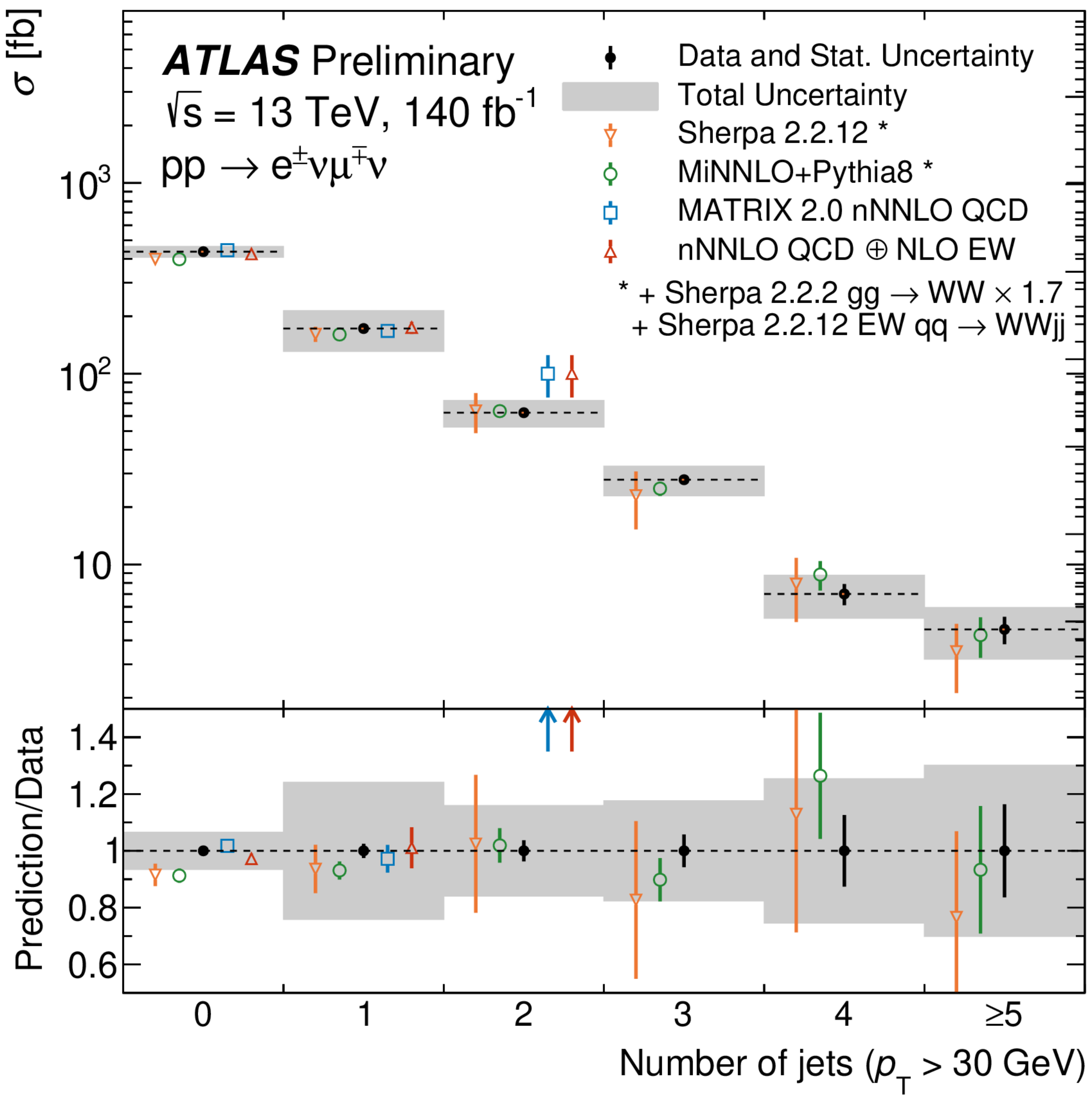
Nikhef UvA

Statistical Analysis: Final Result

- In measurements: cross section number with uncertainties, or a differential result (in many bins of a variable)

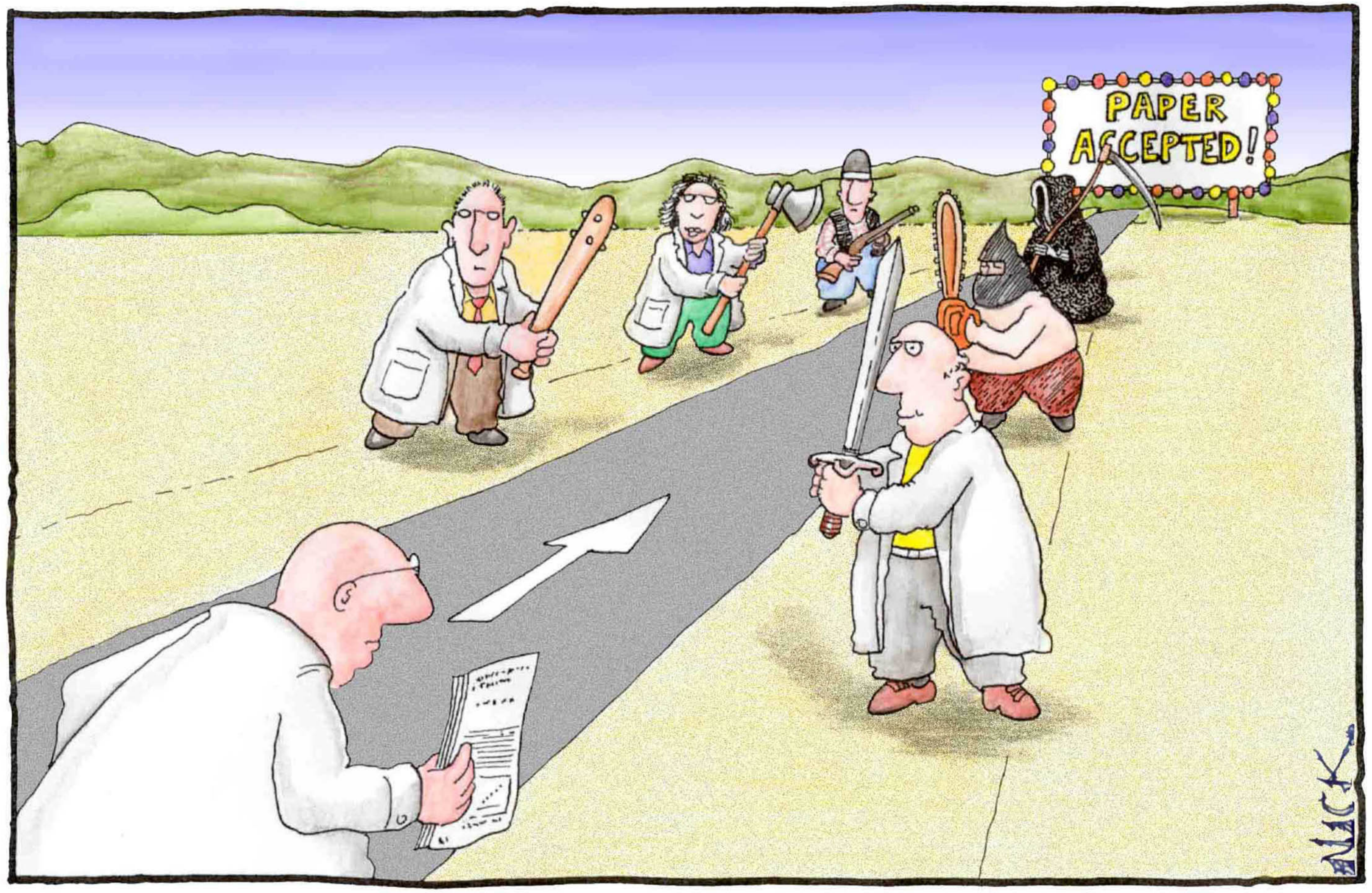
- SM W^+W^- cross section:
  (ATLAS-CONF-2023-12)

➡ 127 ± 4 fb



8. Peer Review

- Definition: *"The process by which scholars assess the quality and accuracy of one another's research paper"*
  - ➡ Quality assurance
  - ➡ Validity and reliability
  - ➡ Enhancing research: constructive feedback (except reviewer #2)
  - ➡ Facilitating communication and collaboration



Most scientists regarded the new streamlined peer-review process as 'quite an improvement.'

ATLAS Review

- ATLAS Collaboration: ~3000 scientific authors, 182 institutions from 42 countries

Everyone needs to agree with your result!

Published Paper!

- Spread your new scientific results to the world!
  - ➡ New measurements as inputs to theory and other experimentalists
  - ➡ Compare results across experiments
  - ➡ Important to report null results as well
- Other relevant things:
  - ➡ Open access
  - ➡ Open data

Nikhef

Thank You!
Questions?

University of Amsterdam
Institute of Physics

ATLAS
EXPERIMENT