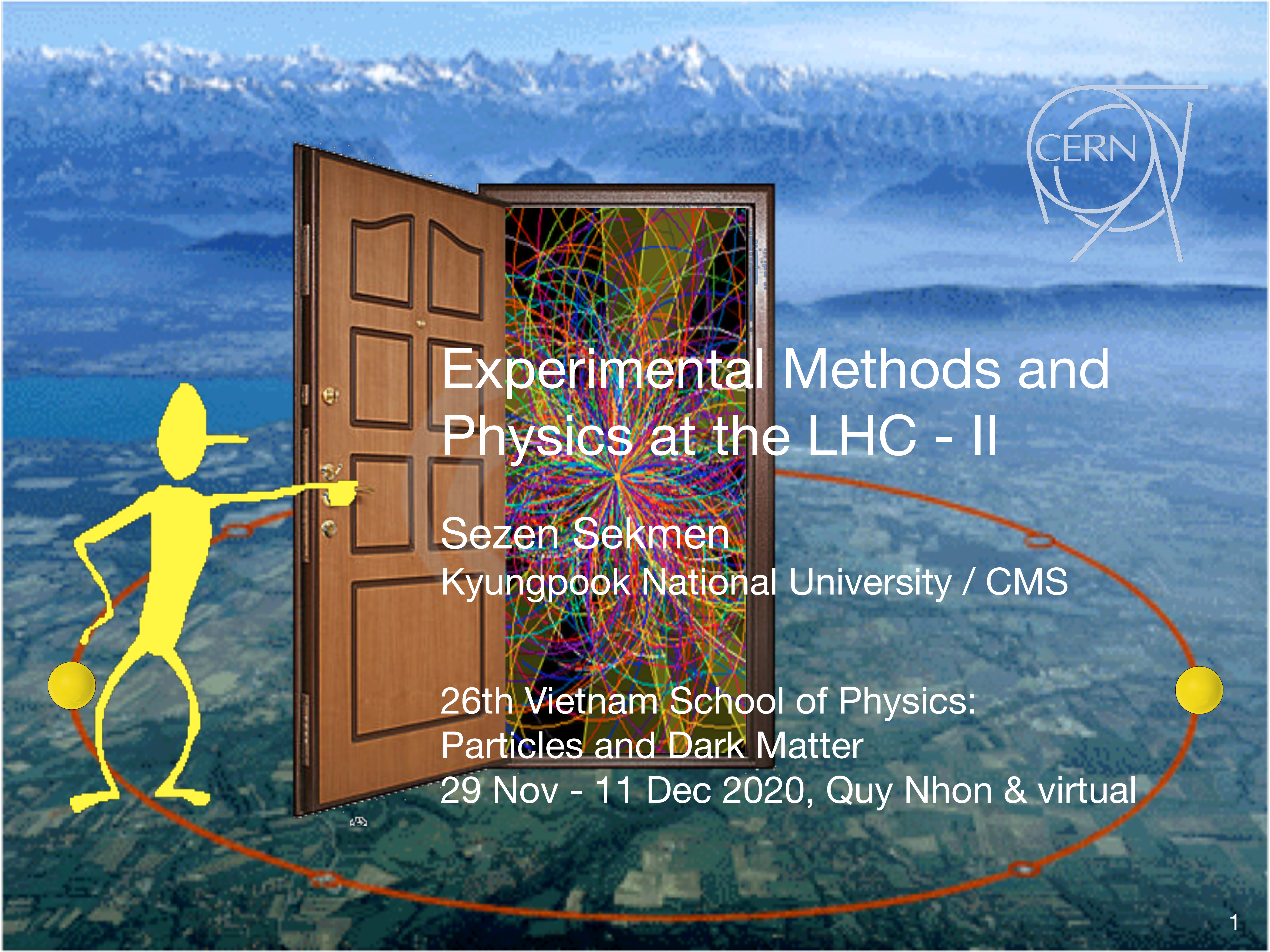


Experimental Methods and Physics at the LHC - II

Sezen Sekmen
Kyungpook National University / CMS

26th Vietnam School of Physics:
Particles and Dark Matter
29 Nov - 11 Dec 2020, Quy Nhon & virtual

Disclaimer

About me: I am a physicist working in the CMS Collaboration. I have been involved in CMS data analysis since the beginning of data taking, in 2009.

Most of the plots in these lectures are taken from CMS, as it was easier for me to find recent versions in a very limited time. However please note that similar plots have been produced by ATLAS.

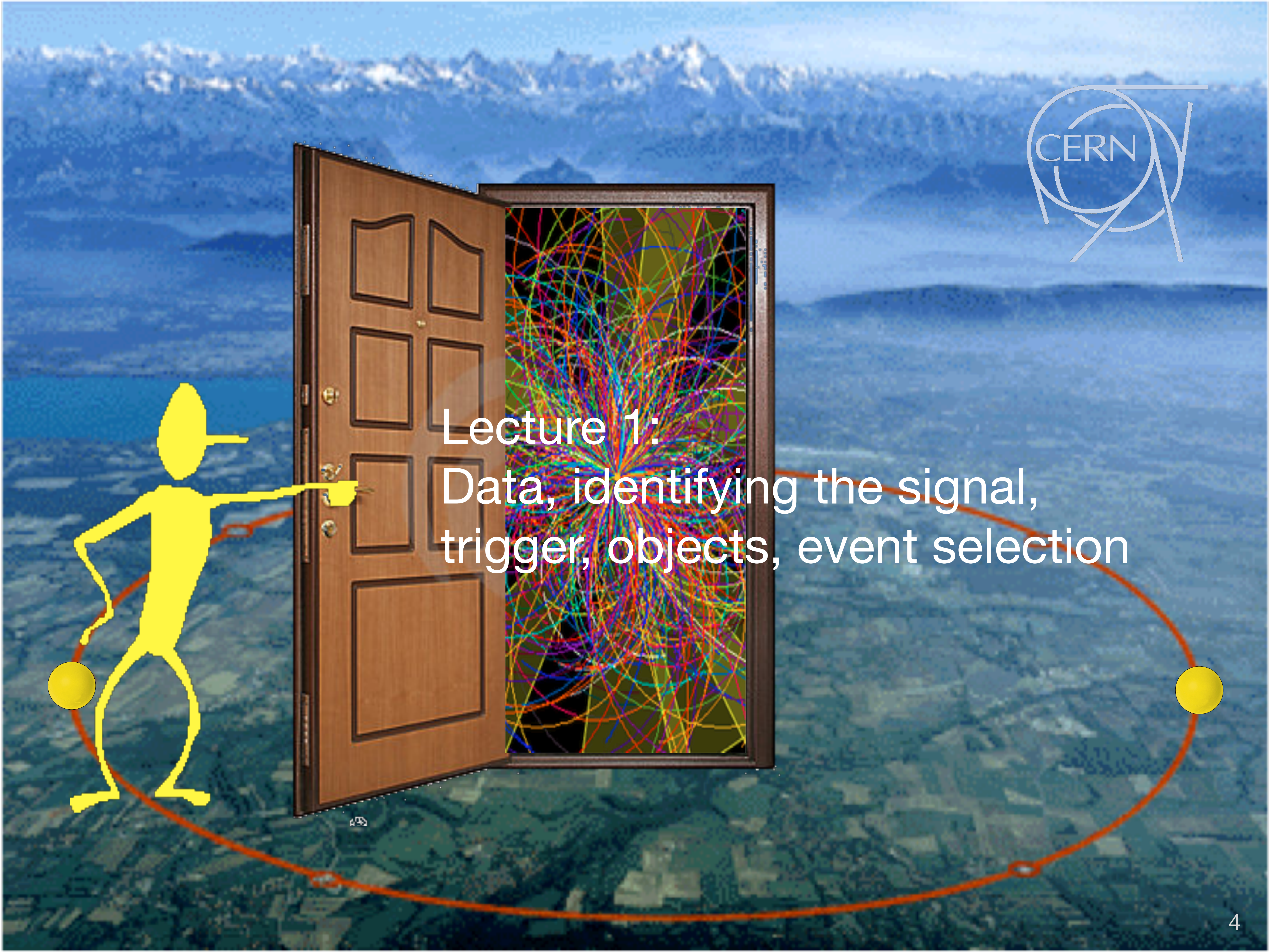
Color highlighting:

Maroon color: New concept introduced / defined.

Blue: Emphasized information.

Overview of the lectures

- Lectures + hands-on exercises focusing on LHC data analysis for physics
- Lecture 1:
  - Data, identifying the signal, trigger, objects, event selection
- Lecture 2:
  - Selection optimization, background estimation
- Lecture 3:
  - Systematic uncertainties, results, interpretation
  - A few highlights from LHC searches
- Hands-on exercises 1 and 2:
  - Looking into signal and background events, plotting variable distributions, applying simple event selections.



Lecture 1:
Data, identifying the signal, trigger, objects, event selection

Elements of data analysis

Data analysis at the LHC takes us from detector output to an experimental result.

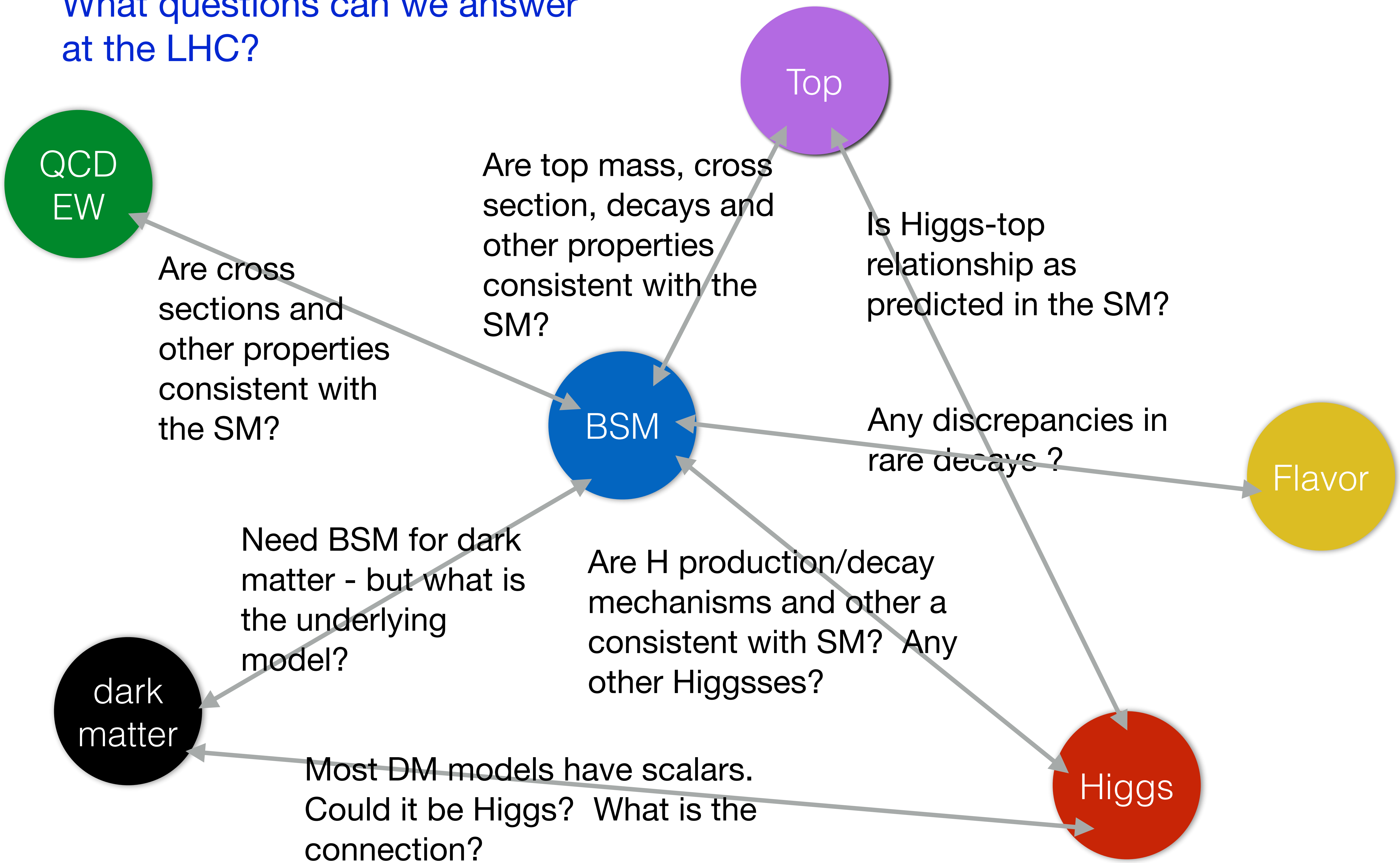
Experimental result is the empirical outcome of an experiment, such as an event count or the measurement of a physical quantity.

Our focus will be LHC physics analyses, where the experimental results are expected to improve our understanding of physics.

Hundreds of different analyses are performed with the LHC data for different purposes, for answering different physics questions.

Overview of the lectures

What questions can we answer at the LHC?



Elements of an LHC physics analysis

0. The data
1. Identifying the signal and designing the search method
2. Designing triggers to collect data
3. Object reconstruction and identification
4. Event selection to discriminate signal from backgrounds
5. Background estimation
6. Systematic uncertainties
7. Statistical analysis
8. Experimental results
9. Interpretation

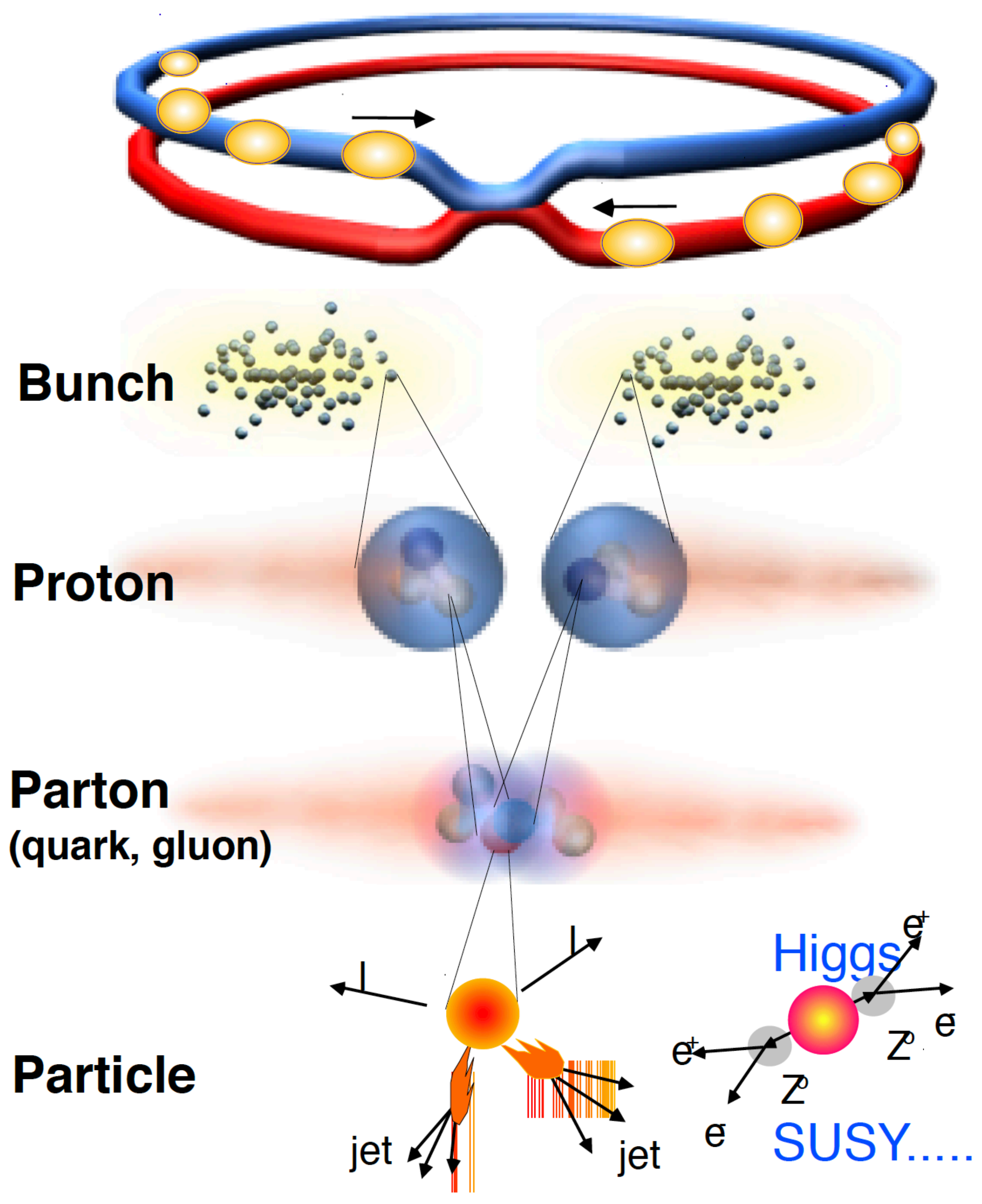
(0) The data

How much is a process produced?

LHC physics analyses are done using collision events.

A collision event at the LHC is the outcome of collision of two protons.

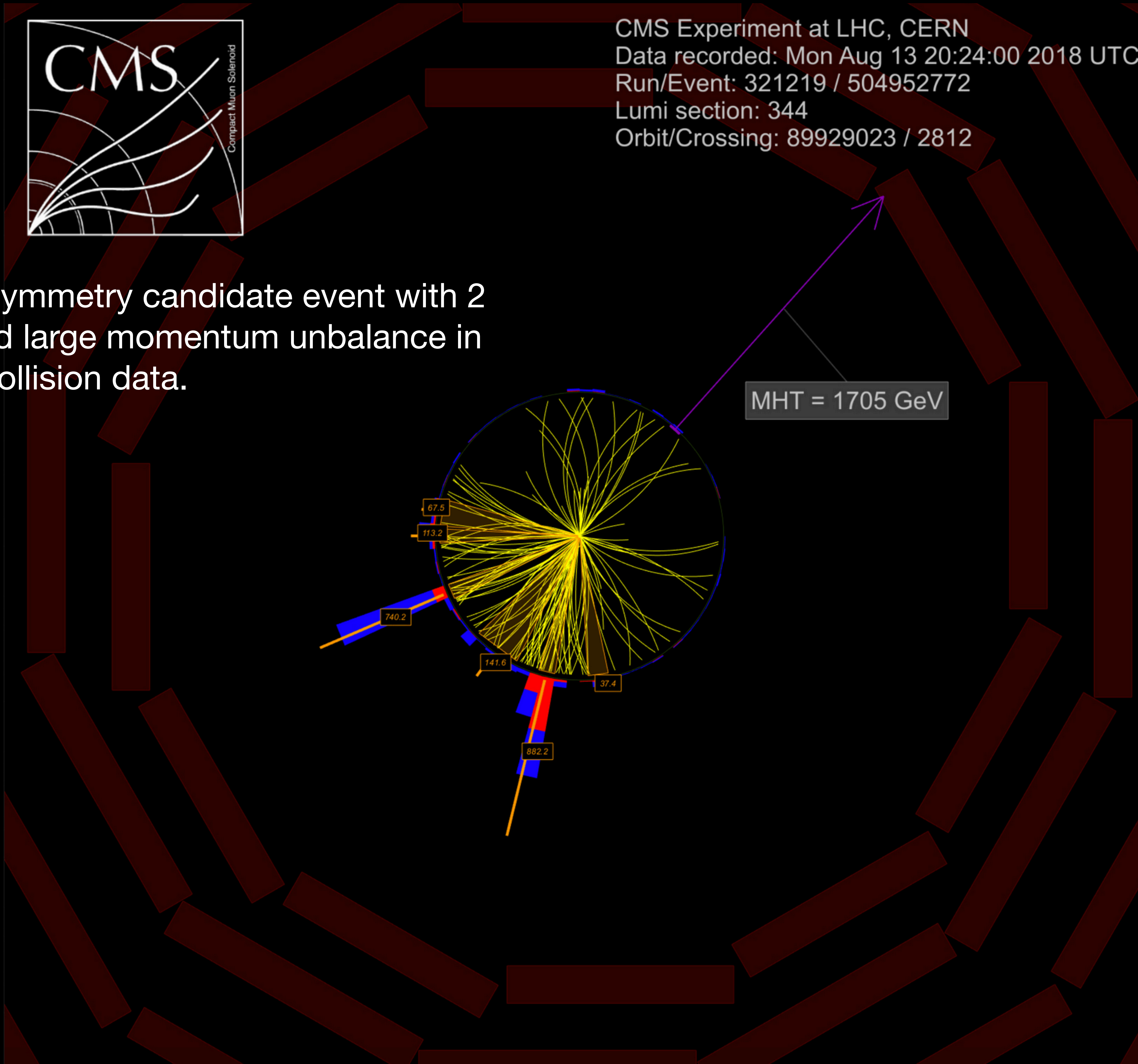
Proton - Proton	**~3600 bunch/beam**
Protons/bunch	**~10^{11}**
Beam energy	**~6.5 TeV (6.5×10^{12} eV)**
Luminosity	**>10^{34}cm^{-2}s^{-1}**
Collision rate	**40MHz**



How much is a process produced?

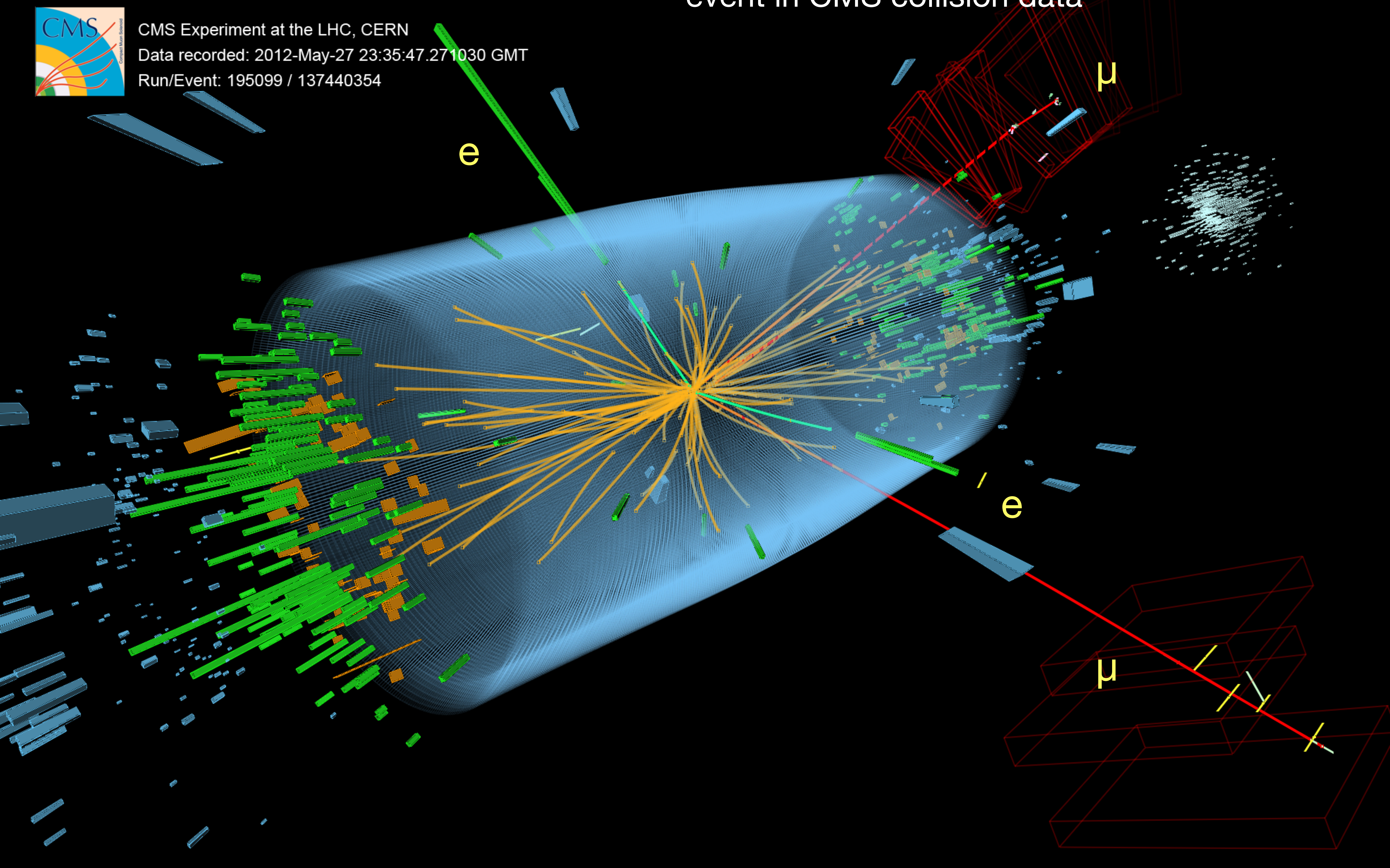
In a high energy physics data analysis, we use two types of events:

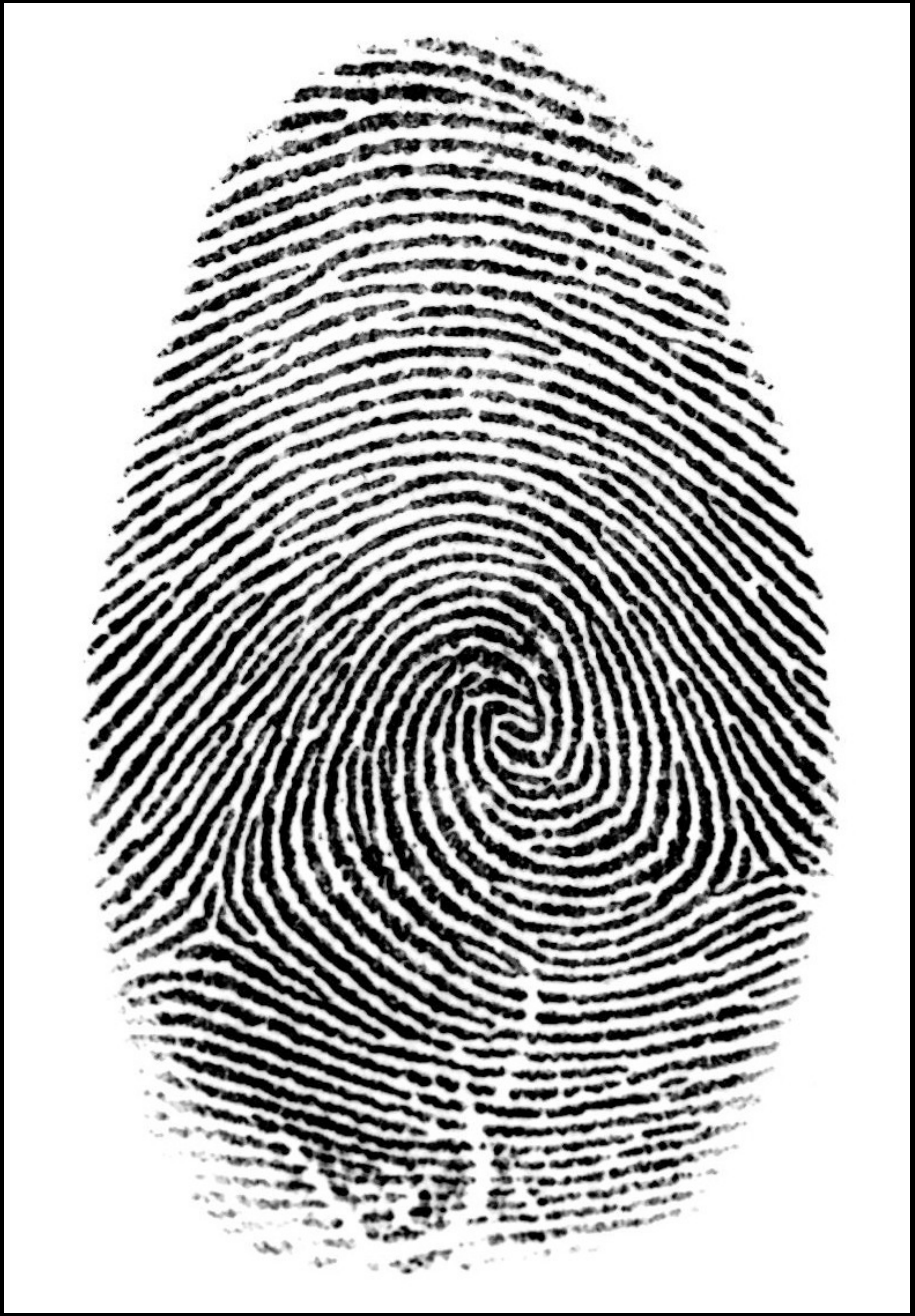
- Real collision events: Collision events produced at the accelerator and recorded by the detector according to the so-called trigger selection.
- Simulated collision events: Computer modelling of collision events. Needed to model physics processes, and are used in almost all analyses. Unlike real data, we have control over which process we generate. Simulation is done in several phases.
  - Generation of the physics process: e.g. production, decay, showers, hadronization. Use tools like Pythia, MadGraph, etc.
  - Simulation of the detector effects: Simulation of the passage of generated particles through a detector. Use tools like experimental simulation packages or the public fast simulator Delphes.
- Reconstruction of the objects: Both real and simulated events must undergo reconstruction, where real or simulated detector output is converted into objects such as jets, electrons, muons, taus, photons, etc.



Supersymmetry candidate event with 2 jets and large momentum unbalance in CMS collision data.

Higgs —> ZZ —> eeμμ candidate event in CMS collision data



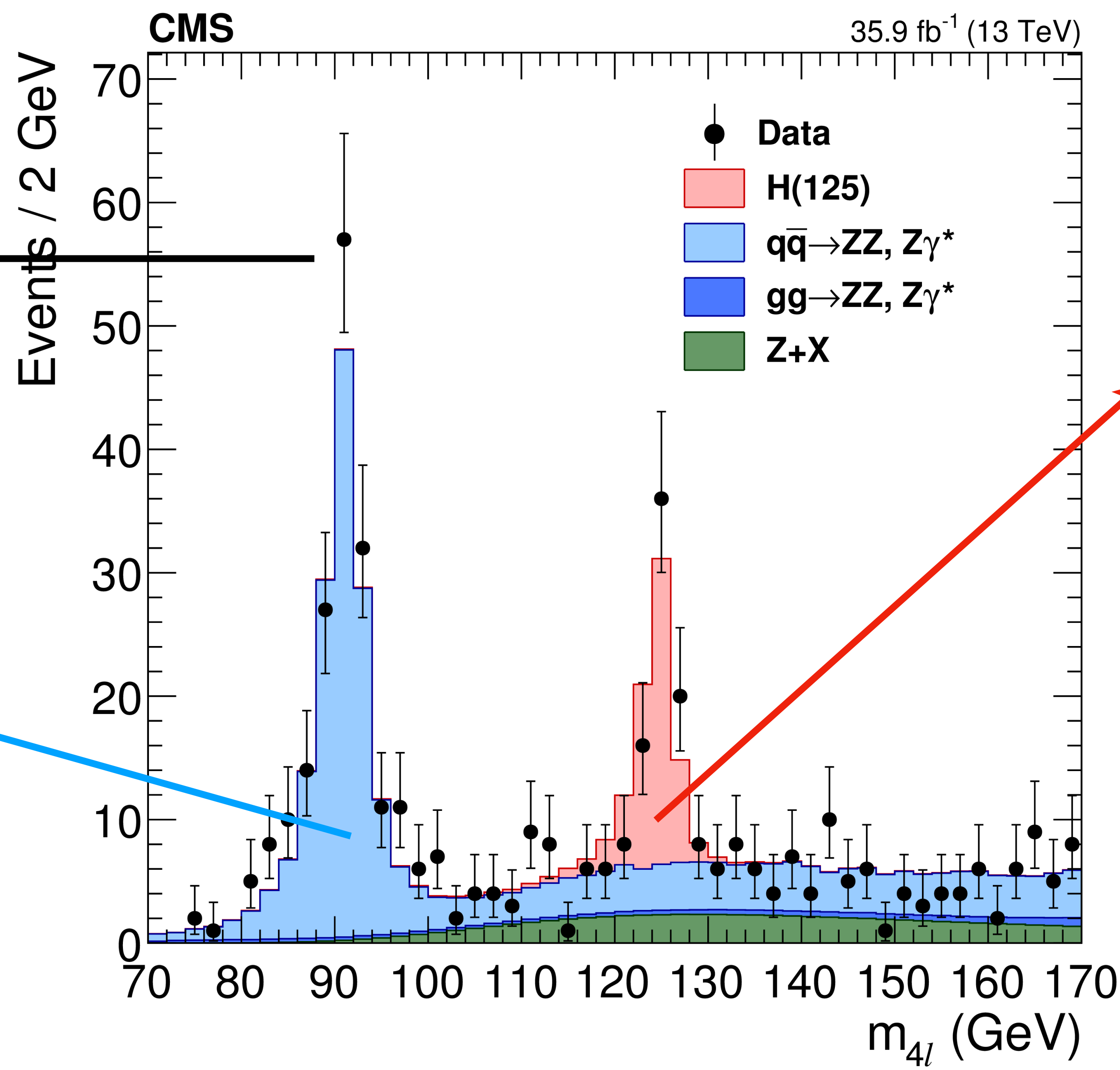


(1) Identifying the signal and designing the search method

What is a signal?

From the CMS Higgs observation in H —> ZZ —> 4 leptons:

Real data: What the experiment gives us.



Signal: the process we are interested in observing.

Background: the processes we are not interested in observing. We either cut these away or estimate their amount.

Is real data consistent with the signal or the background?

Get to know the signal

- Is the signal worth looking at?
  - Let's put our physics model in a calculator.
  - How many events do we expect for the run?
  - What signature does the signal have?
    What are the final state particles?
  - What are the background processes that have a similar final state?

What is the most significant property of the signal?

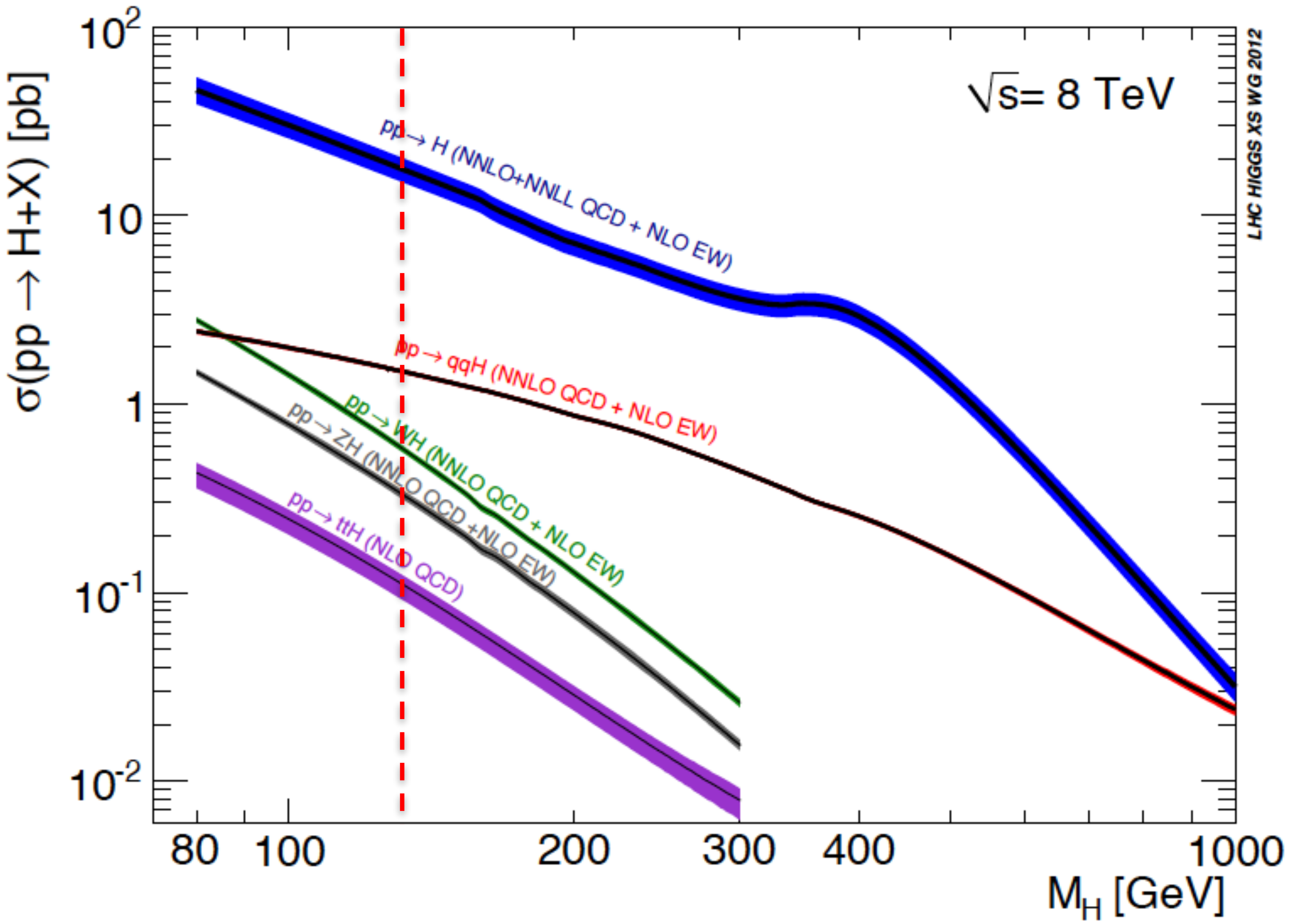
- Produce MC for the signal and investigate:
  - Generate and simulate signal events.
  - How can the signal be discriminated from the background? Find the characteristic kinematics. Can we trigger the signal?
- Produce MC for the backgrounds and investigate:
  - Generate and simulate background events.
  - Is the signal visible after the candidate event selection?

If the answer is "yes", let's start the analysis!

How much is a process produced?

- Cross section (σ): probability that a particle interaction will occur. Cross section depends on the type, properties and interaction energy of the interacting particles. Has area units (barn (10^{-28}m^2) - picobarn, femtobarn relevant for LHC energies).

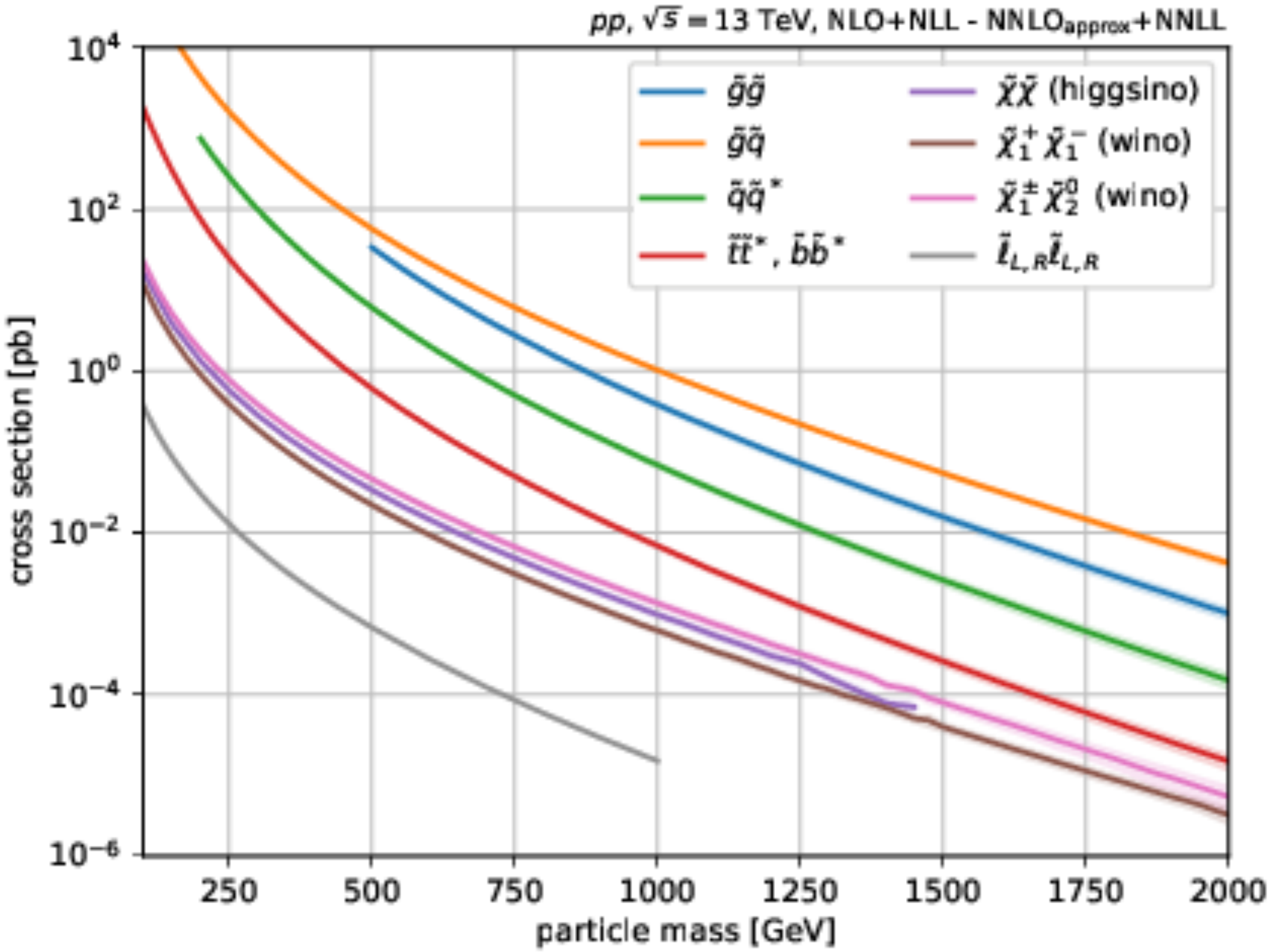
Example: Cross sections for different production mechanisms of the Higgs boson plotted versus the Higgs boson mass.



How much is a process produced?

- Cross section (σ): probability that a particle interaction will occur. Cross section depends on the type, properties and interaction energy of the interacting particles. Has area units (barn ($10^{-28}m^2$) - picobarn, femtobarn relevant for LHC energies).

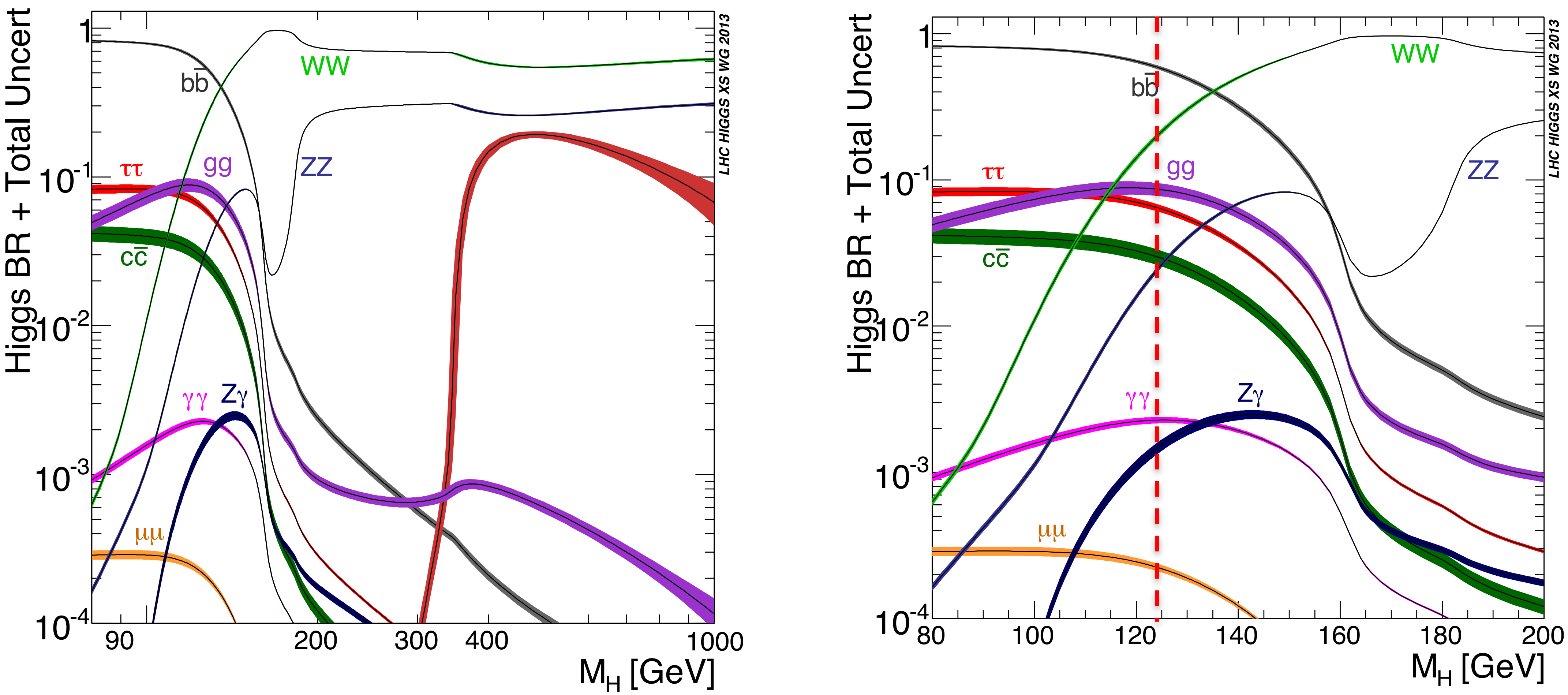
Example: Production cross sections for different supersymmetric particle types versus the particle mass.



Signal identification and search method

How much is a process produced?

- Branching ratio (BR): probability of a particle decaying to a certain mode.
- Sum of BRs for all decay modes equals 1.

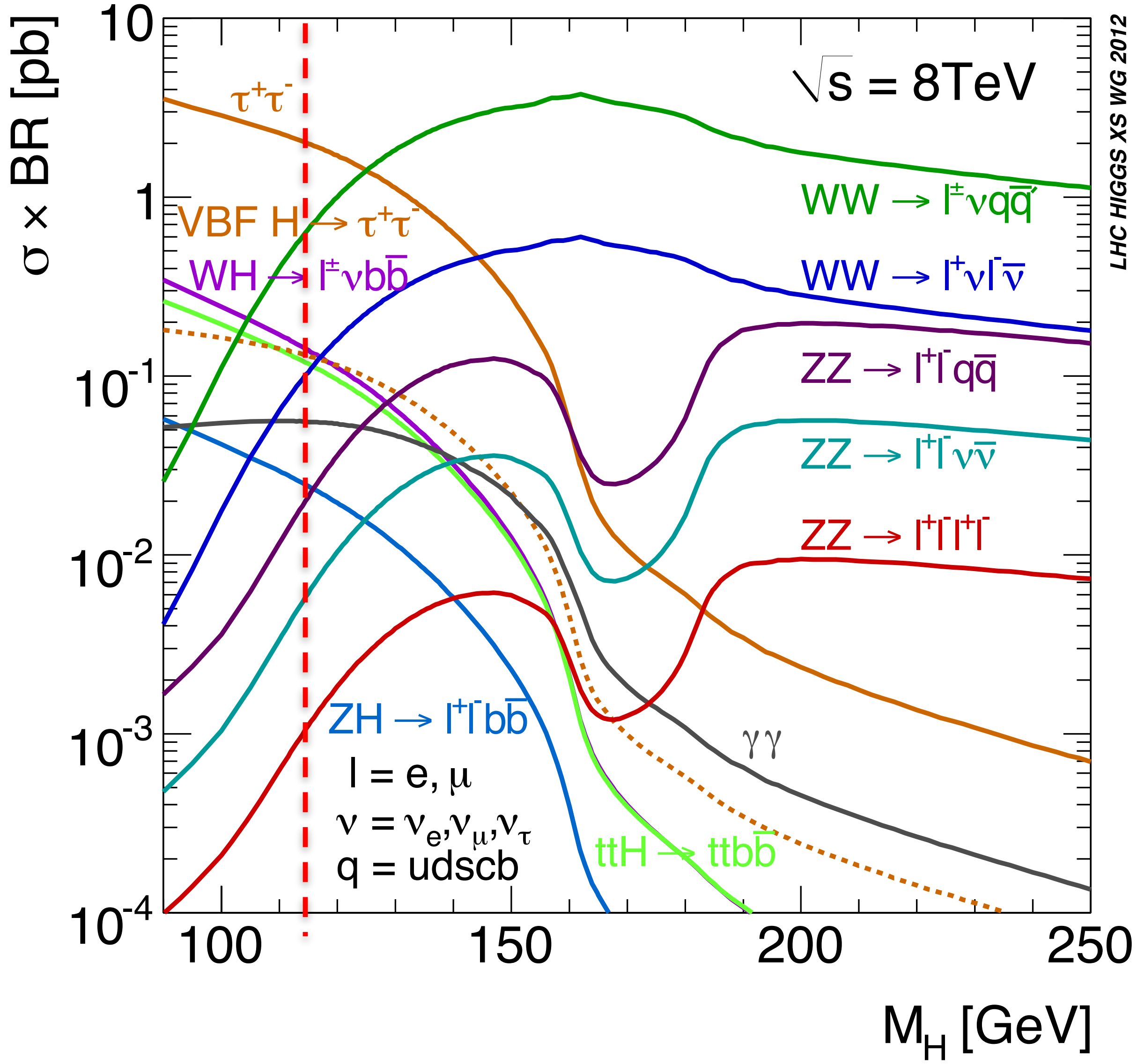


Example: Branching ratios for the different decay modes of the Higgs boson plotted versus the Higgs boson mass.

How much is a process produced?

- Effective cross section σ_e :

$$\sigma_e = \sigma \times BR$$



Example: Effective cross sections for varying production and decay mechanisms of the Higgs boson plotted versus the Higgs boson mass.

The higher $\sigma \times$ BR is. the more Higgs bosons are expected to be produced.

This is an important plot in deciding which search channels to use.

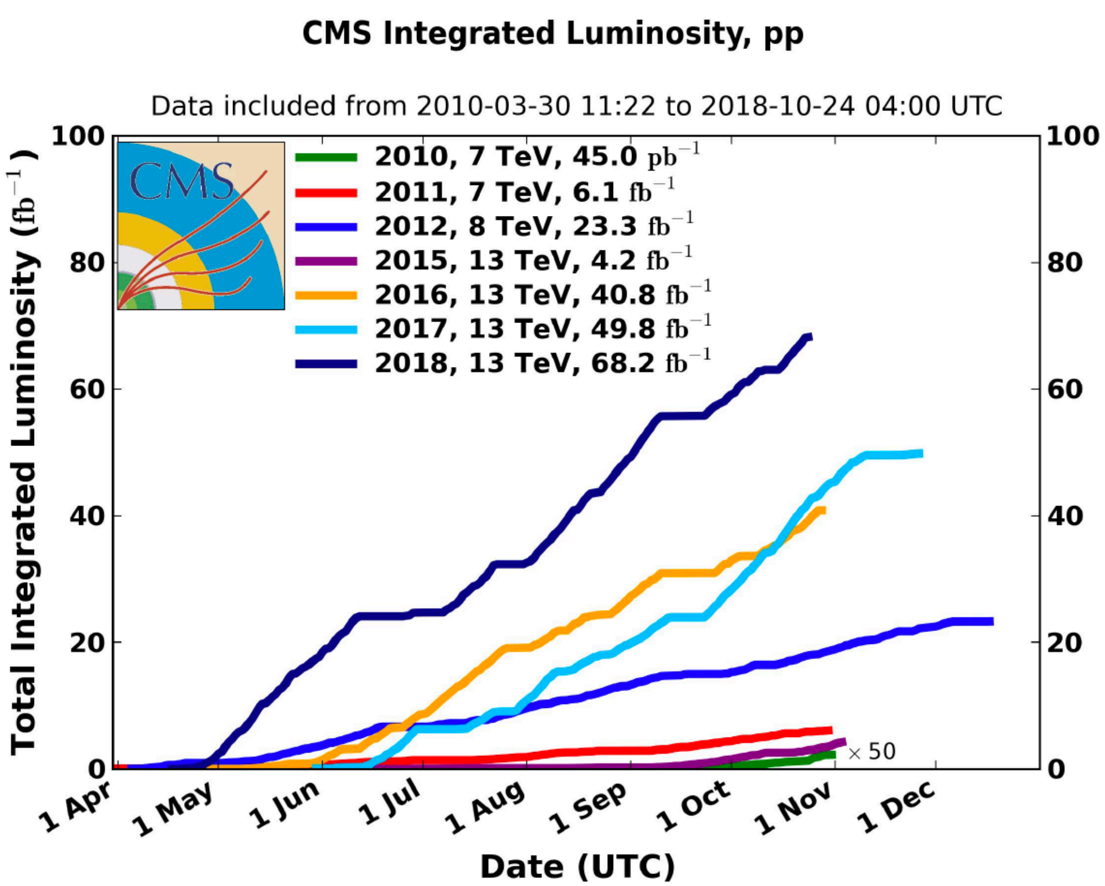
How much is a process produced?

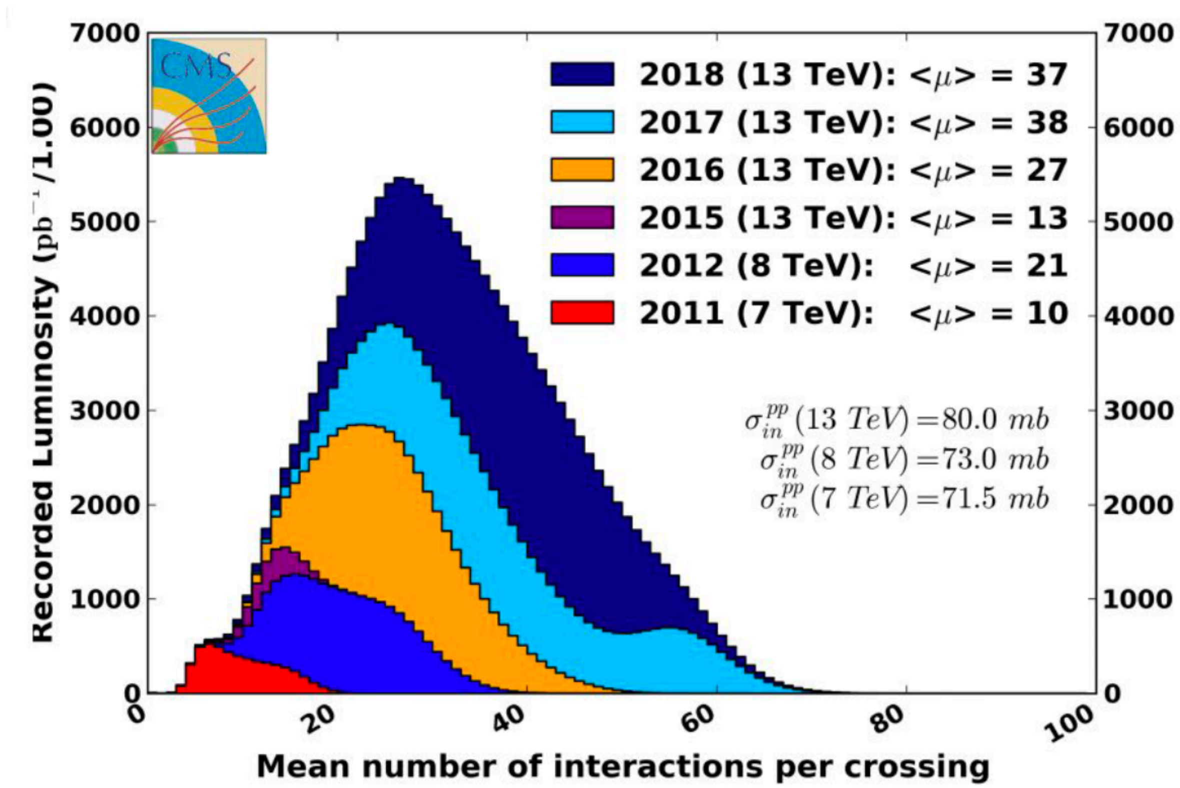
- Instantaneous luminosity, L: number of collisions that take place in 1 second in unit area. Has units of $cm^{-2}s^{-1}$. LHC luminosity at the end of 2012 was: $10^{33}\ cm^{-2}s^{-1}$.
  - Integrated luminosity : Number of collisions per unit area in a given period of time. Has unites of fb^{-1}, pb^{-1}.

$$L = \int L dt$$

- Number of expected events N for a process is given as:

$$N = \sigma_e \times L = \sigma \times BR \times L$$



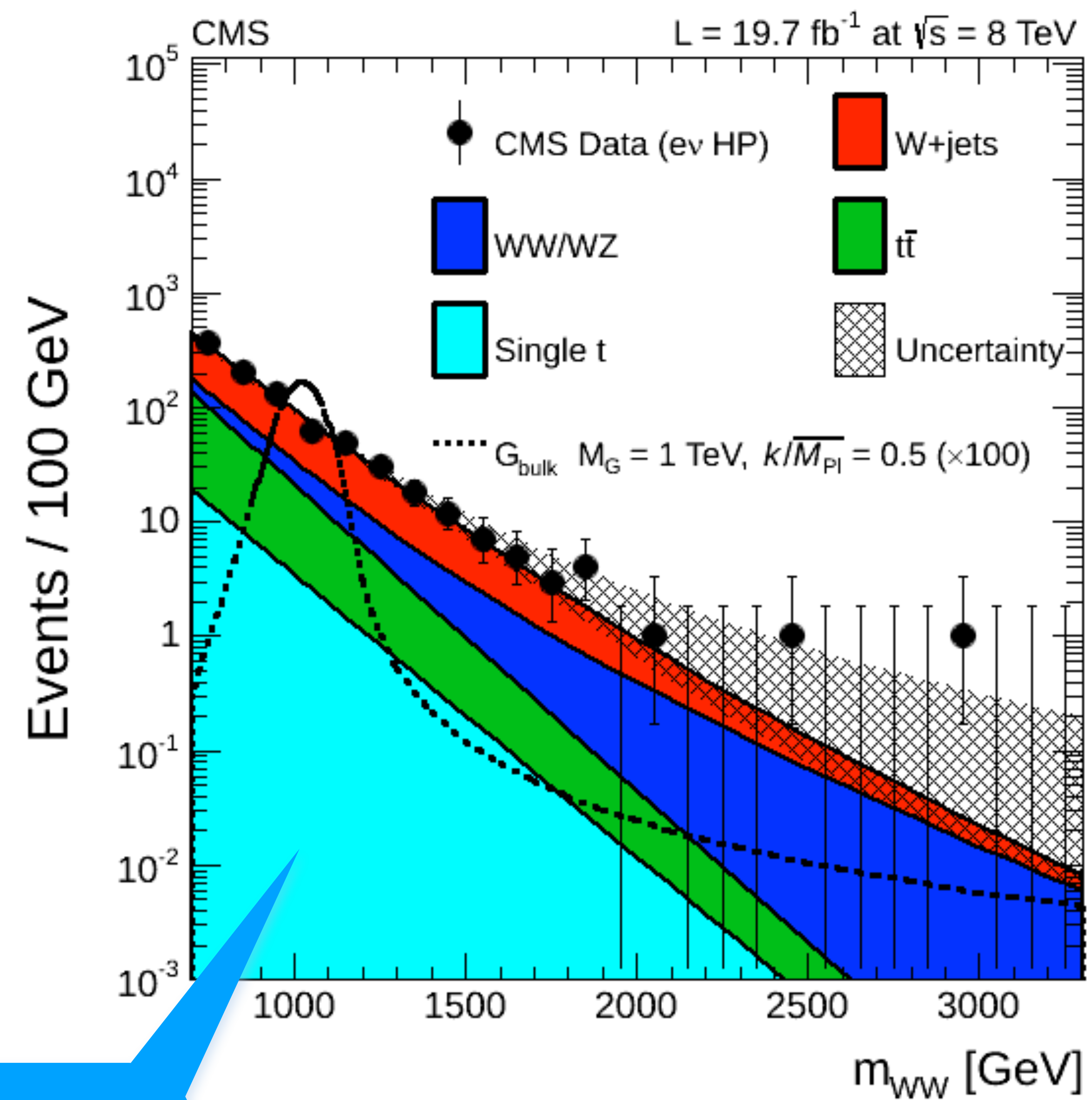


How to discriminate signal from background

Suppose we are looking for a heavy resonance (e.g., a bulk graviton G_{bulk}) decaying into WW.

Signal will give us a peak at the WW invariant mass distribution around the G_{bulk} mass.

On the other hand, since there is no resonance in the SM decaying to WW, we will not get a similar peak for the SM.



Seeing such a peak in the invariant mass would point out to new physics discovery (but it was not the case…).

We will discuss signal discriminating variables soon.

(2) TRIGGER

With what do we trigger?

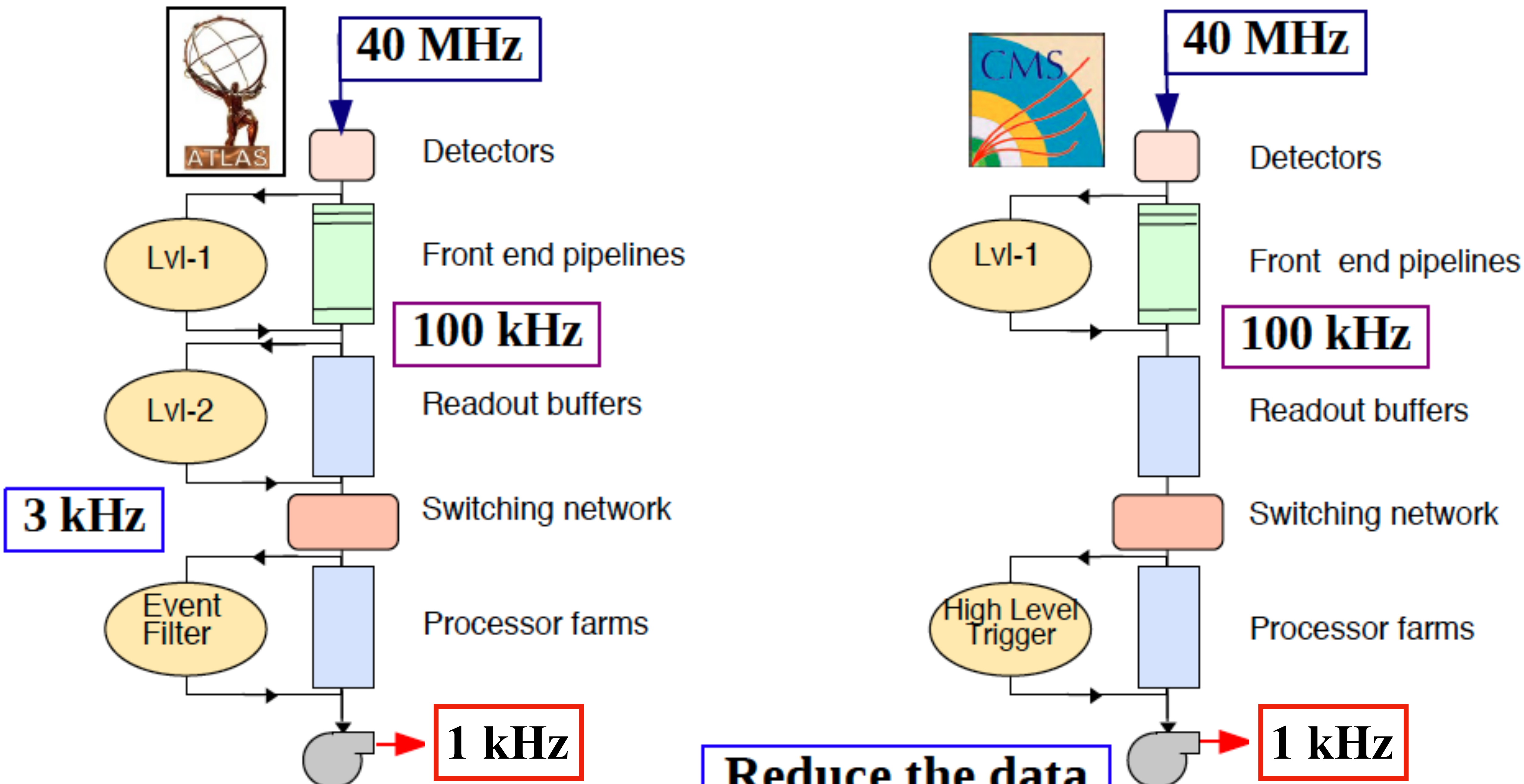
There are two main phases in working with data:

- Online: During data taking.
- Offline: After data taking.

A trigger is a system that uses criteria to rapidly decide which events in a particle detector to keep when only a small fraction of the total can be recorded.

- Triggers are fast online filters that select the most interesting events during data taking and save them for offline analysis. Events not recorded during data taking are lost.
- Trigger systems are necessary due to real-world limitations in computing power, data storage capacity and rates.
- Since experiments are typically searching for "interesting" events that occur at a relatively low rate, trigger systems are used to identify the events that should be recorded for later analysis.
- Triggers are rough sketches of the offline selection. We select events with final states representative of the physics we are looking for.

Trigger systems in ATLAS and CMS



Reduce the data volume in stages

Triggering for analysis

We have 4 trigger tasks related to data analysis:

1. Design trigger paths: Do triggers capable of collecting data with the properties of our signal exist? Otherwise we design them.

What do we use while designing trigger paths?

- Trigger on objects (jets, b-jets, e, μ, γ, missing transverse energy)
  - Trigger objects should satisfy certain identification (e.g. number of hits and χ^2 for the track of a muon) and isolation criteria
  - Thresholds (i.e. kinematic selection criteria) on trigger objects (e.g.: $p_T > 100$ GeV, $|\eta| < 2.4$)
  - Requirement on trigger object multiplicities (e.g.: =2 μs, =4 jets)
- Trigger on event variables: dijet invariant mass, hadronic transverse momentum, other complex new physics search variables (α_T, razor, etc.)

Triggering for analysis

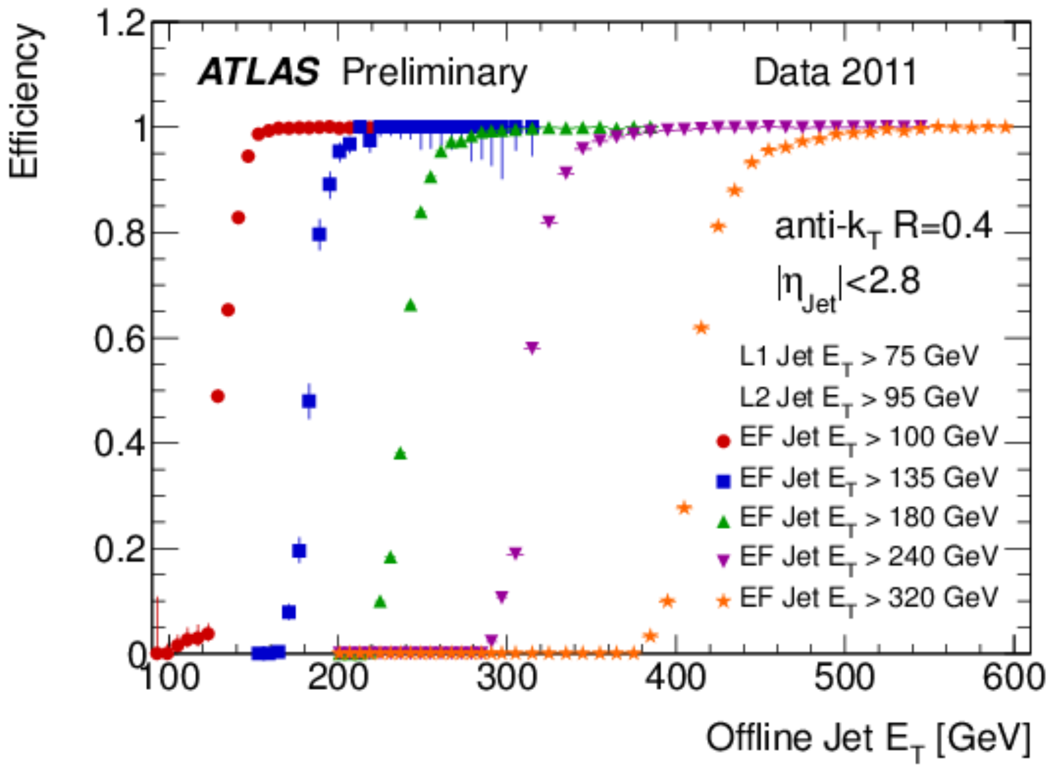
$$\epsilon_{trigger} = \frac{\text{number of events that pass the trigger}}{\text{number of total events}}$$

2. Measure trigger efficiency:

Trigger efficiency (turn on) curve shows trigger efficiency versus an offline variable.

Measuring trigger efficiency:

- Use events collected by a looser trigger
- Use events collected by an orthogonal trigger (e.g. use muon trigger to measure jet efficiency)
- Use the tag-and-probe method (a generic efficiency measurement method).



3. Estimate the trigger rate: Estimate how many events will the trigger path collect per second. If the rate is too high, we cannot record all events. Control the rate by modifying the trigger threshold (e.g. increase the jet pT threshold.)

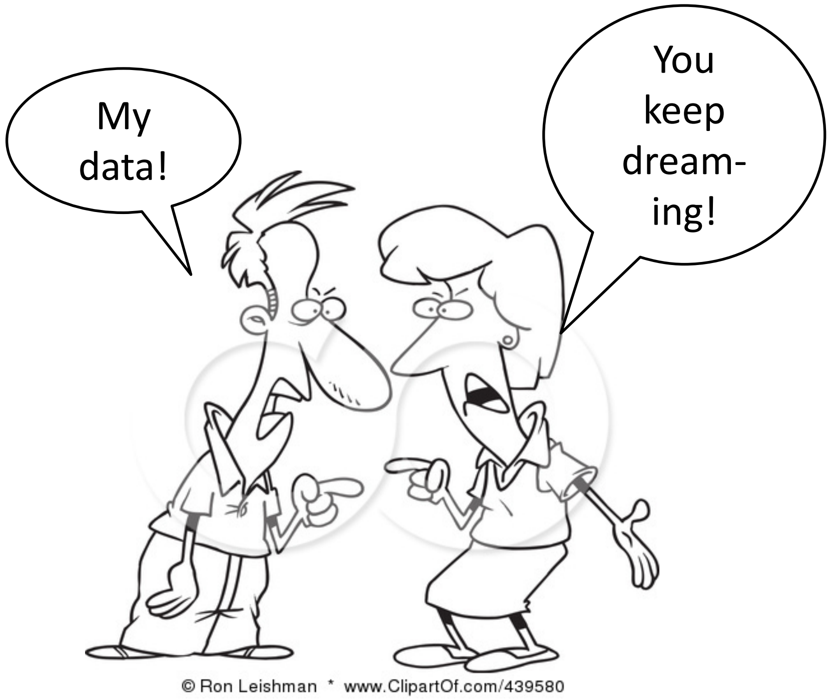
3. Estimate trigger timing: Estimate the CPU time it takes to process the trigger path. Shorter the better!

Putting all together: the trigger menu

Everyone wants more (unbiased!) data for their analysis, hence lower trigger thresholds, hence higher rates.

But the experiment can only collect a certain amount of data: total bandwidth is constant (~1 kHz)!

How to share the bandwidth among the many triggers? Depends on the experiment's physics priorities.



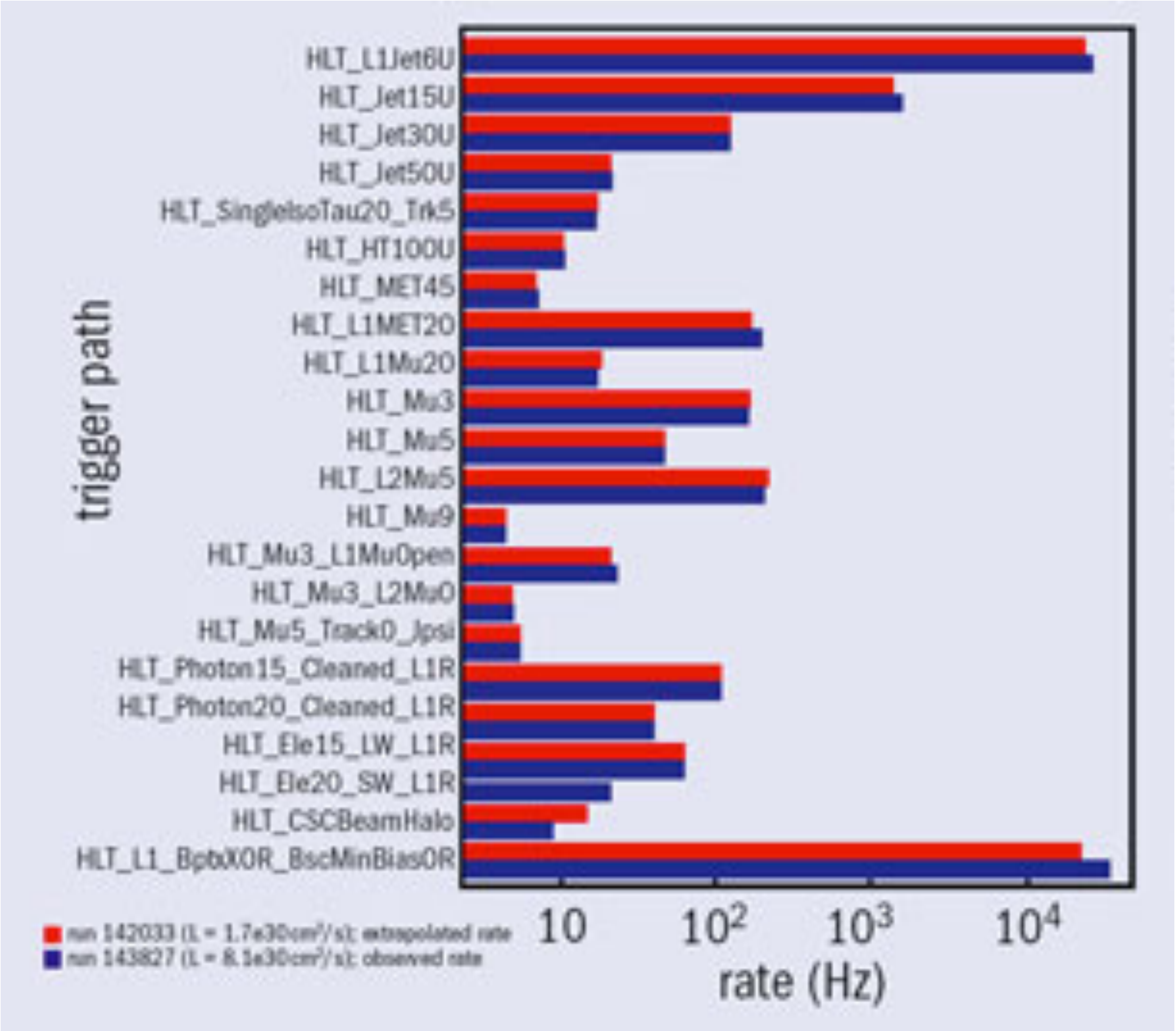
A trigger menu is the set of all triggers used for collecting data in a given run period.

Data is eventually partitioned into smaller sets based on groups of triggers (e.g. single muon dataset, jets+HT dataset, etc.)

Some concepts:

- Prescaling: Randomly taking 1 out of N events that pass a certain trigger.
- Trigger evolution: As the energy and luminosity increase, we keep collected data constant by increasing the trigger thresholds or prescaling them.

Putting all together: the trigger menu



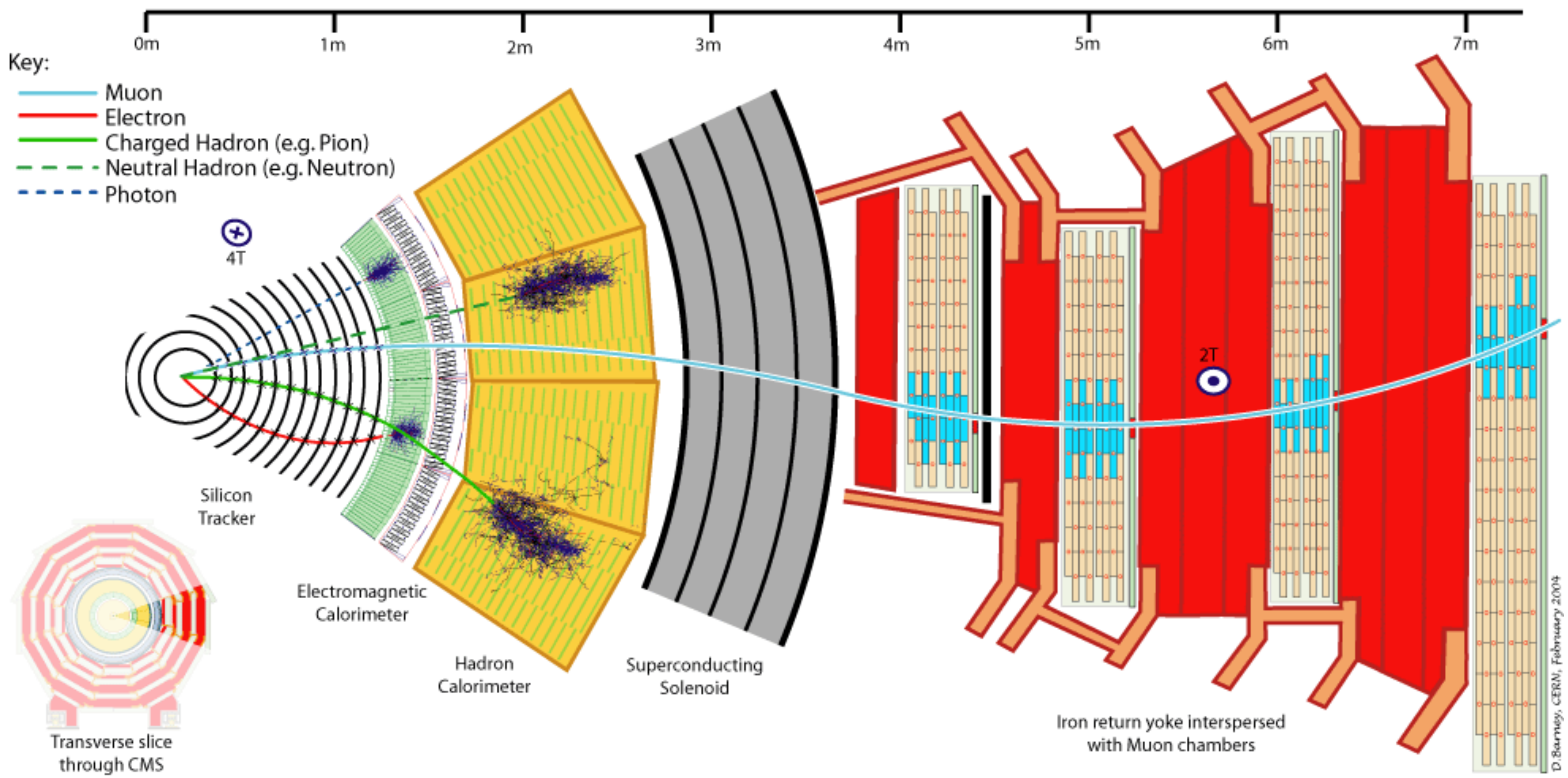
Example trigger menu from CMS with predicted and observed rates.

(3) Object reconstruction and identification

Reconstructing objects from subdetectors

Objects are particles coming out of interaction as they are seen in the detector: jet, b-jet, e, μ, τ, γ, missing transverse energy (MET).

——> described in detail in Nhan Tran’s lectures.



Objects for an analysis

Defining and using objects in an analysis:

Determine which objects would be present in your final state:

- Do I want to reconstruct a Z boson? I need electrons or muons.
- Do I want to reconstruct top quarks? I need b jets.
- Am I looking for supersymmetry? I need missing transverse energy.

Decide on the identification and isolation criteria:

- Will I do a very precise measurement of the Higgs boson in the 2 photon channel? I need tightly identified and isolated photons.
- Do I want no events with top quarks? Let me select loosely identified b jets (i.e. anything that resembles a b jet) and eliminate events containing them.

(4) Event selection

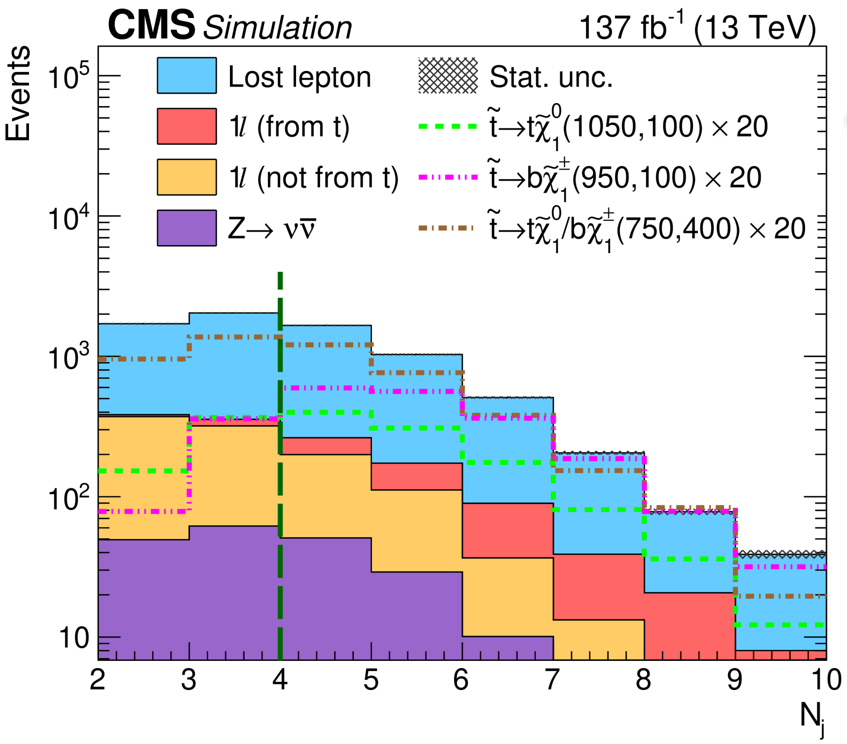
Event selection
Signal region

Signal region: An event selection where the expected signal is significant with respect to the backgrounds. To find a good signal region:

- Characterize the signal:
  - What are the kinematic properties specific to the signal? Multijets? Opposite sign dileptons? Peaks/tails in kinematic variables?
- Find a candidate event selection containing sufficient amount of signal.
  - Do our selection variables efficiently eliminate the backgrounds?
- Check the backgrounds in the signal regions:
  - Is the signal statistically significant, i.e. does signal appear with sufficient significance above background?
  - Can the backgrounds be estimated? (soon...)
- Can we trigger the signal region? Make sure that offline selection corresponds to a region where trigger efficiency is well-modelled.
- Make sure that the selection variables are reconstructed and identified in well-defined regions of the detector (e.g. not feasible to design a search with forward electrons).

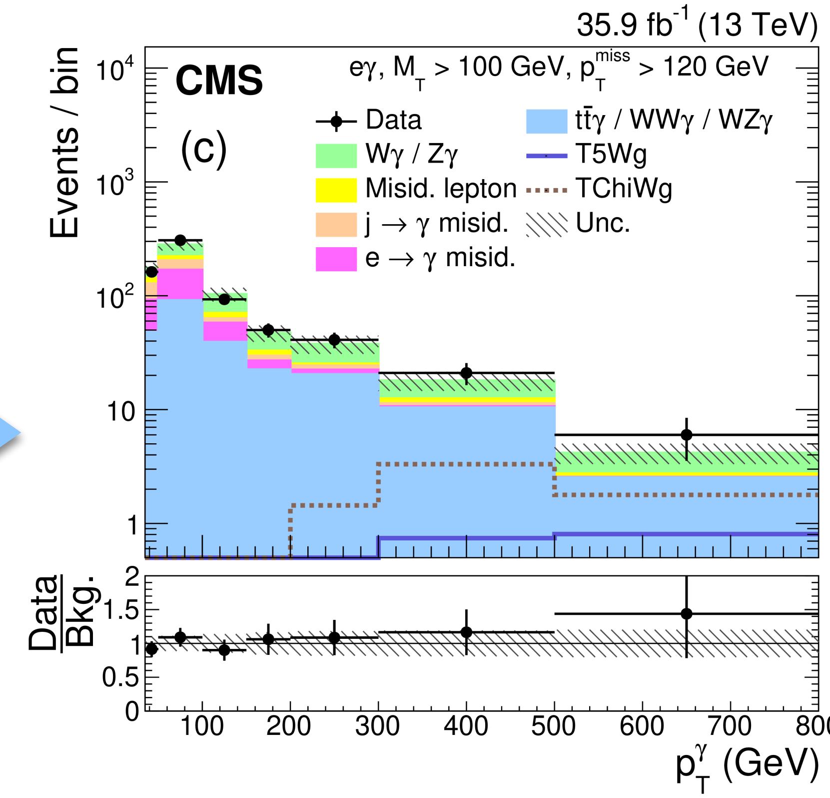
Characterizing the signal

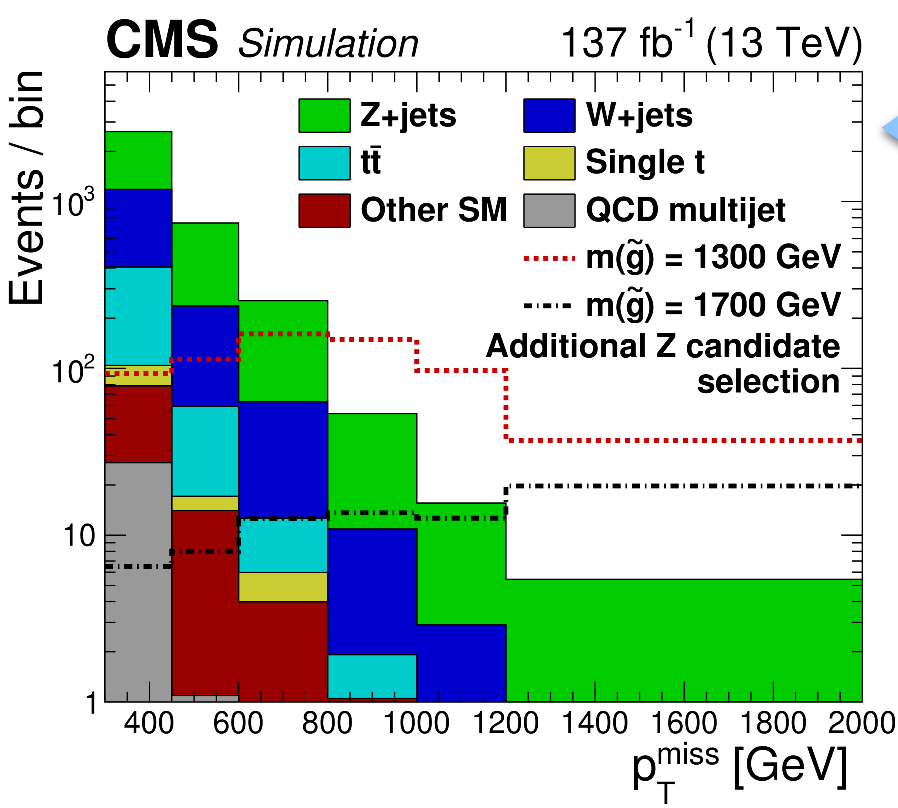
Simple variables



Object multiplicities, e.g. number of jets, leptons, b-jets, etc.

Object transverse momenta pT, e.g. 1st jet pT, 2nd muon pT, photon pT, etc.

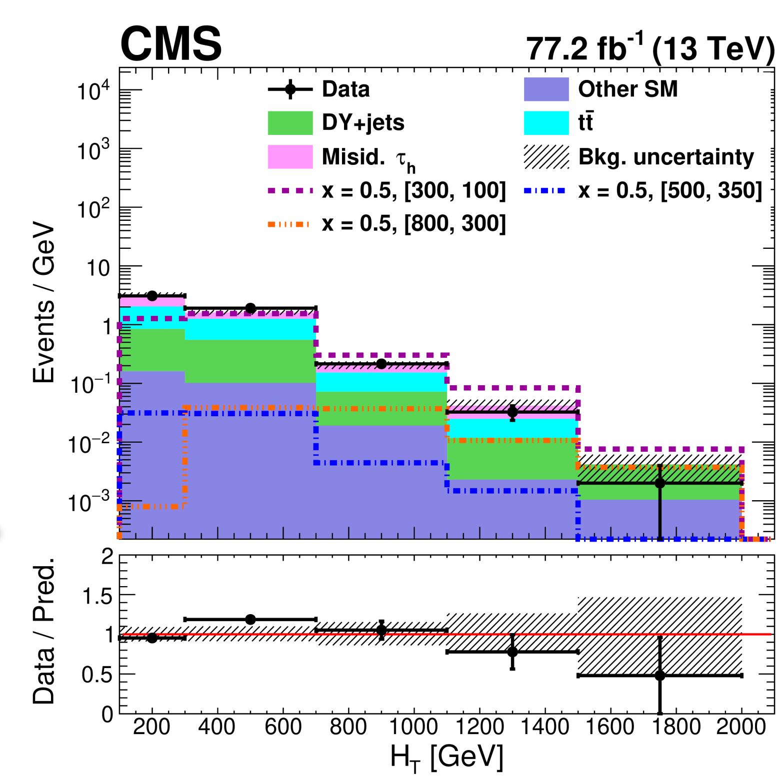




Missing transverse energy/ momentum ETmiss / pTmiss (known as MET!)

Hadronic transverse energy

$$H_T = \sum_{i}^{n\,jets} p_T^{jet_i}$$

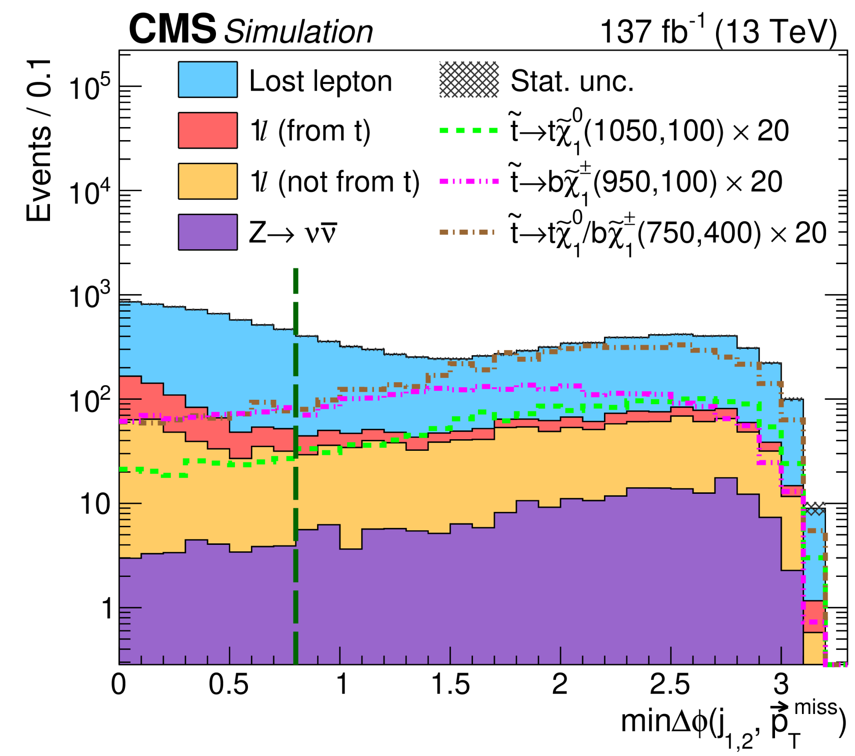


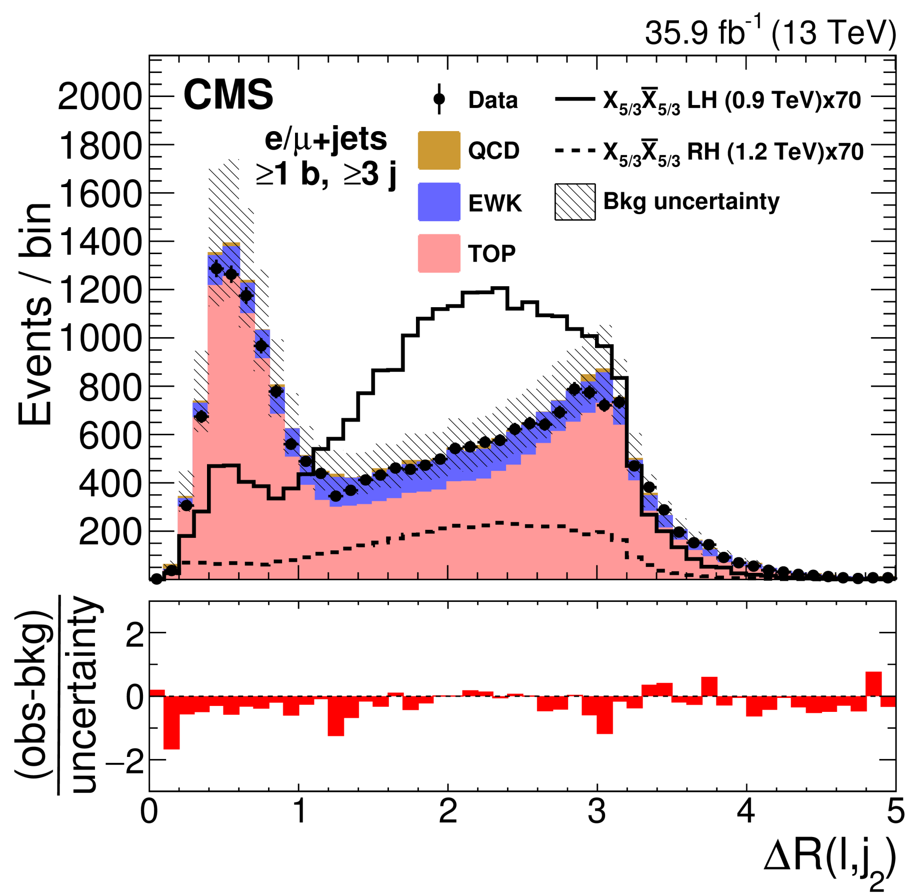
Characterizing the signal
Angular variables

Azimuthal angular separation Δϕ between 2 objects, or objects with MET.

When there are >1 objets of the same type in an event, use the minimum Δϕ.

Good in eliminating events with fake MET.





Angular separation ΔR between 2 objects:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$$

Looks at the overall proximity of objects, e.g. to see if they decayed from the same mother.

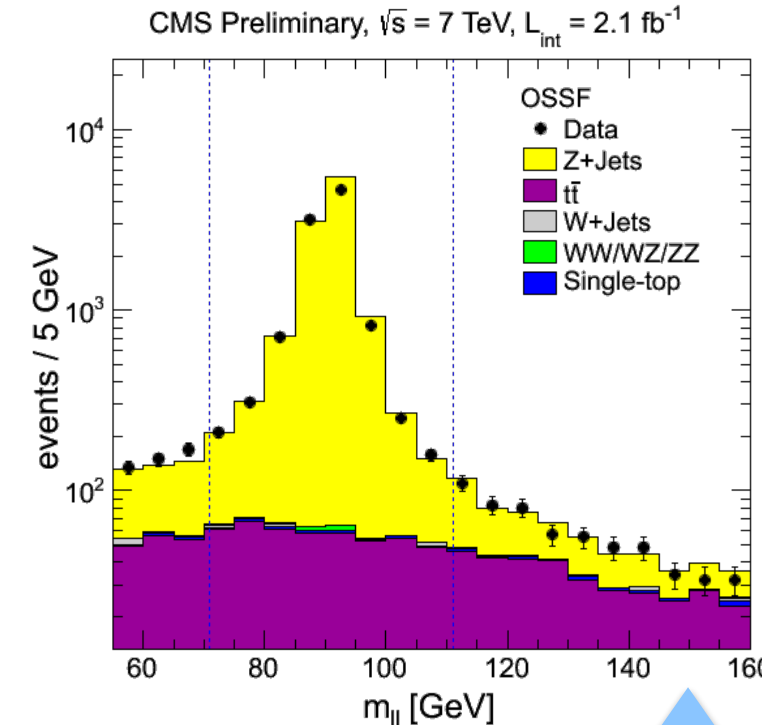
Characterizing the signal
Good old invariant mass

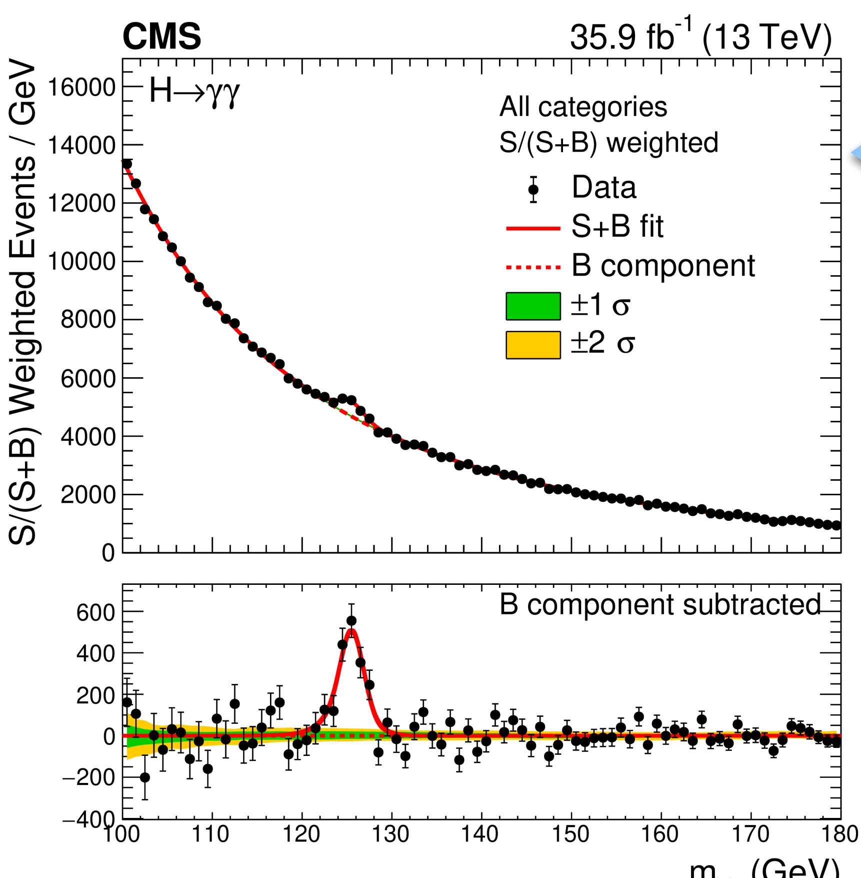
A mother particle decaying into I final state particles has the invariant mass:

$$m = \sqrt{\left(\sum_i E^i\right)^2 - \left(\sum_i \vec{p}^i\right)^2}$$

Inv. mass for a mother particle can be reconstructed if the 4-momenta of all its daughter particles are known. This happens when the decays products are visible.

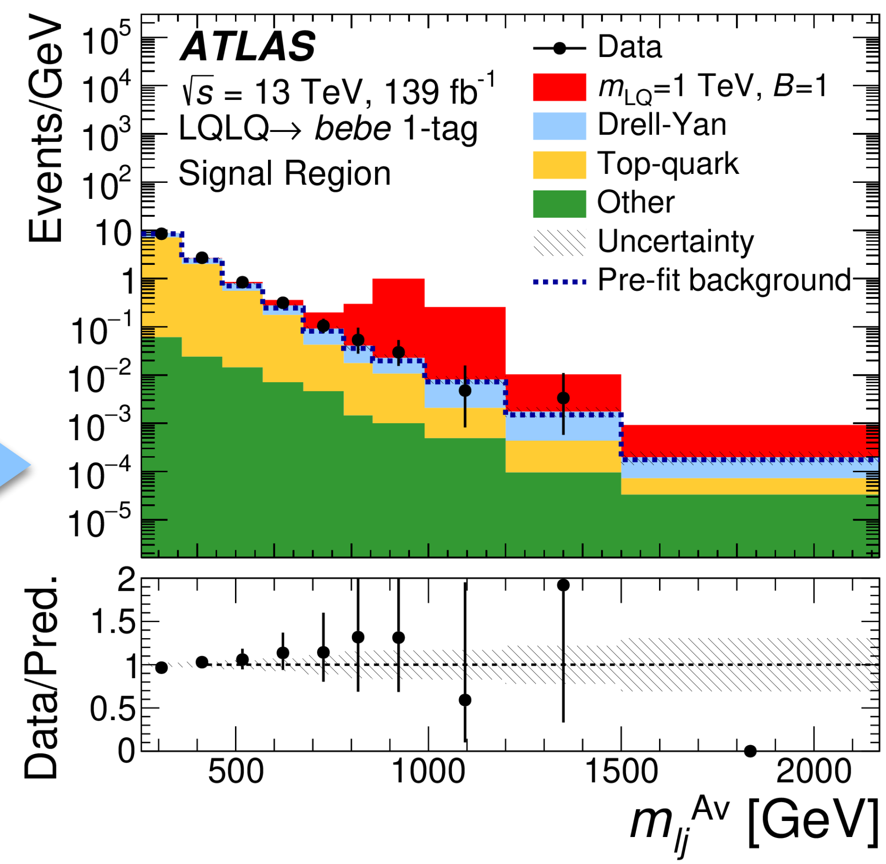
Inv. mass is used when requiring the particles with a known mass (e.g. Zs) in selection, or when looking for new states.





Higgs observed at $m_{\gamma\gamma}$ = 125 GeV.

However no excess for leptoquarks in invariant mass distribution.



Inv. mass also helps new physics searches indirectly.

$m_{Dilepton}$ was used above in a search for SUSY with Z + jets + MET to find events with Zs.

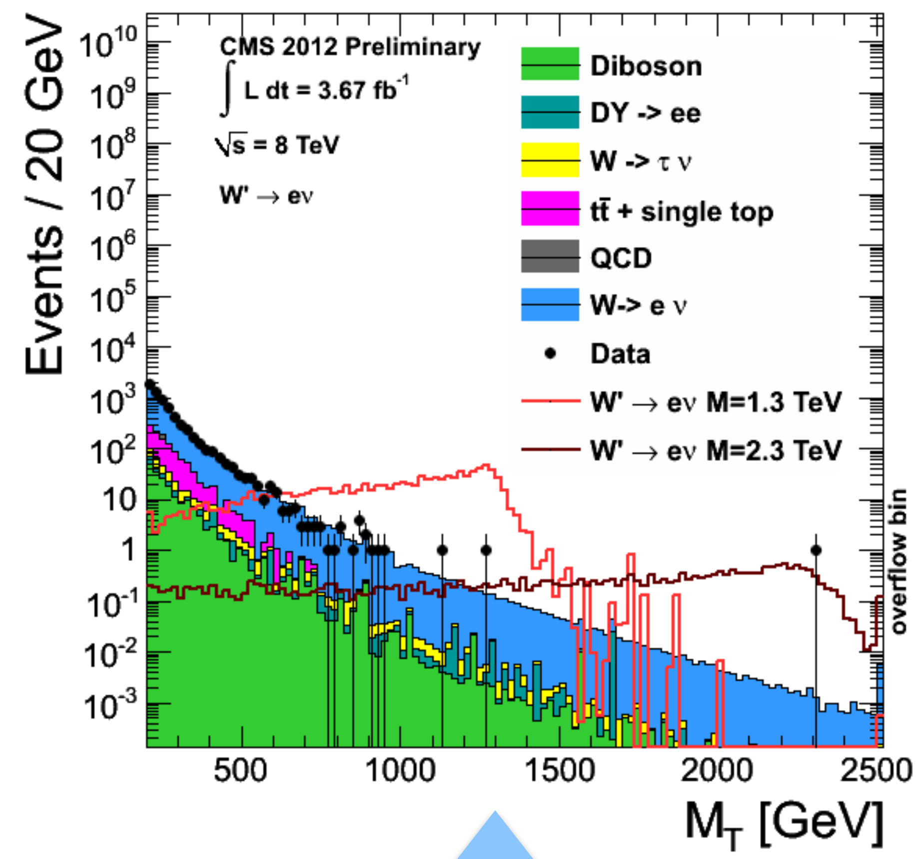
Characterizing the signal
W transverse mass

BUT…we do not always have access to full 4-momenta of the final state particles.

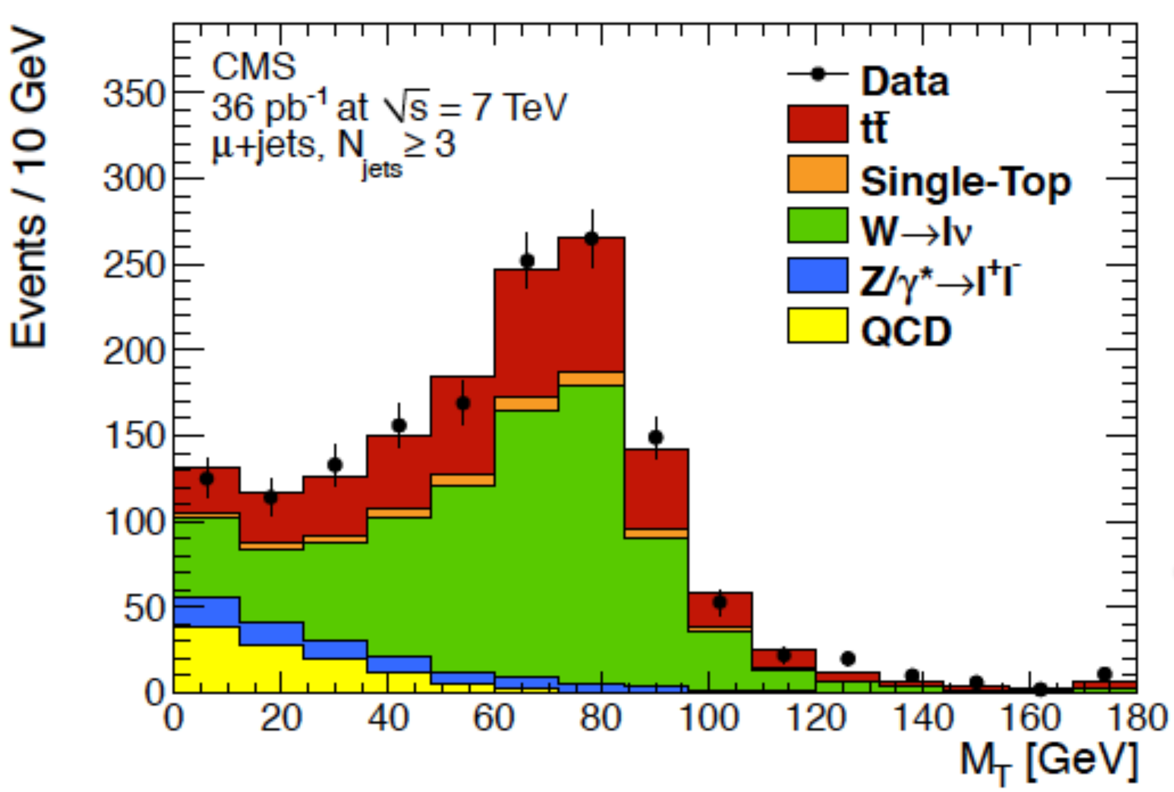
For example, in W→lν decays, invisible neutrinos escape the detector. If there is only one ν in the event, we can approximate ν transverse momentum p_T^ν by the MET. We define the transverse mass for W as:

$$m^2_{T,W} = m^2_\ell + m^2_\nu + 2(p^\ell_T p^\nu_T - \vec{p}^\ell_T \vec{p}^\nu_T)$$
$$(m_\ell, m_\nu \sim 0 \rightarrow) \simeq 2p^\ell_T p^\nu_T(1 - \cos\Delta\phi(\ell,\nu))$$

where $m_{T,W}^{max}$ gives m_W because $m_{T,W} < m_W$.



Used in new physics searches. M_T distribution for hypothetical W' particles where W'→eν.



W M_T is used extensively in top searches and searches for new physics with top-like particles as a discriminating variable in the event selection (Left: from a ttbar cross section measurement in leptons+jets channel).

Characterizing the signal

The "s"transverse mass

BUT…what if we have more than one invisible particle in the final state? Take the typical case

$$pp \to \tilde{q}_1\tilde{q}_2 \to j_1\tilde{\chi}_1 j_2\tilde{\chi}_2$$

where ~χs are invisible. Two invisible particles make up MET. The stransverse mass

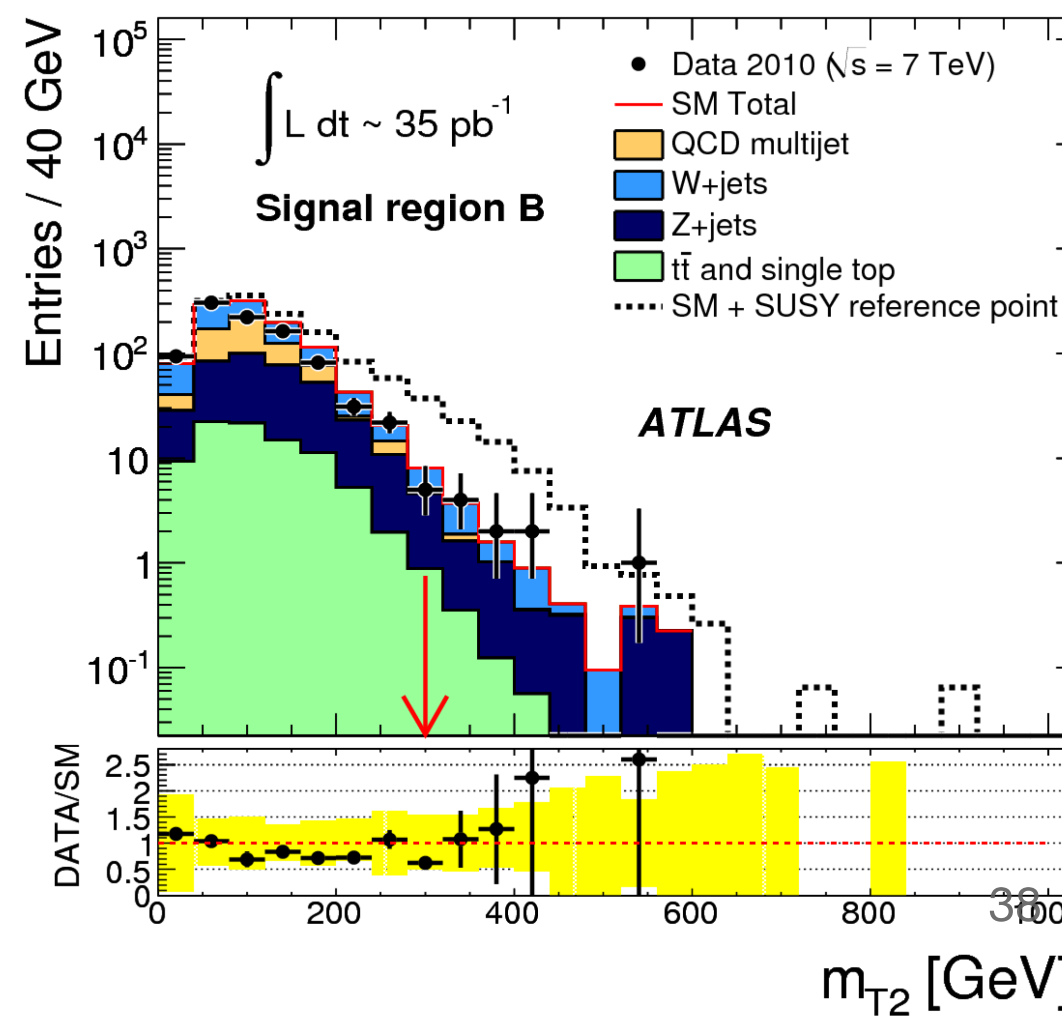
$$m_{T2}(m_{\tilde{\chi}}) = \min_{\vec{p}_T^{\tilde{\chi}1}+\vec{p}_T^{\tilde{\chi}2}=\vec{p}_T^{miss}} \left[\max\left(m_T(\vec{p}_T^{j1},\vec{p}_T^{\tilde{\chi}1}), m_T(\vec{p}_T^{j2},\vec{p}_T^{\tilde{\chi}2})\right)\right] \leq m_{\tilde{q}}^2$$

suggests a way to decompose the MET into these particles.

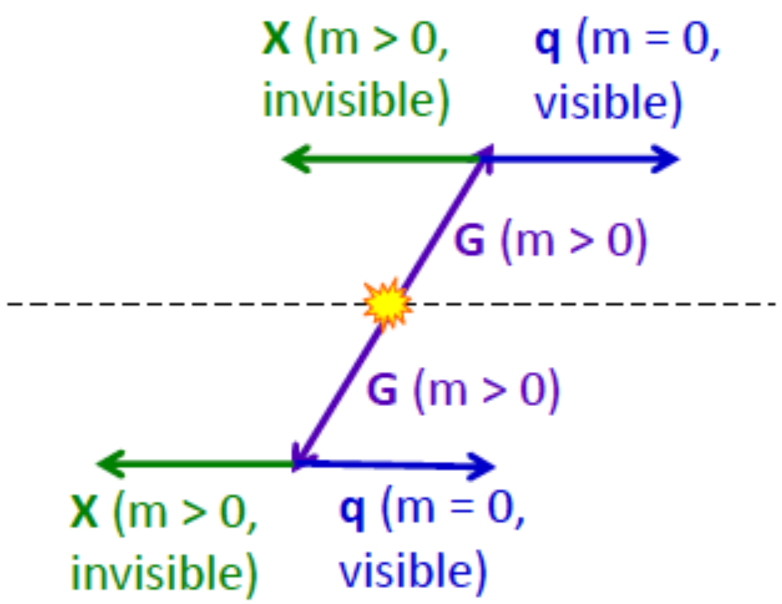
The minimization is over all possible partitions of the measured MET.

However, for massive ~χ, we need the ~χ mass for calculating m_{T2}. It is shown that for different input $m_{\sim\chi}$ values, maximum m_{T2} vs. $m_{\sim\chi}$ curve has a kink at the correct $m_{\sim\chi}$ value.

MT2 is used as a selection variable in SUSY searches in ATLAS and CMS



Razor kinematic variables



Suppose a signal with pair production of heavy particles G, each decaying to a massless visible particle χ and a massive invisible particle q.

In the G rest frame, the momentum of Q is a constant depending on heavy particle masses

$$|\vec{p}^{q}| = \frac{m_G^2 - m_\chi^2}{2m_G} = \frac{m_\Delta}{2}$$

Razor variables estimate the momentum of Q in the G rest frame using lab frame observables.

using longitudinal lab fr. observables:

$$M_R = \sqrt{\frac{(\vec{p}_z^{\,q_1} E^{q_2} - \vec{p}_z^{\,q_2} E^{q_1})^2}{(\vec{p}_z^{\,q_1} - \vec{p}_z^{\,q_2})^2 - (E^{q_1} - E^{q_2})^2}} \approx m_\Delta$$

For a signal with heavy G and χ, M_R distribution peaks at m_Δ. When there are no heavy particles M_R falls exponentially.

using transverse lab fr. observables:

$$M_T^R = \sqrt{\frac{E_T^{miss}}{2}(p_T^{q_1} + p_T^{q_2}) - \frac{1}{2}\vec{E}_T^{miss} \cdot (\vec{p}_T^{\,q_1} + \vec{p}_T^{\,q_2})} < m_\Delta$$

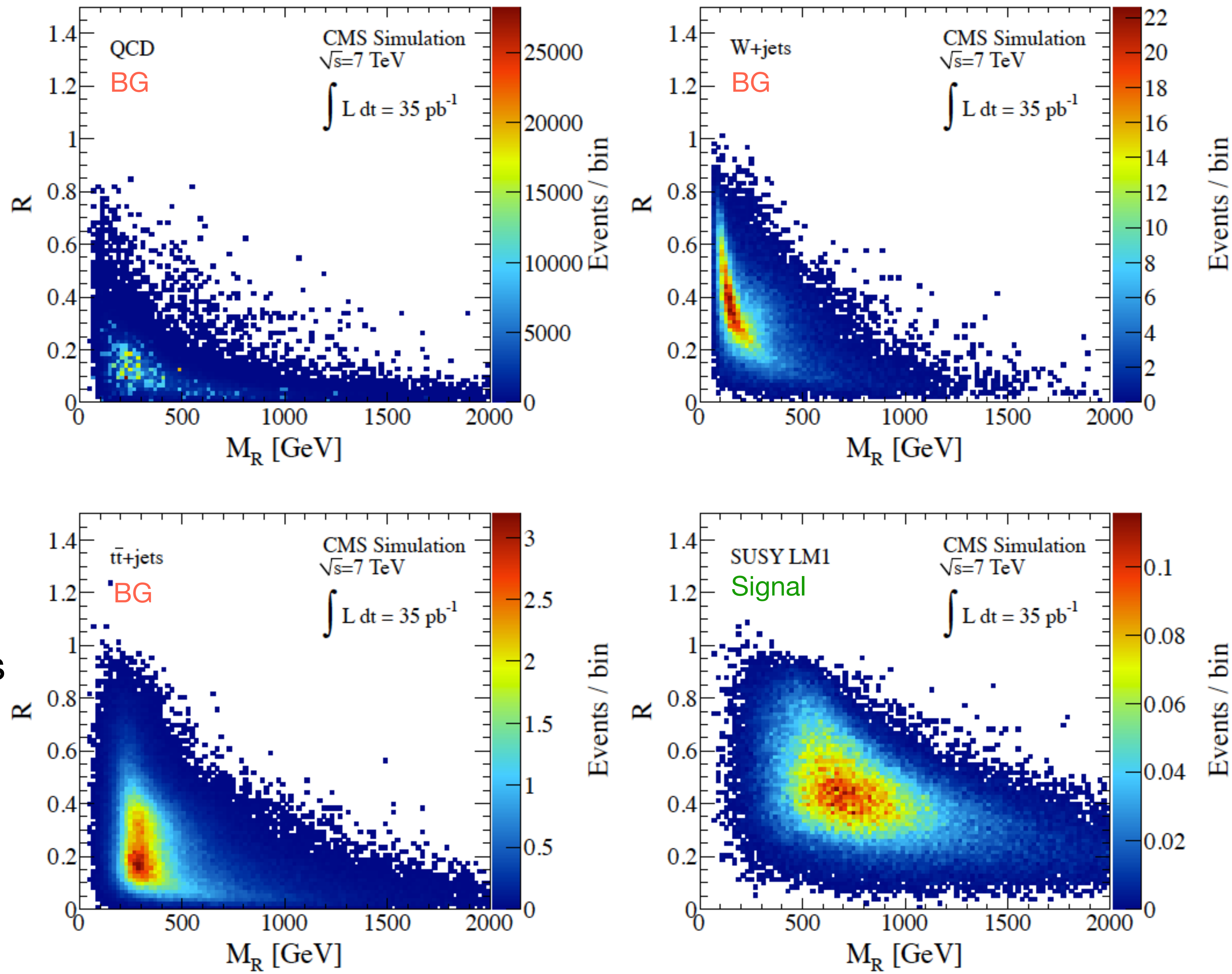
$$R \equiv M_T^R / M_R$$

M_T^R distribution has an endpoint at m_Δ.

Razor variables

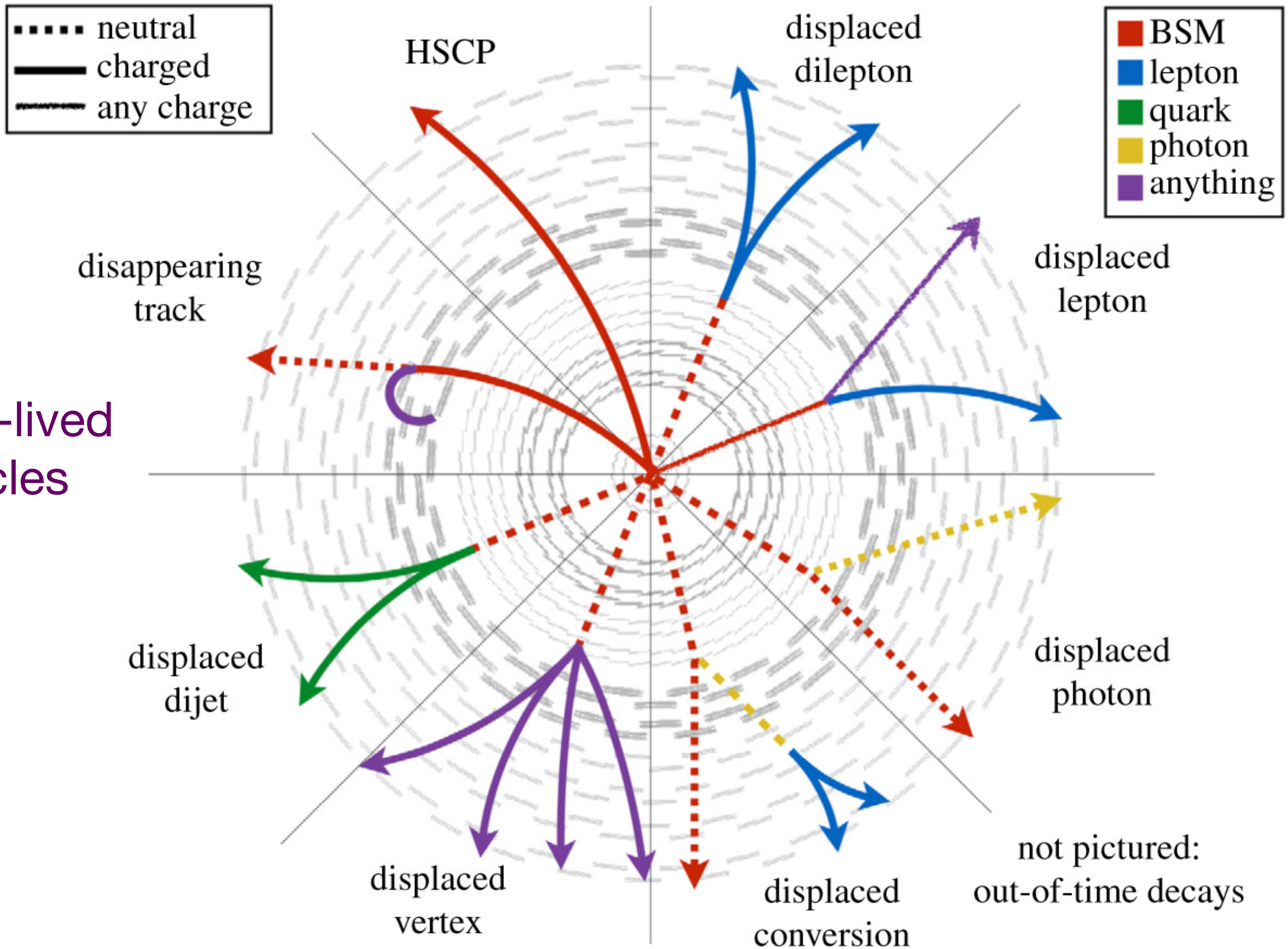
Most kinematic discriminators give an excesses in the tails (e.g. MET), but razor variables define a “bump”, hence they provide very good signal-BG discrimination.

For events with >2 visible objects, we partition these objects into 2 megajets, then compute M_R and R^2.



Long-lived particles

Long-lived particles

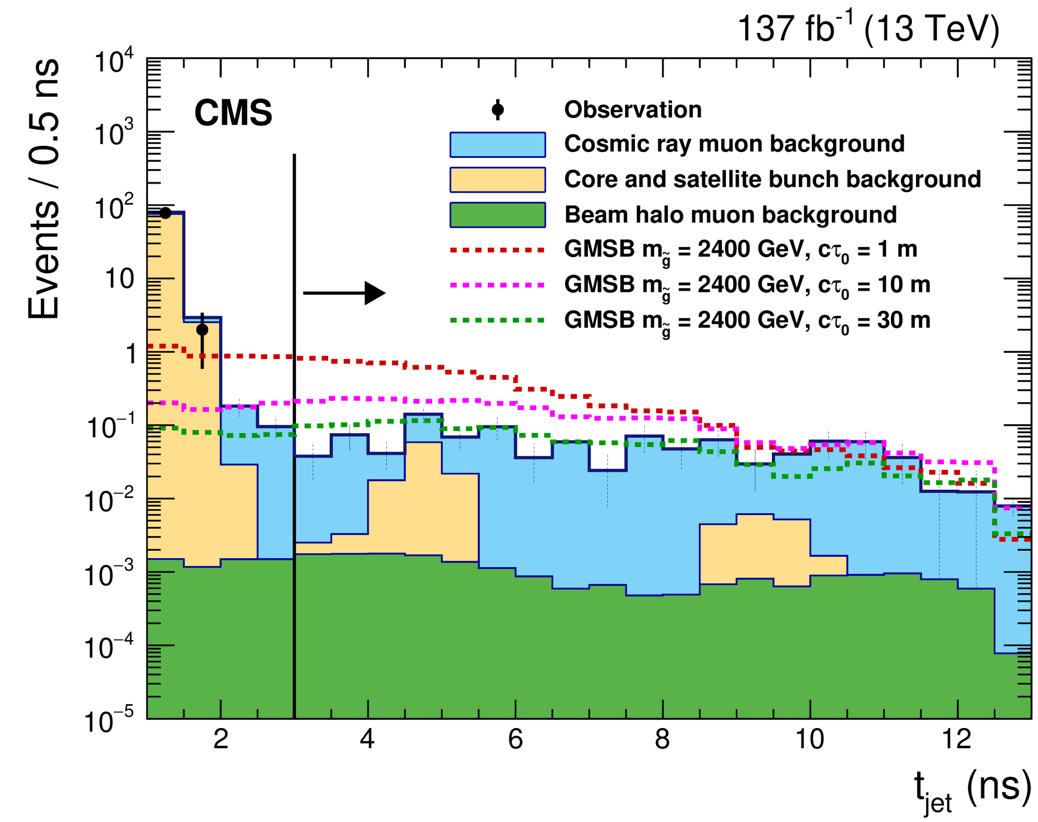


Characterizing the signal

Long-lived particles

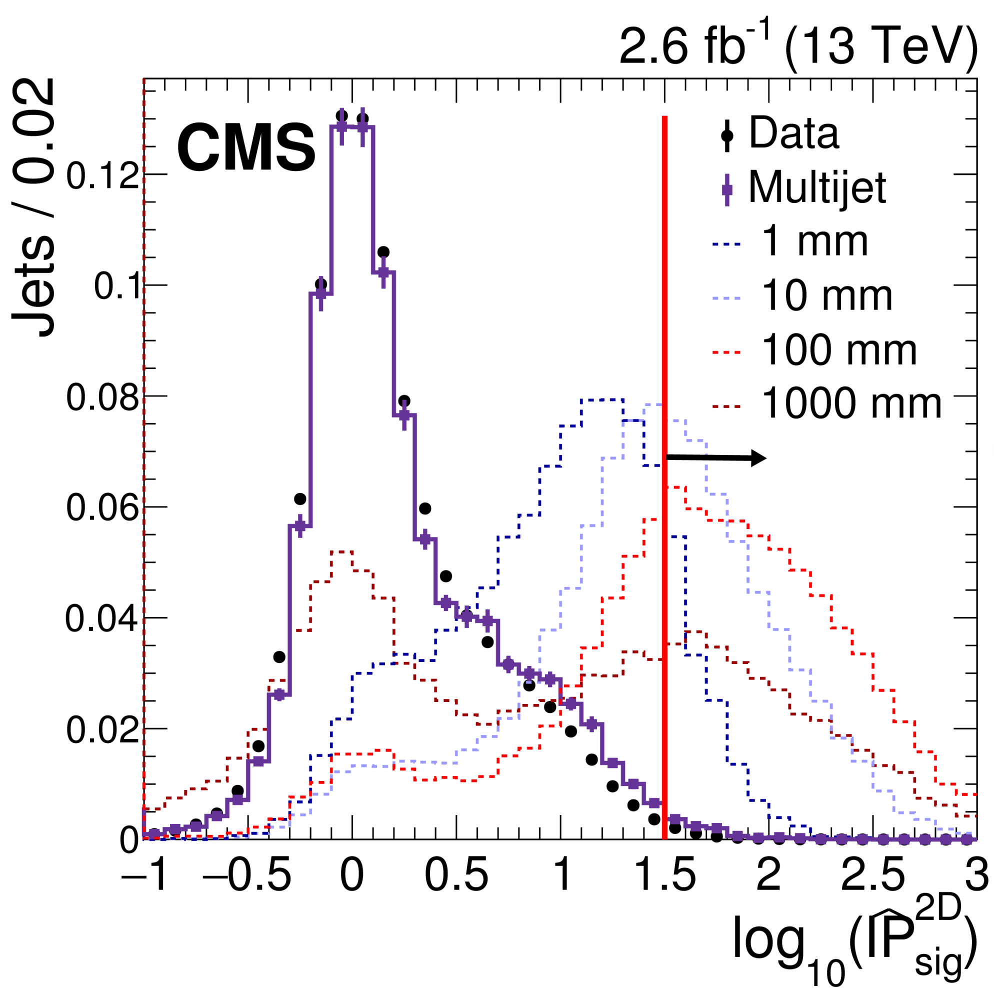
Timing information of the object.

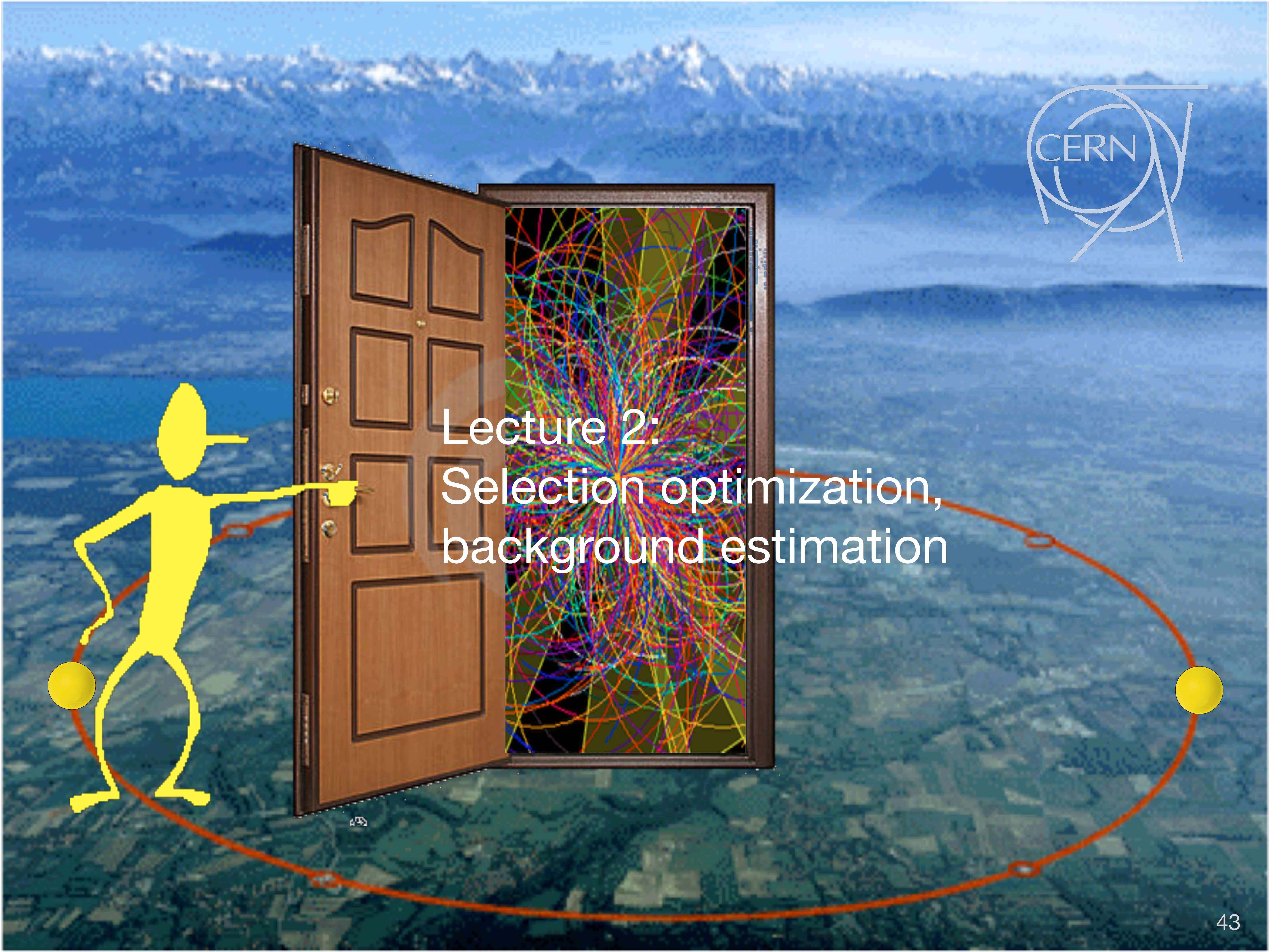
A long-lived BSM particle has bigger timing compared to SM particles.



Displacement information of the object from the interaction point.

A long-lived particle can decay far away from the interaction point.





Lecture 2:
Selection optimization, background estimation