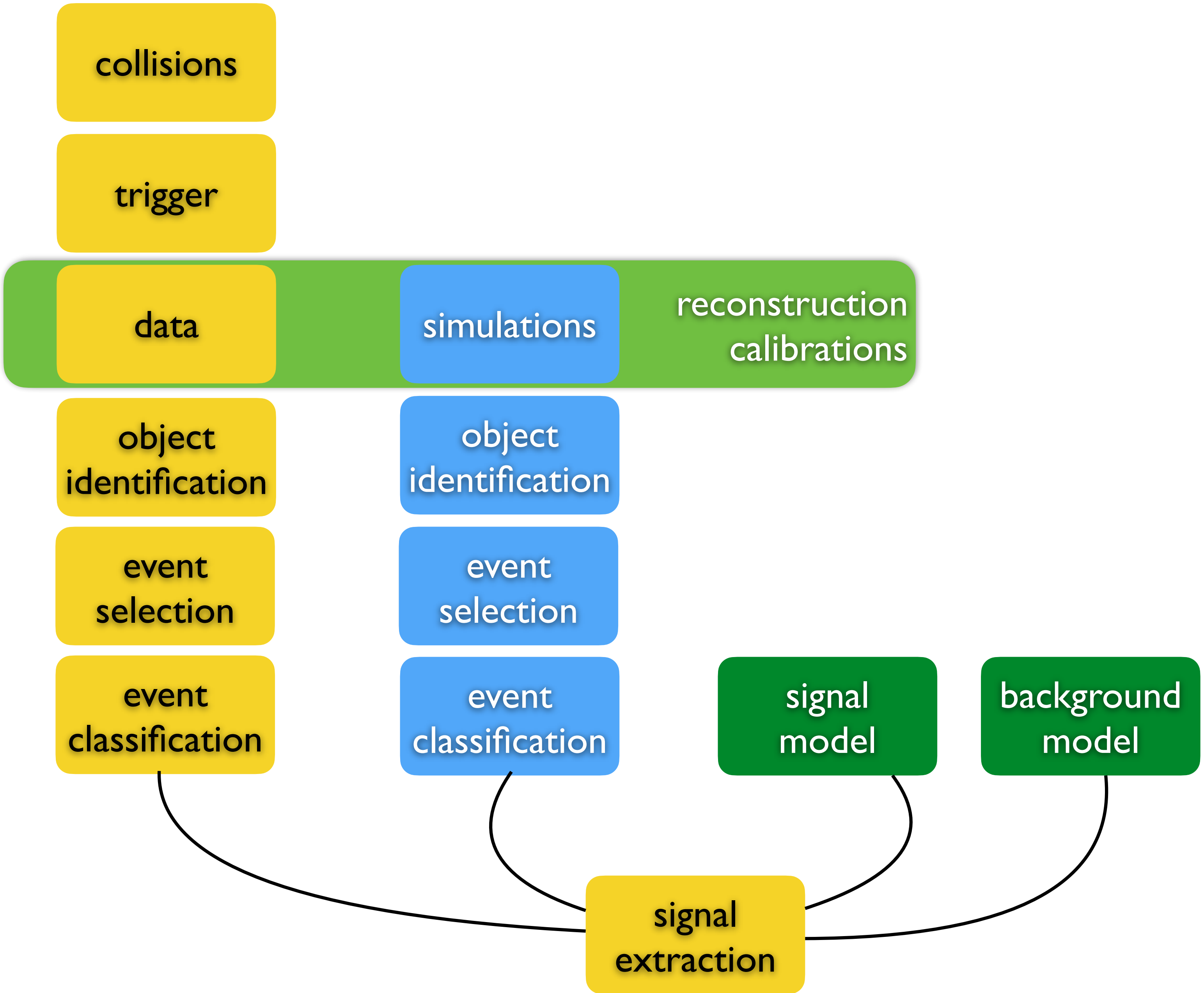
Typical HEP analysis steps

Standard analysis skeleton



Standard skeleton

collisions

@LHC 40 MHz (one bunch crossing every 25 nsec)
—> cannot store all collisions, you need to select *what you want*

Standard skeleton

collisions

trigger

Typically your trigger definition operates at two levels:

Level 1 Trigger (LV1)
working on single objects typically poor resolution :

e.g.

energy deposit in calorimeters: E_T > x GeV
muon stubs in the muon chambers: p_T > y GeV

40MHz —> 100 Khz
(events are buffered and you have3.2 μs to take a decision)

The events that pass the L1 selection will be filtered by another trigger layer

High Level Trigger (HLT):
reconstruct the full event (online resolution)

e.g.

photons pT> x GeV
muon pT > y GeV
jets > z GeV
MET > w GeV

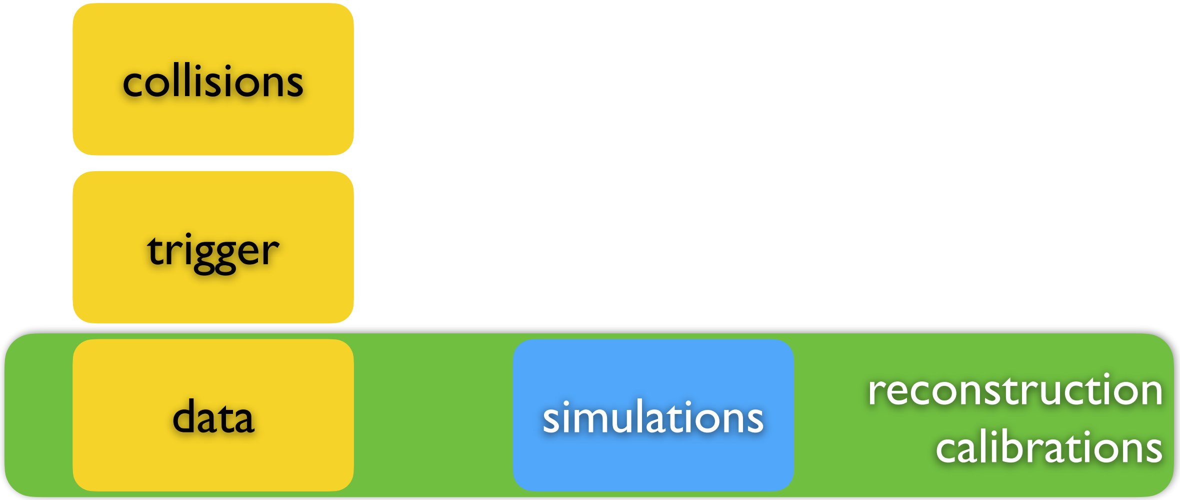
100kHz —> 1 KHz (200 msec to take a decision)

If an event is discarded it is lost forever…

Standard analysis skeleton

Once an event is accepted you reconstruct it with the best available algorithms and calibrations (offline resolution) and store it. This is (one of) your dataset

Standard analysis skeleton



You will need several simulated data set to:

- build a model of the signal you are searching / measuring
- (model background)
- tune object selection
- etc…

Simulations in HEP

Generator: from theory to final states. E.g. gluon gluon —> H —> photons

Hard scatter (matrix elements + pdf), initial/final state radiation, parton showering / hadronization and decays to the final states

Compute the full kinematics of the process and get the distribution of the observables (typically impossible to derive with analytical methods)

Some examples: Pythia8 and Herwig, Powheg, MadGraph5_aMCatNLO, Algpen , …

—> particle level events

Detector simulation: “simulation + digitization”

Simulation = propagate the particles through the detector simulating their interactions (energy deposits) with all the sensitive and passive materials. This is done using “physics libraries” describing the interactions

Some examples: Geant 4, FLUKA, MARS, etc…

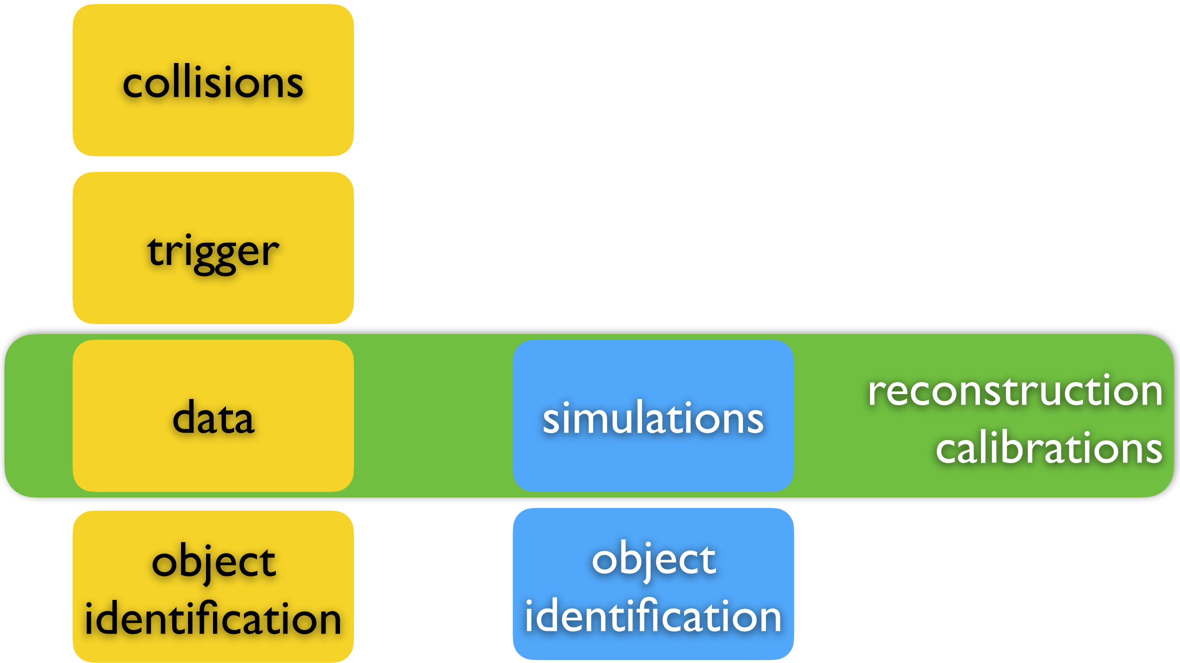
Digitization = simulate how the energy deposited in the different sensitive materials is transformed into electrical/optical signals (deterministic effects (amplifiers, shapers, …) and random effects (noise, calibration inaccuracies, …)

o(100 sec)/ ttbar event

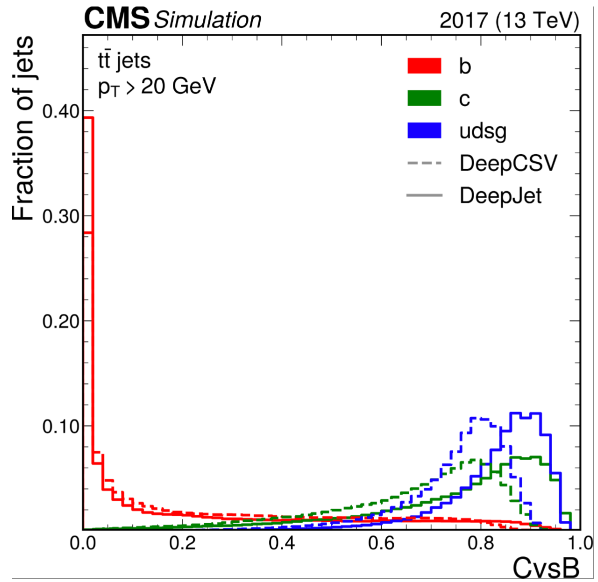
Reconstruction: low level objects: hits/clusters reconstruction

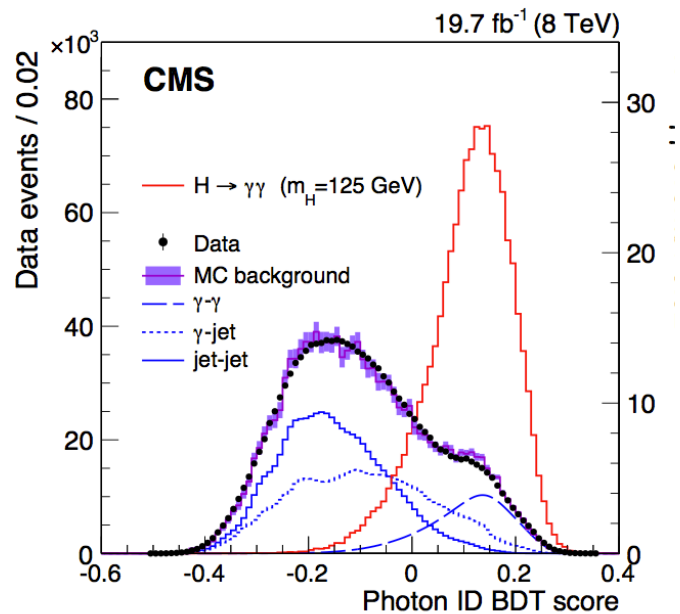
high level objects: tracks, energy superclusters, jets, etc…

Standard analysis skeleton

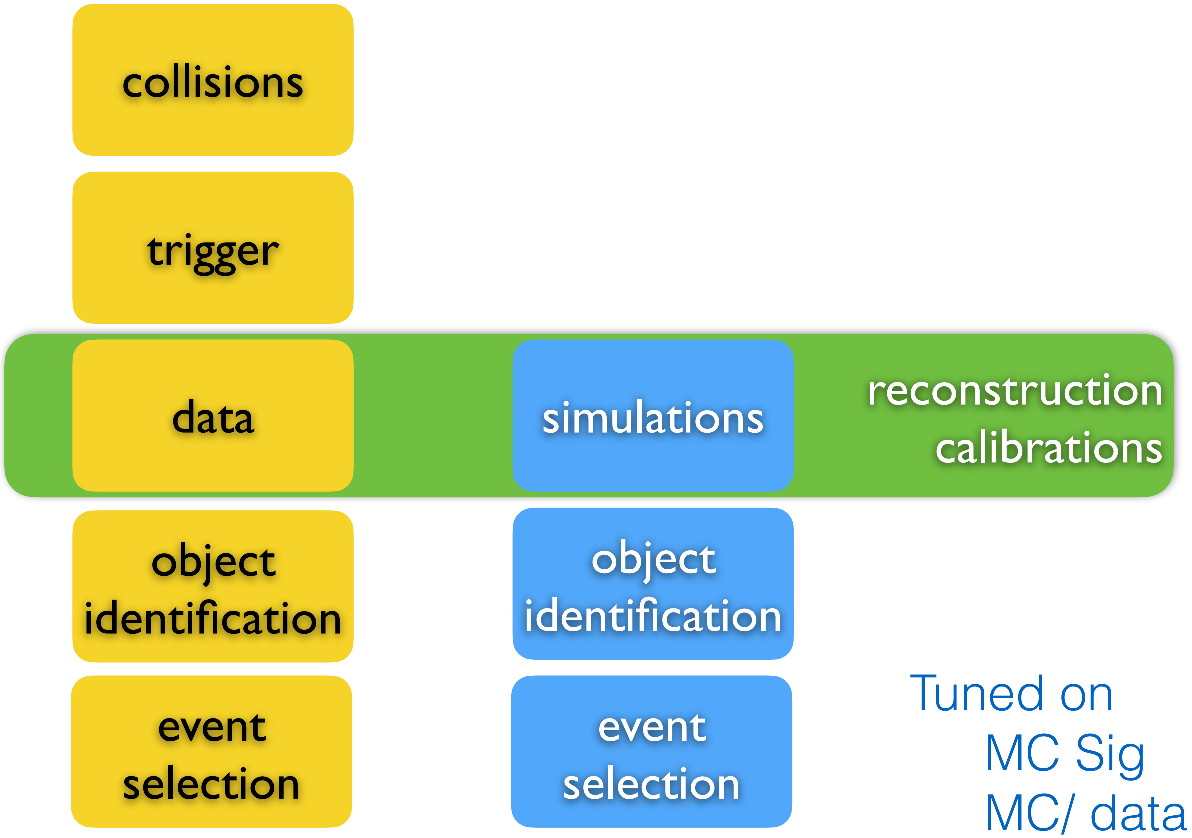


photon, muon, electron, tau identification
b/c tagging vs light jets
. . .





Standard analysis skeleton



Tuned on

MC Sig

MC/ data Bkg

Example H—>WW event selection:

2 oppositely charged isolated leptons

(e/mu min pT > 20/10 GeV)

Missing Transverse Energy from neutrinos (>20GeV)

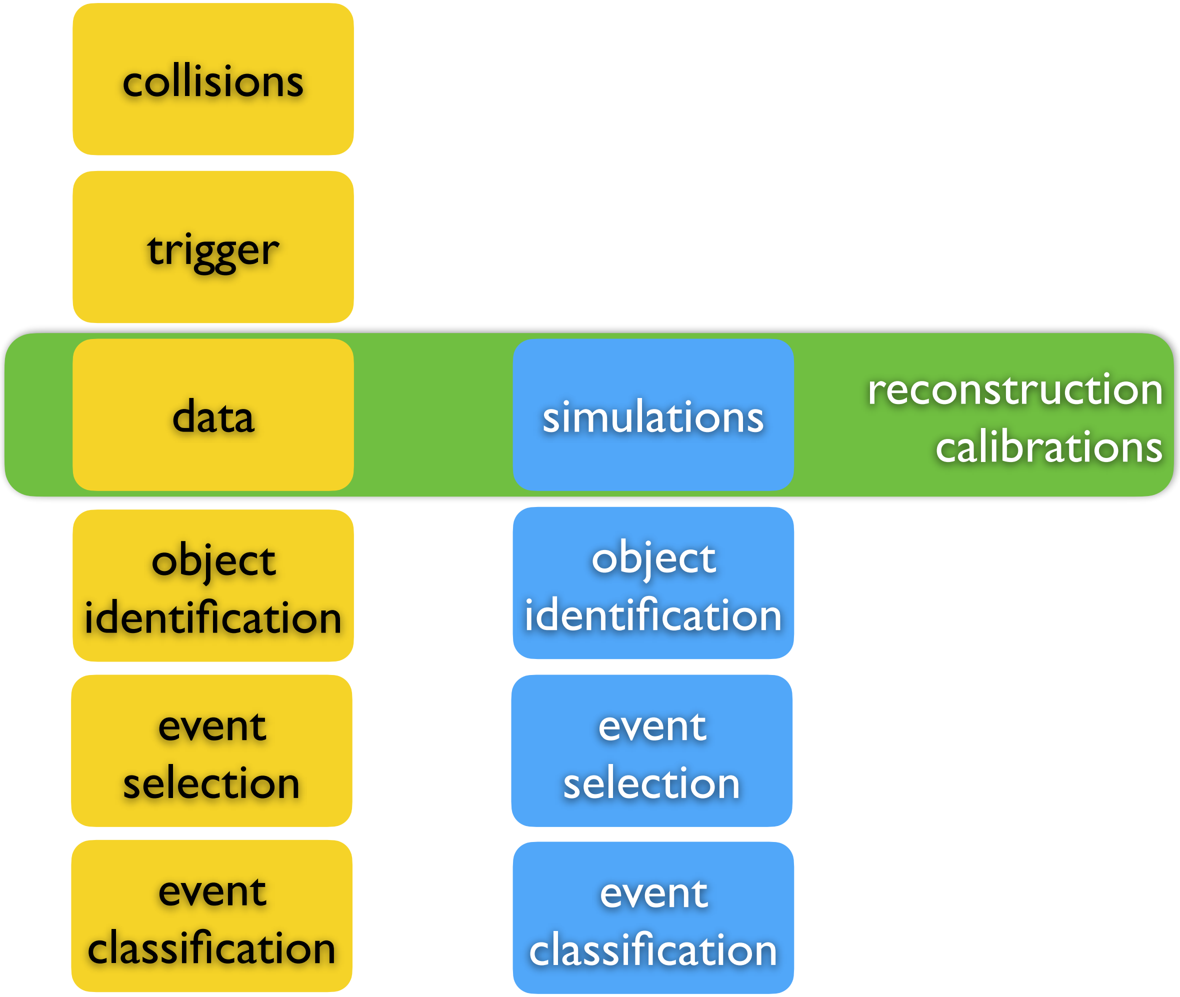
jets pT > 30 GeV (veto b-jets)

m_{ll} > 12 GeV

pT_{ll} > 30 GeV

mT > 30 GeV where $m_T^2 = 2 p_T^{\ell\ell} E_T^{miss}(1 - \cos\Delta\phi(\ell\ell, \vec{E}_T^{miss}))$

Standard analysis skeleton



Split events into mutually exclusive categories allows one to improve the analysis sensitivity

Events classification

Classify events allows the analysis to achieve better sensitivity / better parameters constraints. The reason is that you can make more accurate assumptions on how to model the data in one category (locally) than on the overall sample (globally).

Assume you have large enough statistics, the significance (see later in this lectures) is

given by $Z = \dfrac{S}{\sqrt{B}}$

Now, suppose you data in two categories: $Z_i = \dfrac{S_i}{\sqrt{B_i}}$ i = 1,2 (S_i, B_i>0)

The combined statistical significance of the two categories is: $Z_{cat} = \sqrt{\dfrac{S_1^2}{B_1} + \dfrac{S_2^2}{B_2}}$

The statistical significance of the signal without categories is: $Z = \dfrac{S_1 + S_2}{\sqrt{B_1 + B_2}}$

(total Signal = S_1+S_2, Total background B_1+B_2)

If you compare the two: $Z_{cat}^2 - Z^2 = \dfrac{S_1^2}{B_1} + \dfrac{S_2^2}{B_2} - \dfrac{(S_1 + S_2)^2}{B_1 + B_2} = \dfrac{B_1 B_2}{B_1 + B_2}\left(\dfrac{S_1}{B_1} - \dfrac{S_2}{B_2}\right)^2$

which is always > 0 unless S1/B1 = S2/B2. From a statistics point of view you always improve your analysis sensitivity by categorising the events.

Standard analysis skeleton