Karlsruher Institut für Technologie

Vorlesung 13a:

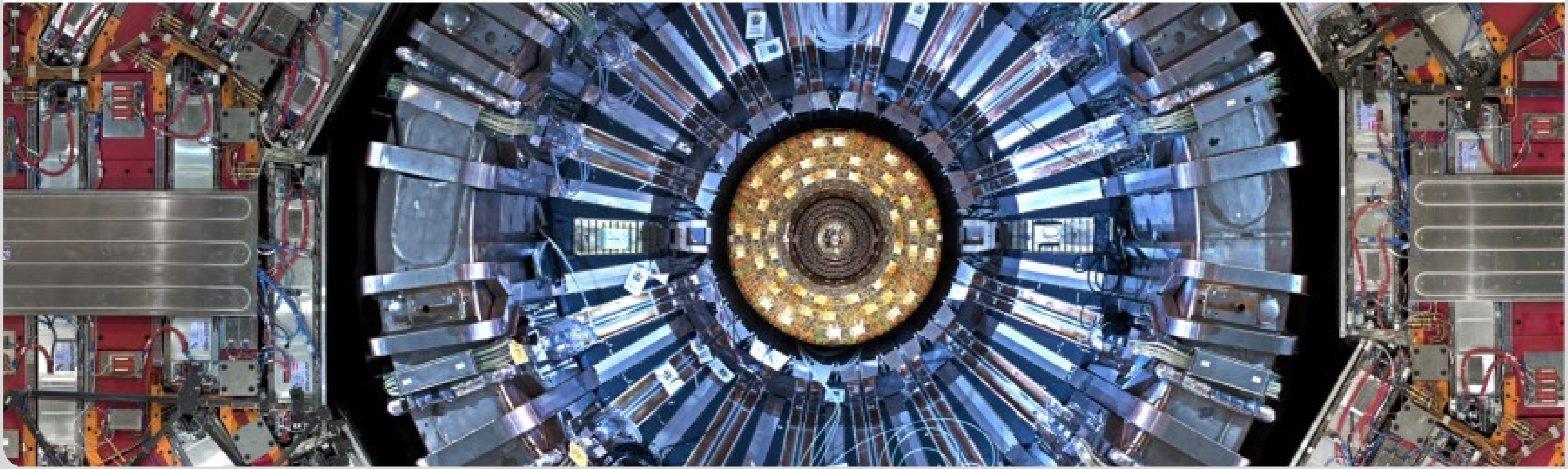
Teilchenphysik I (Particle Physics I)

Analysis Chain

Günter Quast

Fakultät für Physik
Institut für Experimentelle Kernphysik

WS20/21



Organisatorisches

Online-Umfrage zu Vorlesung und praktischen Übungen

vom **14. - 23.12. 2020**

Wir bitten um reges Feedback – insb. zu den neuen digitalen Formaten.

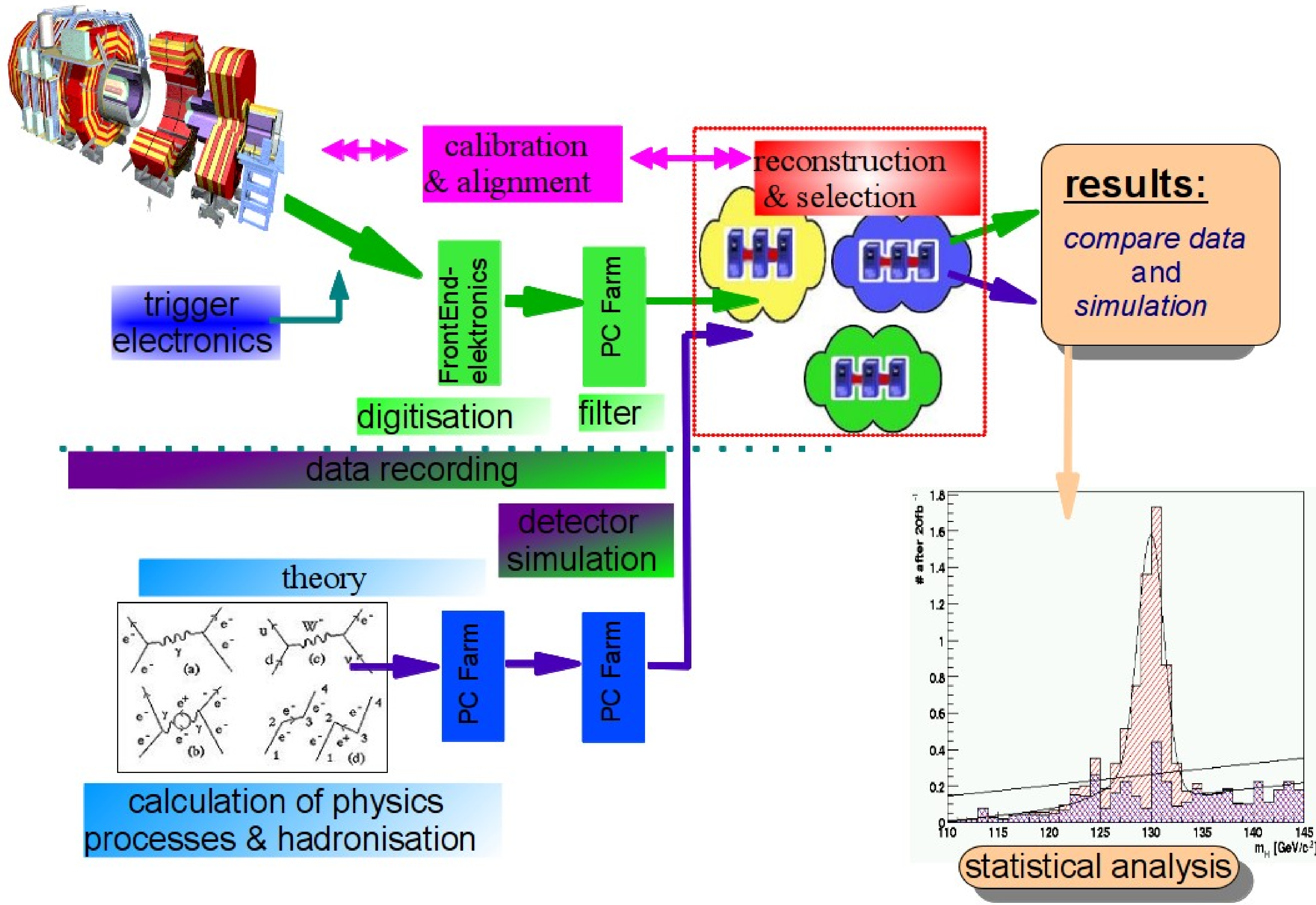
Link: Umfrage zur Vorlesung

Link: Umfrage zu den praktischen Übungen

Die Links finden Sie auch auf den Ilias-Seiten zu Vorlesung und Übung

1. History
2. Basics principles
3. Detectors and Accelerators
4. Theoretical Foundations
    1. Relativistic Quantum Mechanics
    2. Quantum Field Theory and symmetries
    3. Elctroweak Symmetry and Higgs Mechanism
5. QCD and Jets
6. Anlaysis Chain
7. Flavour Physics
    1. Quark mixing and CKM Matrix
    2. Meson-Antimeson Mixing
    3. CP Violation
8. Tests of Electroweak Theory
    1. Discoveries
    2. Precicion Physics with Z bosons
    3. W Boson and electroweak fit

Overview: Components of Analysis Chain



Components of Analysis Chain

- **Digitizers** record data from detector cells
  - remove empty cells („zero-suppression“ and „noise reduction“
- **Trigger** and **Filter** select „interesting“ events „on-line“
  to be stored for „off-line“ analysis
  (events not stored at this point are lost forever !)
- **Reconstruction** process constructs physical objects
  (electrons, muons, jets, …)
  (this and subsequent steps can be repeated many times)
- **Pre-selection** identifies interesting events and objects in
  events for further processing and analysis
- **Analysis** compares measured distributions with
  theoretical expectations

Theory

- **theoretical calculation** of production cross sections
- **hadronisation** of quarks and gluons into jets

Experiment

- **Detector simulation**

same reconstruction, selection and analysis steps
for simulated events as for real events

Reprise: Cross section

cross section:

transition rate initial → final state

in theory	experimentally

Fermi's golden rule

$$\lambda_{i\to f} = 2\pi|M_{fi}|^2\rho$$

amplitude or
"matrix element"
of underlying process

phase space

Cross Section

$$\sigma = \frac{|\mathcal{M}|^2 \cdot [\text{Phase space}]}{[\text{Colliding particle flux}]}$$

$$\sigma = \frac{N_{cand} - N_{bkg}}{\epsilon \cdot \int L}$$

N_{cand}: number of observed events
N_{bkd}: number of expected background events
ε : accetance · efficiency
L: Luminosity

by error propagation

$\rightarrow$

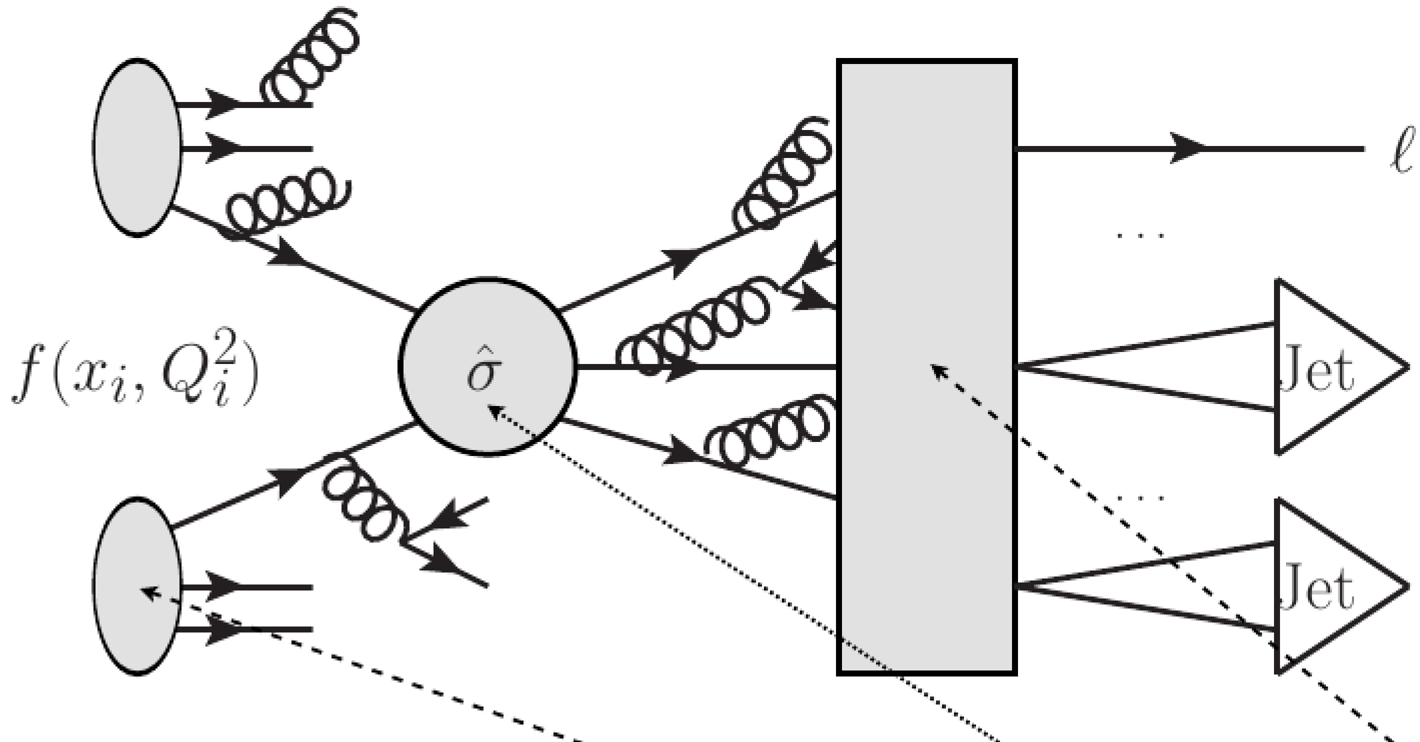
$$\frac{\delta\sigma}{\sigma} = \sqrt{\frac{\delta N_{cand}^2 + \delta N_{bkg}^2}{(N_{cand} - N_{bkg})^2} + \left(\frac{\delta\epsilon}{\epsilon}\right)^2 + \left(\frac{\delta\int L}{\int L}\right)^2}$$

This is the error you want to minimize

- with signal as large as possible
- background as small as possible
- nonetheless, want large efficiency
- luminosity error small (typically beyond your control, also has a “theoretical” component)

Recap: Calculation of Cross sections

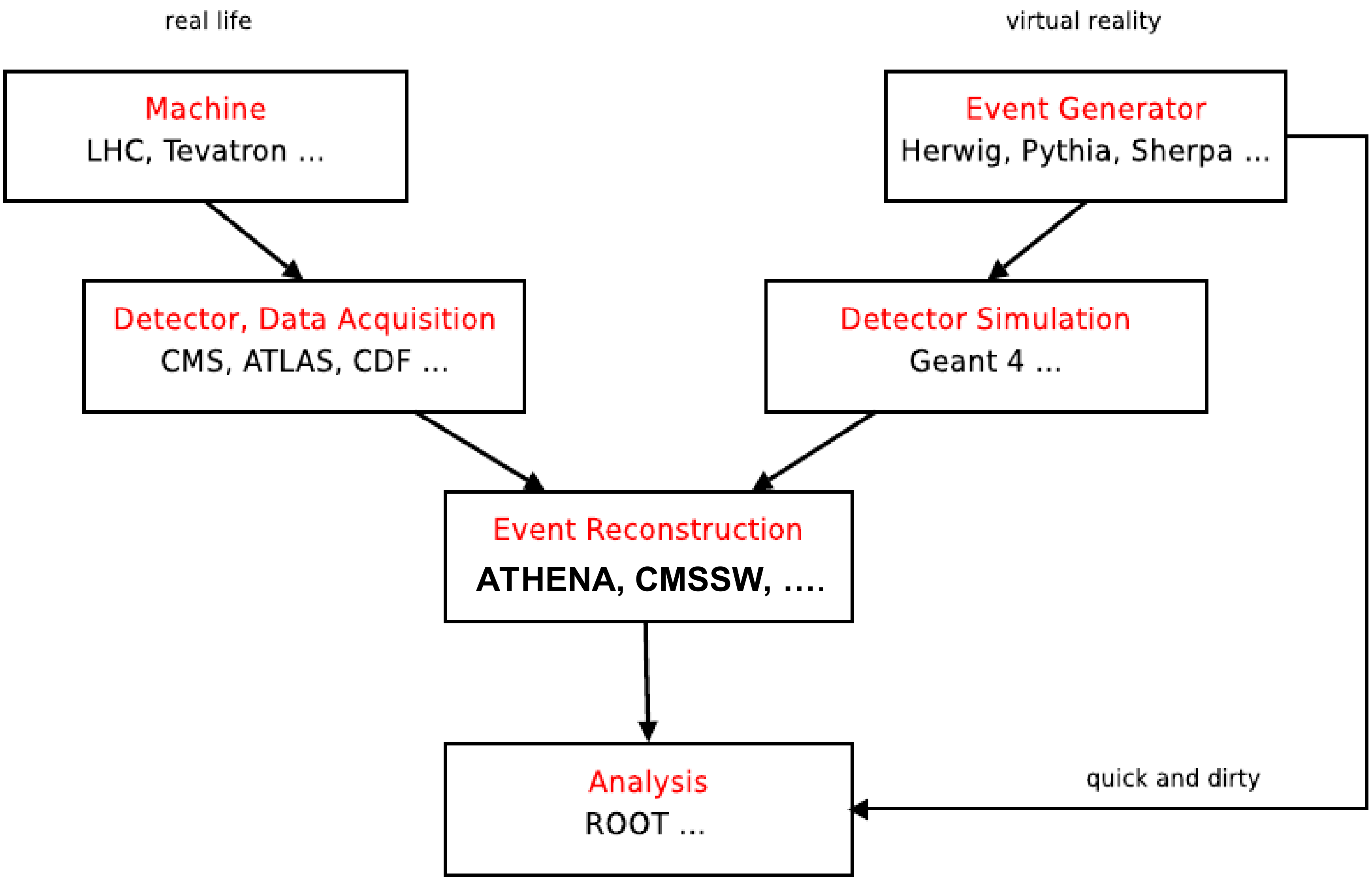
σ = PDFs ⊗ 2→n process ⊗ hadroniszation



$$\sigma_{QCD} = \sum_{jk} \int dx_j \, dx_k \, \boxed{f_j(x_j, \mu_F^2)\, f_k(x_k, \mu_F^2)} \cdot \boxed{\hat{\sigma}(x_j x_k s, \mu_F^2, \mu_R^2)} \otimes \boxed{\text{hadronization}}$$

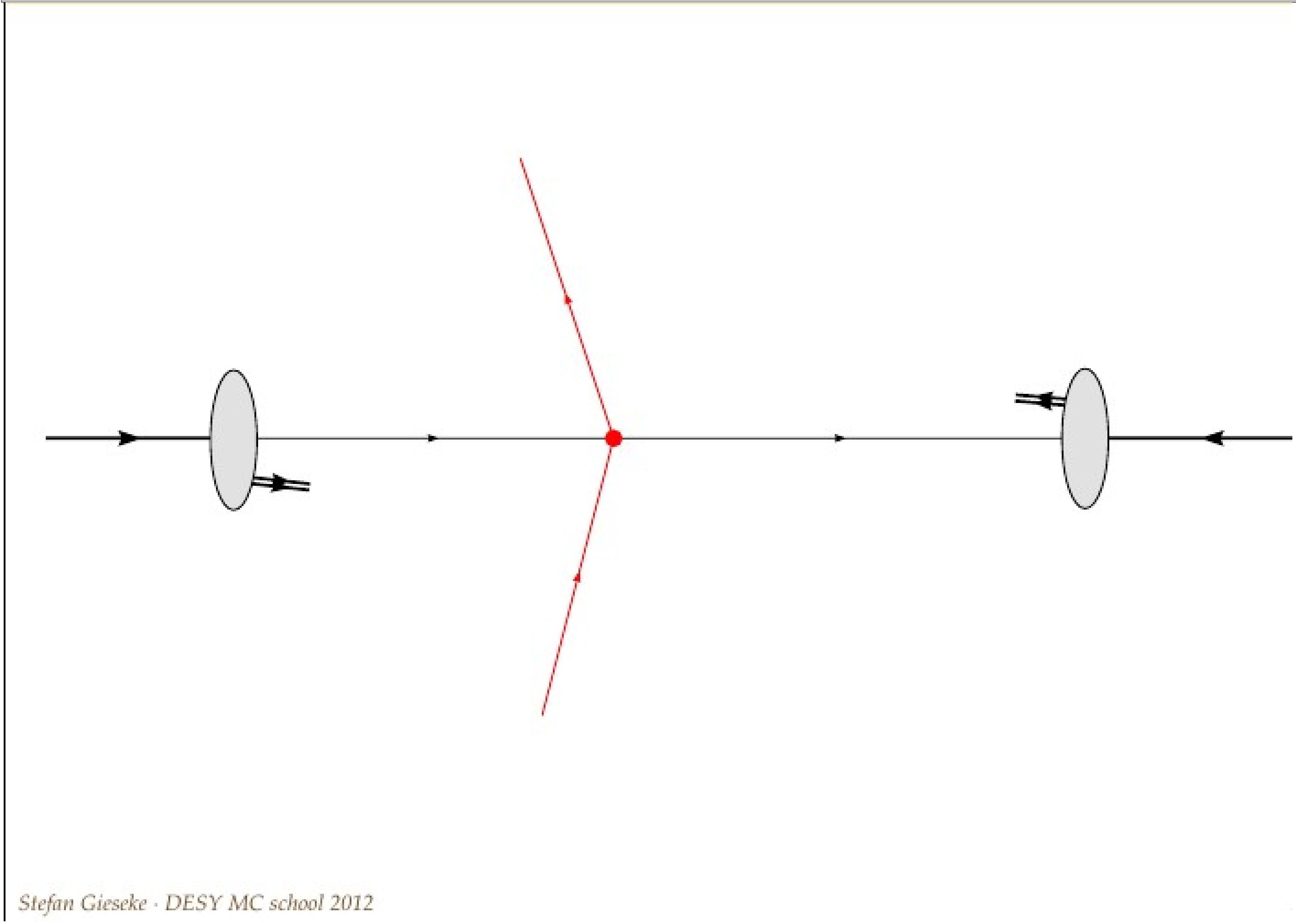
Complicated process – use **MC techniques** to calculate cross sections, **phenomenological modes** to describe hadronization process (quarks → jets)

Steps of MC simulation



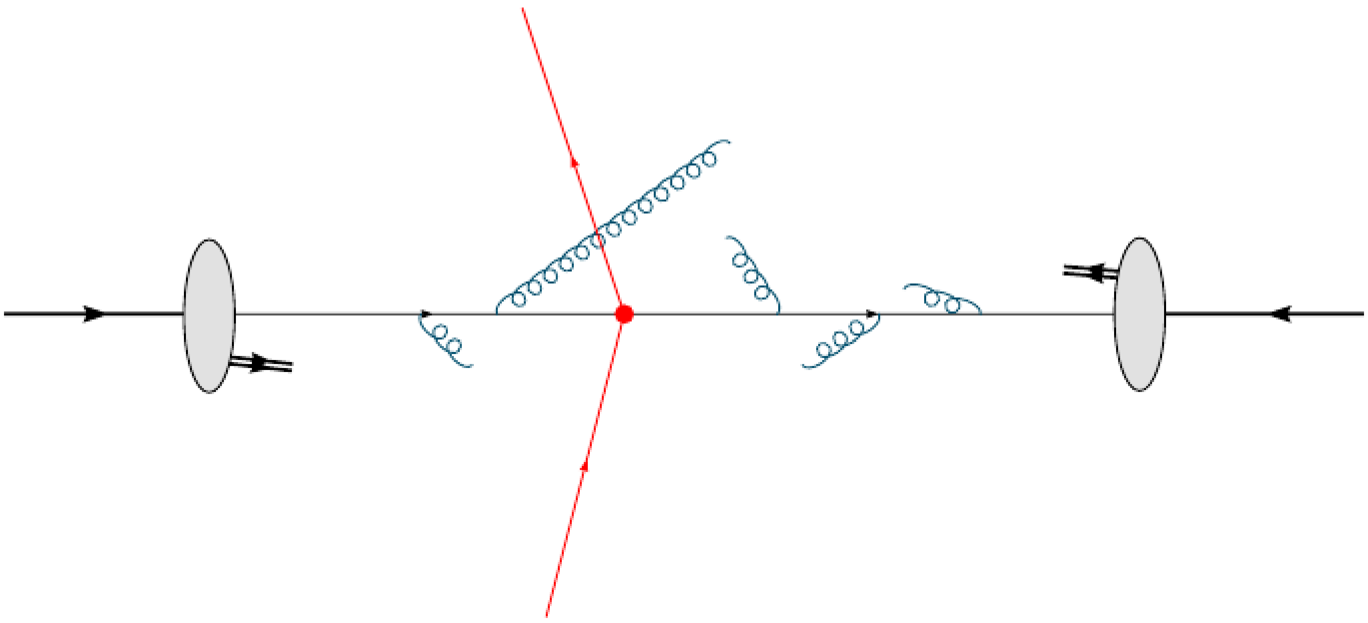
Example: pp collision

Example: pp collision



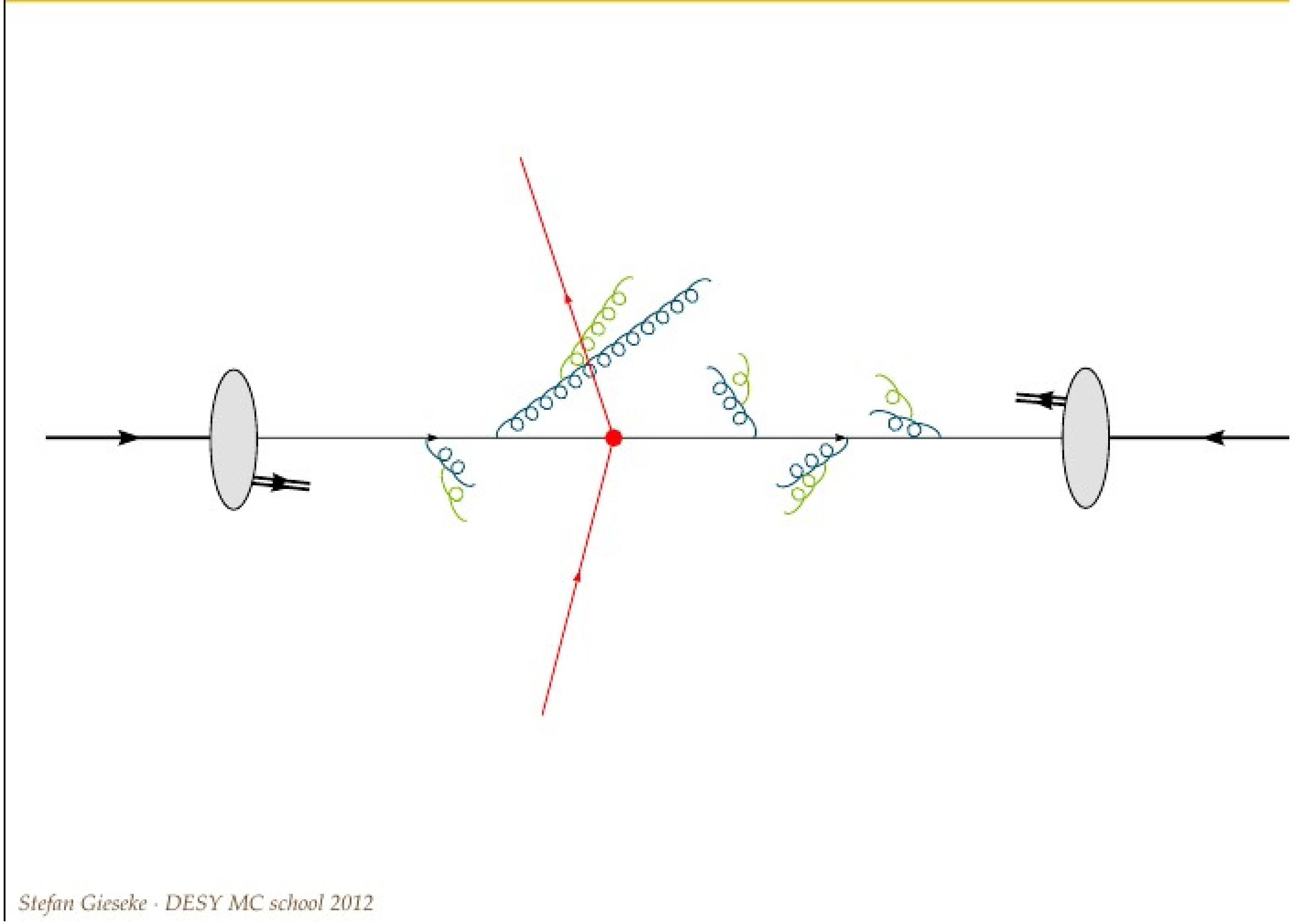
matrix element of hard process

Example: pp collision



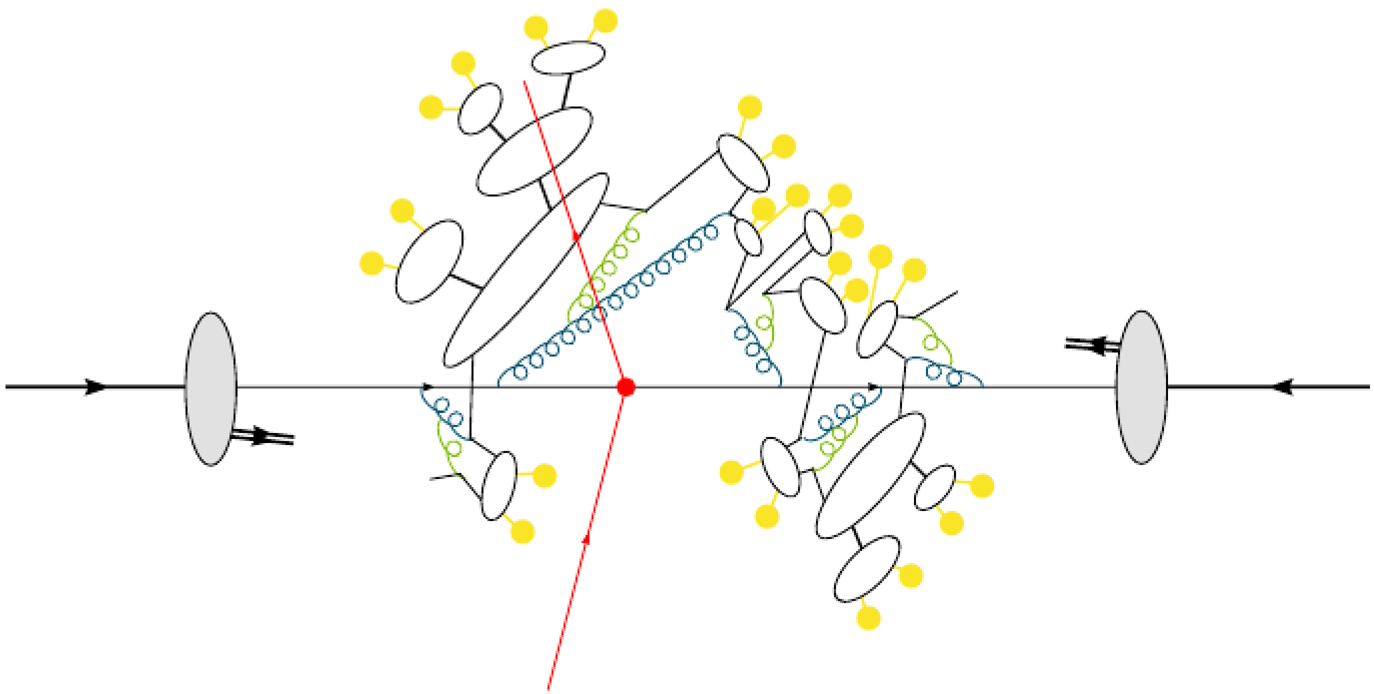
parton shower

Example: pp collision



parton shower

Example: pp collision

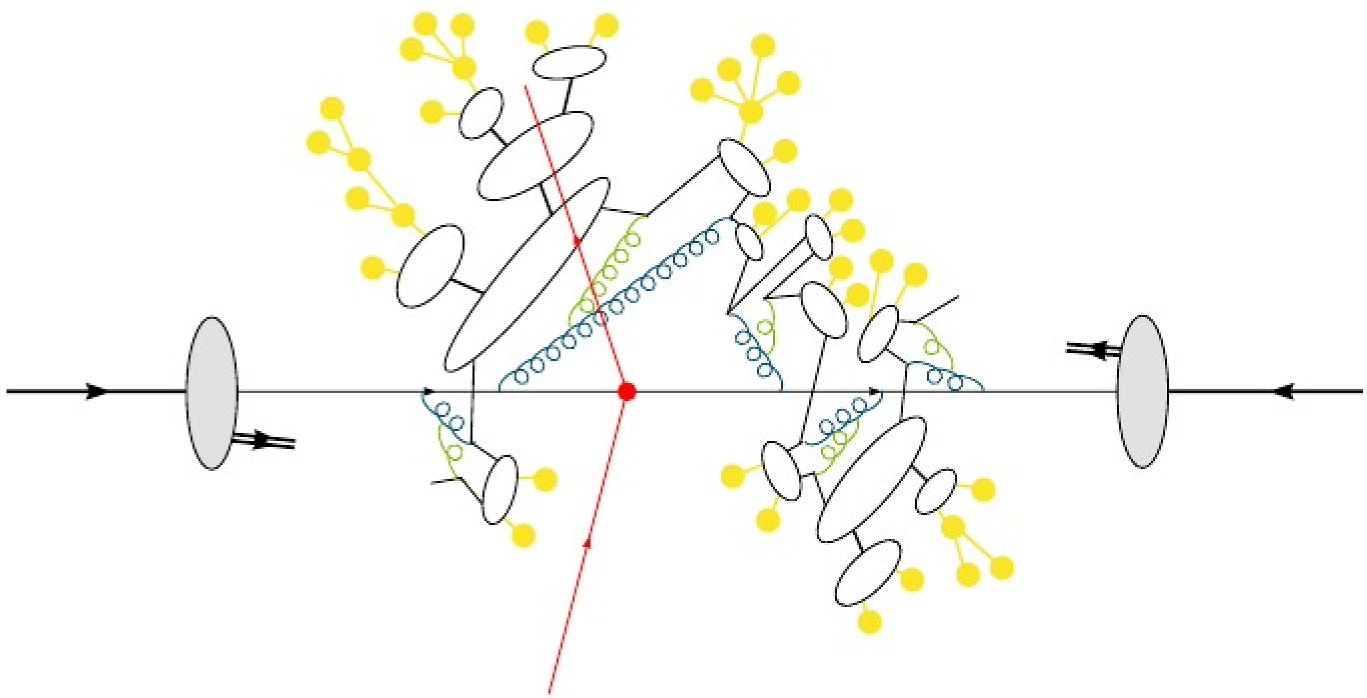


Stefan Gieseke · DESY MC school 2012

hadronization

phenomenological:
Lund string model
(Pythia)
or
cluster hadronisation
(Herwid(++))

Example: pp collision

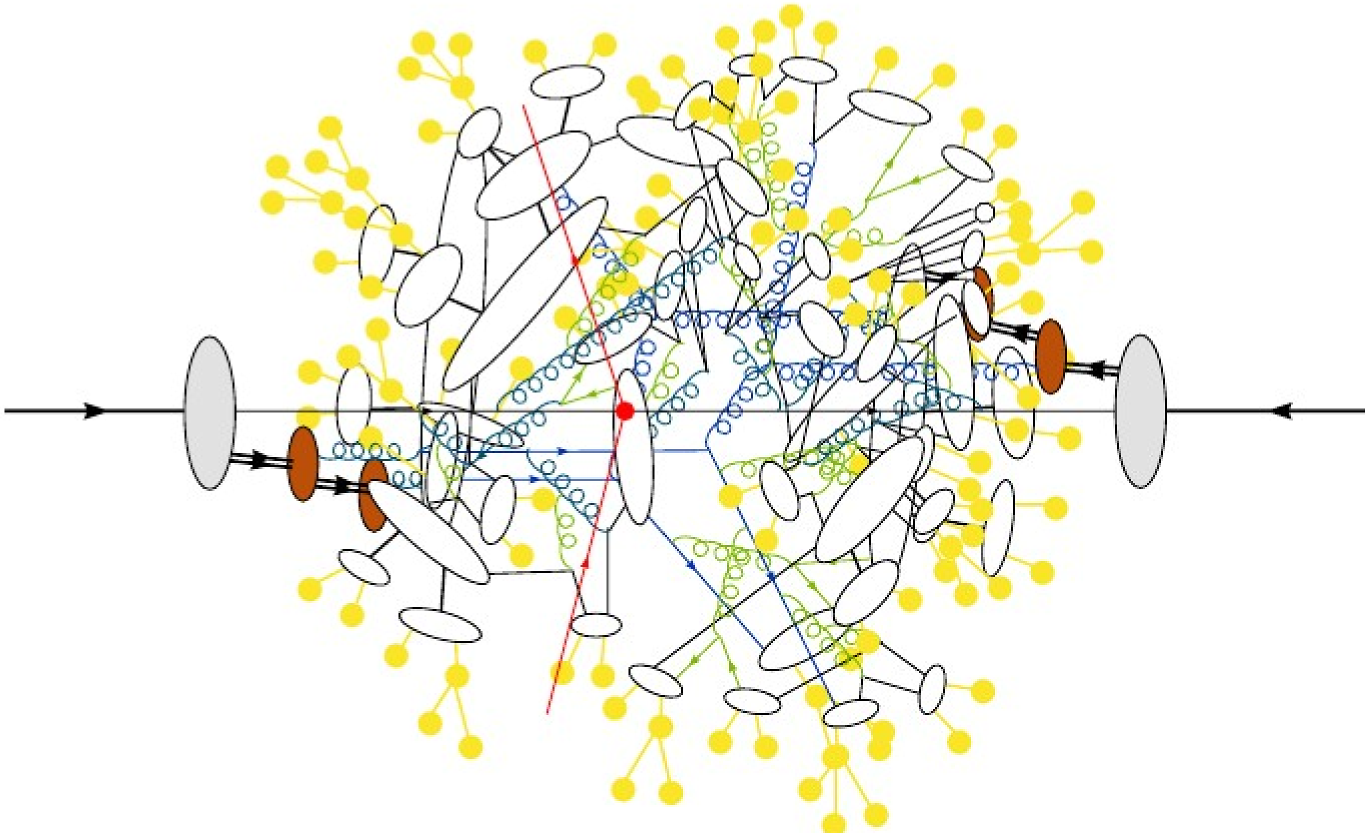


Stefan Gieseke · DESY MC school 2012

hadron decays

tedious -
relies on
measurements

Example: pp collision

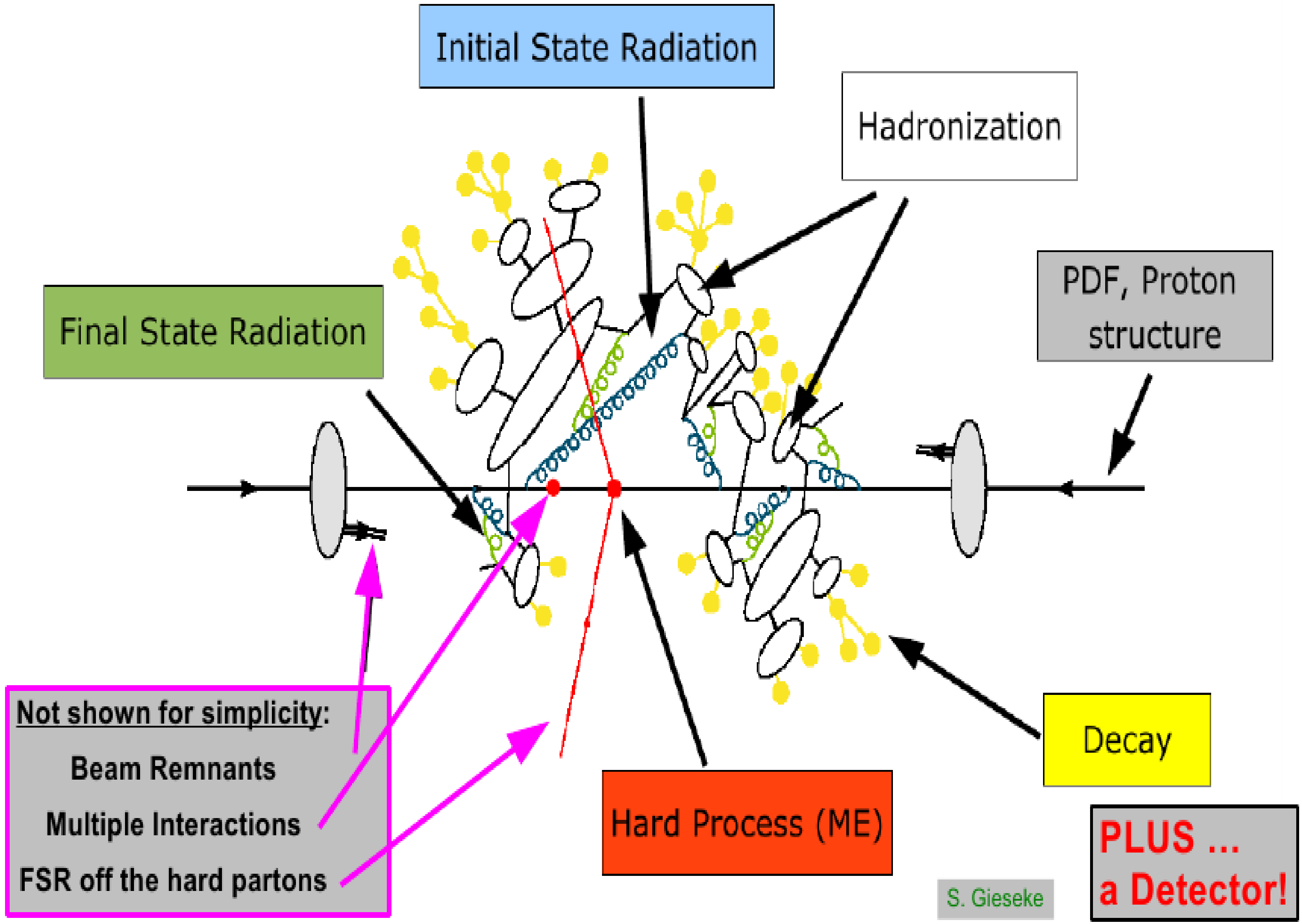


Stefan Gieseke · DESY MC school 2012

relies on models
& measurements
→ needs „tunig“

Multi-parton interactions and underlying event

Summary: pp collision



Example: pp collision

last step:

- process stable particles through detector simulation
  to obtain „hits“ in detector cells;
- run reconstruction software
  to obtain „reconstructed objects“
- run selection procedures („Analysis“)
  to obtain „identified reconstructed objects“

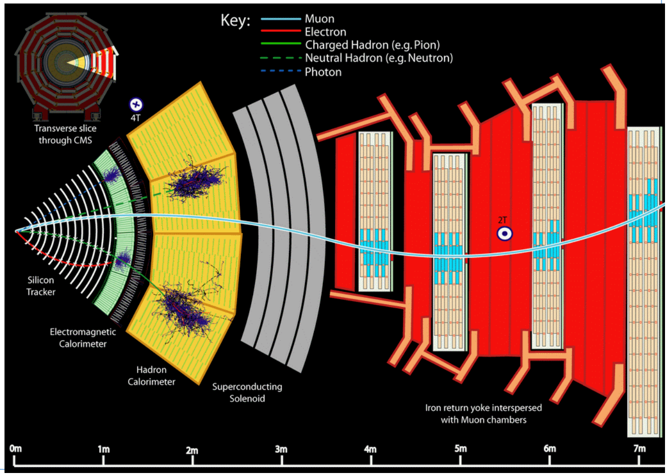
see Lecture 4:
GEANT 4

in total:

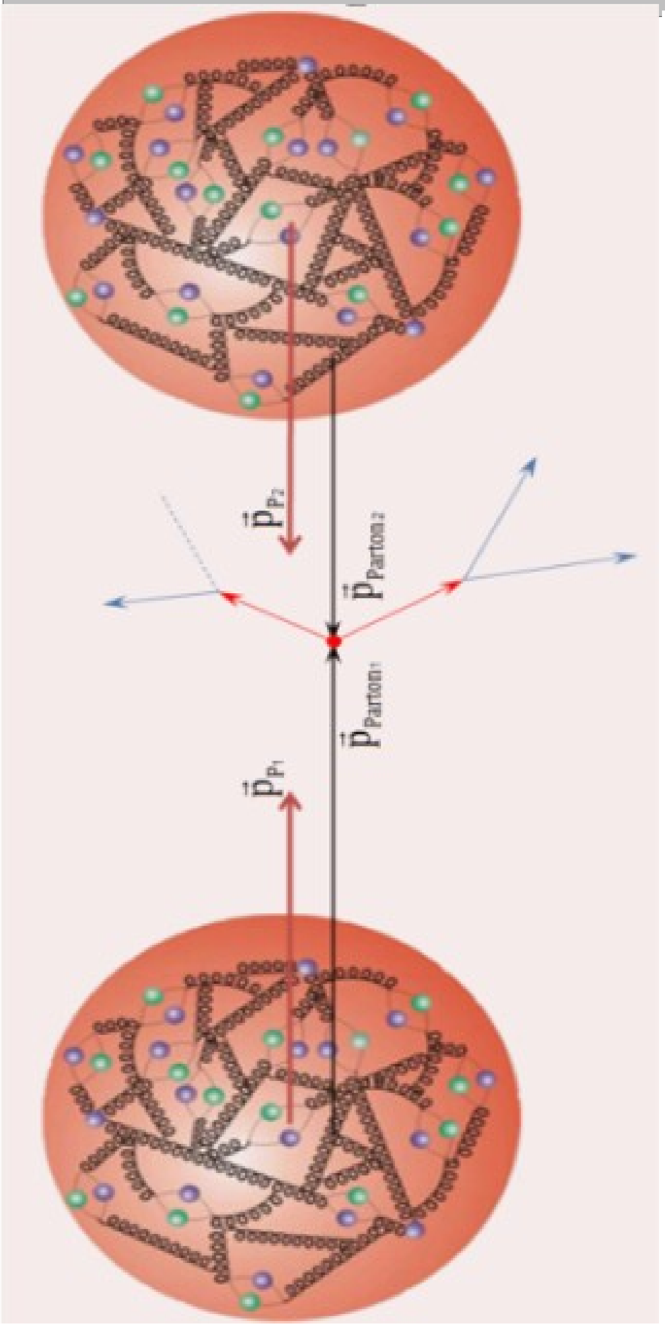
<u>true properties</u> of objects from hard process at parton level
are **folded** with

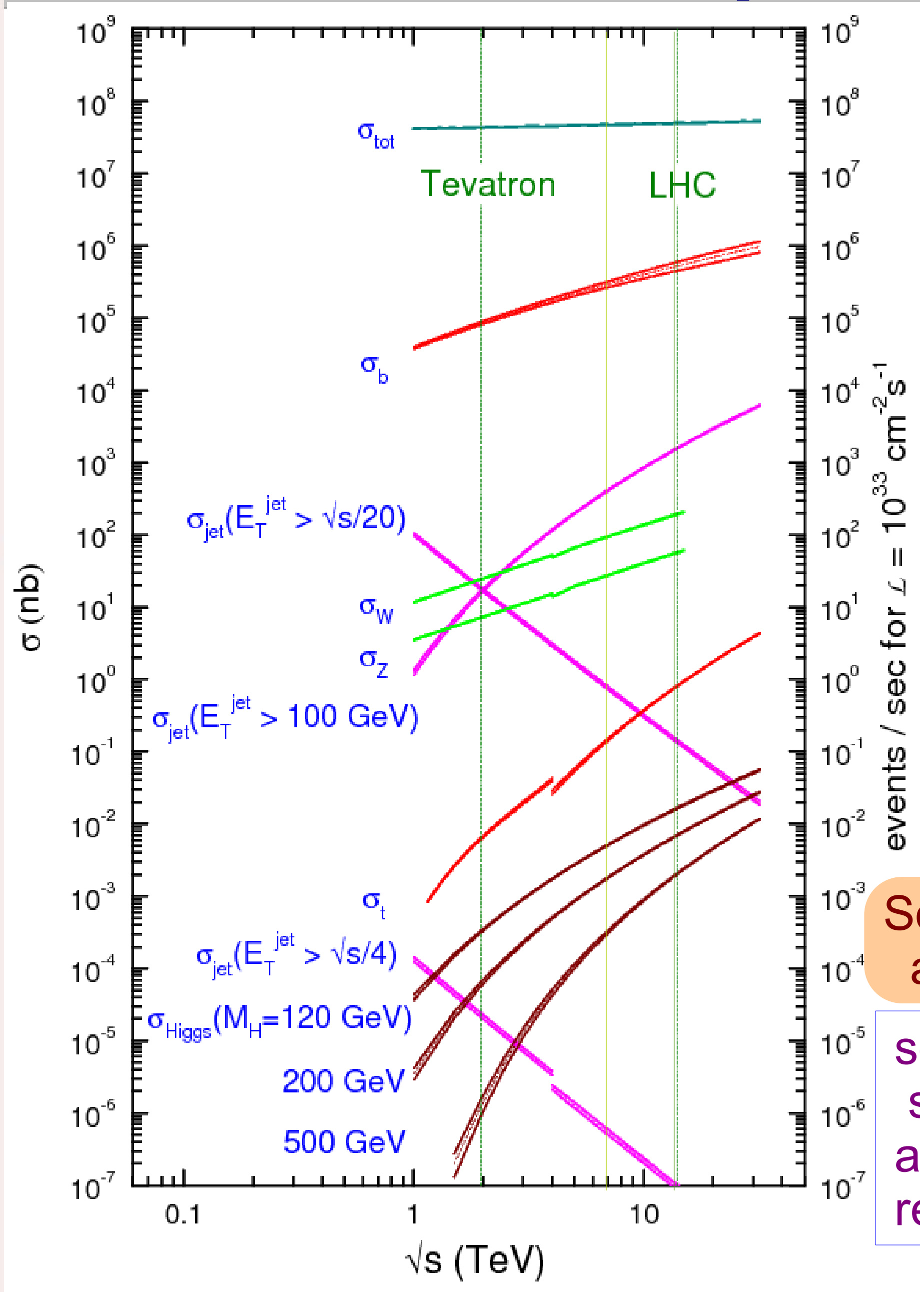
- parton distribution functions,
- hadronization effects,
- detector acceptance and efficiency,
- reconstruction efficiency and resolution,
- identification efficiency and purity

to obtain **<u>reconstructed properties</u>**



Event Selection in the Analysis





Some processes are very rare !

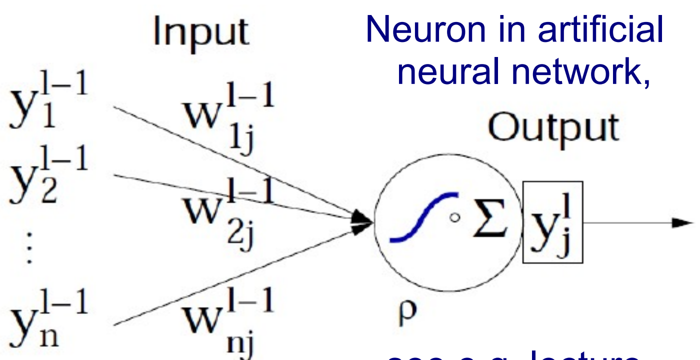
sophisticated signal selection and background rejection needed.

Analysis Steps

- recorded events are **reconstructed**: "detector hits" → physical objects like electrons, muons, photons, hadrons, jets, missing energy …
  need to know reconstruction efficiency and resolution

- **selection** of "interesting events" and objects for a particular analysis
  affected by selection efficiencies for signal and background processes

- last step of analysis involves advanced algorithms for the optimal **separation of signal from background** and **extraction of parameters** of interest from the background-corrected signal distribution
  (multivariate analysis, MVA, like discriminant methods, decorrelated likelihood, artificial neural networks, boosted decision trees)
  understanding the systematics involved is required !



Neuron in artificial neural network,

see e.g. lecture "Datenanalyse"

Finally, arrive at a result with statistical and systematic errors
evaluation of systematics requires much hard work
Much use of **simulated data** is made in this process to evaluate known or suspected sources of uncertainties and propagate them to the final results.

Steps of Event selection

- **hardware Trigger** and **on-line selection** identify „interesting“ events with particles in the sensitive area of the detector
  (events not selected are lost)

  $\longrightarrow$ detector acceptance and online-selection efficiency

- physics objects are **reconstructed** off-line

  $\longrightarrow$ reconstruction efficiency

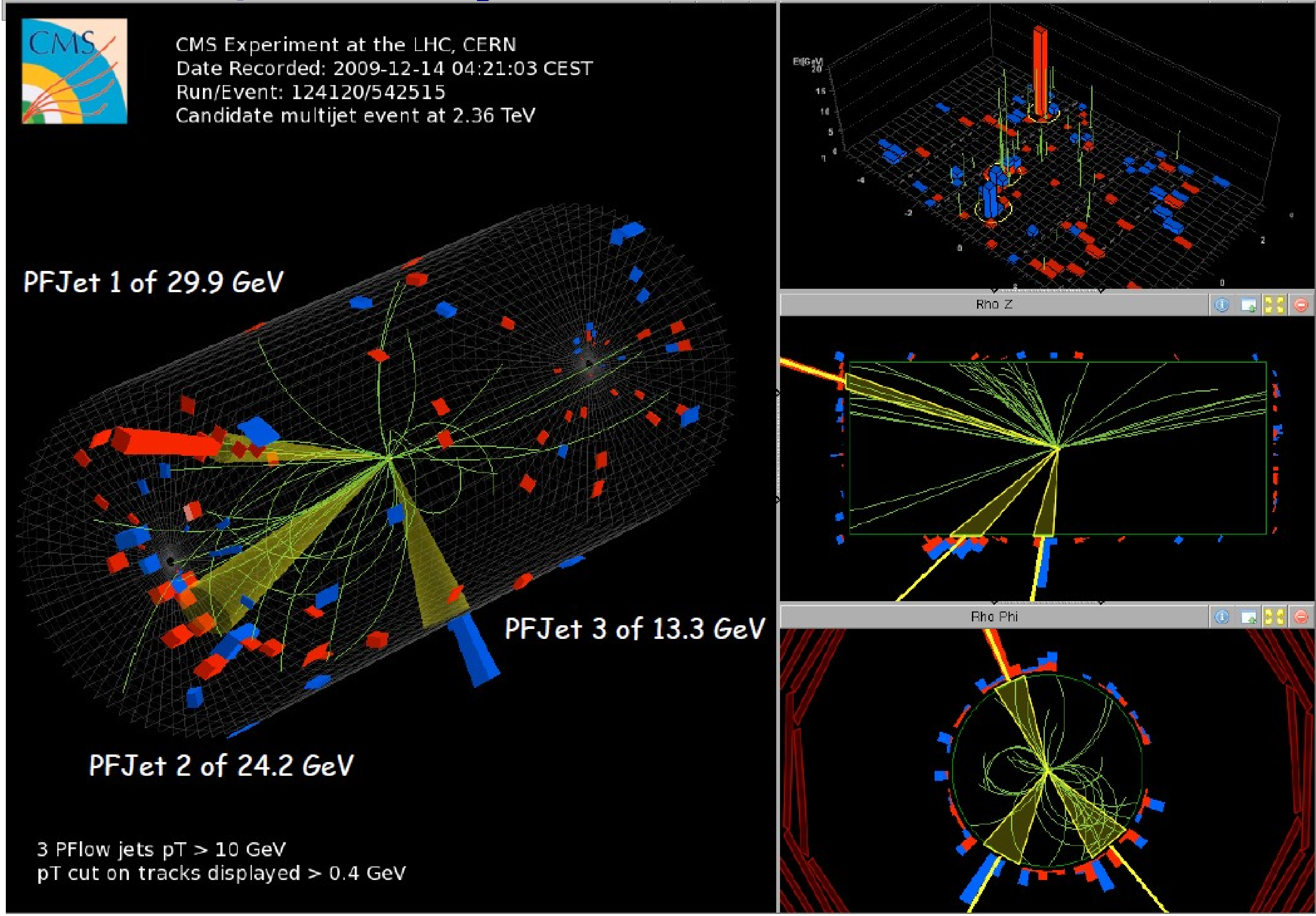
- **Analysis** procedure identifies physics processes and rejects backgrounds

  $\longrightarrow$ selection efficiency and purity

- **statistical inference** to determine confidence intervals of interesting parameters (production cross sections, particle properties, model parameters, ...)

All steps are affected by systematic errors !

example: three-jet event in the CMS detector



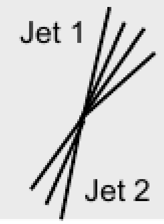
Example: Calibration of the Jet Energy scale

Jets and missing transverse energy must be calibrated

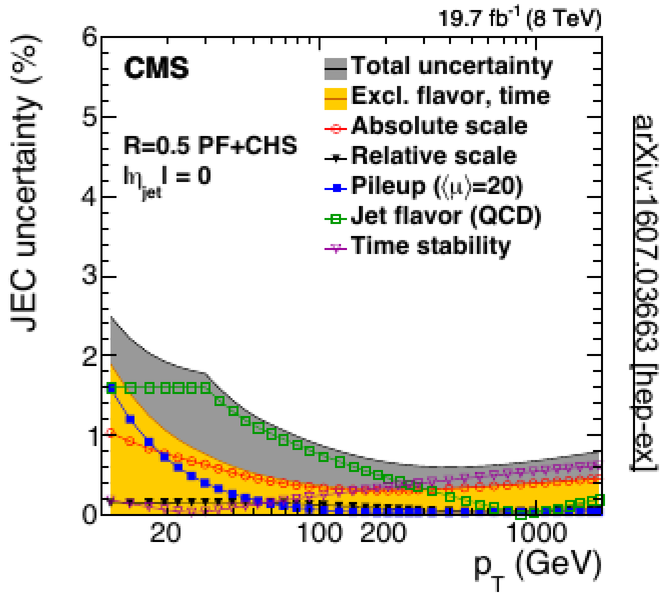
relies on special topologies:

- **di-jet events** to
  equalize detector response
- **Z or γ balanced by a jet** to
  determine absolute scale
- events with
  genuine missing energy
  (Z →$\nu\bar{\nu}$, W, Top)

Example calibration method: jet transverse momentum balance



$$f = \frac{p_{T,1} - p_{T,2}}{(p_{T,1} + p_{T,2})/2}$$



Precision of jet energy calibration reaches level of a few % !

Determination of efficiencies

two options:

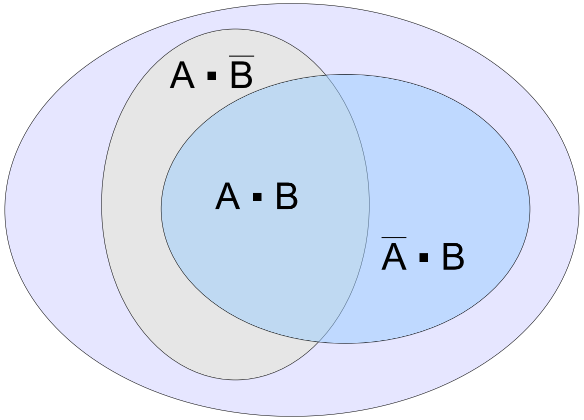
1. take efficiencies from simulation   not always believable !
   check classification in simulated data vs. truth, i.e. determine
   ε_{MC} = fraction of correctly selected objects
   (probability to select background determined in the same way)

2. **design data-driven methods** using redundancy of at least two variables discriminating signal and background
   – **tag & probe method**:
   select very hard on one criterion, even with low efficiency, check result obtained by second criterion

Illustration: two **independent** criteria A, B

$$\epsilon_B = \frac{n(A \cdot B)}{n(A \cdot \bar{B})}$$

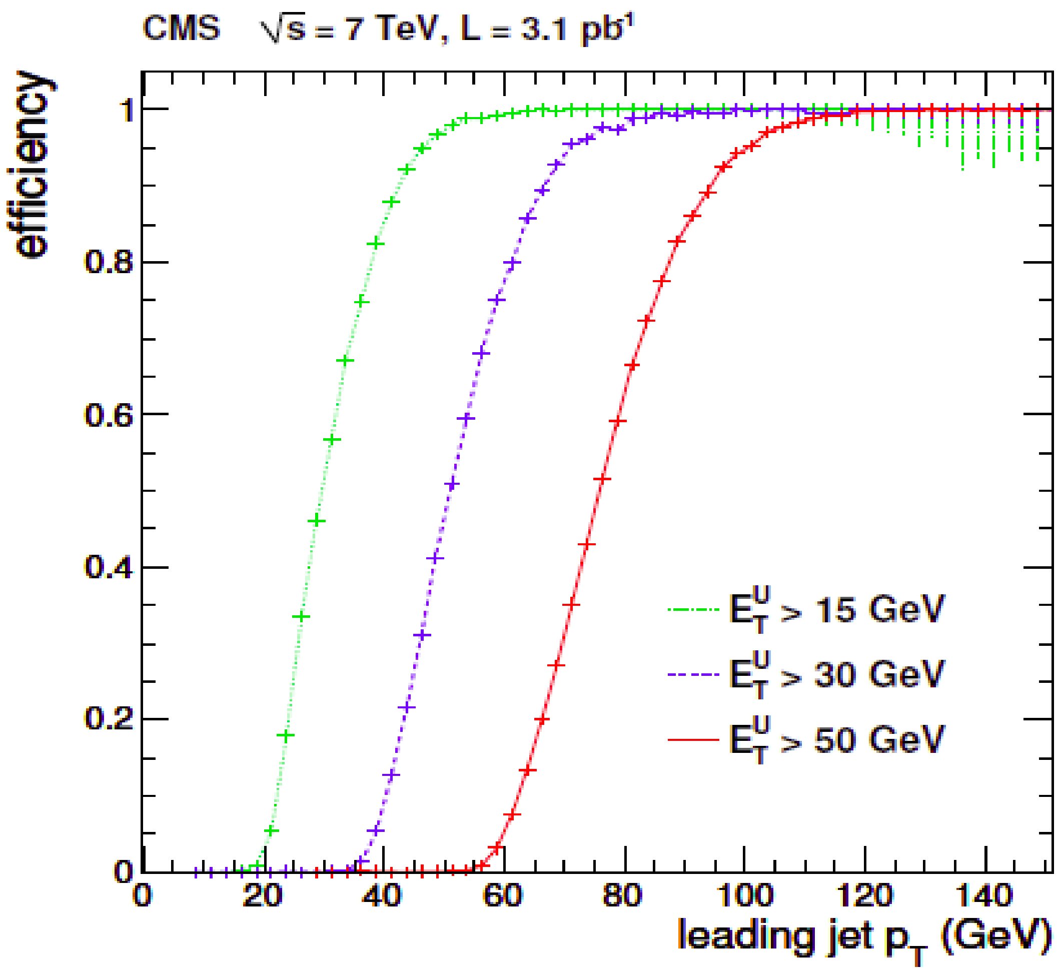
(statistical errors governed by Binomial distribution)



Example: one clear muon and one loose muon with tight selection on Z mass (“tag”) allows to measure the selection or trigger efficiency of second muon (“probe”)

Example: Trigger efficiencies

Typical “turn-on” curves of trigger efficiencies
(calorimeter jet trigger on transverse energy of jets, CMS experiment)



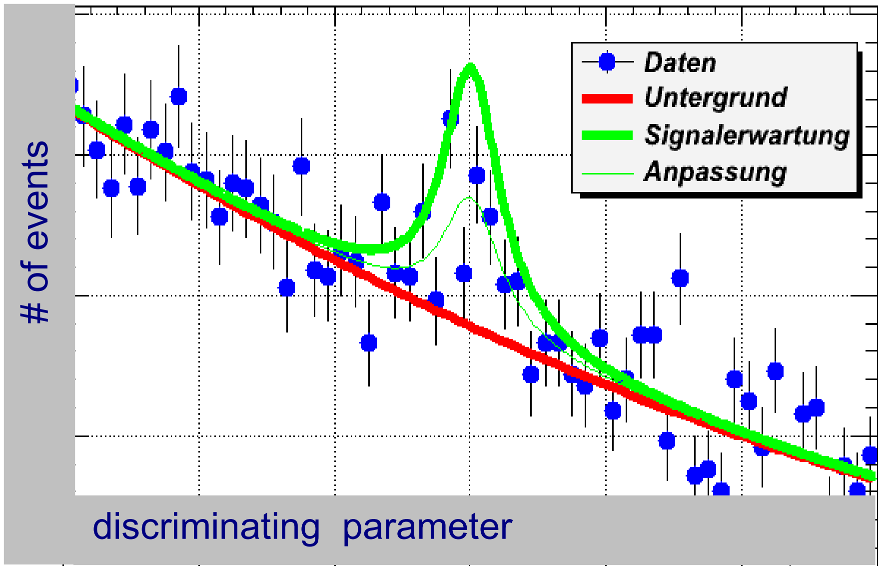
Remarks:

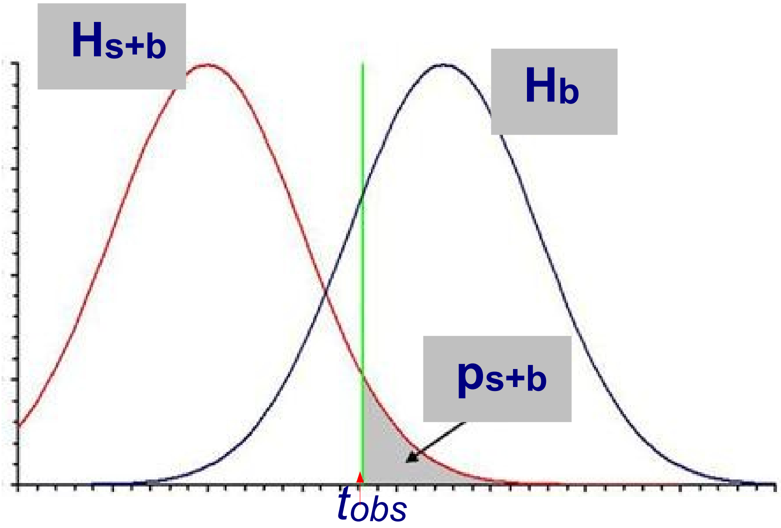
- efficiency at 100% only far beyond “nominal” threshold
- trigger efficiencies vary with time (depend on “on-line” calibration constants)
- to be safe and independent of trigger efficiencies, analyses should use cuts on reconstructed objects that are tighter than trigger requirements

Statistical analysis

The Problem: an excess of observed events can have two sources:

1. signal in addition to expectation
2. a statistical upward fluctuation or insufficient understanding of background distribution (systematic error)





from the statistical view point, a **Hypothesis Test** H_{s+b} vs. H_b

- Definition of a suitable teststatistic t as a function of the data: t_{obs}
- calculation of probabilities $p_{s+b}= \mathrm{Prob}(t>t_{obs} \mid H_{s+b})$

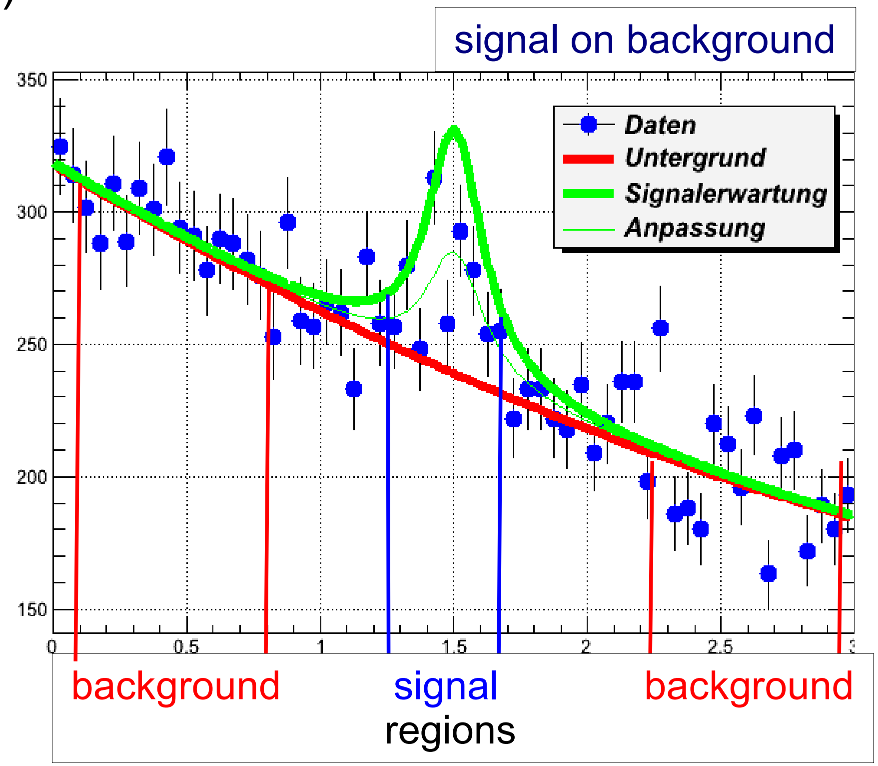
resp. $p_b= \mathrm{Prob}(t < t_{obs} \mid H_b)$

„p-values“ w.r.t. of S+B resp. B-only hypothesis

To postulate the observation of a **new signal**, background fluctuations must be excluded with very high probability !

Determination of background

- **take from MC** (same comments as above)
- **extrapolation from "side band"**
  assuming "simple" background
  shape or by taking background
  shape from simulation
  - **event counting** in background regions, extrapolation under signal assuming (simple) model
  - **fit** of signal + background model to the observed data

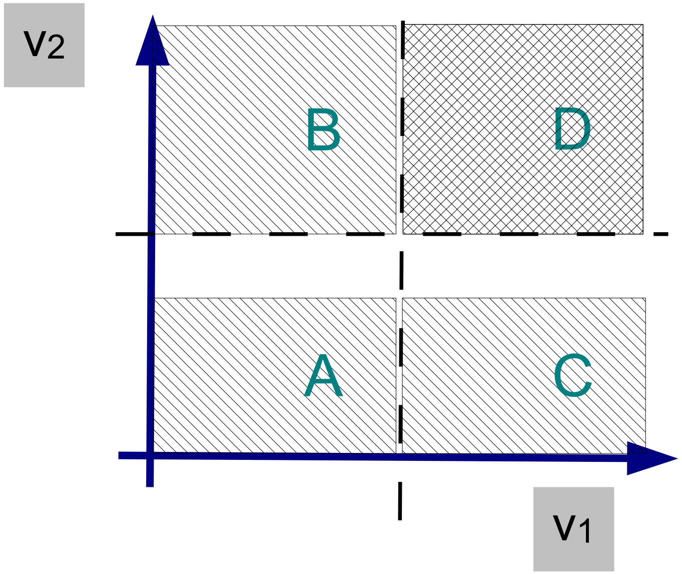


- **if a second, independent variable for separation of signal from background can be found, background determination purely from data becomes possible**

→ **ABCD method**

Determination of background

ABCD – Method ...



Assumptions:

– two independent variables v1 and v2 for background

– signal only in region D

$$\rightarrow \quad n_D^{bkg} = n_C \frac{n_B}{n_A}$$

... a **data driven estimate** of *background under a signal*

Example: invariant mass of two unlike-sign particles, combinatorial background from sample with like-sign particles.

More advanced methods exist to **exploit two uncorrelated variables** to predict the background shape under a signal, see e.g. “sPlot method” in ROOT.

Example of improved background modelling

Hybrid events: data + Monte Carlo

example: $Z \rightarrow \tau\tau$ to background in $H \rightarrow \tau\tau$ search

- $H \rightarrow \mu\mu$ has very low cross section, hence there is no $H \rightarrow \mu\mu$ under $H \rightarrow \mu\mu$
- $Z \rightarrow \mu\mu$ and $Z \rightarrow \tau\tau$ are very similar (lepton universality of weak decay)

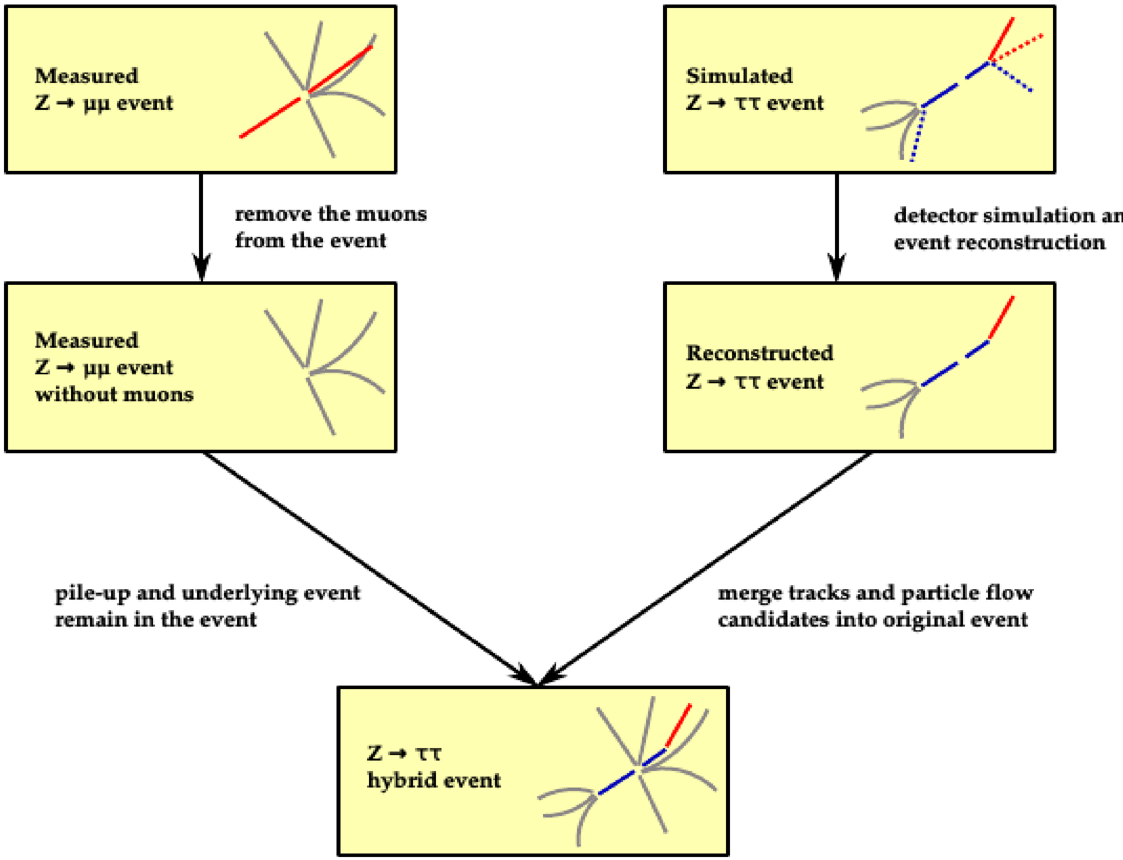
Idea:

replace μ in $Z \rightarrow \tau\tau$ events to model Z background under H signal

advantages:

– non-leptonic part of event is from real data, esp. important in presence of pile-up

- leptonic part can be well and easily modelled
- important cross check of full simulation via MC



PhD thesis Manuel Zeise, Karlsruhe, July 2011