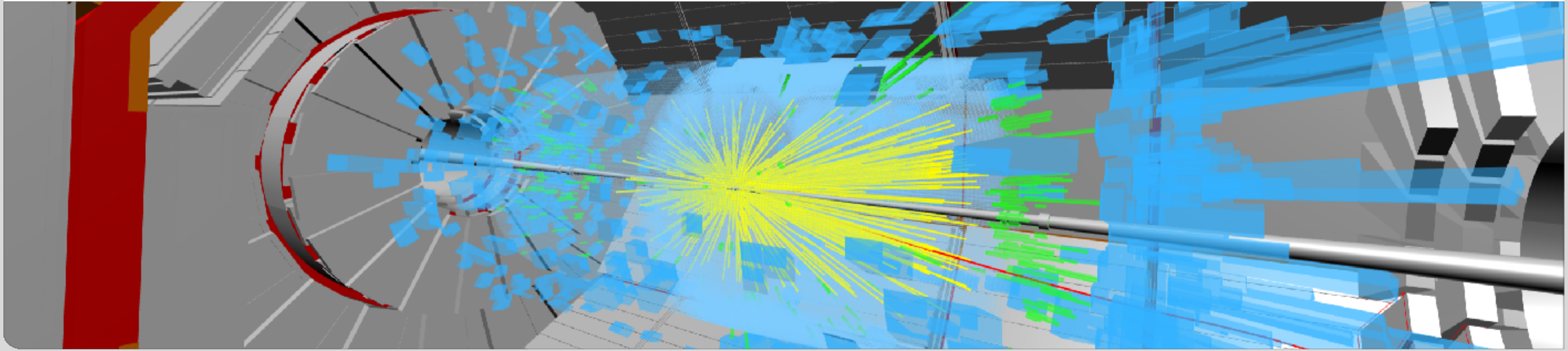
Karlsruhe Institute of Technology

Tools and Techniques for High-p_T Physics

7th ENHEP School on High Energy Physics
Ain Shams University, Cairo, January 26–31, 2019

Ulrich Husemann, Institute of Experimental Particle Physics, Karlsruhe Institute of Technology



Scope of this Lecture

- Introduction the the most important **tools and techniques** for experimental particle physics at colliders
  - Two one-hour **lectures**: discussion of ideas and concepts
  - One **tutorial** session: deeper understanding by hands-on exercise (at least for one example)
- Please **ask questions** – during and after the lecture and anytime you can grab me otherwise

Outline

From Raw Data to Physics Results

Monte-Carlo Event Generation

Physics Objects

Background Estimation

Advanced Signal Analysis

Overview

FROM RAW DATA TO PHYSICS RESULTS

Cross Section: Master Formula

Number of events observed in detector
just count…

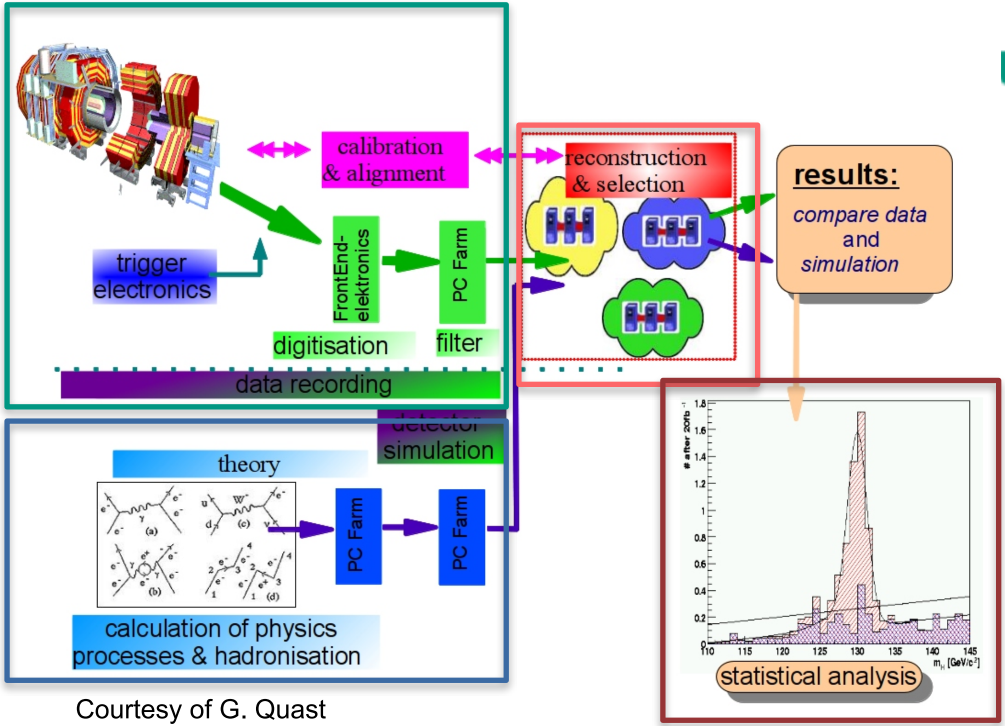
Background
measured in data or from theory prediction

$$\sigma = \frac{N^{\text{obs}} - N^{\text{bkg}}}{\int \mathcal{L}\,\mathrm{d}t \cdot \varepsilon}$$

Luminosity
depending on accelerator, trigger, …

Efficiency
Acceptance and efficiency of detection, analysis-specific optimization

Detector & Analysis Chain



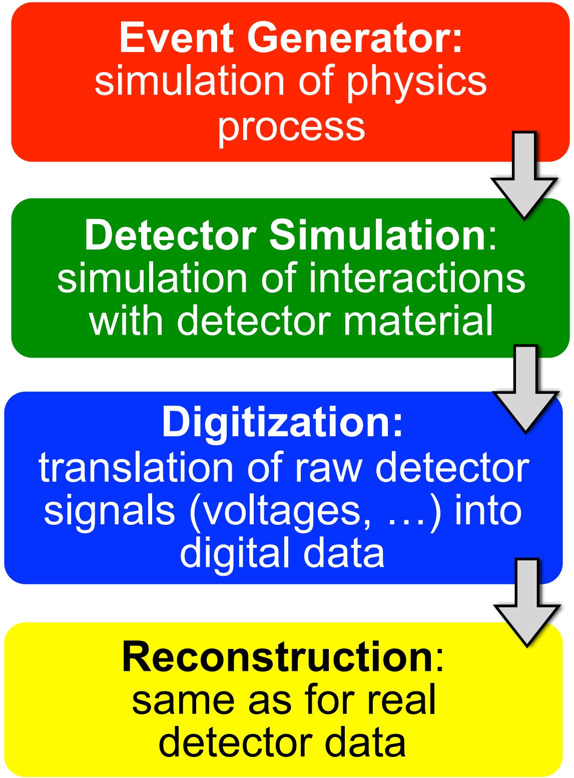
Courtesy of G. Quast

- Collider-based particle physics: complicated analysis chain
  - **Detector** input: calibrated digitized data, online selection ("trigger")
  - **Theory** input: calculations and simulations
  - Physics object **reconstruction** and **selection**
  - **Statistical** analysis & comparison with theory

MONTE-CARLO EVENT GENERATION

MC Simulations in Particle Physics

- Goal: **comprehensive simulation** of collision events based on best knowledge of all physics processes (collision events and interactions in detector)
- Main tools based on **Monte-Carlo (MC) method**:
  - In general: MC = **numerical** techniques to compute **probabilities** using **random numbers**
  - Excellent tool to generate **physics events** (QM = probabilistic theory) and simulate **particle interactions** in detectors



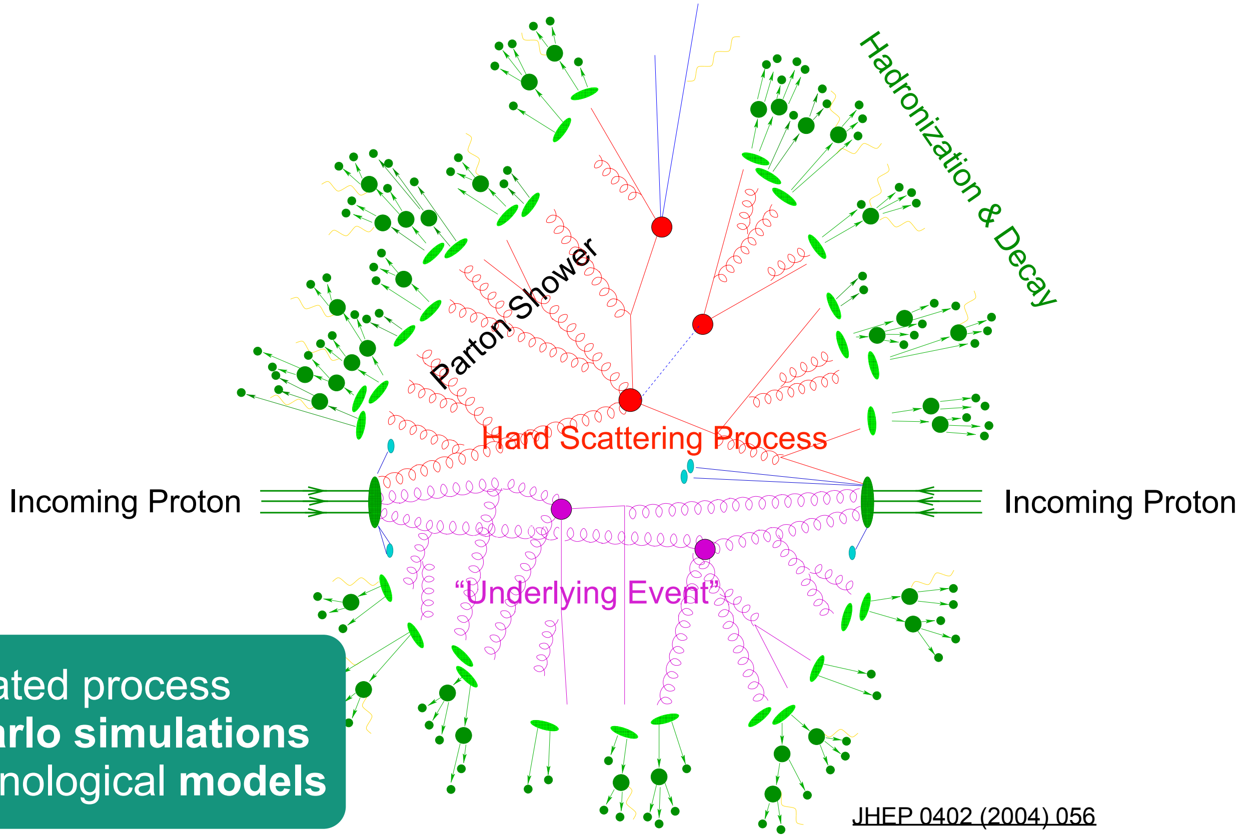
Event Generator: simulation of physics process

Detector Simulation: simulation of interactions with detector material

Digitization: translation of raw detector signals (voltages, …) into digital data

Reconstruction: same as for real detector data

Simulation of a Collision Event



JHEP 0402 (2004) 056

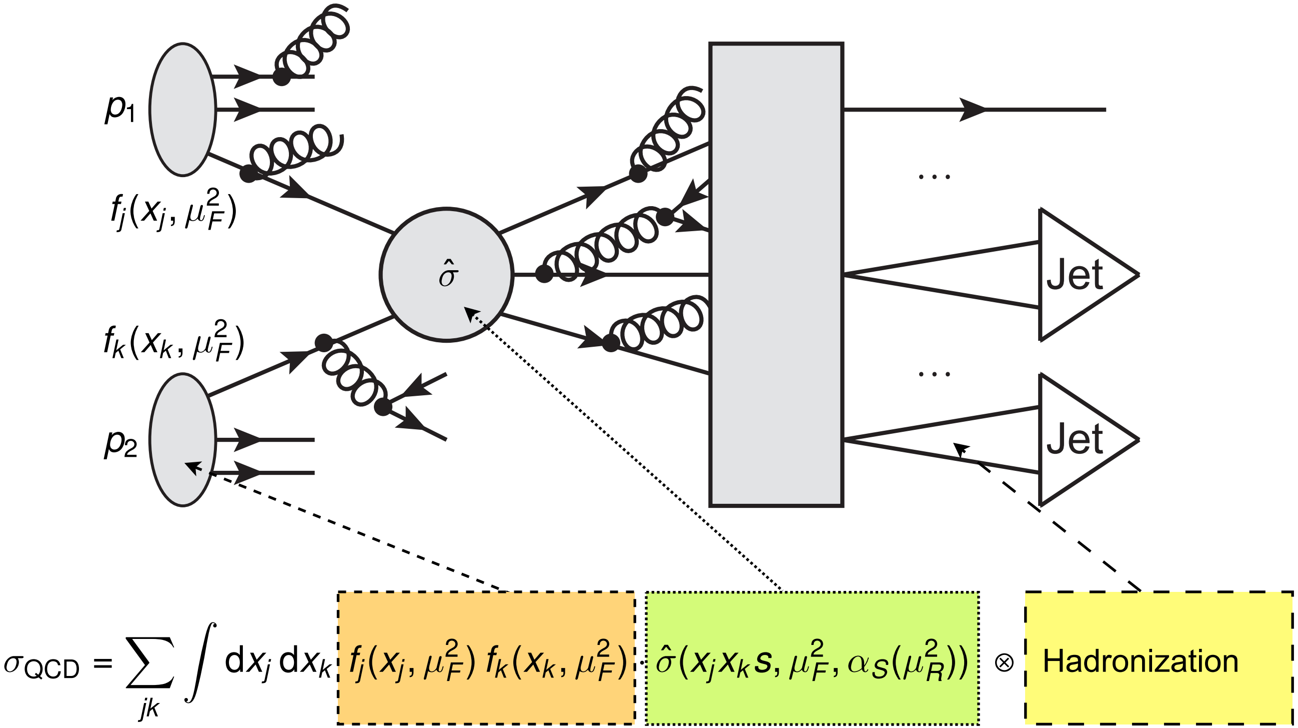
Complicated process
→ **Monte-Carlo simulations**
and phenomenological **models**

MC Event Generators

- Goal: **realistic simulation** of all relevant physics processes in a particle collision
- Problem: **complexity** of hadron-hadron collisions
  - Initial state: hadrons = compound objects, constituents (quarks & gluons) confined in hadron (QCD interactions at low four-momentum transfer Q^2)
  - Final state: many hadrons and leptons
- Way out: **QCD factorization**
  - **Separate treatment** of processes at low and high Q^2
  - High Q^2 (“hard scattering process”, short-distance physics): **perturbation theory** in leading order (LO) or higher orders (NLO, NNLO, …)
  - Low Q^2 (“soft physics”, long-distance physics): phenomenological **models**

QCD Factorization in Pictures

Cross Section = PDFs ⊗ Hard Process ⊗ Hadronization



$$\sigma_{\mathrm{QCD}} = \sum_{jk} \int \mathrm{d}x_j\, \mathrm{d}x_k\, \boxed{f_j(x_j, \mu_F^2)\, f_k(x_k, \mu_F^2)}\, \boxed{\hat{\sigma}(x_j x_k s, \mu_F^2, \alpha_S(\mu_R^2))} \otimes \boxed{\text{Hadronization}}$$

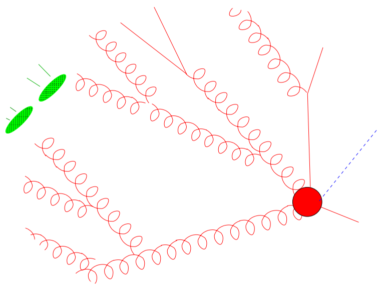
Overview of MC Generators

- Central step in any MC generator: **MC integration** of cross section for hard process in fixed order of perturbation theory (using PDFs as inputs) (Fermi's Golden Rule: cross section = |matrix element|^2 ⊗ phase space)
- **Parton-level** MC generators:
  - Simulation stops at the level of partons (quarks and gluons)
  - No hadronisation, only events weighted with the differential cross section → no full event simulation (still useful for theoretical studies)
- **Particle-level** MC generators:
  - Full event simulation: parton level + unweighting (→ number of MC events corresponds to theoretical expectation) + parton shower + hadronization
  - May be provided as a single comprehensive package or as a combination of ME provider and parton shower MC (SMC) program (examples: later)

Lectures on Event Generators

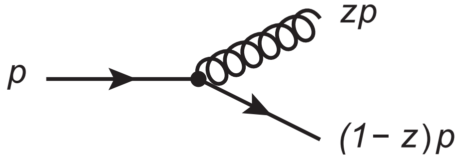
- T. Sjöstrand, Introduction to Event Generators, CTEQ/MCnet School, DESY 2016
  - http://home.thep.lu.se/~torbjorn/talks/desy16a.pdf
  - http://home.thep.lu.se/~torbjorn/talks/desy16b.pdf
  - http://home.thep.lu.se/~torbjorn/talks/desy16c.pdf
  - http://home.thep.lu.se/~torbjorn/talks/desy16d.pdf
- S. Gieseke, Parton Shower Monte Carlos, UA Madrid 2016: https://www.itp.kit.edu/~gieseke/uam/madrid16.pdf
- F. Maltoni, CERN Academic Training Lectures 2012: https://indico.cern.ch/conferenceDisplay.py?confId=181765
- B. Webber, Parton Shower Monte Carlo Event Generators, Scholarpedia, 6(12):10662

Parton Shower



- Parton shower:
  - **Coherent** emission of soft colored particles
    → quantum interference effect
  - Emission process can be modeled by a sequence of **1→2 parton splitting** processes
  - Parton shower = **probabilistic model** of quark fragmentation
- Central object of parton shower: **Sudakov form factor**
  - Probability for a parton i to emit a parton j: **splitting function** P_{ji}
  - Solution of **DGLAP equation** for parton shower: Sudakov form factor

$$\Delta_i(t) = \exp\left[-\sum_j \int_{t_0}^{t} \frac{\mathrm{d}t'}{t'} \int_0^1 \mathrm{d}y\, \frac{\alpha_S}{2\pi} P_{ji}(y)\right]$$

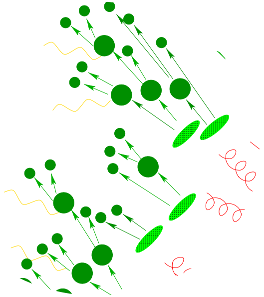


Modern SMC Codes

- General purpose SMC event generators: completely new **C++ developments** based on experience with established FORTRAN codes and new C++ codes
    - **Pythia 8** (http://home.thep.lu.se/~torbjorn/Pythia.html): default: p_T-ordering, other options available
    - **Herwig 7** (http://herwig.hepforge.org): default: angular ordering, optionally: dipole showering
    - **Sherpa** (https://gitlab.com/sherpa-team/sherpa): dipole showering
- LHC Run 2: Pythia 8, Herwig 7, Sherpa as **new standards** for SMC (also in combination with other packages)

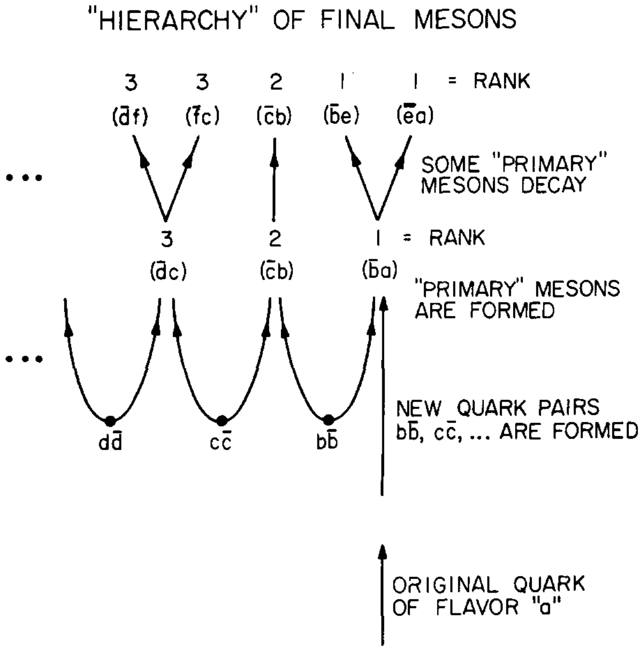
Hadronization Models

- Parton-hadron transition: **non-perturbative** processes
- Phenomenological MC models very successful:
  - Basic assumption: **parton-hadron duality** → very close relation between parton dynamics and properties of final state hadrons
  - Advantage: **full** event simulation → can be used **directly** for experiments
  - Disadvantage: many **ad-hoc** parameters in some models → (rather extensive) **tuning** required
- Most well-known hadronization models:
  - Independent fragmentation (not used any more)
  - Lund **string** model (Pythia)
  - **Cluster** model (Herwig, Sherpa)



Independent Fragmentation

- Fragmentation for each parton **independent** (as the name says…) (R.D. Field, R.P. Feynman, 1978)
- Algorithm (starting from original quark):
  - Quark-antiquark pairs formed out of vacuum → primary mesons with energy fraction *z*
  - New starting point: remaining quark with energy fraction 1–*z*
  - Stop at lower energy threshold
- **Fragmentation function *D*(*z*)**: Probability to find hadron with energy fraction *z* in jet (*D*: non-perturbative object)



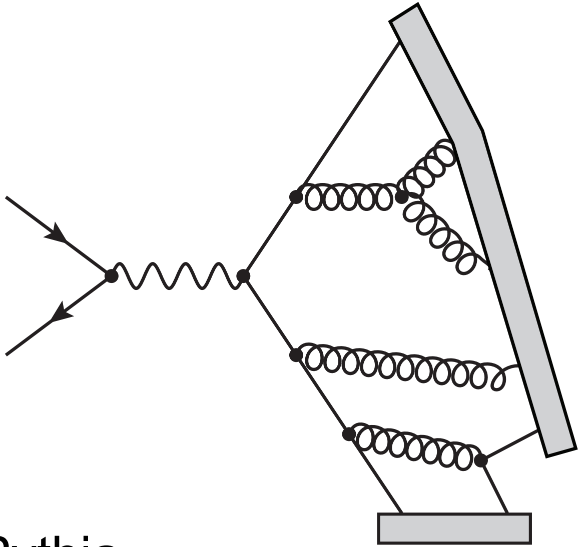
Nucl. Phys. B136 (1978) 1

Lund String Model

- Quark-antiquark pairs form **strings**: (B. Andersson et al., Phys. Rept 97 (1983) 31)
  - Static QCD potential between quark and antiquark: **linear** increase with distance → gluon field collapses to a **string**

$$V(r) = -\frac{4}{3}\frac{\alpha_S(1/r^2)}{r} + kr$$

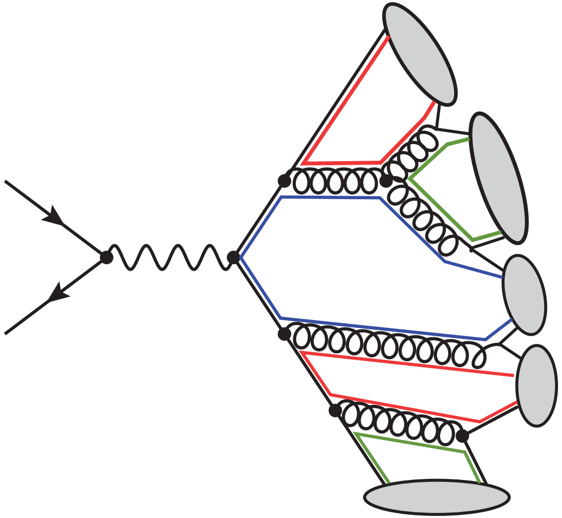
  - Strings **break** once $V(r)$ large enough to form new quark-antiquark pairs
  - Effect of gluons: **"kinks"** in strings
  - Lower energy threshold: formation of hadrons
- Invented at Lunds universitet, Sweden, used in Pythia



after: Ellis et al., QCD and Collider Physics

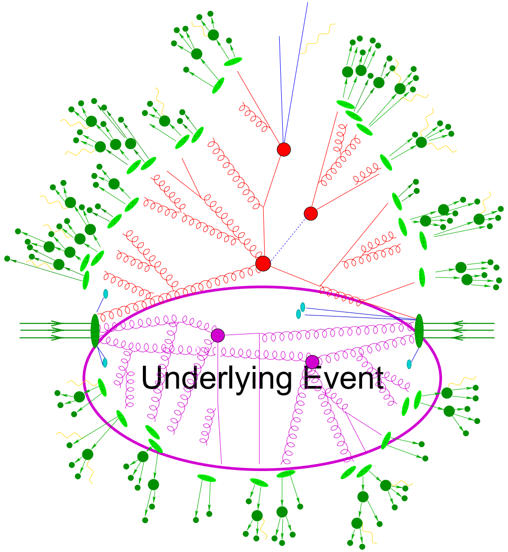
Cluster Model

- Color flow during hadronization governed by **pre-confinement** property of QCD → tendency for partons to form **color-neutral clusters** (B. R. Webber, Nucl. Phys. B238 (1984) 492)
  - Gluons (color-anticolor): splitting into quark-antiquark pairs → **new clusters** with color partners
  - Clusters decay into **hadrons** according to available phase space
  - Advantage: (essentially) no tunable parameters
- Used in **Herwig**



after: Ellis et al., QCD and Collider Physics

Underlying Event

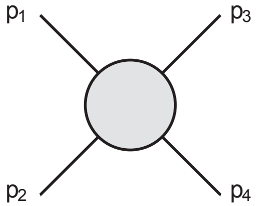


- Total parton-parton cross section ⊗ PDF at the LHC **larger than total proton-proton cross section**
- Reason: **several** parton-parton collisions within the same proton-proton collision
- Definition: **underlying event** = all interactions except hard scattering process (which includes initial state and final state radiation, and beam remnant)
- MC codes include underlying event **models** (tuned to data)

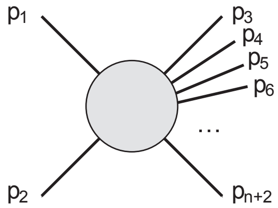
Hard-Scattering Matrix Elements

- First generation of MC codes: **LO** ME for **2→1 and 2→2 processes** → available for all SM and BSM processes
- Improvement I: **LO** matrix elements for important **2→*n* processes**, mostly: **additional real emission** of quarks/ gluons → jets (approximation of higher orders)
- Improvement II: **NLO matrix elements** (= **real** emission + **virtual** corrections = loops) → 2019: available for all SM processes and many BSM processes
- Improvement III: **automation** of ME computations (especially at NLO)

2→2 Process
(e.g. pp → $t\bar{t}$)



2→*n* Process
(e.g. pp → $t\bar{t}$+(n–2) jets)



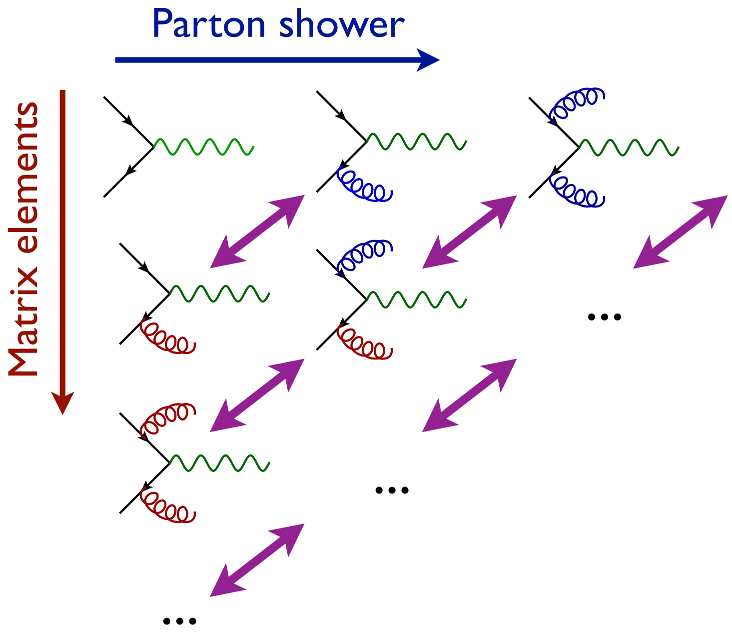
Modern MC Frameworks

- General-purpose SMC codes: **Pythia**, **Herwig**, **Sherpa**
- **LO** for **2→*n*** processes → combination with SMC (non-trivial)
  - **ALPGEN**
  - **MadGraph5_aMC@NLO** (MG5aMC)
  - **Sherpa** (2→*n* processes included in SMC)
- **NLO** (comb. with SMC even less trivial)
  - **POWHEG BOX**
  - **MG5aMC**

Further examples:

- Modern SMC codes also allow using **external ME providers** (e.g.: OpenLoops, GoSam, MG5aMC)
- **Specialized MC code** for important processes (e.g. VBFNLO = parton-level MC for processes with electroweak bosons)

Double Counting and MC Matching



F. Maltoni

- Real emission from both LO 2→n matrix element (ME) and parton shower (PS) → **double counting**
- Solution: **matching** between ME and PS, **removal** of overlap
- Matching algorithms:
  - **MLM** matching (Mangano), e.g. in ALPGEN, MG5aMC
  - **CKKW(-L)** matching (Catani, Krauss, Kuhn, Webber; Lönnblad), e.g. in Sherpa, MG5aMC

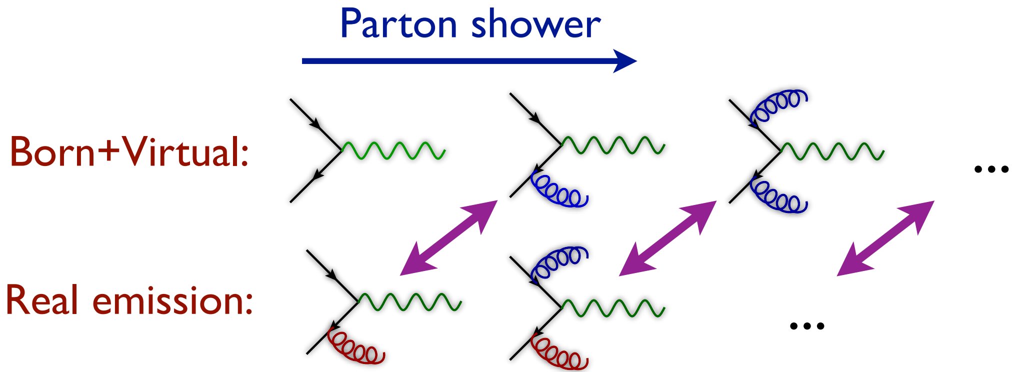
How to Match

- General idea: use ME (PS) only in those areas of phase space where ME (PS) works best

Matrix Element	Parton Shower
fixed order	resummation
numerically expensive	numerically cheap
number of final-state particles limited	arbitralily many final-state particles
works best for hard, well-separated partons	works best for soft/collinear partons

NLO Generators and NLO Merging

- NLO matrix elements: **real** and **virtual** corrections
- Problem: real emission does not only include hard, well separated partons, radiation may also be **soft/collinear**
    - Different separation between NLO ME and PS compared to LO
    - Correct for **two-fold double counting**: real emission & parton shower, virtual corrections and Sudakov form factor
    - Double counting removed by **NLO merging** algorithms (several methods)



F. Maltoni

Short Summary

- MC generators may be classified by:
  - Available **physics processes**
  - **Highest order** in perturbation theory for hard scattering matrix elements
  - **Number** of outgoing particles
  - **Partonic** or **hadronic** final state
  - **“Link”** (=matching/merging) between matrix element and parton shower
- Classes of MC generators:
  - Pure parton-level MC generator (LO or NLO)
  - General-purpose parton shower MC generator (SMC)
  - LO matrix element provider combined with parton shower (ME+PS)
  - NLO matrix element provider combined with parton shower (NLO+PS)

PHYSICS OBJECTS

Reminder: Cross Section Formula

Number of events observed in detector

just count…

Background

measured in data or from theory prediction

$$\sigma = \frac{N^{\text{obs}} - N^{\text{bkg}}}{\int \mathcal{L}\,\mathrm{d}t \cdot \varepsilon}$$

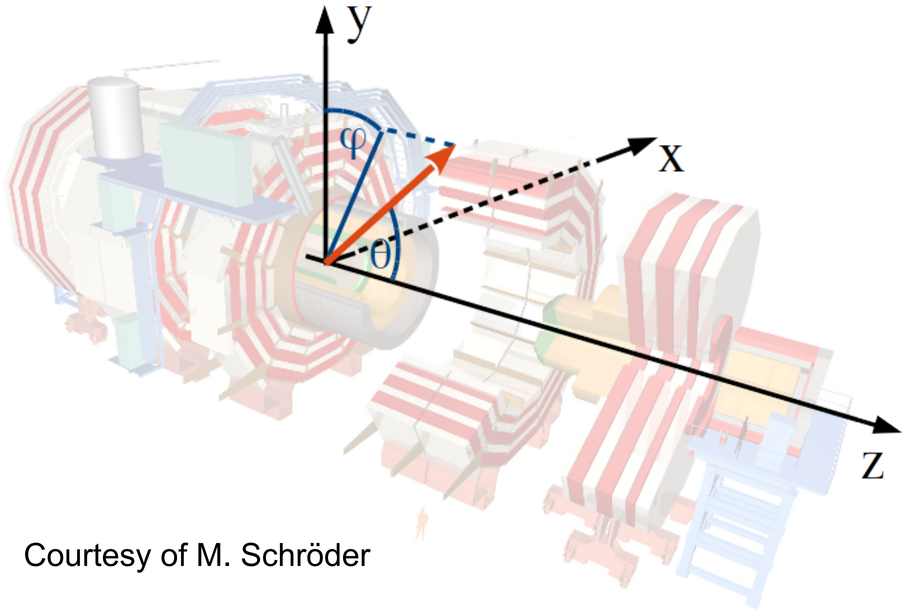
Luminosity

depending on accelerator, trigger, …

Efficiency

Acceptance and efficiency of detection, analysis-specific optimization

Reminder: Collider Kinematics



Courtesy of M. Schröder

- Coordinate system:
  - **Right-handed** system, origin: **center of detector** = nominal collision point
  - **Cartesian** coordinates: *z* axis along counter-clockwise beam direction, *x* axis pointing towards center of the ring, *y* axis pointing upwards
  - **Polar** coordinates: polar angle θ = angle with *z* axis, azimuthal angle ϕ in *xy* plane, measured from *x* axis

Reminder: Collider Kinematics

- Typical **hadron collider observables**:
  - Partonic center-of-mass frame unknown → use **transverse** quantities (invariant under Lorentz boosts along beam axis)
  - Particle-level observables: **transverse momentum** and **pseudorapidity**

$$p_T = \sqrt{p_x^2 + p_y^2} \qquad \eta = -\ln\tan\left(\frac{\theta}{2}\right)$$

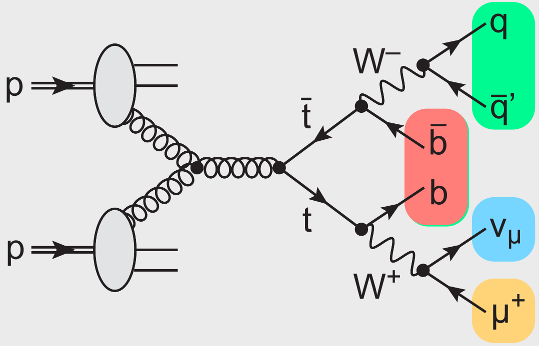
→ four-momenta of particles: particle mass (hypothesis), p_T, η, ϕ

  - **Energy/momentum-sum** observables: **missing transverse momentum**

$$\mathbf{p}_T^{\text{miss}} = -\sum_{\text{rec. particles}} \begin{pmatrix} p_x \\ p_y \end{pmatrix} \quad \rightarrow \quad E_T^{\text{miss}} \equiv p_T^{\text{miss}} \equiv \text{MET} \equiv \left|\mathbf{p}_T^{\text{miss}}\right|$$

Event Selection

Example of cut-based event selection: top lepton+jets selection



Lepton with $p_T > 20$–30 GeV

Neutrino: MET > 30 GeV

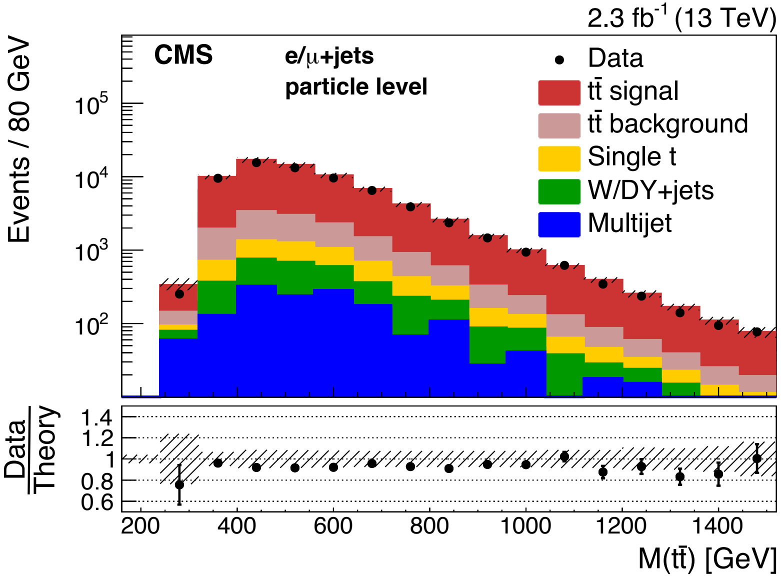
4 jets with $p_T > 40$ GeV

2 of the jets from *B* decays: *b*-tags

- “Art” of data analysis: select events such that signal and background are **optimally separated**
- **Optimization** of selection:
  - Criterion: signal/background ratio $N^{\text{sig}}/N^{\text{bkg}}$
  - Criterion: signal significance $N^{\text{sig}}/\sqrt{N^{\text{sig}} + N^{\text{bkg}}}$
  - Note: use simulated data for optimization, **never optimize the signal on data!**
- **Many methods**, from simple (“cuts”) to sophisticated (multivariate methods)

Background

- **Physics**: which processes **cannot be distinguished** from signal?
- **Instrumental** backgrounds:
  - Detector noise
  - Misidentified objects (“fakes”), e.g. jet misreconstructed as electron
- Example: most important background processes in Higgs and top physics
  - *W* and *Z* boson production in association with jets (“*W/Z*+jets”)
  - QCD multijets



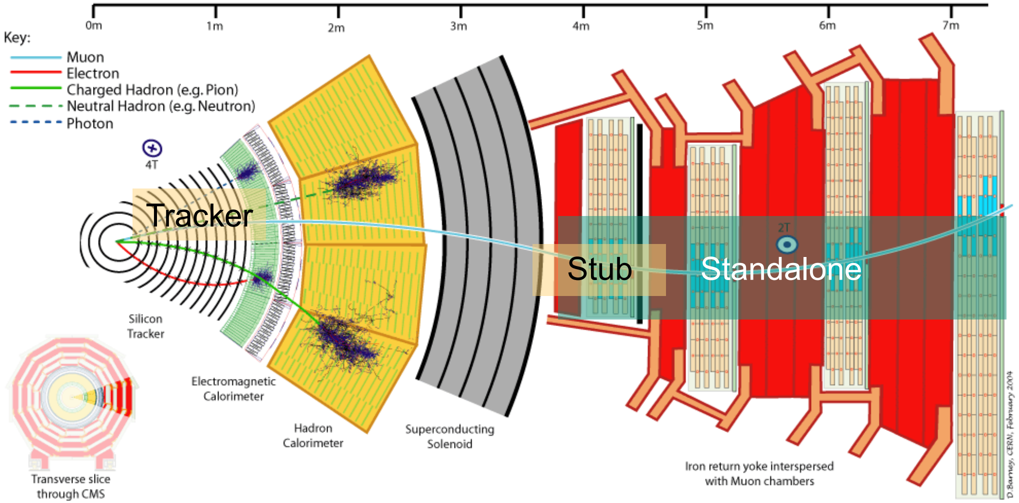
Phys. Rev. D95 (2017) 092001

Hierarchy of Analysis Objects

- Most important physics objects:
  - **Muon**: matching tracks in tracking detector and muon detector
  - **Electron**: ECAL cluster matched to track
  - **Photon**: ECAL cluster without track, conversion into e^+e^-
  - **Charged hadron**: HCAL cluster matched to track
  - **Neutral hadron**: HCAL cluster without track
  - **Jet**: bundles of calorimeter clusters and/or tracks/hadrons
  - **b-Jet**: jet containing B-hadrons
  - **Hadronic tau**: thin jet-like bundle, e.g. one or three charged hadrons
- All physics objects must be **calibrated**:
  - General rule: know where to trust your simulation (and where not)
  - In practice: mixture of **MC-based** and **data-driven** methods

Muon Reconstruction

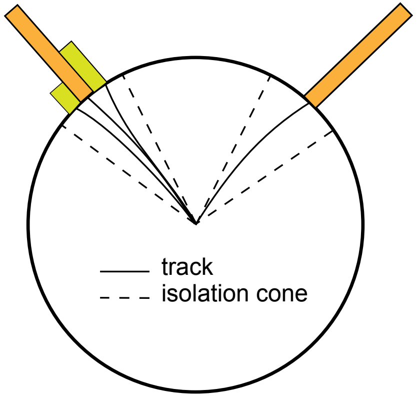
- Typical muon reconstruction strategies (CMS jargon):
  - **Standalone** muons: only reconstructed with muon detector
  - **Tracker** muons: track segment in tracker, “stub” in muon detector
  - **Global** muons: track segments in tracker and muon detector combined



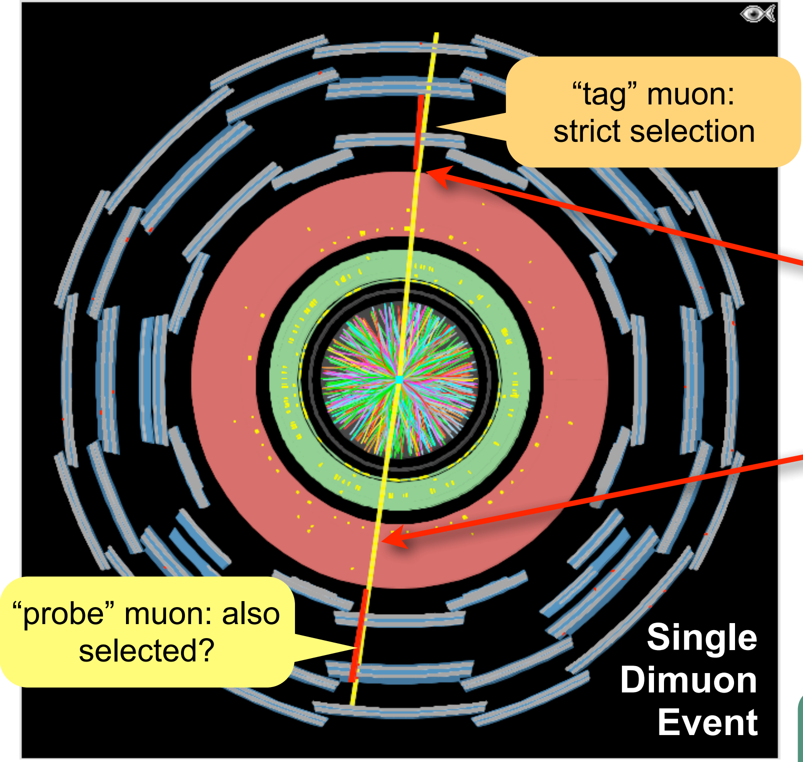
D. Barney, CMS Detector Slice, CC-

Muon Identification

- Muon identification (ID) algorithms:
  - Criteria: quality of track fit, number of tracker layers with hits, impact parameters w.r.t. primary vertex, low energy deposit in calorimeters (MIP)
  - Muons from electroweak decays (e.g. *W*, *Z*): improved ID using track or calorimeter **isolation** (low transverse momentum sum of tracks/energy deposits around muon)
  - Today: combined in **multivariate muon ID** variable
- Muon selection: pre-defined **working points,**
  - Typically: “loose” for high efficiency, “tight” for high purity
  - Advantage: muon ID efficiency determined **centrally**, e.g. by **tag&probe** method (→ next slide)



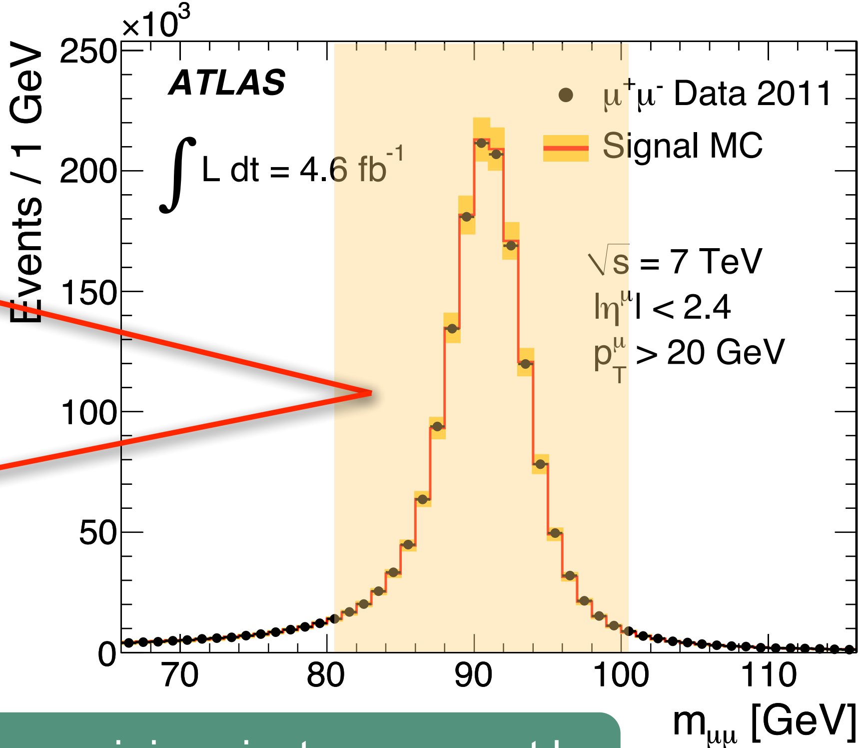
Muon Efficiency: Tag&Probe



“tag” muon: strict selection

“probe” muon: also selected?

https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone



https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/STDM-2012-06/

Muon pair invariant mass: must be in window around *Z*-boson mass

Tag&Probe Method

- Kinematic **constraint**: muon pair invariant mass in **window around *Z*-boson mass** (width ≥ natural width ⊕ detector resolution)
- **Tag&probe** algorithm:
  - “**Tag**” muon: **strict** selection, e.g. muon trigger and clean reconstruction
  - “**Probe**” muon: opposite charge, **looser** selection criteria (e.g. no trigger)
  - **Tag&probe efficiency** (often as a function of kinematics, e.g. p_T and η):

$$\epsilon_{\text{tag\&probe}} = \frac{N_{\text{tag\&probe}}}{N_{\text{tag}}}$$

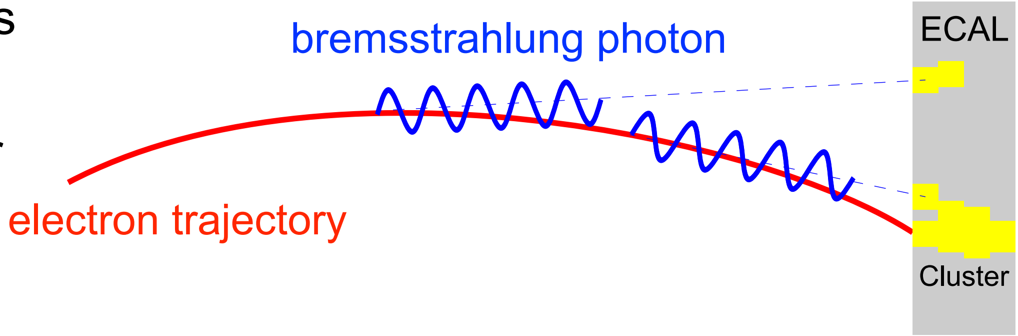
(proper determination of uncertainties → tutorial)

  - Compute tag&probe efficiency in data and simulation: determine **correction factor** (“scale factor”) for simulation (again: as a function of kinematics)

$$\text{SF} = \frac{\epsilon^{\text{data}}_{\text{tag\&probe}}}{\epsilon^{\text{MC}}_{\text{tag\&probe}}}$$

Electron Reconstruction and ID

- Typical electron reconstruction strategy:
  - **Clustering** of energy deposits in calorimeter cells
  - **Matching** of track and cluster
  - **Bremsstrahlung** recovery



- Electron identification:
  - Criteria: number of tracker layers with hits, impact parameter w.r.t. primary vertex, shape of electromagnetic shower, leakage into HCAL, …
  - **Isolation**: similar to muons
  - Today: combined in **multivariate electron ID** variable (loose, tight, …)

Summary

- Hierarchy of analysis objects:
  - Hits/clusters → tracking and calorimeter objects → **physics objects** (charged leptons, jets, b-jets, boosted objects …)
  - Lepton ID: **multivariate** discriminant, isolation
- All objects must be properly calibrated:
  - Tracks: **alignment** of tracking detectors
  - Leptons: **scale factors**, e.g. via tag&probe method