Application of Machine Learning Techniques in Collider Physics Reconstruction & Analysis

Peter Loch

Department of Physics
University of Arizona
Tucson, Arizona, USA

CSI Workshop
BNL
July 26, 2018

ATLAS

THE UNIVERSITY OF ARIZONA
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

Preliminaries

Topics in this talk

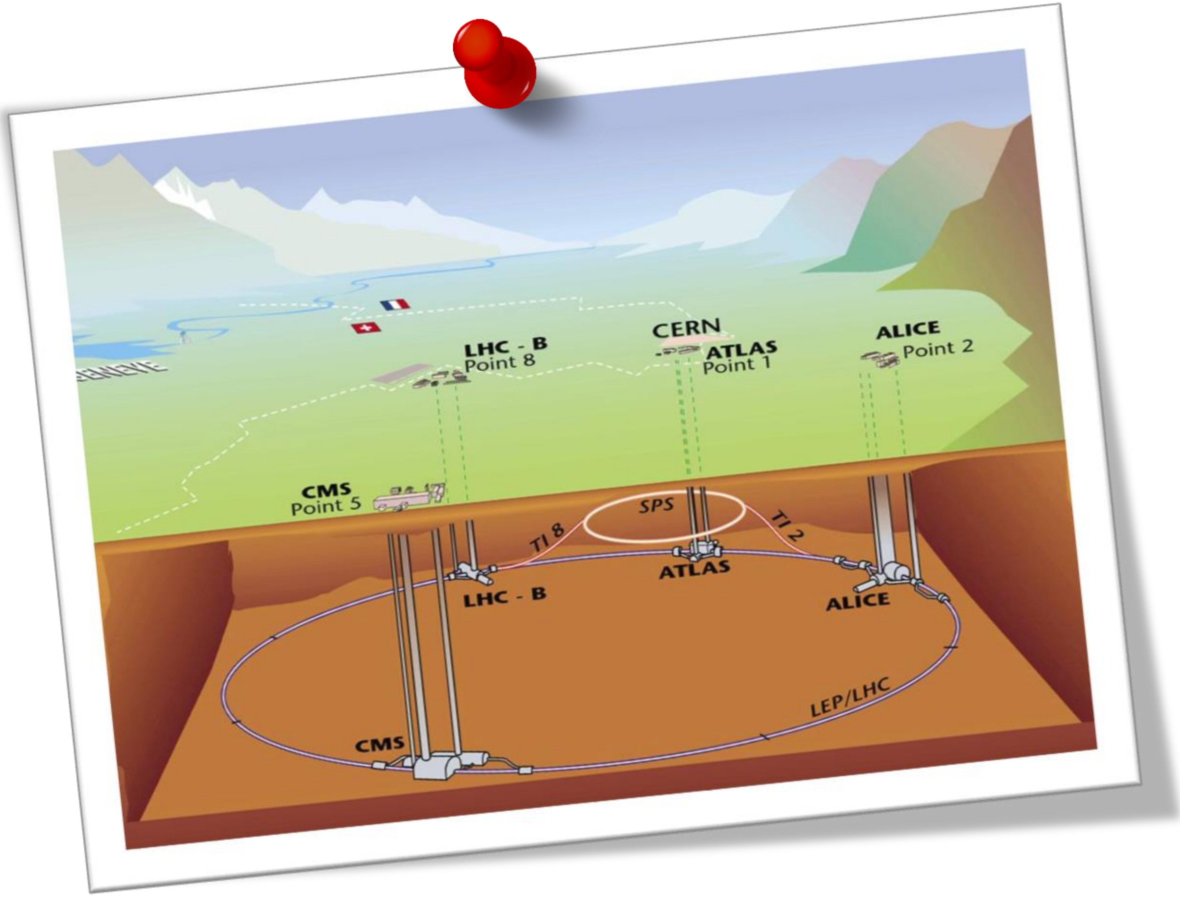
Highly selective & biased choice of topics where machine learning (ML) can be useful in reconstruction of detector signals and physics objects

Calorimeter centric – complex signals especially in the presence of pile-up

Jet focus - composite physics object representing momentum flow generated by partons emitted in a proton–proton collision at LHC

Not at all comprehensive!

Proton–Proton Collisions at the LHC



Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

Schematic model of a proton–proton collision

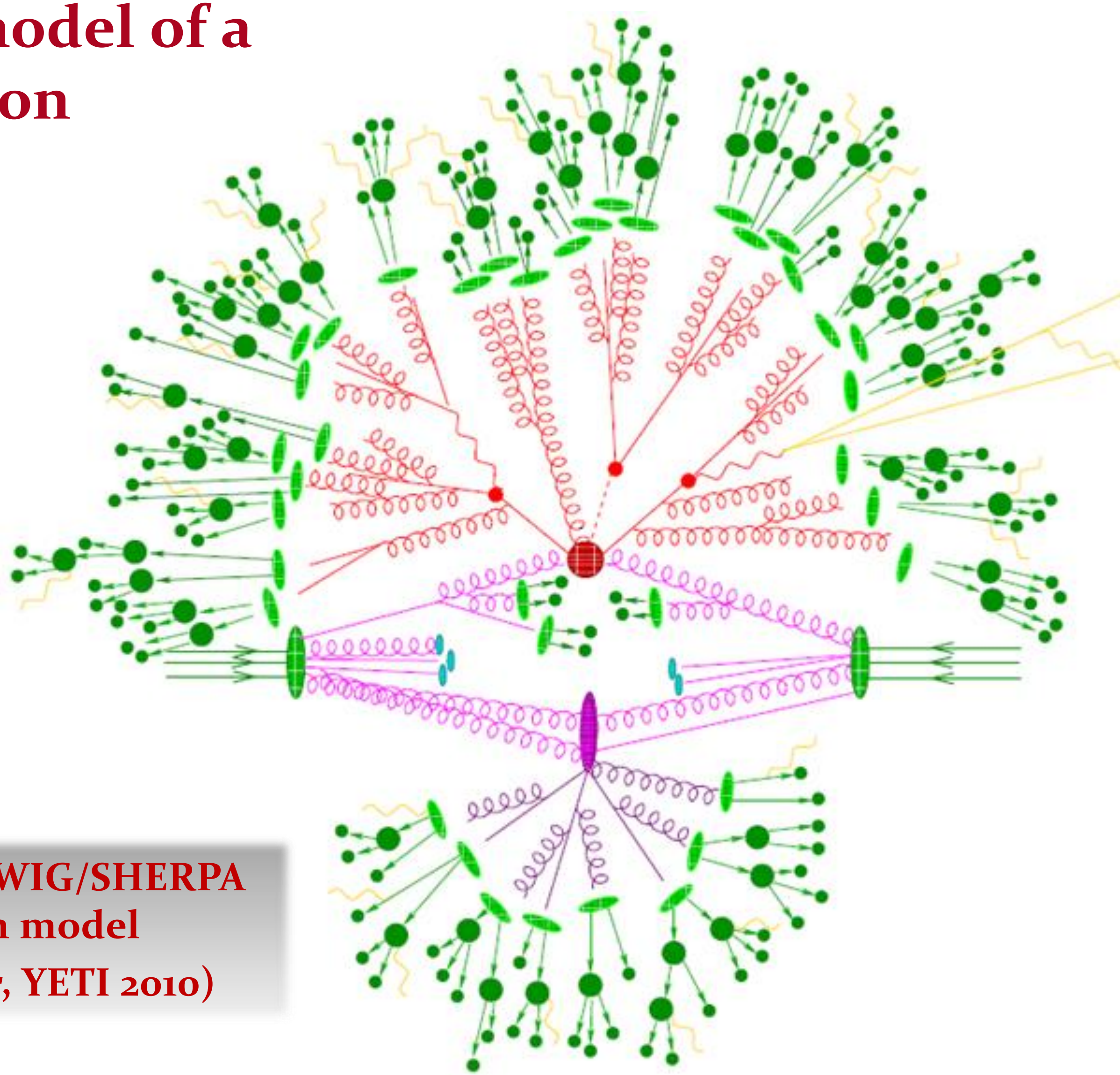
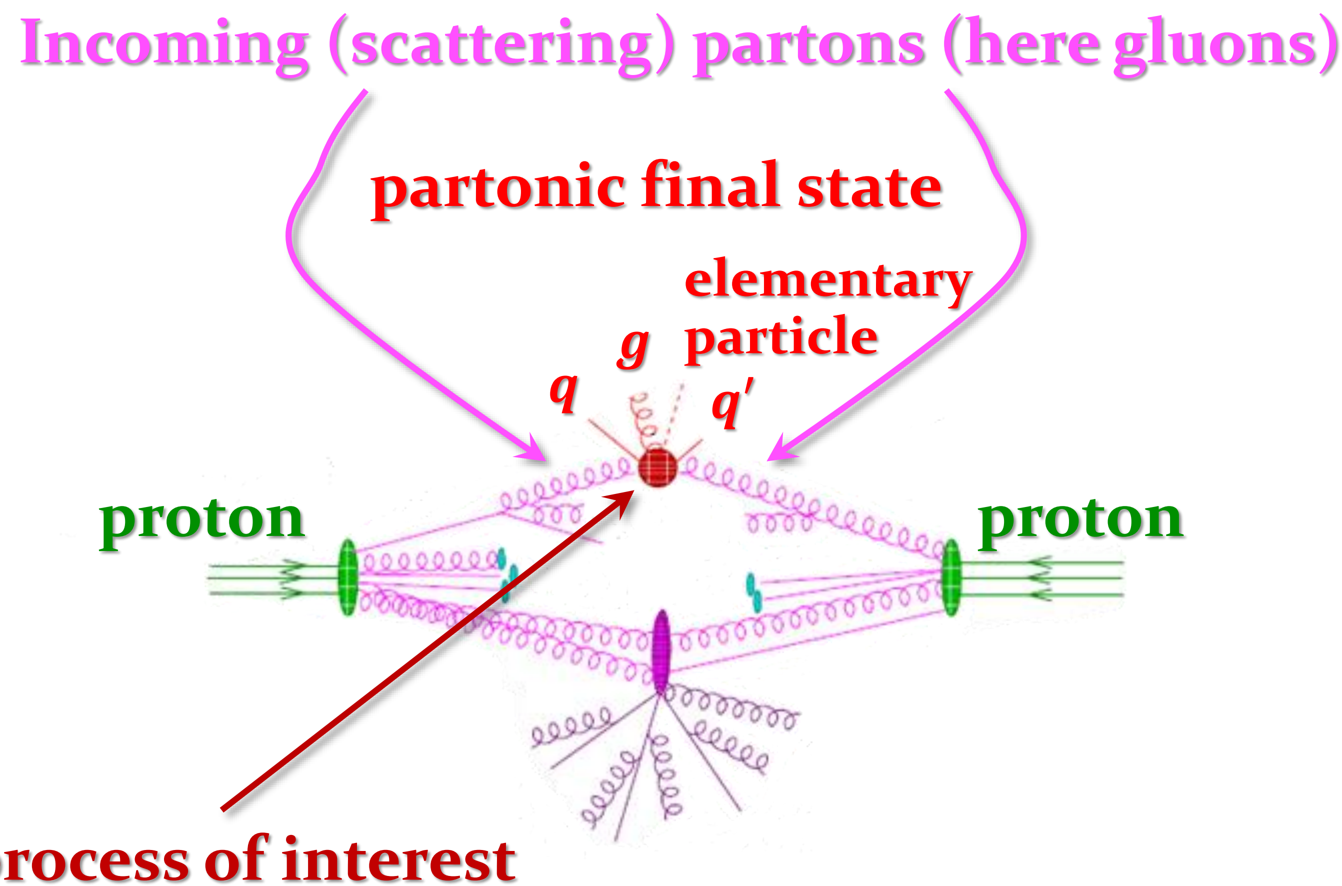


Illustration of HERWIG/SHERPA fragmentation model (Marek Schönherr, YETI 2010)

Schematic model of a proton–proton collision



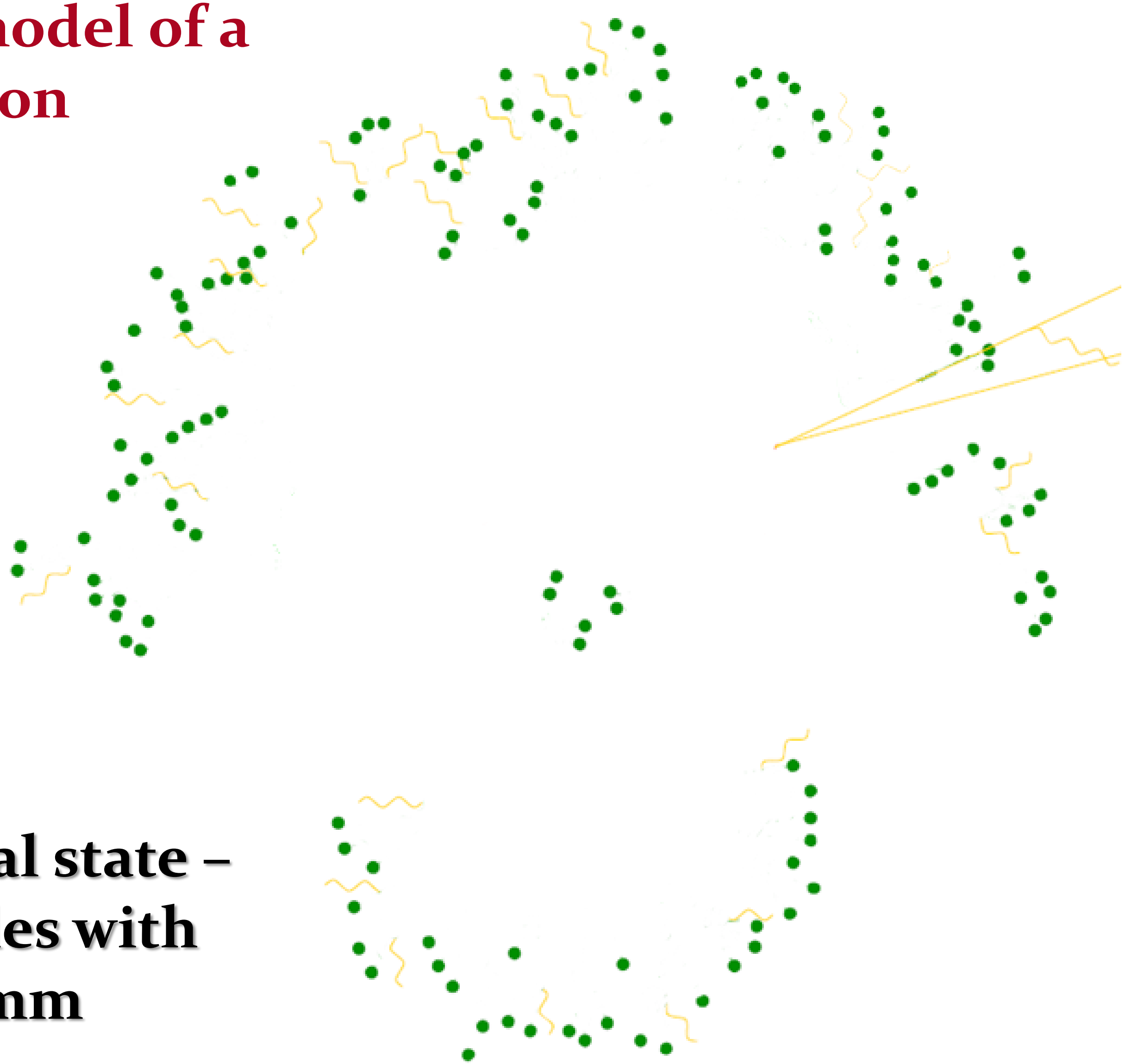
Peter Loch
UAPhysics
The University of Arizona®
College of Science

Schematic model of a proton–proton collision

photons

leptons

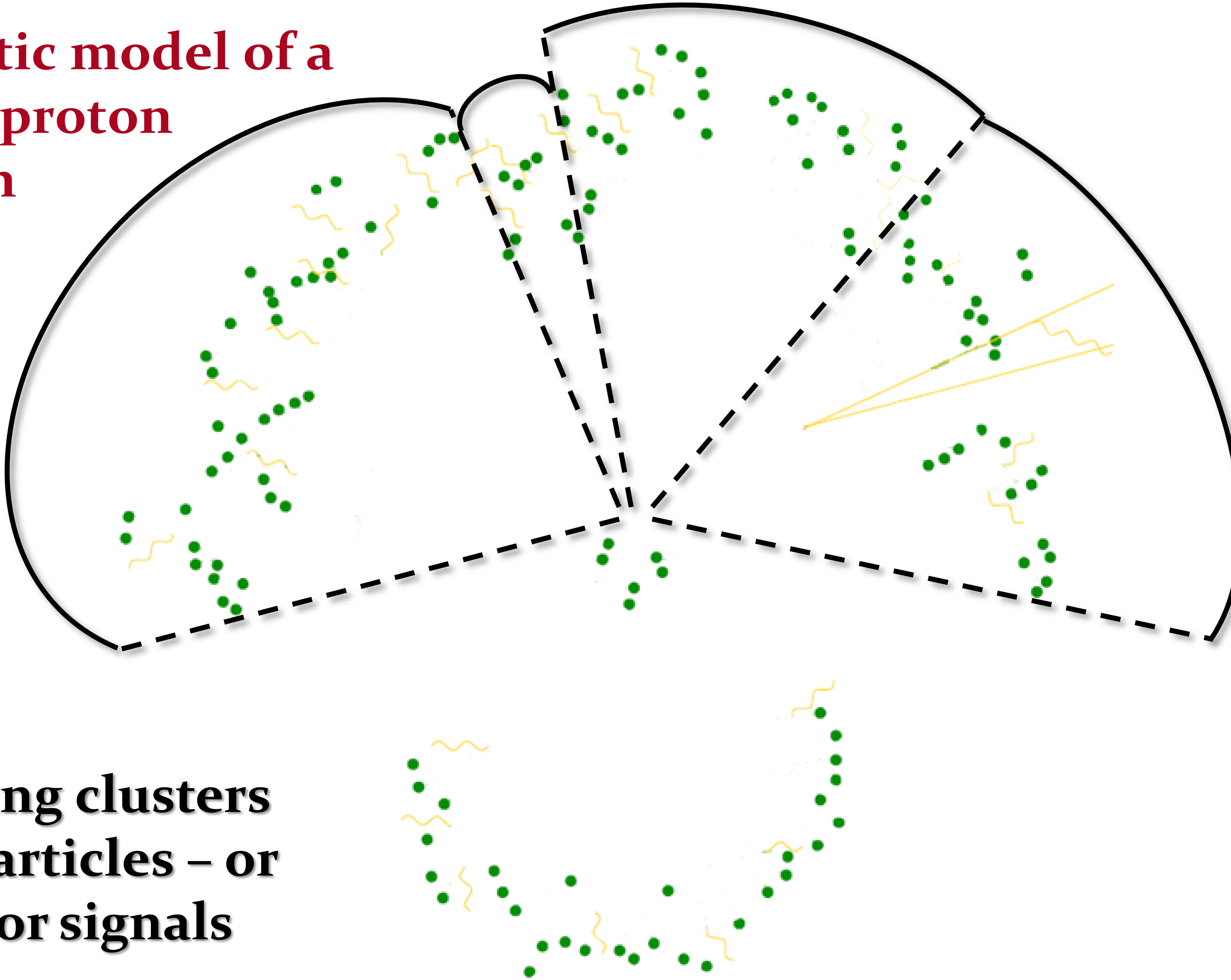
hadrons



detectable final state – stable particles with $c\tau > 10$ mm

Proton-Proton Collisions @ LHC

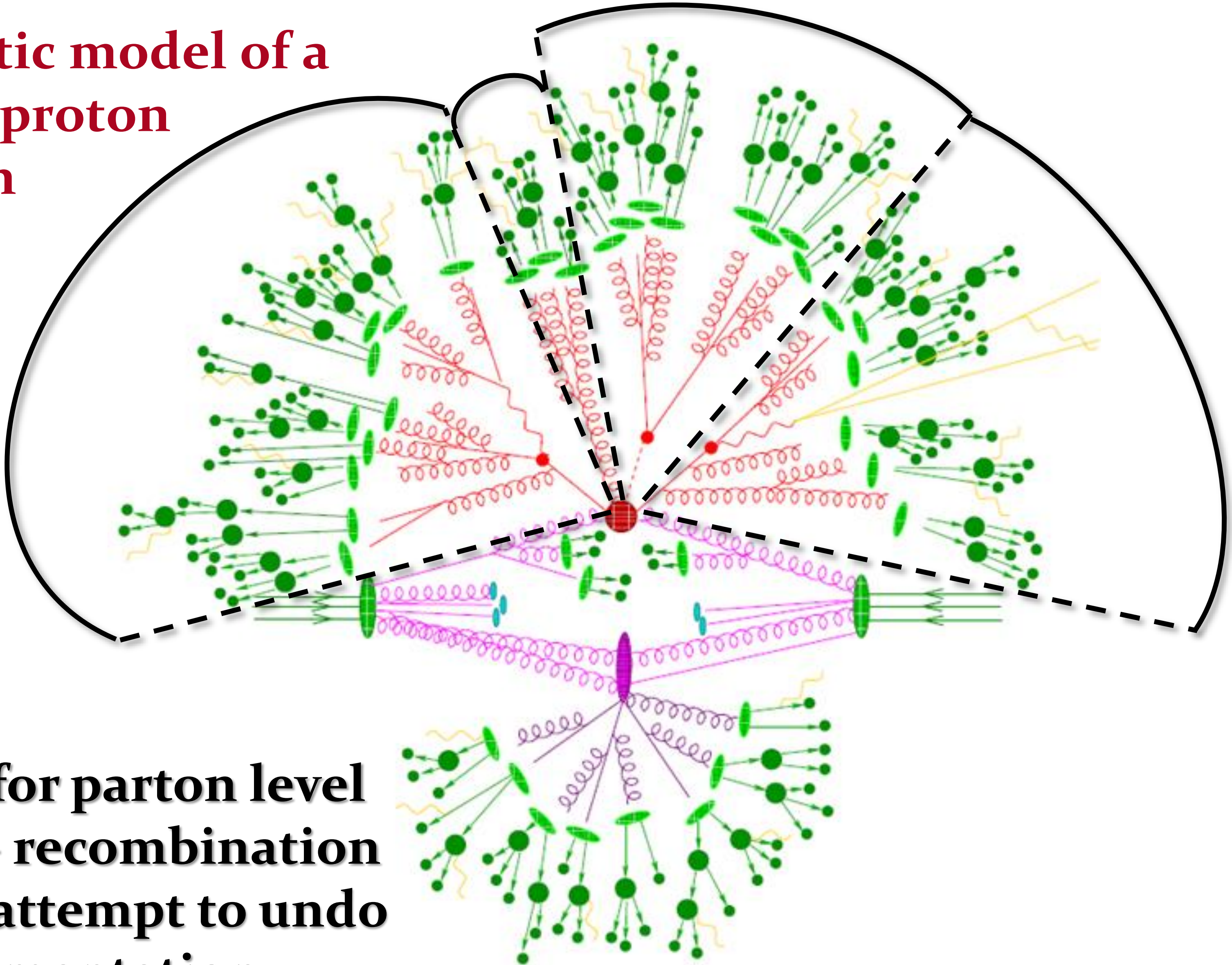
Schematic model of a proton–proton collision



jet finding clusters stable particles – or detector signals

Proton-Proton Collisions @ LHC

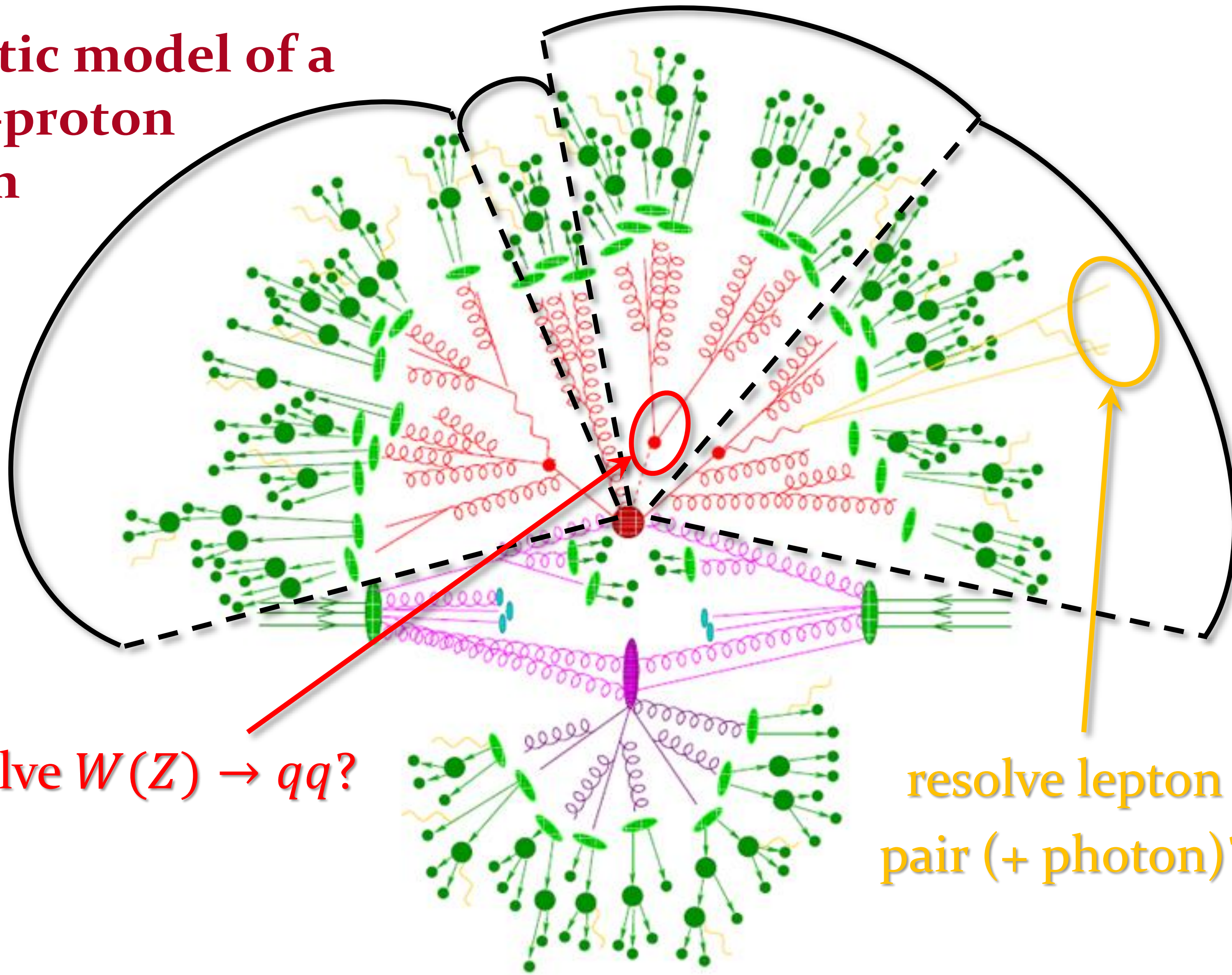
Schematic model of a proton–proton collision



proxies for parton level physics – recombination into jets attempt to undo fragmentation

Proton-Proton Collisions @ LHC

Schematic model of a proton–proton collision



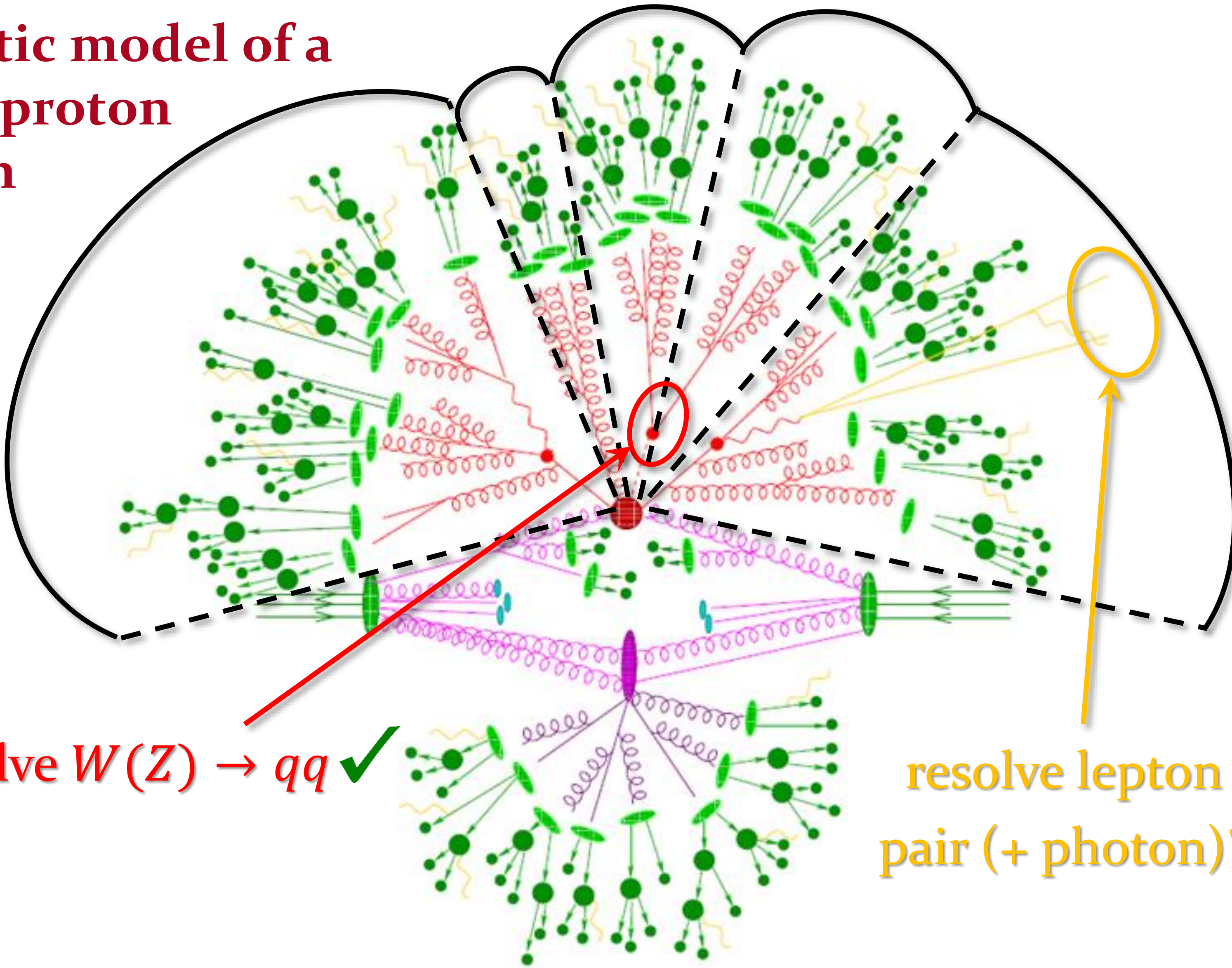
resolve $W(Z) \rightarrow qq$?

resolve lepton pair (+ photon)?

Proton-Proton Collisions @ LHC

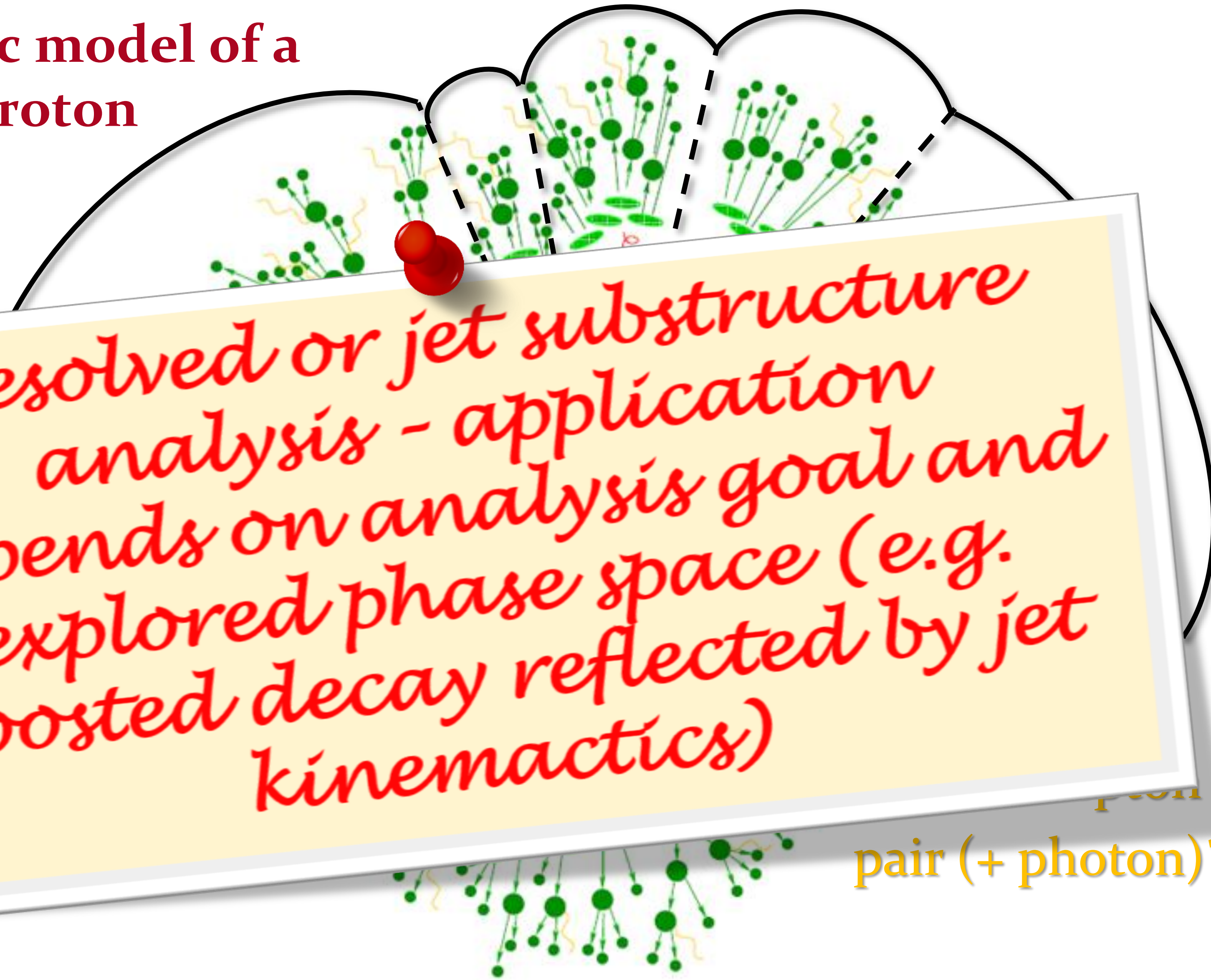
Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

Schematic model of a proton–proton collision



Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

Schematic model of a proton–proton collision



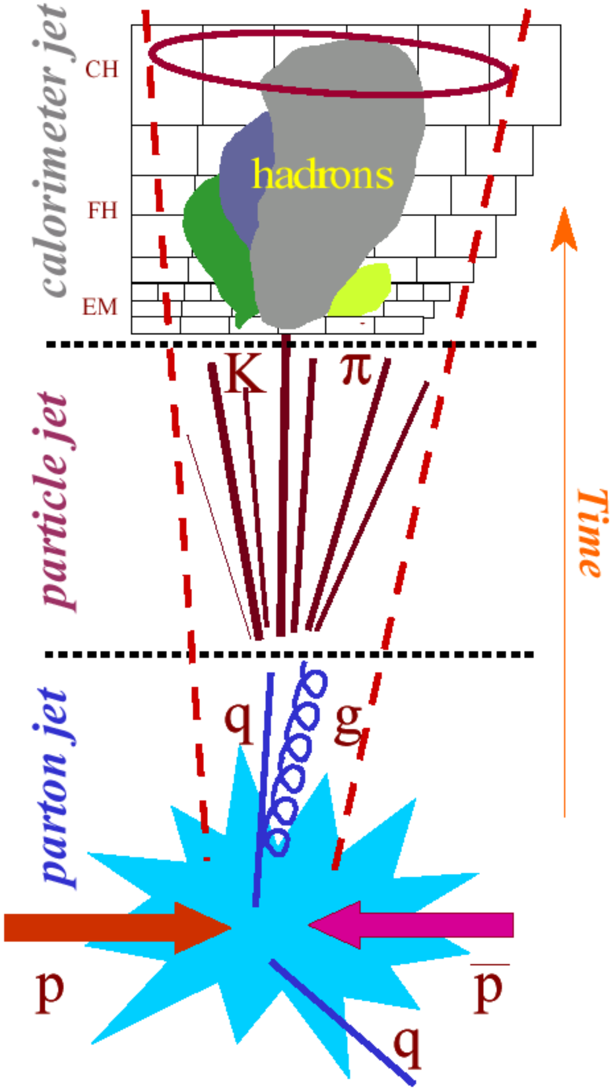
Resolved or jet substructure analysis - application depends on analysis goal and explored phase space (e.g. boosted decay reflected by jet kinemactics)

pair (+ photon)?

Final State (Jet) Reconstruction

Peter Loch
UAPhysics
The University of Arizona
College of Science

Experiment ("Nature")



Jet Reconstruction Challenges

longitudinal energy leakage
detector signal inefficiencies (dead channels, HV…)
pile-up noise from (off- and in-time) bunch crossings
electronic noise
calo signal definition (clustering, noise suppression…)
dead material losses (front, cracks, transitions…)
detector response characteristics (e/h ≠ 1)
jet reconstruction algorithm efficiency
lost soft tracks due to magnetic field

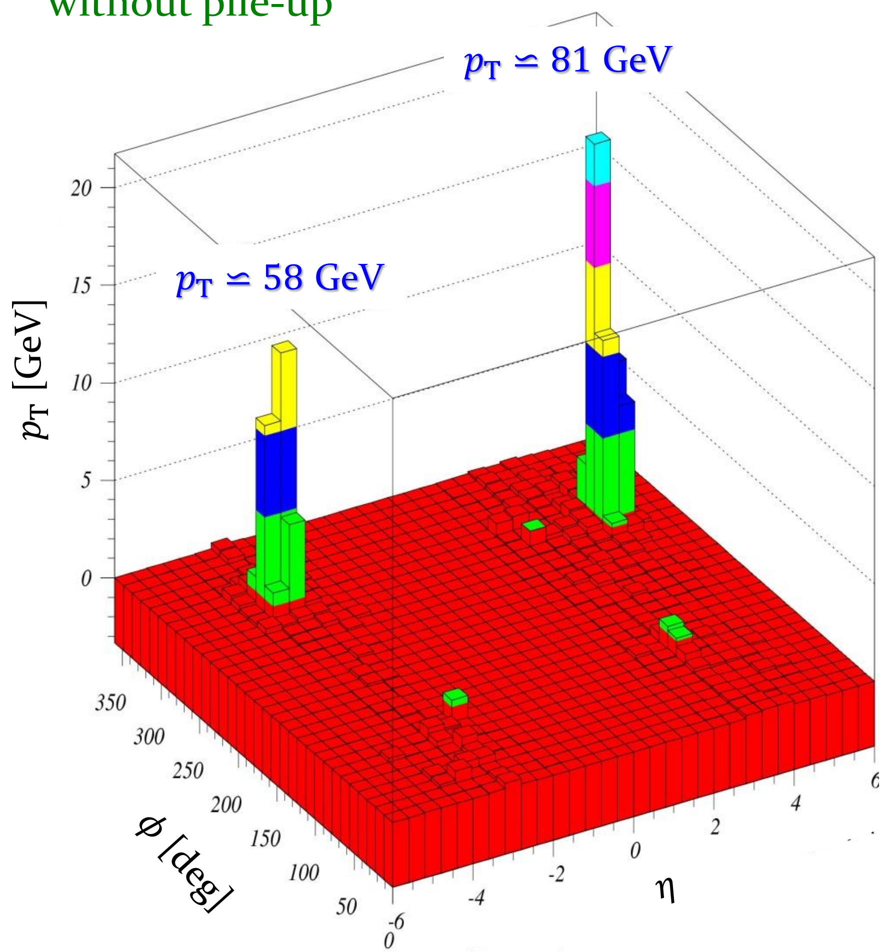
added tracks from underlying event
added tracks from in-time (same trigger) pile-up event
jet reconstruction algorithm efficiency

physics reaction of interest (interaction or parton level)

Pile-up

- Affects jet kinematics and shapes
  - Introduces additional energy/transverse momentum in jets
  - Deteriorates reconstruction of jet mass & internal particle flow features
- Number of pile-up collisions μ
  - $\boldsymbol{\mu = (\mathcal{L}_{\text{inst}} \times \sigma^{pp}_{\text{inel}})/(N_{\text{b}} \times f_{\text{LHC}})}$
  - $\mathcal{L}_{\text{inst}}$ – instantaneous luminosity
  - $\sigma^{pp}_{\text{inel}}$ – inelastic pp cross section
  - N_{b} – number of colliding bunches in LHC
  - f_{LHC} – LHC revolution frequency

without pile-up



$qq \rightarrow qqWW \rightarrow qqH \rightarrow qq\nu\nu\nu\nu$

Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

Pile-up

Affects jet kinematics and shapes

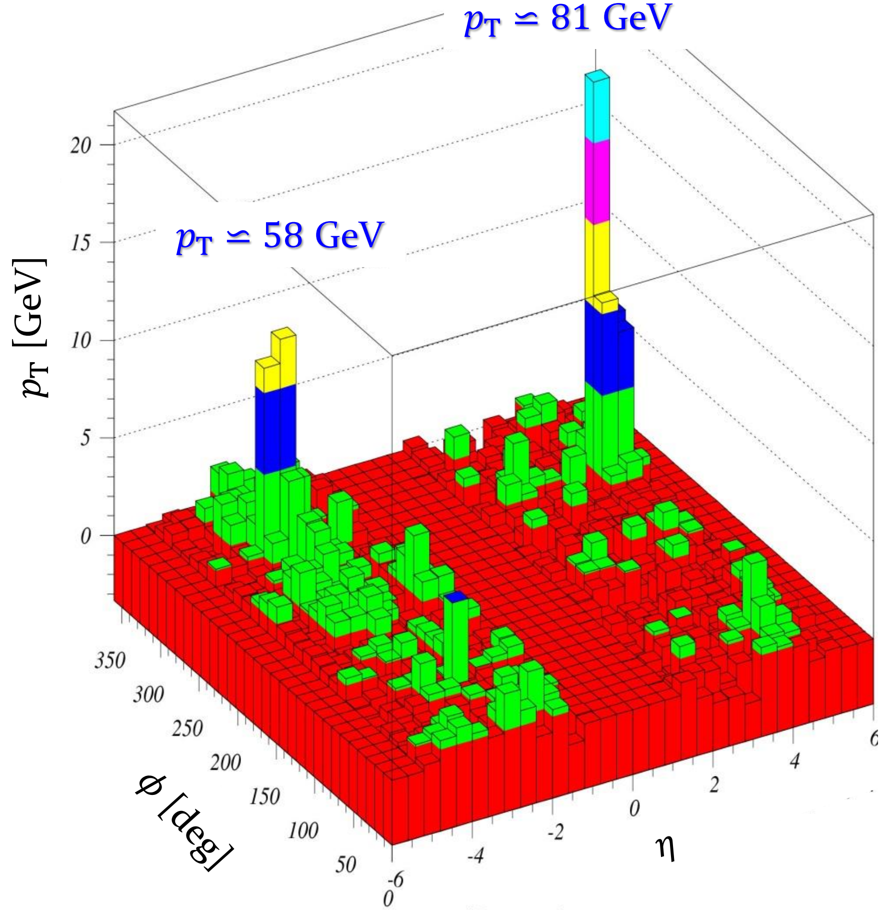
- Introduces additional energy/transverse momentum in jets
- Deteriorates reconstruction of jet mass & internal particle flow features

Number of pile-up collisions μ

$$\boldsymbol{\mu = (\mathcal{L}_{\mathrm{inst}} \times \sigma_{\mathrm{inel}}^{pp})/(N_{\mathrm{b}} \times f_{\mathrm{LHC}})}$$

- $\boldsymbol{\mathcal{L}_{\mathrm{inst}}}$ – instantaneous luminosity
- $\boldsymbol{\sigma_{\mathrm{inel}}^{pp}}$ – inelastic pp cross section
- $\boldsymbol{N_{\mathrm{b}}}$ – number of colliding bunches in LHC
- $\boldsymbol{f_{\mathrm{LHC}}}$ – LHC revolution frequency

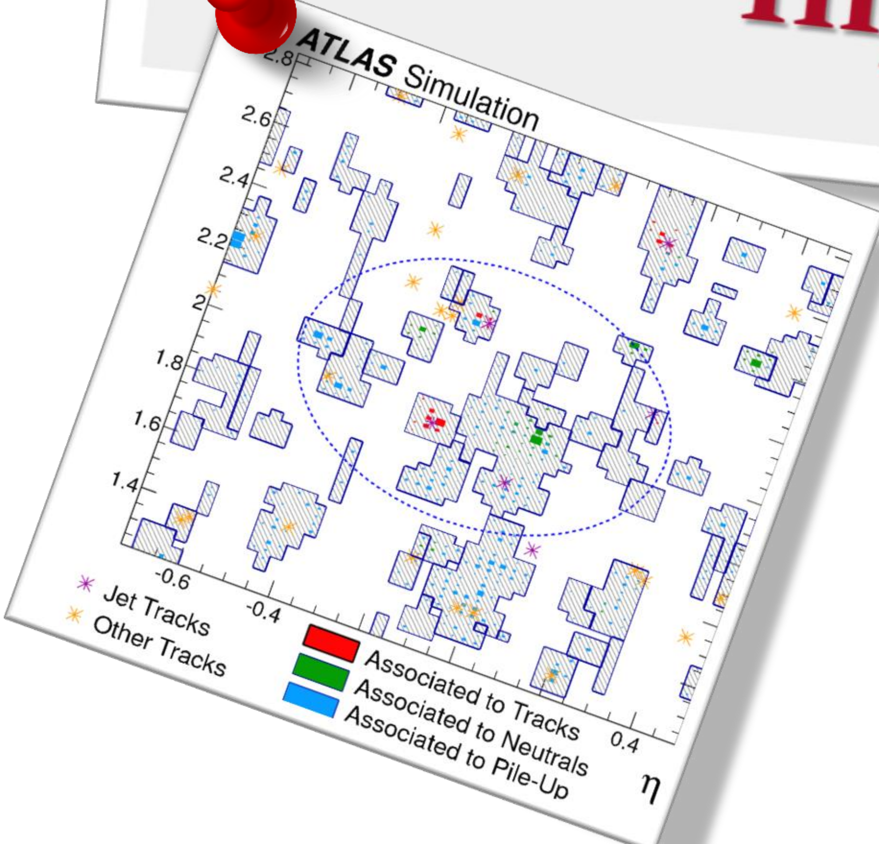
with pile-up @ $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$

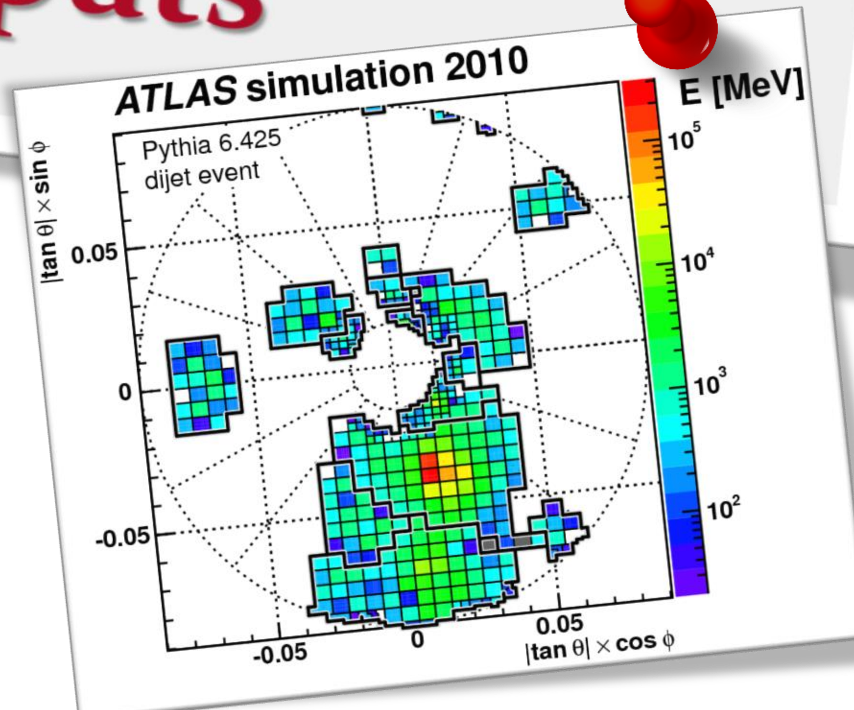


$qq \to qqWW \to qqH \to qq\nu\nu\nu\nu$

Prog.Part.Nucl.Phys.60:484-551,2008

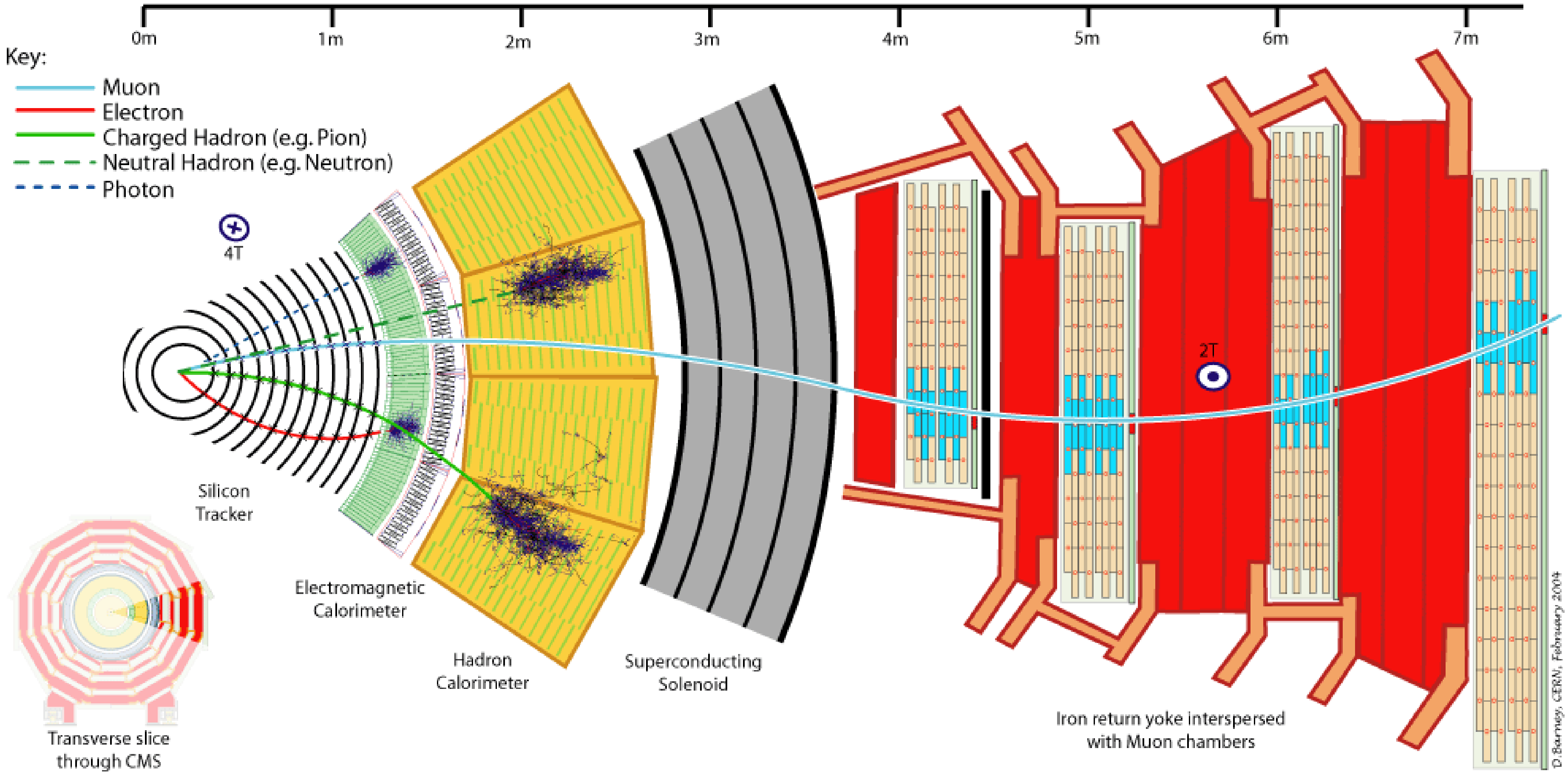
Jet Reconstruction Inputs





Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

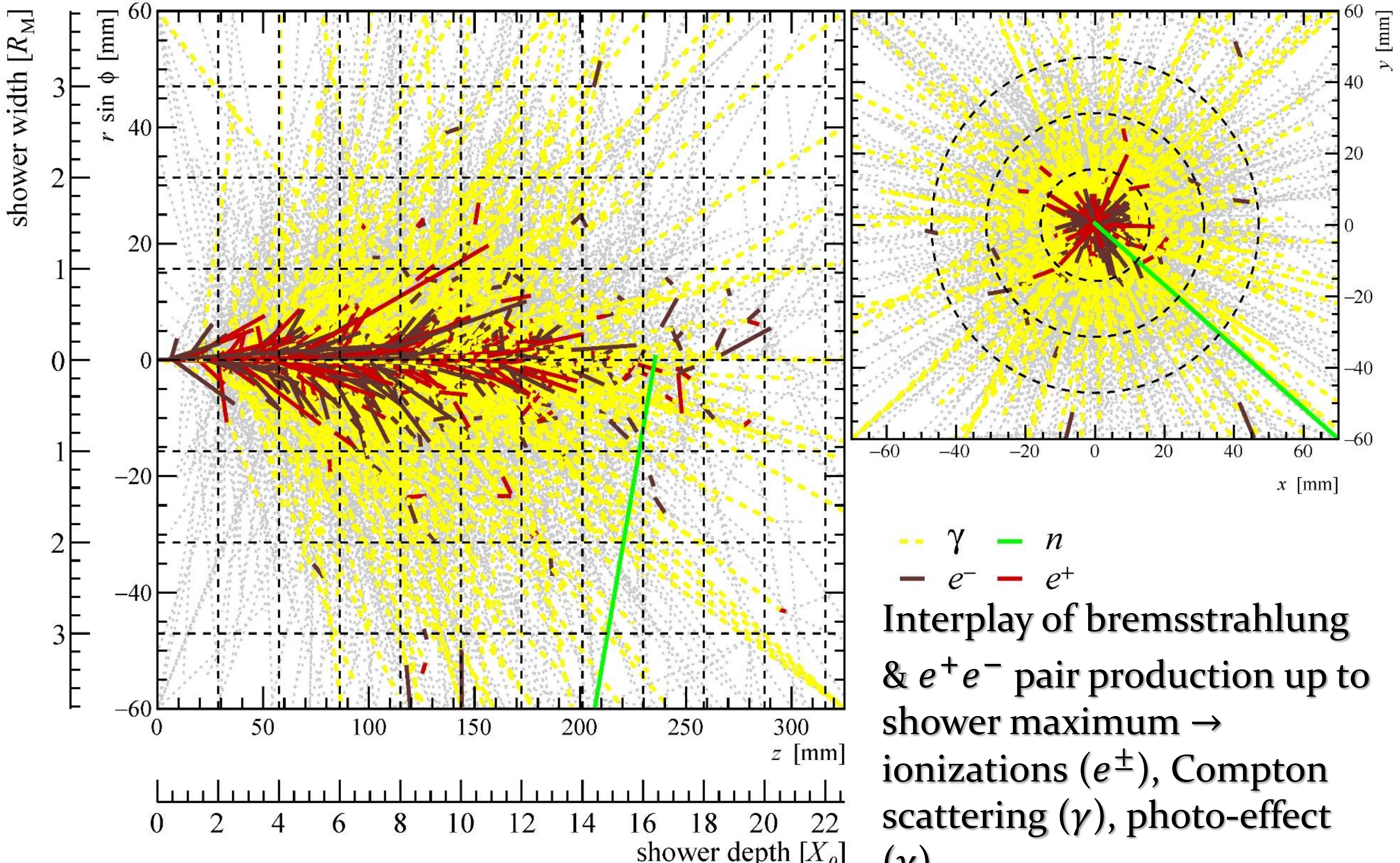
Basic Detector Design



Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

Electromagnetic Showers

GEANT4 Simulation: 10 GeV e^- in copper

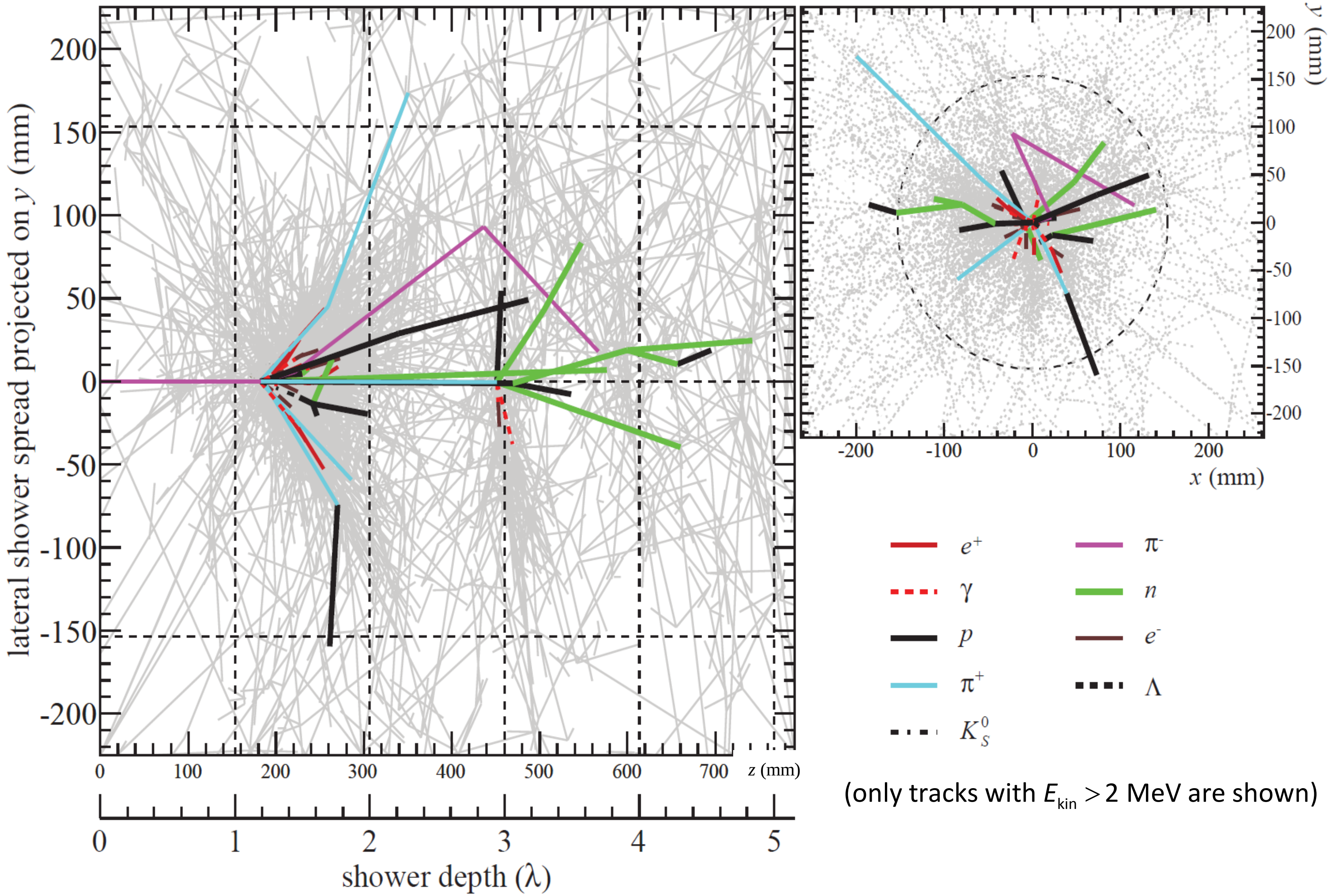


Interplay of bremsstrahlung & e^+e^- pair production up to shower maximum → ionizations ($e^\pm$), Compton scattering (γ), photo-effect (γ)

Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

Hadronic Showers

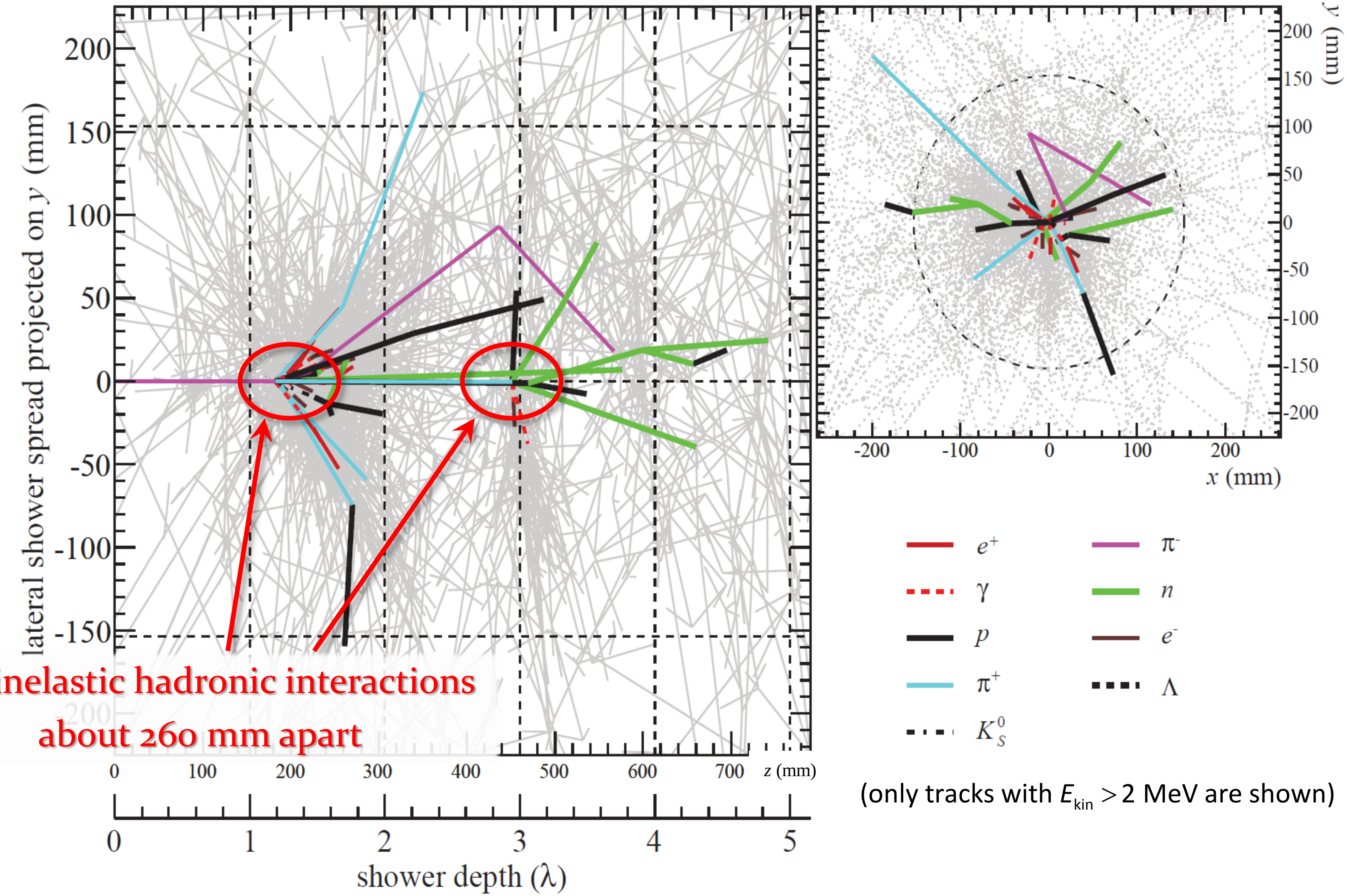
GEANT4 Simulation: 10 GeV π^+ in copper



(only tracks with $E_{\mathrm{kin}} > 2$ MeV are shown)

Hadronic Showers

GEANT4 Simulation: 10 GeV π^+ in copper



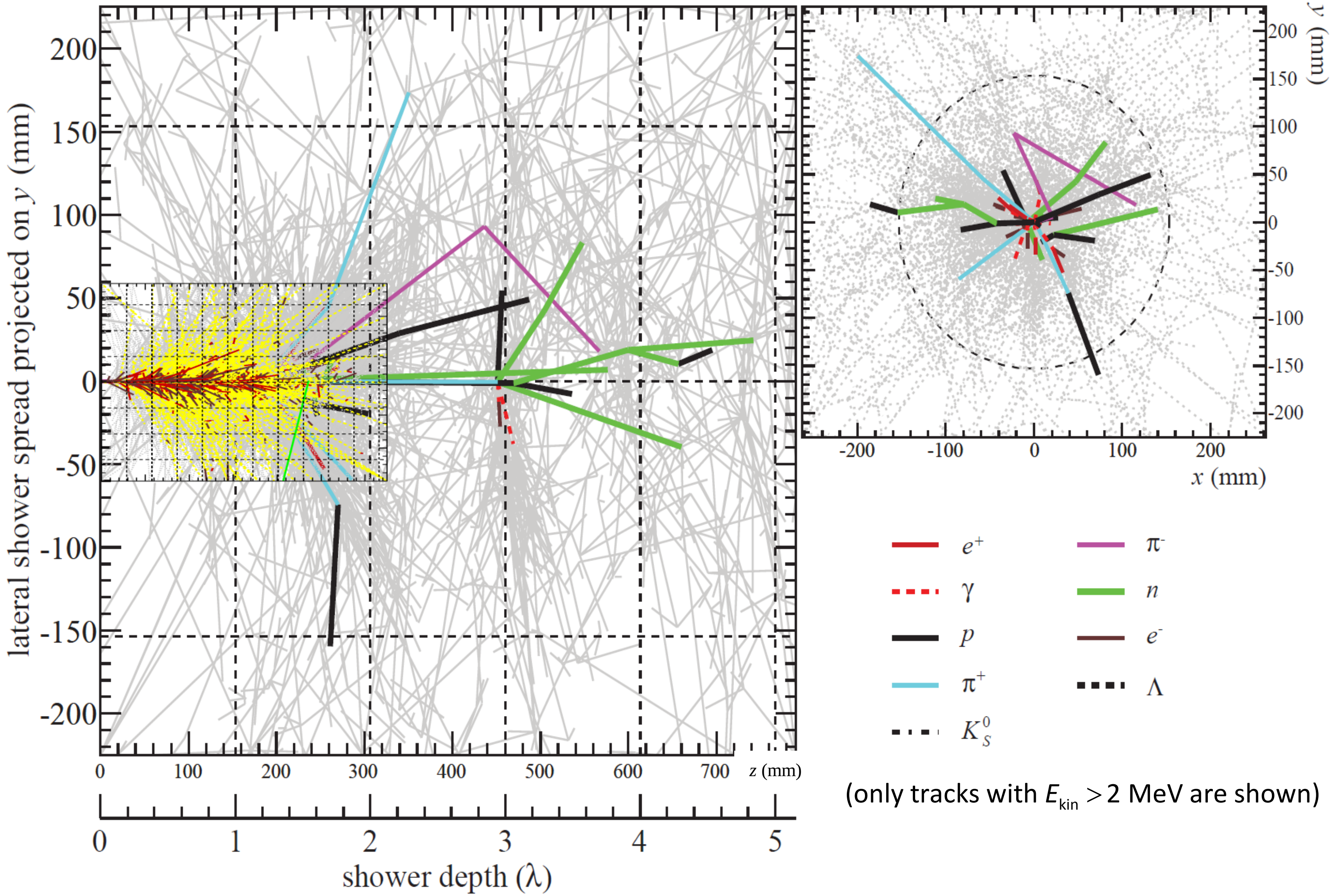
two inelastic hadronic interactions
about 260 mm apart

(only tracks with $E_{\text{kin}} > 2$ MeV are shown)

Hadronic Showers

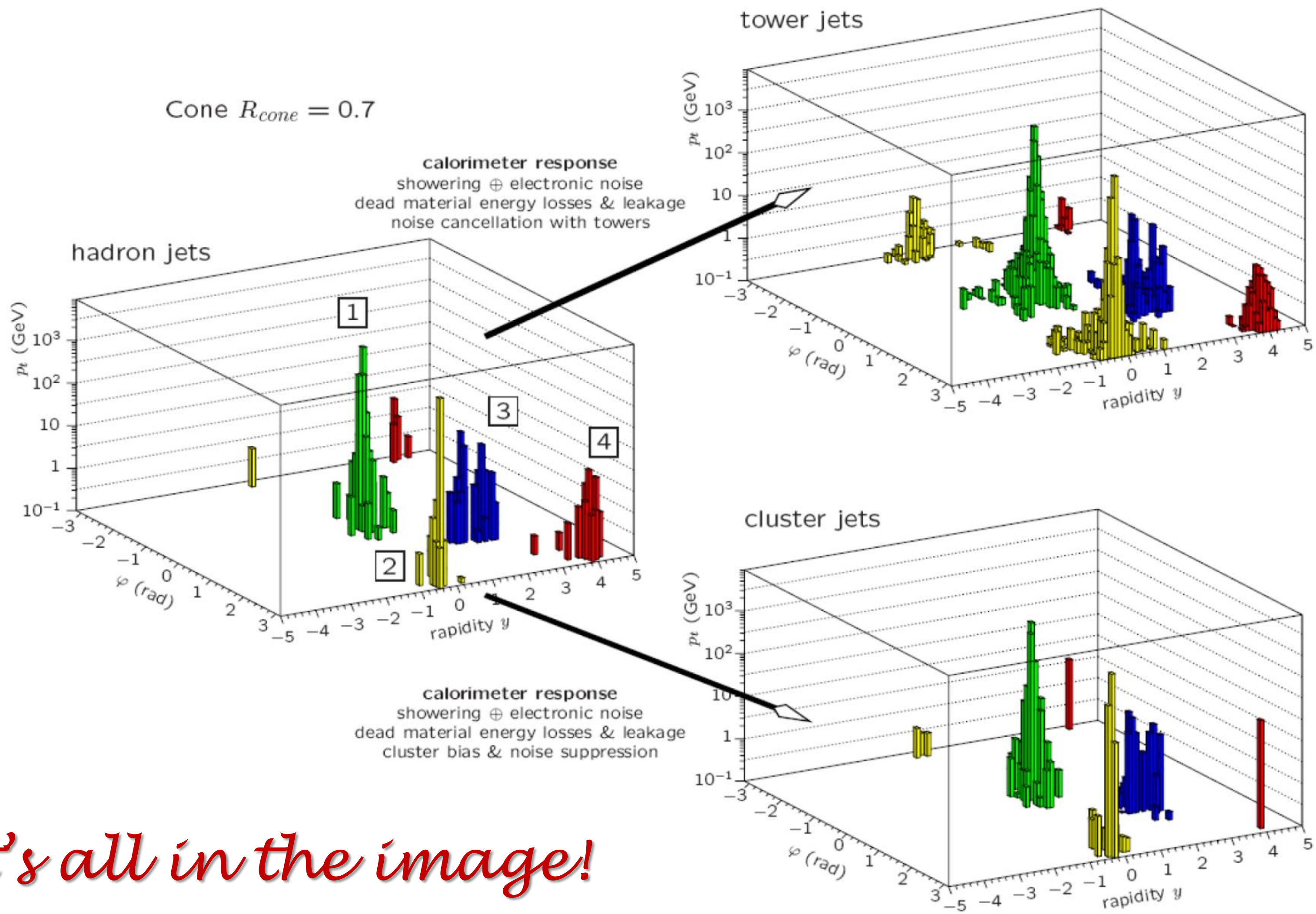
Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

GEANT4 Simulation: 10 GeV π^+ in copper



(only tracks with $E_{\mathrm{kin}} > 2$ MeV are shown)

Calorimeter Jets



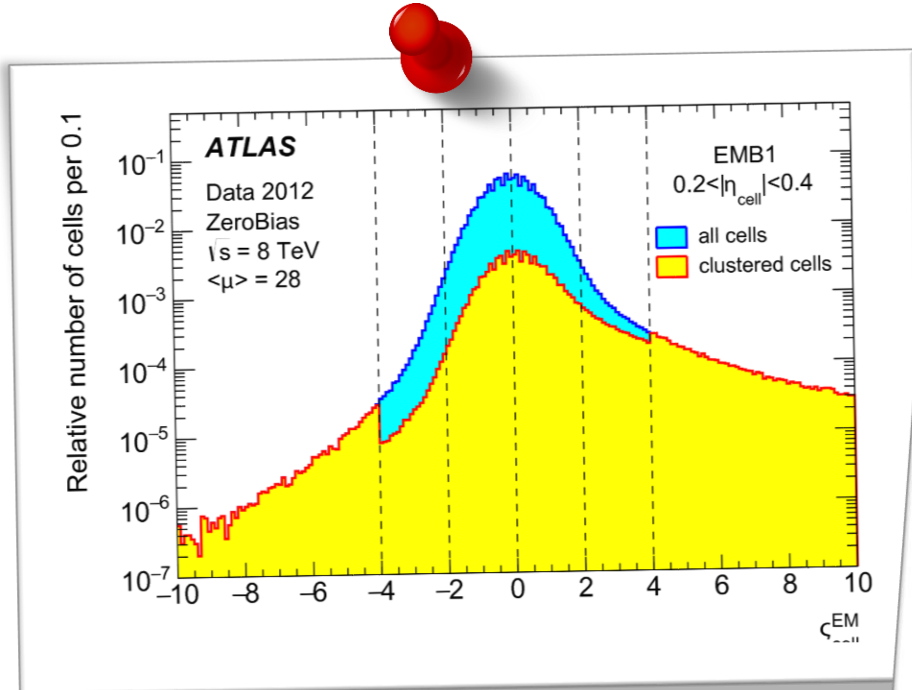
It's all in the image!

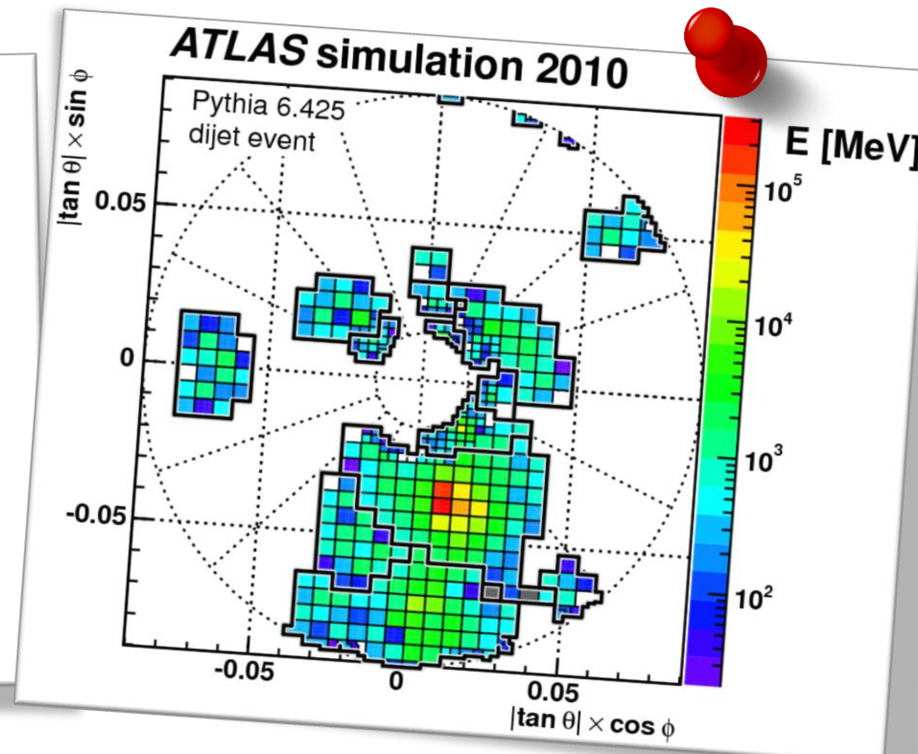
S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch, M. Toennesmann, **Prog.Part.Nucl.Phys.60:484-551,2008**

Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

Follow signal significance patterns

Extract calorimeter cells with significant signals or topologically connected to significant signals (high signal-over-noise)

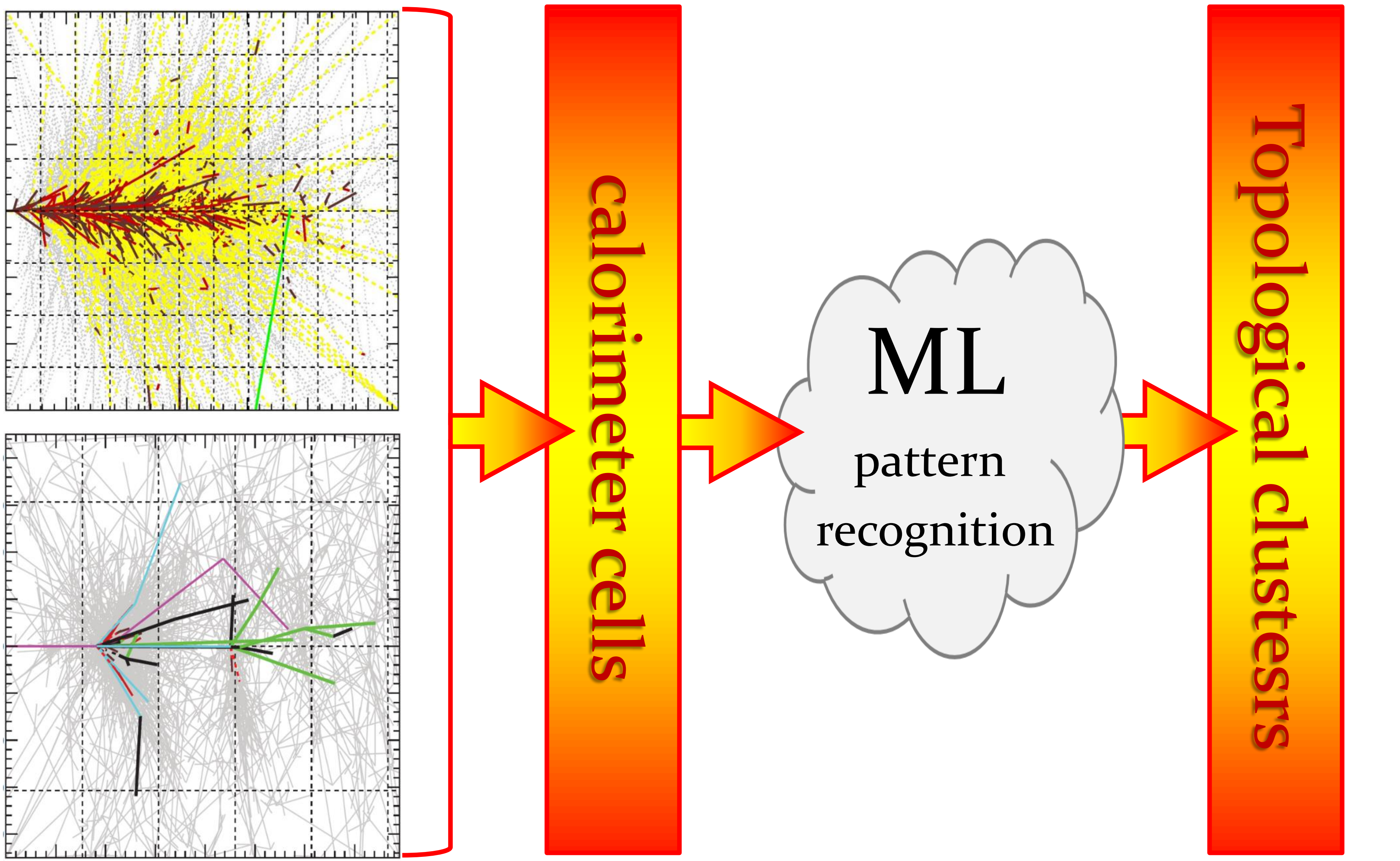




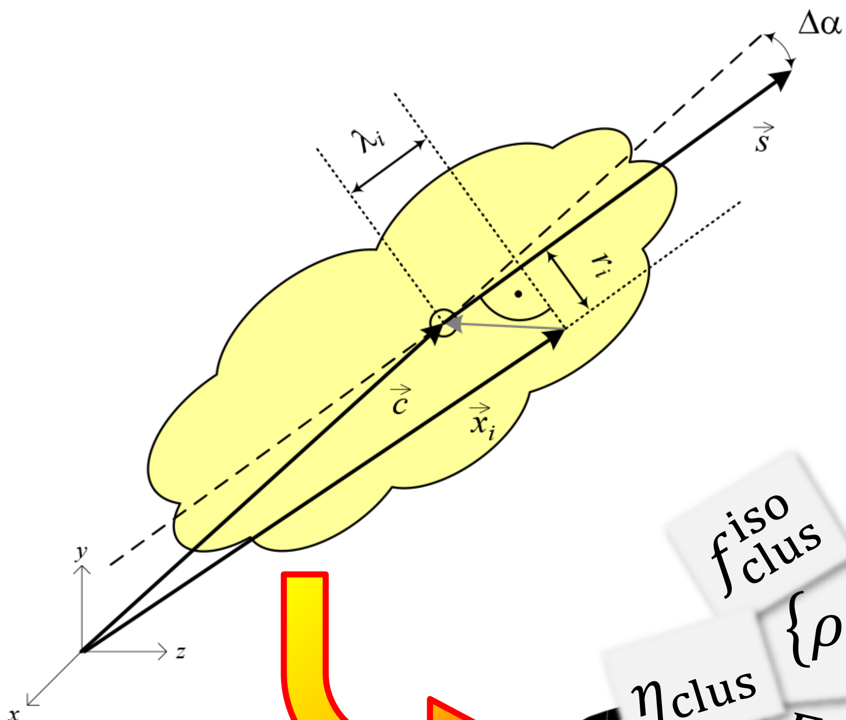
Applies splitting algorithm to resolve spatial energy flow

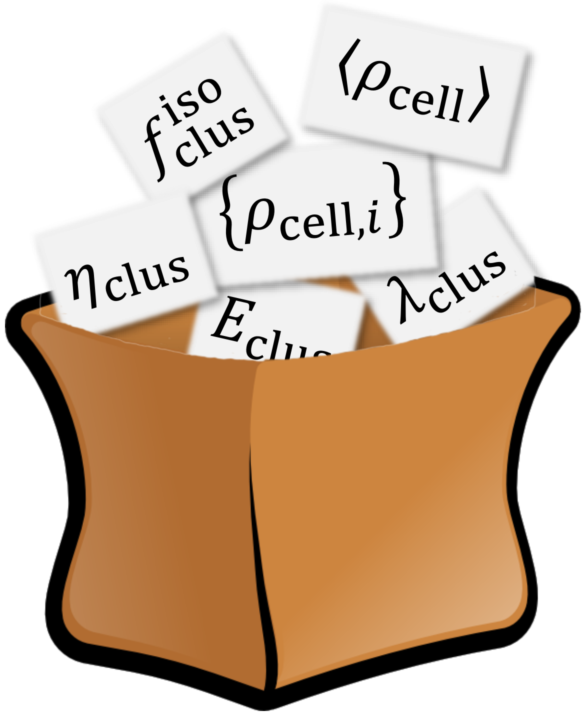
Peter Loch
UAPhysics
The University of Arizona
College of Science

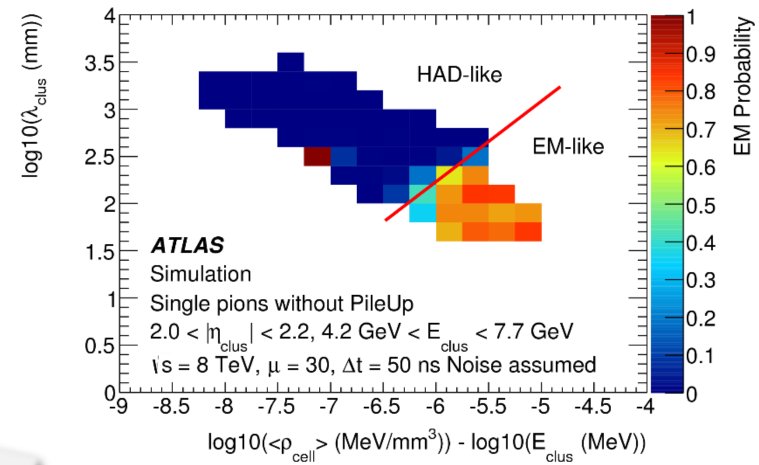
ML for Topo-Cluster Formation



Peter Loch
UAPhysics
The University of Arizona®
College of Science







$$(E,\eta,\varphi)_{\mathrm{clus}} \mapsto (E,\eta,\varphi)_{\mathrm{true}}$$

Topo-Cluster Calibration Parameter Choices

Physics & detector

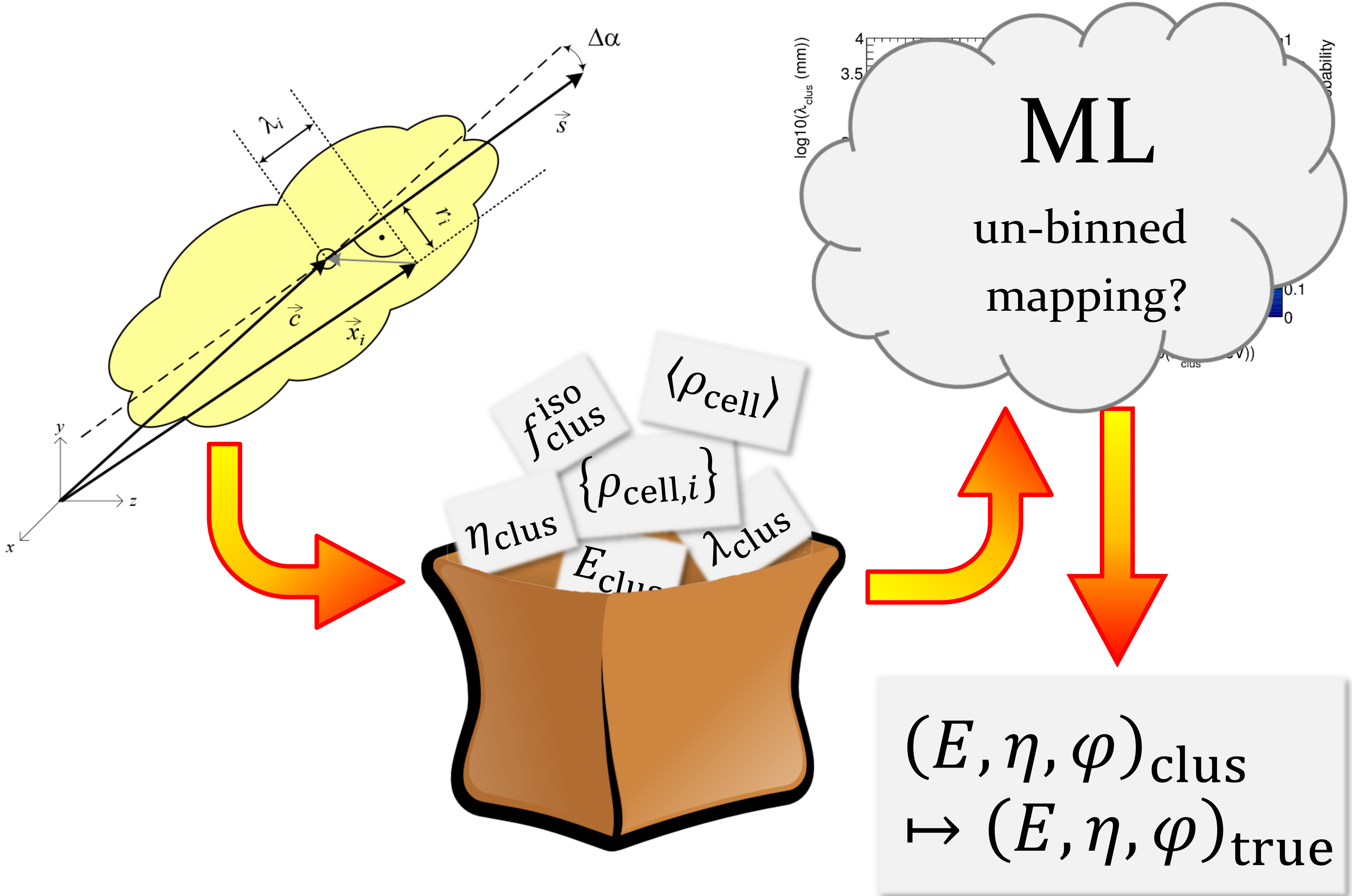
- Sensitive to differences in electromagnetic and hadronic cascades – different response (≡ signal from electrons/photons larger than signal from hadrons at the same deposited energy)
- Reflecting local response characteristics in calorimeter regions and boundaries

Finding observables

- Checked if simulation and experimental data agree – chose best described parameters also in pile-up signal environment
- Ran multi-variate tests (Fischer discriminant) to verify choices – seem to have found the best ones right away

Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

Topo-Cluster Calibration With ML



Jet Substructure

Why Looking Inside Jets at LHC?

Kinematic reach at LHC allows production of (highly) boosted particles

- All decay products of hadronically decaying Standard Model particles can be collected into single jet
  - $W \to qq$, Higgs bosons $H \to bb$, top quarks $t \to Wb \to qqb$
- Searches for new heavy particles with boosted (SM) decay products
  - Single jet mass indicative observable for new particle production

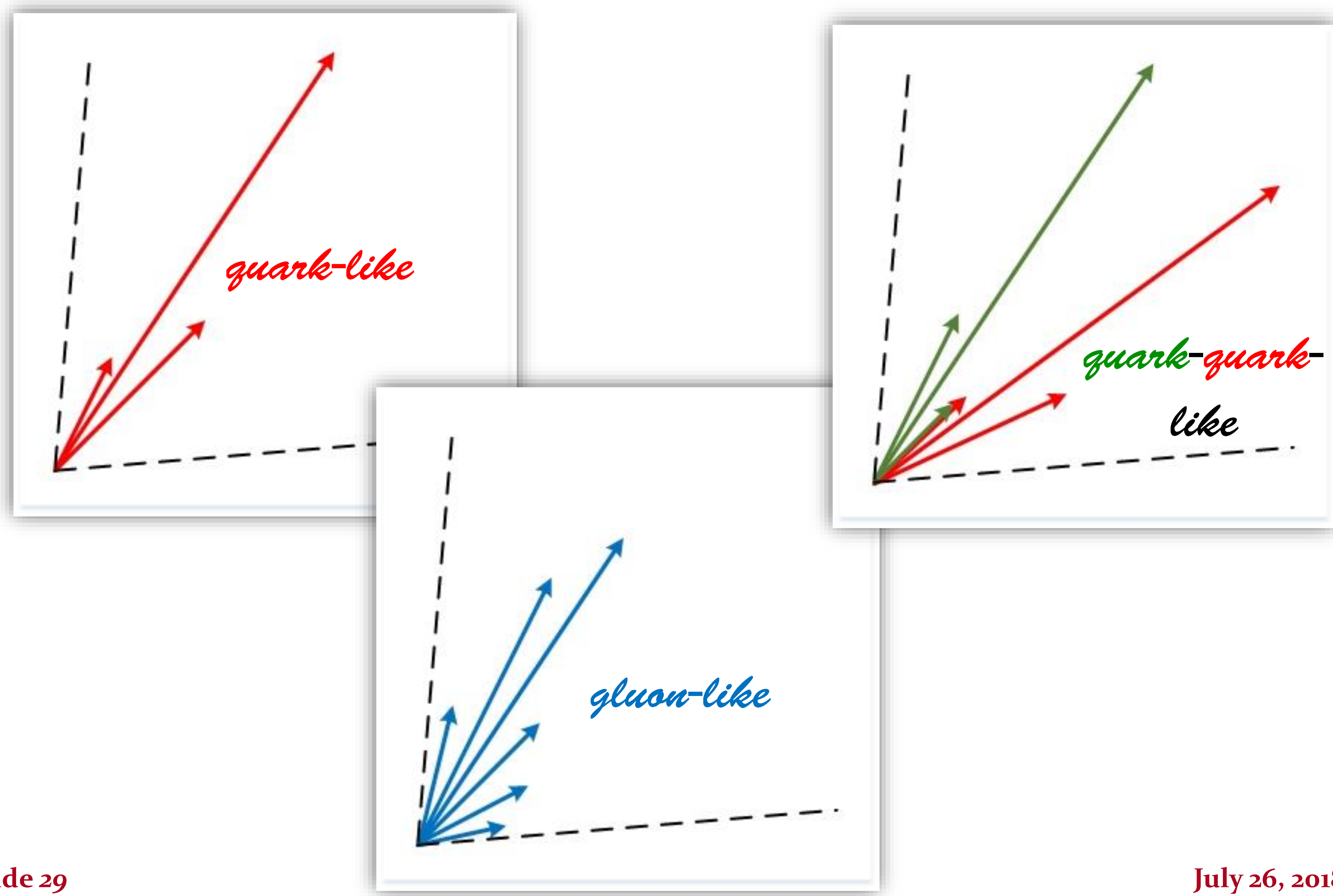
Experimental challenges – detector granularity & pile-up

- Limited granularity introduces limits in resolving sub-structure in a jet
  - E.g., signals from near-by particles can easily overlap in calorimeters
- Presence of pile-up can disturb single jet mass reconstruction from internal flow measurements
  - Needs to extract relevant internal jet energy flow structures for mass reconstruction etc. from diffuse pile-up contributions

Jet substructure analysis

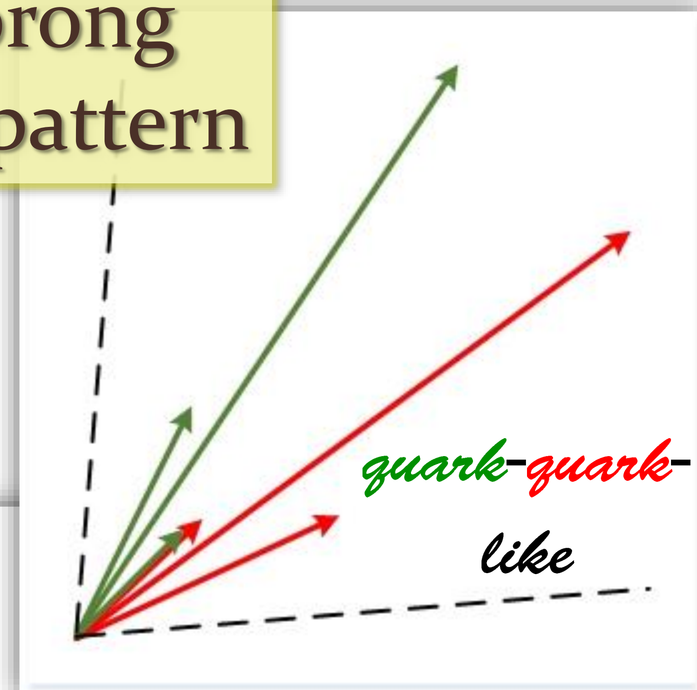
- Collection of techniques aiming at enhancing two- or three-particle (“prongs”) decay patterns in single jets
  - Typically leads to suppression of QCD-like backgrounds from quark- and gluon jets with their typical parton shower and fragmentation driven internal flow structure

Separation of Flow Patterns

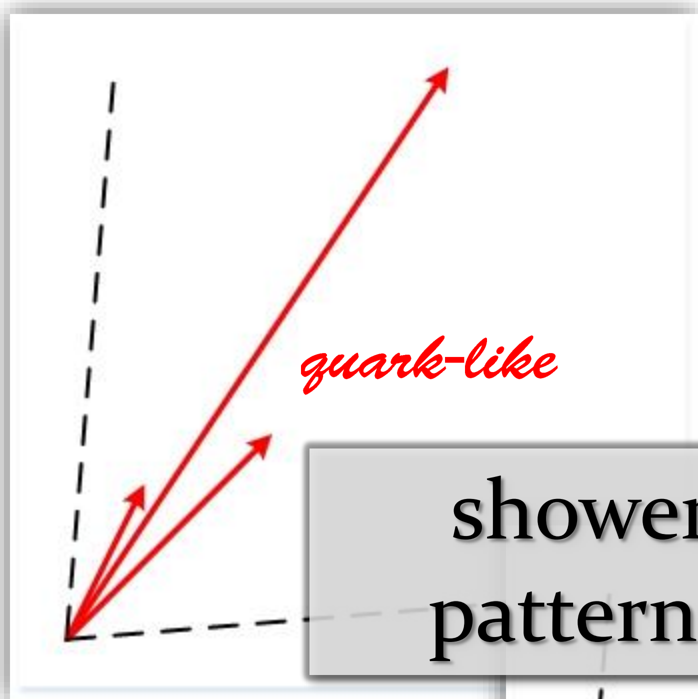


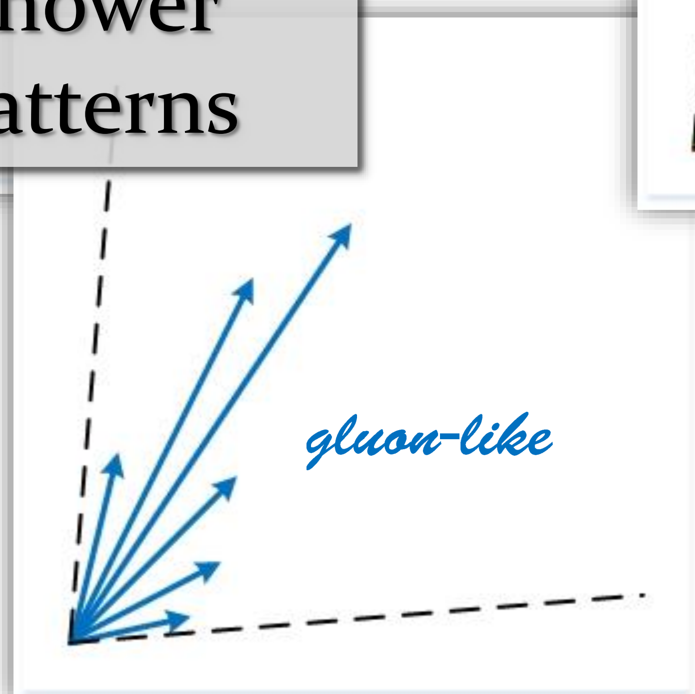
Separation of Flow Patterns

N*-prong decay pattern



shower patterns

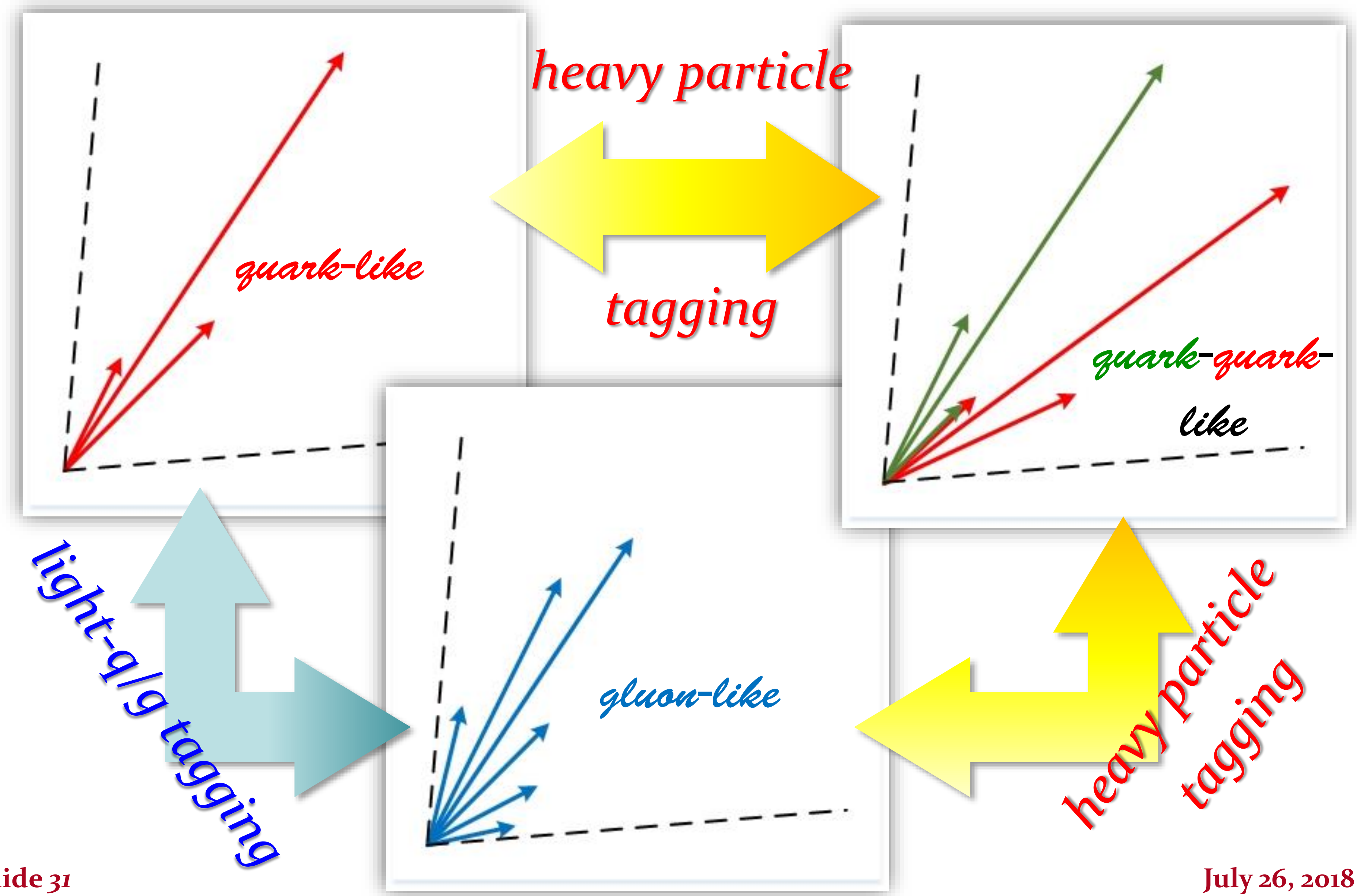




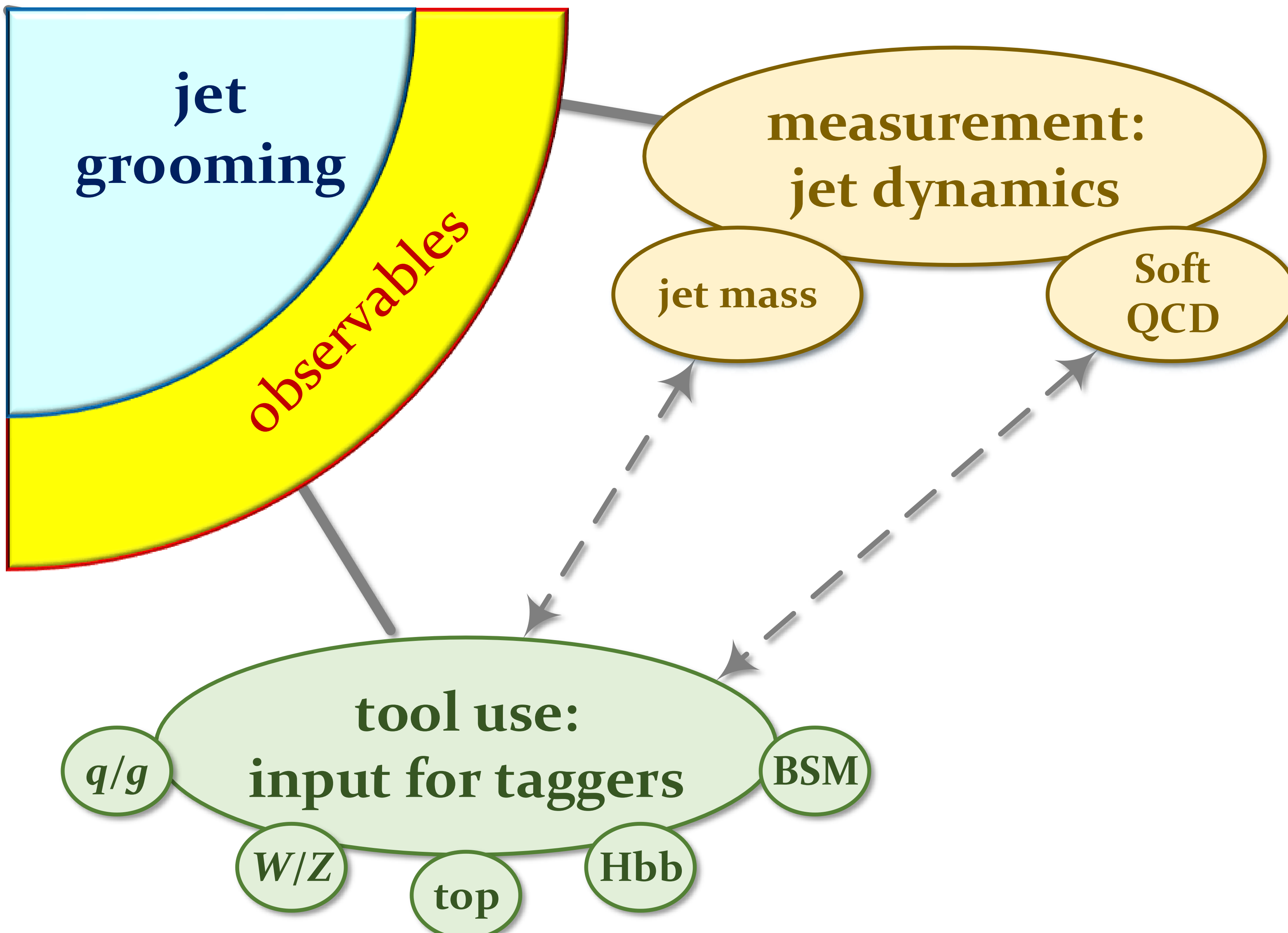
*staged fast decays like $t \to Wb \to qqb$, boosted system of > 1 decaying particle, like in $HH \to bbbb$ etc.

Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA®
College of Science

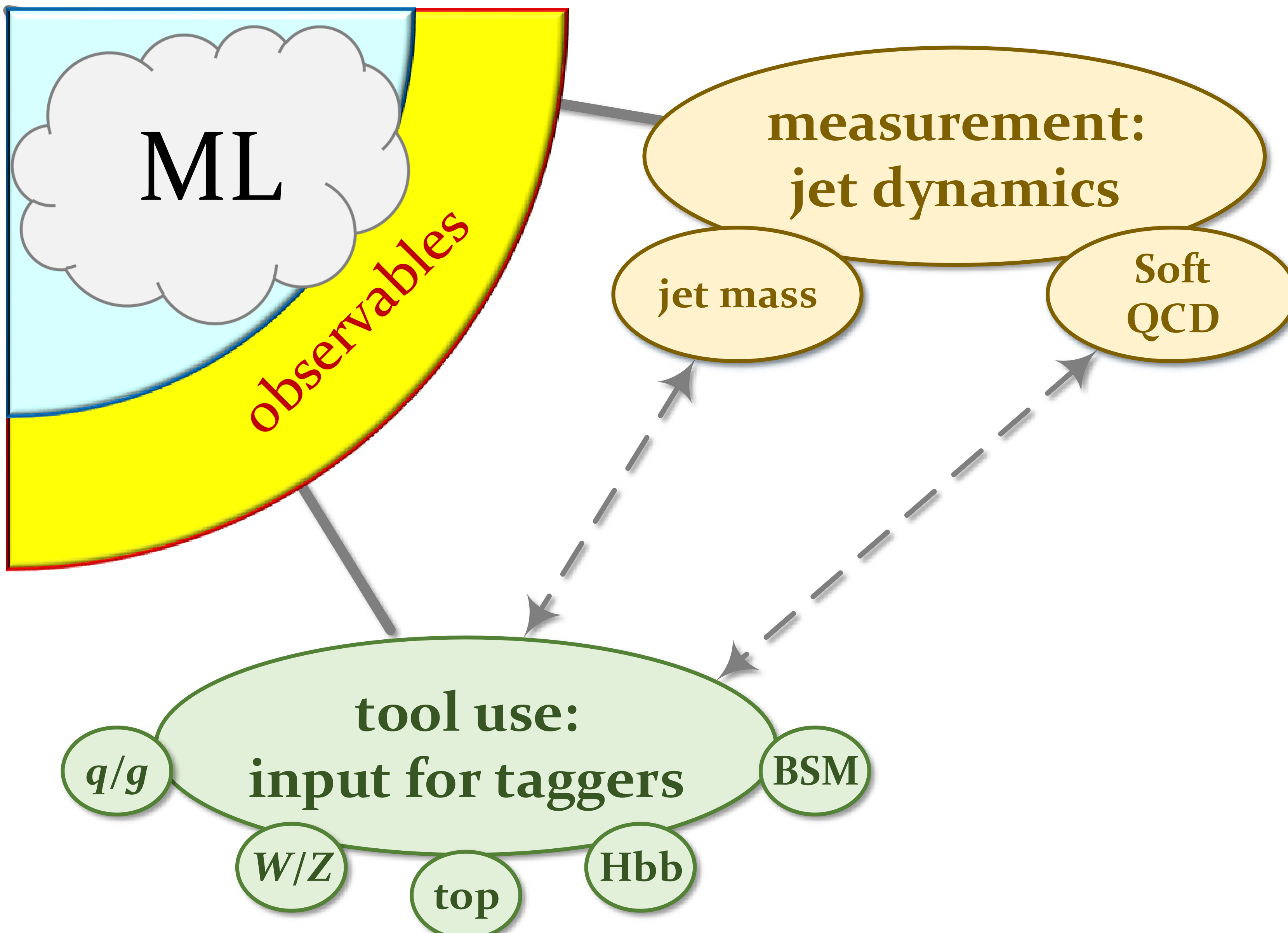
Separation of Flow Patterns



Peter Loch
UAPhysics
The University of Arizona®
College of Science

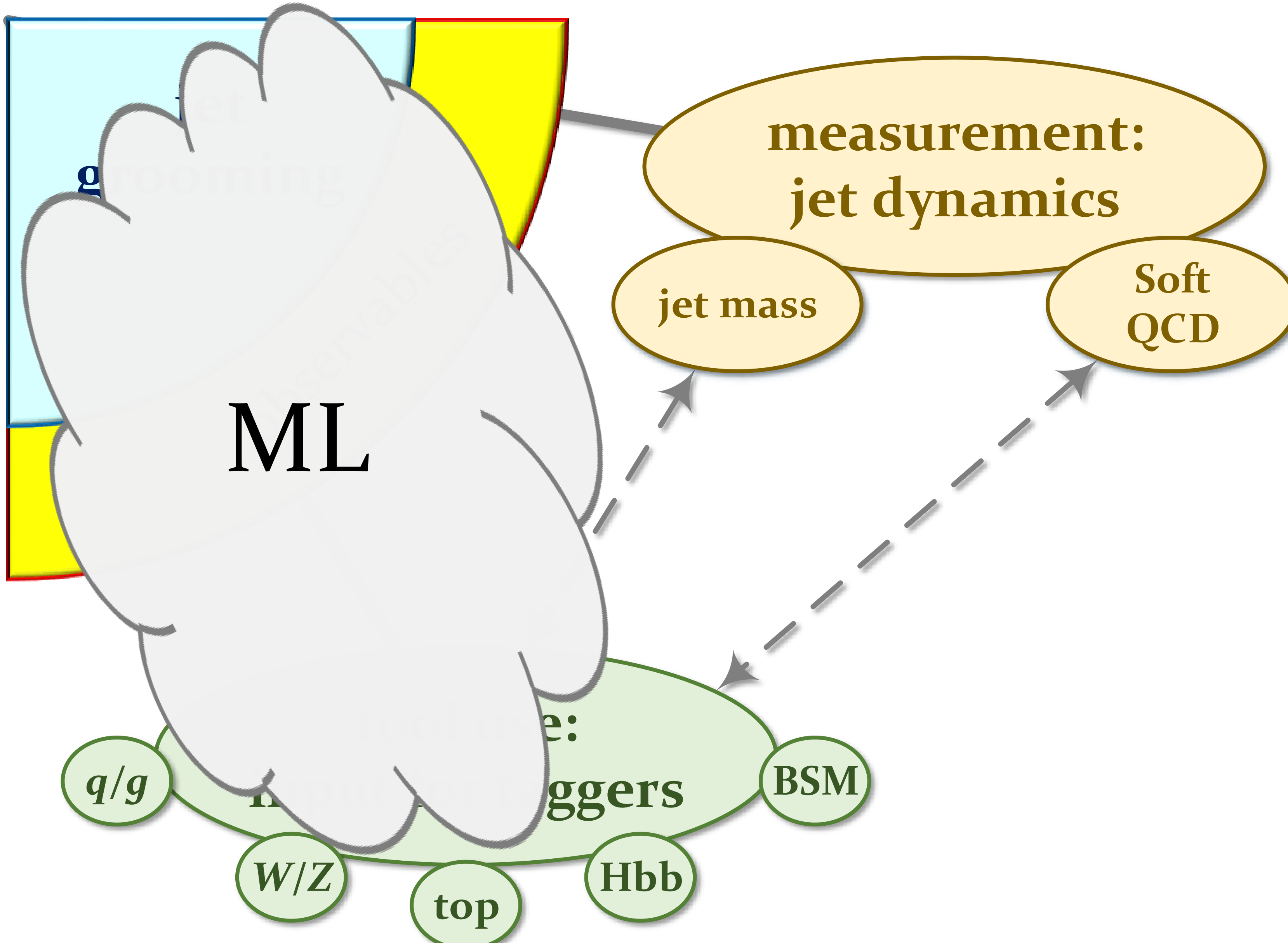


Jet Substructure in Physics Analysis



Jet Substructure in Physics Analysis

Peter Loch
UAPhysics
The University of Arizona®
College of Science



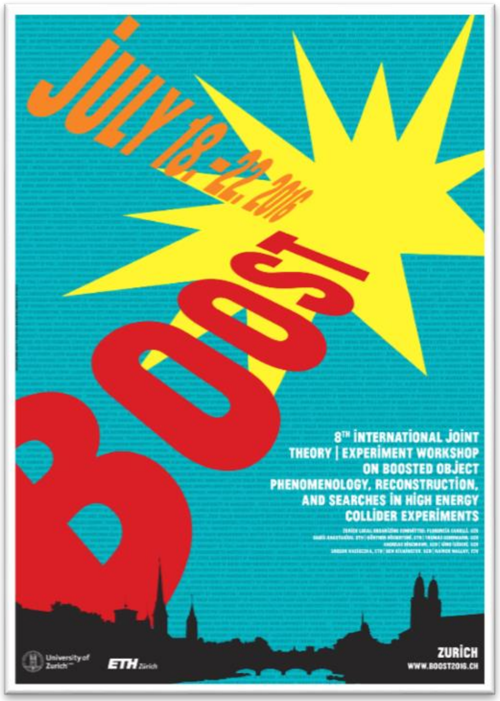
ML in Jet Substructure Analysis

Peter Loch
UAPhysics
THE UNIVERSITY OF ARIZONA
College of Science

Already widely explored – mostly by theorists

Can clearly help to characterize jets beyond simple cuts – better exploration of residual uncorrelated features in relevant observables leads to improved measurements of structure

BOOST	#talks	#ML talks
2016	48	3 (6.3%)
2017	46	3 (6.5%)
2018	46	7 (15.2%)



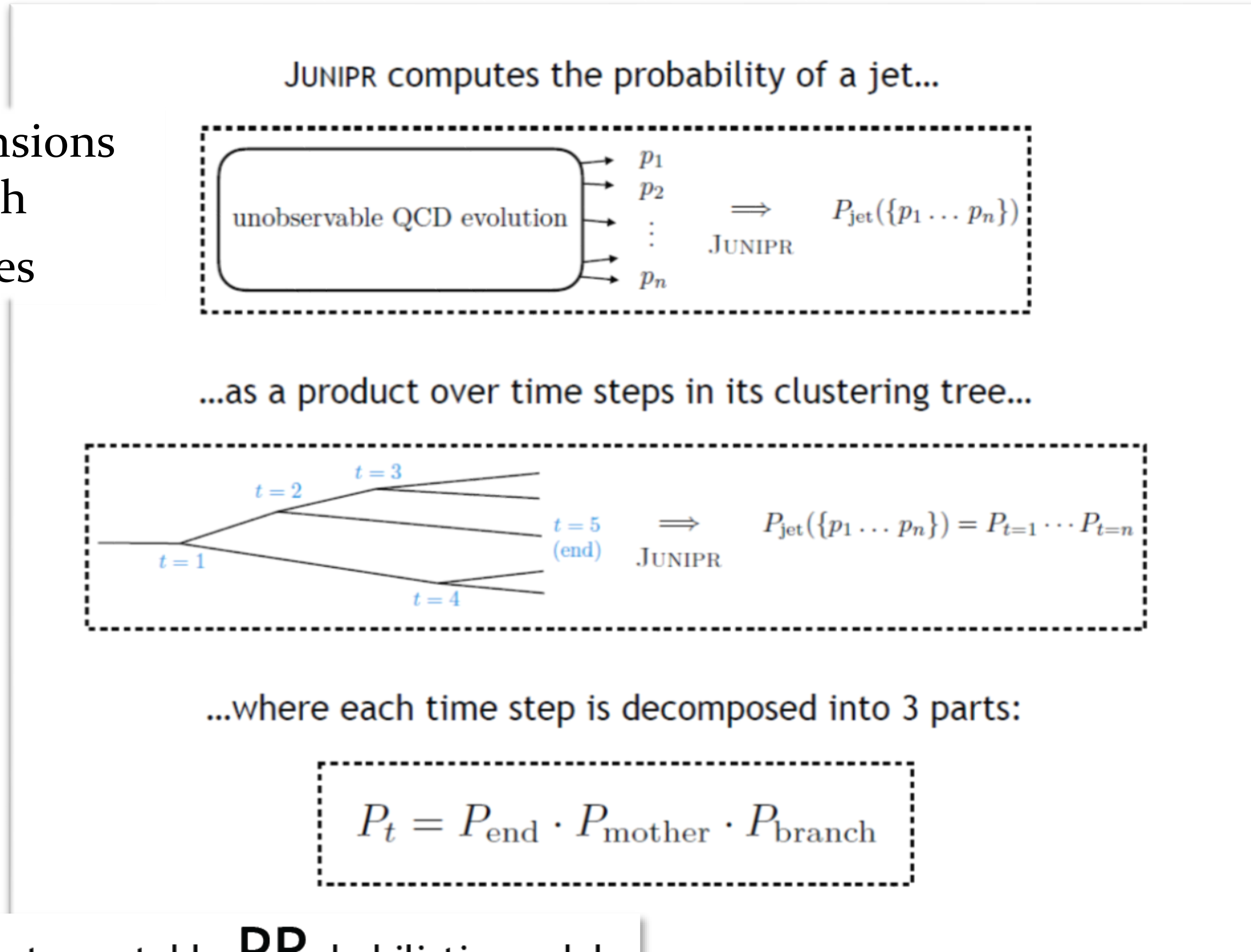
https://indico.cern.ch/event/649482/

Example: JUNIPR

ML analysis of jet branching by investing QCD

$\mathcal{O}(100)$ dimensions for jet with 30 particles

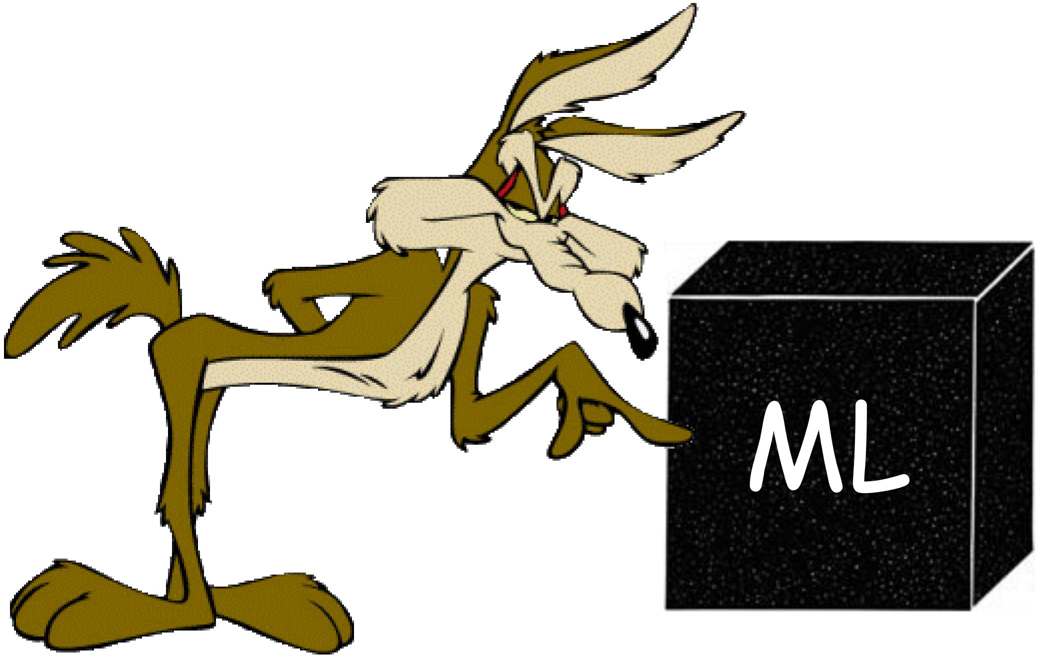
Andreassen, Feige, Fry, Schwartz
[arXiv:1804:09720]



Jets using UNsupervised Interpretable PRobabilistic models

https://indico.cern.ch/event/649482/contributions/3009032/attachments/1689533/2718078/frye-boost-2018.pdf

Experimental Issues (?)



Systematic uncertainties

- Pattern recognition – tagging efficiencies
  - Characterization of individual jets, full event, …
  - Bottoms-up approach starts with uncertainties for input observables – possibly non-trivial (unknown?) propagation through ML kernel?
  - Non-trivial due to correlations through common nuisances – not always possible to determine
  - Often can only test assumptions no (0%) or full (100%) correlations
  - Impractical for large number of inputs – just heard 650 (!) at BOOST2018
  - Hidden biases – how to detect/avoid?
- Signal extraction & measurements
  - Pile-up often overlapping signal features in simple phase spaces – suppression using ML? Jet-by-jet? Full event?
  - Precision measurement of e.g. jet features – how to test/avoid observation biases, control of systematics, …

No fast track here

- But clearly promising first impressions
  - Most obvious application for pattern recognition/image-based analysis

ML in Signal Reconstruction

Calorimeter signal extraction & calibration

Topological cluster formation

- Subject to large fluctuations in number of clusters, cluster shapes, and cluster response in ATLAS calorimeter for hadrons of the same energy – mostly due to fluctuations in shower development (process) and (weak) shape correlations
- More obvious for electromagnetic response – fairly regular and well correlated shapes with small process fluctuations

Cluster calibration

- Can help to reduce bin effects in lookup tables – presently addressed by interpolations (limited to 2-dim)
- Systematic uncertainties can probably be determined similar to present strategies comparing response expectations with calibrated signals

Similar issues with calorimeter jets …

Computing Resources

ML may have to run on large test samples

- Scanning large data samples for signal (and background) features
  - Extract events with a given likelihood to contain e.g. boosted decays, certain jet pattern, etc.
  - May include signal ambiguity resolution – same detector signals contributing to different physics objects in the same event
- Presently using derived data for event characterization
  - Minimal information content to include signal and background in a given phase space
  - Often more than one "view" of the same event in the same data set – derived data production may apply ML for object and event classification and/or filtering