European Research Council
Established by the European Commission

Pier Monni (CERN)
[pier.monni@cern.ch]

Precision Higgs Physics, QCD & Parton Showers [3/3]

PSI summer school - Zuoz, Aug 2024

Introduction to Parton Showers & Jets

What is an event generator (a.k.a. Monte Carlo generator)?

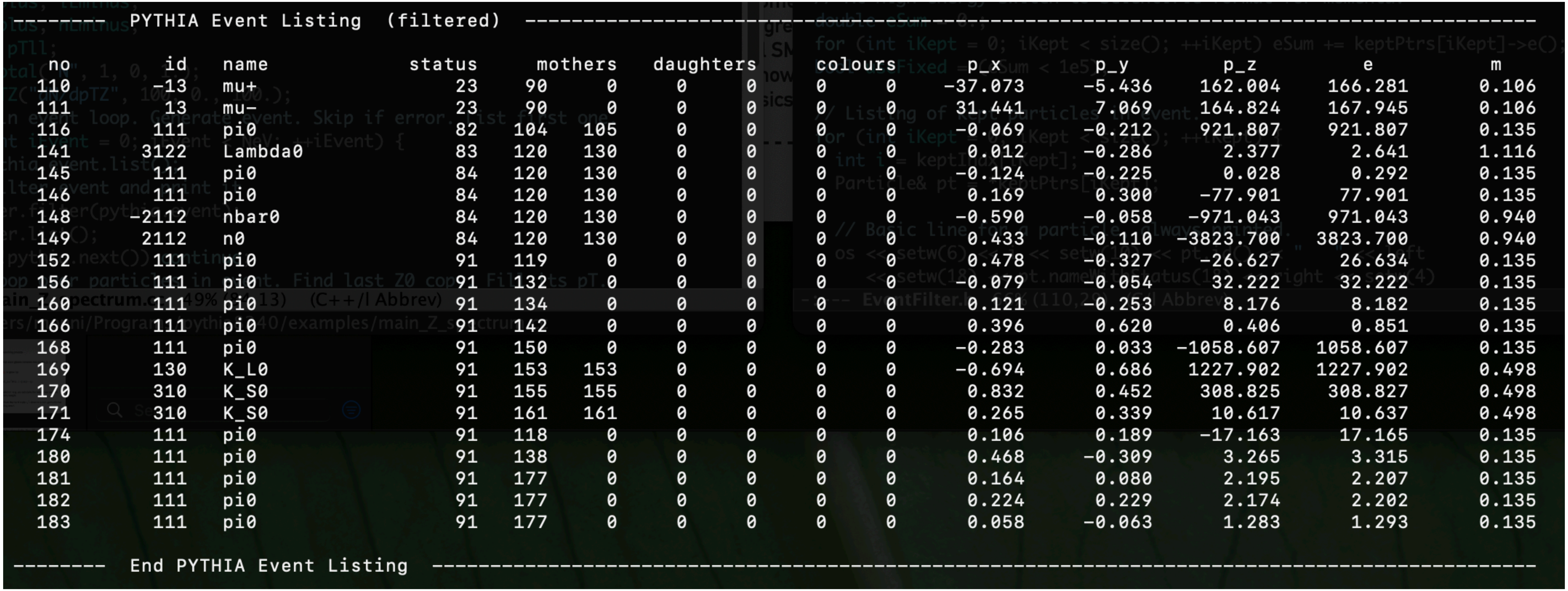
- Event generators are the central simulation tools in modern collider phenomenology
  - Based on Monte Carlo technology, they can simulate realistic collider events (= list of particles) with a physical probability distribution. Generated events can be run through a detector simulation, hence providing a rather accurate (though not perfect!) description of the aftermath of a real-world collision.
  - They generate all stages of the collision:
    - ✓ The hard scattering: the Matrix Element generator
    - ✓ Multi-scale evolution: the Parton Shower stage
    - ✓ Hadronization models & non-perturbative physics
    - ✓ Multiple parton scatterings modelling
    - ✓ Beyond QCD: QED radiation, EW corrections, …

Beware:

Such great versatility comes with limitations. MC generators are often less accurate than state-of-the-art perturbative predictions. Always consult an expert when using them for physics analyses.

An “event” example: e.g. Drell-Yan lepton pair production

- Let’s examine the final state particles produced by Pythia8 (w/ hadronization, no MPI) in a Drell-Yan event



```
 --------  PYTHIA Event Listing  (filtered)  ------------------------------------------------------------------------------------------------------

    no        id   name            status     mothers   daughters     colours      p_x        p_y        p_z         e          m 
   110       -13   mu+                 23     90     0     0     0     0     0    -37.073     -5.436    162.004    166.281      0.106
   111        13   mu-                 23     90     0     0     0     0     0     31.441      7.069    164.824    167.945      0.106
   116       111   pi0                 82    104   105     0     0     0     0     -0.069     -0.212    921.807    921.807      0.135
   141      3122   Lambda0             83    120   130     0     0     0     0      0.012     -0.286      2.377      2.641      1.116
   145       111   pi0                 84    120   130     0     0     0     0     -0.124     -0.225      0.028      0.292      0.135
   146       111   pi0                 84    120   130     0     0     0     0      0.169      0.300    -77.901     77.901      0.135
   148     -2112   nbar0               84    120   130     0     0     0     0     -0.590     -0.058   -971.043    971.043      0.940
   149      2112   n0                  84    120   130     0     0     0     0      0.433     -0.110  -3823.700   3823.700      0.940
   152       111   pi0                 91    119     0     0     0     0     0      0.478     -0.327    -26.627     26.634      0.135
   156       111   pi0                 91    132     0     0     0     0     0     -0.079     -0.054     32.222     32.222      0.135
   160       111   pi0                 91    134     0     0     0     0     0      0.121     -0.253      8.176      8.182      0.135
   166       111   pi0                 91    142     0     0     0     0     0      0.396      0.620      0.406      0.851      0.135
   168       111   pi0                 91    150     0     0     0     0     0     -0.283      0.033  -1058.607   1058.607      0.135
   169       130   K_L0                91    153   153     0     0     0     0     -0.694      0.686   1227.902   1227.902      0.498
   170       310   K_S0                91    155   155     0     0     0     0      0.832      0.452    308.825    308.827      0.498
   171       310   K_S0                91    161   161     0     0     0     0      0.265      0.339     10.617     10.637      0.498
   174       111   pi0                 91    118     0     0     0     0     0      0.106      0.189    -17.163     17.165      0.135
   180       111   pi0                 91    138     0     0     0     0     0      0.468     -0.309      3.265      3.315      0.135
   181       111   pi0                 91    177     0     0     0     0     0      0.164      0.080      2.195      2.207      0.135
   182       111   pi0                 91    177     0     0     0     0     0      0.224      0.229      2.174      2.202      0.135
   183       111   pi0                 91    177     0     0     0     0     0      0.058     -0.063      1.283      1.293      0.135

 --------  End PYTHIA Event Listing  ------------------------------------------------------------------------------------------------------------
```

Let’s consider a live example…

Why are event generators so important?

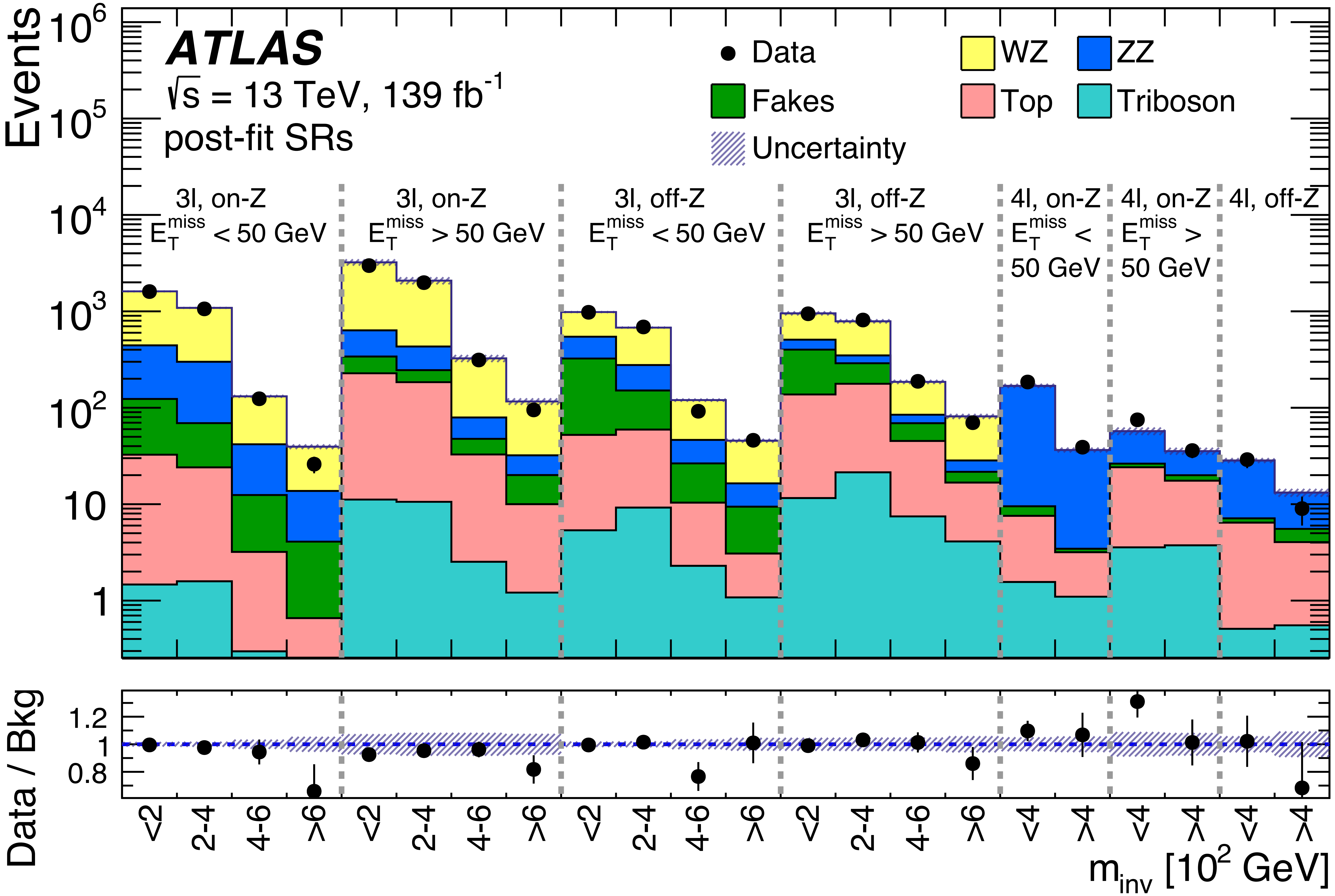
- Their versatility makes them vital for an array of applications: estimation of signal/background events, calibration of measurements (e.g. unfolding, jet energy scale, …), training of machine learning tools (e.g. taggers), experimental extraction of SM parameters (e.g. template fits), and so forth! Effectively used in nearly the totality of LHC analyses

e.g. a typical LHC analysis

Astonishing description of broad classes of collider processes.

e.g. search for new phenomena in multi-lepton final states: agreement with simulation of all SM background processes shows the absence of new physics

ATLAS 2022



Key aspect 1: Description of multi-scale observables

- Through the parton shower (PS) stage, event generators provide a description of the multi-scale evolution that connects the hard scattering to the observation. This is crucial to infer reliably information on the hard scattering (which is what often we want to measure) from the distribution of final-state particles.

→ This stage is also what described by resummations (cf. previous lecture). However, PS are much more flexible tools, albeit so far with lower perturbative accuracy

Resummations

☒ **Low flexibility:** Each calculation is tailored to a given (class of) observable. Mostly perturbative physics (simple NP corrections can be included)

☑ **High accuracy:** Existing technology allows for very accurate theory calculations

Parton Showers

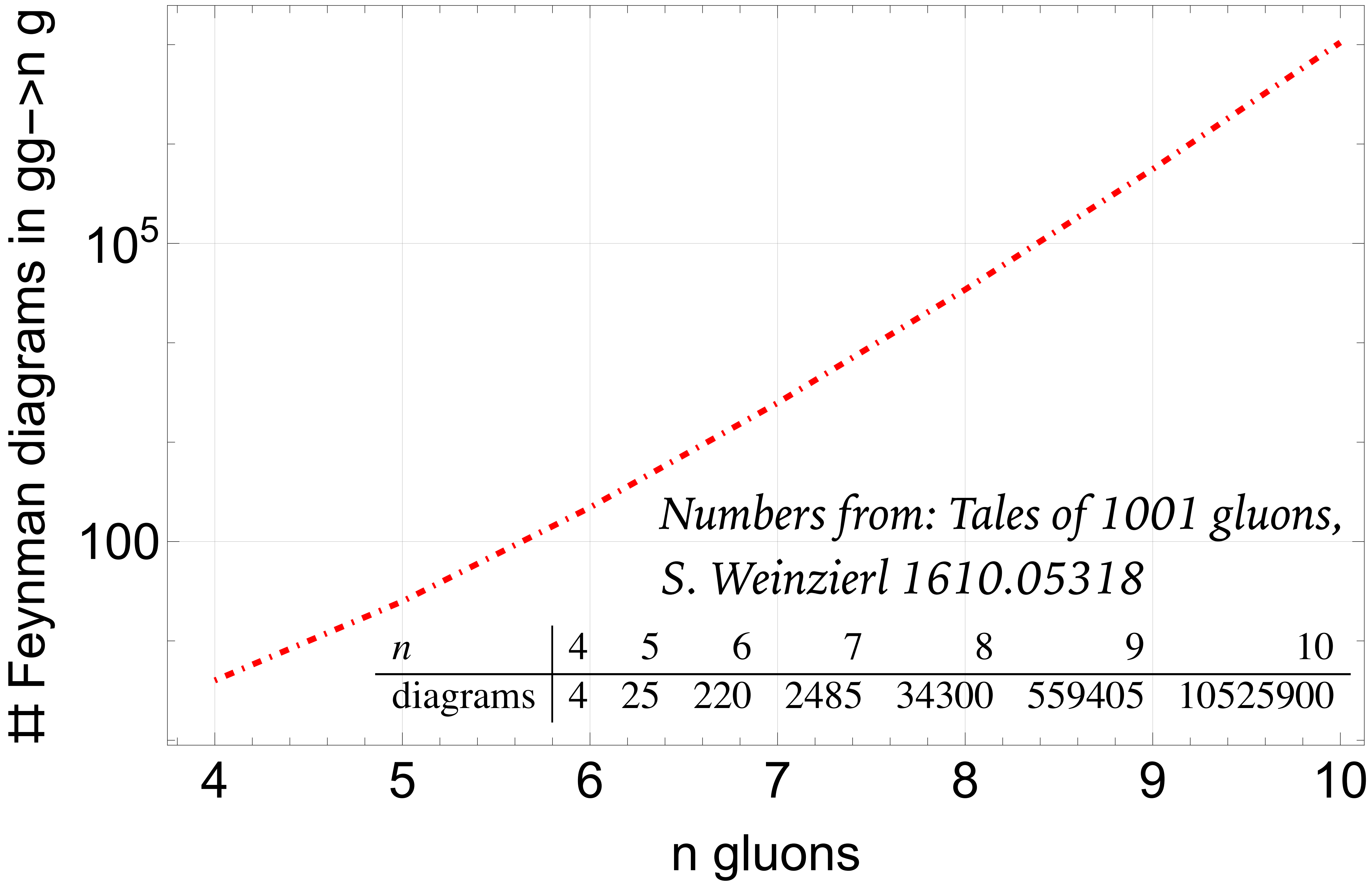
☑ **High flexibility:** Realistic physical events; Simulate any collider reaction/observable (can be measured on each event)

☒ **Low accuracy:** often only LL (~50% error), though ongoing programme to improve their accuracy

Key aspect 2: Generation of many-particle events

- Parton Showers also provide an approximate description of many particle events in regimes with very high particle multiplicity

e.g. Factorial growth of complexity of LO amplitude calculations with multiplicity



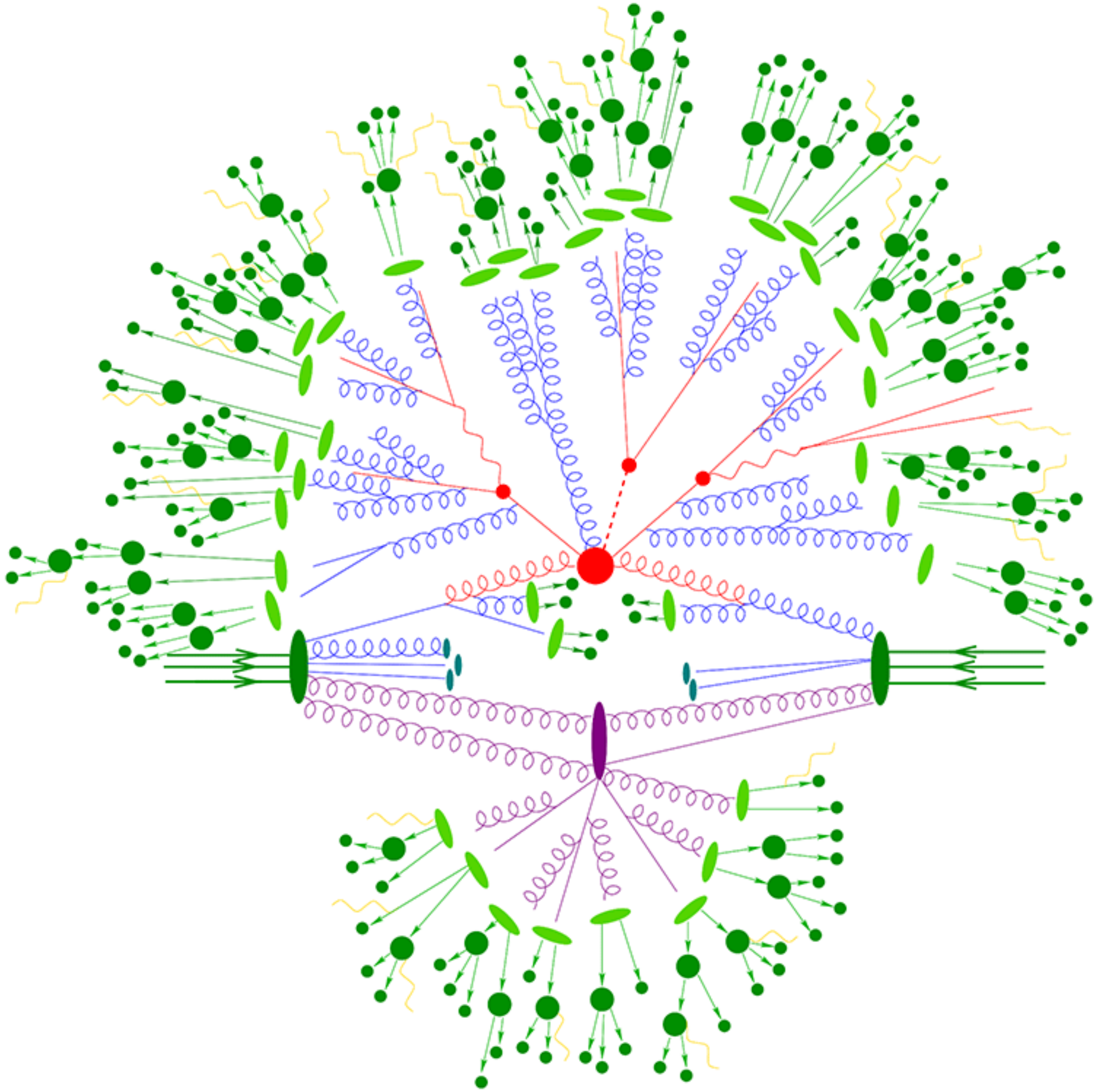
Numbers from: Tales of 1001 gluons, S. Weinzierl 1610.05318

n	4	5	6	7	8	9	10
diagrams	4	25	220	2485	34300	559405	10525900

Even with modern recursive methods (lower computational complexity), one still has to integrate over the n-particle phase space (3n-4 dimensional). This quickly becomes prohibitive even at the LO!

PS provide an efficient way of describing high-multiplicity events and observables.

Key aspect 3: Modelling of non-perturbative physics



Event generators can model rather well the full structure of a collider event, including hadronization dynamics, non-perturbative d.o.f. inside the proton (intrinsic p_T) and multiple scattering in a single pp collision (more later)

Figure Credits: Stefan Höche

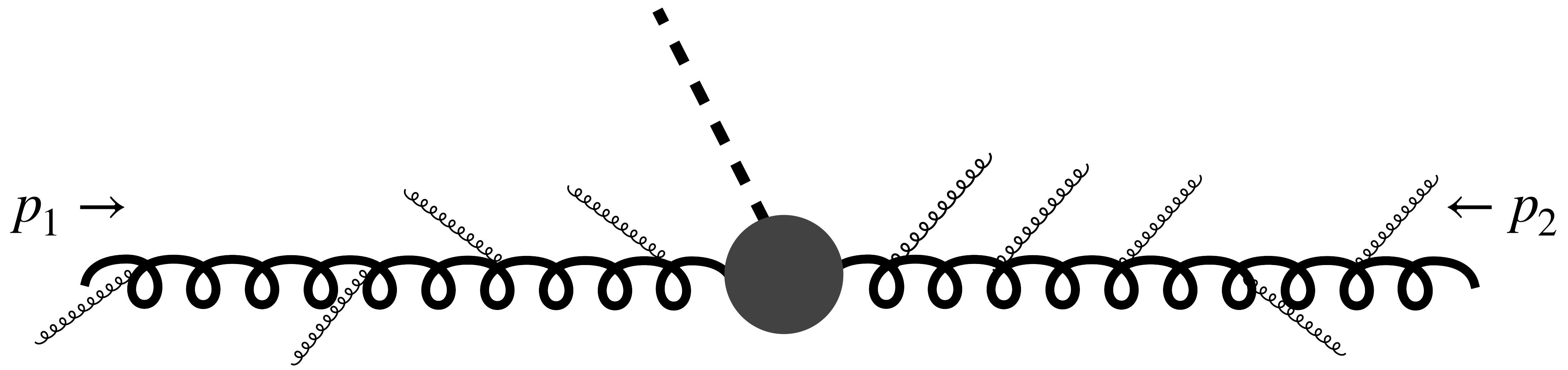
Building a toy PS for $gg \to h$

(Slow) Mathematica code available at this URL

Multi-parton QCD squared amplitudes built recursively

- We can now build a toy parton shower (lowest order, no hadronization/MPI) to simulate the Higgs q_T in gluon fusion. This will allow us to get the gist of this simulation method
  - We work in DL approximation: recall from the previous lecture that we can approximate each emission with the approximate (soft-collinear) matrix element

$$|\mathcal{M}_{\rm sc}(k_i)|^2 = 4\pi\alpha_s C_A \frac{p_1\cdot p_2}{p_1\cdot k_i\, p_2\cdot k_i}$$



  - Our aim is to build the multi-parton final state recursively, starting from a Born configuration $gg \rightarrow h$. This requires rethinking the formulation of the DL resummation in probabilistic terms (very different mathematical language)

The Sudakov form factor

- Our toy parton shower entails three central ingredients:
  - A splitting kernel: our $|\mathscr{M}_{\rm sc}(k_i)|^2$ squared amplitude
  - A kinematic map to absorb the recoil of each emission: we will assign the transverse recoil to the Higgs boson
  - An ordering variable: emissions are radiated in an ordered sequence (causality) as part of the recursive procedure. We choose to order emissions in their transverse momentum w.r.t. the beam, i.e. emissions with a larger k_t are emitted before emissions with a smaller k_t
- The central element of our shower algorithm is the Sudakov form factor encoding the no-emission probability between two resolution scales $Q_1 \geq Q_2$

$$\Delta(Q_1, Q_2) = \exp\left\{ -\int [dk]\,|\mathscr{M}_{\rm sc}(k)|^2\,\Theta(Q_1 \geq k_t \geq Q_2)\right\} \leq 1$$

Emission probability and unitarity

- The Sudakov FF encodes the core information of the branching process
  - The probability of not having any emissions off the initial-state gluons between two scales $Q_1 \geq Q_2$ is given by $\Delta(Q_1, Q_2)$
  - The probability of having an emission at a scale $Q \leq Q_1$ is given by

$$\frac{d\Delta(Q_1, Q)}{dQ} = \Delta(Q_1, Q) \int [dk] \, | \mathscr{M}_{\rm sc}(k) |^2 \Theta(Q_1 \geq Q) \delta(Q - k_t)$$

  - We can use $\Delta(Q_1, Q_2)$ to generate emissions starting from the hard scale $\sqrt{\hat{s}}$ down to a non-perturbative scale Λ (shower cutoff). We then measure our observables (e.g. q_T) on the resulting event.

Emission probability and unitarity

- The Sudakov FF encodes the core information of the branching process

➡ The **probability of not having any emissions** off the initial-state gluons between two scales $Q_1 \geq Q_2$ is given by $\Delta(Q_1, Q_2)$

➡ The **probability of having an emission at a scale $Q \leq Q_1$** is given by

$$\frac{d\Delta(Q_1, Q)}{dQ} = \Delta(Q_1, Q) \int [dk] \, | \mathscr{M}_{\rm sc}(k) |^2 \Theta(Q_1 \geq Q) \delta(Q - k_t)$$

➡ **Evolution**: The probability P_g for the initial state gluons to evolve between $\sqrt{\hat{s}}$ and Λ can be obtained by solving (e.g. with a Monte Carlo algorithm) the recursive equation

$$P_g(\sqrt{\hat{s}}) = \Delta(\sqrt{\hat{s}}, \Lambda) + \int_{\Lambda}^{\sqrt{\hat{s}}} dQ \, \frac{d\Delta(\sqrt{\hat{s}}, Q)}{dQ} P_g(Q)$$

Markov-chain process
(like in LQCD lecture)

Emission probability and unitarity

- The Sudakov FF encodes the core information of the branching process
  - The **probability of not having any emissions** off the initial-state gluons between two scales $Q_1 \geq Q_2$ is given by $\Delta(Q_1, Q_2)$
  - The **probability of having an emission at a scale $Q \leq Q_1$** is given by

$$\frac{d\Delta(Q_1, Q)}{dQ} = \Delta(Q_1, Q) \int [dk] \, |\mathcal{M}_{\rm sc}(k)|^2 \, \Theta(Q_1 \geq Q) \delta(Q - k_t)$$

  - **Unitarity**: If we integrate inclusively over the radiation phase space (e.g. the total cross section), real and virtual corrections exactly cancel giving a unitary weight, i.e. if we don't constrain radiation then $P_g = 1$

$$\Delta(Q_1, \Lambda) + \int_{\Lambda}^{Q_1} dQ \, \frac{d\Delta(Q_1, Q)}{dQ} = \Delta(Q_1, \Lambda) + \Delta(Q_1, Q_1) - \Delta(Q_1, \Lambda) = 1$$

A 2D Monte-Carlo algorithm to simulate the DL Higgs q_T

- Start with a configuration $g(p_1)g(p_2) \to h(p_1+p_2)$ with weight $w = 1$. The initial scale is $Q = \sqrt{\hat{s}} = m_h$

1. Generate an emission: determine its transverse momentum k_t by solving

Importance sampling

$$\Delta(Q, k_t) = \exp\left(-8C_A \frac{\alpha_s}{2\pi}\int_{k_t}^{Q}\frac{dq}{q}\ln\frac{m_h}{q}\right) = x\,(\text{random number}) \in [0,1]$$

2. If $k_t \leq \Lambda$ stop the shower, else generate the emission's azimuthal angle uniformly in $\phi \in [-\pi, \pi]$

3. Assign the transverse recoil to the Higgs by imposing momentum conservation $\vec{q}_T = \sum_i \vec{k}_{t,i}$

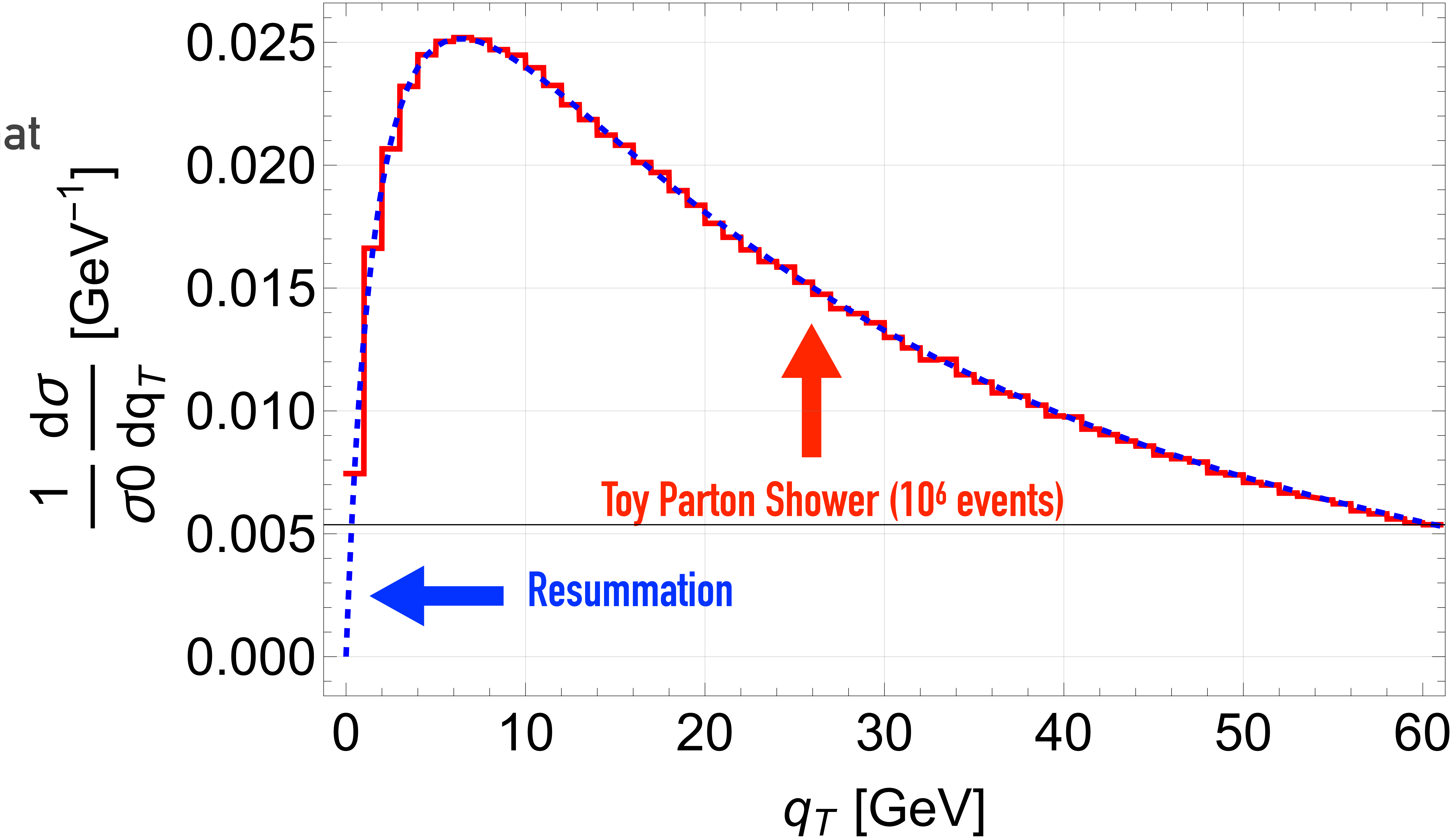
4. Set $Q = k_t$ and go to step 1.

- When the shower stops, measure the observable. Fill histograms with the weight w/N_{events}, and repeat the above algorithm N_{events} times. The resulting distribution is $\frac{1}{\sigma_0}\frac{d\sigma}{dq_T}$

Simulating the Higgs q_T & comparison to DL resummation

A real-life shower would run a hadronization model at Λ, while we will simply stop the shower at the parton level.

- One can demonstrate analytically that the shower is equivalent to our DL resummation (try it!)
- Modern parton showers are much more sophisticated than this toy example, and their connecting to resummation is an active area of research



Toy Parton Shower (10^6 events)

Resummation

Realistic QCD parton showers (PS)

- The main elements of the simplest PS are similar to those in our toy example, but their description of a physical event is much more accurate (e.g. secondary gluon branchings, full momentum conservation). Characterisation in terms of:
  - Splitting probability: defines the matrix element and phase space governing the emission of an extra parton (and corresponding virtual corrections). The branching element can be either a parton (like in our toy example → Parton Showers) or a dipole (= pair of colour connected partons → Dipole Showers).
  - Ordering variable: determines how radiation is sequentially ordered and the coverage of the phase space. Variants span from angular ordering (common in the case of Parton Showers) to transverse momentum or (~) virtuality-like ordering (both used for dipole showers).
  - Kinematic map: determines how the recoil due to the emission of a parton is shared among the other particles in the event. Solutions adopt either a local scheme (recoil is taken from the emitting particle(s) only) or a global scheme (recoil is taken from the whole event).

Realistic QCD parton showers (PS)

- The main elements of the simplest PS are similar to those in our toy example, but their description of a physical event is much more accurate (e.g. secondary gluon branchings, full momentum conservation). Characterisation in terms of:

  - **Splitting r** … extra parton (… e in our toy exa … **s**).

  - **Orderi** … e space. … e momen …

  - **Kinematic** … e other particles in the … ng particle(s) only) or a global scheme (recoil is taken from the whole event).

The workhorses at the LHC (used in nearly all experimental analyses):

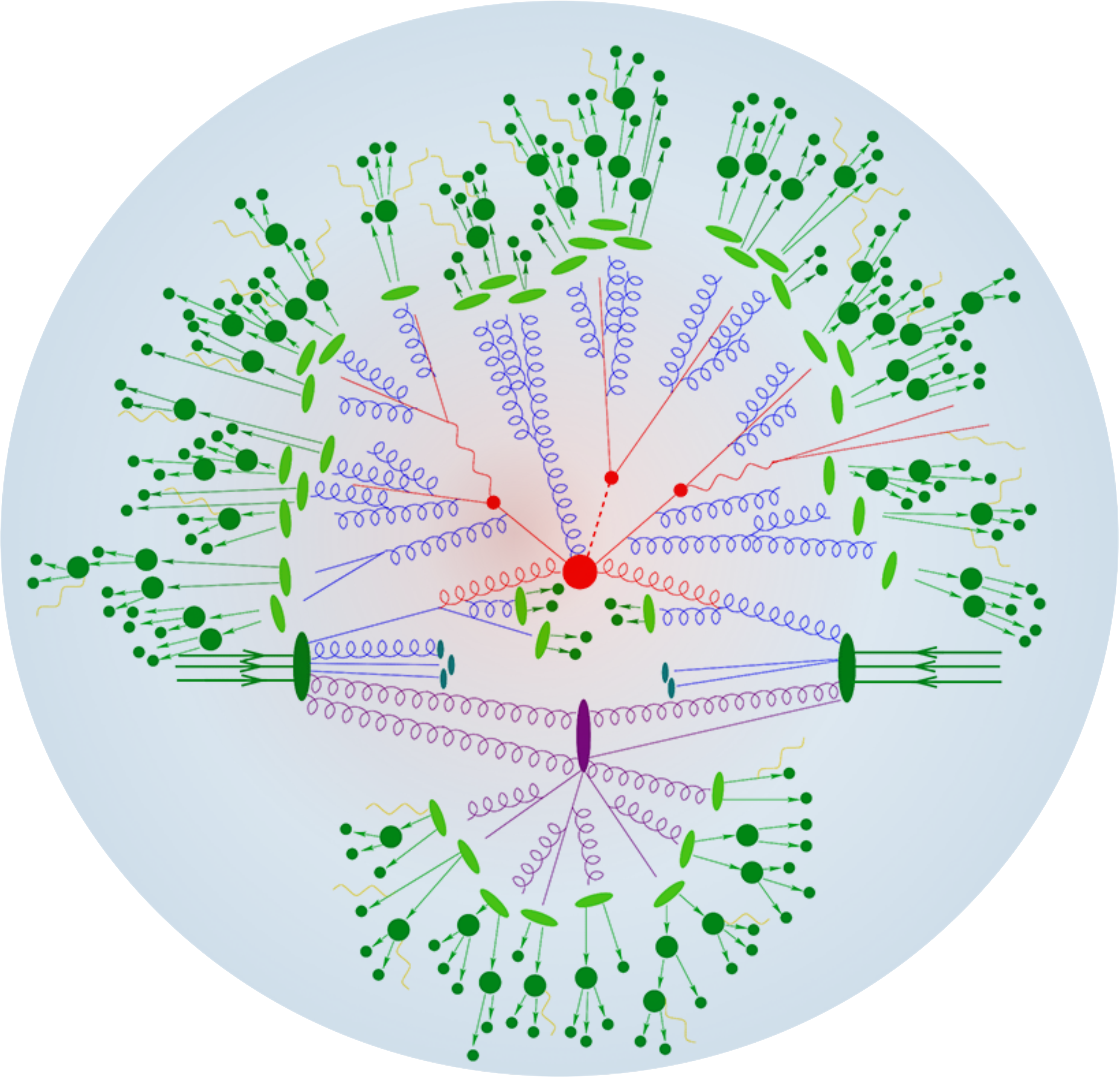
Herwig

Pythia

Sherpa

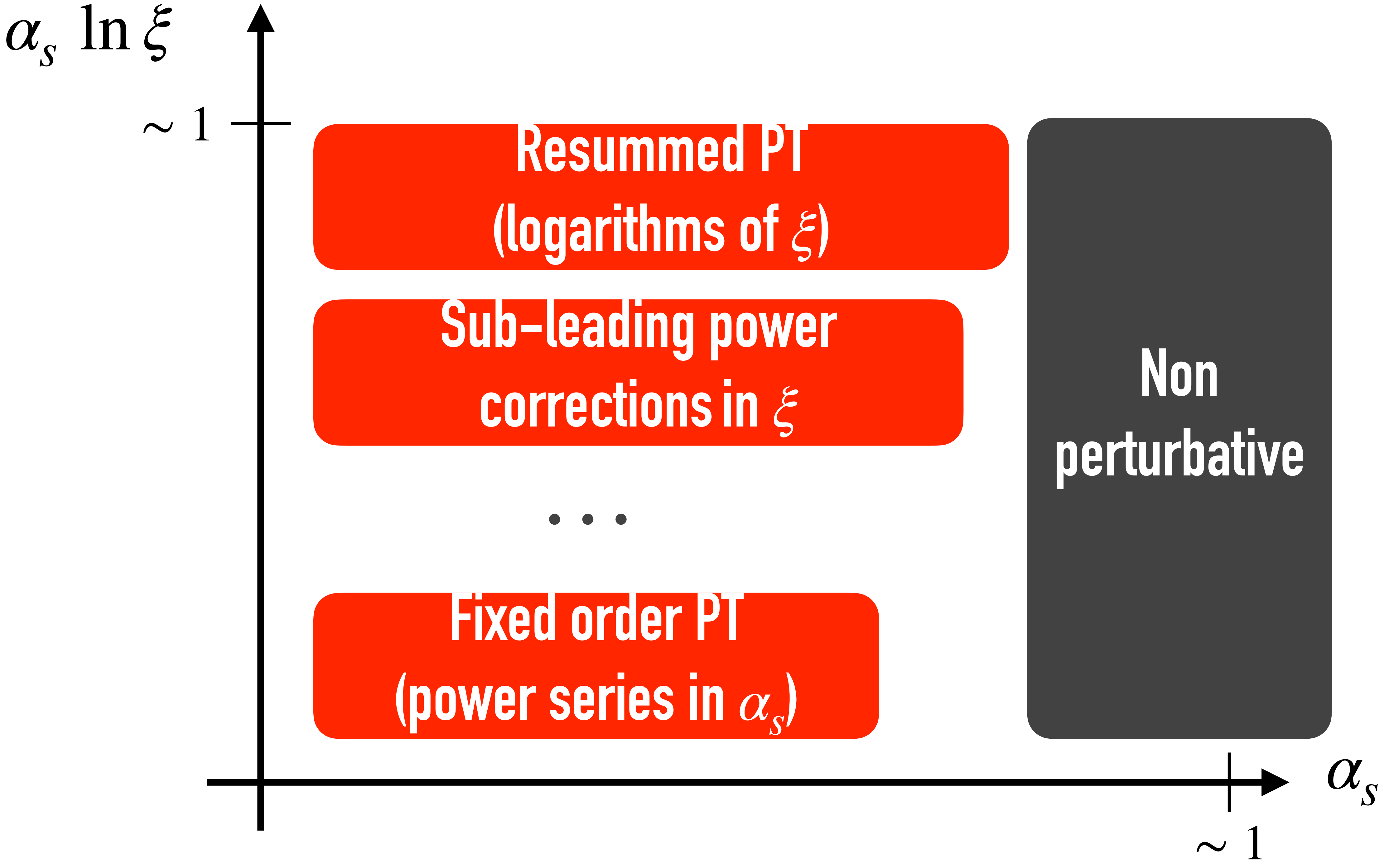
What is the (perturbative) accuracy of an event generator?

- The meaning of the question itself depends on the kinematic regime at which is asked!



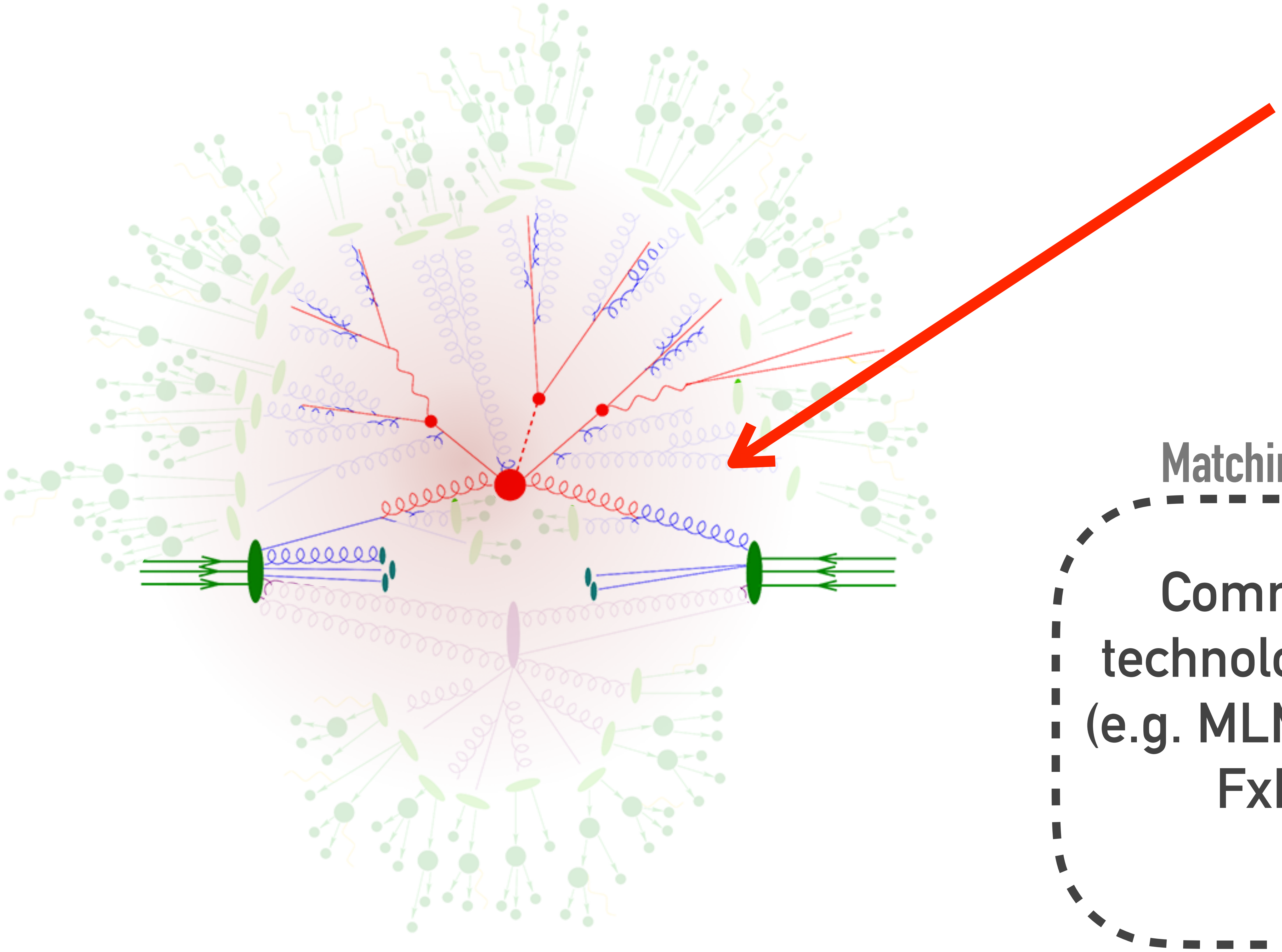
- Figure of merit is the value of coupling and scale ratio(s) ξ (e.g. scale of the measurement / hard scale)

e.g. single scale ratio & coupling constant



What is the (perturbative) accuracy of an event generator?

- The meaning of the question itself depends on the kinematic regime at which is asked!



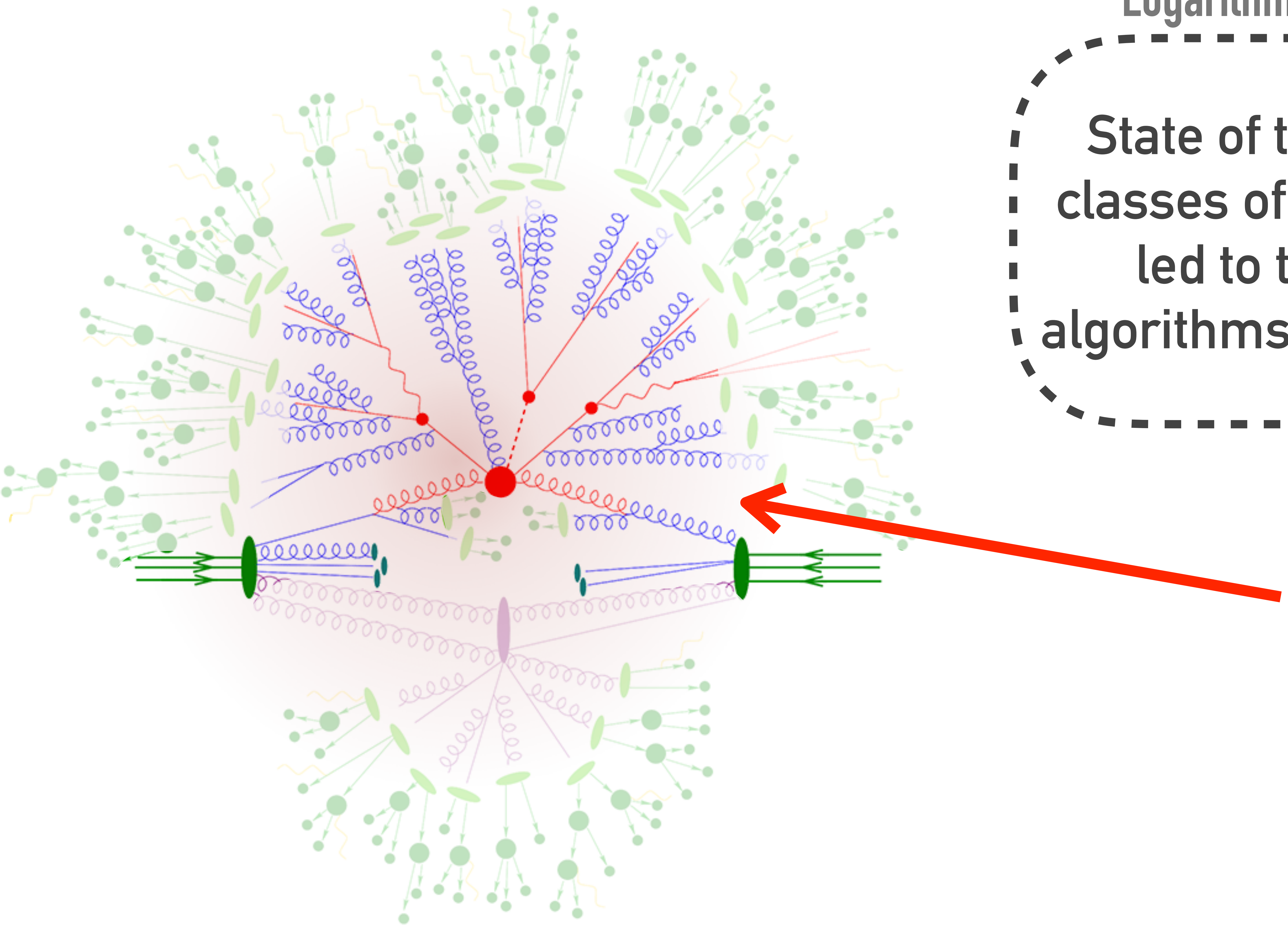
- Short distance (hard)
  - scales probed at LHC: $O(10^2)$-$O(10^3)$ GeV
  - Accuracy in terms of fixed order perturbation theory (LO, NLO, NNLO, …)

Matching & merging

Common PS are only LO for the hard scattering. A lot of technology is now available for their matching/merging to LO (e.g. MLM, CKKW, …), NLO (e.g. POWHEG, MC@NLO, UNLOPS, FxFx, MiNLO,…) or NNLO (e.g. MiNNLO$_{PS}$, Geneva, UN^{2}LOPS,…) QCD calculations.

What is the (perturbative) accuracy of an event generator?

- The meaning of the question itself depends on the kinematic regime at which is asked!



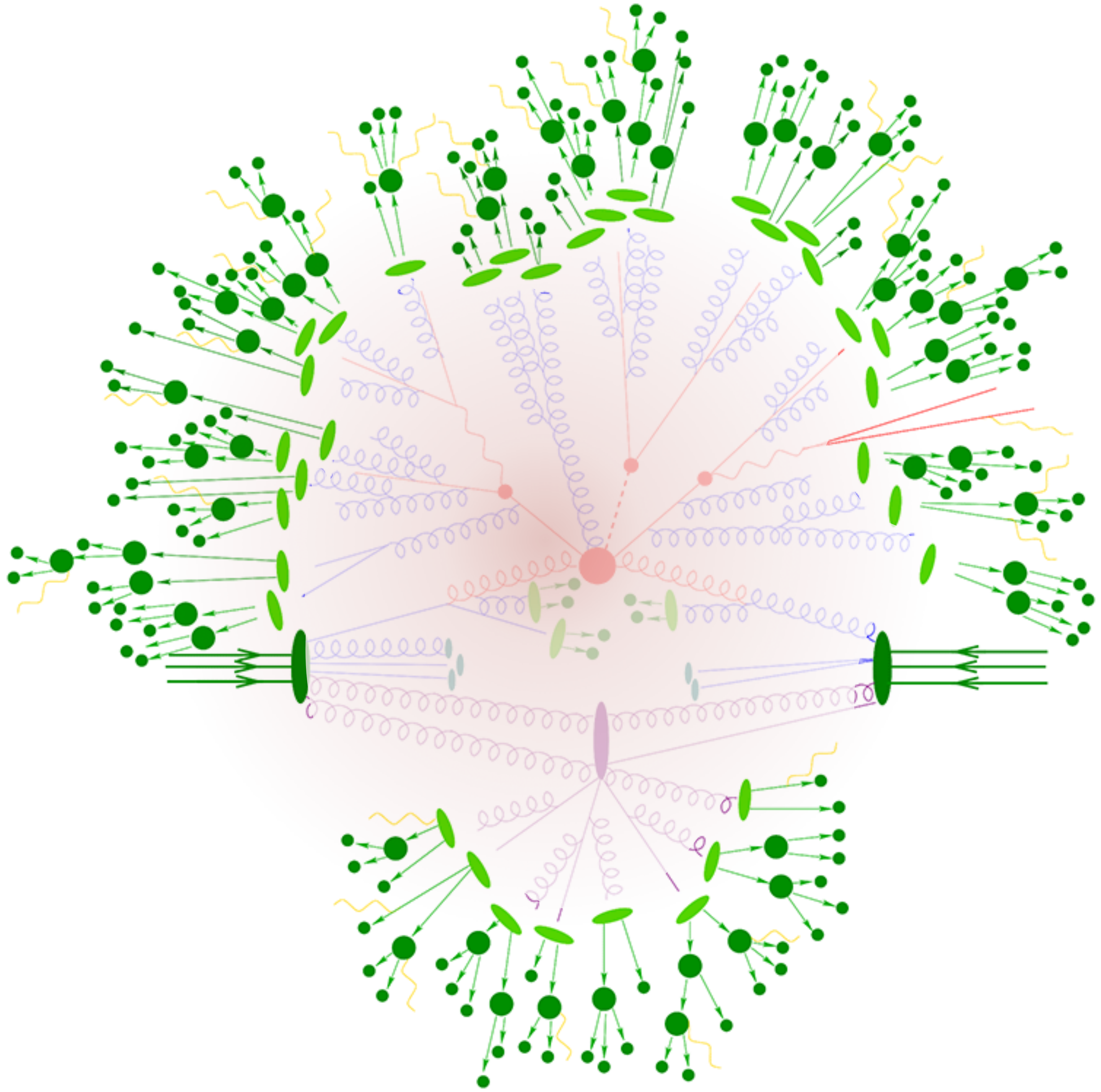
Logarithmic accuracy of Parton Showers

State of the art showers can reach NLL accuracy for large classes of collider observables. Current research in 2024 has led to the formulation of the first NNLL parton shower algorithms (will be available for phenomenology in the future)

evolution (radiation)
towards observable state

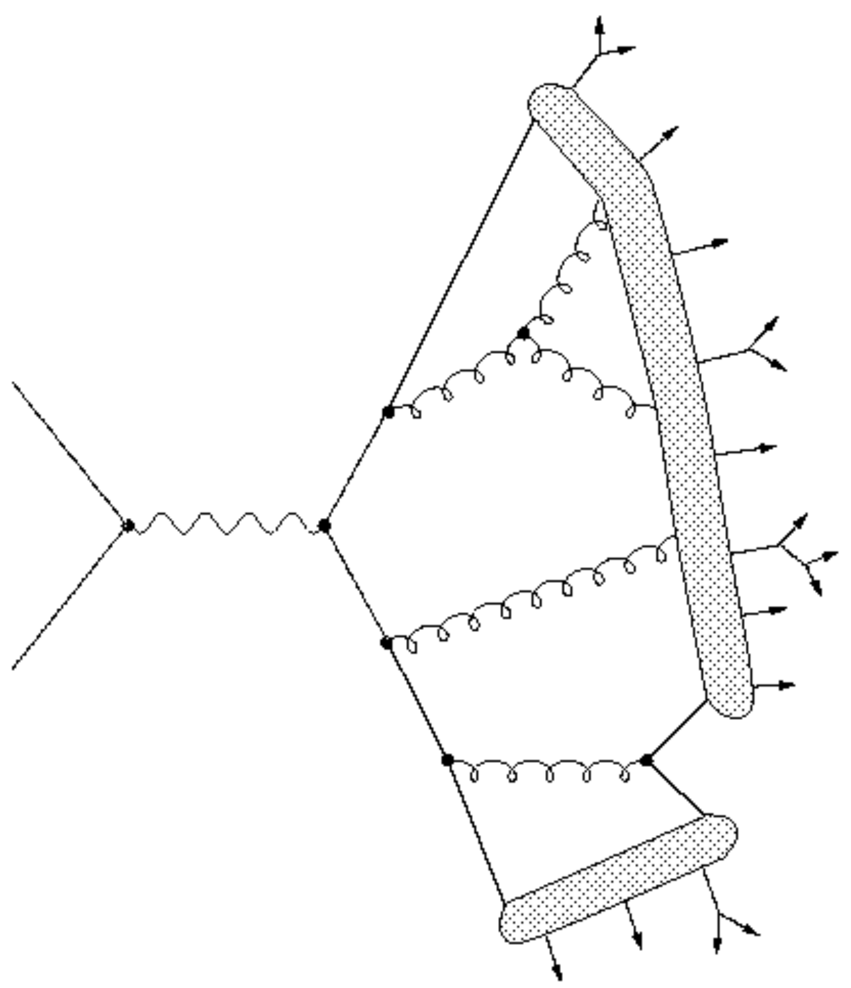
- Long distance (soft and/or collinear)
  - transition from $O(10^2)$-$O(10^3)$ GeV to O(1) GeV
  - Accuracy in terms of resummed perturbation theory (LL, NLL, NNLL, …)

Hadronization models (in event generators)

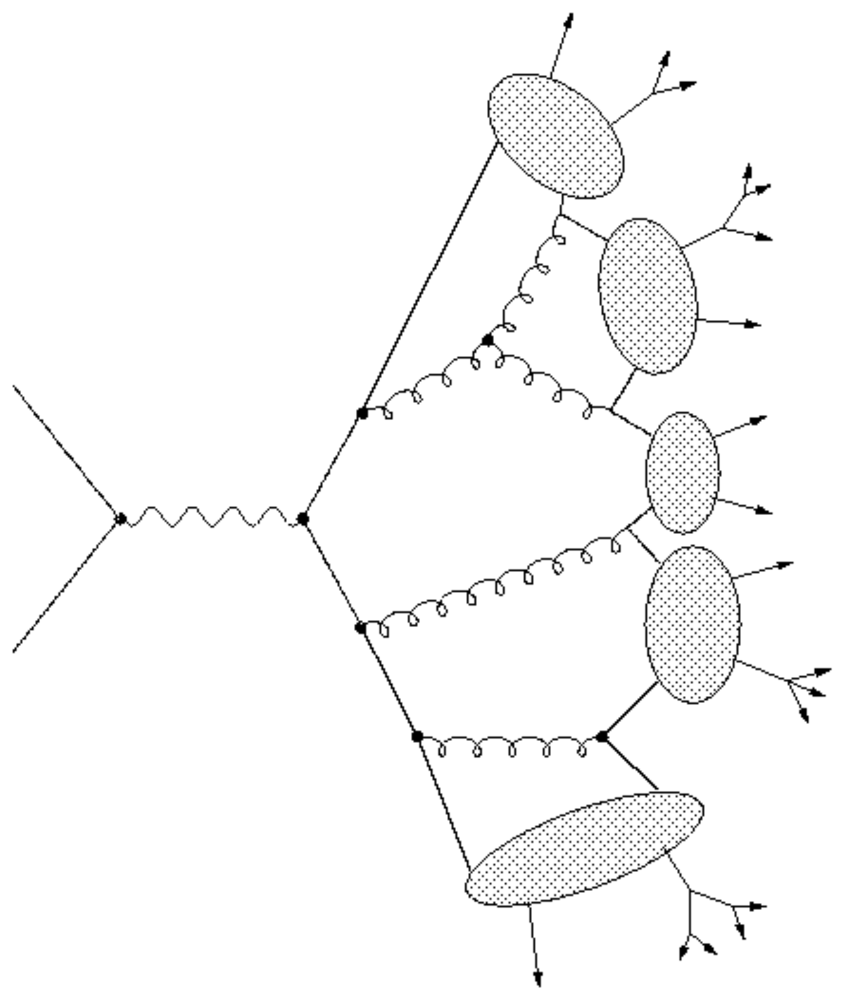


- Non-perturbative transition from a partonic to a hadronic state is described using phenomenological (i.e. not first principles) models
  - ➔ Depend on a number of free parameters which are extracted from experimental data (tuning). Common models:

String model (e.g. in Pythia8)

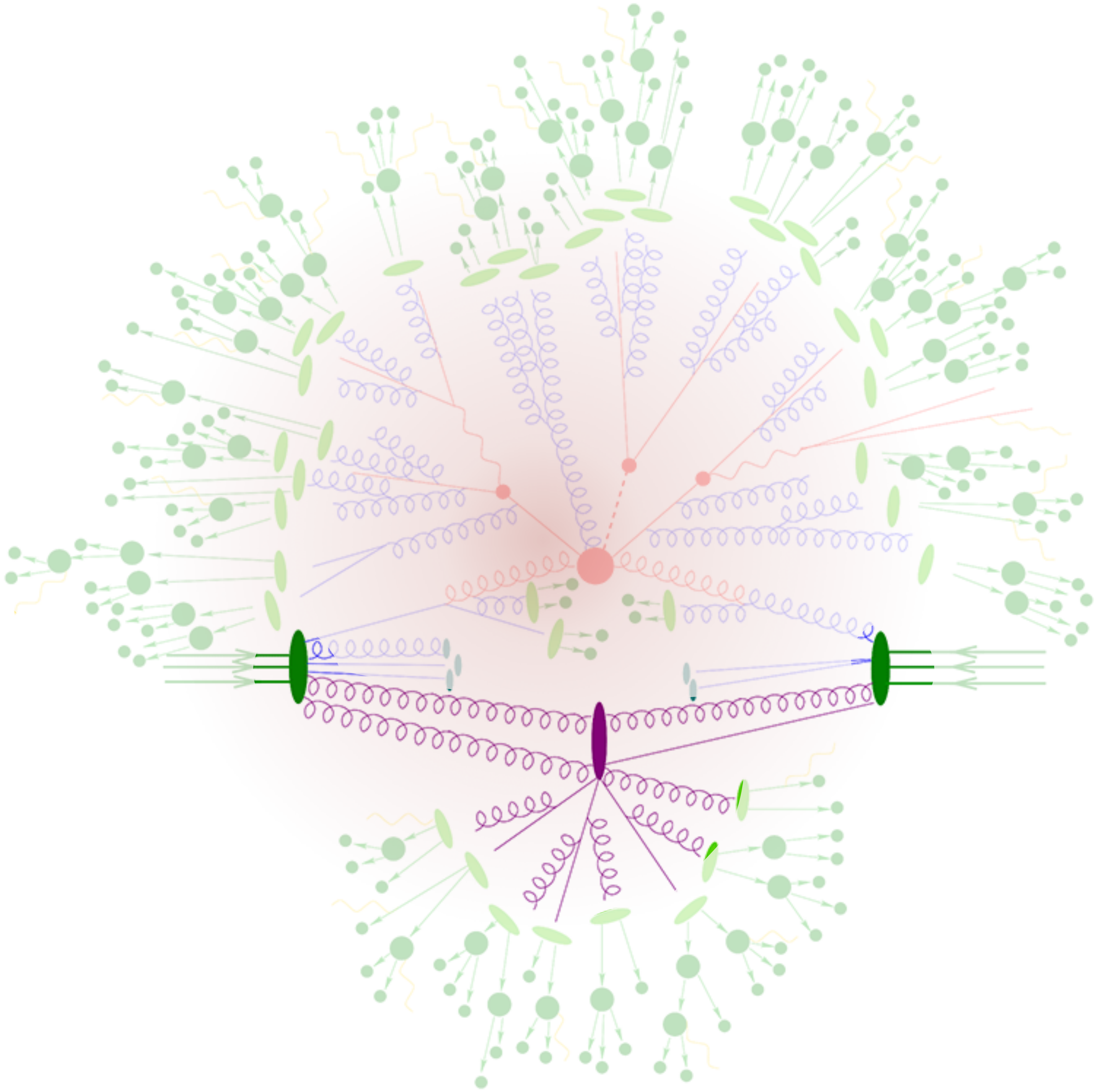


Cluster model (e.g. in Herwig)

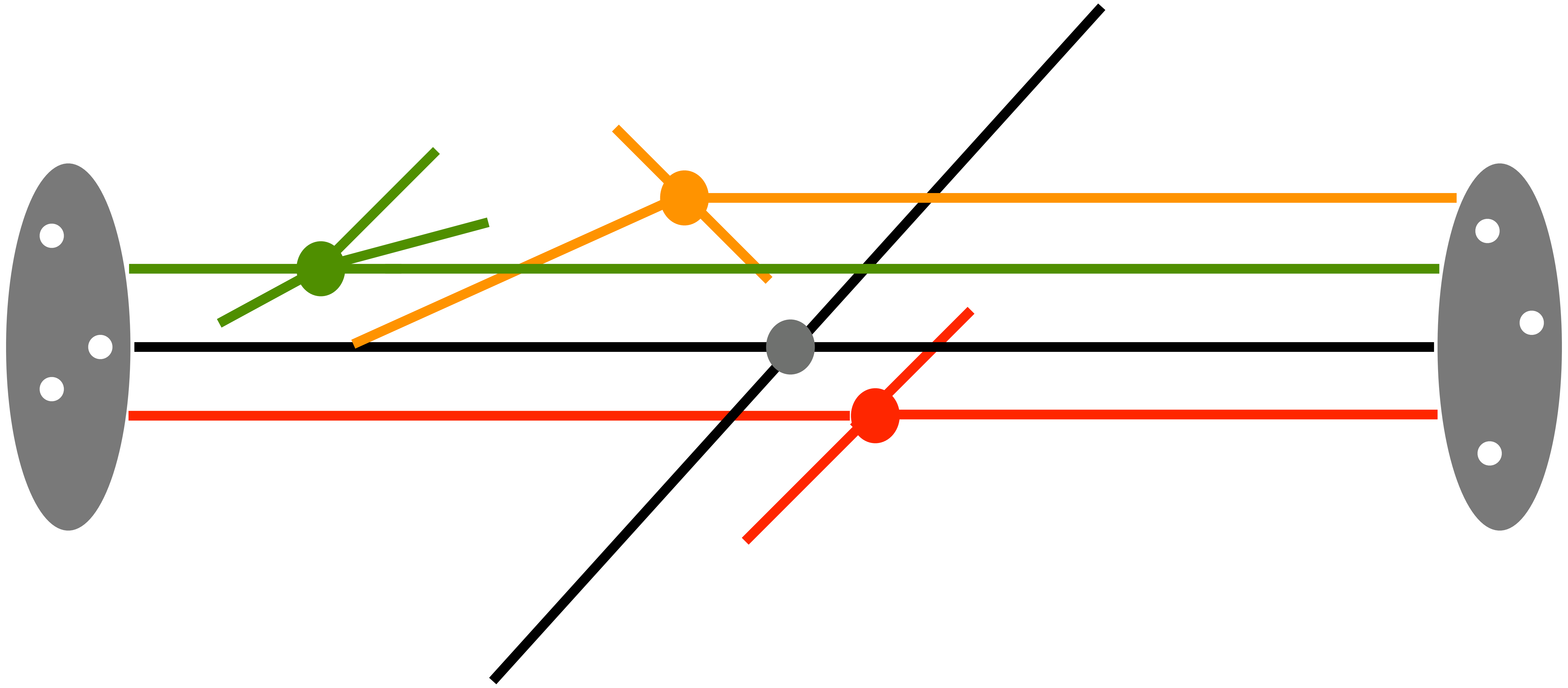


  - ➔ Development of more accurate models (e.g. based on machine learning tech or with better treatment of colour) is an active area of research

Multi-parton interactions (MPI / Underlying Event)



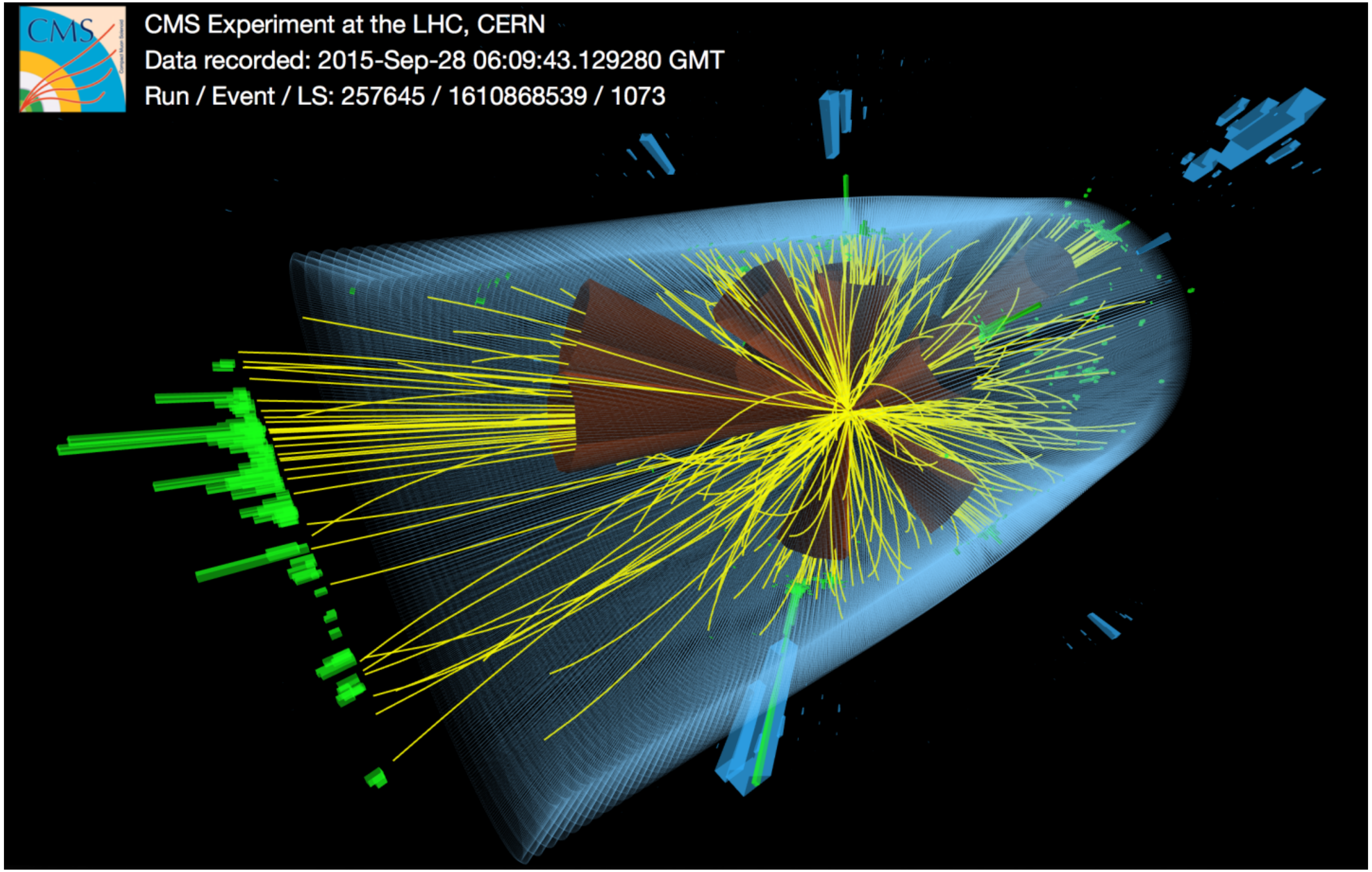
- The main hard scattering (the part described by perturbative calculations) is accompanied by secondary scatterings (~10 at the LHC!) between the partonic content of the two protons
  - Although the momentum transfer is within the perturbative regime, our understanding is limited and we resort to phenomenological models to simulate them
  - Their field theoretical and experimental study is a very active field at the LHC



Hadronic jets in a realistic LHC collision

- Now we have all the elements to describe an LHC event … how can we organise the information encoded in this complexity?

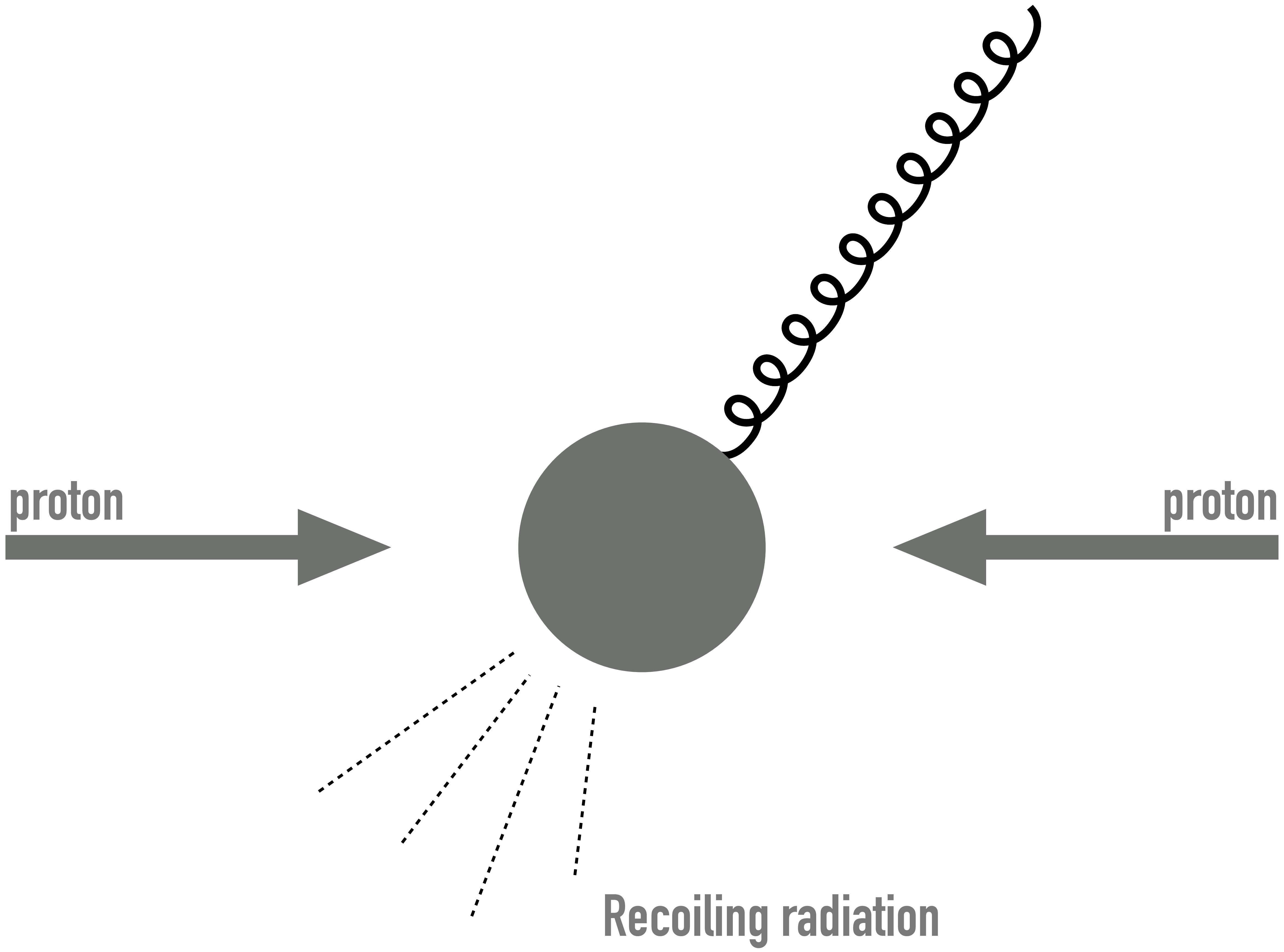
e.g. multi-jet event at the LHC with up to 12 jets (!) with transverse momentum larger than 50 GeV



Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

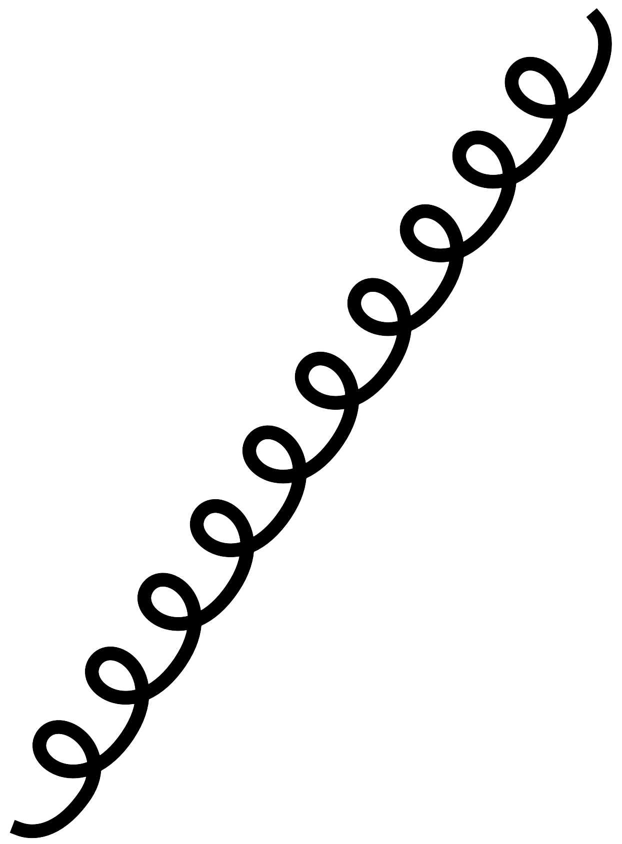
e.g. Fragmentation of a hard gluon



Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

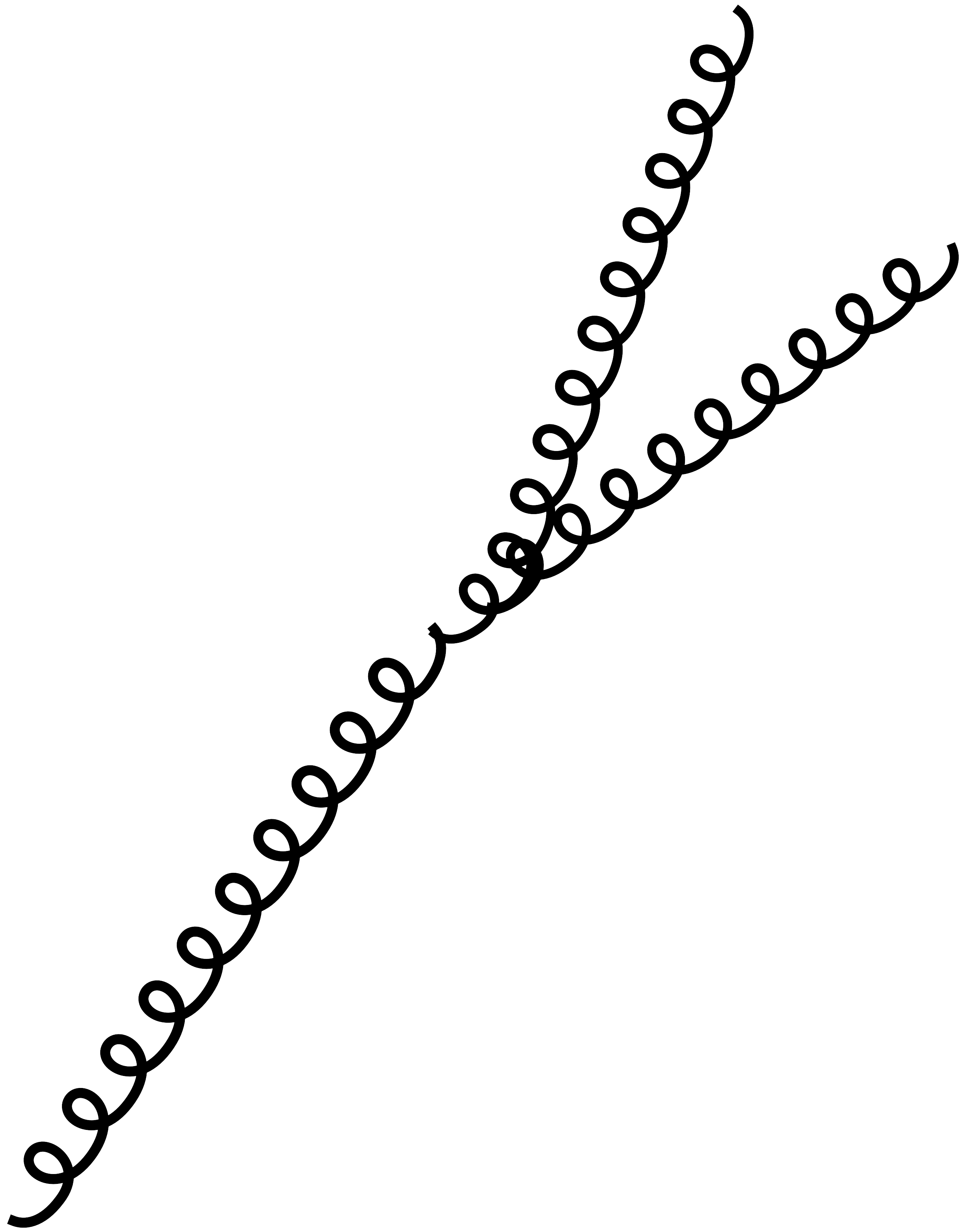
e.g. Fragmentation of a hard gluon



Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

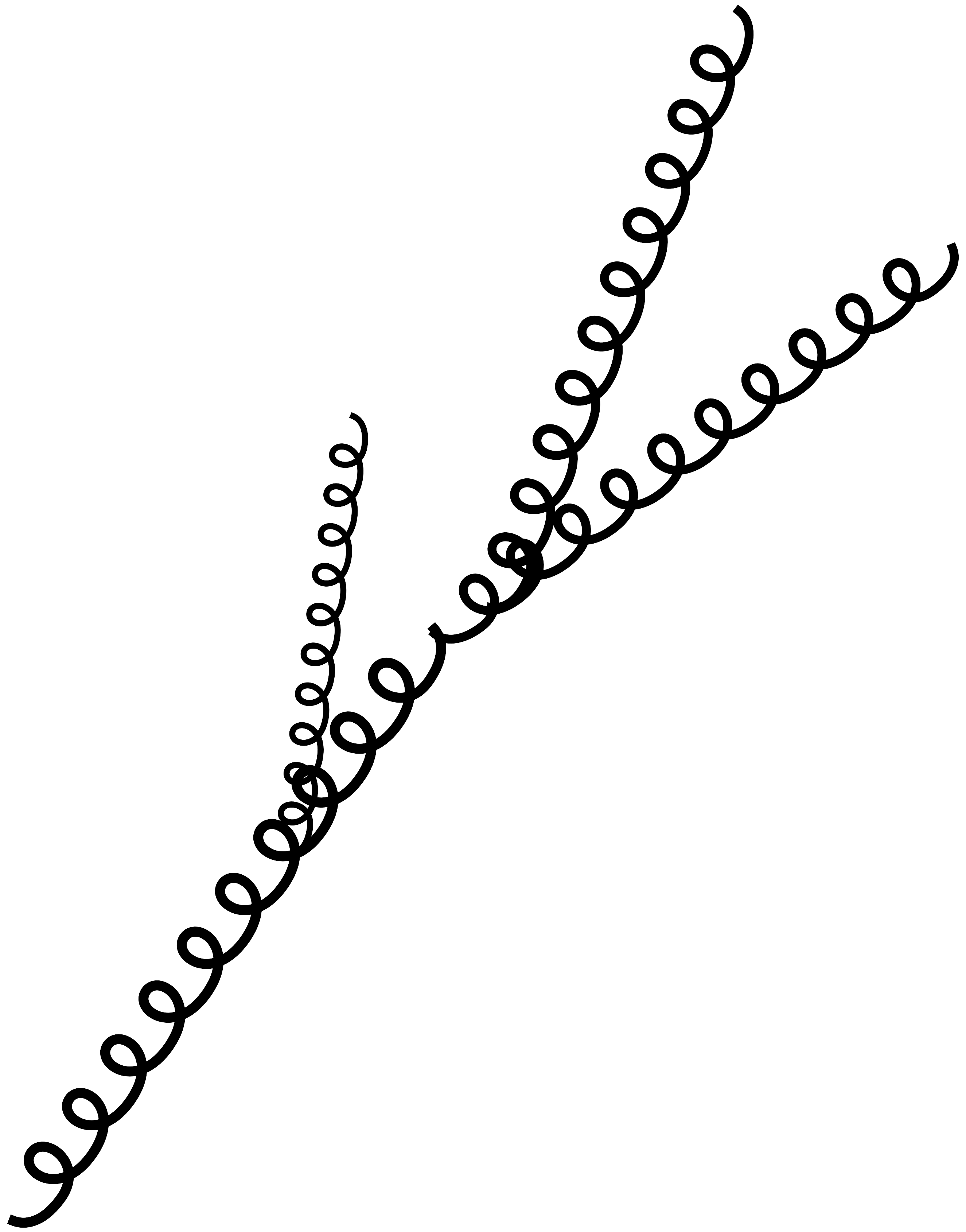
e.g. Fragmentation of a hard gluon



Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

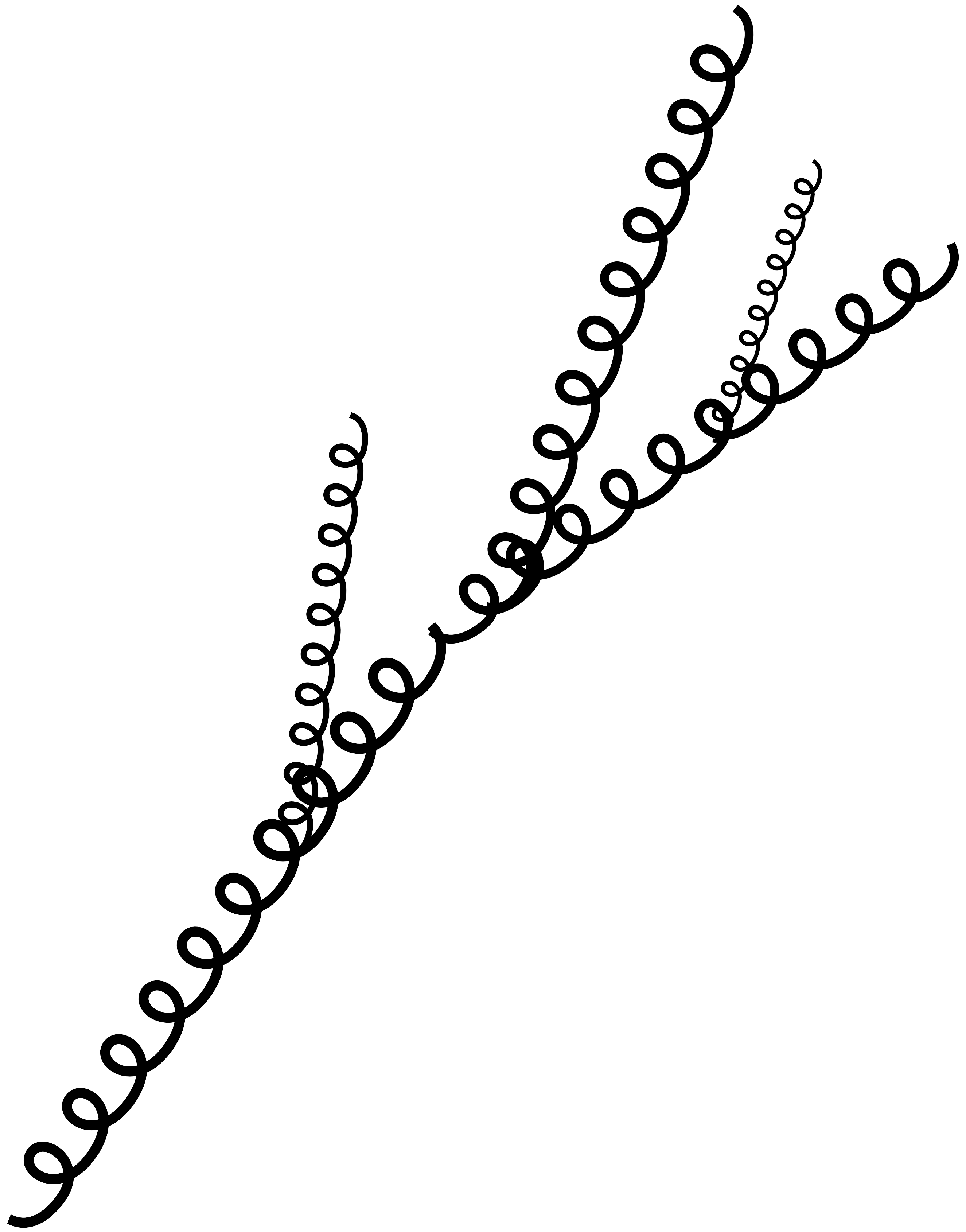
e.g. Fragmentation of a hard gluon



Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

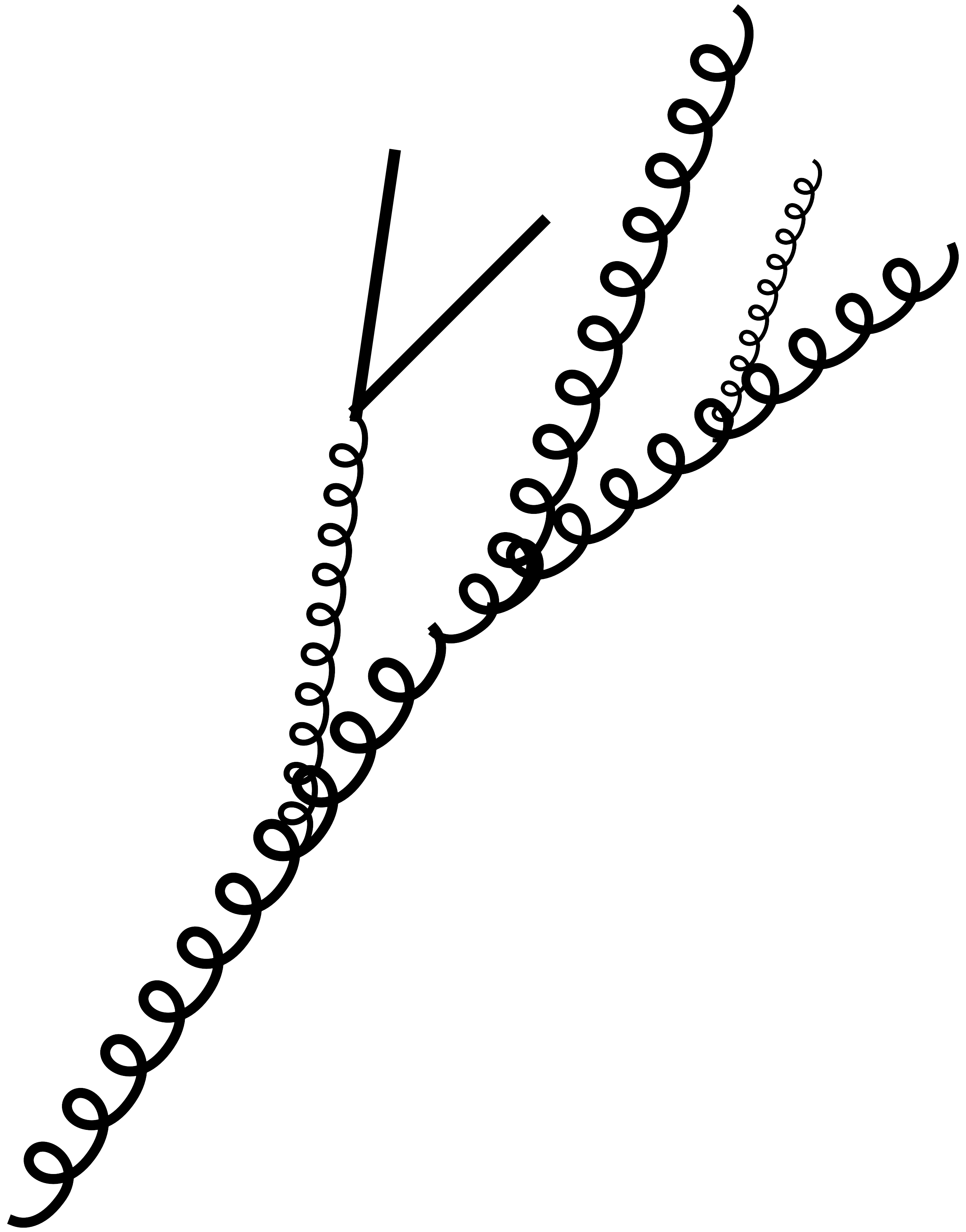
e.g. Fragmentation of a hard gluon



Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

e.g. Fragmentation of a hard gluon

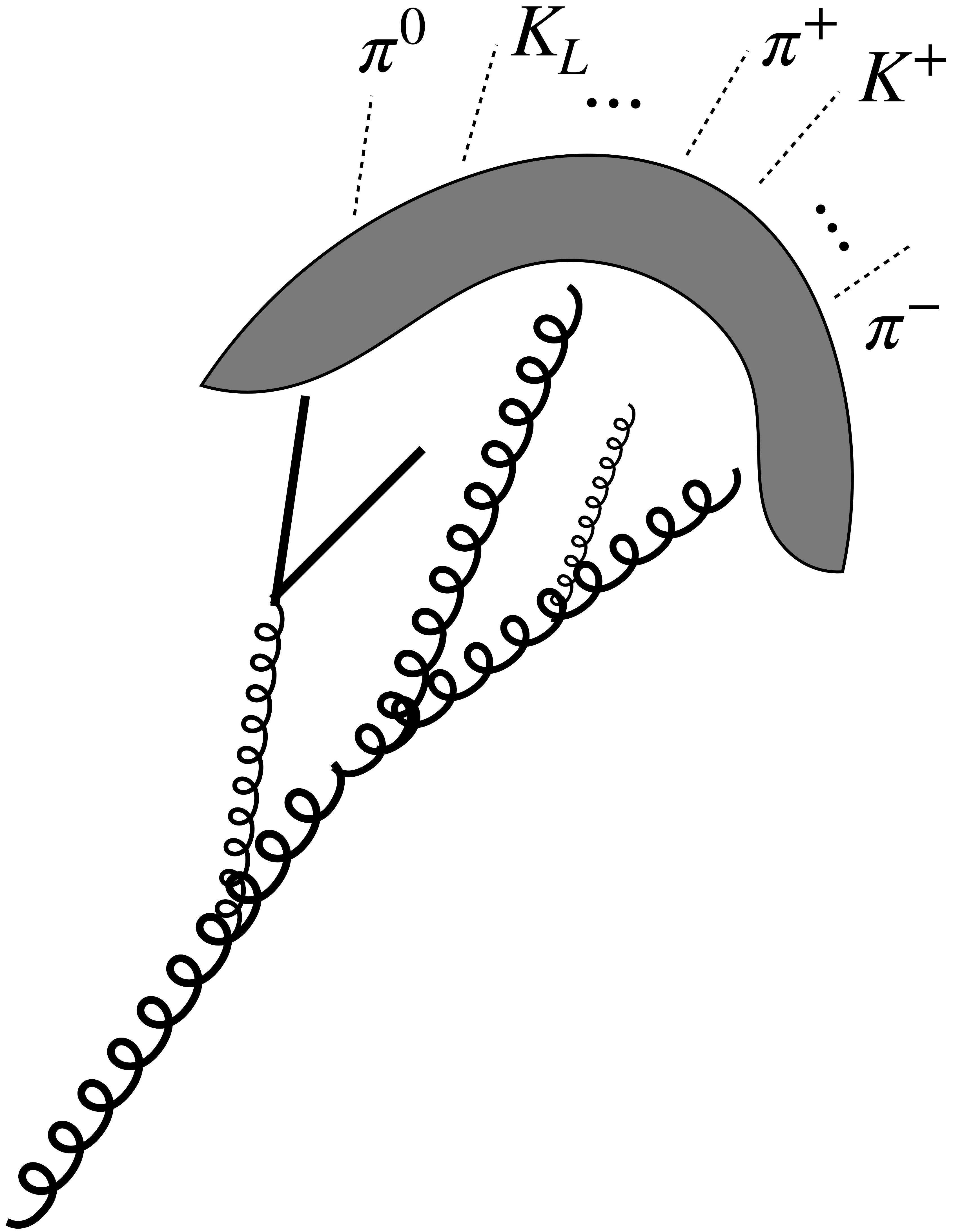


Why do we see jets?

- We have seen that a strongly interacting high-energy system emits large amount of radiation in the soft ($E \to 0$) and in the collinear ($\theta \to 0$) regimes

e.g. Fragmentation of a hard gluon

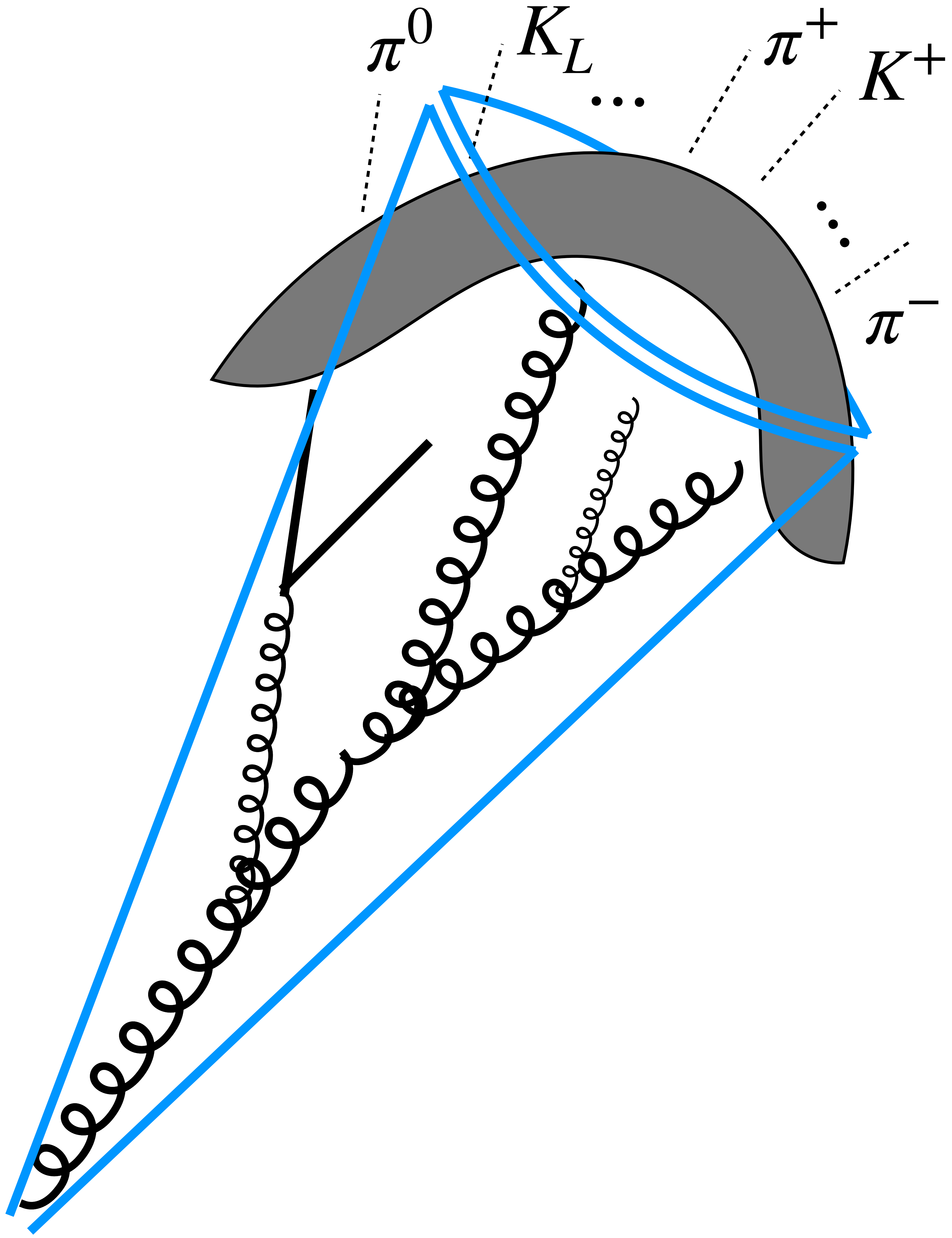
The pattern of QCD particles in the final state is organised (mainly) according to this principle, leading to rather collimated sprays of hadrons, called jets. Additional contamination comes from radiation stemming from MPI & pile-up.



Why do we need jets?

- Why can we rely on perturbative predictions if the final hadrons we observe are intrinsically non-perturbative objects?
  - The perturbative IRC structure of the theory defines most of the kinematic details of the event (in fact, hadrons carry a lot of the quantum properties of the parent parton). Hadronization is a mild kinematic reshuffling of the event (Local Parton-Hadron Duality)
  - Yet, pQCD cannot predict, e.g., the fraction of a certain hadron species or its kinematic properties. We need a sensible method to describe multi-hadron final states in a way that makes sense in perturbation theory

> *"Jets are are legal contracts between theorists and experimentalists"* [Tennenbaum]



IRC safe definition of jets

- A sensible jet (or clustering) algorithm must satisfy our calculability criterion of IRC safety
  - That is, we need a IRC-safe map between partons and jets
  - To reduce the sensitivity to hadronisation, one would like jets to form around "hard" cores (~partons)

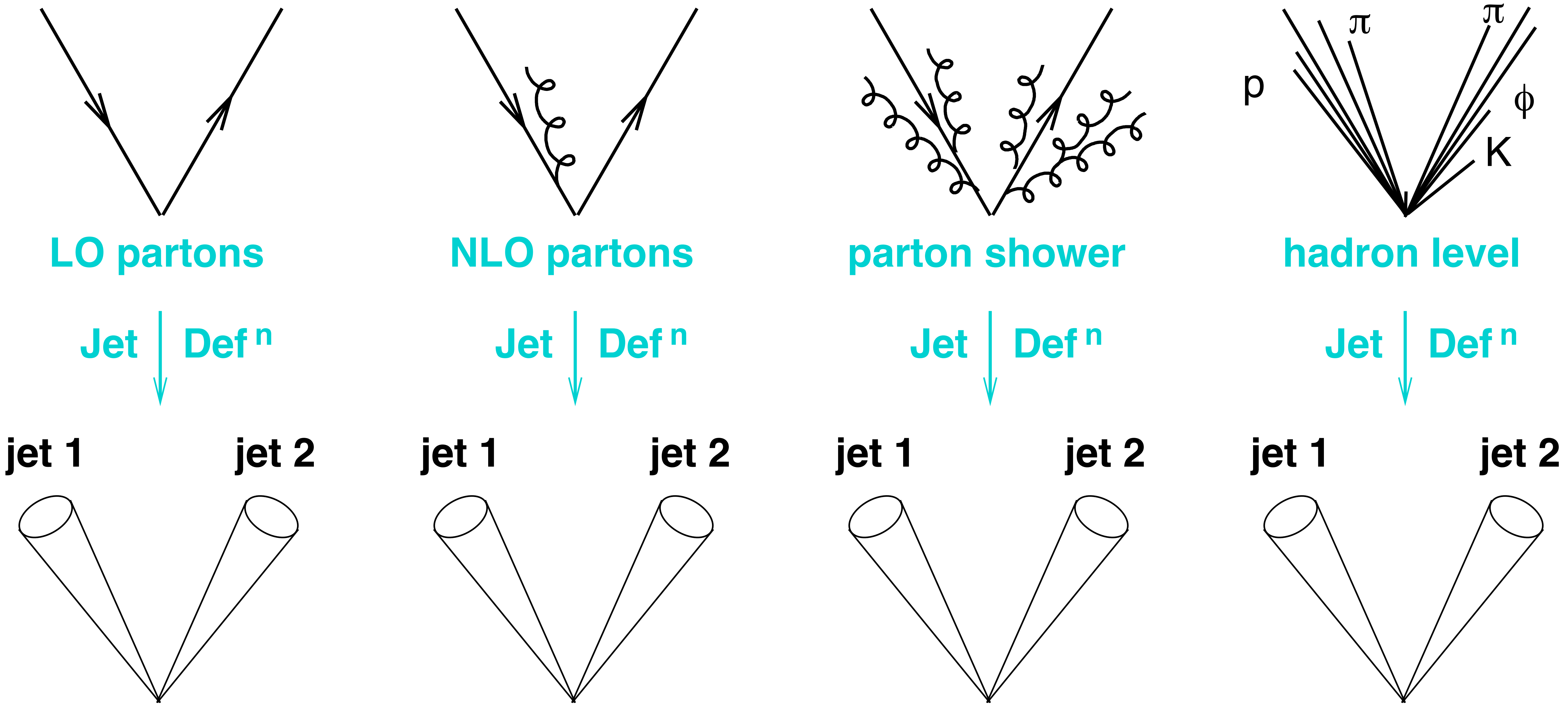


Figure Credits: Gavin Salam

A good jet algorithm should be resilient to "QCD effects" to define a good map to the underlying hard partonic structure.

The precise projection between jets and partons is ambiguous and depends on the specific algorithm

Generalised k_t algorithms

References:

- *kt clustering algorithm: Nucl. Phys. B 406 (1993) 187–224*
- *Cambridge/Aachen jet clustering algorithm* JHEP 08 (1997) 001/[*hep-ph/9907280]*
- *Anti-k_t jet clustering algorithm JHEP 04 (2008) 063*

- Common choice for LHC phenomenology

➡ Ingredients: a transverse momentum threshold $p_{t,min}$, a jet radius R, and the distance measures

Parameter p defines the algorithm:

$p = 1$ k_t algorithm
$p = 0$ Cambridge/Aachen (C/A) algorithm
$p = -1$ Anti-k_t algorithm

$$d_{ij} = \min\left(k_{t,i}^{2p}, k_{t,j}^{2p}\right)\frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2}, \quad d_{iB} = k_{t,i}^{2p}$$

Distance of particle i from particle j

Distance of particle i from the beam

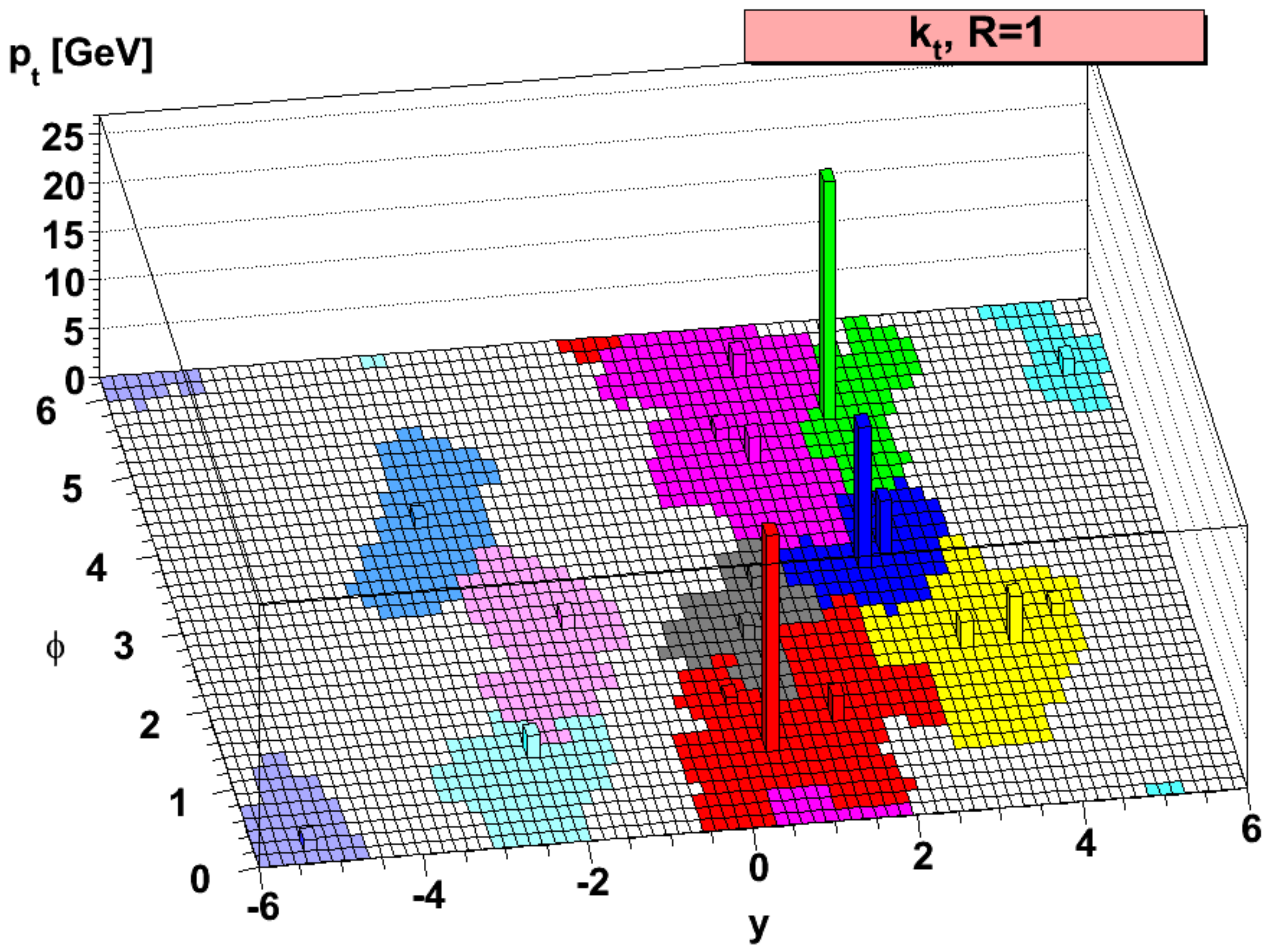
1. Find smallest of d_{ij} or d_{iB}
2. If d_{ij}, recombine them: e.g. replace p_i and p_j with $p_i + p_j$ (other recombination schemes are possible)
3. If d_{iB}, call p_i a jet and remove it from the list of particles
4. Go back to 1 and repeat until no particles are left. Consider only jets with $k_t > p_{t,min}$

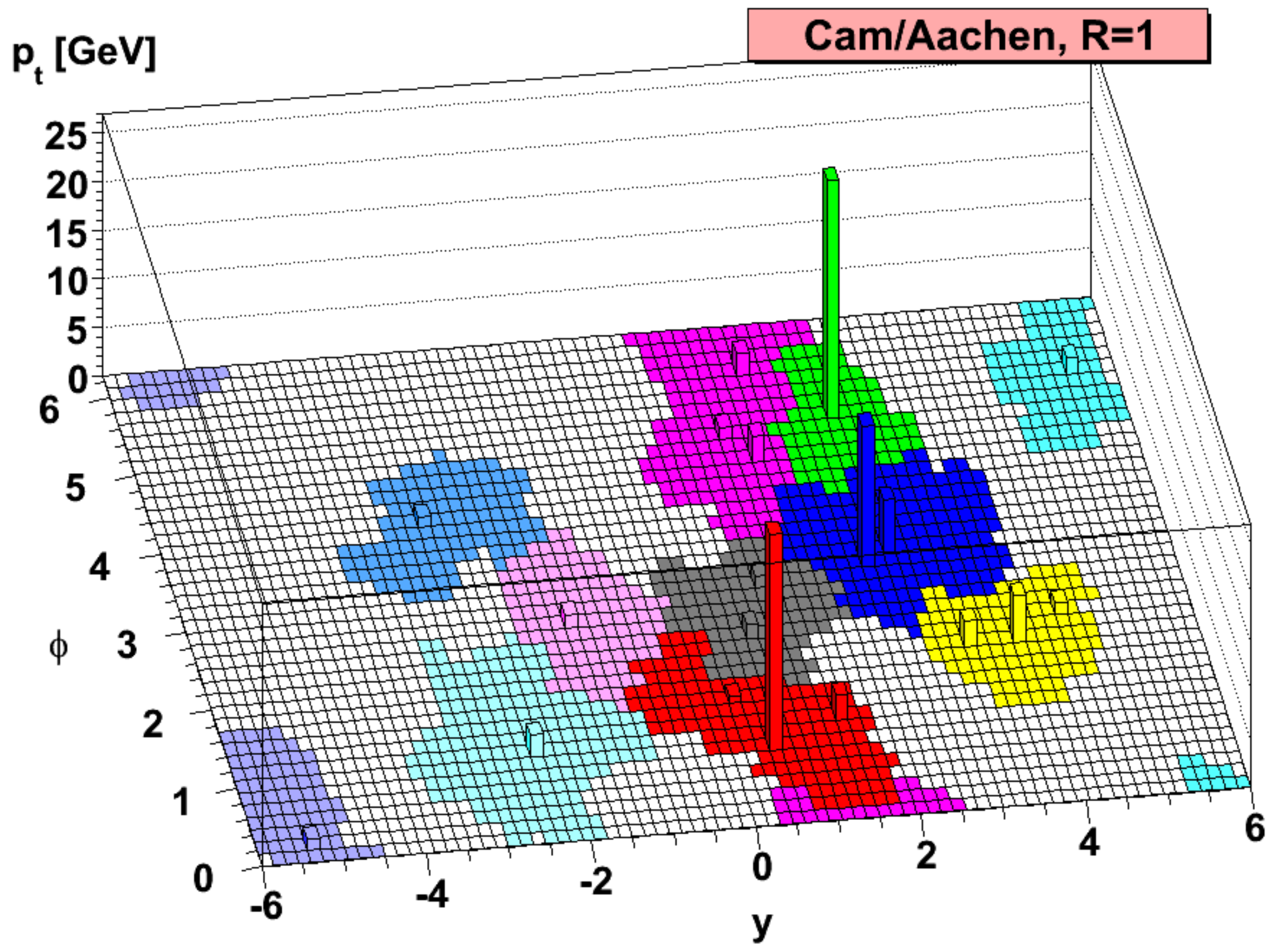
Generalised k_t algorithms

Figures from:

- *The anti-k_t jet clustering algorithm (Cacciari, Salam, Soyez)* 08021189
- *see also Towards Jetography (G. Salam)* 0906. 1833

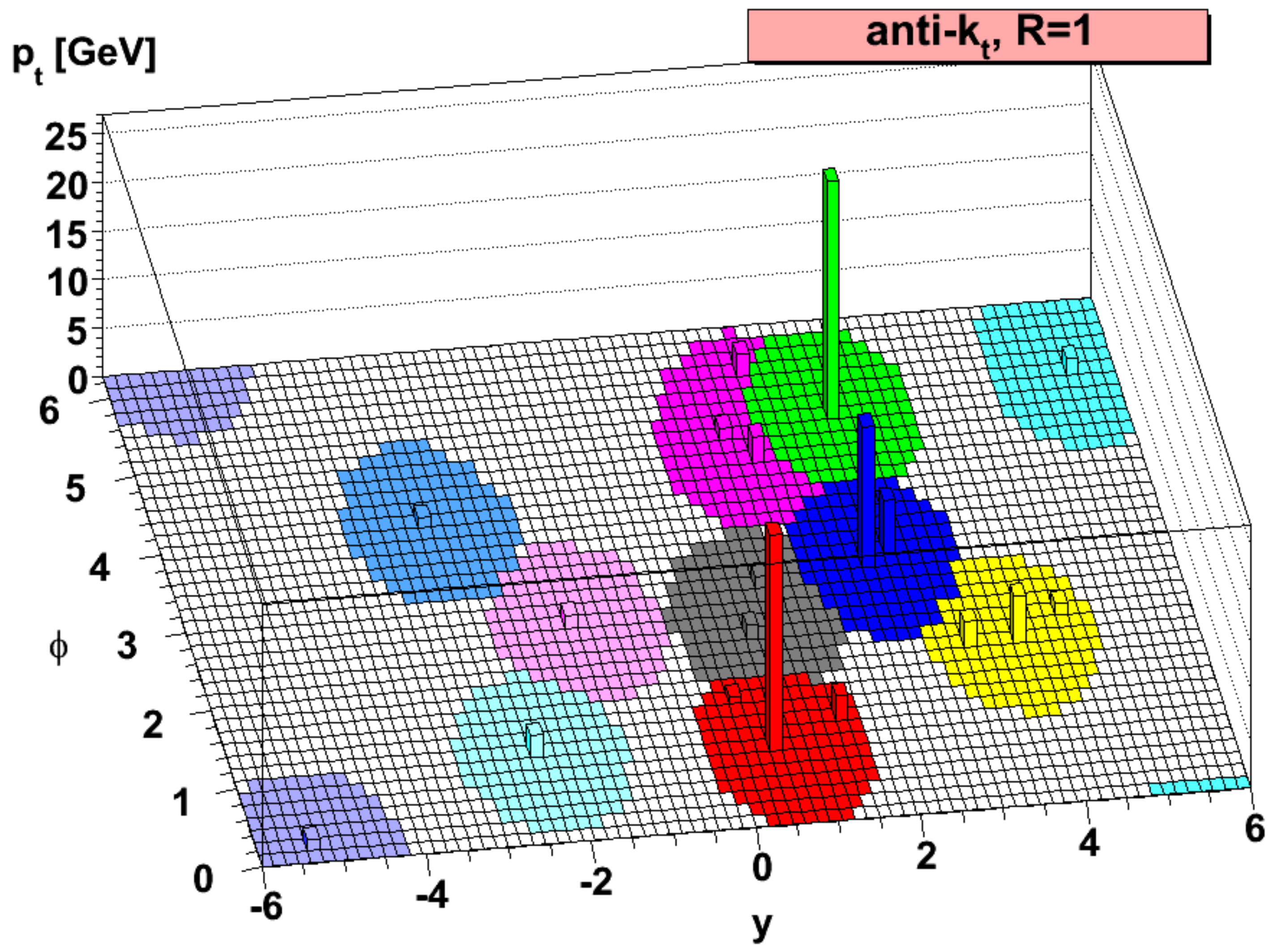
e.g. jets produced in the presence of hard particles surrounded by random soft radiation





Precise parton ↔ jets map depends on the algorithm

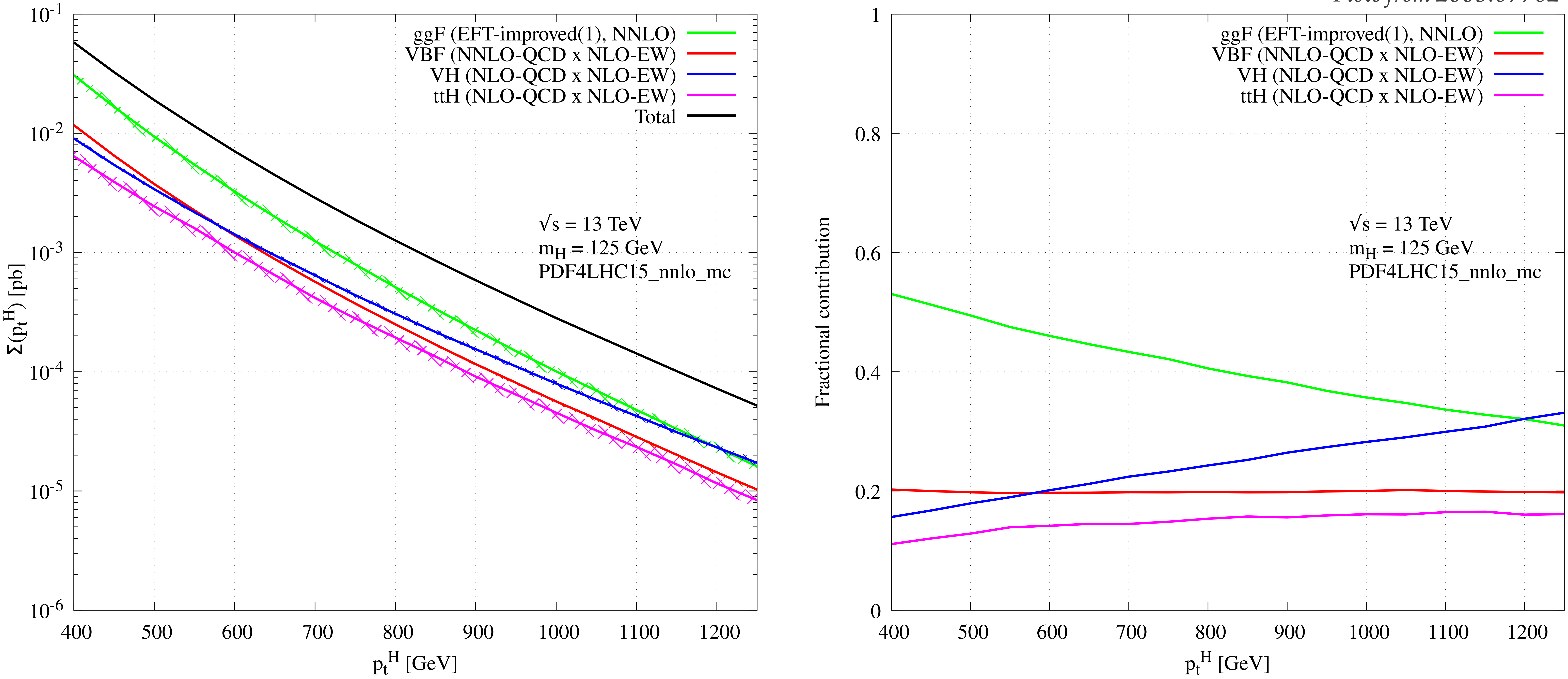
While for k_t and C/A the “catchment area” depends on the specific set of softer particles in the event, anti-k_t produces stable cone-like jets built around the hard core (unless two hard particles are close). This makes it the standard choice for LHC jets



Example: boosted Higgs search/detection

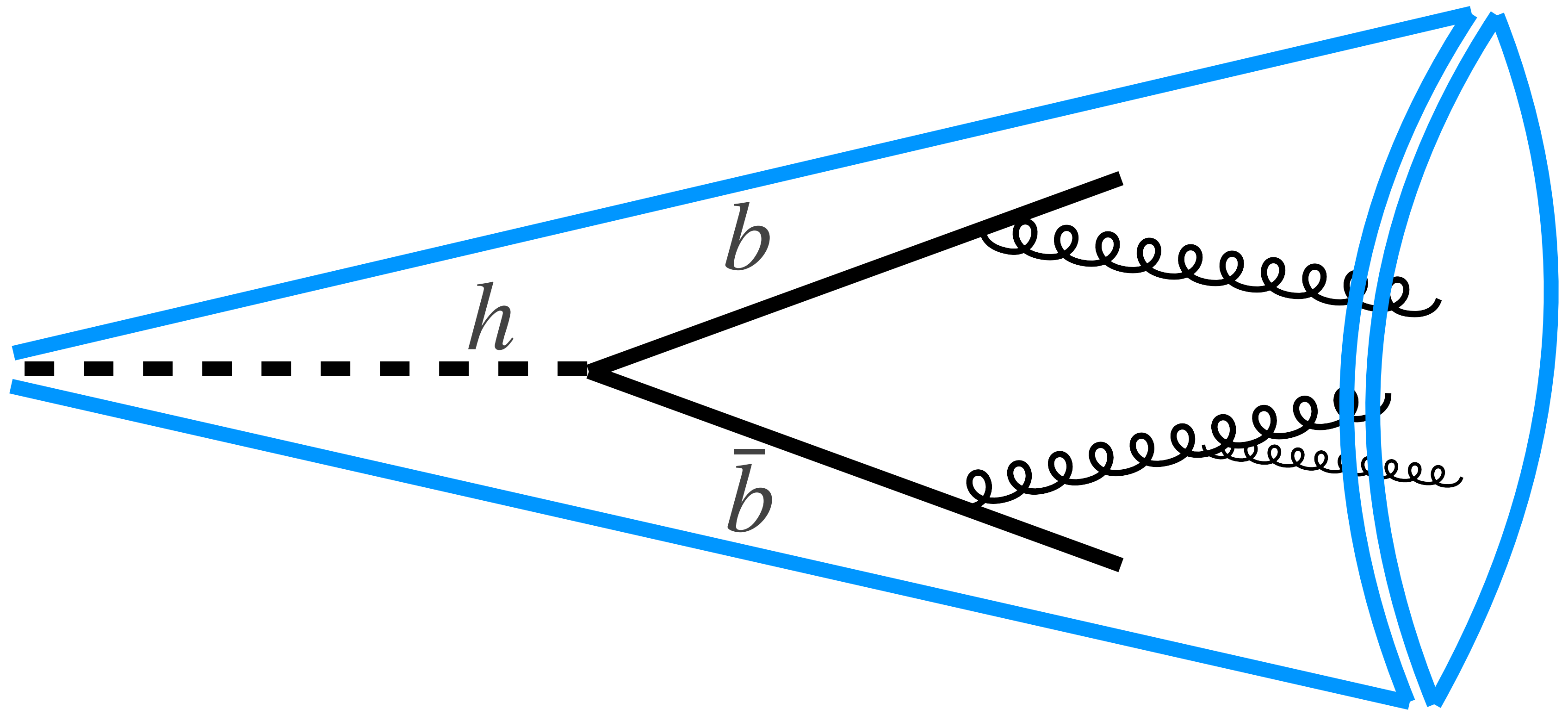
- Large q_T region crucial to search for NP effects. Many production modes yield a comparable contribution

Plots from 2005.07762

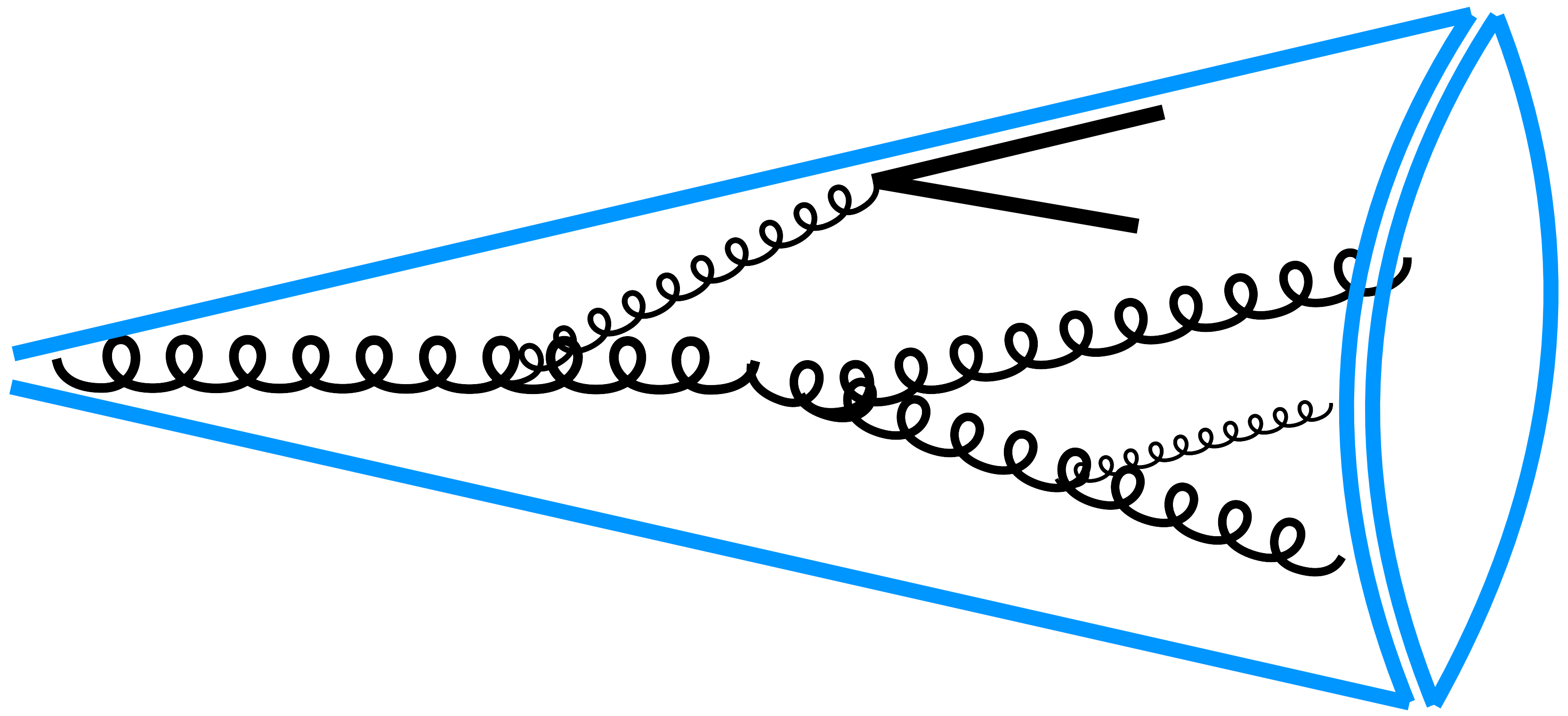


Boosted Higgs & “fat” jets

- We want to single out the Higgs decays (e.g. into two b quarks) from the overwhelming QCD background



vs.



(Over) Simplified procedure:

- Search for a fat jet (anti-k_t, R=0.8) with $p_T > 400$ GeV, containing two b tags (DDB tagger). Now look inside the jet to find the Higgs
- Clean up smaller sub-jets (C/A, R=0.4) removing soft radiation (grooming). This reduces pile-up and UE contamination
- Higgs candidate required to be consistent with a $b\bar{b}$ decay with invariant mass $m_{\mathrm{SD}} \simeq 125\,\mathrm{GeV}$. Signal is peaked at this mass, while QCD background (e.g. gluon jets) is smoothly falling

From CMS 2407.08012

Boosted Higgs & “fat” jets

A clear signal of the Higgs boson decay starts to show up!

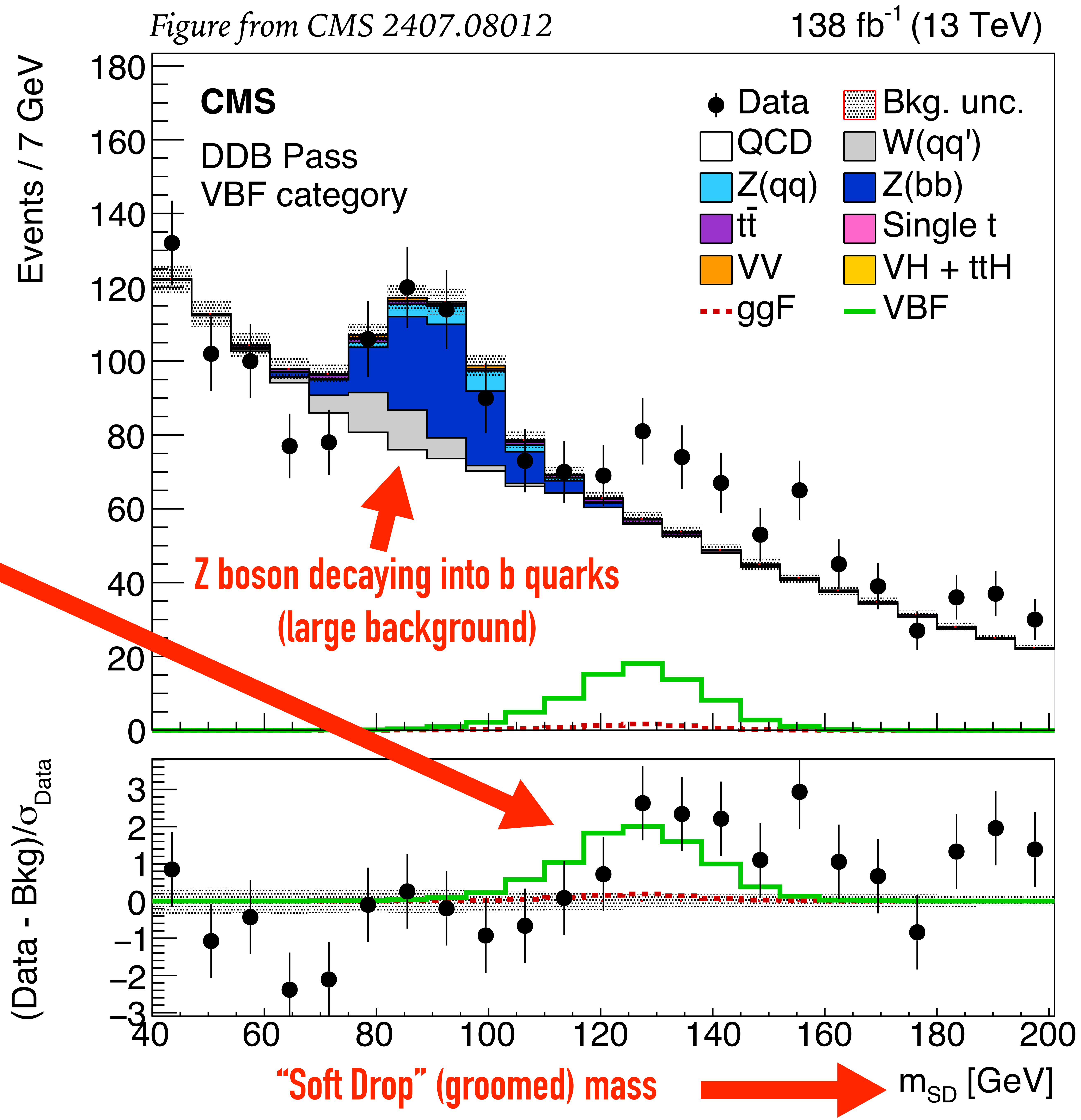


Figure from CMS 2407.08012

Z boson decaying into b quarks (large background)

“Soft Drop” (groomed) mass

Concluding remarks

- The success of QCD at the LHC relies on the key concepts of factorisation & asymptotic freedom. These concepts are at the core of the perturbative treatment of the theory, which sets the stage for the development of a lot of computational technology to predict the outcome of a high-energy scattering

- The scrutiny of precise experimental data with accurate theoretical calculations is an instrumental element of the modern collider-physics landscape. These investigations will reveal invaluable information about the structure of nature at the quantum level and may give us a hint of where to look for NP phenomena

- Moving forward (e.g. HL-LHC, future colliders), it is crucial that we come up with original ideas to tackle increasingly complicated calculations (QCD, EW) and explore the complex structure of gauge theories; as well as with new strategies to extract information from experimental data.

- Have a lot of fun with particle physics!