Introduction to Event Generators

Lecture 3: Hadronisation and Jets

Peter Skands (Monash University)
11th MCnet School, Lund 2017

**QCD**

\[ g^a \]

\[ (-ig_a t^a_{ij} \gamma^\mu) \]

**THEORY**

**PHENOMENOLOGY**

**INTERPRETATION**

**EXPERIMENT**

"Jets"

Figure by T. Sjöstrand

QCD

\[ g^a \]

\[ (-ig_a t^a_{ij} \gamma^\mu) \]
A Monte Carlo program for quark jet generation

T. Sjöstrand, B. Söderberg

A Monte Carlo computer program is presented, that simulates the fragmentation of a fast parton into a jet of mesons. It uses an iterative scaling scheme and is compatible with the jet model of Field and Feynman.

**Note:**
Field-Feynman was an early fragmentation model now superseded by the String (in PYTHIA) and Cluster (in HERWIG & SHERPA) models.

PYTHIA anno 1978
(then called JETSET)

LU TP 78–18
November, 1978

A Monte Carlo Program for Quark Jet Generation

A Monte Carlo computer program is presented, that simulates the fragmentation of a fast parton into a jet of mesons. It uses an iterative scaling scheme and is compatible with the jet model of Field and Feynman.

**Note:**
Field-Feynman was an early fragmentation model now superseded by the String (in PYTHIA) and Cluster (in HERWIG & SHERPA) models.
Here’s a fast parton

**Fast:** It starts at a high factorization scale

\[ Q = Q_F = Q_{\text{hard}} \]

It showers (bremsstrahlung)

It ends up at a low effective factorization scale

\[ Q \sim m_\rho \sim 1 \text{ GeV} \]
Here’s a fast parton

**Fast:** It starts at a high factorization scale

\[ Q = Q_F = Q_{\text{hard}} \]

It showers (bremsstrahlung)

It ends up at a low effective factorization scale

\[ Q \sim m_p \sim 1 \text{ GeV} \]

How about I just call it a hadron?

→ “Local Parton-Hadron Duality”
Early models: “Independent Fragmentation”

Local Parton Hadron Duality (LPHD) can give useful results for inclusive quantities in collinear fragmentation

Motivates a simple model:

“Independent Fragmentation”

But …

The point of confinement is that partons are coloured

Hadronisation = the process of **colour neutralisation**

→ Unphysical to think about independent fragmentation of a single parton into hadrons

→ Too naive to see LPHD (inclusive) as a justification for Independent Fragmentation (exclusive)

→ More physics needed
A physical hadronization model

Should involve at least TWO partons, with opposite color charges (e.g., think of them as $R$ and anti-$R$)*

*) Really, a colour singlet state $\frac{1}{\sqrt{3}} (|RR\rangle + |GG\rangle + |BB\rangle)$

Strong “confining” field emerges between the two charges when their separation $> \sim 1\text{fm}$
**Colour flow in parton showers** (leading-colour approximation)

**Example:** $Z^0 \rightarrow qq$

- **System #1**
  - Time progression
  - Colour flow development

- **System #2**
  - Colour flow development
  - Knots in colour flow

- **System #3**
  - Colour flow development
  - Knots in colour flow

**Coherence** of pQCD cascades → not much “overlap” between systems → Leading-colour approximation pretty good

(LEP measurements in $e^+e^- \rightarrow W^+W^- \rightarrow$ hadrons confirm this (at least to order $10\% \sim 1/N_c^2$))

**Note:** (much) more color getting kicked around in hadron collisions. More tomorrow.
Quark-Antiquark Potential

As function of separation distance

\[ F(r) \approx \text{const} = \kappa \approx 1 \text{ GeV/fm} \quad \leftrightarrow \quad V(r) \approx \kappa r \]

~ Force required to lift a 16-ton truck

What physical system has a linear potential?

Long Distances ~ Linear Potential

“Confined” Partons (a.k.a. Hadrons)
Motivates a model:

Let color field collapse into a narrow flux tube of uniform energy density

\[ \kappa \sim 1 \text{ GeV} / \text{fm} \]

Limit → Relativistic 1+1 dimensional worldsheet

In “unquenched” QCD

\[ g \rightarrow \text{qq} \rightarrow \text{The strings will break} \]

→ Gaussian suppression of high \( m_T^2 = m_q^2 + p_T^2 \)

Heavier quarks suppressed. \( \text{Prob(d:u:s:c)} \approx 1 : 1 : 0.2 : 10^{-11} \)

---

Schwinger Effect

Non-perturbative creation of \( e^+e^- \) pairs in a strong external Electric field

\[ \mathcal{P} \propto \exp \left( \frac{-m^2 - p_T^2}{\kappa / \pi} \right) \]

(\( \kappa \) is the string tension equivalent)

---

Map:

- **Quarks** → String Endpoints
- **Gluons** → Transverse Excitations (kinks)
- Physics then in terms of string worldsheet evolving in spacetime
- Probability of string break (by quantum tunneling) constant per unit area → AREA LAW

Simple space-time picture
Details of string breaks more complicated (e.g., baryons, spin multiplets)

Gluon = kink on string, carrying energy and momentum → STRING EFFECT
Having selected a hadron flavor

How much momentum does it take?

The meson \( M \) takes a fraction \( z \) of the quark momentum, How big that fraction is, \( z \in [0,1] \), is determined by the fragmentation function, \( f(z,Q_0^2) \)

\( m_T^2 = 0.25 \)

\( m_T^2 = 1 \)

\( m_T^2 = 4 \)
Causality → Left-Right Symmetry
→ Constrains form of fragmentation function!
→ Lund Symmetric Fragmentation Function

\[ f(z) \propto \frac{1}{z} (1 - z)^a \exp \left( -\frac{b(m_h^2 + p_{\perp h}^2)}{z} \right) \]

**Note:** In principle, \( a \) can be flavour-dependent. In practice, we only distinguish between baryons and mesons.
**ITERATIVE STRING BREAKS**

**Causality → May iterate from outside-in**

*Note: using light-cone coordinates: \( p_+ = E + p_z \)*

On average, expect energy of \( n^{th} \) “rank” hadron \( \sim E_n \sim \langle z \rangle^n E_0 \)
In Spacetime:

String tension $\approx 1$ GeV/fm $\rightarrow$ a 5-GeV quark can travel 5 fm before all its kinetic energy is transformed to potential energy in the string.

Then it must start moving the other way ($\rightarrow$ "yo-yo" model of mesons. Note: string breaks $\rightarrow$ several mesons)

In Rapidity:

\[ y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) = \frac{1}{2} \ln \left( \frac{(E + p_z)^2}{E^2 - p_z^2} \right) \]

(for $m \rightarrow 0$ : $\frac{1}{2} \ln \left( \frac{1 + \cos \theta}{1 - \cos \theta} \right) = -\ln \tan(\theta/2) = \eta$)

\[ \Rightarrow \Delta y \text{ difference is invariant} \]

Particle Production:

If the quark gives all its energy to a single pion traveling along the z axis

\[ y_{\text{max}} \sim \ln \left( \frac{2E_q}{m_\pi} \right) \]

Increasing $E_q \rightarrow$ logarithmic growth in rapidity range $\langle n_{\text{ch}} \rangle \approx c_0 + c_1 \ln E_{cm}$, $\sim$ Poissonian multiplicity distribution

Scaling in lightcone $p_\pm = E \pm p_z$

$\Rightarrow$ flat central rapidity plateau (+ some endpoint effects)

Rapidity is useful because it is additive under Lorentz boosts (along the rapidity axis)

\[ y' = y + \ln \sqrt{\frac{1 - \beta}{1 + \beta}} \]

"Pseudorapidity" (convenient variable in momentum space)
1980: string (colour coherence) effect

Predicted unique event structure; inside & between jets. Confirmed first by JADE 1980. Generator crucial to sell physics! (today: PS, M&M, MPI, ...)

string motion in the event plane (without breakups)
1980: string (colour coherence) effect

Predicted unique event structure; inside & between jets.
Confirmed first by JADE 1980.

Generator crucial to sell physics!
(today: PS, M&M, MPI, ...)

string motion in the event plane (without breakups)
Differences between Quark and Gluon Jets

More recent study (LHC)
Gluon connected to two string pieces

Each quark connected to one string piece
→ expect factor $2 \sim C_A/C_F$ larger particle multiplicity in gluon jets vs quark jets

Can be hugely important for discriminating new-physics signals (decays to quarks vs decays to gluons, vs composition of background and bremsstrahlung combinatorics)
Starting observation: “Preconfinement”

+ Force $g \rightarrow qq$ splittings at $Q_0$

→ high-mass $q$-$\bar{q}$bar “clusters”

Isotropic 2-body decays to hadrons

according to $PS \approx (2s_1+1)(2s_2+1)(p^*/m)$

Large clusters $\rightarrow$ string-like. (In PYTHIA, small strings $\rightarrow$ cluster-like).
Think of jets as **projections** that provide a universal view of events.

I’m not going to cover the many different types of jet clustering algorithms (k_T, anti-k_T, C/A, cones, ...) – see e.g., lectures & notes by G. Salam.

➤ Focus instead on the physical origin and MC modeling of jets.
Jet clustering algorithms

Map event from low E-resolution scale (i.e., with many partons/hadrons, most of which are soft) to a higher E-resolution scale (with fewer, hard, IR-safe, jets)

Jet Clustering
(Deterministic*)
(Winner-takes-all)

Parton Showering
(Probabilistic)

Many soft particles

\[ Q \sim \Lambda \sim m_{\pi} \sim 150 \text{ MeV} \]

Hadronization

\[ Q \sim Q_{\text{had}} \sim 1 \text{ GeV} \]

A few hard jets

\[ Q \sim E_{\text{cm}} \sim M_{X} \]

Parton shower algorithms

Map a few hard partons to many softer ones

Probabilistic \rightarrow\text{closer to nature.}

Not uniquely invertible by any jet algorithm*

(*) See “Qjets” for a probabilistic jet algorithm, arXiv:1201.1914
(*) See “Sector Showers” for a deterministic shower, arXiv:1109.3608
Definition: an observable is infrared safe if it is insensitive to

**SOFT** radiation:
Adding any number of infinitely soft particles (zero-energy) should not change the value of the observable

**COLLINEAR** radiation:
Splitting an existing particle up into two comoving ones (conserving the total momentum and energy) should not change the value of the observable

Note: some people use the word “infrared” to refer to soft only. Hence you may also hear “infrared and collinear safety”. Advice: always be explicit and clear what you mean.
Counting the number of particles/tracks is ... ?

The number of tracks, weighted by energy times angle*?

angle*: with respect to some principal axis representing the “collinear” direction (e.g., jet axis or “event-shape” axis)
Real life does not have infinities, but pert. infinity leaves a real-life trace

\[ \alpha_s^2 + \alpha_s^3 + \alpha_s^4 \times \infty \rightarrow \alpha_s^2 + \alpha_s^3 + \alpha_s^4 \times \ln p_t/\Lambda \rightarrow \alpha_s^2 + \alpha_s^3 + \alpha_s^3 \]

\text{BOTH WASTED}
YOU decide how to look at event

The construction of jets is inherently ambiguous

**Jet Definition**

1. Which particles get grouped together?
   - **JET ALGORITHM**
     (+ size/resolution parameters)

2. How will you combine their momenta?
   - **RECOMBINATION SCHEME**
     (e.g., ‘E’ scheme: add 4-momenta)

Ambiguity complicates life, but gives flexibility in one’s view of events

→ At what resolution / angular size are you looking for structure(s)?

→ Do you prefer “circular” or “QCD-like” jet areas? (Collinear vs Soft structure)

→ Sequential clustering → substructure (veto/enhance?)
1. Sequential Recombination

Take your 4-vectors. Combine the ones that have the lowest ‘distance measure’

Different names for different distance measures:

- Durham $k_T$: $\Delta R_{ij}^2 \times \min(k_{Ti}^2, k_{Tj}^2)$
- Cambridge/Aachen: $\Delta R_{ij}^2$
- Anti-$k_T$: $\Delta R_{ij}^2 / \max(k_{Ti}^2, k_{Tj}^2)$
- ArClus (3→2): $p_{\perp}^2 = s_{ij}s_{jk}/s_{ijk}$

→ New set of (n-1) 4-vectors

Iterate until A or B (you choose which):
- A: all distance measures larger than something
- B: you reach a specified number of jets

Look at event at:
- specific resolution
- specific $n_{jets}$
WHY $K_T$ (OR $P_T$ OR $\Delta R$)?

Attempt to (approximately) capture universal jet-within-jet-within-jet... behavior

Recall: Approximate full matrix element

$$\frac{|M^{(0)}_{X+1}(s_{i1}, s_{1k}, s)|^2}{|M^{(0)}_X(s)|^2} = 4\pi\alpha_s C_F \left( \frac{2s_{ik}}{s_{i1}s_{1k}} + ... \right)$$

by Leading-Log limit of QCD $\rightarrow$ universal dominant terms

Rewritings in soft/collinear limits

"smallest" $k_T$ (or $p_T$ or $\theta_{ij}$, or ...) $\rightarrow$ largest Eikonal (and/or most collinear)
2. “Cone” type

Take your 4-vectors. Select a procedure for which “test cones” to draw

Different names for different procedures

Seeded (obsolete): start from hardest 4-vectors (and possibly combinations thereof, e.g., CDF midpoint algorithm) = “seeds”

Unseeded: smoothly scan over entire event, trying everything

Sum momenta inside test cone → new test cone direction

Iterate until stable (test cone direction = momentum sum direction)

Warning: to optimise speed, seeded algorithms were sometimes used in the past. INFRARED UNSAFE
May look pretty similar in experimental environment …
But IR unsafe is not nice to your (perturbative) theory friends …

**Unsafe**: badly divergent in pQCD → large IR corrections:

\[
\text{IR Sensitive Corrections} \propto \alpha_s^n \log^m \left( \frac{Q_{\text{UV}}^2}{Q_{\text{IR}}^2} \right), \quad m \leq 2n
\]

Even if we have a hadronization model which computes these corrections, the dependence on it is larger → uncertainty

**Safe** → IR corrections power suppressed:

\[
\text{IR Safe Corrections} \propto \frac{Q_{\text{IR}}^2}{Q_{\text{UV}}^2}
\]

Can still be computed (MC) but can also be neglected (pure pQCD)

Let’s look at an example …
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

"Seeded Cone Algorithm"
Start from "hardest" seeds

ICPR iteration issue
Iterative Cone Progressive Removal

$p_T$ (GeV/c)

cone iteration

cone axis
cone

rapidity
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

"Seeded Cone Algorithm"
Start from “hardest” seeds

ICPR iteration issue
Iterative Cone Progressive Removal

Slides from G. Salam
“Seeded Cone Algorithm”
Start from “hardest” seeds

Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe \(\Rightarrow\) perturbative calculations give \(\infty\)

Slides from G. Salam
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

"Seeded Cone Algorithm"

Start from "hardest" seeds

ICPR iteration issue

Iterative Cone Progressive Removal

Slides from G. Salam
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

"Seeded Cone Algorithm"
Start from "hardest" seeds

ICPR iteration issue
Iterative Cone Progressive Removal

Slides from G. Salam
"Seeded Cone Algorithm"
Start from "hardest" seeds

ICPR iteration issue
Iterative Cone Progressive Removal

Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

Slides from G. Salam
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$.

"Seeded Cone Algorithm"
Start from “hardest” seeds

ICPR iteration issue
Iterative Cone Progressive Removal

Slides from G. Salam
“Seeded Cone Algorithm”
Start from “hardest” seeds

Collinear splitting can modify the hard jets: ICPR algorithms are
collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

"Seeded Cone Algorithm" 
Start from "hardest" seeds

ICPR iteration issue
Iterative Cone Progressive Removal

Slides from G. Salam
"Seeded Cone Algorithm"
Start from "hardest" seeds

Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$
"Seeded Cone Algorithm"
Start from "hardest" seeds

Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

Slides from G. Salam
“Seeded Cone Algorithm”
Start from “hardest” seeds

Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe → perturbative calculations give $\infty$. 

Slides from G. Salam
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$. 

"Seeded Cone Algorithm"
Start from "hardest" seeds

ICPR iteration issue
Iterative Cone Progressive Removal

Slides from G. Salam
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$
"Seeded Cone Algorithm"
Start from “hardest” seeds

Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe \(\Rightarrow\) perturbative calculations give \(\infty\)
Collinear splitting can modify the hard jets: ICPR algorithms are collinear unsafe $\Rightarrow$ perturbative calculations give $\infty$

Note: none of the jet algorithms in use at LHC are seeded. But worth understanding issue if/when you consider proposals for new observables
Use IR Safe algorithms

To study short-distance physics

Recombination-type algos → “inverse shower”
→ can study jet substructure → test shower properties & distinguish BSM?

“Cone-like”: SiSCon (unseeded)
“Recombination-like”: $k_T$, Cambridge/Aachen
“Hybrid”: Anti-$k_T$ (cone-shaped jets from recombination-type algorithm; note: clustering history not ~ shower history)

Use IR Sensitive observables

E.g., number of tracks, identified particles, …
To explicitly study hadronisation and models of IR physics

→ message is not to avoid IR unsafe observables at all costs. But to know when and how to use them.
SUMMARY

Jets: Discovered at SPEAR (SLAC ’72) and DORIS (DESY ’73): at $E_{CM} \sim 5$ GeV
Collimated sprays of nuclear matter (hadrons).
Interpreted as the “fragmentation of fast partons” -> MC generators

PYTHIA (and EPOS): Strings enforce confinement; break up into hadrons
Based on linear confinement: $V(r) = \kappa r$ at large distances + Schwinger tunneling
Powerful energy-momentum picture, with few free parameters
Not very predictive for flavour/spin composition; many free parameters

HERWIG and SHERPA employ ‘cluster model’
Based on universality of cluster mass spectra + ‘preconfinement’
Algorithmically simpler; flavour/spin composition largely from hadron masses

NB: many indications that confinement is more complicated in pp
~ well understood in “dilute” environments (ee: LEP) ~ vacuum
LHC is providing a treasure trove of measurements on jet fragmentation,
identified particles, minimum-bias, underlying event, … tomorrow’s lecture!
Extra Slides
Generally, expect few-hundred MeV shifts by hadronisation

Corrections to IR safe observables are “power corrections”

\[ \propto \frac{\Lambda_{QCD}^2}{Q_{OBS}^2} \]

Corrections for jets of radius \( R = \Delta \eta \times \Delta \phi \)

\[ \propto \frac{1}{R} \]

See
Korchemsky, Sterman, NPB 437 (1995) 415
Seymour, NPB 513 (1998) 269
Dasgupta, Magnea, Salam, JHEP 0802 (2008) 055

Simple analytical estimate \( \to \sim 0.5 \text{ GeV} / R \) correction from hadronisation
(scaled by colour factor)

Significant differences between codes/tunes \( \to \) important to pin down with precise QCD hadronisation measurements at LHC

The effects of hadronisation

Monte Carlo tunes:
- Herwig 6 (AUET2)
- Pythia 8 (Monash 13)

Jet radius, flavour:
- R=0.2, quarks
- R=0.4, quarks
- R=0.2, gluons
- R=0.4, gluons

pp, 7 TeV, no UE

\[ \Delta p_t \times R \frac{C_F}{C} \text{ (scaled by } R \frac{C_F}{C}) \]

\[ p_t \text{ (parton)} \text{ [GeV]} \]

\[ \Delta p_t \text{ [GeV]} \]

Dasgupta, Dreyer, Salam, Soyez, JHEP 1606 (2016) 057

Simple analytical estimate

\[ \sim 0.5 \text{ GeV} / R \text{ correction from hadronisation} \]
M. Strassler, K. Zurek, Phys. Lett. B651 (2007) 374; ...
Figure 1: A schematic depiction of pair production of dark quarks forming two emerging jets.

Emerging Jets

Dark Mesons

Requirements for a model to produce emerging jet phenomenology:
- Hierarchy between the mediator mass and hidden sector mass.
- Strong coupling in hidden sector → large particle multiplicity.
- Macroscopic decay lengths of hidden sector fields back to the visible sector.

Schwaller, Stolarski, Weiler
JHEP 1505 (2015) 059
⇒ **Pythia** allows for hadronization of 3 generic states:

- **colour octet uncharged**, like \( \tilde{g} \), giving \( \tilde{g}u\bar{u}, \tilde{g}u\bar{d}, \tilde{g}g, \ldots \),
- **colour triplet charge** \(+2/3\), like \( \tilde{t} \), giving \( \tilde{t}\bar{u}, \tilde{t}u\bar{d}_0, \ldots \),
- **colour triplet charge** \( -1/3 \), like \( \tilde{b} \), giving \( \tilde{b}\bar{c}, \tilde{b}s_1u, \ldots \).

Most hadronization properties by analogy with normal string fragmentation, but **glueball formation** new aspect, assumed \( \sim 10\% \) of time (or less).

R-hadron interactions with matter: part of detector simulation, i.e. GEANT, not **Pythia**

Freight-train BSM particle surrounded by light pion/gluon cloud → little dE/dx + charge flipping!