Heavy quark fragmentation and hadronization (and gluon splitting)

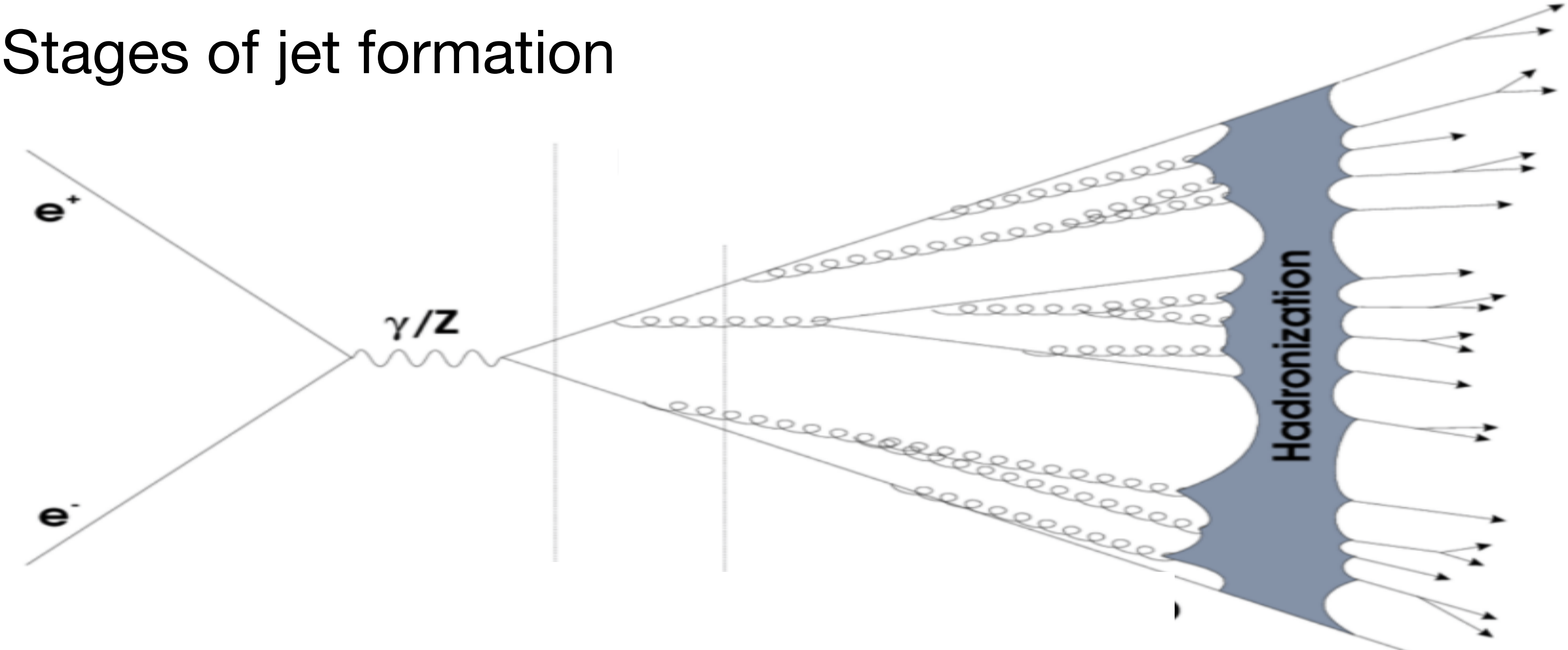
Loukas Gouskos (Brown University)
on behalf of the WG1-PREC group

3rd ECFA Workshop on e^+e^- EW/Higgs/Top Factories

Paris, Oct 9-11, 2024

Introduction

- Stages of jet formation



Hard Scattering
q/g production

Parton showering
gluon radiation
then: gluon splitting

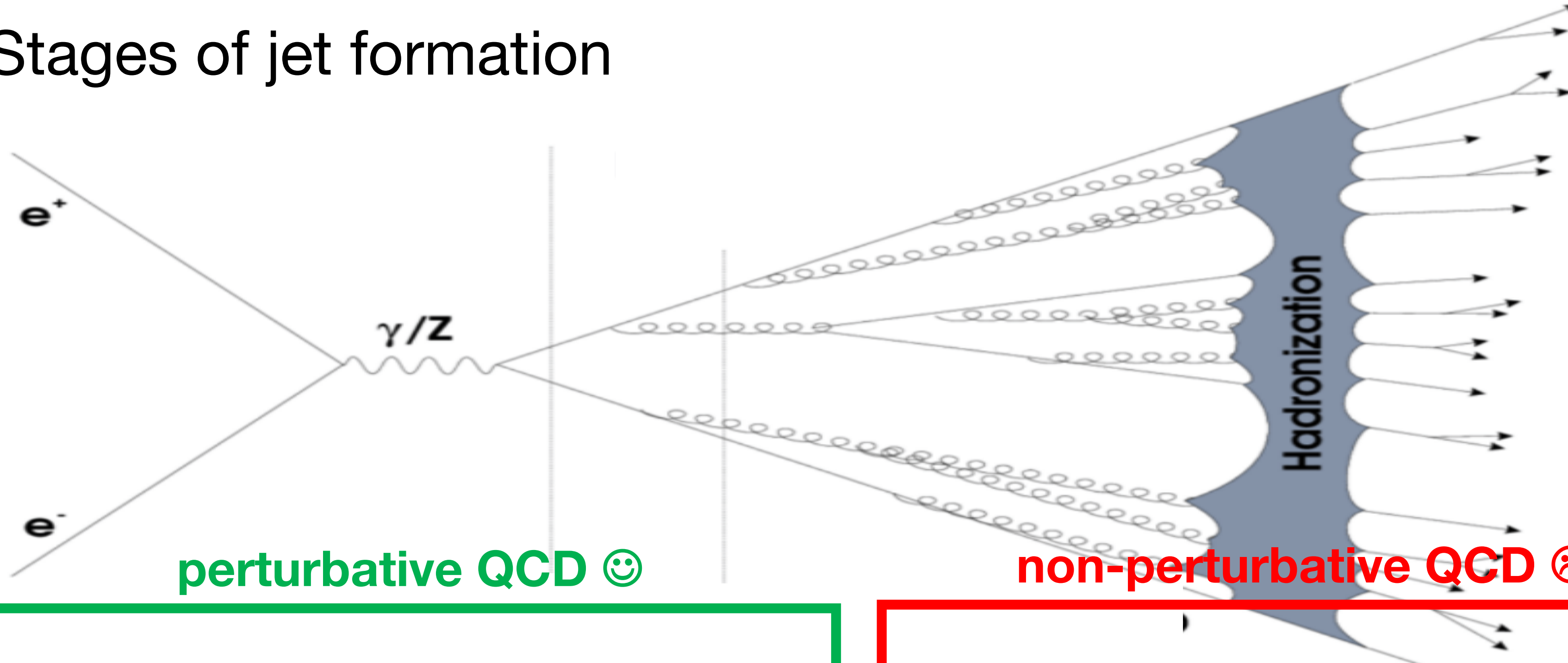
Fragmentation
hadrons from g/g

Hadronization
color neutral
hadrons

Higher Energy scale → **Lower energy scale**

Introduction

- Stages of jet formation



perturbative QCD ☺

non-perturbative QCD ☹

Hard Scattering	Parton showering	Fragmentation	Hadronization
q/g production	gluon radiation then: gluon splitting	hadrons from g/g	color neutral hadrons

→

Higher Energy scale **Lower energy scale**

Fragmentation/Hadronization

- **Non-perturbative** processes
  - We rely on different phenomenological models to simulate them
    - predictivity and uncertainty obtained from consistency between different models and datasets.
  - Model parameters tuned using experimental data
    - particularly from e^+e^- and ep collisions
- Extra complexity for bottom and charm (Heavy flavor) quarks:
  - Large mass [m_b/m_c O(GeV)] → fragmentation and hadronization more sensitive to the QCD scale [Λ_{QCD} O(200 MeV)]
- Usually important differences between generators → Large systematic uncertainties
  - Often: the dominant source

e.g., EW physics

- W-mass measurement
  - σ_{WW} threshold
    - Fragmentation/hadronization unc. subdominant
  - Kinematic reconstruction of Semileptonic & fully hadronic decays
    - Fragmentation/hadronization uncertainties dominate

ALEPH (EPJ C 38, 2004)

Source	uncertainty (fb)			
	$\ell\nu\ell\nu$	$\ell\nu qq$	qqqq	total
Tracking	4	19	31	54
Simulation of calorimeters	–	9	26	31
Hadronization models	–	27	8	35
Z peak $q\bar{q}$ fragmentation	–	–	20	20
Inter-W final state interaction	–	–	28	28
Background contamination	9	5	31	35
Lepton identification	1	2	–	3
Beam-related background	10	17	37	22
$\mathcal{O}(\alpha)$ corrections DPA	2	9	12	6
Luminosity	8	35	44	87
Simulation statistics	6	20	14	25
Total	17	57	87	126

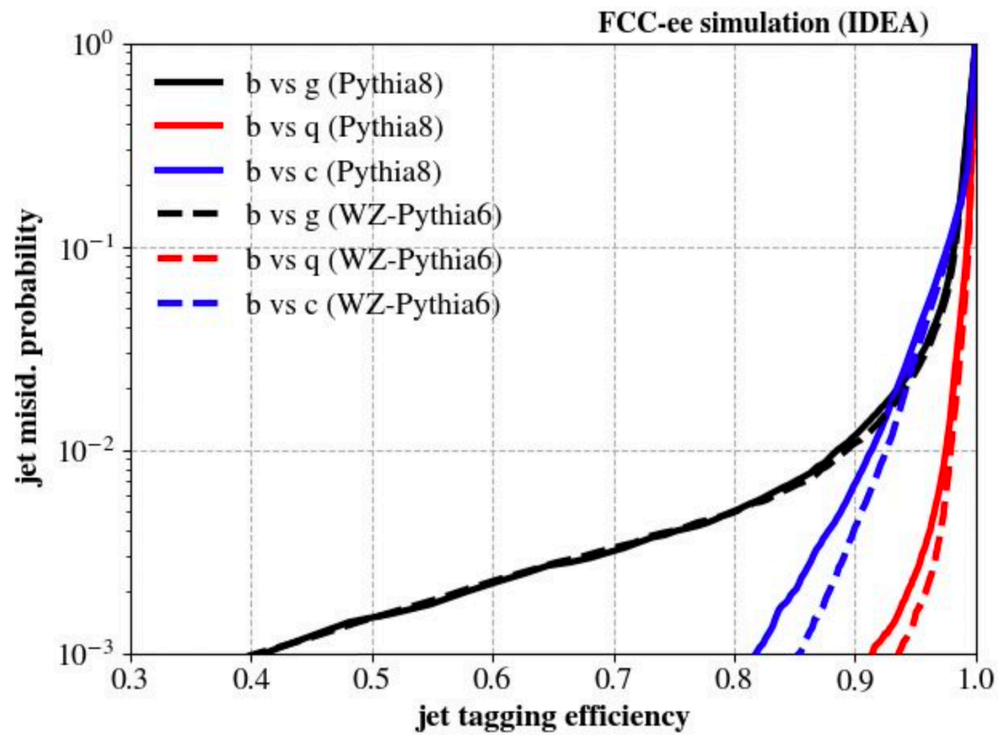
ILC (ref)

ΔM_W [MeV]	ILC	ILC	ILC	ILC
$\sqrt{s}$ [GeV]	250	350	500	1000
$\mathcal{L}$ [fb^{-1}]	500	350	1000	2000
$P(e^-)$ [%]	80	80	80	80
$P(e^+)$ [%]	30	30	30	30
jet energy scale	3.0	3.0	3.0	3.0
hadronization	1.5	1.5	1.5	1.5
pileup	0.5	0.7	1.0	2.0
total systematics	3.4	3.4	3.5	3.9
statistical	1.5	1.5	1.0	0.5
total	3.7	3.7	3.6	3.9

e.g., Higgs couplings

- Impact on jet tagging performance → Directly affects Higgs couplings
  - Nominal version (**solid**)
    - Trained and tested on Pythia 8 events
  - Alternative version (**dashed**)
    - Trained on Pythia 8 Tested on WZ+Pythia 6
- Visible differences (~30%)
  - particularly b-vs-c (blue)
- More interesting: use a completely different Parton Shower generator
  - in progress

bottom-tagging



ECFA WG1-PRECision group

- Precision in theory & experiment [Group's page]
  - Conveners: Paolo Azzurri, Ayres Freitas, Adrian Irles, Andreas Meyer

- Within WG1: Dedicated "focus team" to address the challenge of Heavy Quark Fragmentation & Gluon Splitting (BCFrag & GSplit)
  - Dedicated Twiki: https://gitlab.in2p3.fr/ecfa-study/ECFA-HiggsTopEW-Factories/-/wikis/FocusTopics/BCfrag

- Experts team: Eli Ben Haim, LG, Simon Plätzer, Andrzej Siodmok, Torbjörn Sjöstrand, Maria Ubiali + WG1 conveners

ECFA WG1-PRECision group (II)

- **Goals** of "BCFrag & Gsplit" team:
  - Strategy to measure b-/c-fragmentation functions & hadronization rates
    - @ Z-pole and higher energies
  - g→bb/cc and understand interplay with H→gg & H→bb/cc
    - @ Z-pole and higher energies

- Meetings:
  - Mar 22, 2024: Mini-workshop [https://indico.cern.ch/event/1387393/]
  - Oct 3, 2023 [https://indico.cern.ch/event/1318673/]
  - Aug 31, 2023 [https://indico.cern.ch/event/1318673/]

Direct connection to other Focus Topics

- **H→ss** [add link]:
  Potential to fully establish Higgs couplings to 2nd-G Ferminos
  - Needs excellent control of H→gg/bb/cc
  - Yet, current knowledge of g→bb/cc results in large systematic uncertainties

- **m_W** [add link]:
  Aim O(0.5 MeV) precision
  - LEP: non-perturbative QCD effect → dominant syst. uncertainties

- And others..

Where are we today?

- Three main axes of work
  - Review current state-of-the-art
  - Theory developments needed/envisioned
    - How do they compare with statistical uncertainty anticipated @Future e^+e^- Colliders?
  - EXP requirements (Higgs couplings, m_W (kin fit), R_b)
    - Detectors, simulations, analysis techniques

Long story short: Future e^+e^- colliders require developments in both TH and EXP fronts to reduce these b/c fragmentation and hadronization uncertainties

Main points: Theory

- High precision limit (i.e., Future e^+e^-):
  - Fragmentation functions are not universal but depend on observable and initial state
  - We need to disentangle perturbative and non-perturbative parts
    - but needs dedicated tuning of the fragmentation model
  - Can we do it? Do we need new approach? (e.g., ML-based)

- Special challenge: heavy flavor quarks (b/c)
  - g→bb/cc only modelled in the perturbative step of the process
    - i.e., not in the string/cluster fragmentation
      but m_b/m_c are parameters in the shower
  - Several development in the perturbative part
    - Open question: How to supplement it with the non-perturbative part?

Main points: Theory

- Work started

- Plan of action
  - Systematic understanding of the modeling parameters
  - Define observables that probe the interface between parton showering and hadronization
    - Extend the study to correlations between jet constituents

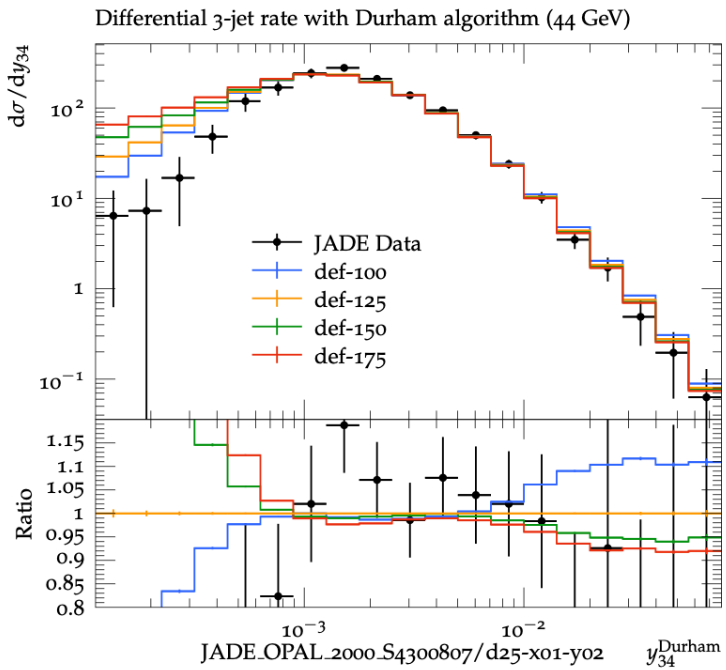
Effort intensified since @Les Houches'23 [details]

- Test them using old data (LEP/SLC)

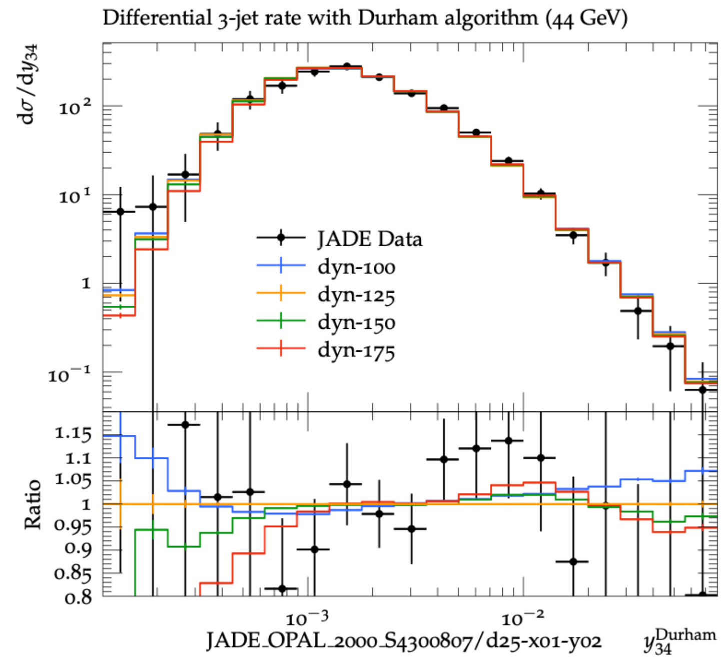
Recent developments

[e.g., 2404.09856](e.g., 2404.09856)

- Disentangling perturbative and non-perturbative effects



Default version:
Q_0: free parameter



New version:
Q_0: Dynamic parameter

- Smaller dependence on Q_0
- Improve data description data [far from done though]

List of observables

2401.07564

Observable	e^+e^-	pp
Event shapes and angular distributions		
Inclusive B/D production cross section	primary production is well known from theory, so any ”excess” is from gluon splitting	combines primary production, gluon splitting, and MPI (multiparton interactions) contributions, each with significant theoretical uncertainties
Flavour composition as far back in decay chains as can be traced (even equal D^{*0} and D^{*+} rates gives unequal D^0 and D^+ ones)	we do not expect sizeable momentum dependence, but interesting to contrast mesons and baryons for smaller ones	significant p_T dependence observed and to be studied further, also high- vs. low-multiplicity events, rapidity, ..., which is important for development/tuning of colour reconnection models
Particle-antiparticle production asymmetries	none expected, except tiny from CP-violation in oscillations	asymmetries expected and observed from p flavour content, increasing at larger rapidities; relates to how string (and cluster?) fragmentation connects central rapidities to beam remnants
Momentum spectra	dn/dx_E with $x_E = 2E_{\rm had}/E_{\rm cm}$; basic distribution for tuning of ”fragmentation function”	dn/dp_T and dn/dy give basic production kinematics, but the many production channels give less easy interpretation

List of observables (II)

2401.07564

B/D hadron momentum fraction of total E or p_T in a jet, with $x = p_{\mathrm{T}}^{\mathrm{had}} / p_{\mathrm{T}}^{\mathrm{jet}}$, as a test of the fragmentation function combined with almost collinear radiation, suitably for some slices of p_{T} (and in addition with a veto that no other B/D should be inside the jet cone, so as to suppress the gluon splitting contribution)	draw a jet cone in θ around B/D and measure x	draw a jet cone in R around B/D and measure x
B/D hadron multiplicity, as a measure of how often several pairs are produced		
Separation inside B/D pairs, where large separation suggests back-to-back primary production, while small separation suggests gluon splitting	separation in θ	separation both in ϕ and in R, since for primary production $\phi = \pi$ is hallmark with η/y separation less interesting, while gluon splitting means R is small while ϕ and y/η individually are less interesting
Hardness difference within (reasonably hard) pairs, $\Delta = \left(p_{\mathrm{T}}^{\max} - p_{\mathrm{T}}^{\min}\right) / \left(p_{\mathrm{T}}^{\max} + p_{\mathrm{T}}^{\min}\right)$, where for gluon splitting $x^2 + (1-x)^2$ translates to $1 + \Delta^2$	separately for small or large θ	separately for large or small ϕ

EXP front

- Detector:
  - Tracker with large acceptance
  - very good vertexing and flavour tagging capabilities (including light quarks and gluon quarks) and calorimeter
  - For Color reconnection studies: efficient tracking and calorimeter @low momenta/Energy is very important
- Measurements:
  - Jet charge and excellent charge hadron identification
  - Fit a representative set of observables for hadronisation calibration
- Simulated samples:
  - Comprehensive list produced using using different hadronization models and parameters
  - Full-SIM: necessary to understand flavour tagging capabilities

EXP front (II)

- Access to LEP Archived Data
  - Ideally we would like to test developments in TH and EXP fronts using LEP data
    - In principle we can improve existing measurements
  - However, understanding and reprocessing of the LEP data is challenging
  - On going effort to systematically export the LEP (ALEPH) data and software tools to the Key4hep environment
    - This would allow the validation of newer calculations and MC tools with existing data
- [Probably] not within the targeted timeline

Giga-Z, Tera-Z?

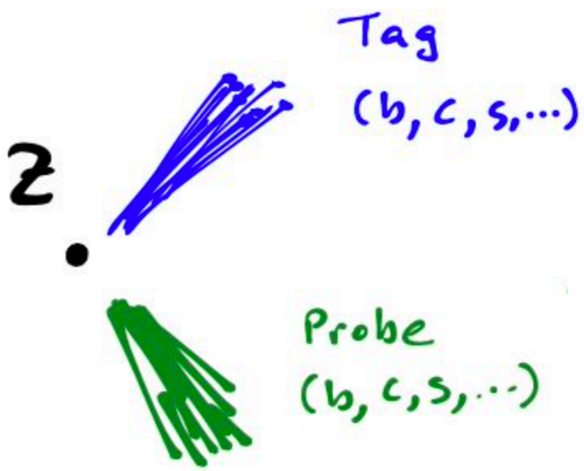
- Short answer – no dedicated studies carried out so far
  - Existing work more generic

- Aims for a future e^+e^- collider:
  - Measure EW observables ~O(100) KeV
  - Measure Higgs couplings ~O(0.1%)

- Next: Back-of-the-envelop calculations assuming Tera-Z stats
  - And latest-greatest taggers

Tera-Z potential

- Strategy: Tag-and-probe @ Z pole
  - First: **Tag** one of the two jets with high purity
  - Then: obtain an **unbiased** sample using the **2nd jet (probe)**



Events @Tera-Z

Z→hadrons	~70%	0.7×10^6 M
→ uu/cc	~12%/flavor	8.4×10^4 M/ flavor
→ dd/ss/bb	~15%/flavor	**1.1×10^5 M/ flavor**

Tera-Z potential (II)

- Taking into account tagging performance

Best case: b-tagging

WP	Eff (b)	Mistag (g)	Mistag (ud)	Mistag (c)
Loose	90%	2%	0.1%	2%
Medium	80%	0.7%	<0.1%	0.3%

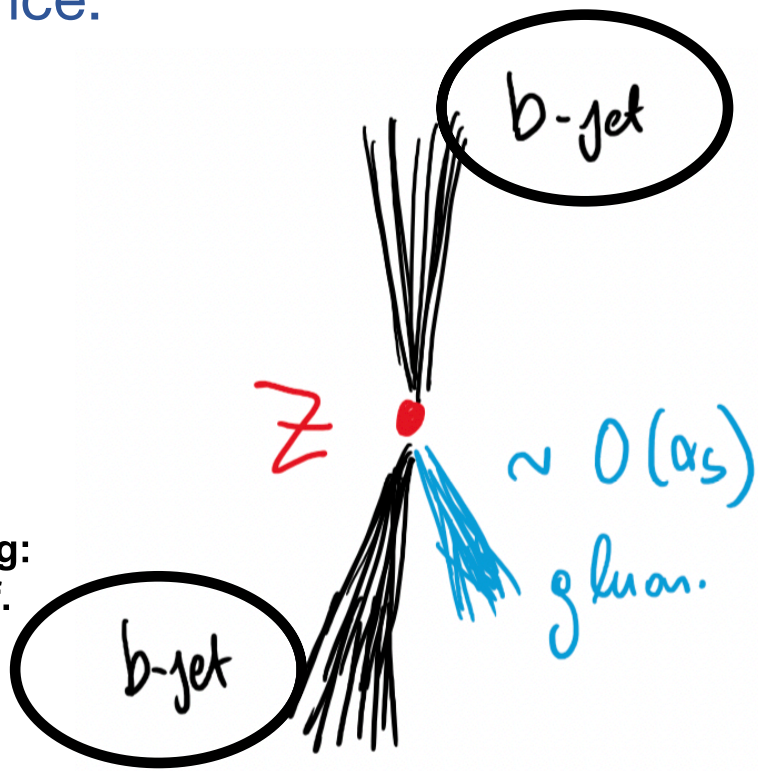
"Worst" case: s-tagging

WP	Eff (s)	Mistag (g)	Mistag (ud)	Mistag (c)	Mistag (b)
Loose	90%	20%	40%	10%	1%
Medium	80%	9%	20%	6%	0.4%

- Event/jet samples:
  - **bottom jets:** >~10^5 M, **strange jets:** >~10^4 M
    - all other jet flavors in between
- **Much larger** than LEP dataset
  - MC tuning, calibration etc.. with unprecedented precision

Tera-Z potential (III)

- Challenging… topic of discussion and brainstorming
  - For instance:



b-tagging: ~90% eff.

b-tagging: ~90% eff.

- 3rd jet is a gluon: O(1-10%) depending on momentum, angle
- Still more than enough for developing the gluon-category

Tera-Z: Huge potential to exhaustively study b/c fragmentation/hadronization and gluon splitting

Towards the Report

- The topic is "open-ended"
  - Also: schedule accelerated by 1y does not help to conclude many of these studies
- Summarize: Open questions and future directions
  - Theory
    - Non-universal fragmentation function; parton flavor, process..?
    - Disentangle the perturbative and non-perturbative components
    - New observables and multi-particle correlations
  - Experiment
    - Detector design
    - Z-pole, How many Z's? Giga-Z, Tera-Z?
    - Access to LEP data under Key4Hep
    - Analysis techniques; more data-driven?
- Lot's of room for innovation and exciting+important work to be done
  - email: ECFA-WHF-FT-BCFRAG@cern.ch