SOLVING THE INVERSE PROBLEM FOR HADRONIZATION

Jure Zupan
U. of Cincinnati

MLHAD

based on 2203.04983, 2308.13459, 2311.09296, 2410.nnnnn, ...;
in collaboration with Christian Bierlich, Phil Ilten, Tony Menzo, Steve Mrenna, Manuel Szewc, Michael K. Wilkinson, Ahmed Youssef

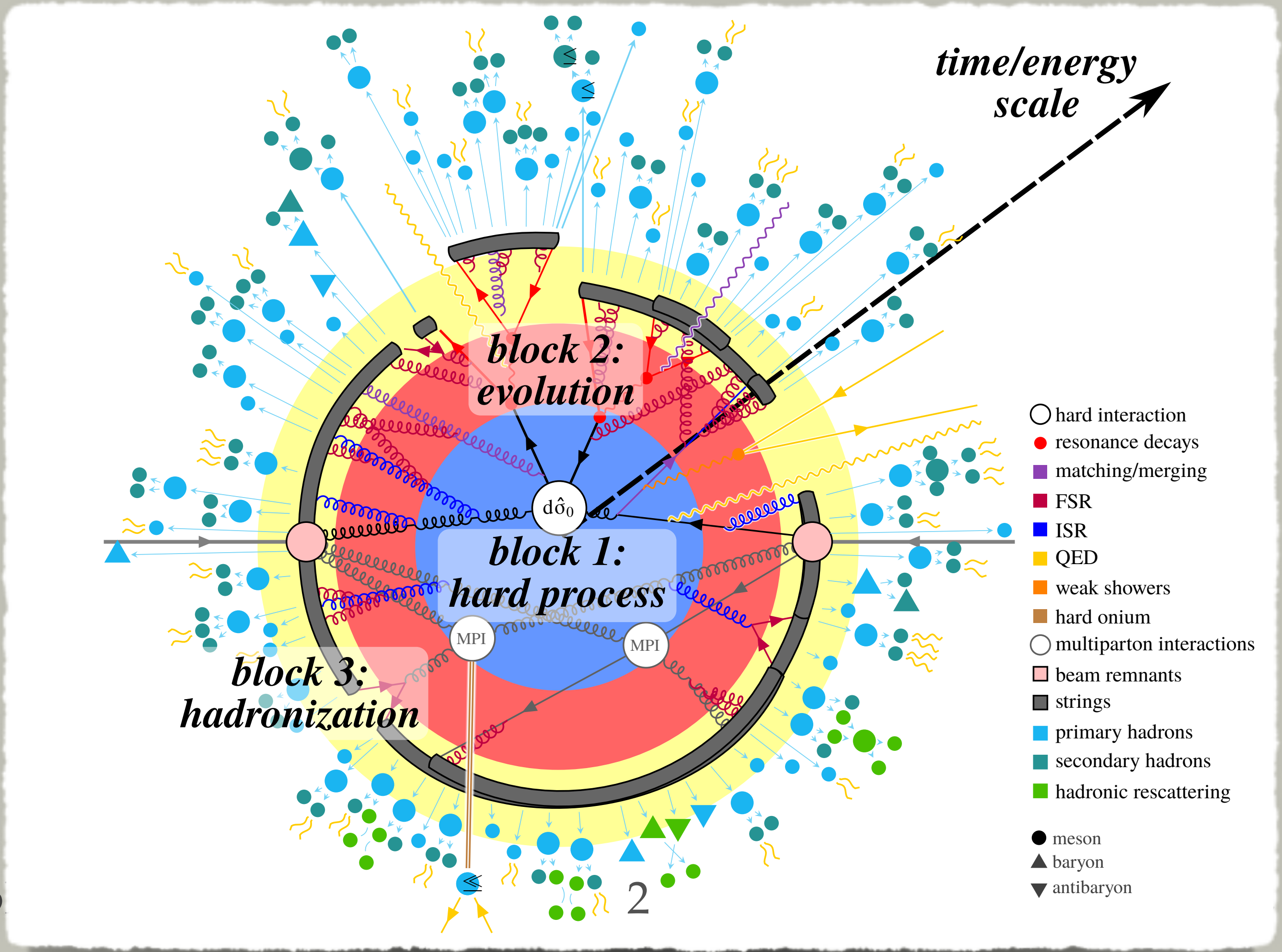
Belica, Oct 3 2024

MONTE CARLO HEP EVENT

- block structure of HEP Monte Carlo
  - hard process
  - shower/evolution
  - hadronization
  - (detector simulation)

} under good perturbative control and systematically improvable (hard process, shower/evolution)

} modeling of nonperturbative physics (hadronization)



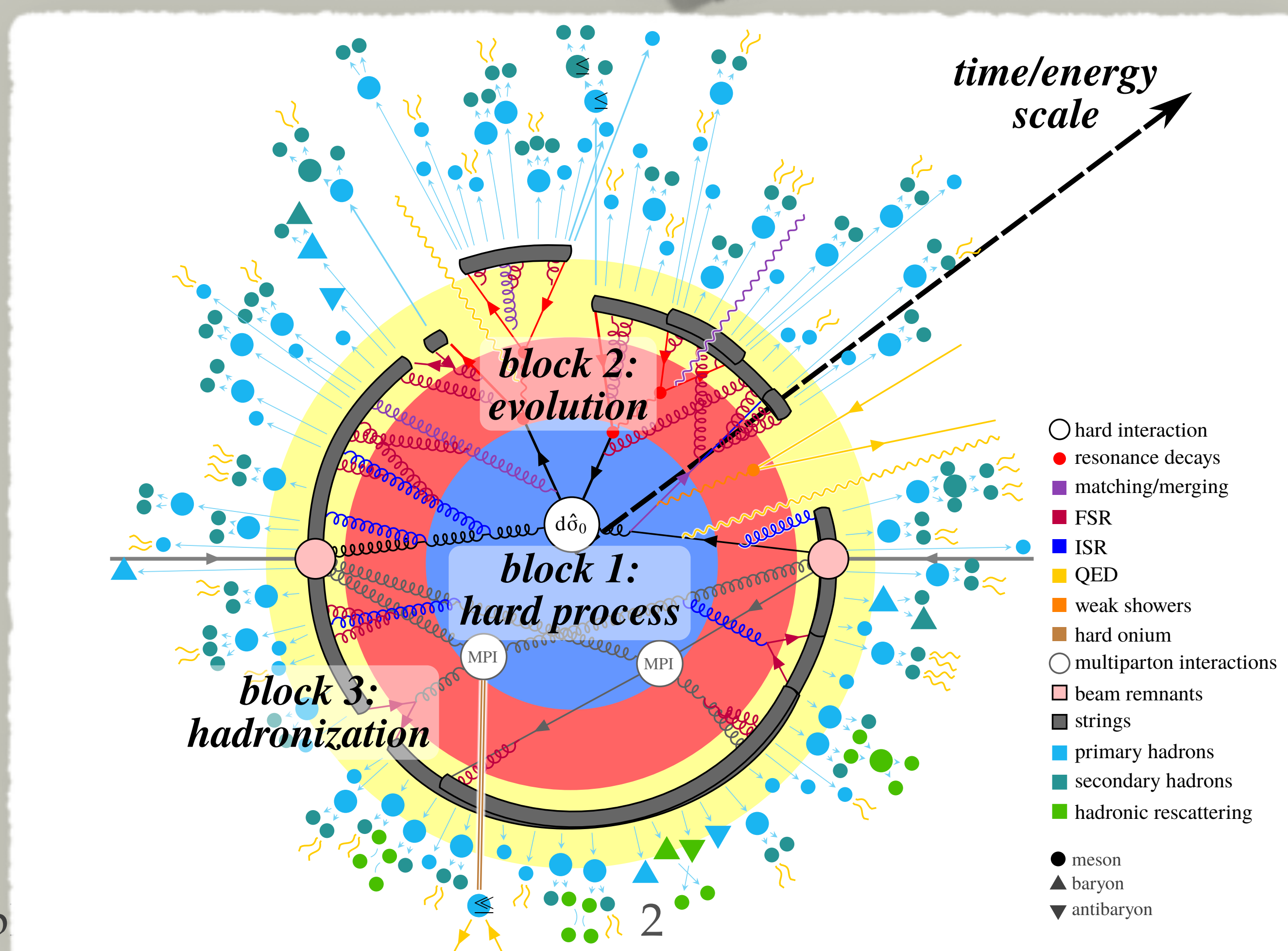
MONTE CARLO HEP EVENT

- block structure of HEP Monte Carlo
  - hard process
  - shower/evolution
  - hadronization
  - (detector simulation)

} under good perturbative control and systematically improvable (hard process, shower/evolution)

} modeling of n… …tive physics (hadronization)

use Machine Learning

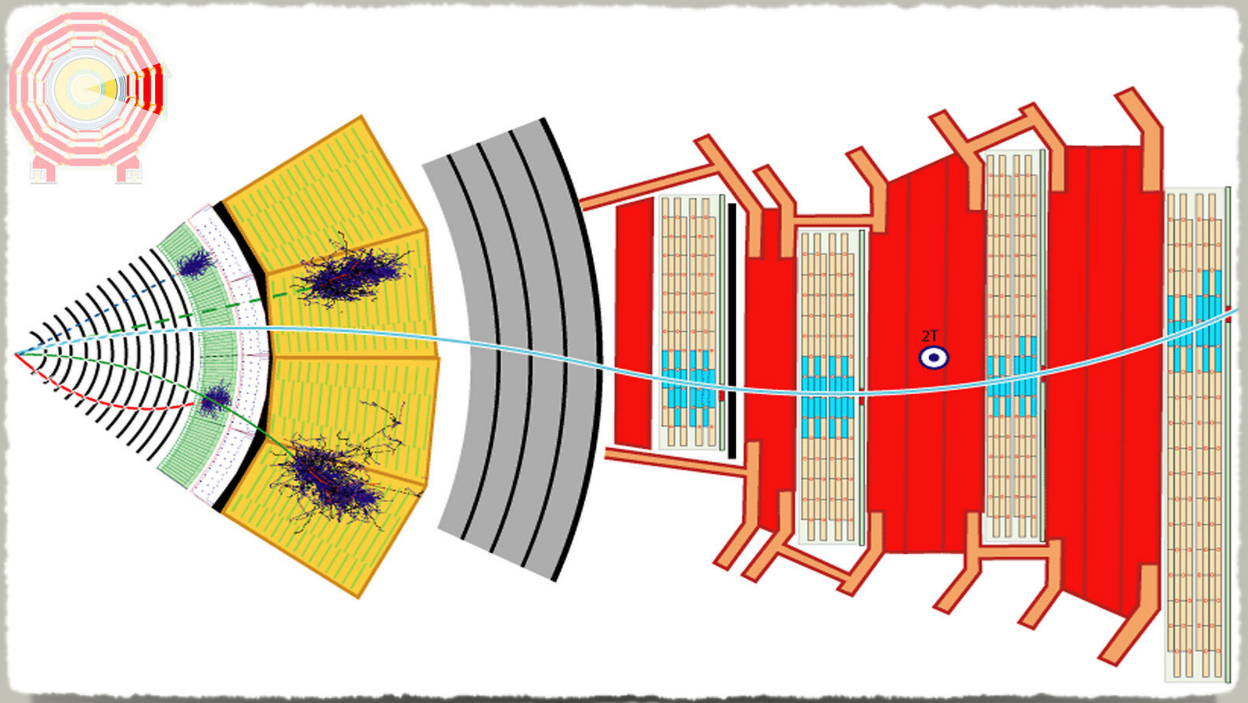


OUTLINE

- should one care about hadronization?
- hadronization models
- ML hadronization (`Mlhad`/`HadML`)
- results of immediate relevance to `Pythia`

SHOULD ONE CARE ABOUT HADRONIZATION?

- if observables/measurements inclusive enough no need for modeling hadronization
  - for the past ~50 years observables have been constructed explicitly to remove any depend. on hadronization
- not the situation in the real world
  - experimental cuts, detectors not perfect, resonances decay in different ways
  - modeling well hadronization step essential for precision studies
- some measurements more sensitive than others
  - e.g., number of charged particles, correlations between exclusive states, etc.



SHOULD ONE CARE ABOUT HADRONIZATION

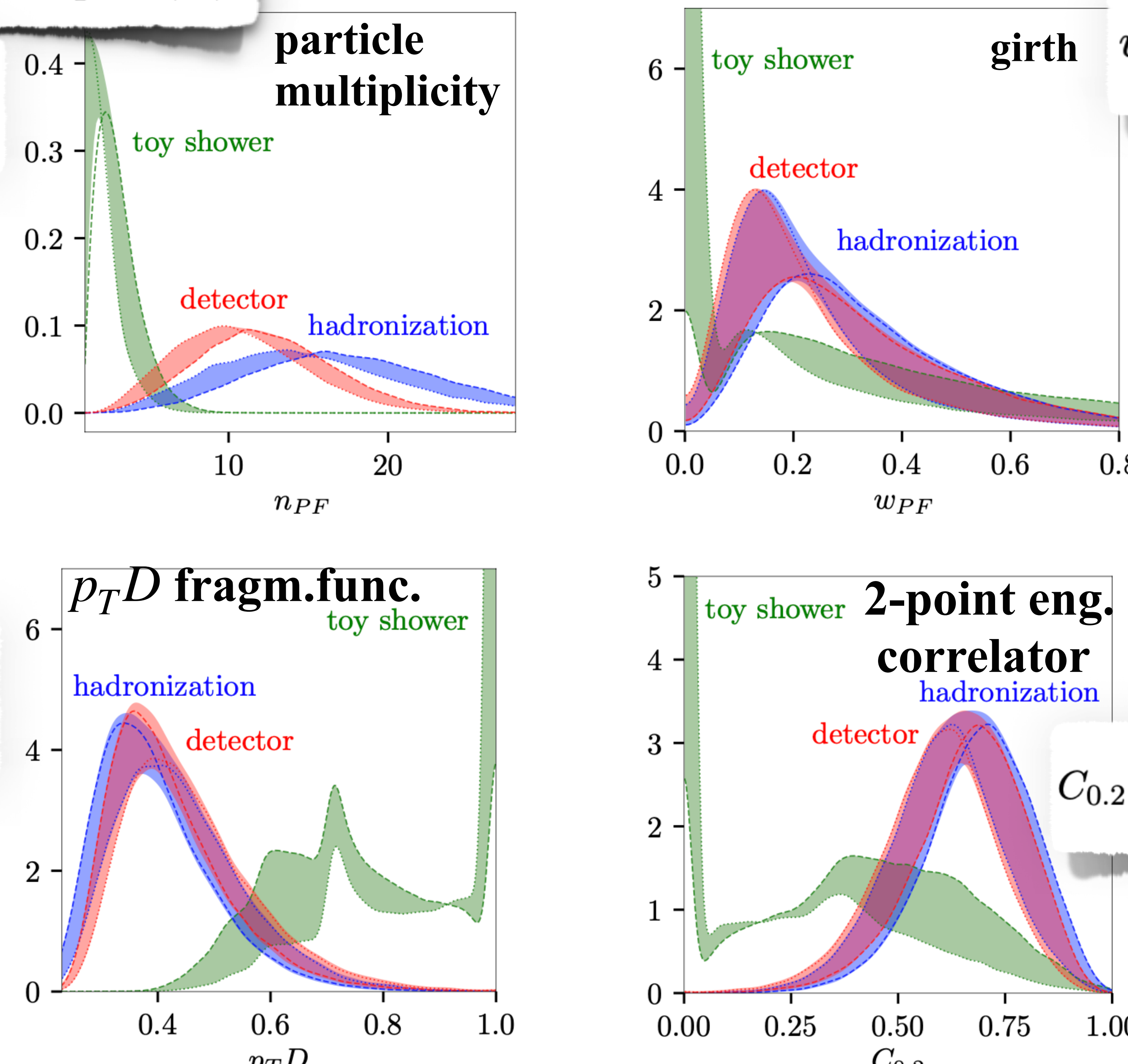
$e^+e^- \to q\bar{q}$ with $q = u, d, s$

$$n_{\rm PF} = \sum_i 1$$

$$w_{\rm PF} = \frac{\sum_i p_{T,i} \Delta R_{i,\rm jet}}{\sum_i p_{T,i}}$$

$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$$

$$C_{0.2} = \frac{\sum_{ij} E_{T,i} E_{T,j} (\Delta R_{ij})^{0.2}}{\sum_i E_{T,i}^2}$$



Bieringer et al, 2012.09873

SHOULD ONE CARE ABOUT HADRONIZATION

- understanding hadronization for precision measurements
- top mass:

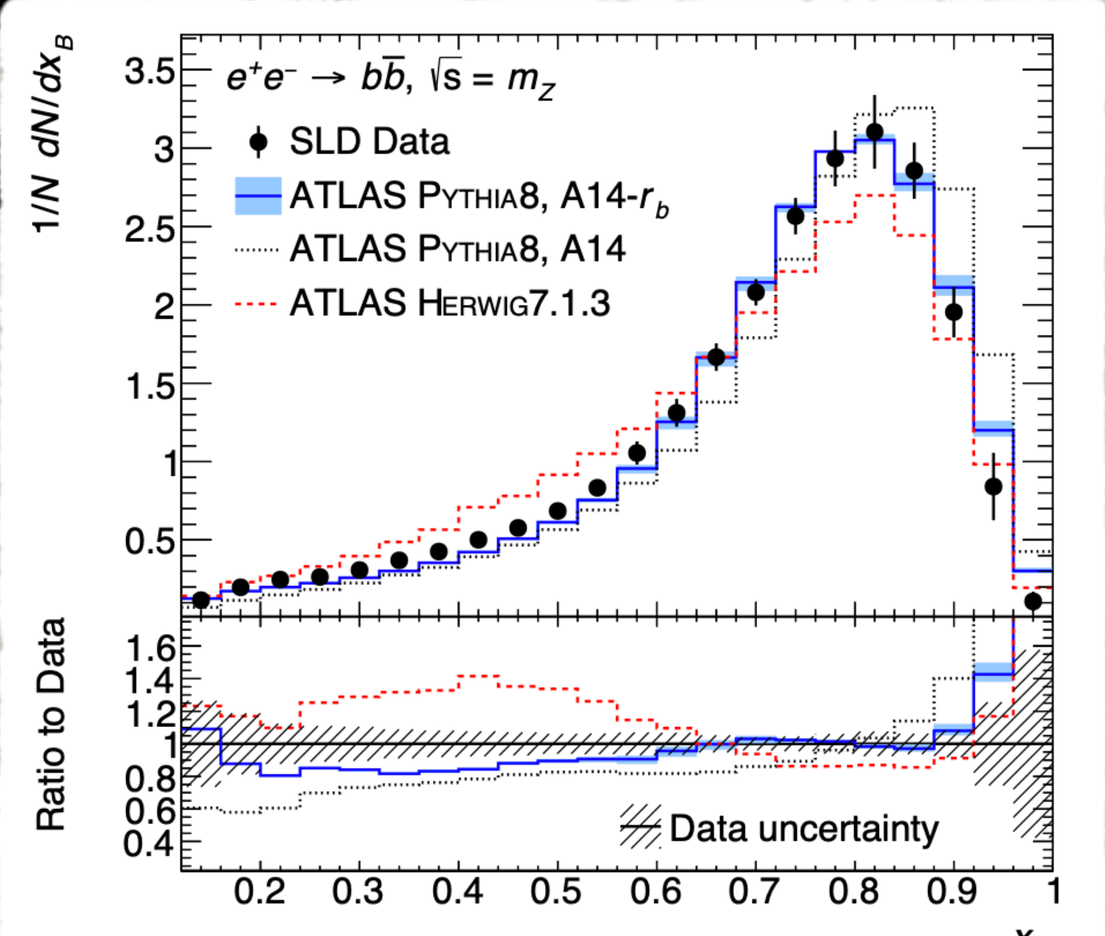
pdglive.lbl.gov

VALUE (GeV)		DOCUMENT ID		TECN	COMMENT
172.69 ± 0.30	**OUR AVERAGE** Error includes scale factor of 1.3. See the ideogram below.				
$172.13\ ^{+0.76}_{-0.77}$		[1] TUMASYAN	2021G	CMS	t-channel single top production
172.6 ± 2.5		[2] SIRUNYAN	2020AR	CMS	jet mass from boosted top
$172.69 \pm 0.25 \pm 0.41$		[3] AABOUD	2019AC	ATLS	7, 8 TeV ATLAS combination
$172.26 \pm 0.07 \pm 0.61$		[4] SIRUNYAN	2019AP	CMS	lepton+jets, all-jets channels
$172.33 \pm 0.14\ ^{+0.66}_{-0.72}$		[5] SIRUNYAN	2019AR	CMS	dilepton channel ($e\mu$, $2e$, 2μ)
$172.44 \pm 0.13 \pm 0.47$		[6] KHACHATRYA..	2016AK	CMS	7, 8 TeV CMS combination
$174.30 \pm 0.35 \pm 0.54$		[7] TEVEWWG	2016	TEVA	Tevatron combination

- *b* quark hadronization model uncert. on indiv. measurements ~0.1-0.2 GeV

SHOULD ONE CARE ABOUT HADRO

- understanding hadronization for precision measurements
- top mass:



ATLAS, 2209.00583

pdglive.lbl.gov

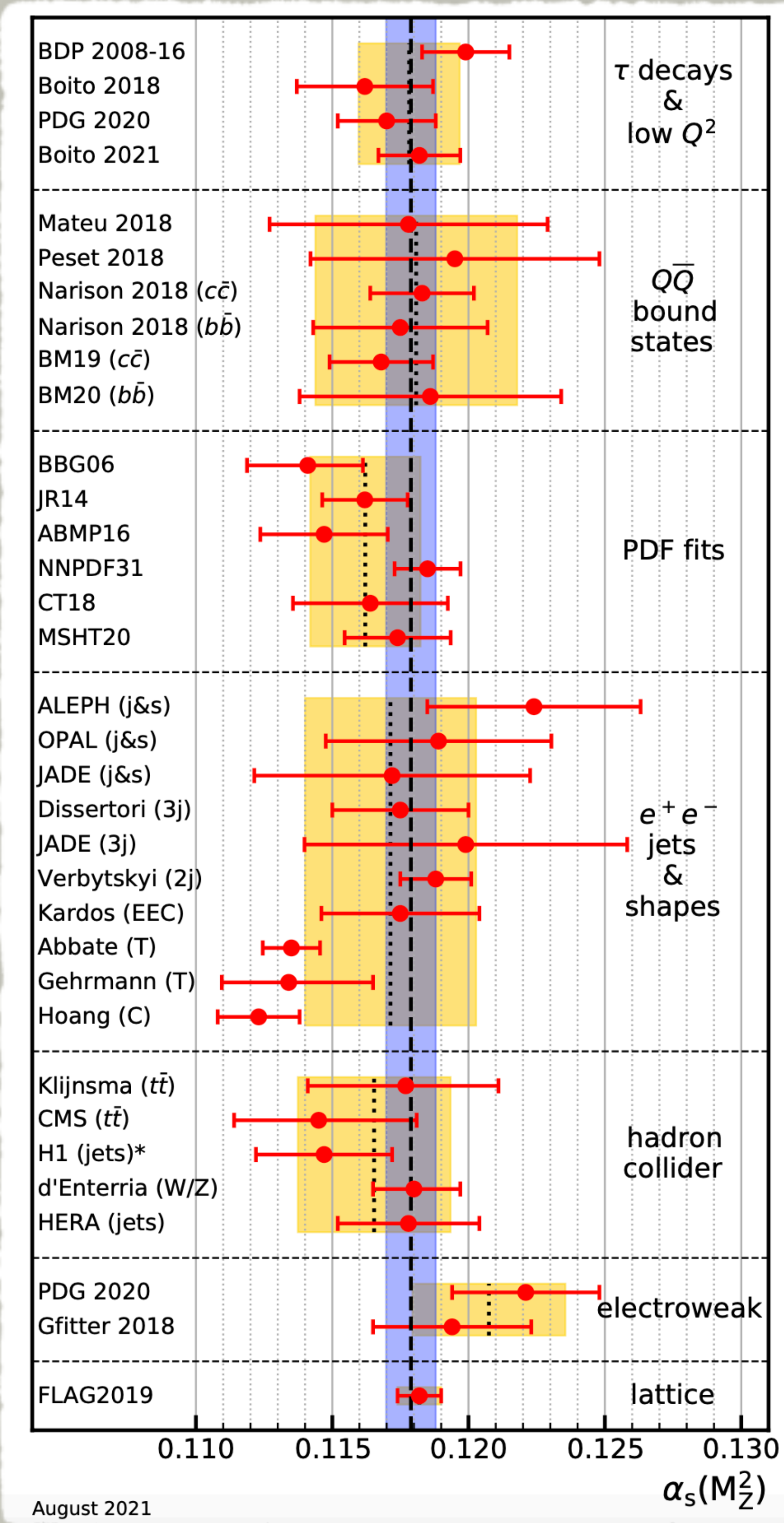
VALUE (GeV)	*DOCUMENT ID*			
172.69 ± 0.30	OUR AVERAGE Error includes scale factor of 1.3. See the			
$172.13^{+0.76}_{-0.77}$	[1] TUMASYAN	2021C		
172.6 ±2.5	[2] SIRUNYAN	2020AR	CMS	jet mass from boosted top
172.69 ±0.25 ±0.41	[3] AABOUD	2019AC	ATLS	7, 8 TeV ATLAS combination
172.26 ±0.07 ±0.61	[4] SIRUNYAN	2019AP	CMS	lepton+jets, all-jets channels
$172.33 \pm 0.14 ^{+0.66}_{-0.72}$	[5] SIRUNYAN	2019AR	CMS	dilepton channel ($e\mu$, $2e$, 2μ)
172.44 ±0.13 ±0.47	[6] KHACHATRYA..	2016AK	CMS	7, 8 TeV CMS combination
174.30 ±0.35 ±0.54	[7] TEVEWWG	2016	TEVA	Tevatron combination

- *b* quark hadronization model uncert. on indiv. measurements ~0.1-0.2 GeV

SHOULD ONE CARE ABOUT HADRONIZATION

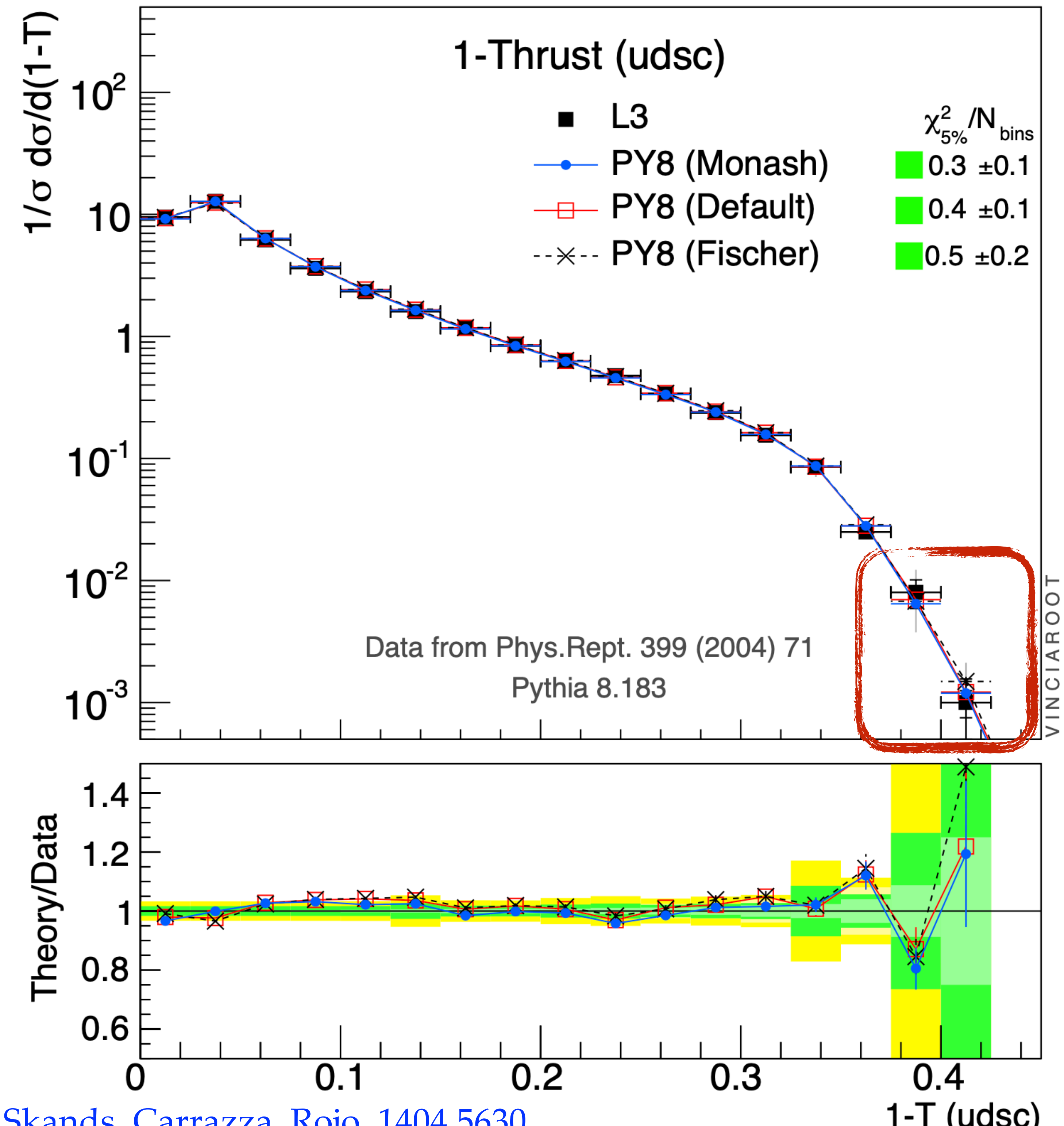
- understanding hadronization for precision measurements
- $\alpha_S(M_Z)$ determinations

hadronization corr. scale as $\sim\Lambda/Q$, can be dominant uncertainties



SHOULD ONE CARE ABOUT HADRONIZATION

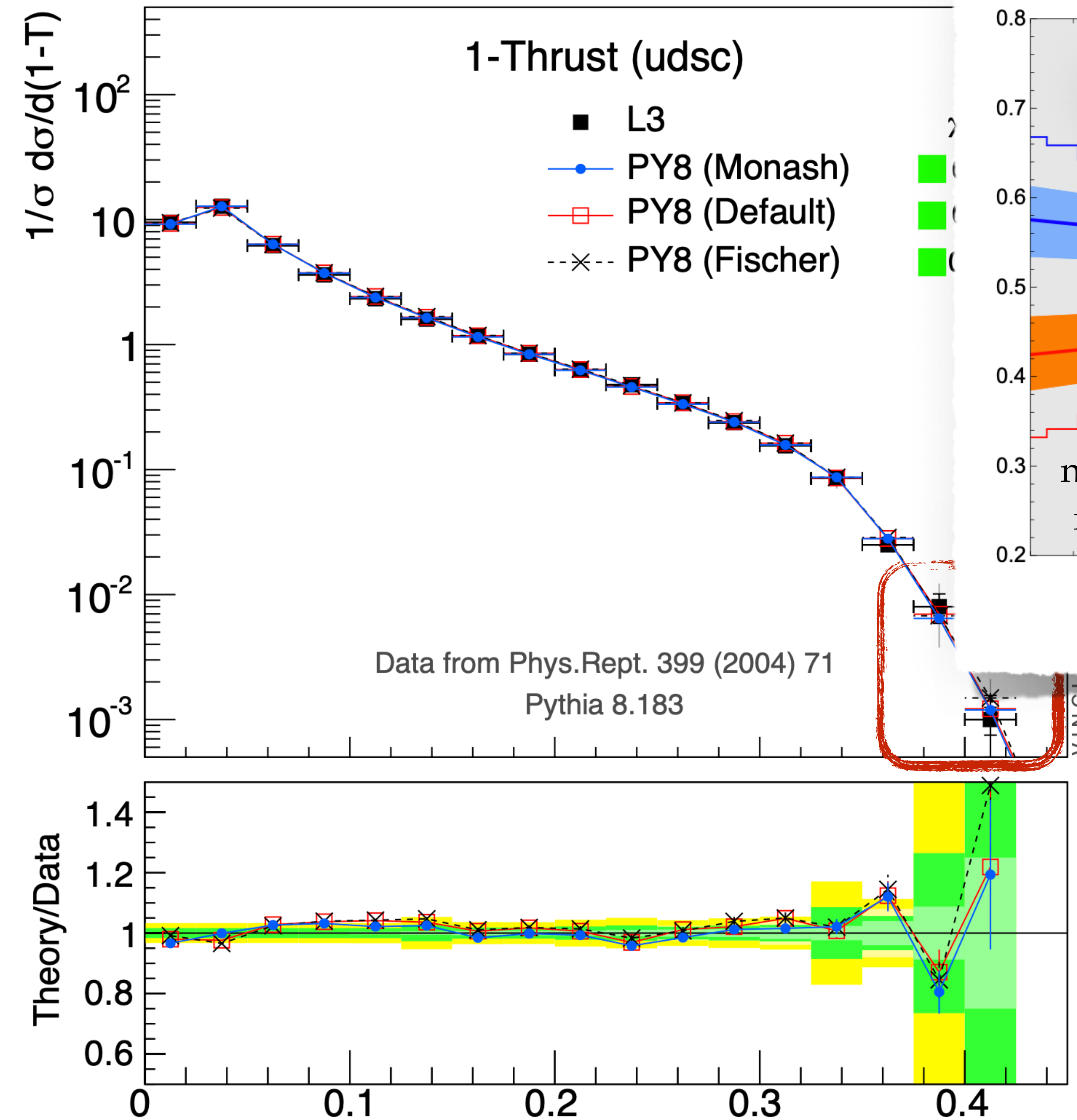
see 2203.11110 for more examples



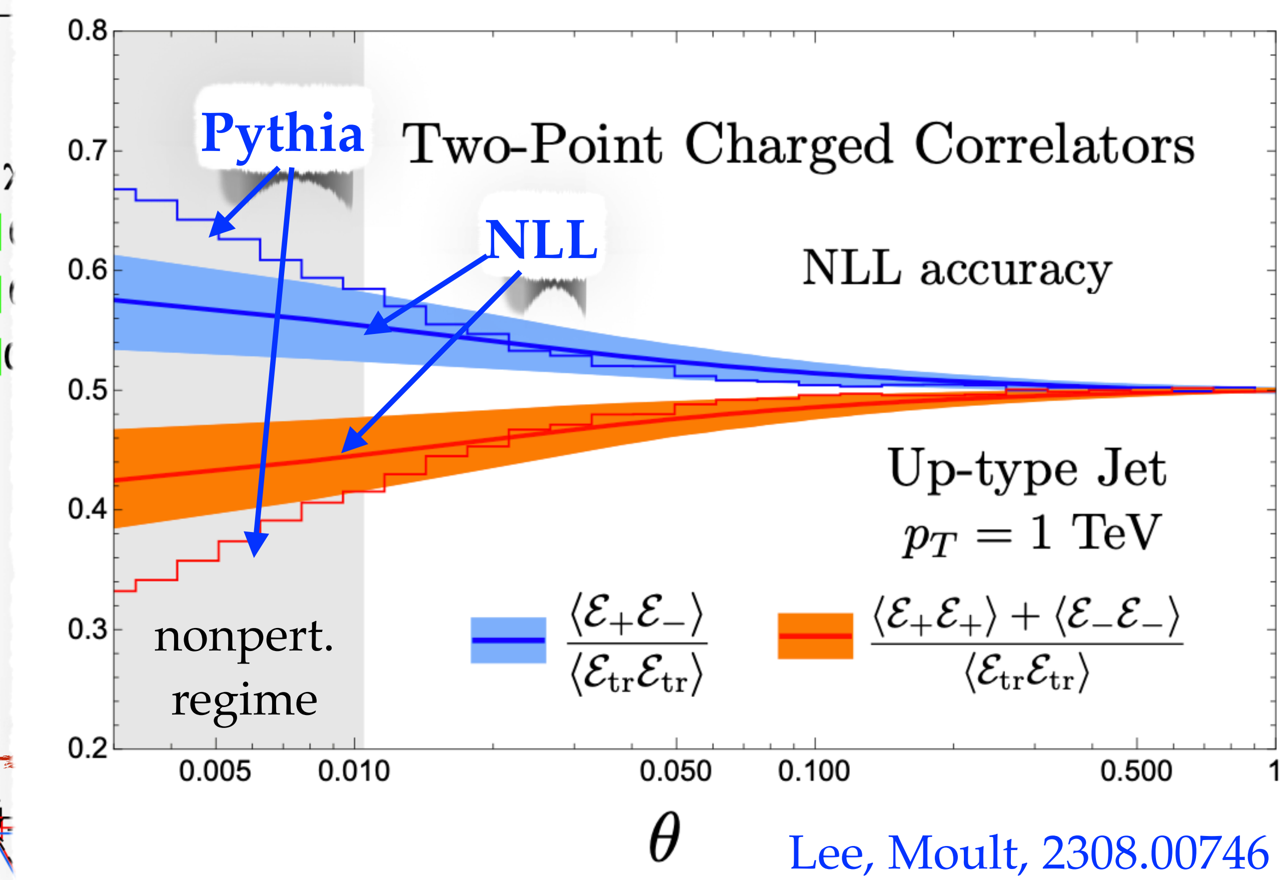
Skands, Carrazza, Rojo, 1404.5630

SHOULD ONE CARE ABOUT HADRONIZATION

see 2203.11110 for more examples



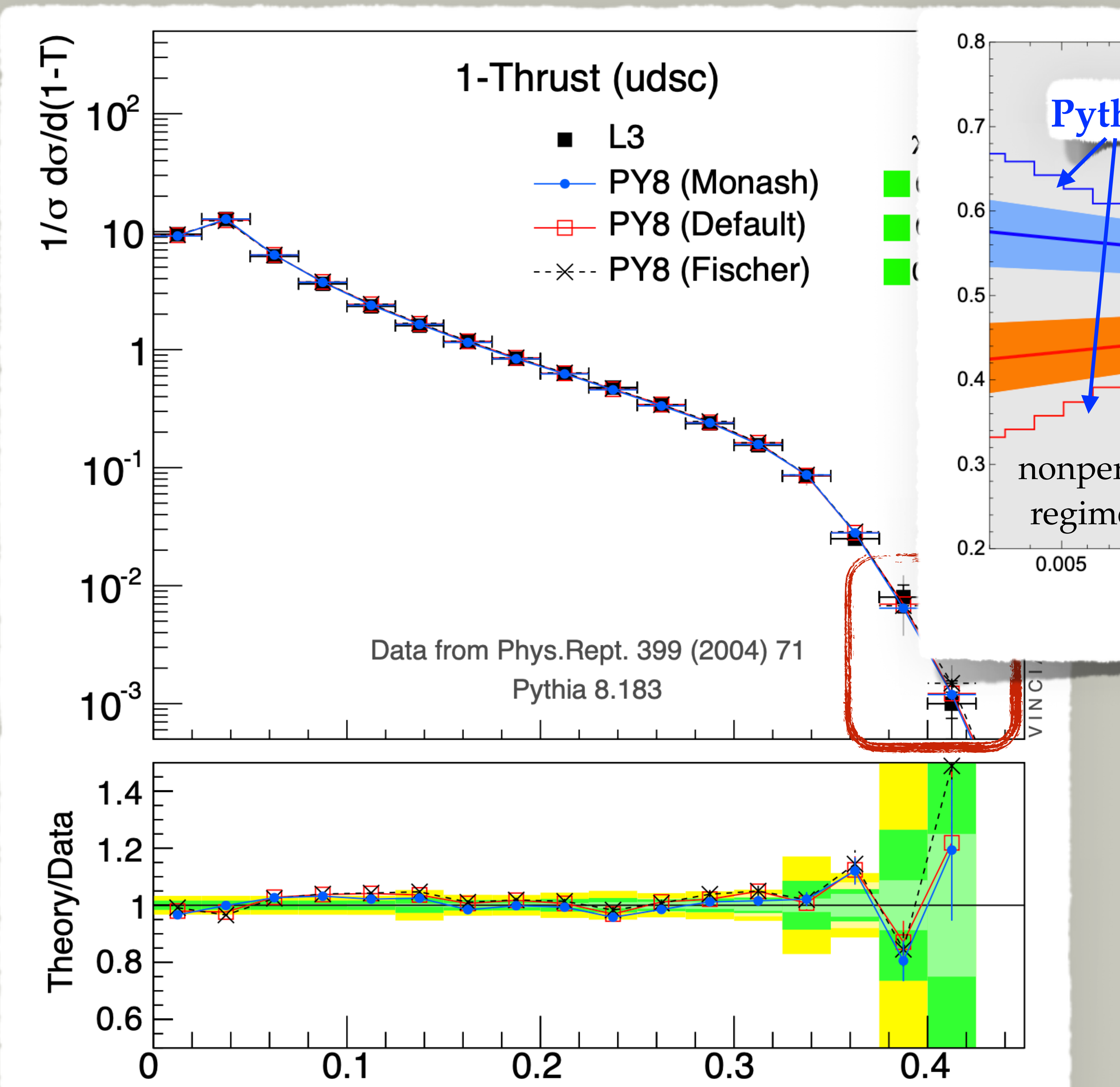
Skands, Carrazza, Rojo, 1404.5630



Lee, Moult, 2308.00746

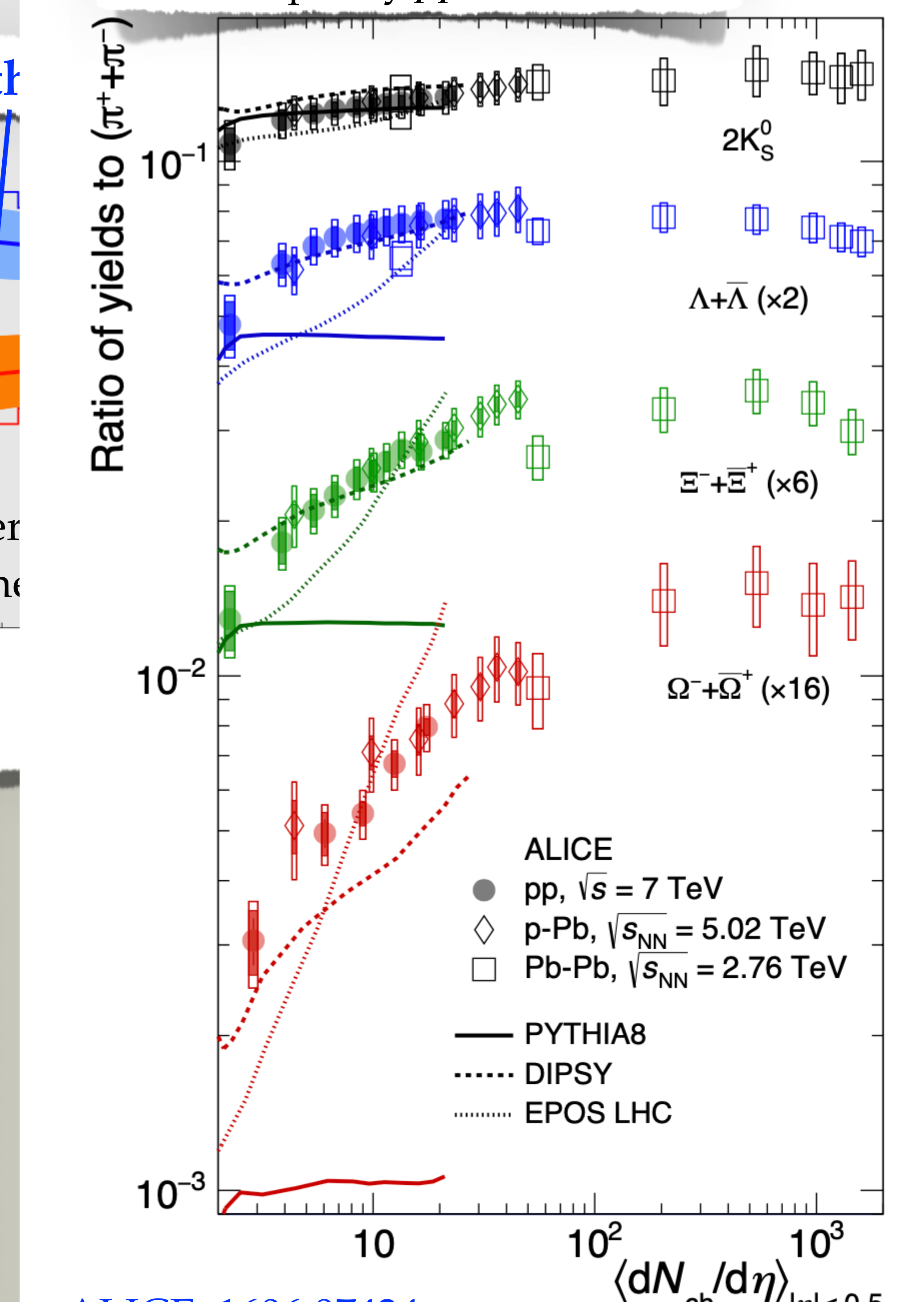
SHOULD ONE CARE ABOUT HADRONIZATION

see 2203.11110
for more examples



Skands, Carrazza, Rojo, 1404.5630

strangeness enhancement in high multiplicity *pp* collisions



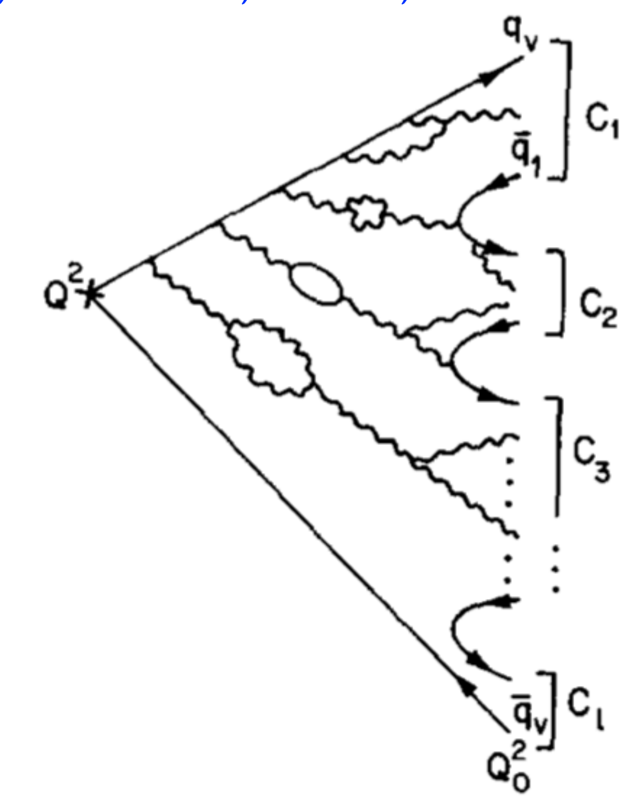
ALICE, 1606.07424

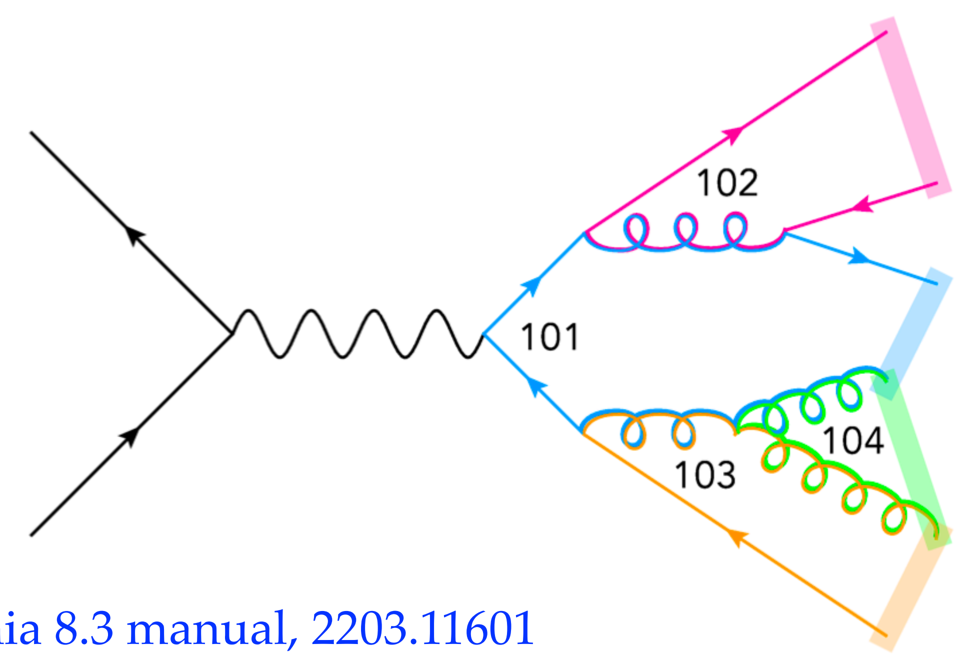
HADRONIZATION MODELS

HADRONIZATION

- two main models for hadronization
  - Lund string model (Pythia)
  - cluster hadronization model (Herwig)
- both have as a starting point stage the final stage of QCD shower
  - stop shower at some scale Q_0
  - in large $N_c \to \infty$ limit planar graphs
  - groups final $q, \bar{q}, g$ in QCD singlet clusters/string pieces
  - either "color preconfinement" or "leading color dipoles/strings"

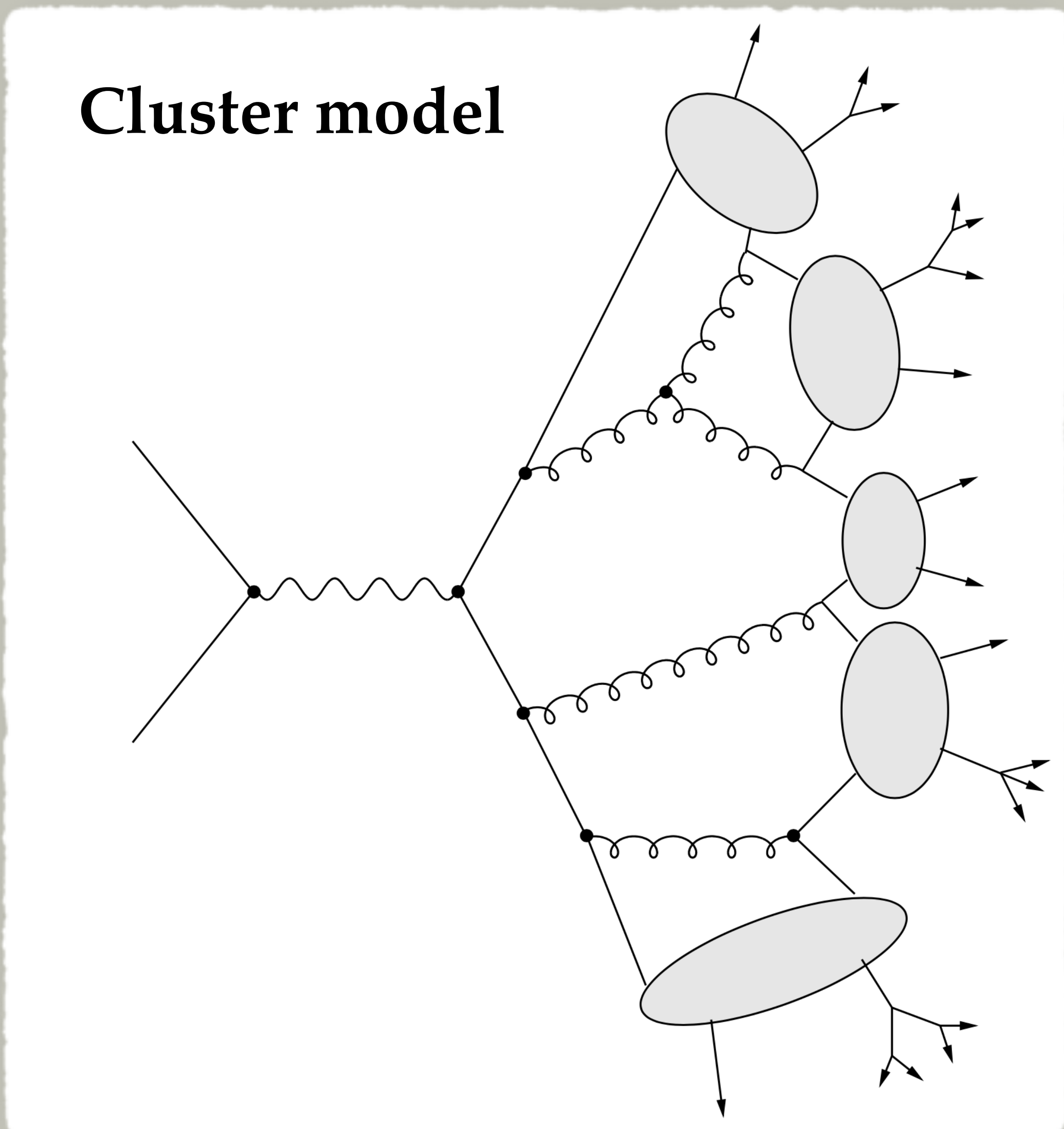
Amati, Veneziano, PLB83, 1979

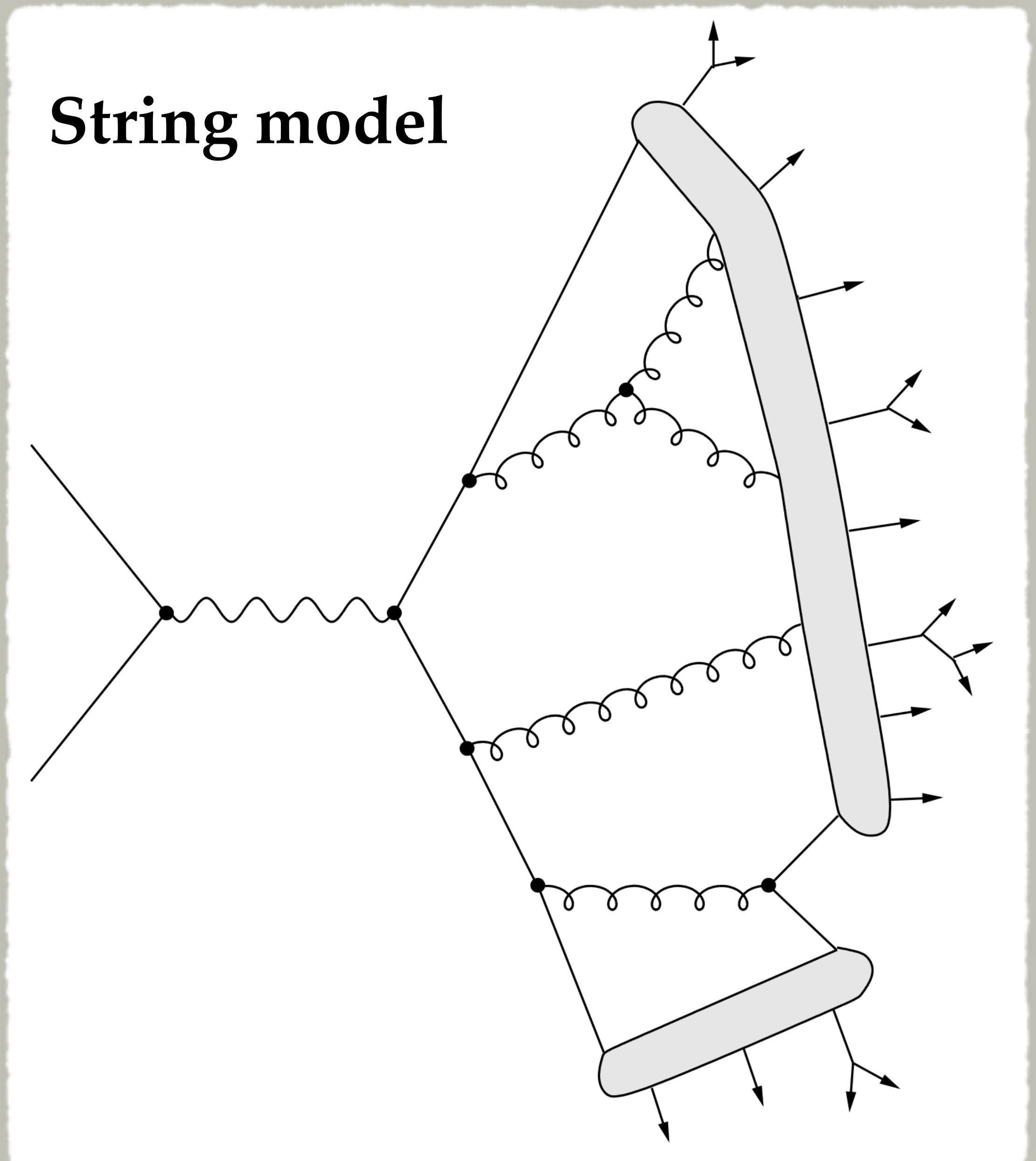




Pyhia 8.3 manual, 2203.11601

CLUSTER VS. STRING MODEL

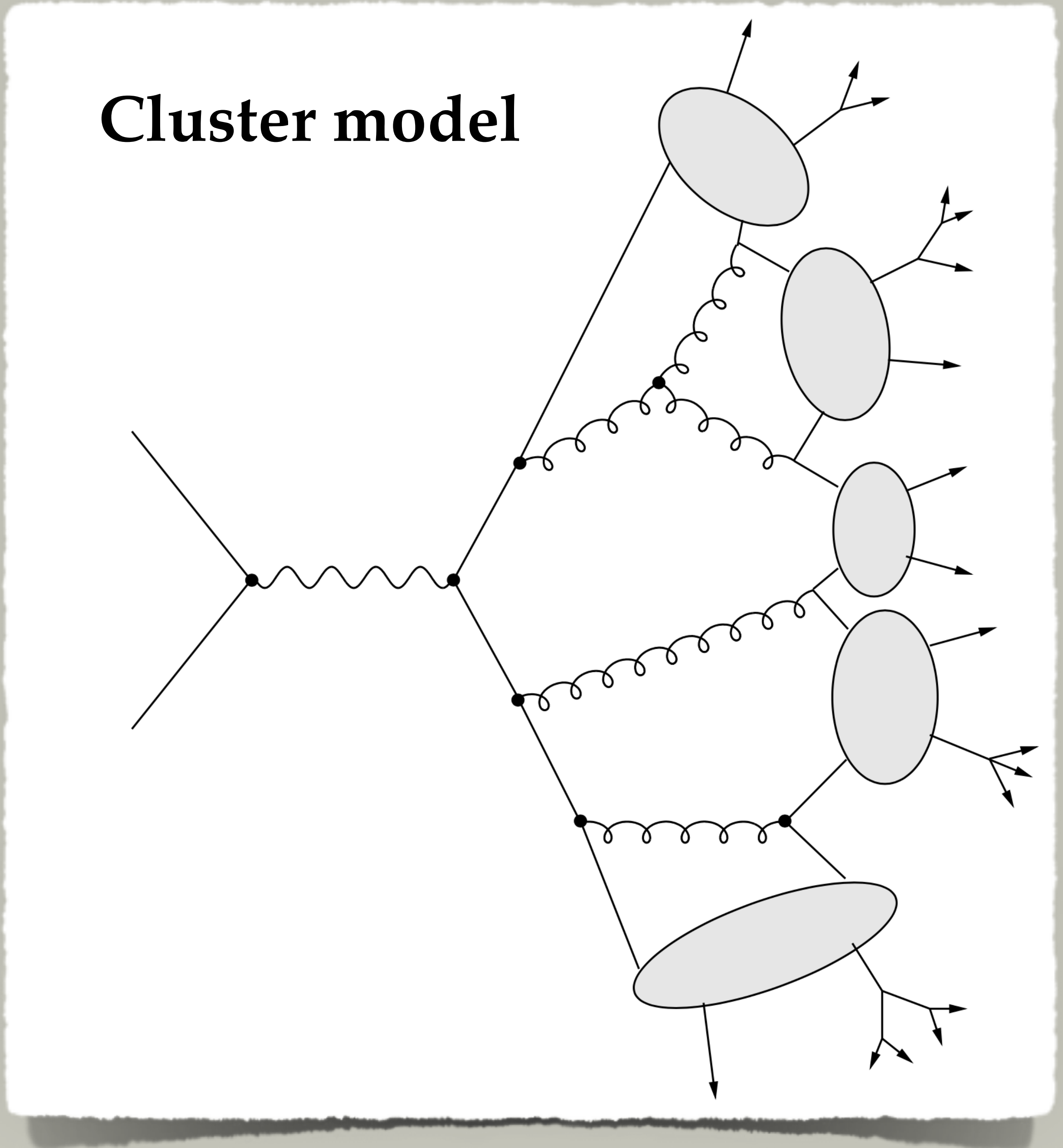


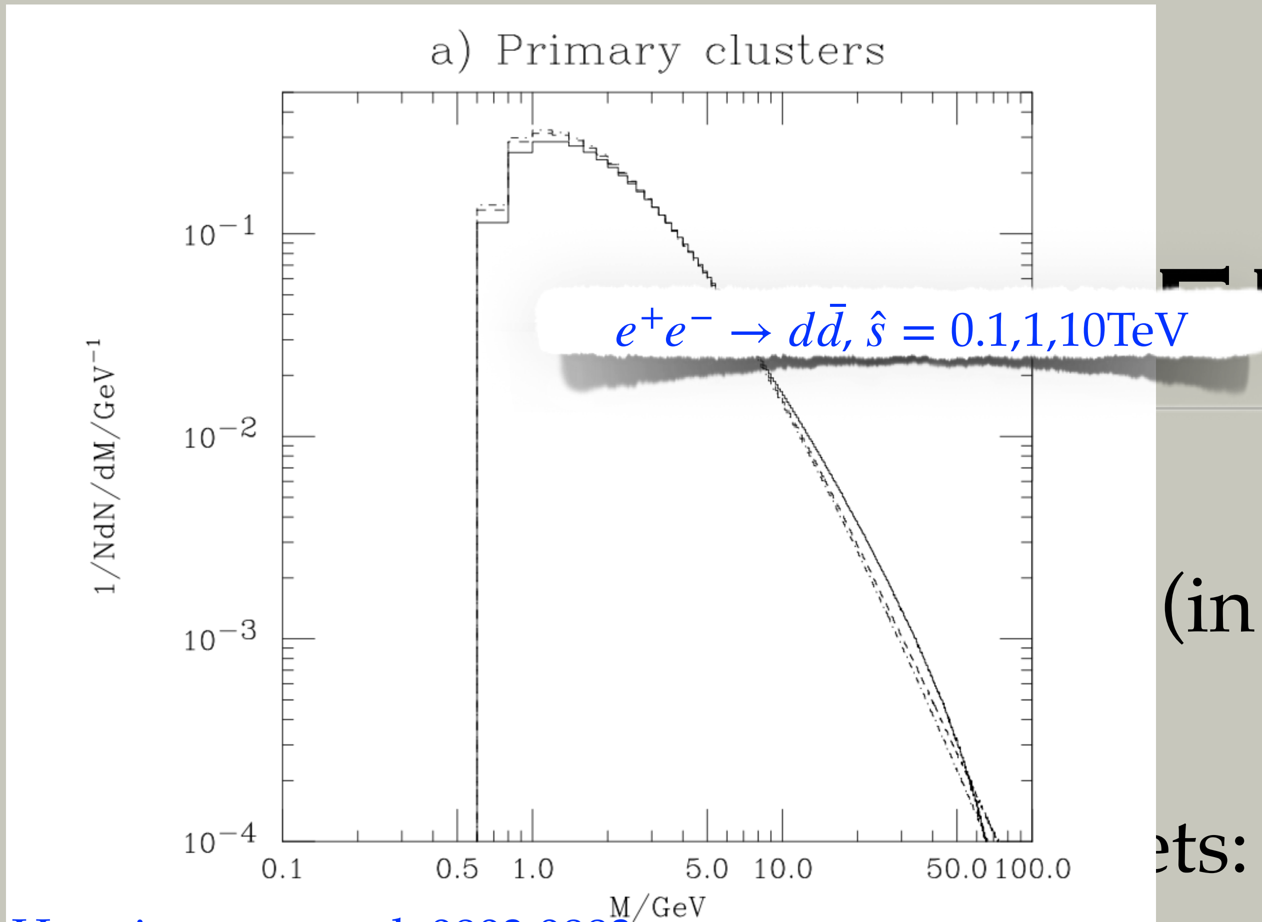


Figs from Webber, 1999

CLUSTER MODEL

- assign mass to gluons, decay them to $q\bar{q}$ pairs (in large N_c limit)
  - these are color singlets: *primary clusters*
  - primary clusters have universal mass distrib
- heavier clusters are decayed to lighter ones (fission/model dep. step)
- relatively small set of params, $\mathcal{O}(30)$





Herwig++ manual, 0803.0883

ER MODEL

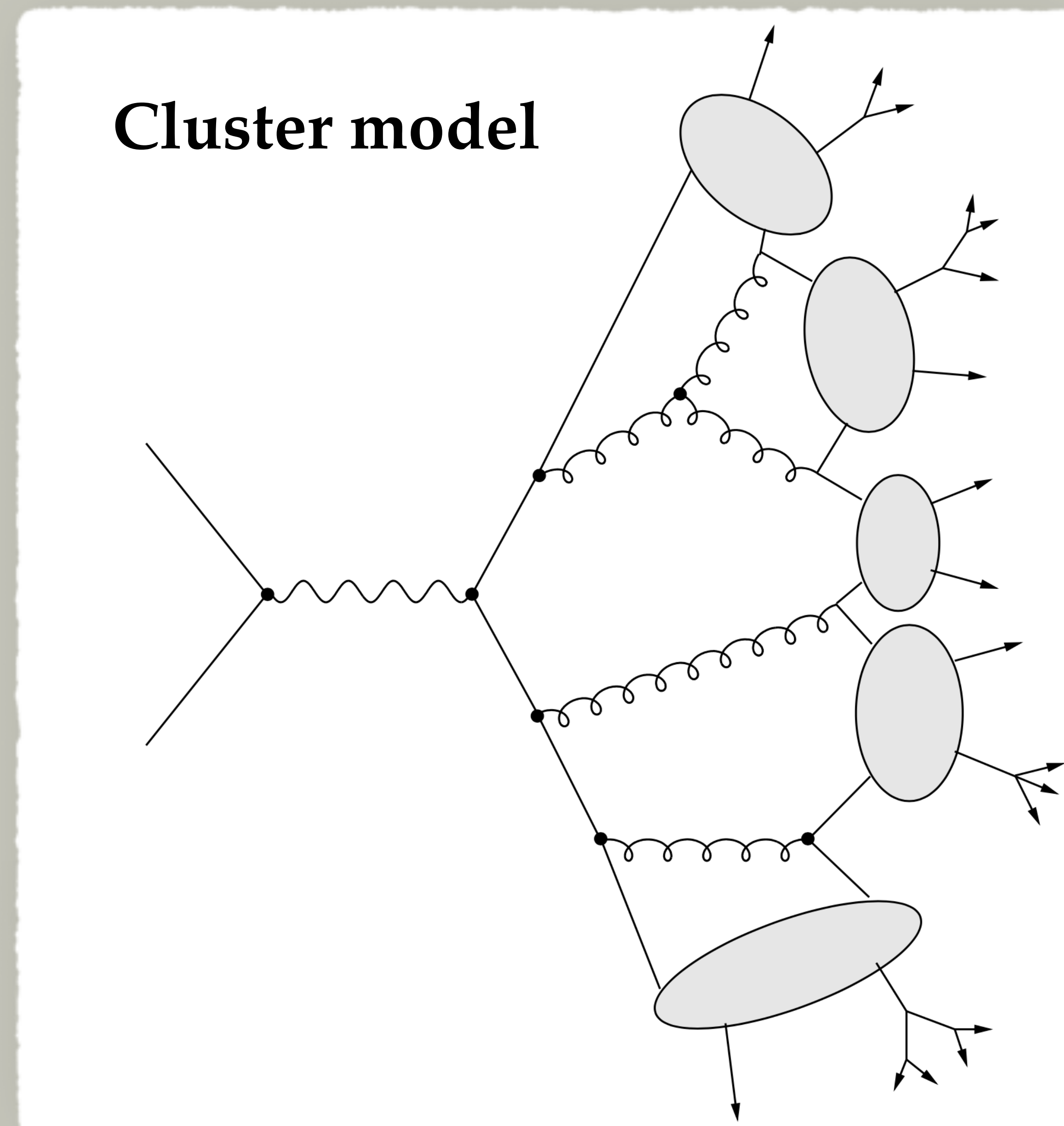
(in

ets:

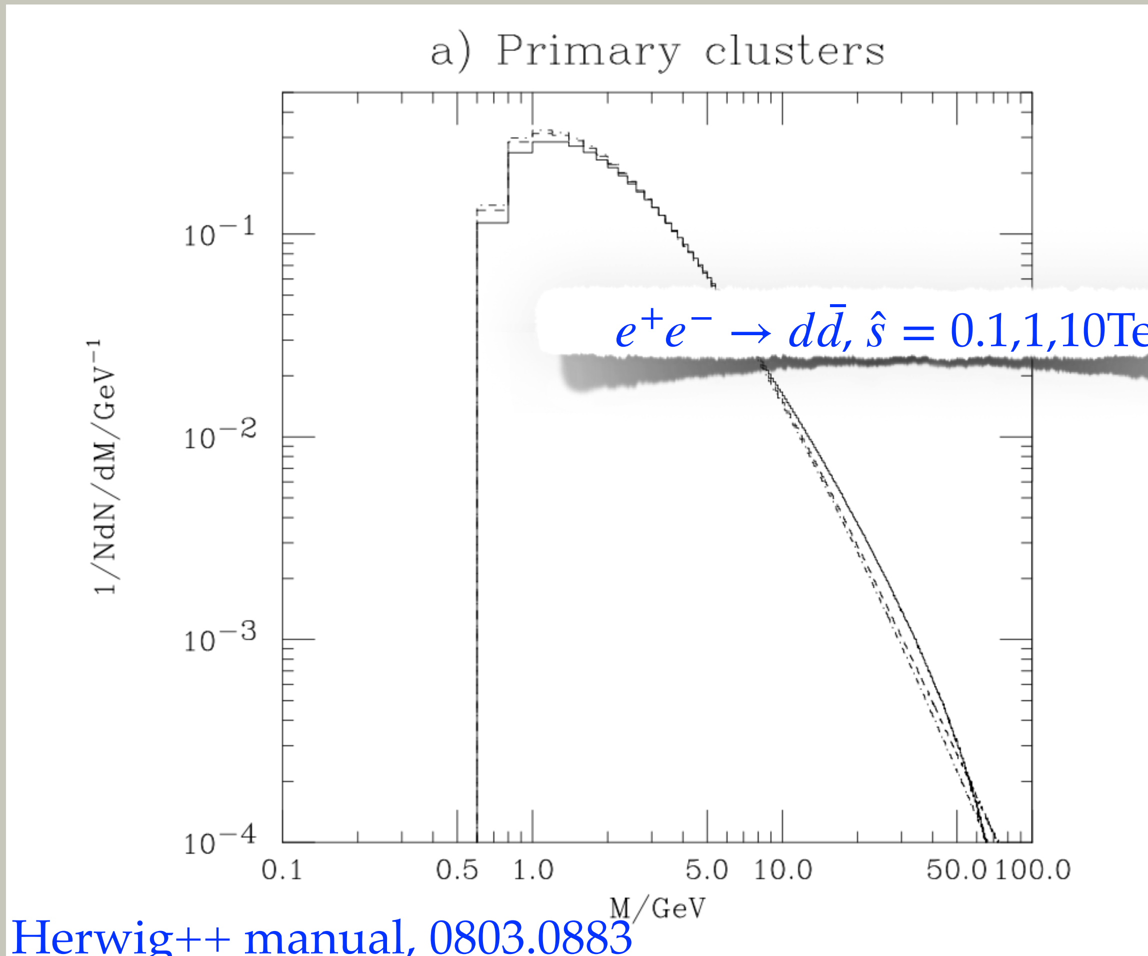
primary clusters

- primary clusters have universal mass distrib

- heavier clusters are decayed to lighter ones (fission/model dep. step)
- relatively small set of params, $\mathcal{O}(30)$



ER MODEL



(in

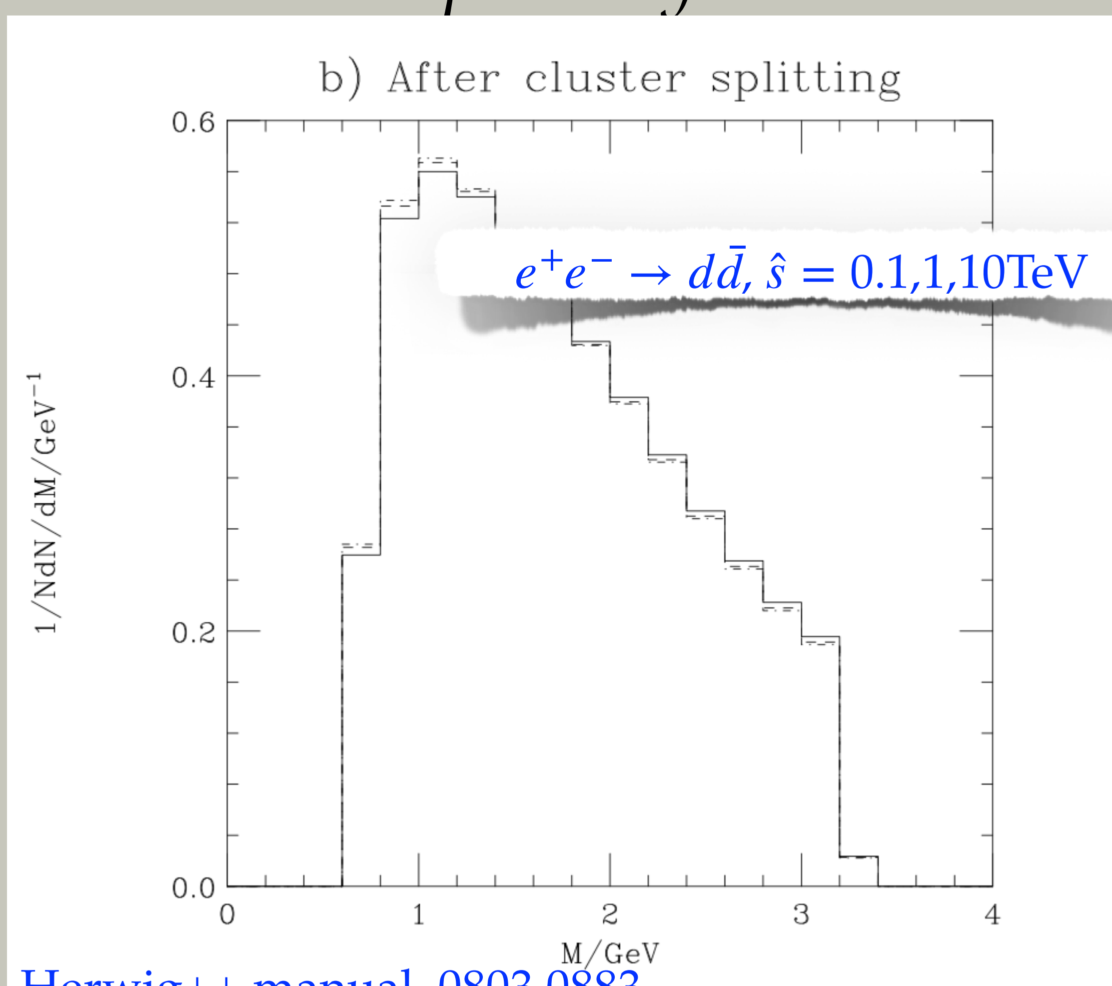
ets:

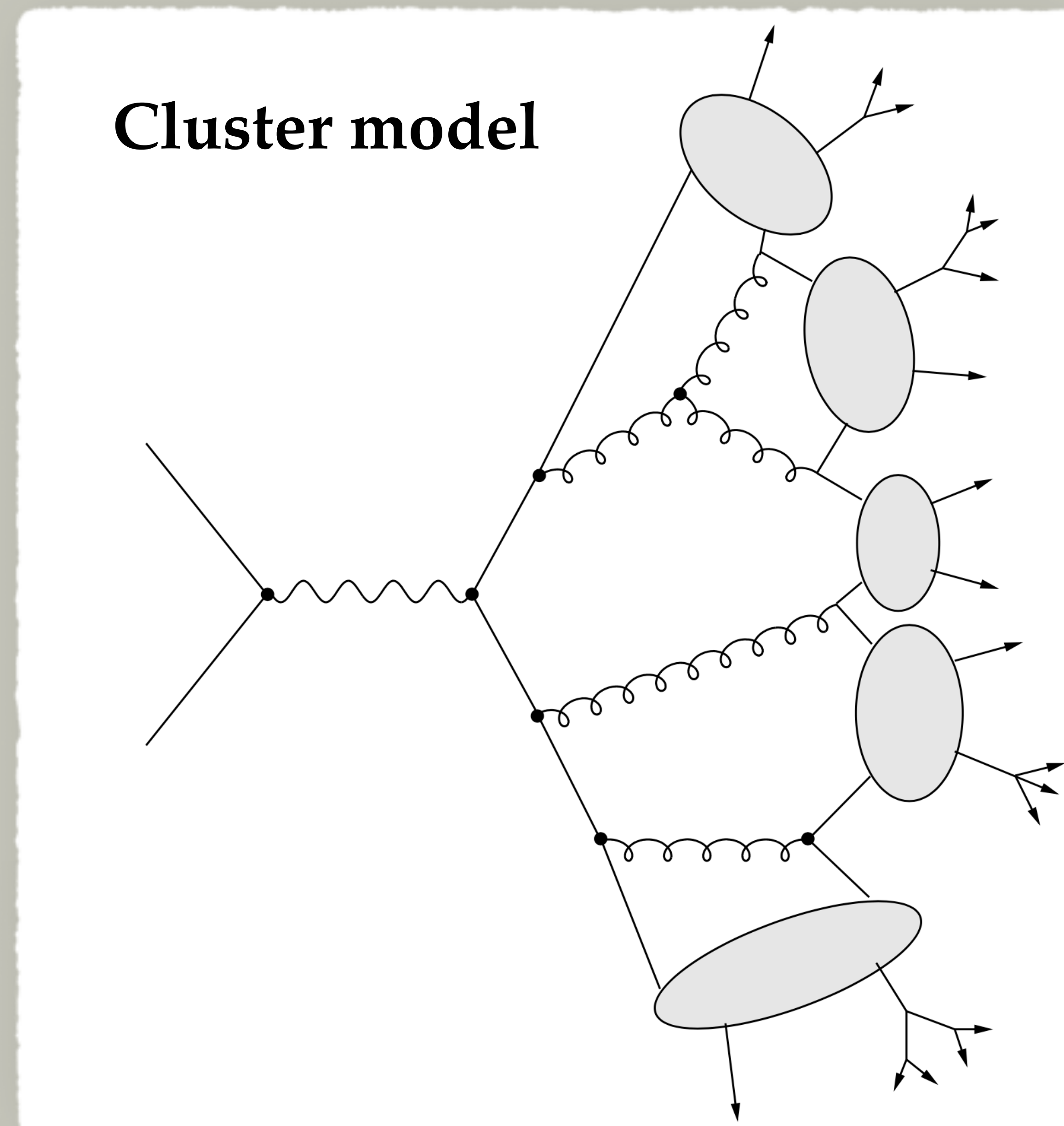
ve

rib

s

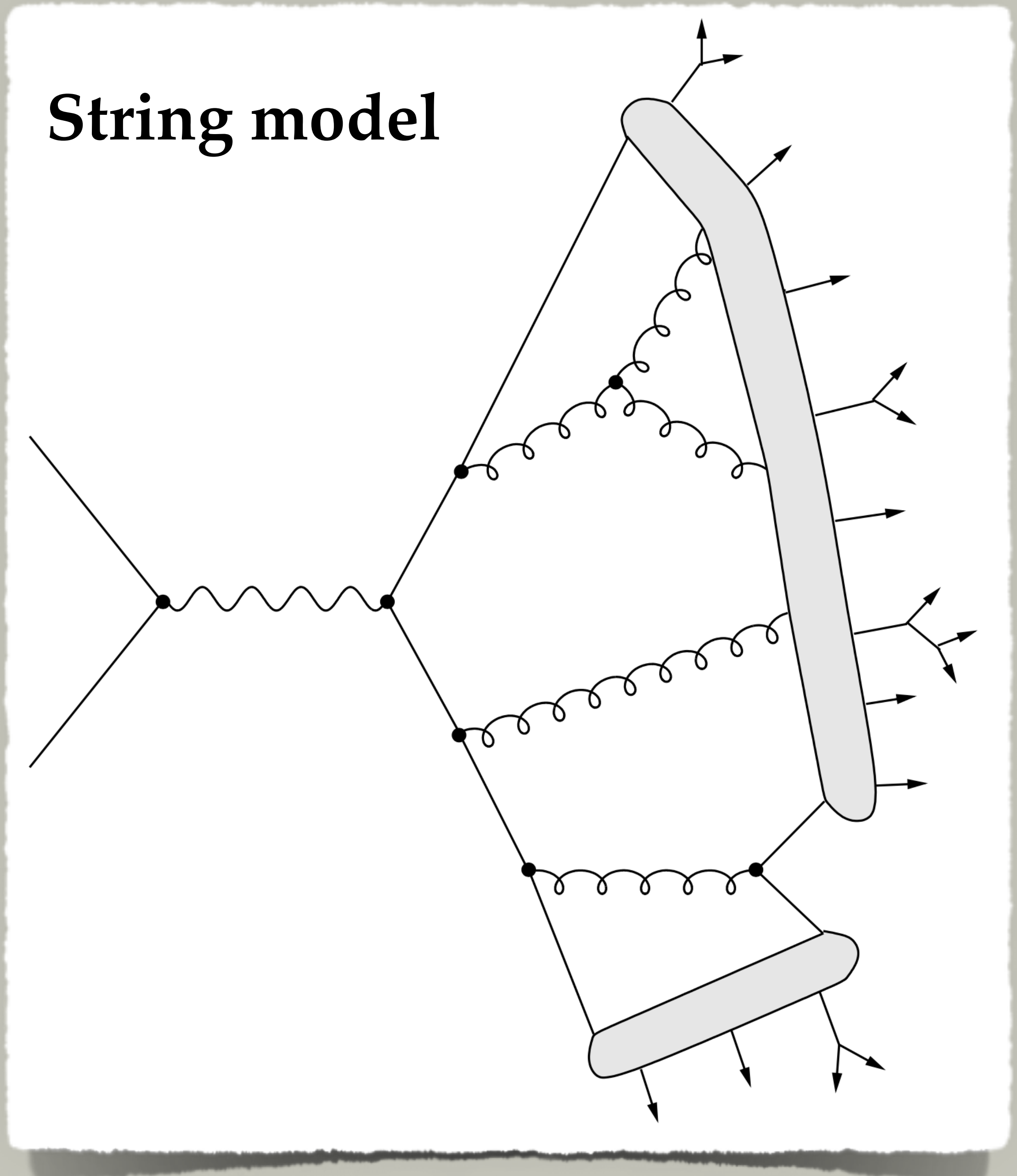
ep)



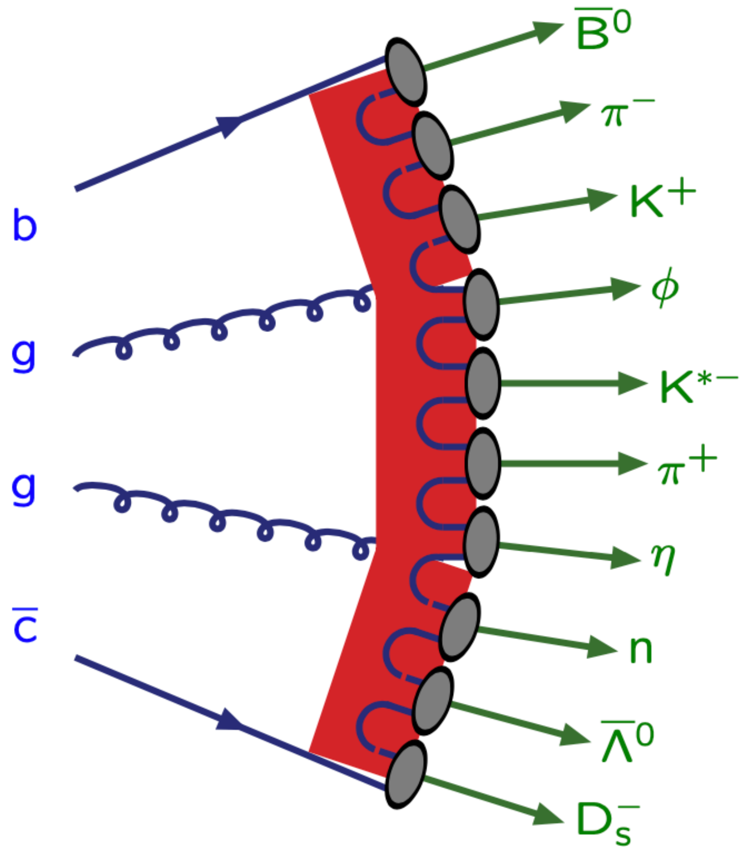


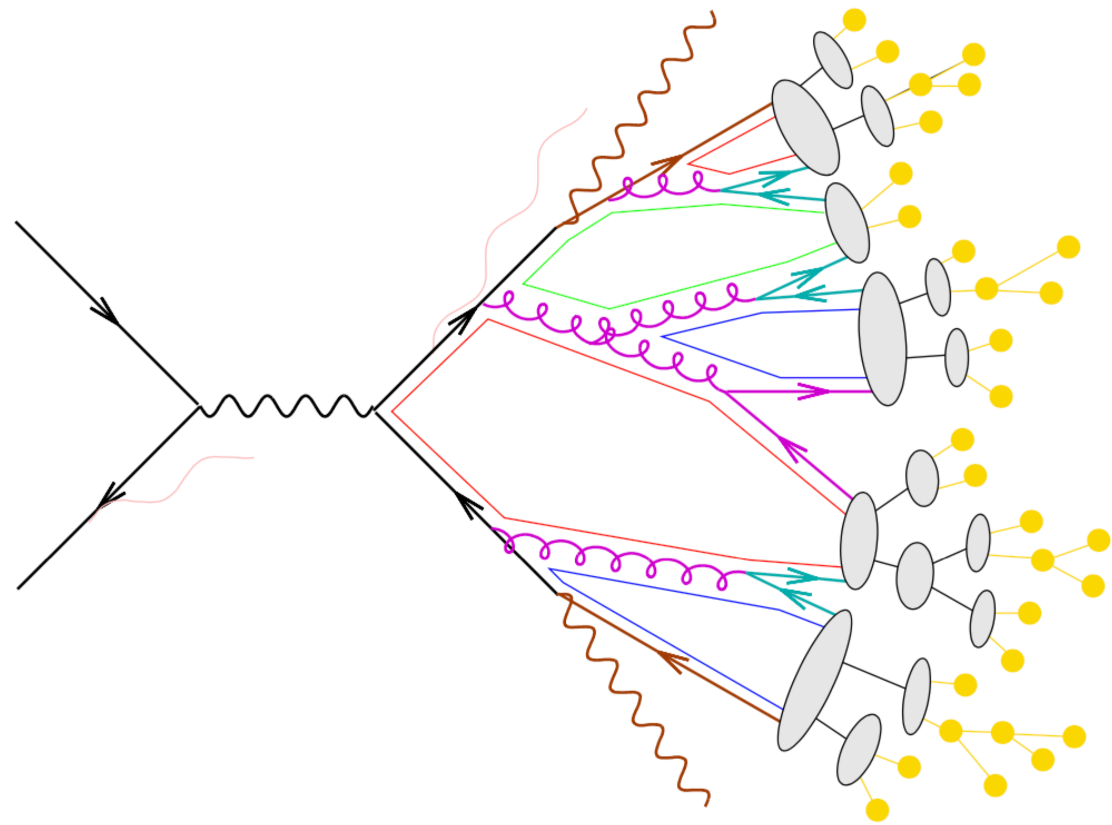
LUND STRING MODEL

- strings connect $q\bar{q}$ systems
  - gluons kinks in strings
- strings break into hadrons
  - controlled by Lund symmetric string fragmentation function *f(z)*
  - flavor selection modeled with tunable parameters
- Pythia Lund string model: many parameters, $\mathcal{O}(100)$
  - 4 params for kinematics
  - most params related to flavor selection and color reconnection



String vs. Cluster





program	PYTHIA	HERWIG
model	string	cluster
energy–momentum picture	powerful	simple
	predictive	unpredictive
parameters	few	many
flavour composition	messy	simple
	unpredictive	in-between
parameters	many	few

"There ain't no such thing as a parameter-free *good* description"

Sjöstrand, 2009

MACHINE LEARNING HADRONIZATION

MLhad: Bierlich, Ilten, Menzo, Mrenna, Wilkinson, Youssef, JZ, 2203.04983, https://gitlab.com/uchep/mlhad

see also HadML: (Chan, Ghosh,) Ju, (Kania), Nachman, (Sangli,) Siodmok, 2203.12660, 2305.17169

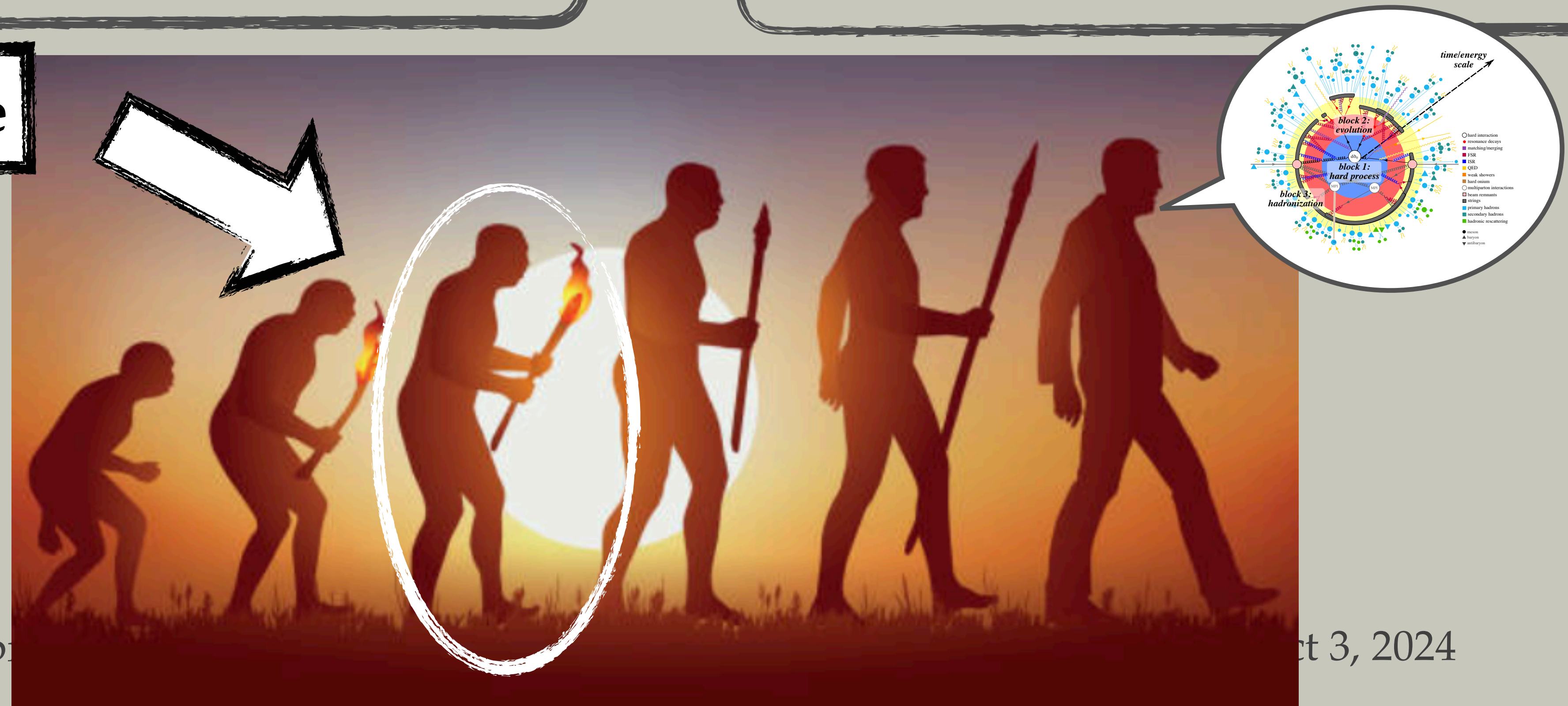
- MLhad: the long term goal
  - use ML to "parametrize our ignorance" about hadronization, use data

MORE IMMEDIATE GOALS

- ML hadronization
  - develop ML architectures
  - reproduce simplified Pythia model
  - learn how to train on data

- Pythia focused
  - easier / faster parameter variation
  - improve on Lund string fragmentation function determin.

we are here



ML hadronization

ML hadronization: MLhad roadmap

- a series of progressive steps to be done before practically useful in Pythia/MC simulations
  - ML architecture that mimicks a simplified Lund string hadronization model
    - train ML on truth level Pythia output (not obs. in exp)
  - develop a framework to propagate errors
  - improved ML architecture with full hadron flavor selector
  - train on mock data (i.e. just observable information)
  - train on real data (i.e. just already measured information)
  - replace/supplement Pythia string model

we are here

MLhad/hadML status

- MLhad: two architectures
  - MLhad cSWAE (conditional Sliced Wasserstein Auto Encoder)
    - latent space distribution need not be analytically known ⇒ could use Pythia output
  - MLhad NF (Bayesian Norm. Flow)
    - incorporation of errors
- trained on a simplif. Pythia string model
- trained on first emissions
  - uses hadronization history ⇒ information that cannot be measured
  - present work: relaxing this assumption

MLhad: 2203.04983, 2311.09296, 2410.nnnnn

- HadML: single architecture
- GAN (Generative Adversarial Network)
  - latent space need not be analytically known
  - two implementations HadML_v1 and HadML_v2
- trained on a simplifed Herwig cluster model
- HadML_v1 trained on first emissions
- HadML_v2 trained on particle flow (point clouds)
  - information that can in principle be measured

HadML: 2203.12660, 2305.17169

MLhad/hadML status

- MLhad: two architectures
  - MLhad cSWAE (conditional Sliced Wasserstein Auto Encoder)
    - latent space distribution need not be analytically known ⇒ could use Pythia output
  - MLhad NF (Bayesian Norm. Flow)
    - incorporation of errors
- trained on a simplif. Pythia string model
- trained on first emissions
  - uses hadronization history ⇒ information that cannot be measured
  - present work: relaxing this assumption

MLhad: 2203.04983, 2311.09296, 2410.nnnnn

- HadML: single architecture
- GAN (Generative Adversarial Network)
  - latent space need not be analytically known
  - two implementations HadML_v1 and HadML_v2
- trained on a simplifed Herwig cluster model
- HadML_v1 trained on first emissions
- HadML_v2 trained on particle flow (point clouds)
  - information that can in principle be measured

HadML: 2203.12660, 2305.17169

THE STRATEGY

- MLhad architectures capture well the (simplified) Pythia Lund string model
- proof of principle - need to see how this ports to training on data
- we want to achieve this in steps
    - modify only parts of Pythia Lund string model

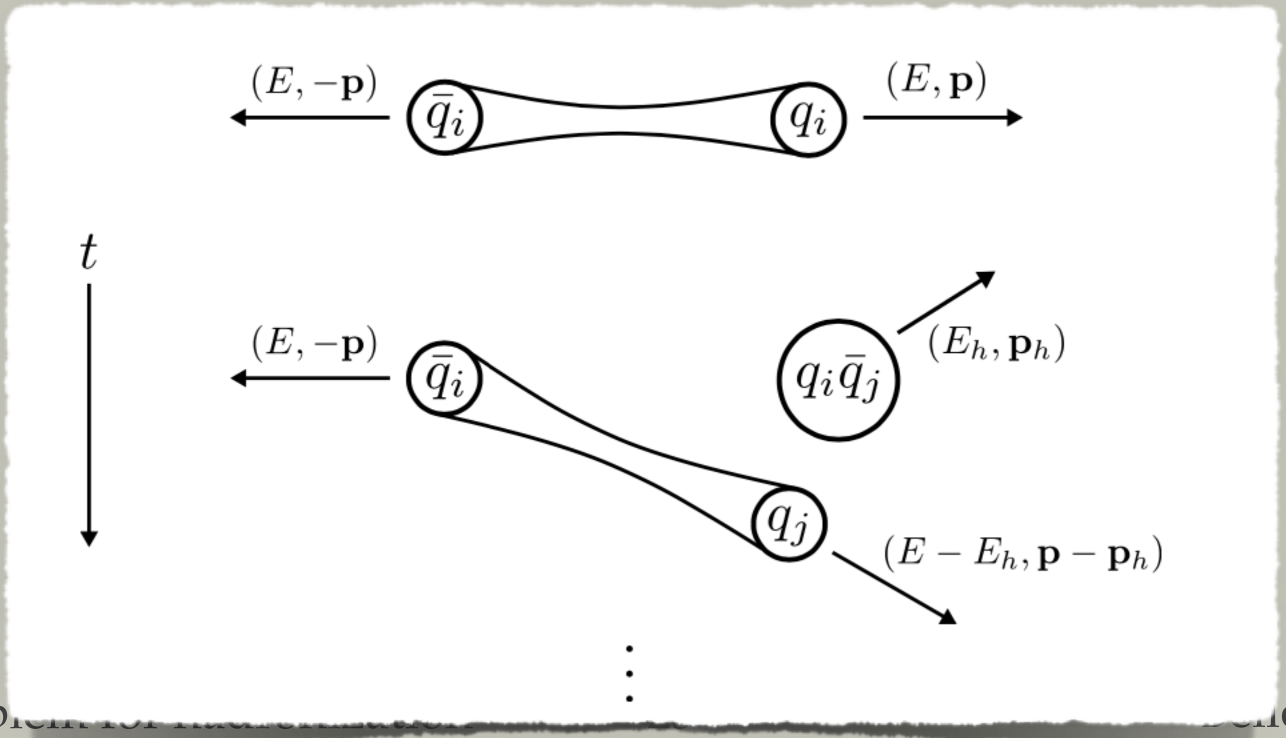
PYTHIA FOCUSED

TWO DIFFERENT GOALS

- solving the inverse problem for hadronization
  - extracting string fragmentation function from data
- speeding up Pythia hadronization when varying parameters
  - reweighting method $\Rightarrow$ backup slides

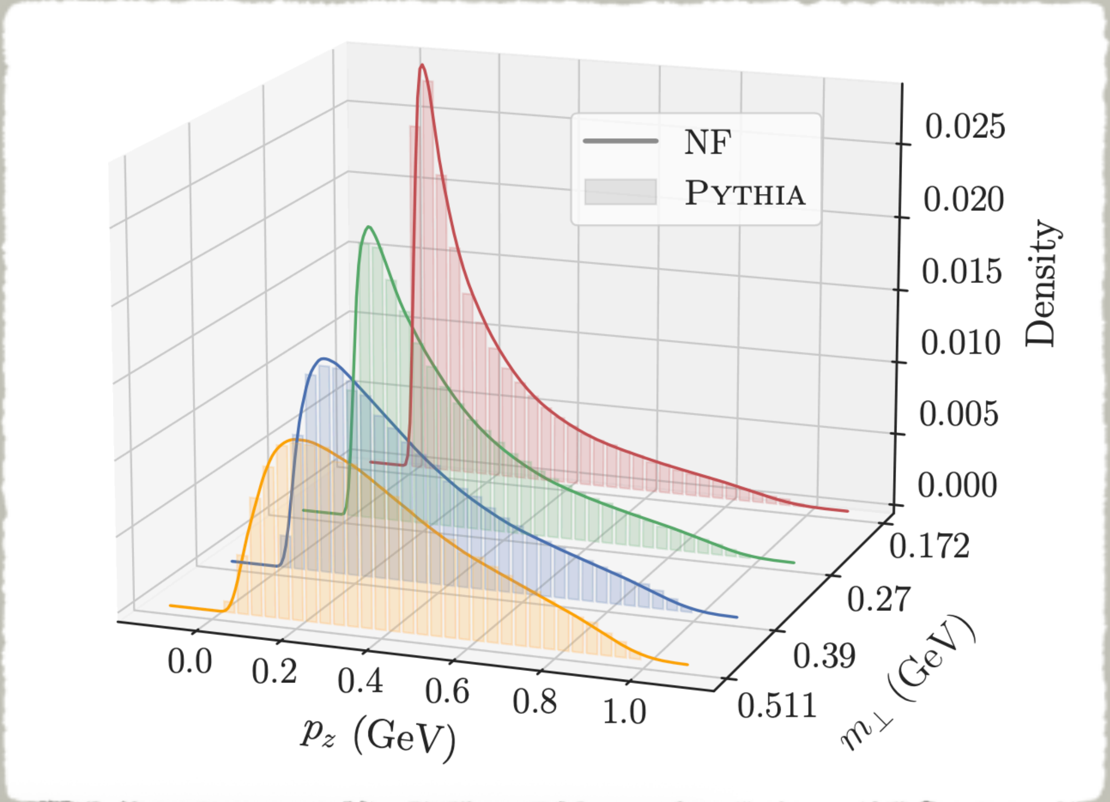
SIMPLIFIED STRING HADRONIZATION MODEL

- assume that color flow done correctly by Pythia
  - including splitting gluons, so that only strings with $q, \bar{q}$ ends
- hadron emission from a string piece controled by fragmentation function $f(z)$
  - the whole hadronization chain is then reproduced by iterating
  - the string is labeled by $q, \bar{q}$ flavor and its energy in cms, $2E$
- for now only u,d quarks, uses Pythia flavor selector



MLhad: 2311.09296

STRING FRAGMENTATION FUNCTION



- in c.o.m. frame of the string:
  - hadron emission described by p_z, p_T + uniform distrib. in azimuthal angle
- p_z distribution from Lund string fragmentation funtion

$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(-\frac{b\, m_\perp^2}{z}\right)$$

$$z = (p_{h,z} + E_h)/E_{\text{string}}$$

$$m_\perp^2 \equiv m^2 + p_T^2$$

- for light quark flavored hadrons only three params.: a,b and mass, m
- p_T from random Gaussian distributions; width another param.

INVERSE PROBLEM

- inverse problem for hadronization
  - can one learn $f(z)$ from data*?
    - *without asking for its parametric form
- compare: NNPDF determinations of PDFs from data

A HARDER PROBLEM

- PDFs: appear at most in quadratic form in observables

$$\sigma_X(s, M_X^2) = \sum_{a,b} \int_{x_{\min}}^1 dx_1\, dx_2\, f_{a/h_1}(x_1, M_X^2) f_{b/h_2}(x_2, M_X^2) \hat{\sigma}_{ab\to X}\left(x_1 x_2 s, M_X^2\right).$$

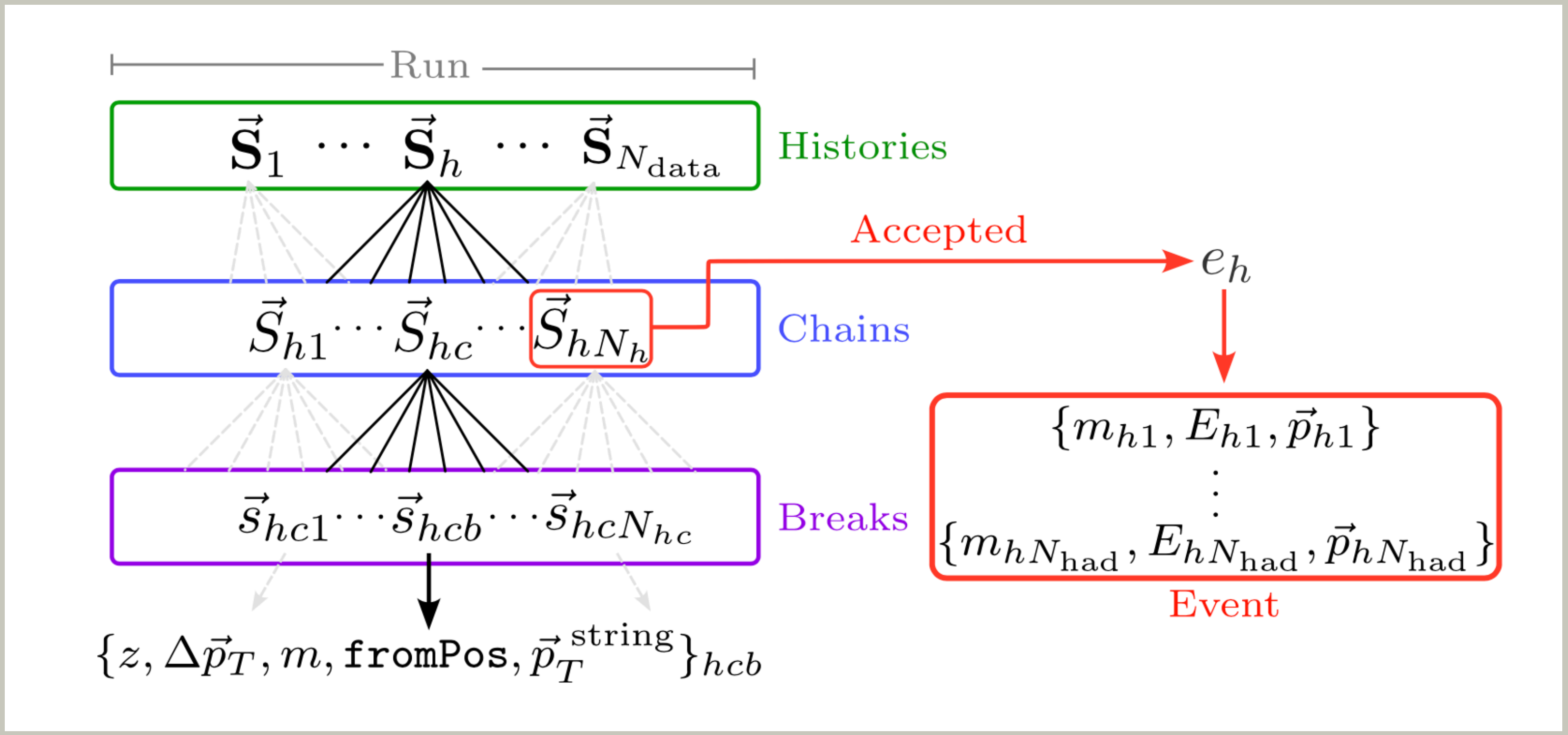
- string fragmentation function: each event a different number of hadrons
  - different number of samplings of $f(z)$

$$\text{Prob.} \sim \prod_{i=1}^{N_{\text{hadr.}}} f(z_i)$$

A HARDER PROBLEM

MLhad: 2410.nnnnn

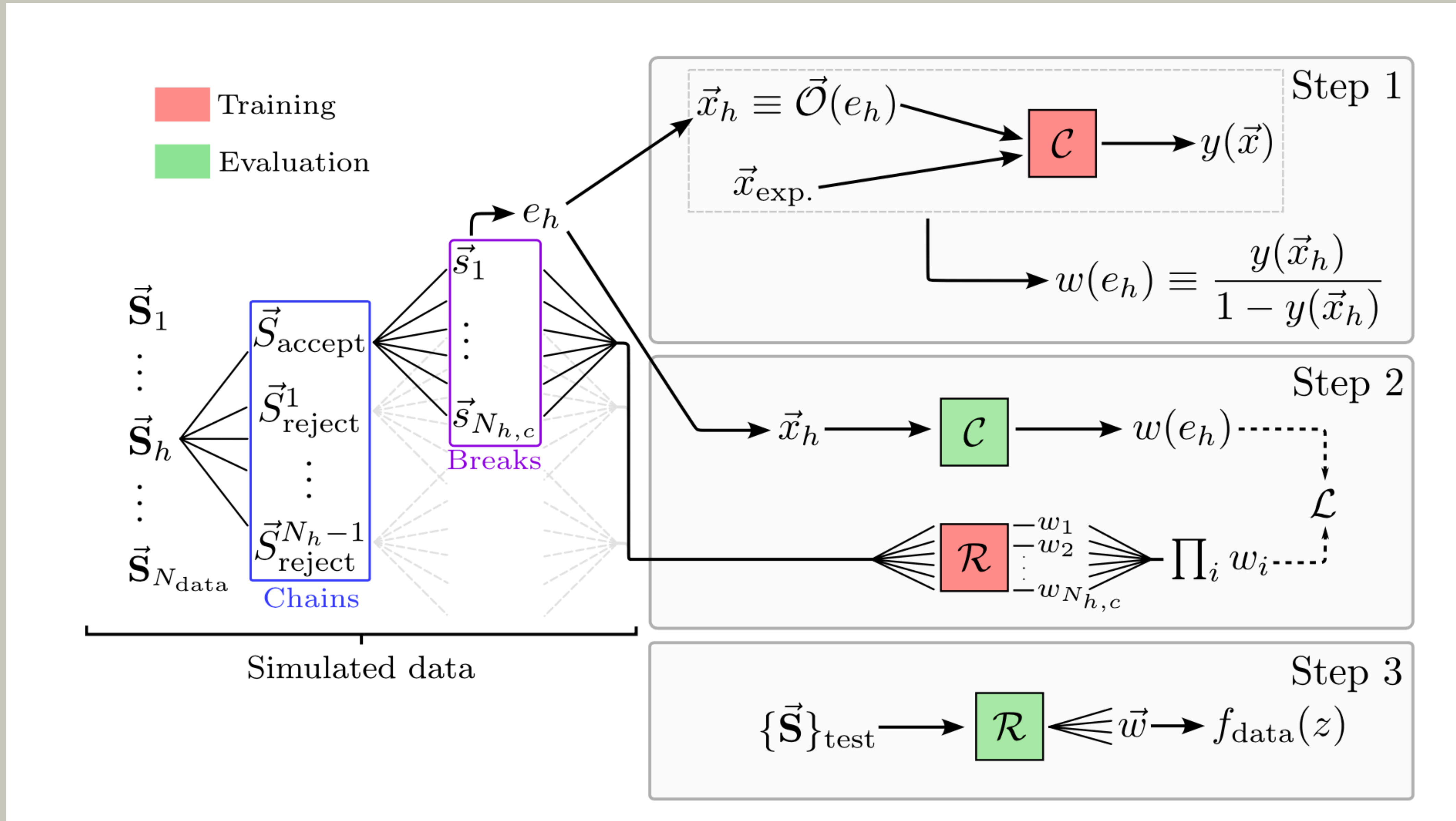
- further complications
  - a complicated permutation symmetry: swapping orders of emissions + appropriate boosts ⇒ lead to the same event
  - in Pythia the final emission may or may not lead to viable kinematics ⇒ $O(1)$ fraction of simulated fragmentation chains is rejected



THE HOMER METHOD

MLhad: 2410.nnnnn

- the HOMER (Histories and Observables for Monte-Carlo Event Reweighting) method
- 3 step approach
  - Step 1: train a classifier on events ⇒ likelihood of an event
  - Step 2: can use this on simulated events to find a neural-net representation of $f(z)$
  - Step 3: simulate using this new $f(z)$
- in all steps use ratios of probabilities (weights)
  - always reweighting from base Pythia simulation results



2410.nnnnn

- Step 2: can use this on simulated events to find a neural-net representation of $f(z)$
- Step 3: simulate using this new $f(z)$

- in all steps use ratios of probabilities (weights)
  - always reweighting from base Pythia simulation results

HOMER RESULTS

MLhad: 2410.nnnnn

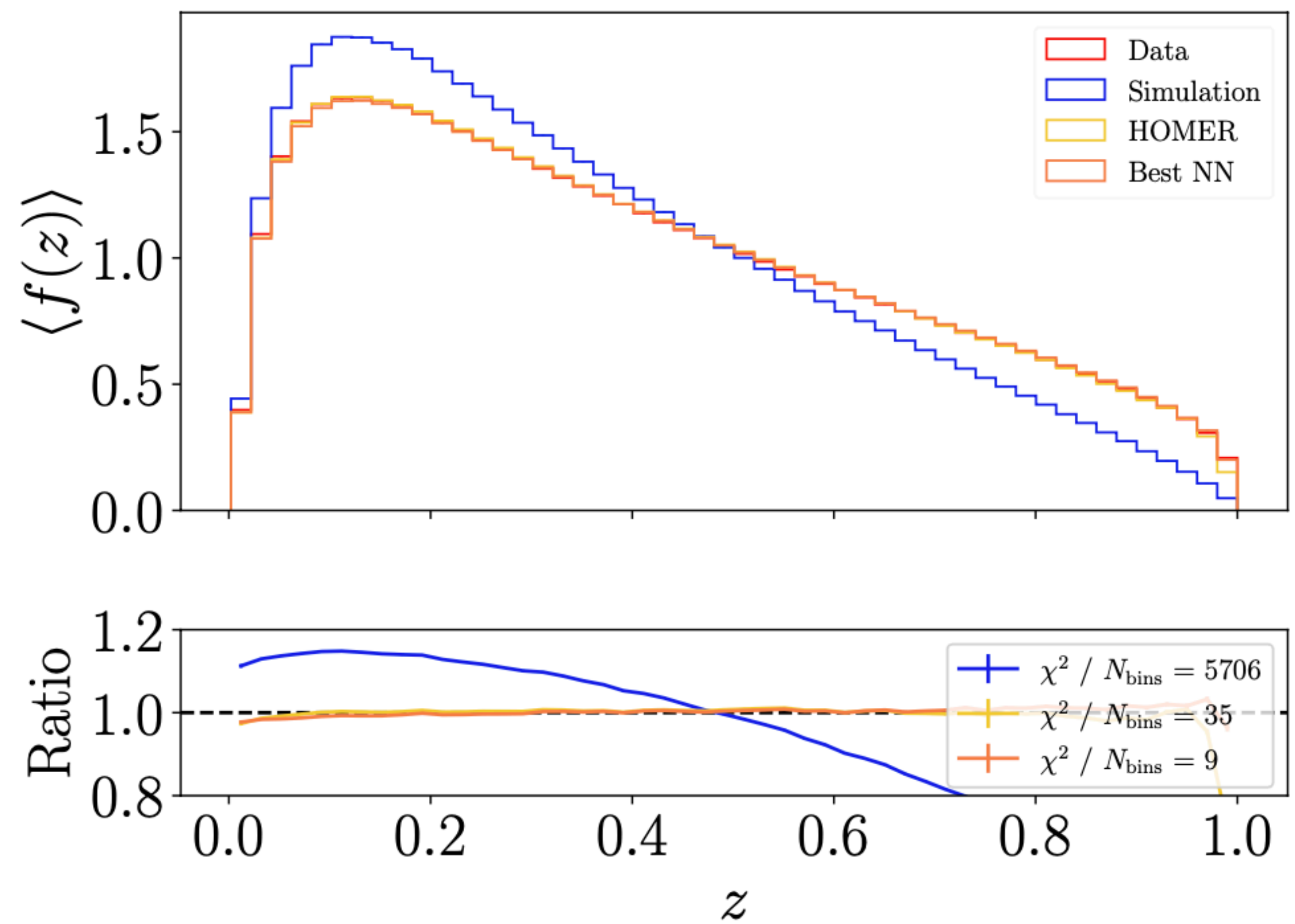
- Note: HOMER uses only measurable quantities - three versions
  - (i) binned high-level obs.;
  - (ii) unbinned high level obs.;
  - (iii) point cloud
- so far a simplified case: $q\bar{q}$ string of fixed energy

HOMER RESULTS

MLhad: 2410.nnnnn

- in this simplified case
  - binned high-level observables (multiplicities, shape observ.,...) suffice
  - additional gain for unbinned high-level obs. case
  - point cloud harder to train on

binned high-level observable case



MLhad: 2410.nnnnn

- in
  - b
    (
    s
  - additional gain for unbinned high-level obs. case
  - point cloud harder to train on

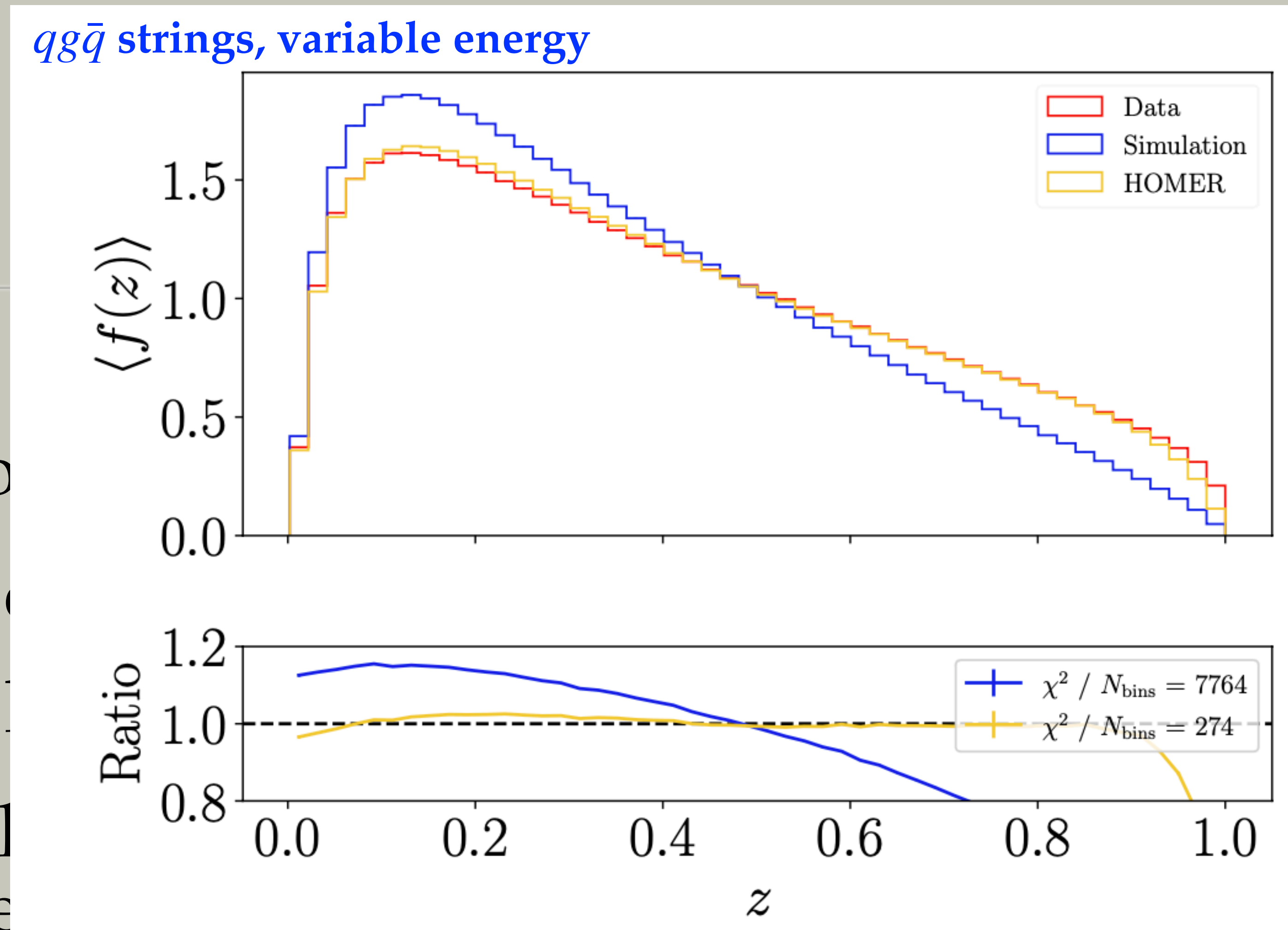
ADDING GLUONS

MLhad: 2412.nnnnn

- working in progress: adding gluons to the string
  - energies and number of gluons as one would get from parton shower
- approximation that went into HOMER for $q\bar{q}$ string breaks down
  - one needs to calculate averages over several simulated fragmentation chains to obtain the estimate for an event weight
  - works well for a single gluon, even if strings with many energies in a sample
  - for many gluons further work required

s

MLhad: 2412.nnnnn



- wo … string
    - … ould get from
- ap … $q\bar{q}$ string bre …
    - one needs to calculate averages over several simulated fragmentation chains to obtain the estimate for an event weight
    - works well for a single gluon, even if strings with many energies in a sample
    - for many gluons further work required

PYTHIA-8 PLUGIN MODULE

- not directly related to hadronization, but is an output of MLhad effort:
  - Pythia8 user contribution plugin platform Pythia8-contrib will be available
    - similar to FastJet-contrib in concept
  - MLhad NF and MLhad cSWAE as test packages
  - contact me, if you are interested in submitting a package/beta test

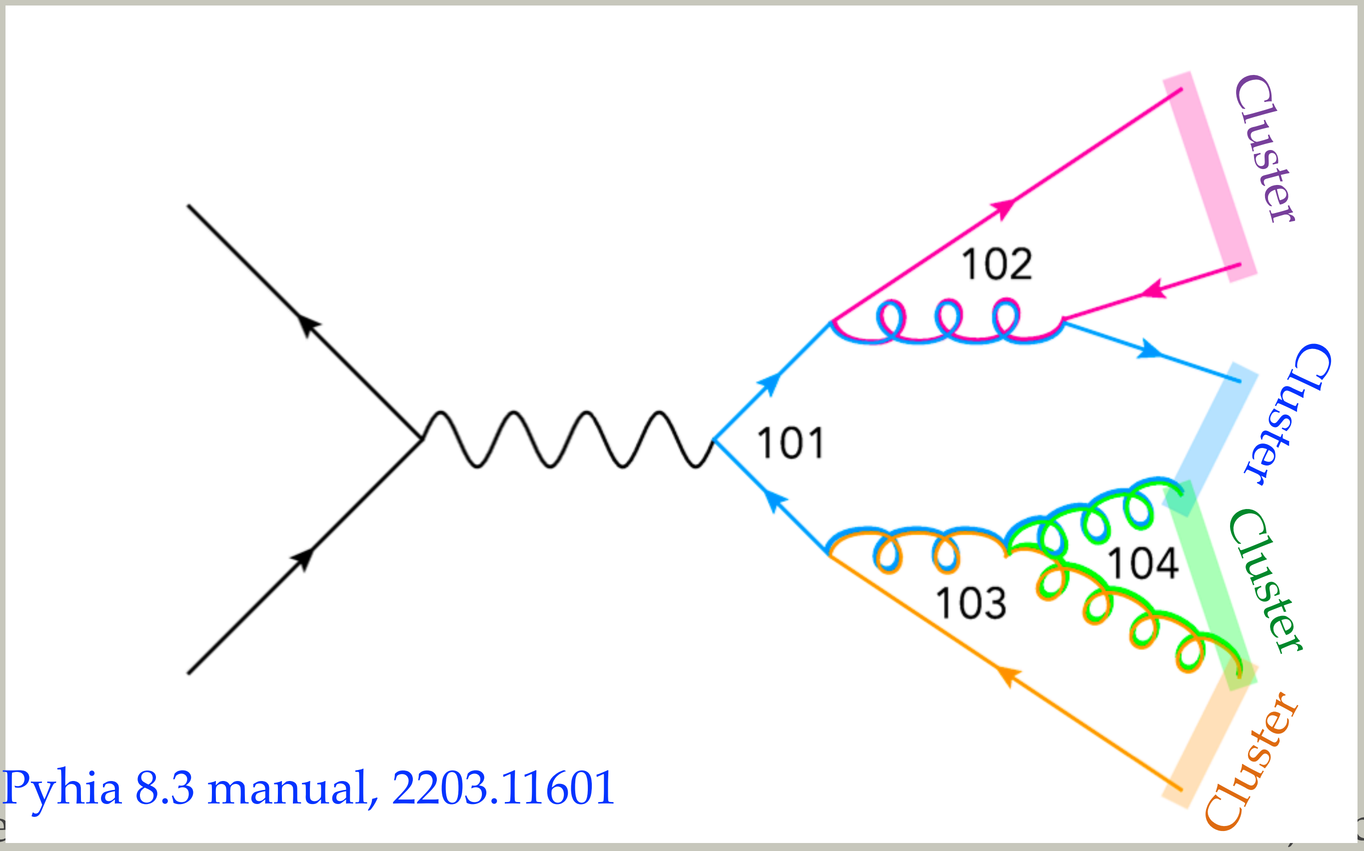
CONCLUSIONS

- MLhad/HadML: first steps in creating ML based hadronization description
- of immediate use but not shown in the talk
  - reweighting algorithms in Pythia for faster variation of hadronization params.

BACKUP SLIDES

CLUSTER MODEL

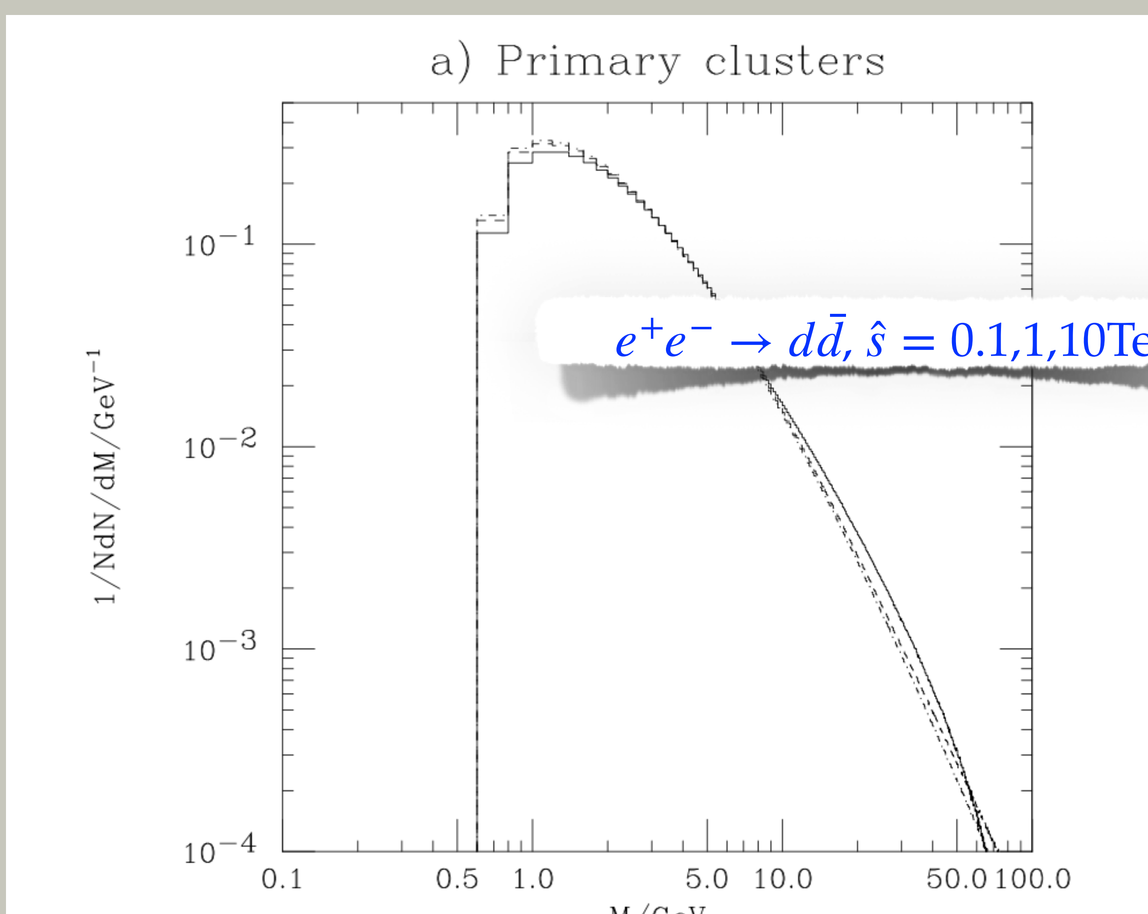
- assign mass to gluons, decay them to $q\bar{q}$ pairs
  - these are color singlets: *primary clusters*
  - primary clusters have universal mass distrib
- heavier clusters are decayed to lighter ones (model dep. step)
- relatively small set of params, $\mathcal{O}(30)$



Pyhia 8.3 manual, 2203.11601

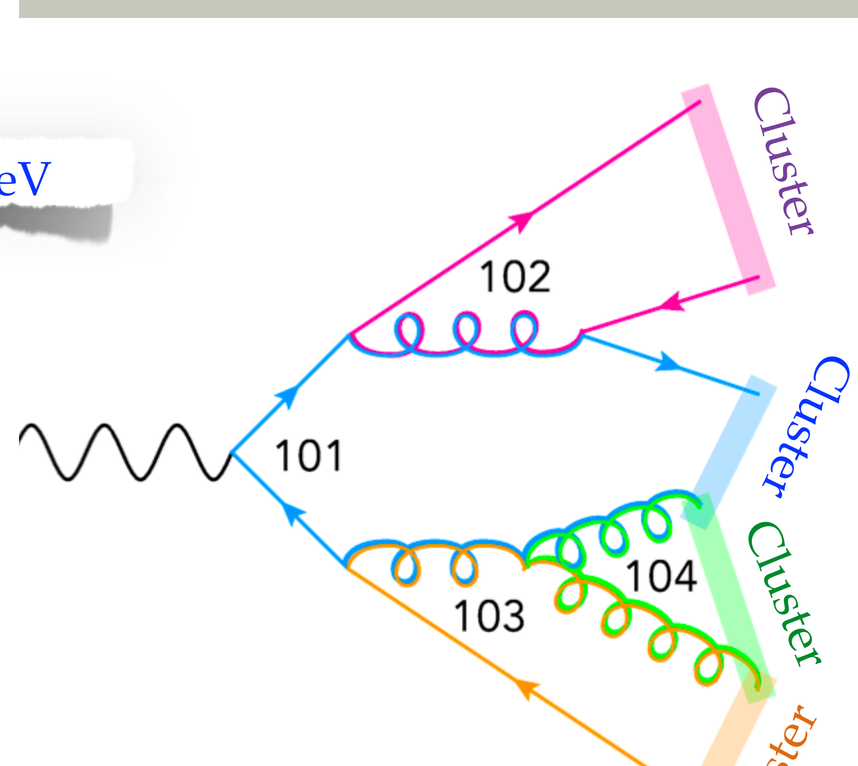
CLUSTER MODEL

- assign mass to gluons, decay them to $q\bar{q}$ pairs
  - these are color singlets: *primary clusters*
  - primary clusters have universal mass distrib
- heavier clusters are decayed to lighter ones (model dep. step)
- relatively small set of params, $\mathcal{O}(30)$



$e^+e^- \to d\bar{d}, \hat{s} = 0.1,1,10$TeV

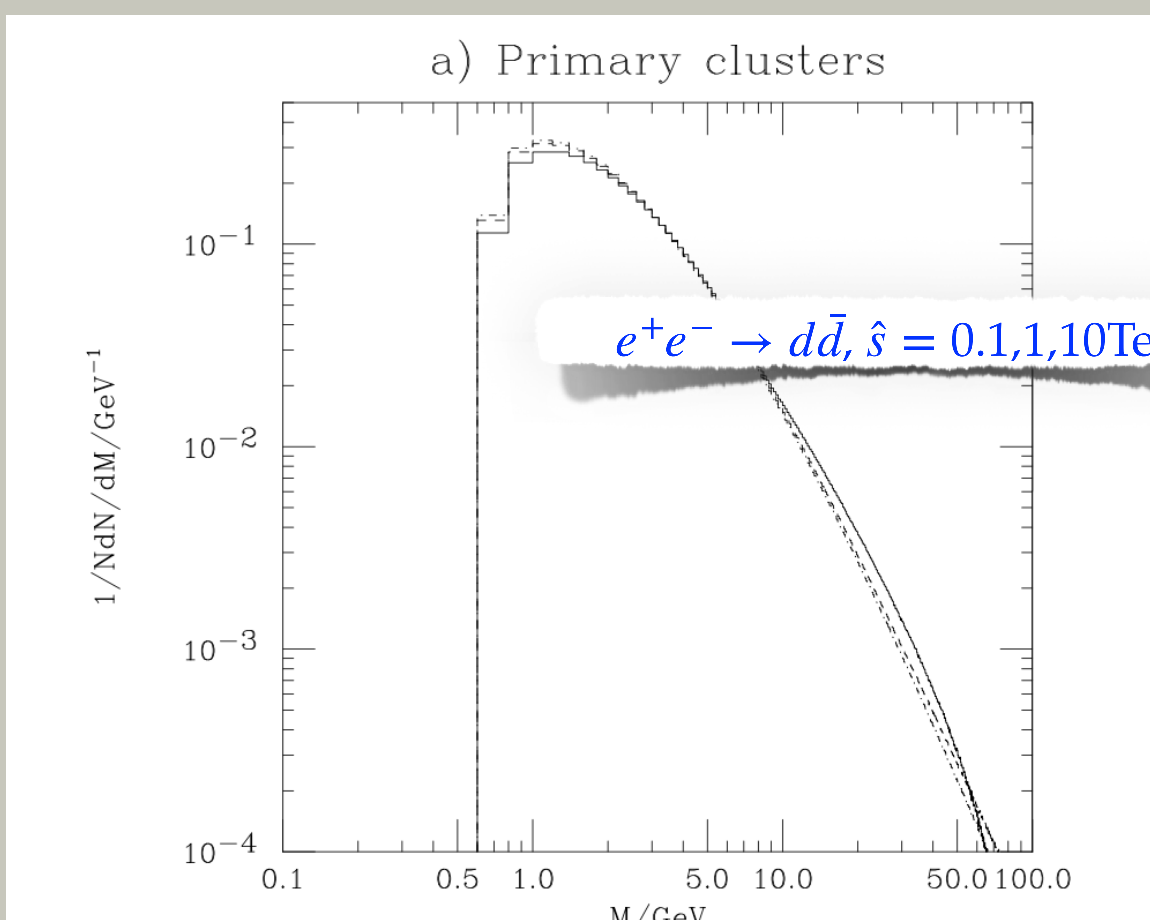
Herwig++ manual, 0803.0883



.11601

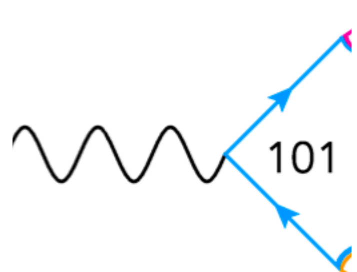
CLUSTER MODEL

- assign mass to gluons, decay them to $q\bar{q}$ pairs
  - these are color singlets: *primary clusters*
  - primary clusters have universal mass distrib
- heavier clusters are decayed to lighter ones (model dep. step)
- relatively small set of params, $\mathcal{O}(30)$

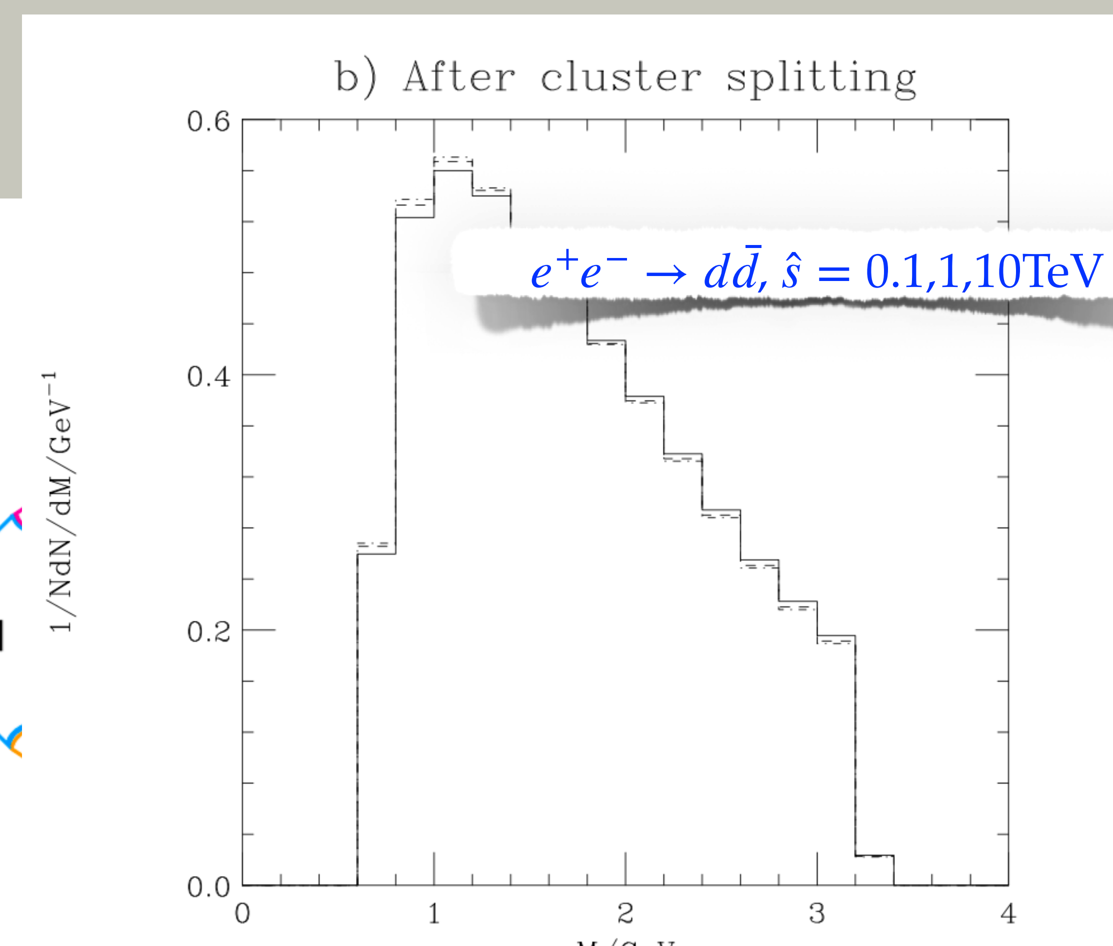


$e^+e^- \to d\bar{d}$, $\hat{s} = 0.1,1,10$TeV

Herwig++ manual, 0803.0883



.11601

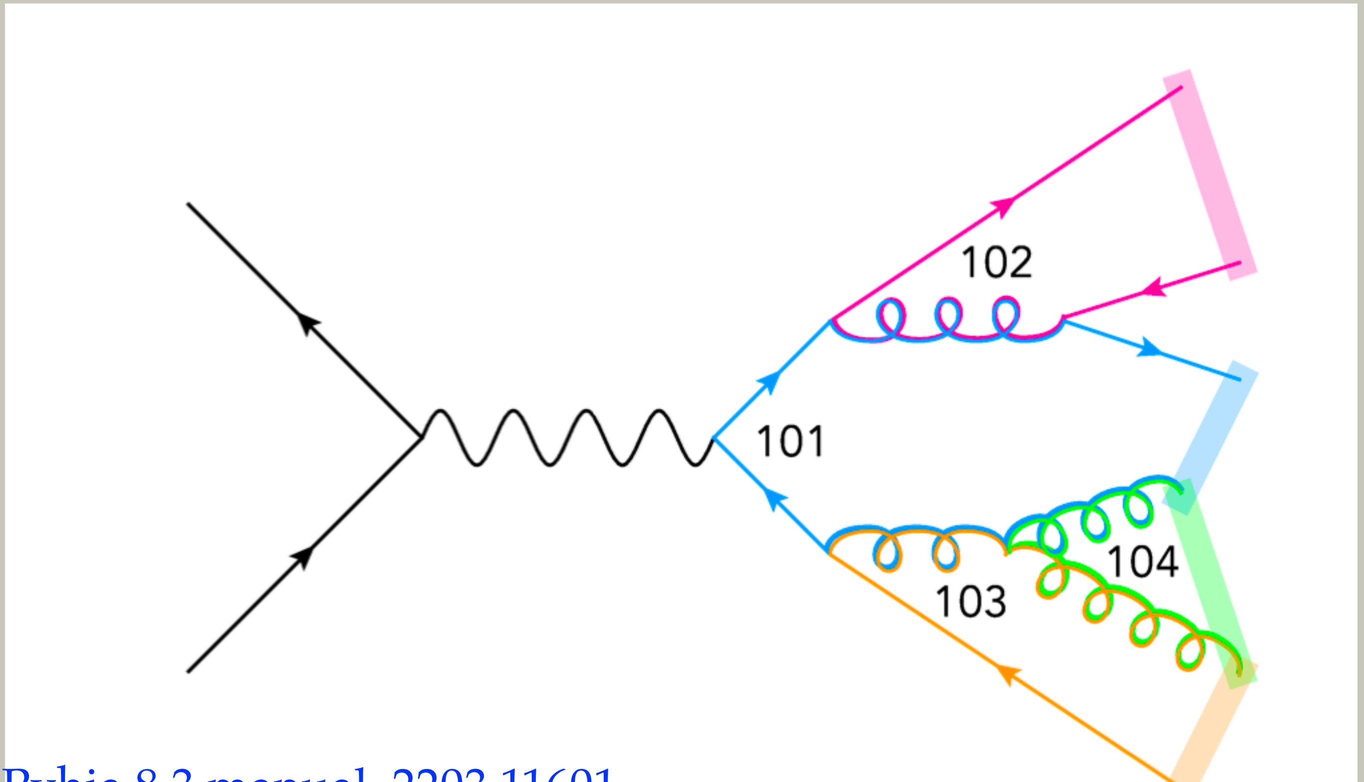


$e^+e^- \to d\bar{d}$, $\hat{s} = 0.1,1,10$TeV

Herwig++ manual, 0803.0883

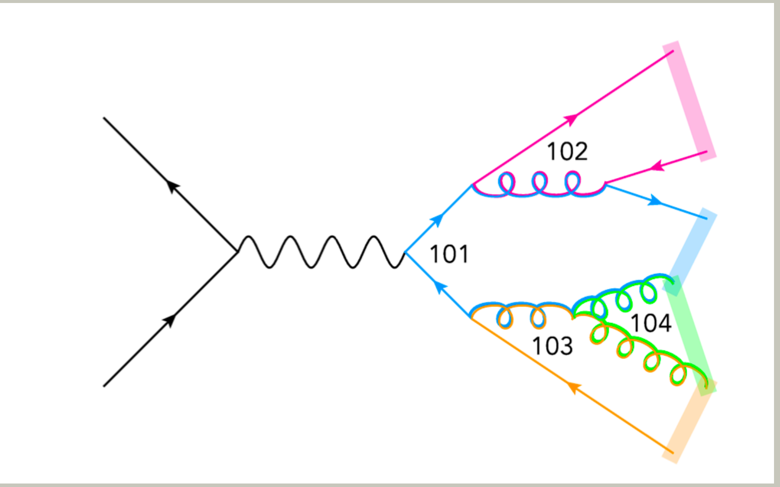
LUND STRING MODEL

- strings connect $q\bar{q}$ systems
- gluons kinks in strings
  - split gluons to a collinear $q\bar{q}$ pair ⇒string pieces
- string pieces break into hadrons (model dep.)
  - controlled by Lund string fragmentation function
- Pythia Lund string model: many parameters, $\mathcal{O}(100)$



Pyhia 8.3 manual, 2203.11601

COLOR RECONNECTION



- all perturbative predictions in leading color approximation ($N_c \to \infty$ with $\alpha_s N_c$ fixed)
  - direct mapping of color flow to strings
- color reconnection: inclusion of $1/N_c$ suppressed terms (model dep.)
  - reassing colors, no change in parton momenta — Pyhia 8.3 manual, 2203.11601
  - several examples where important
    - first historic mention: for charmonium production in B decays — Fritzch, 1977; Ali et al, 1979
    - for multiple parton interactions (Pythia MPI model) — Sjöstrand, Zijl, 1987
    - $e^+e^- \to W^+W^- \to 4j$ at LEP 2 excludes no CR hypothesis — 1302.3415
    - top quark mass determination from hadronic tops
  - several color reconnection models in Pythia
- computationally expensive, especially at high multiplicities

CHALLENGES FOR HADRONIZATION MODELS

Fischer, Sjostrand, 1610.09818

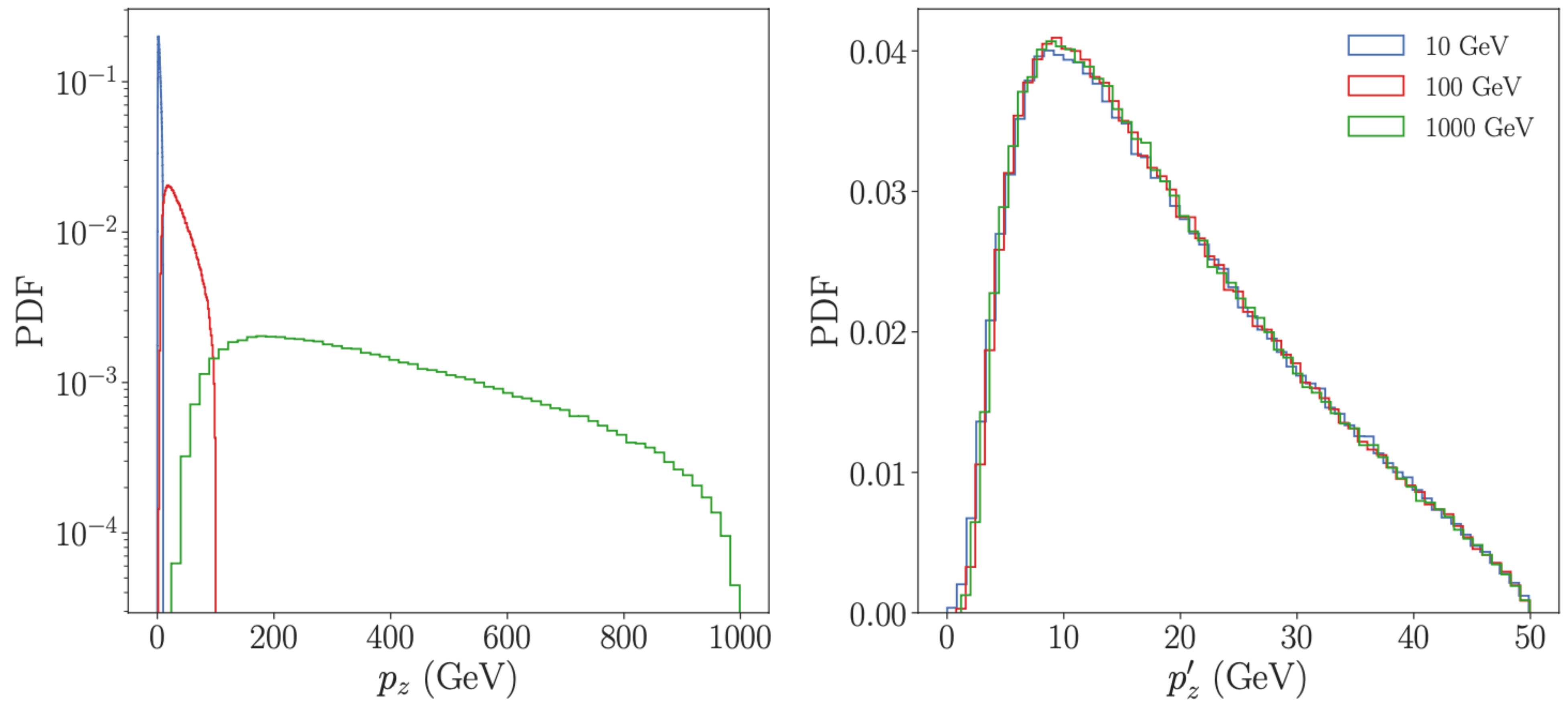
- in general out of the box hadronizations models work within *20-50%*
- some challenges for Pythia
  - change of flavor composition with event multiplicity
    - high multiplicity events have higher strangenesss content
    - no mechanism in Pythia to mimic it
  - average $\langle p_T \rangle$ larger for heavier particles, trend ok in Pythia, but numerically not large enough
  - charge particle p_T spectrum not correctly modelled at low p_T
    - partially can be fixed by tunes, but then a problem at interm. p_T
  - there is a peak in Λ/K p_T spectrum at $p_T \sim 2.5$ GeV, not reproduced by Pythia
  - the observation of the ridge in *pp* requires collective effects
- at least some of them addressed in Pythia 8.3 by introducing more involved models of string interactions, thermodynamical string fragmentation model, etc.
- Herwig has a different set of challenges, e.g., predicting heavy baryon distributions

MLhad

- right now trained directly on Pythia first emission output
  - hadron mom. described by p_z, p_T
- the IR cut-off has two effects
  - p_z and p_T distributions are uncorellated
  - makes the problem scale invariant in p_Z
    - enough to train at one string mass, $2E_{\text{ref}}$
    - for other energies can rescale

$$p_z' \equiv E_{\text{ref}} \frac{p}{E},$$

- this is relaxed in the end, E dependence can be recovered



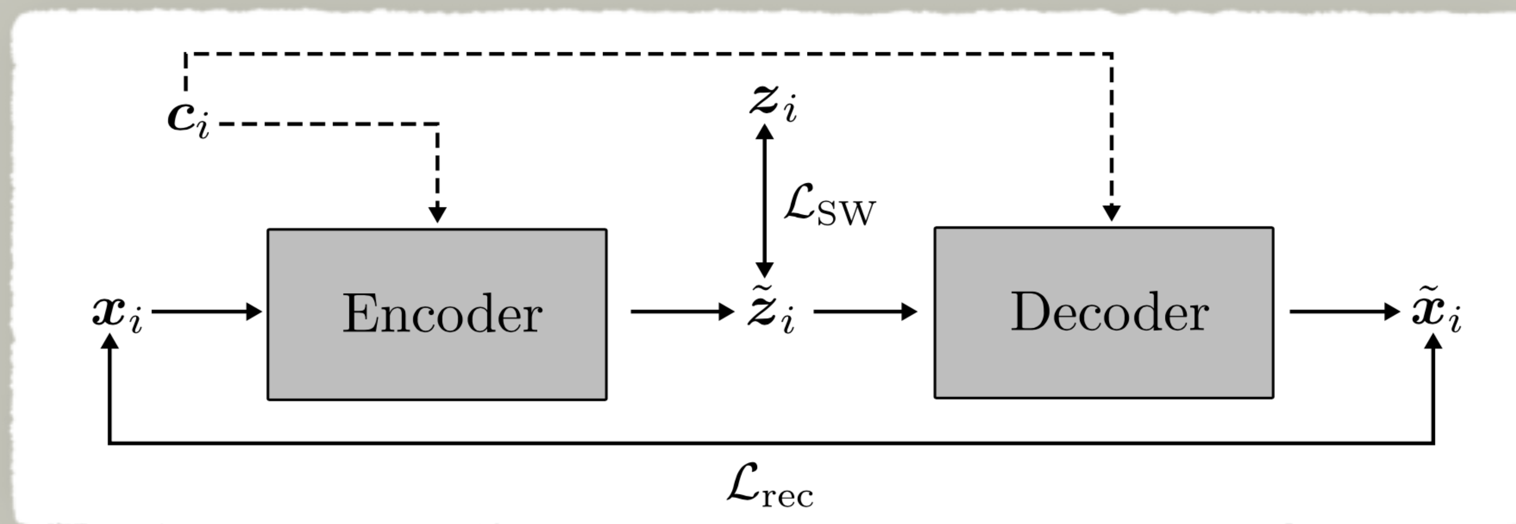
- p_z and p_T distributions are uncorellated
- makes the problem scale invariant in p_Z
  - enough to train at one string mass, $2E_{\text{ref}}$
  - for other energies can rescale

$$p'_z \equiv E_{\text{ref}} \frac{p}{E},$$

- this is relaxed in the end, E dependence can be recovered

cSWAE

- use conditional Sliced-Wasserstein Autoencoder
  - SW gives flexibility in the use of latent space distributions



- string energy E_i is encoded in a label $\bar{c}_i$

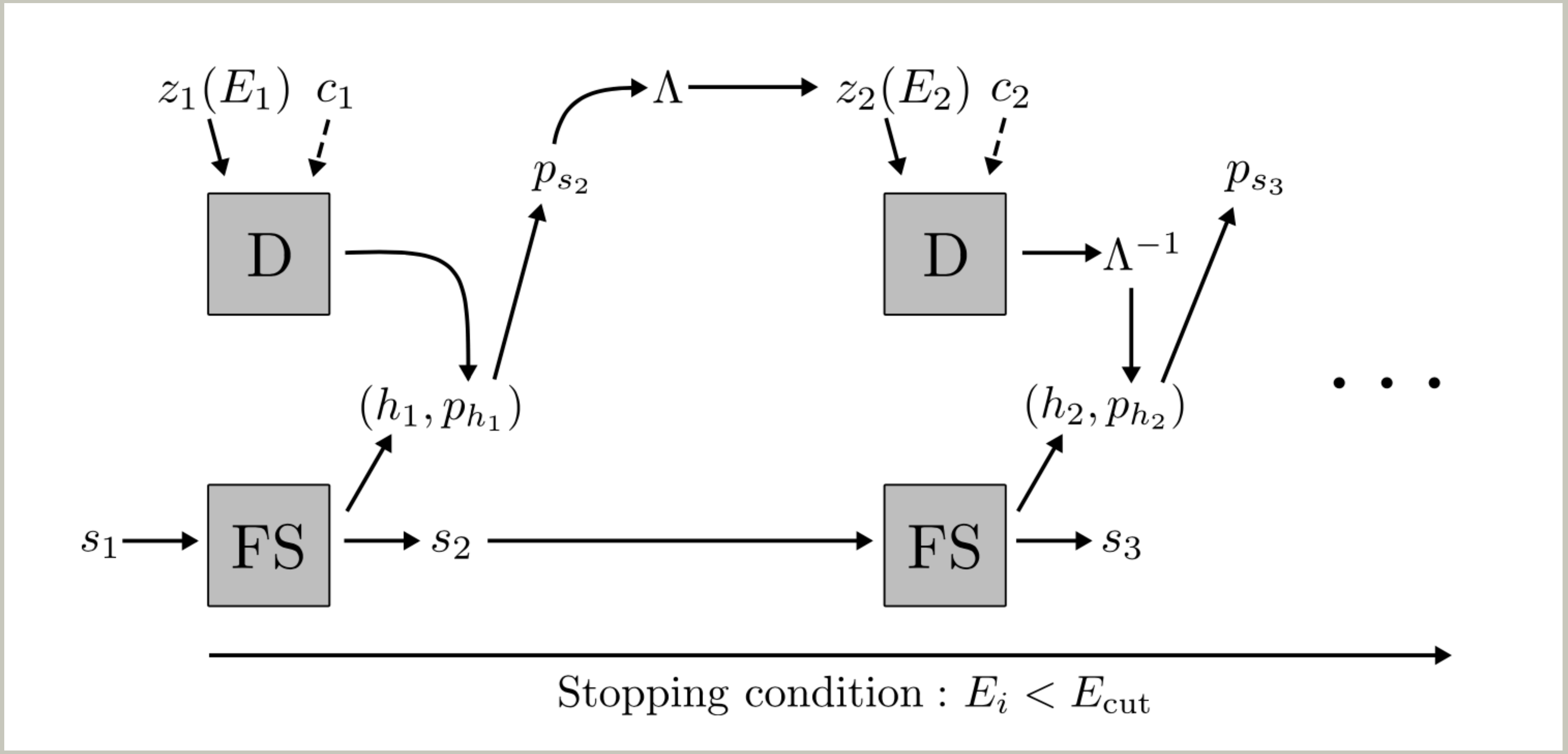
$$\bar{c}_i = \frac{E_{\rm max} - E_i}{E_{\rm max} - E_{\rm min}},$$

- training data: $\mathbf{x}_i$ sorted vector of 100 first emission
  - either p_z or p_T values
- loss function

$$\mathcal{L}(\psi, \phi) = \mathcal{L}_{\rm rec} + \mathcal{L}_{\rm SW},$$

MLhad as a generator

- MLhad as a generator of the hadronization chains



RESULTS - FIRST EMISSIONS

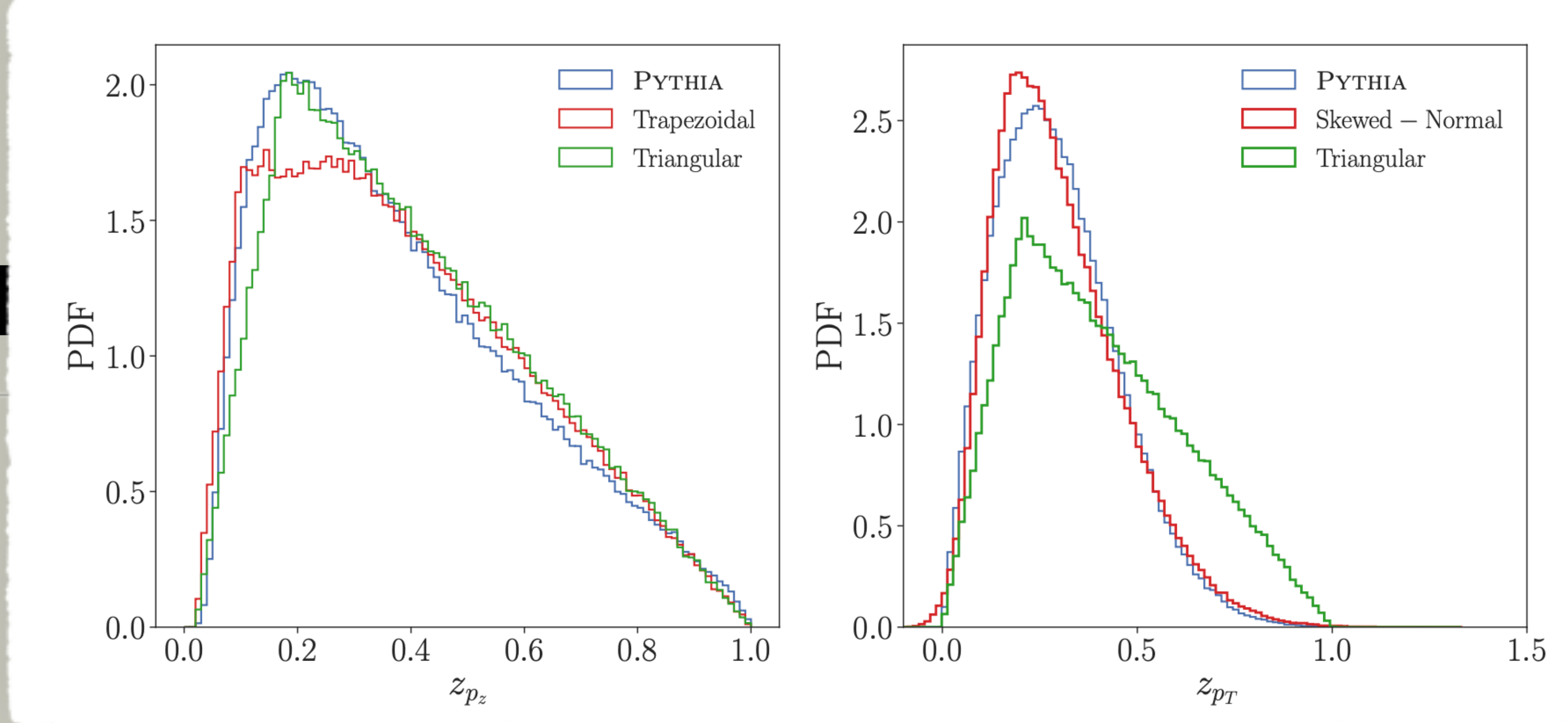
- three different latent space distributions used
- cSWAE training configurations

latent space dim.

L_{SW} vs. L_{rec}

# of SW slices

Variable $\boldsymbol{x}$	Target $\boldsymbol{z}$	t (epochs)	d_z	λ	L
p'_z	Pythia	150	35	35	15
	Trapezoidal	300	2	20	30
	Triangular	150	2	30	25
p_T	Pythia	100	20	30	30
	Skew-norm	120	4	20	25
	Triangular	120	4	15	25



RI … ONS

- … ns used
- cSWAE training configurations

latent space dim.

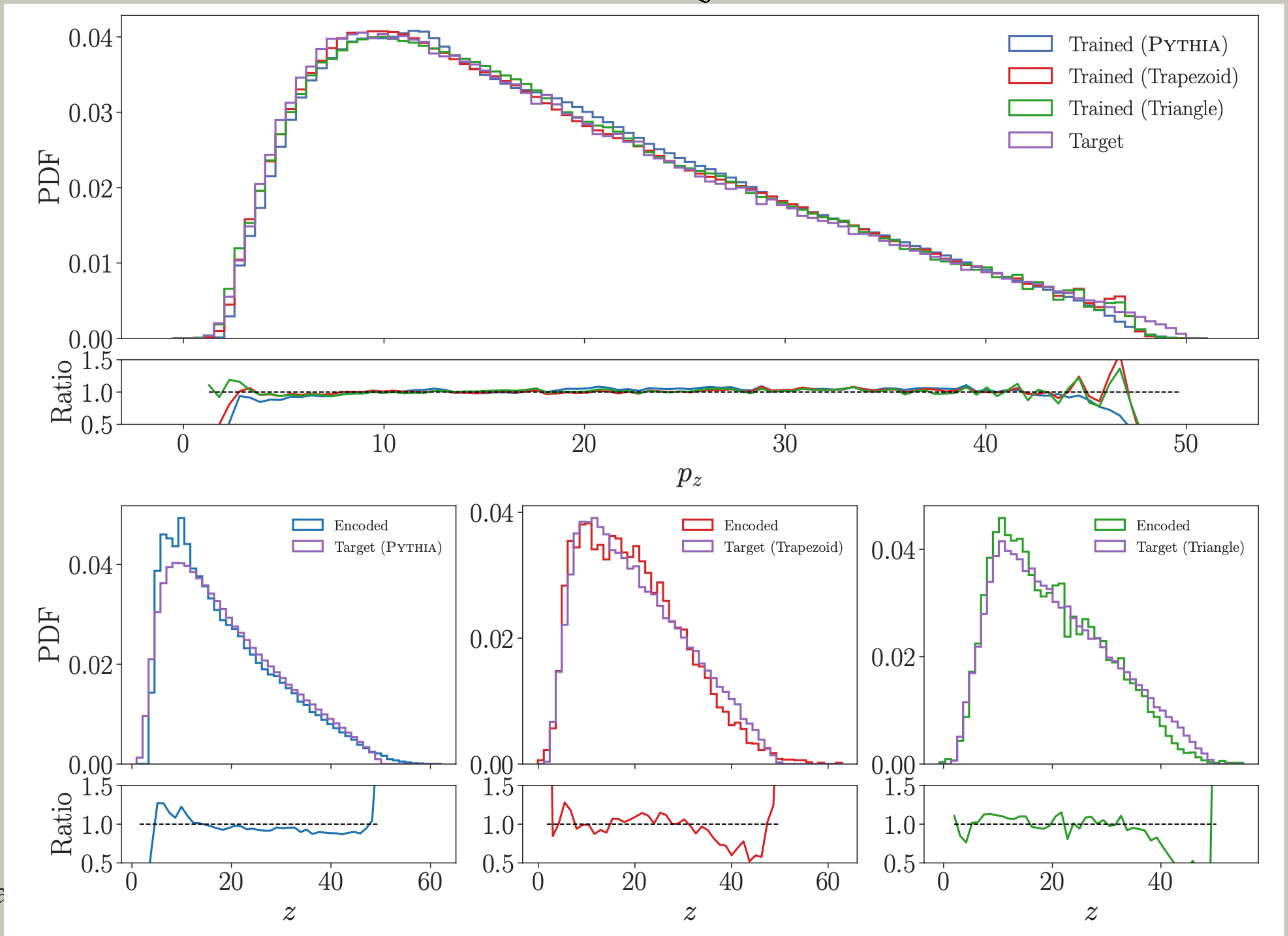
L_{SW} vs. L_{rec}

# of SW slices

Variable x	Target z	t (epochs)	d_z	λ	L
p'_z	PYTHIA	150	35	35	15
	Trapezoidal	300	2	20	30
	Triangular	150	2	30	25
p_T	PYTHIA	100	20	30	30
	Skew-norm	120	4	20	25
	Triangular	120	4	15	25

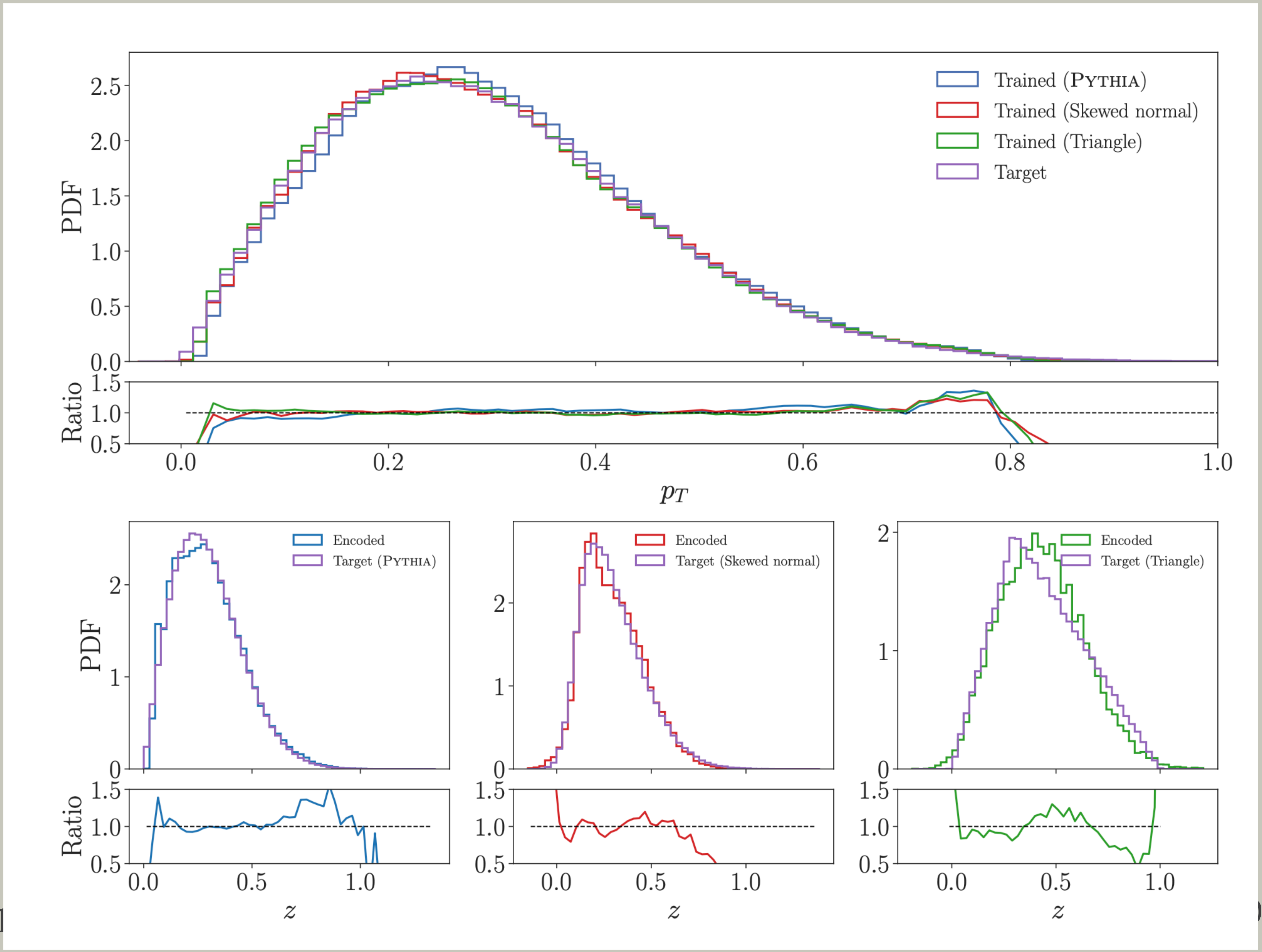
RESULTS - FIRST EMISSION

- MLhad generated p_z distribs.



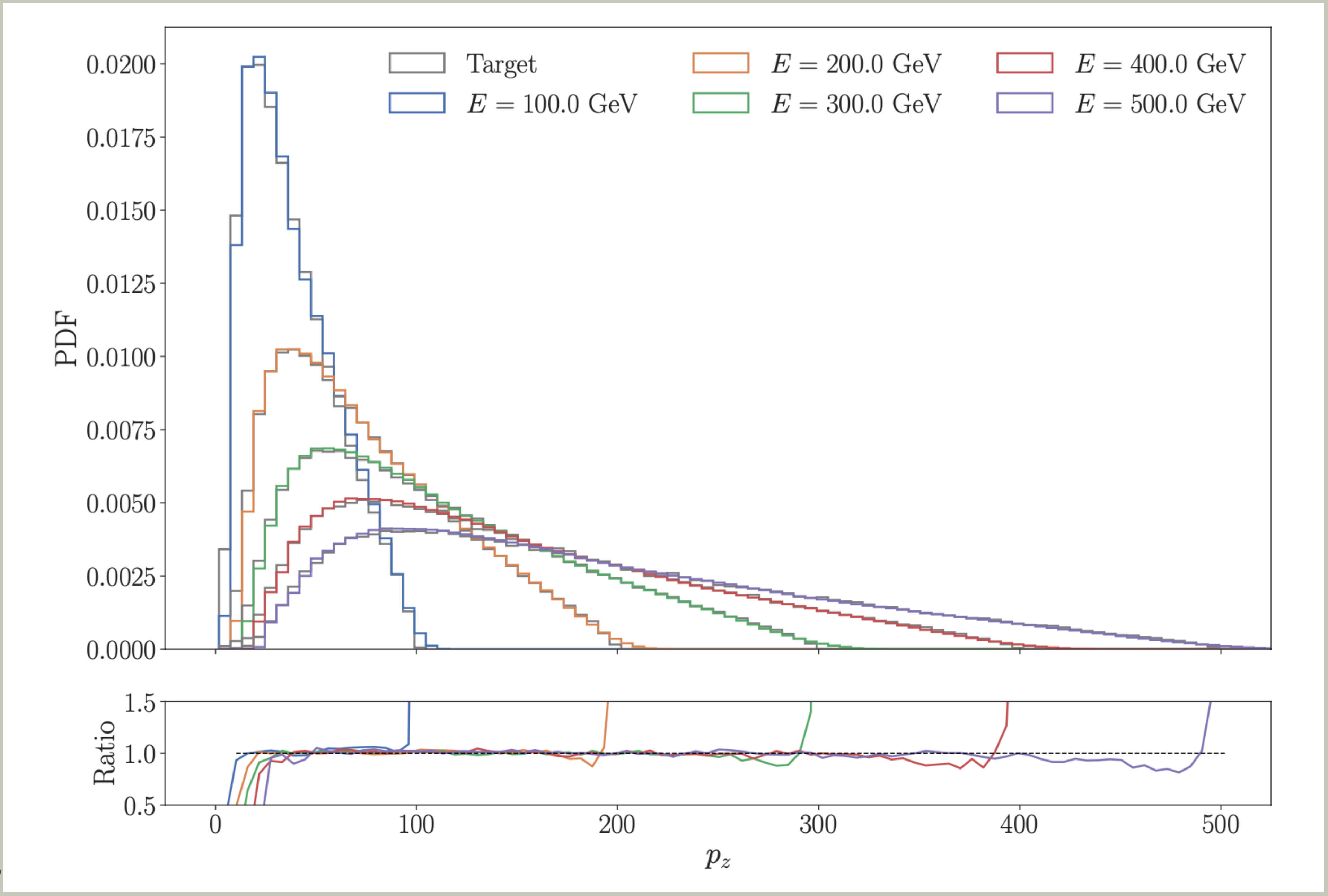
RESULTS - FIRST EMISSION

- MLhad generated p_T distribs.



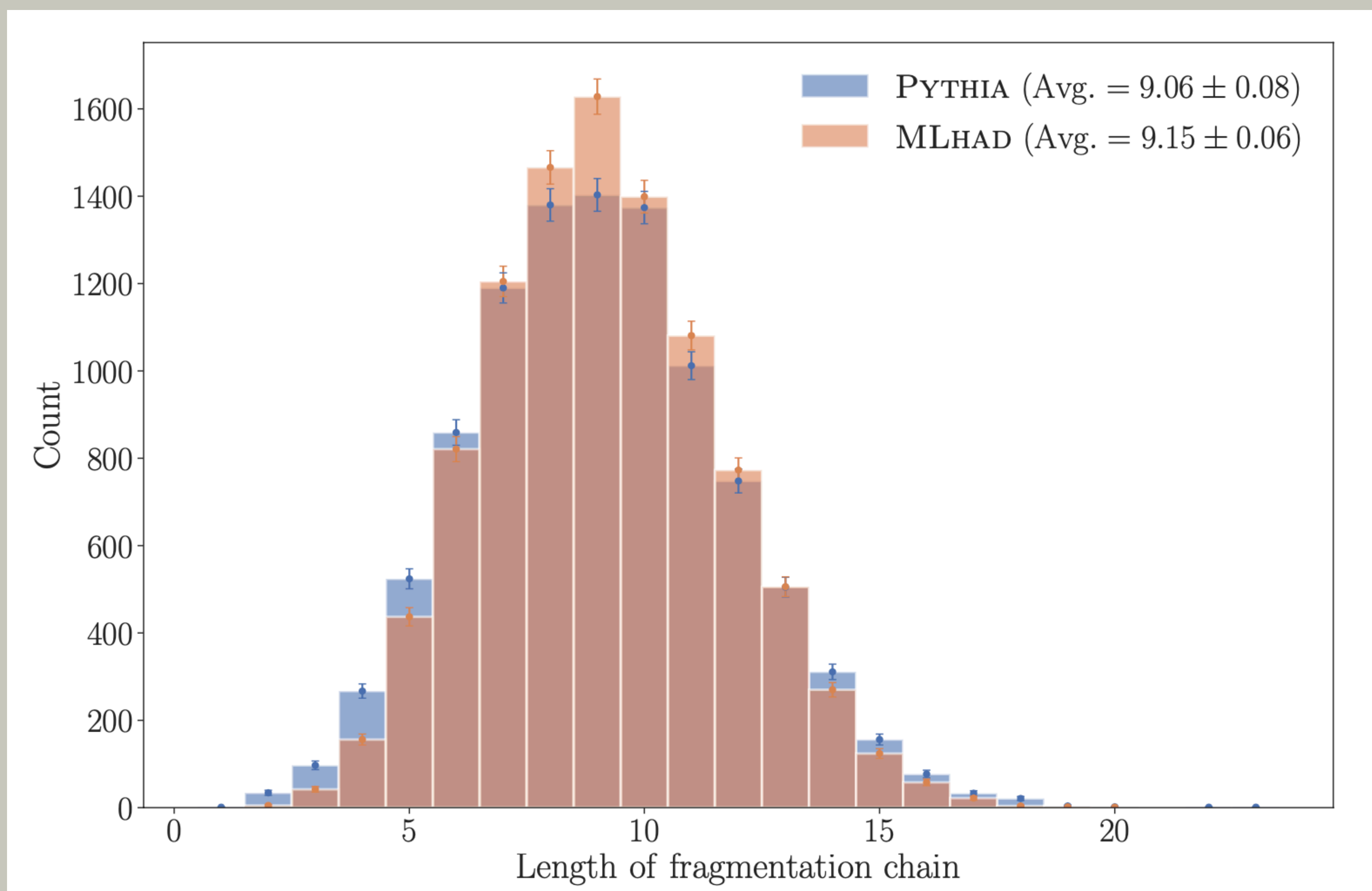
E dependent distributions

- train on first hadron emissions at $E = \{5, 30, 700, 1000\}$ GeV
- generate at a different set of string energies



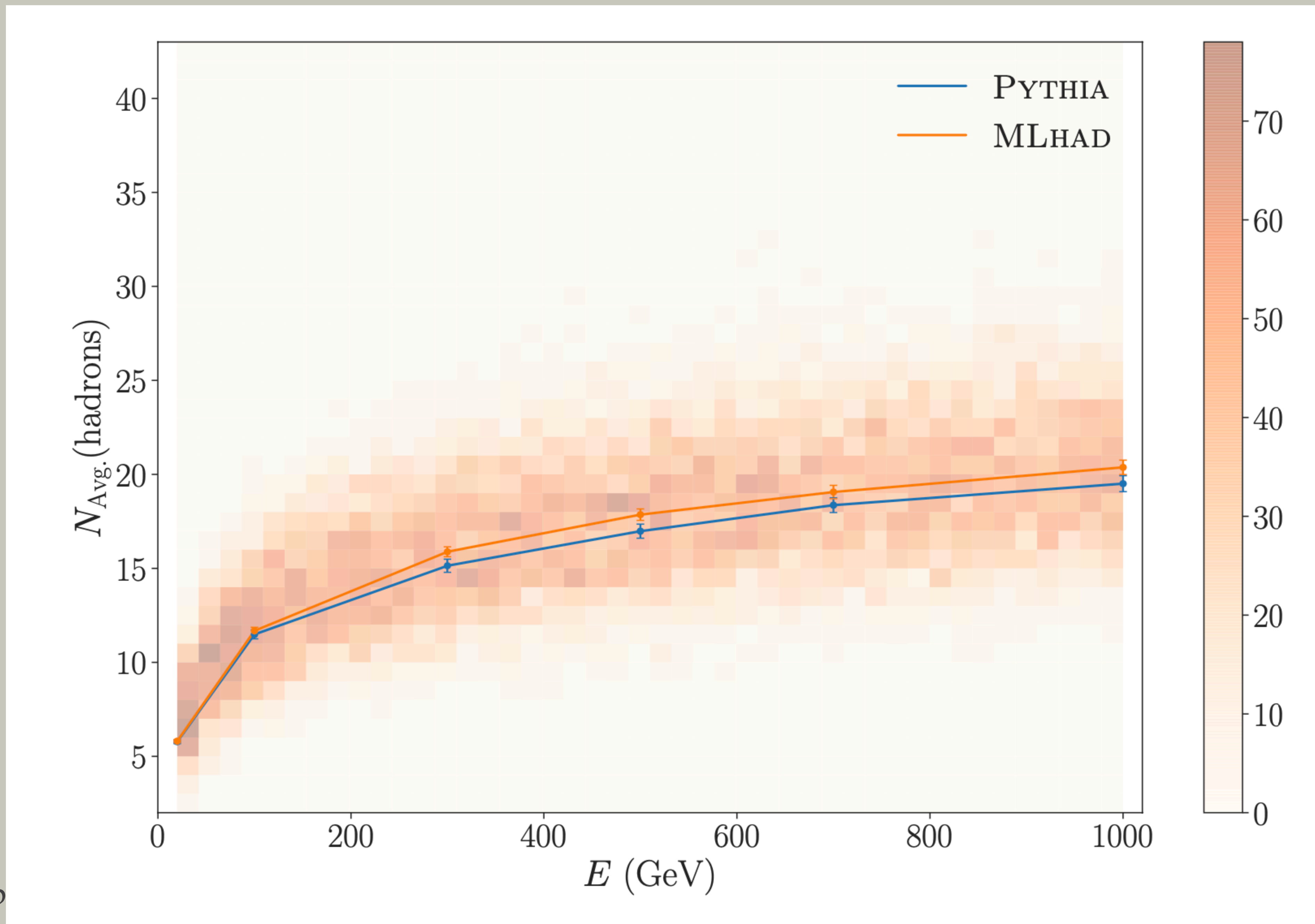
GENERATING HADRONIZATION CHAINS

- number of hadrons produced in hadronization of 50 GeV string



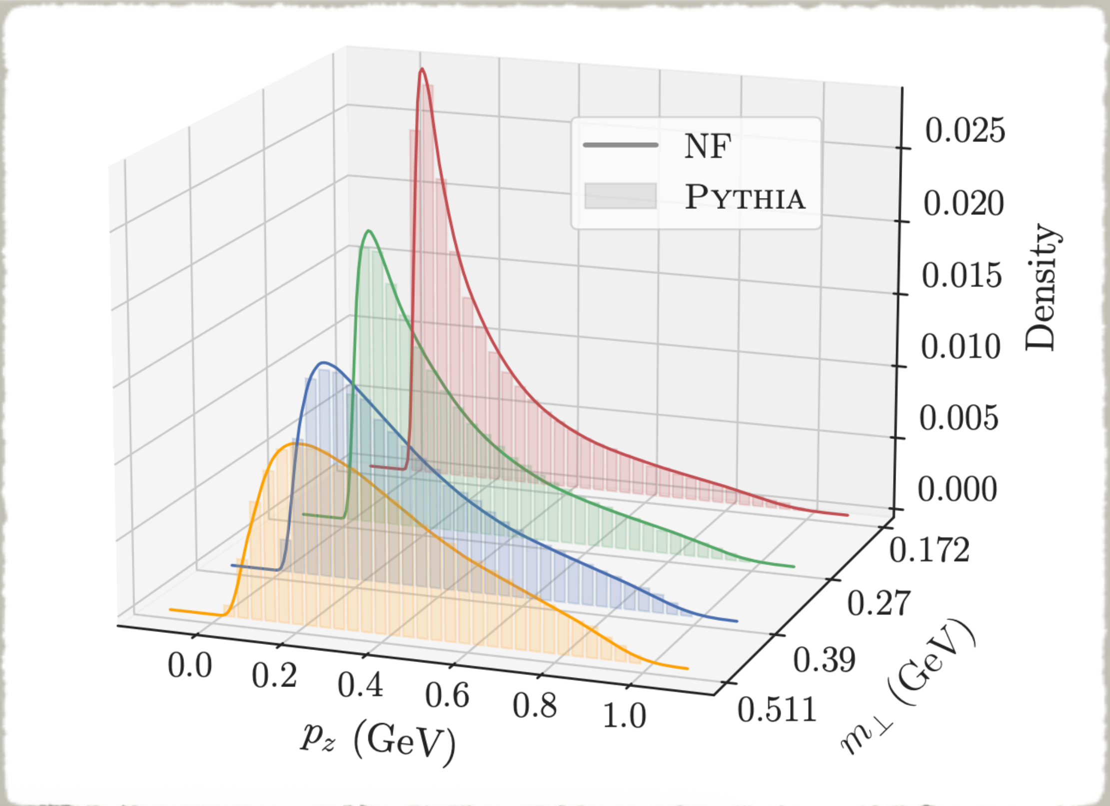
GENERATING HADRONIZATION CHAINS

- the distributions match over a range of string energies



STRING FRAGMENTATION FUNCTION

- in c.o.m. frame of the string:
  - hadron emission described by p_z, p_T + uniform distrib. in azimuthal angle
- p_z distribution from Lund string fragmentation funtion



$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(-\frac{b\, m_\perp^2}{z}\right)$$

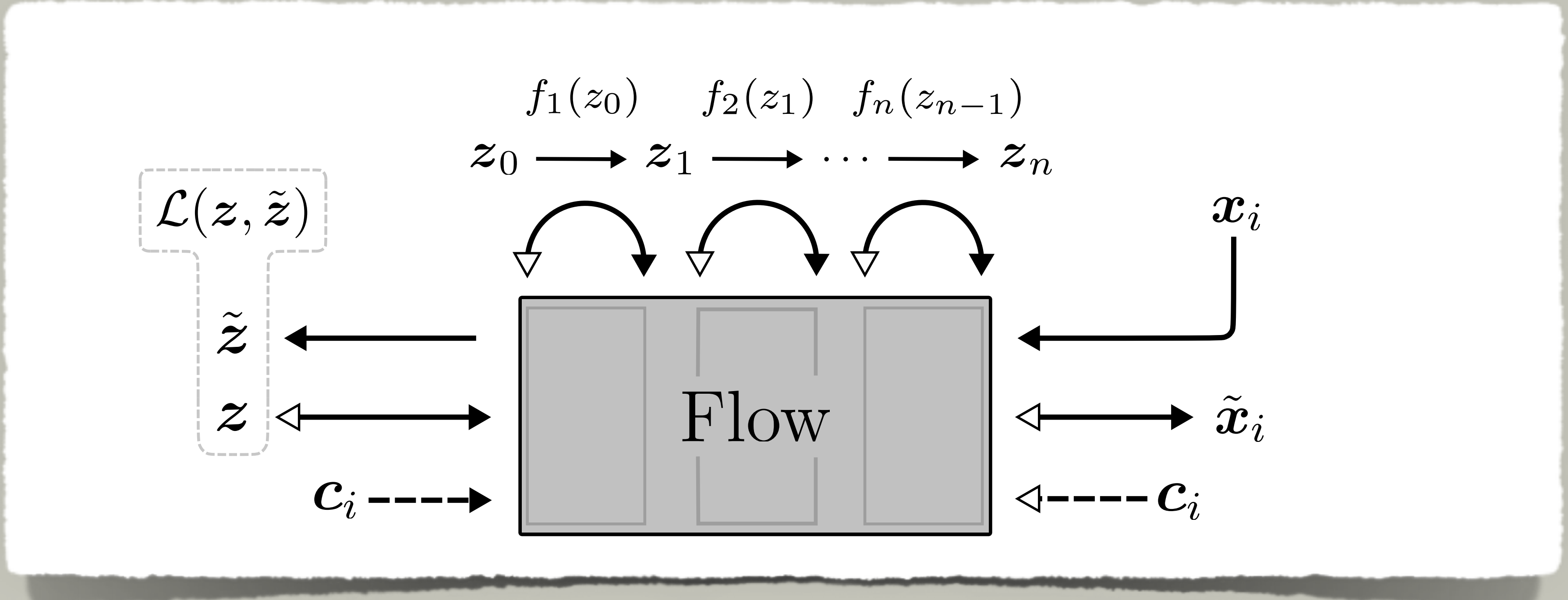
$$z = (p_{h,z} + E_h)/E_{\text{string}}$$

$$m_\perp^2 \equiv m^2 + p_T^2$$

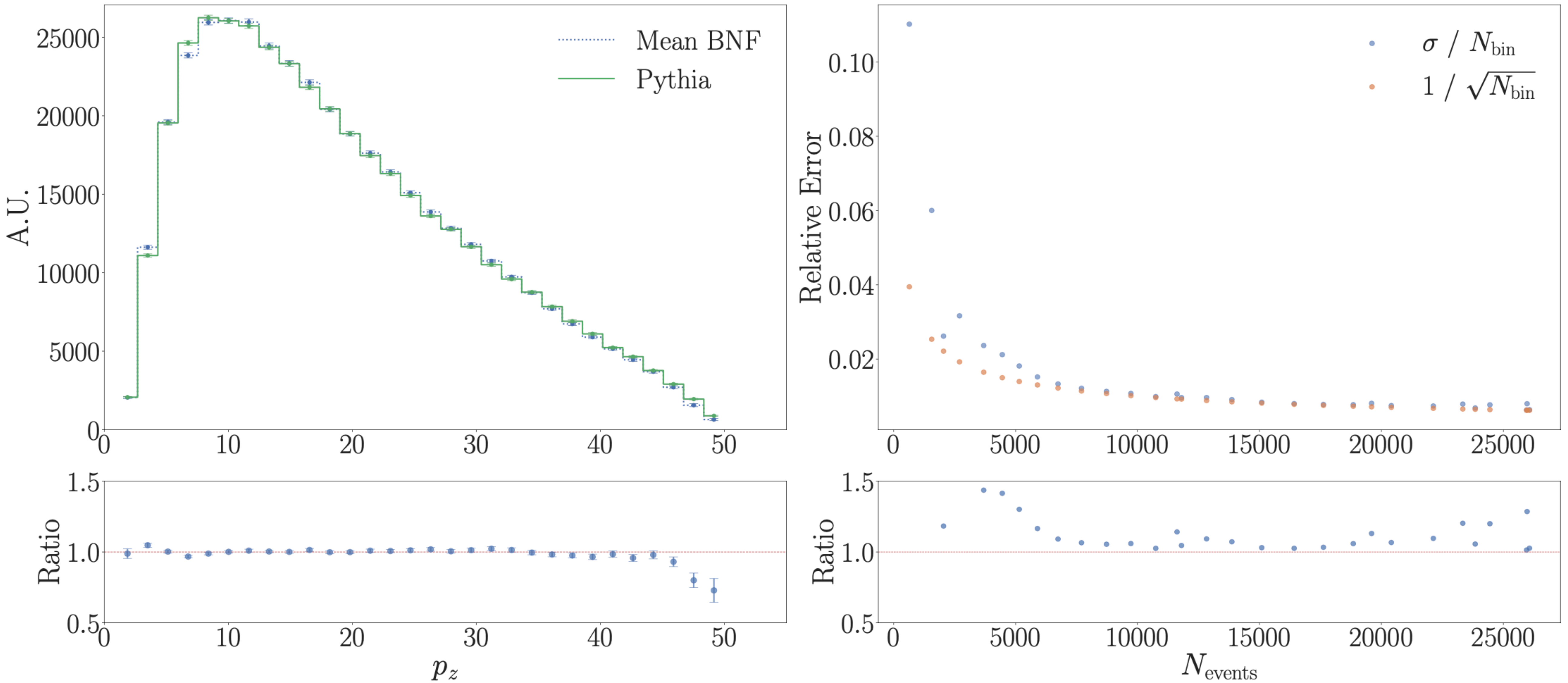
- for light quark flavored hadrons only three params.: a,b and mass, m
- p_T from random Gaussian distributions; width another param.

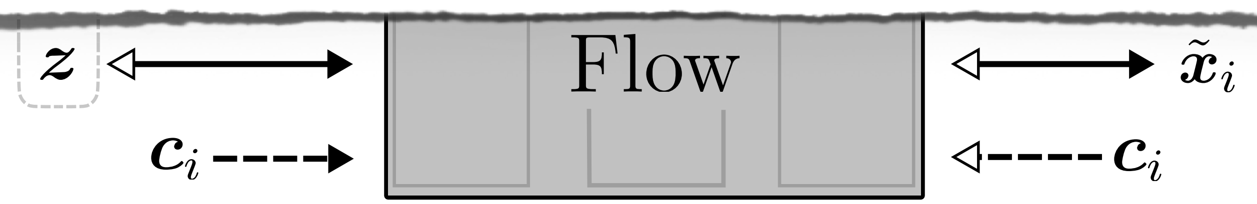
MLHAD NF

- NF: analytic transform. from latent to feature space
  - for us feature space 2D: $\mathbf{x}_i = (p_z, p_T)$



- BNF: NN params. θ are normal random vars.
  - mean and widths trained such that statist. errors reproduced

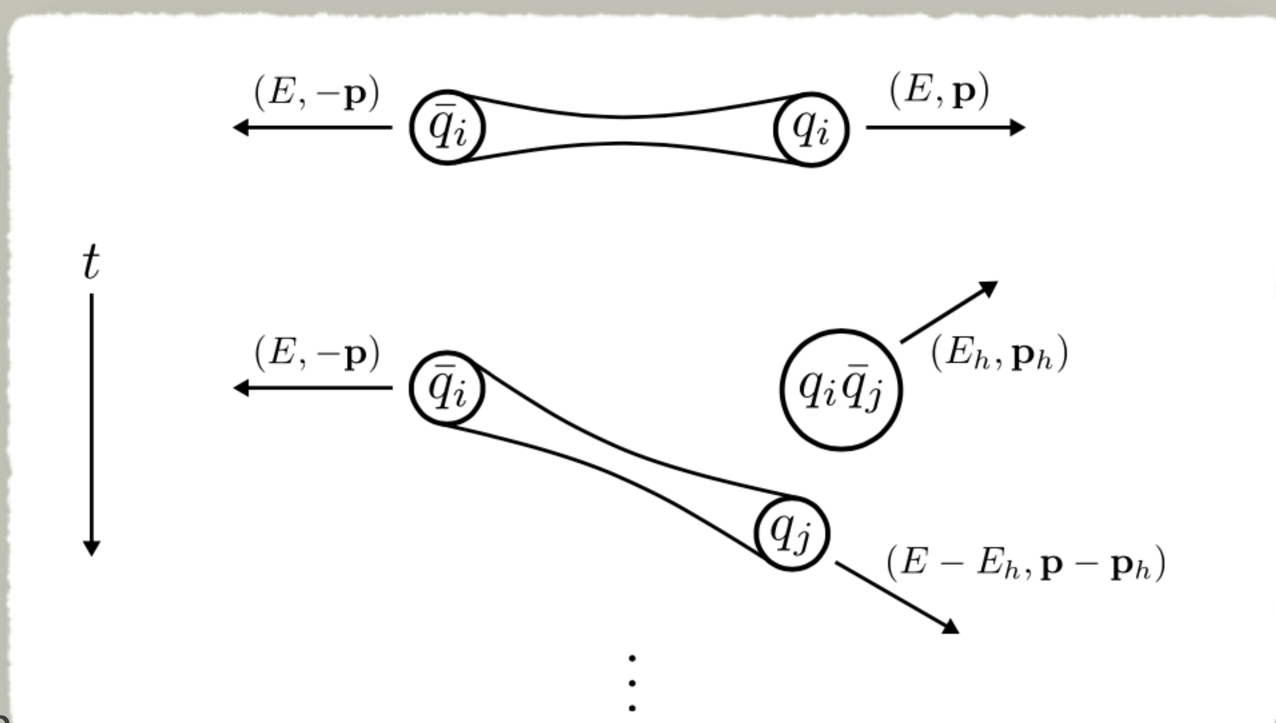




- BNF: NN params. θ are normal random vars.
  - mean and widths trained such that statist. errors reproduced

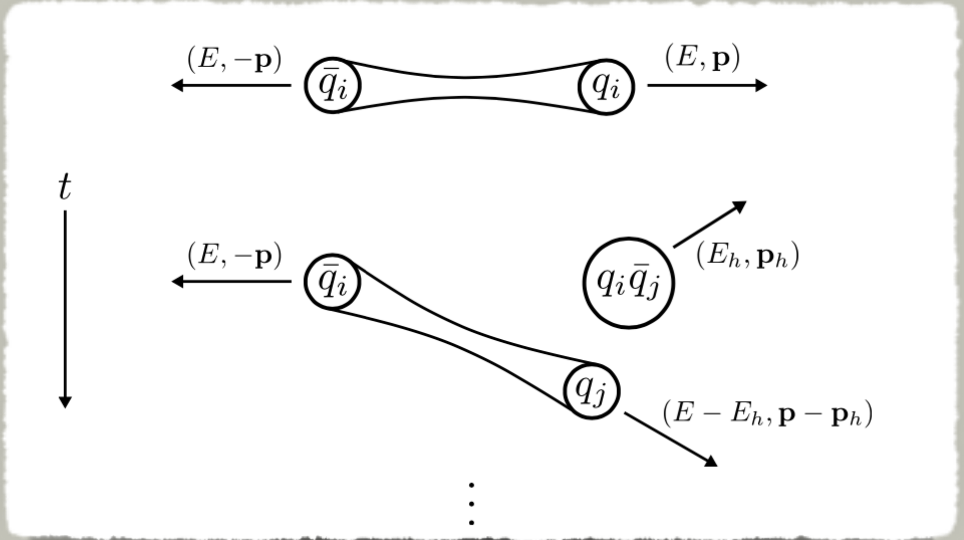
SIMPLIFIED STRING HADRONIZATION MODEL

- assume that color flow done correctly by Pythia
  - including splitting gluons, so that only strings with $q, \bar{q}$ ends
- want to reproduce first hadron emission from a string piece
  - the whole hadronization chain is then reproduced by iterating
  - the string is labeled by $q, \bar{q}$ flavor and its energy in cms, $2E$
- only u,d quarks, uses Pythia flavor selector
- have an IR cut-off of 25 GeV, at which hadronization chain terminates

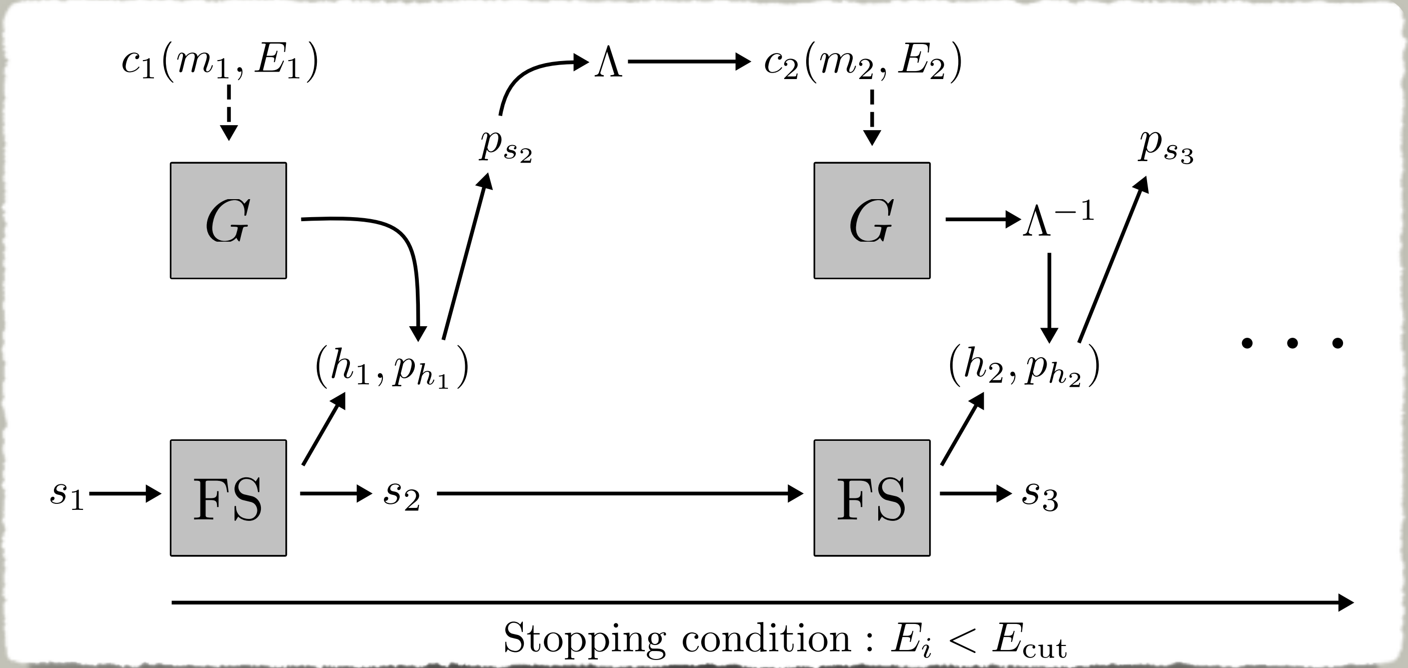


MLHAD NF

- when used as a generator repeat single emissions in c.m.s.'s +boosts

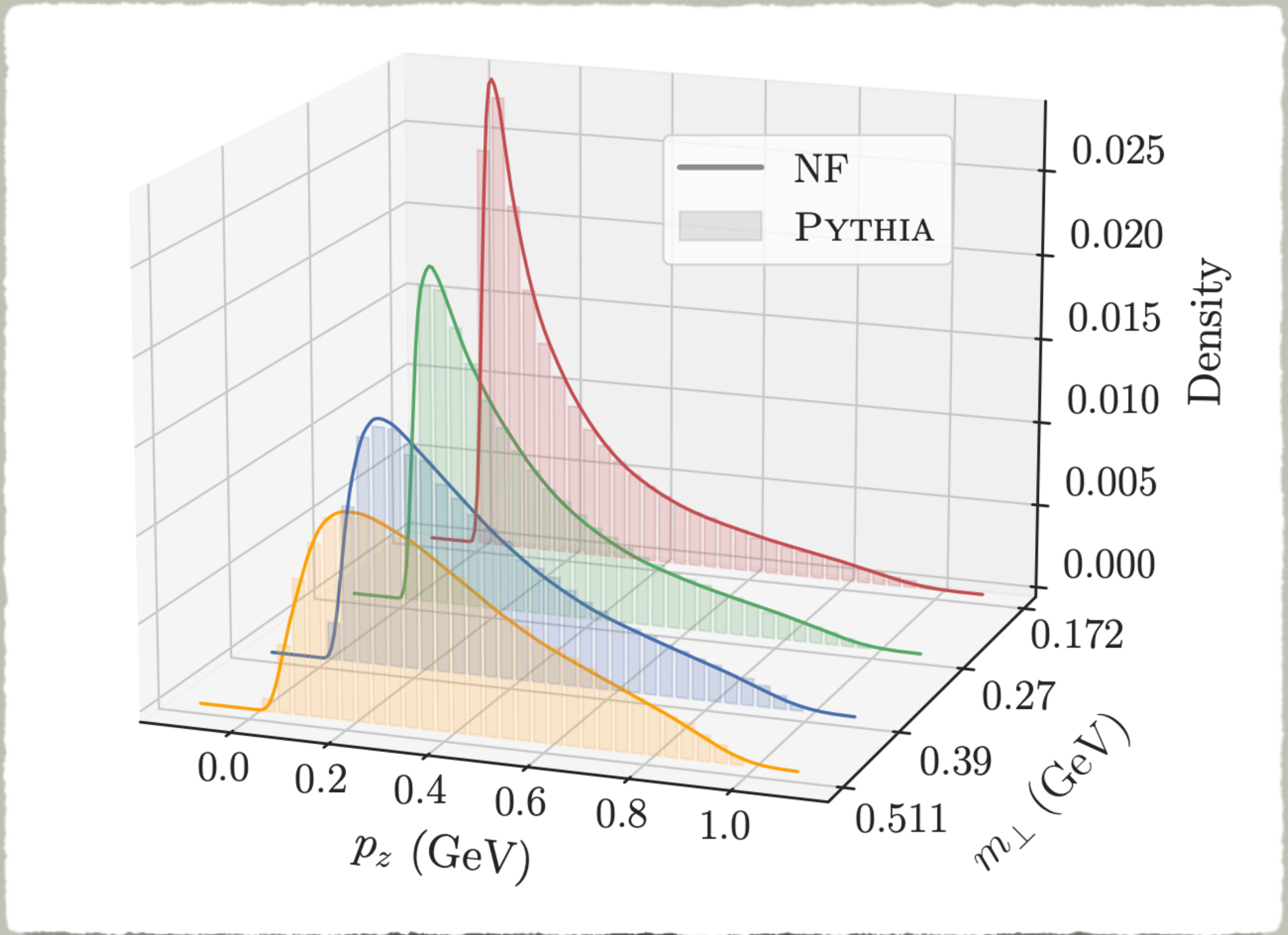


- in general conditioned on string eng. (c_i), hadron mass / flavor (s_i)



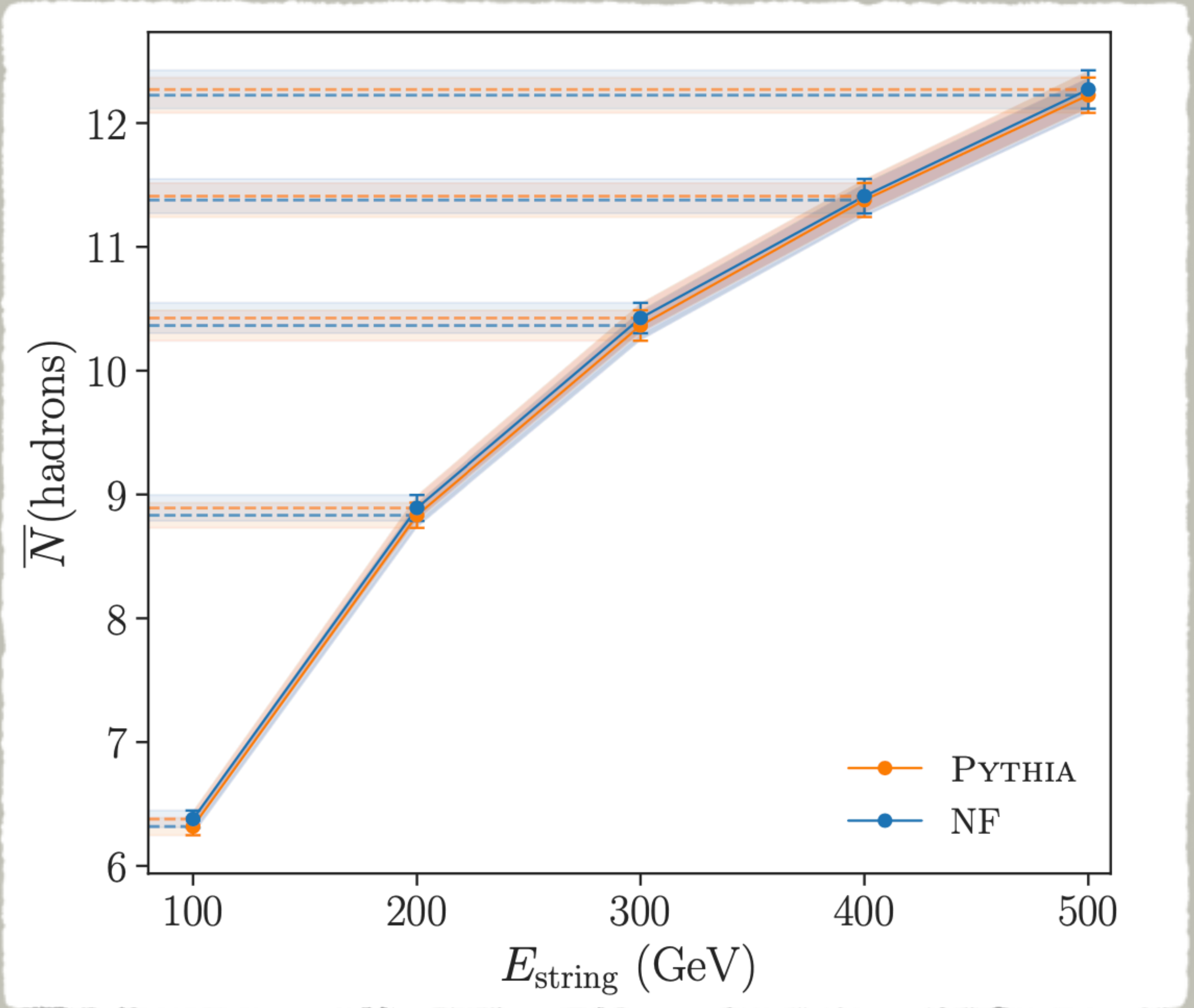
MLHAD NF

- single hadron emissions well reproduced
  - note: in simplified Pythia string model



MLhad NF

- charge multiplicities well reproduced
  - note: in simplified Pythia string model

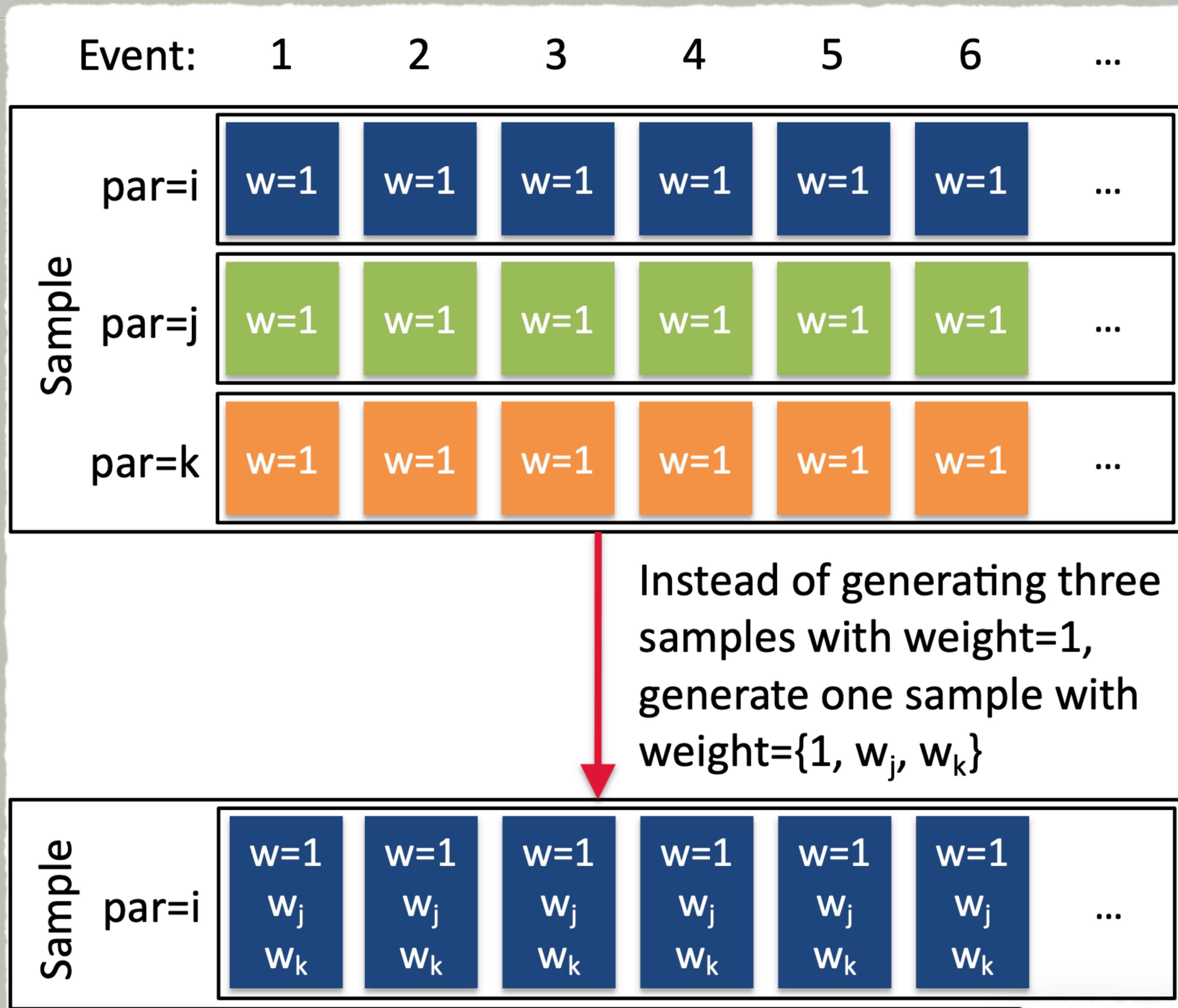


REWEIGHTING HADRONIZED PYTHIA EVENTS

Bierlich et al [MLhad], 2308.13459

- event generation is time-consuming
  - want to reweight events without regenerating
- in Pythia the Lund string fragm. function sampled via standard accept/reject algorithm
  - if rejected instances are kept ⇒ a modified accept/reject algorithm
  - ⇒ new event weights for diff. hadronization params.

REWEIGHTING HADRONIZED PYTHIA EVENTS

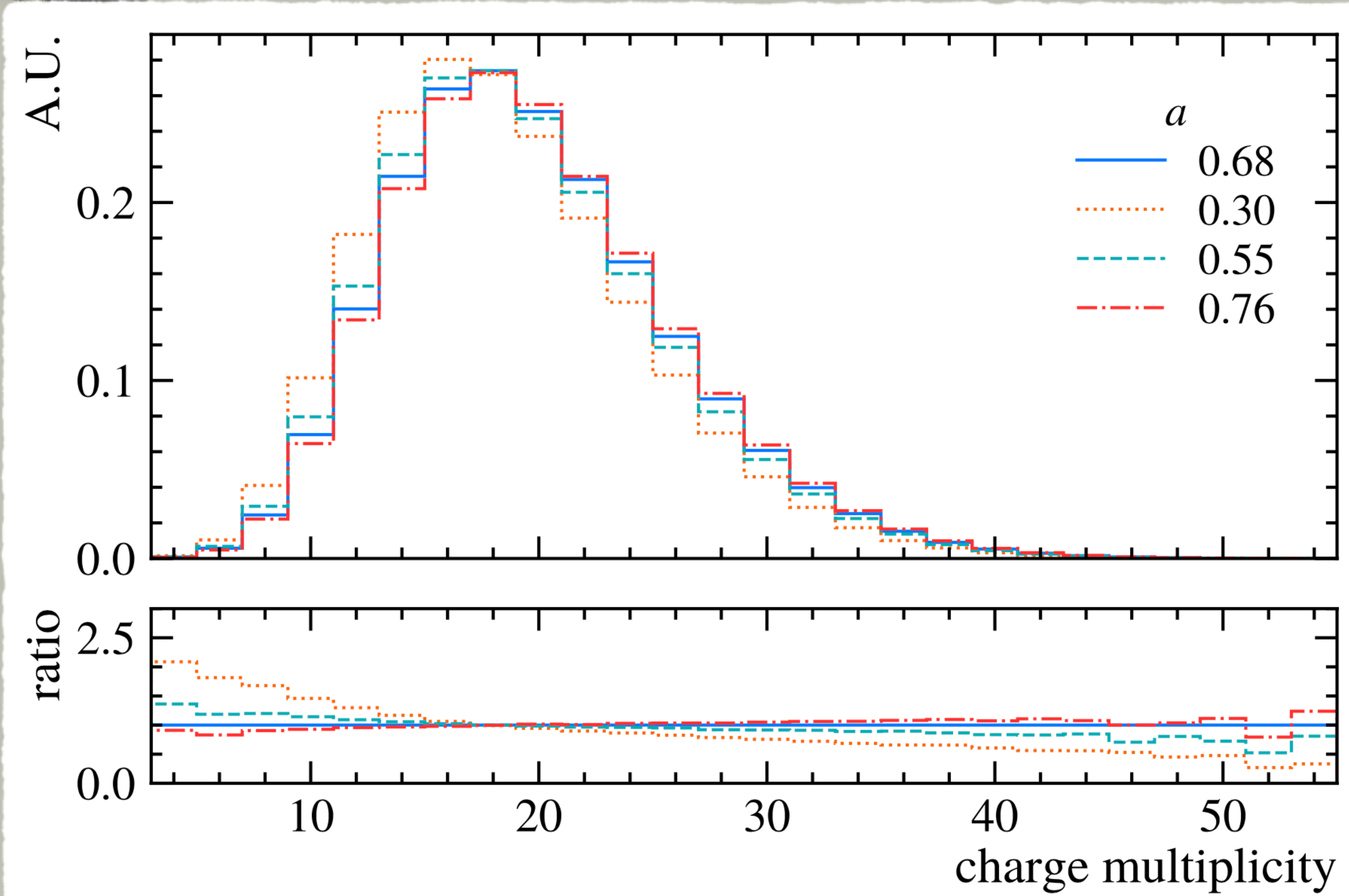


graphics by M. K. Wilkinson

REWEIGHTING HADRONIZED PYTHIA EVENTS

$e^+e^- \to Z \to \text{jets}$

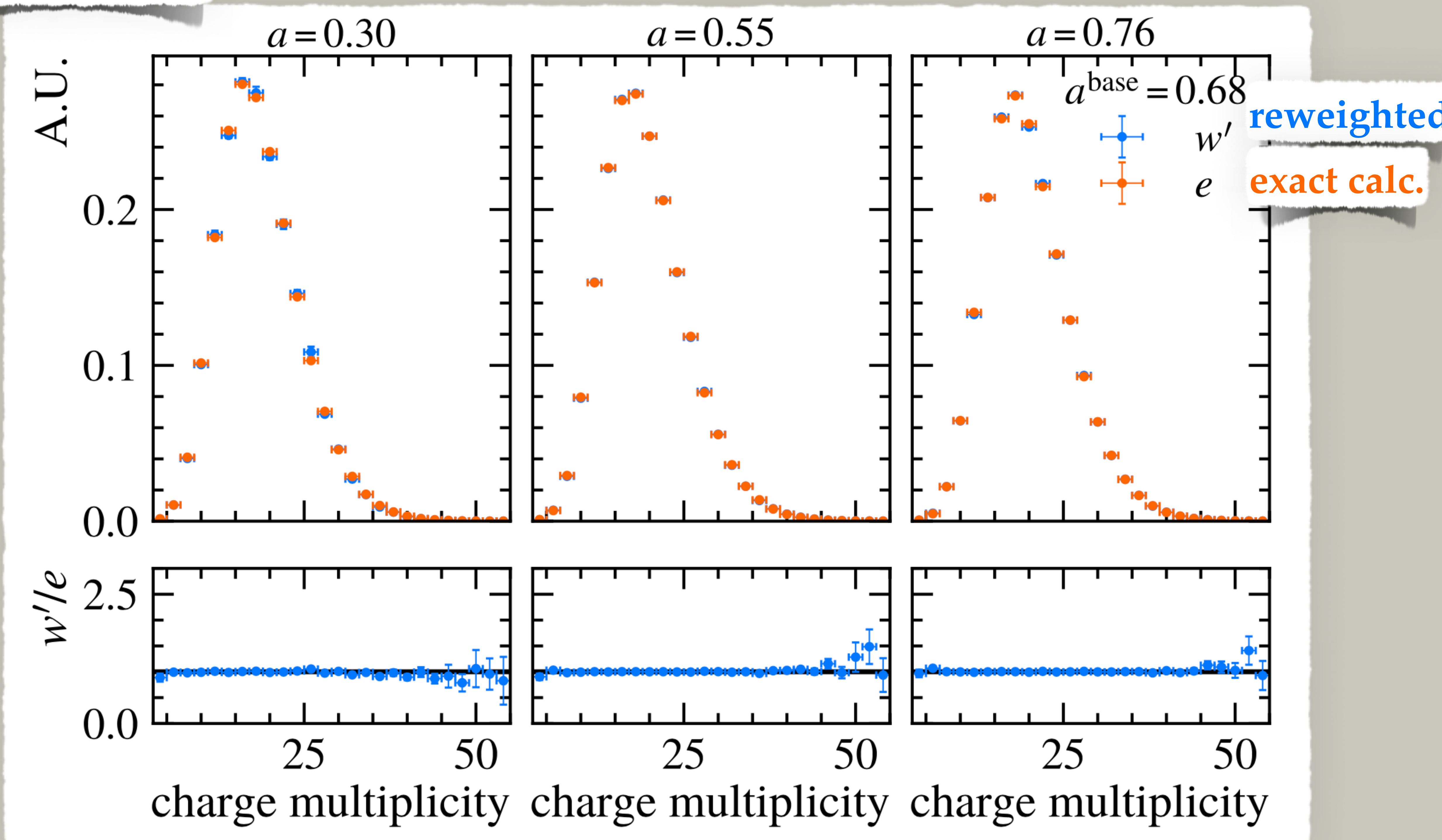
Bierlich et al [MLhad], 2308.13459



REWEIGHTING HADRONIZED PYTHIA EVENTS

$e^+e^- \to Z \to$ jets

Bierlich et al [MLhad], 2308.13459



REWEIGHTING HADRONIZED PYTHIA EVENTS

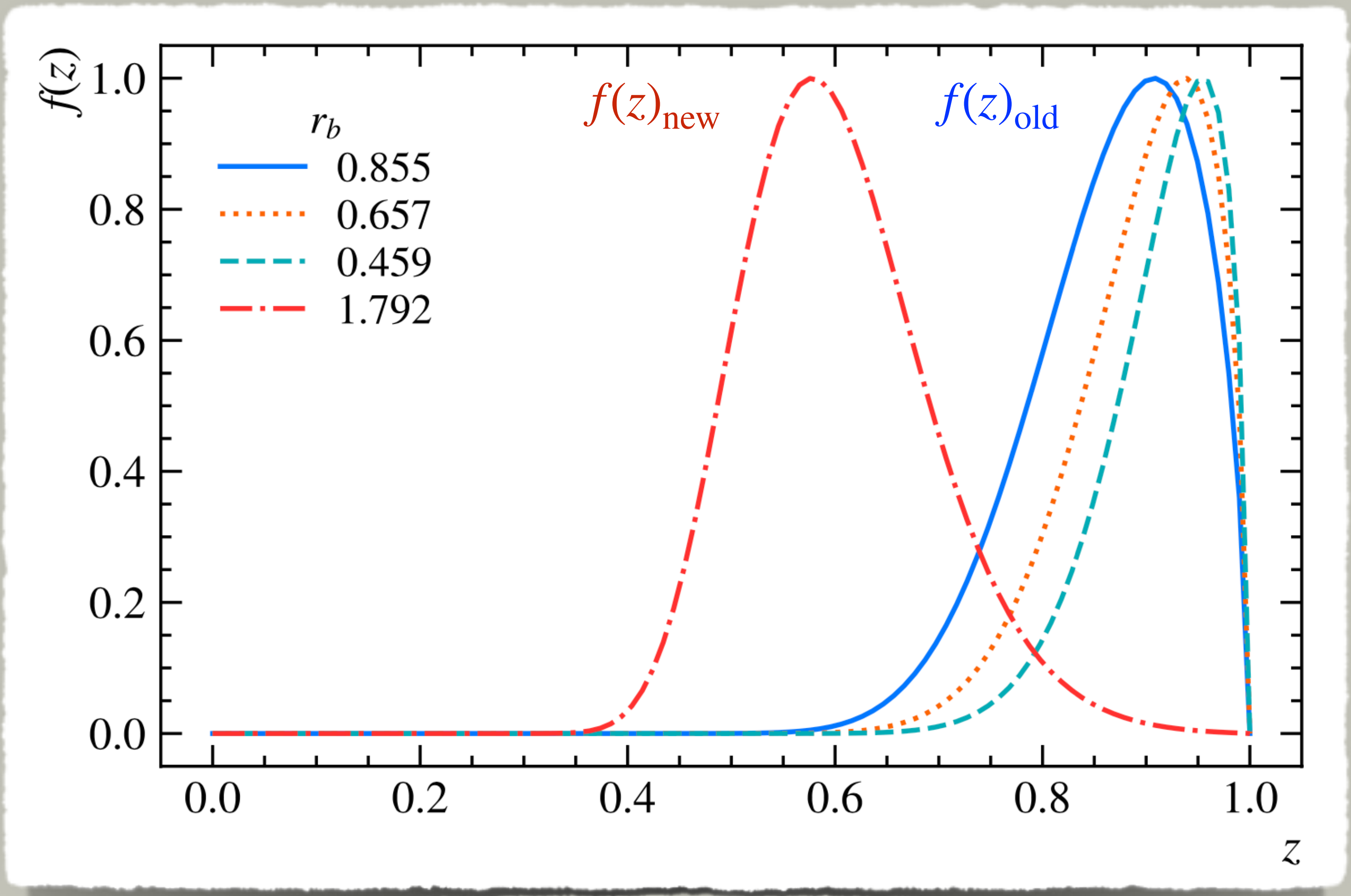
Bierlich et al [MLhad], 2308.13459

- implemented for
  - a, b Lund string fragmentation params.
  - also for heavy flavor param. r_b, and the width parameter for Gaussian sampling of p_T
- for full detector simulations can expect up to several orders of magnitude speed-ups
  - if many variations of hadroniz. params are needed, e.g., in m_t measurements
- caveat: new and old $f(z)$ need to have large enough overlap/area of support

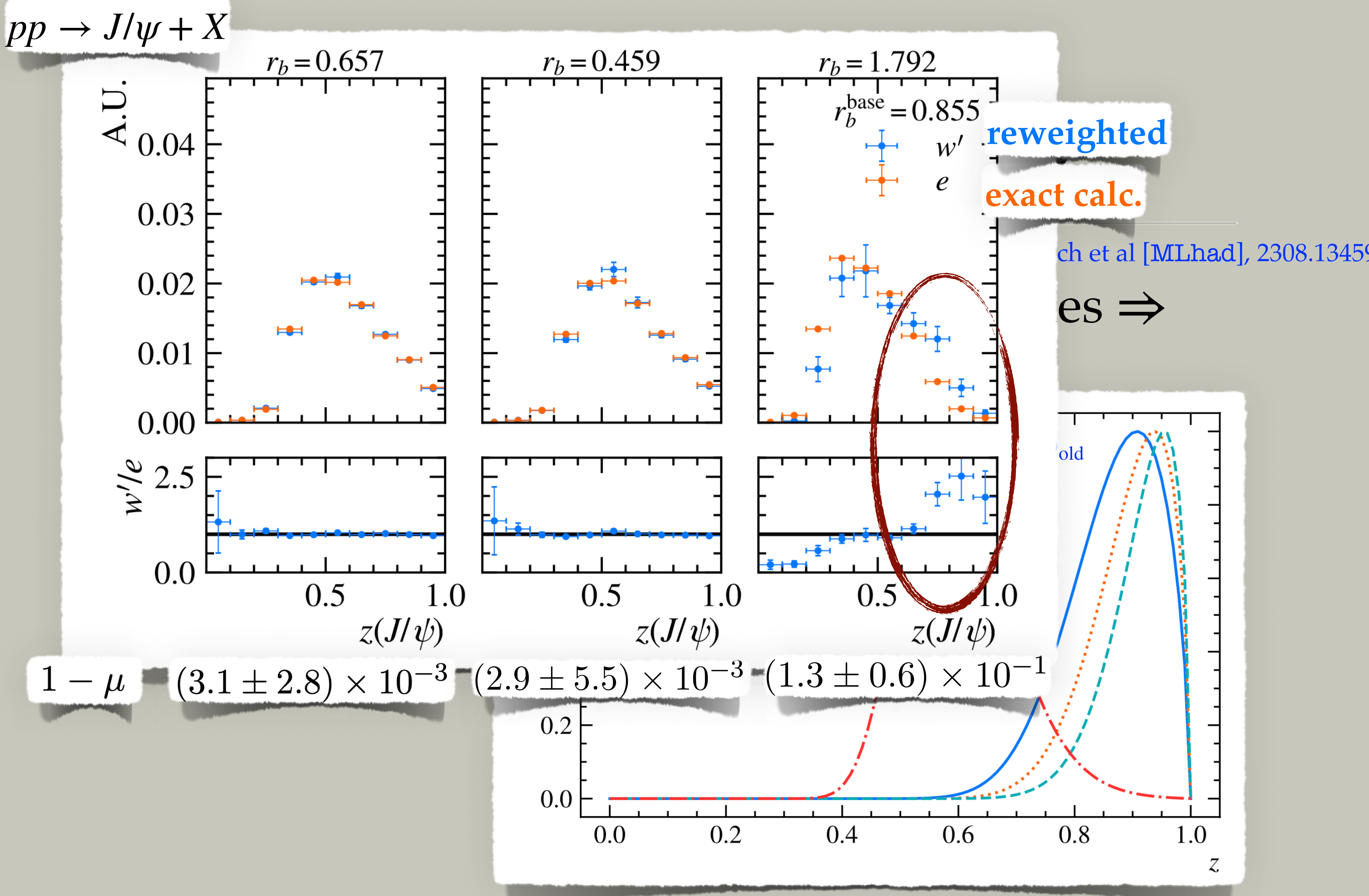
CAVEAT: LARGE PARAM. VARIATIONS

Bierlich et al [MLhad], 2308.13459

- if $f(z)_{\text{new}}$ nonzero where $f(z)_{\text{old}}$ ~vanishes ⇒ large errors



- mean weight μ can be a useful diagnostics tool

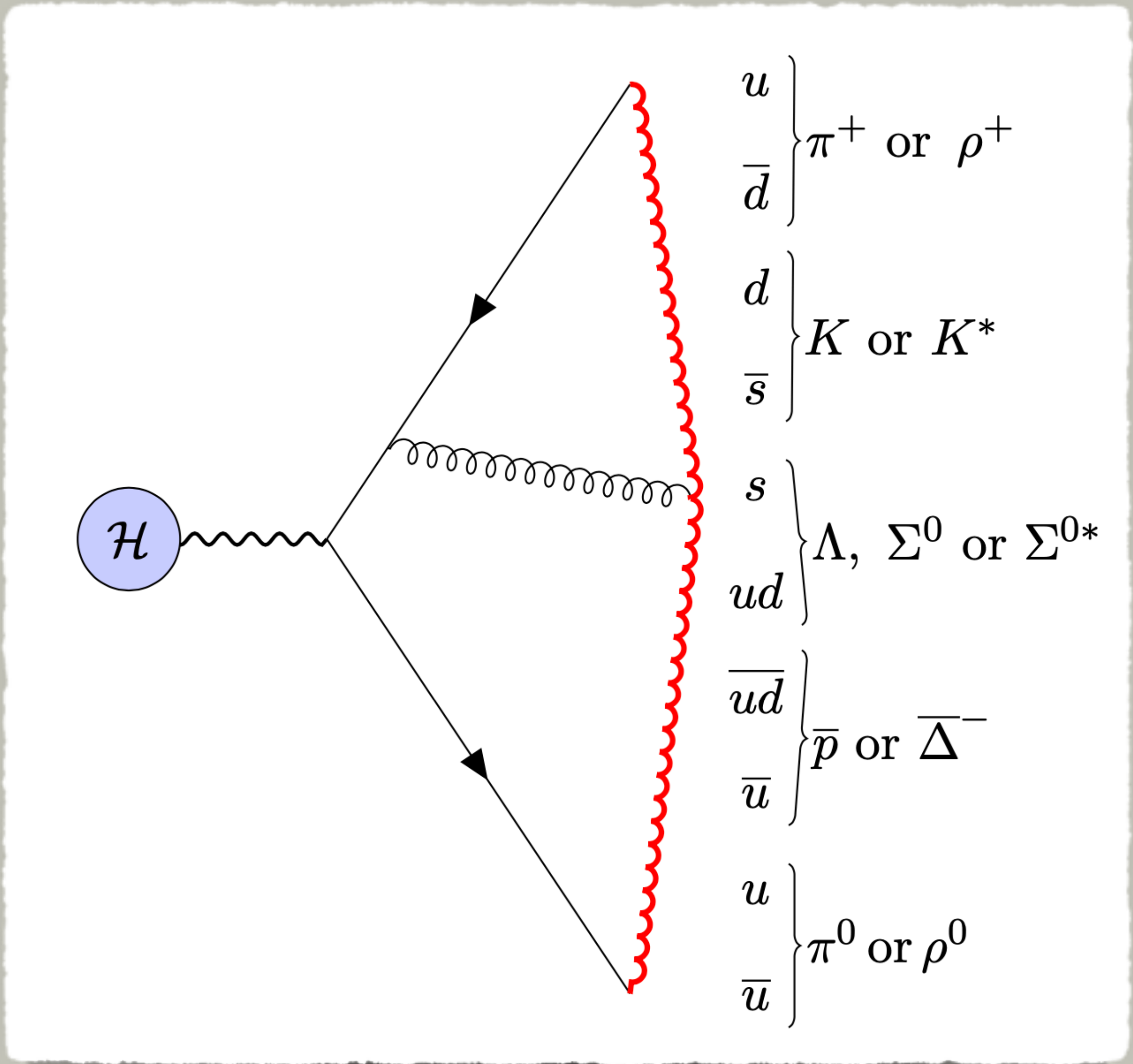


- mean weight μ can be a useful diagnostics tool

FLAVOR REWEIGHTING

Bierlich et al [MLhad], 2411.nnnnn

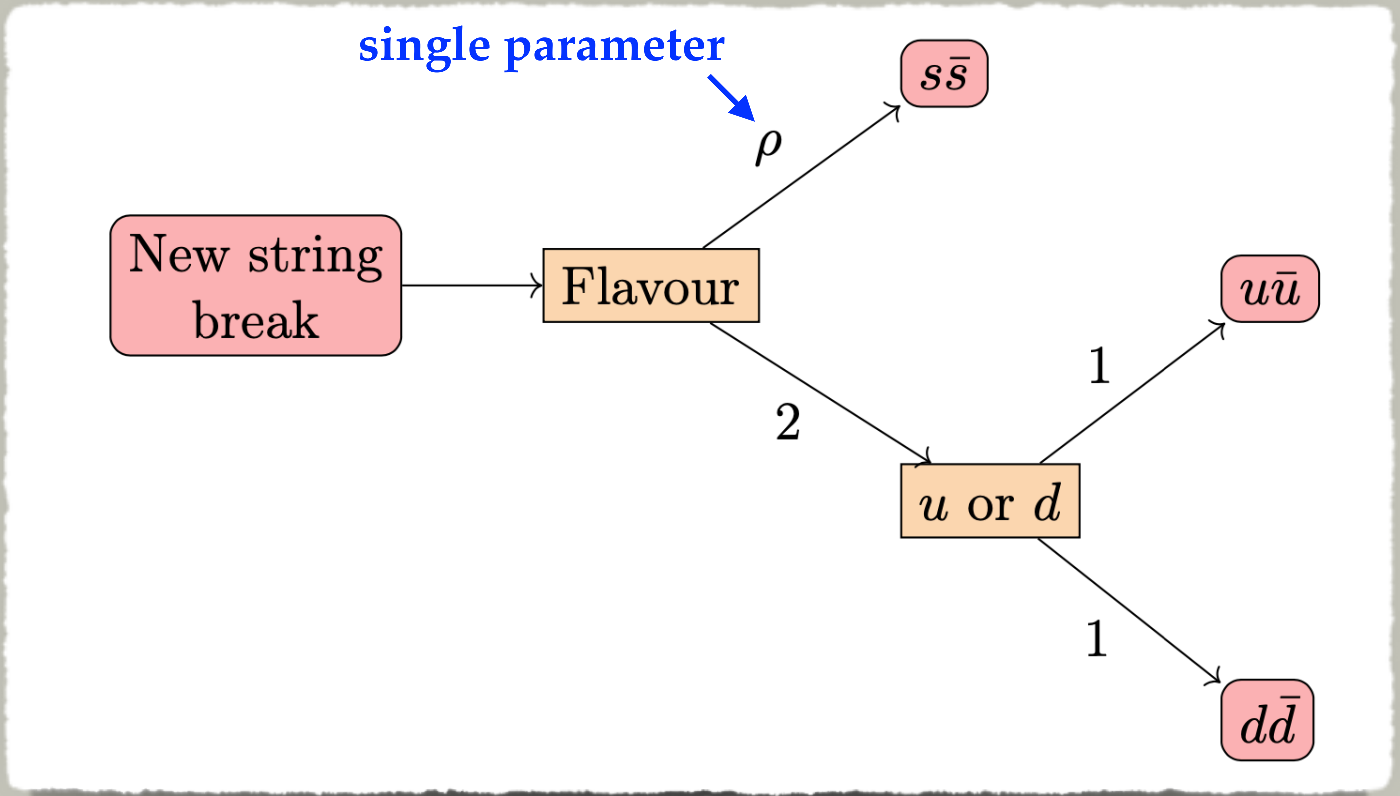
- next step
  - reweighting for flavor parameter variations in Pythia hadronization
- baryons make everything more complicated



FLAVOR REWEIGHTING

Bierlich et al [MLhad], 2311.nnnnn

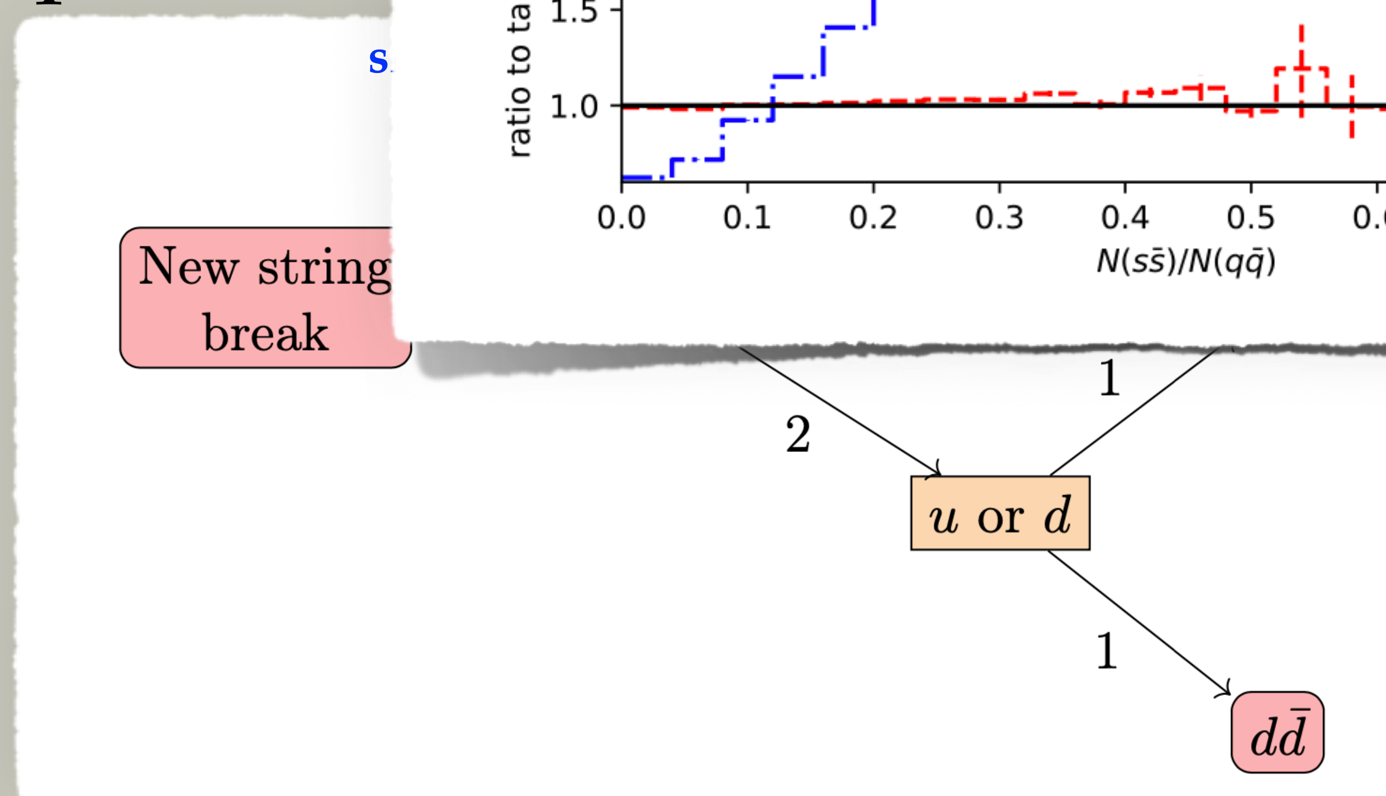
- simple flavor decision flow for string breaks if no diquarks

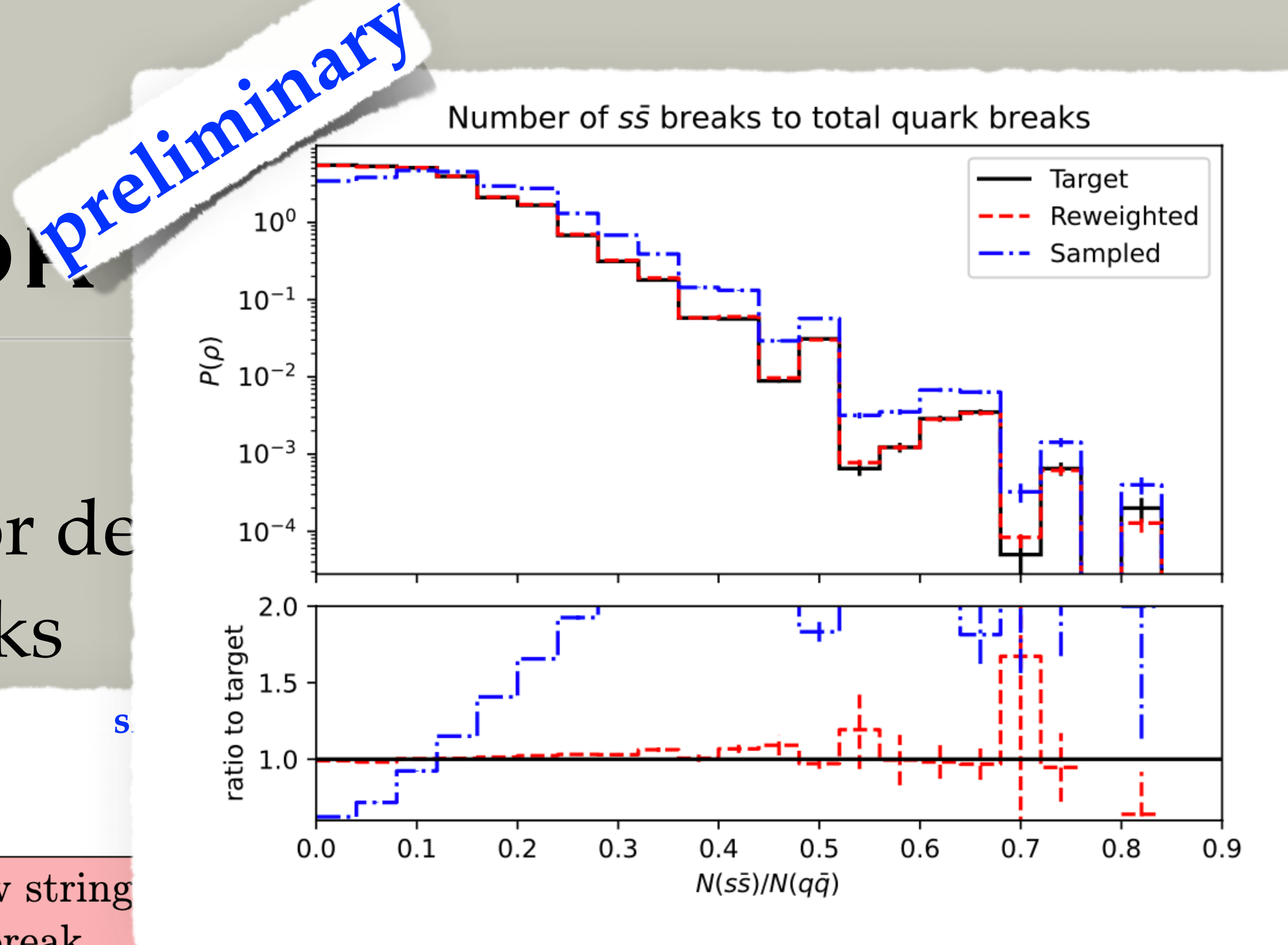


- becomes much more involved with diquarks

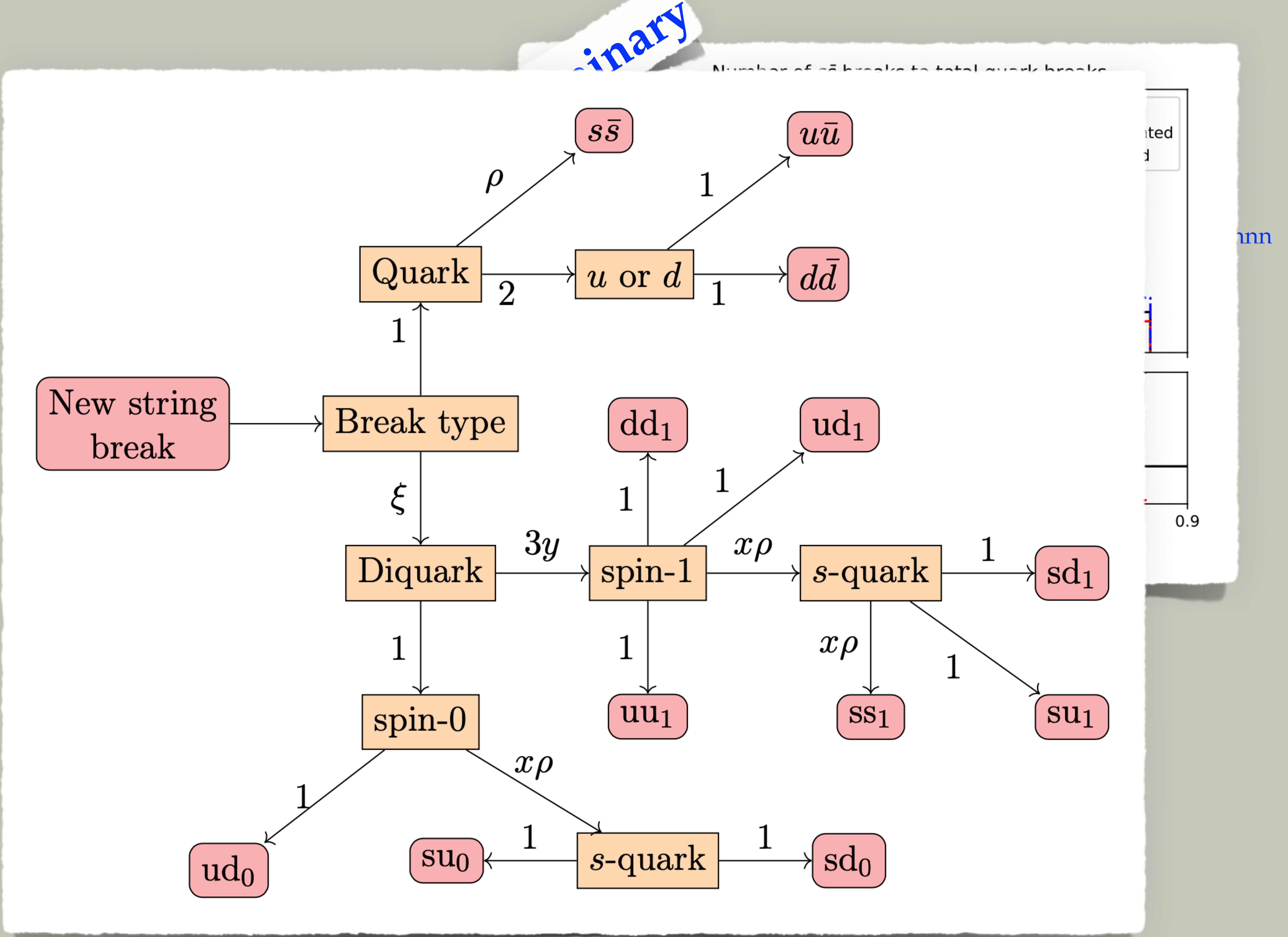
FLAVOR

- simple flavor de
  if no diquarks





- becomes much more involved with diquarks


New string break
Break type
Quark
Diquark
u or d
$s\bar{s}$
$u\bar{u}$
$d\bar{d}$
spin-1
spin-0
s-quark
s-quark
dd$_1$
ud$_1$
uu$_1$
sd$_1$
ss$_1$
su$_1$
ud$_0$
su$_0$
sd$_0$
ρ
ξ
$3y$
$x\rho$
1
2
0.9