Machine Learning for Hadronization

ML4Jets 2026
18/08/2026

Manuel Szewc

ICAS

MLHAD

https://uchep.gitlab.io/mlhad-docs/

Instituto de
Ciencias Físicas
ICIFI-ECYT_UNSAM-CONICET

Plan of the talk

1. What is hadronization?
2. Why do we care about it?
3. How do we tune a model?
4. ML-based hadronization models
5. Perspectives

Disclaimer:
This will be a partial exploration, and I'll skip the details to focus on what is not understood and could be tackled (it's the best stuff!)

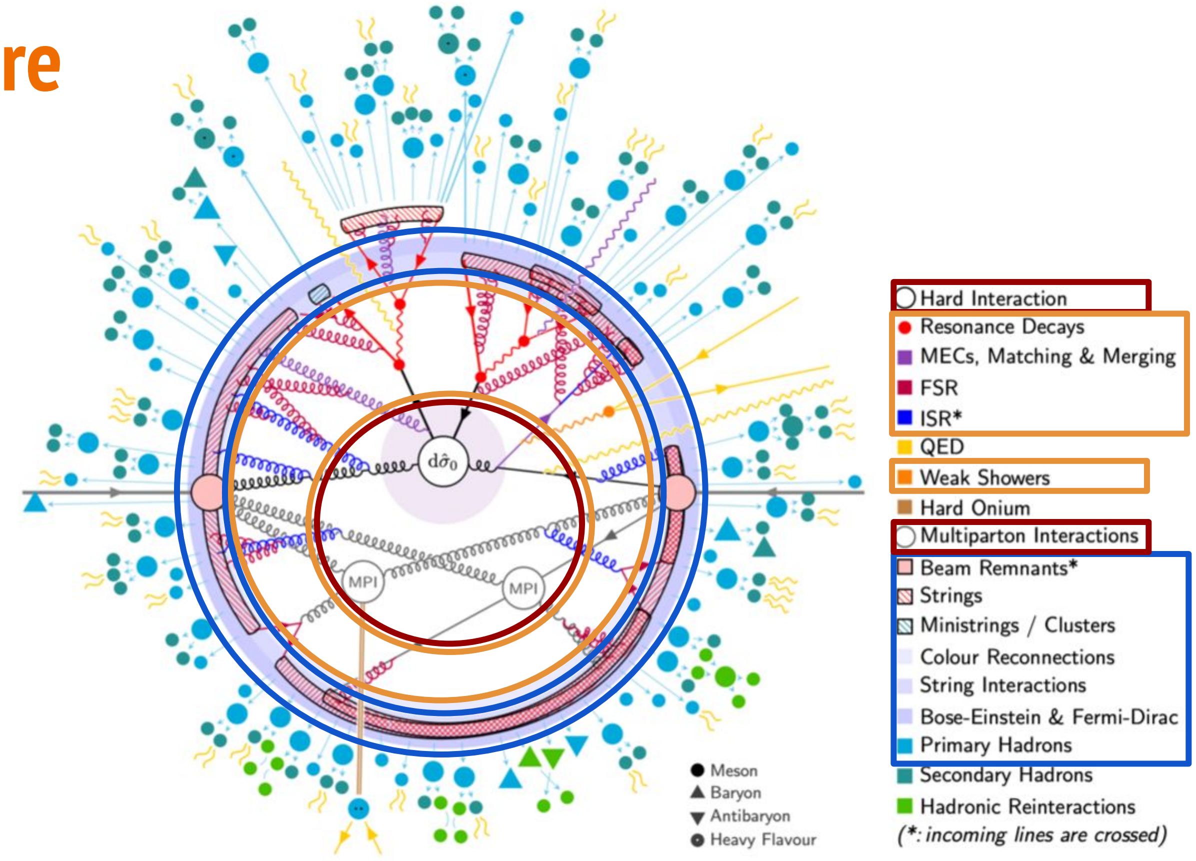
The generator picture

Image from Pythia 8.3 manual. The radial coordinate is time or 1/energy scale.

Hard process dσ: **perturbatively calculated**, directly related to underlying lagrangian, describes **partons hidden to experiment**.

Shower: **perturbative evolution** of partons from hard to hadronization scale, **hidden to experiment.**

Hadronization: combining partons into **measurable hadrons**, non perturbative → **Empirical models**.

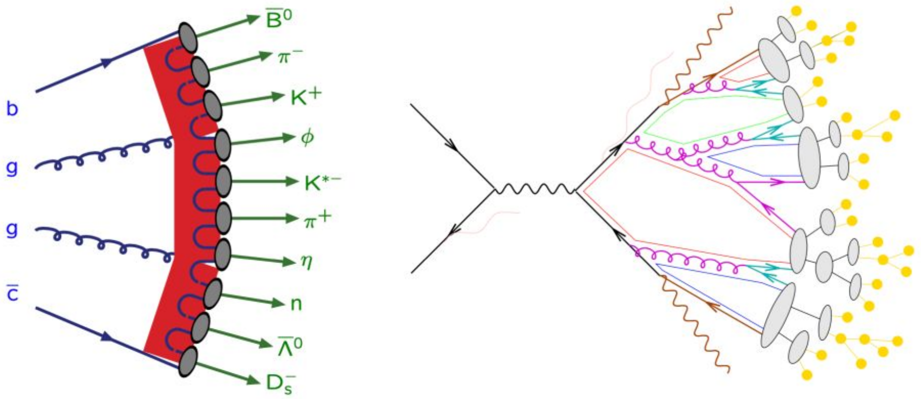


What is hadronization?

Two main empirical models: **string** and **cluster**. Same starting point: **the final stage of QCD shower**.

- Stop shower at some scale $\mathbf{Q_0}$.
- In large $N_c \to \infty$ limit, only **planar graphs**
- Group q, q̄ and g in **QCD singlets** (cluster/string pieces)
- Gluon decay / Gluon as kinks
- Either "**color preconfinement**" or "**leading color dipoles/strings**"

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 with modifications

What is hadronization?

Inclusive fragmentation functions exist! D(z) defined in analogy to Parton Distribution Functions.

How do we **match them with exclusive predictions** from the hadronization models in simulations?

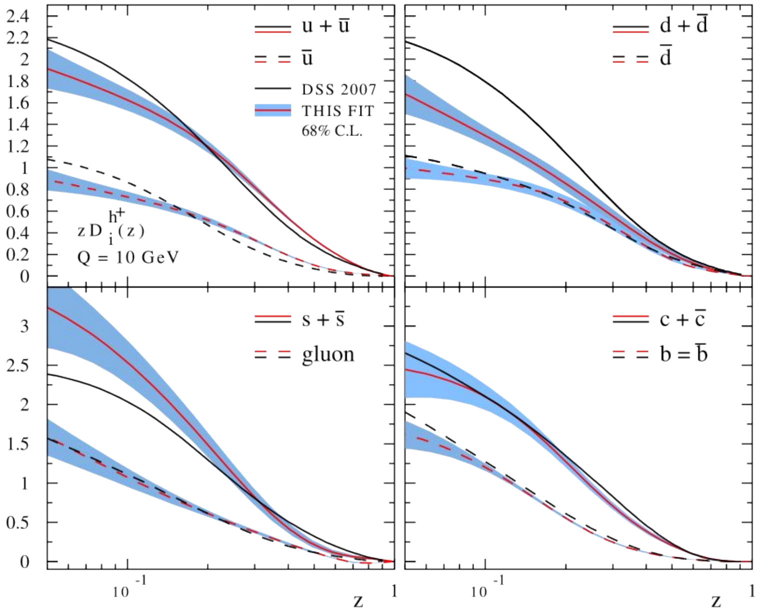


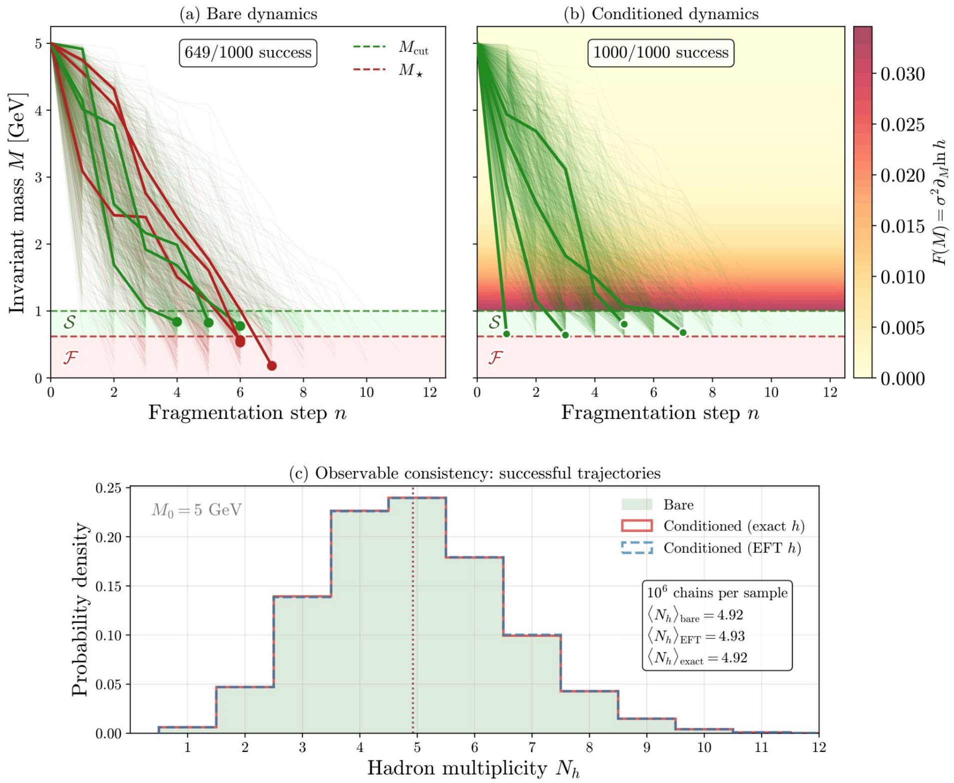
Figure from Borsa et al.

What is hadronization?

Hadronization in generators has several extensions. For the string in Pythia, see *Rope hadronization, String shoving, Thermal string model, non-Gaussian pT tails, Baryon junctions, StringSpinner*. Also, see the recent Effective Hadronization Model by T. Menzo

Non-trivial deviations from the standard linear potential obtained from 1+1 QCD and validated by Lattice. **There is interesting physics in hadronization, let’s not profile it away!**

From T. Menzo

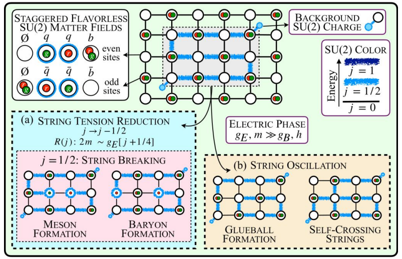


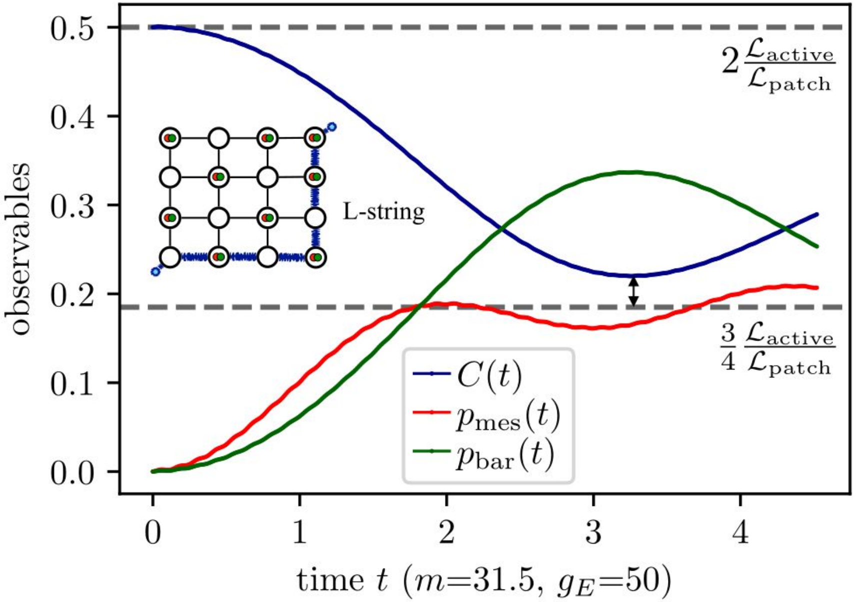
What is hadronization?

But which physics?

Lattice + Quantum Computing input needed (e.g. From String Dynamics to Event Generation with Quantum Simulation workshop). **Can we match Quantum Simulation and Event Generation?**

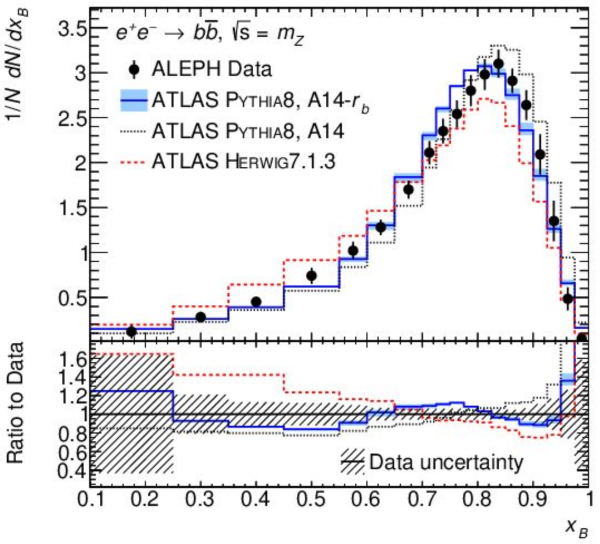
Cataldi et al

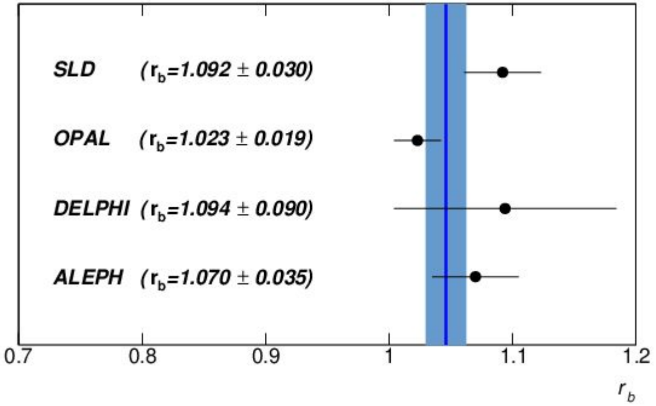




Why care about hadronization?

It impacts the current LHC program. See e.g. this [ATLAS top mass measurement](). Hadronization uncertainties are ~0.2 GeV (~⅓ of total syst.)





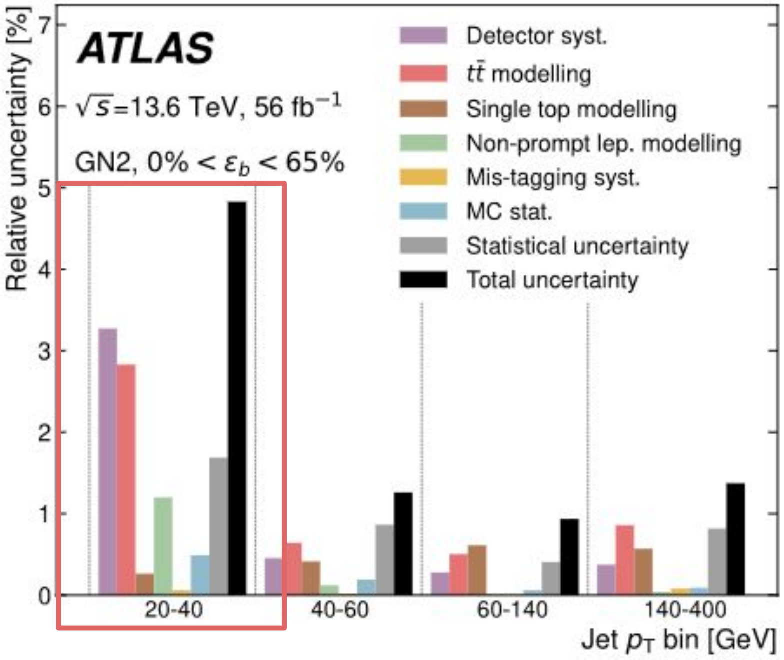
Why care about hadronization?

Reconstruction and tagging are a key part of the LHC program (see e.g. [N. Hartman's talk]) → Hadronization uncertainties become more important the more fine-grained we go

Will become even more important in future colliders!

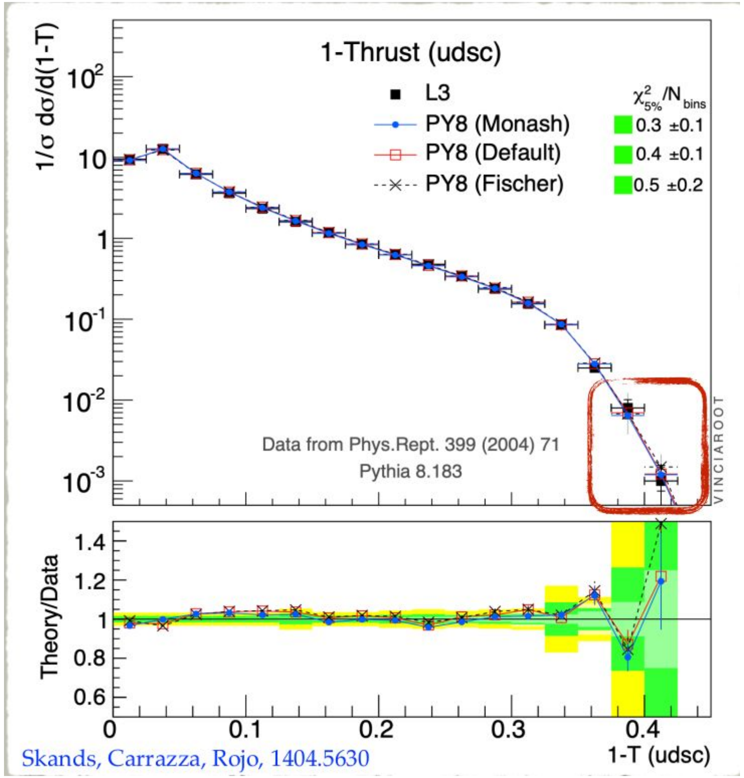
[FCC: QCD Physics report]

Jet flavour tagg. probab.	b quark	c quark	s quark	u quark	d quark	gluon
ϵ_j^b	0.8	1.2×10^{-3}	1.0×10^{-4}	1.6×10^{-4}	1.8×10^{-4}	5.5×10^{-3}
ϵ_j^c	0.014	0.8	1.6×10^{-3}	1.5×10^{-3}	1.6×10^{-3}	0.015
ϵ_j^s	2.4×10^{-3}	4.5×10^{-3}	0.8	0.21	0.24	0.12
ϵ_j^u	3.0×10^{-3}	0.029	0.21	0.8	0.67	0.18
ϵ_j^d	3.1×10^{-3}	0.030	0.24	0.67	0.8	0.18
ϵ_j^g	0.012	0.041	0.12	0.18	0.18	0.8



Taken from [ATLAS]

Why care about hadronization?



It impacts past programs. We want to measure α_s (see e.g. R. Mastandrea's talk) using event shape observables such as thrust.

$\alpha_s(M_Z)$ measurements from LEP can be dominated by hadronization uncertainties (scale as Λ_{had}/Q)

Why care about hadronization?

Will become even more important in future colliders!

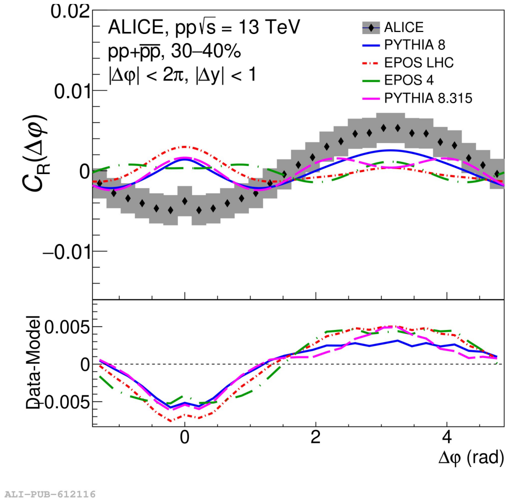
From FCC: QCD Physics report

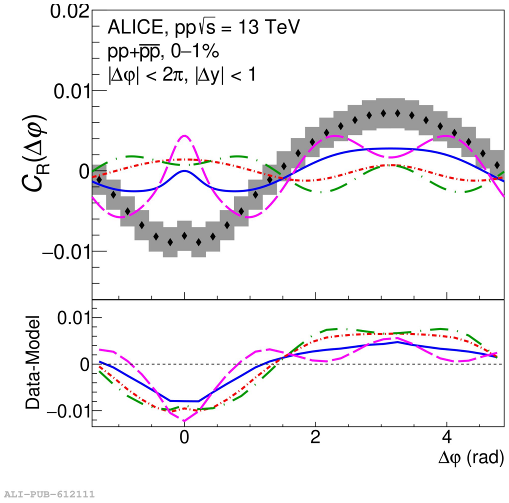
Table 5: Foreseen α_S extractions from e^+e^- data at FCC-ee: Present world-average and extraction method [27], expected FCC-ee statistical and systematic uncertainties (the two bottom categories need dedicated studies), and theory developments required to match the experimental precision.

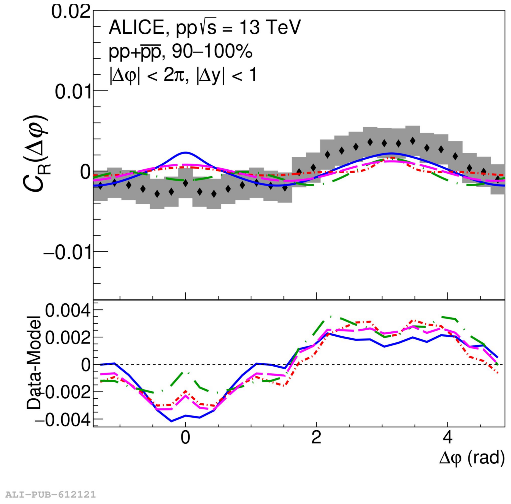
Category	present $\alpha_S(m_Z^2)$ ($\times 10^4$) value $\pm$ uncertainty	FCC-ee Stat.	FCC-ee Syst.	Extraction method Theory developments needed
Z hadronic decays	1 208 $\pm$ 28	0.1	1	Combined R_ℓ^Z, Γ_Z, $\sigma^0_{\rm had}$ fit $\mathcal{O}(\alpha_S^5)$, $\mathcal{O}(\alpha^3)$, $\mathcal{O}(\alpha_S,\alpha^3)$, $\mathcal{O}(\alpha_S^2,\alpha^2)$ corrs.
W hadronic decays	1 070 $\pm$ 350	2	2	Combined R_ℓ^W, Γ_W fit $\mathcal{O}(\alpha_S^5)$, $\mathcal{O}(\alpha^2)$, $\mathcal{O}(\alpha^3)$, $\mathcal{O}(\alpha_S,\alpha^3)$, $\mathcal{O}(\alpha_S^2,\alpha^2)$ corrs.
τ hadronic decays	1 178 $\pm$ 19	$\ll$1	<10	Combined $\Gamma^{\tau,\rm had}$ and τ lifetime fit $\mathcal{O}(\alpha_S^5)$, hadronisation corrections
Event shapes & jet rates	1 171 $\pm$ 31	$\ll$1	<10	Combination of event shapes and jet rates $\mathcal{O}(\alpha_S^4)$, hadronisation corrections

Why care about hadronization?

ALICE reports problems in the flavor structure. Hard to explain with current event generators.





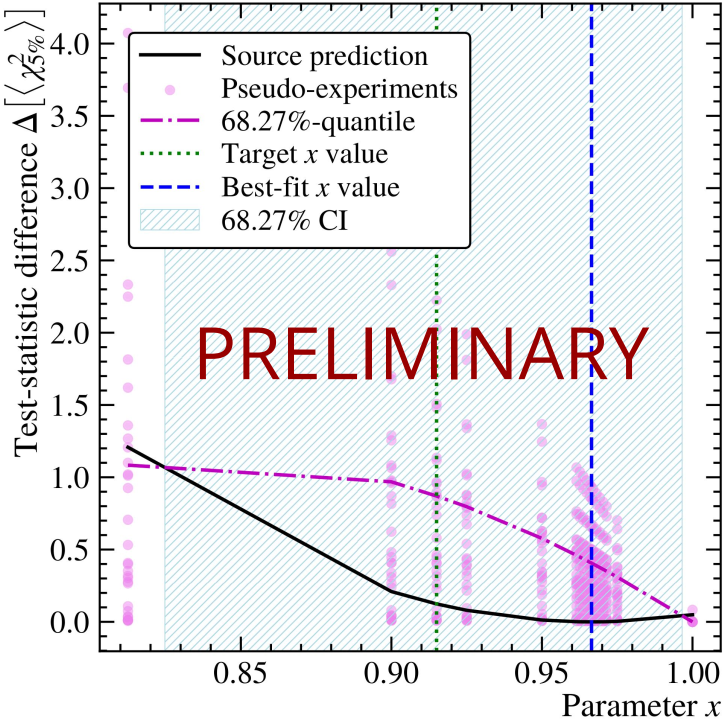


How to tune hadronization models?

Current, state of the art: Not that different to Monash 2014. Hard to get a very good tune, with quantifiable uncertainties that is reproducible.

Attempts have been made (with and w/o ML). See LPCC MCWG Tuning forum!

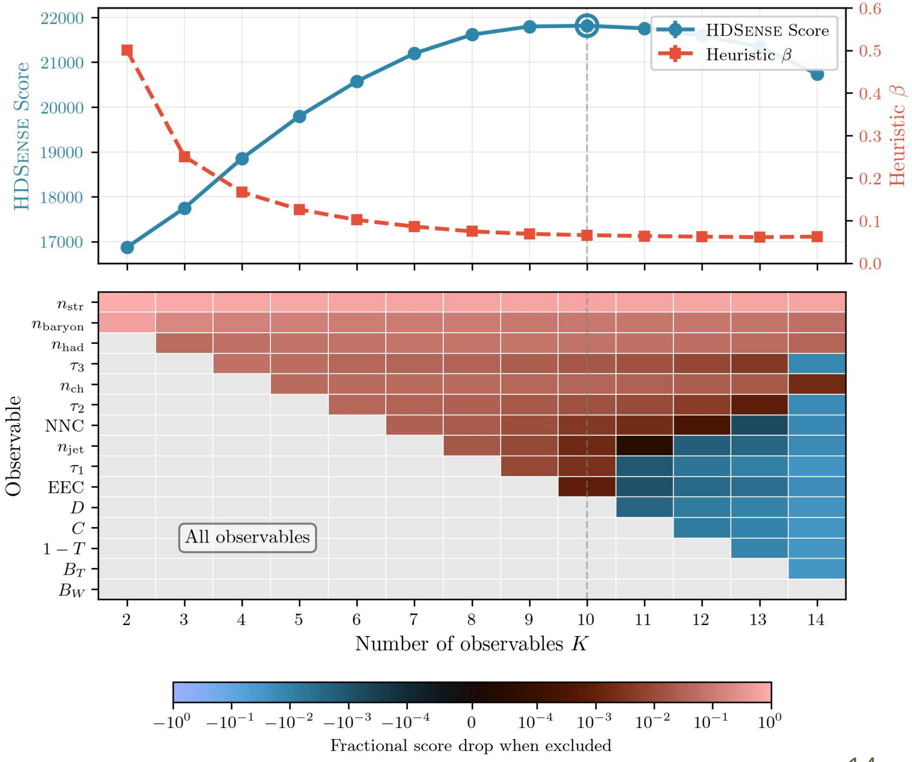
MLHad soon: FRAM, uses Feldman-Cousins Unified construction to accurately provide uncertainties, possible due to reweighting (+ more numerical tricks).



How to tune hadronization models?

Which observables to use? → What do we care about? In this talk, we care about hadronization sensitivity.

HDSense: Model dependent ranking, selects a minimum set of observables that capture the Fisher dependence with reduced simulations.

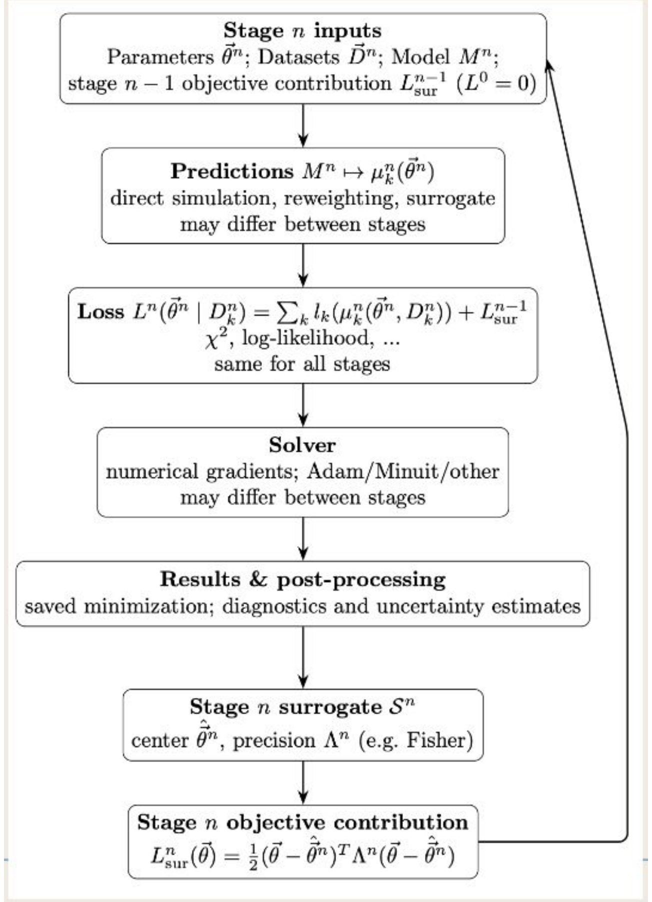


How to tune hadronization models?

Kvasir is a framework for reproducible generator tuning with as few assumptions as possible*.

Should be amenable to all ML models, with good linking to RIVET and YODA

PRELIMINARY

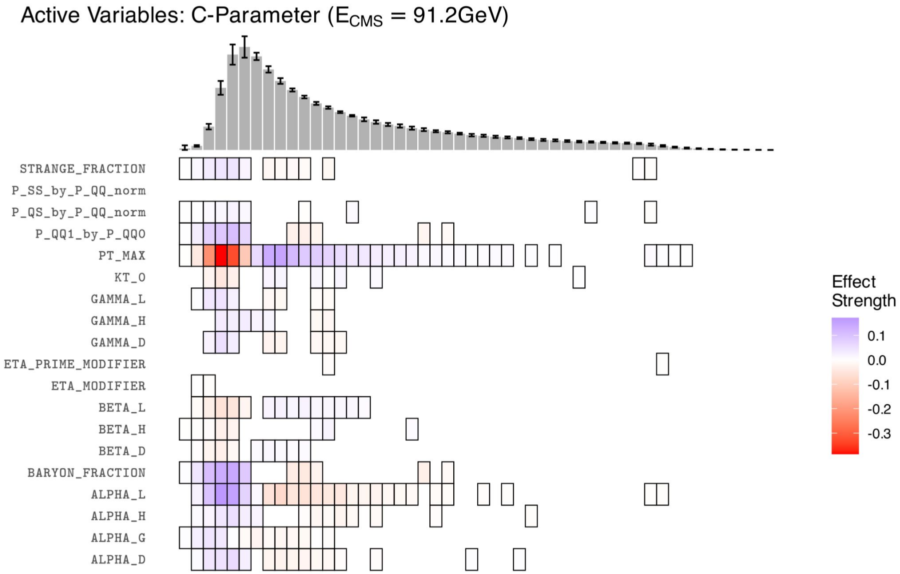


How to tune hadronization models?

$$g(x) = \sum_{i=1}^{p} \beta_i h_i(x_A) + u(x_A) + w(x)$$

Worrying/interesting question: What if we couldn't choose a single, best parameter fit to take as "nominal"?

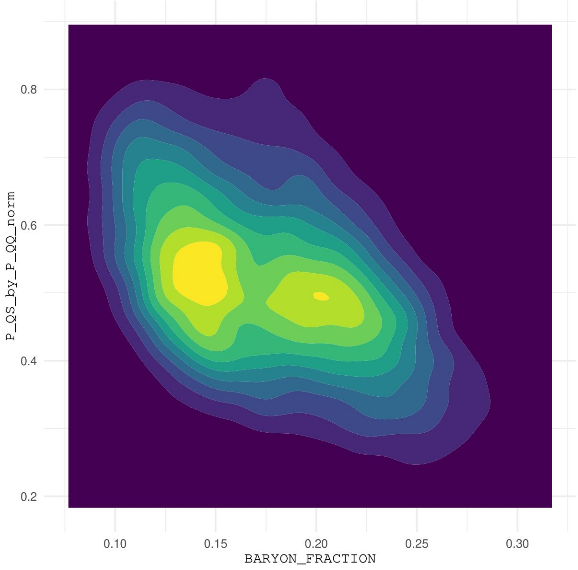
Iskauskas et al define compatible regions for both string and cluster models in Sherpa using History Matching and Bayesian Linear Models → **ML for smart emulation for efficient model selection**.

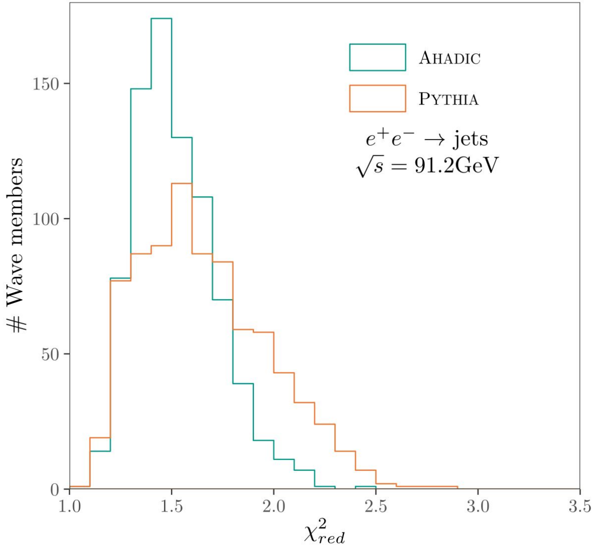


How to tune hadronization models?

Compatible regions are multi-peaked, and do not have a “perfect” fit over the full dataset ($\chi^2_{\text{red}} \sim 1.5\text{-}2$).

Different perspective on what a “good model" is.

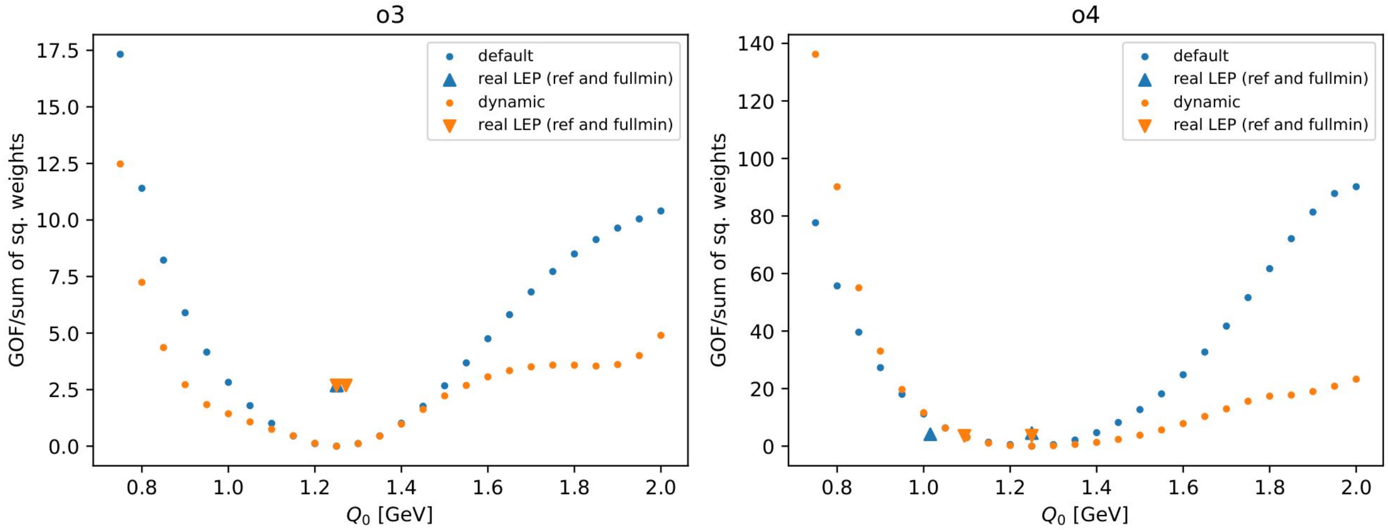




Parton shower impacts hadronization tuning

Gluon emissions affect hadronization, since it is IRC-unsafe. This introduces a correlation between shower cut-off and tuned hadronization parameters.

Hoang et al: impact for the cluster model and how to modify it accordingly!

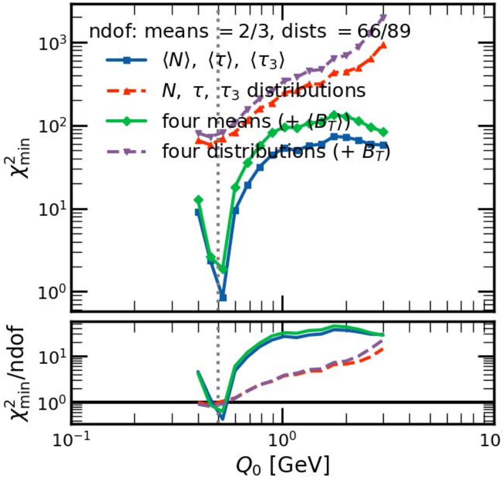


Parton shower impacts hadronization tuning

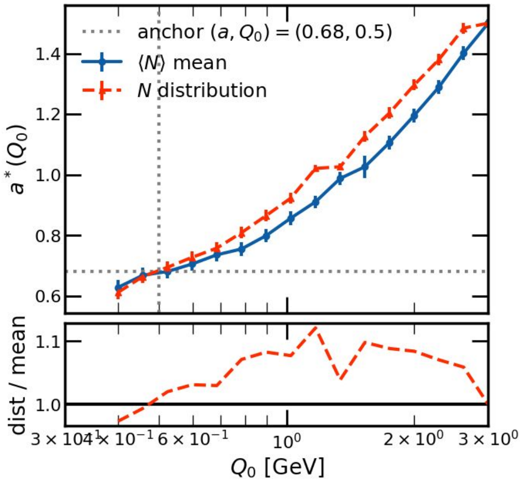
MLHad soon: RGFlow, similar (for Pythia) but different end goal. Changes in Q_0 define a "flow" in parametric space for the Lund model

Ensemble of models yield theory uncertainty, propagated as fragmentation weights.

Some observables have a theoretical scaling, but most general scaling will be numeric. Ask Rikab!



PRELIMINARY



Questions I have about tuning

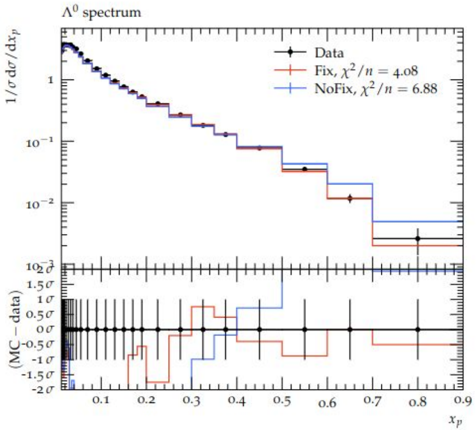
Usual, hard questions: When is a model good enough?* When are its uncertainties meaningful? How to deal with model incompatibility?

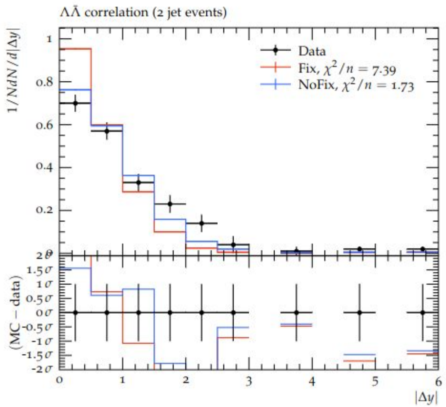
* Mrenna Maxim: “When it begins to disagree with data and drive new experimental analyses!”

Questions I have about tuning

Usual, hard questions: When is a model good enough?* When are its uncertainties meaningful? How to deal with model incompatibility?

Awkward, harder? question: How do we know our model does what we think it does? e.g.: [bug in the aExtra parameter in Pythia]()



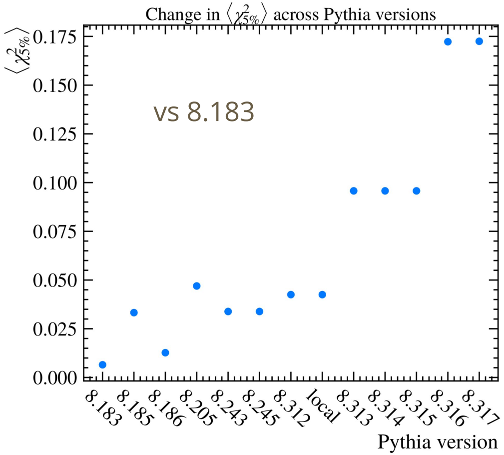


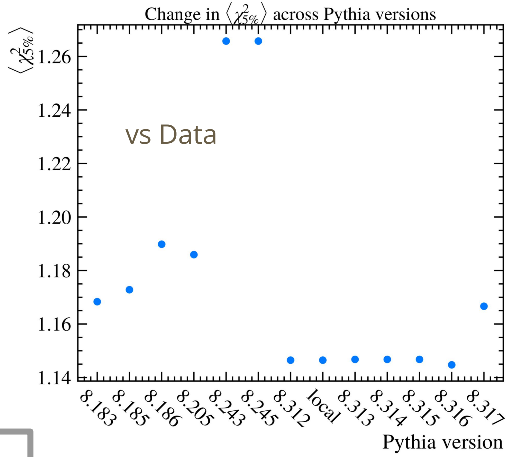
* Mrenna Maxim: "When it begins to disagree with data and drive new experimental analyses!"

Questions I have about tuning

Usual, hard questions: When is a model good enough?* When are its uncertainties meaningful? How to deal with model incompatibility?

Awkward, harder? question: How do we know our model does what we think it does? e.g.: [bug in the aExtra parameter in Pythia] + the Pythia drift. **In both cases, we were not made aware of the issue due to mismatch with data.**

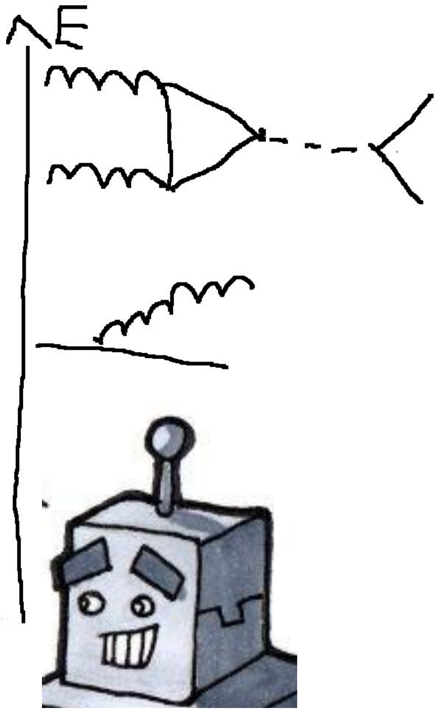




* Mrenna Maxim: "When it begins to disagree with data and drive new experimental analyses!"

Of course… what about ML-based models?

Two groups have ~~recently~~ tackled the subject: **MLHAD** (arxiv:2203.04983, arxiv:2311.09296, arxiv:2410.06342, arxiv:2503.05667,…) and **HADML** (arxiv:2203.12660, arxiv:2305.17169, arxiv:2312.08453). Different **generators** (Pythia, Herwig) and different **architectures** (cSWAE,BNF,GNNs, GAN) with different **degrees of implementation**.



pQFT

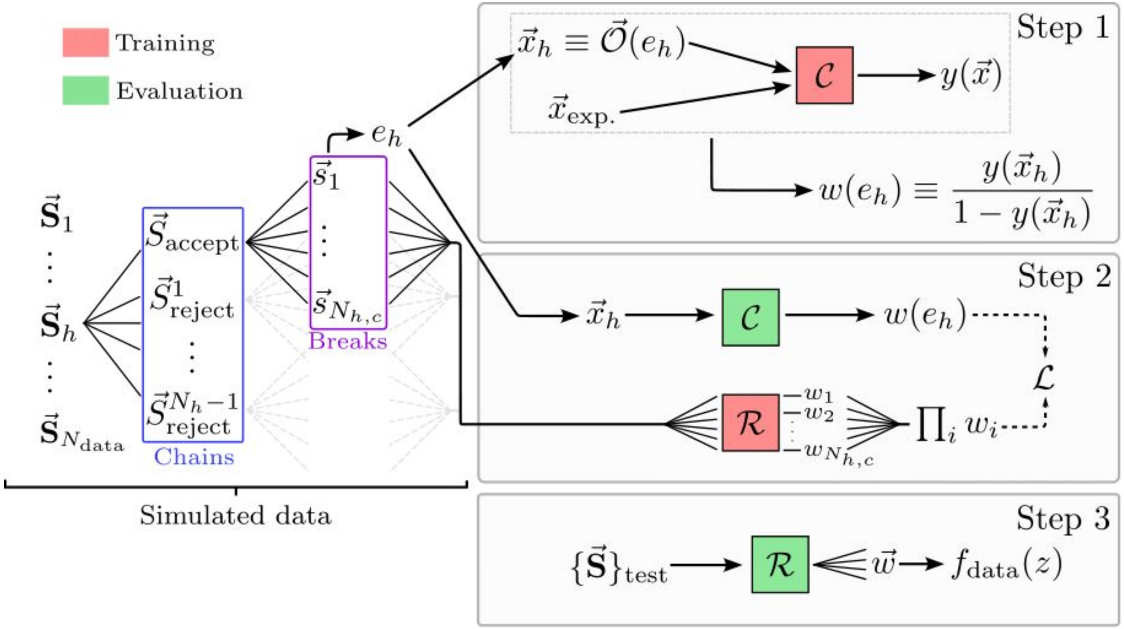
Showers

Empirical **ML** models

Learning from data: HOMER

Histories and **O**bservables for **M**onte-Carlo **E**vent **R**eweighting: We take **Pythia as baseline** and reweight that.

Two step procedure: We generate **once** and learn the appropriate **reweighting function**. New model is **Pythia + weight.** Eventually released as a Pythia plug-in.

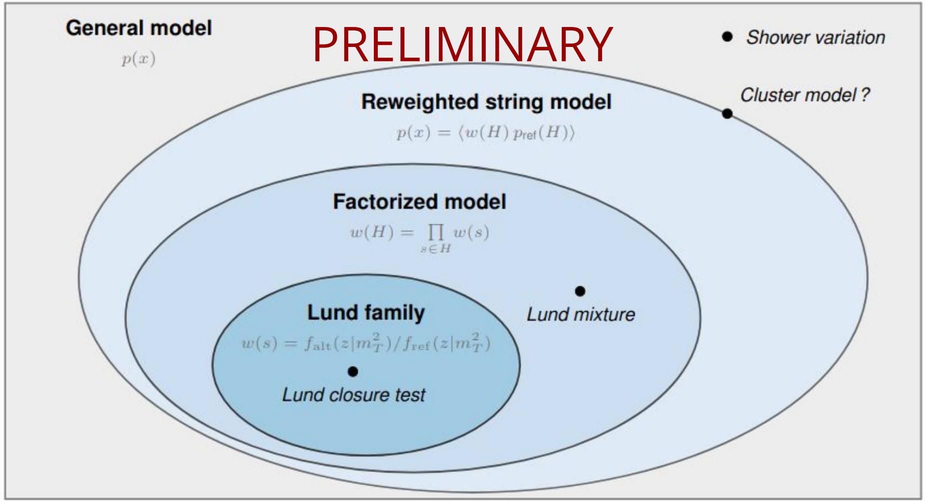


Latest on HOMER

As seen in S. Kim's talk, performance improved by using multiple algorithms, including AUSSIE by Ore at al, tested on a family of datasets

We observe a limitation on what we can learn (related to the inductive bias in choosing the baseline!)

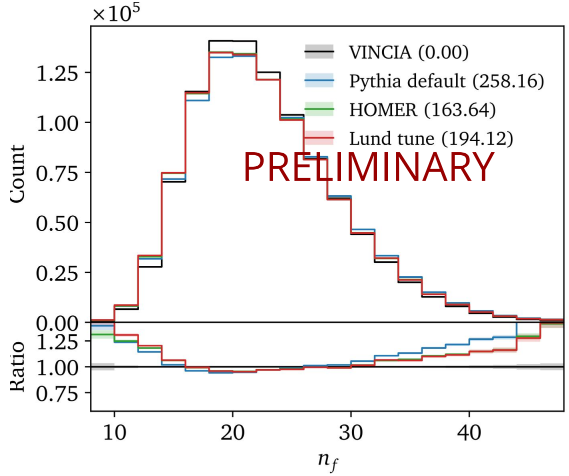
S. Kim, A.Ore, S. Palacios Schweitzer, T. Plehn + MLHad: Paper soon!

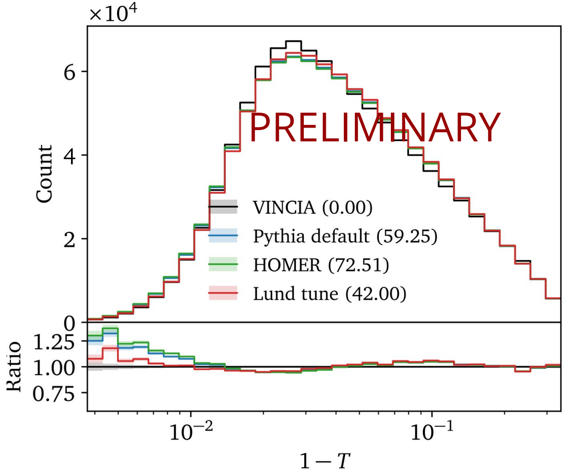


Parton shower and ML-based hadronization models

PS effects may impact any ML-based model even more! We do not want to absorb arbitrary effects, but the converse implies the ML-model fit is dependent on the shower choice and also "flows".

S. Kim's talk: HOMER cannot absorb all differences between Default and VINCIA Shower due to its inductive bias → By design.



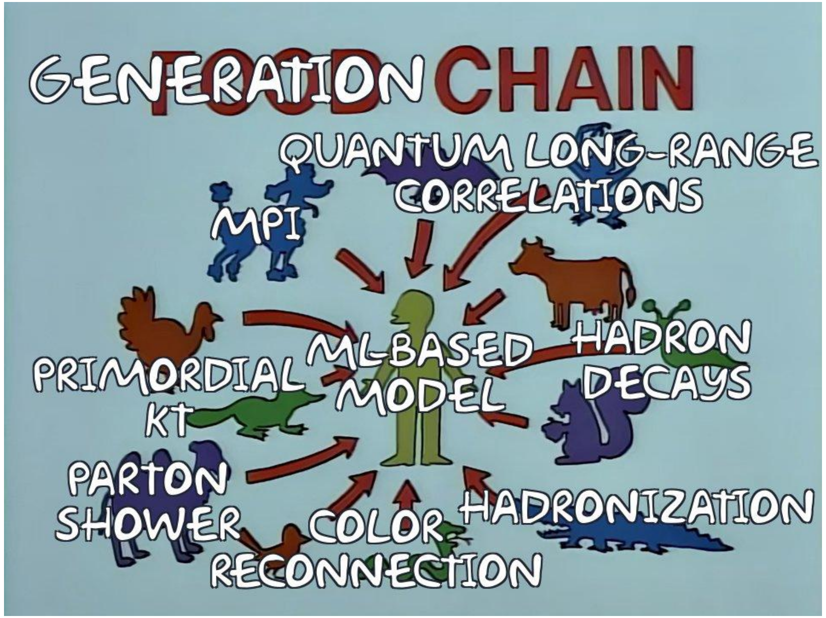


Perspectives on ML-based models

We started with a gradual perspective (let's build carefully). Very good for learning, but... Not really how things work. The hadronization model works in tandem with everything else.

Partial wishlist:

- How to best incorporate flavor? (efforts underway)
- How to go beyond the factorized structure while keeping as much physics as possible? Look at parton shower (started) and softer physics (relevant to pp colliders specially)
- Train on data and see what breaks.



Outlook: What does the future look like? (who knows?)

Theory will continue to be the largest uncertainty in many measurements → tuning will be more and more important. **ML for model fit, model uncertainties and reproducibility.**

Current HOMER seems beyond its comfort zone. Flavor + shower dependence underway but more might be needed. **We need to look at open data.**

Looking at theory input and experimental performance: How much do we want to profile non-perturbative physics away? What does a good hadronization tune looks like? **ML can help with the tuning, the model, or both, but it cannot decide on its quality for us.**

Backup slides

The string kinematics

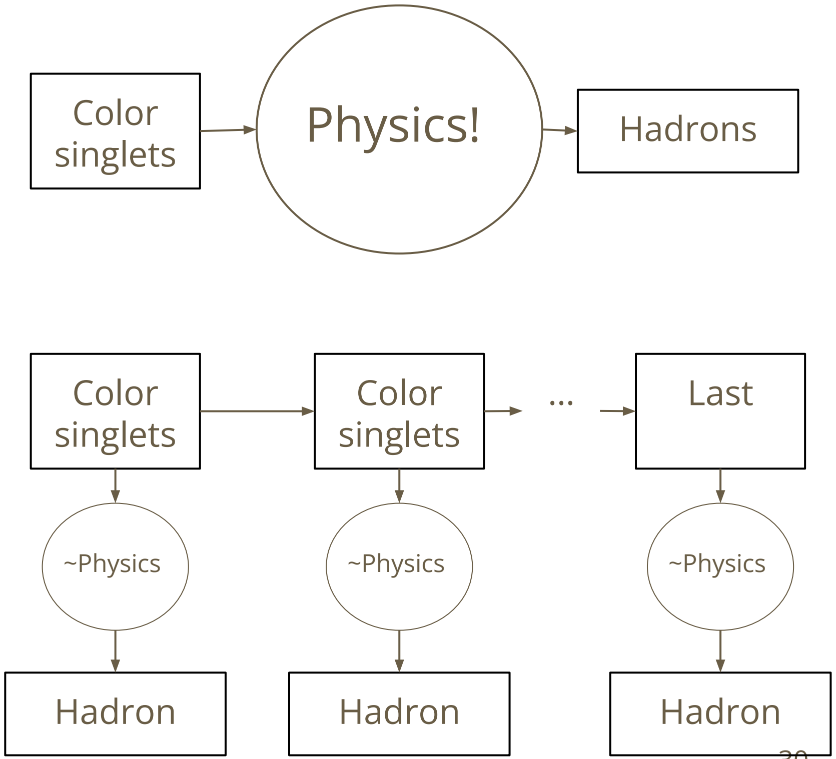
The hadronization pdf can be written explicitly as

$$dP_n = \prod_{j=1}^{n} N_j dp_j \delta(p_j^2 - m_j^2) \delta \left(\sum_{j=1}^{n} p_j^\mu - P_{\text{string}}^\mu \right) \exp(-b\mathcal{A}),$$

where A is the area of the string before the breakup, and evidences the tunneling suppression.

However, this is not practical, and event generators use **iterative fragmentations** where the string breaks are modelled as a **random walk**.

Care has to be taken so that things remain physical (**meet in the middle, finalTwo**).

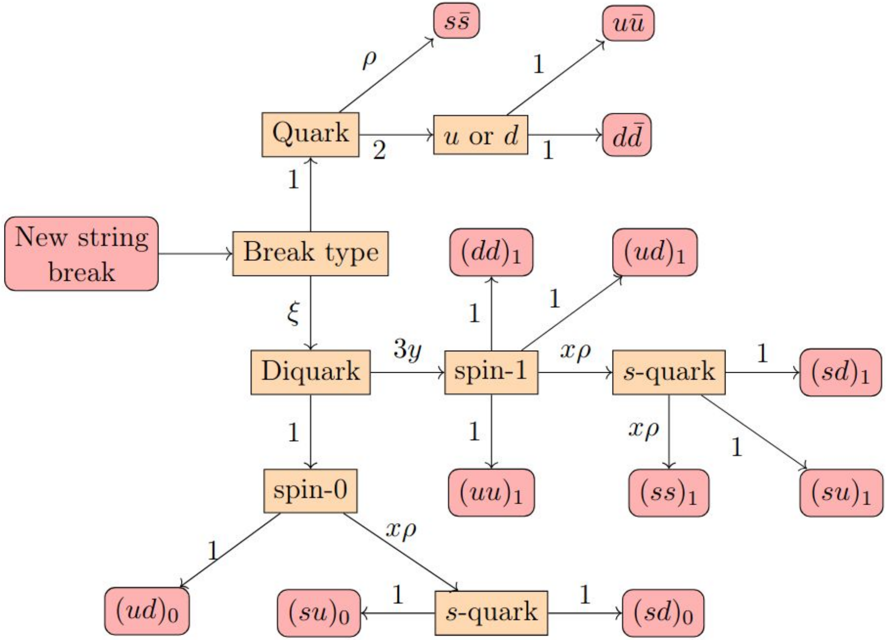


Flavor structure

In the Lund string model, highly parametric. Ensures SU(6) spin-flavor symmetries through accept-reject

Baryon content needs additional “popcorn model”

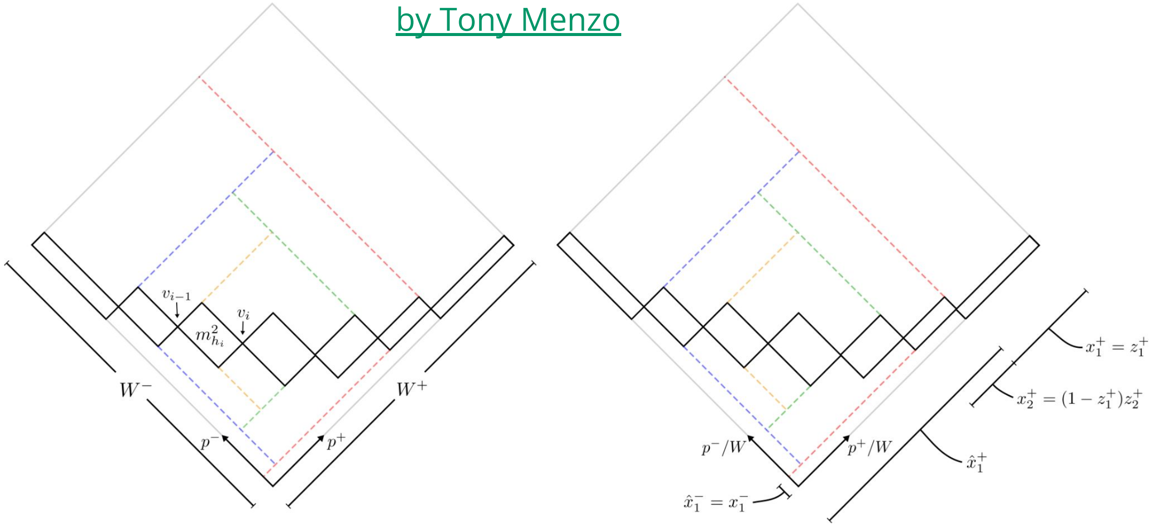
From arxiv:2505.00142, where flavor reweighting for Pythia was introduced.



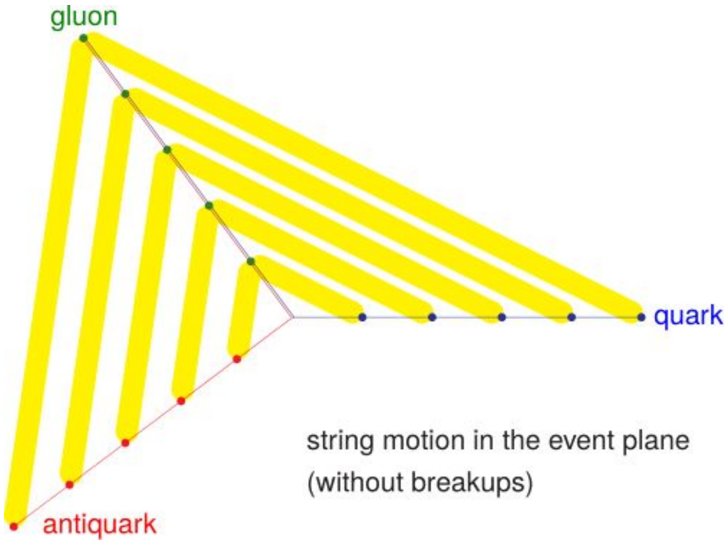
Momentum space for finding next hadronization

In momentum space, describe the string breaking procedure in terms of momentum fractions x_i

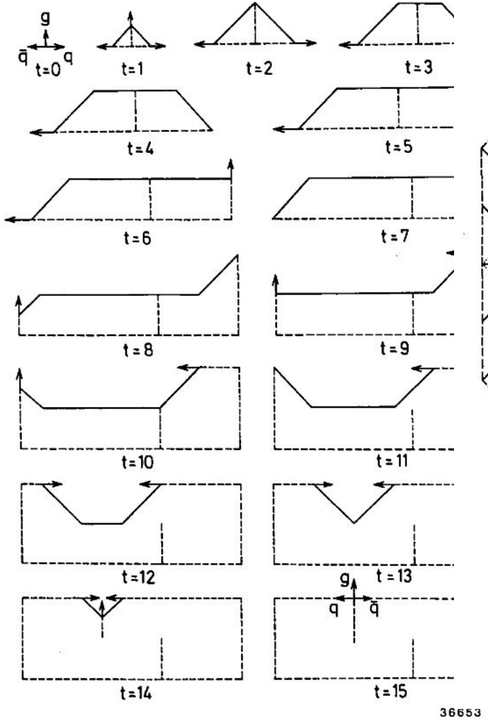
[Highly recommended tutorial by Tony Menzo]



Gluons!



string motion in the event plane
(without breakups)



T. Sjöstrand / Jet fragmentation

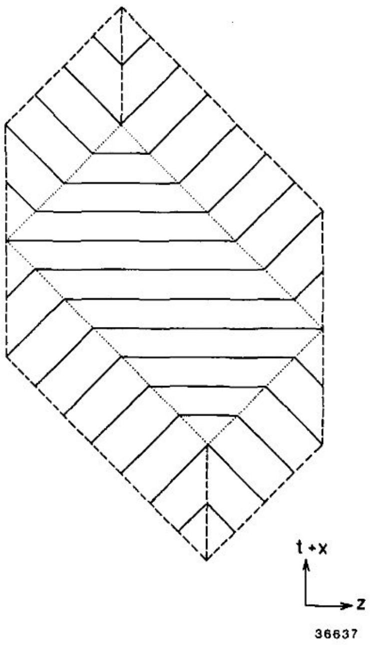


Fig. 5 (continued).

Relevant variables?

Things become more complex, and the hadron momenta depends on proper time and transverse mass

$$p = \sum_{j=j_1}^{j_2} x_+^{(j)} p_+^{(j)} + \sum_{k=k_1}^{k_2} x_-^{(k)} p_-^{(k)} + p_{x1}\hat{e}_x^{(j_1k_1)} + p_{y1}\hat{e}_y^{(j_1k_1)} + p_{x2}\hat{e}_x^{(j_2k_2)} + p_{y2}\hat{e}_y^{(j_2k_2)}$$

However, the sampled variables remain the same as in the simple qq string! Two conditions to be solved.

$$\Gamma_i = (1-z)\left(\Gamma_{i-1} + \frac{m_\perp^2}{z}\right)$$

See here

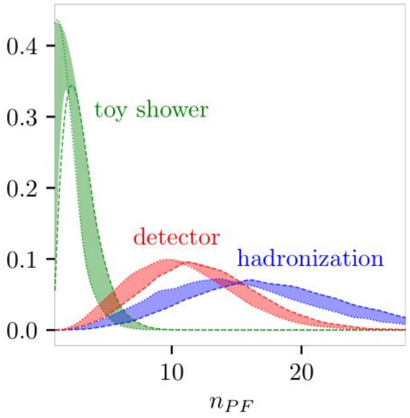
$$p^2 = m^2$$

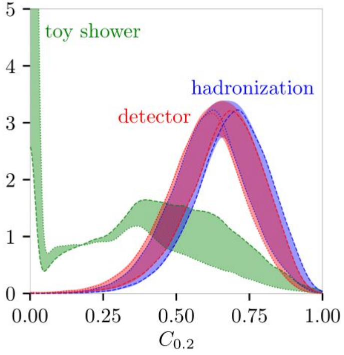
Why care about hadronization?

Bieringer et al

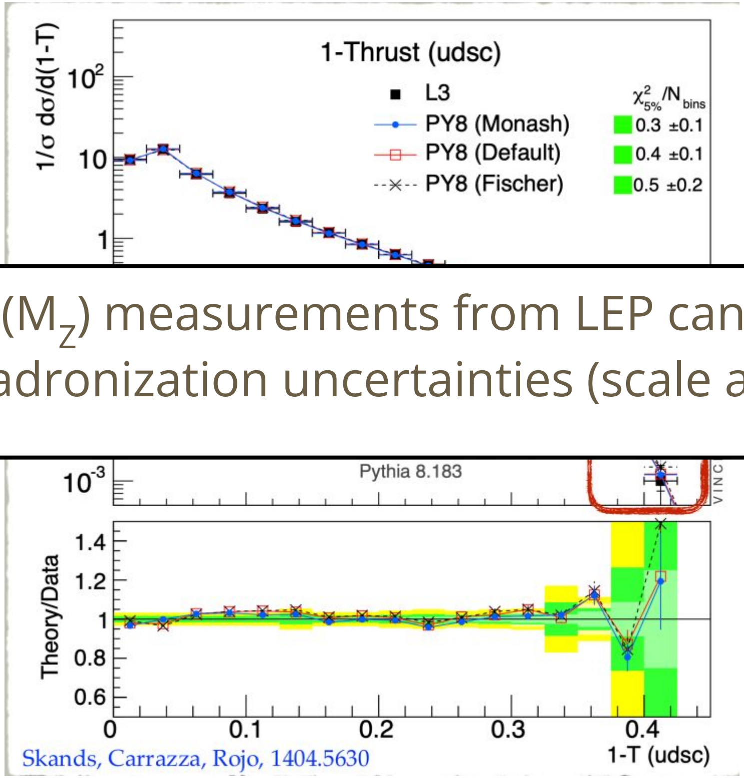
Theory has consistently designed observables to be as insensitive to hadronization as possible to test the Standard Model (are we missing some physics?). See Ben's talk.

However, real measurements have residual dependence even then (e.g. due to experimental cuts) → Hadronization modelling essential for precision studies.

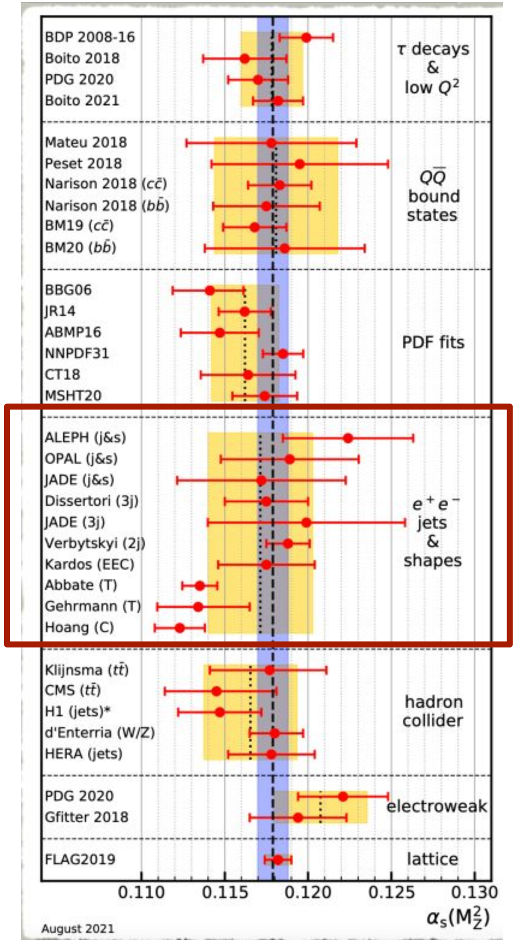




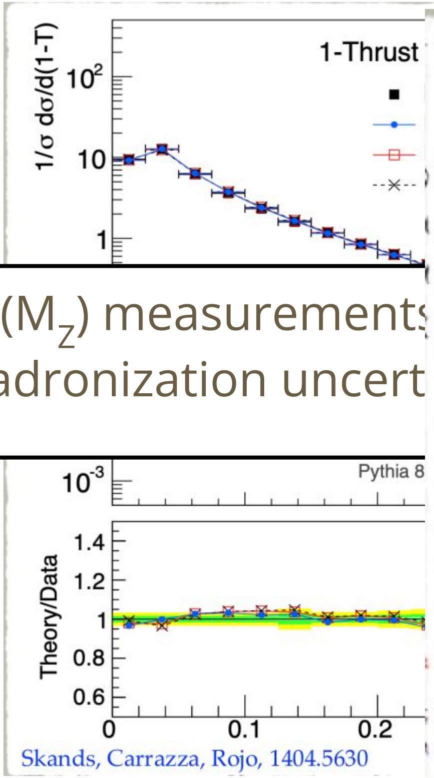
Why care about hadronization?



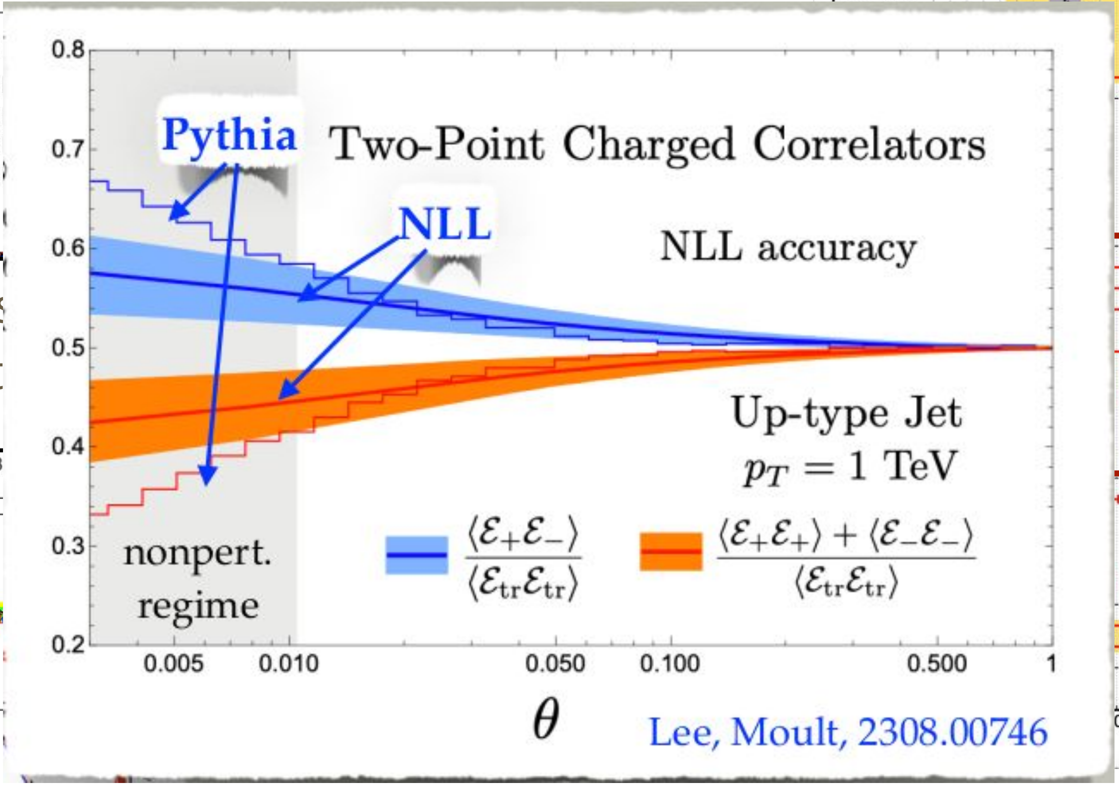
$\alpha_s(M_Z)$ measurements from LEP can be dominated by hadronization uncertainties (scale as Λ_{had}/Q)

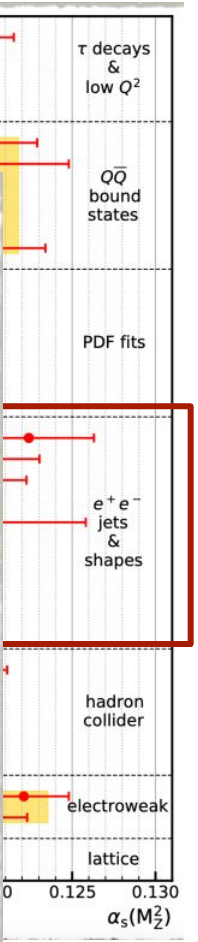


Why care about hadronization?



$\alpha_s(M_Z)$ measurements
hadronization uncert

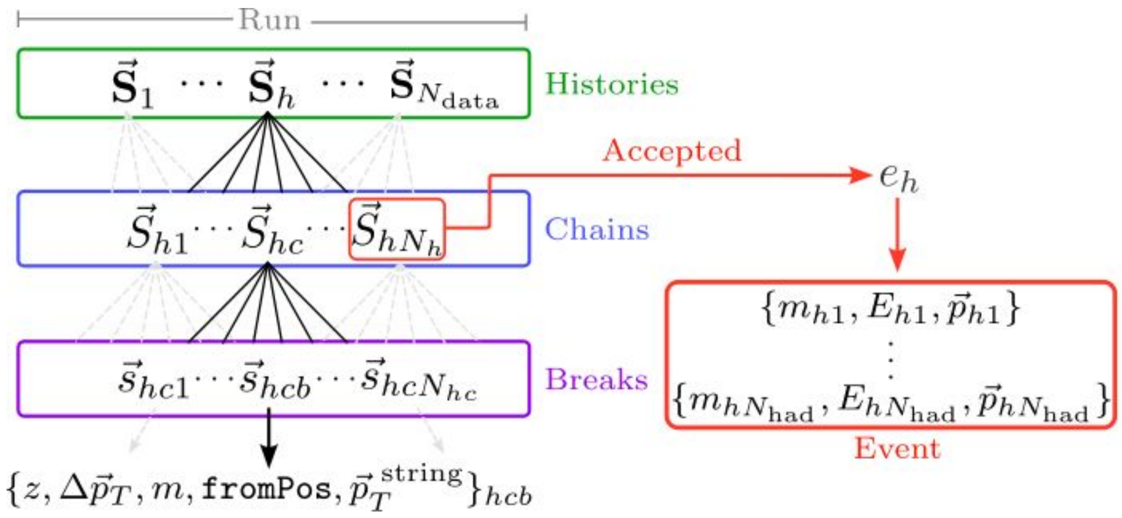




Learning from data: HOMER

Histories and **O**bservables for **M**onte-Carlo **E**vent **R**eweighting: We take **Pythia as baseline** and reweight that.

For simulation, we record everything. **For (pseudo-)data, only observable quantities**.

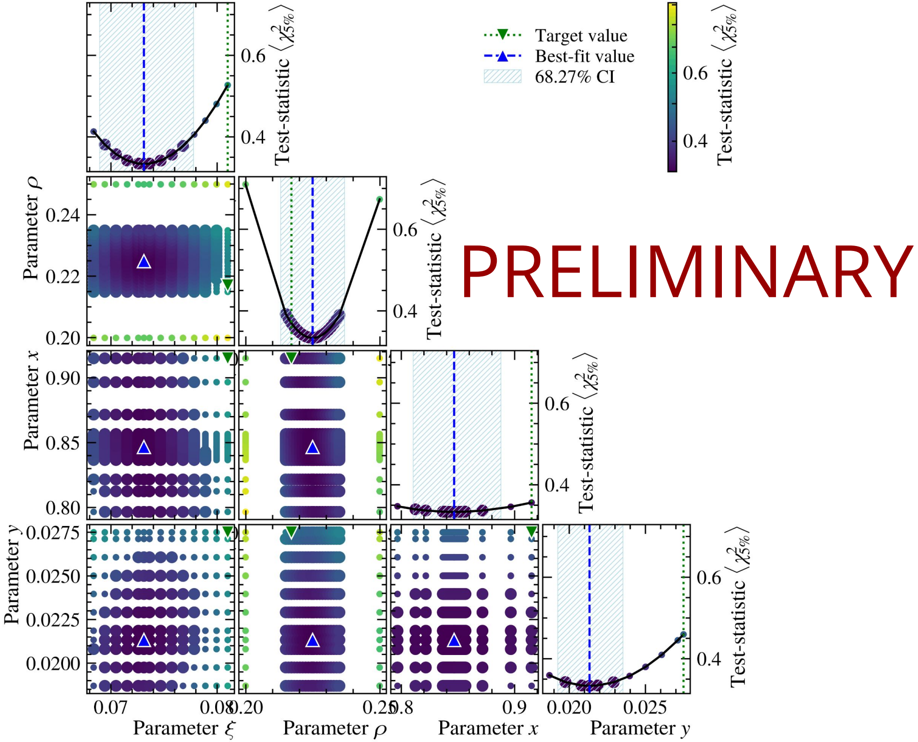


How to tune hadronization models?

Current, state of the art… Not much different to Monash 2014. Hard to get a very good tune, with quantifiable uncertainties that is reproducible.

Part of the issue is simulation costs, non-differentiability. Attempts have been made to use surrogate models / reweighting / ML here. See dedicated LPCC MCWG Tuning forum!

MLHad soon: FRAM, uses Feldman-Cousins Unified construction to accurately provide uncertainties, possible due to reweighting (+ more numerical tricks).

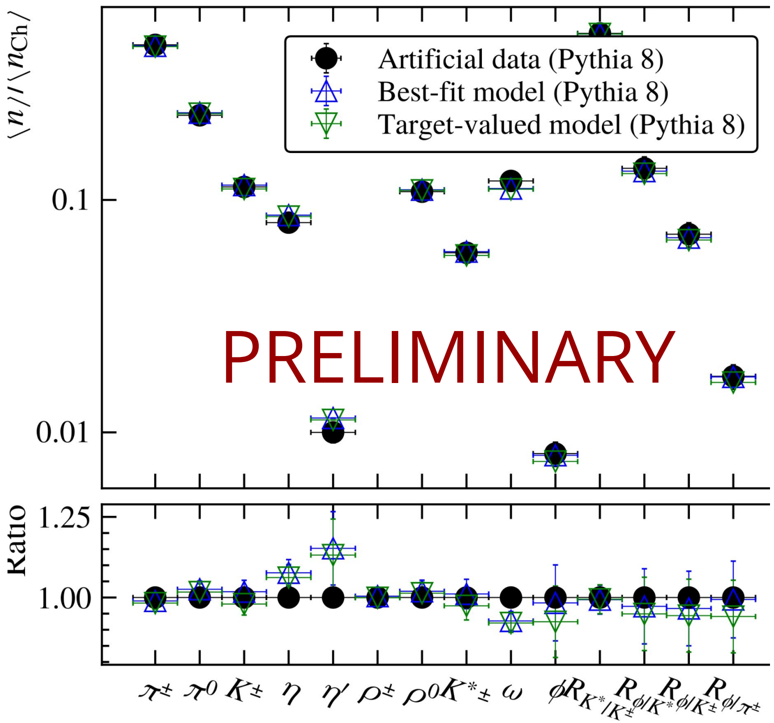


How to tune hadronization models?

Current, state of the art… Not much different to Monash 2014. Hard to get a very good tune, with quantifiable uncertainties that is reproducible.

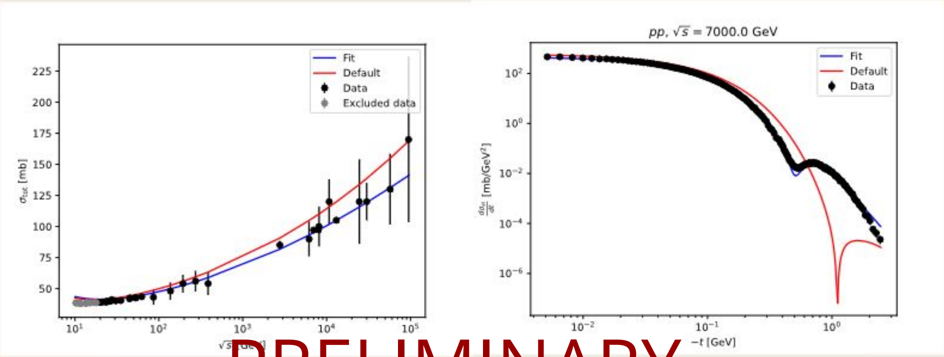
Part of the issue is simulation costs, non-differentiability. Attempts have been made to use surrogate models / reweighting / ML here. See dedicated LPCC MCWG Tuning forum!

MLHad soon: FRAM, uses Feldman-Cousins Unified construction to accurately provide uncertainties, possible due to reweighting (+ more numerical tricks).



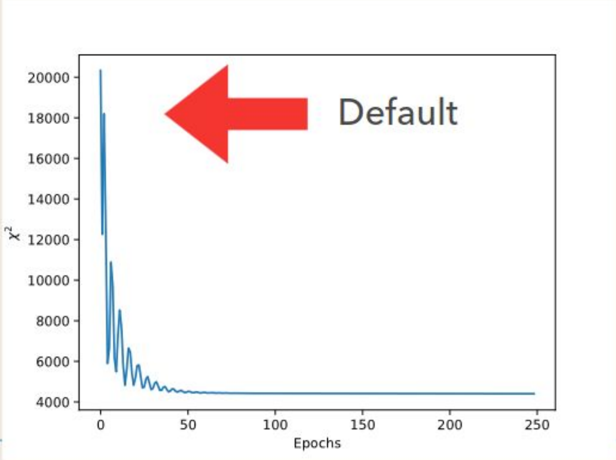
Selected results

- **Observables**
  - $\sigma_{tot}(\sqrt{s}), \sigma_{el}(\sqrt{s}), \rho(\sqrt{s})$ **(pp)**
  - $\frac{d\sigma_{el}}{dt}$ **(pp and ppbar, fixed energy)**
- Energy coverage: Low (19 GeV) → LHC → High (Cosmics)
- More details in backup



PRELIMINARY

Name	Tune	Round to 1σ	Error (1σ)	Monash
SigmaElastic:ABMSTaGluon	0.4176	0.42	0.28	0.3
SigmaElastic:ABMSTtzero	4.07	4.07	0.85	5.03
SigmaElastic:ABMSTlambda	0.5108	0.5108	0.0042	0.3
SigmaElastic:ABMSTepsi[0]	0.08587	0.0859	0.0053	0.106231
SigmaElastic:ABMSTepsi[1]	0.09731	0.0973	0.0012	0.0972043
SigmaElastic:ABMSTepsi[2]	-0.4499	-0.450	0.023	-0.510662
SigmaElastic:ABMSTepsi[3]	-0.1719	-0.172	0.081	-0.302082
SigmaElastic:ABMSTnormXi[0]	200.2	200	36	228.359
SigmaElastic:ABMSTnormXi[1]	199.9	200	15	193.811
SigmaElastic:ABMSTnormXi[2]	499.7	500	78	518.686
SigmaElastic:ABMSTnormXi[3]	12.14	12.1	8.9	10.7843
SigmaElastic:ABMSTalphapi[0]	0.06048	0.060	0.023	0.0449029
SigmaElastic:ABMSTalphapi[1]	0.2661	0.2661	0.0068	0.278037
SigmaElastic:ABMSTalphapi[2]	0.7361	0.736	0.054	0.821595
SigmaElastic:ABMSTalphapi[3]	0.7101	0.71	0.57	0.904556



How to tune hadronization models?

Which observables to use? → What do we care about? In this talk, we care about hadronization sensitivity.

But what about IRC safety? Is there a proper IRC-based ranking of observables? Can we quantify how IRC-unsafe observables are? Is it useful?