SIMULATING PARTICLE PHYSICS HADRONIZATION WITH MACHINE LEARNING

Jure Zupan
u. cincinnati

2nd SMASHING, Vipava, Oct 20, 2026

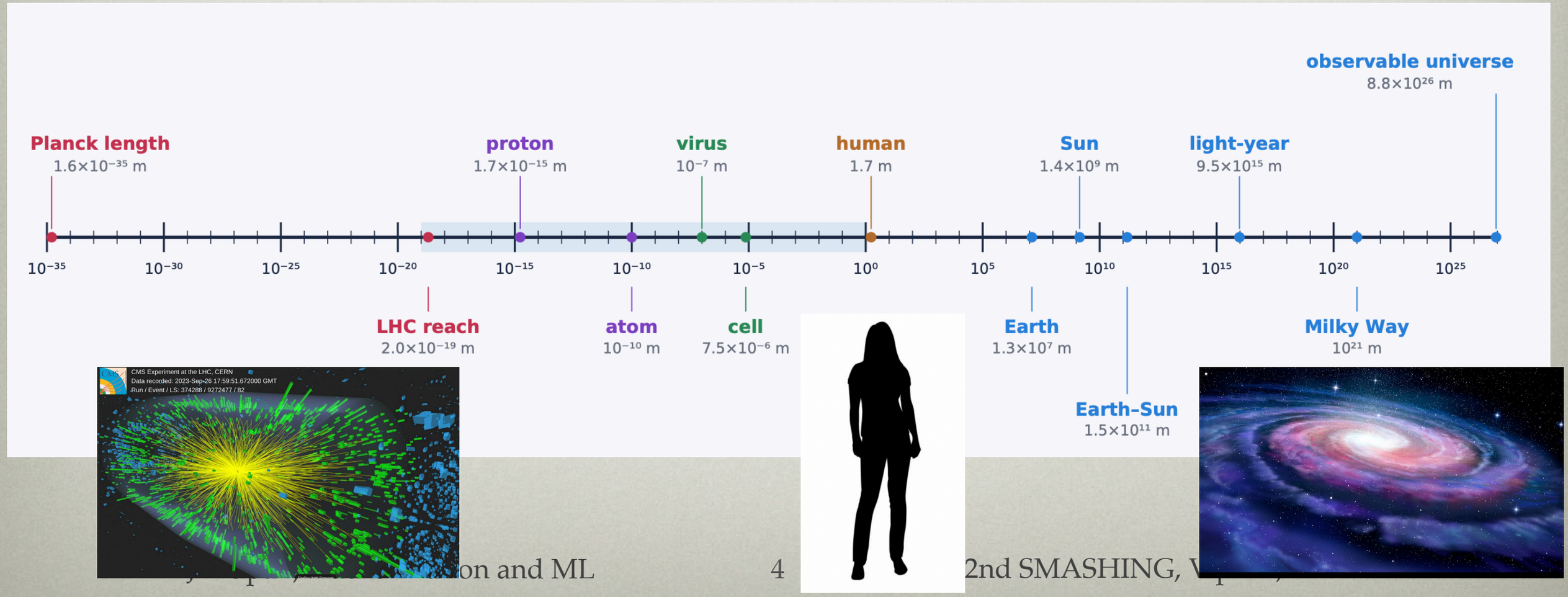
OUTLINE

- High energy physics
  - what?
  - why?
  - where?
  - how? ⇐ Machine Learning
- simulating hadronization with ML

What?

HIGH ENERGY PHYSICS

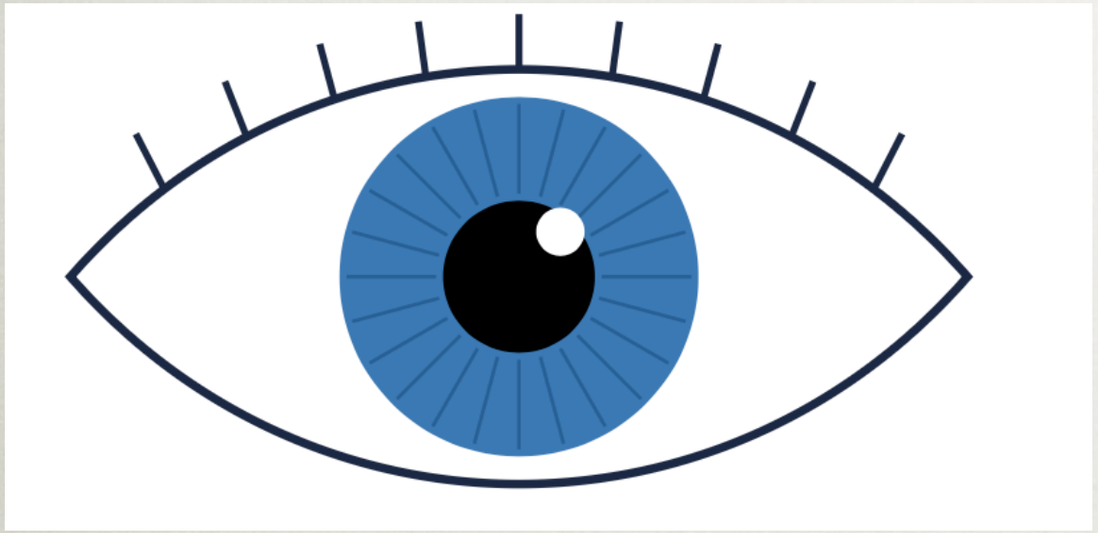
- high energy physics:
  - a way to resolve the smallest distances



EXPERIMENTAL TOOLS:

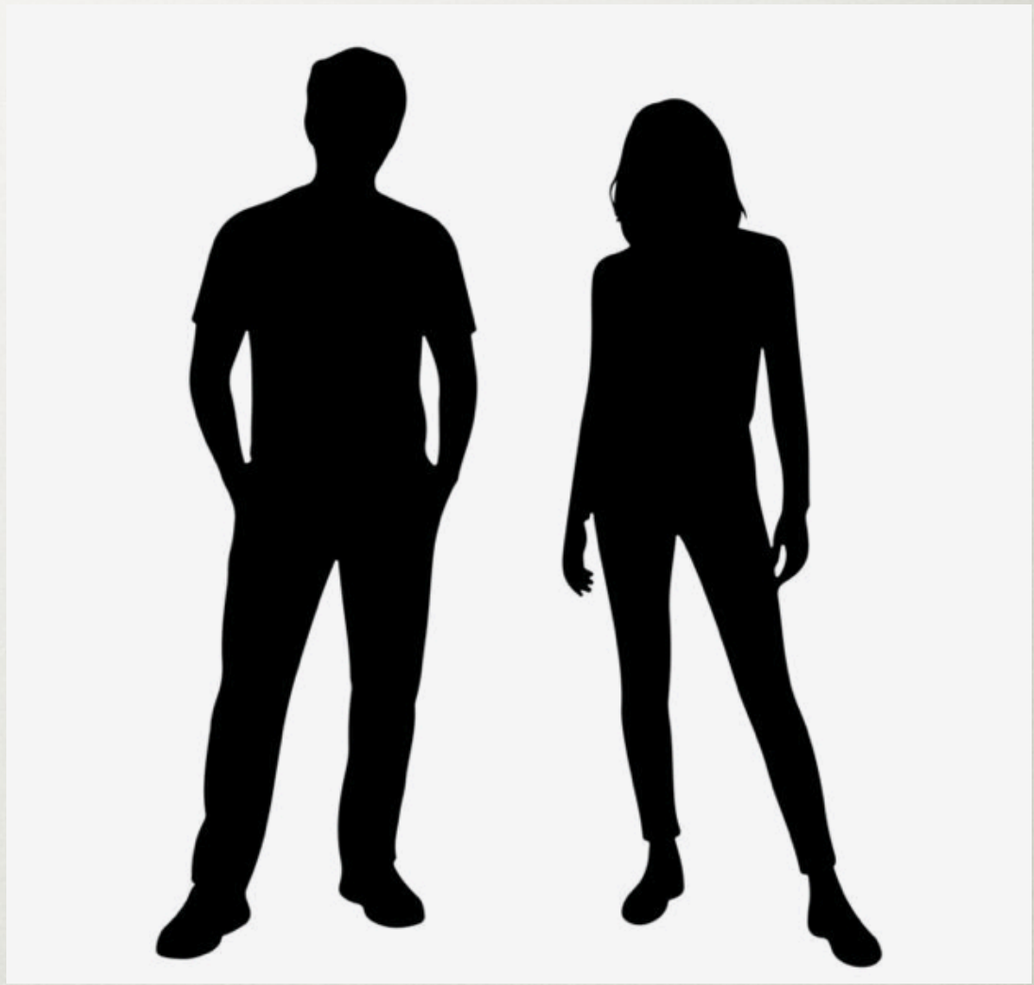
$$10^0\,\mathrm{m} = 1\,\mathrm{m}$$

Instrument



human eye

Object

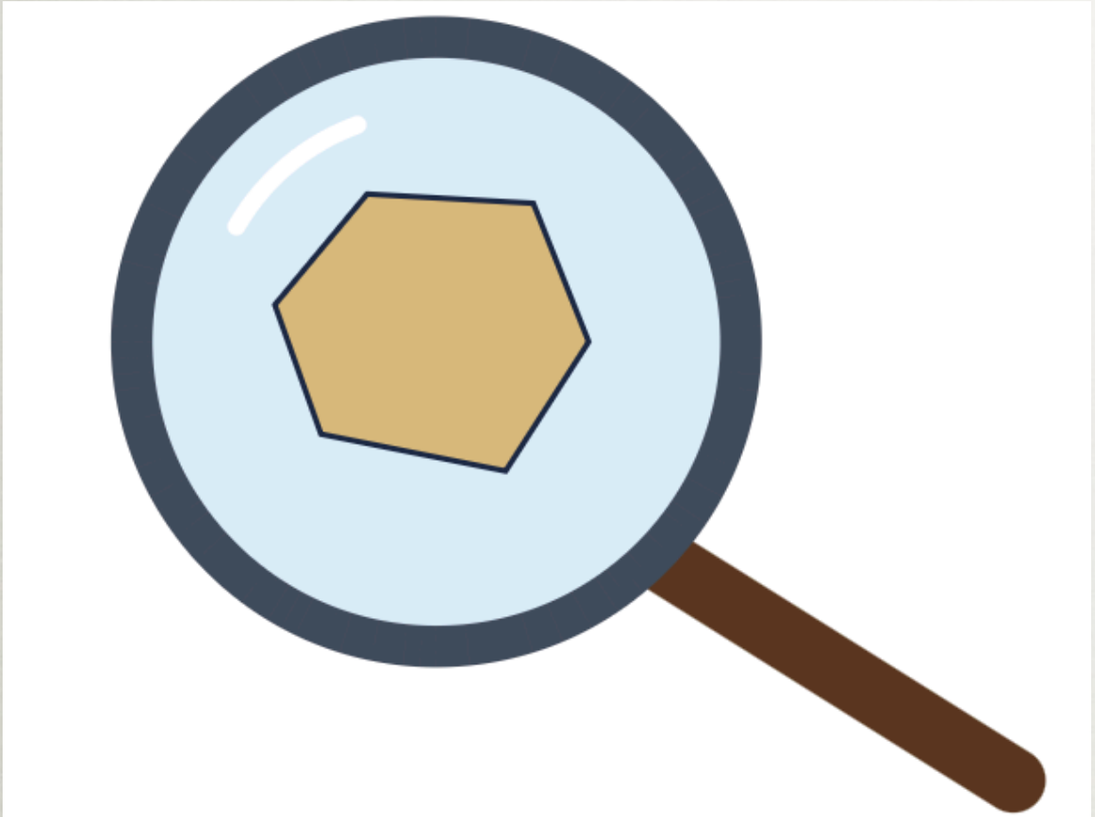


person

EXPERIMENTAL TOOLS:

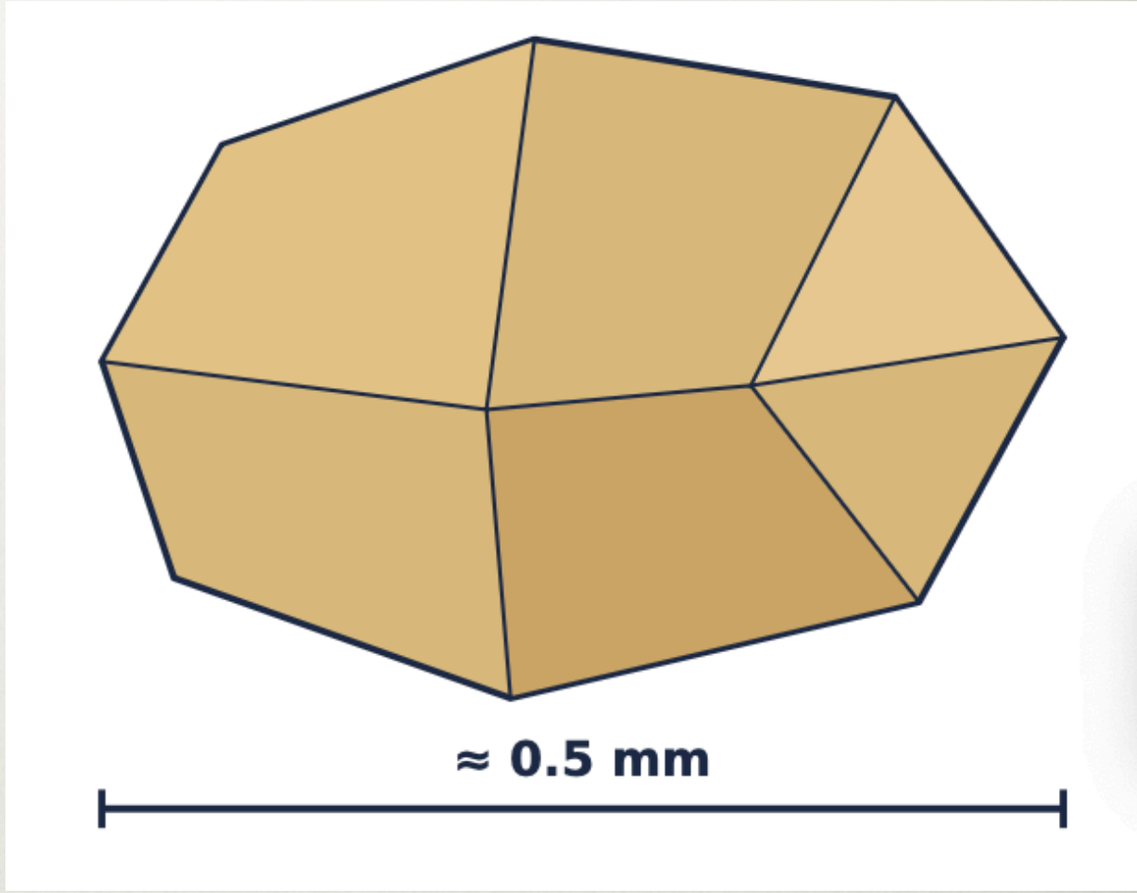
$$10^{-3}\,\mathrm{m} = 1\,\mathrm{mm}$$

Instrument



magnifying glass

Object

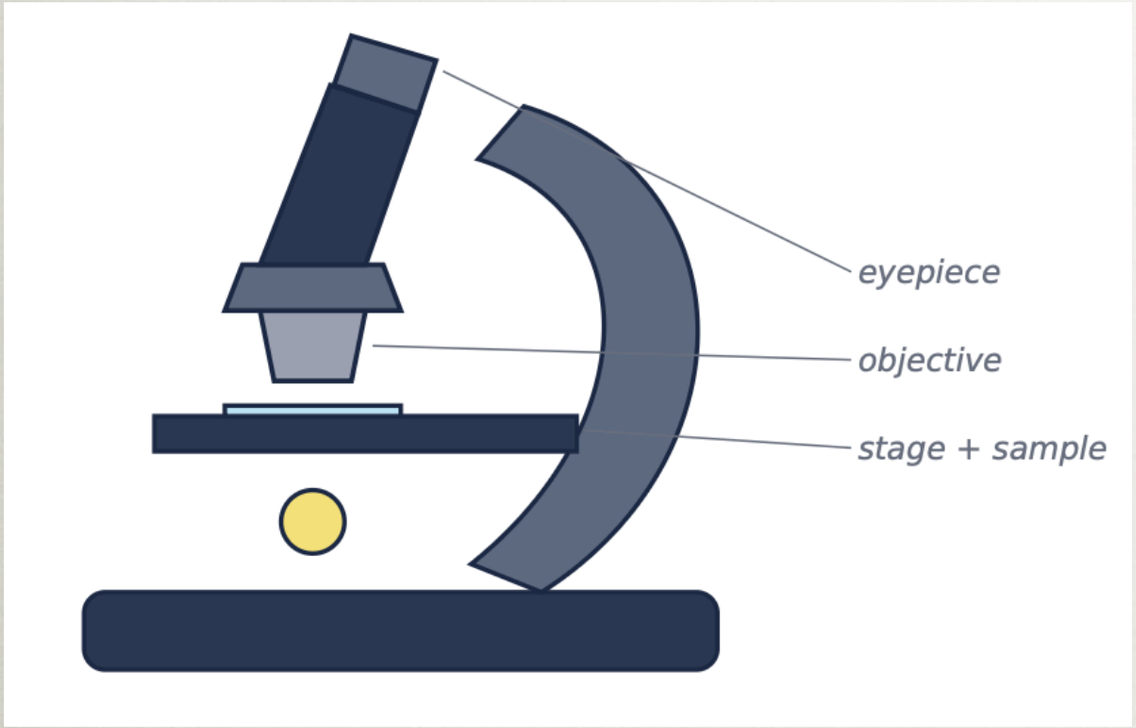


grain of sand

EXPERIMENTAL TOOLS:

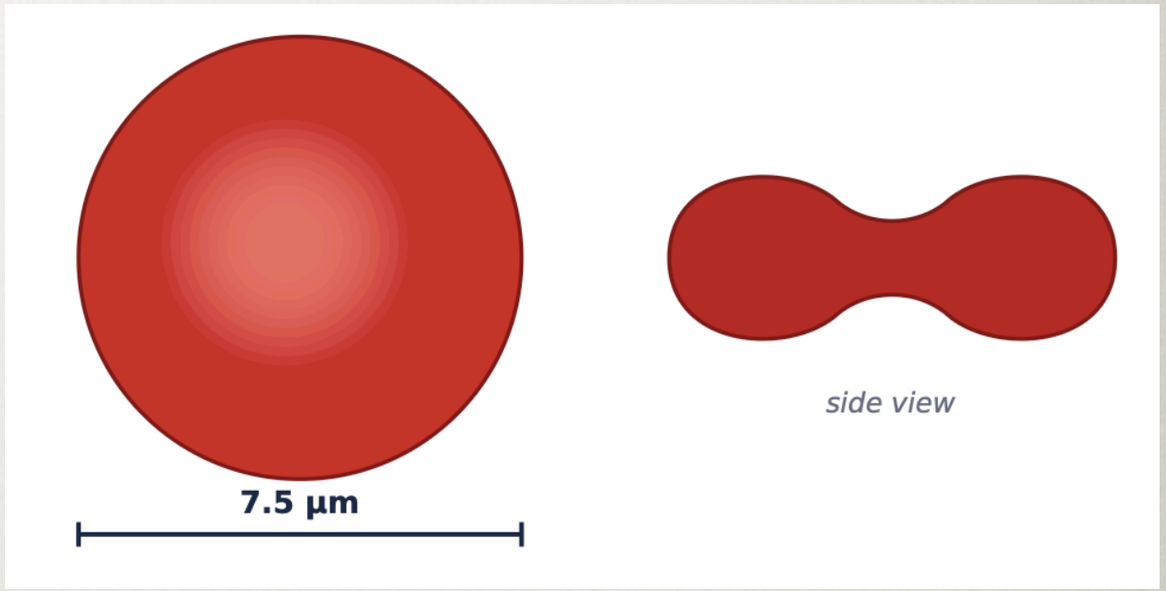
$$10^{-6}\,\mathrm{m} = 1\,\mu\mathrm{m}$$

Instrument



optical microscope

Object

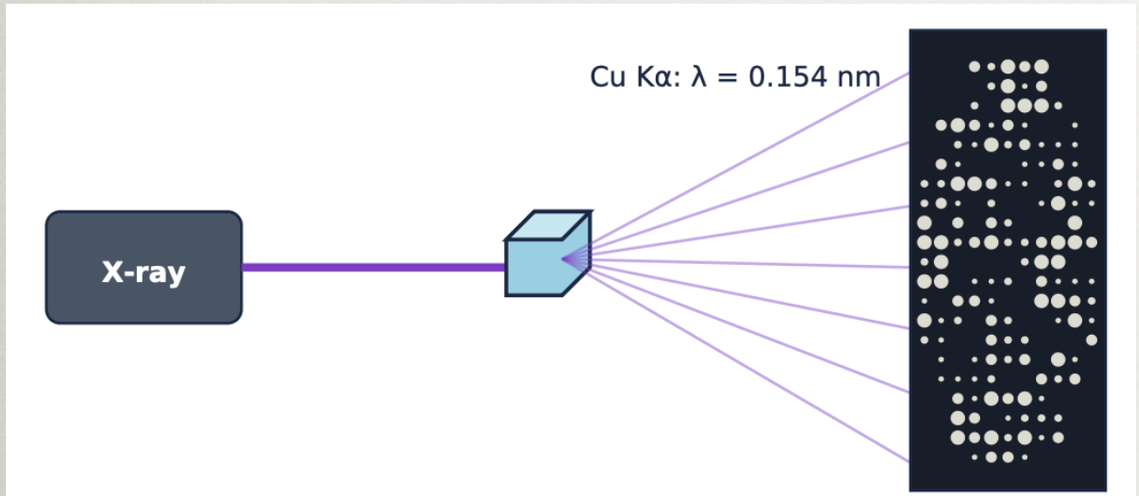


red blood cell

EXPERIMENTAL TOOLS:

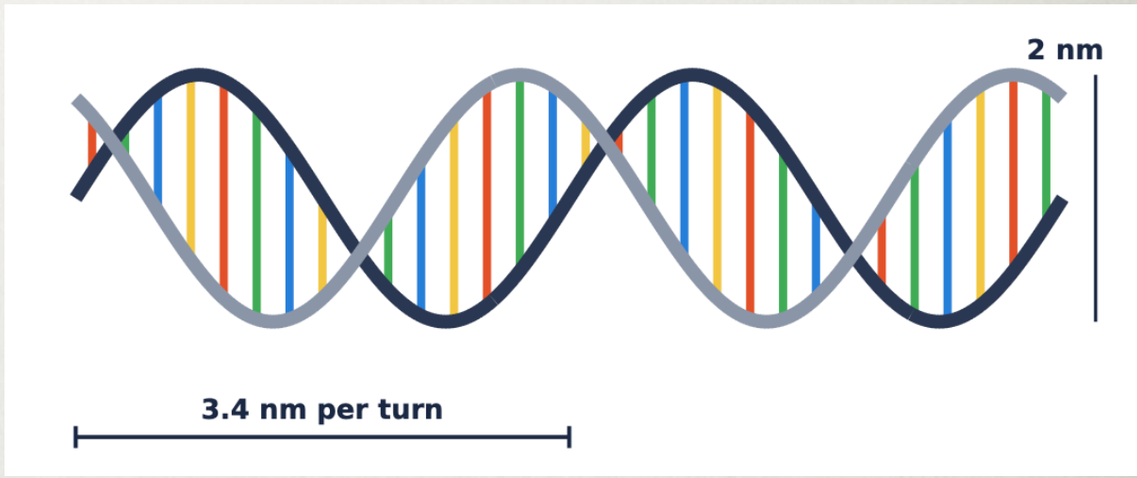
$$10^{-9}\,\mathrm{m} = 1\,\mathrm{nm}$$

Instrument



x-ray crystallography

Object

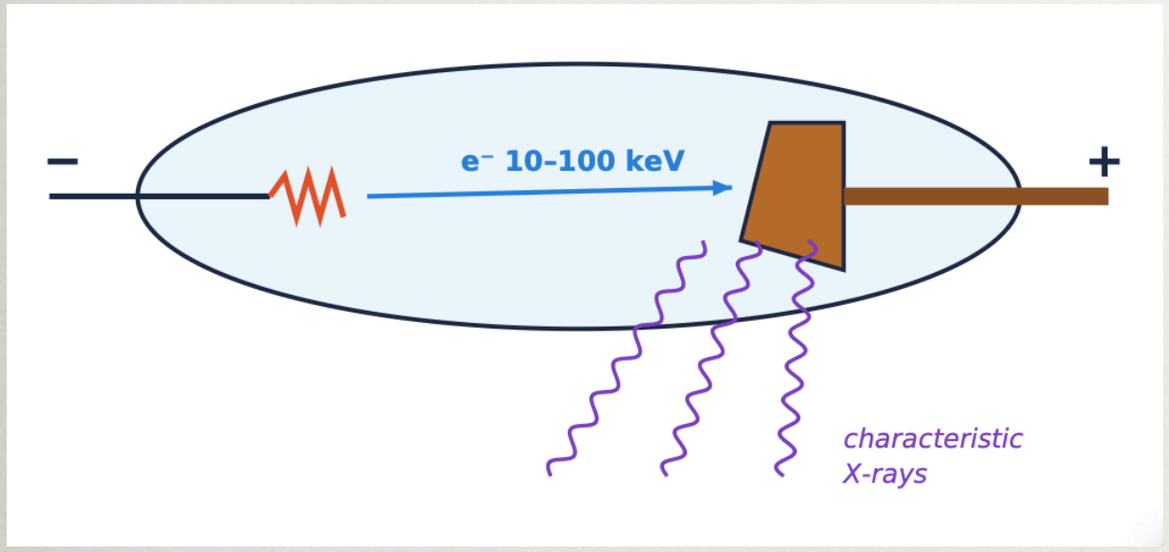


DNA double helix

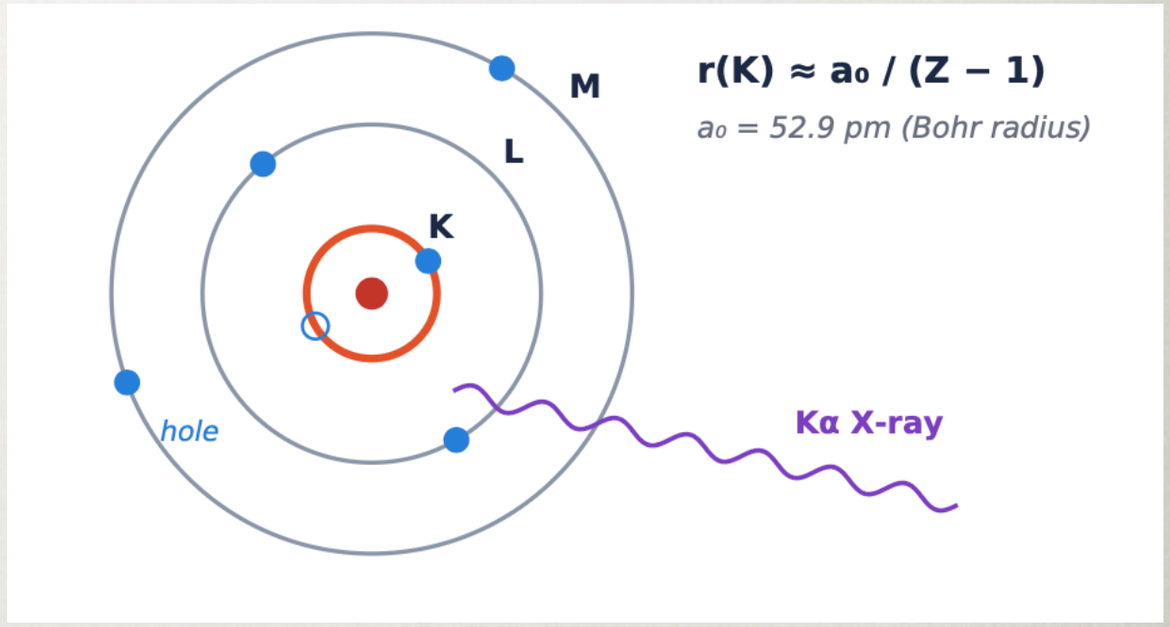
EXPERIMENTAL TOOLS:

$$10^{-12}\,\mathrm{m} = 1\,\mathrm{pm}$$

Instrument



Object



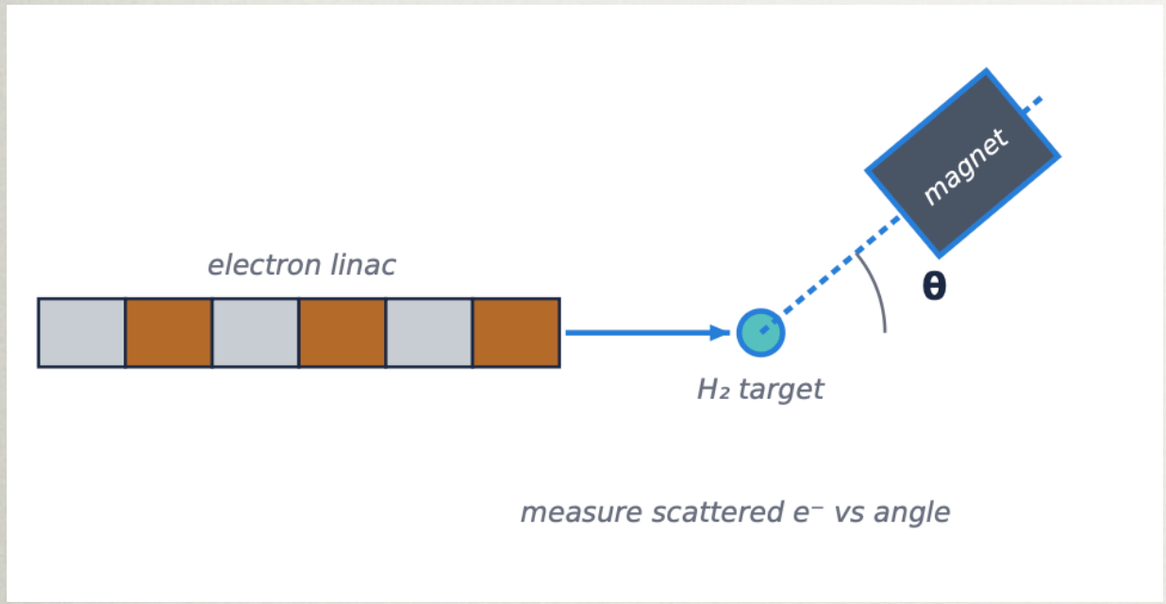
x-ray spectrometry

inner shells of atoms

EXPERIMENTAL TOOLS:

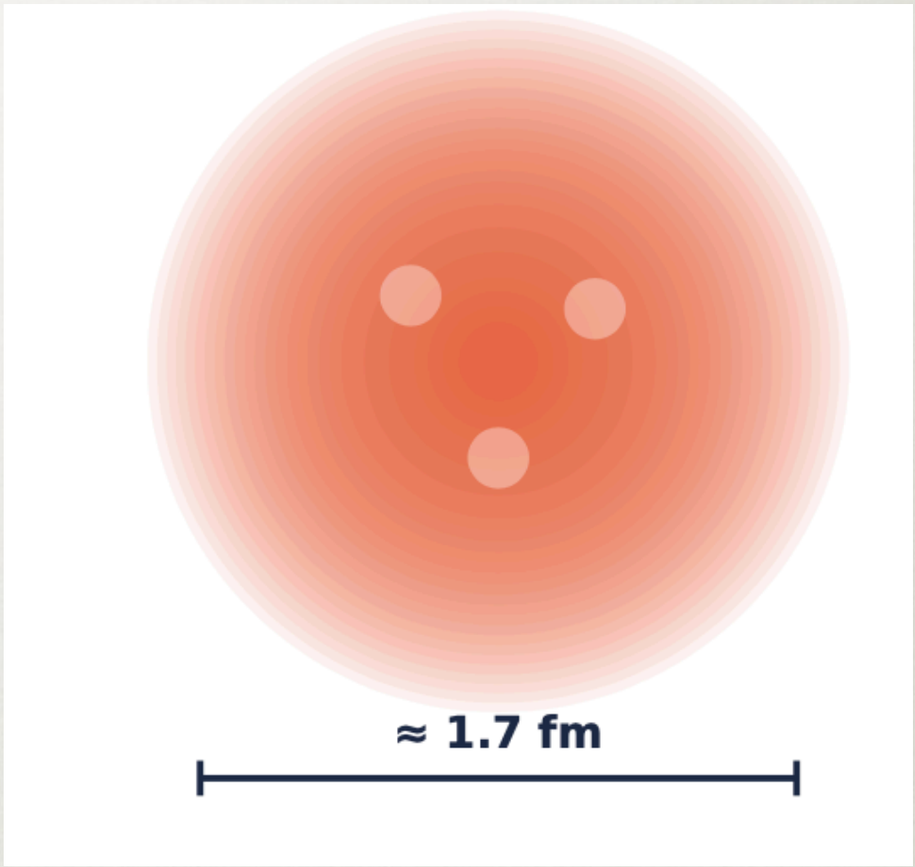
$$10^{-15}\,\mathrm{m} = 1\,\mathrm{fm}$$

Instrument



electron linac + spectrometer

Object

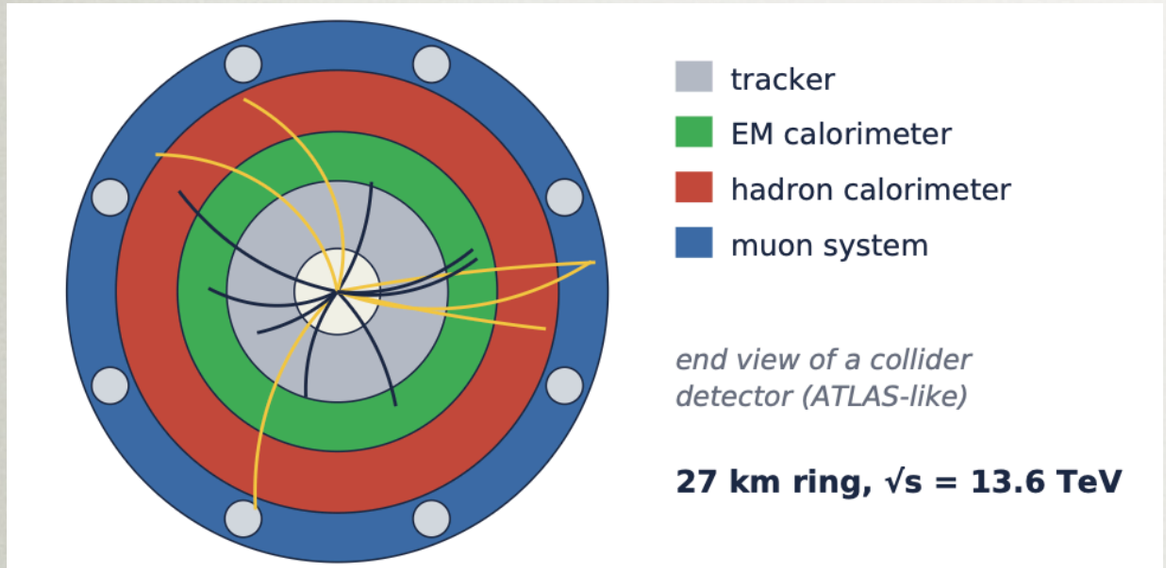


proton

EXPERIMENTAL TOOLS:

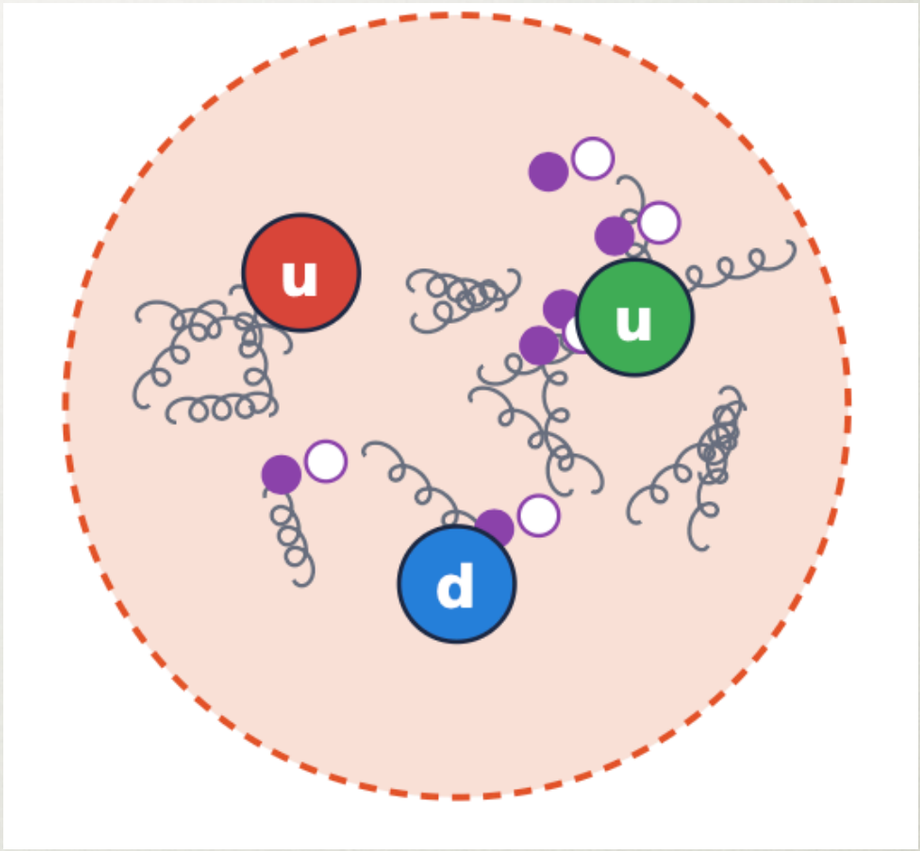
$$10^{-18}\,\mathrm{m} = 1\,\mathrm{am}$$

Instrument



Large Hadron Collider

Object

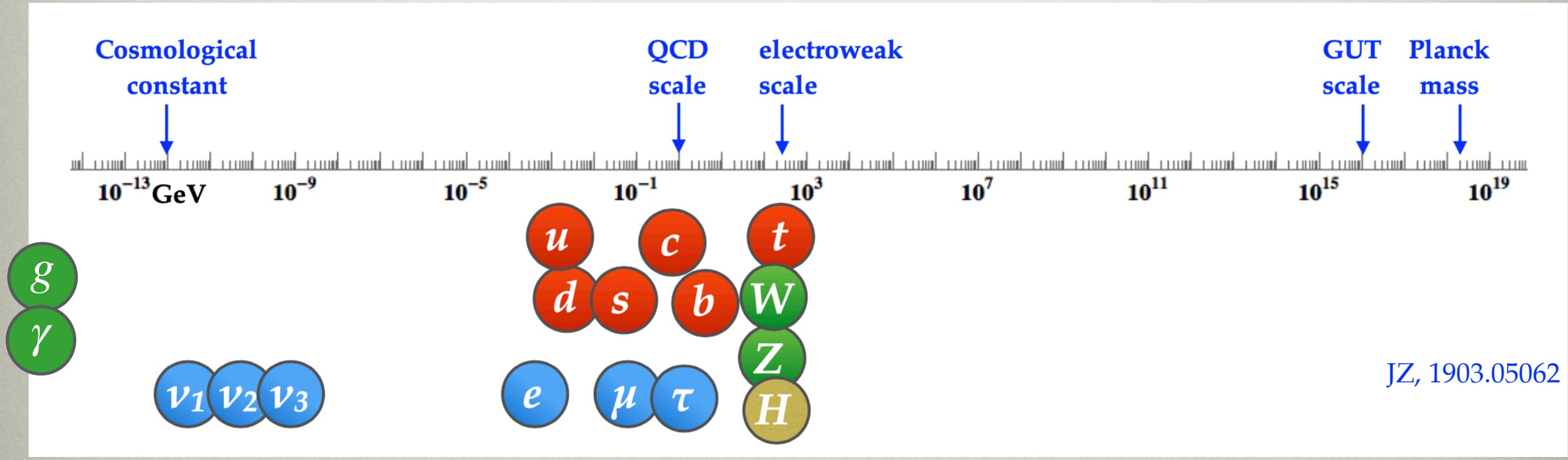


Higgs boson, point-like quarks

Why?

PARTICLE PHYSICS

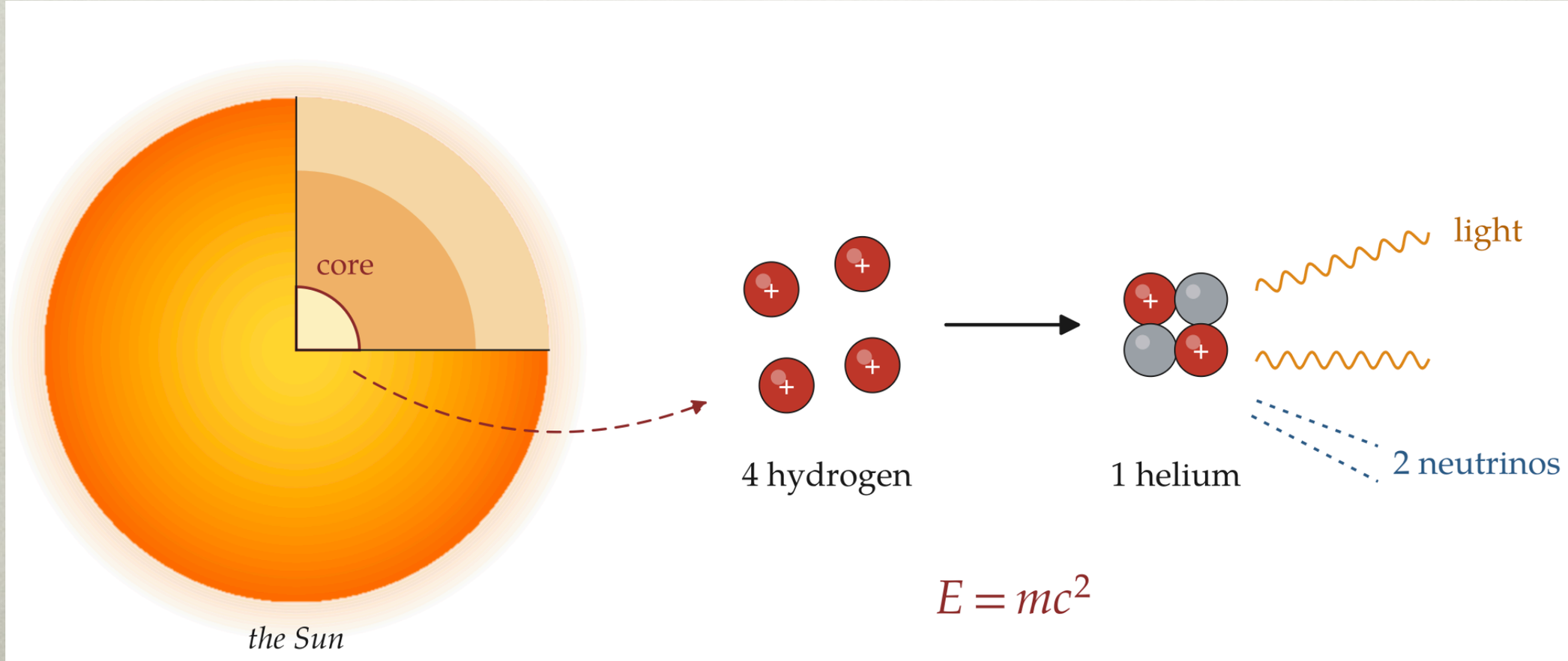
- goal of particle physics: discover laws of nature
  - have we discovered all the forces?
  - all the particles?



JZ, 1903.05062

SOME ANSWERS TO FUNDAMENTAL QUESTIONS

- answers to selected fundamental questions
  - why does the sun shine?

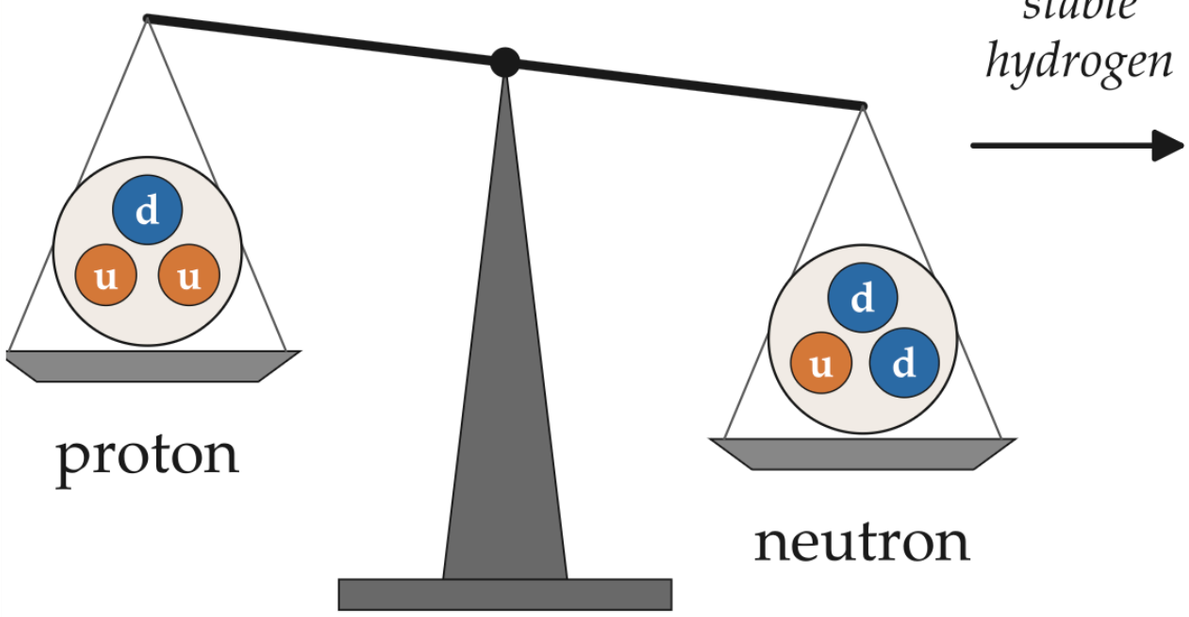


SOME ANSWERS TO FUNDAMENTAL QUESTIONS

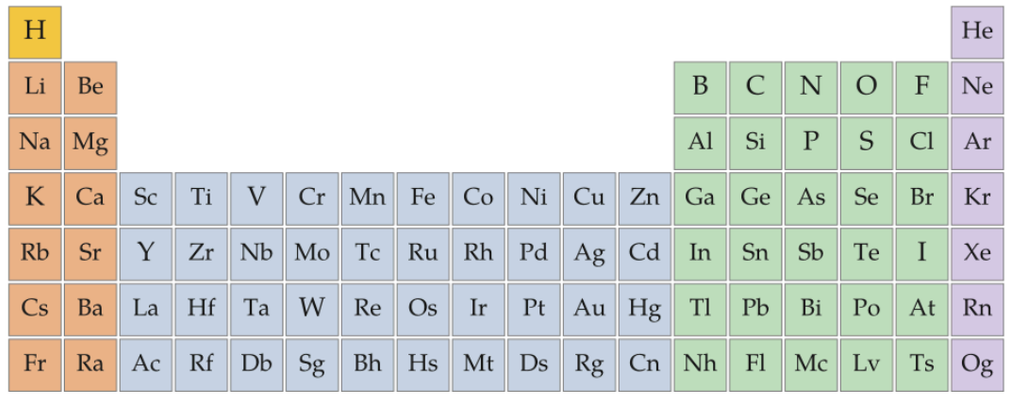
- answers to selected fundamental questions
  - why is it possible for chemistry to exist?
    - why is the neutron heavier than the proton?

down quark heavier than up quark

(*Higgs couplings*)



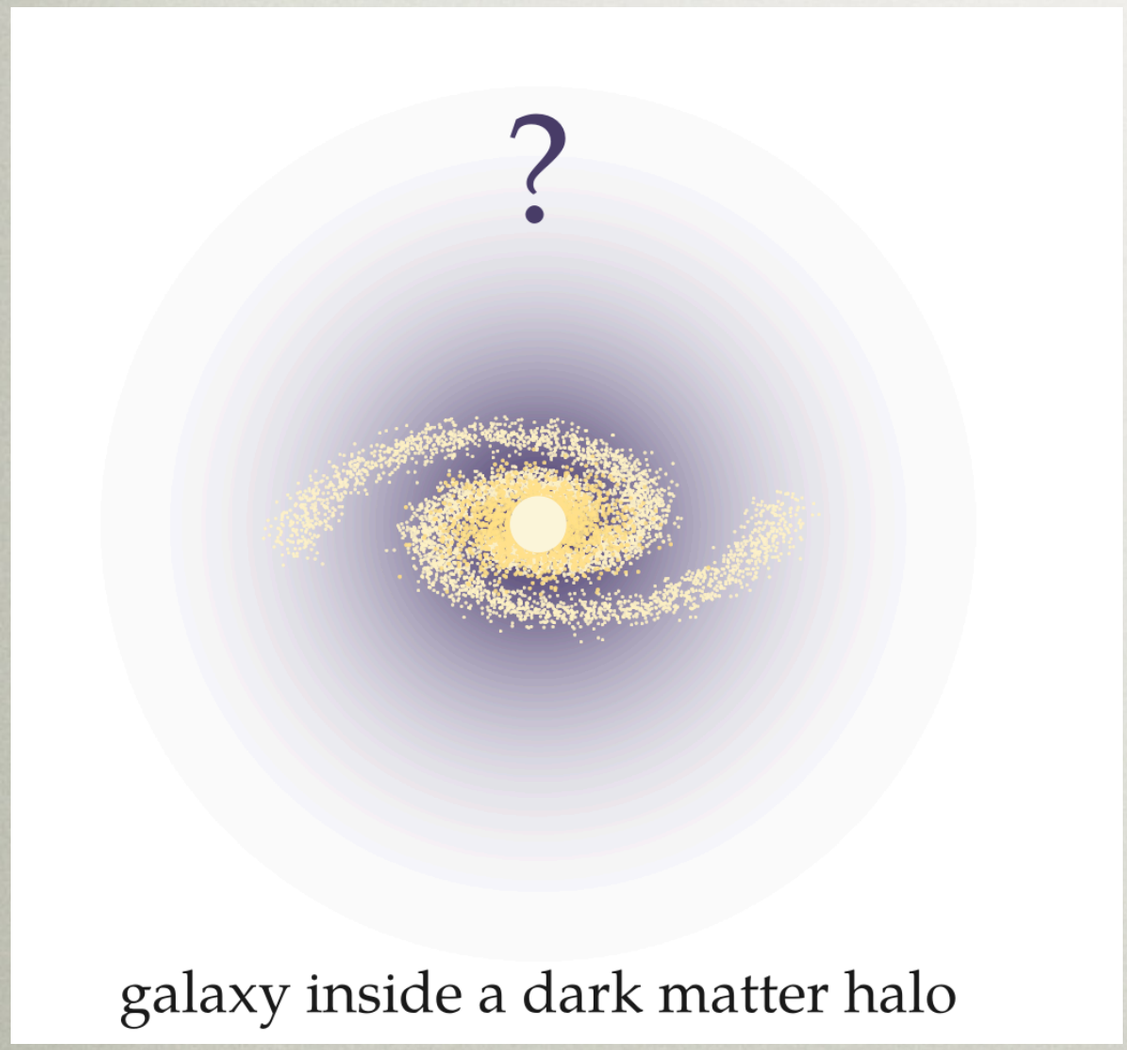
stable hydrogen

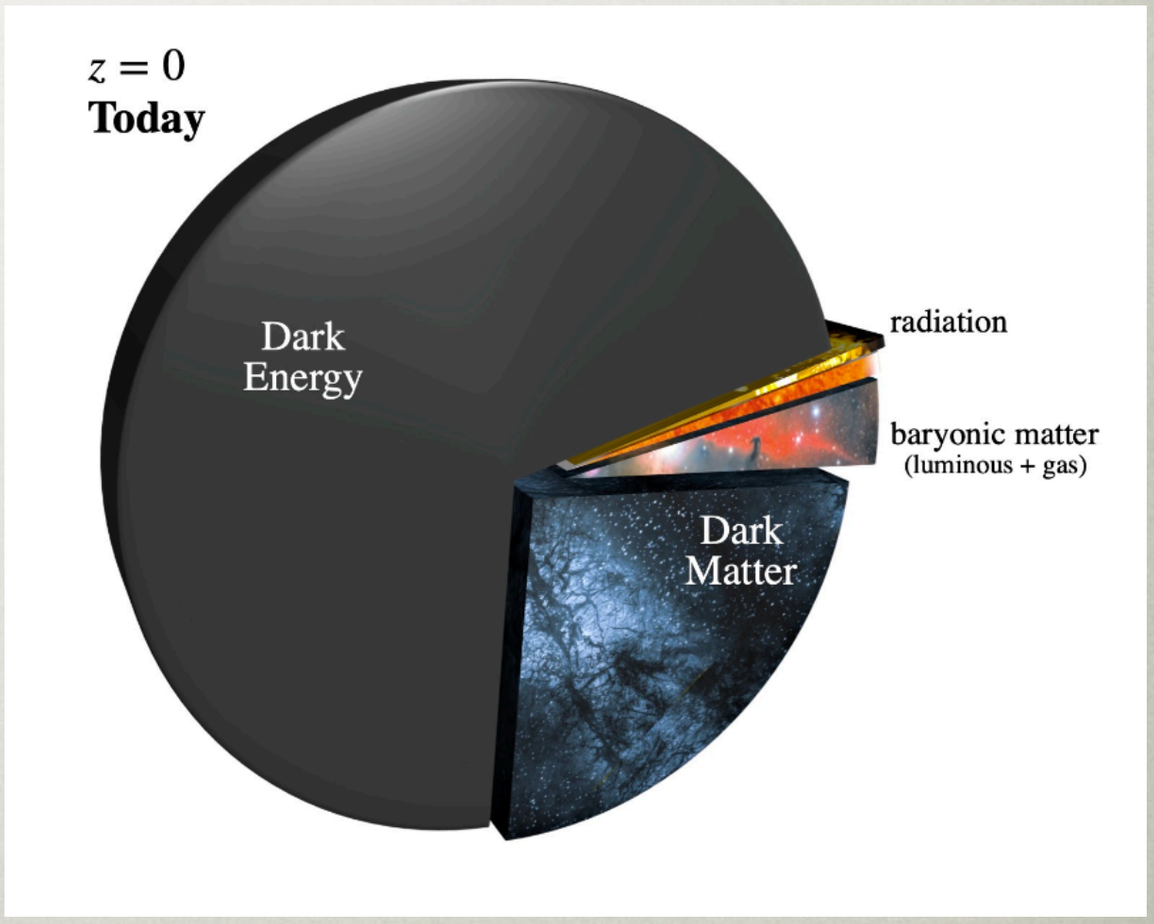


atoms and chemistry

SOME OPEN QUESTIONS

- what is dark matter?

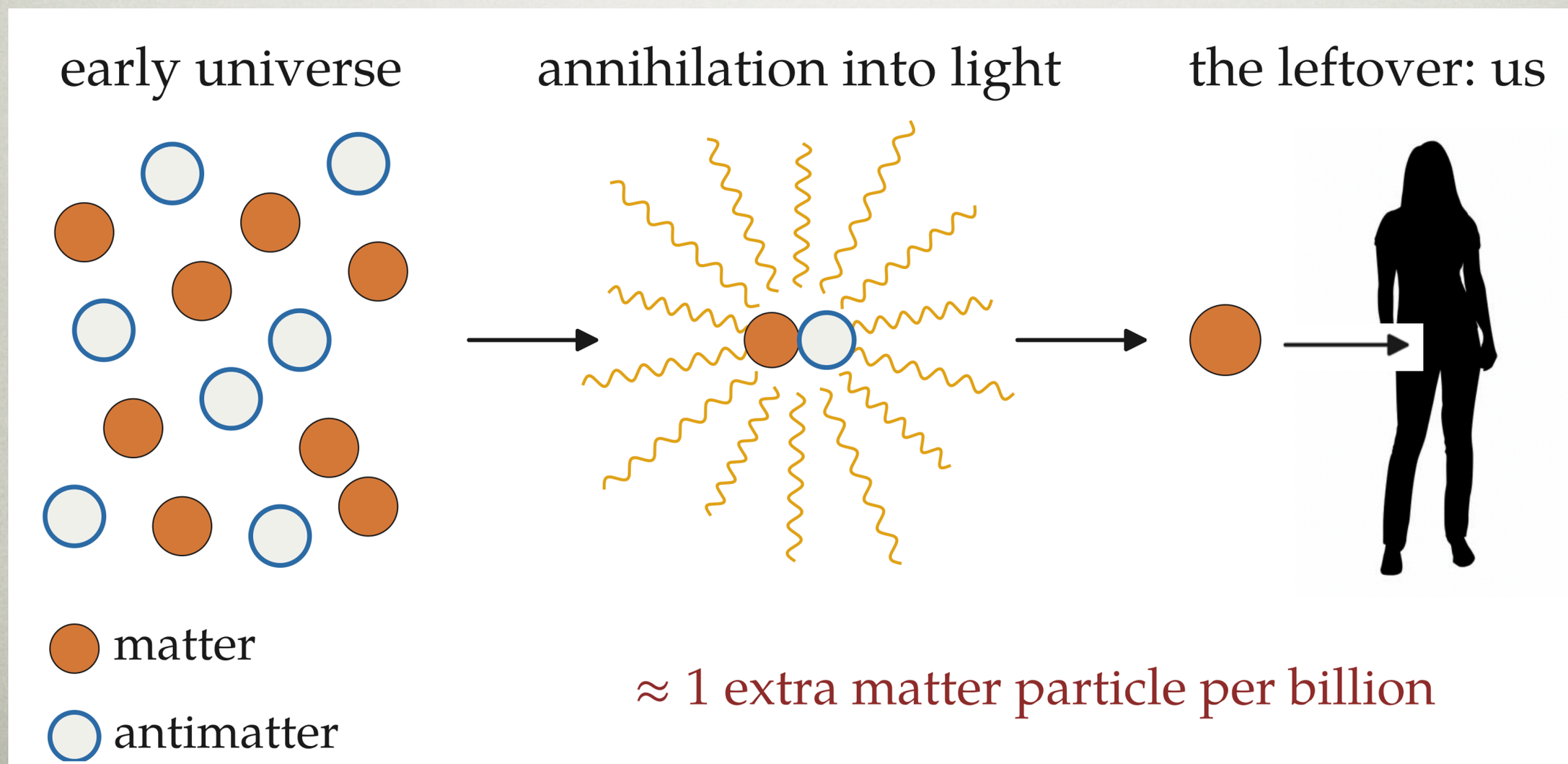




Cirelli, Strumia, JZ, 2406.01705

SOME OPEN QUESTIONS

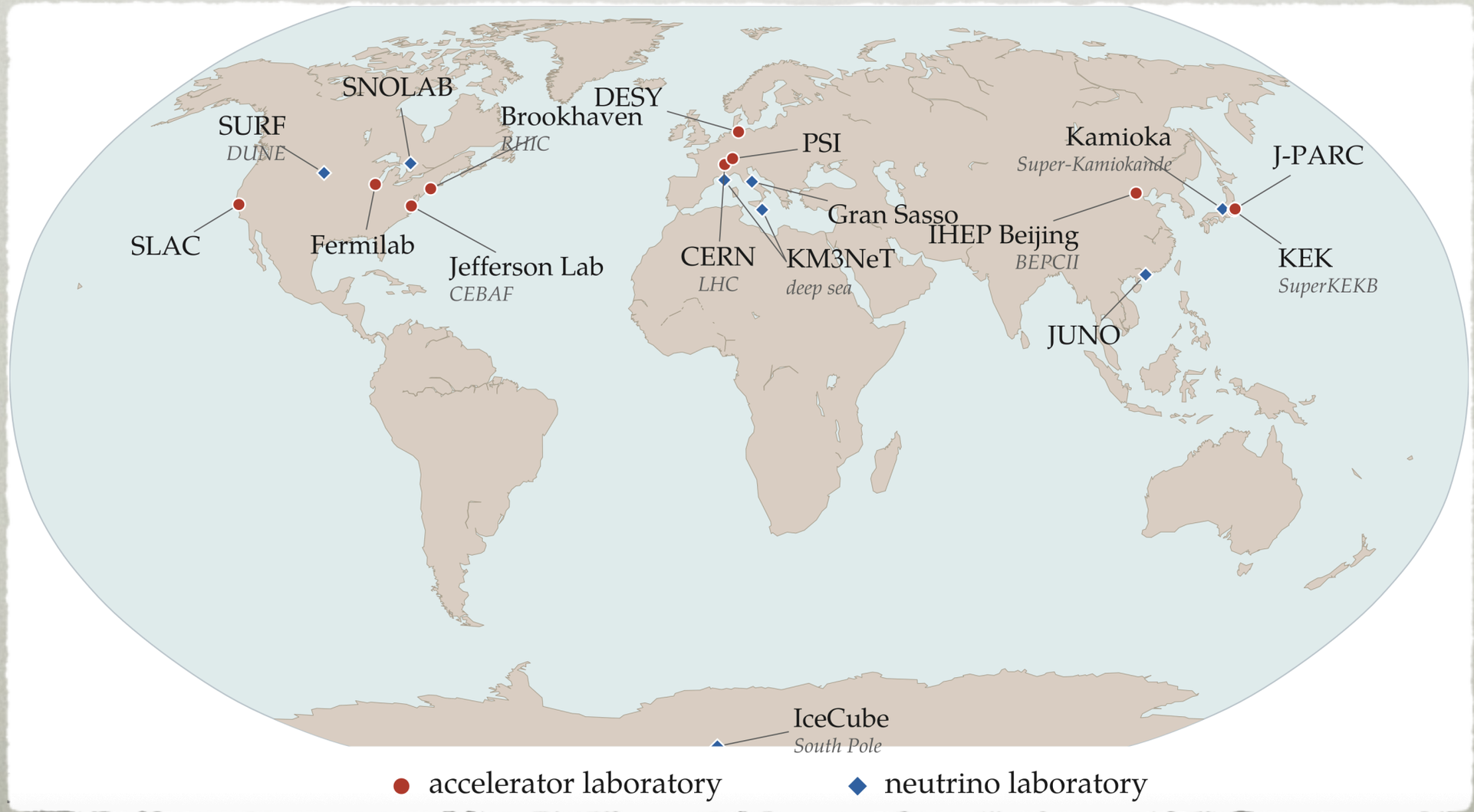
- why are we composed of matter and not antimatter?
  - how is it possible that we exist?
  - baryogenesis



WHERE?

TYPICAL HEP EXPERIMENTS

- typical HEP experimental collaborations large $\mathcal{O}(10^3)$ people
  - experiments concentrated in large international laboratories, CERN, Fermilab, KEK,…
- theoretical groups smaller, typically up to $\mathcal{O}(10)$ people

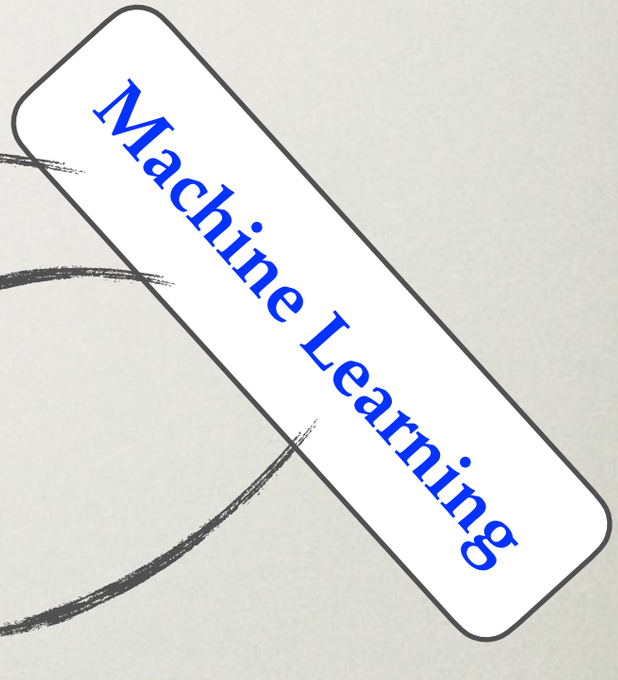


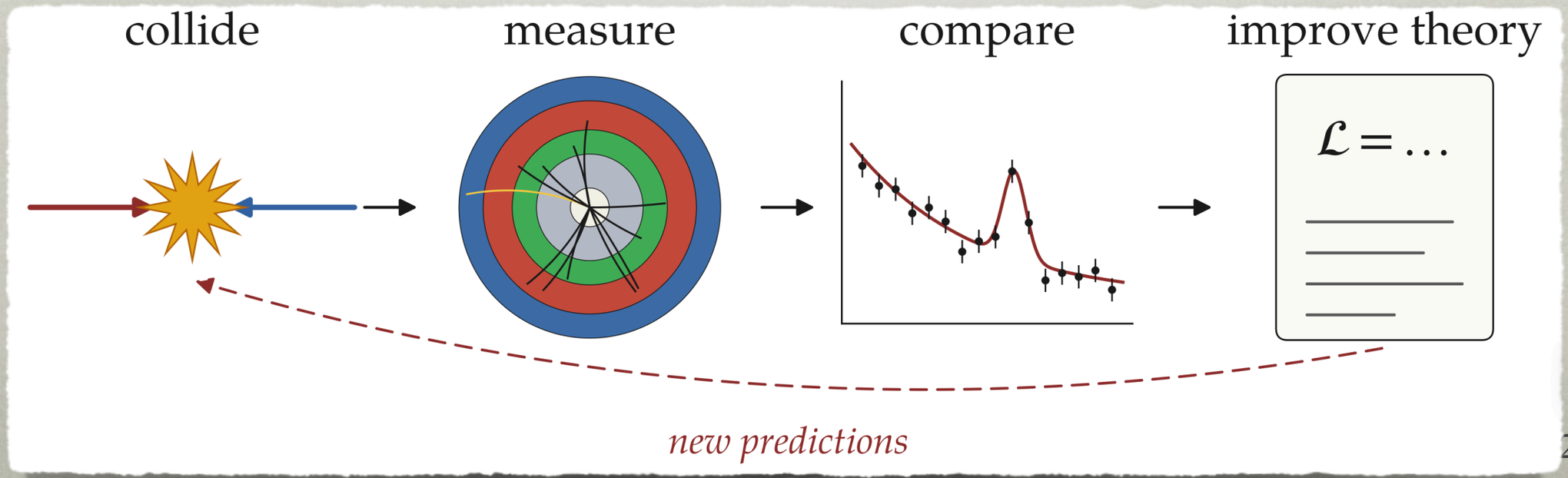
2026

HOW?

HEP measurements

- typical high energy experiment
  - collide particles
  - measure outcomes
  - interpret / compare to theory
  - improve / adjust theory





ROAD TO DISCOVERY

predictions

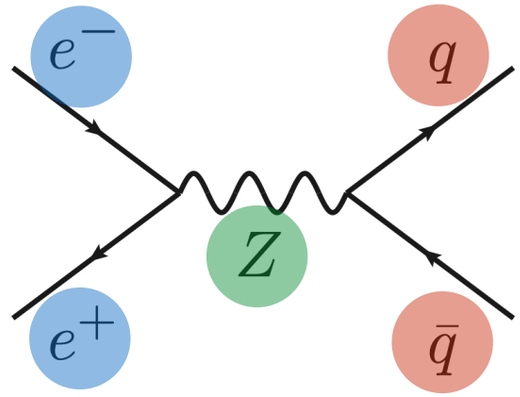


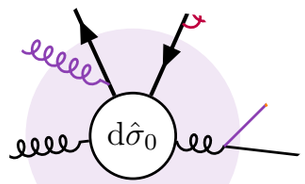


measurements

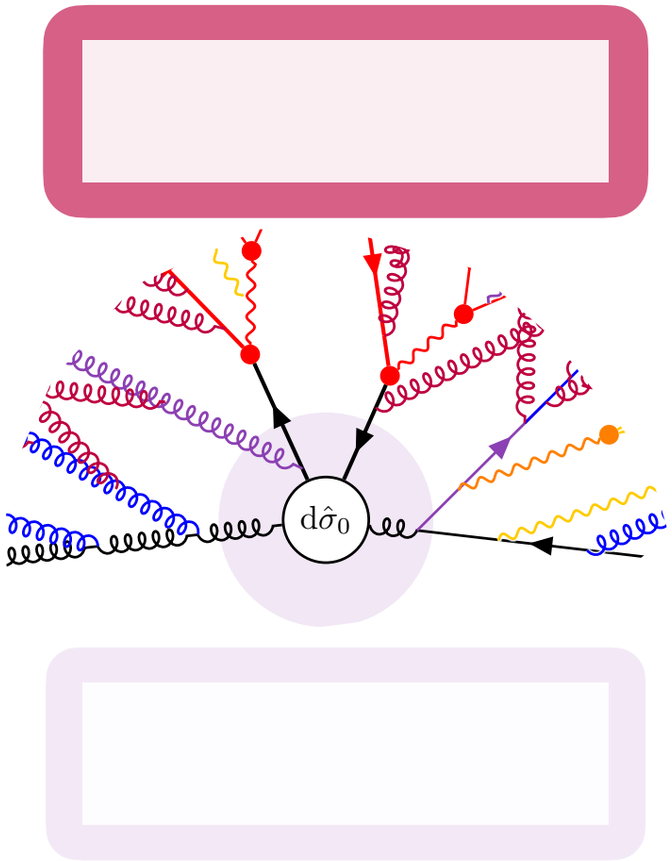
- any deviations of experimental results from predictions?
  - reliable theory predictions essential

THEORY PREDICTIONS IN COLLIDER PHYSICS

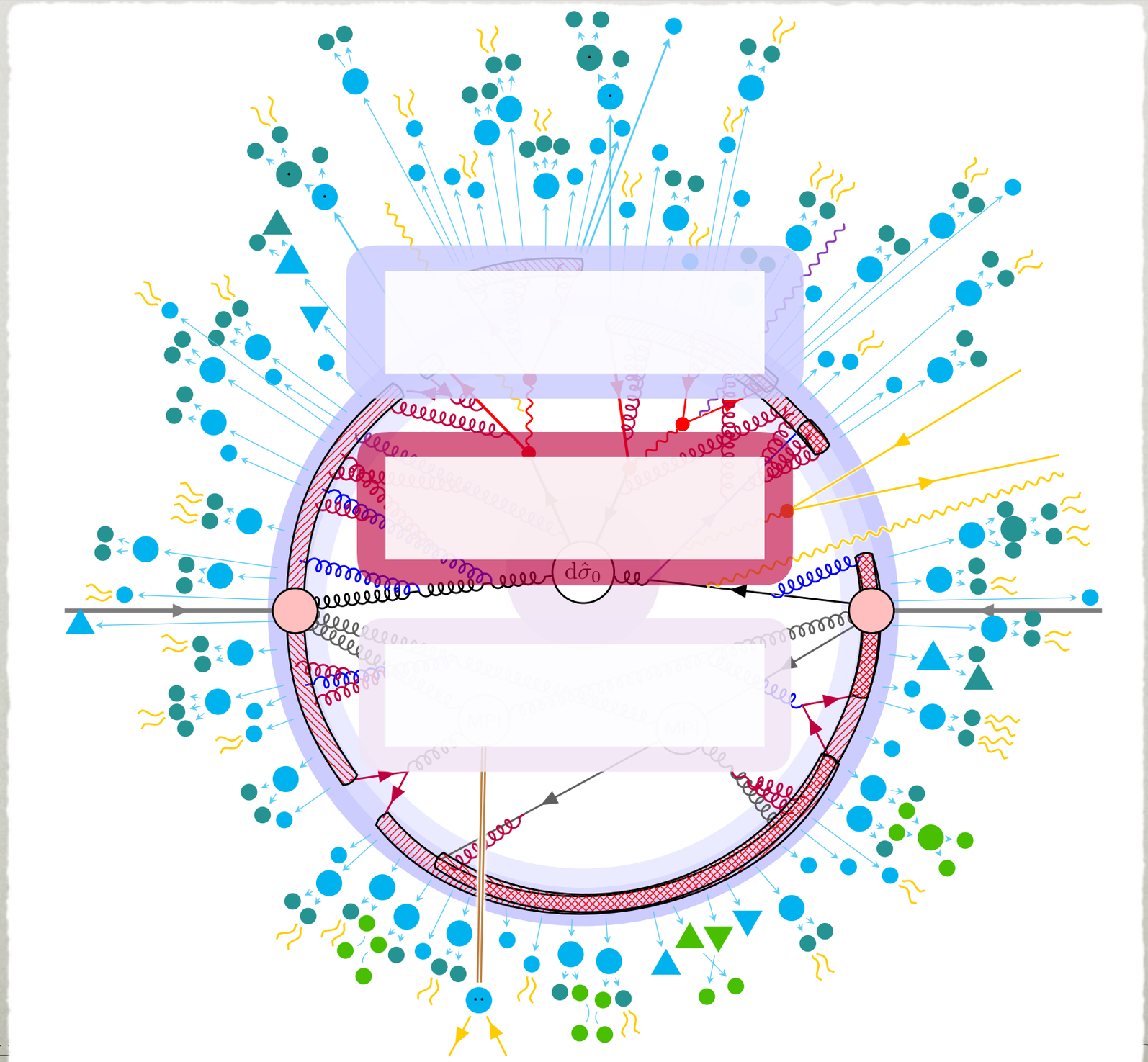




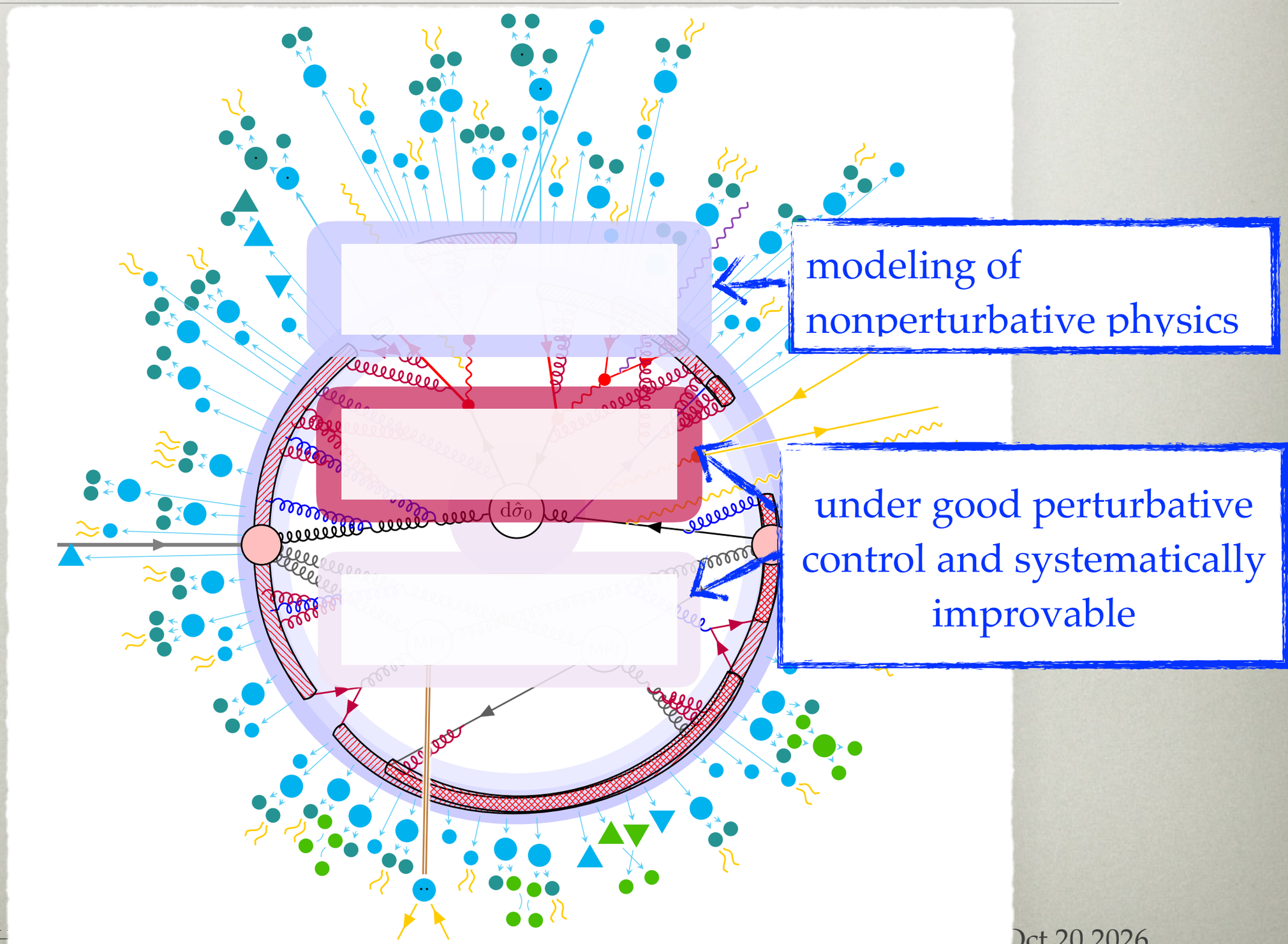
THEORY PREDICTIONS IN COLLIDER PHYSICS



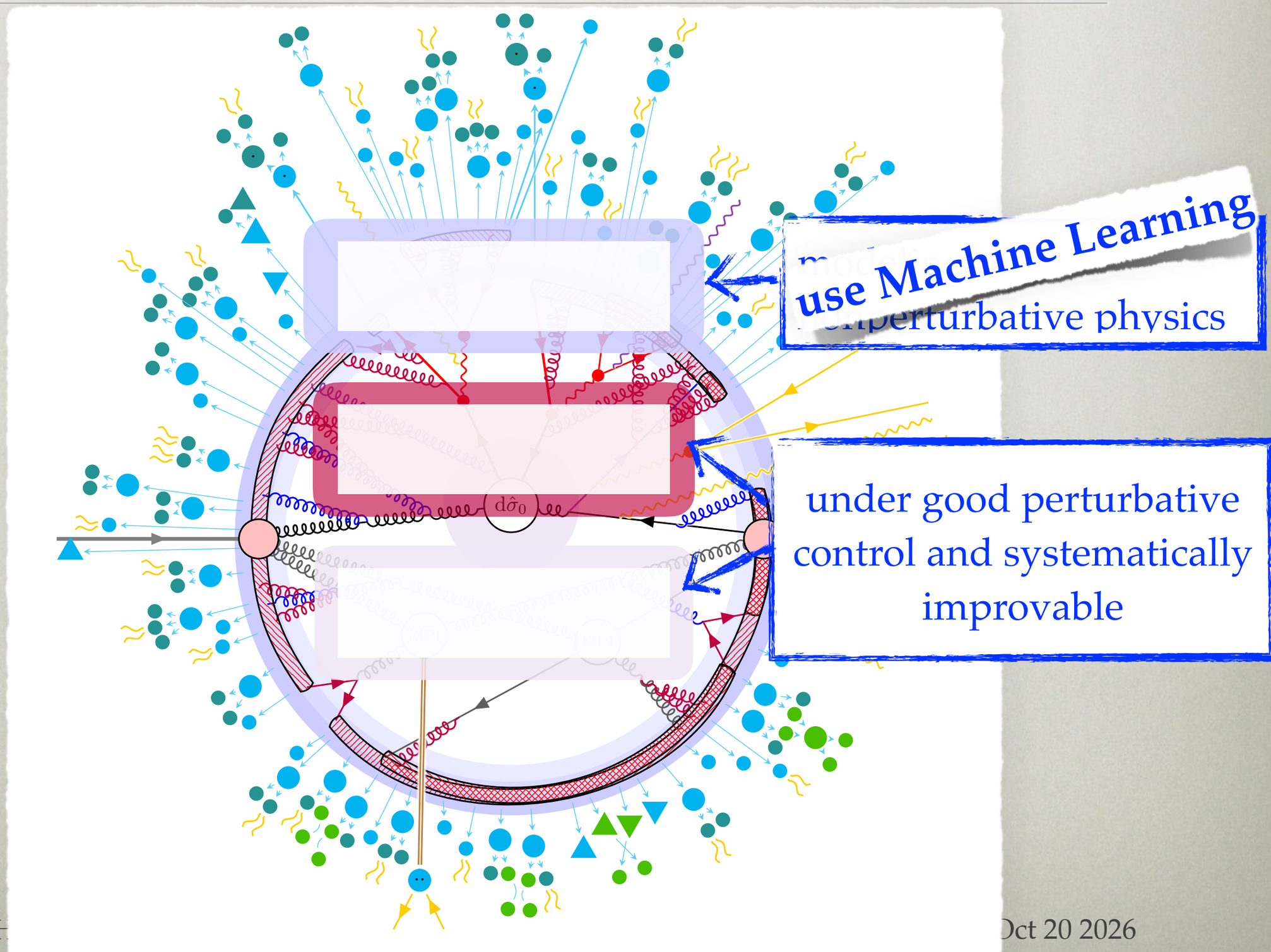
THEORY PREDICTIONS IN COLLIDER PHYSICS



THEORY PREDICTIONS IN COLLIDER PHYSICS



THEORY PREDICTIONS IN COLLIDER PHYSICS



SIMULATING HADRONIZATION WITH ML

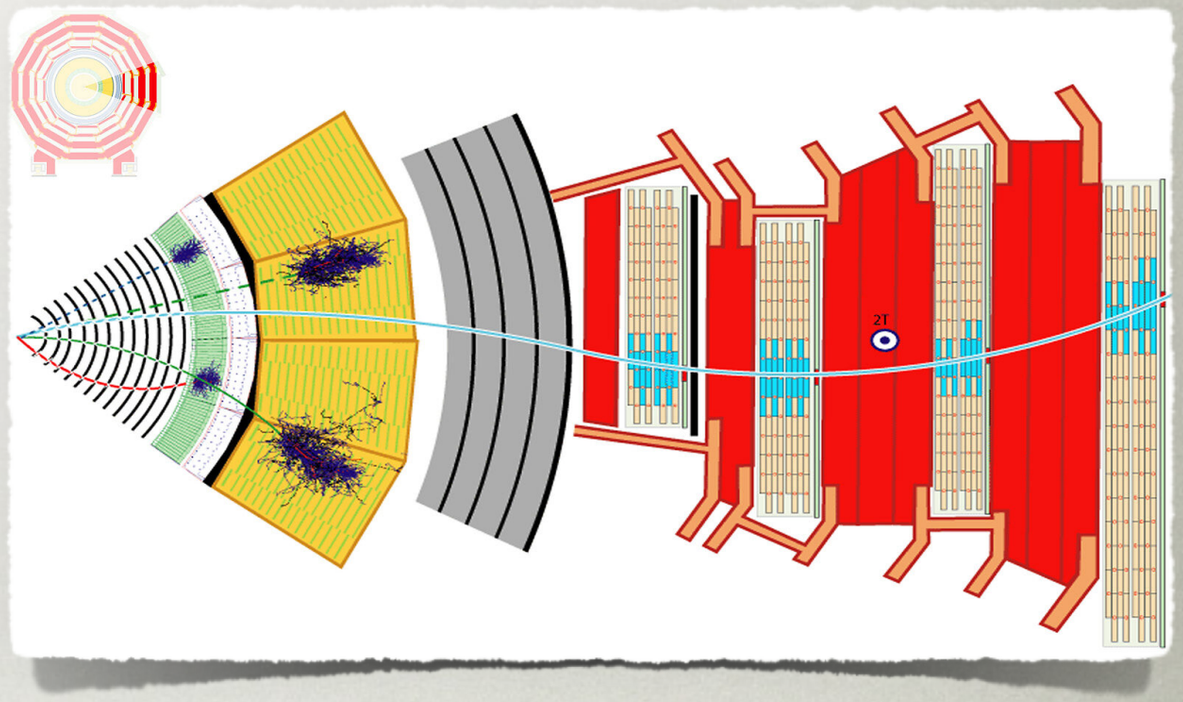
THE REST OF THE TALK

- should one care about hadronization?
- how to improve hadronization models
- inverse problem for hadronization
- HD-sense score
  - of wider utility…

SHOULD ONE CARE ABOUT HADRONIZATION?

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 hadronization well 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

- understanding hadronization important for precision measurements
- top mass
  - b quark hadronization model uncertainties on indiv. measurements ~0.1-0.2 GeV
  - only a factor of a few below WA
- $\alpha_S(M_Z)$ determinations from $e^+e^- \to$ hadrons
  - shape observables dropped from WA due to hadroniz. uncertainties

Huston, Rabbertz, Zanderighi, QCD review in 2025 PDG

- electroweak physics at FCC-ee, Z-pole measurements
  - projections assume improvement on hadronisation uncertainties
    - conservative scenario: $5\times - 10\times$
    - aggressive scenario: $50\times$

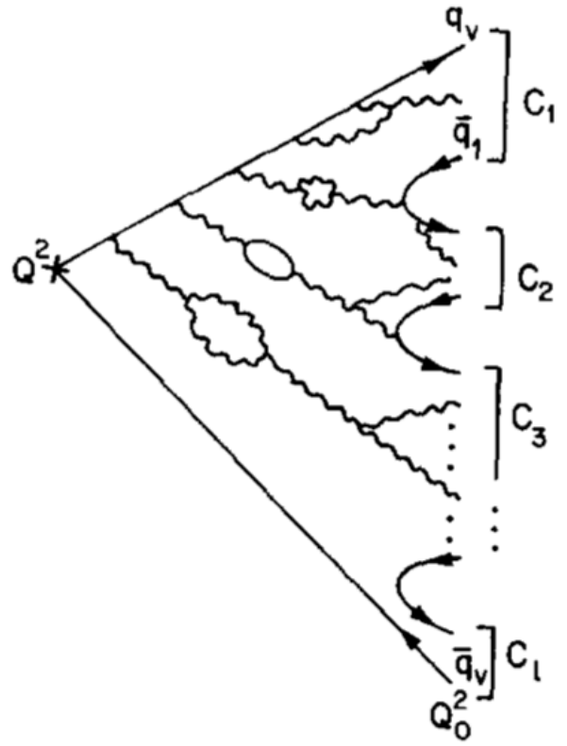
Physics Briefing Book, 2511.03883

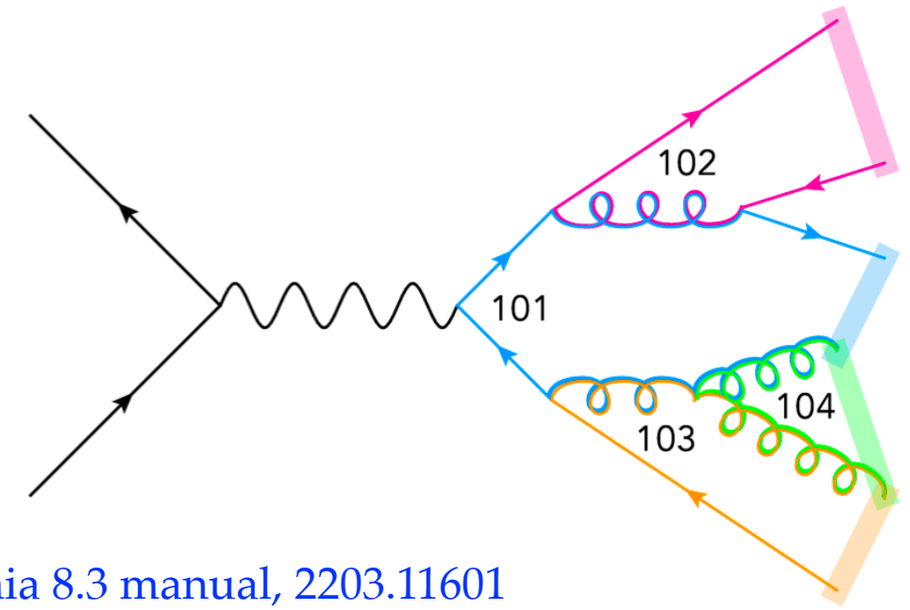
How to improve hadronization models

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

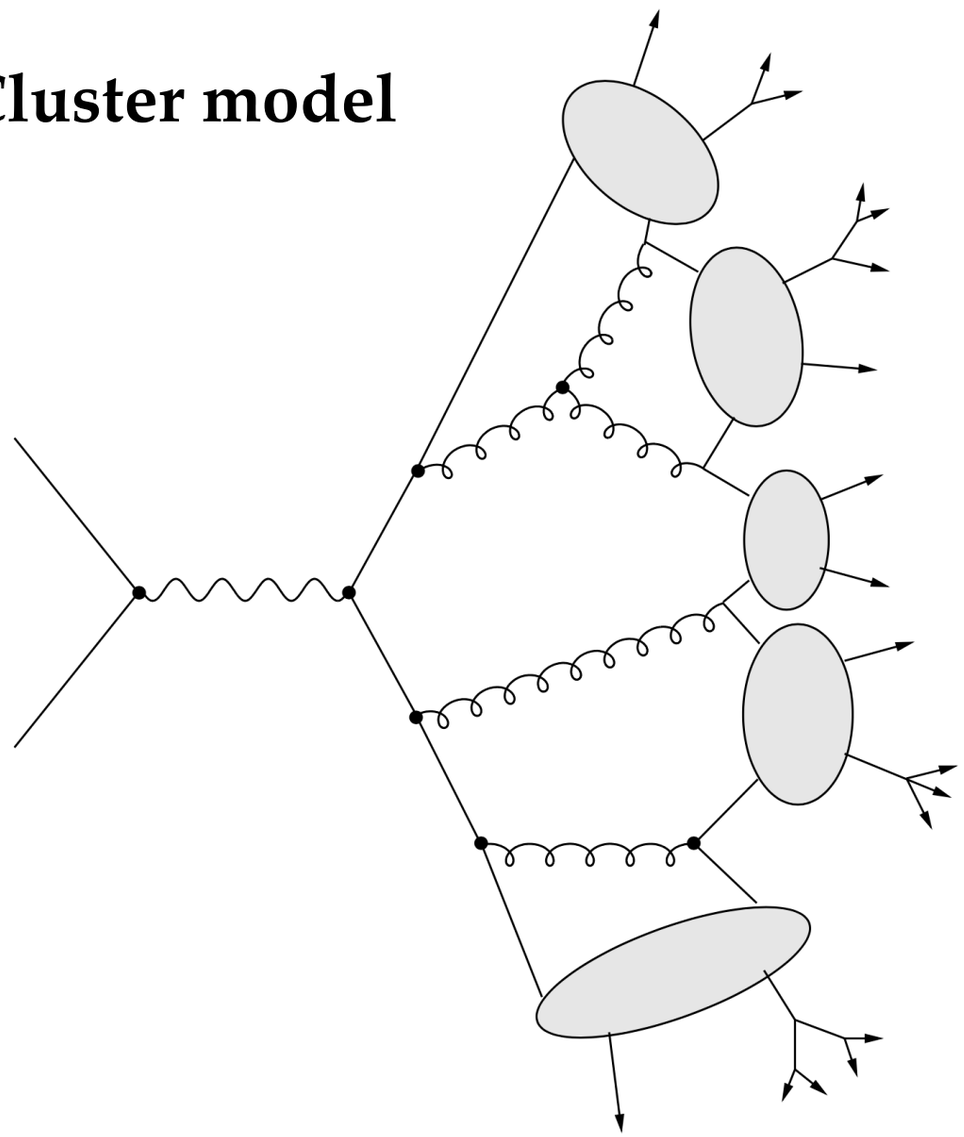




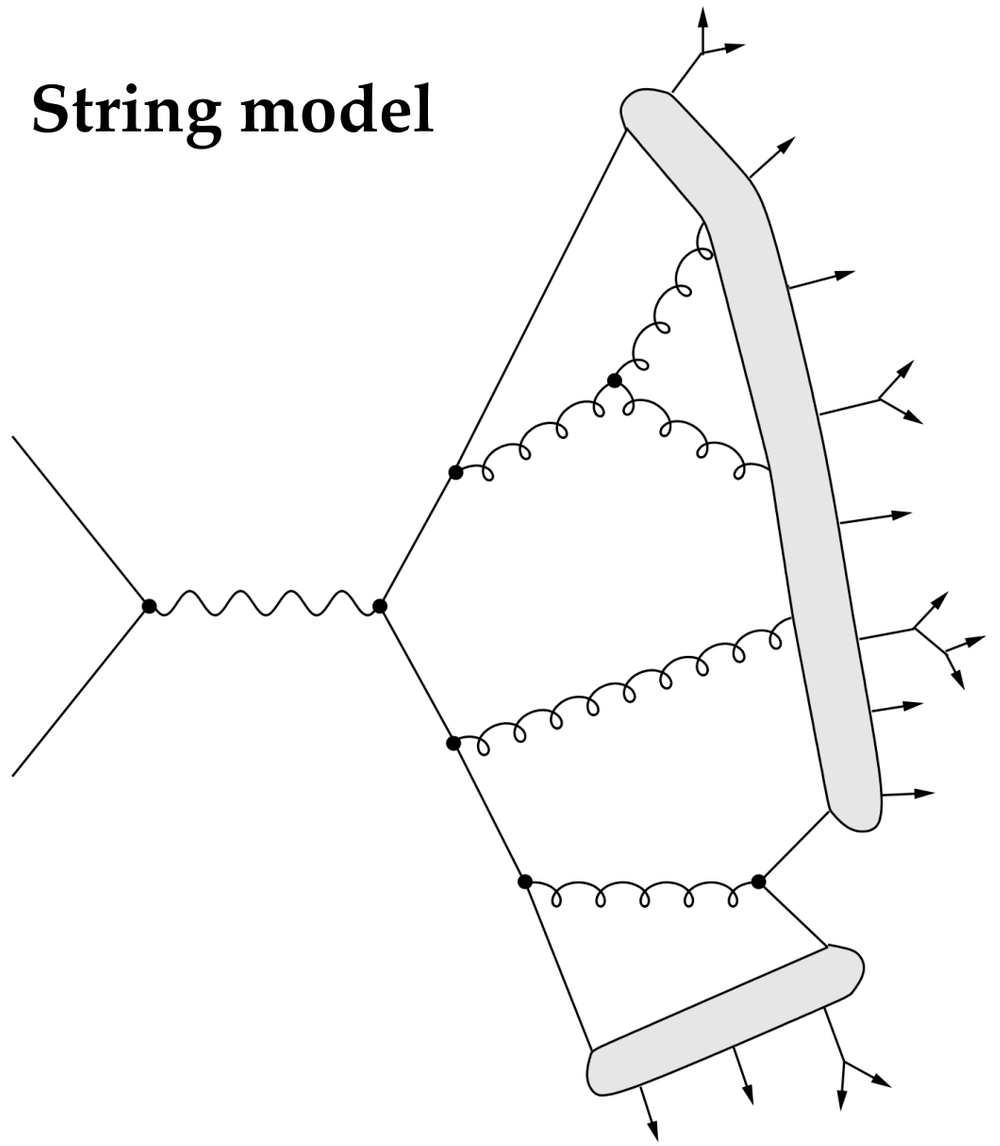
Pyhia 8.3 manual, 2203.11601

CLUSTER VS. STRING MODEL

Cluster model



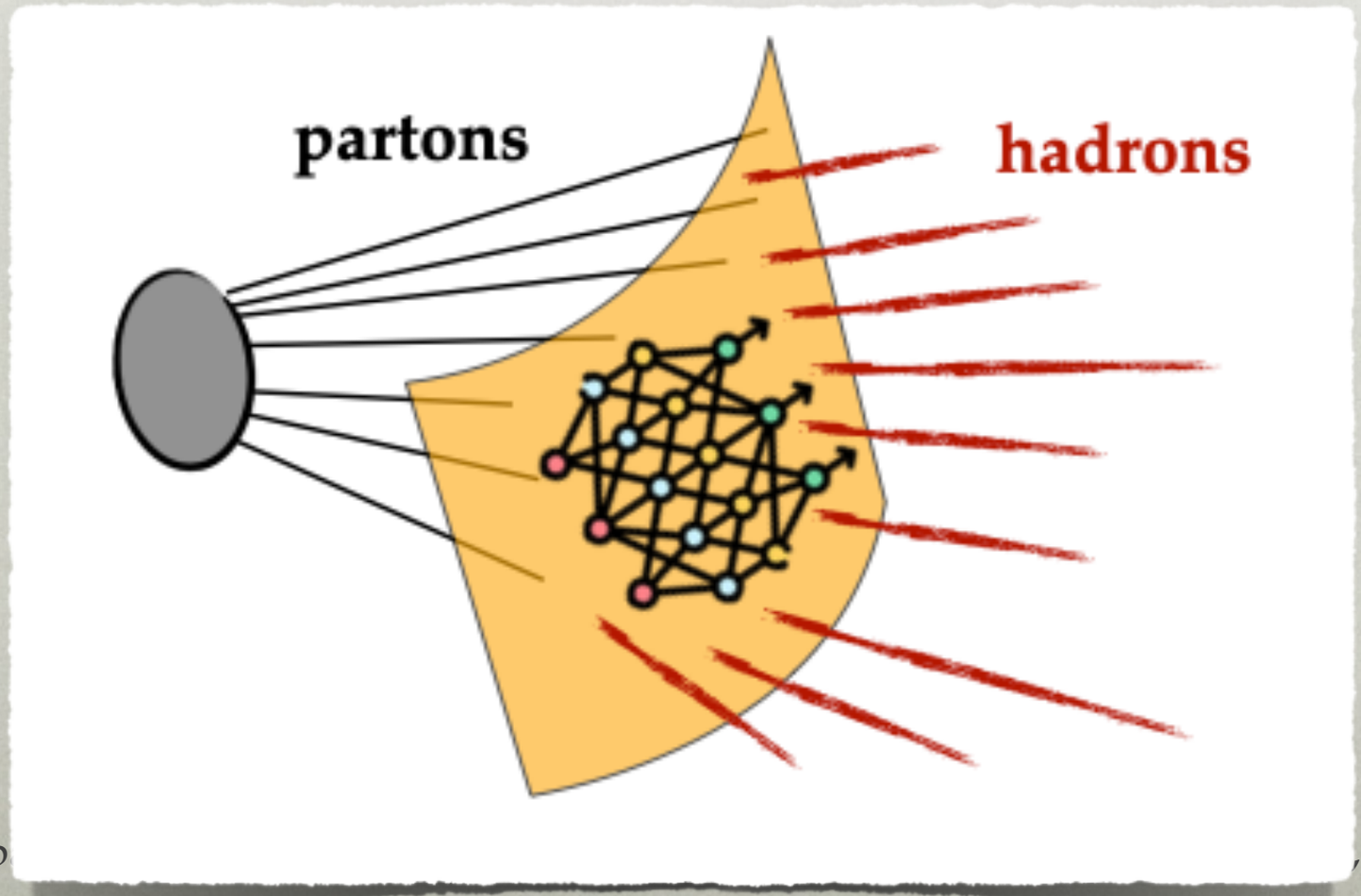
String model



Figs from Webber, 1999

ML alternative

- well suited for ML based description
- point cloud (partons) to point cloud (hadrons)
  - the challenge is how to impose theory constraints
    - e.g. local quark hadron duality



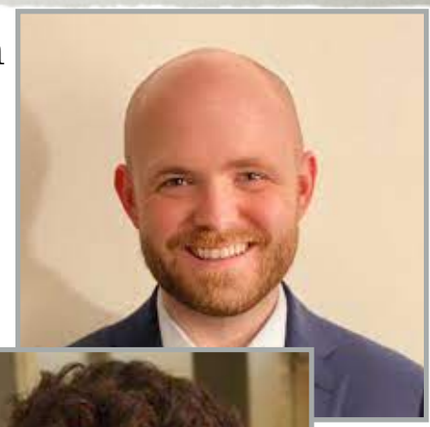
MACHINE LEARNING HADRONIZATION

MLhad: Assi,Bierlich, Gambhir, Ilten, Menzo, Mrenna, Szewc, Wilkinson, JZ, https://gitlab.com/uchep/mlhad


B. Assi

MLHAD

M. Wilkinson




C. Bierlich


R. Gambhir


P. Ilten


T. Menzo


S. Mrenna


M. Szewc +JZ

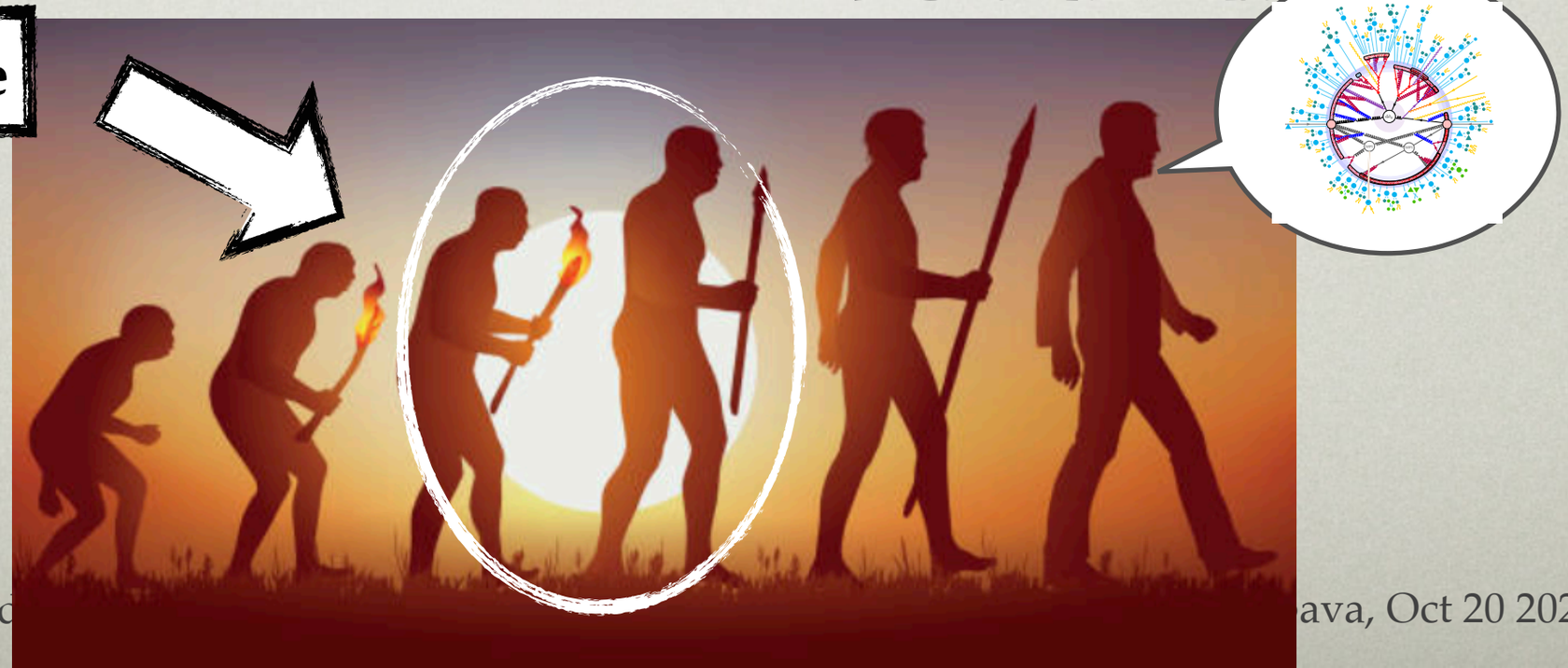
- MLhad: the long term goal:
  - use ML to "parametrize our ignorance" about hadronization, use data + theory constraints

MORE IMMEDIATE GOALS

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

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

we are here



MORE IMMEDIATE GOALS

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

- Pythia event gen. focused
  - easier/faster parameter variation
  - improve on Lund string fragmenta
    function det

inverse problem for hadronization

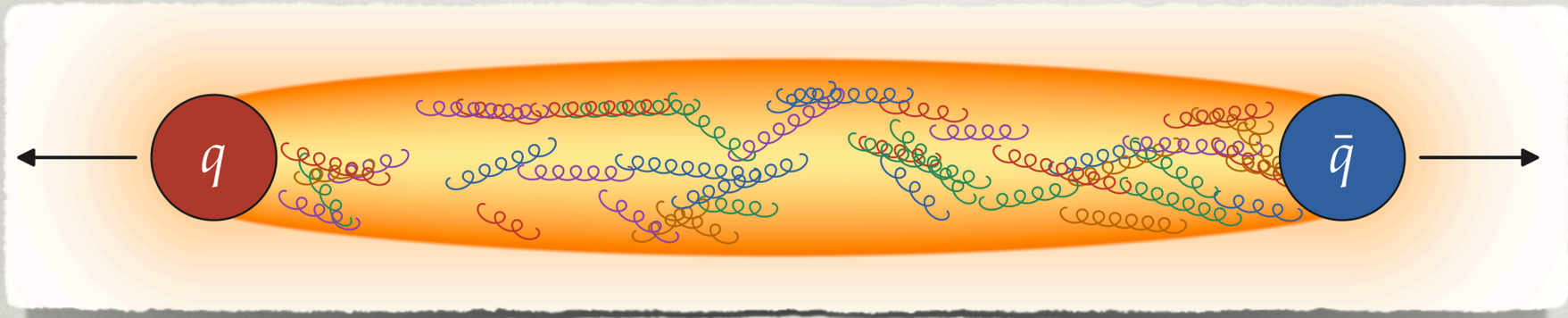
we are here



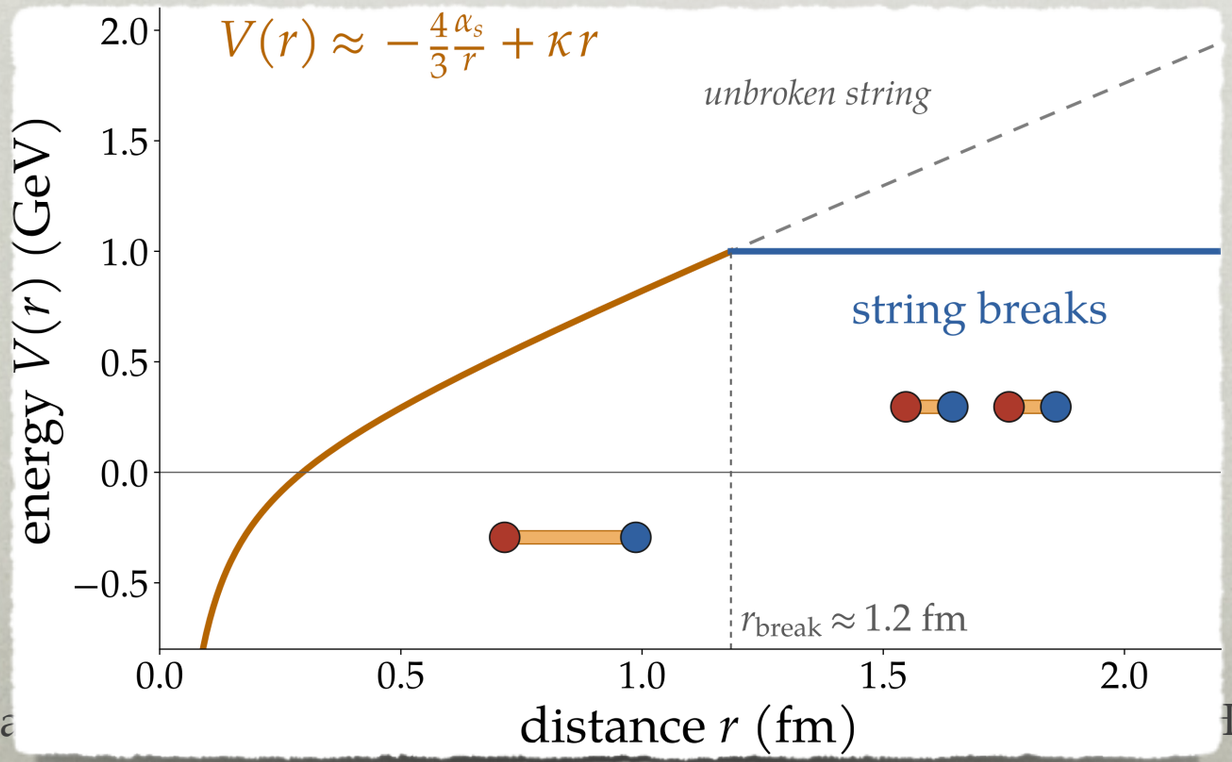
INVERSE PROBLEM FOR HADRONIZATION

Lund string hadronization model

- $q, \bar{q}$ pair connected by a QCD flux tube

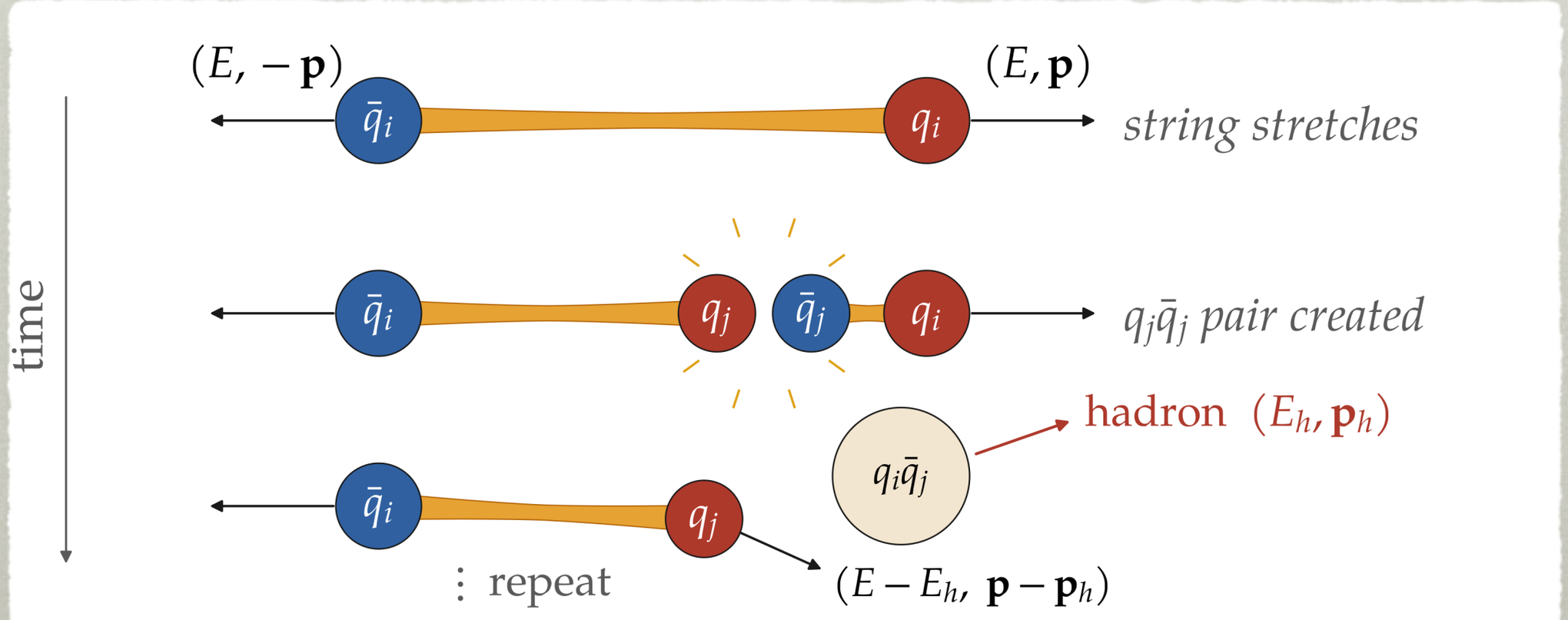


- energy stored in the flux ∝ distance
  - if $q, \bar{q}$ pair too separated ⇒ eng. favorable to break the string



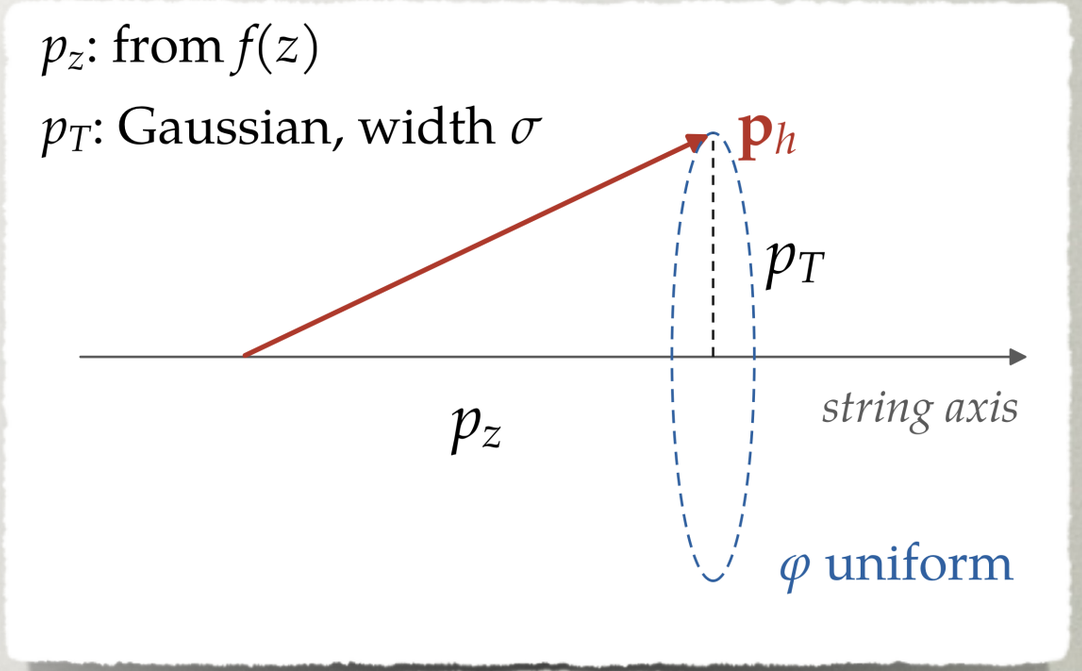
Lund string hadronization model

- end of shower: a collection of strings connecting $q\bar{q}$ pairs
- q and $\bar{q}$ moving apart
- string breaks on one side (randomly chosen)
  - creates a hadron + remnant string



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 function $f(z)$



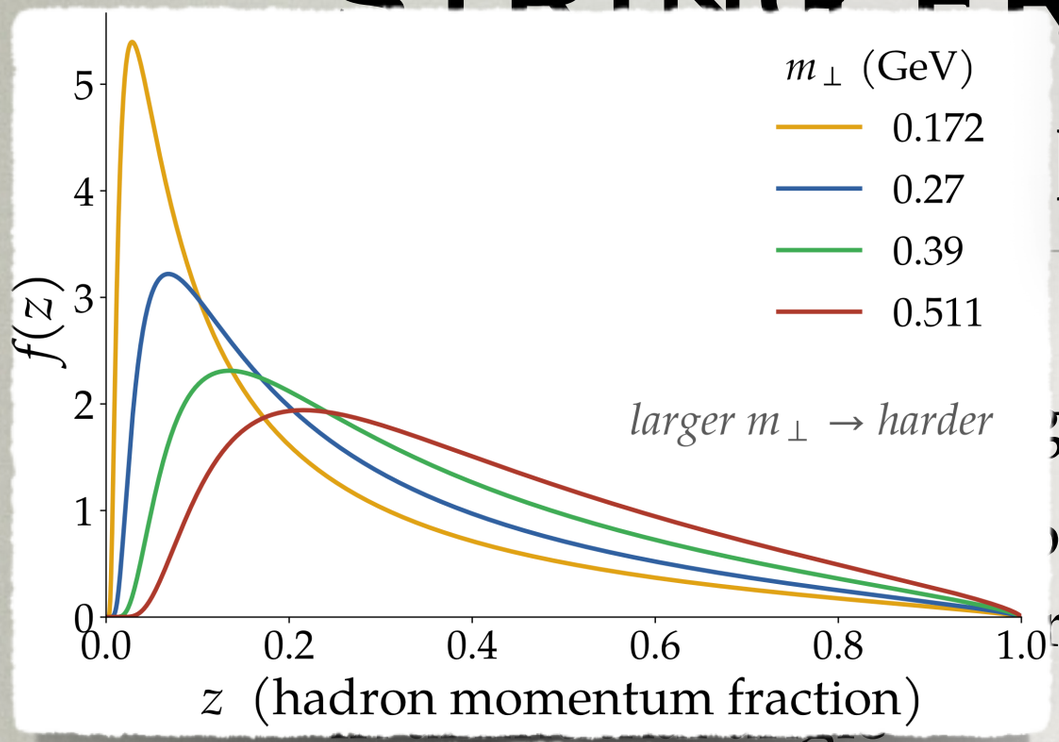
$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(-\frac{b\, m_\perp^2}{z}\right)$$

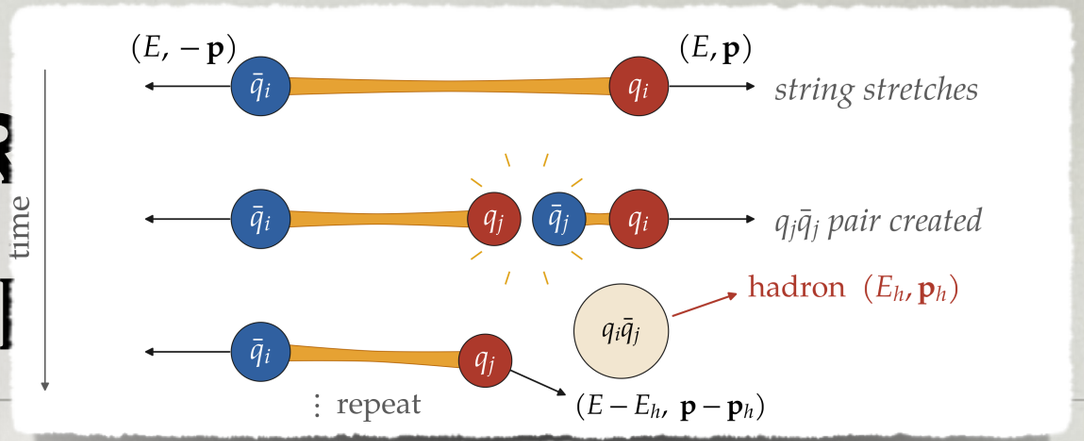
$$z = (p_{h,z} + E_h)/E_{\text{string}}$$

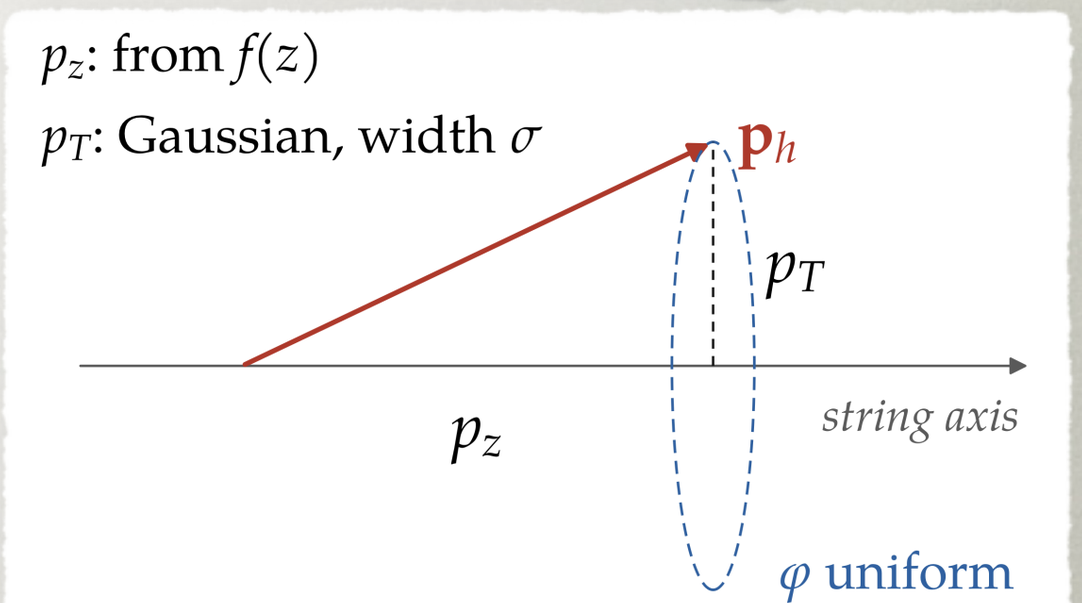
$$m_\perp^2 \equiv m^2 + p_T^2$$

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

STRING FR







- p_z distribution from Lund string fragmentation function $f(z)$

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

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

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

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

INVERSE PROBLEM FOR HADRONIZATION

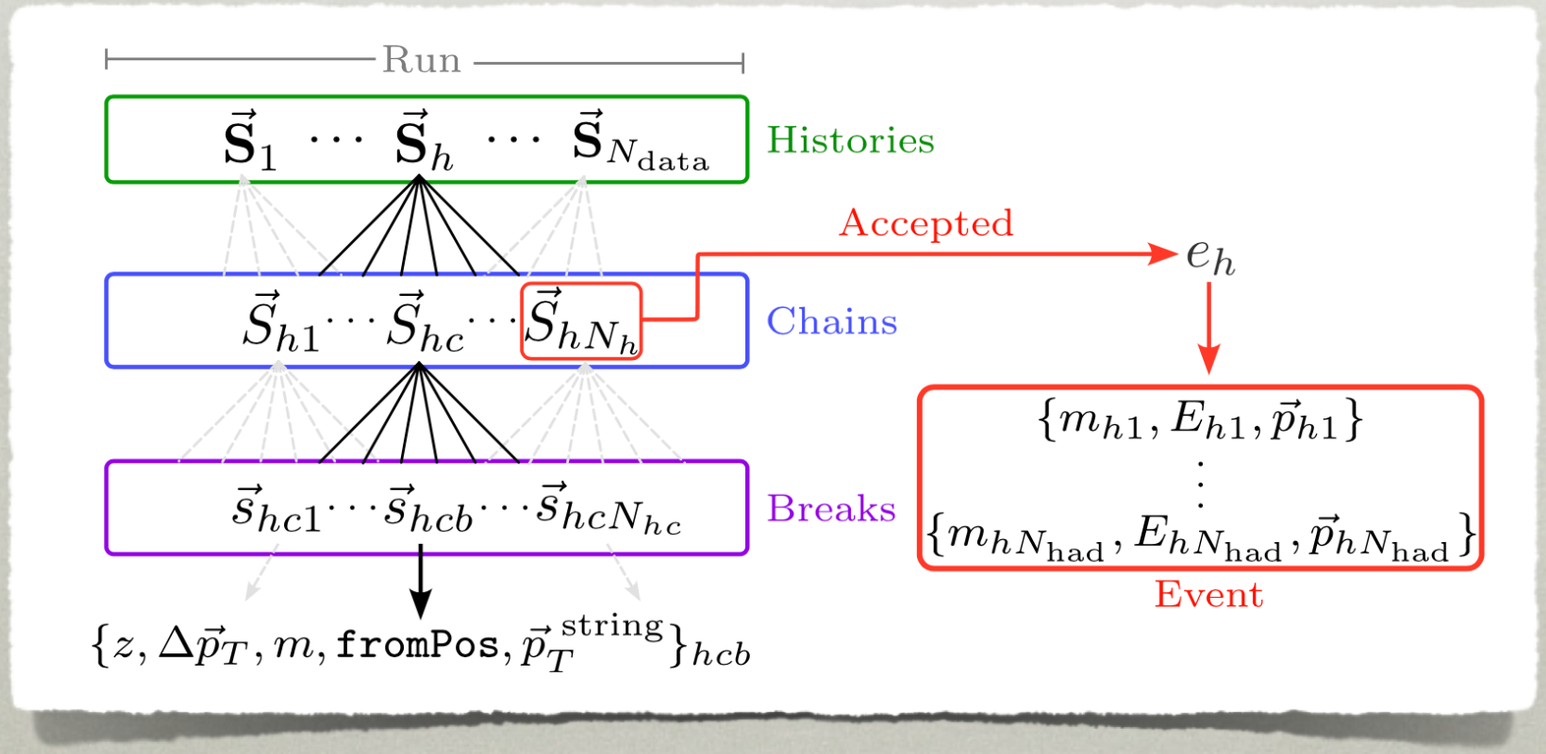
- inverse problem for hadronization
  - can one learn $f(z)$ from data*?
    - *without asking for its parametric form
- challenge: different events sample $f(z)$ a different amount of times

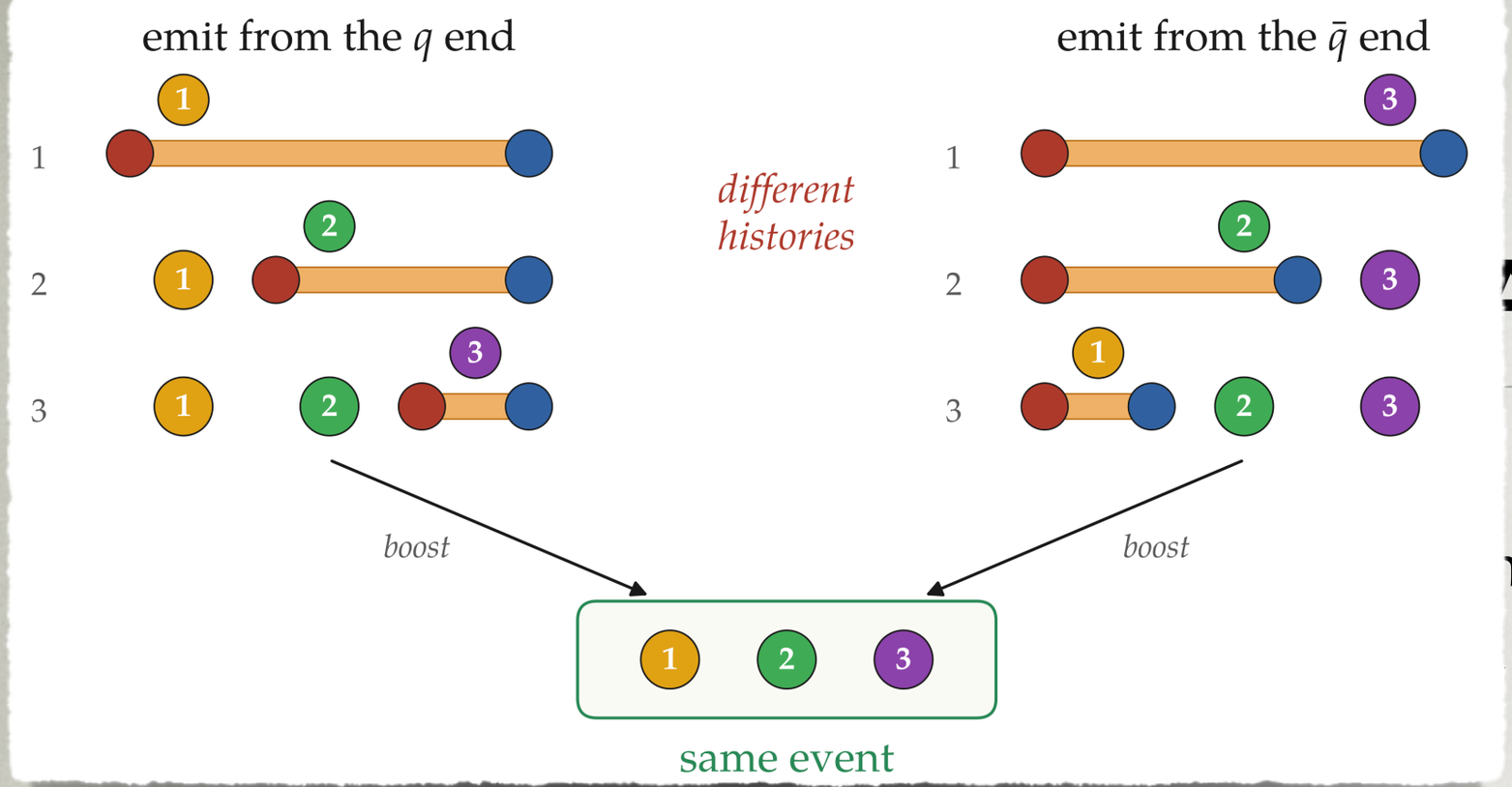
WHY IS THIS HARD?

- further complications

MLhad: 2410.06342, 2503.05667

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



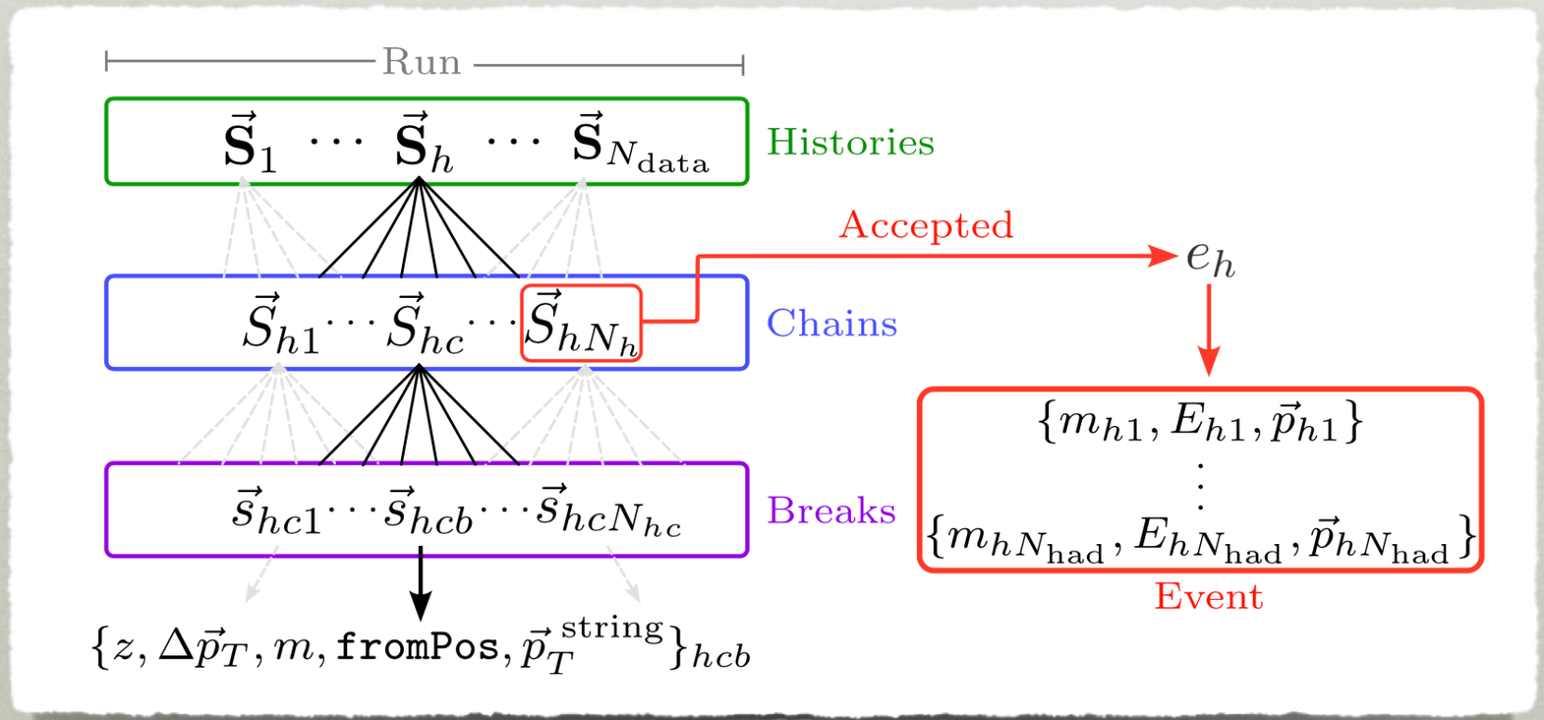


ARD?

MLhad: 2410.06342, 2503.05667

not lead to viable
fragmentation chains is

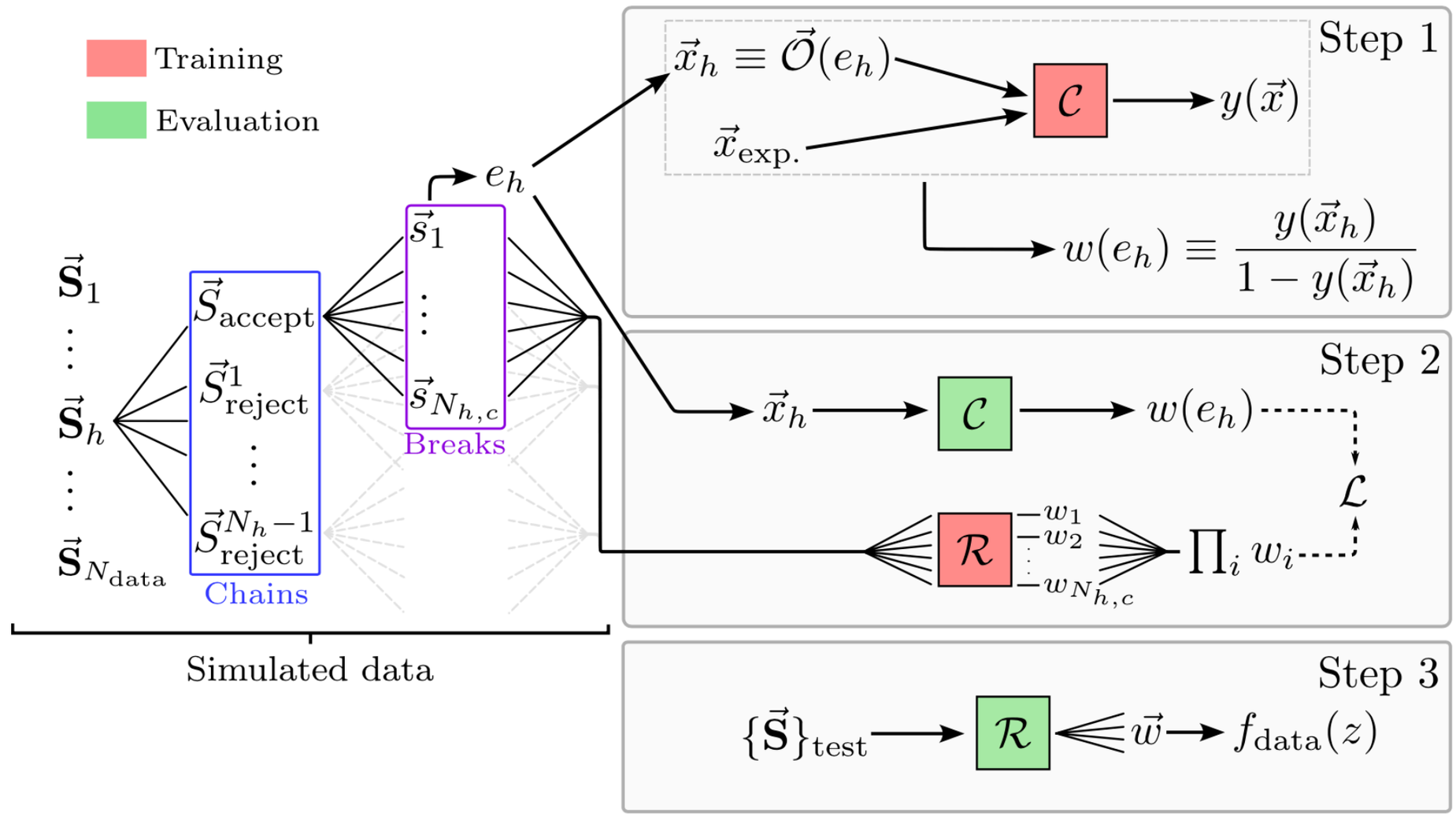
- a complicated permutation symmetry: swapping orders of emissions + appropriate boosts ⇒ lead to the same event



THE HOMER METHOD

MLhad: 2410.06342

- 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
- iterative procedures improve performance



2410.06342

net representation of $f(z)$

- Step 3: simulate using this new $f(z)$

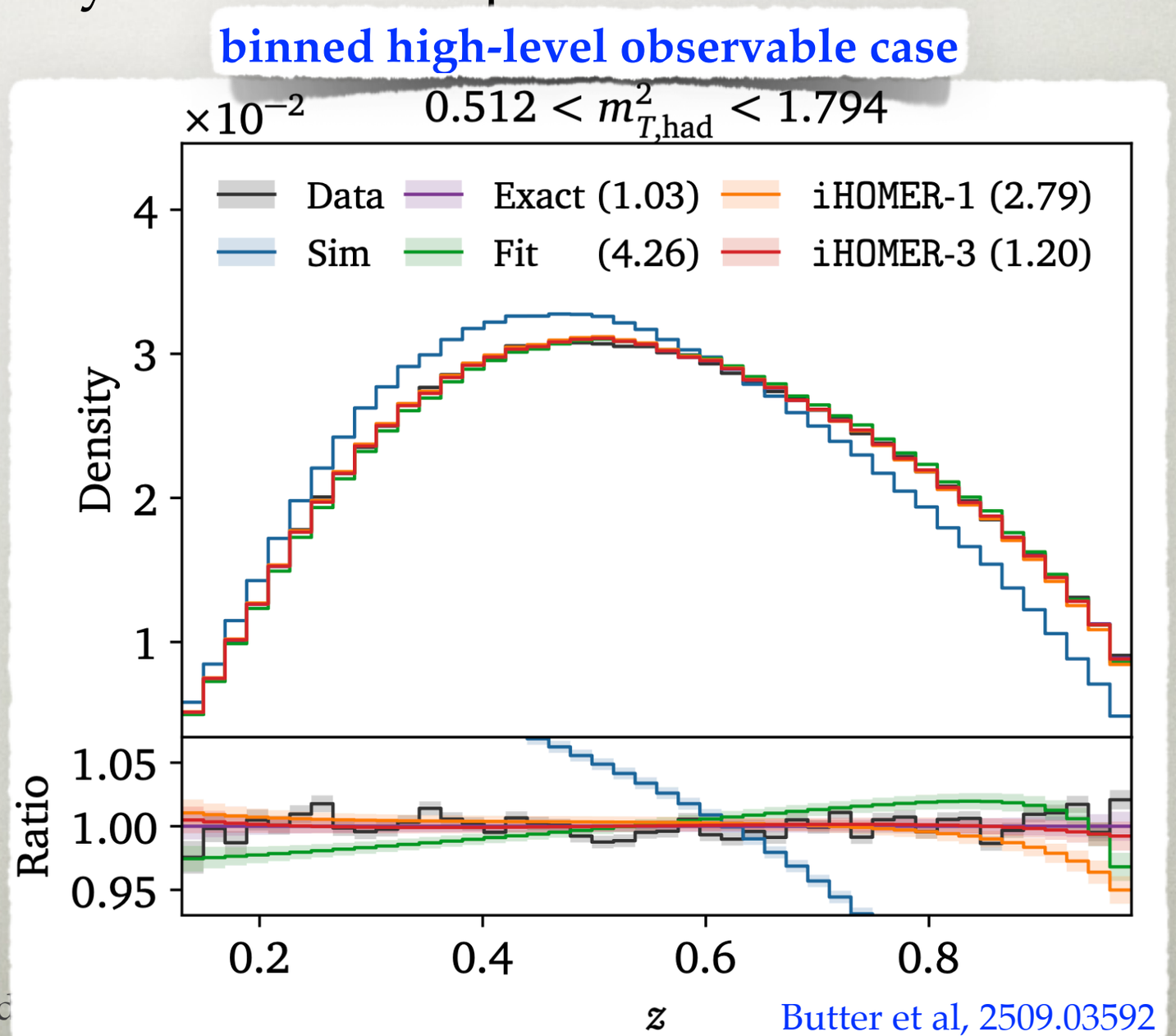
- in all steps use ratios of probabilities (weights)
  - always reweighting from base Pythia simulation results
- iterative procedures improve performance

MLhad: 2410.06342

MLhad: 2503.05667

SOLVING INVERSE PROBLEM WITH HOMER

- main message: inverse problem appears to be solvable
  - HOMER used only measurable quantities
  - recovered $f(z)$



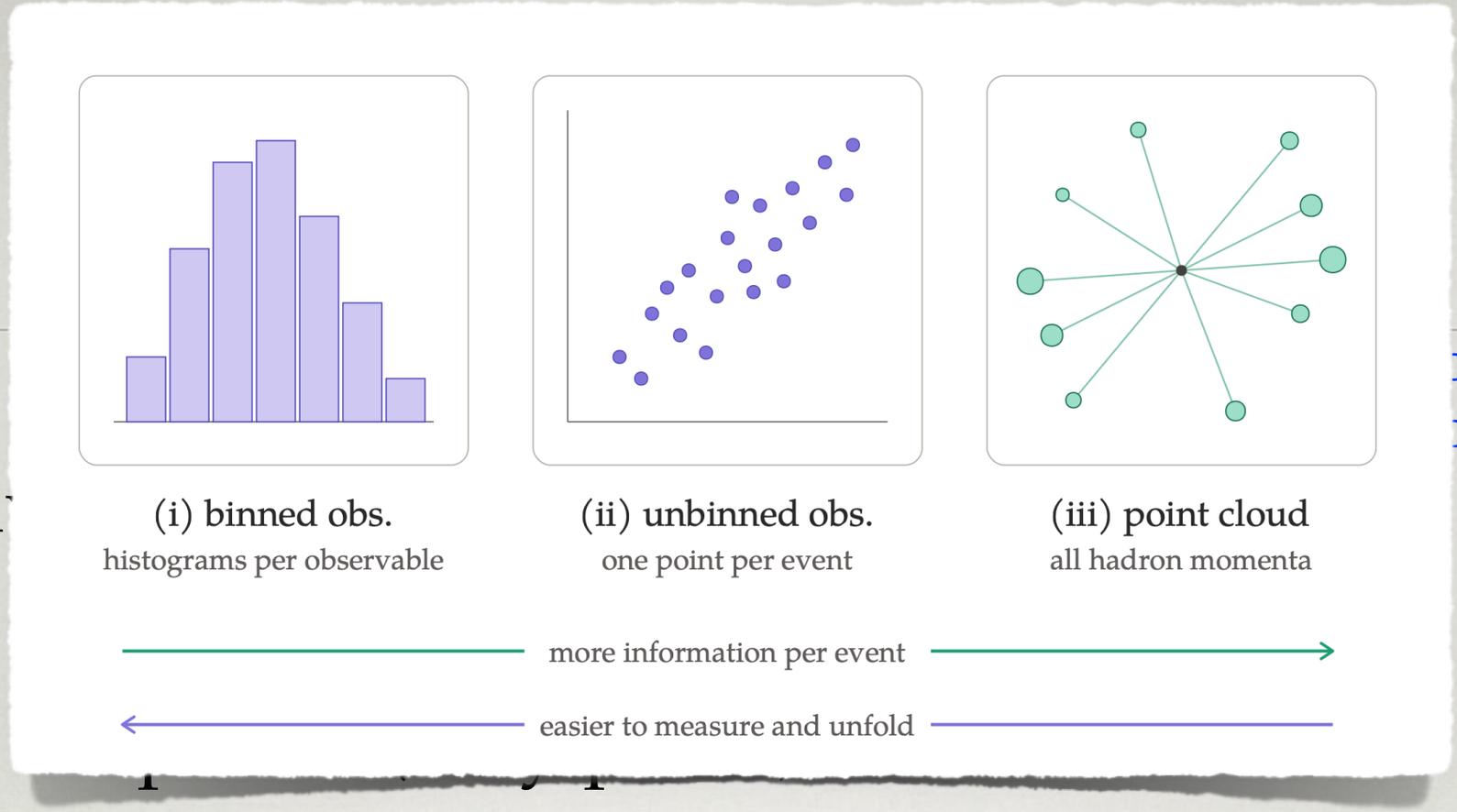
HOMER RESULTS

MLhad: 2410.06342

MLhad: 2503.05667

- caveats
  - Pythia hadronization model used was simplified (only pions)
  - fidelity of results depends on which observables (features) used
    - (i) binned high-level obs.;
    - (ii) unbinned high level obs.;
    - (iii) point cloud
- how to choose optimal set of observables?

- ca



MLhad: 2410.06342

MLhad: 2503.05667

  - fidelity of results depends on which observables (features) used
    - (i) binned high-level obs.;
    - (ii) unbinned high level obs.;
    - (iii) point cloud
- how to choose optimal set of observables?

WHICH OBSERVABLES?

WHICH OBSERVABLE

MLhad, 2602.01509

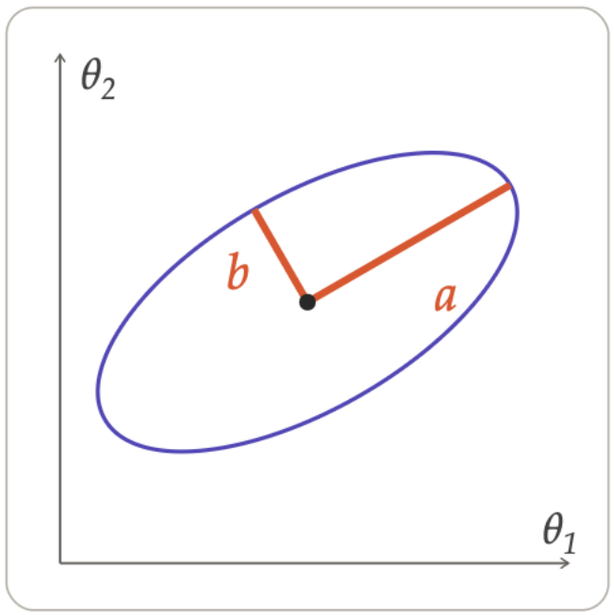
- which measurements are worth making?
  - a classical optimal experiment design problem
- already many criteria based on Fisher information matrix

Cramer-Rao bound

$$I_{ab}(\boldsymbol{\theta}) = \mathbb{E}_{p(\mathcal{O}|\boldsymbol{\theta})}\left[\frac{\partial \log p(\mathcal{O}|\boldsymbol{\theta})}{\partial \theta_a}\frac{\partial \log p(\mathcal{O}|\boldsymbol{\theta})}{\partial \theta_b}\right]$$

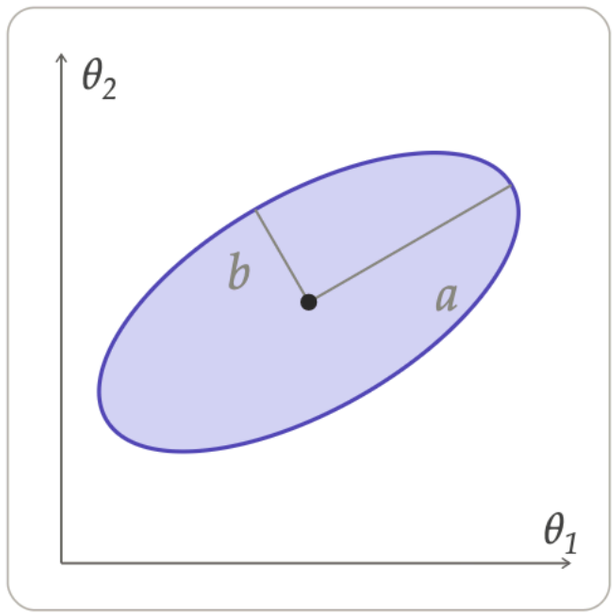
$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq I^{-1}(\boldsymbol{\theta})$$

  - minimize $\mathrm{Tr}(I^{-1})$ (A score)
  - minimize $\det(I^{-1})$ (D score)
  - minim. $\lambda_{\max}(I^{-1})$ (E score)
- why a new score?
  - note: all above need full information matrix
  - often hard to obtain, numerically expensive



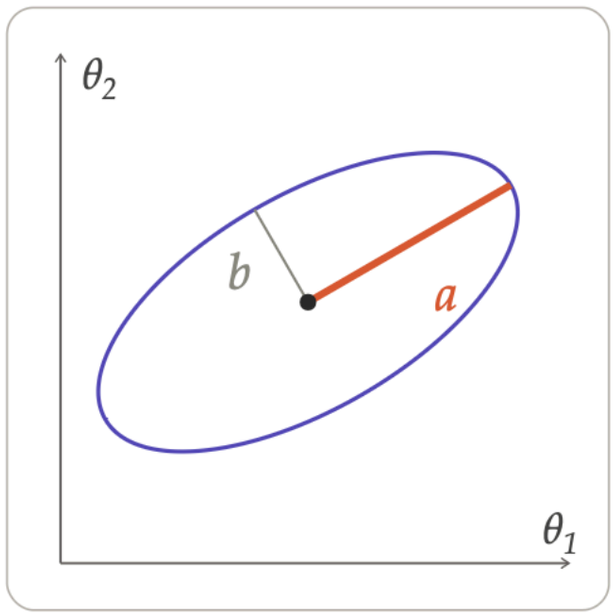
A score

$\mathrm{Tr}(I^{-1}) = a^2 + b^2$



D score

$\det(I^{-1}) = a^2 b^2 \propto (\text{area})^2$



E score

$\lambda_{\max}(I^{-1}) = a^2$

- minimize $\mathrm{Tr}(I^{-1})$ (A score)
- minimize $\det(I^{-1})$ (D score)
- minim. $\lambda_{\max}(I^{-1})$ (E score)

- why a new score?
  - note: all above need full information matrix
  - often hard to obtain, numerically expensive

HD-Sense score

MLhad, 2602.01509

- a heuristic alternative involving only single observable information matrices $I^{(i)}$

$$\mathcal{S}_{\rm HD}(\mathcal{X}) \;=\; {\rm Info}(\mathcal{X}) \left[1 - \beta\, \mathcal{P}_{\rm overlap}(\mathcal{X})\right]$$

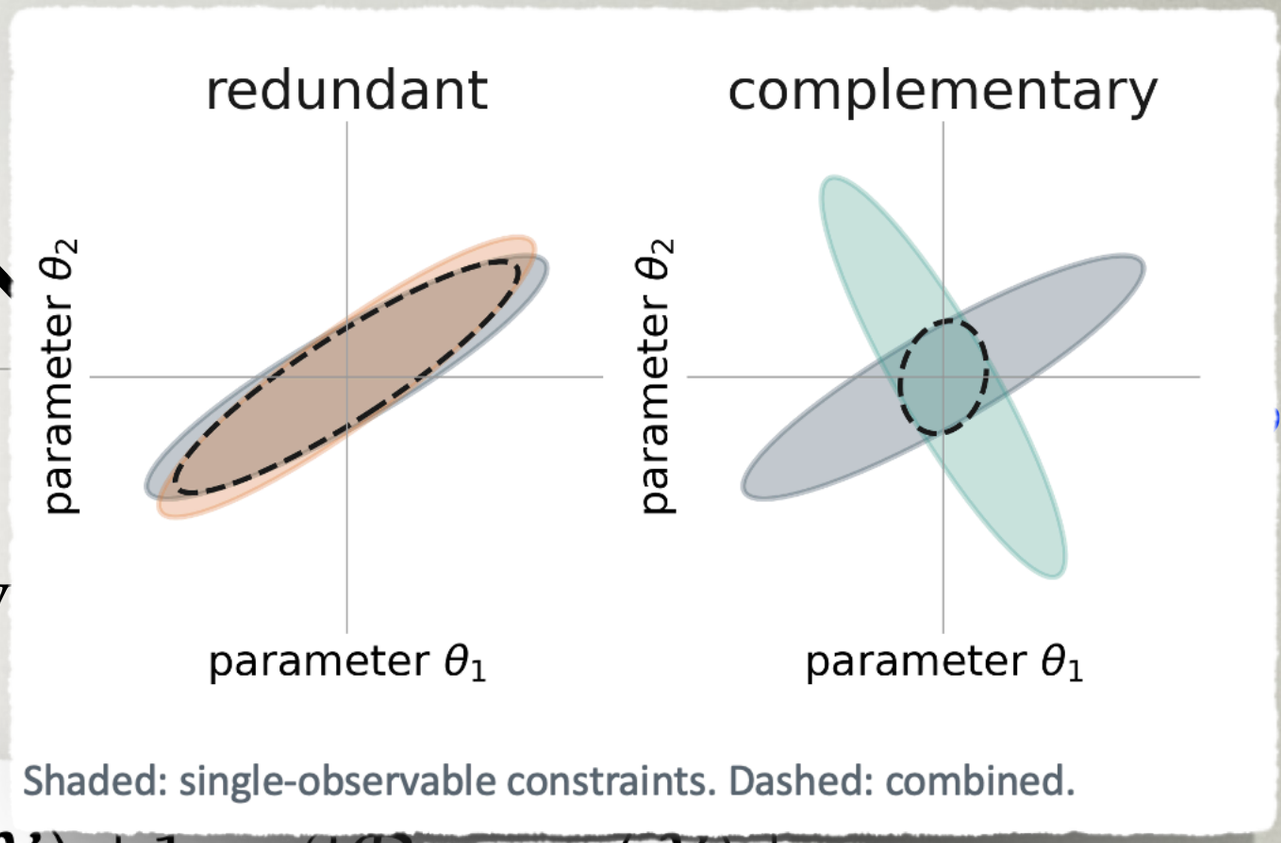
- Info: how strongly each observable alone constrains θ

$${\rm Info}(\mathcal{X}) = \sum_{i\in\mathcal{X}} {\rm Tr}\, I^{(i)}$$

- penalty for overlapping constraints
  - depends on hyper-parameter β
  - can use heuristic choice for it

HD-Sen

- a heuristic alternative inv
  information matrices $I^{(i)}$



Shaded: single-observable constraints. Dashed: combined.

$$\mathcal{S}_{\mathrm{HD}}(\mathcal{X}) = \mathrm{Info}(\mathcal{X}) \left[1 - \beta\, \mathcal{P}_{\mathrm{overlap}}(\mathcal{X})\right]$$

- Info: how strongly each observable alone constrains θ

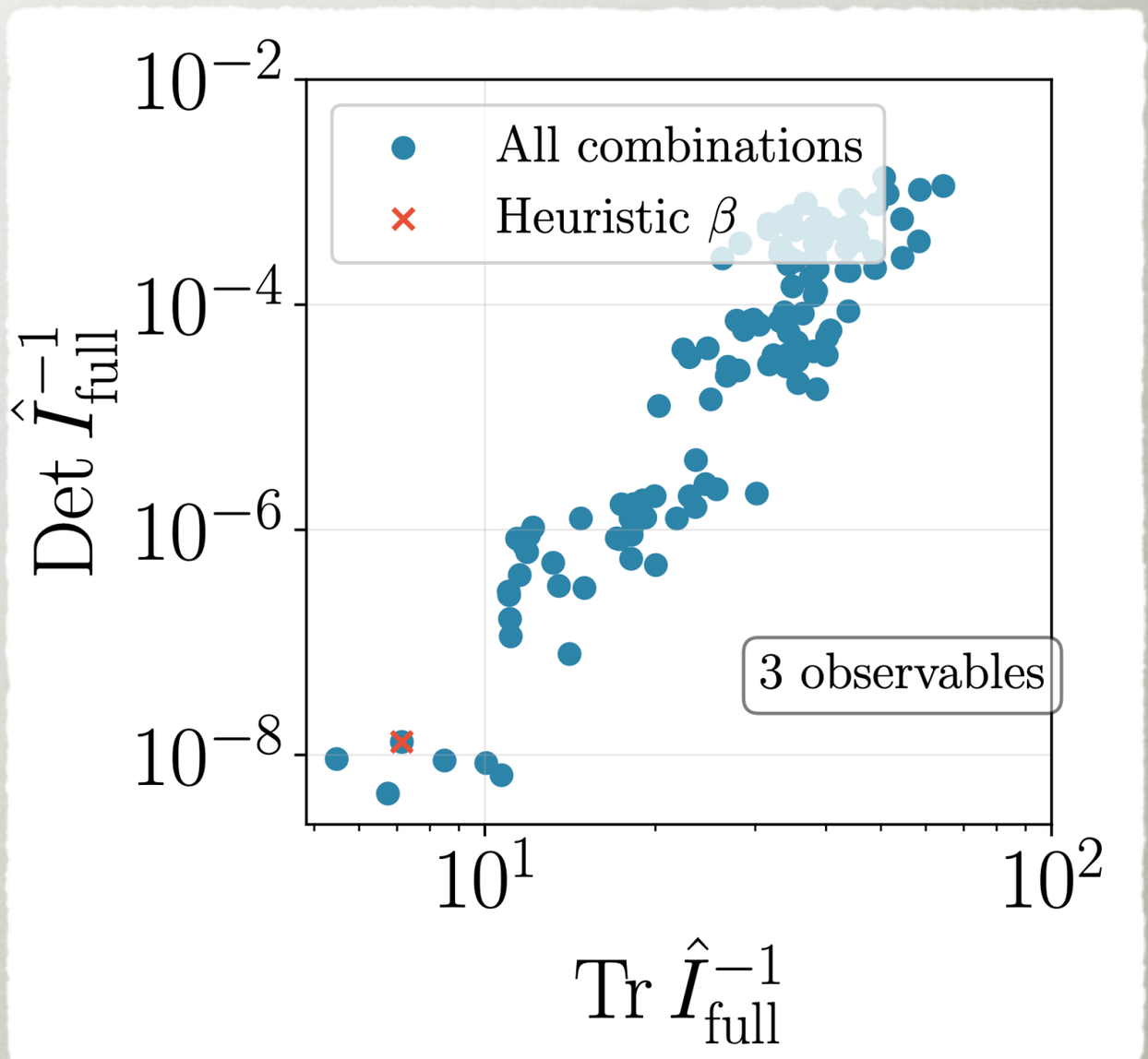
$$\mathrm{Info}(\mathcal{X}) = \sum_{i \in \mathcal{X}} \mathrm{Tr}\, I^{(i)}$$

- penalty for overlapping constraints
  - depends on hyper-parameter β
  - can use heuristic choice for it

HD-Sense score

MLhad, 2602.01509

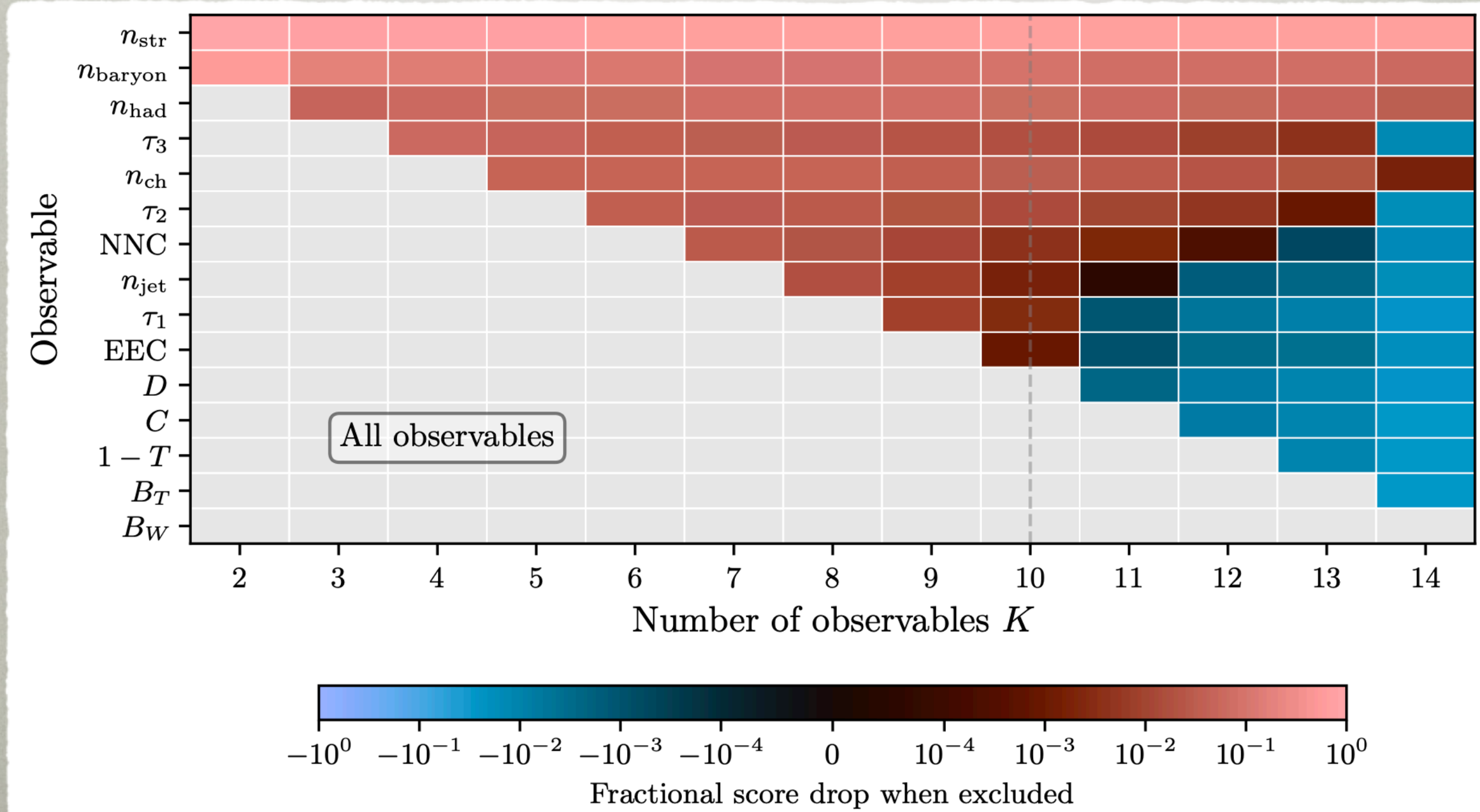
- choosing 3 "most constraining" observables for hadronization (out of 15)
- S_{HD} does as well as other more involved scores



MLhad, 2602.01509

HD-Sense score

- ranking of hadronization observables



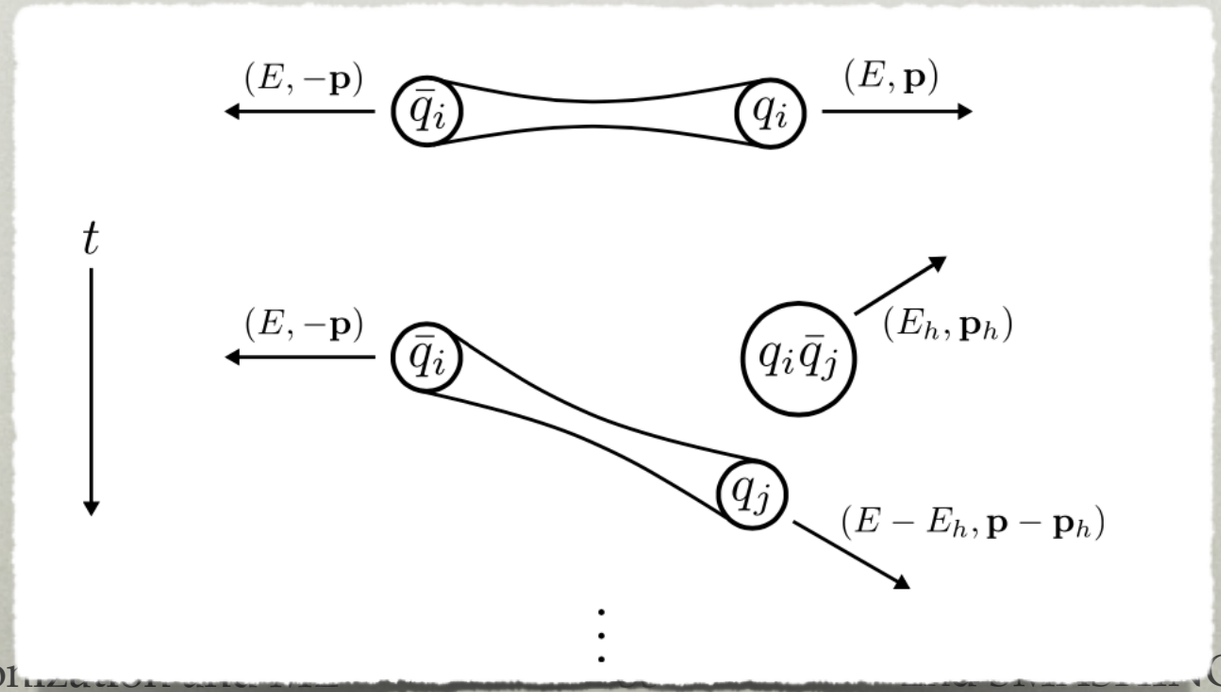
CONCLUSIONS

- hadronization an important systematic uncertainty for precision HEP
- covered part of the MLhad program
  - proof of principle for hadronization inverse problem: $f(z)$ via HOMER
  - HD-Sense score for optimal observable choice

BACKUP SLIDES

SIMPLIFIED STRING HADRONIZATION MODEL

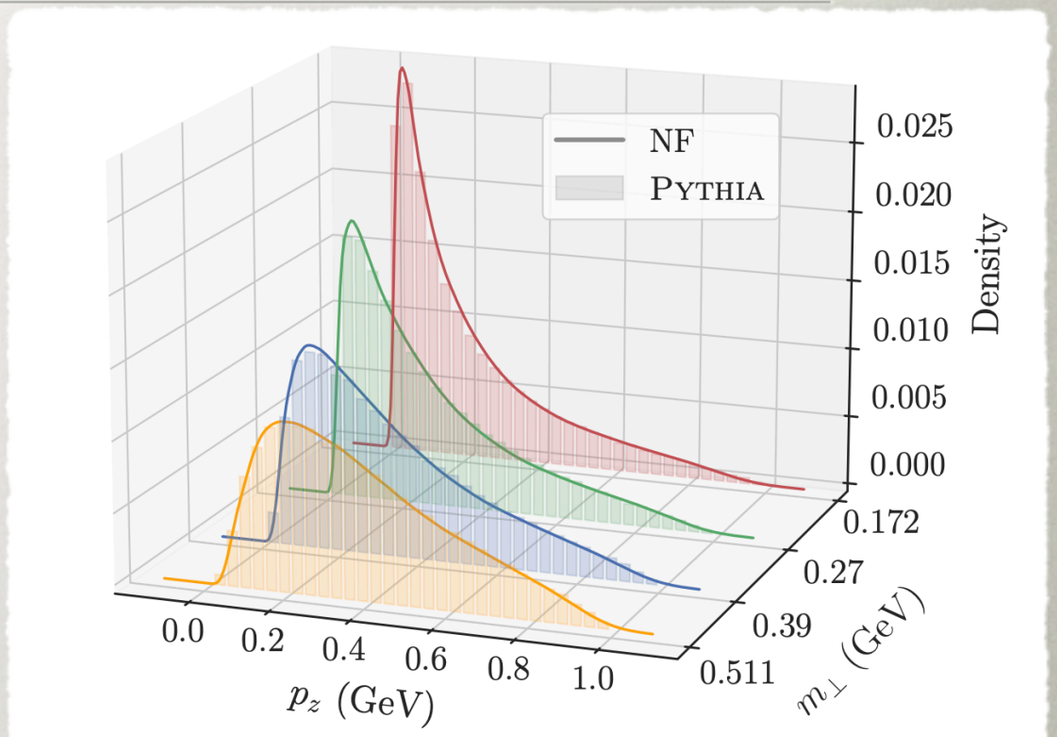
- assume that color flow done correctly by Pythia
- strings with $q, \bar{q}$ ends
  - hadron emission from a string piece controlled by fragmentation function $f(z)$
    - the whole hadronization chain is then reproduced by iterating
    - the string is labeled by $q, \bar{q}$ flavor and its energy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



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 function



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

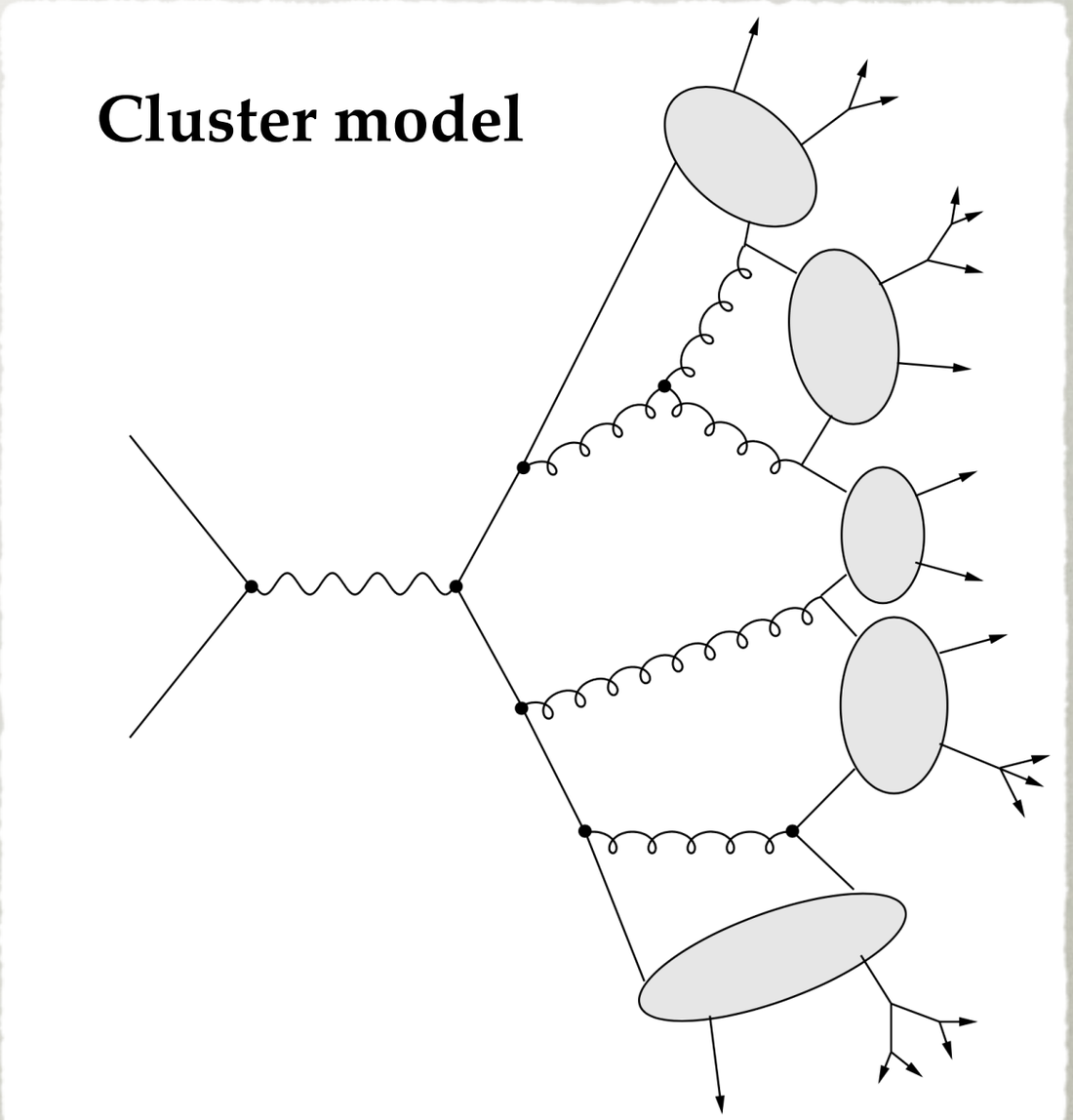
$$z = (p_{h,z} + E_h)/E_{\text{string}}$$

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

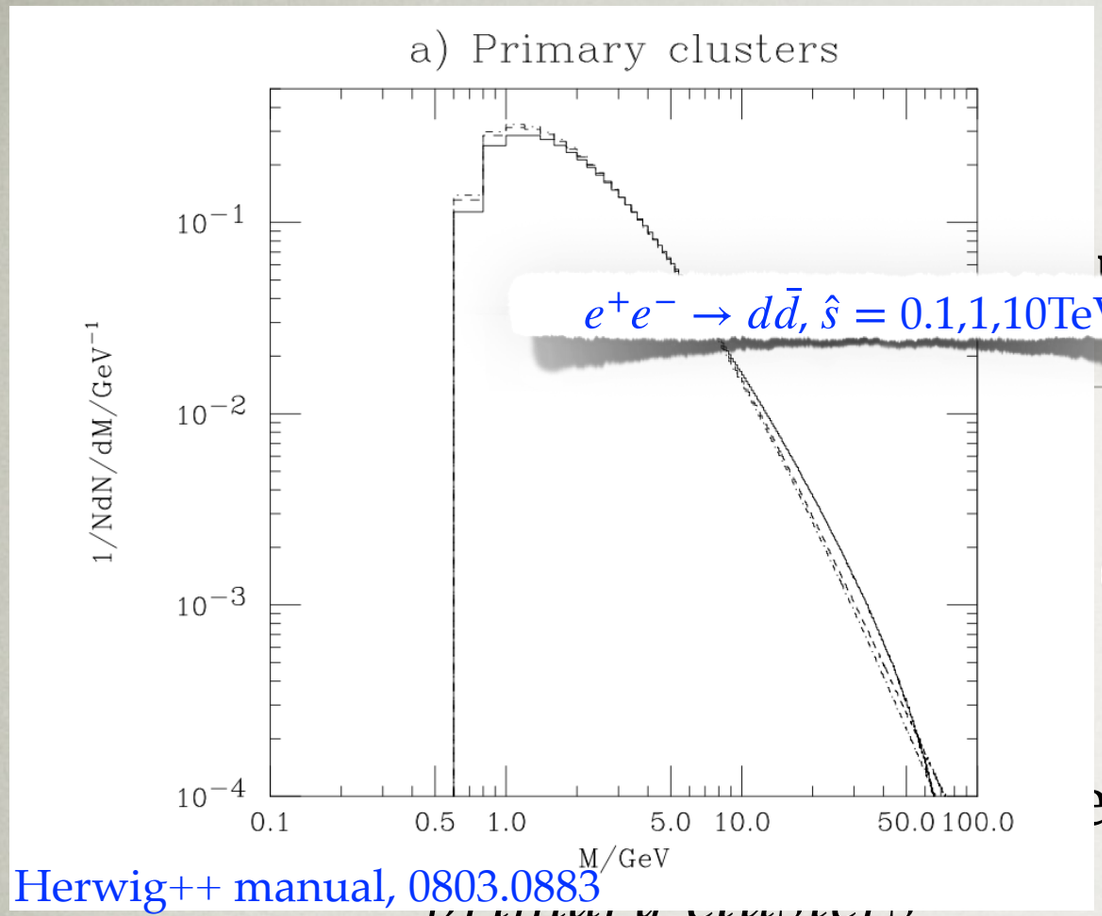
- for light quark flavored hadrons only three params.: a,b and mass, m
- p_T from random Gaussian distributions; width another param.

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, $\mathcal{O}(30)$



TER MODEL



Herwig++ manual, 0803.0883

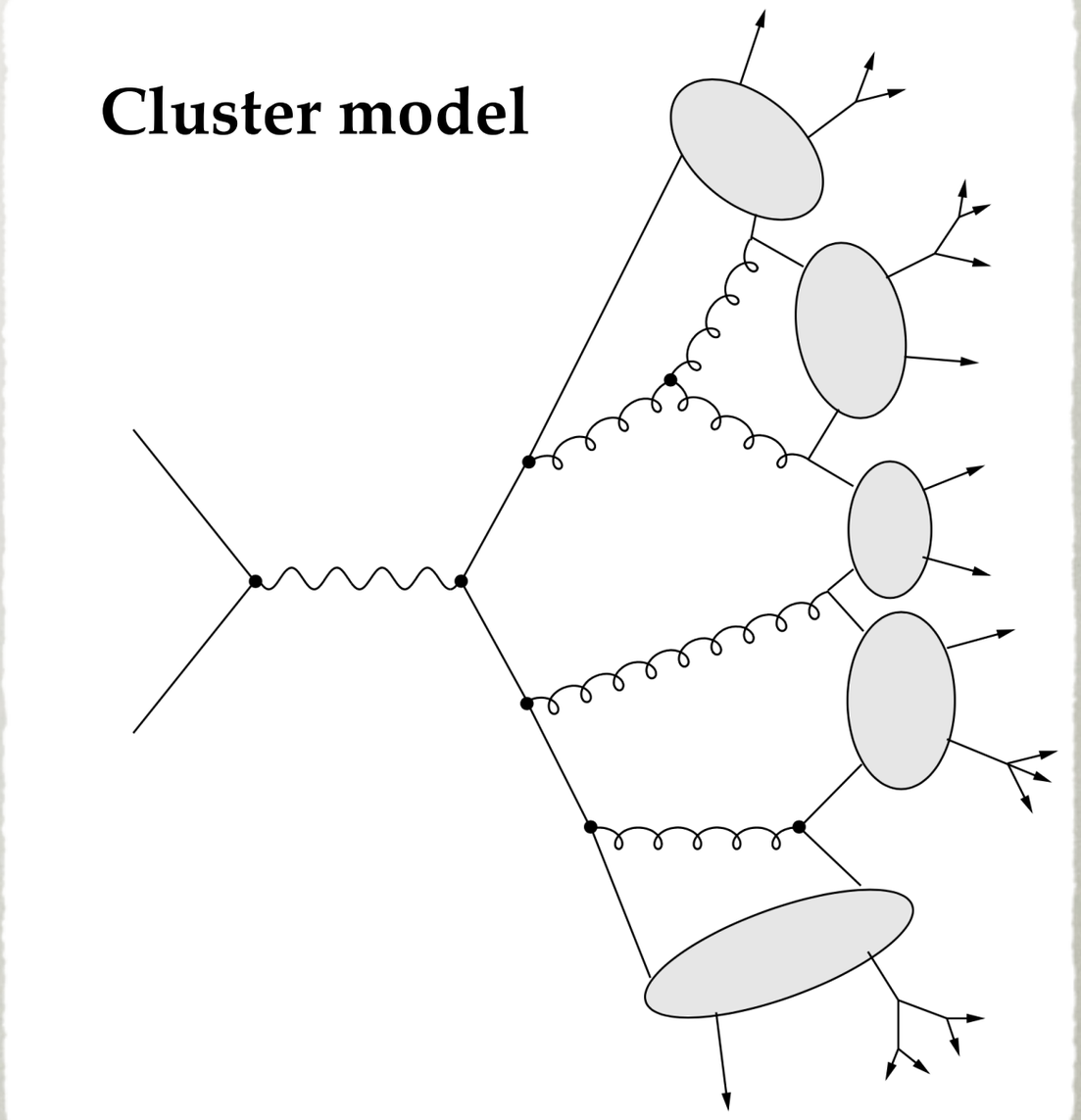
(in

ets:

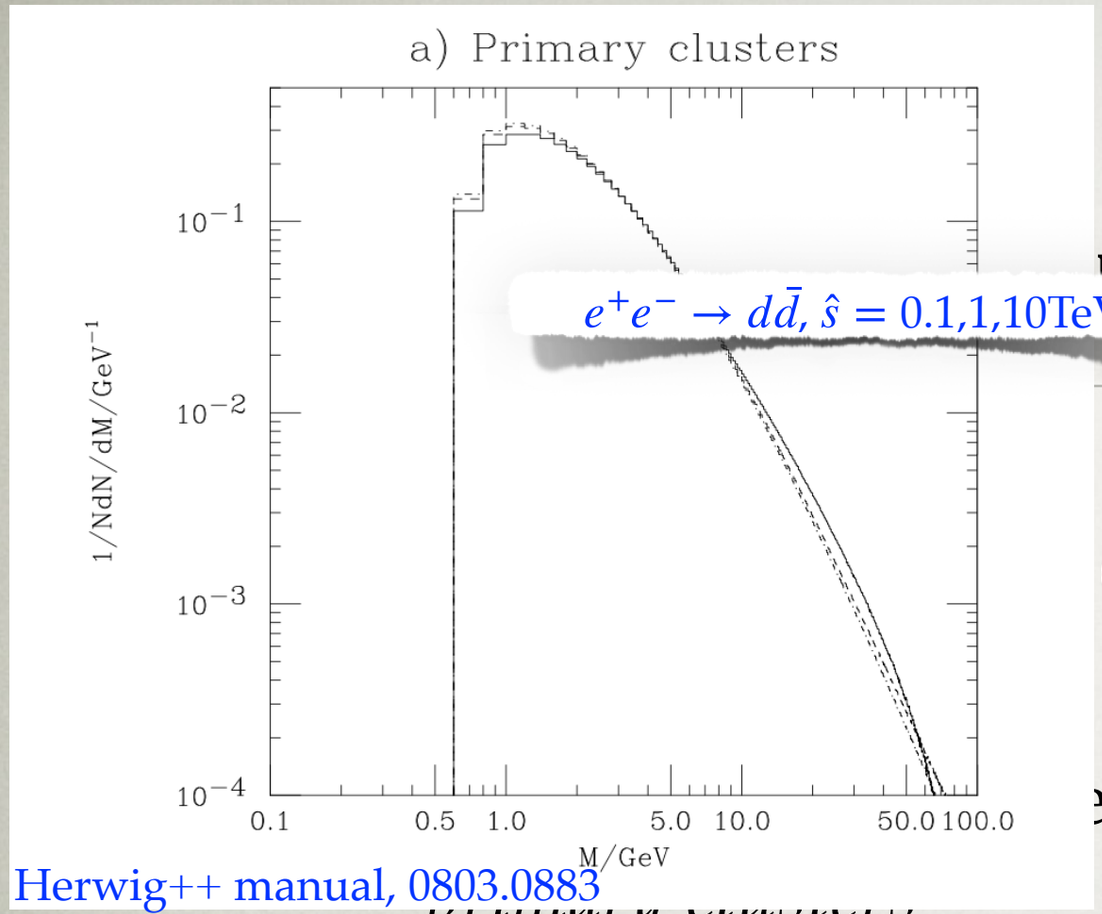
primary clusters

- primary clusters have universal mass distrib

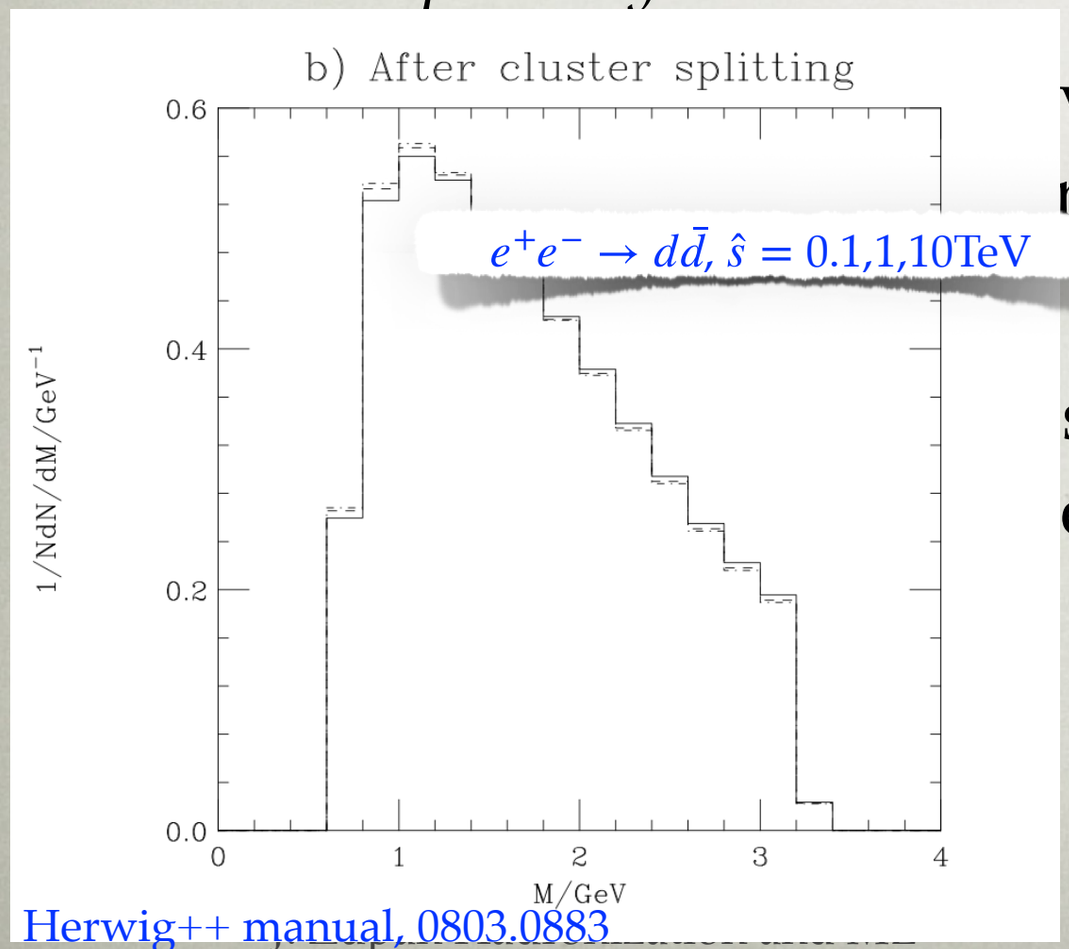
- heavier clusters are decayed to lighter ones (fission/model dep. step)
- relatively small set of params, $\mathcal{O}(30)$



TER MODEL



Herwig++ manual, 0803.0883



Herwig++ manual, 0803.0883

(in

ets:

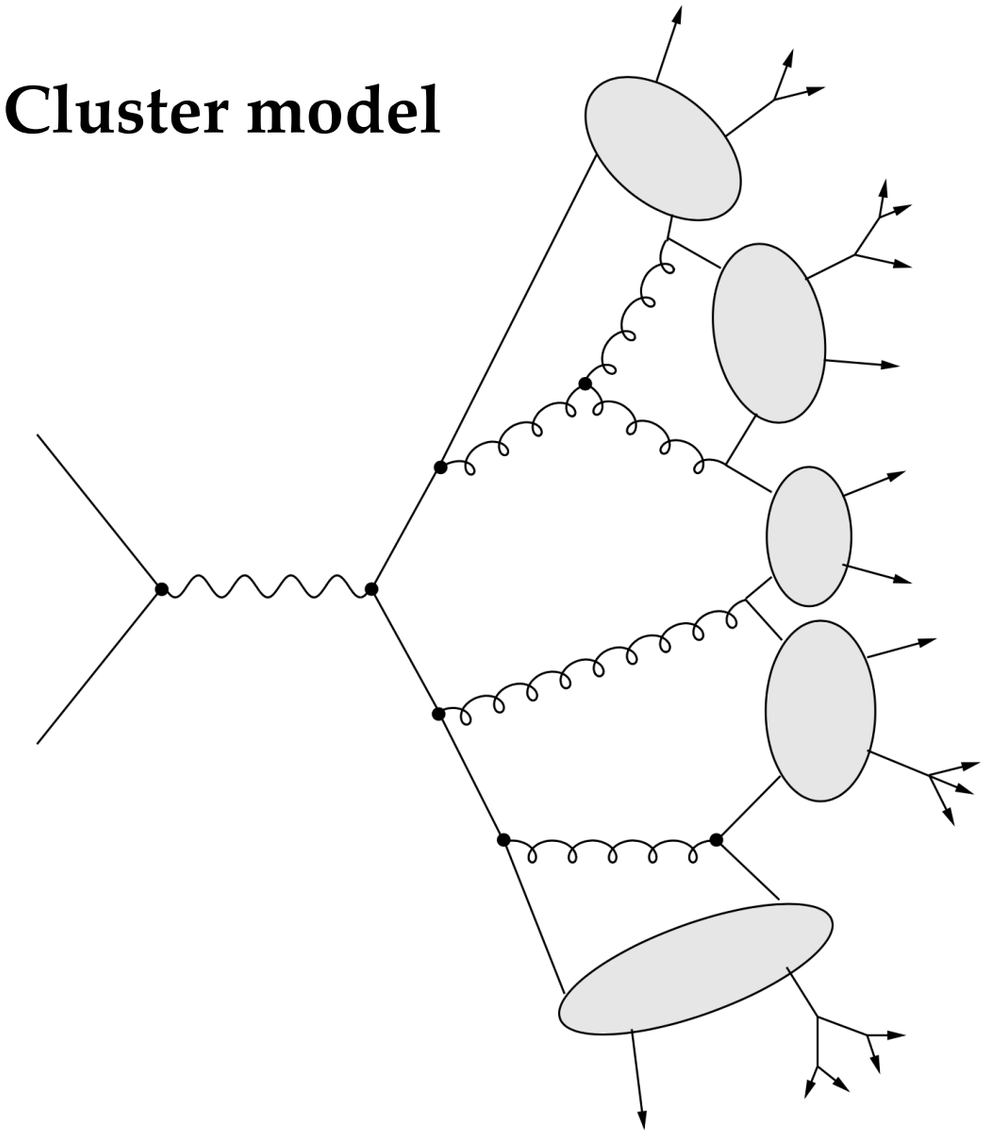
ve

rib

s

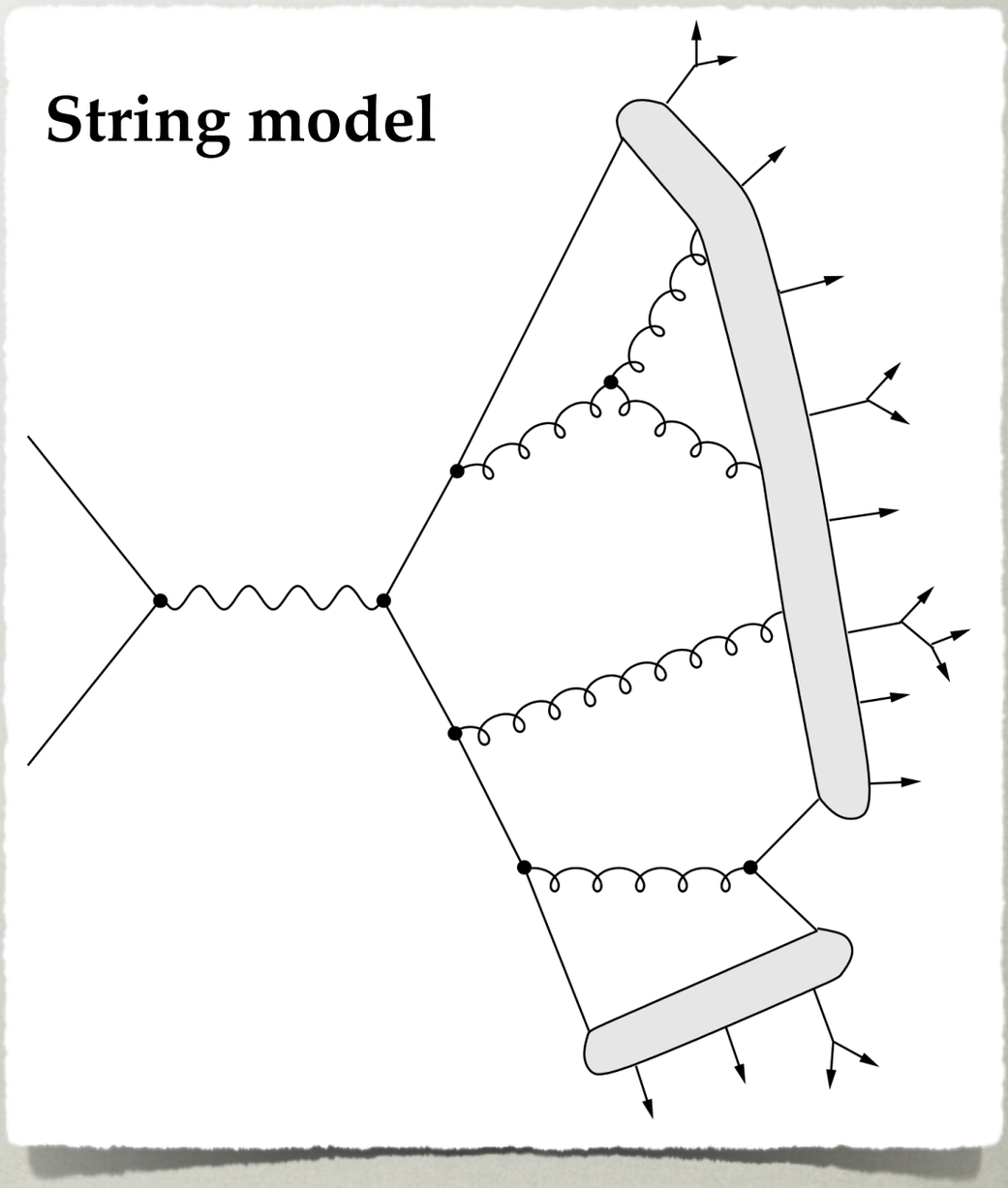
ep)

Cluster model



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, color reconnection, etc



ML hadronization: MLhad roadmap

- a series of progressive steps to be done before practically useful in Pythia/MC simulations
  - ML architecture that mimics a simplified Lund string hadronization model — MLhad, 2203.04983
    - train ML on truth level Pythia output (not obs. in exp)
  - develop a framework to propagate errors — MLhad, 2311.09296
  - train on mock data (i.e. just observable information) — MLhad, 2410.06342, 2503.05667
  - improved ML architecture with full hadron flavor selector — Butter et al, 2509.03592
  - train on real data (i.e. just already measured information) — **we are here**
  - replace/supplement Pythia string model, including color reconnections

A HARDER PROBLEM

- PDFs: appear at most in quadratic form in observables

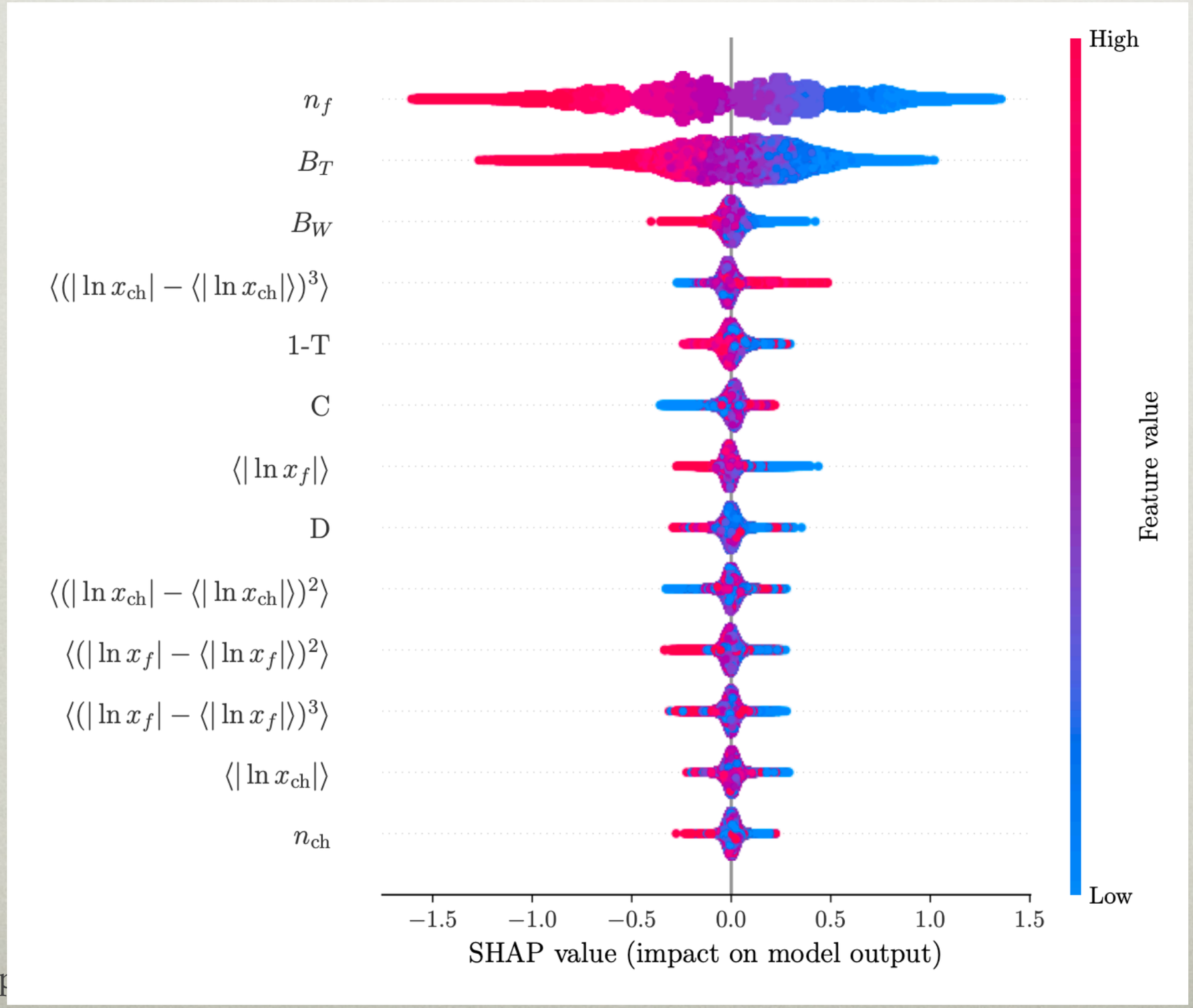
$$\sigma_X(s, M_X^2) = \sum_{a,b} \int_{x_{\min}}^{1} dx_1\, dx_2\, f_{a/h_1}(x_1, M_X^2) f_{b/h_2}(x_2, M_X^2) \hat{\sigma}_{ab\to X}\left(x_1 x_2 s, M_X^2\right).$$

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

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

MLhad: 2410.06342

WHICH OBSERVABLES?



J. Zup

HOMER RESULTS

MLhad: 2410.06342

- HOMER uses only measurable quantities
- for a simplified case, $q\bar{q}$ string of fixed energy: three versions
  - (i) binned high-level obs.;
  - (ii) unbinned high level obs.;
  - (iii) point cloud
- including gluons degrades performance
  - extraction of $f(z)$ still possible

MLhad: 2503.05667

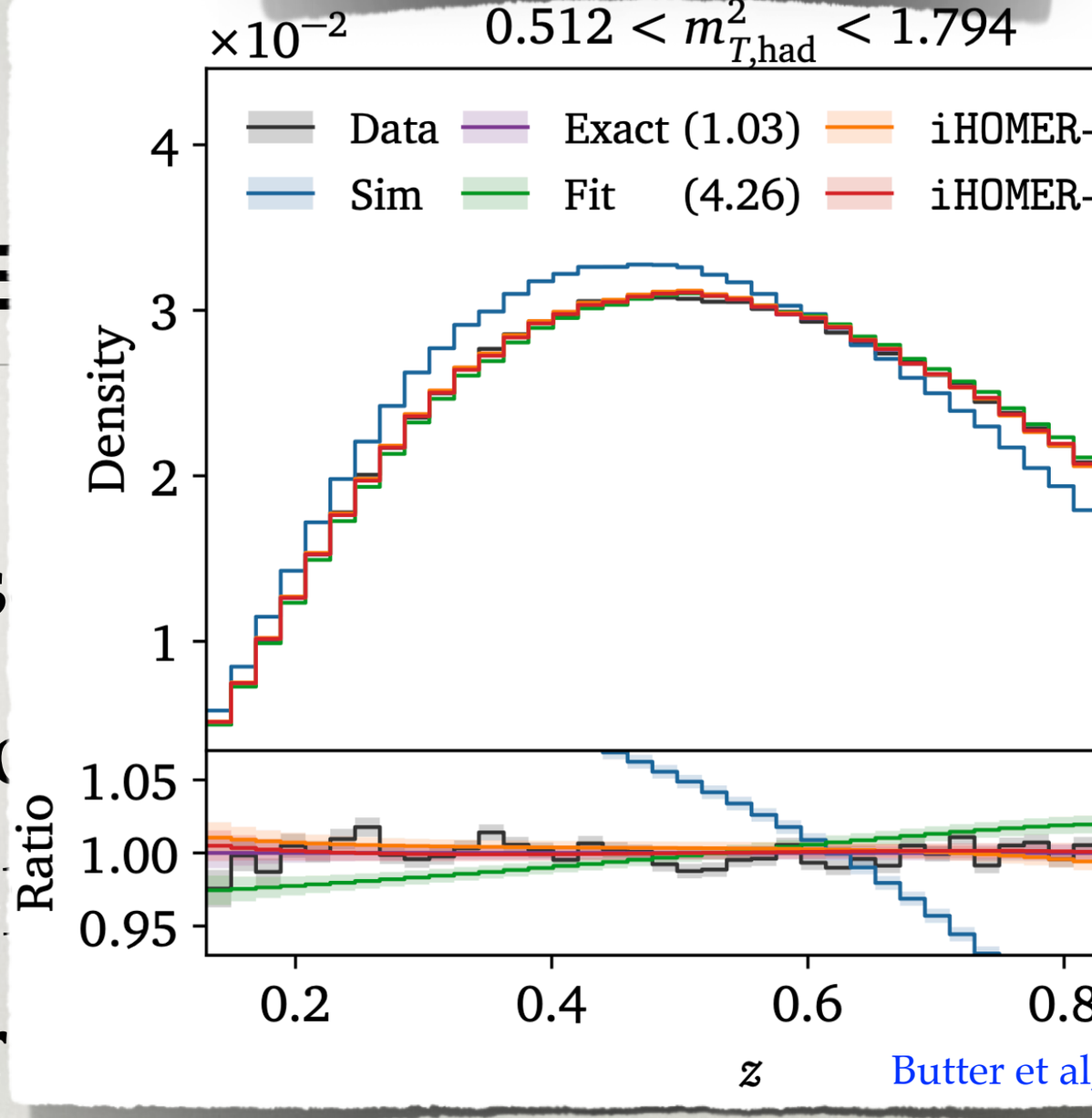
HOMER RESULTS

MLhad: 2410.06342

- 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
- iterative training improves performance

HOMER RE

- in this simplified cas
  - binned high-level (
    (multiplicities, sha
  - additional gain for
    level obs. case
  - point cloud harder to train on
- iterative training improves performance

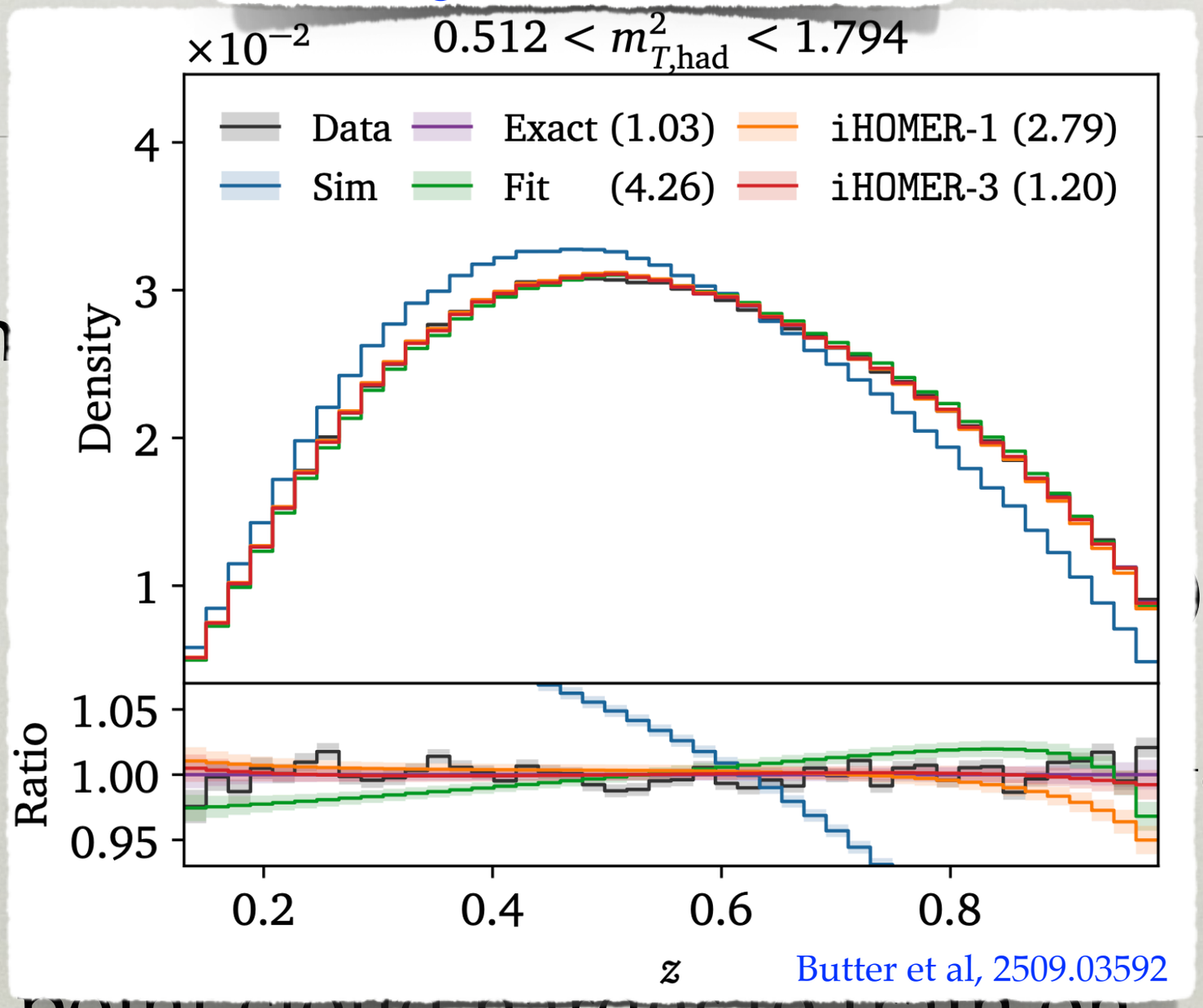


HOMER RESULTS

MLhad: 2410.06342

- 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
- iterative training improves performance

binned high-level observable case



Butter et al, 2509.03592

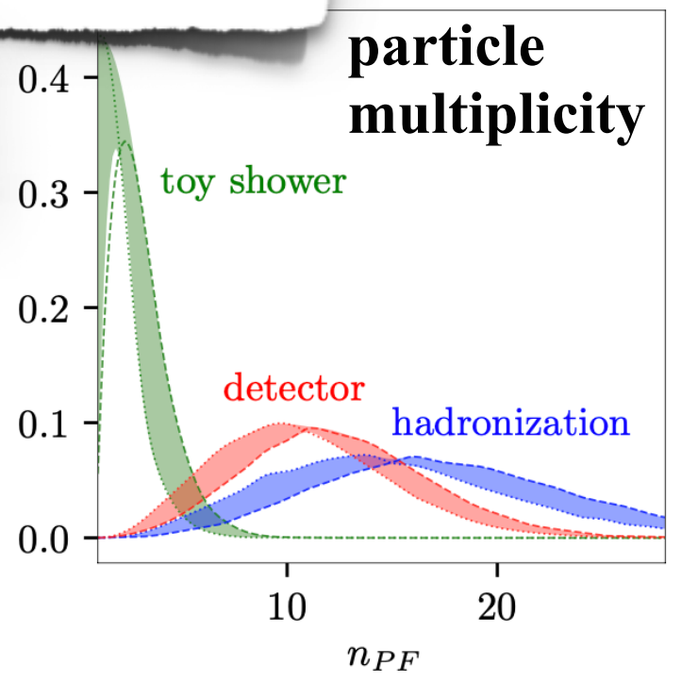
MLhad: 2410.06342

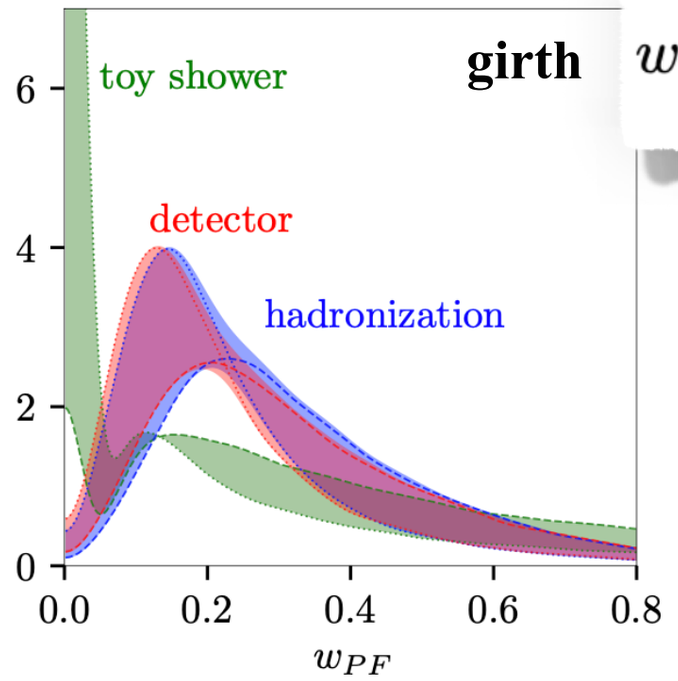
- in
  - suffice
  - gh-
  - point cloud harder to train on
- iterative training improves performance

SHOULD ONE CARE ABOUT HADRONIZATION

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

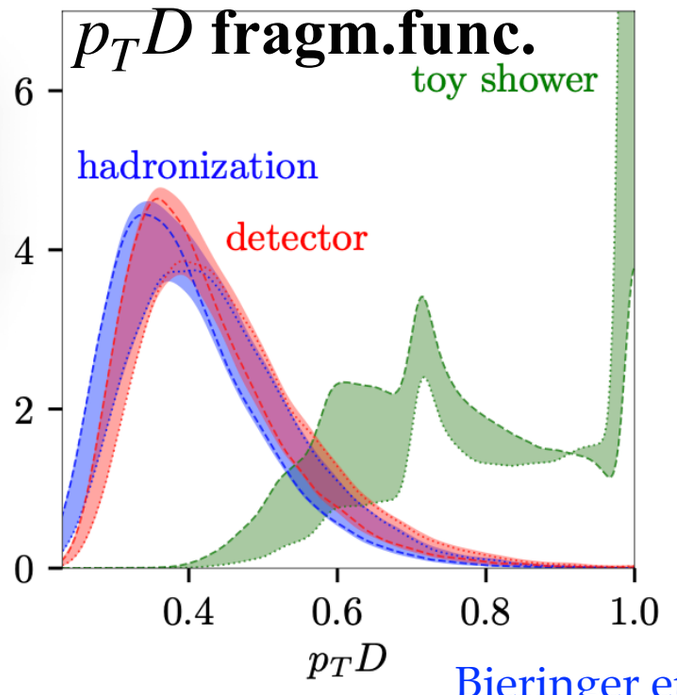
$$n_{\mathrm{PF}} = \sum_i 1$$

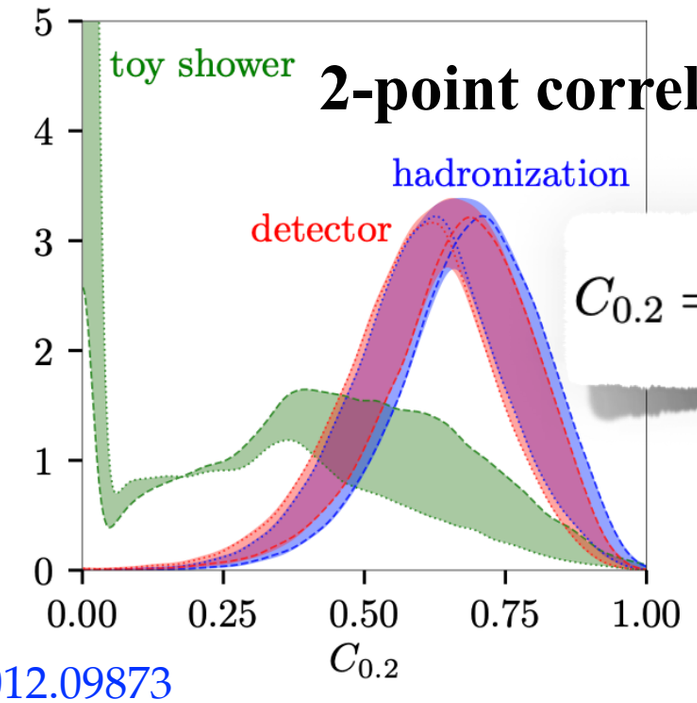




$$w_{\mathrm{PF}} = \frac{\sum_i p_{T,i} \Delta R_{i,\mathrm{jet}}}{\sum_i p_{T,i}}$$

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





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

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*
172.56 ± 0.31	**OUR AVERAGE**	Error includes scale factor of 1.6. See the ideogram below.		
172.52 ±0.14 ±0.30	[1] HAYRAPETYAN	2024F	LHC	ATLAS+CMS combined
174.41 ±0.39 ±0.71	[2] AAD	2023N	ATLS	leptonic invariant mass in ℓ+jets channel
171.77 ±0.37	[3] TUMASYAN	2023BB	CMS	ℓ + $\geq$ 4j (2b)
173.06 ±0.24 ±0.80	[4] TUMASYAN	2023Z	CMS	boosted top; ℓ+jets channel
$172.13\ ^{+0.76}_{-0.77}$	[5] TUMASYAN	2021G	CMS	t-channel single top production
172.6 ±2.5	[6] SIRUNYAN	2020AR	CMS	jet mass from boosted top
172.34 ±0.20 ±0.70	[7] SIRUNYAN	2019AP	CMS	$\geq$ 6 jets ($\geq$ 2b)
$172.33 \pm 0.14\ ^{+0.66}_{-0.72}$	[8] SIRUNYAN	2019AR	CMS	dilepton channel ($e\mu$, 2e, 2μ)
174.30 ±0.35 ±0.54	[9] TEVEWWG	2016	TEVA	Tevatron combination

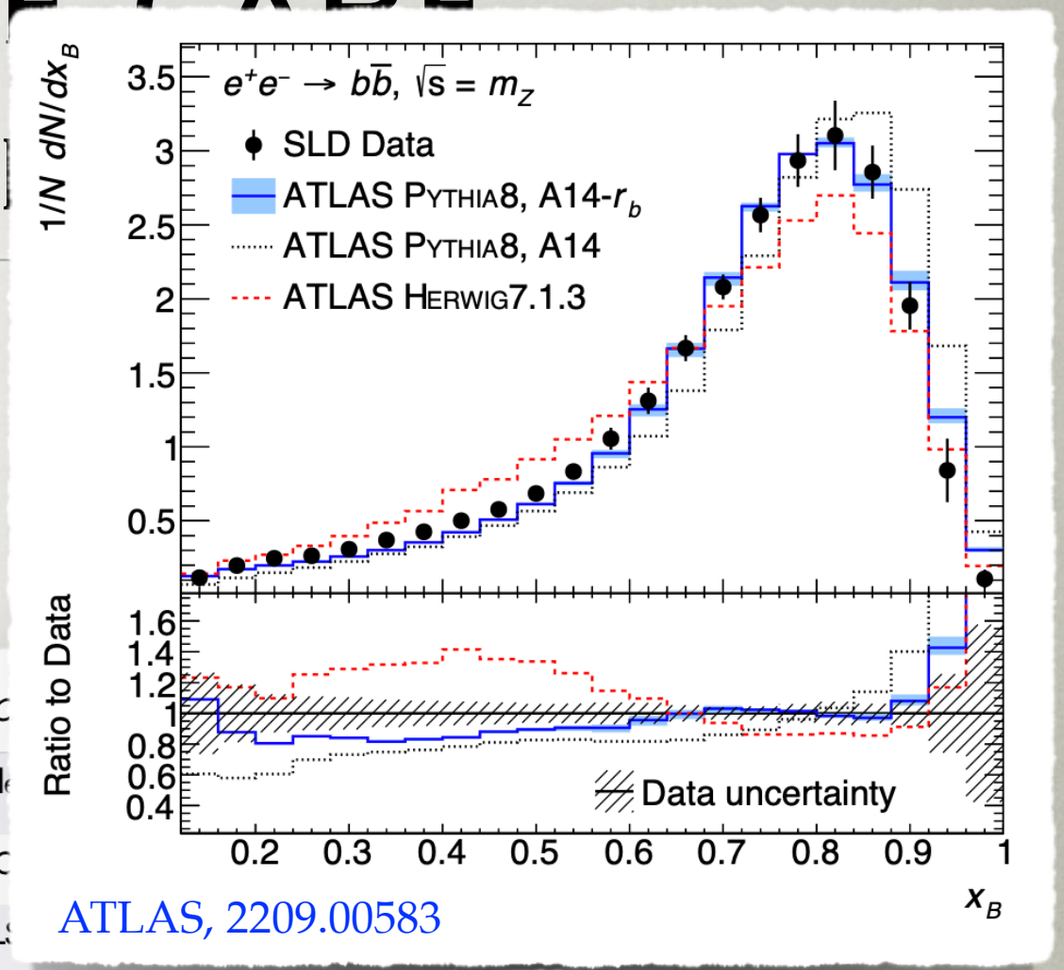
- b quark hadronization model uncert. on indiv. measurements ~0.1-0.2 GeV

SHOULD ONE CARE ABOUT HADRO

- understanding hadronization for precision measurements
- top mass:

pdglive.lbl.gov

VALUE (GeV)	DOCUMENT ID			TEC	
172.56 ± 0.31	**OUR AVERAGE**	Error includes scale factor of 1.6. See the ide			
$172.52 \pm 0.14 \pm 0.30$	[1] HAYRAPETYAN	2024F	LHC		
$174.41 \pm 0.39 \pm 0.71$	[2] AAD	2023N	ATL		
171.77 ± 0.37	[3] TUMASYAN	2023BB	CMS	ℓ + $\geq$ 4j (2b)	
$173.06 \pm 0.24 \pm 0.80$	[4] TUMASYAN	2023Z	CMS	boosted top; ℓ+jets channel	
$172.13\ ^{+0.76}_{-0.77}$	[5] TUMASYAN	2021G	CMS	t-channel single top production	
172.6 ± 2.5	[6] SIRUNYAN	2020AR	CMS	jet mass from boosted top	
$172.34 \pm 0.20 \pm 0.70$	[7] SIRUNYAN	2019AP	CMS	$\geq$ 6 jets ($\geq 2b$)	
$172.33 \pm 0.14\ ^{+0.66}_{-0.72}$	[8] SIRUNYAN	2019AR	CMS	dilepton channel ($e\mu$, 2e, 2μ)	
$174.30 \pm 0.35 \pm 0.54$	[9] TEVEWWG	2016	TEVA	Tevatron combination	

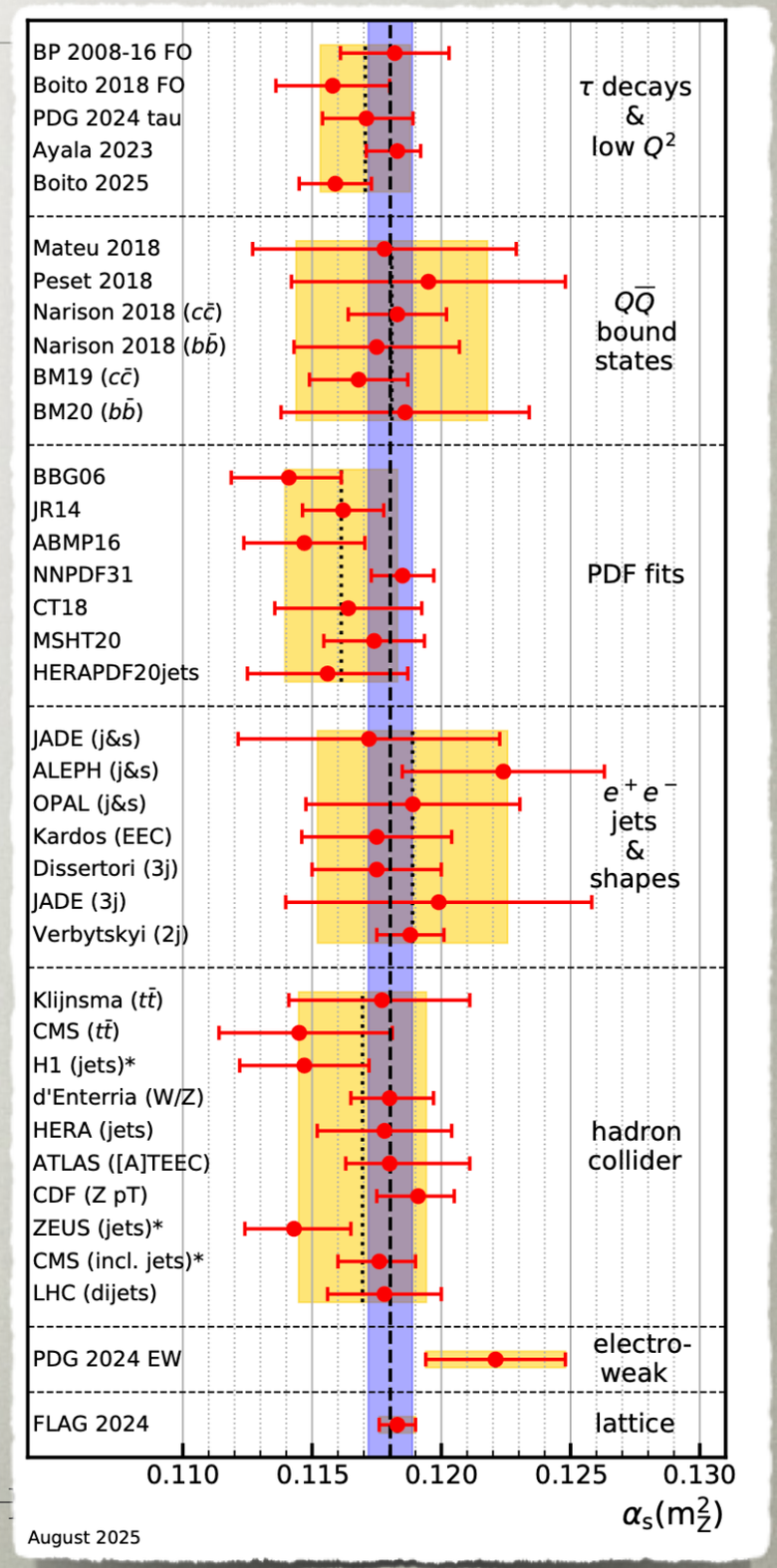


ATLAS, 2209.00583

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

SHOULD ONE CARE ABOUT HADRONIZATION

- $\alpha_S(M_Z)$ determinations from $e^+e^- \to$ hadrons
  - non-perturbative corrections to shape obs. scale as Λ/Q
  - difference when analytic modeling of hadronization (power corr.) within SCET vs. Monte Carlo models
- active area of research ⇒ premature to include in PDG averages

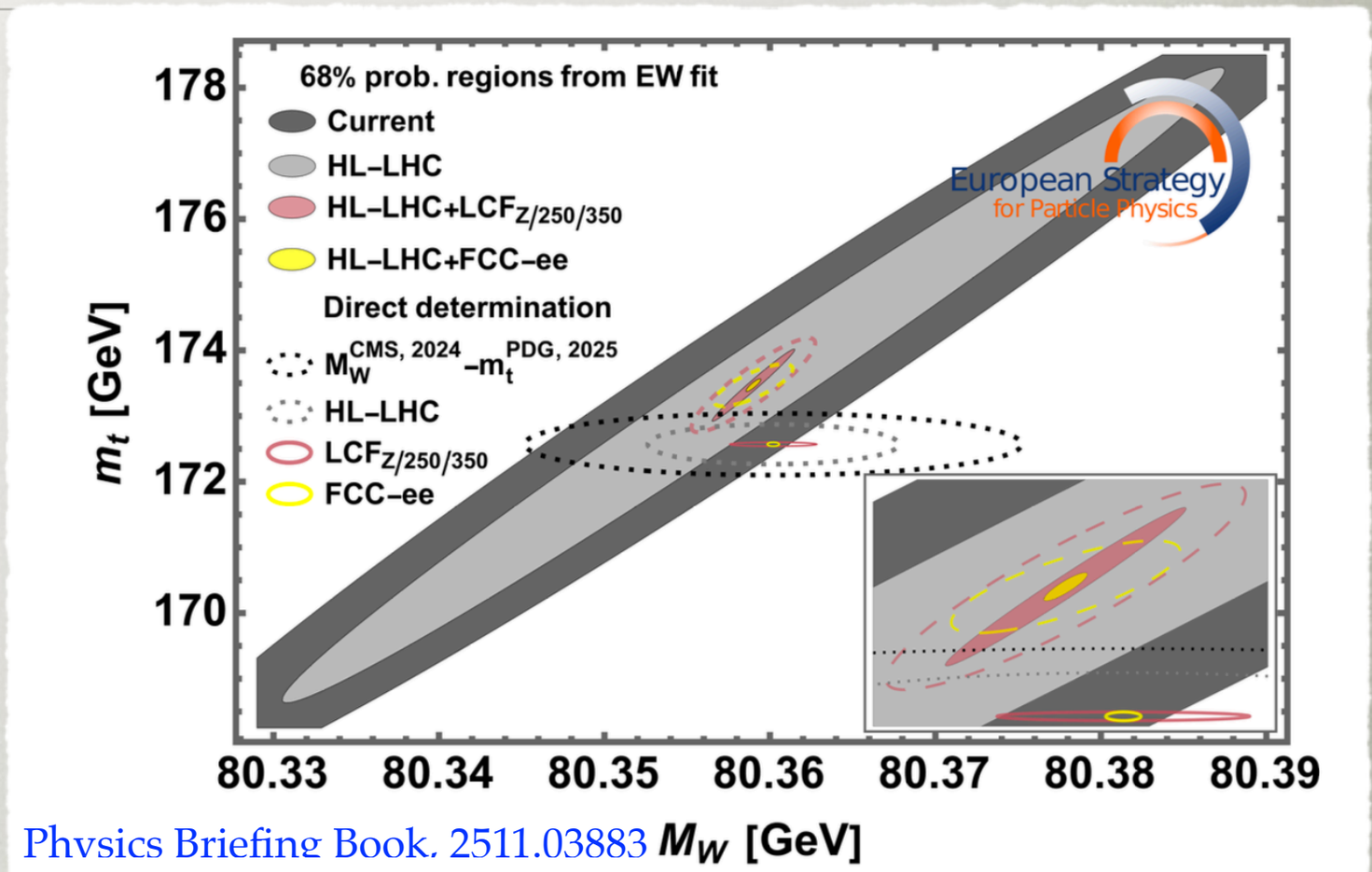


Huston, Rabbertz, Zanderighi, QCD review in 2025 PDG

SHOULD ONE CARE ABOUT HADRONIZATION

Physics Briefing Book, 2511.03883

- electroweak physics at FCC-ee
- Z-pole measurements: projections assume improvement on hadronisation uncertainties
  - conserv. sc.: $5\times - 10\times$
  - aggress. sc.: $50\times$



Physics Briefing Book, 2511.03883

- m_W from $e^+e^- \to WW$ at threshold
  - for $(W \to 2j)(W \to 2j)$ color reconn. unc. $\Delta m_W \approx 1\,\text{MeV}$
  - compare with (semi-)leptonic channels: $\Delta m_W \approx 0.3\,\text{MeV}$