Machine Learning Before the Collision: AI for Theory and Phenomenology

Benoît Assi

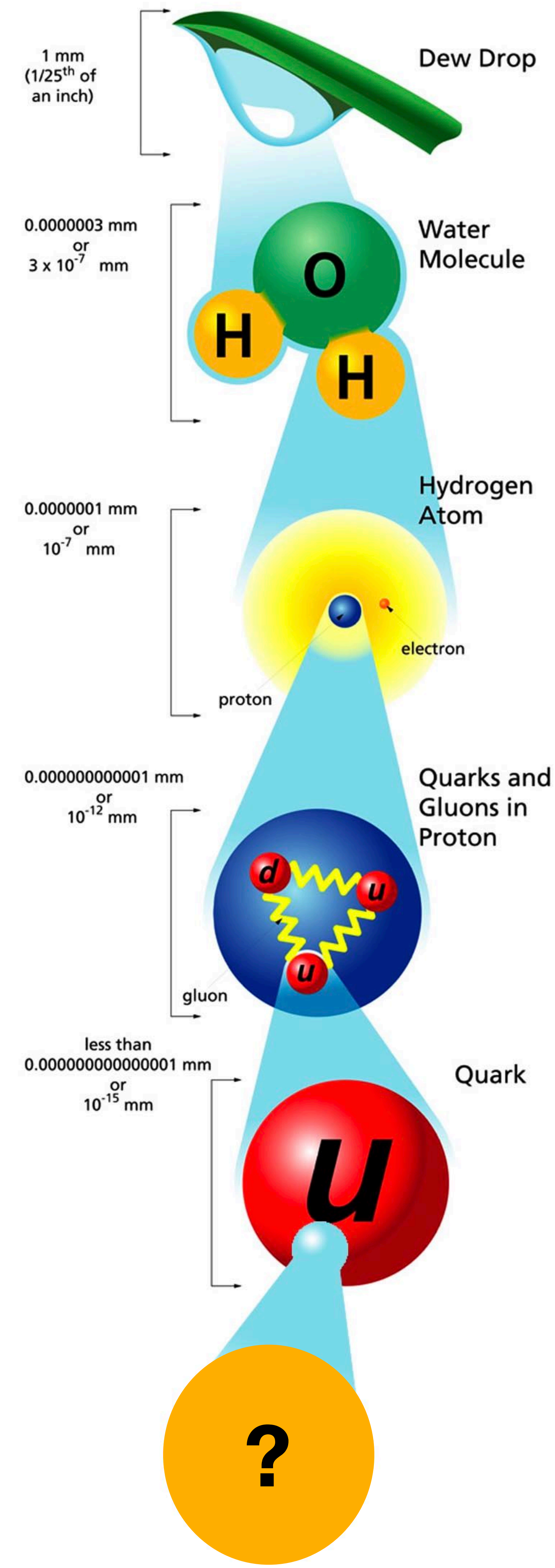
ML4Jets - Wien, Sep 14th, 2026

Three thrusts

I. **AI for amplitudes.** Transformers and symbolic regression assist bootstraps, integral reduction, analytic structure.

II. **AI for precision QCD.** High precsion for a few observables, data needs full events. Information theory bridges them.

III. **AI for BSM phenomenology.** Agents search model space, learned likelihoods sharpen EFT fits.

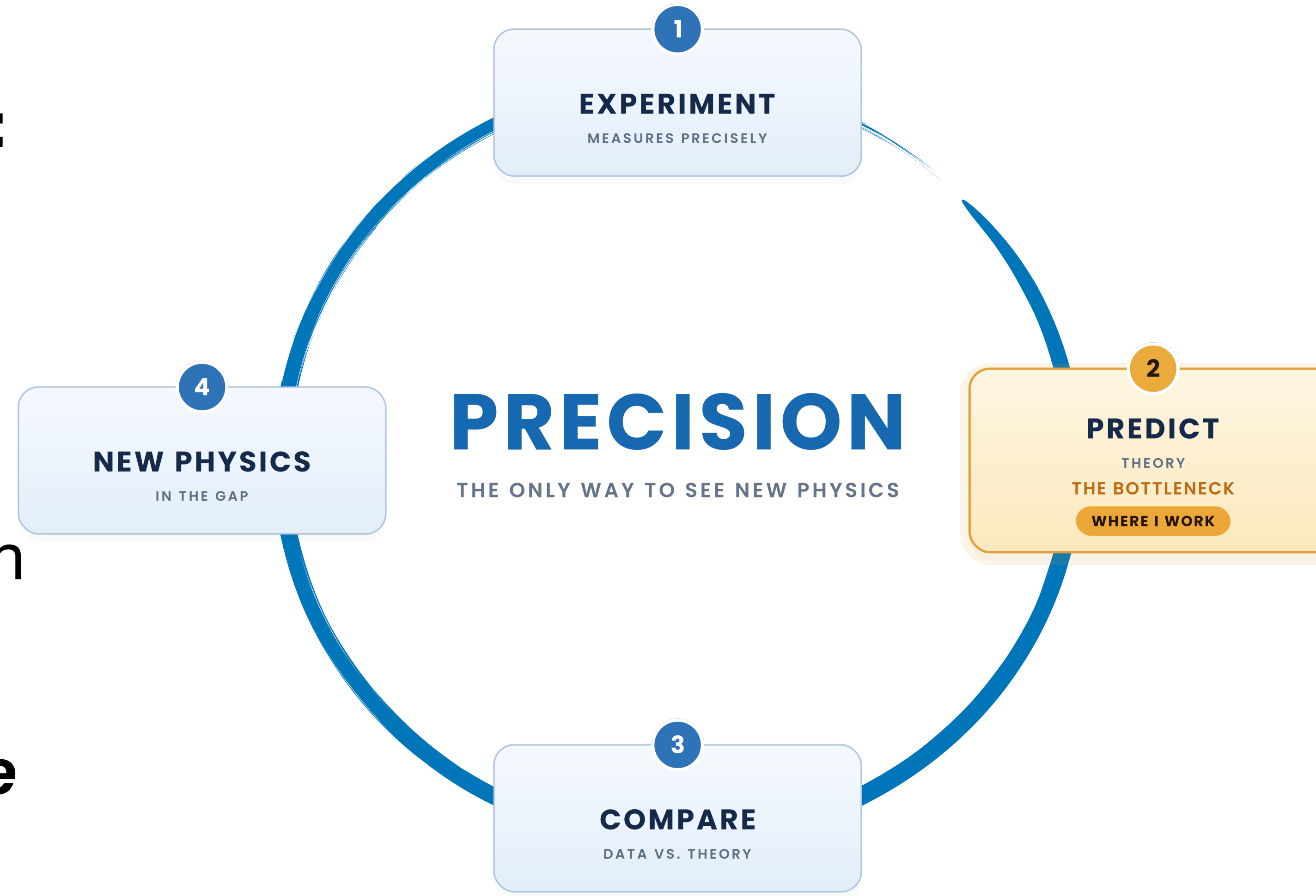


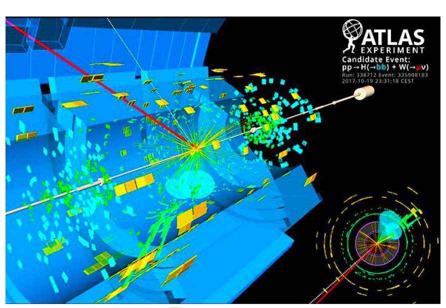
High-precision era

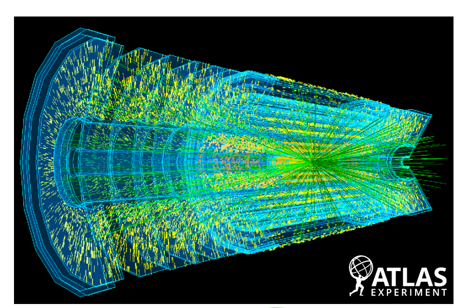
The Standard Model is incomplete: dark matter, matter-antimatter asymmetry, neutrino masses, …

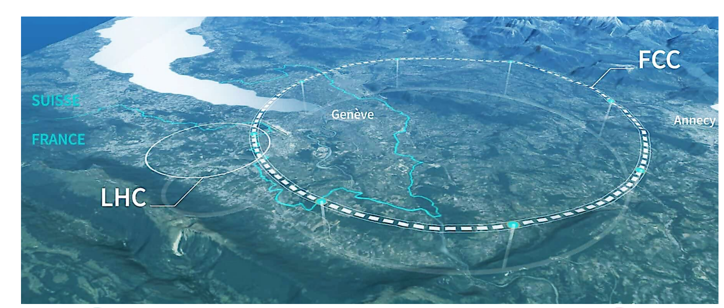
New physics is too heavy or too weakly coupled to produce directly. Hiding as small distortions in known processes.

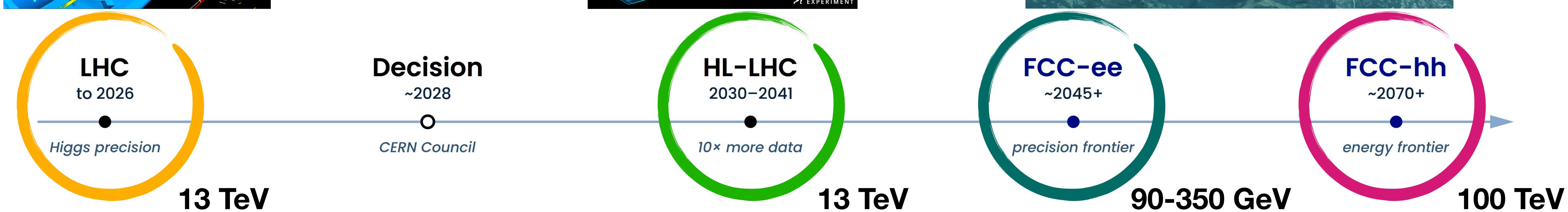
Soon the experiments will measure better than we can predict.



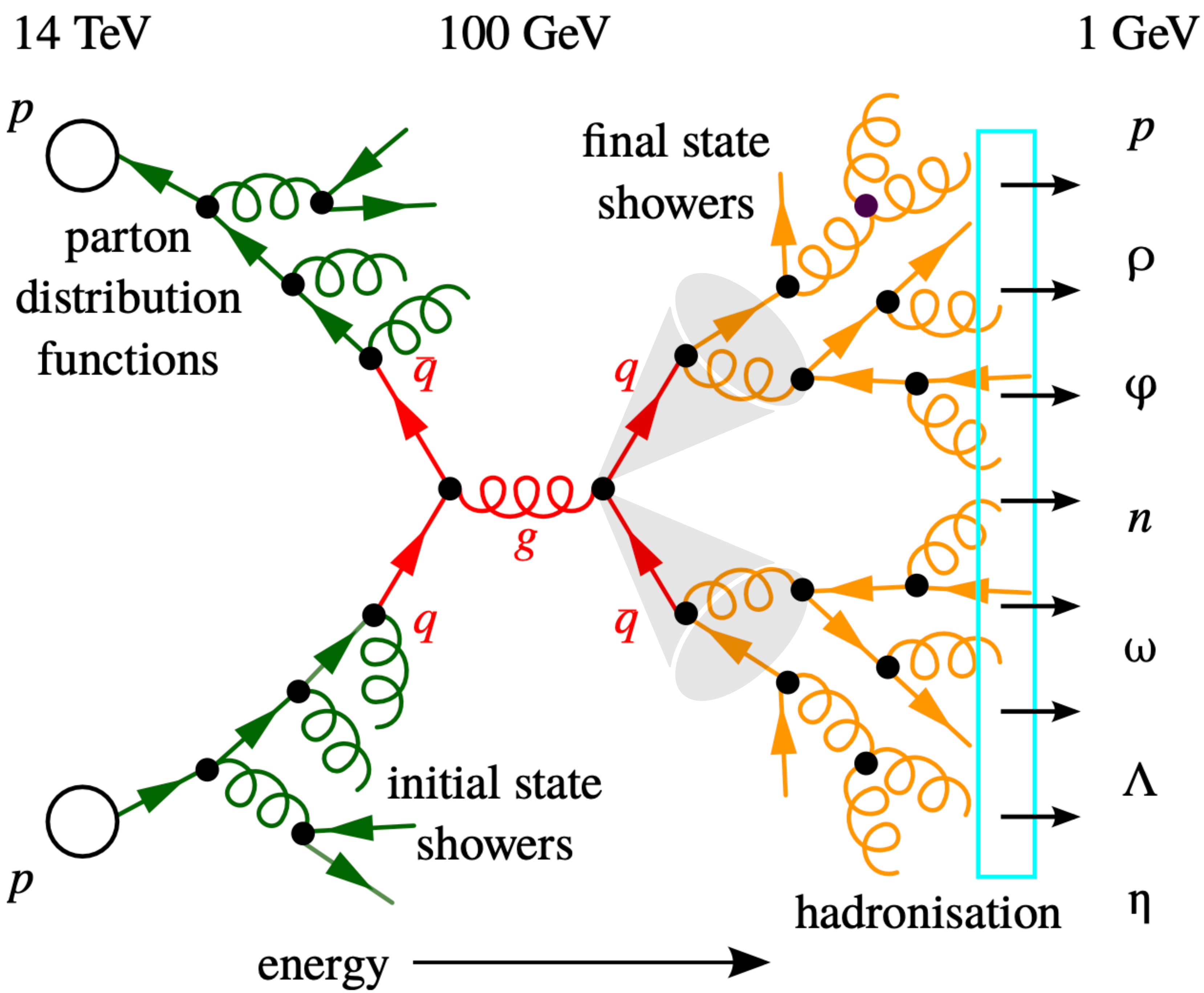








Anatomy of a collision event



The complete final state, every particle, must be calculated with controlled QFT accuracy

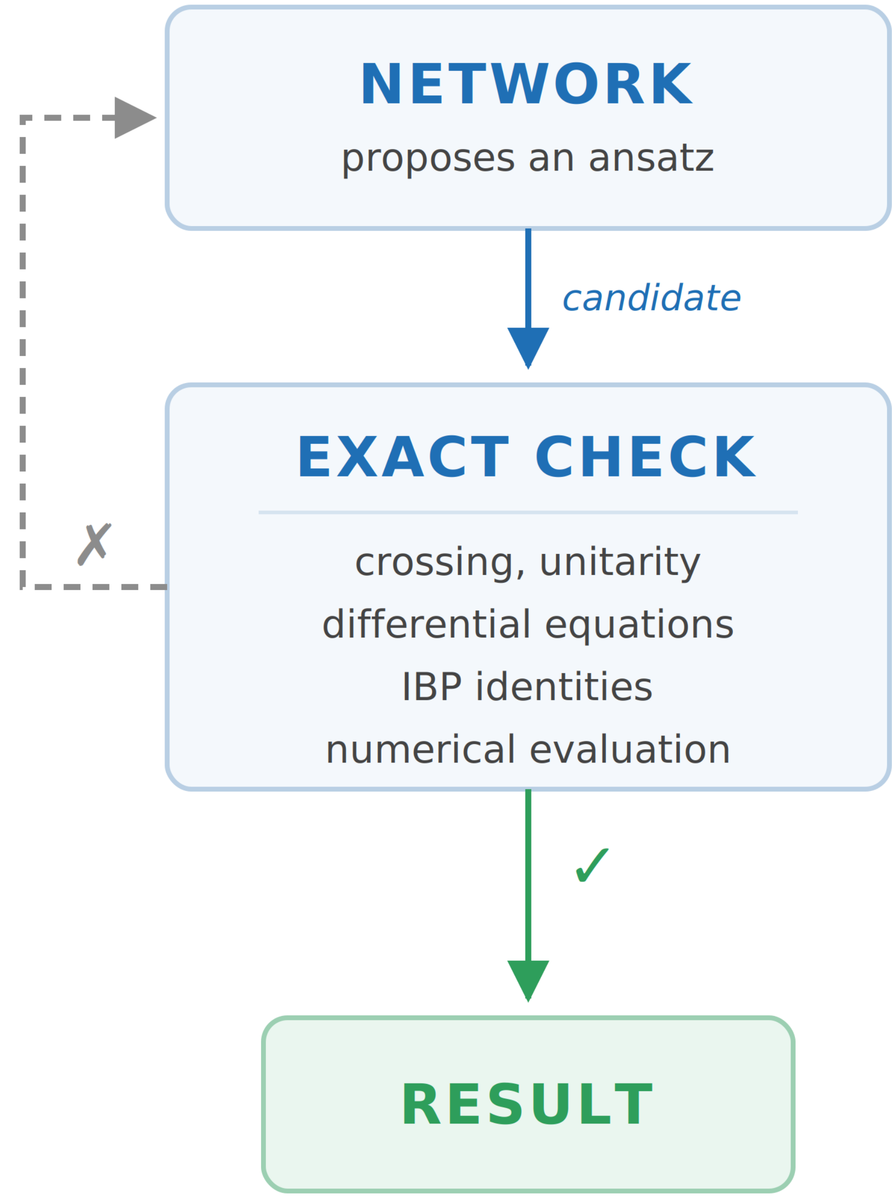
AI for Amplitudes

Where the answer can be checked

Bootstrap ansätze
guessing the function basis and its coefficients [Cai et al., MLST 5 (2024) 035073]

Simplification
compact forms from sprawling but correct expressions [Cheung, Dersy, Schwartz, SciPost Phys. 18 (2025) 040]

Integral reduction
IBP orderings, master integral bases [Zeng, PRD (2025); von Hippel & Wilhelm, JHEP 05 (2025) 185; Shih (2026) 2604.05034]



Fun with Calabi–Yau: *the metric Yau proved exists and nobody could write down, ML made cheap enough to find numerically so that first-principles Yukawa couplings and quark masses became possible* [Larfors et al., arXiv:2205.13408; Constantin et al., Nucl. Phys. B 1010 (2025) 116778]

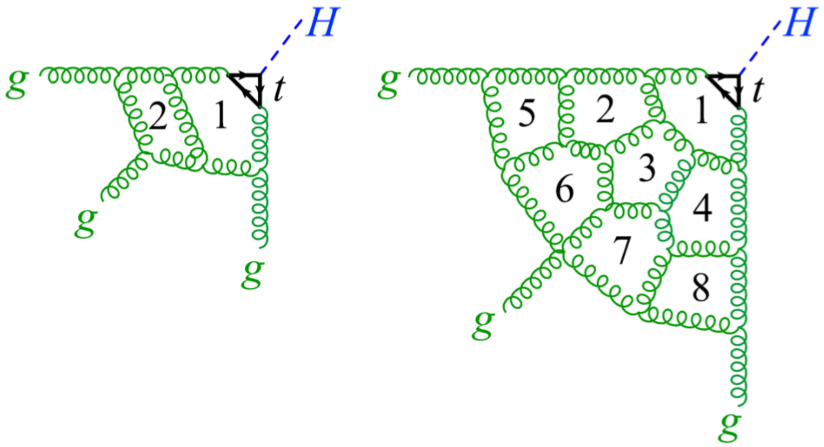
Bootstrapping with transformers

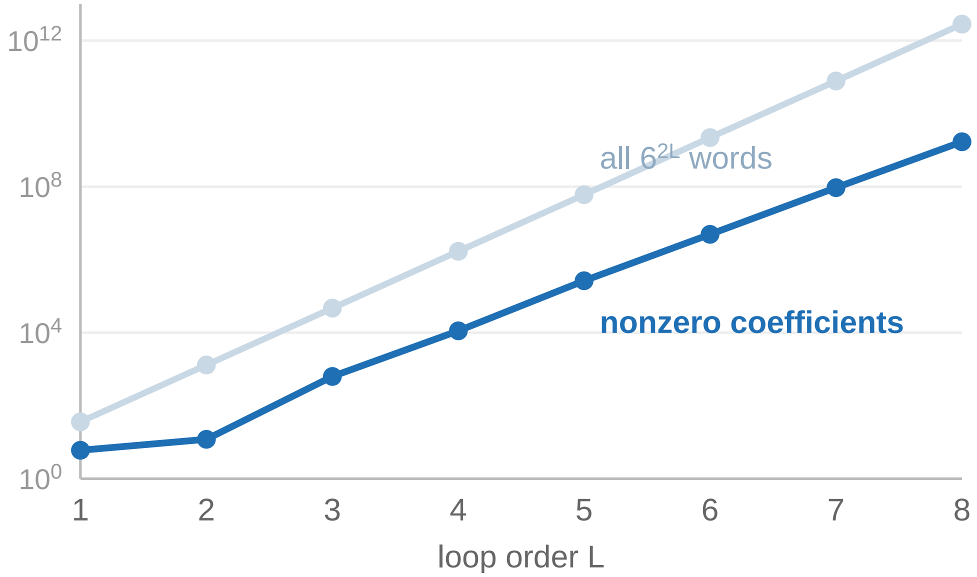
Every loop buys precision but it costs a factorial number of diagrams

No diagrams — most general answer the symmetries allow and let analyticity do the rest. In *planar* **N = 4 super Yang–Mills** — gluons plus a

$$S[F^{(L)}] = \sum_w C^w \; l_1 \otimes \cdots \otimes l_{2L}$$

*The amplitude is a list of integers **C**. They are not independent.*





Unknowns ×4 per loop, solve time ×16. Dead by **L ≈ 9**.

Why a transformer?

A sparse integer linear system — the shape transformers already solve.

Every guess gets checked. **98% is fine when the 2% can't hide.**

Where it's going

Past nine loops.

Reading back what it learned as new relations. Not done yet.

What it isn't

QCD. Same functions, plus lower-weight terms and rational prefactors SYM lacks.

Basis, alphabet, constraints — all still human.

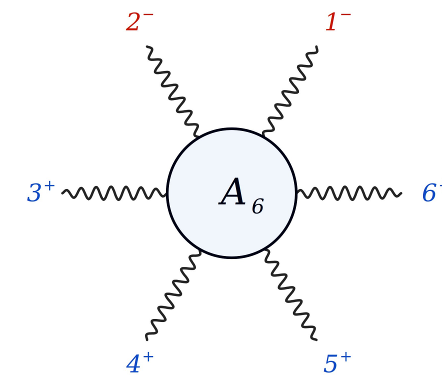
[Cai et al., MLST 5 (2024) 035073]

Simplification is guesswork

Every amplitude is a rational function of spinor brackets — 220 diagrams for gg → gggg, and the answer is *one line*.

The path isn't unique — Schouten and momentum conservation mean equal expressions *don't look equal*. The order of moves is the whole problem.

A transformer does the guessing — trained by scrambling known answers backwards. A second network learns which terms to look at, the way a person does.



Parke–Taylor, 1986

$$A(1^-2^-3^+4^+5^+6^+) = \frac{\langle 12\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 45\rangle\langle 56\rangle\langle 61\rangle}$$

Schouten — any three 2-spinors are dependent

$$\langle ij\rangle\langle kl\rangle + \langle ik\rangle\langle lj\rangle + \langle il\rangle\langle jk\rangle = 0$$

↑ same brackets, different pairing

298 terms in, after one pass

$$\langle 13\rangle\langle 24\rangle[13][25] - \langle 14\rangle\langle 23\rangle[13][25] + \langle 14\rangle\langle 34\rangle[25][34] + \langle 14\rangle\langle 34\rangle[23][45] +$$

↓ transformer, checked numerically

$$\langle 12\rangle\langle 34\rangle[13][25] + \langle 14\rangle\langle 34\rangle[13][45] + \cdots$$

Every candidate is checked numerically — the network proposes, the check disposes.
Unprompted, the embedding sorted terms by mass dimension.

Cheung, Dersy, Schwartz, SciPost Phys. 18 (2025) 040
Dersy, Schwartz, Zhang, IJDSMS 1 (2024) 135 · shown at ML4Jets 2024

Integration is a memory wall

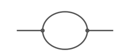
Every loop amplitude reduces to a few master integrals — the reduction, not the integrals, is the bottleneck. *Hundreds of GB.*

Laporta solves everything at once — memory grows with the integral. **Or one move at a time** — but most moves make it *worse*.

A transformer picks the move — trained by scrambling known answers backwards. It learned to step *back* to go forward.

Memory flat where Kira climbs. Slower, until it isn't. One topology.

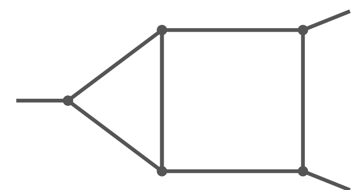
one loop: a vanishing total derivative relates neighbours



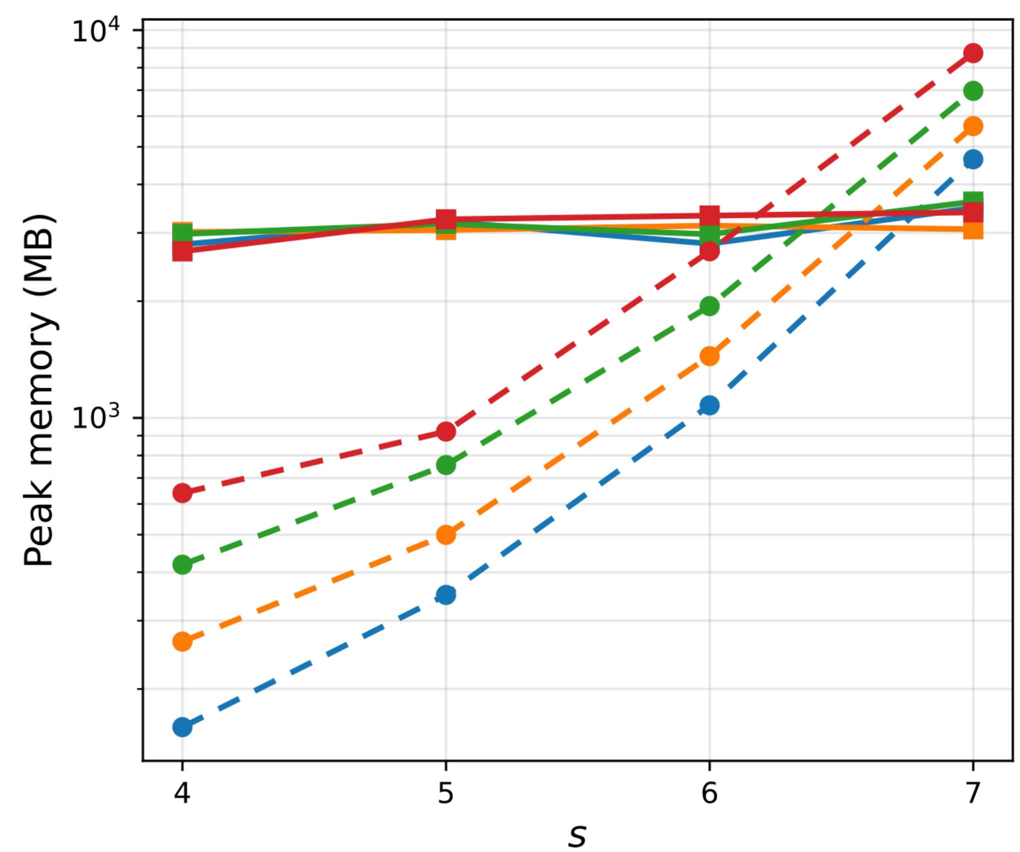
$$(d-2a-b)\,I[a,b] \;-\; b\,I[a-1,\,b+1] \;+\; b\,p^2\,I[a,\,b+1] \;=\; 0$$

$$I[1,2] \;=\; \frac{3-d}{p^2}\,I[1,1]$$ one move, done

two loops, six propagators



$I[2,1,2,1,2,2,\,-4]$ thousands of moves



Shih, arXiv:2604.05034

Shih, arXiv:2604.05034 · von Hippel & Wilhelm, JHEP 05 (2025) 185

AI for precision theory

Where the check is partial

Filling the gap

Theory fixes a few moments. Data needs every event. Put the calculation's accuracy, and its uncertainty, on the whole event [Assi, Höche, Lee, Thaler, PRL 135 (2025) 131901, Assi, Lee, Thaler, JHEP (2026), Mastandrea, Gambhir, 2512.04160]

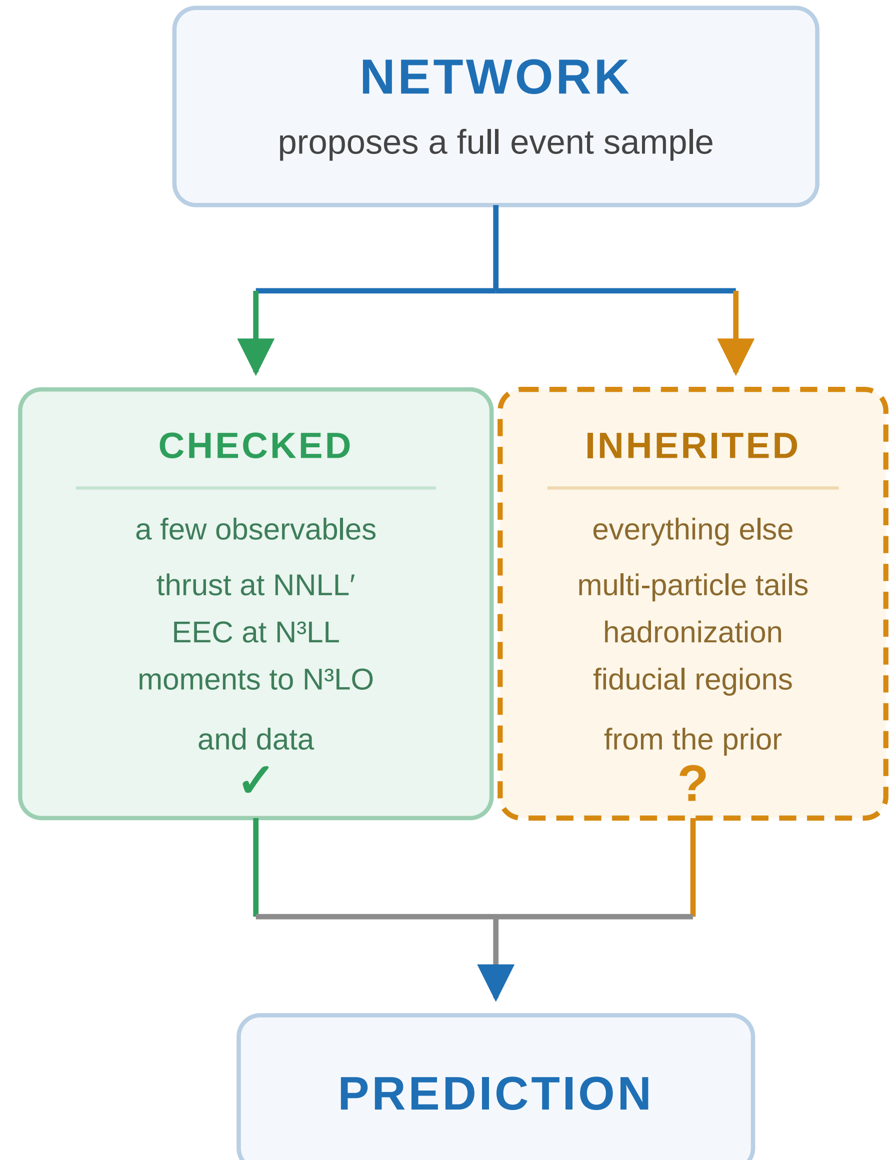
Emulating the calculation

Higher order matrix elements cost too much per phase space point. Learn the amplitude, and learn where to sample [Heimel et al., SciPost 15 (2023) 141]

Machine-learned observables

Thrust was a guess. Optimize the observable for what you want to measure, keeping it IRC-safe so theory can still compute it [Komiske, Metodiev, Thaler, JHEP 07 (2020) 006, SHAPER, JHEP 06 (2023) 195, Larkoski, Thaler, JHEP 08 (2023) 107]

MLHAD ***See Manu's talk on Fri!***



The check is partial.

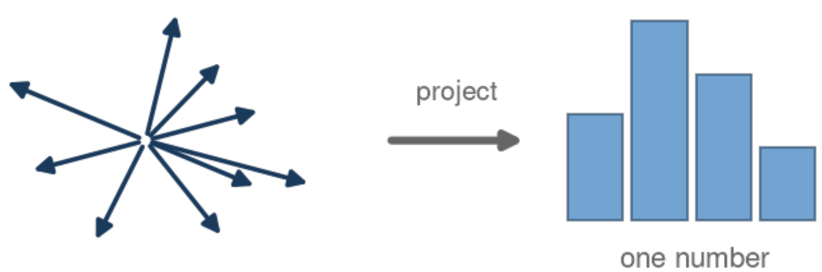
What theory doesn't pin down, the prior decides.

Precision in a hurry

From momenta to observables

Observable is ***one number*** projected from the spray of momenta from an event

jet substructure · missing energy · reconstructed mass · event shapes · multiplicities …

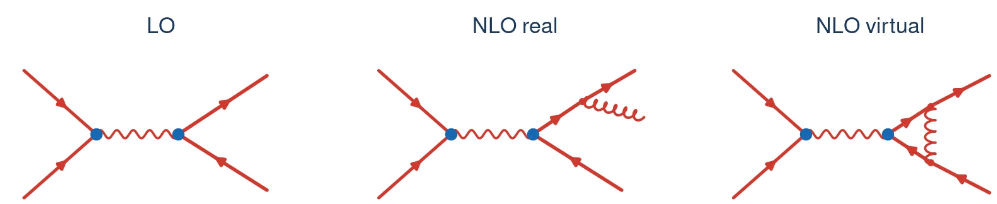


Fixed order (analytic)

$$\frac{d\sigma}{dv} = \underbrace{c_0(v)}_{\text{LO}} + \underbrace{\alpha_s c_1(v)}_{\text{NLO}} + \underbrace{\alpha_s^2 c_2(v)}_{\text{NNLO}} + \cdots$$

Expanded in the coupling. Each term ~10× smaller

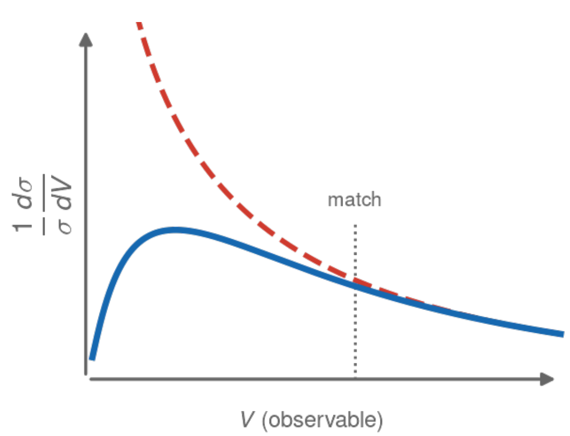
*Many SM processes and observables **automated** at* NNLO *·*
some known at $\mathrm{N^3LO}$



Resummed (analytic)

$$\Sigma(v) \propto \exp\Big[\underbrace{\alpha_s L^2}_{\text{LL}} + \underbrace{\alpha_s L}_{\text{NLL}} + \cdots\Big], \qquad L = \log\frac{1}{v}$$

large logs in the distribution, summed to all orders



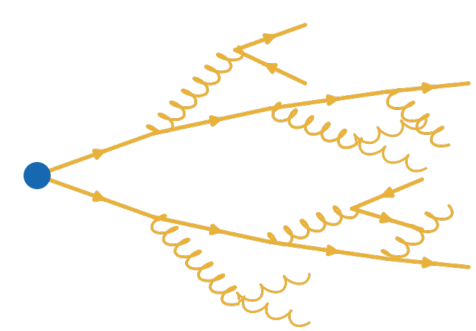
Bespoke multi-year effort for each observable

Some observables known at
NLL *· fewer at* $\mathrm{NNLL/N^3LL}$

Parton Shower (Monte-Carlo)

$$\ln\Delta(t) = -\int \frac{\alpha_s}{2\pi}\frac{dk_t^2}{k_t^2}P(z)\,dz$$

The splitting probability of one emission, iterated.



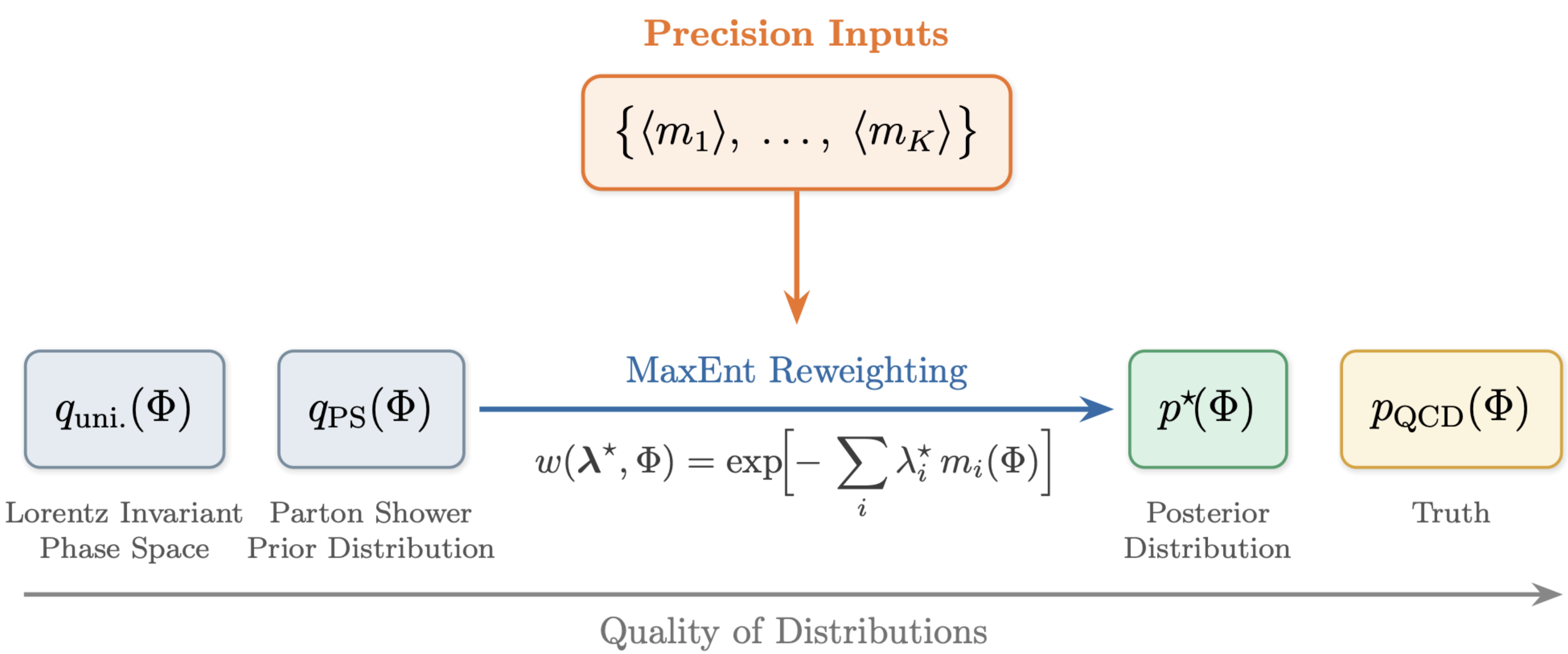
Sums the logs and builds the whole event, all momenta

*All observables **automated** and known at* LL *· **recently*** NLL

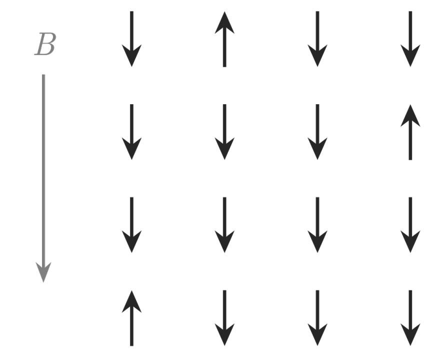
Herren, Höche et al. JHEP 10 (2023)
BA and Höche PRD 109 (2024),…

Parton showers most useful but weakest accuracy

Can we directly teach event generators higher-order theory?



Information-Theoretic optimization



Boltzmann factor:

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

Maximum entropy solution subject to moment constraint on average energy

$$S = -\sum_i p_i \ln p_i, \qquad \sum_i p_i E_i = \langle E \rangle$$

Generalization:

$$p(x) = \overset{\text{Prior}}{q(x)} \exp\Big[-\overset{\text{Sets Partition Function}}{\beta_0} - \underset{\text{Lagrange Multiplier}}{\beta_1} \underset{\text{Constrained Quantity}}{f_1(x)} - \beta_2 f_2(x) - \ldots \Big]$$

If you can perform precision QCD calculations of moment constraints, then you can improve prior MC predictions with learned Lagrange multipliers

Information-Theoretic Optimization

Sudakov structure: take any event-shape distribution

$$r(\tau)_{\mathrm{LL,f.c.}} = \frac{-2\alpha_s C_F}{\pi} \frac{\ln\tau}{\tau} \exp\left[-\frac{\alpha_s C_F}{\pi}\ln^2\tau\right]$$

Sudakov factor

ML perspective: Sudakov is a kind of **Boltzmann factor** and cusp anomalous dimension is like a **Lagrange multiplier** that enforces a **constraint on the 2nd log moment** of the distribution

log moments not previously measured or calculated before in QCD

BA, Höche, Lee, Thaler PRL 132 (2025)

Relative entropy

Take **relative entropy**

$$S_{\mathrm{KL}}(P\,||\,Q) = -\int dx\, p(x)\, \log \frac{q(x)}{p(x)}$$

Continuous version of entropy

Normalized Prior

Plus precision **moment constraints**

$$+\,(\lambda_0 - 1)\big(\langle m_0\rangle_p - 1\big) + \sum_i \lambda_i \left(\langle m_i\rangle_p - c_i\right)$$

Like Lagrangian you have in Stat. mech.

MC prior moment

Precise theory moment

Unique **stationary solution**

$$w(\Phi;\lambda_0,\boldsymbol{\lambda}) = \exp\left[-\lambda_0 - \sum_i \lambda_i\, m_i(\Phi)\right]$$

Convex, strictly positive per-event weights, full event exclusivity preserved

BA, Höche, Lee, Thaler PRL 132 (2025)
BA, Lee, Thaler 2604.00084

HEP ? ML

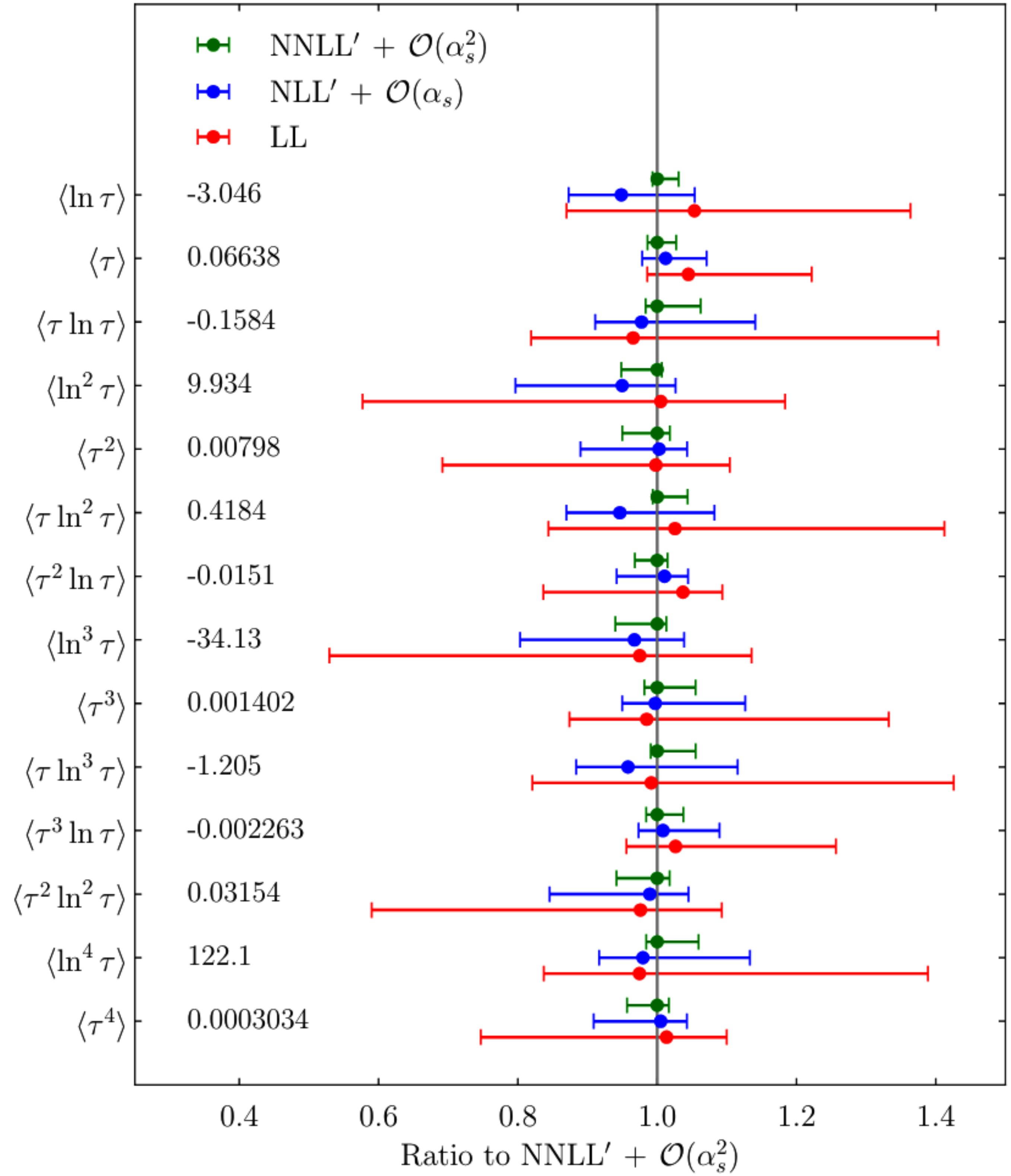
Moments of thrust at LEP

Moments probe analytic *Sudakov structure* of thrust

$$r(v) = \delta(v) + \sum_{m=1}^{\infty}\sum_{n=1}^{2m-1} k_{mn}^{\mathrm{LP}}\left(\frac{\alpha_s}{4\pi}\right)^m \left[\frac{\ln^n v}{v}\right]_+ + \cdots + \sum_{m=1}^{\infty}\sum_{n=1}^{2m-1} k_{mn}^{\mathrm{N^kLP}}\left(\frac{\alpha_s}{4\pi}\right)^m \frac{\ln^n v}{v} v^{k-1}$$

Computed to NNLL + NNLO. ***First logarithmic moments computed in QCD.***

Current showers LL and carry large uncertainty, NNLL + NNLO **shrinks it** down dramatically **on every moment**



BA, Höche, Lee, Thaler, PRL 132 (2025)

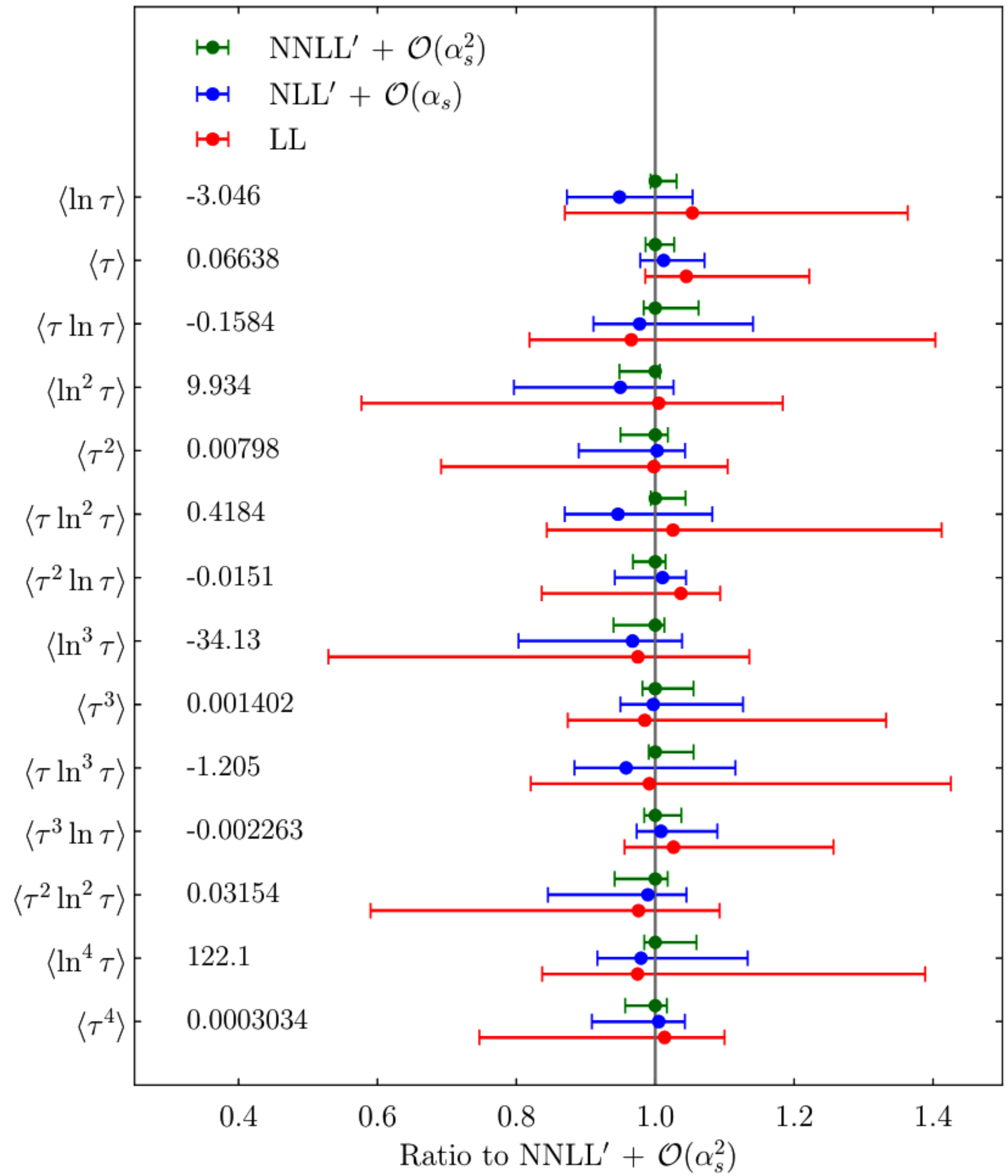
Moments of thrust at LEP

Moments probe analytic *Sudakov structure* of thrust

$$g_{mn}(\tau) = \tau^m \ln^n \tau$$

Computed to NNLL + NNLO.
First logarithmic moments computed in QCD.

Current showers LL and carry large uncertainty, NNLL + NNLO **shrinks it** down dramatically **on every moment**



BA, Höche, Lee, Thaler, PRL 132 (2025)

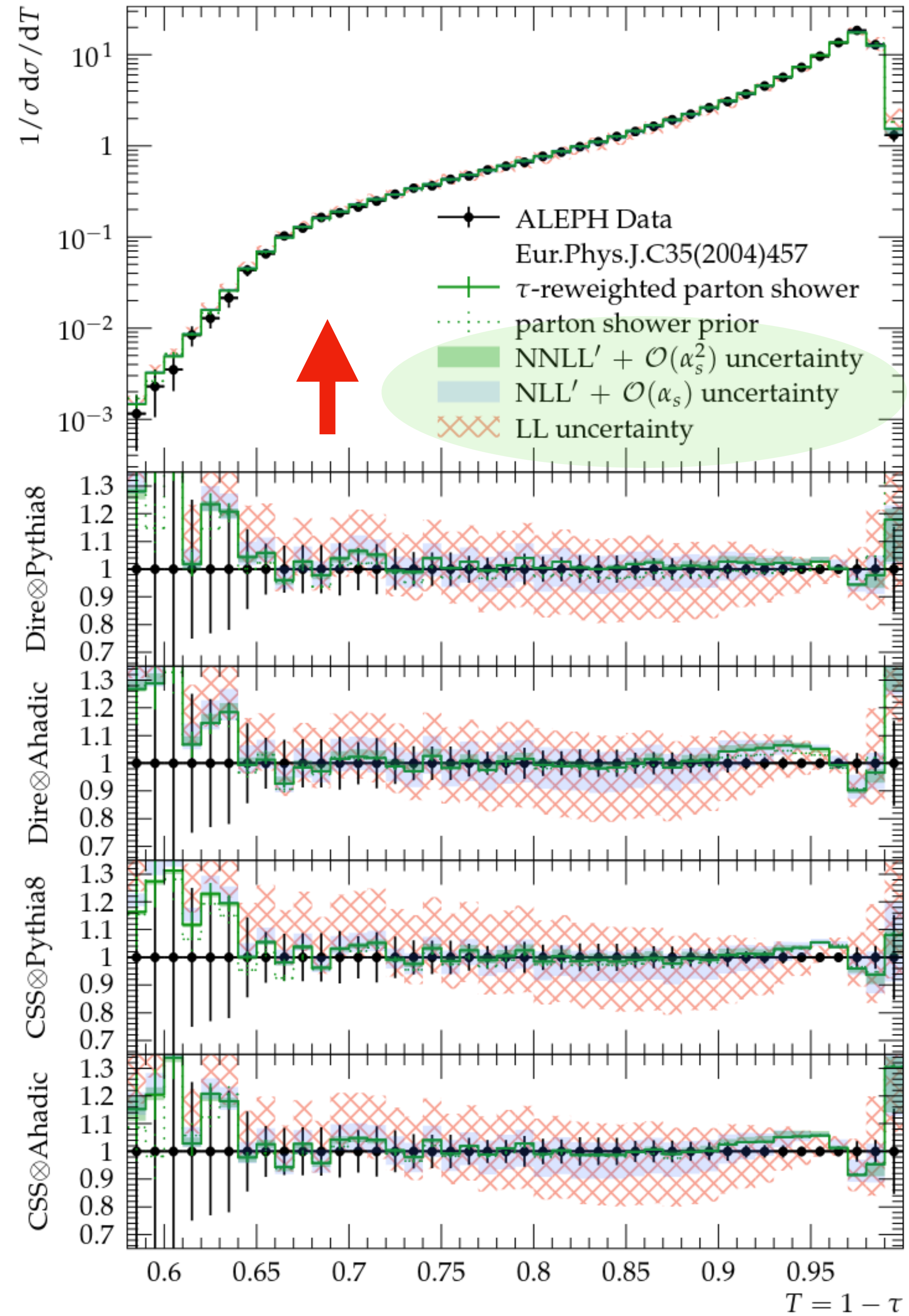
A new benchmark for event shapes

Four phase space priors, two showers by two hadronization models

Converges form LL to NNLL + NNLO, a **new accuracy** benchmark

Theory uncertainties propagate directly, moments → multipliers → predictions

Improvement systematic: more moments & observables (how many?), more accuracy



State of the art theory, taught event by event, with interpretable uncertainties

BA, Höche, Lee, Thaler PRL 132 (2025)

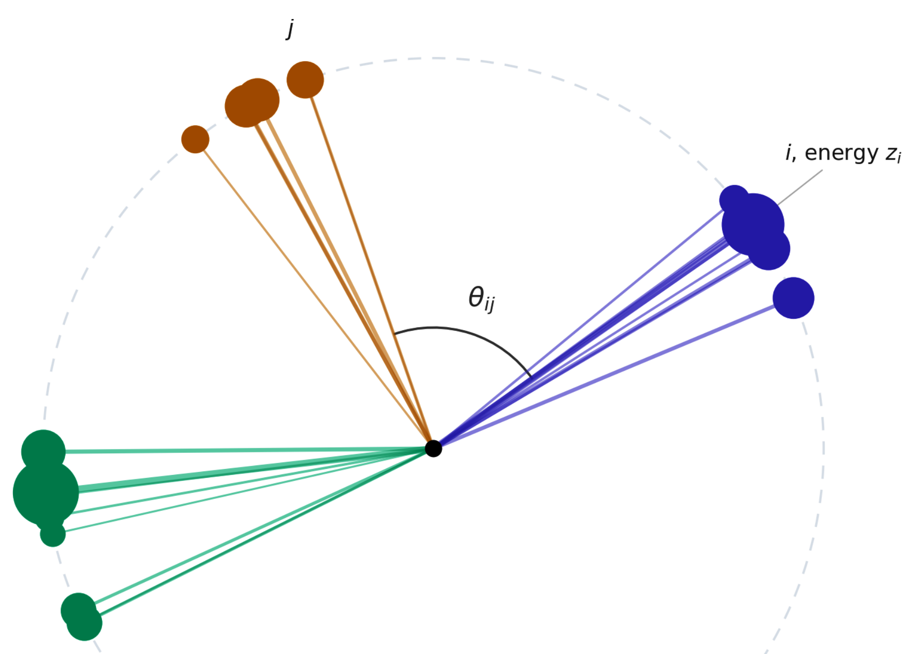
All observables at once

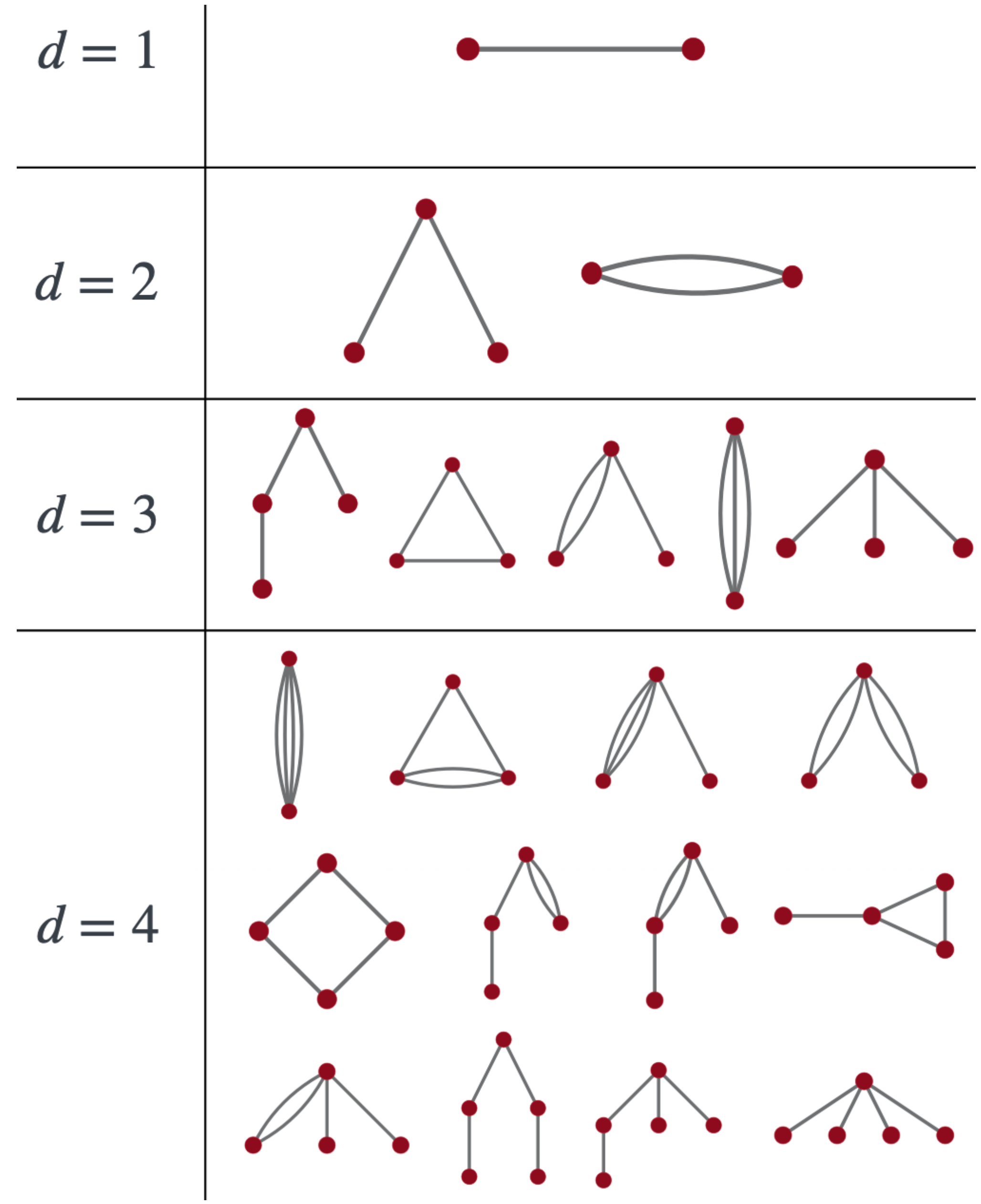
A complete basis of observables — **Energy flow polynomials**

$$\text{EFP}_G^{(\beta)} \equiv \sum_{i_1,\ldots,i_k \in R} \prod_{v \in V} z_{i_v} \prod_{(a,b)\in E} \theta_{i_a i_b}^{\beta}$$

Calculable observables = particle energies and angles — EFPs span them

Low degree captures the bulk — a few moments reach every observable at once





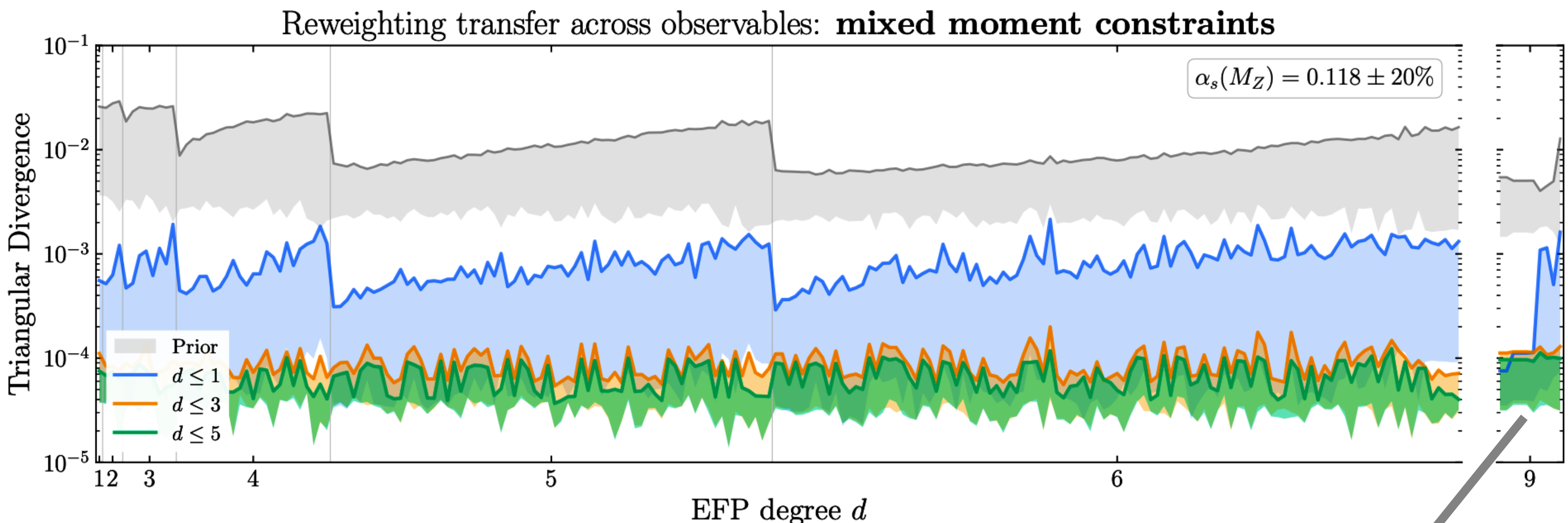
Vertices → energy fractions: $z_i = \dfrac{E_i}{\sum_j E_j}$

Edges → pairwise angles: $\theta_{ij} = \sqrt{2(1 - \cos\angle_{ij})}$

Komiske, Metodiev, Thaler JHEP (2018)

Information saturation

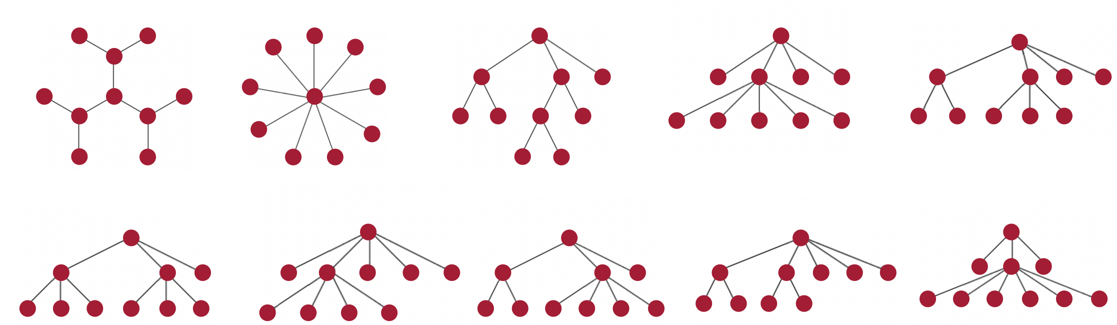
Saturation test setup: prior ***phase space generator***, momentum conserving but *wrong radiation pattern* → IT reweight *to* ***accurate parton shower***



Train on EFPs up to $d \leq 5$, test out to $d \leq 9$

With moment types: $\langle \mathrm{EFP}_G^m \ln^n \mathrm{EFP}_G \rangle$

Saturation by $d \leq 3$ primes and composites, *robust under large prior variation*

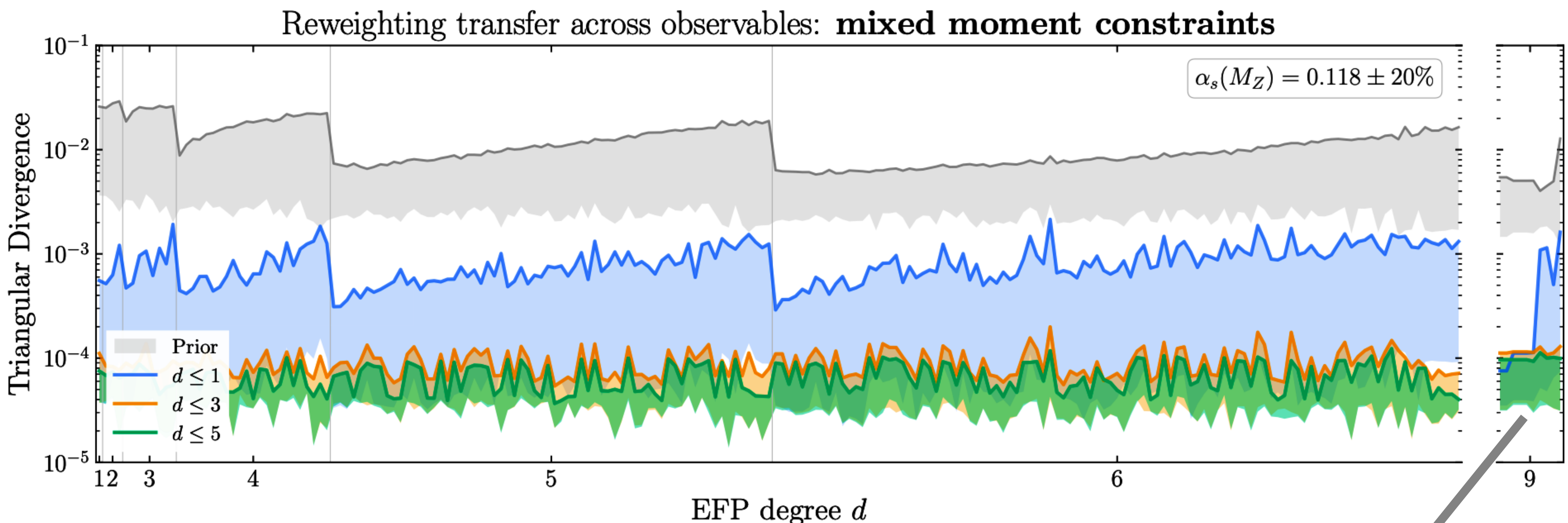


*We focus on the **Cinematic** degree 9 subset of homeorphically irreducible trees*

BA, Lee, Thaler 2604.00084

Information saturation

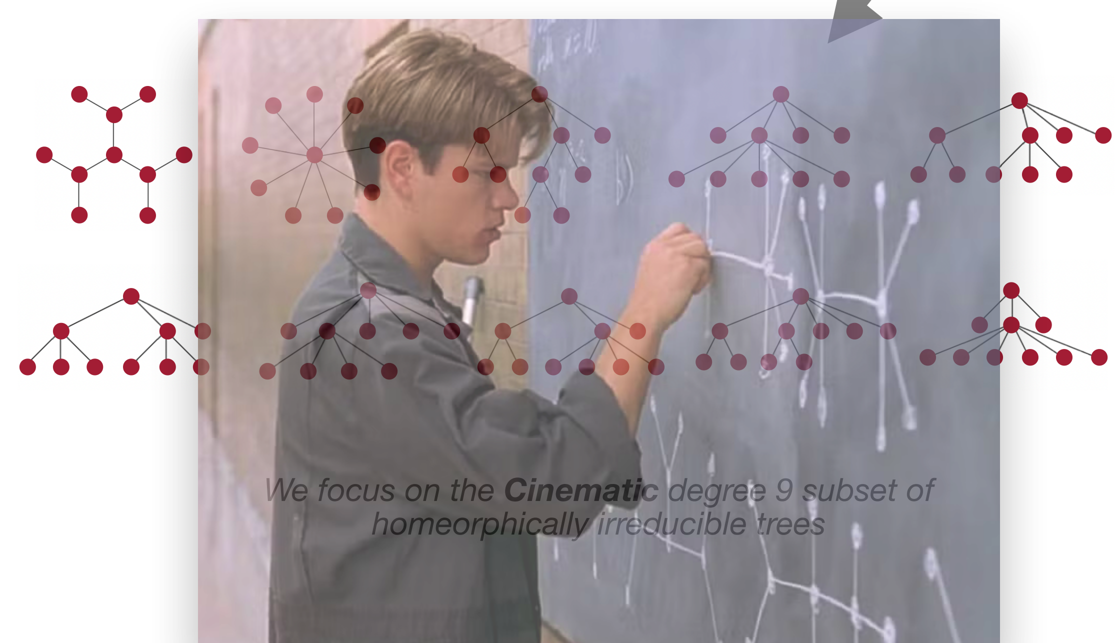
Saturation test setup: prior ***phase space generator***, momentum conserving but *wrong radiation pattern* → IT reweight *to* ***accurate parton shower***



Train on EFPs up to $d \leq 5$, test out to $d \leq 9$

With moment types: $\langle \mathrm{EFP}_G^m \ln^n \mathrm{EFP}_G \rangle$

Saturation by $d \leq 3$ primes and composites, *robust under large prior variation*



BA, Lee, Thaler 2604.00084

Outlook

One event generator that respects every best QCD constraint at once, with *rigorous uncertainties*.

Cross-domain impact: The same framework brings theory-controlled uncertainties to nuclear and neutrino generators, DUNE, EIC, heavy-ion.

Upcoming

Highest orders: $\mathrm{N^3LO} + \mathrm{N^4LL}$ *plus ab-initio LQCD across processes and observables, starting with Drell-Yan for ATLAS/CMS*

BA, Campbell, Höche, Thaler et al. *In preparation*

Code will be made public

ML for MC variation: *combine simulation variation with efficient sampling for full prior-plus-target uncertainty for precise predictions*

BA and Thaler, *In preparation*

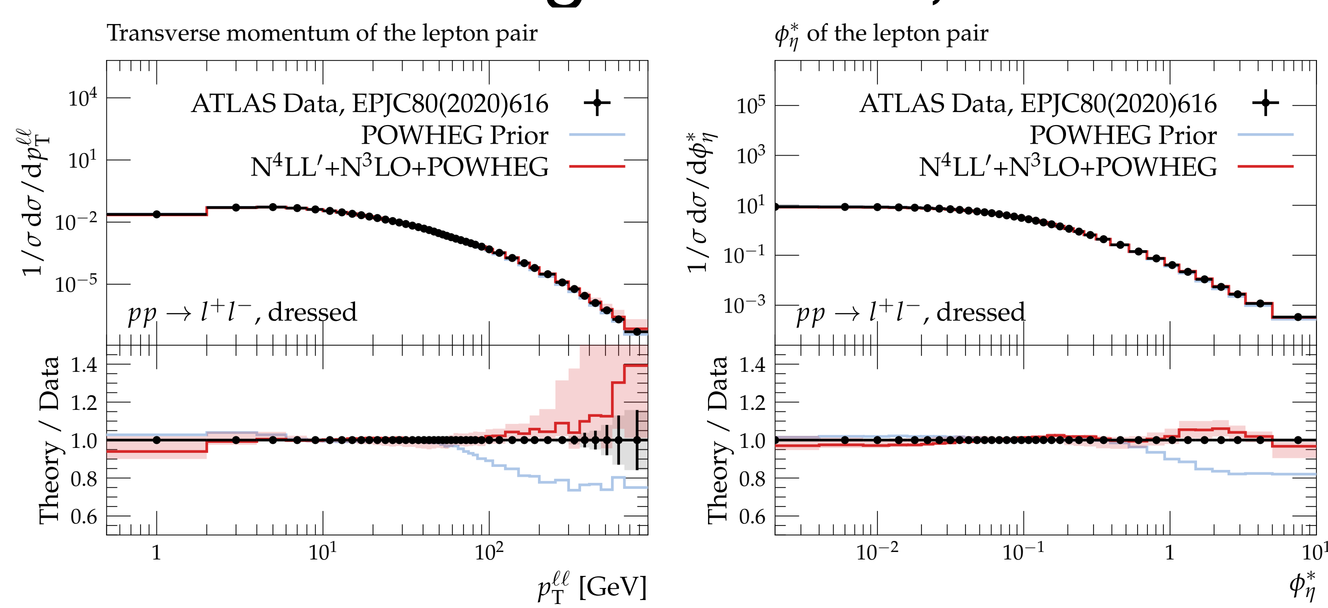
Negative weight alleviation: *efficient matching procedure intrinsically positively weighted for PS* $+ \mathrm{N^kLO}$

BA, Poncelet, Whitehead, *In preparation*

Outlook

One event generator that respects every best QCD constraint at once, with *rigorous uncertainties*.

Cross-domain impact: The same framework brings theory-controlled uncertainties to nuclear and neutrino generators, DUNE, EIC, heavy-ion.



Upcoming

Highest orders: N^{3}LO + N^{4}LL *plus ab-initi*
with Drell-Yan for ATLAS/CMS

Code will be made public

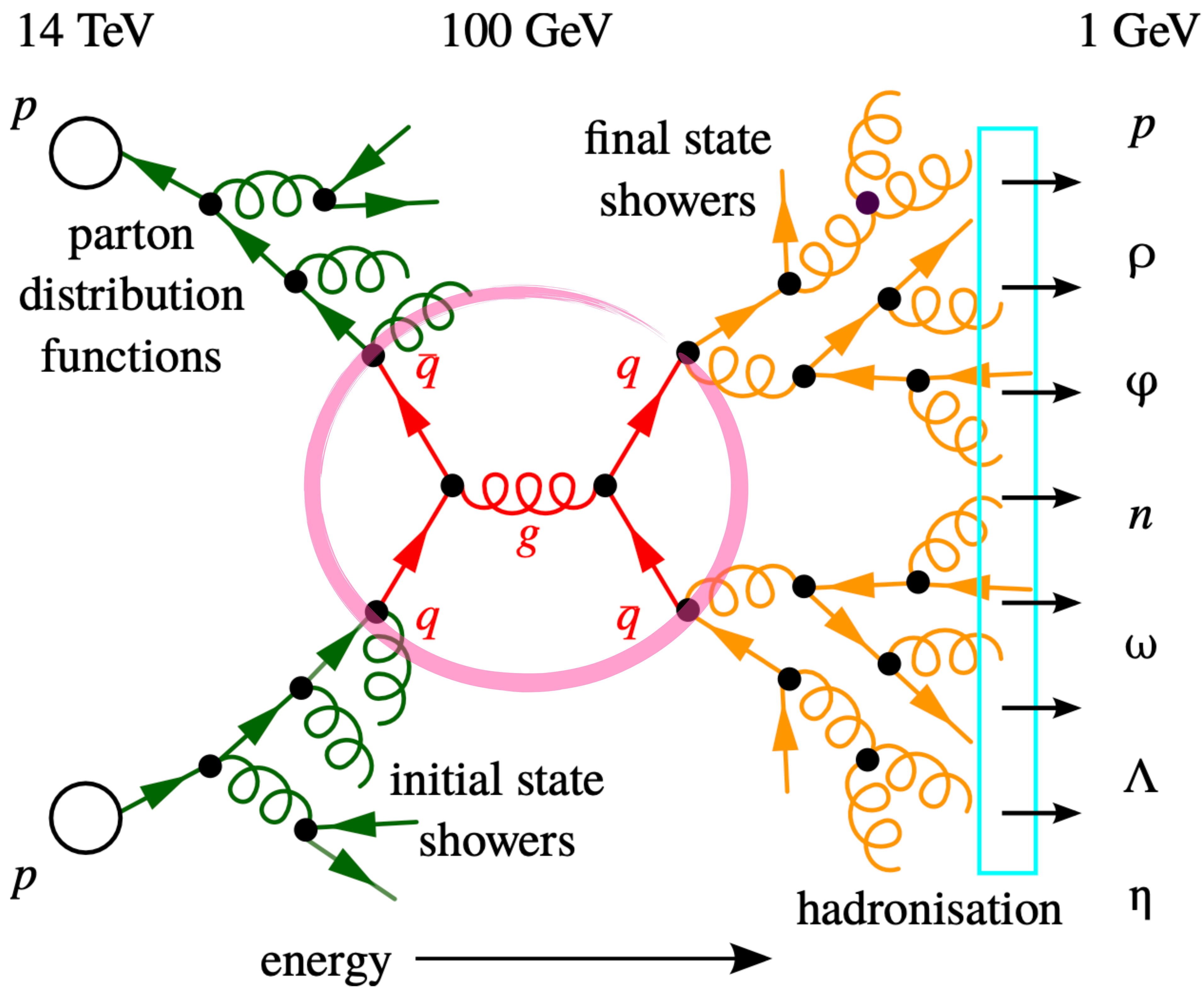
ML for MC variation: *combine simulation variation with efficient sampling for full prior-plus-target uncertainty for precise predictions*

BA and Thaler, *In preparation*

Negative weight alleviation: *efficient matching procedure intrinsically positively weighted for PS* + N^{k}LO

BA, Poncelet, Whitehead, *In preparation*

Anatomy of a collision event



AI for BSM Phenomenology

Where the check can only say no

Proposing models

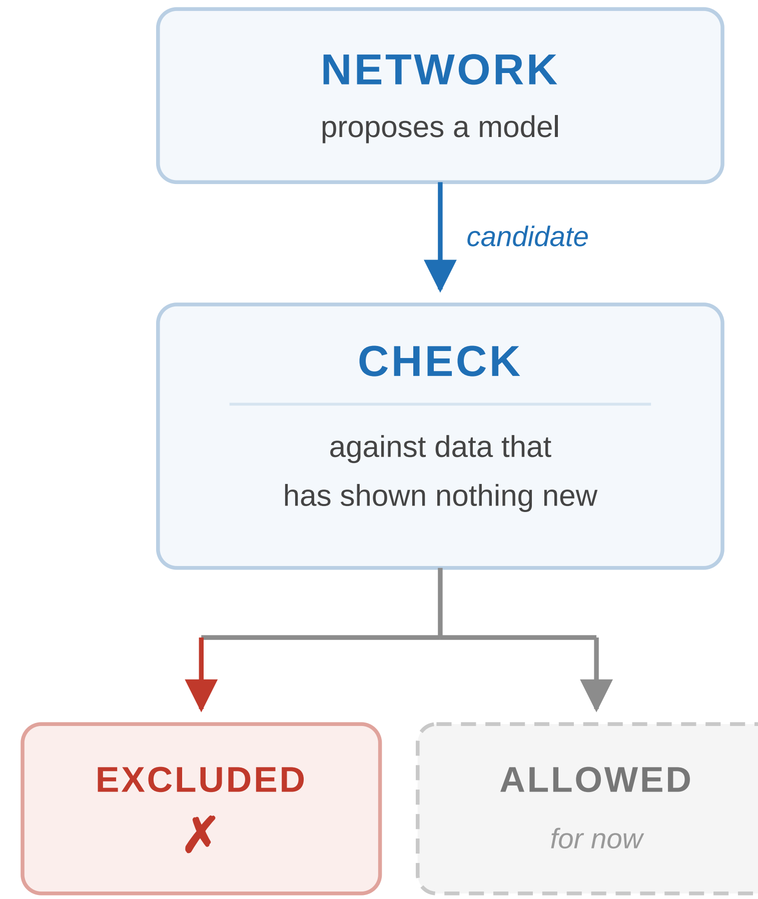
There are far more possible theories than people can write down. Have a robot propose them, keep the ones that pass the known constraints.
Baretz et al., Commun. Phys. (2026) · Hammad & Sanz, arXiv:2608.04100 · Agrawal et al., arXiv:2603.22538

Fitting models to data

Binning the data loses information. Train a network to compare simulation and data event by event, and fit every parameter of the model at once. (Jay's talk)
Brehmer et al., PRD 98 (2018) 052004 · Cranmer et al., PNAS 117 (2020) 30055

Mapping what is still allowed

A model with twenty parameters cannot be scanned point by point. Let a network decide which points to simulate next, so the allowed region is found with a fraction of the simulations.
Caron et al., EPJC 79 (2019) 944 · Goodsell & Joury, EPJC 83 (2023) 268



The check can only say no.

A model is never confirmed, only not yet excluded.

*A fit only tests the model you gave it.
Anything outside it
is invisible.*

What makes this different: *Data can rule a model out but never prove it right. And a fit only tests the model you gave it. Anything outside that model is invisible.*

An agent does the model building

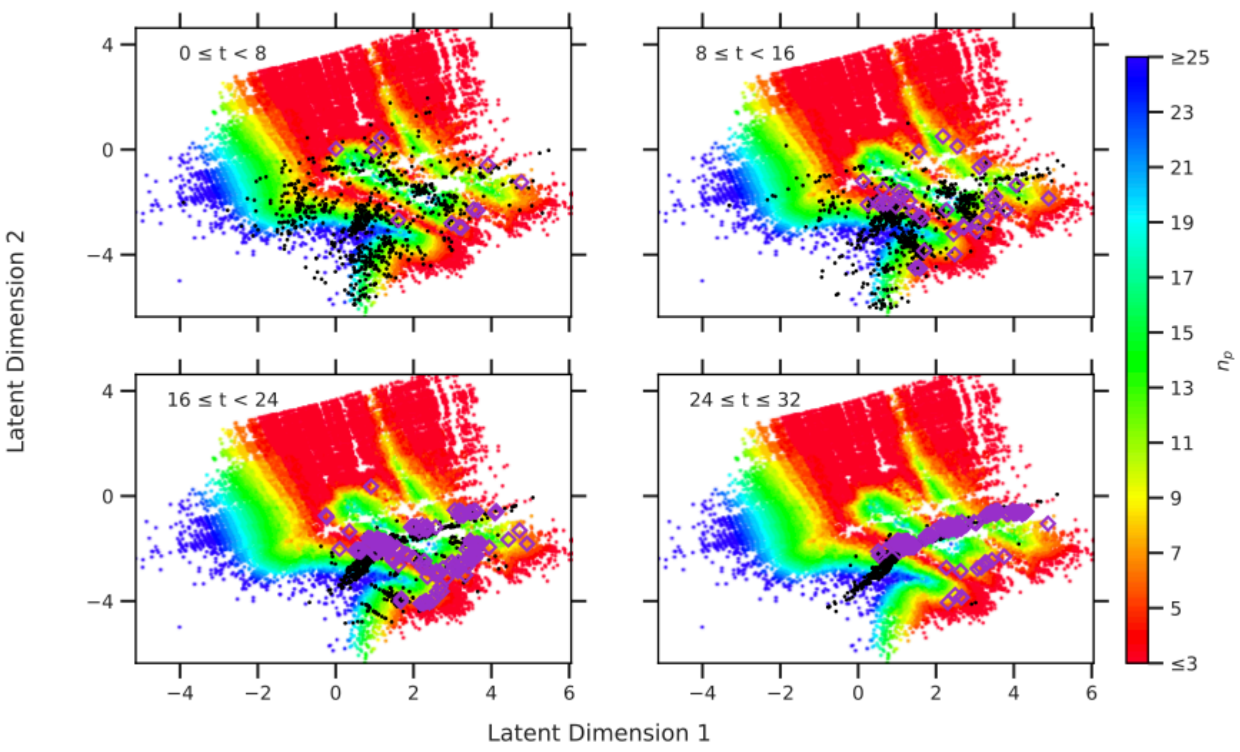
The setup. Neutrinos need mass. Add heavy partners:

$$M_\nu = -\,M_D^T\,M_M^{-1}\,M_D$$

54 parameters, seven measured. A flavor symmetry relates most of them. A *model* is the symmetry, the representations, and the scalars that break it. Discrete, no gradient.

The resolution. A reinforcement learning agent edits one choice at a time. Reward: fits the data with few parameters. **4 million** models per run, ten CPU-hours.

6341 *models that fit the data with at most seven parameters, all released*



Background colour is the number of free parameters, red few and blue many. Early it wanders. Late it finds solutions where the parameters are few.

It rediscovered the known models, then found a four-parameter one under a symmetry nobody had tried.

No scalar potential, no RG running, two mixing angles off. A genetic algorithm with no training has since found 100,000 viable models, faster.

Baretz et al., Commun. Phys. (2026), arXiv:2506.08080 · Crispim Romão & King, arXiv:2606.20000

Fitting new physics to every event

The problem. Wilson coefficients θ to fit from LHC events:

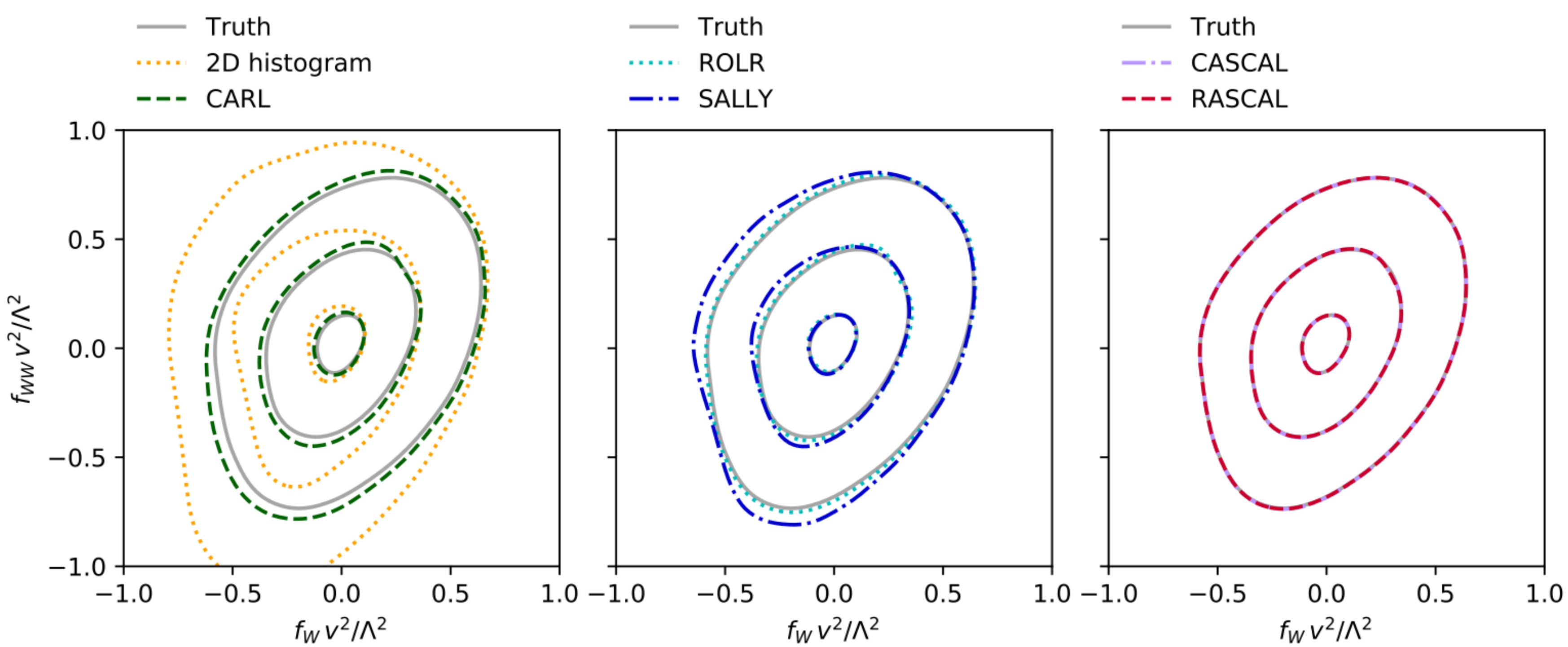
$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \frac{f_W}{\Lambda^2}\,\mathcal{O}_W + \frac{f_{WW}}{\Lambda^2}\,\mathcal{O}_{WW}, \qquad \theta = \left(\frac{f_W v^2}{\Lambda^2}, \frac{f_{WW} v^2}{\Lambda^2}\right)$$

p(x|θ) integrates over shower, hadronization, detector. Simulable, not evaluable. So bin two variables, discard the rest.

The solution. θ lives in the matrix element only. The parton-level ratio is exact and the rest never needs evaluating. A network regressed on it converges to the true ratio.

$$p(x|\theta) = \int dz\; p(x|z)\, p(z|\theta)$$

z: parton momenta. x: reconstructed. p(x|z): blind to θ.



Inference is then optimal. WBF Higgs, idealized detector. Network contours sit on the true ones; the two-variable histogram is much weaker.

The catch. The shower and detector never get evaluated, but the samples carry them. Exact for your simulation, and mismodeling is inherited in full.

Brehmer, Cranmer, Louppe, Pavez, PRD 98 (2018) 052004 · PRL 121 (2018) 111801

Summary

Amplitudes
the check is exact

A transformer guesses, an identity checks. A wrong guess costs you compute and nothing else.

Precision QCD
the check is partial

Precision fixes a few numbers. Reweighting puts them on the whole event. The shower supplies the rest.

BSM
the check only says no

Agents write models, networks fit them. The data rules models out. It has not yet pointed to one.

WHERE EACH IS GOING

Amplitudes. We still choose the letters, the basis, the constraints. Those get learned next. We decide what to compute and read the answer.

Precision QCD. Every calculation folds into the shower as it appears, uncertainty included. Information theory says what each one taught it and which to do next.

BSM. Propose, fit, map become one loop. We decide what a good model is. The loop still only excludes, until something is found.

Back-up

Anatomy of a high-energy collision

Standard perturbative QFT expansion

Radiative effects due to **large energy differences** between collision point and detector

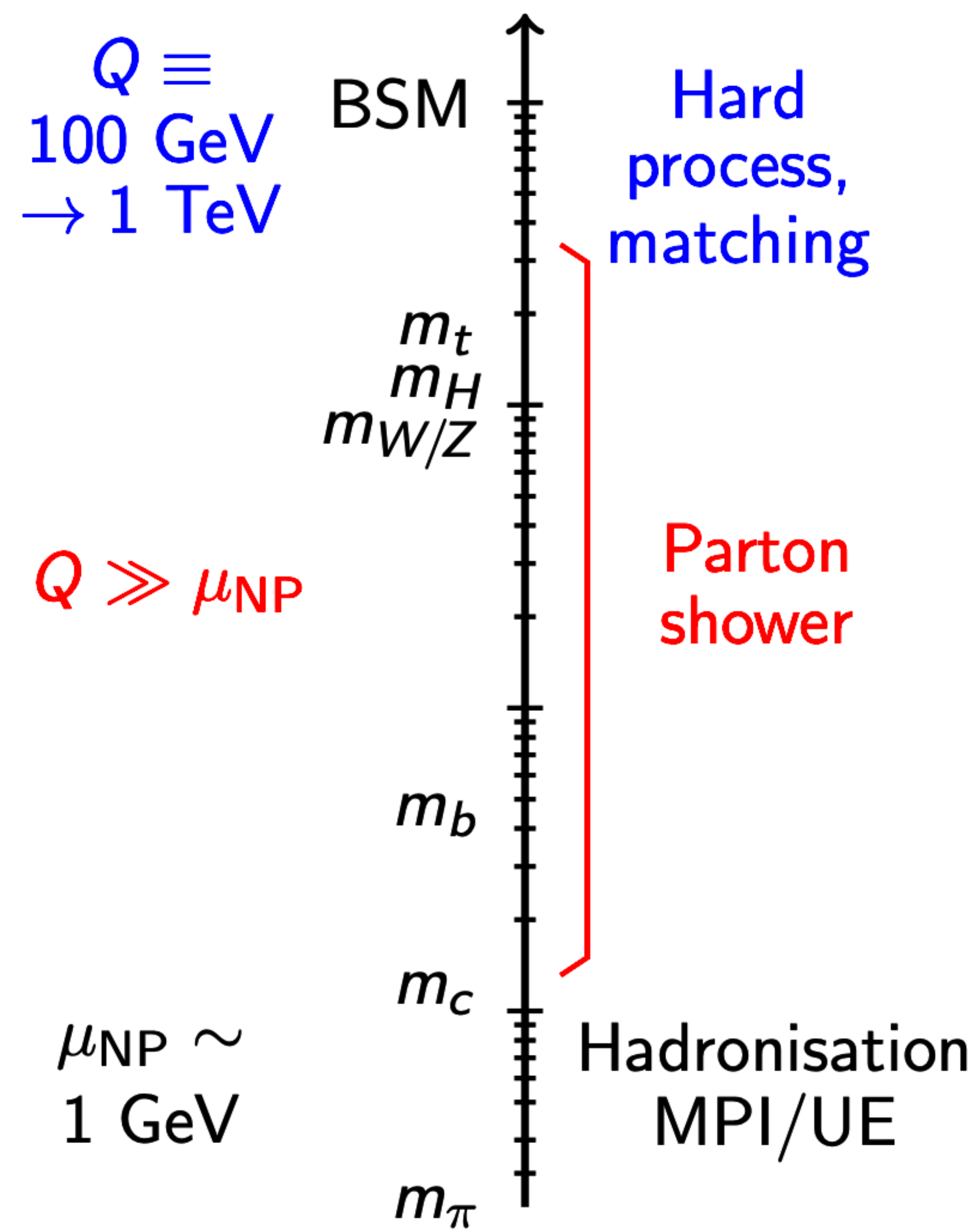
Expect logarithms between disparate scales:

$$\alpha_s \log^2(Q/\mu_{\mathrm{NP}}), \alpha_s \log(Q/\mu_{\mathrm{NP}}), \ldots$$

LL NLL ...

Large logs to **re-sum** analytically or implicitly by shower

Non-perturbative (requires modelling)



Parton Shower

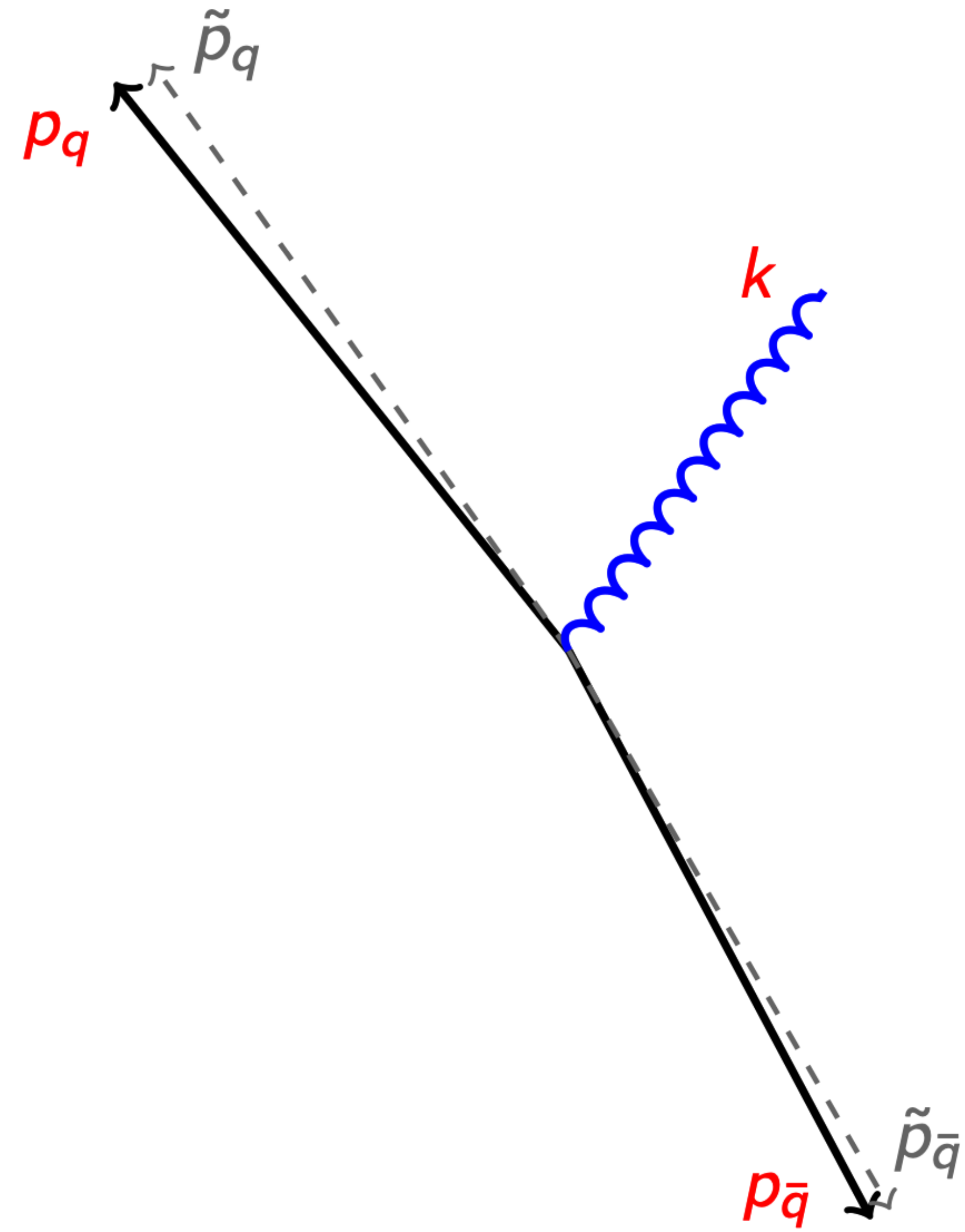
Mechanism: Gluon emission ↔ dipole splitting at large N_c

Branching requires going from a more massive to a less massive state

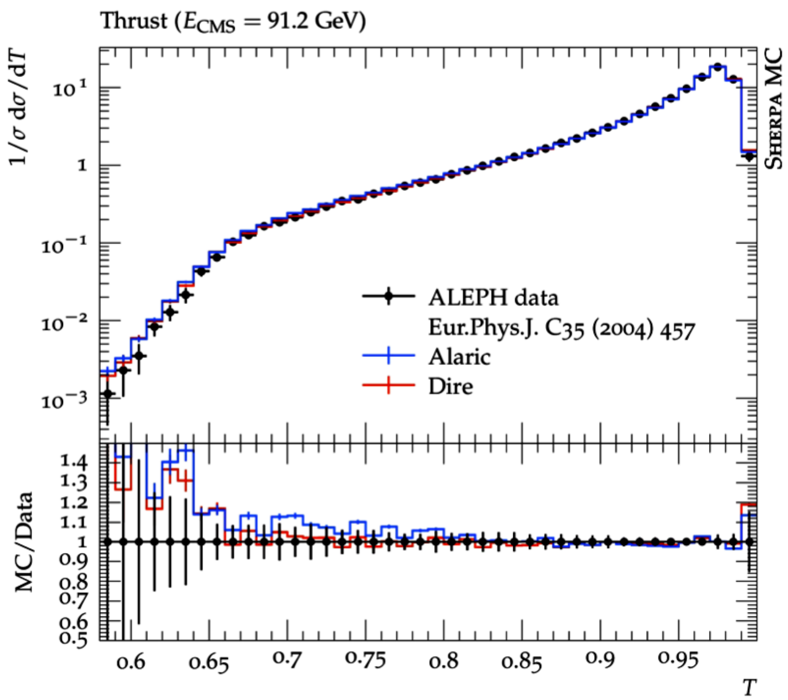
Energy is taken to boost the mass but should not affect the topology of the event

Parton masses must **stay the same** before and after branching

Careful **momentum mapping** needed from pre- to post-branching

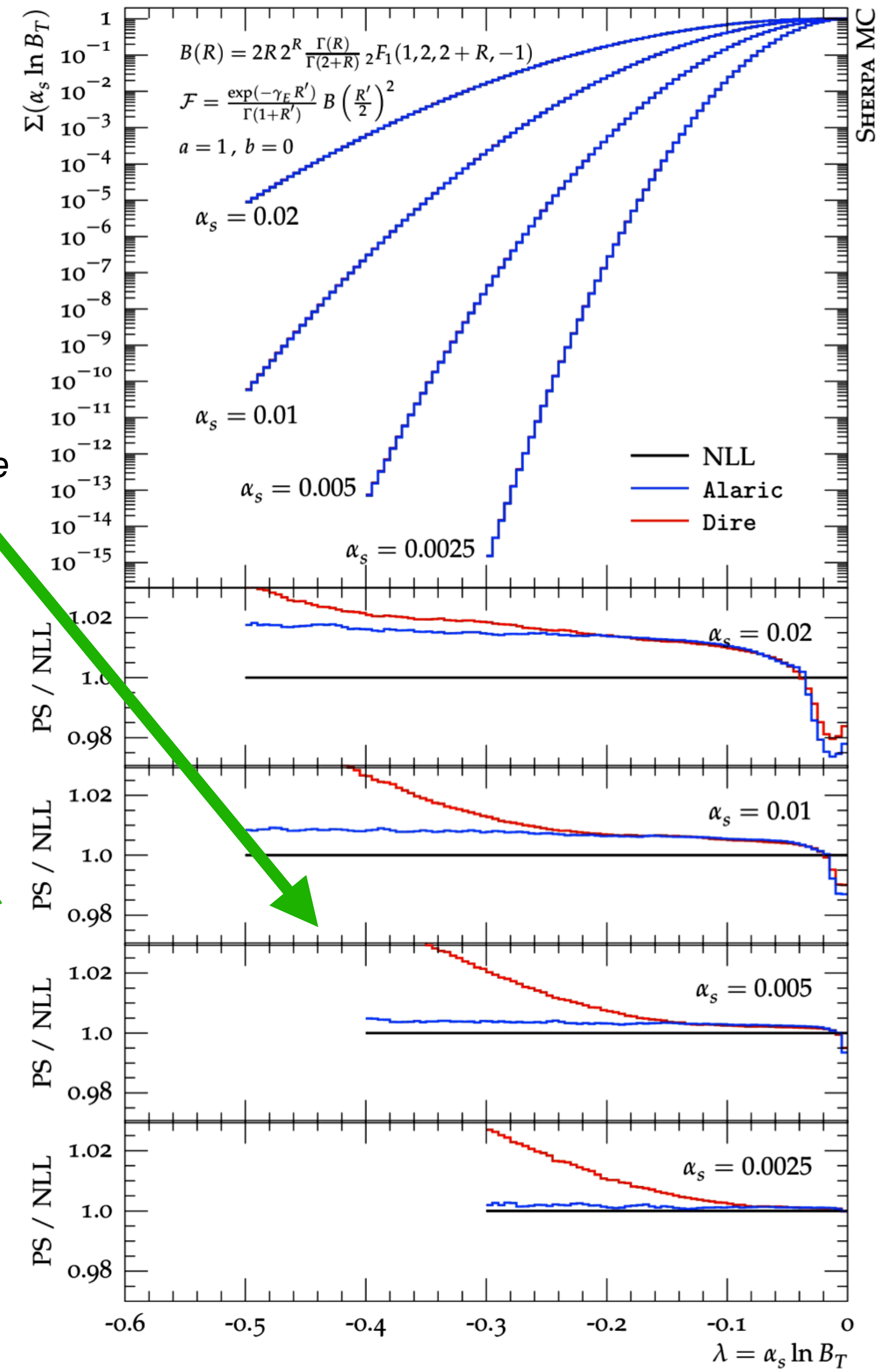


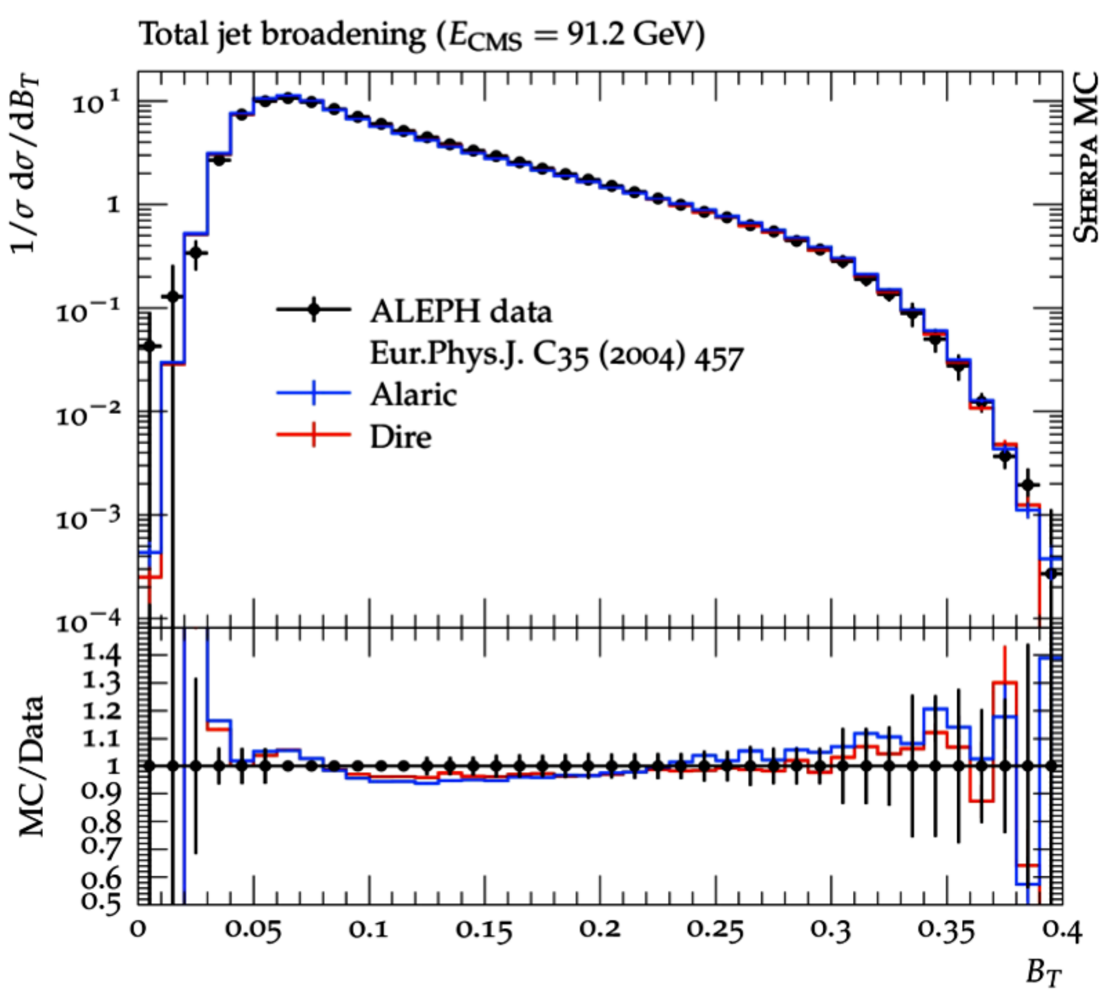
Numerical test and data



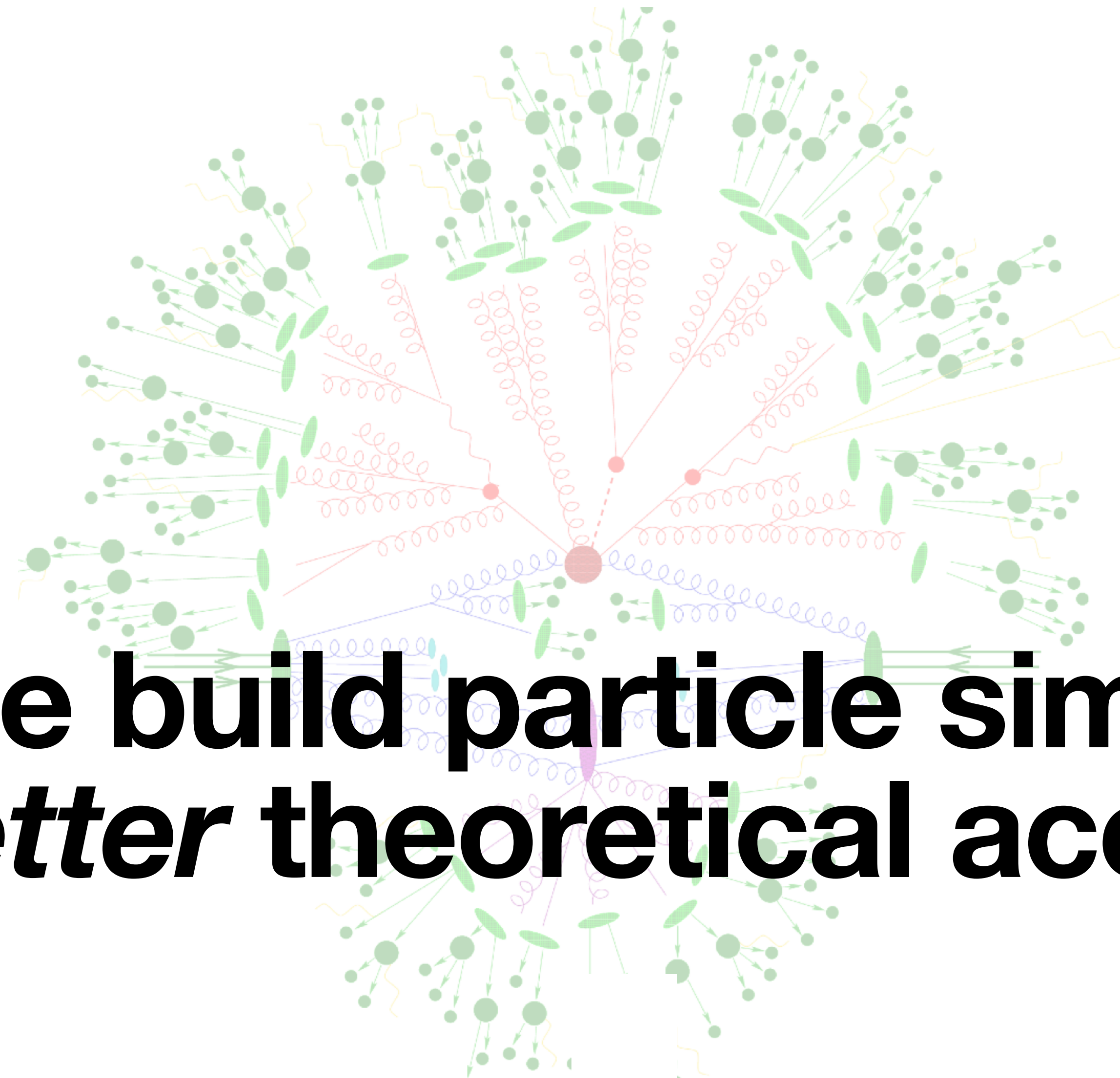
For some observables like thrust ***both*** *ALARIC & standard showers NLL accurate*

For others like broadening ***only ALARIC*** *is - sensitive to transverse momentum (soft) recoil effects*

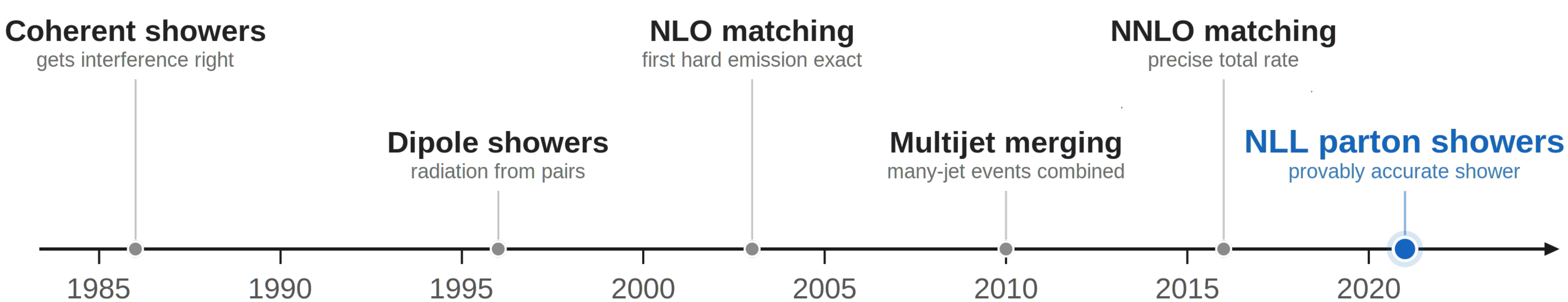




Note: *Perturbative region right - deviations for broadening and a-planarity*



I. Can we build particle simulations with *better* theoretical accuracy?



What makes a parton shower accurate

Emissions factorize in QCD (Markovian) so probabilities independent (multiply)

$$\Delta = \prod_i (1 - dp_i) = e^{-\sum_i dp_i} = e^{-\int P\, d\Phi}$$

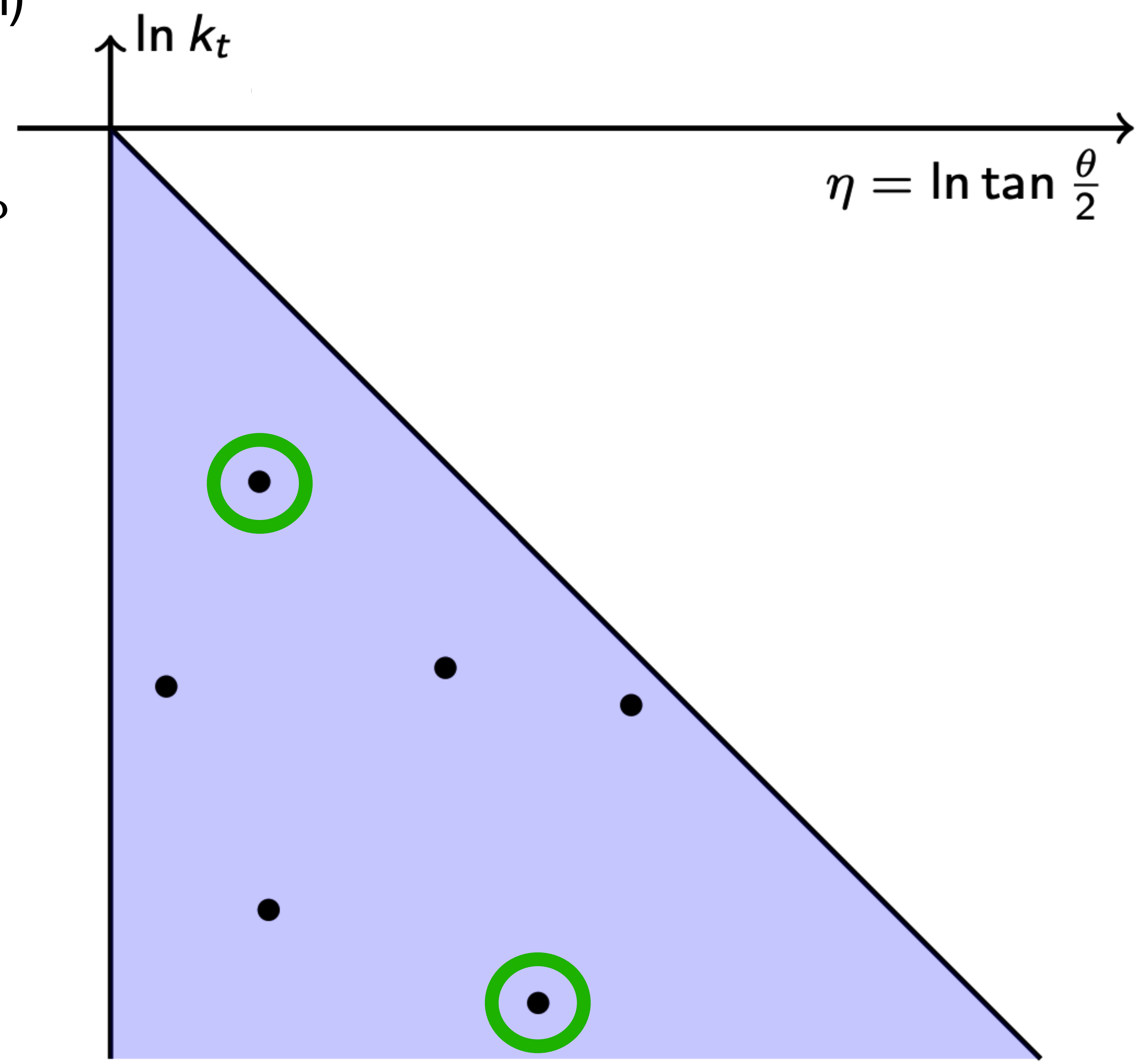
like in nuclear decay

Accuracy = to what order we know the exponent:

$$-\ln \Delta = \int P\, d\Phi \sim \alpha_s L^2 + \alpha_s L + \cdots$$

- **LL:** fill the plane evenly → Area = $\alpha_s L^2$
- **NLL:** emissions widely-separated on Lund plane shouldn't affect each other → $\alpha_s L$

Banfi, Salam, Zanderighi JHEP 03 (2005)



Lund plane

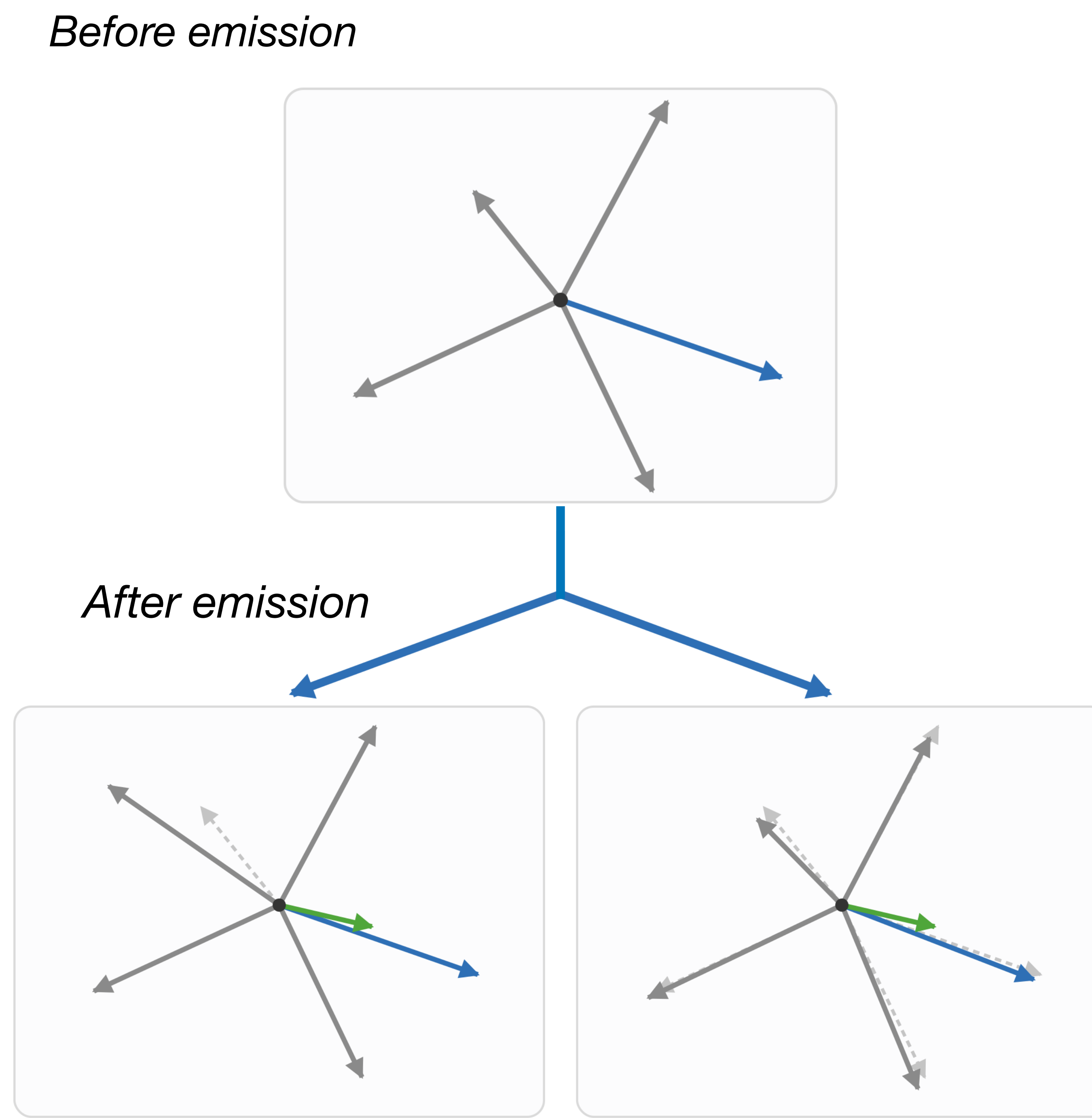
A provably accurate algorithm

Every **emission** **puts** **emitter** **off-shell** — *restoring on-shell* with *4-momentum conserved* forces another parton to recoil

Choice isn't unique — local recoil breaks factorization, exponential is lost

Global recoil preserves geometry — *provably NLL*, and extended to massive partons

Herren, Höche et al. JHEP 10 (2023)
BA and Höche PRD 109 (2024)



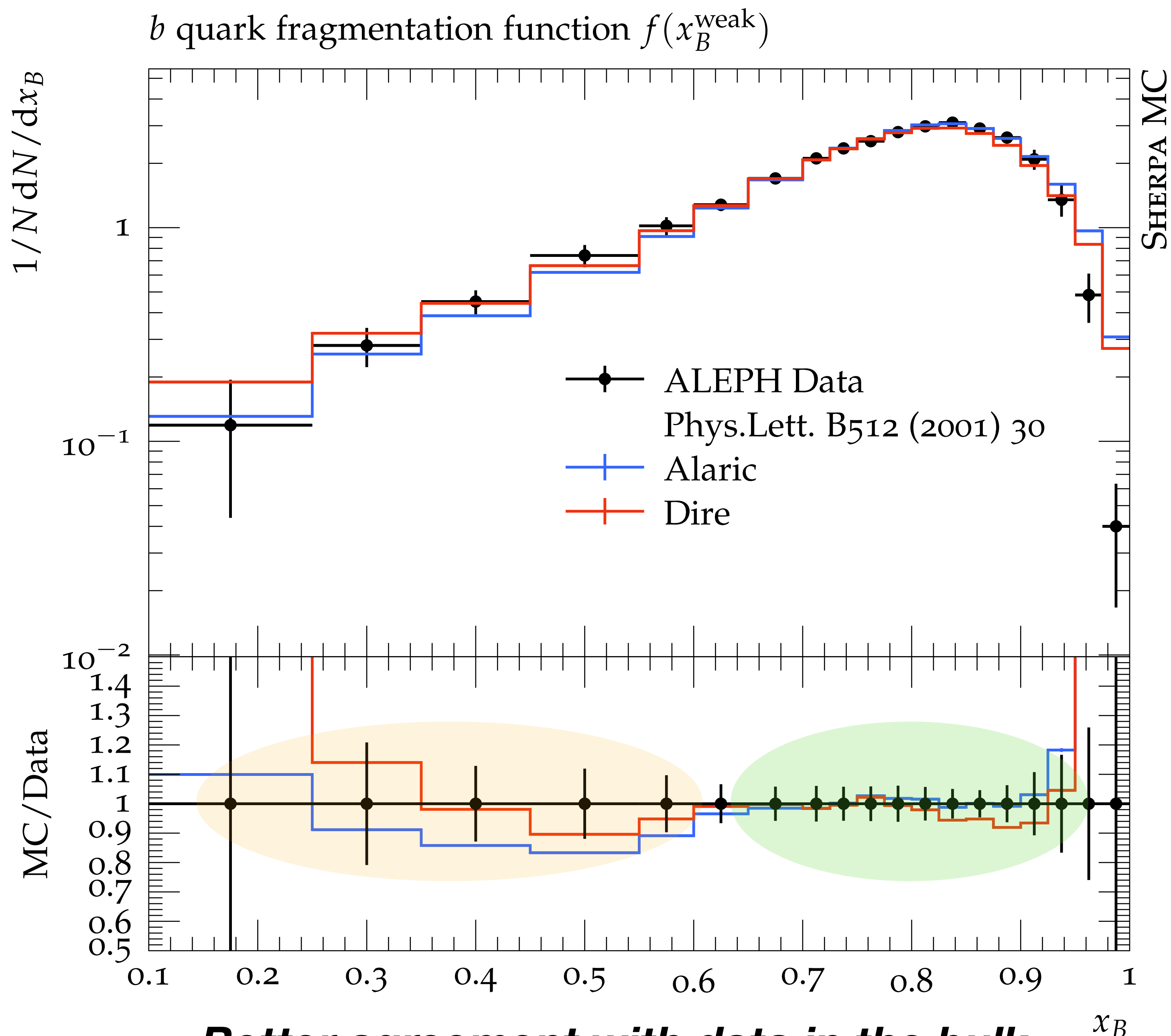
Local recoil (Standard)

one parton absorbs it all → emissions no longer independent

Global recoil (ALARIC)

Each barely moves → stays independent

Massive Improvements



Better agreement with data in the bulk

Newly released in one of the LHC's major event generators
SHERPA 3.0, NLO matched, both massive and massless

BA and Höche PRD 109 (2024)