Jets and Jet Substructure

Matthew Schwartz
Harvard University
Aug 10, 2023

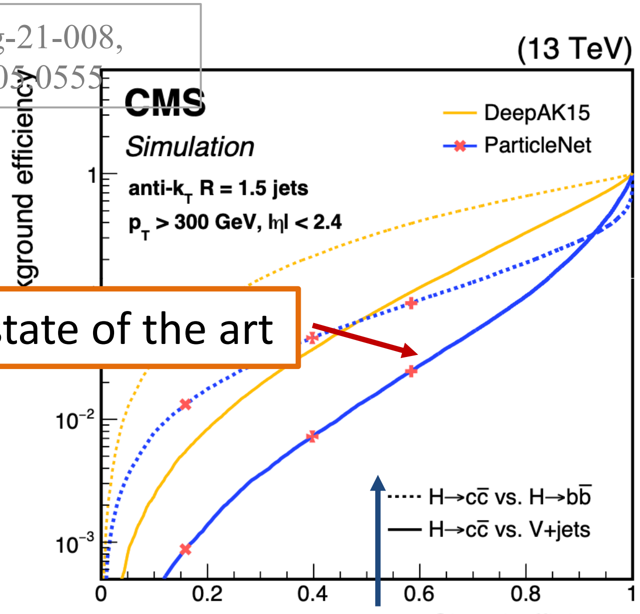
Recent progress in Jets

- Experimental progress
  - Improved calibration/modeling for run 3
  - Machine Learning methods expanding reach
  - Measuring many new observables

- Theory progress
  - Machine learning for tagging, measurement, searches
  - Resummed and fixed order calculations
  - Improvements in jet mass predictions
  - Energy-energy correlators
  - Top mass measurement improments

Charm tagging

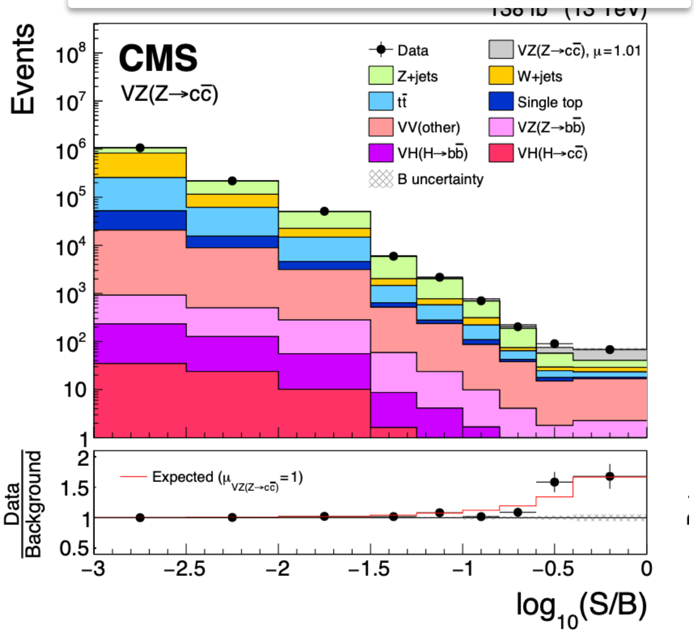
CMS-Hig-21-008,
arXiv:2205.0555



previous state of the art

“particle net”
ML algorithm

observation of Z->cc at 5.7 σ



ratio to SM expecation

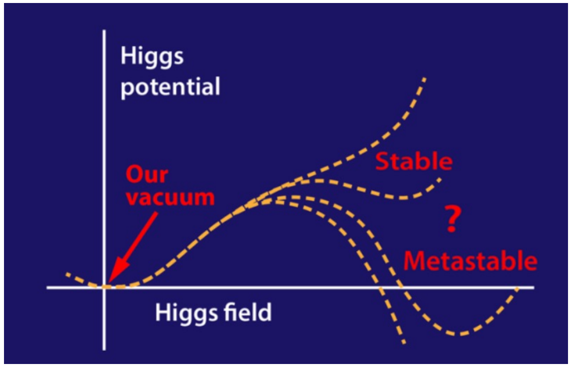
$$\mu_{\mathrm{VZ(Z\to c\bar{c})}} = 1.01^{+0.23}_{-0.21}$$

Particle net (arXiv:1902.08570)

- Graph neural network
- takes momenta of particles
- Includes tracks, particle id, etc

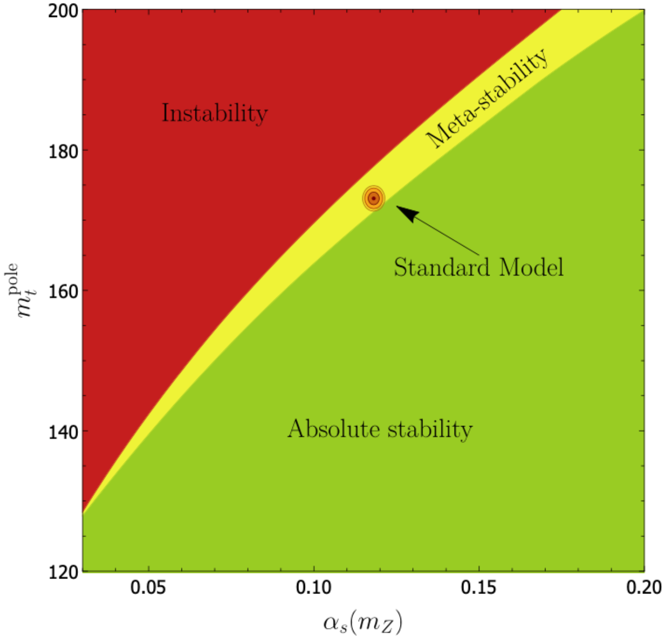
- Higgs to cc is now possible
  - Current limit is 14 x Standard Model
  - Statistics limited
  - Could be seen in run 3

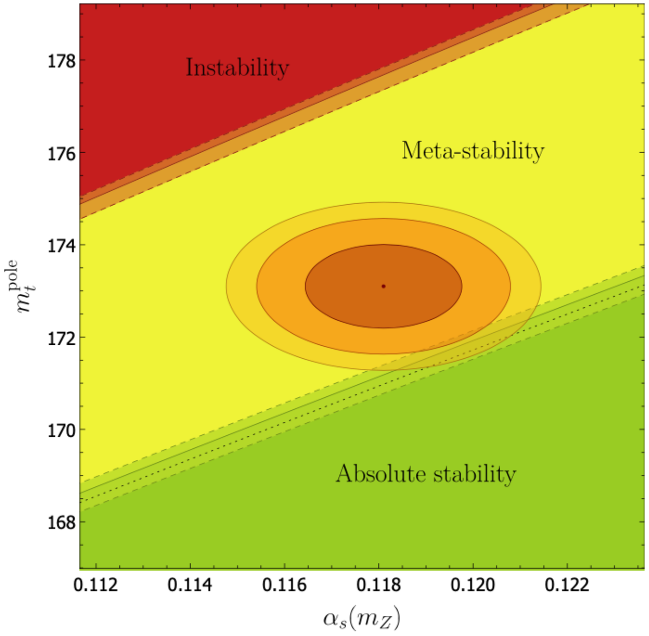
Standard Model measurements are important!



Need to improve measurements of top quark mass and α_s

- Lifetime of the universe depends on their values
- Important measurements within LHC reach



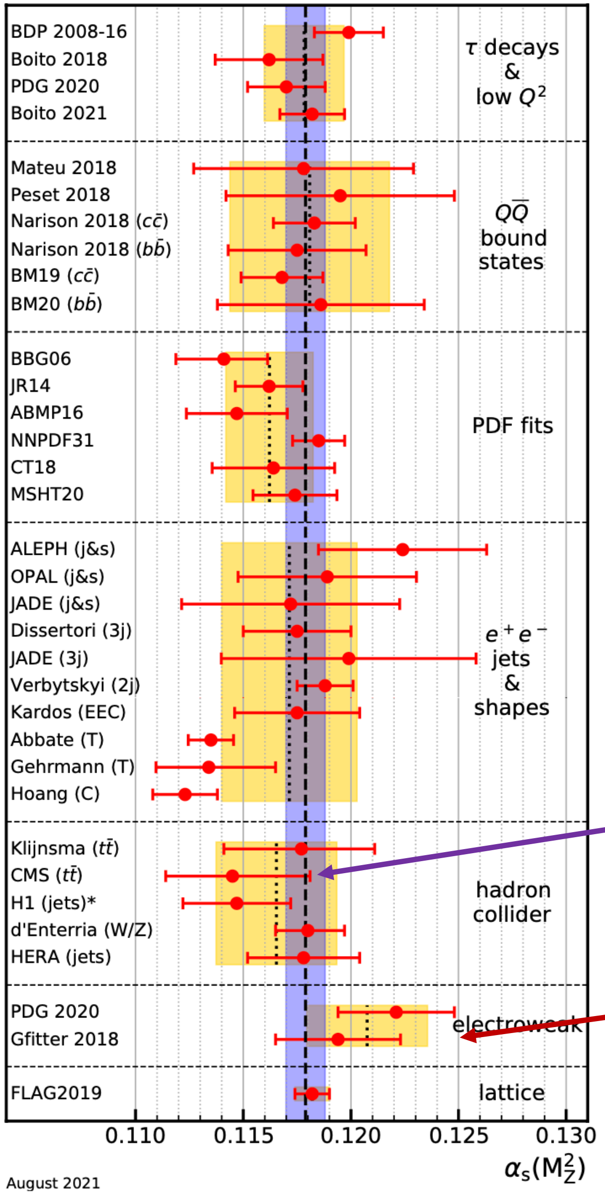


Andreassen, Frost, MDS arXiv:1707.08124

Status of α_s, PDG 2021

$$\alpha_s(M_Z^2) = 0.1179 \pm 0.0009$$

0.8% uncertainty



Energy-energy correlators at e+e-

e+e- event shapes (thrust, C parmater)

Only LHC measurement included in 2021 PDG average

CMS tt total cross section

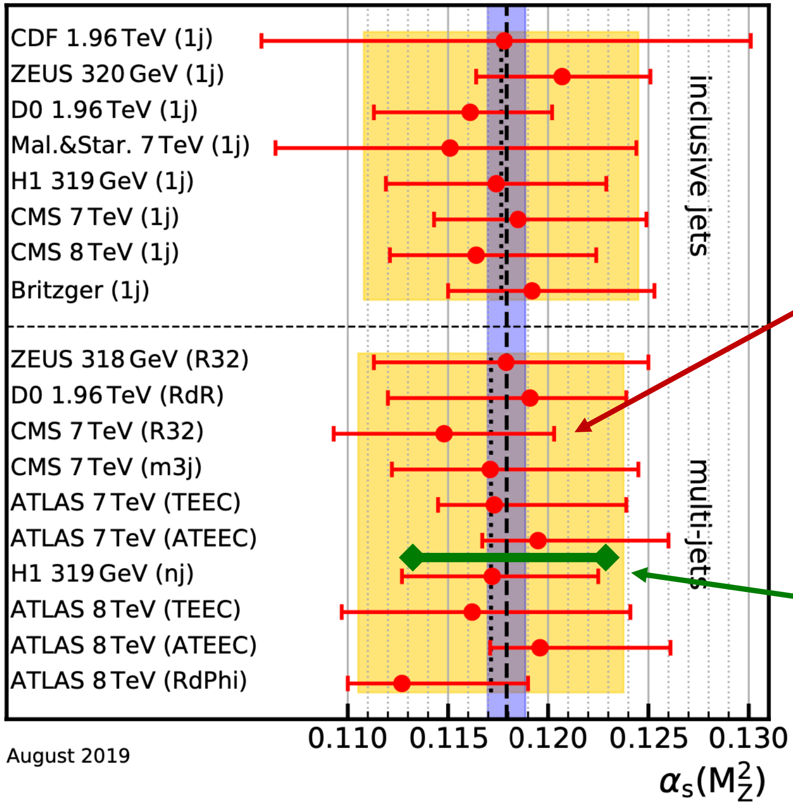
$$\alpha_s(M_Z^2) = 0.1145^{+0.0036}_{-0.0031}.$$

Dominant measurements from Flavor Lattice Averaging Group

$$\alpha_s(M_Z^2) = 0.1182 \pm 0.0008\,;$$

Hadron collider jet measurements

PDG 2021: inclusive and multi-jet measurments



CMS arXiv:1304.7498

- 3 to 2 jet cross section ratios (R32)
- 5 fb^{-1} @ 7 TeV CMS

$$\alpha_S(M_Z) = 0.1148 \pm 0.0014\,(\text{exp.}) \pm 0.0018\,(\text{PDF}) \pm 0.0050\,(\text{theory})$$

$$= 0.1148 \pm 0.0055$$

5% uncertainty

ATLAS arXiv:2301.09351

- Transverse energy-energy correlators (TEECs)
- 139 fb^{-1} @ 13 TeV ATLAS

$$\alpha_s(m_Z) = 0.1175 \pm 0.0006\,(\text{exp.})^{+0.0034}_{-0.0017}\,(\text{theo.})$$

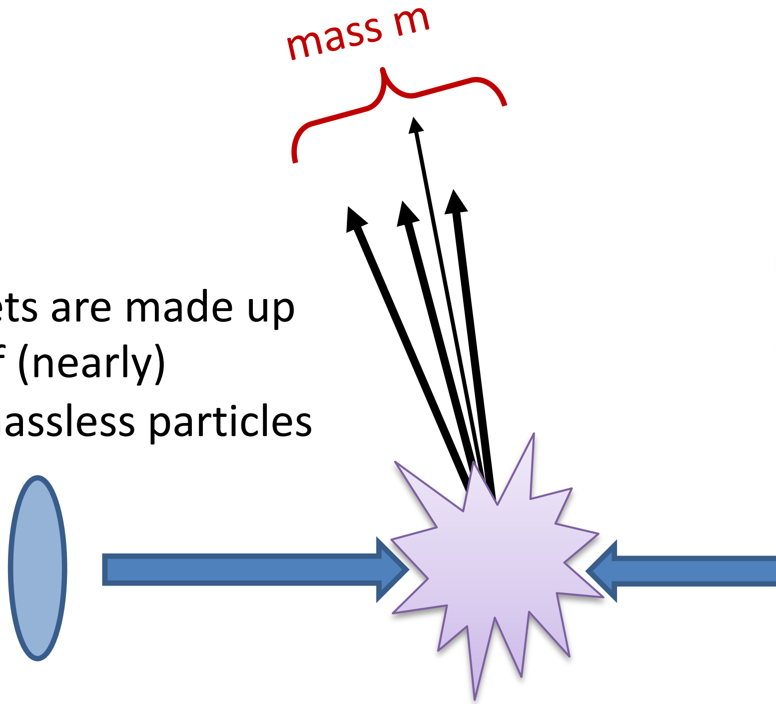
3% uncertainty

Can we get down to the 1% level with jets?

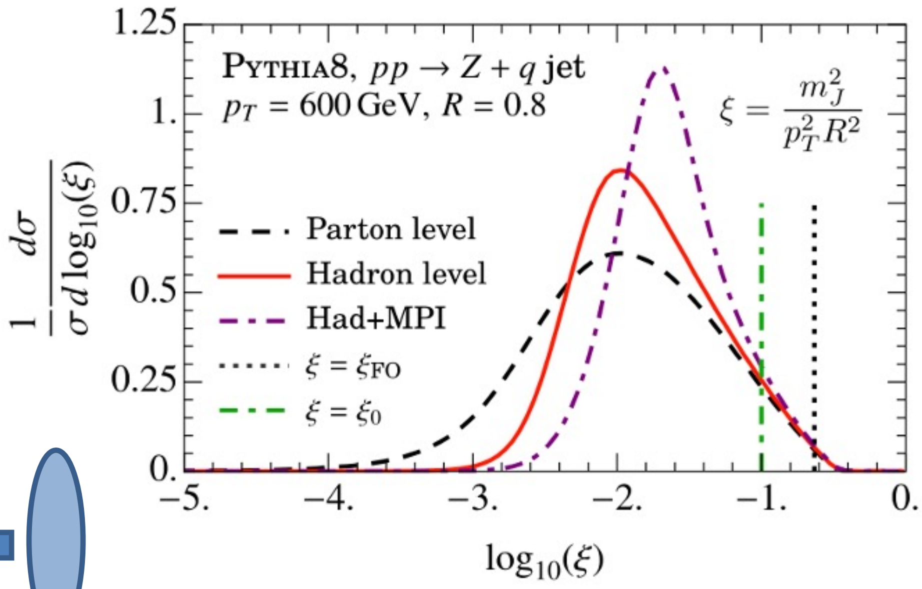
Jet mass measurments

Can we calculate the jet-mass distrubution from first principles?

- Must avoid MC as much as possible to measure α_s



Jets are made up of (nearly) massless particles



Very challenging theory calculation

- Mass is senstitive to many things under poor theoretical control
  - Underlying event, pileup, hadronization corrections, etc.

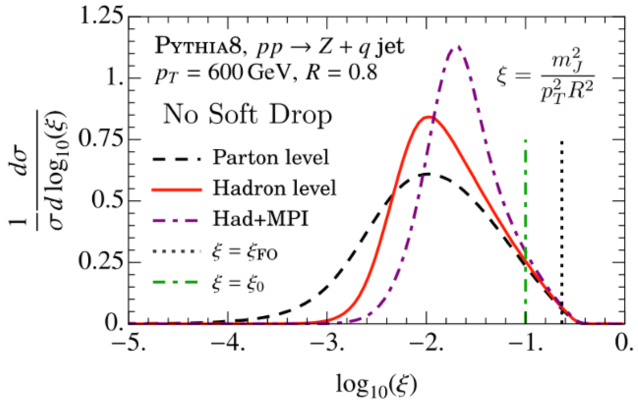
Soft drop

- Removes soft (low energy) particles from a jet in a systematic way

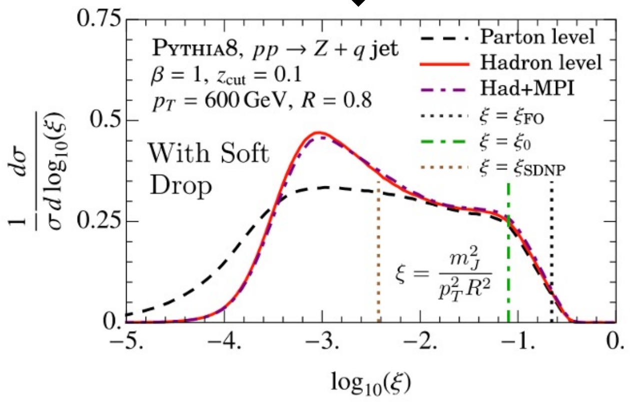
- Undo the clustring, starting from small angles
- Drop a particle if it is soft, meaning

$$\frac{p_{Ta}}{p_{Ta}+p_{Tb}} < z_{\rm cut}\left(\frac{\theta_{ab}}{R}\right)^{\beta}$$

- If neither particle is soft at a given step,
  stop declustering and return the soft-drop jet



apply soft drop



After soft drop

- Effect of MPI (underlying event) and pileup are tiny
- Region exists where hadronization corrections are small

Soft-drop jet mass offers potential for α_s measurement at the LHC

Factorization formula

$$\frac{d^3\sigma}{dp_T\,d\eta\,d\xi} = \sum_{abc}\int \frac{dx_a\,dx_b\,dz}{x_a x_b z} f_a(x_a,\mu) f_b(x_b,\mu) H^c_{ab}\left(x_a, x_b, \eta, \frac{p_T}{z}, \mu\right) \mathcal{G}_c(z, \xi, p_T, R, \mu)$$

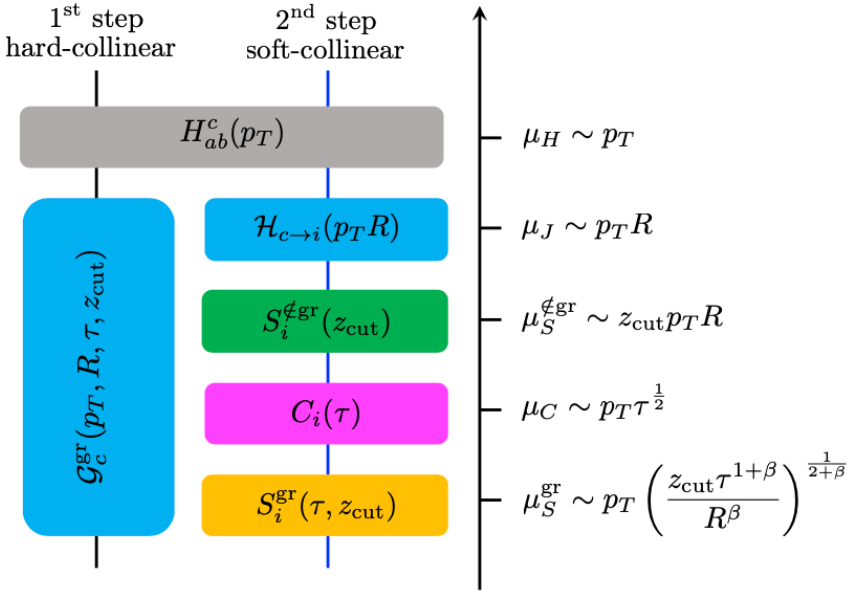
PDFs

Hard function

jet function

$$\xi = \frac{m_J^2}{p_T^2 R^2}$$

soft-drop jet mass



Physics at many scales relevant

- Jet mass, jet energy (pT)
- Collinear scale, soft scale, soft-collinear scale
- Soft drop cutoff scale (z τ)
- …

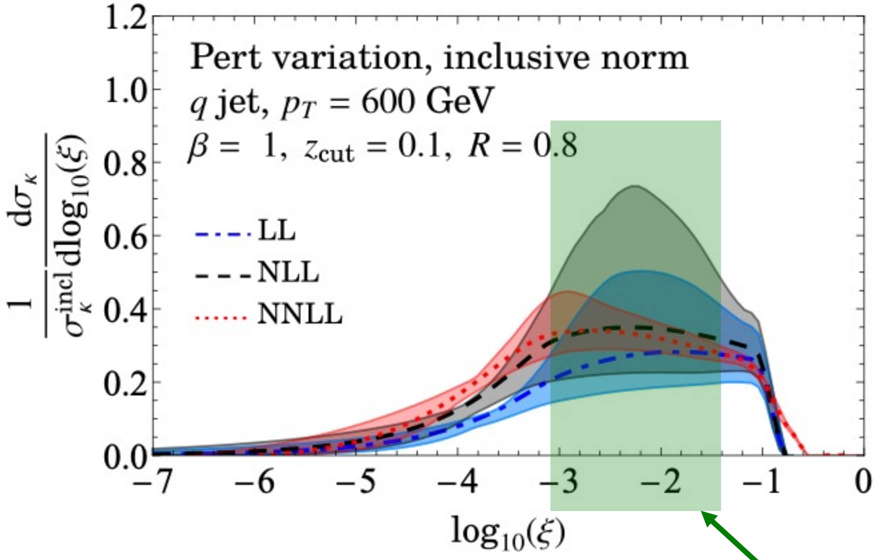
Although complicated, we can still understand it

- Frye, Larkoski, MDS, Yan 1603.09338 factorization
- Marzani et al. 1704.02210 power corrections
- Stewart, Hannesdottir, Pathak, MDS, Stewart 2210.04901
  - NNLL resummation with power corrections

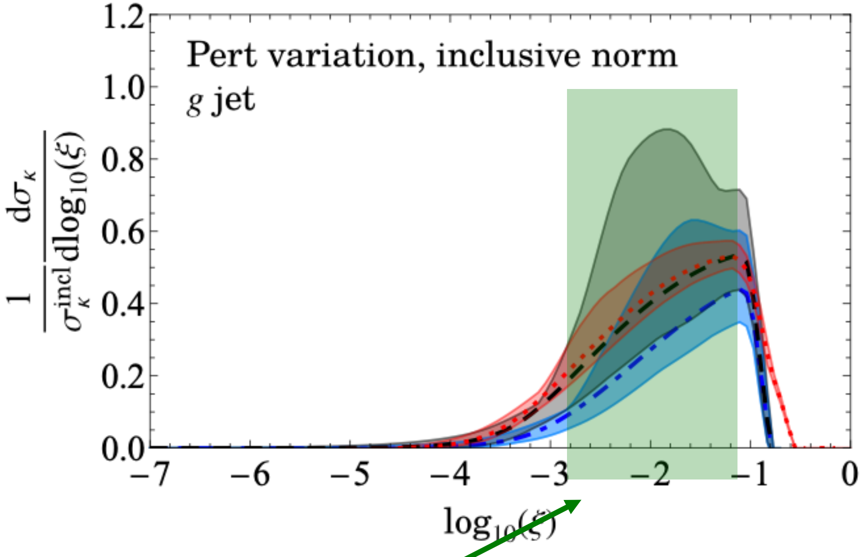
Pertubative Uncertainties

Stewart, Hannesdottir, Pathak, MDS, Stewart
arXiv:2210.04901

quarks



gluons



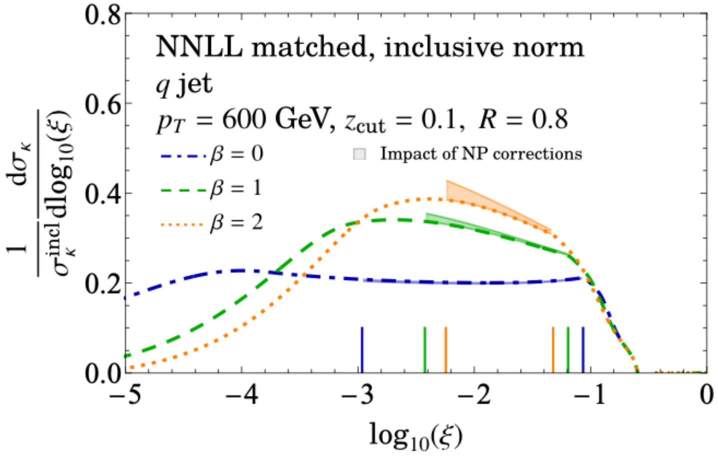
- Good perturbative control in fit region ($-3 < \log10\ \xi < -1$)
- Good convergence from LL -> NLL -> NNLL

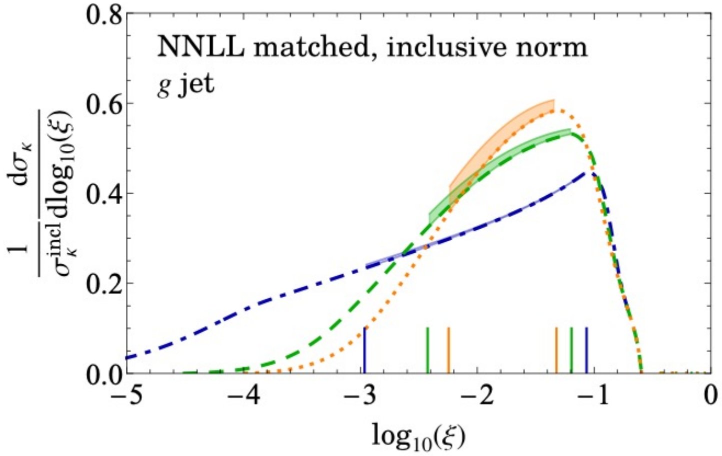
Non-pertubative Uncertainties

- Six non-perturbative shape-function parameters
- Central values fit to pythia

$$\Omega^{\circledcirc}_{1,q} = 0.55\,\mathrm{GeV}\,, \qquad \Upsilon^{\circledcirc}_{1,0q} = -0.73\,\mathrm{GeV}\,, \qquad \Upsilon^{\circledcirc}_{1,1q} = 0.90\,\mathrm{GeV}\,, \qquad \text{for quarks,}$$

$$\Omega^{\circledcirc}_{1g} = 0.91\,\mathrm{GeV}\,, \qquad \Upsilon^{\circledcirc}_{1,0g} = -0.24\,\mathrm{GeV}\,, \qquad \Upsilon^{\circledcirc}_{1,0g} = 0.90\,\mathrm{GeV}\,, \qquad \text{for gluons.}$$



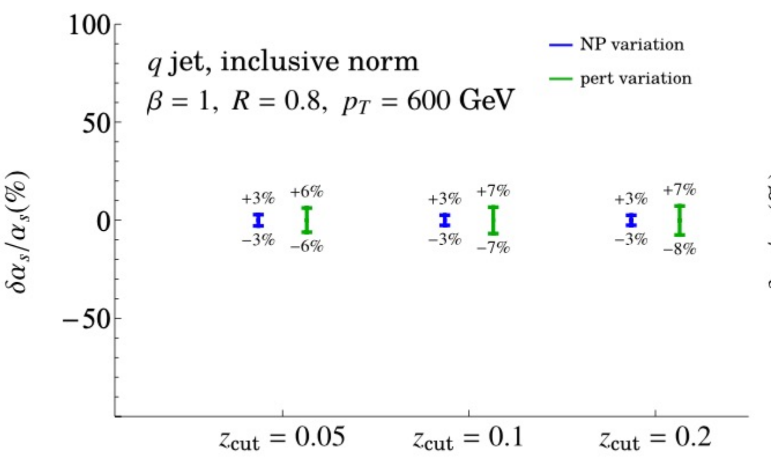


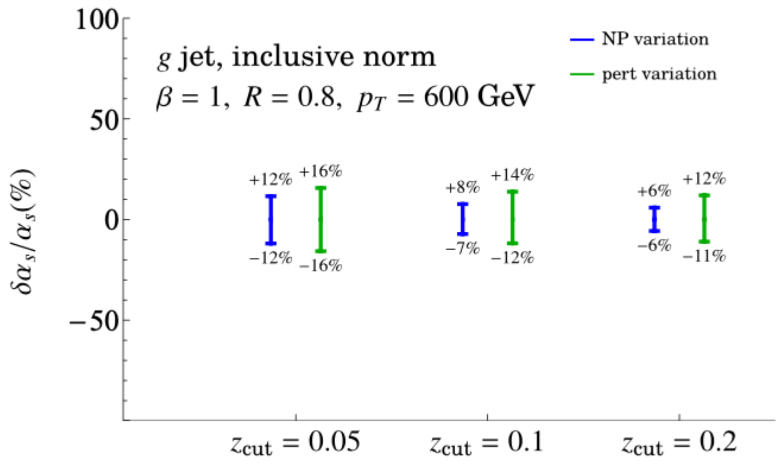
Non-perturbative uncertainty relatively small

- As expected – that's why we're using soft-drop

α_s measurement prospects

Stewart, Hannesdottir, Pathak, MDS, Stewart
arXiv:2210.04901



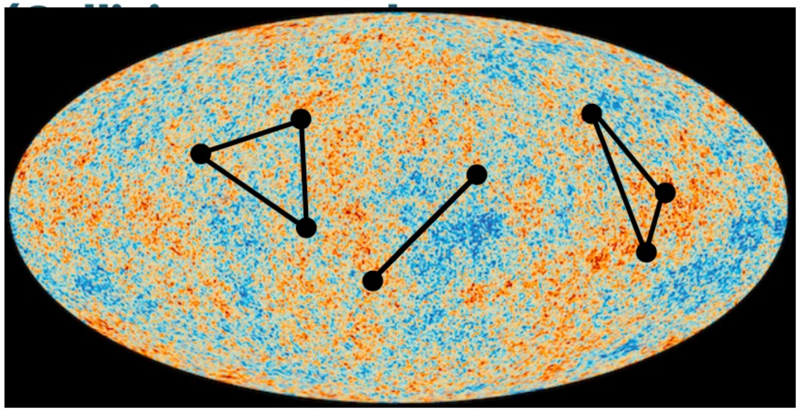


- Should be able to measure α_s at the < 10% level now
- Dominated by perturbative uncertainty
  - Fit range chosen to minimize non-perturbative effects
- Different parameter values and different energies can help reduce overall uncertainty

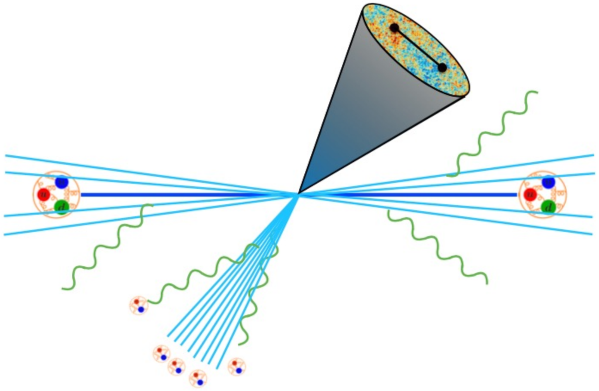
Possible 5% measurement in the future
- Diffuclt to get to < 1% level competitive with world average
- Non-perturbative effects are irreducible below ~ 3%

Energy-energy correlators

An alternative approach to studying jets



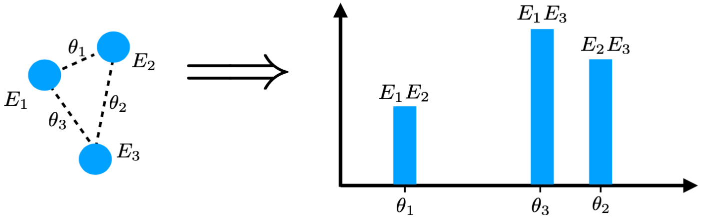
Correlation functions are standard tools in condensed matter and astonomy



Can also measure at colliders

Each event contributes multiple values of the observable

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\theta} = \sum_{i,j}\int d\sigma\,\frac{E_iE_j}{Q^2}\,\delta\left(\theta-\theta_{ij}\right) \sim \left\langle\Psi\left|\mathcal{E}\left(\hat{n}_1\right)\mathcal{E}\left(\hat{n}_2\right)\right|\Psi\right\rangle$$



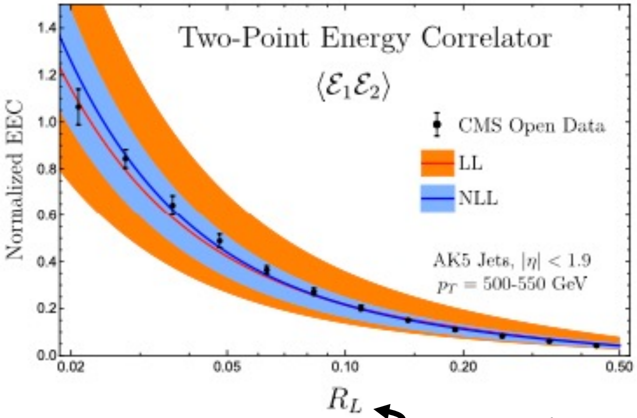
Energy-energy correlators

Lee, Mecaj and Moult arXiv:2205.03414

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}p_T\,\mathrm{d}\eta\,\mathrm{d}\{\zeta\}} = \sum_i \mathcal{H}_i\left(p_T/z,\eta,\mu\right) \otimes \int_0^1 dx\, x^N\, \mathcal{J}_{ij}(z,x,p_T R,\mu)\, J_j^{[N]}(\{\zeta\},x,\mu)\,. \tag{5}$$

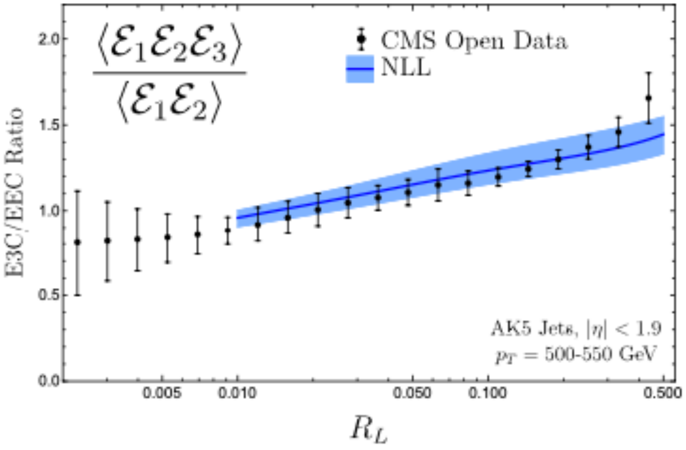
EECs factorize and can be resummed

- 2-point function
- Good agreement with theory and CMS open data

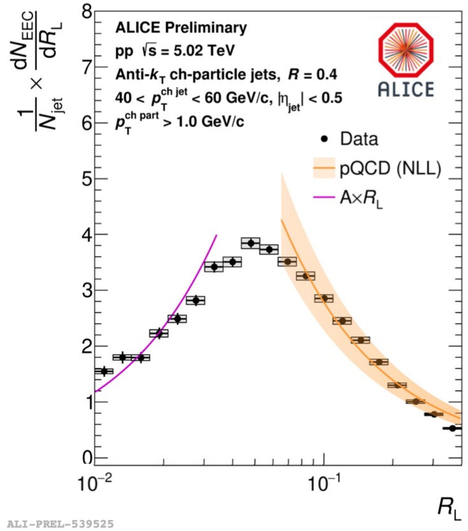


R_L = maximum angle between energy flow operators

3 point function predicted too



ALICE `23



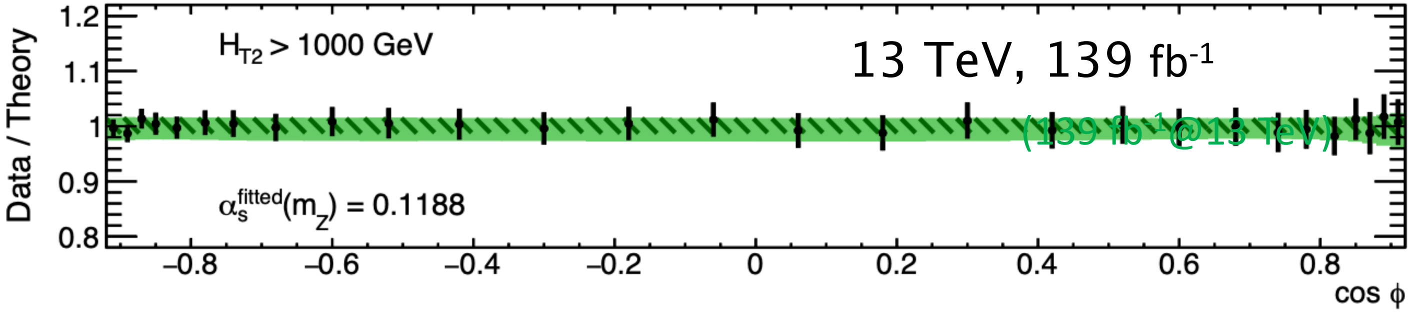
- Measured by ALICE
- good agreement with theory

EECs for α_s

- Transverse energy-energy correlators (EECs)
    - Measures EEC using p_T instead of energy
    - Fit to NNLO theory

$$\frac{1}{N}\sum_{A=1}^{N}\sum_{ij}\frac{E^{A}_{\mathrm{T}i}E^{A}_{\mathrm{T}j}}{\left(\sum_k E^{A}_{\mathrm{T}k}\right)^2}\delta(\cos\phi-\cos\varphi_{ij}),$$

ATLAS arXiv:2301.09351

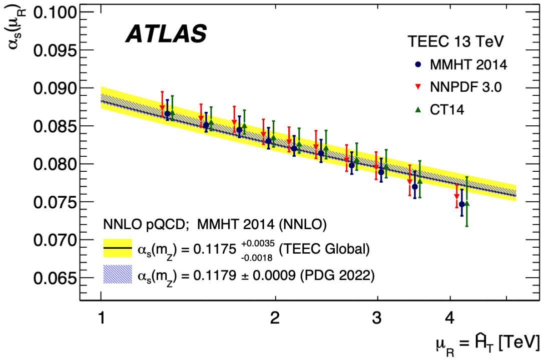


3% uncertainty

$$\alpha_s(m_Z) = 0.1175 \pm 0.0006\ (\mathrm{exp.})^{+0.0034}_{-0.0017}\ (\mathrm{theo.})$$

$$0.1175 \pm 0.0001\ (\mathrm{stat.}) \pm 0.0006\ (\mathrm{sys.})^{+0.0032}_{-0.0011}\ (\mu) \pm 0.0011\ (\mathrm{PDF}) \pm 0.0002\ (\mathrm{NP}) \pm 0.0005\ (\mathrm{mod.})$$

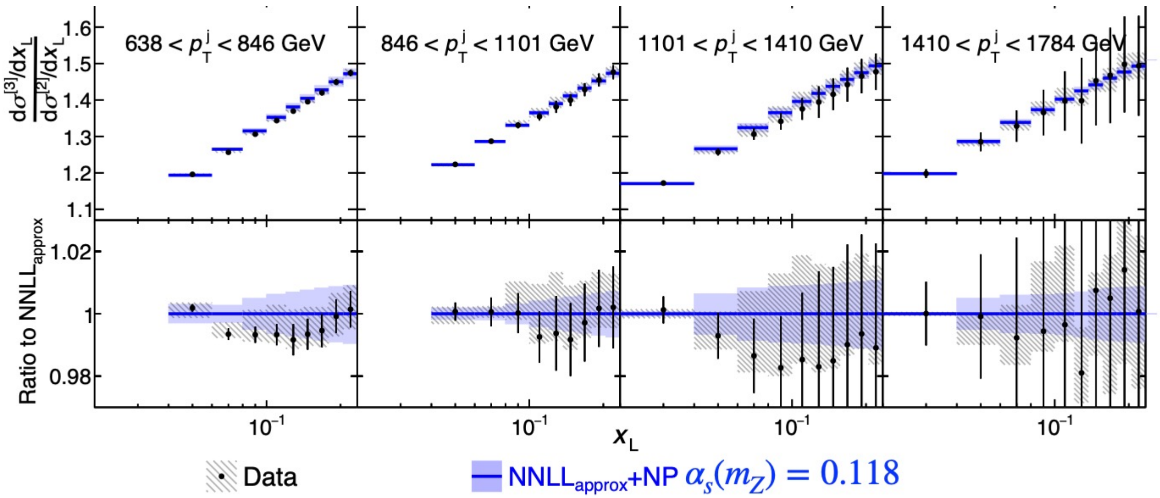
- Does not include any resummation
- Monte carlo used to include non-perturbative effects
- Scale uncertaintities dominate

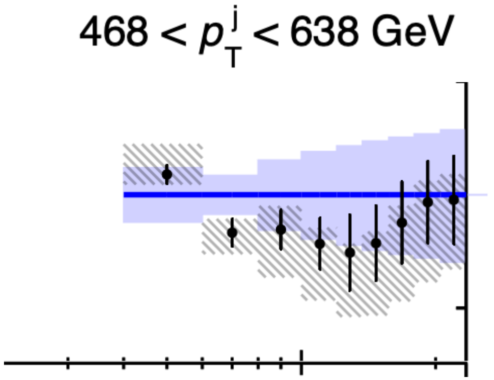


EECs for α_s

Chenfeng Lu CMS, July 31, 2023
(talk at Boost)

- E3C/E2 in high-p_T jets
  - compared to NNLL theory (Chen et al. arXiv:2307.07510)





$$\alpha_s(m_Z) = 0.1229^{+0.0014(stat.)+0.0030(theo.)+0.0023(exp.)}_{-0.0012(stat.)-0.0033(theo.)-0.0036(exp.)}$$

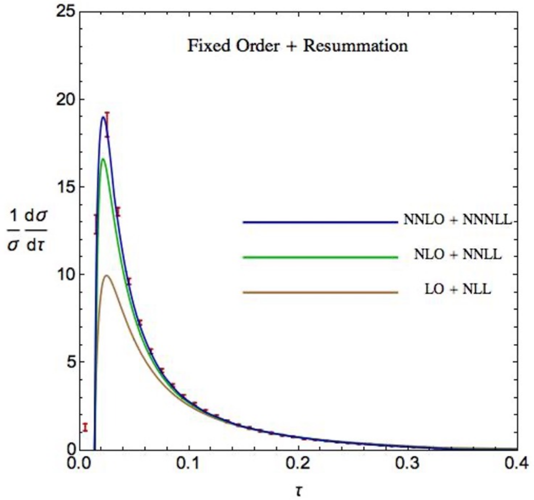
4% uncertainty

- Hadronization taken from average of pythia + herwig (3%)
  - Should use theory model
- Paper not published, hard to assess

e^+e^- event shapes

Thrust, C parameter

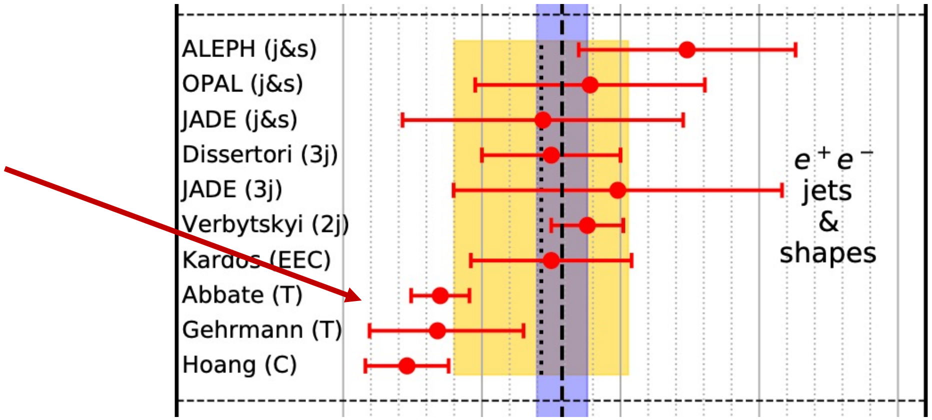
- NNLO fixed order Gerhmann et al, 0711.4711
- NNNLL resummation Becher, MDS 0803.0342.
- Power corrections Abbate et al 1006.0308
- Renormalon subtractions
- heavy quark corrections
- Electroweak corrections



Thrust: $\alpha_s(m_Z) = 0.1135 \pm 0.0011$ 1.0% uncertainty

C parameter: $\alpha_s(M_Z^2) = 0.1123 \pm 0.0015$ 1.3% uncertainty

Values are low compared
to world average

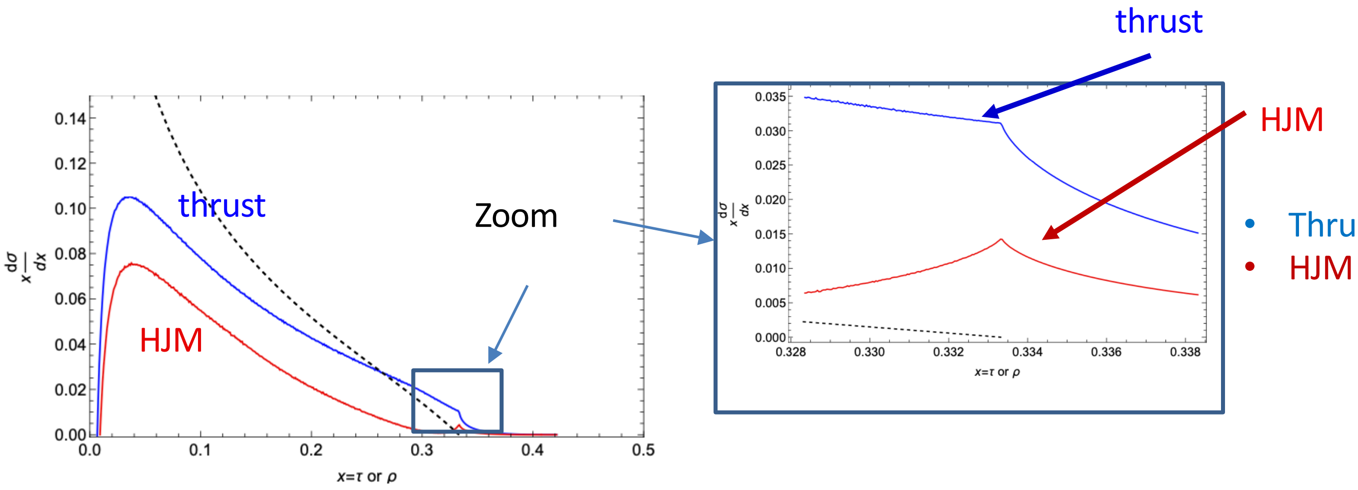


Heavy jet masss

- NNLO fixed order Gerhmann et al, 0711.4711
- NNNLL resummation Chien, MDS 1005.1644
- Salam and Wicke 0102343
    - ”Fits for αs from Heavy Jet Mass come out 10% smaller than for thrust”
- Constency with other event shapes needed to **validate** methodology

Heavy jet mass is qualitatively different from other event shapes

- Differs from thrust and C parameter in the 3-jet region
- Has a left “**Sudakov Shoulder**” Catani and Webber 9710333



- Thrust has right shoulder only
- HJM has left and right shoulders

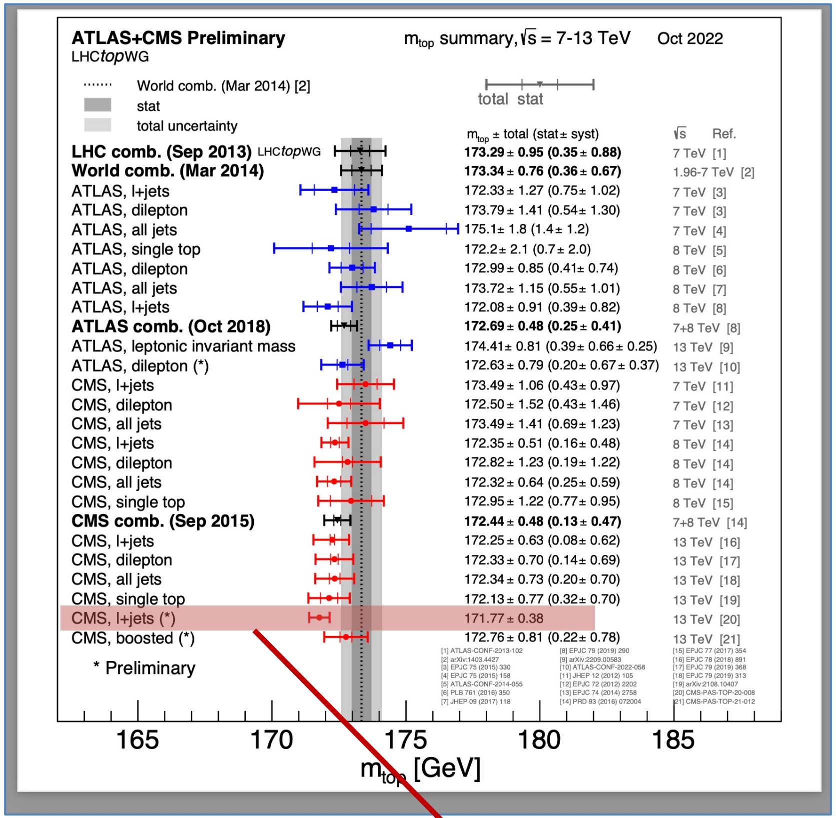
Power corrections in 3-jet region Nason & Zanderighi 2301.03607

- Sudakov Shoulder resummation MDS et al 2205.05702, 2306.08033
- Consistency with thrust and C parmater at 1% level would be convincing
- Stay tuned...

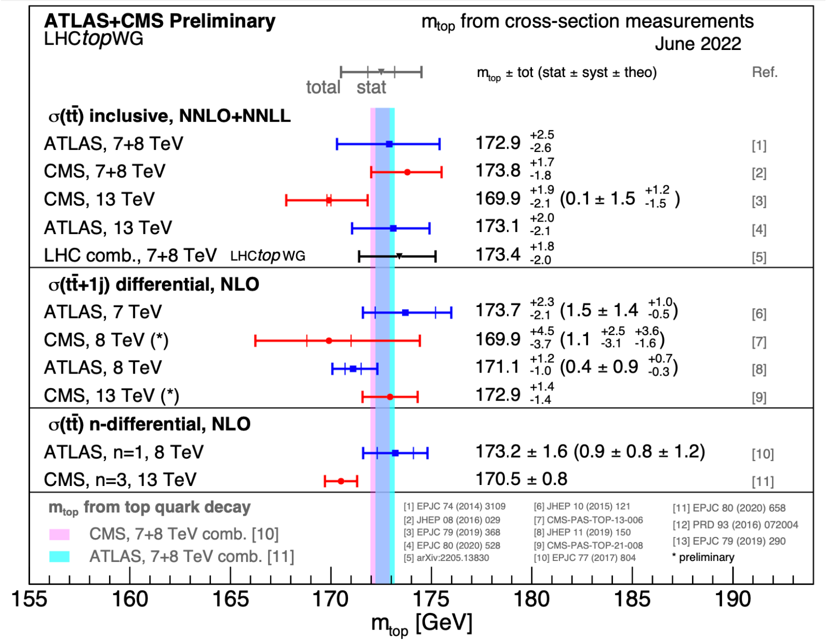
Top quark mass

https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots

Direct measurments



Indirect measurments (cross sections)



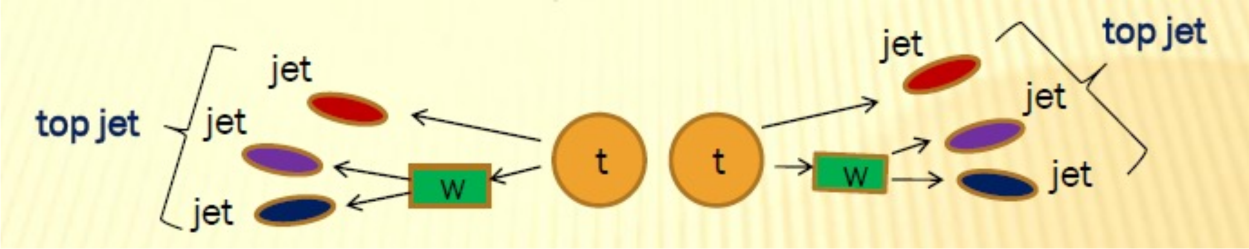
Best measurement in lepton+jets channel

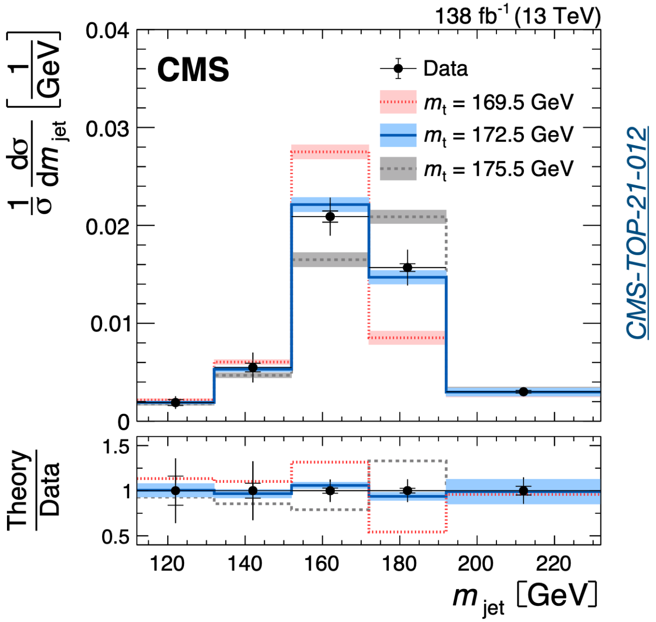
$$m_t = 171.77 \pm 0.38\text{GeV}$$

Top quark mass with fat jets

Fully hadronic channel is very challenging

- Huge multijet backgrounds make tt event identification impossible
- In boosted regime, tops become collamated and easier to see





CMS 2023 arXiv:2211.01456

top quark mass measurement in boosted tt events

$$m_t = 173.06 \pm 0.24\,(\text{stat}) \pm 0.61\,(\text{exp}) \pm 0.47\,(\text{model}) \pm 0.23\,(\text{theo})\,\text{GeV}$$
$$= 173.06 \pm 0.84\,\text{GeV}.$$

jet energy scale, etc.

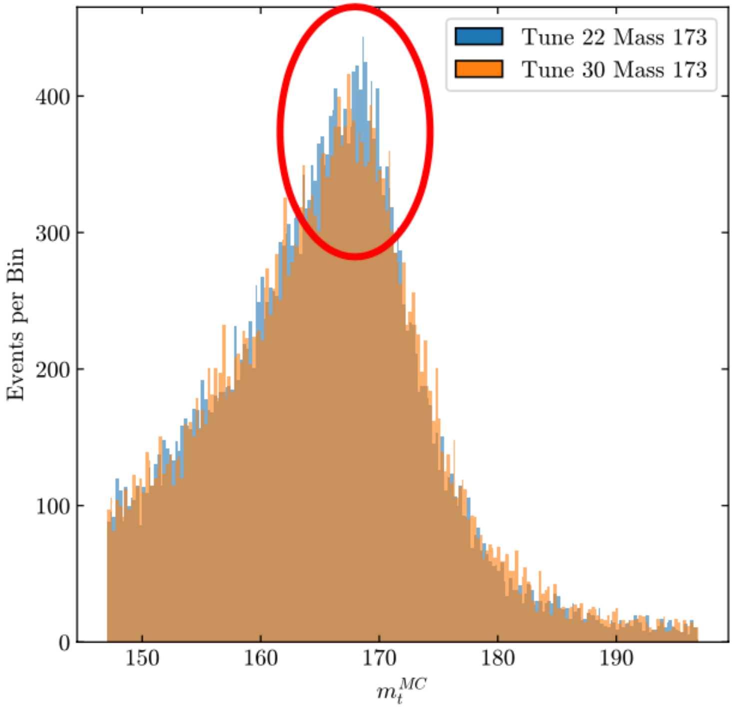
Monte Carlo modeling

perturbaive theory

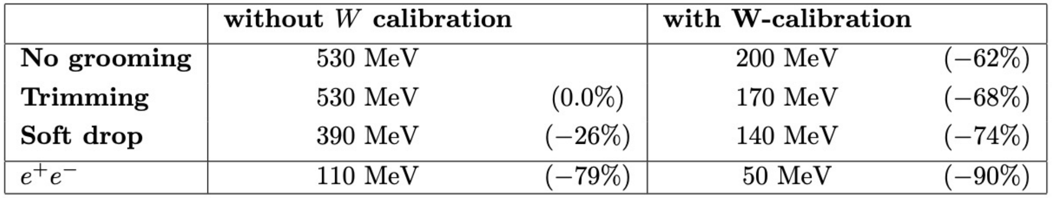
What is the top quark mass?

- Top quark is unstable, has color and charge
  - No well-defined pole-mass in quark propagator
  - MS-bar top mass is well-defined but hard to relate to data
    - useful for indirect measurements like cross section
  - Most experiments measure the **"Monte Carlo" mass**

Different MC tunes with same top mass give different distributions



- differences largely soft physics
- tuning uncertainty reduced with jet grooming (trimming, soft-drop)

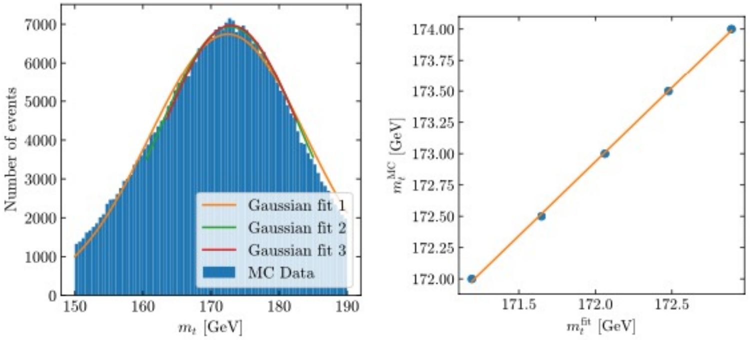


	without W calibration		**with W-calibration**	
No grooming	530 MeV		200 MeV	(−62%)
Trimming	530 MeV	(0.0%)	170 MeV	(−68%)
Soft drop	390 MeV	(−26%)	140 MeV	(−74%)
e^+e^-	110 MeV	(−79%)	50 MeV	(−90%)

Andreassen, MDS arXiv:1705.07135

Event ensembles

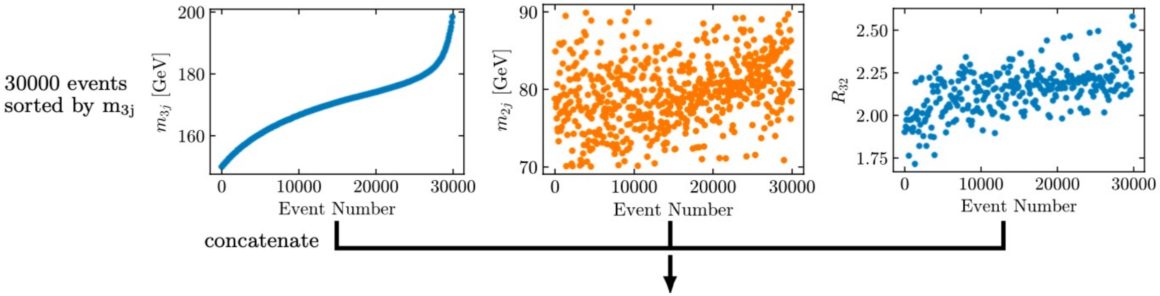
Flescher, Fraser, Hutchison, Osdiek, MDS arXiv:2011.04666



Fitting to peak/histogram shapes is inefficient

- Peak throws out useful information in tails
- Often need awkward parameterization of shape
- Why not just use all the information?

- For each event, measure m_{3j} (top mass), m_{2j} (W mass) and m_{3j}/m_{2j}
- Combine into one large array, sorted by m_{3j}

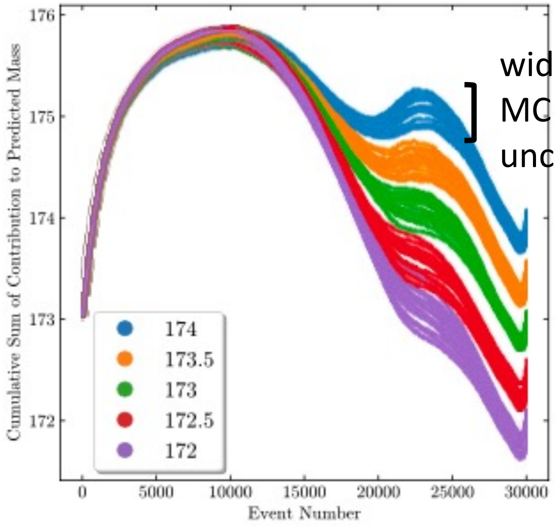


input into regression method

Can use all the information, not just peak

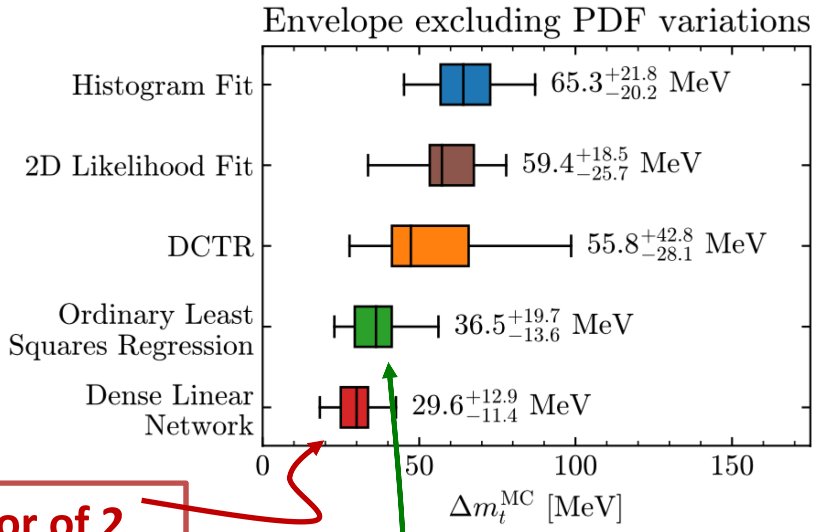
Event ensembles

Predicted mass
mulitpled by network weight



width is MC uncertainty

- full distribution gives good discrimination



Can reduce Monte-Carlo uncertainty by a factor of 2

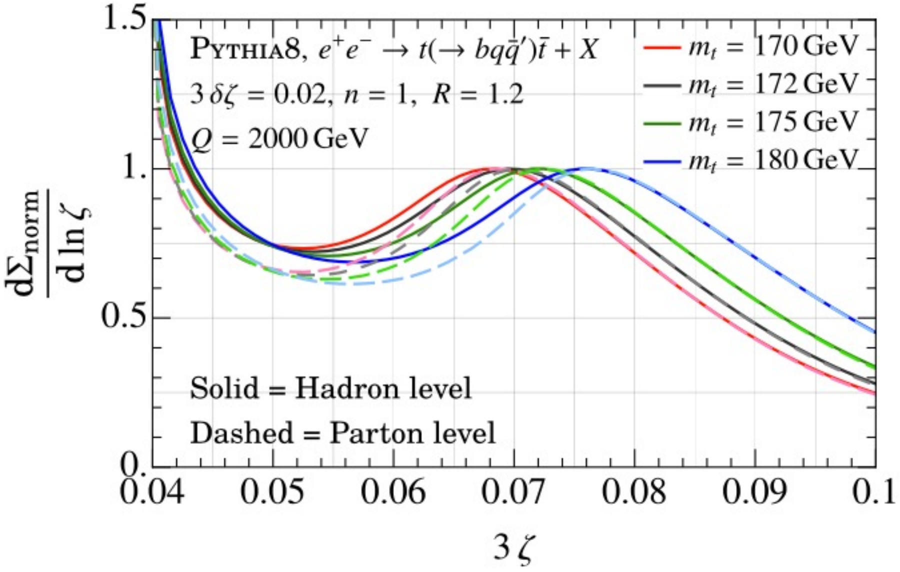
- **Learn from ensemble of events**
- Works better than fitting histograms
- Simple and effective way to use all avaiable data

- Can get similar performance with linear weights = ordinary least-squares regression (not machine learning)
- super fast, no training

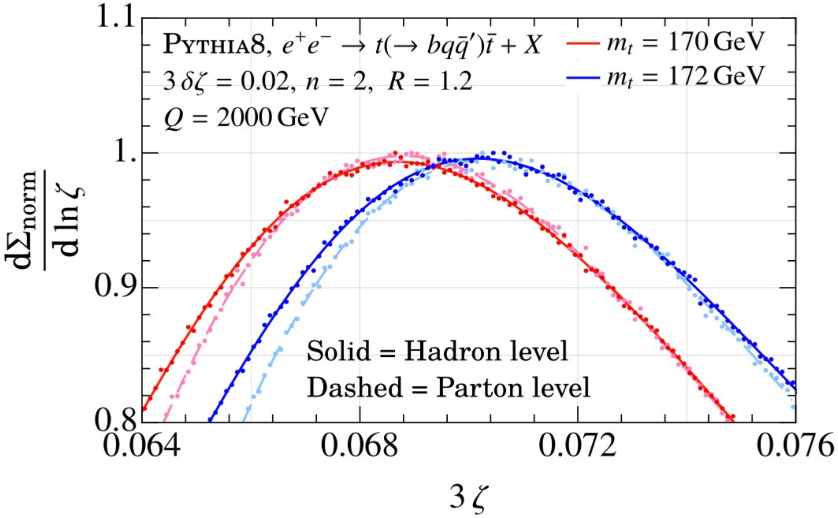
Flescher, Fraser, Hutchison, Osdiek, MDS arXiv:2011.04666

EECs for top mass

Holguin et al. arXiv:2201.08393



Measure the 3-point function in boosted top events



Insensitive to hadronization

$$\widehat{\mathcal{M}}^{(n)}(\zeta_{12},\zeta_{23},\zeta_{31}) = \sum_{i,j,k} \frac{E_i^n E_j^n E_k^n}{Q^{3n}} \delta\left(\zeta_{12}-\hat{\zeta}_{ij}\right)\delta\left(\zeta_{23}-\hat{\zeta}_{ik}\right)\delta\left(\zeta_{31}-\hat{\zeta}_{jk}\right)$$

Factors of energy in definition suppress soft radiation

Looks promising for m_t measurment

- In principle, direct theory-experiment comparision with short-distance top definition
- Early days, but worth watching

Conclusions

A lot of exciting progress in jet physics

- Machine learrning
- Precision measurements
- Top mass determination
- Energy-energy correlators
- Heavy ion physics
- Lund plane kinematics
- Improvements in unfolding
- Jet energy calibration
- Antenna showers
- Hadronization models
- Fixed order matching
- Anomaly detection
- Heavy flavor tagging (e.g. h->cc)
- …