Experimental Introduction, part 1

Petar Maksimovic, Johns Hopkins

ML4Jets, 2022

What to expect from this talk

- I am not as expert on ML as many of you are
- So, *what this talk is not*: a comprehensive review of the state of the field
  - Not even of the stuff developed by the experiments (by some of you in the audience!)
- However, I am an avid and enthusiastic consumer
- *This talk is essentially a shopping list*
  - Shine spotlight on some less-discussed and not-yet-resolved issues
  - Necessarily somewhat personal and biased

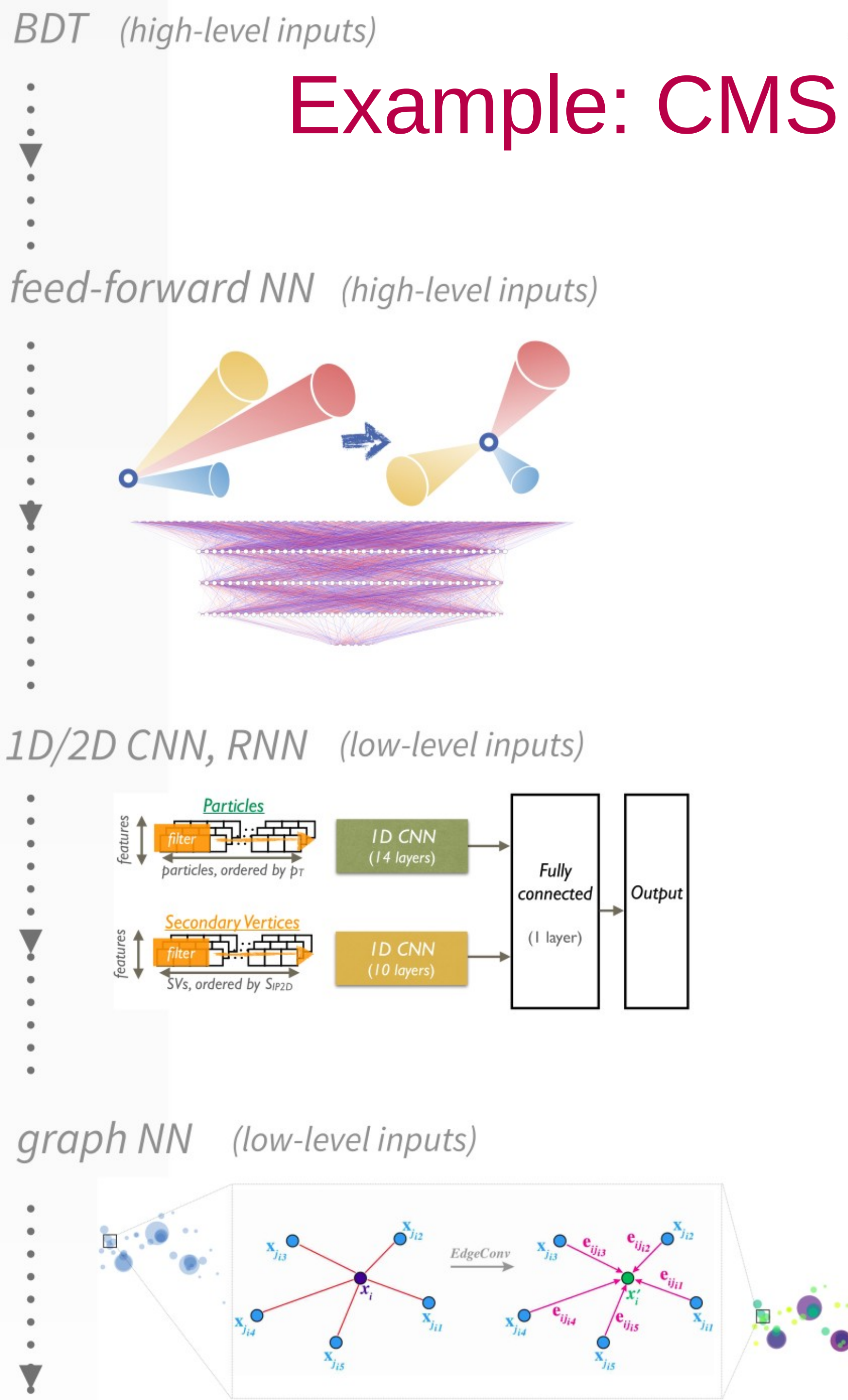
Outline

- First things first: the physics we're after
  - Jet taggers for t, W, Z, H
  - Supervised training for `exotic' substructure
  - `Anomalous' jet substructure (more in Tobias's talk)
- Remember the systematics!
- Resolving jet combinatorics
- Reco: Pixels, Tracking, B-tagging
- Fun with computing
  - (aka "Do we have enough MC?", fast simulation, etc.)
- Closing thoughts

The physics we're after: vanilla

- Boosted hadronic decays of top, W/Z, Higgs → end up in a large-R jet
- Jet taggers: now ML significantly outperforms the `classical' ones
- Everybody here has either heard of these taggers, used them, or even built even more powerful ones
- Searches for BSM:
  - Original raison d'etre of jet taggers
  - Backbone of large chunk of ATLAS and CMS BSM program

Example: CMS Boosted H → bb algorithms

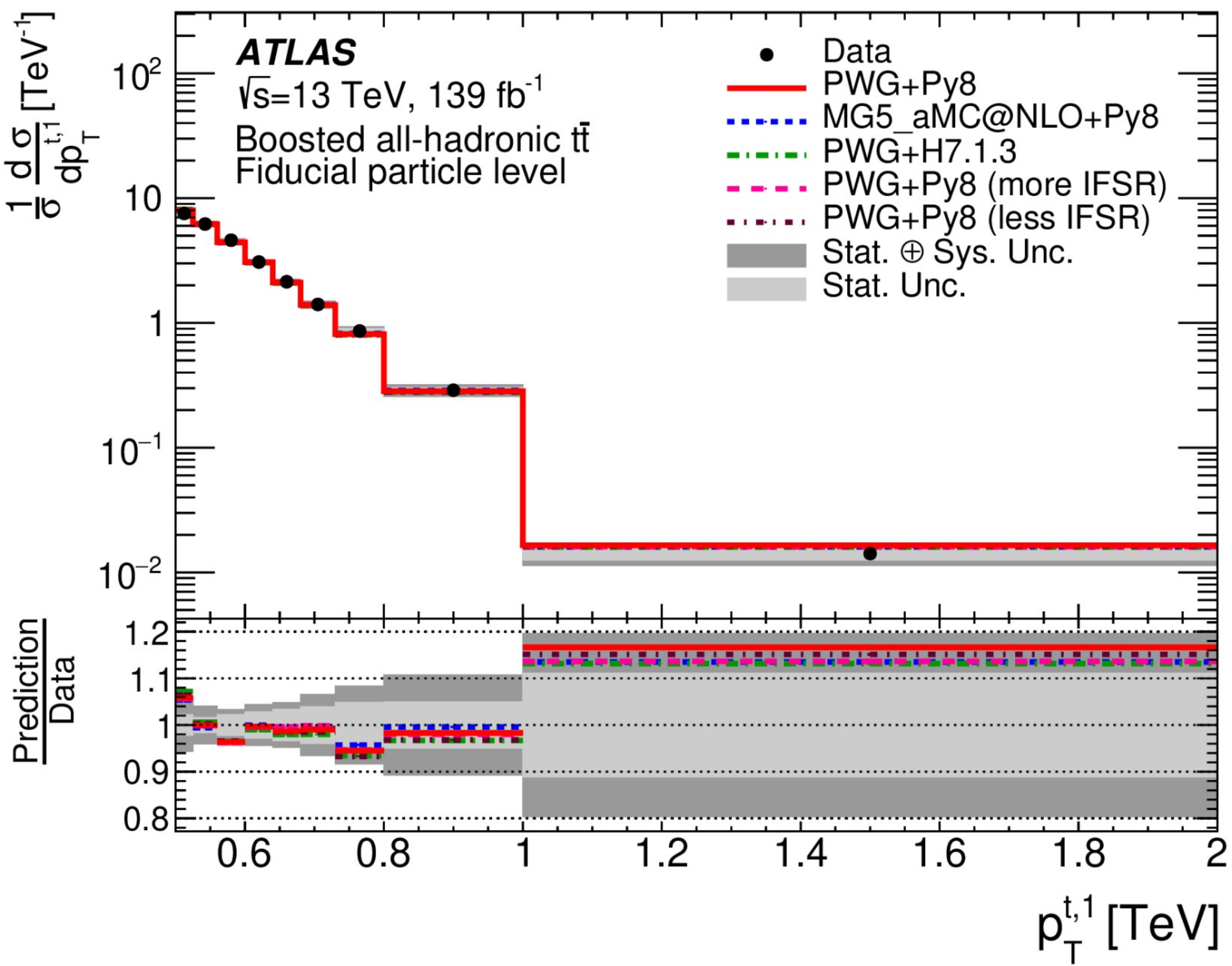


- “double-b”: ~2016
- “Deep Double-b”: single boosted X → bb, cc, etc.
  - Outperforms double-b by ~ x2
- “DeepAK8”: many taggers, analyses still coming out
  - Slightly better than DeepDblB
- “ParticleNet”: ~2021, cutting edge at CMS
  - Outperforms DeepAK8 by ~x2
  - Now $t\bar{t}$ dominates QCD

The physics we're after: vanilla (2)

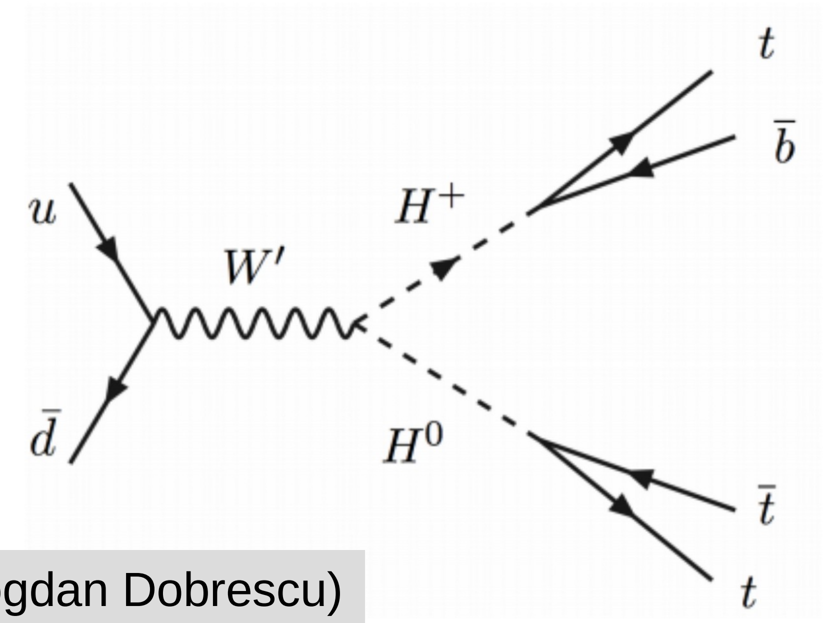
arXiv:2205.02817

- Standard Model physics:
  - Use jet taggers to access high- p_T regime
  - Various differential cross-sections
  - Jet masses, etc.



The physics we're after: strawberry

- Hadronically decaying BSM objects → large-R jets
- Supervised search since we know what we are looking for
- So far, only scratched the surface:
  - Searches for resonance decays to scalars, where $S \rightarrow b\bar{b}$
  - $W_{KK} \rightarrow W\phi$ where radion forms a jet $\phi \rightarrow WW \rightarrow 4q$
- Many interesting models → can publish many cool papers
  - "If you build it, they will come": new unusual taggers can attract new users



(from Bogdan Dobrescu)

The physics we're after: strawberry

- Hadronically decaying BSM objects → large-R jets
- Supervised search since we know what we are looking for
- So far, only scratched the surface:
  - Searches for resonance decays to scalars, where $S \to b\bar{b}$
  - $W_{KK} \to W\phi$ where radion forms a jet $\phi \to WW \to 4q$
- Many interesting models → publish many cool papers
  - "If you build it, they will come": new unusual taggers can attract new users
- But… how to calibrate the signal efficiency of an jet with `exotic' substructure?
- How to estimate uncertainty on data/MC Scale Factor?

The physics we're after: strawberry

- Hadronically decaying BSM objects → large-R jets
- Supervised search since we know what we are looking for
-

$\to b\bar{b}$

$\to 4q$

- ers
  - ers can attract new users

This is a high-priority problem, worthy of your attention!

Need input / help from theory!

- But… how to calibrate the signal efficiency of an jet with \`exotic' substructure?
- How to estimate uncertainty on data/MC Scale Factor?

The physics we’re after: mystery flavor

- What if the BSM physics is being copiously produced, but we’re not simply looking for the correct signature?
  - “Anomalous” jet substructure = “none of the above”
- Train on data, learn what “SM backgrounds” are, then look for what is different from it
- Used (so far by ATLAS) in conjunction with other handles, like heavy resonance mass, or presence of a H jet.
- More in Tobias’s talk, followed by 1st results from ATLAS!

The physics we're after: mystery flavor

- What if the BSM physics is being copiously produced, but we're not simply looking for the correct signature?
  - "Anomalous" jet substructure = "none of the above"
- Train on data, learn what "SM backgrounds" are, then look for what is different from it
- Used (so far by ATLAS) in conjunction with other handles, like heavy resonance mass, or presence of a H jet.
- *Strategically important*: run dedicated searches first, then an anomaly search to clean up everything that's left
  - Ensures we at the LHC are doing our `due diligence'

What we need to search for BSM

- Data
- Tools (taggers, new variables) to suppress background and isolate the signal
  - Most ML taggers still trained on MC...
- Background estimate (minimize uncertainty)
  - If dominated by ttbar, W+jets – get away with MC…
  - QCD: tricky and messy (after lots of work…)
- Signal efficiency (minimize uncertainty)
  - For top, W/Z, Higgs tagging – use standard candles
  - For exotic signatures – ???

We still rely on QCD MC

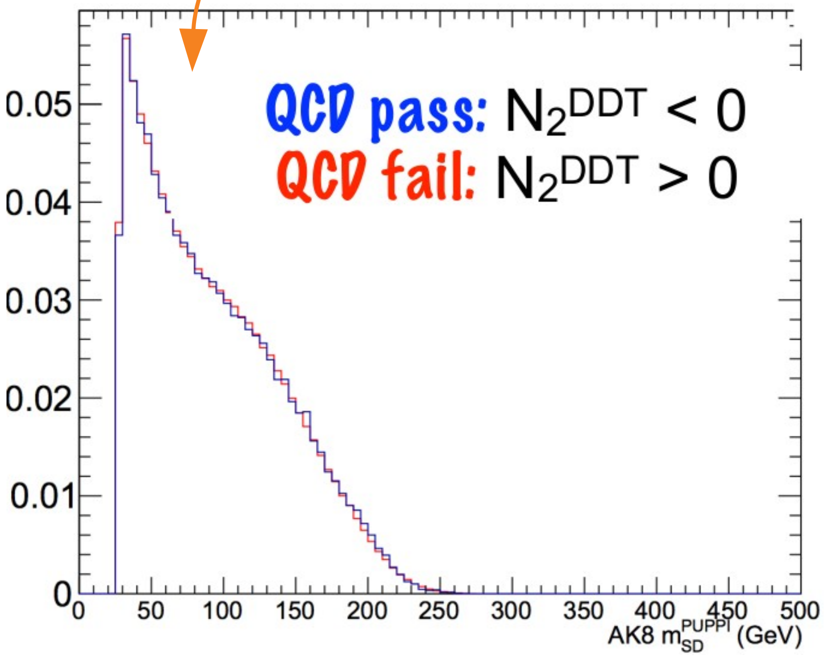
- We try hard to limit use of QCD simulation – elaborate schemes to estimate backgrounds from data
- However!
  - Sometimes data-driven background estimates are “helped” by MC
  - We train ML jet taggers on MC
    - How do we know what we are not picking up on data/MC disagreement?
  - If decorrelation is not perfect, we use MC to make a DDT map (= cut value vs (mass,pt))
  - We use MC for signal efficiency
    - Calibrate in data, when possible
    - How to evaluate signal systematic for exotic substructure?

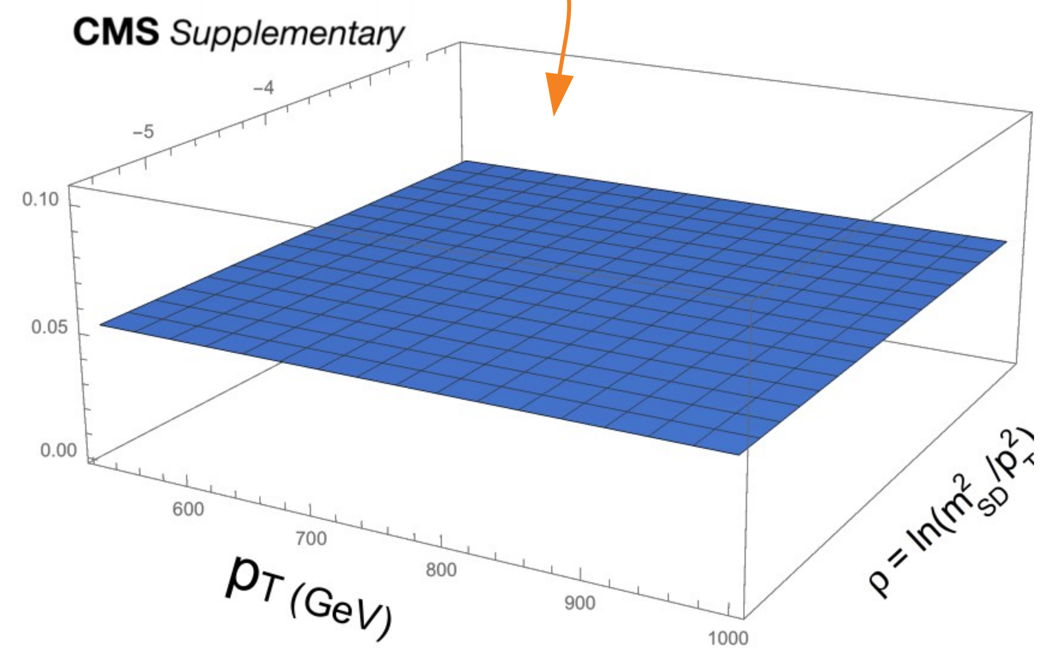
QCD simulations are not perfect

EXO-18-012

- Fully decorrelated tagger (via “DDT map”)
  - Doesn’t mess up the jet mass shape

$\Rightarrow$ Pass (tag) and fail (anti-tag) region have same shape $\Leftrightarrow$ background efficiency …is flat in MC



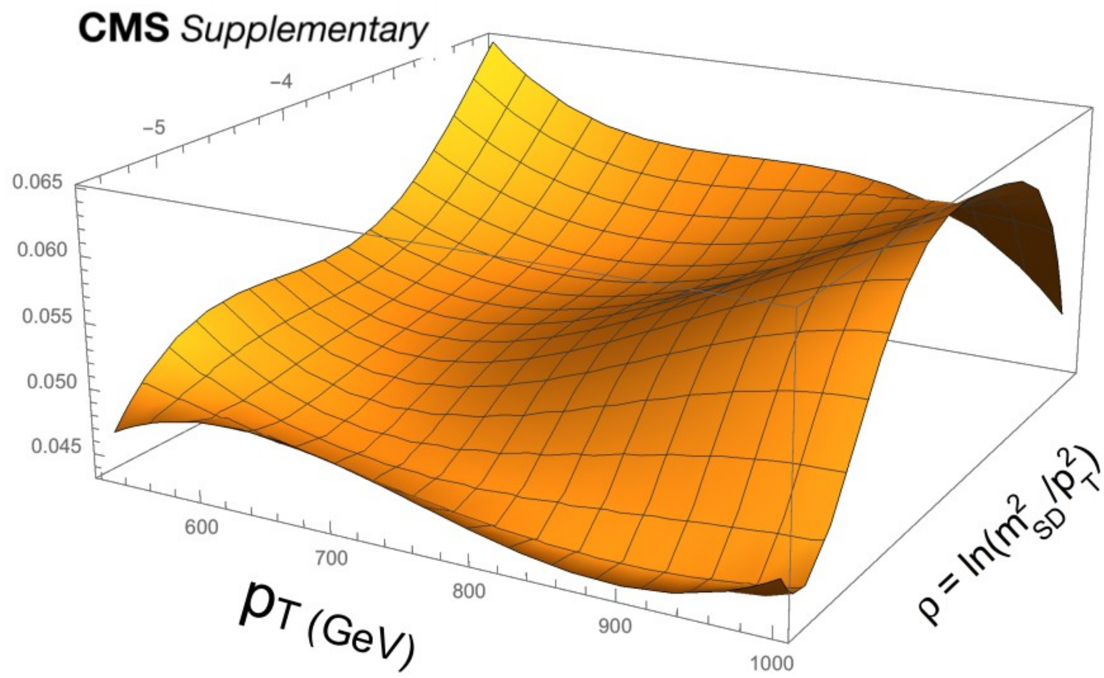


QCD simulations are not perfect

EXO-18-012

- Account for Data/MC discrepancy by a smooth surface fit
- … can be very complicated if Data/MC differences are not trivial!

$$n_{\text{pass}}^{\text{QCD}} = R_{\text{p/f}}\, n_{\text{fail}}^{\text{QCD}}$$



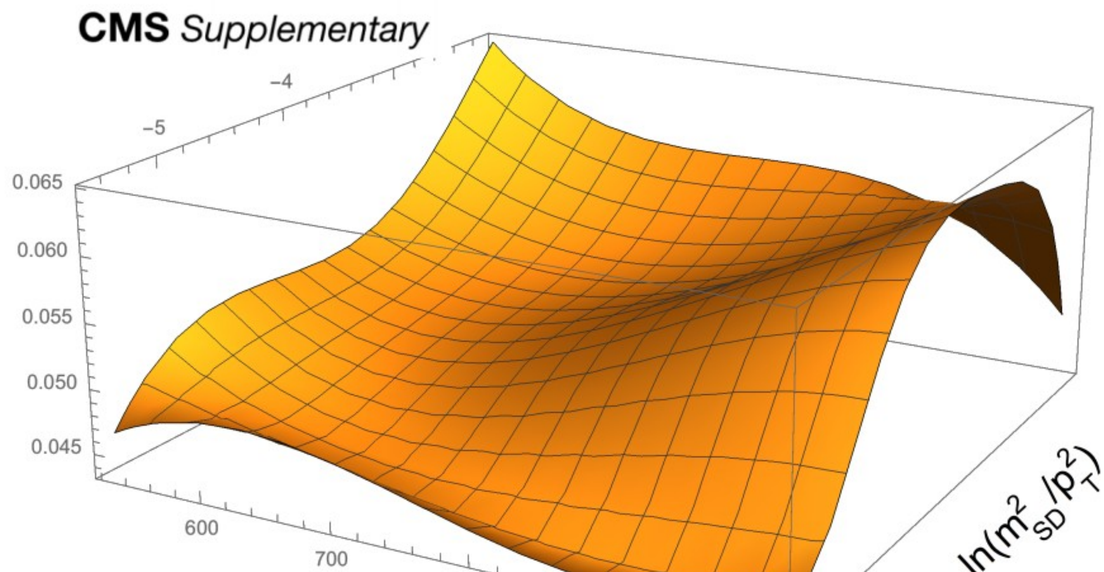
- For light $Z' \to q\bar{q}$:
- Surface is a product of
  - 3rd degree poly in p_{T}
  - 5th degree poly in ρ !!!

QCD simulations are not perfect

EXO-18-012

- Account for Data/MC discrepancy by a smooth surface fit
- … can be very complicated if Data/MC differences are not trivial!

$$n^{\mathrm{QCD}}_{\mathrm{pass}} = R_{\mathrm{p/f}}\, n^{\mathrm{QCD}}_{\mathrm{fail}}$$



- For light $Z' \to q\bar{q}$:
- Surface is a product of
  - 3rd degree poly in p_{T}
  - 5th degree poly in ρ !!!

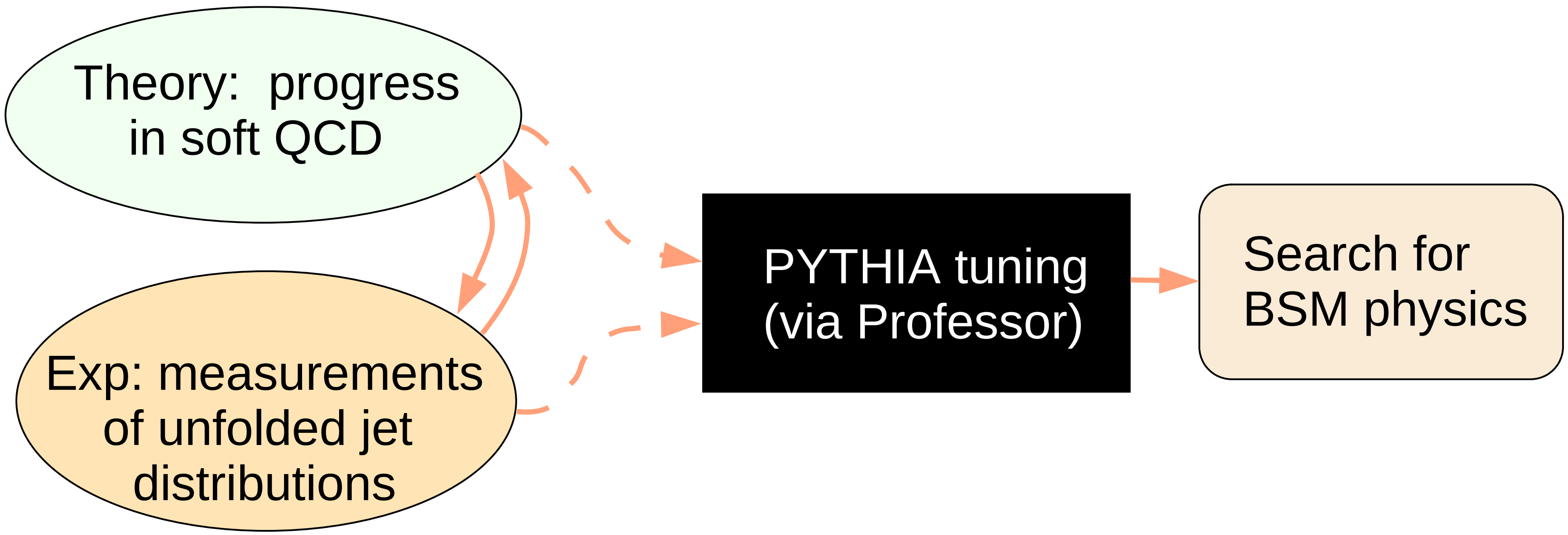
- Neither scalable with luminosity
- Nor easily transferable to other searches!

QCD modeling for the future

- With a better QCD modeling, we could:
  - Train ML algorithms
    - → better data/MC agreement
    - → minimize signal efficiency systematics
  - Decorrelate taggers
    - → well-behaved background shapes → better bkg estimates
    - → if there’s a BSM excesses, it would be “easier” to see
  - Estimate efficiencies of tagging jets with exotic substructure

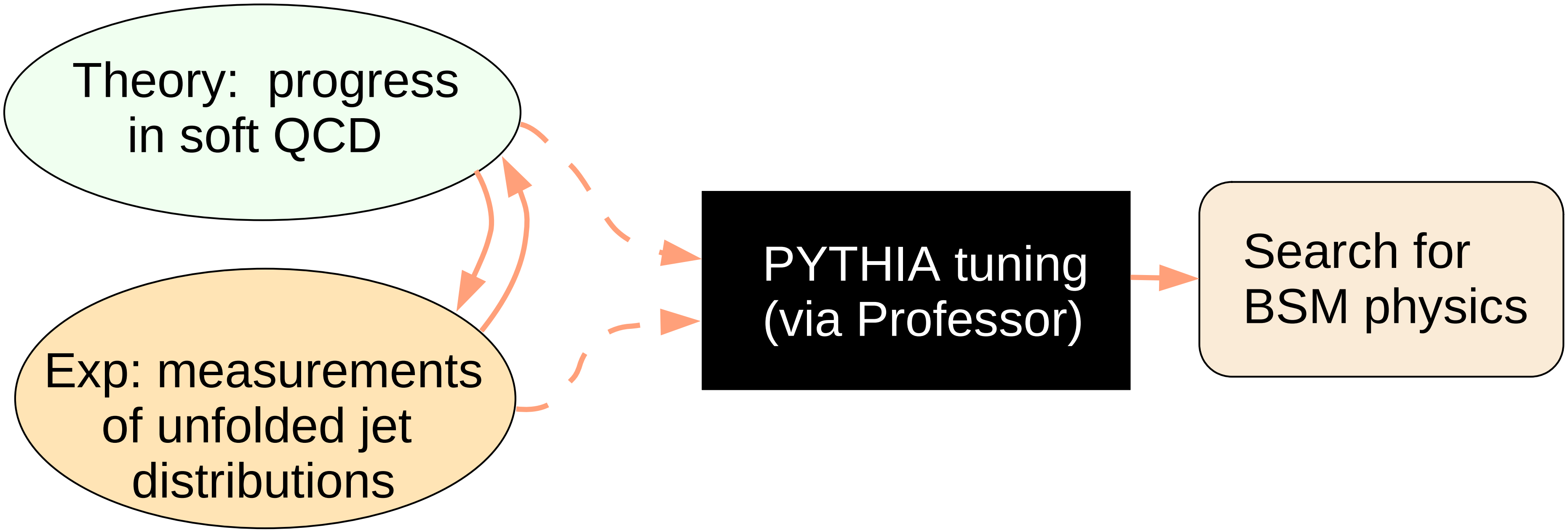
    (see above)
- In general, experimentalist’s life would become a lot easier

Why can't I have that?!



- Somehow, theoretical and experimental progress in soft QCD does not seem to propagate to MC samples we use.
  - Not enough measurements fed into "Professor"?
  - Can't tune both UE and substructure???
  - PYTHIA is insufficient for shower/hadronization?

Why can't I have that?!



- Somehow, theoretical and experimental progress in soft QCD does not seem to propagate to MC samples we use.
  - Not enou
  - Can't tu
  - PYTHIA

How do we know that ML improvements demonstrated on MC will in fact also be there in data?

Reweighting QCD MC vs training on data

- One way around imperfect QCD modeling: reweight MC to match data in low-level quantities
  - Reweight jet image or primary Lund jet plane (no ML)
  - JUNIPR, DCTR (uses ML)
  - SALAD (reweights MC in sidebands – nice!), etc.

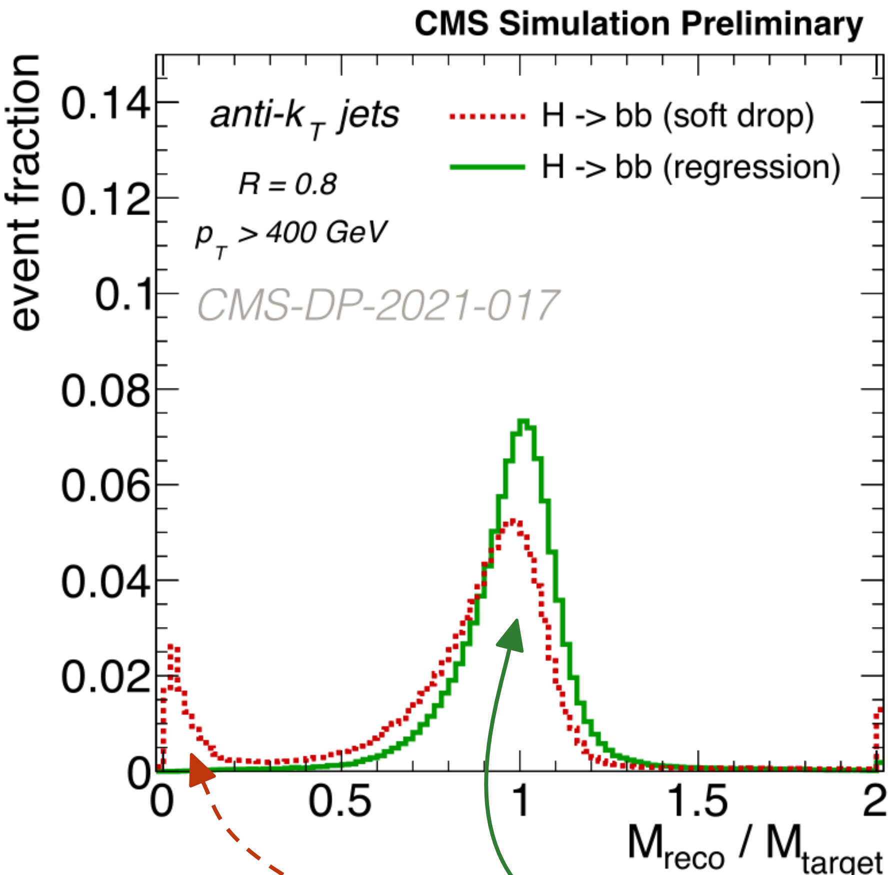
$\Rightarrow$ This clearly requires input / blessing from theory

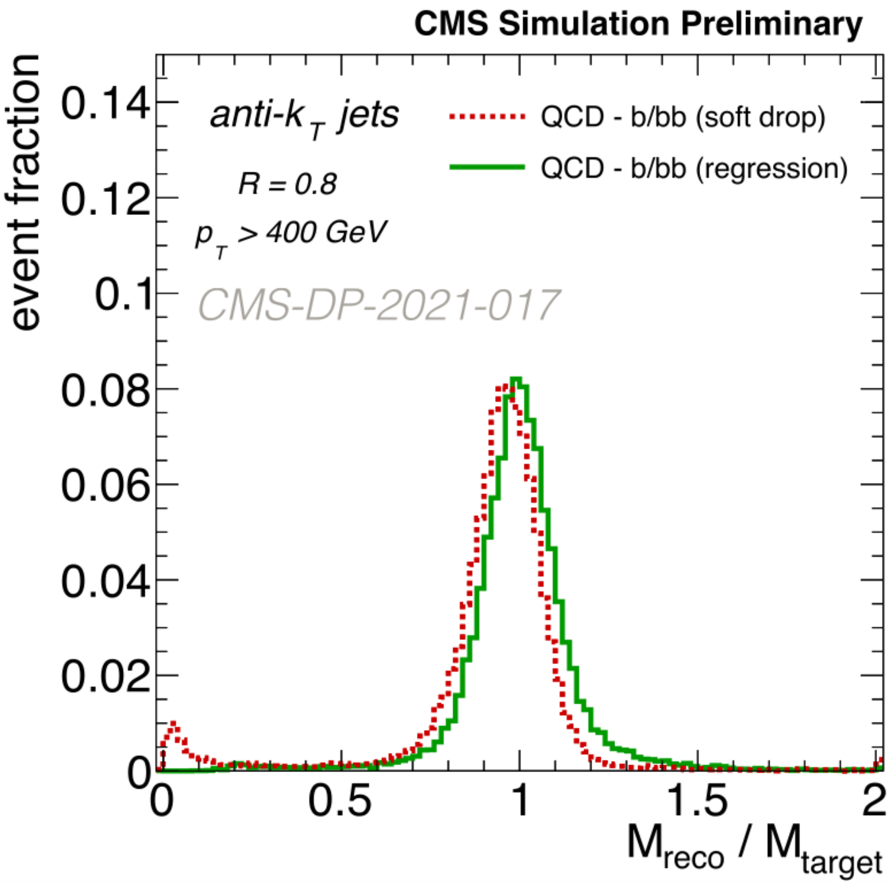
- Another option: use CWoLa to train on data
  - Likely most promising in the long run
  - But needs to be worked out

    include uncertainties, combine with a likelihood fit(?), deploy machinery to be widely used within experiments...

Jet mass regression with ParticleNet

- Use same architecture, inputs, training as for PN tagger
- Training target: pole mass (signal), gen-level mass (QCD)





No signal at 0, narrower peak ⇒ Improvement ~ 20% !

Do we care about IRC safety?

- Initially, we did not (jet pruning in CMS, trimming in ATLAS)
- Then we did (soft drop, at least on CMS)
- But now, we have second thoughts: particle-based taggers and this mass regression are really nice!
- If the ML tool (tagger, mass regression) is calibrated in data, does it matter?
  - Tried taggers built from IRC safe observables, but (for example) ParticleNet outperforms it!

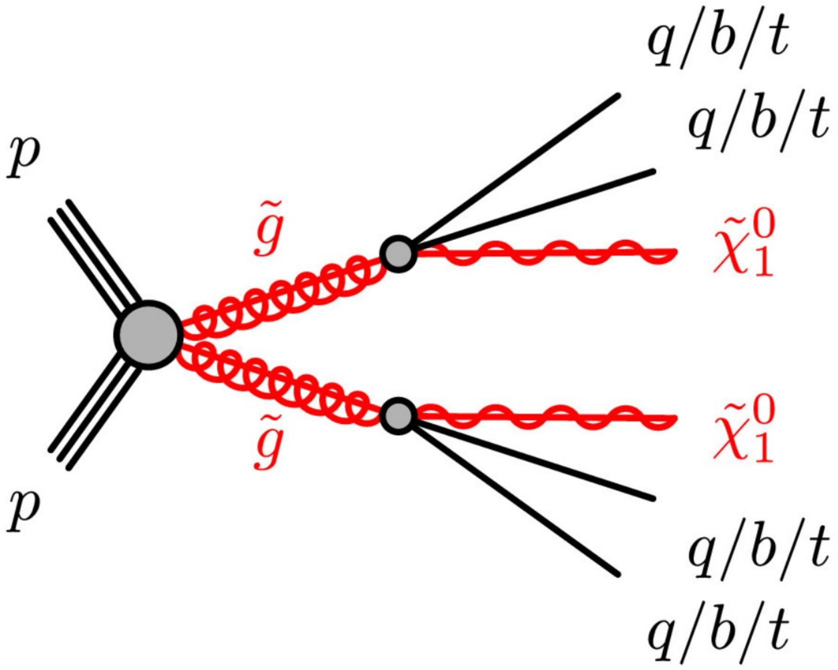
Do we care about IRC safety?

- Initially, we did not (jet pruning in CMS, trimming in ATLAS)
- Then we did (soft drop, at least on CMS)
- But now, we have second thoughts: particle-based taggers and this mass regression are really nice!
- If the ML tool (tagger, mass regression) is calibrated in data, does it matter?
  - Tried taggers built from IRC safe observables, but (for example) ParticleNet outperforms it!

Another topic ready for theoretical input:
If IRC safety helps us bypass limitations from QCD MC, then it may be worth it!

Resolving jet combinatorics

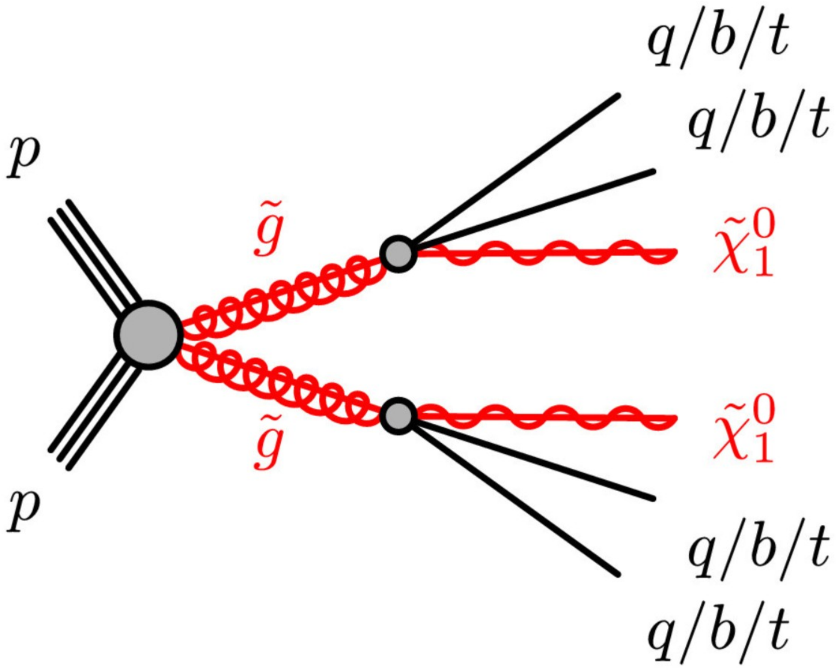
- BSM with many jets in the final state. Example:



- Ubiquitous in searches
- A very old problem, with many solutions
  - Various kinematic fits (old)
  - Recursive Jigsaw Reco (new)
- This is ripe for deployment of clever ML
  - See Larry Lee's talk
  - ML could be grafted on top of RJR
  - But… Need to account for jet energy scale uncertainties
    - (As that may flip the interpretation...)

Resolving jet combinatorics

- BSM with many jets in the final state. Example:
- Ubiquitous in searches
- A very old problem, with many solutions
  - Various kinematic fits (old)
  - Recursive Jigsaw Reco (new)
- This is ripe for deployment of clever ML



ATLAS and CMS spend ~dozens of FTE months / year solving this again and again.

This is another super-important problem that should be solved once and for all!

ML in pixels and tracking

- Pixel clusters are 2D images, ideal for ML
  - ATLAS has been using ML in pixel local reconstruction for a long time (position, error, cluster splitting)
  - CMS: still in R&D (but looks good)
- Unlike QCD – we have very realistic simulation of the charge deposition and drift, calibrated on data
- Tracking pattern recognition is also excellent candidate for ML
  - first ideas in 1988
  - many projects making progress

B tagging

- The first application of ML to jets (although small-R ones)
- Both ATLAS and CMS use well-performing algorithms
- Future: further use of ML in pixel cluster splitting/merging, removing wrong hits pile-up, using “trackless” algorithms
- Computing limitation: hard to study on data, since all easily accessible samples are compressed, without low-level pixel info
- ATLAS is using EvtGen (!) and CMS still Pythia (?!)
- Not sure if / how the measurements of hadronization propagate back to ML training

Detector simulation

- Official party line: there won't be enough computing in HL-LHC to generate full simulation MC
  - (I still have hard time buying this, but let's assume it's true)
- If that's the case, can we trust Fast Simulation for that?
- One way out: make Fast Simulation more realistic (= less ideal) and useful for training ML
  - Not deserving enough attention by experiments
  - So far, good progress with Calorimeter (CaloGAN, etc.)
  - Can we use a similar approach for tracking?

More on computing vs. ML

- (If there won't be enough computing) do we need to be mindful of size and complexity of models, and size of training datasets?
- Typically, a delay of ~1.5-2 years (or more) between a new ML tagger and its adoption in analyses:
  - Needs to be in an official release
  - Discriminants need to be in a slimmed dataset (otherwise people won't use it)
  - Needs to be calibrated
  - Chicken-and-egg: officially supported if there is demand, but no demand if not officially supported...
- Can we make turnover shorter?

Lots of progress… Or, is it?

- Alison Lister's Musings from ML4Jets 2020:

- It seems to take 3-4 years from an idea to appear (e.g. arXiv study on Delphes) until our large collaborations publish results with it
- Why?
  - New methods are hard to calibrate?
  - New methods aren't developed by people who do the calibrations? [c.f. leaky pipeline]
  - Man-power issues?
  - Reticence to change?
  - Technical difficulties (computing infrastructure,…)? [c.f. lwdnn]
  - It's more fun to play with Delphes and new ideas than go through collaboration approval?
- Thoughts for the next few days
  - How can we bring closer together the three communities of ML developers, Object Performance (CP/POG) experts, search/measurement analysers?
    - In particular reduce the inertia and reduce the latency from idea to physics paper
- Simulation improvement efforts do seem to be collaboration driven (at least on ATLAS)

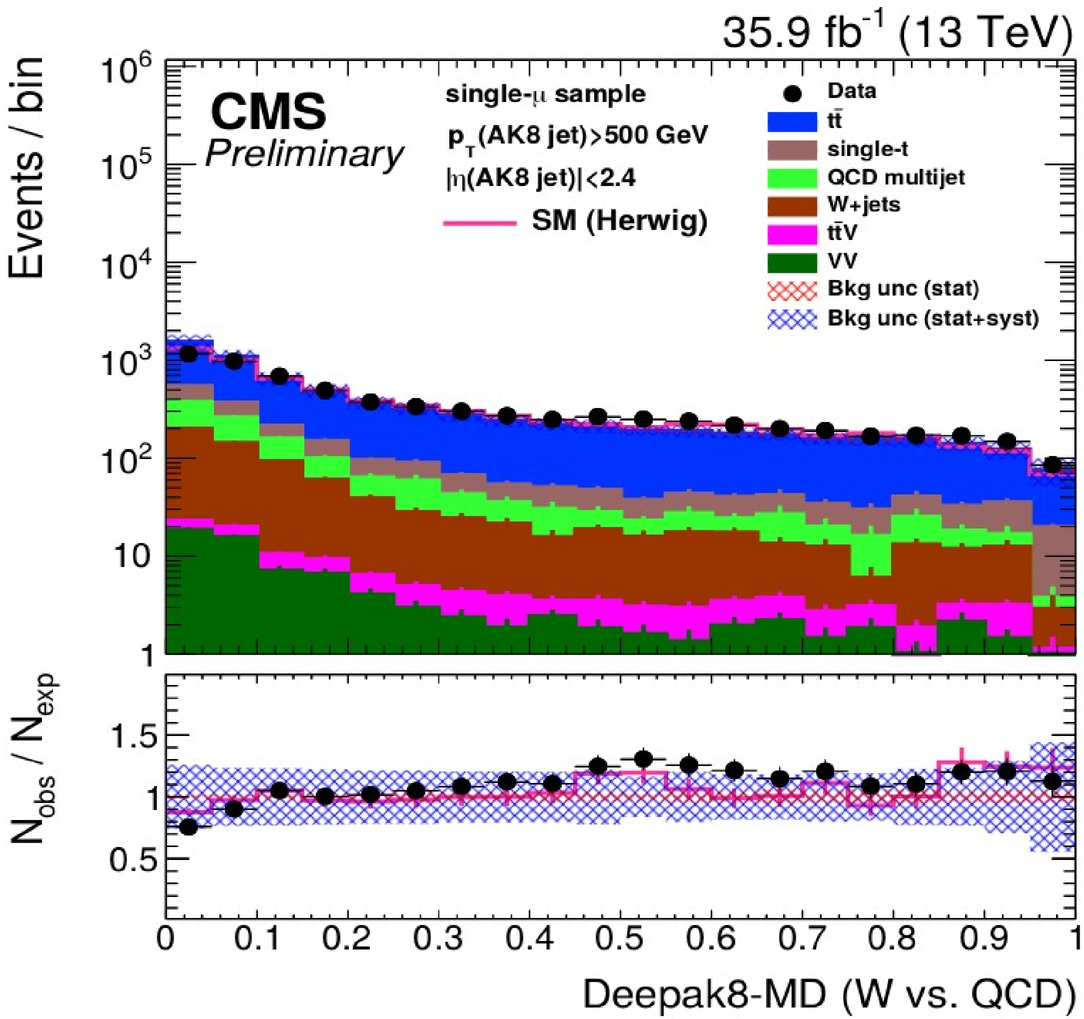
My take

- Rapid development of all sorts of ML tools for HEP
  - Measurements with jets are a prime target, receiving most attention
  - Top / W/Z / Higgs jet taggers still keep improving!
- Nevertheless, some important issues (related to MC) are not being resolved, despite excellent ideas and initial progress
  - Although I would love to be proven wrong!
  - Likely more coordination (here!) is needed
- Adoption of new tools still slow. Computational and sociological challenges still remain
  - Most of what Alison wrote 2.5 years ago is still true! :(

BACKUP MATERIAL

Imperfect MC is used to train ML

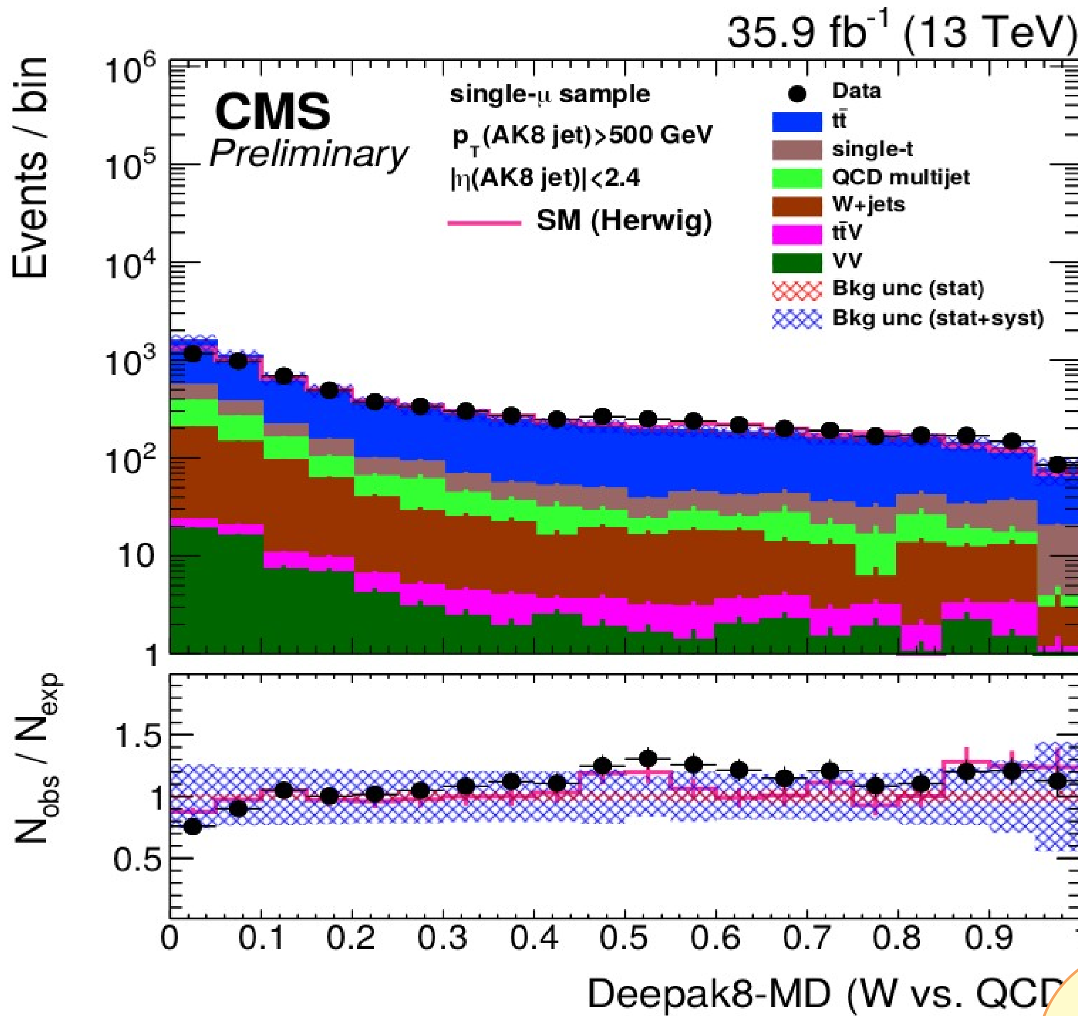
JME-18-002



- Powerful taggers, but…
- Nominally "within errors" from data
- Need to be careful:
  - Scale factors must be measured...
  - And they may be different from 1…
  - … with large error bars

Imperfect MC is used to train ML

JME-18-002

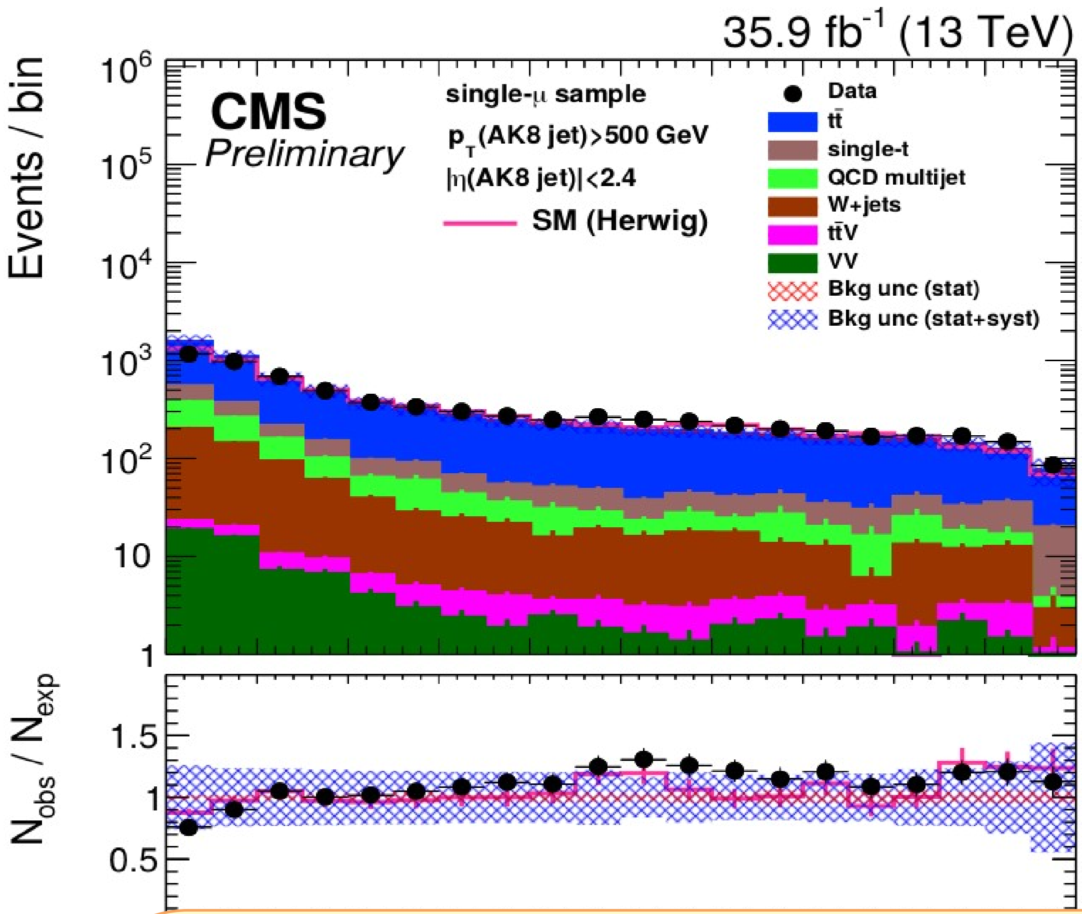


- Powerful taggers, but…
- Nominally "within errors" from data
- Need to be careful:
  - Scale factors must be measured...
  - And they may be different from 1…
  - … with large error bars

Can erase some of the gains from an improved tagger!

Imperfect MC is used to train ML

JME-18-002



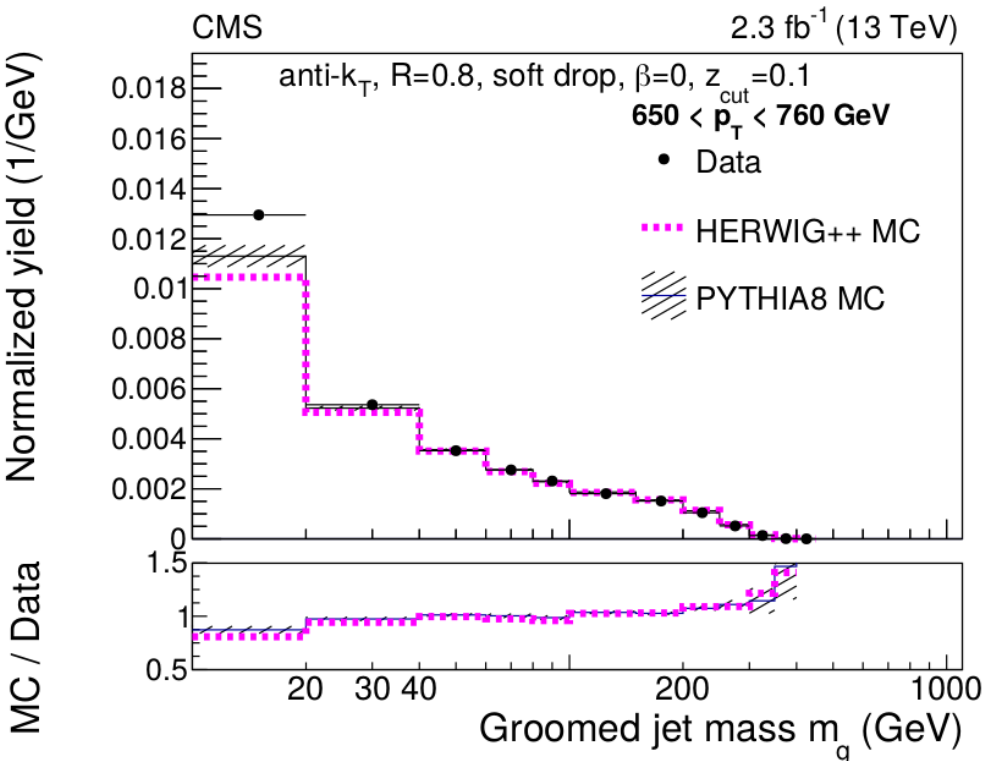
- Powerful taggers, but…
- Nominally “within errors” from data
- Need to be careful:
  - Scale factors must be measured...
  - And they may be different from 1…
  - … with large error bars

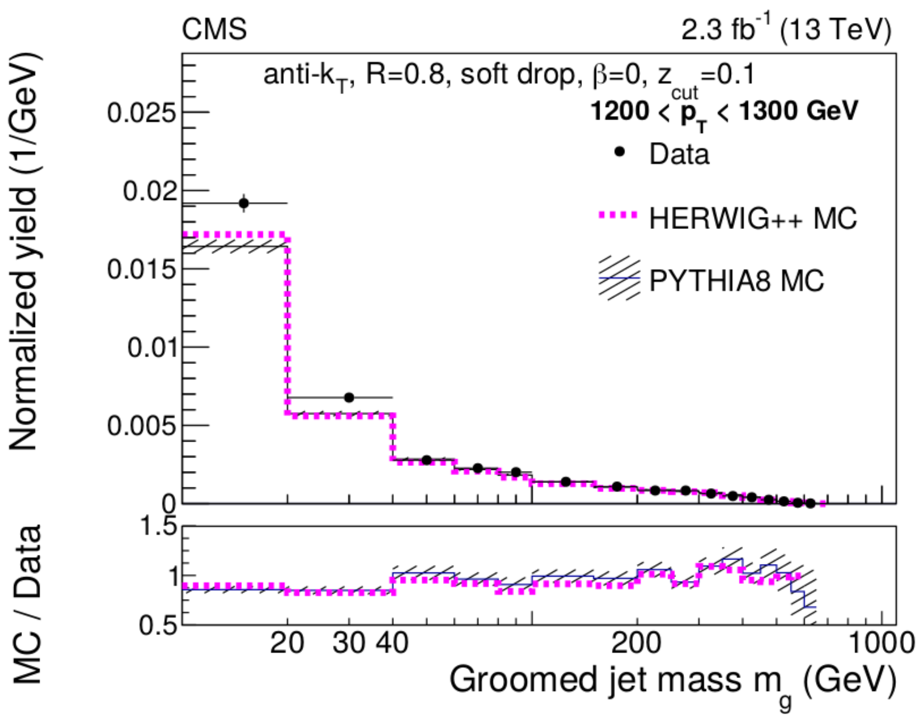
Hard to tell whether DNN is focusing on features poorly modeled in top/W/Z/H MC...

Measurements of jet mass

JHEP 11 (2018) 113

- Both experiments have measurements of jet mass (ungroomed and groomed).





- These are then unfolded.

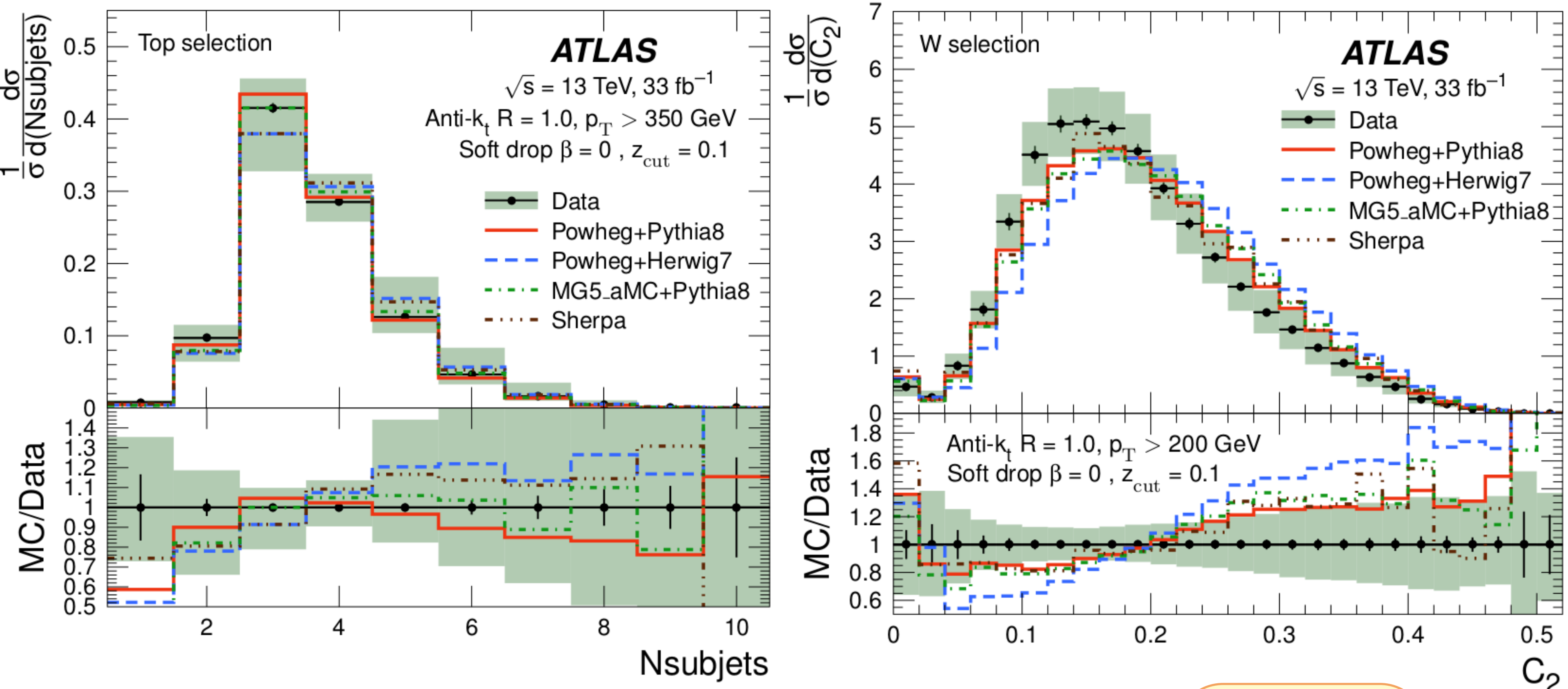
Detector level

Measurements of substructure

arXiv:1903.02942
(accepted by JHEP)

- ATLAS compared a bunch of substructure-related variables to data. Examples:

$$C_2 = \frac{e_3}{(e_2)^2}$$



- These (and others) are then unfolded.

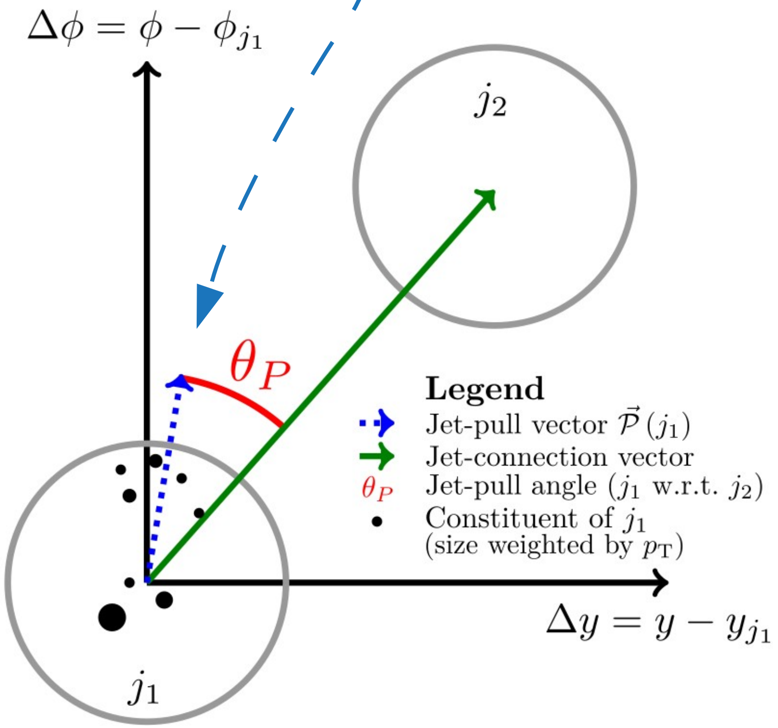
Unfolded

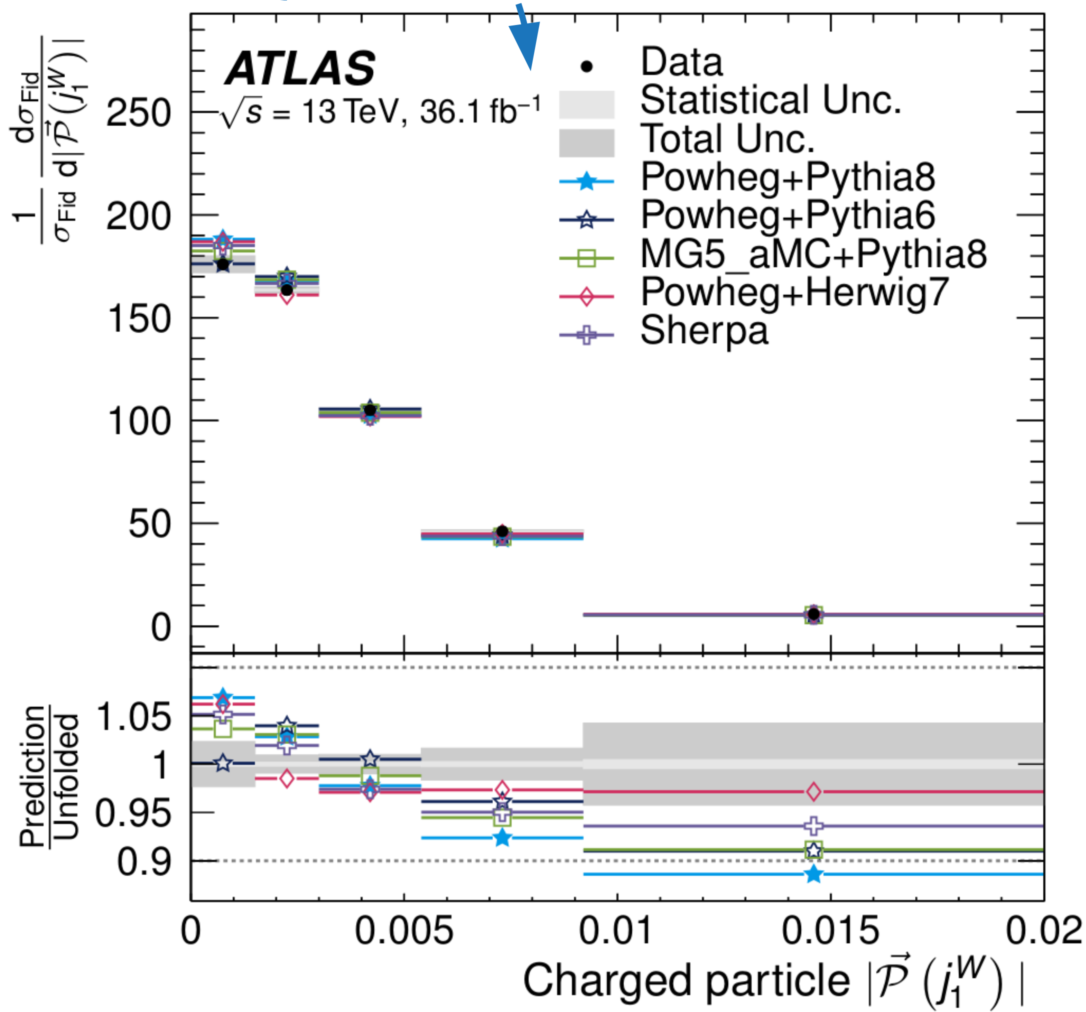
Substructure: jet pull (color flow)

EPJC78 (2018) 10, 847

- Encodes color connections between partons
  - Jet pull vector: angle and magnitude

Unfolded

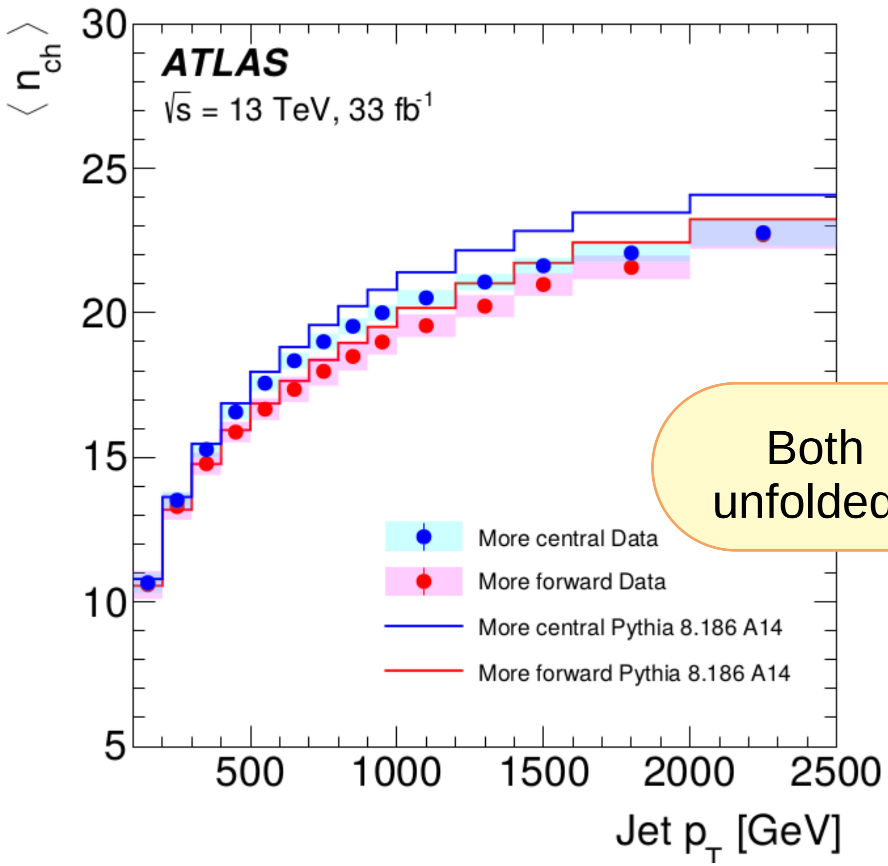


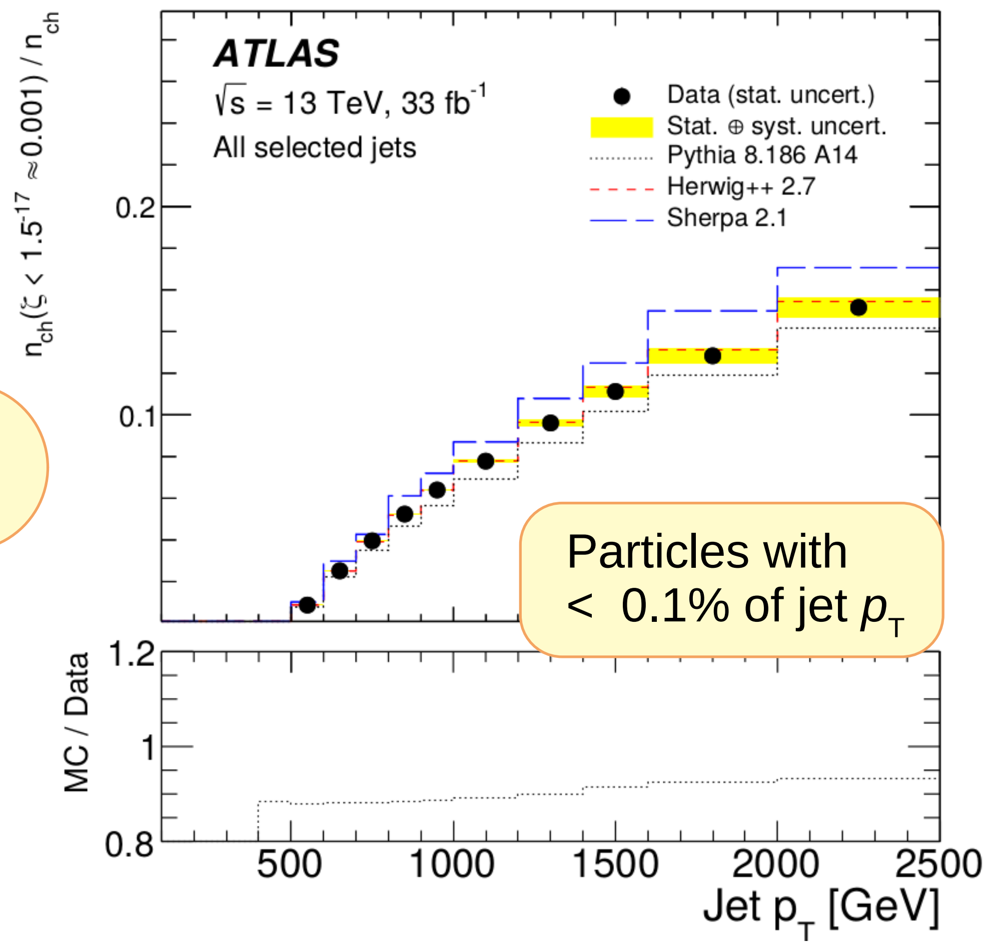


Hadronization

arXiv:1906.09254

- Using charged particles in jets to calc. several variables
  - reasonable description overall
  - also noticeable differences, especially for gluon-like jets





What if PYTHIA doesn't have enough knobs?

- Make a better PYTHIA, or another shower program. Then tune to data. (Repurpose the Professor?)

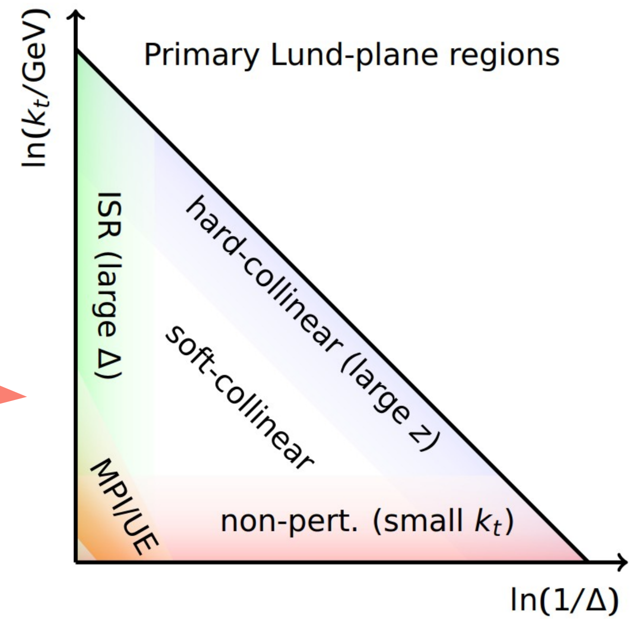
- Or, correct simulation *a posteriori*.
  - Reweight using the Lund Plane?
  - JUNIPR?
  - …. or something else?

  - Maybe the best:
    measure → tune PYTHIA → reweight residual differences.

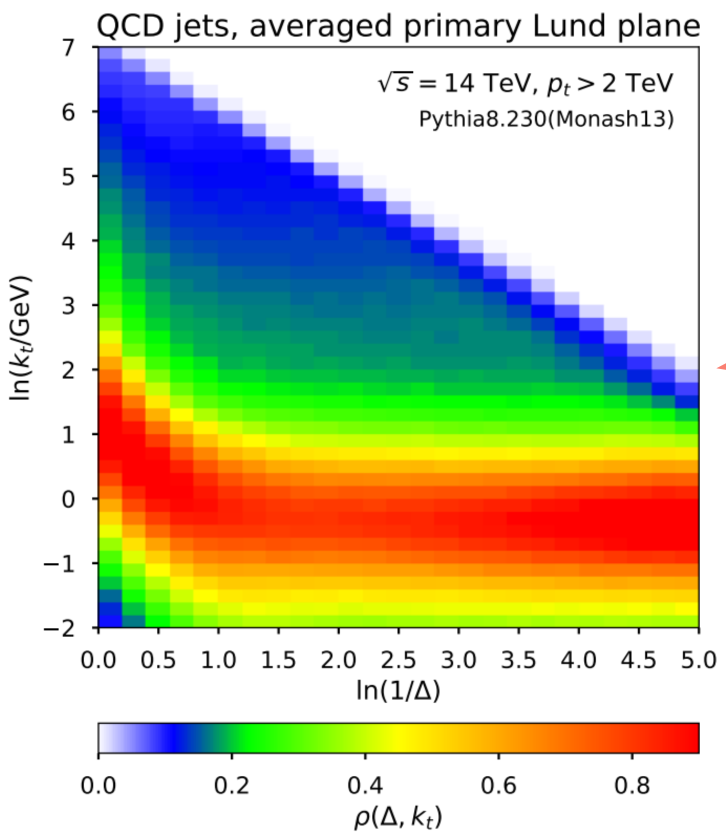
  - Experimentally, **the key question**: what are the *uncertainties* on the result of this procedure?

Primary Lund Plane

- “Jet is an unordered set”, I know… but...



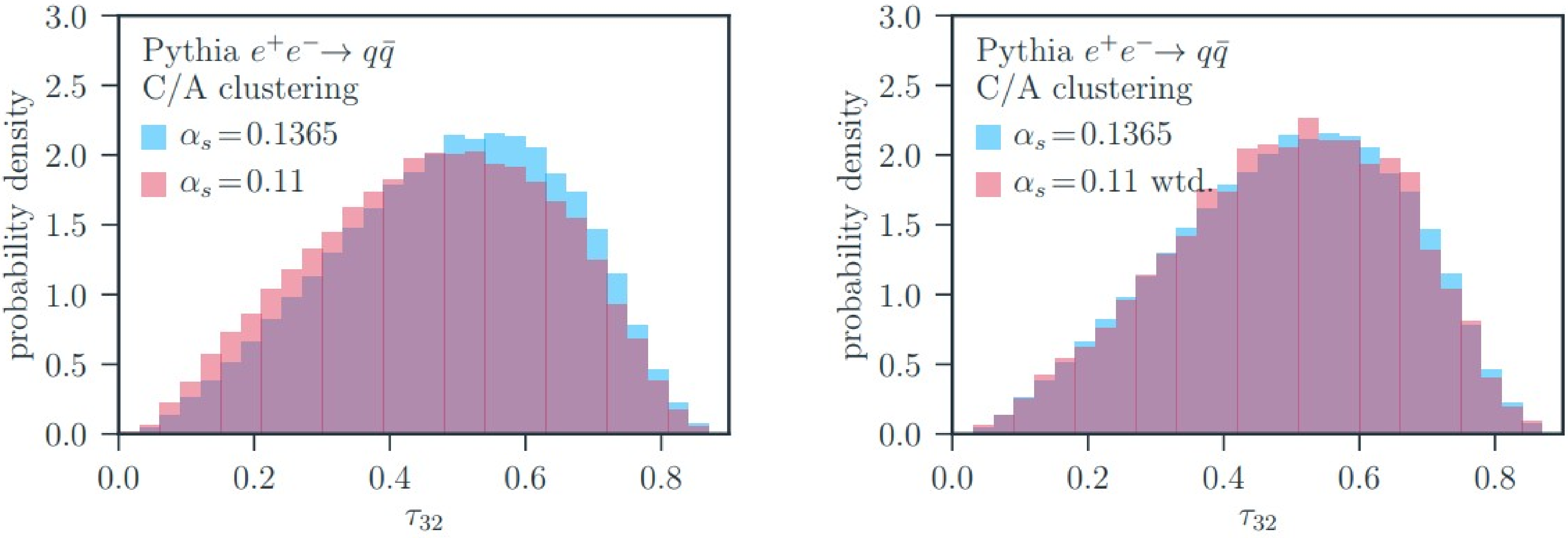
- Access to low-level physics directly
- Intuitive and thus appealing



- Average Lund Plane (many jets)
  - one jet = set of point in this plane
- Can we reweight jets using the ratio of Lund Planes in data and MC?
- Unfolded 2D distribution of the Lund Plane????

JUNIPR

- Recursive NN, unsupervised learning on data
  - (A talk on more advance JUNIPR reweighting later this week!!!)
- Data/MC reweighting was one of its main goals!



- Works in MC: turns one PYTHIA into another.
- Will it work in data?

Do we need QCD MC at all?

- For multijet background estimates, we don't need MC
  - Have been data-driven anyway
  - Although there could be subtle correlations…
- Unsupervised learning from data…
  - Learns QCD: e.g., autoencoder with LoLa
  - Learns QCD in the presence of other backgrounds: e.g., CWoLa
- Can we interpolate between two sidebands
  - e.g., CWoLa hunting
- Can we *extrapolate* from one CR to another???

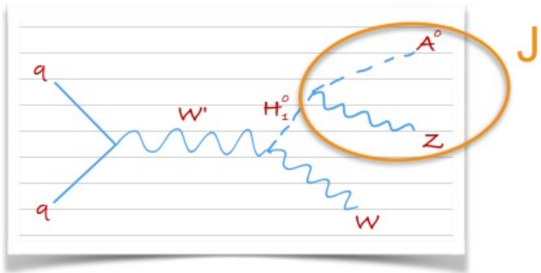
N-pronged jets?

- The future of searches with substructure?

(from Juan-Antonio Aguilar Saavedra)

Case II: Merged multibosons

If intermediate particles are `light´, their decay products are merged

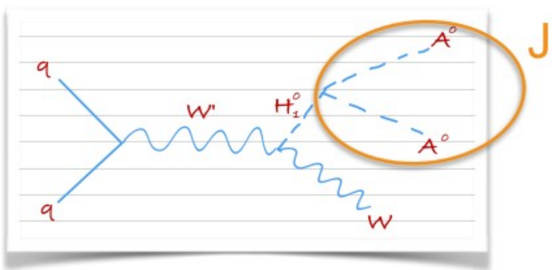


$$M_{W'} \gg M_{H_1^0} \gtrsim M_Z + M_{A^0}$$

$Z \to qq$ / …

$A^0 \to bb$

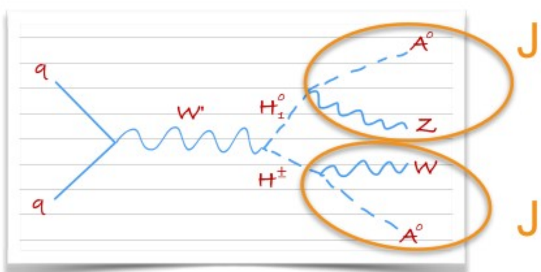
W + fat jet J



$$M_{W'} \gg M_{H_1^0} \gtrsim 2M_{A^0}$$

$A^0 \to bb$

W + fat jet J



$$M_{W'} \gg M_{H_1^0}, M_{H^\pm} \gtrsim M_Z + M_{A^0}$$

$W, Z \to qq$ / …

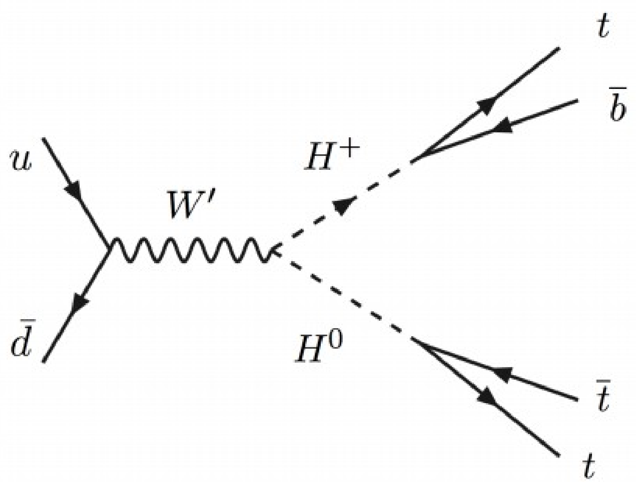
$A^0 \to bb$

two fat jets J

More N-pronged jets?

- More cool signatures with 4- and 6-pronged jets

Heavy Higgs bosons decay directly into a pair of the heaviest fermions:



For $M_{W'} \gg M_{H^+}$:

$(t\bar{b})$-tagged jet + $(t\bar{t})$-tagged jet

For $M_{W'} > M_{H^+} \gg m_t$:

three t-tagged jets + b

(from Bogdan Dobrescu)

- Easy to do a cut-based analysis (let alone a DNN)
- But how to get the efficiency?

What to do about N-pronged jets?

- Give up, can't be done:
  - Can't measure efficiency in data
  - These analyses are always going to be out of reach
- Report limit on $\sigma_X \cdot \mathcal{B}(X) \cdot \underline{\epsilon_X}$ (David Miller's suggestion)
  - Let the consumers of the paper worry about the signal efficiency
  - Would not affect the discovery, only limits
  - May actually spur progress in this area :-/
- Or try to make it work?
  - Learn how to reweight single quark jets from MC
  - Verify that the procedure works for W and top (2,3-prong)
  - Assign further systematics for 4,5,6-prong...