Heavy flavor jets

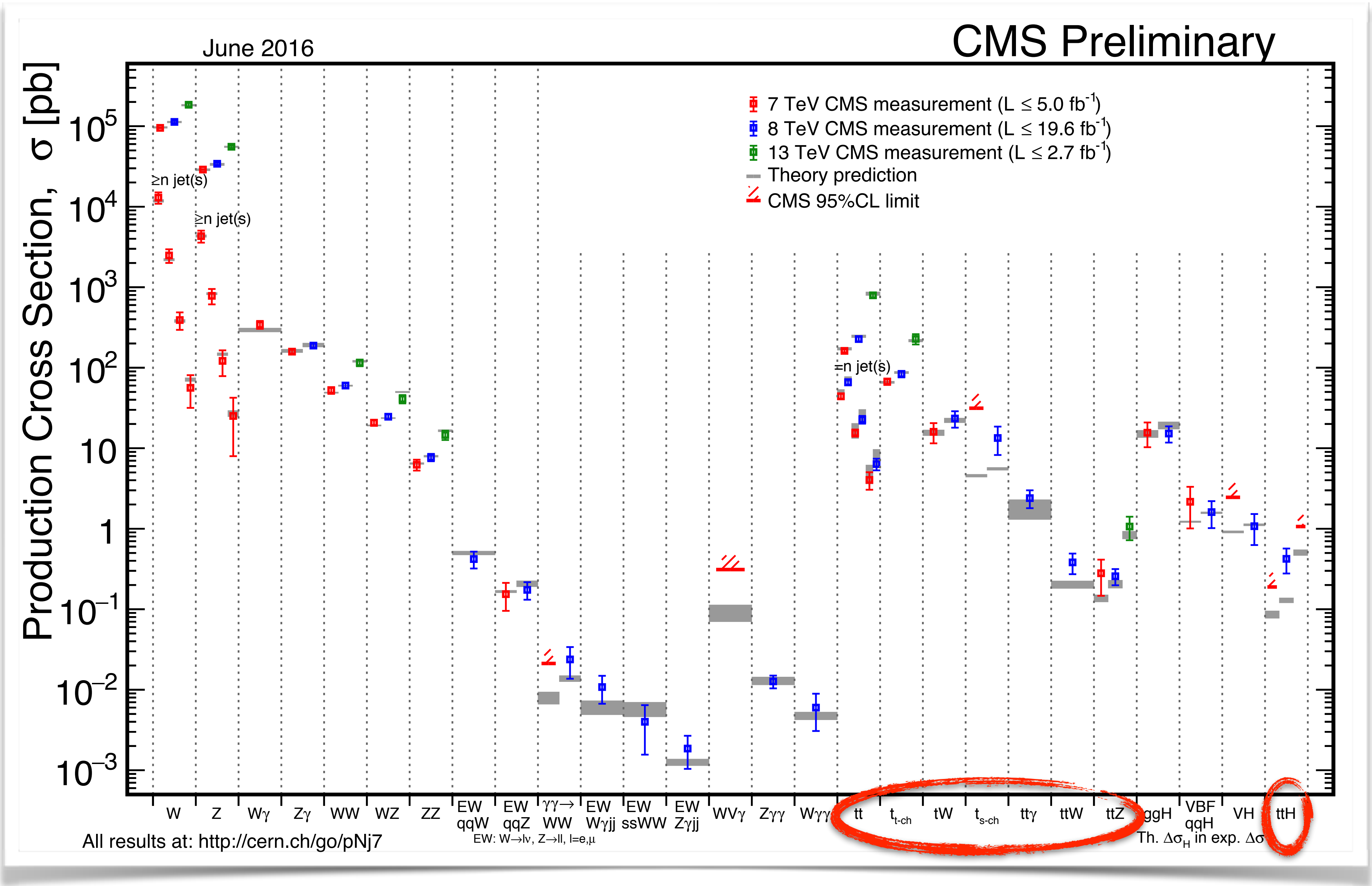
Adam Kardos

University of Debrecen,
MTA-DE Particle Physics Research Group

SMP-J annual workshop
25th, January, 2017

Tops are all around us

Cross section measurements are already done for several top related processes:



Source: http://cern.ch/go/pNj7

Too hot to handle

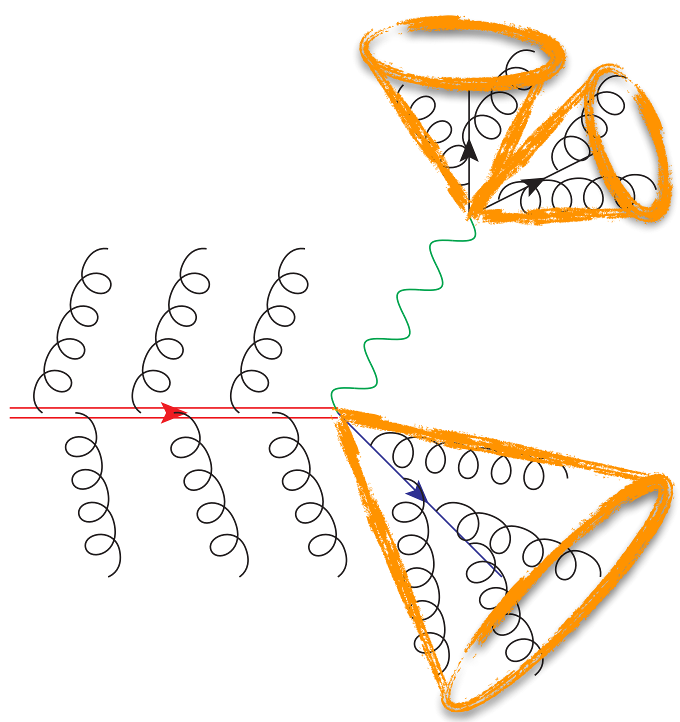
- Top decays before hadronization with really short lifetime
  ⇒ no secondary vertex to look for
- Reconstructing top 4-momentum is tough!

The traditional way: semileptonic channel, lepton trigger + b-tagging

- Issue: all MET is associated to the neutrino coming from W-decay!
- The MET can come from elsewhere too:

  Heavy top partner decaying into top, $t\bar{t}$ + dark matter production,…

In the hadronic channel the behavior is different at low and high p_T.



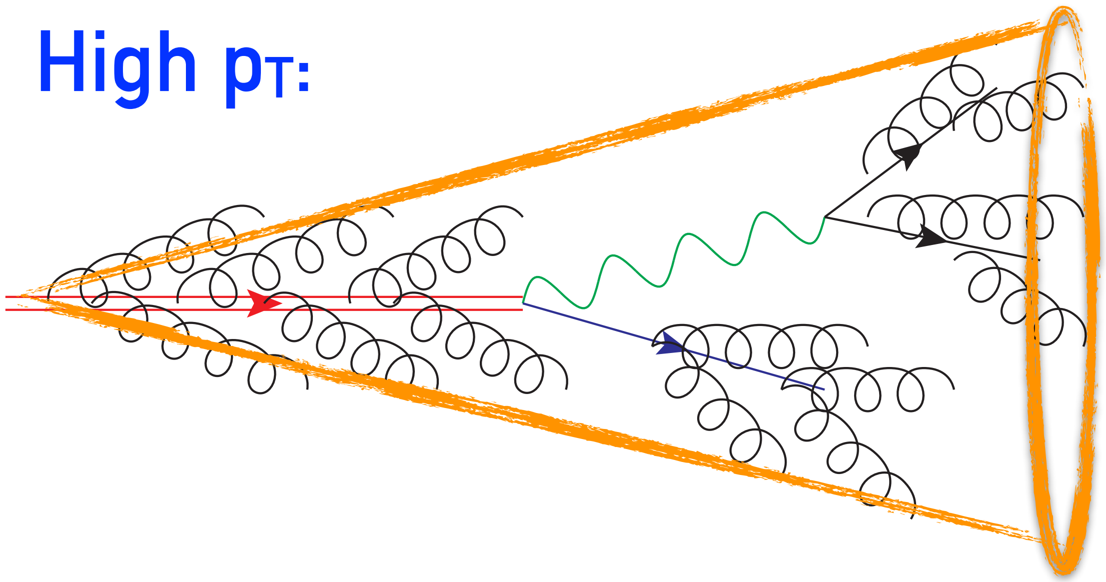
Low p_T:

- individual jets with R ~ O(0.1)
- radiation off top is limited due to the dead cone

$$R_{\rm d.c.} \propto \frac{m_t}{p_{\perp,t}}$$

Boosted region

High p_T:



top decay products form one jet: R~O(1)

Within the fat jet products form subjets

⇒ Non-trivial internal structure develops

Structure can be extracted from calorimeter towers and/or tracks.

With increasing p_T collimation increases

⇒calorimeter granularity sets upper limit on p_T range

⇒At very high p_T only track info can be used

Top tagging efficiency breaks down at very high p_T (~ 1TeV):

- Calorimeter granularity is not enough to resolve structure of fat jet
- Radiation off top fills up area between subjets

The idea of a top tagger

- At high p_T top decay products form one single fat jet.
- The fat jet should contain three subjets with non-QCD splitting origin
  - Subjets are found, e.g. by performing unclustering
    - ⋆ Subjets should fulfill certain criteria: separation, mass drop, symmetry, hardness,. . .
  - QCD noise removal in clustered subjets
    - ⋆ Trimming, filtering and pruning or a combination of these
- Sum of 4-momenta of the three selected subjets is the 4-momentum of the top candidate
  - Possible additional cut on the mass, p_T and η of the top candidate

For a comprehensive review and historical recount see, e.g.,
Plehn & Spannowsky, J.Phys. G39 (2012) 083001

The usual suspects

Processes deserving analysis in boosted regime:

- $t\,\bar{t}$
- $t\,\bar{t}$ + jet(s)
- $t\,\bar{t} + b\,\bar{b}$
- $t\,\bar{t} + H$
- $t\,\bar{t} + V\,(+\,V')$, $V, V' = Z, W$
- $t\,\bar{t} + \gamma$(s)

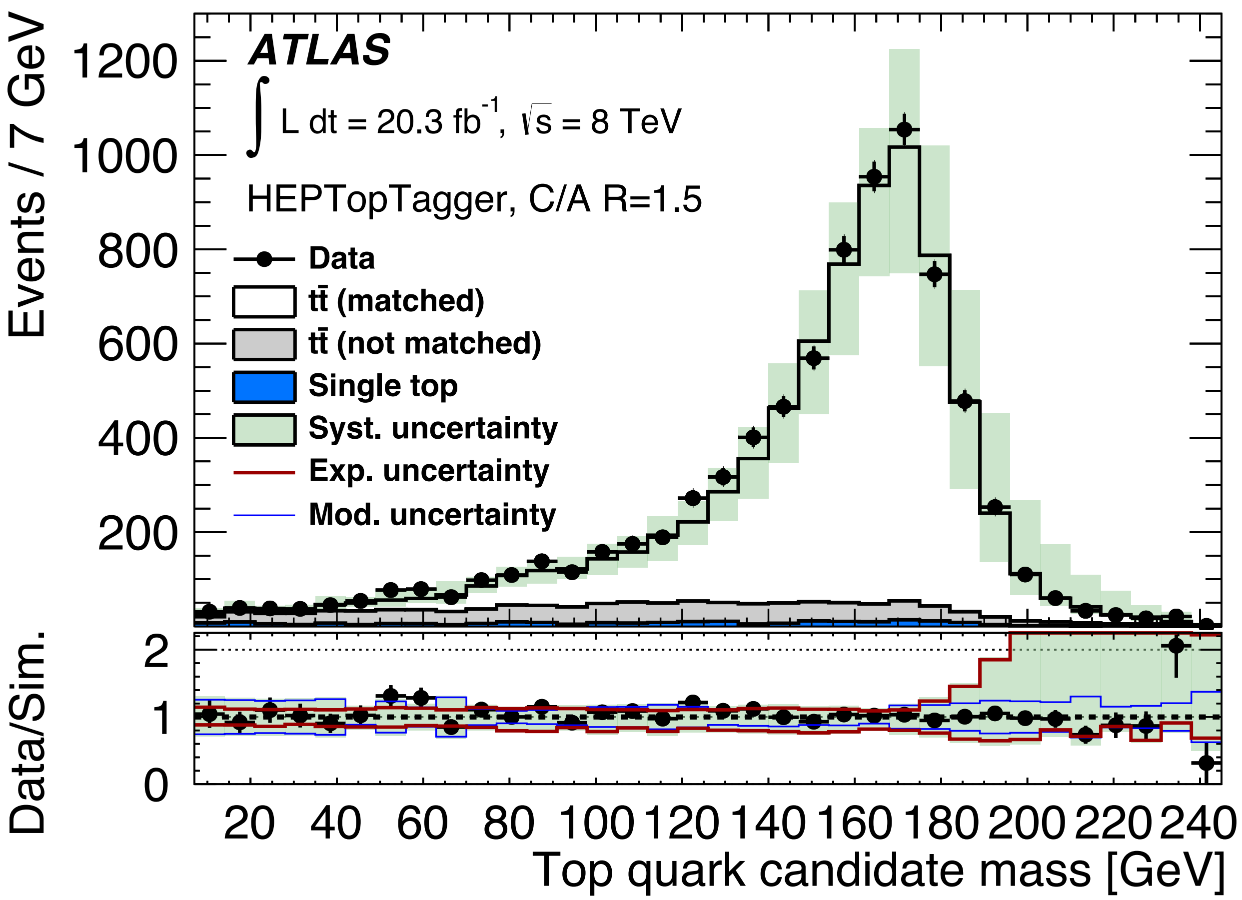
⋆ $t\,\bar{t}\,t\,\bar{t}$

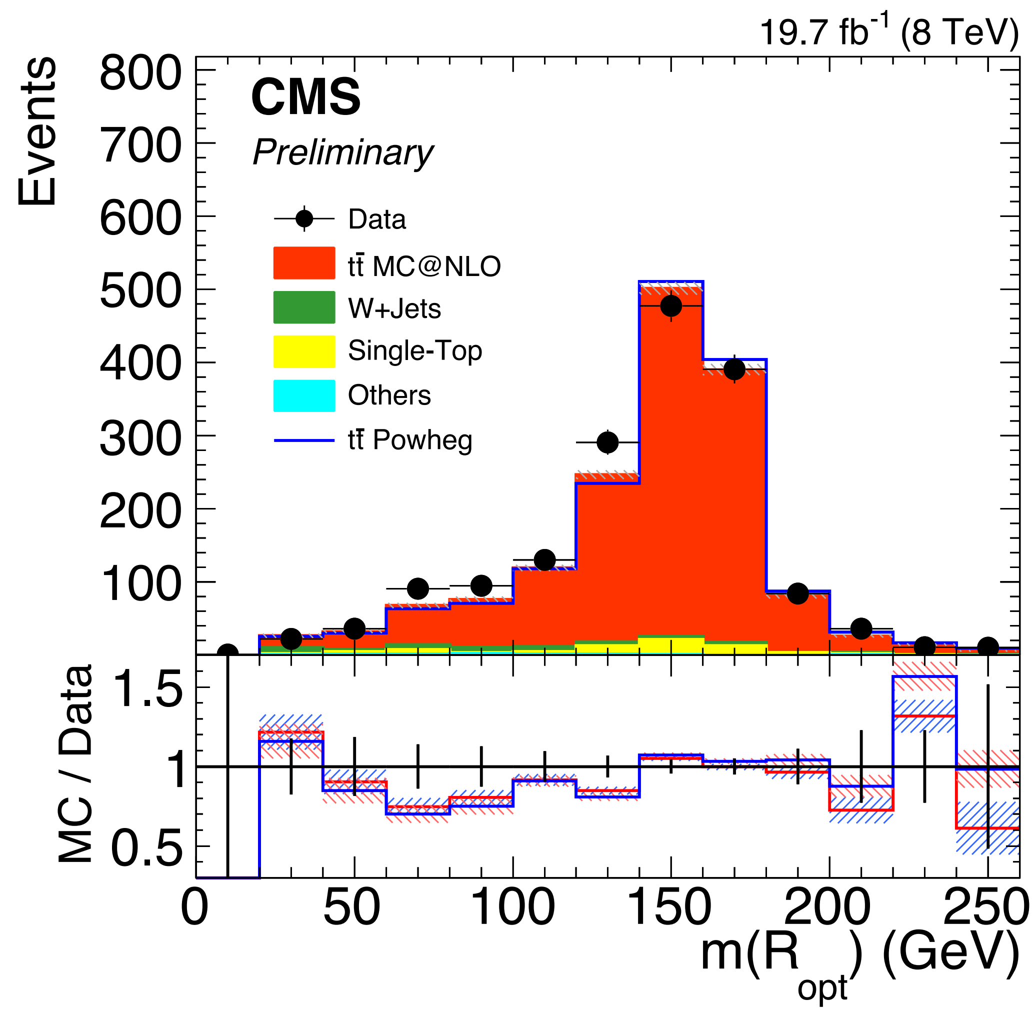
To provide meaningful comparisons predictions have to be on the hadron level $\Longrightarrow$ NLO + PS is needed for all these processes and matched to some hadronization models too!

$t\bar{t}$ production

Already studied in the boosted regime both by ATLAS and CMS:

(JHEP 1606 (2016) 093,…, CMS PAS JME-15-002,…)





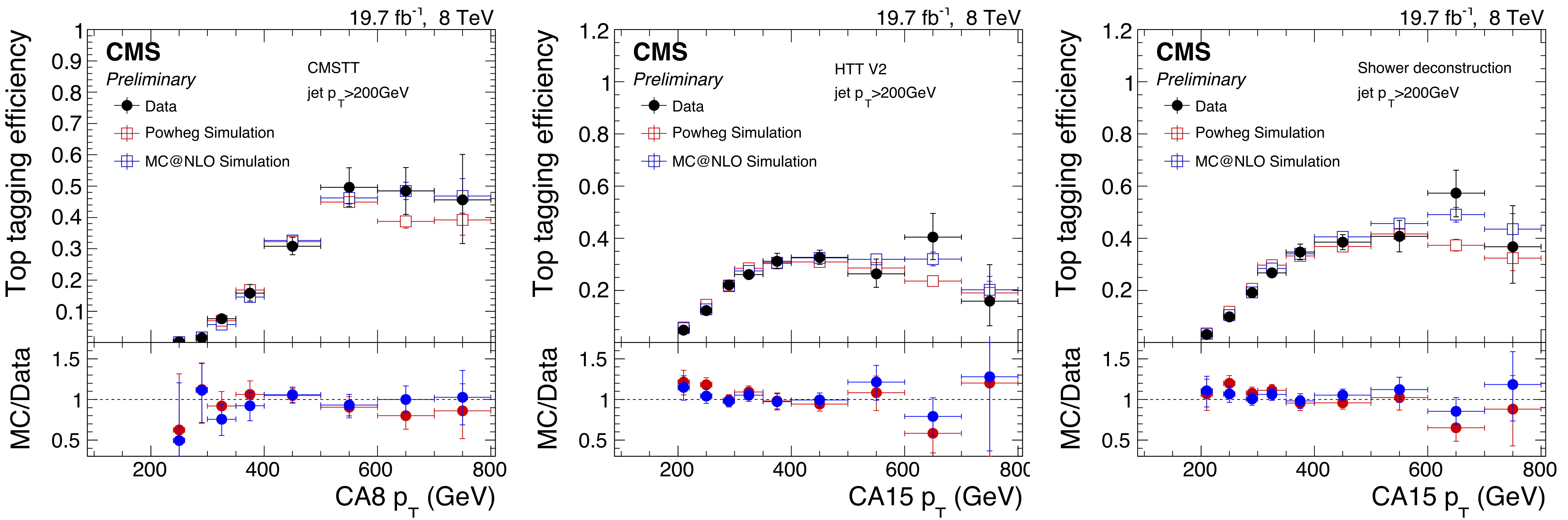
Several top tagging techniques are being used with very good efficiencies ~30-50%

ATLAS claims a ~70% b-tagging efficiency with a neural network driven algorithm!

High b-tagging efficiency is very welcome when tagging multiple b's

$t\,\bar{t}$ production

The tagging efficiencies seem to saturate at high p_T:



From theoretical studies we know there is break-down around 1 TeV

(Schatzel & Spannowsky, Phys.Rev. D89 (2014) no.1, 014007)

At least for HEPTopTagger v1, though expected due to finite granularity of the calorimeter

A simulation with p_Ts above 1TeV would be interesting with detector effects and comparison to HPTTopTagger.

$t\bar{t}$ + jet(s) production

- Just like $t\bar{t}$ production but with one or more extra energetic jet(s).
- Predictions are available at hadron level:
  - `PowHel`, $t\bar{t}j$ (Phys.Lett. B705 (2011) 76-81)
  - `POWHEG-BOX`, $t\bar{t}j$ + decay(JHEP 1201 (2012) 137)
  - `MadGraph5_aMC@NLO`, $t\bar{t}j$ + decay (JHEP 1407 (2014) 079)
  - `Sherpa` + `OpenLoops`, $t\bar{t}j$ & $t\bar{t}jj$ + decay (Phys.Lett. B748 (2015) 74-78)
  - `HELAC-NLO`, $t\bar{t}j$ matched to the Nagy-Soper shower (JHEP 1506 (2015) 033)
- It is possible to generate events with on-shell tops or decay them in the NWA
- Interested in the boosted regime and top p_T distribution spin-correlation among top decay products does not have significant effect.
- Generate on-shell tops and decay in the SMC (cheaper in CPU time)!

$t\,\bar{t} + b\,\bar{b}$ production

- Predictions are available at hadron level:
  - `PowHel`, with massless b's (J.Phys. G41 (2014) 075005, JHEP 1503 (2015) 083)
  - `Sherpa + OpenLoops`, with massive b's (Phys.Lett. B734 (2014) 210-214)
  - `MadGraph5_aMC@NLO`, with massive b's (JHEP 1407 (2014) 079)
- Most important background to $t\,\bar{t}\,H$ production when $H \to b\,\bar{b}$
- For QCD people this process is interesting in its own!
- It is not so easy to get measured:
  - At Born level: $\sim\alpha_S^4$
  - Requires 4 b-tags, high efficiency is key in b-tagging
- Some people state that NLO+PS is not enough, better brace yourselves: NNLO+PS for this process is far-far away.

"The darkest hour is just before the dawn"

$t\bar{t} + b\bar{b}$ production

A note on b's:

b mass acts as a regulator: collinear singularity → Sudakov-log

No hard separation is needed for massive b-pair. In the quasi-collinear case a $t\bar{t}b$ configuration can emerge (two b's in one b jet).

A $t\bar{t}b$ configuration can still give contribution to 4 b-jet case: gluon splitting into a b-pair in shower (Phys.Lett. B734 (2014) 210-214)

We can have b's from different places:

- Top decay
- From hard process (from gluon splitting)
- From shower (also from gluon splitting)

It would be nice to distinguish between b-jets coming from different sources

What if a b-jet is composed of two b's? Is there a way to detect this?

$t\bar{t} + H$ production

- Predictions are available at the hadron level:
  - `MadGraph5_aMC@NLO` (Phys.Lett. B701 (2011) 427-433, JHEP 1407 (2014) 079)
  - `PowHel` (JHEP 1211 (2012) 056)
  - `Sherpa` + `OpenLoops` (JHEP 0902 (2009) 007, Phys.Rev.Lett. 108 (2012) 11160 JHEP 1401 (2014) 046)
  - `POWHEG-BOX` (Phys.Rev. D91 (2015) no.9, 094003)
- All frameworks can provide events with on-shell tops or decayed through NWA
- If $H \to b\bar{b}$ is considered boosted analysis is needed for the top(s) and for the Higgs: two fat jets, 4 b-tags
- In case of $H \to \gamma\gamma$ only one fat jet is requested, photon isolation is key in NLO+PS, the isolation has to be IR-safe and free of fragmentation contribution (no perturbative calculation, hard to measure)

$t\bar{t} + H$ production

- When top and Higgs are in the boosted region with a hadronically decaying Higgs tagging is needed for both!
- Current setup: find two fat jets and apply tagging separately
- Since tagging is needed for both top and Higgs a multi-object tagger would be a better choice
- Since b jets should appear in both fat jets a multi-object tagger seems a better choice to minimise ambiguities coming from wrongfully assigning b-jets to decays.
- In the development of multi-object taggers deep learning can help us (see the talk by Matthew Schwartz and also JHEP 1607 (2016) 069)

$t\bar{t}$ + V (+ V') production

- Predictions are available at the hadron level:
  - `PowHel` (Phys.Rev. D85 (2012) 074022, JHEP 1211 (2012) 056)
  - `Sherpa` + `OpenLoops` (JHEP 1401 (2014) 046)
  - `MadGraph5_aMC@NLO` both $t\bar{t}V$ and $t\bar{t}VV'$ (JHEP 1407 (2014) 079)
- Important background to $t\bar{t}H$ production in various channels ($H \to b\bar{b}$, $H \to WW^*$, $H \to \tau\bar{\tau}$)
- Also important in BSM searches (e.g. same sign lepton + MET + jets)
- Event generation is technically more difficult than for $t\bar{t}$ but conceptually not
- Problem can arise to generate enough events in high top p_T region, in POWHEG-BOX you can use suppression to enrich generation in tail:

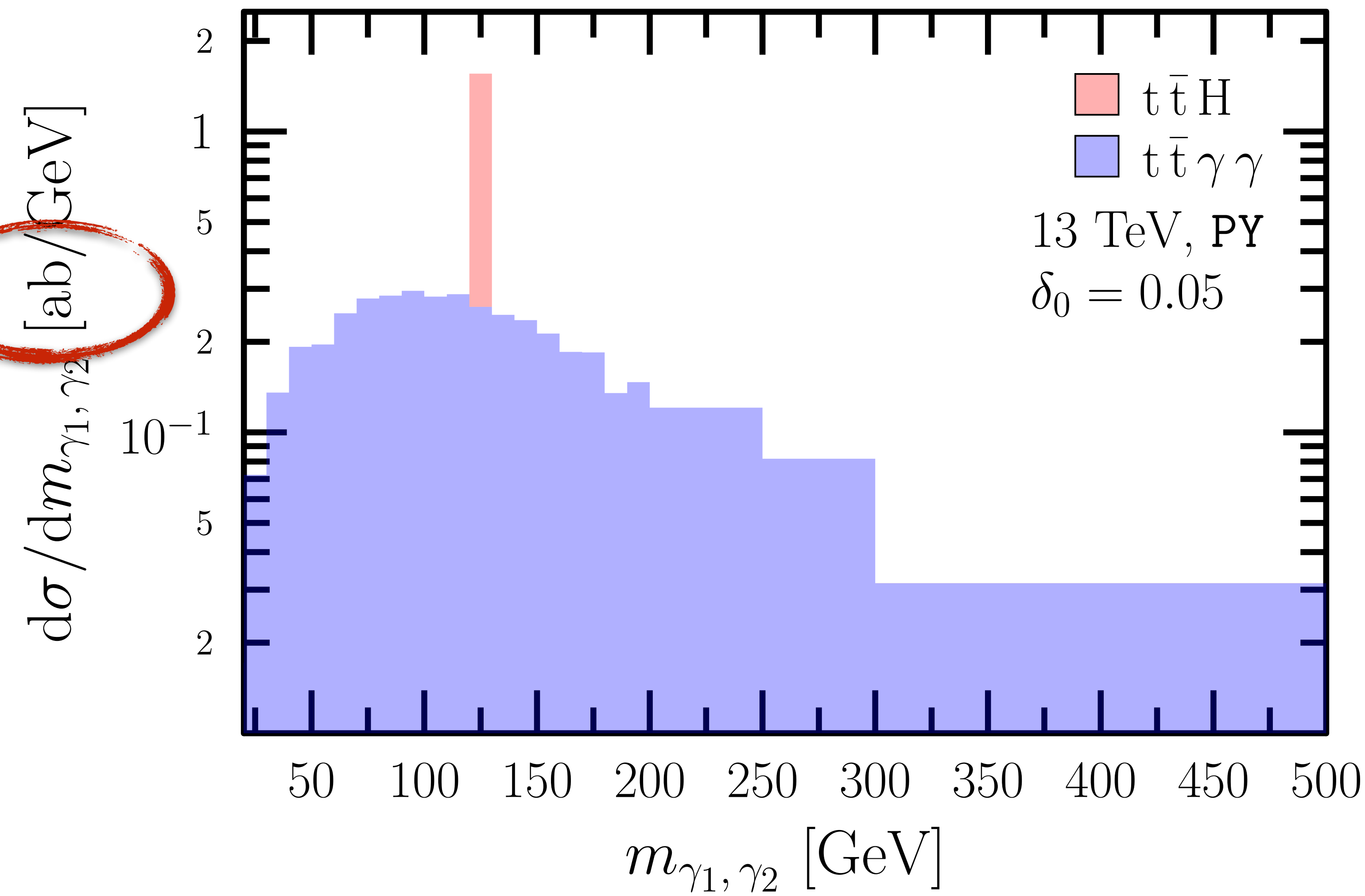
set1: $$\mathrm{supp}_t = \frac{1}{\left(1 + \frac{\left(p_\perp^{\mathrm{supp}}\right)^2}{p_{\perp,t}^2}\right)^i}$$ set2: $$\mathrm{supp}_{\bar{t}} = \frac{1}{\left(1 + \frac{\left(p_\perp^{\mathrm{supp}}\right)^2}{p_{\perp,\bar{t}}^2}\right)^i}$$ i={1,2}

$t\bar{t} + (2)\gamma$ production

- Predictions are available at the hadron level:
  - `PowHel` (JHEP 1505 (2015) 090, Nucl.Phys. B897 (2015) 717-731)
  - `MadGraph5_aMC@NLO` (JHEP 1407 (2014) 079)
- Most crucial background to $t\bar{t}H$ when $H \to \gamma\gamma$
- Due to the photon(s) isolation is needed. Fragmentation is hard to measure $\Longrightarrow$ try to avoid it $\Longrightarrow$ use Frixione-isolation
- Since detector granularity is finite a discrete version is needed: Eur.Phys.J. C31 (2003) 491-502
- Use loose isolation to generate events and more stringent in analysis
- The cross section should be independent of a variation of isolation used during generating events! (a way to check event sample)
- It is easy to go beyond reach with too stringent cuts:

$t\bar{t} + 2\gamma$ production

$t\bar{t} + 2\gamma$ production @ 13 TeV:



- HEPTopTagger v1 with C/A R=1.5, $p_T > 200$ GeV, $|y| < 5$ jets
- $p_{T,\gamma} > 30$ GeV, $|y_\gamma| < 2.5$, $\Delta(\gamma_1,\gamma_2) > 0.4$, jets: anti-$k_T$, R=0.4, $p_{T,j} > 30$ GeV
- One hard lepton: $p_{T,l} > 30$ GeV , $|y_l| < 2.5$
- Separations: $\Delta(\gamma,l) > 0.4$, $\Delta(\gamma,j) > 0.4$
- Minimal (3 GeV) hadronic activity is allowed in lepton isolation cone

Conclusions

- Great to see all this activity around top jets!
- Increasing efficiency in b- and t-tagging makes analyses in the boosted region more and more feasible
- For all important top-pair related processes NLO+PS results are available
- Predictions were usually made for tops with p_T less than ~500 GeV
- For boosted analysis events with higher top p_T are needed up to ~800 GeV
- Event generation efficiency is important, use suppression to enrich sample with events in analysis region
- Sophisticated final states demand sophisticated methods to tag them $\Longrightarrow$ looking into the dawn of multi-object taggers

Thank you for your attention!