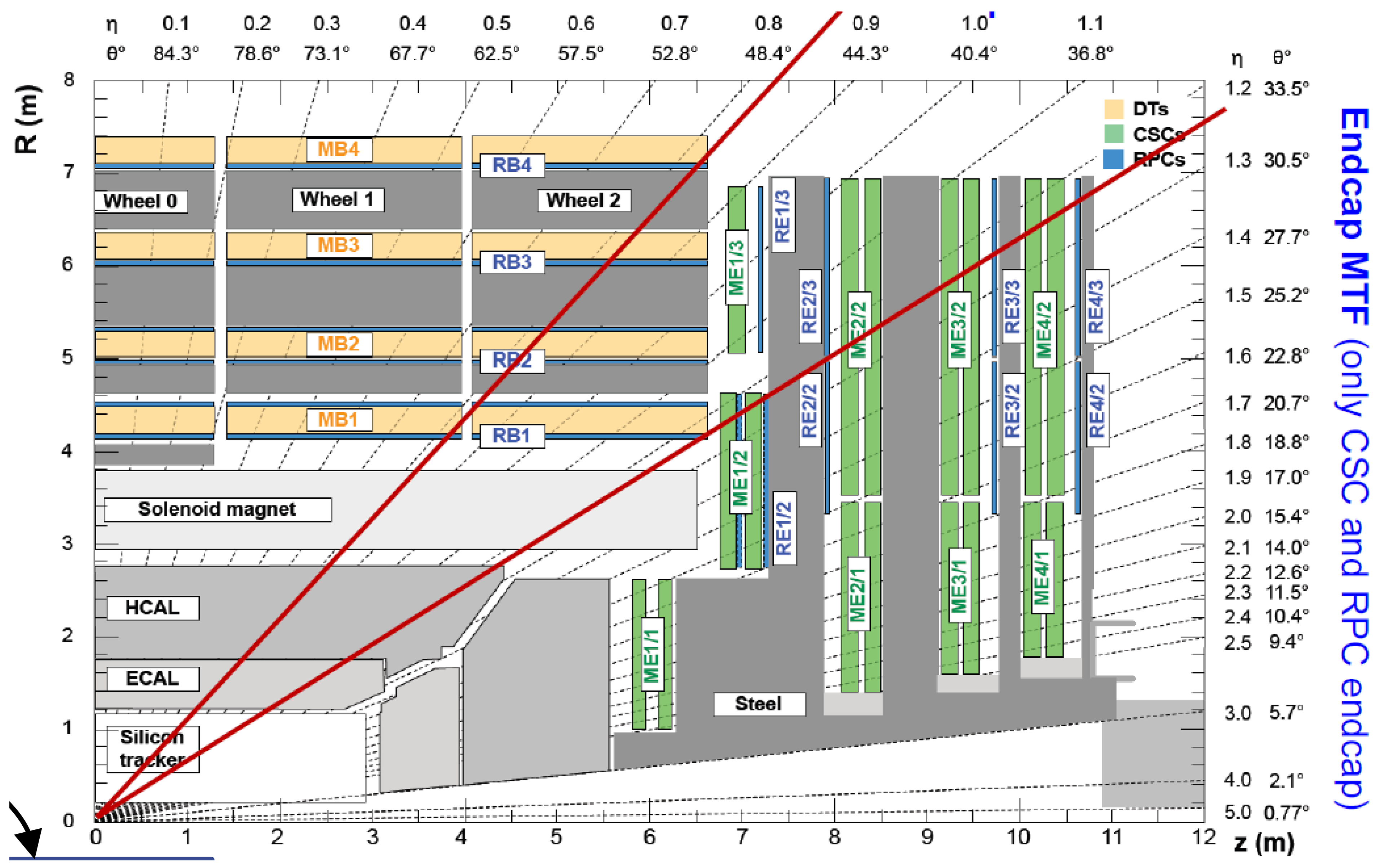
BDTs in the CMS Level 1 Muon Endcap Trigger

Darin Acosta, Andrew Brinkerhoff, Elena Busch, Andrew Carnes,
Ivan Furic, Khristian Kotov, Jia Fu Low, Alexander Madorsky,
Jamal Rorie, Bobby Scurlock, Wei Shi



CMS Level 1 trigger ("L1T") selects 100k interesting events out of 40 million collisions every second at the LHC

- Goal is ≥ 90% efficiency for events we want to keep
- Keep high transverse momentum ("pT") particles, reject low pT events

Quick trigger decision: only 3.0 μs for the entire process

- Fast logic algorithms in FPGAs perform complex analysis: from raw detector inputs, reconstruct, select, and assign pT to particles

Muons leave tracks as they pass through the CMS detector

- Assign pT based on curvature in magnetic field: more bend ↔ low pT
- In L1 trigger, can only use outer muon chambers, not inner tracker

Typical "interesting" event has a single muon with pT > 25 GeV

- For every real 25 GeV muon, there are 1000 muons with pT = 5 GeV!
- Critical to reject as many low-pT muons as possible

Endcap Muon Track Finder (EMTF) creates tracks from hits in 4 forward "stations"

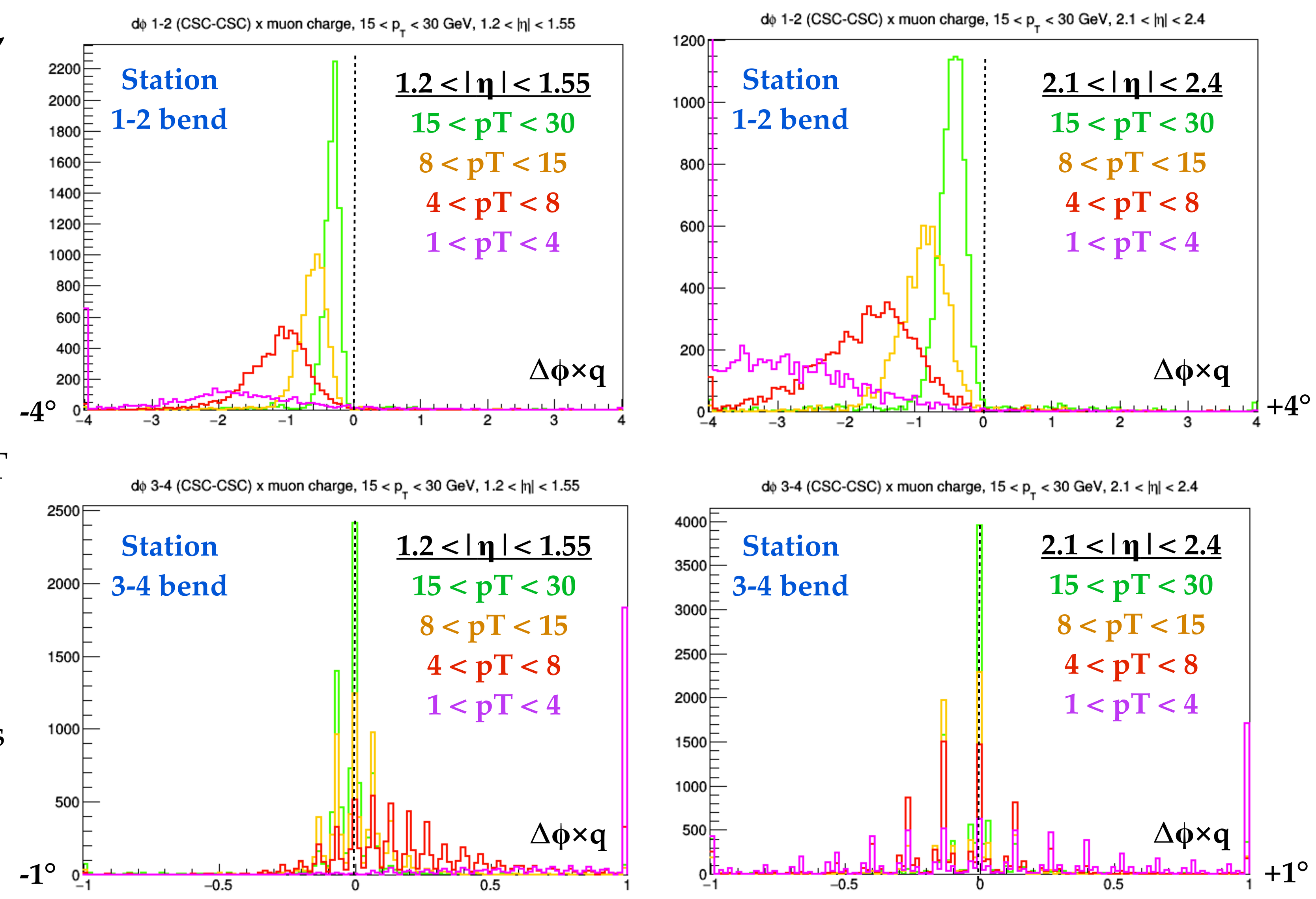
- CSC or RPC hits correlated in θ and ϕ between stations belong to the same muon
- Typically < 1° of bending in θ, up to 12° in ϕ due to magnetic field

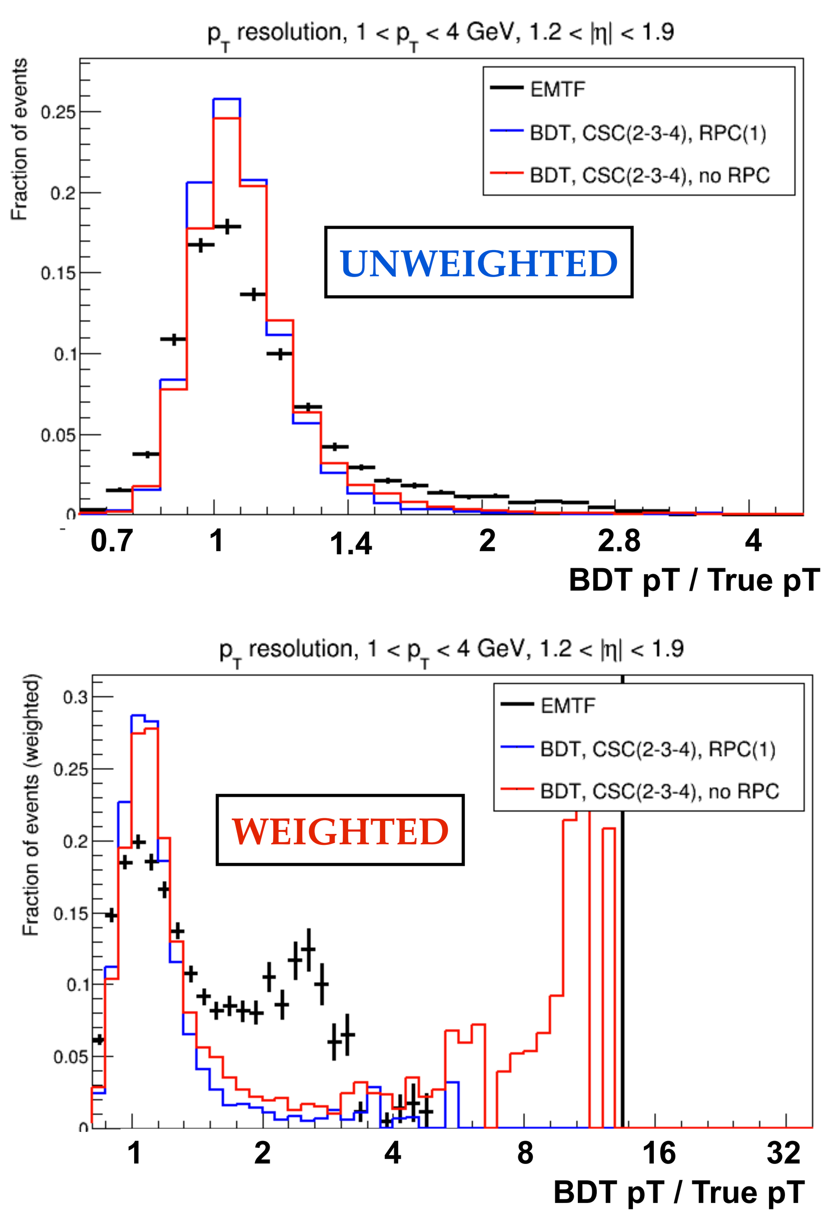
Assign pT mostly using bend in ϕ between stations vs. track θ (or η)

- Magnetic field strength varies with η, but always strongest between stations 1 and 2
- Δϕ(1-2) is consistently "negative" (w.r.t. the muon charge), larger for low pT
- Δϕ(2-3) and Δϕ(3-4) much smaller, more scattering (positive or negative bend) at low pT
- Very different behavior between |η| < 1.55 (more central) and |η| > 2.1 (forward)

Complicated dependencies make this an ideal case for machine learning

- But have only ~500 ns to run entire EMTF algorithm: track-building + pT assignment
- Instead, use Boosted Decision Trees (BDTs) offline to assign pT for 1 billion track profiles
- Write values to 1.2 GB look-up table (LUT), with 30-bit address based on input variables
- Requires some compression: Δϕ(1-2) into 7 bits, Δϕ(2-3) into 5 bits, etc.





Training Feature: Resolution

Optimization may seem straightforward

- Choose the algorithm that assigns pT closest to the true pT (best "resolution")

Reality is much more complicated

- Background muon pT spectrum falls as pT^{-4} or pT^{-5}, so very-low-pT muons with over-assigned pT cause the most rate
- Compare unweighted resolution (top) to resolution weighted by $(\text{BDT} / \text{true pT})^3$ -- suddenly see huge difference in tails

Strategies to reduce high-pT tails

- Re-weight training sample to fall as pT^{-2}
- Target 1/pT in regression instead of pT -- inflates difference for low pT muons

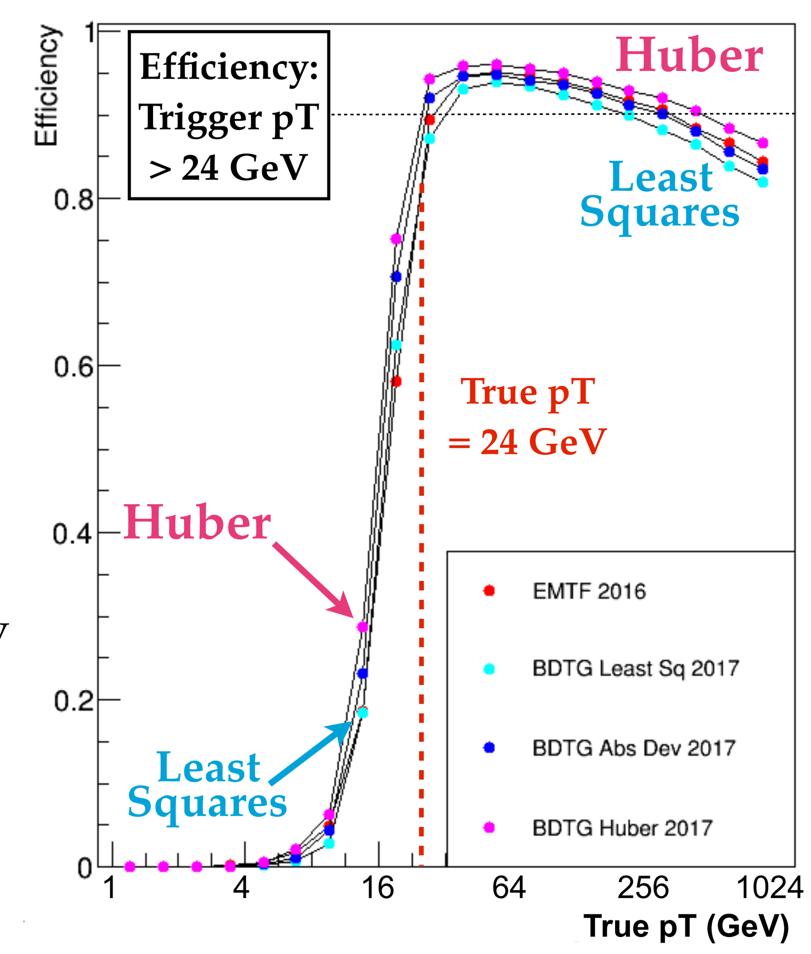
Training Feature: Loss Function

Choice of "loss function" also plays a role

- BDT minimizes Σ (loss) in training events
- loss = $(\text{true pT}^{-1} - \text{BDT pT}^{-1})^2$, or
  loss = $|\text{true pT}^{-1} - \text{BDT pT}^{-1}|$, or
  "Huber" : $(\,)^2$ in the center, $|\,|$ in the tails

Playing off rate vs. efficiency

- Functions that don't penalize tails (e.g. Huber) better in the "plateau" -- efficiency to assign pT > 24 GeV when true pT > 30
- But Least Squares has sharper "turn-on" -- fewer muons with over-assigned pT, and thus a lower background rate



RESULTS

In 2016, reduced background rate by a **factor of two** compared to 2015!

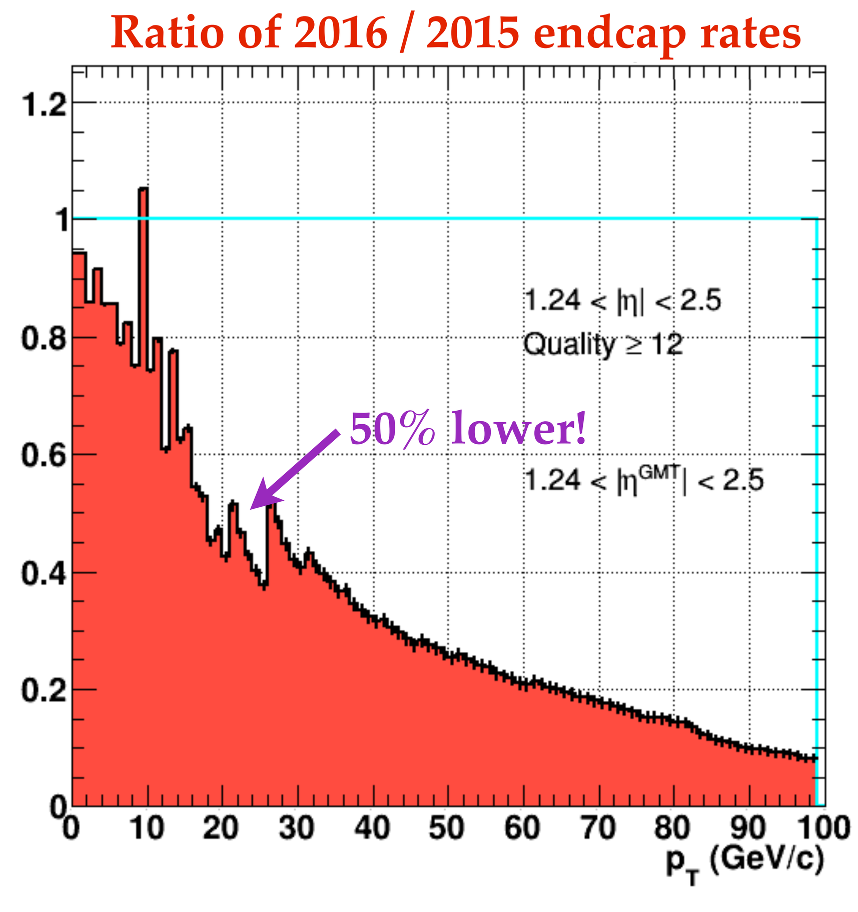
- Most of rate reduction comes from using BDTs to assign pT
- Small efficiency loss because RPC hits were not yet included in EMTF

In 2017, with RPC hits and new BDTs: even **lower rates**, **higher efficiency**

Plus many ideas for future improvements!

- More intelligent algorithms (e.g. DNNs), custom rate vs. efficiency loss functions, training directly on high-statistics data ... and much more!

Ratio of 2016 / 2015 endcap rates



EMTF trigger efficiency