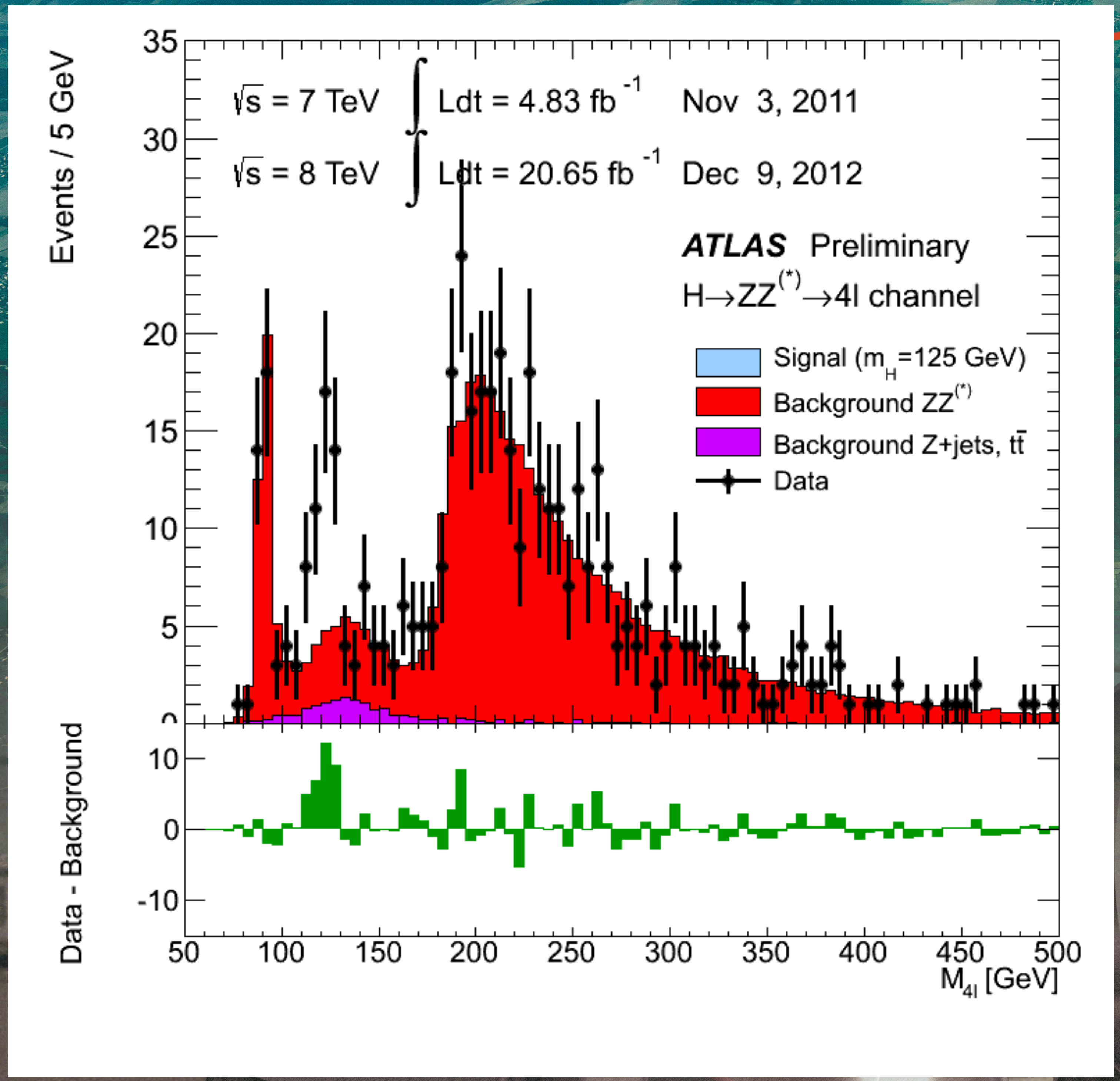
Uncovering New Massive Particles through Boosted Object Tagging at the LHC

Application of the Boosted Event Shape Tagger In Searches for Vector Like Quarks and Top-Coupling Heavy Resonances

Johan Sebastian Bonilla Castro
(They/Them)
University of Chicago, Rising Stars Symposium
09/23/2021

Large Hadron Collider at CERN



13 TeV proton collider
27 km

Run 1 uncovered
Higgs-like scalar
@125 GeV

Run 2 analyses
underway

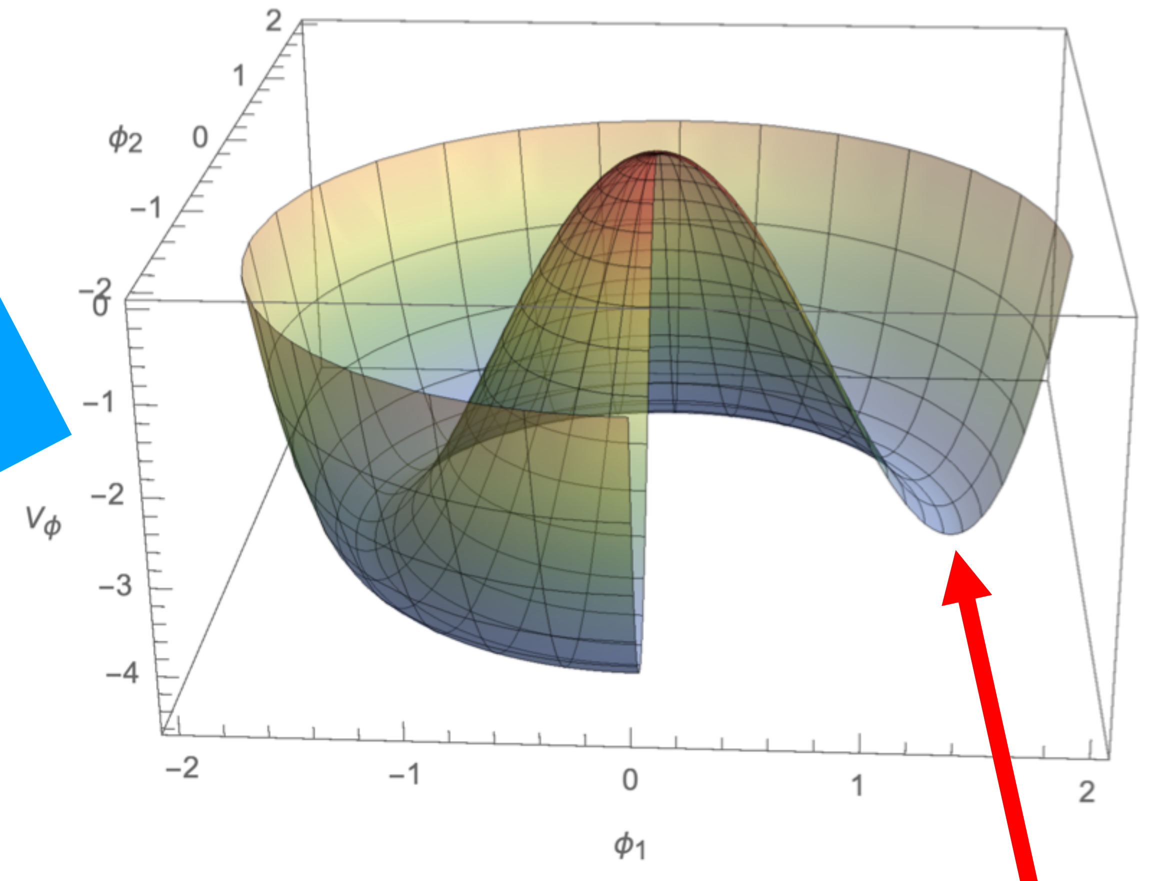
Run 3 starts
Spring 2022

(Un-)Naturalness of the Standard Model

Motivation for BSM Physics

$$V(\Phi_{\text{Higgs}}) = -\frac{1}{2}\mu^2|\Phi_{\text{H}}|^2 \ldots$$

$$\langle\Phi\rangle = \nu \ldots$$



ν

Can be measured from Weak Interactions

$$m^2_{h,phys} = \ldots \sum_{Fermions} k_f m_f^2 + \sum_{Bosons} k_b m_b^2$$

$Tune$ – $\sim O(\Lambda^2_{UV})$ + $\sim O(\Lambda^2_{UV})$

$\sim \text{GeV}^2$

For ANY choice of Λ_{UV}!

But, what if this 125 GeV scalar is non-fundamental?

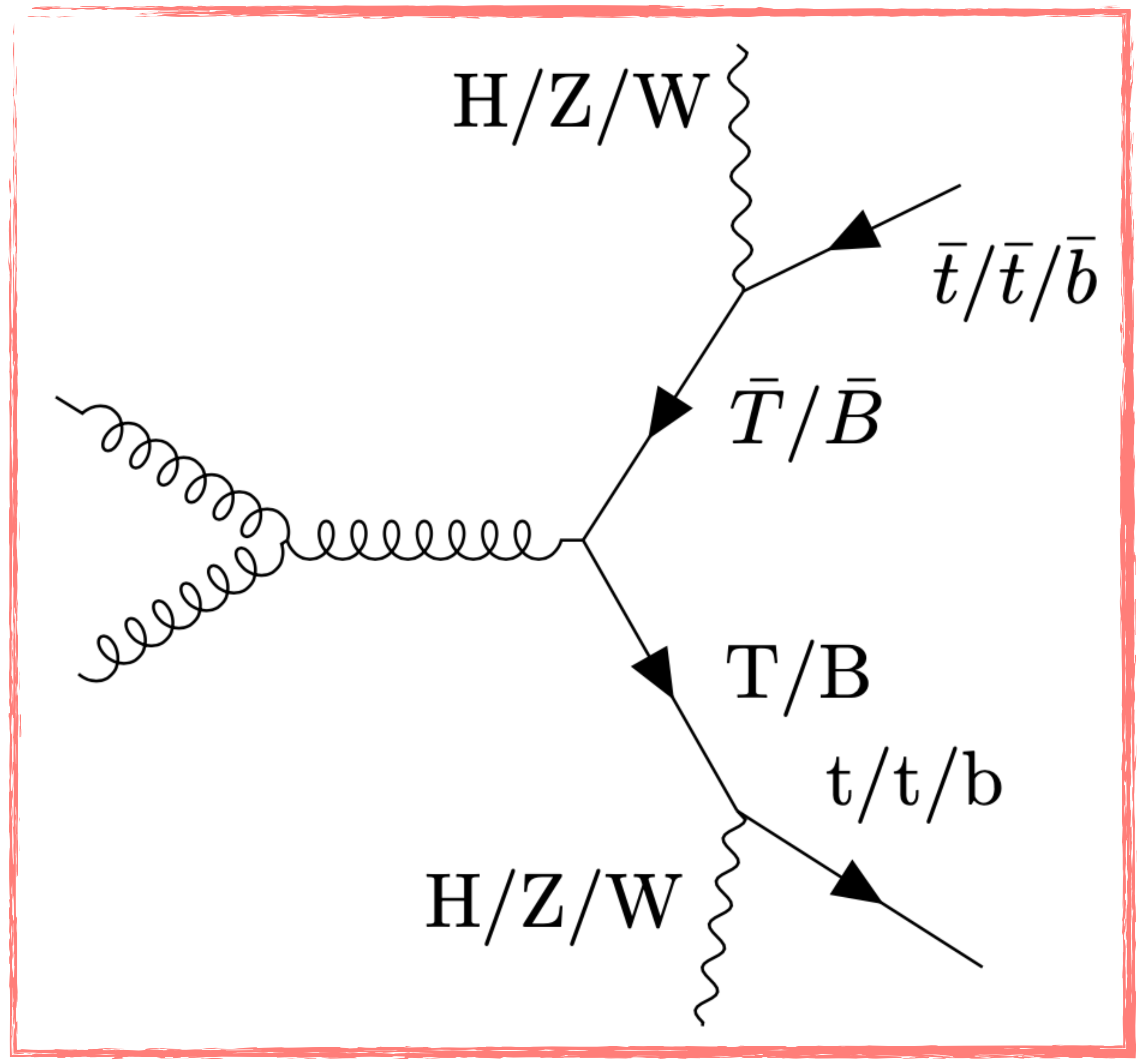
Fixing the Higgs Loop Divergence

With New Heavy Massive Top Partners

Loop contributions from SM particles $\propto y_{SM} \sim m_{SM}^2$

— The top quark -> largest term

— Idea: mitigate top contribution with top partner



Question: is the 125 GeV scalar fundamental?

— Pseudo-Nambu-Goldstone boson -> Composite Higgs Models

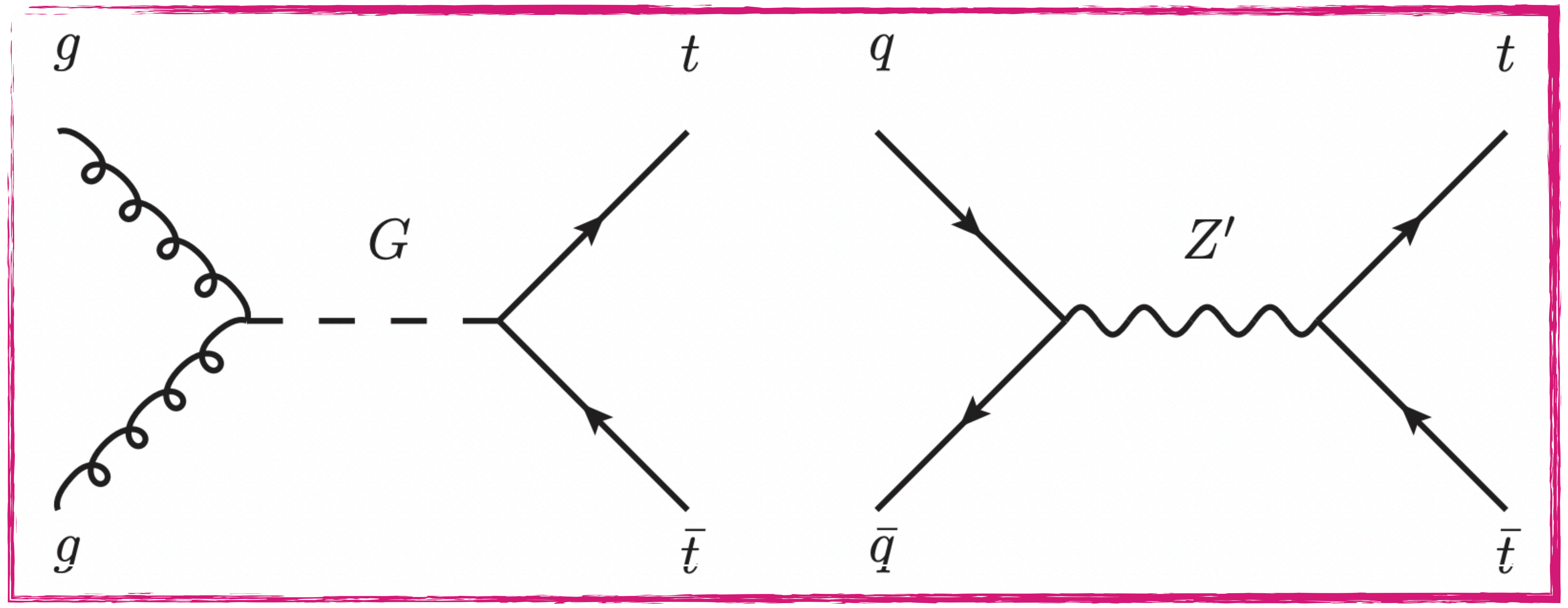
— Extra-Dimensions, RSS Models

Motivates searches for new massive particles

— VLQs consequence of composite Higgs

— SM extensions -> new EW-like bosons

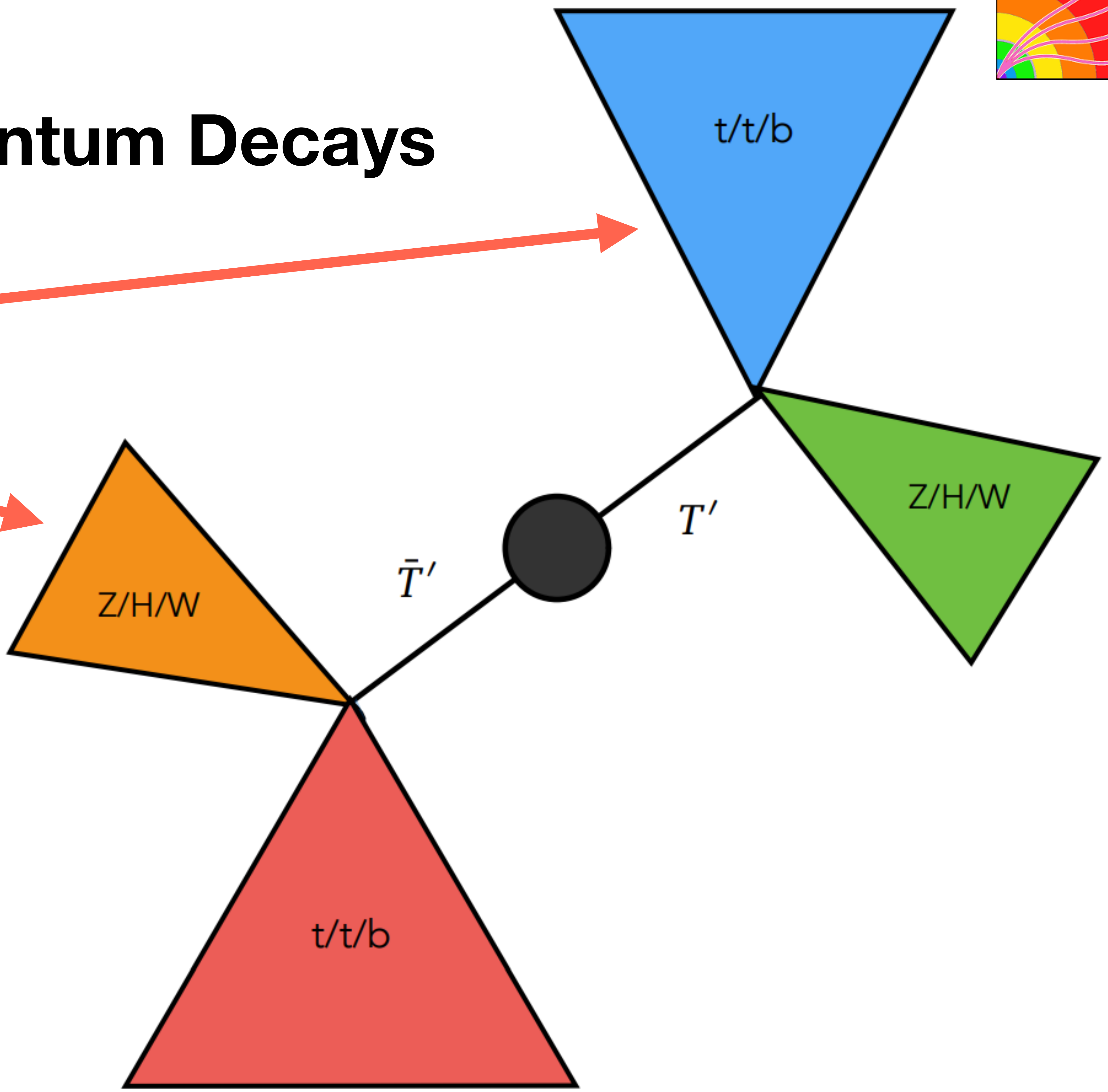
— Massive particles ~TeV scale



Boosted Final States

Heavy New Physics -> High Momentum Decays

‘Fat Jets’ contain boosted daughters



Final state topology ideal for

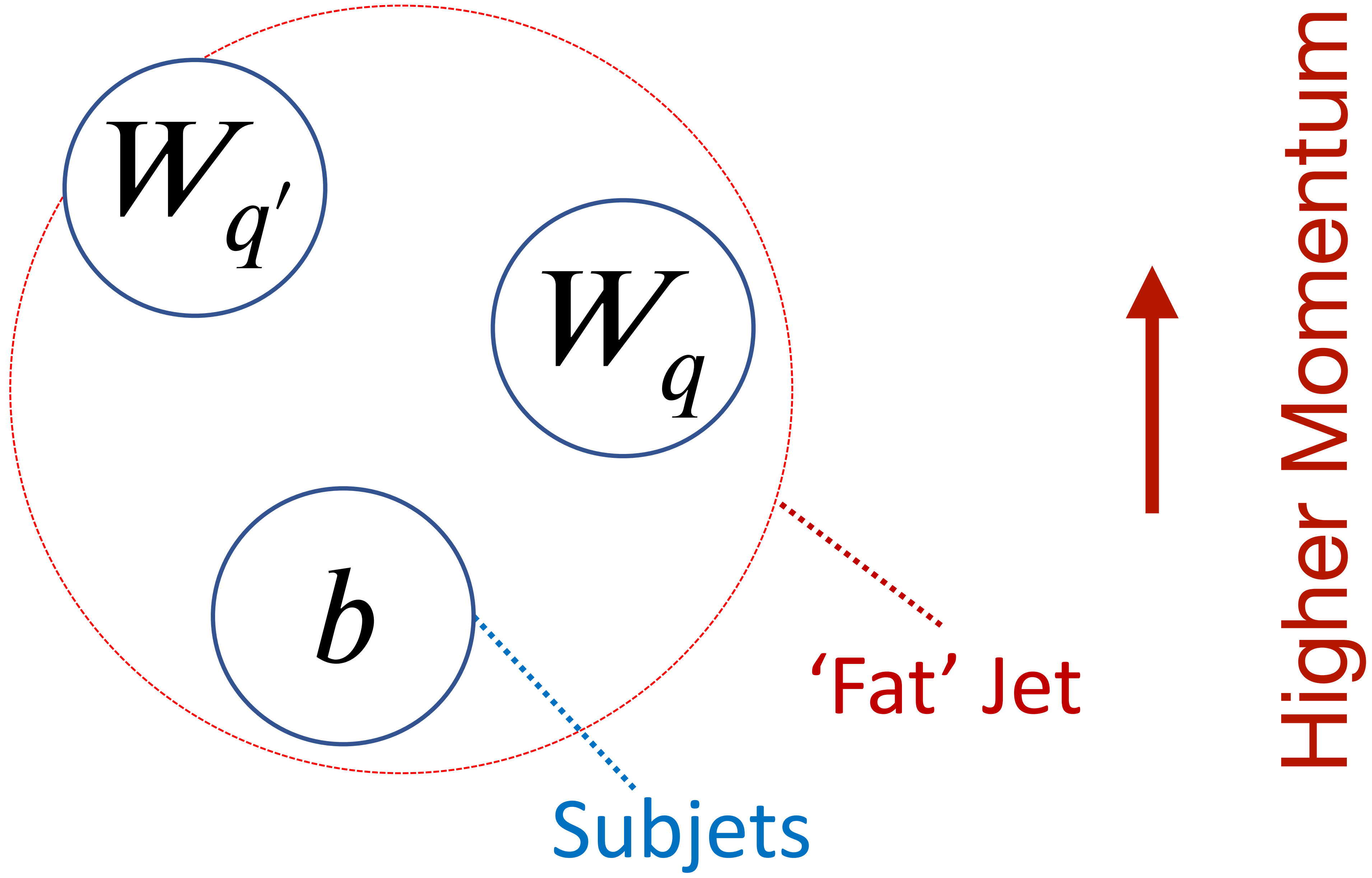
Multi-Classifier Tagger

- Types of daughter particles: W, Z, H, b, t
- Pair-produced VLQs -> 4 high pT jets, many combinations of daughters

[1]: PhysRevD 100, 072001

Jet Substructure for Object Tagging

Evolving Our Methods

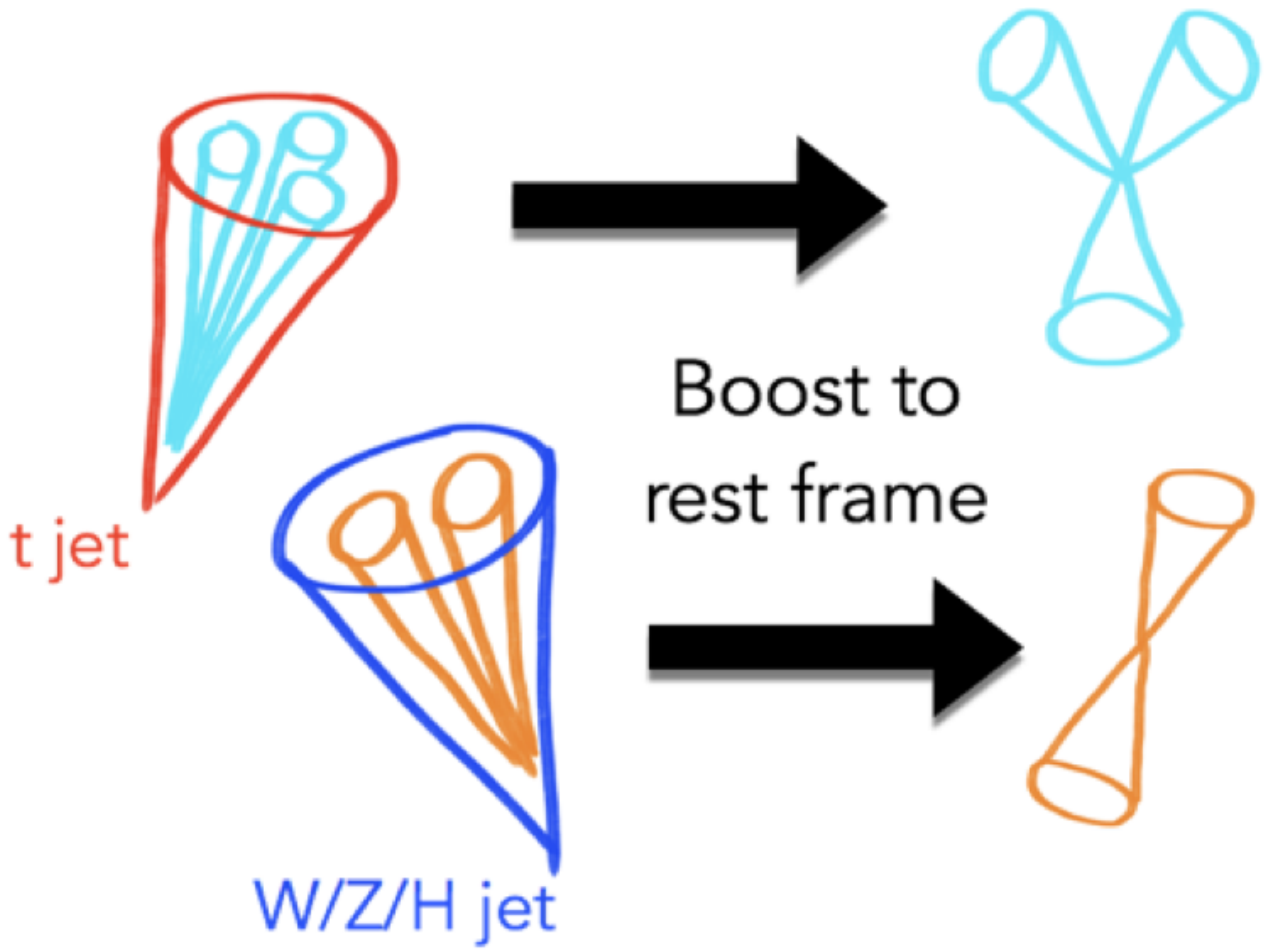


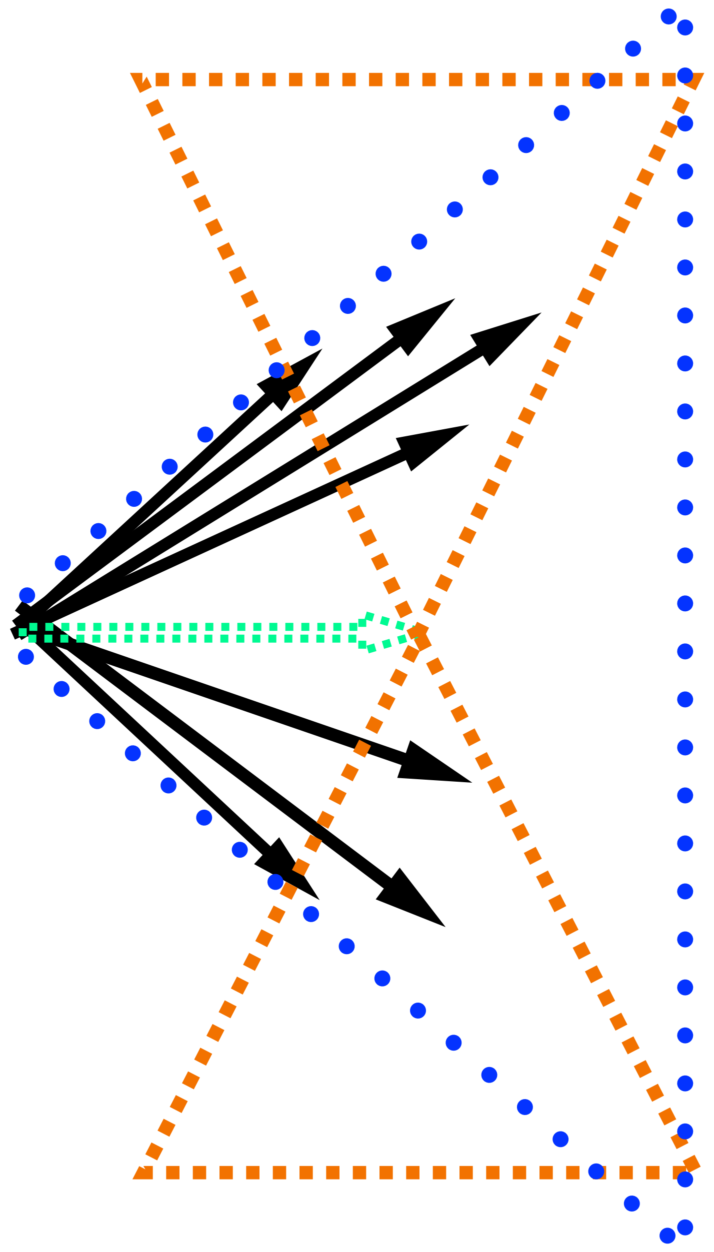
- Traditional observables
  - N-Subjettiness (jet-moments)
  - Jet mass
  - Secondary Vertex info (b-tagging)

- Challenging at very high momentum
  - Opening angle: $\sim \dfrac{2 * mass}{momentum}$
  - Massive resonances or parent particles -> highly boosted decays

- Need to disambiguate close-by objects, make use of new optimizers (ML)

Boosted Event Shape Tagger





Key Idea: Boost jet constituents into hypothetical rest frames

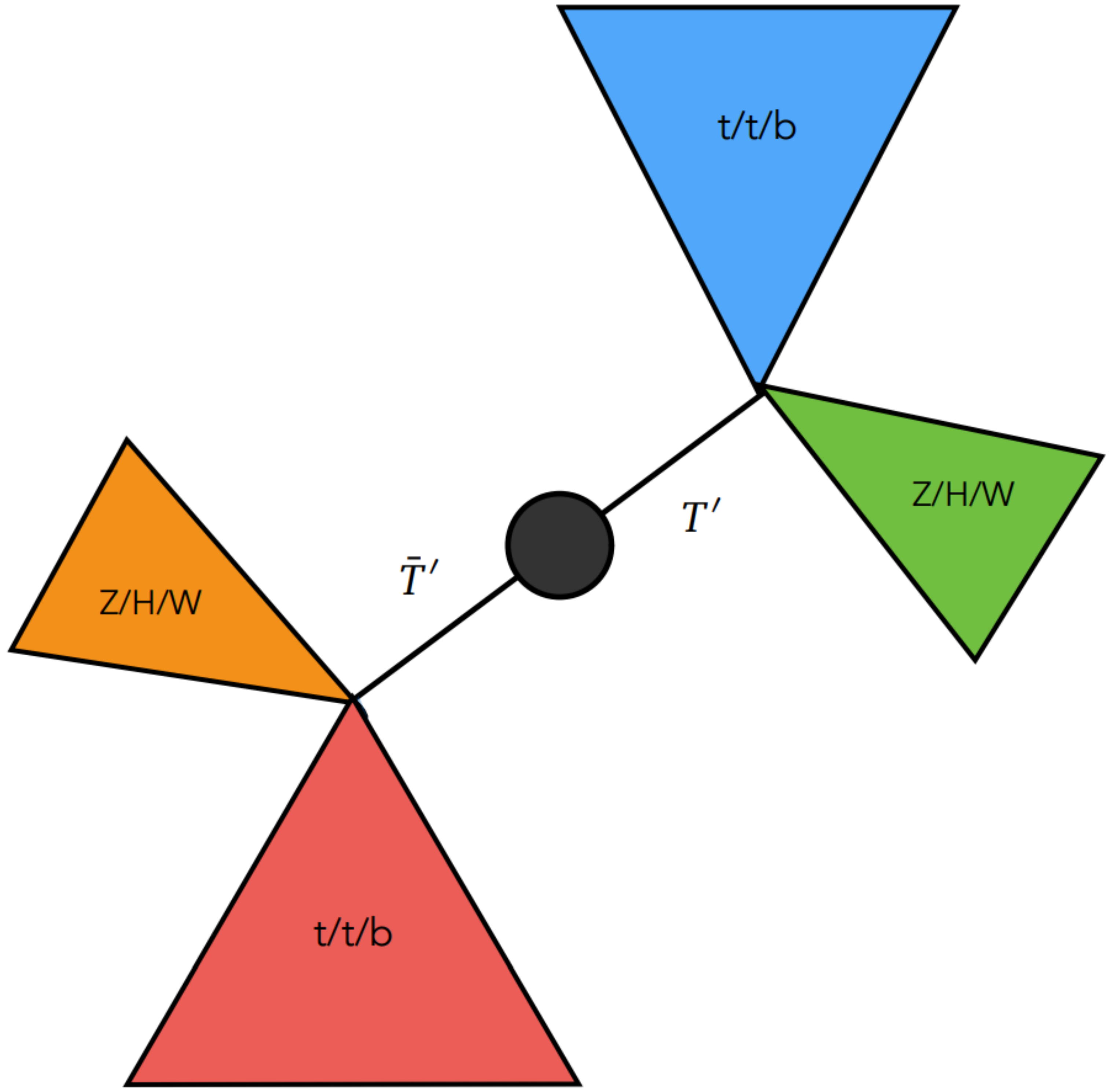
— Lab frame: jet constituents merged into fat-jet cone of R=0.8

— Boost into correct frame: jet constituents become isotropic

— Calculate <u>**Boosted Event Shape (BES) vars**</u> in each hypothetical frame:

- — Fox-Wolfram moments
- — Sphericity tensor
- — Re-clustered jet -> invariant masses
- — Traditional jet substucture
- — And more!

Search for Vector-Like Quarks using the Boosted Event Shape Tagger in the All-Hadronic Channel



All-Hadronic Channel

- At-least 4 AK8 Jets
- $p_T > 400$ GeV, $|\eta| < 2.4$
- $m_{SD} > 10$ GeV

Background Estimation

- QCD: Data-driven control region
- V+Jets, dibosons, ttbar, ttV, 4t shapes from simulation

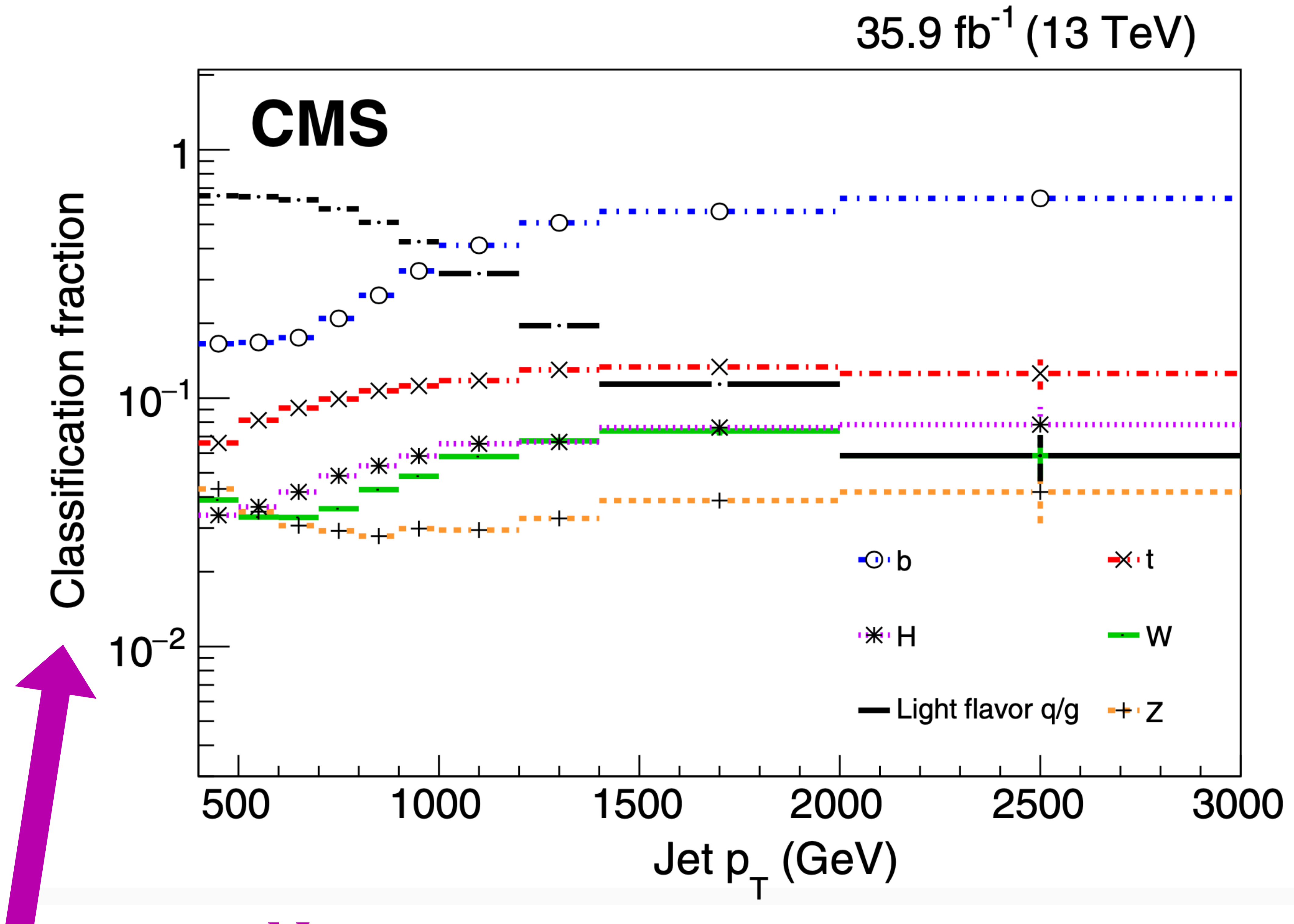
Analysis Strategy

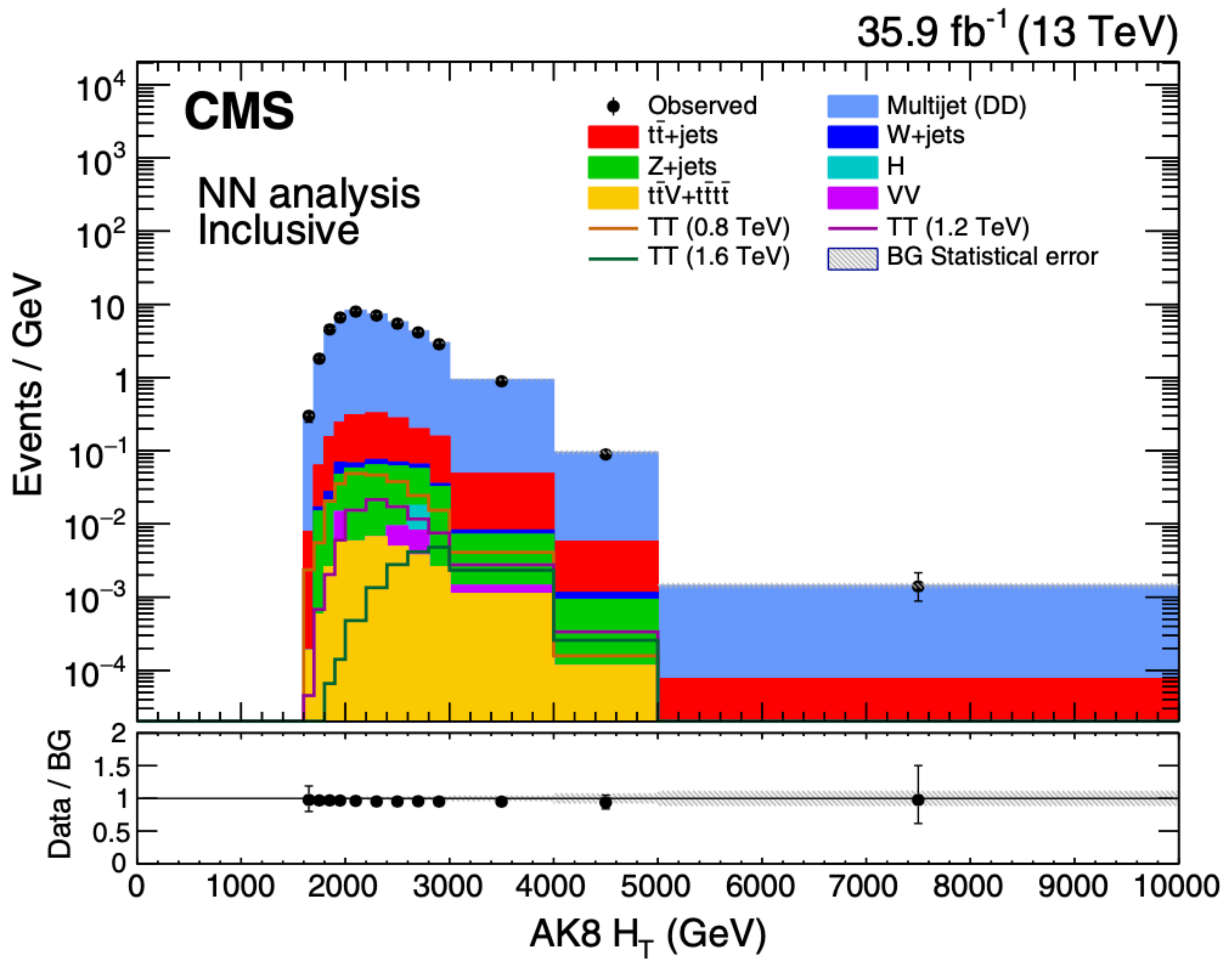
- Tag 4 jets w/ BEST
- Classify event into SR
- Search for excess events at high $H_T = \sum |p_T|$

Background Estimation

Tagging Rates in Data-MC for QCD

Subdominant BGs well-modeled, shapes from simulation





$$\epsilon_X(p_T) = \frac{N_X}{N_{tot}}$$ Measure tagging rate in QCD-rich region to obtain weights for Signal Region

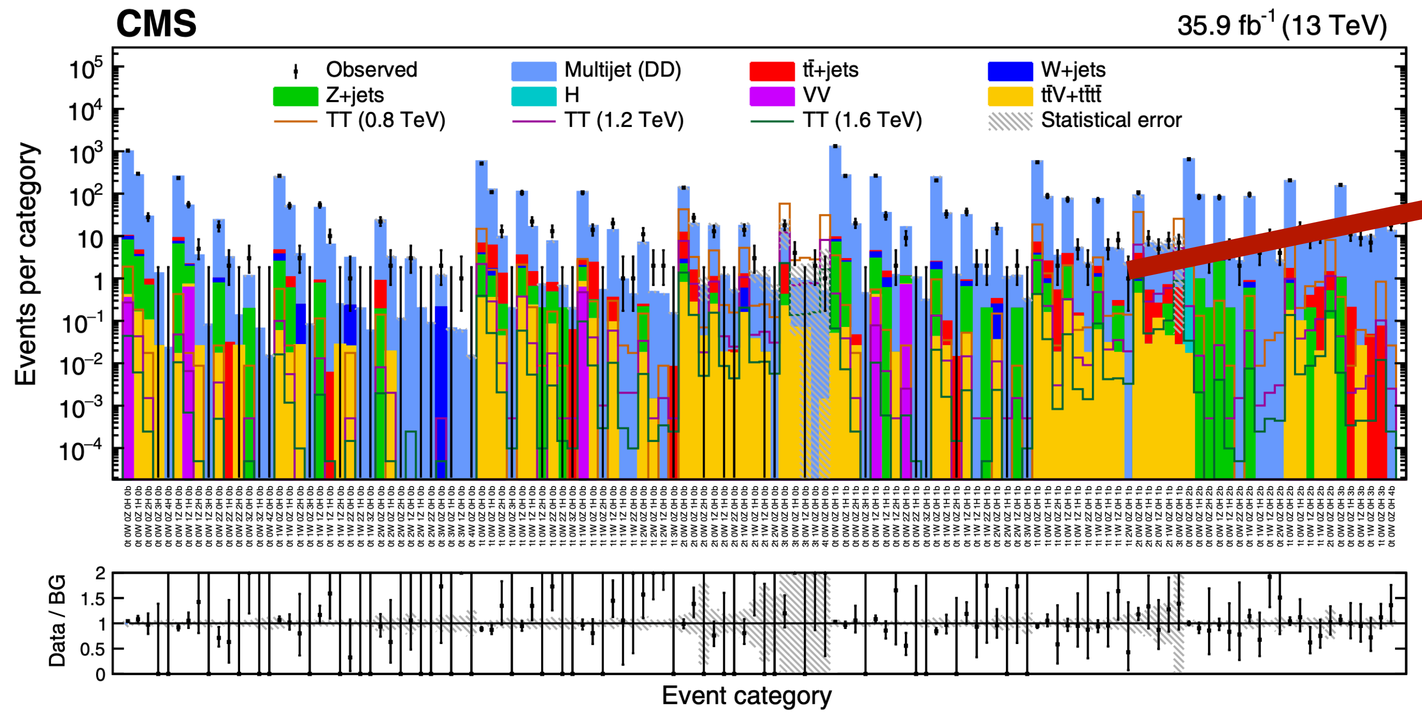
[1]: PhysRevD 100, 072001

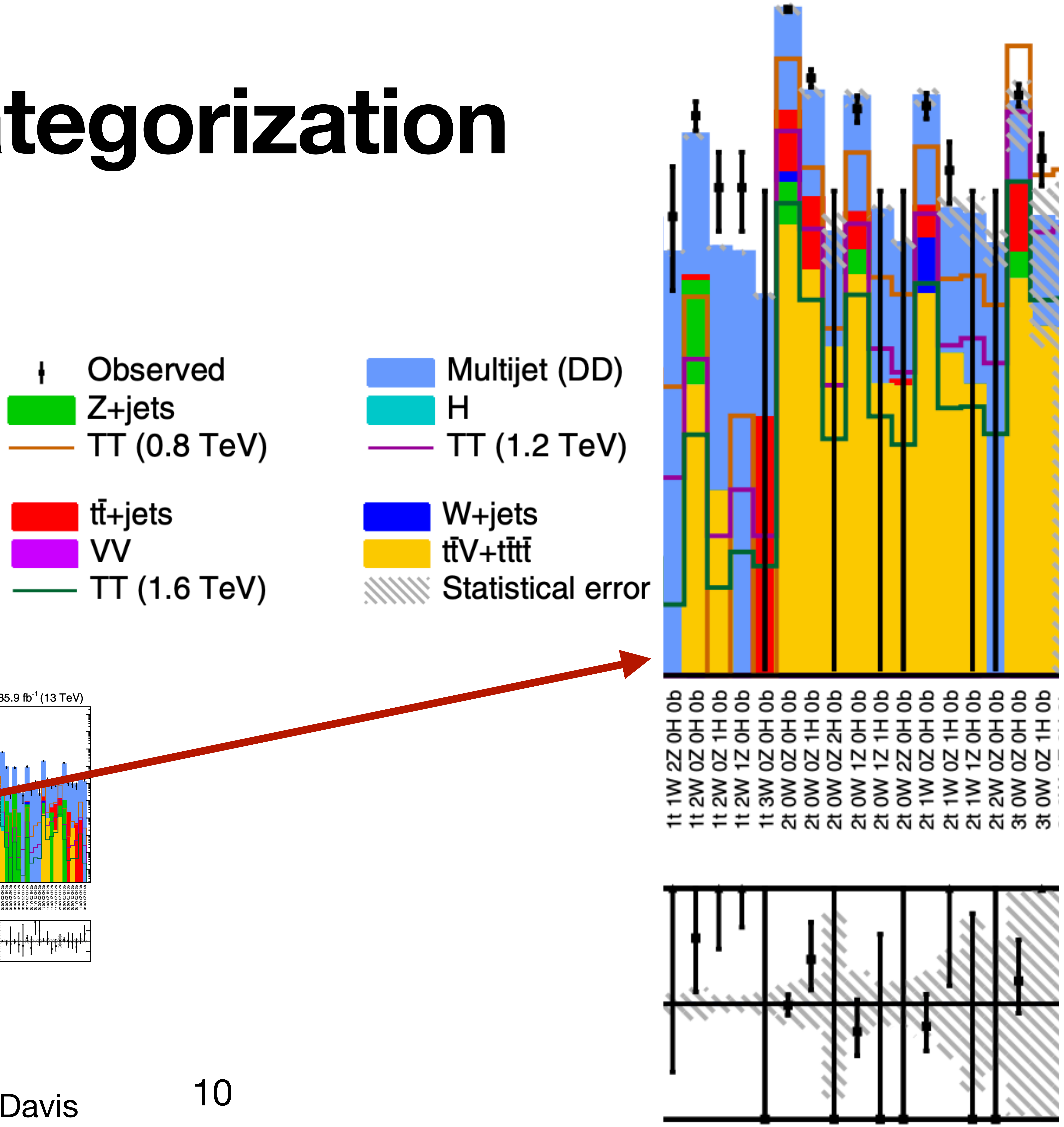
Signal Region Categorization

Tag -> Sort -> Count

- Select >=4 jets and tag
  - 6 classes each
  - 126 possible combinations

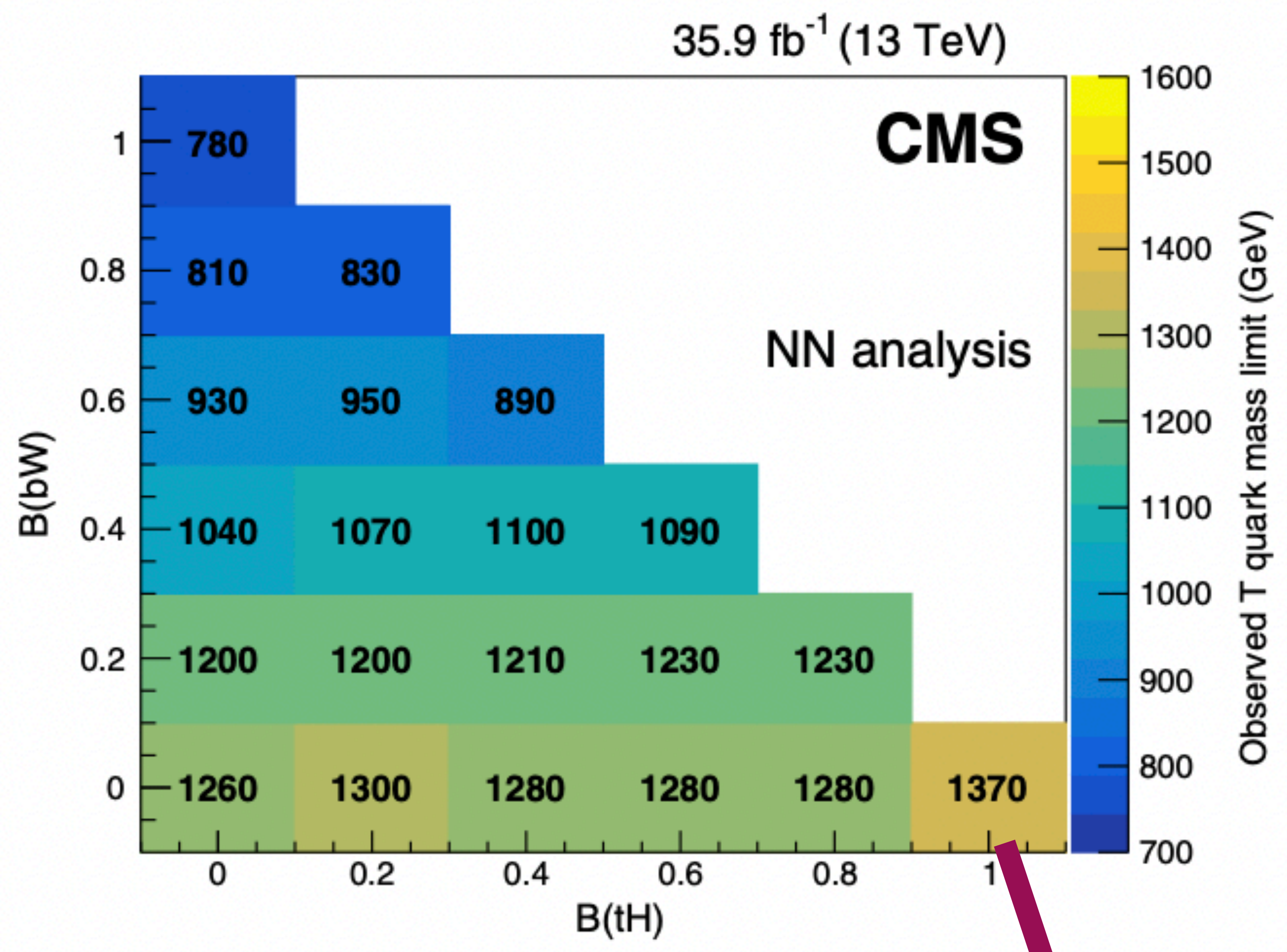
Orthogonal SRs provide powerful combined result

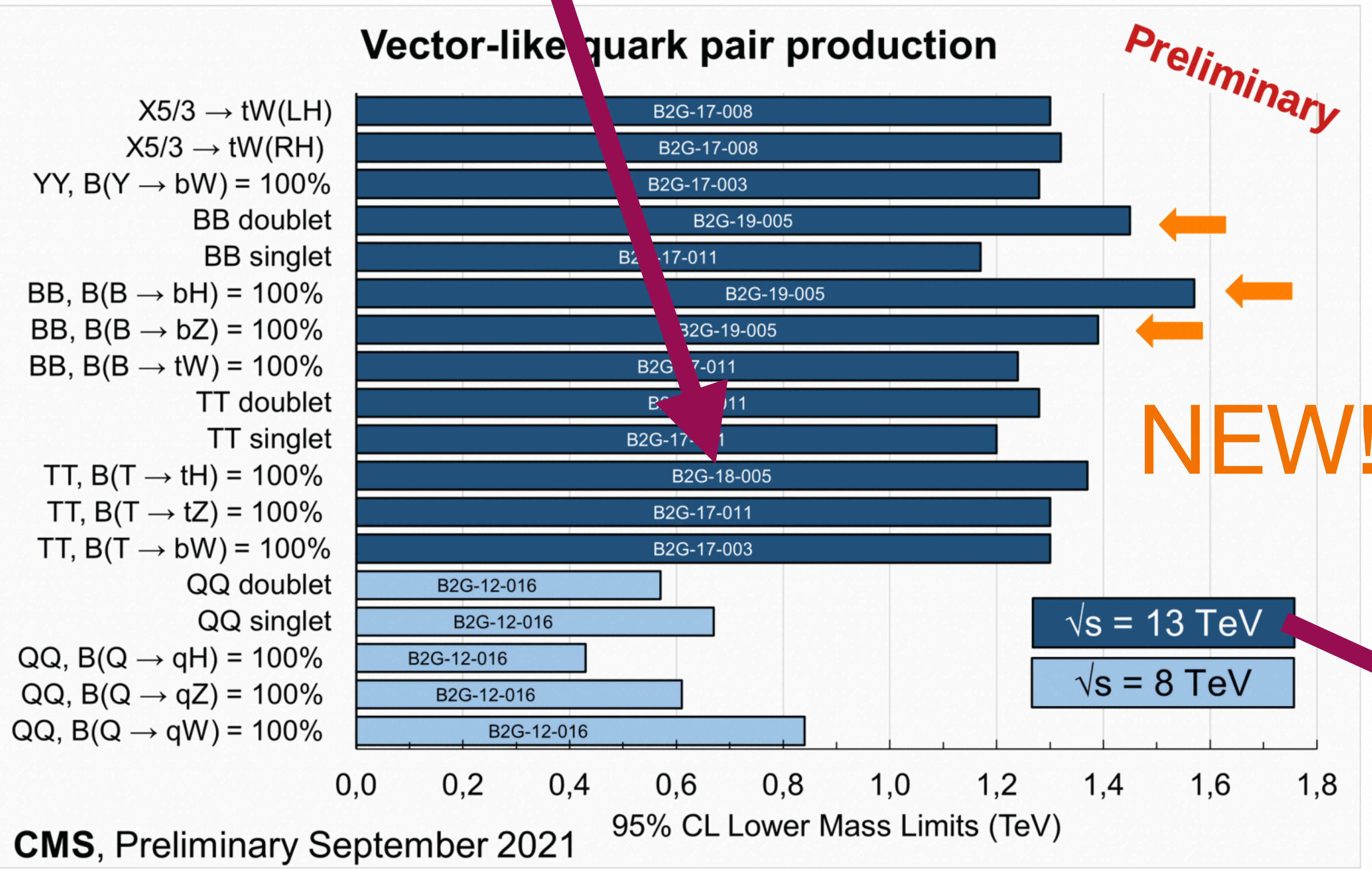




[1]: [PhysRevD 100, 072001]

Public Results of B2G Physics Group

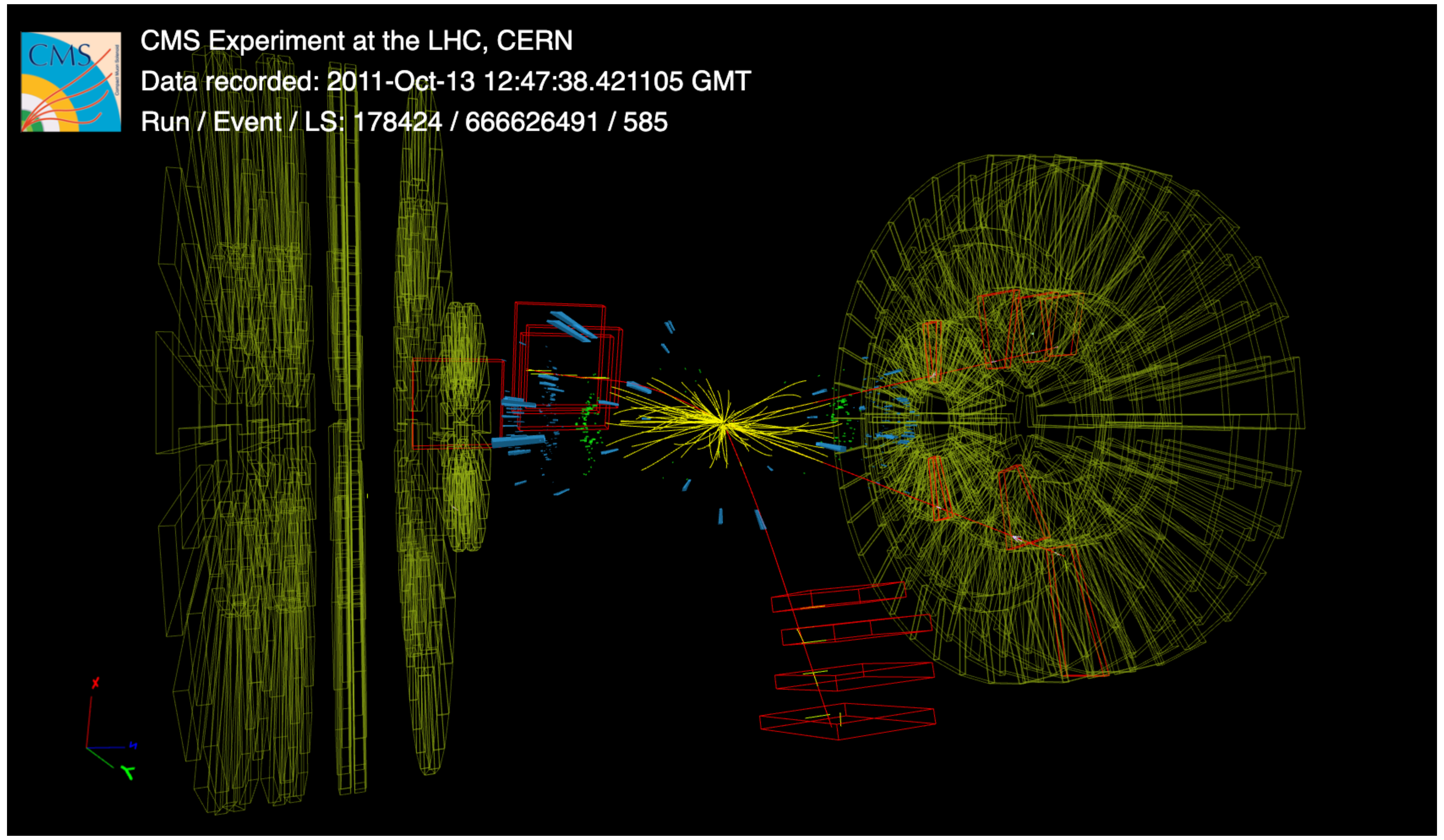




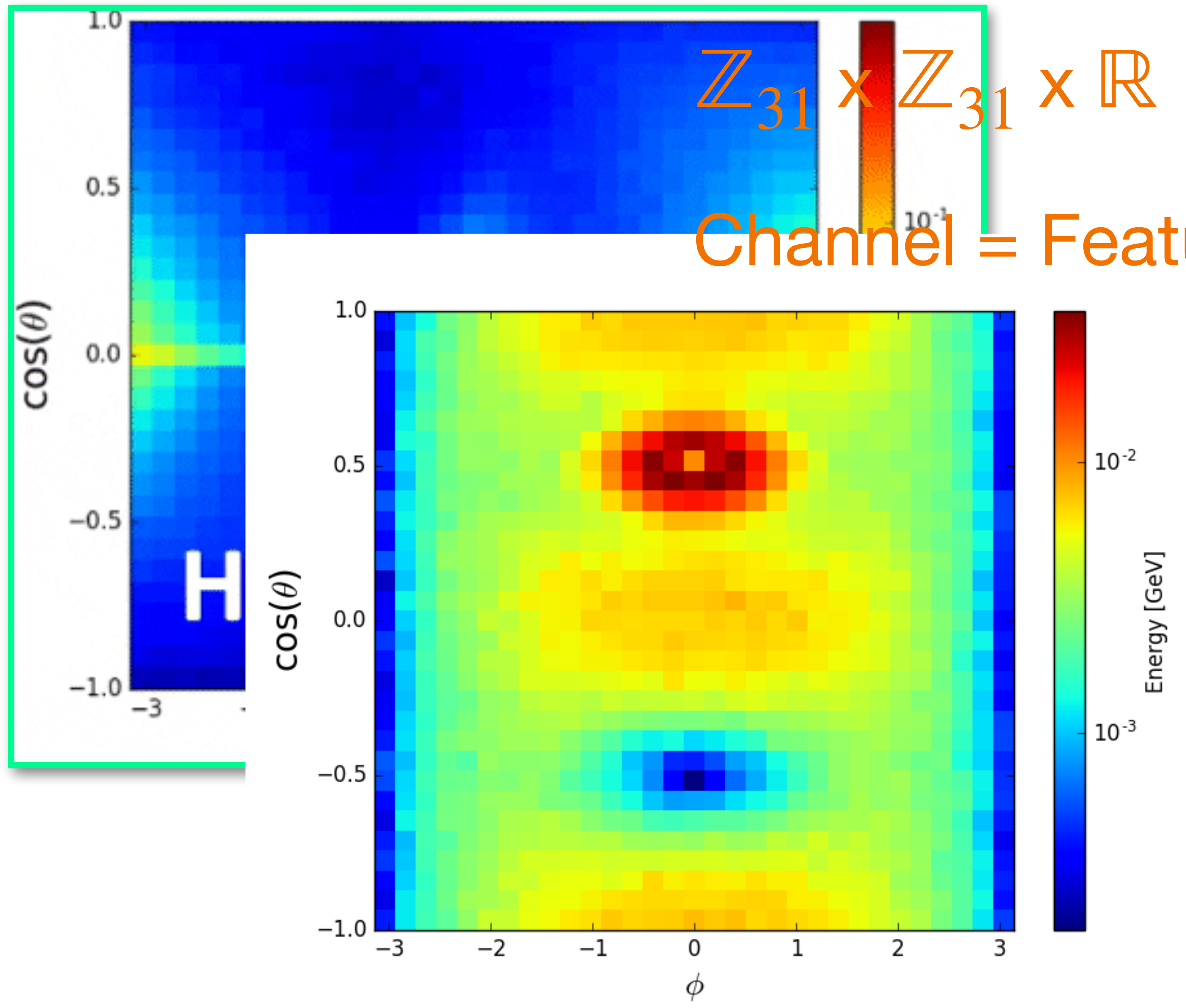
NEW!

Data Representation

Extending BEST with Computer Vision



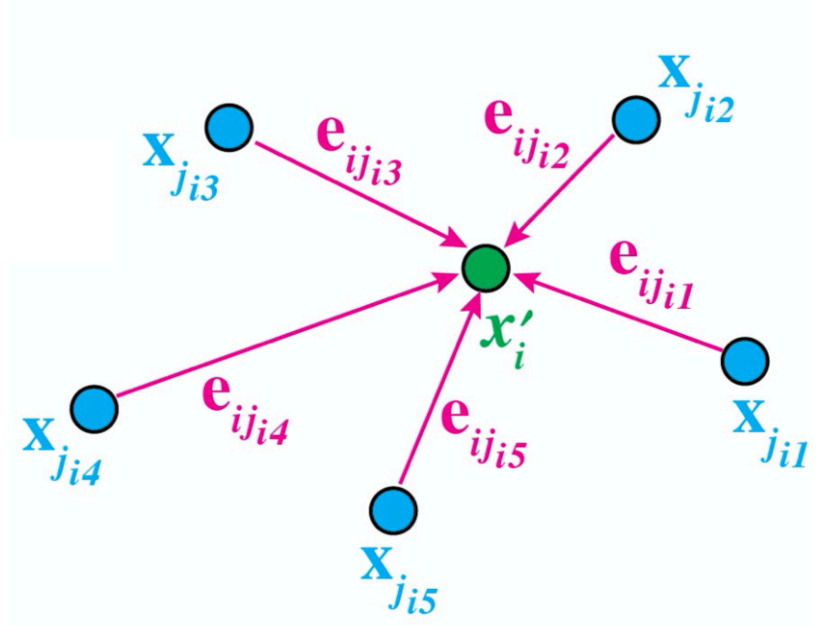
Jet Images



$\mathbb{Z}_{31} \times \mathbb{Z}_{31} \times \mathbb{R}$

Channel = Feature

Point Cloud



$\mathbb{R}^N$ for N features

arxiv:1801.07829

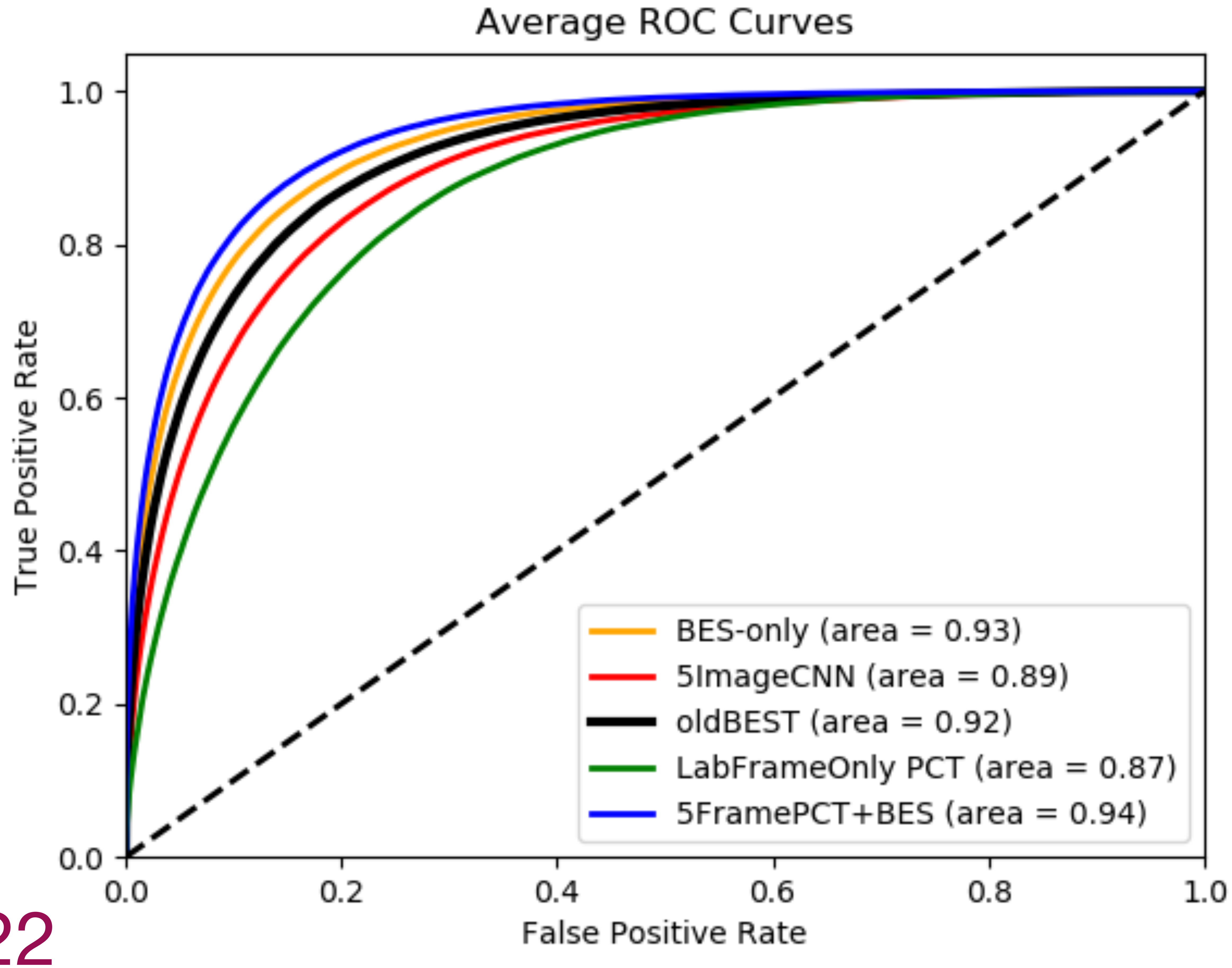
Improvements to BEST in Full Run 2 Analyses

BES (DNN) powerful,

— physics-driven high-level variables saturating learning with CNNs

Data representation crucial, generalization is key!

Full Run 2 BEST: Multi-boost graph network with BESvars vector input



Publications expected late '21, early '22

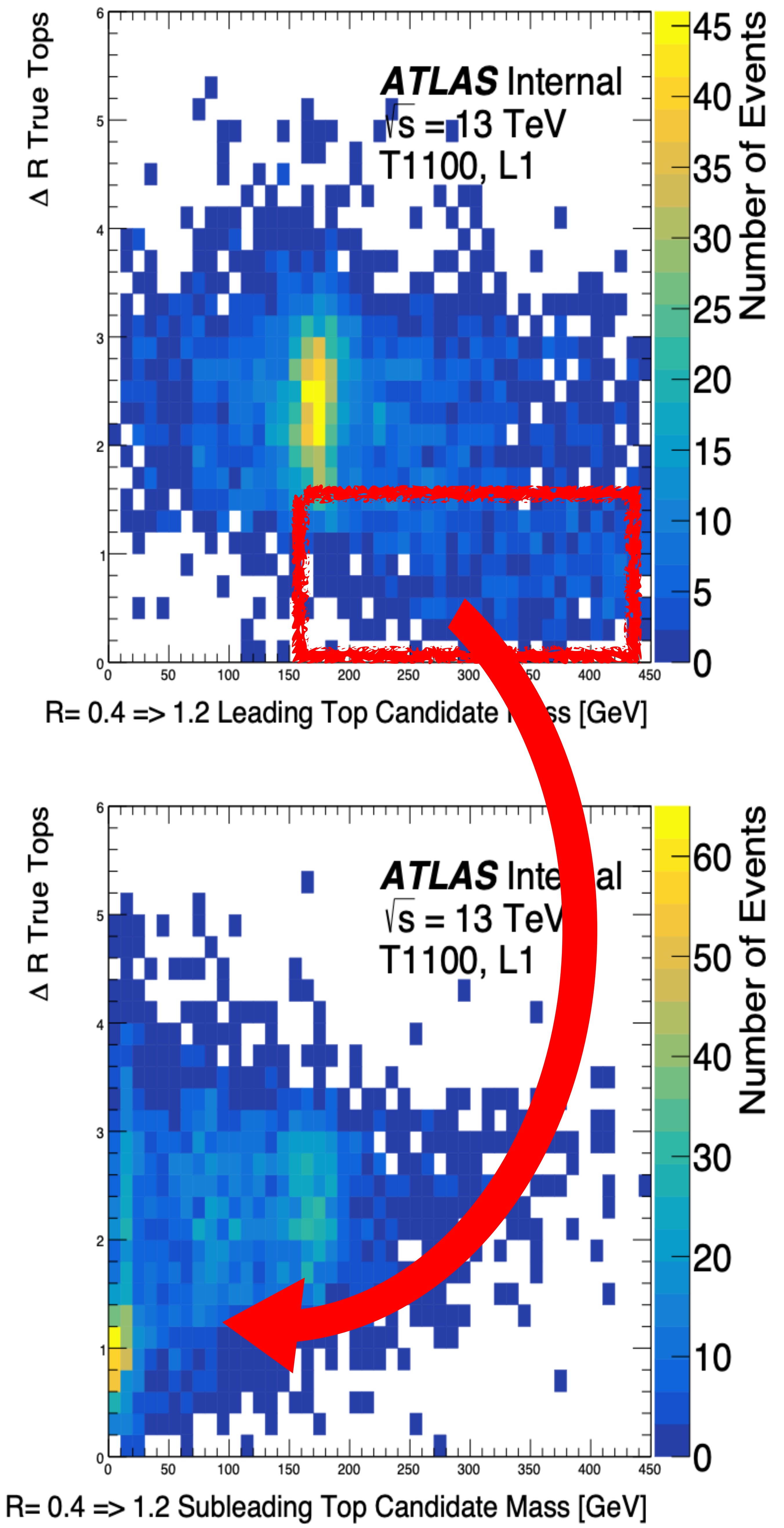
— Combination analyses in formation

— Possibly first LHC results for VLQs @ NLO

Future Tagger Ideas

Plans and Aspirations

- Overlapping Objects
  - Motivation from SUSY top-partner search (PhD work)
  - Tops recoil against neutralinos, blind spot
  - Similar effects in other BSM searches (e.g. VLQs)
- Learn the boost!
  - Current BEST boosts optimized by hand
  - For Run 3, in-situ boosts or boost spectrum
- W/Z discrimination
  - Difficult to differentiate quark flavors in ATLAS/CMS
  - ParticleFlow helps, but far from enough
- Ensembling vs piece-wise taggers



Applications of BEST in CMS

]

Direct optimization for analyses

— All-Hadronic Pair-Produced VLQ Search
6-class tagger, winner-take-all
— Heavy Resonances -> ttbar -> all-hadronic
top-tagger, 1-D score cut
— Combination analyses of TT/BB
and ttbar resonances

CMS-wide uses

— Developing common framework for fast-analysis
— User development of taggers for specific uses

New interpretations for Run 3

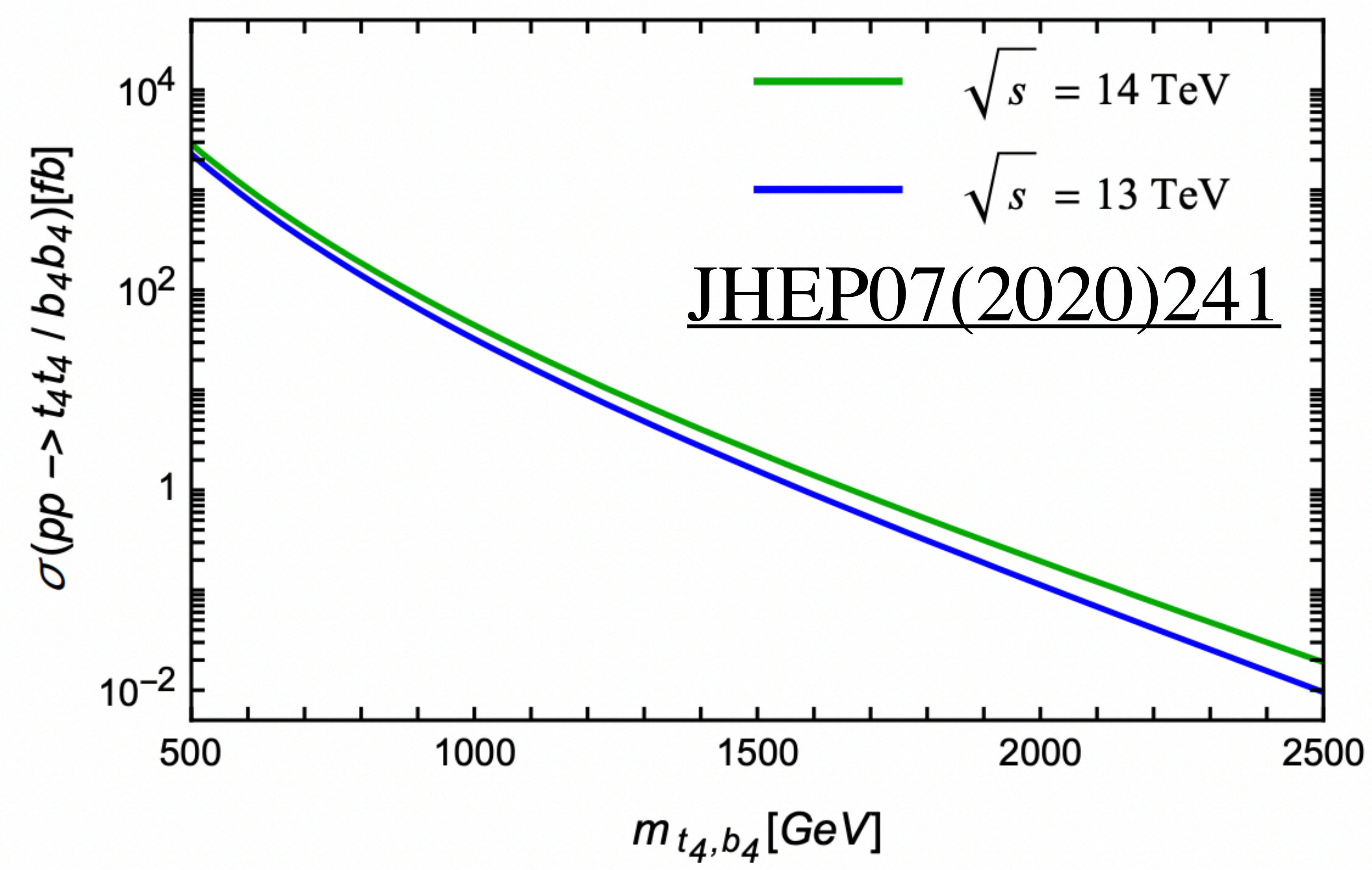
— Wide-width models
— Contact interaction interpretations
— Actively looking for new methods and models

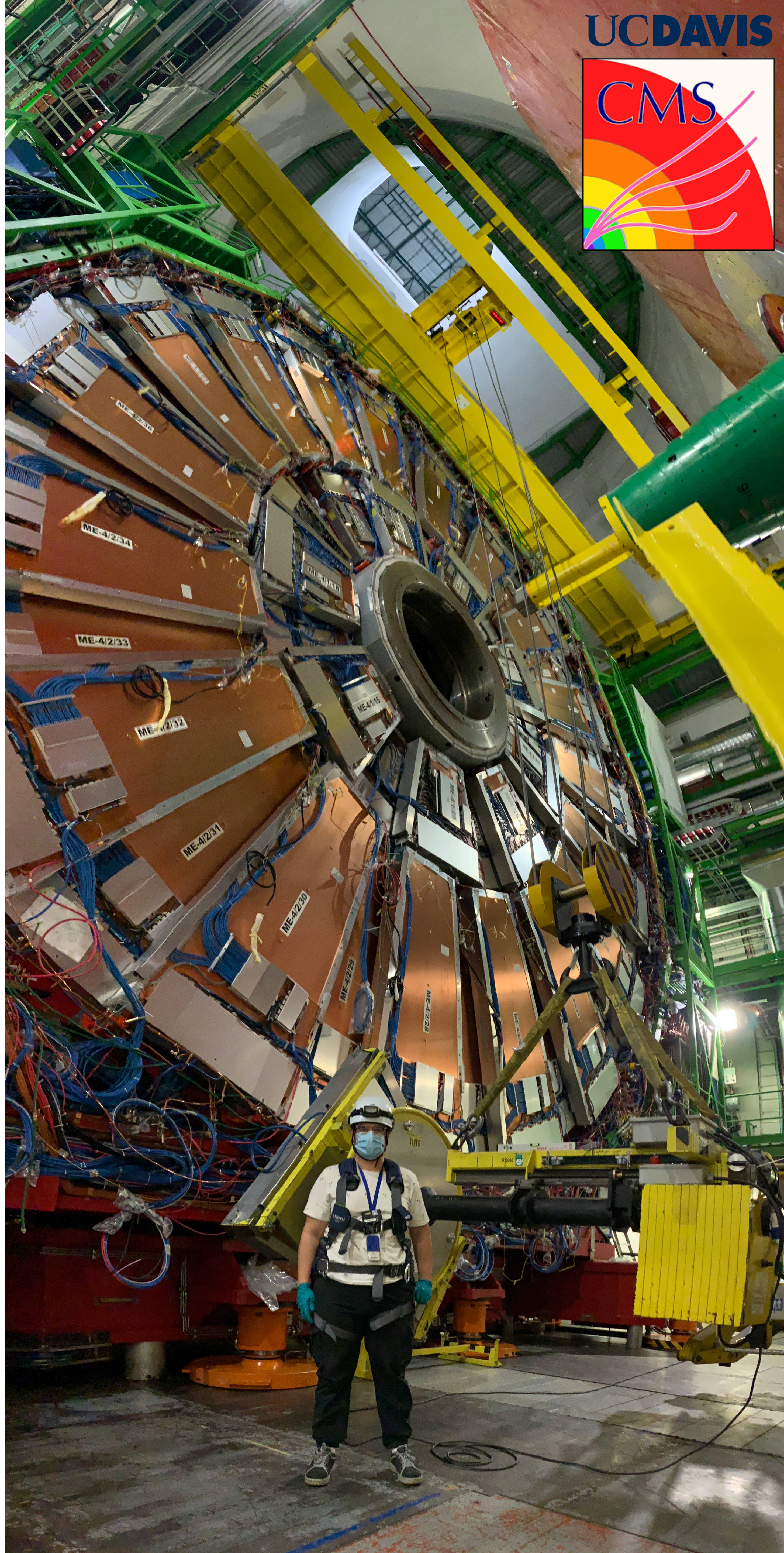
Looking Towards HL-LHC

- HL-LHC will bring 10x data
  - Busy environment!
    30->200 interactions per bunch crossing
  - Higher energy -> Higher boosts

- Higher center of mass energy enhances production cross-section

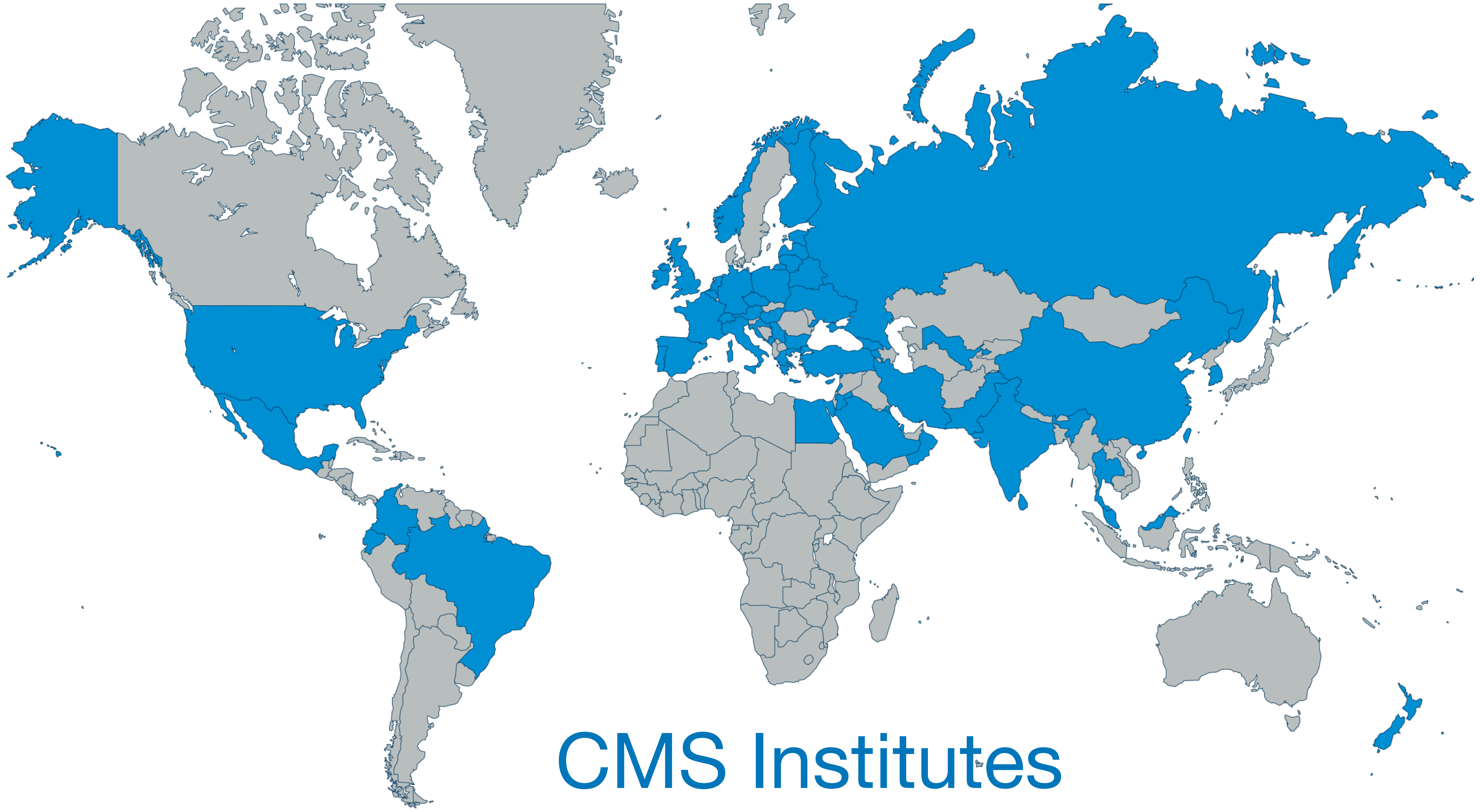
- Investing in better detectors and tools too!





Costa Rica as new CERN Collaborator

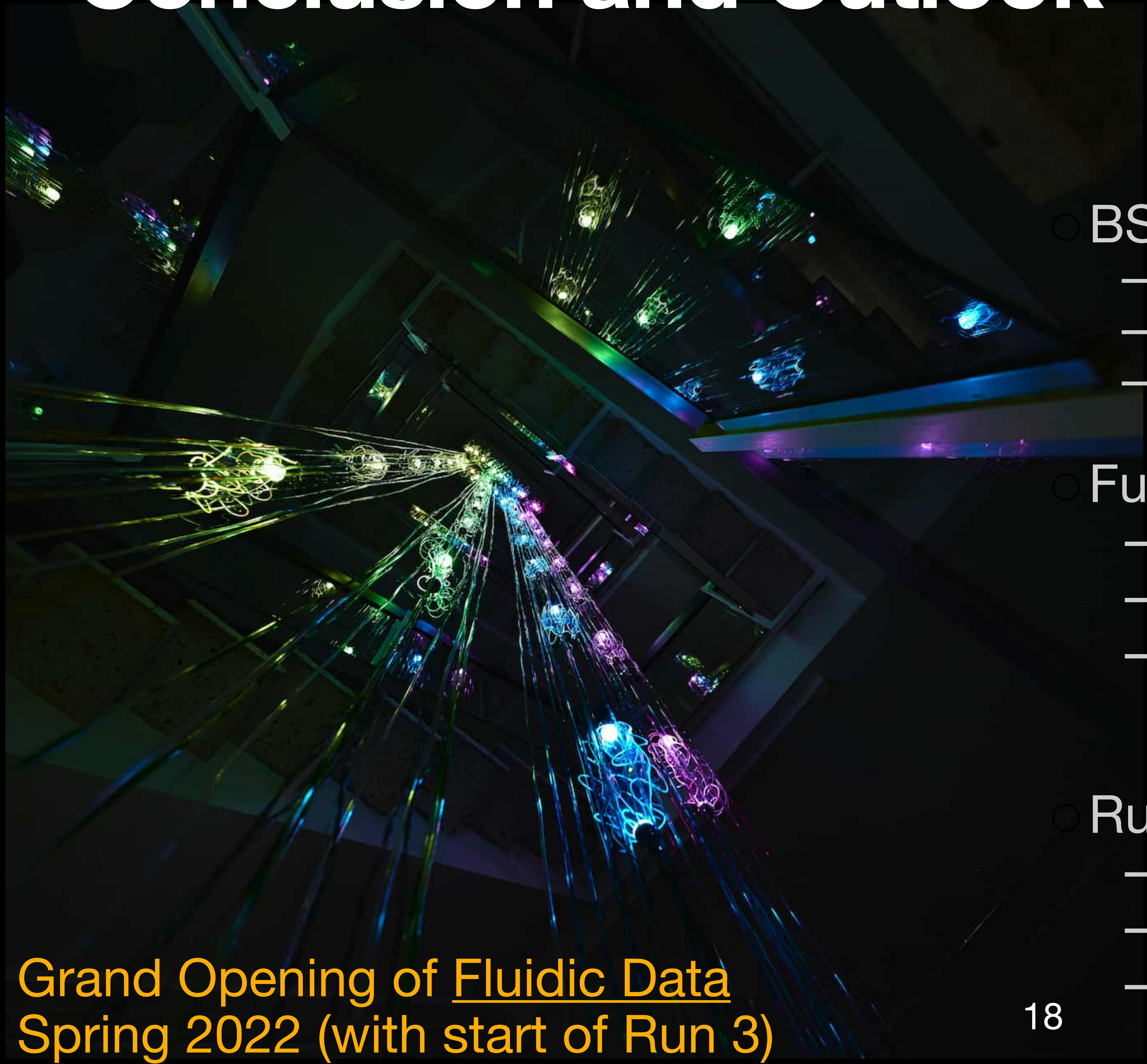
Reducing the Entrance Barrier to HEP for Latin Americans



Conclusion and Outlook

[CMS Stories]

"I learn because I want to know,
I succeed because I have ganas."

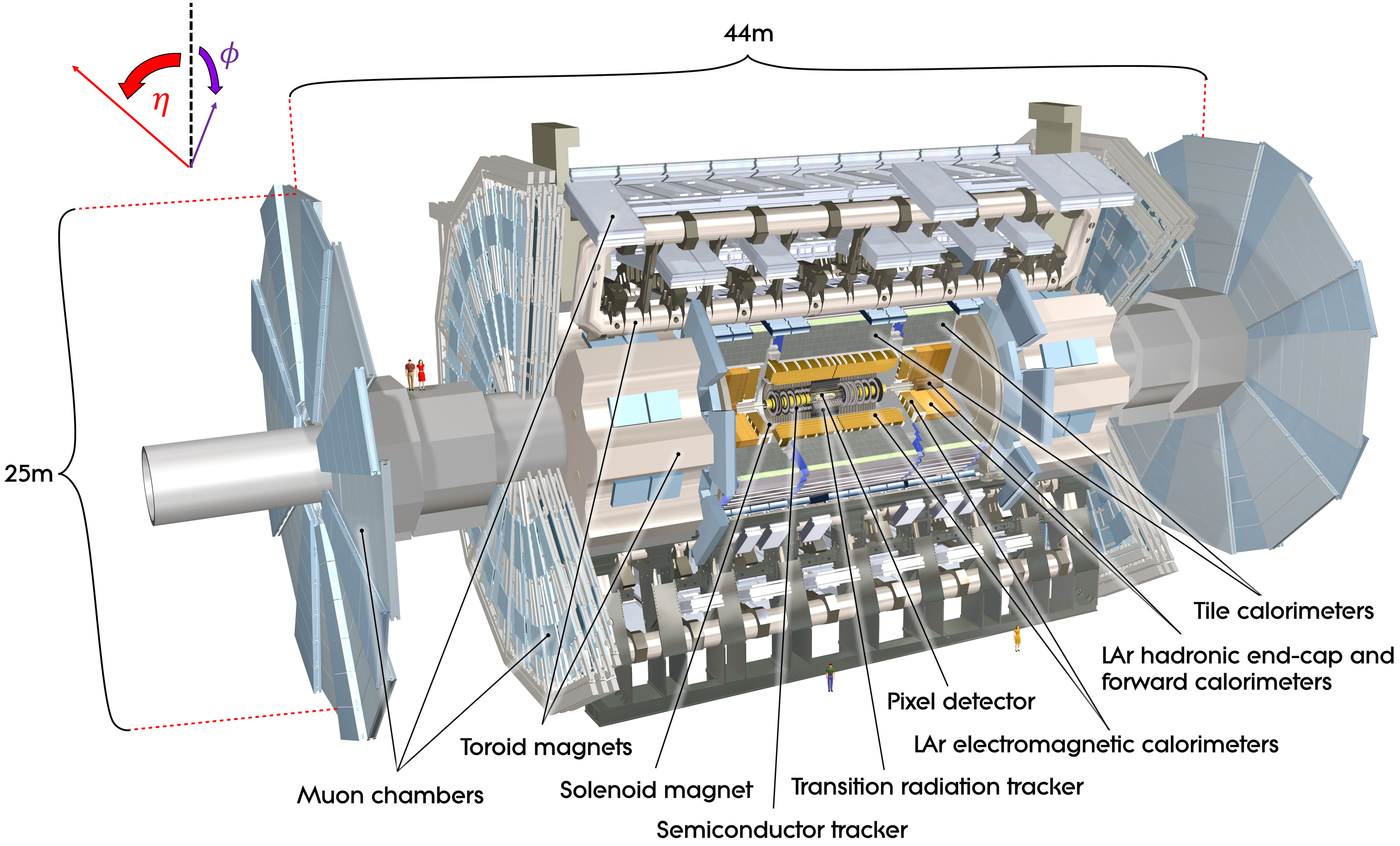


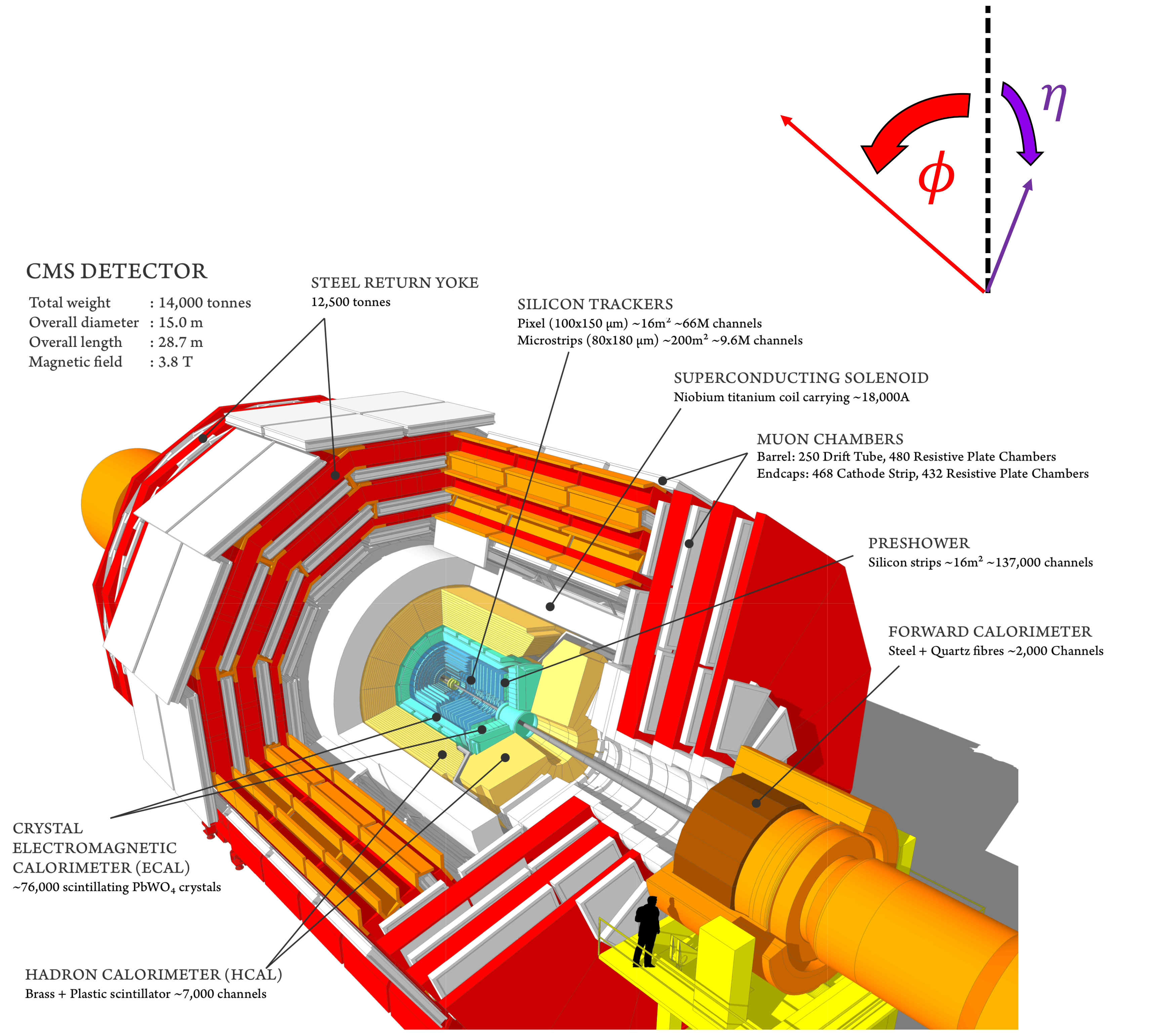
- BSM Physics is out there!
  - Non-SM 125 GeV scaler can be exciting!
  - Tops well-motivated portal to BSM
  - Top partners can solve hierarchy problem
- Future Ideas:
  - Can we *learn* the boost?
  - Discrete->Continuous frames?
  - Piece-wise categorization (QCD-tagger then W/Z/H/top/b)
- Run 3 to start in Spring 2022
  - 13.6 TeV (TBC)
  - Expected ~200 fb^-1
  - Build collaboration tools for HL-LHC!

Grand Opening of Fluidic Data
Spring 2022 (with start of Run 3)

Backup

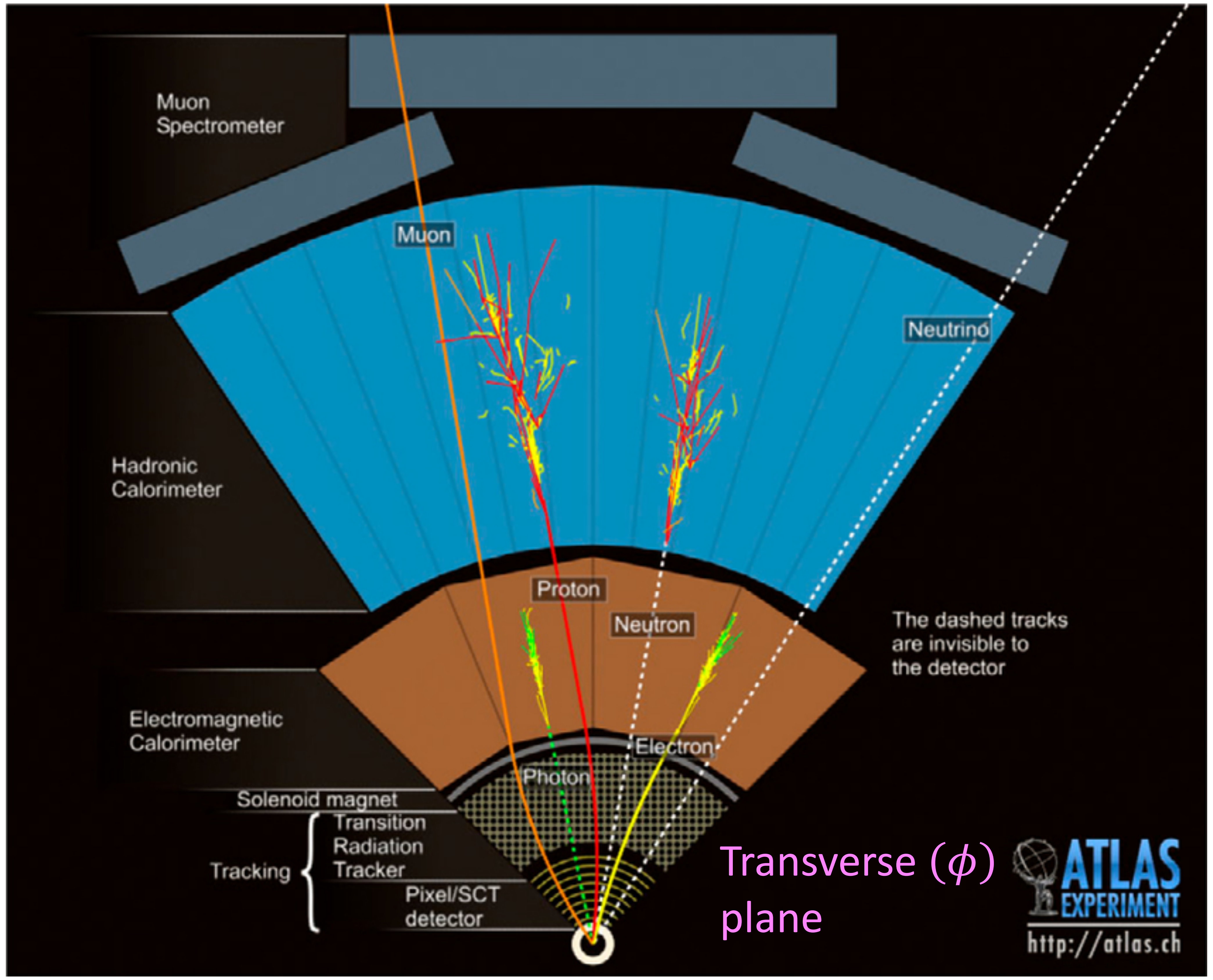
The ATLAS and CMS Detectors

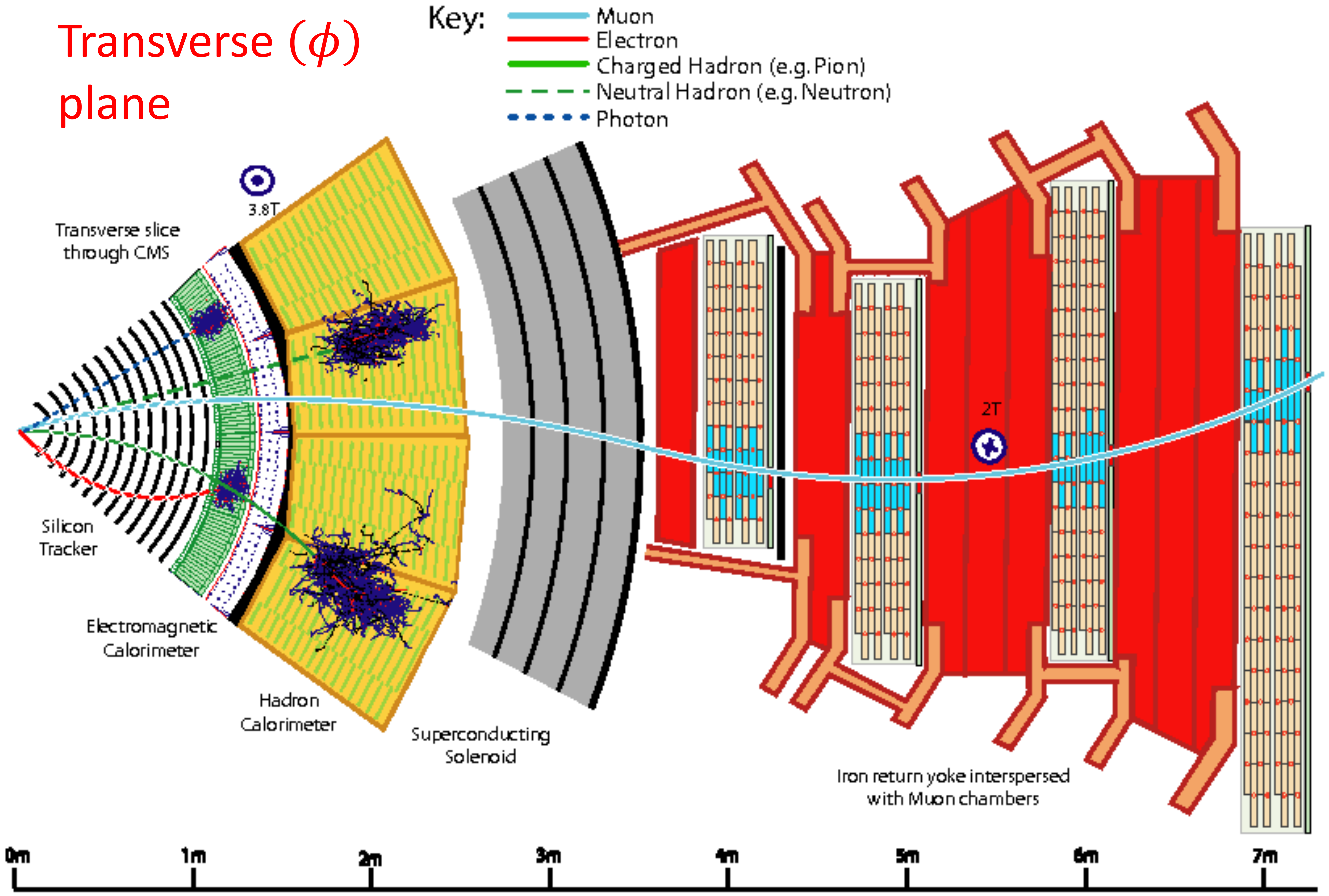




*Approximate Relative Scale

Detecting Particles in ATLAS and CMS

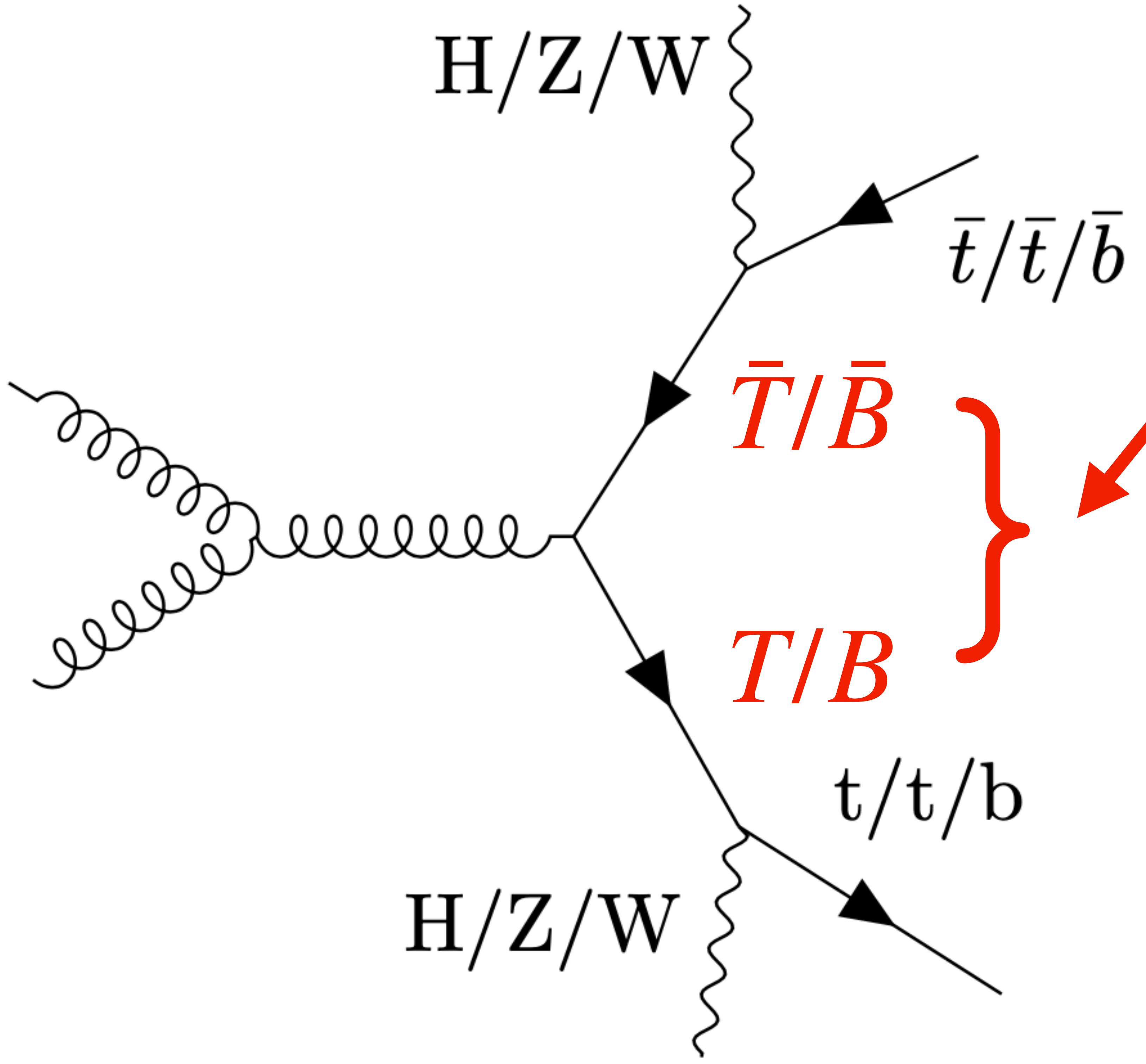




Search for Vector-Like Quarks using the Boosted Event Shape Tagger in the All-Hadronic Channel

Motivations from Higgs

- Why so light at 125 GeV?
- Fundamental or composite particle?
- Precise measurements of Higgs decays rule out new chiral fermions



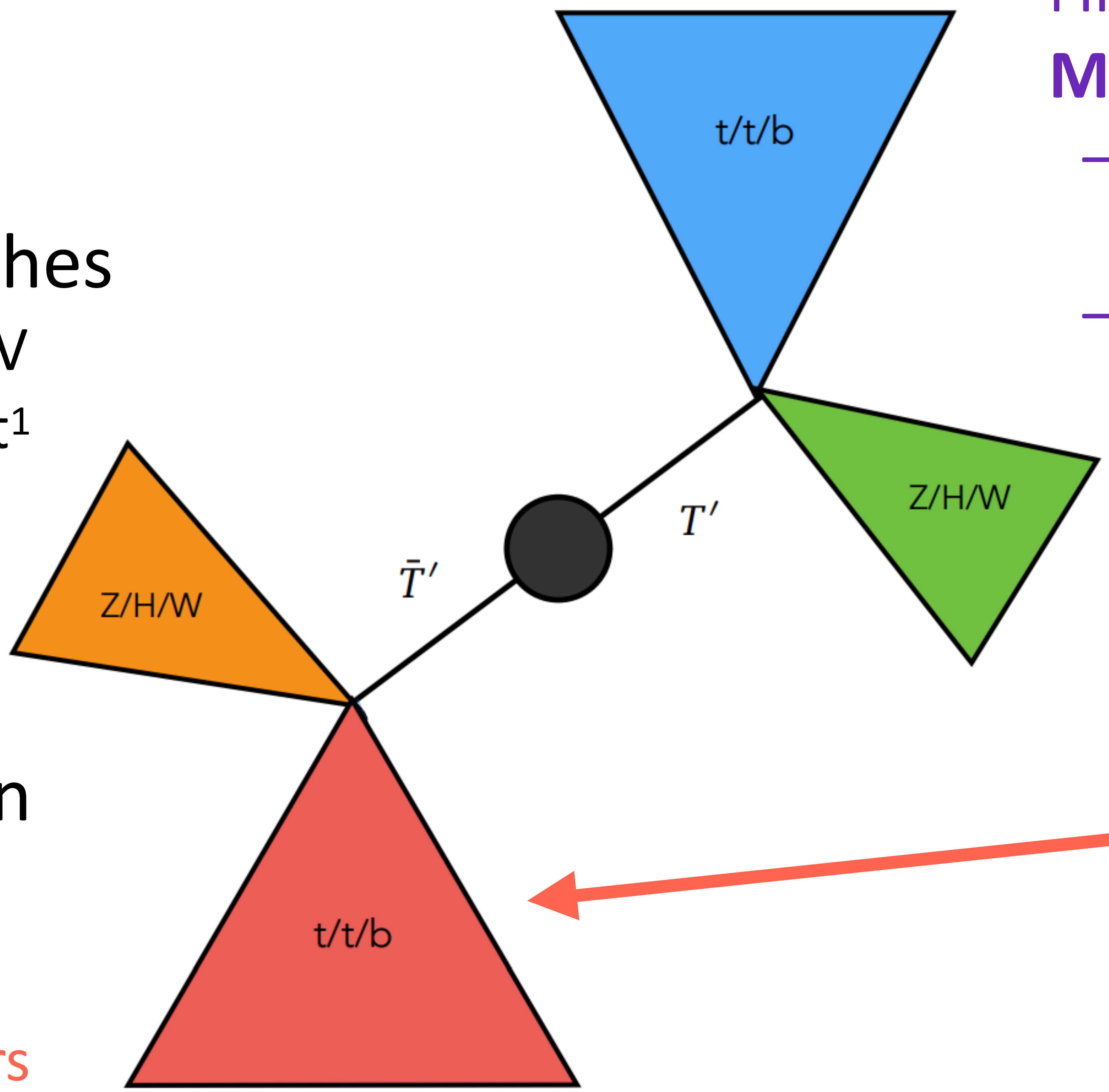
Vector-Like Quarks (VLQ)

- Models with Higgs as pseudo-Nambu-Goldstone
- Flavor Partial Compositeness
- Mass NOT acquired through coupling to Higgs doublet

Analysis Motivation
and Signal Topology

Improving on past VLQ searches
- Exclusions $m_{T/B} \sim 700 - 1200$ GeV
- 3.8x more data than 2016 result[1] $(36 \rightarrow 137\, fb^{-1})$
- Refined Tagger

Decay to SM fermion + boson
- High branching fraction of all final state decays
- Heavy VLQ -> Boosted Daughters



Final state topology ideal for **Multi-Classifier Tagger**
- Types of daughter particles: W, Z, H, b, t
- At least 4 high pT jets, many combinations of daughters

'Fat Jets' contain boosted daughters

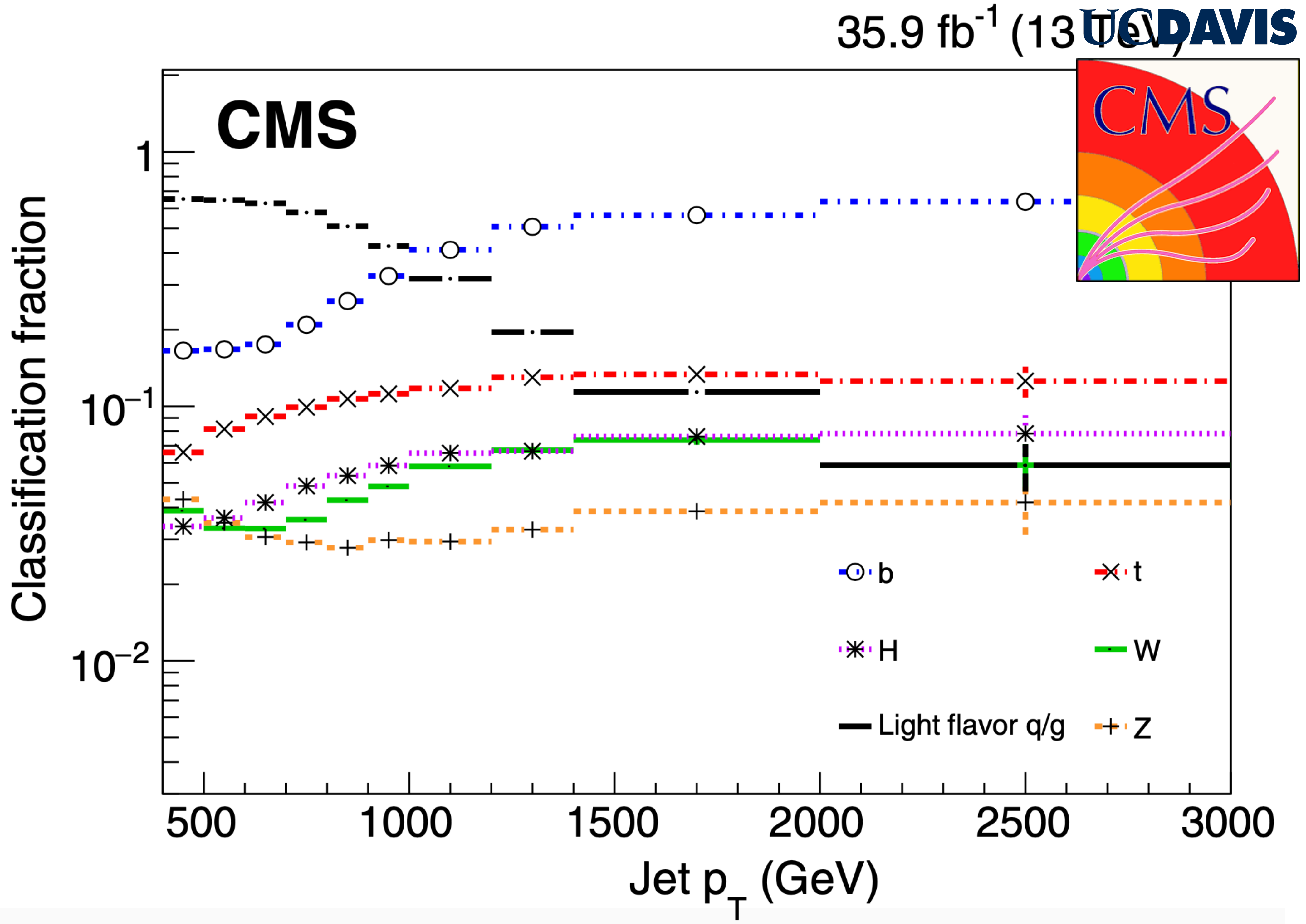
[1]: PhysRevD 100, 072001

Background Estimation
and Tagging Rates in Data-MC

Estimation Strategy for QCD using BEST

— Measure mistag rate w/ data in multi-jet enriched CR, $\epsilon_X(p_T) = \frac{N_X}{N_{tot}}$

— Obtain shape and normalization from VR/SR-inclusive (preselection, 4-jet) region



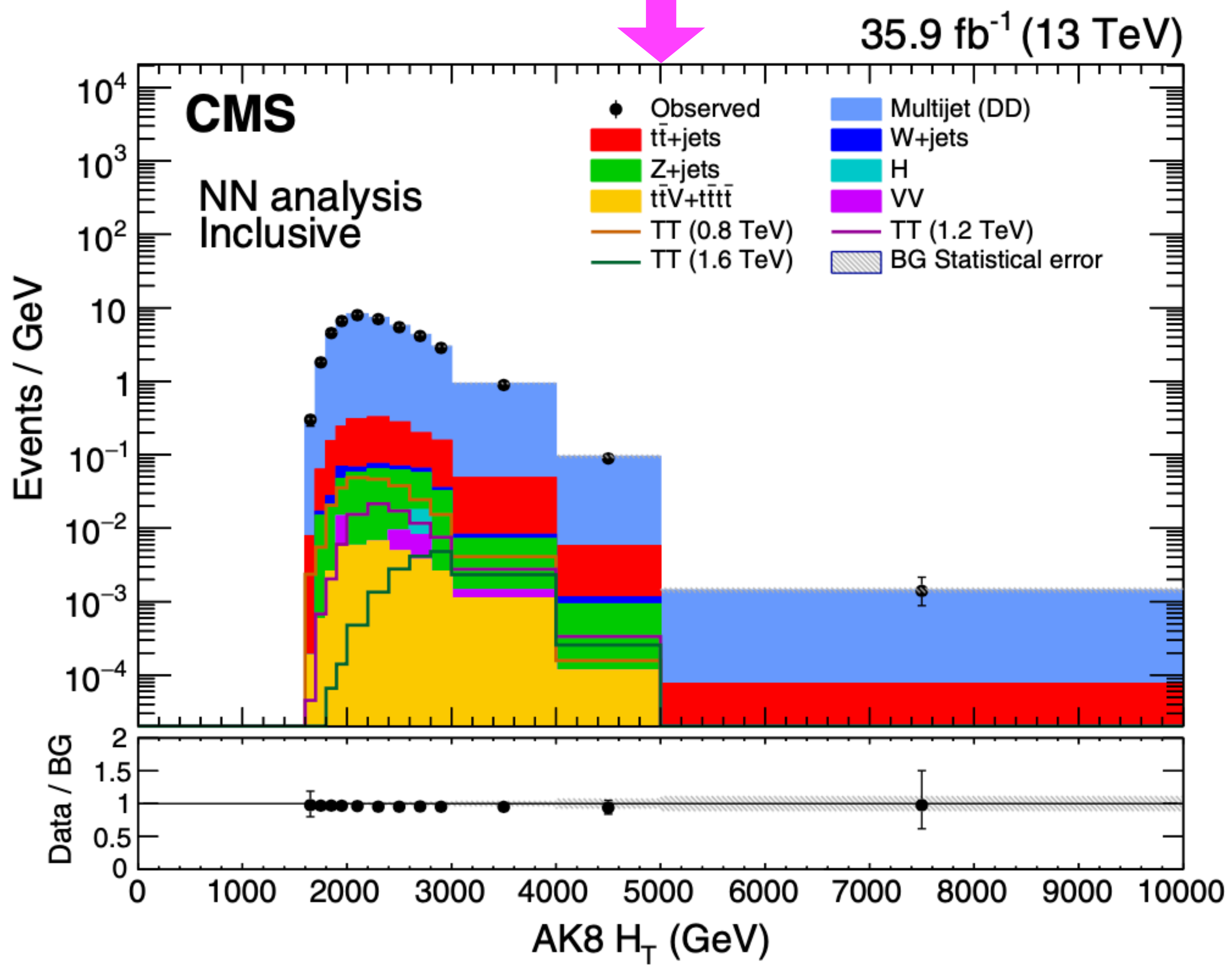
r is total number of expected BG events in each SR/VR

$$r = \sum_{events} \sum_{permutaions} \prod_{i=1}^{4} \epsilon_X(p_{T,i})$$

— For each SR/VR, there is a separate r

— Event weight is calculated by multiplying the leading four jets' tagging rate obtained from the multi-jet-enriched CR, but a sum over all possible permutations of the tags is needed, since relative p_T order of tags is unknown

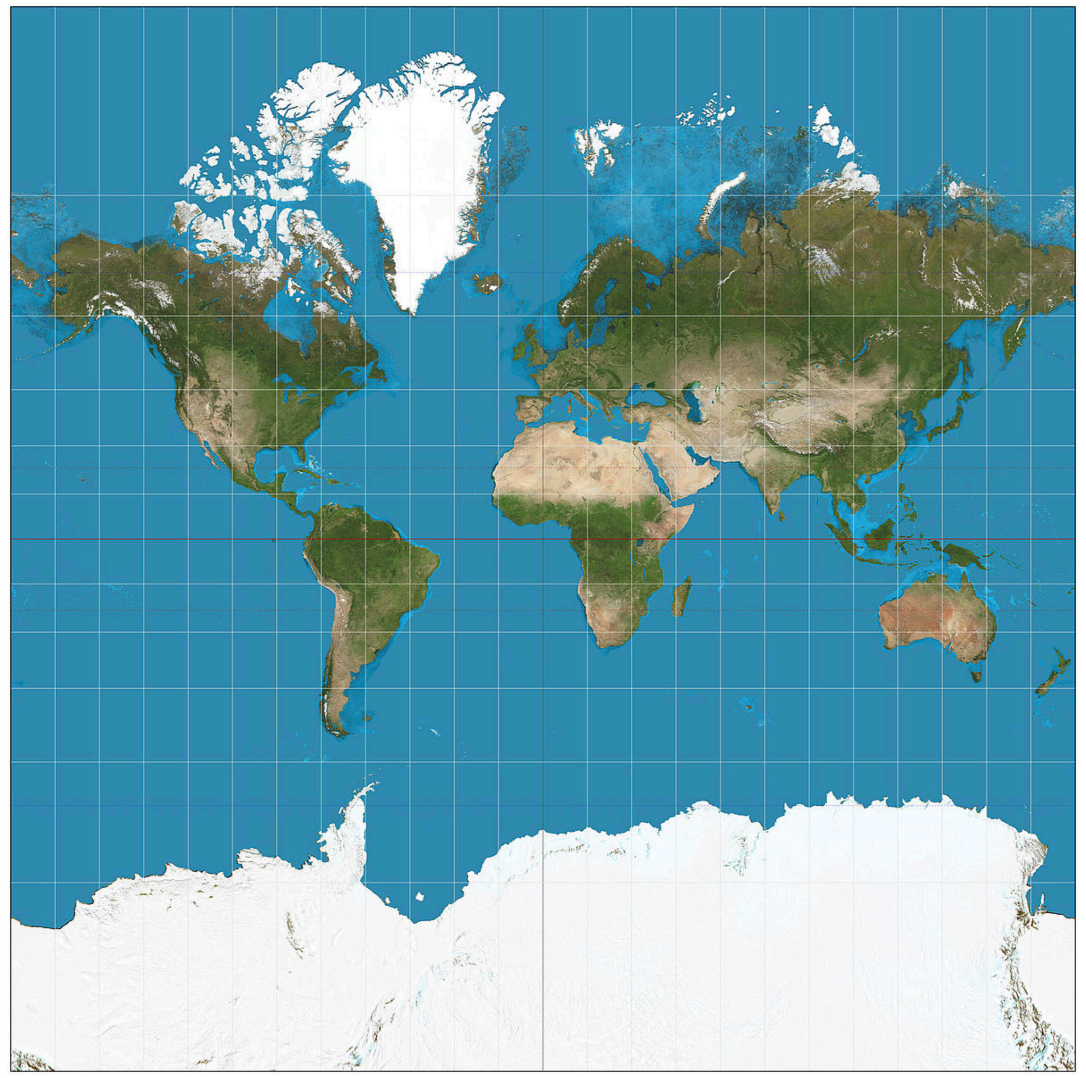
— Sum over events can be interpreted through an Ht distribution



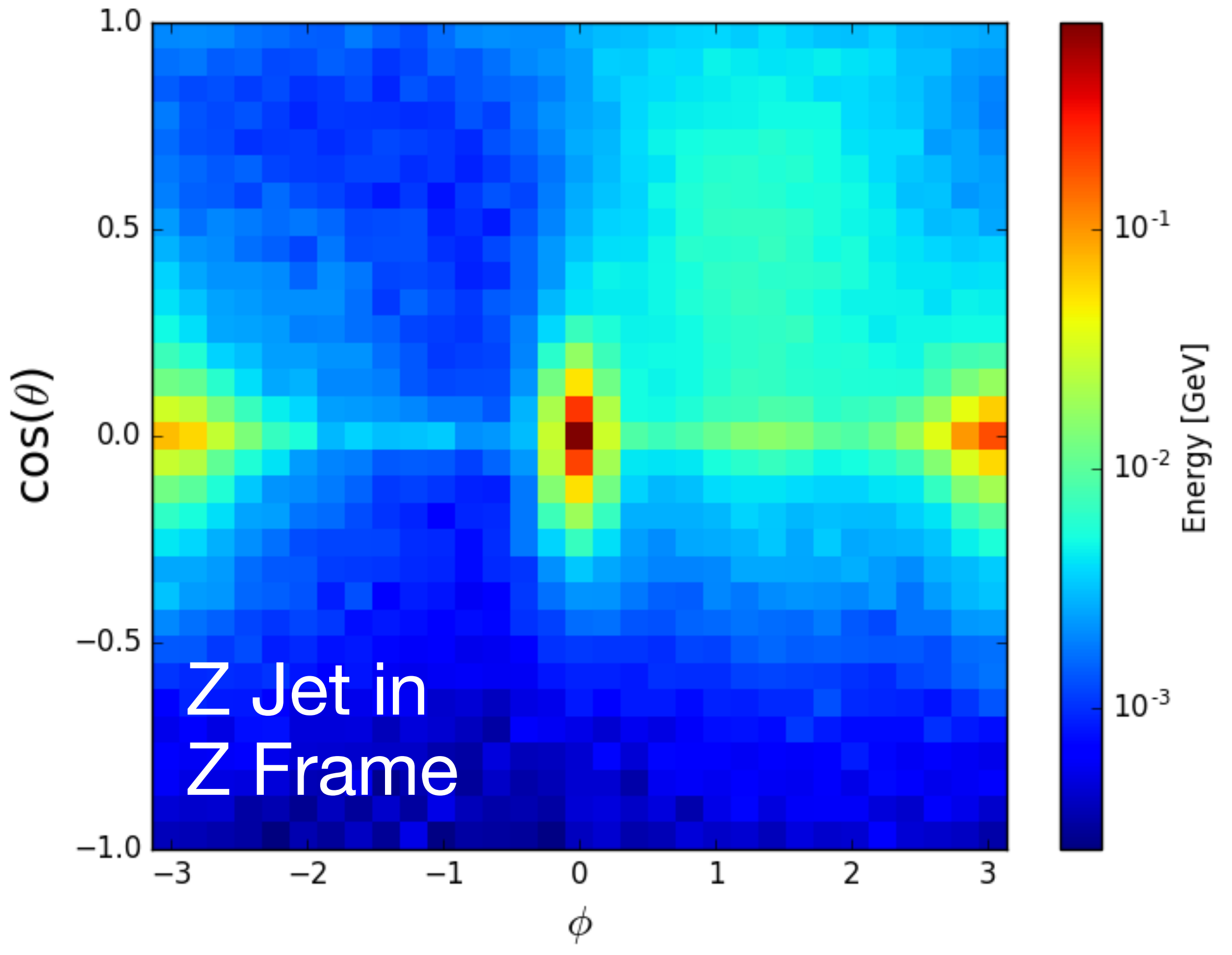
[1]: PhysRevD 100, 072001

Projecting Jet Images

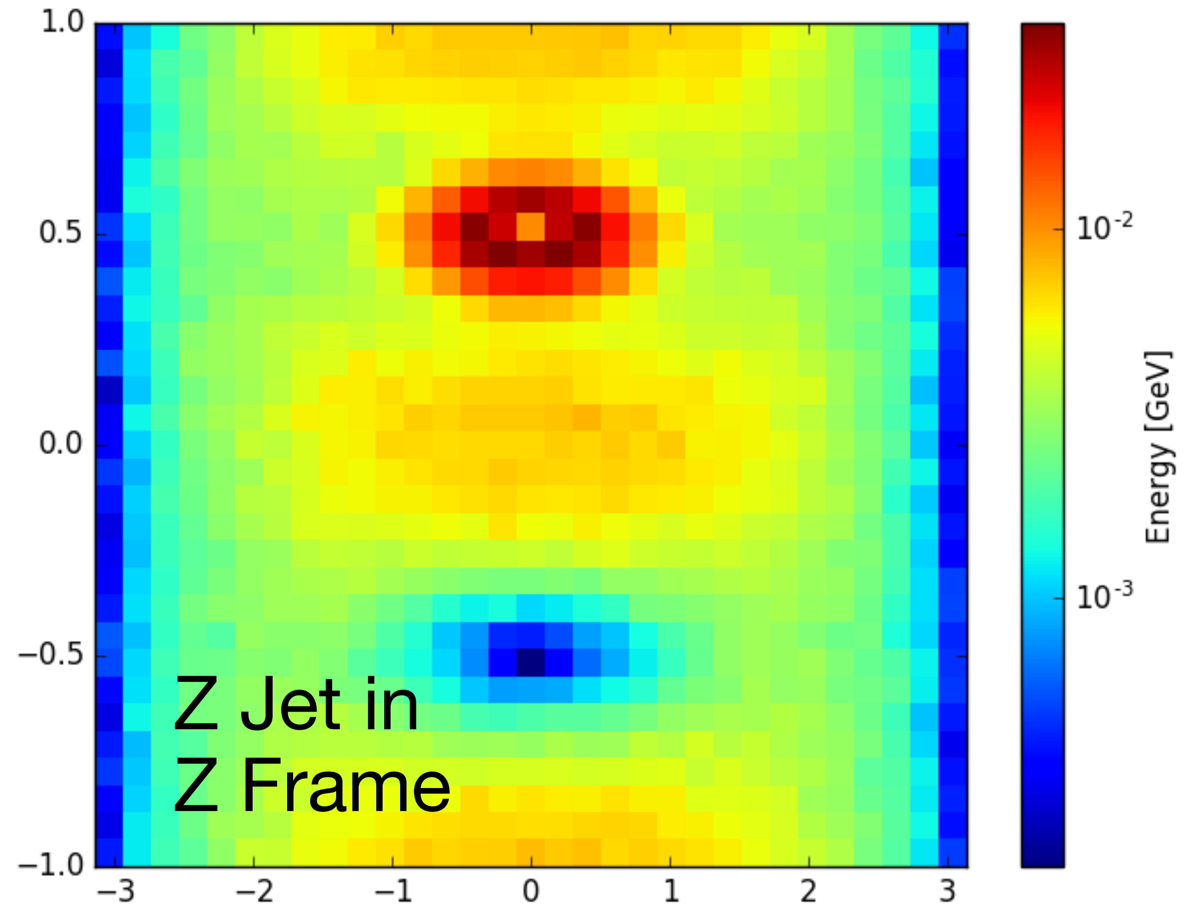
Representing 3D -> 2D



Cylindrical Projection



Cassini Projection



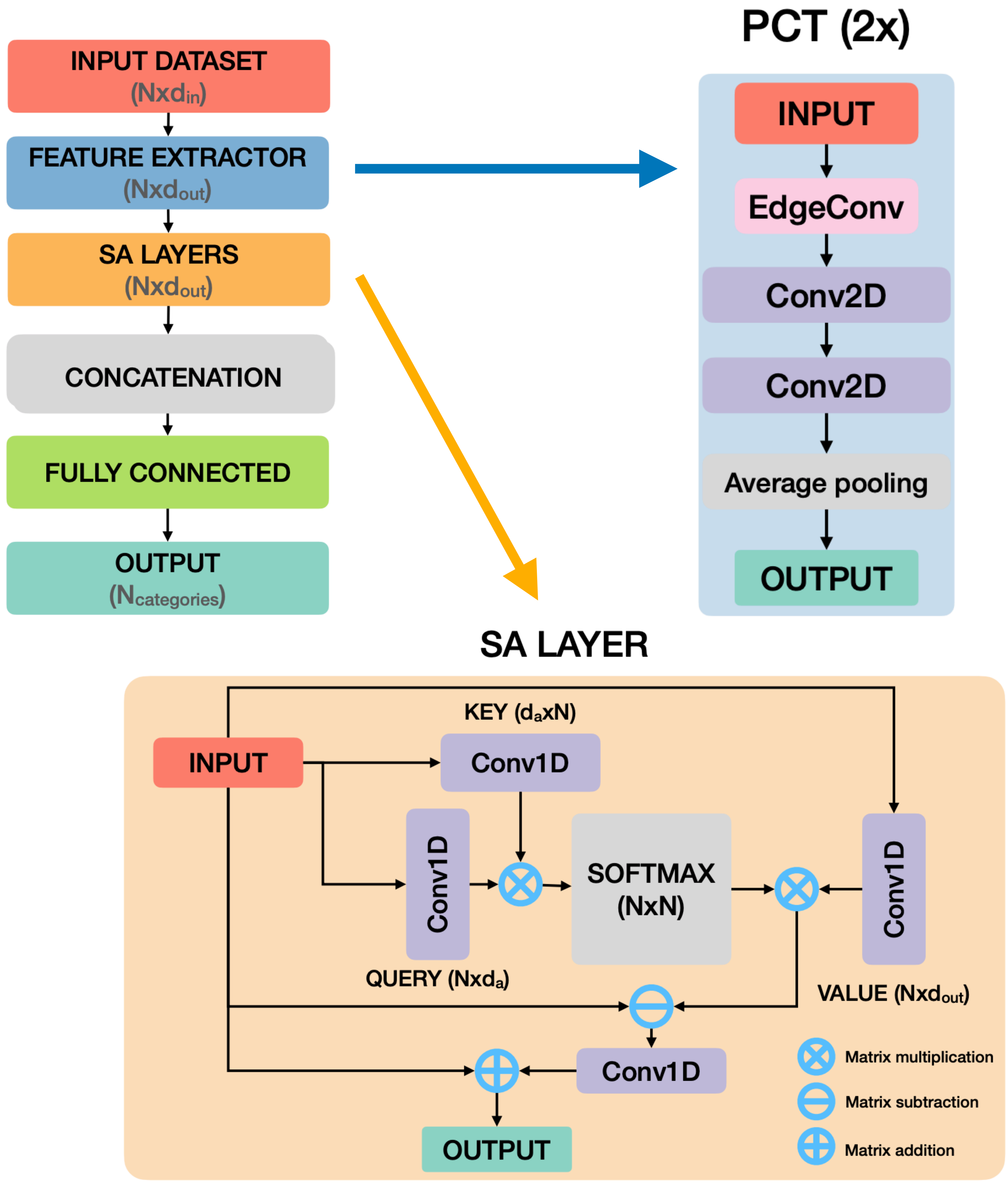
Projecting forces a choice of granularity and transformation

Note: 31x31 ~1k pixels for relatively small (~30) occupancy

More Flexible Data Representation

Application of Point Cloud Transformers[2]

- Point clouds are graph networks relating edges between input features of vectorized data
- Transformer -> attention mechanism
- PCT outperforms Image (CNN) network
  - k-nearest-neighbors: $\Delta\eta, \Delta\phi$ with respect to jet
  - Further optimizations could be done
- WIP: Integrating BES vars into multi-frame PCT
  - Generalize ‘order’ of constituents for clustering
  - Take advantage of ‘order’ of frame boosts, i.e. $m_b < m_W < m_Z < m_H < m_t$



[2]: arxiv:2102.05073

BESTagger (DNN)

Improvements

Extracting information beyond BES

— Other physically-motivated vars?

— Can networks learn more w/ jet constituents

— How to best represent data?

Inputs

59 → 123-142 BES vars

NN Architecture

40 x 40 x 40 → 512 x 256 x 256 x 256 x 144 Dense Layers

Outputs

6 Dense Layer (classes)

More BES vars and additional ‘b’ frame

Integrate low-level jet constituent info

BESTagger

Considerations and Improvements

For 2016 result, 59 BES variables and simple DNN chosen

- Vars: Aplanarity, asymmetry, isotropy, FW, jet charge/eta/m_{SD}/$\tau_{N=1,2,3}$, sphericity, subjet CSV/$m_{i,j}$, thrust
- DNN arch: 40x40x40x6 on 500k events

Outstanding Questions:

- Any other physics-motivated BES variables?
- Can architecture improve performance?
- Are we 'missing' information?

'New' BEST, 142 BES variables and deeper NN

- New Frame: aim for B-hadrons @ 6 GeV
- Adding Vars: nSecondaryVertices, relative coordinates w.r.t. jet centroid, re-clustered jet multiplicity, leading four subjet energies
- DNN Arch: 512x256x256x256x144x6 on 2.2M events

Extracting information beyond BES:

- Integrate low-level jet candidate information, but how to most efficiently represent data?
- CNN (image) vs RNN (sequence) vs PointCloud (graph)

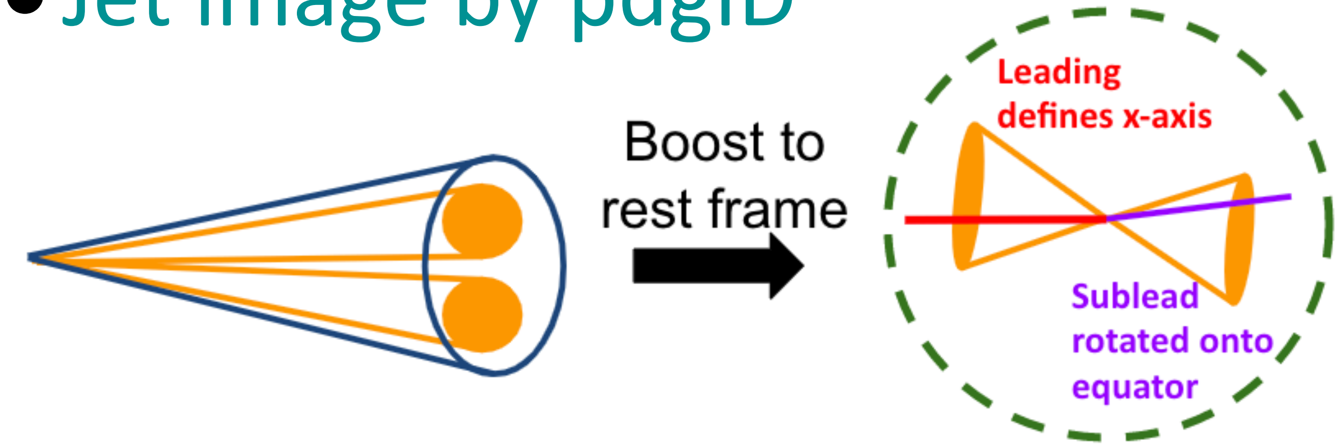
BEST+CNN
and Imaging Procedure

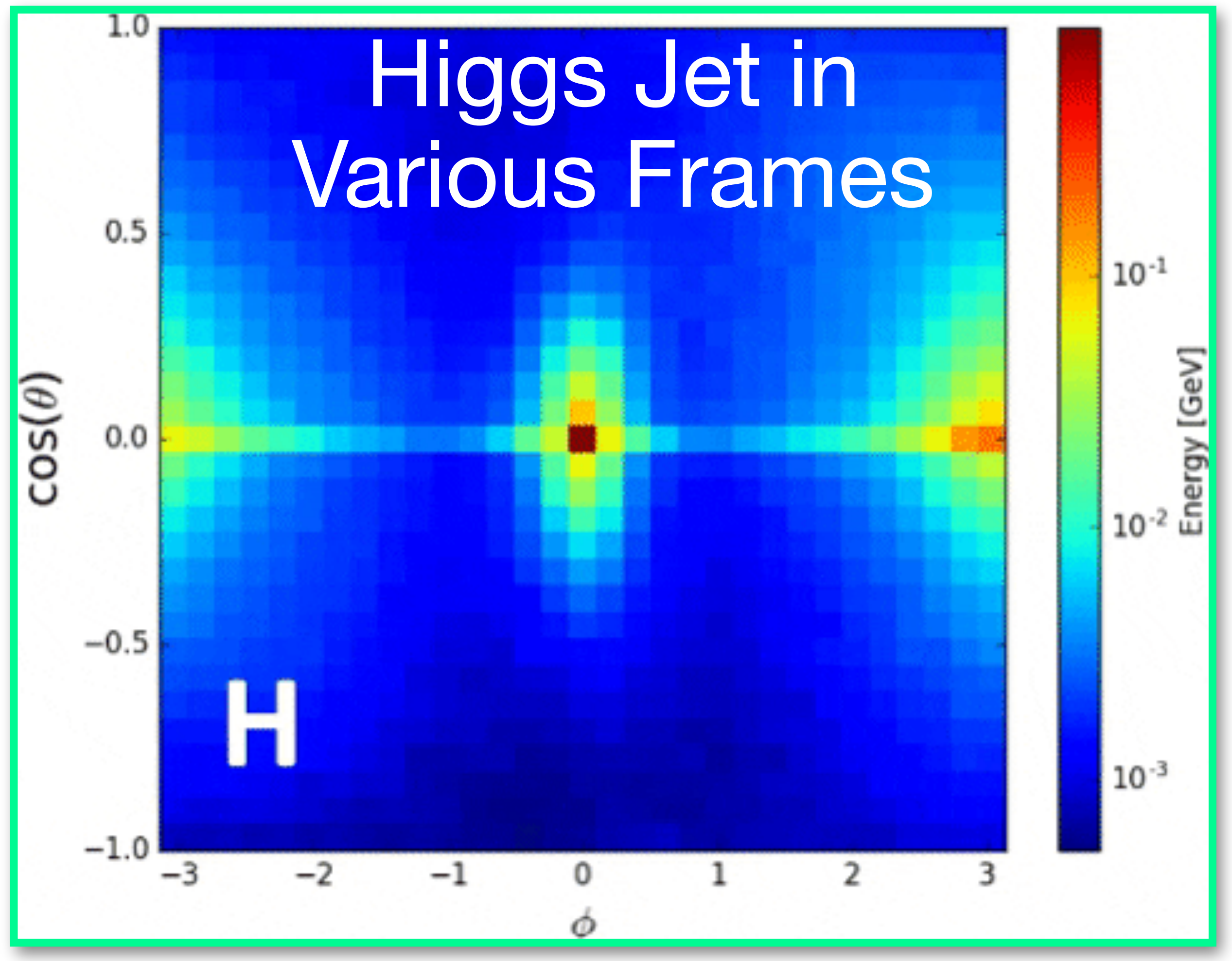
Key Idea: Study substructure in boosted frames

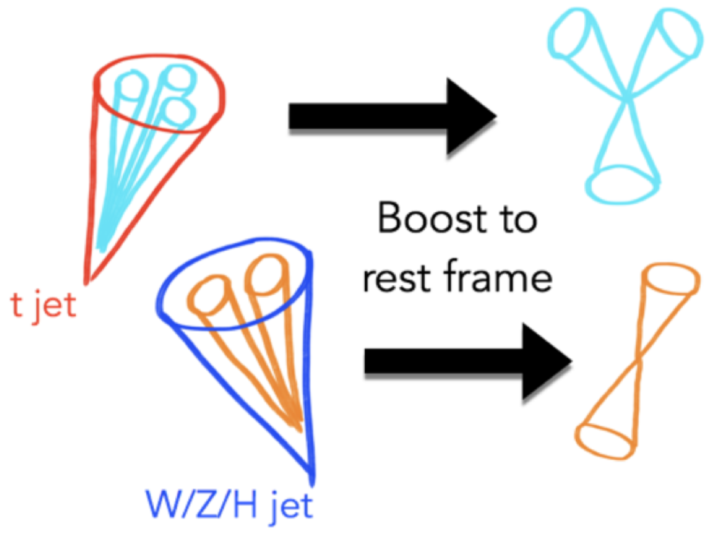
- Neural network based AK8 jet tagger
- Six Classifications: H, W, Z, t, b, light-jet

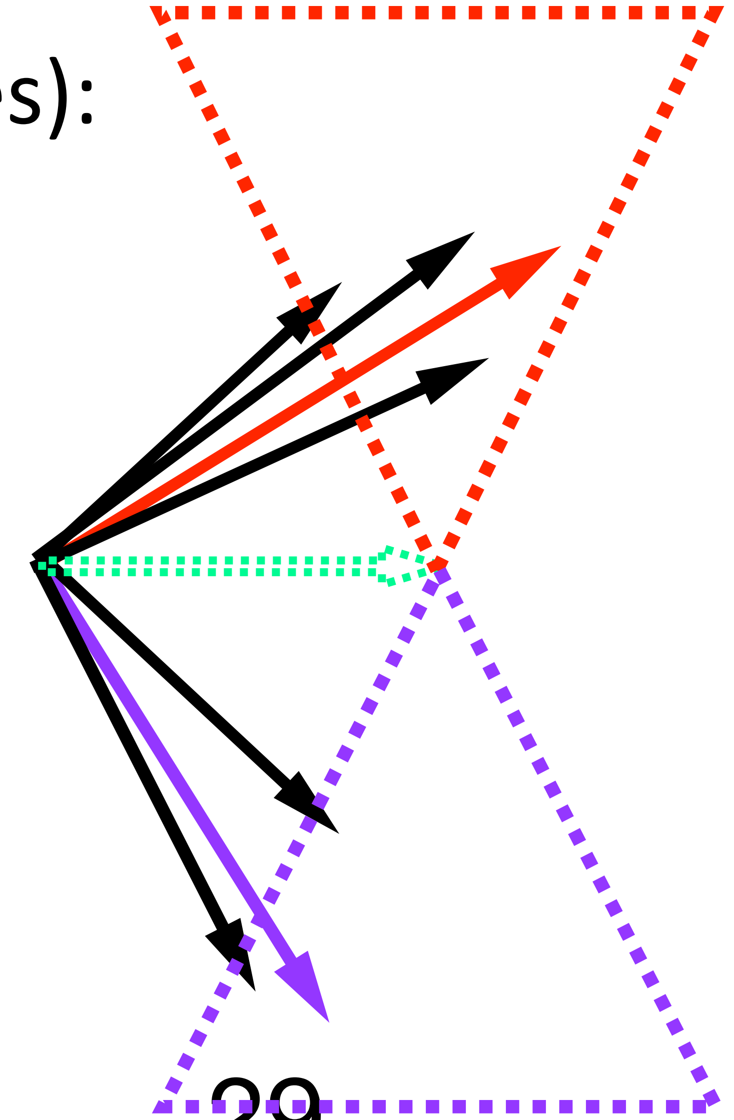
Input features in various frames (b, H, W, Z, t, lab)

- Boosted Event Shape variables:
  Fox-Wolfram moments, sphericity tensor, thrust
- Jet Substructure (reclustered jets in boosted frames):
  m_{ij} , $\cos\left(\theta_{ij}\right)$, $m_{eff,1-4}$, $p_{L,1-4}/p_{T,1-4}$
- deepCSV-bdisc (NEW!)
- Jet image by pdgID









Jet Images

Four frames: H, Z, W, t

— First, boost in frame of choice

— Reclustered PFcands into AK5 jets

— Define x-axis with leading-E PFcand ($\hat{x}$ becomes into-the-page in image)

— Define y-axis by rotating sublead PFcand onto $\theta = 90$ (equator)

— Cylindrical projection of 'sphere'-space onto 'flat' $\left(\phi, \cos\theta\right)$-space

— Perform Reflections s.t. highest-E quadrant is in upper right

— Pixel Value: Energy of PF candidate, normalized by leading pixel (range: 0-1)