Highlights from

ATLAS
Exotics
+
HDBS
Workshop
11-14 June 2019

Exotics: 11-14 June 2019
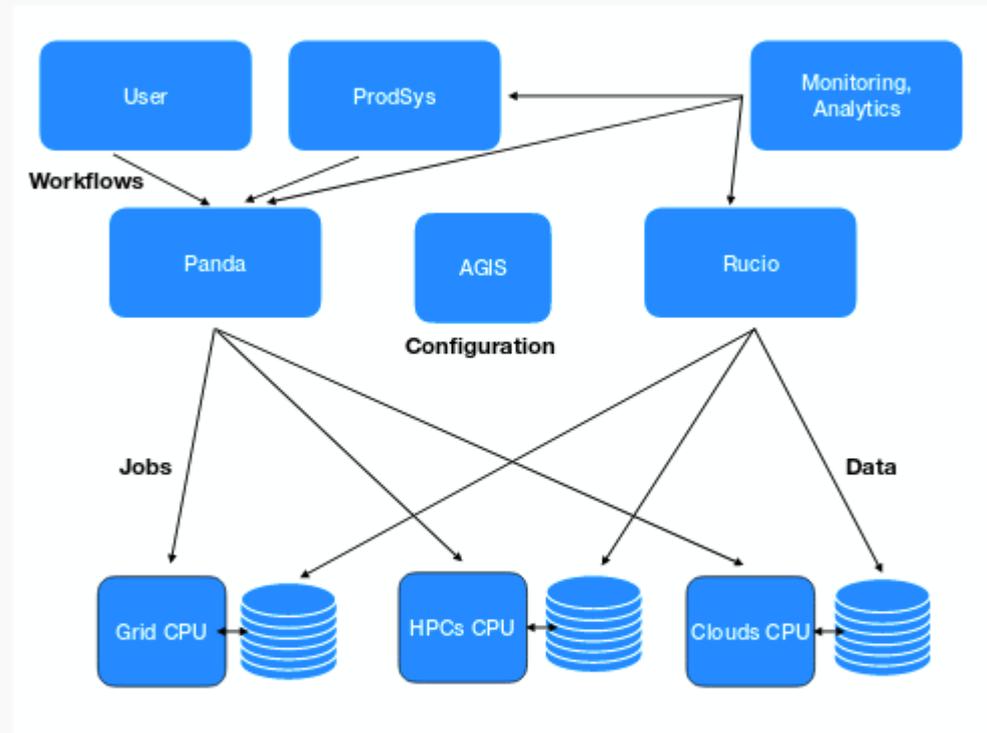HDBS: 11-13 June 2019

VILLA DORIA D'ANGRI
Via Francesco Petrarca 80 Napoli

Current Computing model

The ATLAS distributed computing system is centered around:

- **Workflow management system**: Panda
- **Data management system**: Rucio
- Many **additional components**: AGIS, ProdSys, Analytics, ...
- **Resources**: WLCG grid sites, Tier0, HPCs, Boinc, Cloud
- **Shifters**: Grid, Expert and Analysis (ADCoS, CRC, DAST)

Aim for a 30% reduction in disk space for Run-3, see talk here

How do we achieve that?

Formats	Introduce **DAOD_PHYS** with $\sim$50 kB/event
	Introduce **DAOD_PHYSLITE** with $\sim$10 kB/event and **calibrated objects**
	Reduce number DAODs formats, use these for CP, systematic and R&D studies
Production	Stop open-ended production for data DAODs
	Use a **tape carousel model for AOD** inputs in parts of the DAOD production
	Consider caps on sizes of individual DAOD type datasets
	Bring Rucio redirector with global name space into production
	Smart DAOD replica placement on the grid sites
	Increase usage of docker/singularity containers for analysis
	and group ntuple production
AOD/DAOD content	Significantly **reduced track, trigger, truth** information, use **calibrated objects**
	Apply **lossy compression** for most variables in AOD/DAODs
	Avoid any information duplication in the AOD/DAODs containers

Generally aim for a smaller DAOD format
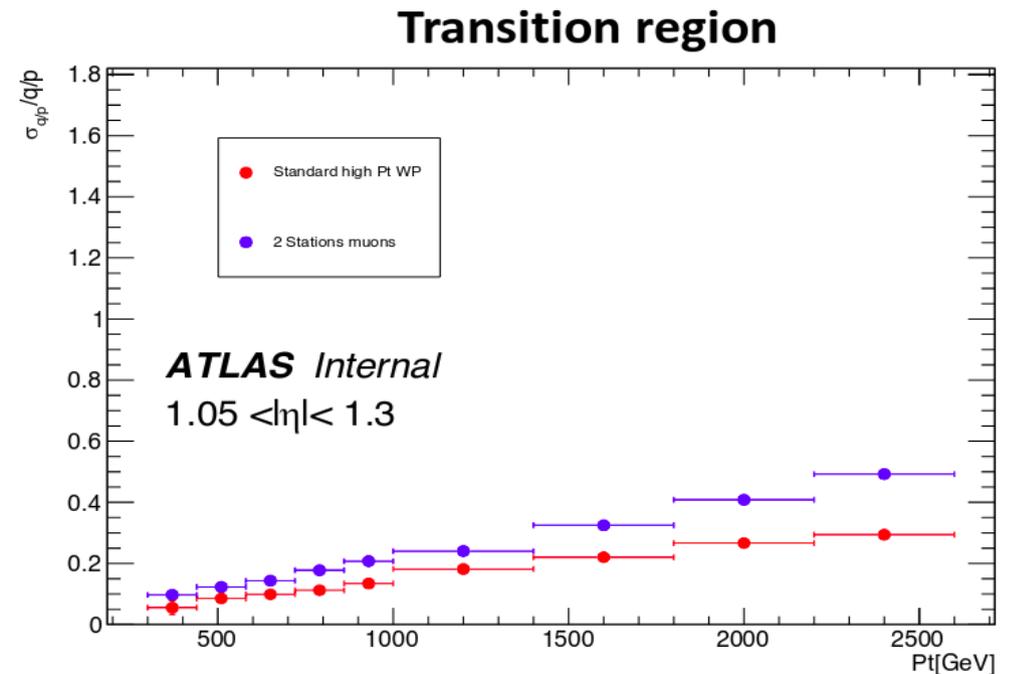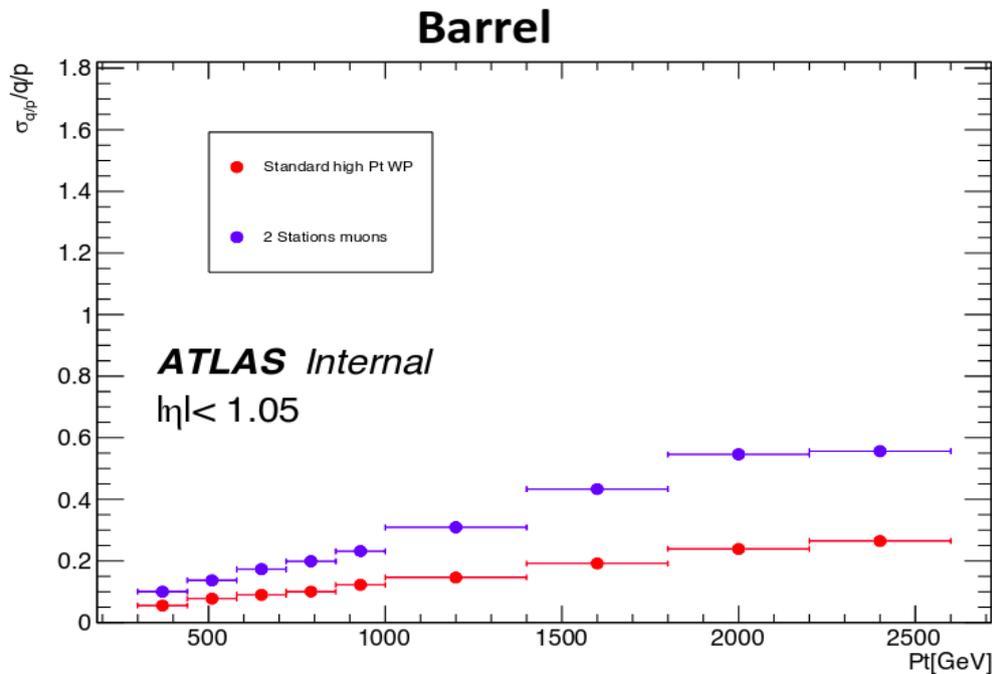
Muon Reconstruction WPs

Low p_T :

- good muon reconstruction efficiency down to p_T of ≈3
- keeping the fake rate under control
- only CB tracks are used
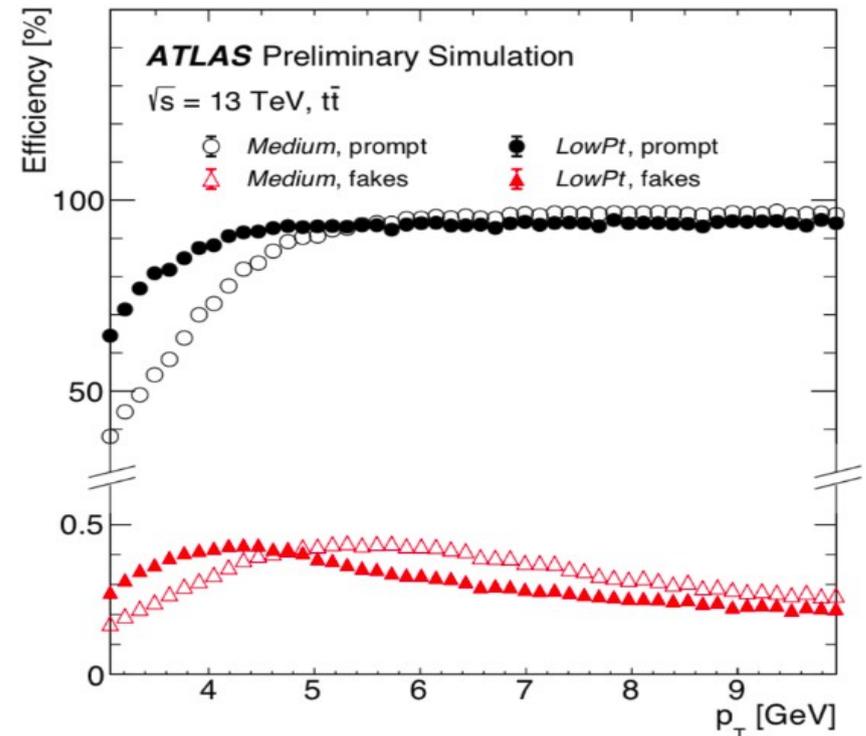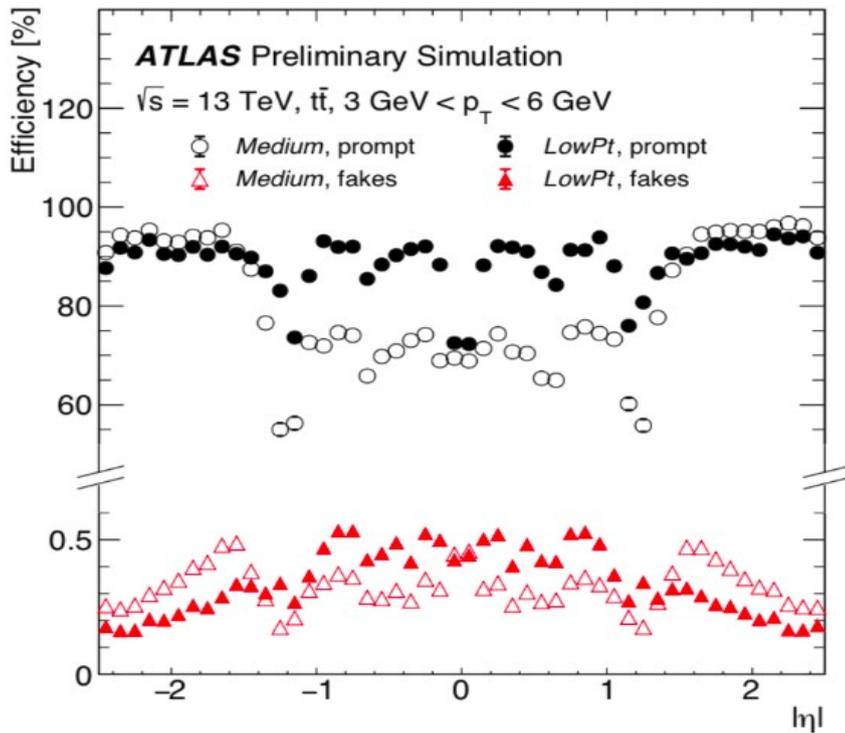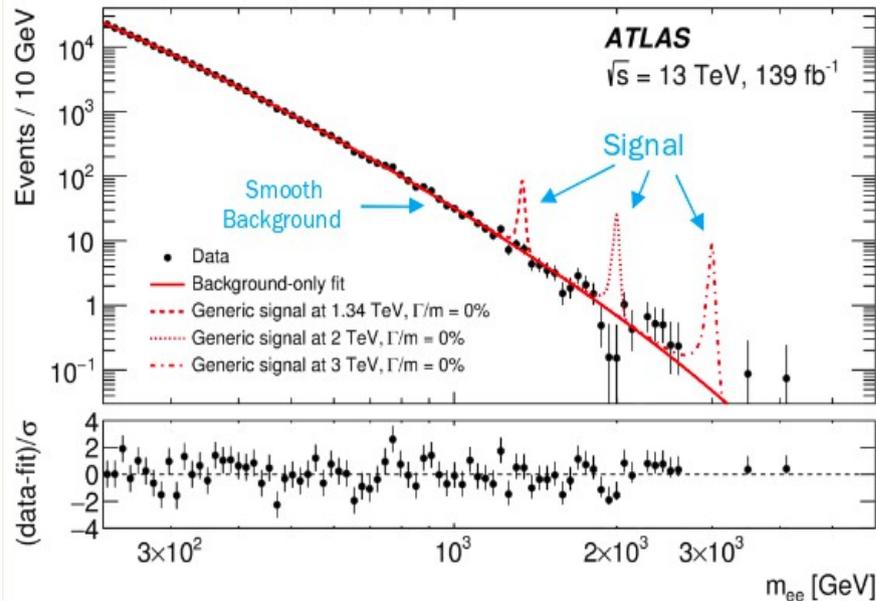
Muon Reconstruction WPs

High p_T:

- optimize to provide good momentum resolution for very high p_T muons, of O(1) TeV
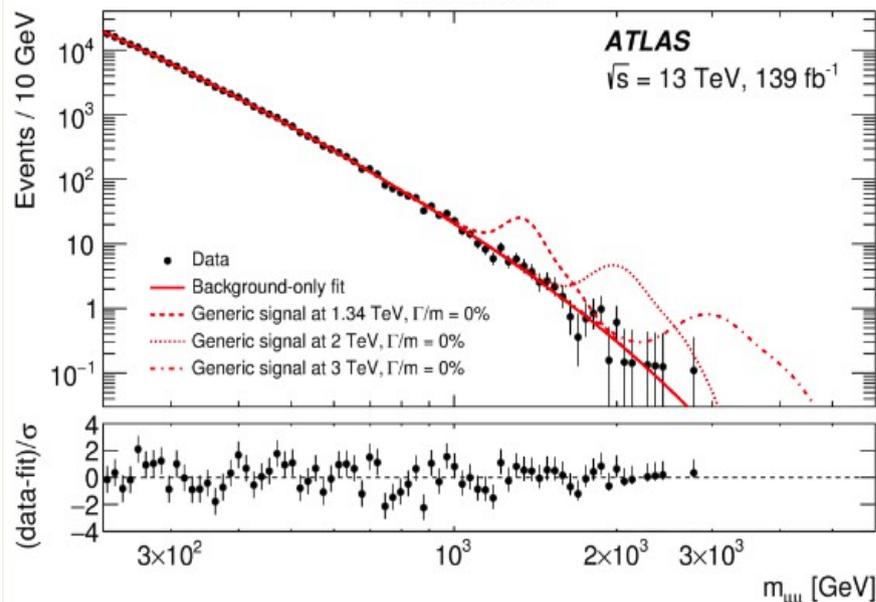- only CB tracks are used

2 Station muons p_T resolution ~2 times worse!

Future searches with leptons

Dilepton search recently published the Full Run-2 dataset analysis, now aiming for exclusive final states and combinations with other analyses

Dilepton+b-jets: Some of the observed B-meson decay anomalies could be explained by Z' resonances that couple mostly to third generation quarks and second Generation leptons. An exclusive channel with 1 or 2 b-tagged jets allows to search for such resonances. A reinterpretation along those lines can be found in http://arxiv.org/abs/1904.10954

Clockwork analysis: narrowly-spaced series of effectively zero-width resonances is predicted starting above a certain mass threshold, usually much above the SM Z-peak. Allows for a solution for the hierarchy problem

Lepton-MET+Dilepton Combination
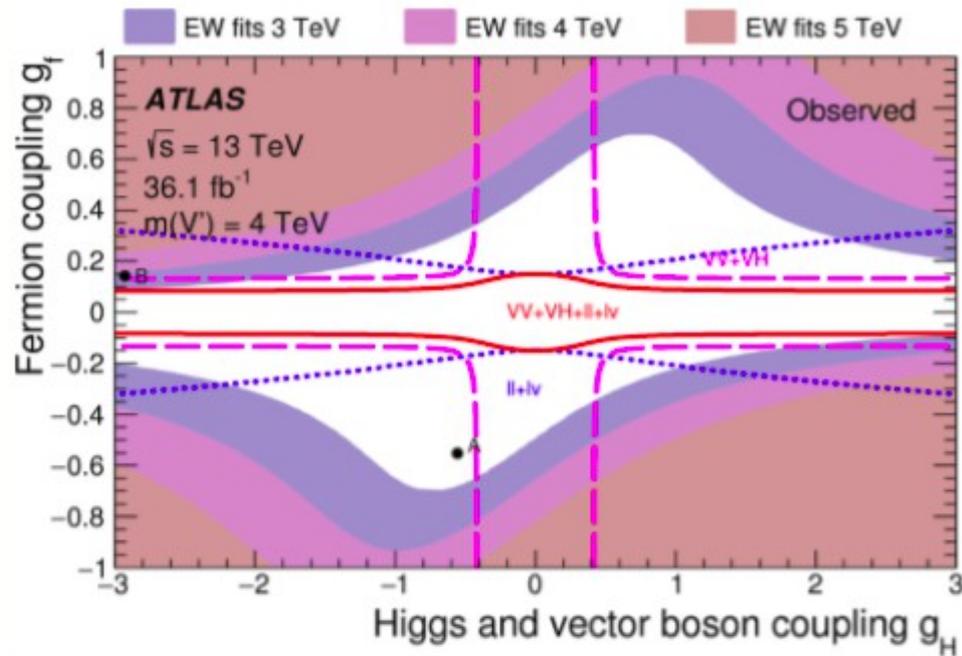
Limits extracted on the Fermion-Higgs/Vector Boson and Quark-lepton coupling parameter space

▪Area outside the curve is excluded▪HVT bosons can couple to fermions (f), leptons (l), and Higgs (h)

▪Currently preparing inputs for heavy resonance combination•Standalone HVT dilepton results weaker than 2015+16 combination, so still a lot to be gained.
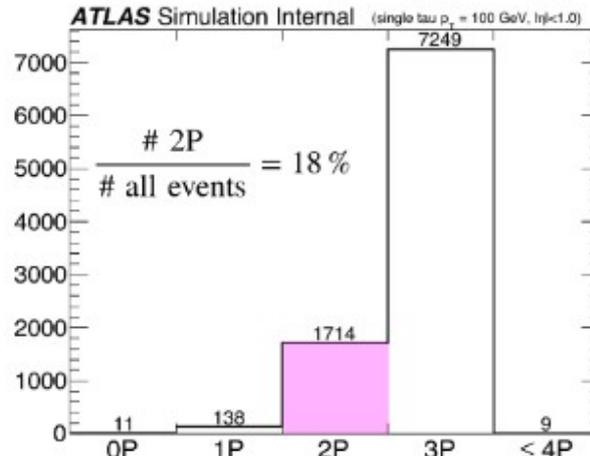
▪A W' Roostatimplementation is now being worked on•

Aiming for an even larger combination with VV+VH, jets and VLQ
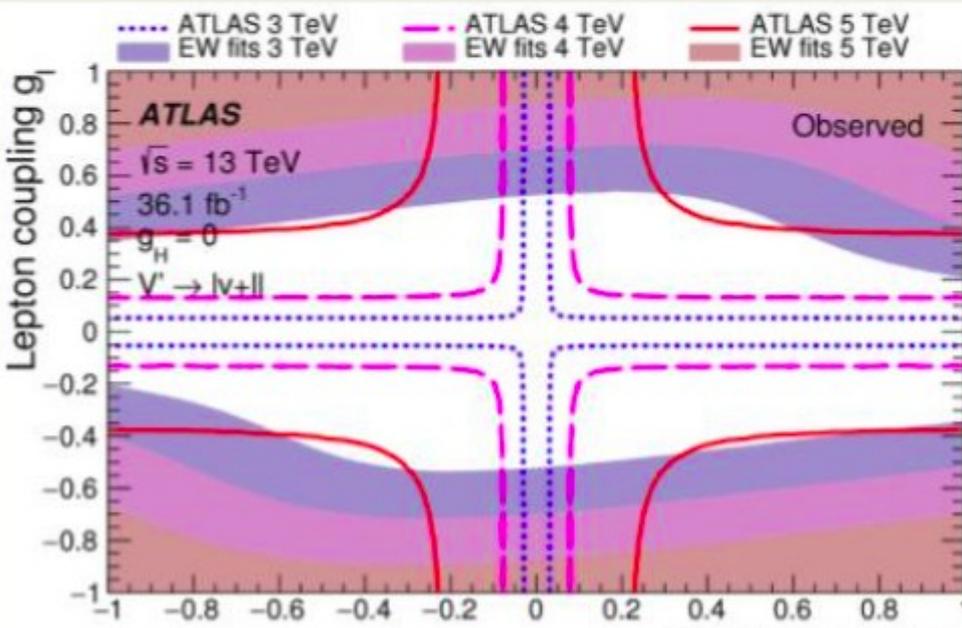
Tau Reconstruction Updates

- Some Taus are reconstrcuted as 2-prong:
  - Unphysical, usually rejected
  - Loose stats - esp. at high p_T

truth 1-prong tau

"Truth1-2P"

reconstruction

truth 3-prong tau

reconstruction
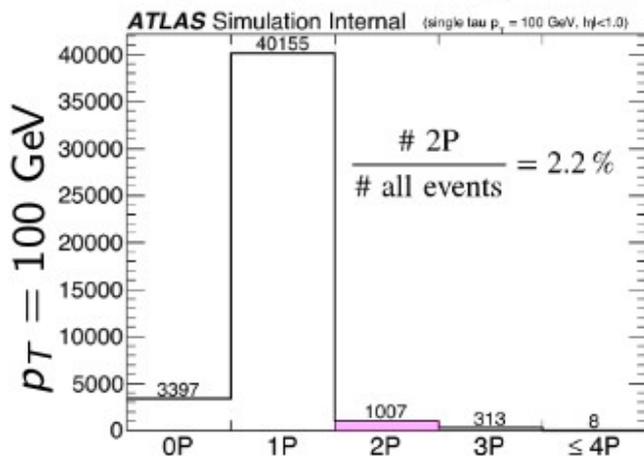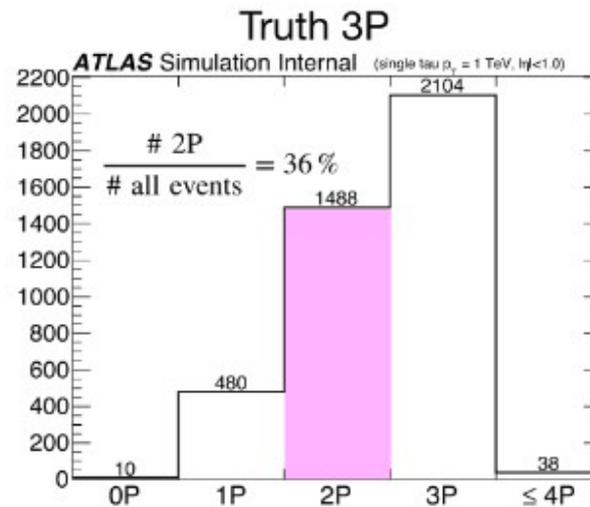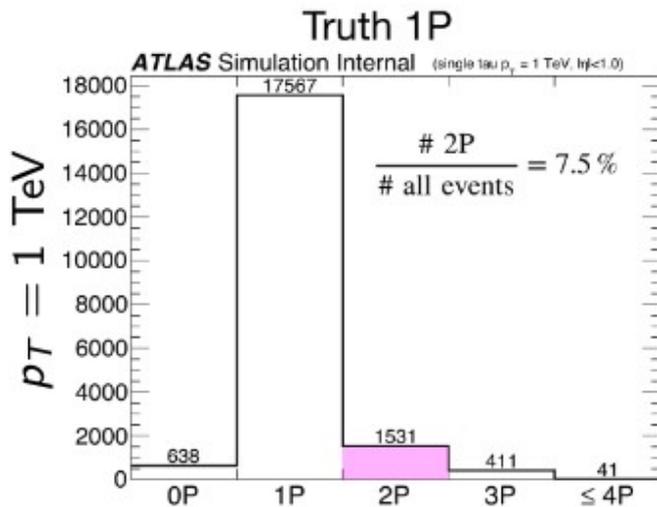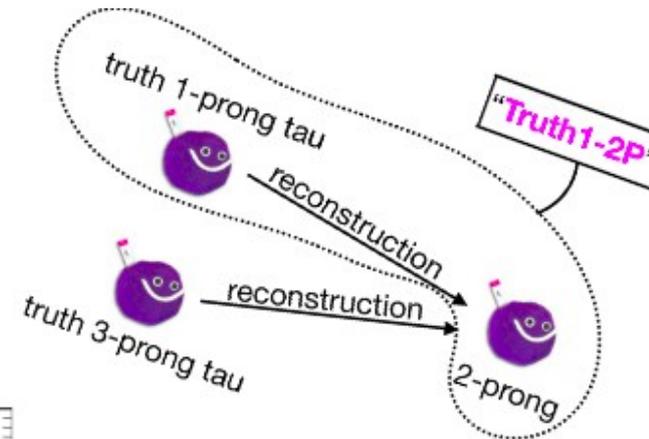
2-prong

Plots by Kosuke Takeda

Two kinds of 2-prong Taus

- Truth 1-prong aka 'Truth1-2P'
- Truth 3-prong aka 'Truth3-2P'

Two-Prong Tau ID BDT

- T1vT3 uses 8 input variables, leading one
  - EleMatchLikelihoodScore

- Current situation for 2-prong Taus:
  - <u>No</u> dedicated 2-prong TauID algorithm
  - Use 3-prong BDT to discriminate against QCD jets

- Possible new approach:
  - Train dedicated 2-prong TauID BDT
    - Use T1vT3 output as input
  - Significantly improves BKG rejection for 2-prong Taus at high p_T
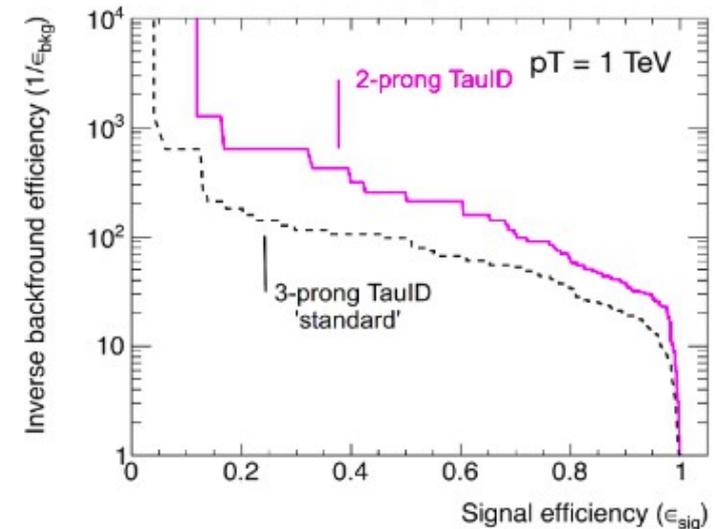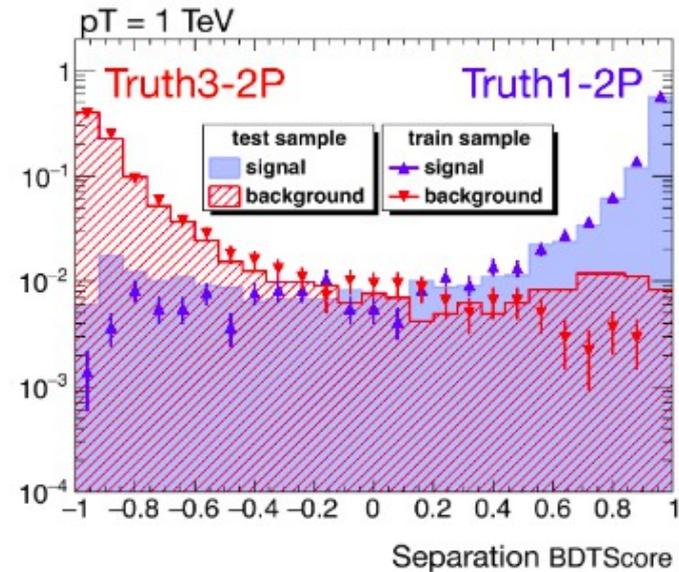
RNN Tau ID for fake rejection

BDT TauID

- 12 'high-level' input variables

RNN TauID

- BDT input variables
- Track-level variables
- Cluster-level variables

RNN clearly outclasses BDT ID

- Expect $\approx 30\%$ higher di-Tau yield

But: No Scale Factors provided for RNN ID by tauWG yet
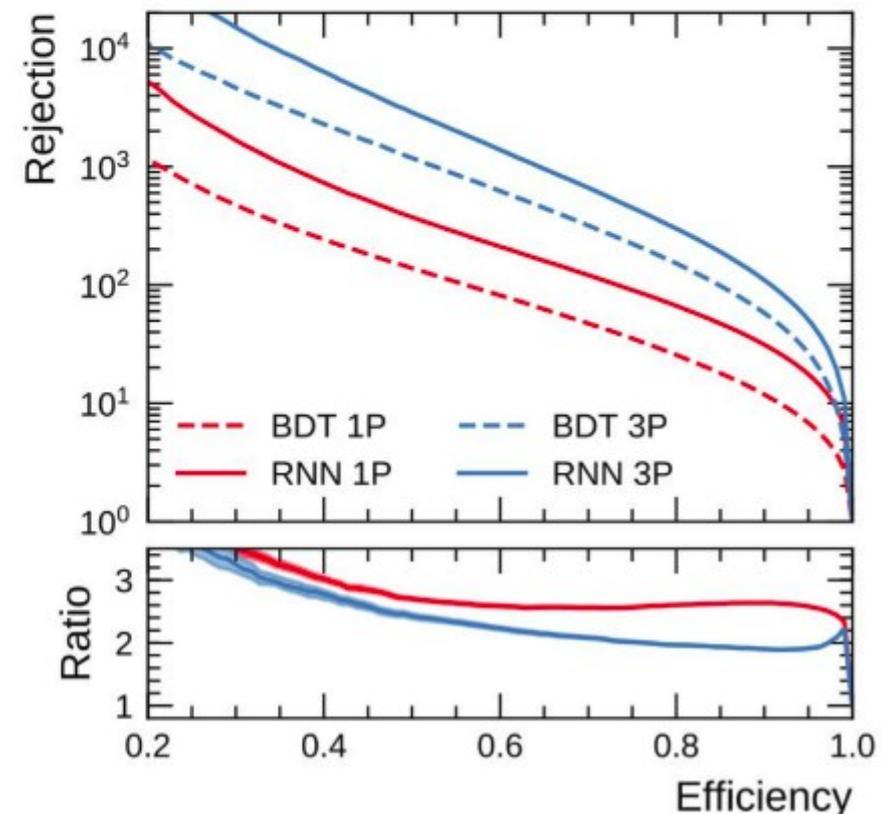
How to run on the grid efficiently

- Submit as little tasks as possible — it takes a long time to submit and process each task (and also causes a lot of bookkeeping).
- Do not set files per job, events per job! Panda can learn from the scouting jobs in most cases (only set N GB per job if really necessary).
- Panda learned a while ago how to properly submit a single task that would process/produce multiple datasets:
  - `--addNthFieldOfInDSToLFN` argument of `pathena/prun` to name your files based on the input
  - `--useContElementBoundary` to not merge the output between the datasets.
- EventLoop python example:

```
driver.options().setString(EL.Job.optSubmitFlags,
    '--addNthFieldOfInDSToLFN=2,3,6 --useContElementBoundary')
```

Heavy Resonance Combination

Harmonization/planning at the analysis level makes combination easier:

- **Orthogonality checks:** want to address overlap between analyses as early as possible.

  - Additional cuts can be implemented; in most cases, goal is to migrate, not remove events.

- **Interference:** addressed in a couple of analysis (tt, $\ell\ell$, ℓv), negligible in others.

- **Object harmonization and systematic uncertainties:**

  - Common naming, pruning, smoothing schemes. Should already be consistent for analyses using same tools.

  - Different taggers or working points requires input from CP groups. E.g. b-tagging: if same tagger, can use uncertainties from pseudo-continuous b-tagging to take correlations correctly into account for different WPs.

  - Large number of NP makes fits unstable: highlights importance of pruning. Toy limits can take a long time (>weeks).

- **Signal scaling** is error prone:

  - All signals should be scaled assuming a cross-section of 1 fb. <u>see twiki!</u>

- **Production mechanisms:** want to approach e.g. VBF production in a more consistent way across analyses. Most general approach is to fit all signals in all regions, for more flexibility during combination.

  Different timescales for each team: important to address these as early as possible.

Hands on session: Rivet

What is Rivet?

Robust **I**ndependent **V**alidation of **E**xperiment and **T**heory

- generator-agnostic, efficient and fast way to produce lots of physics plots
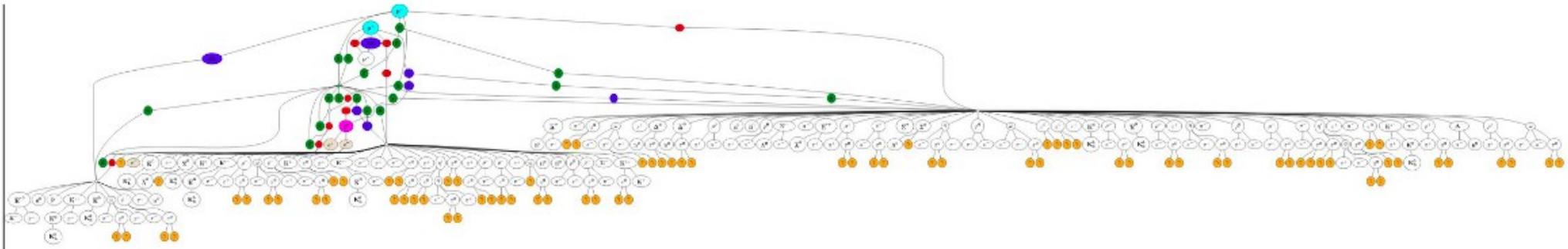
- lightweight means of exchanging analysis details and ideas
- used extensively for validation and tuning of MC generators as well as reinterpretation of measurements (e.g. using **Contur**)
- comes with its own smearing machinery for fast detector simulation

Hands-on session: Unfolding

What is unfolding?

Particle-level (aka Truth-level)

We know what this looks like (for bkg or signal) in Simulation

What happens when two protons collide

We want to know what this looks like in nature (=data)... eg:
- does this signal exist?
- is this process well-modelled ?

Reconstruction
aka
'Forward Folding'

ATLAS: imperfect resolution (migrations in energy, position) imperfect reconstruction (mis-identification of particles) acceptance, efficiency

Unfolding

Detector-level (aka Reco-level)

We can estimate (imperfectly) what this looks like in Simulation using reconstruction, GEANT4..

What can actually be measured in the detector

We know what this looks like for data. This is what we observed.