

Tau studies with TauDPDMaker in FDR-1

David Côté



DESY-ATLAS meeting
January 28th 2008



What's the Full Dress Rehearsal (FDR)?

- ❑ FDR is a test of the complete ATLAS data distribution and analysis chain
- ❑ Fully realistic simulation of...
 - 10-hour fill at average luminosity of 10^{31} and 1-hour fill at average luminosity of 10^{32} (FDR-1)
 - ❑ 0.72 pb^{-1} , ~ 8 million events accepted by EF
 - FDR-2: 10-hour fill at 10^{32} luminosity
- ❑ Pre-simulated and mixed RDO events downloaded at ATLAS "Point 1"
 - During the FDR-1 week:
 - ❑ Run triggers and send accepted events to Tier 0
 - ❑ Reconstruct the events, make ESD, AOD and **DPDs**
 - ❑ Send data to Tier 1, Tier 2 ...
 - ❑ **Analyze** the data!

<https://twiki.cern.ch/twiki/bin/view/Atlas/FullDressRehearsal>

What's a Derived Physics Data (DPD)?

- The “DPD” terminology was introduced in the Analysis Model Forum report (AMF)
- DPD is a “distilled” ESD or AOD collection, potentially incremented with DPD-specific “Analysis Data” (or “UserData”)
 - “primary DPD”:
 - pool.root format, centrally produced
 - “secondary DPD”:
 - pool.root format, privately produced
 - “tertiary DPD”:
 - classic root ntuple, privately produced

Why do we need DPDs ?

- ❑ Skimming will be mandatory to analyze monster ATLAS dataset
 - Example for FDR-1 data (0.72 pb^{-1})
 - ❑ nb $Z \rightarrow \ell\ell$ events: $< 3670^*$
 - ❑ nb events accepted by EF: ~ 8 Millions
- ❑ Container slimming reduces event-size
 - ESD: $\sim 680 \text{ kB} / \text{evt}$
 - AOD: $\sim 170 \text{ kB} / \text{evt}$ (including MC)
 - Primary TauDPD: $\sim 15 \text{ kB} / \text{evt}$ (excluding MC)
- ❑ Skimming, slimming and thinning are intrinsically specific to analysis groups
 - Hence the need for group-specific DPD!

$^* \sigma_{Z \rightarrow \ell\ell} = 3 \times 1.7 \text{ nb} = 5100 \text{ pb}$

Goals of the Tau WG for FDR-1

- ❑ Establish early data validation procedure of tau-ID
 - Key analyses of data control samples:
 - ❑ Tau fake rate from dijets sample
 - ❑ Tau efficiency and energy scale from W and Z samples
 - Need to decide on skim selections and DPD-event content
- ❑ Validate tool: TauDPDMaker
 - Makes primary, secondary & tertiary DPD
 - ❑ specific to the Tau WG
 - Does skimming, slimming, thinning, ...
 - ❑ as recommended by Analysis Model Forum
 - TauDPD requirements:
 - ❑ event content sufficient to complete control sample analyses
 - ❑ average event size < 10 kB
 - ❑ tau re-calibration doable on skimmed data (FDR-2 ?)

Let's take advantage of the high-statistics FDR samples!

What is/does TauDPDMaker?

- Python script making DPD for the Tau WG

- Primary DPD:

- data control samples ($W \rightarrow lv$, $Z \rightarrow ll$, $W/Z \rightarrow jj$, dijets)

- Secondary DPD:

- TauDPDMaker can re-run over its primary TauDPD...

- Tertiary DPD (“TauView” ntuple):

Sylvie →

- analysis of data control samples (Z , W , dijets)

Björn →

- performance comparison of tau algorithms

- Other:

- release validation
 - core tau-ID development

<https://twiki.cern.ch/twiki/bin/view/Atlas/TauDPDMaker>

Tau-ID validation analyses

- Tau fake rate from dijets (very first data!)
 - tag & probe style: apply tau-ID on probe jet
 - fully documented in Tau CSC note (DESY folks: Sylvie, Philip, Sebastian, David)
 - will be done on FDR-1 data
- Tau efficiency from $W \rightarrow \tau \nu$ and $Z \rightarrow \tau \tau$
 - compare $W/Z \rightarrow \tau$ vs. $W/Z \rightarrow e/\mu$ rates
 - newer analyses, should be done on FDR-1 data
- Tau energy scale from $W \rightarrow \tau \nu$ and $Z \rightarrow \tau \tau$
 - probably waits for FDR-2...

In all cases: compare data sample results with MC predictions !

Output streams of TauDPDMaker

```
"athena FileNames1.py TauDPDMaker/TauDPDMaker.py  
athena FileNames2.py TauDPDMaker/TauDPDMaker.py"
```

```
InputCollections=["File2.pool.root"]  
OutputPoolRootFileNamePrefix="TauDPD2"
```

**input will be
AOD and ESD
in FDR-1**

TauDPDMaker

**centrally
produced
in FDR-1**

- TauDPD2_WZStream.pool.root
- TauDPD2_JetsStream.pool.root
- TauDPD2_DevelopmentStream.pool.root
- TauDPD2_ReleaseValidationStream.pool.root
- TauViewNtuple.root } **FDR-1**

Multiple streams written simultaneously in separated output files.
Each stream has its specific skimming, slimming and thinning.

Primary skim selections

□ Dijets: (back-to-back jet pair)

- jet: $p_T > 15 \text{ GeV}$, $\eta < 2.5$
- $\text{abs}(p_T[i] - p_T[j]) < p_{T\text{Max}} / 1.5$
- $2.74 < (\text{jet}[i].\text{phi}() - \text{jet}[j].\text{phi}()) < 3.54$

□ $Z \rightarrow ll$, $W \rightarrow lv$ & $Z/W \rightarrow jj$:

- $\text{MET} > 30 \text{ (10) GeV}$ for W (Z)
 - suppress dijets with low-medium p_T
- $\text{SumET} < 400 \text{ GeV}$
 - suppress dijets with high p_T (J4, J5, ...)
- $Z/W \rightarrow \text{lep}$: at least 1 (2) lepton(s) in the event for W (Z)
 - lepton = ElectronAOD, StacoMuon, TauRec or Tau1p3p
 - $p_T > 15 \text{ GeV}$, $\eta < 2.5$
 - accept $W \rightarrow e\nu/\mu\nu/\tau\nu$, $Z \rightarrow \mu\mu/ee/\tau\tau/\mu\tau/\mu e/e\tau$ final states
- $Z/W \rightarrow \text{jets}$: at least 2 Cone4H1TopoParticleJets
 - $p_T > 15 \text{ GeV}$, $\eta < 2.5$

Primary TauDPD content (slimming)

- Current DPD content is sufficient for superficial W/Z/dijets analyses...
 - But, can't do tau re-calibration with current TauDPD
 - minimal tau calibration input not yet established...
 - present event content:
 - TauContainer and TauDetails: TauRec, Tau1p3p
 - ElectronAODCollection, StacoMuonCollection
 - Cone4TopoParticleJets
 - MET_RefFinal
 - TrackParticles
 - Truth (GEN_AOD and Cone4TruthJets)
 - missing:
 - **trigger**, tau calibration input, ...

Average event size of primary TauDPD

	J1: 17-35 GeV $\sigma = 1380 \mu\text{b}$	J2: 35-70 GeV $\sigma = 93 \mu\text{b}$	J3: 70-140 GeV $\sigma = 6 \mu\text{b}$
W→lv	0.02%	0.6%	1.8%
Z→ll	1.8%	20%	41%
combined rate	1.8%	20%	41%
avg evt size	0.4 kB	3.4 kB	7.2 kB
Z→jj	9.9%	32%	43%
Dijets	17%	49%	58%
combined rate	18%	57%	76%
avg evt size	3 kB	8 kB	14 kB

Limit is 10 kB / avg_evt: acceptable average event size!¹¹

Tertiary TauDPD (ntuple)

- Due to their customized content, TauDPDs need to be read by a customized ntuple maker...
 - “TauView” is built in TauDPDMaker
 - tau-specific configuration of existing EV tools
 - Björn (DESY) extended TauView to compare the performances of several tau algorithms (see his talk today!)
 - Sylvie (DESY) is improving TauView to use it to perform the $Z \rightarrow ll$ and dijets analyses with FDR-1 data
 - Helpful for both EventView and Tau WG communities
 - “ControlSample” ntuple already in TauDPDMaker HEAD
 - Interest already manifested by people from INFN
 - For $W \rightarrow \tau \nu$, people at Krakow are migrating code from CBNT in a new TauDPD-specific package
- For release validation, AthenaRootAccess code is being developed at Simon Fraser University
 - ARA is ideal for such simple analyses

Content of TauViewNtuple.root

- “TauView” can produce three root Trees: RecTau, Truth & ControlSamples

- RecTau: all major tau-ID algorithms

- Full Tau-ID information

- still partly release-12 fashion, will be updated

- Labels for overlaps with reco/truth particles

- Truth: true generated tau, mu, electron, photon, jets

- Kinematic infos of all particles

- Extra infos for taus and their visible decay products

- Labels for matches with other reco/truth particles

- ControlSamples: reco & truth leptons, jets, MET

- Recently started by Sylvie, important for FDR-1

- Will include all necessary infos to analyze $Z \rightarrow ll$ & dijets

- Will be fully-updated to release 13 content

FDR-1

Conclusion and outlook

- ❑ TauDPDMaker will be run in central production, over AODs and ESDs, during FDR-1
 - important test of the new AMF-recommended DPD-making technology
 - DESY group appears to be a front-runner for FDR-1 activities related to DPD-making and analysis! ;-)
 - TauDPD will be usable to analyze $Z \rightarrow l\bar{l}$, $W \rightarrow l\nu$ or dijets
 - ❑ primary DPD already used successfully by people from several institutions and working groups
 - ❑ tertiary DPD being actively developed by Sylvie *et al.*
- ❑ To do before FDR-1 starts:
 - Add “JobTransform” (when available from PAT)
 - Finalize skim cuts and DPD content (optional)
 - Deadline: $\sim 1^{\text{st}}$ February



Backup slides

TauDPDMaker execution flow

"athena FileNames.py TauDPDMaker/TauDPDMaker.py"

pool.root input file (AOD, ESD, TauDPD, ...)

Primary DPD-making
is independent from
TauView, and vice versa.

TauDPD_RecExCommon_PreProcessing (optional)

- creates not already-existing containers, subsequently used "on the fly".
 - e.g.: AOD-level truth and ParticleJet if reading an ESD, custom tau reco, AODtoAOD corrections, ...
- only executes required algorithms

TauDPD_Skimming → TauDPD_Thinning → TauDPD_PoolOutput

- skimming & poor man slimming identical to AODtoDPD.py
- thinning done with private Athena-ARA-usable TauTool
- Multiple output streams written in separated root files, with different content

TauDPD_EventViewInput → TauDPD_EventViewAnalysis

- configuration of EventView tools from scratch
- makes a "TauView" ntuple
- RecTau and Truth Trees contain detailed information about the various tau selectors and their overlap with the various types of truth particles

Early-data calibration scenario

Data/MC comparison (TauDPD)
fake rate, efficiency...

large disagreement!

Fast reco \leftrightarrow root
turnaround very useful!
→requires ESD input←

new MC corrections

modify reco algorithm

change reco seed

change ID selection

tauRec: new jet (ESD)
Tau1p3p: new track (ESD)

New E calibration or detector
alignment also implies
redoing the tau reco seed

use existing variable (AOD)
create new variable (ESD)

new multi-variate analysis (root)

redo data/MC comparison