

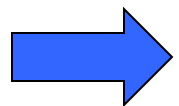
# ***PWG3 Analysis: status, experience, requests***

*Andrea Dainese  
on behalf of PWG3*

- ◆ Analysis groups in PWG3 (11 papers in progress, using 2010 data)
  - ⊕ D2H: vertexing
  - ⊕ HFE: single electrons
  - ⊕ JPSI2E: dielectrons
  - ⊕ MUON: muons and dimuons
- ◆ Analysis modes
- ◆ The special case of muon analyses
- ◆ User feedback
- ◆ Central trains

# Analyses & Their Input

- ◆ D2H (vertexing): entirely based on AOD
  - ⊕ about 15 analyses
- ◆ Electrons: mostly based on ESD; tests with AOD ongoing
  - ⊕ about 10 analyses
  - ⊕ main point to stay with ESD for a while: electron ID
    - still being understood/optmized
    - many detectors (TPC, TOF, TRD, EMCAL)
- ◆ MUON: mostly based on muon-AOD (1% of std AOD), but can use also ESD
  - ⊕ about 10 analyses
  - ⊕ some analyses on AAFs



>25 analyses on AODs + ~10 on ESDs

# ***Muon Analyses: prompt data availability issues***

- ◆ The key concern lately is about “availability”
  - ⊕ (we realize offline is not responsible for all the delays)
- ◆ The data needed for the muon analyses are accumulating very quickly since the increase of the luminosity (equivalent of full 2010 statistics every day of data taking)
- ◆ Prompt availability is crucial to keep the pace of other experiments (eg LHCb). Typical example:  $J/\psi$  polarization
- ◆ All prerequisites are there:
  - ⊕ MUON cluster calibration is ready before reconstruction
  - ⊕ MUON cluster reconstruction is faster than the whole ALICE
  - ⊕ output ESD is small
  - ⊕ muon ESD-filtering is fast and muon-AOD is  $\sim 1\%$  of std-AOD (0.1% for PbPb)

➤ eg: 18 GB for full PbPb 2010

# Example: LHC11c

(logbook : runs with MCH+MTR, >10K events, > 10 minutes ; temp. list, pending e.g. final QA checks)

S = single  $\mu$  low  $p_T$  SH = single  $\mu$  high  $p_T$  L = like-sign  $\mu\mu$  low  $p_T$  U = unlike-sign  $\mu\mu$  low  $p_T$

Laurent

X (Y) runs : X from logbook, Y = not bad in RCT

period 1 <153056			period 2 [ 153056 ; 154724 [			period 3 > 154724		
75 (58) runs			103 (46) runs			47 (8) runs (80% w/o TPC)		
	LB	RCT		LB	RCT		LB	RCT
CINT1or7	212 M	180 M $\times 1e-5$ = 2K J/ $\psi$	CINT7	51 M	39 M	CINT7	2 M	0.3 M
CMUS1or7 0.5 GeV/c	39 M	33 M $\times 2e-4$ = 7K J/ $\psi$				CMUS7 ~ 1 GeV/c	23 M	0
			CMUSH7 ~ 1.7 GeV/c	4.8 M	1.9 M	CMUSH7 ~ 4.2 GeV/c	7 M	1 M
			CMUL7 ~ 1 GeV/c	3.3 M	1.2 M	CMUL7 ~ 1 GeV/c	2 M	0.3 M
			CMUU7-X ~ 1 GeV/c	1.4 M	0.5 M $\times 2e-2$ = 10 K J/ $\psi$	CMUU7 ~ 1 GeV/c	1 M	0.1 M $\times 2e-2$ = 2K J/ $\psi$

x 3

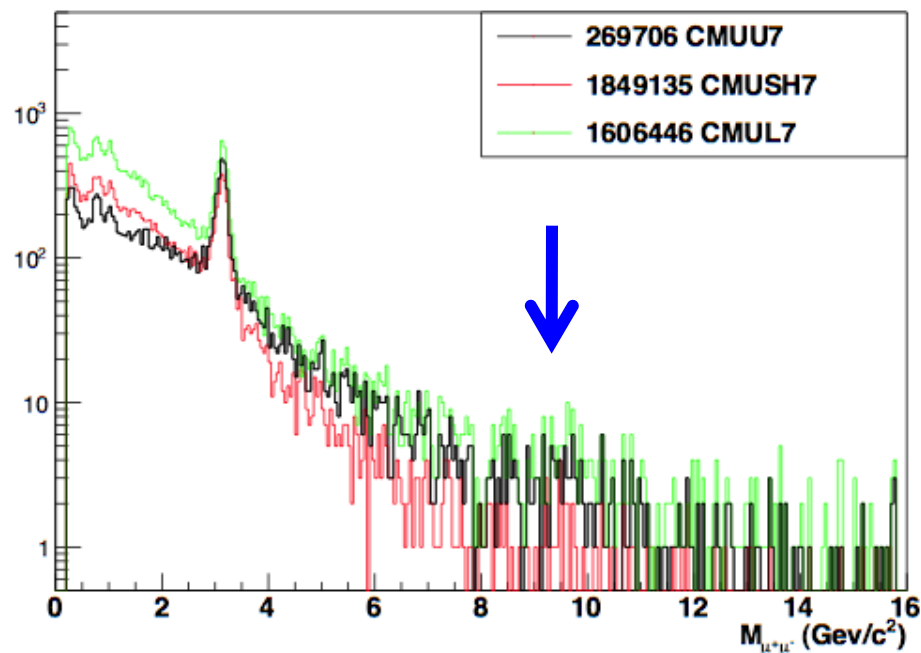
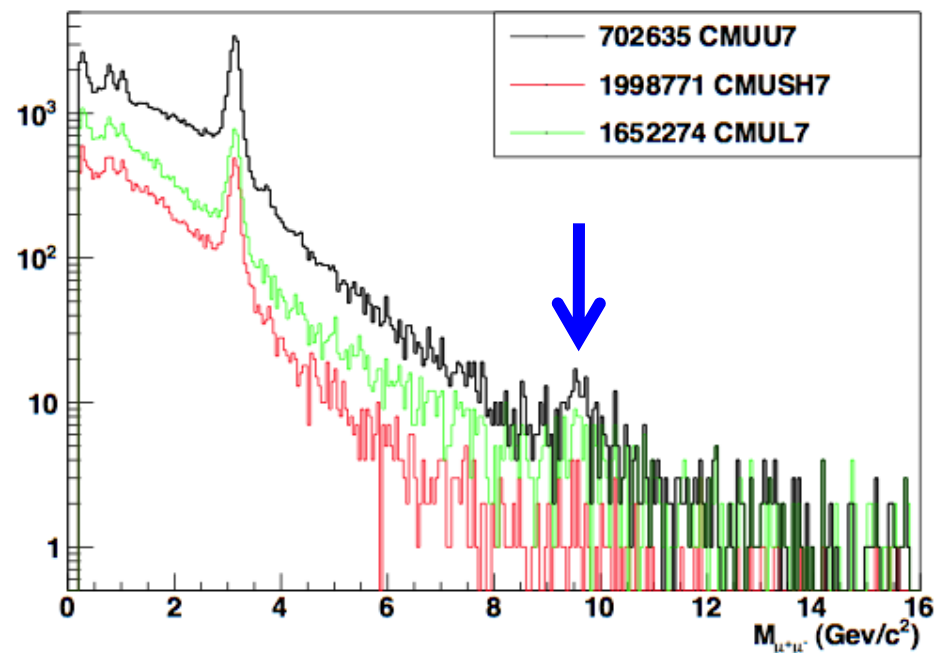
x 10

# Yet, the $Y$ is at hand ...

from ESDs

(matched+Rabs) tracks

from AODs



First quick look at LHC11c data

51 runs from 153232-154211 from pass1 ESDs & AOD055

# Muon-related requests

- ◆ All targeted to the goal of improving the availability of data to analyze

Request	Reason	Outcome
muon-cluster only reco savannah task #20946	increase ESDs availability during the current period (LHC11c at that moment)	refused once by PB
produce AODs from pass1 savannah task #20980	get the smallest possible objects to analyze	agreed (but to be redone with latest tag)
write all events in AODs (i.e. flag not physics selected one, but do not discard them)	increase useability of produce AODs	pending PB discussion ?

# User feedback

- ◆ Analysis (on AODs) on the Grid mainly as single user jobs
- ◆ Two different “modes”:
  - ⊕ pp analysis: strategy and cuts established; run when new sets are available; few iterations
    - could go to a central train
  - ⊕ PbPb analysis: still experimenting/exploring (tune cuts and PID strategies); chaotic analysis; continuous changes in the tasks code; frequent iterations
    - more difficult to go to a central train
- ◆ Users lately very happy with Grid performance. Few problems:
  - ⊕ merging: sometimes very painful
  - ⊕ limited quota: limited #subjobs (problematic in PbPb, where long CPU times require finer splitting)



- ◆ Central train macro kindly prepared and tested by Mihaela
  - ⊕ committed to PWG3/centraltrain
- ◆ Can be configured for AOD or ESD, pp or PbPb
- ◆ Contains ~10-15 tasks that run on AODs + 2 tasks on ESDs
- ◆ Natural splitting in 3 trains:
  1. ESD (electrons)
  2. AOD min. bias (vertexing)
  3. AOD muon (much smaller set of interesting events in muon-AOD, different runs,...)
- ◆ We started to identify and train a team of “conductors”
  - ⊕ first cluster from D2H, coordinated by Zaida

- ◆ 7 analysis tasks (2 x D0, D+, D\*, Ds, Lc, CF)
- ◆ The configuration is already quite complex:
  - ⊕ 2 systems (pp, PbPb) x 2 data type (real, MC)
  - ⊕ data sample (one train = one period, for now): would be useful to be able to run same train on few periods (eg LHC10b,c,d,e) and merge the output per period a posteriori
  - ⊕ task configuration: the D2H tasks are configured using a “cuts” object
    - at the moment, single users add their task many times in the same job to study different cuts and analysis modes
    - the central train should provide this “functionality”: each task attached few times with different config (passed to the AddTask, maybe via a path to a root file on alien); the OADB doesn’t seem to be a good option to store task config and analysis cuts
- ◆ For MC, we run the CF (THnSparse): also many instances per analysis → issue with memory for merging

## ◆ Frequency of train runs:

- ⊕ in order to be useful for the users, the trains should run quite frequently (weekly?), possibly with a subset of the tasks
- ⊕ testing and configuring should become semi-automatic / fast
- ⊕ however, there are also several datasets being analyzed in parallel...
  - eg. in D2H, currently: LHC10bcde, LHC11a, LHC10h, + 3 MCs  
→ 9 datasets

## ◆ First experience in train setup:

- ⊕ run in test-mode locally to optimize the splitting (n. input files)
  - recommendation to stay < 5h, however in test-mode the time is dominated by file access ... not really the time is takes on the WNs
- ⊕ check the memory profiles of all tasks → OK for pp and Pb-Pb