# Stripping requirements FWG

Raluca Mureşan

EPFL

on behalf of FWG

# Outline

- Very first data (VFD)
  - *all inclusive distributions*
  - *strangeness and charm production*
  - *MC tuning and validation*
- $J/\psi$ and charmonia
- EW physics
- Exotica

I apologize in advance if I forgot to include something

# VFD - mandate

To produce physics results starting from the

VERY FIRST $10^8$ events written on the tape with:

- Stable colliding beams with $\sqrt{s} >$ 4TeV and stable detector
- Random (truly unbiased) triggers, or for the VVFD ($\nu \ll 1$) 10% random triggers + MB triggers: SPD multiplicity, HCAL (cut on the sum over highest $E_T$ 2$\times$2 clusters in all HCAL, cut on the largest $E_T$ hadron), ECAL.

# VFD - mandate

Minimal requirements:

- We propose to use only the information provided by the tracking devices; the tracking devices need to be time- and space-aligned. Same data will be used for PID calibration using very clean samples of $D^*$, $\Lambda$ and $K_S$.
- We aim to publish early (before we see large J/psi signals) and we may have to content ourselves with reasonably looking mass peaks for $V^0$ s and charm mesons.
- Many of the distributions of interest are essential for the MC tuning that might be done in collaboration with the other LHC groups $\Longrightarrow$ essential to be as unbiased as possible.

*We plan to use data recorded before the detector is fully calibrated/tuned*

*Our cherish wish is to have access to all the $10^8$ events - only 1% of the data volume we are expecting for a nominal year.*

$10^8$ MB events full DST = 3 TB $\sim$ 300 CHF

# VFD - $10^7$ events unstripped

If it is not possible to access $10^8$ events ☹we would like to access in full at least $10^7$ events:

- OPG report Sept 24: “Maximum size for analysis is typically $10^7$ events”
- Analysis of $10^7$ events MC09 was already performed. Typical running time on the Grid, including waiting was less than 1 day and root files managable as dimensions - even if ntuples were very “heavy” in MC information over 75% MC truth for efficiency studies and purity estimation.
- Estimated yield about 400 of each of the charmed mesons, about 50000 $\Lambda$ (double differential distributions).

# VFD - $10^7$ events unstripped

If it is not possible to access $10^8$ events ☹we would like to access in full at least $10^7$ events:

For some of the all inclusive distributions e.g $\frac{d^2n}{dp_t d\eta}$:

- interesting in their own as one CDF article was *"$p_T$ distribution of charged particles produced in $p\bar{p}$ interactions at $\sqrt{(s)}$ = 630 and 1800 GeV"*, PRL 61, 1819 (1988);
- Can give the key to "New (QCD) physics" see Rick Field *Introduction to UE and MB studies*;
- essential for MC tuning/validation.

Might be even enough for allowing to get some (publishable) results, although for the high $p_t$ region we might not have enough statistics.

# VFD - $10^7$ events unstripped

- LHCb VFD silver bullet for tuning/validating the MC generators at LHC due to the unique rapidity range.

  see Rick Field *Introduction to UE and MB studies* or Peter Skands *"Constraining the underlying event with strangeness and forward baryons"*

  The phenomenological models are extrapolated not only in energy, but also in rapidity for the LHCb kinematical range $\Longrightarrow$ diverge more

- Very useful input to the theorists providing that we understand our data and detector $\Longrightarrow$ we need access to as much information as possible for a minimum of $10^7$ events.

# VFD - stripping scenario

- If it is not possible to access $10^8$ events ☹we would like to access in full at least $10^7$ events
- Since we plan to use only the tracking information we need to keep for all the other events information on:
  - long tracks - first strangeness and charm production studies
  - all the other tracks with VELO segments for charge particle production studies and to veto the diffractive events

# VFD - stripping scenario

Keep all the high level tracking information

Two technical solutions:

| | |
|---|---|
| MicroDST | ☺standard LHCb procedure |
| | ☹not investigated yet (11 kB/event - 2.64 pRec/ProtoP/Charged, 8.18, pRec/Track/Best, 0.08 pRec/Vertex/Primary, 0.03 DAQ/ODIN, 0.07 Rec/Status and Rec/Header ) |
| zooNtuple | ☺Already investigated by people in Heidelberg |
| | ☹Need new code for vertexing and IP calculation. |
| | ☺2 kB/event |

Done as many times as necessary as our requirements might evolve with time.

Open question on how the MB analysis will be thought once we start the data taking in the nominal running conditions.

# $J/\psi$ and charmonia

- Save in full all the events containing a loosely selected $J/\psi$ candidate together with the $J/\psi$ candidates;
- $J/\psi$ stripping selection, can be optimized with the first $10^7$ events;
- F WG will use the $J/\psi$ stripped samples for all the charmonia studies, as well as for the first B physics (first mass $b - J/\psi X$ peaks and lifetimes);
- Similar (microDST) samples for $\psi(2S) \to \mu\mu$ and $Y(1S, 2S, 3S) \to \mu\mu$.

**Beginner's view on MicroDst with 5000 inclusive J/ψ sample(2080 selected)**

**File size for selected events:**

**MicroDST: 12.8MB**
**DST: 458MB (459M with J/ψ information written)**
**Tuple.root: 1.02MB (nearly all the info needed)**

**Bigger than ROOT file but Much smaller than DST file**

**For the time spent some "ratio" with 2.8GHz Xeon machine**

**Time spent in writing selected events:**
**From DST-->MicroDST: 8.45 Min ratio =2.06**
**From DST--->DST : 8.99 Min ratio = 2.07**

**Time spent for analysis on DST and MicroDST:**
**MicroDST to ROOT : ?? not known**
**Selected DST(2080 events) to ROOT : 4.53 Min from DST-->ROOT ratio = 2.05**
**Full DST(5000 events) to ROOT : 10.7 Min from DST-->ROOT ratio = 1.84**
**(Time used on inclusive J/Ψ sample, much more if on Mini-Bias sample)**

**For Analysis, Compare with DST**

**1. Not easy to switch for those who are used to C++ and ROOT for analysis (like me).**
**: Need an easy way to transfer the old C++ programe from DST to MicroDST**
**: More examples and supports**

**2. Give us a standard form to store information.**

**3. With more information needed, no need to re-run DaVinci again for small changes.**

**For storing data, comparison with DST**
**1. Much smaller**

**2. Reduce the frequency of accessing to DST, faster in Analysis**

**My opinion is:**

**1. Offering a similar way of C++ analysis to MicroDST so as to DST.**

**2. During our usual analysis, we access to MicroDST, only when necessary, we go to DST.**

# EW physics and Exotica

- For the first year all the info on the DST (including all trigger information) for the muon HLT2 triggered events (ie. single muon with and without IP cut, di-muon trigger)
- Later on, might want to move to two streams (muon triggers with IP, and muon triggers without IP, or divide the no IP streams further into high and low $p_t$ threshold subsets).
- No selection currently implemented for EW physics;
- Some selections are already in place (Neal, Marcin) for Exotica;
- Any guidance / templates / instruction would be appreciated.

# Summary

- Our cherish wish for the VFD analysis is to have access to all the $10^8$ events in full - only 1% of the data volume we are expecting for a nominal year
- We need at least $10^7$ unstripped events for VFD analysis and tuning the selections
- VFD stripping stream keeping only the high level tracking information for VFD analysis
- All the event information + the $J/\psi$ candidates for events containing a very loose selected $J/\psi$ for the $J/\psi$ production studies and early b-physics studies
- All the event information for events passing the IP muon trigger and dimuon trigers for the EW physics group and Exotica + few specific selections for the Exotica.