CMS winter school 2021
December 18-20
Peking University, China

Short exercise on muon studies

Measuring the muon selection efficiencies using Tag and Probe Method

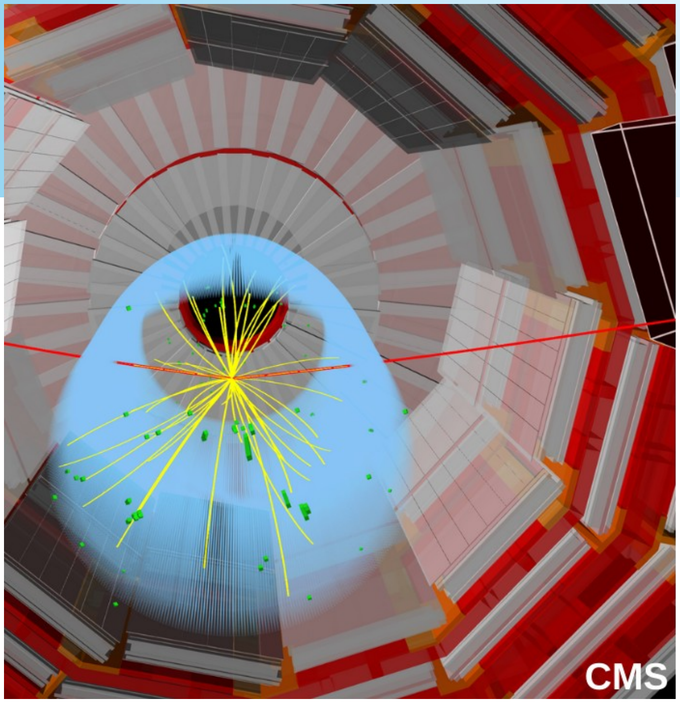
Tahir Javaid (tahir.javaid@cern.ch)
Beihang University, China

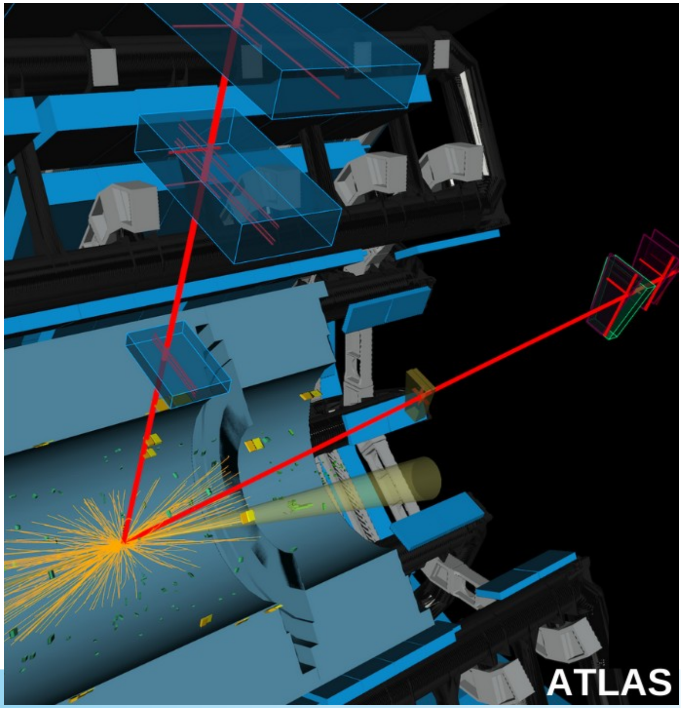
Outline

- **Introduction** to Muons
- **Motivation**: Why we measure muon efficiencies ?
- Tag and Probe **method**
- Muon Identification efficiency **measurement**

Muon

- Charged particle, just like electron and positron, but is 200 times heavier
- Produced in the decay of a number of potential new particles e.g. Higgs (H → 2mu)
  - Key object for various important measurements e.g. Higgs production cross section, mass, decay width and etc.
- Can penetrate several metres of iron without interacting, unlike most of particles
  - dedicated muon detection system in LHC experiments e.g. CMS, ATLAS, ALICE etc.
- In the detector, a muon is **identified** from its track information in the tracker and muon system alongwith its kinematics.





Higgs decay in 2 muons (2020) link

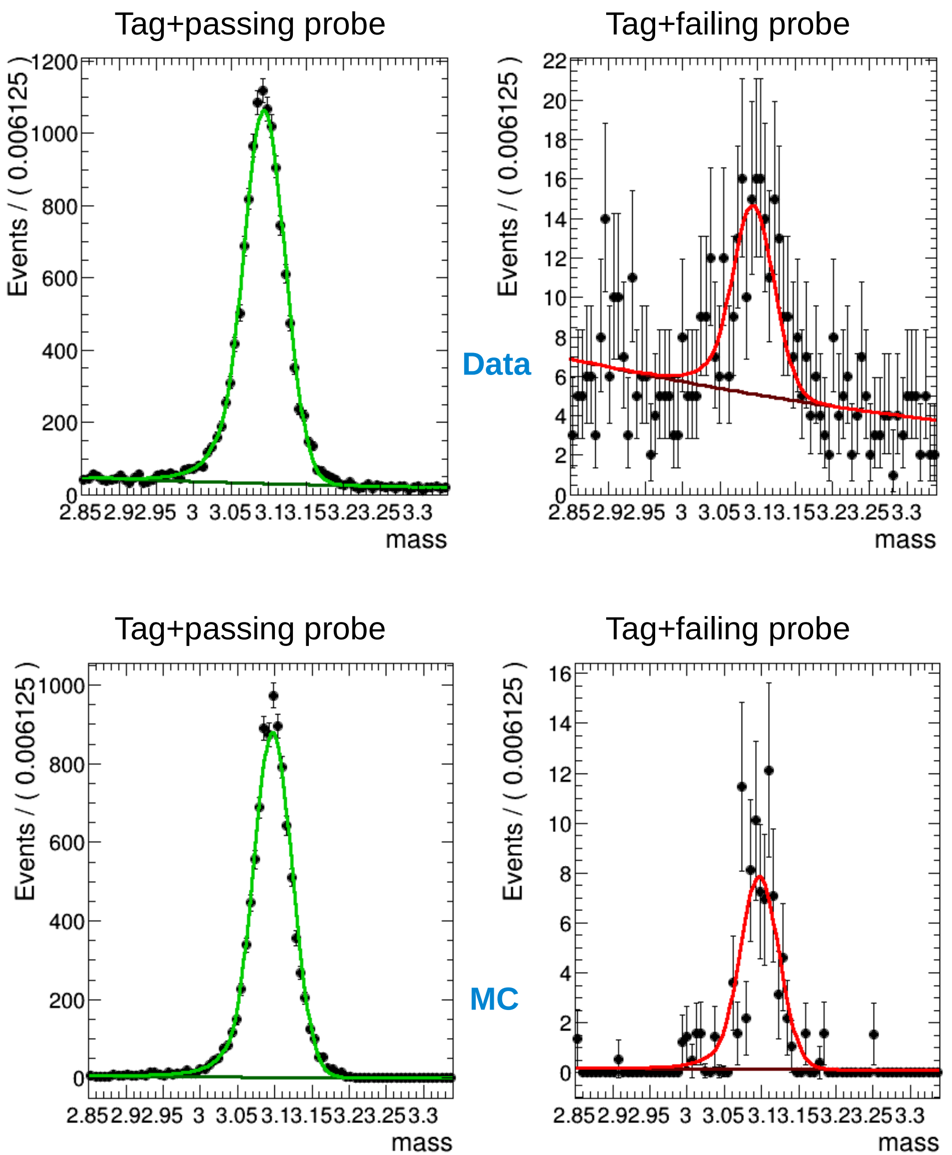
Please look for:
Other decay mode(s) of Higgs where there is muon in final state.

Motivation for muon Efficiency measurement

- MC does not describe the real data well
- In the measurement, MC is corrected with correction factors
- Done by measuring the selection efficiencies
- Correction factors are applied to MC
  - Correction factor or Scale factor = $\frac{Selection\ efficiency\ data}{Selection\ efficiency\ MC}$
- After applying the correction factors, MC is supposed to match the data
  - Introduces **additional** uncertainty source to the measurement

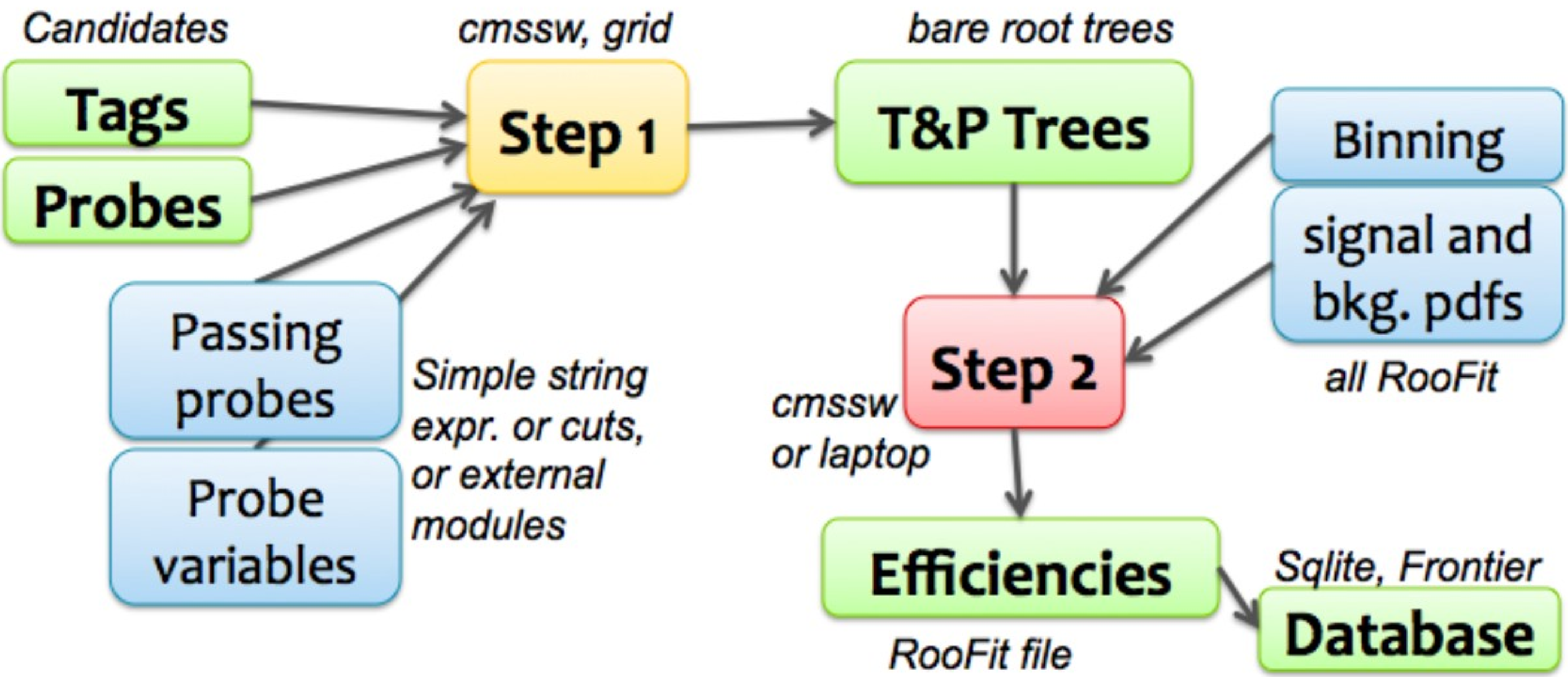
Tag and Probe method

- **Definitions:**
  - **Tag muons** are usually **good quality** muons matched to a dedicated muon Trigger
  - **Probe muons** are inclusive calo. muons or just tracks in the tracker or the muon system
- **Methodology**
  - Processed data and MC samples of di-muon resonances e.g. Z, J/Psi, Y **(root files)**
    - Possess all possible information e.g. muon kinematics **(root branches)**
    - Definition of Tag and probe muons
    - Compare the muon from the probe muon pool with tag muon
      - Grab passing and failing probes
  - Construct mass distribution of tag+passing probe and tag+failing probes
  - Fit the distributions using suitable polynomials **(RooFit)** in bins of **pt or eta**
  - Compute the integral from fit in each bin
  - Efficiency is defined by:
    - € = (probes passing the selections) / all probes



Do you know?
Why do not we consider simple cut and count method to compute efficiencies and instead we fit the distributions ?

Overall Workflow



Hands-on Exercises

- Dump object information from CMS official dataset
  - Get object info from CMS 2017 Jpsi datasets
- Prepare the tag & probe root files
  - Select preliminary tag & probe pair.
- Exercise1
  - Measure muon Loose ID efficiency **(Today)**
- Exercise 2
  - Measure muon Medium ID efficiency **(Homework)**
- Exercise 3
  - Add more systematic sources in Exercise 1 and Exercise 2 **(Homework)**

Link to muon IDs definitions

Please look for:
Other efficiency types we measure using Tag and probe method

Hands-on Exercises : Setting framework and running

1 ssh -XY username@hepfarm02.phy.pku.edu.cn -p 9002

2 mkdir muon; cd muon

3 source /cvmfs/cms.cern.ch/cmsset_default.sh # repeat with every login

4 cmsrel CMSSW_10_2_5

5 cd CMSSW_10_2_5/src

6 cmsenv # repeat with every login

7 cp -r /data/pubfs/pku_visitor/tahirjavaid/pek_school/material/TnP-scripts .

8 cp -r /data/pubfs/pku_visitor/tahirjavaid/pek_school/material/HiggsAnalysis .

9 cp -r /data/pubfs/pku_visitor/tahirjavaid/pek_school/material/samples .

10 scramv1 b # setup complete! Now ready to do measurement

11 cd TnP-scripts # navigates to the directory containing the scripts / macros

12 sh runTnP_MuPOG_jpsi.sh # performs data and MC fits and computes efficiencies

- **Please note:** Output of this step is the input to the next step.

13 sh harvestTnP_MuPOG_ID_2017.sh # adds systematic sources and computes efficiency ratios

Please note, “#” is to add comments, do not include it and the text after it in command line

Exercise 1: Measuring Loose ID efficiency (2017)

- **Data :** 2017 Run BCDEF re-reco (twiki page for details)
- **MC:** eos/cms/store/group/phys_muon/TagAndProbe/Run2017/94X/JPsi/**MC**/TnPTreeJPsi_94X_JpsiToMuMu_Pythia8.root
- **Tag Selection:** tag_pt>8 && abs(tag_eta)<=2.4 && tag_Tight2012 && tag_Mu7p5_Track2_Jpsi_MU
- **Probe Selection:** general tracks **(all probes)**, muonPOG Loose ID **(passing probes)**
  - Loose ID efficiency = $\frac{passing\ probe\ muons}{all\ probe\ muons\ i.e.\ passing\ probes + failing\ probes}$
- **Systematic variations**

  1) Choice of PDFs: Default = JDGauss, expo Alternate= JCB, bern5

  2) Number for mass bins: Default = 40, Extended = 45 , Reduced = 35

  3) Mass range: Default = 2.9-3.3, Extended = 2.95-3.25 , Reduced = 2.85-3.35

***Do you know** what information about MC sample can be extracted from its root file name?*

TnPTreeJPsi_94X_JpsiToMuMu_Pythia8.root

Looking at macro: runTnP_MuPOG_jpsi.sh

- Exploit the core **tnpEfficiency.py** script which helps to fit the dimuon distributions of tag+passing probe and tag+failing probe, and compute the efficiencies in data and MC
- Provides it with several **inputs** e.g. data and MC samples, tag and probe muon definitions, bins of pT, signal and background PDFs and etc.

```bash
#!/bin/bash

PDIR="test/"
JOB="mupog_RecoId"                    ## FULL DATA & MC

XBINS="[3,4,5,6,7,8,10,12,15,20]"

EBINS="[-2.4,-2.1,-1.6,-1.2,-0.9,-0.6,-0.3,-0.2,0.2,0.3,0.6,0.9,1.2,1.6,2.1,2.4]"
VBINS="[0.5,5.5,6.5,7.5,8.5,9.5,10.5,11.5,12.5,13.5,14.5,15.5,17.5,20.5]"
MC='../samples/TnPTreeJPsi_94X_JpsiToMuMu_Pythia8_skimmed_weightAdded.root'
DATA='../samples/TnPTreeJPsi_17Nov2017_Charmonium_Run2017B2F_skimmed.root'
echo "$DATA"
echo "$MC"
PDS="$DATA --refmc $MC"
OPTS=" --doRatio  --pdir $PDIR/$JOB -j 5 --mcw weight"
OPTS="$OPTS -t tpTree/fitter_tree  --mc-cut 1 --mc-mass mass   "
if [[ "$1" != "" ]]; then SEL=$1; OPTS="$OPTS --reqname $1 "; shift; fi
if [[ "$1" != "" ]]; then OPTS="$OPTS $* "; shift; fi
MASS="  -m mass 80,2.85,3.34"
CDEN=" tag_Mu7p5_Track2_Jpsi_MU && pair_drM1 > 0.5 "
#for ID in Loose Medium Tight2012 Reco LooseIdOnly; do
for ID in Loose ; do
  if [[ "$SEL" != "" ]] && echo $SEL | grep -q -v $ID; then continue; fi
  NUM="$ID"
  if [[ "$ID" == "Reco" ]]; then   NUM="(Glb || TM)"; fi
  if [[ "$ID" == "LooseIdOnly" ]]; then NUM="Loose"; CDEN="$CDEN && (Glb || TM)"; fi
echo $NUM
echo $CDEN
  for BMOD in expo bern3; do  # other alternate models are bern4, bern5, bern6, bern7, etc....
    if [[ "$SEL" != "" ]] && echo $SEL | grep -q "_" && echo $SEL | grep -q -v $BMOD; then continue; fi
    for SMOD in  JDGauss JCB; do  # other alternate model is JGauss
        if [[ "$SEL" != "" ]] && echo $SEL | grep -q "_" && echo $SEL | grep -q -v $BMOD; then continue; fi
        DEN="$CDEN"; POST=""
        python tnpEfficiency.py $PDS -d "abs(eta)<1.2 && $DEN" -n "$NUM" $OPTS --x-var pt $XBINS -N mu_${SMOD}_${BMOD}${POST}_${ID}_barrel -b $BMOD -s $SMOD $MASS --xtitle "p_{T} (GeV)" ;
        python tnpEfficiency.py $PDS -d "abs(eta)>1.2 && $DEN" -n "$NUM" $OPTS --x-var pt $XBINS -N mu_${SMOD}_${BMOD}${POST}_${ID}_endcap -b $BMOD -s $SMOD $MASS --xtitle "p_{T} (GeV)";
#         python tnpEfficiency.py $PDS -d "pt > 7 && $DEN" -n "$NUM" $OPTS --x-var eta $EBINS -N mu_${SMOD}_${BMOD}${POST}_${ID}_pt7   -b $BMOD -s $SMOD $MASS  --xtitle "#eta";
#         python tnpEfficiency.py $PDS -d "pt > 7 && $DEN" -n "$NUM" $OPTS --x-var tag_nVertices $VBINS -N mu_${SMOD}_${BMOD}${POST}_${ID}_pt7_vtx   -b $BMOD -s $SMOD $MASS --xtitle "N(ver
```

How will you *modify the macro to measure Medium ID efficiency ?*

Looking at macro: harvestTnP_MuPOG_ID_2017.sh

- Exploit the core **tnpHarvest.py** script which is capable to add systematics to bin by bin data and MC efficiencies and computes their ratio with overall uncertainty
- Provides it with **inputs** of systematic sources

```
#!/bin/bash
P="test"
IN="mupog_RecoId"
#MEAS="mu_Loose mu_Medium mu_Tight2012"                         ##can be the type measurement  ............................OK
MEAS="mu_Loose"                          ##can be the type measurement  ............................OK
for sig in JDGauss JCB; do
    for bkg in bern3 expo; do
        for salt in JDGauss JCB; do
            if [[ "$salt" != "$sig" ]]; then
            for balt in bern3 expo ; do
                if [[ "$balt" != "$bkg" ]]; then
                if [[ "$1" != "" ]]; then MEAS="$*"; fi
                for M in $MEAS; do
                    case $M in
                        mu_Loose) MODS=" -s "${sig}" -b "${bkg}"  --balt "${balt}" --salt "${salt}"  ";   #
#                        mu_Medium) MODS=" -s "${sig}" -b "${bkg}"  --balt "${balt}" --salt "${salt}"  ";   #
#                        mu_Tight2012) MODS=" -s "${sig}" -b "${bkg}"  --balt "${balt}" --salt "${salt}"  ";   #
#                        mu_Loose) MODS=" -s "${sig}" -b "${bkg}"  --balt "${balt}" --salt "${salt}" --alt massExtended --alt massReduced --alt binsExte
nded --alt binsReduced ";   #
                        OUT="$IN/${M}_2017_harvest_${sig}_${bkg}_${salt}_${balt}_mupogSysts"
                        TIT='Muon Id efficiency' ;;
                    esac;
                    OPTS=" --doRatio --pdir ${P}/$OUT --idir ${P}/$IN  --rrange 0.97 1.01  --yrange 0.9 1.01 "; XTIT="p_{T} (GeV)"
                    for BE in barrel endcap; do
                        python tnpHarvest.py -N ${M}_${BE} $OPTS $MODS --ytitle "$TIT" --xtit "$XTIT"
                    done
#                    python tnpHarvest.py -N ${M}_pt7 $OPTS $MODS --ytitle "$TIT" --xtit "#eta"
#                    python tnpHarvest.py -N ${M}_pt7_vtx $OPTS $MODS --ytitle "$TIT" --xtit "N(vertices)"
                done
                fi;
done;
fi;
done;
done
done

"harvestTnP_MuPOG_ID_2017.sh" 35L, 1886C                                                                      1,1           All
```

Output (fits of distributions and efficiency plots) is stored in **"test/mupog_RecoId/"** directory


Thank you

Muon Reconstruction and Identification at CMS

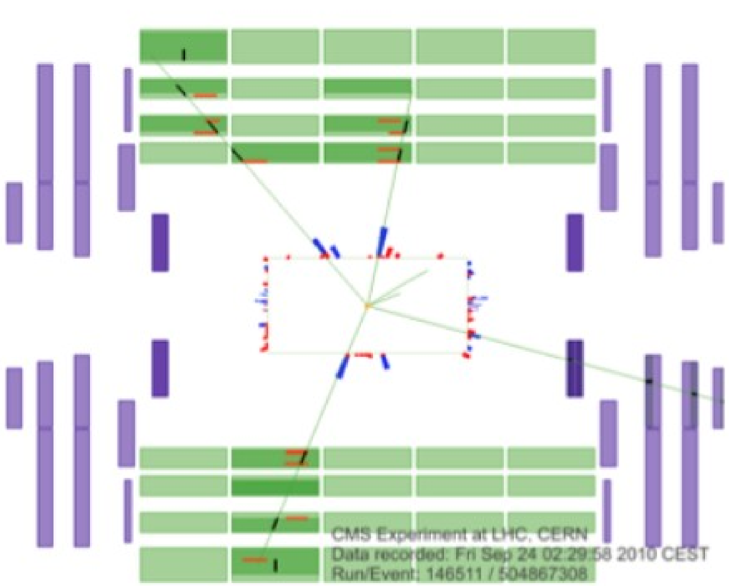
- Independent reconstruction: tracker track and standalone muon track
  - Global muon refitted using hits from two trackers using Kalman filter techniques
  - Tracker Muon: tracker track with at least one matching segment
  - Dedicated algorithms for high-pT muons (TPFMS,Picky)
- Robust and efficient muon reconstruction
  - 99% efficiency within acceptance
  - Candidates with same inner track merged into one collection

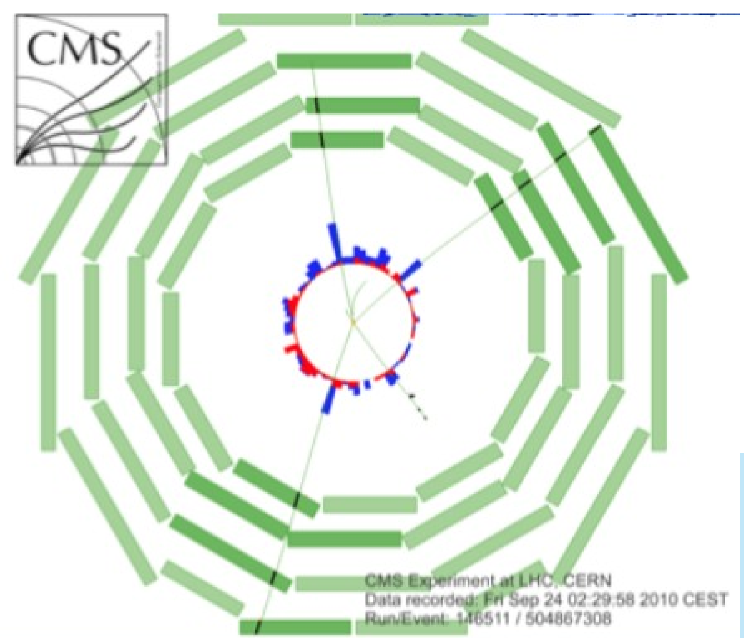
Identification

- Tight: global muon reconstruction + quality criteria on hits, segments and impact parameters. Used for most analysis like W/Z or Higgs.
- Soft: tracker muon matched with muon segment not used for other muon tracks, dedicated for muons with pT<10 GeV.

…………………………..

…………………………..

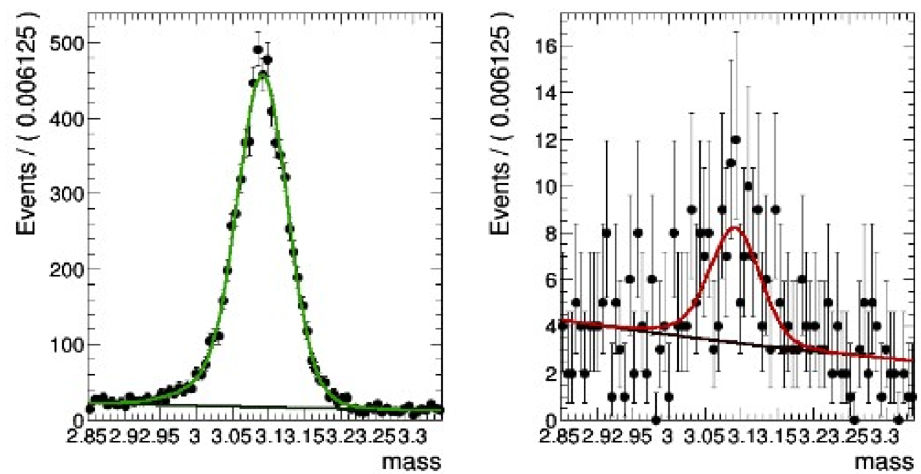




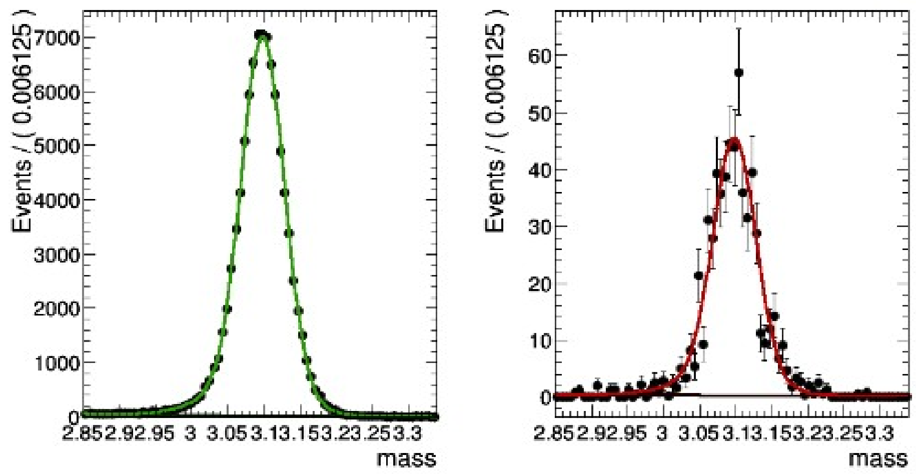
How final efficiency plots look like

Passing probes Data | **Failing probes Data** | **Passing probes MC** | **Failing probes MC**

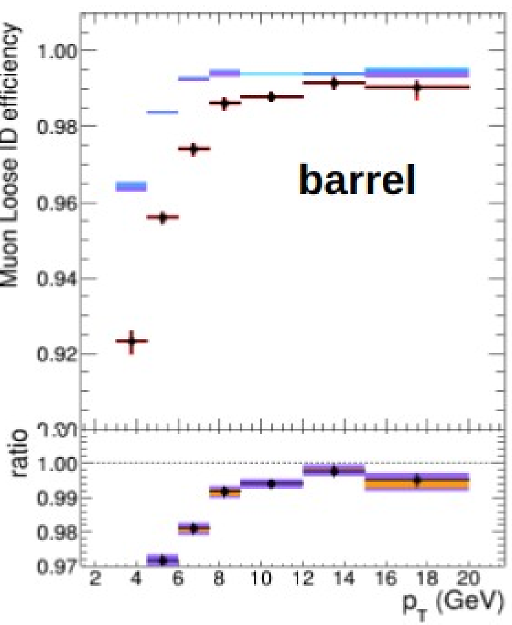
2017

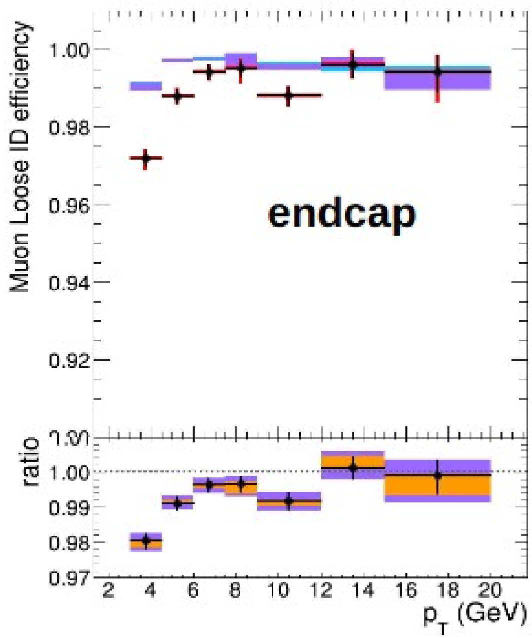


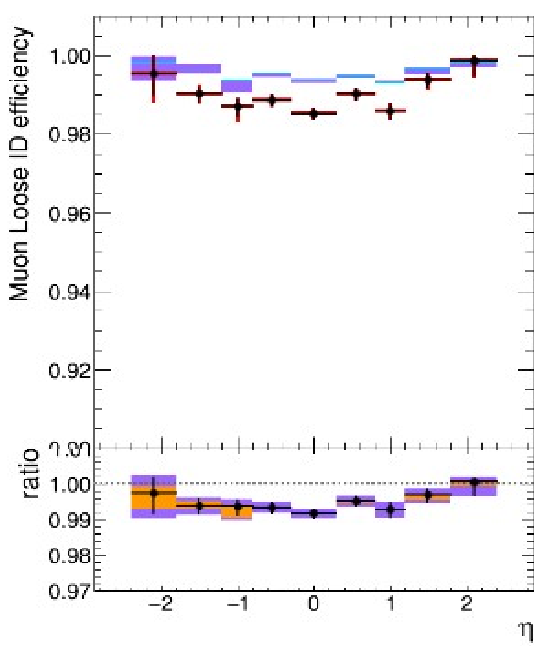
mu_JDGauss_expo_Loose_pt7_bin3



mu_JDGauss_expo_Loose_pt7_bin3_ref

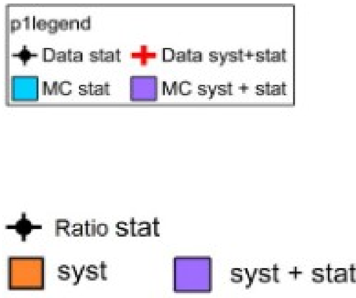






Fit Plots Link

Loose ID



pT Bins=[3,4.5,6,7.5,9,12,15,20]

eta bins= [-2.4,-1.8,-1.2,-0.8,-0.3,0.3,0.8,1.2,1.8,2.4]