Simulation

Ajeeta Khatiwada

August 2019

1 Embed Simulation

For the pp analysis, the real data used as a background in the embed simulation can be found in the following area:
$/gpfs02/phenix/fvtx/subsys/fvtx/shlim/DST/Run15pp200\_FVTX/DST$
There are total of 2760 DST root files corresponding to 276 runs. Prompt J/ψ, $b\bar{b}$ and $c\bar{c}$ embed samples are generated using the same run list (realdataBG-run15pp_file.list) and hadron samples are generated using the same runs but split into smaller files with 300 events each per file (realdataBG-run15pp_file_split.list). When using unsplit files for hadron embedding simulation, some memory leakage issue was seen, whose origin could not be understood. Therefore, to get around the memory issue, we produce smaller root files for hadron samples that are later combined into larger files during reconstruction. The scripts and configuration files corresponding to the production of the simulation files can be found in the path provided below:
$/gpfs/mnt/gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/Simulation$

The simulation can be run by using the following command:
csh event_gen.csh <line_number_in_the_data_run_list>
Environment variable LD_LIBRARY_PATH is set at the top of the event_gen.csh such that the following libraries are available:

libPHHepMCNode.so	libmuon_util.so	libmutoo_classes.so
libSubsysReco_muons.so	libmutoo_alignment.so	libmutoo_core.so
libmuon_subsysreco.so	libmutoo_base.so	libmutoo_display.so
libpicodst_object.so.	libPHG3toG4.so	libg4root.so
libBaseVGM.so	libClhepVGM.so	libGeant4GM.so
libXmlVGM.so	libmutoo_interface.so	libmutoo_modules.so
libmutoo_subsysreco.so	libgeant4vmc.so	libRootGM.so

The above library files are currently produced from a local build using the following procedure:

- Checkout a relevant package from the cvs as below:
  cvs co offline/AnalysisTrain/picoDST_object

- Make necessary changes to the existing scripts.
- If want to make changes to the cvs then commit to cvs as below:
cvs commit -m "Message explaining what the commit includes" files_modified.cc files_modified.h
- Authenticate and get AFS token if prompted:
klog
- Enter kerberos password when prompted.
- Make install and build directories:
mkdir /gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/install
mkdir /gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/build/picoDST
- Go to build directory and configure:
cd /gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/build/picoDST
/direct/phenix+u/workarea/ajeeta/offline/AnalysisTrain/picoDST_object/autogen.sh
–prefix=/gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/install
- Make:
make
- Copy the build programs, libraries and documentations to the correct location:
make install
- Load and use the libraries in macros and shell scripts:
setenv LD_LIBRARY_PATH "/gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/install/lib"

Embedding is done in three steps by running the following three C scripts in root:

1.1 g3tog4_samplename.C

In this step Pythia events are generated using PYTHIA8 using the following configuration for all samples:

- p-p beam
- CM energy: 200 GeV
- 7 ! CTEQ6L PDF

For hadron, J/ψ and $b\bar{b}/c\bar{c}$ events HardQCD:all, Charmonium:all, and HardQCD:hardbbbar flags are turned on respectively. The configuration files are in .cfg format located inside of the pythia_configuration folder.

The events are triggered based on the particle and its parent's pdg ID. Loose cuts on the minimum value of the momentum and range of the pseudo-rapidity is applied. The particles generated by Pythia are passed through GEANT4 and a GEANT based trigger is also applied to the event. This is done using virtual geant PHG3toG4 class that requires root files for detector geometry and magnetic field information. The files are named geom.root and Sim3D++ root respectively and are located in the Simulation directory.

Only the events that pass through our detector system and create a hit in one of the detectors are recorded. The vertex of the event originally produced by Pythia at (0,0,0) is shifted according to the primary vertex determined by FVTX based on the DST file provided for the real data. Basic information from rest of the events are saved in a normalization file. For the events that pass through the GEANT, detailed information of the HepMC interfaced Pythia events is recorded in a MCHepMCParticleContainer. The hits for the particles that pass through our detectors are recorded in PISA (PHENIX Integrated Simulation Application) files. In this step, the PISA files are first created inside PISA/<line_number_in_the_data_run_list>, and needs to be moved to a separate location so that consecutive jobs do not overwrite the file created previously.

1.2 Fun4FVTX_Pisa.C

Detector response is implemented at this stage of simulation. Ideal FVTX geometry is considered for simulation and alignment corrected geometry for the data. The FVTX geometry is applied using a local file fvtx_geometry.root. The dead channels for different sub-detector systems, including the FVTX, are obtained from their respective database and applied accordingly corresponding to the data run numbers set in the simulation. All the required inputs are available locally in the Simulation directory. The geometry for SVX detector is applied based on a local file svxPISA.par.

The primary vertex is smeared according to the event-by-event FVTX+SVX error calculated using embedded data sample. The hit information from different sub-detector systems are saved as DST files.

For hadron samples, there is an additonal step where the contents of pythia output and DST output are combined together to a single file such that a list of these combined files can be provided as input in the reconstruction stage. This process is done such that single root file contains more than 300 events, which is necessary in order to produce meaningful results when applying reassociation using mFvtxMuTrAssociation module during reconstruction.

1.3 Fun4FVTX_RecoDST_sim.C

Event reconstruction takes place at this stage. Hits from data background are combined with the simulated signal hits into a single container that is used for track reconstruction. Track reconstruction is performed on the simulated hits with the muon tracker (MuTr) and FVTX separately. The tracks in FVTX and MuTr are matched and refit. The information is filled in SingleMuonContainer. Furthermore, MCSingleMuon module associates the HepMC signal information with the Monte Carlo hits and tracks, resulting in a container with information on tracks in the muon arm. In the next step, reconstructed tracks are synchronized with the MC tracks from the MCSingleMuon container. The final ntuples that are used in the analysis are saved as picoDST files.

During this process, the muon id efficiencies are read separately for north and south arm using the following local files:
muid_tube_eff_north_Run15pp200.txt and, muid_tube_eff_south_Run15pp200.txt. The dead channel map and geometry files are also obtained locally as in the previous step.

Background related to wrongly associated FvTX and MuTr tracks is also estimated at this stage using mFvtxMuTrAssociation class. The class performs matching between tracks in the different arms of FVTX and MuTr subsystem, and the resulting events are flagged using set of variables in the SingleMuon container. It was observed that applying this reassociation module on a DST file with small number of events resulted in wrong estimate for the misassociation background due to reduced probability of finding FVTX tracks that can be wrongly associated with the MuTr tracks. Therefore, caution has to be taken when running reconstruction on a small number of events with an intention of combining multiple picoDST files together to produce one large file. Furthermore, it was also observed that applying this module at a filtering stage as described in Section 2 resulted in artificial broadening of the DCA_R distributions for the simulations, whose cause has not been understood until the time of writing this note.

The picoDST files produced from above steps are ready for analysis. However, since the files can be large and contain many events outside of the desired phase space. Therefore, we apply filter the picoDST files using some loose cuts that are also applied to data picoDST files, which is detailed in the Section 2.

The simulation procedure, particularly for hadron sample, is a time consuming and computationally taxing process, and is ongoing at the time of writing this draft of the analysis note. Therefore, the number of events produced for the simulation is an ever changing value. There are approximately 9M reconstructed events in hadron simulation, 13M J/ψ events, 7M $b\bar{b}$ events, and 4M $c\bar{c}$ events after the filtering cuts are applied but before requiring any generator level cuts.

2 Filtering

Loose cuts are applied to the picoDST files prior to using them in the analysis such that the files sizes are reduced. This process of filtering is done by using mFilterSingleMuon class in offline/AnalysisTrain/picoDST_object. Following cuts are applied to all the events in data and simulation:

Variable Name	North	South
MuID Last Gap	2–4	
p	3–30 GeV	
p*DG0	≤50 GeV	≤80 GeV
p*DDG0	≤40 GeV	
No. of MuTr hits	≥10	
MuTr Track χ^2	≤10	
No. of FVTX hits	≥2	
FVTX-MuTr matching χ^2	≤3	
FVTX station 4 dϕ	≤0.1	
FVTX track χ^2 p value	≥0.05	

An example macro for used for applying the filtering can be found here:
/gpfs/mnt/gpfs02/phenix/fvtx/subsys/fvtx/ajeeta/condor_jobs/Run_filterSingleMuons_ajeeta.C

3 Tuning the simulation

Detector response may not be properly modeled in the simulation. This can be reflected by difference in DCA_R resolution in the MC as compared to data. We perform smearing of the DCA_R distribution in simulation to match the data distribution using events in lastgap 2 and 3 that are dominated by prompt hadrons. Comparing data distributions in lastgap 2 and 3 with analogous distributions in MC can provide handle on the amount of smearing that needs to be applied to the simulation to reproduce the detector response in data. The DCA_R resolution can be expressed as,

$$\sigma_{DCA_R} = \sqrt{\sigma_a^2 + \frac{\sigma_b^2}{p^2}}, \tag{1}$$

where σ_a is a constant term that includes the momentum independent term of the track resolution and the vertex resolution, and σ_b is the momentum dependent term of the track resolution. To that end, we measure σ_{DCA_R} as a function of $1/p$ for various combinations of 1st hit layer bins and rapidity (north or south) bins. The corresponding values can be found in Table 1. In the central peak closest to the DCA_R value of 0, prompt muons and prompt hadrons are expected to have similar distribution; therefore, allowing us to compare prompt muon distribution from J/ψ decay in a $|DCA_R| < 0.2$ to the data distribution.

Arm		σ_a	σ_b
1^{st} Hit Layer = VTX			
North	Data	2.11132e-02 $\pm$ 2.79009e-04	3.03681e-02 $\pm$ 3.16364e-03
	Hadron MC	2.37849e-02 $\pm$ 4.13079e-04	2.50069e-02 $\pm$ 4.01060e-03
	J/ψ MC	2.29506e-02 $\pm$ 6.48986e-05	2.44838e-02 $\pm$ 8.82327e-04
South	Data	2.12375e-02 $\pm$ 2.19200e-04	2.21809e-02 $\pm$ 3.39627e-03
	Hadron MC	2.20712e-02 $\pm$ 3.86018e-04	3.15350e-02 $\pm$ 2.87670e-03
	J/ψ MC	2.27519e-02 $\pm$ 6.68374e-05	2.57171e-02 $\pm$ 9.10938e-04
1^{st} Hit Layer = 1^{st} Layer FVTX			
North	Data	2.21752e-02 $\pm$ 3.28379e-04	4.83921e-02 $\pm$ 3.94041e-03
	Hadron MC	2.11437e-02 $\pm$ 3.49097e-04	4.04573e-02 $\pm$ 2.18023e-03
	J/ψ MC	2.14364e-02 $\pm$ 6.30536e-05	3.88521e-02 $\pm$ 7.11475e-04
South	Data	2.30856e-02 $\pm$ 3.77483e-04	5.87745e-02 $\pm$ 4.51381e-03
	Hadron MC	2.23485e-02 $\pm$ 4.59066e-04	4.79936e-02 $\pm$ 2.73511e-03
	J/ψ MC	2.18768e-02 $\pm$ 7.61599e-05	4.79502e-02 $\pm$ 8.11312e-04
1^{st} Hit Layer = $1^{st} + 2^{nd} + 3^{rd}$ Layer FVTX			
North	Data	2.21273e-02 $\pm$ 1.31833e-03	5.83695e-02 $\pm$ 8.52053e-03
	Hadron MC	1.96208e-02 $\pm$ 2.27394e-03	4.20388e-02 $\pm$ 1.10755e-02
	J/ψ MC	2.12986e-02 $\pm$ 2.80754e-04	2.68994e-02 $\pm$ 2.85520e-03
South	Data	2.31743e-02 $\pm$ 1.25166e-03	4.58073e-02 $\pm$ 1.14926e-02
	Hadron MC	2.12575e-02 $\pm$ 2.09485e-03	4.08716e-02 $\pm$ 1.19216e-02
	J/ψ MC	2.19632e-02 $\pm$ 3.24307e-04	4.46253e-02 $\pm$ 2.80286e-03

Table 1: The constant and momentum dependent terms of DCA_R resolution for data and MC in various categories.

The resolution values obtained from prompt muons are also provided.

The corresponding plots are provided in 1 and 2. The mismatching background contribution from FVTX tracks matching incorrectly with the MuTr tracks have been subtracted from the data prior to the resolution measurement. The values obtained for prompt muon MC and prompt hadron MC are comparable. However, because of the larger statistics in J/ψ simulation as compared to hadron simulation, the values obtained from the former is utilized to smear the data distributions.

FIXME. From 1 and 2 it is seen that the DCA_R resolution in MC is worse than in data when the first hit is in the VTX. This could be a result of oversmearing of the vertex during the embedding, and is being investigated.

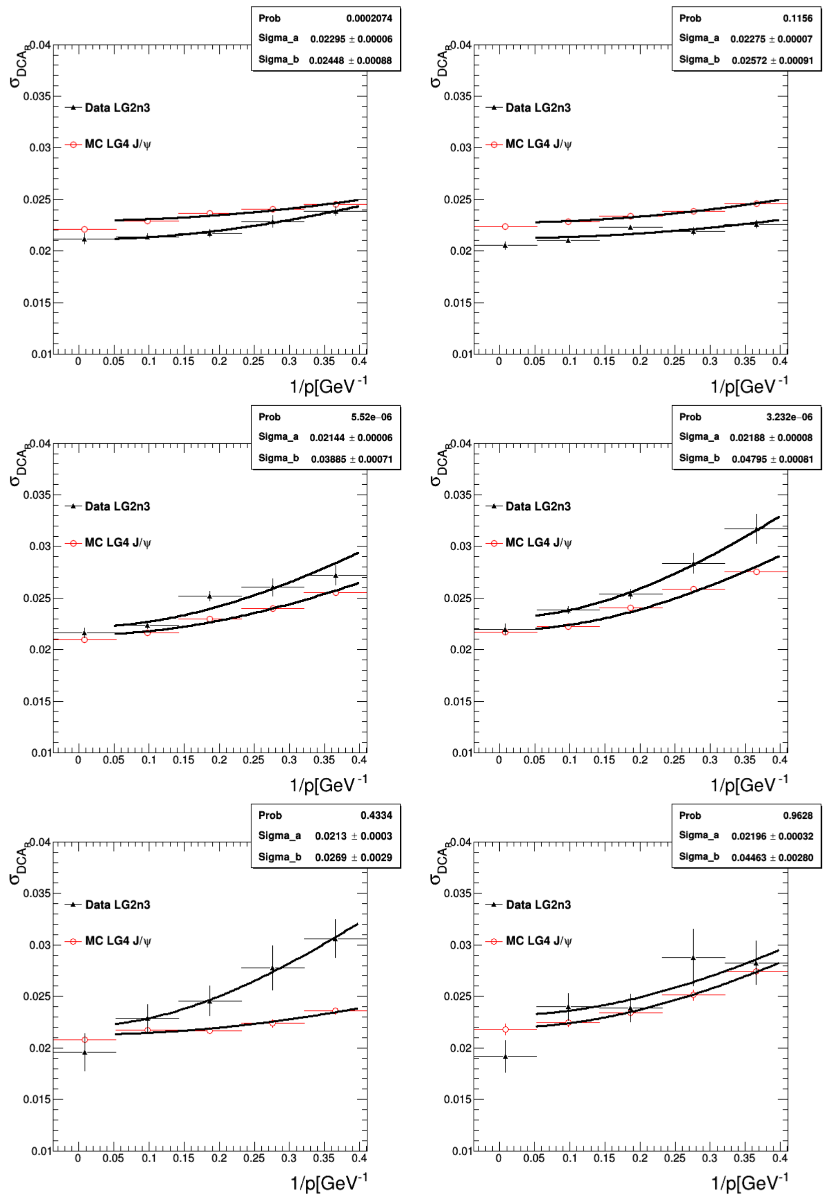

Figure 1: DCA_R resolution as a function of 1/p fitted to obtain σ_a and σ_b in data (black marker) and J/ψ MC (red) for north (left) and south (right) arms and for 1st hit in the Vtx (top), 1^{st} layer of FVTX (middle) and $2^{nd}+3^{rd}+4^{th}$ layers of FVTX (bottom).

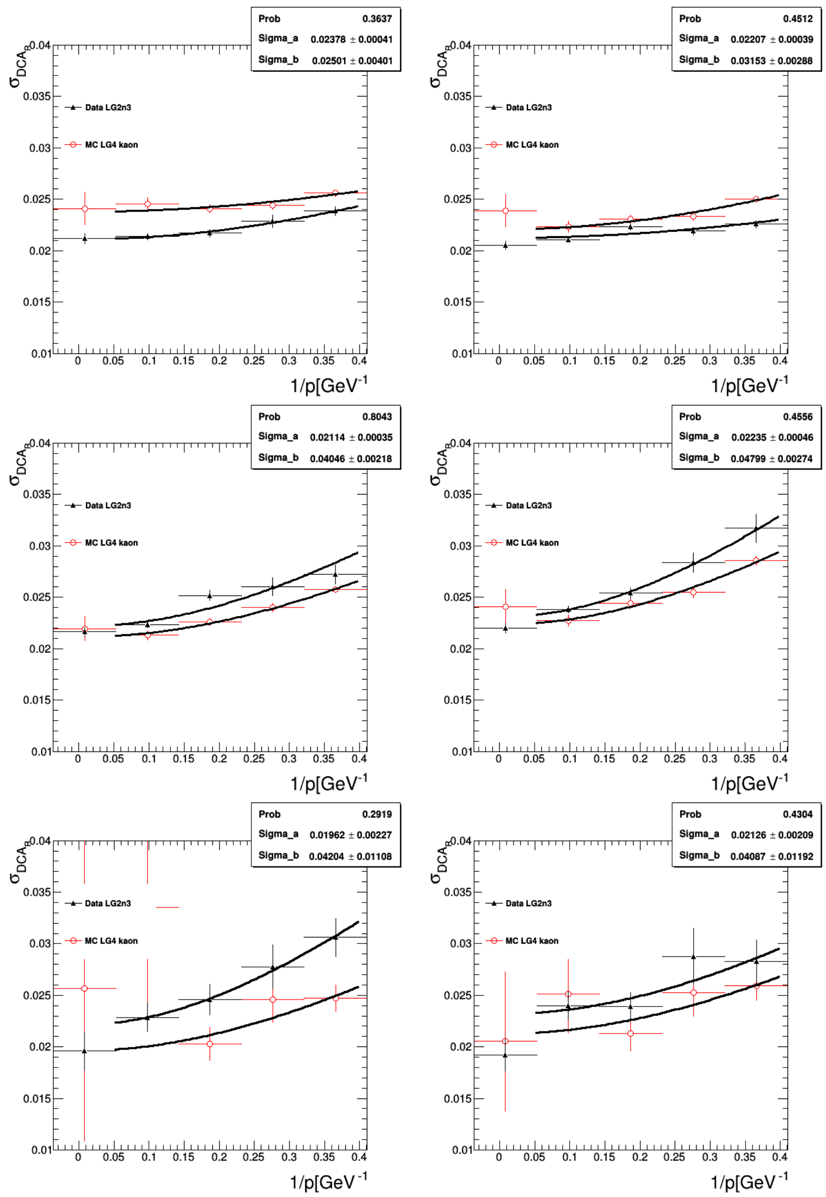

Figure 2: DCA_R resolution as a function of 1/p fitted to obtain σ_a and σ_b in data (black marker) and prompt hadron MC (red) for north (left) and south (right) arms and for 1st hit in the Vtx (top), 1^{st} layer of FVTX (middle) and $2^{nd} + 3^{rd} + 4^{th}$ layers of FVTX (bottom).