

A Tool Package for the Topology Analysis of Inclusive MC Samples

Xingyu Zhou¹, Suxian Li,

Chengping Shen

Beihang University

June 5, 2018

¹zhouxy@buaa.edu.cn

Outline

1 Why

- 1 Why the topology analysis of inclusive MC samples?
- 2 Why the tool package?

2 What

- 1 What does the TopoAna program do?
- 2 What does the TopoAna package contain?

3 How

- 1 How to prepare the input for the TopoAna program?
- 2 How to run the TopoAna program?

4 Summary & Outlook

Why the topology analysis of inclusive MC samples?

- One of the most important tasks in physics analyses is event selection, or in other words, to veto backgrounds.
- As for background analyses, inclusive MC samples are very helpful, because they contain a basic knowledge of events from collision.
- However, the information in the inclusive MC samples is overwhelming, which makes it difficult for us to see main backgrounds directly from the raw MC information.
- To find the main backgrounds quickly, topology analysis programs to categorize, count and tag events are developed.
- With the main backgrounds identified, one can optimize the selection criteria to suppress backgrounds by analyzing the differences between the main background and signal events.

Why the tool package?

- A program called “Topo”, developed by Prof. Shuxian Du of BESIII collaboration, is widely used in BESIII experiment.
- I learned the idea of topology analysis and a lot of programming techniques from it several years ago when I was a PhD student working on BESIII experiment.
- To meet my own needs and also to practice developing analysis tools with C++, ROOT and LaTeX, I wrote a new topology analysis package — “TopoAna” — from scratch.
- Since the “TopoAna” program could also be a useful tool for Belle II experiment, we revised and extended it in the recent months, making it more well-rounded and suitable for Belle II experiment.

What does the TopoAna program do? – Basics

Using the following truth information of MC generated particles saved in the TTree objects of input ROOT files:

- Number of the particles,
- Array of the PDG codes of the particles,
- Array of the mother indices of the particles,

to:

- categorize, count and tag events,
- summarize the statistics of topology in the output txt/tex/pdf files,
- insert tags of topology in all the entries of the TTree object of the output root file. Besides, the topology tags are identical with those listed in the txt/tex/pdf files.

Basically, every event is categorized by complete event tree and by only event final states. Accordingly,

- two tables are listed in the output pdf file
- two tags are inserted in the output root file.

What does the TopoAna package contain?

- You can find the TopoAna package in the following directory on KEKCC servers:
`/home/belle2/zhouxy/workarea/tools/topoana/v1.4.0`
- One script and six directories are contained in the package:
 - `setup.sh` — script to set the path and compile the package
 - `include` — directory of the header file
 - `src` — directory of the source files
 - `bin` — directory of the executable file
 - `share` — directory of common data, style and card files
 - `doc` — directory of related documents
 - `test` — directory of a test example

```
[zhouxy@cw09 v1.4.0]$ pwd
/home/belle2/zhouxy/workarea/tools/topoana/v1.4.0
[zhouxy@cw09 v1.4.0]$ ls
bin/  doc/  include/  setup.sh*  share/  src/  test/
[zhouxy@cw09 v1.4.0]$
```

What does the TopoAna package contain?

include/src/bin

By executing the script `setup.sh`, the executable file can be obtained from the header and source files.

- `topoana.h` — the header file
- `topoana.cpp` — the main source file
- `topoana.exe` — the executable file

It is strongly recommended to execute `setup.sh` first after you copy the package to your own directory and before you use it.

```
[zhouxy@cw14 vl.4.0]$ pwd
/home/belle2/zhouxy/workarea/tools/topoana/v1.4.0
[zhouxy@cw14 vl.4.0]$ ls
bin/ doc/ include/ setup.sh* share/ src/ test/
[zhouxy@cw14 vl.4.0]$ ls include/
topoana.h
[zhouxy@cw14 vl.4.0]$ ls src/
countLiaInVlib.cpp      getRslt.cpp              sortBy1stFromLrgToSml.cpp  writeInfOnRslt.cpp
countPFSts.cpp          makeMapsOnPdata.cpp     sortByPidAndPchrg.cpp     writePnmFromPid.cpp
countSeqEvtBrsInEvtTr.cpp read1stLineOrCloseCurly.cpp sortBySzPidAndPchrg.cpp  writeRsltIntoTexFl.cpp
getEvtIFSts.cpp         readCard.cpp            sortPs.cpp                writeRsltIntoTexFl.cpp
getEvtTr.cpp            readCloseCurly.cpp     topoana.cpp               trim.cpp
getPdfFlFromTexFl.cpp  readExtraLinesOrCloseCurly.cpp writeErrInfOnPid3PchrgMap.cpp
getPidFromTxtPnm.cpp   readOpenCurly.cpp
[zhouxy@cw14 vl.4.0]$ ls bin/
topoana.exe*
[zhouxy@cw14 vl.4.0]$
```

How to prepare the input for the TopoAna program?

To prepare the input for the TopoAna program, a ntuple tool called **MCGenTruthForTopoAna** is developed with reference to the existing ntuple tool **MCGenKinematics**. Compared with **MCGenKinematics**, **MCGenTruthForTopoAna**

- only saves the numbers, PDG code arrays and mother index arrays of MC generated particles,
- won't save unnecessary zero values in the arrays of PDG codes and Mother indexes.

Steps to prepare the input for the TopoAna program:

- install **MCGenTruthForTopoAna**,
- insert the statement on **MCGenTruthForTopoAna** in analysis steering files,
- execute the analysis steering files.

How to prepare the input for the TopoAna program?

Install MCGenTruthForTopoAna

```
[zhouxy@cw05 release-00-09-00]$ pwd
/home/Belle2/zhouxy/workarea/releases/release-00-09-00
[zhouxy@cw05 release-00-09-00]$ ls analysis/NtupleTools/include/NtupleMCGenTruthForTopoAnaTool.h
analysis/NtupleTools/include/NtupleMCGenTruthForTopoAnaTool.h
[zhouxy@cw05 release-00-09-00]$ ls analysis/NtupleTools/src/NtupleMCGenTruthForTopoAnaTool.cc
analysis/NtupleTools/src/NtupleMCGenTruthForTopoAnaTool.cc
[zhouxy@cw05 release-00-09-00]$ grep -C 2 "NtupleMCGenTruthForTopoAnaTool" analysis/NtupleTools/src/NtupleToolList.cc
#include <analysis/NtupleTools/NtupleMCGenKinematicsTool.h>
#include <analysis/NtupleTools/NtupleMCGenCMSKinematicsTool.h>
#include <analysis/NtupleTools/NtupleMCGenTruthForTopoAnaTool.h>

using namespace Belle2;

--
    else if (strToolName == "MCGenKinematics") return new NtupleMCGenKinematicsTool(tree, d, strOption);
    else if (strToolName == "MCGenCMSKinematics") return new NtupleMCGenCMSKinematicsTool(tree, d, strOption);
    else if (strToolName == "MCGenTruthForTopoAna") return new NtupleMCGenTruthForTopoAnaTool(tree, d, strOption);
    B2WARNING("NtupleTool " << strToolName << " is not available!");
    return NULL;
[zhouxy@cw05 release-00-09-00]$ scons
scons: Reading SConscript files ...
Checking for Belle II environment setup...(cached) yes
Checking for analysis setup...(cached) no

scons: done reading SConscript files.
scons: Building targets ...
scons: building associated VariantDir targets: build/Linux_x86_64/opt
*** symlinking : include/analysis/NtupleTools/NtupleMCGenTruthForTopoAnaTool.h
*** compiling : /cvms/belle.cern.ch/sl6/releases/release-00-09-00/framework/io/src/RootIOUtilities.cc
*** linking : lib/Linux_x86_64/opt/libframework_io.so
*** compiling : analysis/NtupleTools/src/NtupleMCGenTruthForTopoAnaTool.cc
*** compiling : analysis/NtupleTools/src/NtupleToolList.cc
*** linking : bin/Linux_x86_64/opt/test_framework
*** linking : modules/Linux_x86_64/opt/libseqroot.so
*** linking : modules/Linux_x86_64/opt/libbeast_microtpc_modules.so
*** linking : modules/Linux_x86_64/opt/libBGOverlayInput.so
*** linking : modules/Linux_x86_64/opt/librootio.so
*** linking : bin/Linux_x86_64/opt/create_dedx_PDFs
*** linking : bin/Linux_x86_64/opt/merge_basf2_files
*** linking : lib/Linux_x86_64/opt/libmva.so
*** linking : modules/Linux_x86_64/opt/libbeast_analysis_modules.so
*** linking : bin/Linux_x86_64/opt/test_all
*** linking : lib/Linux_x86_64/opt/libanalysis_NtupleTools.so
*** linking : modules/Linux_x86_64/opt/libNtupleMaker.so
scons: done building targets.
[zhouxy@cw05 release-00-09-00]$
```

How to prepare the input for the TopoAna program?

Insert the statement on MCGenTruthForTopoAna in analysis steering files

```
[zhouxy@cw05 mixed]$ cat NtupleTools_MCGenTruthForTopoAna_mixed_test.py
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

from basf2 import *
from modularAnalysis import inputMdst
from modularAnalysis import ntupleFile
from modularAnalysis import ntupleTree
from modularAnalysis import analysis_main

# load input ROOT file
inputMdst('default', 'mdst_000018_prod00002218_task00000018.root')

# define Ntuple tools
toolsEvt = ['EventMetaData', 'e+']
toolsEvt += ['MCGenTruthForTopoAna', '^e+']

ntupleFile('mixed1.root')
ntupleTree('evt', '', toolsEvt)

# Process the events
process(analysis_main)

# print out the summary
print(statistics)
```

● Though, the example is illustrated now with no cuts applied in the analysis steering files, it should be used in practice with cuts applied in the analysis steering files.

How to prepare the input for the TopoAna program?

Execute the analysis steering files

```
[zhouxy@cw05 mixed]$ pwd
/home/belle2/zhouxy/workarea/releases/release-00-09-00/analysis/mytests/NtupleTools/MCGen
TruthForTopoAna/mixed
[zhouxy@cw05 mixed]$ ls
NtupleTools_MCGenTruthForTopoAna_mixed_test.py
mdst_000018_prod00002218_task00000018.root
[zhouxy@cw05 mixed]$ basf2 NtupleTools_MCGenTruthForTopoAna_mixed_test.py >NtupleTools_MC
GenTruthForTopoAna_mixed_test.out 2>NtupleTools_MCGenTruthForTopoAna_mixed_test.err
[zhouxy@cw05 mixed]$ ls
NtupleTools_MCGenTruthForTopoAna_mixed_test.err
NtupleTools_MCGenTruthForTopoAna_mixed_test.out
NtupleTools_MCGenTruthForTopoAna_mixed_test.py
mdst_000018_prod00002218_task00000018.root
mixed1.root
[zhouxy@cw05 mixed]$
```

How to prepare the input for the TopoAna program?

Results

```
[zhouxy@cw05 mixed1$ root -l
root [0] TFile f("mixed1.root")
(TFile &) Name: mixed1.root Title:
root [1] f.ls()
TFile**          mixed1.root
TFile*           mixed1.root
KEY: TTree      evt;1
root [2] evt->Show(0,200)
=====> EVENT:0
  exp_no          = 0
  run_no          = 0
  evt_no          = 170002
  Nps             = 33
  Pid             = 300553,
                  511, -511, -411, 223, 211, 113, 413, -211, -211, 211,
                  111, 313, 13, -14, 111, 22, 211, -211, 421, 211,
                  22, 22, 321, -211, 22, 22, 20213, -321, 113, 211,
                  211, -211
  Midx            = 0,
                  0, 0, 1, 1, 1, 1, 2, 2, 2, 2,
                  2, 3, 3, 3, 4, 4, 6, 6, 7, 7,
                  11, 11, 12, 12, 15, 15, 19, 19, 27, 27,
                  29, 29
  m_nCands        = 0
  m_iCand         = 0
```

- Nps — Number of particles
- Pid — Array of particle identifications
- Midx — Array of mother indices of particles

How to run the TopoAna program?

Steps to run the TopoAna program:

- 1 prepare the input for TopoAna as discussed above.
- 2 fill out the input card file
 - an empty card template (**topoana.card**) can be found in the share directory
- 3 execute the TopoAna program
 - command: **path/topoana.exe input_card_file_name**
 - the **default** input card file name is **topoana.card**

How to run the TopoAna program? — Basics

Fill out the input card file

```
[zhouxy@cw08 test]$ sed -n '1,31p' mixed_topoana.card
# The following six items are indispensable, and they can't be left empty.

% Names of input root files
{
  mixed*.root
}

% Tree name
{
  evt
}

% Branch name of the number of particles
{
  Nps
}

% Branch name of the array of particle identifications
{
  Pid
}

% Branch name of the array of the mother indices of particles
{
  Midx
}

% Main name of output files
{
  mixed_topoana
}
[zhouxy@cw08 test]$
```

Wildcards are supported in the first card item just as that for the method Add() of TChain class.

How to run the TopoAna program? — Basics

Execute the TopoAna program

```
[zhouxy@cw06 test]$ pwd
/home/belle2/zhouxy/workarea/tools/topoana/v1.4.0/test
[zhouxy@cw06 test]$ ls
mixed1.root  mixed2.root  mixed_topoana.card
[zhouxy@cw06 test]$ ./bin/topoana.exe mixed_topoana.card >mixed_topoana.OUT 2>mixed_topoana.ERR
[zhouxy@cw06 test]$ ls
mixed1.root  mixed_topoana.ERR  mixed_topoana.card  mixed_topoana.root  mixed_topoana.txt
mixed2.root  mixed_topoana.OUT  mixed_topoana.pdf  mixed_topoana.tex
[zhouxy@cw06 test]$
```

- **mixed_topoana.pdf/tex/txt** — The output pdf/tex/txt file² containing the statistics of topology.
- **mixed_topoana.root** — The output root file³ containing data in the input root files but with the tags of topology.

²Although they are in different formats, they have the same information. The pdf file is the easiest to read. It is converted from the tex file by the "pdflatex" command. If necessary, one could check and see the txt file (with text processing commands).

³Except for this, the TTree object of the output root file is entirely the same as that (those) of the input root file (files). Besides, the topology tags are identical with those listed in the txt/tex/pdf files.

The first page of the first table in the output PDF file

- 

event tree
(event initial-final states)

nCmltEvs

How to run the TopoAna program? – Basics

The first page of the second table in the output PDF file

iEvtIFSts – index of event initial-final states

Table 2: Event initial-final states.			
index	event initial-final states	iEvtIFSts	nEvts
1	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	3	11
2	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	166	10
3	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1169	10
4	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1406	10
5	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2470	10
6	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2733	10
7	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2836	10
8	$e^+e^- \rightarrow e^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	3068	10
9	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1663	9
10	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1950	9
11	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	650	9
12	$e^+e^- \rightarrow \mu^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2061	8
13	$e^+e^- \rightarrow e^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	32	8
14	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1654	8
15	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	707	8
16	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	488	8
17	$e^+e^- \rightarrow e^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	3181	8
18	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	3289	8
19	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	5590	8
20	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	6410	8
21	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1946	7
22	$e^+e^- \rightarrow \mu^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	276	7
23	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2007	7
24	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	792	7
25	$e^+e^- \rightarrow e^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2301	7
26	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2309	7
27	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1038	7
28	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	361	7
29	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1380	7
30	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	2898	7
31	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1402	7
32	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	392	7
33	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	1418	7
34	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	3893	7
35	$e^+e^- \rightarrow \mu^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	4410	7
36	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	4459	7
37	$e^+e^- \rightarrow K_L \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	5365	7
38	$e^+e^- \rightarrow e^+ \mu_\mu \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	198	7
39	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	254	7
40	$e^+e^- \rightarrow \pi^+ \pi^+ \pi^+ \pi^+ \pi^- \pi^- \pi^- \pi^- K^+ K^-$	290	6

How to run the TopoAna program? — Basics

The first entry of the TTree object in the output ROOT file

```
[zhouxy@cw01 test]$ pwd
/home/belle2/zhouxy/workarea/tools/topoana/v1.4.0/test
[zhouxy@cw01 test]$ ls
mixed1.root  mixed_topoana.ERR  mixed_topoana.card  mixed_topoana.root  mixed_topoana.txt
mixed2.root  mixed_topoana.OUT  mixed_topoana.pdf   mixed_topoana.tex
[zhouxy@cw01 test]$ root -l
root [0] TFile f("mixed_topoana.root")
(TFile &) Name: mixed_topoana.root Title:
root [1] f.ls()
TFile**          mixed_topoana.root
TFile*           mixed_topoana.root
KEY: TTree       evt;1
root [2] evt->Show(0)
=====> EVENT:0
exp_no           = 0
run_no           = 0
evt_no           = 170002
Nps              = 33
Pid              = 300553,
                  511, -511, -411, 223, 211, 113, 413, -211, -211, 211,
                  111, 313, 13, -14, 111, 22, 211, -211, 421
Midx             = 0,
                  0, 0, 1, 1, 1, 1, 2, 2, 2, 2,
                  2, 3, 3, 3, 4, 4, 6, 6, 7
m_nCands         = 0
m_iCand          = 0
iEvtTr           = 0
iEvtIFSts       = 0
```

- iEvtTr — index of event tree
- iEvtIFSts — index of event initial-final states

Summary & Outlook

- The program **TopoAna** is developed for the topology analysis of inclusive MC samples.
- The ntuple tool **MCGenTruthForTopoAna** is implemented to prepare the input for **TopoAna**.

Due to the limited time, the following features are not covered in the report and will be presented at the next meetings:

- Tagging of the signal events,
- Other options in the input card file related to the execution of the TopoAna program,
- The module **EnableMyNtupleToolsOnTheGrid** to run **MCGenTruthForTopoAna** on the grid.

Welcome to TopoAna

Please let us know:

- if you have any problems with it;
- if you have any questions about it;
- if you have any suggestions on how to improve it;
- if you find any bugs in it;
- if you want to extend its functions.

I hope it can be helpful to your studies.