

LCFIPlus update

T. Suehara, T. Tanabe
ICEPP, The University of Tokyo

April 25, 2012
KILC12 @ Daegu, Korea

- Introduction
- Framework
- Improvements in jet/vertex finding
- Flavor tagging performance

Introduction

NIM A 610 573 (2009)

LCFIVertex

- ★ vertex finder & flavor tagger for LOI
- ★ **neural net difficult to extend**

rewrite

LCFIPlus

- ★ vertex finding, jet finding, flavor tagger in one package
- ★ exploit **TMVA**
- ★ flexible XML configuration

arXiv:1110.5785

Jet Finding

- ★ need to improve for **multi-jet events**
- ★ vertex first, jet second approach

implement

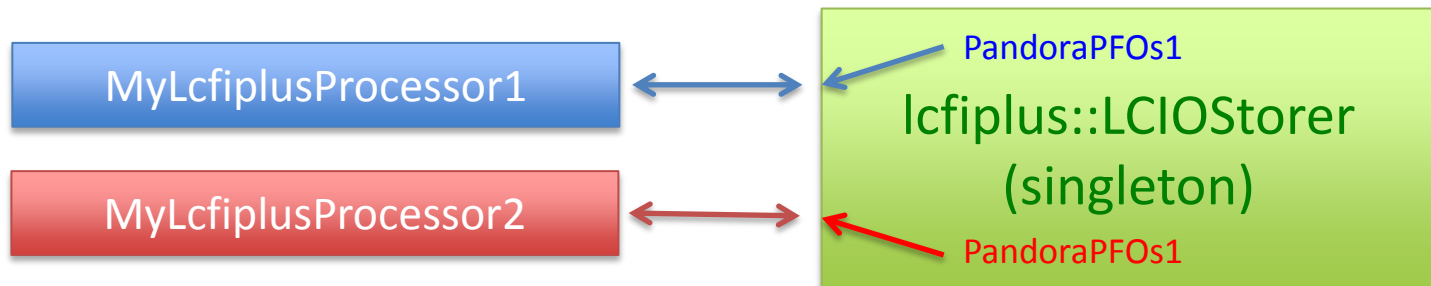
Included in ilcsoft since v01-13

Data Types & Event Model

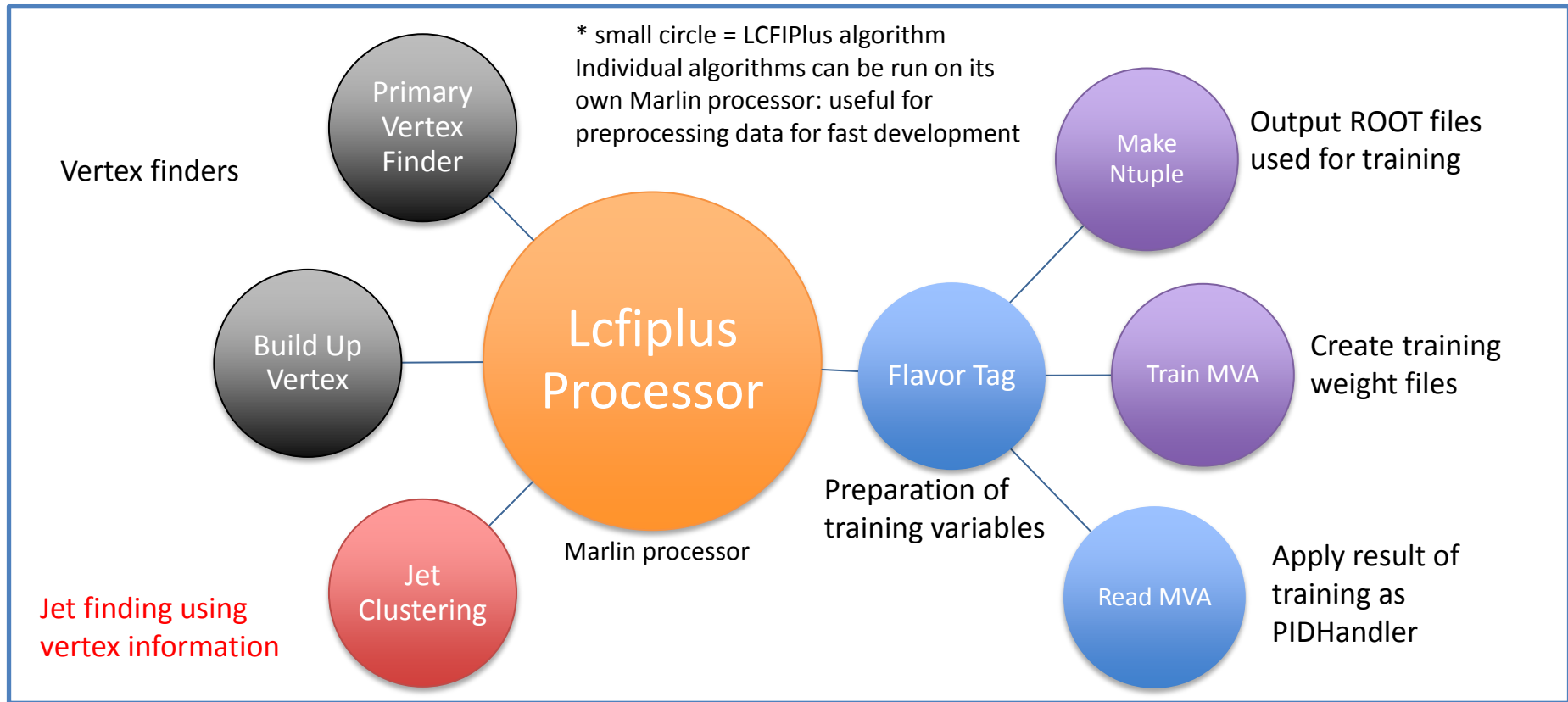
LCIO	LCFIPlus
ReconstructedParticle (PFOs)	Track, Neutral
ReconstructedParticle (Jets)	Jet
Vertex	Vertex
MCParticle	MCParticle (optional)

← configure with XML →

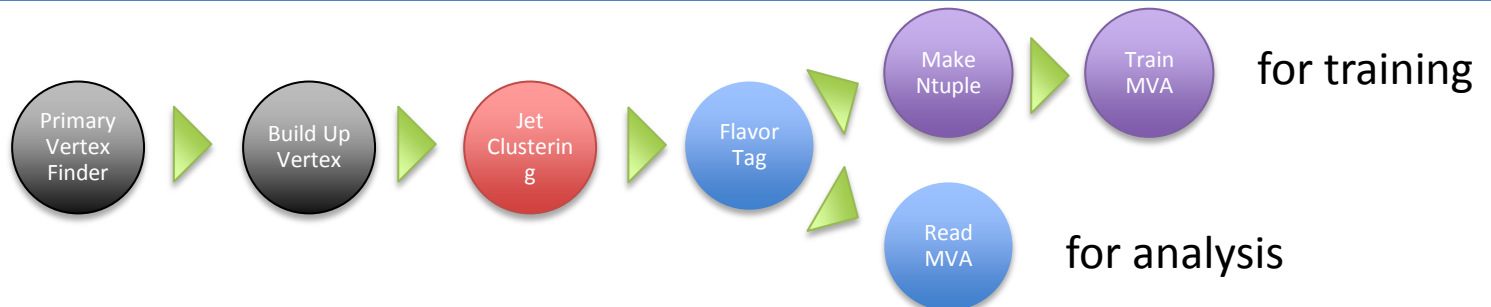
- LCFIPlus has its own event model for the convenience of its algorithms
 - has automatic conversion of data types from/to LCIO: “LCIOStorer” class
 - takes advantage of **LCIO persistence** model for read/write
 - performs intelligent type conversion: pfo vs. jet vs. vertex
 - same collection names are used as in LCIO
- latest changes will be included in the next ilcsoft release (v01-13-06)
 - ability to run multiple instances of LcfixplusProcessor (Marlin processor) with different input lists
 - improvements in vertex finding and flavor tagging variables
 - first version of training weight files for Z->qq / 6f events



Algorithms



Data Flow:



Documentation & Feedback

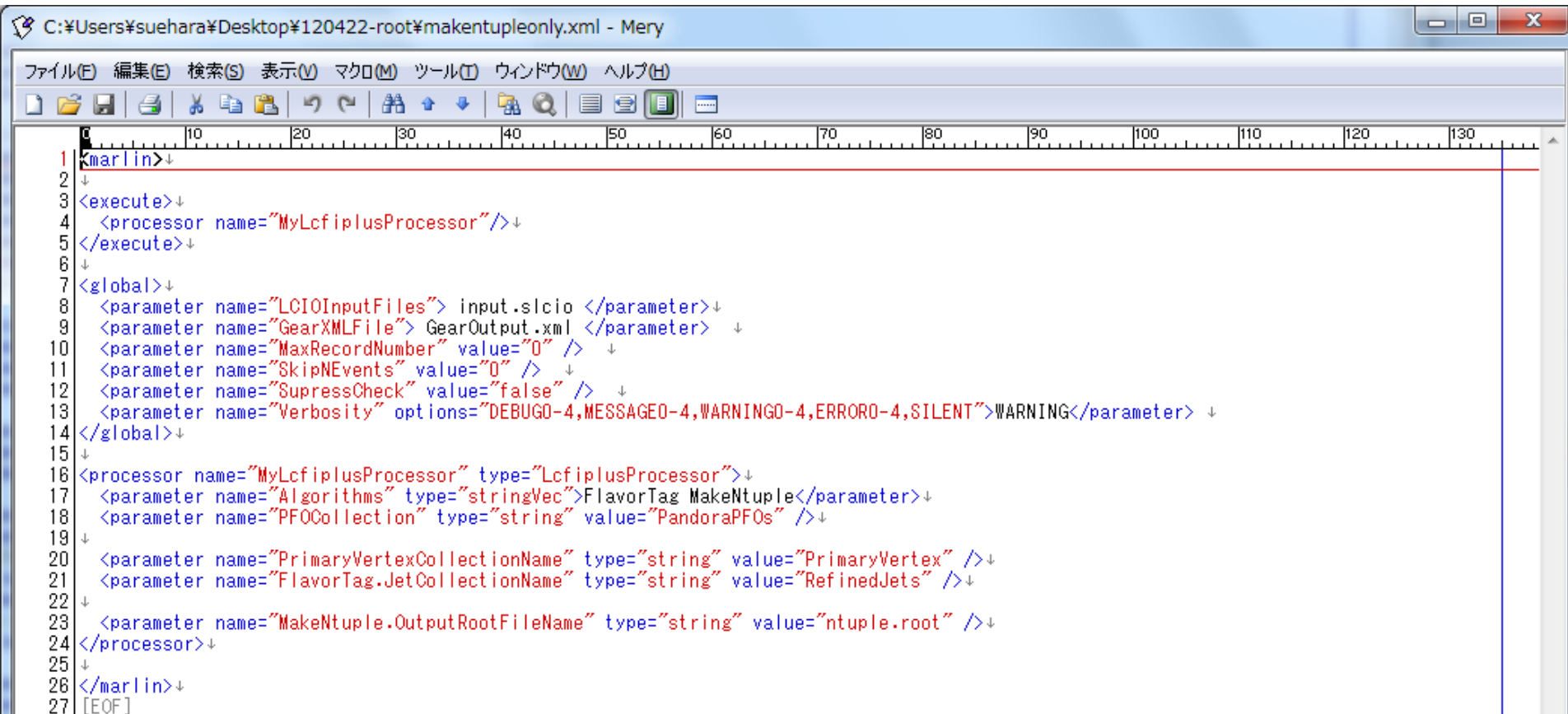
- Doxygen class reference
- User feedback + documentation system has been setup @ CERN (J. Strube):
 - Documentation **wiki** hosted at SLAC (thanks N. Graf)
 - bug tracker (JIRA) also available
 - <https://confluence.slac.stanford.edu/display/ilc/LCFIPlus>
 - some documentation present, will be described in more detail later
- Early bug reports (Thanks: J. Engels, F. Gaede, J. Strube, A. Sailer)
- Nightly builds and check input variables (J. Strube)
- Feedback and support from LC community has been very helpful. **Will continue working with SiD for a smooth transition from LCFIVertex to LCFIPlus.**

Steering

- Typical user experience:
 - obtain DST files
 - reconstruct primary and secondary vertices (if missing)
 - select weight files
 - run analysis with JetClustering + FlavorTag + ReadMVA
- For training:
 - prepare training samples in DST format (with vertices)
 - ntuple preparation step:
 - JetClustering + FlavorTag + MakeNtuple
 - (concatenate the ROOT files with “hadd” if necessary)
 - training step:
 - FlavorTag + TrainMVA

```
1 <marlin>↓
2 ↓
3 <execute>↓
4   <processor name="MyLcfiplusProcessor"/>↓
5   <processor name="MyLcfiplusProcessor2"/>↓
6   <processor name="MyLCIOOutputProcessor"/>↓
7 </execute>↓
8 ↓
9 <global>↓
10  <parameter name="LCIOInputFiles"> input.slcio </parameter>↓
11  <parameter name="GearXMLFile"> GearOutput.xml </parameter> ↓
12  <parameter name="MaxRecordNumber" value="0" /> ↓
13  <parameter name="SkipNEvents" value="0" /> ↓
14  <parameter name="SupressCheck" value="false" /> ↓
15  <parameter name="Verbosity" options="DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT">WARNING</parameter> ↓
16 </global>↓
17 ↓
18 <processor name="MyLcfiplusProcessor" type="LcfiplusProcessor">↓
19   ↓
20   <!-- run primary and secondary vertex finders -->↓
21   <parameter name="Algorithms" type="stringVec"> PrimaryVertexFinder BuildUpVertex </parameter>↓
22   <parameter name="ReadSubdetectorEnergies" type="int" value="1"/> <!-- true for ILD -->↓
23   <parameter name="UpdateVertexRPDaughters" type="int" value="0"/> <!-- false for non-updative PandoraPFOs -->↓
24   <parameter name="PrintEventNumber" type="int" value="1"/>↓
25   <!-- specify input collection names -->↓
26   <parameter name="PFOCollection" type="string" value="PandoraPFOs" />↓
27   <parameter name="PrimaryVertexCollectionName" type="string" value="PrimaryVertex" />↓
28   <parameter name="BuildUpVertexCollectionName" type="string" value="BuildUpVertex" />↓
29   <parameter name="BuildUpVertex.V0VertexCollectionName" type="string" value="BuildUpVertex_V0" />↓
30   <!-- nip -->↓
31 </processor>↓
32 <processor name="MyLcfiplusProcessor2" type="LcfiplusProcessor">↓
33   <!-- run primary and secondary vertex finders -->↓
34   <parameter name="Algorithms" type="stringVec"> JetClustering, JetVertexRefiner </parameter>↓
35   <!-- general parameters -->↓
36   <parameter name="PFOCollection" type="string" value="PandoraPFOs" /> <!-- input PFO collection -->↓
37   <!-- nip -->↓
38 </processor>↓
39 ↓
40 <processor name="MyLCIOOutputProcessor" type="LCIOOutputProcessor">↓
41   <parameter name="LCIOOutputFile" type="string"> jet.slcio </parameter>↓
42   <parameter name="LCIOWriteMode" type="string" value="WRITE_NEW"/>↓
43 </processor>↓
44 ↓
45 </marlin>↓
```

MakeNtuple



The screenshot shows a Mery text editor window with the title bar "C:\Users\suehara\Desktop\120422-root\makentupleonly.xml - Mery". The menu bar includes "ファイル(F)", "編集(E)", "検索(S)", "表示(V)", "マクロ(M)", "ツール(T)", "ウィンドウ(W)", and "ヘルプ(H)". The toolbar contains icons for file operations and editing. The text area displays an XML configuration file with the following content:

```
1 <marlin>↓
2 ↓
3 <execute>↓
4   <processor name="MyLcfiplusProcessor"/>↓
5 </execute>↓
6 ↓
7 <global>↓
8   <parameter name="LCIOInputFiles"> input.slcio </parameter>↓
9   <parameter name="GearXMLFile"> GearOutput.xml </parameter> ↓
10  <parameter name="MaxRecordNumber" value="0" /> ↓
11  <parameter name="SkipNEvents" value="0" /> ↓
12  <parameter name="SupressCheck" value="false" /> ↓
13  <parameter name="Verbosity" options="DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT">WARNING</parameter> ↓
14 </global>↓
15 ↓
16 <processor name="MyLcfiplusProcessor" type="LcfiplusProcessor">↓
17   <parameter name="Algorithms" type="stringVec">FlavorTag MakeNtuple</parameter>↓
18   <parameter name="PFOCollection" type="string" value="PandoraPFOs" />↓
19 ↓
20   <parameter name="PrimaryVertexCollectionName" type="string" value="PrimaryVertex" />↓
21   <parameter name="FlavorTag.JetCollectionName" type="string" value="RefinedJets" />↓
22 ↓
23   <parameter name="MakeNtuple.OutputRootFileName" type="string" value="ntuple.root" />↓
24 </processor>↓
25 ↓
26 </marlin>↓
27 [EOF]
```



```
C:\Users\suehara\Desktop\120422-root\train.xml - Mery
ファイル(F) 編集(E) 検索(S) 表示(V) マクロ(M) ツール(T) ウィンドウ(W) ヘルプ(H)
[Icons]
0 10 20 30 40 50 60 70 80 90 100 110 120 130
1 <marlin>↓
2 ↓
3 <execute>↓
4   <processor name="MyLcfiplusProcessor"/>↓
5 </execute>↓
6 ↓
7 <global>↓
8   <parameter name="LCIOInputFiles"> </parameter>↓
9   <parameter name="MaxRecordNumber" value="0" /> ↓
10  <parameter name="SkipNEvents" value="0" /> ↓
11  <parameter name="SupressCheck" value="false" /> ↓
12  <parameter name="GearXMLFile"> GearOutput.xml </parameter> ↓
13  <parameter name="Verbosity" options="DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT">WARNING</parameter> ↓
14 </global>↓
15 ↓
16 <processor name="MyLcfiplusProcessor" type="LcfiplusProcessor">↓
17   <parameter name="Algorithms" type="stringVec">TrainMVA</parameter>↓
18   <parameter name="PFOCollection" type="string" value="PandoraPFOs" />↓
19   <parameter name="FlavorTag.JetCollectionName" type="string" value="RefinedJets" />↓
20 ↓
21   <parameter name="FlavorTag.WeightsDirectory" type="string" value="lcfiweights" />↓
22   <parameter name="FlavorTag.WeightsPrefix" type="string" value="zpole_v00" />↓
23   <parameter name="FlavorTag.BookName" type="string" value="bdt" />↓
24 ↓
25   <parameter name="FlavorTag.CategoryDefinition1" type="string">nvtx==0</parameter>↓
26   <parameter name="FlavorTag.CategoryVariables1" type="stringVec">↓
27     trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz↓
28   </parameter>↓
29   <parameter name="FlavorTag.CategorySpectators1" type="stringVec">↓
30     aux nvtx↓
31   </parameter>↓
32 ↓
33   <parameter name="FlavorTag.CategoryDefinition2" type="string">nvtx==1&&nvtxall==1</parameter>↓
34   <parameter name="FlavorTag.CategoryVariables2" type="stringVec">↓
35     trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz↓
36     vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmulti vtxmasspc vtxprobz↓
37   </parameter>↓
38   <parameter name="FlavorTag.CategorySpectators2" type="stringVec">↓
39     aux nvtx↓
40   </parameter>↓
41 ↓
42   <parameter name="FlavorTag.CategoryDefinition3" type="string">nvtx==1&&nvtxall==2</parameter>↓
```

HTML

8行, 48桁

改行

日本語 (シフト JIS)

CR + LF

C:\Users\suehara\Desktop\120422-root\flavtagtmp.xml * - Mery

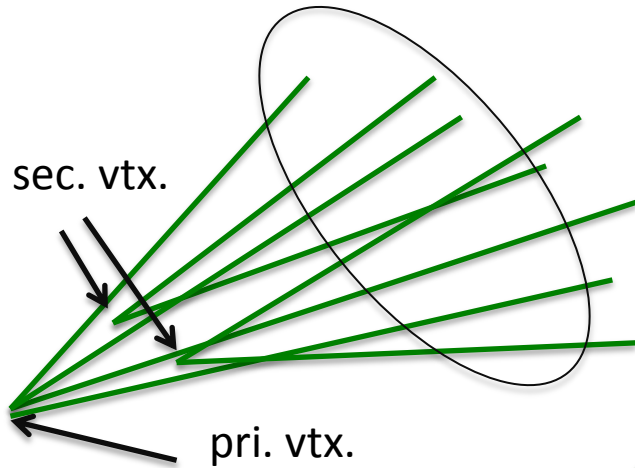
ファイル(F) 編集(E) 検索(S) 表示(V) マクロ(M) ツール(T) ウィンドウ(W) ヘルプ(H)

```
1 <marlin>↓
2 ↓
3 <execute>↓
4 <processor name="MyLcfiplusProcessor"/>↓
5 <processor name="MyJetFlavtagProcessor"/>↓
6 </execute>↓
7 ↓
8 <global>↓
9 <parameter name="MaxRecordNumber" value="1000" />↓
10 <parameter name="SkipNEvents" value="0" /> ↓
11 <parameter name="SupressCheck" value="false" /> ↓
12 <parameter name="GearXMLFile"> ILD_00.xml </parameter> ↓
13 <parameter name="Verbosity" options="DEBUG-4,MESSAGE-4,WARNING-4,ERROR-4,SILENT"> SILENT </parameter> ↓
14 <parameter name="LCIOInputFiles">input.slcio</parameter>↓
15 </global>↓
16 ↓
17 <processor name="MyLcfiplusProcessor" type="LcfiplusProcessor">↓
18 <parameter name="Algorithms" type="stringVec">JetClustering FlavorTag ReadMVA</parameter>↓
19 <parameter name="PrimaryVertexCollectionName" type="string" value="PrimaryVertex" />↓
20 ↓
21 <parameter name="PFOCollection" type="string" value="PandoraPFOs" />↓
22 <parameter name="MCPCollection" type="string" value="MCParticlesSkimmed" />↓
23 <parameter name="MCPFORelation" type="string" value="RecoMCTruthLink" />↓
24 <!--nip-->↓
25 ↓
26 <parameter name="FlavorTag.WeightsDirectory" type="string" value="lcfiweights" />↓
27 <parameter name="FlavorTag.WeightsPrefix" type="string" value="zpole_v00" />↓
28 <parameter name="FlavorTag.BookName" type="string" value="bdt" />↓
29 ↓
30 <parameter name="FlavorTag.CategoryDefinition1" type="string">nvtx==0</parameter>↓
31 <parameter name="FlavorTag.CategoryVariables1" type="stringVec">↓
32   trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz↓
33 </parameter>↓
34 <parameter name="FlavorTag.CategorySpectators1" type="stringVec">↓
35   aux nvtx↓
36 </parameter>↓
37 ↓
38 <parameter name="FlavorTag.CategoryDefinition2" type="string">nvtx==1</parameter>↓
39 <parameter name="FlavorTag.CategoryVariables2" type="stringVec">↓
40   trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz↓
41   vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob↓
42 </parameter>↓
43 <parameter name="FlavorTag.CategorySpectators2" type="stringVec">↓
44   aux nvtx↓
45 </parameter>↓
46 ↓
```

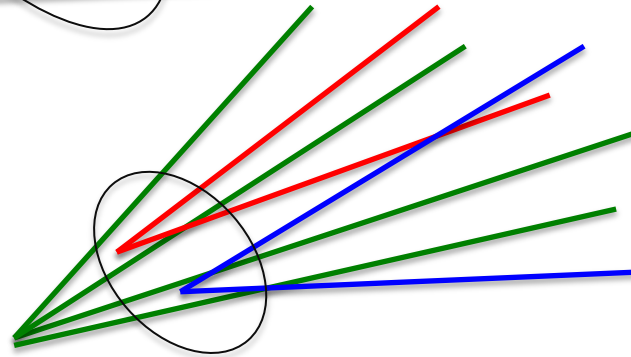
HTML 1行, 1桁 0x003C 日本語 (シフト JIS) CR + LF

Vertex-Jet Finding Overview

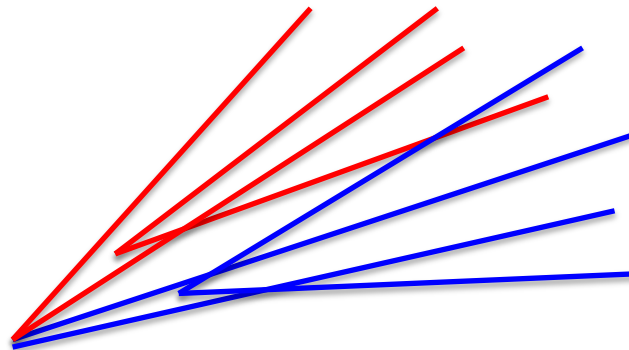
arXiv:1110.5785



1. Difficult to separate two b-jets which are close. Ordinary kt algorithm tends to merge them.



2. To overcome this, find secondary vertices first, and use them as *seeds* for jet finding.



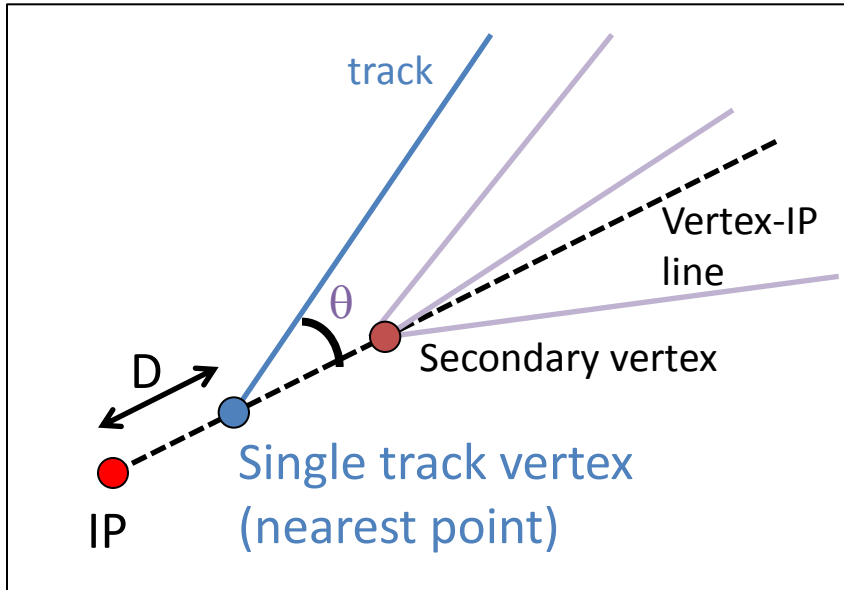
3. This results in an increased chance of correct jet separation.

This effect is particularly pronounced for final states with many b jets, e.g. Zhh

Jet/Vertex Refining Strategy

- Selection of single tracks consistent with a secondary vertex. Treat them as pseudo-vertex.
- Lots of tuning work to improve b/c separation
 - V0 reconstruction the right way → reject V0
 - K-short, Lambda0, photon conversions
 - Vertex clustering and refining, to aim
 - #vtx =2 for b-jet (also counting single track pseudo-vertex)
 - #vtx=1 for c-jet
 - #vtx=0 for uds jet
 - as much as possible
 - Unfortunately there is *leakage*, due to jet clustering mistakes → use multivariate analysis (TMVA) to cope with this. Currently multiclass BDT is implemented. **Will need to study other options.**

Single Track Selection



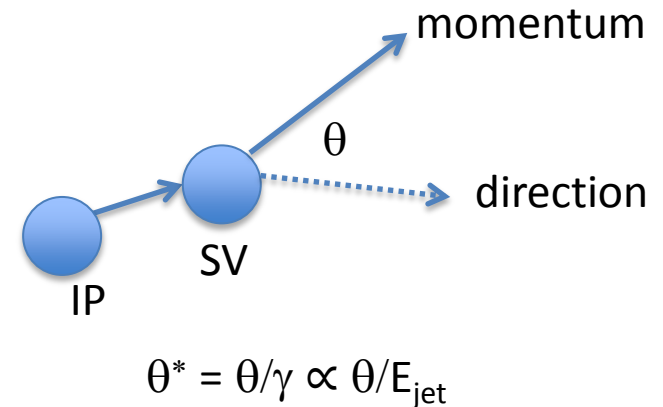
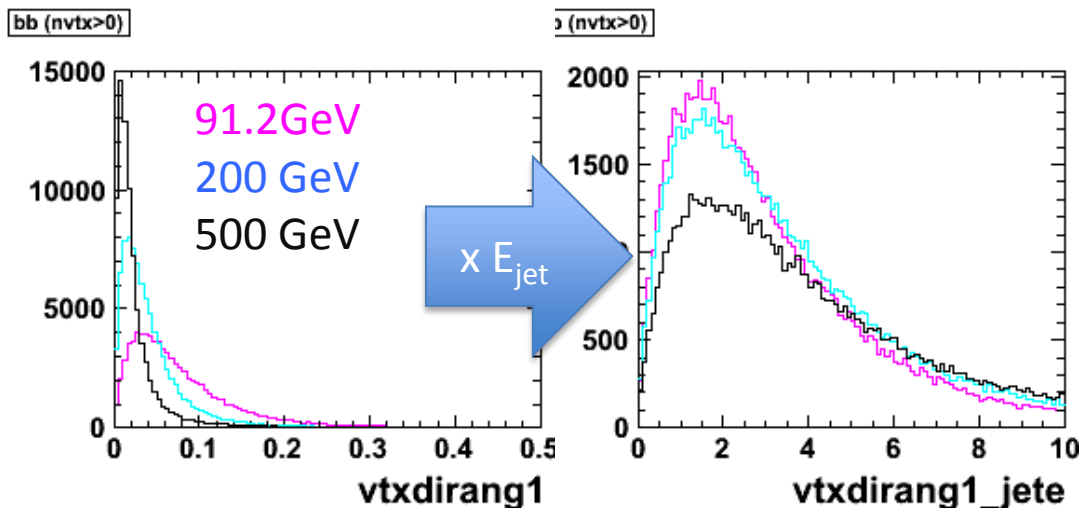
- Normal vertex finder needs > 2 tracks
-> loose many vertices
- Single track vertex can be found by using other vertex direction
- Improves b-tagging performance

Event	0 vtx	1 vtx	≥ 2 vtx
bb normal	322	1052	426(24%)
bb +single	322	459	1019(57%)
cc normal	1003	779	18(1.0%)
cc +single	1003	715	82(4.6%)

Flavor Tagging Input Variables

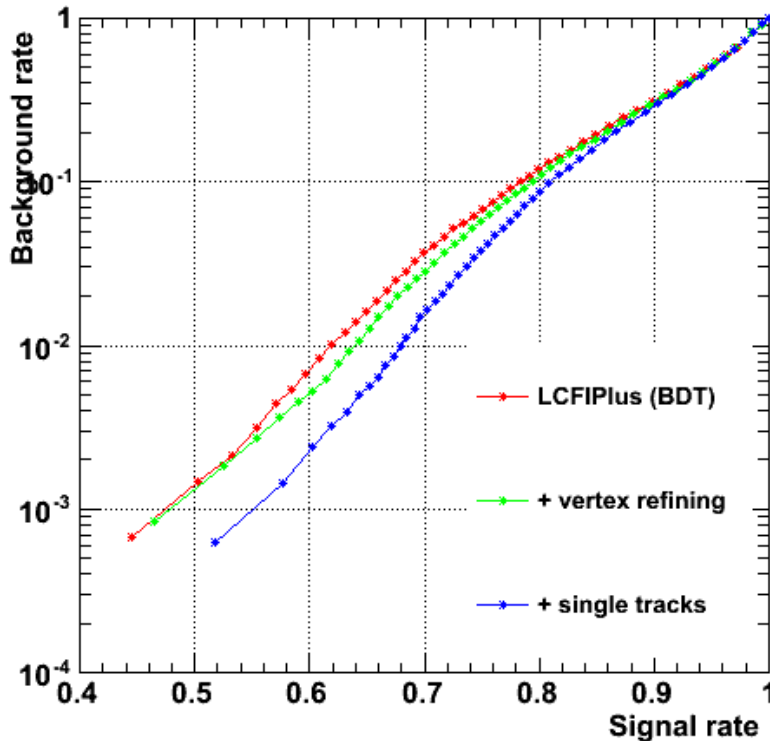
- Previously LCFIPlus used variables which were energy-dependent → degradation in real events with varying energies. Need “boost-invariant” input variables.
 - known procedure in LCFIVertex, reimplemented normalization by **jet energy**
- Do we need normalization of output? → need to be studied**

nvtx=0	trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_ jete trk2pt_ jete jprobr jprob
nvtx=1	vtxlen1_ jete vtxsig1_ jete vtxdirang1_ jete vtxmom1_ jete vtxmass1 vtxmult1 vtxmasspc vtxprob (+ above)
nvtx>=2	vtxlen2_ jete vtxsig2_ jete vtxdirang2_ jete vtxmom2_ jete vtxmass2 vtxmult2 vtxlen12_ jete vtxsig12_ jete vtxdirang12_ jete vtxmom_ jete vtxmass vtxmult (+ above)

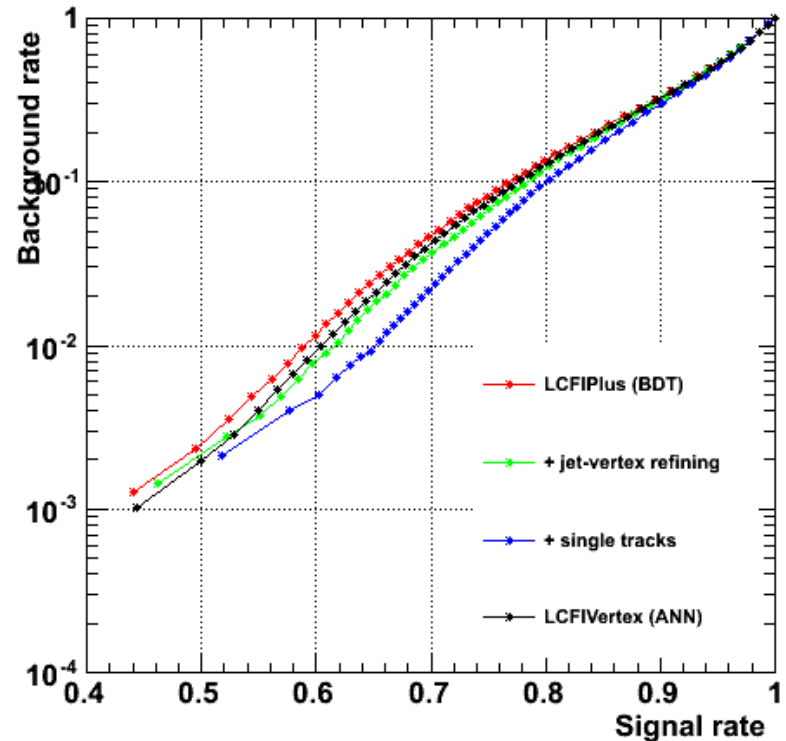


Training vs. Testing

b-tag: Z→qq, c bkg (TRAIN)



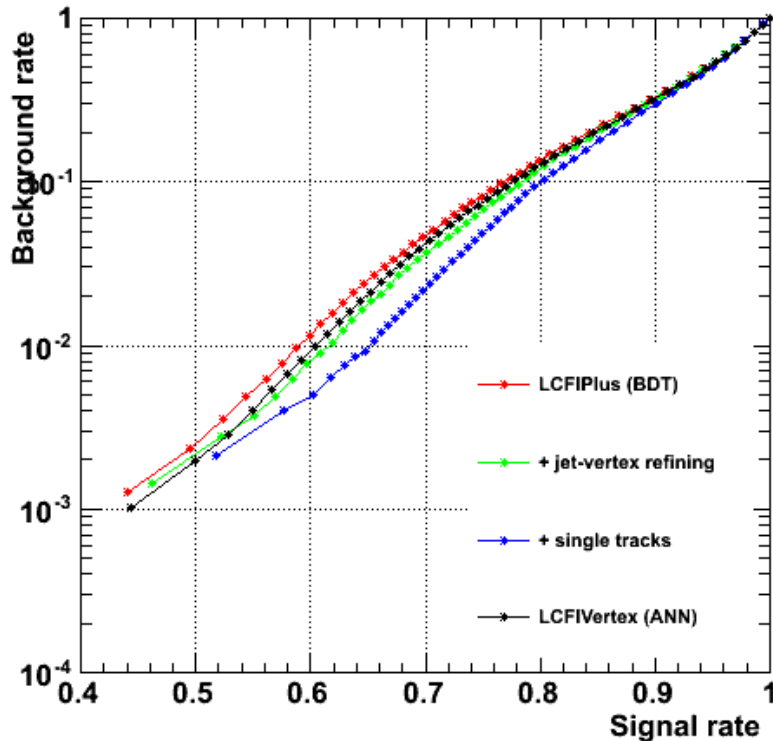
b-tag: Z→qq, c bkg (TEST)



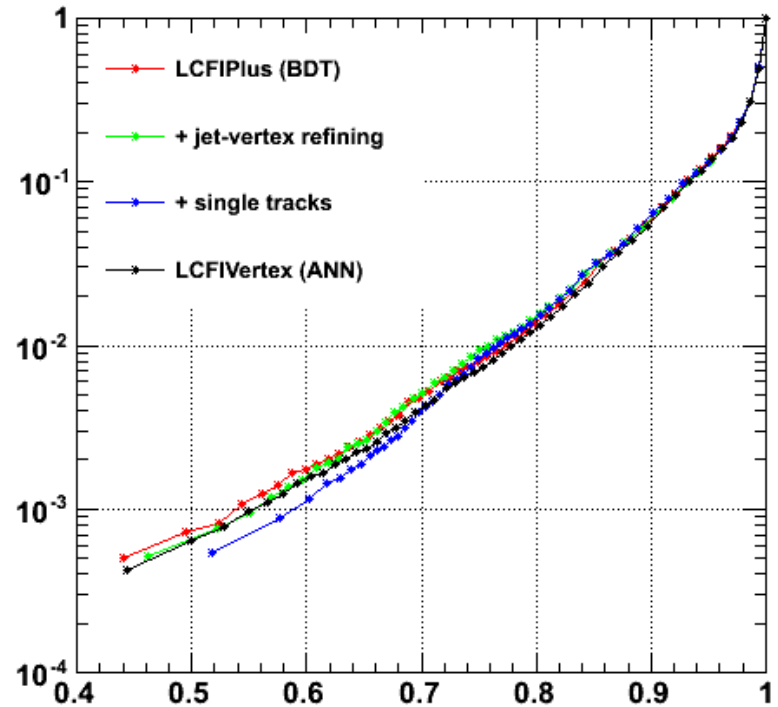
We use independent samples to evaluate the performance of the training.

b-tag vs. c-tag

b-tag: Z→qq, c bkg (TEST)



b-tag: Z→qq, uds bkg (TEST)



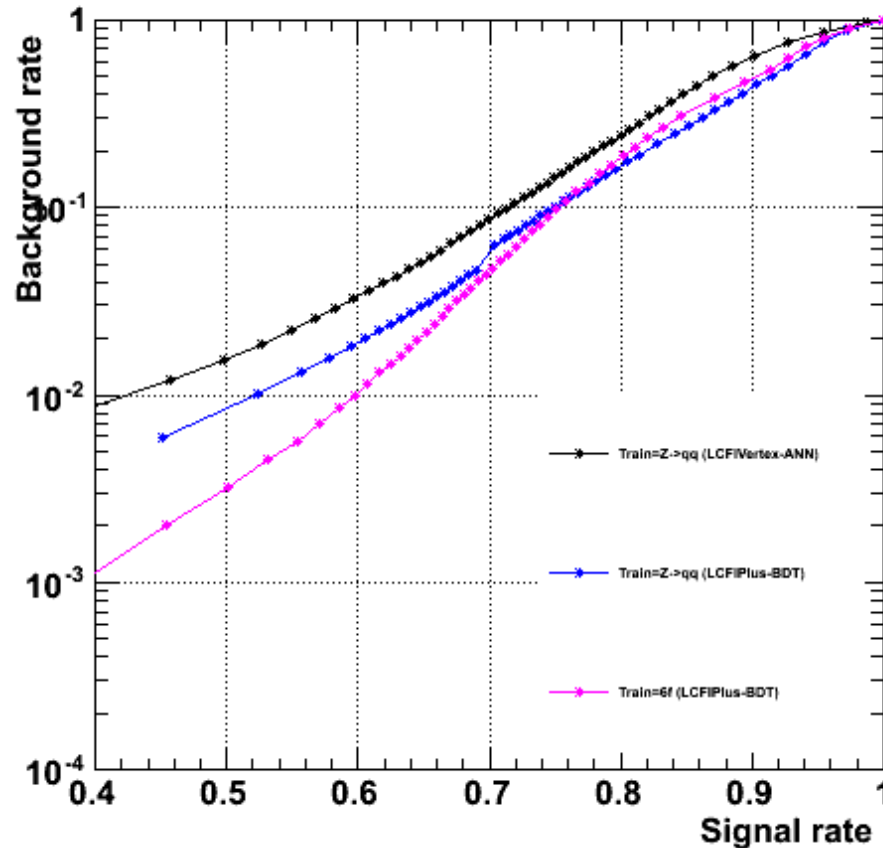
This is for Z→qq sample at E_{cm}=91.2GeV.

A small improvement is seen for b-tag and c-tag.

Improvement is highest for b-tagging in the high purity region.

Performance in 6-jet environment

b-tag: Sample=6f, c bkg, (TEST)



LCFIVertex

2 jet training on 6 jet sample

LCFIPlus

2 jet training on 6 jet sample

LCFIPlus

6 jet training on 6 jet sample

Training and testing performed using 6f samples with 6b, 6c, and 6q with $q=uds$.

Improvement over old algorithm seen in all regions because of vertex-assisted jet finding. Performance in high efficiency region still needs to be understood.

Summary and Outlook

- Software infrastructure now in place for ILD DBD production
→ have complied with technical requests, to be included into the next ilcsoft cycle (~few weeks)
- Will continue working with SiD for a smooth transition from LCFIVertex to LCFIPlus
- Physics performance not yet satisfactory. Need a better understanding of **physics** and **TMVA behavior**. In particular:
 - additional tuning of vertex/jet finding & input variables
 - alternative multivariate algorithms ?
 - normalization of output classifiers ?