350 GeV Flavour Tag Training Steps

To prepare the vertexing jobs and jobs for producing the n-tuples needed for the BDT training, one starts by making a list of the reconstruction output files made from the vvZ (350 GeV) stdhep files:

ls -1 /nfs/slac/g/lcd/ilc_data4/snowmass/ILC350/flavorTraining/sidloi3/slcio/reco/*.slcio > ilc350flav.lst

The following submits the jobs to the batch queue:

source subdstlst-tup-from-rec-job ilc350flav.lst

```
WHERE:

subdstlst-tup-from-rec-job:
cat $1 | xargs -n 1 bsub -q long -R rhel50 "echo sleep 0.1;source subflavtup-from-rec-job.sh"

subflavtup-from-rec-job.sh: export mydir=`echo $1 | awk -F '/' '{print substr($0,1,length($0)-length($NF))}'`
 export myfil=`echo $1 | awk -F '/' '{print $NF}'`
 echo "mydir = "$mydir
 echo "myfil = "$myfil
 mkdir /u/ey/homer/sidhome/lcfi/steering/tupparts/$myfil-dir
 cd /u/ey/homer/sidhome/lcfi/steering/tupparts/$myfil-dir
 rm /u/ey/homer/sidhome/lcfi/steering/tupparts/$myfil-dir/*.{slcio,log}
 ln -s $1 input-rec.slcio
 Marlin /u/ey/homer/sidhome/lcfi/steering/vertexing-basic.xml >  /u/ey/homer/sidhome/lcfi/steering/tupparts
/$myfil-dir/pretup-vtx.log
 Marlin /u/ey/homer/sidhome/lcfi/steering/makentuple-all-batch.xml >  /u/ey/homer/sidhome/lcfi/steering/tupparts
/$myfil-dir/tup.log
 rm input-dst.slcio
 ln -s output-ntpl.root $myfil-ntpl.root




The XML's accessed above are shown below:
```

```
 vertexing-basic.xml:
<marlin>

<execute>
        <processor name="VertexFinder"/>
        <processor name="MyLCIOOutputProcessor"/>
</execute>

<global>
        <parameter name="LCIOInputFiles">
input-rec.slcio
 </parameter>
        <parameter name="GearXMLFile">/u/ey/homer/sidhome/lcfi/steering/clic_sid.gear</parameter>
        <parameter name="MaxRecordNumber" value="-1" />
        <parameter name="SkipNEvents" value="0" />
        <parameter name="SupressCheck" value="false" />
        <parameter name="Verbosity" options="DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT">WARNING</parameter>
</global>

<processor name="VertexFinder" type="LcfiplusProcessor">

        <!-- run primary and secondary vertex finders -->
        <parameter name="Algorithms" type="stringVec"> PrimaryVertexFinder BuildUpVertex </parameter>
        <parameter name="ReadSubdetectorEnergies" type="int" value="0"/> <!-- true for ILD -->
        <parameter name="UpdateVertexRPDaughters" type="int" value="0"/> <!-- false for non-updative
PandoraPFOs -->
        <parameter name="PrintEventNumber" type="int" value="10"/> <!-- 0 for not printing event number, n for
printing every n events -->

        <!-- specify input collection names -->
```

```
        <parameter name="PFOCollection" type="string" value="PandoraPFOCollection" />
        <parameter name="PrimaryVertexCollectionName" type="string" value="PrimaryVertex" />
        <parameter name="BuildUpVertexCollectionName" type="string" value="BuildUpVertex" />
        <parameter name="BuildUpVertex.V0VertexCollectionName" type="string" value="BuildUpVertex_V0" />
        <parameter name="MagneticField" type="float" value="5"/>
        <parameter name="BeamSizeX" type="float" value="335E-6"/>
        <parameter name="BeamSizeY" type="float" value="2.7E-6"/>
        <parameter name="BeamSizeZ" type="float" value="0.225"/>

        <!-- parameters for primary vertex finder -->
        <parameter name="PrimaryVertexFinder.TrackMaxD0" type="double" value="20." />
        <parameter name="PrimaryVertexFinder.TrackMaxZ0" type="double" value="20." />
        <parameter name="PrimaryVertexFinder.TrackMaxInnermostHitRadius" type="double" value="20." />
        <parameter name="PrimaryVertexFinder.TrackMinVtxFtdHits" type="int" value="3" />
        <parameter name="PrimaryVertexFinder.Chi2Threshold" type="double" value="25." />
        <parameter name="PrimaryVertexFinder.UseBeamConstraint" type="int" value="1" />

        <!-- parameters for secondary vertex finder -->
        <parameter name="BuildUpVertex.TrackMaxD0" type="double" value="10." />
        <parameter name="BuildUpVertex.TrackMaxZ0" type="double" value="20." />
        <parameter name="BuildUpVertex.TrackMinPt" type="double" value="0.1" />
        <parameter name="BuildUpVertex.TrackMaxD0Err" type="double" value="0.1" />
        <parameter name="BuildUpVertex.TrackMaxZ0Err" type="double" value="0.1" />
        <parameter name="BuildUpVertex.TrackMinTpcHits" type="int" value="4" />
        <parameter name="BuildUpVertex.TrackMinFtdHits" type="int" value="3" />
        <parameter name="BuildUpVertex.TrackMinVxdHits" type="int" value="3" />
        <parameter name="BuildUpVertex.TrackMinVxdFtdHits" type="int" value="0" />
        <parameter name="BuildUpVertex.PrimaryChi2Threshold" type="double" value="25." />
        <parameter name="BuildUpVertex.SecondaryChi2Threshold" type="double" value="9." />
        <parameter name="BuildUpVertex.MassThreshold" type="double" value="10." />
        <parameter name="BuildUpVertex.MinDistFromIP" type="double" value="0.3" />
        <parameter name="BuildUpVertex.MaxChi2ForDistOrder" type="double" value="1.0" />
        <parameter name="BuildUpVertex.AssocIPTracks" type="int" value="1" />
        <parameter name="BuildUpVertex.AssocIPTracksMinDist" type="double" value="0." />
        <parameter name="BuildUpVertex.AssocIPTracksChi2RatioSecToPri" type="double" value="2.0" />
        <parameter name="BuildUpVertex.UseV0Selection" type="int" value="1" />

</processor>

<processor name="MyLCIOOutputProcessor" type="LCIOOutputProcessor">
        <parameter name="LCIOOutputFile" type="string">
input-dst.slcio
 </parameter>
        <parameter name="LCIOWriteMode" type="string" value="WRITE_NEW"/>
</processor>

</marlin>
```

```
makentuple-all-batch.xml:
<marlin>

<execute>
        <processor name="JetClustering"/>
        <processor name="MakeNtuple"/>
</execute>

<global>
        <parameter name="LCIOInputFiles">
input-dst.slcio
        </parameter>
        <parameter name="GearXMLFile">/u/ey/homer/sidhome/lcfi/steering/clic_sid.gear</parameter>
        <parameter name="MaxRecordNumber" value="-1" />
        <parameter name="SkipNEvents" value="0" />
        <parameter name="SupressCheck" value="false" />
        <parameter name="Verbosity" options="DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT">WARNING</parameter>
</global>
```

```
<processor name="JetClustering" type="LcfiplusProcessor">

        <!-- run primary and secondary vertex finders -->
        <parameter name="Algorithms" type="stringVec"> JetClustering JetVertexRefiner</parameter>

        <!-- general parameters -->
        <parameter name="PFOCollection" type="string" value="PandoraPFOCollection" /> <!-- input PFO collection
-->
        <parameter name="UseMCP" type="int" value="0" /> <!-- MC info not used -->
        <parameter name="MCPCollection" type="string" value="" /> <!-- not used -->
        <parameter name="MCPFORelation" type="string" value="" /> <!-- not used -->
        <parameter name="ReadSubdetectorEnergies" type="int" value="0"/> <!-- true for ILD -->
        <parameter name="UpdateVertexRPDaughters" type="int" value="0"/> <!-- false for non-updative
PandoraPFOs -->
        <parameter name="MagneticField" type="float" value="5"/>

        <!-- jet clustering parameters -->
        <parameter name="JetClustering.InputVertexCollectionName" type="string" value="BuildUpVertex" /> <!--
vertex collections to be used in JC -->
        <parameter name="JetClustering.OutputJetCollectionName" type="stringVec" value="VertexJets" /> <!--
output collection name, may be multiple -->
        <parameter name="JetClustering.NJetsRequested" type="intVec" value="2" /> <!-- Multiple NJets can be
specified -->

        <parameter name="JetClustering.YCut" type="doubleVec" value="0." /> <!-- specify 0 if not used -->
        <parameter name="JetClustering.UseMuonID" type="int" value="1" /> <!-- jet-muon ID for jet clustering --
>
        <parameter name="JetClustering.VertexSelectionMinimumDistance" type="double" value="0.3" /> <!-- in mm
-->
        <parameter name="JetClustering.VertexSelectionMaximumDistance" type="double" value="30." /> <!-- in mm
-->
        <parameter name="JetClustering.VertexSelectionK0MassWidth" type="double" value="0.02" /> <!-- in GeV -->
        <parameter name="JetClustering.YAddedForJetVertexVertex" type="double" value="100"/> <!-- add penalty
for combining vertices -->
        <parameter name="JetClustering.YAddedForJetLeptonVertex" type="double" value="100"/> <!-- add penalty
for combining lepton and vertex -->
        <parameter name="JetClustering.YAddedForJetLeptonLepton" type="double" value="100"/> <!-- add penalty
for combining leptons -->

        <!-- vertex refiner parameters -->
        <parameter name="JetVertexRefiner.InputJetCollectionName" type="string" value="VertexJets" />
        <parameter name="JetVertexRefiner.OutputJetCollectionName" type="string" value="RefinedJets" />
        <parameter name="JetVertexRefiner.PrimaryVertexCollectionName" type="string" value="PrimaryVertex" />
        <parameter name="JetVertexRefiner.InputVertexCollectionName" type="string" value="BuildUpVertex" />
        <parameter name="JetVertexRefiner.V0VertexCollectionName" type="string" value="BuildUpVertex_V0" />
        <parameter name="JetVertexRefiner.OutputVertexCollectionName" type="string" value="RefinedVertex" />

        <parameter name="JetVertexRefiner.MinPosSingle" type="double" value="0.3" />
        <parameter name="JetVertexRefiner.MaxPosSingle" type="double" value="30." />
        <parameter name="JetVertexRefiner.MinEnergySingle" type="double" value="1." />
        <parameter name="JetVertexRefiner.MaxAngleSingle" type="double" value="0.5" />
        <parameter name="JetVertexRefiner.MaxSeparationPerPosSingle" type="double" value="0.1" />
        <parameter name="JetVertexRefiner.mind0sigSingle" type="double" value="5." />
        <parameter name="JetVertexRefiner.minz0sigSingle" type="double" value="5." />
        <parameter name="JetVertexRefiner.OneVertexProbThreshold" type="double" value="0.001" />
        <parameter name="JetVertexRefiner.MaxCharmFlightLengthPerJetEnergy" type="double" value="0.1" />
</processor>

<processor name="MakeNtuple" type="LcfiplusProcessor">
        <parameter name="Algorithms" type="stringVec">FlavorTag MakeNtuple</parameter>

        <!-- general parameters -->
        <parameter name="PFOCollection" type="string" value="PandoraPFOCollection" /> <!-- input PFO collection
-->
        <parameter name="UseMCP" type="int" value="0" /> <!-- MC info not used -->
        <parameter name="MCPCollection" type="string" value="" /> <!-- not used -->
        <parameter name="MCPFORelation" type="string" value="" /> <!-- not used -->
        <parameter name="ReadSubdetectorEnergies" type="int" value="0"/> <!-- true for ILD -->
        <parameter name="UpdateVertexRPDaughters" type="int" value="0"/> <!-- false for non-updative
PandoraPFOs -->
```

```
        <parameter name="PrimaryVertexCollectionName" type="string" value="PrimaryVertex" />
        <parameter name="FlavorTag.JetCollectionName" type="string" value="RefinedJets" />
        <parameter name="MakeNtuple.AuxiliaryInfo" type="int" value="-1" />
        <parameter name="FlavorTag.D0ProbFileName" type="string" value="/u/ey/homer/sidhome/lcfi/steering
/d0prob_zpole.root"/>
        <parameter name="FlavorTag.Z0ProbFileName" type="string" value="/u/ey/homer/sidhome/lcfi/steering
/z0prob_zpole.root"/>

 <parameter name="MakeNtuple.OutputRootFileName" type="string" value="output-ntpl.root" />
</processor>

</marlin>
```

To do the training:

Marlin training-350.xml

where:

```
training-350.xml:

<marlin>

<execute>
        <processor name="MyLcfiplusProcessor"/>
</execute>

<global>
        <parameter name="LCIOInputFiles"> </parameter>
        <parameter name="MaxRecordNumber" value="-1" />
        <parameter name="SkipNEvents" value="0" />
        <parameter name="SupressCheck" value="false" />

        <parameter name="GearXMLFile">clic_sid.gear</parameter>

        <parameter name="Verbosity" options="DEBUG0-4,MESSAGE0-4,WARNING0-4,ERROR0-4,SILENT">WARNING</parameter>
</global>

<processor name="MyLcfiplusProcessor" type="LcfiplusProcessor">
        <parameter name="Algorithms" type="stringVec">TrainMVA</parameter>
        <parameter name="ReadSubdetectorEnergies" type="int" value="0"/> <!-- true for ILD -->
        <parameter name="PFOCollection" type="string" value="PandoraPFOCollection" />
        <parameter name="MagneticField" type="float" value="5"/>

        <parameter name="FlavorTag.WeightsDirectory" type="string" value="/u/ey/homer/sidhome/lcfi/steering
/weights-350" />
        <parameter name="FlavorTag.WeightsPrefix" type="string" value="flavwgts" />
        <parameter name="FlavorTag.BookName" type="string" value="bdt" />

        <parameter name="FlavorTag.CategoryDefinition1" type="string">nvtx==0</parameter>
        <parameter name="FlavorTag.CategoryPreselection1" type="string">trk1d0sig!=0</parameter>
        <parameter name="FlavorTag.CategoryVariables1" type="stringVec">
                trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
        </parameter>
        <parameter name="FlavorTag.CategorySpectators1" type="stringVec">
                aux nvtx
        </parameter>

        <parameter name="FlavorTag.CategoryDefinition2" type="string">nvtx==1&&nvtxall==1</parameter>
        <parameter name="FlavorTag.CategoryPreselection2" type="string">trk1d0sig!=0</parameter>
        <parameter name="FlavorTag.CategoryVariables2" type="stringVec">
                trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
                vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob
                 d0bprob d0cprob d0qprob z0bprob z0cprob z0qprob
                 trkmass
        </parameter>
        <parameter name="FlavorTag.CategorySpectators2" type="stringVec">
                aux nvtx
        </parameter>
```

```
        <parameter name="FlavorTag.CategoryDefinition3" type="string">nvtx==1&&nvtxall==2</parameter>
        <parameter name="FlavorTag.CategoryPreselection3" type="string">trk1d0sig!=0</parameter>
        <parameter name="FlavorTag.CategoryVariables3" type="stringVec">
                trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
                vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob
                1vtxprob vtxlen12all_jete vtxmassall
        </parameter>
        <parameter name="FlavorTag.CategorySpectators3" type="stringVec">
                aux nvtx
        </parameter>

        <parameter name="FlavorTag.CategoryDefinition4" type="string">nvtx>=2</parameter>
        <parameter name="FlavorTag.CategoryPreselection4" type="string">trk1d0sig!=0</parameter>
        <parameter name="FlavorTag.CategoryVariables4" type="stringVec">
                trk1d0sig trk2d0sig trk1z0sig trk2z0sig trk1pt_jete trk2pt_jete jprobr jprobz
                vtxlen1_jete vtxsig1_jete vtxdirang1_jete vtxmom1_jete vtxmass1 vtxmult1 vtxmasspc vtxprob
                vtxlen2_jete vtxsig2_jete vtxdirang2_jete vtxmom2_jete vtxmass2 vtxmult2
                vtxlen12_jete vtxsig12_jete vtxdirang12_jete vtxmom_jete vtxmass vtxmult
                1vtxprob
        </parameter>
        <parameter name="FlavorTag.CategorySpectators4" type="stringVec">
                aux nvtx
        </parameter>


        <parameter name="TrainMVA.Verbose" type="bool" value="true" />
        <parameter name="TrainMVA.BookType" type="string" value="BDT" />
        <parameter name="TrainMVA.BookOptions" type="string">
        !H:!V:NTrees=1000:BoostType=Grad:Shrinkage=0.10:UseBaggedGrad:GradBaggingFraction=0.50:nCuts=20:
NNodesMax=8
        </parameter>

        <parameter name="TrainMVA.InputRootFileB" type="string" value="tupparts/bb350ntp.root" />
        <parameter name="TrainMVA.InputRootFileC" type="string" value="tupparts/cc350ntp.root" />
        <parameter name="TrainMVA.InputRootFileO" type="string" value="tupparts/qq350ntp.root" />
        <parameter name="TrainMVA.TreeNameB" type="string" value="ntp" />
        <parameter name="TrainMVA.TreeNameC" type="string" value="ntp" />
        <parameter name="TrainMVA.TreeNameO" type="string" value="ntp" />

</processor>

</marlin>
```

Results:

```
 The training/testing event counts used were:
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : Number of training and testing events after
rescaling:
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           :
-------------------------------------------------------
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetB -- training entries            : 14418
(sum of weights: 14418)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetB -- testing entries             : 14419
(sum of weights: 14419)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetB -- training and testing entries: 28837
(sum of weights: 28837)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetC -- training entries            : 30296
(sum of weights: 30296)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetC -- testing entries             : 30297
(sum of weights: 30297)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetC -- training and testing entries: 60593
(sum of weights: 60593)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetO -- training entries            : 85074
(sum of weights: 85074)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetO -- testing entries             : 85074
```

```
(sum of weights: 85074)
[ VERBOSE "MyLcfiplusProcessor"] --- DataSetFactory           : jetO -- training and testing entries: 170148
(sum of weights: 170148)

The performance for each BDT category is shown below:
Category 0:
: : Evaluation results ranked by best signal efficiency times signal purity :
---------------------------------------------------------------------------------
: MVA Method     jetB         jetC         jetO :
---------------------------------------------------------------------------------
: bdt            0.286        0.272        0.750 :
---------------------------------------------------------------------------------

Category 1:
: : Evaluation results ranked by best signal efficiency times signal purity :
---------------------------------------------------------------------------------
: MVA Method     jetB         jetC         jetO :
---------------------------------------------------------------------------------
: bdt            0.715        0.673        0.266 :
---------------------------------------------------------------------------------

Category 2:
: : Evaluation results ranked by best signal efficiency times signal purity :
---------------------------------------------------------------------------------
: MVA Method     jetB         jetC         jetO :
---------------------------------------------------------------------------------
: bdt            0.940        0.264        0.151 :
---------------------------------------------------------------------------------

Category 3:
: : Evaluation results ranked by best signal efficiency times signal purity :
---------------------------------------------------------------------------------
: MVA Method     jetB         jetC         jetO :
---------------------------------------------------------------------------------
: bdt            0.983        0.143        0.068 :
---------------------------------------------------------------------------------

where:
[noric02] ~/sidhome/lcfi/steering $ grep CategoryDefinition training-350.xml
        <parameter name="FlavorTag.CategoryDefinition1" type="string">nvtx==0</parameter>
        <parameter name="FlavorTag.CategoryDefinition2" type="string">nvtx==1&&nvtxall==1</parameter>
        <parameter name="FlavorTag.CategoryDefinition3" type="string">nvtx==1&&nvtxall==2</parameter>
        <parameter name="FlavorTag.CategoryDefinition4" type="string">nvtx>=2</parameter>

The training output files are in:
/nfs/slac/g/lcd/mc/prj/users/homer/lcfi/steering/weights-350
total 112192
-rw-r--r-- 1 homer ey  5711212 Jun 30 21:39 flavwgts_c0_bdt.weights.xml
-rw-r--r-- 1 homer ey   522220 Jun 30 21:39 flavwgts_c0_bdt.class.C
-rw-r--r-- 1 homer ey 31472363 Jun 30 22:11 flavwgts_c0.root
-rw-r--r-- 1 homer ey  5784088 Jun 30 22:28 flavwgts_c1_bdt.weights.xml
-rw-r--r-- 1 homer ey   529682 Jun 30 22:28 flavwgts_c1_bdt.class.C
-rw-r--r-- 1 homer ey 20522889 Jun 30 22:39 flavwgts_c1.root
-rw-r--r-- 1 homer ey  5750585 Jun 30 22:46 flavwgts_c2_bdt.weights.xml
-rw-r--r-- 1 homer ey   527296 Jun 30 22:46 flavwgts_c2_bdt.class.C
-rw-r--r-- 1 homer ey 26826174 Jun 30 22:50 flavwgts_c2.root
-rw-r--r-- 1 homer ey  5660157 Jun 30 22:58 flavwgts_c3_bdt.weights.xml
-rw-r--r-- 1 homer ey   523184 Jun 30 22:58 flavwgts_c3_bdt.class.C
-rw-r--r-- 1 homer ey 10526460 Jun 30 23:02 flavwgts_c3.root
```