Easy e^+e^- event analysis with LCIO & mini-DST

Jenny List (DESY), Frank Gaede (DESY),
Mikael Berggren (DESY), Shin-ichi Kawada (DESY)

2020/July/7 - 8

Preparatory Joint Sessions on
"Open Questions and New Ideas"

mini-DST project: Introduction

- Triggered by Snowmass, but also good for beginners.
- We want to involve many newcomers (exp + theo) to the ILC.
- Our approach: provide MC data in a high-level format
  - e.g.: isolated leptons, photons, jets ...
  - directly readable in ROOT
  - can be filled from full or fast simulation
  - allows seamless transition to full output format of the ILD/SiD full simulations

Contents of mini-DST

- Particle flow object (reconstructed particle)
  - including event shape variables, proton/kaon/pion ID, particles reconstructed with forward detectors
- MC truth information
- Links between reconstructed particle and MC truth
- Isolated electrons/muons/taus/photons
- Jets (N = 2, 3, 4, 5, 6)
  - Durham algorithm applied to remained particle flow objects
- Flavor tagging
  - Sophisticated MVA approach, giving three output classifications (b-, c-, other-tag)

Various Types of mini-DST

- Since mini-DST is just a data format, we can create various types of mini-DST.
  - ILD-mini-DST: based on ILD full simulation (can also be done by SiD full simulation)
  - SGV-mini-DST: based on SGV fast simulation (SGV can model either ILD or SiD)
  - Delphes-mini-DST: based on MC samples created with Delphes
    - “Generic ILC detector”: details can be found at “offline flash talks”

Generic ILC detector model for **Delphes**

Fast simulation framework, taking into account detector **acceptance**, **resolution**, reconstruction **efficiency** and providing basic event reconstruction.

No technical details are taken into account

➔ generic ILC detector model

based on earlier experience with ILD and SiD modeling

Main features:

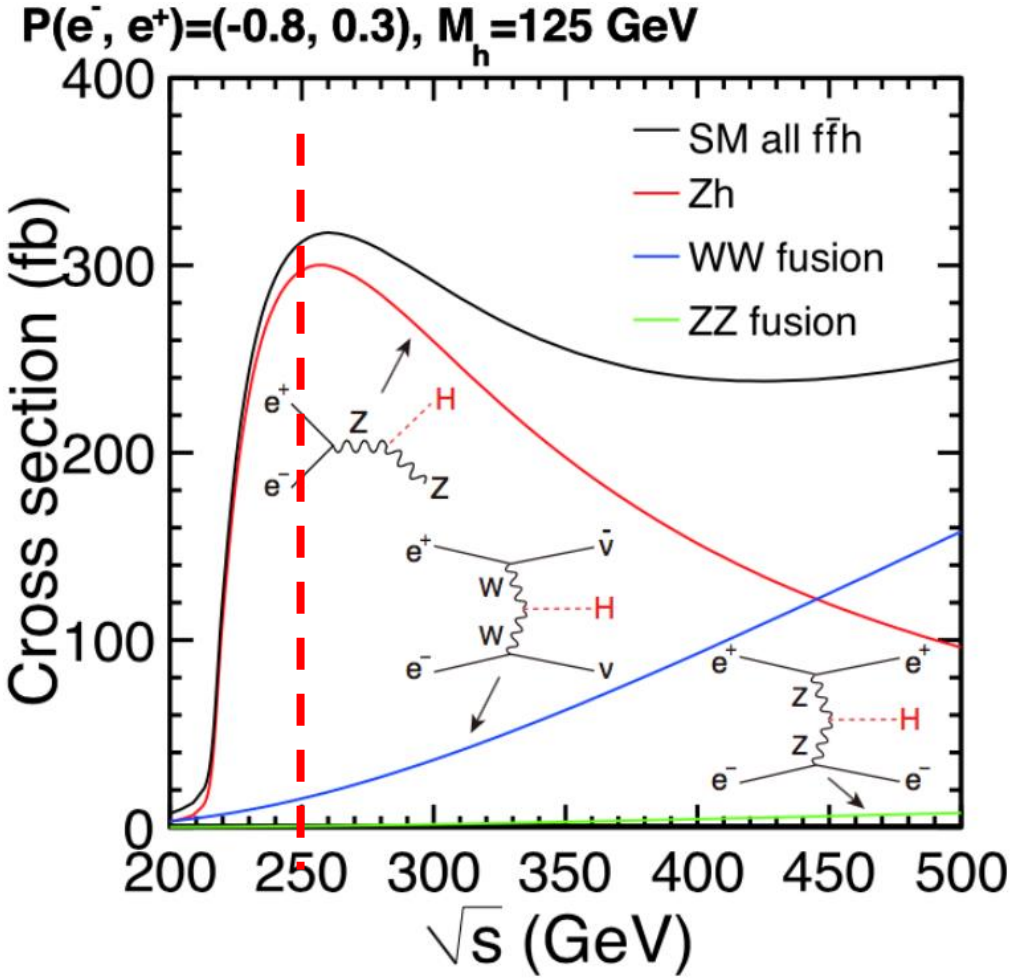
- full detector coverage incl. forward and beam calorimeters
- particle flow reconstruction up to $|\eta| \leq 4$, exclusive jet clustering with N=2...6
- including multiple options for b- and c-tagging
- events can be stored in LCIO format (mini-DST compatible)

Usage of mini-DST

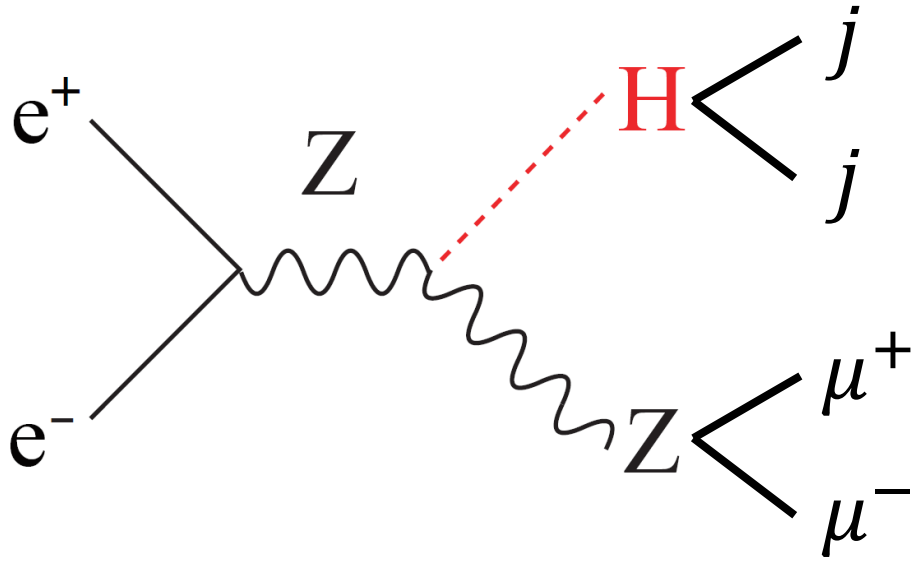
- We are also developing a simple macro to create histograms using mini-DST file.
  - Require ROOT environment, and LCIO (Linear Collider I/O) library which will also be provided
  - You don’t need to install entire iLCSoft to read mini-DST and create your histogram!

```
/*
put this into your .rootlogon.C file

{
 gInterpreter->AddIncludePath("$LCIO");
 gSystem->Load("${LCIO}/lib/liblcio.so");
 gSystem->Load("${LCIO}/lib/liblcioDict.so");
}

for the LCIO API documentation see:
http://lcio.desy.de/v02-09/doc/doxygen_api/html/index.html
*/
```

Example Study: Recoil mass at ILC250



$$e^+e^- \to ZH \to (\mu^+\mu^-)(jj)$$

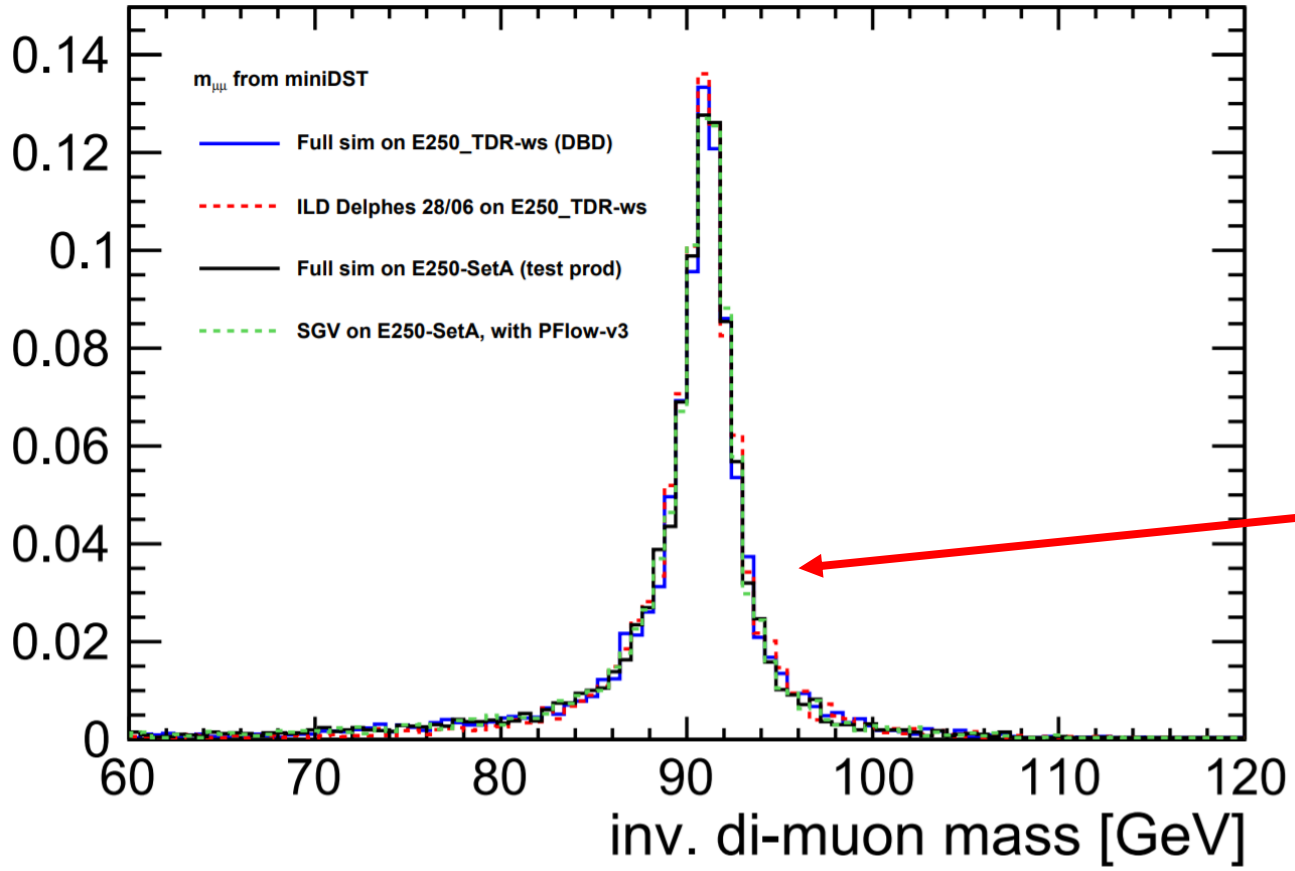


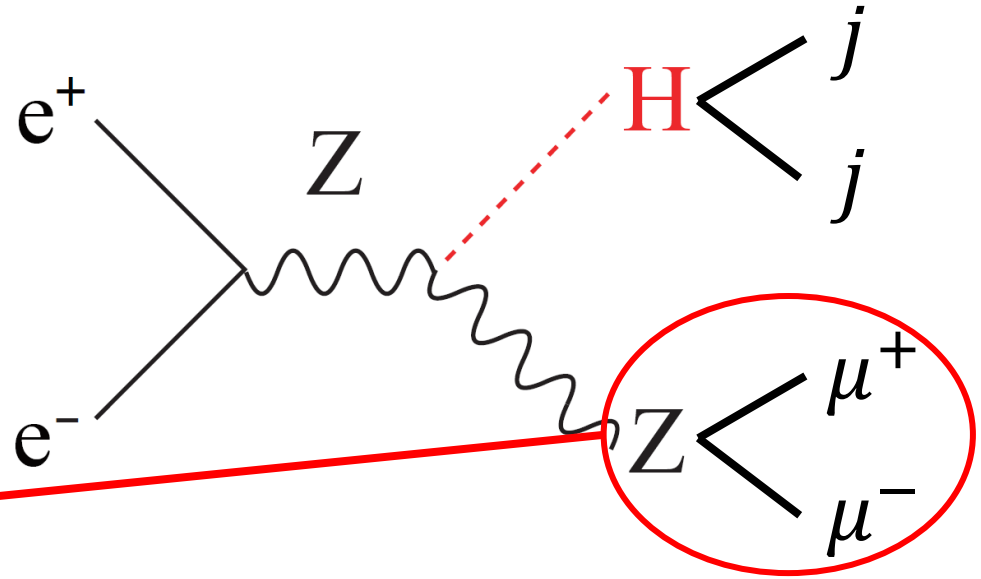
$M_H^2 = \left(p_{\text{initial}} - \left(p_{\mu^+} + p_{\mu^-}\right)\right)^2$ (p: four-momentum)

- known initial state
- recoil technique: only need to measure muons

Example Study: Recoil mass at ILC250

$$e^+e^- \to ZH \to (\mu^+\mu^-)(jj)$$





plots with various types of mini-DST

ILD-mini-DST (old)

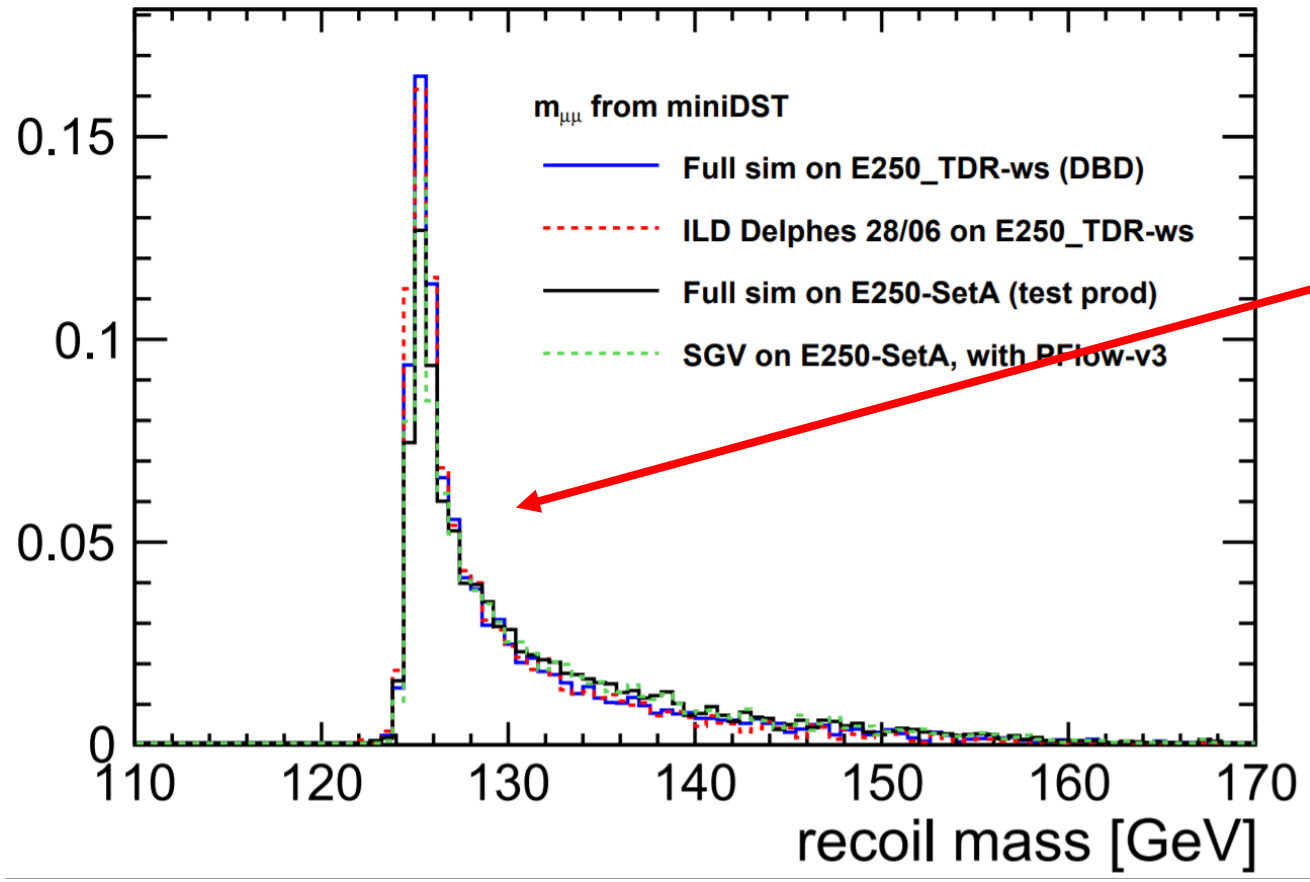
ILD-mini-DST (new)

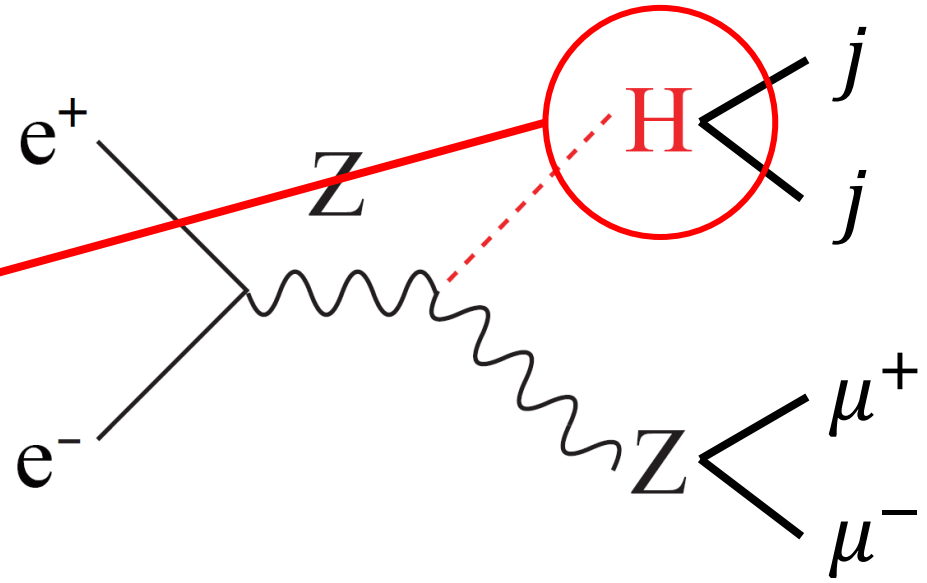
SGV-mini-DST

Delphes-mini-DST

Example Study: Recoil mass at ILC250

$$e^+e^- \rightarrow ZH \rightarrow (\mu^+\mu^-)(jj)$$





plots with various types of mini-DST

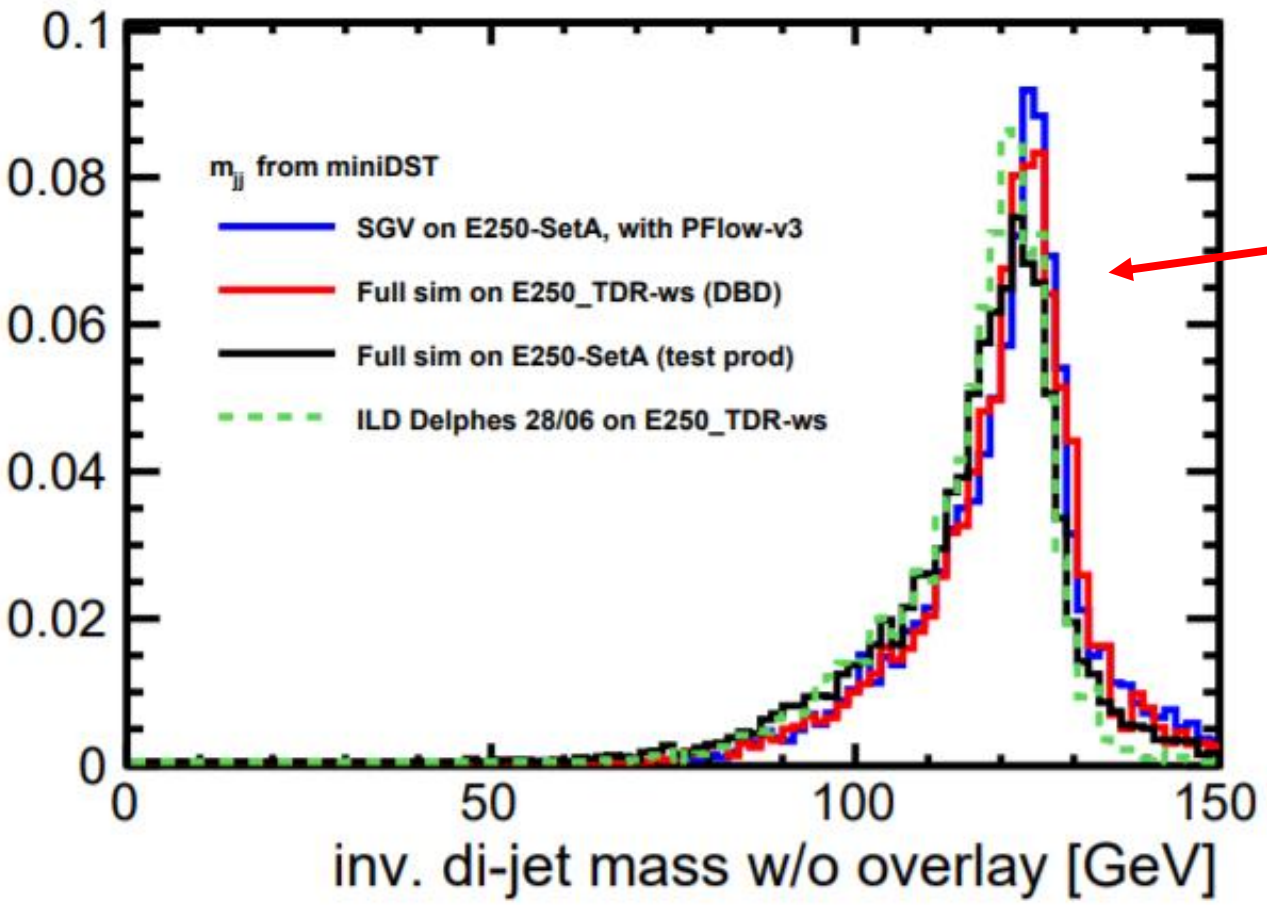
ILD-mini-DST (old)

ILD-mini-DST (new)

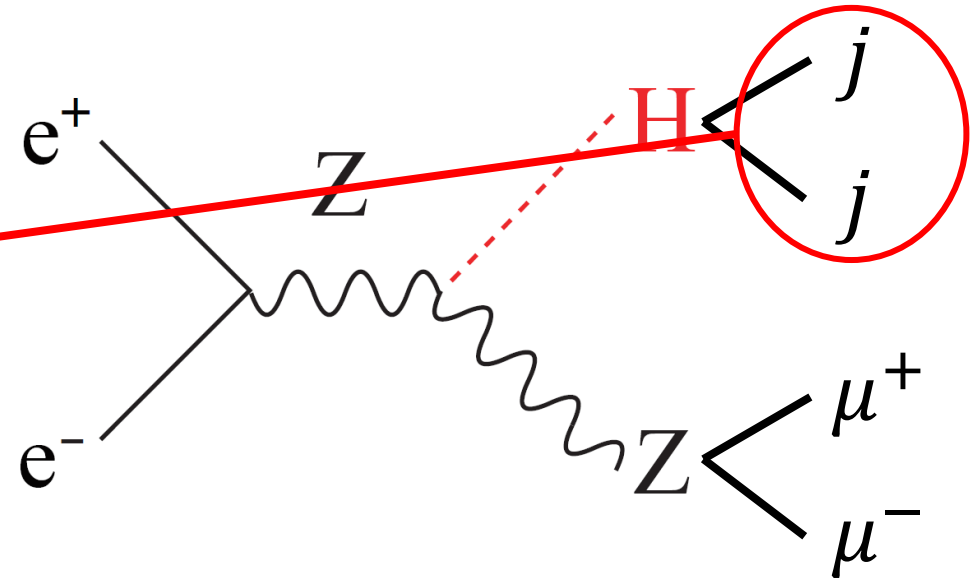
SGV-mini-DST

Delphes-mini-DST

Example Study: Recoil mass at ILC250



$$e^+e^- \to ZH \to (\mu^+\mu^-)(jj)$$



plots with various types of mini-DST

ILD-mini-DST (old)

ILD-mini-DST (new)

SGV-mini-DST

Delphes-mini-DST

Simple Macro

- The plots in previous pages were made by the same macro.
  - Just require ROOT, LCIO library, and mini-DST file
  - Usage of mini-DST format enables a seamless transition among Delphes, SGV, and full simulation

```
//----------  the actual event processing

LCIterator<ReconstructedParticle> jets( evt, "Jets" ) ;
LCIterator<ReconstructedParticle> muons( evt, "Muons" ) ;

if( jets.size() != 2)
  continue;

if( muons.size() != 2)
  continue;

auto mu1 = muons.next();
auto mu2 = muons.next();
hmuonmass->Fill( inv_mass( mu1, mu2) ) ;

auto j1 = jets.next();
auto j2 = jets.next();
hjetmass->Fill( inv_mass( j1, j2) ) ;


// the recoil mass
const auto& vm1 = v4(mu1) ;
const auto& vm2 = v4(mu2) ;
TLorentzVector ecms(0.,0.,0.,250.) ;
TLorentzVector recoil = ecms - ( vm1 + vm2 ) ;
hrecoilm->Fill( recoil.M() ) ;
```

Time Scale

- Hopefully ~ end of July
  - SGV-mini-DST of the generator-level samples including usage examples
- During ~ autumn
  - ILD-mini-DST of new 250 GeV samples
  - To be decided for other center-of-mass energies

Summary

- The mini-DST project has started, triggered by Snowmass, to reduce the complexity for newcomers and beginners.
- Contains output of high-level reconstruction (isolated leptons, photons, jets), particle flow objects, MC truth with relations to reconstructed objects.
- Enables seamless transfer of analysis code among Delphes, SGV and full simulation.
- Not necessary to install iLCSoft to use mini-DST.
- More developments are work in progress.

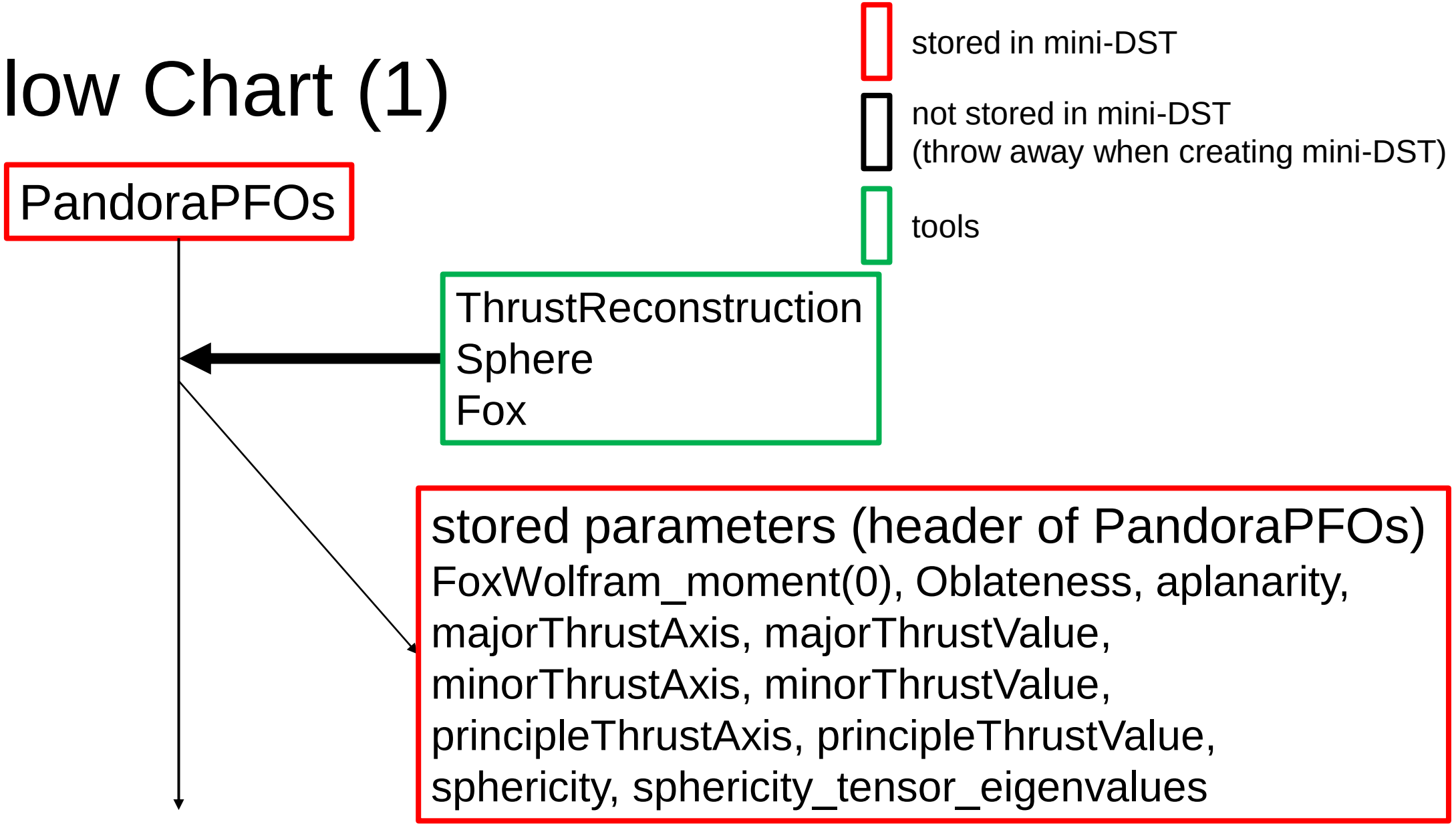
BACKUP

Red: stored in mini-DST
Green: tools

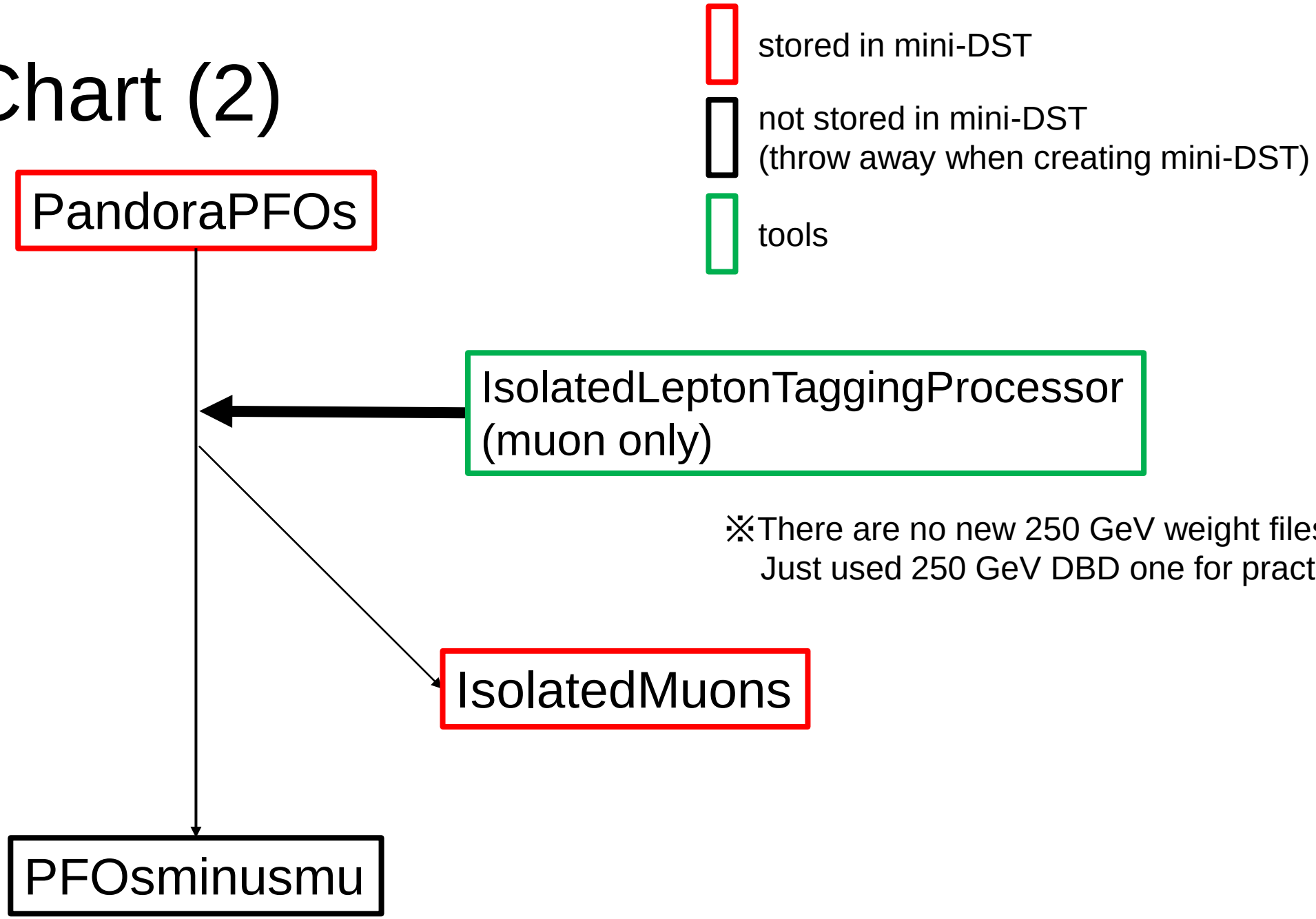
Details of v0 mini-DST

- In v0 mini-DST, the following collections are kept from original DST.
  - PandoraPFOs, BCalRecoParticle (not merged, not simplified anything)
  - MCParticle
  - PrimaryVertex, PrimaryVertex_RP
  - RecoMCTruthLink, MCTruthRecoLink (kept full relation)
- The following collections are added to the mini-DST file.
  - event shape variables (used ThrustReconstruction, Sphere, Fox): stored at the header of PandoraPFOs
  - IsolatedMuons, IsolatedElectrons (used IsolatedLeptonTagging, not tuned)
  - IsolatedTaus (used TaJetClustering, not tuned)
  - RefinedNJets (N = 2, 3, 4, 5, 6) (used LCFIPlus: JetClustering, JetVertexRefiner, FlavorTag, not tuned)
  - The ErrorFlow is applied to RefinedNJets to calculate covariance matrix for jets.

Flow Chart (1)

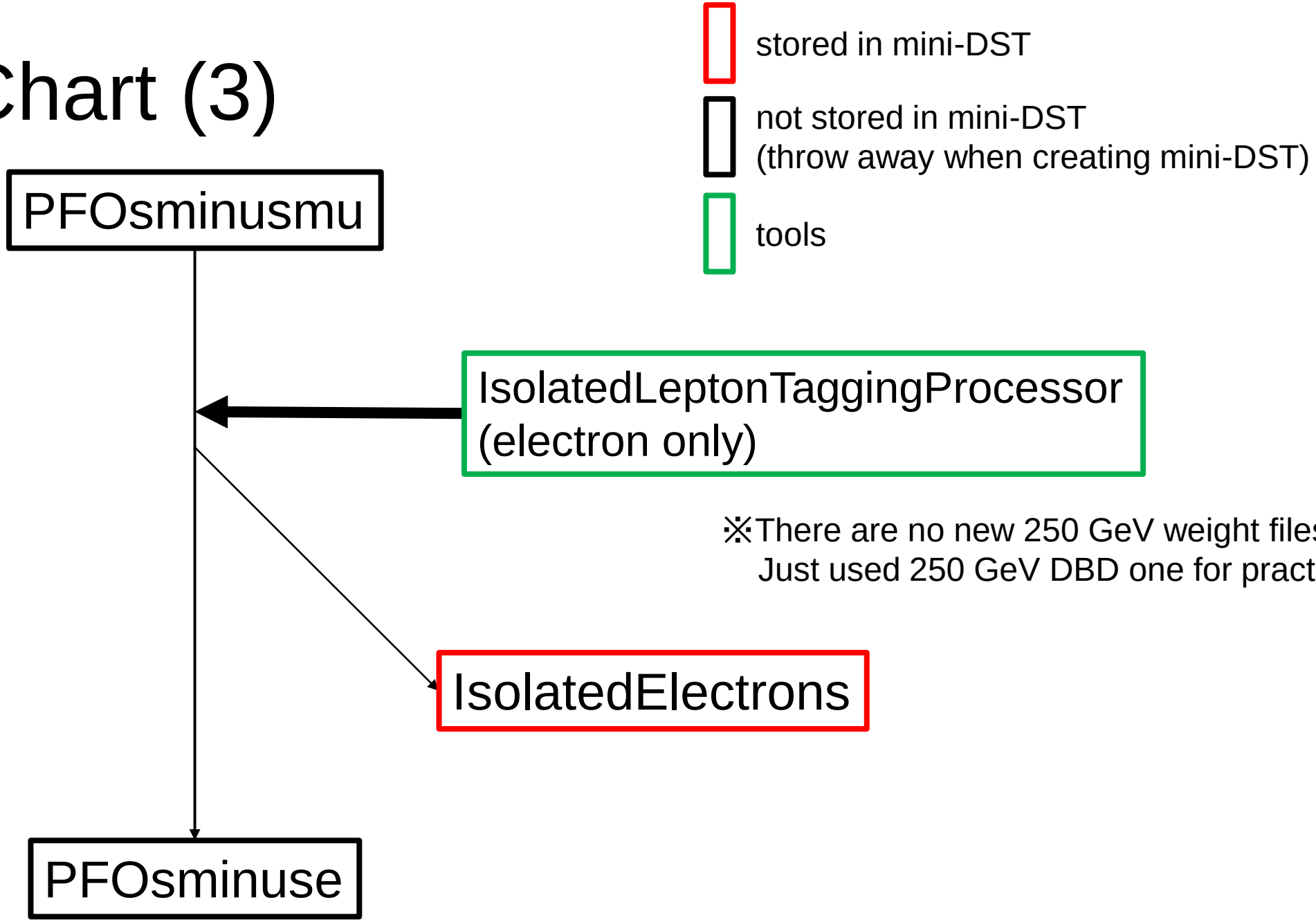

stored in mini-DST
not stored in mini-DST
(throw away when creating mini-DST)
tools
PandoraPFOs
ThrustReconstruction
Sphere
Fox
stored parameters (header of PandoraPFOs)
FoxWolfram_moment(0), Oblateness, aplanarity,
majorThrustAxis, majorThrustValue,
minorThrustAxis, minorThrustValue,
principleThrustAxis, principleThrustValue,
sphericity, sphericity_tensor_eigenvalues

Flow Chart (2)



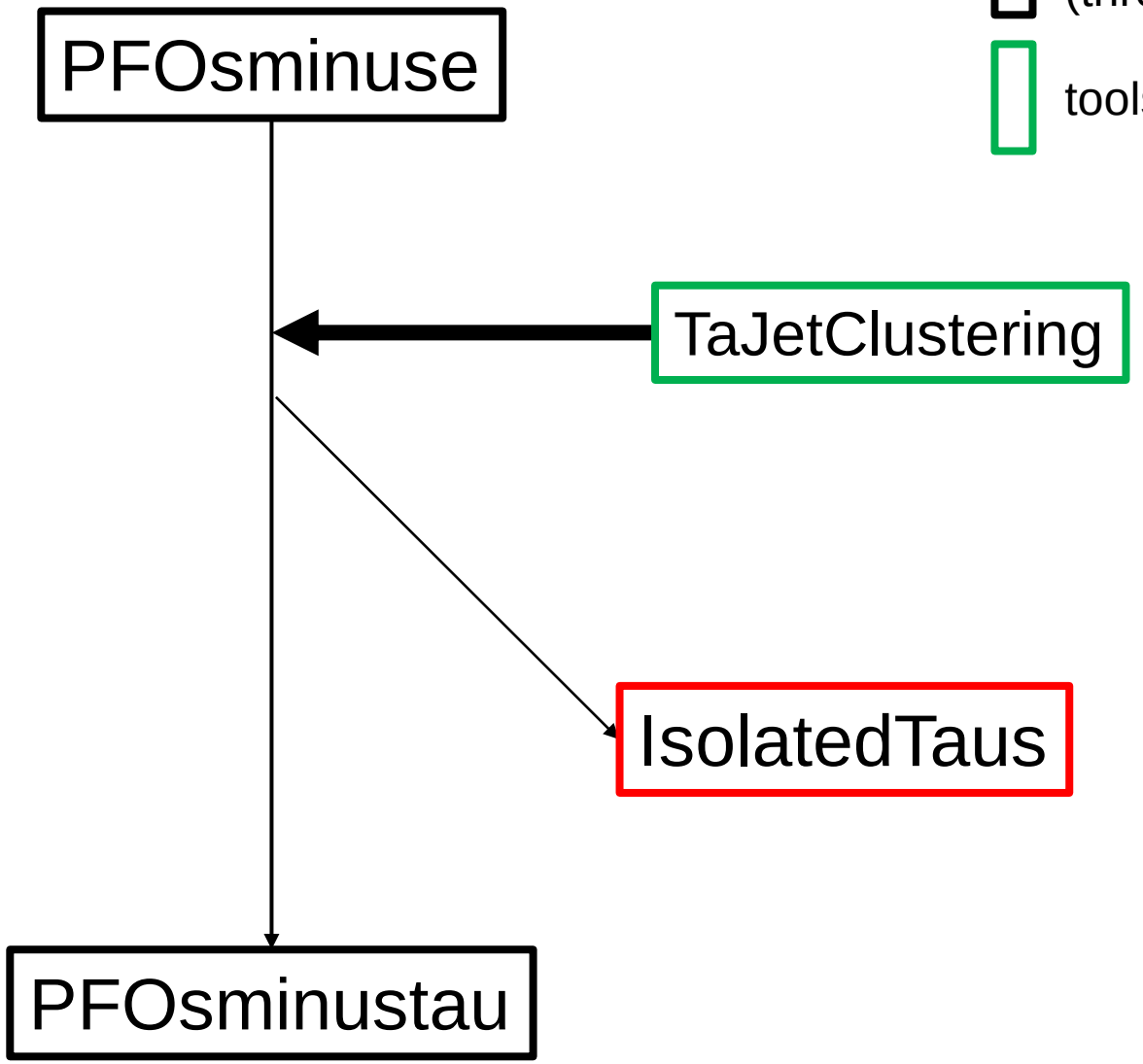
※There are no new 250 GeV weight files.
Just used 250 GeV DBD one for practice.

Flow Chart (3)



※There are no new 250 GeV weight files.
Just used 250 GeV DBD one for practice.

Flow Chart (4)



stored in mini-DST

not stored in mini-DST
(throw away when creating mini-DST)

tools

※I realized that the impact parameters are not used in TaJetClustering now.
“following cuts are found not useful now, so commented out for simplicity”
following cuts = impact parameter cuts and lepton ID cuts

Flow Chart (5)

stored in mini-DST

not stored in mini-DST
(throw away when creating mini-DST)

tools

PFOsminustau

LCFIPlus
(JetClustering, JetVertexRefiner, FlavorTag)

RefinedJets

VertexJets
RefinedJets_rel
RefinedJets_vtx
RefinedJets_vtx_RP
RefinedVertex
RefinedVertex_RP

※There are no new 250 GeV weight files.
Just used 500 GeV one for practice.

Final outcome is RefinedJets
ErrorFlow is applied to RefinedJets
to calculate covariance matrix of jets

File size / event

- Checked with physics test samples ($e^+e^- \to \mu^+\mu^- h$, $e^+e^- \to q\bar{q}h$ w/ $h \to ZZ^*$, $e^+e^- \to f\bar{f}$, $e^+e^- \to e\nu W$ w/ $W \to qq'$, $e^+e^- \to W^+W^-$ w/ semileptonic decay)

- Original DST: 50 - 100 kB / event
- mini-DST: 30 - 50 kB / event (relatively ~40 - 60% smaller)
- The case without MC truth information: 10 - 25 kB / event
  - The last case is for the test of file size, but still useful if you are not interested in MC truth.

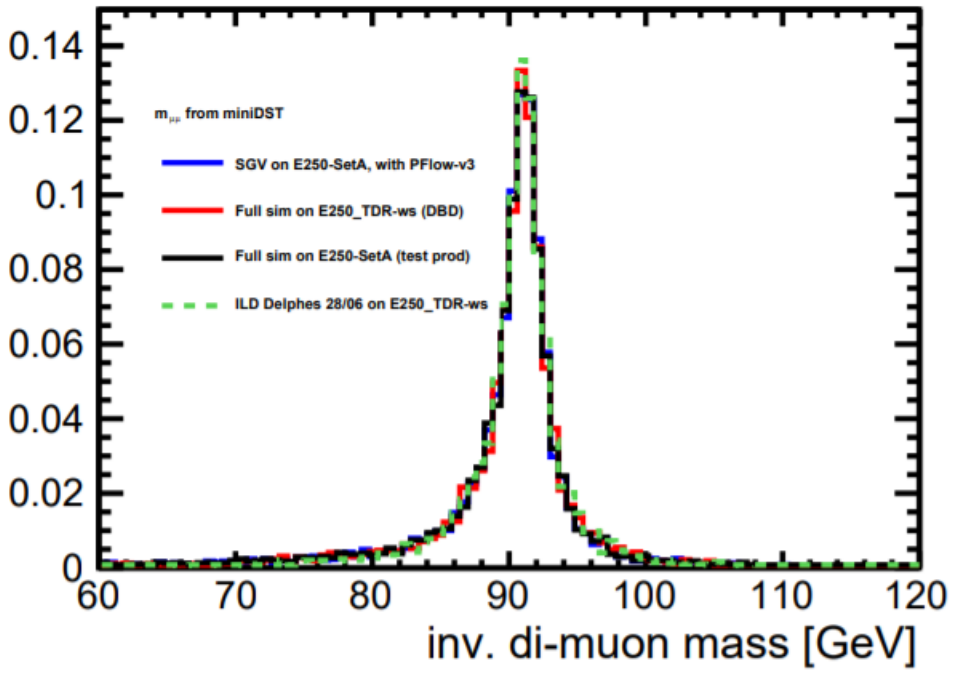
Removed from original DST (CUTSET)

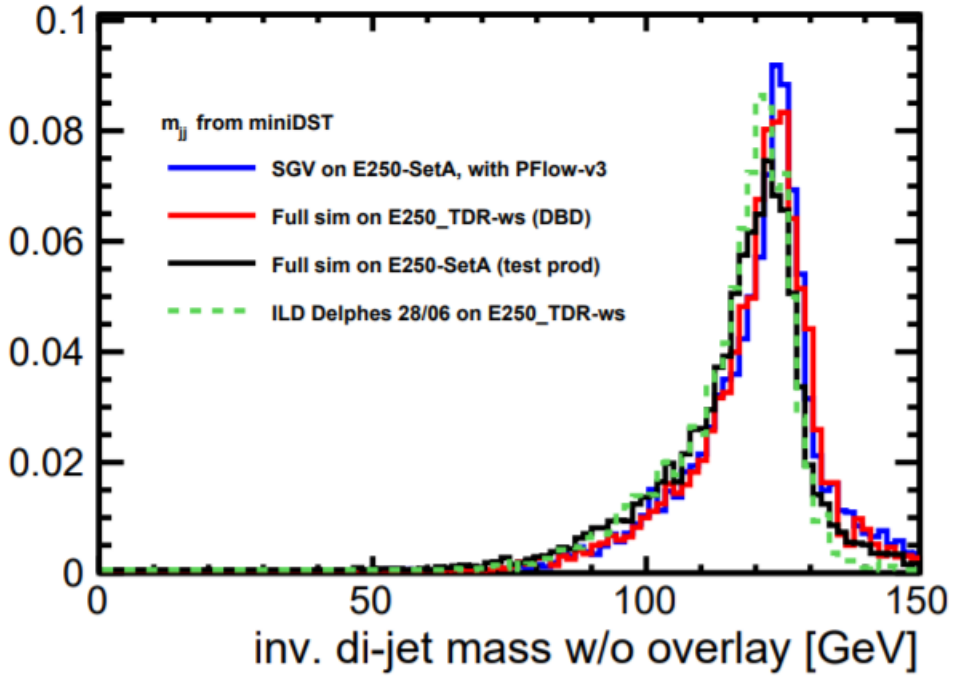
- Collection Types
  - Track
  - Cluster
- Collection Names
  - ClusterMCTruthLink
  - MCTruthClusterLink
  - MCTruthMarlinTrkTracksLink
  - MarlinTrkTracksMCTruthLink
  - DistilledPFOs
  - GammaGammaCandidateEtaPrimes
  - GammaGammaCandidateEtas
  - GammaGammaCandidatePi0s
  - GammaGammaParticles
  - V0RecoParticles
  - V0Vertices
  - ProngRecoParticles
  - ProngVertices
  - KinkRecoParticles
  - KinkVertices
  - SplitRecoParticles
  - SplitVertices
  - BuildUpVertex_RP
  - BuildUpVertex
  - BuildUpVertex_V0_RP
  - BuildUpVertex_V0

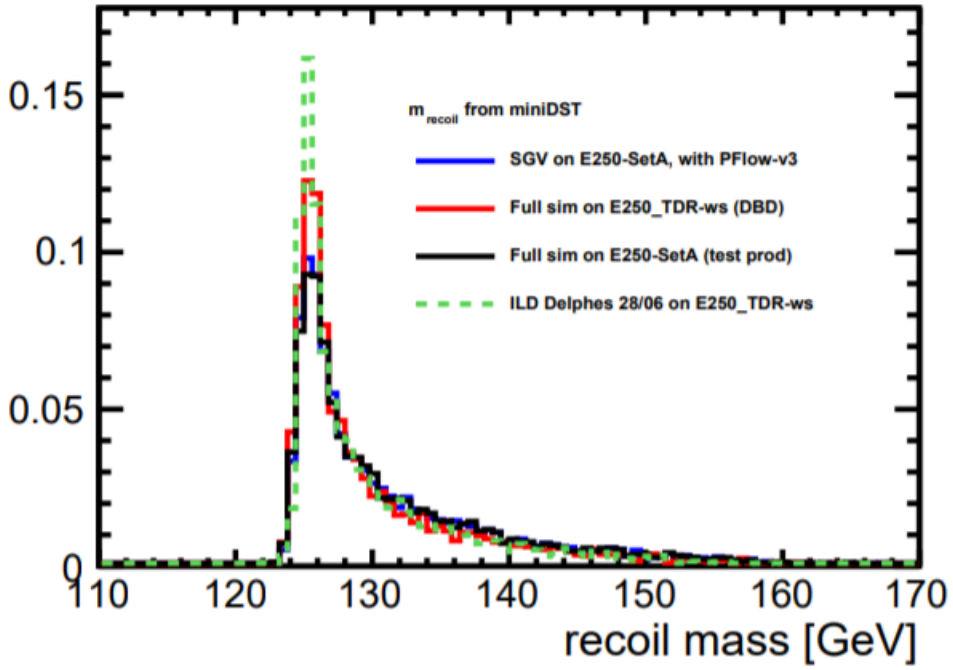
File size (measured with “wc -c filename”)

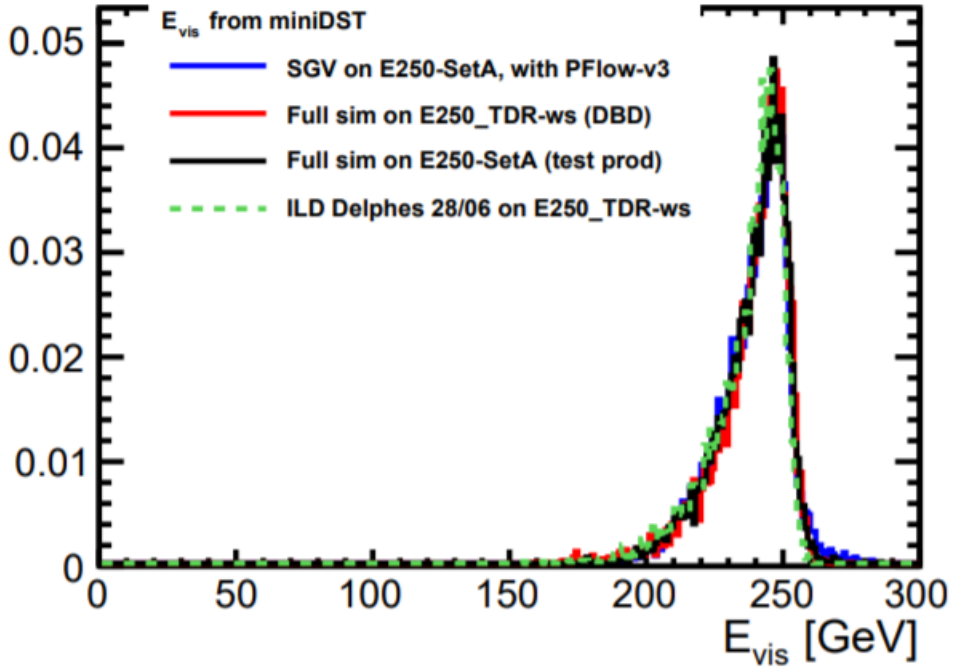
		e2e2h									qqh_zz		2f_z_l	4f_ww_sl	4f_sw_sl
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
	events	12500	12500	12500	11000	10500	10500	10500	10500	8000	5500	4500	10000	10000	10000
no cut	no cut (MB)	663	663	660	584	761	762	762	764	584	680	551	91	595	603
	no cut / event (MB)	0.053	0.053	0.0528	0.0531	0.0725	0.0726	0.0726	0.0728	0.073	0.1236	0.1224	0.009	0.0595	0.0603
case1	remove CUTSET	267	267	265	234	302	302	303	303	232	265	214	50	248	256
	(MB) / event	0.0214	0.0214	0.0212	0.0213	0.0288	0.0288	0.0289	0.0289	0.029	0.0482	0.0476	0.005	0.0248	0.0256
case2	case1 + following event shape numbers IsolatedM/E/T JCandFT (N=2,3,4,5,6)	285	285	284	251	319	319	319	320	245	275	223	56	264	271
	(MB) / event	0.0228	0.0228	0.0227	0.0228	0.0304	0.0304	0.0304	0.0305	0.0306	0.05	0.0496	0.006	0.0264	0.0271
case3	only have event shape numbers PandoraPFOs (w/ shape) IsolatedM/E/T JCandFT (N=2,3,4,5,6) PrimaryVertex PrimaryVertex_RP all others dropped	162	162	162	143	173	173	173	173	132	137	111	36	139	139
	(MB) / event	0.013	0.013	0.013	0.013	0.0165	0.0165	0.0165	0.0165	0.0165	0.0249	0.0247	0.004	0.0139	0.0139

※case3: test for file size, and useful sets if you only interested in final ReconstructedParticle/Vertex collections.
※I could not remove PandoraPFOs due to runtime error. But PandoraPFOs will be stored in mini-DST.


$m_{\mu\mu}$ from miniDST
SGV on E250-SetA, with PFlow-v3
Full sim on E250_TDR-ws (DBD)
Full sim on E250-SetA (test prod)
ILD Delphes 28/06 on E250_TDR-ws
inv. di-muon mass [GeV]


m_{jj} from miniDST
SGV on E250-SetA, with PFlow-v3
Full sim on E250_TDR-ws (DBD)
Full sim on E250-SetA (test prod)
ILD Delphes 28/06 on E250_TDR-ws
inv. di-jet mass w/o overlay [GeV]


m_{recoil} from miniDST
SGV on E250-SetA, with PFlow-v3
Full sim on E250_TDR-ws (DBD)
Full sim on E250-SetA (test prod)
ILD Delphes 28/06 on E250_TDR-ws
recoil mass [GeV]


E_{vis} from miniDST
SGV on E250-SetA, with PFlow-v3
Full sim on E250_TDR-ws (DBD)
Full sim on E250-SetA (test prod)
ILD Delphes 28/06 on E250_TDR-ws
E_{vis} [GeV]