Table of Contents

Full Hadron Analysis Framework

Prerequisites

You should use root version 5.34 or above. The input format for the analysis are the FHD3PDs which are a slimmed/skimmed version of the official TopD3PDs. You do NOT NEED RootCore or TopRootCore to run the code. The data and MC files are located on the grid and also replicated to MPPMU_LOCALGROUPDISK in Garching. It is recommended to run the analysis code on the machines at RZG in Garching (mppuiN.t2.rzg.mpg.de, N=1,2,3).

Getting the code

The analysis code is located in the svn area:

```
svn+ssh://svn.cern.ch/reps/atlasinst/Institutes/MPI/HEC/analysis/FullHadronicTopAnalysis
```

You can either check out the trunk version

```
svn co svn+ssh://svn.cern.ch/reps/atlasinst/Institutes/MPI/HEC/analysis/FullHadronicTopAnalysis/t
```

or a tagged version (**RECOMMENDED**)

```
svn co svn+ssh://svn.cern.ch/reps/atlasinst/Institutes/MPI/HEC/analysis/FullHadronicTopAnalysis/t
```

The latest tagged version is normally the one to be used. If in doubt, contact us. You can browse the trunk and tag versions of the code here: FullHadronicTopAnalysis on SVN

Compiling the code

To compile the code setup root (once in a new shell), cd into the where you installed the package (normally cd FullHadronicTopAnalysis if you are still in the dir where you issued the svn command) and type

```
cd cmt; make
```

If everything went fine there will be the executable `FHAnalysisExe.exe` in the base dir (i.e. cd .. to run it).

Running the code

There are three main steps involved when running the local analysis. By simply typing `./FHAnalysisExe.exe` you can run the code, but note that by default *all* steps are turned off; you can specify which steps you would like turned on by using two of several input parameters. These input parameters are used to steer the code (this list might get larger) and avoid having to re-compile just by turning options on or off. You can type

```
./FHAnalysisExe.exe -usage
```

to see the list:

- **Which steps to run**: options specifying which of the three main steps you wish to run (again note that these steps are all off (0) by default):
    - **-runAll**: runs **all** steps of the analysis (this is how it should be run!), so turns on several of the separate steps below and runs in the correct order **[0 or 1, default = 0]**
    - **-runMCSignalMassVariation**: runs over MC FHD3PD fast sim samples for 7 different generator mtop values. Output is txt file with slopes and intercepts of reconstructed R32 shape parameters vs generator top mass which is used later on. Also saved is a root file with these initial fits, covariance matrices, etc. for the signal shapes (global fit is then performed in final step). (*Required for full analysis*). **[0 or 1, default = 0]**
    - **-runAnalysisInitial**: runs over all events in MC and data from the FHD3PDs input ntuples, does basic event selection (trigger, jet isolation cut, minimum number of required jets, etc.) as well as top reconstruction (the MinChiSquared is used by default), but no cuts are done - just to have the top candidates in place and to do the jet-quark assignment. Corrected jets (adding muon 4-vectors to the jets) are used to determine the jet-quark associations, but strictly for this purpose (i.e. the original jet 4-vectors are used in all other places in the program). (*Required for full analysis*). **[0 or 1, default = 0]**
    - **-runAnalysisFinal**: runs over the output from the **-runInitial** step above, performs the final event selection, does the ABCDEF method for background estimation, produces a final output root file with data histo and also the background shape fit parameters and covariance matrix from the fit. (*Required for full analysis*) **[0 or 1, default = 0]**
    - **-runFinalTopMassExtraction**: runs at the very end to perform the final chi2 minimization to extract the measured values of mtop and Fbkgd with statistical uncertainties (*Required for full analysis*).
    - **-runJetResponseResolution**: runs over MC FHD3PD which have had no cuts applied. It creates an output (txt file) which has b- and light-jet responses and resolutions as a function of jet eta and energy or transverse momentum. The point at which these are calculated is right before the event reconstruction

so that one has a good sense of what the true response and resolution are as well as the mW and mTop distributions. It is this information that is fed into the Chi2 reconstruction (for any stage of the analysis). This should be run if you want to re-generate the txt file, but otherwise it's included in the package. Matched muons are added to jets for the response and resolution and they are added to jets to do the jet-quark assignment in the reconstruction, but not in the final analysis - it is similar in this sense to the transfer functions of the KLFitter. This improves the top reconstruction purity a small amount. **[0 or 1, default = 0]**
  - **-runMCSignalFullPlots**: runs over MC FHD3PD (with no cuts applied upstream) to produce a more exhaustive list of plots of jet kinematics including trigger turn-on curves. **[0 or 1, default = 0]**
  - **-runPLCDerivation**: runs over MC FHD3PD ntuple to derive jet parton-level correction factors for jets (as a function of reco jet energy and pseudorapidity). The output is a `.txt` file located in `/files` containing the jet energy and eta bin ranges as well as the actual correction factors. These are then later read in (if desired) and applied to jets in the analysis (NB: this does now not need to be run at all - it still can be but it is not considered part of the official analysis). Note that by default this really computes correction factors vs quark-level pT or E, and then uses numerical inversion to convert to corrections vs reco-level jet pT or E. DO NOT USE FOR NOW **[0 or 1, default = 0]**

- **Specify inputs/outputs**: options to specify names of input and output files to be used (all set by default, so do not have to be changed):
  - **-inFileMC**: input file path for MC (do not use together with -inFileTxt) **[runInitial]**
  - **-inFileData**: input file path for Data (do not use together with -inFileTxt) **[runInitial]**
  - **-initialOutFile**: initial program output file path (default: tTNtuple_InclTopRecoIndices(MC/Data)_nominal.root **[runInitial]**
  - **-inFileFinalMC**: input file path for final run MC (default: tTNtuple_InclTopRecoIndicesMC_nominal.root) **[runFinal]**
  - **-inFileFinalData**: input file path for final run Data (default: tTNtuple_InclTopRecoIndicesData_nominal.root) **[runFinal]**
  - **-inFileJetCF**: input txt file containing jet correction factors to be applied in final stage (default = files/JetCF_nominal.txt) **[runFinal]** -- do not use
- **Additional options for final analysis**: options to turn on/off several extra features in the final analysis (more are available but currently hard-coded):
  - **-applySLBJetVeto**: cut on top candidates with b-jet suspected to have decayed semileptonically **[0 or 1, default = 1]** -- do not use
  - **-applyJetPLCOption**: apply parton-level correction factors to jet 4-vectors **["X", "A", "B", "C", "D", default = "X" which means not applied]** -- do not use
  - **-sysType**: string identifier for particular run (nominal or systematic) **[default = "nominal"]**

Examples of how to run the code

- To run over everything with all four main steps steps - mc mass variation (to get slopes and intercepts), initial run (which does some extra pre-selection runs top reco as well), final (to do ABCDEF method and produce final plots) and also the final chi2 minimization taking into account signal and bkgd parameter uncertainties and correlations:

```
./FHAnalysisExe.exe -runAll 1 -inFileTxt FileLists/filelist_RZG_nominal.txt -isBatch 1
```

- To do the final stage only (ABCD method), but not re-do the mass variation, initial stage, etc.:

```
./FHAnalysisExe.exe -runAnalysisFinal 1 -inFileTxt FileLists/filelist_RZG_nominal.txt -isBatch 1
```

- To run the analysis to be able to get all of the *extra MC signal plots* (including the top reconstruction response/resolutions and b-tagging efficiency, etc), you need to do (from the beginning, i.e. not just the final step at the end):

```
./FHAnalysisExe.exe -runJetResponseResolution 1 -runMCSignalFullPlots 1 -inFileTxt FileLists/file
```

Getting the FHD3PDs to run locally

The slimmed FHD3PDs are only available on the grid for a short time, then they are auto-deleted. At the moment we keep three copies of the ntuples. One at RZG in Garching, one at MPP and one at Carleton (up in Canada where the bears are at home). Have a look at the text files in `FileLists` to see where they are located and copy them from there if you need them elsewhere (i.e. on you laptop). Create a separate filelist for these and use it in the run examples.

Content of the analysis package

There are several files/classes in the package

- physics.h/.C ... this is the base class to read the ntuple. It was created with root MakeClass mechanism and should not be edited.
- physicsFinal.h/C ... same as above but is used to be able to read the input for the final stage (based on the output ntuple from the initial stage)
- FHAnalysisLoop.h/.C ... inherits from physics and implements the actual event loop. It can access all variables of the slimmed FHD3PD ntuple directly. This is where the base event selection and top reconstruction is done.
- FHAnalysisLoopFinal.h/C ... inherits from physicsFinal and implements the final event loop. It also calls functions from the CombinedPlots helper class to produce final plots (with mc signal, data and the estimated background from the ABCDEF method) or background-subtracted plots
- PlotStylesAndFits.h/C ... is a helper class for setting up 1d and 2d histograms, setting labels for plots, the fit functions used by various programs, etc.
- PartonLevelJetCorrection.h/C ... is a larger bit of code containing everything needed to do the parton-level jet correction derivation (based on the iterative W method for light jets) - do not use
- FHAnalysisExe.C ... the top level file implementing the main() method. It deals with the input and processes the command line arguements.
- TopReco.h/.C ... this is a helper class to reconstruct the tops.
- CombinedPlots.h/.C ... helper class to produce the final data plots / and data/MC comparison, ABCDEF plots, etc. It is called in FHAnalysisLoopFinal.
- EventSelectionTools.h/.C ... all event selection is controlled here (except the pre-selection which is done at the slimming stage).
- JetResponseResolution.h/.C ... as described above
- LorentzVectorTools.h/.C ... for operations on TLorentzVectors, or vectors of such objects as well as iterative look-up of jet PLC values when deriving or applying the parton-level jet corrections.
- SignalPlotLoop.h/.C ... large file in which all plots for MC signal including mass variation, jet kinematics, b-tagging, trigger, etc. are created. The reason it's all in one file is that for each event it prepares vectors of all of the objects, does top reco, etc. so that that is all in place before filling various histos.

More info

For more info you can also checkout the talks from MPI hadronic top analysis meetings:
Meeting on 23 January 2013
Meeting on 11 January 2013

Example of Program Flow Outputs to Expect

- 1. Running Jet Response/Resolution Produces the Following:
    - **reco-level** distributions with gauss fit (jets matched to quarks) to know the values of responses and resolutions that should be used in the chi2 fit. The reconstructed-level quantities are generated by matching jets to the quarks (so that there are no wrong permutations) and building the relevant distributions. They are constructed with the same cuts that would have been applied in the full analysis immediately before the top reconstruction. (Base histo showing ΔMjj at the reconstructed level to get its expected resolution, Same for ΔR32, Base histo for Mjj used for its response and resolution)
    - **output text file** as an example of what it should look like. File is read in for top reconstruction if using chi-squared method (default) and is part of the package that is checked out, but it can also be reproduced (Sample txt file as input for chi squared reconstruction)
- 2. Running MC Mass Variation Produces the Following:
    - **slope of fitted mtop vs generator mtop** (and same for R32) as well as all other parameters from gauss plus landau fit vs top mass (Sample R32 Plot)
    - **output text file** with the slope and intercept from the above (Sample Output Text File)
- 3. Running Initial Step (Data and MC): has no output other than root files in which one already has access to which jets have been selected to form tops, Ws, etc. This makes the next (final) step much faster.
- 4. Final Step with ABCDEF Background-Estimation and Final Fits:
    - **control plots** such as jet pT but many others showing data/MC+bkgd agreement (for each, ABCDEF method was used to estimate bkgd) (Sample Output Final Data/MC Plot for Jet 1 pT)
    - **control plots** such as jet pT but many others (as above) but showing normalized shapes of distributions for different bins of ABCDEF observable 1 (to show if one should expect good performance or not from ABCDEF method) (Sample Output Bkgd Evolution Plot for Jet 1 pT)
    - **qcd shape** Extraction from ABCDEF method (for mtop, mW, R32 only) plus a fit together with the non all-hadronic ttbar sample (Gauss+Landau) (Sample Output QCD Shape for R32)
    - **final data distribution** together with results of the fit for extracted values of mtop and Fbkgd (Sample Output Final Data Distribution with Fit overlayed for R32 observable, Final contour plot showing 1- and 2-sigma contours from the fit)
    - *final value of the *measured* value of mtop is printed to the screen as well as the statistical uncertainty.

Recipe for cut value studies

Log in in Garching, go to some directory of your choice and check out the framework
svn co svn+ssh://svn.cern.ch/reps/atlasinst/Institutes/MPI/HEC/analysis/FullHadronicTopAnalysis/t

Setup the ATLAS environment and ROOT:
setupATLASUI
localSetupROOT 5.34.24-x86_64-slc6-gcc48-opt

Compile the framework:
cd FullHadronicTopAnalysis
cd cmt
make

Now you are ready to run (locally or on batch).

On batch:
make sure ROOT is setup.
cd rzgbatch

There is a file in this dir called:
createBatchScriptsAndPbsFiles.py

open the file and modify those two lines (close to top):
emailAddress = 'wildauer@mppmu.mpg.de'
cutSetup = '-CutVal_Njet 12 -CutVal_Jet5pT 60 -CutVal_Jet6pT 25'

The second line specifies the cut setup to be used. The values shown are default.
Note that - for the moment - one can only run one setup at a time (previous run will be overwritt

Explanation of cut setup:
-CutVal_Njet 12 ... means use 6 to 12 jets. If you put e.g. 7 it will only consider events with 6
-CutVal_Jet5pT 60 ... pt in GeV of the first 5 jets
-CutVal_Jet6pT 25 ... pt in GeV of the 6th jet

After you modified the file type:
python ./createBatchScriptsAndPbsFiles.py

It will create several scripts and two folders (log and pbsFiles).

To run on batch all you need to do is execute first:
./submitFirstJobsToBatch.sh
this will submit one job to batch which runs baseline analyses (needed for the next job(s)).

Once it is finished (you will get an email, will take several hours) you can execute
./submitJobsToBatch.sh
this will submit many jobs to the batch to run in parallel for all the systematics. Each job will

When all these are finished execute the final step which calculates all systematics and produces
There is no batch job needed for this. In the top level dir of the framework just execute
./evaluateSystematics.exe

It will only take a few minutes.

Running PDF Uncertainty (*Expert Only*)

This is an abridged set of steps taken from the LHAPDF installation page which can be found (here). It is assumed that one already has access to all of the various PDF sets (CT10, MSTW, NNPDF) such that they can simply be copied over. The base directory (the full path to the location of your particular version of the trunk or tagged version of the code) will be assumed to be ***baseDir*** (*change this to whatever you use!*). All other standard ROOT, asetup steps are assumed to have been done prior to this point. The steps to then be able to obtain the PDF systematic are given as follows:

- 1. Copy over the tarred file containing the LHAPDF source code, untar in ***baseDir*** directory and move to the newly created LHAPDF-6.1.4 directory:

```
* cd <full path to baseDir>
* cp <path to location of tarred file>/LHAPDF-6.1.4.tar.gz .
* tar xf LHAPDF-6.1.4.tar.gz
* cd LHAPDF-6.1.4
```

- 2. Build LHAPDF packages, and set environment variables (after these steps the directory lhapdf6.1.4 will also be created in baseDir)

```
* ./configure --prefix=$PWD/../lhapdf6.1.4
* make -j2 && make install
* cd ..
* export PATH=$PWD/lhapdf6.1.4/bin:$PATH
* export LD_LIBRARY_PATH=$PWD/lhapdf6.1.4/lib:$LD_LIBRARY_PATH
* export PYTHONPATH=$PWD/lhapdf6.1.4/lib64/python2.6/site-packages:$PYTHONPATH
```

- 3. Just test a few LHAPDF commands to make sure things are working up to this point

```
* lhapdf-config --help
* lhapdf list
```

- 4. Assuming things are ok up to this point, now copy the PDF sets over to where they will need to be accessed (either this or get them yourself as per instructions above)

```
* cp -r <path to PDF sets>/CT10* lhapdf6.1.4/share/LHAPDF/.
* cp -r <path to PDF sets>/MSTW* lhapdf6.1.4/share/LHAPDF/.
* cp -r <path to PDF sets>/NNPDF* lhapdf6.1.4/share/LHAPDF/.
```

- 5. Uncomment the relevant portions of the makefile and python script (two lines in **makefile**, 6 lines in **createBatchScriptsAndPbsFiles.py**)

- 6. Cleanup, recompile from scratch, and then you're ready to run (the new script **rzgbatch/submitJobsToBatch.sh** will then contain the necessary lines to submit jobs which run over the various PDF sets, as with the other systematics. A separate .root file will exist for each PDF set).

```
* cd cmt
* make clean
* make
```

- 7. When batch jobs have all finished successfully, run the **./evaluateSystematics** step as before (this time however the PDF systematics will have been included and the corresponding entry in the systematics table will appear).

Ongoing Analysis Tasks

Description of Task	Priority	Assigned To	Status
Update ABCD correlation plots so that they are (data-signal) rather than data (to reflect expected correlation in multi-jet background)	Medium	Tom	***Completed***
Perform closure tests when drawing pseudo events from 2D R32 distribution and look at pull mean and width plots	Medium	Tom	***Completed***
Adjust fitting procedure for signal shape to allow for tighter cuts as per Sven's suggestion (two-step fitting process)	Low	Tom	Currently on hold due to other priorities
Add input argument to allow for different ABCD options (**ABCD**, **ABCDEF**, **ABEF**, **CDEF**)	Medium	Tom	***Completed***
Add systematic error bars to remaining control plots	Medium	Tom	***Completed***
Investigate data/MC agreement in top quark pT as per Jim's suggestion and investigate effect on measurement	Medium	Tom	***Completed***
Re-do trigger efficiency systematic uncertainty based on pt-dependent scale factors and redo pseudo experiments	High	Tom	***Completed***
Investigate disagreement in template shape fits for R32 values near 2.6 for some mass samples	Medium	Tom/Teresa	***Mostly Understood***
Produce normalized histos comparing the nominal signal vs JER systematic to show size of expected broadening of R32	Medium	Tom	***Completed***
Investigate statistical component of systematic uncertainties (using Barlow reference)	Medium	unassigned	***Completed***
Replace 0th order polynom fit with 1st order to investigate slope (for pull tests)	Medium	unassigned	Not yet begun
Adapt code to be able to run with b-tagging in top reconstruction (mostly works now)	Low	Tom	***Completed***
	Medium	unassigned	***Completed***

Adjust framework to be able to run on only 5 mass points for new FS mass variation samples (involved and needs to remain backwards-compatible)			
Produce quark-gluon flavour root file to feed into JES uncertainty provided (so as to use all-hadronic ttbar-specific q-g fractions rather than defaults)	Low	unassigned	Not yet begun

| Validate adding muon four-vectors to jets by performing study on simulated ttbar events with different b-quark fragmentation if such samples are available, otherwise will likely have to discontinue adding muons in this way | Medium | unassigned | Not yet begun |

This topic: Main > FullHadronAnalysisFramework
Topic revision: r34 - 2017-01-26 - ThomasMcCarthy