

MadAnalysis 5: a framework for phenomenological analysis

Jack Y. Araz

INSTITUTE FOR PARTICLE PHYSICS PHENOMENOLOGY
DURHAM UNIVERSITY

The 2021 EFT School on Collider Phenomenology
April 13th 2021



Outline

- ❖ Introduction
- ❖ MadAnalysis in a nutshell
- ❖ A simple phenomenological study
 - Parton level analysis
 - Hadron level analysis
- ❖ What else can I do with MadAnalysis beyond this tutorial?

MonteCarlo tools & discoveries at the LHC

Exploration

We still need to uncover the veil of the new physics!

Hitchhiker's guide to new physics

Apriori preparation

- ❖ Theoretical model building
- ❖ Phenomenological studies
- ❖ Collider analyses

A posteriori reactions to announcements

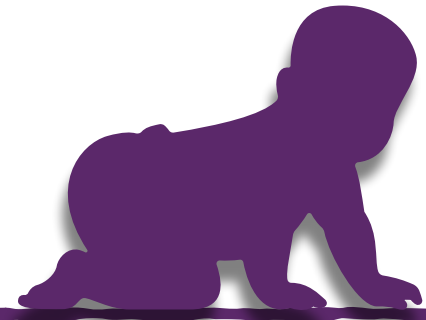
- ❖ Theoretical model building
- ❖ Recasting experimental analyses
- ❖ Precision measurements

- ✦ MonteCarlo tools are essential
- ✦ How to analyze the MC output easily?

MadAnalysis 5 in a nutshell

Normal Mode

- ❖ Intuitive commands typed in the PYTHON interface
- ❖ Analysis performed behind the scenes (black box)
- ❖ Human readable output: HTML and LaTeX



Expert Mode

- ❖ C++ programming within the SampleAnalyzer framework
- ❖ Support for multiple sub-analyses, an efficient way for handling cuts and histograms, etc.



MadAnalysis 5 in a nutshell

What is MadAnalysis 5?

- ❖ A Python and C++ based framework for phenomenological analyses
- ❖ Any level of sophistication: partonic, hadronic, detector, reconstructed
- ❖ Several input format: STDHEP, HEPMC, LHE, LHEO, ROOT (from Delphes)
- ❖ User-friendly, flexible and fast
- ❖ Interfaces to other HEP packages (fast detector simulation, jet clustering etc.)

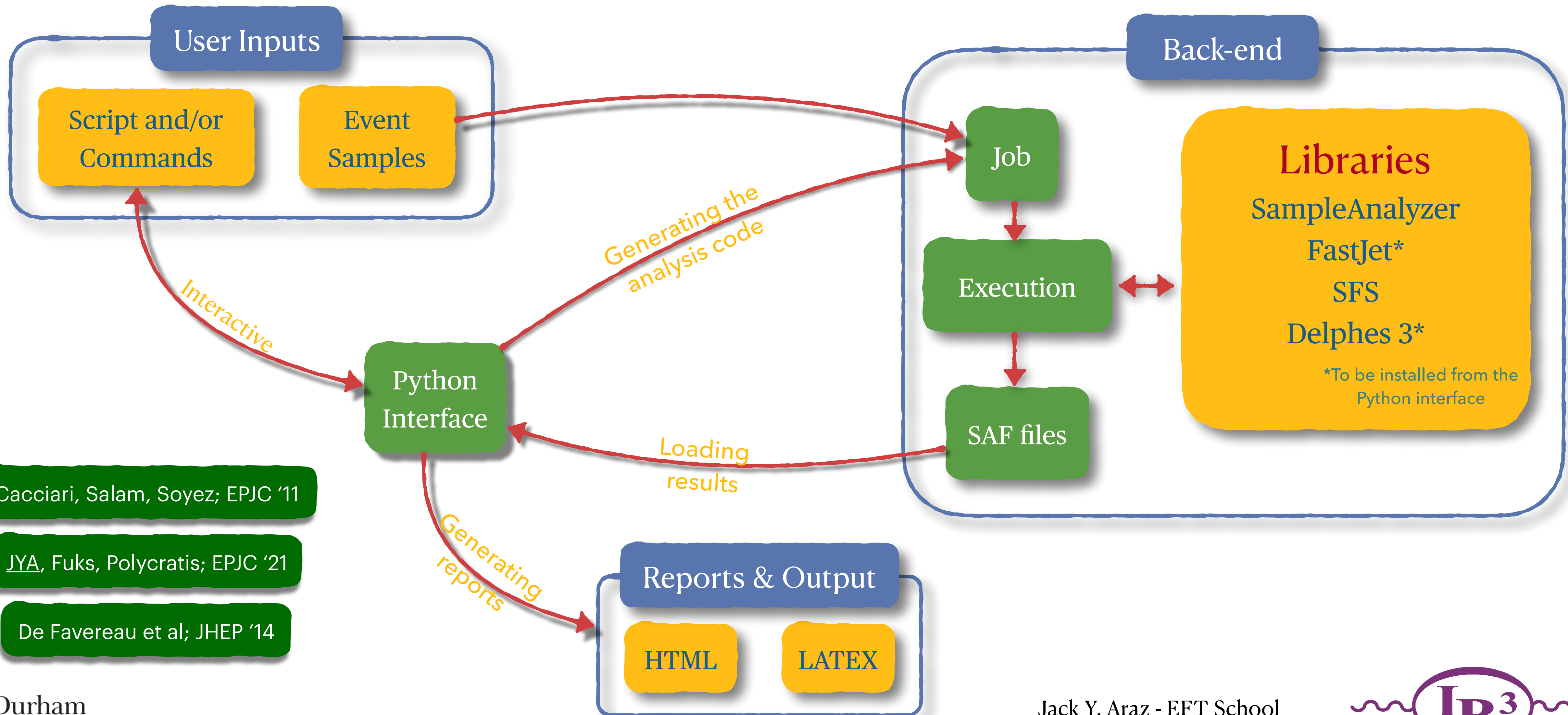
Two Modules

- ❖ A Python command line interface (interactive)
- ❖ A C++ core module, SampleAnalyzer

Today's Tutorial

MadAnalysis 5 : Normal Mode

MadAnalysis 5: Normal Mode



Cacciari, Salam, Soyez; EPJC '11

JYA, Fuks, Polyratis; EPJC '21

De Favereau et al; JHEP '14

MadAnalysis 5: Normal Mode

`$. /bin/ma5`

Main python and C++ libraries to run Ma5

```
ma5> install fastjet
ma5> install zlib
ma5> install matplotlib
```

Optional External libraries

Reco Mode

`$. /bin/ma5 -R`

Libraries for reinterpretation studies

Expert Mode

`$. /bin/ma5 -e`

External plotting libraries

Libraries for report generation

```
MA5: Reading user settings ...
MA5: Checking mandatory packages:
MA5: - Python [OK]
MA5: - GNU GCC g++ [OK]
MA5: - GNU Make [OK]
MA5: Checking optional packages devoted to data processing:
MA5: - Zlib [OK]
MA5: - FastJet [OK]
MA5: - Root [OK]
MA5: - Delphes [OK]
MA5: - Delphes-MA5tune [DISABLED]
MA5: Checking the MadAnalysis 5 core library:
MA5: => MadAnalysis libraries found.
MA5: => MadAnalysis test program works.
MA5: Reading user settings ...
MA5: Checking optional packages devoted to reinterpretation:
MA5: - SciPy [OK]
MA5: - PAD [OK]
MA5: - PADForMA5tune [DISABLED]
MA5: - PADForSFS [OK]
MA5: - pyhf [DISABLED]
MA5: Checking optional packages devoted to histogramming:
MA5: - Root [OK]
MA5: - Matplotlib [OK]
MA5: - gnuplot [DISABLED]
MA5-WARNING: gnuplot disabled. Plots in gnuplot format file will not be produced.
MA5: - pdflatex [OK]
MA5: - latex [OK]
```

You will only need these for today's tutorial

A simple phenomenological study with MadAnalysis 5

Parton Level Analysis (LHE file)

```
$ ./bin/ma5  
ma5> install samples
```

Importing samples

```
ma5> import /PATH/my_smp.lhe.gz as fancy_name
```

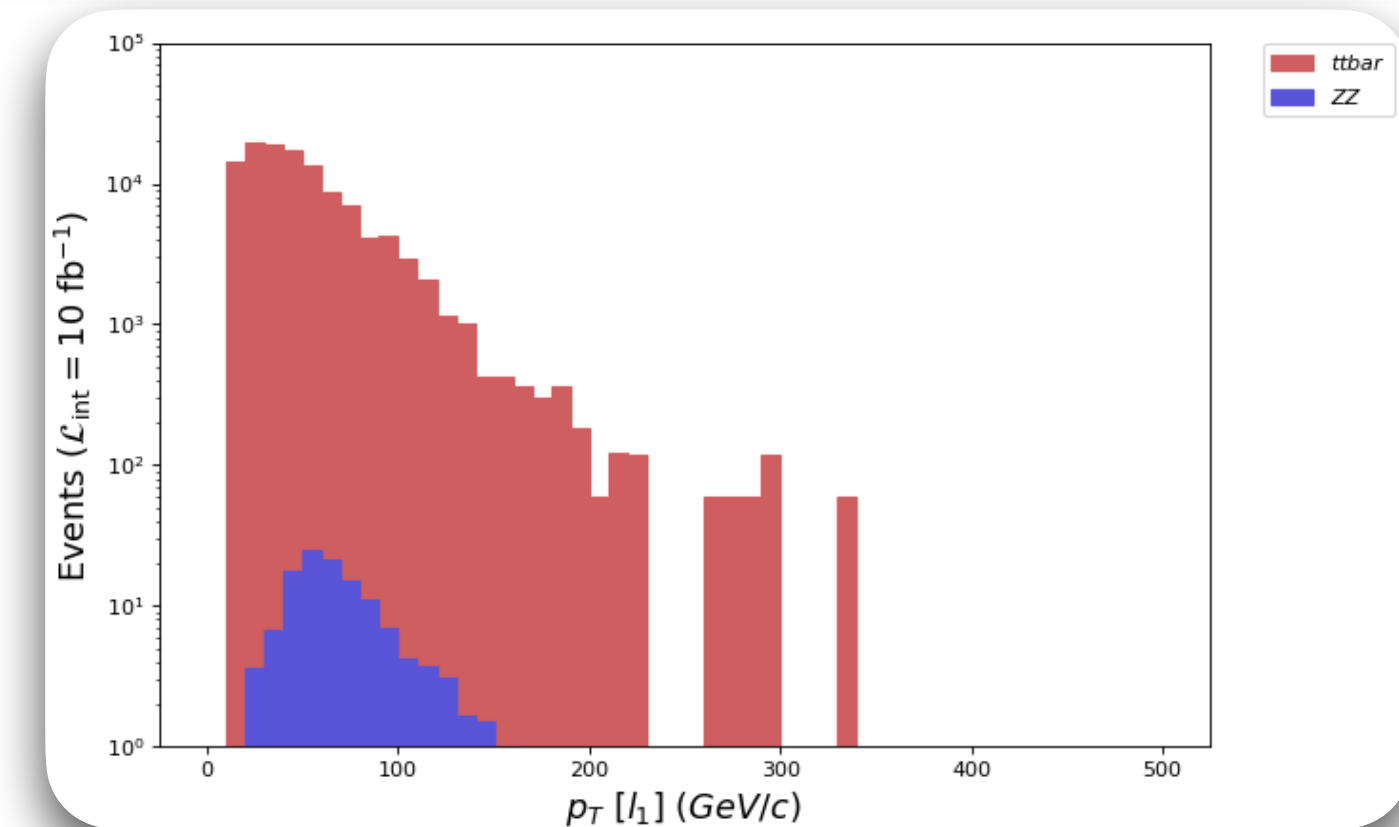
Location of the sample

User-defined identity of the sample

- ❖ Multiple samples with same “Fancy Name” will stored as the same sample
- ❖ Multiple samples with different names will be stored as different samples

Multiple samples for ttbar

```
ma5>import samples/zz.lhe.gz as ZZ  
MA5: -> Storing the file 'zz.lhe.gz' in the dataset 'ZZ'.  
ma5>import samples/ttbar*lhe* as ttbar  
MA5: -> Storing the file 'ttbar_fh.lhe.gz' in the dataset 'ttbar'.  
MA5: -> Storing the file 'ttbar_sl_1.lhe' in the dataset 'ttbar'.  
MA5: -> Storing the file 'ttbar_sl_2.lhe' in the dataset 'ttbar'.  
MA5: -> Storing the file 'ttbar_sl_1.lhe.gz' in the dataset 'ttbar'.  
MA5: -> Storing the file 'ttbar_sl_2.lhe.gz' in the dataset 'ttbar'.  
ma5>define l = l+ l-  
ma5>select (l) PT > 10  
ma5>select (j) PT > 20  
ma5>select (j) DELTAR (l) > 0.4  
ma5>plot PT (l[1]) 50 0 500 [logY]  
ma5>set selection[4].ymin = 1  
ma5>set selection[4].ymax = 1e5
```



```
ma5> submit my_first_analysis
```

Parton Level Analysis (LHE file)

A sample named "ZZ" and another named "ttbar"

```

ma5>import samples/zz.lhe.gz as ZZ
MA5: -> Storing the file 'zz.lhe.gz' in the dataset 'ZZ'.
ma5>import samples/ttbar*lhe* as ttbar
MA5: -> Storing the file 'ttbar_sl_1.lhe.gz' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_2.lhe' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_1.lhe.gz' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_2.lhe.gz' in the dataset 'ttbar'.
ma5>define l = l+ l-
ma5>select (l) PT > 10
ma5>select (j) PT > 20
ma5>select (j) DELTAR (l) > 0.4
ma5>plot PT (l[1]) 50 0 500 [logY]
ma5>set selection[4].ymin = 1
ma5>set selection[4].ymax = 1e5
    
```

Press Tab

```

ma5>import samples/zz.lhe.gz as fancy_name
    
```

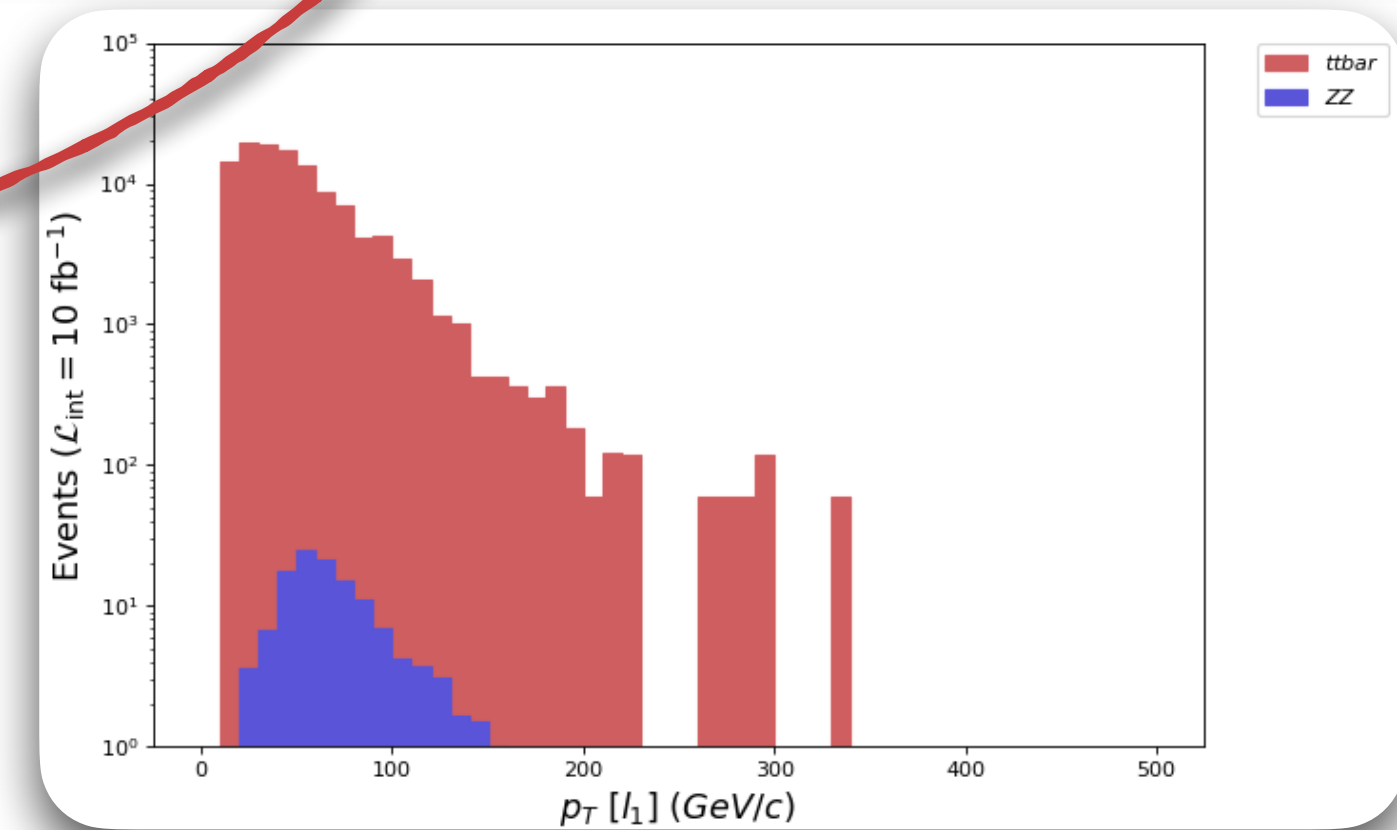
Location of the sample

User-defined identity of the sample

```

ma5>set ttbar.
ttbar.backcolor      ttbar.linewidth      ttbar.scale_down_variation  ttbar.type
ttbar.backstyle      ttbar.pdf_down_variation  ttbar.scale_up_variation    ttbar.weight
ttbar.linecolor      ttbar.pdf_up_variation   ttbar.scale_variation       ttbar.weighted_events
ttbar.linestyle      ttbar.pdf_variation      ttbar.title                 ttbar.xsection
    
```

❖ Multiple samples with different names will be stored as different samples



Parton Level Analysis (LHE file)

What does it mean?

- ❖ `select PT (l) > 20`: selects those for which at least one lepton has a $p_T^l > 20$ GeV
- ❖ `reject PT(1) < 20`: rejects those events for which $p_T^l < 20$ GeV is found.
- ❖ `select (l) PT > 20`: keep in each event only the leptons with $p_T^l > 20$ GeV and ignore any other lepton

Multi-particle definition

Cut-flow

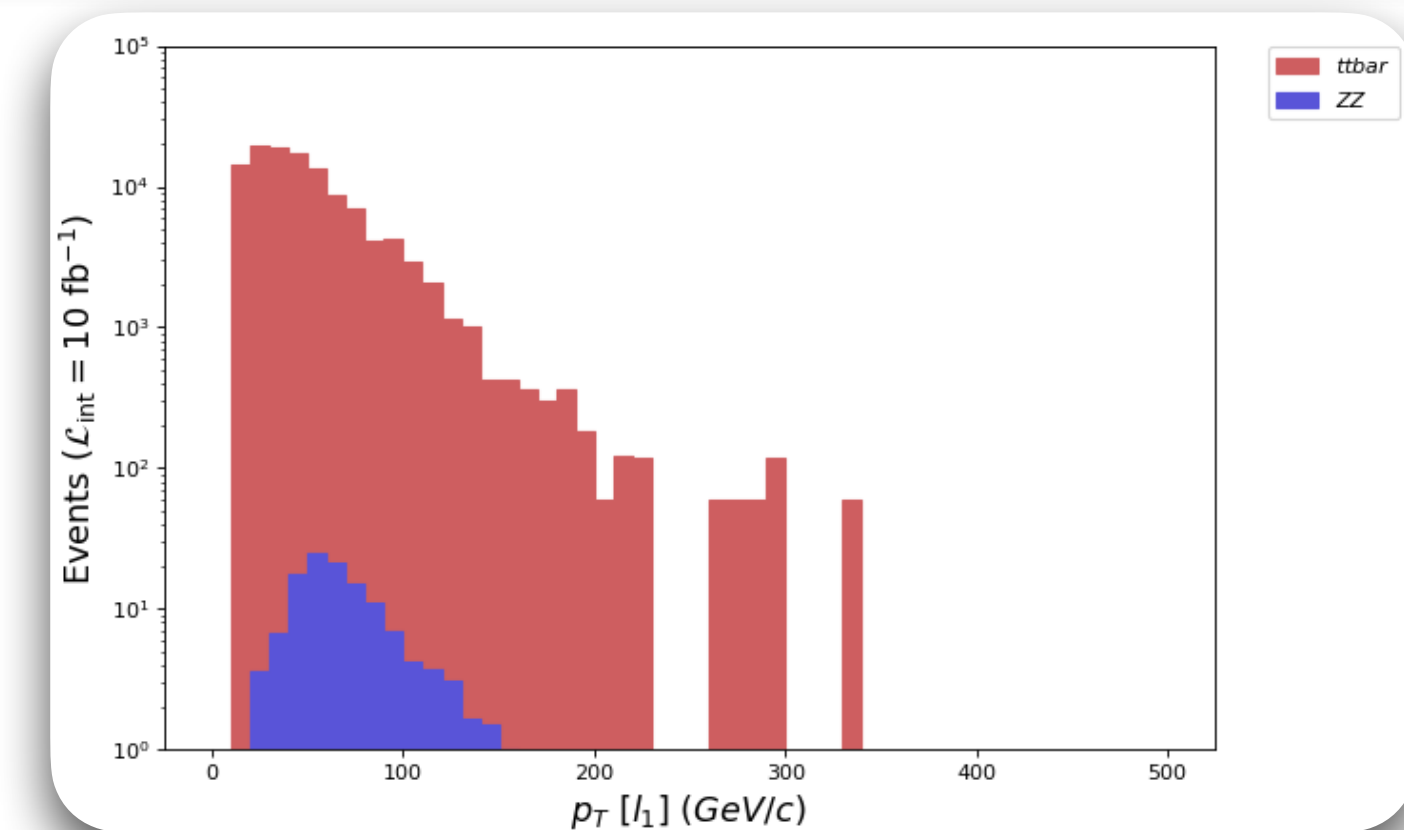
Histogram definition

```

ma5>import samples/zz.lhe.gz as ZZ
MA5: -> Storing the file 'zz.lhe.gz' in the dataset 'ZZ'.
ma5>import samples/ttbar*lhe* as ttbar
MA5: -> Storing the file 'ttbar_fh.lhe.gz' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_1.lhe' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_2.lhe' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_1.lhe.gz' in the dataset 'ttbar'.
MA5: -> Storing the file 'ttbar_sl_2.lhe.gz' in the dataset 'ttbar'.
ma5>define l = l+ l-
ma5>select (l) PT > 10
ma5>select (j) PT > 20
ma5>select (j) DELTAR (l) > 0.4
ma5>plot PT (l[1]) 50 0 500 [logY]
ma5>set selection[4].ymin = 1
ma5>set selection[4].ymax = 1e5
    
```

Press tab after plot command!

ABSETA	N	dABSETA	dTHETA	dsR	dvPY	rPX	sPT	sdPHI	vN	vdMT
ALPHAT	NAPID	dBETA	dY	dsTHETA	dvPZ	rPY	sPX	sdPT	vP	vdMT_MET
ALPHA_QCD	NPID	dDELTAR	dsABSETA	dsY	dvR	rPZ	sPY	sdPX	vPHI	vdN
ALPHA_QED	P	dE	dsBETA	dvABSETA	dvTHETA	rR	sPZ	sdPY	vPT	vdP
BETA	PHI	dET	dsE	dvBETA	dvY	rTHETA	sR	sdPZ	vPX	vdPHI
DELTAR	PT	dETA	dsET	dvDELTAR	rABSETA	rY	sTHETA	sdR	vPY	vdPT
DPHI_0_2PI	PX	dGAMMA	dsETA	dvE	rBETA	sABSETA	sY	sdTHETA	vPZ	vdPX
DPHI_0_PI	PY	dM	dsGAMMA	dvET	rE	sBETA	sdABSETA	sdY	vR	vdPY
E	PZ	dMT	dsM	dvETA	rET	sE	sdBETA	vABSETA	vTHETA	vdPZ
ET	R	dMT_MET	dsMT	dvGAMMA	rETA	sET	sdE	vBETA	vY	vdR
ETA	RECOIL	dN	dsMT_MET	dvM	rGAMMA	sETA	sdET	vDELTAR	vdABSETA	vdTHETA
GAMMA	SCALE	dP	dsN	dvMT	rM	sGAMMA	sdETA	vE	vdBETA	vdY
M	SQRTS	dPHI	dsP	dvMT_MET	rMT	sM	sdGAMMA	vET	vdDELTAR	
MEFF	TET	dPT	dsPHI	dvN	rMT_MET	sMT	sdM	vETA	vdE	
MET	THETA	dPX	dsPT	dvP	rN	sMT_MET	sdMT	vGAMMA	vdET	
MHT	THT	dPY	dsPX	dvPHI	rP	sN	sdMT_MET	vM	vdETA	
MT	WEIGHTS	dPZ	dsPY	dvPT	rPHI	sP	sdN	vMT	vdGAMMA	
Y		dR	dsPZ	dvPX	rPT	sPHI	sdP	vMT_MET	vdM	



Parton Level Analysis (LHE file)

What does it mean?

- ❖ select $PT(1) > 20$: selects those for which at least one lepton has a $p_T^l > 20$ GeV
- ❖ reject $PT(1) < 20$: rejects those events for which $p_T^l < 20$ GeV is found.
- ❖ select $(1) PT > 20$: keep in each event only the leptons with $p_T^l > 20$ GeV and ignore any other lepton

Press tab after plot command!

Multi-particle definition

Cut-flow

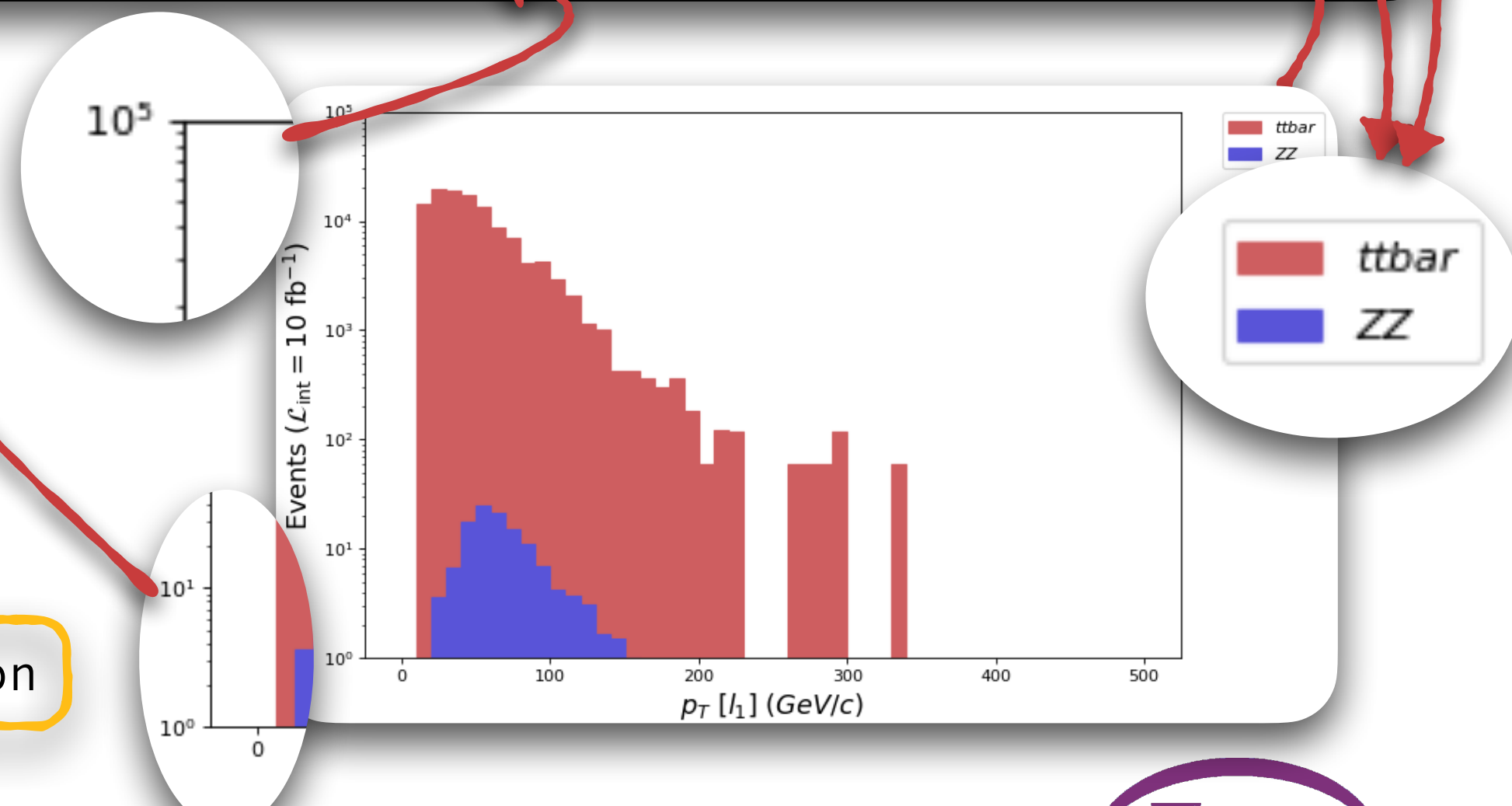
Histogram definition

```

ma5>import samples/zz.lhe.gz as ZZ
MA5:  -> Storing the file 'zz.lhe.gz' in the dataset 'ZZ'.
ma5>import samples/ttbar*lhe* as ttbar
MA5:  -> Storing the file 'ttbar_fh.lhe.gz' in the dataset 'ttbar'.
MA5:  -> Storing the file 'ttbar_sl_1.lhe' in the dataset 'ttbar'.
MA5:  -> Storing the file 'ttbar_sl_2.lhe' in the dataset 'ttbar'.
MA5:  -> Storing the file 'ttbar_sl_1.lhe.gz' in the dataset 'ttbar'.
MA5:  -> Storing the file 'ttbar_sl_2.lhe.gz' in the dataset 'ttbar'.
ma5>define l = l+ l-
ma5>select (1) PT > 10
ma5>select (j) PT > 20
ma5>select (j) DELTAR (1) > 0.4
ma5>plot PT (1[1]) 50 0 500 [logY]
ma5>set selection[4].ymin = 1
ma5>set selection[4].ymax = 1e5
    
```

dABSETA	N	dABSETA	dTHETA	dsR	dvPY	rPX	sPT	sdPHI	vN	vdMT
ALPHAT	NAPID	dBETA	dY	dsTHETA	dvPZ	rPY	sPX	sdPT	vp	vdMT_MET
ALPHA_QCD	NPID	dDELTA	dsABSETA	dsY	dvR	rPZ	sPY	sdPX	vPHI	vdN
ALPHA_QED	P	dE	dsBETA	dvABSETA	dvTHETA	rR	sPZ	sdPY	vPT	vdP
BETA	PHI	dET	dsE	dvBETA	dvY	rTHETA	sR	sdPZ	vpX	vdPHI
DELTA	PT	dETA	dsET	dvDELTA	rABSETA	rY	sTHETA	sdR	vpY	vdPT
DPHI_0_2PI	PX	dGAMMA	dsETA	dvE	rBETA	sABSETA	sY	sdTHETA	vpZ	vdPX
DPHI_0_PI	PY	dM	dsGAMMA	dvET	rE	sBETA	sdABSETA	sdY	vr	vdPY
E	PZ	dMT	dsM	dvETA	rET	sE	sdBETA	vABSETA	vTHETA	vdPZ
ET	R	dMT_MET	dsMT	dvGAMMA	rETA	sET	sdE	vBETA	vY	vdR
ETA	RECOIL	dN	dsMT_MET	dvM	rGAMMA	sETA	sdET	vDELTA	vdABSETA	vdTHETA
GAMMA	SCALE	dP	dsN	dvMT	rM	sGAMMA	sdETA	vE	vdBETA	vdY
M	SQRTS	dPHI	dsP	dvMT_MET	rMT	sM	sdGAMMA	vET	vdDELTA	
MEFF	TET	dPT	dsPHI	dvN	rMT_MET	sMT	sdM	vETA	vdE	
MET	THETA	dPX	dsPT	dvP	rN	sMT_MET	sdMT	vGAMMA	vdET	
MHT	THT	dPY	dsPX	dvPHI	rP	sN	sdMT_MET	vm	vdETA	
MT	WEIGHTS	dPZ	dsPY	dvPT	rPHI	sP	sdN	vMT	vdGAMMA	
MET	Y	dR	dsPZ	dvPX	rPT	sPHI	sdP	vMT_MET	vdM	

ma5> display selection

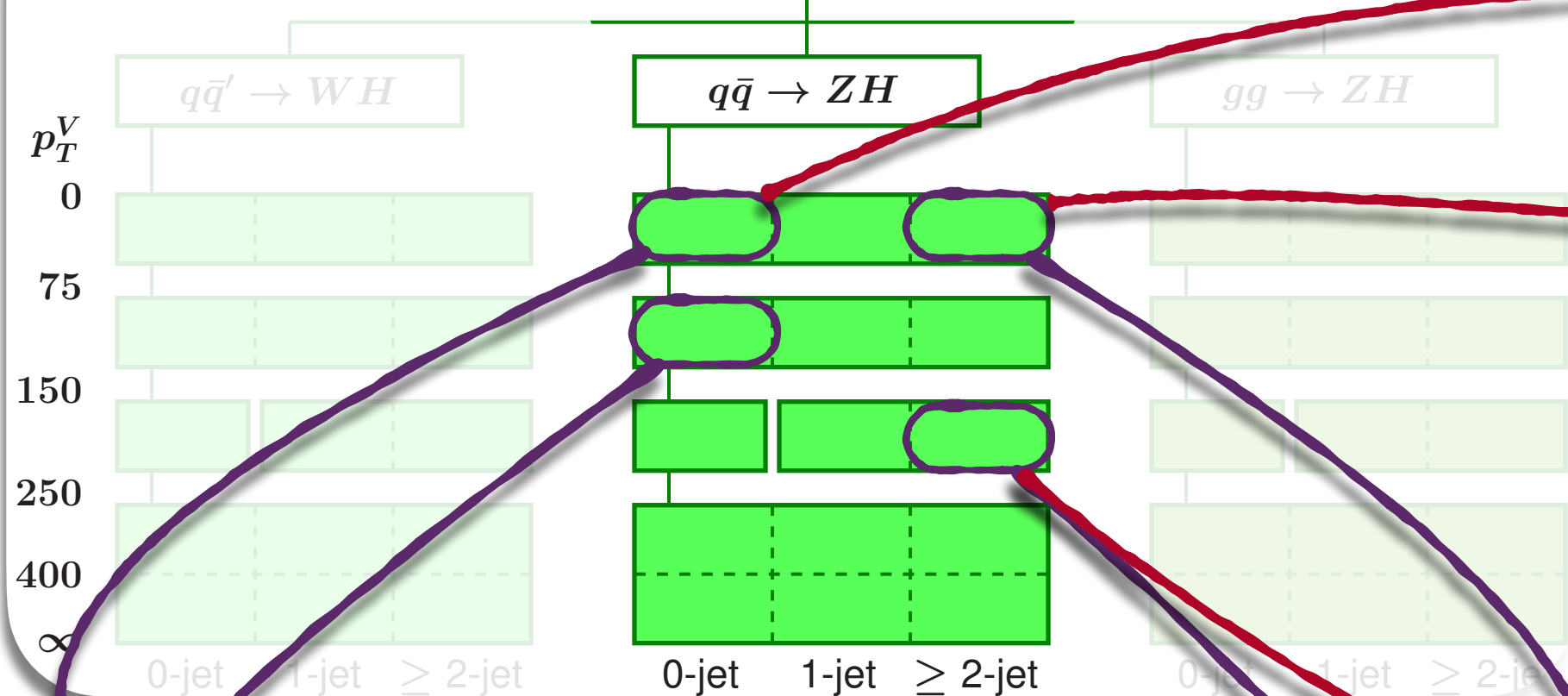


Defining Signal Regions in Normal Mode

ma5> define_region

Stage 1.2

$VH = V(\rightarrow \text{leptons})H$



```
ma5> define_region NJ0_0_75
ma5> define_region NJ2_0_75
ma5> define_region NJ0_75_150
ma5> define_region NJ2_150_250
```

Fill the region definitions with cuts

```
ma5> select <condition> {region_name}
```

Two regions with no jets

```
select N (j) == 0 {NJ0_0_75 NJ0_75_150}
```

Two regions with at least two jets

```
select N (j) >= 2 {NJ2_0_75 NJ2_150_250}
```

One region with $0 < p_t(l_1 \oplus l_2) \leq 75$ [GeV]

```
select vPT(l[1] l[2]) > 0 and \
vPT(l[1] l[2]) <= 75 {NJ0_0_75}
```

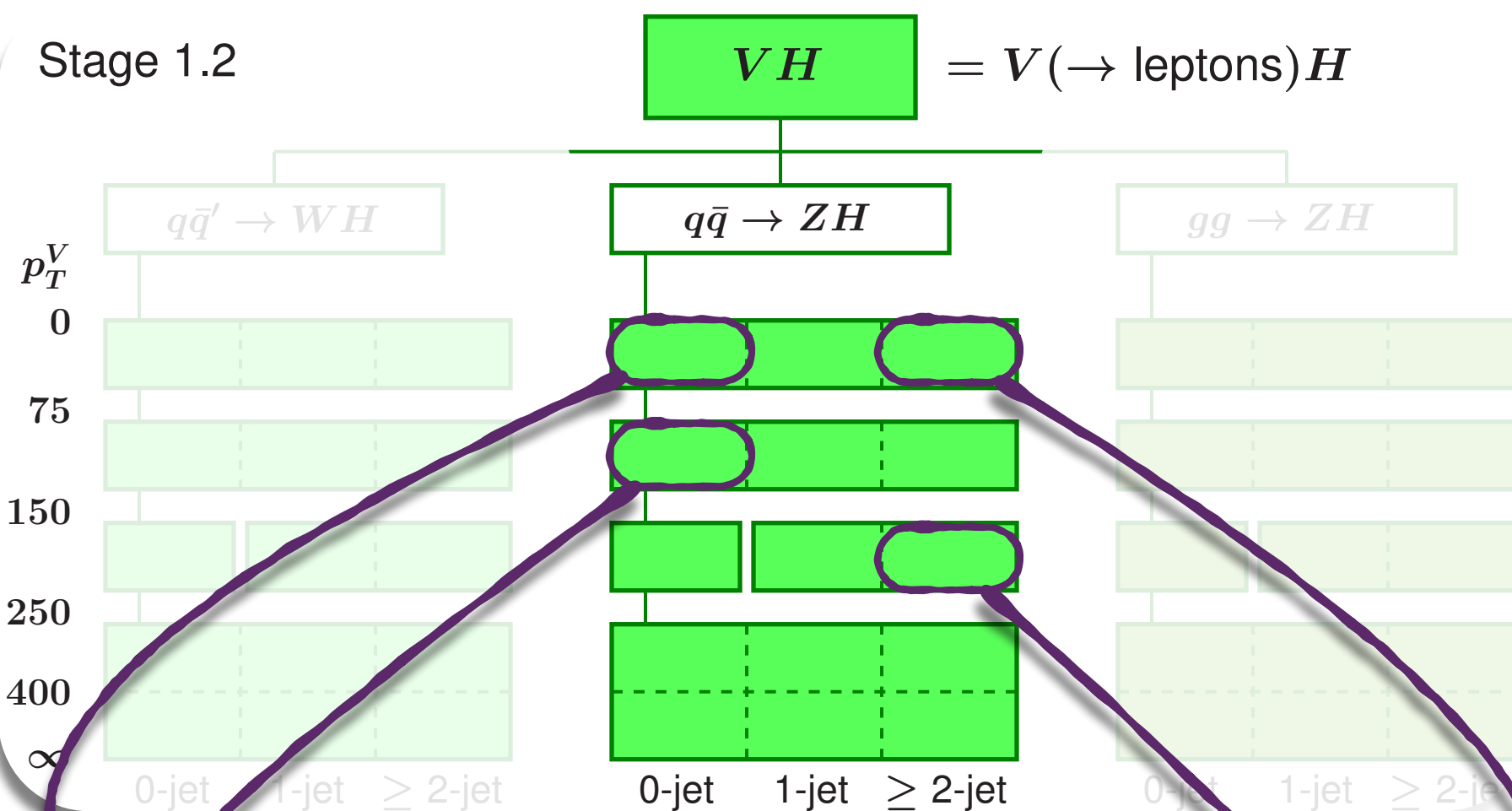
One region with $150 < p_t(l_1 \oplus l_2) \leq 250$ [GeV]

```
select vPT(l[1] l[2]) > 150 and \
vPT(l[1] l[2]) <= 250 {NJ2_150_250}
```

$vPT(l[1] l[2])$:= transverse momentum of the two leading lepton system

Defining Signal Regions in Normal Mode

ma5> define_region



```
ma5> define_region NJ0_0_75
ma5> define_region NJ2_0_75
ma5> define_region NJ0_75_150
ma5> define_region NJ2_150_250
```

Fill the region definitions with histograms

```
ma5> plot <definition> {region_name} [option]
```

A histogram for $p_T^V := p_T(l_1 \oplus l_2)$

```
plot vPT (l[1] l[2]) 50 0 500 {NJ0_0_75} [logY]
```

A histogram for $m_{Vh} := m(l_1 \oplus l_2 \oplus b_1 \oplus b_2)$

```
plot vM (l[1] l[2] b[1] b[2]) 50 0 500 \
{NJ2_0_75} [logY]
```

Hadron Level Analysis (HEPMC file)

```
$ ./bin/ma5 -R  
ma5> set main.fastsim.package = fastjet  
ma5> set main.fastsim.algorithm = antikt  
ma5> set main.fastsim.radius = 0.4  
ma5> set main.fastsim.ptmin = 20
```



Jets are clustered using anti- k_T algorithm with clustering radius of $R = 0.4$ and minimum jet transverse momentum at 20 GeV

Rest is the same, just import HEPMC files this time and Voilà!

How to work with reconstructed final states

- HEPMC files can be too large to work with. Depending on the process and number of events they can easily reach to tens of GB.
- MadAnalysis can write a compressed LHE file by reconstructing the showered events. This file will include only the reconstructed final state objects like electrons, muons, hadronic taus, jets, photons and missing energy.

Define Jet reconstruction and any defector inefficiencies

```
$ ./bin/ma5 -R
```

```
ma5> set main.fastsim.package = fastjet  
ma5> set main.fastsim.algorithm = antikt  
ma5> set main.fastsim.radius = 0.4  
ma5> set main.fastsim.ptmin = 20
```

```
ma5> set main.outputfile = "my_output.lhe.gz"
```

```
ma5> import /PATH/sample.hepmc.gz
```

```
ma5> submit reduced_sample
```

Set output file name

Important: Save the cross section in your notes!

Import sample(s)

What else can I do with MadAnalysis 5?

LHC Recasting with MadAnalysis 5

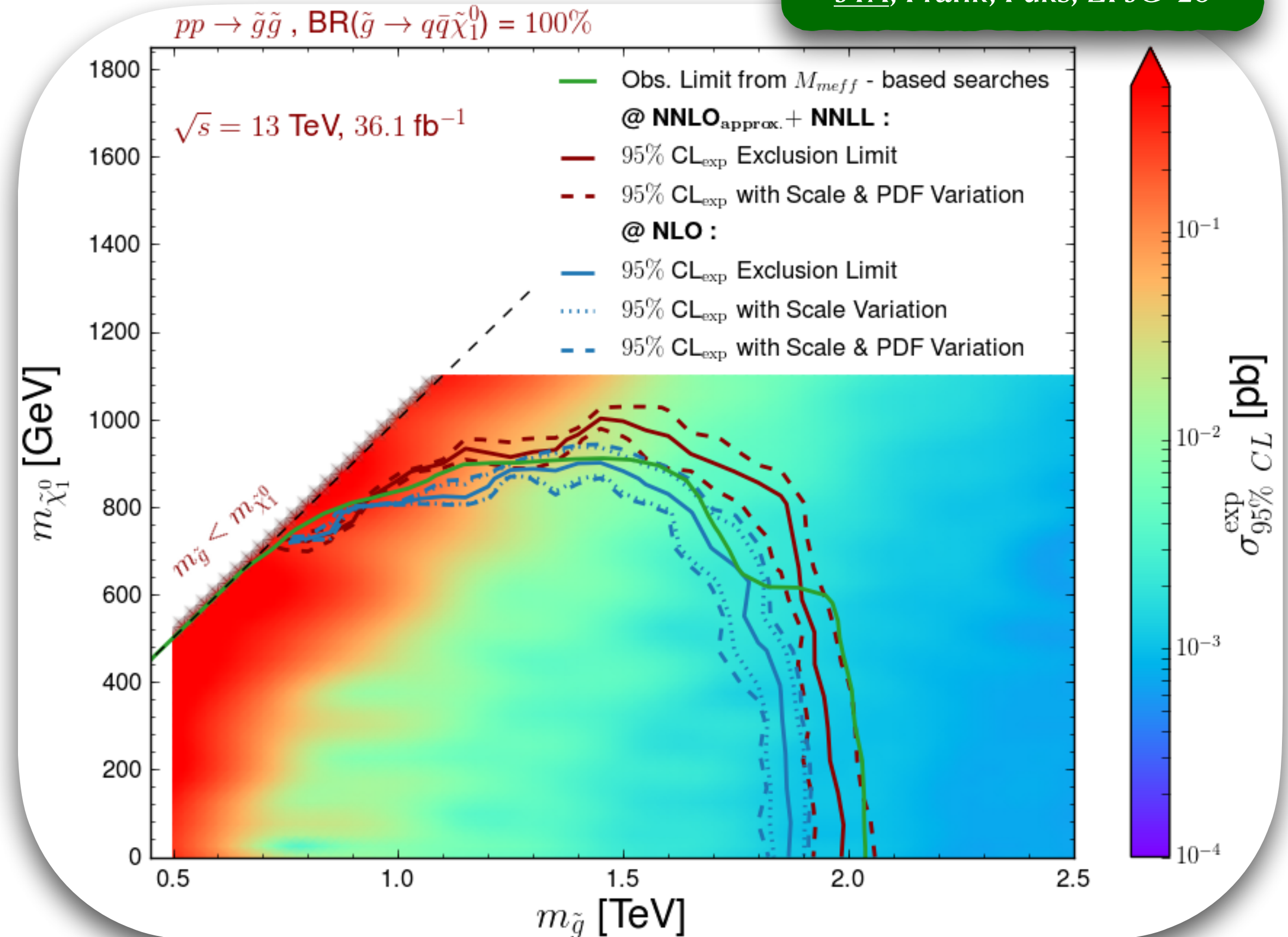
MadAnalysis 5 has ever growing Public Analysis Database, which includes various LHC analysis validated against original analysis performed by ATLAS and CMS.

WANTED: Analysis codes

Scientific reproducibility and data preservation solely depend on preserving analysis logic in a reinterpretable form. You can contribute to the HEP community by sharing the LHC recast you have implemented in the MadAnalysis 5 framework, through [Public Analysis Database \(PAD\)](#)! Please send us your analysis code, detector card, info file and validation note to be included in PAD for public use.

More information and examples can be found in the proceedings of [the second MadAnalysis 5 Workshop on LHC recasting in Korea](#). Analysis codes have been published, documented and got a DOI so that they can now be cited.

JYA, Frank, Fuks; EPJC '20

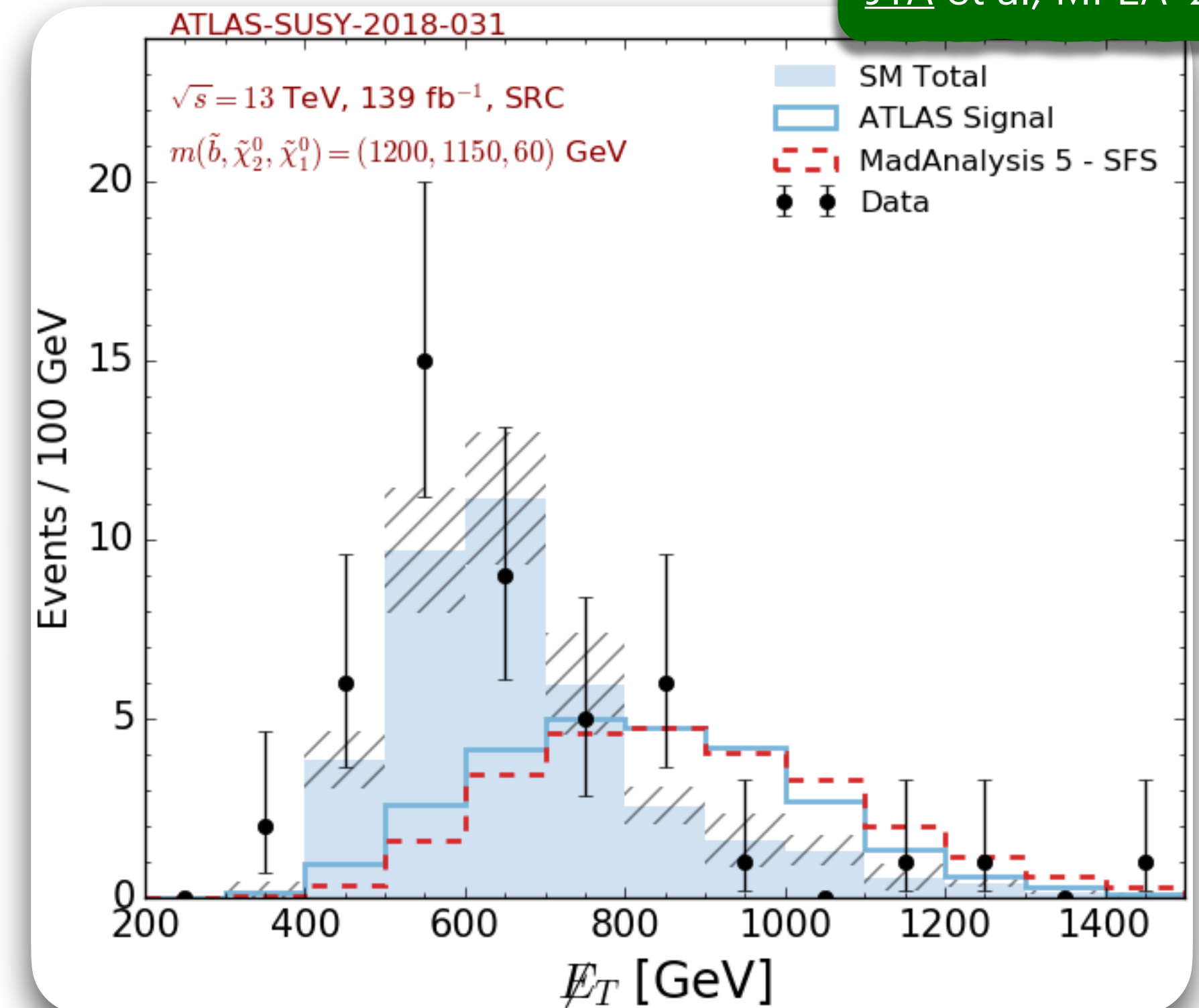


Simplified Detector Simulations

JYA, Fuks, Polcratis; EPJC '21

- Prerequisite : Beginner level English
- Simulation based on reconstruction efficiencies, four momentum smearing and flexible particle (mis)tagging for FS jets, leptons, photons & tracks.
 - Smearer
 - Reconstruction efficiencies
 - (mis)Tagger
 - Observable Scaling (JES, energy scaling, general scaling)
 - Substructure smearing
- Fully integrated to Public Analysis Database `ma5> install PAD`

JYA et al; MPLA '21



Q

Where can I find more information for normal mode functionalities?

- Visit [MadAnalysis 5 home page](#), there you can find a [reference card](#) for normal mode.

The screenshot shows the MadAnalysis 5 Wiki page. A callout bubble points to the 'Reference cards' section, which contains three links: 'Reference for normal mode (for MadAnalysis 5 v1.6): [here](#)', 'Reference for normal mode (for MadAnalysis 5 v1.8): [here](#)', and 'Reference for expert mode (for MadAnalysis 5 v1.6): [here](#)'. Another callout bubble points to the 'Documentation' section, which lists various documents and external add-ons.

MadAnalysis 5

Overview Code Bugs Blueprints Translations Answers

Registered 2013-04-13 by [Eric Conte](#)

MadAnalysis 5 is a framework for phenomenological investigations at particle colliders. Based on a C++ kernel, this program allows to efficiently perform, in a straightforward and user-friendly fashion, sophisticated physics analyses of event files such as those generated by a large class of Monte Carlo event generators.

MadAnalysis 5 can also be used for the recasting of existing LHC analyses. These features are documented on the MA5 PAD (public analysis database), together with instructions to implement new analyses (see <http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase>).

Download

The latest stable version of the MadAnalysis 5 package can be obtained in two ways:

- directly from the Bazaar versioning system by typing in a shell:
`bzr branch lp:madanalysis5`
- as a tarball (to be downloaded from the right side of this page).

Installation

MadAnalysis 5 requires several external libraries in order to properly run. The following libraries have to be installed:

- Python 2.6 or a more recent version (but not the 3.X series)
- The GNU GCC compiler (or clang for Mac OS X users).
- GNU make

To benefit from all options coming with the MadAnalysis 5 program, the following libraries have to be installed:

- ROOT v5.27 or a more recent version. Root 6 is recommended.
- Zlib headers and libraries (automatic installation: typing 'install zlib' in shell)
- FastJet 3.0 or more recent (automatic installation: typing 'install fastjet' in shell)
- LaTeX and pdflatex compilers

MA5 PAD (see <http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase>)

Get Involved

[Report a bug](#)

[Ask a question](#)

[Help translate](#)

- Change details
- Sharing
- Subscribe to bug mail
- Edit bug mail

Get Involved

[Report a bug](#)

[Ask a question](#)

[Help translate](#)

Configuration Progress

[Configuration options](#)

Downloads

Latest version is v1.8

[MadAnalysis5_v1.8.tgz](#)

released on 2020-07-02

[All downloads](#)

Announcements

Q

Where can I find more information?

- Ask questions in [Launchpad](#)

Questions?