# Beyond the Standard Model phenomenology with MadAnalysis 5

Fuks Benjamin

CERN - IPHC - U. Strasbourg

E. Conte, B. Fuks & C. Wymant

HEP@VUB meeting

Jan 27-31, 2014

# From FEYNRULES to event analyses

0. Implementation of the model in FEYNRULES and generation of the Monte Carlo model files

**First talk**

1. Event generation with any Monte Carlo event generator

**Second talk**

- Both signal and backgrounds
- Precision in the normalization: (N)NLO inclusive results
- Generator choice: beware of restrictions (supported Lorentz and color structures)
  → e.g., MADGRAPH 5

2. Parton showering and hadronization
- Precision in the shapes: multiparton matrix-element merging techniques (at least at leading-order)
  → e.g., MADGRAPH 5 + PYTHIA (automated merging)

3. Detector simulation

4. Event analysis with MADANALYSIS 5
- Parton-level, hadron-level and reconstructed-level analyses (and more)

**This talk**

# Outline

# MADANALYSIS 5 in a nutshell

[ Conte, BenjFuks, Serret (CPC '13); Conte, BenjFuks (arXiv:1309.7831) ]

## What is MADANALYSIS 5?

- A framework for phenomenological analyses
- Multiple input format: STDHEP, HEPMC, LHE, LHCO, ROOT
- Any level of sophistication: partonic, hadronic, detector, reconstructed
- User friendly, flexible and fast
- Interfaces to several HEP packages to process events (fastsim, showering, clustering, *etc.*)

➡ Professional analyses in an easy way
➡ No limit on the analysis complexity

## Two modules

- A PYTHON command line interface (interactive and soon independent of ROOT)
- A C++/ROOT core module, SAMPLEANALYZER

## Normal mode

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

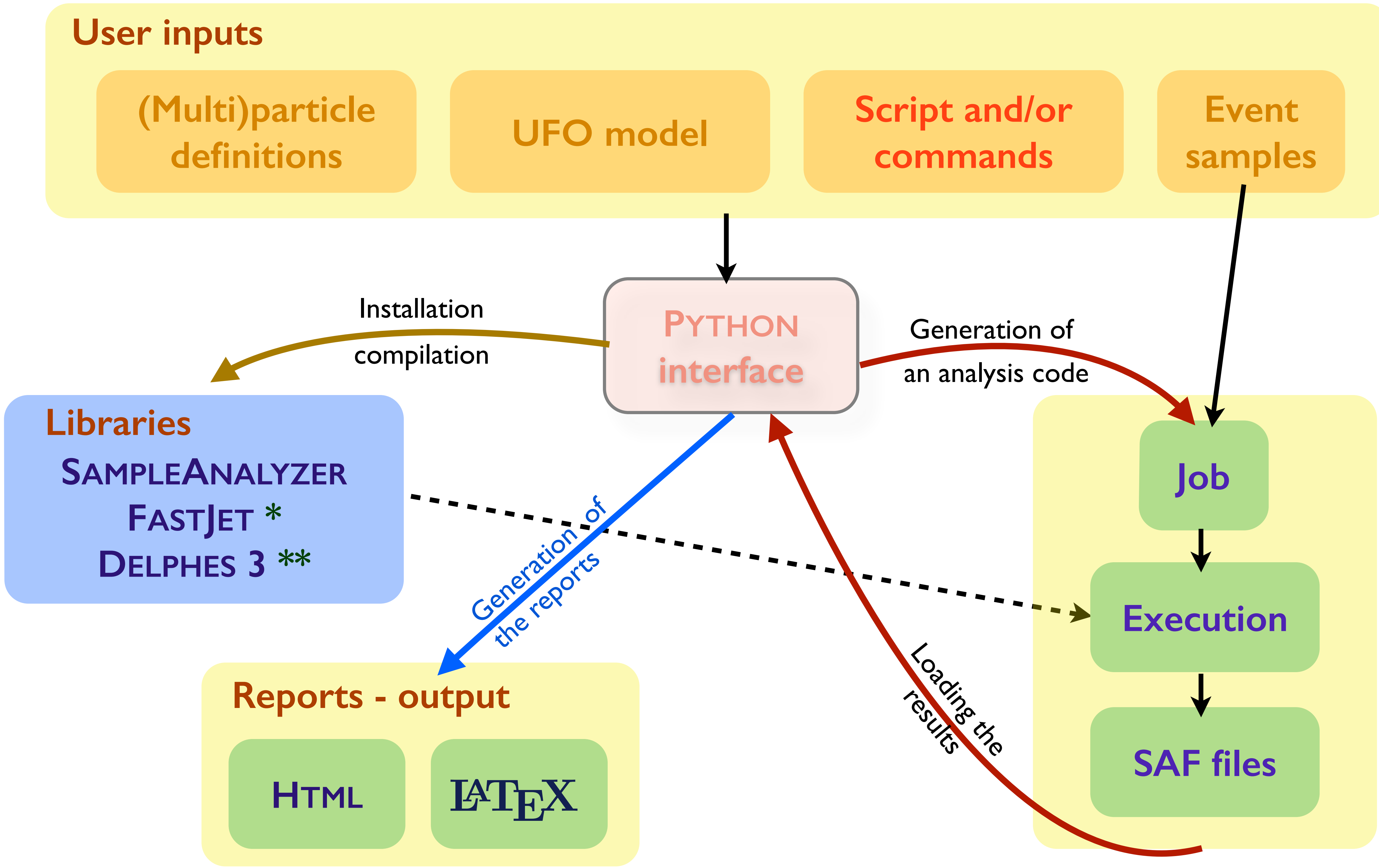
## Expert mode (not covered here)

[ Conte, Dumont, Fuks, Wymant (*in prep.*) ]

- C++/ROOT programming within the SAMPLEANALYZER framework
- New release very soon (multi-analysis, new objects, etc.)

# MADANALYSIS 5: normal running mode

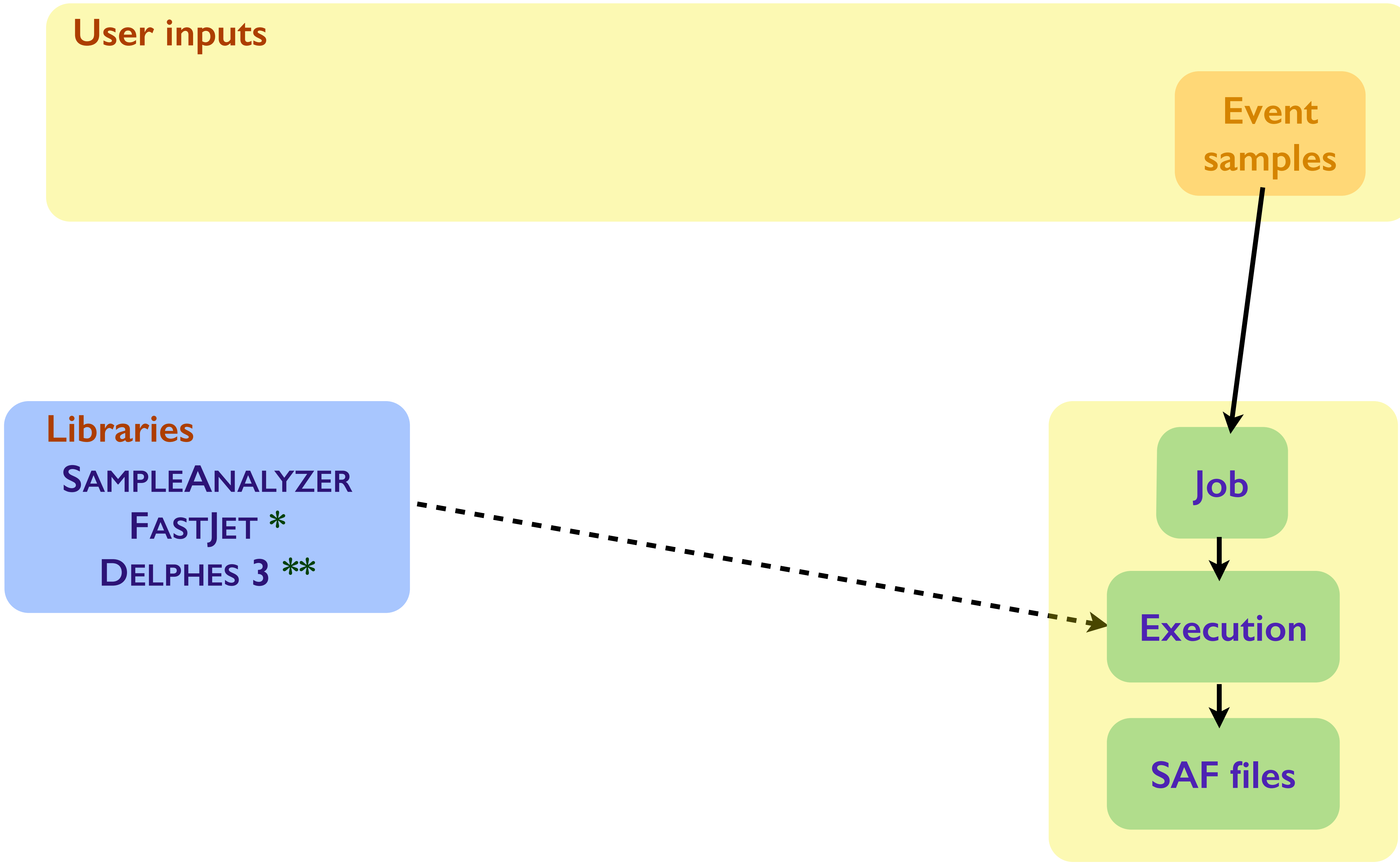
[ Conte, BenjFuks, Serret (CPC '13); Conte, BenjFuks (arXiv:1309.7831) ]




* [ Cacciari, Salam (PLB '06) ]
** [ de Favareau, Delaere, Demin, Giammanco, Lemaitre, Mertens, Selvaggi (arXiv:1307.6346) ]

# MADANALYSIS 5: expert running mode

[ Conte, Dumont, Fuks, Wymant (*in prep.*) ]




* [ Cacciari, Salam (PLB '06) ]
** [ de Favareau, Delaere, Demin, Giammanco, Lemaitre, Mertens, Selvaggi (arXiv:1307.6346) ]

# Basic concepts (I)

## Looking for help...

- ✤ In-line help from the command-line interface
- ✤ Auto-completion using the tab key

```
ma5>help

Documented commands (type help <topic>):
========================================
EOF               display_multiparticles  history  plot    reset     shell
define            display_particles       import   quit    resubmit  submit
display           exit                    install  reject  select    swap
display_datasets  help                    open     remove  set

ma5>
```

## Datasets

- ✤ Events samples are defined through a label.
- ✤ Supported file formats: LHE, STDHEP, HEPMC, LHCO, ROOT
- ✤ Several samples can be grouped (e.g., to increase statistics)
- ✤ Wildcards can be used

```
ma5>import ttbar* as top-antitop
   -> Storing the file 'ttbar.hep.gz' in the dataset 'top-antitop'.
   -> Storing the file 'ttbar2.hep.gz' in the dataset 'top-antitop'.
ma5>import Wjets.lhe.gz as Wboson
   -> Storing the file 'Wjets.lhe.gz' in the dataset 'Wboson'.
ma5>import VV.hep as diboson
   -> Storing the file 'VV.hep' in the dataset 'diboson'.
```

# Basic concepts (2)

## Particles and multiparticles

- ✤ Particles are defined through their PDG code.
- ✤ One can associate labels with particles (makes our lives easier)
- ✤ One can define multiparticles
- ✤ Default: SM + MSSM (as in MADGRAPH 5) + invisible + hadronic
- ✤ Can be defined from a UFO model

```
ma5>define TheMuon = 13
ma5>define TheAntiMuon = -13
ma5>define AllMuon = TheMuon TheAntiMuon
ma5>display l+
   The multiparticle 'l+' is defined by the PDG-ids -13 -11.
ma5>display e+
   The particle 'e+' is defined by the PDG-id -11.
ma5>display invisible
   The multiparticle 'invisible' is defined by the PDG-ids -16 -14 -12 12 14 16 1000022.
ma5>remove TheMuon
ma5>display TheMuon
 ** ERROR: no object called 'TheMuon' found.
```

# Basic concepts (3)

## Histograms - the command *plot*

- Typing plot implies the creation of an histogram (check the *display* command once created)
- Global observables: related to the full event (MET, $H_T$, *etc.*)
- Properties of a particle species (the $p_T$ of the jets, *etc.*)
- Particle ordering can be used
- Particles can be combined
- Log scales can be employed
- Different ways to normalize the histogram
- Virtual particle properties can be studied

```
ma5>plot MET [
ETAordering     PTordering     PZordering     allstate      interstate    normalize2one
ETordering      PXordering     Pordering      finalstate    logX          stack
Eordering       PYordering     ]              initialstate  logY          superimpose
ma5>plot MET [ logY ]
ma5>plot N(mu)
ma5>plot PT( mu[1] )
ma5>plot ETA(t) [ interstate ]
ma5>plot M(t t~)
ma5>plot dPHI( mu[1] mu[2] ) [logX logY]
```

# Basic concepts (4)

## Selection cuts - the commands *reject/select*

- ✤ Events can be selected/rejected whether a condition is satisfied or not
- ✤ Particles can be selected/rejected from the analysis whether a condition is satisfied or not

```
ma5>reject MHT < 200
ma5>select N(j) > 3
ma5>reject (j) PT < 20
ma5>reject (j) DELTAR(mu) < 0.4
```

## Executing the analysis - the command *submit*

- ✤ Create a C++ code with the analysis
- ✤ Create all the histograms
- ✤ Apply all the cuts
- ✤ Generate the reports (*open* to open them)

```
ma5>submit
   Creating folder 'ANALYSIS_0'...
   Copying 'SampleAnalyzer' source files...
   Inserting your selection into 'SampleAnalyzer'...
   Writing the list of datasets...
   Writing the command line history...
   Creating Makefiles...
   Compiling 'SampleAnalyzer'...
   Linking 'SampleAnalyzer'...
   Running 'SampleAnalyzer' over dataset 'top-antitop'...
    ************************************************
    * SampleAnalyzer for MadAnalysis 5 - Welcome.
    * Initializing all components
      - version: 1.1.8 (2013/08/06)
      - general: everything is default.
      - extracting the list of event samples...
      - analyzer 'MadAnalysis5job'
```

# Outline

# The event samples analyzed in this talk

- Setup for this talk
  - LHC collider at a center-of-mass energy of 8 TeV, 20 $fb^{-1}$
  - No lepton cut (pseudorapidity, transverse momentum, *etc.*)
  - Jet cuts: $p_T$ > 20 GeV, $\Delta R_{jj}$ > 0.4, no pseudorapidity cut
  - Hadron-level simulation

- Standard Model background for a dilepton + missing energy + jets signature
  - top-antitop + jets: two leptonic decays
  - WW + jets: two leptonic decays
  - single top (tW) + jets: two leptonic decays
  - ZZ + jets: one leptonic and one invisible decay

- We include a dilepton + missing energy + jets signature new physics signal
  - Disclaimer: we do not care about its nature

- Cross section for the Standard Model background
  - NNLO: top-antitop pairs (≈ 27 pb)
  - NLO + leading NNLO contributions: single top (≈2.5 pb)
  - NLO: diboson (≈5.8 pb for WW and ≈0.3 pb for ZZ)
- Multiparton matrix element merging: up to two jets

# Jet clustering (I)

## ✦ The output of PYTHIA is non-practical for an analysis

- It contains tons of hadrons
- We employ jets rather than each individual hadron
- Jets have to be reconstructed
- The event file is non-readable with human eyes (STDHEP)
- The event file size is very large

## ✦ Jet reconstruction with FASTJET

- Large selection of jet algorithms ($k_T$, anti-$k_T$, *etc.*)
- FASTJET can be used within MADANALYSIS 5
- If FASTJET is installed on the system, ready-to-be-used by MADANALYSIS 5; otherwise:

```
ma5>install fastjet
```

## ✦ Jet reconstruction with MADANALYSIS 5 and FASTJET

- The output can be saved to a LHE or a LHCO file (*set main.outputfile* = ...)
- Human-readable, smaller file size
- Can be reemployed later

# Jet clustering (2)

✦ Jet reconstruction with MADANALYSIS 5 (and FASTJET) for the five considered samples

- ✤ The four background (top-antitop, WW, single top and ZZ) and the signal samples
- ✤ MADANALYSIS 5 must be run in the reconstructed mode: ./bin/ma5 -R

```
ma5>set main.clustering.algorithm =
antikt       cambridge    cdfjetclu    cdfmidpoint  genkt        gridjet      kt           none         siscone
ma5>set main.clustering.algorithm = antikt
ma5>set main.clus
main.clustering.algorithm             main.clustering.bjet_id.misid_cjet   main.clustering.radius
main.clustering.bjet_id.efficiency    main.clustering.bjet_id.misid_ljet   main.clustering.tau_id.efficiency
main.clustering.bjet_id.exclusive     main.clustering.exclusive_id         main.clustering.tau_id.misid_ljet
main.clustering.bjet_id.matching_dr   main.clustering.ptmin
ma5>set main.clustering.radius = 0.5
ma5>set main.clustering.bjet_id.efficiency = .6
ma5>set main.clustering.bjet_id.misid_cjet = .1
ma5>set main.clustering.bjet_id.misid_ljet = .01
ma5>import ttbar.hep.gz
   -> Storing the file 'ttbar.hep.gz' in the dataset 'defaultset'.
ma5>submit
```

Many clustering algorithms available

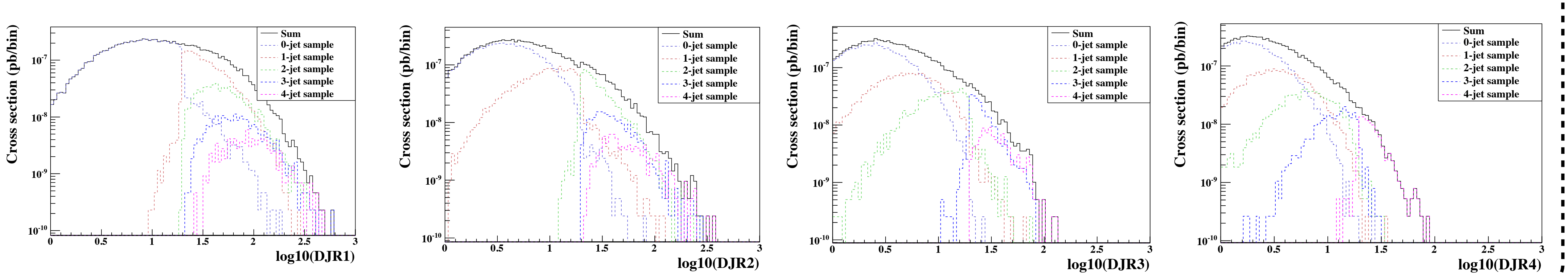
Many options (clustering parameters, b-tagging, tau-tagging, *etc.*)

# Checking the merging procedure

- **Example with a Z plus jets sample** (Z+ 0,1,2,3,4 jets at the parton level + parton showering )
  - MLM merging: removal of the double counting consistently
  - Check of that procedure: the differential jet rate (DJR) distributions
    - the scale for which one event goes from a N to a N+1 jet configuration
    - Extremely sensible to the merging procedure
    - Gives strong check of the choices for the parameters of the merging technique

- **With MADANALYSIS 5**
  - MADANALYSIS 5 must be run in hadron mode: ./bin/ma5 -H
  - The maximum value for N can be entered (default: 4)

```
ma5>set main.merging.check = true
ma5>set main.merging.njets = 4
ma5>import zjets.hep.gz
   -> Storing the file 'zjets.hep.gz' in the dataset 'defaultset'.
ma5>submit
```



# Importing and defining the samples

- Importing the reconstructed LHE samples, setting their properties
  - We define the type (signal or background) of each dataset
  - We assign the cross section associated with each dataset (necessary for a correct normalization)

```
ma5>import samples/signal.lhe as signal
   -> Storing the file 'signal.lhe' in the dataset 'signal'.
ma5>import samples/t_tw.lhe.gz as singletop
   -> Storing the file 't_tw.lhe.gz' in the dataset 'singletop'.
ma5>import samples/ttb2.lhe.gz as ttbar
   -> Storing the file 'ttb2.lhe.gz' in the dataset 'ttbar'.
ma5>import samples/ww2l2v.lhe.gz as ww
   -> Storing the file 'ww2l2v.lhe.gz' in the dataset 'ww'.
ma5>import samples/zz2l2v.lhe.gz as zz
   -> Storing the file 'zz2l2v.lhe.gz' in the dataset 'zz'.
ma5>set ttbar.xsection = 27
ma5>set ww.xsection = 5.8
ma5>set zz.xsection = 0.3
ma5>set singletop.xsection = 2.5
ma5>set signal.xsection = 0.021
ma5>set ww.type = background
ma5>set zz.type = background
ma5>set singletop.type = background
ma5>set ttbar.type = background
```

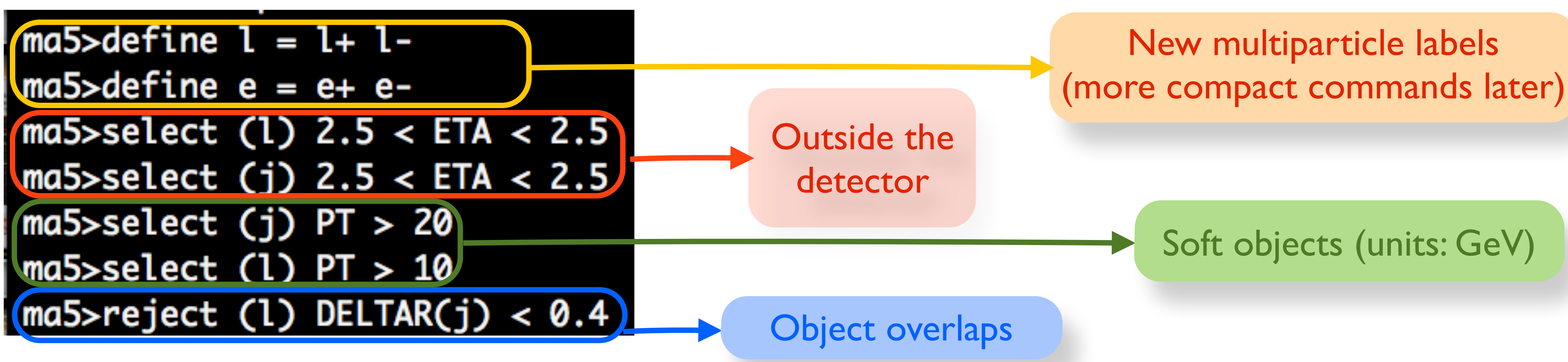
Importing the samples

Cross sections in pb

Signal and background definitions

# Getting closer to the detector...

- We have not simulated any detector response
  - Include reasonable selections getting us close to the experiment
  - Soft objects are not detected
    - Remove from each event any soft jet, lepton, *etc*.
  - Objects lying outside the detector are not detected
    - Remove from each event any jet, lepton, *etc*, whose pseudorapidity is too large.
  - Remove objet overlaps
    - Any charged lepton too close to a jet is removed (we are interested in isolated leptons)



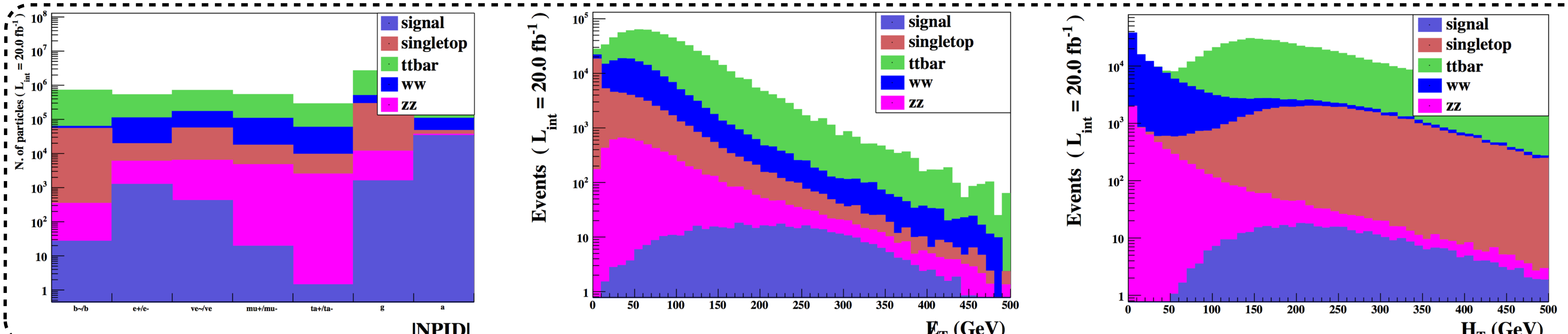
# Investigating some global event properties

- Some observables are related to the full event (called global)
  - Missing and visible energy (*MET, TET*)
  - Missing and visible hadronic energy (*MHT, THT*)
  - The partonic center-of-mass energy (*SQRTS*)
  - The $\alpha_T$ variable (*ALPHAT*): depends on the missing energy, $H_T$ and jet configuration
  - The particle content of the event (*NPID, NAPID, N*)

- General setup for the histograms
  - The luminosity in fb$^{-1}$ (*set main.lumi = ...*)
  - Histogram format (*set main.stacking_method = …; etc.*)

```
ma5>set main.stacking_method = stack
ma5>set main.lumi = 20
ma5>plot NAPID [logY]
ma5>plot MET 50 0 500 [logY]
ma5>plot THT 50 0 500 [logY]
```

- Executing the analysis and browsing the results
  - The command *submit* (the progress can be followed on the screen)
  - The command *open* (open a webpage with the report containing all results)



# Selection requirements

✦ Four selections:

- ✤ The missing energy must be greater than 150 GeV
- ✤ The $H_T$ must be greater than 100 GeV
- ✤ We want exactly two charged leptons
- ✤ We want two or three light jets

```
ma5>reject MET < 150
ma5>select THT > 100
ma5>reject N(l) != 2
ma5>select 2 <= N(j) <= 3
```

✦ MADANALYSIS 5 provide the efficiencies for each sample and for each selection

**Cut: reject MET < 150.0**

| Dataset | Events kept: K | Rejected events: R | Efficiency: K / (K + R) | Cumul. efficiency: K / Initial |
|---|---|---|---|---|
| signal | 297.78 +/- 9.31 | 122.22 +/- 9.31 | 0.7090 +/- 0.0222 | 0.7090 +/- 0.0222 |
| singletop | 1801.5 +/- 41.7 | 48198.5 +/- 41.7 | 0.036030 +/- 0.000833 | 0.036030 +/- 0.000833 |
| ttbar | 67347 +/- 242 | 472652 +/- 242 | 0.12472 +/- 0.00045 | 0.12472 +/- 0.00045 |
| ww | 5804.9 +/- 74.3 | 110195.1 +/- 74.3 | 0.05004 +/- 0.00064 | 0.05004 +/- 0.00064 |
| zz | 598.5 +/- 23.2 | 5401.5 +/- 23.2 | 0.09974 +/- 0.00387 | 0.09974 +/- 0.00387 |

Investigating the best analysis

# The signal over background ratio

✦ The user can enter any formula (using *S*, *B*, *ES*, *EB*)

```
ma5>set main.SBratio = "S/sqrt(S+B)"
Checking the formula ...
Formula corresponding to the uncertainty calculation has been found and set to the variable main.SBerror :
  1./pow(S+B,3./2.)*sqrt((S+2*B)**2*ES**2+S**2*EB**2)
ma5>
```

## Cut-flow chart

- How to compare signal (S) and background (B): S/sqrt ( S+B ) .
- Associated uncertainty: 1./pow(S+B,3./2.)*sqrt((S+2*B)**2*ES**2+S**2*EB**2).

| Cuts | Signal (S) | Background (B) | S vs B |
|---|---|---|---|
| Initial (no cut) | 420 | 712000 | 0.498 |
| Cut 1 | 4.20e+02 +/- 2.38e-07 | 712000 | 4.98e-01 +/- 5.65e-10 |
| Cut 2 | 4.20e+02 +/- 2.38e-07 | 712000 | 4.98e-01 +/- 5.65e-10 |
| Cut 3 | 4.20e+02 +/- 2.38e-07 | 712000 | 4.98e-01 +/- 5.65e-10 |
| Cut 4 | 4.20e+02 +/- 2.38e-07 | 712000 | 4.98e-01 +/- 5.65e-10 |
| Cut 5 | 4.20e+02 +/- 2.38e-07 | 712000 | 4.98e-01 +/- 5.65e-10 |
| Cut 6 | 4.20e+02 +/- 2.38e-07 | 712000 | 4.98e-01 +/- 5.65e-10 |
| Cut 7 | 297.78 +/- 9.31 | 75552 +/- 258 | 1.0812 +/- 0.0676 |
| Cut 8 | 288.16 +/- 9.51 | 69996 +/- 249 | 1.0869 +/- 0.0717 |
| Cut 9 | 240.2 +/- 10.1 | 32182 +/- 174 | 1.334 +/- 0.112 |
| Cut 10 | 187.7 +/- 10.2 | 15172 +/- 121 | 1.514 +/- 0.164 |

Not relevant (selections on objects)

Our selections improve the sensitivity

# Investigating particle properties

- **Many kinematical properties of a given particle can be studied**
  - *BETA, E, ET, ETA, GAMMA, M, MT, P, PHI, PT, PX, PY, PZ, R, THETA, Y*
  - Each of these functions take a single argument (a particle)

```
ma5>plot PT(l[1]) 50 0 500 [logY]
ma5>plot MT(j[1]) 50 0 500 [logY]
```

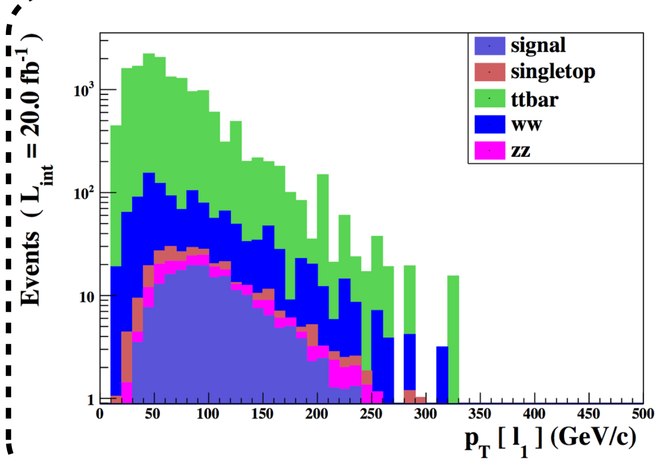
- **The particles are ordered**
  - Use squared brackets to select the right one
  - Several ordering variables are available (*PT, E, PX, etc.*)

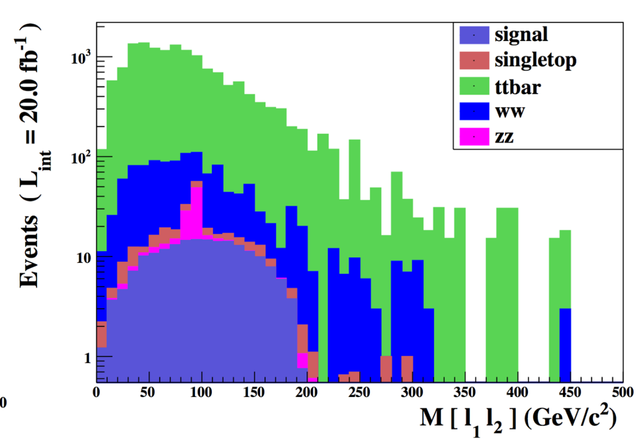
- **Combining particles**
  - Replace the argument by several particles
  - Their four-momenta are summed and the relevant observable is then computed
  - Vectorial and scalar sums/differences as well as ratios are available (*s, v, ds, dv, r* prefixes)

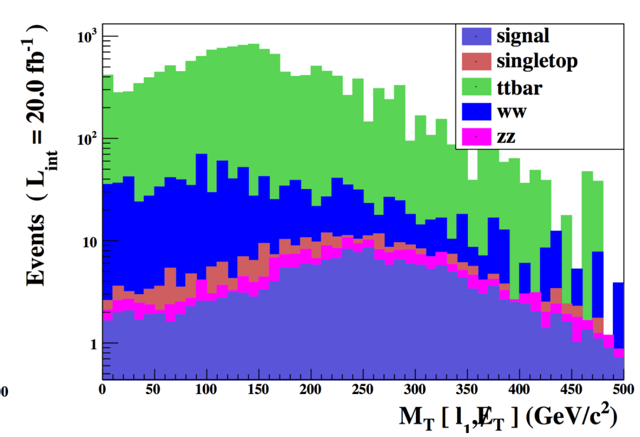
```
ma5>plot M(l[1] l[2]) 50 0 500 [logY]
ma5>plot dPHI(l[1] l[2]) 15 0 6.28 [logY]
```

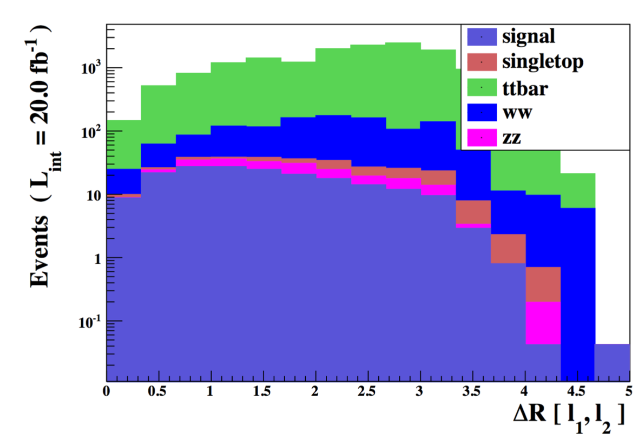
- **Two special functions**
  - *DELTAR*: takes two arguments
  - *MT_MET*: transverse mass obtained when combining one particle and the missing momentum

```
ma5>plot DELTAR(l[1],l[2]) 15 0 5 [logY]
ma5>plot MT_MET(l[1]) 50 0 500 [logY]
ma5>plot MT_MET(j[2]) 50 0 500 [logY]
```



# Outline

# The final words (for all three talks)

- The quest for new physics at the LHC has started
  - Rely on Monte Carlo event generators for background and signal modeling (like MADGRAPH 5)
  - Satellite tools have been intensively developed (like FEYNRULES, MADANALYSIS 5)

- FEYNRULES: http://feynrules.irmp.ucl.ac.be/
  - Straightforward implementation of new physics model in the Monte Carlo tools
  - Has its own computational modules
  - Being interfaced to NLO tools

- MADGRAPH 5 - AMC@NLO: http://launchpad.net/madgraph5
  - Event generation and cross section calculations at the parton-level
  - LO and NLO accuracy are possible
  - Contains a lot of useful packages (MADSPIN, MADWEIGHT, *etc*…)
  - Interface to showering + merging techniques

- MADANALYSIS 5: http://launchpad.net/madanalysis5/
  - Analysis of event samples generated by Monte Carlo tools
  - Useful interfaces to detector simulators, jet reconstruction tools, *etc*.

Automation and precision for new physics phenomenology are on their way

We are almost there!