MadAnalysis 5
A new framework for collider phenomenology

Benjamin Fuks (IPHC Strasbourg / Université de Strasbourg)

In collaboration with E. Conte & G. Serret.

The FeynRules/MadGraph 2012 School
on LHC Phenomenology @ Natal (Brazil)

Sep 30 - Oct 05, 2012

Outline

Comprehensive particle physics phenomenology.

1. Implementation of a new physics model in FeynRules.

   Model building.

2. Event generation with MadGraph **5**.

   Parton-level phenomenology.

3. Parton showering and hadronization with Pythia or Herwig.

   Hadron-level phenomenology.

4. Fast detector simulation with Delphes or PGS.

   Reconstructed-level phenomenology.

Need for a new framework for collider phenomenology.

- **Several levels of sophistication for phenomenological analyses**.
  * **Parton** level.
  * **Hadron** level.
  * **Reconstructed** level.
- **Analysis skeleton**.
  * **Reading** of signal and background event files.
  * Application of **selection cuts**.
  * Creation of **histograms** and **cut-flow charts**.
  * **Extraction of information on the signal** [usually swamped by backgrounds].
- **Drawbacks**.
  * The procedure above is in general based on **home-made tools**.
    - ▶**Lack of traceability**.
    - ▶**Validation of the tools**?
    - ▶**Reproducibility of the results**?
  * These tools can in general only be used at a **specific sophistication level**.
    - ▶**Lack of flexibility**.
  * These tools can in general only be used with a **specific event file format**.
    - ▶**Lack of flexibility**.

Introducing MadAnalysis 5.

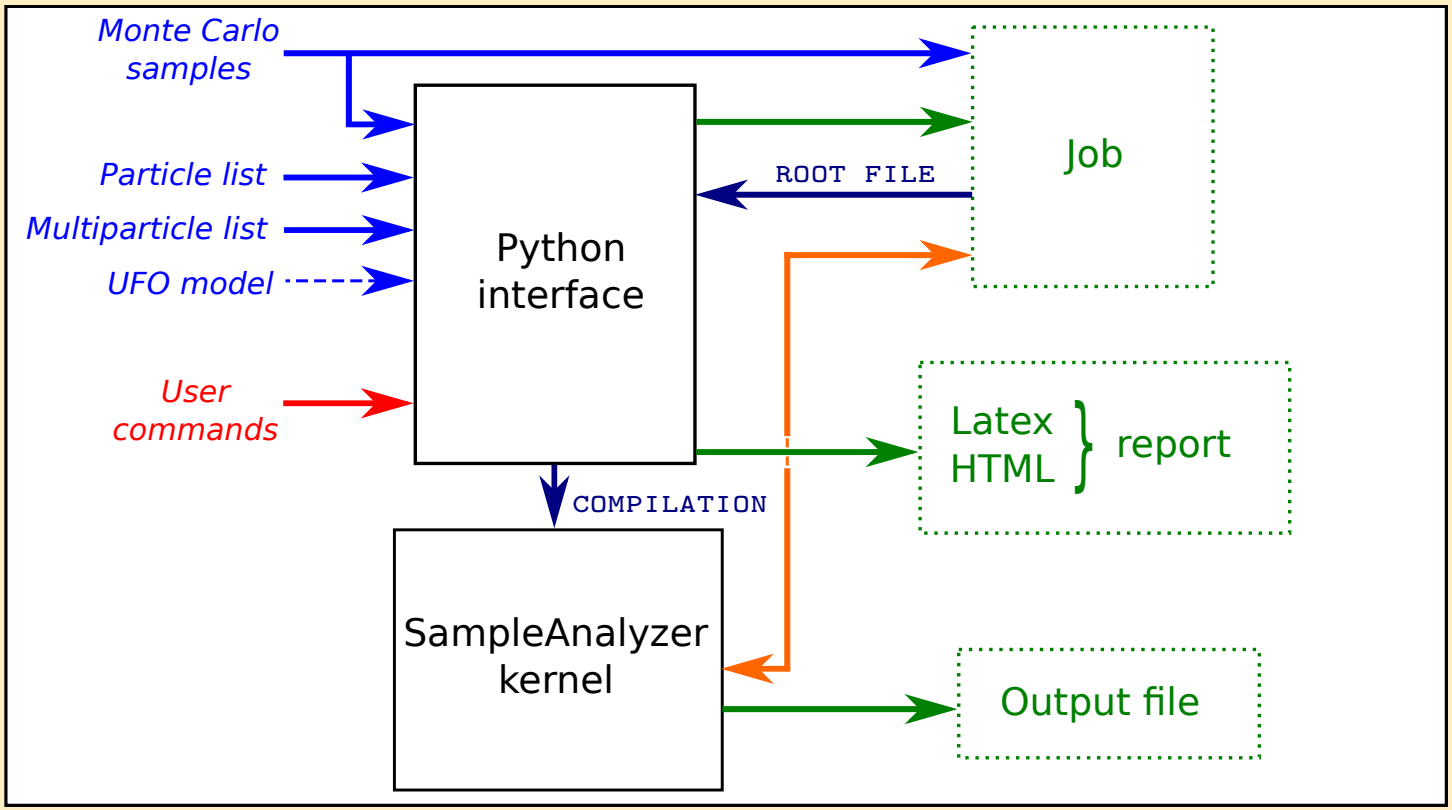
Alleviation of these issues.

- A **new unique** framework for phenomenological analyses.
  - **Any sophistication level** (parton, hadron, reconstructed).
  - **Any event file format** (StdHep, HepMC, LHE, ...).
  - **User-friendly** $\Rightarrow$ professional analyses in a simple way.
  - **Fast**: less than a minute for analyzing 100.000 events.
  - **Flexible** $\Rightarrow$ no limit on the analysis complexity.
  - **Easy to maintain**.
  - **Easy to validate**.

This framework is called MadAnalysis 5.

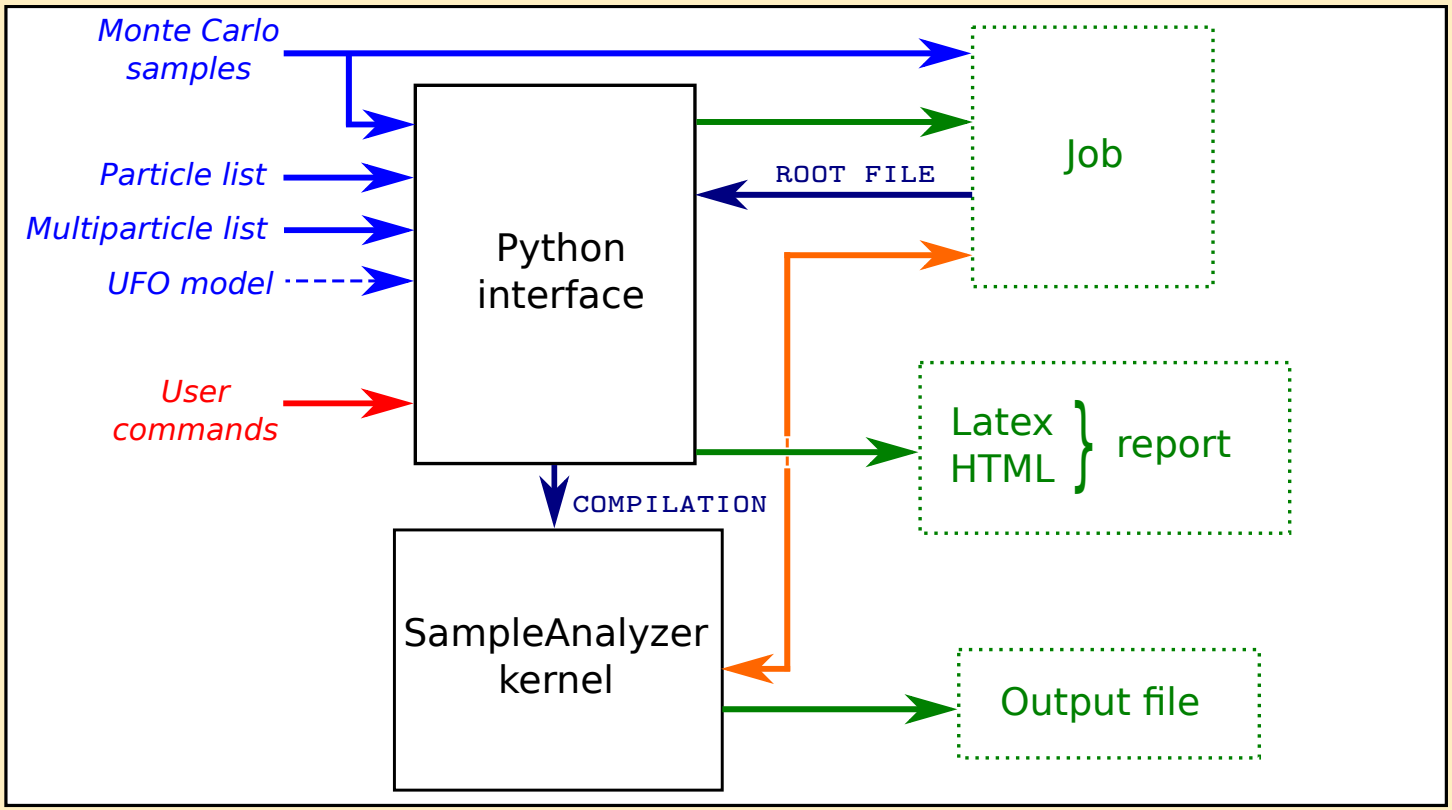
Outline.

The MadAnalysis 5 scheme.



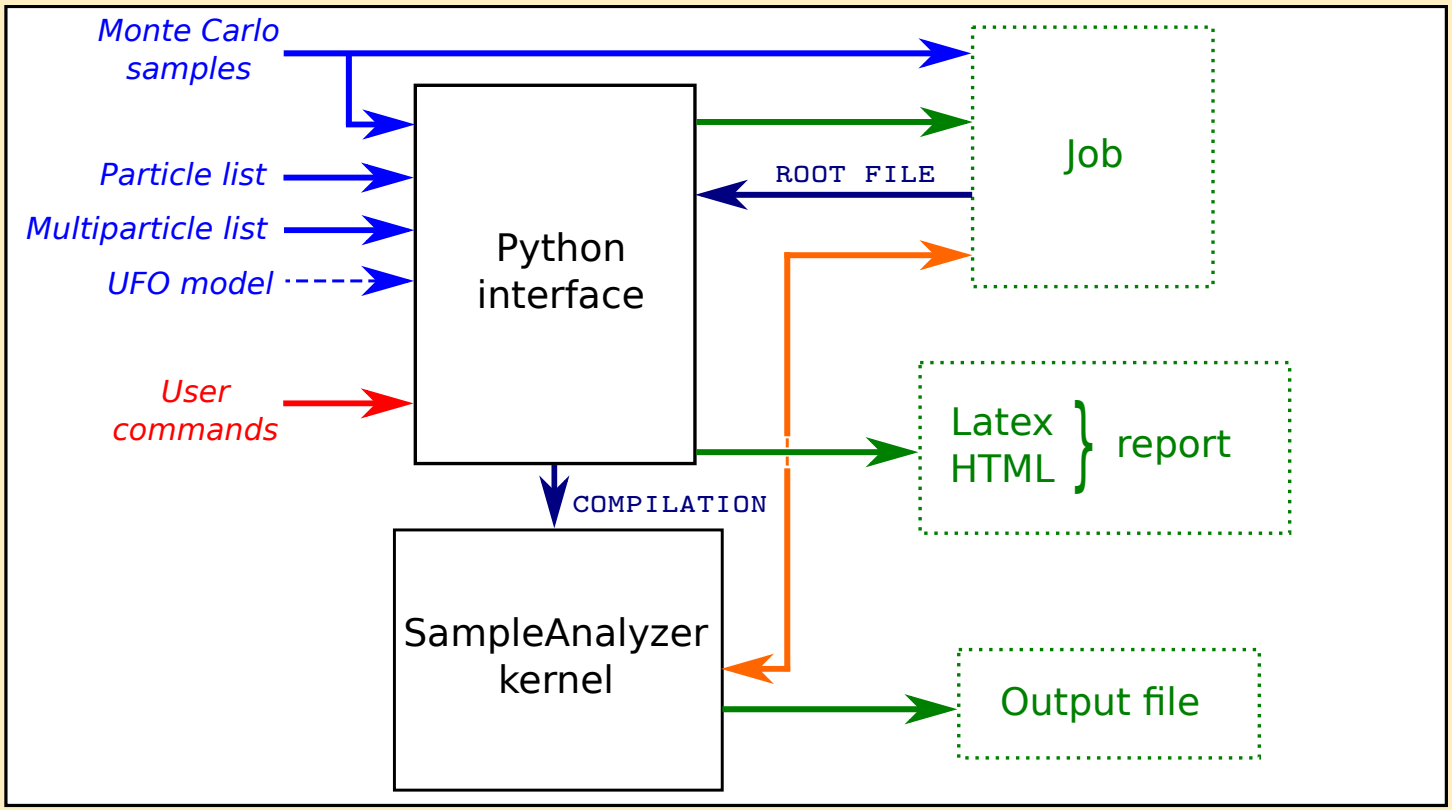
- **Two modules**.
  - A Python **command line interface**: **interactive commands**.
  - A C++/Root **module,** SampleAnalyzer: **performs the analysis**.

The MadAnalysis 5 scheme.



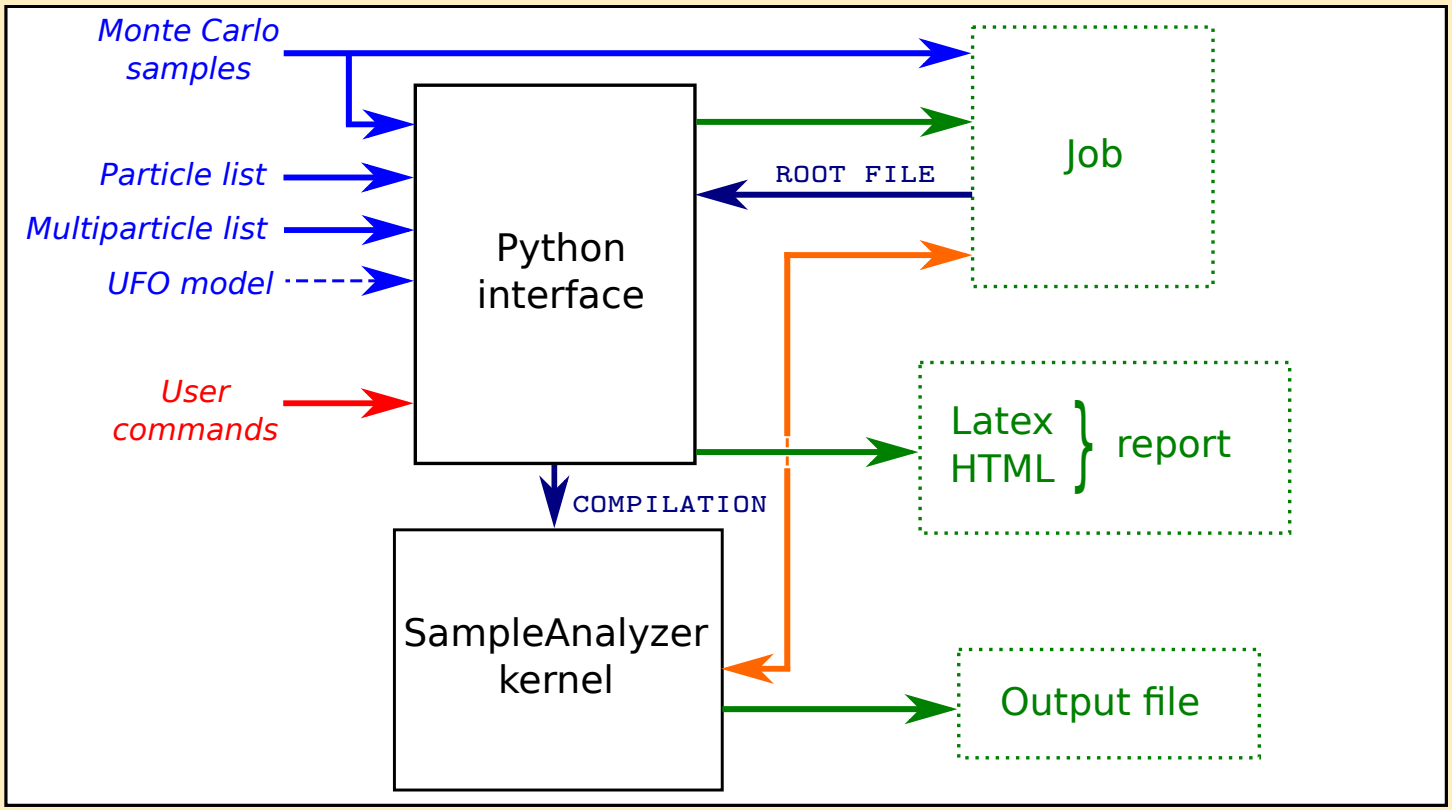
- **Normal mode of running** (user-friendly).
  - **Commands typed in the Python interface**.
  - Analysis performed **behind the scene** (black box).
  - **Human readable output**: HTML, LaTeX.

The MadAnalysis 5 scheme.



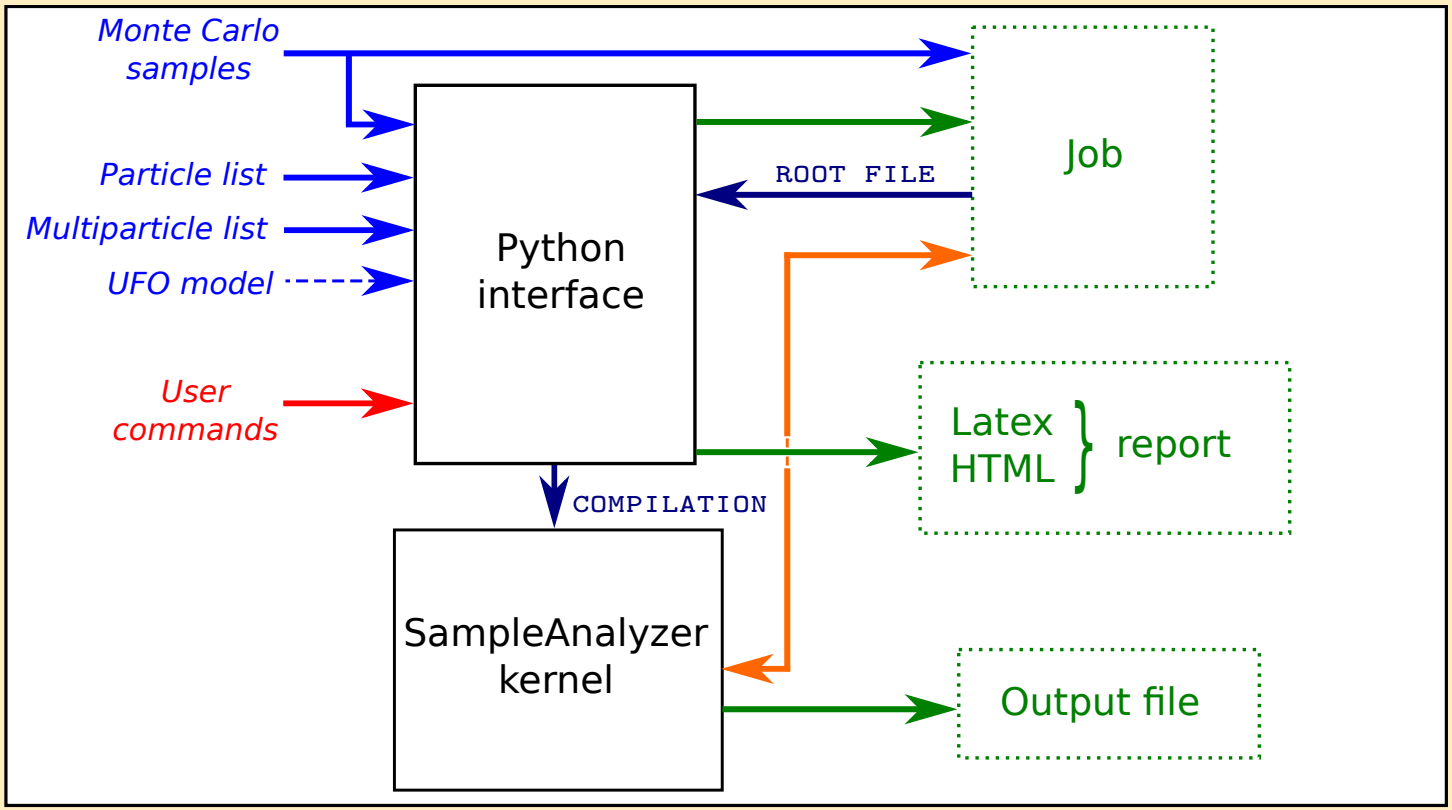
- **Expert mode** (developer-friendly; not covered oin this tutorial).
  - C++ programming within the SampleAnalyzer framework.
  - C++ **and** Root **skills required**.
  - The Python interface creates a **blank analysis as a starting point**.

The MadAnalysis 5 scheme.



- **Inputs**.
  - Monte Carlo samples (zipped or not) $\Leftrightarrow$ **datasets**.
  - **Particle and multiparticle** labels.
  - **User commands**.

The MadAnalysis 5 scheme.



- **Jobs and results**.
  - Translation of the commands by the interface ⇒ C++ **job**.
  - Uses the SampleAnalyzer kernel.
  - Generation of the **results**; conversion of the events to a **compact format**.

Basic concepts.

- **Command line interface**.
  * **In-line help**.
  * **Auto-completion**.

```
ma5> help <command>
```

- **Particles and multiparticles**.
  * Particle are defined by **labels**.
  * A label points to one or several **PDG-id(s)**.
  * **MSSM + SM labels**: automatic.
  * Can be loaded from **UFO files**.
  * Labels can be **created and deleted**.
    ▶`define` and `remove`.

```
define tau = tau+ tau-
define mytau+ = -15
remove mytau+
```

- **Datasets**.
  * A dataset is a **label**.
  * **Collects** similar event samples.
  * Treated **in the same way** by MadAnalysis 5.
  * **Formats**: LHE, LHCO, StdHep, HepMC.

```
import tt1.hep as ttbar
import tt2.hep as ttbar
import Wj1.hep as Wjets
import Wj2.hep as Wjets
```

Plots and cuts.

- **The command** `plot` (more detailed examples in the rest of the tutorial).
  - Creation of an **histogram**.
  - **Global observables** $\Leftrightarrow$ the entire event.
  - **Properties of the particles** in the event.
  - **Ordering** of the particles.
  - **Combining** particles
    - ►Sum and differences.
    - ►Vectorial or scalar.
  - Linear or logarithmic scales.

```
plot MET
plot N(mu)
plot PT(mu[1])
plot ETA(mu) [logY]
plot M(mu[1] mu[2])
plot dM(mu+ mu-)
```

- **Cuts** (more detailed examples in the rest of the tutorial).
  - **Selecting/rejecting** events.
  - **Selecting/rejecting** particles.
    - ►not rejecting the event.

```
reject MHT < 50
select (mu) PT > 50
```

- **Executing the analysis**: `submit`.
- **Reports**.
  - HTML reports.
  - LaTeX reports.

```
generate_html <dir>
generate_latex <dir>
generate_pdflatex <dir>
```

Outline.

Installation.

- **Requirements**.
  - ►These programs are assumed to be already installed.
    - ◇ Python 2.6 or a more recent version (but not the 3.X series).
      `http://www.python.org/`
    - ◇ **The GNU GCC compiler** version 4.3.0 or more recent.
      `http://gcc.gnu.org/`
    - ◇ ROOT v5.27 or a more recent version.
      ►with the Python libraries
      ```
      root-config --version
      ./configure --with-python
      ```
    - ◇ Zlib **headers and libraries**.
      `http://zlib.net/`
- **Installing MadAnalysis 5**.
  - ◇ Download:
    `http://madanalysis.irmp.ucl.ac.be`
  - ◇ **Unpacking** the tar-ball:
    ```
    mkdir madanalysis5
    cd madanalysis5
    tar xvf ma5_v1.1.2.tgz
    ```

Starting MadAnalysis 5 (1).

- **First start of MadAnalysis 5 ⇒ typing in a shell:** `bin/ma5`

```
**************************************************************
*                                                            *
*        W E L C O M E  to  M A D A N A L Y S I S  5         *
*                          ______  ______                    *
*                    /'\_/`\/\  __ \/\  ___\                 *
*                   /\      \ \ \_\ \ \ \__/                 *
*                   \ \ \__\ \ \  __ \ \___``\               *
*                    \ \ \_/\ \ \ \/\ \/\ \_\ \              *
*                     \ \_\\ \_\ \_\ \_\ \____/              *
*                      \/_/ \/_/\/_/\/_/\/___/               *
*                                                            *
*   MA5 release : 1.1.2                          2012/10/01  *
*                                                            *
*   The MadAnalysis Development Team - Please visit us at    *
*   http://madanalysis.irmp.ucl.ac.be                        *
*                                                            *
*              Type 'help' for in-line help.                 *
*                                                            *
**************************************************************
```

Starting MadAnalysis 5 (2).

- **First start of MadAnalysis 5 ⇒ testing all the dependencies.**

```
Checking ROOT libraries ...
Loading ROOT libraries ...
Checking g++ libraries ...
Checking zlib libraries ...
Checking fastjet libraries ...
 ** WARNING: FastJet configuration program is not found.
      JetClustering algorithm will be disabled.
 ** WARNING: To enable this functionnality,
      please type 'install fastjet'.
```

  - ◇ **Warning messages** are printed if relevant.
  - ◇ **FastJet is not installed here...**
    ▶to be addressed later...
  - ◇ **If you get error messages, please use the Virtual Box.**
    ▶ probably an issue with the installation of Root.

Starting MadAnalysis 5 (3).

- **First start of MadAnalysis 5 ⇒ compiling SampleAnalyzer.**

```
Checking MadAnalysis library ...
First use of MadAnalysis detected (or the library is missing)
   Creating a 'Makefile'...
   Compiling the MadAnalysis library...
How many cores would you like to use for the compilation ?
   default = max = 16
Number of cores used for the compilation = 16
   Linking the MadAnalysis library...
   Checking the MadAnalysis library presence...
************************************************************
```

  - ◇ **Compilation** of the core library.
  - ◇ **Linking** of the core library.
  - ◇ **Core library** then ready to be used.

Starting MadAnalysis 5 (4).

- **First start of MadAnalysis 5 ⇒ locating MadGraph 5.**

```
MadGraph 5 NOT found => default particle names from:
  /madanalysis5/madanalysis/input/particles_name_default.txt
84 particles have been successfully exported.
MadGraph 5 NOT found => default multiparticle definitions from:
  madanalysis5/madanalysis/input/multiparticles_default.txt
Creation of a multiparticle labelled by 'invisible'
  (related to missing energy).
Creation of a multiparticle labelled by 'hadronic'
  (related to jet transverse energy).
8 multiparticles have been successfully exported.
```

  - ◇ MadGraph 5 **not found**
    ⇒ MadAnalysis 5 used as a **standalone package**.
  - ◇ **Particle and multiparticle** labels loaded.

Installation of FastJet.

- **Typing in the interpreter**: `install fastjet`.

```
ma5>install fastjet
How many cores would you like to use for the compilation ?
   default = max=16
Answer:
Number of cores used for the compilation = 16

Testing the access to MadAnalysis 5 website ...
'tools' folder is already created
Creating temporary folder '/tmp/ma5install_bfuks' ...
1/1 Downloading the file 'fastjet.tar.gz' ...
Extracting the package ...
Configuring the package ...
Compiling the package ...
Copying headers and libraries into 'tools/fastjet' ...
Checking installation ...
Installation complete.
```

  - ◇ Fully **automated**.

Outline.

Setup of the analysis.

- **Sample(s) to be analyzed:**
  - * **Test samples provided with MadAnalysis 5**:
    ►`install samples`
  - * **Your favorite sample(s)**:
    ►Please generate it yourself
- **In these slides: four different samples.**
  - ◇ $t\bar{t}$ **production** (two event files; simplified LHE plus HEP).
    ►**dileptonic mode**: LHC-8, merging up to 2 extra jets.
    ►**semileptonic mode**: LHC-8, merging up to 2 extra jets.
  - ◇ **Z+jets; dileptonic and invisible modes** (simplified LHE plus HEP).
    ►LHC-8, merging up to 4 extra jets.
  - ◇ **W+jets** (simplified LHE plus HEP).
    ►LHC-8, merging up to 4 extra jets.

Try do produce similar analyses with your own samples

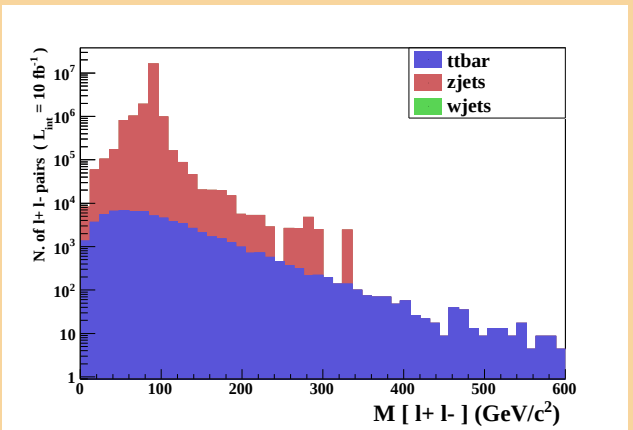
OR

ask for the four samples above.

Particle properties.

- **Kinematical distributions related to particle species.**
  * **Available observables**:
    `BETA, DELTAR, E, ET, ETA, GAMMA, M, MT, P, PHI, PT, PX, PY, PZ, R, THETA, Y.`

```
import ttbar_2l.lhe.gz as ttbar
import ttbar_1l.lhe.gz as ttbar
import wjets.lhe.gz as wjets
import zjets.lhe.gz as zjets
plot M(l+ l-) [logY]
submit mydir
generate_html mydir_html
open mydir_html
```



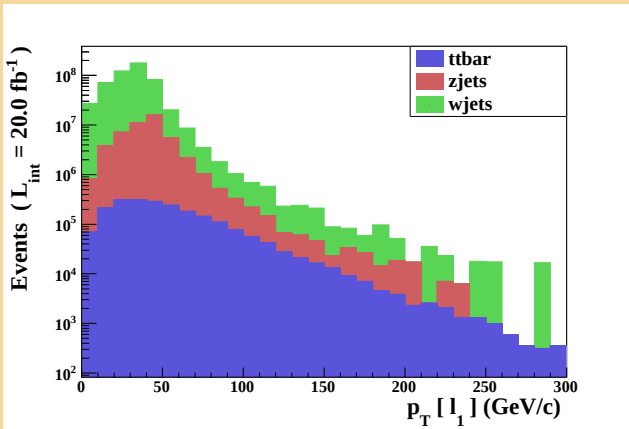
Exercises

* Where are the W-events?
* How to get rid of the Drell-Yan background?

Leading lepton properties.

- **Particle ordering**.
  - Can be access with the **squared brackets** `[<i>]` .
  - Several possible **ordering variables**. `E`, `ET`, `ETA`, `P`, `PT`, `PX`, `PY`, `PZ`.
- Check the **transverse momentum** of the **leading lepton**, using **energy ordering**.
- **Other features**: **cross sections**, **integrated luminosity**.

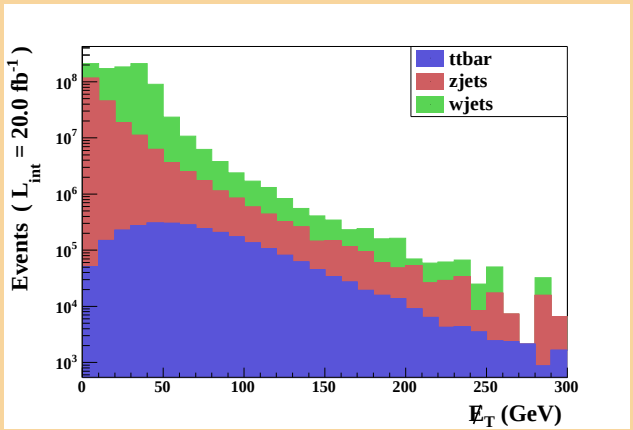
```
set ttbar.xsection = 139.6
set wjets.xsection = 35678
set zjets.xsection = 10319
set main.lumi = 20
define l = l+ l-
plot PT(l[1]) 20 0 200 [logY]
set selection[2].rank = PTordering
resubmit
generate_html mydir_html
open mydir_html
```



Global event observables.

- **Global event kinematical observables**.
  * **Missing and visible energy** of the event `MET`, `TET`.
  * **Missing and visible hadronic energy** of the event `MHT`, `THT`
  * **Partonic center-of-mass energy** `SQRTS`.

```
plot MET 30 0 300 [logY]
resubmit
generate_html mydir_html
open mydir_html
```



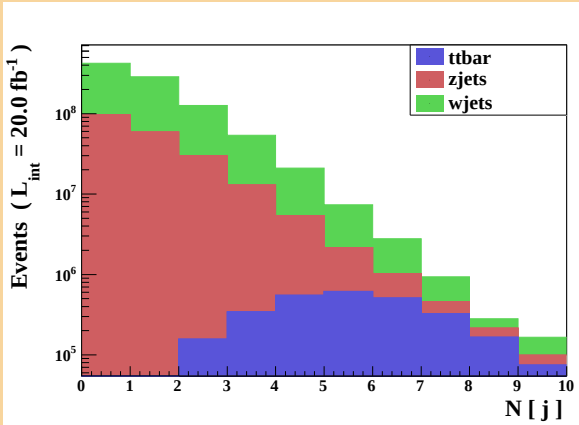
Exercises

* Do we have enough statistics?

Multiplicities.

- **Particle content**.
  - * **Particle content** of the event `NPID`, `NAPID`.
  - * **Particle multiplicity** `N`

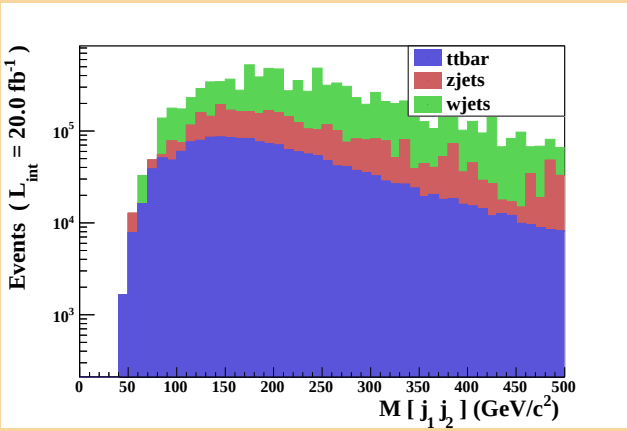
```
define j = j b b~
plot N(j)
resubmit
generate_html mydir_html
open mydir_html
```



Selection cuts.

- **Cuts**.
  - ◇ Through the commands `select` and `reject` followed by a condition.
  - ◇ **Particle candidates**.
    - * **Lepton candidates:** $p_T > 10$ GeV.
    - * **Jet candidates:** $p_T > 20$ GeV.
  - ◇ **Events**.
    - * **Selected events:** $H_T > 200$ GeV.

```
select (l) PT > 10
reject (j) PT < 20
reject THT < 200
plot M(j[1] j[2])
resubmit
generate_html mydir_html
open mydir_html
```



Signal over background ratios.

- **Automated computation of the signal over background ratio**.
  * Samples can be tagged as `signal` or `background`.
  * **Formula** for the signal over background ratio can be provided.
  * **Automatic cut-flow chart** with uncertainties.

```
set wjets.type = background
set zjets.type = background
set main.SBratio = 'S/B'
set main.SBerror = '1./(B**2)*sqrt(B**2*ES**2+S**2*EB**2)'
resubmit
generate_html mydir_html
open mydir_html
```

Cuts	Signal (S)	Background (B)	S vs B
Initial	2792000	919940000	0.00303
cut 1	2792000	919940000 +/- 0.000173	3.034981e-03 +/- 5.7e-16
cut 2	2792000	919940000 +/- 0.000173	3.034981e-03 +/- 5.7e-16
cut 3	1928561 +/- 772	9583745 +/- 3079	0.201233 +/- 0.000103

Outline.

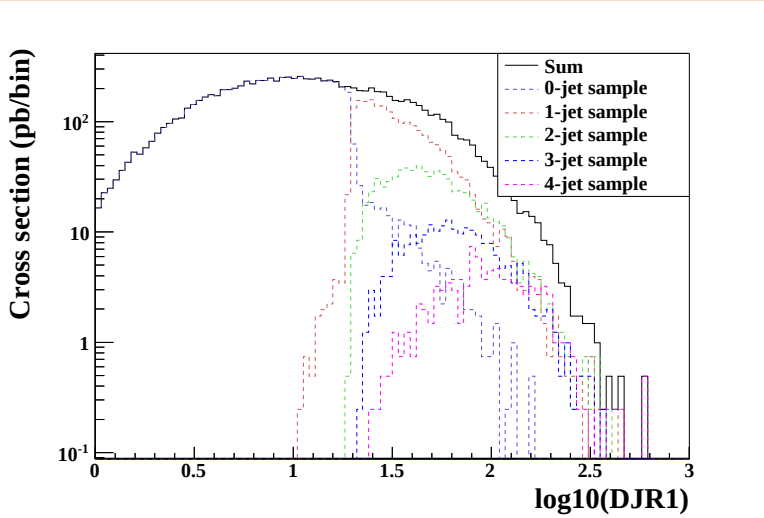
Checking the merging procedure (1).

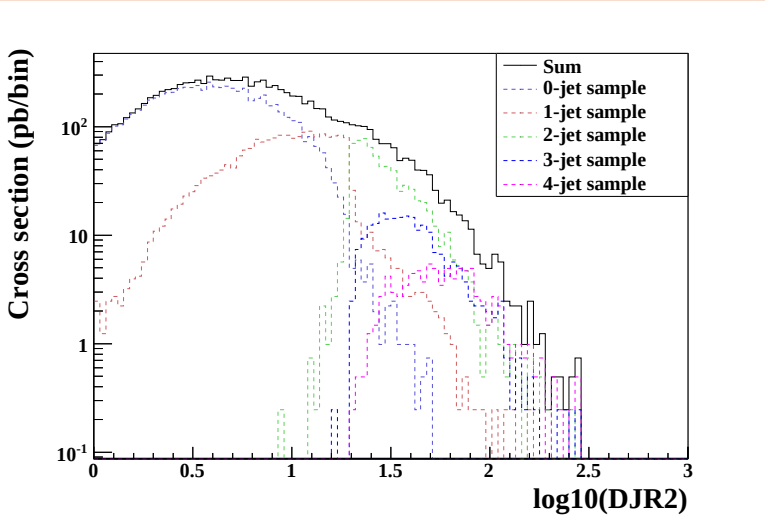
- **Merging matrix-elements with 0,1,2,3,... extra jets**.
  - ◇ Study of the smoothness of the **differential jet rate distributions**.
    - ▶ The scale for which an event goes from a $N \rightarrow N+1$ jet configuration.
    - ▶ Extremely **sensible to the merging procedure**.
  - ◇ This **validates** the choices for the merging parameters.
  - ◇ See Fabio's lecture.
- **Running MadAnalysis 5 in hadron-level mode:** `bin/ma5 -H`

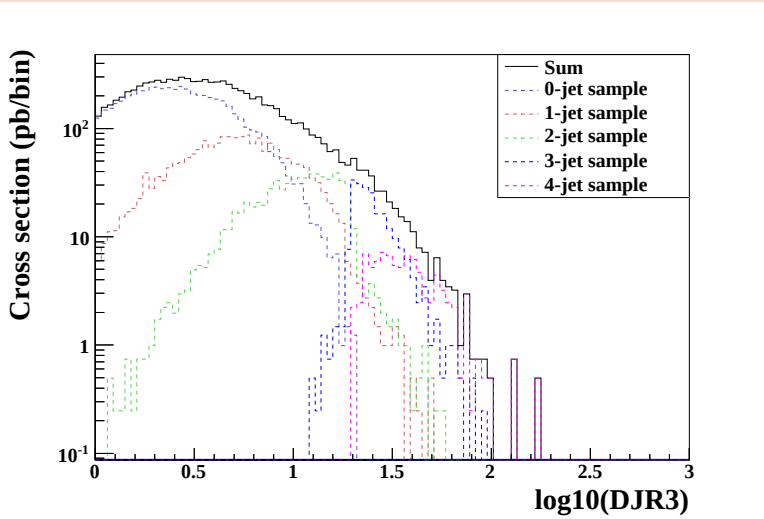
```
import zjets.hep.gz as zjets
set zjets.xsection=10319
set main.lumi = 20
set main.matching.check = true
set main.matching.njets = 4
submit mydir
generate_html mydir_html
open mydir_html
```

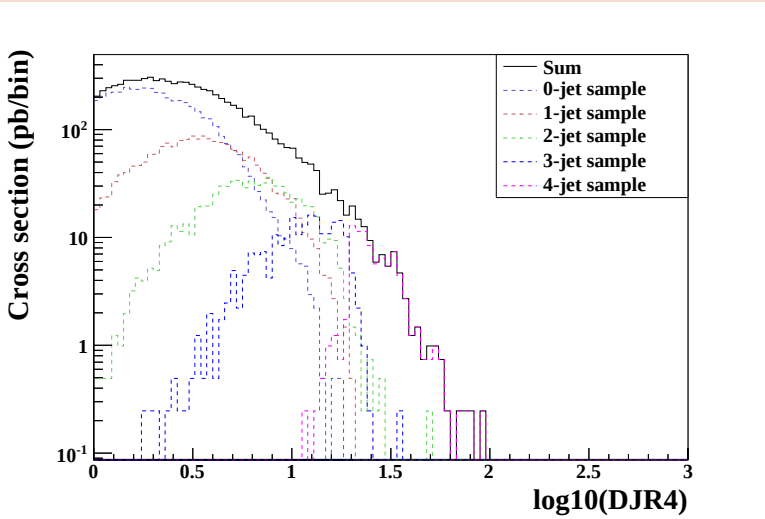
  - ◇ We can choose $N_{\max}$ $\Rightarrow$ the number of desired histograms.

Checking the merging procedure (2).









Reconstructing and analyzing hadron-level files (1).

- **The (StdHep or HepMC) event files contain tons of hadrons.**
  ▶ **Jet clustering** is required.

This is a task for MadAnalysis 5.
The reco mode: `bin/ma5 -R`

- **MadAnalysis 5 is interfaced to FastJet.**
  ▶ **Large selection** of jet algorithms

```
ma5>set main.clustering.algorithm =
antikt        cdfjetclu    genkt         kt             siscone
cambridge     cdfmidpoint  gridjet       none
```

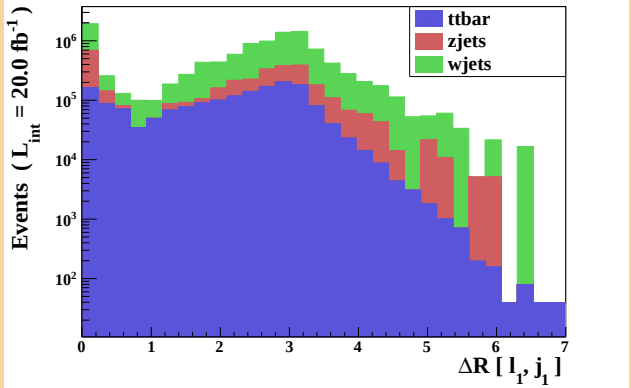
- **Adopting a jet algorithm ⇒ new options (the algorithm parameters).**

```
set main.clustering.algorithm = antikt
set main.clustering.ptmin = 5
set main.clustering.radius = 1
```

Reconstructing and analyzing hadron-level files (2).

- **A small example**.

```
set main.clustering.algorithm=antikt
set main.clustering.ptmin = 5
set main.clustering.radius = 1
import ttbar_2l.hep.gz as ttbar
import ttbar_1l.hep.gz as ttbar
import wjets.hep.gz as wjets
import zjets.hep.gz as zjets
set ttbar.xsection=139.6
set wjets.xsection=35678
set zjets.xsection=10319
set main.lumi = 20
set main.normalize = lumi
select (l) PT > 20
reject (j) PT < 50
reject THT < 200
plot DELTAR(l[1],j[1]) 30 0 7 [logY]
submit mydir
generate_html mydir_html
open mydir_html
```



Outline.

Summary.

- MadAnalysis **5 is a new framework for collider phenomenology**.
  - **Unique** $\Rightarrow$ partonic, hadronic or reconstructed events.
  - **User-friendly** $\Rightarrow$ Python command line interface.
  - **Flexible** $\Rightarrow$ a C++ kernel.
- **A special mode for expert users also exists**.
  - **Developer-friendly** $\Rightarrow$ C++ and Root skills required.
  - **No limitations**.
  - **See the manual**.

Try the code (and love it).

http://madanalysis.irmp.ucl.ac.be
ma5team@iphc.cnrs.fr