MadAnalysis 5

A tool devoted to phenomenological investigations

Eric CONTE, Benjamin FUKS, Manuel KRAFT, Guillaume SERRET

CMS France @ IPHC Strasbourg
23 – 25 January 2013

Outlines

Overview

Writing an analysis step-by-step in the normal mode

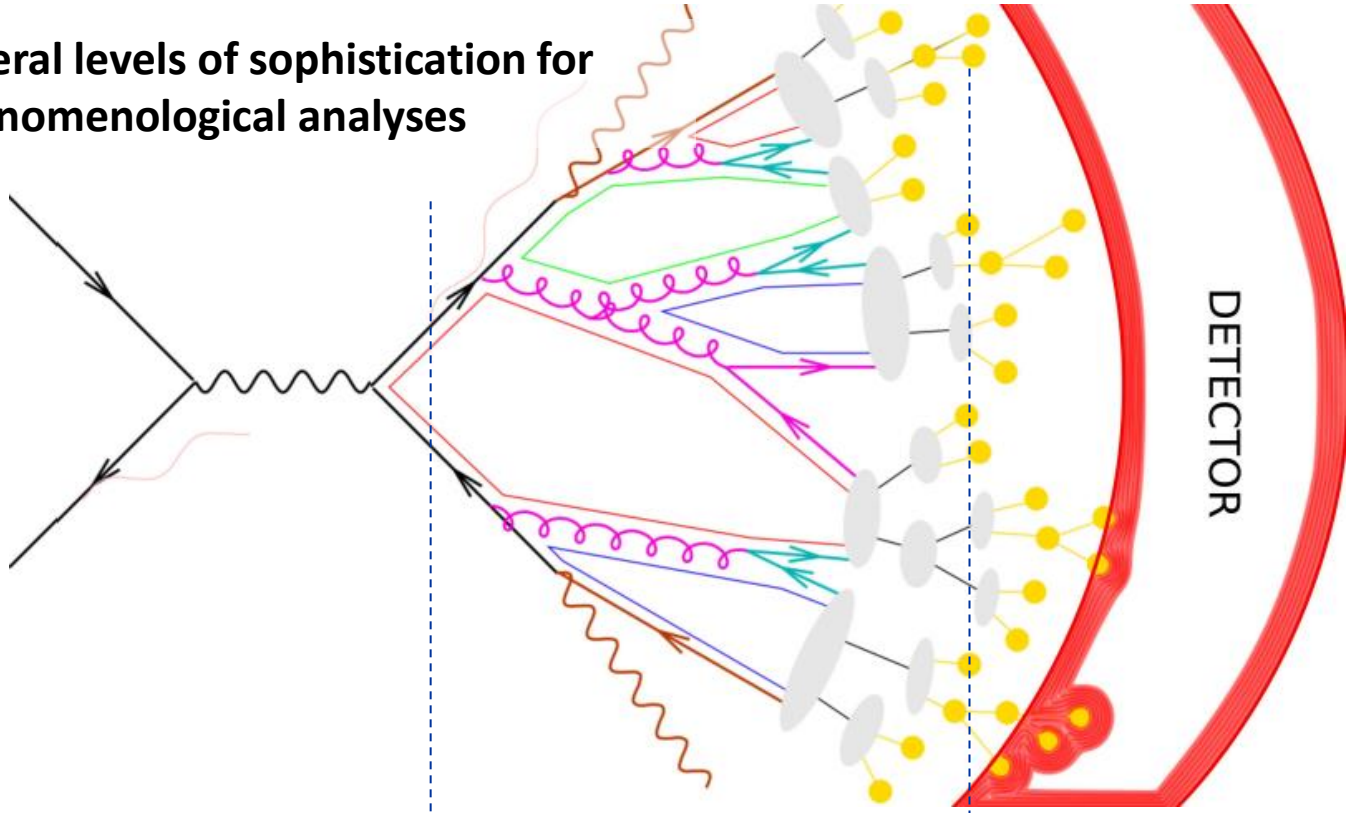
Interface to the FastJet package

Spoilers: what's new in the next releases ?

Summary

Overview

Several levels of sophistication for phenomenological analyses



Parton level | **Hadron level** | **Reconstructed objects level**

LHE files | **STDHEP/HEPMC files** | **LHCO files**

A unique framework : MadAnalysis 5

Overview

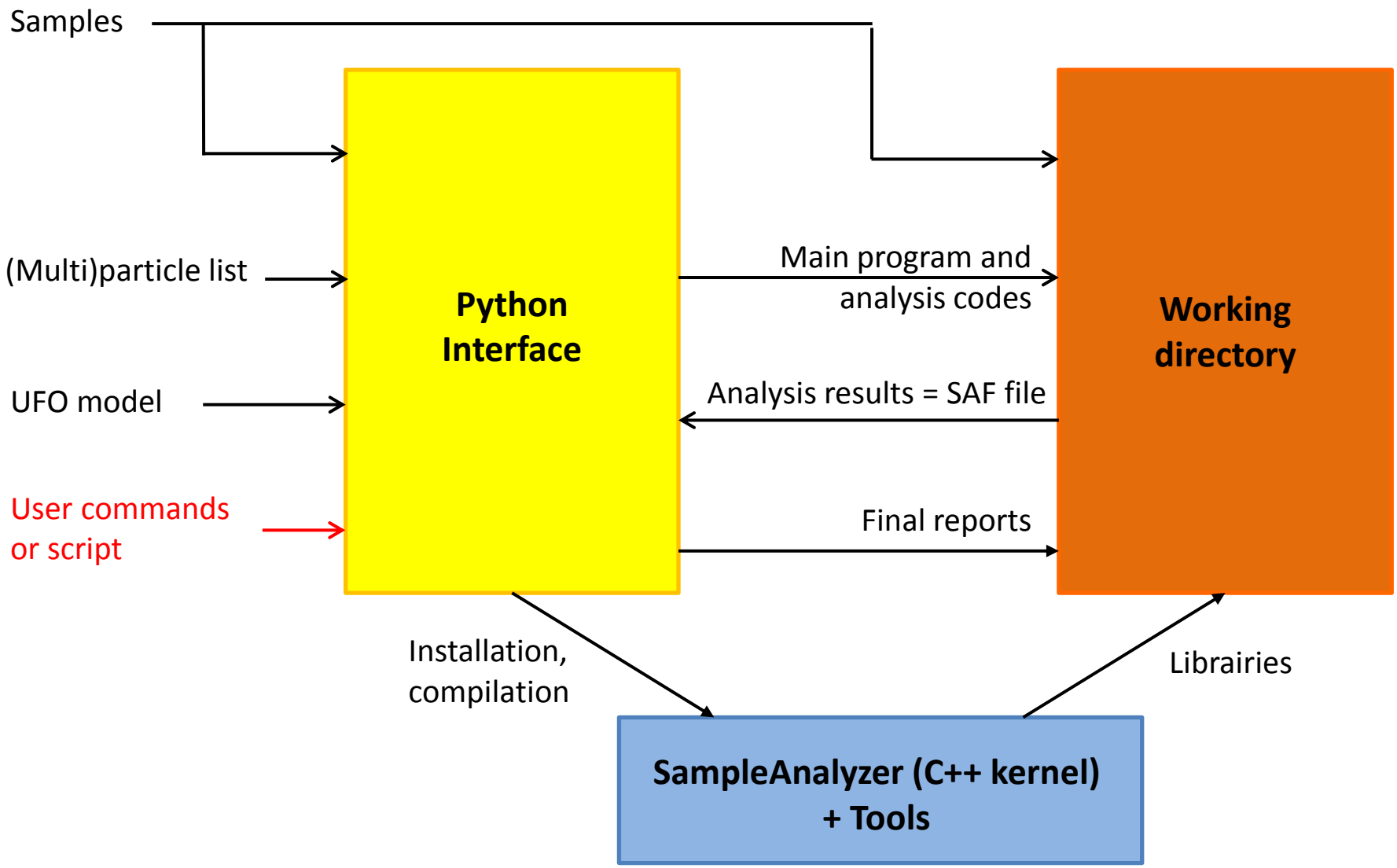
Scope:

- Reading of signal and background event files
- Definition of various selection cuts on the input samples.
- Production of histograms for different distributions.
- Results of the analysis summed up by a S/B-like ratio table.

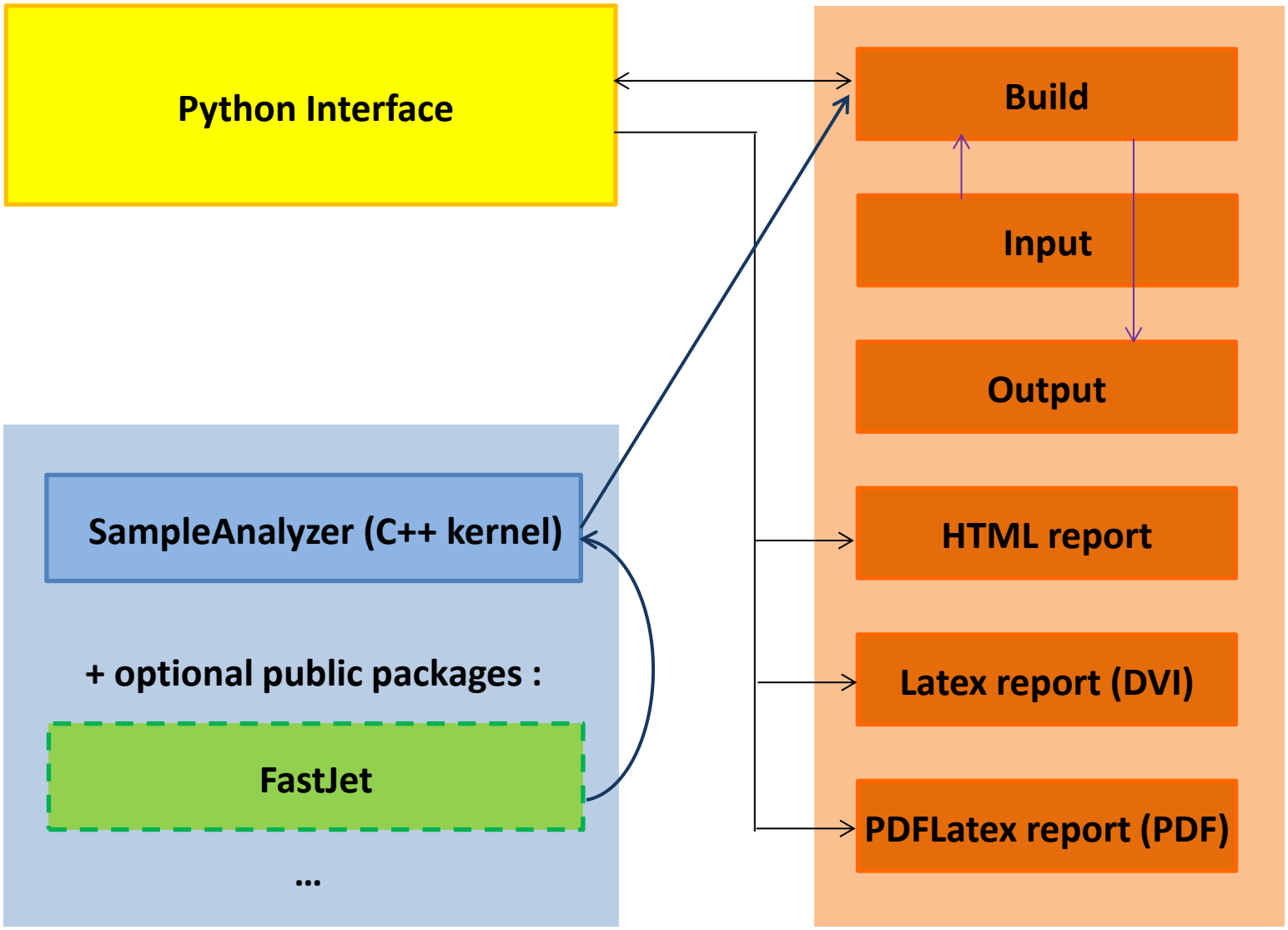
Relevant features of MadAnalysis 5 design:

- Study at any sophistication level (parton, hadron, reconstructed)
- Supporting any event file format (STDHEP, HEPMC, LHE, …)
- **User-friendly** → professional analyses in a simple way
- **Flexible**: no limit on the analysis complexity
- **Easy** to maintain and to validate

Overview


Samples
(Multi)particle list
UFO model
User commands
or script
Python
Interface
Main program and
analysis codes
Analysis results = SAF file
Final reports
Working
directory
Installation,
compilation
Librairies
SampleAnalyzer (C++ kernel)
+ Tools

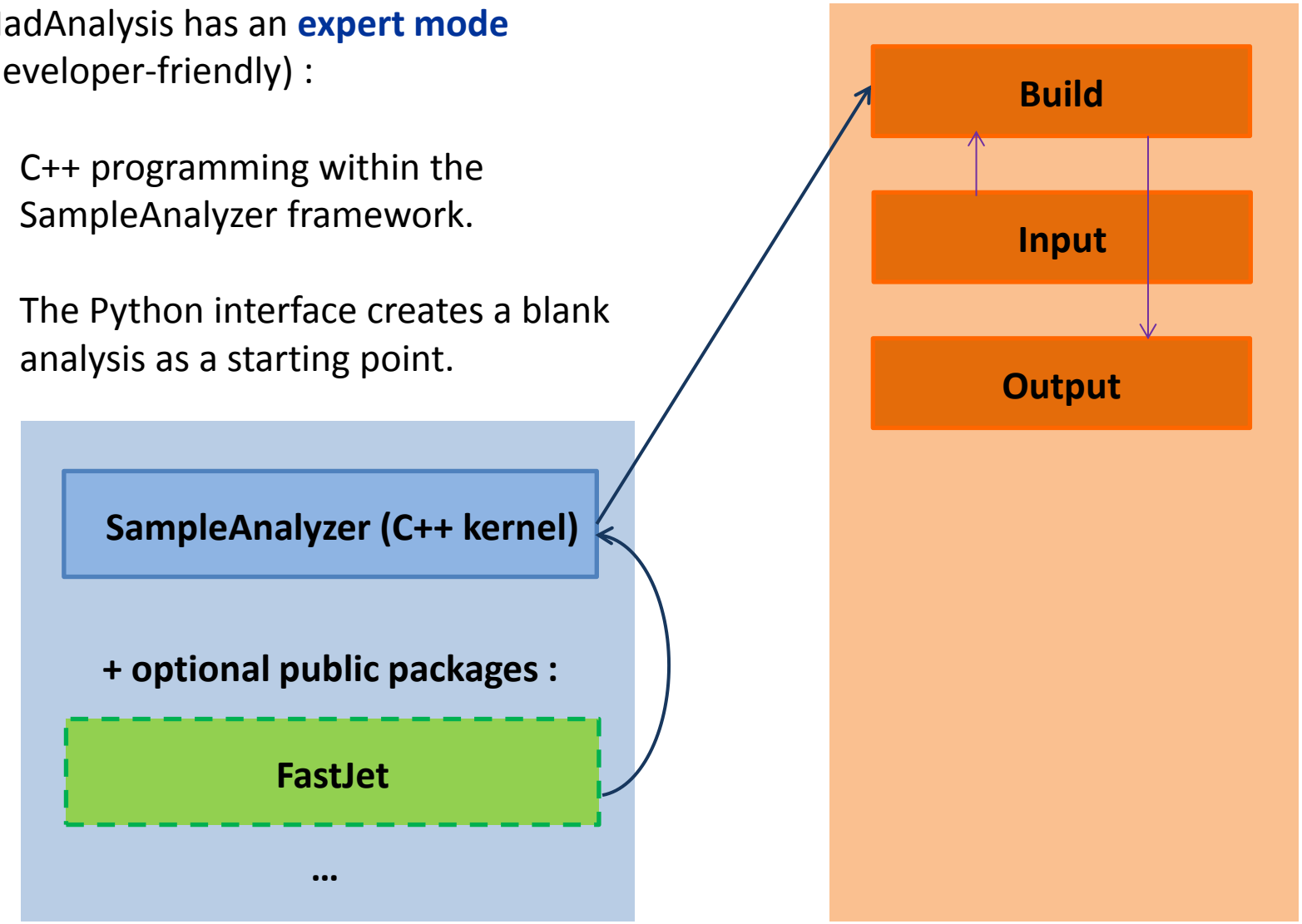
Overview


Python Interface
Build
Input
Output
SampleAnalyzer (C++ kernel)
HTML report
+ optional public packages :
Latex report (DVI)
FastJet
PDFLatex report (PDF)
...
Working directory

Overview

MadAnalysis has an **expert mode** (developer-friendly) :

- C++ programming within the SampleAnalyzer framework.
- The Python interface creates a blank analysis as a starting point.



Installation step

- **Requirements**

Mandatory	Optional
Python 2.6 or a more recent version (but not the 3.X series) **GNU GCC** compiler **ROOT** 5.27 or a more recent version	**zlib** **Latex** / PDFLatex **FastJet** 3.0 or a more recent version

- **Downloading MadAnalysis 5:**
  - From the official website http://madanalysis.irmp.ucl.ac.be (tarball to untar)
  - From MadGraph 5 interface **(available soon)**

Writing an analysis step-by-step

First start of MadAnalysis 5:

- **Execution**

Parton level	Hadron level	Reconstructed objects level
bin/ma5 or bin/ma5 –P	bin/ma5 –H	bin/ma5 -R

- **Initial sequence:**
  - **Step 1: Testing all dependencies**
  - **Step 2: Compiling the static library of SampleAnalyzer**
  - **Step 3: Locating MadGraph and importing the list of particles and multiparticles**

Writing an analysis step-by-step

Defining new particles and multiparticles

- Particles are defined by **labels**, which could point to one or several **PDG-id**.
- SM and MSSM labels are automatically loaded at the starting of MadAnalysis.
- The user can define his own labels :

```
ma5> define mu = mu+ mu-
```

- All labels defined in a UFO model can be loaded too.

Importing datasets

- For MadAnalysis, a **dataset** is a collection of samples which will be merged.
- All sample files are stored in a dataset.

```
ma5> import tt*.lhe
```

```
ma5> import tt*.lhe as ttbar
ma5> import Wj*.lhe as Wjets
```

- Possibility to tag datasets as **signal** or **background**.

Writing an analysis step-by-step

Defining a selection : plots and/or cuts

- **Histograms**
  - Observable can be related to the event or the properties of a particle
  - Plethora of observables: N, E, ET, M, MT, P, PT, PX, PY, PZ, THETA, ETA, …, ALPHAT
  - Combining particles

```
ma5> plot MET
ma5> plot PT(mu)
```

```
ma5> plot M(mu+ mu-)
```

- **Cuts : selecting / rejecting events**

```
ma5> reject MHT < 50
ma5> select N(mu) >= 2
```

- **Cuts : selecting / rejecting a particle or a combination**

```
ma5> select (mu) PT > 50
ma5> select 80 < M (mu+ mu-) < 100
```

Writing an analysis step-by-step

Defining a selection : plots and/or cuts

Several options or syntaxes allow to extend the potential of MadAnalysis.
Some examples:

- By defaut, a combination is interpreted as the vector sum of momenta.
  This interpretation can be changed by adding a prefix to the observable label.
  For instance : vPT, sPT, dsPT, dvPT, rPT

- List of observables specific to the reconstructed object level : ISOL, HE_EE, NTRACKS, …

- Selecting a particle according to its rank in energy (or to other observables)

```
ma5> plot(mu+[1])
```

- Selecting a particle according to its history (requirements on mother, grand-mother …)

```
ma5> plot(mu+ < w+ < t~)
```

Writing an analysis step-by-step

Launching the analysis:

This can be done by the command **`submit`**

- Creating a working directory (with a default name if no name is specified)
- Compiling the C++ job
- Launching the analysis over the different samples contained in the datasets

```
* SampleAnalyzer 2.0 for MadAnalysis 5 - Welcome.
* Option choices: selecting analysis = 'MadAnalysis5job'.
* Extracting the following sample files:
* 1/4  ~/samples/ttbar_sl_1.lhe.gz
*    => sample produced by MadGraph.
*    => Number of processed events: 1000.
. . .
```

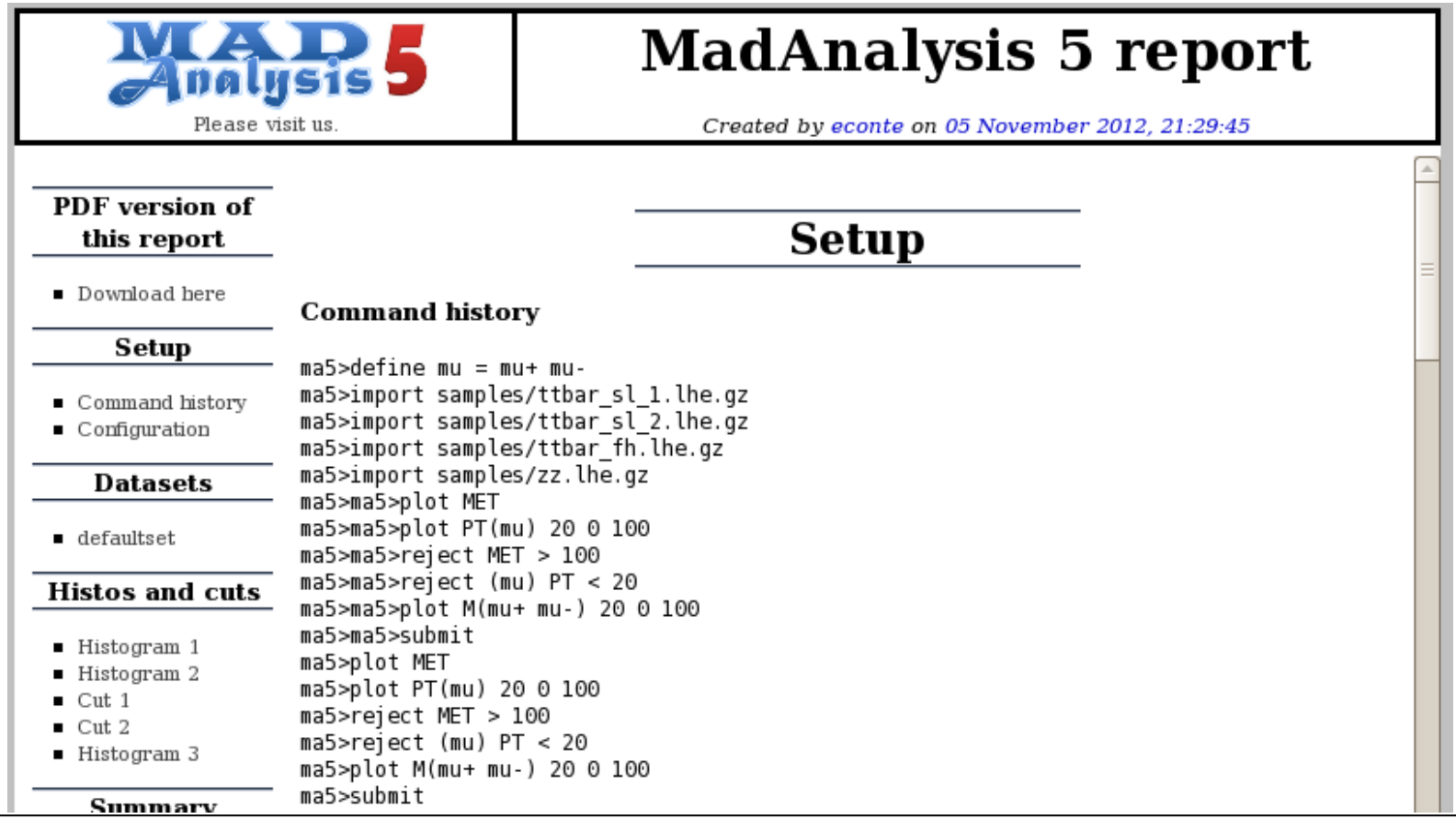
If you modify, after the submission, the analysis or the layout of the plots , the results can be updated in an optimized way by the command **`resubmit`**.

Writing an analysis step-by-step

Opening a generated report:

The command **open** displays the HTML report of the last job created.

Other reports can be opened by: **open workindir/PDF** or **open workingdir/DVI**

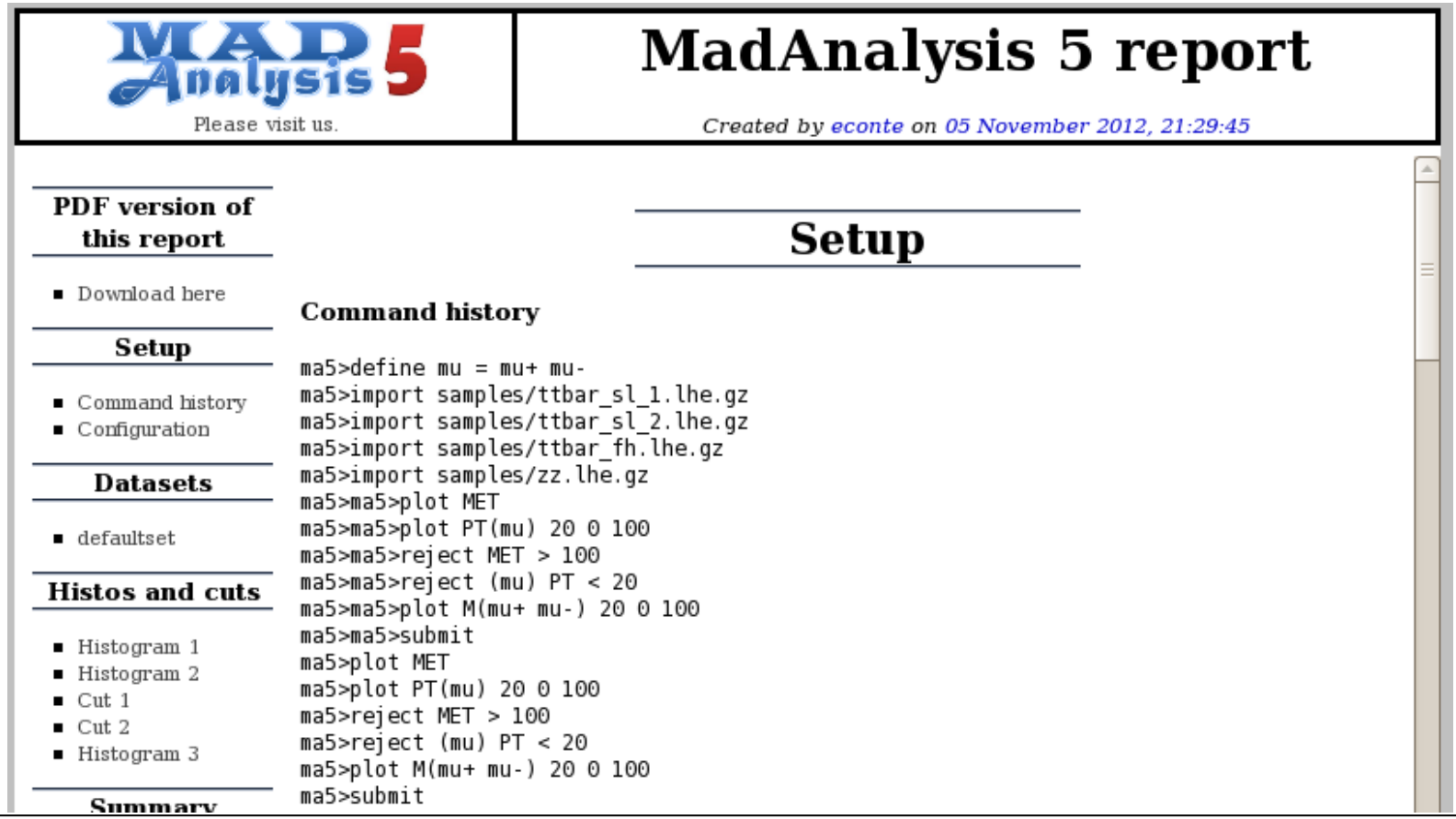


Writing an analysis step-by-step

Opening a generated report:

The command **open** displays the HTML report of the last job created.

Other reports can be opened by: **open workindir/PDF** or **open workingdir/DVI**



Writing an analysis step-by-step

Opening a generated report: Details on sample information

Path to the event file	Nr. of events	Cross section (pb)	Negative wgts (%)
mg5_merged.hep.gz	5116	162.0	0.0

Path to the event file	Nr. of events	Cross section (pb)	Negative wgts (%)
amcatnlo.hw.hep.gz	9993	313	8.4

- By default, MadAnalysis5 takes into account the **event-weights** contained in the samples. If no event-weights are available, there are set to one.

- MadAnalysis5 handles **negative event-weights** produced by NLO generators and propagates them properly into the uncertainties calculation (whatever the ROOT version).

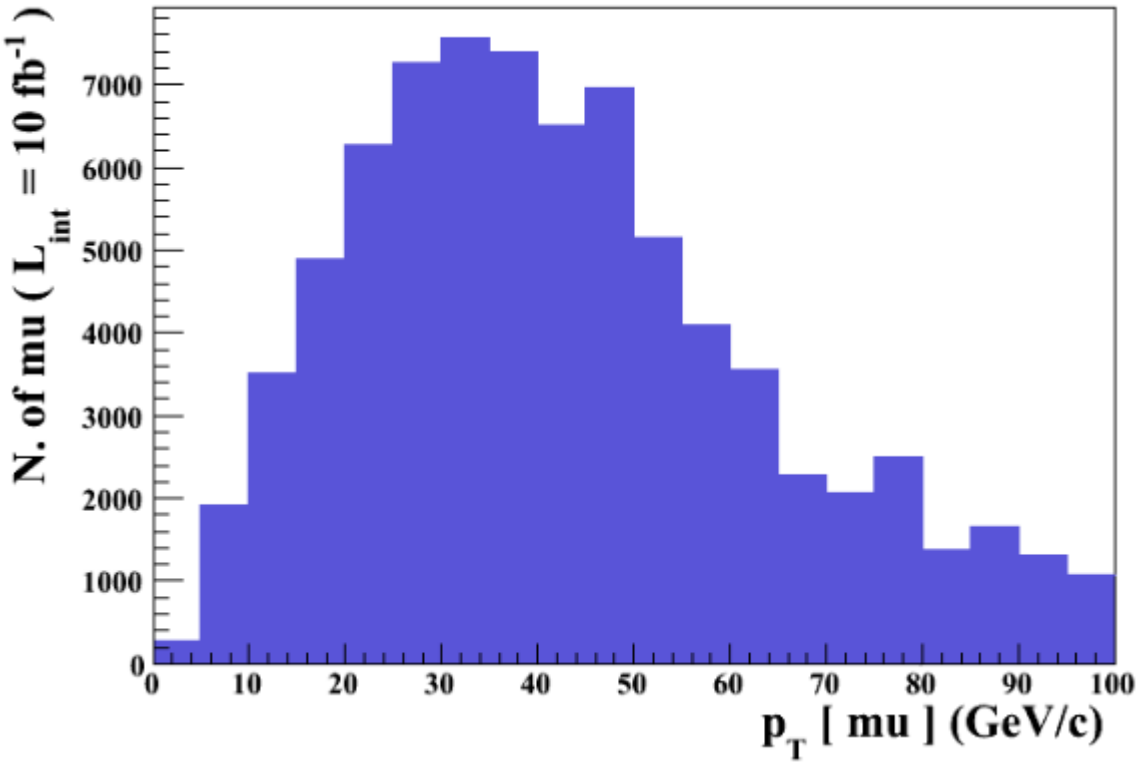
Writing an analysis step-by-step

Opening a generated report:

Details on histogramming

Dataset	Integral	Entries / events	Mean	RMS	Underflow	Overflow
defaultset	82747	0.752	42.8177	21.36	0.0	1.296

Statistics table



- the cross section of the sample is automatically extracted from the sample
- Integrated luminosity is by default 10 pb^{-1}. This value can be set by the user:

```
ma5> set main.lumi =
```

Interface to the FastJet package

Jet clustering algorithms:

This can be done with MadAnalysis 5 in reco mode
bin/ma5 -R

- Need to install **FastJet** and interface it to MadAnalysis

→ Just one command line from the Python interface !

```
ma5>  install fastjet
```

- Large selection of jet algorithms

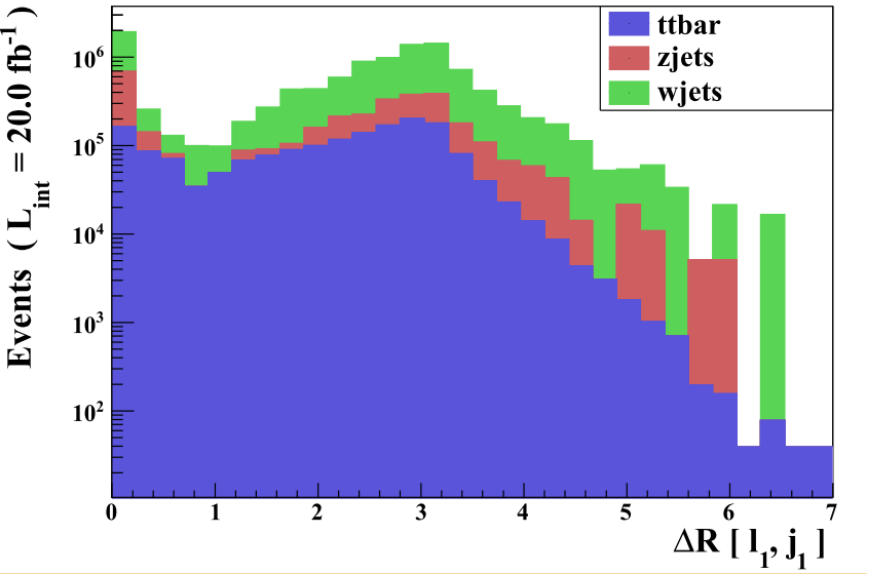
```
ma5>  set  main.clustering.algorithm  =

antikt            cdfjetclu         genkt    kt       siscone
cambridge         cdfmidpoint       gridjet  none
```

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

```
ma5>  set  main.clustering.algorithm  =  antikt
ma5>  set  main.clustering.ptmin   =  5
ma5>  set  main.clustering.radius  =  1
```

Interface to the FastJet package

Jet clustering algorithms:




```
ma5> set main.clustering.algorithm = antikt
ma5> set main.clustering.ptmin   =  5
ma5> set main.clustering.radius  =  1

ma5> import ttbar*.hep.gz as ttbar
ma5> import wjets.hep.gz  as wjets
ma5> import zjets.hep.gz  as zjets

ma5> set ttbar.xsection = 139.6
ma5> set wjets.xsection = 35678
ma5> set zjets.xsection = 10319
ma5> set main.lumi = 20

ma5> select (l) PT > 20
ma5> reject (j) PT < 50
ma5> reject THT < 200
ma5> plot DELTAR(l[1],j[1]) 30 0 7 [logY]

ma5> submit
ma5> open
```

Interface to the FastJet package

Simplified LHE output:

- After the jet clustering procedure, events could be saved in a simplified LHE.
- Enabling this option can be done with one command line in your analysis:

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

- The conventions used in the simplified LHE format are :

Light jets	PDG id = 21
c jets	PDG id = 4
b jets	PDG id = 5
MET	PDG id = 12

Electrons	PDG id = 11
Muons	PDG id = 13
Hadronic taus	PDG id = 15

This functionnality replaces the old module called "HEP2LHE", provided in the MadGraph package

Thanks to Adam Alloul for validation

Interface to the FastJet package

Jet-merging validation plots:

Matrix elements

hard partons

2 complementary approaches

Need to merge them
but avoid double counting.

Parton showers

soft partons

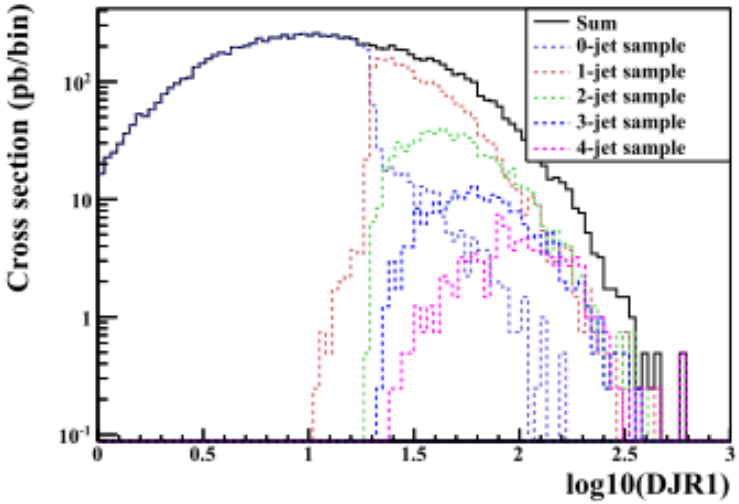
- **Merging matrix-elements with 0, 1, 2, 3, …. extra jets**
  - Study of the smoothness of the differential jet rate (DJR) distributions.
    - The scale for which an event goes from a N → N+1 jet configuration.
    - Extremely sensible to the merging procedure.
  - This validates the choices for the merging parameters.
- **Running MadAnalysis 5 in hadron-level mode: bin/ma5 -H**

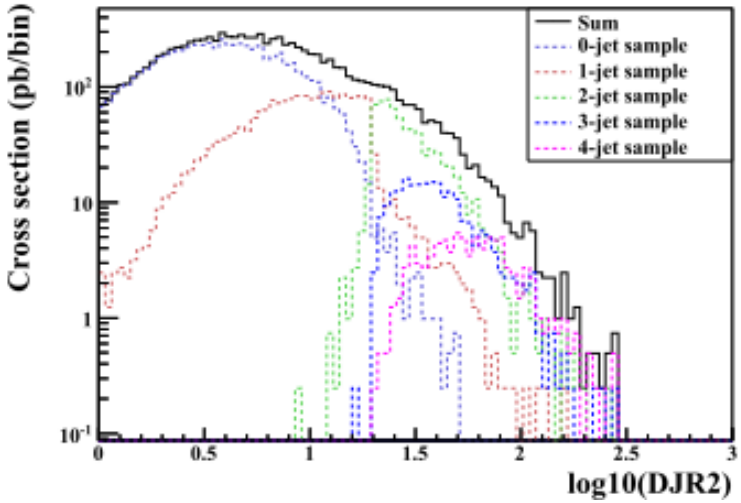
```
ma5> set  main.merging.check  =  true
ma5> set  main.merging.njets  =  4
```

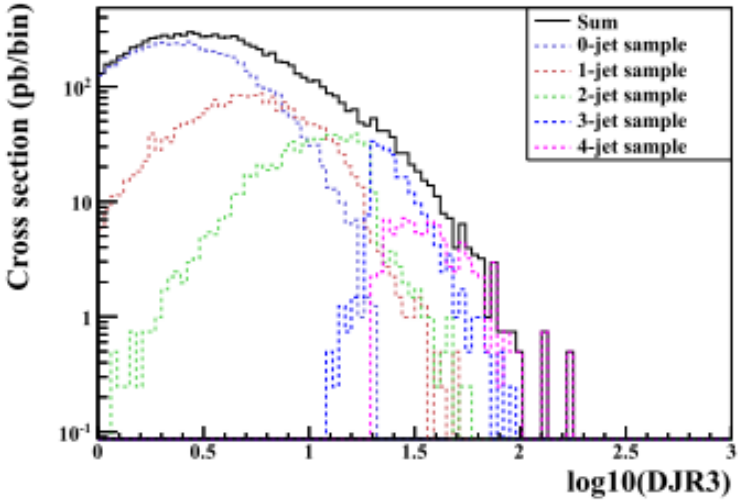
Interface to the FastJet package

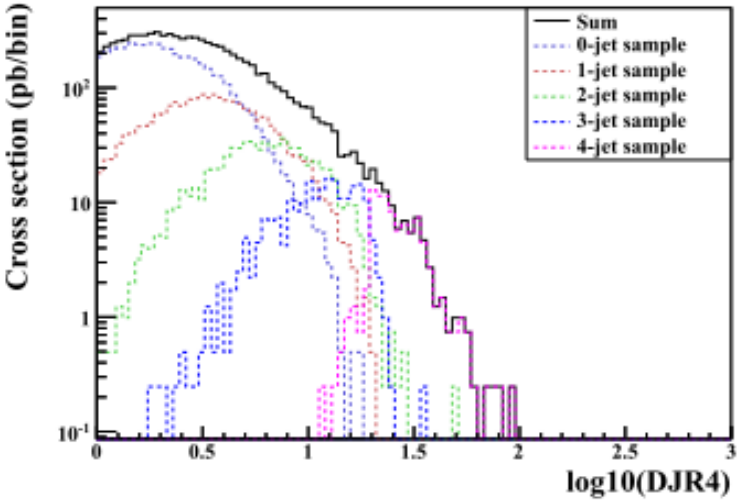
Jet-merging validation plots:

Example of Z production with 0, 1, 2, 3, 4 extra jets





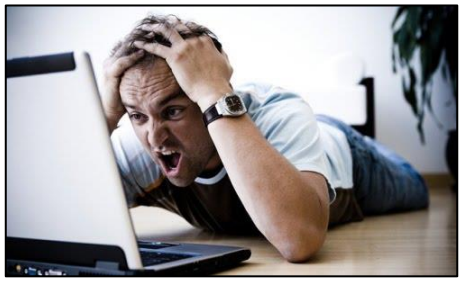




What's new in the next releases ?

In the next release (soon):

- **New internal structure → the expert mode is enriched and more flexible:**
  - Plot and cut data are saved as a simplified XML file (**SAF format**).
  - The main loop of the analysis can be modified by the user. Possibility to add several analyses, filters or jet clustering algorithms in a same job.
- **Removing ROOT libraries from python interface: adopting a new graphical style for plots.**
- **Fixing minor bugs.**
- **Completing the user manual by tutorials.**



What's new in the next releases ?

In the longer-term: the main milestones

- **Ready to perform phenomenological investigations @ NLO:**
  - Using information associated to events produced by **aMC@NLO**, in particular reweighting parameters.
  - Interfacing showering/hadronization package (**Pythia 6/8, Herwig6/++**) and configuring it automatically wrt the generation conditions.
- **Interfacing the existing detector fast-simulation packages (Delphes) and developing new ones related to top physics (in priority: *boosted top reconstruction*).**
- **Pipelining the data-flow "showering-fastsim-analysis": the full processing chain is applied event by event (avoiding intermediate sample files).**
- **Optimizing the processing time by paralleling jobs.**

Summary

- **MadAnalysis 5 = a unique framework for different levels of analysis:**
  Parton level , Hadron level and Reconstructed objects level

- Designed to be **fast** and **user-friendly.**

- Two ways of using the program
  - **Normal mode:** python interface with intuitive commands.
  - **Expert mode:** requiring programming skills (C++, ROOT).

- **Interfaced to FastJet**, MadAnalysis 5 can:
  - launch a specified **jet clustering** sequence to hadronic events.
  - display the **plots dedicated to check the jet merging procedure**.

- Exepected for 2013: interfaces to **fast-simulation** and **showering** packages.

MADAnalysis 5

ma5team@iphc.cnrs.fr
http ://madanalysis.irmp.ucl.ac.be
Comput. Phys. Commun. 184 (2013) 222