Status and new developments in MadAnalysis 5

Eric CONTE, Benjamin FUKS, Guillaume SERRET

GDR Terascale @ LPNHE Paris
5 – 7 November 2012

Outlines

Overview

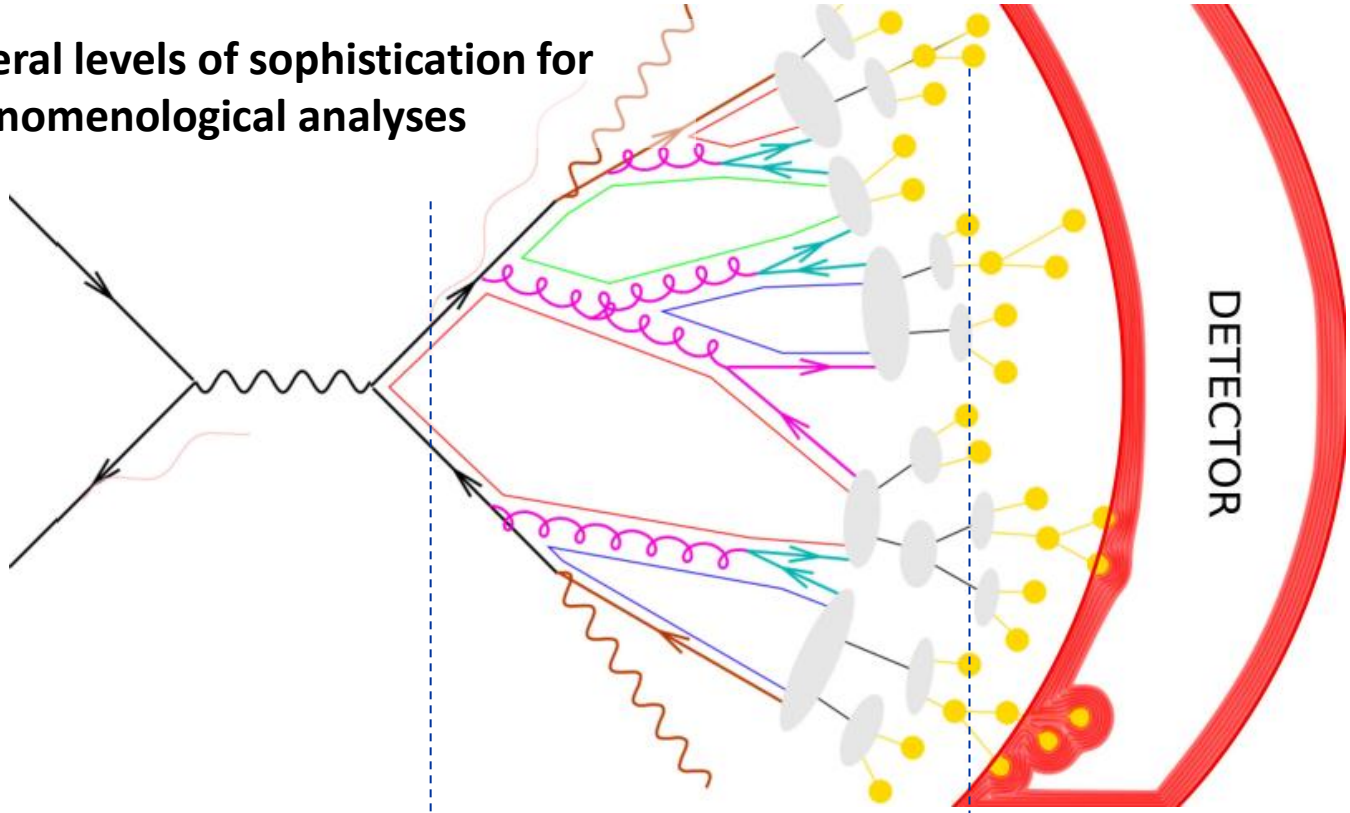
Writing an analysis step-by-step in the normal mode

New functionalities since MadAnalysis 5.1.2

Summary and perspectives

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

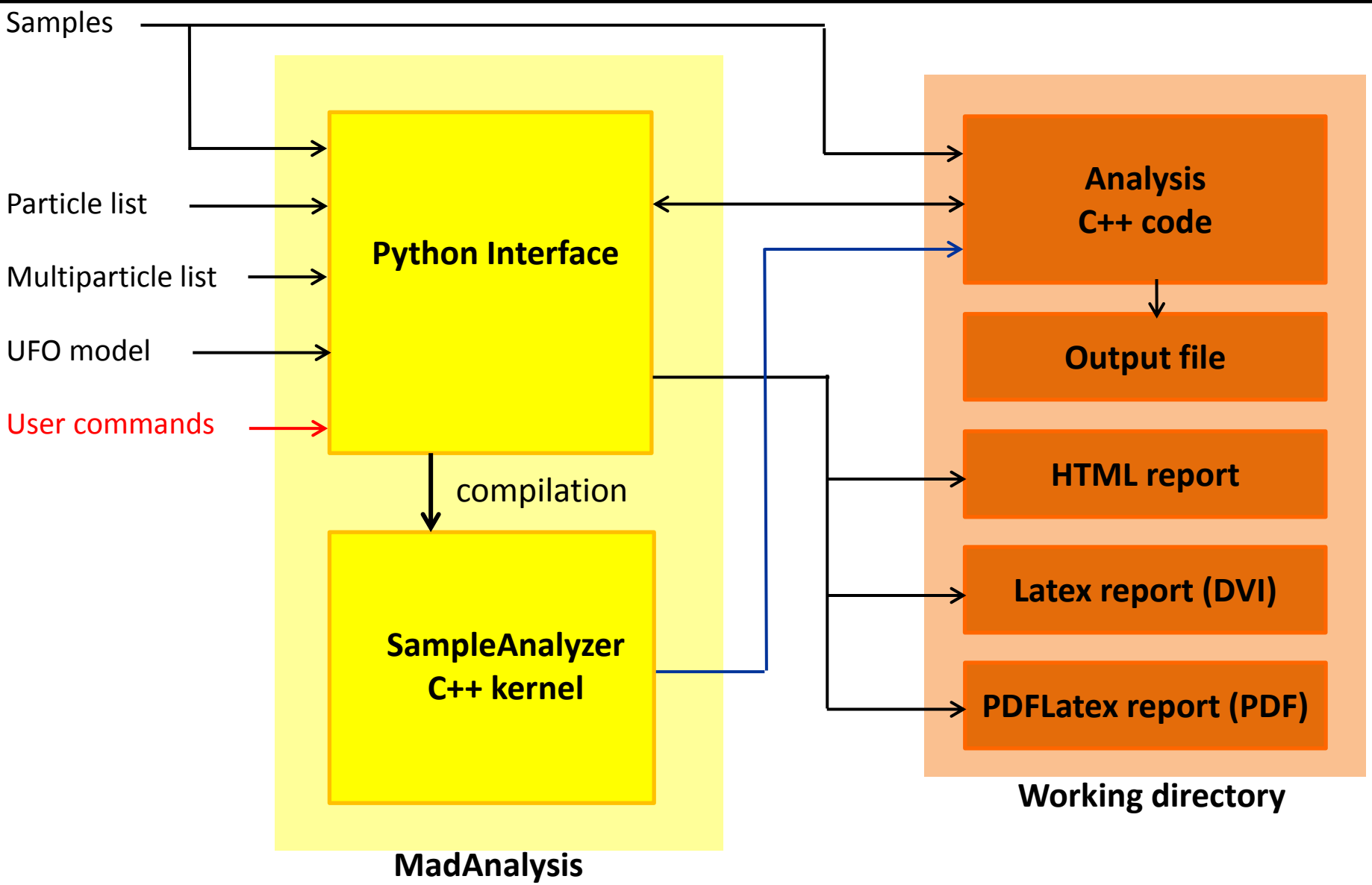
Scope:

- 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.

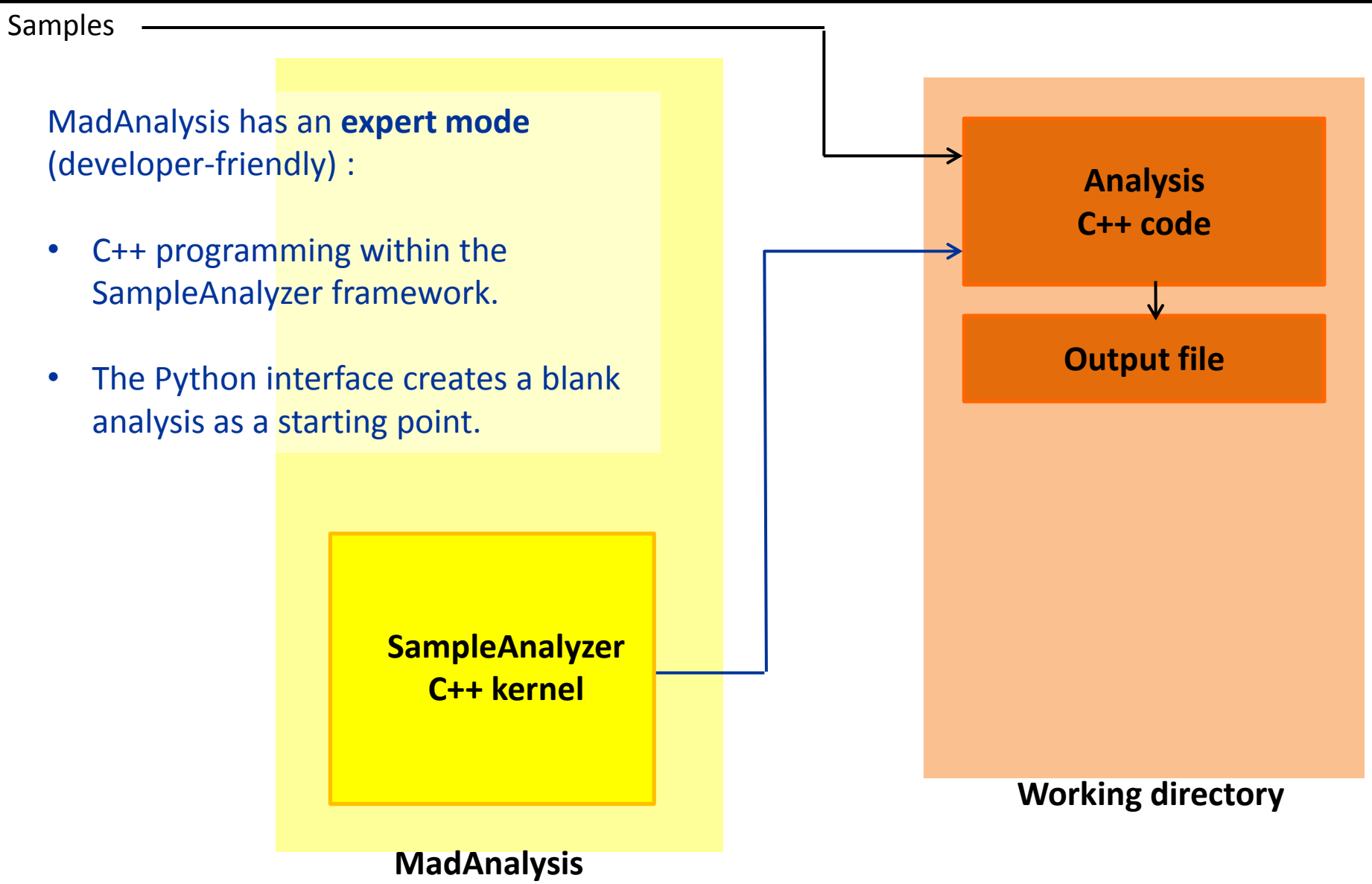
Relevant features of MadAnalysis design:

- 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
Particle list
Multiparticle list
UFO model
User commands
Python Interface
compilation
SampleAnalyzer
C++ kernel
MadAnalysis
Analysis
C++ code
Output file
HTML report
Latex report (DVI)
PDFLatex report (PDF)
Working directory

Overview


Samples
MadAnalysis has an **expert mode** (developer-friendly) :
• C++ programming within the SampleAnalyzer framework.
• The Python interface creates a blank analysis as a starting point.
Analysis
C++ code
Output file
SampleAnalyzer
C++ kernel
Working directory
MadAnalysis

Overview

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, …
  - 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 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.
. . .
```

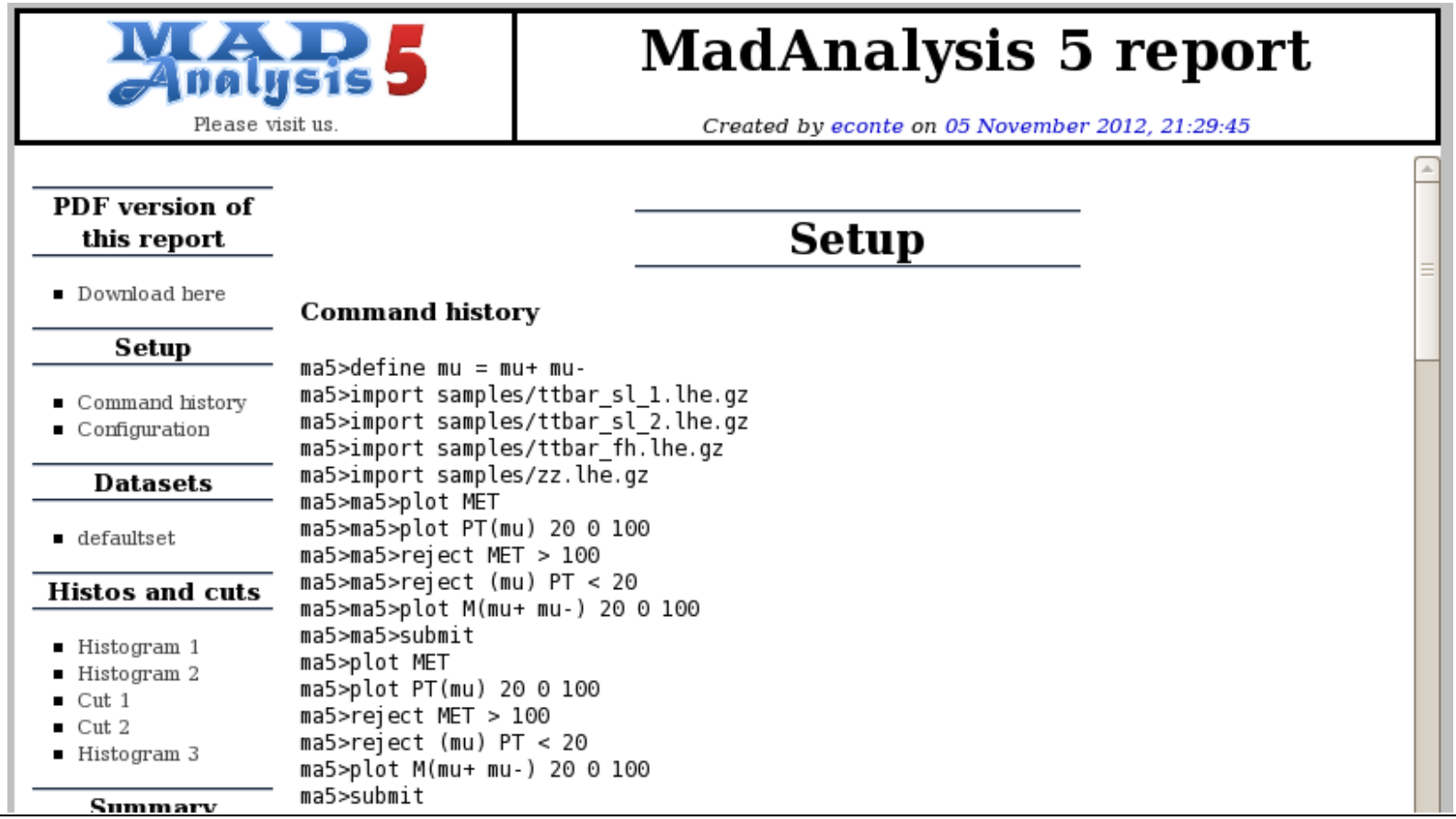
If you modify, after the submission, the analysis or the layout of the plots , the results can be updated in an optimized way by the command **`resubmit`**.

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**



MadAnalysis 5 report

Please visit us.

Created by econte on 05 November 2012, 21:29:45

PDF version of this report

- Download here

Setup

- Command history
- Configuration

Datasets

- defaultset

Histos and cuts

- Histogram 1
- Histogram 2
- Cut 1
- Cut 2
- Histogram 3

Summary

Setup

Command history

```
ma5>define mu = mu+ mu-
ma5>import samples/ttbar_sl_1.lhe.gz
ma5>import samples/ttbar_sl_2.lhe.gz
ma5>import samples/ttbar_fh.lhe.gz
ma5>import samples/zz.lhe.gz
ma5>ma5>plot MET
ma5>ma5>plot PT(mu) 20 0 100
ma5>ma5>reject MET > 100
ma5>ma5>reject (mu) PT < 20
ma5>ma5>plot M(mu+ mu-) 20 0 100
ma5>ma5>submit
ma5>plot MET
ma5>plot PT(mu) 20 0 100
ma5>reject MET > 100
ma5>reject (mu) PT < 20
ma5>plot M(mu+ mu-) 20 0 100
ma5>submit
```

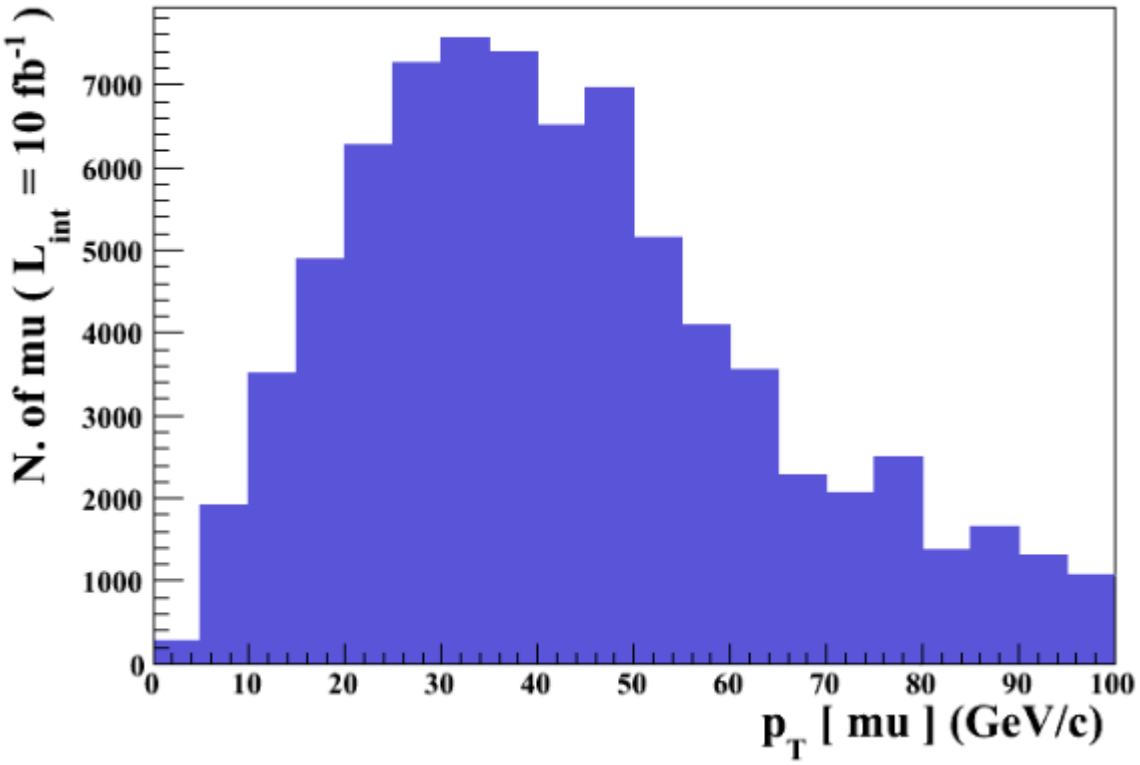
Writing an analysis step-by-step

Opening a generated report:

one extract

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 =
```

New functionality: jet clustering

(STDHEP or HEPMC) event files contain plethora of hadrons.
Jet clustering is required !

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
```

New functionality: jet clustering

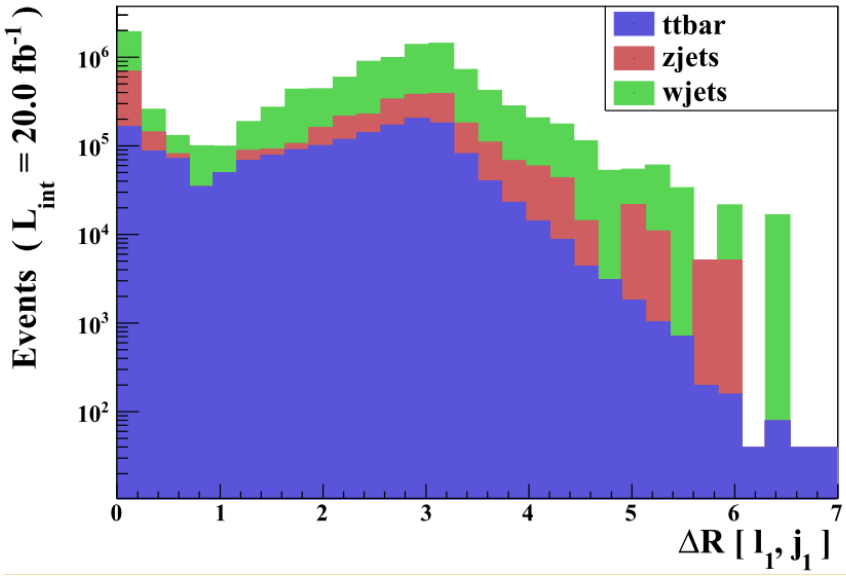
```
ma5> set main.clustering.algorithm = antikt
ma5> set main.clustering.ptmin   =  5
ma5> set main.clustering.radius  =  1
ma5> set main.outputfile = «output.lhe.gz»

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
```



New functionality: merging plots

Matrix elements

hard partons

2 complementary approaches

Need to merge them
but avoid double counting.

Parton showers

soft partons

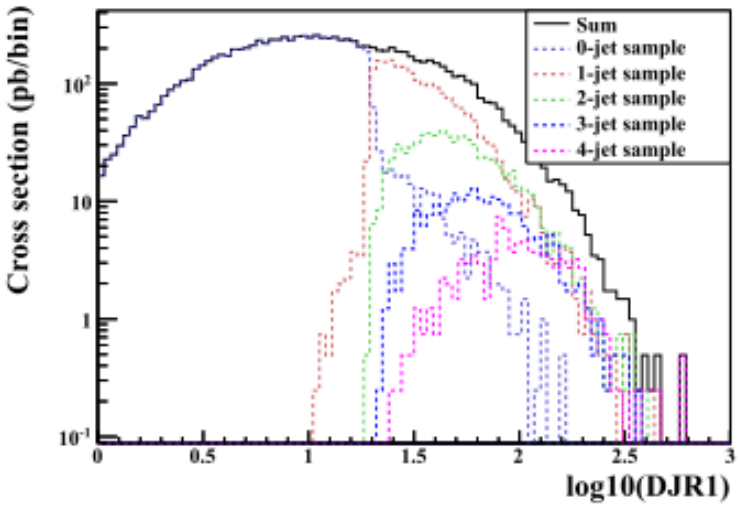
- **Merging matrix-elements with 0, 1, 2, 3, …. extra jets**
  - Study of the smoothness of the differential jet rate (DJR) distributions.
    - The scale for which an event goes from a N → N+1 jet configuration.
    - Extremely sensible to the merging procedure.
  - This validates the choices for the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

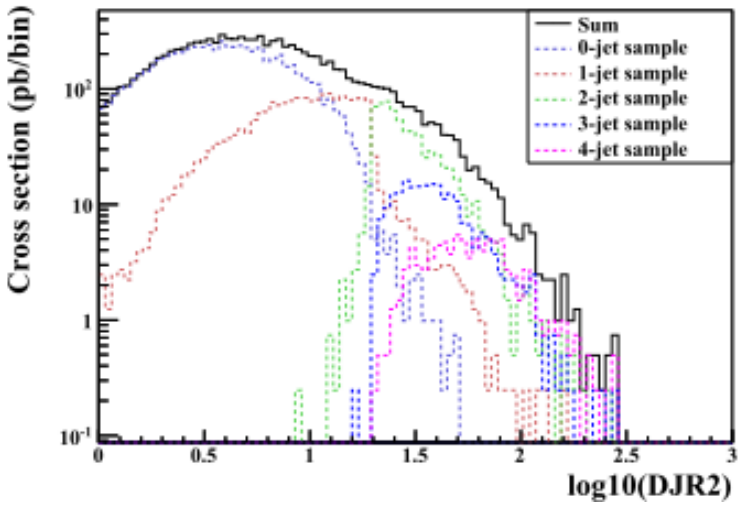
```
ma5> set  main.matching.check  =  true
ma5> set  main.matching.njets  =  4
```

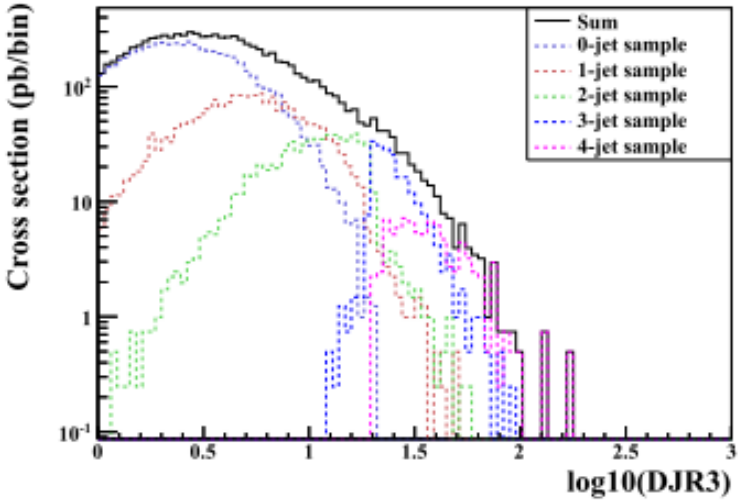
We can choose N_{max} ⇒ the number of desired histograms.

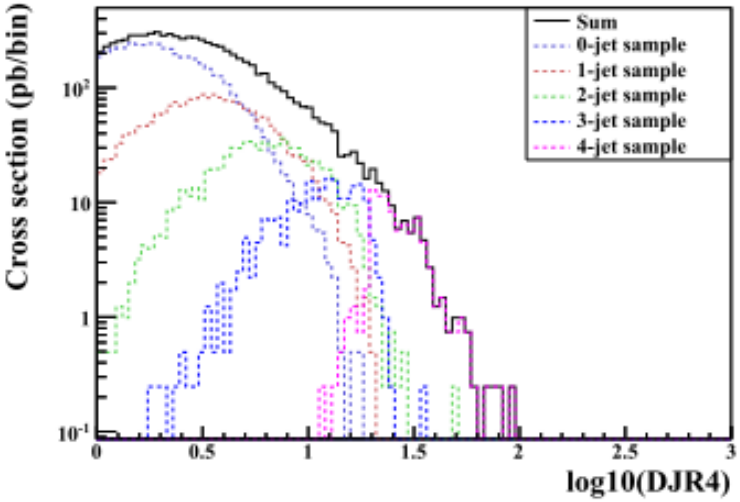
New functionality: merging plots

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









New functionality: weighted events

- Up to now, MadAnalysis could only treat samples with **unweighted events**:
  - Each event has the same weight.
  - Histograms could be rescaled by using “the” cross-section value.

- In the **next release (this week)**, MadAnalysis addresses **weighted events**.
  Weights are taken into account in :
  - Plots
  - Cuts and their statistical uncertainties.

- In particular, **negative weights** are considered too (NLO generators)

ROOT interprets correctly the negative weights from the **version 5.30**.

→ MadAnalysis 5 overcomes this restriction:
Correct results are obtained for **ALL supported versions** of ROOT.

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 ME/PS jet merging**.

- Ready for reading aMC@NLO samples (in particular **negative event-weights**)

MADAnalysis 5

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

Perspectives

Completing the current features of MadAnalysis 5:

- Reading on-fly aMC@NLO events
  → Need to "pipelinize" the data flow aMC@NLO - MadAnalysis

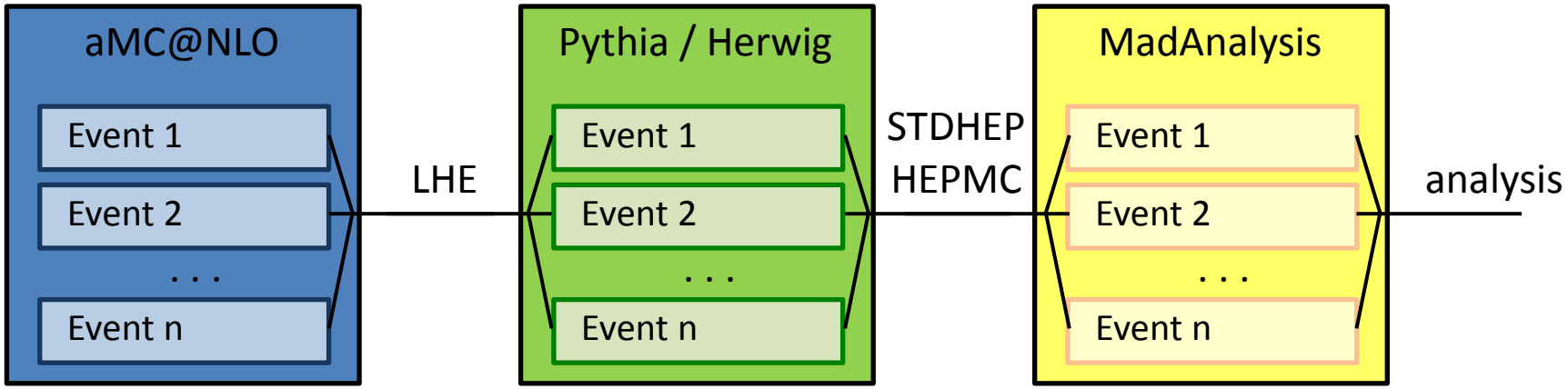
- Submitting several analyses at the same time
  → Reading the event only one time and applying successively several analyses

- No more hadronic event format : STDHEP or HEPMC (files too heavy for the storage)
  → interfacing MadAnalysis to Pythia 6, Pythia 8 and Herwig
  → pipeline scheme is required too

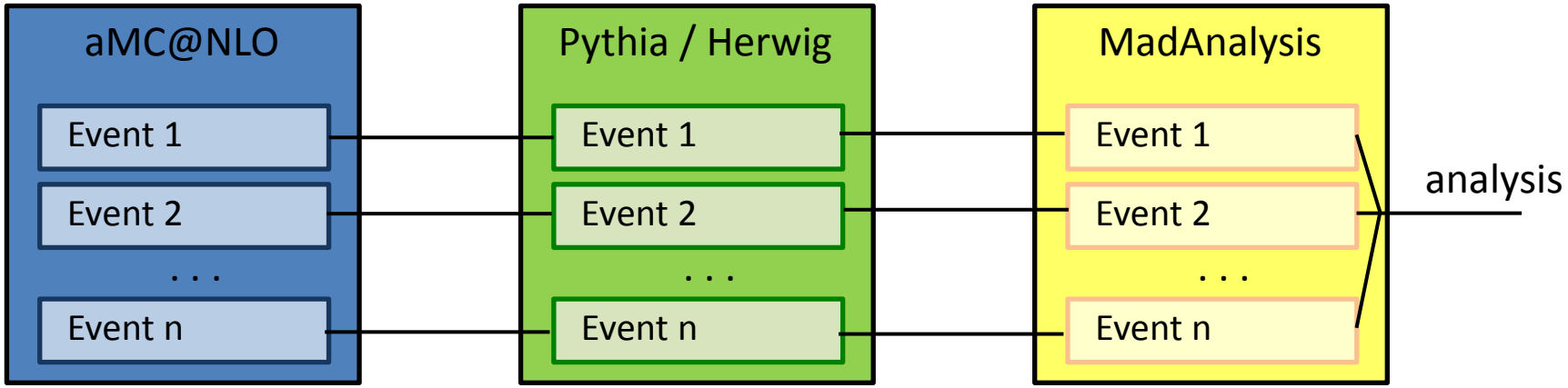
- Extending the generation of output file to other formats than LHE:
  → user-friendly mode: LHCO (text format)
  → expert mode: ROOT (binary format)

Perspectives

Current data-flow:



Pipelined data-flow:



Perspectives

(Very-)Fast simulation project :

The MadAnalysis team was actively involved in the CERN meeting about

Fast simulators for the LHC (June 2012)

Designing with the Delphes team a unified framework for fast simulation

- *Splitting the simulation into several modules*
- *Description on how these modules communicate*
- *Data format: how data are stored in the memory ?*
- *Storage format: how data are stored on disk ?*

Back-Up

Installation and launching MA5

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

- **Step 1: Testing all dependencies**

```
Checking ROOT libraries ...
Loading ROOT libraries ...
Checking g++ libraries ...
Checking zlib libraries ...
Checking fastjet libraries ...
 ** WARNING: FastJet configuration program is not found.
 ** WARNING: To enable this functionality, please type 'install fastjet'.
Checking if pdflatex is installed...
Checking if latex is installed...
```

Installation and launching MA5

First start of MadAnalysis 5:

- **Step 1: Testing all dependencies**
- **Step 2: Compiling the static library of 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
Answer:
Number of cores used for the compilation = 16
   Linking the MadAnalysis library...
   Checking the MadAnalysis library presence...
************************************************************
```

Installation and launching MA5

First start of MadAnalysis 5:

- **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**

```
MadGraph 5 NOT found => default particle names from the file:
  /home/econte/madanalysis/input/particles_name_default.txt
84 particles have been successfully exported.

MadGraph 5 NOT found => default multiparticle definitions from the file:
  /home/econte/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.
```