MadAnalysis 5
A new framework for collider phenomenology

Benjamin Fuks (IPHC Strasbourg / Université de Strasbourg)

In collaboration with E. Conte & G. Serret.

Fast simulators for the LHC @ CERN
June 11-12, 2012

Outline

1. Introduction.
2. Overview of MadAnalysis 5.
3. Examples for normal and expert users.
4. Towards a fast detector simulation.
5. Conclusions.

Comprehensive particle physics phenomenology.

1. Implementation of new physics models in tools such as FEYNRULES.
   [Christensen, Duhr (CPC '09); Christensen, Degrande, Duhr, BenjF (in prep)]
   - or in any **Monte Carlo model generator program**.
2. Event generation with MADGRAPH **5**. [Alwall, Herquet, Mattelaer, Stelzer (JHEP '11)]
   - or with any **Monte Carlo event generator**.

Parton-level phenomenology.

3. Parton showering and hadronization with tools such as PYTHIA or HERWIG.
   [Sjostrand, Mrenna, Skands (JHEP '06; CPC '08); Corcella *et al.* (JHEP '01); Bahr *et al.* (EPJC '08)]
   - or with any **parton showering tool**.

Hadron-level phenomenology.

4. Fast detector simulation with tools such as DELPHES or PGS.
   [Ovyn, Rouby, Lemaitre ('09); Conway ('06)]
   - or with any **fast detector simulation program**.

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 **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.

[Conte, BenjF, Serret (arXiv:1206.1599 [hep-ph])]

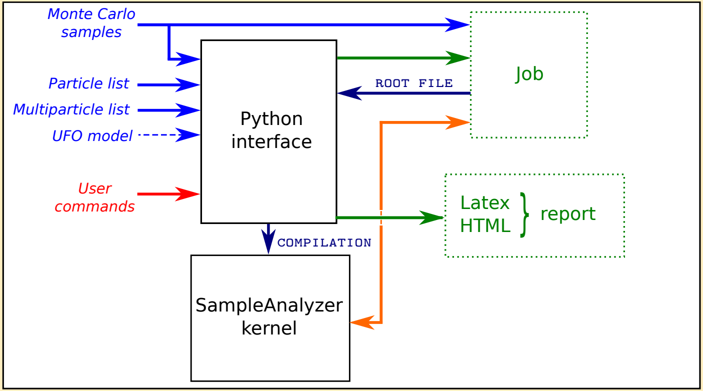
Outline.

1. Introduction.
2. Overview of MadAnalysis 5.
3. Examples for normal and expert users.
4. Towards a fast detector simulation.
5. Conclusions.

From MadAnalysis 4 to MadAnalysis 5.

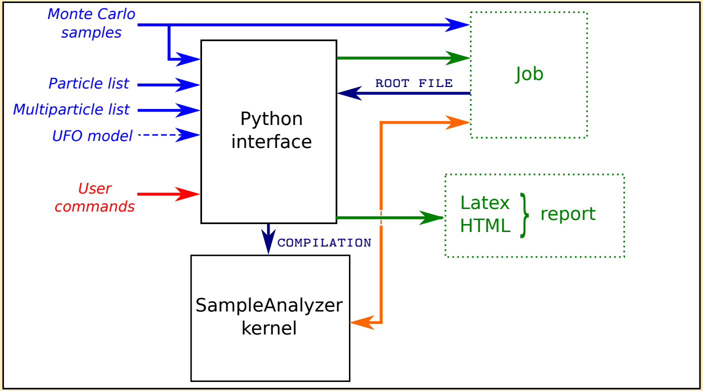
- **Object-oriented programming language**.
  - MadAnalysis 4: Fortran.
  - MadAnalysis 5: C++ core; Python interface; uses Root.
- **Flexibility**.
  - MadAnalysis 4: **No**.
  - MadAnalysis 5: **Yes**.
- **User-friendly**.
  - MadAnalysis 4: **A complicated plot card**.
  - MadAnalysis 5: **Intuitive** Python **commands**.
- **Limitations**.
  - MadAnalysis 4: **What is implemented**.
  - MadAnalysis 5: **The user's imagination**.
- MadAnalysis **5 is going beyond the** Mad**-series of programs.**
  - Can be used as a **standalone package**.

The MadAnalysis 5 scheme.



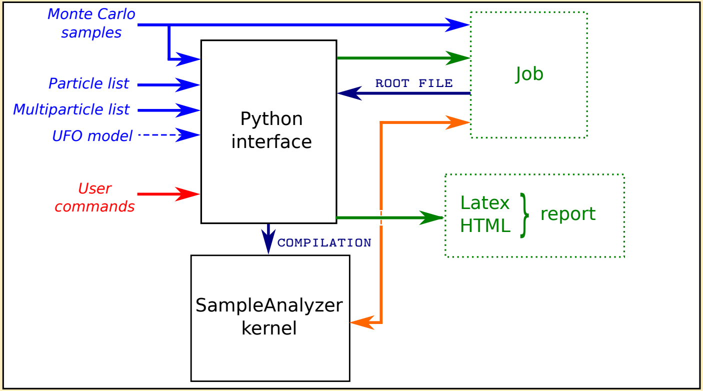
- **Two modules.**
  - * A Python **command line interface**: **interactive commands**.
  - * A C++/Root **module,** SampleAnalyzer: **performs the analysis**.

The MadAnalysis 5 scheme.



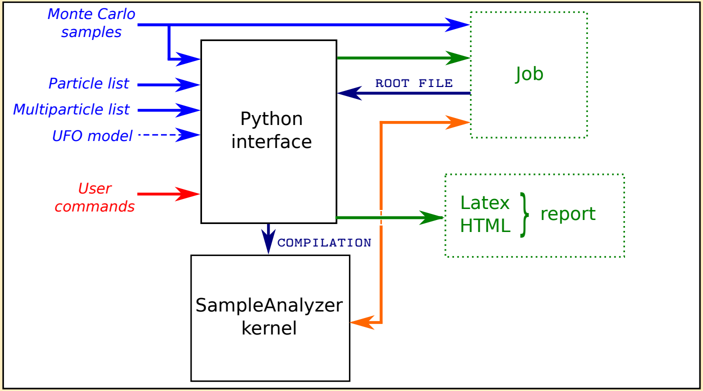
- **Normal mode of running** (user-friendly).
  - * **Commands typed in the Python interface**.
  - * Analysis performed **behind the scene** (black box).
  - * **Human readable output**: HTML, LaTeX.

The MadAnalysis 5 scheme.



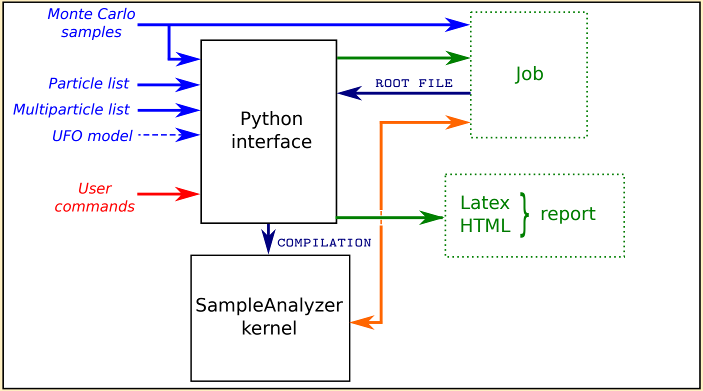
- **Expert mode** (developer-friendly).
  - * C++ programming within the SampleAnalyzer framework.
  - * **C++ and Root skills required**.
  - * The Python interface creates a **blank analysis as a starting point**.

The MadAnalysis 5 scheme.



- **Inputs**.
  - * Monte Carlo samples ⇔ **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**.

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** [Degrande, Duhr, BenjF, Grellscheid, Mattelaer, Reiter (CPC '12)].
  * 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`.
  - * Creation of an **histogram**.
  - * **Global observables** ⇔ 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**.
  - * **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.

1 Introduction.

2 Overview of MadAnalysis 5.

3 Examples for normal and expert users.

4 Towards a fast detector simulation.

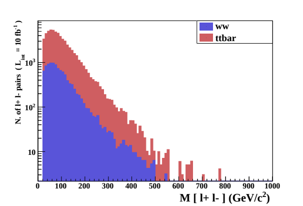
5 Conclusions.

Particle properties.

- **Kinematical distributions related to particle species**.
  - * **Intuitive** commands.
  - * **Available observables**:
    BETA, DELTAR, E, ET, ETA, GAMMA, M, MT, P, PHI, PT, PX, PY, PZ, R, THETA, Y.
  - * Scalar and vectorial sums/differences are implemented.

► $t\bar{t}$ (dileptonic mode).
► WW (dileptonic mode).
► LHC @ 8 TeV; 10 fb^{-1}.
► Parton-level.
► Dilepton invariant mass M(l+ l-).

```
import ttbar_ll.lhe.gz as ttbar
import ww_ll.lhe.gz as ww
plot M(l+ l-) [logY]
submit tempdir
generate_latex temp_tex
```

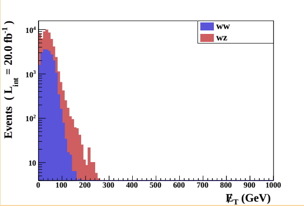


Global event observables.

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

- ▶ *WW* (dileptonic mode).
- ▶ *WZ* (decay to at least one lepton).
- ▶ LHC @ 8 TeV; 20 fb^{-1}.
- ▶ Parton-level.
- ▶ Missing energy distribution.

```
import ww_ll.lhe.gz as ww
import wz_l.lhe.gz as wz
plot MET [logY]
set main.lumi = 20
submit tempdir
generate_latex temp_tex
```

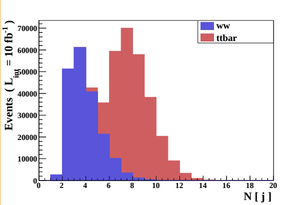


Multiplicities.

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

▶ $t\bar{t} + 0, 1, 2$ jets (hadronic mode).
▶ $WW + 0, 1, 2$ jets (semileptonic mode).
▶ LHC @ 8 TeV; 10 fb^{-1}.
▶ Hadron-level.
▶ Jet multiplicity.

```
import ttbar_hh.lhe.gz as ttbar
import ww_l.lhe.gz as ww
define j = j b b~
plot N(j)
submit tempdir
generate_latex temp_tex
```

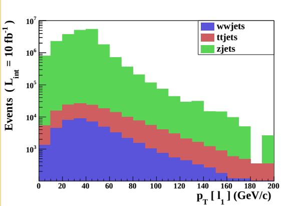


Leading lepton properties.

- **Particle ordering**.
  - * Can be access with the **squared brackets** [<i>] .
  - * Several possible **ordering variables**. E, ET, ETA, P, PT, PX, PY, PZ.

- ►$Z + 0, 1, 2, 3, 4$ jets (dileptonic mode).
- ►$WW + 0, 1, 2$ jets (dileptonic mode).
- ►$t\bar{t} + 0, 1, 2$ jets (dileptonic mode).
- ►LHC @ 8 TeV; 10 fb^{-1}.
- ►Hadron-level.
- ►**Energy ordering**.
- ►**Leading lepton p_T**.
- ►The **binning is specified**.

```
import z.lhe.gz as zjets
import ttbar.lhe.gz as ttjets
import ww.lhe.gz as wwjets
define l = l+ l-
plot PT(l[1]) 20 0 200 [logY]
set selection[1].rank = Eordering
submit tempdir
generate_latex temp_tex
```

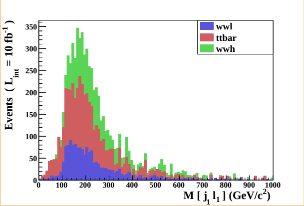


Cuts and signal over background ratios (1).

- **Cuts**.
  - * Through the commands `select` and `reject` followed by a condition.
  - * Samples can be tagged as `signal` or `background`.
  - * **Formula** for the signal over background ratio can be provided.

- **Background**: $WW + 0, 1, 2$ jets (dileptonic and semileptonic modes).
- **Signal**: $t\bar{t} + 0, 1, 2$ jets (dileptonic mode).
- **Lepton candidates:** $p_T > 50$ GeV;
- **Reject events if** $E_T(j) < 50$ GeV.
- **Signal over background ratio**: S/B.

```
import ww_ll.lhe.gz as wwl
import ww_hl.lhe.gz as wwh
import ttbar_ll.lhe.gz as ttbar
set wwl.type = background
set wwh.type = background
set main.SBratio = 'S/B'
define l = l+ l-
select (l) PT > 50
reject ET(j) < 50
plot M(l[1] j[1])
submit tempdir
generate_latex temp_tex
```



Cuts and signal over background ratios (2).

- MadAnalysis **5 output for signal over background comparison**.
  - * Formula for S-B comparison: S/B.
  - * Formula for uncertainty on S-B comparison: 1./(B**2)*sqrt(B**2*ES**2+S**2*EB**2).
- **Reminding the cuts**.

```
select (l) PT > 50
reject ET(j) < 50
```

- **Results**.

Cuts	S	B	S vs B
Initial (no cut)	133780	243725	0.549
cut 1	133780	243725	0.549
cut 2	19134 +/- 128	20738 +/- 127	0.9226 +/- 0.0084

Expert users: W-boson polarization (1).

- **Property to be investigated**.
  * **Polarization of the W issued from a top leptonic decay**.
  * Process: $t\bar{t}$ **in the semileptonic decay channel**.
  * Property investigated to an **angular distribution** $\mathrm{d}\sigma/\mathrm{d}\cos\theta^\star$.
- **The angle $\theta^\star$ is the angle between**:
  * The momentum of the W evaluated in the top rest frame.
  * The momentum of the lepton evaluated in the W rest frame.
- **Developer-friendly implementation**:
  * Only the **relevant part** of the analysis is presented here.
    - ► Event processing, particle identification, histogram creation ⇒ manual.
  * Employing the built-in `ToRestFrame` and `angle` methods.
  * The momentum of the **lepton** is evaluated in the W rest frame.

```
PHYSICS->ToRestFrame(lepton,w);
```

  * The momentum of the W is evaluated in the top rest frame.

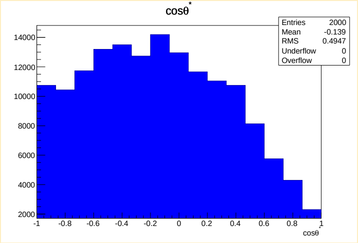
```
PHYSICS->ToRestFrame(w,top);
```

  * Filling the **histogram**:

```
histo->Fill( cos(lepton.angle(w)));
```

Expert users: W-boson polarization (2).

- **Property to be investigated**.
  - * **Polarization of the W issued from a top leptonic decay**.
  - * Process: **$t\bar{t}$ in the semileptonic decay channel**.
  - * Property investigated to an **angular distribution** $\mathrm{d}\sigma/\mathrm{d}\cos\theta^\star$.
- **The angle $\theta^\star$ is the angle between**:
  - * The momentum of the W evaluated in the top rest frame.
  - * The momentum of the lepton evaluated in the W rest frame.
- **Parton-level results**.



Outline.

1. Introduction.
2. Overview of MadAnalysis 5.
3. Examples for normal and expert users.
4. Towards a fast detector simulation.
5. Conclusions.

Why a fast detector simulation in MadAnalysis 5? (1)

- **Status of the existing tools?**
  - * There are several **existing codes which do a pretty good job**.
  - * Each has **limitations** (jet energy scale, complicated efficiencies, ...)
- **For our studies, we need to overcome these limitations**.
  - ▶ Fakes, triggers, τ-reconstruction, ...
- **Three options are offered to us**.
  1. **Hack** those codes.
  2. **Implement** a fastsim in MadAnalysis 5 with the required features.
  3. **The maximal mixing scenario**.
     - ▶ Take the **best from these codes**.
     - ▶ **Import them** in MadAnalysis 5 as shared libraries.
     - ▶ **Implementation** of the missing features in MadAnalysis 5.
- **Option #1 is not really an option**.
  - * **Need to dig into other's code**.
  - * Issues with **validation**.
  - * We might also **ask** to the codes' authors.
    - ▶ Do they want to do it?
    - ▶ What is the timescale then?

Why a fast detector simulation in MadAnalysis 5? (2)

- **Option #2**
  - * That means **reinventing the wheel** for many features.
    - ► A however **low necessary time budget**.
    - ► It might be better to **use this time for developing novelties**.
- **Option #3.**
  - * **No need to dig into other's code**.
  - * We can focus on **developing novel features**.
  - * Users can **benefit from the user-friendliness, flexibility and interactivity** of MadAnalysis 5.
  - * **Existing codes should provide dynamical libraries**.
    - ► We need to work all together (otherwise $\Rightarrow$ option #2).

Functionalities of a fast detector program.

- **Extension the functionalities of the existing codes.**
  - **Smearing**.
  - **Rescaling** (jet and lepton energy scales).
  - **Reco-gen matching.**
  - **Charge (mis)identification** for e, μ and τ.
  - **Particle tagging.**
  - τ **reconstruction**.
- **Not present. Physics studies required**.
  - **Fakes**.
  - **Calorimeter noise** (*e.g.*, random hot cells, ...).
  - **Pile-up.**
  - **Displaced vertices.**
  - **Trigger resolutions and structure.**
  - **Cosmic rays.**

Timeline.

- **Fall '12 release**.
    * **Interface with MadGraph 5**.
        - Automatic installation.
        - Automatic plots after event generation.
    * **Weighted events** (negative weights, aMC@NLO).
    * **Matching** plots.
    * **Interface with FastJet**.
        - New hep2lhe-like package.
    * **Tutorials**.
        - FeynRules-MadGraph 5 school in Natal.
- **Christmas '12: towards a fast detector simulation.**
  ►**We want to be ready for the LHC shutdown**.

Outline.

1. Introduction.
2. Overview of MadAnalysis 5.
3. Examples for normal and expert users.
4. Towards a fast detector simulation.
5. Conclusions.

Summary.

- **MadAnalysis 5 is a new framework for collider phenomenology**.
  * **Unique** ⇒ partonic, hadronic or reconstructed events.
  * **User-friendly** ⇒ Python command line interface.
  * **Flexible** ⇒ a C++ kernel.

- **A special mode for expert users exists**.
  * **Developer-friendly** ⇒ C++ and Root skills required.
  * **No limitations** ⇒ *e.g.*, the W polarization.

- **Major development plans**.
  * Interface with MadGraph **5**.
  * **Weighted events**.
  * **Matching** plots.
  * Interface with `FastJet`.
  * **Fast detector simulation**.

Try the code (and love it).

`http://madanalysis.irmp.ucl.ac.be`

`ma5team@iphc.cnrs.fr`