Collider phenomenology and LHC recasting studies with MadAnalysis 5

Benjamin Fuks

LPTHE / Sorbonne Université

2019 MadGraph School

November 20, 2019

Outline

New physics at the LHC

- The quest for physics beyond the Standard Model has started!
  - How to get hints of new physics?
    - Confront data to the Standard Model expectation in search channels
    - Observe unexplained deviations at a good confidence level
  - Ingredient 1: predictions for the Standard Model background
  - Ingredient 2: predictions for the new physics signals

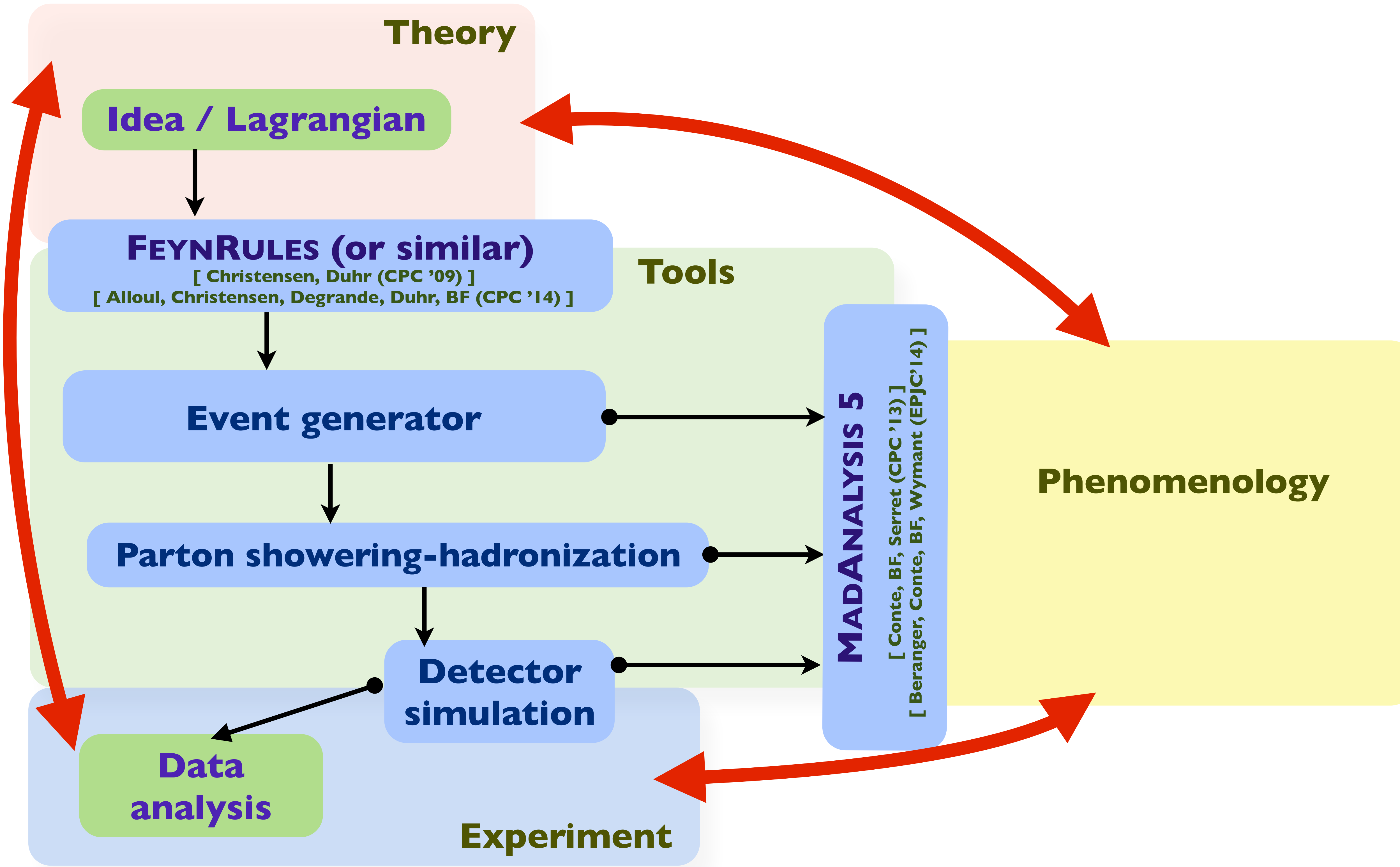
- More on the new physics nature
  - Fitting deviations by new physics signals
    - Designing new analyses to probe new ideas

      Predictions; signal and background analysis
    - Reinterpretation of data in possibly different theoretical frameworks

      Confronting models to data

A framework for LHC analyses: a modern way

[Christensen, de Aquino, Degrande, Duhr, BF, Herquet, Maltoni, Schumann (EPJC '11)]



MadAnalysis 5 and MG5_aMC are interfaced

[BF, Hirschi ('16)]

- Event analysis and recasting is now automated within MG5_aMC
  - All the lecture and tutorial exercises can be done within MG5

```
The following switches determine which programs are run:
/------------------------------------------------------------------------------\
|  1. Choose the shower/hadronization program:                    shower = OFF         |
|  2. Choose the detector simulation program:                   detector = OFF         |
|  3. Run an analysis package on the events generated:          analysis = MADANALYSIS_5 |
|  4. Decay particles with the MadSpin module:                   madspin = OFF         |
|  5. Add weights to events for different model hypothesis:     reweight = Not available (requires NumPy)|
\------------------------------------------------------------------------------/
```

- Parton-level, hadron-level et reco-level: everything is possible within MG5_aMC

```
# Uncomment the line below to skip this analysis altogether
# @MG5aMC skip_analysis

@MG5aMC stdout_lvl=INFO

@MG5aMC inputs = *.lhe
@MG5aMC analysis_name = analysis1

# Multiparticle definition
define vl = 12 14 16
define vl~ = -16 -14 -12
define invisible = ve ve~ vm vm~ vt vt~ vl vl~

# Histogram drawer (options: matplotlib or root)
set main.graphic_render = root

# Global event variables
plot THT   40 0 500 [logY]
plot MET   40 0 500 [logY]
plot SQRTS 40 0 500 [logY]
```

```
Do you want to edit a card (press enter to bypass editing)?
  1 / param       : param_card.dat
  2 / run         : run_card.dat
  3 / pythia8     : pythia8_card.dat
  4 / madanalysis5_parton : madanalysis5_parton_card.dat
  5 / madanalysis5_hadron : madanalysis5_hadron_card.dat
```

Outline

MadAnalysis 5 in a nutshell

[Conte, BF, Serret (CPC '13); Conte, Dumont, BF, Wymant (EPJC '14)]

✦ What is MadAnalysis 5?

- ✤ A framework for phenomenological analyses
- ✤ Any level of sophistication: partonic, hadronic, detector, reconstructed
- ✤ Several input format: StdHep, HepMC, Lhe, Lhco, Root (from Delphes)
- ✤ User-friendly, flexible and fast
- ✤ Interfaced to other HEP packages (detector simulation, jet clustering, …)

✦ Normal pythonic mode

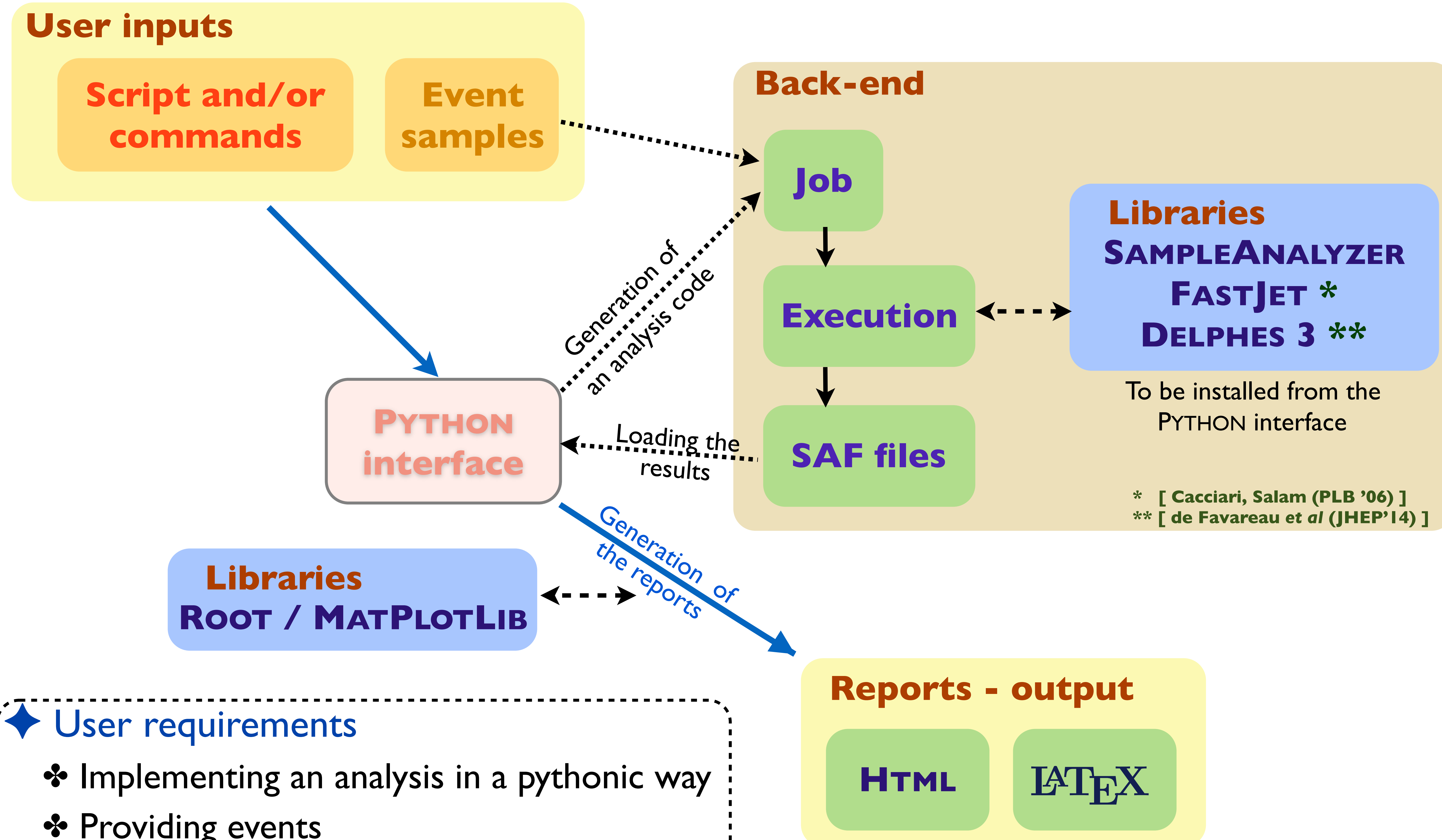
- ✤ Intuitive commands typed in a Python interface
- ✤ Analysis performed behind the scenes (a C++ black box)
- ✤ Human readable output: Html and LaTeX

✦ Expert mode

- ✤ C++ programming within the SampleAnalyzer framework (the MA5 core)

The (normal) user-friendly mode

[Conte, BF, Serret (CPC '13); Conte, Dumont, BF, Wymant (EPJC '14)]

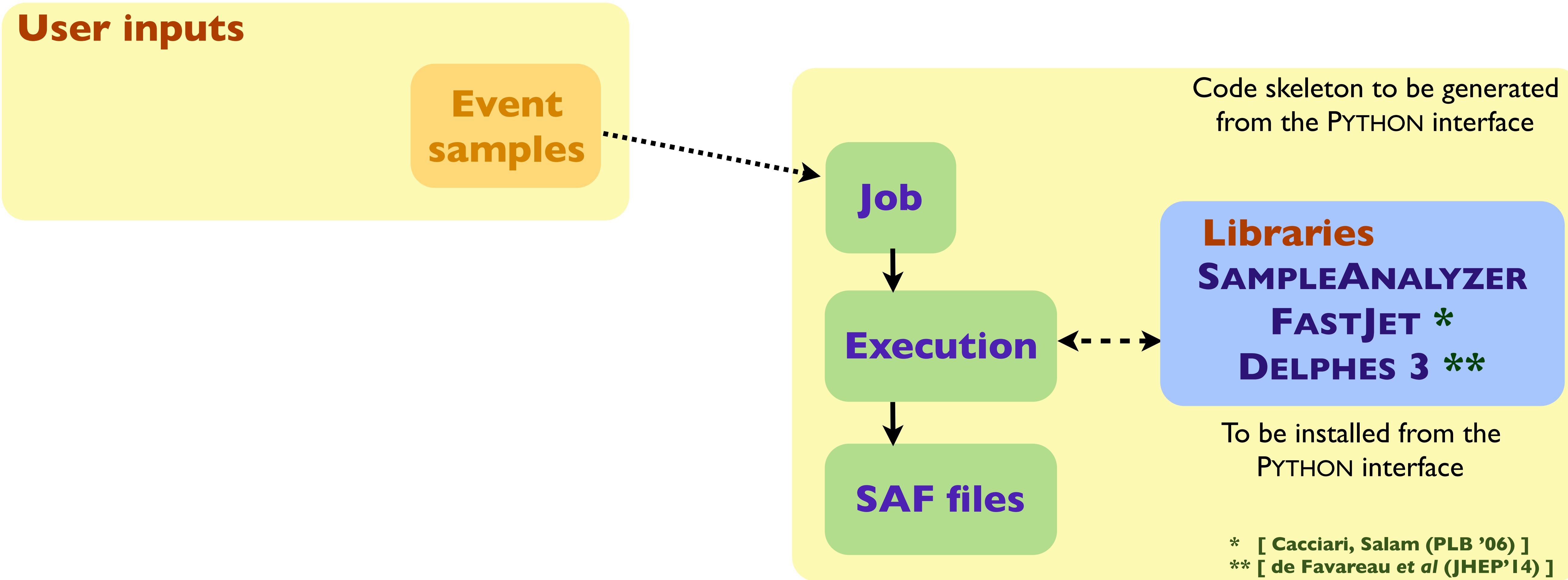


✦ User requirements

- ✤ Implementing an analysis in a pythonic way
- ✤ Providing events

The expert (developer-friendly) mode

[Conte, BF, Serret (CPC '13); Conte, Dumont, BF, Wymant (EPJC '14)]



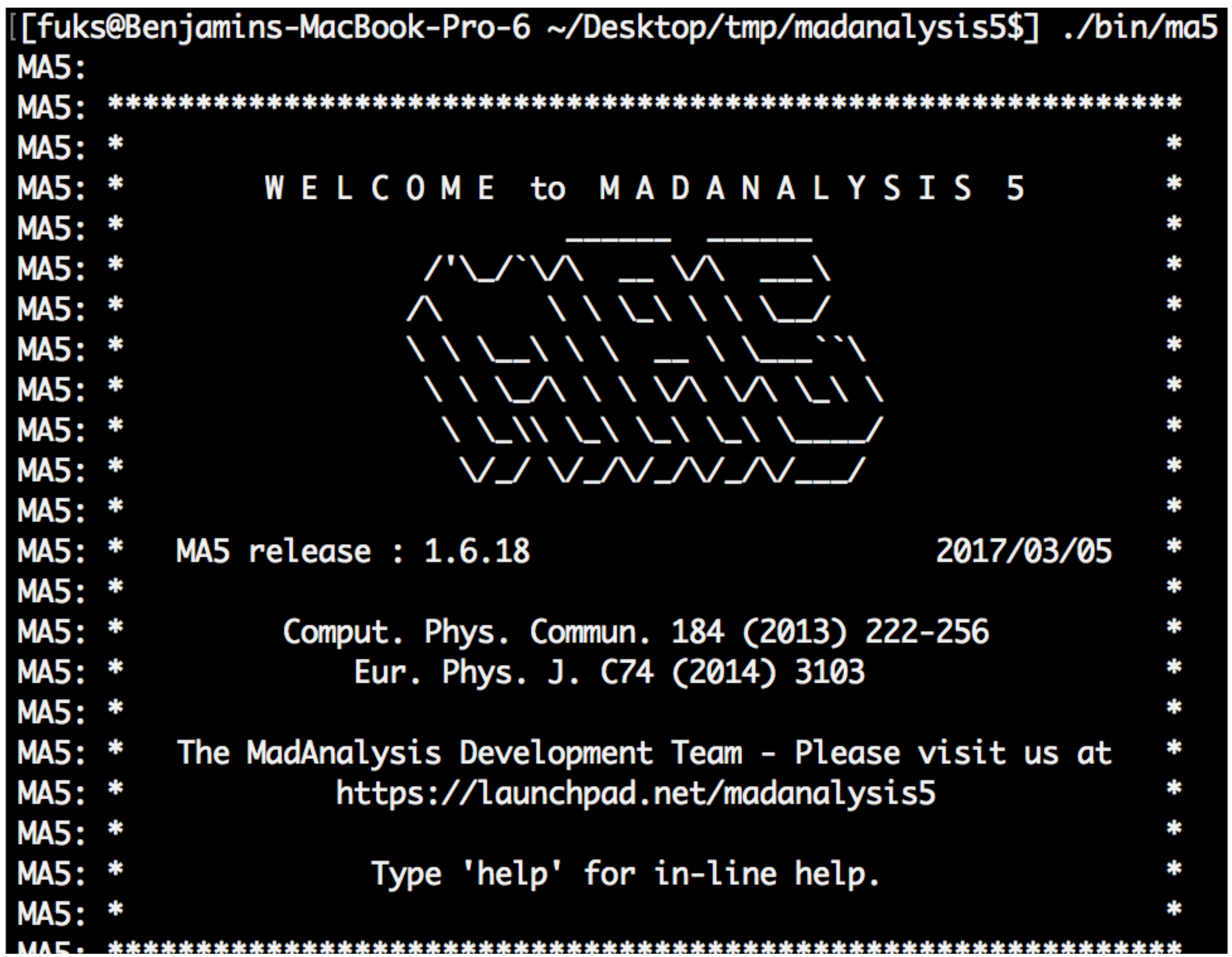
✦ User requirements

- ✤ Implementing a C++ analysis: a skeleton can be automatically generated
- ✤ Providing events
- ✤ Treatment of the output by the user

Getting started in half a second

✦ Installing the program

- ✤ Use version v1.6 beta
- ✤ Unpacking the tar-ball: tar xf ma5_xxx.tgz
- ✤ Start the code: ./bin/ma5



Getting started: dependencies

✦ Requirements (checked on start-up)

- ✤ PYTHON 2.6 or more recent (but not 3.X)
- ✤ The GCC compiler
- ✤ gmake

✦ Optional add-ons

- ✤ ZLIB (compressed event files)
  (needed for the course ➢ see later)
- ✤ FASTJET 3.2.X or more recent
  (needed for the course ➢ see later)
- ✤ ROOT 6.X (or 5.27 or more recent)
  (needed for the course ➢ see later)
- ✤ DELPHES 3 (detector simulation)
  (needed for the course ➢ see later)
- ✤ DELPHES 3-MA5Tune (deprecated)
- ✤ LATEX, PDFLATEX, DVIPDF (reports)
- ✤ The NUMPY library (recasting)

The requirements

```
MA5: Platform: Darwin 16.4.0 [MAC/OSX mode]
MA5: Reading user settings ...
MA5: Checking mandatory packages:
MA5:      - Python                      [OK]
MA5:      - GNU GCC g++                 [OK]
MA5:      - GNU Make                    [OK]
MA5: Checking optional packages devoted to data processing:
MA5:      - Zlib                        [OK]
MA5:      - FastJet                     [DISABLED]
MA5:      - Root                        [OK]
MA5:      - Delphes                     [DISABLED]
MA5:      - Delphes-MA5tune             [DISABLED]
MA5: Checking optional packages devoted to histogramming:
MA5:      - Root                        [OK]
MA5:      - Matplotlib                  [OK]
MA5:      - pdflatex                    [OK]
MA5:      - latex                       [OK]
MA5: Package used for graphical rendering: Root
```

The options

Getting started: compiling the MA5 core

- If the MA5 core is not up-to-date: compilation
  - In case of errors, check the log files

```
MA5: Checking the MadAnalysis 5 core library:
MA5:   => First use of MadAnalysis (or the library is missing).
MA5:
MA5:    ***********************************************************
MA5:                 Building SampleAnalyzer libraries
MA5:    ***********************************************************
MA5:    How many cores for the compiling? default = max = 8
MA5:    => Number of cores used for the compilation = 8
MA5:    Writing the setup files ...
MA5:    Writing all the Makefiles ...
MA5:    ***********************************************************
MA5:    Component 1/13 - test program: SampleAnalyzer configuration
MA5:      - Cleaning the project before building the test program ...
MA5:      - Compiling the source files ...
MA5:      - Linking the test program ...
MA5:      - Checking that the test program is properly built ...
MA5:      - Cleaning the project after building the test program ...
MA5:      - Running the test program ...
MA5:      - Checking the program output...
MA5:       => Status: [OK]
MA5:    ***********************************************************
    ⋮
```

Getting started: invisible and hadronic stuff

✦ Creation of useful containers

✤ Important for hadronic and MET-related observables

```
MA5:    ************************************************************
MA5:    Component 13/13 - test program: SampleAnalyzer core
MA5:      - Cleaning the project before building the test program ...
MA5:      - Compiling the source files ...
MA5:      - Linking the test program ...
MA5:      - Checking that the test program is properly built ...
MA5:      - Cleaning the project after building the test program ...
MA5:      - Running the test program ...
MA5:      - Checking the program output...
MA5:       => Status: [OK]
MA5:    ************************************************************
MA5:
MA5: ***************************************************************
MA5: Particle labels exported from madanalysis/input/particles_name_default.txt
MA5:   => 87 particles successfully exported.
MA5: Multiparticle labels exported from madanalysis/input/multiparticles_default.txt
MA5:   => Creation of the label 'invisible' (-> missing energy).
MA5:   => Creation of the label 'hadronic' (-> jet energy).
MA5:   => 8 multiparticles successfully exported.
```

Getting started: installing external libraries

- Installing the missing libraries: FASTJET
  - Automated installation
  - Type '*install fastjet*' in the interpreter

```
 - FastJet                    [DISABLED]
```

```
ma5>install fastjet
MA5:
MA5:    ***************************************************
MA5:                       Installing fastjet
MA5:    ***************************************************
MA5:    Detecting a previous installation ...
MA5:    => not found. OK
MA5:    How many cores would you like to use for the compilation ? default = max = 8
```

  - Restart MADANALYSIS 5

- Works for many of the optional packages:
  - Tab completion works, for the list

```
ma5>install
delphes        fastjet       matplotlib    PAD            root          zlib
delphesMA5tune gnuplot       numpy         PADForMA5tune  samples
```

```
MA5: Checking optional packages devoted to data processing:
MA5:       - Zlib                       [OK]
MA5:       - FastJet                    [OK]
MA5:       - Root                       [OK]
MA5:       - Delphes                    [OK]
MA5:       - Delphes-MA5tune            [DISABLED]
```

Basic concepts of the normal mode (I)

Looking for help...

- In-line help (in the interpreter)
  - ★ Listing all possible commands
- Auto-completion using the tab key

```
ma5>
define                  help          reset
define_region           history       restart
display                 import        resubmit
display_datasets        install       select
display_multiparticles  open          set
display_particles       plot          shell
display_regions         quit          submit
EOF                     reject        swap
exit                    remove
```

Datasets

- Event file format automatically detected
- Events files associated with labels
- Supported file formats: LHE, STDHEP, HEPMC, LHCO, ROOT (DELPHES 3)
- Several samples can be grouped (e.g., to increase statistics)
- Wildcards can be employed

```
ma5>import ttbar* as ttbar
MA5:    -> Storing the file 'ttbar.hep.gz' in the dataset 'ttbar'.
MA5:    -> Storing the file 'ttbar2.hep.gz' in the dataset 'ttbar'.
ma5>import Wjets.hep.gz as W
MA5:    -> Storing the file 'Wjets.hep.gz' in the dataset 'W'.
ma5>import VV.hep as diboson
MA5:    -> Storing the file 'VV.hep' in the dataset 'diboson'.
```

Basic concepts of the normal mode (2)

Particles and multiparticles

- Particles and multiparticles are defined via their PDG code (*labels*)
  (multi)particle labels make our life easier
- Default:
  - ★ Standard Model labels: as in MadGraph
  - ★ MSSM labels: as in MadGraph
  - ★ invisible: computation of observables related to the missing energy
  - ★ hadronic: computation of observables related to the hadronic activity
- Can be imported from a UFO model

```
ma5>define TheMuon = 13
ma5>define TheAntiMuon = -13
ma5>define AllMuons = TheMuon TheAntiMuon
ma5>display l+
   The multiparticle 'l+' is defined by the PDG-ids -15 -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 1000039.
ma5>remove TheMuon
ma5>display TheMuon
 ** ERROR: no object called 'TheMuon' found.
```

Basic concepts of the normal mode (3)

Histograms - the command *plot*

- *plot*: creation of an histogram
- Global observables, related to the full event (MET, H_T, *etc.)*
- Properties of a particle type (p_T, E, *etc.*)
- Particle ordering can be used
- Particles can be combined
- Virtual particles can be studied
- Log scales can be employed
- Ways to normalize an histogram

```
ma5>plot MET [
]             ETAordering   initialstate  logY          PTordering    PZordering
allstate      ETordering    interstate    normalize2one PXordering    stack
Eordering     finalstate    logX          Pordering     PYordering    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 of the normal mode (4)

Selection cuts - the commands *reject/select*

- Events can be selected/rejected
- Particle candidates to consider

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

All ingredients are there

Basic concepts of the normal mode (5)

Executing the analysis - the command *submit*

- ✤ Creates a C++ code
- ✤ Compilation / execution
- ✤ Histogramming / cuts
- ✤ Reports generation

```
ma5>submit
MA5:    Creating folder 'ANALYSIS_3'...
MA5:    Copying 'SampleAnalyzer' source files...
MA5:    Inserting your selection into 'SampleAnalyzer'...
MA5:    Writing the list of datasets...
MA5:    Writing the command line history...
MA5:    Creating Makefiles...
MA5:    Compiling 'SampleAnalyzer'...
MA5:    Linking 'SampleAnalyzer'...
MA5:    Running 'SampleAnalyzer' over dataset 'defaultset'...
MA5:     *******************************************************
    * SampleAnalyzer for MadAnalysis 5 - Welcome.
    * Initializing all components
      - version: 1.6.18 (2017/03/05)
      - general: everything is default.
      - extracting the list of event samples...
      - analyzer 'MadAnalysis5job'
    * Running over files ...
    * 1/1  /Users/fuks/Work/tools/madanalysis/bzr/v1.6beta/samples/jack_unw.lhe
        => file size: 312.12 ko
        => sample format: LHE file produced by MadGraph5.
        => progress: [====================================]
        => total number of events: 100 ( analyzed: 100 ; skipped: 0 )
    * Finalizing all components ...
    * Total number of processed events: 100.
    * Goodbye.
MA5:     *******************************************************
MA5:    Checking SampleAnalyzer output...
MA5:    Extracting data from the output files...
MA5:    Preparing data for the reports ...
MA5:    Generating all plots ...
MA5:    Generating the HMTL report ...
MA5:      -> To open this HTML report, please type 'open'.
MA5:    Generating the PDF report ...
MA5:      -> To open this PDF report, please type 'open ANALYSIS_3/PDF'.
MA5:    Generating the DVI report ...
MA5:    Well done! Elapsed time = 6 seconds
```

Outline

Toy event samples

✦ Setup

- ✤ LHC collider at a center-of-mass energy of 13 TeV, 2.3 fb^{-1}
- ✤ No lepton cut (pseudorapidity, transverse momentum, *etc.*)
- ✤ Jet cuts: p_T > 20 GeV, ΔR_{jj} > 0.4, no pseudorapidity cut

✦ Toy samples: parton-level (no shower/hadronization)

- ✤ Top-antitop: two leptonic decays $t\bar{t} \to (b\ell^+\nu_\ell)(\bar{b}\ell'^-\bar{\nu}_{\ell'})$
- ✤ WW: two leptonic decays $W^+W^- \to (\ell^+\nu_\ell)(\ell'^-\bar{\nu}_{\ell'})$
- ✤ ZZ = signal: one leptonic and one invisible decay $ZZ \to (\nu_\ell\bar{\nu}_\ell)(\ell'^+\ell'^-)$

✦ NLO cross sections (beware: LO generation, NLO normalization)

- ✤ Top-antitop pairs ≈ 33 pb
- ✤ WW ≈ 4.2 pb
- ✤ ZZ ≈ 0.3 pb

Importing and defining the samples

- Importing the LHE samples, setting their properties
  - We define the type (signal or background) of each dataset
  - We assign cross section numbers and the luminosity ➢ correct normalization

```
ma5>import zz.lhe.gz as zz
   -> Storing the file 'zz.lhe.gz' in the dataset 'zz'.
ma5>import ww.lhe.gz as ww
   -> Storing the file 'ww.lhe.gz' in the dataset 'ww'.
ma5>import ttbar.lhe.gz as ttbar
   -> Storing the file 'ttbar.lhe.gz' in the dataset 'ttbar'.
ma5>set ttbar.xsection = 33
ma5>set ww.xsection = 4.2
ma5>set zz.xsection = 0.3
ma5>set ttbar.type=background
ma5>set ww.type=background
ma5>set zz.type=signal
ma5>set main.lumi = 2.3
```

Importing the samples

Cross sections [in pb]

Signal and background definitions

Luminosity [in fb^{-1}]

Getting closer to a detector (at parton-level)

- ✦ We have not simulated any detector response
  - ✤ Include reasonable selections getting us closer to a real experiment

- ✤ Soft objects are not detected
  - ★ Removal of any jet and lepton that is softer than some threshold

```
ma5>define l = l+ l-
ma5>select (l) PT > 10
ma5>select (j) PT > 20
```

(a new multiparticle label *l* is created)

- ✤ Objects lying outside the detector are not detected
  - ★ Removal of any object lying outside the detector acceptance

```
ma5>select (l) -2.5 < ETA < 2.5
ma5>select (j) -2.5 < ETA < 2.5
```

- ✤ Objet overlap removal
  - ★Any charged lepton too close to a jet is removed (isolation)

```
ma5>reject (l) DELTAR(j) < 0.4
```

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 (unphysical) partonic center-of-mass energy (*SQRTS*)
  - The α_T variable (*ALPHAT*)
  - The particle content of the event (*NPID, NAPID, N*)

- General setup for drawing histograms
  - Superimposing curves on a single histogram (*set main.stacking_method* = ...)

```
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
  - Submitting and executing: *submit*
  - Getting to the results: *open* (open the HTML report)

```
ma5>submit; open
```

Results: sample information

✦ Top-antitop sample information

ttbar

- Sample consisting of: background events.
- Generated events: 10000 events.
- * Cross section imposed by the user: 33.0 pb.
- Normalization to the luminosity: 75900 +/- 0 events.
- Ratio (event weight): 7.6 - warning: please generate more events (weight larger than 1)!

Path to the event file	Nr. of events	Cross section (pb)	Negative wgts (%)
ttbar.lhe.gz	10000	33.0	0.0

- ✤ Signal/background nature
- ✤ Number of events, cross section, normalization, event weight
- ✤ Negative weight information

✦ Example 2: signal sample information

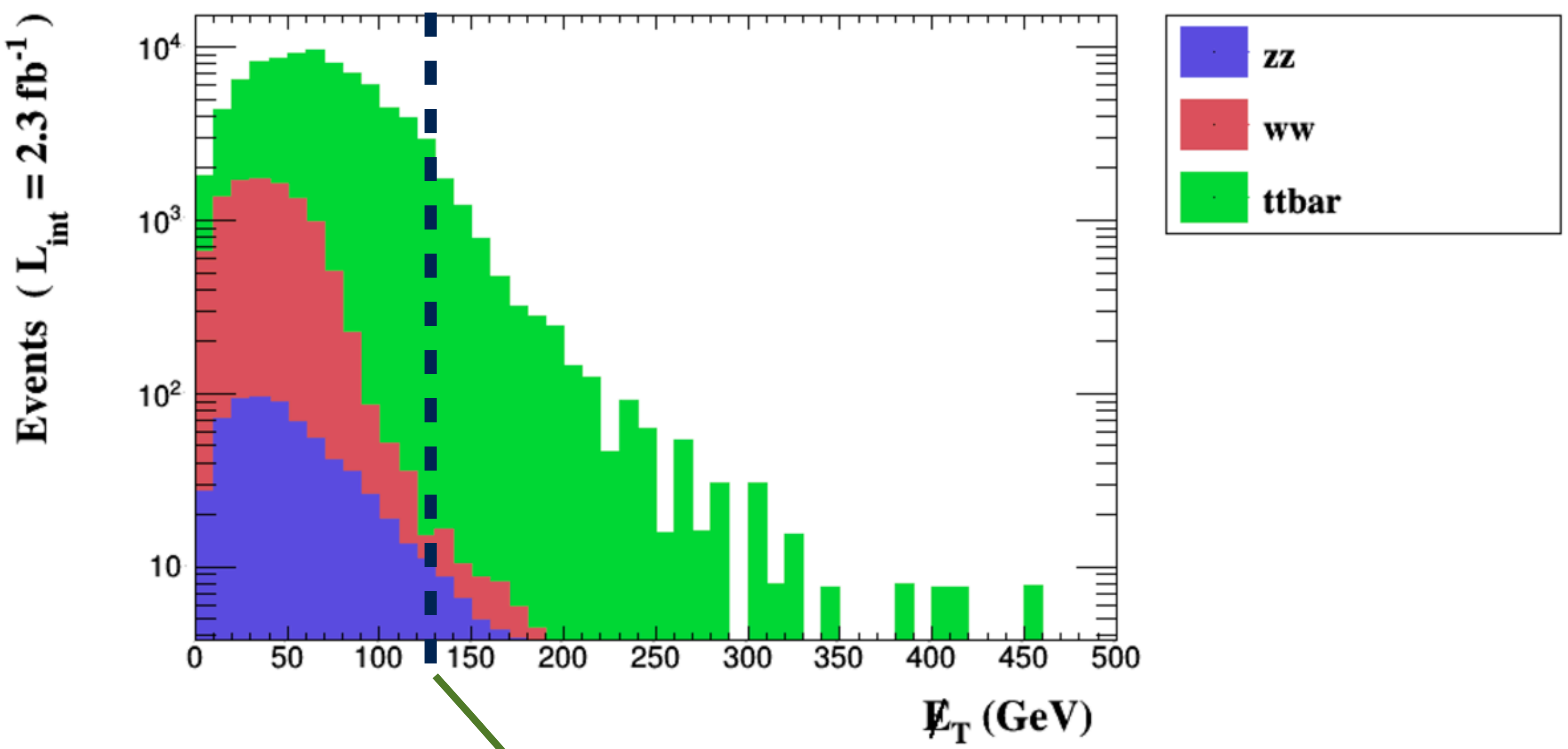
zz

- Sample consisting of: signal events.
- Generated events: 10000 events.
- * Cross section imposed by the user: 0.3 pb.
- Normalization to the luminosity: 690 +/- 0 events.
- Ratio (event weight): 0.069 .

Path to the event file	Nr. of events	Cross section (pb)	Negative wgts (%)
zz.lhe.gz	10000	0.3	0.0

Results: histograms

✦ Missing energy distribution



✤ This suggest a selection cut: MET < 125 GeV

✦ MadAnalysis 5 also returns statistics information for each histogram

Dataset	Integral	Entries / events	Mean	RMS	%Underflow	%Overflow
zz	690	1.0	57.9969	47.37	0.0	0.05
ww	9660	1.0	39.7918	21.81	0.0	0.0
ttbar	75899	1.0	73.7078	39.9	0.0	0.0

Cut implementation

✦ Implementing the missing energy cut `ma5>select MET < 125`

How to choose a cut?
- ★ Large signal efficiency
- ★ Small background efficiencies

Cut 6

Cut: select MET < 125.0

Dataset	Events kept: K	Rejected events: R	Efficiency: K / (K + R)	Cumul. efficiency: K / Initial
zz	640.80 +/- 6.76	49.20 +/- 6.76	0.9287 +/- 0.0098	0.9287 +/- 0.0098
ww	9633.9 +/- 5.1	26.1 +/- 5.1	0.997300 +/- 0.000528	0.997300 +/- 0.000528
ttbar	68826.1 +/- 80.1	7073.9 +/- 80.1	0.90680 +/- 0.00106	0.90680 +/- 0.00106

✦ S *vs* B evolution - the formula can be changed (*set main.fom.formula = …*)

Cut-flow chart

- How to compare signal (S) and background (B): S/sqrt(S+B).

Cuts	Signal (S)	Background (B)	S vs B
Initial (no cut)	690	85559	2.35
Cut 1	690	85559	2.35
Cut 2	690	85559	2.35
Cut 3	690	85559	2.35
Cut 4	690	85559	2.35
Cut 5	690	85559	2.35
Cut 6	640.80 +/- 6.76	78460.1 +/- 80.3	2.278 +/- 0.024

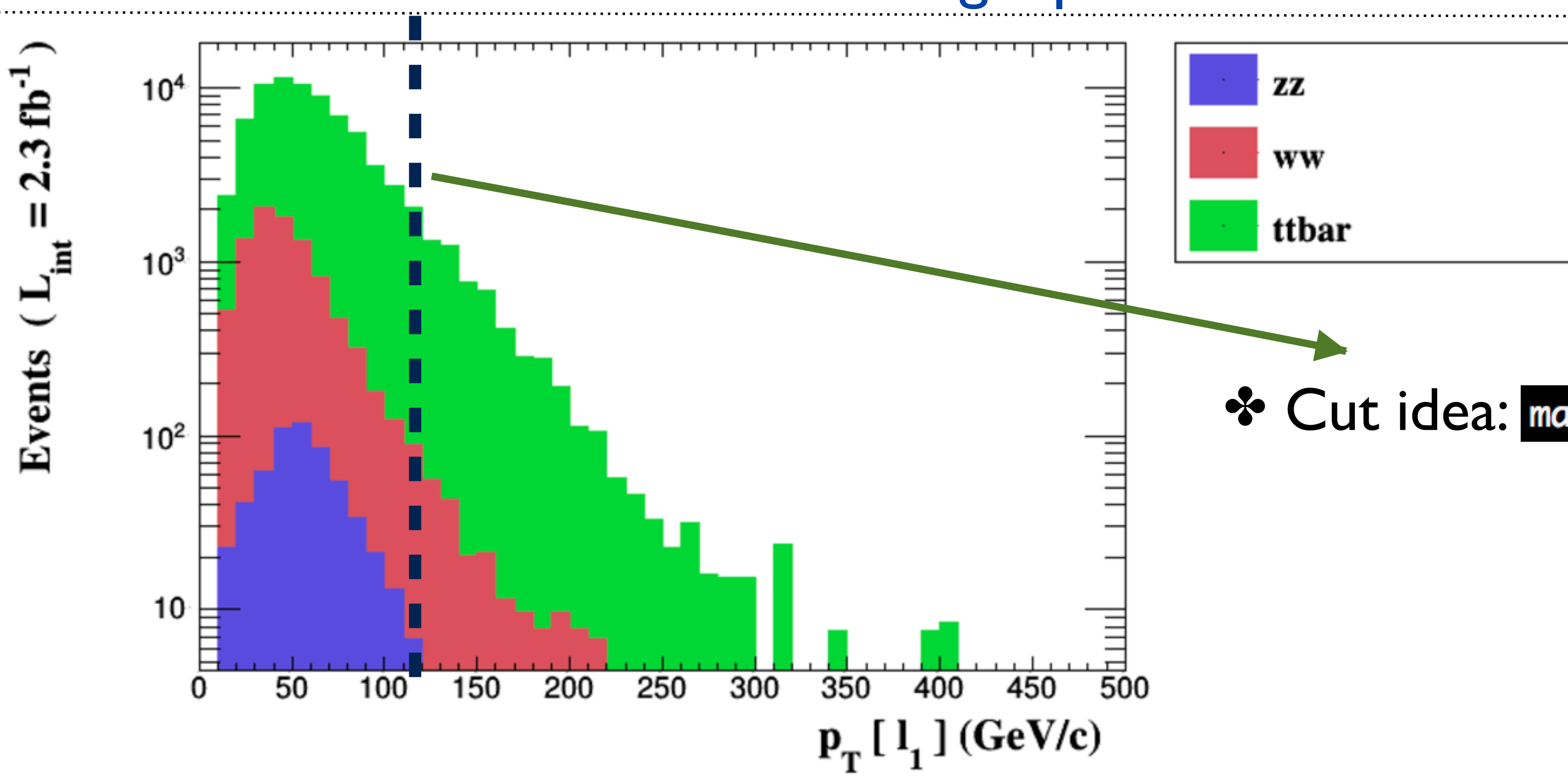
✤The first cuts are due to the object definitions (no modification of the event count)

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)
- The particles are ordered
  - Use squared brackets to select a specific particle
  - Several ordering variables are available (*PT, E, PX, etc.*)

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

- Transverse momentum of the leading lepton



- Cut idea: `ma5>reject PT(l[1]) > 125`

Particle properties: special features

✦ Combining particles

- Replace the single argument by several particles
- Four-momenta are summed before computing the observable
- 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]
```

✦ Four special functions

- *DELTAR*, *DPHI_0_PI*, *DPHI_0_2PI*: take two arguments
- *MT_MET*: transverse mass when combining a particle with the MET

```
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]
```

Novelty: signal regions can be implemented

✦ Several signal regions can be defined

✤ Cuts and histograms can be assigned to given regions

```
ma5>define_region SR1 SR2 CR1
ma5>select MET > 150 {SR1 CR1}
ma5>plot MET {SR1 SR2}
MA5-WARNING:    Histogram found to be attached to distinguishable regions -> multiple declaration:
MA5-WARNING:    * Plot: MET { SR2 }
MA5-WARNING:    * Plot: MET { SR1 }
```

✦ Cutflows for each region

Region: "SR1"			
Cuts	Signal (S)	Background (B)	S vs B
Initial (no cut)	0.021076 +/- 0.000542		
SEL: MET > 150.0	0.000146 +/- 0.012041		

Region: "SR2"			
Cuts	Signal (S)	Background (B)	S vs B
Initial (no cut)	0.021076 +/- 0.000542		

Region: "CR1"			
Cuts	Signal (S)	Background (B)	S vs B
Initial (no cut)	0.021076 +/- 0.000542		
SEL: MET > 150.0	0.000146 +/- 0.012041		

Outline

Reconstructing showered/hadronized events

- The output of a parton shower code is non-practical for an analysis
  - It contains tons of hadrons
  - We prefer to employ jets rather than individual hadrons
    - ➢ Jets have to be reconstructed
  - The event file is non-readable with human eyes (StdHep, HepMC)
  - 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
    - ★ If not, can be locally installed: `ma5>install fastjet`

Basic concepts for jet clustering

Jet clustering and basic detector effects

- Running of FASTJET via the MADANALYSIS 5 interpreter (in the *reco* mode)
- b/tau-tagging efficiencies/mistag rates can be included
- The reconstructed events can be redirected to a file
  - The output file can be used for post-processing
- Can also be used for checking the merging procedure
  - Differential jet rate distributions

```
ma5>set main.fastsim.package =
delphes        delphesMA5tune fastjet        none
ma5>set main.fastsim.package = fastjet
ma5>set main.fastsim.algorithm =
antikt      cambridge   cdfjetclu   cdfmidpoint genkt       gridjet     kt          none        siscone
ma5>set main.fastsim.algorithm = antikt
ma5>set main.fastsim.
main.fastsim.algorithm           main.fastsim.bjet_id.misid_cjet  main.fastsim.ptmin
main.fastsim.bjet_id.efficiency  main.fastsim.bjet_id.misid_ljet  main.fastsim.radius
main.fastsim.bjet_id.exclusive   main.fastsim.exclusive_id        main.fastsim.tau_id.efficiency
main.fastsim.bjet_id.matching_dr main.fastsim.package             main.fastsim.tau_id.misid_ljet
ma5>set main.fastsim.bjet_id.efficiency = 0.60
ma5>set main.fastsim.bjet_id.misid_cjet = 0.10
ma5>set main.fastsim.bjet_id.misid_ljet = 0.01
ma5>set main.outputfile = blabla.lhe
ma5>
```

Jet clustering in practice

- Jet reconstruction with MADANALYSIS 5 (and FASTJET)
  - MADANALYSIS 5 must be run in the reconstructed mode: ./bin/ma5 -R

```
ma5>set main.fastsim.package =
delphes        delphesMA5tune fastjet        none
ma5>set main.fastsim.package = fastjet
ma5>set main.fastsim.
main.fastsim.algorithm           main.fastsim.bjet_id.misid_cjet  main.fastsim.ptmin
main.fastsim.bjet_id.efficiency  main.fastsim.bjet_id.misid_ljet  main.fastsim.radius
main.fastsim.bjet_id.exclusive   main.fastsim.exclusive_id        main.fastsim.tau_id.efficiency
main.fastsim.bjet_id.matching_dr main.fastsim.package             main.fastsim.tau_id.misid_ljet
ma5>set main.fastsim.algorithm =
antikt      cambridge   cdfjetclu   cdfmidpoint genkt       gridjet     kt          none        siscone
ma5>set main.fastsim.algorithm = antikt
ma5>set main.fastsim.radius = 0.4
ma5>set main.fastsim.bjet_id.efficiency = 0.60
ma5>set main.fastsim.bjet_id.misid_cjet = 0.10
ma5>set main.fastsim.bjet_id.misid_ljet = 0.01
```

Many clustering algorithms available (with their options)

Realistic b-tagging

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

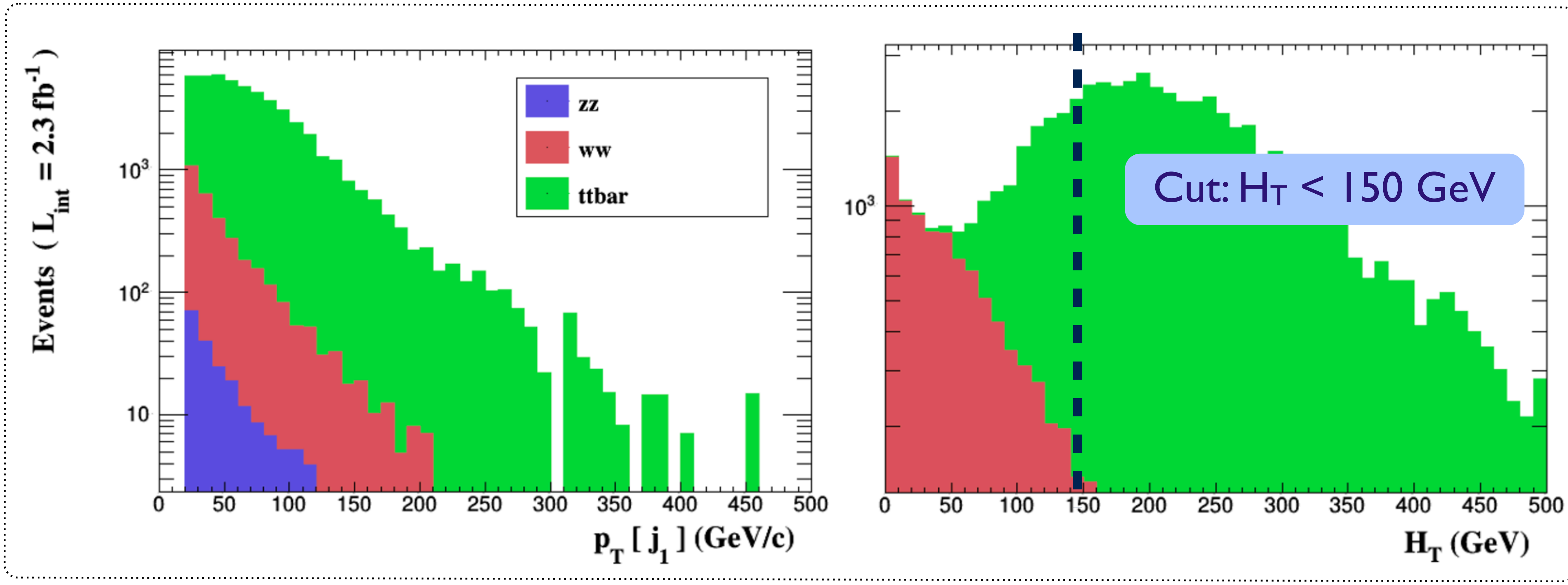
- LHE/LHCO files are created (and can be further used)

Investigating the event jet activity

- ✦ Studying the kinematical properties of the jets
  - ✤ *BETA, E, ET, ETA, GAMMA, M, MT, P, PHI, PT, PX, PY, PZ, R, THETA, Y*
- ✦ Reminder: the particles (and thus the jets) are ordered
  - ✤ Use squared brackets to select a specific particle
- ✦ Cuts: we keep the previously mentioned cuts

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

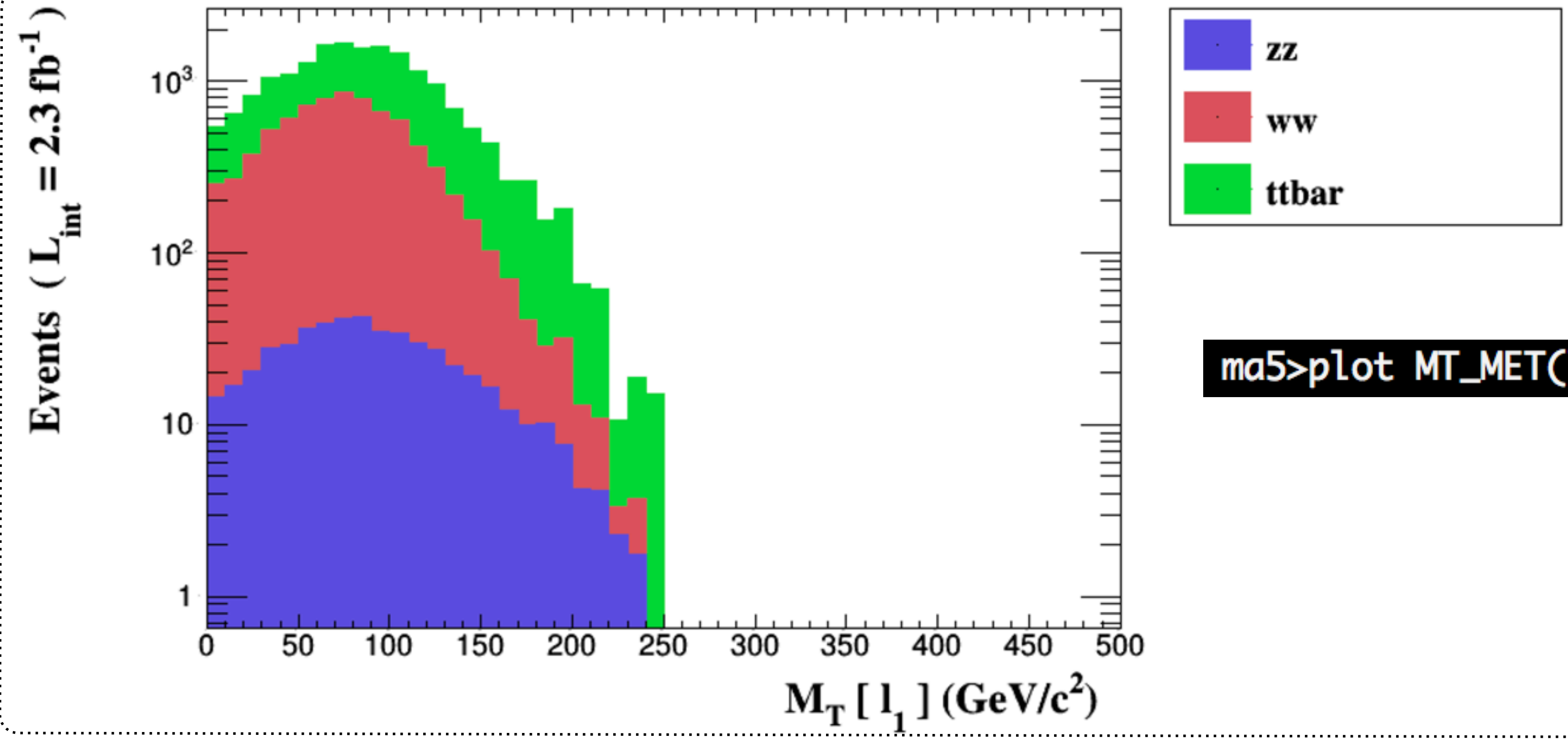
- ✦ Transverse momentum of the leading jet and total hadronic activity



- ✤ H_T distribution: the signal is hidden (*cf.* *y*-axis range)

Final results: missing momentum and cutflow

✦ Reconstructed 'W transverse mass'




```
ma5>plot MT_MET(l[1]) 50 0 500 [logY]
```

✦ Cut flow chart: 4σ is reachable!

Cuts	Signal (S)	Background (B)	S vs B
Initial (no cut)	690	85559	2.35
Cut 1	690	85559	2.35
Cut 2	690	85559	2.35
Cut 3	690	85559	2.35
Cut 4	690	85559	2.35
Cut 5	690	85559	2.35
Cut 6	624.32 +/- 7.71	70481 +/- 110	2.3413 +/- 0.0288
Cut 7	611.39 +/- 8.35	65997 +/- 121	2.3689 +/- 0.0323
Cut 8	562.3 +/- 10.2	18996 +/- 102	4.0209 +/- 0.0727

Realistic detector simulation

- Event reconstruction with DELPHES (ROOT is mandatory)
  - DELPHES 3 (possibly with MA5TUNE detector cards) is interfaced to MADANALYSIS 5
  - Can be installed easi `ma5>install delphes`
  - Easy to use

```
ma5>set main.fastsim.package = delphes
ma5>set main.fastsim.detector = cms-ma5tune
```

  - Output: DELPHES ROOT file with all reconstructed objects (can be further used)

Fast simulation of the detector with DELPHES 3

- Running of DELPHES via the MADANALYSIS 5 interpreter (in the *reco* mode)
- Choice of ATLAS or CMS; pile-up can be included
- The ROOT output file is stored

```
ma5>set main.fastsim.package = delphes
ma5>set main.fastsim.
main.fastsim.detector          main.fastsim.pileup          main.fastsim.skim_genparticles
main.fastsim.output            main.fastsim.rootfile        main.fastsim.skim_towers
main.fastsim.package           main.fastsim.skim_eflow      main.fastsim.skim_tracks
ma5>set main.fastsim.detector =
atlas         atlas-ma5tune cms           cms-ma5tune
ma5>set main.fastsim.detector = cms-ma5tune
```

Outline

The expert mode of MADANALYSIS 5

[Conte, BF, Serret (CPC '13); Conte, Dumont, BF, Wymant (EPJC '14)]

✦ MADANALYSIS 5 without its PYTHON interface

- ✤ More freedom in the observables (only some can be called from the PYTHON console)
- ✤ Complicated cuts can be implemented
- ✤ More suitable for large numbers of events (using several cores)

✦ The expert mode is developer-friendly

The analysis is a C++ class

The SAMPLEANALYZER internal data format

- ★ Readers for LHE, STDHEP, HEPMC, LHCO and DELPHES
- ★ Many classes and methods for particle and object properties
- ★ Specific methods for histograms and cuts

Services

- ★ Physics observables (transverse variables, object identification, isolation)
- ★ Streamers
- ★ Exceptions

Interfaces

- ★ FASTJET
- ★ DELPHES 3
- ★ ROOT

Scripts

- ★ Compilation
- ★ Linking
- ★ Analysis skeleton generator

The new extension of the expert mode

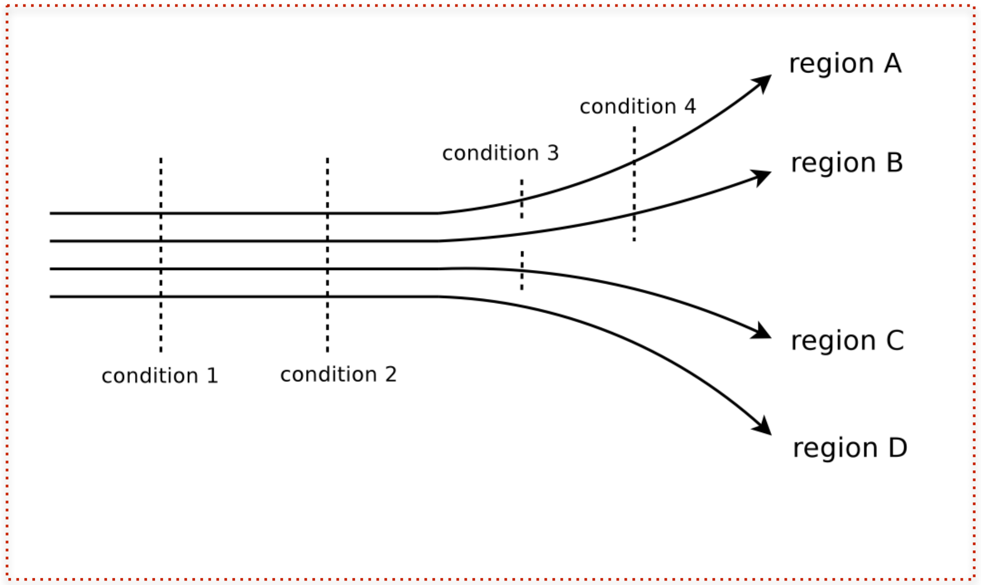
[Conte, BF, Serret (CPC '13); Conte, Dumont, BF, Wymant (EPJC '14)]

✦ Main features (enable a recast of most cut-based LHC analyses)

- ✤ Support for multiple sub-analyses (signal and control regions)
- ✤ New ready-to-use observables (M_{T2}, M_{T2W}, *etc.*)
- ✤ New optimized handling of cuts and histograms

✦ Handling cuts and histograms

- ✤ Naive approach not efficient (see cut #4 for instance)




```
count the event in region D
if (condition 3)
{
  count the event in region C
  if (condition 4)
  {
    count the event in region A
  }
}
if (condition 4)
{
  count the event in region B
}
```

- ✤ A more efficient algorithm has been implemented
  - ★ Each cut condition is only evaluated once
  - ★ It is applied to all surviving regions simultaneously
- ✤ Similar treatment for histograms

Implementing a new analysis (I)

✦ Recommendation: start from any existing recasted LHC analysis

- More information:
  http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase
- Installation: `ma5>install PAD`

✦ A pre-existing script allows for the generation of a new analysis skeleton

- The *newAnalyzer.py* script located in PAD/Build/SampleAnalyzer
- Add an empty analysis in PAD/Build/SampleAnalyzer/User/Analyzer/
- To be modified on the basis of the other analyses in that folder

✦ Declaring regions, cuts and histograms via the *Manager* (*Initialize* method)

- Region: `Manager()->AddRegionSelection("signal");`
- Histogram: `Manager()->AddHisto("MET_selected_mu",15,0,1500);`
- Cut:

```
Manager()->AddCut("2leptons", "SR1");
std::string SRs[] = {"SR1", "SR2"};
Manager()->AddCut("3leptons", SRs);
```

Number of bins, range

Weight management

- The *Execute* method must include an initialization of the weights
  - Correct cutflow management
  - Correct weights in the histograms
  - Efficiency of the application of the cuts

```
double myEventWeight;
if(Configuration().IsNoEventWeight()) myEventWeight=1.;
else if(event.mc()->weight()!=0.) myEventWeight=event.mc()->weight();
else
{
  WARNING << "Found one event with a zero weight. Skipping...\n";
  return false;
}
Manager()->InitializeForNewEvent(myEventWeight);
```

Key command

To be included in any single analysis

Objects

✦ Object definitions

- ✤ Key C++ class: the *event.rec()*
  - ★ Key collections: *electrons*(), *muons*(), *jets*(), …
- ✤ Illustrative loose and tight electron definition

```
std::vector<const RecLeptonFormat*> SignalLeptons, LooseElectrons;
for(unsigned int ii=0; ii<event.rec()->electrons().size(); ii++)
{
  const RecLeptonFormat *MyElec = &(event.rec()->electrons()[ii]);
  double eta = fabs(MyElec->eta());
  double pt = MyElec->pt();
  double iso_var = PHYSICS->Isol->eflow->sumIsolation(MyElec,
      event.rec(),0.4,0.,IsolationEFlow::ALL_COMPONENTS);

  if( (eta>1.44) and (eta<1.57) ) continue;
  if(eta>2.5) continue;

  if(iso_var>0.15*pt) continue;

  if (pt>20)  SignalLeptons.push_back(MyElec);
  if (pt>10)  LooseElectrons.push_back(MyElec);
}
```

Loop over all electrons

Acceptance

Isolation

p_T cuts

Cuts and histograms

✦ Cuts must be implemented in a very specific way

- ✤ Ensures the correct treatment of the cutflows
- ✤ Ensures an efficient running of the code

```
if(!Manager()->ApplyCut(cut1_cond,"2_leptons")) return true;
```

Cut condition

Declared cut name

✦ Filling a declared histograms is straightforward

- ✤ Ensures the correct treatment of the cutflows
- ✤ Ensures an efficient running of the code

```
Manager()->FillHisto("MET",MET);
```

Histogram name

Observable value

✦ More info: existing analyses or the manual (1405.3982 and 1407.3278)

Reinterpreting LHC physics analyses

- ✦ Exploit the full potential of the LHC (for new physics)
  - ✤ Priority #1 of the European strategy for particle physics
  - ✤ Designing new analyses to probe new ideas — Prospectives (based on MC simulations)
  - ✤ Recasting LHC analyses to study models not considered — The LHC legacy

- ✦ LHC data has been collected with significant human and financial efforts
  - ✤ Important for on-going analyses (within popular theoretical contexts of today)
  - ✤ Important for future opportunities (within future scientific contexts)
- ✦ Data preservation in high-energy physics is mandatory **[Kogler, South & Steder (JPCS'12)]**

- ✦ Related tools need to be supported by the entire community **[Kraml *et al.* (EPJC'12)]**
  - ✤ Both theorists and experimentalists
  - ✤ Allowing for the reinterpretation of the LHC analysis results

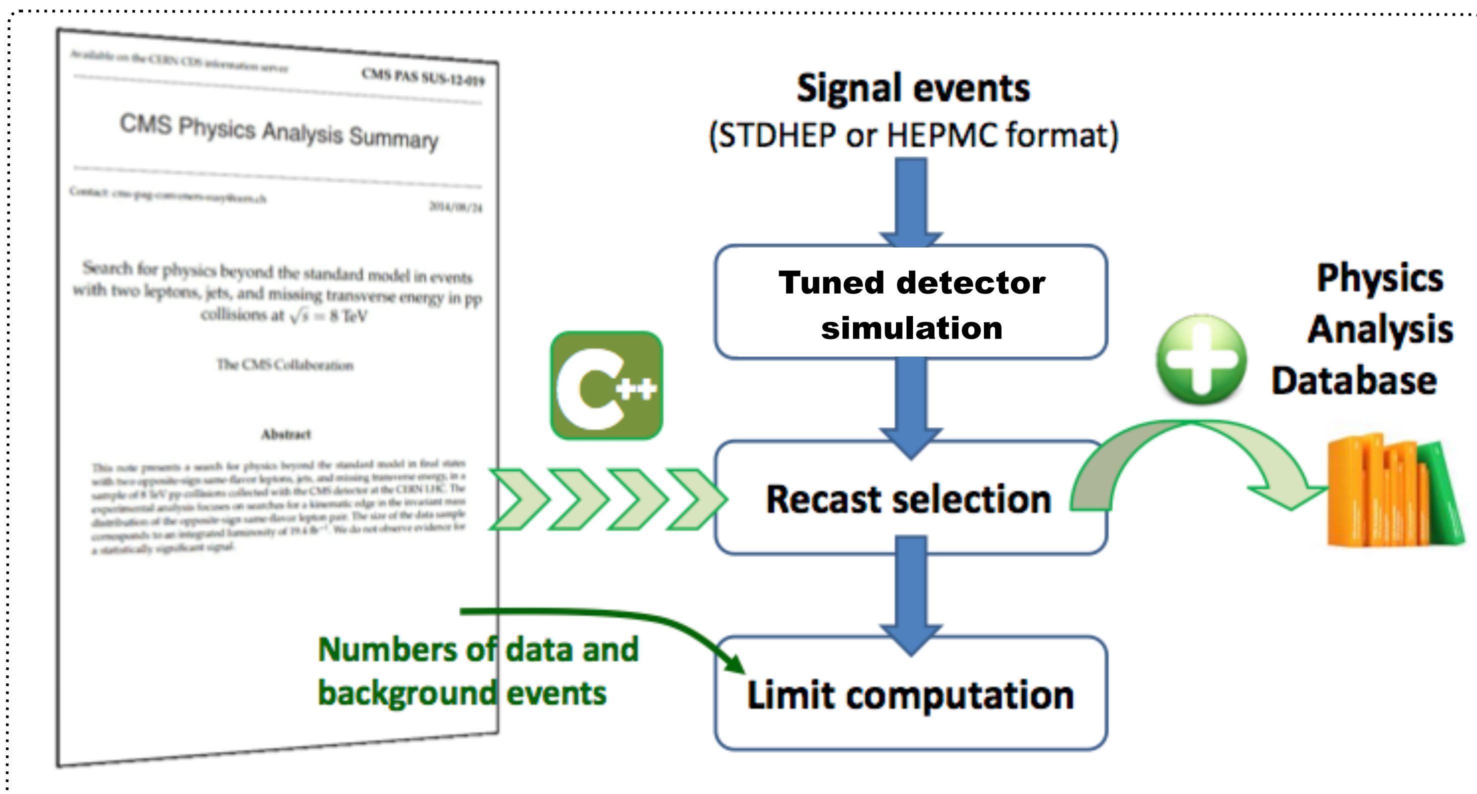
Recasting in MADANALYSIS 5

[Conte, Dumont, BF, Wymant (EPJC '14); Dumont, BF, Kraml *et al.* (EPJC '15)]

✦ There are plethora of new physics realizations that deserve to be studied

- Experimentalists cannot study all the options
- Our choice: rely on a public detector simulator mimicking ATLAS and CMS
- Need for a (public) framework where LHC analyses can be easily implemented

✦ The recasting strategy in MADANALYSIS 5



- 2 options for detector effects
  - ★ DELPHES/PGS-like (resolutions, efficiencies, *etc.*)
  - ★ RIVET-like (transfer functions)

Recasting made easy with MadAnalysis 5 (I)

[Conte, Dumont, BF, Wymant (EPJC '14); Dumont, BF, Kraml *et al.* (EPJC '15)]

✦ Confronting a BSM signal to LHC analyses is straightforward

- ✤ Starting point: a showered/hadronized event file
- ✤ Installation of the detector simulators: 'install Delphes'
- ✤ Installation of the analysis libraries: 'install PAD' (more analyses with the MA5tune)

✦ In practice:

```
ma5>set main.recast = on
MA5-WARNING: DelphesMA5tune and/or the PADForMA5tune are not installed (or deactivated): t
he corresponding analyses will be unavailable
ma5>import samples/stops.hep.gz
MA5:    -> Storing the file 'stops.hep.gz' in the dataset 'defaultset'.
ma5>submit
MA5:    Creating folder 'ANALYSIS_0'...
MA5: Would you like to edit the recasting Card ? (Y/N)
Answer: Y
```

✦ Snippet of the recasting card (only on/off switches to be set by the user)

✤O(20) 8 TeV ATLAS and CMS analyses; 2 13 TeV ATLAS analyses (+10 new analyses soon)

```
atlas_1605_03814     v1.2    on    delphes_card_ATLAS_1604_07773.tcl        # ATLAS - 13 TeV - multijet (2-6 jets) + met
ATLAS_1604_07773     v1.2    on    delphes_card_ATLAS_1604_07773.tcl        # ATLAS - 13 TeV - monojet
ATLAS_EXOT_2014_06   v1.2    on    delphes_card_atlas_sus_2013_05_pad.tcl   # ATLAS - 8 TeV - monophoton
cms_exo_12_047       v1.2    on    delphes_card_cms_b2g_12_012.tcl          # CMS   - 8 TeV - monophoton
cms_exo_12_048       v1.2    on    delphes_card_cms_b2g_12_012.tcl          # CMS   - 8 TeV - monojet
cms_b2g_14_004       v1.2    on    delphes_card_cms_b2g_14_004.tcl          # CMS   - 8 TeV - Dark matter production with a ttbar pair
cms_b2g_12_022       v1.2    on    delphes_card_cms_b2g_14_004.tcl          # CMS   - 8 TeV - Monotop search
CMS_B2G_12_012       v1.2    on    delphes_card_cms_b2g_12_012.tcl          # CMS   - 8 TeV - T5/3 partners in the SSDL channel
```

Recasting made easy with MADANALYSIS 5 (2)

[Conte, Dumont, BF, Wymant (EPJC '14); Dumont, BF, Kraml *et al.* (EPJC '15)]

- Snippet of the output file (example: low statistics ➢ lots of '-1' in the example)
  - CLs if a signal cross section is provided
  - Cross sections excluded at the 95% CL

ATLAS_1604_07773	EM1	25.8538538	27.4980471	||	0.0100000	0.0099499	0.0000000	0.0099499
ATLAS_1604_07773	EM2	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	EM3	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	EM4	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	EM5	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	EM6	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	EM7	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	IM1	58.3118133	52.7020233	||	0.0100000	0.0099499	0.0000000	0.0099499
ATLAS_1604_07773	IM2	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	IM3	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	IM4	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	IM5	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000
ATLAS_1604_07773	IM6	-1	-1	||	0.0000000	0.0000000	0.0000000	0.0000000

Implementing a new analysis in MADANALYSIS 5

- ✦ Picking up an experimental publication
  - ✤ Reading
  - ✤ Understanding

✅ **Relatively easy**

- ✦ Writing the analysis code in the tool internal language

✅ **Relatively easy**

- ✦ Getting the information missing from the publication for a proper validation
  - ✤ Efficiencies (trigger, electrons, muons, b-tagging, JES, *etc.*)
    - ★ Including p_T and/or η dependence
    - ★ Accurate information
  - ✤ Detailed cutflows for some well-defined benchmark scenarios
    - ★ Exact definition of the benchmarks (spectra)
    - ★ Event generation information (cards, tunes, *etc.*)
  - ✤ Expected number of events in each region and cross sections
  - ✤ Digitized histograms (e.g., on HEPDATA)

⚠️ **Essential**
❌ **Often difficult!**

- ✦ Comparing theory tools and real life

Recasting CMS-EXO-12-048

[Conte, BF, Guo ('16)]

✦ Missing information for the validation

- Discussion with CMS to get validation benchmarks
- Cutflows and Monte Carlo information for given benchmarks

✅ **Discussions with CMS needed**

✦ Validation:

	Selection step	CMS	$\epsilon_i^{\rm CMS}$	MA5	$\epsilon_i^{\rm MA5}$	$\delta_i^{\rm rel}$
0	Nominal	84653.7		84653.7		
1	One hard jet	50817.2	0.6	53431.28	0.631	5.2%
2	At most two jets	36061	0.7096	38547.75	0.721	1.61%
3	Requirements if two jets	31878.1	0.884	34436.35	0.893	1.02%
4	Muon veto	31878.1	1	34436.35	1.000	0
5	Electron veto	31865.1	1	34436.35	1.000	0
6	Tau veto	31695.1	0.995	34397.54	0.998	0.3%
	$\not{E}_T > 250$ GeV	8687.22	0.274	7563.04	0.219	20.00%
	$\not{E}_T > 300$ GeV	5400.51	0.621	4477.67	0.592	4.66%
	$\not{E}_T > 350$ GeV	3394.09	0.628	2813.70	0.628	0.00%
	$\not{E}_T > 400$ GeV	2224.15	0.6553	1753.71	0.623	4.93%
	$\not{E}_T > 450$ GeV	1456.02	0.654	1110.92	0.633	3.21%
	$\not{E}_T > 500$ GeV	989.806	0.679	722.83	0.650	4.27%
	$\not{E}_T > 550$ GeV	671.442	0.678	487.54	0.674	0.59%

✅ **Validated at the 20% level**

Issue with the low-MET modelling in DELPHES

✦ The $t\bar{t}$+MET analysis (CMS-B2G-14-004) was validated to the 2-3% level

MADANALYSIS 5 analyses on INSPIRE

[BF, Martini ('16)]

- ✦ Implementation of LHC analyses can be uploaded on INSPIRE
  - ✤ DOI are assigned: can be cited, searched for, *etc*.

Information | Citations (1) | Files

Files are versioned, can be downloaded

MadAnalysis5 implementation of the CMS search for dark matter production with top quark pairs in the single lepton channel (CMS-B2G-14-004)

DOI and citations

Fuks, Benjamin; Martini, Antony

Description: This is the MadAnalysis5 implementation of the CMS search for dark matter in a channel where a pair of dark matter particles is produced in association with a top-antitop system. This search targets events featuring a single lepton originating from the top decays and a large amount of missing transverse energy.

Information how to use this code and a detailed validation summary are available at http://madanalysis.irmp.ucl.ac.be/wiki/PhysicsAnalysisDatabase. The CMS analysis is documented at https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsB2G14004.

Cite as: Fuks, B., Martiny, A. (2016). MadAnalysis5 implementation of the CMS search for dark matter production with top quark pairs in the single lepton channel (CMS-B2G-14-004). doi: 10.7484/INSPIREHEP.DATA.MIHA.JR4G

Automatic installation of all implemented analyses from MADANALYSIS 5

Record added 2016-05-09, last modified 2016-05-09

The Public Analysis Database of MADANALYSIS

[Dumont, BF, Kraml *et al.* (EPJC '15)]

- A database with MADANALYSIS 5 implementations of LHC analyses exists
  - http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase
- Snippet of the webpage

ATLAS analyses, 13 TeV

Analysis	Short Description	Implemented by	Code	Validation note	Version
ATLAS-EXOT-2015-03	monojet + missing transverse energy	D. Sengupta	Inspire	PDF	v1.3/Delphes3
ATLAS-SUSY-2015-06	jets + missing transverse momentum	S. Banerjee, B. Fuks, B. Zaldivar	Inspire	PDF	v1.3/Delphes3

Delphes card for ATLAS-EXOT-2015-03

Dedicated DELPHES cards

Code from INSPIRE

Validation information (cutflows, distributions, *etc.*)

- Can be automatically installed within MADANALYSIS 5

NLO effects on a CLs: top-philic dark matter

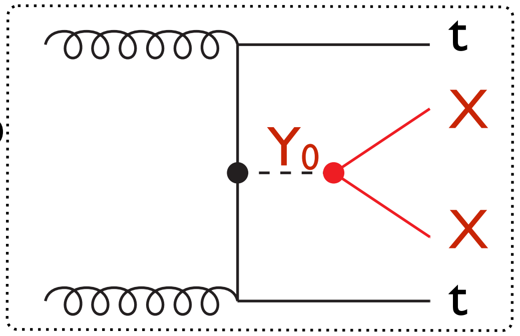
[Arina, Backovic, Conte, BF, Guo *et al.* (JHEP'16)]

✦ A simplified model for top-philic dark matter

[BF & Martini (2016)]

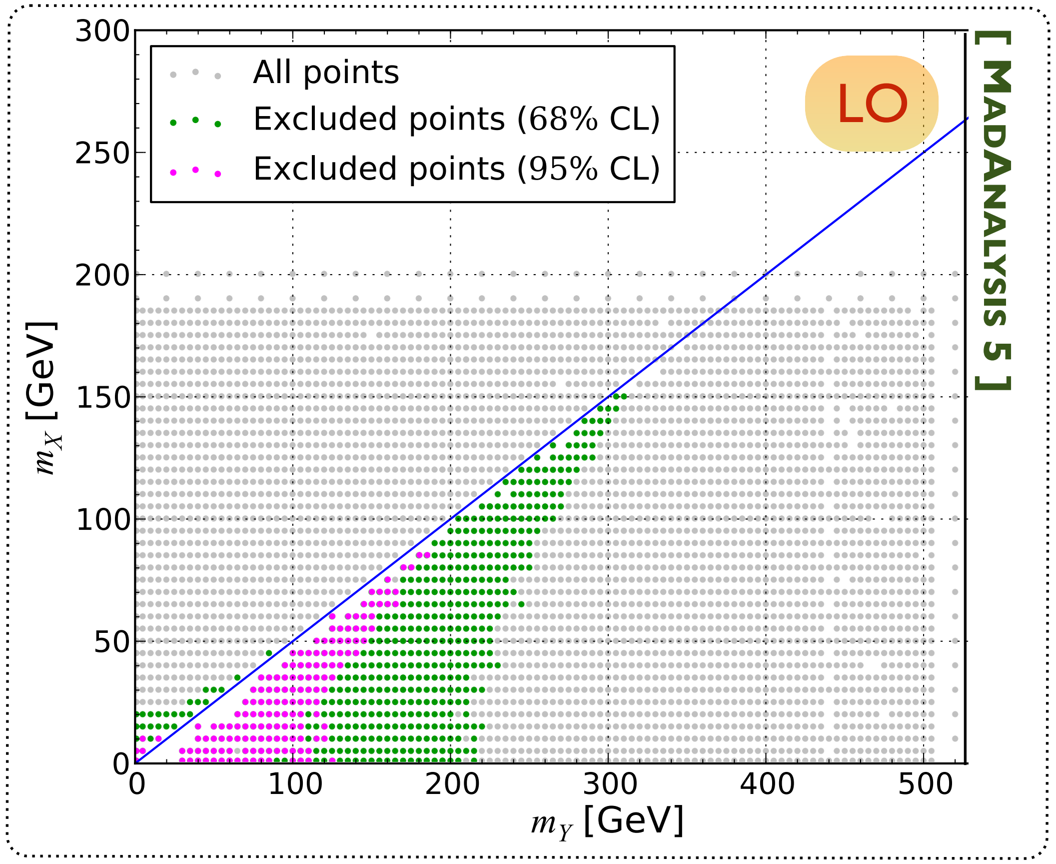
- ✤ A dark sector with a fermionic dark matter candidate X
- ✤ A (scalar) mediator Y_0 linking the dark sector and the top

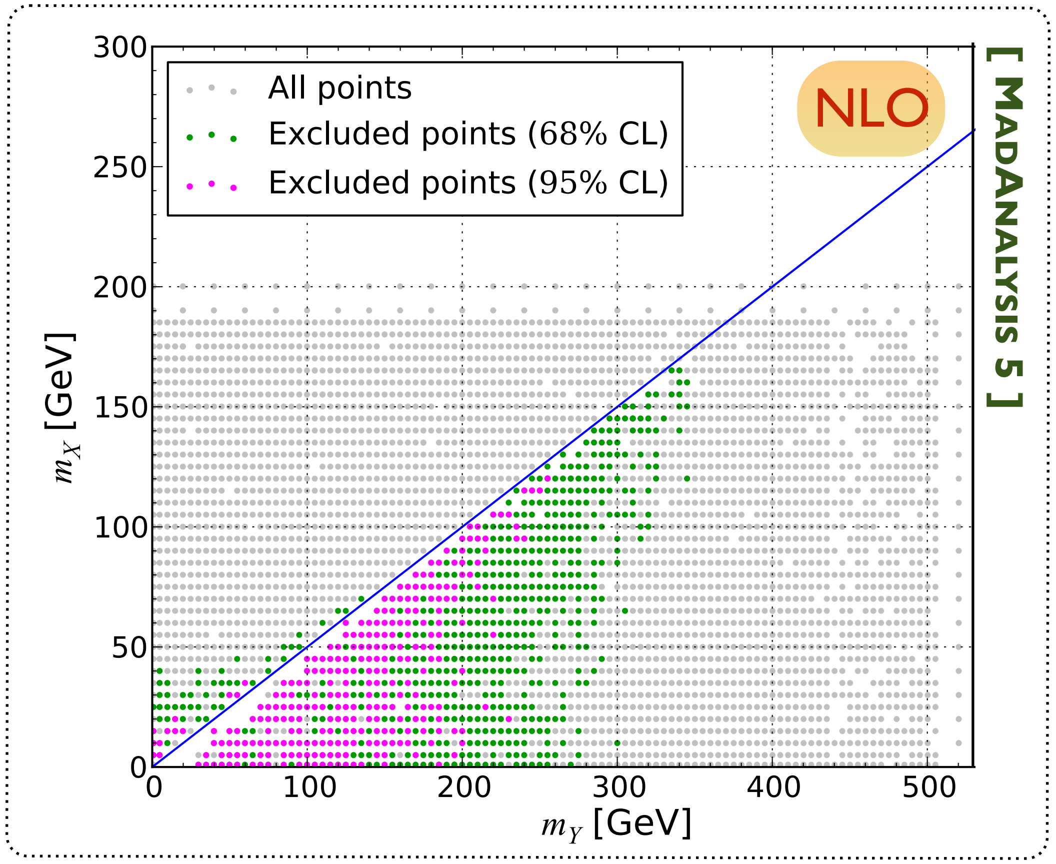
$$\mathcal{L}_{t,X}^{Y_0} = -\Big(g_t \frac{y_t}{\sqrt{2}} \bar{t}t + g_X \bar{X}X\Big) Y_0$$



- ✤ Could be probed with tt+MET events (CMS-B2G-14-004)

✦ For central scales: mild (but visible) NLO effects on the exclusions





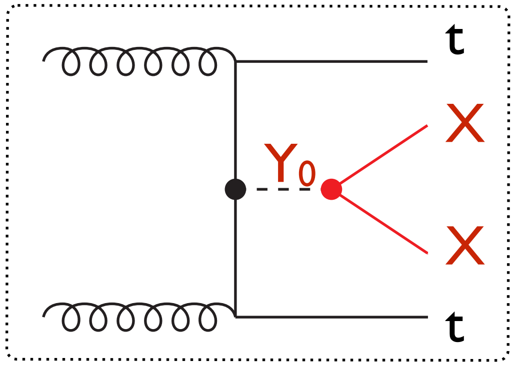
- ✤ How is the picture changing when including scale variations?

NLO effects on a CLs: top-philic dark matter

[Arina, Backovic, Conte, BF, Guo *et al.* (JHEP'16)]

✦ There are theoretical uncertainties on a CLs number

	$(m_Y,\, m_X)$	$\sigma_{\rm LO}$ [pb]	$\text{CL}_{\rm LO}$ [%]	$\sigma_{\rm NLO}$ [pb]	$\text{CL}_{\rm NLO}$ [%]
I	(150, 25) GeV	$0.658^{+34.9\%}_{-24.0\%}$	$98.7^{+0.8\%}_{-13.0\%}$	$0.773^{+6.1\%}_{-10.1\%}$	$95.0^{+2.7\%}_{-0.4\%}$
II	(40, 30) GeV	$0.776^{+34.2\%}_{-24.1\%}$	$74.7^{+19.7\%}_{-17.7\%}$	$0.926^{+5.7\%}_{-10.4\%}$	$84.2^{+0.4\%}_{-14.4\%}$
III	(240, 100) GeV	$0.187^{+37.1\%}_{-24.4\%}$	$91.6^{+6.4\%}_{-18.1\%}$	$0.216^{+6.7\%}_{-11.4\%}$	$86.5^{+8.6\%}_{-5.5\%}$



- ✤ An excluded point (95% CL) may not be excluded when accounting for errors
- ✤ The CLs number can increase / decrease at NLO
- ✤ The error band is reduced

Outline

Summary

✦ MadAnalysis 5:

- ✤ A framework for collider phenomenology (parton, hadron & reco level)
- ✤ User-friendly by means of its Python interface
- ✤ Flexible thanks to its C++ kernel
- ✤ Interfaced to several HEP packages

http://launchpad.net/madanalysis5

✦ Two modes of the code

- ✤ Pythonic: intuitive commands typed in a Python interface
  - ★ Analyses performed behind the scenes
  - ★ Human readable reports as output
- ✤ C++: programming in the SampleAnalyzer framework (the MadAnalysis 5 core)

✦ The LHC legacy

http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase

- ✤ Crucial to be able to reinterpret the LHC results in any theoretical context
- ✤ MadAnalysis 5 has been actively developed along these lines
  - ★ User-friendly way to confront any MC-simulated BSM signal to LHC results
- ✤ **Reproducibility** is the ability of an entire experiment to be reproduced (possibly by an independent theoretical study)