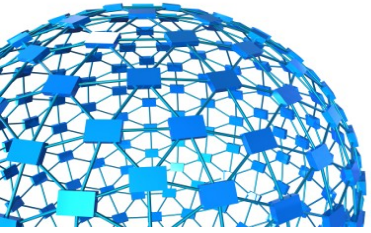
A toolkit for HepSim event samples with Monte Carlo predictions

S. Chekanov (ANL)

(thanks to Ed May for validation work)



Public MC samples: HepSim

HepSim: http://atlaswww.hep.anl.gov/hepsim/

- **Current features**:
  - Stores Monte Carlo (LO+PS) and NLO using unified file format
  - High data compression (~ 40% smaller than ROOT and faster to process)
  - Data "streaming" over a network
  - High-availability of data for public access (http)
  - Data analysis tools that (typically) do not require installation
  - Support for all major platforms (Windows, Linux, Mac, etc.)

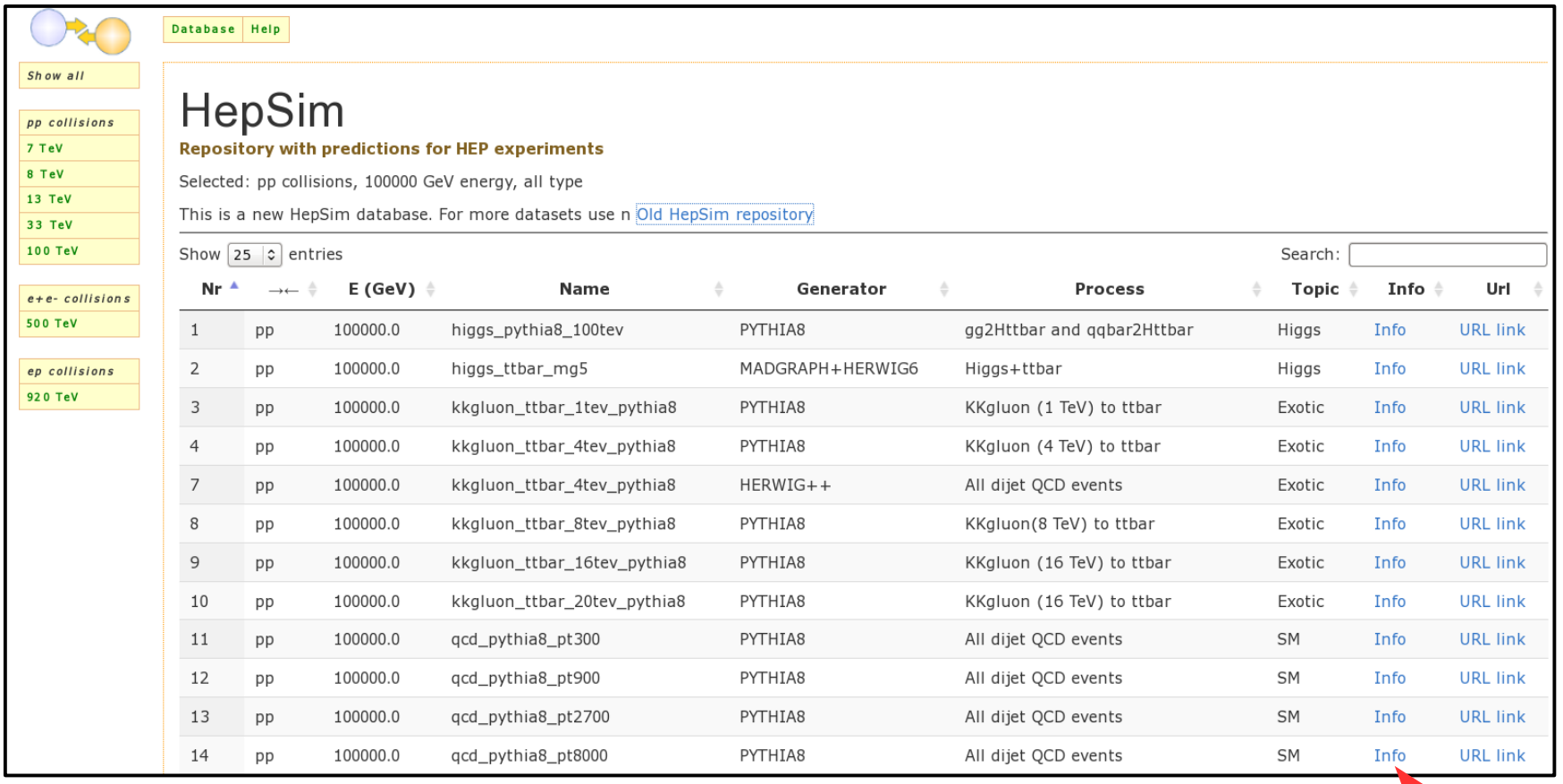
- **The HepSim can be used as a source for fast & full detector simulation, calculations of cross sections, kinematic distributions etc..**

HepSim structure

- **(1) Web front-end based on a SQL database with user login**
  - Describes samples
  - Provide macro to illustrate analysis / data access
- **(2) Data storage back-end.**
  - Can be distributed on multiple servers & continents & clouds
- **(3) Platform-neutral software toolkit that allows:**
  - list available files
  - download files in multiple threads
  - validate and view files and separate events
  - GUI for event scan
  - analyzing data using downloaded files (or) streaming data from a server

HepSim front-end: http://atlaswww.hep.anl.gov/hepsim/

Current statistics: ~150 samples, 21309 files, 213090000 events (approx.), 635G used storage



Database | Help

Show all

pp collisions
7 TeV
8 TeV
13 TeV
33 TeV
100 TeV

e+e- collisions
500 TeV

ep collisions
920 TeV

HepSim

Repository with predictions for HEP experiments

Selected: pp collisions, 100000 GeV energy, all type

This is a new HepSim database. For more datasets use n Old HepSim repository

Show 25 entries Search:

Nr	→←	E (GeV)	Name	Generator	Process	Topic	Info	Url
1	pp	100000.0	higgs_pythia8_100tev	PYTHIA8	gg2Httbar and qqbar2Httbar	Higgs	Info	URL link
2	pp	100000.0	higgs_ttbar_mg5	MADGRAPH+HERWIG6	Higgs+ttbar	Higgs	Info	URL link
3	pp	100000.0	kkgluon_ttbar_1tev_pythia8	PYTHIA8	KKgluon (1 TeV) to ttbar	Exotic	Info	URL link
4	pp	100000.0	kkgluon_ttbar_4tev_pythia8	PYTHIA8	KKgluon (4 TeV) to ttbar	Exotic	Info	URL link
7	pp	100000.0	kkgluon_ttbar_4tev_pythia8	HERWIG++	All dijet QCD events	Exotic	Info	URL link
8	pp	100000.0	kkgluon_ttbar_8tev_pythia8	PYTHIA8	KKgluon(8 TeV) to ttbar	Exotic	Info	URL link
9	pp	100000.0	kkgluon_ttbar_16tev_pythia8	PYTHIA8	KKgluon (16 TeV) to ttbar	Exotic	Info	URL link
10	pp	100000.0	kkgluon_ttbar_20tev_pythia8	PYTHIA8	KKgluon (16 TeV) to ttbar	Exotic	Info	URL link
11	pp	100000.0	qcd_pythia8_pt300	PYTHIA8	All dijet QCD events	SM	Info	URL link
12	pp	100000.0	qcd_pythia8_pt900	PYTHIA8	All dijet QCD events	SM	Info	URL link
13	pp	100000.0	qcd_pythia8_pt2700	PYTHIA8	All dijet QCD events	SM	Info	URL link
14	pp	100000.0	qcd_pythia8_pt8000	PYTHIA8	All dijet QCD events	SM	Info	URL link

Few samples: dijet QCD, ttbar at high pT have 1 ab-1 lumi for a 100 TeV collider!
A few samples are used for ATLAS analysis (exc. Higgs, direct photons)
Many samples are done on BlueGene/Q

HepSim front-end: http://atlaswww.hep.anl.gov/hepsim/

Information about *"higgs_gamgam_mcfm"* dataset

Name:	*higgs_gamgam_mcfm*
Collisions:	pp
CM Energy:	100000.0 GeV
Entry ID:	51
Topic:	Higgs
Generator:	MCFM
Calculation level:	NLO
Process:	Higgs-> +g+g (NLO)
Total events:	26000000
Number of files:	512
Cross section (σ):	632.4 ± 7.0 pb
Luminosity (L):	41,113.2195 pb-1 *(or)* 41.1132 fb-1 *(or)* 0.0411 ab-1
Format:	ProMC
Submission date:	Fri Jun 6 21:35:06 CDT 2014
Download URL:	http://mc.hep.anl.gov/asc/hepsim/events/pp/100tev/higgs_gamgam_mcfm
Dataset size:	2.07 GB
Description:	Higgs to gamma+gamm at NLO QCD. CT10 NLO +52 PDF error (included). MCFM 6.7
Dataset files:	View files

Analysis scripts:

Nr	Macro	Image	Data
1	higgs_gamgam_mcfm.py	higgs_gamgam_mcfm.png	higgs_gamgam_mcfm.jdat

Author: S.Chekanov

What can be done with it?

- Download files in multiple threads
- View event metadata, look at separate events
- Run analysis code using C++/ROOT, Python, Java, etc.
- Run analysis code without downloading files (“data streaming”)
  - similar to “video” streaming: user's front-end does calculations
- Data can be analyzed in a GUI and a “batch” mode.
- Data can be converted to ROOT and other formats (HEPMC, etc)
- Data can be processed with fast / full simulations

Installing HepSim toolkit:

```
wget http://atlaswww.hep.anl.gov/asc/hepsim/hs-toolkit.tgz -O - | tar -xz
source hs-toolkit/setup.sh
```

(or)

```
curl  http://atlaswww.hep.anl.gov/asc/hepsim/hs-toolkit.tgz  | tar -xz
source hs-toolkit/setup.sh
```

→ Java 7 or 8 should be installed. No other requirements.
→ Should work on Linux, Mac, Windows etc.

List and download Monte Carlo samples:

Show the number of files in the Madgraph 5 (Higgs+ttbar) sample:

```
hs-ls http://mc.hep.anl.gov/asc/hepsim/events/pp/100tev/higgs_ttbar_mg5
```

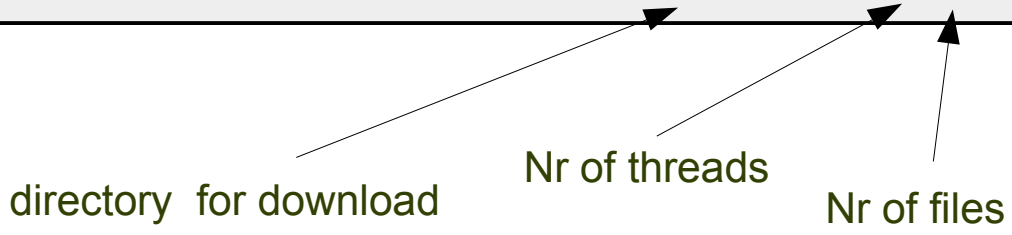
```
mg5_Httbar_100tev_001.promc          24464
.....
.....

mg5_Httbar_100tev_020.promc          24420
-> Summary: Nr of files=20  Total size=489 MB
```

Note: Currently, HepSim uses an explicit URL location. In future it will use sample name and a redirection service

Download files using multiple threads:

```
hs-get http://mc.hep.anl.gov/asc/hepsim/events/pp/100tev/higgs_ttbar_mg5 higgs_ttbar_mg5 2 5
```



File validation

Show metadata and perform a few sanity checks (works faster using downloaded file!)

```
hs-info http://mc.hep.anl.gov/asc/hepsim/events/pp/14tev/higgs/pythia8/pythia8_higgs_1.promc
```

```
ProMC version = 2
Last modified = 2013-06-05 16:32:18
Description   = PYTHIA8;PhaseSpace:mHatMin = 20;PhaseSpace:pTHatMin = 20;
             ParticleDecays:limitTau0 = on;
             ParticleDecays:tau0Max = 10;HiggsSM:all = on;
Events         = 10000
Sigma     (pb) = 2.72474E1 ± 1.92589E-1
Lumi   (pb-1) = 3.67007E2
Varint units  = E:100000 L:1000
Log file:     = logfile.txt
The file was validated. Exit.
```

Used for conversion to variable-number of bytes (varint)

Complete log file from generator is usually included

Print an arbitrary event using a random access feature:

```
hs-info http://mc.hep.anl.gov/asc/hepsim/events/pp/14tev/higgs/pythia8/pythia8_higgs_1.promc 100
```

Event number

Viewing Monte Carlo events / logfiles

One can open and view large files and separate events using random access feature:

```
hs-view [local file or URL]
```

Still not sure what events are inside? Extract a complete Monte-Carlo specific logfile:

```
unzip -p [promc file]  logfile.txt
```

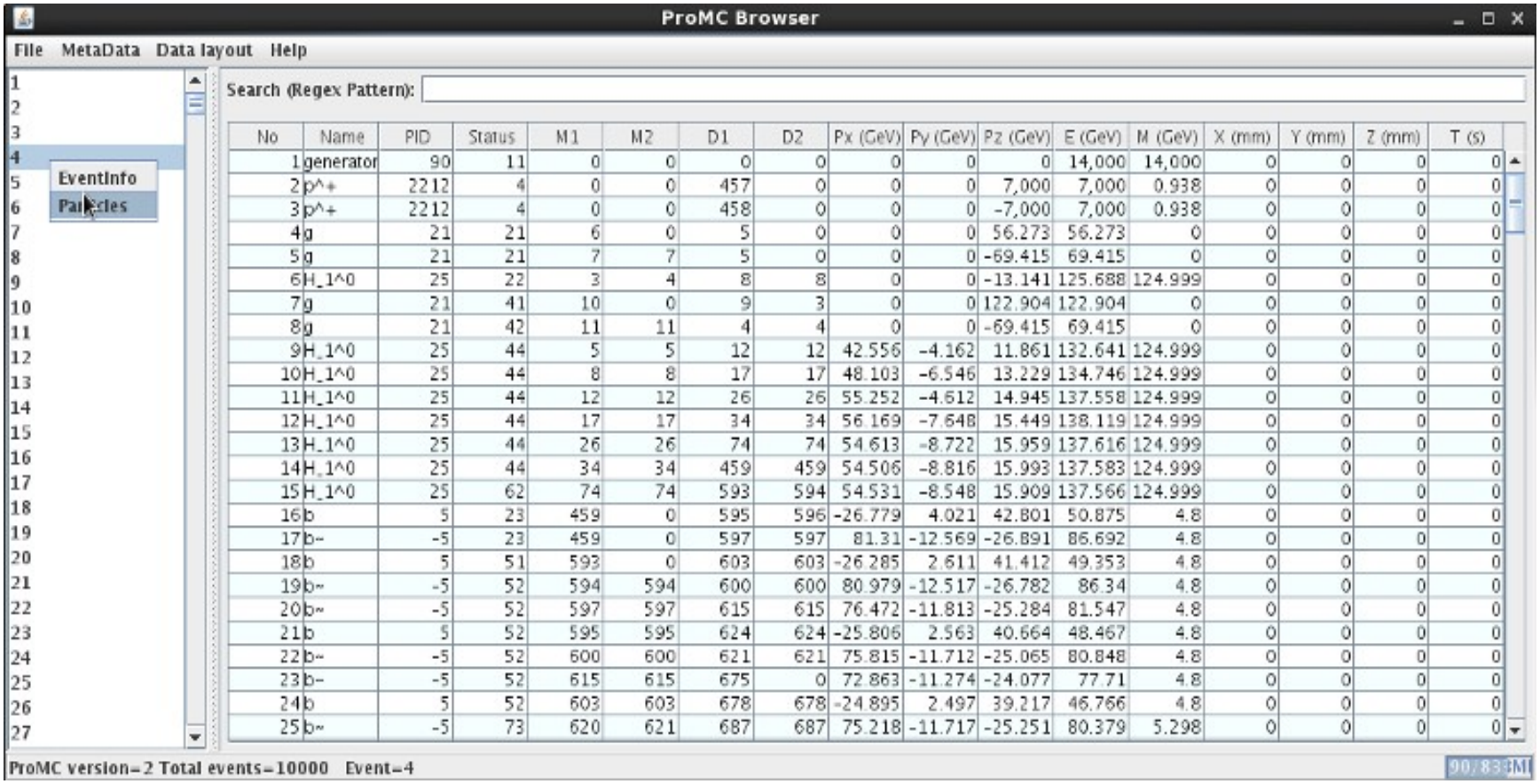
Or look at all events using a GUI mode (tested on >4 GB files)

```
hs-view http://mc.hep.anl.gov/asc/hepsim/events/pp/14tev/qcd/pythia8/pythia8_qcd_1.promc
```

This works faster after downloading the file on local disk:

```
hs-view pythia8_qcd_1.promc
```


Output of "hs-view" (based on http://atlaswww.hep.anl.gov/asc/promc/)



No	Name	PID	Status	M1	M2	D1	D2	Px (GeV)	Py (GeV)	Pz (GeV)	E (GeV)	M (GeV)	X (mm)	Y (mm)	Z (mm)	T (s)
1	generator	90	11	0	0	0	0	0	0	0	14,000	14,000	0	0	0	0
2	p^+	2212	4	0	0	457	0	0	0	7,000	7,000	0.938	0	0	0	0
3	p^+	2212	4	0	0	458	0	0	0	-7,000	7,000	0.938	0	0	0	0
4	g	21	21	6	0	5	0	0	0	56.273	56.273	0	0	0	0	0
5	g	21	21	7	7	5	0	0	0	-69.415	69.415	0	0	0	0	0
6	H_1^0	25	22	3	4	8	8	0	0	-13.141	125.688	124.999	0	0	0	0
7	g	21	41	10	0	9	3	0	0	122.904	122.904	0	0	0	0	0
8	g	21	42	11	11	4	4	0	0	-69.415	69.415	0	0	0	0	0
9	H_1^0	25	44	5	5	12	12	42.556	-4.162	11.861	132.641	124.999	0	0	0	0
10	H_1^0	25	44	8	8	17	17	48.103	-6.546	13.229	134.746	124.999	0	0	0	0
11	H_1^0	25	44	12	12	26	26	55.252	-4.612	14.945	137.558	124.999	0	0	0	0
12	H_1^0	25	44	17	17	34	34	56.169	-7.648	15.449	138.119	124.999	0	0	0	0
13	H_1^0	25	44	26	26	74	74	54.613	-8.722	15.959	137.616	124.999	0	0	0	0
14	H_1^0	25	44	34	34	459	459	54.506	-8.816	15.993	137.583	124.999	0	0	0	0
15	H_1^0	25	62	74	74	593	594	54.531	-8.548	15.909	137.566	124.999	0	0	0	0
16	b	5	23	459	0	595	596	-26.779	4.021	42.801	50.875	4.8	0	0	0	0
17	b~	-5	23	459	0	597	597	81.31	-12.569	-26.891	86.692	4.8	0	0	0	0
18	b	5	51	593	0	603	603	-26.285	2.611	41.412	49.353	4.8	0	0	0	0
19	b~	-5	52	594	594	600	600	80.979	-12.517	-26.782	86.34	4.8	0	0	0	0
20	b~	-5	52	597	597	615	615	76.472	-11.813	-25.284	81.547	4.8	0	0	0	0
21	b	5	52	595	595	624	624	-25.806	2.563	40.664	48.467	4.8	0	0	0	0
22	b~	-5	52	600	600	621	621	75.815	-11.712	-25.065	80.848	4.8	0	0	0	0
23	b~	-5	52	615	615	675	0	72.863	-11.274	-24.077	77.71	4.8	0	0	0	0
24	b	5	52	603	603	678	678	-24.895	2.497	39.217	46.766	4.8	0	0	0	0
25	b~	-5	73	620	621	687	687	75.218	-11.717	-25.251	80.379	5.298	0	0	0	0

Note: particle names are shown using a look-up table (not stored!)
Also one can access event information and a complete log file.

Fast simulation using Delphes & Conversions

Delphes (https://cp3.irmp.ucl.ac.be/projects/delphes) should be installed with DelphesProMC

```
DelphesProMC delphes.tcl output.root higgs_ttbar_mg5.promc
```

Stream 5 generator-level files, run a fast simulation and put output ROOT files inside local "hepsim-output" directory:

```
hs-exec DelphesProMC delphes.tcl output.root \
http://mc.hep.anl.gov/asc/hepsim/events/pp/100tev/higgs_ttbar_mg5 5
```

Conversion to ROOT (without fast simulation). ProMC package should be installed

```
promc2root higgs_ttbar_mg5.promc higgs_ttbar_mg5.root
```

(conversion time < typical download time)

```
promc2hepmc higgs_ttbar_mg5.promc higgs_ttbar_mg5.hepmc
```

HEPMC files have sizes 8-10 larger than the input!

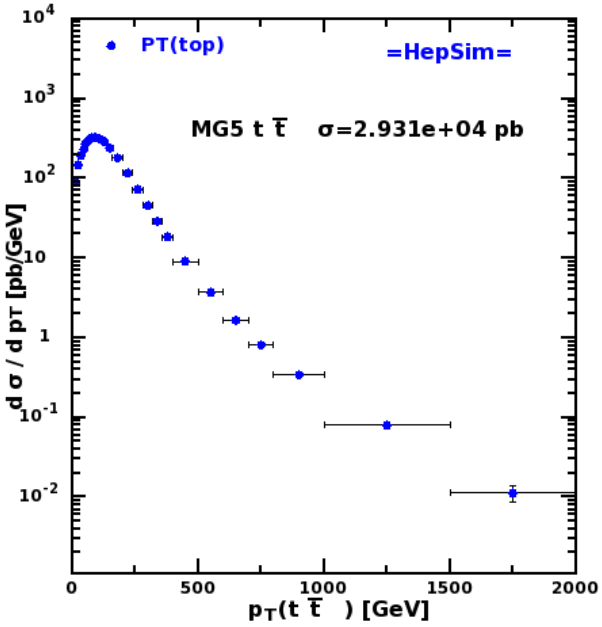
Processing events using Python

```
wget http://mc.hep.anl.gov/asc/hepsim/events/pp/100tev/ttbar_mg5/macros/ttbar_mg5.py
wget -O scavis.zip http://sourceforge.net/projects/scavis/files/latest/download
unzip scavis.zip
./scavis/scavis_batch.sh ttbar_mg5.py \
http://mc.hep.anl.gov/asc/hepsim/events/pp/100tev/ttbar_mg5/ 10000
```

Calculate differential ttbar cross section using Madgraph5 (NLO)

- get a Python analysis script
- get SCaVis from sourceforge (unzip it)
- run a batch script that streams 10000 events:

You may first download data files to make the program faster.



Offline analysis code (requires ProMC C++ package)

- ProMC files are “self-describing”. Language-neutral data layouts are kept inside the files. One can generate C++, Java and Python analysis code for reading and writing data from existing ProMC file.

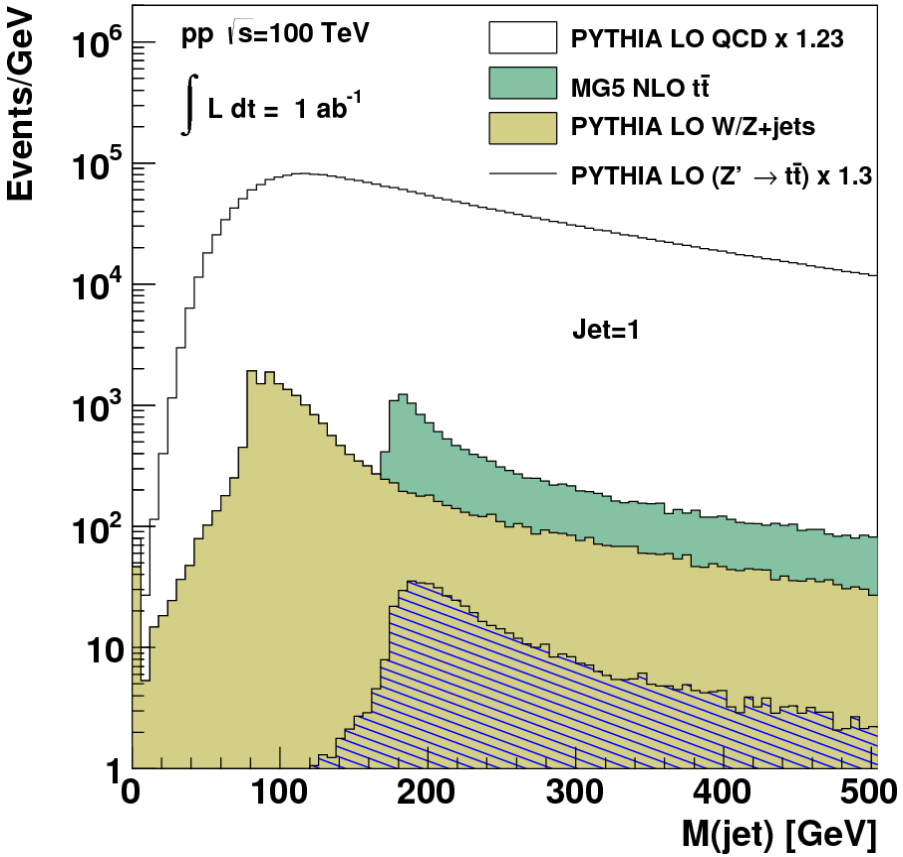
- Example: Create analysis code from a downloaded file
  - **Step 1:** Download a ProMC file from http://atlaswww.hep.anl.gov/hepsim/
  - **Step 2:** Generate analysis code from its structure:

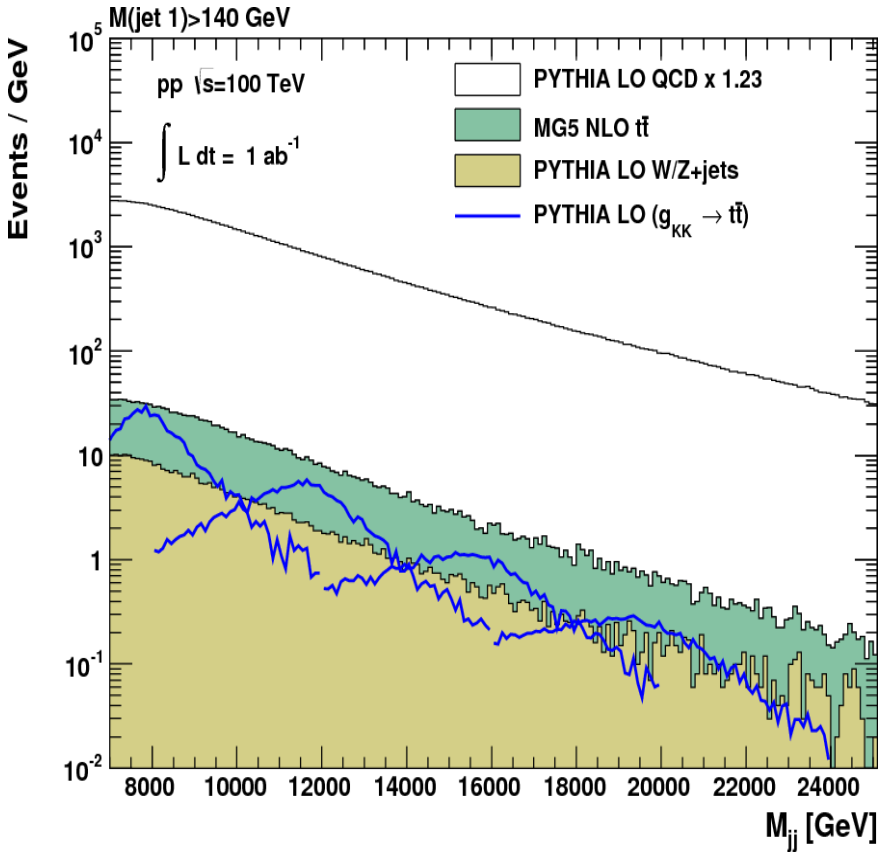
```
promc_info   file.promc    # check information about this file
promc_proto file.promc     # extracts data layouts into the directory "proto"
promc_code                 # generate C++, Java and Python code in src/, java/, python/
make                       # compiles C++ code “reader.cc”
./reader file.promc        # runs over all truth events
```

- **Step 3:** Use ROOT to fill histograms etc.
- **Step 4:** Try Python (slower!) and Java code (faster!) from python/ and java/ directories

Read more: https://atlaswww.hep.anl.gov/asc/wikidoc/doku.php?id=asc:promc:examples

Example: High-pT dijets using 1 ab^{-1} (!)





Need to process with a fast simulation !