Monte Carlo generators for CMS

Roberto Chierici – CNRS IPN Lyon
Fabio Cossutti – INFN Trieste
Fabian Stöckli – CERN
On behalf of the CMS Collaboration

Generators and experimental setup

✓ **Physics and software requirements**
✓ **Main choices in CMS**

Introduction

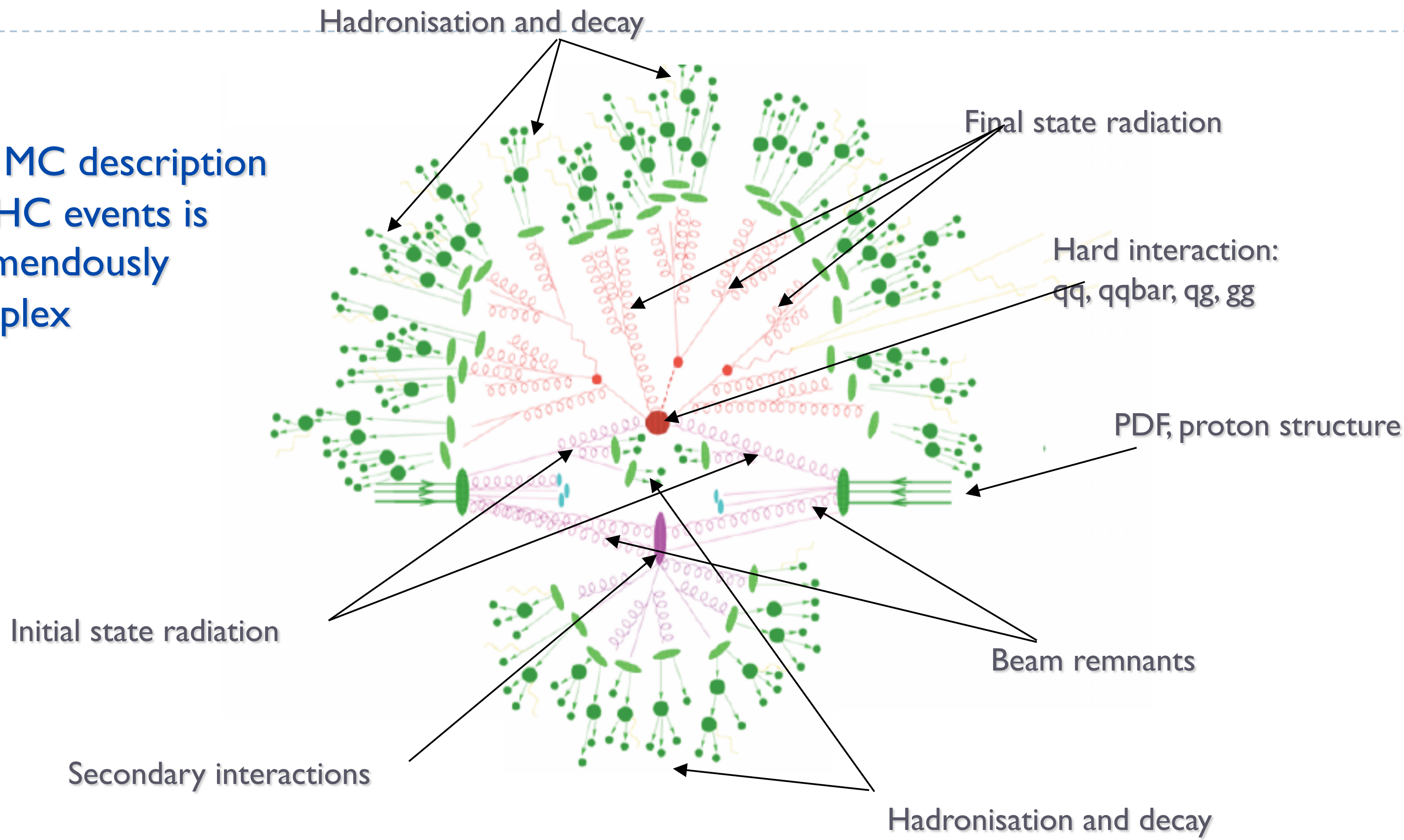
- With the beginning of LHC data taking, we move from physics potential studies and computing challenges to the real work of comparing with data
- A long preparatory work to define a set of generators to be used has been done, but the real data analysis will drive next steps
  - And the possible reaction of the theory community to it
- MC generators are the interface between experiment and theory
  - Setting up generators in an experimental environment brings a number of physics and software problems

✓ build a complete and coherent set of generated samples for doing physics
✓ steer the generator tuning with data
✓ collect the physics requirements, communicate/collaborate with the theory
✓ plan the event generation from the different physics groups' requests
✓ interface, validate, maintain the generator in the experiment's software framework

- **In this talk we try to give an overview of the current CMS situation in this respect**

Physics requirements: basics

The MC description of LHC events is tremendously complex



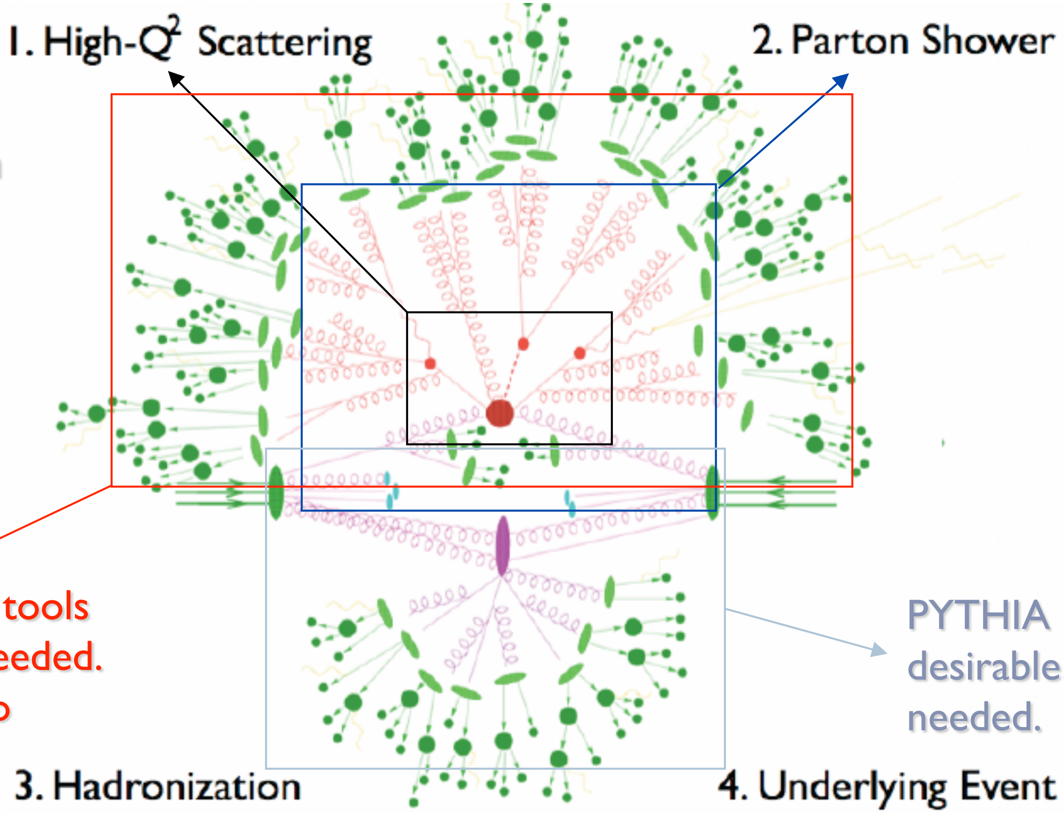
This is a schematization to be able to cut down the problem in pieces and model them in a different way. The "pieces" are correlated !

Physics requirements: needs

Extra partons emission described with ME at the highest possible order (+matching). Spin correlations needed.

Interface to dedicated tools (PYTHIA/HERWIG) is needed. Tuning with data is also needed.

The MC description of LHC events is tremendously complex



Interface to dedicated tools (PYTHIA/HERWIG) needed. Tuning with data is also needed.

PYTHIA MPI. HERWIG/JIMMY desirable. Tuning with data is needed.

Other desirable features, from the experimentalist's viewpoint:

- ✓ output in the Les Houches standard format
- ✓ as much complete as possible coverage of SM phase space
- ✓ user friendly inclusion of new physics signals
- ✓ support ☺

ME vs PS: criteria for choices

- No generator adequately satisfy all the physics needs for the CMS program
- Essential to understand what is best for each kinematic regime
  - Parton Shower: infinite series in α_S keeping only singular terms (collinear approx.)
    - Excellent at low p_T, with emission at any order, simple interface with hadronization
    - Large uncertainties away from singular regions
    - To be used for soft (compared to signal scale) jets
  - Fixed order matrix elements: truncated expansion in α_S
    - Full helicity structure to the given order
    - To be used for hard (compared to signal scale) jets
- High jet multiplicity events are bound to be better described by ME
  - They must be an essential part of the plan

ME and higher orders

- NLO generators with PS now exist: predict correct normalization and shape at NLO
  - Techniques for correcting the first emission in PS to get the shapes of the first additional hard jet in the event correct also exist
- ME/PS matching fundamental for multi parton-LO ME generators
  - Essential a procedure to avoid double counting
    - ME_N+PS has parts of ME_{N+1}+PS
  - In practice we have recipes, independent on the actual generator
    - MLM and CKKW schemes
- The PS tuning should be working properly according to its actual usage
  - E.g. PS only or ME+PS, and which ME
  - Important to verify and possibly re-tune

Importing a generator in CMS

- In order to integrate a generator beyond standalone studies:
  - E.g. integrate in the official framework for massive production of fully simulated and reconstructed events

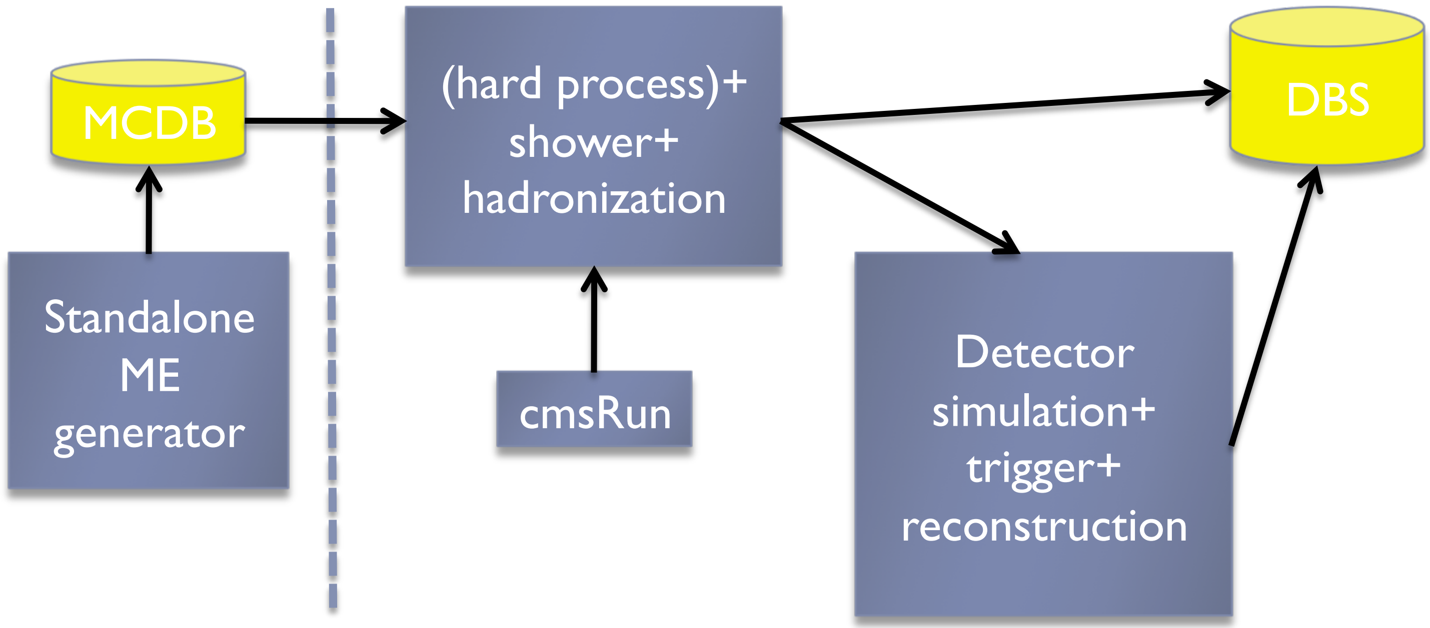
- A motivated request must be made by a physics group
  - Not single user initiatives
- There must be a responsible for an interface
  - A technical validation of it is customary before the import become official
  - Show physics results make sense and are consistent with the standalone expected behavior

Software requirements: the CMS environment

- CMS software distribution: all the components (CMS code + external tools as generators) are recompiled
  - gcc 4.3.4 / SL5 current choice
    - other compilers not an option for official productions, compiler itself part of the distribution
  - FORTRAN handled with gfortran
    - Core SW experts would love to get rid of it, but if a serious physics case for a FORTRAN code is present we use it
- Building an rpm for each external component from source
  - For generators mostly from GENSER repository

Software requirements: the CMS environment

- Going from the theoreticians' code to a massive production in CMS requires a number of steps
- 2 basic ways of integrating it: partonic event interface or plugin + library
  - Event stored as a wrapper around HepMC::GenEvent
  - Plus additional settings information



Generators integration in CMS(SW)

- Generator integration strategy depends on generator way of working
  - No process dependent pre-generation stage, only static code
    - library interfaced to CMSSW through a plugin module (with possible C++/FORTRAN interface)
    - Can be directly used in the official production flow as first step
  - Process dependent generation stage, like ME code generation and phase space integration
    - Code run standalone (either included in distribution or not)
    - results stored in LHE format in MCDB, source for the remaining standard production chain
- LHE format is the key for insertion of complex ME tool
  - Running them in the standard flow far from trivial

Generators currently used in CMS

- Choice is physics topic dependent
  - only tools used so far for massive official productions are mentioned
- Standard Model
  - Soft QCD: PYTHIA6, PYTHIA8, HERWIG6, HERWIG++, SHERPA
  - High p_T QCD: PYTHIA6, PYTHIA8, HERWIG6, HERWIG++, ALPGEN, MadGraph
  - W+jet/Z+jet: PYTHIA6, ALPGEN, MadGraph, MC@NLO, SHERPA, POWHEG
  - Top: ALPGEN, MadGraph, MC@NLO
  - Diffractive physics: POMWIG, Exhume
  - Higgs: PYTHIA6, MC@NLO
- Beyond the Standard Model
  - SUSY: PYTHIA, MadGraph
  - Exotica: PYTHIA, MadGraph
- Heavy Ions
  - Hydjet, PYQEN
- Plus all non collision phenomena
  - Cosmics, beam halo/gas, particle guns for test beams

Generators setup and validation

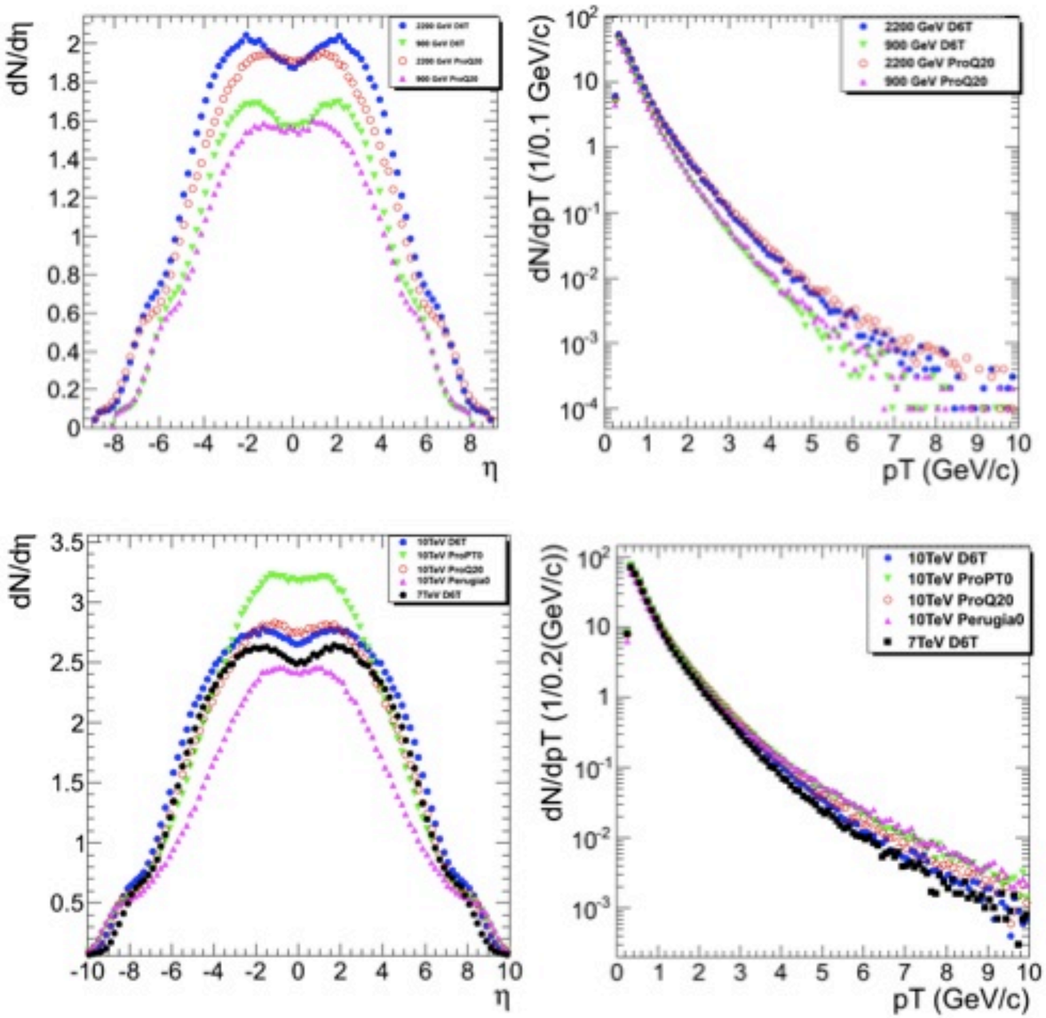
- ✓ **MB/UE issues and tuning**
- ✓ **Forward physics**
- ✓ **MPI issues**
- ✓ **ME issues and ME-PS matching**
- ✓ **Decay handling**

MB/UE tunes: current CMS setup

- Pythia6.4
  - CMS reference until now.
    - Tunes DW(T), D6(T) (R.Field, CDF, CMS UE team).
      - Virtuality ordered Showers, "old" MPI.
  - New CMS references.
    - Automated tunes considering LEP fragmentation data (MCnet/Professor).

      [Systematic event generator tuning for the LHC, **arXiv:0907.2973,** published in EPJC]
      - Considered with interest for the availability of both Virtuality- and pT- ordered Showers.
    - Perugia tunes (P. Skands) adopting also fragmentation results from Professor tunes.
      - pT ordered Showers, "new" MPI.
- Herwig6.5/Jimmy.
  - Tune A, B (A.Moraes, ATLAS).
  - CTEQ66 and MRST-LO* (MCnet/Professor).
- Pythia8.
  - Tunes from MCnet/Professor and P.Skands (available from Pythia 8.130).
- Lack of alternative tunes in other MCs on top of the "optimal" one(s) provided by the authors, but following several ongoing activities.
  - Herwig++, Sherpa.
- CMS would like to promote more activity on the Tune of ME+PS Hybrids.

PYTHIA6 tune studies activities



On the basis of studies of P. Skands and extensive validations, moving towards Perugia0 as new baseline choice

✓on top of PYTHIA 6.422

✓Moving also other generators to their most recent versions

✓No official feedback from first measurements on the tuning activities for the time being

Tuning strategy in CMS

- Effort leaded by P. Bartalini
- The first CMS Tunes will rely on the comparison of detector level results to Full-sim MCs.
  - Reduced botanic of tunes. Limited usage of automated tools. Manual Tuning "a la Rick Field".
- In the meantime CMS is preparing the ground for the maintenance of particle level analyses and results.
  - Evaluation of Rivet/Professor.
    - Rivet seems to be appropriate to document/host the CMS generator level analyses.
    - Internal (CMS) instance would be propagated to the official Rivet branch as soon as the analysis becomes public.
  - Integration of Rivet/Professor to CMMSW through the UNIX fifo
    - Tests ongoing.
  - Work under discussion, ideally to produce results later this year
  - Further functionalities (related to large scale computing) may be added at a later stage.

MPI and PYTHIA8

✓Compare PYTHIA8 with/without MPI with PYTHIA6, HERWIG, ALPGEN for γjjj

✓Simple selection: γ from MC, jets with midPointCone, $p_T(\gamma,\text{jets})>30$ GeV, $\Delta R_{ij}>0.8$, choose pairing on the basis of the minimum sum of relative p_Ts

✓PYTHIA 6.413, DWT or S0 tuning (more colour reconnection)

✓PYTHIA 8.1 default (tuning~S0), with and without MPI

✓HERWIG 6.510, JIMMY 4.2 (and no MPI)

✓ALPGEN γ+3j inclusive, PS with PYTHIA 6.409, DWT tuning

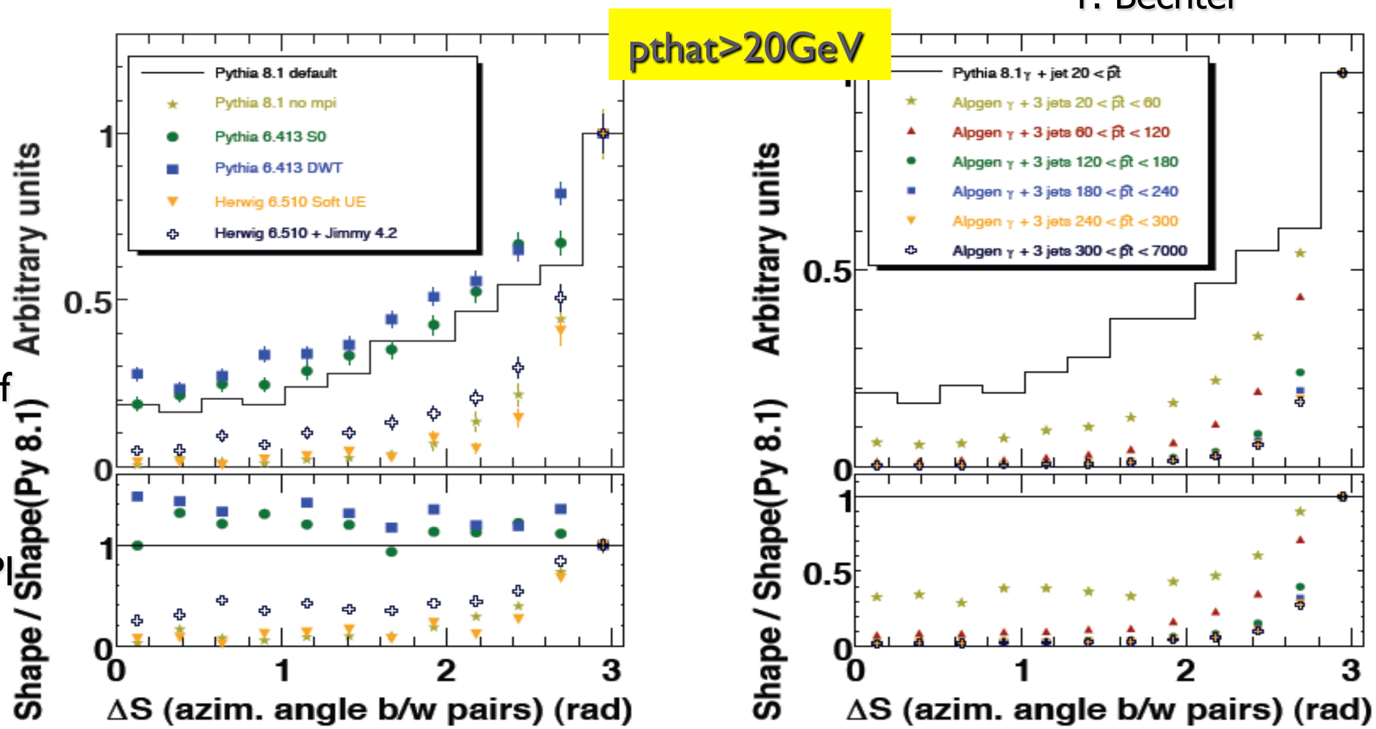
F. Bechtel

✓Large difference in HERWIG and PYTHIA

✓MPI-no MPI give a sizable difference

✓A better ME description of the process has no effect

✓The S0, DWT tunings similar to PYTHIA8 with MPI

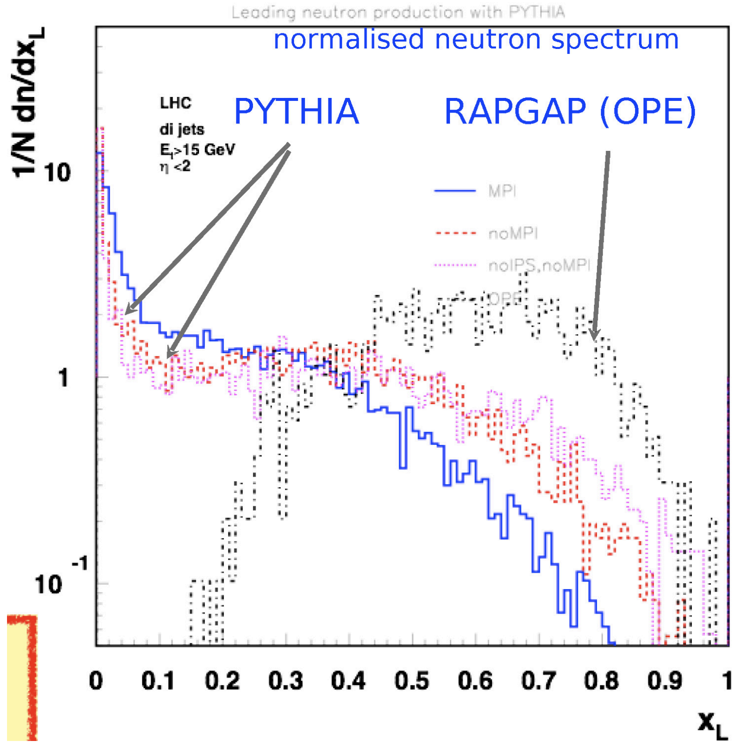


Forward physics

- Large variety of dedicated codes used, on top of basic ones
- For instance:
  - Forward jets and energy flow, e.g. $3 < |\eta| < 6.5$
    - PYTHIA6, PYTHIA8, ARIADNE, CASCADE
  - Ratio of single diffraction (also in association to central jets)
    - PYTHIA6, PYTHIA8, PHOJET, POMWIG, RAPGAP (diff)
  - Leading neutron in minimum bias and double jets
    - PYTHIA6, RAPGAP (OPE)
  - Also EXHUME, HARDCOL

✓Pion PDF from πN data

look at jet production with
$E_t > 15\ \mathrm{GeV}, |\eta| < 2$



Forward jets

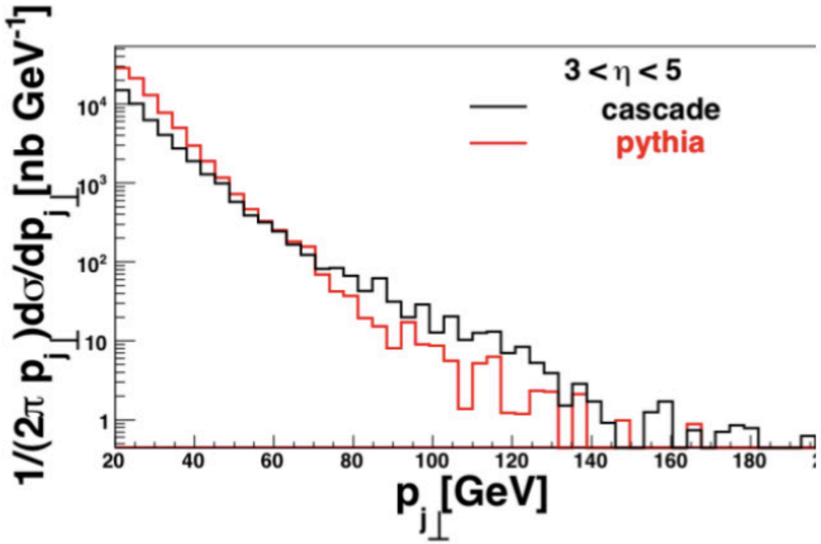
✓Small x effect treatment needed, CASCADE studies used to complement PYTHIA/HERWIG

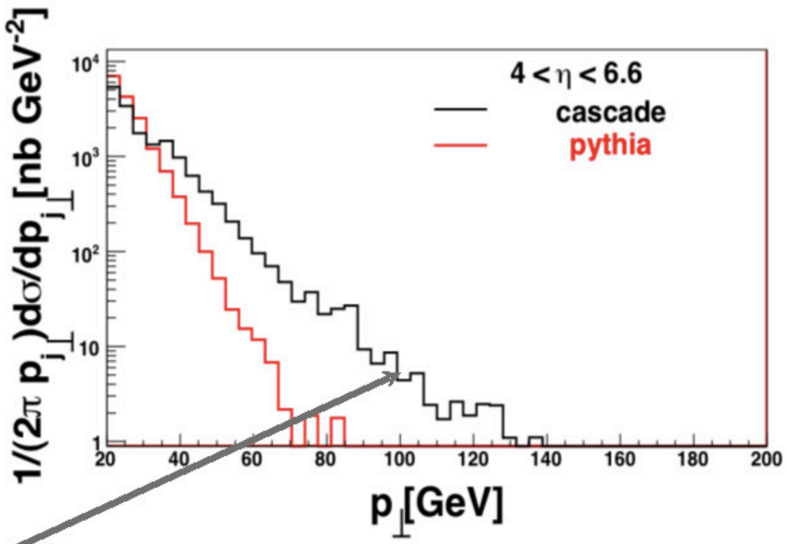
- Estimates for LHC

Estimates by M. Deak, K. Kutak
Calcualtions: **Forward Jet Production at the Large Hadron Collider.**
M. Deak, F. Hautmann, H. Jung, K. Kutak JHEP 0909:121,2009.

Central jet $|y_j| < 2$ $p_{j\perp} > 20\ GeV$ — Forward jet $3 < y_j < 5$ $p_{j\perp} > 20\ GeV$

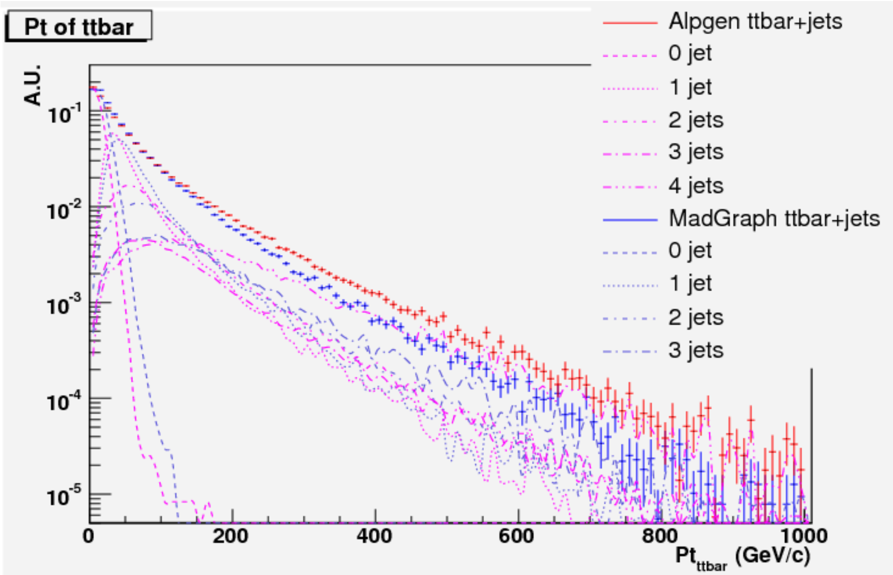
Central jet $|y_j| < 2$ $p_{j\perp} > 20\ GeV$ — Forward jet $4.4 < y_j < 6$ $p_{j\perp} > 20\ GeV$





Significantly larger pt tail at large pt, signal for new parton dynamic effects

ALPGEN vs MadGraph with matching



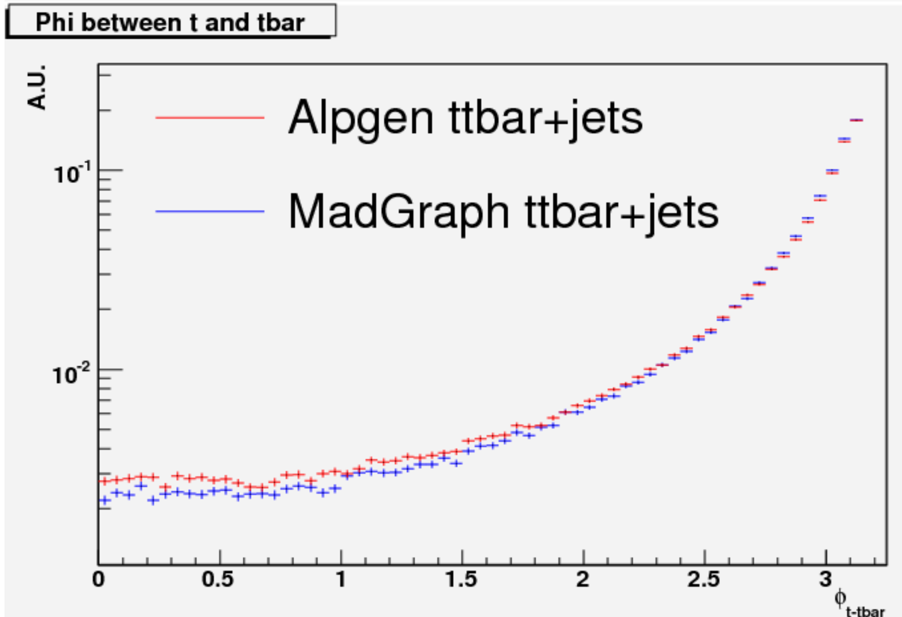
ALPGEN and MadGraph differ by at most 50% on the p_T prediction

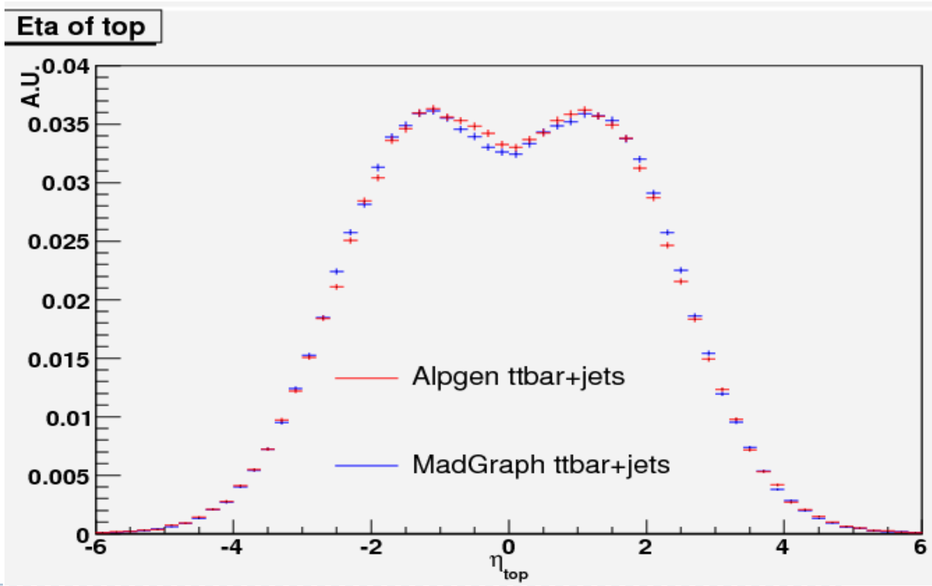
Important to understand the residual theory error on the distributions:

✓Effect of renormalisation and factorisation scales on the predictions

✓Effect of the chosen ME-PS matching scale

Excellent agreement on other variables



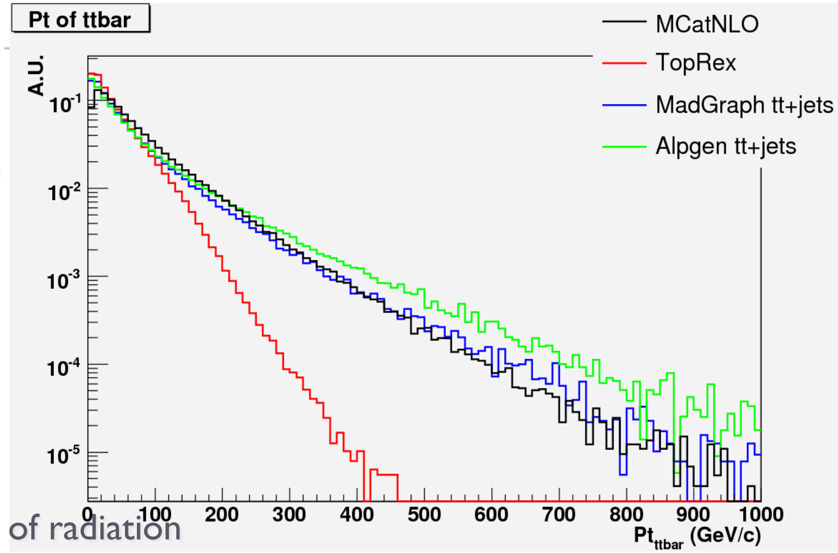


Comparing with MC@NLO

Comparisons to MC@NLO ongoing in CMS

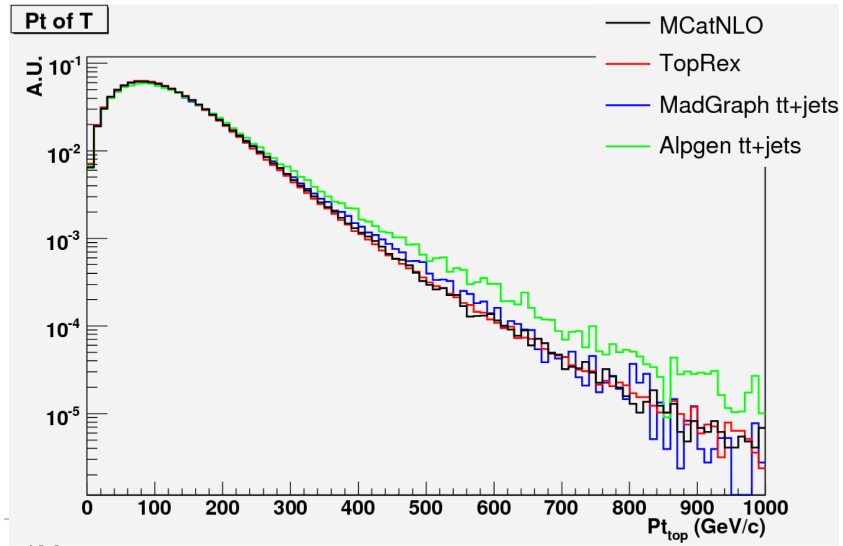
Care in interpreting the results

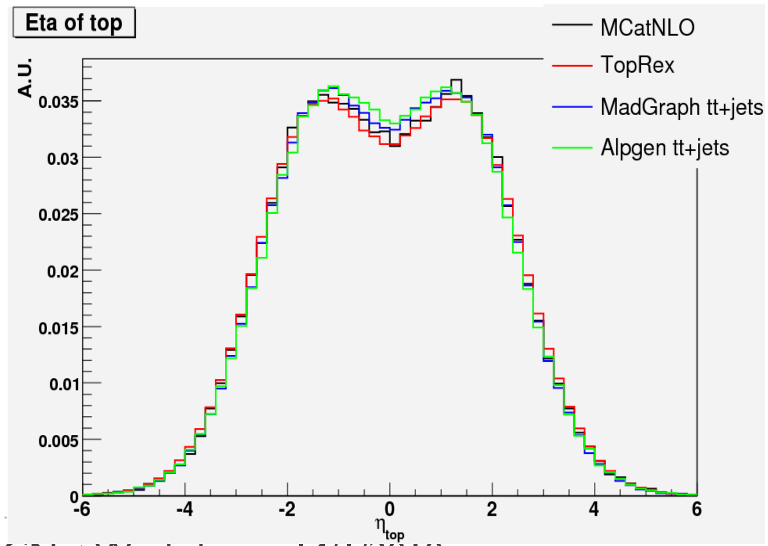
- ✓ Non perturbative part treated by HERWIG/JIMMY.
- ✓ Should compare to a matched tt0j(exc)+tt1j(inc) production



Still a very important step in understanding high p_T radiation and increase our confidence in the process description. Also gives indications on

- ✓ Relative importance of first emission
- ✓ Normalization
- ✓ Indication of systematic errors associated to the description of radiation





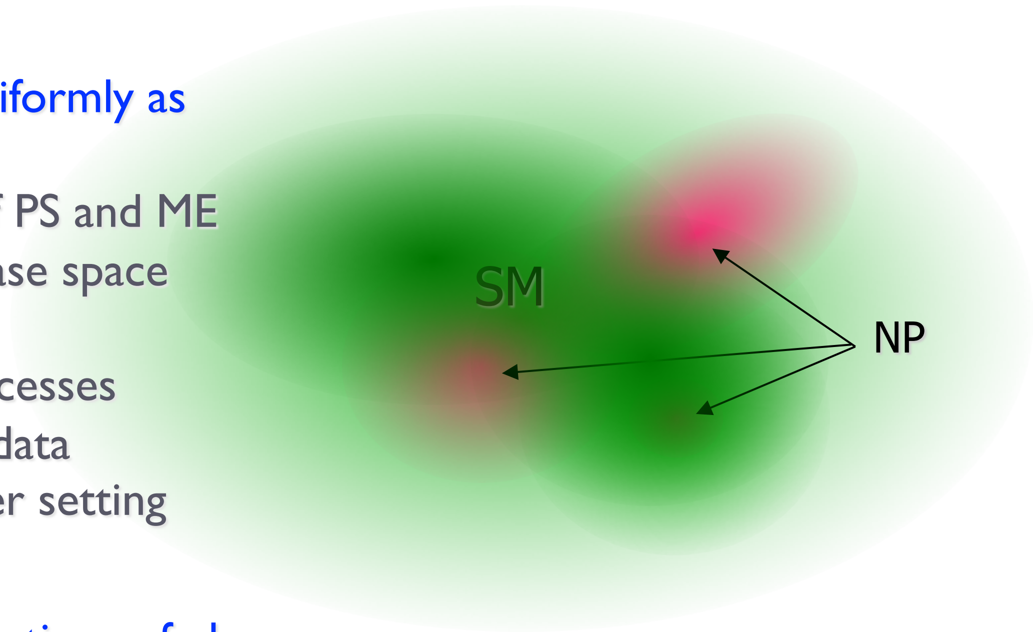
Decay handling

- Discussion ongoing in CMS
  - can we have a complete standard treatment everywhere?
- τ decays: TAUOLA de facto standard
- B decays: EvtGen de facto standard
- Ideally we want both
  - What is crucial where?
    - B-tagging people obviously want EvtGen
    - Analyses sensitive to τ polarization want TAUOLA
    - EvtGen can treat τ decays as well
  - Compare EvtGen and TAUOLA to understand impact of real differences in τ decays treatment
  - One could run TAUOLA and EvtGen in sequence everywhere
    - Technical work on avoiding bad interference on interfaces needed in case

CMS roadmap towards physics

- ✓ **Strategy for productions in CMS**
- ✓ **Current productions**
- ✓ **Conclusions**

Generation strategy



- ✓ Take care of the SM as correctly and uniformly as possible
  - ✓ Determine optimal combination of PS and ME for a complete coverage of the phase space
  - ✓ Coherent interface to showering, fragmentation, decay across all processes
  - ✓ Tuned underlying event. Use LHC data
  - ✓ Uniform choice for input parameter setting and PDFs
- ✓ Add generator redundancy in crucial portions of phase space
  - ✓ One prediction is always not enough: LO vs NLO vs different approaches
  - ✓ Different interface to showering – prepare to tune with data
  - ✓ Different settings to study systematics (tunings, PDFs,…)
  - ✓ Sensitivity of analyses and reconstruction methods to "theory/modeling" effects
- ✓ Then add New Physics samples (SUSY, BSM)
- ✓ Check the overlap with SM in poorly populated regions
  - ✓ "tails" of relevant distributions
- ✓ Improve there

The CMS way

- A large MC production effort has been done in view of the data taking
  - trigger studies, definition and overlaps between primary datasets
  - full SM coverage for training the analyses, especially for QCD studies and rejection
  - systematic studies where relevant, redundancy of generators for validation
- We want to use the right tools in the right portions of the phase space
  - calculations can sometime become very CPU intensive (e.g. W+4jets with ME)
- Large Monte Carlo production managed centrally
  - Especially since generate/compare/regenerate will be an iterative procedure
- How to make a complete and coherent generation?
  - partition the phase space, avoiding double counting of processes or duplication of MC samples
  - use as much as possible a reference generation setup
  - enforce as much as possible common input parameters/setups (PDFs, cuts,…)
    - Homogeneity among different samples, SM vs BSM, simplify detector effects unfolding

CMS current choices for ME

- MadGraph+PYTHIA as a reference ME generator for SM and BSM
  - can treat all phase space coherently, including SM+BSM interferences
  - do not give up higher leading order matched QCD contribution
  - flexibility of including any new physics
- Use ALPGEN+PYTHIA and MC@NLO+HERWIG as primary comparisons for the analyses
- Definition of different portions of phase space in collaboration with the MG/ME team, with theory-validated LHE files and corresponding binaries for Monte Carlo productions
- Trying SHERPA and POWHEG

The massive 2009 CMS production

- Both at $\sqrt{s}$ = 7 and 10 TeV, fully simulated (Geant4) and reconstructed events
  - For instance at 10 TeV (7 TeV roughly similar):
    - O(200 M) with PYTHIA 6
      - MB, QCD, γ +j, Drell-Yan and Onia, e/ μ from b or in-flight decays, W+j,Z+j
    - O (50 M) MadGraph+PYTHIA, O(30 M) ALPGEN+PYTHIA
      - TTbar, Z+j, W+j, other EWK
      - QCD control with HERWIG6 (~ 6 M), HERWIG++ (6 M)
      - W+j, Z+j control with SHERPA (12 M)
      - Forward samples with EXHUME, POMWIG (~ 2 M)
      - + signal samples…
- For the 0.9 and 2.36 TeV data taking:
    - O(50 M) at each energy with PYTHIA6
      - MB, QCD, Onia, detector studies

Conclusions

- It is essential to favor a consistent (same generation strategy for signal and background) and coherent (same settings, full phase space coverage) configuration for the reference MC samples
- We consider crucial to have generator redundancy for cross-checks and validation
  - we do not want to rely on just one prediction
  - learn about sensitivity to theory modeling when data alone is not sufficient
- Follow-up new developments
  - Although taking into account constraints imposed by production/analysis cycles
  - Avoid freezing on some well known setup "just because we are used to it"
- Get ready for tuning with data (MB/UE, radiation, fragmentation, PDFs,…)
  - Strategy in place
- Keep alive a good communication channel with the theory community