Input for the MC4LHC Workshop, March 2010
Collected within CMS Collaboration: *'Status, Experiences and Questions'*

1. **MC Event generators**
    1. **List of used Tools**
        a) Multi-Purpose Generators
        - PYTHIA6 (in past year v. 6.420, now migrated to v. 6.422)
            - Main multi-purpose Generator, used to generate ~400 M fully simulated 7 TeV events in past half year
            - used for various processes, EWK, QCD, Higgs, SUSY, ...
            - standard tune/shower used was D6T with Q^2 shower, plans/discussions to move to pt-ordered shower and Perugia tunes as default
        - PYTHIA8 (v. 8.108, now migrated to v. 8.130)
            - No dedicated production for 7 TeV, used in former productions for QCD (Di-Jet) production
            - used by EXOTICA group for un-particle processes
                - this is the only generator providing such processes, to the best of our knowledge
        - HERWIG6 (v. 6.510)
            - Used almost exclusively for QCD studies (comparison to PY6)
            - used as shower/hadronizer for MC@NLO
            - used together with JIMMY
        - HERWIG++ (v. 2.3.2, now migrated to v. 2.4.2)
            - used almost exclusively for QCD studies (comparison to HW6/PY6)
        - SHERPA (v. 1.1.2, now migrated to v. 1.2.0)
            - used almost exclusively in EWK studies (W/Z production)
            - used for EXOTICA , especially extra-dimensional models

        a) Multi-Leg Matrix-Element Generators
        - MADGRAPH (v. 4.4.13)
            - used extensively for QCD, VB(-pairs)+Jets, HQ (-pairs)+Jets, (di-)photon(+VB)+Jets, Z'
            - 'main' CMS LO Multi-Leg generator
            - interfaced to PY6 for shower/hadronization
        - ALPGEN (v. 213)
            - used mainly for top-pair, VB(+HQ)+Jets, QCD high jet multiplicity
            - serves in many cases as systematic comparison to MADGRAPH
            - ... or where many legs (more then MADGRAPH can do in reasonable time) are needed

        a) NLO Event Generators
        - MC@NLO (v. 3.41)
            - used for top-pair, single top, gluon-fusion

            Higgs, Drell-Yan, W-pairs

   - POWHEG
     - used for Drell-Yan, Higgs

a) Specialized Tools
   - Forward Physics:
     - POMWIG, EXHUME, HARDCOL
     - in preparation: CASCADE, POMPYT, RAPGAP
   - Heavy Ion Physics:
     - PYQUEN, HYDJET
     - in preparation/discussion: AMPT, EPOS, ...

1. **Event Generation Strategies in CMS**
   a) Most multi-purpose generators are completely integrated in the CMS software framework (CMSSW) by linking to external shared object libraries. This allows the users to generate the full events within the CMSSW run commands and can be configured via so called CMSSW configuration files, including any native generator commands.
   b) Event generators that provide information in the format of LHE files, that further need to be processed by a 'shower MC/hadronizer' (like MadGraph) are interfaced via a so called LHEInterface. This interface allows to read in events from LHE files (either stored in a MCDB article or in a local file) and process them using any tool interfaced to CMSSW as explained in a) above.
   c) The generation of the input LHE files happens then completely decoupled from the CMSSW framework, usually using out-of-the box tools (ALPGEN) or dedicated farms (MadGraph).
   d) A somewhat special case is ALPGEN, which provides the hard-scattering events in its native format. To cope with this we provide an AlpgenInterface in CMSSW, that translates back and forth ALPGEN native format into standard LHE format. After this ALPGEN event generation follows the same strategy as MadGraph explained in b) above.
   e) Tools that cannot be interfaced as explained above are, when possible, interfaced in specialized ways. This almost always results in
      - additional efforts by the generator group,
      - non-standard workflows for the computing group and
      - usually less 'interest' of the users, since additional expertise is needed to generate events.
   f) An example for this is Sherpa.
2. **Comments, Questions and Plans**
   a) At the moment event generation in CMS is rather PY6 (with respect to HW6, HW++ and PY8) centered. There are several reasons for this. In general there is a certain amount of confidence in the tool, as it has been used successfully in many analyses and experiments before. Secondly it comes, at least to our knowledge, so far with the most sophisticated and tested tunes.
   b) For the future, we are planning to enhance step-wise the usage of the newer C++ based tools PY8 and HW++. This

will need a certain amount of 'pressure' on the users from the GEN group in CMS.

c) To give more emphasis to this 'migration', it would be useful to **discuss at the workshop the pros and cons of the new tools** (i.e. What can PY8 do that PY6 cannot?) and stress (if this is the case) whether or not updates/support are planned in the future for the 'old' FORTRAN based tools (**maybe some overview talks by the authors would be good)**.

d) In this, also an **overview on tuning activities, and possibly an overview on the already available tunes for the C++ programs** would be desirable.

e) SHERPA, though doubtless a very powerful tool, has not yet that much impact in CMS. Especially the EWK group has however stated clear interest in using SHERPA.

f) As above, **it would be good to emphasize the advantages of SHERPA over other tools in a dedicated session, putting again some weight on tunes for the UE.**

g) Another reason for the lack of 'interest' in SHERPA is the rather complicated generation procedure. While other ME event generators (like MadGraph, see also above in point 1.2) can be interfaced easily via LHE files, for SHERPA we could not find such a solution. **If possible, such an interface would be highly desirable, i.e. To do in a first step the hard scattering event a la MadGraph and then to be able to do the subsequent steps starting from a ASCII LHE file.** Although we're not sure if this is technically possible, we wanted to address this wish (see also above).

h) In general we would like to emphasize that tools fulfilling one or more of the following technical requirements are very easy to interface to the CMSSW framework:
   - package available as shared object library, providing interfaces in C++ to (at least) the main routines
   - package ideally provides output (allows input) in HepMC format
   - package provides as output/accepts as input standard LHE format
   - all relevant parameters can be set via input files, i.e. no compilation step is needed
   - memory consumption of the tool should be under control

i) Understandable & detailed documentation is very important for the users of a generator. We believe this is one of the reasons why PY6 is very well received. It contains details on the physics behind the tool, as well as detailed documentation on the usage (examples, parameters, etc.). While we understand that writing detailed documentation is a major task, we think that other tools (HW++, and in particular PY8) would benefit from investing in this.

j) Special comments on HI tools:
   - HI collisions MC should take into account many

complicated collective phenomena. Simultaneous treatment of various nuclear effects is absent in most of current MC HI codes. Moreover, only a few HI MC generators are runnable at LHC energies and/or publicly available.

1. **(N)NLO Codes & Cross-Sections**
    1. **Inclusive Cross-Sections**
        a) Cross-Sections for processes are so far computed in an 'ad-hoc' way, i.e. each CMS subgroup (EWK, QCD, HIGGS, etc.) provides cross-sections for the processes of their interest.
        b) In most cases, these cross-sections do no correspond to the numbers provided by the Event Generation tools described in 1. above (these numbers are usually LO and not accurate enough).
        c) Which tools are used to compute these numbers depends on the process of interest. A tool used very often is MCFM (for many EWK processes), then the Spira (HDECAY, HIGLU, ...) tools for Higgs cross-sections. Prospino is used in the SUSY community. In addition we use NNLO coded where available (e.g. gg->H, HggTotal, HNNLO, Fehip, FEWZ).
        d) The experience with these tools is rather good, and they are used completely decoupled from the CMSSW framework. The resulting numbers are in the simplest case used to inclusively re-scale the distributions from the event generators to the desired amount of integrated luminosity.
        e) In special cases, or where no tools are publicly available, numbers are taken directly from dedicated publications.
        f) There are efforts ongoing (within CMS, but also in a wider community) to synchronize the numbers. An example for this is the Higgs@LHC working group, which aims to synchronize the Higgs cross-sections for all possible channels between ATLAS and CMS. Other similar efforts are ongoing in other fields.
    2. **Theoretical Errors on Cross-Sections**
        a) Errors on Cross-Sections are usually evaluated by
            - varying the renormalization & factorization scales in a range [mu/2, 2 mu] around some default central, process dependent scale mu,
            - varying the PDF sets (usually within the error sets of a default set).
        b) The total error is then computed as the square-root of the sum of the squares of the individual errors.
    3. **Exclusive Cross-Sections and K-Factors**
        a) In cases where higher order QCD corrections have impact on shapes of distributions, differential reweighting is used.
        b) An example for this is Higgs production in gluon-fusion, where Higgs pt-dependent K-factors are used to re-weight the PY6 events to the MC@NLO Higgs pt spectrum.
    4. **Comments & Questions**
        a) In general it is very desirable to have access to the dedicated programs for cross-section computation, i.e. the

programs should be available. While a provided number can be useful, it's usually necessary to being able to change some parameters (PDF's, center of mass energy, etc.).

b) This implies that easy installation and good documentation is always very welcome.

c) We have not yet found a standardized way on how to estimate theoretical uncertainties on cross-sections. We think a discussion on the following topics might be useful:
   - **combining PDF and scale uncertainties**
   - **evaluating uncertainties on exclusive observabels (i.e. on cross-sections in different jet-multiplicity bins)**
   - **evaluation of alpha_s uncertainties**

d) A general desire is that tools to compute cross-sections (e.g. MCFM) would provide a possibility to compute errors (e.g. from PDF error sets) in a standard format (i.e. Without the need of re-running the code for all error PDF sets).

2. **Decay Packages**
   1. **Dedicated Decay Packages used in CMS**

      a) We make use of the following tools, usually uniquely interfaced to PY6:
         - TAUOLA (v. 27.121)
         - Used for sample generation where emphasis lies on spin-correlations in tau decays.
         - EVTGEN (evtgenlhc v. 9.1)
         - Used for samples where decays of B-hadrons are of special interest.
         - PHOTOS (v. 215, within Tauola, standalone usage in preparation/discussion)
         - Planned to be used for leptonic decays of vector-bosons
   2. **Comments & Questions**

      a) As described above, the dedicated decay packages (Tauola and EvtGen) are
         - only interfaced with PY6 (although technically usable with HW and PY8 as well)
         - only activated for dedicated samples, where emphasis should be put on the decays

      b) However, within CMS we are trying to define a default decay treatment, that should be applied in every sample production.

      c) Unfortunately there is very little experience on the decay packages in HW++ and SHERPA.

      d) The following questions are occupying us the moment, and might be useful to be discussed
         - **how can Tauola and EvtGen be sensibly combined?**
           - **E.g. Tauola for prompt Taus from the hard scattering interaction, EvtGen for the Hadron decays (including Taus from heavy hadron decays)**

- **Comparison of EvtGen and Tauola in tau decays: 'Is EvtGen sufficient? Is the Tau decays treatment homogeneous enough among the two tools?'**
- **Interfacing EvtGen and Tauola to C++ generators: We understand that, to our knowledge at least SHERPA and HW++ have their own sophisticated decay modules. How do they compare to EvtGen/Tauola? And how the QED FSR radiation in HW++ and Sherpa compares with PHOTOS?**

3. **Other Comments & Questions Collection**
   1. **Jet-Multiplicity for W+Jets in ME+PS Generators**
      a) "can we extrapolate the cross-section from the N Jets bin to the N+1 Jets bin? How to treat related systematic uncertainties?"
      b) Heavy Flavor content in W+qq+Jets:
         - there is the need to understand sigma(Wbb(cc) +Jets)/sigma(Wqq+Jets) for different jet multiplicity bins. How well is this understood, tested at Tevatron? How should uncertainties be treated?
      c) Charge asymmetry in W+Jets:
         - is the charge asymmetry sensitive to the Jet multiplicity?
      d) Treatment of ISR/FSR/matching uncertainties: Suggestions? Recipes?
         - Should the matching scale be understood as a tunable parameter?
   2. **PDF packages & LHAPDF**
      a) Although we're aware of the dedicated working group, we think a discussion on PDF sets would be useful. What recommendations for default (N)LO sets could be done? Pros and cons on so called LO* PDFs...
   3. **Generating Events for new models**
      a) Having the possibility of quickly include new models to compare with is very important especially for any beyond the Standard Model study. The combination of tools like FeynRules and LanHEP with CompHEP can allow to quickly produce LO partonic level results in LHE format which can be easily feeded in the CMS simulation, as explained in section 1.b. It is therefore of **great help if new models come directly with a FeynRules calculation available.** Another example of such needs is the usage of the BRIDGE decay package in combination with MadGraph to facilitate the inclusion of SUSY decays.
   4. To the reader: Consider this document as a snapshot of the situation in CMS. We cannot cover all topics, nor the input of all CMS members, thus the topics are naturally biased by the authors. Understand the points raised as topics that keep us busy and considered to be worth being discussed (again) in an expert environment.
   5. **Finally, we believe a review of Tevatron results (focusing on Data-MC comparison) in the main areas would be**

interesting and useful. Please consider this suggestion.