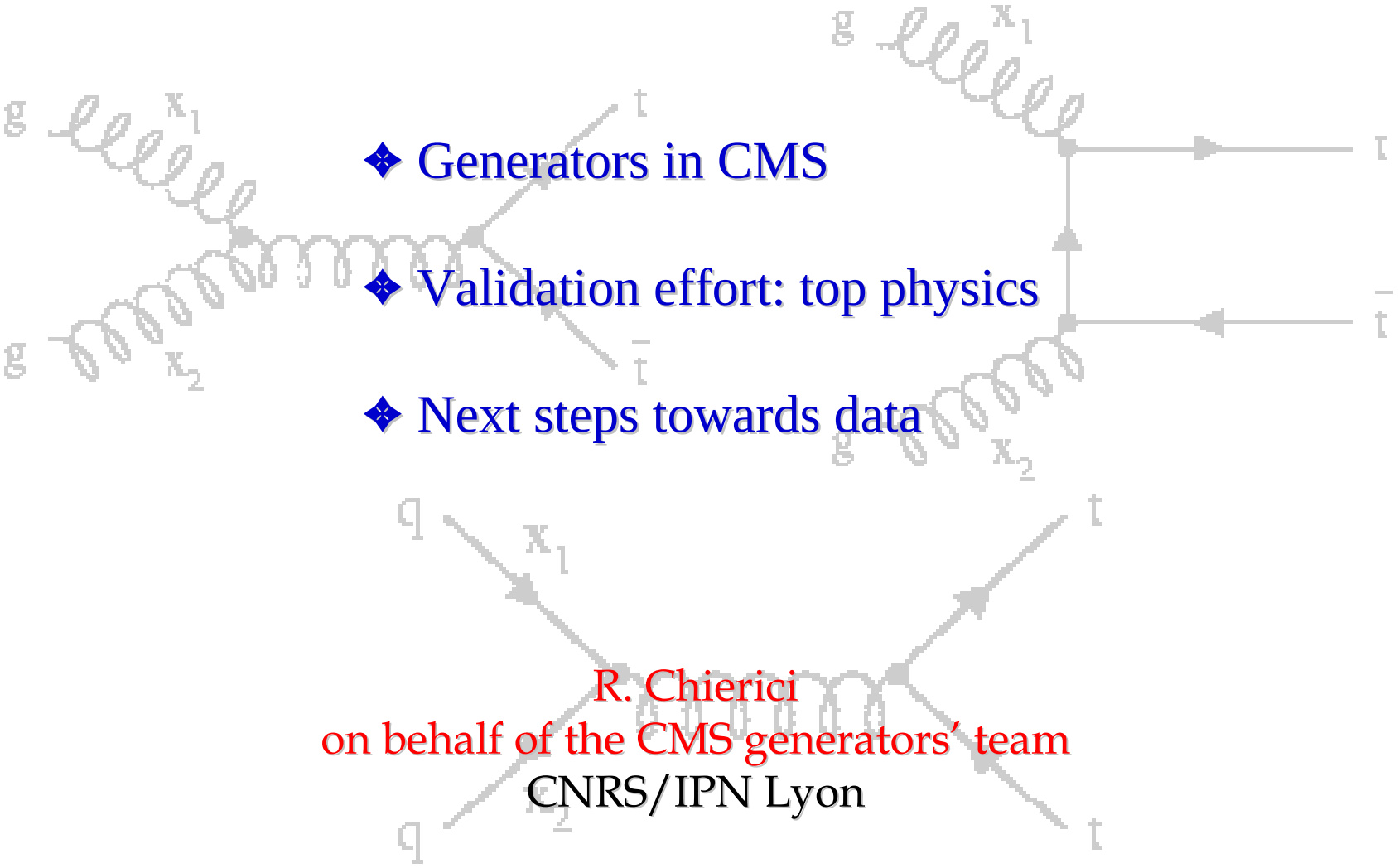
Event generation in CMS



- Generators in CMS
- Validation effort: top physics
- Next steps towards data

R. Chierici
on behalf of the CMS generators' team
CNRS/IPN Lyon

Generators in CMS

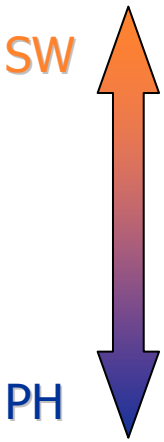
- Organizational issues
- Interfacing a generator to the CMS software

Roberto Chierici

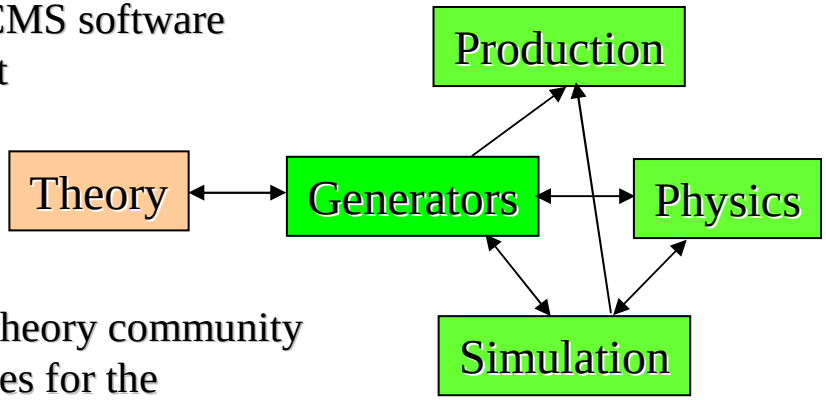
Introduction: the generator team and CMS

2008 is a crucial year for completing all choices related to Monte Carlo productions.
It will be the year of the first MC generation for data taking.

The PH-CMS generators' team does not provide a mere technical support to the other CMS physics teams. In its mandate it should (not in order of importance):



1. be responsible of the generator interfaces to the CMS software
2. coordinate the event generation from the different physics groups' requests
3. decide the best generation setup (what generator, what tuning) in strict communication with the CMS physics teams
4. strictly communicate (and collaborate) with the theory community
5. contribute to the generator validation and strategies for the generator tuning with data



This work implies quite a lot of software and quite a lot of physics.

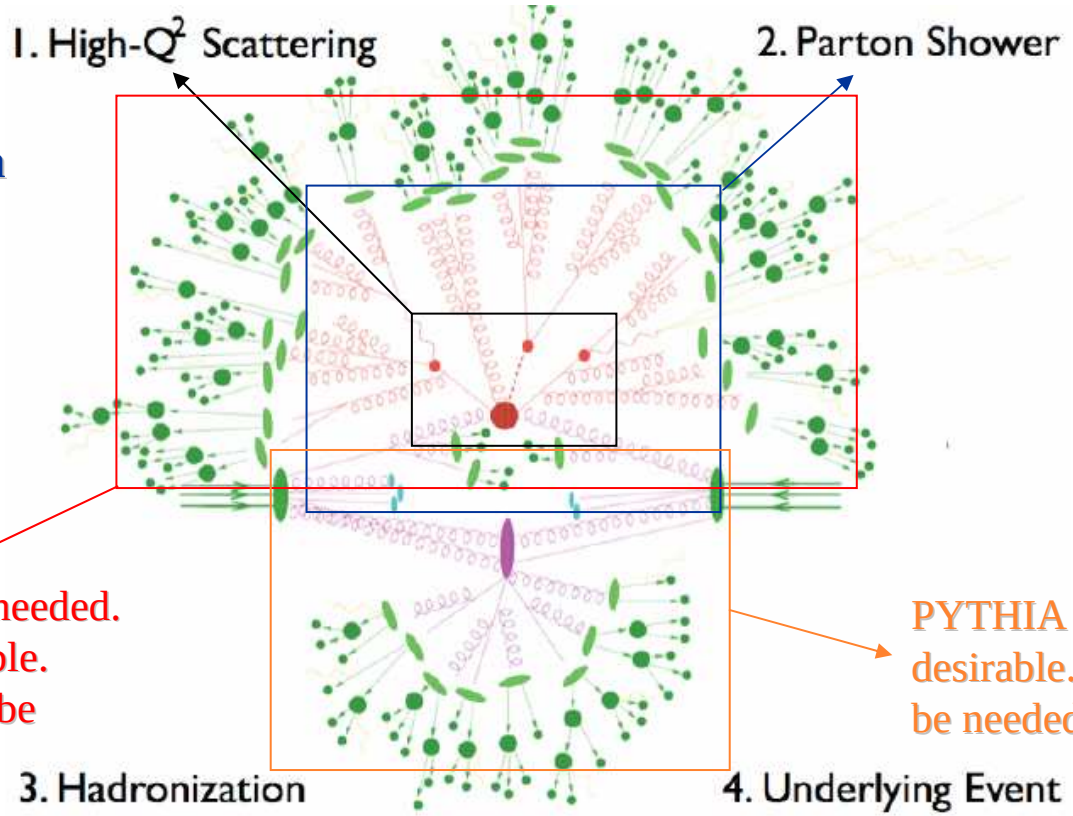
This team should also represent the natural window towards the theory world.
Conversely: theory should communicate to CMS in preference via our group.

Desiderata for a Monte Carlo tool

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

Interface to PYTHIA needed. HERWIG very desirable. Tuning with data will be needed.

The MC description of LHC events is tremendously complex



Interface to PYTHIA needed. HERWIG very desirable. Tuning with data will be needed.

PYTHIA MPI. HERWIG/JIMMY desirable. Tuning with data will be needed.

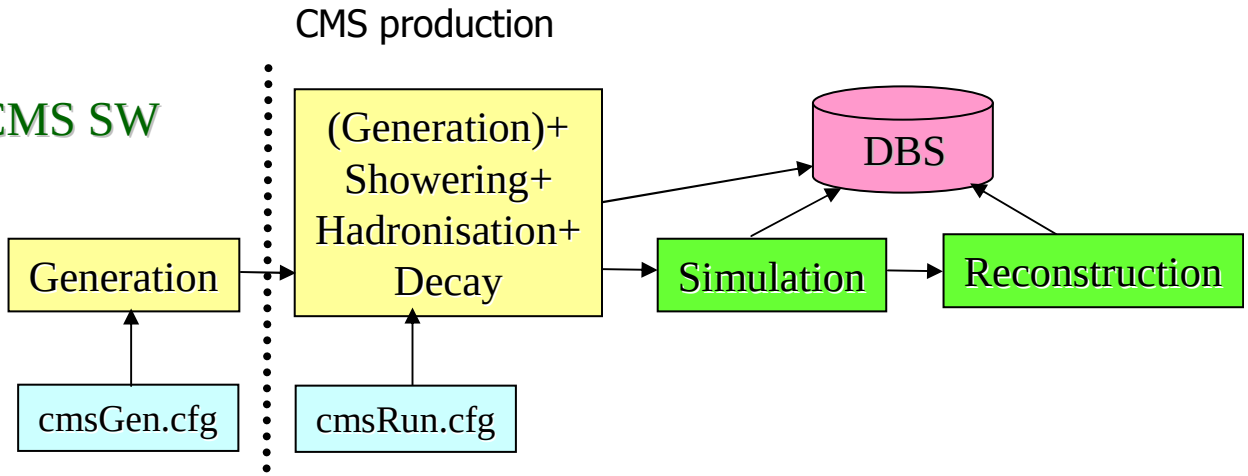
Other desirable features, from the experimentalist's viewpoint:

- 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 ☺

Generators in CMS(SW)

- CMS have preference for the generators integrated in the experiment software, that can directly be used in production. Methods also exist to start from an externally produced parton level (LHE) file

- The generator interfaced to the CMS SW must undergo:
  - ✓ a technical validation to show its ability to run standalone and in production
  - ✓ a physics validation wrt a similar content generator



- The physics program at the LHC is very rich:
  - ✓ pp General purpose: Pythia6, Herwig6, (Pythia8, Herwig++), Sherpa
  - ✓ pp HLO: Alpgen, MadGraph, Helac, Sherpa
  - ✓ pp NLO: MCatNLO
  - ✓ pp Others: CompHEP, TopRex, Phantom
  - ✓ Diffractive physics: Pomwig, Exhume, EDDE
  - ✓ Decayers: EvtGen, Tauola, Photos
  - ✓ Heavy Ions: Hydjet, Pyquen
  - ✓ Detector specific: Cosmic muons, particle guns, beam halo, beam-gas
  - ✓ New physics specific: Charybdis

Interfaced generators

Plethora of generators for the very rich physics program of CMS. The list is not exhaustive.

Generator	View CVS	Documentation	Responsible	Status
Pythia6	Pythia6Interface	View Twiki	Julia Yarba	ready
Herwig6	Herwig6Interface	View Twiki	Fabian Stoeckli	ready
ALPGEN	AlpgenInterface	View Twiki	Maurizio Pierini, Maria Spiropulu	ready
MadGraph	MadGraphInterface	View Twiki	Maria Hansen, Dorian Kcira	ready
CompHEP	CompHEPInterface	View Twiki	Sergey Slabospitsky, Dimitri Konstantinov, Lev Dudko	in progress
MC@NLO	MCatNLOInterface	View Twiki	Fabian Stoeckli	ready
TopRex	TopRexInterface	View Twiki	Sergey Slabospitsky	advanced (but no doc)
StaGen	StaGenInterface	View Twiki ?	Sergey Slabospitsky	advanced (but no doc)
Charybdis	CharybdisInterface	View Twiki	Halil Gamsizkan	advanced (but no doc)
Hydjet	HydjetInterface	View Twiki	Camelia Mironov	in progress
Pyquen	PyquenInterface	View Twiki	Camelia Mironov	in progress
EvtGen	EvtGenInterface	View Twiki	Aniello Nappi, Roberto Covarelli	in progress
Phantom	MadGraphInterface	View Twiki	Sara Bolognesi	ready
ResBos	ResBosInterface	View Twiki	NN	??
Cosmic Muon Generator	CosmicMuonGenerator	View Twiki	Philipp Biallass	ready
Beam Halo Muon Generator	BeamHaloGenerator	View Twiki	Emmanuelle Perez	advanced (but no doc)
Beam Gas Generator	BeamGasGenerator ?	View Twiki	NN	??
Pythia8	Pythia8Interface	View Twiki	Mikhail Kirsanov	in progress
Herwig++	Herwig++Interface	View Twiki ?	Oliver Oberst	??
ExHume ?	ExHumeInterface	View Twiki	Antonio Vilela Pereira	ready
Pomwig	PomwigInterface	View Twiki	Antonio Vilela Pereira	ready
EDDE	EDDEInterface	View Twiki	Andrei Sobol et al.	in progress
SHERPA	to come	View Twiki Validation	Martin Niegel, Markus Merschmeyer, Altan Cakir	in progress

2007 Monte Carlo production

The 2007 Monte Carlo production (with full detector simulation) is approaching the end, will serve for the next round of analyses. Organized in:

- A SM event "cocktail" (100M events), processed ad the Tier-0, covering the bulk of the SM processes

Min bias (20 Mevt)
QCD jets (25 Mevt) and γ+jets (5 Mevt)
Electrons/muons from b-decays (20 Mevt)
Drell-Yan and Onia (10 Mevt)
W/Z (plus jets) (15 Mevt)
Top (5 Mevt)

PYTHIA, pthat bins for QCD

ALPGEN, matched production

- Many "signal" samples (order of 100M), processed at the Tier-2, covering all Higgs and BSM physics signal, as well as SM redundancy and tails

Made with PYTHIA, ALPGEN, MC@NLO

For the 2008 productions we plan to make much more use of MC@NLO and MadGraph. New (in the sense of their use in CMS) generators must be validated first…

Generator validation: an important example

- The importance of matching in tt
- Matched generators vs LO ones

Generation validation for top (SM) physics in CMS

Aim: test generators and validate their physics content (for top pair physics) in the frame of the CMS software. This will test them in the way we use them (debug !) and help understanding their features and make generation choices.

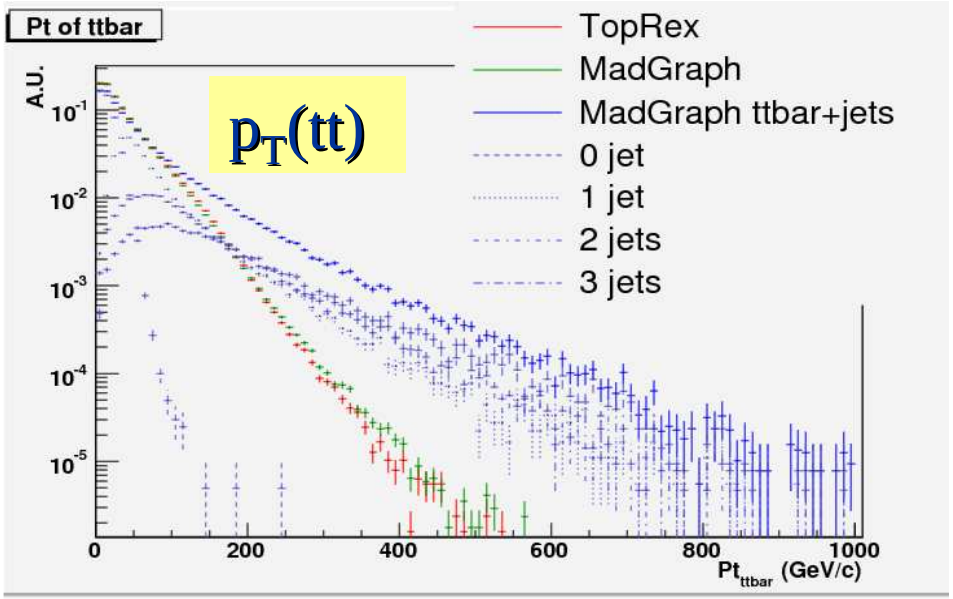
Input parameter settings (cuts, PDFs, scales) as uniform as possible.

- ✓ LO "standard": TopRex, MadGraph
  - $pp \to tt \times tt \to bf_1f_2bf_3f_4$; $pp \to (tt \to) bf_1f_2bf_3f_4$
  - top spin correlations included
- ✓ LO 6 fermions: Phantom, MadGraph
  - $pp \to bf_1f_2bf_3f_4$
  - interferences properly included
- ✓ HLO with ME-PS matching: ALPGEN, MadGraph
  - $pp \to ttNj$, N=0,…3+; $tt \to bf_1f_2bf_3f_4$; j=u,d,s,g
  - matching scheme implemented
  - inclusion of leading higher order terms
  - top spin correlations included
- ✓ NLO: MC@NLO
  - $pp \to ttg$; $tt \to bf_1f_2bf_3f_4$
  - NLO description
  - spin correlations added for all top final states

Work being completed: further statistics from MC@NLO and HELAC.

Matched ME-PS vs standard generation+PS

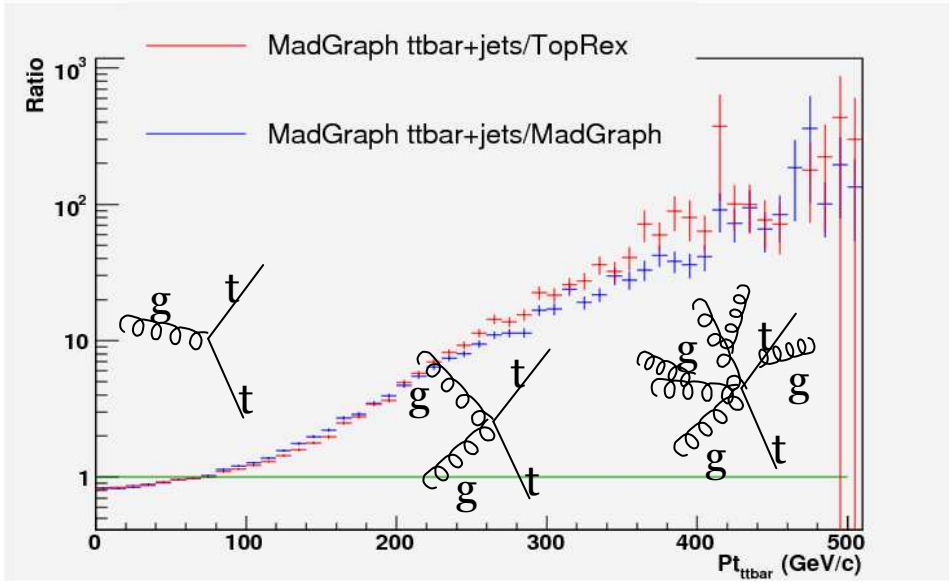
Spectacular differences in transverse variable connected to global radiation

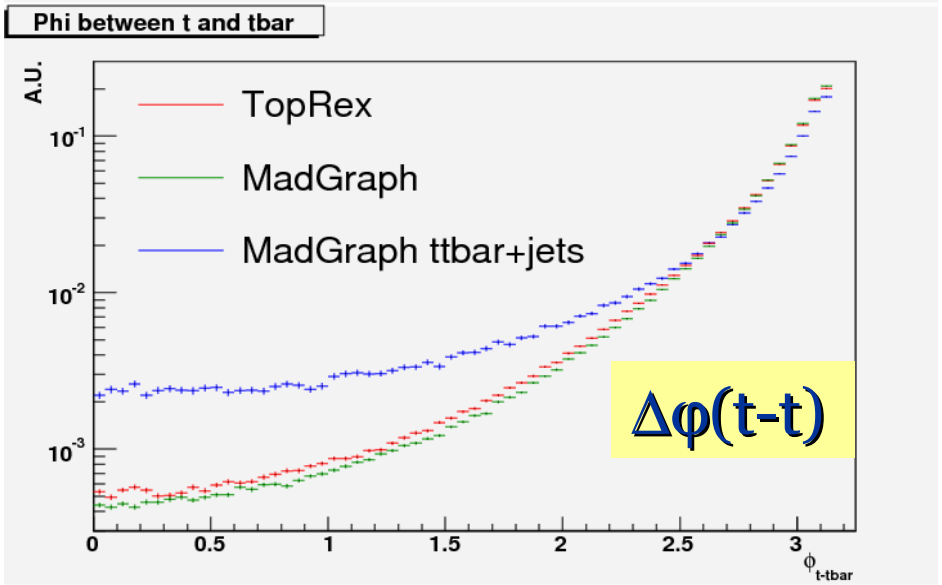


- Large effects at high $p_T(tt)=p_T(radiation)$
- Average $p_T(tt)$~60-70 GeV !
- 40% probability that a tt system recoils against a radiation larger than 50 GeV

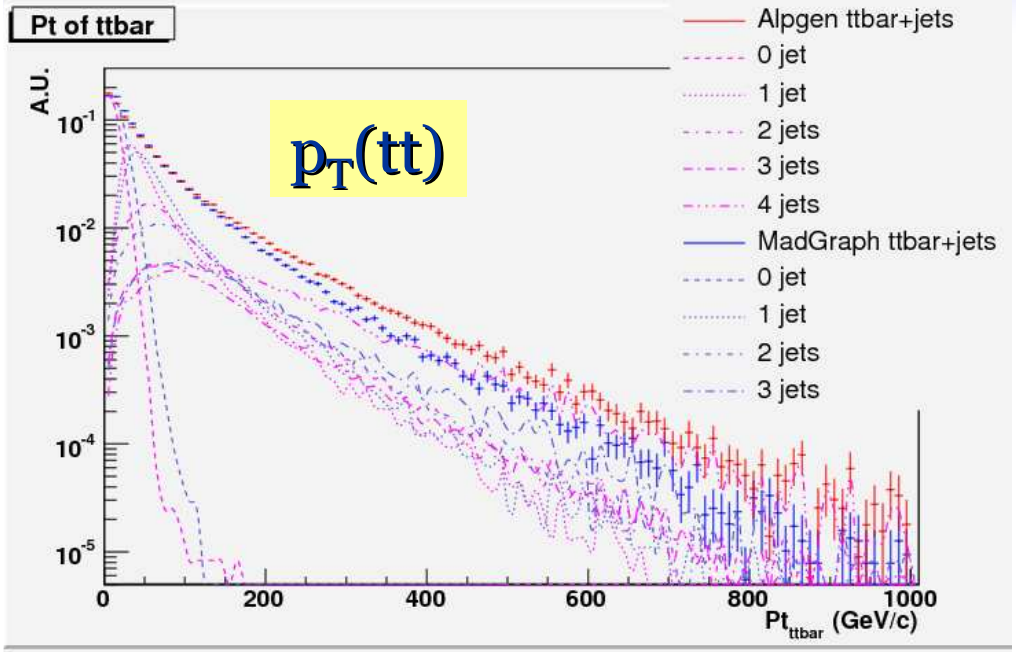
→ effect on reconstruction

→ Mandatory to use the same strategies for physics backgrounds like W/Z+Njets





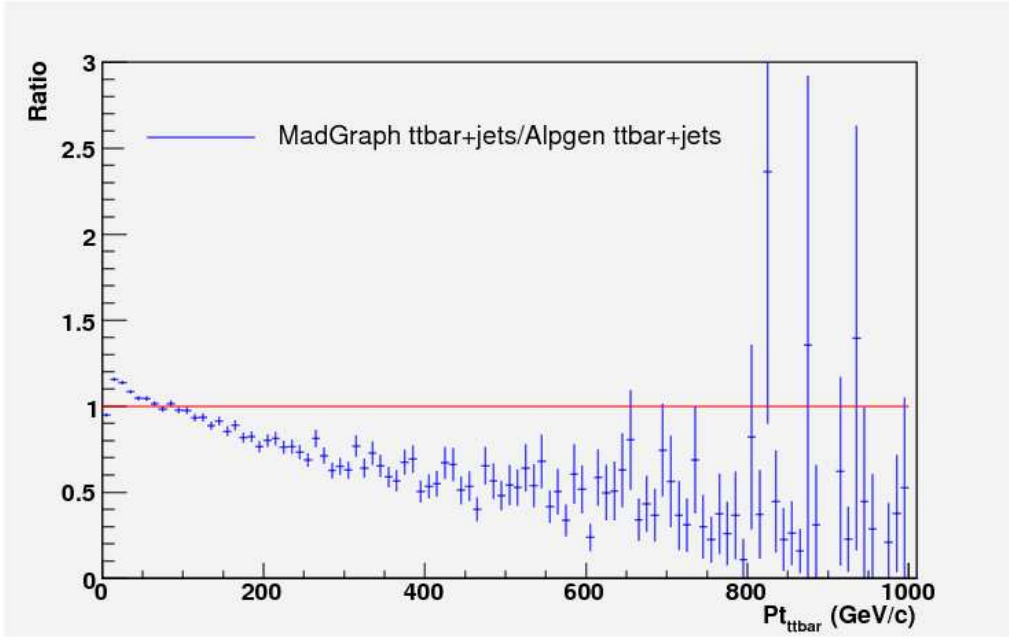
ALPGEN vs MadGraph matched

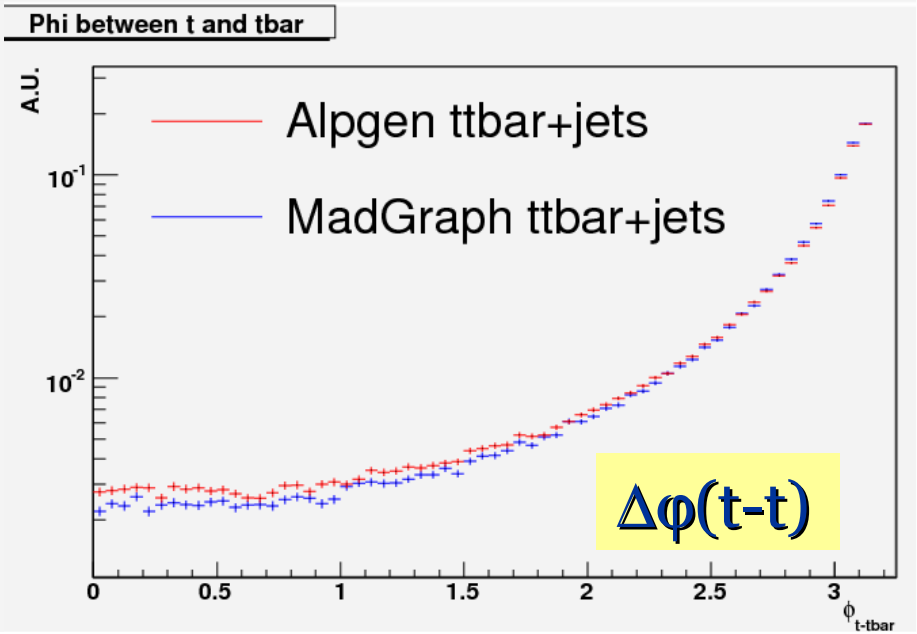


ALPGEN and MadGraph differ by at most 50% on the p_T prediction (several orders of magnitude away from the PS description)

Important to estimate these effects for the analyses:

- Effect of renormalisation and factorisation scales on the predictions
- Effect of the chosen ME-PS matching scale
- Comparison with MC@NLO

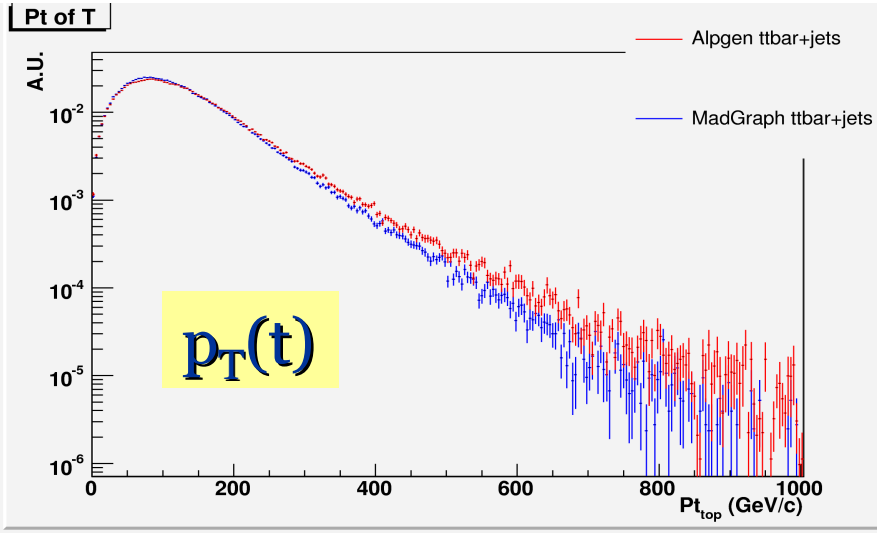


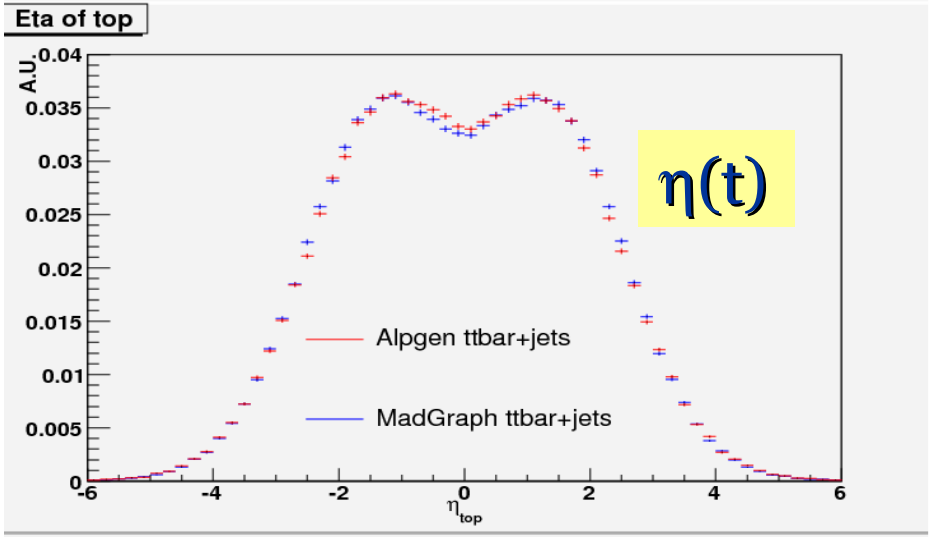


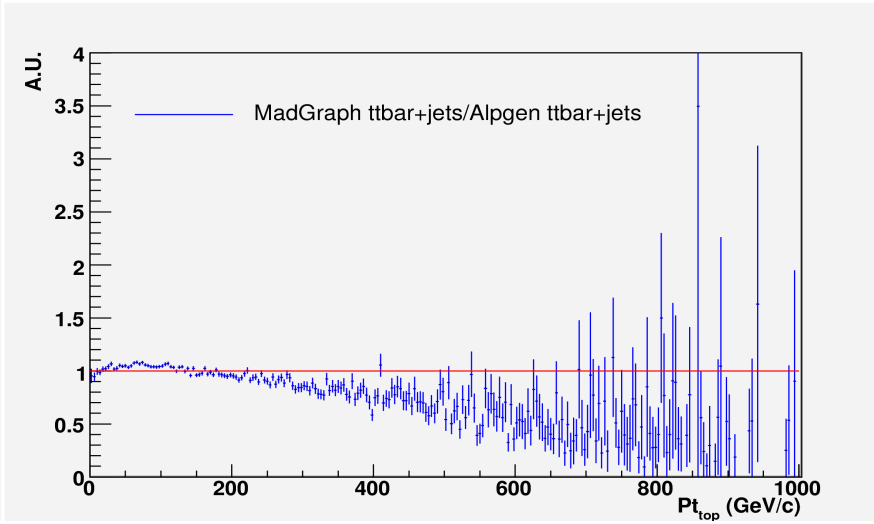
ALPGEN vs MadGraph matched

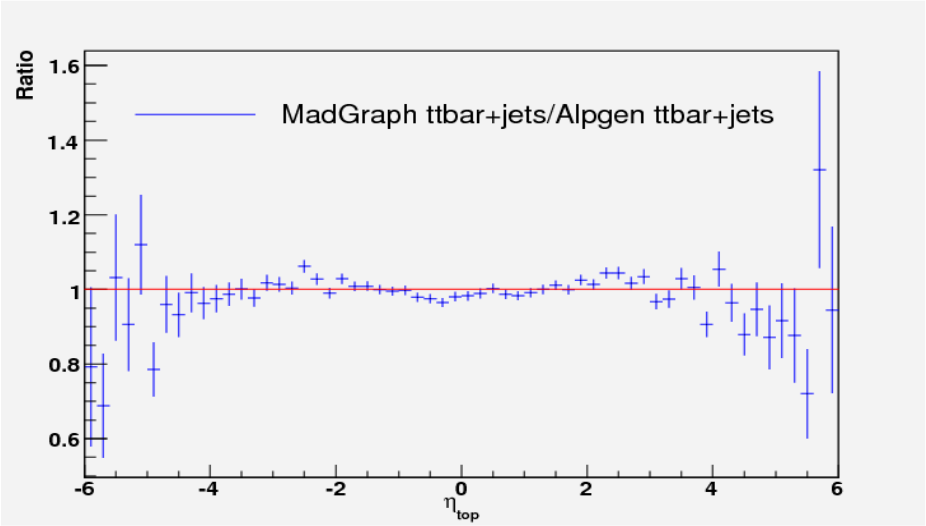
Excellent description of other variables for top physics

Residual differences due to different generator input settings and scales chosen?









Comparison with MC@NLO: anticipation

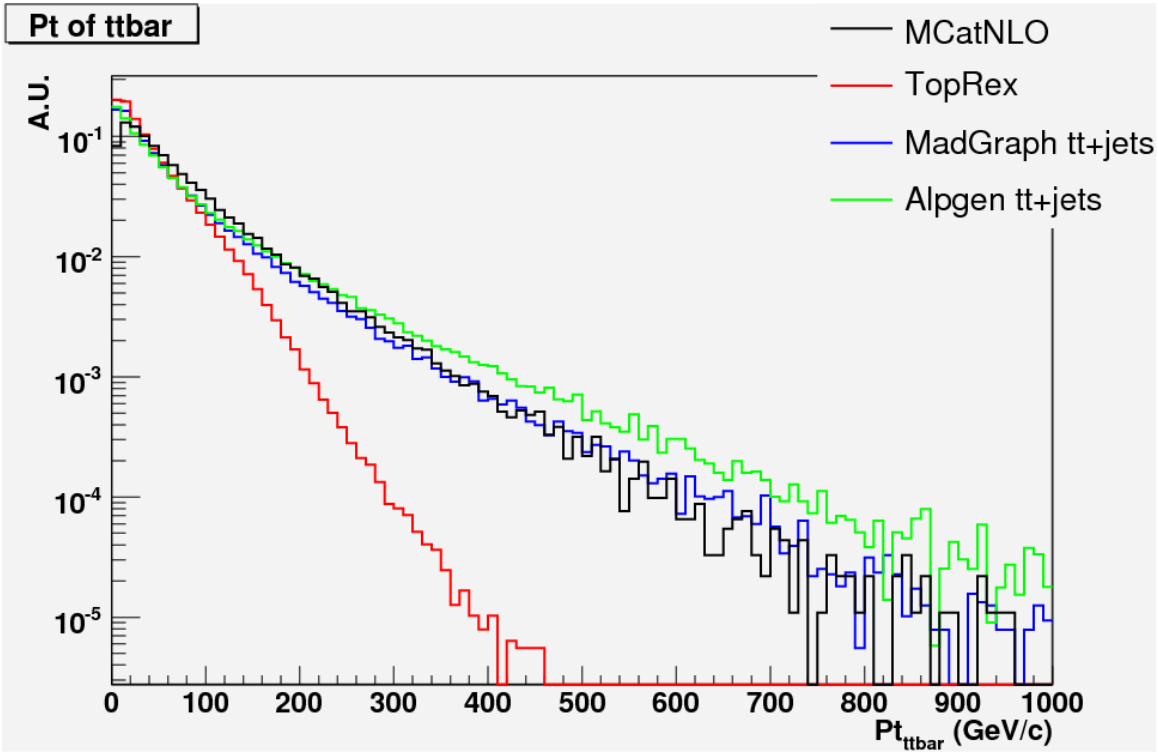
Comparisons to MC@NLO ongoing in CMS. Work just started.
Different conceptual problems in interpreting the results:

- 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. It can also give indications of:

- Relative importance of first emission in the description of the process
- Indication of systematic errors associated to the description of radiation.

Only 130K MCatNLO events processed so far. Indications of agreement to matched tools in the tail…



Roadmap towards physics

- Plans for next Monte Carlo generation

Next generation in CMS

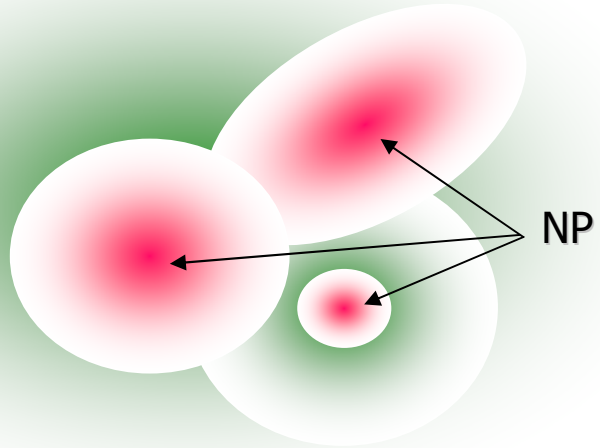
- The next Monte Carlo production in CMS (starting in May) should bring us to the interpretation of the first data (hopefully).

- CMS is currently planning to focus on:
  - ✓ a full simulation production of the order of 100M events. This production corresponds to the first weeks of data taking, where the main component is QCD+MB
    - → test of the computing flow and basic object reconstruction
    - → understanding of the first days/weeks of data with startup simulation conditions
  - ✓ a fast simulation production of the order of 1G events, corresponding to 3-6 months of data taking ay 20% efficiency and 300 Hz rate to storage
    - → full SM coverage for training the analyses

- Huge effort ongoing to provide all needed validation (technical mostly) in such a way that all needed generators and corollary packages will be ready for this effort.
  What cannot be renounced (our current choice):
  - ✓ General purpose: Pythia6, Herwig
  - ✓ HLO: Alpgen, MadGraph
  - ✓ NLO: MCatNLO
  - ✓ Decayers: EvtGen, Tauola, Photos
  - ✓ Desired asap: Sherpa, Pythia8

Generation strategy

- Take care of the SM as correctly as possible
  - → Satisfactory ME description
  - → Interface to showering, fragmentation, dec
  - → Tuned underlying event
  - → Best PDFs, accurate tuning of input parameter settings



- Add generator redundancy in crucial portions of phase space
  - → HLO vs NLO
  - → Different interface to showerings – prepare to tune w
  - → Different settings to study systematics (tunings, PDFs,…)

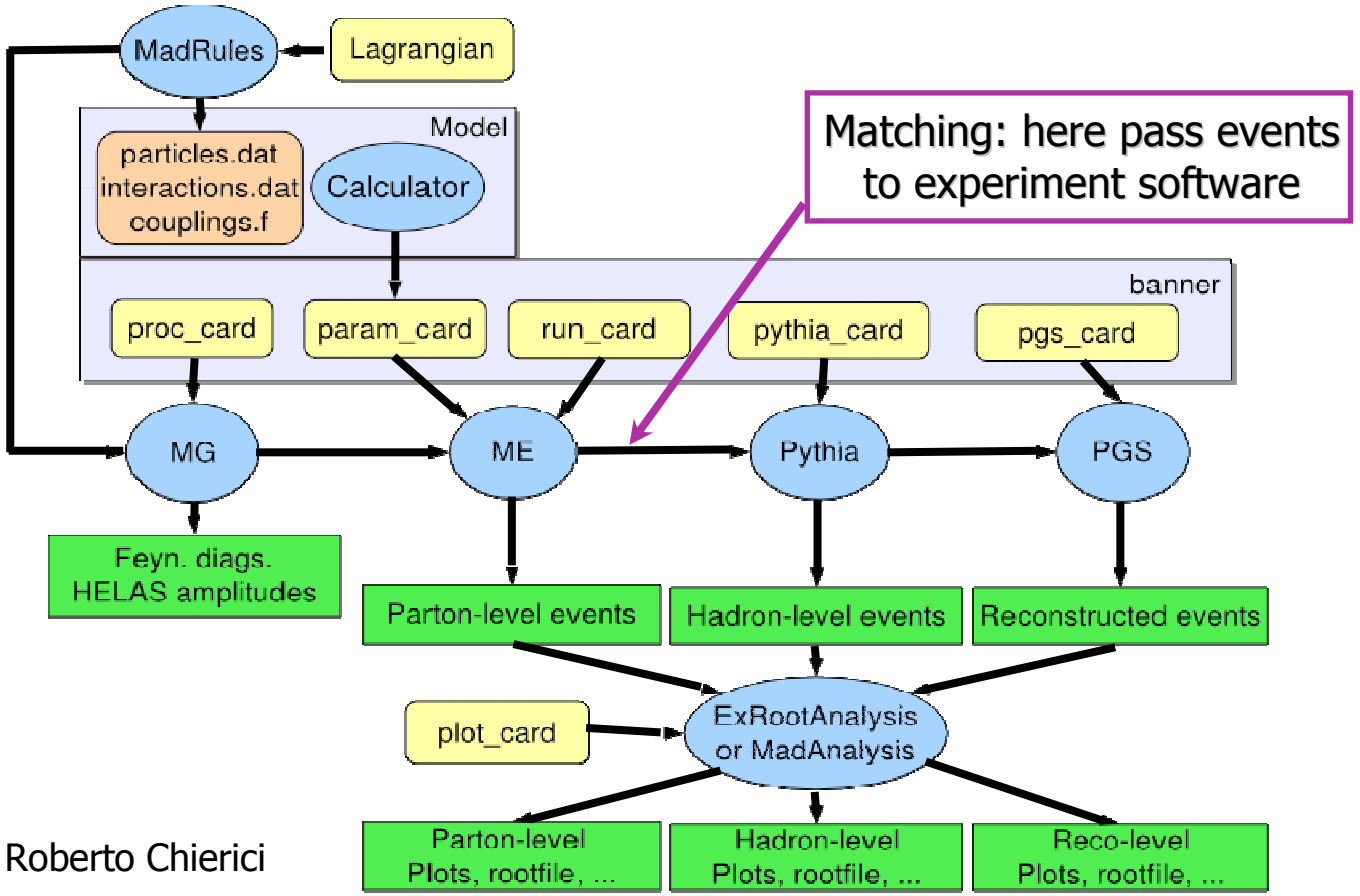
- Add new physics samples
  - → Main SUSY and BSM points to train analyses

- Determine tails
  - → Study what tails are most interesting

- The use of an as much coherent as possible set-up will ease enormously the tasks of the analyses, disentangling detector effects from the physics input to the generation

MadGraph-MadEvent flow

- An interesting proposal was push forward by the MG-ME team, proposing to provide theory-validated LHE files and corresponding binaries to the experiments for Monte Carlo productions.
- Need to agree on the file contents (processes, cuts, settings, …)
- Can treat SM, BSM and exotic models on the same footing, without giving up the nice features of ME-PS matching.



SM: QCD and vector bosons

in collaboration with F. Maltoni and the MG team

Process	Stars	Couplings	Phase space region	Matching	Banner	Event files	Remarks
jets (2)	2	QCD only	pt(at least 1)>X or pt(at least 2)>Y or pt (at least 3)>Z or pt (at least 4)>K	0,1,2,3,4+			light jets are u,d,c,s,g; Need to veto the first gluon splitting into bb in the PS
bb~ + jets	1	QCD only	pt(at least 1)>X or pt(at least 2)>Y or pt (at least 3)>Z or pt (at least 4)>K	0,1,2,3+			massive b; Need to veto the first gluon splitting into bb in the PS
bb~ bb~+ jets	1	QCD only	pt(at least 1)>X or pt(at least 2)>Y or pt (at least 3)>Z or pt (at least 4)>K	0,1+			massive b

Process	Stars	Couplings	Phase space region	Matching	Banner	Event files	Remarks
W (-> l v)+ jets	3	EW=2 + QCD	all	0,1,2,3,4+			W=W+,W- ; l=(e,mu,tau)
Z /a* (-> l+l-)+ jets	3	EW=2 + QCD	m(l+l-)>50 GeV	0,1,2,3,4+			photon is included ; l=(e,mu,tau)
Z (-> vv)+ jets	2	EW=2 + QCD	pt(Z)>100 GeV	0,1,2,3,4+			
V (-> l l')+ QQ~ + jets	1	EW=2+QCD	all	0,1,2+			V=W+,W-,Z ; l=(e,mu,tau,v), (Z->vv included) Q=b
a + jets	1	EW=1 + QCD	pt(a)>20 GeV, abs(eta(a))<2.5, DeltaR(a,jet)>0.3	0,1,2,3,4+			photon
a + QQ~ + jets	1	EW=1 + QCD	pt(a)>20 GeV, abs(eta(a))<2.5, DeltaR(a,jet)>0.3	0,1,2+			photon; Q=b
VV(-> 4l)+ jets	3	EW=2+QCD	all	0,1+			V=W+,W-,Z l=(e,mu,tau,v)
VV (-> 4l) + QQ~	1	EW=1 + QCD	all	no			V=W+,W-,Z l=(e,mu,tau,v), Q=b
aV(-> 2l)+ jets	1	EW=2+QCD	all	0,1+			V=W+,W-,Z l=(e,mu,tau,v)
a a + jets	1	EW=2+QCD	pt(a)>20 GeV, abs(eta(a))<2.5, DeltaR(a,jet)>0.3	0,1,2+			photon
a a + QQ~ + jets	1	EW=1 + QCD	pt(a)>20 GeV, abs(eta(a))<2.5, DeltaR(a,jet)>0.3	no			photon; Q=b
V V V	3	EW=3	all	no			V=W+,W-,Z
a a a	3	EW=3	pt(a)>20 GeV, abs(eta(a))<2.5, DeltaR(a,jet)>0.3	no			

SM: top and Higgs

in collaboration with F. Maltoni and the MG team

Process	Stars	Couplings	Phase space region	Matching	Banner	Event files	Remarks
tt + jets	3	QCD only	all	0,1,2,3+			top decays into everything. Done with DECAY
tt + bb~	3	QCD only	all	no			top decays into everything. Done with DECAY
tjb	3	EW=2, QCD=1	all	no			t-channel, b massive, no top decay
tj	3	EW=2, QCD=0	all	no			t-channel, no top decay
tb	3	EW=2, QCD=0	all	no			s-channel, b massive, no top decay
tW	3	EW=2, QCD=1	all	no			tW-channel, no top decay
tWb	3	EW=2, QCD=2	all	no			tW-channel, b-massive, doub-res diagram subtraction, no top decay

Process	Stars	Couplings	Phase space region	Matching	Banner	Event files	Remarks
Higgs + jets	2	QCD only	all	0,1,2,3+			HEFT, mh=120,140,160,180,200
Higgs + 2 jets	3	EW only	all	no matching			mh=120,140,160,180,200
tt~ + Higgs	3	QCD=2,EW=1	all	no			mh=120
V (-> l l') + Higgs + jets	2	EW=3 + QCD	all	0,1,2			

- All samples will be validated individually by the MG team (matching, sanity checks) and given in a format that can be run in the CMS production.
- SUSY and BSM working points will be added as well when agreed upon.
- Ongoing brainstorming to see what is the best way to generate events, the experimental community believes an event admixture will be very helpful
  - ✓ how to properly mix events?
  - ✓ how to best bias the huge QCD cross-section

Roberto Chierici

Other ideas/activities towards physics

The effort of Monte Carlo validation with new tools in CMS is started and must continue. Pythia8, Herwig++, Sherpa just to mention a few examples.

- from basic sanity checks to physics validation for all supported models
- from local to global validation
- convenient framework are under evaluation

Work in view of interpreting first data. Strategies and tools for understanding and tuning Monte Carlo must be put in place now in collaboration with the physics teams and with theory.

- analysis dependence on different scales
- analyses sensitivity to MPI models
- PDF constraining
- tuning of radiation and fragmentation parameters
- UE-MB tuning – see Paolo's talk

Need to set in place tools/frameworks where needed.

Strategies for assessing theory systematic need also to continue. Extra MC generations must always be accounted for this: for most/all of the analyses we cannot afford and rely on just one generator.

Conclusions

Summary

This is a crucial year for Monte Carlo production in CMS.

We want to organize the Monte Carlo production in such a way to favour a consistent (same generator) and coherent (same settings, full phase space coverage) configuration for the reference generation.

The SM reference should have matched HLO contributions for high multiplicity/p_T physics, maintaining a good description of the softer part. The CMS preference goes for tools also able to easily implement new physics models.

Favour generator redundancy in our production for cross-checks and first evaluation of analyses' sensitivity to systematic effects.

Prepare for tuning with data, define the best use of the first data to constrain the SM description in the Monte Carlo (MB, radiation, scales, PDFs,…)

Keep very much alive the level of communication with a) theory groups, b) ATLAS and c) LCG

Exciting years are in front of us, it is in good part upon the work in the generator team now to shape the way how CMS will do physics. At the startup and not only.