CMS Full Simulation: Evolution Toward the 13 TeV Run

M. Hildreth[1], V. Ivantchenko[2]

[1] Université de Notre Dame du Lac & Fermilab

[2] CERN

Representing the CMS Simulation Group

CHEP2013, Amsterdam, 14 – 18 October 2013

Outline

- CMS Full Simulation (FullSim) status
- Requirements for FullSim for the future 13 TeV run
- Why Geant4 10.0?
- How to speed up Geant4 simulation?
  - CMS input for Geant4 10.0
  - CMS Software improvements
- Summary

CMS FullSim in Production

- 2011 run:
  - 7 TeV Monte Carlo production: $3.0 \cdot 10^9$ events
  - Geant4 9.4 (+ few fixes)
  - Slc5_amd64_gcc434
- 2012 run :
  - 8 TeV Monte Carlo production
    - $6.5 \cdot 10^9$ events produced
    - Started with Geant4 9.4p03 and slc5_amd64_gcc434
    - Continues with slc5_amd64_gcc462 and CHIPS stopping fix
- Legacy re-production
  - 7 TeV and 8 TeV
    - Bremsstrahlung and multiple scattering backported from Geant4 9.5
    - slc5_amd64_gcc462 and slc6_amd64_gcc462
- Production Physics List QGSP_FTFP_BERT_EML
  - A few alternative Physics Lists are available, too

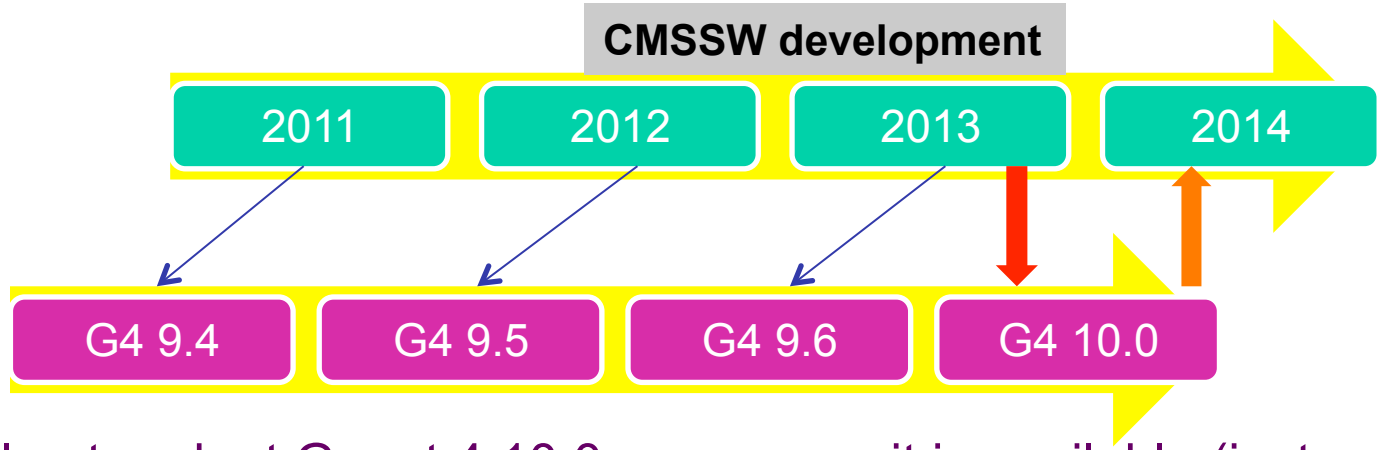
CMS FullSim in Development

- Geant4 9.6p02 is the current development version
  - slc5_amd64_gcc481
  - slc6_amd64_gcc481
- New challenges for 13 TeV run in 2015
  - higher energy, higher multiplicity, higher pileup
  - larger integrated luminosity
  - FastSim (See talk by A. Giammanco) is an option for many analyses
  - Speed up of FullSim necessary to keep up

Process	8 TeV	14 TeV
MinBias	19.3s	21.5s 111%
Z→e+e-	50.9s	116.9s 230%
ttbar	87.1s	115.8s 133%

Why Geant4 10.0?

- Geant4 version 10.0 is under developement
  - Will be multithreading-capable (see talk of Gabriele Cosmo)
  - Physics will be improved
  - CMS requirements for particular improvements may be implemented



- Plan to adopt Geant 4 10.0 as soon as it is available (instead of delay)
  - Start from sequential build and slc6_amd64_gcc481
  - Parallel development with Multi-Threaded build and TTB
  - slc5_amd64_gcc481 and slc6_amd64_gcc481

FullSim Activities: Improving CMSSW

- Profiling of Geant4 9.6 to establish requirements for CPU performance of Geant4 10.0
- Results of profiling for 8 and 14 TeV are similar
  - Except CMS MC truth handling, which grows to 15% of the total time
    - Subject for further optimization
- Standard math library takes up 20% of total CPU
  - std::exp, std::log, std::log10
- The most busy Geant4 methods:

method	% of total time
CrossSectiionDataStore::GetCrossSection	14.0
G4ClassicalRK4::DumbStepper	6.5
G4Mag_UsualEqRhs::EvaluateRhsGivenB	3.3
G4BGGNucleonInelasticXS::CoulombFactor	3.0
G4PolyconeSide::DistanceAway	2.5
G4UrbanMscModel95::ComputeCrossSectionperAtom	2.4

What can be done for FullSim in CMSSW?

- GFlash parameterizations of EM and hadronic showers
  - Used in FastSim (see talk by A.Giammanco)
  - Used for forward calorimeters
  - Parametrisation depends on geometry
  - In CMSSW this can be applied on top of any Physics List
- Gflash can be enabled for high energy e^- , p > 1 GeV/c
- Eta regions between the barrel and the endcap calorimeters are excluded
- Overall CPU effect is about 10%
- Under consideration

Russian Roulette

Well established method for neutron transport :

Lewis, E. E.&Miller, Jr.,W. F. [1984]. Computational Methods of Neutron Transport, John Wiley & Sons, New York.

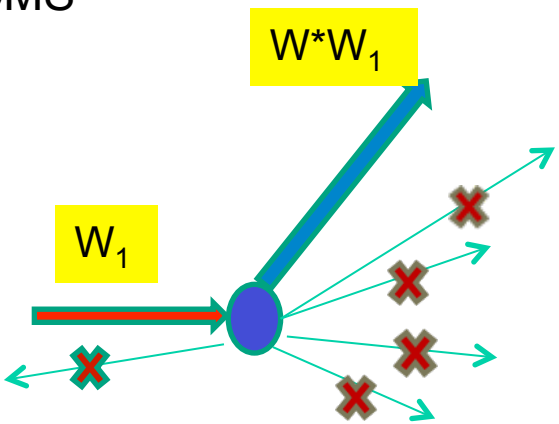
Stephen A. Dupree, S. K. Fraley [2004] A Monte Carlo Primer: A Practical Approach to Radiation Transport, Volume 2

http://www.oecd-nea.org/tools/abstract/detail/nea-0387/

http://www.oecd-nea.org/tools/abstract/detail/ccc-0754/

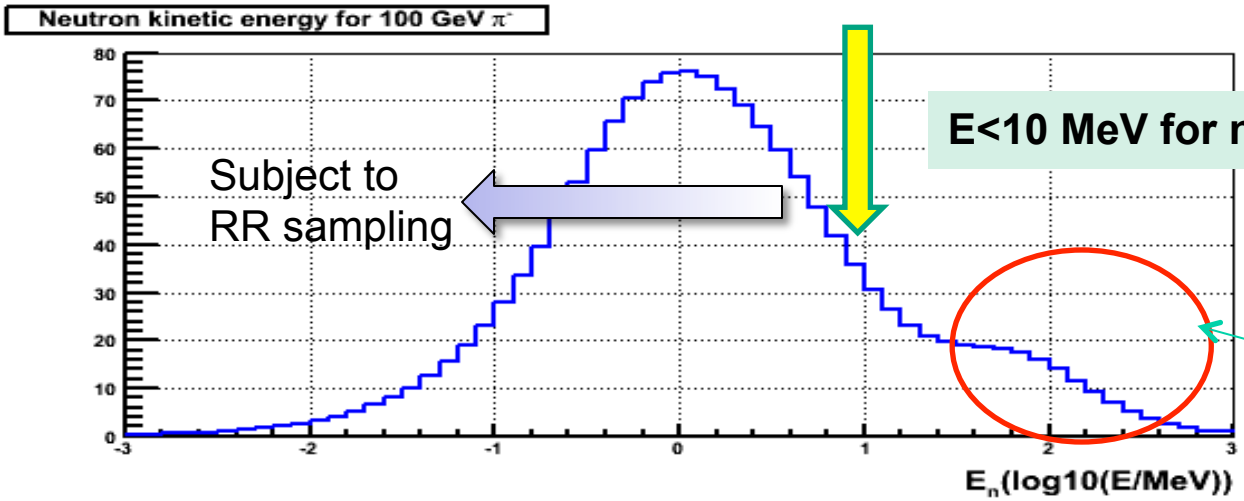
Russian Roulette (RR) in CMSSW

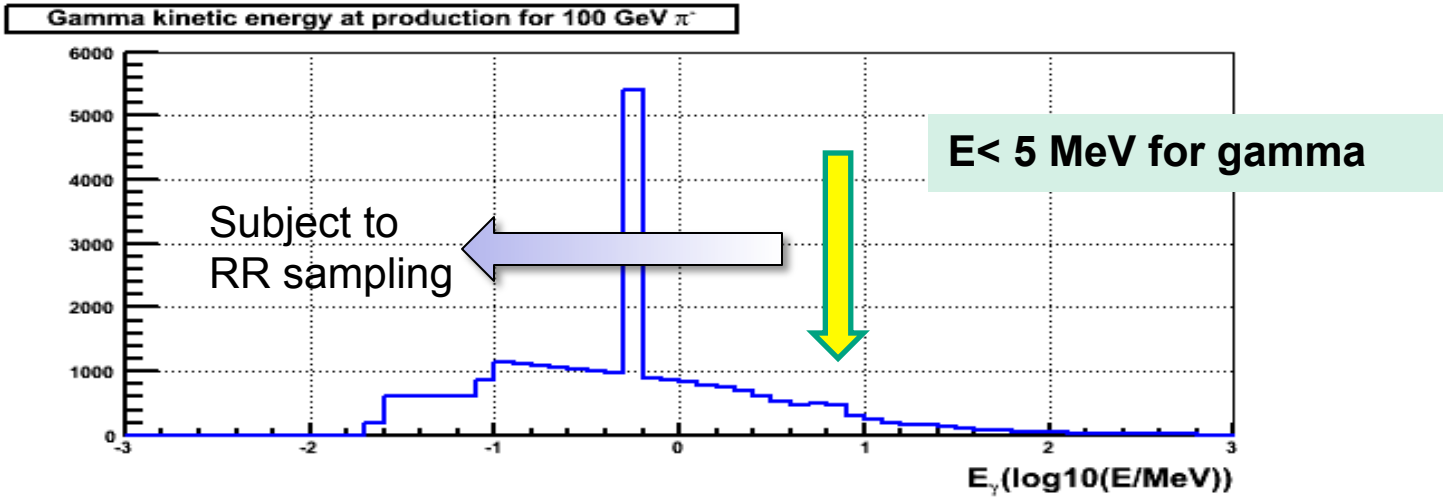
- Method used in neutron shielding calculations for many years
  - Not necessary to track all low-energy particles in a shower
- Some fraction of low-energy particles are killed but remainder get higher weight
  - not suited for tracker, muon systems
  - direct CPU savings (for calorimeter simulation)
  - geometry independent
- RR may be enabled separately per particle type and detector region
  - n, γ - allow significant CPU savings for CMS
  - p, e^- - no visible effect so far
- Two parameters per particle
  - RR factor (1/W)
  - Upper energy limit



RR: Selection of Energy limits

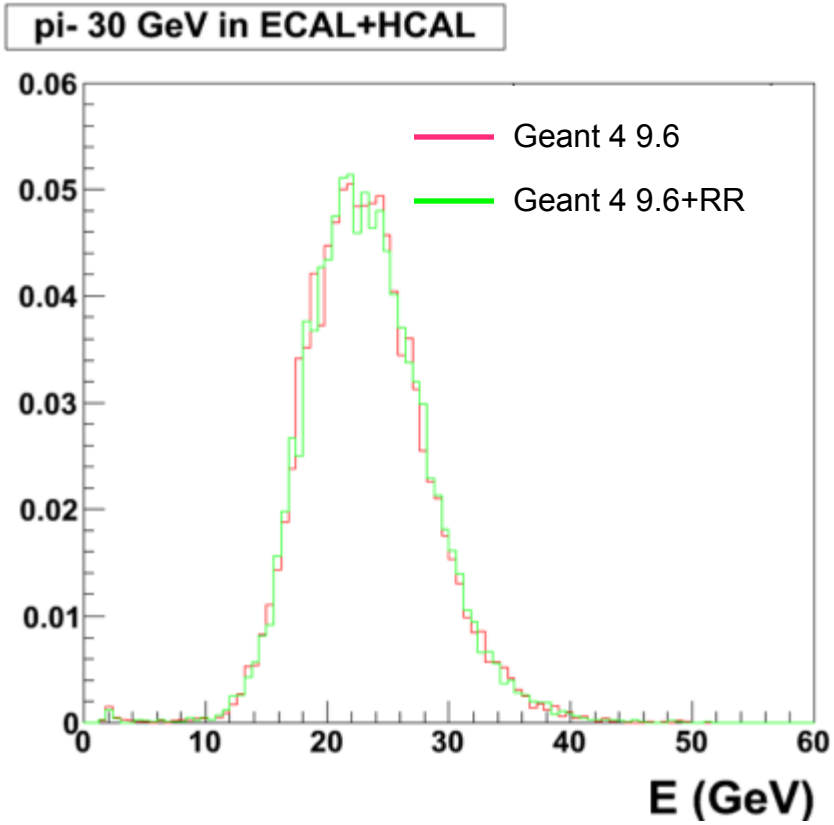
- Using standalone test geometry, find regions where there is a distinction between shower production and individual high-energy particles:

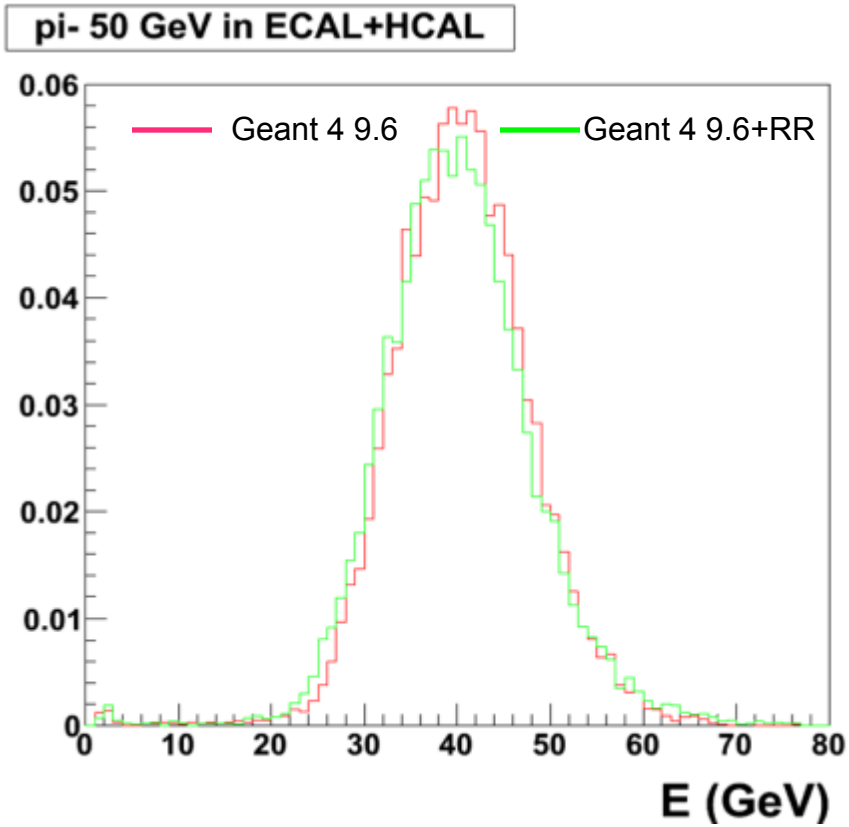




RR: pion beam in simplified Ecal+Hcal

- Checks on total energy, resolutions

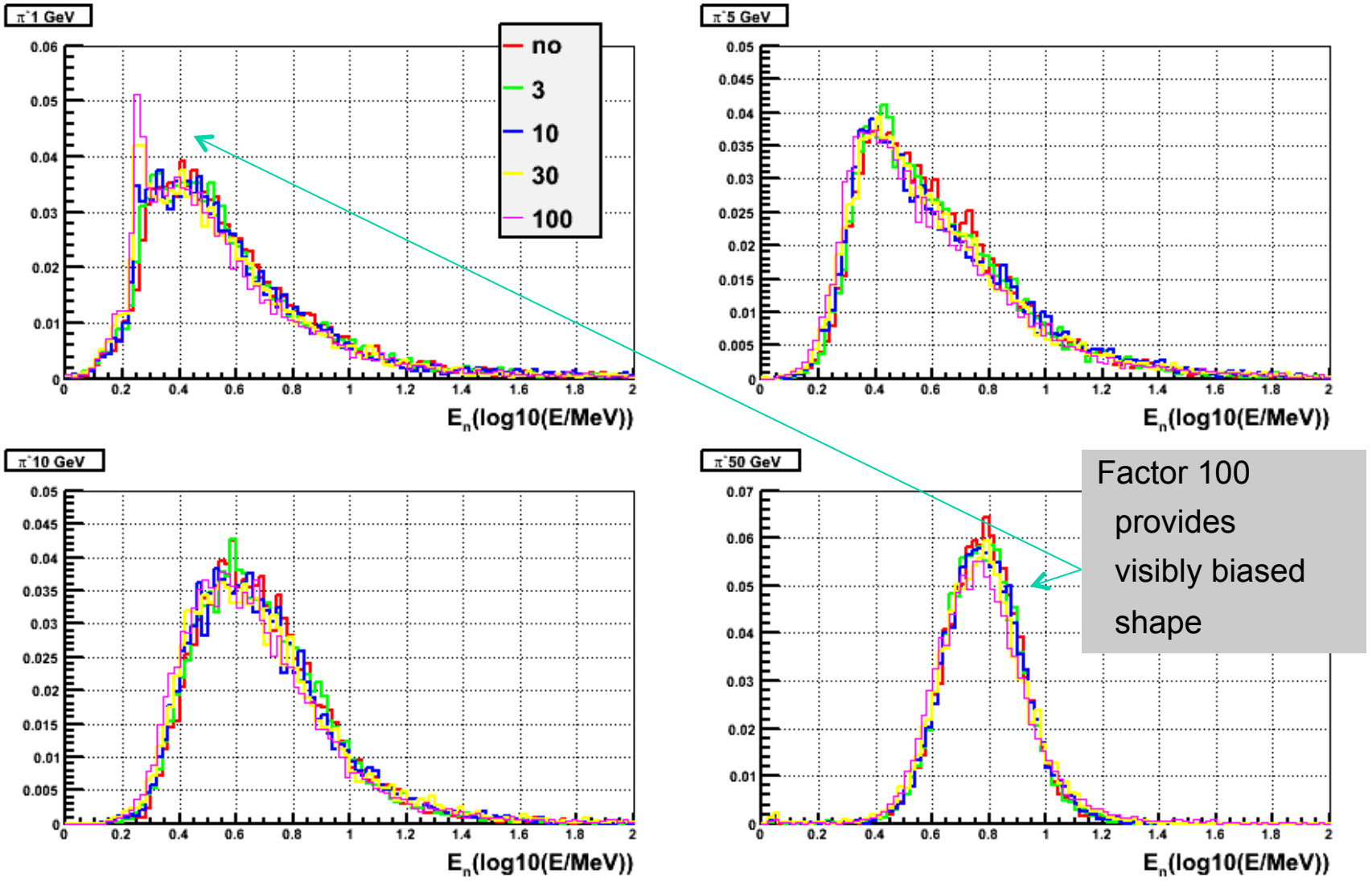




Electrons and γs are subject to RR below 5 MeV - 6% CPU saving

Neutrons are subject to RR below 10 MeV - 40% CPU saving

RR: Comparison of Reconstructed Energy



(using standalone test geometry)

Russian Roulette CPU Usage

- Comparison of CPU time/event for different event types, gradually increasing the fraction of low energy neutrons and photons that are killed

RR Factor	no	3	10	30	100
MinBias	24.8s	21.8s 88%	20.0s 81%	18.7s 75%	18.3.s 74%
Z→e+e-	57.9s	44.7s 77%	39.6s 68%	37.2s 64%	37.3s 64%
ttbar	92.1s	70.9s 77%	57.6s 63%	55.5s 60%	55.6s 60%

← % of default

RR factor 30 seems to be the upper limit – no sense to increase beyond this

CMS software version CMSSW_6_2_0_pre5

Comments on Russian Roulette

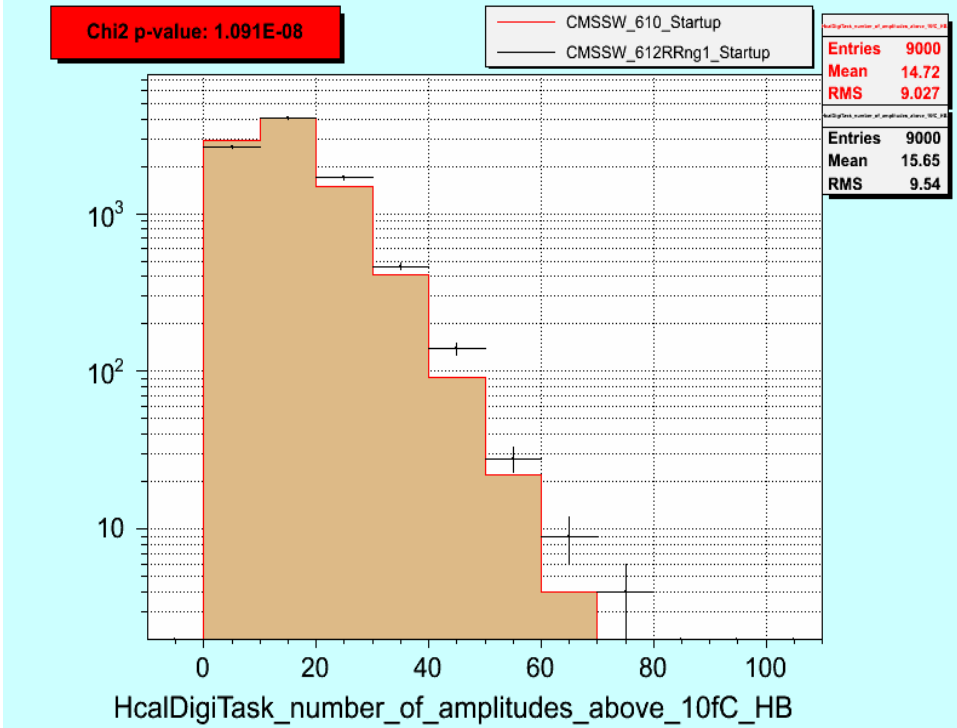
- RR is a biasing method which by itself cannot improve the accuracy of the simulation
  - Studies underway to understand if this simplification is acceptable
- Possible issues:
  - the number of particles considered in each interaction is reduced reduction of number of particles and
  - Energy of each hit is the increased by the weight factor

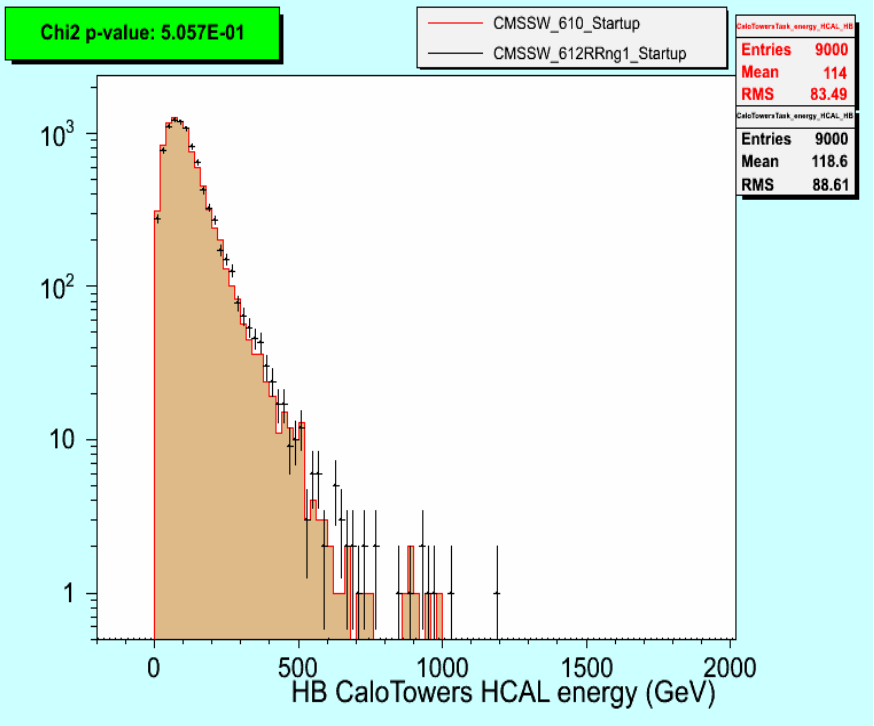
  ⇒ hit distributions may be modified

  ⇒ extra smearing is introduced
    - Mean energy deposition should be ok
    - but width of calorimeter response is increased
      - RR may still be ok if this width is significantly smaller than the intrinsic calorimeter resolution
      - The most sensitive place – ECAL barrel
- Pre-production test samples made to study these effects

Validation: HCAL Response using RR

J.Dittman, S.Abdoullin





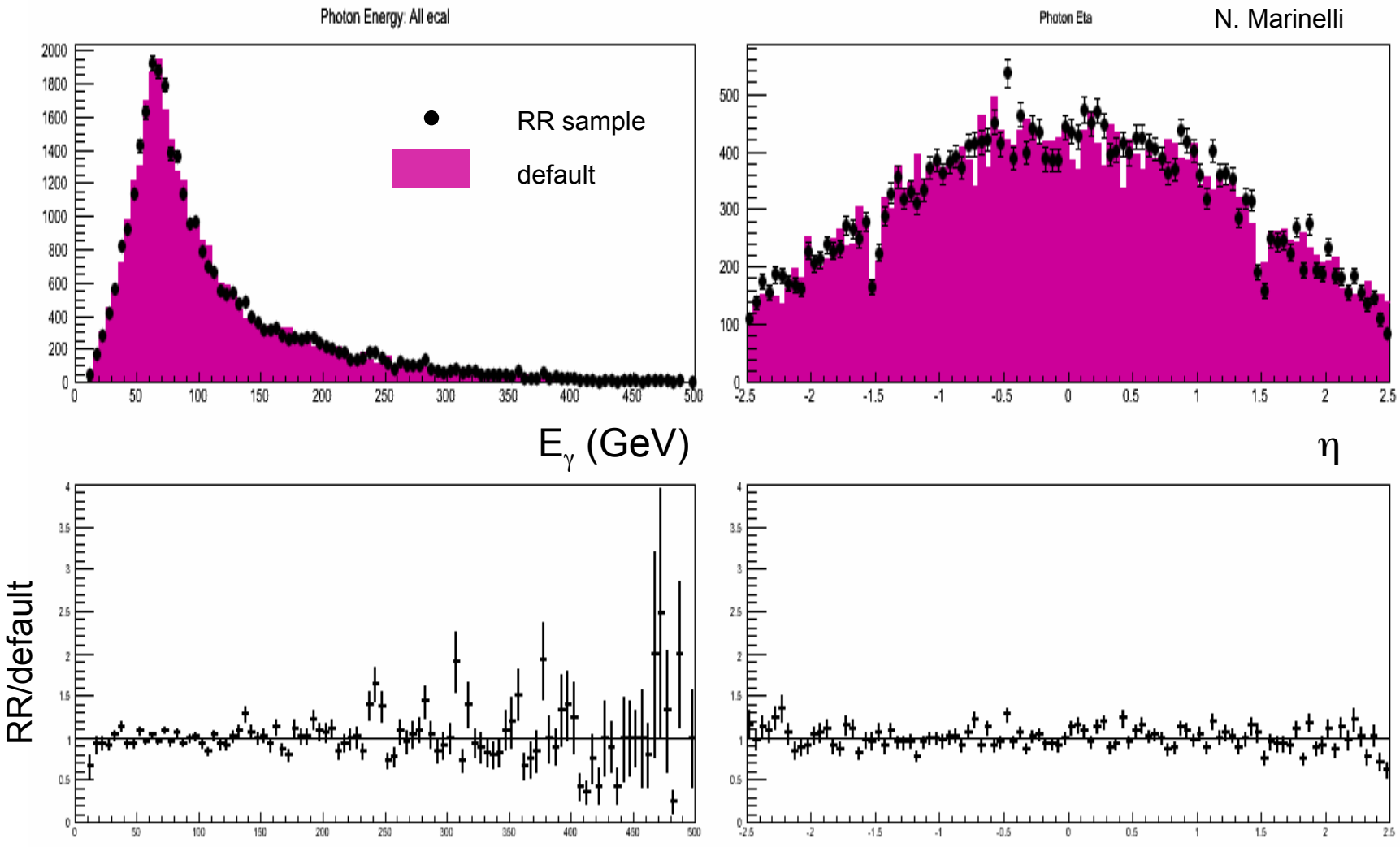
channels with large signal amplitude

reconstructed tower energy

Initial problem of RR method – outliers – solved in this case

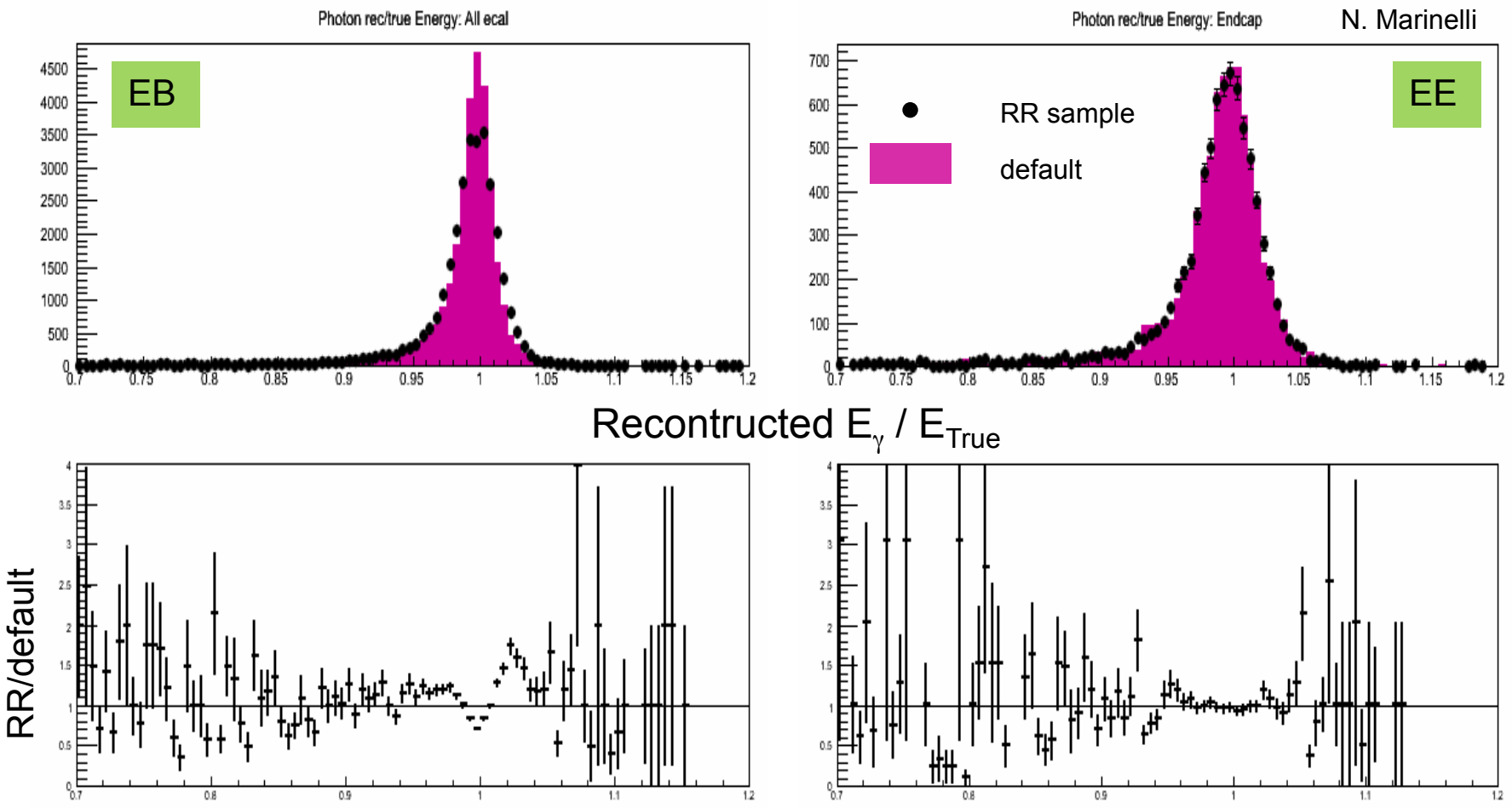
Similar high energy deposition events in HCAL towers and in “standard” MC Samples

Validation: Photon Energies with RR



No visible difference

Validation: Photon Energy Resolution



RR method may introduce some extra smearing which is seen only in ECAL Barrel due to high resolution – RR parameters need to be tuned

Russian Roulette CPU Usage

- Comparison of CPU performance between 8 TeV and 14 TeV Simulation:

Events	Energy (TeV)	No RR	RR=10	Energy (TeV)	No RR	RR=10
MinBias	8	19.3s	15.2s 78.5%	14	21.5s	16.1.s 74.2%
Z→e+e-	8	50.9s	33.4s 65.6%	14	116.9s	92.3s 78.7%
ttbar	8	87.1s	52.8s 60.6%	14	115.8s	74.3s 62.4%

% of default

Only n and γ are biased in ECAL and HCAL; RR Factor 10 is used

At 14 TeV, Zee becomes compatible in CPU with TTbar. Similar RR effects on timing

CMS software version CMSSW_6_2_0_pre6

Summary

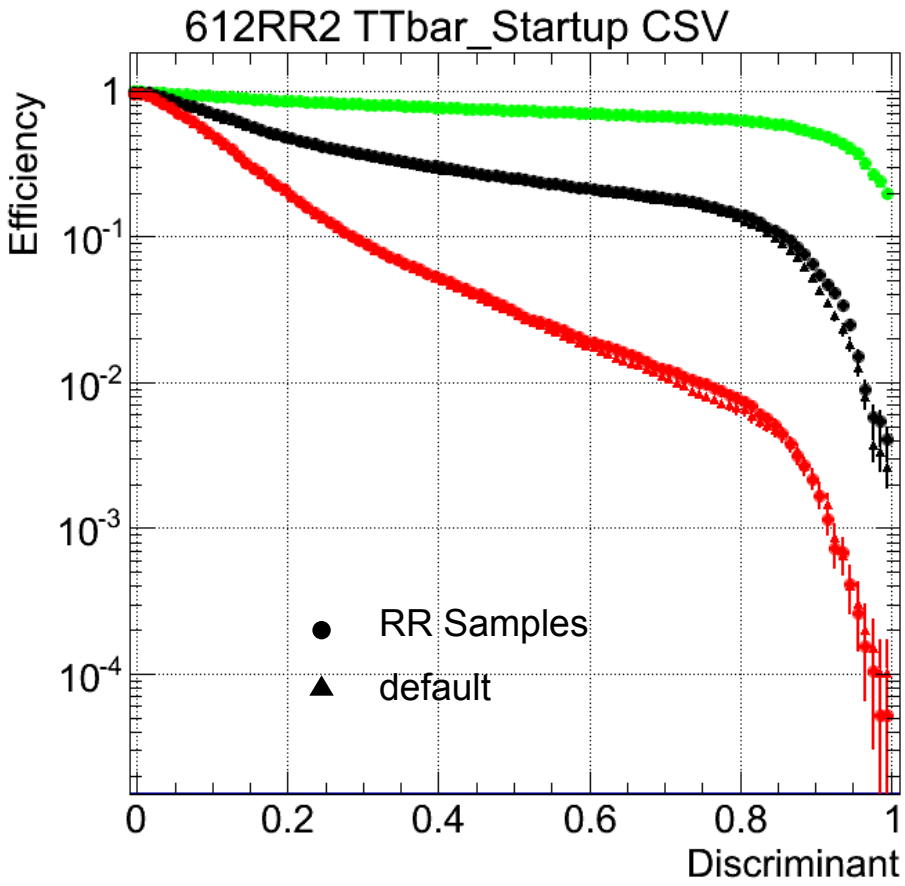
- Full Simulation is still the main workflow for MC production in CMS
  - almost 10B MC events produced for LHC Run I
- CMSSW plan for 13 Tev run includes adoptation of Geant4 10.0
  - “early” adoption so that we have time for complicated integration
  - Multithreading capabilities
  - Performance improvements in sequential mode
- Optimisation of CMS code continues
  - extensive code profiling ongoing
    - math libraries and other bottlenecks identified
  - Gflash still a possibility for speed-up of calorimeter simulation
  - Russian Roulette looks quite viable

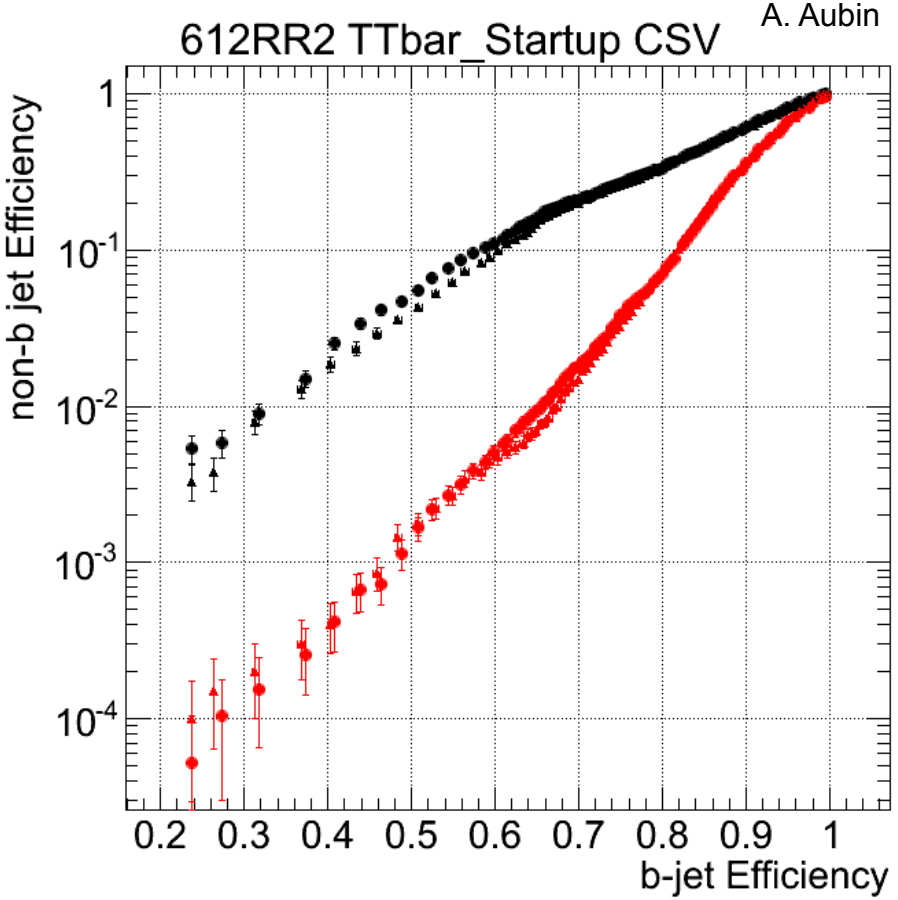
BACKUP: POTENTIAL PLOTS

B-tagging validation

If RR is used corrections to high level objects should be retuned

A. Aubin





Full Sim profiling for 8 TeV

	Functions (Events)	MinBias Cummulative / Self cpu-time (%)	ZEE Cummulative / Self cpu-time (%)	TTBar Cummulative / Self cpu-time (%)
1	G4Mag_UsualEqRhs:: EvaluateRhsGivenB	3.29 / 3.29	3.32 / 3.32	3.33 / 3.33
2	G4PolyconeSide:: DistanceAway	2.66 / 2.34	2.72 / 2.30	2.55 / 2.02
3	G4Navigator:: LocateGlobalPointAndSetup	5.39 / 1.84	5.66 / 2.00	6.21 / 2.29
4	SimTrackManager:: idSavedTrack	1.20 / 1.20	4.03 / 4.03	1.06 / 1.06
5	G4CrossSectionDataStore:: GetCrossSection	15.06 / 1.54	13.70 / 1.60	14.14 / 1.61
6	G4UrbanMscModel95:: ComputeCrossSectionPerAtom	2.84 / 1.70	2.18 / 1.30	1.80 / 1.09
7	G4ClassicalRK4:: DumbStepper	6.65 / 1.55	6.72 / 1.60	6.33 / 1.58
8	G4hPairProductionModel:: ComputeDMicroscopicCrossSection	1.01 / 0.22	0.37 / 0.08	0.20 / 0.05
9	G4PhysicsLogVector:: FindBinLocation	0.62 / 0.11	0.79 / 0.14	0.93 / 0.17
10	G4BGGNucleonInelasticXS:: CoulombFactor	3.43 / 0.53	2.84 / 0.44	3.10 / 0.47
11	G4ElasticHadrNucleusHE:: HadrNucDifferCrSec	1.34 / 0.26	0.63 / 0.12	0.47 / 0.09
12	G4InuclSpecialFunctions:: G4cbrt	0.88 / 0.04	0.72 / 0.03	0.76 / 0.04

The 12 most time consuming methods:

- **1 from CMSSW**
- **11 from Geant4**

Full Sim profiling for 13 TeV

	Functions	Cumulative / Self time spent (%) (MinBias13TeV)	Cumulative / Self time spent (%) (MinBias13TeV + RR)	Cumulative / Self time spent (%) (TTBar13TeV)	Cumulative / Self time spent (%) (TTbar13TeV + RR)	Short Comments
1	SimTrackManager:: idSavedTrack	18.62 / 18.62	16.65 / 16.65	10.08 / 10.08	10.32 / 10.32	The most consuming function. Recursive.
2	G4Mag_UsualEqRhs:: EvaluateRhsGivenB	2.38 / 2.38	2.58 / 2.58	2.82 / 2.82	2.88 / 2.88	-
3	G4PolyconeSide:: DistanceAway	2.41 / 2.19	2.38 / 2.18	2.44 / 2.01	2.21 / 1.85	About 0.3% of cumulative is used by atan2.
4	G4Navigator:: LocateGlobalPointAndSetup	6.67 / 1.76	6.70 / 1.80	6.63 / 2.20	2.27 / 2.12	Calls several functions
5	G4ClassicalRK4:: DumbStepper	5.24 / 1.17	5.64 / 1.25	5.89 / 1.46	5.88 / 1.38	sim::Field:: GetFieldValue and function #2 of the table are called the most.
6	G4ElectroNuclearCrossSection:: GetIsoCrossSection	1.33 / 1.10	1..45 / 1.20	1.49 / 1.23	1.57 / 1.29	About 0.25% of cumulative spent in log

*(clicking the numbers redirects to web profile)

The 6 most time consuming methods:

- **1 from CMSSW (but significantly increased from 2 to 15 %)**
- **5 from Geant4**

CPU profiles for 8 and 13 TeV are similar for Geant4 functions

Requirements for Geant4 10.0

- CPU hot spots in Geant4 for CMS are in simple functions:
  - G4PhysicsVector
  - Cmath library
  - Hadronic cross sections
  - G4Polycone
- Geant4 Collaboration accepts the following requirements
  - Introduce G4Log and G4Exp
    - Extracted from VDT library (see talk by D.Piparo)
  - Introduce G4Pow
  - G4PhysicsVector under review

	Functions	Cumulative time spent (%) (MinBias13TeV)	Cumulative time spent (%) (TTbar13TeV)
1	log	2.55	2.65
2	exp	3.08	3.26
3	atan2	2.85	2.87
4	log10	1.30	1.51
5	pow	0.89	0.96
6	sincos	0.54	0.54
7	atan2f	0.23	0.26
8	atanf	0.11	0.12
9	cos	0.10	0.09
+	TOTAL	11.65 %	12.26 %

*(clicking the numbers redirects to web profile)