Geant4 regression tests on the GRID

J. Apostolakis, I. McLaren,
P. Mendez Lorenzo, A. Ribon
CERN

Outline

- Motivation and strategy
- Description of the suite
- Recent extensions
- Some results for the June 2006 Geant4 release (G4 8.1)
- GRID
- Outlook

Motivation and strategy

Our goal is to detect any difference between two Geant4 versions, by comparing a large number of physics observables.

Only those distributions which are statistically different (i.e. unlikely to be originated from the same parent distribution, whatever it is) should be visually examined by someone, to understand the origin of such difference, either a bug or an improvement.

As a start, driven by the needs of LHC experiments, we consider only calorimeter observables sensitive to hadronic physics.

Some remarks

- It is very important that the regression testing is **automatic** , because only in this way it is possible, in practice, to consider a very large number of distributions, and use it regularly, at least before each new release.

- To do that we need a statistical package that offers several statistical tests (not only Chi2 and Kolmogorov-Smirnov!), to increase the chances to spot any kind of discrepancy between two distributions.

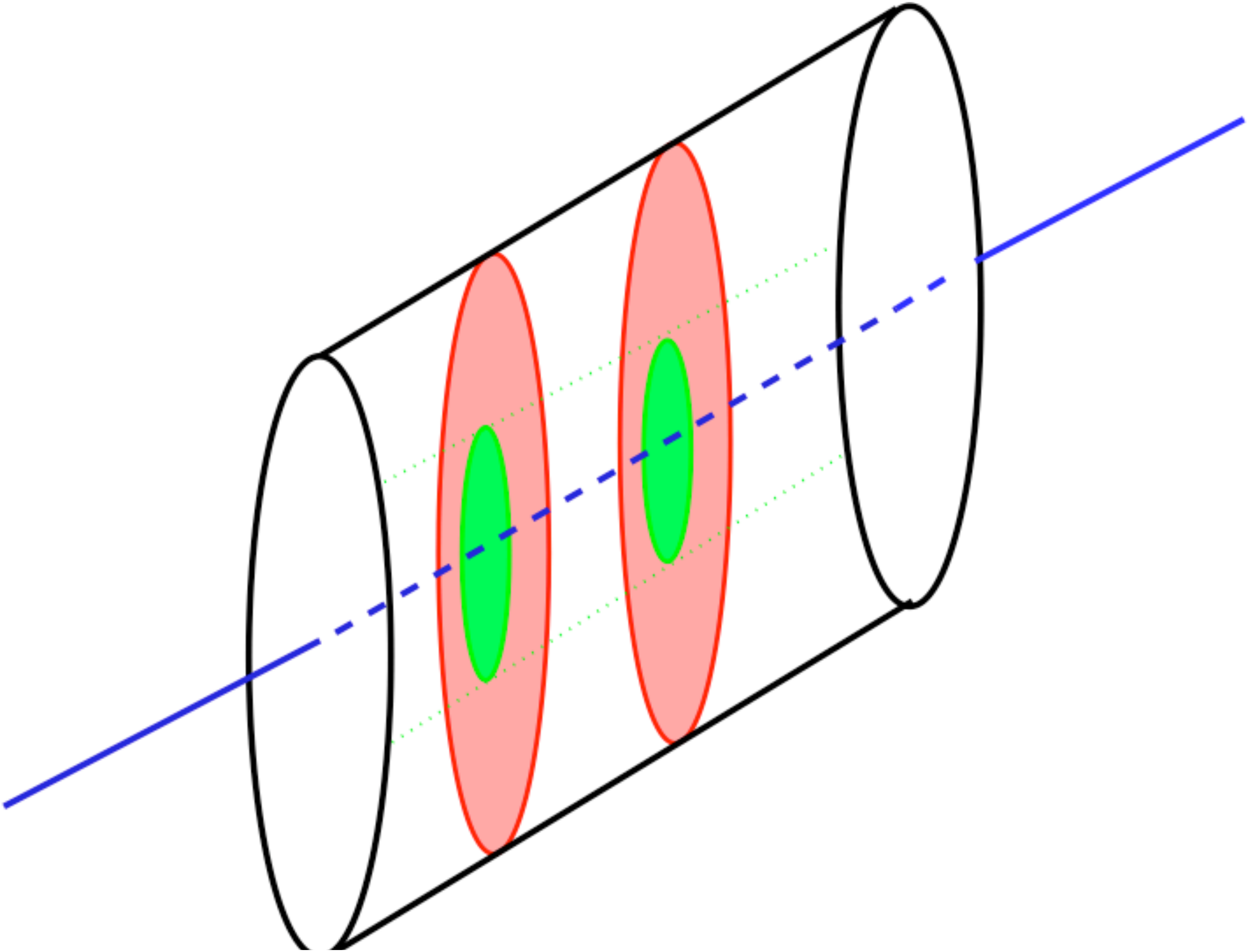
Simplified Calorimeter setup

- It reproduces, in a simplified way, all the LHC calorimeters: Fe-Sci , Cu-Sci , Cu-LAr , W-LAr Pb-Sci , Pb-LAr , PbWO4 .

- Beam particle type: $\pi^{\pm}$, $K^{\pm}$, K_L^0 , p , n , e^- .

- Beam energy: 1, 2, 3, …, 10, 20, 30, 40, 50, 60, 80, 100, 120, 150, 180, 200, 250, 300 (1000) GeV.

- The calorimeter is a tube. The user can choose:
  - the total thickness of the absorber (in [mm] or λ)
  - the radius of the tube (in [mm] or λ)
  - the thickness of the active layer
  - the number of layers
  - the number and the size (in [mm] or λ) of the rings for the lateral shower profile.



A

- **Observables**:
  - total energy deposit in all active layers
  - total energy deposit in the whole calorimeter
  - energy deposit in each active layer (longitudinal shower profile)
  - energy deposit in each ring (i.e. radial bin) (lateral shower profile).

- The program produces in output a HBOOK n-tuple, which stores all the above distributions. Then another program reads two of these n-tuples, makes the statistical tests of the distributions, and produces a **.PS** file whenever the p-value of any of our statistical test is below a certain threshold (currently **1%**).

New observables (I)

- The same setup is useful for studying the impact of physics models on hadronic shower observables.

- Recently, to investigate the Geant4 shower shapes we start monitoring many other variables:
  - average number of steps and tracks per event;
  - average track and step length per event;
  - average number and Ekin of exiting tracks;
  - kinetic energy spectra of tracks entering some active layers;

  each of these is done for different particle types and also for all particle tracks. Finally:
  - contributions to the visible energy and shower shapes for different particle types.

New observables (II)

- When the changes of the physics models are significant, the number of .PS plots produced can be huge. It would take too long to look at them all ! We need a quicker, better and more global way to monitor which observables are affected and by how much.

- We use the new quantities, and their estimated statistical uncertainties, to check for significant changes (e.g. 5 σ or 5%). For the longitudinal shower shapes, we consider 4 fractions; for the lateral shower shapes, 3 fractions.

- We run several cases at 100 GeV without biasing, and only for these we consider the spectra.

CPU issue

- 7 calorimeters x 8 particles x 23 beam energies x 5000 events x 5 Geant4 physics configurations.
- About $\geq$ 0.07 sec/GeV (@1 GHz CPU) to simulate a hadronic shower.
- For higher energies we use biasing techniques to speed up the execution
  - keep 0.25 to 0.5 of $e^-/e^+/\gamma/n$ created with weight 2 to 4;
  - kill neutrons below 1 MeV.
- Overall, it takes 4-6 years of CPU time, but concentrated in 1-2 weeks: need of the **GRID** !

Geant4 release 8.1 (June 2006)

- Checked G4 8.1 candidate tags versus G4 8.0.p01 .

- 10 lambdas, 100 layers (20 readout), 4 mm active layers, 10 rings of increasing size, starting from 0.1 lambdas; biasing for beam energies ≥ 10 GeV .

- SLC 3.0.x g++ 3.2.3 .
  Run on the GRID as **Geant4 Virtual Organization**.
  ≈ 100 CPUs in 7 sites in 7 countries.

- ≈1300 jobs per each Geant4 physics configuration.
  Each job that ends normally produced a tar-ball of about 3 MB size containing the results.

Testing steps for G4 8.1

1. Test recent fixes.

2. Replace stopping process with improved based on CHIPS.

3. Replace the hadronic elastic scattering with a new process
   - that improves low energy behaviour
   - that has new cross-sections for H target.

4. Activate strange-hadrons in Bertini cascade.

5. Activate light-ions in Binary cascade.

6. Check options for faster electromagnetic physics.

GRID (1)

- This was the first production in which Geant4 is a fully official EGEE Virtual Organization, and many sites kindly allowed us to use their resources.

- Some sites are Tier1, the rest are Tier2; some have dedicated CPUs for Geant4 (e.g. CERN, Madrid); some granted higher priority to Geant4 in some CPUs (e.g. SARA, Weizmann); others let Geant4 to compete for resources with other VOs with the same priority (e.g. Italy).

- We were allowed to run, for the first time, on the US Grid (OSG), in 4 sites. This is interesting also for checking the EGEE-OSG compatibility…

GRID (2)

- In spite of a large pool of sites available for Geant4, we were experiencing configuration problems and due to time constraints we were not able to include all of them.
  For the next releases, we are planning to start checking and debugging all the sites available at least one month before the production, so to try to include all of them !

- G4 VO has 1 dedicated Resource Broker (RB) at CERN, and it has access to another 2 (shared); few external RBs are also available but not yet used.

- G4 VO has 1 dedicated, centralized File Catalogue at CERN.

GRID (3)

- A major improvement with respect to previous productions has been achieved:
  - we do only 1 GRID installation of all the software which is stable during the production (e.g. reference version + needed libraries CLHEP, PI, GSL);
  - we copy the candidate version and the application from a Storage Element to the worker node, for each job.

  This allows more flexibility, at the price of a CPU overhead at the beginning of each job (in the future we'll use /**tmp**/ to be more efficient).

- Most of the production has still been done with Patricia's simple framework, but in about 10% of of the jobs we have used GANGA/DIANE which will be our main tool for future productions.

GANGA

- User Interface to the GRID
  designed for ATLAS and LHCb

- Transparent access to multiple backends:
  LCG, LSF, Condor, etc.

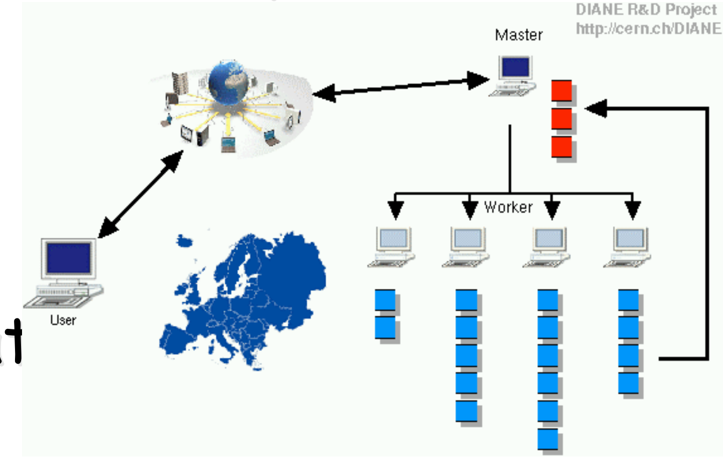
- Keep history of user jobs

- Automatic monitoring of the status of the jobs

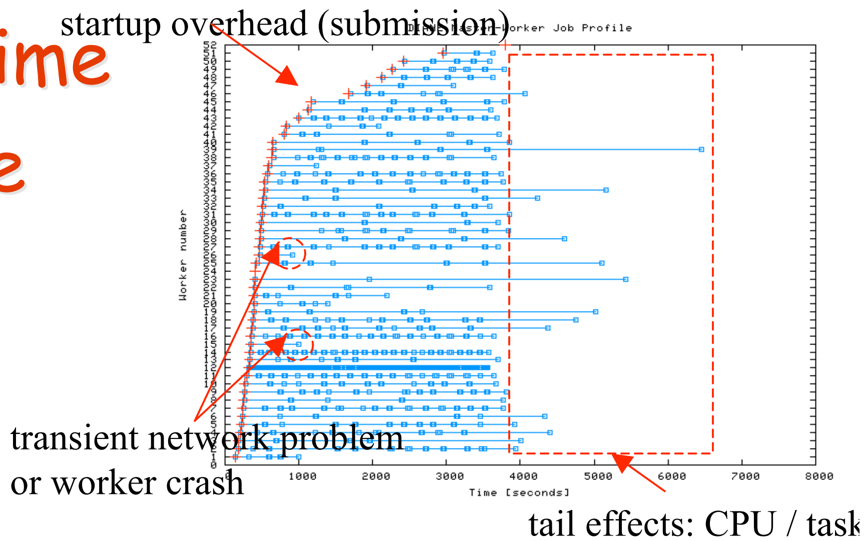
- 3 kinds of interfaces available:
  - programmatic interface (scripts)
  - command line
  - GUI

DIANE : DIstributed ANalysis Environment

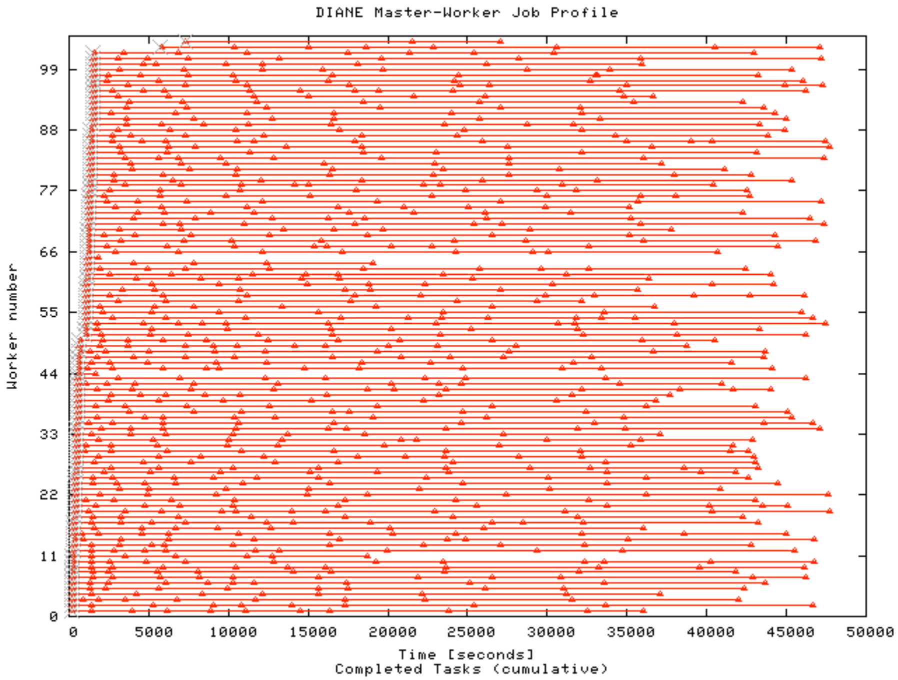
- dynamic user-level scheduler in and beyond master/worker model
- application-independent
  - Atlas Athena AOD analysis
  - BLAST, Autodock: bioinformat
  - ITU frequency planning

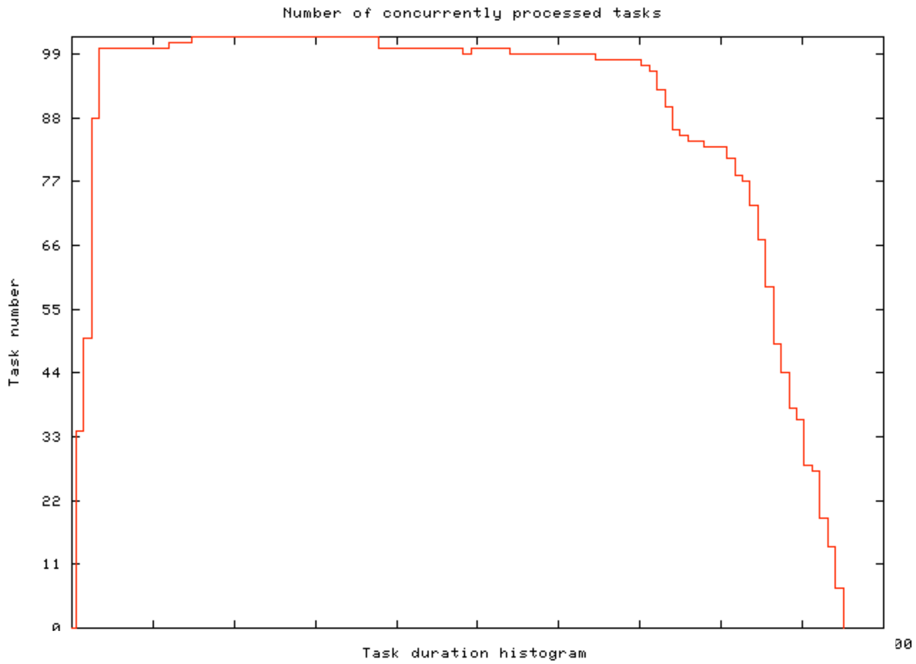


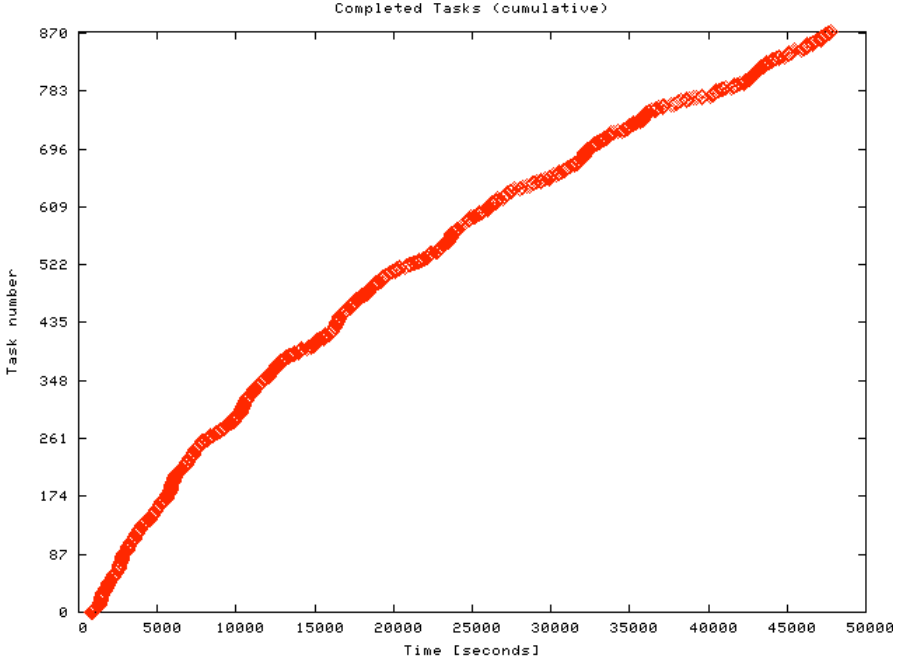
- highlights
  - reduced job turnaround time
  - stable job completion rate
  - efficient error recovery

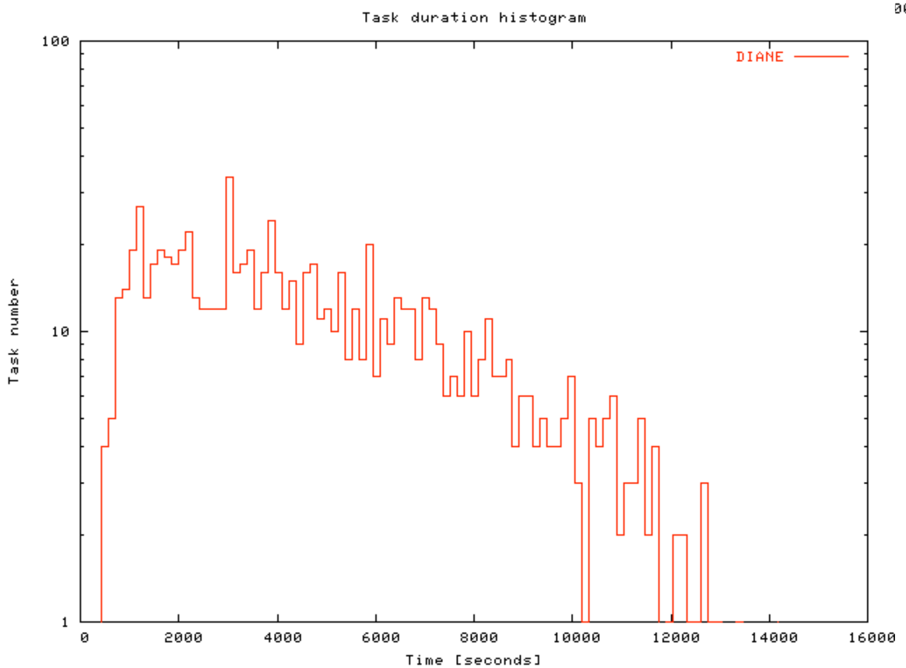


GANGA-DIANE used for 8.1



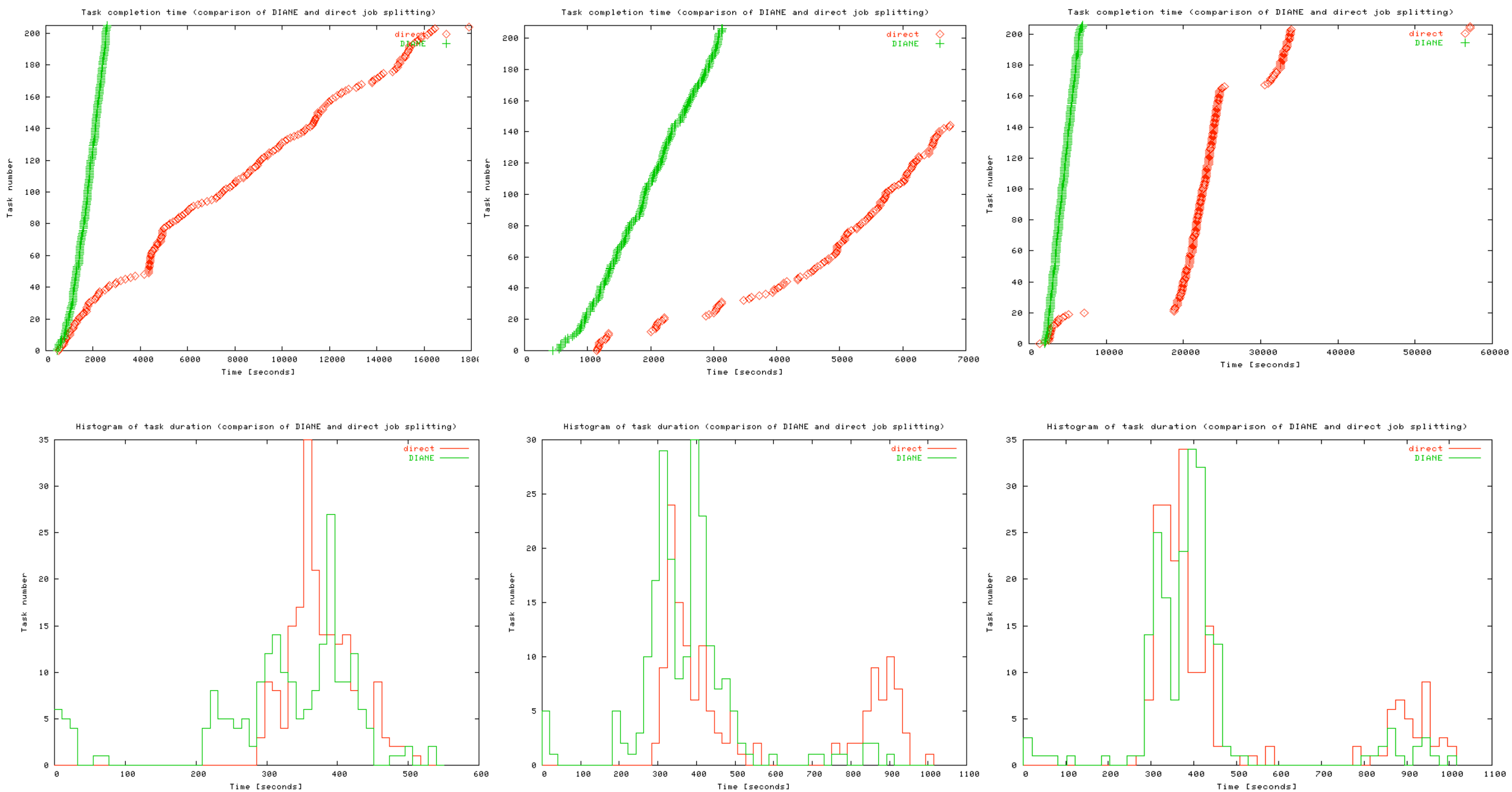






Comparison: simple framework - DIANE

G4 Production 6 sites / 173 CPU ; 207 tasks: 1 task = 1 job



A. Ribon, 13-Oct-2006

Results (1)

- 7 candidate G4 versions tested
- ≈ 20,000 jobs submitted to the GRID
- ≈ 99.9% GRID success rate
  (by a tight selection of the sites)
- a majority of jobs result were analysed
  - earlier results already identified G4 problems (and fixes have been proptly created);
  - sometimes lack of AFS disk space and/or human time.
- ≈ 80,000 PostScript files produced
  of which about 1% were visually examined, less than for 8.0, as summary numbers are now used in their place.

Results (2)

- 6 Geant4 physics configurations tested
- 3 bugs found
- 1 critical issue identified (new elastic scattering), which affects shower shapes, producing narrower and a bit longer showers: this has been finally understood after the release...
- 1 CPU issue identified (+10% due to the new stopping).

Outlook

We have a tool that is useful for robustness testing, regression testing, and for physics studies. We aim to improve and extend this tool in the following directions:

- Include more statistical tests , and study their statistical power.
- Add other ways to spot quickly and effectively changes (use moments; use visual ways to summarize results, etc.).
- For the GRID production, debug the sites earlier, and deploy the GANGA/DIANE framework to use the resources more efficiently.