Computer Science • Vol. 9 • 2008

Anna Kaczmarska*, Paweł Malecki*, Tadeusz Szymocha*,
Elżbieta Richter-Wąs**

THE ATLAS EXPERIMENT SIMULATIONS AS THE COMPUTING CHALLENGE FOR THE ACK CYFRONET AGH

The present High Energy Physics (HEP) experiments require unprecedented amount of computing power and storage space. We present the WLCG structure of the LHC computing, which will be used to perform data processing required by the ATLAS collaboration. We also describe the ATLAS detector full simulation chain. Finally, we quantify the computing needs and up-to-date usage of the ACK CYFRONET AGH resources by the ATLAS detector simulations.

Keywords: *High Energy Physics, HEP, ATLAS, tau lepton physics*

SYMULACJE EKSPERYMENTU ATLAS JAKO WYZWANIE OBLICZENIOWE DLA ACK CYFRONET AGH

Obecne eksperymenty fizyki wysokich energii wymagają niespotykanej mocy obliczeniowej i przestrzeni do przechowywania danych. Prezentujemy strukturę WLCG obliczeń LHC, która będzie użyta do przetwarzania danych na potrzeby współpracy ATLAS. Opisujemy również pełny łańcuch symulacji detektora ATLAS. Na zakończenie, przedstawiamy potrzeby obliczeniowe i aktualne użycie zasobów ACK CYFRONET AGH przez symulacje detektora ATLAS.

Słowa kluczowe: *fizyka wysokich energii, HEP, ATLAS, fizyka leptonów tau*

1. Introduction

The Large Hadron Collider (LHC) is a new paticles accelerator built at CERN, Geneva, Switzerland [1]. It will accelerate and collide proton beams at center of mass energies of 14 TeV/c^2. These collisions are expected to reveal fundamental new processes in particle physics. At points where particles beams cross each other, there are four experimental installations prepared by international collaborations.

---

* Institute of Nuclear Physics Polish Academy of Sciences, Krakow, Poland
`Tadeusz.Szymocha@ifj.edu.pl`
** Institute of Nuclear Physics Polish Academy of Sciences, Krakow, Poland, and Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krakow, Poland

One of these experiments, is the ATLAS (A Toroidal LHC ApparatuS) [2] detector. Its primary physics goal is to search for the Higgs boson – a very important ingredient in understanding the origin of particle masses. In addition it will provide a rich environment to study many aspects of particle physics such as a Charge-Parity (CP) violation in B meson decays, top quark physics and the Quantum Chromodynamics (QCD). Searches for new physics beyond the Standard Model such as supersymmetry and extra dimensions will also be a major focus. The ATLAS collaboration consists from ca. 2000 people from 37 countries and 159 institutes and universities (March 2006), including the Institute of Nuclear Physics Polish Academy of Sciences (IFJ PAN) [3] and the AGH University of Science and Technology [4], where ca. 40 persons are involved in hardware construction and software development.

The information gathered from the detector is organized in so called physics events corresponding to proton-proton collisions. At design intensity these collisions will occur at rate of 1 GHz. The events are then evaluated whether they contain new, interesting physics phenomena. This task is very demanding, therefore the ATLAS experiment is faced with number of challenges. We will investigate very rare processes, and in some cases the preselection system has to find one interesting event out of 10^{11} (see K. Korcyl contribution to this conference). We expect enormous volumes of data, thus ATLAS is also facing a challenge to record and store yearly about 5 PB of information. The ATLAS collaboration, due to its geographical spread, should provide worldwide access to data, therefore high through-put and quality of computing network is necessary. The latter two points are addressed to all computing centers, which form the LHC computing infrastructure, including the Academic Computer Center CYFRONET AGH (ACK CYFRONET AGH) [5].

The physics analysis consists of measuring various observables, related to the proton-proton collision events. It requires that apart of the event reconstruction based on the electronic signals, we need a full computer modelling, including event generation, simulation of the detector response and subsequent analysis and evaluation of acceptance, inefficiency and other corrections. Ahead of the real data taking, it is important to prepare collaboration software for particles reconstruction and identification (e.g. tau leptons) and extensively test them on Monte-Carlo data samples, containing large statistics of events (see M. Wolter contribution to this conference).

2. LHC computing infrastructure

The Worldwide LHC Computing Grid (WLCG) [6] structure is presented in Figure 1. In this scheme, the main centre is CERN (Tier0). It is anticipated that the data reconstruction, i.e. the processing of the full online data into the analysis object data and those used for event tagging, will be performed at Tier0. The Tier1 sites will also perform the reconstruction of these data and make the output available to the smaller Tier2 centers. Another function of those centers will be mirroring of the central database of Tier0. The Tier2 sites are responsible for further data processing. The Tier3 sites (physics institutes) are mainly used to perform specific physics analyses.

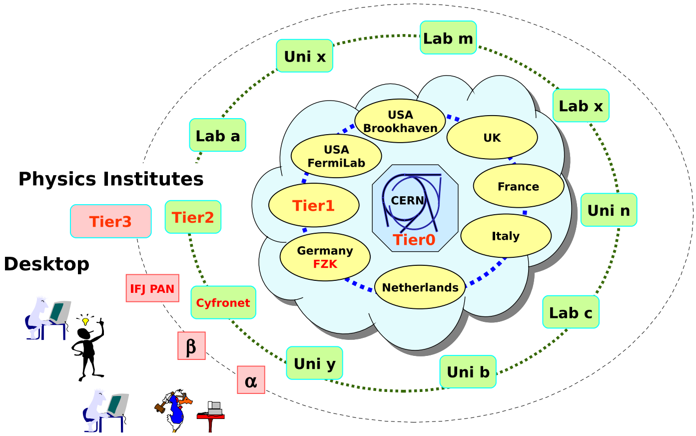

Fig. 1. The Worldwide LHC Computing Grid (WLCG) structure; all elements are connected by fast links. Detailed description in the text

Before the start of LHC, all resources are used for full computing modelling of the detector, as described in the next section. All the centers are connected through fast network. The IFJ PAN (which is Tier3 center) collaborates with the sites on each level of the distributed computing schema in access to the real data: ACK CYFRONET AGH (Tier2) and Forschungszentrum Karlsruhe (FZK–Tier1). The aim of such a distributed computing structure is to optimize the access to the data (real or simulated) and the usage of collaboration resources (central computing, data storage facilities and network bandwith) by all members of the collaboration.

As a result of participation in the EU *CrossGrid* project [7], the ACK CYFRONET AGH acquired new computational power – the *zeus* cluster („Racksaver" PCs), which has been used to generate a large number of Monte Carlo physics events for the ATLAS experiment. When the ACK CYFRONET AGH in 2004 became a Tier2 site, the *zeus* cluster was developed further. Currently about 600 processors and 58 TB of disk storage are devoted to HEP computations.

There are two ways of accessing the ACK CYFRONET AGH resources. The first one is sending grid jobs, usually by the ATLAS central production managers, but also by the users performing their own analyses. The ATLAS simulation software has evolved rapidly (several significant release upgrades within few years) leading to re-simulation of the Monte-Carlo events with different versions and validation of developed software by comparison of the overall performance between different versions of the software. The support of the ATLAS Virtual Organization has assured the availability of the ATLAS collaboration framework (*Athena* [8]) needed at all steps of the full simulation chain described in the next section.

The second method, is interactive access to the worker nodes and sending jobs to the batch system, used by local members of ATLAS IFJ PAN group responsible for the development of the algorithm for the tau leptons identification, *tau1P3P* [9]. A typical activity in that case was validation of the reconstruction software, particularly of the *tau1P3P* algorithm and the development of analysis software (checking for debugging purposes validation histograms on a smaller number of events).

3. ATLAS Full Simulation Chain

The design and construction of ATLAS requires a large amount of simulated Monte-Carlo data in order to optimise the design of the detectors, to estimate the physics performance that will be achieved, and to test the software and computing infrastructure. The ATLAS full simulation chain and corresponding chain for the real data collecting and analysis is presented in Figure 2.

On the right-hand side, starting from label *Physics*: the interactions between colliding protons, lead to creation of particles, that traverse the detector, leave signal hits in its tracking part and deposit energy in calorimeters. This information is used by the reconstruction algorithms to find and identify the particles that have passed the detector (task of Tier0 site). This is accomplished in sub-steps. These results are stored as the Event Summary Data (ESD). Further processing algorithms are used to prepare Analysis Object Data (AOD)(task of Tier1 sites), which contains reduced information from ESD. Then, from AOD data, new containers that will be used in physics analysis are created: Derived Physics Data (DPD) or CBNT (ComBined NTuple) [8] (task of Tier2 sites). The last step of the chain is the analysis, in which cuts on physics quantities (momentum, energy depositions, etc.) are set (task of Tier3 sites).

On the left-hand side of Figure 2, starting from label *Events generation*: particles produced in the proton-proton collisions are generated using a program based on physics theories and phenomenology. Several programs are available and are referred to as the Monte Carlo generators. As an output of events generation, we get a list of particles with their momenta and energy. These particles are propagated, in step called Simulation, through the detector volumes (described in a geometry model) and interactions of particles with detector material are calculated. Then the output of computing simulations is converted to the same format as we get from the real detector (digitalization). Finally, on the simulated data, the same reconstruction and analysis algorithms, which later will be used for real data, are applied.

An exemplary simulated event of the associated Higgs boson production with two bottom quarks and a Higgs boson decay is shown in Figure 3. The view perpendicular to the beam axis (marked red) of the ATLAS detector is presented. The tracking layers of the detectors are marked white. The light blue lines represent the tracks of charged particles. The next broad layers (rings) represent calorimeters and red bars represent energy losses of particles inside these detectors. The outer-most blue structures represent muon chambers, intended for precise measurements of muons momenta.

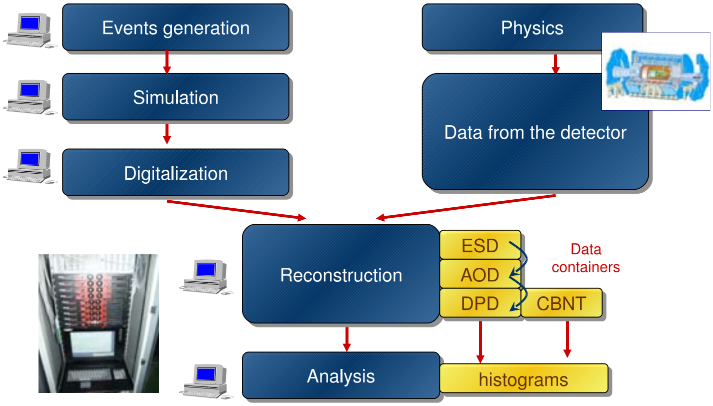

Fig. 2. The ATLAS full simulation and real data analysis chain. Detailed description in the text

The Higgs boson decays promptly, thus on this event display, we can only see final decay chain products: two τ-jets (represented by narrow blue tracks leading to green dots outside muon chambers) and two b-jets (represented by remaining broader tracks).

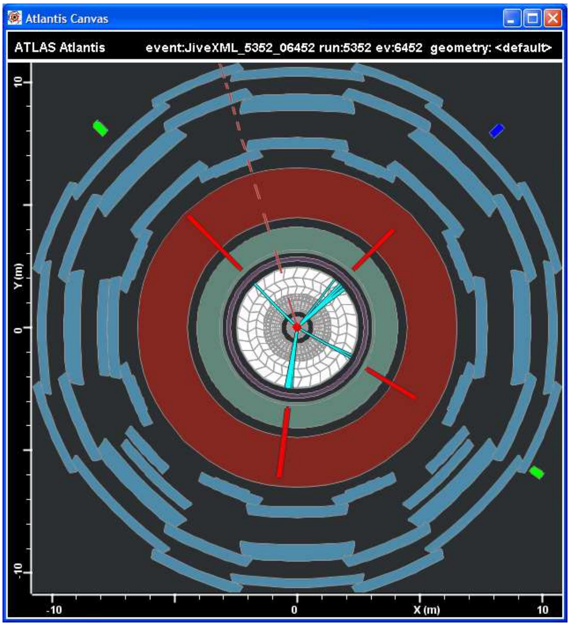

Fig. 3. The exemplary ATLAS event display (view perpendicular to the beam axis) of the associated Higgs boson production with two bottom quarks. The Higgs boson decays into tau leptons pair. Detailed description in the text

We would like to stress, that only a signal event is presented. In real data we will have a much complicated picture, with hundreds of tracks and energy deposition coming from the minimum bias, pile-up and the QCD background processes.

4. ATLAS data production

The large data volumes needed to validate reconstruction and analysis software of the ATLAS collaboration have been produced in so called „data challenges", in which the ACK CYFRONET AGH has participated from the beginning (2002). An other goal of these challenges was the validation of the ATLAS Computing Model, the complete software suite, the data model, and to ensure the correctness of the technical choices to be made.

Table 1

The CPU usage and data size per event in the ATLAS full simulation chain [11]

Stage	CPU [kSI2k-sec/event]	Event size [kB/event]
Generation	1	20
Simulation and Digitalization	400 (aimed 100)	2048
Reconstruction	30 (aimed 15)	10-500
Analysis	0.5	histograms

In order to estimate needed data volume for typical physics analysis performed by the ATLAS physicist, we add (20+2048+500) kB/event (see Table 1), that leads to 2.5 MB/event. We need typically 10^6–10^7 of events, yielding in 2.5–25 TB of data. However, to analyse all the data that will be collected by the ATLAS detector, the ATLAS collaboration requires production petabytes of simulated data per year.

The infrastructure of the ACK CYFRONET AGH was intensively used by the IFJ PAN ATLAS group and the ATLAS production managers. The ATLAS generation, simulation and reconstruction jobs were sent to the ACK CYFRONET AGH cluster as grid jobs and the results were sent back to CERN or to a local storage. As it can be seen from Table 1, the most CPU time consuming task in the full simulation chain is the simulation step. The ATLAS collaboration expects that current processing time of 400 kilo Spec Intel 2000 seconds (kSI2k-sec) [10] time will be reduced to 100 kSI2k-sec, due to an optimization of simulation algorithms.

The ACK CYFRONET AGH *zeus* cluster enabled the development of the various analysis codes, for example a code used for the estimation of the Higgs boson discovery significance in the Minimal Supersymmetric Standard Model [12]. It has been also used in the validation of the tau leptons reconstruction and identification package, developed by the IFJ PAN team. The produced validation Monte Carlo data samples were used by the ATLAS Tau Working Group.

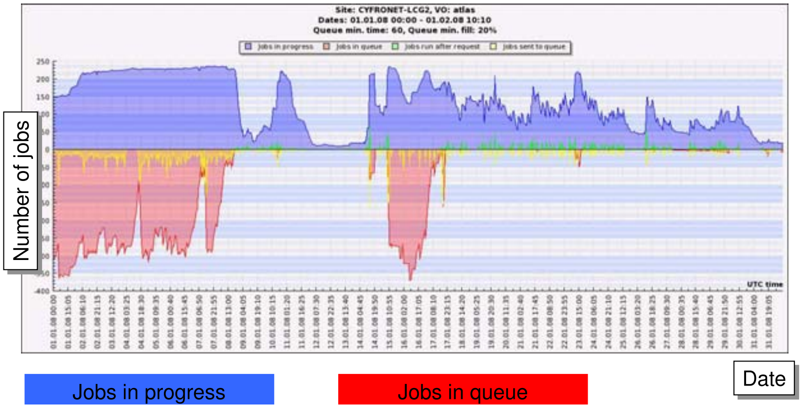

Fig. 4. The number of jobs processed (blue) and queued (red) in period 1–31 January 2008 at the ACK CYFRONET AGH *zeus* cluster

The involvement of ACK CYFRONET AGH as Tier2 center led to an increase of its storage space from ca. 1 TB (2003) to 58 TB (2008). Although it is significant growth, we expect further increase of resources dedicated to the HEP community. The Polish Tier2 Federation plans to pledge 200 TB in year 2008 and 340 TB in year 2009 [13].

The availability and reliability of the ACK CYFRONET AGH resources on the level of 94% at the end of 2007 is satisfactory [14]. The exemplary usage of Tier2 by the ATLAS production jobs is presented in Figure 4. At the beginning of January 2008 over 200 ATLAS jobs were processed at the same time.

5. Conclusions and future work

The cooperation with the ACK CYFRONET AGH has enabled our participation in a world-wide collaboration. The position of the ACK Cyfronet AGH as a Tier2 computing center of the Worldwide LHC Computing Grid required an extension of its storage capacity and delegation of large resources (both CPU and storage) to the Virtual Organisation of ATLAS. In view of the arrival of real data from the ATLAS detector we plan a substantial increase of the CPU and the storage usage, not only by the central production, but mainly by physics analyses performed by the local ATLAS HEP community.

Acknowledgements

We acknowledge a very good cooperation with the ACK CYFRONET AGH. The work was supported in part by Polish Ministry of Science and Higher Education grants 154/6.PRUE/2007 and PBS NR 132/CER/2006/03.

References

[1] The CERN homepage. Available on: `http://www.cern.ch`

[2] The ATLAS experiment homepage. Available on: `http://atlas.ch`

[3] The Institute of Nuclear Physics Polish Academy of Sciences IFJ PAN homepage. Available on: `http://www.ifj.edu.pl`

[4] The AGH University of Science and Technology homepage. Available on: `http://www.agh.edu.pl`

[5] The Academic Computer Center CYFRONET AGH homepage. Available on: `http://www.cyf-kr.edu.pl`

[6] The Worldwide LHC Computing Grid WLCG homepage. Available on: `http://lcg.web.cern.ch/LCG`

[7] The EU Crossgrid project homepage. Available on: `http://www.crossgrid.org`

[8] The ATLAS computing twiki. Available on: `https://twiki.cern.ch/twiki/bin/view/Atlas/AtlasComputing`

[9] Kaczmarska A., Richter-Was E., Wolter M., Janyst L.: *Performance of the tau1p3p algorithm for hadronic tau decays identification with release 12.0.6.* ATLAS Note ATL-PHYS-INT-2008-004; ATL-COM-PHYS-2007-039

[10] The SPEC benchmark homepage. Available on: `http://www.spec.org`

[11] The ATLAS Computing Group, ATLAS TDR–17, CERN-LHCC-2005-022

[12] Szymocha T.: *Prospects for the* $b\bar{b}H$, $H \to \tau^+\tau^-$ *observability for masses below* 200 *GeV with ATLAS detector at LHC.* PhD Thesis October 2007, IFJ PAN Kraków

[13] Memorandum of Understanding for Collaboration in Deployment and Exploitation of the Worldwide LHC Computing Grid, CERN-C-RRB-2005-01/Rev. 07 December 2007

[14] Tier-2 Availability and Reliability Report. Available on: `http://lcg.web.cern.ch/LCG/MB/availability/Tier-2/january2008/`