NEMO's part in investigating the Higgs boson

Ulrike Schnoor and Felix Bührer

Physikalisches Institut, Albert-Ludwigs-Universität Freiburg

After the discovery of a particle compatible with a Higgs boson in 2012, one of the main efforts at the Large Hadron Collider (LHC) at the European Center for Nuclear Research (CERN) has been the investigation of the new particle's properties. This talk describes some of the related studies as well as the general workflow of LHC data analysis as it is carried out at computing centers around the world. Some of this work is performed at the `NEMO` computing cluster at the University of Freiburg, a resource of the *bwHPC-C5* project. It offers additional computing power to the existing Tier-2 and Tier-3 site of the World-wide LHC Computing Grid (WLCG) in Freiburg, the Black Forest Grid (`BFG`), through the deployment of virtual research environments.

1 Introduction

The High-Performance Computing (HPC) cluster `NEMO` [1] at the University of Freiburg is one of the Tier-3 centers within the *bwHPC-C5* project [2]. It provides the communities of elementary particle physics, neuroscience, and microsystems engineering in the state of Baden-Württemberg with a hybrid of cloud computing and HPC resources.

The particle physics community has been driving developments of data-intense research from the start [3], with the most well-known contribution being the 1989 proposal of the World-Wide Web [4]. Even in the era of Big Data, it ranks among the occupants with the largest needs of computing storage [5]. Most of the data production, storage, and analysis is carried out using the World-wide LHC Computing Grid (WLCG) [6].

One of the most important results so far at the Large Hadron Collider has been the observation of a new particle compatible with a Higgs boson in 2012 [7]. This discovery has started the pursuit of a variety of Higgs-related measurements: From the couplings to the particles of the Standard Model to spin and parity properties of the Higgs boson, a large amount of follow-up measurements has already been published, e.g. [8]. So far, all results are in agreement with the Standard Model predictions. However, in many cases, the statistical and systematic uncertainties are rather large, allowing for a range of deviations from the Standard Model which could still be realized in nature. Section 2 briefly outlines the basic theory and current research in particle physics.

The Black Forest Grid (`BFG`) at the University of Freiburg has been a part of the WLCG since 2005. The ATLAS production and analysis jobs for which the `BFG` has provided resources have been and continue to be a part of the discoveries and measurements of ATLAS. Section 3 describes the LHC and the WLCG projects. New approaches to extend the resources for local user jobs and, in an opportunistic way, the Tier-2 resources, are being tested at the `NEMO` HPC cluster and will be described in Sec. 4.

2 The Standard Model of Particle Physics

2.1 Fundamental forces and elementary particles

The Standard Model of Particle Physics (SM) is the most complete theory describing three of the four fundamental forces. It has been developed in the 1960s [9, 10] and has proven to be a very successful theory, accurately describing all observed particle interactions in the currently accessible energy range. The only known exceptions are the gravitational force, which is not included in the SM, and the masses of neutrinos.

The SM is a quantum field theory based on local gauge symmetries which generate the forces between the elementary particles. Each force is mediated by the corresponding gauge bosons: For the electroweak force, which unifies the electromagnetic and weak forces, these are the photon, $W^{\pm}$, and Z^0 bosons. The strong force describing interactions of quarks within nuclei and hadronic bound states is mediated by gluons.

The known fundamental particles can be grouped according to their properties and the interactions in which they take part. Particles with spin 1/2, the fermions, are divided in two groups: Leptons, which interact via the electroweak force, and quarks, which interact via the electroweak and the strong forces. In the lepton sector, doublets of electrically charged leptons and zero electric charge neutrinos can be arranged in three so-called families ordered by mass of the charged lepton (electron, muon, or τ lepton). Similarly, the six quark flavors are ordered in three families where the quarks within one family form a doublet under the weak symmetry.

For each particle of the SM, one antiparticle exists which carries the same mass, same spin, and the exact opposite charges.

2.2 The Higgs mechanism and its discovery at the LHC

In the SM, the fundamental fermions and bosons are described as massless. However, experimental evidence has shown that the weak bosons, charged leptons and the quarks, and even neutrinos, are massive with masses ranging from less than 1 eV (neutrinos) up to 172 GeV (top quark). Introducing a mass for these particles breaks the electroweak symmetry. The Higgs mechanism provides a description to break this symmetry spontaneously, while the underlying Lagrangian still fulfills the symmetries. According to this mechanism, particles gain their mass through the interaction with a scalar field. The excitations of this field generate the Higgs boson, which is massive itself.

The postulation of the Higgs mechanism in 1964 [11, 12, 13] started the hunt for a signal of the Higgs boson. At the predecessors of the LHC, no sign for this boson had been found. Finally, in 2012, the two largest experiments ATLAS and CMS have presented the discovery of a new gauge boson based on data analysis in different decay channels at a mass of 125 GeV [14].

2.3 Scrutinizing the Higgs boson and open questions

Several theories beyond the SM incorporate a Higgs-like boson generated through electroweak symmetry breaking. However, if the Higgs sector contains physics beyond the SM, certain properties of the 125 GeV boson should deviate from the values predicted by the SM.

Several open questions imply the need for physics beyond the SM. The Higgs mechanism does not give a reason for the values of the masses of the elementary particles, which range over many orders of magnitude. Also, several observations have lead to the conclusion, that a new kind of matter called "dark matter" must exist [15, 16]. However, none of the known elementary particles accounts for its effects.

Furthermore, the imbalance of matter and anti-matter in the universe can not be explained within the SM. Today's universe consists mostly of matter, and no large amounts of anti-matter have been found, even though during the Big Bang, equal amounts of anti-matter and matter should have been produced according to the so-called CP symmetry [17]. The SM contains a small asymmetry between matter and anti-matter [18], which is not sufficient to explain the observed ratio of matter to anti-matter though. Hence, searches for new sources of violation of the CP symmetry are being investigated. One such possibility is that the Higgs sector introduces CP violation. Investigations of CP violation in the Higgs sector, using the weak gauge boson fusion production of Higgs bosons, have been carried out at the ATLAS experiment using computing resources in Freiburg [19]. A follow-up study with data from 2015 and 2016 will be using the `NEMO` cluster for parts of the data analysis.

3 The Large Hadron Collider and the World-wide LHC Computing Grid

3.1 The ATLAS experiment at the Large Hadron Collider

The ATLAS experiment is a large multipurpose particle detector measuring proton-proton collisions provided by the Large Hadron Collider (LHC). It is located near Geneva, Switzerland and has been taking data since 2009.

The LHC is situated in a circular tunnel of 27 km circumference. It accelerates two oppositely directed proton beams to an energy of 6.5 TeV each (in 2015 and 2016). The two beams collide at four interaction points, each occupied by a large particle detector. The largest detector is the ATLAS experiment with a length of 45 m and a height of 25 m. It consists of several subdetectors as depicted in Fig. 1.

In the years 2012-2016, the ATLAS data acquisition has produced an output of raw data of 10-15 petabytes per year [20, Fig. 3.7].

3.2 The World-wide LHC Computing Grid

The World-wide LHC Computing Grid (WLCG) is a distributed computing system consisting of more than 170 sites in 42 countries contributing computing power as well as storage to the LHC experiments [6]. The WLCG is organized in a hierarchical way, with the Tier-0 centers at CERN and Wigner Research Center for Physics, Budapest, Hungary in the center (Fig. 1). The Tier-1 level comprises 13 large computing facilities, and 140 centers make up the Tier-2 level.

3.3 The ATLAS computing model

The distributed data management (DDM) and distributed computing services must deal with the vast amounts of collision data that the ATLAS experiment produces when the LHC is running as well as simulated data gained from computational resources. Sufficient backup and fast access to datasets for analyzers must be ensured. Processes of interest usually account for only very small fractions of the collected events. After thorough calibration procedures, data

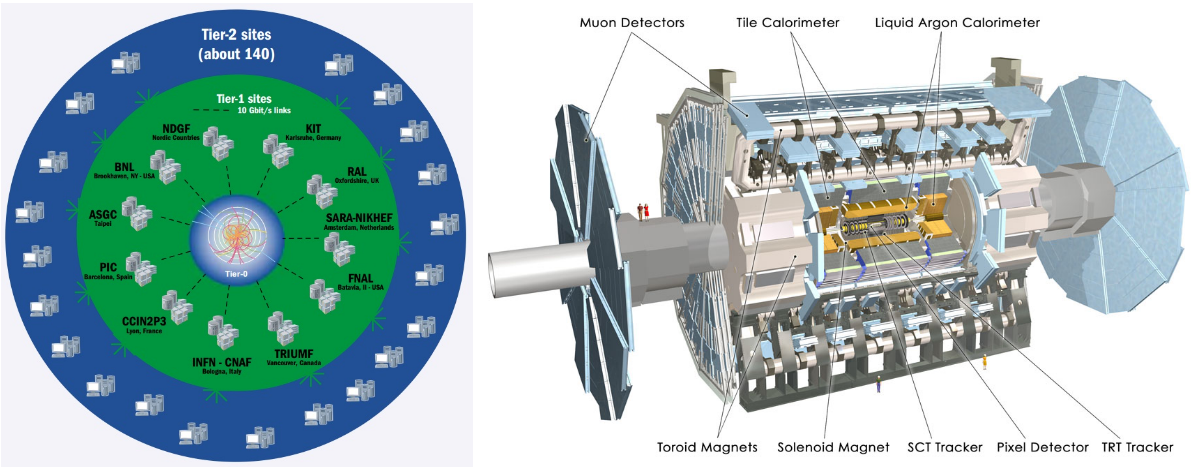

Figure 1: Left: Structure of the WLCG[1]. Right: Schematic of the ATLAS experiment[2].

samples are subsequently reduced through the application of filters optimized for the selection of interesting events.

The ATLAS computing model [21] incorporates all steps from the raw data to the completed statistical analysis of signal events. It is divided in several steps, each reflected in a specific data format described by the ATLAS Event Data Model (EDM) [22].

The primary processing of collision data from the detector is handled at the Tier-0 center. The resulting raw data is archived at the Tier-0 as well as at Tier-1 centers, where it is also reprocessed. The Tier-1 centers provide capacities to store and to analyze the reprocessed data. Derivation algorithms are run on the reconstructed data, providing analysis-specific reduction of the content saved per event. Derived datasets are then copied to Tier-2 facilities where they can be analyzed. Each Analysis Object Data (AOD) dataset is replicated once at a Tier-1 and once at a Tier-2 disk. In addition to data processing capacities, the Tier-2 centers provide capacities for simulation and for calibration of processed raw data. Simulated data is stored at the Tier-1 centers analogously to detector data. Tier-3 facilities provide additional computing resources.

The final element of the chain is the scrutinizing data analysis performed by the researchers [23]. This requires the optimization of selection criteria, simulation of specific event samples, evaluation of systematic uncertainties; furthermore the application of fits and the generation of so-called toy Monte Carlo events for the statistical analysis [23]. These tasks are carried out on Tier-2 and Tier-3 resources or on additional resources provided by institutes and universities around the world.

3.4 The Black Forest Grid in Freiburg

The Black Forest Grid (`BFG`) is a shared Tier-2 and Tier-3 site of the WLCG. It comprises 220 worker nodes providing a total of 3000 cores (hyperthreaded). For distributed WLCG storage, a `dCache` instance provides a total of 1.35 PB. The space provided to ATLAS users corresponds to 1/60 of the total storage of ATLAS in the WLCG. In addition, a parallel storage based on the `lustre` [24] file system is deployed, providing 145 TB of storage to local (Tier-3) users.

The site has been in operation since 2005. It is currently shared with local users from the

[1]Source: `http://wlcg-public.web.cern.ch/tier-centres`
[2]Source: `http://kjende.web.cern.ch/kjende/zpath_files/img/highslide/ATLAS_detector_alles_mittel_EN.png`

communities of physics, biodynamics, and other groups. In the imminent future, these local users are moving to the resources of the *bwHPC-C5* centers.

4 Virtualizing the ATLAS research environment

4.1 The NEMO HPC cluster in Freiburg

The *bwForCluster* NEMO is a High-Performance Computing resource at the Rechenzentrum at Universität Freiburg. It comprises 756 worker nodes, each with 40 cores (hyperthreaded) and a RAM of 128 GB. The nodes are interconnected with a 100 Gbit/s OmniPath [25] system. Each node is supplied with a local 240 GB SSD.

After its installation in July 2016, the NEMO cluster has been ranked 214 in the TOP 500 list [26]. It uses a novel approach for a hybrid of HPC and cloud computing [27] and serves the communities of elementary particle physics, microsystems engineering, and neuroscience in the state of Baden-Württemberg.

4.2 ATLAS software on NEMO

ATLAS data analysis programs currently mainly run on Scientific Linux 6 (SL6) operating systems, which is based on RedHat Enterprise Linux 6. Most of the software environment is provided by the CERN VM File System (cvmfs). Hence, ATLAS data analysis is run ideally on machines with a SL6 operating system providing cvmfs infrastructure.

The *bwForCluster* NEMO runs CentOS7 as operating system and does not provide cvmfs. Therefore, in order to run ATLAS software on NEMO, the virtualization and containerization of the environment is investigated as a means to emulate the ATLAS environment on NEMO as the host system. It can then be used for local (WLCG Tier-3) and grid jobs (WLCG Tier-2).

4.3 Virtualization of the ATLAS infrastructure

The virtualization of ATLAS software on NEMO makes use of its already established hybrid cluster approach. An instance of OpenStack [28] is deployed allowing to run both bare metal jobs and virtual machines (VMs). The latter are using kvm [29] as the hypervisor.

In order to start virtual machines on NEMO on demand for jobs submitted at the BFG, the challenge lies in the connection of the batch schedulers of the two systems, which is illustrated in Fig. 2. The BFG login nodes serve as user interfaces. The scheduler deployed at the BFG, a slurm [30] instance, is used to allow users to schedule jobs running on VMs, while the VMs themselves are scheduled by the Moab [31] instance running at NEMO.

The *elastic computing* feature of the slurm scheduler [32] is used to request the start-up of new VMs on demand.

The VM is based on an SL6 image contextualized using packer [33] and running a dedicated puppet [34] role similar to the one used for setup of regular BFG worker nodes. Using packer and puppet allows to automatize and versionize the generation of a new image whenever an update should be propagated to the VMs.

4.4 NEMO's part in investigating the Higgs boson

With the setup for virtualized research environments described above, NEMO can be used by researchers from the connected institutions for data analysis of ATLAS data. Groups from

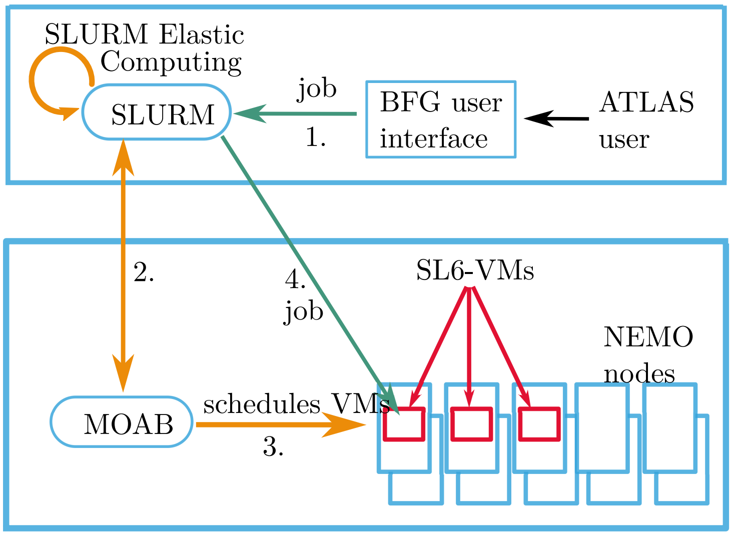

Figure 2: Virtualization environment for ATLAS on NEMO.

Freiburg are taking advantage of this resource. The same setup can potentially be exploited for opportunistic use of NEMO for standalone CPU-intensive tasks such as event generation campaigns [35].

5 Conclusions and Outlook

The ATLAS experiment at the LHC is investigating the nature of the Higgs boson, a particle connected to the mechanism of electroweak symmetry breaking which constitutes an important part of the Standard Model of particle physics. At the same time, it can offer insights into physics beyond the Standard Model, which is investigated through the measurement of its properties, such as its behavior with respect to the CP symmetry.

The ATLAS computing model relies on distributed computing to handle the up to 15 PB of data output from the experiment deploying the World-wide LHC Computing grid.

One of the Tier-2 sites of the WLCG is the BFG at Universität Freiburg. Efforts to virtualize the ATLAS research environment are currently on-going, with the aim of allowing jobs using Tier-3 resources of the BFG to run on the *bwHPC-C5* NEMO center at Universität Freiburg. Main challenges include the integration into the schedulers of NEMO and BFG to start virtual machines on demand, providing access to the local user environment, as well as the automatic generation of a fully functional, versionized virtual machine image.

Acknowledgements

The *bwHPC-C5* project has been funded by the Ministry of Science, Research and the Arts of the state of Baden-Wuerttemberg, Germany.

The authors thank Konrad Meier (Computing Center Universität Freiburg) for collaboration on the virtualization setup. Many sincere thanks to Anton J. Gamel and Michael Böhler for successful operation of the Black Forest Grid and to Markus Schumacher for the project leadership. Cooperation with the HPC team (Michael Janczyk, Konrad Meier, Dirk von Suchodoletz, Bernd Wiebelt) at the Computing Center at Universität Freiburg is gratefully acknowledged.

References

[1] "bwForCluster NEMO at Universität Freiburg." https://www.hpc.uni-freiburg.de/nemo, accessed 2017-01-05

[2] "bwHPC-C5 Project." https://www.bwhpc-c5.de/, accessed 2017-01-05

[3] Paolo Zanella. "30 Years of Computing at CERN." http://ref.web.cern.ch/ref/CERN/CNL/2001/002/cern-computing/, accessed 2017-01-05

[4] https://www.w3.org/History/1989/proposal.html, accessed 2017-01-13

[5] https://www.wired.com/2013/04/bigdata/, accessed 2017-01-13

[6] The LCG TDR Editorial Board. "LHC Computing Grid: Technical Design Report." LCG-TDR-001, CERN-LHCC-2005-024. 20 June 2005. http://cds.cern.ch/record/840543/files/lhcc-2005-024.pdf accessed 2017-01-09

[7] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." Phys. Lett. B 716 (2012) 1-29; CMS collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." Phys. Lett. B 716 (2012) 30

[8] ATLAS and CMS Collaborations. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC *pp* collision data at $\sqrt{s} = 7$ and 8 TeV." JHEP 08 (2016) 045

[9] S.L. Glashow. "Partial-symmetries of weak interactions." Nuclear Physics. 22 (4): 579–588.

[10] S. Weinberg. "A Model of Leptons." Phys. Rev. Lett. 19 (21): 1264–1266.

[11] F. Englert, R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." Physical Review Letters. 13 (9): 321–323.

[12] P.W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." Physical Review Letters. 13 (16): 508–509.

[13] G.S. Guralnik; C.R. Hagen; T.W.B. Kibble. "Global Conservation Laws and Massless Particles." Physical Review Letters. 13 (20): 585–587.

[14] ATLAS and CMS Collaborations. "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments." Phys. Rev. Lett. 114, 191803 (2015)

[15] Jacobus Cornelius Kapteyn. "First attempt at a theory of the arrangement and motion of the sidereal system." Astrophysical Journal. 55: 302–327 (1922)

[16] J.H. Oort. "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems." Bulletin of the Astronomical Institutes of the Netherlands, 6 : 249-287 (1932)

[17] L. Landau. "On the conservation laws of weak interactions, Nuclear Physics." Band 3, 127–131 (1957)

[18] J. H. Christenson, J. W. Cronin, V. L. Fitch, R. Turlay. "Evidence for the 2π Decay of the K_2^0 Meson System." Physical Review Letters. 13: 138 (1964)

[19] ATLAS Collaboration. "Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector." Eur. Phys. J. C76 (2016) 658

[20] "Update of the Computing Models of the WLCG and the LHC Experiments." WLCG-ComputingModel-LS1-001, http://cds.cern.ch/record/1695401/files/LCG-TDR-002.pdf, accessed 2017-01-09

[21] ATLAS Collaboration. "ATLAS Computing Technical Design Report." https://cdsweb.cern.ch/record/837738/files/lhcc-2005-022.pdf, accessed 2017-01-09

[22] A. Buckley, T. Eifert M. Elsing, D. Gillberg, K. Koeneke, A. Krasznahorkay, E. Moyse, M. Nowak, S. Snyder, P. van Gemmeren. "Implementation of the ATLAS Run 2 event data model." 2015 J. Phys.: Conf. Ser. 664 072045 (http://iopscience.iop.org/1742-6596/664/7/072045)

[23] O. Behnke, K. Kroninger, G. Schott, T. Schorner-Sadenius. "Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods." Wiley-VCH 2013. ISBN:3527410589

[24] "Lustre File System." http://lustre.org/, accessed 2017-01-09

[25] "OmniPath." http://www.intel.de/content/www/de/de/high-performance-computing-fabrics/omni-path-architecture-fabric-overview.html, accessed 2017-01-10

[26] NEMO on the TOP500 Supercomputer Statistics List https://www.top500.org/system/178839, accessed 2017-01-10

[27] https://www.bwhpc-c5.de/wiki/index.php/Category:BwForCluster_NEMO, accessed 2017-01-10

[28] OpenStack Open Source Cloud Computing Software https://www.openstack.org/, accessed 2017-01-10

[29] KVM (Kernel-based Virtual Machine) www.linux-kvm.org/, accessed 2017-01-10

[30] https://slurm.schedmd.com/overview.html, accessed 2017-01-13

[31] http://www.adaptivecomputing.com/products/hpc-products/moab-hpc-basic-edition/, accessed 2017-01-10

[32] https://slurm.schedmd.com/elastic_computing.html, accessed 2017-01-13

[33] Packer: tool for creating machine and container images for multiple platforms from a single source configuration. https://www.packer.io/, accessed 2017-01-13

[34] Puppet Enterprise. "IT automation for cloud, security, and DevOps." https://puppet.com/, accessed 2017-01-10

[35] B. Wiebelt, M. Janczyk, D. v. Suchodoletz, A. Aertsen, S. Rotter, G. Quast, M. Schumacher, A. Greiner. "Strukturvorschlag für eine bwHPC-Governance der ENM-Community." in "Kooperation von Rechenzentren: Governance und Steuerung – Organisation, Rechtsgrundlagen, Politik." De Gruyter, Oldenburg, 2016