Physics Analysis Tools for Beauty Physics in ATLAS

Christos Anastopoulos[8], Eva Bouhova-Thacker[2], James Catmore[2], Steve Dallison[5], Frederick Derue[6], Brigitte Epp[3], Patrick Jussel[3], Anna Kaczmarska[7], Lee de Mora[2], Holger von Radziewski[4], Pavel Řezníček[1], Thorsten Stahl[4]
(on behalf of the ATLAS collaboration)

[1]IPNP, Faculty of Mathematics and Physics, Charles University in Prague
V Holešovičkách 2, 18000 Prague-8, Czech Republic
[2]Department of Physics, Lancaster University
Lancaster LA1 4YB, UK
[3]Institute for Astro- and Particle Physics, University of Innsbruck
Technikerstr. 25, A-6020 Innsbruck, Austria
[4]Department of Physics, University of Siegen
Emmy-Noether-Campus, Walter-Flex-Str. 3, D-57068 Siegen, Germany
[5]Particle Physics Department, CCLRC Rutherford Appleton Laboratory
Chilton, Didcot, Oxfordshire, OX11 0QX, UK
[6]LPNHE, IN2P3 - CNRS - Universités Paris VI et Paris VII,
4 place Jussieu, Tour 33, Rez-de-Chaussée, 75252 Paris Cedex 05, France
[7]Institute of Nuclear Physics, Polish Academy of Sciences
Radzikowskiego 152, 31-342 Kraków, Poland
[8]Physics Department, Aristotle University Of Thessaloniki, 541 24, Greece

E-mail: `pavel.reznicek@cern.ch`

Abstract. The Large Hadron Collider experiments will search for physics phenomena beyond the Standard Model. Highly sensitive tests of beauty hadrons will represent an alternative approach to this research. The analyses of complex decay chains of beauty hadrons will require involving several nodes, and detector tracks made by these reactions must be extracted efficiently from other events to make sufficiently precise measurements. This places severe demands on the software used to analyze the B-physics data. The ATLAS B-physics group has written series of tools and algorithms for performing these tasks, to be run within the ATLAS offline software framework Athena. This paper describes this analysis suite, paying particular attention to mechanisms for handling combinatorics, interfaces to secondary vertex fitting packages, B-flavour tagging tools and finally Monte Carlo truth association to pursue simulation data in process of the software validations which is important part of the development of the physics analysis tools.

1. Introduction

ATLAS is a general purpose detector [1] designed to exploit full discovery potential of Large Hadron Collider (LHC) - $7+7\,\mathrm{TeV}$ proton accelerator with ultimate luminosity of $10^{34}\,\mathrm{cm^{-2}s^{-1}}$. The experiment is expected to begin data taking in summer 2008 [2]. Its wide program consists of both discovery physics and precision Standard Model measurements.

Physics of beauty quark systems is non-negligible part of ATLAS program [3]. It involves measurement of CP-violation, B-oscillations, B-hadron properties and rare decays with the aim to observe new physics effects. Main part of the B-physics program is concentrated for the initial low-luminosity ($L = 2 \cdot 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$) stage of LHC. This implies that B-physics will serve as a first test bed for understanding the detector properties. Therefore the very first studies of the ATLAS B-group include detector commissioning with early data, trigger, tracking and muon system calibrations and precision mass and lifetime measurements of well known B-particles. Expected rate of $b\bar{b}$ quark pairs produced within ATLAS detector acceptance during the low luminosity stage of LHC is $\sim 20\,\mathrm{kHz}$. Since for whole the ATLAS physics $\sim 100\,\mathrm{Hz}$ are committed to disk out of which $\sim 10\,\mathrm{Hz}$ are devoted for B-physics, highly selective trigger must be used. The B-physics triggers are generally based on detection of single-muon, di-muon or muon and calorimetry cluster [4, 5].

In order to make measurements precise enough for extraction of new physics effects, the analysis has to efficiently separate the tracks of complex B-hadron decay chains from other reactions. An analysis suite involving several nodes and consisting of a set of common tools, specialized data structures and a list of algorithms for each exclusive decay channel, has been written by ATLAS B-physics group to fulfill this requirement.

2. ATLAS Software and Data Flow Overview

The B-physics analysis suite is implemented within ATLAS software framework Athena [6] - an object oriented based software, primarily using C++ programming language, but also with some FORTRAN and Java components. Athena applications are built up from collections of plug-compatible components driven by variety of configuration files. The framework provides common data-processing support, with major components including list of tools (e.g. vertexing tools), services (e.g. logging facility, histogram and ntuple services), data representation converters, common interface for per-event processing algorithms etc., all driven by application manager component. B-physics analyses are coded in pure C++. The main processing classes are derived from an Athena algorithm extended by robust handling of ROOT-ntuple output [7].

The physics analysis runs on reconstructed data devoted for physics analysis - Analysis Object Data (AOD), the final product of the ATLAS reconstruction chain [6]. RAW data (output of the final stage of ATLAS High Level Trigger) are primarily stored in Tier-0 facility in CERN and copied to Tier-1 sites (~ 10 for ATLAS around the world). The RAW data are reconstructed producing Event Summary Data (ESD), intended to make access to RAW data unnecessary for most physics applications except for some calibration and re-reconstructions. The AOD are then derived from ESD, representing reduced event information suitable for analysis - contain physics objects and other elements of analysis interest. Along side with AOD also Tag Data (TAG) are extracted collecting event-level metadata to support efficient identification and selection of events of interest to a given analysis. Both AOD and ESD are held on Tier-1 sites, while AOD are copied to Tier-2 sites that are accessible for physicist to perform the analyses. Either ESD and AOD have an object-oriented representation readable by Athena.

Tier-2 sites also provide all the required simulation capacity for the experiment. Monte Carlo (MC) simulation chain consists of event generation, Geant4 [8] simulations of the particles passage through ATLAS detector and finally detector response simulation. The output of the last stage (digits) is then used to produce ESD and the rest of the chain is similar as with real data.

To avoid large amount of data being copied over to local computers, the analysis software is being sent to the data on Tier-2 sites. The jobs are distributed via a GRID middleware to which several interface exist, B-physics group is presently using Ganga [9]. Large Computing System Commissioning (CSC) MC production was performed via this interface in 2007: 450k of signal and 700k of background events generated, simulated and reconstructed with ongoing analysis by individual physicists on selected data samples appropriate to studied channels.

3. B-Physics Analysis

At the LHC, the B-physics program consists of study of exclusive decay channels, therefore the analysis is characterized by many different topologies and constraints. In order to minimize duplicated effort, common tools, data structures and calculations were collected and unified over all the B-physics channels. The main task in B-physics analysis is to identify a B-decay chain typically consisting of cascade of several vertices, e.g..:

$$B_s^0 \to \begin{cases} J/\psi \to \begin{cases}\mu^+\\ \mu^-\end{cases} \\ \phi \to \begin{cases}K^+\\ K^-\end{cases}\end{cases} \qquad B_s^0 \to \begin{cases}\mu^+\\ \mu^-\end{cases} \qquad B_s^0 \to \begin{cases}\pi^+\\ D_s^- \to \begin{cases}\pi^-\\ \phi \to \begin{cases}K^+\\ K^-\end{cases}\end{cases}\end{cases}$$

This requires the usage of offline vertex finder, management of combinatorics of tracks to form the candidates of elements of the decay chain and extraction of various properties of the fitted particle-like objects. In order to check the efficiencies, ideal performance and sources of background, MC truth and its associations to reconstructed objects are used and there were tools developed to ease manipulation with these relations.

Since the B-physics analysis suite is a part of Athena software framework, Athena services are used to access the objects stored in the AOD files. Presently, the B-physics analyses rely on reconstructed Inner Detector [10] tracks, combined muon and electron objects, trigger decision information, reconstructed primary vertex and particle jets.

This Athena analysis is used to produce ROOT ntuples, that can be transfered to user's private locations and the final tuning is performed using collection of ROOT scripts.

3.1. Typical Analysis Procedure

Analyses of the various decays have a common structure, that is here illustrated on reconstruction of $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ channel. The sequence of the analysis steps is following:

(i) Get the collections of the analysis objects in a given event.

(ii) Select tracks matching kinematics cuts for muon candidates and consult the muon combined objects to retain tracks belonging to identified muons only. Combine such muon candidates into opposite charged pairs (J/ψ pre-candidates) and possibly cut out pairs not satisfying some additional conditions like e.g. the invariant mass (not in vertex) window.

(iii) Perform vertexing of all the preselected muon pairs and apply appropriate quality, position, mass window and other cuts on the J/ψ candidates.

(iv) Select ϕ candidates using similar procedure (combine and put into vertexing tracks being identified neither as muons nor as electrons).

(v) Combine the J/ψ and ϕ candidates to create tracks quadruplets and perform vertex-fitting of whole the B_s^0 decay topology and again reject low-quality, out of mass window, etc. B_s^0 candidates.

(vi) Calculate additional variables for B_s^0 isolation, flavour tagging etc., to be either used to further rejections or just stored alongside with the B_s^0 candidates in the output ntuple for consequent ROOT analysis.

4. B-Physics Analysis Tools

The analysis suite consists of set of predefined decay identification algorithms, tools for building them and data structures simplifying management of the selected candidates. The software is organized into three separated packages [11]:

(i) `BPhysExamples` - holds the main analysis algorithms and an example skeletons. The structure of the code has one algorithm per one decay channel, with the present implementation of:
   - $B^+ \to J/\psi(\mu^+\mu^-)K^+$
   - $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$
   - $B_s^0 \to D_s^-(\phi(K^+K^-)\pi^-)\pi^+$
   - $B_s^0 \to D_s^-(\phi(K^+K^-)\pi^-)a_1$
   - $\Lambda_b^0 \to \Lambda^0(p, \pi^-)\mu^+\mu^-$

(ii) `BPhysAnalysisTools` - collection of tools for building decay algorithm. This includes tracks combinatorics and kinematic cuts based tracks selection, secondary vertex fitters, truth finding and associations, B-flavour tagging and helper routines to calculate proper time, transverse decay length and other variables.

(iii) `BPhysAnalysisObjects` - contains non-trivial data structures used by the analysis program and the tools: vertex fitting and flavour tagging results and class collecting properties of composite (decayed) particle.

Finally, the suite is completed by users ROOT scripts for analysis of the output ntuples.

4.1. The Data Structures

There are three classes extensively used in the analysis algorithms and tools:

(i) `Vertex`
   Collects results of vertex fitting, using data members natural for analysis. Its important function is also to unify (when possible) the output of various vertexers, so that their differences in output format are not visible from the main analysis algorithms. In comparison to similar objects at the reconstruction stage, this class provides physics particle-like objects - e.g. tracks are having mass hypotheses associated, the vertex holds its invariant mass, etc. The class members contain following information:
   - vertex position, fit covariance matrix, fit quality, fit error flag, type of used vertexer
   - total vertex (composite particle) momentum, charge mass
   - refitted tracks (tracks in vertex) parameters and their χ^2
   - separation to mother vertex
   - refitted primary vertex
   - variables specific to selected vertexer (see section 4.3)

   In order to be kept as lite and independent as possibly, but without impact on functionality, CLHEP [12] objects are used to represent more complex members.

(ii) `VertexAndTracks`
   Designed to represent a composite particle candidate, keeping relevant reconstructed, truth and genealogy information. It holds also results which are used to decide about whether the candidate should be rejected (is out of requested cuts). The present contents is:
   - the fitted vertex
   - original tracks of the stable daughters
   - reference to composite mother and children candidates
   - MC truth particles associated with the stable daughter tracks
   - variables for cut-based selection of the candidates: (fitted) e.g. invariant mass, proper time, impact parameter, total p_{T}

(iii) `BFlavourParticle`
   Container of lepton flavour tagging results (see section 4.4). It holds some properties of the tagging muon candidate and eventually the nearest jet.

4.2. Tool-Box

`BPhysToolBox` class is collecting all common tools and calculation used among most or all the decay channels analyses. Methods of the Tool-Box are divided into five functionally relative groups:

- tracks selection methods, based on kinematic parameters (basically p_{T}, η, charge)
- tracks combinatorics, creating pre-candidates of composite particles, e.g. making (opposite charge) pairs, triplets, etc.
- particle identification functions returning collection of tracks associated with requested particle type (presently includes muon identification only based on either muon detectors information or MC truth associations)
- list of various parameters calculations, e.g. proper time, forward-backward asymmetry of a B-decay, impact parameter, etc. not present in the analysis objects
- methods for manipulation with MC truth, e.g. tracks association to MC particles, extraction of requested decay chain (and also associated tracks) from the MC truth collection, etc.

4.3. Vertexing Interfaces

Robust offline vertex fitter supporting mass and pointing constraints is one of the key tools for B-physics analysis. There are several vertex fitting algorithms in Athena framework, out of which B-physics presently uses two:

- `VKalVrt` using Kalman method for vertexing [13].
- `CTVMFT` fortran based fitter developed by CDF collaboration, extended by interface to Athena [14].

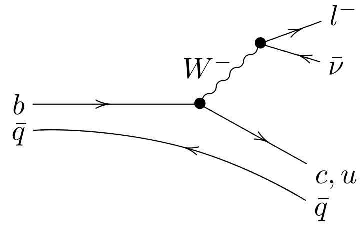
Both the fitters are capable of complex mass, pointing (e.g. to primary vertex) and conversion vertex type constraints. `VKalVrt` works with complete map of ATLAS inner detector magnetic field, in contrast to the CVTMFT fitter which uses constant field approach.

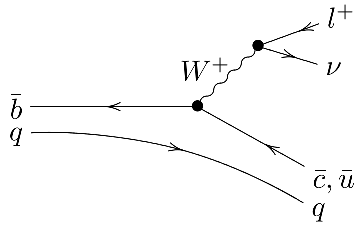
The B-physics interfaces to these fitter are wrappers to their Athena interfaces, which are presently different (unified interface is in development). The purpose of this additional interfacing is to allow analysis like input, e.g. defining decay topology to be fitted, and unifying the output of the two fitters: the results are stored in the `Vertex` class. This approach allows easy connection to any eventual new fitter. Though the input methods for the two vertexer were not yet completely unified, there was created `XtoYZFinder` tool (inheriting from Athena tool service represented by `AlgTool` class [6], p. 29) performing this unification for often fitted two-body decays.

4.4. Tagging Tools

Especially for mixing and CP-violation studies, flavour of the B-meson at the time of production needs to be known. There are two techniques implemented in the B-physics software. The first (*Lepton Tagger*) is using opposite b-quark weak decay characterized by high p_{T} lepton whose charge indicates the B-hadron flavour. The second method (*Jet Charge Tagger*) makes use of associated jet production alongside with production of the B-hadron of interest, when the charge of the jet is correlated with the flavour of the b-quark. Both the implemented tools are inherited from Athena `AlgTool`.

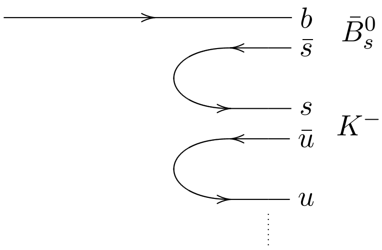
- *Lepton Tagger*

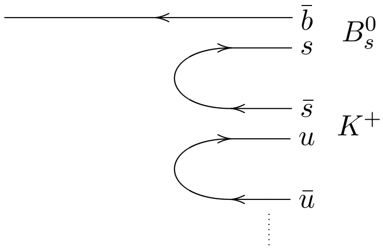




  The lepton tagging algorithm written inside `BFlavourTagger` class is searching for high-p_{T} muons in the event and eventually an associated jet. It scans over input muon collection, and eventually searches for nearest jet from input particle jet collection, and storing the results into a list of `BFlavourParticle`s sorted decreasingly by the muons p_{T}. The inputs and outputs are handled via `set-` and `get`-methods. This lepton tagger tool also allows to perform internal histogramming of the tagging results that can be used e.g. for calculation of wrong-tag fraction.

- *Jet Charge Tagger*





  The method is implemented in `BFlavourJetChargeTagger` class. The core algorithm collects tracks inside predefined cone around a B-hadron, thus building an associate jet candidate and accepting it under certain condition on total weighted charge. The class consists of the `set`-methods for input tracks collection, B-hadron momentum, and configuration of the jet-building/rejection algorithm, then core calculation method and finally `get`-methods providing the tagging result and information about the tagging jet.

5. The Analysis Output

The analysis algorithms use Athena histogramming and ntupling services to produce control histograms and an output ntuple collecting all selected candidates of the analysed decay chain in each event. Applying loose cuts during the Athena analysis allows later selection tunning being performed on the ntuple level. Thus not needing Athena software installed and avoiding to rerun analysis on GRID over full datasets with every slight change of the selection cuts. The Athena analysis is intended to extract the required B-hadron candidates per event, however further statistical analysis over the datasets is performed on the output ntuples, naturally making use of existing statistical and mathematical tools included in ROOT, as the most used tool for ntuple analysis.

5.1. B-Physics Validation

Since the mass, proper time resolution and other properties of the reconstructed decayed particles are very sensitive to tracking and vertexing performance, the B-physics analyses can serve as a reliable validation tool during development of the Athena software, where it can address both reconstruction and simulation issues. For continuous checks of Athena releases, Run Time Tester (RTT) [15] was developed by ATLAS community, running predefined set of tests on each daily build of the development snapshot and reporting the results at CERN WWW pages. B-physics part of the RTT is based on tests of performance of $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ decay evaluated using the B-physics analysis packages. This test suite consists of set of ROOT scripts producing plots and python scripts driving the validation job - repeated analysis with fixed cuts on predefined set of signal events.

6. Summary

Analysis suite for ATLAS B-physics has been developed, covering all the aspects of various decay topologies and constraints of B-physics. The code is integrated within ATLAS software framework Athena, having transparent structure containing a set of main algorithms (one for each studied decay channel) and common tools for building the decay chain. The tools are implemented in separated classes - for vertexing interfaces, flavour tagging and general methods including combinatorics, identification and Monte Carlo truth handling. The suite has be evolving and used by ATLAS B-physics group since 2005, presently heavily applied during Computing System Commissioning ([6], p. 213) tasks. In near future, more decay channels are being included and appropriate tools added and ROOT scripts for the analysis of the output ntuples will be integrated into the analysis suite.

7. References

[1] Atlas Collaboration, "ATLAS: Detector and physics performance technical design report. Volume 1, 2," CERN-LHCC-99-14, ATLAS-TDR-14, May 1999. 475pp., CERN-LHCC-99-15, ATLAS-TDR-15, May 1999. 519pp.

[2] `http://lhc.web.cern.ch/lhc/`, `http://atlas.web.cern.ch/Atlas/index.html`

[3] P. Ball *et al.*, "B decays at the LHC," arXiv:hep-ph/0003238. CERN-TH-2000-101, Mar 2000. 112pp.

[4] J. Baines [ATLAS Collaboration], "ATLAS trigger for B physics," Nucl. Instrum. Meth. A **462** (2001) 259.

[5] N. Panikashvili [ATLAS Collaboration], "The ATLAS B-physics trigger," Nucl. Phys. Proc. Suppl. **156** (2006) 129.

[6] G. Duckeck *et al.* [ATLAS Collaboration], "ATLAS computing: Technical design report," CERN-LHCC-2005-022, ATLAS-TDR-017, Jun 2005. 246pp.

[7] `https://twiki.cern.ch/twiki/bin/view/Atlas/AthenaAwareNTuple`

[8] S. Agostinelli *et al.* [GEANT4 Collaboration], "GEANT4: A simulation toolkit," Nucl. Instrum. Meth. A **506** (2003) 250.

[9] K. Harrison *et al.*, "GANGA: A user-Grid interface for ATLAS and LHCb," *In the Proceedings of 2003 Conference for Computing in High-Energy and Nuclear Physics (CHEP 03), La Jolla, California, 24-28 Mar 2003, pp TUCT002* [arXiv:cs/0306085].
`http://ganga.web.cern.ch/ganga/`

[10] ATLAS Collaboration, "ATLAS inner detector: Technical design report. Vol. 1,2," CERN-LHCC-97-16, ATLAS-TDR-4, Apr 1997. 270pp., CERN-LHCC-97-17, Apr 1997. 663pp.

[11] C. Anastopoulos *et al.*, "Writing B-physics analysis code in Athena," ATL-COM-SOFT-2006-014, 20 Oct 2006. 38 p. (not publicly available),
`https://twiki.cern.ch/twiki/bin/view/Atlas/BPhysWorkingGroupAnalysisAndVertexing`,
`http://atlas-sw.cern.ch/cgi-bin/viewcvs-atlas.cgi/offline/PhysicsAnalysis/BPhys/BPhysExamples/doc`

[12] L. Lonnblad, "CLHEP: A project for designing a C++ class library for high-energy physics," Comput. Phys. Commun. **84** (1994) 307.

[13] V. Kostyukhin, "VKalVrt - package for vertex reconstruction in ATLAS," ATL-PHYS-2003-031, 25 Aug 2003. 18 p.

[14] J. Marriner, "Secondary vertex fit with mass and pointing constraint (CTVMFT)," CDF Internal Note 2724 (1994).

[15] `http://www.hep.ucl.ac.uk/atlas/AtlasTesting`