# ACAT 2010

Monday 22 February 2010 - Saturday 27 February 2010

Jaipur, India

# Book of Abstracts

# Contents

Likelihood-based Particle Flow Algorithm at CDF for Accurate Energy Measurement and Identification of Hadronically Decaying Tau Leptons . . . . . 1

Unstable-particles pair production in modified perturbation theory in NNLO . . . . . 1

Interoperating AliEn and ARC for a distributed Tier1 in the Nordic countries. . . . . . 1

EU-IndiaGrid2 - Sustainable e-infrastructures across Europe and India . . . . . 2

Visual Physics Analysis - Applications in High-Energy- and Astroparticle-Physics . . . . 3

Teaching a Compiler your Coding Rules . . . . . 4

Automated Computation of One-loop Scattering Amplitudes . . . . . 4

PROOF on Demand . . . . . 4

Analysis of Photoluminescence measurement data from interdiffused Quantum Wells by Real coded Quantum inspired Evolutionary Algorithm . . . . . 5

SFrame - A high-performance ROOT-based framework for HEP analysis . . . . . 6

mc4qcd: web based analysis and visualization tool for Lattice QCD . . . . . 6

Computing at Belle II . . . . . 7

FAST PARALLELIZED TRACKING ALGORITHM FOR THE MUON DETECTOR OF THE CBM EXPERIMENT AT FAIR . . . . . 7

Building Efficient Data Planner for Peta-scale Science . . . . . 8

VIRTDOM: a dynamic Virtual Machine manager for customized execution environments. 8

Absorbing systematic effects to obtain a better background model in a search for new physics . . . . . 9

Pattern recognition and estimation methods for track and vertex reconstruction . . . . . 9

Debbie: an innovative approach for the CMS Pixel Tracker web-based configuration DB 10

Distributed parallel processing analysis framework for Belle II and Hyper Suprime-Cam 10

Parallelization of Neutron Transport Code ATES3 on BARC's Parallel System . . . . . 11

Tools to use heterogeneous Grid schedulers and storage system . . . . . 11

PROOF - Best Practices . . . . . 12

High Volume data monitoring with RootSpy . . . . . 12

Online Filtering for Radar Detection of Meteors . . . . . 13

ATLAS Second-Level Electron/Jet Neural Discriminator based on Nonlinear Independent Components . . . . . 14

TMVA - Toolkit for Multivariate Data Analysis . . . . . 14

PROOF - Status and New Developments . . . . . 15

Optimizing CMS software to the CPU . . . . . 15

Fourier Transforms as a tool for Analysis of Hadron-Hadron Collisions. . . . . . 16

Fast Parallel Ring Recognition Algorithm in the RICH Detector of the CBM Experiment at FAIR . . . . . 16

WatchMan Project - Computer Aided Software Engineering applied to HEP Analysis Code Building for LHC . . . . . 17

Optimization of Grid Resources Utilization: QoS-aware client to storage connection in AliEn . . . . . 18

ATLAS Physics Analysis Tools . . . . . 18

Recursive reduction of tensorial one-loop Feynman integrals . . . . . 19

A T3 non-grid end-user analysis model based on prior installed Grid Infrastructure. . . . . 19

BNL Batch and DataCarousel systems at BNL: A tool and UI for efficient access to data on tape with faireshare policies capabilities . . . . . 20

Status of the FORM project . . . . . 20

Classifying extremely imbalanced data sets . . . . . 21

LHC Cloud Computing with CernVM . . . . . 21

Analysis of medical images: the MAGIC-5 Project . . . . . 22

The SHUTTLE: the ALICE Framework for the extraction of the conditions Data . . . . . 22

AliEn2 and beyond . . . . . 23

How to Navigate Next Generation Programming Models for Next Generation Computer Architecture . . . . . 24

IR subtraction schemes . . . . . 24

Calculating one loop multileg processes. A program for the case of $gg \to t\bar{t} + gg$ . . . . . 25

Implementation of new WLCG services into the AliEn Computing model of the ALICE experiment before the data taking . . . . . 25

HepData - the HEP data archive reloaded . . . . . 26

The ALICE Online Data Quality Monitoring . . . . . . . . . . . . . . . . . . . . . . . . . 26

Parallel versions of the symbolic manipulation system FORM . . . . . . . . . . . . . . . 27

Feynman Integral Evaluation by a Sector decomposiTion Approach (FIESTA) . . . . . . . 27

FATRAS –A Novel Fast Track Simulation Engine for the ATLAS Experiment . . . . . . . 28

Alignment of the ATLAS Inner Detector . . . . . . . . . . . . . . . . . . . . . . . . . . 28

Parallel approach to online event reconstruction in the CBM experiment . . . . . . . . . 29

RAVEN - a Random Access, Visualisation and Exploration Network for the Analysis of Petabyte Sized Data Sets . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

Data access in the High Energy Physics community . . . . . . . . . . . . . . . . . . . . 30

Studies of the performances of an open source batch system / scheduler (TORQUE / MAUI) implemented on a middle sized GRID site. . . . . . . . . . . . . . . . . . . . . . . . . 30

Sector decomposition via computational geometry . . . . . . . . . . . . . . . . . . . . . 31

Using TurboSim for Fast Detector Simulation . . . . . . . . . . . . . . . . . . . . . . . 31

Multiple Polylogarithms and Loop Integrals . . . . . . . . . . . . . . . . . . . . . . . . 32

Automating CMS Calibrations using the WMAgent framework . . . . . . . . . . . . . . 32

"NoSQL" databases in CMS Data and Workflow Management . . . . . . . . . . . . . . . 32

Parallelization of likelihoood function data analysis software based on RooFit package . 33

The FeynSystem: FeynArts, FormCalc, LoopTools . . . . . . . . . . . . . . . . . . . . . 33

Contextualization in Practice: The Clemson Experience . . . . . . . . . . . . . . . . . . 33

Monitoring the software quality in FairRoot . . . . . . . . . . . . . . . . . . . . . . . . 34

Parallelization of the SIMD Ensemble Kalman Filter for Track Fitting Using Ct . . . . . . 34

The RooStats project . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

Two-Loop Fermionic Integrals in Perturbation Theory on a Lattice . . . . . . . . . . . . 36

An improvement in LVCT cache replacement policy for data grid . . . . . . . . . . . . . 36

Deterministic numerical box and vertex integrations for one-loop hexagon reductions . 37

Data Transfer Optimization - Going Beyond Heuristics . . . . . . . . . . . . . . . . . . 37

Scientific Computing with Amazon Web Services . . . . . . . . . . . . . . . . . . . . . 38

Data Management Panel . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

Multicore Panel . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

Tools for Dark Matter in Particle Physics and Astrophysics . . . . . . . . . . . . . . . . 39

How computing centres of Alice connect? A social network analysis of cooperative ties 39

Statistics challenges in HEP . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

Official Openning . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

High Tea . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

Application of Many-core Accelerators for Problems in Astronomy and Physics . . . . . 40

Numerical approach to Feynman diagram calculations: Benefits from new computational capabilities . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

Automation of multi-leg one-loop virtual amplitudes . . . . . . . . . . . . . . . . . . . . 41

Lattice QCD simulations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

Simulation and Visualisation Techniques . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

Statistical Methods, Multivariate Analysis and Pattern Recognition . . . . . . . . . . . . 42

Multicore Computing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

Software Development in High Energy Physics: a Critical Look . . . . . . . . . . . . . . 43

Internet Law: What Students, Professors, and Software Developers Need to Know . . . . 43

History of the ROOT System: Conception, Evolution and Experience by Rene BRUN . . 43

Computing Outside the Box: On Demand Computing & its Impact on Scientific Discovery by Ian FOSTER . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

Poster session . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

Ian FOSTER - Public Lecture . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44

Computing Technology for Physics Research Summary . . . . . . . . . . . . . . . . . . . 44

Data Analysis - Algorithms and Tools Summary . . . . . . . . . . . . . . . . . . . . . . . 44

Methodology of Computations in Theoretical Physics Summary . . . . . . . . . . . . . . 44

ACAT 2010 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

The automation of subtraction schemes for next-to-leading order calculations in QCD . . 45

New developments in event generator tuning techniques . . . . . . . . . . . . . . . . . . 45

Applying CUDA Computing Model To Event Reconstruction Software . . . . . . . . . . 45

**Tuesday, 23 February - Data Analysis - Algorithms and Tools / 0**

## Likelihood-based Particle Flow Algorithm at CDF for Accurate Energy Measurement and Identification of Hadronically Decaying Tau Leptons

**Authors:** Alexei Safonov[1]; Andrey Elagin[1]

[1] *Texas A&M University (TAMU)*

**Corresponding Author:** andrey.elagin@cern.ch

We present a new technique for accurate energy measurement of hadronically decaying tau leptons. The technique was developed and tested at CDF experiment at the Tevatron. The technique employs a
particle flow algorithm complemented with a likelihood-based method for separating contributions of
overlapping energy depositions of spatially close particles. In addition to superior energy resolution provided by the method and improved discrimination against backgrounds, this technique provides a direct estimate of the uncertainty in the energy measurement of each individual hadronic tau jet. The estimate of the likelihood of the observed detector response for a given particle hypothesis allows improving rejection against difficult light lepton backgrounds. This new technique is now being deployed to improve sensitivity of the $H \rightarrow \tau\tau$ search at the Tevatron. With appropriate adjustments, the algorithm can be further extended to the case of generic (quark or gluon) jets as well as adopted at other experiments.

**Friday, 26 February - Methodology of Computations in Theoretical Physics / 2**

## Unstable-particles pair production in modified perturbation theory in NNLO

**Author:** Maksim Nekrasov[1]

[1] *Institute for High Energy Physics*

**Corresponding Author:** maksim.nekrasov@ihep.ru

We consider pair production and decay of fundamental unstable particles in the framework of a modified perturbation theory (MPT) which treats resonant contributions of unstable particles in the sense of distributions. The cross-section of the process is calculated within the NNLO of the MPT in a model that admits exact solution. Universal massless-particles contributions are taken into consideration. The calculations are carried out by means of FORTRAN code with double precision which ensures a per mille accuracy of the computations. A comparison of the outcomes with the exact solution demonstrates an excellent convergence of the MPT series at the energies close to and above the maximum of the cross-section. Near the maximum of the cross-section a discrepancy of the NNLO approximation makes up a few per mille.

**3**

## Interoperating AliEn and ARC for a distributed Tier1 in the Nordic countries.

**Author:** Philippe Gros[1]

[1] *Lund University*

**Corresponding Author:** philippe.gros@hep.lu.se

For the intensive offline computation and storage needs of LHC, the Grid has become a necessary tool. The grid software is called middleware, and comes in different flavors. The ALICE experiment has developed AliEn, while ARC has been developed in the Nordic countries. The Nordic community has pledged to LHC a large amount of resources distributed over four countries, where the job management should be done in a unified way using ARC. We have developed an interoperation module in AliEn to this purpose, so that an group of sites can look like a single site to AliEn. A prototype has been completed and tested out of production. This talk will present implementation details of the system and its performance in tests.

**Tuesday, 23 February - Computing Technology for Physics Research / 4**

## EU-IndiaGrid2 - Sustainable e-infrastructures across Europe and India

**Author:** Alberto Masoni[1]

[1] *INFN - Cagliari*

**Corresponding Author:** alberto.masoni@ca.infn.it

EU-IndiaGrid2 - Sustainable e-infrastructures across Europe and India capitalises on the achievements of the FP6 EU-IndiaGrid project and huge infrastructural developments in India. EU-IndiaGrid2 will act as a bridge across European and Indian e-Infrastructures, leveraging on the expertise obtained by partners during the EU-IndiaGrid project. EU-IndiaGrid2 will further the continuous e-Infrastructure evolution in Europe and India, to ensure sustainable scientific, educational and technological collaboration across the two continents. In particular the Large Hadron Collider (LHC) program represents one of the unique science and research facilities to share between India and Europe in the field of Scientific Research in general and in the ICT domain in particular. The Indian partners in the project represent both the ALICE and the CMS communities actively engaged in the LHC program. The role of the EU-IndiaGrid project in this specific activity has been widely recognised within the European Commission and the Indian Government and EU-IndiaGrid2 will continue its action in sustaining this community.
The project, approved within the call FP7-INFRASTRUCTURES-2009-1, starts in January 2010 with a duration of 24 months.

**Summary**:

EU-IndiaGrid2 - Sustainable e-infrastructures across Europe and India capitalises on the achievements of the FP6 EU-IndiaGrid project and huge infrastructural developments in India. EU-IndiaGrid2 will act as a bridge across European and Indian e-Infrastructures, leveraging on the expertise obtained by partners during the EU-IndiaGrid project. EU-IndiaGrid2 will further the continuous e-Infrastructure evolution in Europe and India, to ensure sustainable scientific, educational and technological collaboration across the two continents. EU-IndiaGrid2 will leverage on the EU-IndiaGrid project achievements and the strong cooperation links established with the foremost European and Indian e-Infrastructure initiatives paving the way for successful sustainable cooperation across European and Indian e-Infrastructures. EU-IndiaGrid2 aims to achieve sustainable EU-Indian e-Infrastructure cooperation by supporting a set of relevant applications, promoting their results, as well as the benefits and impacts of e-Infrastructures for Euro-Indian collaboration. In particular EU-IndiaGrid2 will address relevant Institutions and policy makers relying on the leading role of its partners in e-Infrastructure activities both in Europe and India

Objectives:
O1) Consolidate & enhance cooperation between European and Indian e-Infrastructures for the benefit of EU-Indian collaboration in e-Science
O2) Support specific user communities in the exploitation of grid infrastructure in areas strategic for EU-Indian collaboration
O3) ensure a sustainable approach to e-Infrastructures across Europe and India through dissemination actions, meetings & workshops

O4) foster and enhance cooperation with other European Initiatives in the Asian region and worldwide

Action plan: EU-IndiaGrid2 will leverage on the EU-IndiaGrid project achievements and the strong cooperation links established with the foremost European and Indian e-Infrastructure initiatives paving the way for successful sustainable cooperation across European and Indian e-Infrastructures. EU-IndiaGrid2 aims to achieve sustainable EU-Indian e-Infrastructure cooperation by supporting a set of relevant applications, promoting their results, as well as the benefits and impacts of e-Infrastructures for Euro-Indian collaboration. In particular EU-IndiaGrid2 will address relevant Institutions and policy makers relying on the leading role of its partners in e-Infrastructure activities both in Europe and India

Support activities: The project activities are organised within four main work packages.
Liaison with other Projects & Organisations (WP2) will pursue cooperation activities with the most relevant e-Infrastructure initiatives and Institutions both in Europe and India, capitalising on the leading role of its partners in the most relevant European & Indian e-Infrastructures projects and on the close links already established with projects and relevant actors by the EU-IndiaGrid project. WP3, Operational infrastructure and interoperability support, will deliver support to the Interconnection and Interoperation of e-Infrastructures across Europe (EGEE -EGI) and India (WLCG, GARUDA and National Knowledge Network (NKN)). WP4, User communities support, will sustain a set of applications within user communities strategic for EU-Indian collaboration fostering the benefits of technology achievements on scientific collaborations. Close cooperation with WP2, WP3, WP5 will contribute to identifying and performing the necessary actions to favour optimal resources exploited by such user communities. WP5, Communication and Awareness raising will reinforce the global relevance and impact of the results achieved by the project and will contribute to increasing awareness and use of joint e-Infrastructures, identifying engaging and supporting a Euro-Indian e-infrastructure community.

User communities: EU-IndiuaGrid2 targets a number of user communities in the project:
Climate change is a worldwide concern and climate change studies are among the priorities in European and Indian research programs. In particular climate change is one of the flagship activities within the NKN program. EU-IndiaGrid2 aims to support climate change modelling studies on European and Indian e-Infrastructures thanks to the involvement of premier research groups with leading international reputations and a solid collaboration basis enhanced and strengthened in the course of EU-IndiaGrid.
High Energy Physics through the Large Hadron Collider (LHC) program represents one of the unique science and research facilities to share between India and Europe in the field of Scientific Research in general and in the ICT domain in particular. The Indian partners in the project represent both the ALICE and the CMS communities actively engaged in the LHC program. The role of the EU-IndiaGrid project in this specific activity has been widely recognised within the European Commission and the Indian Government and EU-IndiaGrid2 will continue its action in sustaining this community.
Biology and Material Science: these broad areas require computational tools and techniques spawning different disciplines: they will challenge the project in setting up and providing cross disciplinary research services. The successful work of its predecessor EU-IndiaGrid, performed in these areas, allowed the establishment and the reinforcement of relevant EU-Indian collaborations supported by premier Institutions within the Consortium.
The enlargement of such significant user communities is the key to sustainability since motivate the e-Infrastructures existence and then drive their development. EU-IndiaGrid2 will sustain a set of applications strategic for EU-Indian collaboration which can exploit the possibilities offered by network and grid infrastructures.
The project, approved within the call FP7-INFRASTRUCTURES-2009-1, starts in January 2010 with a duration of 24 months.

**Friday, 26 February - Data Analysis - Algorithms and Tools / 5**

## Visual Physics Analysis - Applications in High-Energy- and Astroparticle-Physics

**Authors:** Andreas Hinzmann[1]; Dennis Klingebiel[1]; Gero Müller[1]; Jan Steggemann[1]; Martin Erdmann[1]; Michael Brodski[1]; Robert Fischer[1]; Tatsiana Klimkovich[1]; Thomas Münzer[1]; Tobias Winchen[1]

[1] *III. Physikalisches Institut A, RWTH Aachen University, Germany*

VISPA (Visual Physics Analysis) is a novel development environment to support physicists in prototyping, execution, and verification of data analysis of any complexity. The key idea of VISPA is developing physics analyses using a combination of graphical and textual programming. In VISPA, a multipurpose window provides visual tools to design and execute modular analyses, create analysis templates, and browse physics event data at different steps of an analysis.
VISPA aims at supporting both experiment independent and experiment specific analysis steps. It is therefore designed as a portable analysis framework, supporting Linux, Windows and MacOS, with its own data format including physics objects and containers, thus allowing easy transport of analyses between different computers. All components of VISPA are designed for easy integration with experiment specific software to enable physics analysis within the same graphical tools.
VISPA has proven to be an easy-to-use and flexible development environment in high energy physics as well as in astroparticle physics analyses. In this talk, we present applications of advanced physics analyses, and thereby explain the underlying software concepts.

**Tuesday, 23 February - Computing Technology for Physics Research / 6**

## Teaching a Compiler your Coding Rules

**Author:** Axel Naumann[1]

[1] *CERN*

**Corresponding Author:** axel.naumann@cern.ch

Most software libraries have coding rules. They are usually checked by a dedicated tool which is closed source, not free, and difficult to configure. With the advent of clang, part of the LLVM compiler project, an open source C++ compiler is in reach that allows coding rules to be checked by a production grade parser through its C++ API. An implementation for ROOT's coding convention will be presented, demonstrating how to interface with clang's representation of source code, and explaining how to define rules.

**Tuesday, 23 February - Methodology of Computations in Theoretical Physics / 7**

## Automated Computation of One-loop Scattering Amplitudes

**Author:** Giovanni Ossola[1]

[1] *New York City College of Technology (CUNY)*

**Corresponding Author:** gossola@citytech.cuny.edu

The problem of an efficient and automated computation of scattering amplitudes at the one-loop level for processes with more than 4 particles is crucial for the analysis of the LHC data.
In this presentation I will review the main features of a powerful new approach for the reduction of one-loop amplitudes that operates at the integrand level. The method, also known as OPP reduction, is an important building block towards a fully automated implementation of this type of calculations. I will illustrate the existing numerical codes available for the reduction and discuss the ongoing efforts to target important issues such as stability, versatility and efficiency of the method.

**8**

## PROOF on Demand

**Author:** Anar Manafov[1]

[1] *GSI Helmholtzzentrum für Schwerionenforschung GmbH, Germany*

**Corresponding Author:** a.manafov@gsi.de

PROOF on Demand (PoD) is a set of utilities, which allows starting a PROOF cluster at user request, on any resource management system. It provides a plug-in based system, to use different job submission frontends, such as LSF or gLite WMS. PoD is fully automated and no special knowledge is required to start to use it.
Main components of PoD are pod-agent and pod-console. pod-agent provides the communication layer between a PROOF master on a local machine and PROOF workers on remote resources, possibly behind a firewall. pod-console provides a user-friendly GUI, which is used to setup, manage, and shutdown the dynamic PROOF cluster. Installation is simple and doesn't require administrator privileges, and all the processes run in user space. PoD gives users, who don't have a centrally-administrated static PROOF cluster at their institution, the possibility to enjoy the full power of interactive analysis with PROOF. It is also a very good alternative to static PROOF clusters.
PoD is a specially designed solution to provide a PROOF cluster on the fly.

**Tuesday, 23 February - Data Analysis - Algorithms and Tools / 10**

## Analysis of Photoluminescence measurement data from interdiffused Quantum Wells by Real coded Quantum inspired Evolutionary Algorithm

**Author:** Ashish Mani[1]

**Co-author:** Patvardhan C [1]

[1] *Dayalbagh Educational Institute*

**Corresponding Author:** ashish_mani@dei.ac.in

Reliable analysis of any experimental data is always difficult due to the presence of noise and other types of errors. This paper analyzes data obtained from photoluminescence measurement, after the annealing of interdiffused Quantum Well Hetrostructures, by a recently proposed Real coded Quantum inspired Evolutionary Algorithm (RQiEA). The proposed algorithm directly measures interdiffusion parameters without using Arrhenius plot. Further, the results obtained are better than those with Genetic Algorithm and Least Square Method. The RQiEA is better suited than other state of art techniques of data analysis as it uses real coding rather than binary coding and its search process is inspired by quantum computing. It has also reliably detected extrinsic interdiffusion process.
Photoluminescence is a widely used process for measurement of interdiffusion parameters in semiconductor quantum well heterostructures. This method correlates the changes of the confined energy levels (PL peak energy) into characteristic diffusion length (LD) of the quantum well structure by a linear theoretical model. The correlated LD2 is plotted against annealing time, t, to determine the interdiffusion coefficient, D (T), by using the following equation:
LD2 = 4*D(T)*t

The interdiffusion parameters viz., activation energy, Ea, and the interdiffusion prefactor, Do, are determined by using Arrhenius equation:
D(T) = Do *exp(-Ea/(K*T))
Where K is Boltzmann Constant and T is annealing temperature in Kelvin.

Evolutionary Algorithm (EA) mimics process of natural evolution. RQiEA has been designed by integrating superposition and entanglement ideas from quantum computing in EA. It uses adaptive quantum inspired rotation gates to evolve population qubits. It has been shown that QiEAs are more powerful than EAs as they can better balance Exploration and Exploitation during search process.

**Summary**:

Reliable analysis of any experimental data is always difficult due to the presence of noise and other types of errors. This paper analyzes data obtained from photoluminescence measurement, after the annealing of interdiffused Quantum Well Hetrostructures, by a recently proposed Real coded Quantum inspired Evolutionary Algorithm (RQiEA). The results obtained are better than those with Genetic Algorithm and Least Square Method.

**Tuesday, 23 February - Data Analysis - Algorithms and Tools / 11**

## SFrame - A high-performance ROOT-based framework for HEP analysis

**Authors:** Attila Krasznahorkay[1]; David Berge[2]; Johannes Haller[3]

[1] *New York University*

[2] *CERN*

[3] *Hamburg Univeristy*

**Corresponding Authors:** johannes.haller@physik.uni-hamburg.de, attila.krasznahorkay@cern.ch, david.berge@cern.ch

In a typical offline data analysis in high-energy-physics a large number of collision events are studied. For each event the reconstruction software of the experiments stores a large number of measured event properties in sometimes complex data objects and formats. Usually this huge amount of initial data is reduced in several analysis steps, selecting a subset of interesting events and observables. In addition, the same selection is applied to simulated MC events and the final results are compared to the data. A fast processing of the events is mandatory for an efficient analysis.

In this paper we introduce the SFrame package, a ROOT-based analysis framework, that is widely used in the context of ATLAS data analyses. It features (i) consecutive data reduction in multiple user-defined analysis cycles performing a selection of interesting events and observables, making it easy to calculate and store new derived event variables; (ii) a user-friendly combination of data and MC events using weighting techniques; and in particular (iii) a high-speed processing of the events. We study the timing performance of SFrame and find a highly superior performance compared to other analysis frameworks.

More information can be found at: http://sourceforge.net/projects/sframe/

**Summary**:

The structure and performance of SFrame is presented, which is a light-weight analysis-framework built around ROOT.

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 13**

## mc4qcd: web based analysis and visualization tool for Lattice QCD

**Author:** Massimo Di Pierro[1]

[1] *DePaul University*

**Corresponding Author:** mdipierro@cs.depaul.edu

mc4qcd is a web based collaboration for analysis of Lattice QCD data. Lattice QCD computations consists of a large scale Markov Chain Monte Carlo. Multiple measurements are performed at each MC step. Our system acquires the data by uploading log files, parses them for results of measurements, filters them, mines the data for required information by aggregating results in multiple forms, represents the results as plots and histograms, and it further allows refining and interaction by fitting the results. The system computes moving averages and autocorrelations, builds bootstrap samples and bootstrap errors, and allows modeling the data using Bayesian correlated constrained linear and non-linear fits. It can be scripted to allow real time visualization of results form an ongoing computation. The system is modular and it can be easily adapted to automating the workflow of other types of computations.

**Tuesday, 23 February - Computing Technology for Physics Research / 15**

# Computing at Belle II

**Authors:** Takanori Hara[1]; Thomas Kuhr[2]

[1] *KEK*

[2] *Institut fuer Experimentelle Kernphysik, KIT*

The Belle II experiment, a next-generation B factory experiment at KEK, is expected to record a two orders of magnitude larger data volume than its predecessor, the Belle experiment. The data size and rate are comparable to or more than the ones of LHC experiments and requires to change the computing model from the Belle way, where basically all computing resources were provided by KEK, to a more distributed scheme. While we adopt existing grid technologies for our baseline design, we also investigate the possibility of using cloud computing for peaking resource demands. An important task of the computing framework is to provide easy and transparent access to data and to facilitate the bookkeeping of processed files and failed jobs. To achieve this we set up a metadata catalog based on AMGA and plan to use it in a bookkeeping service that is based on concepts implemented in the SAM data handling system used at CDF and D0.
In this talk we summarize the expected Belle II performance and the resulting computing requirements and show the status and plans of the core components of the computing infrastructure.

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 16**

# FAST PARALLELIZED TRACKING ALGORITHM FOR THE MUON DETECTOR OF THE CBM EXPERIMENT AT FAIR

**Author:** Andrey Lebedev[1]

**Co-authors:** Claudia Hoehne [2]; Gennady Ososkov [3]; Ivan Kisel [2]

[1] *GSI, Darmstadt / JINR, Dubna*

[2] *GSI, Darmstadt*

[3] *JINR, Dubna*

**Corresponding Author:** andrey.lebedev@gsi.de

Particle trajectory recognition is an important and challenging task in the Compressed Baryonic Matter (CBM) experiment at the future FAIR accelerator at Darmstadt. The tracking algorithms have to process terabytes of input data produced in particle collisions. Therefore, the speed of the tracking software is extremly important for data analysis. In this contribution, a fast parallel track reconstruction algorithm which uses available features of modern processors is presented. These

features comprize a SIMD instruction set and multithreading. The first allows to pack several data items into one register and to operate on all of them in parallel thus achieving more operations per cycle. The second feature enables the routines to exploit all available CPU cores and hardware threads. This parallelized version of the tracking algorithm has been compared to the initial serial scalar version which uses a similar approach for tracking. A speed up factor of 140 was achieved (from 630 msec/event to 4.5 msec/event) for an Intel Core 2 Duo processor at 2.26 GHz.

**Thursday, 25 February - Computing Technology for Physics Research / 17**

## Building Efficient Data Planner for Peta-scale Science

**Authors:** Jerome LAURET[1]; Michal SUMBERA[2]; Michal ZEROLA[3]; Roman BARTAK[4]

[1] *BROOKHAVEN NATIONAL LABORATORY*

[2] *Nuclear Physics Inst., Academy of Sciences, Praha*

[3] *Nuclear Physics Inst., Academy of Sciences*

[4] *Faculty of Mathematics and Physics, Charles University, Praha*

**Corresponding Author:** jlauret@bnl.gov

Unprecedented data challenges both in terms of Peta-scale volume and concurrent distributed computing have seen birth with the rise of statistically driven experiments such as the ones represented by the high-energy and nuclear physics community. Distributed computing strategies, heavily relying on the presence of data at the proper place and time, have further raised demands for coordination of data movement on the road onwards achieving high performance. Massive data processing will be hardly "fair"to users and hardly using network bandwidth efficiently whenever diverse usage patterns and priorities will be involved unless we address and deal with planning and reasoning about data
movement and placement. Although there exist several sophisticated and efficient point-to-point data transfer tools, the lack of global planners and decision makers, answering questions such as "How to bring the required dataset to the user?"or "From which sources to grab the replicated data" , is for most part lacking.
We present our work and a status of the development of an automated data planning and scheduling system, ensuring fairness and efficiency of data movement by focusing on the minimal time to realize data movement (delegating the data transfer itself to existing transfer tools). Its principal keystones are self-adaptation to the network/service alteration, optimal selection of transfer channels, bottlenecks avoidance and user fair-share preservation. The planning mechanism is built on constraint based model, reflecting the restrictions from reality by mathematical constraints, using Constraint Programming and Mixed Integer Programming techniques. In this presentation, we will concentrate on
clarifying the overall system from a software engineer's point of view and present the general architecture and interconnection between centralized and distributed components of the system. While the framework is evolving toward implementing more constraints (such as CPU availability versus storage for a better planing of massive analysis and data production), the current state of our implementation in use for STAR within multi-user environment between multiple sites and services
will be presented and the benefit and consequences summarized.

**18**

## VIRTDOM: a dynamic Virtual Machine manager for customized execution environments.

**Author:** Leonello Servoli[1]

**Co-authors:** Mirko Mariotti [2]; Riccardo Cefalà [1]; Riccardo Sbirrazzuoli [2]

[1] *INFN - Sezione di Perugia*

[2] *Univ. Perugia*

**Corresponding Author:** leonello.servoli@pg.infn.it

Distributed computer systems pose a new class of problems, due to increased heterogeneity either from the hardware than from the user's request point of view.
One possible solution is to create on demand virtual working environments tailored on the user's requirements, hence the need to manage dynamically such environments. This work proposes a solution based on the use of Virtual Machines (Xen) coupled with a Virtual Machine Manager to create, destruct and migrate the virtualized working environments according to a customized policy.
The informations will be collected using a client-server mechanism, to allow the manager to deploy preconfigured Virtual Machines on the available hardware resources. When a new execution environment became active, it is automatically recognized by the Batch System Manager and is then ready to be used.
This prototype has been put in production at Perugia INFNGRID site and the first results will be presented.

**Tuesday, 23 February - Data Analysis - Algorithms and Tools / 19**

## Absorbing systematic effects to obtain a better background model in a search for new physics

**Authors:** Eilam Gross[1]; Glen Cowan[2]; Jan Erik Sundermann[3]; Sascha Caron[3]; Stephan Horner[3]

[1] *Weizmann Institute, Rehovot*

[2] *Royal Holloway, University of London*

[3] *Albert-Ludwigs-Universitaet Freiburg*

This contribution discusses a novel approach to estimate the Standard Model backgrounds based on modifying Monte Carlo predictions within their systematic uncertainties. The improved background model is obtained by altering the original predictions with successively more complex correction functions in signal-free control selections. Statistical tests indicate when sufficient compatibility with data is reached. In this way, systematic effects are absorbed into the new background model. The same correction is then applied on the Monte Carlo prediction in the signal region. Comparing this method to other background estimation techniques shows improvements with respect to statistical and systematic uncertainties. The proposed method can also be applied in other fields beyond high energy physics.

**Tuesday, 23 February - Plenary Session / 20**

## Pattern recognition and estimation methods for track and vertex reconstruction

**Author:** Rudolf Frühwirth[1]

[1] *Institute of High Energy Physics, Vienna*

**Corresponding Author:** rudolf.fruehwirth@oeaw.ac.at

The reconstruction of charged tracks and interaction vertices is an important step in the data analysis chain of particle physics experiments. I give a survey of the most popular methods that have been employed in the past and are currently employed by the LHC experiments. Whereas pattern recognition methods are very diverse and rather detector dependent, fitting algorithms offer less variety and can be applied to both track and vertex estimation with minimal changes. In particular, I trace the development from standard least-squares estimators to robust and adaptive estimators in both contexts. I end with an outlook to what I consider the most important issues to be addressed

by experiments at future colliders such as the SuperLHC, the upgraded B-factory at KEK, and the ILC.

**21**

## Debbie: an innovative approach for the CMS Pixel Tracker web-based configuration DB

**Author:** Marco Rovere[1]

**Co-author:** Menasce Dario [2]

[1] *CERN*

[2] *INFN, Milano Bicocca*

**Corresponding Author:** marco.rovere@cern.ch

The configuration of the CMS Pixel detector consists in a complex set of data that uniquely define its startup condition. Since several of these conditions are used to both calibrate the detector over time and to properly initialize it for a physics run, all these data have been collected in a suitably designed database for historical archival and retrieval. In this talk we present a description of the underlying database schema with a particular emphasis on the architecture and implementation of the web-based interface that allows for very sophisticated browsing/editing operations of detector data using a graphical representation of its topology. This interface employs state-of-art technology such as Ajax transactions, svg-based vector graphics and an extensive use of the Extjs JavaScript library. The GUI represents a novel approach to web-based interfaces, since it features a very complex set of widgets, dynamically generated on the fly upon user-demand, thus mimicking the behavior of a stand-alone program specifically designed to this extent, but avoiding portability and interactive-login issues of the latter solution.

**Thursday, 25 February - Computing Technology for Physics Research / 22**

## Distributed parallel processing analysis framework for Belle II and Hyper Suprime-Cam

**Authors:** Nobu KATAYAMA[1]; Ryosuke ITOH[1]; Sogo Mineo[2]; Soohyung Lee[3]

[1] *KEK*

[2] *University of Tokyo*

[3] *Korea University*

**Corresponding Author:** ryosuke.itoh@kek.jp

The real time data analysis at next generation experiments is a challenge because of their enormous data rate and size. The Belle II experiment, the upgraded Belle experiment, requires to manage a data amount of O(100) times the current Belle data size collected at more than 30kHz. A sophisticated data analysis is required for the efficient data reduction in the high level trigger farm in addition to the offline analysis. On the other hand, a telescope survey with Hyper Suprime-Cam at Subaru Observatory for the search of dark energy also needs to handle a large number of CCD images whose size is comparable with that of Belle II. The feed-back of the measurement parameters obtained by the real time data processing

has never been performed in the past where the parameter tuning entirely relies on an empirical method.

We are now developing a new software framework named "roobasf" to be shared both by Belle II and Hyper Suprime-Cam. The framework has the well-established software-bus architecture and the object persistency interface with ROOT IO. In order to achieve the required real-time performance, the parallel processing technique is widely used to utilize a huge number of network-connected PCs with multi-core CPUs. The parallel processing is performed not only in the trivial event-by-event manner, but also in the pipeline of the application software modules which are dynamically placed on many PCs. The object data flow over the network is implemented using the Message Passing Interface (MPI) which also provides the system-wide control scheme. The framework adopts Python as the user interface language. The detailed design and the development status of the framework is presented at the conference.

**24**

## Parallelization of Neutron Transport Code ATES3 on BARC's Parallel System

**Authors:** Anurag Gupta[1]; Kislay Bhatt[1]; Rajesh Kalmady[2]; Vibhuti Duggal[3]

[1] *Scientific Officer 'F'*

[2] *Scientific Officer 'G'*

[3] *Scientific Officer 'D'*

The most fundamental task in the design and analysis of a nuclear reactor core is to find out the neutron distribution as a function of space, direction, energy and possibly time. The most accurate description of the average behavior of neutrons is given by the linear form of Boltzmann transport equation. Due to massive number of unknowns, the solution of the transport equation imposes severe demands on computer processors / memory and requires best of numerical and computational schemes.

A code ATES3 (Anisotropic Transport Equation Solver in 3D) had been developed in BARC for the deterministic solution of 3-D steady-state neutron transport problems. The code makes use of advanced Krylov subspace based schemes for the solution. To use ATES3 for practical reactor core simulations, it has been parallelized on BARC's ANUPAM Parallel Supercomputer using Message Passing Parallel Programming model.

The most time consuming step in the ATES3 code is transport sweep, which was targeted for the parallelization. We used three different data-decomposition techniques for parallelization, namely, parallelization in angular variable, parallelization using Diagonal Sweep approach and in the third approach, we are applying optimization techniques in Diagonal Sweeping.

In the paper we discuss the above parallelization techniques and present the speed-up and efficiency figures obtained with each approach.

**Friday, 26 February - Computing Technology for Physics Research / 25**

## Tools to use heterogeneous Grid schedulers and storage system

**Author:** Mattia Cinquilli[1]

**Co-authors:** Alessandra Fanfani [2]; Daniele Spiga [3]; Eric Vaandering [4]; Fabio Farina [5]; Federica Fanzago [6]; Giuseppe Codispoti [2]; Hassen Riahi [7]; Jose Hernandez [8]; Stefano Lacaprara [6]; Stuart Wakefield [9]

[1] *INFN, Sezione di Perugia*

[2] *INFN and Universita di Bologna*

[3] *CERN*

[4] *FNAL*

[5] *INFN Milano*

[6] *INFN Padova*

[7] *INFN and Universita di Perugia*

[8] *CIEMAT*

[9] *Imperial College London*

**Corresponding Author:** mattia.cinquilli@cern.ch

The Grid approach provides an uniform access to a set of geographically distributed heterogeneous resources and services, enabling projects that would be impossible without massive computing power. Different storage projects have been developed and a few protocols are being used to interact with them such as GsiFtp and SRM (Storage Resource Manager). Moreover, during last few years different Grid projects have developed different middleware such as EGEE, OSG, NorduGrid and each one typically implements its own interface and workflow. For a user community which needs to work through the Grid, interoperability is a key concept. To handle different Grid interfaces, the resource heterogeneity and different workflows, in a really transparent way, we have developed two modular tools: BossLite and Storage Element API. These deal with different Grid schedulers and storage systems respectively, by providing a uniform standard interface that hides the differences between the systems they interact with. BossLite transparently interacts with different Grid systems, working as a layer between an application and the middleware. Storage Element API implements and manages the operations that can be performed with the different protocols used in the main Grid storage systems. Both the tools are already being used in production in the CMS computing tools for distributed analysis and Monte Carlo production. In this paper we show their implementation, how they are used and performance results.

**Friday, 26 February - Computing Technology for Physics Research / 26**

# PROOF - Best Practices

**Author:** Fons Rademakers[1]

**Co-author:** Gerardo Ganis [1]

[1] *CERN*

With PROOF, the parallel ROOT Facility, being widely adopted for LHC data analysis, it becomes more and more important to understand the different parameters that can be tuned to make the system perform optimally. In this talk we will describe a number of "best practices" to get the most out of your PROOF system, based on feedback from several pilot setups. We will describe different cluster configurations (CPU, memory, network, HDD, SSD), PROOF and xrootd configuration options, running a dedicated PROOF system or using PROOF on Demand (PoD) on batch systems. This talk will be beneficial for people setting up Tier-3/4 analysis clusters.

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 27**

# High Volume data monitoring with RootSpy

**Author:** David Lawrence[1]

**Co-author:** James Codega [2]

[1] *Jefferson Lab*

[2] *Virginia Governor's School*

**Corresponding Author:** davidl@jlab.org

The GlueX experiment will gather data at up to 3GB/s into a level-3 trigger farm, a rate unprecedented at Jefferson Lab. Monitoring will be done using the cMsg publish/subscribe system to transport ROOT objects over the network using the newly developed RootSpy package. RootSpy can be attached as a plugin to any monitoring program to "publish" its objects on the network without modification to the original code. A description of the RootSpy package will be presented with details of the pub/sub model it employees for ROOT object distribution. Data rates obtained from tests using multi-threaded monitoring programs will also be shown.

**Tuesday, 23 February - Data Analysis - Algorithms and Tools / 28**

## Online Filtering for Radar Detection of Meteors

**Author:** Gustavo Alves[1]

**Co-authors:** Christina Vianna [1]; Eric LEITE [1]; Fernando Marroquim [1]; Helio Takai [2]

[1] *Federal University of Rio de Janeiro*

[2] *Brookhaven National Lab*

**Corresponding Author:** seixas@lps.ufrj.br

The penetration of a meteor on Earth's atmosphere results on the creation of an ionized trail, able to produce the forward scattering of VHF electromagnetic waves. This fact inspired the RMS (Radio Meteor Scatter) technique, which consists in the meteor detection using passive radar. Considering the characteristic of continuous acquisition inherent to the radar detection technique and the generation of a significant amount of data, composed mainly of background noise, an online filtering system is very attractive. Therefore, this work addresses the development of algorithms for online automatic detection of these signals. In time-domain, the optimal filtering technique is applied. The model assumes that the received signal is masked by additive noise and both signal and noise statistics are used to design a linear filter that maximizes the signal-to-noise ratio. This filter is known as the matched-filter, as detection is performed by correlating the incoming signal with replicas of the target signal components in the receiver end. In frequency-domain, two possibilities are being studied using Short-time Fast Fourier Transform: a narrowband demodulation, which basically consists in performing demodulation in filtered data in order to obtain only the envelope of the signal, and cumulative power spectrum analysis. Demodulation is attractive, as phase delays are produced by the reflection of VHF wave to the various points in the meteors trails and the different paths the traveling wave finds between the transmitting and receiving antennas. The cumulative spectral power is obtained from integrating the power spectral density function, which drastically reduces the noise effect. Sets of experimental data are being analyzed and preliminary results of these techniques with their current status of development will be shown.

**Summary**:

This work addresses the development of algorithms for online automatic detection of meteors using radar technique. In time-domain, the optimal filtering technique is applied. In frequency-domain, two possibilities are being studied using Short-time Fast Fourier Transform: narrowband demodulation and cumulative power spectrum analysis.

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 29**

# ATLAS Second-Level Electron/Jet Neural Discriminator based on Nonlinear Independent Components

**Authors:** Eduardo Simas[1]; Jose Seixas[2]; Luiz Caloba[2]

[1] *Federal Institute of Bahia, Brazil*

[2] *Federal University of Rio de Janeiro*

**Corresponding Author:** seixas@lps.ufrj.br

The ATLAS online filtering (trigger) system comprises three sequential filtering levels and uses information from the three subdetectors (calorimeters, muon system and tracking). The electron/jet channel is very important for triggering system performance as interesting signatures (Higgs, SUSY, etc.) may be found efficiently through decays that produce electrons as final-state particles. Electron/jet separation relies very much on calorimeter information, which, in ATLAS, is segmented into seven layers. Due to differences both in depth and cell granularity of these layers, trigger algorithms may benefit from performing feature extraction at the layer level.
This work addresses the second level (L2) filtering restricted to calorimeter data. Particle discrimination at L2 is split into two phases: feature extraction, in where detector information is processed aiming at extracting a compact set of discriminating variables, and an identification step, where particle discrimination is performed over these relevant variables.
The Neural Ringer is an alternative electron/jet L2 discriminator. Through Neural Ringer, the feature extraction is performed by building up concentric energy rings from a Region of Interest (RoI) data. At each calorimeter layer, the hottest (most energetic) cell is defined as the first ring, and the following rings are formed around it, so that all cells belonging to a ring have their sampled energies added together and normalized. A total of 100 ring sums fully describes the ROI. Next, a supervised neural classifier, fed from the ring-structure, is used for performing the final identification.
Independent Component Analysis (ICA) is a signal processing technique that aims at finding linear projections (s=Ax) of the multidimensional input data (x) in a way that the components of s (also called sources) are statistically independent (or at least as independent as possible). The nonlinear extension of ICA (NLICA) provides a more general formulation, as the sources are assumed to be generated by a nonlinear model: s=F(x), where F(.) is a nonlinear mapping. The Post-nonlinear (PNL) mixing model is a class of NLICA model that restricts the nonlinear mapping to a cascaded structure, which comprises a linear mapping followed by component-wise nonlinearities (cross-channel nonlinearities are not allowed).
In this work, a modification on the Neural Ringer discriminator is proposed by applying the PNL model to the ring-structure for both feature extraction and signal compaction. In order to cope with different characteristics of each calorimeter layer, here the feature extraction procedure is performed in a segmented way (at the layer level). The neural discriminator is then fed from the estimated nonlinear independent components. The proposed algorithm is applied to different L2 datasets. Compared to the Neural Ringer, the proposed approach reduces the number of inputs for the neural classifier (contributing to reduce the computational requirements) and also produces higher discrimination performance.

**Summary**:

This work addresses the second level (L2) filtering of ATLAS detector restricted to calorimeter data. It is proposed a modification on the Neural Ringer algorithm, which is an alternative for electron/jet discrimination. A nonlinear (PNL) independent component analysis model is applied to a ring-structure calorimeter data description for both feature extraction and signal compaction. A neural discriminator is then fed from the estimated nonlinear independent components. The proposed algorithm is successfully applied to different L2 datasets.

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 31**

# TMVA - Toolkit for Multivariate Data Analysis

**Authors:** Andreas Hoecker[1]; Eckhard von Toerne[2]; Helge Voss[3]; Jan Therhaag[2]; Joerg Stelzer[4]; Peter Speckmayer[1]

[1] *CERN, Switzerland*

[2] *Bonn University, Germany*

[3] *MPI Munich, Germany*

[4] *DESY, Germany*

**Corresponding Author:** joerg.stelzer@cern.ch

At the dawn of LHC data taking, multivariate data analysis techniques have become the core of many physics analyses. TMVA provides easy access to sophisticated multivariate classifiers and is widely used to study and deploy these for data selection. Beyond classification, most multivariate methods in TMVA perform regression optimization which can be used to predict data corrections, e.g. for calibration or shower corrections. The tightening of the integration with ROOT provides a common platform for discussion between the user community and the TMVA devolopers. The talk gives an overview of the new features in TMVA such as regression, multi-class classification and cathegorization, the extented pre-processing capabilities, and planned further developments.

**32**

## PROOF - Status and New Developments

**Author:** Gerardo GANIS[1]

**Co-author:** Fons RADEMAKERS [1]

[1] *CERN*

**Corresponding Author:** gerardo.ganis@cern.ch

The Parallel ROOT Facility, PROOF, is an extension of ROOT enabling interactive analysis of large sets of ROOT files in parallel on clusters of computers or many-core machines. PROOF provides an alternative to the traditional batch-oriented exploitation of distributed computing resources. The PROOF dynamic approach allows for better adaptation to the varying and unpredictable work-load during the end-user analysis phase.

PROOF is currently being widely adopted as Tier-3 solution by the LHC experiments. The ALICE collaboration is also using PROOF for prompt calibration, reconstruction and analysis of selected samples for very fast QA response.

In this talk we will present the many new developments and performance enhancements made during the last year and we will discuss the issues we are currently working on.

**Friday, 26 February - Computing Technology for Physics Research / 36**

## Optimizing CMS software to the CPU

**Author:** Peter Elmer[1]

[1] *PRINCETON UNIVERSITY*

**Corresponding Author:** peter.elmer@cern.ch

CMS is a large, general-purpose experiment at the Large Hadron Collider (LHC) at CERN. For its simulation, triggering, data reconstruction and analysis needs, CMS collaborators have developed many millions of lines of C++ code, which are used to create applications run in computer centers around the world. Maximizing the performance and efficiency of the software is highly desirable in order to maximize the physics results obtained from the available computing resources.

In the past the code optimization effort in CMS has focused on improving algorithms, basic C++ issues, excessive dynamic memory use and memory footprint. Optimizing software today, on modern multi-/many-core 64bit CPU's and their memory architectures, requires however a more sophisticated approach. This presentation will summarize efforts in CMS to understand how to properly optimize our software for maximum performance on modern hardware. Experience with various tools, lessons learned and concrete results achieved will be described.

**Friday, 26 February - Data Analysis - Algorithms and Tools / 37**

## Fourier Transforms as a tool for Analysis of Hadron-Hadron Collisions.

**Authors:** James William Monk[1]; Mario Campanelli[1]

[1] *Department of Physics and Astronomy - University College London*

**Corresponding Author:** james.william.monk@cern.ch

Hadronic final states in hadron-hadron collisions are often studied by clustering final state hadrons into jets, each jet approximately corresponding to a hard parton. The typical jet size in a high energy hadron collision is between 0.4 and 1.0 in eta-phi. On the other hand, there may be structures of interest in an event that are of a different scale to the jet size. For example, to a first approximation the underlying event is a uniform emission of radiation spanning the entire detector, colour connection effects between hard partons may fill the region between a jet and the proton remnant and hadronisation effects may extend beyond the jets. We consider the possibility of performing a Fourier decomposition on individual events in order to produce a power spectrum of the transverse energy radiated at different angular scales. We attempt to identify correlations in the emission of radiation over distances ranging from the full detector size to approximately 0.2 in eta-phi.

**Summary**:

In a sample of Monte Carlo di-jet events we find that taking the Fourier transform of the distribution of radiation in each event shows some interesting features, even in the simplest one-dimensional transformation in the phi direction. Such features include the suppression of certain of the Fourier coefficients when radiation between the di-jets is suppressed and the effect of the underlying event on the smallest of the Fourier coefficients.

**Friday, 26 February - Data Analysis - Algorithms and Tools / 39**

## Fast Parallel Ring Recognition Algorithm in the RICH Detector of the CBM Experiment at FAIR

**Author:** Semen Lebedev[1]

**Co-authors:** Claudia Hoehne [2]; Gennady Ososkov [3]; Ivan Kisel [2]

[1] *GSI, Darmstadt / JINR, Dubna*

[2] *GSI, Darmstadt*

[3] *JINR, Dubna*

**Corresponding Author:** s.lebedev@gsi.de

The Compressed Baryonic Matter (CBM) experiment at the future FAIR facility at Darmstadt will measure dileptons emitted from the hot and dense phase in heavy-ion collisions. In case of an electron measurement, a high purity of identified electrons is required in order to suppress the background. Electron identification in CBM will be performed by a Ring Imaging Cherenkov (RICH) detector and Transition Radiation Detectors (TRD).
Very fast data reconstruction is extremely important for CBM because of the huge amount of data which has to be handled. In this contribution, the parallel ring recognition algorithm is presented. Modern CPUs have two features, which enable parallel programming. First, the SSE technology allows using the SIMD execution model. Second, multi core CPUs enable to use multithreading. Both features have been implemented in the ring reconstruction of the RICH detector. A speed up factor of 20 has been achieved (from 750 ms/event to 38 ms/event) for an Intel Core 2 Duo processor at 2.13 GHz.

**Friday, 26 February - Data Analysis - Algorithms and Tools / 41**

## WatchMan Project - Computer Aided Software Engineering applied to HEP Analysis Code Building for LHC

**Authors:** Renaud Bruneliere[1]; Riccardo Maria Bianchi[1]; Sascha Caron[1]

[1] *Physikalisches Institut-Albert-Ludwigs-Universitaet Freiburg-Unk*

A lot of code written for high-level data analysis has many similar
properties, e.g. reading out the data of given input files, data
selection, overlap removal of physical objects, calculation of basic
physical quantities and the output of the analysis results.
Because of this, too many times, writing a new piece of code, one starts copying and pasting from old code, modyfing it then for specific purposes, ending up with a plethora of classes to maintain, debug and validate.
Moreover nowadays the complexity of software frameworks of HEP experiments needs that the user gets many technical details before starting writing the code.

Writing such code for each new analysis is error prone and time
consuming.

A solution of this problem is WatchMan, a "data analysis construction
kit" and highly automated analysis code generator.

WatchMan takes as inputs user-settings from a GUI or from a text-like steering
file, and in few easy steps it dynamically generates the complete analysis code, ready to be run over data, locally or on the GRID.

The package has been implemented in Python and C++, using CASE (Computer Aided Software Engineering) principles.

As a first example we interfaced the tool to the framework of the ATLAS experiment and it has been used for various analyses in ATLAS by several users.

The package is nevertheless independent of the experimental framework, and modular interfaces will be provided for other experiments as well.

**Summary**:

- Main Idea: Why Automating the Analysis Code Building? Concept, needs and advantages.
- Layout of the package
- User Interface: GUI, steering file, web interface
- Customization by the User
- Dynamical Code Building
- Characteristics of the Generated Code
- Validation of the Generated Code
- Running the generated code locally and on the GRID
- Output Ntuple file storing the results from the generated code
- Modular Interfaces to Software Frameworks of HEP Experiments
- Examples of Real Life Code Building: from realistic HEP analysis ideas to Analysis Code in few easy steps.

**Friday, 26 February - Computing Technology for Physics Research** / **43**

## Optimization of Grid Resources Utilization: QoS-aware client to storage connection in AliEn

**Authors:** Costin Grigoras[1]; Latchezar Betev[1]; Pablo Saiz[1]; Steffen Schreiner[1]

[1] *CERN*

**Corresponding Author:** costin.grigoras@cern.ch

In a World Wide distributed system like the ALICE Environment (AliEn) Grid Services, the closeness of the data to the actual computational infrastructure denotes a substantial difference in terms of resources
utilization efficiency. Applications unaware of the locality of the data or the status of the storage environment can waste network bandwidth in case of slow networks or fail accessing data from remote or inoperational storage elements. In this paper we present an approach to QoS-aware client to storage connection by introduction of a periodically updated Storage Element Rank Cache. Based on the MonALISA monitoring framework, a Resource Discovery Broker is continuously assessing the status of all available Storage Elements in the AliEn Grid. Combining availability with network topology information, rated lists of Storage Elements are offered to any client requesting access to remote data. The lists are centrally
cached by AliEn and filtered in the course of user-based authorization and requested QoS flags. This approach shows significant improvements towards an optimized storage and network resource utilization and enhances the client resilience in case of failures.

**48**

## ATLAS Physics Analysis Tools

**Author:** Roger Jones[1]

[1] *Physics Department-Lancaster University-Unknown*

**Corresponding Author:** roger.jones@cern.ch

The ATLAS experiment at the Large Hadron Collider is expected to start colliding proton beams in September 2009. The enormous amount of data produced (˜1PB per year) poses a great challenge to the ATLAS computing. ATLAS will search for the Higgs boson and Physics beyond the standard model. In order to meet this challenge, a suite of common Physics Analysis Tools (PAT) has been developed as part of the Physics Analysis software project. These tools run within the ATLAS software framework, ATHENA, covering a wide range of applications. There are tools responsible for event selection based on analysed data and detector quality information, tools responsible for specific physics analysis operations including data quality monitoring and physics validation, and complete analysis toolkits (frameworks) with the goal to aid the physicist to perform his analysis hiding the details of the ATHENA framework.

**Tuesday, 23 February - Methodology of Computations in Theoretical Physics / 52**

## Recursive reduction of tensorial one-loop Feynman integrals

**Author:** Tord Riemann[1]

[1] *DESY*

**Corresponding Author:** tord.riemann@desy.de

A new reduction of tensorial one-loop Feynman integrals with massive and massless propagators to scalar functions is introduced. The method is recursive:
n-point integrals of rank R are expressed by n-point and (n-1)-point integrals of rank (R-1).
The algorithm is realized in a Fortran package.

**55**

## A T3 non-grid end-user analysis model based on prior installed Grid Infrastructure.

**Authors:** Andre Maier[1]; Fabrizio Furano[2]; Renato Santana[3]; VInicius Bucard De Castro[3]; roberto santinelli[1]

[1] *CERN/IT/GS*

[2] *Conseil Europeen Recherche Nucl. (CERN)*

[3] *CBPF*

**Corresponding Author:** roberto.santinelli@cern.ch

An unprecedented amount of data will soon come out of CERN's Large Hadron Collider (LHC). Large user communities will immediately demand data access for physics analysis. Despite the Grid and the distributed infrastructure allowing geographically distributed data mining and analysis, there will be an important concentration of user analysis activities where the data resides, nullifying, to some extent, the grid paradigm itself. The LHCb (Large Hadron Collider beauty) experiment's computing model envisages data distribution to be restricted to selected centers, known as Tier-1 centers. More general, due to the limited storage capability, none of the LHC experiment's computing models envisages the distribution of all data across all sites. Driven by the need to avoid unnecessary over-usage of a few sites where data resides and also by the need to exploit storage facilities at non Tier-1sites (e.g Tier-2 and Tier-3 sites), this work proposes a model to copy, on demand, data from grid centers for local usage. This will allow tapping into storage facilities, otherwise unused.

Once available on the site, the data is used by the institute's scientific community to perform a local analysis. This can be done by using dedicated computing resources, accessible via special site batch system queues already in place via the site's middleware installation. Through this solution, local physics communities will be in the position to define their own priorities by running on their own resources and, at the same time, the risk to have crowded batch queues on remote systems (e.g. the LSF at CERN) are minimized. Conscious of the need to keep a consistent interface for the end-user analysis in both the LHCb and ATLAS user communities, some work and studies have been done to integrate it and customize the Ganga interface to submit local non-grid jobs. Ganga allows, for the user, to transparently change on where to submit the job, either the local cluster or the Grid, without the need to change the job description.

```
Finally, this paper presents a first working prototype, proof of concept of a new model for the end
```

**Tuesday, 23 February - Computing Technology for Physics Research / 56**

## BNL Batch and DataCarousel systems at BNL: A tool and UI for efficient access to data on tape with faireshare policies capabilities

**Authors:** David YU[1]; Jerome LAURET[1]

[1] *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** jlauret@bnl.gov

The BNL facility, supporting the RHIC experiments as its Tier0 center and thereafter the Atlas/LHC as a Tier1 center had to address early the issue of efficient access to data stored to Mass Storage. Random use destroys access performance to tape by causing too frequent, high latency and time consuming tape mount and dismount. Coupled with a high job throughput from multiple RHIC experiments, in the early 2000, the experimental and facility teams were forced to consider ingenuous approaches. A tape access "batch"system integrated to the production system was first developed, based on the initial OakRidge National Lab (ORNL) Batch code. In parallel, a highly customizable layer and UI known as the DataCarousel was developed in-house to provide multi-user fairshare with group and user level policies controlling the sharing of resources. The simple UI, based on a perl module, allowed to create user helper script to restore datasets on disks as well as had all the features necessary to interface with higher level storage aggregation solutions. Hence, beyond the simple access at data production level, the system was also successfully used in support of numerous data access tools such as interfacing with the Scalla/Xrootd MSS plugin back end, similarly the dCache back end access to MSS. Today, all RHIC and Atlas experiments use a combination of the Batch system and the Datacarousel following a 10 years search for efficient use of resources.

In 2005, BNL's HPSS team decided to enhance the new features such as improve the HPSS resource management, enhance the visibility of realtime staging activities, statistics of historical data for performance analysis. BNL Batch provides dynamic HPSS resource management and scheduled read job efficiently while the staging performance can still be further optimized in user level using the DataCarousel to maximize the tape staging performance (sorting by tape while preserving fareshareness policies).

In this presentation, we will present an overview of our system and development and share the findings of our efforts.

**Summary**:

Efficient access to data on tape tool & UI interfacing with storage aggregator for coordinated, fair and policy driven data access.

**Tuesday, 23 February - Methodology of Computations in Theoretical Physics / 57**

# Status of the FORM project

**Author:** Irina Pushkina[1]

[1] *NIKHEF*

**Corresponding Author:** irinap@nikhef.nl

Currently there is a lot of activity in the FORM project. There is much progress on making it open source. Work is done on simplification of lengthy formulas and routines for dealing with rational polynomials are under construction. In addition new models of parallelization are being studied to make optimal use of current multi-processor machines.

**Tuesday, 23 February - Data Analysis - Algorithms and Tools / 60**

# Classifying extremely imbalanced data sets

**Author:** Markward Britsch[1]

**Co-authors:** Michael Schmelling [2]; Nikolai Gagunashvili [3]

[1] *Max-Planck-Institut fuer Kernphysik (MPI)-Unknown-Unknown*

[2] *Max-Planck-Institut fuer Kernphysik (MPI)*

[3] *University of Akureyri*

**Corresponding Author:** markward@mpi-hd.mpg.de

Imbalanced data sets containing much more background than signal instances are very common in particle physics, and will also be characteristic for the upcoming analyses of LHC data. Following up the work presented at ACAT 2008, we use the multivariate technique presented there (a rule growing algorithm with the meta-methods bagging and instance weighting) on much more imbalanced data sets, especially a selection of D0 decays without the use of particle identification. It turns out that the quality of the result strongly depends on the number of background instances used for training. We discuss methods to exploit this in order to improve the results significantly, and how to handle and reduce the size of large training sets without loss of result quality in general. We will also comment on how to take into account statistical fluctuation in receiver operation curves (ROC) for comparing classifier methods.

**Tuesday, 23 February - Plenary Session / 61**

# LHC Cloud Computing with CernVM

**Authors:** Ben Segal[1]; Predrag Buncic[1]

**Co-authors:** Artem Harutyunyan [2]; Carlos Aguado Sanchez [1]; David Garcia Quintas [1]; David Weir [3]; Jakob Blomer [1]; Jarno Rantala [4]; Pere Mato [1]; Yushu Yao [5]

[1] *CERN*

[2] *Yerevan Physics Institute, Armenia*

[3] *Imperial College, London, UK*

[4] *Tampere University of Technology, Finland*

[5] *Lawrence Berkeley National Laboratory, USA*

**Corresponding Author:** b.segal@cern.ch

Using virtualization technology, the entire application environment of an LHC experiment, including its Linux operating system and the experiment's code, libraries and support utilities, can be incorporated into a virtual image and executed under suitable hypervisors installed on a choice of target host platforms.

The Virtualization R&D project at CERN is developing CernVM, a virtual machine designed to support the full range of LHC physics computing on a wide range of hypervisors and platforms including end-user laptops, Grid and cluster nodes, volunteer PC's running BOINC, and nodes on the Amazon Elastic Compute Cloud (EC2). CernVM interfaces to the LHC experiments' code repositories by means of a specially tuned network file system CVMFS, ensuring complete compatibility of the application with the developers' native version. CernVM provides mechanisms to minimize virtual machine image sizes and to keep images efficiently up to date when code changes.

CernVM also provides interfaces to the LHC experiments' job submission systems and workload management systems (e.g. ATLAS/PanDA, LHCb/DIRAC, ALICE/Alien), allowing clouds of CernVM-equipped worker nodes to be accessed by the experiments without changing their job production procedures.

Currently supported clouds include Amazon EC2, private clusters, Tier3 sites, and a cloud of BOINC volunteer PC's which represents a very large potential resource, so far untapped by the LHC experiments.

This paper presents the current state of development of CernVM support for LHC cloud computing.

**Tuesday, 23 February - Plenary Session / 62**

## Analysis of medical images: the MAGIC-5 Project

**Author:** Piergiorgio Cerello[1]

[1] *INFN - TORINO*

**Corresponding Author:** cerello@to.infn.it

The MAGIC-5 Project focuses on the development of analysis algorithms for the automated detection of anomalies in medical images, compatible with the use in a distributed environment.
Presently, two main research subjects are being addressed: the detection of nodules in low-dose high-resolution lung computed tomographies and the analysis of brain MRIs for the segmentation and classification of the hyppocampus as an early marker of the Alzheimer's disease.
MAGIC-5 started as a spin-off of high energy physics software development and involves a community of developers in constant contact with - some of them also involved in - HEP projects.
The most relevant results will be presented and discussed, together with a new model, based on virtual ant colonies, for the segmentation of complex structures. The possible used of such a model in HEP is addressed.

**63**

## The SHUTTLE: the ALICE Framework for the extraction of the conditions Data

**Author:** Chiara Zampolli[1]

**Co-authors:** Alberto Colla [2]; Federico Carminati [3]

[1] *CERN & CNAF-INFN*

[2] *Universita' di Roma Sapienza and INFN Rome*

[3] *CERN*

**Corresponding Author:** chiara.zampolli@cern.ch

C. Zampolli for the ALICE Collaboration.

ALICE will collect data at a rate of 1.25 GB/s during heavy-ion runs, and of 100 MB/s during p-p data taking. In a standard data taking year, the expected total data volume is of the order of 2PB. This includes raw data, reconstructed data, and the conditions data needed for the calibration and the alignment of the ALICE detectors, on top of simulated data. The raw data produced in the DAQ system are stored in the Grid, in order to be subsequently reconstructed. On its turn, the conditions data needed for reconstruction should be extracted from the raw data themselves. The Shuttle is the ALICE Online-Offline framework that handles and coordinates among the 18 ALICE detectors and the 5 online systems (DAQ, DCS, ECS, HLT and Trigger) the gathering, processing, and publication of the conditions data. The online systems and their databases are subject to stringent access control and have very limited outside exposure. The Shuttle is providing an interface between the protected online world and the external computing resources. All collected conditions data are exported on the Grid, thus making them accessible for the reconstruction and analysis. The Shuttle conditions data handling process consists of copying the data produced by the online systems for each subdetector (in whatever format), preprocessing them (e.g. performing consolidation, fitting…), reformatting them in ROOT files, and storing them in the Offline Condition Database (OCDB) in the Grid. The reconstruction of a given run is started automatically when the Shuttle processing of that run is finished, including the storage of the conditions data in the Grid. A constant monitoring of the Shuttle operation is performed through a MonALISA service. The talk describes the features of the Shuttle framework. The performance of such a complex system during the commissioning phase and at the LHC startup is presented. Operational statistics, issues, and problems encountered are discussed.

**64**

# AliEn2 and beyond

**Authors:** Alina Grigoras[1]; Andreas Peters[1]; Costin Grigoras[1]; Fabrizio Furano[1]; Federico Carminati[1]; Latchezar Betev[1]; Olga Datskova[1]; Pablo Saiz[1]; Patricia Mendez[1]; Sehoon Lee[2]; Stefano Bagnasco[3]; Steffen Schreiner[1]

[1] *CERN*

[2] *KISTI*

[3] *INFN Torino*

**Corresponding Author:** pablo.saiz@cern.ch

At the time of this conference, the ALICE experiment will have already data from the LHC accelerator at CERN. ALICE uses AliEn to be able to distribute and analyze all this data among the more than eighty sites that participate in the collaboration. AliEn is a system that allows the use of distributed computing and storage resources all over the world. It hides the differences between the heterogeneous components, providing a simple and intuitive interface to the end users. AliEn is also used by other high energy physics experiments like CBM and PANDA.

AliEn has been improved over the last eight years to be able to cope with computing projects like the ones of the LHC experiments. During this time, a number of exercises of increasing scale and complexity have been performed to evaluate if AliEn would be up to the challenge. All those exercises helped to identify the components that needed to be improved. Finally, the time for tests and exercises is over. This contribution will describe if AliEn was ready for the data taking, and the main lessons that were learned once the detector started collecting data.

**Summary**:

S. Bagnasco, L. Betev, F. Carminati, O. Datskova, F. Furano, A. Grigoras, C.

Grigoras, S. Lee, P. Mendez Lorenzo, A. Peters, P. Saiz, S. Schreiner

bagnasco@to.infn.it bagnasco@to.infn.it, Alina Gabriela Grigoras alina.gabriela.grigoras@cern.ch, Costin Grigoras Costin.Grigoras@cern.ch, Federico Carminati Federico.Carminati@cern.ch, Olga Vladimirovna Datskova olga.vladimirovna.datskova@cern.ch, Sehoon Lee sehooi@gmail.com, Patricia Mendez Lorenzo Patricia.Mendez@cern.ch, Andreas Joachim Peters Andreas.Joachim.Peters@cern.ch, Steffen Schreiner Steffen.Schreiner@cern.ch, Fabrizio Furano Fabrizio.Furano@cern.ch
References: 4AFD950B.1070500@cern.ch

**Thursday, 25 February - Plenary Session / 65**

## How to Navigate Next Generation Programming Models for Next Generation Computer Architecture

**Author:** Anwar Ghuloum[1]

[1] *Intel Corporation*

**Corresponding Author:** anwar.ghuloum@intel.com

Something strange has been happening in the slowly evolving, placid world of high performance computing. Software and hardware vendors have been introducing new programming models at a breakneck pace. At first blush, the proliferation of parallel programming models might seem confusing to software developers, but is it really surprising? In fact, programming models have been rapidly evolving for the better part of two decades, thanks in no small part to the boom in web-based application development frameworks and tools. The fundamental forces driving this are twofold: the first is the importance of domain-specific specialization and optimizations productively use modern hardware infrastructure and the second a base of software developers that is better able to adapt to these programming models and use "the right tool for the job". They key question that remains which must be answered by vendors of these tools is: What tool is right for you?
Intel provides a broad set of programming tools and programming models that is a microcosm of the broader diversity available in the software ecosystem. I will discuss how these tools and programming models relate and interoperate with each other in a way that developers can use to navigate the their choices. I will pay particular attention to our work in adding data parallelism in C++ via Intel's Ct technology in ways that eliminate the traditional modularity tax associated with C++ frameworks. I will also show how this work has been applied in particles physics workloads.

**Friday, 26 February - Methodology of Computations in Theoretical Physics / 67**

## IR subtraction schemes

**Author:** Paolo Bolzoni[1]

[1] *DESY*

**Corresponding Author:** paolo.bolzoni@desy.de

To compute jet cross sections at higher orders in QCD efficiently one has to deal with infrared divergences. These divergences cancel out between virtual and real corrections once the phase space integrals are performed. To use standard numerical integration methods like Monte Carlo the divergences' cancellation must be performed explicitly. Usually this is done constructing appropriate counterterms which are integrated over the unresolved region of the phase space. We will show some new approches to the infrared subtraction techniques for computing NNLO jet cross sections in QCD and the future possible phenomenological applications.

**Thursday, 25 February - Methodology of Computations in Theoretical Physics / 68**

## Calculating one loop multileg processes. A program for the case of $gg \to t\bar{t} + gg$

**Author:** Theodoros Diakonidis[1]

[1] *DESY,Zeuthen*

**Corresponding Author:** theodoros.diakonidis@desy.de

Processes with more than 5 legs are added to experimentalists' wish list for a long time
now. This study is targeted to the NLO qcd corrections of such processes in the LHC.
Many Feynman diagrams are contributing, including those with five- and
six-point functions.
A Fortran code for the numerical calculation of one-loop corrections for the process
$gg \to t\bar{t} + gg$ is reviewed.
A variety of tools like Diana, Form, Maple, Fortran are used in combination.

**Thursday, 25 February - Computing Technology for Physics Research / 69**

## Implementation of new WLCG services into the AliEn Computing model of the ALICE experiment before the data taking

**Authors:** Alina Grigoras[1]; Andreas Joachim Peters[2]; Costin Grigoras[1]; Fabrizio Furano[2]; Federico Carminati[1]; Latchezar Betev[1]; Olga Datskova[3]; Pablo Saiz[4]; Patricia Mendez Lorenzo[4]; Stefano Bagnasco[5]; Steffen SCHREINER[1]

[1] *CERN PH/AIP*

[2] *CERN IT/DM*

[3] *CERN PH/UAI*

[4] *CERN IT/GS*

[5] *INFN/Torino*

**Corresponding Author:** patricia.mendez@cern.ch

By the time of this conference the LHC ALICE experiment at CERN will have collected a significant amount of data. To process the data that will be produced during the life time of the LHC, ALICE has developed over the last years a distributed computing model across more than 90 sites that build on the overall WLCG (World-wide LHC Computing Grid) service. ALICE implements the different Grid services provided by the gLite middleware into the experiment computing model.

During the period 2008-2009 the WLCG project has deployed new versions of some services which are crucial for the ALICE computing as the gLite3.2 VOBOX, the CREAM-CE and the gLite3.2 WMS. In terms of Computing systems, the current LCG-CE used by the four LHC experiments is about

to be deprecated in benefit of the new CREAM service (Computing Resource Execution And Management). CREAM is a lightweight service created to handle job management operations at the CE level. It is able to accept requests both via the gLite WMS service and also via direct submission for transmission to the local batch system.
This flexible duality provides the users with a large level of freedom to adapt the service to their own computing models, but at the same time it requires a careful follow up of the requirements and tests of the experiments to ensure that their needs are fulfilled before real data taking.
ALICE has been the first Grid community to implement the CREAM into the experiment computing model and to test it to a production level. Since 2008 ALICE is providing the CERN Grid deployment team and the CREAM developers with important feedback which have lead to the identification of important bugs and issues solved before the real data taking. In addition ALICE has been also leader in testing and implementing other generic services as the gLite3.2 VOBOX and WMS before their final deployment.

In this talk we present a summary of the ALICE experiences by using these new services also including the testing results of the other three LHC experiments. The experiments requirements and the expectations for both the sites and the services themselves are exposed in detail. Finally, the operations procedures, which have been elaborated together with the experiment support teams will be included in this presentation

**70**

## HepData - the HEP data archive reloaded

**Authors:** Andy Buckley[1]; Mike Whalley[2]

[1] *University of Edinburgh*

[2] *Durham University*

**Corresponding Author:** a.g.buckley@gmail.com

The HEPDATA repository is a venerable collection of major HEP results from more than 35 years of particle physics activity. Historically accessed by teletype and remote terminal, the primary interaction mode has for many years been via the HEPDATA website, hosted in Durham, UK. The viability of this system has been limited by a set of legacy software choices, in particular a hierarchical database with no type distinction and no bindings to modern development frameworks.

As part of a large coherent programme of work for Monte Carlo generator validation and tuning, the database and front-ends have been rebuilt using a modern database engine, an automated object-relational Java API, and a high-level approach to Java servlet applications. The data migration process has increased the semantic content of many records and provided a new format for data input, and the API has allowed a much more powerful set of Web interfaces to be developed, including flexible output formatting. This last feature means that HepData can now be used for automatic single-sourcing and synchronisation of data records used for MC validation, and opens the door to a range of wider applications.

**Summary**:

?

**Thursday, 25 February - Computing Technology for Physics Research / 71**

## The ALICE Online Data Quality Monitoring

**Authors:** Adriana Telesca[1]; Barthelemy von Haller[1]; Sylvain Chapeland[1]

**Co-authors:** Csaba Soós [1]; David Rodriguez Navarro [1]; Filipo Costa [1]; Franco Carena [1]; Giuseppe Simonetti [1]; Irina Makhlyueva [1]; Marc Frauman [1]; Ornella Rademakers-di Rosa [1]; Pierre Vande Vyvre [1]; Roberto Divià [1]; Ulrich Fuchs [1]; Vasco Chibante Barroso [1]; Wisla Carena [1]

[1] *CERN*

**Corresponding Author:** barthelemy.von.haller@cern.ch

ALICE (A Large Ion Collider Experiment) is the detector designed to study the physics of strongly interacting matter and the quark-gluon plasma in Heavy-Ion collisions at the CERN Large Hadron Collider (LHC).

The online Data Quality Monitoring (DQM) is a critical element of the data acquisition's software chain. It intends to provide shifters with precise and complete information to quickly identify and overcome problems, and as a consequence to ensure acquisition of high quality data. DQM typically involves the online gathering, the analysis by user-defined algorithms and the visualization of monitored data.

This paper describes the final design of ALICE's DQM framework called AMORE (Automatic MOnitoRing Environment), as well as its latest and coming features like the integration with the offline analysis and reconstruction framework, a better use of multi-core processors by a parallelization effort, and its interface with the eLogBook.
The concurrent collection and analysis of data in an online environment requires the framework to be highly efficient, robust and scalable. We will describe what has been implemented to achieve these goals and the benchmarks we carried on to ensure appropriate performance.

We finally review the wide range of usages people make of this framework, from the basic monitoring of a single sub-detector to the most complex ones within the High Level Trigger farm or using the Prompt Reconstruction and we describe the various ways of accessing the monitoring results. We conclude with our experience, before and after the LHC restart, when monitoring the data quality in a real-world and challenging environment.

**Tuesday, 23 February - Methodology of Computations in Theoretical Physics / 75**

## Parallel versions of the symbolic manipulation system FORM

**Author:** Mikhail Tentyukov[1]

[1] *Karlsruhe University*

**Corresponding Author:** tentukov@particle.uni-karlsruhe.de

The symbolic manipulation program FORM is specialized to handle very large algebraic expressions. Some specific features of its internal structure make FORM very well suited for parallelization.

We have now parallel versions of FORM, one is based on POSIX threads and is optimal for modern multicore computers while another one uses MPI and can be used to parallelize FORM on clusters and Massive Parallel Processing systems. Most existing FORM programs will be able to take advantage of the parallel execution without the need for modifications.

**Friday, 26 February - Methodology of Computations in Theoretical Physics / 76**

## Feynman Integral Evaluation by a Sector decomposiTion Approach (FIESTA)

**Author:** Mikhail Tentyukov[1]

**Co-authors:** Alexander Smirnov [2]; Vladimir Smirnov [3]

[1] *Karlsruhe University*

[2] *Scientific Research Computing Center of Moscow State University*

[3] *Nuclear Physics Institute of Moscow State University*

**Corresponding Author:** tentukov@particle.uni-karlsruhe.de

Sector decomposition in its practical aspect is a constructive method used to evaluate Feynman integrals numerically. We present a new program performing the sector decomposition and integrating the expression afterwards. Also the program can be used in order to expand Feynman integrals automatically in limits of momenta and masses with the use of sector decompositions and Mellin–Barnes representations. The program is parallelizable on modern multicore computers and even to multiple computers.

Also we demonstrate some new numerical results for four-loop massless propagator master integrals.

**Friday, 26 February - Data Analysis - Algorithms and Tools / 78**

## FATRAS –A Novel Fast Track Simulation Engine for the ATLAS Experiment

**Authors:** Sebastian Fleischmann[1]; Stephen Haywood[2]

[1] *U. Bonn*

[2] *RAL*

**Corresponding Author:** s.haywood@rl.ac.uk

Monte Carlo simulation of the detector response is an inevitable part of any kind of analysis which is performed with data from the LHC experiments. These simulated data sets are needed with large statistics and high precision level, which makes their production a CPU-cost intensive task. ATLAS has thus concentrated on optimizing both full and fast detector simulation techniques to achieve this goal within the computing limits of the collaboration. At the early stages of data-taking, in particular, it is necessary to reprocess the Monte Carlo event samples continuously, while integrating adaptations to the simulation modules to improve the agreement with the data taken from the detector itself.

We present a new, fast track simulation engine which establishes a full Monte Carlo simulation which is based on modules and the geometry of the ATLAS standard track reconstruction application. This is combined with a fast parametric-response simulation of the Calorimeter. This approach shows a high level of agreement with full simulation, while achieving a relative timing gain of about 100. FATRAS was designed to provide a fast feedback cycle for tuning the MC simulation with real data: this includes the material distribution inside the detector, the integration of misalignment and conditions status, as well as calibration at the hit level. We present the concepts of the fast track simulation, although will concentrate mainly on the performance after integrating the feedback from first data taken with the ATLAS detector during the 2009-10 winter months.

**79**

## Alignment of the ATLAS Inner Detector

**Author:** Stephen Haywood[1]

[1] *RAL*

**Corresponding Author:** s.haywood@rl.ac.uk

CERN's Large Hadron Collider (LHC) is the world's largest particle accelerator. It will make two proton beams collide at an unprecedented centre-of-mass energy of 14 TeV. ATLAS is a general purpose detector which will record the products of the LHC proton-proton collisions. At the inner radii, the detector is equipped with a charged-particle tracking system built on two technologies: silicon and drift-tube based detectors, comprising the ATLAS Inner Detector (ID). The performance of this detector will be optimized in order to enable ATLAS to achieve its physics goals.

The alignment of this tracking system poses a real challenge as one needs to determine almost 36000 degrees of freedom. The precision one needs to attain for the alignment of the most sensitive coordinates of the silicon sensors is just few microns. This limit comes from the requirement that the misalignment should not worsen significantly the track parameter determination far beyond that due the intrinsic sensor resolution. Therefore the alignment of the ATLAS ID requires the application of complex algorithms which demand extensive CPU and memory usage, as large matrix inversion and many iterations of the algorithms are required. So far, the proposed alignment algorithms have been exercised already in several situations, such as: a Combined Test Beam, Cosmic Ray runs and a large-scale computing simulation of physics samples mimicking the ATLAS operation. For the later samples, the trigger of the experiment has been simulated and the event filters applied in order to produce an alignment stream. The full alignment chain has been tested on that stream and alignment constants have been produced and validated within 24 hours, providing a complete test of the first-pass reconstruction of physics events.

The ATLAS ID commissioning is being finalized currently and many thousands of cosmic-ray tracks have already been reconstructed with the final operating system. These cosmic ray data served to produce an early alignment of the real ATLAS Inner Detector even before the LHC start-up.

**Friday, 26 February - Data Analysis - Algorithms and Tools / 80**

## Parallel approach to online event reconstruction in the CBM experiment

**Author:** Ivan Kisel[1]

[1] *Gesellschaft fuer Schwerionen forschung mbH (GSI)-Unknown-Unknow*

Future many-core CPU and GPU architectures require relevant changes in the traditional approach to data analysis. Massive hardware parallelism at the levels of cores, threads and vectors has to be adequately reflected in mathematical, numerical and programming optimization of the algorithms used for event reconstruction and analysis.
An investigation of the Kalman filter, which is the core of the reconstruction algorithms in modern HEP experiments, has demonstrated a potential several orders of magnitude increase of the speed of the algorithms, if properly optimized and parallelized. The Kalman filter based track fit is used as a benchmark for monitoring the performance of novel CPU and GPU architectures, as well as for investigating modern parallel programming languages.
In the CBM experiment at FAIR/GSI all basic reconstruction algorithms have been parallelized. For maximum performance all algorithms use variables in single precision only. In addition, a significant speed-up is provided by localizing data in a high-speed cache memory. Portability of the parallel reconstruction algorithms with respect to different CPU and GPU architectures is supported by the special headers and vector classes, which have been developed for using SIMD instruction sets. The reconstruction quality is monitored at each stage in order to keep it at the same level as for the initial scalar versions of the algorithms.
Different reconstruction methods, implemented in CBM, show different degrees of intrinsic parallelism, thus the speed-up varies up to few orders of magnitude. The speed-up factors for each stage of the algorithms parallelization are presented and discussed.

**81**

## RAVEN - a Random Access, Visualisation and Exploration Network for the Analysis of Petabyte Sized Data Sets

**Authors:** Hans Kristian Gudmundson[1]; Helmut Neukirchen[2]; Markward Britsch[3]; Michael Schmelling[4]; Nicola Whitehead[1]; Nikolai Gagunashvili[1]

[1] *University of Akureyri, Iceland*

[2] *University of Iceland, Reykjavik, Iceland*

[3] *MPI for Nuclear Physics,Heidelberg, Germany*

[4] *MPI for Nuclear Physics, Heidelberg, Germany*

**Corresponding Author:** michael.schmelling@mpi-hd.mpg.de

The analysis and visualisation of the LHC data is a good example of human interaction with petabytes of inhomogenous data. A proposal is presented, addressing both physics analysis and information technology, to develop a novel distributed analysis infrastructure which is scalable to allow real time random access to and interaction with peta-bytes of data. The proposed hardware basis is a network of intelligent "CSR"-units, which combine Computing, data Storage and Routing functionalities. At the software level the project would develop effcient protocols for broadcasting information, data distribution and information collection upon such a network, together with a middleware layer for data processing, client applications for data visualisation and an interface for the management of the system.

**Wednesday, 24 February - Plenary Session / 83**

## Data access in the High Energy Physics community

**Author:** Fabrizio Furano[1]

[1] *Conseil Europeen Recherche Nucl. (CERN)*

**Corresponding Author:** fabrizio.furano@cern.ch

In this talk we pragmatically address some general aspects about massive data access in the HEP environment, starting to focus on the relationships that lie among the characteristics of the available technologies and the data access strategies which are consequently possible. Moreover, the upcoming evolutions in the computing performance available also at the personal level will likely pose new challenges for the systems that have to feed the computations with data. The talk will introduce then some ideas that will likely constitute the next steps in the evolution of this kind of worldwide distributed systems, towards new levels of performance, interoperability and robustness. Efficiently running data-intensive applications can be very challenging in a single site, depending on the scale of the computations; running them in a worldwide distributed environment with chaotic user-related random access patterns needs a design which avoids all the pitfalls which could harm its efficiency at a major degree.

**86**

## Studies of the performances of an open source batch system / scheduler (TORQUE / MAUI) implemented on a middle sized GRID site.

**Authors:** Claudio Tanci[1]; Francesco Cantini[2]; Leonello Servoli[3]; Mirko Mariotti[4]

[1] *Universita' degli Studi di Perugia*

[2] *INFN - Sezione di PErugia*

[3] *INFN - Sezione di Perugia*

[4] *Universita' degli Studi di Perugia*

**Corresponding Author:** leonello.servoli@pg.infn.it

Open source computing clusters for scientific purposes are growing in size, complexity and eterogeneity; often they are also included in some geographically distributed computing GRID. In this framework the difficulty of assessing the overall efficiency, identigying the bottlenecks and tracking the failures of single components is increasing continously.
In previous works we have formalized and proposed a set of metrics to make such an evaluation and built an analysis and simulation infrastructure to find a quantitative measure of the efficiency of a TORQUE/Maui based system. The analysis presented in this work is based on the comparison of real data and simulated data, the latter assumed as maximum theoretical limit.

**Friday, 26 February - Methodology of Computations in Theoretical Physics / 87**

## Sector decomposition via computational geometry

**Author:** Toshiaki KANEKO[1]

**Co-author:** Takahiro UEDA [2]

[1] *KEK, Computing Research Center*

[2] *University of Tsukuba*

**Corresponding Author:** toshiaki.kaneko@kek.jp

One of the powerful tools for evaluating multi-loop/leg integrals is sector decomposition, which can isolate infrared divergences from parametric representations of the integrals. The aim of this talk is to present a new method to replace iterated sector decomposition, in which the problems are converted into a set of problems in convex geometry, and then they can be solved by using algorithms in computational geometry. This method never falls into an infinite loop, and some examples show that it gives the relatively small number of generated sectors.

**88**

## Using TurboSim for Fast Detector Simulation

**Author:** Anil P. Singh[1]

**Co-authors:** Harrison B. Prosper [2]; Lovedeep K. Saini [3]; Suman B. Beri [3]

[1] *Panjab Univerisity*

[2] *Florida State Univerisity*

[3] *Panjab University*

The new physics searches like SUSY in the CMS detector at the LHC will require a very fine scanning of the parameter space over a a large number of the points. Accordingly we need to address the problem of developing a very fast setup to generate and simulate large MC samples. We have

explored the use TurboSim as a fast and the standalone setup for generating such samples. TurboSim does not intend to replace full simulation rather it uses the full simulation to create a very large lookup table for mapping the generator level particles into the particle level objects. We have used centrally generated winter-09 datasets to populate the lookup database. Here we present the details of validation exercise performed by comparing TurboSimulated events with the full detector simulation using (Z->ee) candle analysis. We also compare the CPU requirements for turbosim and cms fastsim and examine the potential for extensive usage of turbosim for fast and largescale MC production for SUSY searches.

**Friday, 26 February - Methodology of Computations in Theoretical Physics / 90**

## Multiple Polylogarithms and Loop Integrals

**Author:** Yoshimasa Kurihara[1]

[1] *KEK*

**Corresponding Author:** yoshimasa.kurihara@kek.jp

An importance of the multiple-polylog function (MLP) for the calculation of loop integrals was pointed out by many authors.
We give some general discussions between MLP and multi-loop integrals from view point of computer algebra.

**93**

## Automating CMS Calibrations using the WMAgent framework

**Author:** Hassen Riahi[1]

[1] *University and INFN Perugia*

Particle beams are now circulating in the world's most powerful particle accelerator LHC at CERN and the experiments are ready to record data from beam. Data from first collisions will be crucial for sub-detector commissioning, making alignment and calibration high priority activities. Executing the alignment and calibration workflow represents a complex and time consuming task, with intricate data dependencies and stringent latency requirements. For this reason CMS Workload Management has defined an architecture and strategy to automate the workflows to minimise the required human effort. In this paper we discuss the WMCore and WMagent components used in the actual CRABSERVER framework showing the prototype tuned over the first use case, the Beam Spot. We also present the longer term strategies for moving alignment and calibration workflows into the standard T0 processing and automating the analysis workflow.

**Friday, 26 February - Computing Technology for Physics Research / 95**

## ”NoSQL” databases in CMS Data and Workflow Management

**Author:** Simon Metson[1]

**Co-authors:** Andrew Melo [2]; Dave Evans [3]; Steve Foulkes [3]; Valentin Kuznetsov [4]

[1] *H.H. Wills Physics Laboratory*

[2] *Vanderbilt*

[3] *FNAL*

[4] *Cornell*

**Corresponding Author:** simon.metson@cern.ch

In recent years a new type of database has emerged in the computing landscape. These "NoSQL" databases tend to originate from large internet companies that have to serve simple data structures to millions of customers daily. The databases specialise for certain use cases or data structures and run on commodity hardware, as opposed to large traditional database clusters.

In this paper we discuss the current usage of "NoSQL" databases in the CMS data and workload management tools, discuss how we expect our systems to evolve to take advantage of them and how these technologies could be used in a wider context.

**96**

## Parallelization of likelihoood function data analysis software based on RooFit package

**Author:** Alfio Lazzaro[1]

[1] *Universita degli Studi di Milano & INFN, Milano*

With the startup of the LHC experiments, the community will be focused on the data analysis of the collected data. The complexity of the data analyses will be a key factor to find eventual new phenomena. For such a reason many data analysis tools are being developed in the last years. allowing the use of different techniques, such as likelihood-based procedures, neural networks, boost decision trees. In particular the likelihood-based procedures allow the estimation of unknown parameters based on a given input sample. Complex likelihood functions, with several free parameters, many independent variables and large data sample, can be very CPU-time consuming for their calculation. Furthermore for a good estimation it is required the generation of several simulated samples of events from the probability density functions, so the whole procedure which can be CPU-time consuming. In this presentation I will show how the likelihood calculation, the normalization integrals calculation, and the events generation can be parallelized using MPI techniques to scale over multiple nodes or multi-threads for multi-cores in a single node. We will present the speed-up improvements obtained in typical physics applications such as complex maximum likelihood fits using the RooFit and RooStats packages. We will also show results of hybrid parallelization between MPI and multi-threads, to take full advantage of multi-core architectures.

**Thursday, 25 February - Methodology of Computations in Theoretical Physics / 97**

## The FeynSystem: FeynArts, FormCalc, LoopTools

**Author:** Thomas Hahn[1]

[1] *MPI Munich*

**Corresponding Author:** hahn@mppmu.mpg.de

The talk describes the recent additions in the automated Feynman diagram computation system FeynArts, FormCalc, and LoopTools

**Thursday, 25 February - Computing Technology for Physics Research / 100**

## Contextualization in Practice: The Clemson Experience

**Author:** Michael Fenn[1]

**Co-authors:** Jerome LAURET [2]; Sebastien Goasguen [1]

[1] *Clemson University*

[2] *BROOKHAVEN NATIONAL LABORATORY*

**Corresponding Author:** mfenn@clemson.edu

Dynamic virtual organization clusters with user-supplied virtual machines (VMs) have advantages over generic environments. These advantages include the ability for the user to have a priori knowledge of the scientific tools and libraries available to programs executing in the virtualized environment well as the other details of the environment. The user can also perform small-scale testing locally, thus saving time and conserving computational resources.
However, user-supplied VMs require contextualization in order to operate properly in a given cluster environment. Two types of contextualization are necessary per-environment and per-session. Examples of per-environment contextualization include one-time configuration tasks such as ensuring availability of ephemeral storage, mounting of a cluster-provided shared filesystem, integration with the cluster's batch scheduler, etc. Also necessary is per-session contextualization such as the assignment of MAC and IP addresses.
This paper discusses the challenges and techniques used to overcome those challenges in the contextualization of the STAR VM for the Clemson University cluster environment. Also included are suggestions to VM authors to allow for efficient contextualization of their VMs.

**101**

## Monitoring the software quality in FairRoot

**Authors:** Florian Uhlig[1]; Mohammad Al-Turany[1]

[1] *GSI Darmstadt*

**Corresponding Author:** mohammad.al-turany@cern.ch

Up-to-date informations about a software project helps to find problems as early as possible. This includes for example information if a software project can be build on all supported platforms without errors or if specified tests can be executed and deliver the correct results.

We will present the scheme which is used within the FairRoot framework to continuously monitor the status of the project. The tools used for these tasks are based on the open source tools CMake and CDash.
CMake is used to generate standard build files for the different operating systems/compiler out of simple configuration files and to steer the build and test processes. The generated information is send
to a central CDash server. From the generated web pages information about the status of the project at any given time can be obtained.

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 102**

## Parallelization of the SIMD Ensemble Kalman Filter for Track Fitting Using Ct

**Authors:** Anwar Ghuloum[1]; Michael D. McCool[2]; R. Gabriel Esteves[2]

[1] *Intel*

[2] *Intel/University of Waterloo*

**Corresponding Author:** rodolfo.g.esteves@intel.com

A great portion of data mining in a high-energy detector experiment is spent in the complementary tasks of track finding and track fitting. These problems correspond, respectively, to associating a set of measurements to a single particle, and to determining the parameters of the track given a candidate path [Avery 1992]. These parameters usually correspond to the 5-tuple state of the model of motion of a charged particle in a magnetic field.

Global algorithms for track fitting have been superceded by recursive least-square estimation algorithms that, assuming the measurement noise is Gaussian, result in an optimal estimate [Fr\"uwirth and Widl 1992]. However, this assumption hardly ever holds due to energy loss and multiple scattering. Extensions to the Kalman filter have been proposed and implemented [Fr\"uwirth 1997] that use sums of Gaussians to model non-Gaussian distributions. In addition, non-linear filtering is necessary to eliminate the effects of outliers including misclassified measurements. Track fitting based on sums of Gaussians can be implemented through an ensemble of single Gaussian fits, each of which is the result of a single Kalman filter [Gorbunov et al. 2008].

Efficient parallel implementations of these algorithms are also crucial due to the large amount of data that needs to be processed. Since many separate tracks need to be fitted, the simplest way to parallelize the problem is to fit many independent tracks at once. However, with the ensemble algorithm is it also possible to parallelize across the ensemble within a single track fit. Parallelism mechanisms available in the hardware include multiple nodes and multiple cores.

Prior implementations can also make use of the SIMD vector units present in most modern CPUs [Gorbunov et al. 2008]. We have performed an implementation of the ensemble approach using Ct. Ct is a generalized data-parallel programming platform that can efficiently target both multiple cores and SIMD vector units from a single high-level specification in C++. Compared to the earlier SIMD implementation, Ct allows for hardware and instruction set portability.

We have obtained scalable speedup with respect to a scalar baseline implemented using both single and double precision. In the prior work, a speedup of 1.6x for scalar vs. vectorized versions of the algorithm were reported for double precision. Our results are comparable to these.

In addition, we have explored the implementation of sequential Monte Carlo in Ct using particle filtering to model arbitrary probabality distributions. However, one challenge in this context is the computation of the likelihood function, which requires estimates of the measurement error for every measurement and their correlation.

References
Avery, P. 1992. Applied fitting theory V: Track fitting using the Kalman filter. Tech. rep.
Fr\"uuwirth, R., and Widl, E. 1992. Track-based alignment using a Kalman filter technique. Communications in Computational Physics .
Fr\"uwirth, R. 1997. Track-fitting with non-Gaussian noise. Communications in Computational Physics (Jan.).
Gorbunov, S., Kebschull, U., Kisel, I., Lindenstruth, V., and M\"uller, W. F. J. 2008. Fast SIMDized Kalman filter-based track fit. Computer Physics Communications .

**Thursday, 25 February - Data Analysis - Algorithms and Tools / 105**

# The RooStats project

**Authors:** Gregory Schott[1]; Kyle Cranmer[2]; Lorenzo Moneta[3]; Wouter Verkerke[4]

[1] *Karlsruhe Institute of Technology*

[2] *New York State University*

[3] *CERN*

[4] *Nikhef*

**Corresponding Author:** lorenzo.moneta@cern.ch

RooStats is a project to create advanced statistical tools required for the analysis of LHC data, with emphasis on discoveries, confidence intervals, and combined measurements.
The idea is to provide the major statistical techniques as a set of C++ classes with coherent interfaces, which can be used on arbitrary model and datasets in a common way. The classes are built on top of RooFit, which provides a very convenient functionality for modeling the probability density functions or the likelihood functions, required as inputs for any statistical technique. Furthermore, RooFit provides via the RooWorkspace class, the functionality for easily creating models, for analysis combination and for digital publication of the likelihood function and the data.
We will present in detail the design and the implementation of the different statistical methods of RooStats. These include various classes for interval estimation and for hypothesis test depending on different statistical techniques such as those based on the likelihood function, or on frequentists or bayesian statistics. These methods can be applied in complex problems, including cases with multi parameter of interests and various nuisance parameters. We will also show some example of usage and we will describe the results and the statistical plots obtained by running the RooStats methods.

**Friday, 26 February - Methodology of Computations in Theoretical Physics / 107**

## Two-Loop Fermionic Integrals in Perturbation Theory on a Lattice

**Author:** Roman Rogalyov[1]

[1] *IHEP*

**Corresponding Author:** rogalyov@th1.ihep.su

A comprehensive number of one-loop integrals in a theory with Wilson fermions at $r = 1$ is computed using the Burgio–Caracciolo–Pelissetto algorithm. With the use of these results, the fermionic propagator in the coordinate representation is evaluated, making it possible to extend the Luscher-Weisz procedure for two-loop integrals to the fermionic case. Computations are performed with FORM and REDUCE packages.

**108**

## An improvement in LVCT cache replacement policy for data grid

**Author:** Arti Kashyap[1]

**Co-authors:** A. Rathore [2]; J. P. Achara [2]; V.K. Gupta [2]

[1] *LNM Institute of IT*

[2] *LNMIIT*

**Corresponding Author:** arti.kashyap@lnmiit.com

Abstract: Caching in data grid has great benefits because of faster and nearer data availability of data objects. Caching decreases retrieval time of data objects. One of the challenging tasks in designing a cache is designing its replacement policy. The replacement policy decides which set of files are to be evicted to accommodate the newly arrived file in the cache and also whether a newly arrived file should be cached or not. The traditional replacements policies, based on LRU and LFU algorithms are not suitable for the data grid because of large-and-varying size and retrieving cost of data objects in data grid [1]. Other class of replacement policies used in data grid is based on calculating utility function for each file which also takes care of size and retrieving cost of the file along with Locality

Strength [2, 3, 4]. We can easily deduce that lesser the value of locality strength for a file, it is better to evict that file because of its low probability of re-accessing while other two parameters remaining the same. In LVCT (Least Value Based on Caching Time) policy [1], locality strength in the utility function is estimated using a term "Caching Time (CT)"where CT for a file F is defined to be sum of size of all files accessed after last reference to file F. While this policy showed better performance compared to previous policies [2, 3, 4, 5], it did not take care of the cases where CT computed in the policy is almost same but the number of files accessed after last reference to file F differs largely. For example, two cases where (i) number of files is large but sizes of these files are small and (ii) number of files is small but sizes of these files are big after last reference to a file; can result into nearly same value of CT and hence same value of locality strength. However, locality strength of first case should be low compared to second case because of large number of files accessed after its last reference having the same value of CT for both. Here, we propose that number of files accessed after last reference to file F should also be considered along with CT in estimation of locality strength so that a better decision can be made for the eviction of a file.

References:
[1] Song Jiang and Xiaodong Zhang, "Efficient distributed disk caching in data grid management", Proceedings of IEEE International Conference on Cluster Computing, 2003.
[2] R. Wooster and M. Abrams, "Proxy caching that estimates page load delays", Proceedings of 6th international World Wide Web Conference, April 1997
[3] P. Lorensetti, L. Rizzo and L. Vicisano, "Replacement Policies for a Proxy Cache", http://www.iet.unipi.it/luigi/research.html
[4] P. Cao and S. Irani, "Cost-aware WWW proxy caching algorithms", Proceedings of USENIX Symposium on Internet Technologies and Systems, 1997
[5] Ekow Otoo, Frank Olken and Arie Shoshani, "Disk Cache Replacement Algorithm for Storage Resource Managers in Data Grids", Proceeding of IEEE Conference on Super-Computing, 2002

**Tuesday, 23 February - Methodology of Computations in Theoretical Physics / 110**

## Deterministic numerical box and vertex integrations for one-loop hexagon reductions

**Author:** Elise de Doncker[1]

**Co-authors:** Fukuko Yuasa [2]; Junpei Fujimoto [2]; Nobuyuki Hamaguchi [3]; Tadashi Ishikawa [2]; Yoshimasa Kurihara [2]; Yoshimitsu Shimizu [2]

[1] *Western Michigan University*

[2] *High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki, Japan*

[3] *Hitachi, Ltd., Software Division, Totsuka-ku, Yokohama, Japan*

**Corresponding Author:** elise.dedoncker@wmich.edu

We provide a fully numerical, deterministic integration at the level of the three- and four-point functions, in the reduction of the one-loop hexagon integral by sector decomposition. For the corresponding two- and three-dimensional integrals we use an adaptive numerical approach applied recursively in two and three dimensions, respectively.
The adaptive integration is coupled with an extrapolation for an accurate, automatic treatment of integrand singularities arising from vanishing denominators in the interior of the integration domain. Furthermore, the recursive procedure alleviates extensive memory use as incurred with standard adaptive, multidimensional integration software.
Tensor integrals are handled automatically by this technique and the separation of infrared singularities follows naturally by dimensional regularization.

**Thursday, 25 February - Plenary Session / 112**

## Data Transfer Optimization - Going Beyond Heuristics

**Author:** Roman Bartak[1]

[1] *Charles University in Prague*

**Corresponding Author:** bartak@ktiml.mff.cuni.cz

Scheduling data transfers is frequently realized using heuristic approaches. This is justifiable for on-line systems when extremely fast response is required, however, when sending large amount of data such as transferring large files or streaming video, it is worthwhile to do real optimization. This paper describes formal models for various networking problems with the focus on data networks. In particular we describe how to model path placement problems where the task is to select a path for each demand starting at some source node and finishing at the destination node. The demands can be instantaneous, for example in data streaming, or spread in time in so called bandwidth on demand problems. This second problem is a complex combination of path placement and cumulative scheduling problems. As constraint programming is a successful technology for solving scheduling problems, we sketch the basic principles of constraint programming and illustrate how constraint programming can be applied to solve the above mentioned problems. The main advantage of this solving approach is extendibility where the base model can be augmented with additional constraints derived from the specific problem requirements.

**Friday, 26 February - Plenary Session / 113**

## Scientific Computing with Amazon Web Services

**Author:** Singh Deepak[1]

[1] *Business Development Manager - Amazon EC2*

In an era where high throughput instruments and sensors are increasingly providing us faster access to new kinds of data, it is becoming very important to have timely access to resources which allow scientists to collaborate and share data while maintaining the ability to process vas > quantities of data or run large scale simulations when required. Built on Amazon's vast global computing infrastructure, Amazon Web Services (AWS) provides scientists with a number of highly scalable, highly available infrastructure services that can be used to perform a variety of tasks. The ability to scale storage and analytics resources on-demand has made AWS a platform for a number of scientific challenges including high energy physics, next generation sequencing, and galaxy mapping. A number of scientists are also making a number of algorithms and applications available as Amazon Machine Images, or as applications that can be deployed to Amazon Elastic MapReduce. In this talk, we will discuss the suite of Amazon Web Services relevant to the scientific community, go over some example use cases, and the advantages that cloud computing offers for the scientific community. We will also discuss how we can leverage new paradigms and trends in distributed computing infrastructure and utility models that allow us to manage and analyze big data at scale.

**Friday, 26 February - Data Management Panel / 114**

## Data Management Panel

**Authors:** Alberto Pace[1]; Andrew Hanushevsky[2]; Beob Kyun Kim[3]; Rene Brun[1]; Tony Cass[1]

[1] *CERN*

[2] *Unknown*

[3] *KISTI*

**Thursday, 25 February - Multicore Panel / 115**

## Multicore Panel

**Authors:** Alfio Lazzaro[1]; Anwar Ghuloum[2]; Mohammad Al-Turany[3]; Sverre Jarp[4]

[1] *Universita degli Studi di Milano & INFN, Milano*

[2] *Intel Corporation*

[3] *GSI DARMSTADT*

[4] *CERN*

The multicore panel will review recent activities in the multicore/manycore arena.
It will consist of four people kicking off the session by making short presentations, but it will mainly rely on a good interaction with the audience:
Mohammad Al-Turany (GSI/IT)
Anwar Ghuloum (INTEL Labs)
Sverre Jarp (CERN/IT)
Alfio Lazzaro (CERN/IT)

**Thursday, 25 February - Plenary Session / 117**

## Tools for Dark Matter in Particle Physics and Astrophysics

**Author:** Alexander Pukhov[1]

[1] *Moscow State University, Russia*

**118**

## How computing centres of Alice connect? A social network analysis of cooperative ties

**Authors:** Eric Widmer[None]; Federico Carminati[1]; Giuliana Galli Carminati[None]

[1] *CERN*

The ALICE Core Computing Project conducted in the fall 2009 an inquiry on the social connections between the Computing Centres of the ALICE Distributed Computing infrastructure. This inquiry was based on social network analysis, a scientific method dedicated to the understanding of complex relational structures linking human beings. The paper provides innovative insights on various relational dimensions of the cooperative work within the ALICE project, such as support and advices, interactions by emails or work hindrances existing among centers. The analysis focuses on core-periphery and subgroups issues and reveal some suboptimally organized cluster of ties. Propositions for improvement will be discussed.

**Wednesday, 24 February - Plenary Session / 119**

## Statistics challenges in HEP

**Author:** Eilam Gross[1]

[1] *Weissman Institute of Physical Sciences*

The LHC was built as a discovery machine, whether for a Higgs Boson or Supersymmetry.
In this review talk we will concentrate on the methods used in the HEP community to test hypotheses. We will cover via a comparative study, from the LEP hybrid "CLs" method and the Bayesian TEVATRON exclusion techniques to the LHC frequentist discovery techniques. We will explain how to read all the exclusion and prospective discovery plots with their yellow and green bands and how to include systematics in a significance calculation. The review is aimed to be pedagogical.

**Afternoon session / 121**

## Official Openning

**Afternoon session / 122**

## High Tea

**Friday, 26 February - Plenary Session / 128**

## Application of Many-core Accelerators for Problems in Astronomy and Physics

**Author:** Naohito Nakasato[1]

[1] *University of Aizu*

Recently, many-core accelerators are developing so fast that the computing devices attract researchers who are always demanding faster computers. Since many-core accelerators such as graphic processing unit (GPU) are nothing but parallel computers, we need to modify an existing application program with specific optimizations (mostly parallelization) for a given accelerator.
In this paper, we describe our problem-specific compiler system for many-core accelerators, specifically, GPU and GRAPE-DR. GRAPE-DR is another many-core accelerators device that is specially targeted scientific applications.
In our compiler, we focus a compute intensive problem expressed as two-nested loop.Recently, many-core accelerators are developing so fast that the computing devices attract researchers who are always demanding faster computers.
Since many-core accelerators such as graphic processing unit (GPU) are nothing but parallel computers, we need to modify an existing application program with specific optimizations (mostly parallelization) for a given accelerator.
In this paper, we describe our problem-specific compiler system for many-core accelerators, specifically, GPU and GRAPE-DR. GRAPE-DR is another many-core accelerators device that is specially targeted scientific applications.
In our compiler, we focus a compute intensive problem expressed as two-nested loop. Our compiler ask a user to write computations in the inner-most loop. All details related to parallelization and optimization techniques for a given accelerator are hidden from the user point of view. Our compiler successfully generates the fastest code ever for astronomical N-body simulations with the performance of 2600 GFLOPS (single precision) on a recent GPU.
However, this code that simply uses a brute-force $O(N2)$ algorithm is not practically useful for a system with $N > 100,000$. For more lager system, we need a sophisticated $O(N \log N)$ force evaluation algorithm, e.g., the oct-tree method. We also report our implementation of the oct-tree method on GPU. We successfully run a simulation of structure formation in the universe very efficiently using the oct-tree method. Another successful application on both GPU and GRAPE-DR is the evaluation

of a multi-dimensional integral with quadruple precision. The program generated by our compiler runs at a speed of 5 - 7 GFLOPS on GPU and 3 - 5 on GRAPE-DR.
This computation speed is more than 50 times faster than a general purpose CPU.Recently, many-core accelerators are developing so fast that the computing devices attract researchers who are always demanding faster computers. Since many-core accelerators such as graphic processing unit (GPU) are nothing but parallel computers, we need to modify an existing application program with specific optimizations (mostly parallelization) for a given accelerator.
In this paper, we describe our problem-specific compiler system for many-core accelerators, specifically, GPU and GRAPE-DR. GRAPE-DR is another many-core accelerators device that is specially targeted scientific applications.
In our compiler, we focus a compute intensive problem expressed as two-nested loop. Our compiler ask a user to write computations in the inner-most loop.
All details related to parallelization and optimization techniques for a given accelerator are hidden from the user point of view. Our compiler successfully generates the fastest code ever for astronomical N-body simulations with the performance of 2600 GFLOPS (single precision) on a recent GPU.
However, this code that simply uses a brute-force $O(N2)$ algorithm is not practically useful for a system with $N > 100,000$. For more lager system, we need a sophisticated $O(N \log N)$ force evaluation algorithm, e.g., the oct-tree method.
We also report our implementation of the oct-tree method on GPU. We successfully run a simulation of structure formation in the universe very efficiently using the oct-tree method. Another successful application on both GPU and GRAPE-DR is the evaluation of a multi-dimensional integral with quadruple precision. The program generated by our compiler runs at a speed of 5 - 7 GFLOPS on GPU and 3 - 5 on GRAPE-DR. This computation speed is more than 50 times faster than a general purpose CPU.
Our compiler ask a user to write computations in the inner-most loop.
All details related to parallelization and optimization techniques for a given accelerator are hidden from the user point of view.
Our compiler successfully generates the fastest code ever for astronomical N-body simulations with the performance of 2600 GFLOPS (single precision) on a recent GPU.
However, this code that simply uses a brute-force $O(N2)$ algorithm is not practically useful for a system with $N > 100,000$. For more lager system, we need a sophisticated $O(N \log N)$ force evaluation algorithm, e.g., the oct-tree method. We also report our implementation of the oct-tree method on GPU. We successfully run a simulation of structure formation in the universe very efficiently using the oct-tree method. Another successful application on both GPU and GRAPE-DR is the evaluation of a multi-dimensional integral with quadruple precision.
The program generated by our compiler runs at a speed of 5 - 7 GFLOPS on GPU and 3 - 5 on GRAPE-DR.
This computation speed is more than 50 times faster than a general purpose CPU.

**Friday, 26 February - Plenary Session / 129**

## Numerical approach to Feynman diagram calculations: Benefits from new computational capabilities

**Author:** Fukuko YUASA[1]

[1] *KEK*

**Wednesday, 24 February - Plenary Session / 130**

## Automation of multi-leg one-loop virtual amplitudes

**Author:** Daniel Maitre[1]

[1] *IPPP, Great Britain*

In the last years, much progress has been made in the computation of one-loop virtual matrix elements for processes involving many external particles. I this talk I will show the importance of NLO-accuracy computations for phenomenologically relevant processes and review the recent progress that will make their automated computation tractable and their inclusion in Monte Carlo tools possible.

**Thursday, 25 February - Plenary Session / 131**

## Lattice QCD simulations

**Author:** Karl Jansen[1]

[1] *NIC, DESY, Zeuthen*

The formulation of QCD on a 4-dimensional space-time euclidean lattice is given. We describe, how with particular implementations of the lattice Dirac operator the lattice artefacts can be changed from a linear to a quadratic behaviour in the lattice spacing allowing therefore to reach the continuum limit faster. We give an account of the algorithmic aspects of the simulations, discuss the supercomputers used and give the computational costs. A few examples of physical quantities which are computed today at almost physical quark masses are presented.

**Student Session / 132**

## Simulation and Visualisation Techniques

**Author:** Matevz Tadel[1]

[1] *CERN*

**Student Session / 133**

## Statistical Methods, Multivariate Analysis and Pattern Recognition

**Author:** Liliana Teodorescu[1]

[1] *Brunel University*

This lecture will present key statistical concepts and methods for multivariate data analysis and their applications in high energy physics.
It will discuss the meaning of multivariate statistical analysis and its benefits, will present methods for data preparation for applying multivariate analysis techniques, the generic problems addresses by these techniques and a few classes of such techniques. The applicability of these methods to pattern recognition problems in high energy physics will be demonstrated with examples from specific physics analyses.

**Student Session / 134**

## Multicore Computing

**Author:** Alfio Lazzaro[1]

[1] *Universita degli Studi di Milano & INFN, Milano*

**Student Session / 135**

## Software Development in High Energy Physics: a Critical Look

**Author:** Federico Carminati[1]

[1] *CERN*

**Student Session / 136**

## Internet Law: What Students, Professors, and Software Developers Need to Know

**Author:** Lawrence Pinsky[1]

[1] *UNIVERSITY OF HOUSTON*

**Afternoon session / 137**

## History of the ROOT System: Conception, Evolution and Experience by Rene BRUN

**Author:** Rene Brun[1]

[1] *CERN*

The ROOT system is now widely used in HEP, Nuclear Physics and many other fields. It is becoming a mature system and the software backbone for most experiments ranging from data acquisition systems, controls, simulation, reconstruction and of course data analysis. The talk will review the history of its conception at a time when HEP was moving from the Fortran era to C++. While the original target was a PAW-like system for data analysis, it became rapidly obvious that a more ambitious system had to be developed as a working alternative to the defunct object-oriented data base systems. Thanks to the collaboration of many individuals, ROOT has been gradually extended to include high quality math libraries, statistical analysis tools, visualization tools for statistics objects, detectors and event displays. The current ideas on the evolution of the system will also be presented.

**Afternoon session / 138**

## Computing Outside the Box: On Demand Computing & its Impact on Scientific Discovery by Ian FOSTER

**Author:** Ian Foster[1]

[1] *Unknown*

**Tuesday, 23 February - Poster Session / 139**

## Poster session

Poster list can be found here:
Poster list

**Ian Foster - Public Lecture / 140**

## Ian FOSTER - Public Lecture

**Author:** Ian Foster[1]

[1] *Unknown*

**ACAT 2010 Summary / 141**

## Computing Technology for Physics Research Summary

**Author:** Axel Naumann[1]

[1] *CERN*

**ACAT 2010 Summary / 142**

## Data Analysis - Algorithms and Tools Summary

**Author:** Liliana Teodorescu[1]

[1] *Brunel University*

**ACAT 2010 Summary / 143**

## Methodology of Computations in Theoretical Physics Summary

**Co-author:** Peter Uwer [1]

[1] *Humboldt-Universität zu Berlin*

**ACAT 2010 Summary / 144**

## ACAT 2010 Summary

**Thursday, 25 February - Methodology of Computations in Theoretical Physics / 145**

## The automation of subtraction schemes for next-to-leading order calculations in QCD

**Author:** Rikkert Frederix[1]

[1] *University Zurich*

There has been made tremendous progress in the automation of one-loop (or virtual) contributions to next-to-leading order (NLO) calculations in QCD, using both the conventional Feynman diagram approach as well as unitarity-based techniques. To have rates and distributions for observables at particle colliders at NLO accuracy also the real emission and subtraction terms have to be included in the calculation. Recently, two variations of subtraction schemes have been automated by several groups. In this talk these two schemes will be reviewed and there implementations discussed.

**Thursday, 25 February - Methodology of Computations in Theoretical Physics / 146**

## New developments in event generator tuning techniques

**Authors:** Andy Buckley[1]; Heiko Lacker[2]; Hendrik Hoeth[3]; Holger Schulz[2]; James Monk[4]; Jan-Eike von Seggern[2]

[1] *University of Edinburgh, UK*

[2] *Humboldt-Universität zu Berlin*

[3] *IPPP, Durham, UK*

[4] *MCnet/Cedar*

Data analyses in hadron collider physics depend on background simulations performed by Monte Carlo (MC) event generators. However, calculational limitations and non-perturbative effects require approximate models with adjustable parameters. In fact, we need to simultaneously tune many phenomenological parameters in a high-dimensional parameter-space in order to make the MC generator predictions fit the data. It is desirable to achieve this goal without spending too much time or computing resources iterating parameter settings and comparing the same set of plots over and over again.

I will present extensions and improvements to the MC tuning system, Professor, which addresses the aforementioned problems by constructing a fast analytic model of a MC generator which can then be easily fitted to data. Using this procedure it is for the first time possible to get a robust estimate of the uncertainty of generator tunings. Furthermore, we can use these uncertainty estimates to study the effect of new (pseudo-) data on the quality of tunings and therefore decide if a measurement is worthwhile in the prospect of generator tuning. The potential of the Professor method outside the MC tuning area is presented as well.

**Friday, 26 February - Plenary Session / 147**

## Applying CUDA Computing Model To Event Reconstruction Software

**Author:** Mohammad AL-TURANY[1]

[1] *GSI DARMSTADT*