Summary of Track-2

Offline Software

Andrew Norman, Elizabeth Sexton-Kennedy, Hisaya Kurashige, Ivan Kisel and James Kowalkowski
Overview

Track-2 Sessions:
#1 Mon: Reconstruction
#2 Mon: Exp. Computing Models
#3 Tue: Frameworks
#4 Tue: Simulation
#5 Thu: Analysis
#6 Thu: Tools

Mon-Thu: Enjoyed Okinawa and OIST

95 Track-2 related Abstracts
44 Oral
51 Poster
Track reconstruction software for Run-1 was designed with a lot of redundancy and safety margin.

Run-1 performance convinced us that our system was understood and extremely well modelled by MC.

Re-investigation of track seeding - taking high purity seeds from strip - make optimal use of new innermost Pixel layer (IBL)

Greatly improved the seed purity managed to be more efficient in less time: ~25 % saving

Summary:
• We achieved a factor 4 speedup of the overall event reconstruction time at the same time improving physics performance on all ends mainly achieved by the Inner Detector track reconstruction ready for Run-2 data taking
We re-factored the pileup simulation to process each interaction sequentially
- Required substantial rewrite of digitization code, and the re-organization of internal event processing

- The content of each event is dropped from memory once processed:
  - Only 1 event in memory at any given time, so arbitrarily many pileup events can be included in the digitization
  - Next challenge in pileup simulation for CMS: Reduce the I/O burden from the pileup events to open up more resources for processing
The new seeding and the cluster charge cut reduce timing of PixelLess and TobTec by 2x

- Physics performances and timing in different conditions:
  - TTbar samples with
  - BX=25 ns, <PU>=25, 40, 70, 140
  - BX=50 ns, <PU>=25

- Iterative tracking time reduction @25 ns:
  - 2x at PU=25, 3x at PU=40, 4x at PU=70

Summary:
- Many developments have been included in CMS' tracking code for Run2.
- Timing is now under control.
- Should expect the same or better physics performances as in Run1.
- Work is not over and many other developments are on their way.
Effect of the VeloTT in the HLT1 chain

- 3 times faster reconstruction chain → HLT1 runnable in ~32ms with looser requirements than in run I
  - 500MeV $p_T$ tracks available already at HLT1 level!
- 4 times less ghost rate, i.e. 6%
- 87% single track efficiency for $p > 3$GeV, $p_T > 500$MeV

Summary: Several studies done to improve the performance of the reconstruction chain to face the new challenging condition of Run II.
Slice borders are defined by regions where the neighborhood density drops below some critical value. The algorithm expands from neighbor to neighbor to find all borders.

Summary:
- A series of pattern-recognition algorithms have been adapted and chained into a vertex-first approach that works well for short and long tracks or showers.
- Reconstruction forms basis of neural net for electron neutrino event classification.
Run full track building with combinatorial expansion of candidates
- ultimate physics performance, slower
- 85% (95%) of tracks found with ≥90% (60%) of the hits

Parallelization is implemented by distributing threads across 21 eta bins
- for nEtaBin multiple of nThreads, split eta bins in threads
- for nThreads multiple of nEtaBin, split seeds in bin across nThreads/nEtaBin threads

Large speedup achieved, both on Xeon and Xeon Phi
- up to ~5x on Xeon and >10x Xeon Phi
- speedup saturates above nThreads=42

Summary:
- Significant speedup achieved both on Xeon and Xeon Phi.
- Ideal scaling indicates a large margin for further improvements.
Approximate the Hopfield-Network more closely

- Allowing weights on vertices $\theta_i$ and edges $w_{ij}$
- Weighted evaluation rule picking the neighbor with highest weight gain

$$E_i = \theta_i + \max_{\text{neighbor } j} (w_{ij} + E_j) = \sum_{\text{best path to } i} w_{ij} + \sum_{\text{best path to } i} \theta_i$$

Inherited properties

- High evaluation speed by single pass over the graph (forward pass, $O(n)$ - algorithm)
- Generation of valuable tracks following cells with highest state (backward pass)
- Iteration for additional disconnected tracks
- Emphasis on local connections - robustness against separated background and energy loss
- Agnostic to start position and direction of flight - seeding phase unnecessary

Upgraded properties

- Refined track model by encoding probability of positions and propagations into the weights
Reconstructed tracks clearly represent groups, which correspond to the original events. 83% of single events, no splitted events, further analysis with TOF information at the vertexing stage.
Achievements:

- Significant speedup of reconstruction to cope with 1kHz EF rate: \( \sim 4x \) !!!
- In Run 1 Inner Detector dominated processing time, most work done:
  - more modern Matrix library
    Eigen replaced CLHEP: faster due to SIMD intrinsics, ...
  - general code cleanups, Magnetic field rewritten in C++
  - 'free' lunch: newer gcc, Intel mathlib, SLC6, 64bits

Summary:

- Since end of Run 1, ATLAS updated offline software significantly to match new requirements
  - factor of \( \sim 4 \) speedup of data reconstruction
  - overhauled analysis model including EDM
  - multi-processing software in Production

Since the end of LHC Run 1, ATLAS improved the full analysis chain from RAW data to final ntuples, addressing a few issues seen as potential problems for Run 2.
Analysis-level event data model redesigned for Run 2. ("xAOD")
- Simplify, and make more directly usable with ROOT.
- Based on new "auxiliary data" feature of DataVector.
  - Attach data of arbitrary type to elements of DataVector.
  - Data are stored as vectors, managed via separate "auxiliary store" object via abstract interface.
  - Object data stored as auxiliary data rather than in the object itself.

Almost all object data stored as auxiliary data.

DataVector<Foo>

AuxElement 0
Foo

AuxElement 1
Foo

AuxElement 2
Foo

IAuxStore

vector<int> "anInt"

vector<float> "aFloat"
ATLAS has changed its event data model to unify Athena and ROOT analyses (xAOD).

To allow the reuse of metadata components in downstream analyses that are not utilizing the ATLAS control framework, Dual-Use Tools are being developed to summarize metadata records.

- Transfer some functionality from the framework MetaDataTools to Dual-Use Tools
- Provide generic MetaDataToolWrapper to allow framework integration
  - Listen to incidents
  - Interact with stores

Summary:
- Run 2 conditions further emphasize the importance of metadata for distributed data processing and analysis.
- At the same time, the move to new computing architectures requires extensions to the metadata infrastructure.
The Data Quality Monitoring Software for the CMS experiment at the LHC

#2 Mon: Experiment's Computing Models

M. Rovere

The Data Quality Monitoring Software for the CMS experiment at the LHC

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski

Track-2: Offline Software

DQM framework proved to be extremely flexible and stable: DQM is used everywhere in CMS

Summary:
- DQM Framework adapted and improved in the face of fundamental changes in the Online (DAQ2) and Offline (Multithreading) environments
Mini-AOD: A New Analysis Data Format for CMS

#2 Mon: Experiment’s Computing Models

C. Vuosalo

Overall size for ttbar MC: ~40 kB/event

Summary:
- Solution is compressed Mini-AOD
  - 10% of size of AOD
  - Replaces intermediate ntuples with standard format for most analyses

Previous method of storing AOD and many versions of intermediate ntuples would overflow storage capacity many times over.

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski
MAUS: The MICE Analysis and User Software

D. Rajaram

Summary:
- MAUS provides a simulation, reconstruction, and accelerator physics analysis framework for MICE

Frontend: Document-oriented database

Transform

Merge

Output

Web front-end

Histogram mergers
The NO\(\nu\)A Simulation Chain

#2 Mon: Experiment’s Computing Models

A. Aurisano

Simulation Chain

- Beam simulation → FLUKA/FLUGG
- Neutrino interactions → GENIE
- Detector simulation → GEANT4
- Parametrized front-end simulation
- Cosmic rays → CRY

**Summary:**
- Simulation at NO\(\nu\)A is a multi-stage process.
- Preliminary results show that low level quantities in cosmic ray data and Monte Carlo match well.

**Simulation Chain:**

- Simulates the incoming neutrino flavor and energy spectrum from \(\pi\), \(K\), and \(\mu\) decays
- Produces particle lists and kinematics to be propagated through the detector
- Propagates particles through the detector and produces energy deposits in active material
- Converts energy deposits into scintillation light, transports scintillation light to the APD, and simulates the readout response.
- Output looks like raw data
Heavy photons have recently become popular since they could explain experimental anomalies in particle physics (g-2) and astrophysics (e+ excess).

The Heavy Photon Search (HPS) Experiment is a fixed-target experiment that uses the JLab electron beam to search for heavy photons with masses 10-200 MeV.

Summary:
- Use of existing simulation and reconstruction software (developed for the ILC collider detectors) minimized the time needed to design and optimize the detector.
- Java-based reconstruction software working well.

Simulation & Reconstruction

- EGS5/MadGraph
- Geant4/SLIC
- Readout simulation/pileup
- Hit reconstruction, track finding, track fitting

Simulation

- Physics, matrix element
- Particle-matter interactions & propagation, geometry simulation
- Electronics simulation

Data

- MC particles stdhep
- Energy dep. in Si “G4 hits” slcio
- Raw data evio
- Pulse heights in ADC counts “raw” lcio
- Hits, tracks “recon’ed” lcio

Analysis

- evio-lcio converter
- Recon

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski

Track-2: Offline Software

CHEP 2015, Okinawa, 17.04.2015 18/47
Running ATLAS Workloads within Massively Parallel Distributed Applications using Athena Multi-Process Framework (AthenaMP)

#3 Tue: Frameworks

V. Tsulai

Summary:
- Recent tests on supercomputers demonstrated that AthenaMP can run efficiently within distributed applications at large scale.

Schematic view of ATLAS AthenaMP

Currently is actively used for running ATLAS production jobs on multi-core resources on the Grid.
A New Petabyte-scale Data Derivation Framework for ATLAS

J. Catmore

#3 Tue: Frameworks

Summary:
• The framework is in use already and the output data products are within the resource limits.
• It will be deployed from the start of data taking next month.

The framework is built on the main ATLAS data processing framework (Athena)

The framework is based on the main ATLAS data processing framework (Athena)

Provided by users

Job options

Athena event loop

Kernel

Modified data in StoreGate

Skimming

Thinning

Augmenting

Slimming

Persistency

Used for analysis

Keep event?

Allows corrections to be made to the reconstruction via “AODFix” (configured centrally)

xAOD

Reco fixes

Modified data in StoreGate

Persistency

Job options

Athena event loop

Kernel

Provided by users

xAOD

Reco fixes

Modified data in StoreGate

Skimming

Thinning

Augmenting

Slimming

Keep event?

Persistency

Used for analysis

The framework is built on the main ATLAS data processing framework (Athena)
Requirements for a Next Generation Framework: ATLAS Experience

S. Kama

- Single thread event processing is well established
- Running multiple instances of single threaded process is wasting memory
- Running event processing in multiple threads require separation of global components like multi-event services and local components such as per-event or sub-event algorithms and tools

Summary:
- Multi-threaded multi-event processing is required.
- Same framework should handle both offline and trigger use cases.
- Scheduler needs to be smarter and more complex than serial case.
- Whiteboard should handle multiple events on flight transparently to algorithms.

Future Frameworks Requirements Group started working spring 2014.

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisiel and J. Kowalkowski
Hybrid multiprocess / multi-threading framework implementation

- events are distributed to worker processes via AthenaMP shared event queue
- multi-threading within each worker
  - only one concurrent event, but multiple parallel algorithms in different threads
    - reduce thread safety issues
    - initial modeling on toy simulation shows ~8 parallel data paths for full reconstruction
  - full multi-threading, with multiple concurrent events and algorithms

Maximizes processor utilization while reducing memory footprint compared with pure event level concurrency

- eg 4 concurrent events:

<table>
<thead>
<tr>
<th></th>
<th>time /s</th>
<th>memory /MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>mp: 4 proc, 1 conc. event</td>
<td>32.8</td>
<td>1847</td>
</tr>
<tr>
<td>mp: 2 proc, 2 conc. events</td>
<td>34.4</td>
<td>1241</td>
</tr>
<tr>
<td>mp: 1 proc, 4 conc. events</td>
<td>53.6</td>
<td>935</td>
</tr>
<tr>
<td>hive: 4 concurrent events</td>
<td>47.8</td>
<td>817</td>
</tr>
</tbody>
</table>
Using the CMS Threaded Application in a Production Environment

#3 Tue: Frameworks

C. Jones

Evolved from original single-threaded application

- Updated algorithms can run concurrently (A1)
- Only one legacy algorithm is allowed to run at a time (A2)

Processes events in block concurrently

- Synchronizes on block boundaries (B1)

Summary:

- CMS has successfully transitioned to using multiple threads.
- Job success rate is sufficient for Run 2 data processing.
New Developments in FAIRROOT

F. Uhlig

#3 Tue: Frameworks

FAIR Message Queues (FairMQ)

Parameter File(s)
Database

Input File(s)
Online Data

Topology Definition

Task 1
C

Task 2
Fortran

Task 3
C++

Task 4

Task 5

Task 6

Computer 1
Linux

Computer 2
Windows

GPU on Computer 3

Output File

Parameter Manager

FAIRROOT used by ~10 experiments, mostly at GSI, some at CERN.

Summary:
• Each task is a single independent process.
• Each task sends its results to the next task in the row.
• Allow to setup exactly the needed chain.
ALFA: The New ALICE-FAIR Software Framework

Strategy: massive data volume reduction; coupling between online and offline.

Summary:
- ALFA is a tools set library (transport layer, configuration tools, management and monitoring tools)
- a data-flow based model (Message Queues based multi-processing)
- provides unified access to configuration parameters and databases.

M. Al-Turany
Physics Analysis Software Framework for Belle II

#3 Tue: Frameworks

The framework utilizes the Belle II software framework Basf2.

Steering example: reconstruction of $D^+ \rightarrow D^0\pi^+, D^0 \rightarrow K^-\pi^+$

```python
#!/usr/bin/env python
# -*- coding: utf-8 -*-

from basf2 import *
from modularAnalysis import *
from stdLoosePSParticles import *

inputMdst('DstarSignalMC.mdst.root') # define mdst input file

stdVeryLoosePi()  # make lists of very loosely selected pions (pi+:all, pi-:all)
stdLoosePi()      # make lists of loosely selected pions (pi+:loose, pi-:loose)
stdLooseK()       # make lists of loosely selected kaons (K+:loose, K-:loose)

# reconstruct D0 -> K- pi+ + cc
reconstructDecay('D0 -> K-:loose pi+:loose', '1.7 < M < 2.0 and p_CMS > 2.2')
vertexKFit('D0', 0.001)
applyCuts('D0', '1.81 < M < 1.91')

# reconstruct D*+ -> D0 pi+ + cc
reconstructDecay('D*+ -> D0 pi+:all', 'Q < 0.05')
vertexKFit('D*', 0.001)
applyCuts('D*', 'Q < 0.02 and p_CMS > 2.5')

outputMdst('recDstar.udst.root', ['D*+']) # write selected events to micro dst
process(analysis_main) # process events
```

Summary:

Although the framework is still under development, a user can already perform most of the physics analysis steps like decay reconstructions, vertex fits, tag the flavor of a B meson and perform TMVA-based continuum suppression.
Geant4
Version 10 Series

#4 Tue: Simulation

M. Asai

- Major release Geant4 version 10.0 was released on December 6\textsuperscript{th}, 2013.
  - The first major release since June 2007.
- There are several highlighted features including
  - Multithreading capability with event parallelism
  - Isomer production
  - Enhancements in biasing options
  - Introduction of phonon transport with a new concept of crystal
- Version 10.1 released on December 5\textsuperscript{th}, 2014 has lots of improvements in both physics and computing aspects.
  - Current version : 10.1-patch01 released on April 1\textsuperscript{st}, 2015
- Release of version 10.2 is scheduled on December 4\textsuperscript{th}, 2015
The Virtual Monte Carlo interface (VMC) allows to run simulation with different Monte Carlo codes from the same user application.

**Summary:**
- Geant4 VMC 3.0, multithreading capable, is available since November 2014.
- The interest in Geant4 multithreading in both ALICE and FAIR.
- Several improvements in build system and testing.

**Geant4 VMC History**

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Feature</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>v0.1</td>
<td>Oct 2002</td>
<td>First version</td>
<td>Oct 2002, First version</td>
</tr>
<tr>
<td>v1.0</td>
<td>Jul 2003</td>
<td>Geometry convertors</td>
<td>Root 3.03/09, Geant4 4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(root2g4, ...)</td>
<td></td>
</tr>
<tr>
<td>v0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v1.9</td>
<td>Dec 2006</td>
<td></td>
<td>Root 3.05/06, Geant4 5.2</td>
</tr>
<tr>
<td>v2.0</td>
<td></td>
<td></td>
<td>G4Root Navigation</td>
</tr>
<tr>
<td>2.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v3.0</td>
<td>Nov 2014</td>
<td></td>
<td>Multithreading CMake Root 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td></td>
<td></td>
<td>Root 5.34/23 and 6.02/01 Geant4 10.00.p03</td>
</tr>
</tbody>
</table>
**slic:**

**A Geant4-based Detector Response Simulation Program**

N. Graf

---

The goal of slic is to develop and distribute an executable program, which provides the full functionality and flexibility of the Geant4 toolkit, but insulates the end user from having to write any C++ code.

**Summary:**
- slic provides a complete and flexible detector simulation package capable of simulating arbitrary complex detectors with runtime detector description.
- Being used by LC (ILC & CLIC) detector community and by HPS experiment.

---

same event run with single executable with different input xml files

SiD

GLD

LDC

---

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski

Track-2: Offline Software

CHEP 2015, Okinawa, 17.04.2015 29/47
We developed three main components

1. A multithread scheduler to handle the particle baskets
2. A vectorised geometry library and navigator
3. A vectorised Compton scattering

Our results indicate that

- Basket handling introduces a minimal overhead
- SIMD gains half an order of magnitude in performance

We believe we are on track with our objective

An optimistic success-oriented interpretation of our results could be 1.4 (scheduler) x 4 (physics & geometry) ~ 6

Of course the proof of the cake will be in the eating
CMS Full Simulation for Run-2

#4 Tue: Simulation

D. Lange

Where our simulation CPU goes

- Transportation
- EM physics
- Hadronic physics
- SD/User actions
- Other

Technical performance improvements for Run 2 simulation:

1. Upgrade to Geant4 10.0p03 (~5%)
2. Implementation of Russian Roulette technique (~30%)
3. CMSSW code optimization (~15%)
4. Library repackaging (~10%)

Summary:

- Large gains achieved since Run 1. Now investigating other areas including interesting new components of Geant4.
- Deployment of multithreaded simulation planned for 2016.
CMS Detector Description for Run II and Beyond

#4 Tue: Simulation

G. Boudoul

Example of the Pixel SubDetector within the Tracker

Present Pixel Geometry (3 layers - 2disks)

Present Pixel or the future Pixel description can be exchanged within the Tracker

Future Pixel Geometry (4 layers - 3disks)
Continuous Readout Simulation with FairRoot for the PANDA Experiment

#4 Tue: Simulation

T. Stockmanns

Signal and background-events very similar → no hardware trigger possible.

13. April 2015

Folie 3

Tobias Stockmanns

Event Structure and Detector Response

Slow detector

Wanted data structure

Data structure in Root

Event 1

Event 2

Event 3

Summary:

- Time-based simulation part of FairRoot-Framework.
- Reconstruction based on time slices.

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski

Track-2: Offline Software

CHEP 2015, Okinawa, 17.04.2015 33/47
Motivation: ILC should be equipped with high precision detectors.

Summary:
- Analytic energy reconstruction method: energy resolution reaches < 6% at 80 GeV with satisfactory linearity.
- ANN technique also giving promised results.

#4 Tue: Simulation

Energy Reconstruction Study in a Semi-Digital Hadronic Calorimeter for ILC

S. Mannai

Energy reconstruction: Artificial Neural Networks

- TMultiLayerPerceptron of root package.
- 2 hidden layers with 6 and 2 neurons.
- The input variables: \( N_1, N_2, N_3 \).
- The output variable is the reconstructed energy: \( E_{\text{rec}} \).
- Monte Carlo Simulation
  - Training Samples: Odd energies, 1-99 GeV (50 training samples)
  - Test Samples: Even energies, 10-90 GeV (40 test samples)
Typical Machine Learning goal is event classification; try to minimize e.g. classification error rate.

Goal in HEP search is to establish whether event sample contains only background; rejecting this hypothesis \( \approx \) discovery of signal.

Often approach in HEP is to use distribution of MVA classifier.

Simplest case, use classifier to define “search region” and count:

\[ s = \text{expected number of signal events (assuming it exists)} \]
\[ b = \text{expected number of background events} \]

Goal: Minimize Approximate Median Significance of discovery:

\[
AMS = \sqrt{2 \left( (s + b) \ln \left( 1 + \frac{s}{b} \right) - s \right)}
\]

(Modified in the Challenge to prevent small search region where estimate of \( b \) may fluctuate very low: \( b \rightarrow b + b_{\text{reg}} \).)
In the LHCb experiment, a wide variety of Monte Carlo samples need to be produced for the experiment's physics program. Procedures based on common infrastructures have been set up to handle Monte Carlo productions centrally.

**MC requests workflow**

- **Collects needs and submit requests**
  - **Physics WG Liaison**
  - **New/Submitted**

- **Verification of technical consistency**
  - **Technical Expert**
  - **Tech Accepted/Rejected**

- **Appropriateness of the request for the physics program**
  - **Physics WG conveners**
  - **Phys Accepted/Rejected**

- **Accepted**
  - **Submit Production**
  - **Active**

- **Statistics required Produced**
  - **(MC) Production manager**
  - **Done**

**Summary:** Allows for massive transparent and efficient production on a world-wide distributed system with very little manpower.
The Library Event Matching Classifier for $\nu_e$ Events in NOvA

Library is 77M simulated neutrino interactions
- 18M $\nu_e$ charged-current signal events
- 29M $\nu_{\mu}$ charged-current events and neutral current events
- 30M $\pi^0$-enriched neutral-current events ($\pi^0 \rightarrow \gamma\gamma$ is primary background)

Library events are from NO$\nu$A Monte Carlo
- Flux from FLUKA/FLUGG simulation of beamline
- $\nu$ interactions simulated by GENIE
- Particle propagation in detector from GEANT4
- NO$\nu$A simulation code converts energy depositions to electronic signals

Performance scales roughly logarithmically with library size

Flipping events transversely effectively quadruples library size
- Events can be flipped in either view (x,y) or both
- Symmetry is not perfect, but close enough
- Beam direction is slightly upward ($\sim 3^\circ$)
- Attenuation in fiber varies with distance from readout
Higgs boson rate in $H \to WW^*$

- Simulated samples available in 5 GeV in $m_H$

Higgs boson $CP$ measurement

- Combination of 1d ($H \to WW^*$) and 2d ($H \to ZZ^*$) moment morphing

**ATLAS**

$H \to WW^* \to l\nu l\nu$

- $\sqrt{s} = 7$ TeV, 4.5 fb$^{-1}$
- $\sqrt{s} = 8$ TeV, 20.3 fb$^{-1}$

- Obs ($\hat{m}_H = 128, \hat{\mu} = 0.94$)
- Obs ± 1σ
- Obs ± 2σ
- Obs ± 3σ

### Summary:

- New algorithm to interpolate between Monte Carlo sample distributions; generic, simple, fast, stable and accurate.
CosmoSIS: Modular Cosmological Parameter Estimation

A. Manzotti

**Summary:**
HEP might consider adopting a similar to CosmoSIS attribution concept to help encourage sharing useful software.

CosmoSIS is a modular framework for parameter estimation

- **APIs:**
  - Define module configuration, how to obtain their input data, how they interact with the framework, how outputs are organized, and how new modular components can be added.

- **Runtime environment:**
  - To reliably configure and run programs that use these modules.

- **Development environment:**
  - To write, build, and test analysis software, and which makes it easy to share what they have developed. Registration of attribution information (required citations, etc.) for use of any contributed code.

- **Distribution system:**
  - To make it easy to install the code and ensure compatibility.
Given data from LHC, what are likely values of masses, cross sections...?

**Summary:**
- **BAT well established**
- More powerful sampling algorithms in BAT 2.0

**Bayesian Analysis Toolkit: 1.0 and beyond**

F. Beaujean

\[
P(\theta | D, M) \propto P(D | \theta, M) P_0(\theta | M)
\]

**BAT components**

**USER DEFINED**
- create model
- read data

**COMMON TOOLS**
- Normalize()
- FindMode()
- MarginalizeAll()
- PrintAllMarginalized()
- PrintKnowledgeUpdatePlots()

**DEFINE** MyModel : BCMModel
- AddParameter("mu", 0, 1)
- LogLikelihood()
- LogAPrioriProbability()
Using R in ROOT with the ROOT-R Package

#6 Thu: Tools

L. Moneta

- ROOT plugin for Minimisation implemented using R
  - developed by Kirby Hermann (GSOC student 2014)
  - give access to R optimisation tools when fitting or multi-dimensional function minimisation
  - based on R optim and optimx packages

```c
ROOT::Math::MinimizerOptions::SetDefaultMinimizer("RMinimizer","L-BFGS-B");
hist->Fit("gaus");
```

Example of RMinimizer

- ROOT-R package gives access to ROOT users to the R capabilities and its rich functionality.
- ROOT-R package is ready to be released in the next ROOT production version (6.0.4)
- Easy to use directly R from ROOT prompt
- ROOTR provides easy access to R tools in ROOT and C++
Interpreting results: testing models

Several interpretations/hypothesis tests available (RooStats-based) and macros to interpret and present the results.

- **Hypothesis tests**: test a specific alternative (signal) model

  HistFitter provides tools to run over sets of models and graphically present the results.

- **Upper limit calculations**: performs repeated hypothesis tests

  Determines 95% CL cross-section upper limit for specific signal models

\[ m_{1/2} \text{ [GeV]} \]

\[ m_0 \text{ [GeV]} \]

\[ \text{Expected limit (}\pm 1 \sigma_{\text{exp}}) \]

\[ \text{Observed limit (}\pm 1 \sigma_{\text{theory}}) \]

**Note:** dummy figure!
SciDB is an Array Database management system and analytic platform, scalable out to 1,000s of processors and TBs - PBs of data.

Summary:
- Lots of optimized functions: sampling, matrix operations (ScalAPACK), statistics, user designed functions
- Parallel architecture that is transparent to the user

SciDB scales well

Y. Yao et al., Computing in Science & Engineering, May 2015

- Return for all files
- Return for a specific file
- Estimate of performance at large-scale

Return max(l) for narrow range of m/z and r.t. (sec)

25% of the total dataset

Number SciDB instance in the cluster
DDG4 - a Simulation Framework based on DD4hep and Geant4

#6 Thu: Tools

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski

Track-2: Offline Software

CHEP 2015, Okinawa, 17.04.2015  44/47
• Data stored in DST files vs. data stored in HBase
  • 3.5 times faster with pre-selection

• 20% faster without pre-selection
Breaking the Silos: 
The *art* Documentation Suite

**Introduction**
1. What is a framework
2. Define Prerequisites
3. Overview of documentation
4. ...

**Workbook**
- Exercise 1
  - Activity 1
  - Activity 2
- Exercise 2
  - Activity 1
  - Activity 2
- ... 
- Exercise N

**Reference Manual**

**Technical Reference**

**External Refs**
- Root, C++, STL, G4
  - ...

**Users Guide**
- Table of Contents
  1. Users view of *art*
  2. Modules
  3. Services
  4. FHiCL
  5. Interface to G4
  6. Interface to SAM
  7. ..... 
- Appendices
  1. Best Practices
  2. Trouble Shooting
  3. CLHEP gap Filler Docs
  4. Glossary
  5. Index

**Status of the documentation suite**

- 90% Complete
- 25% Complete
- 5% Complete

Summary:
- There is a plan for an integrated *art* documentation suite, which still a mostly volunteer effort, new volunteers welcome.

A. Norman, E. Sexton-Kennedy, H. Kurashige, I. Kisel and J. Kowalkowski
Track-2: Offline Software
CHEP 2015, Okinawa, 17.04.2015  46/47
Summary

We can more:
• more optimisation
• more compactness
• more vectorization
• more cores
• more events/s
• more expect at CHEP2016