The Large Hadron Collider

From Big Data to Physics Discovery



Markus Elsing

Target Seminar, Groningen, 24.9.2013



Introduction: LHC and Experiments





Introduction: LHC and Experiments





Introduction: LHC and Experiments







Colliding Protons in the Experiments





... at a rate of 40 million events per second

Markus Elsing

The ATLAS Detector as a "Camera"





the detector has has about 90 million channels, 2.5 MB per event at a rate of 40 MHz (!)

Markus Elsing

Introduction: LHC

- LHC is a high energy and high luminosity proton-proton collider
 - → design centre-of-mass energy is 14 TeV and design luminosity is $\mathscr{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$
 - ➡ first collider to reach energy regime of high energy cosmic rays (HECR)
 - ➡ expect ~23 p-p collisions at a bunch crossing frequency of 40 MHz (!)

• LHC is a unique machine

- → first collider to explore the physics at the TeV scale
- ➡ excellent sensitivity to rare (new physics) processes

expected production cross-sections

- → large inclusive b, W/Z and top production rates
 - LHC is a combined b-, W/Z- and top-factory
- cross-section for jet and W/Z production orders of magnitude larger than e.g. expected for Higgs
- → total cross-section dominated by soft interactions







Introduction: LHC Physics Programme

- proton-proton programme:
 - I. mass and electroweak symmetry breaking
 - search for the Higgs Boson, measurement of its properties
 - II. hierarchy in the TeV domain
 - search for new phenomena moderating the hierarchy problem
 - search for the unexpected at the high-energy frontier
 - III.electroweak unification and strong interactions
 - precision measurements (m_{top}, M_W) and tests of the Standard Model
 - tests of perturbative QCD at the high-energy frontier IV. flavour
 - B-,D-mixing, rare decays and CP violation as tests of the Standard Model
- heavy ion programme: (not covered here)
 - study quark-gluon plasma in Pb+Pb collisions at up to 5.5 TeV per colliding nucleon





Peter Higgs visiting CERN in 2008

LHC Run-1 from 2010 until 2012

• first LHC running period

- → 2010+2011at **7** TeV and **8** TeV in 2012
- ➡ increase in centre-of mass energy yields increase in parton luminosity, especially for large M_X processes
- but jet, W/Z and top cross-sections scale fast, background for new physics searches

outstanding LHC performance

- ⇒ peak luminosity of 7.7 × 10³³ cm⁻²s⁻¹ with half the number of bunches
- → more than 25 *fb*⁻¹ in total recorded in Run-1
- currently in 2 year shutdown
 - → restart in 2015 at 13-14 TeV

presented in the following



→ 7 TeV and latest 8 TeV physics results

➡ status mostly summer 2013



W.J. Stirling, private communication



Markus Elsing

High Luminosity comes at a Price

• typical LHC event in 2012

- ➡ large number of interactions in 1 event
 - so-called event "pileup"
- → exceeding detector design levels (!)

• challenge for the experiments

- → detector (including readout)
- trigger: select interesting interactions, keeping acceptable total rate
- data volume: from the detector recorded on tape and to be processed/analyzed on computing GRID worldwide
- reconstruction and analysis: make sense out of these very complex events and extracting interesting physics information

let's look at processing chain

how do we deal with all the data ?







Basics: Event Reconstruction "in a Nutshell"





Basics: Event Reconstruction "in a Nutshell"





Basics: Event Reconstruction "in a Nutshell"







... a bit more complicated





Run Number: 183081, Event Number: 10108572

Date: 2011-06-05 17:08:03 CEST

Online and Offline Data Processing

event selection during data taking - the Trigger

➡ experiments produced ~ 10 PB of raw data every year during Run-1



40 MHz, 100 TB/s

Level-1 Trigger: special electronics



20 kHz, 50 GB/s

Level-2+3 Trigger: software, PC farm



400 Hz , 1 GB/s



Online and Offline Data Processing

event selection during data taking - the Trigger

➡ experiments produced ~ 10 PB of raw data every year during Run-1





40 MHz, 100 TB/s

Level-1 Trigger: special electronics



20 kHz, 50 GB/s

Level-2+3 Trigger: software, PC farm



400 Hz , 1 GB/s



event reconstruction

- → sophisticated pattern recognition algorithms
 - especially for track reconstruction
 - → Trigger: only Regions of Interests (Rol)
 - → Offline: full event reconstruction
- CPU limited due to pileup (!)



Markus Elsing

Trigger Selection Strategy

total p-p cross section

➡ dominated by soft interactions (QCD)

W,Z,top,Higgs production

- ➡ orders of magnitude smaller cross section
- \Rightarrow events with high p_T jets/e/μ/τ/γ or missing E_T
 - they define Regions of Interests (Rol) in events

• Trigger strategy:

- → Level-1: (special) fast readout, search for candidates
- → Level-2: partial event readout to reconstruct Rols
- ➡ Level-3: full event building and use offline software



fast rejection

- ➡ reject events early
- ➡ minimize data movement
- selected events (raw data) stored for offline processing



W.J. Stirling, private communication

LHC Raw Data vs Commercial World





LHC Raw Data vs Commercial World





➡ LHC total data volume not so different from Google and Facebook (@ science budget)

Processing and Analysis Model

real data processing chain



- ➡ output of reconstruction input to analysis
- ➡ next step is data reduction in group analysis
- → selected data is subject of final analysis



Final Analysis



Processing and Analysis Model

complemented by simulation chain



- → physics generators simulate physics event
- ➡ particles tracked through detector simulation
- ➡ detector response simulated in digitization





Processing and Analysis Model



The Simulation Chain

Monte Carlo physics generators

- ➡ sophisticated modeling of p-p interactions
 - proton parton distribution functions (PDFs)
 - matrix elements calculations
 - fragmentation and particle decay models
- → used to produce all kinds of physics channels

• very detailed detector simulation

- track particles through detector material
 - particle transport in b-field
 - complex geometry model
 - emulate full physics of passage of particles through matter (secondaries, showers...)
- ➡ energy deposits in sensors

detector digitization

- sophisticated emulation of sensor response
 - to energy (ionization) signals and of readout
 - output "looks like" real data (incl. pileup)



	model	placed volumes	
ALICE	Root	4.3 M	
ATLAS	GeoModel	4.8 M	
CMS	DDD	2.7 M	
LHCb	LHCb Det.Des.	18.5 M	



Computing Model of the Experiments



• tasks of Offline Computing:

- → reconstruction of raw data events (mainly at Tier-0)
- ➡ production of large Monte-Carlo (simulated data) samples
 - generators for different physics processes
 - very detailed simulation of detector response (Geant4)
 - same reconstruction applied to Monte-Carlo samples
- physics analysis of data sets by research teams

Worldwide LHC Computing Grid (WLCG)

- → federation of more than 150 computing centers worldwide
 - GRID middleware as common layer, experiment specific Production Systems and Data Management Systems
- ➡ hierarchically organized (Monarc Model):
 - Tier-0 at CERN (for prompt reconstruction of raw data taken)
 - Tier-1 centers worldwide, regional associated Tier-2 centers
 - small Tier-3 farms (for local data analysis at institutes)
- data distributed worldwide from CERN
 - Tier-1 act as centers for processing of large data sets
 - Tier-2 act mainly CPU farms for simulation





WLCG

NIKHEF



WLCG - truly Worldwide Computing



• 150 computing centers distributed over 40 countries

- → 300+ k CPU cores (~ 2M HEP-SPEC-06)
 - the biggest site with ~50k CPU cores, 12 Tier-1 with 2-30k CPU cores
- → each Tier center is small compared to a typical High Performance Centers
 - WLCG is optimized for data intensive processing with distributed data, services and operation infrastructure



Example: ATLAS GRID Statistics



→ ATLAS: ~ 3 billion raw and 3-5 billion simulation events per year of data taking

Markus Elsing

Evolution of WLCG Resources

upgrades of existing centers

- ⇒ additional resources expected mainly from advancements in technologie (CPU or disk)
- → will not match additional needs in coming years

todays infrastructure

- ➡ X86 based, 4 GB per core, commodity type CPU servers
- ➡ applications running "event" parallel on separate cores
- ➡ jobs are send to the data to avoid transfers

technology is evolving fast

- → network bandwidth fastest growing resource
 - data transfer to remote jobs is less of a problem
 - strict Monarc Model no longer necessary
 - flexible data placement with data popularity driven replication
- → modern processors: vectorization of the applications
- "many core" processors like Intel Phi (MIC) or GPGPUs
 - much less memory per core !

HEP Applications on modern Processors

high energy physics applications

- → optimized for precision and physics performance
- → more than 10 years of software development
 - projects have several 100 developers (students...)
 - complex applications, highly specialized

vectorization (SIMD)

- ➡ modern processors have long registers
 - big potential in vectorizing the code
 - requires good C++ knowledge from developers
- → experience with auto-vectorization in compiler
 - HEP software not easy to optimize
- → dedicated vectorization of crucial parts of the software
 - i.e. Runge-Kutta b-field transport of charge particles in simulation and reconstruction
- → vector libraries for linear algebra or trigonometric functions

• optimize data model for complex calculations

Development towards Multi-Threading

todays applications are "event" parallel

- → typically 2-4 GB of memory required per core
- ➡ not yet a problem with commodity CPU servers

• future "many core" processors

- → much less memory per core
- → todays applications are problematic
 - memory access and bandwidth limits, etc.

short term developments

- multi-threading to reduce memory requirements
- ➡ applications based on i.e. Intel TBB
 - software frameworks support thread parallelism
 - first prototype applications exit (CMS, ...)

in the long term adapt the applications

- → simulation: dedicated tasks are assigned to cores
- → GaudiHive framework: parallel processing of individual algorithms

High Performance Computing and HEP

infrastructure is getting heterogeneous

- ➡ mostly opportunistic usage of additional resources.
 - commercial Cloud providers (i.e. Google, Amazon)
 - free CPU in High Performance Computing centers
- ➡ big HPC centers outperform WLCG in CPU
 - X86, BlueGene, NVIDIA GPUs, ARM, ...
- → GRID (ARC Middleware) or Cloud (OpenStack) interface

suitable applications

- CPU resource hungry with low data throughput
- → physik generators, maybe detector simulation
- X86 based systems
 - → small overhead to migrate applications

• GPU based systems

→ complete rewite necessary (so far) or dedicated code

➡ i.e. physics generators VEGA or Baes/Spring

Markus Elsing

Physics Groups: Data Analysis on GRID

- → international research teams (students, Postdocs,...) cover physics subjects
 - development of sophisticated techniques
 - comparing data and Monte Carlo simulation, extracting of physics results
- → huge physics spectrum covered by the experiments
 - results are often presented only weeks/months after data taking finished
 - so, let's look at the physics now...

Jets Production and Underlying Events

underlying events

- pre-LHC models predict e.g. too little transverse activity, region sensitive to multi-parton interactions
- ➡ LHC results basis for improved MC tunes
 - good description achieved with modern codes like PYTHIA8 or HERWIG++
 - cosmic air shower models yield as well a good description (EPOS)
- jet production
 - ⇒ excellent description by pQCD over many orders of magnitude (LHC covers huge range in p_T and |y|)

based on PDFs, constrained by HERA and TEVATRON

Standard Model Measurements

• W[±]/Z, W/Z+jets and di-boson production

- ➡ important tests of SM
- ➡ background for searches (Higgs)

• W and Z studies

- ➡ huge event rates
 - heavily used for calibration
- W/Z rapidity distribution sensitive to strange quark sea contribution in proton PDFs
 - ATLAS compatible with no strange sea suppression

di-boson production

can put limits on anomalous
Triple Gauge Couplings

Top Cross-Sections and Mass

• LHC is a top factory

→ tt cross-section is large ~200 pb (~4 million events so far)

rich top physics program

• top pair and single top production

- → several channels accessible, even all hadronic
- → 7 and first 8 TeV results in agreement with SM

precision top mass measurements

- → derive mt from kinematic mass reconstruction
- ➡ already systematically dominated (jet energy scale, ...)

Mt and Electroweak Fit

direct top and W mass measurements

- currently still based on TEVATRON results
- compatible with combined fit to electroweak precision data and a light Higgs
- ➡ as well with MSSM

precise measurements of top mass

experimental observable and pole mass ?

$$m_t^{\exp} = m_{pole} (1 \pm \Delta)$$

• kinematic reconstruction from uncolored final state

- sensitive to hadronisation effects (color reconnection...)
- ➡ determine running mass (MS-scheme) from CDF/DO top pair cross-section at NNLO, yields:

 $m_{pole} = 173.3 \pm 2.8 \, GeV$

- close to world average, factor 4 larger uncertainty
- PDF and α_s uncertainties currently limiting for LHC,

may be reduced in the future ?

Alekhin, Djouadi, Moch, arXiv:1207.0980v2

Heinemeyer, Hollik, Stockinger, Weiglein, Zeune₂₈

Searches for the SM Higgs

• SM Higgs phenomenology

- ➡ precisely predicted, but Higgs mass
 - NLO and NNLO calculations (typical $\sigma \sim$ 5-15%)
 - production dominated by gg fusion, then vector boson fusion (VBF), associated WH and ZH
- ➡ cross-section and branching ratios are strong function of M_H

• Higgs searches using 2011 data

- ➡ both experiments saw hints for a light Higgs
 - around ~3σ each, ignoring "look elsewhere effect"
 - indications as well in TEVATRON data
- → low mass region at LHC
 - many decay modes accessible (γγ,ZZ,WW,ττ,bb)
 - γγ and ZZ yield excellent mass resolution (~1%)
- ➡ challenging to control backgrounds, except for ZZ

experiments "blinded" the 2012 data

- not to bias results on new data
- huge effort to optimize expected sensitivity (pileup)

The Discovery of the Higgs Boson

CERN Seminar 4.July 2012

CERN

Fabiola Gianotti, Joe Incandela

Markus Elsing

Results on 4.July 2012

ATLAS and CMS announced discovery

- → based on 7 TeV data and fraction of 8 TeV data
- ➡ both experiments excluded full mass rage, but window around ~125 GeV
 - including channels sensitive to high m_H
- \rightarrow best sensitivities in $\gamma\gamma$ and ZZ(4I) channels

local significance at min. p₀:

ATLAS	CMS	
5.9 σ obs.	5.0 σ obs.	
4.9 σ exp.	5.8 σ exp.	•

- → since then we more than doubled the data set
 - $\bullet\,$ individual channels now passed 5 σ

Higgs Signal Strength and Production

- full Run-1 data allows more detailed studies
- signal strength ($\mu = \sigma/\sigma_{SM}$) for channels
 - → within uncertainties agreement with Standard Model
 - \Rightarrow combined: ATLAS μ =1.23±0.18, CMS μ = 0.80 ± 0.14

study individual production modes

→ ATLAS 3.3 o evidence for VBF, CMS 3.2 o evidence for VBF, VH

allows to determine Higgs couplings

- → combining information from production and decay
- \rightarrow at LHC this requires some model assumptions (about Γ_{H})
 - deduce Higgs couplings as a function of mass

- σ(stat)

ATLAS

Anomalous Higgs Couplings

popular model for interpretation:

- ➡ scale Higgs-vector boson (KV) and Higgs-fermion couplings (KF)
- combined production and decay results

indirect constraints from EW data

→ observables modified by loop corrections

- ➡ compare result from Higgs production at LHC and fit to LHC + electroweak precision data
 - so far **Ky** from electroweak fit to LEP, Tevatron and LHC data more precise
 - both agree well with Standard Model within uncertainties

Z+X Zγ',Z

• already a precision measurement !

Spin and Parity of Higgs Boson

• experimental approach

- \rightarrow observables sensitive to spin and parity (JP) like decay angles
- \rightarrow test alternative models with 0-, 2+ against 0 4odel)

 - observation of $H \rightarrow \gamma \gamma$ excludes J=1 $q = \log \frac{\mathcal{L}(J^P = 0^+, \hat{\mu}_{0^+}, \hat{\hat{\theta}}_{0^+})}{\mathcal{L}(J^P_{alt}, \hat{\hat{\mu}}_{J^P_{alt}}, \hat{\hat{\theta}}_{J^P_{alt}})}$
- → test hypotheses pairwise against data
- → exclusion of 0- w.r.t. 0+:
 - ATLAS: 97.8% C.L., CMS: 99.8% C.L.
- \rightarrow exclusion of J=2 models:
 - CMS: 99.4% C.L., ATLAS: >99.9% C.L.

• both experiments compatible with 0+!

Higgs studies at LHC just a starting !

- \Rightarrow expect 350 fb⁻¹ until 2021, HL-LHC aims for 3000 fb⁻¹ until 2030
 - LHC Run2+3 will enable us to measure $H \rightarrow \tau \tau$ and bb, ttH, (...) and to much improve the precision on Higgs couplings

- HL-LHC will see more rare channels opening up (e.g. $H \rightarrow \mu\mu$)
- measuring Higgs self coupling λ probably difficult

Searches for Supersymmetry

excluded up to $\sim 1.7 \text{ TeV}$ for $m(\tilde{q})=m(\tilde{g})$

motivations for (minimal) SUSY

- ➡ provides solution for hierarchy problem
- Higgs mechanism for EWSB is built in predicts a light Higgs
- ➡ unification of couplings
- ➡ R-parity: LSP is dark matter candidate

• SUSY is broken

- ➡ plenty of SUSY breaking models (CMSSM, ...)
 - different sets of free SUSY parameters
 - each model has rich phenomenology

• LHC results disfavor CMSSM

- ➡ no light SUSY discovered (so far)
- ➡ Higgs at *125.7 GeV* still within SUSY reach
- → constraints from rare B decays ($B_s \rightarrow \mu \mu$...)
- instead, "bottom up" approaches
 - → phenomenological SUSY model (pMSSM)
 - simplified models to express results for SUSY s-particle searches

excluded $m(\tilde{g}) < 1.3$ TeV for any $m(\tilde{q})$

"Natural" SUSY ?

• not fine tuned Higgs requires:

$$\delta m_H^2 = \cdots \delta M \cdots + \cdots \delta N \cdots \sim 0$$

s-particles linked to Higgs loop need to be light

• 3rd generation squarks

 cross-sections at LHC expected to be smaller than for 1st and 2nd generation

• generic SUSY searches at LHC

- → like: "0-lepton" (signature: jets + missing ET)
 - interpretation in simplified model
- ➡ yield stringent limits on 1st and 2nd gen.
 - excluded up to ~1.7 TeV for $m(\tilde{q})=m(\tilde{g})$
- ➡ not constraining 3rd generation squarks
 - needs specialized \tilde{t} and b searches

Dedicated Stop Searches

simplified models

→ assumes 100% branching ratios

• gluino mediated Stop

- ➡ 4 top squarks in final state
- ➡ modes via virtual/on-shell stop
 - but limit on m(g̃) depends little on m(t̃) above/below m(g̃)

 $m_{\widetilde{\chi}_1^0}$ [GeV]

⇒ sensitive to $m(\tilde{g}) < 1300 \text{ GeV}$ for $m(\tilde{\chi}_1^0) < 550 \text{ GeV}$

direct Stop pair production

- \Rightarrow 2W+2b-jets+missing E_T
- \rightarrow modes with $m(\tilde{t})$ above/below m(t)
 - combination of several signatures to maximize sensitivity

"If you cover the white then weak scale SUSY is probably dead" R. Barbieri (ICHEP'12)

No TeV Scale New Physics (yet)

huge list of experimental signatures and models covered

- ➡ typical limits achieved up to:
 - singly produced objects with QCD couplings ~ 3.5 TeV
 - singly produced objects with EW couplings ~ 4 TeV
 - pair produced objects with QCD couplings ~ 600 TeV
 - unitarity limited rates ~ 4 TeV
 - compositeness scale ~ 8 TeV

• details in figures...

CERN

Indirect Constraints on new Physics

- mixing induced CP violation in $B_s \rightarrow J/\Psi \phi$
 - CP violating phase ϕ_s^{ccs} in B_s mixing-decay interference
 - ➡ golden mode:
 - sensitivity to new physics entering mixing between 2nd and 3rd quark generation
 - precise SM prediction, tiny theoretical uncertainty (assume $\phi_s^{ccs} \approx -2\beta_s$)
 - ➡ LHCb does time dependent analysis (of tagged events)
 - recently (untagged+tagged) ATLAS results
 - ➡ so far consistent with SM prediction
 - remains priority to improve precision

LHCb	current	I0 fb⁻ ^I	50 fb ⁻¹	Theory
2β _s (Β _s →J/Ψφ)	0.1	0.025	0.008	0.003
		arXiv:1208.3355v2		

Indirect Constraints on new Physics

• search for new physics in $B_{(s)} \rightarrow \mu\mu$

- → new physics in loop effects vs precise SM predictions
 - high sensitive to models with extended Higgs sector
 - $B \rightarrow \mu \mu / B_s \rightarrow \mu \mu$ probes minimal flavor violation
- \rightarrow EPS'13: first observation of $B_s \rightarrow \mu\mu$ by LHCb+CMS:

SM

h,A,H

What if SM unchanged up to M_{PI}?

• no new physics up to very high scales ?

- → Standard Model defines running of couplings
- \rightarrow special meaning of $\lambda \approx 0$ at M_{Pl}?

absolute vacuum stability with Higgs self coupling $\lambda(M_{PI}) \ge 0$?

- \rightarrow not quite for current "best" values of M_t and M_H
- → Standard Model vacuum probably metastable with lifetime >> age of universe

 $M_t = 171.0 \, \text{GeV}$

 $(t_7) = 0.1205$

 $(M_7) = 0.1163$

 $M_{\rm e} = 175.3 \, {\rm GeV}$

Summary and Outlook

the LHC, the experiments and their computing are doing fantastically well

- → excellent data taken during Run-1, distributed and processed on WLCG
 - GRID computing is a success for large sale data intensive physics analysis
- → very rich harvest of physics results, much broader than any talk could cover

• the Higgs Boson has been discovered !

→ its properties are (so far) compatible with Standard Model predictions

• LHC is a discovery machine for new physics

- ➡ experiments cover a huge spectrum of signatures and BSM models
- ➡ no signs for TeV scale physics beyond the Standard Model yet

• this is just the start

- → machine upgrade from 8 TeV to close to 13 TeV in 2013/2014 shutdown
 - expect to take ~*350 fb*⁻¹ at *14 TeV* until 2021

 \Rightarrow HL-LHC will take LHC program until 2030, for a total of 3000 fb⁻¹

Acknowledgements

many thanks to all the people who helped in preparing this talk, for the useful discussions and their suggestions...

... and of course, all material presented here is the result of the fabulous work by many, many people...

