



Datenverarbeitung am CERN und den LHC-Experimenten

Niko Neufeld PH/LHCB

Slides (teilweise) von Jan Fiete Grosse-Oetringhaus, CERN
PH/ALICE und Ian Bird, CERN/IT

Oesterreichisches Lehrerprogramm

26. November 2013



Datenverarbeitung

- „Datenverarbeitung (DV) bezeichnet den organisierten Umgang mit Datenmengen mit dem Ziel, **Informationen** über diese Datenmengen **zu gewinnen** oder diese Datenmengen **zu verändern**“ (Wikipedia)
- Beispiele
 - Taschenrechner (Verarbeitung der Daten)
 - Erfassung von Noten
 - Zeugnisdruck (Automatisierung, Archivierung)
 - Statistik: Klassendurchschnitt, historische Entwicklung (Verarbeitung)
 - Flugpassagierlistenauswertung
 - Rasterfahndung (Verarbeitung, Reduktion)



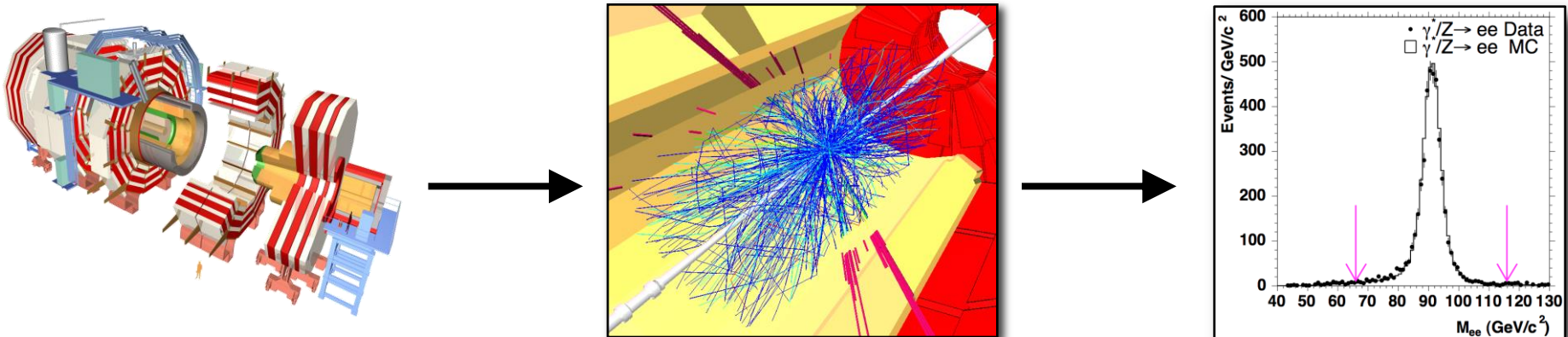
Datenverarbeitung am CERN

- “Klassische” Datenverarbeitung (Personalverwaltung etc)
- Webserver, Email-Infrastruktur
 - Ein typischer CERN-Mitarbeiter bekommt 100-1000 EMail pro Tag und versendet 10-100 (von den erhaltenen emails sind viele automatisch generiert von Ueberwachungssystemen usw)
- Computersysteme zur Steuerung, Ueberwachung und Verwaltung der Beschleuniger und Experimente
- Verarbeitung der experimentellen Daten



Datenverarbeitung an den LHC-Experimenten

- Die Datenverarbeitung ermöglicht aus “den Kollisionen” ein Physikergebnis zu extrahieren
- Benötigte Schritte
 - Auslösen der Datennahme (“Trigger”)
 - Auslese und Speicherung der Daten
 - Interpretation der einzelnen elektronischen Signale → Physikobjekte z.B. Teilchen (“Rekonstruktion”)
 - Interpretation der Physikobjekte → Physikergebnis (“Analyse”)





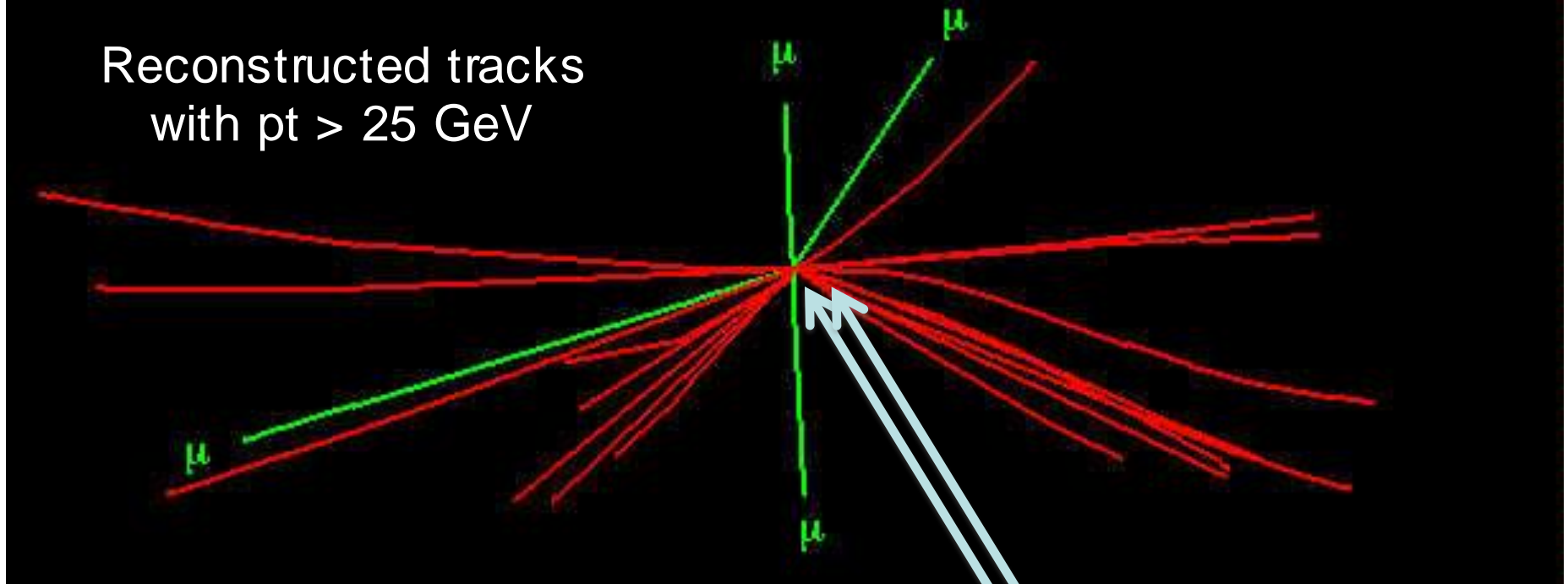
1 Byte	= 1 Buchstabe	
100 Byte	= 1 SMS	
1.024 Byte	= 1 Kilo byte	= 1 eMail
1.048.576 Byte	= 1 Mega byte	= 10 Min. Telefongespräch / 1 Foto / 1 LHC Kollision
734.003.200 Byte	= 700 Megabyte	= 1 CD-ROM
1.073.741.824 Byte	= 1 Giga byte	= Daten die ein LHC-Experiment pro Sekunde produziert
5.046.586.572 Byte	= 4,7 Gigabyte	= 1 DVD
1.099.511.627.776 Byte	= 1 Tera byte	= Bibliothek mit 1 Million Büchern
219.902.325.555.200 Byte	= 200 Terabyte	= 10 Milliarden Webseiten / Größe der amerikanischen Kongressbibliothek
1.801.439.850.948.198 Byte	= 1,6 Peta byte	= Weltweit produzierte Information in Papierform
11.258.999.068.426.240 Byte	= 10 Petabyte	= jährliche LHC Datenproduktion
5.764.607.523.034.234.880 Byte	= 5 Exa byte	= jährliche weltweite Datenproduktion in der Form von Radio, Fernsehen etc.
19.599.665.578.316.398.592 Byte	= 17 Exabyte	= Datenvolumen aller Telefongespräche eines Jahres

Folie teilweise uebernommen von Andreas Hirstius



Die Nadel im Heuhaufen

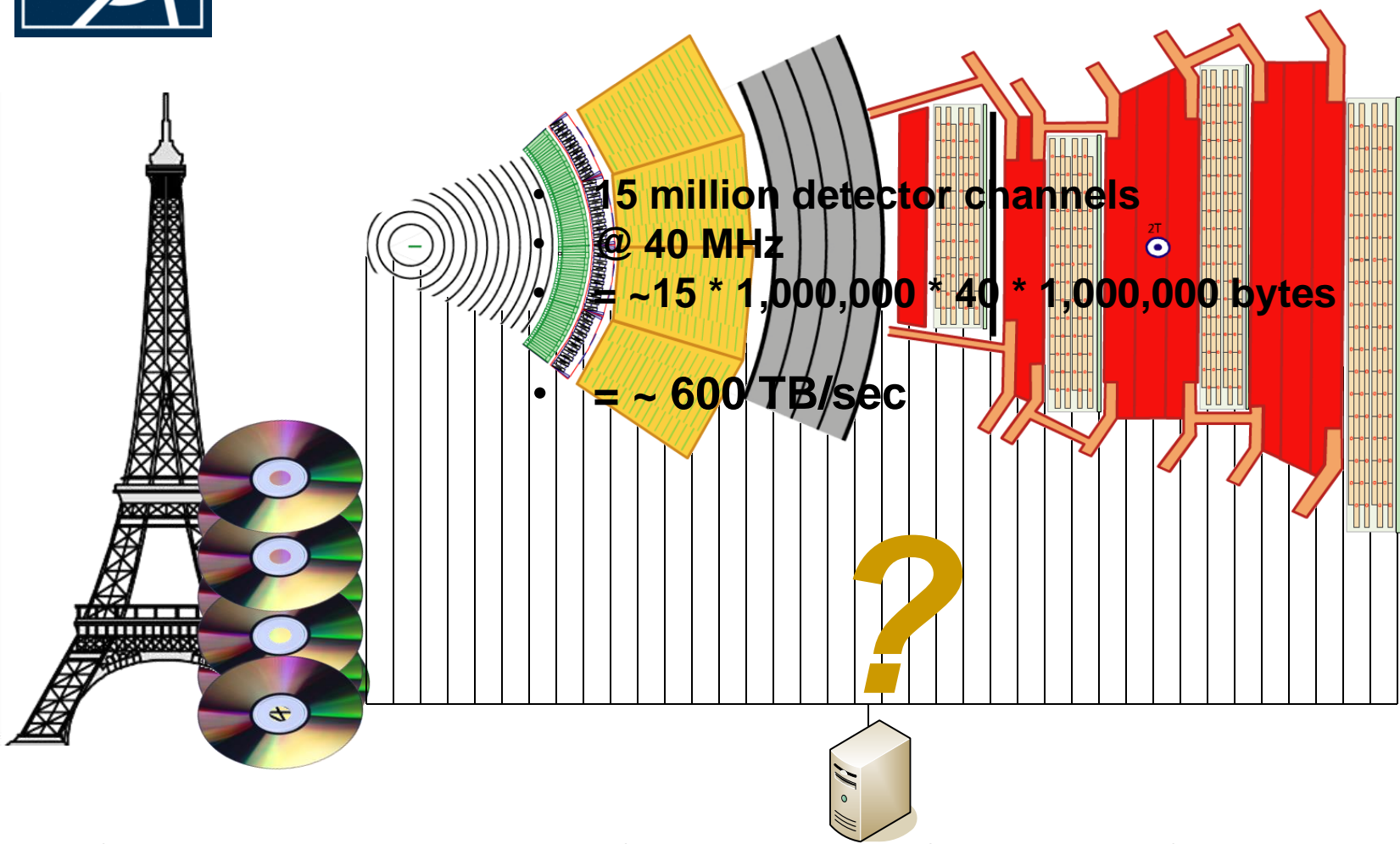
Simulation from CMS



**This is what we find looking for a 100 GeV signal
 We get this 4000000000 times per second!**



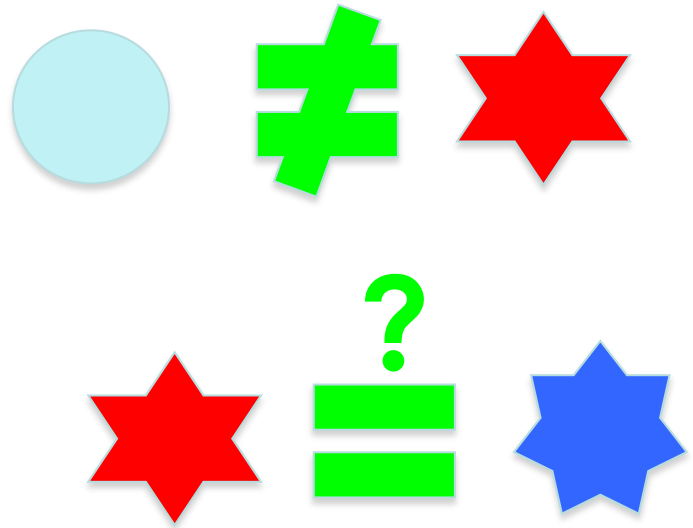
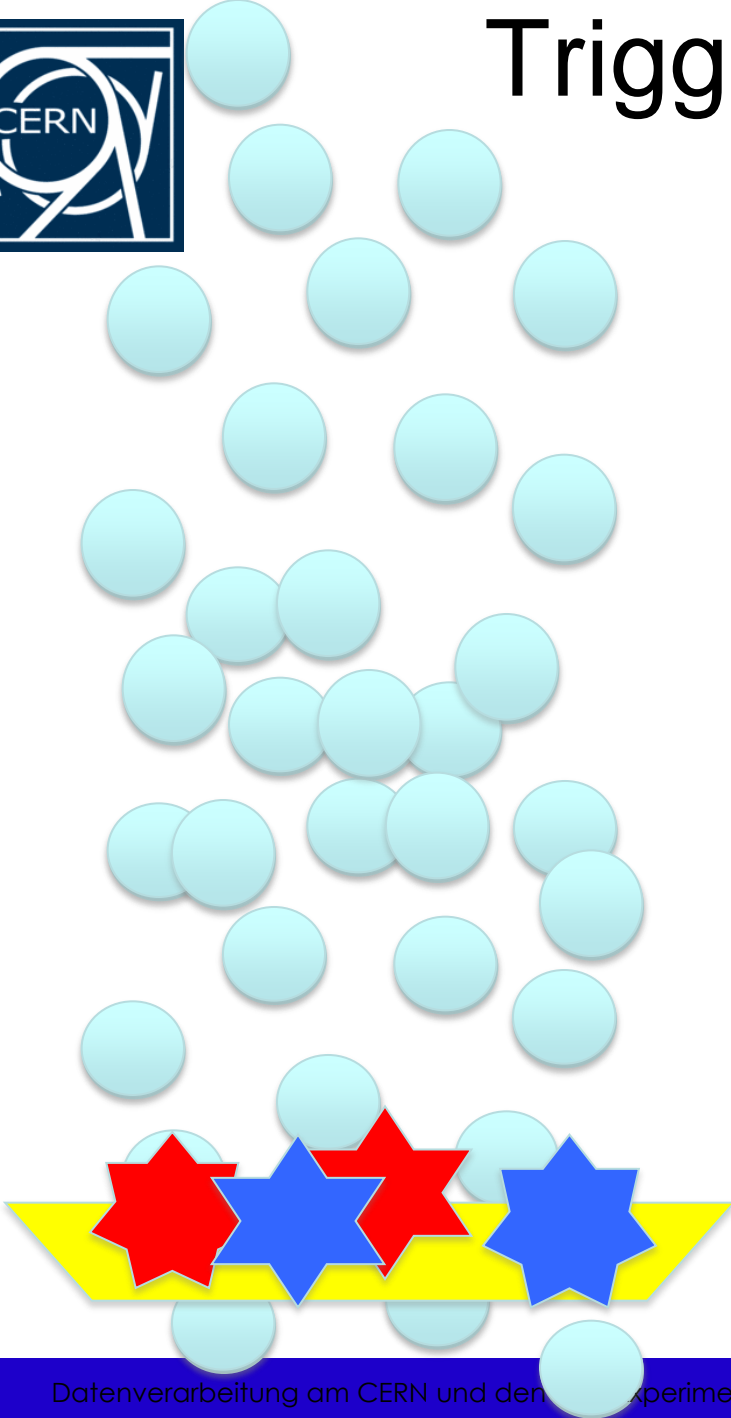
Das Heu: 15 million sensors



**Wie sucht man etwas in 600000 Gigabytes jede Sekunde?
 Das bedeutet eine Suche in einem 100 m Stapel von DVDs**



Triggering – selecting the interesting few

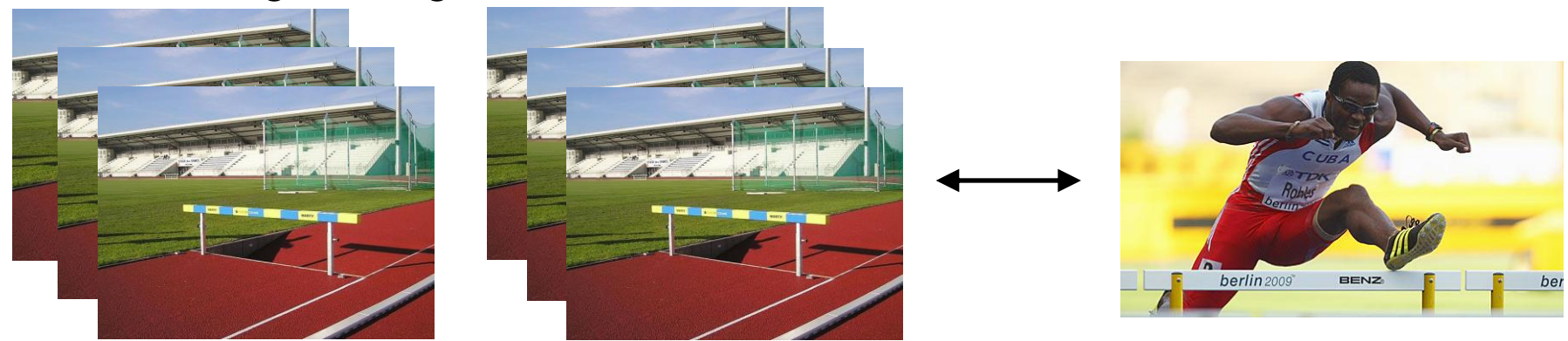


Filter 399 out of 400 collisions
Must keep the good = interesting ones



Trigger

- Trigger = Auslöser
- Das Triggersystem entscheidet, wann eine Kollision (ein Event) aufgezeichnet wird
- Beispiel: Hürdenlauf
 - Zufälliges VS gezieltes Auslösen einer Fotokamera

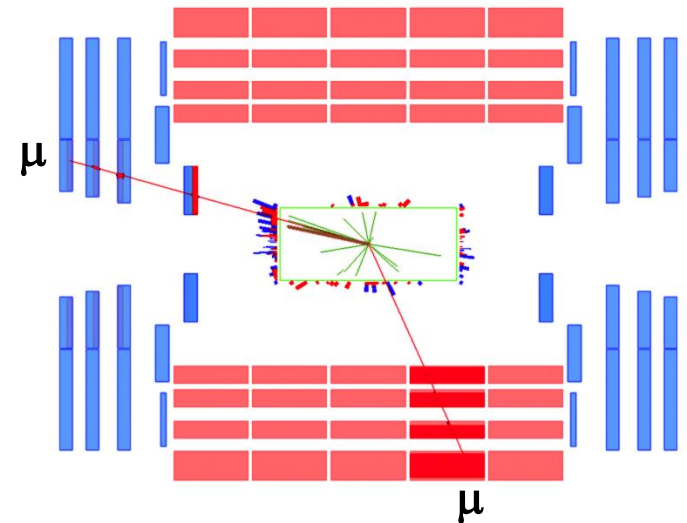


- Ziel: Aufzeichnung von interessanten Kollisionen
 - Aufzeichnung nur zu Zeiten wo Kollisionen im Detektor stattfinden
 - Auswählen interessanter Signaturen



Trigger (2)

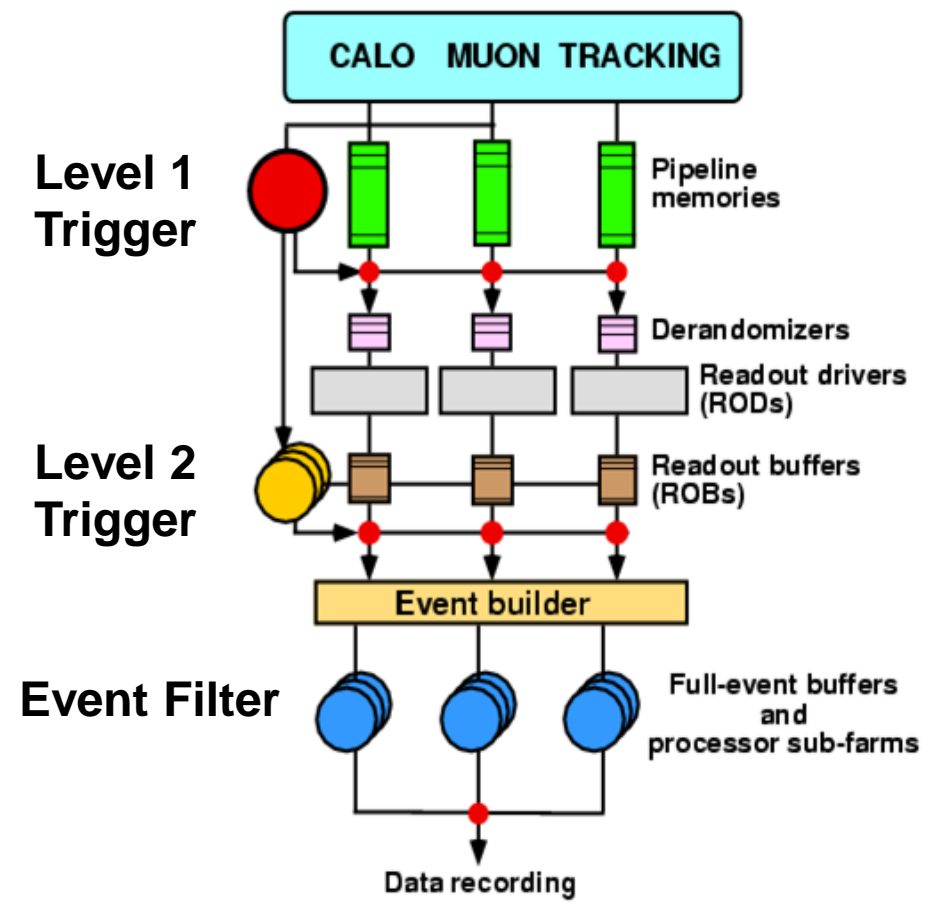
- 1 Milliarde Kollisionen im LHC pro Sekunde
- Nur 100 - 1000 können pro Sekunde aufgezeichnet werden
- Interessante Signaturen müssen während der Datennahme (= "online") erkannt werden
- Beispiel $Z \rightarrow \mu\mu$
 - Suche Kollisionen mit hochenergetischen Muonen
 - Die Detektorspuren von 10^9 Kollisionen müssen innerhalb von 1 Sekunde untersucht werden
- Kombination von Hardware- (dedizierte Hardware) und Softwaretriggern (Rechnerfarmen)





Trigger – Beispiel ATLAS

- Kollisionsrate 40 MHz (jeweils ~ 25 Kollisionen)
- Level 1 Trigger ~ 100 kHz
 - Erste Auswahl
 - Interessante Regionen im Detektor werden erkannt
- Level 2 Trigger ~ 5 kHz
 - Datenanalyse der interessanten Regionen
- Level 3 Trigger (“event filter”) ~ 400 Hz
 - Schnelle Analyse des gesamten Events





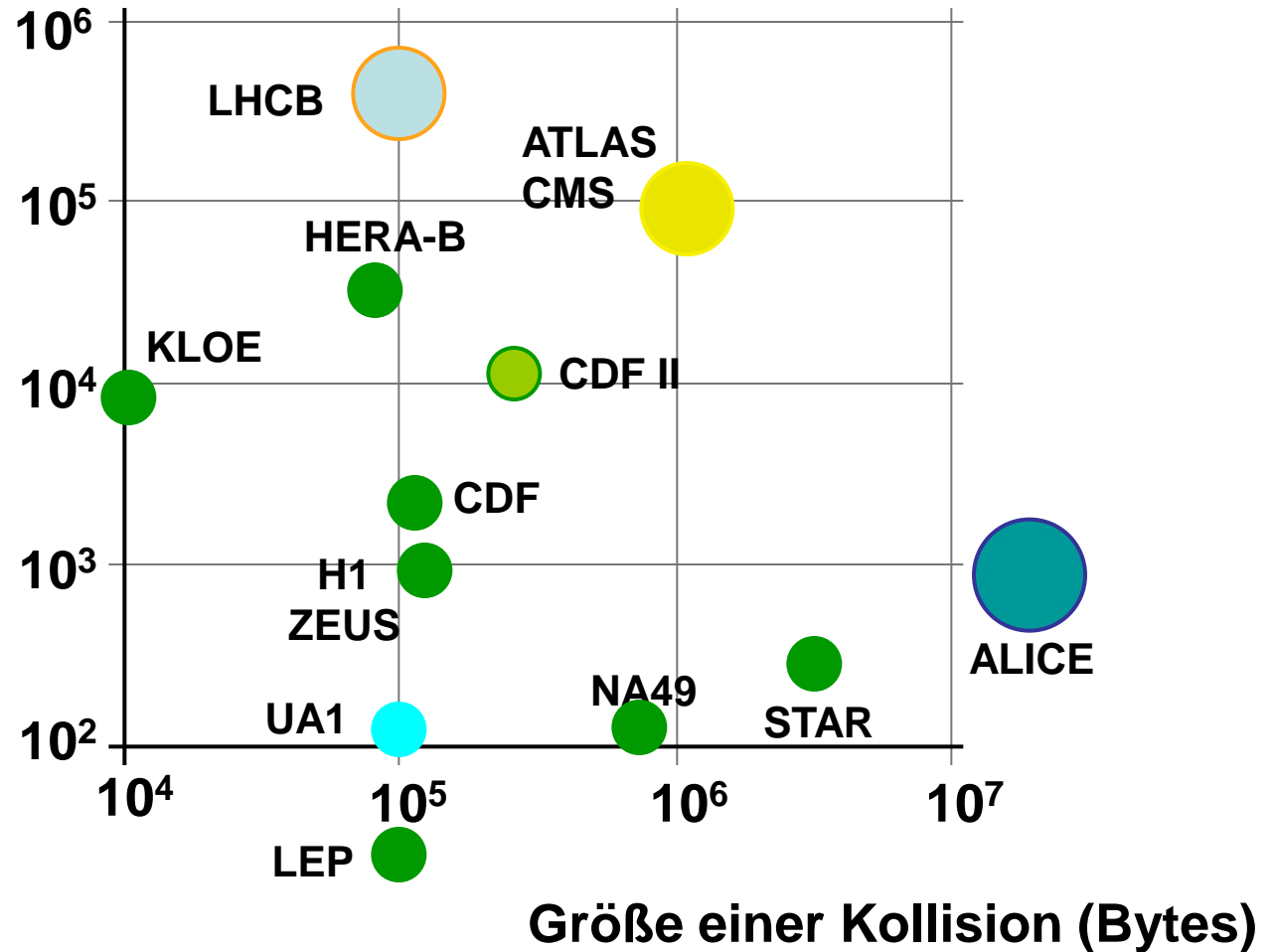
Auslesen der Daten

- Sobald eine Kollision als interessant identifiziert ist, wird sie gespeichert
- Zusammenfassen der Daten von vielen Detektorelementen (“Event building”)
 - Komplexer Schritt, da die Detektorelemente viele Kollisionen gleichzeitig auslesen und nur Teile davon gespeichert werden sollen
 - Größe einer Kollision zwischen 55 kB (LHCb) und 1-2 MB (ALICE, ATLAS, CMS) für pp Kollisionen – bis zu 50 MB für PbPb Kollisionen
- Daten vieler Kollisionen werden in Dateien zusammengefasst und auf Festplatten zwischengespeichert
 - Bis zu 1 GB/s; > 1 CD pro Sekunde
- Daten werden auf Datenbänder exportiert
 - Gesamtdaten pro Jahr 1-2 PB pro Experiment



LHC Datenmengen

Level 1 Rate
(Hz)



Data Handling and Computation for Physics Analysis



Online trigger and filtering

Selection & reconstruction

Offline Reconstruction

Processed Data (Active tapes)



100 %

Raw data

10%

Event reprocessing

Batch physics analysis

1%

Offline Analysis w/ROOT



Analysis objects (extracted by physics topic)



Interactive analysis

Event simulation

Offline Simulation w/GEANT4



Rekonstruktion

Kleiner Teil einer ausgelesenen Kollision:

```

0x01e84ce0: 0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84cf0: 0x01e8 0x87ec 0x01e8 0x85d8 0x7363 0x616e 0x0000 0x0000
0x01e84d00: 0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d10: 0x01e8 0x87e8 0x01e8 0x8618 0x7365 0x7400 0x0000 0x0000
0x01e84d20: 0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d30: 0x01e8 0x87a8 0x01e8 0x8658 0x7370 0x6c69 0x7400 0x0000
0x01e84d40: 0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d50: 0x01e8 0x8854 0x01e8 0x8698 0x7374 0x7269 0x6e67 0x0000
0x01e84d60: 0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d70: 0x01e8 0x875c 0x01e8 0x86d8 0x7375 0x6273 0x7400 0x0000
0x01e84d80: 0x0000 0x0019 0x0000 0x0000 0x0000 0x0000 0x01e8 0x5b7c
0x01e84d90: 0x01e8 0x87c0 0x01e8 0x8718 0x7377 0x6974 0x6368 0x0000

```

Adresse = welcher Detektor hat diesen Wert ausgelesen?

Daten = was wurde ausgelesen?

Rekonstruktion = Umwandlung der elektronischen Signal in "Physikobjekte"

Spur: $\phi = 0.23$, $\eta = 0.75$, $p_T = 2.3 \text{ GeV}/c$

Wahrscheinlichkeit für Pion: 70%, Kaon: 28%, Proton: 2%



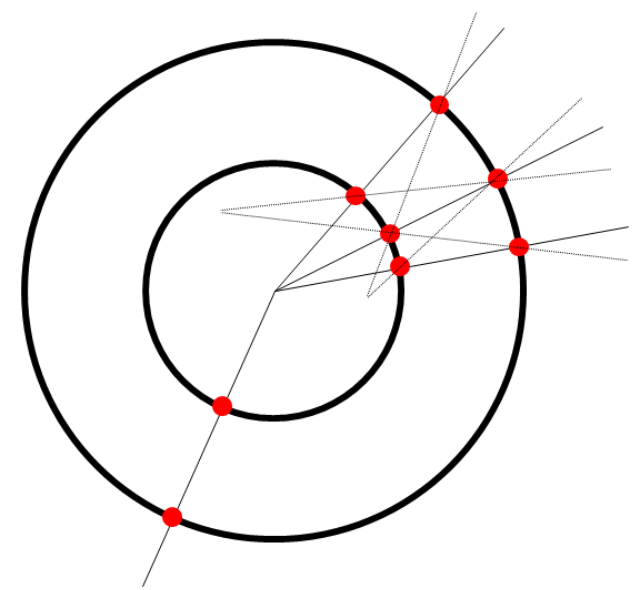
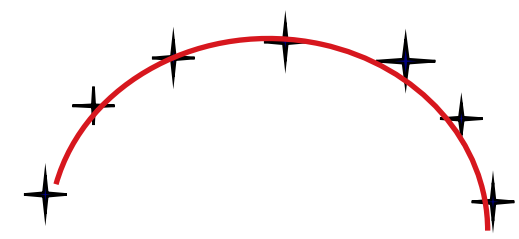
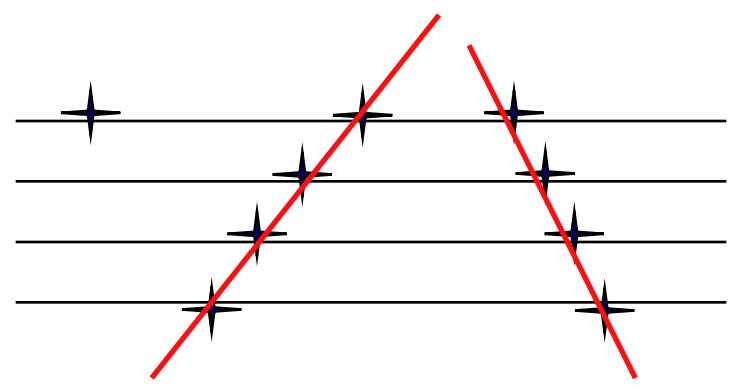
Schritte der Rekonstruktion

- Abstrahierung der elektronischen Signale vom Detektor
 - Detektorelement 1244 hat zum Zeitpunkt 1333096259.344245 ein Signal von 120 gemessen
 - Signal an Position $x = 1.2$ cm, $y = 4.5$ cm, $z = 3.2$ cm, deponierte Energie 100 keV
 - Benötigt Informationen über genaue Position des Detektorelements (Alignment) und die Kalibration
- Teilchen hinterlassen häufig Signale in angrenzenden Detektorelementen
 - Zusammenfassung von Detektorsignalen zu "Clustern"
- Zusammenfassung der Cluster zu Spuren (Mustererkennung)
- Zusammenfassung der Spuren (Kollisionsvertex, Teilchenzerfälle)



Mustererkennung

- Erkennung von Spuren
 - Geradlinig in Magnetfeldrichtung
 - Ablenkung auf Kreisbahn rechtwinklig zum Magnetfeld
→ Erlaubt die Bestimmung des Impulses
- Kollisionsvertex
- Sekundärvertices
 - Erlaubt das Erkennen von Teilchenzerfällen



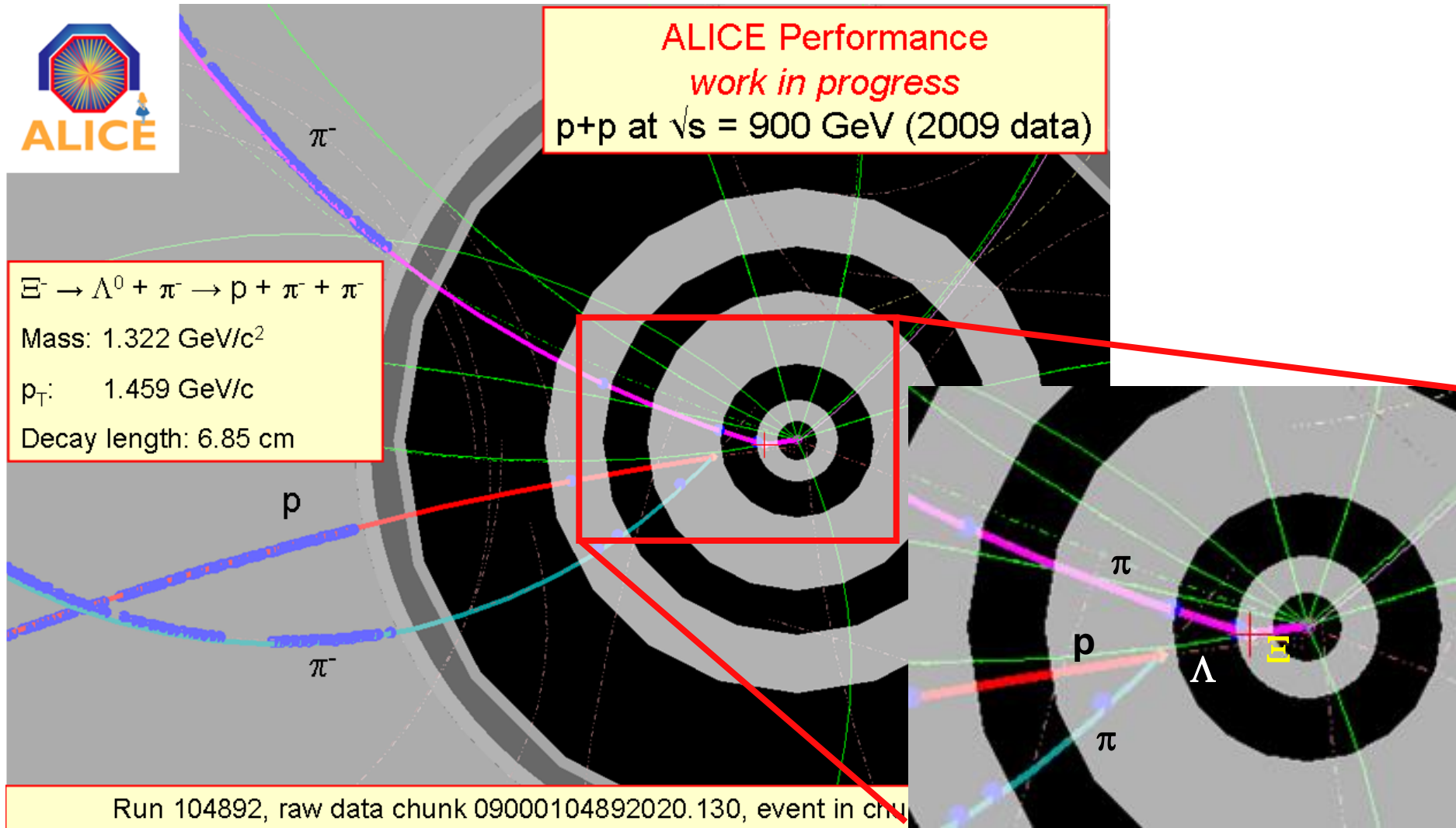


Mustererkennung (2)



ALICE Performance
work in progress
p+p at $\sqrt{s} = 900$ GeV (2009 data)

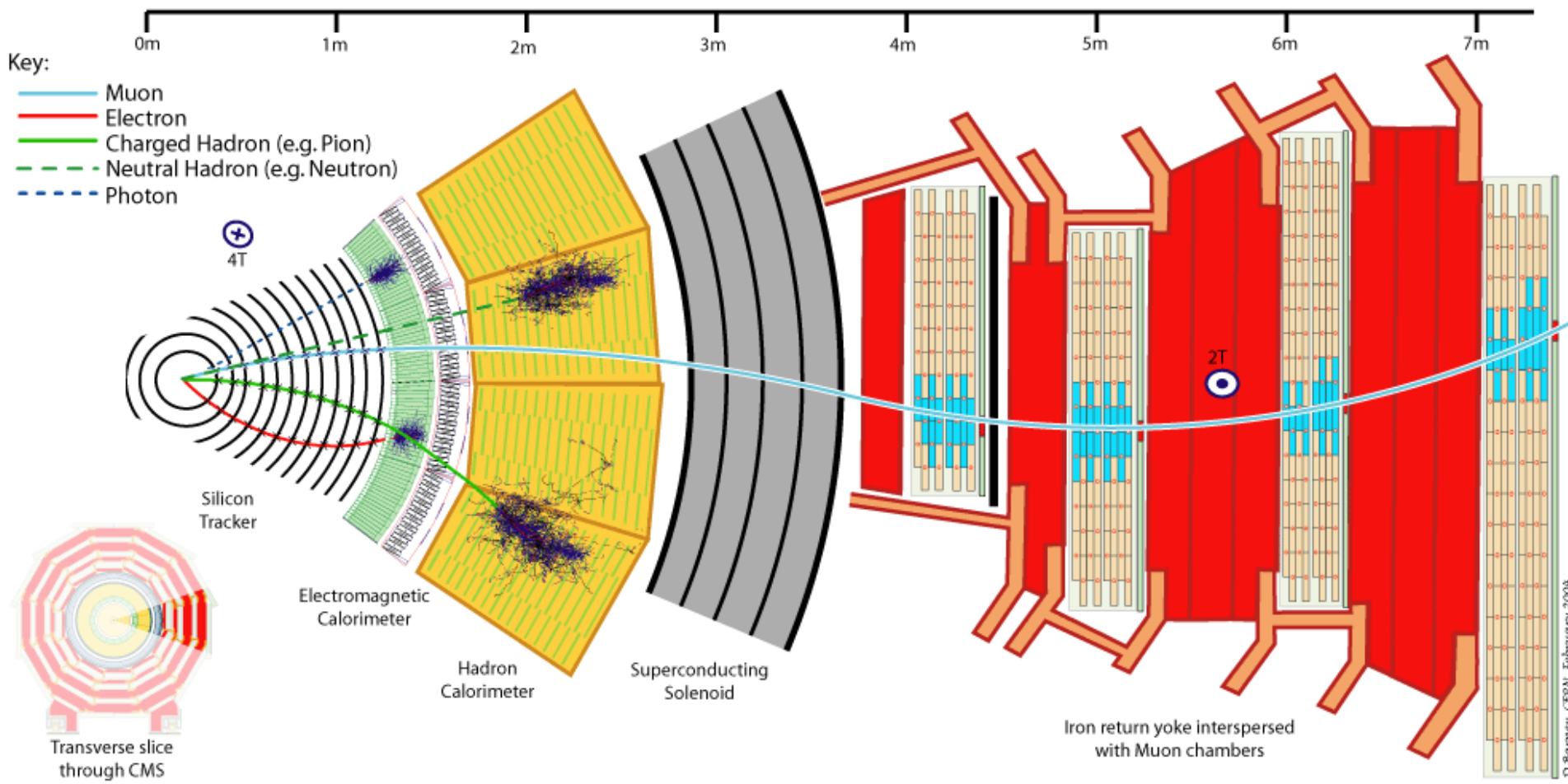
$\Xi^- \rightarrow \Lambda^0 + \pi^- \rightarrow p + \pi^- + \pi^-$
Mass: 1.322 GeV/c²
 p_T : 1.459 GeV/c
Decay length: 6.85 cm



Run 104892, raw data chunk 09000104892020.130, event in ch1



Kombination von Detektorsystemen





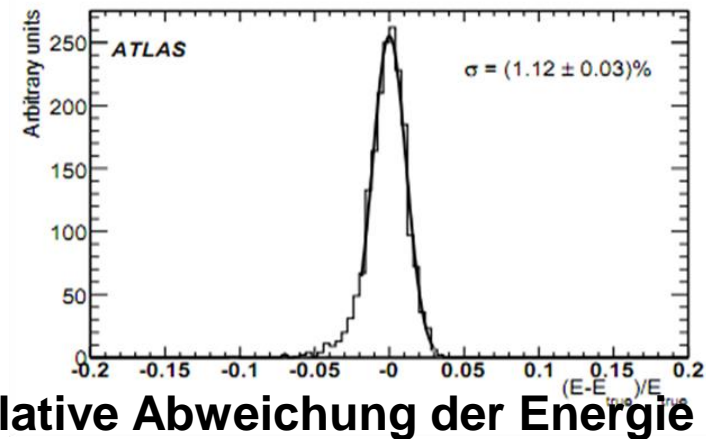
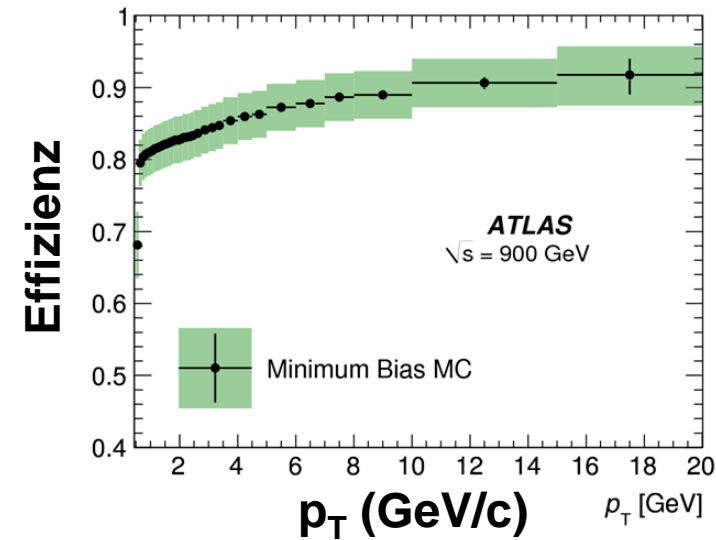
Simulation

- Zum Verständnis des Detektors werden detaillierte Simulationen benötigt
- Eine Detektorsimulation reproduziert das Verhalten des Detektors so gut wie möglich
- Elemente einer Simulationssoftware
 - Kollisionsgenerator: Simuliert welche Teilchen bei einer Kollision im Detektor entstehen. Z.B. Pythia
 - Geometrie: Genaue Beschreibung aller aktiven (= Signale werden gemessen) und inaktiven (= Abstützung, Kühlung, Stromversorgung, ...) Elemente im Detektor
 - Transportsoftware: Simuliert die Wechselwirkung von Teilchen mit dem Detektormaterial. Z.B. GEANT4
 - Teilchen können zerfallen, abgelenkt werden, stecken bleiben
 - Energieabgabe an das aktive Detektormaterial wird aufgezeichnet
- Anschließend werden die simulierten Signale von der normalen Rekonstruktionssoftware rekonstruiert



Simulation (2)

- Ergebnisse von Simulationen erlauben die Effekte des Detektors auf Messgrößen abzuschätzen
- Wichtige Beispiele
 - Spurrekonstruktionseffizienz – wie wahrscheinlich ist es ein Teilchen zu messen?
 - Auflösung – wie genau wird eine Größe gemessen?
 - Fehlerrate ("fake rate") – wie oft kommt es vor, dass ein Teilchen gefunden wird wo keins war?





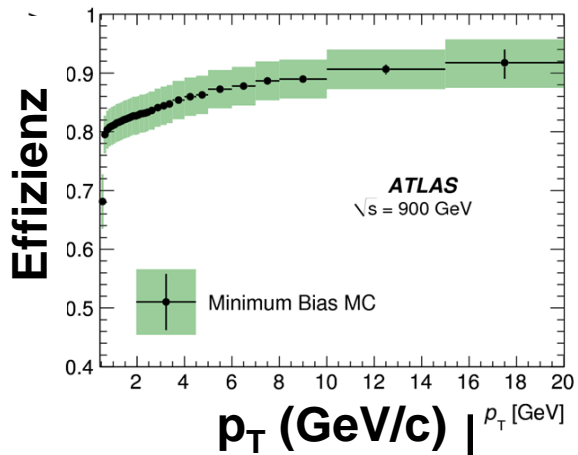
Analyse

- Interpretation der Physikobjekte (Teilchenspuren, Vertices, Kalorimeterinformationen) die während der Rekonstruktion gefunden wurden
 - Verarbeitung von vielen Kollisionen um ein statistisch signifikantes Ergebnis zu erhalten
 - Häufigste Aufgabe des/der PhysikerIn
- Arbeitsablauf
 - Selektion von geeigneten Kollisionen
 - Zusammenfassung der benötigten Größen aus den Kollisionen (oft in so genannten "Histogrammen")
 - Korrigieren von Detektoreffekten
 - Extraktion von Physikgrößen
 - Abschätzung der Messunsicherheiten
 - Vergleich mit theoretischen Vorhersagen und Simulationen
- Während Rekonstruktion hauptsächlich eine "zentrale" Aufgabe ist (Programmierung und Ausführung in wenigen Gruppen), wird Analyse von fast allen Gruppen und Einzelpersonen in einem Experiment durchgeführt



Analyse – Beispiel Impulsverteilung

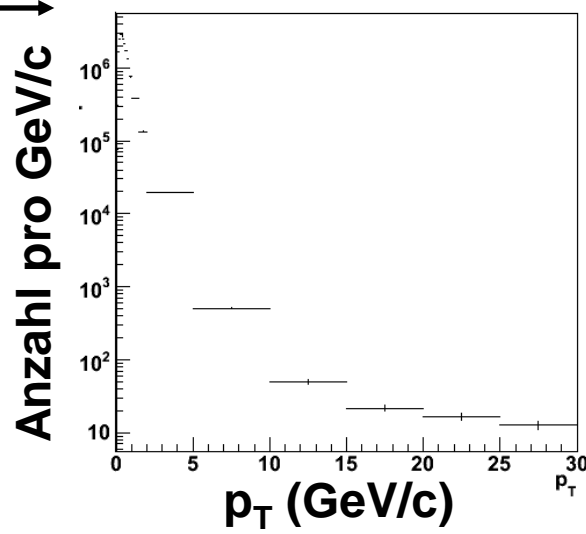
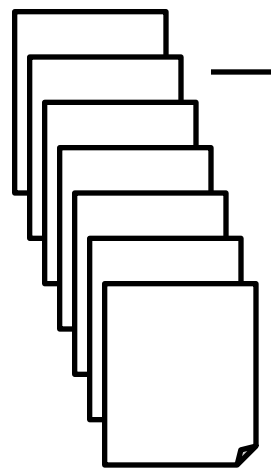
Frage: Wie viele Teilchen haben wie oft welchen Impuls?



Rekonstruktions-effizienz (aus Simulationen)

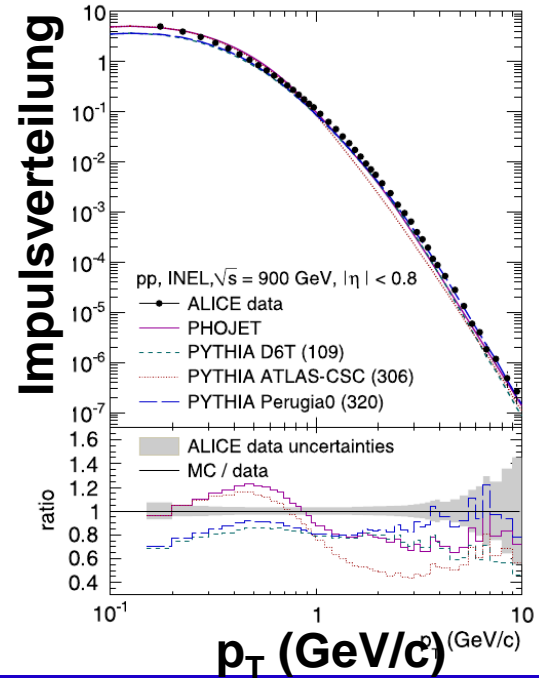
Rekonstruierte Kollisionen

Extraktion der Messgröße in ein "Histogramm"



Korrektur Detektor-effekte

Ergebnis + Modellvergleich



Anmerkung: Dieses Schema ist stark vereinfacht

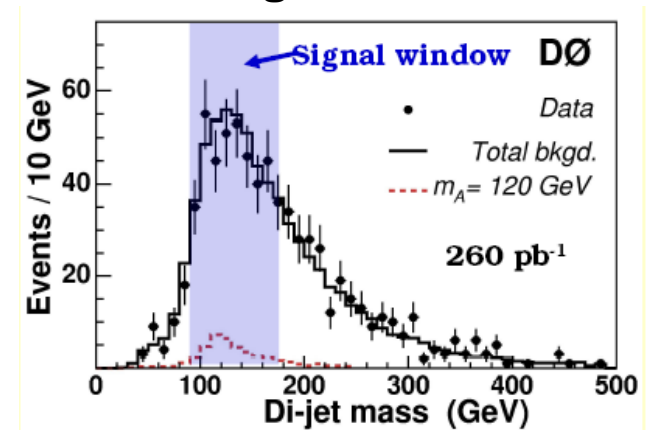


Das ROOT Framework

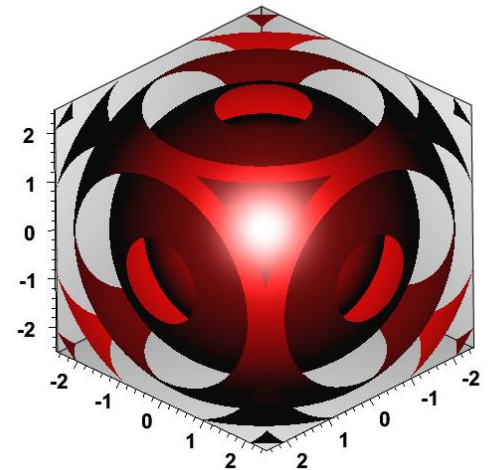
- ROOT ist ein umfangreiches, objektorientiertes Framework für Datenverarbeitung und Analyse
- Wird seit 1995 am CERN entwickelt, mittlerweile 10 Entwickler
- Basiert auf C++
- Umfangreiche 2D und 3D Tools für die Visualisierung
- Mathematik- und Statistikbibliotheken
- Kann programmatisch (Skripte, Programme) und via graphischer Benutzeroberfläche (GUI) verwendet werden
- Verfügbar für Linux, Mac, Windows

Information & Download
→ <http://root.cern.ch>

Histogramm:



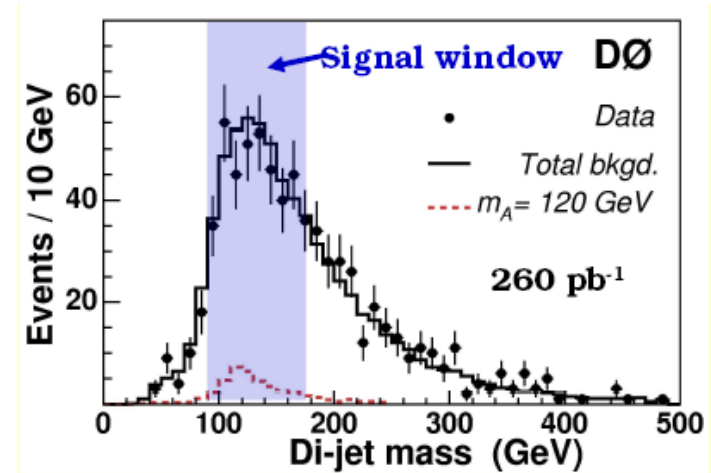
3D Funktion:





Beispiel: ROOT Histogram

- Histogramme speichern Daten in Wertebereichen ("Bins")
 - Zuteilung eines Messwertes zu einem Bin mit gewisser Breite
 - Erlaubt Visualisierung & statistische Analyse

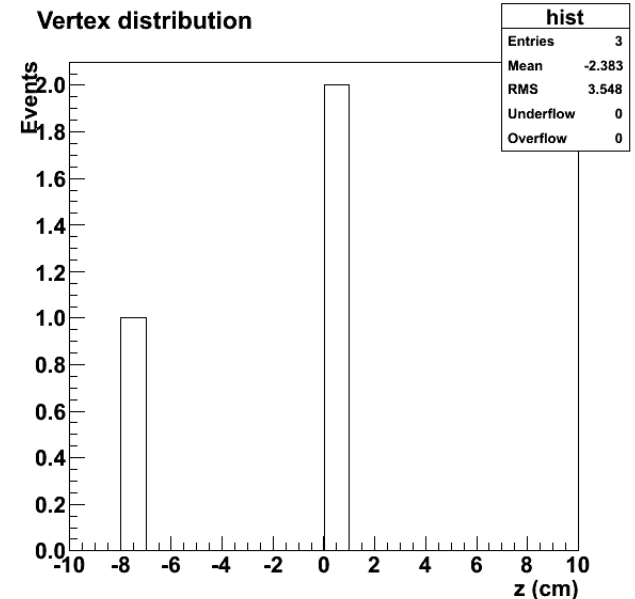


Histogrammbeispiel (C++ code)

```

hist = new TH1F("hist", "Vertex distribution
                ;z (cm);Events", 20, -10, 10);
hist->Fill(0.05);
hist->Fill(-7.4);
hist->Fill(0.2);
hist->Draw();

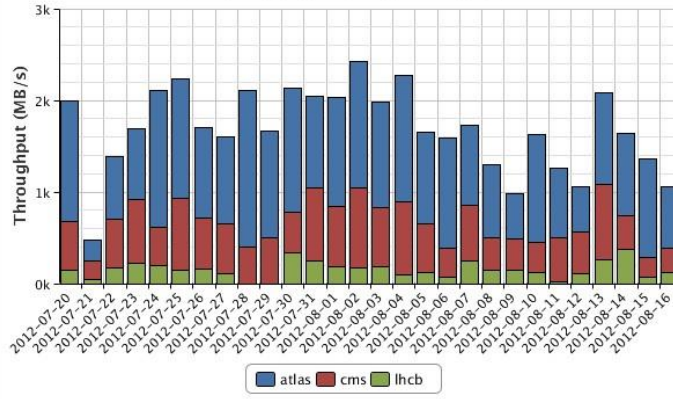
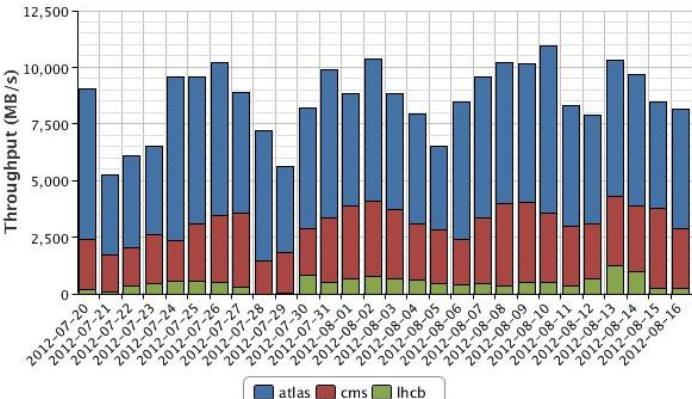
```





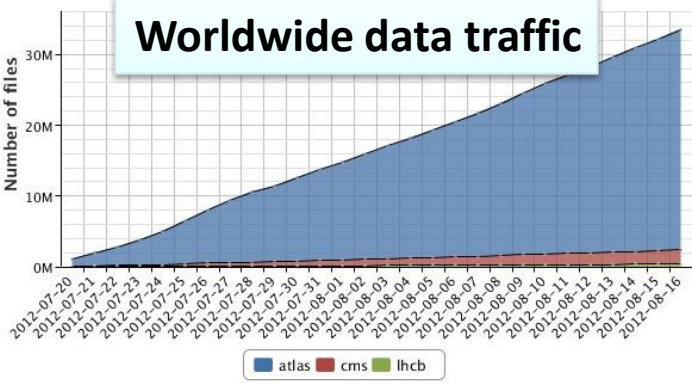
Datenmengen

- Alle vorgestellten Tätigkeiten benötigen eine große Menge Rechenleistung und Speicherplatz
- Die Rekonstruktion eines pp Events dauert 5-15 Sekunden (wird einige Male wiederholt)
- Die Simulation eines pp Events dauert 5-30 Minuten
- Die Analyse eines pp Events dauert < 1 Sekunde (wird hunderte Male wiederholt)
- In 2011 hat z.B. ATLAS
 - 2 Milliarden Events aufgezeichnet
 - 2 Milliarden Events simuliert
- Dies würde auf einem Computer $\sim 40\,000$ Jahre dauern ...oder auf 40 000 Computern 1 Jahr

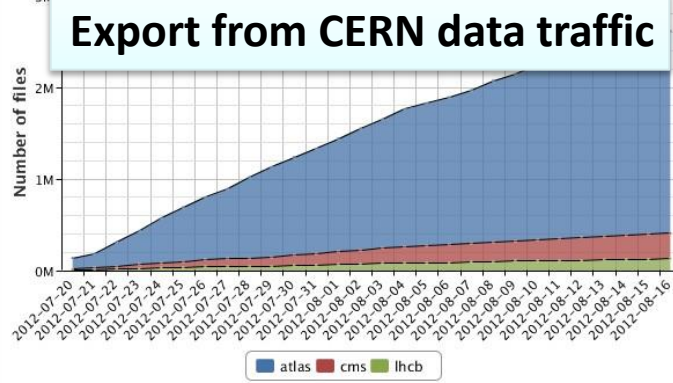


Data Transfers
A key technology for LHC

Worldwide data traffic



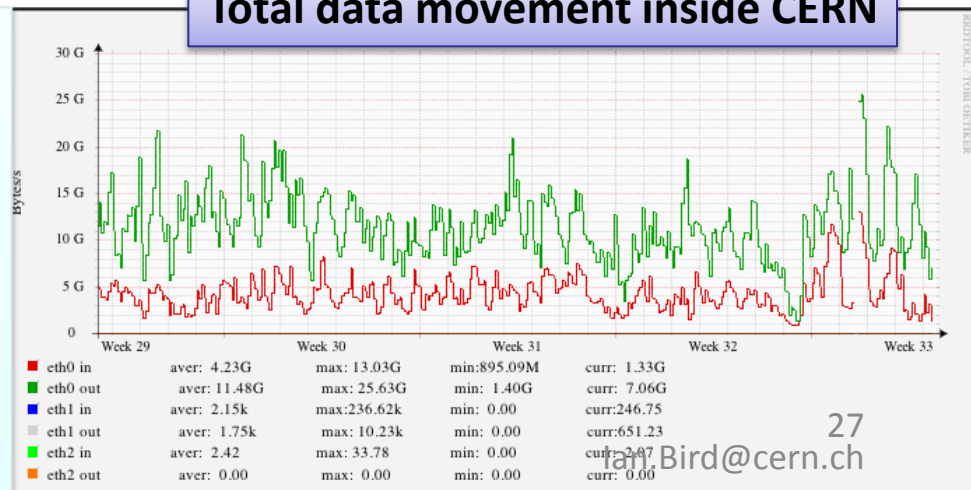
Export from CERN data traffic



Last month:
4 PB, 2.6M files exported from CERN to Tier 1s
20 PB, 33M files moved worldwide

- Data movement inside Tier 0:**
- Data in: 4.3 GB/s (peak 13 GB/s)
 - Data out: 11.5 GB/s (peak 25 GB/s)
 - ~1.5 PB/day data movement at CERN

Total data movement inside CERN



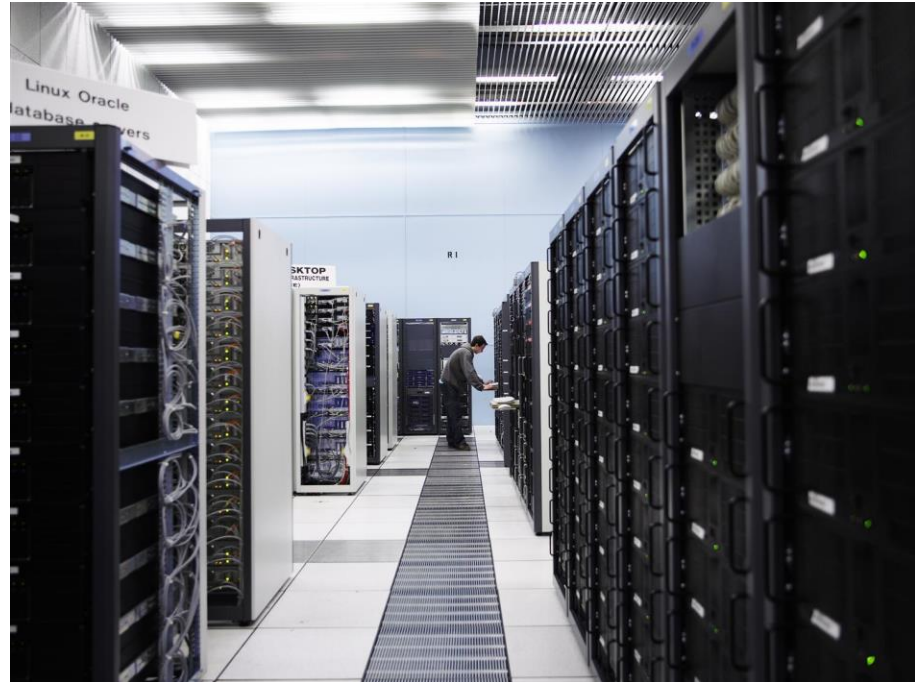


Datenmengen (2)

- Verwenden vieler Computer notwendig
 - Das Verteilen der Daten auf viele laufende Programme muss organisiert werden ("Jobs")
- So genannte Batchsysteme verteilen Jobs auf die Rechner eines Rechenzentrums
- Für die Verarbeitung der Daten der LHC-Experimente werden viele Rechenzentren verwendet
 - Anforderungen an CPU und Speicherplatz immens
 - Viele beteiligte Länder → Ressourcen sind vor Ort vorhanden
 - Diese Rechenzentren sind über das Internet verbunden
- Software wird benötigt, die Job und Datenfluss zwischen den beteiligten Zentren verwaltet
 - "Grid middleware" (zwischen Rechenzentrum und BenutzerIn)



CERN Rechenzentrum



15 000 CPUs
65 000 CPU Kerne
~60 PB Festplattenkapazität
~45 PB auf Datenbändern
3.5 MW Stromverbrauch



The CERN Data Centre in Numbers

- Data Centre Operations (Tier 0)
 - 24x7 operator support and System Administration services to support 24x7 operation of all IT services.
 - Hardware installation & retirement
 - ~7,000 hardware movements/year; ~1800 disk failures/year
 - Management and Automation framework for large scale Linux clusters

Racks	828
Servers	8,750
Processors	15,694
Cores	68,443
HEPSpec06	482,507

Disks	67,320
Raw disk capacity (TiB)	97,096
Memory modules	52,377
Memory capacity (TiB)	205
RAID controllers	3,749

Tape Drives	160
Tape Cartridges	45000
Tape slots	56000
Data on Tape (PiB)	65

High Speed Routers (640 Mbps → 10 Tbps)	25
---	----

Ethernet Switches	500
10 Gbps/100Gbps ports	1200/50

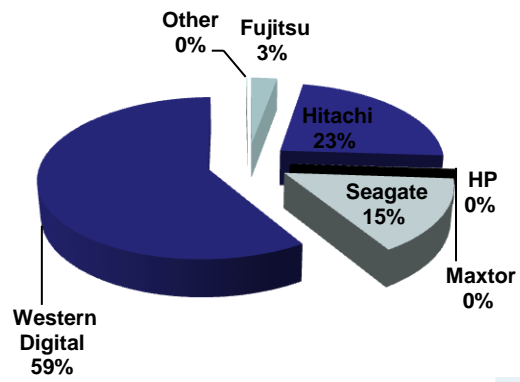
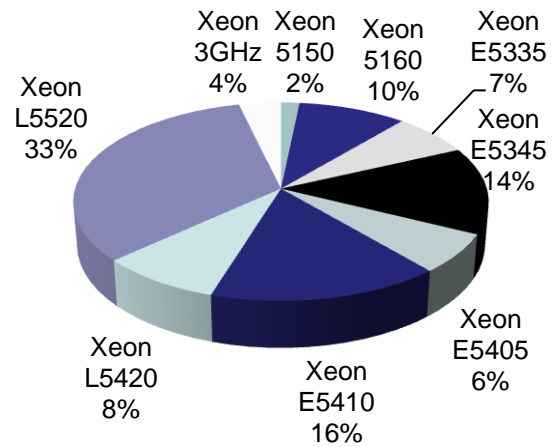
Switching Capacity	6 Tbps
--------------------	--------

1 Gbps ports	16,041
--------------	--------

10 Gbps ports	1,884
---------------	-------

IT Power Consumption	2456 KW
----------------------	---------

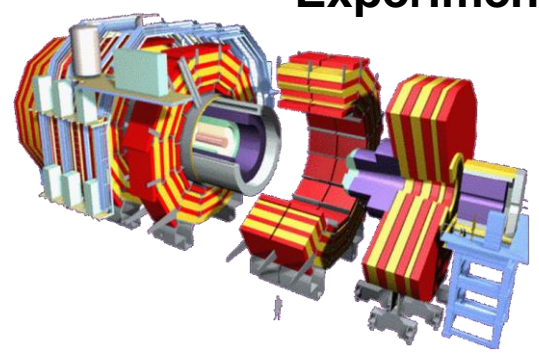
Total Power Consumption	3890 KW
-------------------------	---------



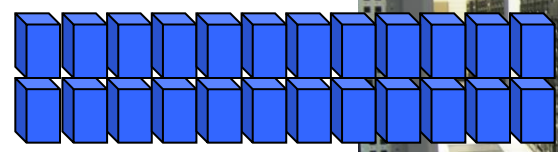


Datenfluss im CERN Rechenzentrum

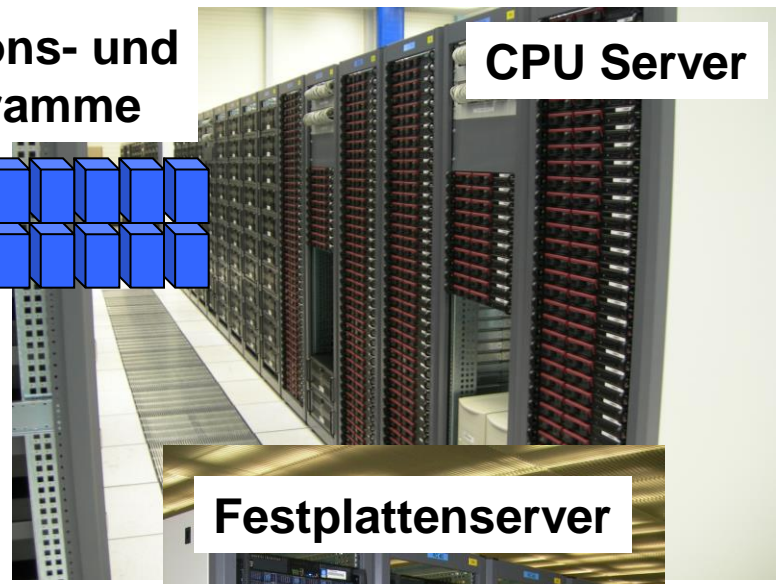
Experiment



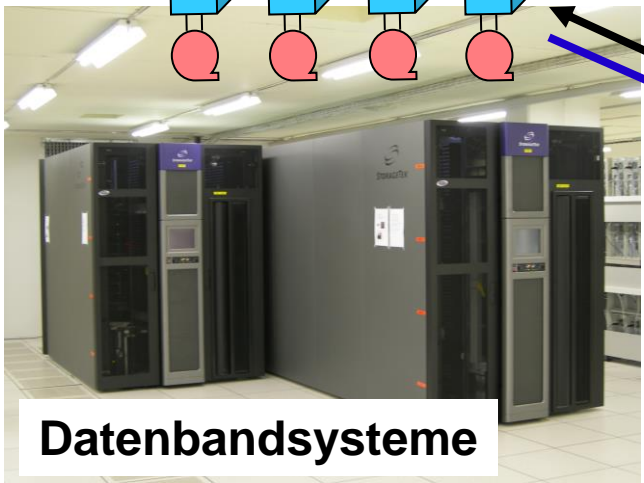
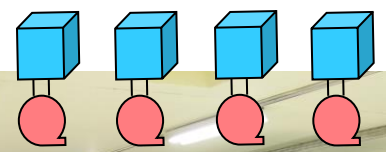
Rekonstruktions- und Analyseprogramme



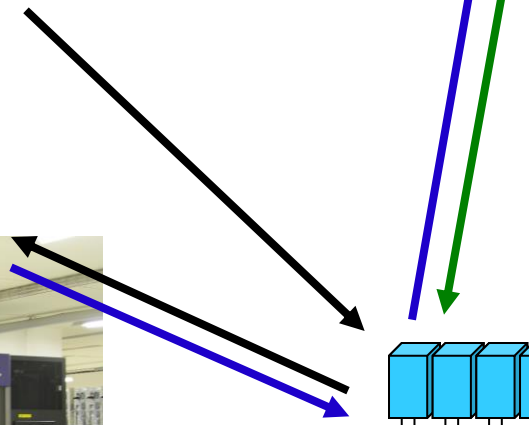
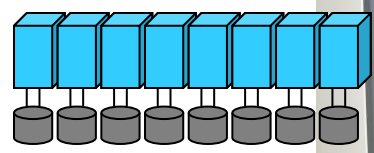
CPU Server



Festplattenserver

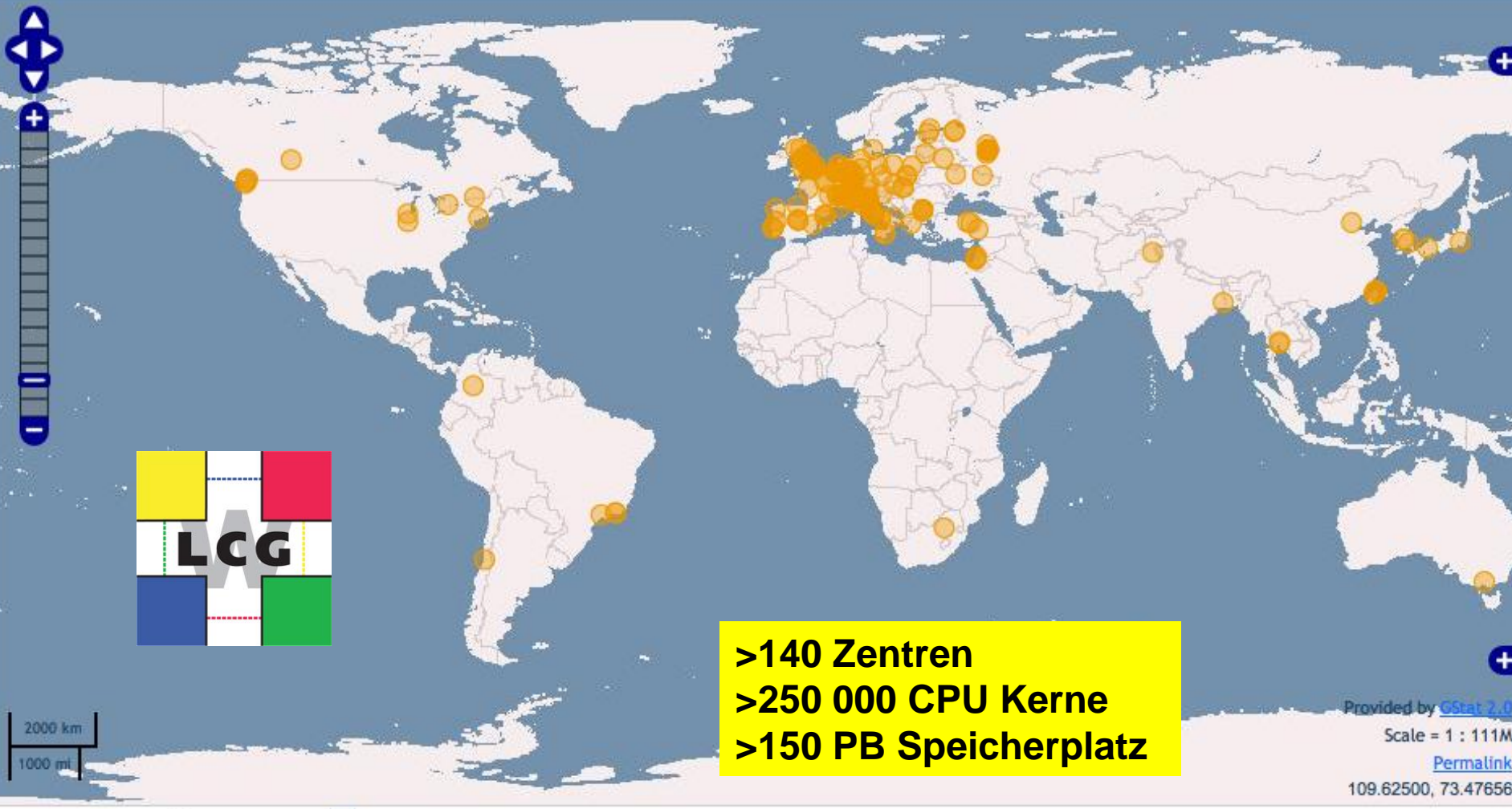


Datenbandsysteme





WLCG (Worldwide LHC Computing Grid) Kollaboration





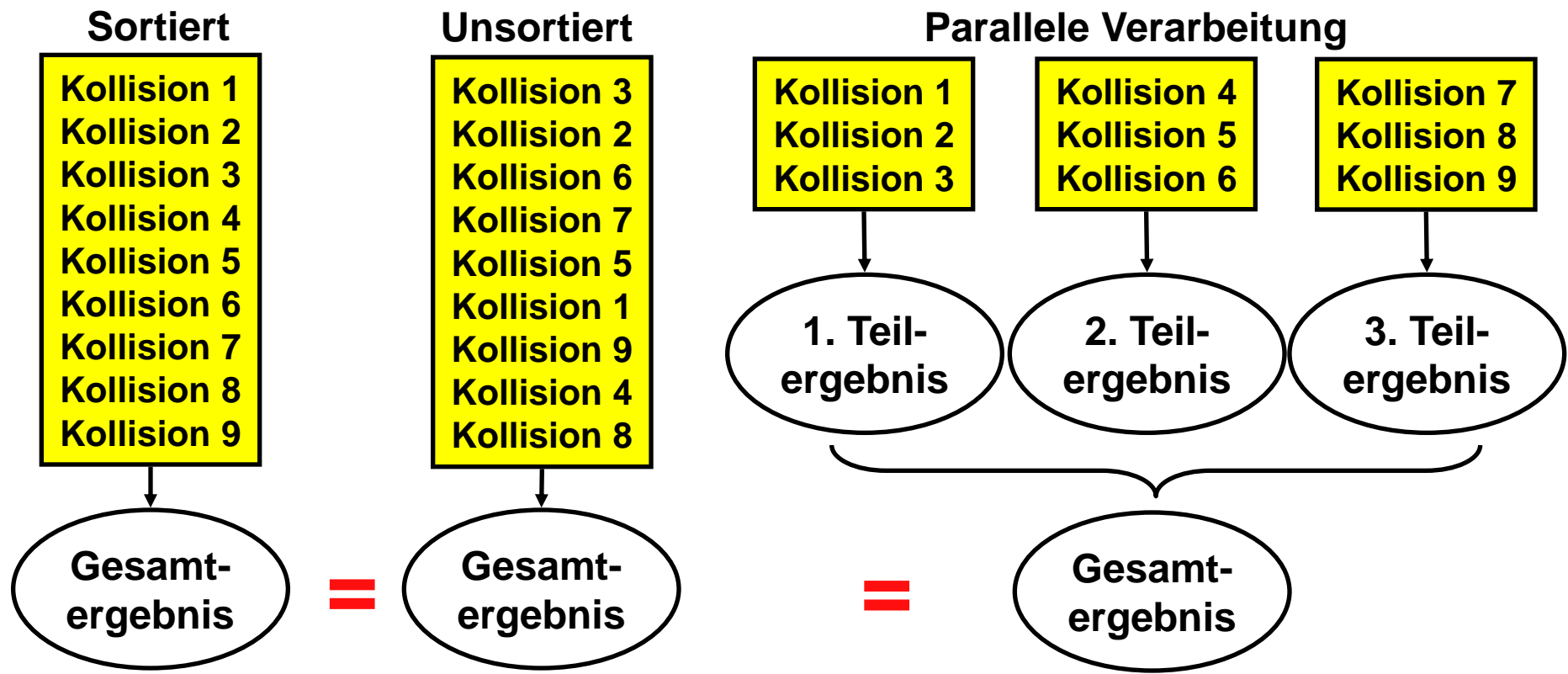
Analyse im Grid

- Der/die BenutzerIn möchte gewisse Daten analysieren (Liste von Dateien)
- Die Gridsoftware stellt fest, in welchen Rechenzentren diese Daten vorliegen und startet die Analysejobs in diesen Zentren ("data locality")
- Statt der Daten selbst (große Datenmenge), muss nur das Analyseprogramm und das Analyseergebnis übers Netzwerk verschickt werden (kleine Datenmenge)
- Die Analyse kann mehrere Zentren gleichzeitig nutzen
- Die Grid middleware ermöglicht die Analyse durch eine Reihe von Programmen (so genannte "Services")
 - In der Praxis sind das Verteilen der Jobs und das Zusammenfügen der einzelnen Analyseergebnisse sehr komplexe Aufgaben



Parallelisierung

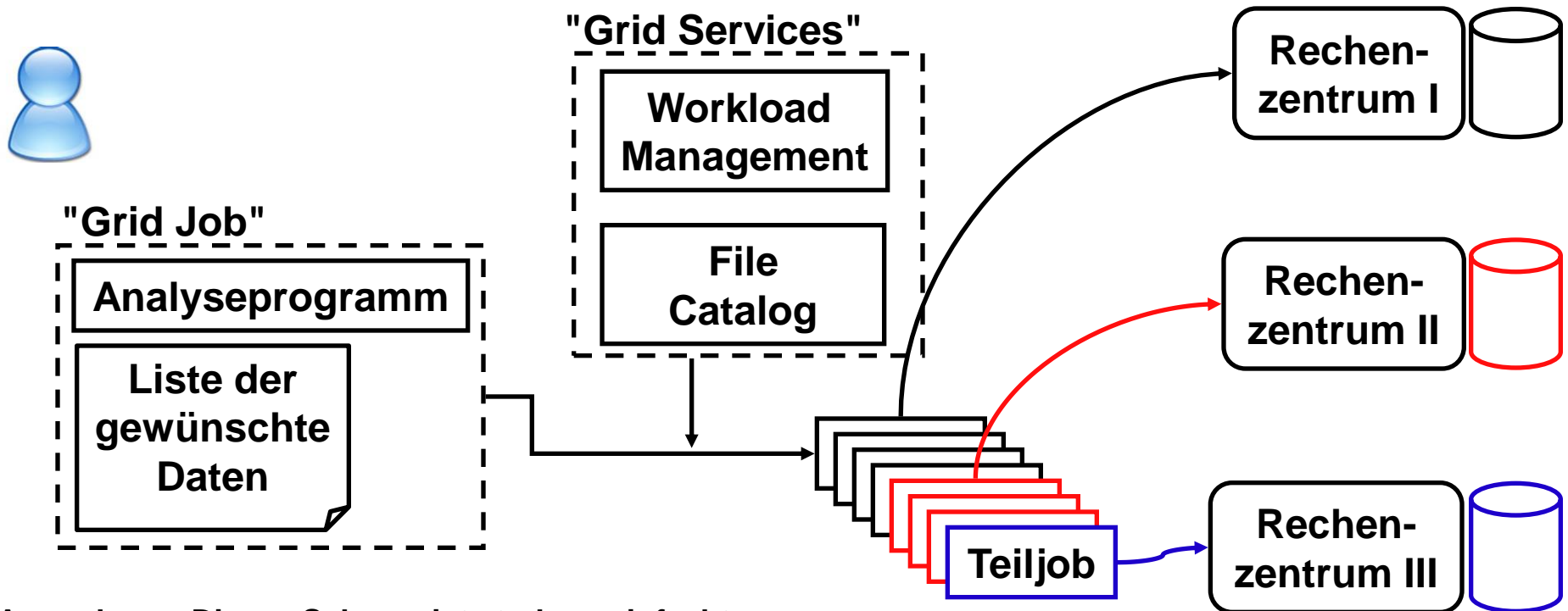
- Durch die Unabhängigkeit der einzelnen Kollisionen sind die Reihenfolge und die Aufteilung der Daten für die Analyse für das Gesamtergebnis unerheblich





Ablauf Analyse im Grid

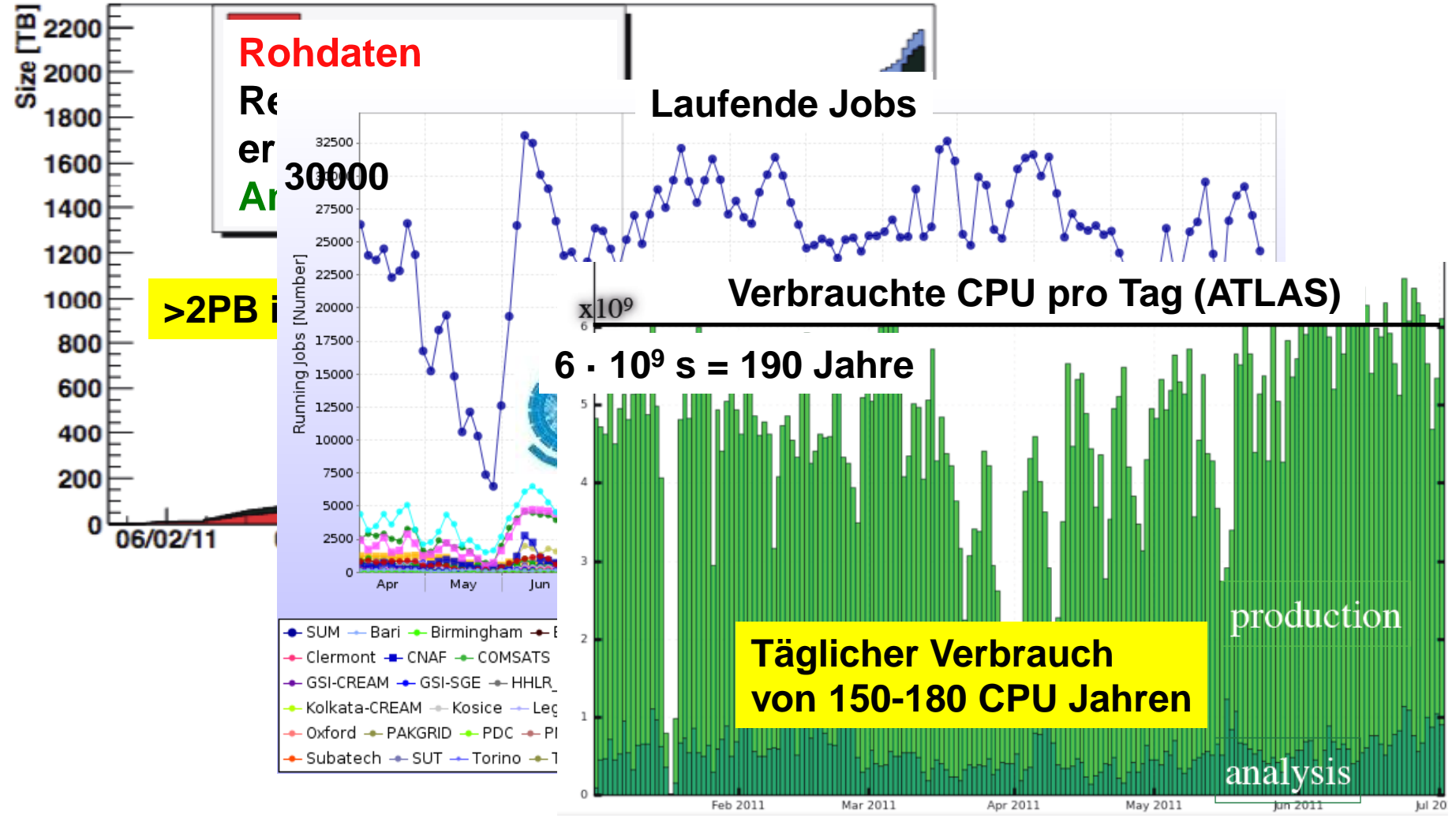
- Der/die BenutzerIn schickt einen Grid Job
- Unter Verwendung der Information wo die gewünschten Daten vorliegen, wird der Job in viele Teiljobs aufgeteilt
- Diese Teiljobs werden an den entsprechenden Rechenzentren ausgeführt
- Anschließend werden die einzelnen Analyseergebnisse zusammengefasst



Anmerkung: Dieses Schema ist stark vereinfacht



Grid Statistiken



Quelle: I. Fisk, WLCG Workshop DESY 2011



LHC@Home



- Nur ein Bruchteil der verfügbaren Rechenzeit eines Computers zu Hause oder am Arbeitsplatz wird normalerweise verwendet
- LHC@Home ist ein System, das jeder Person auf der Erde ermöglicht, die sowieso nicht verwendete Rechenzeit für LHC und CERN-Projekte zur Verfügung zu stellen
- Der eigene Computer löst Berechnungen für das LHC Upgrade. Diese Ergebnisse bestimmen, wie der LHC erweitert wird
- Bereits 100 000 angemeldete Benutzer
- Informationen: <http://lhathome.web.cern.ch/LHCathome>



Zusammenfassung

- Die heutige Teilchenphysik benötigt immense Rechen- und Speicherkapazitäten
 - Datennahme, Rekonstruktion, Analyse
- Das Grid erlaubt hunderte von Rechenzentren weltweit zu nutzen
 - Zur Zeit können über 250 000 Programme gleichzeitig ausgeführt werden
- Am CERN sind Physik und Informatik eng verknüpft
 - Gutes Computergrundwissen ist essentiell

Dank an Jamie Boyd, Ian Fisk und Andreas Hirstius von denen einige Ideen, Abbildung und Folien uebernommen wurden