Summary of Track 3: Data Analysis and Visualisation

Lucas Taylor
Northeastern University, Boston

<u>*Outline:*</u>

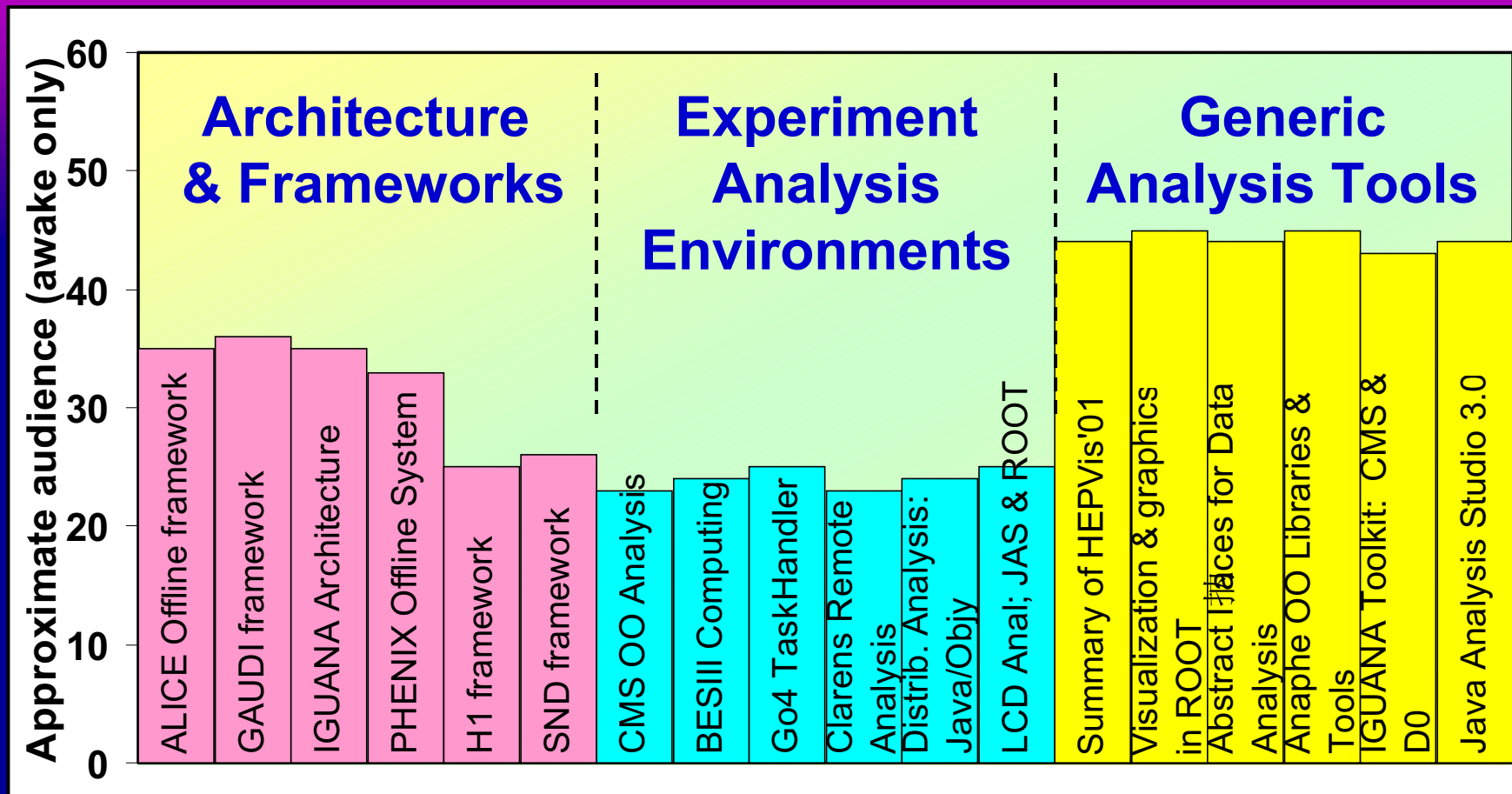*Topics and program*
Session 1: Architecture and Frameworks
Session 2: Experiments' Analysis Environments
Session 3: Generic Analysis Tools
Impressions, trends, and platitudes...

Topics and Program

Bar chart — "Approximate audience (awake only)" (y-axis, 0 to 60) across three categories:

Architecture & Frameworks (pink bars):
- ALICE Offline framework
- GAUDI framework
- IGUANA Architecture
- PHENIX Offline System
- H1 framework
- SND framework

Experiment Analysis Environments (cyan bars):
- CMS OO Analysis
- BESIII Computing
- Go4 TaskHandler
- Clarens Remote Analysis
- Distrib. Analysis: Java/Objy
- LCD Anal; JAS & ROOT

Generic Analysis Tools (yellow bars):
- Summary of HEPVis'01
- Visualization & graphics in ROOT
- Abstract Ifaces for Data Analysis
- Anaphe OO Libraries & Tools
- IGUANA Toolkit: CMS & D0
- Java Analysis Studio 3.0

- ■ Spawned reconstruction session ⇒Track 5
- ■ Extensive Poster Session (~ 20 papers)
- ■ Cancelled: 3-047, 3-049, 3-051, 3-074

Architecture and Frameworks

(Session 1, Monday PM)

Outline:

Topics and program
Session 1: Architecture and Frameworks
Session 2: Experiments' Analysis Environments
*Session 3: Generic Analysis Tools*
Impressions, trends, and platitudes...

The ALICE Offline framework, status and perspectives
(Federico Carminati)

- **ALICE heavy ions at LHC**
  - Will need to write data at 1.25GB/s to tape

- **AliROOT: C++ framework based on ROOT**
  - Virtual MC interface: GEANT3, GEANT4, Fluka
  - Need a common geometrical modeler
  - Whiteboard & "Alien" distributed file catalog (SQL,SOAP)

- **Testing approach with data challenges**
  - DATE+ROOT+CASTOR : 120 MB/s, <85> MB/s
  - 110 TB written into CASTOR tape system

- **DataGrid and ROOT: working on PROOF**

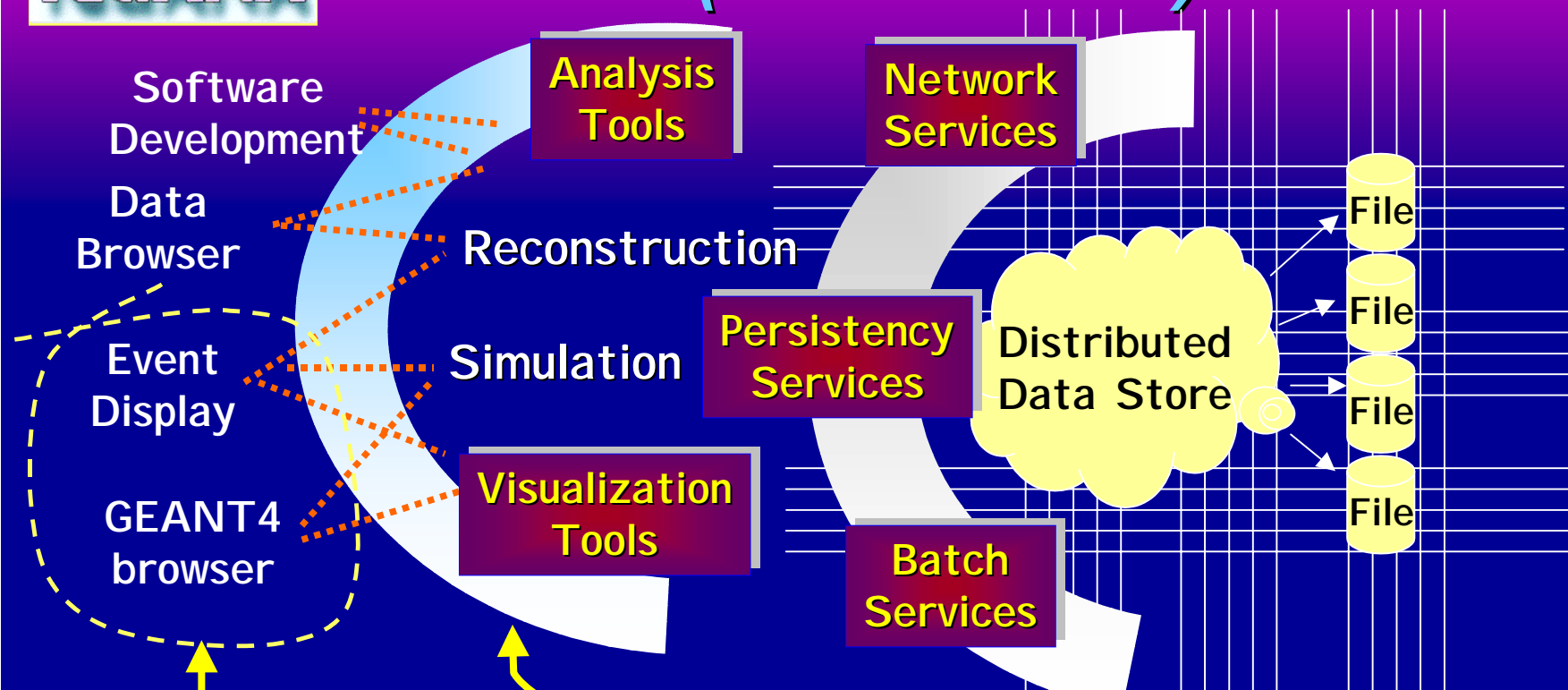See also: [3-070,3-075,4-028,7-013,8-051,10-021]
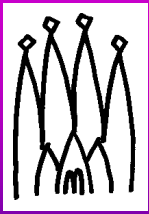
Status of the GAUDI event-processing framework (Pere Mato)

- **Initially LHCb, now also ATLAS ("Athena")**
  - ❖ Also used by GLAST and HARP

- **Architecture-centric; emphasizes interfaces**
  - ❖ Separates: "Data-handling" and "algorithms"
  - ❖ Separates: "Transient" and "persistent" representations

- **Many new and improved services**
  - ❖ Resource monitoring; dynamic-loading, Histo / Ntuple persistency, scripting (with Python) [3-065]
  - ❖ Object Definition Language for automatic code generation, introspection, … XML (LHCb), IDL (ATLAS) [4-051]
  - ❖ Integration with GEANT4 [5-009]

- **Plan: "Grid-capable" not "Grid-dependent"**

IGUANA Open Architecture
(Lassi Tuura)

Software Development

Data Browser

Event Display

GEANT4 browser

IGUANA components [3-040]

Analysis Tools

Network Services

Reconstruction

Simulation

Persistency Services

Visualization Tools

Batch Services

Distributed Data Store

File

File

File

File

Currently implemented IGUANA Framework and Services
- ❖ Thin **portability layer**
- ❖ Tiny **kernel**
- ❖ Variety of **plug-ins**
  - ➢ Application personalities (drivers)
  - ➢ Session and application extensions
  - ➢ Browsers and sites
  - ➢ Data models and representations

IGUANA example: Interactive GEANT4 Browser
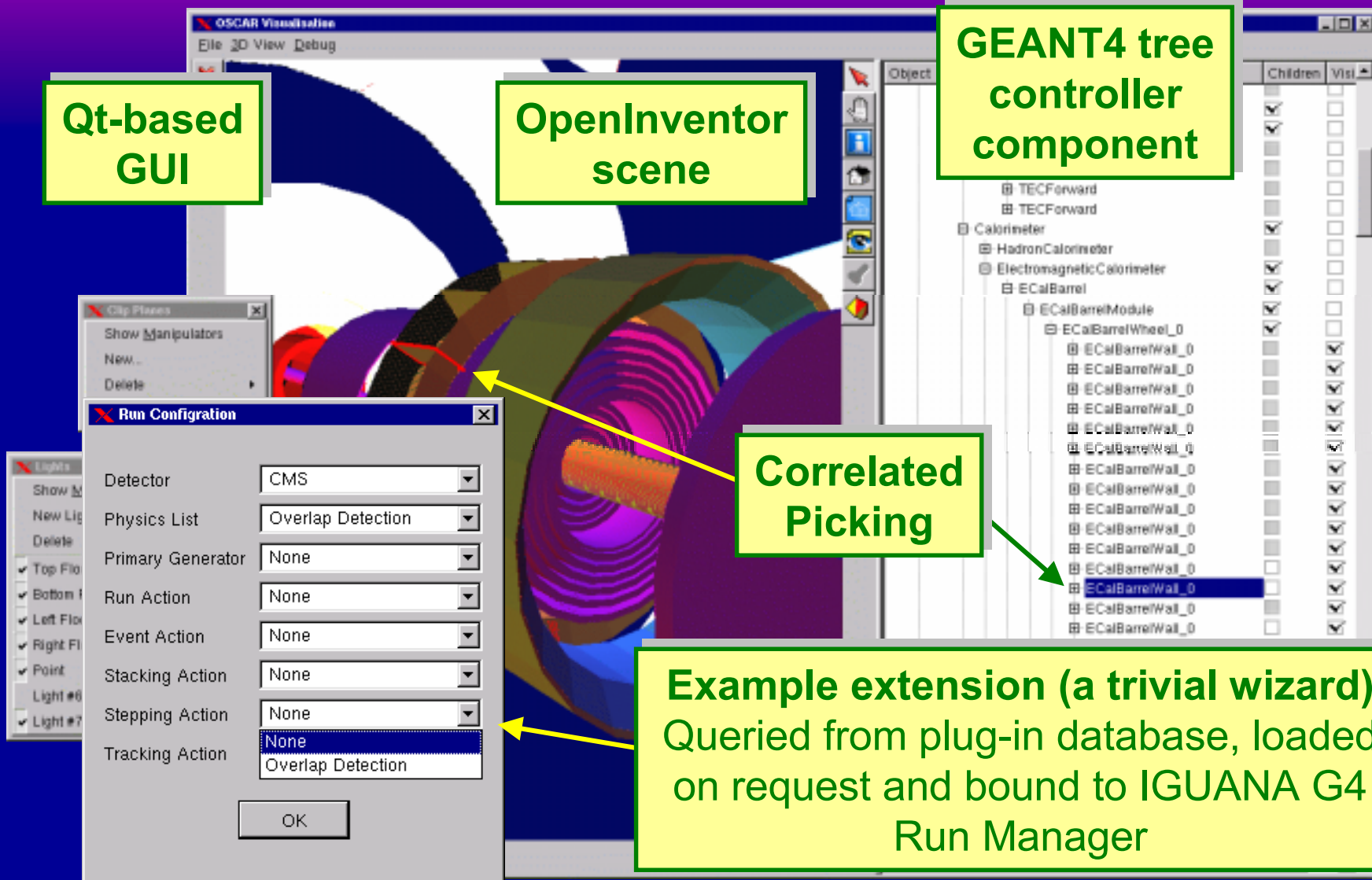
The PHENIX Offline Computing System
(Martin Purschke)

- ## Heavy ions at RHIC $(10^9$ Au+Au evt/yr)
  - ❖ Running experiment! **20 MB/sec** for few months / year

- ## ROOT-based C++ framework
  - ❖ Raw data in PHENIX (PRDF) format
  - ❖ Objectivity as main archival DB
  - ❖ ROOT DST's "wrapped" with templated accessors

- ## Large computing facility
  - ❖ **1.2 PB** capacity in HPSS STK silos ("data carousel" model)
  - ❖ **40 TB** RAID and **1200 CPU's**

- ## Year 1 data analyzed (QM'01, publications)
  - ❖ Year 2 data: gearing up for production

A new OO physics analysis framework for the H1 experiment
(Ursula Berthon)

- **H1 running until ~2006 (shutdown '99/'00)**
  - ❖ **6 TByte / year**   (50 million raw events / year)

- **Implemented new (ROOT-based) framework**
  - ❖ Exploiting established f77 code
  - ❖ Able to read legacy H1 data format

- **A number of extensions to ROOT required**
  - ❖ H1Tree: treatment of parallel trees
  - ❖ H1Pointer: persistent relations across files
  - ❖ RunCatalog: file/event handling (mySQL)

SND offline framework
(Dmitry Bukin)

■ **VEPP e+e– φ-factory is upgrading**
  ❖ energy and luminosity
  ❖ software to OO

■ **New C++ framework**
  ❖ Transient / persistent data separation
  ❖ SWIG-generated Python interface (and perl)
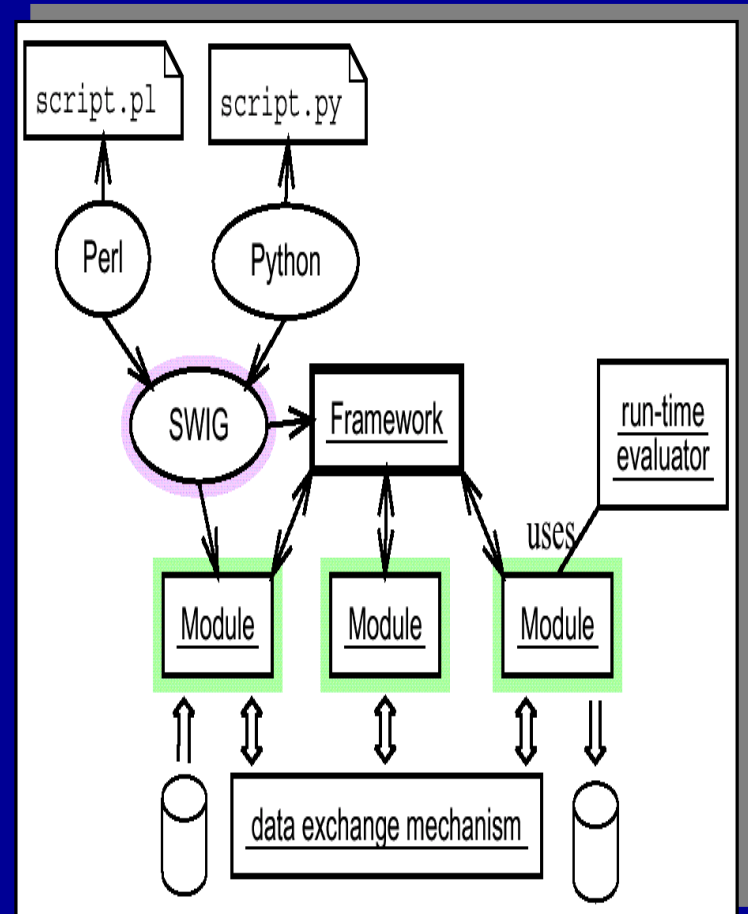
Experiments' Analysis Environments

(Session 2, Tuesday PM)

<u>*Outline:*</u>

Topics and program
Session 1: Architecture and Frameworks
Session 2: Experiments' Analysis Environments
Session 3: Generic Analysis Tools
Impressions, trends, and platitudes...

CMS Object-Oriented Analysis
(Stephan Wynhoff)

- **LHC beam in 2006 but already a large effort**
  - ❖ **~100 developers / users**
  - ❖ CERN, Italy, Russia, UK, USA,...
  - ❖ **~ 500 CPUs**

- **50 Tbyte in Objectivity**
  - ❖ Actively read for analysis
  - ❖ Integrated tape systems (CASTOR and ENSTORE)
  - ❖ Objectivity performing well

- **Interactive analysis**
  - ❖ PAW, ROOT, Anaphe/Lizard, IGUANA...

- **Active use and development of GRID tools:**
  [10-051, 10-052, 10-053]

Chart legend:
- CERN T0/T1(shared)
- Regional T1's
- Regional T2's

Y-axis: CMS CPU (kSI95) — 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000
X-axis: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007
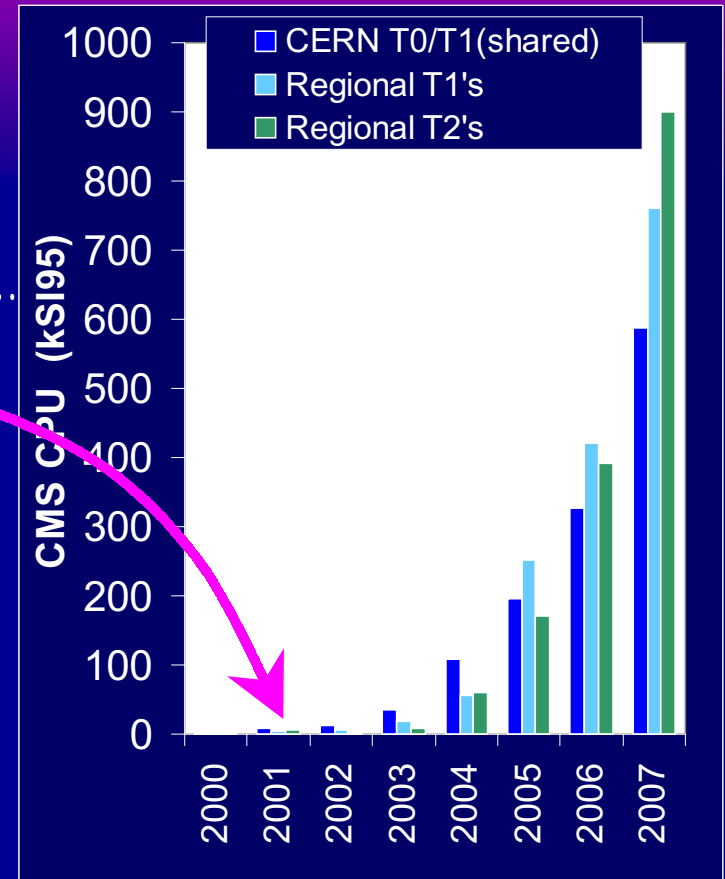
The BES III Offline Analysis Software and Computing Environment
(Mao Zepu)

- **BES III: expects ~400 TByte/year**
  - ❖ Sum of: raw, reconstructed, and simulated data

- **New OO/C++ framework and environment**
  - ❖ Backwards-compatible with BESII software
  - ❖ GEANT4 / CERN SW for simulation/analysis

- **New computing system at IHEP, Beijing**
  - ❖ **500 TByte / year** tape system and **25T Byte** disk
  - ❖ **36,000 MIPS** of CPU (PC farm)
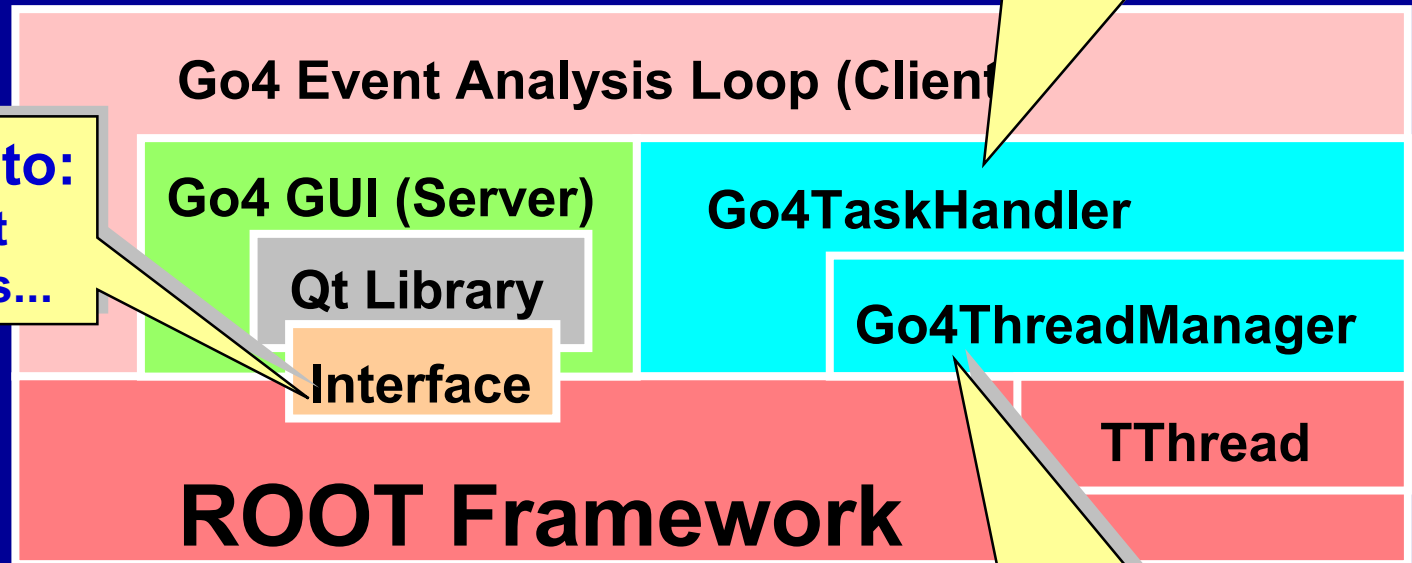
Prototype for a Generic Thin-Client Remote Analysis Environment for CMS
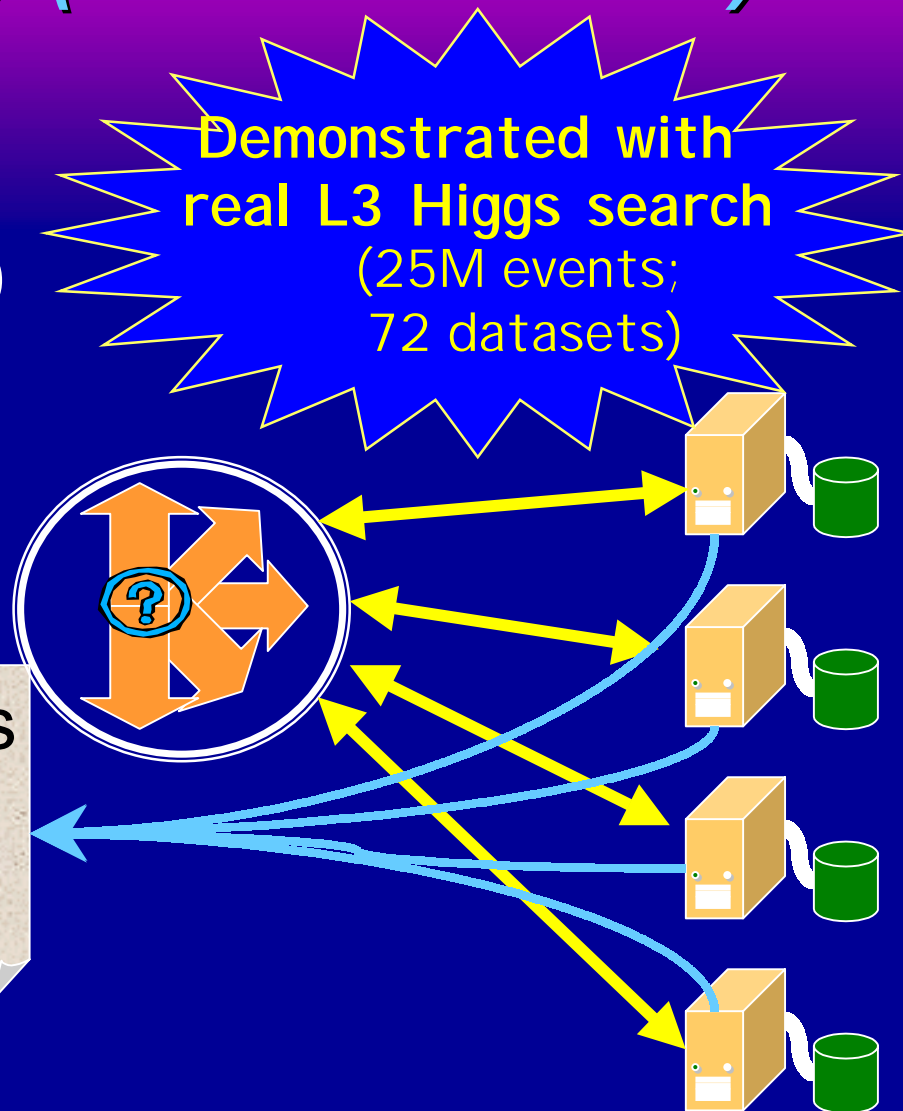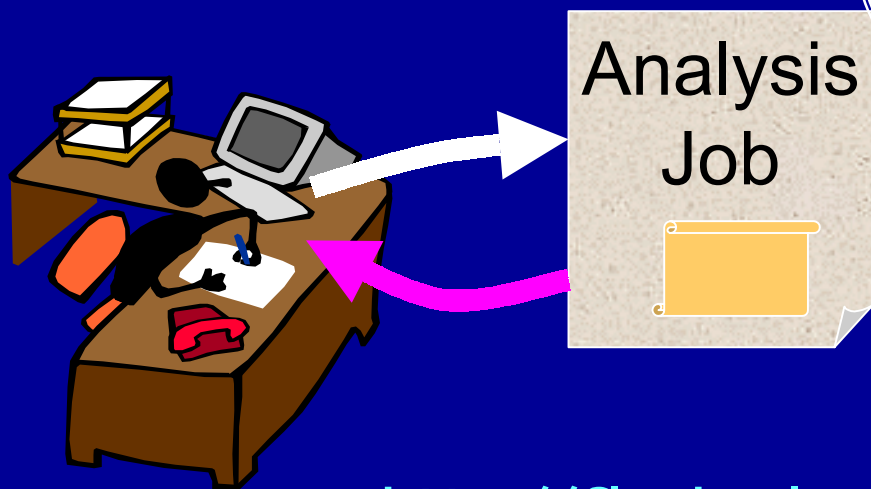## *(Conrad Steenberg)*

- **Data server** of histograms/tags (Clarens)
    - ❖ standalone process or CGI process in web server

- **Client** for remote analysis
    - ❖ Java (JAS) for analysis/histogramming of tags
    - ❖ Python analysis (using Lizard)

- **Communication** based on XML-RPC over http
    - ❖ currently transfers histograms and tags

- **GRID tools** foreseen
    - ❖ GSI authentication and data-moving (Globus)

http://heppc22.hep.caltech.edu

Distributed Analysis with Java & Objectivity (Jeremiah Mans)

BlueOx

- Resource discovery
- Brokering (job-to-data)
- Execution (sandbox)
- Result merging

Demonstrated with real L3 Higgs search (25M events; 72 datasets)

Analysis Job

http://flywheel.princeton.edu/BlueOx

Linear Collider Detector Analysis using JAS & ROOT
(Norman Graf)

- **Full suite of OO tools (C++ and Java)**
  - ❖ Simulation, Reconstruction & Analysis

- **Newcomers generally prefer JAS / Java**
  - ❖ Easier to get started
  - ❖ Extensible using plug-ins (FreeHEP)
  - ❖ Integrated analysis and development environment
  - ❖ Natural support for client-server remote analysis

- **See also:**
  - ❖ FreeHep Java Library   [8-017]
  - ❖ Java Analysis Studio 3.0   [3-022]
  - ❖ LCD Full Simulation & Reconstruction   [3-029]

  http://www-sldnt.slac.stanford.edu/nld

Generic Analysis and Visualisation Tools

(Session 3, Wednesday PM)

Outline:

Topics and program
Session 1: Architecture and Frameworks
Session 2: Experiments' Analysis Environments
Session 3: Generic Analysis Tools
Impressions, trends, and platitudes...

# Summary of the HEPVis'01 Workshop (George Alverson)

Northeastern UNIVERSITY

- **Initially visualisation; now also analysis tools**

- **Common software efforts**
  - ❖ HEPVis   (OpenInventor extensions; C++)
  - ❖ AIDA   (Abstract Interfaces for Data Analysis)
  - ❖ FreeHEP   (HEP related software; mostly-Java)

- **I will not try to (recursively) summarise**
  - ❖ sorry George…

- **…but I have shamelessly plagiarised !**
  - ❖ thanks George

Data Visualization and Graphics in ROOT *(Fons Rademakers)*

■ **Clear message: lots of graphics functionality!**

■ **Quasi-in-house graphical user interface**
  ❖ Functionally quite good (but little documentation)
  ❖ Go4 extension: can use ROOT with free Qt toolkit

■ **Rich canvases and many 2D primitives**

■ **Rather basic in-house 3D**
  ❖ But a virtual interface to OpenGL exists

■ **Technically could migrate to modern graphics packages (e.g. Qt, OpenGL,OpenInventor…)**
  ❖ But it would take quite some work

See also: 3-070, 3-075, 4-028, 7-013, 8-051, 10-021

Data Visualization and Graphics in ROOT *(Fons Rademakers)*
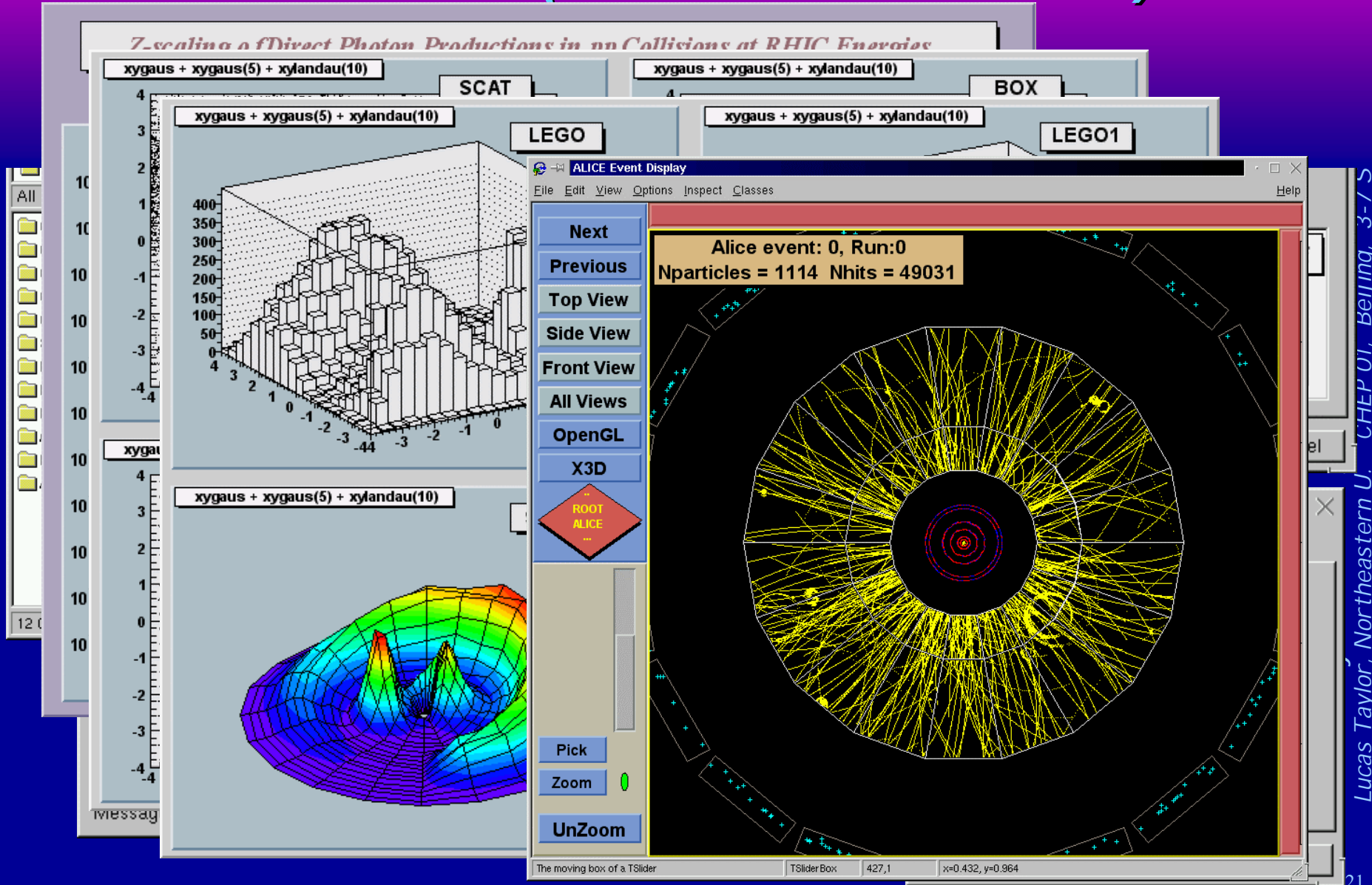
Abstract Interfaces for Data Analysis - Component Architecture for Data Analysis Tools (Andreas Pfeiffer)
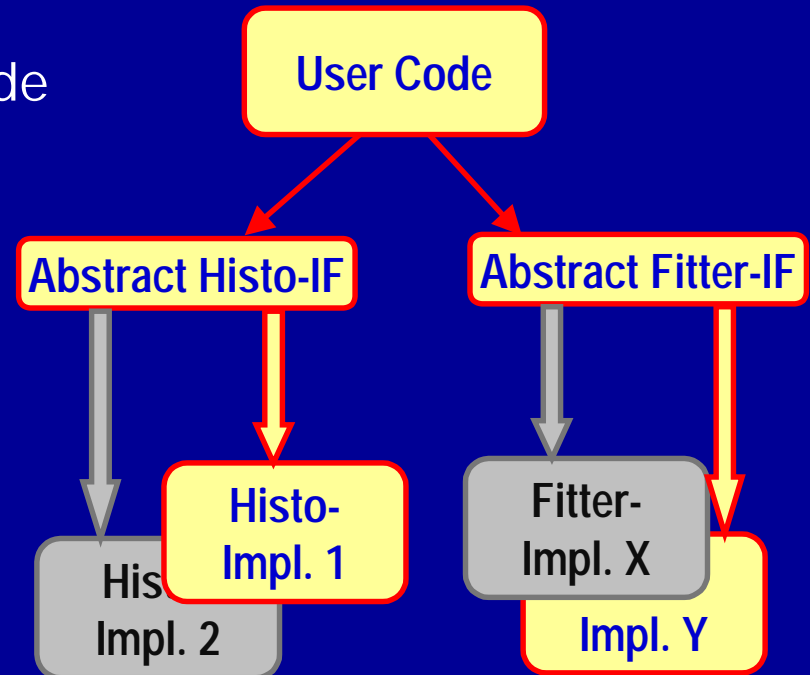
■ **AIDA defines abstract interfaces for common physics analysis tools**
  - ❖ No interference with user code
  - ❖ Minimizes component coupling
  - ❖ Choice of implementations
  - ❖ Replacement of components

■ **Who uses AIDA ?**
  - ❖ Anaphe / Lizard (next slide)
  - ❖ IGUANA (CMS visualization)
  - ❖ GAUDI (LHCB) framework
  - ❖ ATHENA (Atlas) framework
  - ❖ Analyzer modules in Geant 4
  - ❖ JAS
  - ❖ Open Scientist

  ☹ Conspicuously absent: ROOT, Hippodraw

Diagram boxes: User Code → Abstract Histo-IF, Abstract Fitter-IF; Histo-Impl. 1, Histo-Impl. 2, Fitter-Impl. X, Impl. Y

Anaphe - OO Libraries and Tools for Data Analysis (Jakub Moscicki)

- **OO replacement of CERNLIB (was "LHC++")**
- **Full range of (AIDA-compliant) packages**
 - ❖ Still improving, especially Lizard analysis tool

Data Analysis	Lizard - AIDA
Custom graphics (2-D)	Qt - Qplotter
Basic graphics (3-D)	OpenInventor – OpenGL
Basic math	NAG C
HEP foundation	CLHEP
HEP math	FML - Gemini - CLHEP
Histograms	HTL
Database	HepODBMS
Persistency	ODMG/Objectivity DB
C++	Standard Libraries

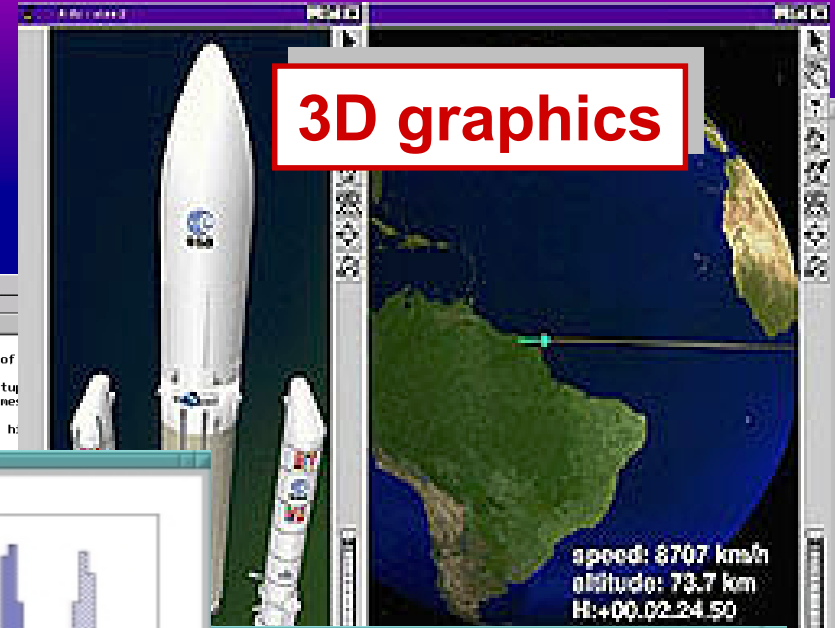**Not "all or nothing"**

If you don't want some bit (e.g. Objectivity) you can still use the rest of it

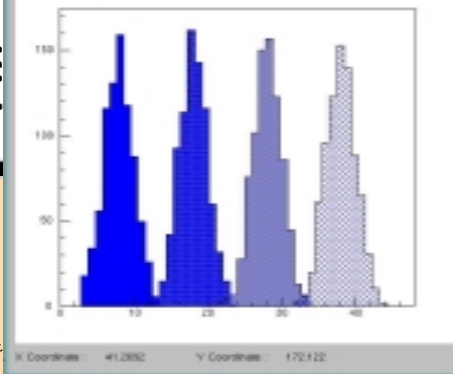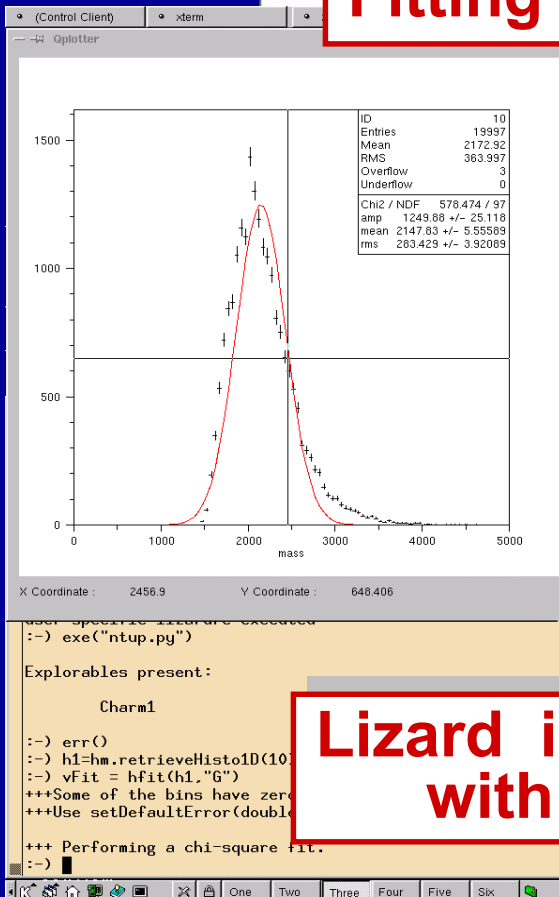# Anaphe - OO Libraries and Tools for Data Analysis (Jakub Moscicki)

3D graphics
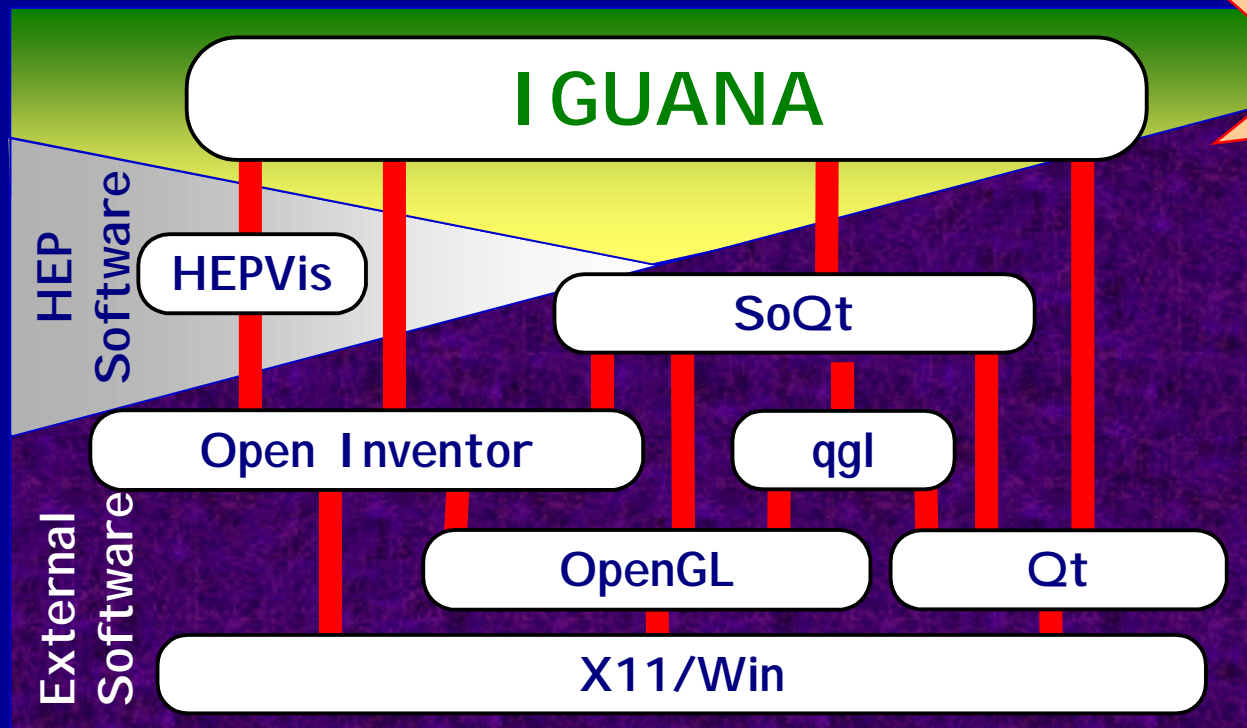
Fitting
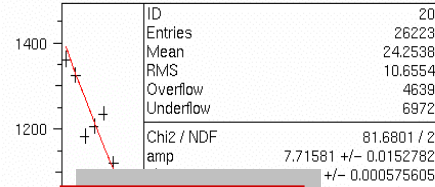
2D plotting

Lizard interactivity with Python

The IGUANA Interactive Graphics Toolkit with Examples from CMS and D0 *(Ianna Osborne)*
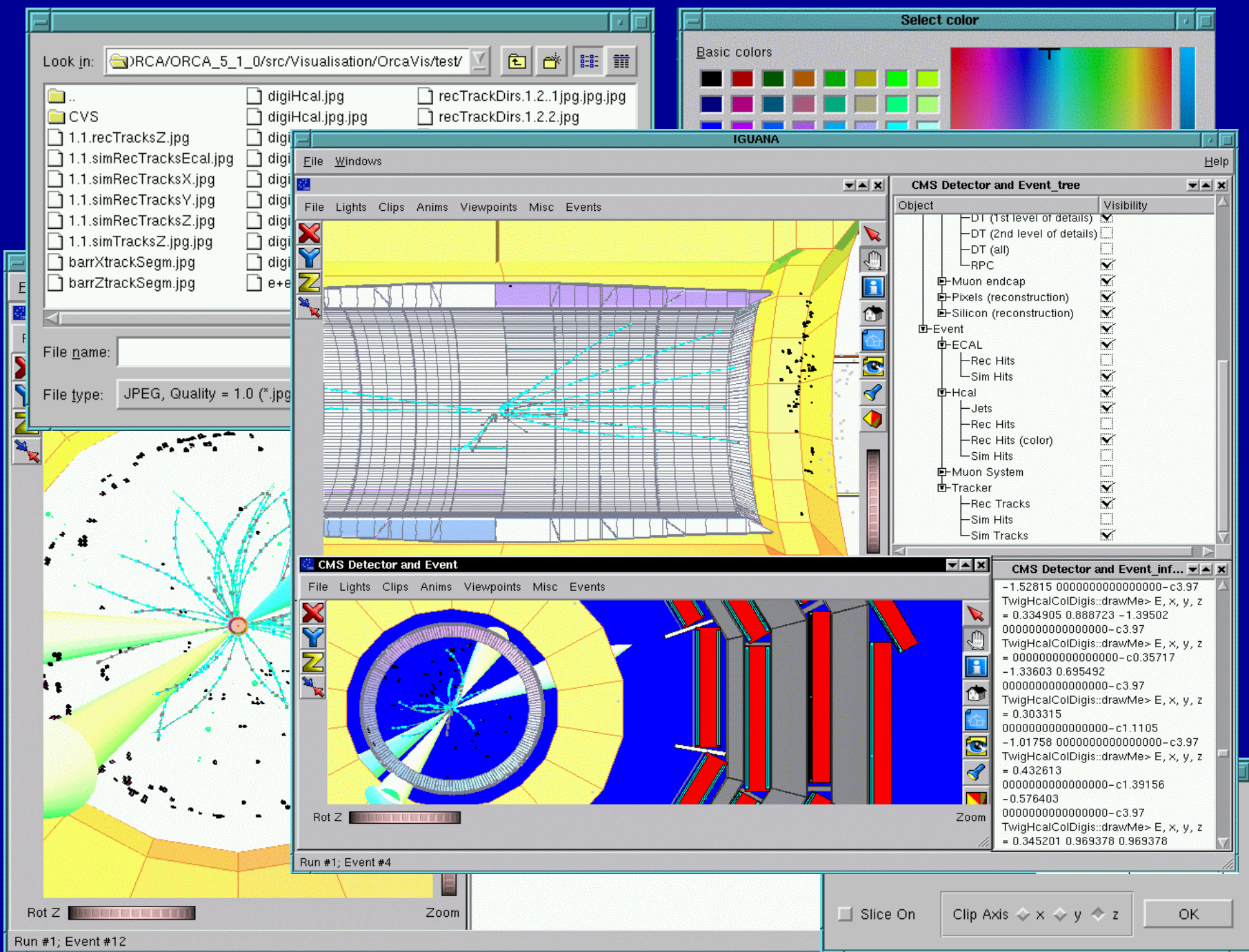
- C++ event display tools (used: CMS, D0,L3...)
- GEANT3 and GEANT4 browsers

All components are now free

IGUANA

HEP Software

HEPVis

SoQt

Open Inventor

qgl

External Software

OpenGL

Qt

X11/Win

See also:
"IGUANA Open Architecture"
[3-039]

IGUANA free from:
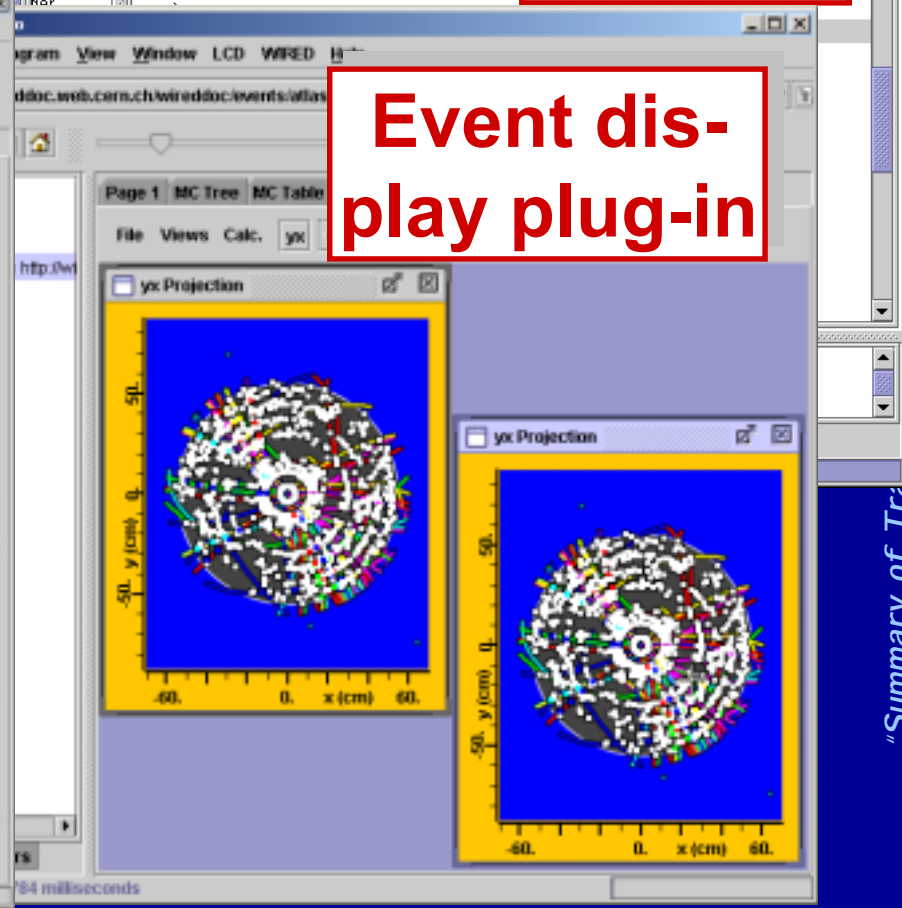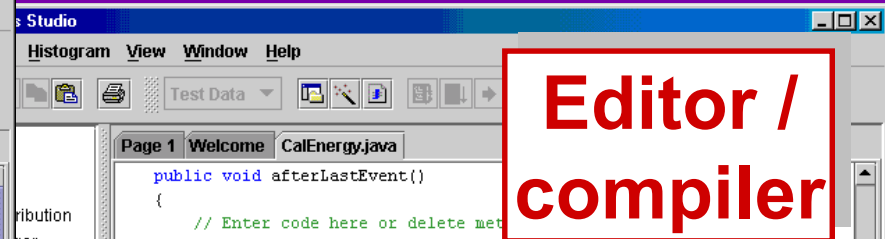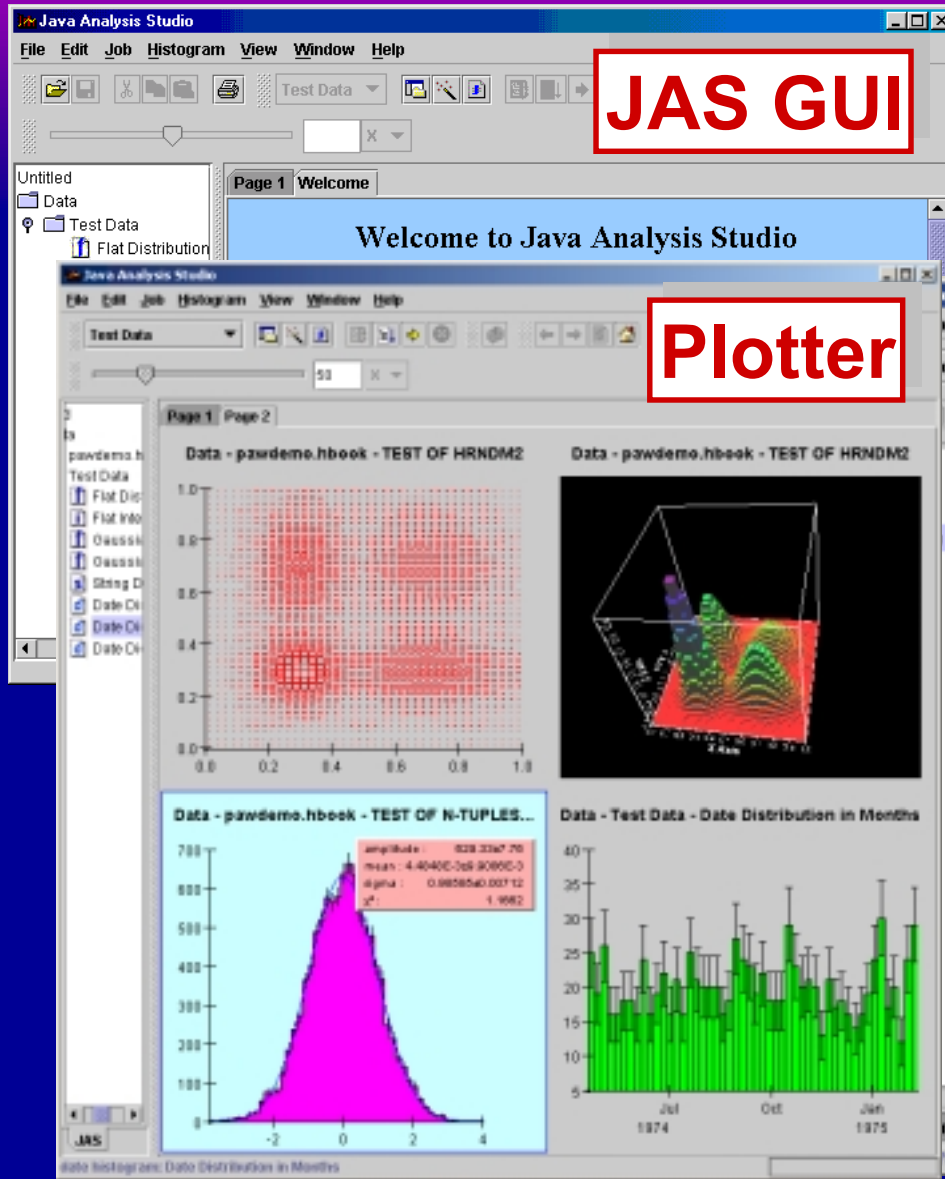
http://iguana.cern.ch

Java Analysis Studio (JAS) 3.0
(Norman Graf)

- **Rich GUI and well-integrated environment**
    - ❖ Built in **editor / compiler**

- **Designed to be easily extended**
    - ❖ **Plug-ins** and **Data Interface Modules** (DIM's)

- **New for the (AIDA-compliant) JAS 3.0:**
    - ❖ Scripting (probably) using **Jython**
    - ❖ Improved plug-in handling (in FreeHEP)
    - ❖ New DIM to **read ROOT files** (version ≥ 3.0)
    - ❖ **tuple explorer** plug-in: PAW, ROOT, SQL, & ASCII files

- **Local and Client-Server Operation**
    - ❖ Prototype system for distributing analysis on a farm
    - ❖ Future: integrate with GRID services (with e.g. "BlueOx")
        - ➤ Authentification, data catalog, resource locator,...

http://jas.freehep.org

Java Analysis Studio (JAS) 3.0
(Norman Graf)

JAS GUI

Editor / compiler

Plotter

Event dis-play plug-in

Impressions, Trends and Platitudes

An attempt to objectively summarise
- not intended to offend
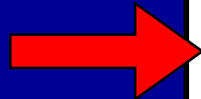
> ## Outline:
>
> Topics and program
> Session 1: Architecture and Frameworks
> Session 2: Experiments' Analysis Environments
> Session 3: Generic Analysis Tools
> Impressions, trends, and platitudes...

Conspicuous by their absence

- **No mention of commercial integrated analysis and visualisation systems**
  - ❖ AVS Express, IDL, IRIS Explorer…assume dead for HEP

- **Graphics performance not mentioned**
  - ❖ Moore's law is doing its job ?

- **Virtual reality**
  - ❖ not useful ? too expensive ? just a passing fad ?

- **No talks by IT groups at: FNAL,DESY,KEK,…**
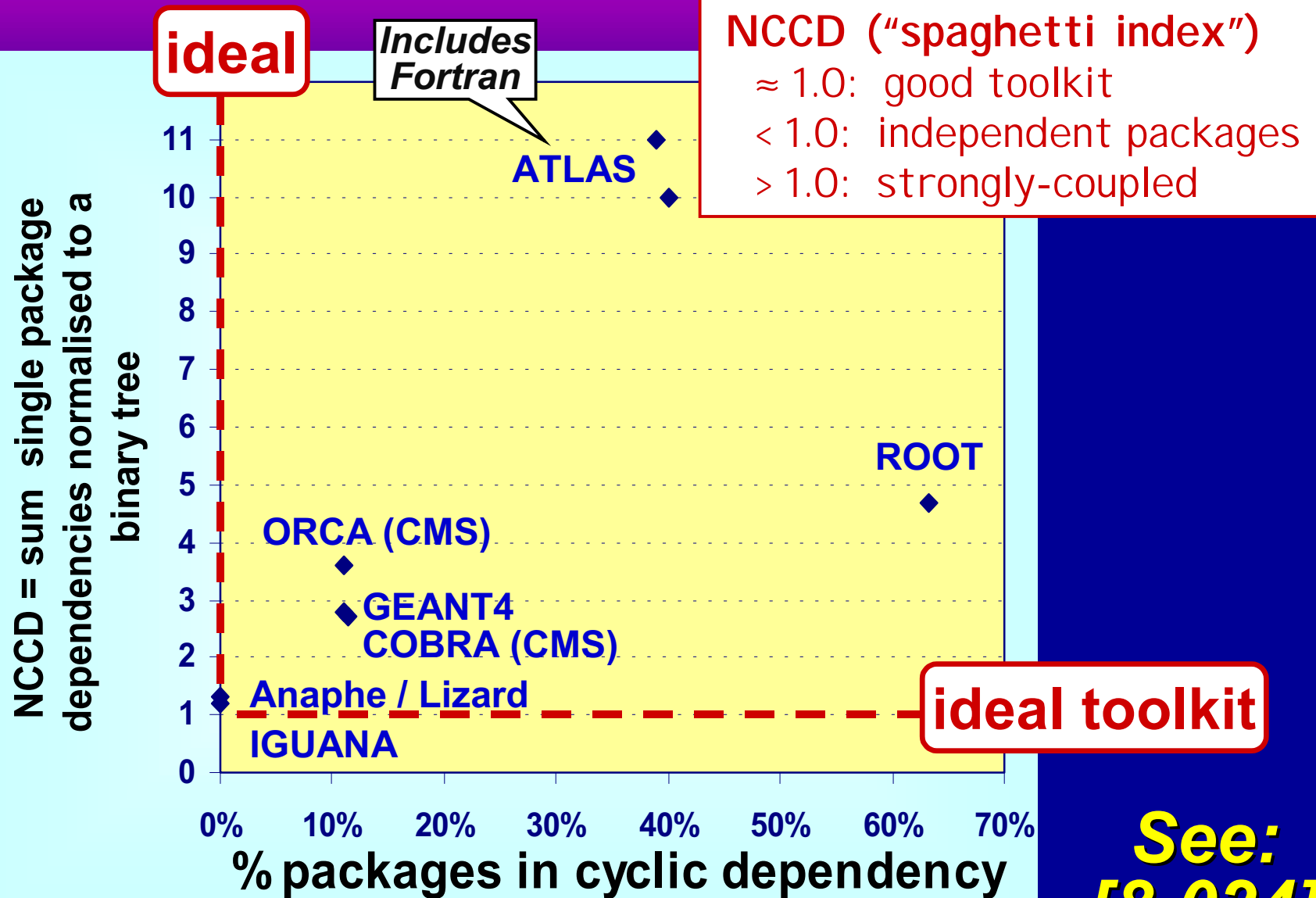 - ❖ Is it significant ? Quite a few from CERN and SLAC

Alive and well but simply not presented here
  - ❖ CMS "COBRA" framework
  - ❖ OpenScientist OnX/Lab packages
  - ❖ WIRED event display

Some Personal Disappointments

- **Generic analysis tools: duplicated effort**
  - ❖ Anaphe / Lizard, HippoDraw, JAS, OnX/Lab, ROOT,…
- **ROOT: well-used and very functional …but…**
  - ❖ **Schizophrenic**… am I a framework ? a persistency mechanism ? an analysis toolkit ? a graphics library ?
  - ❖ No plans to make ROOT AIDA-compliant
  - ❖ Hope: split ROOT into loosely-coupled "ROOT-lets"
    - ➢ Framework, persistency, CINT,…
    - ➢ Histograms, fitting, functions,…
    - ➢ GUI, 2D graphics,3D graphics,…
- **Disturbing: software quality barely mentioned**
  - ❖ Can we believe the handful of Higgs in ~10^9 evts/yr ?
  - ❖ Modularity in design…can it be verified /quantified?

"Ignominy" Tool to Quantify Modularity

ideal

Includes Fortran

NCCD ("spaghetti index")
- ≈ 1.0: good toolkit
- < 1.0: independent packages
- > 1.0: strongly-coupled

NCCD = sum single package dependencies normalised to a binary tree

ATLAS

ROOT

ORCA (CMS)

GEANT4
COBRA (CMS)

Anaphe / Lizard

IGUANA

ideal toolkit

% packages in cyclic dependency

See: [8-024]

Favourable Developments

- **Acceptance of collaborating frameworks rather than a single all-powerful framework**

- **DIM's for reading ROOT files**
  - ❖ `C++:` RioGrande
  - ❖ `Java:` hep.io.root in FreeHEP, used by JAS & Hippodraw

- **Several prototypes for remote analysis**
  - ❖ (Correlated) lightweight Java clients: BlueOx,Clarens,JAS

- **Progress towards modular architectures – buzz-words are turning into realities**
  - ❖ "Abstract Interfaces", "components", "plug-ins",…

- **Some HEP-wide de-facto standards emerging**

Emerging Standard ?
Python as "Software Glue"

- ## Clear trend towards Python
  - ❖ Used by: ATLAS (Athena),CMS, D0, LHCb (Gaudi), SND,…
  - ❖ Used by: Lizard/Anaphe, HippoDraw, JAS (Jython)…
  - ❖ Architecturally, scripting is "just another service"
  - ❖ ROOT is the exception to the "Python rule"
    - ➢ CINT interpreter plays a central role
    - ➢ Developers and users seem happy

- ## Python is popular with developers…
  - ❖ Rapid prototyping;   gluing together code
  - ❖ (Almost) auto-generation of wrappers (SWIG)

- ## …but acceptance by users not yet proven
  - ❖ Another language to learn, syntax,…

Emerging Standards ?
HEP Graphics Toolkits

■ **Some significant commercial graphics software has become free**
- ❖ High-quality, good support, well-documented
- ❖ Linux, other Unixes, and Windows

■ **Qt** **(Graphical User Interface toolkit)**
- ❖ <u>Used by</u>: Anaphe / Lizard, CLEO, CMS, D0, Go4, etc…
- ❖ ROOT does not use Qt but it is interoperable (Go4)

■ **OpenGL/OpenInventor**(low/high-level graphics)
- ❖ <u>Used by</u>: Anaphe / Lizard, CDF, CMS, D0, GEANT4,…
- ❖ ROOT now has a virtual GL interface

■ **If the above holds true, maybe add the IGUANA C++ event display toolkit**

Emerging Standards ?
Analysis Tools Interfaces

- **AIDA** Abstract Interfaces for Data Analysis
  - ❖ Anaphe / Lizard
  - ❖ IGUANA (CMS visualization)
  - ❖ GAUDI (LHCB) framework
  - ❖ ATHENA (Atlas) framework
  - ❖ Analyzer modules in Geant4
  - ❖ JAS
  - ❖ Open Scientist
  - ☹ No participation from ROOT and Hippodraw

- **Less risk, easier maintenance, and (hopefully!) less duplication of effort for e.g.**
  - ❖ Histograms, plotters, fitting, function algebra, canvas, event viewers, GUI toolkits and extensions, etc…
  - ❖ Aside: why not a HEP software project hosting site ?

Acknowledgements

Thank you all for your patience

Thanks to the many people who helped me with this session

❖ Organisers:
  ➢ especially Guo Yanan, Yu Chuansong, and Zhijia Sun
❖ Co-convenors:
  ➢ Irwin Gaines and Takashi Sasaki
❖ Speakers:
  ➢ especially George Alverson, Vincenzo Innocente, Ianna Osborne, Andreas Pfeiffer, Lassi Tuura...