Introduction to the General H1 software and Data structure

S. Levonian
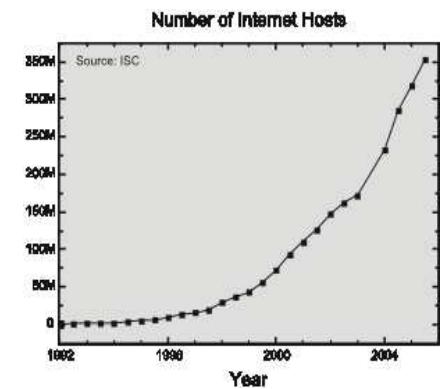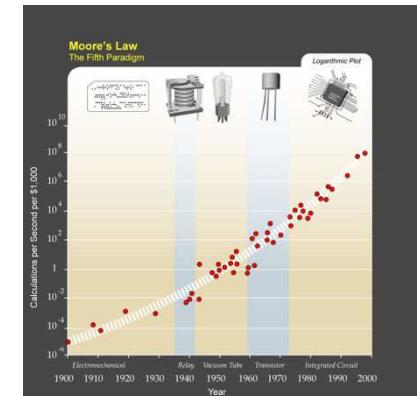
- **The history**

  20 years of the H1 computing

- **General H1 software**

  one million lines of code

- **H1 data**

  data location and access methods

- **Documentation**

  Manuals, info tools, help

How to survive in a rapidly changing environment?

- **Moore's law:** computing power in 20 years raised by $\sim 2^{10} = 1000$

- **Architecture/OS:**
  mainframe era $\Rightarrow$ minicomputers and WS $\Rightarrow$ PC farms
  (H1 saw: AIX, ALLIANT, ALPHA, APOLLO, AXP, Ultrix, HPUX, IBM,
  IBMMVS, IBMRT, MAC, MIPS, OS-9, RTPC, VAX, VAXVMS, VMS,
  UNIX, SGI, SUN, LINUX)

- **Languages:** Fortran(1954), C(1972), SQL(1978), C++(1983)
  Perl(1987), Python(1990), Java(1995)

- **Code managment:** PCM(1986) $\Rightarrow$ CMZ (1990)$\Rightarrow$ CVS(2001)

- **Graphics:** GKS $\Rightarrow$ motif $\Rightarrow$ revolution (infinite capabilities)

- **Mass storage** development $\Rightarrow$ changing paradigm of data access
  H1 logging rate: $5 \Rightarrow 25$Hz,   Data volume: $2 \Rightarrow 50$Tb/year

- **Internet, open source s/w, distributed computing, GRID...**
  Should be flexible enough to be ready for unexpected things.

Extremely non-trivial to plan software for such long lived projects

Moore's Law
The Fifth Paradigm

Number of Internet Hosts

Lesson number 1: Take software seriousely

Cannot simply rely on the general computing progress

Big mistake to consider software projects as 2-nd priority

This does not work and will cause a lot of problems!
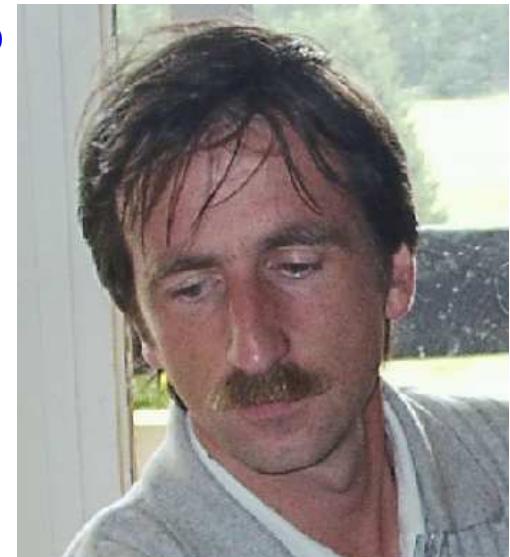
Good data structure and

software organization

is vital for efficient analysis!

H1 s/w rules and recommendations

The basic conventions were worked out in the beginning
(largely thanks to the main contributors: V.Blobel and S.Egli)
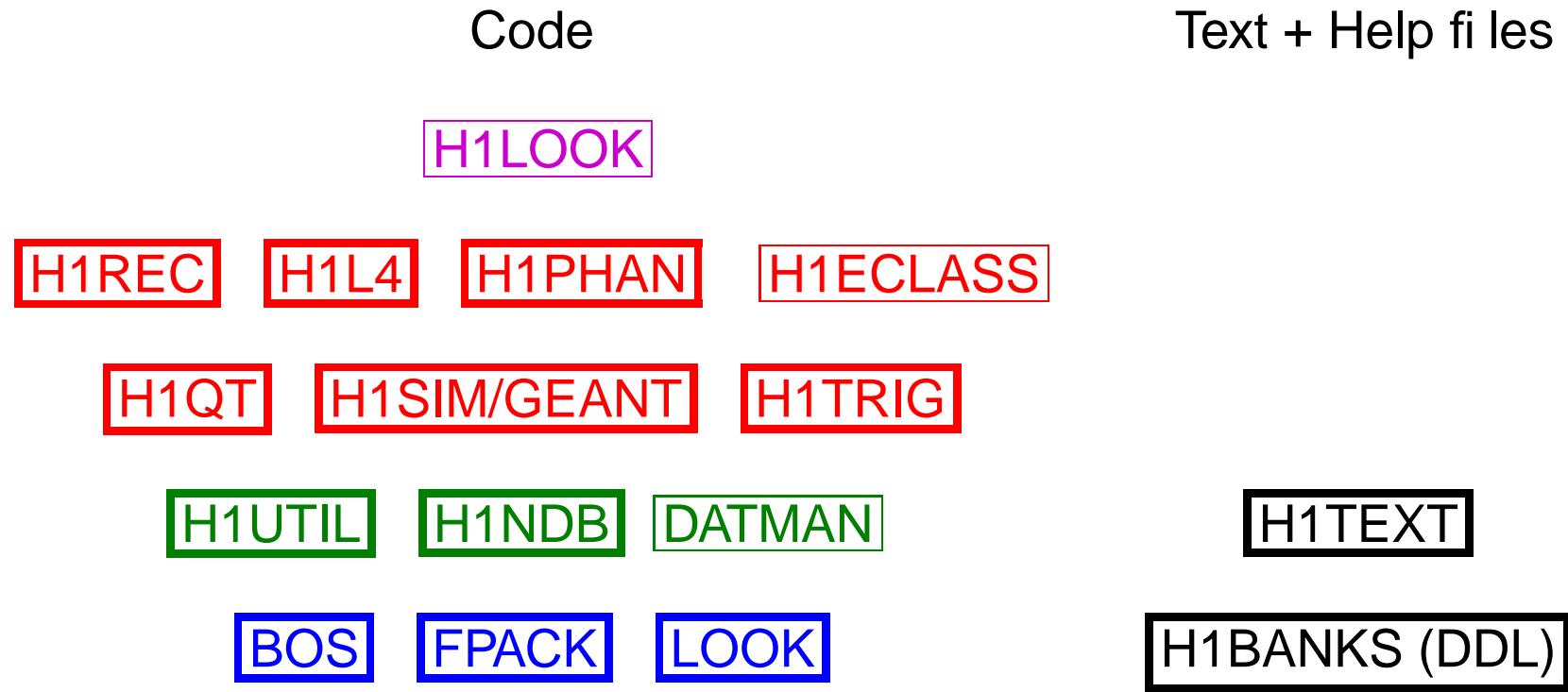
- **Programming language is F77**

  (later an extention to C was accepted, mainly in H1QT and H1TRIG)

  Note: basic H1 s/w amounts to 7000 s/r $\approx 10^6$ lines of code

- **Modular structure (s/w note 12, 1990-92)**

  communications between modules only via BOS banks, standartized

  steering, debugging, error logging, in particular no Fortran STOP (!)

- **H1 data structure and formats (s/w notes 1, 13, 1989-90)**

  naming conventions, bank formats and access methods, machine indep. I/O

- **Maintain backward compatibility and portability**

  (Up to 20 different platforms were used in H1 history!)

  avoid/minimize branches thus making support easier

- **Split general H1 s/w and analysis s/w**

  the former is hard to rewrite, while the latter is easier to adopt

- **Respect general programming culture**

  Avoid hardwired parameters, use DB steering banks instead.

  Provide readable and suffi ciently commented source code.

  ...

Overview of the H1 software packages

The H1 software is organized in packages with the layered structure:

Code   Text + Help fi les

H1LOOK

H1REC   H1L4   H1PHAN   H1ECLASS

H1QT   H1SIM/GEANT   H1TRIG

H1UTIL   H1NDB   DATMAN   H1TEXT

BOS   FPACK   LOOK   H1BANKS (DDL)

Basic utility packages – no H1 specific code, fully backward compatible (author: V.Blobel)
H1 low level packages and higher packages (DATMAN, H1ECLASS and H1LOOK are frozen)
Higher level packages can call routines from lower level packages, but not vise versa.

General H1 software
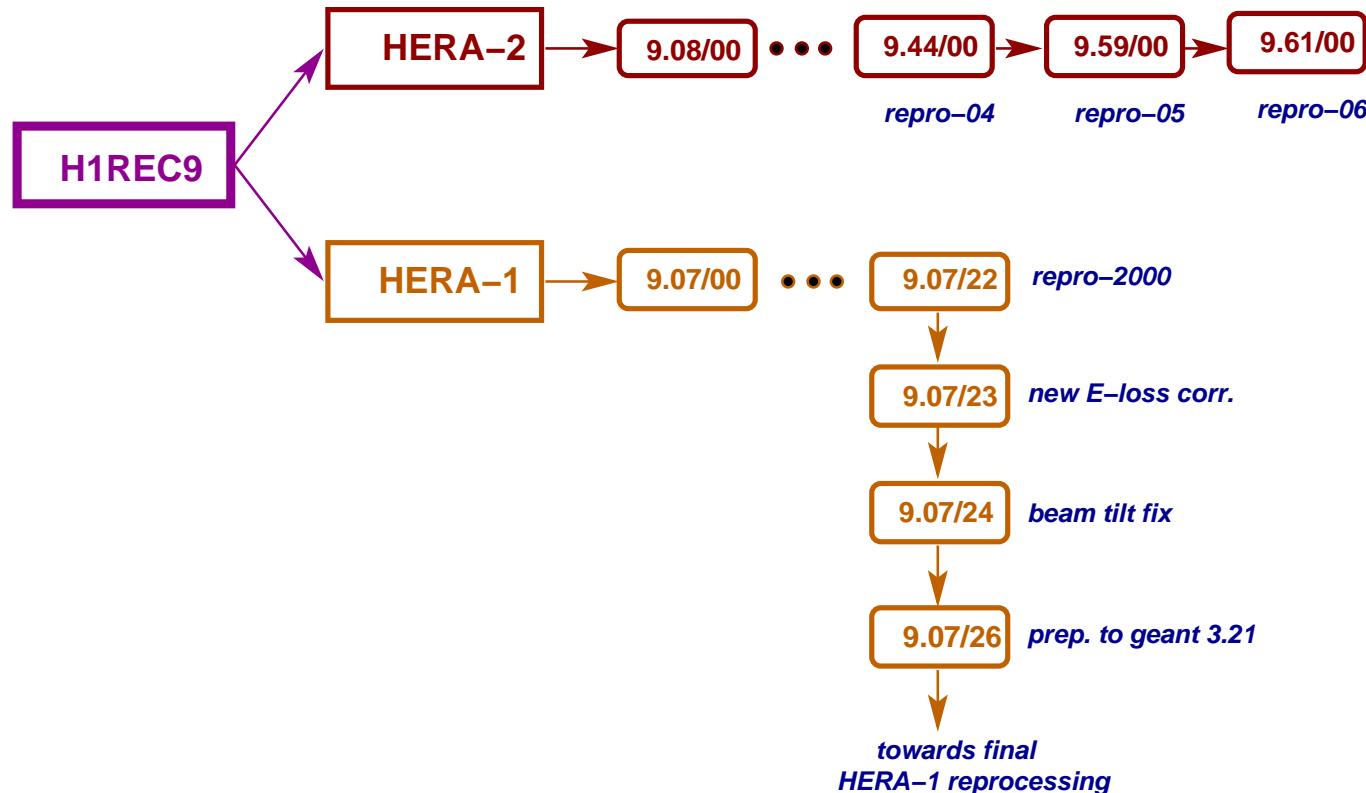
BOS   –   Dynamic data and memory management system *(V.Blobel, 1985)*

FPACK   –   Machine-independent data handling I/O package *(V.Blobel, 1991)*

LOOK   –   General system for graphics applications in physics *(V.Blobel, 1990)*

H1NDB   –   H1 database software package (interface to Oracle) *(L.Criegee, C.Kleinwort)*

H1UTIL   –   Collection of utility functions, used in H1 application packages *(H1 Collab.)*

DATMAN   –   A Data Management Tool package (used mainly in H1ED) *(U.Leberton et al.)*

H1QT   – $Q - t$ analysis code for CT, FT and Fμ (witten in C) *(D. Sankey et al.)*

H1SIM   –   H1 detector simulation based on GEANT3 *(H1 Collab.)*

H1TRIG   –   H1 trigger simulation package *(H1 Collab.)*

H1REC   –   H1 reconstruction package *(H1 Collab.)*

H1L4   –   H1 Level 4 and Event classification software *(H1 Collab.)*

H1ECLASS   –   Old (frozen) Level 5 Event classification package *(H1 Collab.)*

H1PHAN   –   H1 Physics Analysis library (Fortran based) *(H1 Collab.)*

H1LOOK   –   H1 Event Display (frozen) *(U.Berthon, S.Levonian, Th.Naumann)*

Most of the packages are stable. Main developments continue for H1SIM, H1REC, H1TRIG

Lesson number 2: Obey rules

1) H1REC branching due to FTREC backward incompatibility

HERA-1 branch is falling behind at the moment, needs an extra effort for parallel support

2) Violation of the customized access methods to BOS bank content

Direct access to bit-packed information instead of using CALL BKTOW caused 2 man-months work
of cleaning and fi xing the bugs when we had to use the mix of big-endian and little-endian machines

SIMREC Problems and Solutions

Documented technical problems of H1SIMREC

- Crash in h1trig/bigray/brth.F for HERA2 simulation. Fixed in **h1trig** version 1.70/00
- Problems with CJC simulation in h1simrec95800. Reason: name clash between h1rec and h1sim: s/r CJISOC Solved in h1rec 9.59/00
- Funny regular structures (as a function of Event number) in D* MC simulations M.-O.Boenig, 17 October 2006 Under investigation...

PS. If you know more problems, please report to me, preferably with a documented evidence.

Last update: **18.10.2006 by S.Levonian**

Observed problems in data vs MC comparisons

- General H1SIM status (S.Levonian)
- Description of nuclear absorption for low momenta particles (K.Daum)
- dE/dx simulation for HERA1 and HERA2 (K.Daum)
- Pt mismatch between tracks and clusters for high momenta electrons (C.Diaconu, E.Sauvan: (17-Nov-2004, 26-Oct-2005, 2-Nov-2005)
- Forward/backward beamline simulation (V.Andreev, S.Levonian, Diff.group)
- Problem with energy reconstruction in SpaCal in HERA2 MC (at few per mille level) (Boenig, Marc-Oliver) -- see here discussion 1, discussion 2 and explanation
- Wrong Spacal z-position in MC for HERA-2, by approx. 2.9cm (fix is waiting for manpower, until that should be corrected at the analysis level) -- See e.g. ELAN alignment plots for details.

PPS. If you know more problems, please report to me, preferably with a documented evidence.

Last update: **18.10.2006 by S. Levonian**

Version Table: H1 s/w version compatibility database

```
#------------------------------------------------------------------------------------

 Row H1REC8 H1REC9 H1REC H1PHAN H1L4 H1ECLASS H1SIM H1UTIL H1TRIG H1LOOK H1BSTREC
#------------------------------------------------------------------------------------
 31     -      -    96100  31200  63200   20700   35400  63304  18300    -     10800    !test
 30     -      -    96000  31200  63000   20700   34500  63300  18000    -     10500    !new !hera2_06a
 29     -      -    95900  31200  62800   20600   34200  63200  17000    -     10400    !pro !hera2_05b
 28     -      -    90726  31100  50800   20600   33900  63000  12900  11907    -       !hera1_05c
 27     -      -    94700  31200  62600   20600   33900  63000  12900    -     10400    !hera2_05a
 26     -      -    90723  31100  50800   20600   33800  62800  12900  11907    -       !hera1_05b
 25     -      -    90722  31100  50800   20600   33800  62200  12900  11907    -       !hera1_05a
 24     -      -    94400  31100  62500   20600   33800  62800  12900    -     10400    !repro04
 23     -      -    94300  31100  62400   20600   33700  62700  12800    -     10400    !hera2_04b
 22     -      -    94201  31002  62200   20600   33401  62300  12500    -     10300    !hera2_04a
 21     -      -    90722  31003  50800   20600   33101  62200  10112  11907    -       !hera1_04
 20     -      -    93000  31002  61100   20600   33101  62000  12000    -     10100    !hera2_03
 19     -      -    90718  30800  50800   20506   32800  40926  10112  11907    -       !hera1_03  !mchera1
 18     -    90605    -    30110  50611   20506   31003  40924  10112  11907    -       !hera1_02
 16     -    90601    -    30109  50610   20506   30924  40921  10108  11900    -       !repro00
#------------------------------------------------------------------------------------
 15     -    90527    -    30107  50607   20506   30924  40914  10108  11900    -       !mc99_00
 14     -    90421    -    30106  50600   20506   30921  40913  10108  11900    -       !00c
 13   81405  90413    -    30105  50600   20506   30919  40912  10104  11801    -       !00b
 12   81404  90411    -    30104  50600   20506   30918  40911  10101  11800    -       !00a !mc96_97
  9     -    90407    -    30102  50501   20505   30912  40910  10008  11705    -       !99a
  8     -    90303    -    30004  50400   20505   30906  40906  10003  11704    -       !repro98
  7   81403    -      -    30004  50204   20505   30906  40906  10003  11704    -       !old96_97
  6     -    90106    -    30003  50204   20505   30905  40905  10003  11703    -       !98b
  5   81400    -      -    30002  50204   20505   30724  40725    -    11700    -       !98a
  4   81103    -      -    30000  50204   20504   30713  40716    -    11612    -       !mc95 !97b
  3   81103    -      -    20311  50204   20504   30713  40716    -    11612    -       !keep
  2     -      -      -    10908  40317    -      30709  40710    -      -      -       !mc94
  1   80702    -      -    20311    -     20405   30648  40628    -      -      -       !oldvers
#------------------------------------------------------------------------------------
```

H1UNIX – H1 s/w installation toolkit *(U.Berthon,   https://www-h1.desy.de/icas/imanuals/h1unix.html)*

Release policy, installation rules, guidelines for librarians *(H1 s/w notes 54,59)*

Present status and Plans

- **Recent Global releases**

  ▷ 12-Jan-2006 (VT29) → version for reprocessing of HERA2 e^-p data

  ▷ 17-Mar-2006 (VT30) → version for 2006 setup (BST/FST in) + trigger s/w updates + CST bug fi x

  ▷ ??-Oct-2006 (VT31) → version for reprocessed 2006 e^-p data   (imminent!)

- **Intermediate releases since last global release**
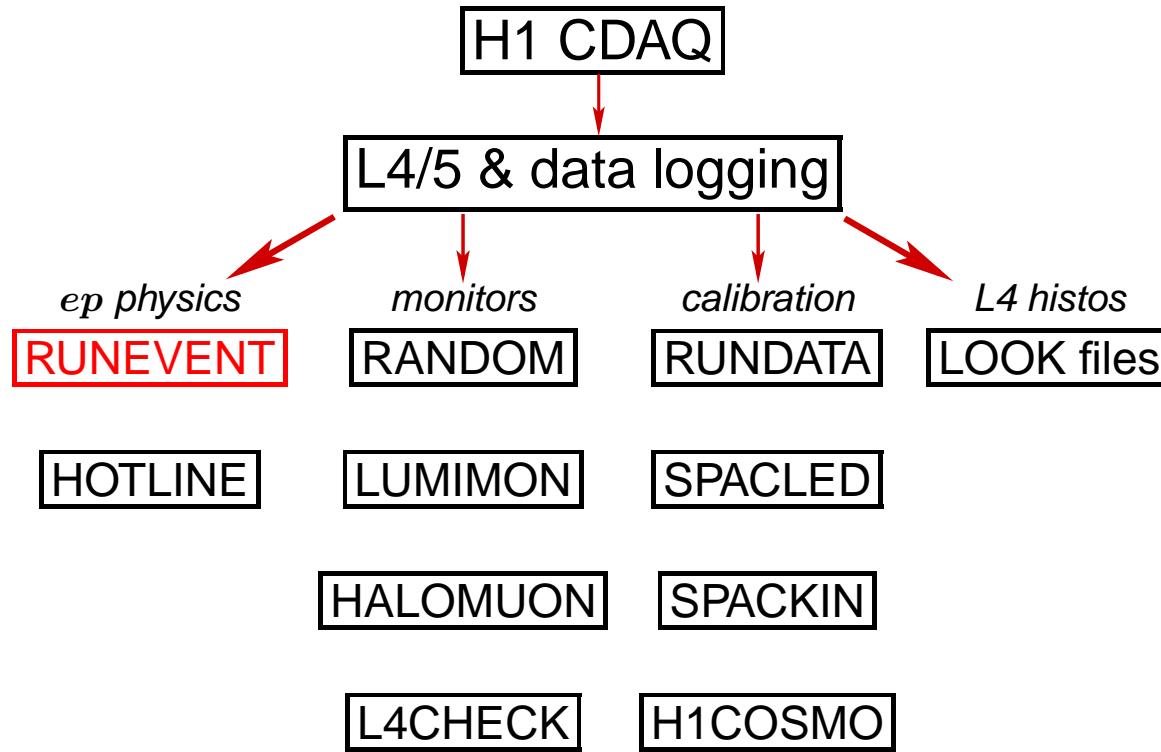
H1REC	11	CSTREC,CJCREC,FTREC,BSTREC,Lumi, COZ cal., DST updates, bug fi xes	9.61/02
H1SIM	6	CST,FTT,FTI2, Run dep., general noise s/w, fwd particle sim., improved steering	3.55/00
H1TRIG	3	Update of L2/L3 utilities, L2NN, CIP2K updates	1.83/00
H1L4	3	Updates of cl. 13, 14, 15,16,18 (HQ, high-y F_L, FNC, VM), online monitoring	6.32/00

- **To be done**

  ▷ H1SIM: Finilise and implement new run-dep noise simulation for all relevant subdetectors

  ▷ H1REC: Yet improve global tracking to achieve ultimate precision; make HERA-1 branch up-to-date

  ▷ H1TRIG: Add Jet trigger, full FTTEMU

H1 data flow

H1 data are sets of BOS banks combined and written as Fpack records to different streams

```
                        ┌──────────┐
                        │ H1 CDAQ  │
                        └──────────┘
                             │
                             ▼
                   ┌────────────────────┐
                   │ L4/5 & data logging│
                   └────────────────────┘
         ┌───────────┬───────────┴───────────┬───────────┐
         ▼           ▼                       ▼           ▼
    ep physics    monitors              calibration   L4 histos
   ┌──────────┐ ┌──────────┐           ┌──────────┐ ┌──────────┐
   │ RUNEVENT │ │ RANDOM   │           │ RUNDATA  │ │LOOK files│
   └──────────┘ └──────────┘           └──────────┘ └──────────┘

   ┌──────────┐ ┌──────────┐           ┌──────────┐
   │ HOTLINE  │ │ LUMIMON  │           │ SPACLED  │
   └──────────┘ └──────────┘           └──────────┘

                ┌──────────┐           ┌──────────┐
                │ HALOMUON │           │ SPACKIN  │
                └──────────┘           └──────────┘

                ┌──────────┐           ┌──────────┐
                │ L4CHECK  │           │ H1COSMO  │
                └──────────┘           └──────────┘
```

H1 raw data files ≡ Sum of all records (except LOOK histos): */acs/data/YY/rawd/*
RUNEVENT (180 kb/ev) ⇒ CDST (18 kb/ev): */acs/data/YY/dst1/*   (+ INDEX files)
Special streams are stored in respective */acs* directories
LOOK L4/5 histos: */acs/data/l4histos/*

Use Fpack SELECT mechanism to access specific data type (records) and Run/Events subset

Documentation, Information, Help

- Web pages:

  ▷ Computing and Software Navigator: *https://www-h1.desy.de/icas/*

  ▷ Manuals for H1 software: *https://www-h1.desy.de/icas/imanuals/*

  ▷ Monte Carlo Mass Production: *https://www-h1.desy.de/imc/*

  ▷ H1 Hypernews forums:

      ∗ H1 Data

      ∗ Computing & Software

      ∗ Bugs, Problems, Frustrations, Fixes

- Unix tools:

  ▷ **lhb** $[-t]$ *bank_name* (ex: "lhb head", "lhb -t OSVX")

  ▷ **man** *package* (works for BOS,LOOK,FPACK,NDB,DATMAN,H1LOOK)

  ▷ **cartl** − H1 Data Information Service (fi les, runs, NDB banks etc.)

Software Manuals

H1 Software

- [BOS manual](#): The H1 Bank Operating System
- [CVS within H1](#)
- [FPACK](#) : Data handling I/O package *(Vsn 0.89/00 Sep.1994)*, [FPACK](#) *(Vsn 1.00/00 Dec.1998)*,
- H1 Data Definition Language: [H1DDL](#) and [DATMAN](#) (H1sw #008)
- [H1ECLASS](#): Event classification (on L5, obsolete >=1998)
- [H1ED](#) : Event display (.ps and .html version)
- [H1RED](#): RooT Event display
- [H1LOOK](#) : General purpose event display
- [LOOK](#) : Graphics system
- **[H1PHAN](#)**: PHysics ANalysis library, containing among others
  - [H1TOX](#): Analysis toolbox
  - [QBGFMAR](#)
  - [QHQTRK](#) (Heavy Flavour track selection code by Lee West)
  - ...
- [H1SIM](#): Detector Simulation based on GEANT3 (geant user's guide: v3.15([ps](#)) , v3.21([ps](#) , [pdf](#) , [html](#)))
- [H1REC](#): Reconstruction program (vsn 9)
- [H1UNIX](#): H1 SW package handling and installation tools
  - Update of [H1UNIX for CVS and H1OO](#)
- [ZUBR](#) : Display L4-histograms (obsolete tool: [KOOL](#))
- **H1 Database:** [NDB](#): H1 Database Software Package
- List of all [H1-Software Notes](#)
  H1OO: [The Object Oriented Physics Analysis Project](#)

Summary

- Thanks to the good design basic H1 software successfully survived 20 years, Moore's law and several computing revolutions

- Most of the problematic areas are due to violations of our own rules and specifi cations, as well as due to the lack of qualifi ed manpower

- Last effort is still needed to achieve adequate quality for the fi nal round of ultimate precision HERA analyses