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)
- **Code managament**: 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 \( \Rightarrow \) 25Hz, Data volume: 2 \( \Rightarrow \) 50Tb/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**
Lesson number 1: Take software seriously

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 extension 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 sufficiently commented source code.
  ...

...
Overview of the H1 software packages

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

**Code**

- H1LOOK
- H1REC
- H1L4
- H1PHAN
- H1ECLASS
- H1QT
- H1SIM/GEANT
- H1TRIG
- H1UTIL
- H1NDB
- DATMAN
- BOS
- FPACK
- LOOK

**Text + Help files**

- H1TEXT
- 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

<table>
<thead>
<tr>
<th>Package</th>
<th>Description</th>
<th>Authors/Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOS</td>
<td>Dynamic data and memory management system</td>
<td>(V.Blobel, 1985)</td>
</tr>
<tr>
<td>FPACK</td>
<td>Machine-independent data handling I/O package</td>
<td>(V.Blobel, 1991)</td>
</tr>
<tr>
<td>LOOK</td>
<td>General system for graphics applications in physics</td>
<td>(V.Blobel, 1990)</td>
</tr>
<tr>
<td>H1NDB</td>
<td>H1 database software package (interface to Oracle)</td>
<td>(L.Criegee, C.Kleinwort)</td>
</tr>
<tr>
<td>H1UTIL</td>
<td>Collection of utility functions, used in H1 application packages</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>DATMAN</td>
<td>A Data Management Tool package (used mainly in H1ED)</td>
<td>(U.Leberton et al.)</td>
</tr>
<tr>
<td>H1QT</td>
<td>Q – t analysis code for CT, FT and Fμ (witten in C)</td>
<td>(D. Sankey et al.)</td>
</tr>
<tr>
<td>H1SIM</td>
<td>H1 detector simulation based on GEANT3</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>H1TRIG</td>
<td>H1 trigger simulation package</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>H1REC</td>
<td>H1 reconstruction package</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>H1L4</td>
<td>H1 Level 4 and Event classification software</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>H1ECLASS</td>
<td>Old (frozen) Level 5 Event classification package</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>H1PHAN</td>
<td>H1 Physics Analysis library (Fortran based)</td>
<td>(H1 Collab.)</td>
</tr>
<tr>
<td>H1LOOK</td>
<td>H1 Event Display (frozen)</td>
<td>(U.Berthon, S.Levonian, Th.Naumann)</td>
</tr>
</tbody>
</table>

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 fixing 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 CJC; SOSC
  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**)
- 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 HERA2, 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

<table>
<thead>
<tr>
<th>Row</th>
<th>H1REC8</th>
<th>H1REC9</th>
<th>H1REC</th>
<th>H1PHAN</th>
<th>H1L4</th>
<th>H1CLASS</th>
<th>H1SIM</th>
<th>H1UTIL</th>
<th>H1TRIG</th>
<th>H1LOOK</th>
<th>H1BSTREC</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>-</td>
<td>-</td>
<td>96100</td>
<td>31200</td>
<td>63200</td>
<td>20700</td>
<td>35400</td>
<td>63304</td>
<td>18300</td>
<td>-</td>
<td>10800 !test</td>
</tr>
<tr>
<td>30</td>
<td>-</td>
<td>-</td>
<td>96000</td>
<td>31200</td>
<td>63000</td>
<td>20700</td>
<td>35400</td>
<td>63300</td>
<td>18000</td>
<td>-</td>
<td>10500 !new !hera2_06a</td>
</tr>
<tr>
<td>29</td>
<td>-</td>
<td>-</td>
<td>95900</td>
<td>31200</td>
<td>62800</td>
<td>20600</td>
<td>34200</td>
<td>63200</td>
<td>17000</td>
<td>-</td>
<td>10400 !pro !hera2_05b</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>-</td>
<td>90726</td>
<td>31100</td>
<td>50800</td>
<td>20600</td>
<td>33900</td>
<td>63000</td>
<td>12900</td>
<td>11907</td>
<td>-         !hera1_05c</td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>-</td>
<td>94700</td>
<td>31200</td>
<td>62600</td>
<td>20600</td>
<td>33900</td>
<td>63000</td>
<td>12900</td>
<td>-</td>
<td>10400 !hera2_05a</td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>-</td>
<td>90723</td>
<td>31100</td>
<td>50800</td>
<td>20600</td>
<td>33800</td>
<td>62800</td>
<td>12900</td>
<td>11907</td>
<td>-         !hera1_05b</td>
</tr>
<tr>
<td>25</td>
<td>-</td>
<td>-</td>
<td>90723</td>
<td>31100</td>
<td>50800</td>
<td>20600</td>
<td>33800</td>
<td>62800</td>
<td>12900</td>
<td>11907</td>
<td>-         !hera1_05a</td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>-</td>
<td>94400</td>
<td>31100</td>
<td>62500</td>
<td>20600</td>
<td>33800</td>
<td>62800</td>
<td>12900</td>
<td>11907</td>
<td>-         !hera1_05a</td>
</tr>
<tr>
<td>23</td>
<td>-</td>
<td>-</td>
<td>94300</td>
<td>31100</td>
<td>62400</td>
<td>20600</td>
<td>33700</td>
<td>62700</td>
<td>12800</td>
<td>-</td>
<td>10400 !hera2_04b</td>
</tr>
<tr>
<td>22</td>
<td>-</td>
<td>-</td>
<td>94201</td>
<td>31002</td>
<td>62200</td>
<td>20600</td>
<td>33400</td>
<td>62300</td>
<td>12500</td>
<td>-</td>
<td>10300 !hera2_04a</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>-</td>
<td>90722</td>
<td>31003</td>
<td>50800</td>
<td>20600</td>
<td>33101</td>
<td>62200</td>
<td>10112</td>
<td>11907</td>
<td>-         !hera1_04</td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>-</td>
<td>93000</td>
<td>31002</td>
<td>61100</td>
<td>20600</td>
<td>33101</td>
<td>62000</td>
<td>12000</td>
<td>-</td>
<td>10100 !hera2_03</td>
</tr>
<tr>
<td>19</td>
<td>-</td>
<td>-</td>
<td>90718</td>
<td>30800</td>
<td>50800</td>
<td>20506</td>
<td>32800</td>
<td>40926</td>
<td>10112</td>
<td>11907</td>
<td>-         !hera1_03 !mchera1</td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>90605</td>
<td>-</td>
<td>30110</td>
<td>50611</td>
<td>20506</td>
<td>31003</td>
<td>40924</td>
<td>10112</td>
<td>11907</td>
<td>-         !hera1_02</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>90601</td>
<td>-</td>
<td>30109</td>
<td>50610</td>
<td>20506</td>
<td>30924</td>
<td>40921</td>
<td>10108</td>
<td>11900</td>
<td>-         !repro00</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>90527</td>
<td>-</td>
<td>30107</td>
<td>50607</td>
<td>20506</td>
<td>30924</td>
<td>40914</td>
<td>10108</td>
<td>11900</td>
<td>-         !mc99_00</td>
</tr>
<tr>
<td>14</td>
<td>-</td>
<td>90421</td>
<td>-</td>
<td>30106</td>
<td>50600</td>
<td>20506</td>
<td>30921</td>
<td>40913</td>
<td>10108</td>
<td>11900</td>
<td>-         !00c</td>
</tr>
<tr>
<td>13</td>
<td>81405</td>
<td>90413</td>
<td>-</td>
<td>30105</td>
<td>50600</td>
<td>20506</td>
<td>30919</td>
<td>40912</td>
<td>10104</td>
<td>11801</td>
<td>-         !00b</td>
</tr>
<tr>
<td>12</td>
<td>81404</td>
<td>90411</td>
<td>-</td>
<td>30104</td>
<td>50600</td>
<td>20506</td>
<td>30918</td>
<td>40911</td>
<td>10101</td>
<td>11800</td>
<td>-         !00a !mc96_97</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>90407</td>
<td>-</td>
<td>30102</td>
<td>50501</td>
<td>20505</td>
<td>30912</td>
<td>40910</td>
<td>10008</td>
<td>11705</td>
<td>-         !99a</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>90303</td>
<td>-</td>
<td>30004</td>
<td>50400</td>
<td>20505</td>
<td>30906</td>
<td>40906</td>
<td>10003</td>
<td>11704</td>
<td>-         !repro98</td>
</tr>
<tr>
<td>7</td>
<td>81403</td>
<td>-</td>
<td>-</td>
<td>30004</td>
<td>50204</td>
<td>20505</td>
<td>30906</td>
<td>40906</td>
<td>10003</td>
<td>11704</td>
<td>-         !old96_97</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>90106</td>
<td>-</td>
<td>30033</td>
<td>50204</td>
<td>20505</td>
<td>30905</td>
<td>40905</td>
<td>10003</td>
<td>11703</td>
<td>-         !98b</td>
</tr>
<tr>
<td>5</td>
<td>81400</td>
<td>-</td>
<td>-</td>
<td>30022</td>
<td>50204</td>
<td>20505</td>
<td>30724</td>
<td>40725</td>
<td>-</td>
<td>11700</td>
<td>-         !98a</td>
</tr>
<tr>
<td>4</td>
<td>81103</td>
<td>-</td>
<td>-</td>
<td>30000</td>
<td>50204</td>
<td>20504</td>
<td>30713</td>
<td>40716</td>
<td>-</td>
<td>11612</td>
<td>-         !mc95 !97b</td>
</tr>
<tr>
<td>3</td>
<td>81103</td>
<td>-</td>
<td>-</td>
<td>20311</td>
<td>50204</td>
<td>20504</td>
<td>30713</td>
<td>40716</td>
<td>-</td>
<td>11612</td>
<td>-         !keep</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>10908</td>
<td>40317</td>
<td>-</td>
<td>30709</td>
<td>40710</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-         !mc94</td>
</tr>
<tr>
<td>1</td>
<td>80702</td>
<td>-</td>
<td>-</td>
<td>20311</td>
<td>-</td>
<td>20405</td>
<td>30648</td>
<td>40628</td>
<td>-</td>
<td>-</td>
<td>-         !oldvers</td>
</tr>
</tbody>
</table>

**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 fix
  - ??-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 fix 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 are sets of BOS banks combined and written as Fpack records to different streams

H1 CDAQ

L4/5 & data logging

ep physics

RUNEVENT

HOTLINE

HALOMUON

L4CHECK

monitors

RANDOM

LUMIMON

H1COSMO

calibration

RUNDATA

SPACLED

SPACKIN

L4 histos

H1 raw data files \equiv \text{Sum of all records (except LOOK histos): } /acs/data/YY/rawd/

RUNEVENT (180 kb/ev) \Rightarrow \text{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/](https://www-h1.desy.de/icas/)
  - **Manuals for H1 software:** [https://www-h1.desy.de/icas/imanuals/](https://www-h1.desy.de/icas/imanuals/)
  - **Monte Carlo Mass Production:** [https://www-h1.desy.de/imc/](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 (files, runs, NDB banks etc.)
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 specifications, as well as due to the lack of qualified manpower.

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