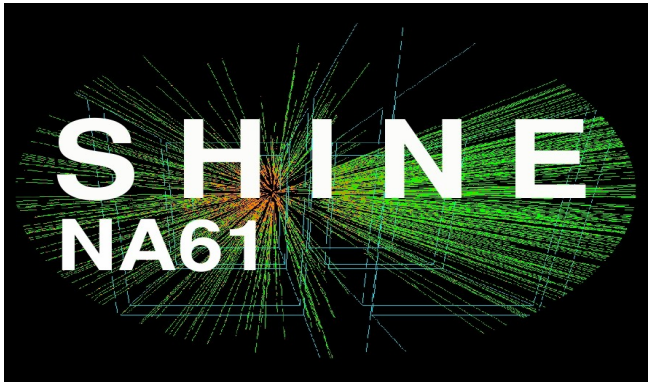
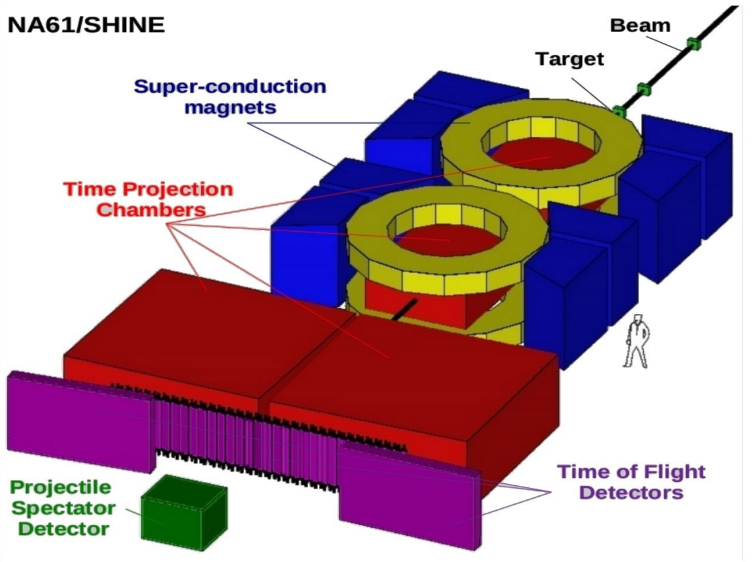
Overview of the NA61 software upgrade proposal

András László



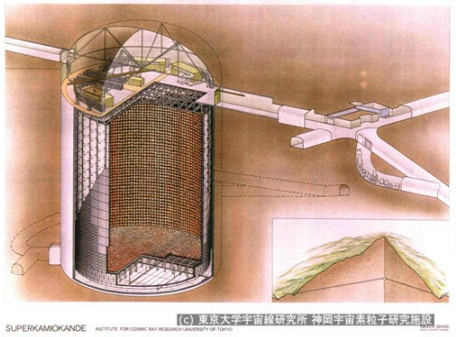
Outline

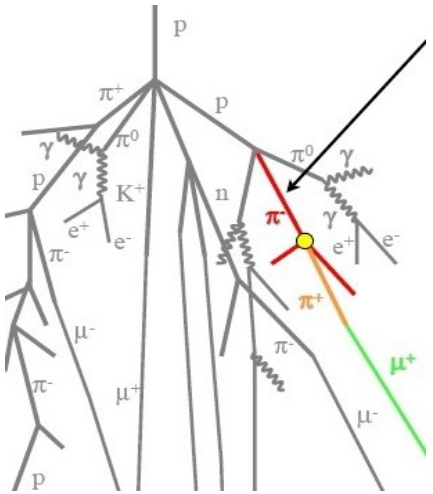
- Introduction
- Motivation for upgrading off-line software
- Outline of the upgraded software
- Upgrade strategy
- Impact on reconstruction
- Impact on simulation
- Impact on analysis
- Relation to virtualization project
- Summary

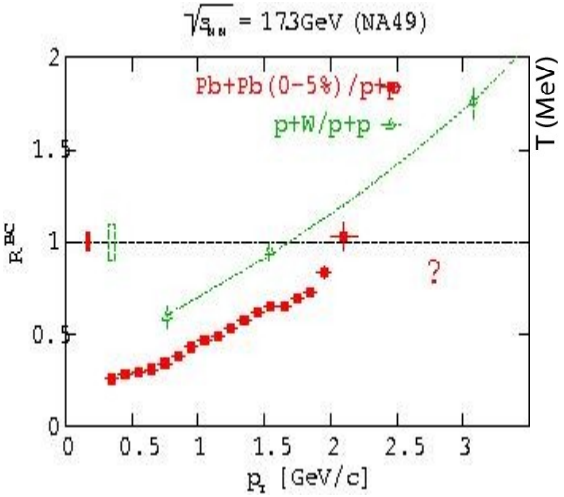


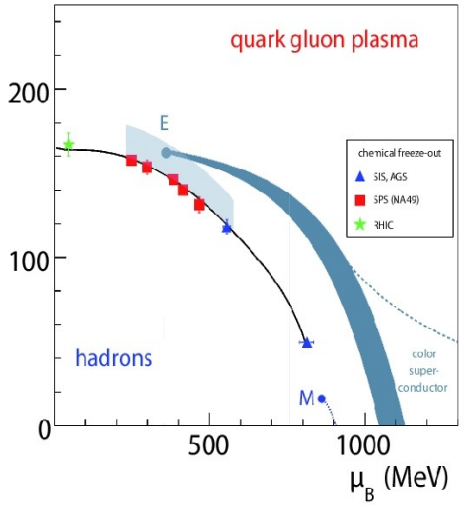
Introduction

- NA61 is a large acceptance hadron spectrometer experiment. Main components of tracking system and software inherited from NA49.
- Main physics goals are to measure:
  - Reference hadron spectra in p+C for T2K experiment.
  - Reference hadron spectra in π^-+C for the Pierre Auger Observatory.
  - High p_T physics in p+p, p+Pb.
  - Fluctuactions in ion-ion for CP&OoD in strong interactions.

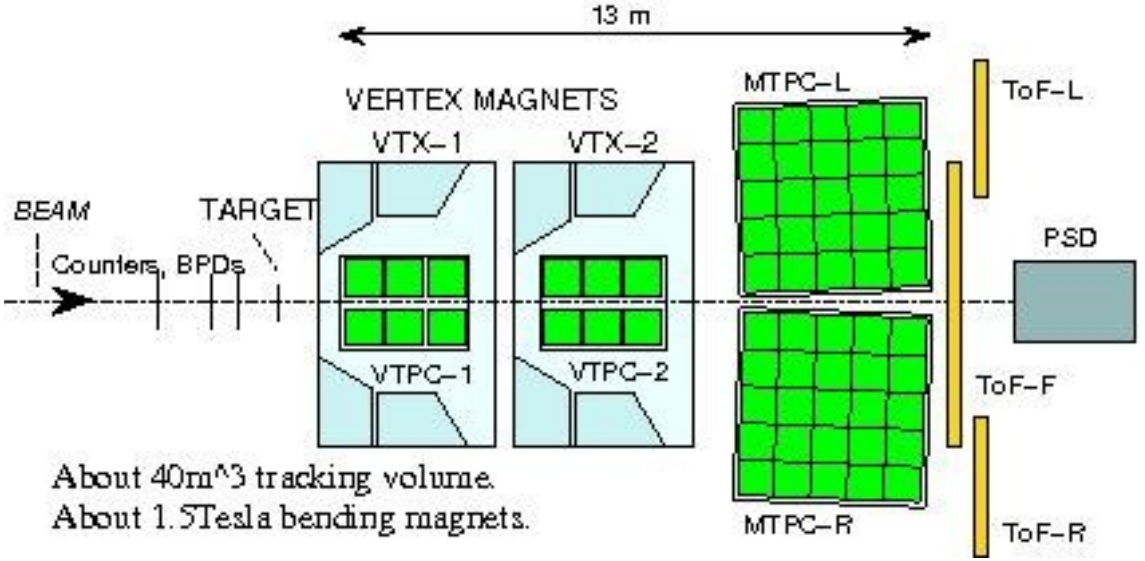




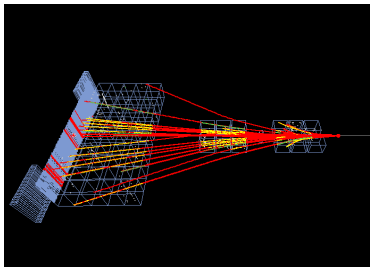




- Main detector components: 4 large volume TPCs (40 m^3), 2 superconducting bending magnets (1.5 Tesla), TOF system (60 psec time resolution), projectile spectator detector (1 nucleon precision, in preparation).



- Full black event: 100 MByte. Event with zero compression: 4 MByte.
- Event rate: 80 Hz. In time average (due to accelerator): 20 Hz.
- Maximum about 7 TByte/day raw data, for 4-5 months/year.



Data taking

- 2007: pilot physics data for neutrino experiments.
- 2009: 40M physics events for neutrinos, cosmic rays, high p_T.
- 2010: 55M physics events for neutrinos and high p_T.
- 2011-2014 (at least): still about 100M physics events, mainly for high p_T and ion-program.

Development

- Several planned hardware upgrades (PSD, LMPD).
- There is demand for algorithmic upgrades (Kalman filter, dE/dx).
- Data analysis/simulation expected until 2020 (at least).

=> Continued development and developable software is needed.

(Up to last year: minimalistic plan, just preserve oldsoftware by Virtualization.)

- After development phase: software environment can be preserved by

CernVM
SoFTware Appliance

Present off-line software

- Inherited from NA49, written in early 90-s, based on DSPACK server concept.
- Reconstruction chain based on sequence of DSPACK client processes written in non-standard FORTRAN, C and C++, steered by a quite extended system of shell-scripts (multi-language paradigm). About 15 key clients.
- Calibration factors accessed by call to MYSQL data base, returning file names of DSPACK files, containing the calibration information.
- Simulation written in GEANT3.
- Explicit modularity (each module is a process: „client”).
- Memory efficiency for weak computers as well (DSPACK, SHM).
- Current scheme is pipelined to some extent (in some parts).
- Complete source tree of simulation+reconstruction is about 0.4 GByte.
- Memory requirement is rather low for simulation and reconstruction (<0.2 GByte /job).
- CPU consumption is also not very demanding (2.8 sec wall clock time / p+p event).
- *Thanks to the developers (Predrag Buncic et al.): it still runs, is usable and is highly efficient from the performance point of view after 15 years!*

Motivation to upgrade off-line software

- Several concurrent data formats (ROOT, DSPACK).
- Different programming languages (C++, C, FORTRAN, shell).
- Obsolete parts in data model. Cleanup is difficult: difficult to follow data dependencies of clients.
- Extra difficulties due to multi-process approach (SHM, IPC, no gdb). Propagation of info in steering routine is difficult to trace.
- Simulation in GEANT3 (FORTRAN, obsolete hadronic generators, not supported by CERN anymore).
- No automated testing.
- Documentation is weak and largely out of date. Support basically not available.
- Difficult to develop (GTPC, new raw data format, future detector and algorithmic developments).

Outline of the upgraded software

- Reconstruction chain as single process, written in C++, steered by XML config file system. Idea based on software of Pierre Auger Observatory.
- Unified data model: based on STL containers, ROOT streaming. Prototype already exists and used: NDST. Idea based on Auger. Should be able to read/write old DSPACK as well for transition.
- Modularity: collection of processing modules, sequenced by a steering routine via XML instructions (later python?). Info within event data model and detector description. Communication via event store interface.
- Client wrapper modules will be introduced for the transition period. These can interface with the event model via the DSPACK read/write capability.
- Simulation in GEANT4.
- Automated testing.
- Semi-automated documentation.

Upgrade strategy

- First, the new event data model has to be established. Translation capability from/to old DSPACK data model necessary for transition. *(to be done by the core group, time estimate: 2-3 months)*
- Secondly, the steering environment and module template has to be developed. *(to be done by the core group, time estimate: 1-2 months)*
- A special module, the client wrapper module has to be developed to be able to use old DSPACK clients in the new framework. Technically, this means wrapping of fixed signature C and FORTRAN routines. *(to be done by the core group, time estimate: 1 months)*
- At this point the new framework will be able to run, powered by the old clients, and the system can be commissioned.
- Gradual upgrades of the modules. *(to be done gradually, on the basis of upgrade demands, by possible new developers as well)*
- External libraries: STL, ROOT and DSPACK (until phaseout).

Impact on reconstruction

- New detectors are planned to be added to the experiment (still GTPC, PSD, LMPD, possibly vertex detector). Easy extendibility is therefore required.
- There is demand for algorithmic upgrades (Kalman filter based tracking, maximum likelihood based dE/dx).
- Expertise is available: both on detector and algorithmic side.
- Calibration automation could be achieved much easier with an easy-to-develop system.

Impact on simulation

- Phasing out of GEANT3 based simulation chain will improve developability of event simulation (now largely undocumented mixture of C and non-standard FORTRAN).
- Some models in GEANT3 are outdated (see MC talk). Especially problem for PSD simulation and in beam mode or at higher energies.
- Some important parts of the present simulation chain is not fully operational. This problem can be eliminated during the upgrade.

Impact on analysis

- Analysis, QA and reconstruction can be done on the very same data model.
- Analysis people can easily investigate reconstruction level details, whenever becomes necessary.
- Support for such an analysis platform is strong within the collaboration: already in use within the NA61 analyzers.
- Existence of documentation will greatly help the end-users.

Relation to virtualization project

- Common PH-SFT and NA61 project on software virtualization is ongoing. (1 project associate)
- Originally, Virtualization was intended as a minimalistic solution for software preservation.
- The upgraded software is expected to have relatively moderate dependencies: C++, STL, ROOT and DSPACK (during transition).
- Nevertheless, on the long term, Virtualization can ensure data and software preservation of the upgraded software, since on the long term even the C++ libraries can change significantly.
- Software environment preservation becomes especially necessary after the finishing of development phase.
- Virtualization is planned to be done via CernVM SoFTware Appliance

Summary

- A strategy for upgrading the NA61 off-line software is proposed.
- Based on the software competence in NA61 and Pierr Auger Observatory.
- Three step upgrade is proposed:
  - establish data model, steering routine and module template,
  - create wrapper module to be able to use old modules (clients),
  - gradually upgrade modules on demand.
- Would be great improvement for the developers of reconstruction, for automation of calibration, and for analysis people as well.
- The upgraded software is planned to be preserved after the development phase using the CernVM virtual machine.
- *For efficient start of the project, support for 2 Technical Students is asked from CERN.*
- *Advice and collaboration of PH-SFT is highly appreciated.*