Opportunities offered by LEP data@EDM4hep for future EW and Higgs factories

3rd DPHEP Collaboration Workshop

June 23, 2021

G Ganis
CERN-EP

Outline

- What can be still done with LEP data?
  - Non-exhaustive list of analysis / optimizations which could be done with LEP data
  - Role in the preparation of next EW/Higgs factories
- EDM4hep / Key4hep reminder
  - Main ideas, status
- LEP data @ EDM4hep
  - Is it feasible?

What can be still done with LEP data? Some examples

- Confirmation/verification of (HL-)LHC discoveries or 3-sigma excess
  - Possible associated effects visible at LEP
- Optimization / generalization of analysis techniques applied only partially
  - E.g. event rotation technique for acceptance determination for Z→hadrons
    - 2-3 order of magnitude more powerful statistically than straight MC
    - Can be generalized to leptons in the context of future EW factories and improve LEP results too
- New analysis ideas and/or tools coming from future experiments studies
  - E.g. Machine learning techniques for Heavy Flavour observables, tau physics
    - Tuning of algorithms and possibly improved LEP measurements (or even new, e.g. mass)
- New analysis
  - E.g. LEP detectors not fully exploited in searching for feebly interacting particles
    - Optimized tools may allow to include channels difficult for LHC, e.g. those w/ taus
- Validation and Tuning of new version of hadronization tools
  - E.g. Pythia8, Herwig7, Sherpa … require genuine e+e- data sets (W. Kilian@ECFA-kickoff)

How can we use LEP data today?

- Need approval from collaboration
  - And typically a member of the collaboration on board
- Need access to the *preserved* software chain
  - Not uniform (anymore)
    - Requires VM or container for SLC6 (ALEPH, OPAL); DELPHI can run on lxplus
  - Requires experiment expertise
- Typical approach
  - Interesting observables extracted to text files and converted into ROOT TTrees
  - It works and detaches from experiment specifics
    - ROOT is de-facto a cross-experiment standard
  - But lacks generality
- Can this be generalized?

Key4hep: Turnkey Software Stack

- Projects rely/benefit more and more on/from well structured software
  - Frameworks gluing together differently customized tools for specific tasks
  - Scattered but recurring landscape with (often) lots of similarities
- Idea: develop a *common low maintenance, customizable, turnkey software stack*
  - Consensus reached among ILC, CLIC, FCC, CEPC, SCTF (and also LHC) communities
  - Bologna 2019, Hong Kong 2020; under the HSF umbrella
  - Part of the CERN EP R&D initiative and AIDAInnova; fostered by ECFA
  - Actively developed: key4hep GitHub project, doc
- Key4hep design goals: easy to use, set up, deploy, extend
  - One framework (Gaudi), based on community developed common tools (Podio, DD4hep, ...)
  - Connection to all relevant HEP (ROOT, Geant4, ...) and external packages (Boost, ...)
- Organic interoperability requires common Event Data Model
  - For transient and persistent storage

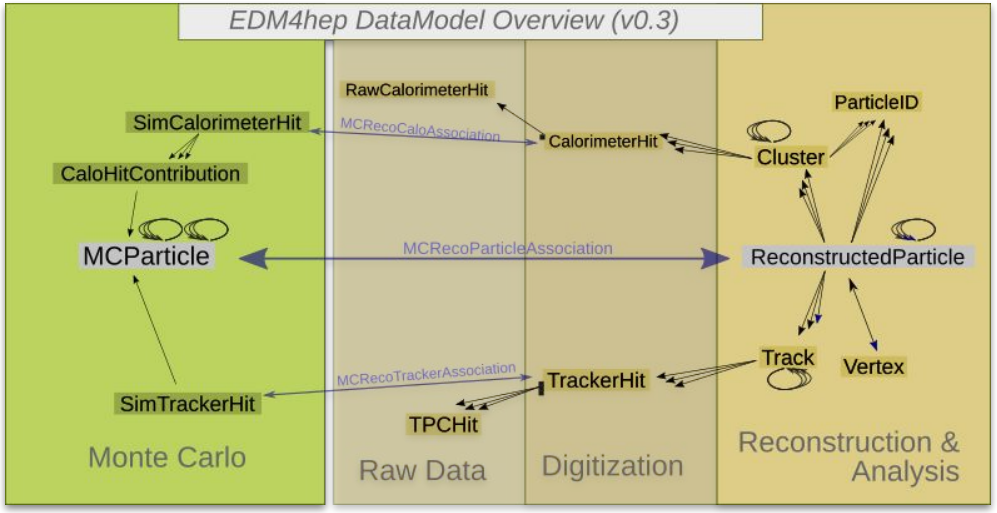
EDM4hep: the common language

EDM provides common language for exchange among framework components

- Heavily inspired by LCIO and FCC-edm
- Generated with PoDIO: detach definition from backend (ROOT as default)

Challenges:

- Same for all HEP experiments?
- Efficient support for detector needs
  - Interaction w/ detector teams from the start
    - Eg. cluster counting for IDEA Drift Chamber
  - Applies also for “old” experiments



T Madlener, vCHEP 2021

Why LEP data @ EDM4hep?

- Key4hep / EDM4hep: framework with longer perspective than a single experiment
  - Not just *another data format*, but one that might become a standard
    - For data structures and file format
      - ROOT (compatible) default today, technology evolve as needed
- Helps achieving the FAIR data principles by improving on
  - Accessibility, detaching from library and OS obsolency
  - Interoperability, single standard framework
  - Re-usability, less specific expertise
- Requires “migration”, which may be a pain
  - But return-on-investment may be large
  - For LEP data, interest from EW/Higgs factories may provide a unique opportunity

Is it feasible?

Conversion of existing data

- Target is at least *level 3 equivalent*
  - *“perform complete analyses when the existing detector reconstruction and simulated data sets are adequate for the pursued goal”*
  - Convert and store objects such as vertex/tracks w/ covariances, reconstructed calo objects, ...
    - Adaptation of existing data structures might be required
  - Level 4 (full chain with hits ...) may require more modification/expansions or new data structures
    - Similar work of what is done with detector concept developers
- Approach: intermediate XML-like file, read by a dedicated EDM4hep converter
  - Fully automatizable

New Monte Carlo productions

- Feed the existing workflow {VM, container} through shared format (e.g. HepMC)
  - Delphes-like approach possible

Concluding remarks

- At least for ALEPH, early investigations for migrating to EDM4hep promising
  - Either ALPHA (Fortran) or even ALPHA++ provides the relevant code for migration
  - Need to investigate more DELPHI and OPAL
- In general, lot of interest from future EW/Higgs factories communities. If the LEP data were migrated to EDM4hep they would add a great value!
  - Allow the same algorithms and techniques to be applied on simulated and real data
  - Give more realistic evaluations of the techniques under study
  - Provide a powerful training basis to the youngsters
  - Opportunity for future (EW/Higgs factories) students for thesis / publications with real data
    - Essential to get a job in HEP!
- Several LEP *experts* involved in EW/Higgs factories studies
  - Unique opportunity, for LEP data to get better preservation perspectives, for future factories to get access to still useful data

- After all, *the best Monte Carlo is the data!* (J.F. Grivaz, 1989)