Outline

- Some introductory slides
- Monte Carlo Generators
- MC production at FCCee
- Concluding remarks

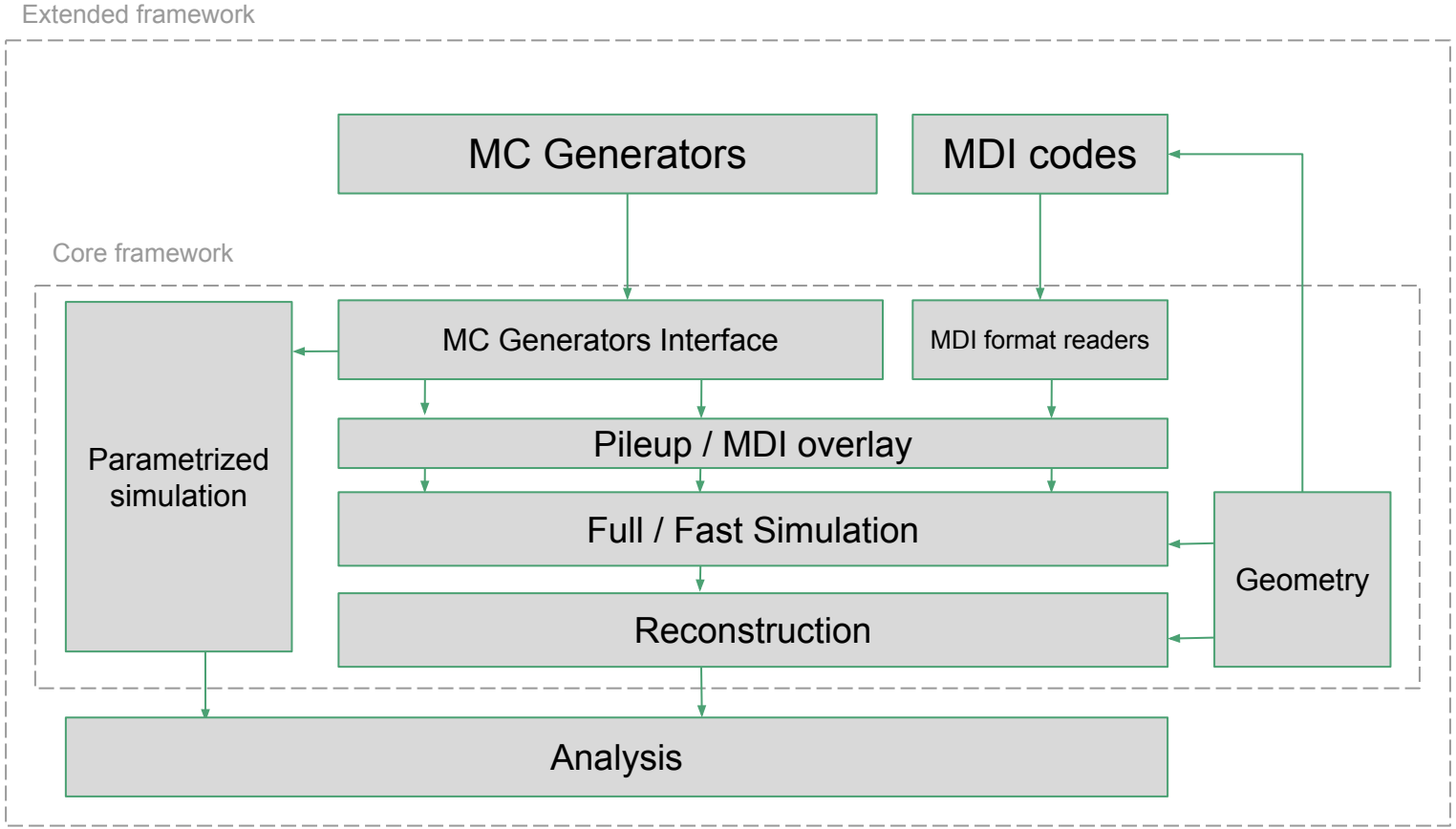
Production Experience FCC

ECFA Higgs Factories: 1st Topical Meeting on Generators

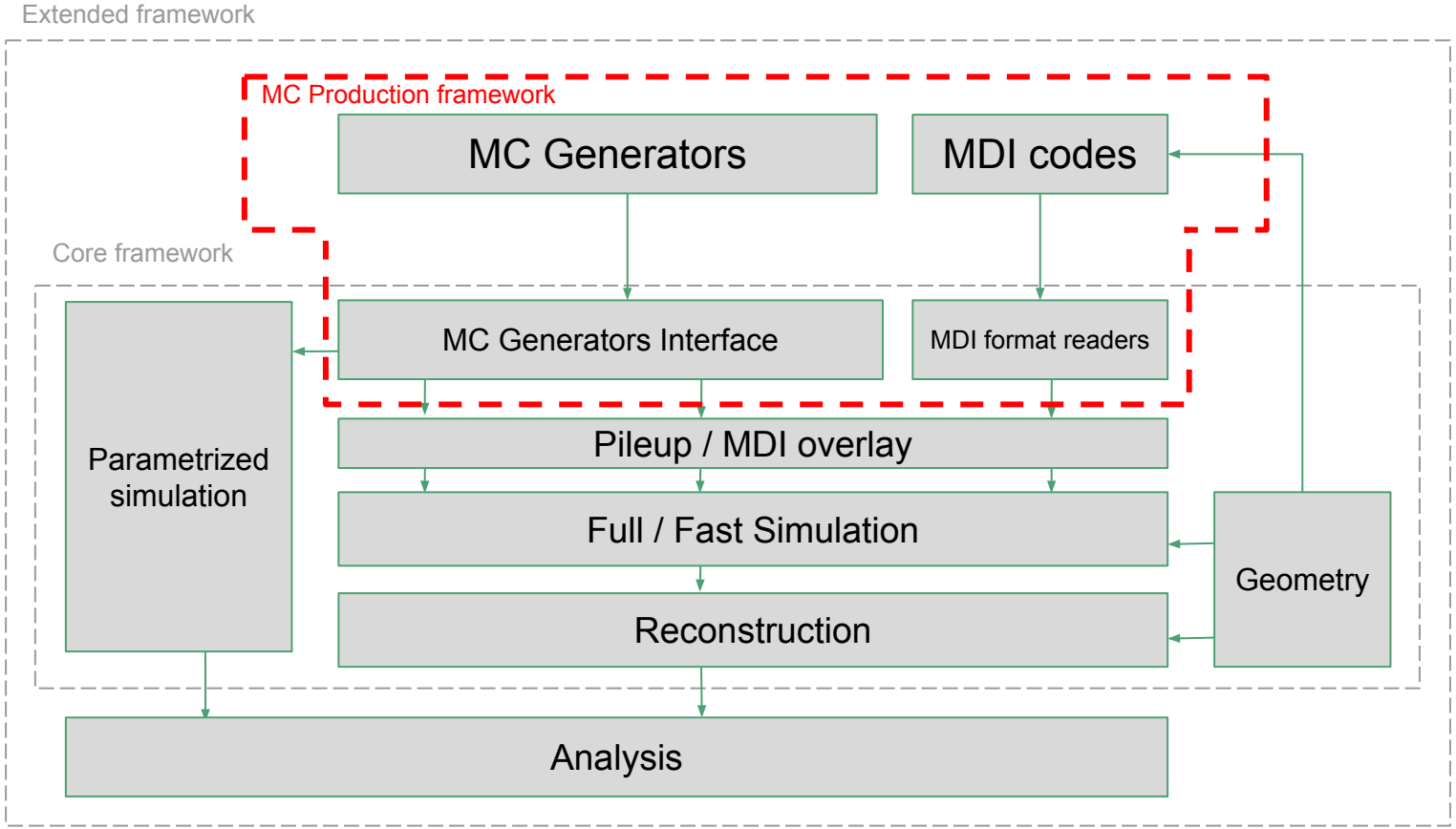
November 10, 2021
C. Helsens (with input from G. Ganis and E. Perez)
CERN-EP

Some introductory slides

Typical workflows to support



Typical workflows to support



The role of the framework

- Provide uniform view on the components
  - Common configuration, log and error reporting, plug-in management, ...

- A good framework adapts to varying landscape to always provide optimal interoperability and use of resources. Today this means:
  - Solid multi-thread support, ability to cope with heterogeneous resources in terms of
    - Different hardware (GPU, FPGA); segregation level (HPC network restrictions, ...); cloud protocols

- HEP experiments have a tendency to start from scratch
  - AliROOT, O2 (ALICE); CMSSW FWCore (CMS); Gaudi (LHCb, ATLAS, HARP, BESIII, Minerva, Daya Bay)

- Adopting a demonstrated solution matching the (projected) needs buys in the experience and future evolution

The role of the framework

Continue and strengthen the outstanding common work started ~2 years ago

The common software vision

Create a software ecosystem integrating in optimal way various software components to provide a ready-to-use full-fledged data processing solution for HEP experiments

Complete set of tools for

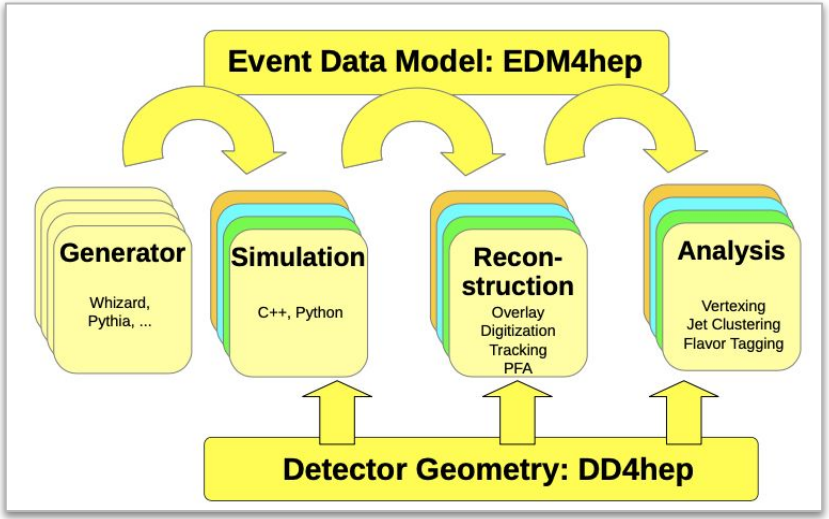
- Generation, simulation, reconstruction, analysis
- Build, package, test, deploy, run

Core Ingredients of current key4hep

- PoDIO for EDM4hep, based on LCIO and FCC-edm
- Gaudi framework, devel/used for (HL-)LHC
- DD4hep for geometry, adopted at LHC
- Spack package manager, lot of interest from LHC

Community project, unifying efforts

- Contributions from CLIC, ILC, FCC, CEPC
- And interest from STCF, muon collider, ...



Kick-off meetings in Bologna, Hong Kong

The common software vision

Extend the common software vision to a common generator vision?

Not within the current scope of Key4Hep, but would be great to develop in k4Gen

Monte Carlo Generators

Monte Carlo Generators

- Recent review of tools for e^+e^- at the ECFA kick-off (W. Kilian)
- Legacy LEP generators still state of the art at the Z and WW energies
  - "Archeological" work to recover KKMC, BHLUMI, BabaYaga, KoralWW, ...
- Main software challenges
  - Interfacing with the framework
    - Typically standalone apps w/ output in common data format, e.g. HepMC, LHEF
      - In rare cases they provide a callable interface, e.g. Pythia8
  - Availability in the shared software stacks
    - Private codes, unversioned tarballs, … version control issues
- Other requirements relevant for FCCee
  - Uniform treatment of beam parameters (beam energy spread, crossing angle)
  - For generation, possibly better achieved at framework level
    - Generate at nominal √s and zero crossing angle apply same boost before simulation

Monte Carlo Generators

Now is a unique opportunity to have ambitious plans for the future and develop common/generic generator interfaces output format suitable for both hadron and lepton machines

Beam and MDI-related backgrounds

- Several processes and codes, including
  - (In)coherent pair creation — GuineaPig
  - Synchtron Radiation — MDISim, SynRad, Sync_Bkg
  - Radiative bhabhas — GuineaPig, BBBrem, BHLUMI, Whizard
  - γγ→hadrons — GuineaPig + Pythia

  Codes not always in public repositories, outputs in different, non-standard formats
- Optimization of the FCCee interaction region design requires deep level of understanding of the detector backgrounds
  - Only achievable with integration in experiment software
- Framework integration to unify/simplify access of each relevant codes
  - Supercode as glue with controlled configuration and normalization; on going effort
- Consistent description of the relevant geometry elements
  - Requires interplay between detector and machine geometry formats (e.g. CAD)

Beam and MDI-related backgrounds

Now is a an other unique opportunity to have ambitious plans for the future and develop common/generic MDI interfaces and output format suitable for both hadron and lepton machines

MC production at FCC

Resources available at CERN for FCC so far

- Storage (eos)
  - 500 TB, fully backed-up, as of today 68% filled
- Computing
  - 3000 HS06 inside the condor batch system

For “free”

Resources available at CERN for FCC so far

- Storage (eos)
  - 500 TB, fully backed-up, as of today 68% filled
- Computing
  - 3000 HS06 inside the condor batch system

For “free”

- And tomorrow?
  - 1 PB of EOS storage
    - period 2021-2024: 44 KCHF (11kCHF/y)
    - Extension 2025-2028: 22kCHF
  - 9000 HS06 HTCondor
    - for the period 2021-2024: 32 kCHF (8 KCHF/y
    - Extension 2025-2028: 15 KCHF.
  - moving data to tape is not a problem, 200 TB would be about 2kCHF, thus negligible

Scale of Monte Carlo productions done

- "Spring2021" production, EDM4hep
  - Delphes events IDEA with *Track Covariance full matrix lower triangle*
  - Total: ~10^{10} events, ~53 TB, mostly at Z peak

- Existing productions for FCC-hh (100 TeV)
  - EDM4hep
    - Total: ~1.3×10^{8} events, ~14 TB
  - Old fcc-edm
    - Delphes: ~2.1×10^{9} events, ~93 TB
    - Full sim, single particle: ~$2.4 \cdot 10^{9}$ events, ~160 TB

- Stored at CERN on EOS
  - /eos/experiment/fcc/{ee,hh}/generation/DelphesEvents/

What we have used so far to produce those events

EventProducer - 1

https://github.com/HEP-FCC/EventProducer
https://github.com/HEP-FCC/FCC-config

- Python code to produce large amount of events using CERN ressources
  - Was running on lxbatch, migrated to HTCondor
  - Always eos as a storage element
- Developed by CH for FCC-hh (with contributions from M. Selvaggi)
  - Adapted to FCC-ee when it was needed
  - E. Perez is actively contributing in developments, validation and production for FCC-ee
  - Only very few people with the expertise...
- What it does?
  - It can directly call some generators (madgraph, KKMCee) and produce LHE events
  - It can directly run Pythia using k4SimDelphes executables under the hood with or without EvtGen, directly or reading, LHE or STDHEP files (Whizard)
    https://github.com/key4hep/k4SimDelphes
  - Thus produces Delphes events in EDM4Hep format, but straightforward to move to integrated framework to support other types of simulations

EventProducer - 2

https://github.com/HEP-FCC/EventProducer
https://github.com/HEP-FCC/FCC-config

Typical command to run:

python bin/run.py --send --FCCee --reco -p p8_ee_Zbb_ecm91_EvtGen_Bd2MuMu -N 10 -n 100000 --type p8 --condor -q microcentury --version spring2021 --detector IDEA

This will send for FCCee using delphes for process Z->bb, B->mumu 10 jobs of 100kevt running pythia+EvtGen using HTCondor and the microcentury queue for the spring2021 production, using the IDEA detector

The database system is based on common branches between EventProducer and FCC-configs

Home | Contact | FCC-ee | Les Houches | STDHEP | IDEA | IDEA 3T | IDEA Full Silicon | Stat

Delphes FCCee Physic events spring 2021 production (IDEA with Track Covariance full matrix lower triangle)

bd2mumu

NO	NAME	NEVENTS	NWEIGHTS	NFILES	NBAD	NEOS	SIZE (GB)	OUTPUT PATH
108	p8_ee_Zbb_ecm91_EvtGen_Bd2MuMu	1,000,000	0	10	0	10	4.87	/eos/experiment/fcc/ee/generation /DelphesEvents/spring2021 /IDEA/p8_ee_Zbb_ecm91_EvtGen_Bd2MuMu/

What we will be using

From batch to distributed computing

- In-house system served well CERN-based productions
  - CDR , Spring 2021
- Inclusion of non-CERN resources desired
  - Main requirements: central file catalogue, replication, remote access
  - Major development for the in-house system
- Among the existing solutions, DIRAC looked the more appropriate
  - Complete: workload management, file catalogue
  - Used by many experiments: LHCb, Belle II, BES III, JUNO, ILC/CLIC, …
- iLCDirac: LC community instance
  - Already serving another VO (CALICE)
  - Extended to serve FCC VO
  - Good example of re-use of generic solution

FCC VO @ iLCDirac

- Re-activated the FCC VO
  - Existing but dormant (never used)
  - Standard registration procedure
- Associated CERN FCC resources to FCC VO
  - HTCondor and EOS area
- Activated FCC VO in iLCDirac
- Added steering applications of interest for FCC workflows
  - Increased flexibility of generic application
  - Dedicated Gaudi application
  - Dedicated application Delphes, KKMCee, …
- Defined a list of workflows to support (as initial target)
  1. MC + Delphes
  2. MC + Full Sim (CLD) + iLCSoft Rec @ MarlinWrapper
  3. ...

Thanks to G. Ganis, A. Sailer and A. Stano

FCC VO @ iLCDirac : distributed cat. example

```
lxplus:~$ source /cvmfs/clicdp.cern.ch/DIRACpreview/bashrc
lxplus:~$ dirac-proxy-init -g fcc_user

lxplus:~$ dirac-dms-lfn-replicas /fcc/user/g/ganis/edm4hep_test_output.root
LFN                                          StorageElement URL
============================================================
/fcc/user/g/ganis/edm4hep_test_output.root   CNAF-DISK       davs://xfer-archive.cr.cnaf.infn.it:8443/fcc/user/g/ganis/edm4hep_test_output.root
                                             CERN-DST-EOS    gsiftp://eospublicftp.cern.ch//eos/experiment/fcc/prod/fcc/user/g/ganis/edm4hep_test_output.root

lxplus:~$ source /cvmfs/sw.hsf.org/key4hep/setup.sh
lxplus:~$ root -l
root [0] TFile *f1 = TFile::Open("davs://xfer-archive.cr.cnaf.infn.it:8443/fcc/user/g/ganis/edm4hep_test_output.root")
(TFile *) 0x2ec28b0
root [1]
```

Concluding remarks

Possible issues

- When a Monte carlo takes into account the BES:
  - the resulting events, in the LHE format, are not understood by Pythia
  - Tweaks are needed to avoid crashes
  - example:
    - as E(beam) can be larger than the nominal energy because of the Gaussian BES, need to tell Pythia to ignore (E,p) conservation
  - But other tweaks are needed too
  - No complete picture yet, and not sure that the tweaks always lead to fully consistent events
  - Way-out used so far for BES-events generated by Whizard: let Whizard run the hadronisation, and write events in the STDHEP format instead of LHE

Outlook

- If the community is ready to adopt a common approach
  - Profit from this unique opportunity to define the needs for long term future
  - Have ambitious goals for common data formats
    - Generation
    - Simulation
    - Analysis
    - EDM4Hep is a good candidate, will be used by key4Hep community
  - Need strong commitment from key4Hep and generator authors
  - People working for FCC will be happy to be prime users

Thank you!

EventProducer - 2

https://github.com/HEP-FCC/EventProducer

- In more details:
  - Once files (lhe, stdhep, root) are added to eos AND the process is properly registered inside the configuration file, a cronjob will validate the files and register them in yaml files. Each physical file on eos has his counter yaml file with relevant informations (date, size, number of events, who produced it etc…)
  - Once all the individual files are checked, bad jobs/files are cleaned, the same cronjob will merge all the individual yaml files into a merge.yaml (one per process for a given production)
  - Once all the processes are checked, a web page will be created summarizing the status of the production
  - The database system is based on common branches between EventProducer and FCC-configs

FCC VO @ iLCDirac : status

- Workflow demonstration
  - Workflow 1 OK: can run, save and retrieve files
  - Workflow 2 not yet demonstrated
    - Pending fixes in CLD full sim in k4SimGeant4
- Storage organization
  - Started definition of metadata keywords
    - Based on ILC experience
- First integration of external site: CNAF (storage only)
  - Can replicate files between CERN-EOS ad CNAF
  - Remote access enabled
  - Common catalogue view
- Required modifications available in iLCDirac DIRACPreview for testing
  - source /cvmfs/clicdp.cern.ch/DIRACpreview/bashrc

FCC VO @ iLCDirac : next

- Short term
  - Consolidate workflows
    - Following priorities of current activities
  - Get experience with metadata
  - Get experience with DIRAC/iLCDirac in general
  - Define procedure for registration “external” productions
    - Possibly run them through iLCDirac

- Define strategies for central and private productions
  - Central person in charge
    - Advisor for private productions
  - Planning needs

A few considerations on the resources needed

- The run at Z peak sets the scale
  - ≈10^{12} evts, 3-6 EByte storage, 10 MHS06 CPU (≈ current ATLAS yearly needs)
- These numbers are similar to the ones expected for (HL-)LHC
  - Do not expect issues for operations in 2040 and beyond
- For the FSR the situation is different
  - Analysis at Delphes level are possible (see $B_c \longrightarrow \tau^+ \nu_\tau$)
  - Full simulation of all components require 10^3-10^4 times more
- Techniques of overcome this limitations are required
  - E.g. interplay of full and parametrized simulation (see next)
- Planned community improvements in fast simulation very welcome
  - Possible improvements of the parametrized simulation treatment of critical parts such as calorimetric object could also be envisaged / investigated
    - E.g. based on improvements of fast simulation à la Geant4/GFlash or Machine Learning / GAN

FCC VO @ iLCDirac : some documentation

- Talks at FCC software meetings
  - G Ganis, Distributed Computing Matters, 16 April 2021
  - A. Tsaregorodtsev, LHCb/DIRAC Data Management, 30 April 2021
  - A Sailer, FCC Dirac Workflows with the Transformation System, 8 October 2021
- Documents
  - Distributed Computing section in the FCC Starterkit Lessons
  - Definition of EOS area (Google Doc)

Outlook

- Current direction in software is towards strengthening the common efforts
  - HEP is ready for a leap towards a common software ecosystem
  - Common R&D (AIDAs, CERN EP's, ECFA's …) have an essential role in this
  - Future collider community engaged around Key4hep and FCC is part and parcel of this
  - Documentation and training for both users and developers is essential

- FCC software requirements towards the next strategy update defined
  - Main challenge is to get realistic performance estimations with limited resources
  - The (FCC) community should all feel concerned and get involved
    - Effort needed to provide dedicated workforce and preserve current expertise
    - The best effort approach may not be sufficient

- The LEP data can still have a role for future EW and Higgs factories
  - Effort should be made to integrated them in the common software effort