Software for PED studies

FCC Computing Challenges

OpenLab Technical Workshop 2025
CERN

March 4, 2025
G Ganis, CERN-EP

About myself

- Leader of the EP-SFT group providing scientific software for the experiments
  - Current main activity lines: simulation (Geant4), Analysis and Data processing (ROOT), software distribution (CernVM-FS), Software stacks for experiments, ML for experiments (ML4EP; just started, focus on inference and fast simulation; connection with NGT/hls4ml)
- In EP-SFT since the creation of the group (2002), working in ROOT, CernVM-FS, Stacks
- Physics background
- A past in ALEPH at LEP (ancestor of LHC)
  - And a present: driving migration ALEPH data to new standard formats
- FCC Software co-coordinator since 2019

FCC: Future Circular Collider

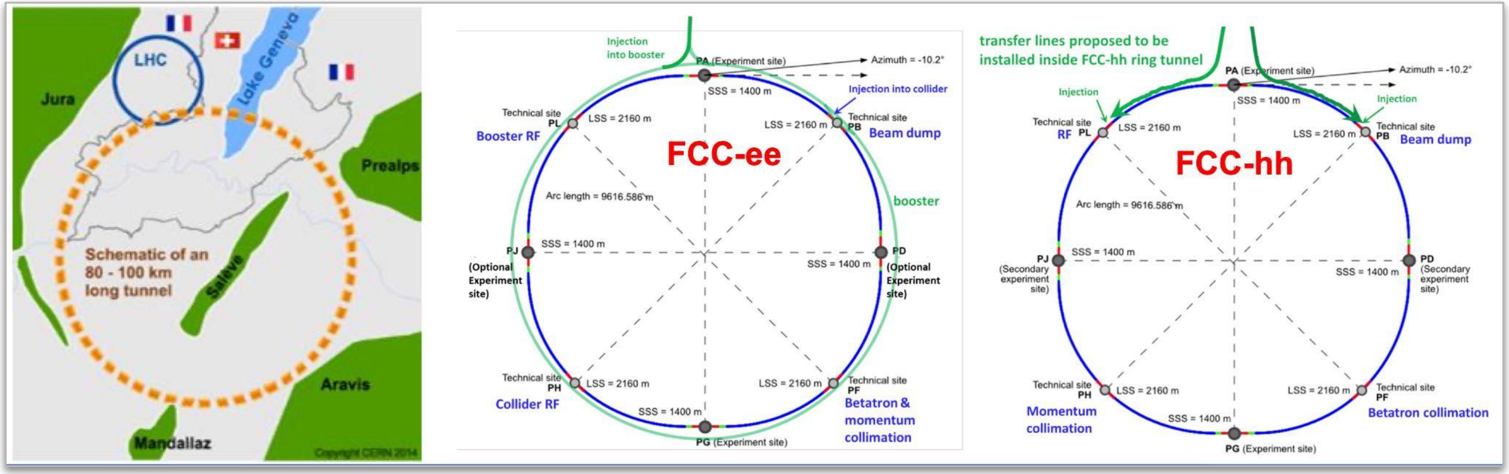
- Proposed particle accelerator a CERN for the after LHC era
    - *Future* = operation not before end of HL-LHC, i.e 2041+
- The project began in 2012 following the LHC's observation of a *light Higgs boson*, reinvigorating interest and potential of circular colliders as the next e^+e^- collider.
- Community (ESUPP, Snowmass) agree that the *HEP community highest priority is an e+e− Higgs factory to complete and deepen the Higgs bosons measurements at HL-LHC, followed by an hadron collider at O(10) TeV parton centre-of mass energy, i.e. O(100) TeV*
- Not yet unanimous consensus about the shape of the Higgs factory
    - But the hadron collider must be circular

> *The FCC project is an integrated program starting with a luminosity-frontier e+e- machine spanning from E_{CM} below the Z pole to top pair production threshold (FCC-ee) later evolving to an energy-frontier hadron collider (FCC-hh)*

ESUPP = European Strategy Update for Particle Physics = process updating european HEP strategy every 6-7 years
Snowmass = Similar process in USA, with similar frequency

FCC integrated program

- Comprehensive long-term program maximising physics opportunities, enabling the operation of a new major facility a few years after the end of HL-LHC
- Colliders share common civil engineering, dedicated and existing CERN infrastructure
- Tunnel length (~100 km) allows e+e- at O(350 GeV) and to reach O(100 TeV) for the hadron collider(*), boosting the physics reach of both colliders



(*) Pushing current dipole magnet technology to fields around 15 T

The FCC project preparation timeline

- 2012: project start
- 2018: Conceptual Design Report, submitted to ESUPP 2019
- 2021: ESUPP recommends e+e- Higgs factory as next collider, followed by hadron collider at energy frontier
- 2021: CERN council ask for FCC Feasibility Study Report for ESUPP 2026
    - Machine, physics/experiment/detectors(*), civil engineering, infrastructures, environmental and financial sustainability
- 2025: FSR being finalised
- 2025-2027: pre-TDR phase, to deepen FSR studies
- 2026: ESUPP 2026 recommendation
- 2028: possible approval by CERN Council

Phase 1

Phase 2

Phase 3

(*) Includes Software and Computing

The FCC project timeline

Fixing Choices

~2040

2045 - - - - - - - - - - - - - - - - 2060

E_{CM} = {91, 160, 240, 350} GeV

L = {140, 20, 7.5, 1.6} $10^{34}cm^{-2}s^{-1}$

Int L = {205, 19, 11, 3} ab^{-1}

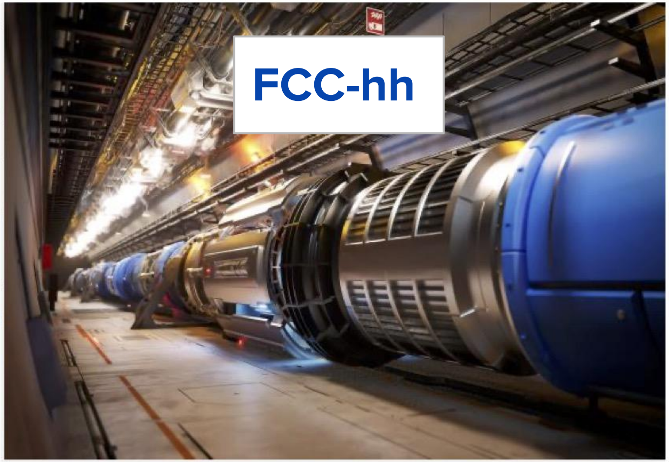
Evts = ~ {10^{12}Z, 10^{8} WW, 10^{6}HZ, 10^{6} TT}

Frontier high instantaneous luminosity to maximise physics potential, requiring a complex interaction region with 35” crossing angle

Dismantling and installing new collider **~10 y**

Fixing Choices

~2065



2070 - - - - - - - - - - - - - - - - 2095

E_{CM} = 100 TeV

L = 30 · $10^{34}cm^{-2}s^{-1}$

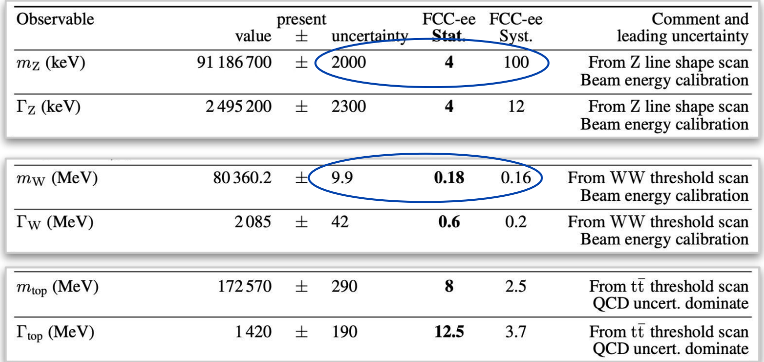
Int L = 30 ab^{-1}

Pile up = 950

High luminosity to maximise physics potential, in par with expectations

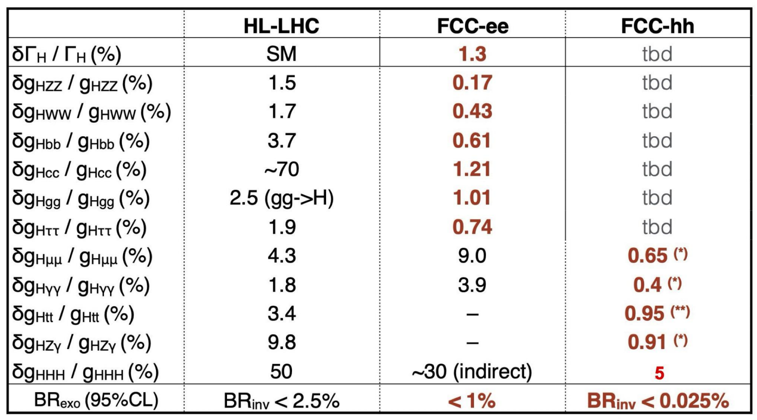
FCC physics expected precisions (excerpt)

FCC-ee: 2-3 order of mag. Improvement for EW observables



Observable	present value	±	uncertainty	FCC-ee Stat.	FCC-ee Syst.	Comment and leading uncertainty
m_Z (keV)	91 186 700	±	2000	**4**	100	From Z line shape scan Beam energy calibration
Γ_Z (keV)	2 495 200	±	2300	**4**	12	From Z line shape scan Beam energy calibration
m_W (MeV)	80 360.2	±	9.9	**0.18**	0.16	From WW threshold scan Beam energy calibration
Γ_W (MeV)	2 085	±	42	**0.6**	0.2	From WW threshold scan Beam energy calibration
m_{top} (MeV)	172 570	±	290	**8**	2.5	From $t\bar{t}$ threshold scan QCD uncert. dominate
Γ_{top} (MeV)	1 420	±	190	**12.5**	3.7	From $t\bar{t}$ threshold scan QCD uncert. dominate

FCC-hh Higgs measurements: $O(10^{-1})$ improvement on specific couplings.



	HL-LHC	FCC-ee	FCC-hh
$\delta\Gamma_H / \Gamma_H$ (%)	SM	**1.3**	tbd
$\delta g_{HZZ} / g_{HZZ}$ (%)	1.5	**0.17**	tbd
$\delta g_{HWW} / g_{HWW}$ (%)	1.7	**0.43**	tbd
$\delta g_{Hbb} / g_{Hbb}$ (%)	3.7	**0.61**	tbd
$\delta g_{Hcc} / g_{Hcc}$ (%)	~70	**1.21**	tbd
$\delta g_{Hgg} / g_{Hgg}$ (%)	2.5 (gg->H)	**1.01**	tbd
$\delta g_{H\tau\tau} / g_{H\tau\tau}$ (%)	1.9	**0.74**	tbd
$\delta g_{H\mu\mu} / g_{H\mu\mu}$ (%)	4.3	9.0	**0.65 (*)**
$\delta g_{H\gamma\gamma} / g_{H\gamma\gamma}$ (%)	1.8	3.9	**0.4 (*)**
$\delta g_{Htt} / g_{Htt}$ (%)	3.4	–	**0.95 (**)**
$\delta g_{HZ\gamma} / g_{HZ\gamma}$ (%)	9.8	–	**0.91 (*)**
$\delta g_{HHH} / g_{HHH}$ (%)	50	~30 (indirect)	**5**
BR_{exo} (95%CL)	$BR_{inv} < 2.5\%$	**< 1%**	**$BR_{inv} < 0.025\%$**

Complementarity FCC-ee/FCC-hh

Keeping the systematics under control to fully exploit the potential of FCC is one of the challenges for the project

FCC Computing

Computing at FCC-ee and FCC-hh

- The studies done so far have been done with software ecosystems using components developed at LHC, integrated with other products coming from explorative R&D projects (e.g. DD4hep for detector description)
    - While the software ecosystem, Key4hep, served and serves the current status of affairs, it is unlikely that his components will be the one used if the project is approved
- Progress has been initially affected limited availability of resources, in particular human
    - Many studies have been done in parametrised simulation
    - Effor now is rumping up, with more detailed / full simulation studies
- What follows is based on what is available now and to some extent speculative
    - It should hopefully give an idea of what kind of challenges FCC will face
- Given the timescale difference, the two colliders will be discussed separately

Software: Key4hep, a common software vision

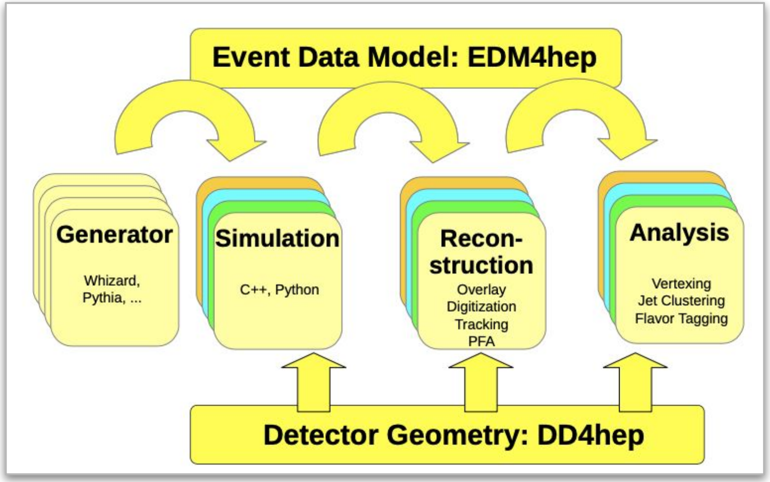
Create a software ecosystem integrating in optimal way various software components to provide a ready-to-use full-fledged data processing solution for HEP experiments

Complete set of tools

- Generation, simulation, reconstruction, analysis
- Build, package, test, deploy, run
- Available through cvmfs

Common Core ingredients

- PoDIO for EDM4hep, based on LCIO and FCC-edm
- Gaudi framework, devel/used for (HL-)LHC
- DD4hep for geometry, adopted at LHC (CMS,LHCb)



Event Data Model: EDM4hep

Generator
Whizard,
Pythia, ...

Simulation
C++, Python

Recon-struction
Overlay
Digitization
Tracking
PFA

Analysis
Vertexing
Jet Clustering
Flavor Tagging

Detector Geometry: DD4hep

Adopted in the Higgs factories community, DRDs, …

Computing at FCC in a nutshell

FCC-ee computing is not a problem, because it has similar or less issues of HL-LHC and it comes after. Final choices will have to be done around 2040, with HL-LHC well consolidated

FCC-hh computing is challenging, but choices to be made in 30-40 years which is huge in the ITC field, so one could expect/hope that the difficulties will be gone

Given the timescale difference, the two colliders will be discussed separately

FCC-ee

FCC-ee computing is not a problem, because it has similar or less issues of HL-LHC and it comes after. Final choices will have to be done around 2040, with HL-LHC well consolidated

- This is in general true. However, storage requirements, and perhaps Monte Carlo Generators, require a bit of attention
- Also, FCC-ee could be the almost ideal test-bed for ML-based developments such as
    - Particle Flow, tracking and/or flavour tagging/particle ID
    - Detector simulation
    - Detector optimisation with Differential Programming

FCC-ee: storage

- Integrated luminosities
  - Nominal: {205, 19, 11, 3} ab^{-1} at $\sqrt{s}$ = {88-94, 157-163, 240, 340-365} GeV
  - # of evts: $6x10^{12}$ visible Z decays, $2.4x10^{8}$ WW events, $2.1x10^{6}$ ZH events, $2x10^{6}$ tt events
- Baseline event sizes / processing time for hadronic evts at Z

Process	E_{CMS} (GeV)	Sizes /evt		Processing time /evt	
		Delphes[2] (kB)	Full[1] (MB)	Delphes[2] (ms)	Full[1] (s)
Z→had, Z→l+l-	91.18	8.3 , 1.2	**1.1 , 0.16**	14 , 0.5	**11 , 1.6**
WW→all, l^+nul^-nu	157-163	9.5 , 1.2	**1.3 , 0.16**	16 , 0.5	**13 , 1.6**
HZ→nunubb, bbbb	240	8.9 , 13	**1.2 , 1.8**	15 , 23	**12 , 18**
ZZ → all	240	10	**1.4**	17	**13**
ttbar → all	365	18	**2.3**	30	**23**

[1] Measured for Z-> had, extrapolated for others

[2] Delphse: parametrized simulation

CERN OpenStack node used for tests: 16 cores, 32 GB RAM.
CERN Openstack Core = 10-15 HEPSpec06 (HS06 ≈ HS23)

Projections/detector - Z,WW,HZ,Top full nom stat

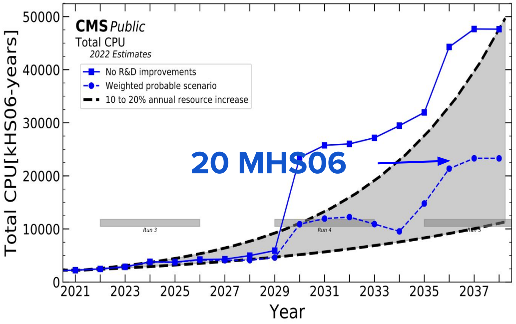
Run	Process	N evts	Delphes		Full Simulation		Comments
			Storage (PB)	Computing (HS06/4y)	Storage (PB)	Computing (HS06/4y)	
Z	qqbar	1500 G	**12.5**	**1.65 k**	**1650**	**1.5 M**	Full nominal statistics ≈ order of magnitude of the **data sample produced by one detector**
	l^+l^-	225 G	0.275	9	**40**	**30 k**	
W	WW	60 M	$\sim 10^{-3}$		0.075	54	
HZ	HZ	500 k	$\sim 10^{-5}$		$\sim 10^{-3}$	0.74	
	VBF-H	16 k	$\sim 10^{-6}$		$\sim 10^{-4}$		
Top	ttbar	500 k	$\sim 10^{-5}$		$\sim 10^{-2}$	0.93	
	HZ	90 k	$\sim 10^{-6}$		$\sim 10^{-4}$		
	VBF-H	23 k	$\sim 10^{-6}$		$\sim 10^{-4}$		
Total		1725 G	13	**1.65 k**	1700	**1.5 M**	
4 exp		6900 G	52	**6.6 k**	6800	**6.0 M**	

4y ≈ Z run

Projections of resource needs of HL-LHC

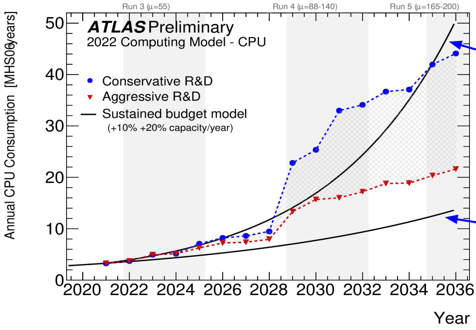
Including MC ≍ Data

Processing power



20 MHS06

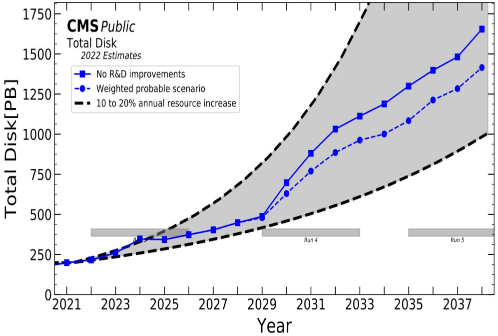
Better software



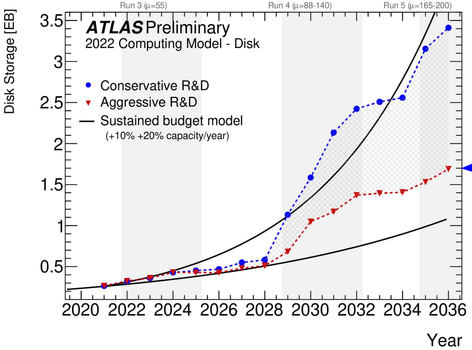
Pledged +20%

Pledged +10%

Storage



Better "use" of storage



1.5 EB

CMS, ATLAS

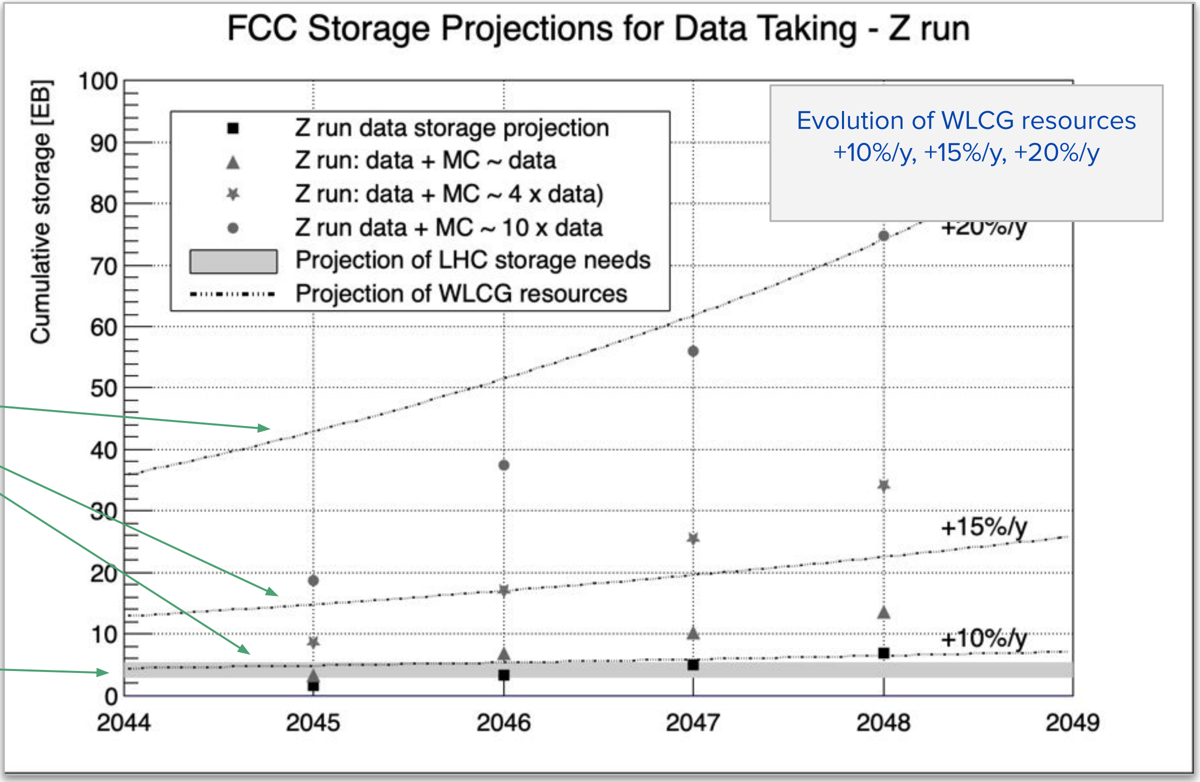
Illustrative Storage Projection for Z Run

4 experiments
4 equal runs {2045, 2046, 2047, 2048}

Evolution of WLCG resources
+10%/y, +15%/y, +20%/y

(starting point: 500 PB in 2020
≃ ATLAS + CMS + 40%)

LHC at the end of HL (≃ 5 EB)



Evolution of WLCG resources
+10%/y, +15%/y, +20%/y

FCC-ee storage remarks

- Raw data storage are comparable to Hl-LHC needs
- MC needs might be larger that at LHC
    - At LEP, 5x-10x data
- HL-LHC will continue needing the space during FCC-ee run for the final analysis
- Might require storage usage strategies/policies

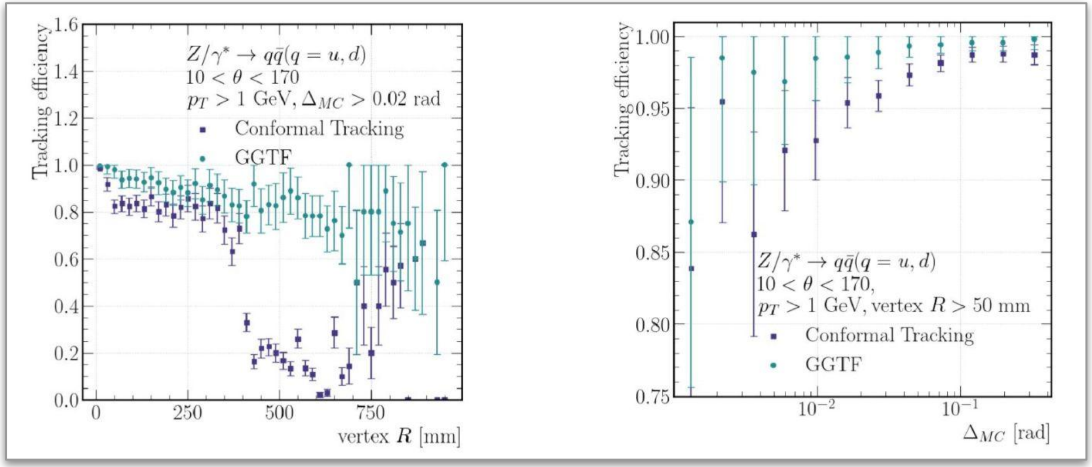
FCC-ee : Monte Carlo generators

- The precision expected at FCC-ee requires a significant improvement in MC generators, with respect to their previous generation, far exceeding the current state of the art
  - This includes new MCs dedicated to luminosity measurements
    - Low angle Bhabha scattering
- Large number of new graphs to be calculated and implemented in efficient way
- MC are GPU friendly but authors able to code GPU need to be trained
  - A new category of programming-skilled theorist will have to emerge
  - And get recognition for their work
- Other technical aspects concerned crossing angle, beam spectra, beam effects,...
  - Given the required precision they need to be implemented in the codes
- Impact of ML is investigated, for example, for phase space calculations

FCC-ee : ML Tracking

- Traditional methods for track finding are often complex and tailored to specific detectors and input geometries. The idea is to use ML to implement detector-agnostic algorithm
- Promising work has already been done in this direction:
  - Better tracking efficiency
  - Also for displaced tracks

A. De Vita
(see also
D Garcia et al)



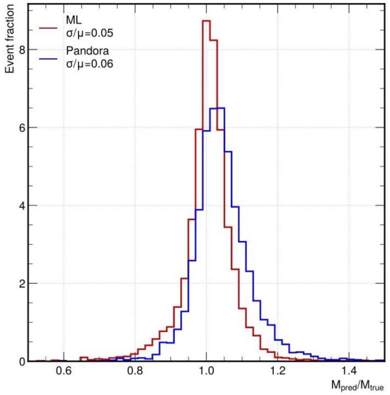
- Very useful for FCC-ee while still optimising the trackers
  - But if further developed might become also useful during for final trackers

FCC-ee : ML Particle flow

- Particle flow attempts to reconstruct the particle content resulting from a collision, using optimally the information from each detector.
- The power of the technology in e+e- was already evident at LEP and has been refined later at LHC and also in a experiment agnostic way (see Pandora)
- ML optimisation capabilities seem very promising for PF and indeed there have already studies in this direction applied to CMS (see J Pata et al, J. Kieseler et al)
- In the context of PF studies for FCCee some work has started comparing with classic approach

A. De Vita
(see also
D Garcia et al)

Jet-like particle
gun



FCC-hh

FCC-hh computing is challenging, but choices to be made in 30-40 years from now, which is huge in the ITC field, so one could expect/hope that the difficulties will be gone

- These difficulties concern
  - Storage requirements, which are O(100x HL-LHC), i.e. O(zettaB)
  - Monte Carlo simulation
    - Both for event generators and detector response
  - Triggering on large production data rates
  - Reconstruction at high pile up

FCC-hh : the trigger challenge

- FCC-hh needs a very high granularity detector to cope with the increase in the rates of interesting processes due to
  - 6x HL-HLC peak luminosity
  - Corresponding pile-up increase from 135 to 1000
- The reference detector will result in
  - 250 TB/s form calorimeters and muons
    - 10x ATLAS and CMS
      - Perhaps manageable by the time of the construction
  - 2 PB/s from the tracker!
    - The manageability of which is an open question

FCC-hh : high PU reconstruction

- Reconstruction in the pile up range expected for FCC-hh is going to be a challenge
- Preliminary studies done in full simulation with CMS at PU much larger that HL-LHC have shown
  - Tracking dominating the reconstruction time
  - A quadratic behaviour as a function of the PU, meaning reconstruction times O(100 x CMS)
- The current technologies are not suited for those value. Remains to see if the technological evolution, the inclusion of timing and the optimisation of algorithms, will bring down the times to bearable conditions

Summary

- FCC is expected to produce a lot of data during the two phases, e+e- and pp
- The storage of these data will be a challenge in itself, particularly for FCC-hh
- Exploiting the full potential of these large data sample several computing related improvements in various fields (event generators, reconstruction algorithms, …) will be beneficial or needed
- The optimal use of ML technologies will play an important role
    - And perhaps, given the timescale, also of QC technologies
- The HL-LHC will certainly help in understanding the extent of the challenge

Thank you for the attention