Software for PED studies

About Computing Resources

8th FCC Physics Workshop
CERN

Jan 15, 2025
G Ganis, CERN-EP

Recommendation from the mid-term review

Full simulations studies with comparison of different options require considerable computing resources. We recommend focussing on a few key channels. We also recommend that appropriate computing resources are allocated at CERN

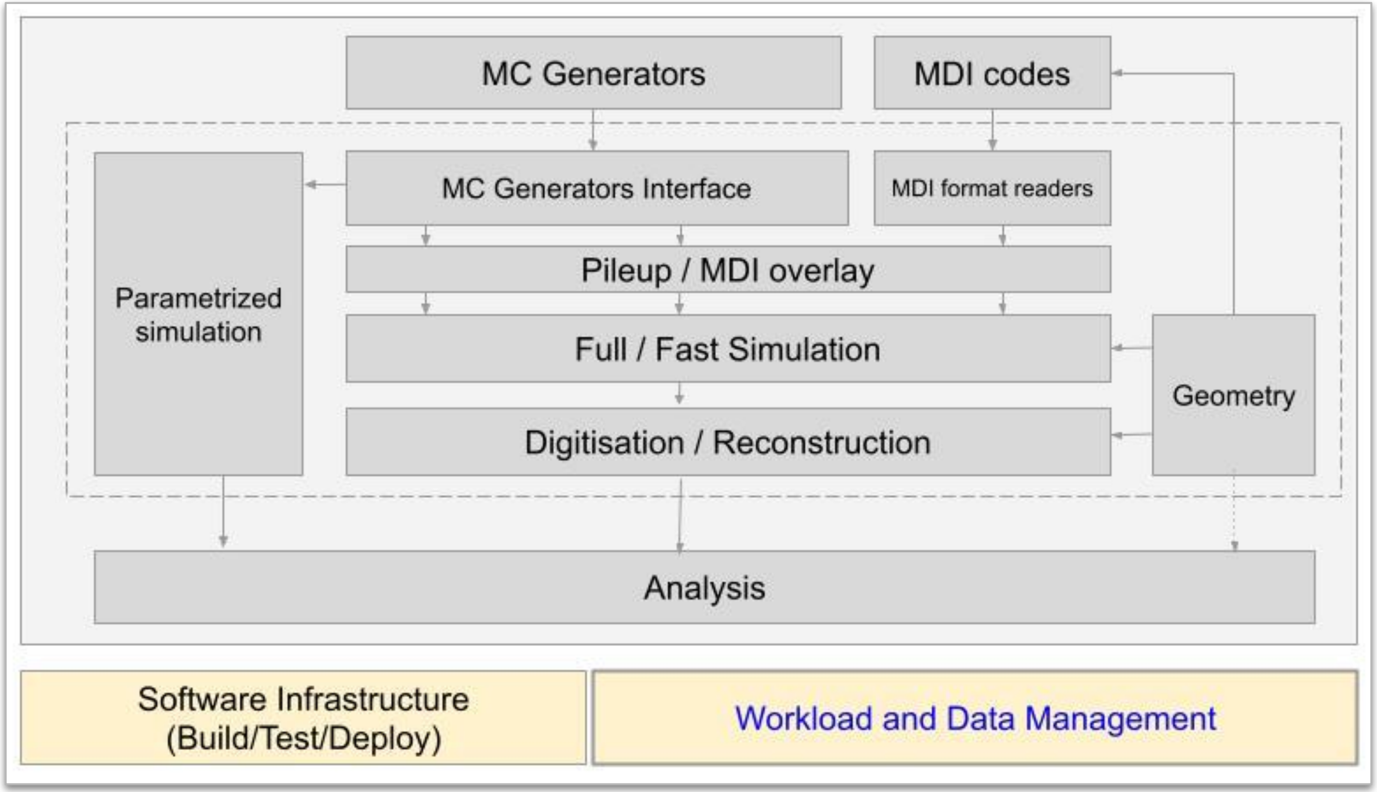
Related talks:
Computing Resource Needs, 7th FCC Physics Workshop 2024, Annecy, France
Computing Resource Needs for the FSR and beyond, FCC Week 2024, San Francisco, US

Today's talk reflects what has been included in the Feasibility Study Report

Preamble

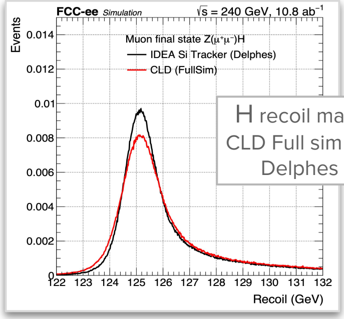
- Estimating the needs for computing resources requires assumptions which are difficult to get right, in particular when there are many moving targets
  - Data structures, detectors, tools, etc
  - Statistical needs
- Focus on two situations
  - Pre-TDR 2025-2027
    - Continuation of feasibility studies with development of detector concepts tested against a given set of benchmark measurements
  - Z run in 2045-2048 (4 equal years)
    - 4 frozen detectors, behaving ≈CLD (for storage and CPU)
      - Data evt sizes ≈ MC evt sizes
      - All data on disk during initial analysis

FCC Workflows



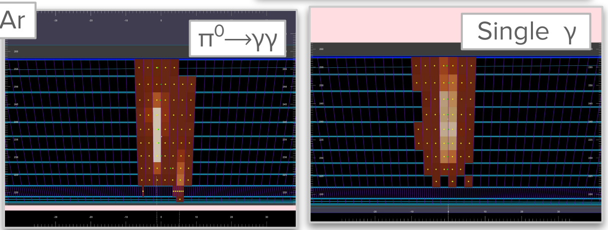
Activities to support during pre-TDR

- **Analysis**
  - Parametrized simulation for measurement feasibility
    - Provision of DELPHES samples
  - Increasing number of full simulation and reconstruction analysis
    - Provision of fully simulated + default reconstruction samples for all available detector benchmarks
      - Ideally 1:1 correspondence w/ Delphes samples

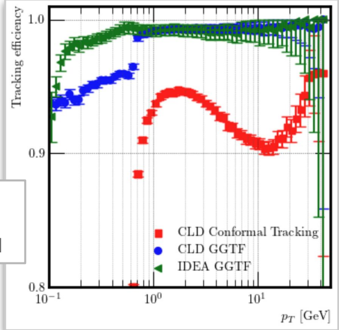


H recoil mass CLD Full sim vs Delphes

LAr



- **Sub-detector development/optimization**
  - Particle guns + samples of collision events
- **Algorithm design/optimization**
  - Particle guns and specific-event enriched samples



Tracking ML vs Classical

- Activities might benefit from access to accelerators (GPU, …)
  - In particular those based on ML

Current situation in terms of resources

- So far, most of the activities have been done on CERN resources
  - Storage: EOS volumes O(1 PB)
    - 600 TB (⟶900 TB) for central productions
    - 100 TB for analysis
  - Processing power
    - CPU: ~1000 cores (9000 HS06) on lxbatch
    - GPU: some resources at CERN, EuroHPC

In 2024 this represents about 1/1000 of what is available to LHC

Modeling the resource needs

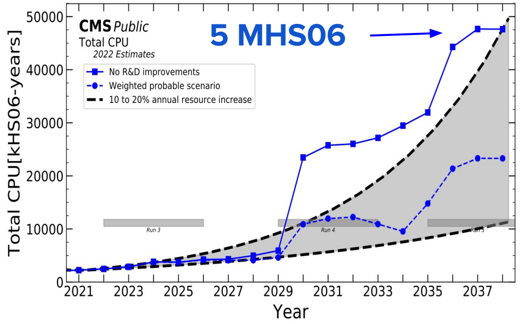
- LHC developed models to quantify needs to monitor the situation in view of HL-LHC and support request for funding
- For the LHC experiments
  - {detector, data formats, core algorithms, …} frozen
  - Main unknown are the collected luminosity, the statistics of the MC samples, reprocessing needs, …
- For future projects such as FCC
  - Target nominal luminosity known
  - Many moving targets and unstructured activities: number and type of detector concepts, data formats, algorithms, …

Appropriate resources requires the determination of how much MC statistics is required to achieve the goals

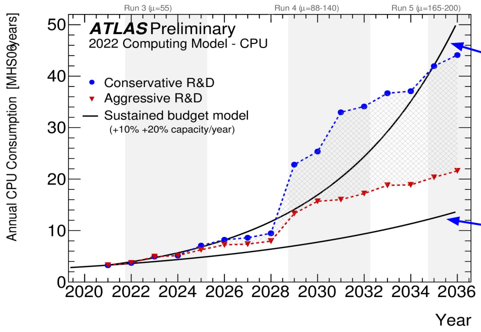
Projections of resource needs of HL-LHC

At LHC MC ≂ Data

Processing power



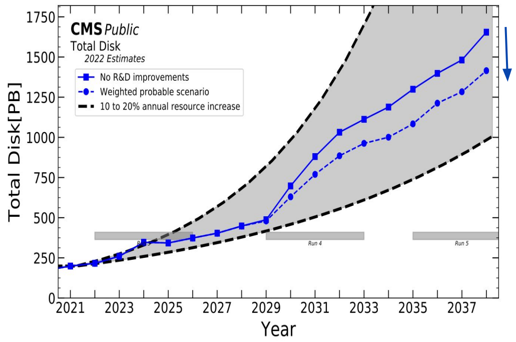
Better software



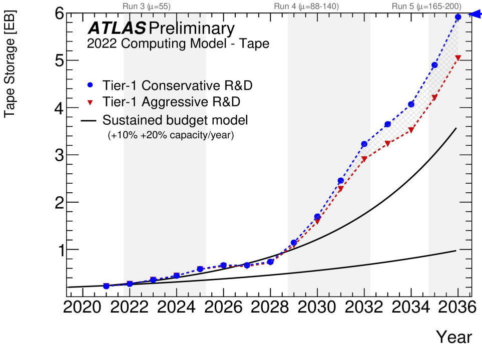
Pledged +20%

Pledged +10%

Storage



Better “use” of storage



6 EB

CMS, ATLAS

FCC-ee assumptions and baseline needs

- Integrated luminosities
  - Nominal: {205, 19.2, 10.8, 0.41, 2.65} ab^{-1} at $\sqrt{s}$ = {88-94, 157-163, 240, 340-350,365} 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 (kB)	Full[1] (MB)	Delphes (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

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 (HS06/3y)	Storage (PB)	Computing (HS06/3y)	
Z	qqbar	1500 G	**12.5**	**2.2 k**	**1650**	**2 M**	Full nominal statistics ≈ order of magnitude of the **data sample produced by one detector**
	l^+l^-	225 G	0.275	12	**40**	**40 k**	
W	WW	60 M	$\sim 10^{-3}$		0.075	72	
HZ	HZ	500 k	$\sim 10^{-5}$		$\sim 10^{-3}$	0.95	
	VBF-H	16 k	$\sim 10^{-6}$		$\sim 10^{-4}$		
Top	ttbar	500 k	$\sim 10^{-5}$		$\sim 10^{-2}$	1.25	
	HZ	90 k	$\sim 10^{-6}$		$\sim 10^{-4}$		
	VBF-H	23 k	$\sim 10^{-6}$		$\sim 10^{-4}$		
Total		1725 G	13	2.2 k	1700	2 M	

3y ≈ pre-TDR

Considerations 1

- Storage will be critical
  - Numbers do not include
    - Reconstruction
      - At LEP and Belle II, ~30% of the full sim CPU time
    - Analysis
      - Mostly of chaotic resources, but central ntuple reduction can be envisaged
  - For Delphes, in principle we can reuse the sample for a different detector “concept”
- Full nominal statistics for Z run unrealistic as it would require numbers similar to the ones of HL-LHC
- What can be done with amounts of the order of what we have now?
  - Full nominal statistics for WW, HZ, ttbar
  - *Large enough* Z run samples
    - *Large enough* to be defined

Example: Z LEPx100, {WW,HZ,Top} full nominal stat

Run	Process	N evts	Delphes		Full Simulation		Comments
			Storage (TB)	Computing (HS06/3y)	Storage (TB)	Computing (HS06/3y)	
Z	qqbar	400 M	3.25	~ 1	440	0.475	100x LEP
	l^+l^-	42.5 M	0.05		6.5	7	
W	WW	60 M	~ 1		75	71.5	Full nominal statistics
HZ	HZ	0.5 M	$\sim 10^{-2}$		~ 1	~ 1	
	VBF-H	16.25 k	$\sim 10^{-3}$		0.25		
Top	ttbar	0.5 M	$\sim 10^{-2}$		~ 10	~ 1	
	HZ	0.09 M	$\sim 10^{-3}$		0.2		
	VBF-H	23 k	$\sim 10^{-3}$		0.25		
Total		**500 M**	**4 TB**	**~ 1**	**0.53 PB**	**0.55**	
4 exp		**2000 M**	**16 TB**	**~ 4**	**2.1 PB**	**2.20 k**	

Considerations 2

- A baseline scenario with NxLEP Z run + full statistics for {WW,HZ,top} looks doable with the amount of resources realistically achievable
    - Whether this is enough needs to be worked out by the relevant working groups
- Ways to effectively increase the amount of resources available need to be pursued in parallel
    - Increase in the amount of resources
        - Pledged
        - Opportunistic (HPC, National/Institutes, …)
    - Software quality improvements (more efficient data structures, tools, etc)

Pledged resources

- Currently FCC gets resources from the quota reserved to the SME projects
    - Can probably be increased by large factors
- Discussions ongoing to integrate FCC in WLCG already in 2025
    - This would be justified considering FCC as the natural continuation of LHC

- A target of 5 PB, 100 kHS06 looks attainable in the time scale of the pre-TDR
    - This is about 10x current quotas and should enable, for example, the baseline scenario, with some contingency
- Note that the current tools (simulation, …) have not yet been tested at the statistics required by FCC ee

Opportunistic resources

- HPC or alike
  - In contact with CERN OpenLab to exploit EuroHPC
    - Development calls: transparent access through CERN OpenLab interface
    - CPU: AMD, Intel, ARM GPU, NVidia, AMD
    - O(1000) cores, 6 m to 1 year; investigating possibility for Dirac Integration
    - Use case: specific studies needing intense use limited in time
  - AI calls for ML specific developments

- Institute or national departments resources
  - To be possibly included in the FCC VO through DIRAC
- …

Distributed Workload and Data Management

- Distributed model required to include additional resources
  - Main requirements: central file catalogue, replication, remote access
- iLCDirac: LC community DIRAC instance
  - DIRAC: workload management, file catalogue used by LHCb, Belle II, BES III, JUNO, ILC/CLIC, ...
- FCC Virtual Organisation part of iLCDirac
  - CERN FCC resources (HTCondor, EOS area) associate to FCC VO @ iLCDirac
  - Added steering applications of interest for FCC workflows
  - Productions are starting using the instance
  - Inclusion for Storage Elements has been done with
    - CNAF, Bari and Glasgow
    - So far only for testing purposes

How to optimize use of resources?

- Software
  - Increase quality and efficiency of code
    - Long and expensive process, in general, use latest versions
      - Recent Geant4 is up to a factor of 2 faster for ATLAS
    - New ML-based faster simulation techniques promising
      - Gaining momentum in LHC (ATLAS: ML sim of parts of EM cal in production)
        - Need to understand how to optimally integrate them
  - Investigate possibility selective/filtered simulations
    - Filters at generation level
    - Simulate only parts of relevance (ALICE)
  - Reconstruction/analysis could benefit from using heterogeneous resources
    - Synergies with LHC
  - Investigate end-to-end ML-based techniques for large statistics studies
    - À la CMS Flashsim

How to optimize use of resources?

- Physics Performance
  - Investigate variance reduction technologies to go beyond the rule of thumb

    MC Sample = Expected data sample

    reducing the number of events required
    - Could be useful also in perspective, when data will be there
  - Facilitate interplay targeted full simulation and parametrized simulation
    - E.g. Automatic/optimal creation of Delphes configurations
  - LHC is testing at large statistics
    - Use this to identify processes requiring more attention

These optimization will strengthen requests to the funding agencies

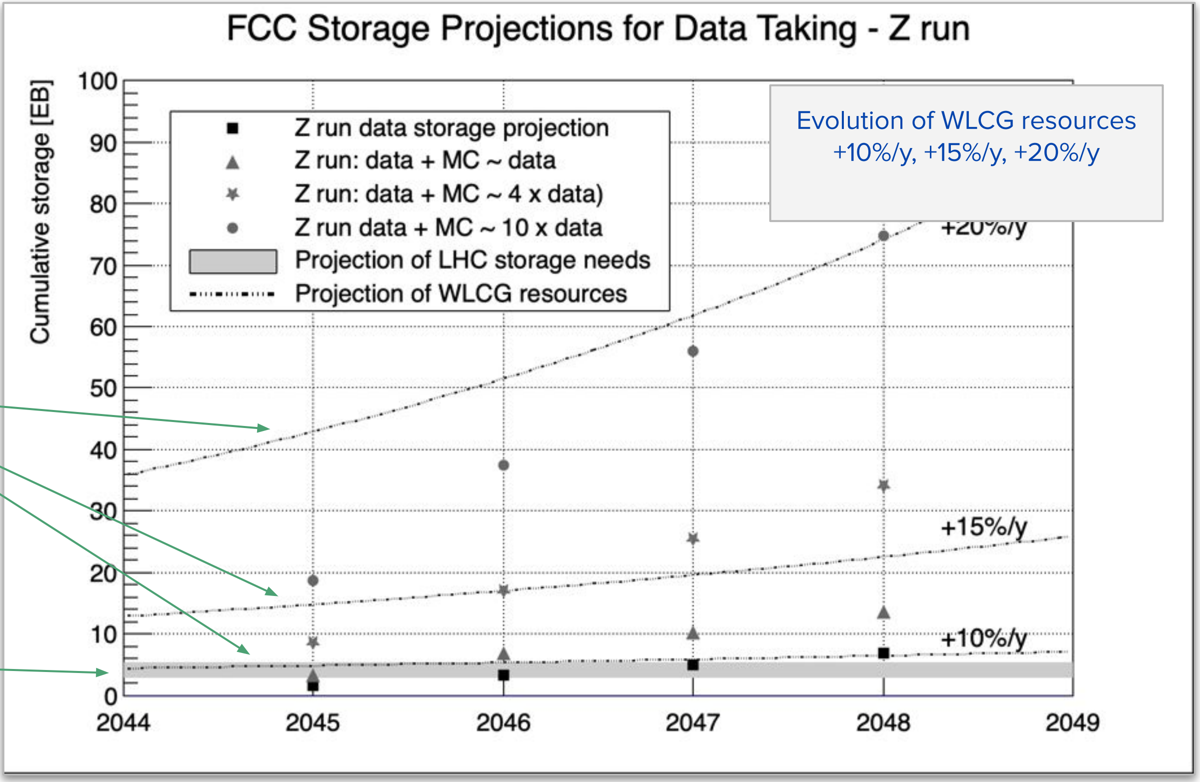
Illustrative Storage Projection for Z Run

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 + 10%)

LHC at the end of HL (≈ 5 EB)



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

Summary

- FCC PED studies ahead will need an increasing amount of computing resources
  - Exact requirements to be worked out by working groups
- Current CERN centric resources should enable minimal working scenarios
- In parallel, various fronts are being considered, following the LHC approach (WLCG, HPC calls, opportunistic usage, national resources) to increase available amount
  - Efforts are needed both on the software and on the physics interpretation side for optimal use, and to identify possible criticalities
    - Synergies with LHC activities need to be exploited
  - Storage requires attention: optimal use might require policies for retention