Software for PED studies

Estimating FCC computational needs from near term through the 2040s

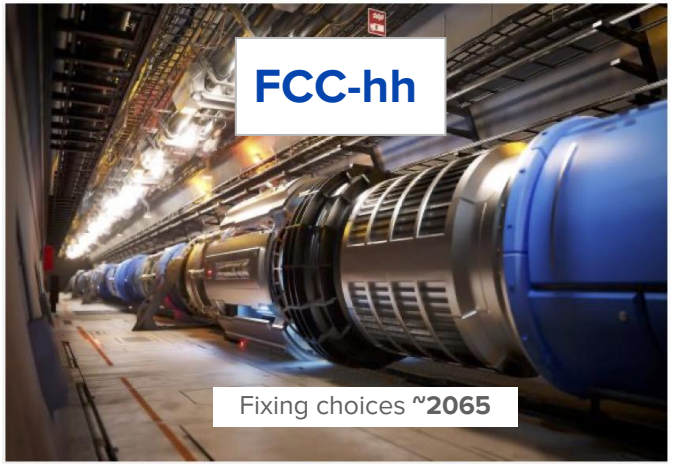
2026 Update of European Strategy for Particle Physics
Open Symposium, Venice Lido

Jun 23, 2025

G Ganis, CERN-EP

EP-SFT
Software Frameworks And Tools

The FCC project



Dismantling and installing new collider ~10 y

NOW
Pre-TDR

2025-2027

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, 10^6HZ, 10^6 TT}

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

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

Context

- The presentation addresses two questions raised during the FSR

> Are adequate computing resources available for the current and next phase (pre-TDR)?
> Or: how many resources are needed for the FCC pre-TDR (2025-2027) studies?

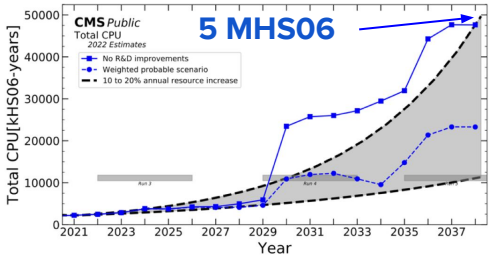
> How would the FCC-ee run in 2040's look like in terms of computing resources?

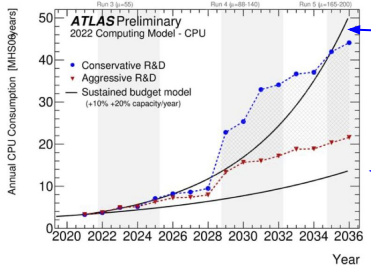
- Based on the results reported in the Feasibility Study Report, Vol 1, PED
  - Some updates followed subsequent discussions, in particular concerning missing processes and operation sequence
- Estimating the needs for computing resources requires knowledge about data structures, detectors, tools, etc, and statistical needs.
  - For a project with many moving targets we can only identify the relevant factors, for which we can make assumptions
- Focus on FCCee (no FCChh) offline needs (discussion for online/TDAQ just started)

Projections of resource needs of HL-LHC

Processing
power

Better software



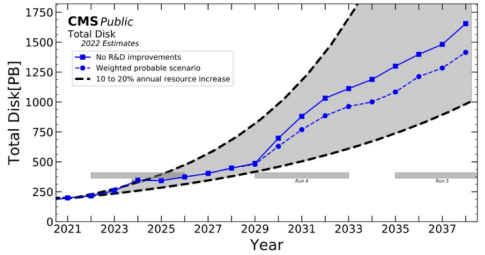


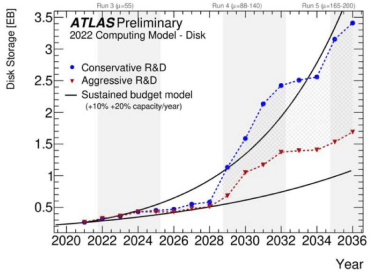
Pledged +20%

Pledged +10%

Disk
Storage

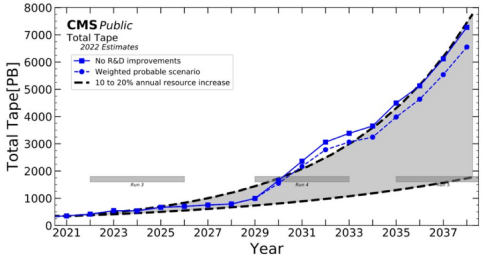
Better "use" of storage

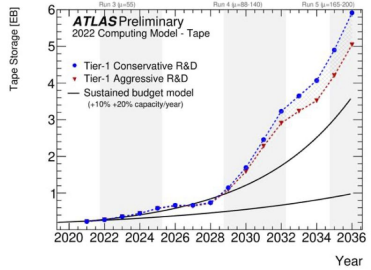




3.5 EB

Tape
Storage





6 EB

≈ 8-9 EB/exp
(at LHC MC ≃ Data)

[CMS](), [ATLAS]()

Modeling FCCee needs: factors / assumptions

- Workflows
  - During the pre-TDR phase the main activities are
    - **Analysis** of benchmark cases, usually started with parametrized simulation, and followed by an increasing number of full sim+reco samples
    - **Sub-detector development / algorithm design / optimization**, starting with samples with particle guns, followed by event (enriched) samples of fully simulated events
- Event sizes, processing times
  - Detectors concepts far from being frozen, but **# of channels ≈ 1 G** (LHC ≈ 100 M, LEP ≈ 1 M)
  - Use reference detector for scaling: CLD, FCC version of the CLIC detector
    - **Event sizes** (leptonic-hadronic): **parametrized 1.2-18 kB**, **full sim 0.16-2.3 MB**
    - **Processing time** (leptonic-hadronic): **parametrized 0.5-30 ms**, **full sim 1.6-23 s**
- Statistics needs
  - Baseline assumption: **MC ≈ expected data sample**
    - Needs to be reviewed case by case, both for current studies and the eventual experiment running

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

Full nominal statistics ≈ order of magnitude of the **data sample produced by one detector**

Run	Process	N evts	Delphes		Full Simulation		Comments
			Storage (PB)	Comp (HS06/3y)	Storage (PB)	Comp (HS06/3y)	
Z *Nominal* **50 ab^{-1}** 1.5·10^{12}Z	qqbar	1500 G	**12.5**	**2.2 k**	**1650**	**2 M**	**Resource needs completely dominated by the Z run**
	l$^+$l$^-$	225 G	0.275	12	**40**	**40 k**	
	e+e- lumi	1200 G	**1.9**	64	**250**	**200 k**	
	gg -> had	18 G	0.04	1.6	**6**	**3 k**	
W Nominal **5 ab^{-1}** 0.6·10^{8}WW	WW	60 M	~10^{-3}		0.075	72	Full simulation needs are well in the reach of todays resources
	e+e- lumi	18 G	0.05	~1	**6**	3 k	
	gg -> had	2 G	~10^{-3}	~0.1	**1**	300	
HZ *Nominal* **2.7 ab^{-1}** 0.6·10^{6}HZ	HZ	500 k	~10^{-5}		~10^{-3}	0.95	
	e+e- lumi	8 G	~10^{-2}	~0.5	**2**	1.5 k	
	gg -> had	1.5 G	~10^{-3}	~0.1	0.75	2.25	
Top *Nominal* **0.8 ab^{-1}** 0.5·10^{6}ttbar	ttbar	500 k	~10^{-5}		~10^{-2}	1.25	
	e+e- lumi	1 G	~10^{-3}	~10^{-2}	0.2	150	
	gg -> had	0.5 G	~10^{-3}	~10^{-2}	0.2	~0.1	
Total		~3000 G	**13 PB**	**2.3 k**	**~1955 PB**	**2.25 M**	

3y ≃ pre-TDR

Wrt FSR Table 21, added e+e- lumi and e+e-→ɣɣ →qq

≃ **2 EB / det**

Reco and analysis not included (at ALEPH@LEP ~10-20% of the full sim, for storage and computing)

Question 1

Are adequate computing resources available for the current and next phase (pre-TDR)?
Or: how many resources are needed for the FCC pre-TDR (2025-2027) studies?

Question 1: resources for current studies

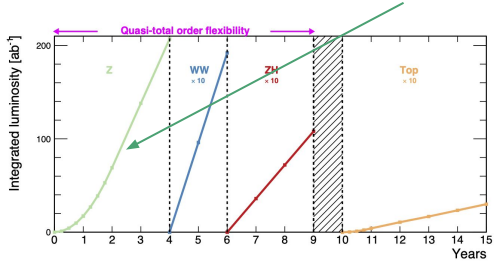
- The **FSR** has been done with O(**1 PB**) storage, O(**10 kHS06**) computing extracted from CERN non-LHC pool. This can be increased by O(10) in the time scale of pre-TDR
- The Z run requirements are ≈ HL-LHC: full nominal statistics unrealistic. {WW, HZ, ttbar} runs have little demands in term of resources
  - Heading towards a scenario in which we treat differently the Z run from the rest
- Can we have performance projections at nominal lumi scale w/ limited MC stat? Handles are
  - Software quality and efficiency; biased / filtered samples to reduce variance; new fast techniques (flashsim, …); interplay parametrized and full simulation; access to more heterogeneous resources (EuroHPC, national resources, …)
  - Several potential synergies with LHC
- *En passant*, 5-10 PB and 0.1 MHS06 enable a 100xLEP scenario (Z run with 100x LEP nominal luminosity), likely matching the precision of the current tools
  - Full sim storage: ≈500 TB, 600 HS06

Question 2

How would the FCC-ee run in 2040’s look like in terms of computing resources?

Illustrative Storage Projection for Z Run

4 experiments, 4 equal years at Z

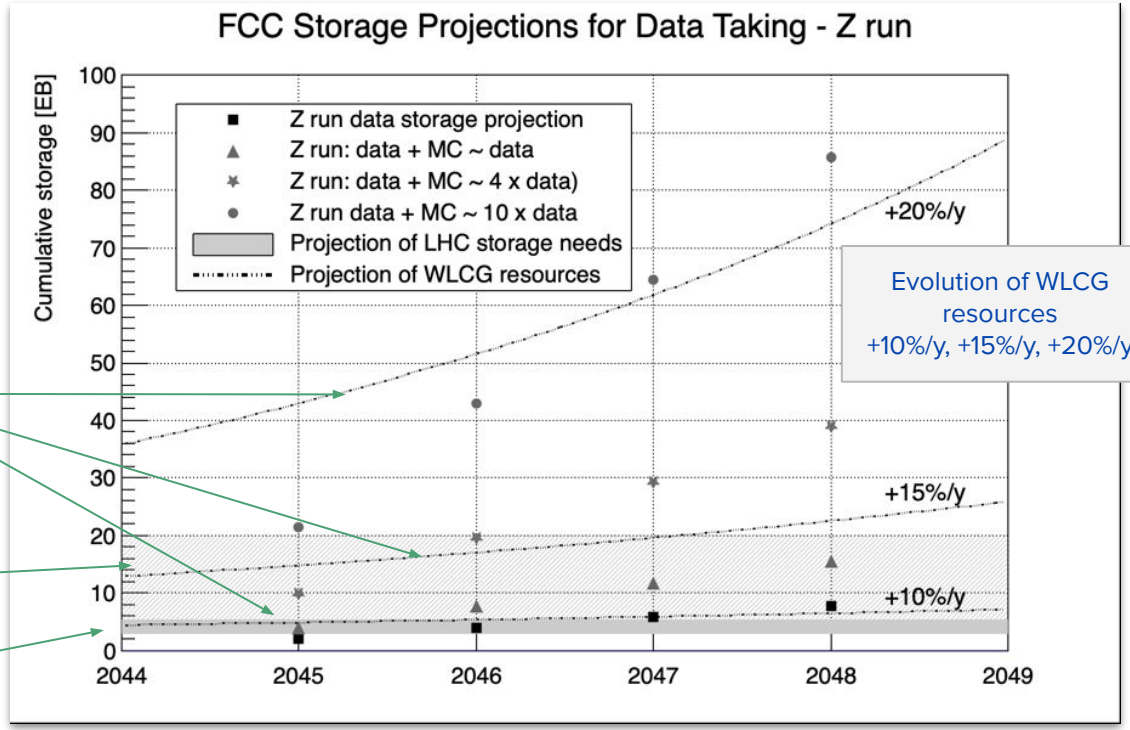


Evolution of WLCG disk resources
+10%/y, +15%/y, +20%/y

(starting point: 500 PB in 2020
≃ ATLAS+CMS + 10%)

LHC **tape** at the end of HL (≃ 5 EB)

LHC **disk** at the end of HL (≃ 5 EB)



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

More realistic Storage Projection for FCC-ee

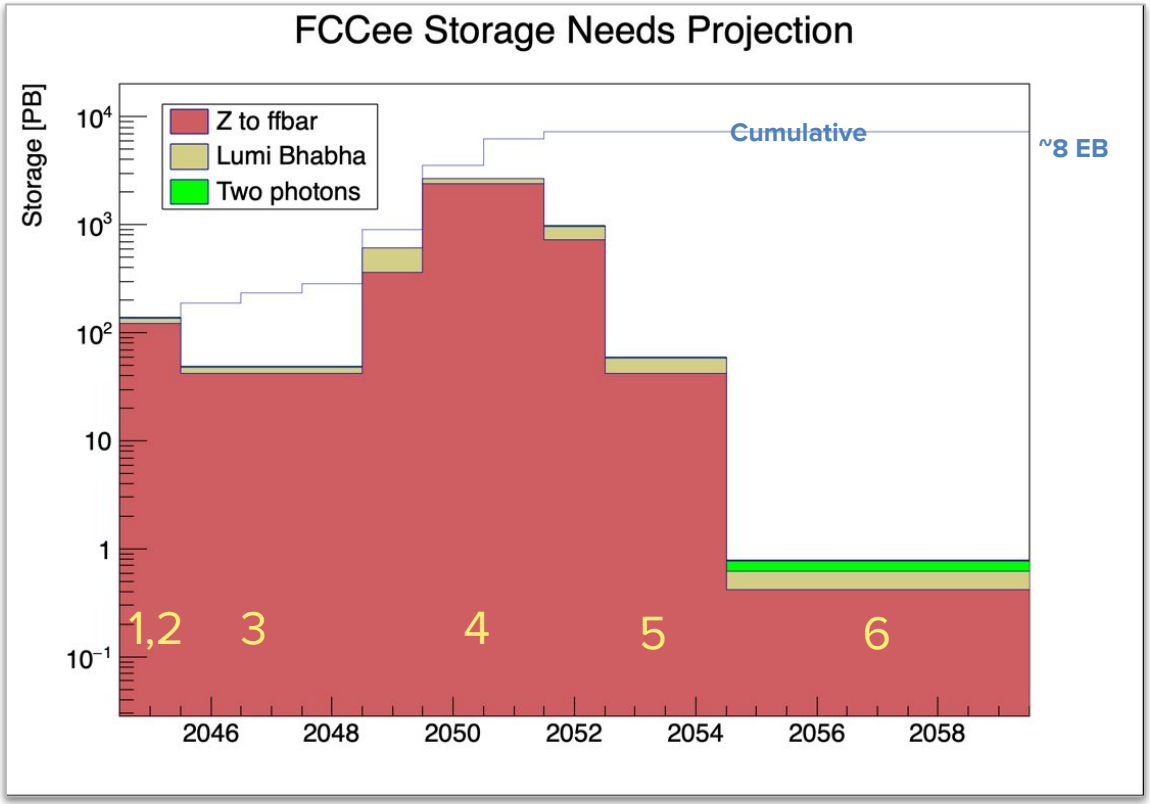
4 equally demanding experiments

More realistic sequence (15 y operations)

1. 1-2 m 0.25 * nom_lumi run at Z
2. 1-2 m 0.25 * nom_lumi run at WW
3. 3.5 y HZ run
4. 4 y Z run (peak and off-peak)
5. 2 y WW run
6. 5 y ttbar run

1 day Z calibration runs every month
during WW, HZ, ttbar runs

Based on P Janot et al (and P Janot private communication)



Summary

- Ultimate FCCee computational needs are O(HL-LHC)
  - FCCee operations in 2040s-2050s should not be particularly challenging
- During pre-TDR, evolution of current CERN centric resources should enable minimal investigating scenarios
  - Various fronts are being considered to increase available amount of resources, following the LHC approach (EuroHPC,...)
  - Ongoing efforts both on the software and on the physics interpretation side for optimal use, and to identify possible criticalities
- Exploitation of the several synergies with the LHC activities essential

Thank you!

Backup

More realistic Storage Projection for FCC-ee

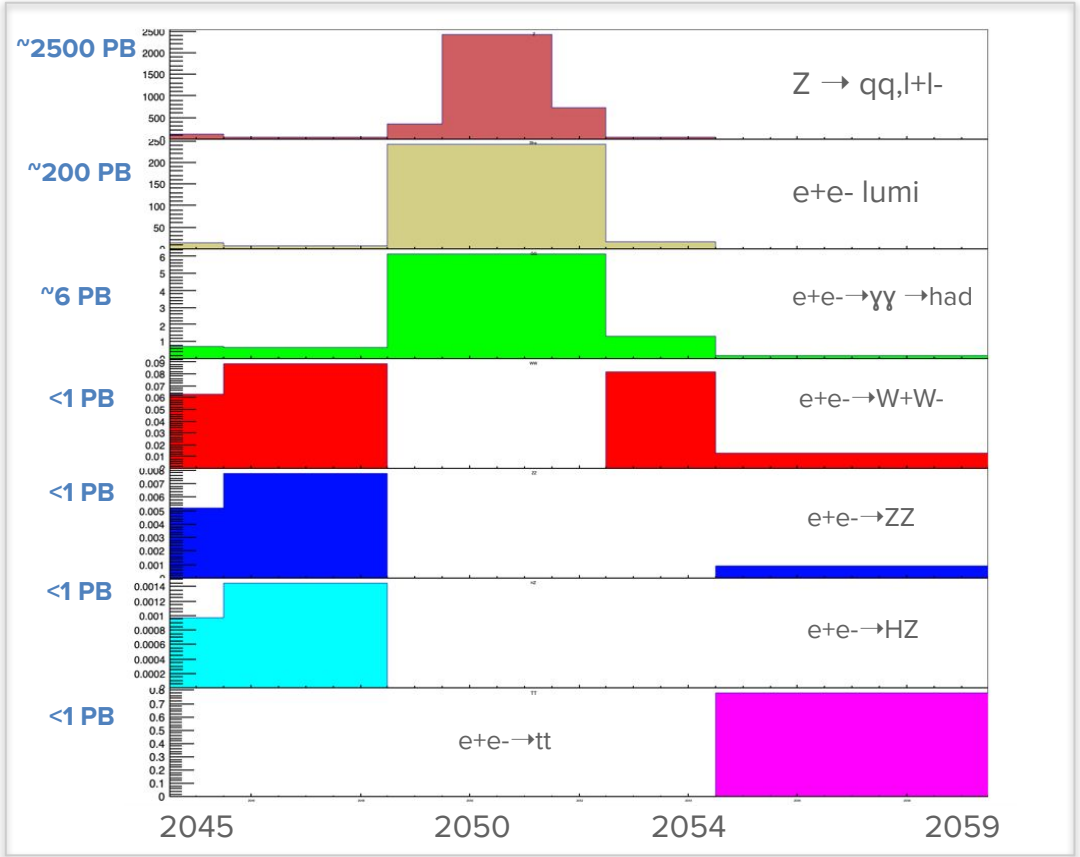
4 equally demanding experiments

More realistic sequence (15 y operations)

1. 1-2 m 0.25 * nom_lumi run at Z
2. 1-2 m 0.25 * nom_lumi run at WW
3. 3.5 y HZ run
4. 4 y Z run (peak and off-peak)
5. 2 y WW run
6. 5 y ttbar run

1 day Z calibration runs every month
during WW, HZ, ttbar runs

Based on P Janot et al (and P Janot private communication)



Modeling the resource needs

Table 1: The baseline FCC-ee operation model with four interaction points.

Working point	Z pole	WW thresh.	ZH	$t\bar{t}$	
$\sqrt{s}$ (GeV)	88, 91, 94	157, 163	240	340–350	365
Lumi/IP ($10^{34}\,cm^{-2}s^{-1}$)	140	20	7.5	1.8	1.4
Lumi/year (ab^{-1})	68	9.6	3.6	0.83	0.67
Run time (year)	4	2	3	1	4
Integrated lumi. (ab^{-1})	205	19.2	10.8	0.42	2.70
Number of events	6×10^{12} Z	2.4×10^{8} WW	2.2×10^{6} ZH + 65k WW $\rightarrow$ H	2×10^{6} $t\bar{t}$ + 370k ZH + 92k WW $\rightarrow$ H	

For given **target nominal luminosity**, estimating resources requires assumptions about type of detector concepts, data formats, algorithms, … i.e. moving targets and unstructured activities

And the MC statistics required to achieve the goals of the studies

To get an idea, use CLD (FCC version of the CLIC detector) as a reference detector, the full chain being available

FULL = 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)

Table 21: Baseline event sizes and processing times.

Process $e^+e^- \rightarrow$	$\sqrt{s}$ (GeV)	Size / event		Processing time / event	
		DELPHES (kB)	FULL (MB)	DELPHES (ms)	FULL (s)
$Z \rightarrow q\bar{q}, \ell^+\ell^-$	91.18	8.3, 1.2	1.1, 0.16	14, 0.5	11, 1.6
$W^+W^- \rightarrow$ all, $\nu\bar{\nu}\ell^+\ell^-$	157–163	9.5, 1.2	1.3, 0.16	16, 0.5	13, 1.6
$HZ \rightarrow \nu\bar{\nu}b\bar{b}$, $b\bar{b}b\bar{b}$	240	8.9, 13	1.2, 1.8	15, 23	12, 18
$ZZ \rightarrow$ all	240	10	1.4	17	13
$t\bar{t} \rightarrow$ all	365	18	2.3	30	23

Large enough example: Z, lumi, γγ: LEPx100

{WW,HZ,Top} still nominal stat

Run	Process	N evts	Delphes		Full Simulation		Comments
			Storage (TB)	Computing (HS06/3y)	Storage (TB)	Computing (HS06/3y)	
Z	qqbar	400 M	3.3	~1	440	500	100x LEP
	l^+l^-	60 M	0.07		10	10	
	e+e- lumi	320 M	0.5	$\sim 10^{-2}$	66	50	
	γγ →had	5 M	$\sim 10^{-2}$		2	1	
W	WW	60 M	~1		75	72	Nominal stat
	e+e- lumi	320 M	0.5	~1	66	50	100x LEP
	γγ →had	5 M	$\sim 10^{-2}$	~0.1	2	1	
HZ	HZ	500 k	$\sim 10^{-2}$		$\sim 10^{-3}$	0.95	Nominal stat
	e+e- lumi	320 M	0.5	~0.5	66	50	100x LEP
	γγ →had	5 M	$\sim 10^{-2}$	~0.1	2	1	
Top	ttbar	500 k	$\sim 10^{-2}$		$\sim 10^{-2}$	1.25	Nominal stat
	e+e- lumi	320 M	0.5	$\sim 10^{-2}$	66	50	100x LEP
	γγ →had	5 M	$\sim 10^{-2}$	$\sim 10^{-2}$	2	1	
Total 1 exp		**~1800 M**	**~6 TB**	**~3**	**~0.8 PB**	**~800**	
4 exp		**~ 7 G**	**~24 TB**	**~12**	**~3.2 PB**	**~3200**	

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

Table 21: Baseline event sizes and processing times. Parametrised simulation and full simulation with GEANT4 are denoted DELPHES and FULL, respectively. For FULL, the values are extrapolated from measurements performed with $Z \to q\bar{q}$ events, under the assumptions described in the text.

Process $e^+e^- \to$	$\sqrt{s}$ (GeV)	Size / event		Processing time / event	
		DELPHES (kB)	FULL (MB)	DELPHES (ms)	FULL (s)
$Z \to q\bar{q}, \ell^+\ell^-$	91.18	8.3, 1.2	1.1, 0.16	14, 0.5	11, 1.6
$W^+W^- \to$ all, $\nu\bar{\nu}\ell^+\ell^-$	157–163	9.5, 1.2	1.3, 0.16	16, 0.5	13, 1.6
$HZ \to \nu\bar{\nu}b\bar{b}$, $b\bar{b}b\bar{b}$	240	8.9, 13	1.2, 1.8	15, 23	12, 18
$ZZ \to$ all	240	10	1.4	17	13
$t\bar{t} \to$ all	365	18	2.3	30	23

FULL = Measured for Z-> had, extrapolated for others

FULL = 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

Table 22: Projected needs for the nominal integrated luminosity, for one experiment. The amount of HS06 is shown for a reference period of three years, i.e., roughly the duration of the next phase of the study.

Run	Process	Number of events	DELPHES Storage (PB)	DELPHES CPU (HS06)	FULL Storage (PB)	FULL CPU (HS06)
Z	$q\bar{q}$	1500 G	12.5	2.2 k	1650	2 M
	$\ell^+\ell^-$	225 G	0.275	12	40	40 k
W	W^+W^-	60 M	$\sim 10^{-3}$		0.075	72
HZ	HZ	500 k	$\sim 10^{-5}$		$\sim 10^{-3}$	~ 1
	VBFH	16 k	$\sim 10^{-6}$		$\lesssim 10^{-3}$	
top	$t\bar{t}$	500 k	$\sim 10^{-5}$		$\sim 10^{-2}$	~ 1
	HZ	90 k	$\sim 10^{-6}$		$\lesssim 10^{-3}$	
	VBF H	23 k	$\sim 10^{-6}$		$\lesssim 10^{-3}$	
Total		1725 G	**13**	2.2	**1690**	**2 M**

Over 3y ≈ pre-TDR

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

Table 23: Projected needs for one experiment, for the scenario with nominal integrated luminosity samples for the W, HZ and $t\bar{t}$ runs, and event samples 100 times larger than the LEP samples for the Z run. The total corresponds to four experiments requiring the same resources. The amount of HS06 is shown for a reference period of three years, i.e., roughly the duration of the next phase of the study. The totals in bold are beyond today's availability.

Run	Process	Number of events	DELPHES Storage (TB)	DELPHES CPU (HS06)	FULL Storage (TB)	FULL CPU (HS06)
Z (100×LEP)	$q\bar{q}$	400 M	3.25	∼ 1	440	475
	$\ell^+\ell^-$	42.5 M	0.05		6.5	7
W	W^+W^-	60 M	0.6		75	72
HZ	HZ	500 k	0.0065		1	∼ 1
	VBF H	16 k	∼ 0.001		0.25	
top	$t\bar{t}$	500 k	0.009		9	∼ 1
	HZ	90 k	∼ 0.001		0.2	
	VBFH	23 k	∼ 0.001		0.25	
Total		500 M	4	∼ 1	**530**	**550**
4 experiments		2000 M	16	∼ 4	**2100**	**2200**

Over 3y ≃ pre-TDR