

Computing for ILD experiment

Akiya Miyamoto, KEK

1 December, 2017
LCTPC workshop

Introduction

- Computing resources for ILD data analysis were not described in LOI/TDR, because , computing technologies and analysis software will evolve a lot
- After DBD, 2014-2015, first estimation (guestimate) was made by a request to identify items not mentioned by DBD. Using H20 scenario, 2 documents
 - ◆ LCC Software&Computing WG (<http://www.linearcollider.org/P-D/Working-groups>)
<http://www-jlc.kek.jp/~miyamoto/SoftwareCommonTask/docs/ILCComputing-EDMS1130485.A.1.1.pdf>
 - ◆ ILD computing estimate by A.M.
<http://www-jlc.kek.jp/~miyamoto/SoftwareCommonTask/docs/ILDComputing-EDMS1123105.A.1.1.pdf>
 - Roughly speaking, ILC is similar to Belle2 but smaller than HL-LHC
 - Inputs for an estimation of costs and spaces
- At LCWS2017, LCC requested to refine estimation in preparation for a possible query from a MEXT committee,
 - ◆ including new progress. With 250 GeV scenario.
 - ◆ by end of December
 - Too early for full revision, but a good opportunity to revisit very old estimate on data size and data processing.
 - Computing cost :
 - Partially by lab, a part of detector cost. mainly during data taking.
 - Infrastructure (building, power, cooling, network, ...) would be by lab.

Bases of estimation: ILD raw data size in TDR (@500 GeV)

2014

raw data size per train estimated @ 500 GeV

Sub-detector	Channels [10 ⁶]	Beam induced [Hits/BX]	Noise [Hits/BX]	Data volume per train [MB]
VTX (CPS)	300	1700	1.2	< 100
VTX (FPCCD)	4200	1700	1200	135
TPC	2	216	2000	12
FTD	1	260	0.3	2
SIT	1	11	0.3	6
SET	5	1		1
ETD	4			7
SiECAL	100	444	29	3
ScECAL	10	44	40	
AHCAL	8	18000	640	1
SDHCAL	70	28000	70	
MUON	0.1		8	≤ 1
LumiCal	0.2			4
BeamCal	0.04			126**

VXD : ~ 100MB
 BeamCal : 126 MB →
 reduced to 5% = 6MB
 Others < 40MB

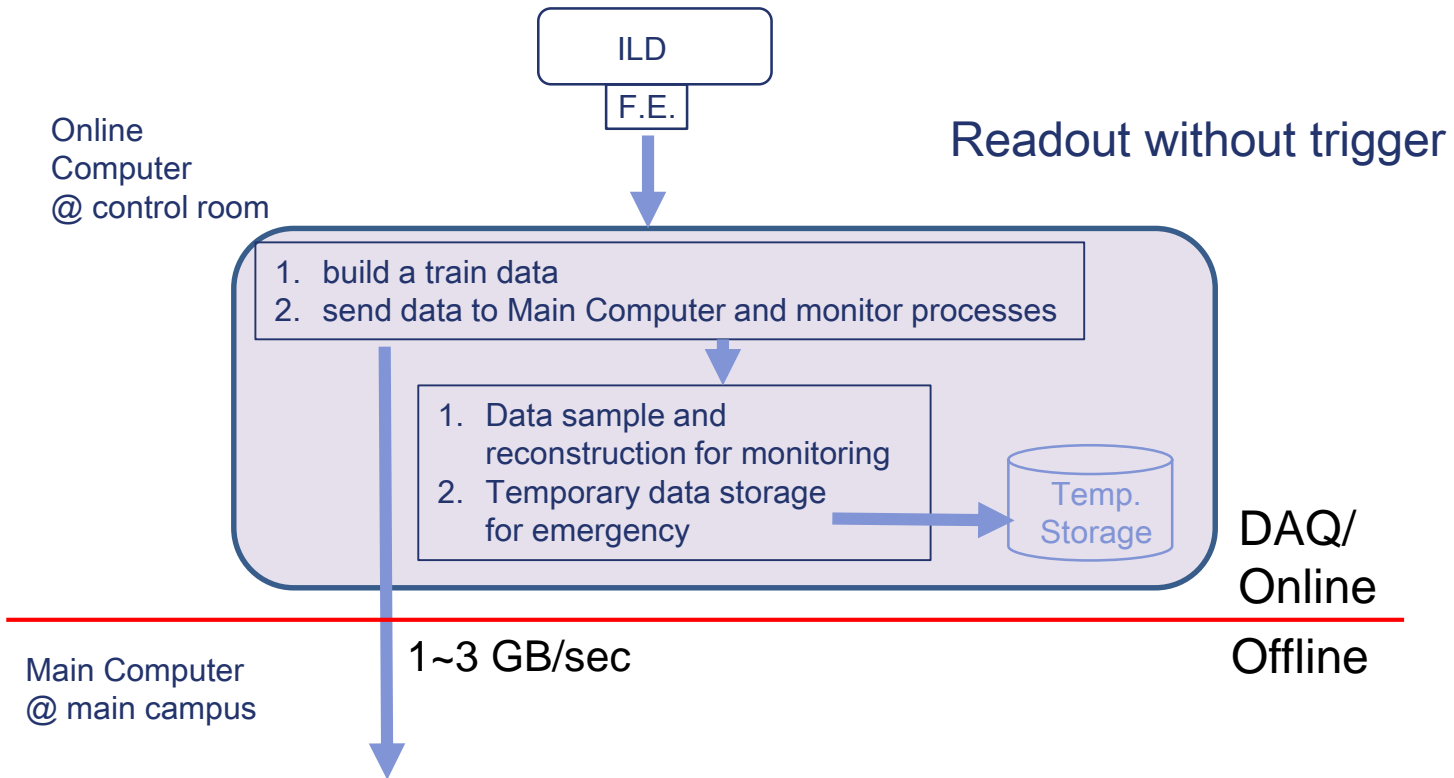
Dominated by lowE
 e+/e- background
 due to beamstrahlung

→ 130

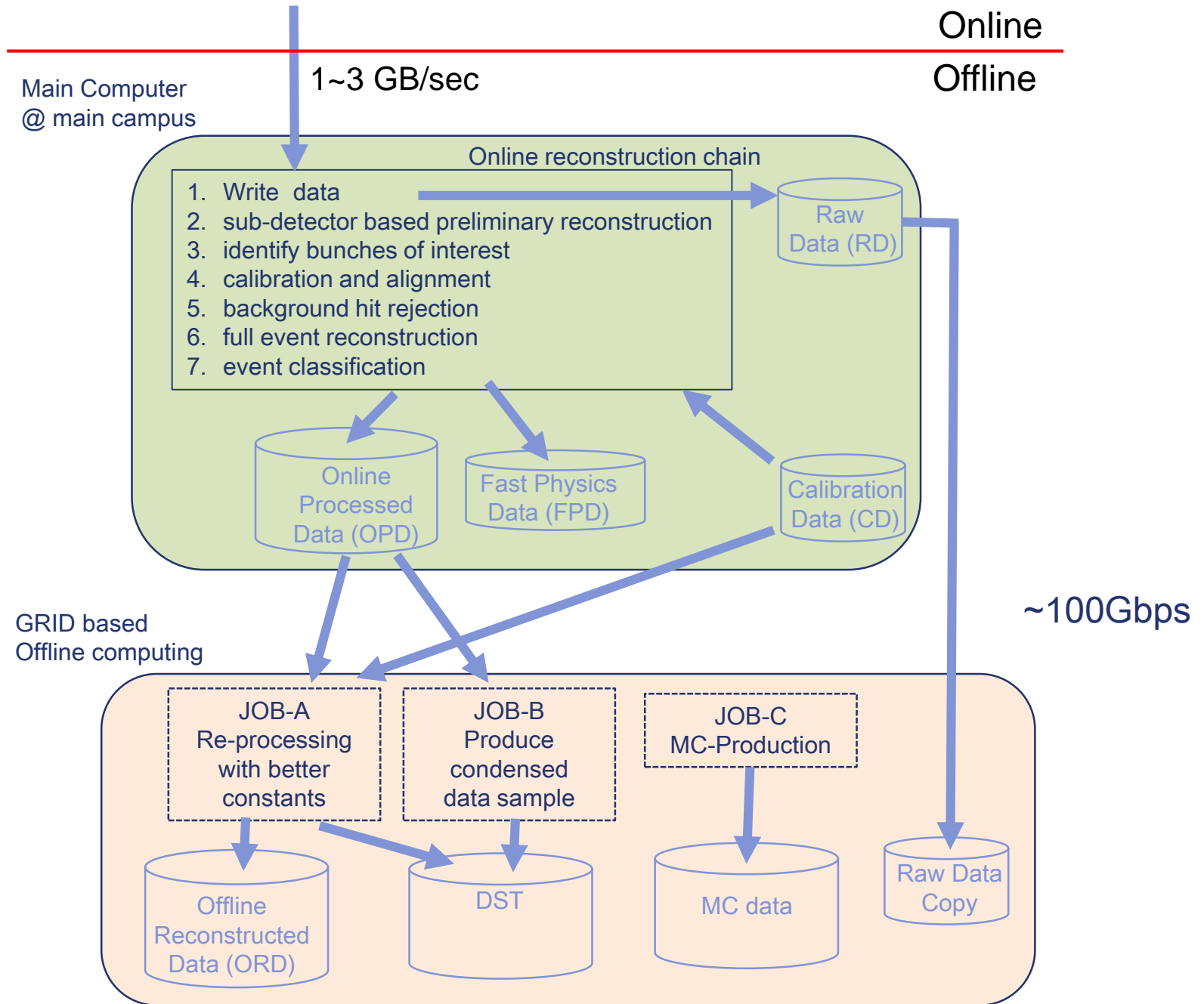
Total data size : < 150MB/train = 750MB/sec ~ 6Gbps (bit per sec)

~ 7.5PB/1 year (10⁷ sec) for ILD

A model of ILD data processing



DAQ/
Online
Offline



Assumptions : Data size

- Run scenario: 250 GeV. 250fb⁻¹/4 years, 750fb⁻¹/6 years
 - Data size:
 - ◆ Raw data:
 - TDR 500GeV with AHCAL corr.: ~280(~560 MB) /Train
 - Original + 1 copy somewhere in the world.
 - ◆ Filtering/Analysis
 - 2~3% of BXs kept. Processed size : x2 ~3
 - 9 copies
 - ◆ Simulation data:
 - x10 statistics of signal events
-

Assumptions : CPU

■ MC Simulation

- ◆ x10 real data, on GRID (not at lab.)
- ◆ CPU time estimation: based on DBD production, x2 for bhabha
- ◆ Data reconstruction: a fraction of simulation

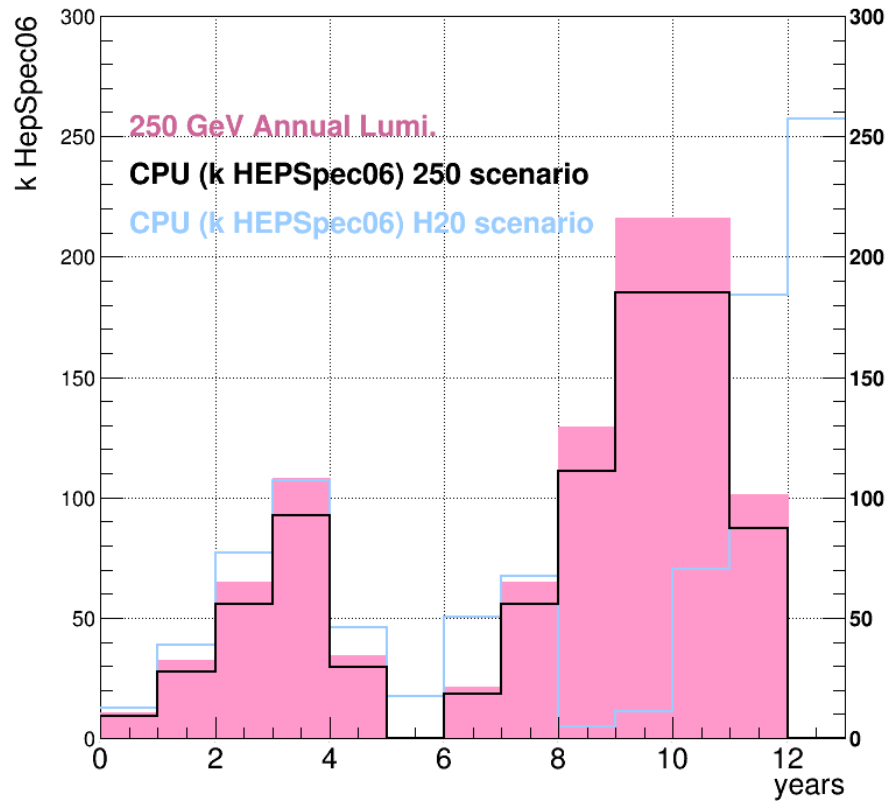
■ Real data processing:

- ◆ Data filtering: all BXs, same CPU time as data reconstruction
→ Major part of CPU demands
- ◆ Reconstruction : Filtered event (a few % of all BXs)
- ◆ Another reconstruction after re-calibration, on GRID

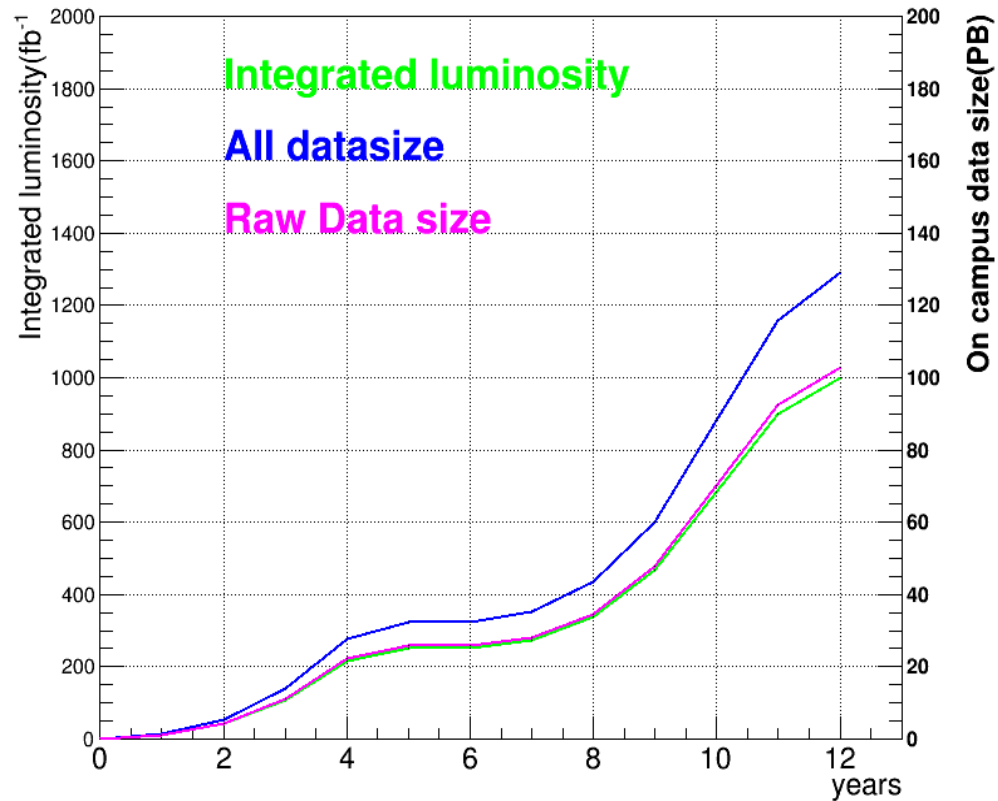
■ User analysis, detector calibration, are not counted.

ILD demands on Site (preliminary)

250 GeV : Campus comp for ILD



ILD: 250 Integrated luminosity and data size on Campus



Roughly 2 x KEKCC(now) for ILD & SiD

Questions to LCTPC

- Is the estimation of TPC data size in DBD still valid ?
- How much CPU will be required to
 - calibrate hits and remove background hits
 - identify signal bunches from a train data ?
- Comments ?