Computing Requirements of the ILC experiments

LCC Software and Computing Working Group
Akiya Miyamoto, Andre Sailer, Frank Gaede, Norman Graf, Jan Strube, Marcel Stanitzki

Abstract

A concept of computing for the two ILC detectors (ILD and SiD) and a preliminary estimation of the required resources for the ILC computing infrastructure at the ILC site are presented. The resource estimation is based on studies of the ILD detector concept as described in the Detailed Baseline Document, with input from SiD. More detailed studies should be carried out after the detector concepts have completed their TDR.

Introduction

This paper has been prepared as an input to the LCC Infrastructure Working Group chaired by Sakue Yamada. The computing cost for the experiments was not included in the ILC Detailed Baseline Design (DBD) but is needed in order to have a complete financial overview of the ILC project. This document summarizes the common ILC computing needs, however they are still preliminary. The current estimation is based on a significantly-sized computing center on the main campus. Reducing the requirements on this primary site by efficient sharing of resources with off-site centers is currently being studied. A prerequisite for this is the establishment of significant networking capabilities as currently planned for SINET5.

Overview of ILC Computing

HEP experiments commonly rely heavily on the computing infrastructure provided by a host laboratory (CERN, DESY, Fermilab, KEK, SLAC) for their data taking and analysis. ILD and SiD propose that the future ILC host laboratory will follow this tradition and provide the essential on-site computing resources. However, due to the planned dramatic increase in networking connectivity between Japan and the European and US infrastructures[1], the processing capabilities of the ILC campus are expected to be on the same order of magnitude as several other, remote computing centers. The LHC experiments have already established the usage of distributed computing (GRID) for reconstruction and Monte-Carlo production. Both ILD and SiD will build on this experience and envision that a significant part of the computing happens off-site using existing data centers such as the current LHC Tier-1 sites.

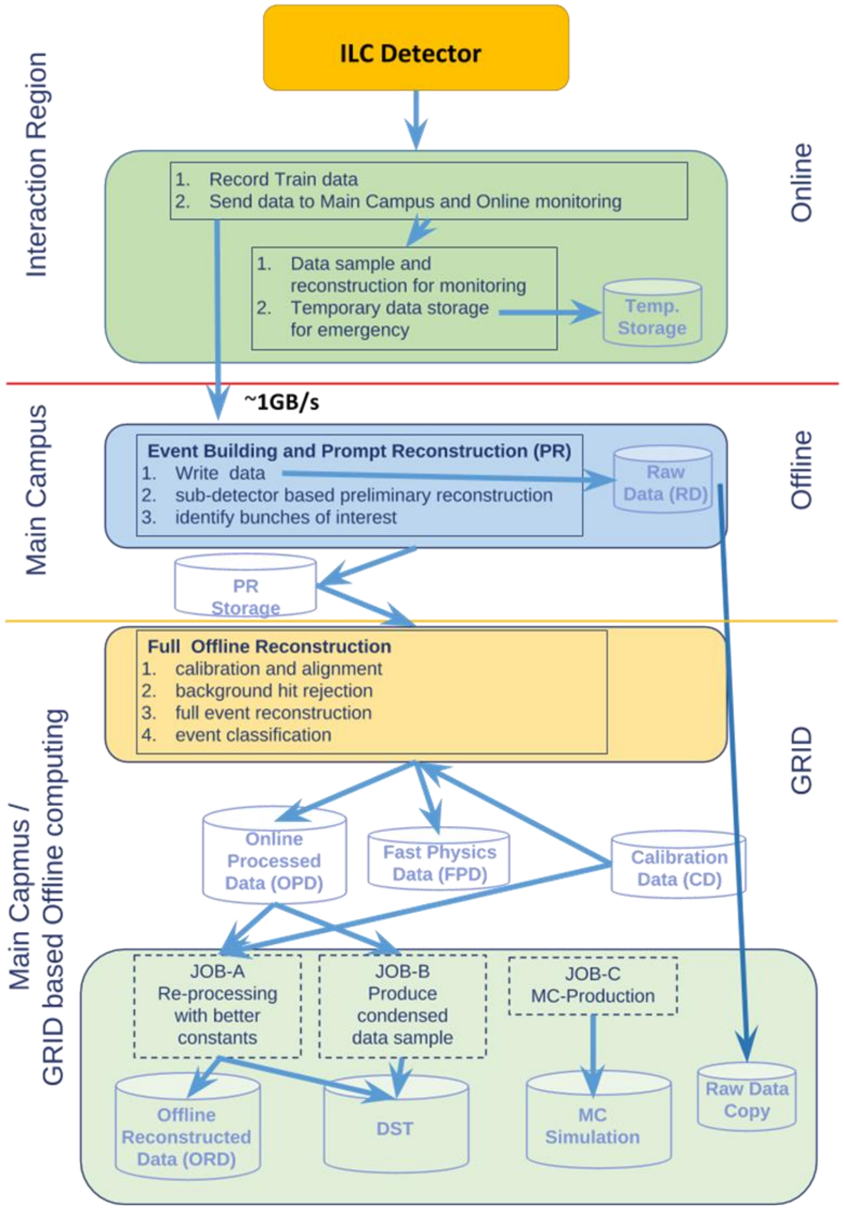
ILD and SiD have estimated the computing and human resources needed for data-taking and data-analysis. These numbers will evolve over time, but ILD and SiD currently envision the following model:

- At the IP, we require storage capacity in order to continue data taking for up to ten days in case of a severe network outage. Additionally, limited computing capacity is need for the online monitoring of the incoming data. The requirements for storage and CPU are small and are expected to be provided by the experiments.
- The raw data rate from the detector at nominal energy of 500 GeV at nominal luminosity of 1.47×10^{34} $cm^{-2}s^{-1}$ will be about 1 GB/s including safety factors on the amount of beam background. This rate is approximately the same for ILD and SiD.
- The raw data size will be up to ten PB per year per detector with the ILC running at the design luminosity at 500 GeV. The total data size accumulated strongly depends on the assumed running scenario of the ILC machine[2].

[1] "APAN 38 Meeting", Taiwan, August 14th, 2014. <http://www.jp.apan.net/meetings/1408-TW/SINET-updates_APAN38_20140814_kozono.pdf>
[2] "[1506.07830] ILC Operating Scenarios - arXiv.org." 2015. 26 Jun. 2015 <http://arxiv.org/abs/1506.07830>

Computing Requirements of the ILC experiments

LCC Software and Computing Working Group
Akiya Miyamoto, Andre Sailer, Frank Gaede, Norman Graf, Jan Strube, Marcel Stanitzki



- The network from the Interaction Point (IP) to the main campus needs to provide enough bandwidth to enable a real-time transfer of the raw data set.
- The network connection from the main campus to the rest of the world has less stringent requirements.
- Event-building -of the raw data and identification of the event(s) in a bunch train that will be selected for full reconstruction will happen at a computing farm at the main campus. It should be noted that while not all events will be selected for full reconstruction immediately, the intention is to store all bunch crossings in the full bunch train. Timely data analysis is crucial to maintain quality of data and a possible feedback to detector and accelerator operation. Significant CPU resources will be necessary in order to reconstruct events under beam background conditions. Based on a preliminary estimation, several thousand to a couple ten thousands cores of CPU, depending on running energies, will be necessary on site. An update to this estimation with more detailed models for detector simulation and reconstruction software should happen as part of the Technical Design Report for the detectors.
- ILD and SiD will store one set of the complete raw data at the ILC laboratory (main campus) and one complete, possibly distributed, copy at other sites.
- Full offline reconstruction of the events will happen using the GRID infrastructure. Only events that have

been tagged by the prompt processing will be reconstructed using the full chain.

- A total data taking period of about twenty years has been proposed by the ILC Parameters Working Group. In order to keep the computer system up-to-date and healthy, a regular system upgrade is necessary. For example, KEK replaces its computer farms after a rental period of four to five years. The cost of the system rental and replacement should be included as a part of the ILC running cost.
- Monte Carlo simulation and re-analysis of data with updated calibration constants will require significant computing resources. These will be provided by systems in the collaborating institutes, possibly using the already established GRID infrastructure.

Summary of required ILC computing resources

The preliminary requirements for the ILC computing infrastructure are summarized in Table 1 both for the first five years. The estimation is based on the currently available technology (2014). If one assumes an annual performance growth of 20% for the CPU and 15% for the storage at fixed cost (based on current LHC assumptions[3]), one arrives at cost reductions of factors of six and four, respectively, in ten years. The development speed in high-performance computing is so fast that a further significant cost reduction is to be expected at the start of data-taking. Therefore the numbers shown in Table 1 are subject to change. The running time is assumed to be equally shared between ILD and SiD.

Table 1: Summary of the required computing resources at the ILC laboratory for both ILD and SiD experiments. Note that the computing resources other than at ILC laboratory are not included in this table. We assume that CPU is predominantly used for event building and reconstruction. Sharing of resources between the two detectors in a push-pull scenario could lead to a reduced requirement for CPU.

Items	**unit**	**Total**
Storage before luminosity upgrade	PB	100
Max. CPU needs before luminosity upgrade	kHEPSpec	250
Raw data rate from detector to the main campus	GB/s	1.5
Network bandwidth required from detector to the main campus	Gbps	40
Network bandwidth from the main campus to outside	Gbps	40

Table 2 Global resources (including those at the ILC campus, shown in Table 1), raw data copy, including resources for user analysis. The storage required for user analysis is expected to be approximately 10% of the data size and is not included explicitly in this estimation. The estimation for CPU for user analysis is assumed to be comparable to the prompt processing of recorded data based on current Belle II planning.

Items	**unit**	**Total**
Storage before luminosity upgrade	PB	250
Max. CPU needs before luminosity upgrade	kHEPSpec	400
Max. CPU needs for user analysis	kHEPSpec	250

Comparison with Belle-II and LHC Computing needs

We compare the ILC computing needs with the requirements by the Belle-II (refer to footnote 4) and the LHC (refer to footnote 3) experiments. Belle-II currently plans their computing needs up to 2022 and the total required resources are given here. For the large LHC experiments, ATLAS and CMS, we list the estimation for Run II of the LHC with an expected 300 fb^{-1} per experiment.

Table 3 Comparison of the ILC requirements with the requirements of the Belle-II experiment at KEK and the

[3] Ian Bird. WLCG Workshop, Copenhagen. 12th November 2013 <https://indico.cern.ch/event/251191/session/2/material/slides/1?contribId=9>

ATLAS and CMS experiments at CERN[4].

Item	**unit**	**Belle-II**	**ATLAS**	**CMS**	**ILC**
Date	year	2016-2022	2015-17	2015-17	2029-2036
Storage	PB	485	337	295	250
CPU	kHEPSpec	1650	1695	1690	650
Raw data rate from dete to the main campus	GB/s	1.8			1.5

Table 4 Development of data rates and CPU requirements for BELLE-II over time.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Tape (PB)	2.8	2.8	2.8	2.8	19.24	54.43	103.55	153.89	204.64	255.39
Disk (PB)	4	4	5	8	27.98	79.17	115.68	153.1	190.82	228.55
CPU (kHepSPEC)	45	45	50	55	328.31	568.98	567.54	609.45	643.14	672.6

Support for ILC computing

We suppose that the following functions are included as the computing infrastructure and provided by the ILC laboratory.

Table 5 Services expected to be provided by ILC campus.

Task	**Details**
Management	Operation in general, negotiation among sites, maintain rules for computing, information delivery
Network	Maintenance, monitor, trouble handling develop faster network
Certificate	Maintenance, person ID, issue certificate, trouble handling
Security	Maintenance, Monitor, incident handling, prevention
Storage	Maintenance, trouble handling, improvements
CPU servers	Maintenance, trouble handling, improvements
Software maintenance	Maintenance, trouble handling, improvements
Core Software development	Design, implementation, maintenance, trouble handling, improvements
User contacts	User portal, monitor, dispatch problems, trouble handling
Misc. server management	Operation, maintenance, trouble handling of various servers such as EDMS, Indico, Web, cloud, mail, etc
Videoconferencing Support	Maintenance & User support

We assume that personnel to maintain the computer hardware and the operating system itself are included in the rental cost of the computing farm.

[4] http://indico.cern.ch/event/274974/contribution/73/material/slides/0.pdf