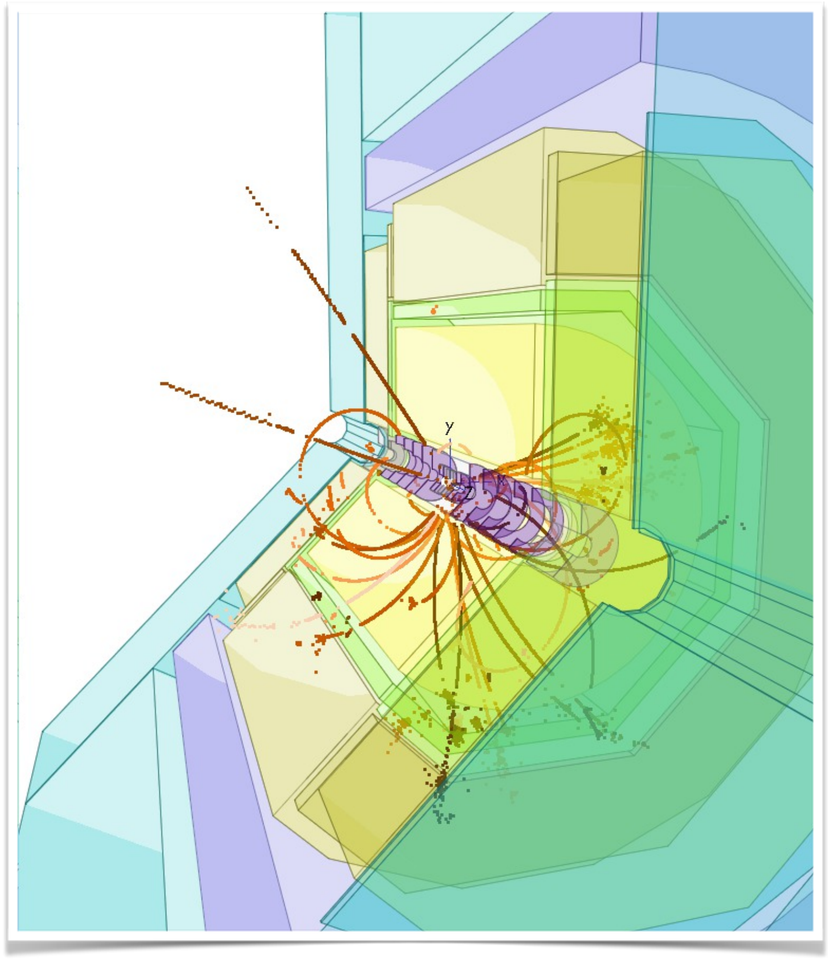
Software Coordinator Report

ILD Software and Analysis Meeting

Frank Gaede, DESY, 2.12.20

Outline

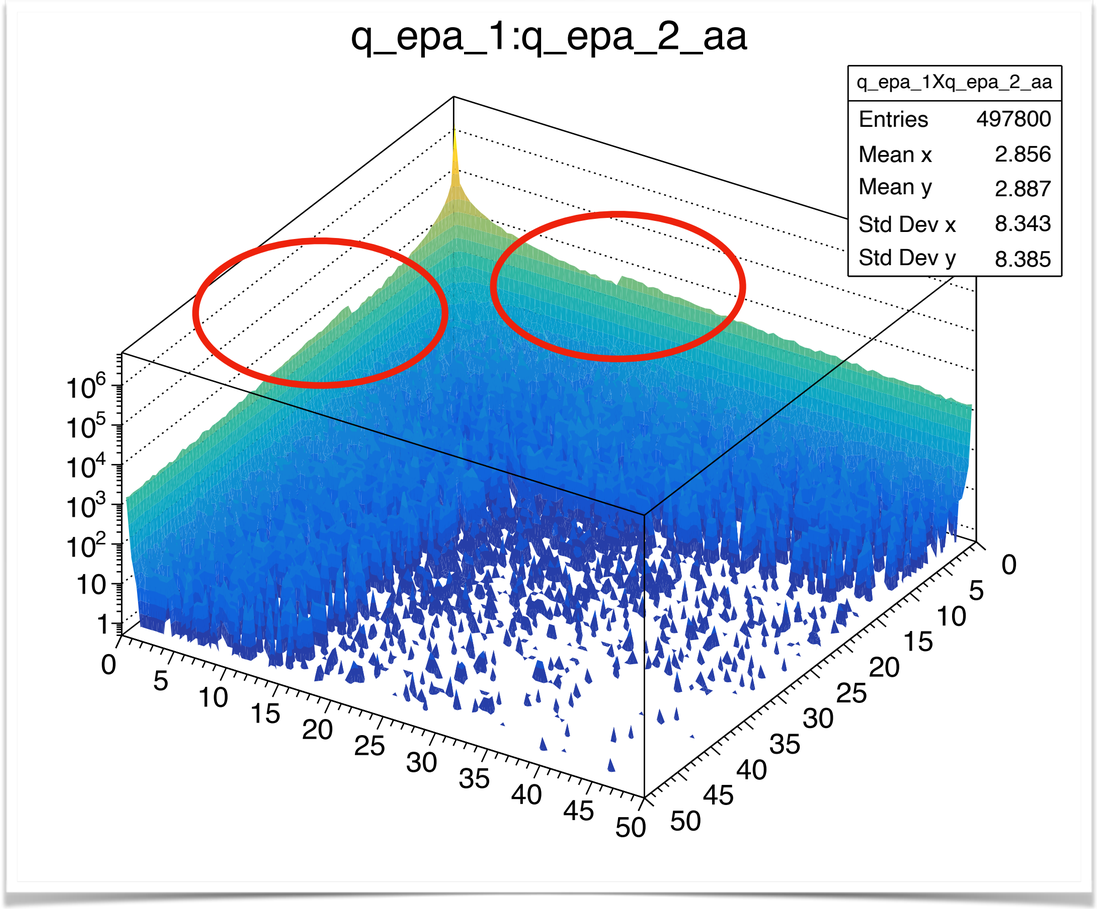
- Generator
- Simulation
- Reconstruction
- Monte Carlo Production
- Announcement: Julia Tutorial



Generator

M.Berggren, J.Tian

- no news since last meeting
- working on a problem with cross section of samples with virtual photons
  - 4 cases w/ different Q^2 cuts
  - observe factor two too large cross sections in two cases and jump in Q^2
  - reported to Whizard authors
  - waiting for reply/fix
  - …
- still work in progress w/ Whizard authors



Reconstruction

R.Ete

- started work on more realistic time measurement in Calorimeter Digitizer
  - see: https://github.com/iLCSoft/MarlinReco/pull/83
- features:
  - sum of hit energy contributions over threshold define the hit time
  - take electronic shaping time (fast/slow) into account
  - allow for optional resolution via smearing
- work in progress
  - will have a report in a future SW&Ana meeting

Simulation

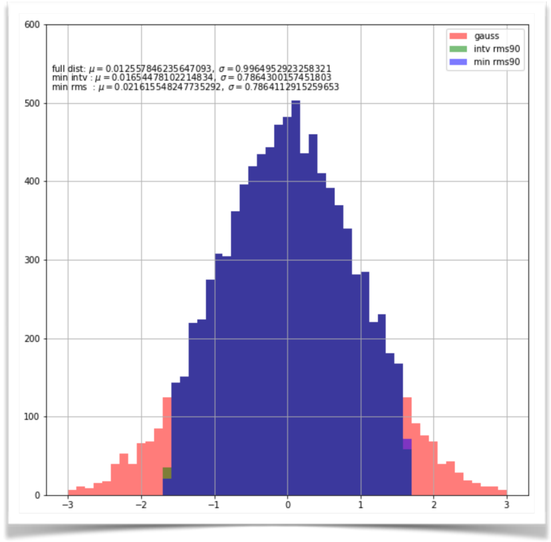
D.Jeans

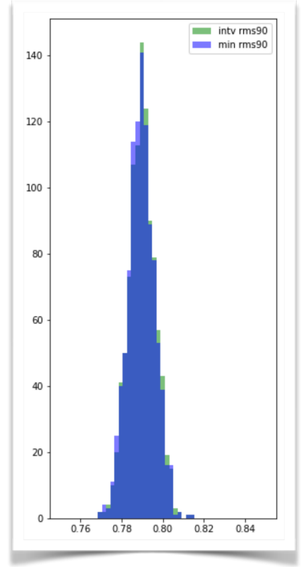
- started to investigate spurious crashes observed in running mass production
  - ~ one per mille of 2f -> had jobs crashed
- able to reproduce on one particular file (independently on different systems w/ different versions,…)
  - so far no conclusive results
    - intermittent crash w/o clear and changing stack trace, depending on random seed, debug mode, …
- given the low occurrence of this, SW convenors agreed that this is not a problem for current production

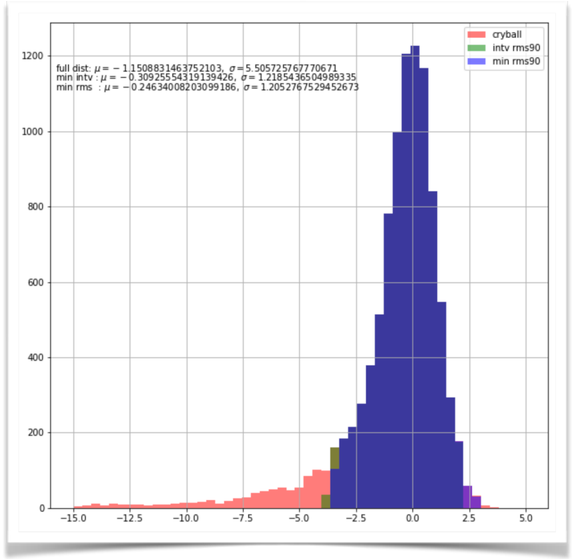
Reconstruction

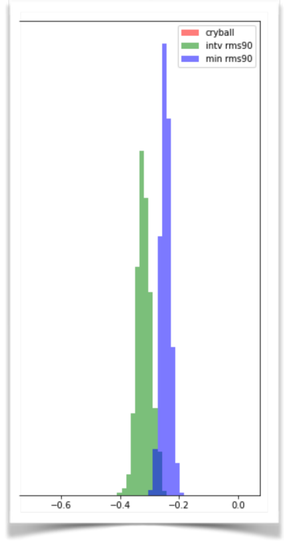
FG

- in context of generative machine learning for fast calorimeter shower simulation, started a discussion on definition of RMS90
  - definition: **RMS of smallest quantile that contains 90% of the data**
  - implementation: **smallest RMS** of any quantile that contains 90% of the data
- the implementation (used throughout the linear collider community, Pandora, ILDPerformance,….) is equivalent to the definition
  - for a symmetric distribution, eg. a Gaussian
  - **it is not for a skewed distribution**, e.g. a crystal ball function
- need to fix and check (likely small) effect on JER, etc









Monte Carlo Production

A.Miyamoto, H.Ono

Status

- New samples produced since last meeting
  - Higgs exclusive process is about to complete.
  - Stopped job submission in order to
    - Prepare next submission
    - Moving SIM files (~70TB) at DESY-SRM to KEK-SRM
      (Moving job is occupying DESY-KEK band width)
  - Latest information is in https://ild.ngt.ndu.ac.jp/elog/dbd-prod/
  - New web page for summarizing of DST-Merged files produced/planned
    - https://ild.ngt.ndu.ac.jp/mc-prod/prodmon/prodsum-mc2020.html
- Storage
  - About O(200 TB) of tapes added to KEK-SRM
    - ~ 380 TB free space left as of 2 Dec.

Yet to be produced

(A) Short & Small SIM file processes
- A couple of failed channels (qqh_bb.eL.pR, 5f, aa_4f)
- Flavortag-250-qq

(B) Channels of large cross section
- aa_2f : CPU ~5 kdays, SIM ~ 130TB
- 3f : CPU ~40 kdays, SIM ~ 270TB

CPU and SIM are resouces needed to process all remaining channels

(C) Rest of high cross section channel (done fraction is shown by %)
- 2f_had (2%): CPU ~203 kdays, SIM~ 2400TB
- 2f_lep (50%): CPU ~ 9 kdays, SIM~85TB
- 4f ww_h, W&Z_h, ww_sl, sw_sl (10%) : CPU 71 kdays
- 4f_sze_l (60%) : CPU 4 kdays

(D) Other 4f : negligible

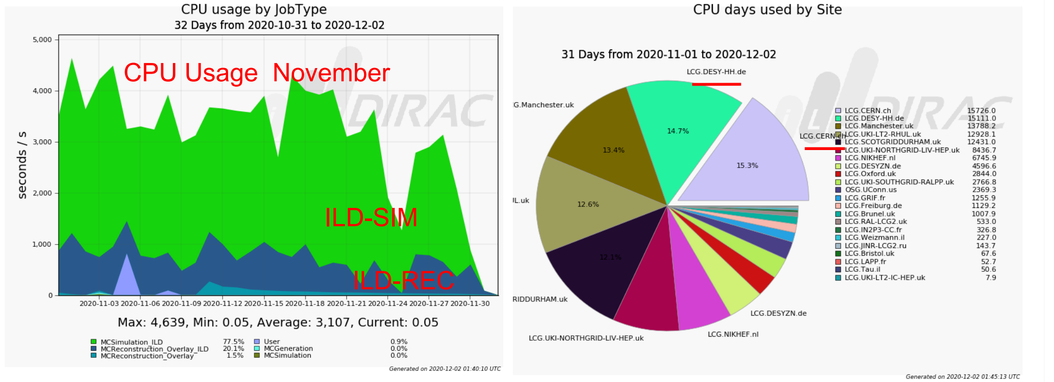
(E) Reconstruction of existing SIM files with o2 option

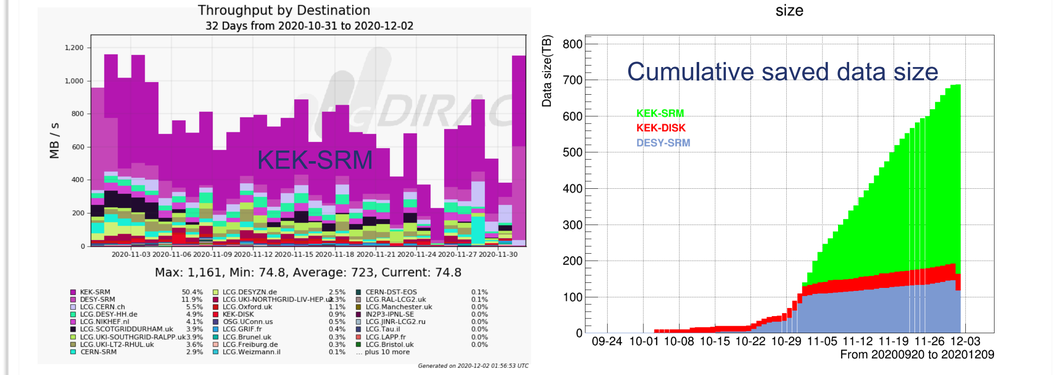
Plan and questions
- (A) & (D) : No big resource required. ➔ Produce them next, keeping all SIM
- (C) : Create o1 & o2 DSTs simultaneously. Neither SIM not REC are kept.
- (B) : Same as C), but keeps 5%(10% ?) SIM.
- (E) : Priority ? ~ 350 kdays to complete (A)~(D)

check status of production at this web page:
https://ild.ngt.ndu.ac.jp/mc-prod/prodmon/prodsum-mc2020.html

Monte Carlo Production

A.Miyamoto, H.Ono





Data written since mid. October

	Data Size (TB)			
mc-2020	Total	DESY-SRM	KEK-DISK	KEK-SRM
SIM	596.3	50.1	22.4	523.4
REC	31.8	30.7	1.1	---
DST-Merged	31.1	15.7	15.4	
Gen	11.7	5.8	5.9	---
Total	687.6	117.7	46.2	523.4

- might have to drop more than 10% of the SIM files
- started discussion with SDHCal group
  - plan for Higgs analysis w/ ILD_l5_o2_v02 option
  - … need to see how this could be incorporated …

ILD Analysis and ML with Julia

Tutorial by Jan Strube, PNNL

Tuesday, December 8
07:30 - 08:30 CET
15:30 - 16:30 JST
22:30 - 23:30 PST (Mon)

- Topics:
  - introduction to Julia
  - Julia on the DESY Jupyterhub
  - Julia on the DESY NAF / KEKCC as an alternative
  - first steps with LCIO (using the recoil mass / btag examples)
    - plotting with Julia
  - first steps with deep learning (using a simple calorimeter calibration as example)
- Q&A