The ILD Software Tools and Detector Performance.

ICHEP 2020
Prague, Czech Republic

Rémi Été, on behalf of the ILD concept group

DESY

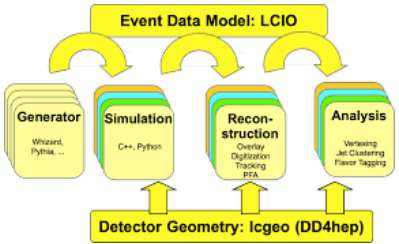
July 31, 2020

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

The iLCSoft software stack

https://github.com/iLCSoft

- Software stack of the ILC experiment
- Nowadays used by many other experiments/collaborations
  → e.g: CLICdp, CEPC, CALICE, LCTPC, EU-Telescope
- Maintained by FLC @ DESY and CLICdp @ CERN



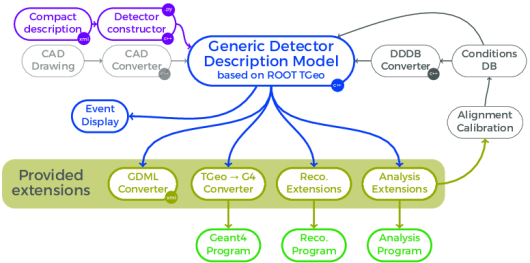
Main components

- **DD4hep**: Geometry description for simulation (Geant4) and reconstruction
- **LCIO**: Linear Collider IO and Event Data Model (EDM)
- **Marlin**: Reconstruction framework based on LCIO
- **PandoraPFA**: Particle flow reconstruction for Linear Colliders (`LCContent`)

DD4hep: detector geometry package

https://github.com/AIDASoft/DD4hep

- Generic detector description for HEP
- Single complete description source for
  - Simulation
  - Reconstruction
  - Analysis
- **DDG4** for simulation
  - Gateway to Geant4
  - Fully customizable: input / output, Geant4 actions, physics list, etc...
- **DDRec** for reconstruction
  - High level view of detectors: # layers, thicknesses, dimensions, etc...
  - Tracking surfaces, material properties, cellID converter



Philosophy:
single source of geometry, **different interfaces**

The ILD detector description

Optimizing ILD: ILD-L vs. ILD-S. arXiv:2003.01116

ILD detector(s) described in detail:

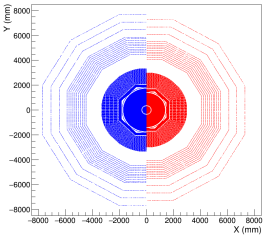
- Materials, extents, sensitive volumes, services, etc...
- ILD-S (small TPC radius) vs. ILD-L (large TPC radius)

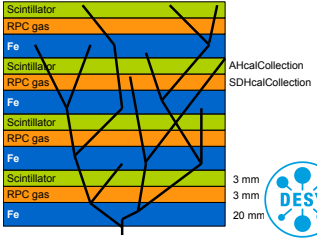
Hybrid simulation with 4 calorimeter options:

Detector	Si-ECal	Sc-Ecal	AHCal	SDHCal
ILD_l5_o1_v02	x		x	
ILD_l5_o2_v02	x			x
ILD_l5_o3_v02		x	x	
ILD_l5_o4_v02		x		x

- Simulate 4 options, reconstruct 1 option
- Save CPU time and minimize storage

For the ILD concept and technologies:
→ ICHEP talk `the ILD detector by T. Tanabe`

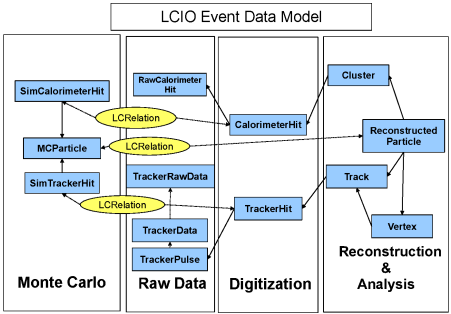




LCIO: the Linear Collider event data model

https://github.com/iLCSoft/LCIO

- Data handling for all steps in HEP workflow:
  - Generator, simulation, reconstruction, analysis
- Standalone IO library:
  - Binary data format (XDR), ZLIB compression
  - Schema evolution (block versioning)
  - Extensible and backward compatible format
  - Endianness agnostic (big / little endian)
  - Recently re-implemented for multi-threading usage
- Very robust: 20 years of usage...



The Marlin framework

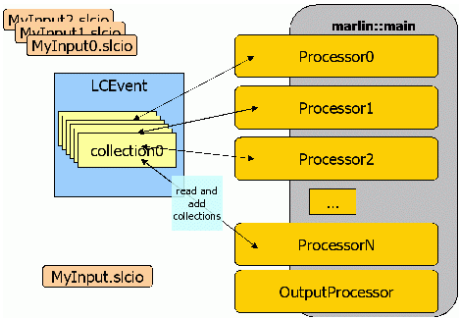
https://github.com/iLCSoft/Marlin

Standard HEP event processing framework

- Based on LCIO event data model
- Reconstruction and analysis
- Handles histogramming and conditions data

The Marlin framework

- Describes a task list (`Processor`) to execute
- Read events and process them through the chain
- Each `Processor` read and/or create new collections in the event
- Standard sequential event processing pipeline in HEP



The Monte Carlo mass production system

The production system

DIRAC system

- Job management, file catalog, ...
- Transformation system for productions
- Written in Python

iLCDirac system

- DIRAC extension for ILC/CALICE VOs
- Specific to iLCSoft applications
- Developed and operated by CLIC @ CERN

Main transformations:

1. GenSplit: split generator files
2. Simulation: runs `ddsim`
3. OverlayBKG: prepare reconstruction for bkg overlay
4. Reconstruction: runs `Marlin`
5. DSTMerge: merge DST files after reconstruction

The Monte Carlo mass production system

Recent data production: dataset and statistics

- Storage $\sim$ 1 PB
- Luminosity $\sim$ 500 fb^{-1}
- $E_{cms} = 500$ GeV

event class	description	events processed
2f	two fermion final states	60.0×10^6
4f	four fermion final states	22.6×10^6
5f	five fermion final states	4.01×10^6
6f	six fermion final states	13.8×10^6
aa_4f	two fermion by $\gamma\gamma$ interaction	1.63×10^6
higgs	higgs process	3.97×10^6
np	new physics process	3.25×10^6
aa_lowpt	$\gamma\gamma \rightarrow$ *hadrons* background	2.50×10^6
seeablepairs	e^+e^--pair background	1.00×10^5 BXs
calibration	single particle, $q\bar{q}$ events	27.71×10^6
6f(WW)	dedicated 6f sample at $E_{cms} = 1$ *TeV*	1.75×10^6

Table: Number of Monte Carlo events produced

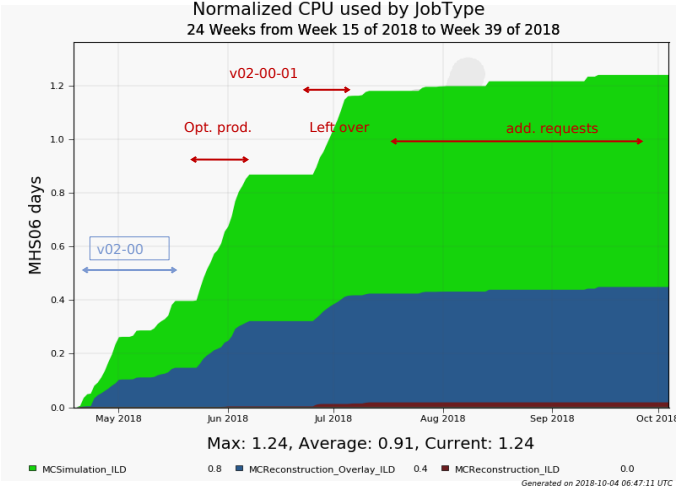
The Monte Carlo mass production system

Recent data production: cummulative CPU

Simulation

Reconstruction

Reconstruction (calib)



The ILD reconstruction chain

https://github.com/iLCSoft/ILDConfig

Background overlay

Digitization

Tracking

Particle Flow

High Level Reco

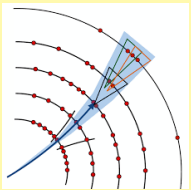
1 BX overlay

~ 1 BKG event / collision

Energy calibration

Inefficiency & noise randomization

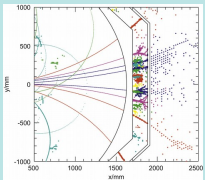
Time integration



TPC : Kalman filter

Vertex : Cellular automaton

Fwd detectors : Triplet seeds + extrapolation



PandoraPFA

Particle ID (de/dX + shower + MVA)

γγ-Finders (kinematic fit)

Primary vertex finder (LCFIPlus)

Time of flight (fast timing calorimeters)

The ILD detector performance

Tracking and jet energy resolution

Tracking

- Goal for momentum resolution:

$$\sigma_{1/p_T} \approx 2 \times 10^{-5}\ \text{GeV}^{-1}$$

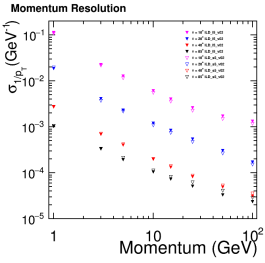
- Tracking efficiency close to 1!

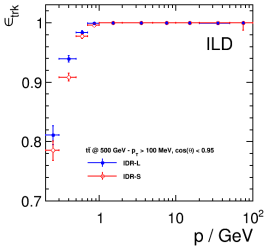
Jet energy resolution

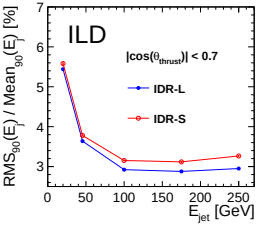
- Jet energy resolution (JER) defined as:

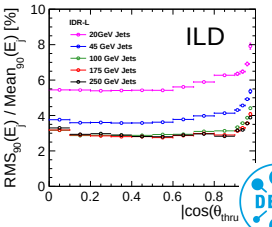
$$\frac{\sigma_{E_{jet}}}{E_{jet}} := \frac{\text{rms}_{90}(E_{jet})}{\text{mean}_{90}(E_{jet})}$$

- JER $\sim 3 - 4\%$
- Getting worse in the forward direction









The ILD detector performance

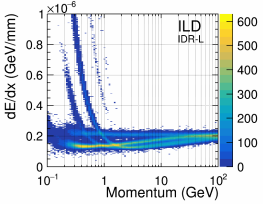
Particle ID and flavour tag

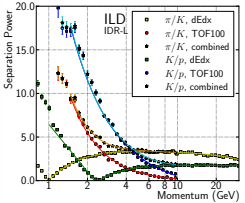
Particle ID

- TPC dE/dX with/without TOF
- Separation power:

$$\eta_{A,B}(p) = \frac{|\mu_A(p) - \mu_B(p)|}{\sqrt{\frac{1}{2}(\sigma_A^2(p) + \sigma_B^2(p))}}$$

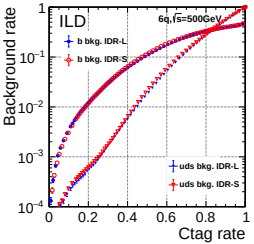
- Improvement at low P with TOF

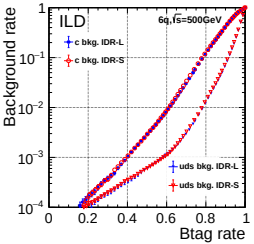




Flavour tag

- c and b jet identification crucial for physics analysis ($H \to c\bar{c}$, $H \to b\bar{b}$)
- Identification using BDTs





Conclusion and outlook

Conclusion:

- iLCSoft: a software stack for future colliders studies
  - Realistic full simulation and reconstruction
- iLCDirac: the ILC Monte-Carlo mass production software
- Recent MC production:
  - 500 GeV CMS, 1 PB produced
  - Learnt a lot about massive data production...
- Excellent ILD detector performance

Outlook

- Software tools evolving towards multi-threading
- Reconstruction performance: still place for improvement
  - Timing detectors?