The ILD Software Tools and Detector Performance.

ICHEP 2019
Prague, Czech Republic

Rémi Ete, on behalf of the ILD software group

DESY

July 31, 2020

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

AIDA 2020

ilc

DESY.

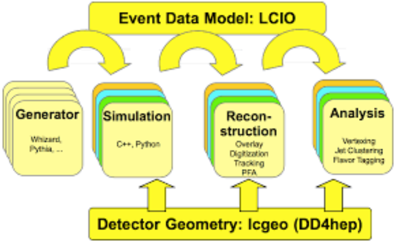
Outline

1. ILC software tools
2. ILD detector and event reconstruction
3. The IDR Monte Carlo mass production
4. Detectors performances
5. Summary and outlook

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



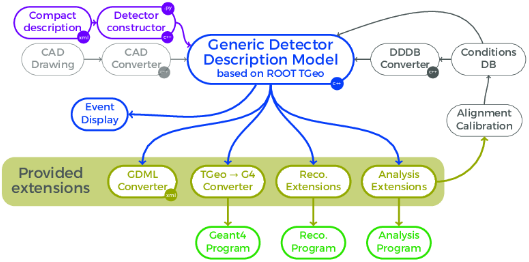
Main components

- **DD4hep**: Geometry description for simulation (Geant4) and reconstruction
- **LCIO**: Linear Collider IO and Event Data Model (EDM) (2003)
- **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

ILD detector(s) described in detail:

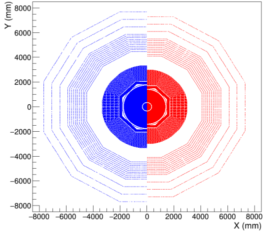
- Materials, extents, sensitive volumes, services, etc...
- ILD-S (small TPC radius) vs. ILD-L (large TPC radius)

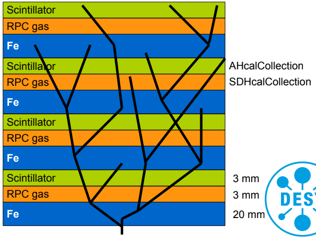
Hybrid simulation with 4 calorimeter options:

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

See detailed talk on the ILD detector by T. Tanabe

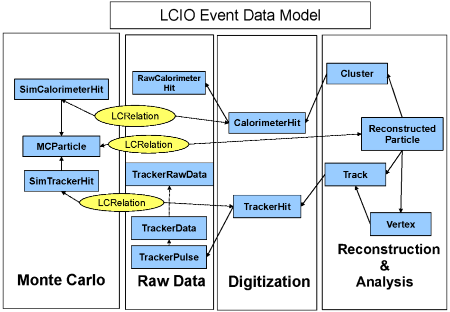




LCIO: the Linear Collider event data model

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
- Handles object relations
  - Weighted link between two objects
  - Very convenient for MC $\leftrightarrow$ Reco navigation



The Marlin framework

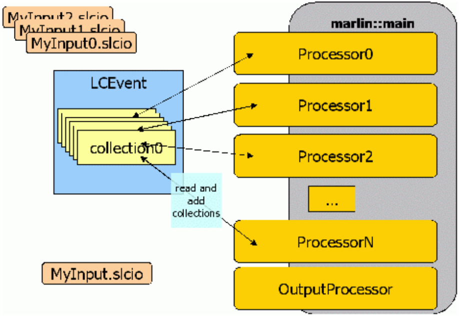
https://github.com/iLCSoft/Marlin

Standard HEP event processing framework

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

<u>The Marlin framework</u>

- 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 IDR Monte Carlo mass production

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. OverlyBKG: prepare reconstruction for bkg overlay
4. Reconstruction: runs `Marlin`
5. DSTMerge: merge DST files after reconstruction

The IDR Monte Carlo mass production

Dataset and statistics

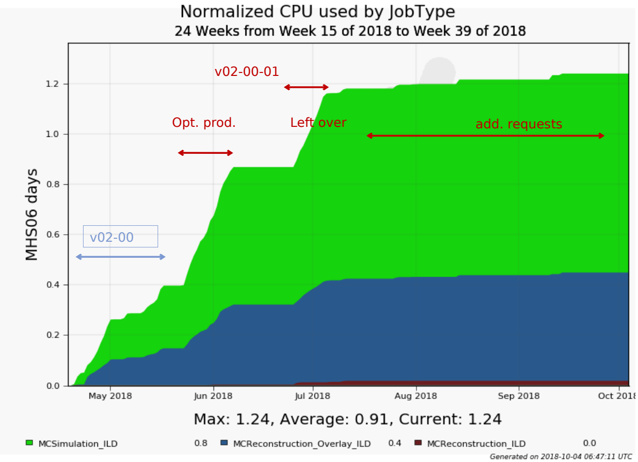
- Storage $\sim$ 1 PB
- Luminosity $\sim$ 500 fb^{-1}
- E_{cms} = 500 GeV

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 IDR Monte Carlo mass production

Cummulative CPU



The ILD reconstruction chain

https://github.com/iLCSoft/ILDConfig

- Background overlay
- Digitization
- Tracking
- Particle Flow
- High level reconstruction

The ILD reconstruction chain

https://github.com/iLCSoft/ILDConfig

- **Background overlay**
- Digitization
- Tracking
- Particle Flow
- High level reconstruction

- Overlay beam induced background with different probabilities
- Currently 1 BX overlaid
- Primary vertex of overlaid events are smeared

BKG source	Probability	Vertex z offset (mm)	Vertex z sigma (mm)
$\gamma\gamma$	0.350	0	0.1698
$e^+\gamma$	0.243	+0.0422	0.186
$e^-\gamma$	0.246	-0.0422	0.186
e^+e^-	0.211	0	0.1968
e^+e^- pair	1	0	0

The ILD reconstruction chain

https://github.com/iLCSoft/ILDConfig

- Background overlay
- Digitization
- Tracking
- Particle Flow
- High level reconstruction

Calorimeter digitization

- Apply energy calibration constants
- Emulate noisy / inefficient channels
- Treatment for timing
- Corrections for gap hits

Tracking detectors digitization

Detector	U resolution	V resolution
Vertex	3 μm	3 μm
SIT	5 μm	5 μm
FTD pixel	3 μm	3 μm
FTD strip	7 μm	7 μm
	R-ϕ resolution	Z resolution
TPC	0.05 mm	0.4 mm

The ILD reconstruction chain

https://github.com/iLCSoft/ILDConfig

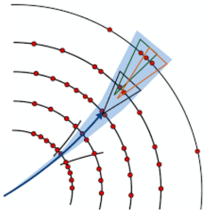
- Background overlay
- Digitization
- Tracking
- Particle Flow
- High level reconstruction

ILD track reconstruction performed in 3 independent steps

- VTX and SIT: triplet seed + extrapolation to next layers
- FTD: cellular automaton for track candidates
- TPC: outer pad rows topological clustering + Kalman filter

Finally:

- Track candidates and segments combination
- Final Kalman filter re-fit



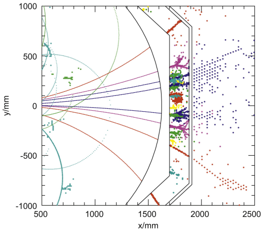
The ILD reconstruction chain

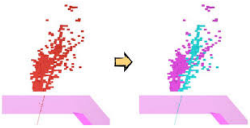
https://github.com/iLCSoft/ILDConfig

- Background overlay
- Digitization
- Tracking
- Particle Flow
- High level reconstruction

PandoraPFA reconstructs particles individually

- Calorimeter clustering algorithm
- Pattern recognition association algorithms
- Iterative re-clustering procedure
- Particle identification: e^{+-}, γ, n^0, μ^{-+}, π^{+-}





Reclustering

The ILD reconstruction chain

https://github.com/iLCSoft/ILDConfig

- Background overlay
- Digitization
- Tracking
- Particle Flow
- High level reconstruction

- Particle identification
  - Combination of TPC dEdX, shower shapes and MVA method
- $\gamma\gamma$-finders
  - Kinematic fits for π^0 and η-mesons identification
- MC thruth linking
  - Create links between MC objects to reconstructed objects
- Primary vertex finder
- TOF estimators
  - Compute various TOF estimates based on calorimeters information

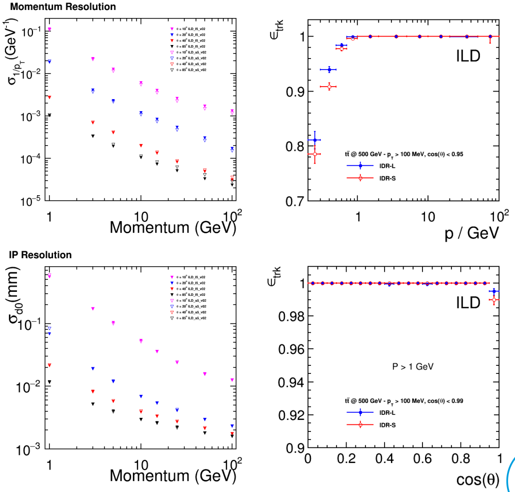
The ILD detector performance

Tracking: momentum resolution and efficiency

- Goal for momentum resolution:

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

- Momentum resolution
  - ILD-L slightly better in barrel
    $\rightarrow$ more hits available in the TPC
  - ILD-S slightly better in forward region
    $\rightarrow$ More curvature because B higher
- Tracking efficiency
  - Very close to 1!
  - ILD-L better at low momentum
    $\rightarrow$ Less curvature because B smaller



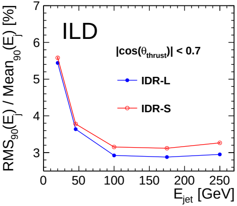
The ILD detector performance

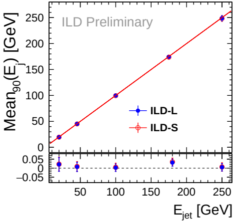
Jet energy: resolution and scale

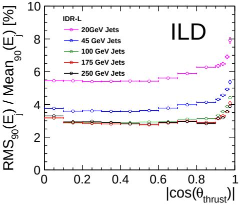
- Jet energy resolution defined as:

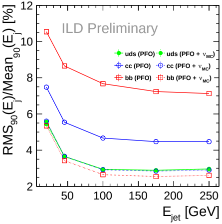
$$\frac{\sigma_{E_{jet}}}{E_{jet}} := \frac{\mathrm{rms}_{90}(E_{jet})}{\mathrm{mean}_{90}(E_{jet})} \quad (1)$$

- Better than 4% for $E_{jet} \geq 45\,GeV$
- Approaching 3% (3.2%) for ILD-L (ILD-S)
- Getting worse in the forward direction
- Jet energy scale better than 5%
- Effect on heavy quark flavor visible:
  - Missing energy due to neutrinos
  - Adding neutrino MC energy → recovery!









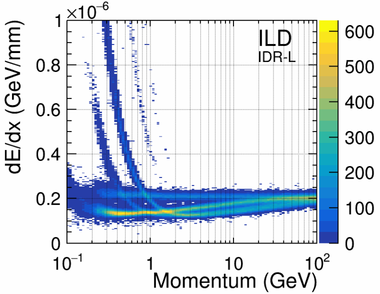
The ILD detector performance

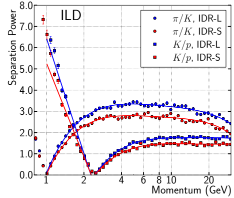
Charged particles identification: dEdX and TOF

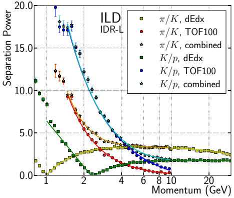
- dEdX in the ILD-TPC
  $\rightarrow$ a powerful tool for PID
- Separation power:

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

- Combination with TOF:
  $\rightarrow$ improvment at low P
- Huge potential with *timing-detector*



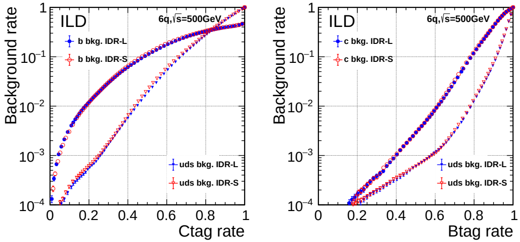




The ILD detector performance

LCFIPlus: flavor-tag performance

- c and b jet identification crucial for physics analysis ($H \to c\bar{c}$, $H \to b\bar{b}$)
- Identification using BDTs
- Trained with $e^+e^- \to 6\ q$
- No difference between ILD-L and ILD-S
- Results varying as a function of jet energy and multiplicity
- Re-training needed for specific event topology



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
- IDR 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
- Detector performance: still place for improvment