Generator samples for the ILD optimization

Philip Bechtle

DESY

January 15th 2008

Outline

Introduction

- Every Analysis used in the context of the optimization will need a good (fully simulated) sample of the SM
- Hence:
  - Provide the SM sample centrally, if possible simulated (with different detector setups) and reconstructed (up to a default PF?)
  - Provide all information and tools necessary to produce specific signal samples individually with exactly the same setup as the SM sample

Background

- Tim Barklow and Norman Graf in a great effort have produced a 2nd generation of a complete SM sample for 500fb^{-1} at all 4 polarizations
- There are very good reasons to use this sample:
  - Whizard is a multi-purpose ME generator. That means:
  - Signals and backgrounds of all types (SM + MSSM) can be produced with the same settings
  - It contains all interferences, hence it is more accurate than generators like Pythia, especially for complex final states (6f and more)
  - Tuning it, on the other hand (Correct inclusion of FSR, gluon radiation, etc), is not so easy and not as well performed as for other generators which have been used at LEP
- A validation of this sample was performed at SLAC and within our efforts

Scope of the Study

- No single MC generator is optimal for everything
- Some conceptual inaccuracies are present in the Whizard sample. Most of these are connected with

  Overlap between fragmentation and the hard ME

  which can cause doublecounting e.g. between

  $ee \to qqqq$ and $ee \to qq\,(+g \to qq$ from frag.$)$

- We have tested that the following goals are met for the processes in question:
- Physical distributions are correct
- Some degree of wrong normalisation would be acceptable for the purpose of optimisation

Outline

Main issues of Discussion: Fragmentation

- Fragmentation
  User defined fragmentation with full fledged Pythia gluon radiation and Jetset fragmentation is implemented
- For details of how this is done, see

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/ilc_fragment_call.f90
- This has been tested by Tim Barklow, the RHUL Group and Marco Verzocchi
- Comparisons witl LEP-tuned Pythia have turned out successfully
- More details maybe in Status of the ZHH analysis in the Analyses 2 session

Main issues of Discussion: Higgs

- A 120 GeV Higgs has originally been included in the simulation
- Since Whizard is a ME generator, this introduces a non-removable contribution to a variety of many final states
- Hence, currently all affected final states are reproduced with $m_h = 2$ TeV
- Higgs Final states can then be added with masses chosen by the analysist
- Samples expected to be available in 3 to 4 weeks

Main issues of Discussion: Tau

- Tauola was not included for the τ final states
- Good enough for the use of the sample as background, if the signal does not use specific τ properties
- Also due to the unfortunate budget cuts, Tim does not have the possibility to set up the user defined fragmentation procedures for all τ final states with Tauola
- Tim would produce the samples if we provide him with a working version of

  `ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/ilc_fragment_call.f90`

  including `tauola`
- Volunteers?

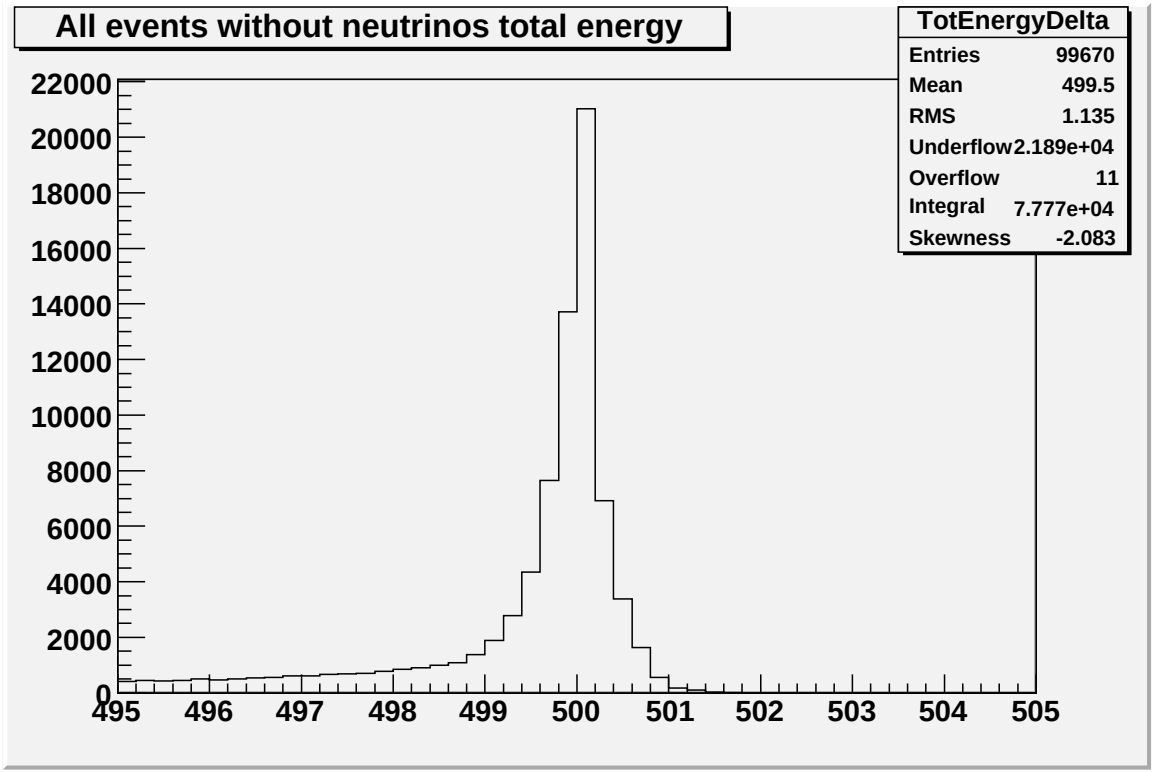
Main issues of Discussion: Tau

- Tauola was not included for the τ final states
- Good enough for the use of the sample as background, if the signal does not use specific τ properties
- Also due to the unfortunate budget cuts, Tim does not have the possibility to set up the user defined fragmentation procedures for all τ final states with Tauola
- Tim would produce the samples if we provide him with a working version of

  `ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/ilc_fragment_call.f90`

  including `tauola`
- Volunteers? If not, DESY will work on that.

Beamstrahlung and Beam Energy Spread



- Beamstrahlung and initial beam energy spread is calculated using `Guinea Pig` and included as documented in

http://confluence.slac.stanford.edu/display/ilc/Standard+Model+Data+Samples

Additional Issues

- p_T for ISR is included
- Double counting in 4e and eeqq final states – the cuts placed on m and q^2 reduce this to acceptable level
- CKM Matrix is unity. Any experience from WW analysis as to how bad that is for our purpose? Probably OK for many optimization studies
- Exact calculation of γ Final state radiation not included for Bhabha scattering – if someone wishes to to studies of e.g. luminosity measurements with Bhabhas, the Bhabha signal needs to be produced differently for these studies
  Bhabhas good enoug hfor backgrouns

Outline

Availability

- Overview and all information needed for the setup of Whizard

  http://confluence.slac.stanford.edu/display/ilc/Standard+Model+Data+Samples
- Comlete mixed SM sample

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/
- Individual Processes in all four polarisation states separately

  ftp://ftp-lcd.slac.stanford.edu/ilc/whizdata/ILC500/
- Log files etc. of the individual processes

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/run_output/
- If everything goes to plan: The simulated (and reconstructed?) events will be available at

  http://www-flc.desy.de/simulation/database/

  See Ivan Marchesinis talk for that

Production of Signals

- Signals outside of the SM sample have to be produced individually
- The site

  http://confluence.slac.stanford.edu/display/ilc/Standard+Model+Data+Samples

  provides all necessary files to set up Whizard in exactly the same way as done for the SM sample
- The information from

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/whizard-src/user.f90

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/guinea-pig/ilc_0500_may05_run05_seed06/

  can be used to set the same beam structure for any other generator
- Information from

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/ilc_fragment_call.f90

  ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/calc_a1sq_a2sq.f90 can be adapted to set up the fragmentation also for other generators
- The resulting generated events can be made available together with the SM sample at

  http://www-flc.desy.de/simulation/database/

Simulated Events

(Too?) Aggressive proposal for production in the rough order of priority:

possible signals or backgrounds:		≈ No. Events
$ee \rightarrow 4f$	50fb-1	5M
$ee \rightarrow 2f$	20fb-1	2.5M
$ee \rightarrow 6f$	50fb-1	100k
$ee \rightarrow hX$	50fb-1	75k
calibration samples:		
light quark 2f at 91.2 GeV	20 000 events	
tt (6f) at 350 GeV	20 000 events	
backgrounds:		
$\gamma\gamma \rightarrow X$	0.1fb-1	1M
$ee \rightarrow \gamma\gamma(n * \gamma)$	10fb-1	0.5M
$\nu\nu(n * \gamma)$	20fb-1	1.5M
$ee \rightarrow ee$	0.1fb-1	0.2M
$e\gamma \rightarrow e\gamma$	0.1fb-1	0.6M
rest	1fb-1	0.6M

How to Produce and Provide the Simulated Samples?

- The luminosity goal on the previous slide is somewhat agressive for one detector
- It is most probably too agressive for optimization
- My personal recommendations:
  - Produce as much as possible of the background mentioned above for one detector model
  - Start analyses on that set
  - Determine exact needs for specific backgrounds for each important optimization analysis
  - Do one dimensional scan (e.g. fixed coil energy content) with a background set tailored to analyses
  - In the minimum of the 1D scan, do 2D/3D scan
- How to do the production technically: See Ivan Marchesinis talk Grid tools for Mokka and Co in the Tools for optimizing ILD session

Summary and Recommendation

- Go for one generator for most final states
  - The work to validate several individual generators against each other by far exceeds our manpower and timescale
  - We have to make sure that our selections select final states and not generators
- My recommendation would be to to use a matrix element generator
  - 6f final states very important for optimization, useless without ME generator like Whizard
- Hence, we're going to use the SLAC sample
  - The machinery for producing hundreds of final states is in place
  - No reason to start from scratch even if included Higgs/Tauola makes re-generation of some part of the sample necessary/desireable
- Sample has been checked extensively, and a machinery for centralized production of simulated events is in place