Requirements for a full simulation analysis on Zhh

Michele Faucci Giannelli, Mike Green,
Veronique Boisvert, Fabrizio Salvatore, Tao Wu

Studying the Zhh process

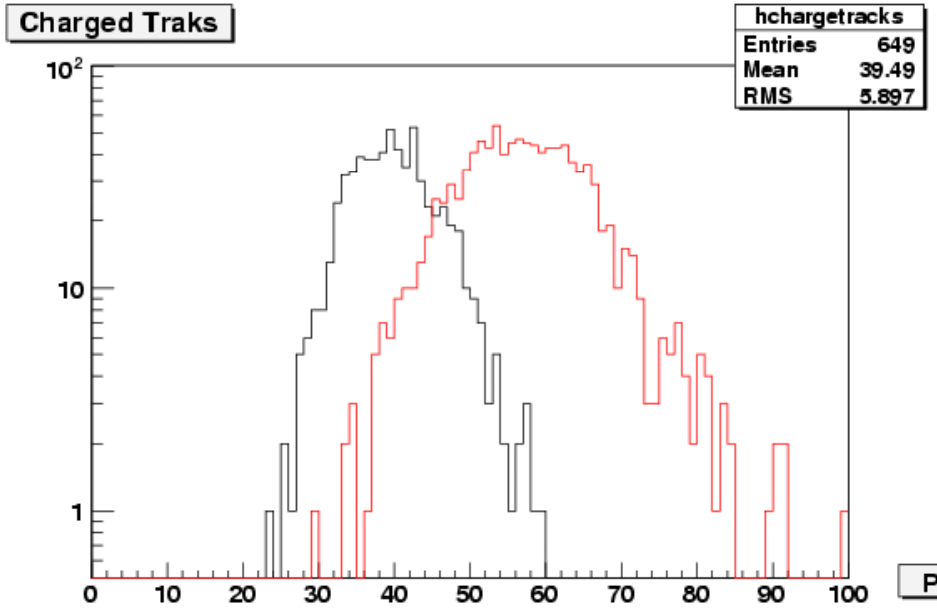
- Zhh is one of the main channels for ILD optimization:
  - It is a good example of precision physics at ILC
  - Having a complex final state can be used to test software and detector performances
- The cross section is very small: 0.16 fb^{-1}
  - The highest BR channel is Z→ qq, h → bb for a total of 29 events for 500 fb^{-1} of data
- Main background: $t\bar{t}$ which has a cross section of 720 fb^{-1} (4500 times the signal !!!)
- Other backgrounds: ZZh, WWZ, tth, Zh

Royal Holloway
University of London

What generator ?

- Generators to be used to simulate signal/bg events have to be decided and fixed ASAP.
- There are several generators available:
  - Pandora Pythia (PP)
  - WHIZARD
  - Other (Sherpa, HERWIG, Grace, …)
    - Not tested in this analysis
- Main difference observed so far: gluon emission
  - WHIZARD has no gluon emission by default
    - Potentially, incorrect multiplicity distribution
  - Pandora Pythia has correct gluon emission (LEP tuned)
- WHIZARD is important mainly for simulating three boson channels that do not contain the Higgs because PP do not include them
  - ZZZ, WWZ, ttZ, (tbW, ttH)
  - Need comparison between generators to assess which one is the right one to use

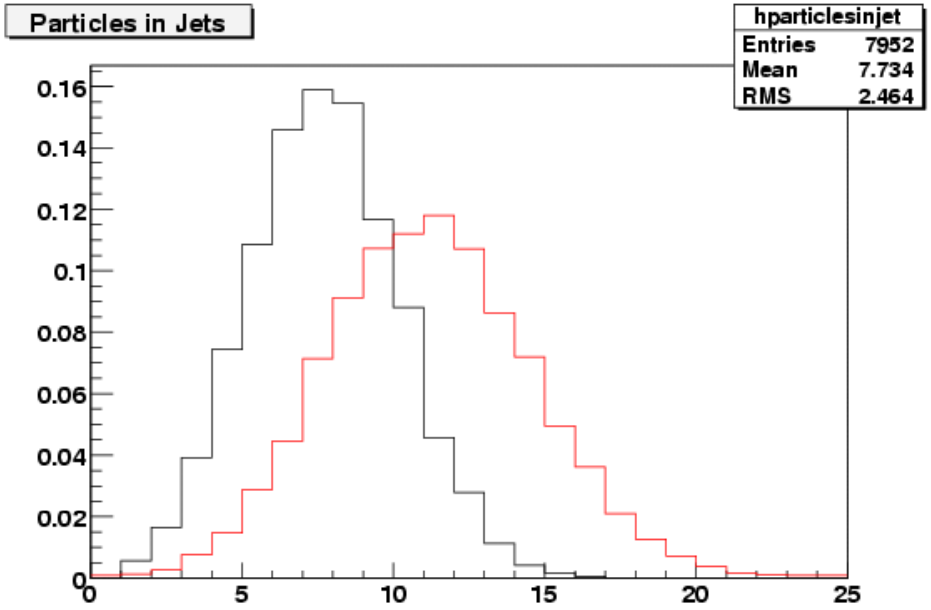
Pandora Pythia vs WHIZARD - I



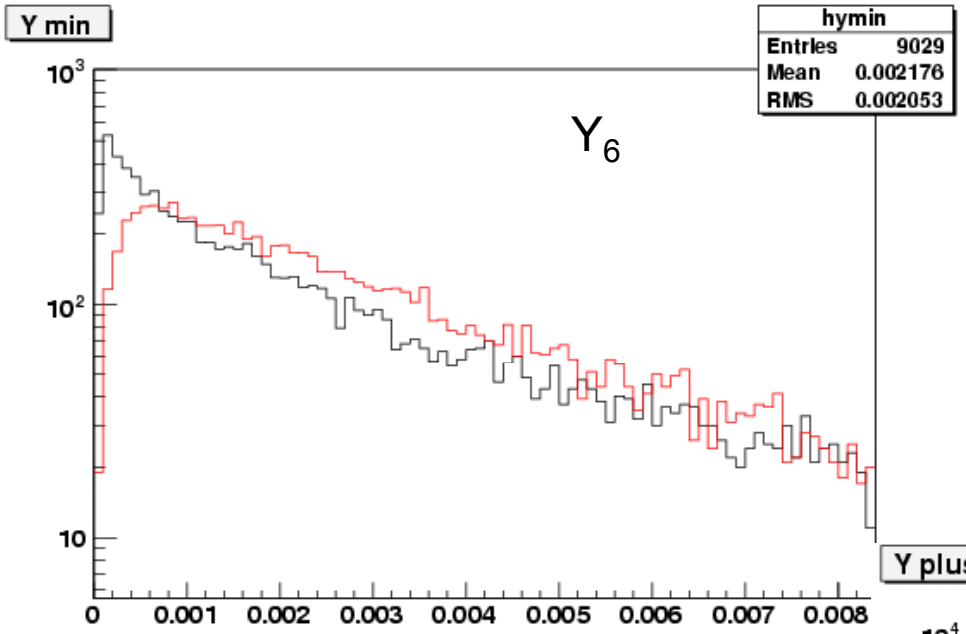
Charged track multiplicity distribution

Zhh from WHIZARD

Zhh from PP

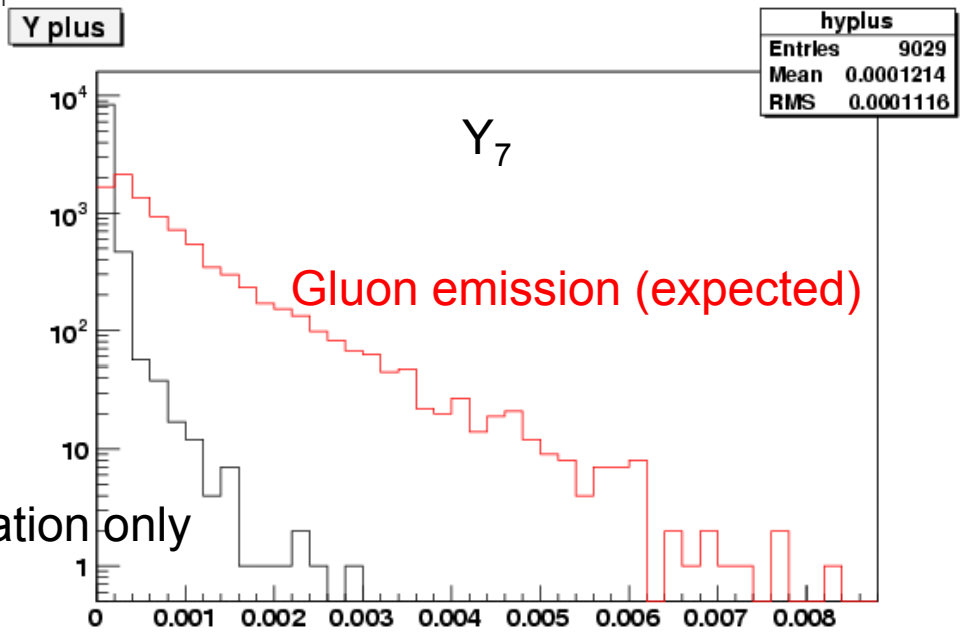


Pandora Pythia vs WHIZARD - II



Distribution of and Y_6 and Y_7

Y_n is the minimum Y value for which an event is reconstructed as an "n jet" event



Zhh from WHIZARD

Zhh from PP

Gluon emission (expected)

Hadronization only

Preliminary comments - I

- Some comments from preliminary comparison between PP and WHIZARD
  - Missing gluon emission in WHIZARD reduces multiplicity distribution in jets → confusion term in jet energy resolution is artificially smaller
  - Default configuration of PP and WHIZARD foresee that all particles are decayed by the generator
    - OK for fast MC simulation (generation + parametric detector smearing), but not correct for full simulation analysis
      - long lived particles (e.g. π, K, Λ, Σ, Ω) have to be decayed by in the detector
- WHIZARD v1.50 has incorrect implementation of CKM matrix
  - Only diagonal terms of the matrix are present (and ≡ 1 !)
    - Wrong W decays
- This "feature" has been corrected in the latest version of WHIZARD (1.51)
  - Need to use events generated with updated version of the generator

Preliminary comments - II

- 2 ab^{-1} sample produced at SLAC using WHIZARD potentially has all "features" described above
  - Need to check with authors before using them for massive MC production
  - SLAC events generated with generator-level cuts a' la SiD (geometrical acceptance) → potential 'bias' when used for ILD ?
- WHIZARD can generate six fermions final states, but is very complex to retrieve intermediate states (e^+e^-→"ZHH"→qqbbbb)
  - Should generation be made "by channel" instead of "by final state"?
- Pandora Pythia is not developed anymore and it does not include some of the 6 fermion channels
  - Use a mixture of the two generators
- Which beam polarization do we use?
  - 0 e^+, 80% e^- (standard Tesla) ?
- Which beam setup do we use ?
  - 500 GeV, 350 GeV ? Which model ?

Samples generated at RHUL

- Generator used: Pandora Pythia
  - ~ 20K events generated in WHIZARD v1.51 to perform comparison showed in previous slides
- Beam line: NLC500
- Polarization: 80% (e^-), 0% (e^+)
- ISR and beamstrahlung ON
- Higgs mass: 120 GeV
- CM Energy: 500 GeV
- Particles not decayed in generator:
  - $\pi^{\pm}$, $K^{\pm}$, K_s, K_l, Λ, $\Sigma^{\pm}$, Ξ^-, Ξ^0, Ω^-
- Mokka v06-04
  - Detector model: LDC00Sc
  - Physics list: LCPhys

Cross sections

Event type	σ (fb)	Events/500fb^{-1}	Generated events (PP)	Simulated events (Mokka)	% of available events/500fb^{-1}
Zhh (tot)	0.16	80			
Zhh→qqbbbb	0.0593	30	1000	1000	3375
ttbar (lept)	73	36500	100000	10000	27
ttbat (mixed)	310	155000	100000	35000	23
ttbar (cqcq)	82	41000	200000	10000	24
ttbar (uquq)	82	41000	200000	10000	24
ttbar (cquq)	164	82000	300000	10000	12
bbh	10.6	5300	30000	16000	302
ZZh	0.174	87	1000	1000	1150
ZZZ	1.05	525	0	0	0
WWZ	35.3	17650	0	0	0
tth	0.15	75	0	0	0
ttZ	0.7	350	0	0	0
tbW	16.8	8400	0	0	0

Not included or have problem in Pandora Pythia

Time needed for simulation (500fb^{-1})

- Used RHUL farm (SLC3)
  - 3.1 GHz Xeon, 1GB ram, jobs submitted to PBS (no grid submission yet)
- Simulated events divided in jobs of 100 events
- Mokka simulation:
  - a 6 jet job requires ~ 40h to be processed
  - a 4 b-jet job requires ~ 25 h
  - a 2 light + 2 b jet job requires ~ 20 h
  - a 2 b jet job requires ~ 10 hours
- For 500 fb^{-1}:
  - 160k 6-jets events → 1600 jobs → 64000h
  - 160k 4-jets events → 1600 jobs → 40000h

  Total of ~100000h with 100 cpu → 42 days

Time needed for reconstruction (500 fb^{-1})

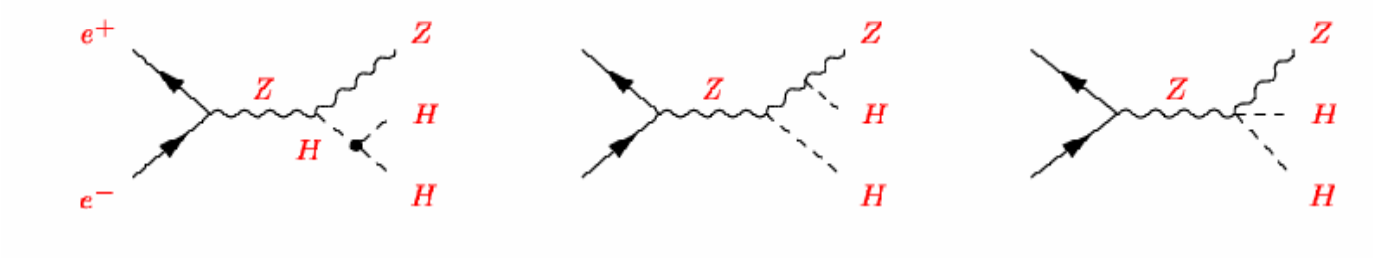
- Use Marlin for reconstruction
  - TrkCheater + Pandora + LCFI
  - 2h per job
    - Loosely dependent on number of jets
- 3200 jobs → 3 days
- Using the Grid tools would reduce the time and allow simulation of much larger sample (1 ab^{-1} would be desirable !)
- A second PFA (TrackBased) for comparison would require a double reconstruction time

Royal Holloway
University of London

Status of the analysis on Zhh to 6 jets at RHUL

Zhh physics

- Self coupling of the Higgs can be measured from the first diagram



- The goal is to repeat the analysis performed at generation level with full simulation
  - is 20% resolution reachable?
- This is precision physics, can not be done at LHC
- BR (Z→qq) 70%
- BR (h→bb) 73%
- Main channel is qqbbbb (40%)
  - νν̄bbbb (16%)
  - qqbbWW (12%)
  - llbbbb (only 4.5%)

Reconstruction

- Reconstruction using Marlin and MarlinReco:
  - TrackCheater
  - Pandora v1.01
  - LCFI v1.0
- Goal is to use:
  - FullLDCTracking
  - Pandora v2.0
  - New version of TrackBasePFA
  - LCFI ??
- Merging of all output in one single file per channel
- Use shape variables to study possible cuts
- Analysis processors

Possible cuts

- Topological cuts:
  - $\mathrm{Cos}(\theta_{thrust})$
  - Thrust
  - Fox-Wolfram moments (R1 and R2)

  } back-to-back vs spherical events
- Missing energy:
  - |P(z)|
  - Total reconstructed energy
- 2 and 4 jets events can be rejected using:
  - Jets EnergyEM/Energy
  - Jet number of particles
  - Y_6
  - Number of charged tracks
- B tagging will play a central role

ttbar rejection

- A possible S/B separation can be performed using kinematic fitting (min χ^2):

mass terms

$$\chi^2 = \frac{\left(M_{12}^{fit} - M_Z\right)^2}{\Gamma_Z^2} + \frac{\left(M_{34}^{fit} - M_H\right)^2}{\Gamma_H^2} + \frac{\left(M_{56}^{fit} - M_H\right)^2}{\Gamma_H^2} +$$

$$\sum_{i=3,4,5,6} \frac{\left(NNbtag(i) - 1\right)^2}{\sigma_{btag}^2} + \sum_{i=1}^{6} \frac{\left(E^{fit} - E^{reco}\right)^2}{\sigma_{ene}^2} + \frac{\left(\sum_{i=1}^{6} E_i^{fit} - 500\right)^2}{\sigma_{beam}^2}$$

b-tagging

Jet energy constraint

Total energy constraint

Summary and Outlook

- Several events generated (and reconstructed) at RHUL using Pandora Pythia (PP) are available
  - Contact Michele/Fabrizio for more info
- Preliminary comparison of events generated with PP and WHIZARD
  - Some "features" of WHIZARD (e.g. gluon emission) need to be taken into account before using it for mass MC production
- Need to contact SLAC people to understand if $2ab^{-1}$ of events produced there are usable for ILD optimization production
  - Possible problem with CKM matrix implementation
  - Generator-level cuts a' la SiD
- If the reconstruction should start in two weeks, we propose to use a combination of PP and WHIZARD to generate the events needed:
  - if there are doubts on the SLAC sample we can provide ttbar events from PP to start
  - We should aim for more than $500fb^{-1}$, a solid analysis needs at least $1ab^{-1}$

Summary and Outlook

- RHUL available for testing the files produced on the GRID
- The analysis on Zhh → 6 jets is ongoing
- Pre-selection almost complete
  - Processor to optimize the cuts is in place
- Several background events have been generated and reconstructed
  - Used for studies at generator level
  - Used for studying preliminary cuts
- Looking into using kinematic fitting for ttbar rejection