Standard Model Data Samples

2000 fb $^{-1}$ SM Data Samples at Ecm= 250, 350, 500 & 1000 GeV

Introduction

Stdhep files for 2000 fb $^{-1}$ Standard Model data samples at Ecm= 250 & 500 GeV have been produced and are currently located on SLAC mass storage. Some
Standard Model data samples at Ecm= 350 GeV have also been produced. A complete list of the processes can be found in ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC250/doc/integ_index_0250_01,
ftp://ftp-lcd.slac.stanford.edu/ilc3/whizdata/ILC350/doc/integ_index_0350_SB2009_nTF_extbunches_01, and ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/doc/integ_index_0500_01 . The five columns are process_id, initial_state,
the variable IDRUPLH, an internal sequence number (1 - 4) and a bit indicating whether or not events were produced for this particular initial state polarization sign combination (0 = events generated, 1= events not generated).
Subsets of these events are available via ftp; please see the table of derived stdhep files.

The events are produced assuming 100% polarization for the initial state electron and positron; different files for the same final state correspond to different polarization sign combinations. The variable IDRUPLH indexes the different final states and polarizations. Assume, for example, that a process has IDRUPLH=14995 . The stdhep file is ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/w14995_01.stdhep and the information about the generation of this file can be found in the directory ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/doc/run_output/w14995/run_01/ .
The log file is ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/doc/run_output/w14995/run_01/whizard.log ,
the whizard input file is ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/doc/run_output/w14995/run_01/whizard.in
and cross section information is in ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/doc/run_output/w14995/run_01/whizard.n3n3n3n3ss_o.out. For Ecm= 350 GeV, the event generation log files can be found in ftp://ftp-lcd.slac.stanford.edu/ilc3/whizdata/ILC350/doc/run_output .

The WHIZARD Monte Carlo version 1.40 is used for parton generation for all processes except SUSY processes. WHIZARD version 1.51 is used for SUSY processes. For processes with (without) an on-shell Higgs boson in the final state the Higgs mass is assumed to be 120 GeV (2000 GeV). The Makefile and build log files for this implementation of WHIZARD can be found in ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/doc/whizard-v1r4p0 .

Electron/Positron Beam Properties: Beamstrahlung and LINAC Energy Spread

The following lines in whizard.in control the properties of the colliding electron/positron beams:

USER_spectrum_on = T
USER_spectrum_mode = -2

The first line indicates that a user-supplied function is used to simulate the beams. A copy of this function can
be found in ftp://ftp-lcd.slac.stanford.edu/ilc3/whizdata/whizard/v1r4p0/whizard-v1r4p0/whizard-src/user.f90 for the WHIZARD 1.40 version and in ftp://ftp-lcd.slac.stanford.edu/ilc3/whizdata/whizard/whizard-1.95/whizard-src/user.f90 for the WHIZARD 1.95 version. This function refers to links with the names ... _linker_000... which can be found in
ftp://ftp-lcd.slac.stanford.edu/ilc3/whizdata/whizard/guinea-pig/energy_spread.

The absolute value of USER_spectrum_mode determines which energy spectrum is used, with the sign +/- indicating
electron/positron beam, respectively. For the Ecm = 500 GeV SM data sample this absolute value is always 2, and corresponds
to the Guinea-Pig data contained in the directory ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/guinea-pig/ilc_0500_may05_run05_seed06/
This spectrum represents the default ILC design for Ecm=500 GeV circa August 2005, and includes both incoming LINAC
energy spread and beamstrahlung.

For the Ecm = 250 GeV SM data sample the absolute value of USER_spectrum_mode is 4 for the LOI samples with do_isr = T, and USER_spectrum_mode is 5 for the post-LOI samples with do_isr = F. The Guinea-Pig data for USER_spectrum_mode=5 is in ftp://ftp-lcd.slac.stanford.edu/ilc3/ILC250/LOI_backgrounds/guineapig/postLOI/guineapig/lumionly/raw . USER_spectrum_mode=6 corresponds to SB2009_350_nTF_extbunches. USER_spectrum_mode=7 corresponds to SB2009_500_nTF_extbunches. To summarize, the USER_spectrum_mode numbers, aka the Lumi_linker numbers, are as follows:

Lumi_linker number	Ecm(GeV)	General Description	Machine Configuration
2	500	RDR (Jul 2005)	rdr
3	350	RDR (Aug 2005)	rdr

4	250	RDR (Aug 2008) but do_isr=T (ISR turned on by mistake)	rdr_isr_on
5	250	RDR (May 2009) (note: beams 1 & 2 are swapped, see user.f90)	rdr_beams_swapped
6	350	SB2009_350_nTF_extbunches	sb2009_ntf
7	500	SB2009_500_nTF_extbunches	sb2009_ntf
8	350	SB2009_350_TF_extbunches	sb2009_tf
9	500	SB2009_500_TF_extbunches	sb2009_tf
10	3000	CLIC_July_2010_C++	clic_cplus
11	3000	CLIC_Aug_2010_C_Schulte	clic_schulte_aug2010
12	1000	ILC_1000_with_TF_Aug_2010	ilc_tf_aug2010
13	500	CLIC_500_Feb_2011_Schulte	clic_shulte_feb2011
14	1000	ILC_1000_5pcBS_no_TF_Sep_2011	5pcBS_notf
15	1000	ILC_1000_10pcBS_no_TF_Sep_2011	10pcBS_notf
16	1000	ILC_1000_B1b_with_TF_Nov_2011	B1b_tf
17	1500	CLIC_1500_Nov_2011	clic_1500_nov2011
18	1000	ILC_1000_Waisty_opt_Jan_2012	B1b_ws
19	1400	CLIC_1400_Jan_2012	clic_1400_jan2012
20	350	CLIC_350_Apr_2012	clic_350_apr2012
21	500	ilc_500_waisty_250_jan_2012	TDR_ws
22	250	ilc_250_waisty_250_jan_2012	TDR_ws
23	350	ilc_350_waisty_250_jan_2012	TDR_ws
24	250	ilc_250_cr_A_jun_2019	eps_x=0.5*TDR

Final State Parton Showering and Fragmentation

PYTHIA 6.205 is used for final state QED/QCD parton showering and for the fragmentation of quarks and gluons. Parton showering is performed for all final state fermions with the exception of electrons. Final state QED showering of electrons is normally turned off because the PYTHIA final state showering code indiscriminately uses the invariant mass of final state fermion-antifermion pairs for the maximum virtuality scale; however, PYTHIA final state QED showering of electrons was turned on for all Ecm = 250 GeV processes so that its effects could be studied in e+e- --> e+e-H. Otherwise default parameters are used for parton showering and fragmentation.

The interface to PYTHIA is contained in ftp://ftp-lcd.slac.stanford.edu/ilc3/whizdata/whizard/whizard-1.95/whizard-src/user.f90 . The source code for referenced subroutines can be found in ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include .

Color flow information was not available in WHIZARD 1.40, and so kinematic and parton id information is used to identify color singlet systems (see ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/ilc_fragment_call.f90 and ftp://ftp-lcd.slac.stanford.edu/ilc/ILC500/StandardModel/a6f/include/calc_a1sq_a2sq.f90).

Kinematic Cuts

Kinematic cuts are applied to massless particle configurations where some cutoff is required to avoid an infinite cross section. In the WHIZARD event generation stage we set the masses of all first and second generation fermions to zero, so that photons, gluons, electrons, muons and u,d,s,c quarks are all subject to kinematic cuts. The cuts are given by the WHIZARD input parameters *default_jet_cut*, *default_mass_cut* and *default_q_cut* . We use the WHIZARD default value
default_jet_cut=10 GeV for the minimum invariant mass of a pair of colored particles. We use *default_mass_cut*=4 GeV for the minimum invariant mass of a pair of colorless particles, and we use *default_q_cut*=4 GeV for the minimum sqrt(-Q**2) for
massless t-channel processes.

The massless assumption for the first and second fermion generations can produce some odd results given the kinematic cut values we have chosen. For example, the cross section for gamma gamma -> tau tau is significantly larger than the cross section for gamma gamma -> mu mu since the cut value of 4 GeV for *default_mass_cut* and *default_q_cut* is much larger than the tau mass.
Also the cross section for gamma gamma -> qq is suppressed relative to the corresponding cross section for lepton pair
production because the cut value of 10 GeV for *default_jet_cut* is larger than the cut value of 4 GeV for *default_mass_cut* .

Mixed stdhep files.

There are over 3500 files in ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/, each produced with 100% electron and positron polarization. To use these stdhep files in practice one must read the correct number of events from a subset of these files, or read a *mixed* stdhep file. A mixed stdhep file is built from the files
in ftp://ftp-lcd.slac.stanford.edu/ilc2/whizdata/ILC500/ and corresponds to a particular subset of final states with a particular initial state electron and positron polarization
combination (such as -80% electron and +30% positron).

Event Weight

Due to the presence of some high cross section processes, the events in a mixed stdhep file are not completely unweighted.
The event weight must therefore always be considered when analyzing events.
This weight is stored in the variable EVENTWEIGHTLH in the stdhep common block HEPEV4.

Process Identification

For each event in a mixed stdhep file the variable IDRUPLH from the common block HEPEV4 is used to identify the process. (See the description of the IDRUPLH variable in the introduction above.)

LOI Mixed stdhep files with randomized final states.

The following table lists Mixed Stdhep files that were produced for the LOI. The Raw Stdhep information corresponds to the original stdhep files, each with a single final state and 100% initial state polarization.

Mixed Stdhep Files	Raw Stdhep Files	Ecm (GeV)	mHiggs (GeV)	Processes	Event Weight Lumi (fb-1)	Pol. (% e / % e)	Mixed Nevents
directory inv_ab directory inv_ab	directory index logs	500	2000	All 0,2,4,6 fermion SM processes	250 250	-80/+30 +80/-30	4,737,499 2,453,865
directory inv_ab directory inv_ab	directory index logs	250	2000	All 0,2,4,6 fermion SM processes	125 125	-80/+30 +80/-30	4,972,958 2,904,045
directory		250	120	ffH run1	125 125	-80/+30 +80/-30	39893 25677
directory		250	120	ffH run2	125 125	-80/+30 +80/-30	39893 25677
directory		250	120	ffH run3	125 125	-80/+30 +80/-30	39893 25677
directory		250	120	ffH run4	125 125	-80/+30 +80/-30	39893 25677
directory		250	120	ffH, H--> e2E2	125 125	-80/+30 +80/-30	61191 39401
	directory index logs	1000	120	vvH, H-> all modes except e2E2		-100/+100 +100/-100	100,000 100,000
	directory index logs	1000	120	vvH, H-> e2E2 only		-100/+100 +100/-100	10,000 10,000
	directory index logs	350	2000	e2e2, ffe2e2			
	directory index logs	350	120	e2e2H			
	directory index logs	500	120	e2e2H			
directory inv_ab directory inv_ab	directory index logs	250	120	ffH, H--> cC	125 125	-80/+30 +80/-30	432,970 433,005
directory inv_ab directory inv_ab	directory index logs	250	119.7	e1E1H, e2E2H	250 250	-80/+30 +80/-30	149,994 150,000
directory inv_ab directory inv_ab	directory index logs	250	120.0	e1E1H, e2E2H	250 250	-80/+30 +80/-30	149,994 150,000
directory inv_ab directory inv_ab	directory index logs	250	120.0	ffH, H--> e2E2, postLOI,	250 250	-80/+30 +80/-30	499278 499564
directory inv_ab directory inv_ab	directory index logs	250	119.7	e1E1H, H--> cC, postLOI, unit weight	250 250	-80/+30 +80/-30	3155 2116
directory inv_ab directory inv_ab	directory index logs	250	119.7	e2E2H, H--> cC, postLOI, unit weight	250 250	-80/+30 +80/-30	2934 1978
directory inv_ab directory inv_ab	directory index logs	250	120.0	e1E1H, H--> cC, postLOI, unit weight	250 250	-80/+30 +80/-30	3133 2104
directory inv_ab directory inv_ab	directory index logs	250	120.0	e2E2H, H--> cC, postLOI, unit weight	250 250	-80/+30 +80/-30	2916 1964
directory inv_ab directory inv_ab	directory index logs	250	120.3	e1E1H, H--> cC, postLOI, unit weight	250 250	-80/+30 +80/-30	3116 2089
directory inv_ab directory inv_ab	directory index logs	250	120.3	e2E2H, H--> cC, postLOI, unit weight	250 250	-80/+30 +80/-30	2899 1953
directory inv_ab directory inv_ab	directory index logs	250	119.7	e1E1H, H--> cC, postLOI	250 250	-80/+30 +80/-30	499,523 499,988

directory inv_ab directory inv_ab	directory index logs	250	119.7	e2E2H, H--> cC, postLOI	250 250	-80/+30 +80/-30	501,011 499,608
directory inv_ab directory inv_ab	directory index logs	250	120.0	e1E1H, H--> cC, postLOI	250 250	-80/+30 +80/-30	499,465 499,465
directory inv_ab directory inv_ab	directory index logs	250	120.0	e2E2H, H--> cC, postLOI	250 250	-80/+30 +80/-30	500,285 500,284
directory inv_ab directory inv_ab	directory index logs	250	120.3	e1E1H, H--> cC, postLOI	250 250	-80/+30 +80/-30	499,562 499,807
directory inv_ab directory inv_ab	directory index logs	250	120.3	e2E2H, H--> cC, postLOI	250 250	-80/+30 +80/-30	498,836 498,069
directory inv_ab directory inv_ab	directory index logs	250	120.0	e1E1H, SM H decay, postLOI	250 250	-80/+30 +80/-30	499,462 499,461
directory inv_ab directory inv_ab	directory index logs	250	120.0	e2E2H, SM H decay, postLOI	250 250	-80/+30 +80/-30	500,285 500,284
directory	directory index logs	250	2000	e1E1, 60<Mee<115 GeV, etc.	125 125	-80/+30 +80/-30	3,296,144 3,019,696
directory	directory index logs	230	120	ffH	250 250	-80/+30 +80/-30	359,754 359,788
directory		500	2000	e2E2+missing, e3E3+missing	250 250	-80/+30 +80/-30	81,310,104 80,704,368
directory		500	2000	e3E3, DESY, correct tau pol.	250 250	-80/+30 +80/-30	1,148,162 1,010,385
directory		500	2000	e3E3	2000 2000	-80/+30 +80/-30	1,443,280 1,186,634
directory inv_ab directory inv_ab	directory index logs	500	2000	six fermions from tT, Mt=173.5 GeV	250 250	-80/+30 +80/-30	741,725 344,362
directory inv_ab directory inv_ab	directory index logs	500	2000	six fermions from tT, Mt=174.0 GeV	250 250	-80/+30 +80/-30	757,724 351,861
directory	directory index map	500	2000	Desy Susy Point5	250 250	-80/+30 +80/-30	106,053 106,293
directory inv_ab	directory index logs	500	2000	SLAC Susy Point5 delM = 0.0	500	-80/+30	1,264,422
directory inv_ab	directory index logs	500	2000	SLAC Susy Point5 delMch1 = 0.5	500	-80/+30	1,293,786
directory inv_ab	directory index logs	500	2000	SLAC Susy Point5 delMneu1 = 0.5	500	-80/+30	1,204,148
directory inv_ab	directory index logs	500	2000	SLAC Susy Point5 delMneu2 = 0.5	500	-80/+30	1,276,559
directory	directory index logs	500	120	ffHH, gHHH=1.00	1000 1000	-80/+30 +80/-30	116486 76540
directory	directory index logs	500	120	ffHH, gHHH=1.25	1000 1000	-80/+30 +80/-30	131396 87442

The lumi values in this table refer to the event weight normalization, and may not correspond in some instances to the ratio of the number of events to the cross section. Each Mixed Stdhep
File directory contains files inv_ab_stdhep_files_XXX_ecmEEE_80L_30R and inv_ab_stdhep_files_XXX_ecmEEE_80R_30L which describe the raw stdhep files used to build the mixed files (for example ftp://ftp-lcd.slac.stanford.edu/ilc/ILC250/Large_Stdhep_SM /inv_ab_stdhep_files_Large_Stdhep_SM_ecm250_80L_30R).
The fields in the ..._inv_ab_stdhep_files are:

filename (at SLAC)		Begin event		End event		Event weight		process_id		1st initial particle polarization if particle is e- or e+; beams/brems flag if photon		2nd initial particle polarization if particle is e- or e+; beams /brems flag if photon

The "index" link in the "Raw Stdhep Files" column takes you to the list of processes in the mixed stdhep file. The five columns in this list are: process_id, initial_state,
the variable IDRUPLH, an internal sequence number (1 - 4) and a bit indicating whether or not events were produced for this particular initial state polarization sign combination (0 = events generated, 1= events not generated). The internal sequence number denotes the initial state electron or positron helicity and/or the type of initial state photon (EPA or bremsstrahlung):

Internal Seq Number					e1	E1					e1	A					A	E1				
1					L	L					L	EPA					EPA	L				
2					L	R					L	beams					beams	L				
3					R	L					R	EPA					EPA	R				
4					R	R					R	beams					beams	R				

DBD Mixed stdhep files with randomized final states.

The following table lists Mixed Stdhep files that have been produced for the DBD.

Mixed Stdhep Files	Ecm(GeV)	mHiggs (GeV)	Processes	Event Weight Lumi (fb-1)	Pol. (% e / % e)	Mixed Nevents
directory inv_ab directory inv_ab	1000	2000	4f_WW	1000 1000	-80/+20 +80/-20	5,135,536 436,591
directory inv_ab directory inv_ab	1000	2000	6f_ttbar	1000 1000	-80/+20 +80/-20	566,454 566,494
directory inv_ab directory inv_ab	1000	2000	all_other_SM_background	1000 1000	-80/+20 +80/-20	3,232,672 2,814,719
directory inv_ab directory inv_ab	1000	2000	eeZ_vvZ_leptonic	1000 1000	-80/+20 +80/-20	10,965,564 10,909,758
directory inv_ab directory inv_ab	1000	2000	evW_eeZ_vvZ_semileptonic	1000 1000	-80/+20 +80/-20	6,570,292 5,080,159
directory inv_ab directory inv_ab	1000	125	higgs_ffh_mumu	1000 1000	-80/+20 +80/-20	316,219 43,429
directory inv_ab directory inv_ab	1000	125	higgs_ffh_nomu	1000 1000	-80/+20 +80/-20	1,544,378 1,544,398
directory inv_ab directory inv_ab	1000	2000	ttbb-6q-all	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	2000	ttbb-ln4q-all	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	125	tth-2l2nbb-hbb	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	125	tth-2l2nbb-hnonbb	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	125	tth-6q-hbb	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	125	tth-6q-hnonbb	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	125	tth-ln4q-hbb	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	125	tth-ln4q-hnonbb	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	2000	ttz-6q-all	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	2000	ttz-ln4q-all	1000 1000	-80/+20 +80/-20	51,000 51,002
directory inv_ab directory inv_ab	1000	2000	aa_lowpt	0.00292 0.00292	-80/+20 +80/-20	2,394,460 2,331,717
directory inv_ab directory inv_ab	500	2000	6f_ttbar_mt173p5	250 250	-80/+30 +80/-30	925,595 411,543
directory inv_ab directory inv_ab	500	2000	6f_ttbar_mt174p0	250 250	-80/+30 +80/-30	938,761 418,120
directory inv_ab directory inv_ab	500	2000	all_SM_background	250 250	-80/+30 +80/-30	2,269,212 1,586,791
directory inv_ab directory inv_ab	500	2000	bhabha_inclusive	0.00036 0.00036	-80/+30 +80/-30	100,334 100,284
directory inv_ab directory inv_ab	500	2000	ea_ea	0.000144 0.000144	-80/+30 +80/-30	1,151 1,151
directory inv_ab directory inv_ab	500	2000	aa_lowpt	0.00347 0.00347	-80/+30 +80/-30	2,199,480 2,199,480
directory inv_ab directory inv_ab directory inv_ab directory inv_ab	250	2000	all_SM_background	250 250 25 25	-80/+30 +80/-30 -80/-30 +80/+30	2,822,661 2,058,374 2,294,529 1,657,611

directory inv_ab directory inv_ab	250	2000	evW_eeZ_vvZ_semileptonic	250 250	-80/+30 +80/-30	2,030,078 1,485,507
directory inv_ab directory inv_ab	250	2000	higgs_ffh_zz	250 250	-80/+30 +80/-30	120,000 120,012

FAQ

Whizard 1.40 has no gluon emission by default, leading to potentially incorrect multiplicity distributions.

The WHIZARD version 1.40 that was used to generate this sample indeed did not include gluon emission. However gluon radiation was simulated using PYTHIA's parton showering algorithm. WHIZARD versions 1.50 and higher include gluon emission, and, starting with version 1.91, WHIZARD has its own parton showering code.

Whizard 1.40 has an incorrect implementation of the CKM matrix. Only diagonal terms of the matrix are present (and = 1!), giving wrong W decays.

Although true for the Whizard version 1.40 that was used to generate this data sample, it is extremely doubtful that this will have any effect on the current analyses. WHIZARD versions 1.51 and higher include the correct CKM matrix, and so future data samples will include the rarer W decays.

This sample has generator level cuts a la SiD, providing a potential bias when used for ILD.

There are, indeed, some kinematical cuts for processes with divergent cross-sections, which can be seen by looking at the whizard.in file as described above. However, the only kinematic cut that leads to a genuine loss of events is a 4 GeV minimum invariant mass cut on final state fermion-antifermion pairs.