

Philip Bechtle, Ivor Fleck, Andreas Meyer, Marcus Rammes, Thomas Schörner-Sadenius

Introduction to Terascale Physics 24 February 2011

THE Z' BOSON

And the relevant final states

Z': massive, electrically neutral resonance. Potentially in extended gauge groups like

- Can mix with the SM Z.
- Supposed to be heavier than 500 GeV.
- Preferred decay mode: to top quark pair with subsequent decays to Wb pairs:

Top final states and W boson decays:

- W bosons can decay leptonically or hadronically
- fully hadronic, semi-leptonic and fully leptonic final states.
- Here concentrate on semi-leptonic final state (moderate background, good branching ratio)





EXISTING LIMITS ON THE Z' BOSON From LEP, HERA, Tevatron, LHC



THE PROJECT: FINDING THE Z'



- At least 4 jets with $|\eta| < 2.5$ and $p_T > 40 \,\text{GeV}$
- Exactly one isolated lepton with $p_T > 20$, GeV for the electrons and $p_T > 25 \text{ GeV}$ for the muons; excluded crack region $1.37 < |\eta| < 1.52^{-1}$. The isolation cuts are $E_T^{\text{cone20}} < C1 + C2 \cdot E_T$ (C1 = 4.0 GeV, C2 = 0.023) for the electrons and $E_T^{\text{cone20}}/p_T < 0.1$ for the muons

Z' Tutorial

- Exactly two *b*-tagged jets (tag weight > 6.0)

ANALYSIS FRAMEWORK AND FLOW

In only a few cc files ... data samples

#include (...)

```
// for debugging purposes: set to -1 if all events should be read int maxReadEventsPerFile = -1;
```

```
// Sample file names
const char* zPrime_file="data/zprime_sample.root";
```

```
// Raw data NTuple declaration
Data *zprimeData,*ttbarData,*zeeData,*zmumuData,*wenuData,*wmunuDat
```

```
// Reco data declaration
RecoData *zprimeRecoData ...
```

```
// Now include all exercise files #include "exercise1.cc"
```

• • •

```
// main() routine
int main(int argnum,char** args)
```

```
// Read sample NTuples
zprimeData=readNTuple(zPrime_file);
```

```
'/' Perform Z' reconstruction on samples
zprimeRecoData=getReconstructedData(zprimeData);
```

```
// Open a text file to log important results
report.open("analysis_report.txt",std::ios_base::out);
// And now it's your turn...
exercisel();
```

```
• • •
```

ANALYSIS FRAMEWORK AND FLOW

In only a few cc files ... data samples

```
*****
#include (...)
                                   RecoData* getReconstructedData(Data* data):
// for debugging purposes: se
int maxReadEventsPerFile = -
                                   Reconstruct the top quarks, calculate Z' mass and return pointer to reconstructed
                                   data
                                                    ******
                                  /*****
// Sample file names
const char* zPrime_file="data_RecoData* getReconstructedData(Data* data)
// Raw data NTuple declaration
Data *zprimeData, *ttbarData, *zeeData, *zmumuData, *wenuData, *wmunuDat
// Reco data declaration
RecoData *zprimeRecoData ...
// Now include all exercise files #include "exercise1.cc"
. . .
// main() routine
int main(int argnum char** args)
  // Read sample NTuples
zprimeData=readNTuple(zPrime_file);
  '/' Perform Z' reconstruction on samples
zprimeRecoData=getReconstructedData(zprimeData);
  // Open a text file to log important results report.open("analysis_report.txt", std::ios_base::out);
  // And now it's your turn...
  exercise1():
  ...
```

ANALYSIS FRAMEWORK AND FLOW

In only a few cc files ... data samples

```
******
#include (...)
                                    RecoData* getReconstructedData(Data* data):
// for debugging purposes: se
int maxReadEventsPerFile = -1
                                    Reconstruct the top quarks, calculate Z' mass and return pointer to reconstructed
                                    data
                                   // Sample file names
const char* zPrime_file="data_RecoData* getReconstructedData(Data* data)
                                                         // How many events to read?
// Raw data NTuple declaration
                                                         unsigned int maxEvents=data->size();
Data *zprimeData *ttbarData *zeeData *zmumuData
                                                         for (unsigned int i=0;i⊲maxEvents;i++)
// Reco data declaration
RecoData *zprimeRecoData ...
                                                         {
                                                                // Event has passed all cuts?
                                                                if (isGoodEvent(data->at(i)))
// Now include all exercise files #include "exercise1.cc"
                                                                       Event event=data->at(i);
. . .
                                                                       RecoObject tmpObj;
// main() routine
int main(int argnum char** args)
                                                                        tmpObj.event=event;
  // Read sample NTuples
zprimeData=readNTuple(zPrime_file);
                                                                       HadronicWBoson hadW=getBestHadronicWFit(event);
                                                                       LeptonicWBoson lepW=getBestLeptonicWFit(event);
                                                                       HadronicTop hadTop=getBestHadronicTopFit(event.hadW);
  '/' Perform Z' reconstruction on samples
zprimeRecoData=getReconstructedData(zprimeData
                                                                       LeptonicTop lepTop=getBestLeptonicTopFit(event,lepW,hadTop);
                                                                        tmpObj.hadTop=hadTop;
  // Open a text file to log important results report.open("analysis_report.txt", std::ios_bas
                                                                        tmpObj.lepTop=lepTop;
                                                                       tmpObj.ZprimeMass=getZprimeMass(hadTop,lepTop);
  // And now it's your turn...
  exercise1():
                                                                       ret_>push_back(tmpObj);
  ...
                                                                       goodEvents++;
```

RUNNING THE ANALYSIS

... for the various exercises ...

000 1	Ferminal — ssh	n — 63×34			
[desy-cms884] ~/tuto	rial \$ ls		Ē		
Makefile	exercise4.cc	readNTuple.cc	4		
analysis_report.txt	exercise5.cc	reconstruction.cc	C		
cuts.cc	exercise6.cc	save			
data	exercise7.cc	solutions			
doc	exercise8.cc	stripped			
exercise1.cc	exercise9.cc	test.cc			
exercise10.cc	kinematics.cc	typedefs.cc			
exercise2.cc	main.cc				
exercise3.cc	plots				
[desy-cms884] ~/tuto	rial \$ make				
g++ -g -O2 -Wall -pthread -m32 -I/opt/products/root/5.26.00/inc					
lude main.cc -I/opt/	products/root/5	.26.00/include _L/opt/produc			
ts/root/5.26.00/lib -L./ -L/opt/products/root/5.26.00/lib -lCor					
e -lCint -lRIO -lNet -lHist -lGraf -lGraf3d -lGpad -lTree -lRin					
t -lPostscript -lMatrix -lPhysics -lMathCore -lThread -pthread					
-lm -ldl -rdynamic -o main -lMathCore					
[desy-cms884] ~/tuto	rial \$ ls				
Makefile	exercise4.cc	plots			
analysis_report.txt	exercise5.cc	readNTuple.cc			
cuts.cc	exercise6.cc	reconstruction.cc			
data	exercise7.cc	save			
doc	exercise8.cc	solutions			
exercise1.cc	exercise9.cc	stripped			
exercise10.cc	kinematics.cc	test.cc			
exercise2.cc	main	typedefs.cc			
exercise3.cc	main.cc				
[desy-cms884] ~/tuto	rial \$./main		П		
+++ NTuple file: dat	a/zprime_sample	.root:			
=> 19020 entries fo	und				
+++ NTuple file: data/ttbar_sample.root:					
=> 58475 entries fo	und		L		
			¥		
[desy-cms884] ~/tuto	rial \$		2		

THE DATA STRUCTURES

used for the analysis, defined in typedefs.cc

struct RecoObject

- Event event: the raw event associated to the reconstruction
- HadronicTop hadTop: the reconstructed hadronic top quark
- LeptonicTop lepTop: the reconstructed leptonic top quark
- double ZprimeMass: the reconstructed invariant Z' mass

THE DATA STRUCTURES

used for the analysis, defined in typedefs.cc

struct HadronicTop

- Jet bJet: b-tagged jet closest in ΔR to the previously reconstructed hadronic W boson
- HadronicWBoson wBoson: the hadronic W boson
- double mass: reconstructed invariant hadronic top mass
- double deltaR: ΔR between the b jet and the hadronic W

struct LeptonicTop

- Jet bJet: b-tagged jet closest in ΔR to the previously reconstructed leptonic W boson
- LeptonicWBoson wBoson: the leptonic W boson
- double mass: reconstructed invariant leptonic top mass (best agreement of the two leptonic W solutions with hadronic top mass)
- double deltaR: ΔR between the b jet and the leptonic W

THE DATA STRUCTURES used for the analysis, defined in typedefs.cc

struct LeptonicWBoson

- double El,pxl,pyl,pzl: energy and 3-momentum components of the charged lepton constituent of the W boson
- double ptnu: transverse momentum of the recovered neutrino
- double pxnu,pynu: x and y 3-momentum components of the recovered neutrino
- std::pair<double,double> pznu: the two soluintions of the p^ν_z recovery (see 4.4)
- std::pair<double,double> Enu: the two energies calculated from the p_z^{ν} solutions of the recovered neutrino
- std::pair<double,double> mass: the two invariant W masses calculated from the p^v_z solutions
- double Enu_final: E_ν for the best W fit
- double pznu_final: p^v_z for the best W fit
- double mass_final: invariant W mass for the best W fit

struct HadronicWBoson

- Jet jet1, jet2: light jet constituents of the hadronic W boson
- double mass: reconstructed invariant W mass

THE DATA STRUCTURES

used for the analysis, defined in typedefs.cc

struct Electron

- E : electron energy
- Et : electron transverse energy
- p : norm of electron 3-momentum
- pt: electron p_T
- px,py,pz : vector components of electron 3-momentum
- double eta, phi: electron η , φ
- charge : charge of the electron
- etCone20 : energy deposition in a ΔR = 0.2 cone around the electron track
- p : norm of jet 3-momentum
- pt: jet p_T
- px,py,pz : vector components of jet 3-momentum
- double eta, phi: jet η , φ
- tag-weight : statistical weight of b tagging
- bjet_prob : b tag weight converted to a probability

struct Jet

- E : jet energy
- Et : jet transverse energy

THE DATA STRUCTURES

used for the analysis, defined in typedefs.cc

struct Event

- std::vector<Jet>: all jets of the event
- std::vector<Electron>: all electrons of the event
- std::vector<Muon>: all muons of the event
- double missingET: missing transverse energy $\not\!\!\!E_T$
- double missingET_phi: azimuthal direction of $\not\!\!E_T$ (angle in x, y-plane)
- weight: the event weight (only important when scaling samples up to a mixed data sample, is 1.0 in the prefabricated samples)

struct Muon

- double E : muon energy
- double Et : muon transverse energy
- double p : norm of muon 3-momentum
- double pt: muon p_T
- double px,py,pz : vector components of muon 3-momentum
- double eta,phi: muon η, φ
- double tag_weight : statistical weight of b tagging
- double charge : charge of the muon
- double etCone20 : energy deposition in a $\Delta R=0.2$ cone around the muon track



TH1F zprimeJetMultiHisto2("zprimeJetMultiHisto2","",16,-0.5,15.5);

for (unsigned int i=0;i<zprimeRecoData->size();i++)
zprimeJetMultiHisto2.Fill(zprimeRecoData->at(i).event.jets.size());

```
if (event.jets[i].bjet_prob>prob)
```

DATA SAMPLES

for ttbar and Z' signals, and for backgrounds

We have generated simple ROOT NTuples for signal and the most important background processes containing only the most relevant information you need for this exercise. They have been produced from official ATLAS MonteCarlo production data sets.

file name	contained data	$\sigma [{ m pb}]$	$\int \mathcal{L} dt [\mathrm{pb}^{-1}]$
zee_jets.root	$Z ightarrow e^+e^- + n{ m jets}$	28.81	1000
zmumu_jets.root	$Z ightarrow \mu^+ \mu^- + n { m jets}$	29.41	1000
$wenu_jets.root$	$W ightarrow e u + n { m jets}$	160.57	200
$wmunu_jets.root$	$W ightarrow \mu u + n { m jets}$	90.77	200
wbb_jets.root	$Wbar{b}+n{ m jets}$	14.64	1000
ttbar_sample.root	$tar{t},{ m semi-lep}$	220.66	250
${\tt zprime_sample.root}$	$Z' \rightarrow t \bar{t}, M_{Z'} = 1000 { m GeV}$	0.634	30000
data_sample.root	mixed $Z'/t\bar{t}$ sample	—	300

The luminosities of the samples were chosen in a way to end up with similar number of events in each sample.

EXERCISE 1: SELECTION

- At least 4 jets with $|\eta| < 2.5$ and $p_T > 40 \,\mathrm{GeV}$
- Exactly one isolated lepton with $p_T > 20$, GeV for the electrons and $p_T > 25$ GeV for the muons; excluded crack region $1.37 < |\eta| < 1.52^{-1}$. The isolation cuts are $E_T^{\text{cone}20} < C1 + C2 \cdot E_T$ (C1 = 4.0 GeV, C2 = 0.023) for the electrons and $E_T^{\text{cone}20}/p_T < 0.1$ for the muons
- Exactly two *b*-tagged jets (tag weight > 6.0)

Exercise 1

Make two plots for the jet multiplicities of the $t\bar{t}$ and the Z' sample out of two overlaid histograms: One histogram filled with the jet multiplicity before and one after applying the cuts. Normalize the histograms to their integrals.

Calculate the event selection efficiencies $\epsilon_{t\bar{t}}^{C1}$ and $\epsilon_{Z'}^{C1}$ for the $t\bar{t}$ and the Z' sample. From the efficiencies of all the background samples (except $t\bar{t}$) calculate an average background suppression factor $\langle b \rangle = 1 - \langle \epsilon \rangle$.

EXERCISE 1: RESULT



EXERCISE 2: RECONSTRUCTED HAD. W

The hadronic W boson in a semi-leptonic top quark event is the W boson that decays into two light flavor (LF) jets. LF jets are all jets that have not been b-tagged (tag weight < 6.0). Among all LF jets, the two ones are used for the kinematic W reconstruction that build up the W mass fitting the best $(M_W = 80.4 \,\text{GeV})$.

Exercise 2

Combine the hadronic W mass spectra of the $t\bar{t}$ and the Z' samples in one plot. Normalize both histograms to their integrals.

Z' Tutorial





EXERCISE 3: COMBINATORIAL BG

The most important background in this process is the so-called "combinatorical background": If you have n LF jets in your event, there are n(n-1) possibilities to combine them two the W boson. For n = 5 LF jets, there are $5 \cdot 4 = 20$ combinations, only one of them is true.

The combinatorical background of the W reconstruction can be described by a 4th degree Tschebyschow polynomial, the correctly combined jets can be assumed to be Gaussian distributed around the W mass. The first four orders of

Exercise 3

$$T_0 = 1$$
 $T_1 = x$ $T_2 = 2x^2 - 1$
 $T_3 = 4x^3 - 3x$ $T_4 = 8x^4 - 8x^2 + 1$

Fit the sum of a 4th degree Tschebyschow and a Gaussian distribution to your hadronic W mass spectrum of the $t\bar{t}$ sample and extract the W mass and width from the mean and the RMS of the Gaussian.

<u>Hint</u>: First define five TF1's representing your Tschebyschoff polynomials $T_0 \ldots T_4$ and one TF1 for the Gauss. In a sixth TF1, add the five polynomials, each one provided with a free fit parameter for scaling, and the Gaussian TF1. Use the standard TH1::Fit(...) method. Maybe you have to set start parameters for the Gauss to bring the fit to converge.

EXERCISE 3: RESULT



EXERCISE 4: THE HADRONIC TOP

After the hadronic W is reconstructed, the b jet that is closest in ΔR to the hadronic W boson is chosen to reconstruct the hadronic top quark.

Exercise 4

Plot the hadronic top mass spectrum for the $t\bar{t}$ and the Z' sample (normalized to their integrals).

Perform the previous exercise with the Tschebyschoff fit to obtain the top mass and width of the $t\bar{t}$ sample.

EXERCISE 4: RESULT



EXERCISE 5: THE "LEPTONIC" W

Taking the W boson mass as a fixed value, one can obtain a quadratic equation in p_z^{ν} :

(1) has either two solutions or none. If it has none, p_T^{ν} is iteratively decreased until the equation has two solutions. Later on, the leptonic top mass will be reconstructed with the leptonic W from both solutions. The leptonic top mass that agrees best with the hadronic top mass will be kept.

Exercise 5

Combine the histograms of the leptonic W mass for both p_z^ν solutions in one plot. Make one plot for each the $t\bar{t}$ and the Z' sample.

EXERCISE 5: RESULT



EXERCISE 6: Z' RECONSTRUCTION

Finally, the Z' invariant mass can be calculated from the reconstructed tops. Since the cross section for SM $t\bar{t}$ production is about a factor 370 higher than $\sigma(Z' \to t\bar{t})$, one has to think about a way to distinguish the production channels.

 $t\bar{t}$ production from a Z' is a resonant decay, i. e. the top quarks will be more boosted in average. As consequence, the opening angle of the cone of the decay products from the top quarks will be more narrow. This should be seen as a clear signature in the ΔR distributions of the b jet and the corresponding W boson.

Exercise 6

Fill two histograms with the invariant Z' mass spectrum from the Z' and $t\bar{t}$ samples and combine them in one plot. Choose logarithmic y scaling.







Exercise 7

For each the hadronic and the leptonic top quark, fill two histograms with the ΔR spectrum of the Z' and $t\bar{t}$ samples. Make two plots out of them (e.g. "hadronic_top_deltaR.eps" and "leptonic_top_deltaR.eps"). Normalize the two histograms in each plot to their integrals.

Choose a cut in ΔR in a way to maximize purity on the on hand, but not cutting away too many events on the other hand. Apply an additional, appropriate cut on the invariant Z' mass; choose a quite naarow window around 1000 GeV

With these additional cuts, again fill two histograms with the invariant Z' mass, this time with linear y scaling. Normalize the histograms to 3 fb^{-1} now. Calculate the efficiencies $\epsilon_{t\bar{t}}^{C2}$ und $\epsilon_{Z'}^{C2}$ for the additional cuts.

EXERCISE 7: RESULT



EXERCISE 8: DISCOVERY POTENTIAL

Let's assume there was no Z' at all (the *null hypothesis*). Then, the number of expected $t\bar{t}$ is $N_{\exp} = \epsilon_{t\bar{t}} \cdot N_{t\bar{t}}$. N_{\exp} will fluctuate according to a Poisson distribution. For large numbers the uncertainty is given by:

$$\Delta N_{
m exp} = \sqrt{N_{
m exp}}$$

We can ask now the question how many events we may count additionally to an expected number N_{exp} of $t\bar{t}$ events after Z' reconstruction and still trust the null hypothesis. Obviously, the number of $t\bar{t}$ we really observe is given by

$$N_{\rm obs} = N_{\rm exp} + \eta \cdot \Delta N_{\rm exp} = N_{\rm exp} + \eta \cdot \sqrt{N_{\rm exp}} \Rightarrow \eta = \frac{N_{\rm obs} - N_{\rm exp}}{\sqrt{N_{\rm exp}}}$$

 η is called the *discovery potential*, it tells you the number of standard deviations the observed number of events is lying above the expected number of events.

- The number of expected (and observed) events depends on the integrated luminosity. The relative deviation becomes smaller with the number of events
- Produce a graph of eta as a function of integrated luminosity
 - For each value of luminosity calculate $\,N_{
 m obs},\,N_{
 m exp},\,\eta\,$ taking the efficiencies into account:

$$N = \sigma \cdot \mathcal{L} \cdot \epsilon$$

EXERCISE 8: DISCOVERY POTENTIAL

Exercise 8

Extract N_{obs} and N_{exp} from your analysis and calculate the discovery potential for a 1000 GeV Z' at an integrated luminosity of 3 fb⁻¹.

Plot the function "discovery portential vs. integrated luminosity". Estimate your value for $N_{\rm exp}$ and $N_{\rm obs}$ from $\sigma_{t\bar{t}}$ and the reconstruction efficiency. Use a TF1 with draw option "AL".

<u>Hint</u>: The overall efficiencies for sequential cuts always factorize, i. e. $\epsilon = \epsilon^{C1} \cdot \epsilon^{C2}$.





EXERCISE 9: LIMIT DERIVATION

 From the measured number of events, knowing the contribution from backgrounds, efficiencies and luminosity, an upper limit on the possible Z' cross section can be given

$$\sigma = \frac{N_{\rm obs}}{\epsilon \cdot L}$$

- N_{obs} is composed of $N_{obs} = N_{obs,t\bar{t}} + N_{obs,Z'}$ and has a statistical uncertainty
- Assuming again that N_{obs} comes entirely from tt, i.e. N_{obs} = N_{exp,tt}, the upper limit N_{max,Z} can be determined and expressed in units of the standard deviation of N_{obs}

$$N_{\max,Z'} = N_{\text{obs}} - N_{\exp,t\bar{t}}$$

- The upper limit of the cross section at % C.L. is calculated from N_{max,Z} for the corresponding point of the (Gaussian) distribution $N_{obs}(\mu, CL)$ where $\mu = N_{exp,t\bar{t}}$
- Make a graph of N_{max,Z} as a function of C.L. using the results for N_{obs} obtained in exercise 7.

Exercise 9

Plot a function "CL vs. upper bound cross section" in the range of CL = [60...99%]. Use a TGraph and add x, y-values in a for loop. Use the draw option "AL".





EXERCISE 10: CROSS SECTION DETERM.

Exercise 10

Make a histogram of the invariant Z' mass of the $t\bar{t}$ sample. Fit a Landau distribution to the histogram.

Make a histogram of the invariant Z' mass spectrum of data_sample.root. Scale the Landau fit by the ratio of luminosities.

With the scaled Landau fit, go through the data histogram bin by bin and subtract the function value at the bin center from the bin content. Fill a new histogram with that difference bin by bin. Take care that you set equal bin numbers for all histograms!

Calculate the cross section of the "unknown" resonance and its uncertainty.

EXERCISE 10: CROSS SECTION DETERM.

- Determine and fit background shape from simulation
- Normalize and overlay background shape on data



Subtract background from data

EXERCISE 10: RESULT

