A data driven background estimate for leptonic analysis

Supervisor: Dr. Patrizia Azzi

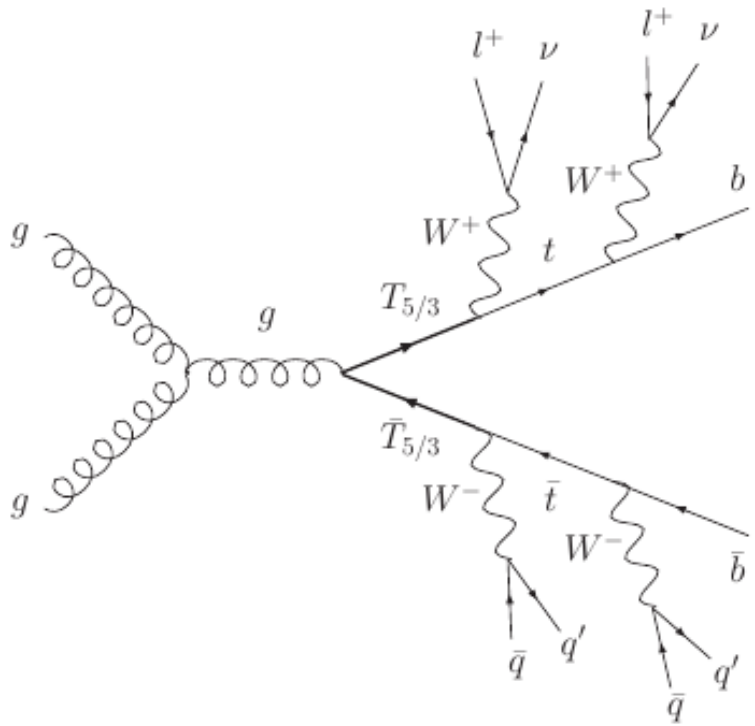
Davide Racco

University of Padova, Italy

16th August 2012

Search for the Top Partners

- Heavy fermions coupled to third generation quarks, predicted in many models BSM.
- Exotic charges: $T_{5/3}$, $B_{1/3}$.
- Decay in tW, with final state $t\bar{t}W^+W^-$.
- We focus on $T_{5/3}$, $\bar{T}_{5/3}$, in the channel with:
  - 2 same-sign, prompt[a] leptons;
  - Jets;
  - MET.

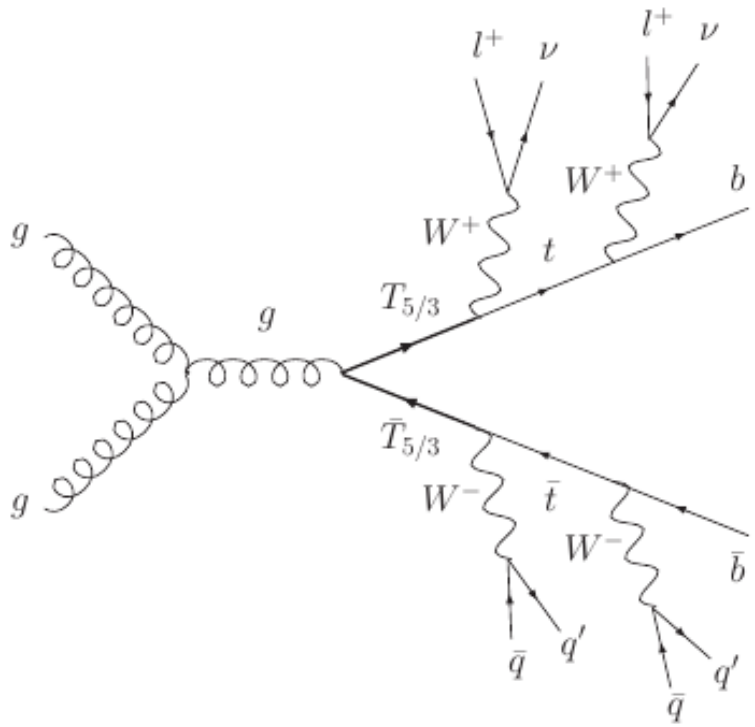


[a] *Prompt* leptons: typical leptons coming from W, Z, with high isolation.

Signature: $T_{5/3}\bar{T}_{5/3} \longrightarrow t\bar{t}W^+W^- \longrightarrow \ell^\pm\ell^\pm + \text{jets} + \not{E}_T$

Search for the Top Partners

- Heavy fermions coupled to third generation quarks, predicted in many models BSM.
- Exotic charges: $T_{5/3}$, $B_{1/3}$.
- Decay in tW, with final state $t\bar{t}W^+W^-$.
- We focus on $T_{5/3}$, $\bar{T}_{5/3}$, in the channel with:
  - 2 same-sign, prompt[a] leptons;
  - Jets;
  - MET.



[a] *Prompt* leptons: typical leptons coming from W, Z, with high isolation.

Signature: $T_{5/3}\bar{T}_{5/3} \longrightarrow t\bar{t}W^+W^- \longrightarrow \ell^\pm\ell^\pm + \text{jets} + \not{E}_T$

Backgrounds in Physics analysis

- In pp collisions, there is also the production of lots of events with SM processes that can pass the chosen selection: they are called *background*.
- In our analysis there are three types of backgrounds:
  1. SM processes with similar final state (2 same-sign prompt leptons): $t\bar{t}W$, $t\bar{t}WW$, WWZ

     $\rightarrow$ rare processes, low cross section
  2. Opposite sign leptons, with charge misreconstruction

     $\rightarrow$ very low probability
  3. Same sign prompt/non-prompt leptons: main background

     $\rightarrow$ this is our focus

Backgrounds in Physics analysis

- In pp collisions, there is also the production of lots of events with SM processes that can pass the chosen selection: they are called *background*.
- In our analysis there are three types of backgrounds:
  1. SM processes with similar final state (2 same-sign prompt leptons): $t\bar{t}W$, $t\bar{t}WW$, WWZ
     $\rightarrow$ rare processes, low cross section
  2. Opposite sign leptons, with charge misreconstruction
     $\rightarrow$ very low probability
  3. Same sign prompt/non-prompt leptons: main background
     $\rightarrow$ this is our focus

Non-prompt leptons in our analysis

- The non-prompt leptons are non-isolated leptons that can, nevertheless, pass the tight quality cuts.
  In our analysis they mostly come from the quark b semi-leptonic decay of $t\bar{t}$ events:

$$t\bar{t} \to WbW\bar{b} \to \underbrace{\ell\nu}_{W}\, b \quad \underbrace{jj}_{W}\, b \begin{cases} \text{prompt lepton} \\ \text{non-prompt lepton (from b)} \end{cases}$$

- There are various methods developed to estimate the probability that such a lepton passes the selection.
  The MC simulations do not work in this case
  $\Longrightarrow$ *data driven* methods.

Non-prompt leptons in our analysis

- The non-prompt leptons are non-isolated leptons that can, nevertheless, pass the tight quality cuts.
  In our analysis they mostly come from the quark b semi-leptonic decay of $t\bar{t}$ events:

$$t\bar{t} \to WbW\bar{b} \to \underbrace{\ell\nu}_{W}\, b \quad \underbrace{jj}_{W}\, b \begin{cases} \text{prompt lepton} \\ \text{non-prompt lepton (from b)} \end{cases}$$

- There are various methods developed to estimate the probability that such a lepton passes the selection.
  The MC simulations do not work in this case
  $\Longrightarrow$ *data driven* methods.

Definition of tight/loose rate method

- We consider the ratio

$$\frac{\text{number of tight leptons}}{\text{number of loose leptons}} .$$

  It represents the probability that a lepton passing the loose selections passes also the tight ones.
- We compute this ratio over a dataset depleted in leptons from Z or W, and rich in non-prompt leptons from $b \to \ell X$.

Example: Z+jets events

Since processes as ZW, ZZ have very low cross sections, we expect that most of the other leptons of the event (apart from the two identified with the Z) are non-prompt.

Other case: W removal

Similar method, requiring one lepton and cuts on $\not{E}_T$, M_T, jets, to select events from QCD where the contribution from Ws and Zs has been removed.

Definition of tight/loose rate method

- We consider the ratio

$$\frac{\text{number of tight leptons}}{\text{number of loose leptons}} .$$

  It represents the probability that a lepton passing the loose selections passes also the tight ones.
- We compute this ratio over a dataset depleted in leptons from Z or W, and rich in non-prompt leptons from $b \to \ell X$.

Example: Z+jets events

Since processes as ZW, ZZ have very low cross sections, we expect that most of the other leptons of the event (apart from the two identified with the Z) are non-prompt.

Other case: W removal

Similar method, requiring one lepton and cuts on $\not{E}_T$, M_T, jets, to select events from QCD where the contribution from Ws and Zs has been removed.

Definition of tight/loose rate method

- We consider the ratio

$$\frac{\text{number of tight leptons}}{\text{number of loose leptons}} .$$

  It represents the probability that a lepton passing the loose selections passes also the tight ones.
- We compute this ratio over a dataset depleted in leptons from Z or W, and rich in non-prompt leptons from $b \to \ell X$.

Example: Z+jets events

Since processes as ZW, ZZ have very low cross sections, we expect that most of the other leptons of the event (apart from the two identified with the Z) are non-prompt.

Other case: W removal

Similar method, requiring one lepton and cuts on $\not{E}_T$, M_T, jets, to select events from QCD where the contribution from Ws and Zs has been removed.

Tight/Loose criteria

ELECTRONS

- $p_T > 30$ Gev
- $|\eta| < 2.4$
- Not in barrel-endcap region
- Charge consistent
- Cut category: "HyperTightMC1" CiC selection
- **Relative Isolation** $< 0.6 \rightarrow < 0.15$
- Conversion rejection cuts
- **Transverse impact parameter** < 0.02 cm

MUONS

- $p_T > 30$ Gev
- $|\eta| < 2.4$
- Classified as GlobalMuon and TrackerMuon
- Normalized $\chi^2 < 50 \rightarrow < 10$
- **Impact parameter** < 2 cm $\rightarrow < 0.02$ cm
- **Relative Isolation** $< 0.4 \rightarrow 0.2$
- Number of muon hits > 0
- Number of valid silicon hits ≥ 11
- At least one matching segment

Tight/Loose criteria

ELECTRONS

- $p_T > 30$ Gev
- $|\eta| < 2.4$
- Not in barrel-endcap region
- Charge consistent
- Cut category: "HyperTightMC1" CiC selection
- **Relative Isolation** $< 0.6 \rightarrow < 0.15$
- Conversion rejection cuts
- **Transverse impact parameter** < 0.02 cm

MUONS

- $p_T > 30$ Gev
- $|\eta| < 2.4$
- Classified as GlobalMuon and TrackerMuon
- Normalized $\chi^2 < 50 \rightarrow < 10$
- **Impact parameter** < 2 cm $\rightarrow < 0.02$ cm
- **Relative Isolation** $< 0.4 \rightarrow 0.2$
- Number of muon hits > 0
- Number of valid silicon hits ≥ 11
- At least one matching segment

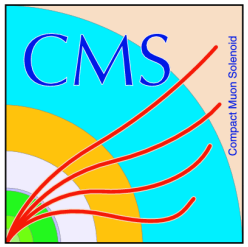
Required tasks

The datasets used in the analysis
are data taken by CMS in 2012 (~ 5 fb^{-1}, 8 TeV).
Analysis completed (2011 data); it was necessary to:

- update code for ROOT Tree production;
- use CRAB to run the programs on the GRID;
- adapt/develop macros
  for new analysis (and begin to learn Python...)

The datasets used have the following triggers:

- SingleMu_2012 (sets A, B1).
- SingleElectron_2012 (sets A, B1, B2).



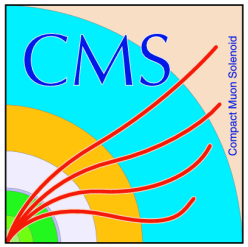
Required tasks

The datasets used in the analysis
are data taken by CMS in 2012 (~ 5 fb^{-1}, 8 TeV).
Analysis completed (2011 data); it was necessary to:

- update code for ROOT Tree production;
- use CRAB to run the programs on the GRID;
- adapt/develop macros
  for new analysis (and begin to learn Python...)

The datasets used have the following triggers:

- SingleMu_2012 (sets A, B1).
- SingleElectron_2012 (sets A, B1, B2).



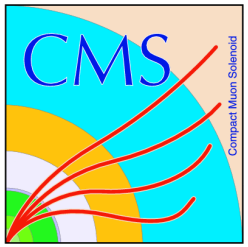
Required tasks

The datasets used in the analysis
are data taken by CMS in 2012 (~ 5 fb^{-1}, 8 TeV).
Analysis completed (2011 data); it was necessary to:

- update code for ROOT Tree production;
- use CRAB to run the programs on the GRID;
- adapt/develop macros
  for new analysis (and begin to learn Python...)

The datasets used have the following triggers:

- SingleMu_2012 (sets A, B1).
- SingleElectron_2012 (sets A, B1, B2).



Z + jets sample definition

- Skim of the data imposing that there is at least one pair of opposite sign, same flavour leptons with $76 < M(\ell\ell) < 106$ GeV.
- Then we want to select the events containing Z+jets :
  1. MET < 25 GeV $\rightsquigarrow$ remove WZ events
  2. At least 1 jet with $p_T > 30$ $\rightsquigarrow$ hadronic activity
  3. At least 1 "extra" lepton $\rightsquigarrow$ "probe" leptons
- We compute the $\frac{\text{tight}}{\text{loose}}$ ratio over the "extra" leptons.

Z + jets sample definition

- Skim of the data imposing that there is at least one pair of opposite sign, same flavour leptons with $76 < M(\ell\ell) < 106$ GeV.
- Then we want to select the events containing Z+jets:

1. MET < 25 GeV $\rightsquigarrow$ remove WZ events
2. At least 1 jet with $p_T > 30$ $\rightsquigarrow$ hadronic activity
3. At least 1 "extra" lepton $\rightsquigarrow$ "probe" leptons

- We compute the $\frac{\text{tight}}{\text{loose}}$ ratio over the "extra" leptons.

Z + jets sample definition

- Skim of the data imposing that there is at least one pair of opposite sign, same flavour leptons with $76 < M(\ell\ell) < 106$ GeV.
- Then we want to select the events containing Z+jets:

1. MET < 25 GeV $\rightsquigarrow$ remove WZ events
2. At least 1 jet with $p_T > 30$ $\rightsquigarrow$ hadronic activity
3. At least 1 "extra" lepton $\rightsquigarrow$ "probe" leptons

- We compute the $\frac{\text{tight}}{\text{loose}}$ ratio over the "extra" leptons.

Distributions after the cuts

ONLY Z SKIM

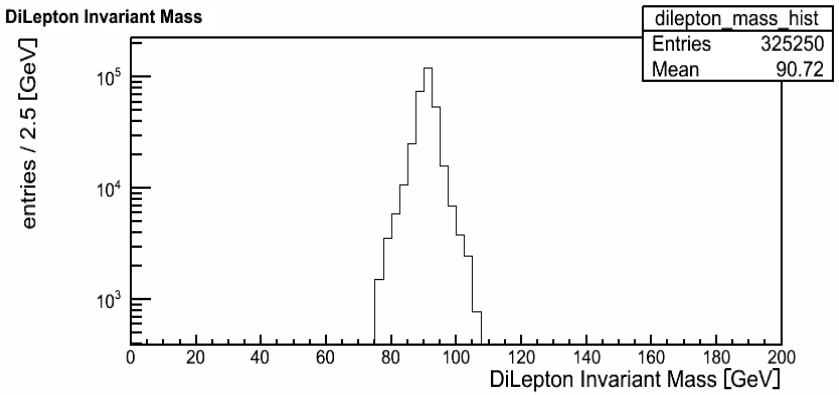


Figure: Invariant mass of the Z $\ell\ell$ pair.

Dataset	μ trigger
Channel	$Z \to \mu\mu$
Probe lepton	e

AT LEAST ONE JET

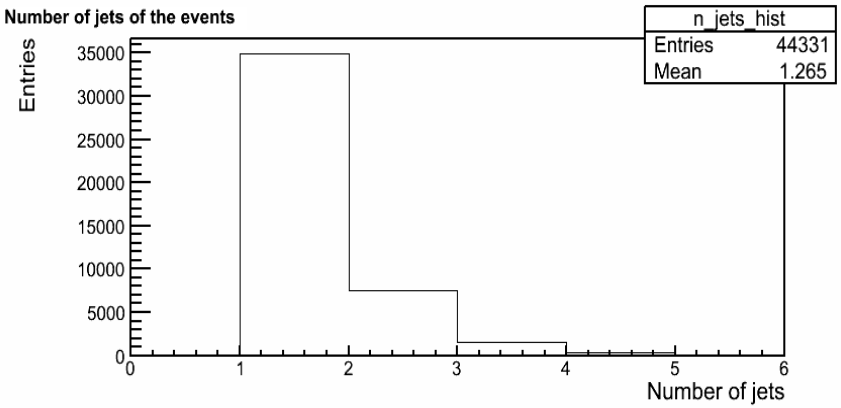


Figure: Number of jets.

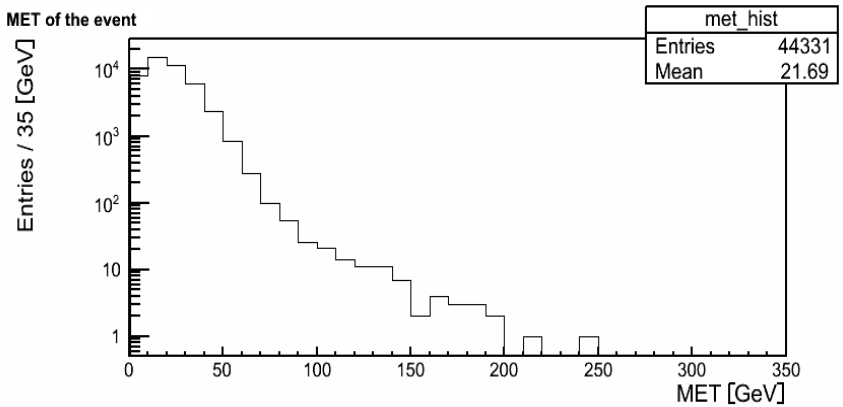


Figure: MET of the passed events.

Distributions after the cuts

AT LEAST ONE JET, MET < 25 GeV

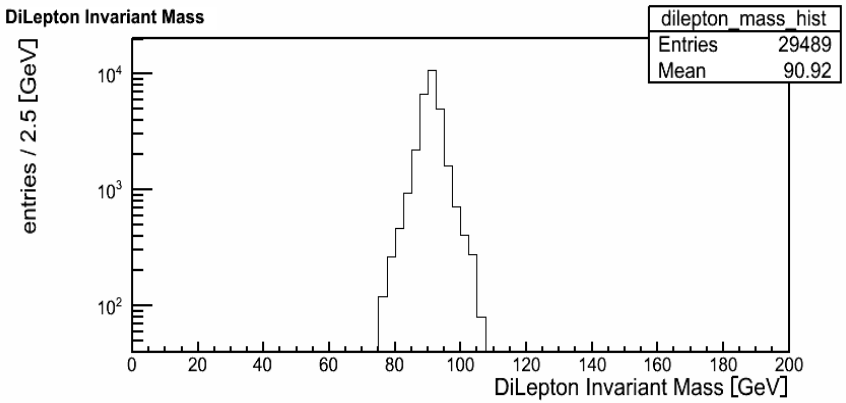


Figure: Invariant mass of the Z $\ell\ell$ pair.

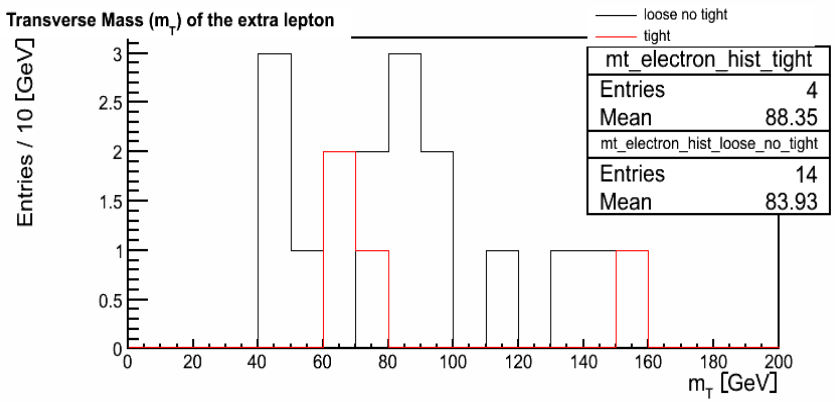


Figure: M_T of the extra lepton.

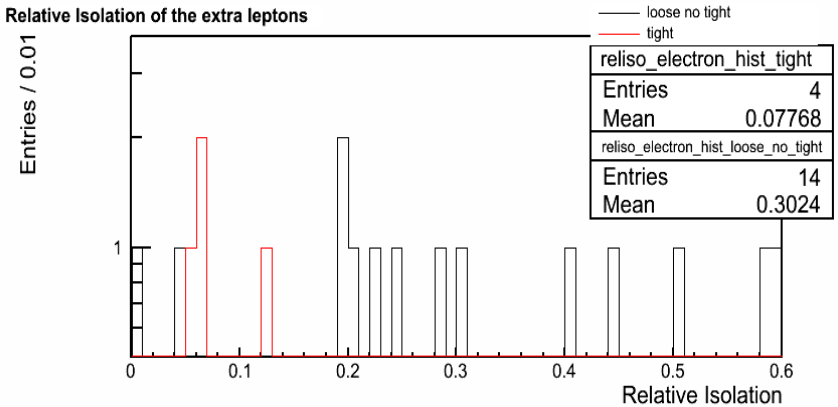


Figure: Relative isolation of the extra lepton.

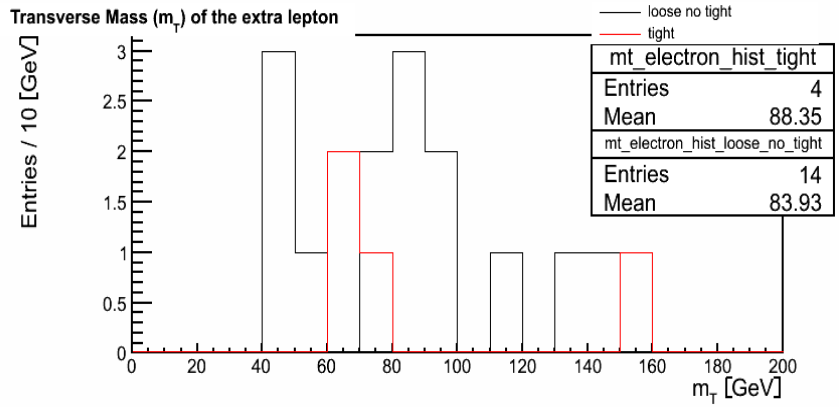


Figure: Invariant mass (extra lepton, jet).

Tight/Loose ratios

SINGLE MUON (Z-Skim: $Z \to \mu\mu$)

e

Cut	Passed	Loose	Tight	Tight/Loose
(none)	325250	76	21	$0.276 + 0.054 - 0.048$
①	238289	46	8	$0.174 + 0.063 - 0.049$
②	44331	30	11	$0.367 + 0.091 - 0.082$
①,②	29489	18	4	$0.222 + 0.111 - 0.082$

1. MET < 25 GeV
2. At least 1 jet with $p_T > 30$

Future tasks

For the remaining half of my Summer Student experience...

- Increase the statistics, using all the available datasets in 2012 and optimizing cuts.
- Compare with the MC simulations of Z + jets and other processes events, and compare them with the data
  $\rightarrow$ obtain the events containing non-prompt leptons (also rare events)!
- Add a similar analysis for the W + jets dataset.
- Repeat the "W removal" method used for the 2011 data on the 2012, and compare.
- Study a possible dependence of the $\frac{\text{tight}}{\text{loose}}$ ratio on variables such as the number of vertices of the event (pile-up).

Future tasks

For the remaining half of my Summer Student experience...

- Increase the statistics, using all the available datasets in 2012 and optimizing cuts.
- Compare with the MC simulations of Z + jets and other processes events, and compare them with the data
  $\rightarrow$ obtain the events containing non-prompt leptons (also rare events)!
- Add a similar analysis for the W + jets dataset.
- Repeat the "W removal" method used for the 2011 data on the 2012, and compare.
- Study a possible dependence of the $\frac{\text{tight}}{\text{loose}}$ ratio on variables such as the number of vertices of the event (pile-up).

Acknowledgements

I would like to thank:

- *Patrizia Azzi*, and the whole group working in the Top Partner Analysis.
- *Matteo Abis*, because without his previous work and his help in my discovery of Python, my task would have been much more difficult.