Search for Pair Production of First Generation Scalar Leptoquarks

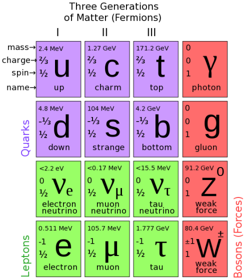
in $LQ\overline{LQ} \rightarrow e^{\pm} q \nu_e q$ at $\sqrt{s} = 1.96$ TeV

August 7, 2009

Xinlu Huang
Nevis Lab, Columbia University / Johns Hopkins University
Advisors:
Lidija Zivkovic
Gustaaf Brooijmans
Nevis Lab, Columbia University

Outline

1. What is a leptoquark?
    - Theoretical background
    - LQ phenomenology
2. How can one find it?
3. My analysis
4. Results and Outlook



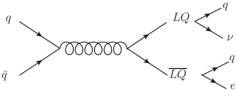
- Leptoquark (LQ) is predicted by many extensions of the Standard Model (SUSY, GUT, technicolor, etc.)
  - LQ can be a mediating boson, allowing interaction between leptons and quarks
    - In the SM, leptons and quarks do not directly interact
  - Can be scalar or vector field; has three generations
  - Short-lived and decays to a lepton and a quark
- Motivations:
  - Predicted by SM extensions
  - Unify leptons and quarks: reduce arbitrariness of the SM
  - Explain why 3 generations of matters
  - Signs of compositeness of quarks and leptons

Outline

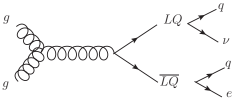
- LQ can be produced in singles or in pairs:
    - My analysis is on LQ pairs; independent of model assumptions
- Produced via quark-antiquark annihilation or gluon-gluon fusion:

$$q + \overline{q} \rightarrow LQ + \overline{LQ}$$
$$g + g \rightarrow LQ + \overline{LQ}$$

- Assume no intergenerational mixing
- LQ pair decays to 1 of 3 final states: $eeqq$, $e\nu_e qq$, and $\nu_e \nu_e qq$.
- Define **branching ratio** $\beta = Br(LQ \rightarrow \nu_e + q)$, then probability of LQ pair decaying to $e^{\pm} q \nu_e q$ is $2\beta(1-\beta)$.



Example of q-$\bar{q}$ annihilation



Example of g-g fusion

Outline

Challenges:

- Only produced at high energy (or else we would have seen it)
    - Wait for nature to produce them or try to make them ⇒ colliders!
- Short-lived
- Low probability of occurance (at least at low energy)

A collider search at DØ...

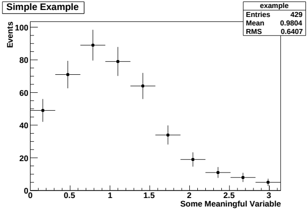
- produce anything possible - hard to isolate events of interest
    - pick a final state and have to deal with all processes that produce the chosen final state
- Compare number and shape between observed and expected, and report differences in term of statistical significance

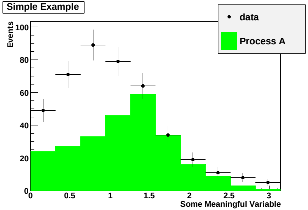
Outline

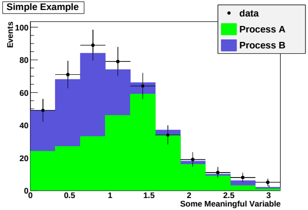
1. What is a leptoquark?
2. How can one find it?
    - The General Idea
    - A Grossly Simplified Example
    - More complicated details
3. My analysis
4. Results and Outlook

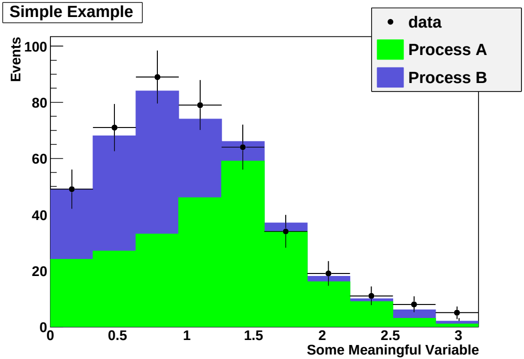
Suppose that...

- We collide 10000 proton-antiproton pairs...
- and it happens that 429 of the collisions end up producing exactly 1 electron, 2 quarks, and 1 neutrino
- We know from the Standard Model that only processes A and B produce the same final state. Then we simulate the detector response...
    - Process A has a probability to happen 2.58% of the times, or $10000 \times 0.0258 = 258$ events.
    - Process B has a probability to happen 1.76% of the times, or $10000 \times 0.0176 = 176$ events.



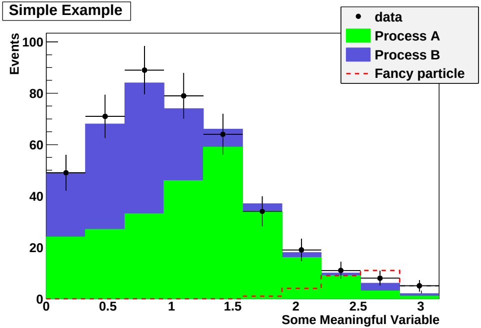






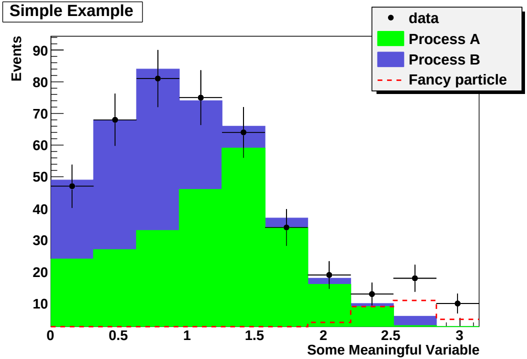
- Note the agreement between data and “background”:
  - at bin 3, for example, we observed ∼90 events, expect ∼35 from process A and ∼52 from process B; the difference is within 1σ
  - shows that our background models are sound (a must!)

Now, the interesting part: adding "signal" events



- Suppose we know how the detector will reponse when the "fancy particle" is produced during collision... then we can compare.
    - At bin 9, for example, we can see $\sim$ 10 obs. vs. $\sim$ 8 exp. events, but if "fancy particle" exists, we expect to see 12 more events above the background or 22 total. 10 vs. 22 $\Rightarrow$ not so likely
    - Then we can proceed to "set limit", i.e. the assumptions behind "fancy particle" probably needs modification.

Now, what if...



- Note the bump at the tail: observed more than 20 events, but expected $\sim 5 \Rightarrow$ statistically unlikely
- Note also the good agreement between data and background in other places, and that the signal just happens to be at where there is excess in data
- We might have found something!

Outline

- ~~Suppose we collide 10000 proton-antiproton pairs~~ ⇒
  - We collide more than we can handle (396 ns per collision):
    - A sophisticated trigger system is used to filter out uninteresting events in real time
    - Having lots of data is a good thing!
  - Commonly measure in term of **integrated luminosity** in fb^{-1}:

$$\mathcal{N} = \sigma_{eff} \times \int \mathcal{L}\, dt, \begin{cases} \sigma_{eff} & \text{the effective inelastic cross section of } p\bar{p} \text{ collision} \\ \mathcal{L} & \text{the instantaneous luminosity} \end{cases}$$

  - The Tevatron has collected 0.16 fb^{-1} during Run I (1992-1996), $\sim 1\ \mathrm{fb}^{-1}$ during Run IIa (2002-2005), and almost 4 fb^{-1} so far during Run IIb (2005-now).
    - My analysis is based on Run IIb data.

- ~~Process A has a probability to happen 2.58% of the times...~~ ⇒
- We measure probability in term of total **cross section** (σ):
  - In classical hit-or-miss scattering, cross section ⇔ target area
  - Can be extended to represent interaction probability - “effective” target area depending on the strength of interaction force field
  - Commonly measured in picobarn (pb)
- In collisions of luminosity 3.9 fb^{-1}, LQ of mass 300 GeV has cross section of 0.013 pb ⇒ 3.9 $\mathrm{fb}^{-1} \times 0.013\mathrm{pb} = 50.7$ events are expected to produce a pair of 300 GeV LQs - not frequent at all!

- ~~end up producing exactly 1 electron, 2 quarks, and 1 neutrino...~~ ⇒
- We can't measure the particles directly, but their energy, momentum, projectory, etc.
  - electron ⇒ clusters of energy deposition in calorimeter with high fraction deposited in the EM sections. $\mathrm{EM\_Likelihood}$ quantify the id. likelihood.
  - quark ⇒ jet: hadronization of quarks produce showers of secondary particles whose energy deposit withint a certain cone of radius $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.5$
  - neutrino ⇒ "missing energy": neutrinos are not detected, but can be inferred from conservation of energy
- Because of composite nature of proton, the colliding quark carry indeterminate fraction of KE of the proton - collisions may happen in a boosted frame.
  - Energy conservation only useful in transverse plane.
  - Only information about neutrinos is missing transverse energy ($\not{E}_{\mathrm{T}}$): the negative vector sum of energies in all detector cells.
- The final state 1 electron, 2 quarks, and 1 neutrino is actually a $e\not{E}_{\mathrm{T}}jj$ events, topologically speaking

- ~~Then we simulate the detector response...~~ ⇒
- To reduce statistical error, lots of events were generated then normalized to correct cross section (by a factor of $\mathcal{N} \times \sigma / N_{generated}$)
- Since MC samples are directly used in analysis, corrections on various levels are applied:
  - Normalization accounting for NLO for LO only generators
  - Reweighting to correct for idiosyncracies of MC generators
  - Luminosity profile reweighting
  - Triggers corrections

Outline

- **Data**: Collected during Run IIb of the Tevatron, 3.89 fb^{-1}
- **SM Background**: W+jets, Z+jets, diboson (WW or WZ), single top, and top pairs
  - Simulated with Monte Carlo technique, using generators ALPGEN (most processes), COMPHEP (single top), and PYTHIA (dibosons).
- **Instrumental background**: QCD multijet (next slide)
- **Signal**: for each assumed LQ masses in 200, 220, 240, 260, 280 and 300 GeV, a Monte Carlo sample is generated by PYTHIA:
  - during generation, LQ pair is allowed to decay to any of the three possible final state: $e^{\pm}q\nu_e q$, $eeq\bar{q}$, $\nu_e\nu_e q\bar{q}$; then $e^{\pm}q\nu_e q$ final state is selected.
  - generated events are processed through a standarized analysis package, Common Analysis Framework (CAFE)

QCD multijet background estimation

- Detector limitation ⇒ a jet can be misidentified as an electron, thus multijet events and other events without e can be identified as having final state $e^{\pm} q \nu_e q$
- Significant background, but not simulated well; estimated from data and background MC using a Matrix Method:
  - divided data a **tight sample** ($\mathrm{EM\_Likelihood} > 0.85$) and a **loose sample** (no constrain on $\mathrm{EM\_Likelihood}$): $\mathcal{S}_{Tight} \subset \mathcal{S}_{Loose}$

$$\text{define} \begin{cases} \varepsilon_{QCD} & \text{efficiency for a fake electron to pass the } \textit{tight} \text{ criterion} \\ \varepsilon_{sig} & \text{efficiency for a real electron to pass the } \textit{tight} \text{ criterion} \end{cases}$$

  - Then solve for $N_{tight}^{QCD} = \varepsilon_{QCD} N_{loose}^{QCD}$, number of QCD events in the tight sample:

$$\begin{cases} N_{loose} & = N_{loose}^{sig} + N_{loose}^{QCD} \\ N_{tight} & = \varepsilon_{sig} N_{loose}^{sig} + \varepsilon_{QCD} N_{loose}^{QCD} \end{cases} \quad \Rightarrow \quad N_{tight}^{QCD} = \varepsilon_{QCD} \frac{\varepsilon_{sig} N_{loose} - N_{tight}}{\varepsilon_{sig} - \varepsilon_{QCD}}$$

Outline

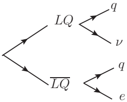
1. What is a leptoquark?
2. How can one find it?
3. My analysis
    - Data set and MC samples
    - Leptoquark reconstruction
    - Preselection
    - Further cuts
4. Results and Outlook

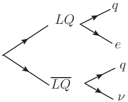
Why?

- allow more efficient cuts
- can measure properties of leptoquarks, if they exist

How?

- Two possible pairings

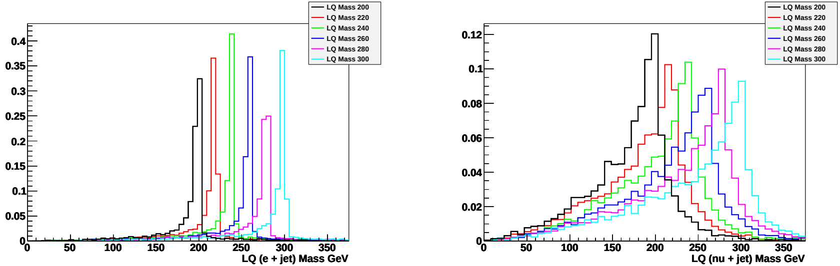




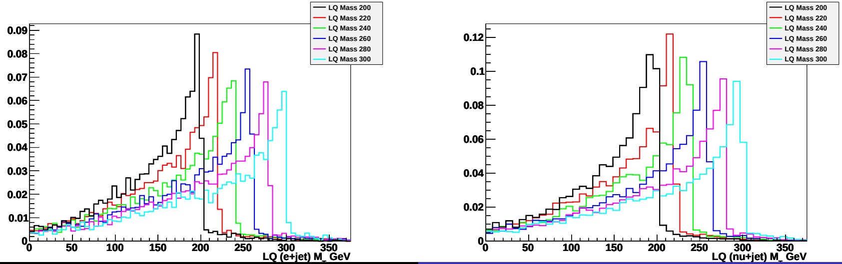
- Several algorithms were attempted; results can be compared to true information in MC
- Algorithm by minimizing differences in M_T is chosen; $\sim$ 75% successful

Reconstructed LQ masses:

(ν is assumed to be a massless particle with no p_z for reconstruction purposes)



Reconstructed LQ M_T:

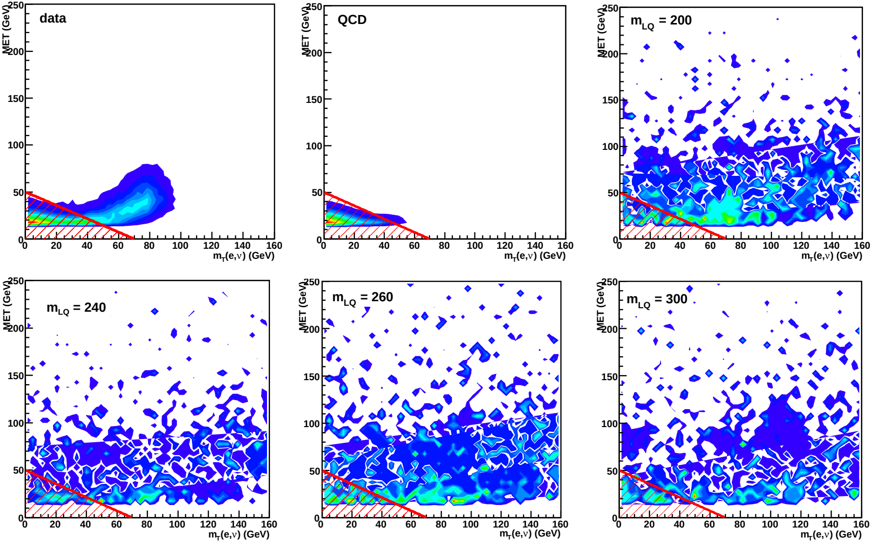


Outline

Preselection

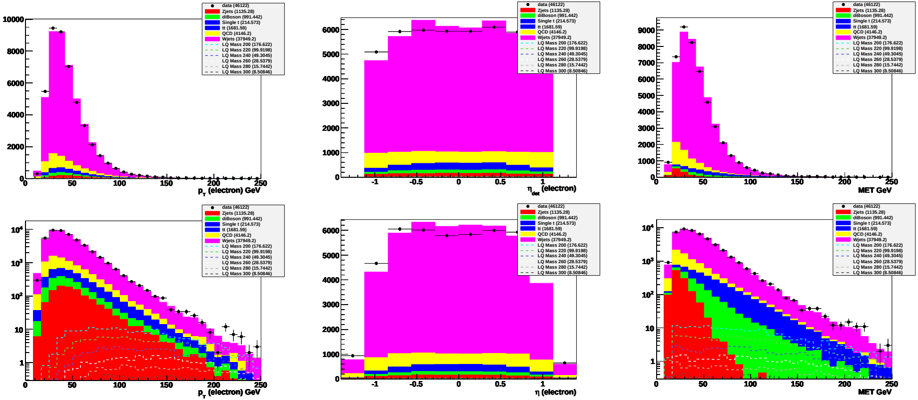
- Entire dataset contains 10^{10} events; analysis on the entire set can takes very long. Preselection ⇒ smaller working data set.
- **Basic cuts:**
  - Single *tight* electron with $p_T \geq 15$ GeV and $|\eta_{det}| \leq 1.1$
  - 2 or more of vertex confirmed jets; both leading jets has $p_T \geq 20$ GeV and $|\eta_{det}| \leq 2.5$
  - $\not{E}_{\mathrm{T}} > 15$ GeV
- **Triangular cut**: $\not{E}_{\mathrm{T}}/50 + m_T/70 \geq 1$ (to eliminate QCD background, next slide)
- **Additional reweighting**: 46122 data events and 43504 expected background events - additional weight factor on W+jets background:

$$S = \frac{N_{data} - (N_{background} - N_{W+jets})}{N_{W+jets}} = 1.074$$

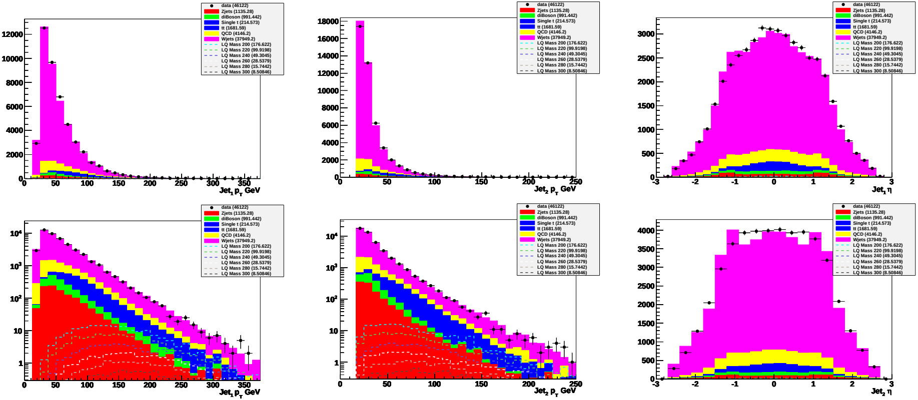


Triangular cut on $\not{E}_T$-$M_T(e,\nu_e)$ plane to eliminate QCD background.

Data and background agreement after preselection cuts - e and ν



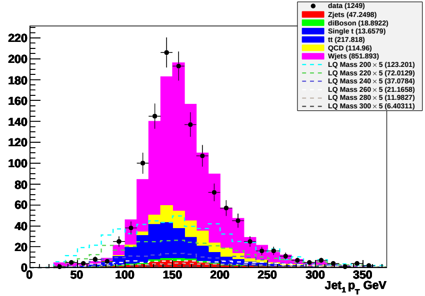
Data and background agreement after preselection cuts - jets



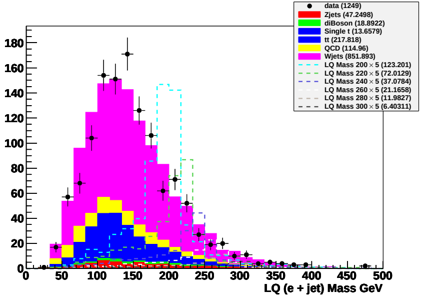
Outline

Futher cuts

- Since no significant excess of data over background is found, the rest of analysis will be focused on optimizing cut to set limit.
- Goal: reduce background/signal ratio
  - Reduce systematics
  - Cut at places where background is prominent while signal is small



example of a bad discriminating variable



example of a good discriminating variable

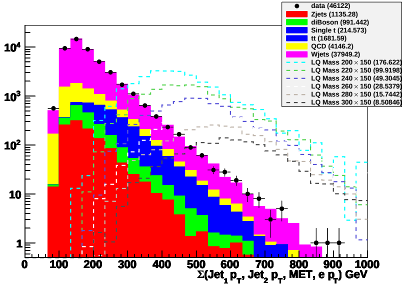
Let $S_T = Jet_1\ p_T + Jet_2\ p_T + \not{E}_T + e\ p_T$.

Cut:

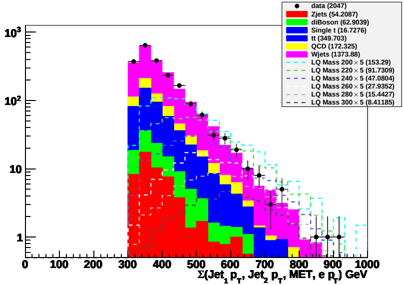
$S_T \geq 320$ GeV

since the signal events produce products with high transverse momentum:

S_T Before cut:



S_T after cut:



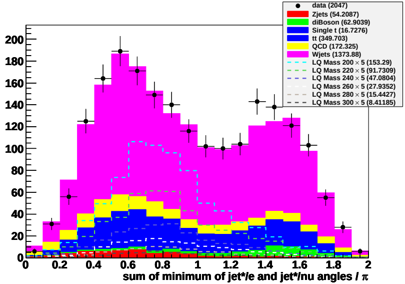
	data	sum bg	Z+jets	diBoson	Single t	tt	QCD	W+jets	m_{LQ} 240 GeV
Before cut	46122.0	46118.3	1135.3	991.4	214.6	1681.6	4146.2	37949.2	49.3
After cut	2047.0	2029.8	54.2	62.9	16.7	349.7	172.3	1373.9	47.1
Eff.	0.044	0.044	0.048	0.063	0.078	0.208	0.042	0.036	0.955

Cut:

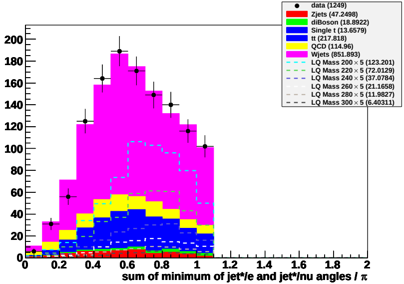
$\min(\Delta\phi(j_1, e), \Delta\phi(j_2, e)) + \min(\Delta\phi(j_1, \nu), \Delta\phi(j_2, \nu)) \leq 1.1\pi.$

Since the LQ produced from rest have high p_T, the daughter products are expected to be close to each other in ϕ. The variable $\min(d\phi(j_1, e), \Delta\phi(j_2, e)) + \min(\Delta\phi(j_1, \nu), d\phi(j_2, \nu))$ provides good separation between background and signal.

Before cut:



after cut:

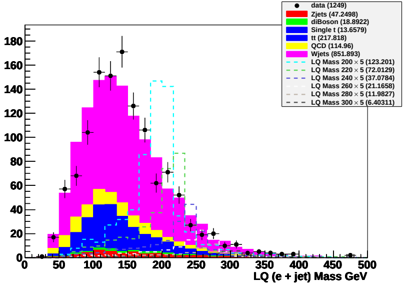


	data	sum bg	Z+jets	diBoson	Single t	tt	QCD	W+jets	m_{LQ} 240 GeV
Before cut	2047.0	2029.8	54.2	62.9	16.7	349.7	172.3	1373.9	47.1
After cut	1249.0	1264.5	47.2	18.9	13.7	217.8	115.0	851.9	37.1
Eff.	0.610	0.623	0.872	0.300	0.816	0.623	0.667	0.620	0.788

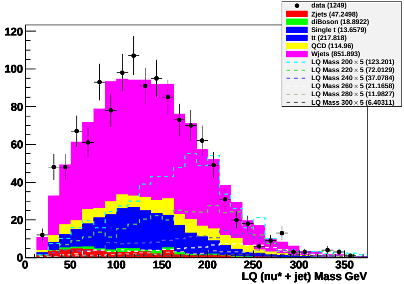
Cut:

$m_{LQ}(\text{reco. from } e + jet) \geq 160$ GeV
$m_{LQ}(\text{reco. from } \nu + jet) \geq 700$ GeV

$m_{LQ}(e, jet)$ before cut:



$m_{LQ}(\nu, jet)$ before cut:



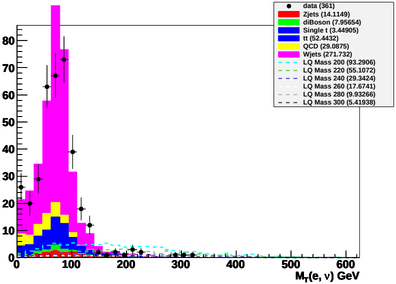
	data	sum bg	Z+jets	diBoson	Single t	tt	QCD	W+jets	m_{LQ} 240 GeV
Before cut	449.0	456.4	20.3	9.5	3.9	59.5	35.4	327.8	31.0
After cut	361.0	378.8	14.1	8.0	3.4	52.4	29.1	271.7	29.3
Eff.	0.804	0.830	0.696	0.841	0.878	0.882	0.822	0.829	0.947

Cut:

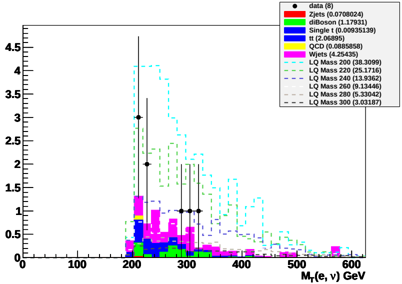
$M_T(e,\nu) \geq 150$ GeV

$M_T(e,\nu)$ is the transverse mass of W in the W+jets background, while the signals have flat shape.

$M_T(e,\nu)$ before cut:



$M_T(e,\nu)$ after cut:



	data	sum bg	Z+jets	diBoson	Single t	tt	QCD	W+jets	m_{LQ} 240 GeV
Before cut	361.0	378.8	14.1	8.0	3.4	52.4	29.1	271.7	29.3
After cut	8.0	7.7	0.1	1.2	0.0	2.1	0.1	4.3	13.9
Eff.	0.022	0.020	0.007	0.150	0.000	0.040	0.003	0.016	0.474

Outline

Systematic Uncertainties

- Systematic uncertainties for signal and MC background are included in final limit calculation.
- Flat (constant) systematics on background, affects normalization:
  - Integrated luminosity: 6%
  - Electron identification and trigger: 4%
  - Cross section on varies processes: 6% (single top and top pairs), 7% (diboson)
  - QCD: 20%
- Shape systematics on background, affects distribution in shape:
  - These are obtained via generation background MC again but varying the respective input paramters by $\pm 1\sigma$
  - Jet energy scale (JES)
  - Jet energy resolution (JER): accounting for differences in jet energy resolution between data and MC
  - Jet identification and efficiency: accounting for differences in jet identification and reconstruction efficiency
- A flat 10% uncertainty is assigned to LQ cross section.

Outline

1. What is a leptoquark?
2. How can one find it?
3. My analysis
4. Results and Outlook
    - Systematic Uncertainties
    - Setting Limit
    - Conclusion and future work

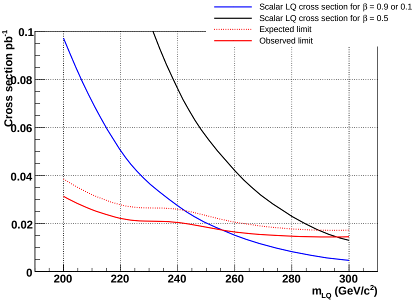
Limit Calculating Method

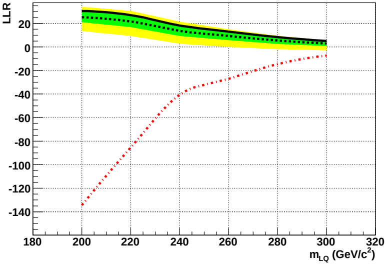
- COLLIE (COnfidence Level LImit Evaluator) is used to evaluate limit
  - COLLIE use a semi-frequentist approach
  - Compares a null hypothesis (just background) to a test hypothesis (background + signal)
- We compare theoretical σ to observed σ at 95% CL
  - Expected σ: observsed σ if background agrees perfectly with data

m_{LQ} GeV	σ ($\beta = 0.5$)	σ ($\beta = 0.9$)	Expected σ	Observed σ
200	0.27	0.097	0.039	0.031
220	0.14	0.050	0.028	0.022
240	0.076	0.002	0.026	0.020
260	0.042	0.015	0.021	0.016
280	0.023	0.007	0.018	0.015
300	0.013	0.005	0.017	0.014

Table: Cross sections: theoretical, expected, and observed

- Previous search in the same channel (DØ Note 5856) has set limit on scalar first generation LQ at 264 GeV for $\beta = 0.5$.
- My limits:





- For $\beta = 0.5$, the limit for scalar LQ_1 mass is about 290 GeV (expected) and 296 GeV (observed).

Outline

1. What is a leptoquark?
2. How can one find it?
3. My analysis
4. Results and Outlook
    - Systematic Uncertainties
    - Setting Limit
    - Conclusion and future work

- I analyzed 3.89 fb^{-1} of DØ data to search for scalar first generation leptoquark using channel $LQ\overline{LQ} \rightarrow e^{\pm} q \nu_e q$
- No significant excess of data over background was found, and I set a limit on scalar leptoquark mass of 296 GeV, improving on the previous result of 264 GeV.
- Future work:
  - Hopefully continue working on the project to bring it to publishable stage
  - Include the 1 fb^{-1} of data from Run IIa in the analysis
  - Change the MC signal samples from private to officially requested ones

Acknoledgement

Many, many thanks to:

- Lidija Zivkovic
- Thomas Gadfort
- Gustaaf Brooijmans
- Mike Shaevitz and John Parsons
- NSF