CMS Physics Analysis Summary

Contact: cms-pag-conveners-exotica@cern.ch 2009/07/10

Search for Pair Production of First Generation Scalar Leptoquarks at the CMS Experiment

The CMS Collaboration

Abstract

We investigate the discovery potential of the CMS experiment for pair production of first generation scalar leptoquarks that decay to an electron and quark, assuming an integrated luminosity of 100 pb^{-1} and pp collisions at $\sqrt{s} = 10$ TeV. We discuss reconstruction and identification of high-energy electrons and jets, and optimization of the event selection. Data-driven techniques are used to determine the main standard model backgrounds. The CMS discovery and exclusion potentials for different leptoquark mass hypotheses are presented.

1 Introduction

This note describes analysis techniques appropriate for the search for first generation scalar leptoquarks using data from the CMS experiment taken with 5 TeV proton beams and corresponding to 100 pb^{-1} of integrated luminosity. The methods described are evaluated using Monte Carlo (MC) simulations.

Leptoquarks (LQ) are new exotic particles conjectured to have both baryon and lepton number and to decay primarily into leptons and quarks. Several approaches to physics beyond the standard model (SM) predict the existence of leptoquarks, including Grand Unified Theory, Technicolor, and composite models [1].

Leptoquarks carry color and fractional electric charge, and can be either scalar or vector particles. The three generations of predicted leptoquarks correspond to the families of quarks and leptons in the SM. Intergenerational coupling of leptoquarks is highly constrained by the experimental limits on flavor-changing neutral currents and proton decay [1, 2].

The parameters of the model are i) M_{LQ}, the LQ mass, ii) β, the branching fraction $LQ \rightarrow l + q$ where l is a charged lepton and q is a quark, and iii) λ_{LQlq}, the coupling between the LQ, lepton and quark ($LQ - l - q$ vertex). The complementary decay $LQ \rightarrow \nu + q$, where ν is a weak-isospin partner of the charged lepton, has branching fraction 1-β. Results from the experiments at the HERA accelerator give upper limits on the LQ production cross section that restrict λ_{LQlq} to be small (comparable or less than the strength of the electromagnetic coupling λ_{EM}) for $M_{LQ} < 300$ GeV [3]. In this analysis we assume the parameter $\lambda_{LQlq} = \lambda_{EM} \approx 0.3$, that leads to Γ_{LQ}/M_{LQ} of about 0.2% for scalar leptoquarks.

In hadron-hadron collisions at Tevatron and LHC, pair production of leptoquarks takes place mainly through gluon-gluon fusion and quark-antiquark annihilation. The cross sections for both of these sub-processes are almost independent of the value of λ_{LQlq}, since there is no $LQ - l - q$ vertex in the Feynman diagram for LQ pair production at leading order. The production of a single leptoquark in association with a lepton is also possible via quark-gluon fusion at a lower rate. The cross section for single leptoquark production becomes comparable to the one for pair production only for $M_{LQ} \approx 1$ TeV [4], which is well above the experimental reach of this start-up analysis. Cross sections for single and pair production of vector leptoquarks are expected to be larger than for scalar leptoquarks [4]. The most stringent limit from Tevatron is 256 GeV/c^2 (assuming $\beta = 1$ for the electron channel) on the mass of first-generation scalar leptoquarks, independent of the coupling λ_{LQlq} [5, 6].

This note examines the potential reach of a search for pair production of first-generation scalar leptoquarks that decay into electrons and (light) quarks with an unknown branching fraction β. This process yields an experimental signature that is quite striking, with two high transverse momentum (p_T) electrons and two high p_T jets (eejj channel), and a peak in the electron-jet invariant mass spectrum that corresponds to the LQ mass. No true missing transverse momentum is expected.

2 Monte Carlo Samples

MC signal samples are generated using PYTHIA version 6.227 with leptoquark masses ranging from 250 to 600 GeV/c^2 and $\beta = 1$.

The "Full Simulation" (FullSim) signal samples are produced using the official CMS software for generation, simulation (based on GEANT4), digitization and reconstruction.

These FullSim samples are supplemented with "Fast simulation" (FastSim) signal samples. The FastSim uses a parameterization of the detector response.

The FullSim SM background used for this analysis is described below.

- $t\bar{t}$ + jets events, generated using MADGRAPH [7], inclusive production (all decays);
- Z/γ + N jets events (with $N_{\text{jet}} \leq 4$), generated using MADGRAPH, Z decaying into charged leptons;
- W + N jets events (with $N_{\text{jet}} \leq 4$), generated using MADGRAPH, W decaying into leptons;
- QCD multi-jet events, generated with MADGRAPH, inclusive production in bins of H_T^{jet} [1] from 100 GeV/c to 1 TeV/c;
- VV + jets events (VV = WW, WZ or ZZ), generated with MADGRAPH, W and Z decaying into leptons.

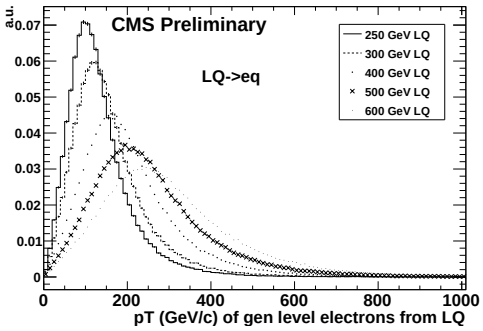
Table 1 lists the signal and background MC samples used in this analysis with number of events and cross sections. All samples were produced with no misalignment or miscalibration of the detector. Pile-up was not included in the simulation of these samples.

Table 1: Signal and background MC samples used in this analysis, with type of simulation used (full or fast), the number of generated events, and the cross section at NLO and LO for $\sqrt{s} = 10$ TeV. The k-factor is also reported. Leptoquark cross sections at NLO and LO are based on the calculation of Kramer *et al.* [8] (thanks to Micheal Kramer for providing the updated numbers for $\sqrt{s} = 10$ TeV). $t\bar{t}$ cross section at NLO is taken from [9] using the central value for the CTEQ6.6 PDF set. The approximate K-factor for Z/W + jets samples is derived from [10].

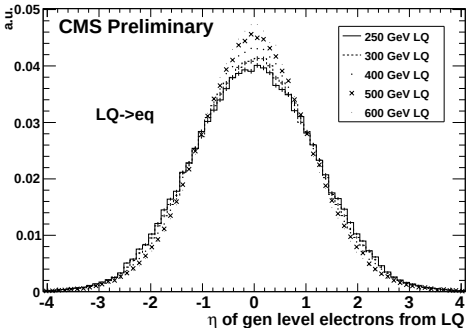
MC Sample	Full/Fast Simulation	N. Events Analyzed	Equivalent Luminosity (pb^{-1})	σ_{NLO} (pb)	σ_{LO} (pb)	K-factor
$M_{LQ} = 250$ GeV/c^2	Full	52k	$5.15{\cdot}10^3$	10.1	6.53	1.547
$M_{LQ} = 400$ GeV/c^2	Full	63k	$84{\cdot}10^3$	0.75	0.462	1.628
$M_{LQ} = 250$ GeV/c^2	Fast	125k	$12.4{\cdot}10^3$	10.1	6.53	1.547
$M_{LQ} = 300$ GeV/c^2	Fast	127k	$33.4{\cdot}10^3$	3.8	2.42	1.57
$M_{LQ} = 400$ GeV/c^2	Fast	150k	$200{\cdot}10^3$	0.75	0.462	1.628
$M_{LQ} = 500$ GeV/c^2	Fast	125k	$635{\cdot}10^3$	0.197	0.118	1.669
$M_{LQ} = 600$ GeV/c^2	Fast	131k	$2.12{\cdot}10^6$	0.0617	0.0617	1.723
$t\bar{t}$ + N jets	Full	0.905M	$2.19{\cdot}10^3$	414	317	1.31
Z/γ + N jets	Full	1.159M	275	4218	3700	1.14 (@14 TeV)
Z/γ + N jets	Fast	8.46M	$2{\cdot}10^3$	4218	3700	1.14 (@14 TeV)
W + N jets	Full	6.107M	134	45600	40000	1.14 (@14 TeV)
VV + N jets	Full	101.8k	$8.63{\cdot}10^3$		11.8	
QCD ($H_T \in [100, 250]$ GeV/c)	Full	13.721M	0.915		$15.0 \cdot 10^6$	
QCD ($H_T \in [250, 500]$ GeV/c)	Full	3.698M	9.25		$400 \cdot 10^3$	
QCD ($H_T \in [500, 1000]$ GeV/c)	Full	3.807M	272		$14 \cdot 10^3$	
QCD ($H_T > 1000$ GeV/c)	Full	0.654M	$1.77{\cdot}10^3$		370	

Figure 1 shows the p_T and η distributions at generator level of the electrons from the LQ decay. Quarks from the LQ decay have almost identical p_T and η distributions.

[1] $H_T^{jet} = \sum_{jets} p_T^{jet}$



a.



b.

Figure 1: Distribution of the p_T (a) and η (b) variables at generator level for the electrons coming from the LQ decay.

3 Trigger Studies

The presence of high energy electrons in the final state is exploited for the online selection of candidate leptoquark-antileptoquark events.

Two trigger menus ("low luminosity" and "high luminosity" menus) are considered for the two possible luminosity scenarios of $8 \cdot 10^{29}$ cm^{-2}s^{-1} and 10^{31} cm^{-2}s^{-1}, respectively, expected for the LHC start-up. The trigger of the low luminosity menu selects events containing an electromagnetic cluster in the calorimeter. The corresponding Level-1 trigger requires an electromagnetic object with $E_T > 8$ GeV and relaxed ID requirements (a cut on the hadronic energy, and a shape cut on the electromagnetic energy deposit in the trigger tower with the highest energy are applied). At the high level trigger (HLT), the applied threshold is $E_T > 15$ GeV and no isolation or track-matching is required. The expected HLT rate of this trigger is around 10 Hz for the low luminosity scenario. At high luminosity this trigger is prescaled, so instead a trigger with an HLT threshold of $E_T > 25$ GeV and no isolation or track-matching is used. The expected HLT rate of this trigger is around 20 Hz for the high luminosity scenario.

The efficiencies for leptoquarks with these triggers are studied at two leptoquark mass points using the FullSim samples, and are close to 100%, as shown in Table 2.

Table 2: Efficiencies of the HLT electromagnetic (EM) triggers for LQ masses of 250 and 400 GeV/c^2 (FullSim samples are used). L1 efficiencies are included. The fractional statistical uncertainties on the trigger efficiencies are less than 0.1%.

M_{LQ} (GeV/c^2)	EM trigger $E_T > 15$ GeV	EM trigger $E_T > 25$ GeV
250	99.6%	99.3%
400	99.7%	99.6%

4 Reconstructed Objects

4.1 Electron Reconstruction and Selection

A reconstructed electron object is a cluster of depositions of energy in the electromagnetic calorimeter (ECAL) that has an extrapolated track pointing to it. This study uses the standard,

basic electron identification requirements [11]. Standard corrections for the electromagnetic (e.m.) energy scale are applied.

Additional selection requirements, including isolation, which are optimized for electrons with energies of hundreds of GeV, are used to further reduce backgrounds from fake electrons [12].

The definitions of the variables are given below:

- H/E: the ratio of the energy in all the hadron calorimeter (HCAL) channels in a cone $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.1$ from the electromagnetic cluster associated with the electron, and the energy of the electromagnetic cluster itself.
- $\sigma_{\eta\eta}$: a variable reflecting the distribution along η of the electron shower

$$\sigma_{\eta\eta} = \frac{\sum_k^{5x5} w_k (\iota_\eta^k - \bar{\iota}_\eta)^2}{\sum_k^{5x5} w_k} \quad ,$$

  where ι_η^k is the index of the η position of the k^{th} crystal in a 5x5 matrix of crystals centered on the seed crystal of the cluster associated to the electron, $\bar{\iota}_\eta$ is the energy weighted mean index of the η position of the 5x5 block, and w_k is a weight given to each crystal as defined by

$$w_k = 4.2 + \ln(\frac{E_k}{E_{5x5}}) \quad ,$$

  where E_k is the energy of the k^{th} crystal and E_{5x5} is the total energy deposited in the 5x5 block. The values for the endcap are corrected for the different crystal size with respect to the barrel region.
- $\Delta\eta^{trk-SC}$ and $\Delta\phi^{trk-SC}$: the difference in η and ϕ between the position of the electromagnetic cluster associated to the electron and the track position extrapolated from the inner tracker to the ECAL surface.
- Number of tracks (N_T) with $p_T > 1.5$ GeV/c in an annulus $0.02 < \Delta R < 0.2$ around the direction of the electron.
- Track isolation (Track iso): the sum of the p_T of tracks in the annulus defined above for N_T.
- Electromagnetic isolation (EM iso): the transverse electromagnetic energy of all reconstruction ECAL energy deposits in a cone $\Delta R < 0.3$, centered on the position of the electron in the calorimeter, excluding those that are part of the electromagnetic cluster.
- Hadronic isolation (HAD iso): the transverse hadronic energy of all the reconstructed HCAL energy deposits in an annulus $0.15 < \Delta R < 0.3$, centered on the position of the electron in the calorimeter.

Table 3 summarizes the cuts used in this analysis.

Because electrons from LQ decays are usually isolated, and have large p_T (hundreds of GeV/c), they can be reconstructed with high efficiency. Table 4.1 shows the electron acceptance, A_{gen} (the acceptance cuts are $p_T^{gen} > 30$ GeV/c and $|\eta^{gen}| < 2.5$), the electron reconstruction efficiency within the acceptance region, $\varepsilon_{\text{reco}}$, the efficiency for a reconstructed electron within the acceptance region to pass also the ID and isolation requirements, $\varepsilon_{\text{ID+iso}}$, and the overall efficiency, $A_{\text{gen}} \cdot \varepsilon_{\text{reco}} \cdot \varepsilon_{\text{ID+iso}}$, for single electrons from LQ decays for a 400 GeV/c^2 LQ mass. The overall single electron efficiency, including acceptance, is $\approx 80\%$.

Table 3: Electron ID and isolation criteria. The barrel (endcap) region covers pseudo-rapidity with $|\eta| < 1.442$ ($1.560 < |\eta| < 2.5$).

ID Variables			Isolation Variables		
Variable	Barrel	Endcap	Variable	Barrel	Endcap
H/E	< 0.05	< 0.1	N_T	< 4	< 4
$\sigma_{\eta\eta}$	< 0.011	< 0.0275	Track iso (GeV/c)	< 7.5	< 15
$\lvert\Delta\eta^{trk-SC}\rvert$	< 0.005	< 0.007	EM iso (GeV)	$< 6 + 0.01 \cdot E_T$	$< 6 + 0.01 \cdot E_t$
$\lvert\Delta\phi^{trk-SC}\rvert$	< 0.09	< 0.09	HAD iso (GeV)	$< 4 + 0.005 \cdot E_T$	$< 4 + 0.005 \cdot E_t$

Table 4: Electron acceptance (A_{gen}), reconstruction efficiency ($\varepsilon_{\text{reco}}$), combined ID and isolation efficiency ($\varepsilon_{\text{ID+iso}}$), and overall efficiency ($A_{\text{gen}} \cdot \varepsilon_{\text{reco}} \cdot \varepsilon_{\text{ID+iso}}$) for single electrons from decays of LQ's with a mass of 400 GeV/c^2 (FullSim sample used). The fractional statistical uncertainties on the efficiencies are less than 0.2%.

LQ mass (GeV/c^2)	A_{gen}	$\varepsilon_{\text{reco}}$	$\varepsilon_{\text{ID+iso}}$	$A_{\text{gen}} \cdot \varepsilon_{\text{reco}} \cdot \varepsilon_{\text{ID+iso}}$
400	96.4%	93.8%	88.9%	80.4%

4.2 Jet Reconstruction and Selection

The jet reconstruction algorithm used in this analysis is an iterative cone algorithm with radius $\Delta R = 0.5$. Jets are removed from the jet collection if they are within a ΔR of 0.5 from an electron passing ID and isolation requirements.

Three types of jet energy corrections (JEC), are applied in this analysis. The relative jet corrections provide uniform jet energy response in η. The absolute jet corrections aim to correct jet energies back to the particle level as a function of jet p_T. In addition, jet energy flavor corrections are applied to take into account the flavor of the originating parton, under the assumption that the leading jets from LQ decay come from light quarks.

5 Event Selection

The basic strategy to identify the pair production of particles that decay to a light quark jet (j1 and j2, respectively) and an electron (e1 and e2, respectively) is to look for events with large S_T, defined as $S_T \equiv p_T(e_1) + p_T(e_2) + p_T(j_1) + p_T(j_2)$. We can use the invariant mass of the electron-jet pairs to help confirm the source of these events, as a leptoquark would produce a bump in the distribution of this invariant mass. The following criteria are used to select candidate signal events with high efficiency while minimizing backgrounds:

1. at least 2 isolated electrons, both with $p_T > 30$ GeV/c
2. at least 2 jets, both with $p_T > 50$ GeV/c and $|\eta| < 3$
3. $M_{ee} > 100$ GeV/c^2
4. $S_T > f(M_{LQ})$, where $f(M_{LQ})$ is a function of the hypothesized LQ mass.

The specific values of the kinematic cuts are determined using a cut-optimization procedure described in section 5.1.

The cut on M_{ee}, the invariant mass of the electron pair, removes background events from the Z/γ+jets process as shown in Figure 2 a). The S_T cut was used by the D0 collaboration in [13]. In that paper, an event selection optimization that considered combinations of several kinematic variables showed that, at Tevatron energies, little is gained by adding cuts on other variables. The distribution of S_T for the present analysis is shown in Figure 2 b).

There are two ways to combine two electrons and two jets to make two electron-jet pairs. For each event, the combination with the minimum difference between the invariant masses of the two electron-jet pairs (ΔM_{ej}) is chosen[2]. The resulting M_{ej} distribution after the full selection is shown in Figure 3 for the signal and backgrounds.

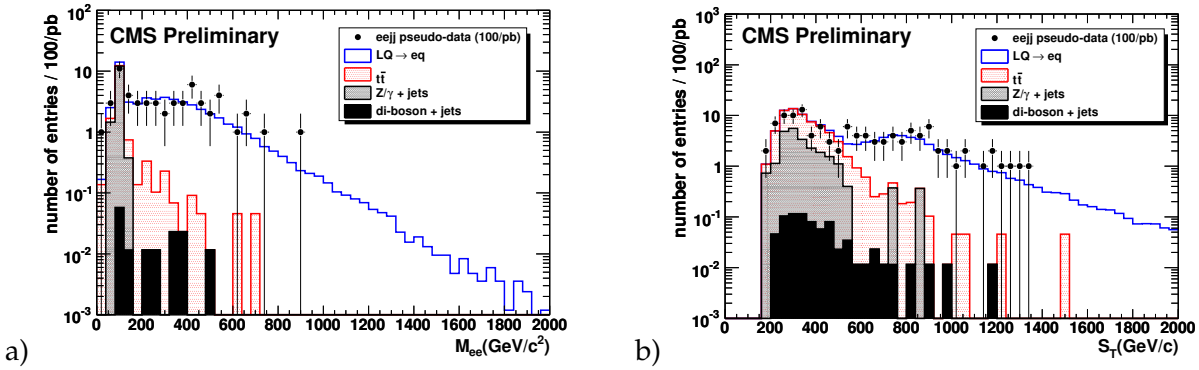


Figure 2: a) Invariant mass of the electron pair, M_{ee}. b) Scalar sum of the p_T of the 2 leading electrons and 2 leading jets. In each histogram, the distributions for the signal with $M_{LQ} = 400$ GeV/c^2 and the contributing backgrounds (with the exception of the QCD multi-jet background, see Section 6.3) are shown after applying all cuts except the one involving the plotted variable. The histograms are cumulative. Black dots indicate pseudo data randomly generated according to the total signal + background distribution, and assuming 100 pb^{-1} of data.

The efficiency of each selection cut is shown in Table 5 for a LQ sample with a mass of 400 GeV/c^2. Tables 6 and 7 show the number of events selected by each cut for $t\bar{t}$ and Z/γ+jets events, respectively, which are the dominant backgrounds in the final sample. A summary of the number of selected signal and background events expected in 100 pb^{-1} of data is reported in Table 8 [3]. The overall signal selection efficiency is around 35–65% for the LQ masses investigated.

5.1 Cut Optimization

Eight variables are studied to optimize the selection of events: the p_T's of the two leading electrons and two leading jets, the restriction on η for two leading electrons, the restriction on η for the two leading jets, the invariant mass M_{ee} of the two leading electrons, and S_T.

[2] A study at MC generator level has been performed to investigate the effect of jets generated by initial state radiation (ISR). The probability that an ISR jet has actually a larger p_T than one or both jets produced by the decay of the leptoquarks is 25% and 12% at LQ mass of 250 GeV, and 20% and 8% at LQ mass of 400 GeV. In such percentage of events, choosing the 2 leading jets as daughters of the leptoquarks is not correct. If the third leading jet were considered, the number of combinations to make 2 electron-jet pairs would become six. The same algorithm used above and based on the minimum difference ΔM_{ej} could be applied to select one among the six combinations. Our study showed that the improvement in finding the correct combination would be at the level of the percent with respect to the case where only the two leading jets are considered. Given the moderate improvement coming from using the third leading jet and considerations of simplicity and robustness of the analysis, the current version of the analysis uses only the two leading jets.

[3] No W+jets events survives the eejj selection with the current MC statistics available($\approx$130 pb^{-1}). Techniques for estimating the QCD multi-jet background using a factorized approach based on MC will be discussed in Section 6.3.

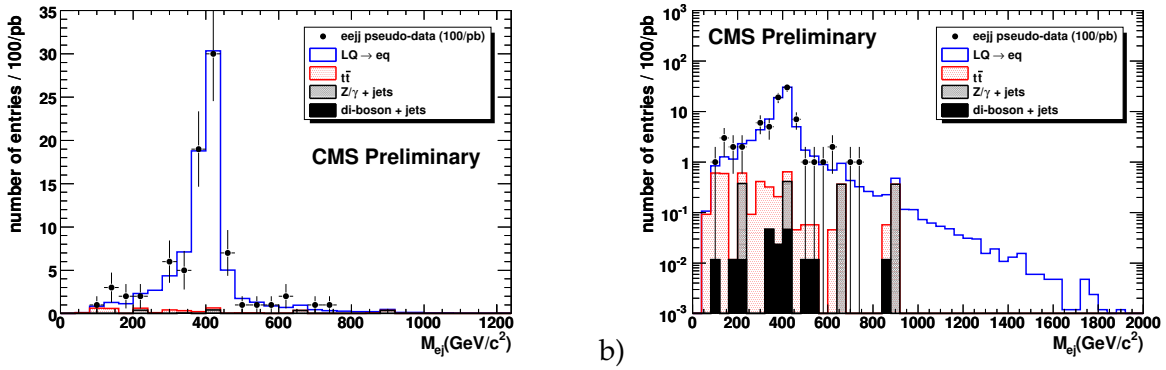


Figure 3: Distribution of the invariant mass, M_{ej}, of the electron-jet pairing with smaller ΔM_{ej}, for signal with $M_{LQ} = 400$ GeV/c^2 and the contributing backgrounds (with the exception of the QCD multi-jet background, see Section 6.3). The complete event selection has been applied. The histograms (two entries per event) are cumulative. Black dots indicate pseudo data randomly generated according to the total signal + background distribution, and assuming 100 pb^{-1} of data. The plot is shown in linear and log scale in a) and b) respectively.

Table 5: Sample of $M_{LQ} = 400\ GeV/c^2$ (FullSim): Sequence of selection cuts with number of events selected in 100 pb^{-1}, efficiency relative to the preceding cut and absolute efficiency. The reported uncertainties on the number of events and efficiencies are due to MC statistics. (*) A skim that requires 2 leptons (electrons and/or muons with loose ID requirements) and 2 jets with loose p_T thresholds is included in this cut.

Cut	N_{evt} passed for 100 pb^{-1}	ε_{rel}	ε_{abs}
None	75	1	1
2 ele $P_T > 30$ GeV (*)	66.75 $\pm$ 0.09	0.890 $\pm$ 0.001	0.890 $\pm$ 0.001
2 ele (ID+Iso) $P_T > 30$ GeV	50.09 $\pm$ 0.14	0.750 $\pm$ 0.002	0.668 $\pm$ 0.002
2 jets (Cleaned),P_T >50 GeV,$\|\eta\|$ <3	46.91 $\pm$ 0.14	0.937 $\pm$ 0.004	0.625 $\pm$ 0.002
M_{ee} >100 GeV	44.88 $\pm$ 0.15	0.957 $\pm$ 0.004	0.598 $\pm$ 0.002
S_T >620 GeV	38.98 $\pm$ 0.15	0.869 $\pm$ 0.004	0.520 $\pm$ 0.002

Table 6: Sample of $t\bar{t}$ + N jets: Sequence of selection cuts with number of events selected in 100 pb^{-1}, efficiency relative to the preceding cut and absolute efficiency. The reported uncertainties on the number of events and efficiencies are due to MC statistics. (*) A skim that requires 2 leptons (electrons and/or muons with loose ID requirements) and 2 jets with loose p_T thresholds is included in this cut.

Cut	N_{evt} passed for 100 pb^{-1}	ε_{rel}	ε_{abs}
None	41400	1	1
2 ele $P_T > 30$ GeV (*)	1754 $\pm$ 9	0.0424 $\pm$ 0.0002	0.0424 $\pm$ 0.0002
2 ele (ID+Iso) $P_T > 30$ GeV	155 $\pm$ 3	0.0883 $\pm$ 0.0016	(3.74 $\pm$ 0.06) 10^{-3}
2 jets (Cleaned),P_T >50 GeV,$\|\eta\|$ <3	77.2 $\pm$ 1.9	0.498 $\pm$ 0.015	(1.86 $\pm$ 0.04) 10^{-3}
M_{ee} >100 GeV	46.2 $\pm$ 1.4	0.60 $\pm$ 0.02	(1.12 $\pm$ 0.04) 10^{-3}
S_T >620 GeV	1.5 $\pm$ 0.3	0.032 $\pm$ 0.006	(3.5 $\pm$ 0.6) 10^{-5}

Table 7: Sample of Z/γ + N jets: Sequence of selection cuts with number of events selected in 100 pb^{-1}, efficiency relative to the preceding cut and absolute efficiency. The reported uncertainties on the number of events and efficiencies are due to MC statistics. (*) A skim that requires 2 leptons (electrons and/or muons with loose ID requirements) and 2 jets with loose p_T thresholds is included in this cut.

Cut	N_{evt} passed for 100 pb^{-1}	ε_{rel}	ε_{abs}
None	422000	1	1
2 ele $P_T > 30$ GeV (*)	2650 $\pm$ 30	(6.29 $\pm$ 0.07) 10^{-3}	(6.29 $\pm$ 0.07) 10^{-3}
2 ele (ID+Iso) $P_T > 30$ GeV	2020 $\pm$ 30	0.759 $\pm$ 0.013	(4.78 $\pm$ 0.06) 10^{-3}
2 jets (Cleaned),P_T >50 GeV,$\|\eta\|$ <3	328 $\pm$ 11	0.163 $\pm$ 0.006	(7.8 $\pm$ 0.3) 10^{-4}
M_{ee} >100 GeV	23 $\pm$ 3	0.070 $\pm$ 0.009	(5.4 $\pm$ 0.7) 10^{-5}
S_T >620 GeV	0.7 $\pm$ 0.5	0.03 $\pm$ 0.02	(1.7 $\pm$ 1.2) 10^{-6}

Table 8: Number of events expected from LQ signal and background samples after the analysis selection for 100 pb^{-1} of data. Signal selection efficiencies are also reported for different LQ masses; the fractional statistical uncertainty on the signal efficiencies is less than 1%. The cut value on the kinematic variable S_T depends on the LQ mass, and it is indicated in the second column. Data samples from FullSim Monte Carlo are used for all backgrounds and for LQ masses of 250 and 400 GeV/c^2. Signal samples marked by (*) are made with the FastSim Monte Carlo. Consistent results between FullSim and FastSim within the quoted systematics (see Section 7) are found.

M_{LQ} (GeV/c^2) in Signal Sample	S_T cut (GeV/c)	Events in Signal Sample	Selection Efficiency	Events in Background Samples		
				$t\bar{t}$ + N jets	Z/γ + N jets	VV + N jets
250	460	342 $\pm$ 2	34%	7.2 $\pm$ 0.6	2.6 $\pm$ 1.0	0.21 $\pm$ 0.05
300 (*)	520	163.4 $\pm$ 0.5	43%	3.9 $\pm$ 0.4	1.1 $\pm$ 0.6	0.15 $\pm$ 0.04
400	620	38.98 $\pm$ 0.15	52%	1.5 $\pm$ 0.3	0.7 $\pm$ 0.5	0.09 $\pm$ 0.03
500 (*)	740	11.56 $\pm$ 0.03	59%	0.69 $\pm$ 0.18	0.4 $\pm$ 0.4	0.05 $\pm$ 0.02
600 (*)	740	4.04 $\pm$ 0.01	66%	as above	as above	as above

A scan of the 8-dimensional parameter space is performed to find the approximate optimized region for the baseline selection cuts using all background samples, a LQ sample with mass 250 GeV/c^2 and assuming 100 pb^{-1} of data. The combination with the highest signal significance, defined as $S = N_S/\sqrt{N_S + N_B}$, is taken as the optimized set of cut values. The result of the optimization prefers as large a value as possible for the requirement on $|\eta^{ele}|$ (2.5, the acceptance of the tracker) and a value of $|\eta^{jets}| < 3$.

The optimal cut values for electron and jet p_T are consistently the lowest considered value in the scanned range (20 GeV/c). However, the p_T cut on electrons is increased to 30 GeV/c since the rate for jets faking electrons will have large uncertainties at startup. The p_T cut on jets is increased to 50 GeV/c in order to reduce the effect of uncertainties in the initial and final state radiation, and the uncertainties on calorimeter response at the start-up (not considered during the optimization procedure). These changes have a negligible effect on the signal significance. No major differences in the baseline cuts (*id est* excluding the S_T cut) have been found by optimizing at different LQ masses; hence, the cuts numbered 1, 2 and 3 in Section 5 are used at all masses.

The final optimization of the S_T cut, which is the most discriminating variable between signal and background, is performed by using a Bayesian approach (described in Section 8) to minimize the upper limit on leptoquark cross section in the absence of observed signal. This optimization indicates the S_T cut should increase with LQ mass. The optimized values of S_T cuts for different mass hypotheses are shown in Table 8 assuming 100 pb^{-1} of data. The optimal cut values are found to be independent of an addition of systematic uncertainties in number of signal and background events.

6 Data-driven techniques for background estimate

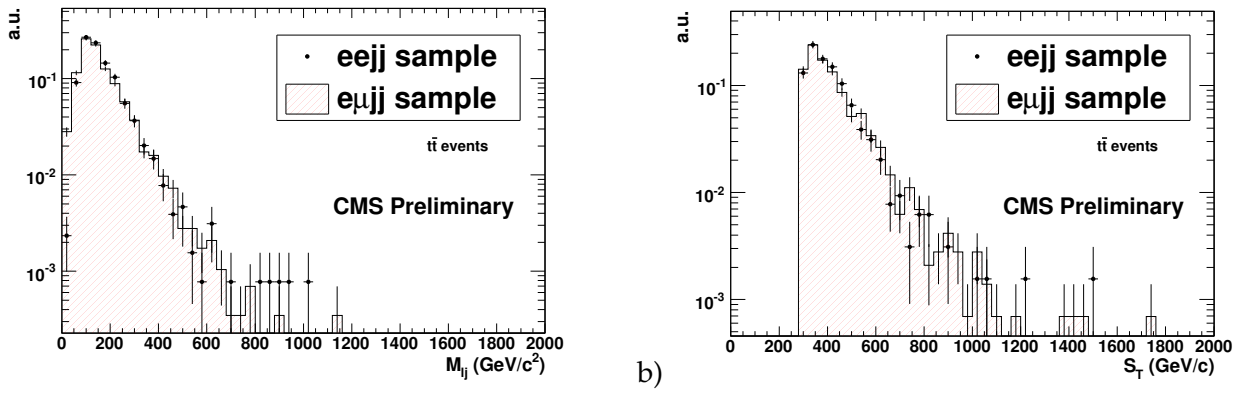
After the event selection the dominant number of SM background events originates from $t\bar{t}$ and Z/γ+jets processes, as summarized in Table 8. Sections 6.1 and 6.2 describe data-driven techniques that use control samples to estimate the absolute normalization for these two backgrounds, and the shapes of the distributions of the selection variables for the $t\bar{t}$ background. Techniques for estimating the QCD multi-jet background using a factorized approach based on MC are discussed in Section 6.3.

6.1 $t\bar{t}$ background control sample

Very strong constraints from rare processes exist on leptoquarks with couplings to both electrons and muons. Therefore a useful control sample can be obtained by requiring at least one electron and one muon, instead of at least 2 electrons, in the final state, in addition to the two jets. For $t\bar{t}$ events, the distributions of the selection variables for the eμjj and the eejj samples are expected to be very similar in shape, since the kinematics of the process does not depend on the nature of the lepton. Figure 4 shows good agreement between the shape of M_{lj} and S_T distributions with the current MC statistics available ($\approx$2 fb^{-1}). Similar agreement is found for the other reconstructed kinematic variables used in the selection.

The eμjj sample is dominated by $t\bar{t}$ events, with a small contamination estimated from MC of less than 5% (for an S_T cut of 460 GeV/c), mainly from di-boson events. For $t\bar{t}$ events, the eμjj sample with 100 pb^{-1} of data is expected to have about 17, 11, and 5 events, respectively, for S_T cuts of 460, 520, and 620 GeV/c (used for LQ masses of 250, 300 and 400 GeV/c^2).

Although the W decays with equal branching fraction to electrons and muons, the trigger filter, the offline selection, and the different reconstruction efficiencies, acceptances, and p_T res-



a) b)

Figure 4: Distributions of the lepton-jet invariant mass (a) and S_T (b) for the eejj and the eμjj samples, for $t\bar{t}$ events. All baseline selection criteria described in Section 5 are applied with the S_T cut set to 300 GeV/c.

olutions bias the number of selected eejj and eμjj events. In this analysis, we correct for the different reconstruction efficiencies (which is the dominant effect for the current selection) by binning the events in the p_T of the leading muon and using

$$N^{est.}_{\mathrm{eejj}} = \frac{1}{2}\sum_{p_T^\mu} N_{\mathrm{e}\mu\mathrm{jj}}(p_T^\mu)\cdot R(p_T^\mu) \quad , \tag{1}$$

where $N_{\mathrm{e}\mu\mathrm{jj}}(p_T^\mu)$ is number of events in the bin, and $R(p_T)$ is the ratio between electron and muon reconstruction efficiencies as a function of lepton p_T. The ratio $R(p_T)$ is obtained using a MC FullSim sample of Z/γ+jets events with an equivalent integrated luminosity of 275 pb^{-1}. The value of $R(p_T)$ is between 0.85–0.95 for $30 < p_T < 500$ GeV/c. Once real data becomes available this ratio could be obtained with *tag&probe* method using $Z \to ee$ and $Z \to \mu\mu$ events.

6.2 Z/γ+jets background control sample

A control sample that can be used to estimate the Z/γ+jets background can be obtained by using the full selection criteria but reversing the M_{ee} cut to select events consistent with a Z boson ($80\ GeV/c^2 < M_{ee} < 100\ GeV/c^2$). This control sample (eejjAtZ sample) is an almost pure sample of Z/γ+jets events (less than 5% background contamination, dominated by $t\bar{t}$ events, for an S_T cut of 460 GeV/c). The signal contamination is less than 10% for all S_T cuts considered. The eejjAtZ sample with 100 pb^{-1} of data is expected to have about 30, 21, and 11 events, respectively for an S_T cut of 460, 520 and 620 GeV/c (used for LQ masses of 250, 300 and 400 GeV/c^2).

The background in the final sample is estimated by rescaling the control sample using a hybrid method which combines the eejjAtZ data with MC information. The number of Z/γ+jets events in the eejj signal sample ($M_{ee} > 100$ GeV/c^2) can be estimated by

$$N^{Z}_{eejj} = N_{\mathrm{eejjAtZ}} \cdot R_{\mathrm{OffZ/AtZ}} \quad , \tag{2}$$

where N_{eejjAtZ} is the number of events in the eejjAtZ control sample, and $R_{\mathrm{OffZ/AtZ}}$ is the ratio between the number of Z/γ+jets events with $M_{ee} > 100$ GeV/c^2 (OffZ events) and

80 GeV/c^2 $<$ M$_{ee}$ $<$ 100 GeV/c^2 (AtZ events) that have passed all the other selection criteria. In this analysis the value of $R_{\mathrm{OffZ/AtZ}}$ is determined directly from MC. Figure 5 shows the value of the inverse of $R_{\mathrm{OffZ/AtZ}}$ as a function of different S_T cuts, obtained from the FastSim MC sample of Z/γ+jets events.

For inclusive $Z/\gamma \rightarrow ee$ production, the MC is expected to predict well the value of the ratio $R_{OffZ/AtZ}$. The level of agreement will be checked with data using a control sample with two electrons and no requirements on jets (which is expected to be dominated by Z/γ events up to values of M_{ee} of at least one TeV/c^2 [12]). The MC description of the jet component of the $Z/\gamma \rightarrow ee + jj$ process is expected to have larger uncertainty than the electron one. The data–MC comparison for jet reconstructed quantities will be performed using the eejjAtZ control sample to evaluate the level of agreement and, if needed, to correct the MC for the differences observed with respect to data. The MC predicts that the jet component of the Z/γ+jets process is similar between the signal (eejjOffZ events) and the control (eejjAtZ events) region. This has been observed by comparing kinematic jet variables like the scalar p_T sum of the two leading jets and the velocity of the jet–jet system between eejjAtZ and eejjOffZ MC events. The latter result indicates that, once a good agreement between data and MC is reached in the control region for both electron and jet quantities, the ratio $R_{OffZ/AtZ}$ (calculated from MC) could be used to reasonably predict the expected number of Z/γ+jets events in the signal region.

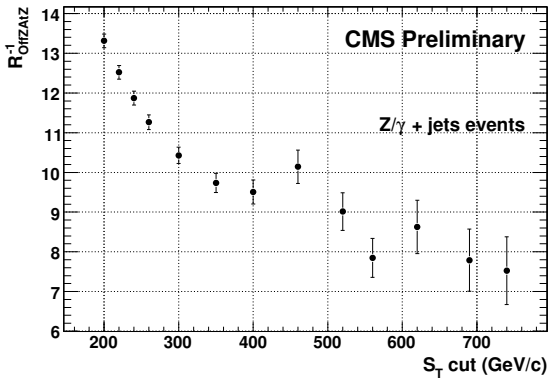


Figure 5: The value of the inverse of $R_{\mathrm{OffZ/AtZ}}$ as a function of different S_T cuts, obtained from the FastSim MC sample of Z/γ+jets events.

6.3 QCD multi-jet Background

We consider a QCD multi-jet sample with two fake electrons and two jets (ffjj sample) that pass all the kinematic selection criteria (see Section 5), and rescale it to estimate the QCD multi-jet contamination in the final sample (eejj) as

$$N^{QCD}_{\mathrm{eejj}} = N^{QCD}_{\mathrm{ffjj}} \cdot P(e|f)^2 \quad , \tag{3}$$

where "f" is a reconstructed electron with loose ID requirements, "e" is a reconstructed electron passing the tight ID and isolation criteria (see Table 3), N^{QCD}_{eejj} (N^{QCD}_{ffjj}) is the number of QCD multi-jet events in the eejj (ffjj) sample (control sample), and $P(e|f)$ is the probability of a fake electron (f) to pass the tight ID and isolation requirements (e).

Table 9 summarizes the number of selected events in the ffjj sample for different QCD multi-jet samples. An S_T cut of 460 GeV/c (the optimized cut for a LQ mass of 250 GeV/c^2) is used in the ffjj selection. The total number of ffjj events in 100 pb^{-1} of data is $N^{QCD}_{\mathrm{ffjj}} = 6130 \pm 47$.

Table 9: The absolute number of events passing the ffjj selection, the number of selected events rescaled to 100 pb^{-1} of data, the cross section at LO and the equivalent integrated luminosity of the generated sample, for each H_T bin of the QCD multi-jet sample.

H_T bin (GeV/c)	N_{evt}^{sel}	N_{evt}^{sel} for 100 pb^{-1}	σ_{LO} (pb)	Equivalent Luminosity (pb^{-1})
$[250, 500]$	4	40 ± 20	$400 \cdot 10^3$	9.25
$[500, 1000]$	12431	4570 ± 40	$14 \cdot 10^3$	272
> 1000	26786	1515 ± 10	370	$1.77 \cdot 10^3$

The probability $P(e|f)$ is estimated from the rate of "f" to "e" in a ffjj sample obtained from the QCD multi-jet samples by applying the selection cuts described in Section 5 with the S_T cut set to 460 GeV/c. In the barrel region ($|\eta| < 1.5$), the probability $P(e|f)$ is almost flat in η and $\approx 2.5 \cdot 10^{-3}$, while in the endcap region it is increasing with η and has an average value of $\approx 1.4 \cdot 10^{-2}$. The overall average value in barrel and endcaps is $P(e|f) = 5.3 \cdot 10^{-3}$. A closure test performed without isolation cuts suggests that this method might underestimate the QCD multi-jet background by approximately 40%.

With the cut $S_T > 460$ GeV/c, the number of QCD multi-jet events in the eejj sample is estimated as $0.17 < N_{\text{eejj}}^{QCD} < 0.26$ events by rescaling the ffjj sample with Equation 3 and taking into account the underestimation suggested by the closure test. Therefore, QCD multi-jet background is expected to be small compared to the $t\bar{t}$ and Z/γ+jet backgrounds for the same S_T cut. It has been verified that this statement remains true even when a harder S_T cut is applied. The contamination of QCD multi-jet background in the eμjj sample is expected to be even smaller than eejj sample, since the probability for jets to fake muons is smaller than for electrons.

7 Systematic Uncertainties

The main sources of systematic uncertainties for this analysis are discussed below.

1. Uncertainty in the jet and electron energy scale

   To quantify the effect of the uncertainty on the reconstructed energy of electrons and jets, the analysis is repeated rescaling the electron (jet) energies with a factor of $\pm$2% ($\pm$10%).

   For selections optimized for different LQ masses, a change in the jet energy by +10% (-10%) leads to a change of +1 to +10% (-3 to -15%) in the signal efficiency and +33 to +52% (-15 to -28%) in the number of background events, while a +2% (-2%) variation in the electron energy scale yields a +1 to +2% (-1 to -4%) change in signal efficiency and a +1 to +9% (-9 to -20%) change in the number of background events.

   Jet and electron energy scale uncertainties are not included for $t\bar{t}$ and Z/γ+jets backgrounds, since they are determined using the data driven methods discussed in Section 6.

2. Uncertainty in the integrated luminosity of the data

   This uncertainty is estimated to be 10% for the first several months of LHC running [14].

3. Statistical uncertainty on the MC data

   The uncertainty on the final number of selected events for the FullSim sample with a leptoquark mass of 400 GeV/c^2 is approximately 0.4%. The number of events produced for

some of the background samples, however, correspond to a much smaller equivalent integrated luminosity. The statistical uncertainty on the number of MC events is summarized for signal and background samples in Table 8 of Section 5.

4. Uncertainties on FastSim selection efficiencies with respect to FullSim

   The signal samples produced with FastSim show a slightly higher selection efficiency than FullSim, mostly due to the higher reconstruction efficiency of the electrons in the FastSim. This leads to a higher final selection efficiency in FastSim compared to FullSim by approximately 5% for leptoquark samples with a mass of 250 and 400 GeV/c^2. FullSim samples at higher mass are not available to perform the comparison. A conservative uncertainty of 10% on the selection efficiency for FastSim samples is used in the whole mass range investigated.

5. Uncertainty from the data-driven background estimates

   Estimates of the background by data-driven techniques are affected by the statistical uncertainties on the size of the control samples (that scales as the square root of the number of events). The number of events expected in each control sample varies with the S_T cut used, as described in Section 6. These uncertainties are calculated for an integrated luminosity of 100 pb^{-1}.

   The uncertainties on the ratios $R(p_T)$ (see Equation 1) and $R_{\mathrm{OffZ/AtZ}}$ (see Equation 2) are expected to be small compared to the statistical uncertainty associated with the size of the control samples.

6. Theoretical uncertainties

   The Parton Density Function (PDF) uncertainties are expected to give a major contribution to the theoretical uncertainties for this analysis. The effect of this systematic uncertainty has been estimated using the method developed by the CTEQ collaboration [15], which is implemented in the CMS software as an event re-weighting technique. For this analysis we have used the PDF set CTEQ6.1.

   The uncertainty on the leptoquark cross section ranges from 5% (for LQ mass of 250 GeV/c^2) to 17% (for LQ mass of 600 GeV/c^2). The uncertainty on the signal efficiency is estimated to be less than 1% in the whole mass range investigated. A 10% uncertainty on the number of selected background events is used for all the LQ mass hypotheses.

8 CMS Discovery and Exclusion Potential

In order to estimate the potential of the CMS detector to discover the first generation leptoquarks in the electron channel, a counting experiment approach is used. The number of signal and background events expected for an integrated luminosity of 100 pb^{-1} listed in Table 8 are used to produce the plots in this section. The effects of the systematic uncertainties described in Section 7 are taken into account by summing them in quadrature, and combining them with the likelihood when calculating discovery and exclusion potential.

To quantify the significance of the leptoquark signal, the S_{cP} significance estimator [16] is used. For a Poisson distribution with mean b, the probability to observe $n = s + b$ events or greater is:

$$P = p(n \geq s + b|b) = \sum_{n=s+b}^{+\infty} \frac{b^n}{n!} e^{-b}, \tag{4}$$

where s and b are the expected numbers of signal and background events, respectively. This probability can be converted into an equivalent number of standard deviations using the one-sided Gaussian probability

$$P = \frac{1}{\sqrt{2\pi}} \int_{S_{cP}}^{+\infty} e^{-\frac{x^2}{2}} \mathrm{d}x, \tag{5}$$

which is the S_{cP} significance. Figure 6 shows the minimum β for a 5σ discovery as a function of leptoquark mass for 100 pb^{-1} of integrated luminosity. The *look elsewhere* effect is not taken into account in this analysis.

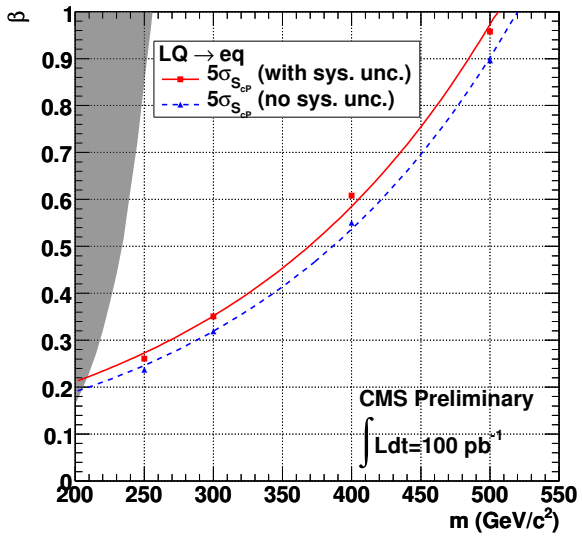


Figure 6: Minimum β for a 5σ discovery as a function of leptoquark mass for 100 pb^{-1} of integrated luminosity. The solid red line includes the systematic uncertainties described in Section 7 (uncertainties on the estimated number of background events come from the data-driven methods of Section 6). The shaded region is excluded by the current Tevatron limits. The *look elsewhere* effect is not taken into account in the results.

For setting upper limits in the absence of the leptoquark signal, the Bayesian approach [17] is used. Figure 7 shows the minimum β for a 95% C.L. exclusion of the leptoquark hypothesis as a function of leptoquark mass for 100 pb^{-1} of integrated luminosity. Figure 8 shows the expected 95% C.L. upper limit on the product of β^2 and the leptoquark pair production cross section as a function of leptoquark mass for 100 pb^{-1} of integrated luminosity.

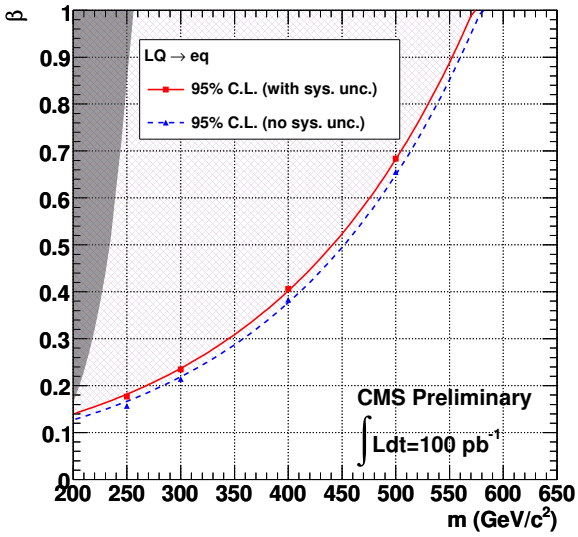


Figure 7: Minimum β for a 95% C.L. exclusion of the leptoquark hypothesis as a function of leptoquark mass for 100 pb^{-1} of integrated luminosity. The solid red line includes the systematic uncertainties described in Section 7 (uncertainties on the estimated number of background events come from the data-driven methods of Section 6). Gray shaded region is excluded by the current Tevatron limits.

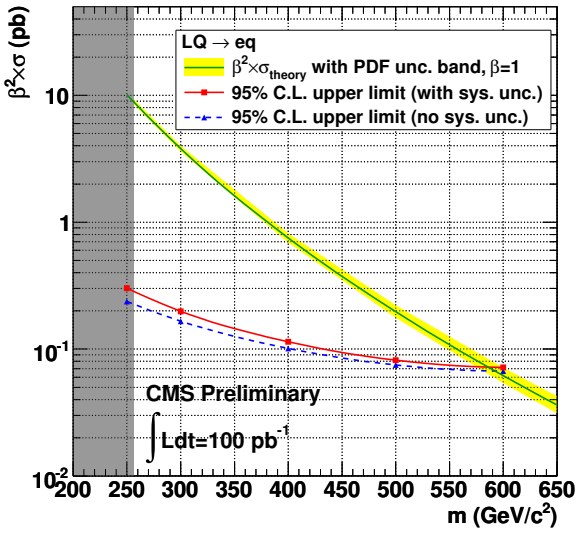


Figure 8: Expected 95% C.L. upper limit on the product of β^2 and the leptoquark pair production cross section as a function of leptoquark mass for 100 pb^{-1} of integrated luminosity. The solid red line includes the systematic uncertainties described in Section 7 (uncertainties on the estimated number of background events come from the data-driven methods of Section 6). The shaded region is excluded by the current Tevatron limits, assuming $\beta = 1$.

9 Conclusion

The search for pair production of first generation leptoquarks that decay to an electron and a jet is studied using a MC simulation. The analysis strategy assumes an integrated luminosity of 100 pb^{-1} and pp collisions at $\sqrt{s} = 10$ TeV. Standard CMS techniques are used for electron and jet identification. An optimized cut-based event selection is applied, and the main expected background sources are studied, $t\bar{t}$ and Z/γ+jets providing a significant contribution after event selection. Data-driven techniques to understand the characteristics of these contributions are developed.

The discovery and exclusion potential in the channel with two electrons plus two jets is determined using two statistical estimators (S_{cP} for discovery and a Bayesian approach for exclusion) suited for a counting experiment in the Poisson regime. The effect of the main systematic uncertainties is studied and taken into account in the final results. This study shows that the analysis is sensitive to the production of leptoquarks with mass above the Tevatron exclusion limit of 256 GeV/c^2 for $\beta = 1$. With an integrated luminosity of 100 pb^{-1}, discovery should be possible up to a leptoquark mass of about 505, 370, and 270 GeV/c^2 assuming, respectively, $\beta = 1$, 0.5, and 0.3. In absence of evidence of a signal, the existence of a scalar leptoquark with mass lower than about 570, 445, and 350 GeV/c^2, assuming respectively $\beta = 1$, 0.5, and 0.3, can be excluded with 100 pb^{-1} of data.

10 Acknowledgements

We would like to thank Albert De Roeck, Greg Landsberg, Sam Harper, Andrey Korytov, Michael Kramer, Danilo Piparo, Gregory Schott, and the CMS group working on the second generation leptoquark analysis for the useful discussions and suggestions.

We thank the technical and administrative staff at CERN and the other CMS Institutions.

References

[1] D. E. Acosta and S. K. Blessing, "Leptoquark searches at HERA and the Tevatron," *Ann. Rev. Nucl. Part. Sci.* **49** (1999) 389–434.

[2] S. Davidson, D. C. Bailey, and B. A. Campbell, "Model independent constraints on leptoquarks from rare processes," *Z. Phys.* **C61** (1994) 613–644, arXiv:hep-ph/9309310.

[3] **H1** Collaboration, A. Aktas et al., "Search for leptoquark bosons in e p collisions at HERA," *Phys. Lett.* **B629** (2005) 9–19, arXiv:hep-ex/0506044.

[4] A. Belyaev, C. Leroy, R. Mehdiyev, and A. Pukhov, "Leptoquark single and pair production at LHC with CalcHEP/CompHEP in the complete model," *JHEP* **09** (2005) 005, arXiv:hep-ph/0502067.

[5] **D0** Collaboration, V. M. Abazov et al., "Search for first-generation scalar leptoquarks in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96-TeV," *Phys. Rev.* **D71** (2005) 071104, arXiv:hep-ex/0412029.

[6] **CDF** Collaboration, D. E. Acosta et al., "Search for first-generation scalar leptoquarks in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV," *Phys. Rev.* **D72** (2005) 051107, arXiv:hep-ex/0506074.

[7] J. Alwall et al., "MadGraph/MadEvent v4: The New Web Generation," *JHEP* **09** (2007) 028, arXiv:0706.2334.

[8] Krämer, M. and Plehn, T. and Spira, M. and Zerwas, P. M., "Pair production of scalar leptoquarks at the CERN LHC," *Phys. Rev. D* **71** (Mar, 2005) 057503.

[9] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, "Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC," *JHEP* **09** (2008) 127, arXiv:0804.2800.

[10] S. Frixione and M. L. Mangano, "How accurately can we measure the W cross section?," *JHEP* **05** (2004) 056, arXiv:hep-ph/0405130.

[11] **CMS** Collaboration, S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, "Electron reconstruction in CMS," Technical Report CMS-NOTE-2006-040. CERN-CMS-NOTE-2006-040, CERN, Geneva, Feb, 2006.

[12] **CMS** Collaboration, "Search for high mass resonance production decaying into an electron pair in the CMS experiment," *CMS PAS* **EXO-08-001** (2009).

[13] **D0** Collaboration, V. M. Abazov et al., "Search for first-generation scalar and vector leptoquarks," *Phys. Rev.* **D64** (2001) 092004, arXiv:hep-ex/0105072.

[14] **CMS** Collaboration, G. L. Bayatian et al., "CMS technical design report, volume II: Physics performance, Appendix B.2: Experimental uncertainties," *J. Phys.* **G34** (2007) 995–1579.

[15] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "Uncertainties of predictions from parton distributions. I: Experimental errors," *Eur. Phys. J.* **C28** (2003) 455–473, arXiv:hep-ph/0211080.

[16] S. I. Bityukov, S. E. Erofeeva, N. V. Krasnikov, and A. N. Nikitenko, "Program for evaluation of significance, confidence intervals and limits by direct calculation of probabilities,". Prepared for PHYSTATO5: Statistical Problems in Particle Physics, Astrophysics and Cosmology, Oxford, England, United Kingdom, 12-15 Sep 2005.

[17] **Particle Data Group** Collaboration, C. Amsler et al., "Review of particle physics, Section 32.3.1: The Bayesian approach," *Phys. Lett.* **B667** (2008) 1.