Data driven background study for new physics searches with same sign dileptons at $\sqrt{s} = 10$ TeV

D. Barge, C. Campagnari, P. Kalavase, D. Kovalskyi, V. Krutelyov, J. Ribnik

University of California, Santa Barbara

W. Andrews, D. Evans, F. Golf, J. Mülmenstädt, S. Padhi, Y. Tu, F. Würthwein, A. Yagil

University of California, San Diego

L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, O. Gutsche

Fermi National Accelerator Laboratory, Batavia, Illinois

Abstract

We discuss data driven methods to estimate the number of same sign Standard Model dilepton events in searches for new physics characterized by large $\not\!\!E_T$ and significant hadronic activity. For these searches, the dominant background is from $t\bar{t}$ events. The study provides a new method to estimate the charge mis-measurement of leptons. The residual flavor enriched background is estimated using the lepton fake rate method. We show sensitivity to several SUSY benchmark points using 100 pb^{-1} of integrated luminosity.

1 Introduction

In this note we describe data driven methods to estimate the backgrounds in searches for new physics in events with two high P_T, isolated, same sign leptons, large $\not\!\!E_T$, and significant hadronic activity. This generic signature is sensitive to several new physics scenarios such as SUSY. For the purpose of this note we restrict ourselves to the ee, $e\mu$, and $\mu\mu$ final states, *i.e.*, we do not consider τ's, except in the case that the τ decays leptonically.

For a reasonable event selection, as shown in Section 4, the dominant background is from $t\bar{t}$ events. We categorize the total background into contributions from charge mis-measurement of leptons and events with fake leptons. The former is estimated using charge fake probability, as discussed in Section 6. The latter, which also includes heavy flavor decays is estimated using the lepton fake rate method described in Section 7.

This note is organized as follows. In Section 2 we list the Monte Carlo data samples, as well as the software tags used in the analysis. In Section 3 we describe the same sign dilepton event selection used in this study. The expected event yields for the dominant Standard Model processes as well as a few SUSY benchmark points are given in Section 4. A brief introduction to classifying the dominant background based on their origin is given in Section 5. In Section 6 we discuss the data driven procedure to estimate the charge mis-measurement followed by lepton fake rate method in Section 7. We apply the above mentioned data driven methods to the Standard Model cocktail in Section 8 and study the effect of applying the background prediction procedure in the presence of new physics. In Section 9 we briefly discuss the associated systematics and finally, in Section 10 we summarize the results.

2 Data Samples

This study is based on the 2_2_X re-reco full simulation data samples listed in Table 1. The Standard Model (SM) data sets have been normalized to the cross-sections compiled by the top group [1]; for the SUSY data sets we used the cross-sections from the Summer 2008 production page [2].

/TTJets-madgraph/Fall08_IDEAL_V11_redigi_v10/GEN-SIM-RECO
/WJets-madgraph/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO
/ZJets-madgraph/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO
/WW/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO
/WZ_incl/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO
/ZZ/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO
/SingleTop_sChannel/Summer08_IDEAL_V11_redigi_v3/GEN-SIM-RECO
/SingleTop_tWChannel/Summer08_IDEAL_V11_redigi_v3/GEN-SIM-RECO
/SingleTop_tChannel/Summer08_IDEAL_V11_redigi_v3/GEN-SIM-RECO
/SUSY_LM0-sftsht/Summer08_IDEAL_V11_v1/GEN-SIM-RECO
/SUSY_LM*-sftsht/Summer08_IDEAL_V11_redigi_v1/GEN-SIM-RECO

Table 1: The data sets used in this study.

Monte Carlo events were analyzed with CMSSW_2_2_10 with the additional tags listed in Table 2.

3 Event Selection

The event selection used is not optimized for any specific SUSY scenario. It is based on small modifications to the dilepton event selections that we used in recently approved WW[3] and $t\bar{t}$ [4] cross-section analyses. A quick summary of the event selection is:

- The event is required to pass the single e or μ triggers.
- Two isolated, same sign leptons (ee, $e\mu$, and $\mu\mu$).
- Leptons must have $P_T > 10$ GeV, $|\eta| < 2.4$ and at least one of them must have $P_T > 20$ GeV.
- We veto the candidate lepton, if an extra lepton in the event, pairs with the candidate lepton to form a Z within the mass range between $76 < m_{\ell\ell}$ (GeV) < 106. This requirement is designed to reject WZ events.
- At least three L2L3 corrected caloJets with $P_T > 30$ GeV and $|\eta| < 2.4$.

V01-08-04 CondFormats/JetMETObjects
V00-06-02-09 DataFormats/METReco
V07-02-12-03 DataFormats/MuonReco
V01-08-02-01 JetMETCorrections/Algorithms
V01-08-15 JetMETCorrections/Configuration
V03-02-06 JetMETCorrections/JetPlusTrack
V02-09-02 JetMETCorrections/Modules
VB04-00-02-04 JetMETCorrections/Type1MET
V01-04-03 RecoJets/JetAssociationAlgorithms
V00-04-02-17 RecoMET/Configuration
V02-05-00-21 RecoMET/METAlgorithms
V02-08-02-17 RecoMET/METProducers
V03-26-04 DataFormats/PatCandidate
V05-05-09 PhysicsTools/PatAlgos
V03-06-03 PhysicsTools/PatUtils
V03-01-16 PhysicsTools/PFCandProducer
V09-30-03 PhysicsTools/HepMCCandAlgos
V05-13-02 DataFormats/HepMCCandidate

Table 2: Additional software tags used in this study.

- The scalar sum of the P_T of all jets passing the requirements above should be $>$ 200 GeV.
- We require $\not{E}_T > 80$ GeV. Track Corrected MET (tcMET) [5] is used as a measure for $\not{E}_T$.

The details of the lepton and trigger selections are given below.

3.1 Electron Selection

- The electron ID is the "e-gamma category based tight", with small modifications to account for changes between CMSSW 1_6_X and 2_2_X; see Reference [3] for details.
- No muon candidate within $\Delta R < 0.1$.
- $|d_0| < 200\ \mu m$ (corrected for beamspot).
- Iso $<$ 0.1, where Iso=Sum/Max(20 GeV, P_T), and Sum = tkIso + hcalIso + Max(0 GeV, ecalIso - 2GeV). All isolation sums are the standard sums used in release 2_2_X from the egamma group (cone of 0.4 for ecal, jurassic, rec-hit based; cone of 0.3 for tracker, and cone of 0.4 for hcal).
- Conversion rejection [6] using tracks within cone of 0.3 of the candidate electron:
  - $|\Delta \cot\theta| < 0.02$; the difference between cotangent azimuthal angles of tracks parallel to each other.
  - $|d_{2d}| < 0.02$ cm; the two dimensional distance between points within nearest tracks.
  - The charge of the associated GSF and CTF tracks must be consistent. If the CTF track is not reconstructed, the electron is kept. This is discussed further in Section 3.4.

3.2 Muon Selection

- Must be a global muon **and** a tracker muon [7].
- GlobalMuonPromptTight (global χ^2/ndof$<$10) [8].
- At least 11 valid hits for the silicon track [8].
- $|d_0| < 200\ \mu m$ (from silicon track, corrected for beamspot).
- Minimum ionizing: EcalVetoEnergy < 4 GeV and HcalVetoEnergy < 6 GeV [9].
- Iso $<$ 0.1, where Iso=Sum/Max(20 GeV, P_T), and Sum = tkIso + hcalIso + ecalIso. All isolation sums are the standard sums stored in the muon object in release 2_2_X, and are calculated in a cone of 0.3.

3.3 Trigger Selection

We use inclusive lepton triggers with no isolation, $i.e.$, the logical OR of `HLT_Ele15_SW_L1R` and `HLT_Mu9`. The combined trigger efficiency is $\sim$ 99% for dilepton events that pass the event selection. These triggers are expected to be present in the data taking trigger table.

3.4 Selection due to charge mis-reconstruction

One of the main backgrounds to the same sign isolated dilepton signature consists of SM opposite sign dilepton events, where the charge of one of the leptons is misidentified. Monte Carlo studies show that the muon charge misidentification rate is negligible, up to very high momenta. On the other hand, because of curvature changes due to bremsstrahlung in the tracker, electrons have a significant probability of being reconstructed with the wrong sign.

In order to reduce the contribution of charge mis-reconstruction, we have studied the charge mis-identification rates in a dedicated "electron gun" sample. The charge mis-identification rate is defined as the ratio of electrons with wrong reconstructed charge compared to the true charge over all reconstructed and truth matched electron candidates. The details of the study can be found elsewhere [10].

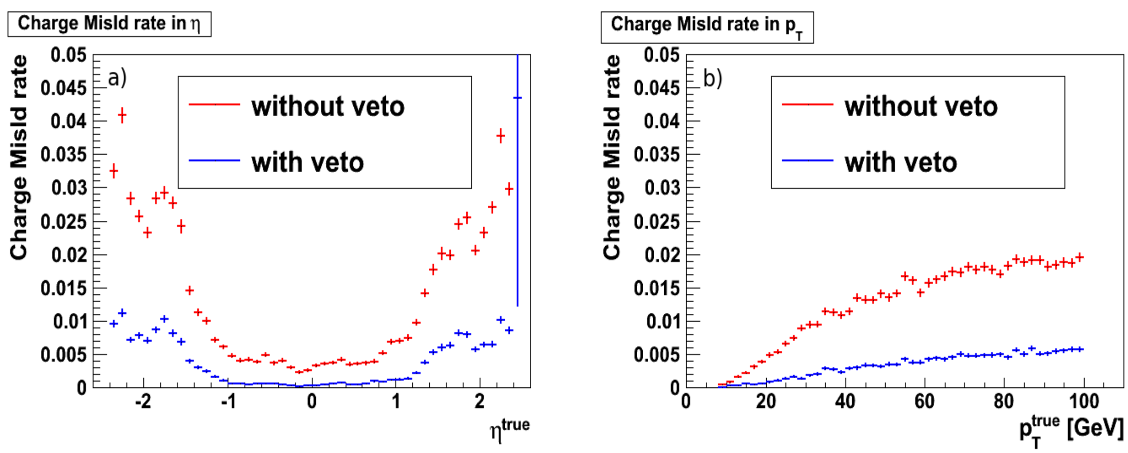

Figure 1: Charge mis-identification rate, as a function of a) η and b) p_T of the generated electron using "electron gun" sample . The solid (blue) distribution shows the rate after the veto.

Electrons are selected with the identification and isolation requirements mentioned earlier. The standard way to define the charge of the electron is to use the charge of the corresponding Gaussian Sum Filter (GSF) track. We have found that the probability of misidentifying the charge can be reduced by a factor of ~ 3.9 with a small $\sim 2.1\%$ loss in efficiency by vetoing electrons where the associated GeneralTrack (CTF track), if found, has a different charge. This is illustrated in Figure 1. The rate after the veto is found to vary between 0.04% to 1% with increase in pseudorapidity. This selection is applied as part of the standard selection for the rest of this document.

The association of CTF tracks to electron objects is performed by counting the number of shared valid hits between the CTF and GSF tracks in the pixels and TIB/TID. The associated CTF track is defined as the CTF track with the highest fraction of shared hits with the GSF track, with the further requirement that the fraction of shared hits be $>$ 50%. The fraction of shared hits is defined as the number of shared hits divided by min(N_{CTF}, N_{GSF}), where N_{CTF} and N_{GSF} are the number of pixel + TIB/TID valid hits for the CTF and GSF tracks respectively.

4 Event Yields

Applying the event selections described in Section 3 to the Monte Carlo data sets described in Section 2, results in the expected event yields in 100 pb^{-1} listed in Table 3 and 4 for SM and SUSY benchmark points, respectively.

This event selection shows sensitivity to several SUSY points. The dominant SM contribution is from $t\bar{t}$ decays.

Same Sign leptons	Total SM	$t\bar{t}$	tW	WZ	ZZ	WW	DY	Wjets
ee	0.45±0.14	0.44±0.14	0.00±0.00	0.00±0.00	0.01±0.01	0.00±0.00	0.00±0.00	0.00±0.00
$\mu\mu$	0.17±0.08	0.13±0.08	0.00±0.00	0.03±0.02	0.01±0.01	0.00±0.00	0.00±0.00	0.00±0.00
$e\mu$	0.48±0.14	0.39±0.13	0.04±0.03	0.04±0.03	0.01±0.01	0.00±0.00	0.00±0.00	0.00±0.00
total	1.10±0.21	0.96±0.21	0.04±0.03	0.07±0.03	0.03±0.01	0.00±0.00	0.00±0.00	0.00±0.00

Table 3: Expected number of SM events passing the event selection in 100 pb^{-1} of integrated luminosity. Uncertainties are from MC statistics.

Same Sign leptons	LM0	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8	LM9
ee	9.93±0.73	1.95±0.17	0.21±0.02	1.45±0.11	0.49±0.05	0.17±0.01	0.45±0.02	0.22±0.03	0.72±0.03	0.50±0.05
$\mu\mu$	11.99±0.81	2.42±0.19	0.30±0.02	1.91±0.12	0.63±0.06	0.18±0.01	0.45±0.02	0.27±0.03	0.88±0.03	0.64±0.06
$e\mu$	22.52±1.11	4.55±0.26	0.48±0.03	3.09±0.15	1.18±0.08	0.34±0.02	0.86±0.03	0.42±0.04	1.62±0.05	1.26±0.09
total	44.44±1.55	8.92±0.37	0.99±0.04	6.45±0.22	2.30±0.12	0.69±0.03	1.76±0.04	0.91±0.06	3.22±0.07	2.40±0.12

Table 4: Expected number of SUSY benchmark simulated events passing the event selection in 100 pb^{-1} of integrated luminosity. Uncertainties are from MC statistics.

5 Discussion of backgrounds

As shown in Section 4, the SM background to this study is dominated by $t\bar{t}$ events. The contribution from single-top, diboson production (WZ, ZZ), is small. The diboson backgrounds will be estimated from Monte Carlo. The Monte Carlo acceptances for these processes will be corrected for differences between data and Monte Carlo lepton identification and trigger efficiencies, as determined using the tag-and-probe method.

We study the same sign dileptons for the $t\bar{t}$ background based on their origin. The dominant source of dileptons in $t\bar{t}$ events are produced via, $t \to Wb$; where $W \to \ell\nu$. We classify them as follows, based on truth matched to their "parents":

- Type-I: both leptons originate from real W (including $W \to \tau \to \ell$) bosons, one with mis-reconstructed charge.
- Type-II a): one of the leptons is from a real W and the other originates from heavy flavor sources (b, c).
- Type-II b): one of the leptons is from a W and the other is a fake lepton.
- Type-III: both leptons are fakes or from heavy flavor sources.

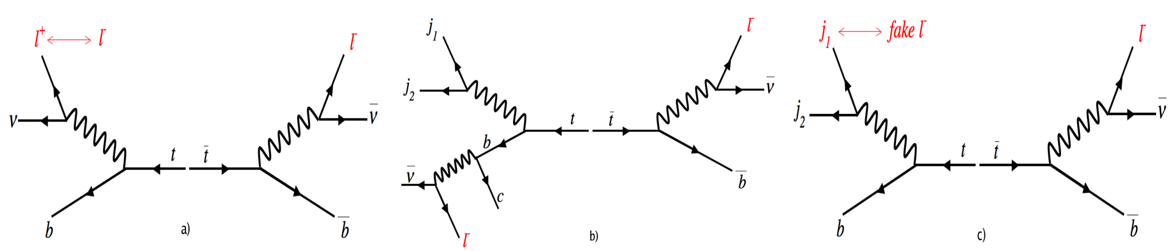

Figure 2: Classification of same sign dileptons from $t\bar{t}$ decays. a) lepton from W with mis-reconstructed charge; b) one of the lepton is from W, the other originating from heavy flavor sources; c) one of the lepton is from W, and the other is a fake lepton.

Figure 2 illustrates contribution from different types in $t\bar{t}$ decays. The MC expectations of these contributions are given in Table 5.

It is interesting to note the following:

- Approximately 23%, of the contribution is from Type-I (Charge mis-identification).
- Almost all of the charge mis-identification is from electrons, in ee and $e\mu$ channel.

Same Sign Leptons	Total	Type-I	Type-II	Type-II a)	Type-II b)	Type-III
ee	0.44±0.14	0.09±0.06	0.35±0.12	0.17±0.09	0.17±0.09	0.00±0.00
$\mu\mu$	0.13±0.08	0.00±0.00	0.13±0.08	0.13±0.08	0.00±0.00	0.00±0.00
$e\mu$	0.39±0.13	0.13±0.08	0.26±0.11	0.26±0.11	0.00±0.00	0.00±0.00
total	0.96±0.21	0.22±0.10	0.74±0.18	0.57±0.16	0.17±0.09	0.00±0.00

Table 5: Expected number of $t\bar{t}$ events of various types in 100 pb^{-1} of integrated luminosity. Uncertainties are from MC statistics.

- The bulk of the $t\bar{t}$ contribution in our event selection is from Type-II. The dominant among them is due to heavy flavor sources ($\sim 60\%$)
- No events with two fake leptons (Type-III) are expected within 100 pb^{-1}.

In the following sections, we will briefly describe two different data driven approaches to estimate these contributions. The Type-I contribution will be estimated using the "charge-flip rate" described below, whereas the Type-II part will be estimated using the lepton fake rate method [11].

6 Estimation of Charge Mis-Identification (Type-I)

In order to understand the Type-I background, we need to know the probability as a function of P_T and η for an electron to have its charge misassigned (flipped). We called this probability the "charge-flip rate", $P_{ChargeFlip}$.

The sample of $Z \rightarrow ee$ is an ideal sample to measure the charge-flip rate. However, the P_T range of electrons from Z decays is limited to (roughly) 20 GeV $< P_T <$ 60 GeV, while we need the flip-rate down to 10 GeV and up to higher momenta. Our plan then is to measure $P_{chargeFlip}$ from Z data in the accessible P_T range; we will then compare the $P_{chargeFlip}$ from Z data with the $P_{chargeFlip}$ from Monte Carlo using a single electron gun. In the analysis we will use the Monte Carlo flip rate, possibly adjusted if we find that data and Monte Carlo do not agree

In this Section we demonstrate that we are able to extract $P_{chargeFlip}$ from a Monte Carlo Drell Yan sample, and that this flip rate agrees with the flip rate from a single electron gun. The analysis described here is based on a dielectron sample within the Z mass region ($76 < m_{ee} < 106$ GeV). No opposite sign requirement is made. Both electrons must pass al the identification and isolation criteria described in Section 3.

6.1 Measurement of charge-flip rate on Z events

We define the charge-flip rate as

$$P_{ChargeFlip} = \frac{N_{Wrong}(P_T, |\eta|)}{N_{Total}(P_T, |\eta|)} \tag{1}$$

where $N_{Wrong}(P_T, |\eta|)$ is the number of wrongly charged electrons and $N_{Total}(P_T, |\eta|)$ is the total number of electrons in the sample. We select events with same sign (SS) and opposite sign (OS) dielectrons within the Z mass range, with both electrons passing the selection described in Section 3.1.

We determine $P_{ChargeFlip}$ in two steps. In the first step we limit ourselves to the barrel region, $|\eta| < 1.0$, where we expect $P_{ChargeFlip}$ to have a small η dependence. Thus, we select $Z \rightarrow ee$, same sign (SS) as well as opposite sign (OS), with both electrons in the barrel. We construct two $P_T - |\eta|$ distributions, one for electrons in the SS sample and one for electrons in the OS samples. Neglecting the small probability of double charge flips, the OS distribution will contain only electrons with the correct charge assignment, while the SS distribution will contain a 50-50 admixture of charge flipped and non-charge flipped electrons. Then, the number of wrongly charged electrons in the barrel as a function of P_T and $|\eta|$ is obtained as

$$N_{Wrong}(P_T, |\eta|) = SS(P_T, |\eta|) - k * OS(P_T, |\eta|) \tag{2}$$

The normalization k is given by the ratio of SS and OS events. The charge-flip rate in the barrel is then obtained by dividing $N_{Wrong}(P_T, |\eta|)$ by the total number of electrons as a function of P_T and $|\eta|$.

Once $P_{ChargeFlip}$ in the barrel has been determined, the second step of the procedure is to extend the measurement to higher rapidities ($|\eta| > 1$). This is done using $Z \to ee$, OS as well as SS, with one electron of $|\eta| < 1$ and one electron of $|\eta| > 1$. $P_{ChargeFlip}$ for $|\eta| > 1$ is determined by taking the ratio of the $P_T - |\eta|$ distributions of electrons with $|\eta| > 1$ in the SS sample to the total. A correction needs to be applied to account for the events with the charge of the $|\eta| < 1$ flipped. This is done using the $P_{ChargeFlip}$ for $|\eta| < 1$ determined in the first step.

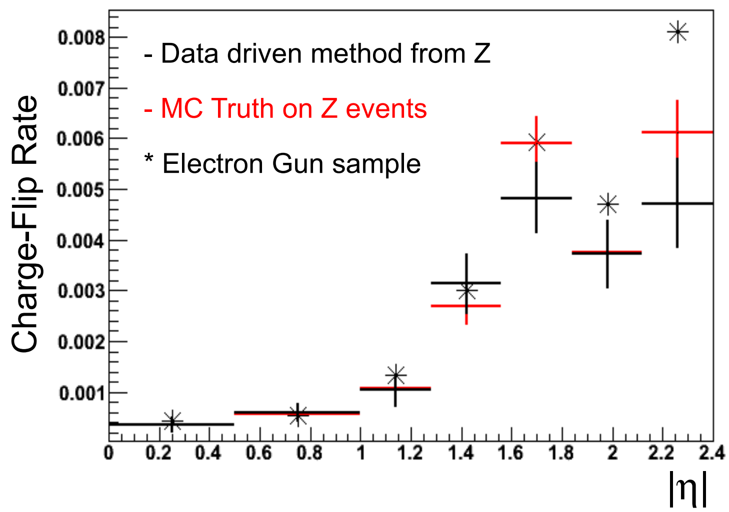

Figure 3: Charge-flip rate as a function $|\eta|$ for the "electron gun" sample (star) compared with distributions using data driven method from Z (black histograms) as well as truth matched Z events (red histogram). The integrated uminosity of the Z sample is about 300 pb^{-1}.

It is important to verify that this procedure yields the correct charge-flip rate. To this end, in Figure 3 we compare three flip rates as a function of $|\eta|$ and integrated over P_T:

1. The flip rate obtained from the Z sample using the two-step procedure outlined above.
2. The flip rate obtained from the Z sample using MC truth.
3. The flip rate obtained from the single electron gun, weighted by the $P_T - \eta$ distribution of electrons in the Z sample.

The three distributions agree quite well. This demonstrates that we are able to measure the flip rate from the data in the P_T region covered by $Z \to ee$.

6.2 Application of charge-flip rate to our analysis

We now perform a test of the charge flip prediction on $t\bar{t}$ MC events by relaxing cuts on $\not{E}_T$and jets. The test is meant to demonstrate that the charge-flip rate as determined from the "electron gun" sample can be applied to $t\bar{t}$ events. Dilepton events are selected without any $\not{E}_T$and jet cuts using an integrated luminosity of 100 pb^{-1}. The observed event yield is obtained by selecting same sign dielectron events. In order to get the estimation we use the following procedure:

- Select opposite sign dielectrons using the standard selection.
- Obtain the $P^1_{ChargeFlip}$ and $P^2_{ChargeFlip}$ for each electron for a given p_T and η. We use the charge-flip rate from the electron gun.
- Assuming either of the electrons can flip signs, the flip probability is given by $F = P_{ChargeFlip}/(1 - P_{ChargeFlip})$.

- Weight each event by $weight * (F^1 + F^2)$.
- Add up all of the weights

Results of the Monte Carlo tests for event yields is given in Table 6. From this study we can conclude that the charge-flip rate does a very good job of reproducing the rate of charge mis-identification of electrons in $t\bar{t}$ events.

Sample	Event yield
$t\bar{t}$ (Observed)	2.4 ± 0.3
$t\bar{t}$ (Predicted)	2.1

Table 6: Monte Carlo test of the electron charge-flip rate. Rates are normalized to 100 pb^{-1} of integrated luminosity.

We are now ready to apply the charge-flip rate to our $t\bar{t}$ sample after the full analysis selection. The same sign dilepton sample will have three different contributions from Type-I, Type-II and Type-III events. The results of the application of the procedure outlined above is summarized in Table 7.

Same Sign leptons	Total	Type-I	Type-II	Type-II a)	Type-II b)	Type-III
ee (predicted)	0.05	0.05	0.00	0.00	0.00	0.00
$\mu\mu$ (predicted)	0.00	0.00	0.00	0.00	0.00	0.00
$e\mu$ (predicted)	0.07	0.07	0.00	0.00	0.00	0.00
total (predicted)	0.12	0.12	0.00	0.00	0.00	0.00

Table 7: The number of events predicted using charge-flip rate in $t\bar{t}$ events for various types. Rates are normalized to 100 pb^{-1}.

Using Table 5 as observed and Table 7 as the prediction, we find that the charge-flip method predicts the bulk of the Type-I events. Out of a total of 0.22 ± 0.10 (roughly corresponds to 5 MC events) expected events, we predict 0.12 (≈ 3) events in 100 pb^{-1}. As expected, the method does not predict any contribution from Type-II events. We consider the agreement to be satisfactory.

7 Data driven Method for Fake Lepton backgrounds (Type-II)

In this analysis, the primary source of events are from Type-II category. As shown in Table 5, approximately 3/4 of these events are expected to be from heavy flavor sources, while the remainder is from fake leptons (Type-II b)). This means that one needs to be careful in defining the lepton fake rates to make sure they predict both sources of fakes with sufficient accuracy.

In Reference [11] we described a data-driven method to predict the fake background in a dilepton analysis. This method is being applied by us also in the WW [3] as well as $t\bar{t}$ [4] analyses. Here we briefly summarize the method, and we apply it to the same sign dilepton study.

7.1 The fake rate definition

The method starts by defining a fake rate (FR) measured in QCD events. We use the Pythia QCD samples with $\hat{P}_T > 30, 80$ GeV. The fake rate is defined as the probability for a lepton passing loose cuts (Fakeable Object, FO) to pass the analysis cuts as a function of p_T and η. The measured probability is then applied to dilepton candidates passing loose cuts to obtain a prediction to the fake lepton contribution. The details of the applications of the FR are given in Section 7.2.

Fakeable Objects are defined as follows:

- Electron Fakeable Object, eFO:
  - GSFElectron with $p_T > 10$ GeV;
  - $|\eta| < 2.4$;

– No reconstructed muon with $\Delta R < 0.1$;
– Electron ID and Conversion rejection defined in Section 3.1 as well as veto used in Section 3.4;
– Iso $<$ 0.4, where Iso=Sum/Max(20 GeV, P_T), and Sum = tkIso + hcalIso + Max(0 GeV, ecalIso - 2 GeV).

- Muon Fakeable Object, μFO:
  – Global and Tracker Muon with $p_T > 10$ GeV;
  – $|\eta| < 2.4$;
  – Global fit χ^2/ndof < 20;
  – $|d_0| <$ 200 μm (from silicon track, corrected for beamspot);
  – Iso $<$ 0.4, where Iso=Sum/Max(20 GeV, P_T), and Sum = tkIso + hcalIso + ecalIso.

The FR for electrons and muons are determined from the QCD sample. The FR does not give a direct measure for an absolute lepton fake rate. It is the probability for a fake lepton passing loose identification requirements as well as isolation to also pass a tighter selection.

7.2 Application of lepton fake rate to our analysis

We evaluate the FR in $t\bar{t}$ Monte Carlo events. This test is meant to check if the FR as determined from the QCD events can be applied to $t\bar{t}$. In order to perform this test we define the following four event selections:

- $t\bar{t} \to WbWb \to \mu + e\ \nu\nu bb$:
  – Require a global muon with $p_T > 10$ GeV, truth matched to $W \to \mu$.
  – Require a same sign electron that passes all the standard identification and isolation requirements.
- $t\bar{t} \to WbWb \to \mu + (eFO \times FR)\ \nu\nu bb$:
  – Require a global muon with $p_T > 10$ GeV, truth matched to $W \to \mu$.
  – Require a same sign eFO; weight each event by the FR for the corresponding eFO.
- $t\bar{t} \to WbWb \to e + \mu\ \nu\nu bb$:
  – Require an electron with $p_T > 10$ GeV, truth matched to $W \to e$.
  – Require a same sign muon that passes all the standard identification and isolation requirements.
- $t\bar{t} \to WbWb \to e + (\mu FO \times FR)\ \nu\nu bb$:
  – Require an electron with $p_T > 10$ GeV, truth matched to $W \to e$.
  – Require a same sign μFO; weight each event by the FR for the corresponding μFO.

Sample	Event yield
$t\bar{t}$ with $\mu + e$ (Observed)	2.1 $\pm$ 0.3
$t\bar{t}$ with $\mu + (eFO \times FR)$ (Predicted)	2.9

Table 8: Monte Carlo test of the electron fake rate using 100 pb^{-1} of integrated luminosity.

Sample	Event yield
$t\bar{t}$ with $e + \mu$ (Observed)	2.4 $\pm$ 0.3
$t\bar{t}$ with $e + (\mu FO \times FR)$ (Predicted)	2.7

Table 9: Monte Carlo test of the muon fake rate using 100 pb^{-1} of integrated luminosity.

The Monte Carlo test for the electron FR consists of comparing event yields and distributions for $\mu + e$ and $\mu + (eFO \times FR)$. Similarly, for muon FR it consists of comparing event yields and distributions for $e + \mu$ and

$e + (\mu DO \times FR)$. Results of the Monte Carlo tests for event yields are given in Tables 8 and 9. The uncertainties are from the MC statistics. From these studies we conclude that the QCD FR parametrization does a good job of reproducing the rate of fake electrons and muons in $t\bar{t}$ events.

In order to obtain a prediction of the fake contribution to our analysis, we proceed in the following way:

- Select lepton $+FO$ events where
  - one of the leptons passes all the standard identification and isolation requirements.
  - the other lepton is a FO but fails the standard identification and isolation requirements.
- The event passes all the standard kinematical cuts as outlined in Section 3.
- Weigh each event by $FR/(1 - FR)$, where FR is the fake rate for the FO under consideration.
- Add up all the weights.

Same Sign leptons	Total	Type-I	Type-II	Type-II a)	Type-II b)	Type-III
ee (predicted)	0.21	0.01	0.20	0.15	0.05	0.00
$\mu\mu$ (predicted)	0.10	0.00	0.10	0.09	0.01	0.00
$e\mu$ (predicted)	0.31	0.01	0.30	0.26	0.04	0.00
total (predicted)	0.62	0.02	0.60	0.50	0.10	0.00

Table 10: The number of events predicted using lepton fake rate method in $t\bar{t}$ events for various types. Rates are normalized to 100 pb^{-1}.

The results of the application of the procedure outlined above is summarized in Table 10. We conclude the following in comparison with the observed events from Table 5:

- We predict within $\sim 20\%$ of the observed Type-II contributions.
- Within Type-II, the contribution from events with heavy flavor sources are largely predicted ($\sim 88\%$).
- The method introduces an overestimate for the true leptons in Type-I. This is at $\sim 2\%$ level, and is negligible compared to the associated statistical as well as systematic uncertainties.

8 Application of the data driven methods to the SM and SUSY benchmark points

We apply the two data driven procedures to predict the backgrounds in the $t\bar{t}$ -dominated SM sample. Table 11 shows the contribution of all SM background. The prediction and observation agree to within $\sim 30\%$.

Same Sign Leptons	Total SM	$t\bar{t}$	tW	WZ	ZZ	WW	DY	Wjets
ee (observed)	0.45±0.14	0.44±0.14	0.00±0.00	0.00±0.00	0.01±0.01	0.00±0.00	0.00±0.00	0.00±0.00
ee (predicted)	0.27	0.26	0.01	0.00	0.00	0.00	0.00	0.00
$\mu\mu$ (observed)	0.17±0.08	0.13±0.08	0.00±0.00	0.03±0.02	0.01±0.01	0.00±0.00	0.00±0.00	0.00±0.00
$\mu\mu$ (predicted)	0.11	0.10	0.01	0.00	0.00	0.00	0.00	0.00
$e\mu$ (observed)	0.48±0.14	0.39±0.13	0.04±0.03	0.04±0.03	0.01±0.01	0.00±0.00	0.00±0.00	0.00±0.00
$e\mu$ (predicted)	0.39	0.38	0.01	0.00	0.00	0.00	0.00	0.00
total (observed)	1.10±0.21	0.96±0.21	0.04±0.03	0.07±0.03	0.03±0.01	0.00±0.00	0.00±0.00	0.00±0.00
total (predicted)	0.77	0.74	0.03	0.00	0.00	0.00	0.00	0.00

Table 11: Observed and predicted number of SM events passing the event selection in 100 pb^{-1} of integrated luminosity. The uncertainties are from MC statistics.

We also apply both of these methods to a combination of SM and SUSY samples to derive the prediction. Table 12 shows the contribution of observed and expected prediction to the sample. Typically, one would compare number of observed events with the prediction to look for excess in “signal” over the total background.

Same Sign	SM+LM0	SM+LM1	SM+LM2	SM+LM3	SM+LM4	SM+LM5	SM+LM6	SM+LM7	SM+LM8	SM+LM9
Observed	45.54±1.57	10.02±0.43	2.09±0.21	7.55±0.31	3.40±0.24	1.79±0.21	2.86±0.21	2.01±0.22	4.32±0.22	3.50±0.24
Predicted	4.10	1.26	0.83	1.24	0.91	0.81	0.84	0.83	1.03	1.00

Table 12: Observed and predicted number of SM and SUSY events passing the event selection in 100 pb^{-1} of integrated luminosity. The uncertainties are from MC statistics.

9 Remarks on systematic uncertainties

The associated systematic uncertainties are not discussed in this document. We plan to measure them in data. The expected dominant sources of systematical uncertainties are due to detector effects, effects of modeling of the contributing processes, uncertainties of the data-driven background prediction methods. Systematic uncertainties from the lepton selection, ID, and reconstruction efficiencies will be estimated based on the corresponding systematics of the tag-and-probe method used to determine these efficiencies $Z \to \ell\ell$ in data. We will assess the uncertainty arising from the jet energy scale using γ/Z balance with jets. The uncertainties of the data-driven background estimate will be studied using a measure of "bias per lepton" as a function of the either charge fake candidate in charge-flip rate or FO in lepton fake rate selections. The current document focuses on reducing and measuring SM backgrounds using the aforesaid data driven methods.

10 Conclusion

We have studied two different data driven methods to predict the backgrounds for searches beyond the standard model, in events with high P_T same sign dileptons (e or μ), significant hadronic activity, and high $\not\!\!E_T$. For these searches, the dominant background is from $t\bar{t}$. We characterized the background into different types based on charge mis-identification, leptons from heavy flavor sources, as well as fake leptons.

We used the charge-flip rate method to predict the number of wrongly charged leptons. The fakes including leptons from heavy flavor sources are predicted using lepton fake rates. The prediction using the methods is shown to work for the considered SM samples. Using both methods on the ensemble of the SM and SUSY samples, we show sensitivity towards several SUSY benchmark points using 100 pb^{-1} of integrated luminosity.

References

[1] http://www.ge.infn.it/t̃osi/cms/topMC.html.

[2] https://twiki.cern.ch/twiki/bin/view/CMS/ProductionSummer2008.

[3] "Prospects for measuring the WW production cross section in pp collisions at $\sqrt{s}$ =10 TeV", CMS AN-2009/042 and PAS EWK-09-002.

[4] "Expectations for observation of top quark pair production in the dilepton final state with the early CMS data", CMS AN-2009/050 and PAS TOP-09-002.

[5] "Correcting Missing Transverse Energy Using Tracks", CMS AN-2009/022.

[6] http://indico.cern.ch/materialDisplay.py?contribId=3&materialId=slides&confId=58309.

[7] https://hypernews.cern.ch/HyperNews/CMS/get/muon/258.html.

[8] "Muon Identification in CMS", CMS AN-2008/098.

[9] https://twiki.cern.ch/twiki/bin/view/CMS/VplusJets.

[10] http://indico.cern.ch/materialDisplay.py?contribId=3&materialId=slides&confId=64436.

[11] "Data-driven methods to estimate the electron and muon fake contributions to lepton analyses", CMS AN-2009/041.

[12] P. Kalavase, et. al, Analysis Note in preparation.