

50 pb^{-1} Z Candle Note Draft
— applied to 2.66 pb^{-1} of data

Yi Chen
California institute of technology
`chen.yi.first@gmail.com`

October 13, 2010

Contents

1	Introduction	2
2	Data & MC samples	2
3	Event cleaning	2
4	Jet flavors	2
5	Selections. Muon, Z. Isolation	3
6	MC sample efficiencies	3
7	Signal extraction (fit strategy)	5
8	Fit result on data	7
9	Toy studies: shape, parameters	7
10	JES uncertainty	10
11	Anti-muon control sample	10

1 Introduction

N/A. I'll fill this in later.

2 Data & MC samples

The monte-carlo simulations used in this note are summarized in table 1.

Table 1: MC samples used.

MC	Sample	Generator	Number of events
$Z \rightarrow \mu\mu + \text{Jets}$???	MADGRAPH	???
$W \rightarrow \mu\nu_\mu + \text{Jets}$???	MADGRAPH	???
$t\bar{t} + \text{Jets}$???	MADGRAPH	???
QCD (with two muons)	???	PYTHIA6	???

Data samples included in this note are listed in table ??.

Table 2: Data samples

Run range	dataset	integrated luminosity
132440-191053	/Run2010/GodDataSetForMuon	1.06 fb^{-1}

3 Event cleaning

Meow.

4 Jet flavors

All the jets are clustered using the anti-Kt algorithm with cone radius 0.5. A few different types of jeta are considered in this analysis.

1. Calorimetric jet (*CaloJet*). The jet constituents are taken from both hadronic calorimeter deposits and electromagnetic calorimeter deposits. The calorimeters are first combined into *calotowers* according the the geometrical locations of the cells. Calotowers are then treated as four vector with zero norm to be included in the jet clustering algorithm. The jets need to be corrected for energy. To estimate the effect of the energy correction, we also perform the analysis in terms of uncorrected energies. The threshold for counting CaloJets is 30 GeV/c, while the threshold for counting uncorrected CaloJets is 20 GeV/c.
2. Track jet. All tracks that pass some basic quality selection criteria are served as input to the jet clustering algorithm, again treating them as massless. We require that the number of hits along the track is greater than 6, and the normalized χ^2 is smaller than 20. The track vertex, compared to the primary vertex, has to be within 0.1cm in z direction, 0.1 cm in magnitude, and 600 μm in the transverse direction. Also, if

there exists another reconstructed vertex that is closer to the best primary vertex, the track is not included in the jet clustering. The pseudorapidity of the track is required to be $-2.4 < |\eta| < 2.4$, and the momentum is required to be within range 0.5-500 GeV/c. The threshold for track jet counting is 20 GeV/c.

3. Particle flow jet (*PF jet*). There is an elaborate set of algorithms that attempts to reconstruct particle candidates as good as possible, and then to start physics analysis from there. The jet algorithms take in the reconstructed candidates as input. The threshold for PF jet counting is 20 GeV/c. Compatibility with the primary vertex is not implemented yet.

5 Selections. Muon, Z. Isolation

In order to retain as much signal as possible, only a minimal selection is applied to data. For candidate muons, we require the following.

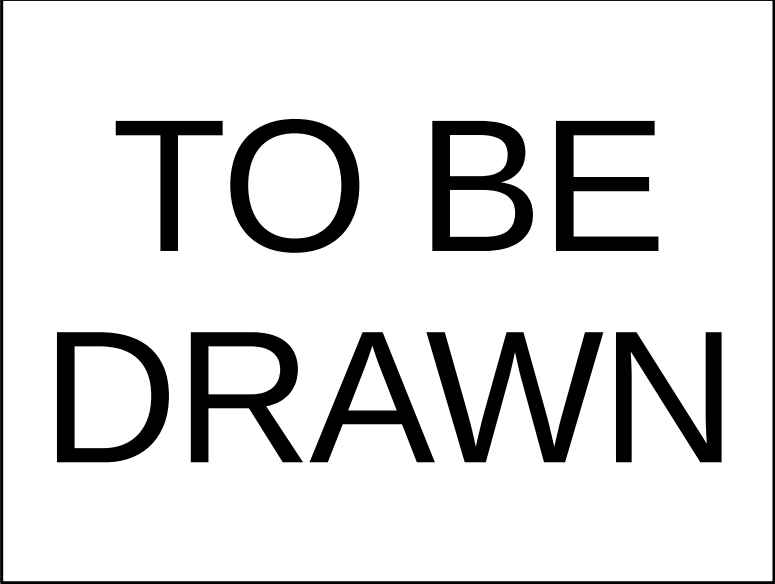
1. The muon is reconstructed both as global muon and tracker muon
2. One valid pixel hit for the muon track is required to be present, and a total of 6 hits in the tracker system is required.
3. We also require that there is one valid hit in the muon chambers that is consistent with the global track.
4. The maximum allowed global muon track fitted χ^2 is 10.
5. The PT threshold in this analysis for the first leg (tighter) muon is set to be 15 GeV/c with pseudorapidity range $-2.1 < |\eta| < 2.1$.
6. The limit on second leg muon is more relaxed for more statistics, and it is set on 10 GeV/c with range $-2.4 < |\eta| < 2.4$.

An additional isolation requirement is required on both muons that make up the Z candidate. This is mainly due to the fact that in the standard particle-flow jet clustering algorithm, the muons are considered as part of the input collections to the clustering algorithm, and because of the high transverse momentum, the muons are clustered as jets, as shown in figure 1. Furthermore, the muon jets tend to take up additional soft particles in the vicinity of the muon (see figure 2), causing extra systematic uncertainty in jet-counting. Therefore, a isolation cut is selected as 30% of relative combined isolation with cone size 0.3. This causes roughly 5% efficiency drop for one-jet case, consistent with expectations, as shown in figure 3.

For any event with at least two good muon candidates as defined in the previous section, the muon pair is formed as a Z candidate if the invariant mass of the two muons are with range 60-120 GeV/ c^2 . An additional requirement to reject fake muon pairs is to require that the two muons are at least 0.01 units apart in $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

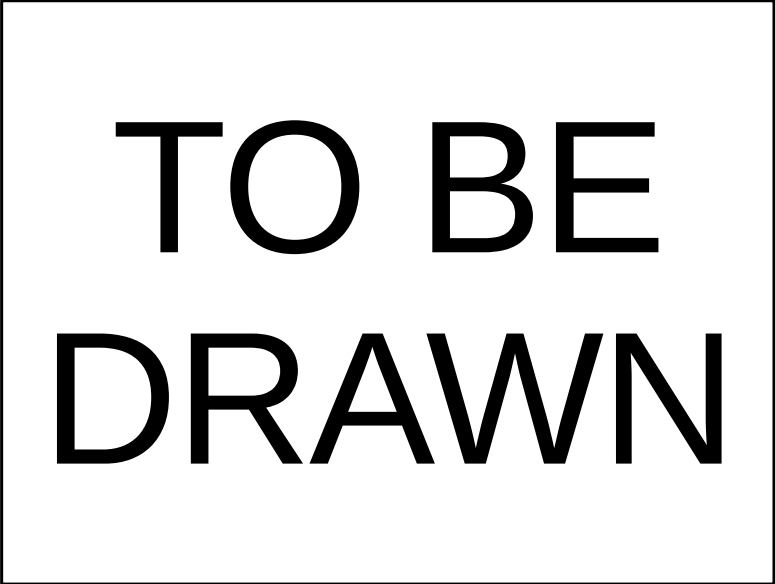
6 MC sample efficiencies

The efficiency of each step in the selection for MC samples are shown in table ??



TO BE
DRAWN

Figure 1: Candidate muon vs. closest jet, $\Delta\phi$ plot (MC signal)



TO BE
DRAWN

Figure 2: Muon PT vs. closest jet PT (MC signal)



Figure 3: Isolation cut vs. percentage that was cut out in MC signal events

Table 3: Efficiency table for MC samples

A	V	D	W	H
---	---	---	---	---

7 Signal extraction (fit strategy)

In order to extract signal yields from data, a fit is performed on the dimuon mass spectrum. The signal shape is taken as the Cruijff function, defined as

$$f(M_{ll}; m, \sigma_L, \sigma_R, \alpha_L, \alpha_R) = N_s e^{-\frac{(M_{ll} - m)^2}{2\sigma^2 + \alpha^2(M_{ll} - m)^2}},$$

where $\sigma = \sigma_L(\sigma_R)$ for $M_{ll} < m(M_{ll} > m)$ and $\alpha = \alpha_L(\alpha_R)$ for $M_{ll} < m(M_{ll} > m)$. The background shape is chosen as a falling exponential, as can be seen in the QCD sample which dominates the background in low jet multiplicity bin. When the number of jets is larger, events with $t\bar{t}$ start to have a significant contribution, and those are also well-modeled by a falling exponential. Even though the exponent is not the same in different samples, the exponent is floated in the fit, and the difference in exponent will be taken account for.

A minimum likelihood fit is performed simulatneously for all different jet counts, with the total likelihood written as



Figure 4: Isolation cut vs. percentage that was cut out in MC signal events

$$L = (Prefactor) \times \sum_i \left\{ \sum_{n_{jet}=1}^{n_{max}-1} \left((N_{S,n_{jet}} - N_{S,n_{jet}+1}) F_S(M_{ll}^i) + N_{B,n_{jet}} F_{B,n_{jet}}(M_{ll}^i) \right) \delta_{n_{jet},n_{jet}^i} \right. \\ \left. + (N_{S,n_{max}} F_S(M_{ll}^i) + N_{B,n_{max}} F_{B,n_{max}}(M_{ll}^i)) \theta(n_{jet}^i - n_{max}) \right\},$$

where F_S is the signal PDF, constrained to be the same for all jet bins. The background PDF, $F_{B,n_{jet}}$, is not constrained to be the same for different jet bins, and the exponents are left floating in the fit. Each term (except the last) is constrained to be with the same exclusive jet bin through the Kronecker delta function $\delta_{n_{jet},n_{jet}^i}$. The last jet bin is inclusive, including all number of jets greater or equal to n_{max} , constrained by the step function $\theta(n_{jet}^i - n_{max})$.

All the parameters for signal PDF are floated except α_L . Since the $\alpha_L(\alpha_R)$ controls how large the tail is, if α_L is floated, in the presense of background events, which is approximated by an exponential function, the tail of the signal PDF will be larger then real signal shape. As a result, the signal yield will include significant contributions from background events, which does not serve our purpose of extracting the signal yield. Fortunately, the signal shape does not depend too much on isolation, as shown in figure 4 in MC studies. Therefore we can apply a very tight isolation cut to remove all backgrounds and some signal events, and fit to get α_L to be used to extract signal yield as a function of jet multiplicity.



Figure 5: Signal yields relative to the perfect case when α_L is varied while keeping others fixed.

8 Fit result on data

The fit strategy described above is applied to data.

9 Toy studies: shape, parameters

In order to demonstrate the dependence of each parameter in the signal fit, a toy study is performed. The fitted Z signal shape is taken as a template, and 1000 toy experiment samples are generated assuming 50/pb of integrated luminosity. Each toy experiment is then fitted with signal functional form with slightly different parameters compared to the input parameters. The pull distribution of the signal yield is plotted and compared to Gaussian of width 1.

For each shape parameter, we vary it by -10%, -5%, -2.5%, -1%, 0%, 1%, 2.5%, 5% and 10% while keeping other parameters fixed at the right (input) value. The signal yield relative to perfect case is plotted in figures 5, 6, 7 and 8. We can see that the signal yield is most sensitive to α_L and negligible in other parameters. This is because of the radiation effects of the muons lower the mass, and naturally there will be a tail in the lower side of the peak. Biases arise in estimating the size of the tail. However, even in the case of α_L , the bias in the signal yield is mild.

Fit with wrong α_L value while floating others is also analyzed with toy experiments. This is an estimate of the bias in realistic conditions as in the fit to extract signal value. The result is shown in figure 9. TODO: INSERT RESULT!

TO BE
DRAWN

Figure 6: Signal yields relative to the perfect case when α_R is varied while keeping others fixed.

TO BE
DRAWN

Figure 7: Signal yields relative to the perfect case when σ_L is varied while keeping others fixed.

TO BE
DRAWN

Figure 8: Signal yields relative to the perfect case when σ_L is varied while keeping others fixed.

TO BE
DRAWN

Figure 9: Signal yields relative to the perfect case when σ_L is varied and floating others.

10 JES uncertainty

Since CaloJets and PF jets are corrected for energy, there will be systematic uncertainties associated with the corrections. The CMS-official numbers for the uncertainties is quoted as

1. 10% overall scale for CaloJets
2. 5% overall scale for PF jets
3. 2% eta-dependent scale ($2\% \times |\eta|$) for both types of jets.
4. The two types of errors (absolute and eta-dependent) are assumed to be independent to each other.

Therefore the strategy to estimate the systematics of the uncertainties is to perform the fit and extract signal yields for different jet multiplicities 4 additional times with different thresholds:

1. Original threshold + 10%
2. Original threshold - 10%
3. Original threshold + $2\% \times |\eta|$
4. Original threshold - $2\% \times |\eta|$

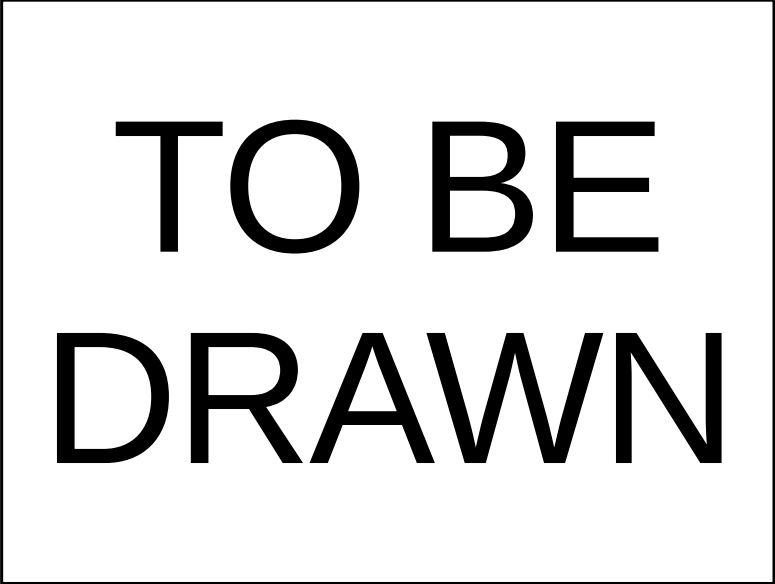
The difference between different thresholds are taken as an estimate to the jet energy correction uncertainty. The result, when applied to data, is shown in figures 10 and 11 for CaloJets and PF jets respectively.

11 Anti-muon control sample

To justify the choice of exponential PDF for background events, an anti-muon control sample check is performed. This is done by comparing the following

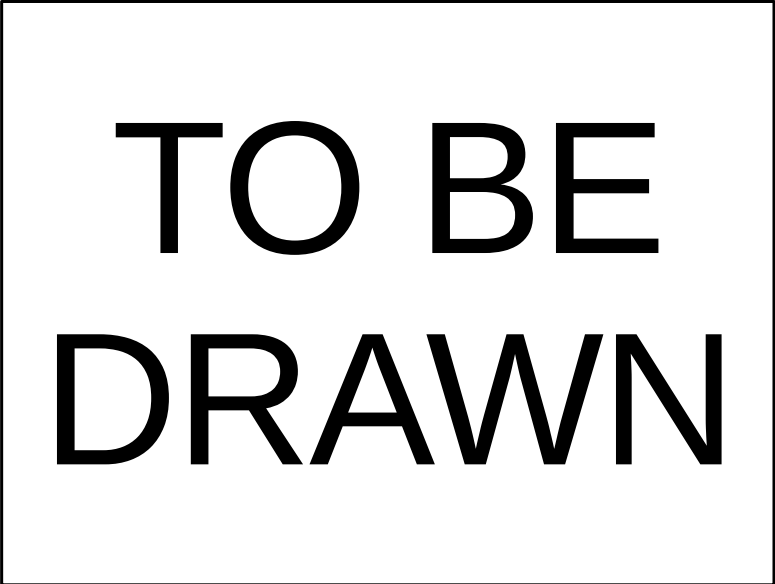
1. MC QCD sample, normal selection
2. MC QCD sample, inverting isolation on the first leg muon
3. Data, inverting isolation on the first leg muon

The result is shown in figure 12. Each histogram is scaled to the same area so that we can compare the shape. Indeed, the shapes are similar to each other. Unfortunately there isn't enough statistics on the QCD sample to accomodate the increasing statistics in data.



TO BE
DRAWN

Figure 10: Jet energy scale uncertainty estimation for CaloJets



TO BE
DRAWN

Figure 11: Jet energy scale uncertainty estimation for PF jets

**TO BE
DRAWN**

Figure 12: Anti-muon check