View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Scientific Open-access Literature Archive and Repository

IL NUOVO CIMENTO Vol. 35 C, N. 6 Novembre-Dicembre 2012
DOI 10.1393/ncc/i2012-11382-6

Colloquia: LaThuile12

Finding a Z → 2 jets signal in W + 3 jets events at CDF

C. Vernieri

Scuola Normale Superiore - Pisa and INFN, Sezione di Pisa - Pisa, Italy

ricevuto il 7 Settembre 2012

Summary. — The observation of WZ associated production at the Tevatron in a final state with a lepton, missing transverse energy and jets is difficult since the signal rate is low and competes with a huge background. In an attempt to increase the acceptance, the sample where three high-energy jets are reconstructed is investigated. In this sample, which within our event selection cuts includes 1/3 of the diboson signal events, rather than choosing the two transverse energy (E_T) leading jets to detect a Z signal, the information carried by all jets is combined.

PACS `14.70.-e` – Gauge bosons.
PACS `13.85.Ni` – Inclusive production with identified hadrons.
PACS `13.85.Qk` – Inclusive production with identified leptons, photons, or other nonhadronic particles.
PACS `13.87.-a` – Jets in large-Q^2 scattering.

1. – Introduction

In the Standard Model (SM) of particle physics, the electroweak bosons are the gauge bosons of the local $SU(2)\otimes U(1)$ symmetry. All couplings of the W, Z bosons and photons are well defined within this symmetry, while the W- and Z-boson masses arise because of the spontaneous breaking of this symmetry.

The study of diboson (W,Z) production at hadron colliders provides a test of the electroweak sector of the SM, since any deviation from the predicted WWZ, WZZ couplings (TGC, Trilinear Gauge Couplings) would be indicative of new physics [1].

Diboson measurements are also instrumental for searches for the SM light Higgs boson. By choosing to focus on the final state where a Z-boson decays into $b\bar{b}$-pairs, the topology of WZ events would be the same as expected for associated production of a W and a light Higgs boson ($M_H < 135\,\mathrm{GeV}$). At the Tevatron, the process WH → W$b\bar{b}$ has an expected cross section times branching ratio ($\sigma \cdot \mathrm{BR}$) about five times lower than WZ → W$b\bar{b}$ for $M_H \simeq 120\,\mathrm{GeV}/c^2$. Therefore, observing that process would be a benchmark for the even more difficult light Higgs search in the WH → W$b\bar{b}$ process.

© Società Italiana di Fisica

2. – Motivations for the three-jets studies

Observing WZ associated production at the 1.96 TeV center-of-mass energy of the Fermilab proton-antiproton Tevatron collider is difficult since the event rate is extremely low. NLO calculations predict WZ production cross section to be about 3.22 pb [2]. Thus, one expects a handful of events per fb^{-1} of integrated luminosity in the $\ell\nu q\bar{q}$ final state, after allowing for trigger and kinematical selection efficiency. This statement remains valid even if the few accepted ZZ events with leptonic decay of one Z, where one lepton is not detected, are included.

Furthermore, the signal to background ratio is very poor, due primarily to the large background contributed by the production of W and associated jets. Since the main signal feature to be exploited to disentangle signal from background is the invariant mass of H-decay jets, the correct selection of the jets to be assigned to H decay and an optimal resolution in jet systems mass is of utmost importance.

In diboson analyses at CDF the standard kinematical cut requires two high-energy jets (*i.e.* $E_T > 20\,\mathrm{GeV}$) in the candidate sample (*two-jets region*). Since simulations show that if a third high-energy jet is allowed (*three-jets region*, as defined by our selection cuts on jet energy), the signal acceptance is increased by 33%, it would be important to be able to detect the Z signal also in events with more than two high-energy jets.

However, the issue is confused because in WZ events additional jets may be initiated by gluon(s) radiated from the interacting partons (Initial State Radiation, ISR) or from the Z-decay products (Final State Radiation, FSR). This work presents a method to overcome this difficulty and by making optimal use of the information on diboson production contained in the sample with 3 associated jets.

Extra-activity produced by spectator partons or by pile-up of events was found to be negligible in our studies.

2·1. *Event selection.* – The experimental signature involves the presence of a charged lepton (electron or muon), a neutrino (identified through the missing transverse energy, $\not{E}_T$) and large-E_T jets.

The offline event selection identifies jets using the JETCLU cone algorithm with radius $\sqrt{(\Delta\phi^2 + \Delta\eta^2)} = 0.4$, in the space of azimuthal angle ϕ and pseudorapidity η, corrected for detector effects as described in [3].

The sample we investigate is selected by the following cuts:

- exactly three jets(1) with $E_T(J_1, J_2, J_3) > 25,\ 15,\ 15\,\mathrm{GeV}$ and $|\eta(J_1, J_2, J_3)| < 2,\ 2,\ 3.6$;
- an isolated triggered electron or muon with $|\eta| < 1.1$ and $E_T > 20\,\mathrm{GeV}$;
- $\not{E}_T > 20\,\mathrm{GeV}$;
- Multi-jet QCD veto:
  - $M_T^W > 10\ (30)\,\mathrm{GeV}$ if the triggered lepton is a muon (electron), M_T^W being the W-invariant mass in the transverse plane,
  - $\not{E}_T$-significance(2) > 1.8 if the triggered lepton is an electron.

---

(1) Events with a fourth jet with $E_T > 10\,\mathrm{GeV}$ are rejected.

(2) $\not{E}_T$-significance $= (-\log_{10}(P(\not{E}_T^{fluct} > \not{E}_T)))$, where P is the probability and $\not{E}_T^{fluct}$ is the expected missing transverse energy arisen from fluctuations in the energy measurements [4].

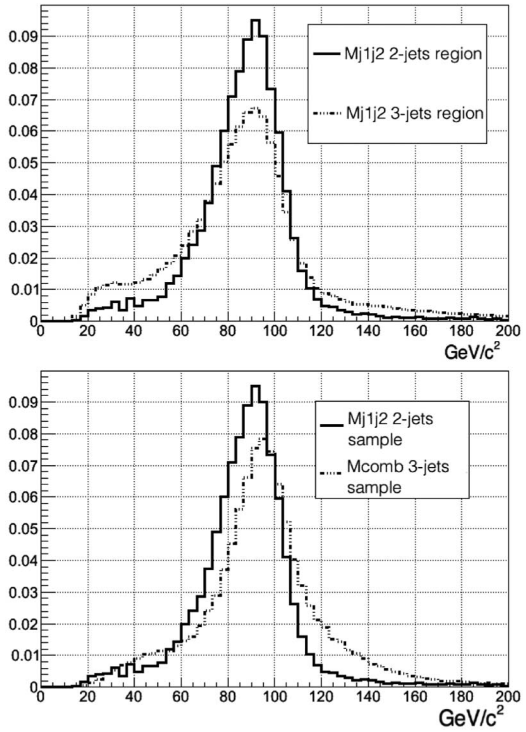

Fig. 1. – Top, $M(J_1J_2)$ in the *three-jets region* (dotted) is compared to $M(J_1J_2)$ in the *two-jets region*. Bottom, $M(J_1J_2)$ in the *two-jets region* is compared to MJJ_{COMB} (dotted) in the *three-jets region*.

Then, two different subsamples corresponding to an integrated luminosity of $6.6\,\text{fb}^{-1}$ are studied separately. One, the tag sample, where two b-jets in the final state are required, represents the golden channel for the light SM Higgs boson search at Tevatron (WH → W$b\bar{b}$). In this analysis the b-tagger employed is the b-ness [5], which is a multivariate, neural network (NN) based tagger. It provides an output value serving as a figure of merit to indicate how b-like a jet appears to be. Jets with increasing b-ness are more b-like.

The second, the notag sample is the sub-sample of the pretag sample(3) where the tag obtained by removing the tag sample. This makes the tag and no-tag samples independent of each other and allows combining the results obtained by analyzing the two samples.

In order to select the tag sample we require the two leading jets to have $b$ness $>$ 0.75, -0.2 respectively. These cuts have been optimized against the sensitivity of the measurement. In fig. 1 the invariant mass built using the two E_T leading jets $M(J_1J_2)$ for WZ MC events in the *two jets region* is compared with the same distribution built in the *three jets region*. In the three jets region, since jets due to initial or final state radiation confuse the choice of the jet system to be attributed to Z decay, $M(J_1J_2)$ has

---

(3) Pretag sample is the one where no constrain on jets flavor are applied.

TABLE I. – *Predicted and observed number of events in the notag and tag samples. W+jets and QCD rates are estimated by fitting data. The expected rates are separated for different triggered lepton type. By construction the expected numbers are equal to the observed ones.*

NOTAG	Process	Rate (Electrons)	Rate (Muons)
	Signal (WZ/ZZ)	66.2 ± 0.9	69.5 ± 0.9
	WW	386.2 ± 3.0	311.1 ± 3.1
	$t\bar{t}$	333.0 ± 1.4	288.5 ± 1.2
	single-top	68.9 ± 0.4	57.8 ± 0.3
	Z+jets	350.0 ± 3.2	1167.8 ± 4.5
	W+jets	10304.2 ± 29.6	8275 ± 22.8
	QCD	1600.4 ± 60.0	352.3 ± 5.4
	Total Observed	13109.0 ± 114.5	10522.0 ± 102.6
TAG	Process	Rate (Electrons)	Rate (Muons)
	Signal (WZ/ZZ)	3.5 ± 0.2	3.6 ± 0.2
	WW	6.2 ± 0.4	4.7 ± 0.3
	$t\bar{t}$	146.4 ± 0.9	127.9 ± 0.8
	single-top	22.5 ± 0.2	18.7 ± 0.2
	Z+jets	8.0 ± 0.4	23.6 ± 0.6
	W+jets	212.0 ± 3.9	189.9 ± 3.2
	QCD	32.5 ± 0.3	5.7 ± 0.0
	Total Observed	431.0 ± 20.8	374.0 ± 19.3

a degraded resolution: high mass and low mass tails due to wrong combinations are present. It is reasonable to expect that choosing the correct jet combination MJJ_{COMB} (to be defined later) for building the Z mass would improve the resolution. (see fig. 1, bottom). This work builds on the analysis methods reported in [6].

2·2. *Composition of the selected events.* – The following processes contribute to a data sample selected within our cuts:

- Electroweak and top (EW): WW, WZ, ZZ, Z+jets, $t\bar{t}$, single *top*. Each of these processes can mimick the signal signature, with one detected lepton, large $\not\!\!E_T$ and jets. The contamination of these processes in the selected data sample is estimated by using their accurately predicted cross sections [2]. The shapes (templates) of a number of observables are obtained from ALPGEN+Pythia [7], Pythia MC [8] after the simulation of the CDF detector.
- W($\rightarrow l\nu$)+jets, $l = e, \mu, \tau$. Due to the presence of real leptons and neutrinos, the $W +$ jets background is the hardest to be reduced. Templates are obtained from ALPGEN+Pythia MC, while the rate normalization is obtained from data [6].
- QCD: multi-jet production with a jet faking the lepton and fake $\not\!\!E_T$. Since the mechanism for a jet faking a lepton or for fake missing transverse energy is not expected to be well modeled in MC events, both rate normalization and templates are obtained from data.

In table I we show the estimated number of events for each process contributing for the $M(J_1 J_2)$ distribution in the notag and tag samples.

3. – Adopted strategy

In order to simulate the WZ $\rightarrow \ell\nu jjj$ process we used the ALPGEN generator interfaced to the generator PYTHIA to include jet fragmentation.

Jets are ordered in decreasing E_T in the notag sample and in decreasing b-ness in the tag sample(4).

We started from studying the three jets sample in WZ MC in which jets are matched in direction to particles produced by the hadronization ("hadrons") of partons from Z. The matching algorithm implemented searches for hadrons rather than quarks in the jet cone and traces back the origin of the hadrons in order to understand if they were produced by a Z-decay. In this way the rate of matching reaches $\sim 99\%$(5) and it allow us to train NNs with a set of events as much as possible similar to the real data.

Since PYTHIA saves all the information related to stable hadrons produced by partons hadronization for each hadron shower we are able to state if it comes from a primary beam parton (ISR) or if it originates from Z (FSR). Then, we look for stable hadrons within the jet cone and for each of the 3 jets in the event, we ask that the total hadron energy originating from a single parton is $> 50\%$ than the jet energy. With this method we are able to label the 99% of jets as ISR or FSR.

Once the origin of each jet is well understood we know event-by-event which jet combination should be used to reconstruct the Z mass (named the right jet combination, RJC). In terms of the frequency of RJC the notag (tag) sample is composed as follows:

1. J_3 is from ISR, J_1 and J_2 from FSR $\mapsto$ RJC $= J_1J_2$: 33.5% (53.4%) of events
2. J_2 is from ISR, J_1 and J_3 from FSR $\mapsto$ RJC $= J_1J_3$: 21.4% (9.5%) of events
3. J_1 is from ISR, J_2 and J_3 from FSR $\mapsto$ RJC $= J_2J_3$: 10.8% (4.9%) of events
4. J_1, J_2, J_3 are from FSR $\mapsto$ RJC $= J_1J_2J_3$: 33.3% (31.2%) of events

Notice that in tag sample J_1J_2 is the RJC in the 53.4% of cases, since jets are ordered in b-ness and we require the two b-ness leading jets to satisfy some criterion. The greater contribution of $M(J_1J_2)$ in the whole sample is the reason why in the tag sample the resolution is already good for the distribution built with the two jets with highest b-ness. Still, even in this sample a better combination than J_1J_2 can be searched for in $\sim 47\%$ of events.

3˙1. *Neural Networks*. – Four different Neural Networks (NNs) have been trained, using MLP method [9], in MC signal events to isolate each of the above cases: NN(J_1J_2), NN(J_1J_3), NN(J_2J_3) and NN($J_1J_2J_3$). These NNs combine kinematical information and some tools developed by CDF Collaboration for discriminating gluon-like and b-like jets from light-flavored jets [5, 10]. Inputs to NNs are:

1. Kinematical variables:
    - $\mathrm{d}\eta_{j_ij_k} = |\eta_{j_i} - \eta_{j_k}|$
    - $\mathrm{d}R_{j_ij_k}$

---

(4) J_1, J_2 would be the two with highest b-ness value, J_3 the one with highest E_T among the others.

(5) The rate of matching jets to quarks is about 60%.

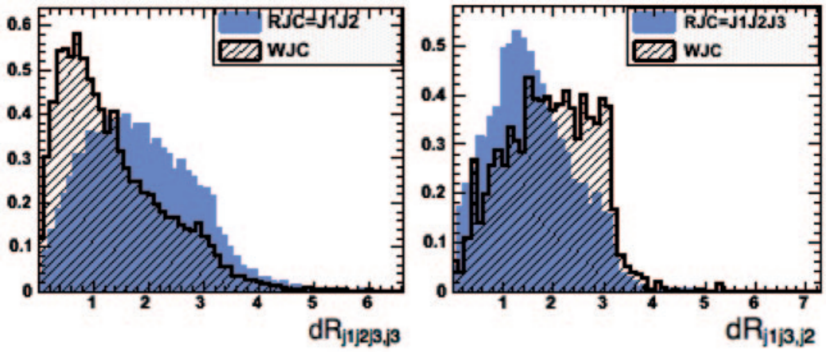

Fig. 2. – Some distributions of the variables used as input to NNs, built for the RJC sample and for the complementary one (shaded).

TABLE II. – *Parameters of the fits to the distributions of* $M(J_1J_2)$ *and* MJJ_{COMB} *in the tag and notag samples. A is the acceptance; p is the purity which is defined as the fraction of events where the corrected jets are selected;* σ *and* μ *are width and average of Gaussian fits to the distributions in the mass window* $[70, 110]\,\mathrm{GeV}/c^2$.

	Notag: $M(J_1J_2)$	MJJ_{COMB}	Tag: $M(J_1J_2)$	MJJ_{COMB}
A	100%	90%	100%	92%
p	35%	65%	53%	72%
σ/μ	0.25	0.13	0.22	0.14

- $\mathrm{d}R_{j_i\ell}$, $\mathrm{d}R_{j_kj_l,j_p}$, $\mathrm{d}R_{j_1j_2j_3,j_k}$ (6)

2. Variables related to the jet systems:
   - $m_{j_ij_k}/m_{j_1j_2j_3}$
   - $\gamma_{j_ij_k} = (E_{j_i} + E_{j_k})/m_{j_ij_k}$
   - $\gamma_{jjj} = (E_{j_1} + E_{j_2} + E_{j_3})/m_{j_1j_2j_3}$
   - "pt-imbalance" $= P_{Tj1} + P_{Tj2} - P_{T\ell} - \not{E}_T$
   - $\eta(j_i + j_k)/\eta(j_p)$, $p_T(j_i + j_k)/p_T(j_p)$
3. b/light quark discriminant, quark/gluon discriminant.

Based on the response of the four NNs, we determine the most likely jet combination for building the Z mass for each event. The method allows to use a different combination from J_1J_2 in about 65% (45%) of cases in the notag (tag) sample.

In fig. 2 some inputs are shown.

Combining by a set of subsequent optimal cuts(7) the information provided by the outputs of the four NNs, we build a MJJ_{COMB} Z-mass [6]. Using MJJ_{COMB} rather than $M(J_1J_2)$, the resolution improves by a factor ~ 2, see fig. 1 and table II.

---

(6) $i, k, p = 1, 2, 3$ are the indices of the jets. $\ell =$ highest E_T lepton.

(7) Cuts have been optimized against the sensitivity of the measurement.

TABLE III. – *Sensitivity of the fits considering only the three jets region.*

Fit Method	$P_{2\sigma}$	$P_{3\sigma}$
Fit signal $WZ/ZZ/WW$ (pretag)		
$M(J_1J_2)$	51.2%	6.4%
MJJ_{COMB}	66.7%	25.9%
	p-value	
Fit signal WZ/ZZ (notag+tag)		
$M(J_1J_2)$	0.44σ	
MJJ_{COMB}	0.54σ	

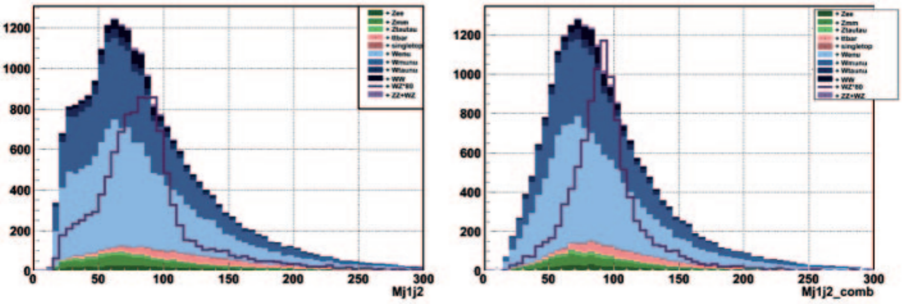

Fig. 3. – Simulation of signal+background for the notag sample. Left, $M(J_1J_2)$. Right, MJJ_{COMB}. The horizontal scale is in GeV/c^2. The signal is multiplied by 80.

We apply the method also to the main sources of background of a typical diboson analysis at CDF (W+jets, Z+ jets, $t\bar{t}$ and single top) and compare the result to WZ events. In figs. 3 and 4 and in table II one observes that MJJ_{COMB} allows a better separation of the WZ/ZZ signal from background in both notag and tag samples.

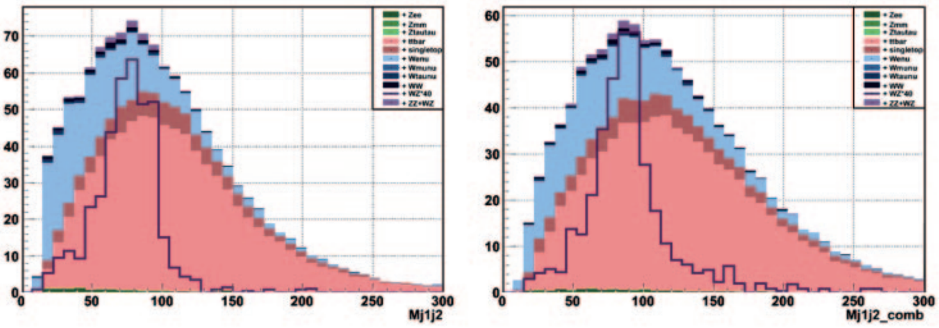

Fig. 4. – Simulation of signal+background for the tag sample. Left, $M(J_1J_2)$. Right, MJJ_{COMB} built with the criterion described in the text. The horizontal scale is in GeV/c^2 and the signal is multiplied by 40.

4. – Tests of the method

To qualify the potential of the method we have studied an experimental data sample accepting events with an isolated large E_T (p_T) lepton, large missing E_T and three large transverse-momentum jets. The selection cuts accept jets of all flavors (*pretag* sample), and all diboson events including WW besides WZ, ZZ may pass the cuts. We estimate the probability at three standard deviations level to extract an inclusive diboson signal. After our procedure for building the Z mass is applied, $P_{3\sigma}$ is about 4 times greater than when building the Z mass "by default" with the two E_T leading jets, as reported in table III.

This attempt represents just a check of our technique. A diboson signal has been observed at CDF using W events with exclusive two jets [11], we performed a test to gauge the probability of revealing a diboson signal also in the pretag three jets sample(8).

In order to discriminate WZ against the WW contribution we apply our technique considering only WZ/ZZ as the signal. We decide to treat separately the notag and tag three jets regions and then combine the results in order to reach a greater sensitivity. The sensitivity increases when MJJ_{COMB} rather than the standard $M(J_1J_2)$ is used: the expected p-value is about 20% greater in the former case (see table III).

In conclusion, our technique allows including the three jets sample in the WZ/ZZ search in order to increase acceptance and sensitivity in the search for the hadronically decaying Z boson.

* * *

The author is grateful to M. Trovato, Prof. G. Bellettini, Dr. G. Latino and Dr. V. Rusu for many fruitful discussions and suggestions.

REFERENCES

[1] Hagiwara K. *et al.*, *Phys. Rev. D*, **41** (1990) 2113.
[2] Campbell J. M. and Ellis R. K., *Phys. Rev. D*, **65** (2002) 113007.
[3] Bhatti A. *et al.*, *Nucl. Instrum. Methods A*, **566** (2006) 375.
[4] Culbertson R. *et al.*, CDF Public Note, 9184 (2008).
[5] Freeman J. *et al.*, *Nucl. Instrum. Methods A*, **663** (2012) 37.
[6] Vernieri C., CDF Public Note (Thesis), 10675 (2011).
[7] Mangano M. *et al.*, *JHEP*, **07** (2001) 001.
[8] Sjöstrand T. *et al.*, *Computer Phys. Commun.*, **135** (2001) 238.
[9] Hoecker A. *et al.*, `http://tmva.sourceforge.net` (2009).
[10] Ketchum W. *et al.*, CDF Public Note, 10643 (2011).
[11] Aaltonen T. *et al.*, *Phys. Rev. Lett. A*, **104** (2010) 101801.

---

(8) We expect the ZZ contribution to be negligible due to the requirement on $\not{E}_T$.