Explicit Jet Veto as a Tool to Purify the Underlying Event in the Drell-Yan Process Using CMS Open Data

Saeid Paktinat Mehdiabad
Faculty of Physics, Yazd University, P.O. Box 89195-741, Yazd, Iran
School of Particles and Accelerators,
Institute for Research in Fundamental Sciences (IPM), P.O.Box 19395-5531, Tehran, Iran

Ali Fahim
Department of Engineering Science,
College of Engineering, University of Tehran,
P.O. Box 11155-4563, Tehran, Iran
(Dated: February 1, 2022)

Abstract

The underlying event is an important part of high-energy collision events. In the event generators, the underlying event is tuned by fits to collision data. Usually, the underlying event observables are affected by the existence of extra jets and it is difficult to find a part of the phase space which is dominated by the underlying event. In this paper, we suggest to veto the jets in the considered region to disentangle these effects. The idea is verified to work on CMS Open Data. To our knowledge, it is the first time that such ideas are tested on real collision data.

Keywords: CMS Open Data, Underlying Event, Drell-Yan Process

*Electronic address: spaktinat@yazd.ac.ir
†Electronic address: a.fahim@ut.ac.ir
I. INTRODUCTION

In the hadron colliders, the physics process of our interest is usually a hard scattering of a single parton from another one. However, in each proton-proton (pp) collision, in addition to such primitive hard-scattering products, there are many particles originating either from the soft color interactions between the rest of the partons, or from additional hard scattering of other partons at the same pp collision. While the former parts are called beam remnants, the latter ones are known as the multiple partonic interactions (MPI). Both processes are called the underlying event (UE). Another intrinsic part of the hadron collisions, is the gluon radiation occurring either before the original hard scattering (initial state radiation (ISR)) or after that (final state radiation (FSR)). Particles coming from ISR and FSR mimic the UE activities. Hence, it is difficult to empirically distinguish UE from the ISR and FSR productions.

A precise measurement or observation of a new physics will not be possible without a good understanding of these processes in a hadron collider. Due to low transverse momentum ($p_T$) of their products, perturbative QCD may not accurately describe these processes. So, the practical understanding of UE is limited to the phenomenological models. The CDF [2–5] experiment at Tevatron and the ATLAS [6–11], ALICE [12] and CMS [13–17] experiments at LHC have carried out several measurements of parameters and observables sensitive to UE at their available center-of-mass energies ($\sqrt{s}$). Due to easy identification of Z-boson production, the experiments usually study the UE activities in the events containing the Drell-Yan (DY) process.

To find UE parameters, it is important to have a clean environment dominated by the UE activities. Usually, it is difficult to find this environment and extra jets affect the UE observables. For example, two of them which are investigated by the ATLAS experiment [10] are shown in Fig. 1. In a pure UE environment, the shown distributions must be the same for different Z-boson $p_T(p_T^Z)$, but obviously it is not the case here. Recently, it is suggested to use the event shape observables to find the UE dominated regions [18]. Here, we propose a new method as an alternative procedure to specify a cleaner UE environment. In the new method, the jets are vetoed explicitly from the region of interest. For the first time, to verify the performance of the methods, Open-Data [19] collected or simulated by the CMS experiment is used. The events contain a muon pair consistent with a Z-boson
FIG. 1: Distributions of the scalar transverse momentum sum density and charged particle multiplicity density. The figure illustrates the distributions in three different intervals of $p_T^Z$ [10].

production in pp collisions at $\sqrt{s} = 7$ TeV.

After the introduction, the paper is structured as follows. Section II describes the simulated and collision data used in this analysis. The method and the results are presented in section III. Then section IV concludes the paper.

II. CMS OPEN DATA

The LHC experiments have released part of their pp collision data and the related Monte Carlo simulations as a project known as the Open Data. In this analysis, we look at the events with two opposite charge muons in $\sqrt{s} = 7$ TeV taken by the CMS experiment in 2011. The datasets for collision data and simulated events are shown in Tab. II. Due to

TABLE I: The path of simulated and collision data in the CMS Open Data storage. The abbreviation and number of events analyzed in the paper are also shown.

| Dataset Path | Abbreviation | Number of Events |
|--------------|--------------|------------------|
| /MonteCarlo2011/Summer11LegDR... | | |
| /DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola... | | |
| /AODSIM/PUS13_START53_LV6-v1/XXXX | $\text{DYJets}$ | 10169936 |
| XXXX $\in \{00000, 010000, 010001, 010002\}$ | | |
| /Run2011A/DoubleMu/AOD/12Oct2013-v1/10000/ | $\text{DoubleMu}$ | 26320428 |
limitation of the resources, only part of the available data is used in the analysis. To read
the data files, the official software of CMS (CMSSW_5_3_32) is used as explained in the
guidelines provided by the experiment[22].

Every event is required to have at least one vertex, which is built from a fit to extrap-
olation of tracks to the nominal collision point. The fitted vertex must have more than 4
degrees of freedom. The distance of the vertex from the nominal collision point can be at
most, 24 cm in z-direction and 2 cm in ρ-direction, where the cylindrical coordinate has an
origin at the center of the detector and z-direction is parallel to the beams.

The objects used in the analysis are defined as follows:

• Muons: The muon candidates must pass isPFMuon and either isGlobalMuon or
  isTrackerMuon.

• Electrons: The electron candidates are required to pass a tight identification
  (electronID("eidRobustTight")).

• Jets: The jets are reconstructed using the anti-\(k_T\) algorithm with a distance param-
  eter 0.5. To clean the jets, the following cuts are applied: numberOfDaughters > 1,
  neutralHadronEnergyFraction < 0.99 and neutralEmEnergyFraction < 0.99. For
  jets in the tracker coverage (pseudorapidity |\(\eta\)| < 2.4) these extra conditions are also
  required: chargedEmEnergyFraction < 0.99, chargedHadronEnergyFraction > 0
  and chargedMultiplicity > 0.

• Tracks: Among all tracks, only the following tracks are selected. The measured values
  of ρ and z with respect to the vertex of the hardest muon over their uncertainty must be
  less than 3 and the uncertainty of \(p_T\) over its value must be less than 0.05. The former
  condition guarantees that only the tracks from the main vertex enter the analysis and
  the tracks from other soft collisions (pile up tracks) are rejected. The tracks have to
  pass also a HighPurity identification quality.

The selected leptons (electrons or muons) need also to have \(p_T > 25\) GeV and |\(\eta\)| < 2.5.
If the sum of energy deposit close to a lepton in the tracker (trackIso) and calorimeters
(caloIso) is less than 20% of lepton \(p_T\), it enters the analysis as an isolated lepton. The
selected jets pass the kinematic cuts of \(p_T > 20\) GeV and |\(\eta\)| < 2.5. The selected tracks have
the same cut on |\(\eta\)| with \(p_T > 0.5\) GeV.
For both \textit{DYJets} and \textit{DoubleMu} datasets, the following selections are applied.

- **DiMuons:** To select the Z boson candidates decaying to dimuon pair, every event is required to have at least one pair of opposite charge muons with an invariant mass close to the Z boson mass (65-115 GeV).

- **Electron veto:** To reject the backgrounds from diboson production, there should not be any isolated electron in the event. It is also useful to further reduce the size of the datasets.

- **Jet:** To have boosted Z bosons, at least one jet must be in the event.

For \textit{DoubleMu} sample, only parts of data, certified by the experiment, are used. For \textit{DYJets} sample, both generated charged particles and reconstructed tracks are used in the analysis. The former part can provide the generated information without the complexities introduced by the non perfect detectors.

### III. ANALYSIS AND RESULTS

The different regions of the azimuthal plane respect to the Z-boson flight direction could have different sensitivity to the UE activities. For the comparability, the same definition of the UE regions as in the ATLAS paper [10] (Fig. 2) is used.

![FIG. 2: Definition of UE regions in the azimuthal plane regarding the Z-boson direction [10].](image-url)
The charged tracks, depending on their azimuthal distance with respect to the Z-boson direction \(|\Delta \phi|\), are divided in the following three azimuthal UE regions:

- \(|\Delta \phi| < 60^\circ\), the **Toward region**, 
- \(60^\circ < |\Delta \phi| < 120^\circ\), the **Transverse region**, and 
- \(|\Delta \phi| > 120^\circ\), the **Away region**.

A considerable momentum of Z-boson could make this separation between the regions a lot more clear. However, such cases have only a small contribution to entire sample. Figure 3 (left) shows the distribution of \(p_T^Z\) for the generated and reconstructed events from \(DYJets\) sample and also the reconstructed events from \(DoubleMu\) sample. As the figure illustrates, for example the events with \(p_T^Z \geq 50\) GeV, make only few percent of the selected events. In Fig. 3 (right), the invariant mass of the selected dimuon is compared in reconstructed events from \(DYJets\) and \(DoubleMu\) samples. There is a good agreement between the distributions, showing a good description of the real data in the Monte Carlo simulated events.

In the analysis, the tracks corresponding to the Z-boson products are omitted. Since the hard scattering particles, which balance the transverse momentum of the Z-boson, are
mostly in the Away region, all remaining tracks in the toward region are dominantly from the UE activities. The charged track multiplicity \( N_{\text{ch}} \) and the scalar transverse momentum sum of the charged tracks \( \sum p_T \) are the main observables which are studied below. It is common to look at the density of these observables, \( N_{\text{ch}}/\delta\eta\delta\phi \) and \( \sum p_T/\delta\eta\delta\phi \), so \( N_{\text{ch}} \) and \( \sum p_T \) are divided by the angular area \( \delta\eta\delta\phi \). In this analysis, the observables are investigated only in the toward region, where the angular area is \( \delta\eta\delta\phi = 10\pi/3 \).

Figures 4 and 5 illustrate the distribution of \( \sum p_T \) and \( N_{\text{ch}} \) densities, respectively for the

toward region. Distributions are provided in four different \( p_T \) ranges, in both the generator level (left) and real collision data (right). The figures depict that in contrast to what is expected for a UE dominated region, distribution of observables are not independent of \( p_T \) ranges. Following the conclusion of Ref. [10], it seems still the toward region is contaminated by the tracks from additional jets. The distributions for the reconstructed events in \( \text{DYJets} \) sample also show the similar features.

A. Jet Veto

In Ref.[18], the event shapes are suggested as a tool to select a part of the phase space which is enriched by UE. It is not confirmed by our investigation on the available Open Data. To disentangle the effect of UE from the extra jets, we suggest to veto the jets in the
FIG. 5: Distributions of $N_{ch}$ density, in four different $p_T^2$ ranges in the toward region. Distributions are in the generator level (left) and real data (right).

toward region explicitly. The UE sensitive distributions are plotted again after this veto.

The $\sum p_T$ density is shown in Fig. 6 (left) for the generated events in DYJets sample.

FIG. 6: The $\sum p_T$ density in generated events from DYJets sample and reconstructed events from DoubleMu sample, when jet veto is applied. The expected UE behaviour is observed.

The figure should be compared to Figs. 1(left) and 4. Apart from the events with high $p_T^2$, which have a low statistics and fluctuations are dominant (See Fig. 3(left)), it is clear that changing the $p_T^2$ range does not affect the UE distribution as it is expected for a pure UE region. It is concluded that, this region is a UE dominated region and can be used safely to tune the Monte Carlo generators to reproduce the UE observables as correctly as
possible. The same distribution is investigated by the reconstructed events from DYJets and DoubleMu samples and the same result is concluded. The latter distribution is shown in Fig. 6 (right).

The $N_{ch}$ density is another UE sensitive observable which is studied in this analysis. Vetoing the jets from the toward region can suppress the effect of the extra jets in $N_{ch}$ density distribution also. The distribution for generated events in DYJets sample is shown in Fig. 7 (left). Comparing to Figs. 1(right) and 5, the expected behaviour for a UE dominated region is seen. The same features are seen for the distributions made by the reconstructed events from DYJets and DoubleMu samples. The latter distribution is shown in Fig. 7 (right). The different behaviour for high $p_T$ sample can be assigned to statistical fluctuations. The jet veto is not perfect, because the jet selection is not perfect. The jet identification and kinematic cuts, can introduce some bias in jet veto, but similar distributions for three $p_T^Z$ bins, below 50 GeV, can guarantee that the effect for high $p_T^Z$ bin is not due to the jet selection. It seems the low $p_T$ jets and misidentified or nonidentified jets, can not be the responsible of this effect.

It is suggested that the LHC experiments try this proposal and compare different MC tunes for UE sensitive observables. It should be emphasized that, since no comparison is done between reconstructed and generated events, unfolding is not necessary and it is not done in this analysis.

FIG. 7: The $N_{ch}$ density in generated events of DYJets sample and reconstructed events of DoubleMu sample, when jet veto is applied. The expected UE behaviour is observed.
IV. CONCLUSION

The underlying event as an important part of high-energy collision events, is tuned in the event generators by fits to collision data. It is important to find a part of the phase space which is dominated by UE. Different methods are proposed to find such environments. Usually, the UE observables are affected by the existence of extra jets. A new method is proposed to disentangle the effects of extra jets from UE. It is suggested to veto the extra jets in the toward region. For the first time, the CMS Open Data is used to investigate the performance of this method and an alternative method. After applying the jet veto, the UE sensitive observables show the expected behavior for a fully dominated UE environment. The result is verified in Monte Carlo simulated events (both generated and reconstructed objects) and also real collision data.

V. ACKNOWLEDGMENTS

The authors are grateful to the CMS and ATLAS collaborations for their fantastic results and releasing the high quality LHC data.

[1] T. Sjostrand and M. van Zijl, Phys. Lett. B188, 149 (1987).
[2] T. Affolder et al. (CDF), Phys. Rev. D65, 092002 (2002).
[3] R. D. Field (CDF), eConf C010630, P501 (2001), hep-ph/0201192.
[4] D. Acosta et al. (CDF), Phys. Rev. D70, 072002 (2004), hep-ex/0404004.
[5] T. Aaltonen et al. (CDF), Phys. Rev. D82, 034001 (2010), 1003.3146.
[6] G. Aad et al. (ATLAS), Phys. Rev. D83, 112001 (2011), 1012.0791.
[7] G. Aad et al. (ATLAS), Eur. Phys. J. C71, 1636 (2011), 1103.1816.
[8] G. Aad et al. (ATLAS), Phys. Rev. D86, 072004 (2012), 1208.0563.
[9] G. Aad et al. (ATLAS), Eur. Phys. J. C74, 2965 (2014), 1406.0392.
[10] G. Aad et al. (ATLAS), Eur. Phys. J. C74, 3195 (2014), 1409.3433.
[11] M. Aaboud et al. (ATLAS), JHEP 03, 157 (2017), 1701.05390.
[12] B. Abelev et al. (ALICE), JHEP 07, 116 (2012), 1112.2082.
[13] V. Khachatryan et al. (CMS), Eur. Phys. J. C70, 555 (2010), 1006.2083.
[14] S. Chatrchyan et al. (CMS), JHEP 09, 109 (2011), 1107.0330.
[15] S. Chatrchyan et al. (CMS), Eur. Phys. J. C72, 2080 (2012), 1204.1411.
[16] S. Chatrchyan et al. (CMS), Eur. Phys. J. C73, 2674 (2013), 1310.4554.
[17] A. M. Sirunyan et al. (CMS), JHEP 07, 032 (2018), 1711.04299.
[18] D. Kar and D. S. Rafanoharana, in 9th International Workshop on Multiple Partonic Interactions at the LHC (MPI@LHC 2017) Shimla, India, December 11-15, 2017 (2018), 1801.05218.
[19] T. McCauley (ATLAS, CMS), PoS ICHEP2016, 335 (2016).
[20] G. Aad et al. (ATLAS), Eur. Phys. J. C76, 375 (2016), 1602.08980.
[21] S. Chatrchyan et al. (CMS), Phys. Lett. B722, 238 (2013), 1301.1646.
[22] http://opendata.cern.ch