

# UC Berkeley

## UC Berkeley Previously Published Works

**Title**

Search strategy using LHC pileup interactions as a zero bias sample

**Permalink**

<https://escholarship.org/uc/item/81h29768>

**Journal**

Physical Review D, 97(9)

**ISSN**

2470-0010

**Authors**

Nachman, Benjamin

Rubbo, Francesco

**Publication Date**

2018-05-01

**DOI**

10.1103/physrevd.97.092002

Peer reviewed

# Search strategy using LHC pileup interactions as a zero bias sample

Benjamin Nachman\*

*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*

Francesco Rubbo†

*SLAC National Accelerator Laboratory, 575 San Hill Road, Menlo Park, California 94025, USA*



(Received 24 March 2017; revised manuscript received 9 June 2017; published 17 May 2018)

Due to a limited bandwidth and a large proton-proton interaction cross section relative to the rate of interesting physics processes, most events produced at the Large Hadron Collider (LHC) are discarded in real time. A sophisticated trigger system must quickly decide which events should be kept and is very efficient for a broad range of processes. However, there are many processes that cannot be accommodated by this trigger system. Furthermore, there may be models of physics beyond the standard model (BSM) constructed after data taking that could have been triggered, but no trigger was implemented at run time. Both of these cases can be covered by exploiting pileup interactions as an effective zero bias sample. At the end of high-luminosity LHC operations, this zero bias dataset will have accumulated about  $1 \text{ fb}^{-1}$  of data from which a bottom line cross section limit of  $\mathcal{O}(1) \text{ fb}$  can be set for BSM models already in the literature and those yet to come.

DOI: [10.1103/PhysRevD.97.092002](https://doi.org/10.1103/PhysRevD.97.092002)

## I. INTRODUCTION

At a proton-proton ( $pp$ ) collider like the Large Hadron Collider (LHC), interesting events are rare. Unlike electron-positron colliders, the partonic center-of-mass energy  $\sqrt{\hat{s}}$  follows a broad distribution set by parton distribution functions (PDF). As such, the total inelastic cross section ( $\mathcal{O}(100) \text{ mb}$  [1,2] at  $\sqrt{s} = 13 \text{ TeV}$ ) is many orders of magnitude above the production cross section for electro-weak scale particles, such as the  $W$  boson ( $\mathcal{O}(10) \text{ nb}$  [3,4] at  $\sqrt{s} = 13 \text{ TeV}$ ). In order to increase the rate of interesting events as much as possible, the LHC is operated at very high luminosities. Proton bunches collide every 25 ns and the bunch density is such that multiple  $pp$  interactions (pileup) occur in each bunch crossing. Due to limited readout and disk space capabilities, it is not possible to fully record every bunch collision. Therefore, both the ATLAS and CMS experiments have developed strategies to trigger on events of interest. Trigger systems are implemented at multiple levels, with ultrafast but simple algorithms in hardware (L1) and increasingly complex algorithms in software at higher levels (HLT), where a more detailed

readout of the detectors is exploited. An event is fully recorded only if it satisfies the selection criteria at all levels of the trigger. The L1 triggers decrease the 40 MHz rate down to  $\mathcal{O}(100) \text{ kHz}$ , which is further reduced to  $\mathcal{O}(1) \text{ kHz}$  after the HLT triggers. In order to achieve these reductions, triggers targeting processes with a very high cross section are prescaled: events are randomly discarded so that only a fraction  $1/p$  ( $p = \text{prescale}$ ) are recorded. For example, at  $\sqrt{s} = 8 \text{ TeV}$ , the ATLAS single jet triggers targeting events with at least one jet with a minimum transverse momentum ( $p_T$ ) between 50 and 100 GeV had prescales of  $p \sim 10^4$  [5]. The lowest unprescaled ( $p = 1$ ) single jet trigger requires a minimum transverse momentum  $p_T \sim 500 \text{ GeV}$ . The prescales for low  $p_T$  processes, such as inclusive jet production, are increased at least linearly with the instantaneous luminosity, in order to keep the rate constant.

While the existing trigger system is very effective at identifying high- $p_T$  objects, there are a plethora of viable models of physics beyond the standard model (BSM) that are not well covered. One broad class of models predicts exotic signatures involving isolated charged particle tracks. Pattern recognition for track reconstruction in the ATLAS and CMS inner detectors is computationally expensive and it only runs in a limited way at the HLT (and possibly at L1 in the future). Tracks with kinks, displaced vertices, high  $dE/dx$ , anomalous timing, intermittent hits, and exotic curvature will not be efficiently reconstructed by L1 tracking and are also not easily (or at all) covered in the HLT (see Ref. [6] for a review). For example, oscillating

\*bnachman@cern.ch

†rubbo@cern.ch

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

pairs of tracks from new strong dynamics [7–9] require dedicated reconstruction algorithms. In addition to models with low multiplicity tracks, BSM processes that predict extreme multiplicities of low energy particles [9–11] are striking signatures that may be largely uncovered by existing or even possible triggering techniques. Another broad class of models predicts exotic structure inside hadronic jets. This includes jets with many displaced vertices [12] as well as jets with large invisible components [13]. There are likely many other models that have yet to be proposed in the literature that would leave extraordinary detector signatures but too exotic to be captured by standard trigger schemes.

All of the models discussed so far have the property that their signature is so exotic that there is likely not a significant background rate from the standard model. The current triggering scheme also limits the sensitivity to models with large SM background, such that large prescales are required. This includes low mass dijet resonances, such as a leptophobic  $Z'$  [14].

There are two existing strategies for recovering model coverage that would otherwise have been lost by the trigger. One strategy is to look for a target process produced in association with another very energetic object that can be used for triggering. For example, a low mass  $Z'$  that decays into jets can be produced in association with a high  $p_T$  photon [15] or jet [16] from initial state radiation (ISR). However, this strategy introduces a large effective prescale due to a reduction in the cross section. In addition, this procedure cannot be used to measure new or standard model processes differentially in their low  $p_T$  phase space. Another powerful strategy, referred to as data-scouting or trigger-level analysis, stores only a smaller relevant fraction of the detector information for the events selected by the L1 trigger [17,18]. These trigger-level analyses are not impacted by the prescales of the HLT triggers, but are limited by the L1 prescales that are often tighter. As only a small fraction of the detector information can be recorded at the L1 accept rate, only the specific final states for which the trigger-level analysis strategy has been designed are accessible.

The new strategy presented in this paper uses each individual pileup interaction for physics analysis. All of these interactions potentially contain interesting physics processes and are recorded by the detector along with the primary interaction that satisfied any arbitrary trigger. Every event passing any trigger can be used for the purpose of studying the pileup interactions, which are recorded with nearly no selection bias. The effective prescale associated to this zero bias sample (ZBS) is inversely proportional to the overall trigger bandwidth. For a sufficiently high bandwidth, this effective prescale can be lower than the one from the ISR strategy. In addition, a trigger-level analysis can be combined with this strategy, thus enhancing even further the physics reach by enabling access to a large

quantity of otherwise unused data. For analysis offline, complex reconstruction algorithms can run without real-time constraints on algorithm speed that would have been required to save the event using a targeted trigger online.

## II. THE ZERO BIAS SAMPLE

Reconstructed tracks from charged particles are the most important handle for identifying pileup interactions. Individual collision vertices are built from tracks and various objects can be associated to these vertices through their associated tracks. For example, the jet vertex tagger (JVT) used in ATLAS is 90% efficient at associating a jet with  $20 < p_T < 50$  GeV to its correct vertex while mis-identifying stochastic or QCD jets from other vertices 1% of the time [19]. Ignoring the small detector inefficiencies and fake rates, the effective luminosity from the pileup collisions collected from a trigger system with bandwidth  $w$  is given by

$$\int \mathcal{L}(\text{ZBS}) dt = \frac{w}{40 \text{ MHz}} \times \int \mathcal{L} dt. \quad (1)$$

The last term in Eq. (1) is the integrated luminosity collected with standard triggers. One way to derive this equation is to consider the number of events recorded by the LHC experiments. Suppose there is a process  $X$  with a cross section  $\sigma_X$ . If  $\Sigma$  is the total inelastic cross section, then a fraction  $\sigma_X/\Sigma$  (on average) of all  $pp$  collisions recorded will contain the process  $X$ . If the bandwidth is  $w$  and assuming that the production of  $X$  does not significantly influence this rate, then the total number of  $X$  events will be  $\langle \mu \rangle \times w \times T \times \sigma_X/\Sigma$ , where  $T$  is the amount of time the LHC is operated for  $pp$  collisions and  $\langle \mu \rangle$  is the average number of  $pp$  collisions per bunch crossing. Now let  $Y$  be any standard model process that can be triggered with 100% efficiency and has a prescale of 1. Analogously to the calculation for  $X$ , the number of  $Y$  events recorded will be  $\langle \mu \rangle \times H \times T \times \sigma_Y/\Sigma$ , where  $H$  is the total rate of bunch crossings (40 MHz at the LHC). This shows that  $\int \mathcal{L} dt = \langle \mu \rangle \times H \times T/\Sigma$  and solving the system of equations to derive the effective luminosity for the process  $X$  results in Eq. (1). The bandwidth  $w$  can go from the 100–500 Hz typical of Run 1 (2010–2012) data-taking to the upper limit of the expected L1 trigger bandwidth for the High Luminosity LHC (HL-LHC) experiments [20,21]. Table I shows the ZBS integrated luminosity for various LHC running conditions in Run 1 and projected for Run 2 (2015–2018) and beyond. The effective prescale for the best HL-LHC data acquisition scenarios (last row in Table I) is between 4000 and 6000. For trigger-level analyses of the ZBS dataset at the HL-LHC, the effective prescale is between 50 and 100. This means that if a particular signature has a primary trigger efficiency that is less than 0.02–0.03% offline or 1–2% at L1, then the ZBS data set will record more signal events. Note that the predicted

TABLE I. The ZBS integrated luminosity for various LHC running conditions in Run 1 and projected for Run 2 and beyond. The last column shows the projected luminosity when performing a trigger-level analysis, before the application of any software prescales.

LHC Run	Total Lumi. [1/fb]	$\langle\mu\rangle$	L1 Rate [MHz]	HLT Rate [kHz]	ZBS [1/fb]	ZBS @ HLT [1/fb]
1	20	20	0.1	0.1	$5 \times 10^{-5}$	0.05
2 + 3	300	80	0.1	1	$7.5 \times 10^{-3}$	0.75
4 + 5 (ATLAS)	3000	200	0.4	10	0.75	30
4 + 5 (CMS)	3000	200	0.75	7.5	0.56	56.3

integrated luminosity during Runs 2 and 3 of the LHC is about  $300 \text{ fb}^{-1}$  while the final HL-LHC data is expected to be about  $3 \text{ ab}^{-1}$ .

### III. CROSS SECTION PROJECTIONS

Figure 1 presents the 95% confidence level cross section limits with the  $\text{CL}_s$  procedure [22], covering a broad class of models, for the ZBS throughout the lifetime of the (HL-)LHC. Limits are computed assuming various signal-to-background ratios. For a zero-background search, at least three signal events are needed to reach the 95% confidence limit.

The offline version of the ZBS can be analyzed at any time, including (well) after the full HL-LHC program has ended, while the trigger-level (online) version requires some level of analysis to be implemented in the software

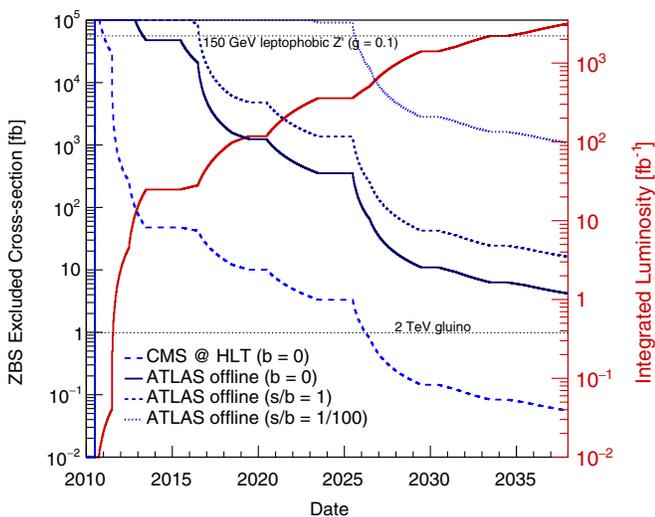


FIG. 1. The excluded cross section and the integrated luminosity are shown throughout the lifetime of the (HL-)LHC. Statistical uncertainties are assumed to dominate over systematic ones in this low background regime. For reference, the cross section for a 2 TeV gluino and a 150 GeV leptophobic  $Z'$  with  $g = 0.1$  are also shown (see text for details).

trigger at run time. For a model that predicts an extraordinary signature with nearly no SM background (see Sec. I), a ZBS analysis will be able to set a cross section limit of nearly 3 fb. For reference, this is the cross section of a 1.8 TeV gluino [23], a 1.1 TeV stop [23], a 1.4 TeV fermionic top quark partner [24], and a 1.7 TeV colored quirk with infracolor representation size 5 [9]. Cross-section limits for the ZBS applied at HLT are nearly a factor of 100 stronger, though would require dedicated algorithms to be put in place before the start of the HL-LHC to achieve the full potential. Figure 1 also shows the limits for searches that have a nontrivial background component. While many of the models described in the introduction are nearly background-free, some may have contributions from low probability events from the standard model. With a signal-to-background ratio of 0.001, a cross section limit of about 1 pb can be set.

The important low mass dijet resonance search is a concrete illustration of the power of the ZBS. Every hadron collider has searches for dijet resonances, which are predicted in a wide variety of BSM models. To start, consider a case with a high signal-to-background ratio. For example, suppose the dark matter consists of an extended sector of quark- and gluon-like objects and a confining QCD-like  $\text{SU}(3)$ -symmetry as in e.g. [12]. In certain regions of the model parameter space, such a model would give rise to emerging jets wherein jets are formed in the dark sector and then after some time, the dark quark and gluon fragmentation products decay into SM particles. Suppose there is a leptophobic  $Z'$  that connects the visible sector QCD with the dark sector QCD. As a minimal but complex realization of this model [25], the dark sector is a nearly exact copy of the SM QCD where, for simplicity, there is only one hadron called the dark  $\rho$ . There are two relevant free parameters of this model for studying various strategies: the mass of the  $Z'$  and the lifetime of the dark hadrons  $c\tau$ , which is set by the couplings of the  $Z'$ . Each dark  $\rho$  resulting from dark quark or gluon fragmentation could result in a displaced vertex. Figure 2 shows the efficiency of various identification algorithms as a function of  $Z'$  mass for various lifetimes. The following paragraph explains and compares the three schemes, where the first (“pixel”) scheme uses the ZBS while the other two use traditional triggering strategies.

The first possibility is to reconstruct displaced vertices in the pixel systems of the ATLAS or CMS detectors, labeled [30] pixel in Fig. 2. It is not possible to accurately estimate the background from simulation, but based on Refs. [31–33], it seems conservative that requiring  $\geq 4$  displaced vertices is near the zero-background regime. The maximum efficiency is for  $c\tau \sim 5 \text{ mm}$  and is  $\gtrsim 50\%$  across the entire mass range. The ZBS may be a powerful tool to target these models because standard approaches are not powerful: there is usually not enough  $E_T^{\text{miss}}$  or  $H_T$  to trigger on, as in the SUSY cases studied in Refs. [31–33]. Instead, if the dark mesons decay in the muon chambers of ATLAS

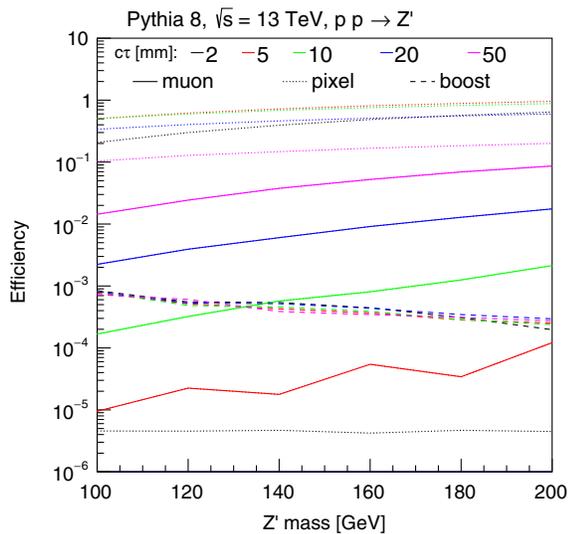


FIG. 2. The efficiency for three analysis strategies for identifying  $Z'$  decays into dark quarks that subsequently decay into SM mesons. The dotted line for the 2 mm muons indicates an upper bound based on nearly one million events per mass point. See the text for further details.

or CMS, then the muon trigger could be used to identify events [34,35], indicated by [36] “muon” in Fig. 2. A third strategy is to identify cases in which the  $Z'$  is sufficiently boosted that all of its decay products are captured inside a single jet. In that case, one could look for a bump in the single jet mass distribution [16], indicated by [37] “boosted” in Fig. 2. There is a large effective prescale from requiring the  $Z'$  to be boosted.

The prescale for the offline ZBS for ATLAS is 4000 (Table I); therefore, the efficiency for the ZBS and the displaced vertex approach is about  $50\%/(4000) = 10^{-4}$  for  $m_{Z'} = 200$  GeV and  $c\tau \sim 5$  mm. The muon trigger has a similar or slightly lower efficiency when  $c\tau$  is  $\lesssim 5$  mm. Both projections have some approximations and to know precisely which is better would require a detailed detector and background simulation. These studies indicate that both methods would result in similar sensitivity and therefore the ZBS strategy is worth pursuing further. This is especially true if one can implement region of interest secondary vertex reconstruction at HLT. The boosted strategy is likely too inefficient, as the necessary condition of the opening angle for the decay is already at the  $10^{-4}$  level at  $m_{Z'} = 200$  GeV.

In addition to low background searches for exotic signatures, the ZBS can also be competitive with searches for SM-like final states that exploit associated production. To illustrate this case, consider the traditional low-mass dijet resonance search. The SM dijet cross section is so large that searching for bumps in the dijet invariant mass ( $m_{jj}$ ) spectrum is plagued by large prescales at low  $m_{jj}$ . Such searches are therefore performed with trigger level and ISR analyses. Figure 3 shows the effective ZBS prescale [38] compared with the effective prescale due to the reduction in

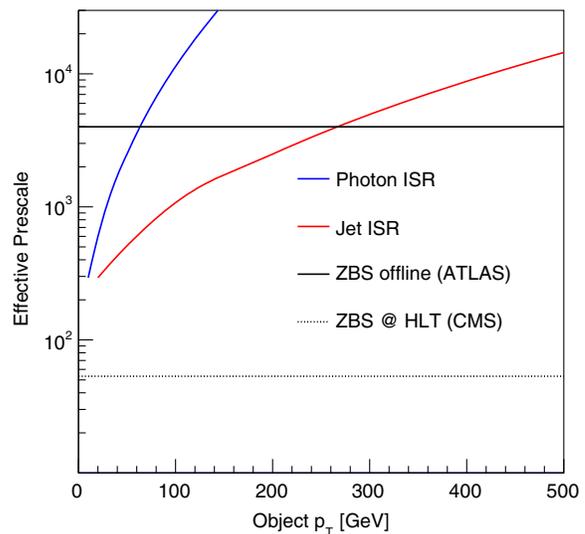


FIG. 3. A comparison of the effective prescale for the ZBS and ISR analyses for a particular leptophobic  $Z'$  model [14]. The ZBS prescales are the same as in Table I.

cross section when requiring the  $Z'$  to be produced in association with a ISR high  $p_T$  photon or jet. By construction, the prescale is independent of  $p_T$  for the ZBS, but grows quickly with the photon or jet  $p_T$  for the ISR analyses. Typical minimum requirements for unprescaled single photon and jet triggers are  $p_T = 100$  and  $p_T = 400$  GeV respectively. At these values the effective prescale is significantly larger than the one expected for the ZBS. The photon and jet thresholds can be lowered when combining the ISR technique with data scouting. However, when the ZBS is combined with data scouting (at the HL-LHC), the solid line in Fig. 3 becomes the dashed one. Therefore, even at trigger level, the ZBS analysis has a lower effective prescale than ISR techniques.

The ZBS is superior to the ISR technique in terms of prescale (at the HL-LHC), but a fair comparison also requires an assessment of the signal-to-background ratio. A loss in events from the effective prescale from an ISR requirement can be partially compensated by better discrimination power. To estimate the approximate sensitivity to a dijet resonance, a benchmark  $Z'$  model [14] is simulated with MG5\_aMC 2.1.1 [39] interfaced with PYTHIA 8.170 [40]. To simulate the detector response, the jet momenta are smeared according to  $\sigma(p_T)/p_T = 1.3/\sqrt{p_T/\text{GeV}}$ . The jet resolution depends on the pileup conditions and is in general worse at trigger level than for fully reconstructed offline jets. Therefore, the resolution function is conservatively chosen to be worse than the typical energy resolution at LHC experiments in Run 2. Events are required to have two jets with  $p_T > 25$  GeV, and the two leading such jets are used to compute the dijet invariant mass,  $m_{jj}$ . More sophisticated approaches could better exploit events with significant initial or final state radiation for an enhanced sensitivity but are beyond the scope of this paper. A simple binned  $\chi^2$

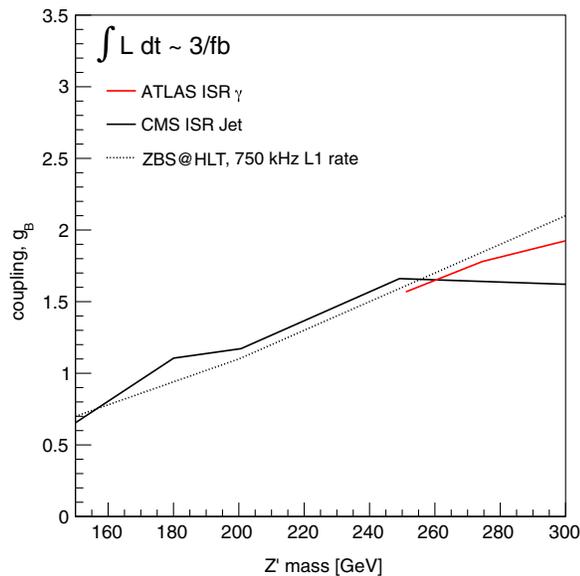


FIG. 4. The coupling limits as a function of the  $Z'$  mass for ISR jet/photon analyses using the 2015 LHC run datasets compared with estimates for the ZBS using a dataset of the same size as collected at the HL-LHC (with a significantly higher bandwidth). Existing limits are from Refs. [15,16].

analysis of the dijet invariant mass spectrum in a window around the target  $Z'$  mass is performed, using toy Monte Carlo simulation to estimate the  $p$ -value. A given mass point is declared excluded if the corresponding  $p$ -value is less than 0.05. As a validation of this procedure, the coupling upper limit is estimated for a 500 GeV  $Z'$  with  $20.3 \text{ fb}^{-1}$  of unrescaled single jet trigger simulated data at  $\sqrt{s} = 8 \text{ TeV}$ . The limit obtained, approximately 1.5, is consistent with the Run 1 ATLAS result [41].

Figure 4 shows a comparison between published ISR limits and our estimate of the ZBS @ HLT based on the standard search for a peak in the  $m_{jj}$  distribution. The ISR result will slowly improve with more integrated luminosity so for a fair comparison, both strategies are evaluated with a data set size corresponding to the 2015 run. The ZBS results are estimated assuming that the same data set is recorded at HL-LHC rates so a relative comparison between strategies near their peak performance is possible. With this setup, the limits are found to be comparable. Given the complementarity of the two strategies, further gain can be achieved by combining the results. Note that a conservative estimate of the jet resolution at the HL-LHC is used for the ZBS analysis (which is also highly simplified). It is likely that the limits shown here are therefore conservative for the ZBS (also supported by Fig. 3).

#### IV. IMPLEMENTATION CHALLENGES

While the ZBS data set holds great promise, using the data in practice will be technically challenging. The first challenge is to access data from all triggers. This is both a challenge for

a ZBS analysis online and offline. Online, the problem is that most HLT items are tied to a single L1 trigger. ZBS analyses working at the trigger level would need to have access to all events that pass L1, which is a bandwidth challenge. Offline, the problem is that the data are separated into streams and most analyses require a single trigger in order to reduce the data volume. Overcoming these challenges would require more sophisticated bookkeeping algorithms.

The second challenge is having access to all of the pileup information. The time to perform track reconstruction does not scale well with  $\mu$  and so current fast algorithms operating at the trigger level reduce the bandwidth by explicitly ignoring pileup and displaced tracks as early as possible in the analysis chain (see e.g. Ref. [42]). Special and possibly time consuming track reconstruction may be required for some analyses that are looking for exotic track signatures in the ZBS. This is a more serious challenge online, where information about the pileup collisions may even be discarded before it can ever be analyzed. For example, the current trigger-level analysis procedure (to reduce the data rate) is to save only what is needed for the final analysis selection for offline processing. If after the data are collected, there is an idea that can be studied with the ZBS, it will not be possible to look at other information in these data. Therefore, analyses that use the ZBS at HLT will need to be designed prior to data taking. Even if they are designed ahead of time, accessing and reconstructing all of the pileup information at the HLT will be a significant technical challenge. Both ATLAS and CMS have ambitious goals for fast event reconstruction and it seems possible, though with much effort, to make the ZBS analysis at HLT work.

Another technical challenge is that most algorithms for event reconstruction are designed for a single primary vertex; this would need to be generalized to handle any vertex. For example, the fraction of track energy from the hard scatter collision that is used to identify pileup jets needs to be recomputable with respect to any vertex labeled as hard scatter. Some of these tools already exist, such as the reassignment of the hard-scatter vertex in the  $H \rightarrow \gamma\gamma$  measurement in ATLAS that can use calorimeter pointing information to identify the correct vertex [43]. The reconstruction resolution will always be a challenge at high  $\mu$ , but a jet with a given  $p_T$  from a pileup collision will have the same resolution as a jet with the same  $p_T$  from the collision that triggered the event. Therefore, this is a challenge that the primary trigger analyses will also face.

None of these challenges exclude the possibility of a ZBS analysis, but they do show that while the data are will be produced “for free,” significant effort will be required to ensure they are collectable and analyzable.

#### V. CONCLUSIONS

The multiple pileup interactions produced in LHC collisions yield unbiased data which can be used to probe physics processes otherwise inaccessible or with limited

acceptance. The effective prescale for this zero bias data set is about 40 000 in Runs 2 + 3, and drops to 400 for trigger-level analyses. If the trigger efficiency for any search is lower than this amount, then the ZBS may be more powerful. In particular, for exotic signatures that are nearly impossible to trigger on due to bandwidth and time constraints in the trigger, the ZBS may be the best strategy. This was illustrated explicitly for a Hidden Valley model with a  $Z'$  and dark sector QCD, where the ZBS data set has a sensitivity that is likely comparable to or better than existing trigger strategies. Of course, the ZBS idea will apply to models that have not yet appeared in the literature. For existing models, one can fully exploit the ZBS by implementing selections in the software trigger to set the most stringent limits despite failing a direct hardware trigger. As discussed in the previous section, using the ZBS data would be technically challenging for both the online and offline versions. The studies and examples presented in earlier sections show that these costs in time and effort are worth serious consideration.

In addition to setting a bottom line for searches with a low background rate, the ZBS may also be competitive with traditional searches that exploit associative production to pass the trigger. Requiring a second object, such as a high- $p_T$  ISR jet or photon introduces a large effective prescale that can be harsher than the prescale from the ZBS. When combined with a trigger-level analysis, the ZBS is expected to provide comparable limits to the ISR technique in the case of the low mass dijet resonance search. These estimates are based on a simplified model of the dijet

resonance searches and could be improved with additional sophistication. The simple model ignores the efficiency for reconstructing primary vertices and any inefficiencies in associating jets to these vertices. For the relatively high masses targeted by the example, these inefficiencies are relatively small. However, for lower mass measurements and searches, these inefficiencies may be important. Both ATLAS and CMS have ongoing studies to improve the tracking performance in high pileup environments, including the use of timing information to distinguish objects from spatially overlapping vertices. All of these interesting developments will be important for the ZBS strategy.

After the full LHC program, the ZBS will have accumulated about  $0.5\text{--}1\text{ fb}^{-1}$  of fully unbiased  $pp$  collision data that would not have been analyzed. We have shown that the novel concept of analyzing all pileup interactions enhances the physics reach of the LHC experiments and could constitute a useful strategy to fully exploit the HL-LHC data set.

## ACKNOWLEDGMENTS

We would like to thank Felix Yu for help simulating the  $Z'$  benchmark model and Nhan Tran for useful discussions about the ZBS. In addition, we are grateful to John Alison for helping us compute the correct luminosity for the ZBS. This work is supported by U.S. Department of Energy (DOE) under Contracts No. DE-AC02-05CH11231 and No. DE-AC02-76SF00515.

- 
- [1] ATLAS Collaboration, *Phys. Rev. Lett.* **117**, 182002 (2016).
  - [2] CMS Collaboration, Report No. CMS-PAS-FSQ-15-005, 2016.
  - [3] ATLAS Collaboration, *Phys. Lett. B* **759**, 601 (2016).
  - [4] CMS Collaboration, Report No. CMS-PAS-SMP-15-004, 2015.
  - [5] ATLAS Collaboration, *Phys. Rev. D* **93**, 052003 (2016).
  - [6] P. Meade, M. Papucci, and T. Volansky, *Phys. Rev. Lett.* **109**, 031801 (2012).
  - [7] L. B. Okun, *Pis'ma Zh. Eksp. Teor. Fiz.* **31**, 156 (1979) [*JETP Lett.* **31**, 144 (1980)].
  - [8] L. B. Okun, *Nucl. Phys.* **B173**, 1 (1980).
  - [9] J. Kang and M. A. Luty, *J. High Energy Phys.* **11** (2009) 065.
  - [10] S. Knapen, S. Pagan Griso, M. Papucci, and D. J. Robinson, *J. High Energy Phys.* **08** (2017) 076.
  - [11] R. Harnik and T. Wizansky, *Phys. Rev. D* **80**, 075015 (2009).
  - [12] P. Schwaller, D. Stolarski, and A. Weiler, *J. High Energy Phys.* **05** (2015) 059.
  - [13] T. Cohen, M. Lisanti, and H. K. Lou, *Phys. Rev. Lett.* **115**, 171804 (2015).
  - [14] B. A. Dobrescu and F. Yu, *Phys. Rev. D* **88**, 035021 (2013).
  - [15] ATLAS Collaboration, Report No. ATLAS-CONF-2016-029, 2016.
  - [16] CMS Collaboration, *Phys. Rev. Lett.* **119**, 111802 (2017).
  - [17] CMS Collaboration, *Phys. Rev. Lett.* **117**, 031802 (2016).
  - [18] ATLAS Collaboration, Report No. ATLAS-CONF-2016-030, 2016.
  - [19] ATLAS Collaboration, *Eur. Phys. J. C* **76**, 581 (2016).
  - [20] ATLAS Collaboration, Report No. CERN-LHCC-2015-020, 2015; URL <https://cds.cern.ch/record/2055248>.
  - [21] CMS Collaboration, Report No. CERN-LHCC-2015-019, 2015; URL <https://cds.cern.ch/record/2055167>.
  - [22] A. L. Read, *J. Phys. G* **28**, 2693 (2002).
  - [23] C. Borschensky, M. Kramer, A. Kulesza, M. Mangano, S. Padhi, T. Plehn, and X. Portell, *Eur. Phys. J. C* **74**, 3174 (2014).
  - [24] M. Czakon and A. Mitov, *Comput. Phys. Commun.* **185**, 2930 (2014).
  - [25] Pythia 8.2 [26] includes a Hidden Valley [27] module that even incorporates the running of the dark  $\alpha_s$  [28,29].

However, when the dark quarks ( $q\bar{v}$ ) are very light (SM quark-like) then the decay and hadronization can be eliminated. As an alternative, two instances of Pythia 8 are used to generate  $Z'$  events, one with  $Z' \rightarrow qq'$  and one with  $Z' \rightarrow \nu\bar{\nu}$ . The first instance has modified SM parameters. The hadrons resulting from the modified SM case (possible since the  $Z'$  is a color-singlet) are identified and then placed in an unmodified event but with the vector sum boosted to match the  $Z' \rightarrow \nu\bar{\nu}$  four-vector. In the modified setup, there are no baryons, quarks do not radiate photons, electromagnetically decaying hadrons are stable, and the dark pions have the same mass as the SM  $\rho$  meson so that they can decay into two SM pions. The modified parameters are `TimeShower:QEDshowerByQ=false`, `StringFlav:decupletSup=0`, `StringFlav:probStoUD=0`, `StringFlav:probQtoQ=0`, `StringFlav:etaSup=0`, `StringFlav:etaPrimeSup=0`, `211:m0:0.775`, `111:m0:0.775`, `111:mayDecay=false`, `223:mayDecay=false`, `221:mayDecay=false`.

- [26] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015).
- [27] M. J. Strassler and K. M. Zurek, *Phys. Lett. B* **651**, 374 (2007).
- [28] L. Carloni and T. Sjostrand, *J. High Energy Phys.* **09** (2010) 105.
- [29] L. Carloni, J. Rathsmann, and T. Sjostrand, *J. High Energy Phys.* **04** (2011) 091.
- [30] For this case, the decay is required to occur between 30 and 100 mm. No material veto is applied, which in practice will lower the efficiency.
- [31] ATLAS Collaboration, *Phys. Rev. D* **92**, 072004 (2015).
- [32] CMS Collaboration, *Phys. Rev. D* **91**, 012007 (2015).
- [33] ATLAS Collaboration, *Phys. Rev. D* **97**, 052012 (2018).
- [34] ATLAS Collaboration, *Phys. Rev. D* **92**, 012010 (2015).
- [35] ATLAS Collaboration, *J. Instrum.* **8**, P07015 (2013).
- [36] For concreteness, this case corresponds to any decay that happens within the ATLAS muon spectrometer (MS) acceptance, between 4 and 7 meters (similar results hold for CMS, where the MS is closer to the interaction point). Pions from the dark  $\rho$  decay must be at least 10 GeV [35]. The actual trigger efficiency will be much lower, because the particles.
- [37] In particular, this case is defined by  $2m^{Z'}/p_T^{Z'} \leq 0.8$ . Further selection criteria on the number of prongs in the jet will further reduce the efficiency and even with strict criteria, the search does not approach zero-background because there is a large background from 'normal' QCD jets. The power of this strategy is further reduced for  $Z'$  with large width (in Fig. 2, the width is assumed to be negligible).
- [38] Prescales for the ISR analyses are computed with MG5\_aMC 2.1.1 [39] by comparing the cross section for  $p\bar{p} > Z' j$  or  $p\bar{p} > Z' a$  to  $p\bar{p} > Z'$  and do not include any further prescale obtained at L1 or HLT.
- [39] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [40] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178**, 852 (2008).
- [41] ATLAS Collaboration, *Phys. Rev. D* **91**, 052007 (2015).
- [42] ATLAS Collaboration, Tech. Report No. CERN-LHCC-2013-007, Report No. ATLAS-TDR-021, 2013, aTLAS Fast Tracker Technical Design Report, URL <https://cds.cern.ch/record/1552953>.
- [43] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **716**, 1 (2012).